

## RESEARCH ARTICLE

# Biophysical characterization of the *Plasmodium falciparum* circumsporozoite protein's N-terminal domain

Rob Geens<sup>1,2</sup> | Jessica Stanisich<sup>3</sup> | Olivier Beyens<sup>4</sup> | Stijn D'Hondt<sup>4</sup> |  
 Jean-Michel Thiberge<sup>5</sup> | Amber Ryckebosch<sup>1</sup> | Anke De Groot<sup>1</sup> |  
 Stefan Magez<sup>3,6</sup> | Didier Vertommen<sup>7</sup> | Rogério Amino<sup>5</sup> | Hans De Winter<sup>4</sup> |  
 Alexander N. Volkov<sup>2,8,9</sup> | Peter Tompa<sup>2,8,10</sup>  | Yann G.-J. Sterckx<sup>1</sup> 

<sup>1</sup>Laboratory of Medical Biochemistry (LMB), University of Antwerp, Antwerp, Belgium

<sup>2</sup>Structural Biology Brussels, Vrije Universiteit Brussel, Brussels, Belgium

<sup>3</sup>Cellular and Molecular Immunology, Vrije Universiteit Brussel, Brussels, Belgium

<sup>4</sup>Laboratory of Medicinal Chemistry (UAMC), University of Antwerp, Antwerp, Belgium

<sup>5</sup>Unit of Malaria Infection & Immunity, Institut Pasteur, Paris, France

<sup>6</sup>Ghent University Global Campus, Incheon, South Korea

<sup>7</sup>de Duve Institute and MASSPROT Platform, UCLouvain, Brussels, Belgium

<sup>8</sup>VIB-VUB Center for Structural Biology, Vlaams Instituut voor Biotechnologie (VIB), Brussels, Belgium

<sup>9</sup>Jean Jeener NMR Centre, Vrije Universiteit Brussel, Brussels, Belgium

<sup>10</sup>Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Budapest, Hungary

## Correspondence

Yann G.-J. Sterckx, Laboratory of Medical Biochemistry (LMB), University of Antwerp, Campus Drie Eiken, Universiteitsplein 1, 2610 Antwerp, Belgium.  
 Email: [yann.sterckx@uantwerpen.be](mailto:yann.sterckx@uantwerpen.be)

## Abstract

The circumsporozoite protein (CSP) is the main surface antigen of the *Plasmodium* sporozoite (SPZ) and forms the basis of the currently only licensed anti-malarial vaccine (RTS,S/AS01). CSP uniformly coats the SPZ and plays a pivotal role in its immunobiology, in both the insect and the vertebrate hosts. Although CSP's N-terminal domain (CSP<sub>N</sub>) has been reported to play an important role in multiple CSP functions, a thorough biophysical and structural characterization of CSP<sub>N</sub> is currently lacking. Here, we present an alternative method for the recombinant production and purification of CSP<sub>N</sub> from *Plasmodium falciparum* (PfCSP<sub>N</sub>), which provides pure, high-quality protein preparations with high yields. Through an interdisciplinary approach combining in-solution experimental methods and in silico analyses, we provide strong evidence that PfCSP<sub>N</sub> is an intrinsically disordered region displaying some degree of compaction.

## KEYWORDS

biophysics, circumsporozoite protein, intrinsic disorder, *Plasmodium falciparum*, recombinant protein production and purification, structural biology

Olivier Beyens and Stijn D'Hondt contributed equally to this work and should be considered joint authors.

Alexander N. Volkov, Peter Tompa, and Yann G.-J. Sterckx contributed equally to this work and should be considered joint senior authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

**Funding information**

Agence Nationale de la Recherche,  
Grant/Award Numbers: ANR-  
19-CE15-0027, ANR-10-LABX-62-IBEID;  
Fonds Wetenschappelijk Onderzoek,  
Grant/Award Number: 1S04221N;  
Universiteit Antwerpen

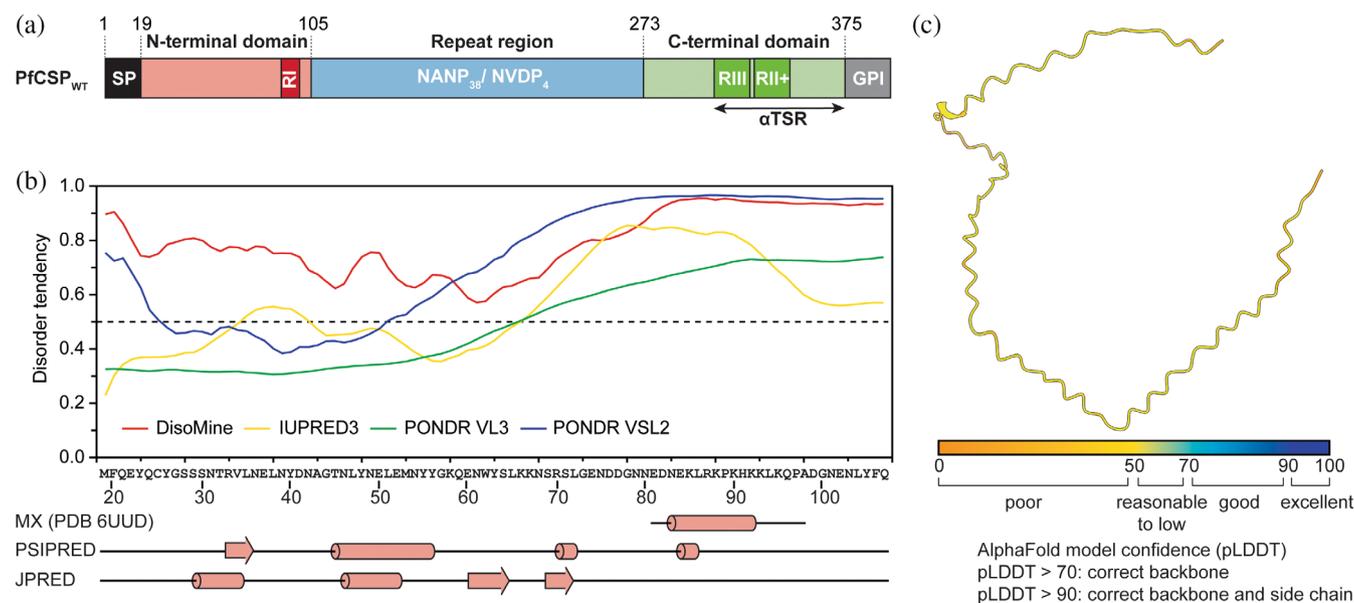
**Review Editor:** Jean Baum

**1 | INTRODUCTION**

With more than 200 million cases and more than half a million deaths annually, malaria remains a major global health threat (WHO, 2022). This devastating disease is caused by parasites of the *Plasmodium* genus, which are injected in their sporozoite (SPZ) form in the host skin by infected female *Anopheles* mosquitoes probing for blood. From their inoculation site, the SPZs travel to the liver using their specialized migratory machinery, where they differentiate into merozoites to initiate the blood stage of the infection that causes the notorious malaria pathology (Cowman et al., 2016).

Upon arrival in the liver sinusoids, SPZs interact with highly-sulfated heparan sulfate proteoglycans (HS-HSPGs) presented by hepatic cells (Coppi et al., 2007, 2011; Pinzon-Ortiz et al., 2001). This specific interaction is governed by the circumsporozoite protein (CSP), the SPZ's

major surface antigen, which uniformly coats the parasite surface. Despite differences in both length and composition of CSP amino acid sequences from various *Plasmodium* species, all have a similar modular build-up (Figure 1a) consisting of (i) an N-terminal domain (CSP<sub>N</sub>) containing the conserved pentapeptide Region I, (ii) a highly flexible linker consisting of tandem repeats, (iii) a C-terminal domain (CSP<sub>C</sub>) containing the conserved Region III and Region II+, and (iv) a C-terminal glycosylphosphatidylinositol (GPI) anchor providing a physical link to the parasite's plasma membrane (McCutchan et al., 1996). CSP plays a central role in SPZ immunobiology, in both the insect and vertebrate hosts (Aly et al., 2009). Within the latter, CSP has been described to (i) be essential for the recognition and invasion of host hepatocytes through interaction with liver-specific HS-HSPGs (Coppi et al., 2007; Pinzon-Ortiz et al., 2001), (ii) constitute the immunodominant SPZ surface antigen



**FIGURE 1** Sequence-based structural analysis of PfcSP<sub>N</sub>. (a) Schematic overview of wild-type PfcSP from *Plasmodium falciparum* NF54. (b) Disorder tendency of PfcSP<sub>N</sub> predicted by several algorithms is shown at the top. Residues with disorder tendencies above the threshold of 0.5 (dotted line) are considered disordered. Secondary structure predictions are plotted below, with red cylinders and arrows representing  $\alpha$ -helices and  $\beta$ -strands, respectively. (c) AlphaFold2 model of PfcSP<sub>N</sub>. The structure is colored according to the pLDDT score, which reflects (local) model quality.

(Kumar et al., 2006), (iii) possess both extra- and intracellular effector functions to dampen host inflammatory responses (Singh et al., 2007; Usynin et al., 2007), and (iv) form a shield around the SPZ to protect its plasma membrane against pore-forming proteins during parasite migration in the skin (Aliprandini et al., 2018).

Over the past decades, it has been documented that CSP<sub>N</sub> is involved in several CSP-mediated events: it is required for correct SPZ localization to both the mosquito salivary glands (Myung et al., 2004) and the vertebrate hepatocytes (Coppi et al., 2011). The molecular mechanism underlying CSP-mediated hepatocyte recognition long remained under debate as two conserved regions of CSP were proposed to contain the HS-HSPG binding site: Region I within CSP<sub>N</sub> (Ancsin & Kisilevsky, 2004; Rathore et al., 2002) and Region II+ within CSP<sub>C</sub> (Gantt et al., 1997). However, more recent studies have demonstrated (i) the inability of CSP<sub>C</sub> to bind heparin (Doud et al., 2012) and (ii) the importance of a stretch of positively charged residues preceding Region I (Zhao et al., 2016). These findings support the model proposed by Coppi and colleagues, in which CSP<sub>N</sub> is the main contributor to the HS-HSPG interaction, while CSP<sub>C</sub> is involved in further cell adhesion and invasion processes, which still remain to be fully elucidated (Coppi et al., 2011). CSP<sub>N</sub> is proteolytically cleaved within Region I upon HS-HSPG binding, an event that is essential for productive SPZ invasion of hepatic cells (Coppi et al., 2011), making this region an interesting target for cross-species anti-malarial strategies. Indeed, studies with anti-CSP<sub>N</sub> antibodies have shown that many target epitopes adjacent to or within Region I (Espinosa et al., 2015; Tan et al., 2021; Thai et al., 2020). However, divergent results have been obtained regarding the ability of these antibodies to bind native CSP on the SPZ surface and their protective efficacy *in vitro* and *in vivo* (Bongfen et al., 2009; Espinosa et al., 2015; Tan et al., 2021; Thai et al., 2020).

In contrast to the known  $\alpha$ -thrombospondin type-1 repeat ( $\alpha$ TSR) structure of CSP<sub>C</sub> (Doud et al., 2012), the structural features of CSP<sub>N</sub> are less well understood. While some studies make reference to this domain's highly flexible/disordered nature (Kucharska, Binter, et al., 2022; Oyen et al., 2018; Thai et al., 2020; Zhao et al., 2016), a thorough biophysical and structural characterization of CSP<sub>N</sub> is currently lacking. Here, we describe an alternative method for the recombinant production and purification of CSP<sub>N</sub> from *Plasmodium falciparum* (PfCSP<sub>N</sub>). In our strategy, we have replaced the repeat region between PfCSP<sub>N</sub> and PfCSP<sub>C</sub> by a truncated linker containing a protease cleavage site. The functionality and biological relevance of the final PfCSP<sub>N</sub> product are demonstrated by a multidisciplinary approach

encompassing heparin affinity chromatography, isothermal titration calorimetry, and HepG2 binding assays. Our method provides a final yield of ~15 mg per liter of bacterial culture, thereby providing sufficient material for biophysical studies to probe the domain's structure–function relationship. Furthermore, the study encompasses *in silico* disorder and secondary structure predictions, *in-solution* studies (AGF, DLS, CD spectroscopy, SAXS, and NMR spectroscopy) and MD simulations. When taken together, these data provide strong evidence that PfCSP<sub>N</sub> is intrinsically disordered.

