scientific reports

Check for updates

OPEN

Identification and biophysical characterization of *Plasmodium* peptide binding by common African HLAs

Marielle B. Frooman[1,2], Klara Choi[1,2], Maya Z. Kahn[1,2], Li-Yen Yang[1], Aubrielle Cunningham[1], Jenna M. RisCassi[1] & Andrew C. McShan[1]✉

Human Leukocyte Antigens (HLA) are immunoreceptors that present peptide antigens at the cell surface to T cells as a primary mechanism of immune surveillance. Malaria, a disease associated with the *Plasmodium* parasite, claims > 600,000 lives per year globally with most deaths occurring in Africa. Development of efficacious prophylactic vaccines or therapeutic treatments for malaria has been hindered by the lack of a basic understanding of the role of HLA-mediated peptide antigen presentation during *Plasmodium* infection. In particular, there is (i) little understanding of which peptide antigens are presented by HLAs in the context of malaria, and (ii) a lack of structural insights into *Plasmodium* peptide antigen presentation by HLAs, which underpins peptide/HLA stability, specificity, cross-presentation across HLA alleles, and recognition by T cell receptors. To begin to address these knowledge gaps, we identify and characterize candidate peptide antigens derived from *Plasmodium falciparum* with potential for presentation by common class I HLA alleles. We computationally screen nine proteins from the *P. falciparum* proteome to predict eight peptides with potential for cross-presentation by common alleles in African populations, HLA-A*02:01 and HLA-B*08:01. We then validate the predictions by producing recombinant HLAs in complex with the eight identified peptides by in vitro refolding. We evaluate the folding and thermal stability of the resulting sixteen peptide/HLA complexes by CD spectroscopy and nanoDSF. In silico modeling of peptide/HLA complexes informs a plausible structural basis for mechanisms for cross-presentation of *P. falciparum* peptides across HLA-A*02:01 and HLA-B*08:01 alleles. Finally, we expand our identified *P. falciparum* peptides to cover a broader range of HLA alleles in malaria endemic populations with experimental validation provided for HLA-C*07:01 and HLA-E*01:03. Together, our results are a step forward towards a deeper understanding of the potential for multi-allele cross-presentation of peptides in malaria. These results further inform future development of multivalent vaccine strategies targeting HLA profiles in malaria endemic populations.

Class I Human Leukocyte Antigen (HLA) immunoreceptors play essential roles in adaptive immunity by presenting self and foreign peptide antigens on the surface of nearly all nucleated cells, which enables cytotoxic $CD8^+$ T cells to surveil cellular health via the proteome[1]. Recognition of pathogen derived peptide antigen/HLA complexes by $CD8^+$ αβ T cell receptors (TCRs) can trigger killing of infected cells[2]. Peptide/HLA assembly occurs in the endoplasmic reticulum, where the invariant light chain (human β2-microglobulin, hβ2m), the heavy chain (HLA), and the peptide antigen non-covalently associate to form a ternary complex in a process guided by immunological chaperones[3–6]. The proteasome and peptidases coordinate to process endogenous and foreign proteins resulting in peptides (typically of lengths of 8 to 15 amino acids) that are trafficked into the endoplasmic reticulum by the transporter associated with antigen processing (TAP) for association with HLA molecules[7,8]. While hβ2m is conserved across peptide/HLA complexes, the HLA heavy chain is one of the most polymorphic genes in the human genome resulting in thousands of different alleles that vary from person to person across populations[9]. The major class I HLA molecules (HLA-A, HLA-B, HLA-C, HLA-E, HLA-G) are grouped based on sequence similarity with thousands of variants within each class[9]. Each HLA contains a specialized antigen binding groove with several pockets (termed A, B, C, D, E, and F) whose chemical

[1]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332, USA. [2]These authors contributed equally to this work: Marielle B. Frooman, Klara Choi and Maya Z. Kahn. ✉email: andrew.mcshan@chemistry.gatech.edu

environment and geometry, varying between different alleles, determines the HLA's affinity for each peptide antigens[10,11]. Ultimately, the HLA antigen binding groove establishes the repertoire of peptide antigens that participate in immune signaling with T cells where each HLA allele is estimated to present ~ 10,000 unique 9-mer peptides[12,13]. Given that HLA-presented peptide antigens can serve as prophylactic vaccines[14] and T cells reactive to disease-specific peptide/HLA complexes can be used for immunotherapy[15], it is critical to identify novel pathogen derived peptide antigens, characterize their HLA binding repertoires in disease susceptible populations, determine the stability of the resulting peptide/HLA complexes, and elucidate atomic structures of the resulting peptide/HLA complex[16–18]. Peptide/HLA structures provide important insights into which amino acids of the peptide antigen anchor it into the HLA groove, and which amino acids of the peptide antigen are surface displayed to TCRs, providing a partial view of the molecular basis for immunity[2,10,17].

Malaria is a disease of urgent public health concern with an estimated ~ 245 million cases resulting in > 600,000 deaths in 84 endemic countries yearly[19]. Current therapeutic strategies to treat or prevent malaria include small molecule drugs[20], vaccines[21], and antibodies[22]. The World Health Organization recommendations for vaccination for children living in malaria endemic areas include RTS, S/AS01[23,24] and R21/Matrix-M[25], although many more vaccine candidates are in the pipeline for further development[26,27]. Despite the availability of the recent malaria vaccines, many African, Asian, and South American countries are experiencing a malaria resurgence[19,28]. Thus, there is a need to understand malaria immunity during pre-erythrocytic, erythrocytic, and post-erythrocytic stages of infection. Recent efforts have focused on uncovering the role that peptide/HLA and $CD8^+$ T cell immune responses play in recognition and targeting of malaria associated epitopes, which remain incompletely understood[29,30].

The causative agents of malaria are parasitic protists from the genus *Plasmodium* (species *P. falciparum*, *P. vivax*, *P. knowlesi*, *P. ovale*, *and P. malariae*) that undergo a complex disease cycle upon infection[30]. *Plasmodium* enter human hosts as sporozoites through the saliva of *Anopheles*, and potentially *Culex*, mosquito vectors[31]. The sporozoites first rapidly invade hepatocytes, which then release merozoites into the bloodstream, where they initiate the symptomatic cycle of invasion and rupturing of erythrocytes[30]. Finally, the sexual cycle is initiated when asexual parasites produce mature gametocytes that are transmitted in the blood to new mosquitoes[32]. The infection is mediated by key proteins from *Plasmodium*, such as circumsporozoite protein (CSP), liver stage antigen 3 (LSA3), merozoite surface protein 1 (MSP1), and sexual stage-specific protein (SSFP also known as Pfs16)[33,34]. CSP and MSP1 are the primary surface proteins of sporozoites and merozoites, respectively, aiding in infiltration of hepatocytes and erythrocytes[35,36]. CSP is an extremely important vaccine target for the pre-erythrocytic stage of infection. In particular, CSP serves as a major feature of the RTS, S/AS01 vaccine (epitopes within repeat region and C-terminal domain of CSP)[23], the R21/Matrix-M vaccine (epitopes within repeat region and C-terminal domain of CSP)[37], and also serves as a monoclonal antibody target[38]. MSP1 is one of the most abundant merozoite surface proteins that participates during invasion of erythrocytes and has been shown to promote both T cell and antibody based immune responses[39,40]. LSA3 is primarily expressed on sporozoite, schizonts, and merozoites, and has been shown to be highly immunogenic[41,42]. SSFP/Pfs16 is primarily expressed in gametocytes and thus represents an attractive target for transmission-blocking therapeutics against malaria[43,44]. However, while *Plasmodium* proteins with critical roles in pathogenesis have shown great promise as therapeutic targets to treat malaria and generally result in robust immune responses[41,45–47], the potential for cross-presentation of epitopes across diverse HLA genotypes within malaria endemic populations remains only partially explored[48,49].

As a result of whole genome sequencing and bioinformatics advances, discovering peptides presented by HLAs has become more streamlined[50]. Several studies have experimentally and computationally predicted *Plasmodium falciparum* epitopes that serve as candidates for presentation by HLA molecules[51–54] (summarized in Supplementary Data 1). CSP has been the focus of the large majority of studies with 111 peptides identified in the Immune Epitope Data Bank (IEDB) spanning many HLA alleles[49,52,55–61]. MSP1 has 6 peptides identified in the IEDB spanning HLA-A*02:01[62]. LSA3 has 7 identified in the IEDB spanning HLA-A*02:01 and HLA-B*53:01 [63,64]. SSFP/Pfs16 has 2 identified in the IEDB spanning HLA-B*07:02, HLA-B*53:01, HLA-B*35:01, HLA-B*51:01, and HLA-B*54:01 [48,49,60]. Many class I HLA alleles (not just those listed above) are prevalent in the malaria endemic population where some are associated with protection from severe malaria while other alleles are associated with increased susceptibility to malaria[9,65–68]. Given that each HLA exhibits unique a peptide binding profile[69], it is essential to define immunogenic signatures across alleles and identify peptide antigens that could be presented by multiple individuals in the population. It is especially important to define the biophysical and structure basis of how candidate epitopes bind to and stabilize HLAs, and the nature of the peptide/HLA surface presented to T cell receptors. While there are hundreds of peptide/HLA structures available in the Protein Data Bank[70], to our knowledge, there is only one published crystal structure of a class I HLA presenting a *Plasmodium* derived peptide antigen: HLA-B*53:01 with 9-mer KPIVQYDNF derived from liver stage antigen 1 residues 1786–1793 [71]. Computational modeling of peptide/HLA complexes is a high-throughput approach that can provide insights into antigen presentation in the absence of experimental structures[70,72,73], but has not yet been extensively applied to presentation of *Plasmodium* peptides by HLAs. Thus, there is a wealth of biophysical and structural information that remains untapped for the numerous *Plasmodium* derived peptides with potential for cross-presentation by different HLA molecules, resulting in a range of distinct $CD8^+$ T cell responses during malaria.

To begin to address these knowledge gaps, we perform a comprehensive bioinformatics, biophysical, and structure modeling analysis to discover, validate, and characterize candidate malaria associated antigens derived from nine key *Plasmodium* proteins with two HLA alleles common in African populations, HLA-A*02:01 and HLA-B*08:01. We computationally predict *Plasmodium* derived 9-mer peptides and validate eight hits from MSP1, CSP, SSFP, and LSA3 proteins as positive binders to both HLA-A*02:01 and HLA-B*08:01 in vitro. The resulting sixteen peptide/HLA complexes are characterized in terms of their apparent refolding yields/

total yields and thermal stabilities. Using TFold, a fine-tuned version of AlphaFold2 optimized for peptide/HLA structure prediction[73], we generate and analyze in silico structural models of experimentally validated peptide/HLA complexes to compare HLA binding modes as well as TCR recognition surfaces between different *Plasmodium* derived antigens and different HLA alleles. Finally, we expand our computational predictions by probing which of the > 10,000 class I HLA alleles, including those found in malaria endemic populations, are expected to present our eight identified *Plasmodium* derived peptides. Experimental validation is provided for *Plasmodium* derived peptide binding to HLA-E*01:03 and HLA-C*07:02, common in malaria endemic African populations. Together, our results provide an expanded holistic view of the biophysical and structural basis behind HLA-mediated cross-presentation of candidate *Plasmodium* antigens across diverse HLA alleles within susceptible populations. These studies provide a step towards elucidating the mechanistic basis for differences in immunogenicity and malaria susceptibility across individuals while informing future studies of HLA-mediated malaria vaccine development.

Materials and methods

Computational prediction of HLA presented plasmodium epitopes

Due to the prevalence of *Plasmodium* infection in Africa resulting in malaria disease, HLA alleles with high frequency were selected from distributions of HLAs in African populations provided by the Allele Frequency Net Database[9]. HLA-A*02:01 and HLA-B*08:01, common in both Sub-Saharan African and North African populations (and many other malaria endemic populations as well), were chosen as our initial representative alleles for study. Identification of candidate HLA-A*02:01 and HLA-B*08:01 binding epitopes was achieved by in silico screening of proteins important for the lifecycle and virulence of *Plasmodium falciparum* (isolate 3D7). The following proteins were screened: circumsporozoite protein (CSP – UniProt ID #Q7K740), sporozoite threonine and asparagine-rich protein (STARP – UniProt ID #Q8IC44), thrombospondin-related anonymous protein (TRAP – UniProt ID #P16893), exported protein 1 (EXP1 – UniProt ID #P04923), liver stage antigen 1 (LSA1 – UniProt ID #A0A143ZZD7), liver stage antigen 3 (LSA3 – UniProt ID #O96275), merozoite surface protein 1 (MSP1 – UniProt ID #Q8I0U8), apical membrane antigen 1 (AMA1 – UniProt ID #Q7KQK5), and sexual stage specific protein (SSFP or PFS16 – UniProt ID #Q6ZMA7) (Supplementary Data 1). Full-length amino acid sequences for these proteins were computationally cut into all possible 9-mer amino acid fragments by netMHCpan-4.1[50]. A list of strong and weak binders was provided by netMHCpan-4.1 along with % rank binding affinities (%Rank BA) using default parameters (Supplementary Data 1). Epitopes were then filtered as predicted to be positive for binding to the TAP transporter using TAPPred[74]. The final eight candidate *Plasmodium falciparum* isolate 3D7 derived peptide antigens for experimental testing were chosen to exhibit strong predicted binding affinity for both HLA-A*02:01 and HLA-B*08:01 (Table 1). Kullback–Leibler sequence (peptide motif) logos for HLA-A*02:01 and HLA-B*08:01 were generated using Seq2Logo v2.0 (https://services.healthtech.dtu.dk/services/Seq2Logo-2.0/) where inputs were amino acid sequences of all netMHCpan-4.1 predicted strong binding 9-mer peptides for each HLA allele[75].

To find previously identified epitopes for CSP, LSA3, MSP1, and SSFP, we searched the IEDB (https://www.iedb.org/) with the following parameters: Epitope – Any; Epitope Source – Antigen – UnitProtID#; Assay – T cell + MHC Ligand; Outcome: Positive; MHC Restriction – Class I; Host – Human; Disease – Any. The results are presented in the "IEDB search results" tab of Supplementary Data 1.

