Cryo-EM reveals the architecture of placental malaria VAR2CSA and provides molecular insight into chondroitin sulfate binding

Kaituo Wang
University of Copenhagen https://orcid.org/0000-0002-5922-7109
Robert Dagil
University of Copenhagen https://orcid.org/0000-0002-5594-0716
Thomas Lavstsen
University of Copenhagen
Sandeep Misra
Department of BioMolecular Sciences, University of Mississippi https://orcid.org/0000-0001-9165-3987
Charlotte Spliid
University of Copenhagen and University of California
Yong Wang
University of Copenhagen https://orcid.org/0000-0001-9156-0377
Tobias Gustavsson
Centre for Medical Parasitology at Department of Immunology and Microbiology, University of Copenhagen and Department of Infectious Diseases, Copenhagen University Hospital
Daniel Sandoval
Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla
Elena Vidal-Calvo
VAR2Pharmaceuticals and University of Copenhagen
Swati Choudhary
Centre for Medical Parasitology at Department of Immunology and Microbiology, University of Copenhagen and Department of Infectious Diseases, Copenhagen University Hospital
Mette Agerbæk
Centre for Medical Parasitology at Department of Immunology and Microbiology, University of Copenhagen and Department of Infectious Diseases, Copenhagen University Hospital
Kresten Lindorff-Larsen
University of Copenhagen https://orcid.org/0000-0002-4750-6039
Morten Nielsen
Centre for Medical Parasitology at Department of Immunology and Microbiology, University of Copenhagen and Department of Infectious Diseases, Copenhagen University Hospital

https://orcid.org/0000-0003-2668-4992
Thor Theander
University of Copenhagen
Joshua Sharp
Department of BioMolecular Sciences, University of Mississippi https://orcid.org/0000-0002-0115-0276
Thomas Clausen
Department of Cellular and Molecular Medicine, University of California, San Diego, La Joll, USA.
Pontus Gourdon
University of Copenhagen https://orcid.org/0000-0002-8631-3539
Ali Salanti (**salanti@sund.ku.dk**)
University of Copenhagen and Department of Infectious Diseases, Copenhagen University Hospital

---

Article

Keywords: placental malaria VAR2CS, cryo-EM

Posted Date: January 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-121821/v1

Version of Record: A version of this preprint was published at Nature Communications on May 19th, 2021. See the published version at https://doi.org/10.1038/s41467-021-23254-1.

Cryo-EM reveals the architecture of placental malaria VAR2CSA and provides molecular insight into chondroitin sulfate binding

Kaituo Wang[1*], Robert Dagil[2*], Thomas Lavstsen[2], Sandeep K. Misra[3], Charlotte B. Spliid[2,4], Yong Wang[5], Tobias Gustavsson[2,6], Daniel R. Sandoval[4], Elena Ethel Vidal-Calvo[2,6], Swati Choudhary[2], Mette Ø Agerbaek[2], Kresten Lindorff-Larsen[5], Morten A. Nielsen[2], Thor G. Theander[2], Joshua S. Sharp[3,7], Thomas Mandel Clausen[4], Pontus Gourdon[1,8#], Ali Salanti[2#]

* shared first authorship

[1] Department of Biomedical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark
[2] Centre for Medical Parasitology at Department for Immunology and Microbiology, Faculty of Health and Medical Sciences, University of Copenhagen and Department of Infectious Disease, Copenhagen University Hospital, 2200 Copenhagen, Denmark
[3] Department of BioMolecular Sciences, University of Mississippi, University, MS, United States
[4] Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92093, United States.
[5] Structural Biology and NMR Laboratory & Linderstrøm-Lang Centre for Protein Science Department of Biology, University of Copenhagen, Denmark
[6] VAR2Pharmaceuticals, Ole Maaloesvej 3, 2200 N Denmark.
[7] Department of Chemistry and Biochemistry, University of Mississippi, University, MS, United States
[8] Department of Experimental Medical Science, Lund University, Lund, Sweden

Corresponding authors: Pontus Gourdon (pontus@sund.ku.dk); Ali Salanti (Salanti@sund.ku.dk)

Abstract

Malaria during pregnancy remains a major health problem in *Plasmodium falciparum* endemic areas. Parasite-infected red blood cells sequester in the placenta through interaction between parasite-expressed protein VAR2CSA and the glycosaminoglycan chondroitin sulfate A (CS) abundantly present in the intervillous space. Placental malaria can have severe consequences for both mother and child by causing maternal anemia, low birth weight and stillbirth. Several VAR2CSA-based vaccines have been developed and clinically tested but they have failed to induce an antibody response that effectively inhibits placental adhesion of different genetic variants of VAR2CSA. The interaction between VAR2CSA and CS represents a unique case of an evolution-driven high-affinity interaction between a protein and an oncofetal carbohydrate. Here, we report cryo-EM structures of the VAR2CSA ectodomain up to 3.1 Å resolution revealing an overall V-shaped architecture and a complex domain organization. Notably, the surface displays a single significantly electropositive patch, compatible with binding of negatively charged CS. Using molecular docking and molecular dynamics

simulations as well as comparative hydroxyl radical protein foot-printing of VAR2CSA in complex with placental CS, we identify the CS-binding groove, intersecting with the positively charged patch of the central VAR2CSA structure. We identify distinctive conserved structural features upholding the macro-molecular domain complex and CS binding capacity of VAR2CSA as well as divergent elements possibly allowing immune escape at or near the CS binding site. These observations enable rational design of second-generation placental malaria vaccines eliciting broadly VAR2CSA-reactive antibodies and novel cancer therapies.

Introduction

Plasmodium falciparum is the most lethal human malaria parasite. During the erythrocytic stage, the parasites infect, multiply, rupture and re-infect red blood cells. Infected erythrocytes are effectively removed from the blood by splenic filtering. However, to avoid splenic clearance the parasites express proteins on the surface of the infected erythrocytes that anchors these cells to the host vasculature. The binding is mediated by members of a family of proteins called *Plasmodium falciparum* Erythrocyte Membrane Protein 1 (PfEMP1)[1]. In each parasite genome, PfEMP1s are encoded by approximately 60 different *var* genes, including *var2csa*[2, 3]. The transcription of *var* genes is regulated in a mutually exclusive manner, ensuring that only one PfEMP1 variant is expressed on the surface of an infected erythrocytes at any given time[4]. VAR2CSA allows infected erythrocytes to sequester in the placental vasculature, causing placental malaria[5]. While infections in pregnant women are often clinically silent, they do cause maternal anemia and significantly impair fetal growth. Thus, placental malaria is estimated to result in 900,000 low birth weight deliveries each year in Africa. It has long been known that in malaria-endemic areas pregnant women have a higher risk of being infected with malaria than non-pregnant women[6]. This is due to the establishment of the placenta, which creates a new niche for binding of infected erythrocytes during pregnancy. Early work from Fried and Duffy showed that the malaria parasites accumulate in the placenta by binding to a glycosaminoglycan of the chondroitin sulfate (CS) A type abundantly present in the placenta[7]. Later, it was identified that infected erythrocytes binding to placental chondroitin sulfate A (plCS) expressed a PfEMP1 gene named *var2csa*[5]. After exposure to malaria in pregnancy, women develop antibodies against VAR2CSA, which inhibit parasites from binding to placental tissue, and protect against placental malaria during subsequent pregnancies. Therefore, VAR2CSA is utilized for the development of vaccines protecting women against placental malaria, and two vaccines are currently in clinical development[8, 9].