## 2 | RESULTS

### 2.1 | *In silico* structure predictions suggest that PfCSP<sub>N</sub> is intrinsically disordered

The highly flexible/disordered character of PfCSP<sub>N</sub> has been proposed in several publications (Kucharska, Binter, et al., 2022; Oyen et al., 2018; Thai et al., 2020; Zhao et al., 2016). Indeed, several structure prediction tools support this statement (Figure 1b,c). Various intrinsic disorder predictors report a high overall tendency for disorder and an AlphaFold2-based structure prediction is characterized by an obvious lack of tertiary structure. In addition, the PfCSP<sub>N</sub> AlphaFold2 model has predicted local distance difference test (pLDDT) values around 50, which is indicative of intrinsic disorder (Tunyasyunakool et al., 2021). Interestingly, secondary structure predictors seem to suggest that PfCSP<sub>N</sub> contains secondary structure elements (mainly  $\alpha$ -helices). Furthermore, a crystal structure of a peptide corresponding to PfCSP<sub>N</sub> residues Glu81-Ala98 in complex with a monoclonal antibody shows that the stretch Asn83-Lys92 adopts an  $\alpha$ -helical conformation (Thai et al., 2020). Taken together, the *in-silico* analyses suggest that PfCSP<sub>N</sub> is an intrinsically disordered region (IDR) harboring secondary structure elements, although this has not been experimentally confirmed thus far.

### 2.2 | A PfCSP construct encoding a shortened, protease-cleavable linker enables high yield recombinant production of PfCSP<sub>N</sub> in *E. coli*

Despite several trials (including variations in culture media, incubation temperatures before and after induction, inducer concentrations, *E. coli* strains, etc.), the recombinant production of PfCSP<sub>N</sub> as a stand-alone protein in *E. coli* did not work in our hands (overexpression

from a PfCSP<sub>N</sub>-encoding construct could not be observed). Given that full-length and truncated linker PfCSP variants are easily produced, we turned to a construct design in which the PfCSP NANP linker region is replaced by a shortened linker (SL) encompassing the TEV protease cleavage site and two NANP repeats (PfCSP<sub>SL/TEV</sub>; Figure 2a). The overall strategy consists in recombinantly producing and purifying C-terminally His-tagged PfCSP<sub>SL/TEV</sub> (Figure 2b), followed by TEV protease treatment and subsequent purification of the PfCSP<sub>N</sub> cleavage product (Figure 2c).

The recombinant production of PfCSP<sub>SL/TEV</sub> is performed in *E. coli* SHuffle Express T7, given the presence of multiple disulfides in CSP's  $\alpha$ TSR domain (Doud et al., 2012). Induction of expression results in clear production of the target protein (MM ~ 21 kDa) as demonstrated by SDS-PAGE (Figure 2b, top gel). SDS-PAGE analysis of cell lysis tests shows that PfCSP<sub>SL/TEV</sub> accumulates in inclusion bodies (Figure 2b, bottom gel, lanes 1 to 3). For this reason, cell lysis is performed in a Triton X-100-containing lysis buffer to wash away contaminants attached to the inclusion bodies through the hydrophobic effect. The inclusion bodies are subsequently solubilized in a buffer containing 8.0 M urea prior to IMAC purification, which results in a single elution peak containing relatively pure PfCSP<sub>SL/TEV</sub> when increasing the imidazole concentration to 0.5 M (Figure 2b, top chromatogram and bottom gel, lanes 4–8). The IMAC elution fractions are then subjected to a desalting step, which is included to reduce the urea concentration to 3.0 M prior to TEV protease treatment. In addition to the buffer change, desalting further polishes the sample, thereby resulting in a highly pure PfCSP<sub>SL/TEV</sub> preparation (Figure 2b, bottom chromatogram and gel, lane 9).

PfCSP<sub>N</sub> is generated by proteolytically processing PfCSP<sub>SL/TEV</sub>, which is performed under mildly denaturing conditions given that our in-house TEV protease retains its activity in 3.0 M urea. Overnight incubation at room temperature results in a close to complete cleavage of PfCSP<sub>SL/TEV</sub> into two products: PfCSP<sub>N</sub> and PfCSP<sub>LC</sub> (Figure 2c, gel, lanes 1 to 2). The solubility and fully reduced state of all proteins present in the sample are ensured by increasing the urea and reducing agent concentrations to 8.0 M and 4.0 mM, respectively. PfCSP<sub>N</sub> is obtained through a reverse IMAC purification step (Figure 2c, top chromatogram). As PfCSP<sub>N</sub> is the only protein in the sample lacking a His-tag (both the TEV protease and PfCSP<sub>LC</sub> are His-tagged), it is readily isolated in the flow-through. SDS-PAGE analysis demonstrates the high purity of the obtained PfCSP<sub>N</sub> fractions (Figure 2c, gel, lane 4). Finally, PfCSP<sub>N</sub> is brought back to native conditions by fast desalting to non-denaturing buffer, followed by incubation at 4°C for 2–3 days

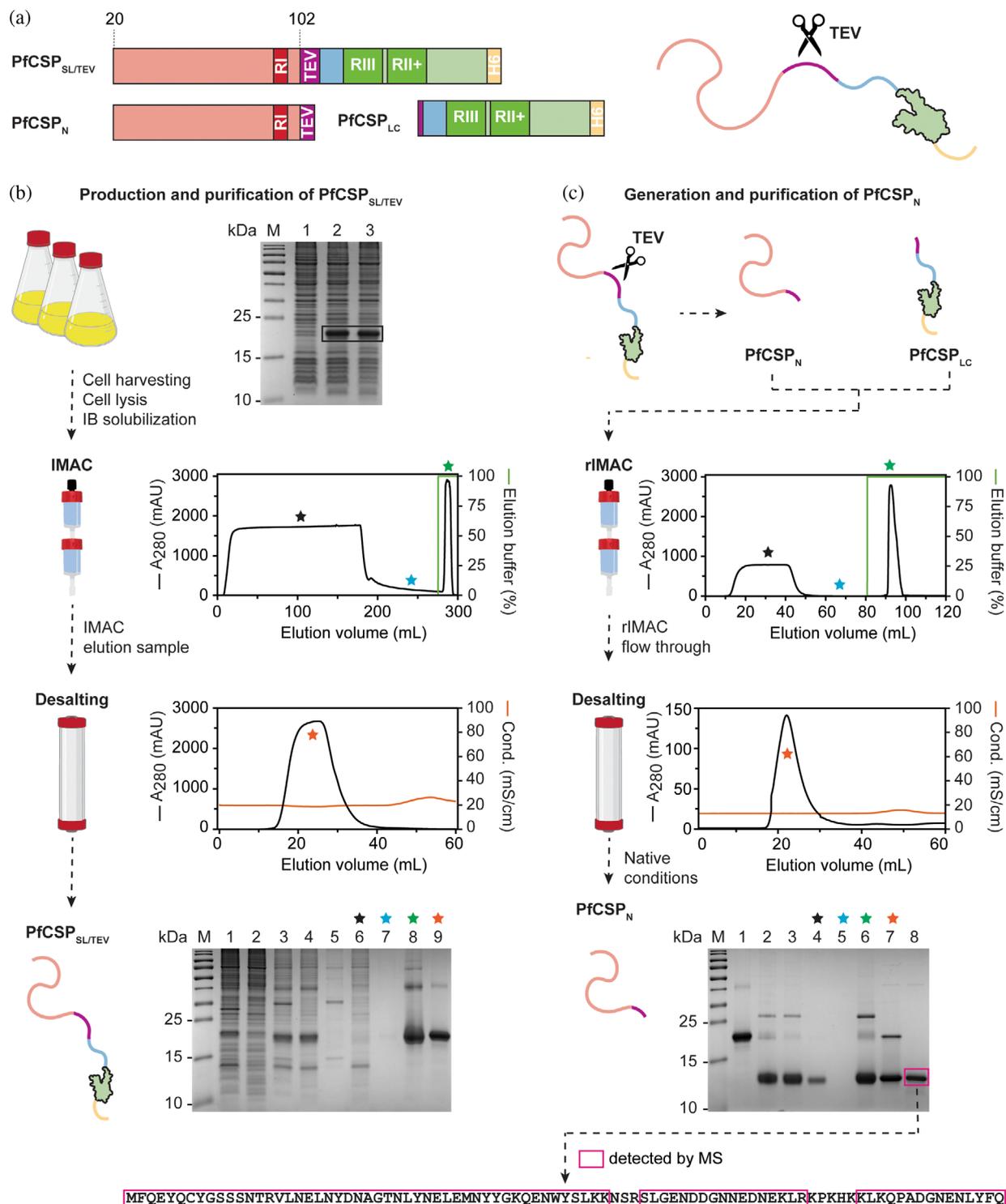
(Figure 2c, bottom chromatogram and gel, lanes 7 to 8). Optimal results are obtained in Tris-HCl buffer at pH 8.0 with 100 mM NaCl at low protein concentration (0.2 mg/mL). Aggregates, if any, are removed by filtration. Mass spectrometry on the excised band confirmed that the final purification product is intact PfCSP<sub>N</sub> (sequence coverage of 91.01%; Figure 2c, bottom panel). This approach typically yields ~15 mg highly pure PfCSP<sub>N</sub> per liter of bacterial culture.

### 2.3 | Recombinant PfCSP<sub>N</sub> is functional and biologically relevant

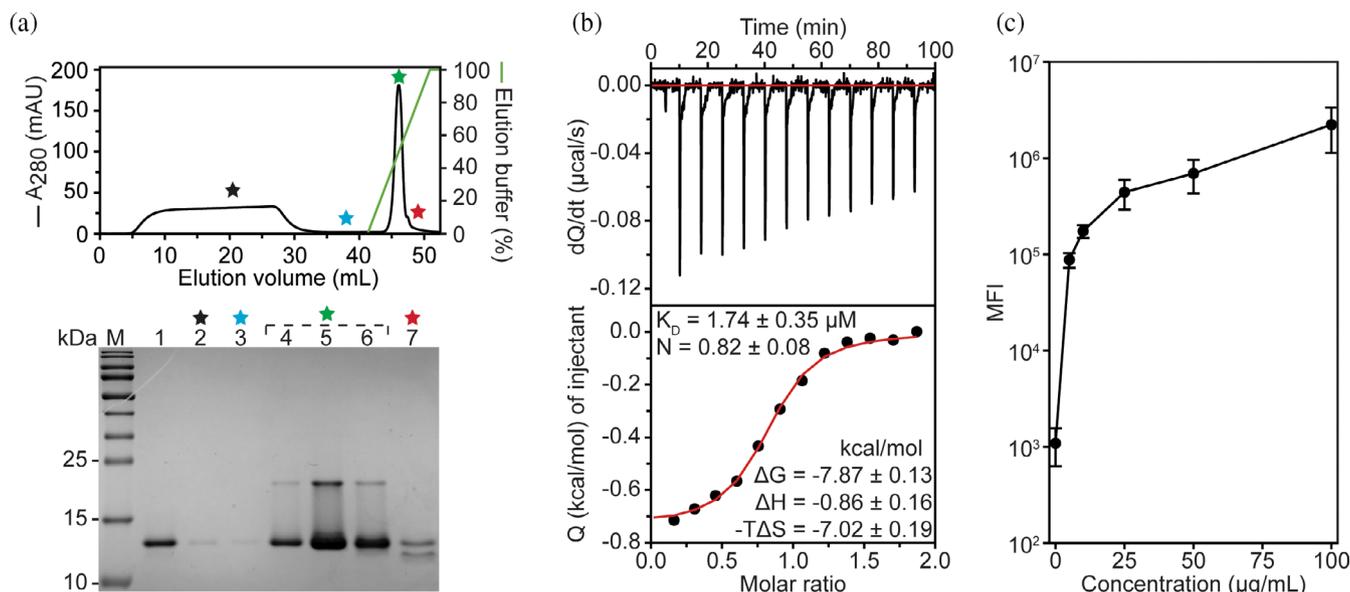
The functionality, and hence biological relevance, of the protein was validated by a combination of heparin affinity chromatography (HAC), isothermal titration calorimetry (ITC), and HepG2 binding assays (Figure 3).