Candidate epitope	Sequence	Predicted BA*	Refolding yield (%)	Thermal stability (°C)	$CD8^+$ T Cell stimulation/antigen presentation
CSP 319–328	YLNKIQNSL	17.98 (A02) 183.57 (B08) 1106.93 (C07)	31.3 ± 0.7 (A02) 58.8 ± 3.0 (B08) 36.9 ± 3.3 (C07)	61.0 ± 0.1 (A02) 54.9 ± 0.4 (B08) 61.2 ± 0.2 (C07)	Yes [56,58,100,101]
LSA3 78–87	YVDKKLNKL	853.47 (A02) 1745.10 (B08)	13.5 ± 2.8 (A02) 14.7 ± 3.7 (B08)	52.5 ± 0.4 (A02) 56.3 ± 0.1 (B08)	N/D*
LSA3 1229–1238	KLGERVESL	43.41 (A02) 532.68 (B08) 13,867.29 (E01)	24.6 ± 2.3 (A02) 47.6 ± 3.1 (B08) 38.1 ± 1.7 (E01)	58.4 ± 0.1 (A02) 60.5 ± 0.1 (B08) 61.0 ± 0.3 (E01)	N/D*
MSP1 511–520	ELLEKFYEM	170.09 (A02) 58.76 (B08)	19.0 ± 0.6 (A02) 44.4 ± 0.6 (B08)	51.6 ± 0.4 (A02) 61.0 ± 0.0 (B08)	N/D*
MSP1 643–652	LLIKKIEDL	130.37 (A02) 397.57 (B08) 22,097.51 (E01)	24.8 ± 1.1 (A02) 32.6 ± 3.1 (B08) 8.0 ± 2.3 (E01)	51.3 ± 0.1(A02) 52.9 ± 0.1 (B08) 38.4 ± 0.2 (E01)	Yes 62
MSP1 823–832	SMDQKLLEV	18.89 (A02) 3910.45 (B08)	39.5 ± 18.7 (A02) 12.5 ± 3.4 (B08)	63.2 ± 0.1 (A02) 52.6 ± 0.1 (B08)	Yes [62]
MSP1 880–889	YQKEMIYYL	9.40 (A02) 271.16 (B08)	44.3 ± 0.1 (A02) 63.3 ± 2.6 (B08)	53.6 ± 0.1 (A02) 51.6 ± 0.0 (B08)	N/D*
SSFP 93–102	NMLDKKTTV	207.29 (A02) 284.64 (B08)	27.1 ± 1.0 (A02) 51.1 ± 1.0 (B08)	52.5 ± 0.4 (A02) 56.3 ± 0.1 (B08)	N/D*

Table 1. Summary of candidate *P. falciparum* derived peptides for cross-presentation by HLA-A*02:01, HLA-B*08:01, HLA-C*07:02, and HLA-E*01:03 alleles. A02 = HLA-A*02:01. B08 = HLA-B*08:01. E01 = HLA-E*01:03. C07 = HLA-C*07:02. *Predicted BA = predicting binding affinity from netMHCpan-4.1 (nM units). *N/D = not yet determined (based on literature and Immune Epitope Database search). Error bars are derived from mean values ± standard deviation for two technical replicates (n = 2).

Towards evaluating the entire HLA class I allele binding landscape, netMHCpan-4.1 was used to predict binding of each of the eight identified *Plasmodium* derived peptides to a total of 10,386 HLA-A-, B-, C-, E-, and G- alleles. A list of strong and weak binders was obtained together with % rank binding affinity (BA) from the netMHCpan-4.1 output using default parameters. BLOSUM62 identity scores[76] of each HLA relative to HLA-A*02:01 for all 10,386 alleles were generated from a Clustal Omega v1.2.4 alignment of residues representing a pseudo MHC groove (MHC_pseudo.dat file located in $netMHCpan-4.1/data). We further filtered the netMHCpan-4.1 results based on 47 alleles that cover 99% of the global population[9,77]. Kullback–Leibler sequence (peptide motif) logos were obtained from netMHCpan-4.1 motif viewer with option "MHC binders (BA)".

Based on allelesfrequenecies.net (http://www.allelefrequencies.net/top10freqs.asp), the top 10 most frequently represented alleles in African populations are broken down as follows:

HLA-A (North Africa): HLA-A*01:01‡, HLA-A*02:01‡, HLA-A*02:04, HLA-A*02:05, HLA-A*03:01‡, HLA-A*23:01‡, HLA-A*24:02‡, HLA-A*29:02, HLA-A*30:01‡, HLA-A*68:02.
HLA-A (Sub-saharan Africa): HLA-A*01:01‡, HLA-A*02:01‡, HLA-A*03:01‡, HLA-A*11:01‡, HLA-A*23:01‡, HLA-A*29:02, HLA-A*30:01‡, HLA-A*30:02‡, HLA-A*68:02, HLA-A*74:01.
HLA-B (North Africa): HLA-B*08:01‡, HLA-B*35:01‡, HLA-B*41:01, HLA-B*42:01, HLA-B*44:02‡, HLA-B*45:01, HLA-B*50:01, HLA-B*51:01‡, HLA-B*52:01‡, HLA-B*53:01‡.
HLA-B (Sub-saharan Africa): HLA-B*07:02‡, HLA-B*07:05, HLA-B*07:06, HLA-B*07:07, HLA-B*07:08, HLA-B*07:12, HLA-B*07:17, HLA-B*08:01‡, HLA-B*08:02, HLA-B*08:03.
HLA-C (North Africa): HLA-C*02:02‡, HLA-C*04:01‡, HLA-C*06:02‡, HLA-C*07:01‡, HLA-C*07:02‡, HLA-C*08:02, HLA-C*12:03‡, HLA-C*15:02‡, HLA-C*16:01‡, HLA-C*17:01‡.
HLA-C (Sub-saharan Africa): HLA-C*02:02‡, HLA-C*02:10, HLA-C*03:04‡, HLA-C*04:01‡, HLA-C*06:02‡, HLA-C*07:01‡, HLA-C*07:02‡, HLA-C*08:02, HLA-C*16:01‡, HLA-C*17:01‡.

The double dagger symbol (‡) above marks alleles that are also present in the top 99% of class I HLA alleles presented in the global population (i.e., would be of interest not only to malaria related immune responses in Africa but other endemic populations).

Recombinant protein production

Recombinant plasmid DNA containing gene sequences that encode for HLA-A*02:01, HLA-B*08:01, and human β2m (hβ2m) were prepared. Constructs were soluble domains of HLA-A*02:01 (residues 25–299 – UniProt ID #P04439), HLA-B*08:01 (residues 25–299 – UniProt ID #P01889), HLA-C*07:02 (residues 26–299 – IMGT/HLA accession #HLA00434), HLA-E*01:03 (residues 22–296 – IMGT/HLA accession #HLA00936), and hβ2m (residues 21–119 – UniProt ID #P61769) as described previously[78]. Briefly, DNA sequences were codon optimized for *Escherichia coli* K-12 expression systems, synthetic DNA was purchased (GenScript), and genes were separately subcloned into pET-22b(+) vectors containing a carbenicillin antibiotic resistance gene using NdeI/BamHI restriction sites. Recombinant plasmid DNA (~ 100 ng) was transformed into BL21(DE3) *E. coli* cells (New England Biolabs) using heat shock for 45 s in a dry heat block at 42 ℃. Each mixture was plated on a 1X carbenicillin LB-agar plate under sterile conditions. The plates were incubated overnight (~ 16 h) at 37℃. Proteins were expressed by the Studier autoinduction method in 500 mL LB-broth media (Miller) supplemented with 1X carbenicillin[79]. To 500 mL LB-broth, the following autoinduction buffers were added: 25 mL solution 1 [0.5 M $(NH_4)_2SO_4$, 1 M KH_2PO_4, 1 M Na_2HPO_4 pH 7], 10 mL solution 2 [25% v/v glycerol, 2.5% w/v glucose, 10% w/v α-lactose], and 0.5 mL solution 3 [1 M $MgSO_4$]. Cultures were incubated overnight (~ 18 h) at 37 ℃ with shaking at 225 RPM. Following incubation, cell growth was verified by checking the OD_{600} nm of each liquid culture. Cells were centrifuged at 6000 RPM for 20 min at 4 ℃ to separate out the spent media from the *E. coli* cell pellet. The supernatants were discarded, and the bacterial pellets were frozen at – 80 ℃ until further processing.

Isolation and solubilization of inclusion bodies

Inclusion bodies were isolated and solubilized as described previously[78,80]. Briefly, bacterial cell pellets were resuspended with 12.5 mL of BugBuster (EMD Millipore) and 130 µL of 200 mM phenylmethylsulfonyl fluoride. Cell suspensions were sonicated at 40% amplitude and chilled on ice in 30 s cycles for a total of 6 min followed by centrifugation for 20 min at 10,000 RPM at 4 ℃. All supernatants were discarded, and inclusion bodies were further isolated through continued sonication and centrifugation cycles with 25 mL of Wash Buffer A [100 mM Tris pH 8, 2 mM EDTA, 0.1% w/v deoxycholate] followed by 25 mL of TE buffer [100 mM Tris pH 8, 2 mM EDTA]. Inclusion bodies were resuspended in 6 mL TE buffer with further sonication. Protein expression was confirmed through SDS-PAGE. For inclusion body solubilization, mixtures of 5.7 g guanidine hydrochloride and 360 µL of fresh 0.1 M dithiothreitol were added to each suspension. Solutions were incubated for one hour at 25 °C with gentle agitation then centrifuged at 10,000 RPM for 20 min at 4 ℃ to remove any insoluble material. Inclusion body concentrations were measured with a NanoDrop spectrophotometer at A_{280} nm. The molecular weight and molar extinction coefficients used for calculations were obtained from ExPASy ProtParam tool[81].

In vitro refolding

In vitro refolding was performed by combining a 1:3 molar ratio (1 is heavy chain, 3 is light chain) of HLA heavy chain to hβ2m light chain inclusion bodies along with 9-mer peptides[78,80]. Erlenmeyer flasks containing 250 mL of refolding buffer [100 mM Tris pH 8, 2 mM EDTA, 0.4 M Arginine-HCl, 4.9 mM L-glutathione reduced, 0.57 mM L-glutathione oxidized] were prepared and stirred at 4 ℃ for 30 min. At 4 ℃, 10 mg of the selected 9-mer peptide, ~ 10X molar excess relative to each HLA heavy chain, was dissolved in 1 mL dimethyl sulfoxide and

added to the refolding buffer in a slow dropwise fashion. Candidate peptides were chemically synthesized and purified to > 95% purity (GenScript). Following peptide addition to the refolding buffer, 50 mg of HLA inclusion bodies was combined with 50 mg of hβ2m inclusion bodies. The HLA/hβ2m mixtures were added slowly in a dropwise fashion into their respective flasks over 5 h. The refolding mixture was left without stirring (~ 4 days) at 4 ℃. For all refoldings, two technical replicates were performed, and data are presented as the mean ± standard deviation ($n = 2$).

Protein purification

Dialysis membranes with 3500 Da molecular weight cutoff (Spectrum) were soaked in water. The 250 mL refoldings of peptide/HLA-A*02:01/hβ2m complexes or peptide/HLA-B*08:01/hβ2m complexes were dialyzed separately in 5 L of dialysis buffer [150 mM NaCl, 25 mM Tris pH 8] for 18 h at 4 ℃ with stirring. The protein complexes were then centrifuged at 6,000 RPM for 15 min at 4 ℃ to remove precipitates. The protein solutions were concentrated with 30 kDa molecular weight cutoff Amicon Ultra centrifugal filter devices (EMD Millipore) until 5 mL of sample was obtained. The concentrated protein complexes were then centrifuged at 10,000 RPM for 10 min at 4 ℃ to remove precipitates then passed through a size exclusion chromatography (SEC) column. SEC was performed using an ÄKTA GO FPLC with a HiLoad 16/600 Superdex 75 pg column at 1 mL/min in SEC buffer [150 mM NaCl, 25 mM Tris pH 8] to separate successfully refolded peptide/HLA/hβ2m complexes from free β2m and protein aggregates. Based on the resulting SEC peaks, fractions of sample containing the purified protein complexes were collected and consolidated. SDS-PAGE of the consolidated fractions was performed to confirm successful refolding and purification of the proteins. Upon completion of SEC purification, the solutions were concentrated to ~ 10 mg/mL for further analysis. Refolding ratios (%) were determined by integrating the total area under the aggregate peak (roughly 45 min) and the refolded peptide/HLA/hβ2m peak (roughly 60 min) from SEC traces and then calculating the ratio of $Area_{Folded}/(Area_{Aggregate} + Area_{Folded})$[82]. Refolding efficiency was also quantified in terms of total yield (milligrams, mg) of properly confirmed peptide/HLA/hβ2m complex obtained. The total yield was determined in Unicorn software with the formula (A * V) / (d * 1000 * ε) where A is $Area_{Folded}$ (roughly 60 min peak) in mL*mAU, d is the UV-cell path length in cm, ε is the extinction coefficient in $mg^{-1}*mL*cm^{-1}$, V is the pooled fraction volume in mL. For all refoldings, two technical replicates were performed and data are presented as the mean ± standard deviation ($n = 2$). Following SEC, properly refolded peptide/HLA/hβ2m complex fractions were further purified by ion exchange chromatography. Ion exchange chromatography was performed using an ÄKTA GO FPLC with a HiTrap Q HP column at 1 mL/min flow rate with a linear gradient of Buffer A [50 mM NaCl, 25mM Tris pH 8] and Buffer B [1 M NaCl, 25mM Tris pH 8] over 5 column volumes. The resulting peptide/HLA/hβ2m complexes were extensively dialyzed into 150 mM NaCl, 25 mM Tris pH 8. Protein identity and purity were determined by SDS-PAGE. Protein concentrations were determined via NanoDrop spectrophotometer at A_{280} nm. The molecular weight and molar extinction coefficients used for calculations were obtained from ExPASy ProtParam tool[81].

Circular dichroism (CD) spectroscopy

Far-UV CD spectra were collected on a JASCO J-815 Spectropolarimeter. CD spectra were measured from 190 nm to 350 with 0.3 mg/mL peptide/HLA/hβ2m complexes in CD buffer [100 mM NaCl, 20 mM sodium phosphate pH 7.2] using 1 mm path length quartz cuvettes (Starna Cells #21-Q-1/CD). CD spectra were acquired in triplicate at 25 °C with a scan rate of 50 nm/min and then averaged. The average CD spectra were smoothed using a Savitzky–Golay filter with convolution width of 7.

NanoDSF

Nano Differential Scanning Fluorometry (nanoDSF) was performed using a Prometheus NT.48 (NanoTemper Technologies). Purified peptide/HLA/hβ2m samples at 10 µM in SEC buffer [150 mM NaCl, 25 mM Tris pH 8] were loaded into nanoDSF grade standard capillaries (NanoTemper Technologies) and exposed to temperatures ranging from 20 °C to 95 °C with a thermal ramping rate of 1 °C/min. The ratio of intrinsic fluorescence emission (F_{350} nm/F_{330} nm) was monitored. Melting temperatures (T_m) were obtained from fitting the first derivative plots in the Prometheus NT.48 Software. Two technical replicates were performed and data is presented as the mean ± standard deviation ($n = 2$).