VAR2CSA is embedded in the erythrocyte membrane through a single C-terminal transmembrane spanning segment. Compared to other PfEMP1s, the ~310 kDa VAR2CSA ectodomain is relatively conserved among parasites[10]. It is comprised of a short N-terminal segment and six Duffy Binding-Like (DBL) domains (DBL1-DBL6) unique to VAR2CSA. Several of the DBL domains are linked by complex inter-domain (ID1-3) regions with limited or no homology among PfEMP1s yielding a DBL1-ID1-DBL2-ID2-DBL3-DBL4-ID3-DBL5-ID4-DBL6 structure[10].

The ligand for VAR2CSA is Chondroitin Sulfate A (CS), a glycosaminoglycan (GAG), which is comprised of repeating disaccharide units of N-acetyl-D-galactosamine (GalNAc) and glucuronic acid (GlcA). CS can be attached to proteins as non-branched linear GAG sidechains, and thereby form proteoglycans. In the placenta the CS modified major receptor is syndecan-1[11, 12]. The CS disaccharide units can be modified by the addition of sulfate groups, such as sulfation of the hydroxyl groups at C2 of GlcA, and/or C4 (4-*O*-sulfation) and C6 (6-*O*-sulfation) of GalNAc. Several studies suggest that VAR2CSA interacts with a CS saccharide of a length between 12-16 saccharides consisting primarily, but not exclusively of 4-*O*-sulfated units[13, 14]. Interestingly, the specific CS signature recognized by VAR2CSA, is not only found in placenta, but also on almost all malignant tumors[13, 15]. This has enabled specific targeting of solid tumors as well as capture of circulating tumor cells using recombinant VAR2CSA[16].

Structures are available for some of the single VAR2CSA DBL domains, and low-resolution envelope structures are available for the full-length ectodomain lacking the C-terminal transmembrane helix[17-19]. However, there are no reports on how the domains are arranged or to what degree they interact, and in particular there is no structural insight into the molecular mechanism of binding to CS. This has complicated vaccine design, which has relied on empirical screening of large panels of N- and C-terminally truncated recombinant proteins. The approach has defined DBL2 and flanking regions as the minimal CSA binding region[20]. Thus, the recombinant VAR2CSA proteins in clinical development are the PrimVac construct comprising the ~100 kDa DBL1-DBL2, and the PamVac construct comprising the ~70 kDa ID1-ID2a region[8, 9]. These VAR2CSA forms maintain high affinity binding to CSA and induces high levels of inhibiting antibodies towards the homolog parasite variants but they are less effective in inhibiting the binding of heterologous variants. This probably reflects that both vaccine proteins include regions that are diverse in sequence among VAR2CSA variants, and therefore likely to confer antigenic diversity. High-resolution structural insight and identification of the residues involved in the CS interaction is therefore needed for the development of more effective vaccines to protect against placental malaria.

Here, we describe the structure of the VAR2CSA ectodomain, displaying an intricate core assembly of N-terminal domains, followed by more loosely arranged C-terminal domains. The structure supports that DBL2 is central for charge-complementation CS binding, exposing a series of positively charged residues on the surface. Using molecular docking of CS oligosaccharides, we show that atypical DBL features of VAR2CSA DBL2 domains form the structural basis for CS binding. By fast photochemical oxidation of proteins (FPOP) analyses of VAR2CSA in complex with placental CS and mutational analyses, we validate the structural and surface exposed regions involved in CS binding and demonstrate that the groove can encompass a CS oligo of around 13 saccharides. Structural loops, polymorphic among VAR2CSA variants, found near the CS binding site may have evolved to escape antibody recognition. These observations enable design of new VAR2CSA-based placental malaria vaccines, providing hope for a vaccine eliciting a broadly reactive, parasite binding inhibiting antibodies.

Results and discussion

Expression and buffer selection of full length VAR2CSA

The extracellular region from the N-terminal methionine (M1) to amino acid F2649 representing the entire ectodomain of the FCR3 VAR2CSA variant was expressed in baculovirus transfected insect cells as a secreted monomeric protein and purified by immobilized nickel affinity chromatography (Fig. 1a). A selection of different post purification formulation buffers and salts were screened and analyzed using size-exclusion chromatography. By comparing neutral pH buffers with either NaCl or KCl as electrolyte, we observed that KCl containing buffers yielded a more compact VAR2CSA configuration (elution at 9.9ml in NaCl and 12.1ml in KCl) (Fig. 1b). To further test the different buffers, we measured the binding kinetics between the CS proteoglycan CSPG and VAR2CSA using a Quartz-Crystal Microscale biosensor. From the fitted association and dissociation rates, we obtained VAR2CSA binding affinities, which were 10-fold stronger in KCl (k_D = 0.1nM) than in NaCl buffer (k_D =1.4nM) (Fig. 1c). This indicates that the VAR2CSA conformation and associated CS-binding are highly dependent on the electrolyte present, which may be important during cellular trafficking to the erythrocyte membrane and for ligand binding in the placenta milieu. Structural analyses were therefore performed in the KCl buffer.

Cryo-EM structure of the VAR2CSA ectodomain

Next, we determined the cryo-EM structure of the isolated VAR2CSA form generated from a map of overall 3.8 Å resolution. The final model represents the ectodomain of VAR2CSA, exhibiting a well-folded conformation with clearly defined DBL domains (Fig. 1d, Fig. S1a) and in agreement with previously published structures of DBL3, DBL4 and DBL6[19, 21, 22]. The region spanning ID1-DBL2-ID2-DBL3-DBL4-ID3 represents the core of the macromolecular structure, as highlighted by a compact structure with a high degree of inter-domain interactions (Fig. 1d). This observation aligns with the structural envelope proposed by SAXS analyses previously[17]. The structure discloses the fold of ID2, forming a separate domain consisting of a bundle of α-helices, a shape which together with ID3, a 30 residues long α-helix (residue 1955-1985), serves as a structural glue for the core. Notably, ID2 and ID3 are directly linked, as the C-terminus of ID2 forms an open-mouth shape conformation, which interacts extensively with the ID3 α-helix (Fig. 1e, Fig. S2a, b). Underscoring the significance of the ID2-ID3 interface, VAR2CSA sequence conservation analysis indicates that the ID2 and ID3 interacting regions are highly conserved (Fig. S3a,b), suggesting a common fold for all VAR2CSA variants. The structure of the core is further maintained through interactions between several DBL domains and inter-domain-links. Specifically, dense contacts are detected between DBL2 and DBL4, ID2 and DBL4, ID2 and DBL3 as well as DBL3 and DBL4. The fold of the DBL3-DBL4 region is similar to the previously described fold (PDB-ID 4P1T)[21] (Fig. S4). DBL1 exhibits only a few inter-domain interaction sites with DBL2 and DBL4, leaving a cleft in-between the core and DBL1. Likewise, the C-terminal DBL5 and DBL6 are more separated from the core, linking the protein to the membrane in a physiological setting. The DBL5 and DBL6 form a rod-like structure with weak inter-domain contacts (Fig. 1d).