HAC is frequently used in CSP research given that heparin is a structural analogue of CSP's natural ligand (HS-HSPG) (Doud et al., 2012; Rathore & McCutchan, 2000a, 2000b; Rathore, McCutchan, et al., 2001; Zhao et al., 2016). SDS-PAGE analysis of the HAC elution profile reveals that PfCSP<sub>N</sub> elutes at relatively high NaCl concentrations (~0.4 M to 0.5 M NaCl; Figure 3a), which is indicative of a strong interaction with heparin. The main elution fractions (lanes 4 to 7) also contain a far less abundant higher molecular mass band, which we interpret as cross-linked PfCSP<sub>N</sub> (artifact formed during SDS sample preparation). Finally, the fraction corresponding to the small shoulder at the trailing edge of the main peak displays an additional low-molecular band on SDS-PAGE, which is probably a PfCSP<sub>N</sub> degradation product.

Given that many proteins have an intrinsic ability to bind heparin (Peysseon & Ricard-Blum, 2014), the ability of PfCSP<sub>N</sub> to interact with HS-HSPGs specifically was further investigated. First, ITC was employed to characterize the interaction between PfCSP<sub>N</sub> and a sulfated heparin dodecasaccharide (HO12), which corresponds to the minimal HS-HSPG moiety binding to PfCSP (Rathore, McCutchan, et al., 2001). Analysis of the thermogram reveals that PfCSP<sub>N</sub> binds to HO12 with micromolar affinity, and that the interaction is mainly entropy-driven (Figure 3b). The stoichiometry of the interaction is close to 1, confirming that it reports on a single binding event. Finally, fluorescently labeled PfCSP<sub>N</sub> was probed for its ability to bind HepG2 cells (a human hepatoma cell line presenting native HS-HSPGs; Figure 3c). PfCSP<sub>N</sub> binds to HepG2 cells in a specific, dose-dependent manner that is similar to observations made by another group for full-length PfCSP and truncates thereof (Rathore et al., 2002). By validating the identity and functionality, we have



**FIGURE 2** Strategy for the recombinant production and purification of PfCSP<sub>N</sub> based on a PfCSP construct encoding a shortened, protease-cleavable linker. (a) Schematic overview PfCSP constructs. (b) Recombinant production of PfCSP<sub>SL/TEV</sub> (panel 1) and its purification (panels 2 to 4) by IMAC and desalting. Normalized culture samples were analyzed by SDS-PAGE: before induction (lane 1), 3 h post induction (lane 2), and 20 h post induction (lane 3). Purification was visualized by SDS-PAGE: lysed cells (lane 1), soluble fraction (lane 2), insoluble fraction (lane 3), solubilized inclusion bodies (lane 4), remaining insoluble material (lane 5), IMAC flow-through, wash and elution fractions (lanes 6 to 8), and desalted PfCSP<sub>SL/TEV</sub> (lane 9). (c) Generation of PfCSP<sub>N</sub> by proteolytic cleavage of PfCSP<sub>SL/TEV</sub> (panel 1) followed by the isolation and purification of PfCSP<sub>N</sub> (panels 2 to 4) by reverse IMAC and desalting. Purification was visualized by SDS-PAGE: PfCSP<sub>SL/TEV</sub> before addition of TEV protease (lane 1), TEV cleavage sample after overnight incubation (lane 2), IMAC input, flow-through, wash and elution samples (lanes 3 to 6), desalted PfCSP<sub>N</sub> (lane 7), and native PfCSP<sub>N</sub> (lane 8). The final panel shows the result of a LC-MS/MS analysis on the excised gel piece from panel 4 (purple box), validating the identity of PfCSP<sub>N</sub>.



**FIGURE 3** Validation of the functionality of recombinant PfcCSP<sub>N</sub>. (a) HPLC. PfcCSP<sub>N</sub> was loaded on the heparin column and eluted with linear salt gradient. An SDS-PAGE analysis is shown below the chromatogram: PfcCSP<sub>N</sub> loaded onto the column (lane 1), flow through and wash fractions (lanes 2 and 3), elution fractions (lanes 4 to 7). (b) ITC binding studies with PfcCSP<sub>N</sub> and HO12. The top panels represent the thermograms with the baseline (red line) and the raw heat output data (black line) as a function of time; the bottom panels depict the isotherms corresponding to the integrated heat peaks (black dots) that are fitted by a one-site binding model (red line). Extracted binding parameters are indicated in the inset. (c) HepG2 binding assay. Binding is indicated by median fluorescence intensity (MFI) signals. The mean and standard deviation of triplicates for each PfcCSP<sub>N</sub>-AF647 concentration are shown and data were corrected for nonspecific binding by BSA-AF647.

demonstrated that our recombinant strategy yields highly pure and biologically relevant PfcCSP<sub>N</sub>.

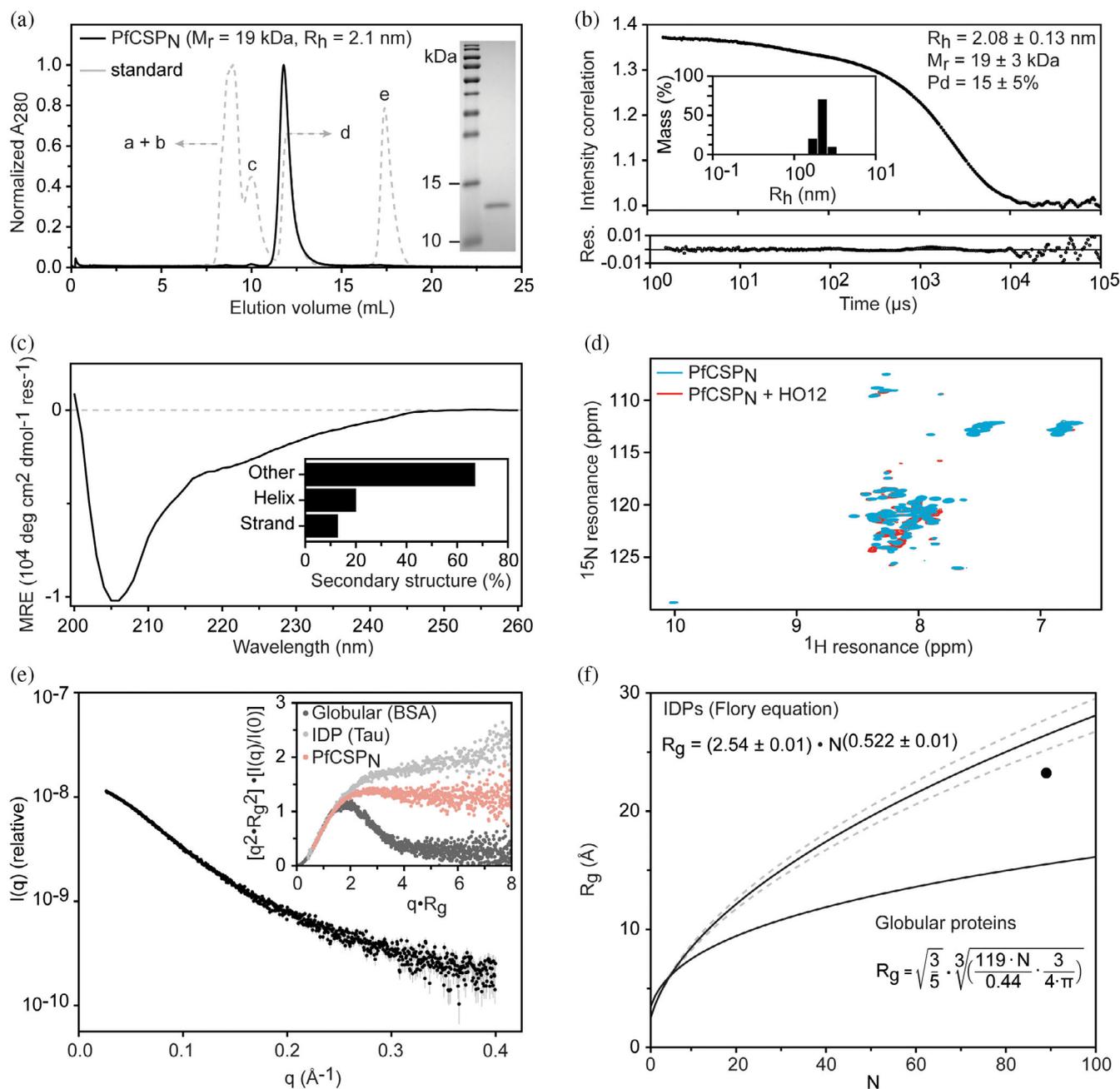
## 2.4 | In-solution biophysical studies confirm that PfcCSP<sub>N</sub> is intrinsically disordered

Interestingly, throughout the purification and validation, we noticed that PfcCSP<sub>N</sub> consistently displays an aberrant electrophoretic mobility in SDS-PAGE. PfcCSP<sub>N</sub> migrates with a higher apparent molecular mass compared to what is expected based on its sequence (10.5 kDa), reminiscent of typical behavior observed for intrinsically disordered proteins (IDPs) (Tompa, 2002). A combination of analytical gel filtration (AGF), dynamic light scattering (DLS), circular dichroism (CD) spectroscopy, nuclear magnetic resonance (NMR spectroscopy) and small-angle X-ray scattering (SAXS) was used to further experimentally confirm the IDP nature of PfcCSP<sub>N</sub>.

AGF and DLS were employed to investigate the hydrodynamic behavior of PfcCSP<sub>N</sub>. Both techniques demonstrate that the apparent molecular mass ( $MM_{app}$ ) and hydrodynamic radius ( $R_{h,app}$ ) are higher than the theoretical values for globular proteins of the same size (1.8 and 1.2 times, respectively; Figure 4a,b). This discrepancy is

characteristic of an IDP and suggests that PfcCSP<sub>N</sub> adopts a nonglobular conformation (Cszimók et al., 2006). To further examine the structural features of PfcCSP<sub>N</sub>, a far-UV CD spectrum was collected (Figure 4c). The spectrum is typical for a protein with a significant amount of intrinsic disorder with a dominant negative peak at 205 nm (Tompa, 2002). However, completely disordered proteins have a mean residue ellipticity (MRE) value close to zero around 220 nm. This is clearly not the case for PfcCSP<sub>N</sub>, suggesting that the protein contains some secondary structure elements despite being mainly disordered. Indeed, secondary structure composition estimated by the BeStSel server (Micsonai et al., 2015) shows that PfcCSP<sub>N</sub> is composed of 20.0%  $\alpha$ -helices, 12.9%  $\beta$ -strands, and 67.1% other structures or disorder.