Peptide/HLA structure modeling with TFold

TFold code obtained from GitHub (https://github.com/v-mikhaylov/tfold-release) was run locally. Default parameters were used with automated register detection. The command used to model the structures was `model_pmhcs.sh $input_file $working_dir`. The structures indicated as the best models by TFold were selected. The predicted error score (100-pLDDT averaged over the peptide core) for the TFold models ranged from 3.6 to 6.8, which is in the range of the authors' defined high predicted accuracy error score[73].

Peptide/HLA interaction analysis

Hydrogen bonds and salt bridges between peptide residues and HLA residues were identified from TFold models using PBDe PISA v1.52 (https://www.ebi.ac.uk/pdbe/pisa/)[83] and PyMOL v2.5.4. Explicit interface analysis calculations (i.e., ΔSASA interface and ΔG separated) were performed using *Rosetta* v2021.16 with Idealize, Relax, and InterfaceAnalyzer applications[84]. The following code was performed on the TFold peptide/HLA models:

```
idealize_jd2.static.macosclangrelease -s *.pdb.
relax.static.macosclangrelease -s *_0001.pdb -relax: ramp_constraints
false -ex1 -ex2 -use_input_sc -flip_HNQ -no_optH false.
```

```
InterfaceAnalyzer.static.macosclangrelease -s *_0001_0001.pdb -out: file:
scorefile interface_scores.sc -overwrite.
awk 'NR>  1{print $NF, $9, $6}' interface_scores.sc.
```

Results

In silico identification of eight unique *P. falciparum* candidate antigens predicted to bind to both HLA-A*02:01 and HLA-B*08:01 alleles

To identify *P. falciparum* derived peptides with the potential to be presented by both HLA-A*02:01 and HLA-B*08:01 alleles (two common HLAs in malaria endemic populations from the Allele Frequency Net Database[9] – see Materials and Methods), we performed in silico screening of a subset of the *P. falciparum* 3D7 proteome[85]. Full length sequences of nine proteins key to the life cycle and virulence of *P. falciparum* infection were screened by netMHCpan-4.1 to identify nonameric (9-mer) peptides predicted to bind to both HLA-A*02:01 and HLA-B*08:01 [50]. Of the nine *P. falciparum* proteins screened, circumsporozoite surface protein (CSP) and sexual stage specific protein (SSFP) each contained one 9-mer sequence predicted to strongly bind to both alleles, liver stage antigen 3 (LSA3) contained two 9-mer sequences predicted to strongly bind to both alleles, and merozoite surface protein 1 (MSP1) contained four 9-mer sequences predicted to strongly bind to both alleles (Fig. 1A, red asterisks and Table 1). The predicted affinities of these eight candidate 9-mer epitopes to either HLA-A*02:01 or HLA-B*08:01 ranged from 9 to 4000 nM (Fig. 1B; Table 1). Sequence logos, which enable visualization of peptide binding motifs to assess sequence similarity among identified peptides, were generated from the amino acid sequences of all 9-mer peptides predicted to strongly bind to either HLA-A*02:01 or HLA-B*08:01 [75]. For *P. falciparum* derived peptides expected to bind strongly to HLA-A*02:01, positions 2 and 9 displayed high frequency of hydrophobic residues Leu, Val, Ile, and Met, while position 4 displayed a moderate frequency of negatively charged residues Glu and Asp (Fig. 1C, top panel). *P. falciparum* derived peptides expected to bind strongly to HLA-B*08:01 displayed high frequency of hydrophobic residues Leu, Val, Ile, Phe, and Met at position 9 (with moderate frequency at position 2), while position 5 displayed a moderate frequency of positively charged residues Lys and Arg (Fig. 1C, bottom panel). These observations are consistent with previous reports of experimentally determined peptide binding motifs for 9-mer peptides with HLA-A*02:01 and HLA-B*08:01 alleles[69,86,87]. Together, these results identify a list of eight candidate *P. falciparum* derived 9-mer epitopes (spanning four different proteins involved in *Plasmodium* pathogenesis/lifecycle) that are predicted to associate with both HLA-A*02:01 and HLA-B*08:01 (Table 1). Based on sequence logos, the defining features for cross-presentation of *P. falciparum* 9-mer peptides on HLA-A*02:01 and HLA-B*08:01 is a delicate balance between the identity of the hydrophobic residues at positions 2/9 (i.e., Leu, Val, Ile) and the charged residues at positions 4/5 (i.e., E, D, K) (Fig. 1C).

In silico identified *P. falciparum* peptides efficiently refold in vitro with both HLA-A*02:01 and HLA-B*08:01

The primary pipeline for experimental validation of netMHCpan-4.1 predicted epitopes is in vitro refolding of candidate peptides together with *E. coli* expressed recombinant heavy chain (HLA-A*02:01 or HLA-B*08:01) and light chain (hβ2m)[17,78,88,89]. Here, the presence of a moderate to strongly binding peptide, provided exogenously in the refolding buffer, is necessary for proper in vitro refolding of the peptide/HLA/β2m complex[78]. To experimentally validate the eight identified candidate *P. falciparum* derived 9-mer peptides with both HLA alleles (Table 1), we performed a total of sixteen in vitro refolding experiments (with two technical replicates each) between chemically synthesized peptide together with either HLA-A*02:01/hβ2m or HLA-B*08:01/hβ2m. To evaluate refolding efficiency, we analyzed (i) apparent refolding yields from size exclusion chromatography (SEC) traces during protein purification and (ii) quantified the total yield of properly conformed peptide/HLA/hβ2m complex generated in milligrams (Fig. 2A, B; Table 1 and Supplementary Fig. 1A-C). HLA-A*02:01/hβ2m or HLA-B*08:01/hβ2m refolded in the absence of exogenous peptide (i.e., empty) did not generate any appreciable amounts of properly folded peptide/HLA complex (Fig. 2B and Supplementary Fig. 1B, C). For seven of the eight peptides, higher refolding yields were obtained from mixtures of *P. falciparum* derived 9-mer peptides with HLA-B*08:01/hβ2m relative to complexing with HLA-A*02:01/hβ2m (Fig. 2A, B and Supplementary Fig. 1A, C). Apparent refolding yields of ~60% (6 to 8 milligrams of protein) were obtained for CSP 319–328/HLA-B*08:01 and MSP1 880–889/HLA-B*08:01 (Table 1; Fig. 2A, B and Supplementary Fig. 1A, C). In contrast, LSA3 78–87 exhibited poor refolding with both HLA-A*02:01 and HLA-B*08:01. MSP1 823–832 was the only peptide that refolded more efficiently with HLA-A*02:01 than HLA-B*08:01. To obtain additional evidence that the in vitro refolding procedures were successful, we performed far-UV circular dichroism (CD) spectroscopy analysis, which has previously been shown to highlight the β-sheet secondary structure characteristics of folded peptide/HLA/hβ2m molecules with negative CD signal at ~217 nm[90]. As expected, CD spectra revealed that *Plasmodium* derived peptide/HLA/hβ2m complexes were properly refolded, and no striking differences in CD spectra due to influences on the identity of the peptide or HLA allele were observed within the resolution of the experiment (Fig. 2C and Supplementary Fig. 2). Together, these results show that netMHCpan-4.1 based predictions of 9-mer *Plasmodium*-derived peptide/HLA pairs are robust, and refolding of all sixteen peptide/HLA/hβ2m complexes was successful with notable peptide-dependent and allele-dependent differences in refolding yields.

P. falciparum peptides stabilize HLA-A*02:01 and HLA-B*08:01 complexes

Thermal stability (T_m or melting temperature) has been previously shown to positively correlate with cytotoxic immunogenicity of peptide/HLA/hβ2m complexes[16,18,91]. Since peptide/HLA/hβ2m complex stability is both peptide or HLA allele dependent, it is important to quantify the thermal stability of each purified complex. To evaluate the thermal stability of our eight *P. falciparum* derived 9-mer peptides in complex with HLA-A*02:01

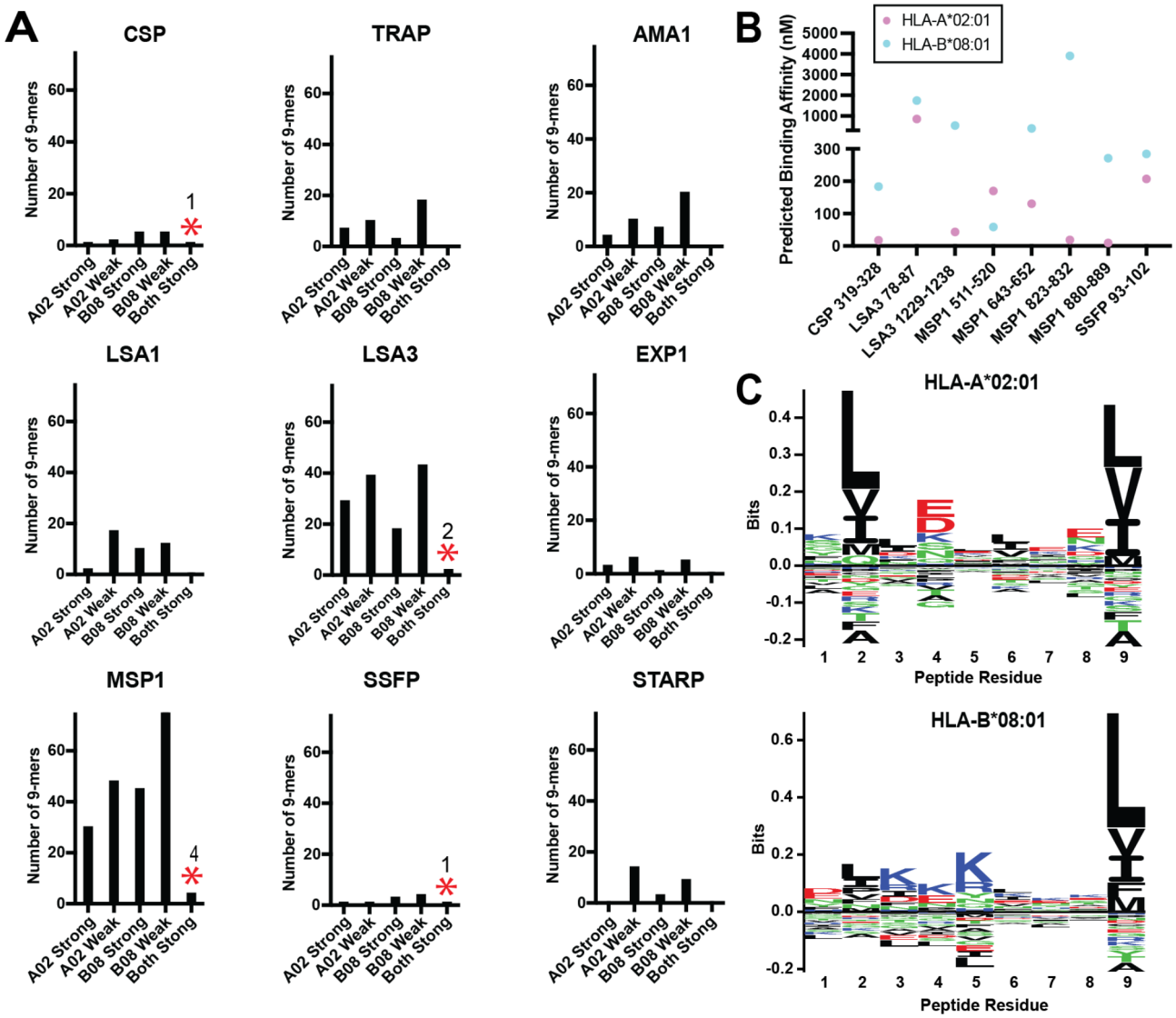

Fig. 1. In silico identification of *P. falciparum* peptides predicted to bind to HLA-A*02:01 and HLA-B*08:01. (**A**) Summary of the number of 9-mer peptides with predicted weak or strong binding (by netMHCpan-4.1) to HLA-A*02:01, HLA-B*08:01, or both alleles across nine *P. falciparum* proteins. The red asterisk and number above indicate the quantity of unique peptides predicted to strongly bind to both alleles. (**B**) Comparison of netMHCpan-4.1 predicted binding affinities (nM units) for candidate *P. falciparum* 9-mer epitopes predicted to bind to both HLA alleles (HLA-A*02:01 in pink, HLA-B*08:01 in cyan). A lower value indicates better predicted binding. (**C**) Sequence logo motifs (generated by Seq2Logo 2.0) derived from peptide sequences of all predicted strong binding *P. falciparum* peptides for either HLA-A*02:01 or HLA-B*08:01. Hydrophobic amino acids are colored black, positively charged amino acids are colored blue, negatively charged amino acids are colored red, polar amino acids are colored green.

or HLA-B*08:01, we determined T_m values for all sixteen refolded complexes with nano differential scanning fluorimetry (nanoDSF). NanoDSF has been shown to be an excellent label-free approach to characterize peptide/HLA/hβ2m thermal stability by probing changes in intrinsic tryptophan and tyrosine side-chain fluorescence as a ratio of emissions at 350 and 330 nm (F_{350}/F_{330}) as a function of temperature[92]. As measured by nanoDSF, all sixteen peptide/HLA/hβ2m complexes exhibited moderate to high T_m in the 51°C to 63°C range (Fig. 2C, D; Table 1 and Supplementary Fig. 3), which is in line with previous reports showing T_m values ranging from 41 to 68 °C[92–94]. For five of the eight test peptides, peptide/HLA-B*08:01/hβ2m complexes displayed a higher T_m relative to the analogous peptide/HLA-A*02:01/hβ2m complex (Fig. 2D; Table 1 and Supplementary Fig. 3), which is consistent with observations from refolding yields (Fig. 2B and Supplementary Fig. 1C). We hypothesized that in vitro refolding yield (a proxy of kinetic stability) and T_m values (a proxy of thermal stability) would be positively correlated. While we did not identify a clear linear correlation between T_m values with apparent refolding yield (%), we did observe a modest correlation between T_m values with total yield of protein obtained (mg) (Supplementary Fig. 4A, C). Here, apparent refolding yields might not provide robust correlations