VAR2CSA Cryo-EM structure in the presence of placental CS

To shed further light on how VAR2CSA binds CS, we solved the cryo-EM structure of VAR2CSA supplemented with CS purified from placental tissue (plCS). Compared to the structure determined in the absence of CS, the new data yielded a 3.1 Å resolution map overall, extending to 2.8 Å in the core region (Fig. S1b, c). By contrast, DBL1 and DBL5/6 were less well resolved compared to the apo structure (Fig. S1a, Fig. S5). Flexibility of these domains is further supported by two separately generated maps based on different subsets of particles, showing the same core structure, but showing different conformations of DBL5/DBL6 (Fig. S1d, Fig. S6). Thus, the plCS-VAR2CSA cryo-EM data indicated that the CS ligand stabilizes the VAR2CSA core region, while the peripheral regions become flexible and potentially displaced. Similarly, these data indicate that plCS do not induce any conformational changes

in the core, as structures of the DBL2-ID2-DBL3-DBL4-ID3 domains were highly similar in the presence or absence of plCS. Although plCS binding affinity to VAR2CSA is in the low nanomolar range, we did not observe any additional density that could represent ligand binding in the presence of plCS. This may be due to the intrinsic heterogeneity of plCS prepared from its natural placental source. Interestingly, analysis of the surface electrostatics revealed a single strongly positively patch, situated in a groove that spans through the VAR2CSA core which could encompass a negatively charged CS oligosaccharide (Fig. 1f). The positive patch is located in the cleft between DBL1 and the core, in-between DBL1, DBL2 and DBL4 and includes the experimentally defined minimal CS binding region, consisting of DBL2 and flanking inter-domain-stretches[17]. Thus, our cryo-EM data were in agreement with a key role of the positively charged surface region around the DBL2 domain for binding of CS. We also noted that the DBL1-core cleft, which extends a possible binding region beyond the electropositive area, to a large extent is formed by DBL4. This is of interest as DBL4 was previously a lead candidate for vaccine development, because recombinant DBL4 protein can elicit adhesion-blocking antibodies[23].Interestingly, we previously demonstrated that to induce functional inhibitory antibodies the DBL4 domain boundaries required the inclusion of the ID3 region, which now appears close to the proposed CS binding groove[23].

VAR2CSA-specific DBL elements confer both CS binding and immune escape

We noted that key features of the VAR2CSA structure are established through distinct adaptations of the DBL domains. DBL domains are comprised of three subdomains (S1-3) and maintained by relatively conserved residues primarily found in helix-forming sequence stretches (homology blocks 1-5 or HB1-5)[10] (Fig. S7). Previously characterized PfEMP1-protein receptor interactions are mediated by residues in, or adjacent to, HB1[24-27]. Atypical for DBL domains in general, the HB1 helix in VAR2CSA DBL2 is broken up mid-helix by an insertion forming a flexible loop (Fig. 2a, Fig. S7). This loop is glycine-rich, but polymorphic between VAR2CSA variants varying from 4-11 amino acid in length. Compared to DBL3 and DBL4, the C-terminal of HB1 in DBL2 forms part of the highly positively charged surface in the core of the structure (Fig. 2b). Thus, while irregular in its makeup, the DBL2 is similar to other PfEMP1 domains, in that the proposed key ligand-binding site is located in the C-terminal of HB1. The insertion of a structural and sequence variable region near this site, may reflect evolutionary selection for antigenic variable elements providing protection and escape from antibody recognition.

At the other end of the CS binding groove, there is a Lysine-rich loop flanked by the WIW-motif that is conserved in all VAR2CSA DBL2s (Fig. 2c, d). This region contains a tryptophan conserved in all DBL domains (W558), which serves the function of stabilizing the overall DBL structure through interactions with another tryptophan in HB2 (W776) of the DBL2 (Fig. 2d). The region is structurally and sequence polymorphic among different DBL domains, but particularly enriched in positively charged amino acids in VAR2CSA DBL2 (Fig. 2b, d). Interestingly, a similar region together with positively charged residues near the HB1 C-terminal forms the glycan binding channels of the *P. falciparum* EBA-175 protein that mediate binding of parasites to glycophorin A, when merozoites invade erythrocytes[28]. Altogether, these observations support the hypotheses that CS binds a positively charged groove sloping through the VAR2CSA core structure, and that the lysine-rich region in the DBL2 N-terminus loops provides an ideal albeit antigenically diverse entry point to the groove.

In DBL4, the HB5 is also unusually elongated and broken up mid-helix. Here, an 8-15 amino acid long loop is inserted, presenting a conserved Arginine-Lysine pair possibly binding the protruding CS chain, and a sequence-variable flexible stretch of 2-7 amino acids, possibly shielding the one end of the CS-binding groove from antibody recognition.

Mapping CS-protected sites by FPOP-MS

To gain additional support for the proposed CS binding groove, we expressed the previously defined CS binding region as a recombinant DBL1-ID2 protein. Using this DBL1-ID2, we applied Fast Photochemical Oxidation of Proteins (FPOP) with and without plCS complexed to the protein. Following treatment of the VAR2CSA or plCS-VAR2CSA complex with chymotrypsin, oxidation differences in the respective peptides were compared. We identified various regions primarily in ID1, DBL2 and ID2 that exhibited protection from solvent upon binding to plCS (Fig. 3a, Fig. S8). The protection of the ID1, DBL2 and ID2 domains support the notions of either a flexible binding site with partial or dynamic engagement of multiple residues of the protein, and/or a binding-induced conformational change resulting in widespread shielding of multiple parts of these domains from solvent. The peptide that showed the largest absolute reduction in oxidation after plCS binding was peptide 543-558. This peptide and the C-terminal adjacent peptide, also protected by plCS, spans the above-mentioned lysine-rich region N-terminal to DBL2 HB4 (Fig. S8). In addition, the FPOP analyses showed protection of the HB1 DBL2 region. Together, these observations support the role of the positively charged DBL2 surface and HB1 groove (Fig. 1f) as the CS binding region.

Mutagenesis of DBL1-ID2 for CSA binding assay

Unusual for DBL domains, the DBL2 N-terminal region contains an additional tryptophan residue, creating the conserved "WIW-motif" (Fig. 2d). In the FCR3 VAR2CSA variant studied here, the WIW-motif is flanked by several lysine residues ^{555}KKK**WIW**KK^{562}. To assess the importance of folding of this region for CS binding, we produced a recombinant DBL1-ID2 protein with a ^{555}KKK**AIA**KK^{562} substitution to alter the configuration of the loop. The integrity and purity of the wild-type and mutant protein was analyzed by SDS-PAGE (Fig. 3b) and measurement of melting temperature. The melting temperature of the mutant showed a decrease of ~4°C compared to the wild-type protein, however, a clear transition from folded to unfolded was observed as for the wild type (Fig. 3c). When assessing the CSPG interaction using a biosensor, binding of the mutated protein was almost abolished compared to wild type (Fig. 3d). Similar results were obtained when binding was evaluated through binding to CS expressed a lung carcinoma cell line evaluated by flow cytometry (Fig. 3e). These results underscore the importance of the WIW-motif in positioning the positively charged residues for CS binding.