The high degree of structural disorder is further illustrated by solution NMR spectroscopy. In particular, the <sup>1</sup>H-<sup>15</sup>N HSQC spectrum of <sup>15</sup>N labeled PfcCSP<sub>N</sub> exhibits poor signal dispersion, with most of the peaks confined to a narrow spectral region (Figure 4d, blue). Such a cluttered <sup>1</sup>H-<sup>15</sup>N HSQC spectrum indicates that PfcCSP<sub>N</sub> backbone amides sample a range of random-coil chemical shift values and is a salient feature of many IDPs (Burke et al., 2015; Conicella et al., 2016; Ryan et al., 2018). To further probe its interaction with heparin, we acquired the <sup>1</sup>H-<sup>15</sup>N HSQC spectrum of <sup>15</sup>N labeled PfcCSP<sub>N</sub> in



**FIGURE 4** Biophysical characterization of PfCSP<sub>N</sub> through various in-solution techniques. (a) AGF. The elution profile of the standard is indicated by the gray dotted line, with elution peaks indicated by small letters: (a) bovine thyroglobulin (MM = 670 kDa,  $R_h = 8.6$  nm), (b) bovine  $\gamma$ -globulin (MM = 158 kDa,  $R_h = 5.1$  nm), (c) chicken ovalbumin (MM = 44 kDa,  $R_h = 2.8$  nm), (d) horse myoglobin (MM = 17 kDa,  $R_h = 1.9$  nm), (e) vitamin B12 (MM = 1.35 kDa,  $R_h = 0.85$  nm). The SDS-PAGE analysis of the PfCSP<sub>N</sub> elution peak is shown in the inset. (b) DLS. The black dots and gray line represent the experimental data and fit, respectively, with the residuals shown at the bottom. The percentage of mass versus the  $R_h$  distribution is plotted in the inset. (c) CD spectrum of PfCSP<sub>N</sub> in mean residue ellipticity (MRE). The estimation of secondary structure composition by BestSel is plotted in the inset. (d) NMR spectroscopy.  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of PfCSP<sub>N</sub> (blue) and PfCSP<sub>N</sub> in complex with HO12 (red). (e) SAXS curve of PfCSP<sub>N</sub>. The inset shows the normalized Kratky plots of PfCSP<sub>N</sub>, a globular protein (BSA), and a random coil IDP (Tau protein). (f) Comparison of the experimental  $R_g$  of PfCSP<sub>N</sub> determined by SAXS (black dot) with the theoretical  $R_g$  values of globular proteins and IDPs (the gray dotted lines represent the confidence interval of the Flory equation) with the same number of amino acids.

complex with HO12 (Figure 4d, red). Upon addition of HO12, several PfCSP<sub>N</sub> backbone amide peaks experience binding shifts, confirming the interaction. At the same

time, the spectral changes are modest and limited to a handful of resonances, ruling out major binding-induced structural changes (e.g., folding upon binding) and

demonstrating that PfCSP<sub>N</sub> retains its IDP character in the heparin-bound form.

The IDP nature of PfCSP<sub>N</sub> was further confirmed by investigating its in-solution structural parameters using SAXS (Figure 4e,f). Macromolecular flexibility in solution is readily visualized by the normalized Kratky plot (Figure 4e, inset). While a globular protein displays a bell-shaped curve with a maximum at (1.732, 1.104), the curve of a completely unfolded protein behaving like a random chain increases steeply until it reaches a plateau at  $\sim 2$  (Durand et al., 2010). The normalized Kratky plot of PfCSP<sub>N</sub> undoubtedly illustrates that it behaves like a highly flexible protein. However, PfCSP<sub>N</sub> does not reach a plateau at 2 but instead stabilizes at  $\sim 1.5$ , suggesting that its conformational ensemble is characterized by high flexibility, yet displays some degree of compaction. This behavior is further confirmed through a comparison of the protein's experimentally determined radius of gyration ( $R_g$ ) with the theoretical values expected for globular proteins and IDPs of similar length (89 aa). PfCSP<sub>N</sub> clearly deviates strongly from typical globular protein behavior and displays many characteristics of an IDP-like protein (Figure 4f). However, the protein does not behave as a complete random coil IDP as its  $R_g$  is not located within the Flory region. Instead, the  $R_g$  lies beneath the Flory region, which is again indicative of a conformational ensemble characterized by a degree of compaction (Bernadó & Svergun, 2012). This conformational trait is supported by calculating the  $R_g/R_h$  ratio: while globular proteins and IDPs, are characterized by ratios of 0.77 and 1.3, respectively, PfCSP<sub>N</sub> has a ratio of 1.1.

When combined, the in-solution methods provide strong experimental evidence for PfCSP<sub>N</sub>'s highly flexible nature.

## 2.5 | MD simulations support the presence of $\alpha$ -helices in the PfCSP<sub>N</sub> conformational ensemble

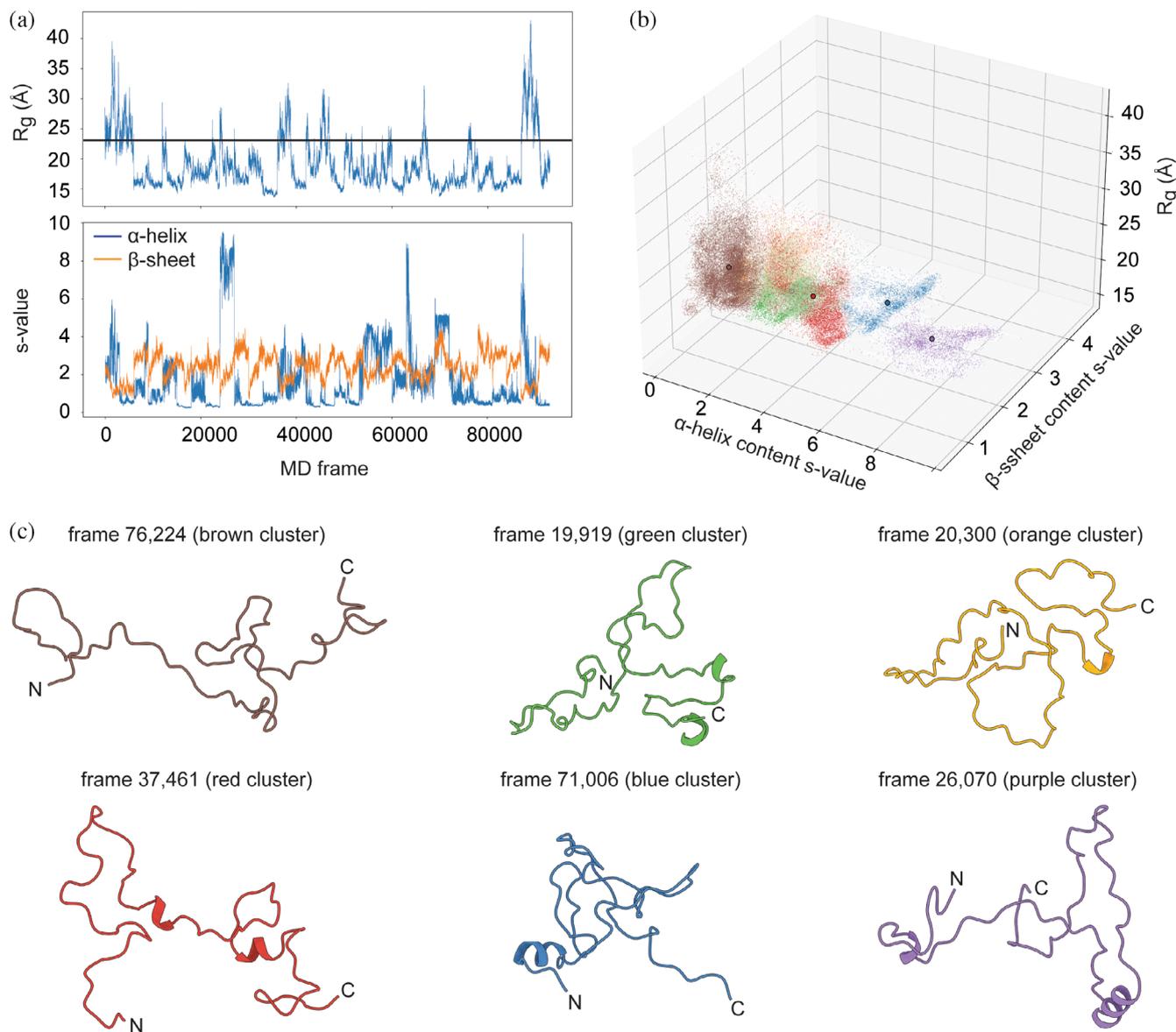
Finally, the conformational ensemble of PfCSP<sub>N</sub> was studied through all-atom molecular dynamics (MDs) simulations. A large trajectory (93,000 frames) corresponding to 1.86  $\mu$ s of simulation data was generated and, for each frame, the  $R_g$  and secondary structure content were analyzed (Figure 5). The  $R_g$  value analysis indicates that a wide variety of conformers is sampled (as would be expected for an IDP), with more compact conformations being more frequently adopted (Figure 5a, top panel). This corroborates well with the experimental SAXS data, thereby demonstrating that the simulation ensemble is representative of PfCSP<sub>N</sub>'s overall in-solution behavior. Secondary structure content observed during the

simulation runs was assessed through  $s$ -value analysis (*cfr.* Section 4 for an in-depth explanation). The analysis reveals that  $\alpha$ -helices are observed in several occurrences with  $s$ -values of 6 and higher (occurrences of at least 7%), whereas  $s$ -values for  $\beta$ -sheets remain below 4 over the entire trajectory (occurrences of 0.06% or less; Figure 5a, bottom panel). It thus appears that  $\alpha$ -helical conformations are more frequently sampled compared to  $\beta$ -sheets, which is fully consistent with the above-mentioned CD data. 3D-clustering of the frames based on  $\alpha$ -helical content  $s$ -value,  $\beta$ -sheet content  $s$ -value, and calculated  $R_g$  distinguishes six clusters from each other (Figure 5b). A representative frame from each cluster is shown in Figure 4c. These structural models support the current view of the protein's conformational ensemble based on the experimental in-solution studies; that is, PfCSP<sub>N</sub> is an IDR containing secondary structure elements (predominantly  $\alpha$ -helices).

## 3 | DISCUSSION

The recombinant production of a variety of *Plasmodium* CSP constructs representing (near) full-length versions (CSP<sub>FL</sub>) or truncated variants corresponding to specific domains or regions have been described in literature, using bacterial (Aliprandini et al., 2018; Beutler et al., 2022; Oyen et al., 2017; Plassmeyer et al., 2009; Singh et al., 2020; Zhao et al., 2016), yeast (Doud et al., 2012; Marques et al., 2020; Plassmeyer et al., 2009; Zhao et al., 2016), and mammalian cell expression systems (Doud et al., 2012; Keitany et al., 2014; Kisalu et al., 2018; Kucharska, Binter, et al., 2022; Kucharska, Hossain, et al., 2022; Kucharska et al., 2020; Scally et al., 2018; Thai et al., 2020; Triller et al., 2017; Vijayan et al., 2021; Wang et al., 2020). Despite this extensive repertoire, detailed reports on the recombinant production and purification of CSP<sub>N</sub> are scarce (Rathore, Kumar, et al., 2001). Indeed, most studies on this domain are peptide-based (Aley et al., 1986; Ancsin & Kisilevsky, 2004; Bermúdez et al., 2008; Myung et al., 2004; Sidjanski et al., 1997), indicating that CSP<sub>N</sub> is difficult to produce recombinantly as a stand-alone protein (as was the case in our hands).

Here, we present an alternative strategy to obtain PfCSP<sub>N</sub> based on the recombinant overexpression of a construct encoding PfCSP's N- and C-terminal domains linked by a TEV protease cleavage site and two tandem repeats of the original linker (PfCSP<sub>SL/TEV</sub>). PfCSP<sub>SL/TEV</sub> is easily produced in *E. coli* SHuffle by standard IPTG induction protocols and relatively high yields are obtained after its purification from inclusion bodies. PfCSP<sub>N</sub> is subsequently obtained by TEV cleavage of



**FIGURE 5** MD simulations on PfCSP<sub>N</sub>. (a) The  $R_g$  (top panel) and  $\alpha$ -helix and  $\beta$ -sheet content (bottom panel) of PfCSP<sub>N</sub>, as calculated per frame of the concatenated and cleaned trajectory of all 31 MD simulations. The black line in the  $R_g$  plot indicates the experimental value determined by SAXS. (b) A 3D-clustering plot of the  $\alpha$ -helix and  $\beta$ -sheet content, as well as the  $R_g$  of PfCSP<sub>N</sub>, as calculated for each frame of the concatenated and cleaned trajectory of all 31 MD simulations. The frames were clustered into six clusters using the KMeans algorithm. (c) Cartoon representations of representative structures from each cluster.