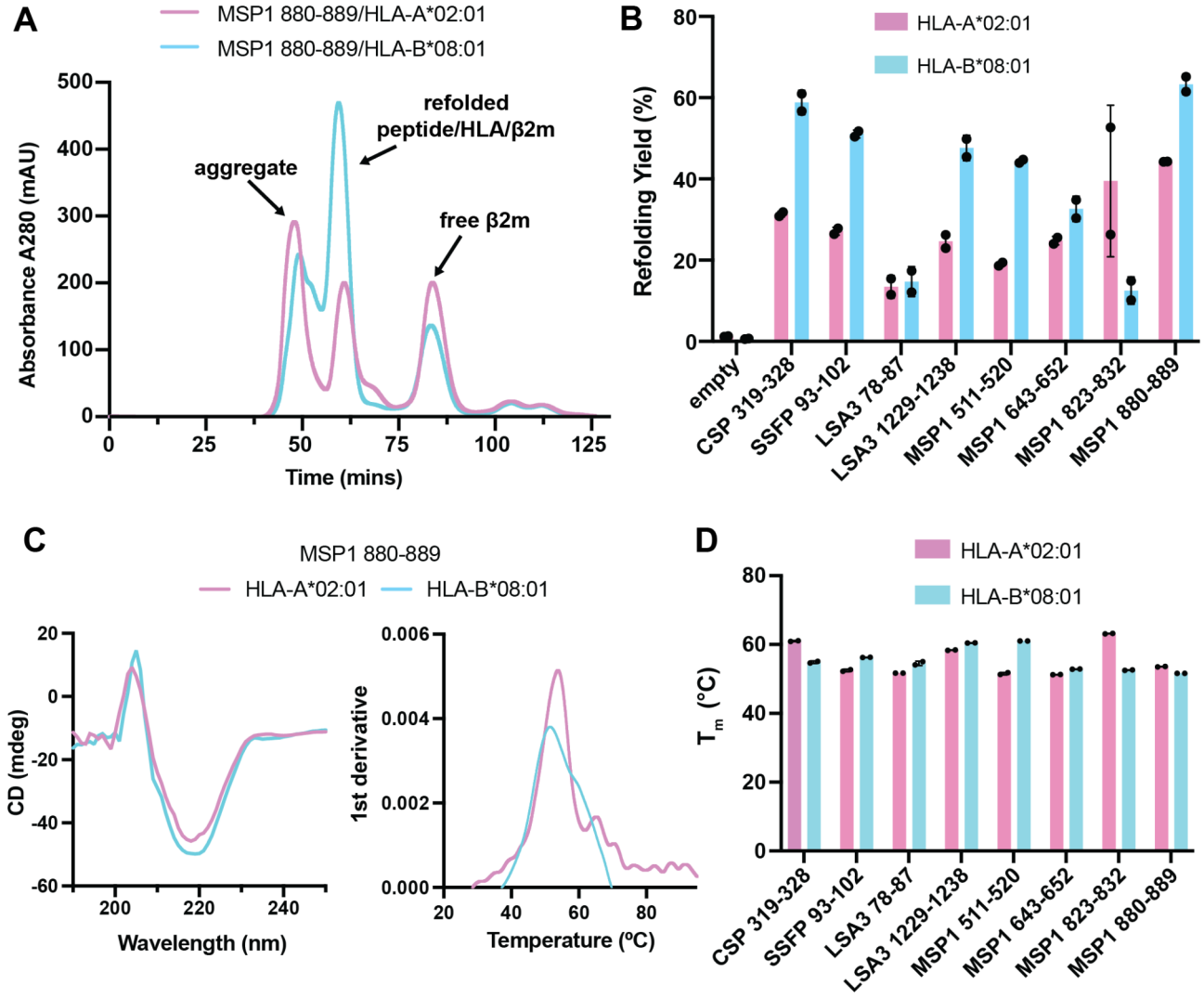

Fig. 2. In vitro validation and biophysical characterization of eight *P. falciparum* derived peptides with HLA-A*02:01 and HLA-B*08:01. (**A**) Size exclusion chromatography traces of HLA-A*02:01/hβ2m (pink) and HLA-B*08:01/hβ2m (cyan) refolded with MSP1 880–889 (YQKEMIYYL). (**B**) Summary of apparent in vitro refolding yields (%), determined from size exclusion chromatography traces, for HLA-A*02:01/hβ2m (pink) or HLA-B*08:01/hβ2m (cyan) without (empty) or with a panel of predicted candidate antigens derived from *Plasmodium*. (**C**) Left: Far-UV circular dichroism spectra and Right: First derivative of the denaturation nanoDSF spectra (F_{350}/F_{330}) for in vitro refolded MSP1 880–889/HLA-A*02:01/hβ2m (pink) and MSP1 880–889/HLA-B*08:01/hβ2m (cyan) complexes. (**D**) NanoDSF derived melting temperatures (T_m, °C) for different *P. falciparum* derived candidate peptide antigens in complex with either HLA-A*02:01/hβ2m (pink) or HLA-B*08:01/hβ2m (cyan). Error bars are derived from mean values ± standard deviation for two technical replicates ($n = 2$).

as they are based on quantification relative the aggregate peak in size exclusion chromatography experiments and do not consider any insoluble materials filtered out before or during the purification (which may be peptide and HLA allele specific). We also identified a modest correlation between in vitro refolding yield (%) or total yield (mg) as a function of netMHCpan-4.1 predicted affinity (Supplementary Fig. 4B, E). There was also a very weak correlation between T_m values and netMHCpan-4.1 predicted affinity (Supplementary Fig. 4D). Why do we only observe weak to moderate trends of correlation between thermal stabilities vs. apparent refolding yields or total protein yields? We suspect that is that there are several kinetic barriers to refolding (peptide solubility, formation of molecular interactions, peptide diffusion away from the HLA groove, conformational plasticity, etc.) that limit refolding, but once the peptide/HLA/ hβ2m complex forms it remains stable. Together, these results reveal that in silico identified *P. falciparum* 9-mer peptides efficiently stabilize both HLA-A*02:01/hβ2m and HLA-B*08:01/ hβ2m complexes, and that the degree of stabilization is both peptide and HLA-dependent such that it correlates modestly with in vitro refolding efficiency.

In silico models provide a structural basis for *P. falciparum* peptide cross-presentation on HLA-A*02:01 and HLA-B*08:01

We attempted to obtain crystal structures of purified *Plasmodium* derived peptide/HLA complexes but were not successful at growing high quality diffracting crystals. Thus, we turned to molecular modeling. Structural modeling of peptide/HLA complexes can provide valuable insights into *(i)* peptide/HLA interactions that govern kinetic and thermal stability of immunogenic complexes, and *(ii)* the defining features of solvent exposed peptide and HLA residues for T cell receptor recognition[2,17,72,95]. TFold is robust AlphaFold2-based pipeline fine-tuned for accurate peptide-HLA structure modeling[73]. To gain insights into their molecular features, we used TFold to model atomic structures of our eight in silico identified and experimentally validated *P. falciparum* 9-mer peptides with HLA-A*02:01 and HLA-B*08:01 – a total of 16 unique peptide/HLA complexes were modeled. In line with observations from landmark studies on peptide binding modes to HLA-A*02:01 and HLA-B*08:01[10,96], the orientation of 9-mer peptide primary anchor residues 2 and 9 (P2 and P9) into the HLA antigen groove pockets A and F were relatively conserved between all eight peptides across the two alleles (Supplementary Fig. 5 and Supplementary Table 1). The major differences in peptide docking modes within the HLA groove occurred between peptide residues 3 to 7 (P3-P7), the central region of the peptide, which we attribute to differences in the HLA C and D pockets between HLA-A*02:01 and HLA-B*08:01[10,87]. Figures 3 and 4, and Supplementary Table 1 summarize all peptide/HLA structural models where the *Plasmodium* derived peptide numbering is relative to the amino acid numbering of the 9-mer peptide from N- to C- terminus rather than the numbering of the full-length peptide (for example, L2 and L9 of LSA3 1229–1238 refer to the 2nd and 9th amino acids of the 9-mer peptide, which are packed into the P2 and P9 pockets of the HLA groove in Fig. 3).

Comparisons of the details of TFold models reveal P3-P7 positions of *P. falciparum* 9-mer peptides undergo more frequent hydrogen bond and salt bridge interactions with residues within in the floor of the antigen binding groove of HLA-B*08:01 relative to HLA-A*02:01 (Figs. 3 and 4, and Supplementary Table 1). This discrepancy is most apparent in the LSA3 1229–1238 peptide that experimentally demonstrated higher refolding yields with the HLA-B*08:01 in compared to HLA-A*02:01 in nearly two-fold magnitude (Fig. 2B). Within the LSA3 1229–1238 peptide, the P3-P7 region did not form any polar contacts with HLA-A*02:01 in contrast to several salt bridges and hydrogen bonds with HLA-B*08:01 (Fig. 3 and Supplementary Table 1). These TFold predicted differences in peptide/HLA docking modes underpin differences in efficiency of in vitro refolding yields. Differences in peptide/HLA docking modes also result in significantly different peptide surfaces displayed to TCRs. For example, in the case of LSA3 1229–1238/HLA-A*02:01 complexes, R5 is predicted to be exposed to the TCR. In contrast, for LSA3 1229–1238/HLA-B*08:01 complexes, R5 is buried in the HLA groove (Fig. 3, bottom panel and Supplementary Fig. 5). The orientation of R5 is dependent on the size and shape complementarity with residues that form the HLA C/D pockets, which differ between HLA-A*02:01 and HLA-B*08:01 isoform (Figs. 3 and 4, and Supplementary Fig. 5). To further analyze each *P. falciparum* peptide/HLA complex, we passed each TFold model through *Rosetta*'s InterfaceAnalyzer application to determine differences in solvent accessible area buried at the interface (ΔSASA interface) and interface free energy (ΔG separated)[84]. In general, TFold models of the eight identified *P. falciparum* peptides exhibited higher ΔSASA interface values and lower ΔG separated values when in complex with HLA-B*08:01 relative to the corresponding HLA-A*02:01 complexes (Supplementary Fig. 6). These observations are consistent with high yields of peptide/HLA-B*08:01 in vitro refolding, increased peptide/HLA-B*08:01 thermal stability, and an increased number of peptide-HLA-B*08:01 polar contacts from TFold models. One notable difference was MSP 823–832, which showed a clear preference for HLA-A*02:01 in experiments and in silico modeling: higher in vitro refolding yields, increased thermal stability, and lower ΔG separated values (Fig. 2B, Supplementary Fig. 3, and Supplementary Fig. 6), which we attribute to an increased number of polar contacts between MSP 823–823 with HLA-A*02:01 relative to HLA-B*08:01 (Fig. 4 and Supplementary Table 1). Thus, coupled with in vitro measurements, in silico TFold structure modeling and Rosetta analysis provide insights into the mechanisms underpinning binding of *P. falciparum* derived peptides to both HLA-A*02:01 and HLA-B*02:01. These models further provide potential immunogenicity profiles of *P. falciparum* derived peptides, suggesting distinct peptide-dependent and HLA allele-dependent differences in the molecular surfaces solvent exposed antigen surfaces to TCRs (Supplementary Fig. 5).

Expanding identified *P. falciparum* candidate epitopes to a broader HLA repertoire inclusive of malaria endemic populations

HLA haplotypes of different individuals in malaria endemic populations can determine whether they are at elevated or reduced risk of developing disease following *Plasmodium* infection, as only a subset of HLA allotypes can display potential peptide antigens[9,65–68]. To expand the repertoire of HLA alleles that could display *P. falciparum* derived peptides beyond HLA-A*02:01 and HLA-B*08:01, we utilized netMHCpan-4.1 to predict binding of our eight identified *P. falciparum* candidate epitopes against *(i)* 47 HLA-A-, B-, C-, E-, and G- alleles covering 99% of the global population, and *(ii)* 10,386 HLA-A-, B-, C-, E-, and G- alleles covering most known class I HLA molecules[17,50,77]. Of the 47 class I alleles that cover 99% of the global population, 25 of them overlap with the top 10 HLA-A, B,C alleles frequently observed in North Africa and Sub-Saharan Africa (Supplementary Data 1 and Materials and Methods). We compared netMHCpan-4.1 derived % rank binding affinity (%Rank BA) with BLOSUM62 identity scores of each allele relative to HLA-A*02:01, an experimentally confirmed binder of each peptide from our experiments (Fig. 5A, Supplementary Fig. 7, and Supplementary Data 1). For three of the eight *P. falciparum* derived 9-mer peptides (SSFP 93–103, MSP1 511–520, and MSP1 643–652), HLA-B*08:01 was among the highest affinity predictions (Fig. 5A), which agrees with our experimental observations that HLA-B*08:01 refolds in vitro with these peptides. Interestingly, we also identified HLA-C alleles (HLA-C*05:01, HLA-C*07:02, HLA-C*17:01) and HLA-E alleles (HLA-E*01:03) as the best predicted binders across the eight *P. falciparum* derived 9-mer peptides (Fig. 5A). Of these, HLA-C*07:02, HLA-C*17:01, and HLA-E*01:03 are commonly found in African populations[9,97–99]. From these predictions, HLA-C*07:02 was among the top binders