Modelling the VAR2CSA-CS complex

Combining all structural and biochemical results, we generated a docking model of VAR2CSA in complex with a CS 20-mer oligosaccharide, whose stability was validated by unbiased atomistic molecular dynamics simulations. Strikingly, the docking model showed a near perfect fitting of the CS oligo chain along the positively charged groove (Fig. 4a, b) identified in Fig. 1e. The MD simulations revealed a stable core region with about 7 disaccharide units (Fig. 4c, d), in good agreement with previous experimental estimation[14]. The simulations also suggested significant flexibility of DBL1 and DBL5/6 in consistent with the relatively poor EM density in these regions (Fig. S9). Combined with the fact that the DBL1 appeared more flexible in the CS-bound cryo-EM structure, it may indicate that DBL1 acts as a cover, which protects the CS binding region from antibody recognition prior to ligand binding, and is displaced once CS is bound.

Conclusion

The cryo-EM analysis of VAR2CSA demonstrates the overall architecture and the domain arrangement achieved through intricate interactions between DBL domains. This assembly facilitates the binding of long chain CS molecules through a positively charged electrostatic groove sloping through the protein core (Fig. 4a). The center of this groove can accommodate a CS oligosaccharide of around 7 disaccharide units (Fig 4c). Our data suggest that the peripheral DBL1, 5 and 6 domains are displaced upon ligand interaction, whereas the core maintains its structure. The binding site is surrounded by flexible loops containing polymorphic VAR2CSA sequences possibly evolved to provide escape from immune recognition. Malaria vaccines based on the CS binding region of VAR2CSA have failed to induce anti-adhesion responses that cross-react with heterologous variants of VAR2CSA. The high-resolution structure and insight into the binding region provides hope for future development of an intelligently designed cross reactive placental malaria vaccine, targeting conserved features of the CS binding groove. Further, our cryo-EM structure could enable engineering of novel cancer therapeutics targeting oncofetal CS.

Methods

Protein cloning, expression and purification

The VAR2CSA extracellular fragment starting from the N-terminal methionine M1 to amino acid F2649, located prior the putative transmembrane region, was amplified from codon optimized FCR3 (GenBank™ accession no. GU249598) and cloned into baculovirus vector pAcGP67-A (BD Biosciences) including a V5 and 6xHis-tag at the C-terminal. The construct was co-transfected into Sf9 cells to generate virus particles used to infect high-five insect cells (as previously described[17, 29]). Expression was induced for 2 days and supernatant cleared by centrifugation at 10,000 g for 15 min at 4°C. The supernatant was concentrated and buffer-exchanged into PBS pH 7.4 buffer using a 50,000 MWCO hollow-fiber filter. Imidazole to a final concentration of 60mM was added and then loaded onto a 5ml HisTrap HP (Cytiva) column. The protein was eluted using a linear gradient toward PBS with 300mM imidazole pH 7.4. Purity and homogeneity were verified by SDS-PAGE and protein was aliquoted and stored at -80°C. VAR2CSA wild-type construct DBL1-ID2 was expressed in *E. coli* and purified as earlier described [30]. DBL1-ID2 ^{555}KKK**A**I**A**KK^{562} mutant was generated using Thermo Fisher site directed mutagenesis kit and verified by sanger sequencing. Protein expression and purification was performed similar to wild-type DBL1-ID2.

Cryo-EM sample preparation

For apo sample preparation, a frozen aliquot of VAR2CSA was subjected to size-exclusion chromatography in 20mM Tris pH 7.5, 125mM KCl. The generated protein peak was concentrated to 8.6mg/ml and fluorinated Fos-Cholin8 (FC8) was added to the protein sample to a final concentration of 1mM immediately before freezing. Quantifoil 1.2/1.3 holy carbon grids were glow-discharged with Leica Coater ACE 200 for 30s using 5mA current. Cryo-EM grids were prepared with a Vitrobot Mark IV operated at 100% humidity and 4°C. A 3.5ul aliquot of purified protein was applied to the grids, incubated for 5s, blotted 3.5s and plunge frozen into liquid ethane. The sample with CS was prepared by incubating VAR2CSA with placenta CS (purified as previously described[31]) in a 1:7 mass ratio for 30 min at room temperature. The sample was chondroitinase ABC treated with 300mU enzyme for 1.5h and further purified using a superdex200 increase 10/300GL column equilibrated in 20mM Tris pH 7.5 and 75mM KCl. Peak fractions were concentrated to 0.5mg/ml. The cryo-freezing settings were identical except no FC8 was added.

Cryo-EM data collection and data processing

The apo sample cryo-EM datasets were collected on a Titan Krios electron microscope (FEI) operated at 300kV with a Falcon K3 direct detector camera. A total of 8,081 movies were recorded under linear mode at pixel size of 0.832Å and a total dose of 60 e/Å2 spread over 19 frames. Cryo-EM data were processed using cryosparc[32]. The initial processing steps were full-frame-motion correction and CTF determination with ctffind4[33] (wrapped in cryosparc). Blob particle picking was done in a subset (2,166) of the micrographs and particles were extracted and 2D classified to prepare a template-based particle selection. All template selected particles were re-extracted with local-motion correction with dose-weighting using a box size of 440 pixels. A total of 940,807 auto-picked particles were subjected to several rounds of reference-free two-dimensional class averaging to clean-up clearly defective particles. The cleaned-up particle set was processed with the standard cryosparc workflow, including ab-initial model reconstitution, multiple rounds of heterogeneous refinement and non-uniform refinement iterations. The final map was calculated using a subset of 102,676 particles to an overall resolution of 3.82 Å, with the best parts stretching to about 3.6Å (Fig S5).

The CS treated sample dataset (in total 5,234 movies) was collected at pixel size of 0.832Å in Falcon K3 counting mode. The total dose was 40e/Å2 over 40 frames. With similar data processing stratagy as for the apo structure, a total of 1,022,972 particles was extracted and a final subset of 266,774 particles was used to produce a map of overall 3.1Å resolution, with the more well-resolved portion reaching about 2.8Å (Fig. S6). The final dataset could be further

classified into three different subsets, with almost identical core-structure but displaying variance in the DBL5/6 domains.

"Imbalance" hetero refinement

During data processing, one special trick was used in cryosparc hetero-refinement step inspired by the "random-phase 3D Classification" method described by Xin et.al.[34]. For example, after one round the NU-refinement, the final map is low-pass-filtered to resolution of 20 Å. In the hetero-refinement step, the map and the 20Å-low-pass-filtered map was used as two initial models and the "Initial resolution" parameter was set to a number 8Å. This will generate two "imbalance" sub-set of particles with different size, the low-pass-filtered map could serve as "Trash can" to put the lower quality particles that have lower agreement with the high-resolution map. By adjusting the low-pass-filter resolution, initial resolution, box-size, and number of iteration refinement, usually 5% to 25% of low-quality particles could be removed each round, and resolution/map quality will increase until reaching consensus (Supplementary Fig S5 and S6).