PfCSP<sub>SL/TEV</sub> followed by reverse IMAC given that it lacks a His-tag. A final desalting serves as a polishing step to deliver highly pure PfCSP<sub>N</sub> samples. Curiously, while PfCSP<sub>FL</sub> versions produced in *E. coli* have been reported to be soluble (Oyen et al., 2017; Patra et al., 2017; Schwenk et al., 2014), PfCSP<sub>SL/TEV</sub> accumulates in inclusion bodies. PfCSP<sub>SL/TEV</sub> and the PfCSP<sub>FL</sub> versions all contain PfCSP's N- and C-terminal domains connected by the NANP/NVDP tandem repeat linker. However, the linker length is significantly different in both construct types, with PfCSP<sub>SL/TEV</sub> and PfCSP<sub>FL</sub> containing two and  $\geq 22$  tetrapeptide repeats, respectively. Interestingly, NANP repeats have previously been described as a

solubilizing fusion partner for the production of recombinant human amyloid- $\beta$ , a protein with a central role in Alzheimer's disease (Finder et al., 2010). These results suggest that the PfCSP repeat region may have a solubilizing function, explaining why severe linker truncation would lead to the accumulation of PfCSP<sub>SL/TEV</sub> in inclusion bodies. Nonetheless, the presented strategy provides high yields of pure, high-quality PfCSP<sub>N</sub> preparations suitable for functional and structural studies.

By combining experimental in-solution methods with in silico analyses, the findings in this paper also provide strong evidence that PfCSP<sub>N</sub> harbors the archetypical features of IDRs. Several in silico predictors indicate that

PfCSP<sub>N</sub> has a high tendency for disorder with total disorder content ranging from 47% to 100%, which is in agreement with previously reported predictions (Guy et al., 2015). In addition, an AlphaFold2 prediction model for PfCSP<sub>N</sub> clearly lacks tertiary structure and has pLDDT values close to 50, which is indicative of intrinsic disorder (Tunyasuvunakool et al., 2021). The in-silico studies were experimentally confirmed by thoroughly characterizing PfCSP<sub>N</sub>'s biophysical and structural properties through various in-solution techniques (AGF, DLS, CD spectroscopy, SAXS, and NMR spectroscopy). When combined, the results demonstrate that the structural features of this CSP domain are best described by a highly flexible/dynamic conformational ensemble displaying some degree of compaction and containing secondary structure elements (mainly  $\alpha$ -helices). This view is further supported by MD simulations performed over a relatively long-time frame (1.86  $\mu$ s), demonstrating that several PfCSP<sub>N</sub> regions have helical propensities. Interestingly, one of these segments (Asp82-Arg87) lays within the Glu81-Ala98 region which was recently shown to adopt an  $\alpha$ -helical conformation in complex with a monoclonal antibody (Thai et al., 2020). The crystal structure shows that the stretch Asn83-Lys92 adopts an  $\alpha$ -helical conformation, although we cannot exclude that this geometry may be induced by the interaction with the antibody. Furthermore, the intrinsically disordered nature of PfCSP<sub>N</sub> explains why electron density for this domain is frequently lacking in X-ray crystallographic and cryo-EM studies (Oyen et al., 2018; Thai et al., 2020). It also explains why its structural and functional integrity are not impacted by a purification protocol that heavily relies on the use of denaturing agents.

The biological relevance of PfCSP<sub>N</sub> was demonstrated through a combination of biophysical and functional assays. The HAC and HepG2 binding assay data are in complete agreement with previously published reports: while HAC shows that PfCSP<sub>N</sub> strongly interacts with heparin (Ancsin & Kisilevsky, 2004; Rathore et al., 2002; Zhao et al., 2016), the HepG2 assays evidence that PfCSP<sub>N</sub> binds to exposed HS-HSPGs in a dose-dependent manner in the expected concentration range (Rathore et al., 2002). Furthermore, ITC experiments show that PfCSP<sub>N</sub> binds to HO12 (a sulfated heparin dodecasaccharide corresponding to a minimal PfCSP-binding HS-HSPG moiety [Rathore, McCutchan, et al., 2001]) with micromolar affinity and a 1:1 stoichiometry. While micromolar binding is often observed for protein—HS-HSPG interactions (Peysseon & Ricard-Blum, 2014), this is weaker compared to the sub-micromolar affinity constant ( $\sim$ 200 nM) reported in a previous study (Rathore, McCutchan, et al., 2001). This discrepancy may be

explained by the difference in chain lengths of the used heparin oligosaccharide and the absence of CSP<sub>C</sub> in our study (it has been proposed that CSP<sub>C</sub> may contribute to the interaction [Zhao et al., 2016]). Interestingly, PfCSP<sub>N</sub>-HO12 interaction is apparently not accompanied by binding-induced structural changes: that is, the folding upon binding often described for IDP-ligand interactions (Malagrino et al., 2022) could not be observed. Instead, the NMR data suggest that PfCSP<sub>N</sub> retains its IDP character in the HO12-bound form, which may be indicative of a so-called “fuzzy type interaction” (Tompa & Fuxreiter, 2008). It is tempting to speculate why PfCSP<sub>N</sub> harbors IDP features in both free and HO12-bound forms. One reason might be that Region I must remain readily accessible for proteolytic cleavage during hepatocyte invasion (Coppi et al., 2005, 2011), which would indeed be facilitated by PfCSP<sub>N</sub>'s highly flexible nature. Further insights will most likely be provided by a detailed structural description of the conformational ensemble of full-length PfCSP, which we are currently actively pursuing.

## 4 | MATERIALS AND METHODS

### 4.1 | Bioinformatics

Disorder tendencies were predicted based on the PfCSP<sub>N</sub> amino acid sequence using DisoMine (Orlando et al., 2022), IUPred3 (Erdős et al., 2021), PONDR VL3-BA (Radivojac et al., 2003), and PONDR VSL2 (Peng et al., 2006). Secondary structures were predicted with AlphaFold2 (Jumper et al., 2021), PSIPRED (Buchan & Jones, 2019; Jones, 1999), and JPRED4 (Drozdetskiy et al., 2015).

### 4.2 | Production and purification PfCSP<sub>SL/TEV</sub> and generation of PfCSP<sub>N</sub>

The PfCSP<sub>SL/TEV</sub> gene is based on *P. falciparum* NF54 CSP (Uniprot ID: P19597) and contains the N-terminal domain (Phe20-Asn101) followed by a TEV cleavage site (ENLYFQSGG), a shortened linker (PNANPNANPNA), the C-terminal domain (Glu310-Ser384), and a hexahistidine tag, and was commercially cloned into the pET21b vector (Genscript). Chemocompetent *E. coli* SHuffle Express T7 cells (NEB) were heat-transformed for gene expression. Pre-cultures were started by inoculation of 2xTY-Gluc-Car medium (2xTY medium at pH 8.2 supplemented with 0.2% glucose and 100  $\mu$ g/mL carbenicillin) with a single colony or cell paste from a glycerol stock

and grown overnight shaking at 37°C. Main cultures of 1 L were inoculated with a 250-fold dilution of pre-culture and grown shaking at 37°C to an OD<sub>600</sub> of 0.6–0.9, at which gene expression was induced with 1 mM IPTG and the temperature was decreased to 20°C for overnight incubation. Cells were harvested 20 h post-induction by centrifugation (Avanti JXN-26, Beckman Coulter) at 5000 rpm and 4°C for 30 min. Cell pellets were resuspended in lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 0.1 mM phenylmethylsulfonyl fluoride, 0.5 mM benzamidine hydrochloride, and 1 tablet of Roche cOmplete EDTA-free protease inhibitor per 50 mL, pH 8.0), flash frozen in liquid nitrogen and stored at –20°C.

For the production of isotopically labeled protein, small cultures of 5 mL LB-Car medium (LB medium supplemented with 100 µg/mL carbenicillin) were inoculated and grown overnight shaking at 37°C. Pre-cultures in adjusted M9 minimal-Car medium (9 g/L K<sub>2</sub>HPO<sub>4</sub>, 5 g/L KH<sub>2</sub>PO<sub>4</sub>, 9 g/L Na<sub>2</sub>HPO<sub>4</sub>, 2.4 g/L K<sub>2</sub>SO<sub>4</sub>, 1 mM MgCl<sub>2</sub>, 0.1% LB medium, 0.25% <sup>13</sup>C D-glucose, 0.1% <sup>15</sup>NH<sub>4</sub>Cl, 6 mg/L FeSO<sub>4</sub>·7H<sub>2</sub>O, 6 mg/L CaCl<sub>2</sub>·2H<sub>2</sub>O, 1.2 mg/L MnCl<sub>2</sub>·4H<sub>2</sub>O, 0.8 mg/L CoCl<sub>2</sub>·6H<sub>2</sub>O, 0.7 mg/L ZnSO<sub>4</sub>·7H<sub>2</sub>O, 0.3 mg/L CuCl<sub>2</sub>·2H<sub>2</sub>O, 0.02 mg/L H<sub>3</sub>BO<sub>4</sub>, 0.25 mg/L (NH<sub>4</sub>)<sub>6</sub>Mo<sub>7</sub>O<sub>24</sub>·4H<sub>2</sub>O, 5 mg/L EDTA, 100 µg/mL carbenicillin) were inoculated with a 100-fold dilution of the small LB culture and grown overnight shaking at 37°C. Main cultures of 0.5 L were inoculated with a 10-fold dilution of the pre-culture and cultured as described above.

Thawed cells were lysed by ultrasonication (180 cycles of 5 s pulses at 20% amplitude and 5 s breaks on Sonics VCX-130) on ice. Cellular debris and inclusion bodies containing PfCSP<sub>SL/TEV</sub> were pelleted by centrifugation at 40,000 × g and 4°C for 1 h. Pellets were solubilized in IB solubilization buffer (20 mM Tris-HCl, 500 mM NaCl, 10–20 mM imidazole, 8 M urea, 2 mM DTT, pH 8.0) by pipetting, ultrasonication on ice (60 cycles), and head-over-head rotation at room temperature. The remaining insoluble cellular debris was pelleted by centrifugation at 40,000 × g for 1 h at 4°C and the supernatant was filtered through a 0.45 µm filter prior to loading on two stacked 5 mL HisTRAP HP columns (Cytiva) at 5 mL/min. Contaminants were removed by washing with 10 column volumes (CVs) Tris buffer A (20 mM Tris-HCl, 500 mM NaCl, 10–20 mM imidazole, 8 M urea, 2 mM DTT, pH 8.0) and PfCSP<sub>SL/TEV</sub> was eluted with Tris buffer B (20 mM Tris-HCl, 500 mM NaCl, 500 mM imidazole, 8 M urea, 2 mM DTT, pH 8.0). The IMAC elution sample was desalted to TEV reaction buffer (50 mM sodium phosphate, 200 mM NaCl, 3 M urea, 2 mM DTT, pH 7.2) using a HiPrep 26/10 Desalting

column (Cytiva) at 5 mL/min and PfCSP<sub>SL/TEV</sub> was cleaved by TEV protease (expressed and purified in-house) during overnight incubation at room temperature. The next morning, the urea and DTT concentrations were increased to 8.0 and 4.0 mM, respectively, and the sample was incubated for 2 h in a head-over-head shaker at room temperature. The solution was filtered through a 0.45 µm filter and loaded on two stacked 5 mL HisTRAP HP columns at 2.5 mL/min to isolate PfCSP<sub>N</sub> in the flow-through. A buffer switch to non-denaturing buffer (50 mM Tris-HCl, 100 mM NaCl, 1 mM TCEP, pH 8.0) was conducted with a HiPrep 26/10 Desalting column at 10 mL/min after which PfCSP<sub>N</sub> was diluted to 0.2 mg/mL and incubated for 2–3 days at 4°C. Aggregates were removed by filtering through a 0.20 µm filter. PfCSP<sub>N</sub> is already highly pure after the reverse IMAC, if necessary, remaining contaminants were removed by size exclusion chromatography (SEC) on a HiLoad 16/60 Superdex 75 column (Cytiva). Purity was analyzed by SDS-PAGE and anti-His Western blotting. The authenticity of PfCSP<sub>N</sub> was verified by excising the band from the SDS-PAGE, in-gel trypsin digestion LC-MS/MS identification as described elsewhere (Kostyuk et al., 2022).

### 4.3 | Heparin-affinity chromatography

PfCSP<sub>N</sub> (~3 mg) in non-denaturing buffer was diluted 1:1 in Tris buffer C (50 mM Tris-HCl, 2 mM DTT, pH 8.0) and loaded on a 1 mL HiTRAP Heparin (Cytiva) column at 0.5 mL/min. Nonbinding proteins were washed out with 5 CVs Tris buffer C and bound protein was eluted with a linear gradient of Tris buffer D (50 mM Tris-HCl, 1 M NaCl, 2 mM DTT, pH 8.0) over 10 CVs.