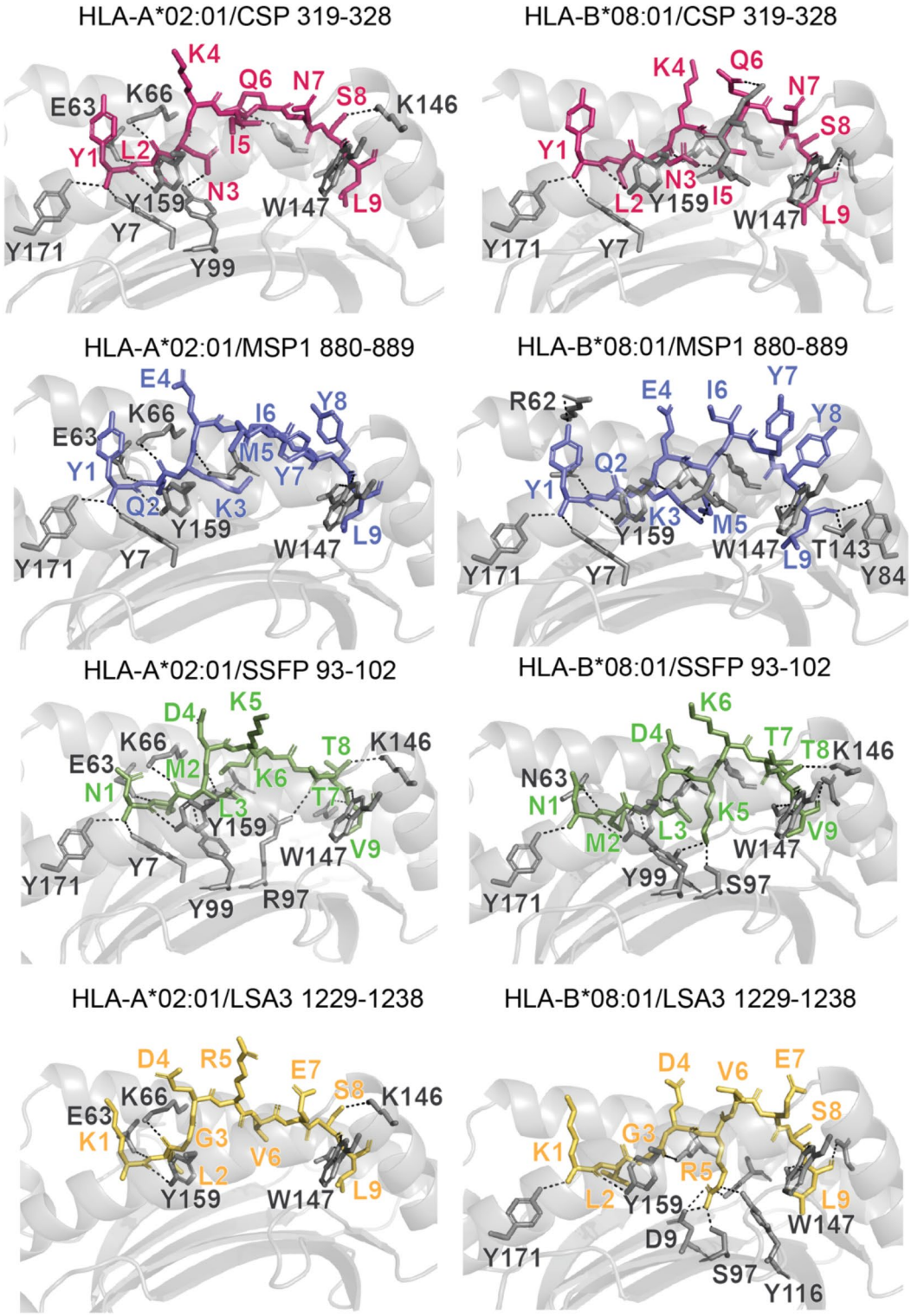

Fig. 3. In silico structural characterization of CSP 319–328, MSP1 880–889, SSFP 93–102, and LSA3 1229–1238 peptide binding to HLA-A*02:01 and HLA-B*08:01 grooves. Hydrogen bonds and salt bridge contacts between P1-P9 positions of *P. falciparum* candidate antigens CSP 319–328, MSP1 880–889, SSFP 93–102, and LSA3 1229–1238 are highlighted between residues in the HLA-A*02:01 or HLA-B*08:01 grooves as modeled by TFold. HLA groove residue numbering is consistent with the literature convention from X-ray crystallography studies.

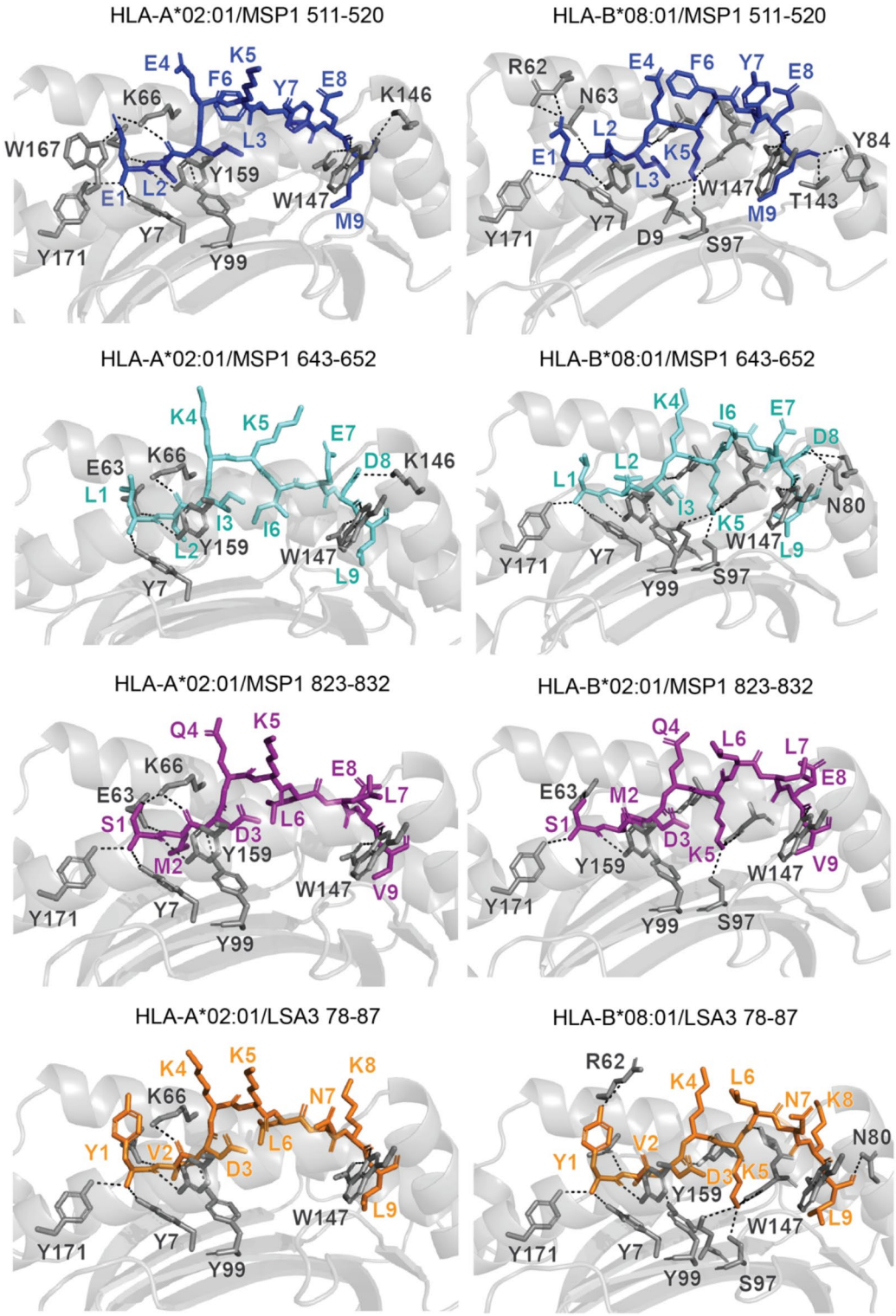

Fig. 4. In silico **structural characterization of MSP1 511–520, MSP1 643–652, MSP1 823–832, and LSA3 78–87 peptide binding to HLA-A*02:01 and HLA-B*08:01 grooves.** Hydrogen bonds and salt bridge contacts between P1-P9 positions of *P. falciparum* candidate antigens MSP1 511–520, MSP1 643–652, MSP1 823–832, and LSA3 78–87 are highlighted between residues in the HLA-A*02:01 or HLA-B*08:01 grooves as modeled by TFold. HLA groove residue numbering is consistent with the literature convention from X-ray crystallography studies.

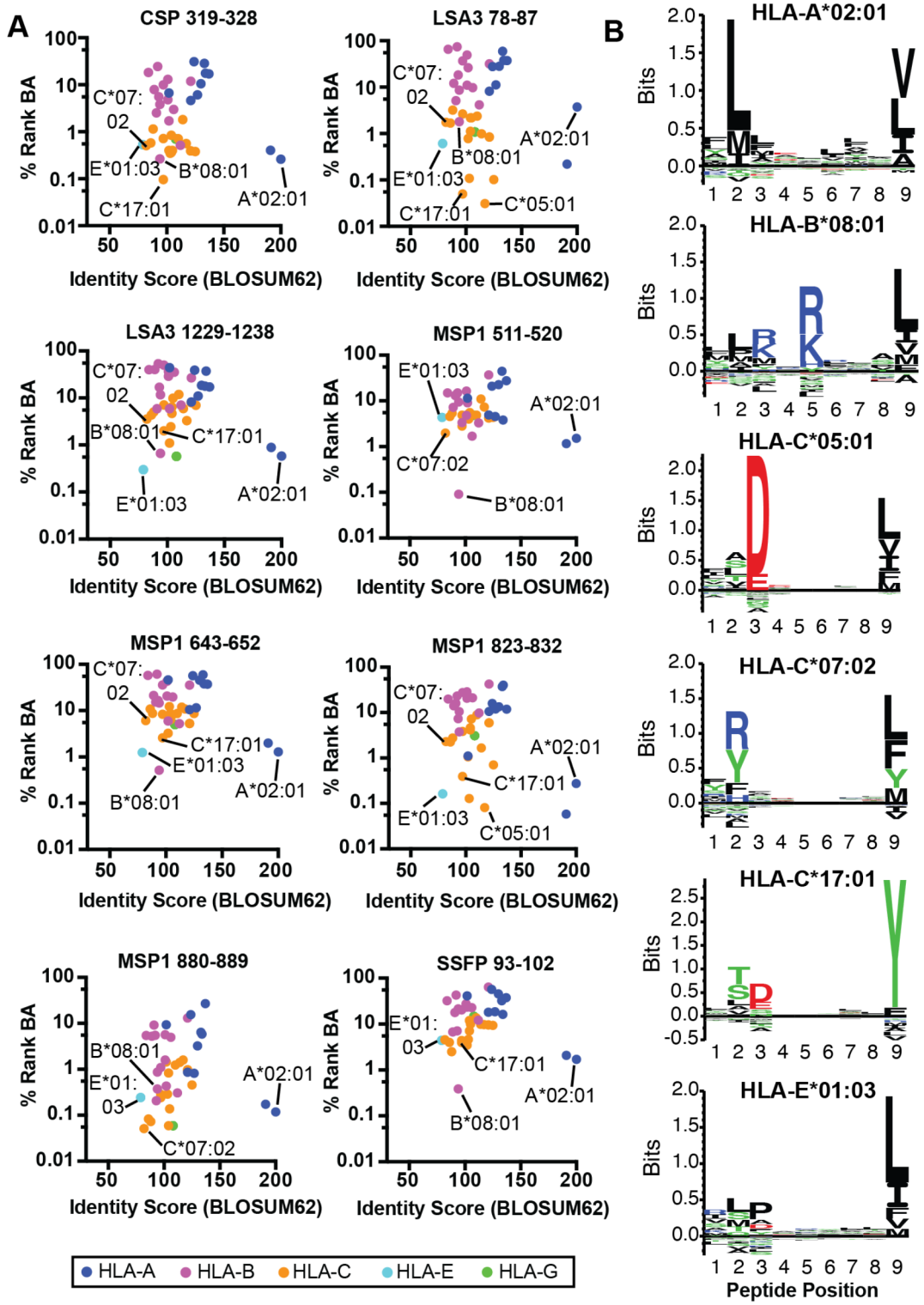

Fig. 5. In silico evaluation of the binding repertoire of eight identified *P. falciparum* candidate antigens across 47 class I HLA alleles covering 99% of the global population. (**A**) In silico netMHCpan-4.1 prediction of binding for eight identified *P. falciparum* candidate antigens with 47 class I HLA alleles covering 99% of the global population, including malaria endemic populations worldwide (see Supplementary Data 1 for full list of alleles). The % rank binding affinity (BA) from netMHCpan-4.1 is compared to the BLOSUM62 identity score for all HLA alleles relative to HLA-A*02:01. A selection of the top predicted peptide binding alleles are labeled for each peptide. (**B**) Peptide binding motifs (generated by netMHCpan-4.1 motif viewer) for top HLA alleles across the eight identified *P. falciparum* candidate antigens. HLA-C*07:02, HLA-C*17:01, and HLA-E*01:03 are commonly found in African populations.

for CSP 319–328 and MSP1 880–889 (Fig. 5A and Supplementary Data 1). HLA-C*17:01 was among the top binders for CSP 319–328, LSA3 78–87, MSP1 823–832, and MSP1 880–889 (Fig. 5A and Supplementary Data 1). HLA-E*01:03 was among the top binders for CSP 319–328, LSA3 78–87, LSA3 1229–1238, MSP1 823–832, and MSP1 880–889 (Fig. 5A and Supplementary Data 1). To identify features across different HLA classes that could be attributed to presentation of putative *P. falciparum* epitopes, we compared peptide motifs across HLA alleles with strongest %Rank BA for each peptide from netMHCpan-4.1. Based on sequence logos, the defining features for cross-presentation of the eight *P. falciparum* 9-mer peptides across all HLA-A-, B-, C-, E-, and G- alleles is a delicate balance between the identity of the hydrophobic residues at positions 2/9 (Leu, Val, Ile, Met, Tyr), the charged residues at position 3 (Asp), and the charge residues at positions 5 (Arg, Lys) (Fig. 5B).

To test these new predictions, we next performed additional in vitro refolding experiments for two pairs predicted to bind well (CSP 319–328/HLA-C*07:02 and LSA3 1229–1238/HLA-E*01:03) and one predicted to bind poorly (MSP1 643–652/HLA-E*01:03) (Table 1; Fig. 5A and Supplementary Data 1). In agreement with netMHCpan predictions, CSP 319–328/HLA-C*07:02/hβ2m and LSA3 1229–1238/HLA-E*01:03/hβ2m refolded efficiency with apparent refolding yields of ~37%, which was in stark contrast to the ~8% for MSP1 643–652/HLA-E*01:03/hβ2m (Fig. 6A, B; Table 1, and Supplementary Fig. 1C). The nanoDSF measured thermal stabilities also differed between the complexes with ~60 °C for CSP 319–328/HLA-C*07:02/hβ2m and LSA3 1229–1238/HLA-E*01:03/hβ2m relative to ~38 °C for MSP1 643–652/HLA-E*01:03/hβ2m (Fig. 6C; Table 1). MSP1 643–652/HLA-E*01:03/hβ2m also exhibited a minor population with higher stability of ~63 °C, indicating only a small population of stabilized HLA molecules (Fig. 6C). TFold modeling predicts several similar and distinct features of the polar contacts between HLA-C*07:02 and HLA-E*01:03 relative to HLA-A*02:01 and HLA-B*08:01 (Figs. 3 and 6D, and Supplementary Table 2). For example, polar contacts between CSP 319–328 residues N7 contacts residues in the HLA-C*07:02 groove but not any in the HLA-A*02:01 or HLA-B*07:01 groove, which results in a different peptide backbone conformation of the epitope (Figs. 3 and 6D). Together, our in silico analysis, supported by in vitro experiments, indicates that the *P. falciparum* 9-mer peptides have potential to be presented by a large proportion of HLA alleles outside of HLA-A*02:01 and HLA-B*08:01, including HLA-C*07:02 and HLA-E*01:03 commonly found in African populations[9,97–99].