Model building and refinement

Published crystal structures of DBL3/4 (PDB-ID 4P1T)[21] and DBL6 (PDB-ID 2Y8D)[35] were directly docked into the corresponding region with good fit. For DBL domains lacking published structures (DBL1/2/4/5), the initial models were generated using the corresponding sequences and PDB-ID 2WAU as a template using the SWISS-MODEL online server. In the 3.1Å map, the core structure (residue AA556-1985) was built *de novo* with high confidence except for some flexible loop regions. The model building was done iteratively using COOT[36] and phenix_real_space_refine of the Phenix software package[37]. Secondary structure restraints and Ramachandran restraints were also imposed during refinement. The resolution of DBL1 and DBL5 was insufficient for *de novo* model building, and these domains were refined using homology model and later MD-assisted model refinement. The quality of the model was validated assessed using Molprobity[38] (see S1 Table for statistics). The apo (EMD-12017, PDB-ID 7B52) and CS (EMD-12018, PDBID 7B54) structures will be released upon publication. The overall structure (Fig. 1d) was generated using the 3.8Å apo map. All other structure figures were generated using the CS structure.

Molecular docking and MD simulations of the VAR2CSA in complex with CS

The refinement models of core domains together with the homology model of low resolution domains were subsequently threaded by adding missing residues using Modeller9.18[39] to construct a full-length model of VAR2CSA. Missing residues within and between domains were modeled as unstructured loops. The full-length model was then used as a template to fit into the Cryo-EM density map of the apo and CS VAR2CSA structures, respectively, using molecular dynamics flexible fitting (MDFF) method with secondary structure, cis-peptide and chirality restraints to prevent overfitting [40]. MDFF was performed using an implicit solvent with a scaling factor of the map potential, g = 0.3. The models were refined by multiple rounds of manual adjustment in PyMol and optimization in MDFF, as well as energy minimization in Gromacs (version 2019)[41] using CHARMM36m force field [42]. CS was modeled as a CSA 20mer which was subsequently used to dock with the highly positively charged regions of VAR2CSA using HADDOCK 2.4 [43]. The docked structures were further refined by energy minimization using Gromacs. Electrostatic surfaces were analyzed using APBS plugin in PyMol.

The docking model of VAR2CSA in complex with CS was subsequently placed into a periodic cubic box with sides of 18.7 nm solvated with TIP3P water molecules containing K+ and Cl− ions at 0.1 M, resulting in 193271 molecules (620728 atoms) in total. The CHARMM36m force field was used for the protein. Force field parameters for CSA were generated using the Glycan Modeler module in the CHARMM-GUI web interface[44] . Neighbor searching was performed every 20 steps. The PME algorithm was used for electrostatic interactions with a cut-off of 1.2 nm. A reciprocal grid of 160 x 160 x 160 cells was used with 4th order B-spline interpolation. A single cut-off of 1.2 nm was used for Van der Waals interactions. Temperature coupling was done with the Nose-Hoover algorithm. Pressure coupling was done with the Parrinello-Rahman algorithm. The hydrogen mass repartitioning technique[45] was employed with a single LINCS iteration (expansion order 4), allowing simulations to be performed with an integration time step of 4 fs. MD simulations of VAR2CSA complex were performed using Gromacs 2019.5. The interactions between VAR2CSA and the bound CS were analyzed by GetContacts scripts (https://getcontacts.github.io/). The flexibility of VAR2CSA domains and CS was analyzed using Gromacs rmsd and rmsf tools.

Attana kinetic measurements

Kinetic analysis of VAR2CSA binding to CSPG decorin was performed on a quartz crystal microbalance biosensor (Attana A200, Attana AB). Decorin CSPG was immobilized on a LNB carboxyl gold-coated sensor chip by amine coupling using S-NHS and EDC. The sensor chip was stabilized at 25µl/min and 22°C in 20mM tris pH 7.5 with either 75mM KCl or 75mM NaCl and VAR2CSA was used as analyte in a 2-fold dilution series from either 200nM or 50nM to 6.25nM. For DBL1-ID2 wild-type and the KKKAIAKK mutant, the binding kinetics were measured in PBS pH 7.4. After each protein injection, the sensor surface was regenerated by 0.1M NaOH. Buffer injections were subtracted from each binding curve and fitted to a 1:1 or 1:2 binding model using TraceDrawer (Ridgeview Instruments AB).

NanoDSF

Protein stability was assessed using NanoDSF instrument (Nanotemper Prometheus NT.48). DBL1-ID2 wild-type and the KKKAIAKK mutant were loaded into capillaries and heated by 1°C/min from 20°C to 90°C. The fluorescence of 330nm and 350nm was monitored during heating and the ratio used to calculate T_m when 50% of the protein was unfolded.

Fast Photochemical Oxidation of Proteins (FPOP):

Glutamine, sodium phosphate, and catalase were purchased from Sigma-Aldrich (St. Louis, MO). LCMS-grade formic acid, water, acetonitrile, hydrogen peroxide and adenine were obtained from Fisher Scientific (Fair Lawn, NJ). Fused silica capillary was purchased from Molex, LLC (Lisle, IL). Sequencing grade modified trypsin was purchased from Promega (Madison, WI).

A final concentration of 2 µM DBL1-ID2 protein was incubated in 10 mM sodium phosphate buffer in the presence or absence of 2 µM placental chondroitin sulfate at pH 7.8 for one hour. FPOP was performed as described previously [46]. Briefly, 20 µl of protein sample mixture 1 mM adenine, 17 mM glutamine, 1 mM adenine and 100 mM hydrogen peroxide was flowed through a fused silica capillary with I.D. 100 µM. The sample was irradiated by a pulsed laser beam from a Compex Pro 102 KrF excimer laser at 248 nm wavelength (Coherent, Germany). The laser fluence at the intersection with the fused silica capillary flow cell was calculated to be ~13.1 mJ/mm2/pulse. Laser repetition rate was set at 20 Hz and the sample flow rate was set at 16 µL/min to ensure a 15% exclusion volume between irradiated volumes. The samples were immediately quenched in a microcentrifuge tube containing 25 µl of quench mixture (0.5 µg/µl H-Met-NH2 and 0.5 µg/µl catalase). The adenine hydroxyl radical dosimetry readings were measured at 265 nm in a Nanodrop (Thermo Scientific) to ensure all the samples were