### 4.4 | Isothermal titration calorimetry

ITC experiments were performed on a MicroCal Peaq-ITC calorimeter system (Malvern Panalytcs) in 50 mM Tris-HCl, 100 mM NaCl, 2 mM TCEP, pH 7.5 at 25°C, a stirring speed of 750 rpm and a reference power of 5 µcal/s. Titrations were performed with 12 successive injections of 3 µL HO12 (550 µM; ΔHexA,2S 1–4 GlcNS,6S 1–4 (IdoUA,2S 1–4 GlcNS,6S)<sub>5</sub>; Iduron) during 6 s, preceded by a dummy injection of 0.4 µL and with 450 s breaks, to a calorimetric cell filled with PfCSP<sub>N</sub> (55 µM; cell volume of 200 µL). Experiments were executed in quadruplicates and the mean and standard deviation are reported. Data analysis was performed with the MicroCal Peaq-ITC software (Malvern Panalytcs), according to a 1:1 binding model.

## 4.5 | HepG2 binding assay

Recombinant PfCSP<sub>N</sub> was fluorescently labeled with Alexa Fluor™ 647 C2 Maleimide (Invitrogen) according to the manufacturer's instructions. HepG2 cells (ATCC) were cultured overnight on 96-well plates (Falcon) at the density of 40,000 per well in DMEM 10% FCS (Gibco) at 37°C, 5% CO<sub>2</sub>. Cells were washed twice with PBS (Gibco), put in suspension using the Cell Dissociation Buffer Enzyme-Free (Gibco), centrifuged at 500 × *g* for 5 min at 4°C and fixed with 2% PFA in PBS for 1 h at 4°C. Fixed cells were washed twice with PBS and incubated with 0, 5, 10, 25, 50, and 100 µg/mL of PfCSP<sub>N</sub> or BSA conjugated with Alexa-Fluor 647 for 1 h at 37°C. Cell suspension was washed twice with PBS, filtered using a 35-µm cell strainer (Falcon) and analyzed by flow cytometry (CytoFlex S, Beckman Coulter) using CytExpert (Beckman Coulter).

## 4.6 | Analytical gel filtration

AGF was performed on an ENrich SEC 70 column (Bio-Rad) in 50 mM Tris-HCl, 150 mM NaCl, 2 mM TCEP, pH 7.5. Samples of 100 µL of PfCSP<sub>N</sub> (at 0.66 mg/mL) and of the Bio-Rad gel filtration standard were injected and eluted at a flow rate of 0.75 mL/min. The apparent molecular mass and hydrodynamic radius of PfCSP<sub>N</sub> were determined according to (Uversky, 1993).

## 4.7 | Dynamic light scattering

DLS data were collected on a DynaPro NanoStar instrument (Wyatt Technology) at 25°C. Samples of 50 µL PfCSP<sub>N</sub> at 1.0 mg/mL (20 mM sodium phosphate, 100 mM NaCl, 2 mM DTT, pH 6.0) were filtered with Costar Spin-X 0.22 µm centrifuge filters (Corning) prior to data collection using disposable DLS cuvettes (Wyatt Technology). Data were processed and analyzed using Dynamics 7.1.9 and RStudio.

## 4.8 | Circular dichroism spectroscopy

Five accumulations were collected on a MOS-500 spectrometer (BioLogic) and averaged. Scans were recorded over a wavelength range of 190–260 nm in steps of 1 nm, a band width of 1.0 nm, and an acquisition time of 0.5 s. The CD spectrum was obtained at room temperature at a concentration of 0.2 mg/mL in 50 mM Tris-HCl, 150 mM NaCl, 2 mM TCEP, 2 µM leupeptin, 125 µM AEBBSF, 100 µM EDTA, pH 7.5. Raw CD data (ellipticity  $\theta$  in

mdeg) were normalized to MRE (MRE in deg.cm<sup>2</sup>.dmol<sup>-1</sup>.res<sup>-1</sup>) according to Equation (1),

$$\text{MRE} = \frac{\theta \cdot \text{MM}}{n \cdot C \cdot l}, \quad (1)$$

with MM, *n*, *C*, and *l* representing the molar mass (Da), the number of residues, the concentration (mg/mL), and the path length (cm), respectively.

## 4.9 | Small-angle X-ray scattering

SEC-SAXS data were collected at the SWING beamline (SOLEIL synchrotron, Saint-Aubin, France) (Thureau et al., 2021) in HPLC mode using a Shodex KW404-4F column pre-equilibrated with SAXS buffer (50 mM MES, 500 mM NaCl, 1 mM TCEP, pH 6.6). The PfCSP<sub>N</sub> sample was concentrated to 10 mg/mL using a centrifugal filter (Amicon Ultra-0.5 3K). Subsequently, a buffer switch was performed using the same device by mixing 50 µL concentrated sample with 450 µL SAXS buffer and concentrating it again to 50 µL. This was repeated five times. Finally, a 50 µL sample was injected at 10 mg/mL and eluted at a flow rate of 0.25 mL/min while scattering data was collected with an exposure time of 990 ms and a dead time of 10 ms. Data were processed and analyzed using the ATSAS package (Manalastas-Cantos et al., 2021), the online SAXSMoW tool (Piiadov et al., 2019) and the RStudio software. The information on data collection and derived structural parameters is summarized in Supplementary Table S1.

## 4.10 | NMR spectroscopy

The NMR spectra were acquired at 298 K on a Bruker Avance III HD 800 MHz spectrometer equipped with a TCI cryoprobe. The samples contained [<sup>13</sup>C, <sup>15</sup>N] PfCSP<sub>N</sub> (50 µM) alone or in the presence of 2 molar equivalents of HO12, 20 mM MES pH 6.6, 100 mM NaCl, 2 mM TCEP, and 6% D<sub>2</sub>O for the lock. The data were acquired, processed, and analyzed in TopSpin 3.6 (Bruker).

## 4.11 | SDS-PAGE

Samples were analyzed by SDS-PAGE under reducing and denaturing conditions using 15% acrylamide gels. Proteins were visualized using PageBlue protein staining solution (Thermo Scientific). PageRuler prestained protein ladder (Thermo Scientific) was used as a molecular reference.

## 4.12 | MDs simulations

The AlphaFold coordinate file of PfCSP was downloaded from UniProt (UniProt ID: P19597), and the *in vitro* PfCSP<sub>N</sub> construct was reproduced from this using PyMOL (Schrödinger, LLC) (Jumper et al., 2021; Varadi et al., 2022). Residue ionization states were generated with MolProbity and H++ (Anandakrishnan et al., 2012; Gordon et al., 2005; Myers et al., 2006; Williams et al., 2018). MD simulations were performed using GROMACS 2021.1 with the CHARMM36m force field (Abraham et al., 2015; Huang et al., 2017; Van Der Spoel et al., 2005). A cubic box of 150 Å in all dimensions and the CHARMM modified TIP3P water model was used (Jorgensen et al., 1983). The box was charge-neutralized and supplemented with sodium and chloride ions to reach a final concentration of 500 mM NaCl. Steepest descent algorithm was used for two sequential energy minimization (EM) procedures. During the second EM procedure, constraints were added to the C-H bonds of the system using the LINCS algorithm (Hess et al., 1997).

Thirty-one runs of the simulations were generated. Of these, 30 runs were set to explore conformational space at 500 K for 10 ns, then lowered to 300 K during 60 ns each. The 31st copy was set to immediately explore conformational space at 300 K for 60 ns. For all replicas, a first 1 ns equilibration step (timesteps of 2 fs using the leap-frog algorithm) was conducted under the canonical ensemble, using the velocity-rescale thermostat (Bussi et al., 2007). The first 30 copies were equilibrated at 500 K and the 31st at 300 K. During equilibration, all protein heavy atoms were position restrained. The second 1 ns equilibration step was ran under the isothermal-isobaric ensemble, using the Berendsen barostat to equilibrate all replicas at 1 bar (Berendsen et al., 1984). A third and final equilibration step was analogous to the second, only this time releasing the position restraints on the protein heavy atoms. Thereafter, pre-simulations of 10 ns (timesteps of 2 fs) were executed for 30 replicas of the system at 500 K. The final frames of these simulations were used as starting points for 30 simulations at 300 K, together with the 31st simulation. 60 ns (timesteps of 2 fs) were simulated for each replica. Every 20 ps, current system data were written out, leading to trajectories of 3000 frames per simulation. In total, 1.86 μs of data were collected at 300 K. These simulations were executed on the UAntwerp Tier-2 supercomputer cluster Leibniz, maintained by the CalcUA infrastructure as part of the Flemish Supercomputer Center.

The 31 simulation trajectories were concatenated and water molecules and ions excluded. PLUMED 2.8.1 was used to calculate the radius of gyration ( $R_g$ ) of the construct at each frame of the concatenated trajectory (Bonomi et al., 2009; The PLUMED consortium, 2019;

Tribello et al., 2014). Secondary structure content, namely  $\alpha$ -helix and  $\beta$ -sheet content, was measured as well using PLUMED. Using NumPy and sci-kit learn, a KMeans clustering algorithm was applied to the  $R_g$ ,  $\alpha$ -helix content and  $\beta$ -sheet content data sets, to subdivide the data points into six clusters (Harris et al., 2020; Hartigan & Wong, 1979; Pedregosa et al., 2011). Secondary structure content observed during the simulation runs was assessed through *s*-value analysis (Pietrucci & Laio, 2009). For  $\alpha$ -helices, *s*-values are calculated by comparing all possible six sequential residue segments of the protein to an ideal  $\alpha$ -helix structure. When the six-residue segment under study closely resembles an  $\alpha$ -helix, a score close to one will be assigned for that segment, and conversely, if no close resemblance is found, a zero score is assigned. All segment scores are then summed into the final *s*-value. The maximum *s*-value for an all  $\alpha$ -helical protein of *n* residues is given by the formula  $(n - 5)$ . An analogous methodology is used to study  $\beta$ -sheets, but with the six sequential residues compared to an ideal  $\beta$ -sheet now being split in two-three residue segments. The maximum *s*-value for an all  $\beta$ -sheet protein of *n* residues containing both parallel and anti-parallel  $\beta$ -sheets is given by the formula  $(n - 7) \times (n - 6)$ . Consequently, high *s*-values likely indicate presence of the secondary structural elements under study, whereas lower values can result from summing many values close to zero. For PfCSP<sub>N</sub> (*n* = 89), maximum *s*-values for  $\alpha$ -helix and  $\beta$ -sheet content are 84 and 6806, respectively. Molecular graphics visualization and analysis were performed with UCSF ChimeraX (Goddard et al., 2018).

### AUTHOR CONTRIBUTIONS

**Rob Geens:** Conceptualization; investigation; writing – original draft; methodology; validation; visualization; writing – review and editing; software; formal analysis; project administration; data curation; supervision; resources. **Jessica Stanisich:** Conceptualization; writing – review and editing; investigation. **Olivier Beyens:** Conceptualization; writing – review and editing; methodology; validation; visualization; investigation; formal analysis; data curation. **Stijn D'Hondt:** Conceptualization; investigation; methodology; validation; visualization; writing – review and editing; formal analysis; data curation. **Jean-Michel Thiberge:** Methodology; investigation; writing – review and editing; validation; formal analysis; data curation; resources. **Amber Ryckebosch:** Investigation; methodology; validation. **Anke De Groot:** Investigation; methodology; validation. **Stefan Magez:** Writing – review and editing; resources; funding acquisition. **Didier Vertommen:** Investigation; methodology; validation; writing – review and editing; formal analysis; resources. **Rogério Amino:** Investigation; writing – review and editing; validation; methodology;

formal analysis; data curation; resources. **Hans De Winter:** Writing – review and editing; investigation; conceptualization; methodology; validation; visualization; formal analysis; data curation. **Alexander N. Volkov:** Conceptualization; investigation; methodology; validation; visualization; writing – review and editing; formal analysis; data curation; supervision; project administration; resources. **Peter Tompa:** Conceptualization; investigation; methodology; validation; visualization; writing – review and editing; formal analysis; data curation; supervision; project administration; resources. **Yann G.-J. Sterckx:** Conceptualization; investigation; funding acquisition; writing – original draft; writing – review and editing; visualization; validation; methodology; software; formal analysis; project administration; resources; supervision; data curation.