Discussion

HLA class I mediated antigen presentation of *Plasmodium* derived peptides in the context of malaria remain underexplored, especially outside of the context of the well-studied HLA-A*02:01 allele[51–54]. The goal of our study was two-fold. First, we sought to obtain a list of candidate *Plasmodium* derived 9-mer epitopes beyond association with HLA-A*02:01. These newly identified peptides could serve as model systems to understand HLA-dependent immune responses during *Plasmodium* infection and have potential to serve as vaccine design candidates. Second, we aimed to decipher the biophysical and structural basis of *Plasmodium* derived peptide binding to HLAs to (i) identify features defining peptide-HLA cross-presentation across a range of malaria endemic populations, (ii) elucidate features underpinning stability of peptide-HLA complexes in malaria, and (iii) define unique molecular surfaces of *Plasmodium* derived epitopes for potential TCR recognition.

Here, by screening a subset of the *P. falciparum* proteome for peptides with potential for cross-presentation, we identified and experimentally characterized eight candidate *P. falciparum* 9-mer epitopes for HLA-A*02:01 and HLA-B*08:01 (Figs. 1 and 2; Table 1). Three of our identified epitopes have been previously shown to be HLA-A*02:01 binders. CSP 319–328 (YLNKIQNSL) has been shown to associate with HLA-A*02:01 resulting in robust $CD8^+$ T cell responses[56,58,100,101]. MSP1 643–652 (LLIKKIEDL) and MSP1 823–832 (SMDQKLLEV) have been shown to be presented by HLA-A*02:01 on the cell surface, though immunogenicity remains unknown[62]. We have expanded these known epitopes of *P. falciparum* beyond the context of HLA-A*02:01 by characterizing their binding to HLA-B*08:01 experimentally (Figs. 2 and 6) and to all known class I HLA alleles computationally (Fig. 5 and Supplementary Fig. 7). We also identified and characterized five new candidate *P. falciparum* 9-mer peptide antigens presented by HLA-A*02:01 and HLA-B*08:01 (Table 1): LSA3 78–87 (YVDKKLNKL), LSA3 1229–1238 (KLGERVESL), MSP1 511–520 (ELLEKFYEM), MSP1 880–889 (YQKEMIYYL), SSFP 93–102 (NMLDKKTTV). While our in vitro experiments confirm the eight *P. falciparum* derived 9-mer peptides associate with HLA-A*02:01 and HLA-B*08:01 (with various degrees of efficiency and stability), future experiments are required to explore their immunological properties. Robust $CD8^+$ T cell responses have been reported for CSP 319–328 [56,59,100,101], a peptide identified in our study. While we expect many of these *P. falciparum* derived peptide/HLA complexes will prove immunogenic through cytotoxic T cell activation assays and flow cytometry staining of $CD8^+$ αβ T cell populations[56,58,100,101], this requires extensive testing since many complicated variables underpin immunogenicity, such as processing of peptides by enzymes[102], peptide competition for the HLA[103], chaperone-mediated editing[4,104], cross-presentation by MHC-II molecules[105], polymorphic variation between species[106], and immune evasion[107,108]. Another caveat is that CSP, LSA3, MSP1, and SSFP proteins are expressed on the parasites at different stages of infection and infiltrate different host cell types (hepatocytes and erythrocytes). While hepatocytes express HLA and are capable of antigen presentation[109], it is commonly accepted that mature human erythrocytes lack appreciable HLA expression, which limits the potential for the peptide/HLA cell surface presentation of blood-stage associated *Plasmodium* proteins[107]. Under certain pathological conditions (i.e., infection by *Plasmodium* parasite), precursors to mature erythrocytes, such as erythroblasts, are capable of expressing $CD8^+$ T cell activating levels of antigen-presenting molecules[110–114]. Normocytes may also exhibit trace amounts of HLA molecules as residual populations resulting from nucleated precursors, but have not been attributed T cell activating properties[110–114]. Thus, given that there remains no clear evidence for a role of CD8 immunity in protection from blood-stage *P. falciparum* infection, the therapeutic potential of any epitopes identified within proteins expressed exclusively on blood-stage parasites is limited.

Proteins expressed in different *Plasmodium* species are likely to exhibit polymorphisms that could influence their HLA binding ability. Such polymorphisms could result in species dependent antigen presentation and

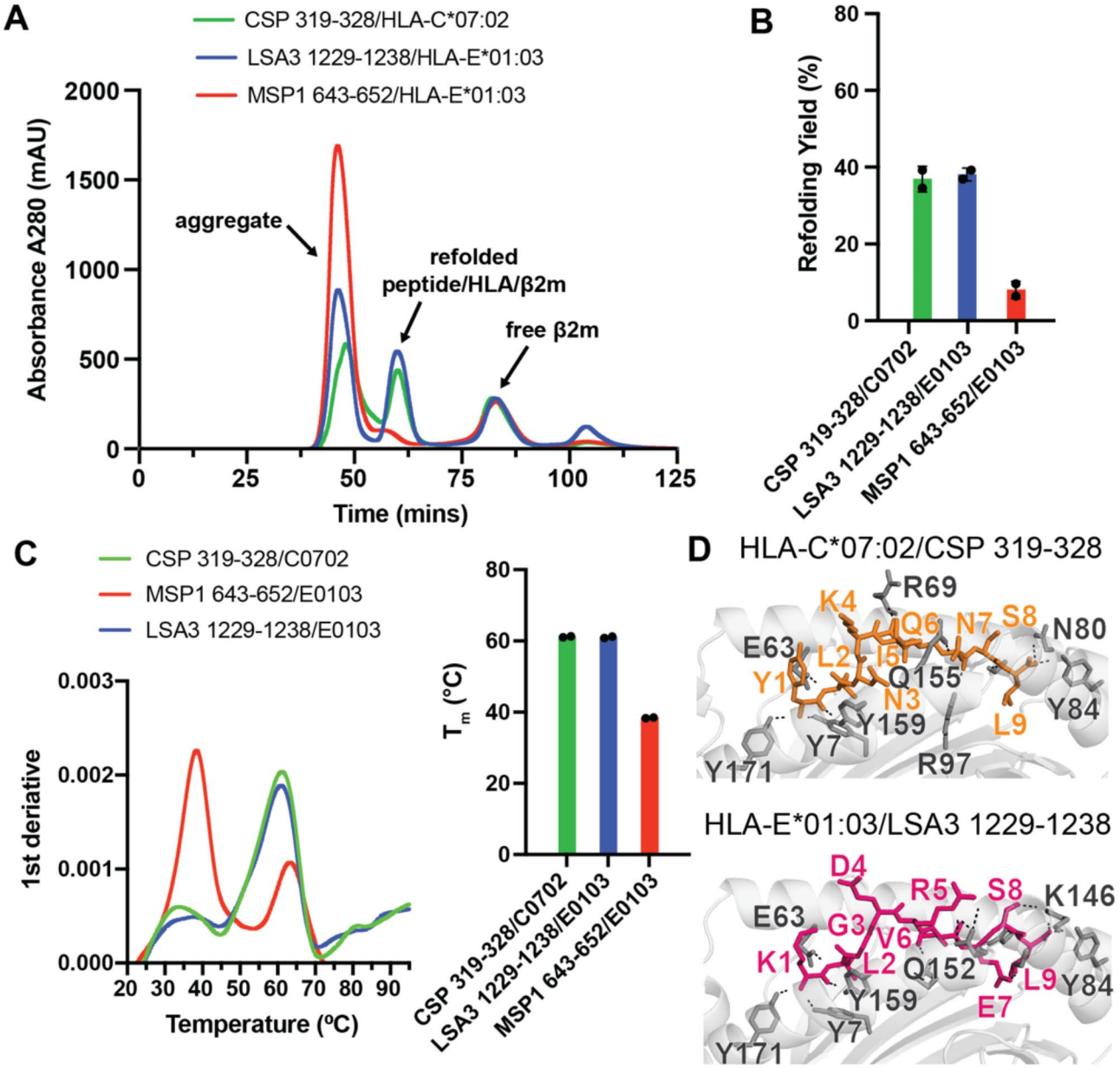

Fig. 6. In vitro validation and modeling of *P. falciparum* derived peptides with HLA-C*07:02 and HLA-E*01:03. (**A**) Size exclusion chromatography traces of in vitro refolded CSP 319–328/HLA-C*07:02/hβ2m (green), LSA3 1229–1238/HLA-E*01:03/hβ2m (blue), and MSP1 643–652/HLA-E*01:03 (red). (**B**) Summary of apparent in vitro refolding yields (%) determined from size exclusion chromatography traces for HLA-C*07:02/hβ2m or HLA-E*01:03/hβ2m with peptides derived from *Plasmodium*. (**C**) Left: First derivative of the denaturation nanoDSF spectra (F_{350}/F_{330}) for in vitro refolded complexes from panel A. Right: NanoDSF derived melting temperatures (T_m, °C) for different *P. falciparum* derived candidate peptide antigens in complex with either HLA-C*07:02/hβ2m or HLA-E*01:03/hβ2m. Error bars are derived from mean values ± standard deviation for two technical replicates ($n = 2$). (**D**) Hydrogen bonds and salt bridge contacts between P1-P9 positions of *P. falciparum* candidate antigens CSP 319–328/HLA-C*07:02 (top) and LSA3 1229–1238/HLA-E*01:03 (bottom) as modeled by TFold. HLA groove residue numbering is consistent with the literature convention from X-ray crystallography studies.

immunogenicity, which would influence vaccine design strategies. To probe this, we evaluated (i) the amino acid sequence conservation and (ii) HLA-A*02:01 and HLA-B*08:01 binding ability (netMHCpan prediction) for several peptides across *Plasmodium* species implicated in malaria: *P. falciparum*, *P. malariae*, *P. vivax*, *P. ovale*, *and P. knowlesi*[115]. We identified two peptides (i.e., MSP1 643–652 and MSP1 880–889) that were quite conserved and exhibited similar netMHCpan predicted affinity to HLA-A*02:01 and HLA-B*08:01 across different *Plasmodium* species (Supplementary Fig. 8). For example, P1 Leu and P9 Leu anchor positions were strictly conserved across *Plasmodium* species for MSP1 643–652, resulting in similar predicted binding to

HLA-A*02:01 and HLA-B*08:01. In contrast, we also identified two peptides (i.e., CSP 319–328 and MSP1 511–520) that were not well conserved and were predicted by netMHCpan to have highly variable affinities with HLA-A*02:01 and HLA-B*08:01 across different *Plasmodium* species (Supplementary Fig. 8). Together, these results highlight the complexity of the problem: a good candidate for malarial vaccine design should exhibit both cross-presentation of HLA alleles in the population of interest while maintaining enough conservation through different *Plasmodium* species to be broadly applicable.

To our knowledge, there is only a single published crystal structure of a class I HLA presenting a *Plasmodium* derived peptide antigen: HLA-B*53:01 with 9-mer KPIVQYDNF derived from liver stage antigen 1 residues 1786–1793 [71]. Our report presents eight candidate *P. falciparum* 9-mer peptide/HLA complexes have been characterized in terms of their in vitro refolding ability, thermal stability, and structural modeling with HLA-A*02:01 and HLA-B*08:01. We observed both peptide- and HLA-dependent behavior of the resulting *P. falciparum* 9-mer peptide/HLA complexes. In terms of yields of protein generated, CSP 319–328, SSFP 93–102, and MSP1 880–889 are likely to serve as the best candidates for cross-presentation on HLA-A*02:01 and HLA-B*08:01 due to their increased refolding yields and high thermal stabilities relative to other peptides, while LSA3 78–87 is the least promising candidate (Fig. 2; Table 1). Computational analysis expands the set of possible alleles for our eight candidate *P. falciparum* 9-mer peptides beyond just HLA-A*02:01 and HLA-B*08:01. We predict that due to similar preferences in peptide anchor positions, HLA-C alleles (HLA-C*05, HLA-C*07, HLA-C*17) and HLA-E alleles (HLA-E*01:03) commonly found in malaria endemic populations will strongly bind and cross-present several of our identified *P. falciparum* derived peptides – two of these predictions were confirmed experimentally – CSP 319–328/HLA-C*07:02 and LSA3 1229–1238/HLA-E*01:03 (Figs. 5 and 6, and Supplementary Fig. 7). In terms of ability to present *P. falciparum* derived 9-mer peptides: HLA-A*02:01 prefers Leu/Met/Ile at position 2 and Val/Leu/Ile at position 9, HLA-B*08:01 prefers Leu at position 2, Arg/Lyse at position 3, Arg/Lyse at position 5, and Leu/Ile/Val at position 9, HLA-C*07:02 prefers Arg, Tyr, Phe at position 2 and Leu/Phe/Tyr at position 9, HLA-C*17:01 prefers Thr/Ser/Leu at position 2 and Tyr/Phe at position 9, and HLA-E*01:03 prefers Leu at position 2 and Leu/Ile/Phe at position 9[11] (Fig. 5B). Thus, the defining features for cross-presentation of the eight *P. falciparum* 9-mer peptides across all HLA-A-, B-, C-, E-, and G- alleles is a delicate balance between the identity of the hydrophobic residues at positions 2/9, the charged residues at position 3, and the charge residues at positions 5 (Fig. 5B). The cross-presentation motifs, together with biophysical stability, structure, and immunogenicity profiles, provide valuable information on the most relevant vaccine candidates for antigen presentation of "public" antigens against HLA diverse populations[100,116]. Future studies in populations with high HLA diversity, such as malaria endemic populations where pathogens drive selection pressure for certain alleles, would benefit from quantifying HLA frequencies in those populations together with the identifying potential epitopes for cross-presentation by combined computational and experimental methods[13,48,100,116,117].