445 exposed to equivalent amounts of hydroxyl radical [47]. Each experimental condition, DBL1-ID2
446 with or without placental chondroitin sulfate, was performed in triplicate for statistical analysis.
447 Placental CS was purified from human placenta as previously described[31].
448 A final concentration of 100 mM Tris, pH 8.0 containing 1 mM $CaCl_2$ and 5 mM DTT was
449 added to the FPOP samples and incubated at 95°C for 15 minutes to reduce and denature the
450 protein. The sample was immediately cooled on ice for 2 min. Chymotrypsin with 1:20 ratio
451 of chymotrypsin:protein was added and incubated at 25°C for 16 hr with rotation. The digestion
452 reaction was stopped by the addition of 0.5% trifluoroacetic acid. The samples were analyzed
453 on a Dionex Ultimate 3000 nano-LC system coupled to an Orbitrap Fusion Thermo Scientific
454 (San Jose, CA). Samples were first injected via autosampler onto a 300 µM id X5 mm PepMap
455 100, 5 µm (Thermo Scientific) C18 trapping cartridge, then back-eluted onto an Acclaim
456 PepMap 100 C18 nanocolumn (0.75 mm × 150 mm, 2 µm, Thermo Scientific). Peptides were
457 separated using gradient of solvent A (0.1% formic acid in water) and solvent B (0.1% formic
458 acid in acetonitrile) at a flow rate of 0.30 µL/min. The gradient consisted of 2 to 10% solvent
459 B over 3 min, increasing to 35% B over 25 min, ramped to 95 % B over 3 min, held for 3 min,
460 and then returned to 2% B over 2 min and held for 7 min. The spray voltage was set to 2,400
461 volts, and the temperature of the heated capillary was set to 300°C. The data were collected in
462 positive ion mode and full scans were collected in Orbitrap with resolution of 60,000 and MS2
463 scans were detected in ion trap. Full MS scans were acquired from m/z 250 to 2,000 followed
464 by eight subsequent MS2 CID scans on the top eight most abundant peptide ions. The peptides
465 were fragmented by collision-induced dissociation CID with an isolation width of 3 m/z units.
466 Peptides were identified using Byonic version v2.10.5 (Protein Metrics). The Byonic search
467 parameters included all possible major oxidation modi*fi*cations as variable modi*fi*cations and
468 the enzyme specificity was set to cleave the protein after tyrosine, phenylalanine, tryptophan,
469 and leucine. The peak intensities of the unoxidized peptides were used to search manually for
470 oxidized peptides in the LC-MS data using accurate mass measurement and relative retention
471 time shift compared to the unmodified peptide. The intensities of each unoxidized peptide and
472 their corresponding oxidation products observed in LC-MS were used to calculate the average
473 oxidation events per peptide in the sample as previously reported [48]. Briefly, peptide level
474 oxidation was calculated by adding the ion intensities of all the oxidized peptides multiplied
475 by the number of oxidation events required for the mass shift (e.g., one event for +16, two
476 events for +32) and then divided by the sum of the ion intensities of all unoxidized and oxidized
477 peptide masses as represented by equation 1.

$$P = [I(+16)oxidized\ X\ 1 + I\ (+32)oxidized\ X\ 2 + I(+48)oxidized\ X\ 3 + \ldots / [Iunoxidized + I(+16)oxidized + I(+32)oxidized + I(+48)oxidized \ldots]$$ (1).

where P denotes the oxidation events at the peptide level and I values are the peak intensities of oxidized and unoxidized peptides.

Cell cultures

A549 cells were cultured in DMEM medium. The media were acquired from Sigma Aldrich and supplemented with GlutaMaxTM, 10% fetal bovine serum and 1% penicillin-streptomycin. A549 cells were regularly detached using Trypsin-EDTA Solution (59417, Sigma-Aldrich) for passaging and all cells were sustained at 5% CO_2 and 37°C.

Flow cytometry

Protein binding to A549 was tested by flow cytometry. A549 cells were detached with StemPro Accutase™ Cell Dissociation Reagent (A1110501, Gibco). Cells were incubated with a two-fold dilution of V5-tagged WT or mutant rVAR2 (400-12.5nM) for 30 minutes at 4°C, followed by an incubation with 200x diluted Penta-His Alexa Fluor 488 Conjugate (35310, Qiagen). Samples were processed on an LSR II flow cytometer (BD) and results were analysed using FlowJo™ software (BD Life Sciences).

Sequence analysis

A total of 3,737 sequences annotated with VAR2CSA-specific DBL domains were extracted from the var database, varDB PF3K[49]. Of these, 1,388 spanned across the NTS-DBL6 domain region. To assess sequence diversity, amino acid alignments were constructed for each domain separately using all sequences spanning a VAR2CSA domain using Muscle and subsequently edited by hand[50]. From these alignments, an alignment spanning NTS-DBL6 of the VAR2CSA ectodomain was constructed from those sequences spanning NTS-DBL6. Sequence LOGOs were made using WebLogo 3[51].

Figure Legends

Fig. 1: Apo-structure of the ectodomain of VAR2CSA. a) SDS-PAGE non-reduced/reduced of full-length VAR2CSA (307kDa). **b)** Gel filtration profile of VAR2CSA in different buffers, top panel KCl, bottom panel NaCl. **c)** QCM biosensor interaction between CSPG decorin and VAR2CSA, top panel KCl (k_D 0.11nM), bottom panel NaCl (k_{D1} 1.4nM). Black curves represent recorded data, red curves fitted data. **d)** Overall structure of VAR2CSA. The separate domains are colored as shown in the overview below. The core region of the structure and minimal CS binding region (AA450-1025) are indicated. **e)** α-helix from ID3 (orange) AA1955-1985 interaction with C-terminal part of ID2 (blue). To the right the structure is turned 90° clockwise. **f)** Surface potential (kT/e) of VAR2CSA shown on apo-structure in a similar orientation as in panel d). The boxed region shows the positively charged binding groove in the minimal CS binding region.

Fig 2: Structural analysis of DBL2 and CS binding groove. a) Comparison of domains DBL2, DBL3 and DBL4. The red loop shows the "WIW-motif"-loop from each domain, the green α-helix shows HB1. In DBL2, an additional loop in HB1 introduces a kink in the α-helix. **b)** DBL2 domain shown as cartoon, the red loop shows S1 (HB4 and "WIW-motif"-loop), blue α-helices show S2 (HB3-HB5) and green α-helices show S3 (HB2-HB1). S1-S3 residue boundaries and colors are shown in legend below. The tip nitrogen atom of Lys/Arg residues in binding groove are shown as blue spheres and labelled according to residues number. **c)** Structural details of the interaction between "WIW-motif" and HB2/HB3 in DBL2. Interacting side chains are shown as sticks and are labelled with residue numbers. **d)** LOGO conservation analyses of "WIW-motif"-loop AA553-583 in *var2csa* variants. Residue number from VAR2CSA sequence shown below.

Fig 3: Hydroxyl radical protein footprinting and CSA binding analysis of DBL1-ID2 mutant. a) FPOP analysis of placental CS binding to wild-type DBL1-ID2. 13 chymotryptic peptides in VAR2CSA were found to exhibit significant protection from oxidation upon binding to placental CS (blue asterisks). Two peptides (red asterisks) exhibit significant exposure ($p < 0.05$, two-tailed Student's t-test). Peptides not detected as oxidized in either condition are not shown. Error bars represent one standard deviation from a triplicate measure. **b)** SDS-PAGE non-reduced/reduced of DBL1-ID2 wild-type (KKKWIWKK) and mutant (KKK**A**I**A**KK). **c)** T_m determination by nanoDSF melting curves of wild-type (T_m 67.7°C) and mutant (T_m 63.7°C) DBL1-ID2. Region between 40°C and 80°C is shown **d)** QCM biosensor

kinetic measurements of wild-type (top, k_{D1} 1.1nM) and mutant DBL1-ID2 (bottom, no binding) to decorin CSPG. **e)** Flow cytometry binding analyses of lung cancer cell line A549 to wild-type and mutant DBL1-ID2.