## ACKNOWLEDGMENTS

Rob Geens is a doctoral fellow supported by a DOCPRO4-NIEUWZAP (code 40043) grant awarded to Yann G.-J. Sterckx by the University of Antwerp “Bijzonder Onderzoeksfonds (BOF).” Olivier Beyens is a PhD fellow supported by FWO-Vlaanderen (1S04221N). Rogerio Amino is funded by the Agence Nationale de la Recherche (SporoSTOP ANR-19-CE15-0027), and the French Government’s Investissement d’Avenir program, Laboratoire d’Excellence “Integrative Biology of Emerging Infectious Diseases” (ANR-10-LABX-62-IBEID). The authors wish to thank the staff of the SWING beam line at SOLEIL synchrotron (Javier Perez and Thomas Bizien) for outstanding beam line support. The authors acknowledge use of the CalcUA and VSC supercomputing facilities and wish to thank the staff for outstanding support.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The SAXS data are available in the SASBDB under the accession code SASDRF8. All other data are contained within this article or in the supplementary materials.

## ORCID

Peter Tompa  <https://orcid.org/0000-0001-8042-9939>

Yann G.-J. Sterckx  <https://orcid.org/0000-0002-7420-0983>

## REFERENCES

- Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1–2:19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
- Aley SB, Bates MD, Tam JP, Hollingdale MR. Synthetic peptides from the circumsporozoite proteins of *Plasmodium falciparum* and *Plasmodium knowlesi* recognize the human hepatoma cell line HepG2-A16 in vitro. *J Exp Med*. 1986;164(6):1915–22. <https://doi.org/10.1084/jem.164.6.1915>
- Aliprandini E, Tavares J, Panatieri RH, Thiberge S, Yamamoto MM, Silvie O, et al. Cytotoxic anti-circumsporozoite antibodies target malaria sporozoites in the host skin. *Nat Microbiol*. 2018;3(11):1224–33. <https://doi.org/10.1038/s41564-018-0254-z>
- Aly ASI, Vaughan AM, Kappe SHI. Malaria parasite development in the mosquito and infection of the mammalian host. *Annu Rev Microbiol*. 2009;63:195–221. <https://doi.org/10.1146/annurev.micro.091208.073403>
- Anandkrishnan R, Aguilar B, Onufriev AV. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res*. 2012;40(W1):W537–41. <https://doi.org/10.1093/nar/gks375>
- Ancsin JB, Kisilevsky R. A binding site for highly sulfated heparan sulfate is identified in the N terminus of the circumsporozoite protein. *J Biol Chem*. 2004;279(21):21824–32. <https://doi.org/10.1074/jbc.M401979200>
- Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys*. 1984;81(8):3684–90. <https://doi.org/10.1063/1.448118>
- Bermúdez A, Vanegas M, Patarroyo ME. Structural and immunological analysis of circumsporozoite protein peptides: a further step in the identification of potential components of a minimal subunit-based, chemically synthesised antimalarial vaccine. *Vaccine*. 2008;26(52):6908–18. <https://doi.org/10.1016/j.vaccine.2008.09.071>
- Bernadó P, Svergun DI. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol Biosyst*. 2012;8(1):151–67. <https://doi.org/10.1039/C1MB05275F>
- Beutler N, Pholcharee T, Oyen D, Flores-Garcia Y, MacGill RS, Garcia E, et al. A novel CSP C-terminal epitope targeted by an antibody with protective activity against *Plasmodium falciparum*. *PLoS Pathog*. 2022;18(3):e1010409. <https://doi.org/10.1371/journal.ppat.1010409>
- Bongfen SE, Ntsama PM, Offner S, Smith T, Felger I, Tanner M, et al. The N-terminal domain of *Plasmodium falciparum* circumsporozoite protein represents a target of protective immunity. *Vaccine*. 2009;27(2):328–35. <https://doi.org/10.1016/j.vaccine.2008.09.097>
- Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, Raiteri P, et al. PLUMED: a portable plugin for free-energy calculations with molecular dynamics. *Comput Phys Commun*. 2009;180(10):1961–72. <https://doi.org/10.1016/j.cpc.2009.05.011>
- Buchan DWA, Jones DT. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res*. 2019;47:402–7. <https://doi.org/10.1093/nar/gkz297>
- Burke KA, Janke AM, Rhine CL, Fawzi NL. Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II. *Mol Cell*. 2015;60(2):231–41. <https://doi.org/10.1016/j.molcel.2015.09.006>
- Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity-rescaling. *J Chem Phys*. 2007;126(1):014101. <https://doi.org/10.1063/1.2408420>

- Conicella AE, Zerze GH, Mittal J, Fawzi NL. ALS mutations disrupt phase separation mediated by  $\alpha$ -helical structure in the TDP-43 low-complexity C-terminal domain. *Structure*. 2016;24(9):1537–49. <https://doi.org/10.1016/j.str.2016.07.007>
- Coppi A, Natarajan R, Pradel G, Bennett BL, James ER, Roggero MA, 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*. 2011; 208(2):341–56. <https://doi.org/10.1084/jem.20101488>
- Coppi A, Pinzon-Ortiz C, Hutter C, Sinnis P. The plasmodium circumsporozoite protein is proteolytically processed during cell invasion. *J Exp Med*. 2005;201(1):27–33. <https://doi.org/10.1084/jem.20040989>
- Coppi A, Tewari R, Bishop JR, Bennett BL, Lawrence R, Esko JD, et al. Heparan sulfate proteoglycans provide a signal to plasmodium sporozoites to stop migrating and productively invade host cells. *Cell Host Microbe*. 2007;2(5):316–27. <https://doi.org/10.1016/j.chom.2007.10.002>
- Cowman AF, Healer J, Marapana D, Marsh K. Malaria: biology and disease. *Cell*. 2016;167(3):610–24. <https://doi.org/10.1016/j.cell.2016.07.055>
- Csizmók V, Szöllösi E, Friedrich P, Tompa P. A novel two-dimensional electrophoresis technique for the identification of intrinsically unstructured proteins\*. *Mol Cell Proteomics*. 2006; 5(2):265–73. <https://doi.org/10.1074/mcp.M500181-MCP200>
- Doud MB, Koksals AC, Mi L-Z, Song G, Lu C, Springer TA. Unexpected fold in the circumsporozoite protein target of malaria vaccines. *Proc Natl Acad Sci*. 2012;109(20):7817–22. <https://doi.org/10.1073/pnas.1205737109>
- Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res*. 2015; 43(W1):W389–94. <https://doi.org/10.1093/nar/gkv332>
- Durand D, Vivès C, Cannella D, Pérez J, Pebay-Peyroula E, Vachette P, et al. NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *J Struct Biol*. 2010;169(1):45–53. <https://doi.org/10.1016/j.jsb.2009.08.009>
- Erdős G, Pajkos M, Dosztányi Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res*. 2021;49(W1):W297–303. <https://doi.org/10.1093/nar/gkab408>
- Espinosa DA, Gutierrez GM, Rojas-López M, Noe AR, Shi L, Tse S-W, et al. Proteolytic cleavage of the *Plasmodium falciparum* circumsporozoite protein is a target of protective antibodies. *J Infect Dis*. 2015;212(7):1111–9. <https://doi.org/10.1093/infdis/jiv154>
- Finder VH, Vodopivec I, Nitsch RM, Glockshuber R. The recombinant amyloid- $\beta$  peptide A $\beta$ 1–42 aggregates faster and is more neurotoxic than synthetic A $\beta$ 1–42. *J Mol Biol*. 2010;396(1):9–18. <https://doi.org/10.1016/j.jmb.2009.12.016>
- Gantt SM, Clavijo P, Bai X, Esko JD, Sinnis P. Cell adhesion to a motif shared by the malaria circumsporozoite protein and thrombospondin is mediated by its glycosaminoglycan-binding region and not by CSVTCG. *J Biol Chem*. 1997;272(31):19205–13. <https://doi.org/10.1074/jbc.272.31.19205>
- Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci*. 2018;27(1):14–25. <https://doi.org/10.1002/pro.3235>
- Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A. H<sup>+</sup>: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res*. 2005;33(suppl\_2): W368–71. <https://doi.org/10.1093/nar/gki464>
- Guy AJ, Irani V, MacRaid CA, Anders RF, Norton RS, Beeson JG, et al. Insights into the immunological properties of intrinsically disordered malaria proteins using proteome scale predictions. *PLoS One*. 2015;10(10):e0141729. <https://doi.org/10.1371/journal.pone.0141729>
- Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature*. 2020;585(7825):357–62. <https://doi.org/10.1038/s41586-020-2649-2>
- Hartigan JA, Wong MA. Algorithm AS 136: a K-means clustering algorithm. *Appl Stat*. 1979;28(1):100. <https://doi.org/10.2307/2346830>
- Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem*. 1997;18(12):1463–72. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H)
- Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods*. 2017;14(1):71–3. <https://doi.org/10.1038/nmeth.4067>
- Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*. 1999;292(2):195–202.
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79(2):926–35. <https://doi.org/10.1063/1.445869>
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–9. <https://doi.org/10.1038/s41586-021-03819-2>
- Keitany GJ, Sack B, Smithers H, Chen L, Jang IK, Sebastian L, et al. Immunization of mice with live-attenuated late liver stage-arresting *Plasmodium yoelii* parasites generates protective antibody responses to preerythrocytic stages of malaria. *Infect Immun*. 2014;82(12):5143–53. <https://doi.org/10.1128/IAI.02320-14>
- Kisalu NK, Idris AH, Weidle C, Flores-Garcia Y, Flynn BJ, Sack BK, et al. A human monoclonal antibody prevents malaria infection by targeting a new site of vulnerability on the parasite. *Nat Med*. 2018;24(4):408–16. <https://doi.org/10.1038/nm.4512>
- Kostyuk AI, Tossounian M-A, Panova AS, Thauvin M, Raevskii RI, Ezerina D, et al. Hypocrates is a genetically encoded fluorescent biosensor for (pseudo)hypohalous acids and their derivatives. *Nat Commun*. 2022;13(1):Article 1. <https://doi.org/10.1038/s41467-021-27796-2>
- Kucharska I, Binter Š, Murugan R, Scally SW, Ludwig J, Prieto K, et al. High-density binding to *Plasmodium falciparum* circumsporozoite protein repeats by inhibitory antibody elicited in mouse with human immunoglobulin repertoire. *PLoS Pathog*. 2022; 18(11):e1010999. <https://doi.org/10.1371/journal.ppat.1010999>
- Kucharska I, Hossain L, Ivanochko D, Yang Q, Rubinstein JL, Pomès R, et al. Structural basis of *Plasmodium vivax* inhibition by antibodies binding to the circumsporozoite protein repeats. *eLife*. 2022;11:e72908. <https://doi.org/10.7554/eLife.72908>
- Kucharska I, Thai E, Srivastava A, Rubinstein JL, Pomès R, Julien J-P. Structural ordering of the plasmodium berghei circumsporozoite protein repeats by inhibitory antibody 3D11. *eLife*. 2020;9:e59018. <https://doi.org/10.7554/eLife.59018>