From a bird's eye view, our in silico models of *P. falciparum* derived peptide/HLA-A*02:01 and HLA-B*08:01 complexes provide novel insights into experimental results while informing the potential for TCR recognition. We find an excellent correlation between in vitro refolding yields, thermal stability, and interface characteristics of TFold models, which we attribute to differences in peptide- and HLA allele-dependent differences in peptide-HLA polar contacts in the HLA groove (Figs. 3 and 4, Supplementary Fig. 5, Supplementary Fig. 6, and Supplementary Table 1). TFold modeling explains experimental evidence that, in general, *P. falciparum* derived peptide/HLA-B*08:01 complexes are much more stable. TFold highlight the presence of increased peptide-HLA interactions in the secondary P5 anchor position, consistent with known peptide binding modes for HLA-B*08:01[11,118]. In terms of potential for TCR recognition, it is extremely important to structurally characterize (either computationally or experimentally) peptide/HLA conformations since in some cases the same peptide can be presented in similar conformations across different HLA alleles, and other cases the same peptide can be presented in different conformations even within the same HLA supertype, which depends on subtle polymorphic variables in the HLA groove[11]. For example, HIV Nef71-79 peptide is presented with a similar conformation by HLA-B*07:02, HLA-B*81:01, HLA-B*42:01, and HLA-B*42:02 [119]. In contrast, HLA-B*42:01 and HLA-B*81:01 present the HIV Gag180–188 in distinct conformations[119]. Here, we find that solvent accessible molecular surfaces of *P. falciparum* derived peptides on HLA-A*02:01, HLA-B*08:01, HLA-E*01:03, and HLA-C*07:02 are extremely HLA allele dependent, which suggests that presentation of *P. falciparum* epitopes likely results in unique recognition of different TCR repertoires (Figs. 3, 4 and 6, and Supplementary Fig. 5). Of the eight identified peptides here, only CSP 319–328 has been shown to simulate $CD8^+$ T cell responses[56,58,100,101]. Our modeling results provide evidence that K4, Q6, N7, and S8 are the peptide antigen residues that could be recognized by $CD8^+$ αβTCRs; this is a special case where the CSP 319–328 molecular surface presented to TCRs is similar between HLA-A*02:01 and HLA-B*08:01 (Fig. 3 and Supplementary Fig. 5). We resorted to computational modeling since the peptide/HLA complexes here did not form well-diffracting crystals. Complementary experimental approaches could be used to characterize peptide/HLA structures, including cryo-electron microscopy (cryoEM) and solution NMR. In terms of cryo-EM, the pHLA complexes (~45 kDa) are an extremely challenging target for the current size limitations of the technique[120]. There are cryo-EM structures of pMHC molecules with other proteins (i.e., chaperones, T cell receptors) are available since the complex is much larger (~90 kDa)[121–123]. Solution NMR is also a potential alternative but requires multiple isotopically labeled samples, a suite of 3D NMR experiments, and challenging NMR resonance assignment[4,17,82]. Computational modeling (i.e., by TFold) comes with several important caveats and limitations. First, while Mikhaylov et al. provide extensive evidence that error scores (100-pLDDT averaged over the peptide core) below 6.8 are reflective of accurate models[73], these are still only in silico predictions and require further experimental probing. In rare cases, TFold can generate fairly inaccurate models with low error scores[73]. Second, it is well known that some peptides within HLA grooves undergo significant conformational flexibility that is important for both accuracy of modeling the TCR receptor interface and potential for immunogencity[17,124], which would

limit the validity of static TFold models. The dynamic features of these peptide/HLA complexes could be probed by solution NMR, molecular dynamics simulations, and HDX-MS[17,124–127].

In summary, we have identified a set of *P. falciparum* derived peptides with the potential for cross-presentation by HLA-A*02:01, HLA-B*08:01, HLA-C*07:02, and HLA-E*01:03 alleles commonly represented in malaria endemic African populations. A combination of in vitro refolding, biophysical characterization, and in silico structure modeling provides insights on the key features for cross-presentation, which requires future experimental probing for cellular immunity and the potential for presentation for polymeric epitopes across different *Plasmodium* species.

Data availability

All source data are provided in Supplementary Data 1. Scripts for global HLA repertoire characterization (via netMHCpan-4.1 and BLOSUM62) and TFold models are available at GitHub: https://github.com/mcshanlab/Frooman-et-al_Malaria_2024.

Received: 11 October 2024; Accepted: 25 February 2025
Published online: 12 March 2025

References

1. Rock, K. L., Reits, E. & Neefjes, J. Present yourself! By MHC class I and MHC class II molecules. *Trends Immunol.* **37**, 724–737 (2016).
2. Rossjohn, J. et al. T cell antigen receptor recognition of antigen-presenting molecules. *Annu. Rev. Immunol.* **33**, 169–200 (2015).
3. Blum, J. S., Wearsch, P. A. & Cresswell, P. Pathways of antigen processing. *Annu. Rev. Immunol.* **31**, 443–473 (2013).
4. McShan, A. C. et al. Peptide exchange on MHC-I by TAPBPR is driven by a negative allostery release cycle. *Nat. Chem. Biol.* **14**, 811–820 (2018).
5. Boyle, L. H. et al. Tapasin-related protein TAPBPR is an additional component of the MHC class I presentation pathway. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 3465–3470 (2013).
6. Blees, A. et al. Structure of the human MHC-I peptide-loading complex. *Nature* **551**, 525 (2017).
7. Trolle, T. et al. The length distribution of class I restricted T cell epitopes is determined by both peptide supply and MHC allele specific binding preference. *J. Immunol. Baltim. Md. 1950.* **196**, 1480–1487 (2016).
8. Bell, M. J. et al. The peptide length specificity of some HLA class I alleles is very broad and includes peptides of up to 25 amino acids in length. *Mol. Immunol.* **46**, 1911–1917 (2009).
9. Gonzalez-Galarza, F. F., Christmas, S., Middleton, D. & Jones, A. R. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Res.* **39**, D913–919 (2011).
10. Saper, M. A., Bjorkman, P. J. & Wiley, D. C. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J. Mol. Biol.* **219**, 277–319 (1991).
11. Nguyen, A. T., Szeto, C. & Gras, S. The pockets guide to HLA class I molecules. *Biochem. Soc. Trans.* **49**, 2319–2331 (2021).
12. Hunt, D. F. et al. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* **255**, 1261–1263 (1992).
13. Sarkizova, S. et al. A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nat. Biotechnol.* **38**, 199–209 (2020).
14. Hamley, I. W. Peptides for vaccine development. *ACS Appl. Bio Mater.* **5**, 905–944 (2022).
15. Kaeuferle, T., Krauss, R., Blaeschke, F., Willier, S. & Feuchtinger, T. Strategies of adoptive T -cell transfer to treat refractory viral infections post allogeneic stem cell transplantation. *J. Hematol. Oncol. J. Hematol. Oncol.* **12**, 13 (2019).
16. Lie-Andersen, O. et al. Impact of peptide:hla complex stability for the identification of SARS-CoV-2-specific CD8 + T cells. *Front. Immunol.* **14**, 1151659 (2023).
17. McShan, A. C. et al. Conformational plasticity of RAS Q61 family of neoepitopes results in distinct features for targeted recognition. *Nat. Commun.* **14**, 8204 (2023).
18. Kaseke, C. et al. HLA class-I-peptide stability mediates CD8 + T cell immunodominance hierarchies and facilitates HLA-associated immune control of HIV. *Cell. Rep.* **36**, 109378 (2021).
19. Poespoprodjo, J. R. Douglas, N. M., Ansong, D., Kho, S. & Anstey, N. M. Malaria. *Lancet Lond. Engl.* **402**, 2328–2345 (2023).
20. Moussaoui, D. et al. Mechanism of small molecule Inhibition of plasmodium falciparum myosin A informs antimalarial drug design. *Nat. Commun.* **14**, 3463 (2023).
21. Sibomana, O. et al. Routine malaria vaccination in Africa: a step toward malaria eradication? *Malar. J.* **24**, 1 (2025).
22. Barrett, J. R. et al. Analysis of the diverse antigenic landscape of the malaria protein RH5 identifies a potent vaccine-induced human public antibody clonotype. *Cell* **187**, 4964–4980e21 (2024).
23. Laurens, M. B. & RTS. S/AS01 vaccine (Mosquirix™): an overview. *Hum. Vaccines Immunother.* **16**, 480–489 (2019).
24. RTS,S Clinical Trials Partnership. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. *Lancet Lond. Engl.* **386**, 31–45 (2015).
25. Datoo, M. S. et al. Safety and efficacy of malaria vaccine candidate R21/Matrix-M in African children: a multicentre, double-blind, randomised, phase 3 trial. *Lancet Lond. Engl.* **403**, 533–544 (2024).
26. Duffy, P. E., Gorres, J. P., Healy, S. A. & Fried, M. Malaria vaccines: a new era of prevention and control. *Nat. Rev. Microbiol.* **22**, 756–772 (2024).
27. da Veiga, G. T. S., Moriggi, M. R., Vettorazzi, J. F., Müller-Santos, M. & Albrecht, L. Plasmodium Vivax vaccine: what is the best way to go? *Front. Immunol.* **13**, 910236 (2022).
28. Schmit, N. et al. The public health impact and cost-effectiveness of the R21/Matrix-M malaria vaccine: a mathematical modelling study. *Lancet Infect. Dis.* **S1473-3099** (23), 00816–00812. https://doi.org/10.1016/S1473-3099(23)00816-2 (2024).
29. Kurup, S. P., Butler, N. S. & Harty, J. T. T cell-mediated immunity to malaria. *Nat. Rev. Immunol.* **19**, 457–471 (2019).
30. da Lima-Junior, J., Pratt-Riccio, L. R. & C. & Major histocompatibility complex and malaria: focus on plasmodium Vivax infection. *Front. Immunol.* **7**, 13 (2016).
31. Molina-Cruz, A., Lehmann, T. & Knöckel, J. Could culicine mosquitoes transmit human malaria? *Trends Parasitol.* **29**, 530–537 (2013).
32. Guttery, D. S., Holder, A. A. & Tewari, R. Sexual development in plasmodium: lessons from functional analyses. *PLoS Pathog.* **8**, e1002404 (2012).
33. Prudêncio, M., Rodriguez, A. & Mota, M. M. The silent path to thousands of merozoites: the plasmodium liver stage. *Nat. Rev. Microbiol.* **4**, 849–856 (2006).
34. Kumar, H., Tolia, N. H. & Getting The structural biology of malaria invasion. *PLoS Pathog.* **15**, e1007943 (2019).
35. Coppi, A. et al. The malaria circumsporozoite protein has two functional domains, each with distinct roles as sporozoites journey from mosquito to mammalian host. *J. Exp. Med.* **208**, 341–356 (2011).

36. Dijkman, P. M. et al. Structure of the merozoite surface protein 1 from plasmodium falciparum. *Sci. Adv.* **7**, eabg0465 (2021).
37. Collins, K. A., Snaith, R., Cottingham, M. G., Gilbert, S. C. & Hill, A. V. S. Enhancing protective immunity to malaria with a highly Immunogenic virus-like particle vaccine. *Sci. Rep.* **7**, 46621 (2017).
38. Wu, R. L. et al. Low-Dose subcutaneous or intravenous monoclonal antibody to prevent malaria. *N Engl. J. Med.* **387**, 397–407 (2022).
39. Holder, A. A. The carboxy-terminus of merozoite surface protein 1: structure, specific antibodies and immunity to malaria. *Parasitology* **136**, 1445–1456 (2009).
40. Parra, M. et al. Characterization of conserved T- and B-Cell epitopes in plasmodium falciparum major merozoite surface protein 1. *Infect. Immun.* **68**, 2685–2691 (2000).
41. Daubersies, P. et al. Protection against plasmodium falciparum malaria in chimpanzees by immunization with the conserved pre-erythrocytic liver-stage antigen 3. *Nat. Med.* **6**, 1258–1263 (2000).
42. Toure-Balde, A. et al. Evidence for multiple B- and T-cell epitopes in plasmodium falciparum liver-stage antigen 3. *Infect. Immun.* **77**, 1189–1196 (2009).
43. Ramelow, J., Keleta, Y., Niu, G., Wang, X. & Li, J. Plasmodium parasitophorous vacuole membrane protein Pfs16 promotes malaria transmission by Silencing mosquito immunity. *J. Biol. Chem.* **299**, 104824 (2023).
44. Keleta, Y., Ramelow, J., Cui, L. & Li, J. Molecular interactions between parasite and mosquito during midgut invasion as targets to block malaria transmission. *NPJ Vaccines.* **6**, 140 (2021).
45. Gibbins, M. P. et al. Importance of the immunodominant CD8 + T cell epitope of plasmodium Berghei circumsporozoite protein in Parasite- and Vaccine-Induced protection. *Infect. Immun.* **88**, e00383–e00320 (2020).
46. Blank, A. et al. Immunization with full-length plasmodium falciparum merozoite surface protein 1 is safe and elicits functional cytophilic antibodies in a randomized first-in-human trial. *NPJ Vaccines.* **5**, 10 (2020).
47. Long, C. A. & Zavala, F. Immune responses in malaria. *Cold Spring Harb Perspect. Med.* **7**, a025577 (2017).
48. Tang, Y., Lin, Y., Mao, Y., Dong, M. & Wang, H. Cross-binding between plasmodium falciparum CTL epitopes and HLA class I molecules. *Immunol. Invest.* **32**, 31–41 (2003).
49. Doolan, D. L. et al. Degenerate cytotoxic T cell epitopes from P. falciparum restricted by multiple HLA-A and HLA-B supertype alleles. *Immunity* **7**, 97–112 (1997).
50. Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif Deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454 (2020).
51. Heide, J., Vaughan, K. C., Sette, A., Jacobs, T. & Schulze Zur Wiesch, J. Comprehensive review of human plasmodium falciparum-Specific CD8 + T cell epitopes. *Front. Immunol.* **10**, 397 (2019).
52. Kusi, K. A. et al. Identification of plasmodium falciparum circumsporozoite protein-specific CD8 + T cell epitopes in a malaria exposed population. *PloS One.* **15**, e0228177 (2020).
53. Belmonte, M. et al. Immunodominant T cell peptides from four candidate malarial antigens as biomarkers of protective immunity against malaria. *Vaccine* **41**, 1265–1273 (2023).
54. Doolan, D. L., Wizel, B. & Hoffman, S. L. Class I HLA-restricted cytotoxic T lymphocyte responses against malaria–elucidation on the basis of HLA peptide binding motifs. *Immunol. Res.* **15**, 280–305 (1996).
55. Sedegah, M. et al. Identification of minimal human MHC-restricted CD8 + T-cell epitopes within the plasmodium falciparum circumsporozoite protein (CSP). *Malar. J.* **12**, 185 (2013).
56. Blum-Tirouvanziam, U. et al. Localization of HLA-A2.1-restricted T cell epitopes in the circumsporozoite protein of plasmodium falciparum. *J. Immunol. Baltim. Md. 1950.* **154**, 3922–3931 (1995).
57. López, J. A. et al. A synthetic malaria vaccine elicits a potent CD8(+) and CD4(+) T lymphocyte immune response in humans. Implications for vaccination strategies. *Eur. J. Immunol.* **31**, 1989–1998 (2001).
58. González, J. M. et al. HLA-A*0201 restricted CD8 + T-lymphocyte responses to malaria: identification of new plasmodium falciparum epitopes by IFN-gamma ELISPOT. *Parasite Immunol.* **22**, 501–514 (2000).
59. Aidoo, M. et al. Identification of conserved antigenic components for a cytotoxic T lymphocyte-inducing vaccine against malaria. *Lancet Lond. Engl.* **345**, 1003–1007 (1995).
60. Hill, A. V. et al. Molecular analysis of the association of HLA-B53 and resistance to severe malaria. *Nature* **360**, 434–439 (1992).
61. Malik, A., Egan, J. E., Houghten, R. A., Sadoff, J. C. & Hoffman, S. L. Human cytotoxic T lymphocytes against the Plasmodium falciparum circumsporozoite protein. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 3300–3304 (1991).
62. Carralot, J. P., Lemmel, C., Stevanovic, S. & Pascolo, S. Mass spectrometric identification of an HLA-A*0201 epitope from plasmodium falciparum MSP-1. *Int. Immunol.* **20**, 1451–1456 (2008).
63. Aidoo, M. et al. Cytotoxic T-lymphocyte epitopes for HLA-B53 and other HLA types in the malaria vaccine candidate liver-stage antigen 3. *Infect. Immun.* **68**, 227–232 (2000).
64. Dodoo, D. et al. Measuring naturally acquired immune responses to candidate malaria vaccine antigens in Ghanaian adults. *Malar. J.* **10**, 168 (2011).
65. Sanchez-Mazas, A. et al. The HLA-B landscape of Africa: signatures of pathogen-driven selection and molecular identification of candidate alleles to malaria protection. *Mol. Ecol.* **26**, 6238–6252 (2017).
66. van Janse, W. J., de Kock, A., Bester, C. & Kloppers, J. F. HLA major allele group frequencies in a diverse population of the free state Province, South Africa. *Heliyon* **7**, e06850 (2021).
67. Garamszegi, L. Z. Global distribution of malaria-resistant MHC-HLA alleles: the number and frequencies of alleles and malaria risk. *Malar. J.* **13**, 349 (2014).
68. Hill, A. V. et al. Common West African HLA antigens are associated with protection from severe malaria. *Nature* **352**, 595–600 (1991).
69. Tadros, D. M., Eggenschwiler, S., Racle, J. & Gfeller, D. The MHC motif atlas: a database of MHC binding specificities and ligands. *Nucleic Acids Res.* **51**, D428–D437 (2023).
70. Gupta, S., Nerli, S., Kutti Kandy, S., Mersky, G. L. & Sgourakis, N. G. HLA3DB: comprehensive annotation of peptide/hla complexes enables blind structure prediction of T cell epitopes. *Nat. Commun.* **14**, 6349 (2023).
71. Smith, K. J. et al. Bound water structure and polymorphic amino acids act together to allow the binding of different peptides to MHC class I HLA-B53. *Immunity* **4**, 215–228 (1996).
72. Yanover, C. & Bradley Large-scale characterization of peptide-MHC binding landscapes with structural simulations. *Proc. Natl. Acad. Sci.* **108**, 6981–6986 (2011).
73. Mikhaylov, V. et al. Accurate modeling of peptide-MHC structures with alphafold. *Struct. Lond. Engl. 1993.* **32**, 228–241e4 (2024).
74. Bhasin, M., Lata, S. & Raghava, G. P. S. TAPPred prediction of TAP-binding peptides in antigens. *Methods Mol. Biol. Clifton NJ.* **409**, 381–386 (2007).
75. Thomsen, M. C. F. & Nielsen, M. Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res.* **40**, W281–W287 (2012).
76. Styczynski, M. P., Jensen, K. L., Rigoutsos, I. & Stephanopoulos, G. BLOSUM62 miscalculations improve search performance. *Nat. Biotechnol.* **26**, 274–275 (2008).
77. Weiskopf, D. et al. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8 + T cells. *Proc. Natl. Acad. Sci. U S A.* **110**, E2046–2053 (2013).