Fig 4. Docking model of VAR2CSA in complex with a CSA 20-mer. a) The docking structure of VAR2CSA in complex with CSA 20-mer. The electrostatic potential surfaces of VAR2CSA with the minimal binding region and binding groove, which is highlighted by a black box. The surface potential (kT/e) color bar is shown below. The CSA molecule is represented by sticks in magenta. The DBL5/6 domains are shown in white cartoon. The right panel shows the structure turned 90° clockwise. **b)** The same model is shown as transparent surface in white and secondary elements in green. The CSA molecule is shown as magenta sticks. **c)** Conformational flexibility of CSA ligand in the binding groove of VAR2CSA. Boxed area shows the core part of the CSA chain with relatively low RMSF values, corresponding to roughly seven disaccharide units. **d)** Binding groove of VAR2CSA with an ensemble of CSA 20-mer with 70 conformers sampled in a 70ns MD simulation with an interval of 1ns.

Fig S1: **Resolution and structural flexibility of determined the cryoEM structure a)** CryoEM map of apo VAR2CSA, colored by local resolution at two different level. **b)** Cryo-EM map of the VAR2CSA in presence of plCS, colored by local resolution at two different levels. **c)** The VAR2CSA-plCS map with DBL1/5/6 part partially transparent and show at a different contour level compared to the core structure. Compared to the apo structure, the core part is conformationally stable, while DBL1/5/6 are flexible. **d)** Two independent maps/models calculated using separate sub-fractions of particles. Alignment using the core structure shows displaced are DBL5/6.

Fig S2: Quality of the cryo-EM density in selected functionally important regions a) ID3-helix (residue AA1955-1985). **b)** C-terminal part of ID2 which is interacting with ID3 α-helix (residue AA1140-1181). **c)** "WIW-motif"-loop in DBL2 (residue AA556-590). **d)** HB1 C-terminal half of α-helix in DBL2 involved in CSA binding (residue AA758-783)

Fig S3: **a)** LOGO conservation of ID3 region (residues AA1955-1985) in VAR2CSA subvariants. **b)** LOGO conservation of ID2 region interacting with ID3 (residues AA1139-1181).

Fig S4: The cryo-EM structure of the DBL3/4 domains compared to the equivalent crystal structure (PDB-ID 4P1T). The crystal structure is colored in gray and the cryo-EM structure of DBL3 in purple and DBL4 in yellow.

Fig S5: Workflow of the cryo-EM structure determination of the VAR2CSA apo structure. **a)** Representative micrograph. **b)** Representative 2D class averages with a box size of 36.6nm. **c)** Brief flowchart of the data processing. The Methods section and Table S1 contains further details. **d)** Gold standard Fourier shell correlation (FSC) curve of the map. **e)** FSC curve of build model vs map. **f)** The final map colored by resolution.

Fig S6: Workflow of the cryo-EM structure determination of the VAR2CSA plCS structure. **a)** Representative micrograph. **b)** Representative 2D class averages with a box size of 36.6nm. **c)** Brief flowchart of the data processing. The Methods section and Table S1 contains further details. **d)** Gold standard Fourier shell correlation (FSC) curve of the map. **e)** FSC curve of build model vs map. **f)** The final map colored by resolution.

Fig S7: LOGO plot of VAR2CSA ectodomain. Domain and homology block borders indicated above LOGO along with data from FPOP analysis of CS-protected/exposed peptides in the DBL1-ID2 protein and CS-interacting residues predicted by Molecular Docking (MD). Small LOGO alignment below in the middle shows atypical HBs in VAR2CSA DBL2. Graphic HB LOGO legend modified from[10].

Fig S8: FPOP-MS peptides mapped on the surface of the VAR2CSA apo structure with docked CSA 20-mer shown as purple sticks. Cyan regions indicate an up to 2x reduction in oxidation and blue regions show more than 2x reduction. Red regions designate an increase in surface exposure when plCS is bound.

Fig S9: Domain flexibility revealed by MD simulations. The RMS fluctuation of VAR2CSA at a residue level revealed by the MD simulation of the VAR2CSA-CS complex. There are significant motions observed in DBL1/5/6 domains and also the loop regions as illustrated by the grey transparent tubes.

Table S1: Cryo-EM data collection data processing and model building statistics.

Acknowledgement

The authors would like to express their deep gratitude to Benjamin Jacobsen, Andreas Frederiksen, Elham Alijazaeri and Susanne Lücking Nielsen, for their excellent technical assistance. CPR for use of nanoDSF instrument. S.K.M. and J.S.S. acknowledge support for this work from the National Institute of General Medical Sciences, National Institutes of Health (P41GM103390 and R01GM127267). K.L-L and Y.W. acknowledge access to computational resources from the Danish National Supercomputer for Life Sciences (Computerome) and the ROBUST Resource for Biomolecular Simulations (supported by the Novo Nordisk Foundation), and financial support from the Lundbeck Foundation BRAINSTRUC initiative. AS, TL have received support from the Danish Research Councils, ERC and Carlsberg Foundation. MØA from Innovation Foundation Denmark. PG was financed by the Lundbeck and the Knut and Alice Wallenberg Foundations as well as by The Independent Research Fund Denmark and the Swedish Research Council.

Author contributions:

Protein expression and purification, flow cytometry, biosensor analyses was designed, performed and analyzed by RD, TG, EEV, SC, MØA, MAN, AS. FPOP analyses was performed and analyzed by SKM, JSS, TMC and DRS. CryoEM and simulations were executed and analyzed by KW, PG, KLL, YW. Placental CS was purified by CS. Sequence analyses were achieved by TL and TGT. Study was designed by KW, AS, RD. Manuscript was written by AS, TGT, PG, KW, TL, RD. All authors contributed to discussing the data and proof-reading the manuscript.

Competing interest:

AS, TGT, MAN, TMC are listed as co-inventors on a patent family covering the use of VAR2CSA to target and diagnose cancer. AS and TGT are listed as co-inventor on a patent on using VAR2CSA as a prophylactic malaria vaccine during pregnancy. J.S.S. has financial interest in an early-stage company commercializing technologies for protein higher order structure analysis. The other authors have no financial conflicts of interest.

References

Reference List

1. Baruch,D.I. *et al.* Cloning the *Plasmodium falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell* **82**, 77-87 (1995).

2. Kraemer,S.M. & Smith,J.D. A family affair: var genes, PfEMP1 binding, and malaria disease. *Curr. Opin. Microbiol.* **9**, 374-380 (2006).

3. Salanti,A. *et al.* Selective upregulation of a single distinctly structured var gene in chondroitin sulphate A-adhering *Plasmodium falciparum* involved in pregnancy-associated malaria. *Mol. Microbiol.* **49**, 179-191 (2003).

4. Dzikowski,R., Frank,M., & Deitsch,K. Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production. *PLoS. Pathog.* **2**, e22 (2006).

5. Salanti,A. *et al.* Evidence for the involvement of VAR2CSA in pregnancy-associated malaria. *J. Exp. Med.* **200**, 1197-1203 (2004).

6. Brabin,B.J. An analysis of malaria in pregnancy in Africa. *Bull. World Health Organ* **61**, 1005-1016 (1983).

7. Fried,M. & Duffy,P.E. Adherence of *Plasmodium falciparum* to chondroitin sulfate A in the human placenta. *Science* **272**, 1502-1504 (1996).

8. Mordmuller,B. *et al.* First-in-human, randomized, double-blind clinical trial of differentially adjuvanted PAMVAC, a vaccine candidate to prevent pregnancy-associated malaria. *Clin. Infect. Dis.*(2019).