- Kumar KA, Sano G, Boscardin S, Nussenzweig RS, Nussenzweig MC, Zavala F, et al. The circumsporozoite protein is an immunodominant protective antigen in irradiated sporozoites. *Nature*. 2006; 444(7121):937–40. <https://doi.org/10.1038/nature05361>
- Malagrino F, Diop A, Pagano L, Nardella C, Toto A, Gianni S. Unveiling induced folding of intrinsically disordered proteins – protein engineering, frustration and emerging themes. *Curr Opin Struct Biol*. 2022;72:153–60. <https://doi.org/10.1016/j.sbi.2021.11.004>
- Manalastas-Cantos K, Konarev PV, Hajizadeh NR, Kikhney AG, Petoukhov MV, Molodenskiy DS, et al. *ATSAS 3.0*: expanded functionality and new tools for small-angle scattering data analysis. *J Appl Cryst*. 2021;54(1):343–55. <https://doi.org/10.1107/S1600576720013412>
- Marques RF, Gimenez AM, Aliprandini E, Novais JT, Cury DP, Watanabe I-S, et al. Protective malaria vaccine in mice based on the plasmodium vivax circumsporozoite protein fused with the mumps nucleocapsid protein. *Vaccines*. 2020;8(2):Article 2. <https://doi.org/10.3390/vaccines8020190>
- McCutchan TF, Kissinger JC, Touray MG, Rogers MJ, Li J, Sullivan M, et al. Comparison of circumsporozoite proteins from avian and mammalian malarias: biological and phylogenetic implications. *Proc Natl Acad Sci*. 1996;93(21):11889–94. <https://doi.org/10.1073/pnas.93.21.11889>
- Miconai A, Wien F, Kernya L, Lee Y-H, Goto Y, Réfrégiers M, et al. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc Natl Acad Sci*. 2015; 112(24):3095–103. <https://doi.org/10.1073/pnas.1500851112>
- Myers J, Grothaus G, Narayanan S, Onufriev A. A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins*. 2006;63(4):928–38. <https://doi.org/10.1002/prot.20922>
- Myung JM, Marshall P, Sinnis P. The Plasmodium circumsporozoite protein is involved in mosquito salivary gland invasion by sporozoites. *Mol Biochem Parasitol*. 2004;133(1):53–9. <https://doi.org/10.1016/j.molbiopara.2003.09.002>
- Orlando G, Raimondi D, Codicè F, Tabaro F, Vranken W. Prediction of disordered regions in proteins with recurrent neural networks and protein dynamics. *J Mol Biol*. 2022;434(12):167579. <https://doi.org/10.1016/j.jmb.2022.167579>
- Oyen D, Torres JL, Cottrell CA, Richter King C, Wilson IA, Ward AB. Cryo-EM structure of *P. falciparum* circumsporozoite protein with a vaccine-elicited antibody is stabilized by somatically mutated inter-Fab contacts. *Sci Adv*. 2018;4(10):eaau8529. <https://doi.org/10.1126/sciadv.aau8529>
- Oyen D, Torres JL, Wille-Reece U, Ockenhouse CF, Emerling D, Glanville J, et al. Structural basis for antibody recognition of the NANP repeats in *Plasmodium falciparum* circumsporozoite protein. *Proc Natl Acad Sci*. 2017;114(48):E10438–45. <https://doi.org/10.1073/pnas.1715812114>
- Patra AP, Sharma S, Ainavarapu SRK. Force spectroscopy of the *Plasmodium falciparum* vaccine candidate circumsporozoite protein suggests a mechanically pliable repeat region. *J Biol Chem*. 2017;292(6):2110–9. <https://doi.org/10.1074/jbc.M116.754796>
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825–30.
- Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinform*. 2006;7(1):208. <https://doi.org/10.1186/1471-2105-7-208>
- Peysseelon F, Ricard-Blum S. Heparin–protein interactions: from affinity and kinetics to biological roles. Application to an interaction network regulating angiogenesis. *Matrix Biol*. 2014;35: 73–81. <https://doi.org/10.1016/j.matbio.2013.11.001>
- Pietrucci F, Laio A. A collective variable for the efficient exploration of protein beta-sheet structures: application to SH3 and GB1. *J Chem Theory Comput*. 2009;5(9):2197–201. <https://doi.org/10.1021/ct900202f>
- Piiadov V, Ares de Araújo E, Oliveira Neto M, Craievich AF, Polikarpov I. SAXSMoW 2.0: online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. *Protein Sci*. 2019;28(2):454–63. <https://doi.org/10.1002/pro.3528>
- Pinzon-Ortiz C, Friedman J, Esko J, Sinnis P. The binding of the circumsporozoite protein to cell surface heparan sulfate proteoglycans is required for plasmodium sporozoite attachment to target cells. *J Biol Chem*. 2001;276(29):26784–91. <https://doi.org/10.1074/jbc.M104038200>
- Plassmeyer ML, Reiter K, Shimp RL, Kotova S, Smith PD, Hurt DE, et al. Structure of the plasmodium falciparum Circumsporozoite protein, a leading malaria vaccine candidate. *J Biol Chem*. 2009; 284(39):26951–63. <https://doi.org/10.1074/jbc.M109.013706>
- Radivojac P, Obradović Z, Brown CJ, Dunker AK. Prediction of boundaries between intrinsically ordered and disordered protein regions. *Pac Symp Biocomput*. 2003;8:216–27. [https://doi.org/10.1142/9789812776303\\_0021](https://doi.org/10.1142/9789812776303_0021)
- Rathore D, Kumar S, Lanar DE, McCutchan TF. Disruption of disulfide linkages of the *Plasmodium falciparum* circumsporozoite protein: effects on cytotoxic and antibody responses in mice. *Mol Biochem Parasitol*. 2001;118(1):75–82. [https://doi.org/10.1016/S0166-6851\(01\)00369-3](https://doi.org/10.1016/S0166-6851(01)00369-3)
- Rathore D, McCutchan TF. Role of cysteines in *Plasmodium falciparum* circumsporozoite protein: interactions with heparin can rejuvenate inactive protein mutants. *Proc Natl Acad Sci*. 2000a; 97(15):8530–5. <https://doi.org/10.1073/pnas.140224597>
- Rathore D, McCutchan TF. The cytotoxic T-lymphocyte epitope of the *Plasmodium falciparum* circumsporozoite protein also modulates the efficiency of receptor-ligand interaction with hepatocytes. *Infect Immun*. 2000b;68(2):740–3. <https://doi.org/10.1128/IAI.68.2.740-743.2000>
- Rathore D, McCutchan TF, Garboczi DN, Toida T, Hernáiz MJ, LeBrun LA, et al. Direct measurement of the interactions of glycosaminoglycans and a heparin decasaccharide with the malaria circumsporozoite protein. *Biochemistry*. 2001;40(38): 11518–24. <https://doi.org/10.1021/bi0105476>
- Rathore D, Sacci JB, de la Vega P, McCutchan TF. Binding and invasion of liver cells by *Plasmodium falciparum* sporozoites. *J Biol Chem*. 2002;277(9):7092–8. <https://doi.org/10.1074/jbc.M106862200>
- Ryan VH, Dignon GL, Zerze GH, Chabata CV, Silva R, Conicella AE, et al. Mechanistic view of hnRNPA2 low-complexity domain structure, interactions, and phase separation altered by mutation and arginine methylation. *Mol Cell*. 2018;69(3):465–479.e7. <https://doi.org/10.1016/j.molcel.2017.12.022>

- Scally SW, Murugan R, Bosch A, Triller G, Costa G, Mordmüller B, et al. Rare PfCSP C-terminal antibodies induced by live sporozoite vaccination are ineffective against malaria infection. *J Exp Med*. 2018;215(1):63–75. <https://doi.org/10.1084/jem.20170869>
- Schwenk R, DeBot M, Porter M, Nikki J, Rein L, Spaccapelo R, et al. IgG2 antibodies against a clinical grade *Plasmodium falciparum* CSP vaccine antigen associate with protection against transgenic sporozoite challenge in mice. *PLoS One*. 2014;9(10):e111020. <https://doi.org/10.1371/journal.pone.0111020>
- Sidjanski SP, Vanderberg JP, Sinnis P. Anopheles stephensi salivary glands bear receptors for region I of the circumsporozoite protein of *Plasmodium falciparum*. *Mol Biochem Parasitol*. 1997;90(1):33–41. [https://doi.org/10.1016/S0166-6851\(97\)00124-2](https://doi.org/10.1016/S0166-6851(97)00124-2)
- Singh AP, Buscaglia CA, Wang Q, Levay A, Nussenzweig DR, Walker JR, et al. Plasmodium circumsporozoite protein promotes the development of the liver stages of the parasite. *Cell*. 2007;131(3):492–504. <https://doi.org/10.1016/j.cell.2007.09.013>
- Singh SK, Plieskatt J, Chourasia BK, Singh V, Bolscher JM, Dechering KJ, et al. The *Plasmodium falciparum* circumsporozoite protein produced in *Lactococcus lactis* is pure and stable. *J Biol Chem*. 2020;295(2):403–14. <https://doi.org/10.1074/jbc.RA119.011268>
- Tan J, Cho H, Pholcharee T, Pereira LS, Doumbo S, Doumtabe D, et al. Functional human IgA targets a conserved site on malaria sporozoites. *Sci Transl Med*. 2021;13(599):eabg2344. <https://doi.org/10.1126/scitranslmed.abg2344>
- Thai E, Costa G, Weyrich A, Murugan R, Oyen D, Flores-Garcia Y, et al. A high-affinity antibody against the CSP N-terminal domain lacks *Plasmodium falciparum* inhibitory activity. *J Exp Med*. 2020;217(11):e20200061. <https://doi.org/10.1084/jem.20200061>
- The PLUMED consortium. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat Methods*. 2019;16(8):667–73. <https://doi.org/10.1038/s41592-019-0506-8>
- Thureau A, Roblin P, Pérez J. BioSAXS on the SWING beamline at synchrotron SOLEIL. *J Appl Cryst*. 2021;54(6):1698–710. <https://doi.org/10.1107/S1600576721008736>
- Tompa P. Intrinsically unstructured proteins. *Trends Biochem Sci*. 2002;27(10):527–33. [https://doi.org/10.1016/s0968-0004\(02\)02169-2](https://doi.org/10.1016/s0968-0004(02)02169-2)
- Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends Biochem Sci*. 2008;33(1):2–8. <https://doi.org/10.1016/j.tibs.2007.10.003>
- Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: new feathers for an old bird. *Comput Phys Commun*. 2014;185(2):604–13. <https://doi.org/10.1016/j.cpc.2013.09.018>
- Triller G, Scally SW, Costa G, Pissarev M, Kreschel C, Bosch A, et al. Natural parasite exposure induces protective human anti-malarial antibodies. *Immunity*. 2017;47(6):1197–1209.e10. <https://doi.org/10.1016/j.immuni.2017.11.007>
- Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Židek A, et al. Highly accurate protein structure prediction for the human proteome. *Nature*. 2021;596(7873):590–6. <https://doi.org/10.1038/s41586-021-03828-1>
- Usynin I, Klotz C, Frevert U. Malaria circumsporozoite protein inhibits the respiratory burst in Kupffer cells. *Cell Microbiol*. 2007;9(11):2610–28. <https://doi.org/10.1111/j.1462-5822.2007.00982.x>
- Uversky VN. Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry*. 1993;32(48):13288–98. <https://doi.org/10.1021/bi00211a042>
- Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible, and free. *J Comput Chem*. 2005;26(16):1701–18. <https://doi.org/10.1002/jcc.20291>
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res*. 2022;50(D1):D439–44. <https://doi.org/10.1093/nar/gkab1061>
- Vijayan K, Visweswaran GRR, Chandrasekaran R, Trakhimets O, Brown SL, Watson A, et al. Antibody interference by a non-neutralizing antibody abrogates humoral protection against *Plasmodium yoelii* liver stage. *Cell Rep*. 2021;36(5):109489. <https://doi.org/10.1016/j.celrep.2021.109489>
- Wang LT, Pereira LS, Flores-Garcia Y, O'Connor J, Flynn BJ, Schön A, et al. A potent anti-malarial human monoclonal antibody targets circumsporozoite protein minor repeats and neutralizes sporozoites in the liver. *Immunity*. 2020;53(4):733–744.e8. <https://doi.org/10.1016/j.immuni.2020.08.014>
- WHO. World malaria report 2022. 2022.
- Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci*. 2018;27(1):293–315. <https://doi.org/10.1002/pro.3330>
- Zhao J, Bhanot P, Hu J, Wang Q. A comprehensive analysis of plasmodium circumsporozoite protein binding to hepatocytes. *PLoS One*. 2016;11(8):e0161607. <https://doi.org/10.1371/journal.pone.0161607>

## SUPPORTING INFORMATION

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

**How to cite this article:** Geens R, Stanisich J, Beyens O, D'Hondt S, Thiberge J-M, Ryckebosch A, et al. Biophysical characterization of the *Plasmodium falciparum* circumsporozoite protein's N-terminal domain. *Protein Science*. 2024;33(1):e4852. <https://doi.org/10.1002/pro.4852>