78. Garboczi, D. N., Hung, D. T. & Wiley, D. C. HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in Escherichia coli and complexed with single antigenic peptides. *Proc. Natl. Acad. Sci.* **89**, 3429–3433 (1992).
79. Studier, F. W. Stable expression clones and auto-induction for protein production in E. coli. *Methods Mol. Biol. Clifton NJ*. **1091**, 17–32 (2014).
80. McShan, A. C. et al. Molecular determinants of chaperone interactions on MHC-I for folding and antigen repertoire selection. *Proc. Natl. Acad. Sci. U S A*. **116**, 25602–25613 (2019).
81. Gasteiger, E. et al. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788 (2003).
82. McShan, A. C. et al. TAPBPR employs a ligand-independent Docking mechanism to chaperone MR1 molecules. *Nat. Chem. Biol.* **18**, 859–868 (2022).
83. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
84. Stranges, P. B. & Kuhlman, B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci. Publ Protein Soc.* **22**, 74–82 (2013).
85. Gardner, M. J. et al. Genome sequence of the human malaria parasite plasmodium falciparum. *Nature* **419**, 498–511 (2002).
86. Drijfhout, J. W., Brandt, R. M., D'Amaro, J., Kast, W. M. & Melief, C. J. Detailed motifs for peptide binding to HLA-A*0201 derived from large random sets of peptides using a cellular binding assay. *Hum. Immunol.* **43**, 1–12 (1995).
87. Perez, M. A. S., Bassani-Sternberg, M., Coukos, G., Gfeller, D. & Zoete, V. Analysis of secondary structure biases in naturally presented HLA-I ligands. *Front. Immunol.* **10**, 2731 (2019).
88. Toor, J. S. et al. A recurrent mutation in anaplastic lymphoma kinase with distinct neoepitope conformations. *Front. Immunol.* **9**, 99 (2018).
89. McMurtrey, C. et al. Toxoplasma gondii peptide ligands open the gate of the HLA class I binding groove. *eLife* **5**, e12556.
90. Gorga, J. et al. (ed, C.) Comparison of the secondary structures of human class I and class II major histocompatibility complex antigens by fourier transform infrared and circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. U S A* **86** 2321–2325 (1989).
91. Harndahl, M. et al. Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity. *Eur. J. Immunol.* **42**, 1405–1416 (2012).
92. Saikia, A. & Springer, S. Peptide-MHC I complex stability measured by nanoscale differential scanning fluorimetry reveals molecular mechanism of thermal denaturation. *Mol. Immunol.* **136**, 73–81 (2021).
93. Hellman, L. M. et al. Differential scanning fluorimetry based assessments of the thermal and kinetic stability of peptide-MHC complexes. *J. Immunol. Methods*. **432**, 95–101 (2016).
94. Sun, Y. et al. Universal open MHC-I molecules for rapid peptide loading and enhanced complex stability across HLA allotypes. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2304055120 (2023).
95. Szeto, C., Lobos, C. A., Nguyen, A. T. & Gras, S. T. C. R. Recognition of Peptide–MHC-I: rule makers and breakers. *Int. J. Mol. Sci.* **22**, 68 (2020).
96. Reid, S. W. et al. Antagonist HIV-1 gag peptides induce structural changes in HLA B8. *J. Exp. Med.* **184**, 2279–2286 (1996).
97. Matte, C., Lacaille, J., Zijenah, L., Ward, B. & Roger, M. HLA-G and HLA-E polymorphisms in an Indigenous African population. The ZVITAMBO study group. *Hum. Immunol.* **61**, 1150–1156 (2000).
98. Nemat-Gorgani, N. et al. Diversity of KIR, HLA class I, and their interactions in seven populations of Sub-Saharan Africans. *J. Immunol. Baltim. Md. 1950*. **202**, 2636–2647 (2019).
99. Tshabalala, M. et al. High resolution HLA ∼A, ∼B, ∼C, ∼DRB1, ∼DQA1, and ∼DQB1 diversity in South African populations. *Front. Genet.* **13**, 711944 (2022).
100. Prato, S., Fleming, J., Schmidt, C. W., Corradin, G. & Lopez, J. A. Cross-presentation of a human malaria CTL epitope is conformation dependent. *Mol. Immunol.* **43**, 2031–2036 (2006).
101. Li, X. et al. Human CD8 + T cells mediate protective immunity induced by a human malaria vaccine in human immune system mice. *Vaccine* **34**, 4501–4506 (2016).
102. Li, L., Batliwala, M. & Bouvier, M. ERAP1 enzyme-mediated trimming and structural analyses of MHC I-bound precursor peptides yield novel insights into antigen processing and presentation. *J. Biol. Chem.* **294**, 18534–18544 (2019).
103. Adorini, L. & Nagy, Z. A. Peptide competition for antigen presentation. *Immunol. Today*. **11**, 21–24 (1990).
104. Lan, H. et al. Exchange catalysis by Tapasin exploits conserved and allele-specific features of MHC-I molecules. *Nat. Commun.* **12**, 4236 (2021).
105. Riley, E. M. et al. MHC and malaria: the relationship between HLA class II alleles and immune responses to plasmodium falciparum. *Int. Immunol.* **4**, 1055–1063 (1992).
106. Huang, H. Y. et al. Genetic polymorphism of plasmodium falciparum circumsporozoite protein on Bioko Island, Equatorial Guinea and global comparative analysis. *Malar. J.* **19**, 245 (2020).
107. Moser, K. A. et al. Strains used in whole organism plasmodium falciparum vaccine trials differ in genome structure, sequence, and Immunogenic potential. *Genome Med.* **12**, 6 (2020).
108. Flanagan, K. L., Wilson, K. L. & Plebanski, M. Polymorphism in liver-stage malaria vaccine candidate proteins: immune evasion and implications for vaccine design. *Expert Rev. Vaccines*. **15**, 389–399 (2016).
109. Chen, M., Tabaczewski, P., Truscott, S. M., Van Kaer, L. & Stroynowski, I. Hepatocytes express abundant surface class I MHC and efficiently use transporter associated with antigen processing, Tapasin, and low molecular weight polypeptide proteasome subunit components of antigen processing and presentation pathway. *J. Immunol. Baltim. Md. 1950*. **175**, 1047–1055 (2005).
110. Imai, T. et al. CD8(+) T cell activation by murine erythroblasts infected with malaria parasites. *Sci. Rep.* **3**, 1572 (2013).
111. Botto, M., So, A. K., Giles, C. M., Mason, P. D. & Walport, M. J. HLA class I expression on erythrocytes and platelets from patients with systemic lupus erythematosus, rheumatoid arthritis and from normal subjects. *Br. J. Haematol.* **75**, 106–111 (1990).
112. Imai, T. et al. Cytotoxic activities of $CD8^+$ T cells collaborate with macrophages to protect against blood-stage murine malaria. *eLife* **4**, e04232 (2015).
113. Everett, E. T., Scornik, J. C., Davis, G. & Kao, K. J. Induction of erythrocyte HLA expression during interferon treatment and HIV infection. *Hum. Immunol.* **29**, 14–22 (1990).
114. de Villartay, J. P., Rouger, P., Muller, J. Y. & Salmon, C. HLA antigens on peripheral red blood cells: analysis by flow cytofluorometry using monoclonal antibodies. *Tissue Antigens*. **26**, 12–19 (1985).
115. Kho, S. et al. Diagnostic performance of a 5-plex malaria immunoassay in regions co-endemic for plasmodium falciparum, P. vivax, P. knowlesi, P. malariae and P. ovale. *Sci. Rep.* **12**, 7286 (2022).
116. Colbert, J. D., Cruz, F. M. & Rock, K. L. Cross-presentation of exogenous antigens on MHC I molecules. *Curr. Opin. Immunol.* **64**, 1–8 (2020).
117. Prugnolle, F. et al. Pathogen-driven selection and worldwide HLA class I diversity. *Curr. Biol. CB*. **15**, 1022–1027 (2005).
118. Papadaki, G. F. et al. Decoupling peptide binding from T cell receptor recognition with engineered chimeric MHC-I molecules. *Front. Immunol.* **14**, 1116906 (2023).
119. Grant, E. J. et al. Broad CD8 + T cell cross-recognition of distinct influenza A strains in humans. *Nat. Commun.* **9**, 5427 (2018).
120. Kimanius, D. et al. Data-driven regularization lowers the size barrier of cryo-EM structure determination. *Nat. Methods*. **21**, 1216–1221 (2024).
121. Saotome, K. et al. Structural analysis of cancer-relevant TCR-CD3 and peptide-MHC complexes by cryoem. *Nat. Commun.* **14**, 2401 (2023).

122. Sun, Y. et al. CryoEM structure of an MHC-I/TAPBPR peptide-bound intermediate reveals the mechanism of antigen proofreading. *Proc. Natl. Acad. Sci. U. S. A.* **122**, e2416992122 (2025).
123. Domnick, A. et al. Molecular basis of MHC I quality control in the peptide loading complex. *Nat. Commun.* **13**, 4701 (2022).
124. Ayres, C. M., Corcelli, S. A. & Baker, B. M. Peptide and Peptide-Dependent motions in MHC proteins: immunological implications and biophysical underpinnings. *Front. Immunol.* **8**, 935 (2017).
125. Pöhlmann, T. et al. Differential peptide dynamics is linked to major histocompatibility complex polymorphism. *J. Biol. Chem.* **279**, 28197–28201 (2004).
126. Insaidoo, F. K., Zajicek, J. & Baker, B. M. A general and efficient approach for NMR studies of peptide dynamics in class I MHC peptide binding grooves. *Biochemistry* **48**, 9708–9710 (2009).
127. van Hateren, A. et al. Direct evidence for conformational dynamics in major histocompatibility complex class I molecules. *J. Biol. Chem.* **292**, 20255–20269 (2017).

Acknowledgements

We acknowledge D. Huard and R. Lieberman for lab space and assistance with attempts at X-ray crystallography. We thank R. Hughley and G. Newnam for assistance with instrument scheduling. A.C.M. acknowledges start-up funding from the Georgia Institute of Technology. A.C. and A.C.M acknowledge funding from NSF FAST REU Award #CHE-185251. K.C. and M.F. acknowledge funding from Georgia Institute of Technology President's Undergraduate Research Salary Award program. L.-Y.Y. acknowledges funds from the Taiwan Ministry of Education Government Scholarship to Study Abroad program.

Author contributions

A.C.M. designed the research and supervised the project. M.B.F., K.C., M.Z.K., L-Y.Y., and A.C.M. performed netMHCpan analysis. M.B.F., K.C., M.Z.K., A.C., J.M.R., and A.C.M. prepared recombinant protein samples. M.B.F., K.C., and M.Z.K. performed CD spectroscopy measurements. M.B.F. performed nanoDSF experiments. K.C., L-Y.Y., and A.C.M. performed in silico structure prediction and analysis. M.B.F., K.C., M.Z.K., and A.C.M. wrote the manuscript with feedback from all authors.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-92191-6.

Correspondence and requests for materials should be addressed to A.C.M.

Reprints and permissions information is available at www.nature.com/reprints.

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