9. Sirima,S.B. *et al.* PRIMVAC vaccine adjuvanted with Alhydrogel or GLA-SE to prevent placental malaria: a first-in-human, randomised, double-blind, placebo-controlled study. *Lancet Infect. Dis.* **20**, 585-597 (2020).

10. Rask,T.S., Hansen,D.A., Theander,T.G., Gorm,P.A., & Lavstsen,T. Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes--divide and conquer. *PLoS. Comput. Biol.* **6**, (2010).

11. Achur,R.N., Valiyaveettil,M., Alkhalil,A., Ockenhouse,C.F., & Gowda,D.C. Characterization of proteoglycans of human placenta and identification of unique chondroitin sulfate proteoglycans of the intervillous spaces that mediate the adherence of *Plasmodium falciparum*-infected erythrocytes to the placenta. *J. Biol. Chem.* **275**, 40344-40356 (2000).

12. Ayres,P.M. *et al.* Placental Sequestration of Plasmodium falciparum Malaria Parasites Is Mediated by the Interaction Between VAR2CSA and Chondroitin Sulfate A on Syndecan-1. *PLoS. Pathog.* **12**, e1005831 (2016).

13. Salanti,A. *et al.* Targeting Human Cancer by a Glycosaminoglycan Binding Malaria Protein. *Cancer Cell* **28**, 500-514 (2015).

14. Alkhalil,A., Achur,R.N., Valiyaveettil,M., Ockenhouse,C.F., & Gowda,D.C. Structural requirements for the adherence of *Plasmodium falciparum*-infected erythrocytes to chondroitin sulfate proteoglycans of human placenta. *J. Biol. Chem.* **275**, 40357-40364 (2000).

15. Seiler,R. *et al.* An Oncofetal Glycosaminoglycan Modification Provides Therapeutic Access to Cisplatin-resistant Bladder Cancer. *Eur. Urol.* **72**, 142-150 (2017).

16. Agerbaek,M.O. *et al.* The VAR2CSA malaria protein efficiently retrieves circulating tumor cells in an EpCAM-independent manner. *Nat. Commun.* **9**, 3279 (2018).

17. Clausen,T.M. *et al.* Structural and Functional Insight into How the Plasmodium falciparum VAR2CSA Protein Mediates Binding to Chondroitin Sulfate A in Placental Malaria. *J. Biol. Chem.* **287**, 23332-23345 (2012).

18. Bewley,M.C., Gautam,L., Jagadeeshaprasad,M.G., Gowda,D.C., & Flanagan,J.M. Molecular architecture and domain arrangement of the placental malaria protein VAR2CSA suggests a model for carbohydrate binding. *J. Biol. Chem.*(2020).

19. Higgins,M.K. The structure of a chondroitin sulfate-binding domain important in placental malaria. *J. Biol. Chem.* **283**, 21842-21846 (2008).

20. Clausen,T.M. *et al.* Structural and functional insight into how the Plasmodium falciparum VAR2CSA protein mediates binding to chondroitin sulfate A in placental malaria. *J. Biol. Chem.* **287**, 23332-23345 (2012).

21. Gangnard,S. *et al.* Structure of the DBL3X-DBL4Îµ region of the VAR2CSA placental malaria vaccine candidate: insight into DBL domain interactions. *Sci. Rep.* **5**, 14868 (2015).

22. Khunrae,P., Philip,J.M., Bull,D.R., & Higgins,M.K. Structural comparison of two CSPG-binding DBL domains from the VAR2CSA protein important in malaria during pregnancy. *J. Mol. Biol.* **393**, 202-213 (2009).

23. Magistrado,P.A. *et al.* High efficacy of anti DBL4varepsilon-VAR2CSA antibodies in inhibition of CSA-binding Plasmodium falciparum-infected erythrocytes from pregnant women. *Vaccine* **29**, 437-443 (2011).

24. Lennartz,F. *et al.* Structure-Guided Identification of a Family of Dual Receptor-Binding PfEMP1 that Is Associated with Cerebral Malaria. *Cell Host. Microbe* **21**, 403-414 (2017).

25. Lennartz,F., Smith,C., Craig,A.G., & Higgins,M.K. Structural insights into diverse modes of ICAM-1 binding by Plasmodium falciparum-infected erythrocytes. *Proc. Natl. Acad. Sci. U. S. A* **116**, 20124-20134 (2019).

26. Hsieh,F.L. *et al.* The structural basis for CD36 binding by the malaria parasite. *Nat. Commun.* **7**, 12837 (2016).

27. Lau,C.K. *et al.* Structural conservation despite huge sequence diversity allows EPCR binding by the PfEMP1 family implicated in severe childhood malaria. *Cell Host. Microbe* **17**, 118-129 (2015).

28. Adams,J.H. *et al.* A family of erythrocyte binding proteins of malaria parasites. *Proc. Natl. Acad. Sci. U. S. A* **89**, 7085-7089 (1992).

29. Khunrae,P. *et al.* Full-length recombinant Plasmodium falciparum VAR2CSA binds specifically to CSPG and induces potent parasite adhesion-blocking antibodies. *J. Mol. Biol.* **397**, 826-834 (2010).

30. Nielsen,M.A. *et al.* The Influence of Sub-Unit Composition and Expression System on the Functional Antibody Response in the Development of a VAR2CSA Based Plasmodium falciparum Placental Malaria Vaccine. *PLoS. One.* **10**, e0135406 (2015).

31. Clausen,T.M. *et al.* Real-time and label free determination of ligand binding-kinetics to primary cancer tissue specimens; a novel tool for the assessment of biomarker targeting. *Sens. Biosensing. Res.* **9**, 23-30 (2016).

32. Punjani,A., Rubinstein,J.L., Fleet,D.J., & Brubaker,M.A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290-296 (2017).

33. Rohou,A. & Grigorieff,N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216-221 (2015).

34. Gong,X. *et al.* Structural Insights into the Niemann-Pick C1 (NPC1)-Mediated Cholesterol Transfer and Ebola Infection. *Cell* **165**, 1467-1478 (2016).

35. Gangnard,S. *et al.* Structural and immunological correlations between the variable blocks of the VAR2CSA domain DBL6Îµ from two Plasmodium falciparum parasite lines. *J. Mol. Biol.* **425**, 1697-1711 (2013).

36. Emsley,P., Lohkamp,B., Scott,W.G., & Cowtan,K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486-501 (2010).

37. Afonine,P.V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D. Struct. Biol.* **74**, 531-544 (2018).

38. Chen,V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 12-21 (2010).

39. Webb,B. & Sali,A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinformatics.* **54**, 5 (2016).

40. Trabuco,L.G., Villa,E., Mitra,K., Frank,J., & Schulten,K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure.* **16**, 673-683 (2008).

41. Páll,S. *et al.* Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J. Chem. Phys.* **153**, 134110 (2020).

42. Huang,J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71-73 (2017).

43. van Zundert,G.C.P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **428**, 720-725 (2016).

44. Lee,J. *et al.* CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory. Comput.* **12**, 405-413 (2016).

45. Hopkins,C.W., Le,G.S., Walker,R.C., & Roitberg,A.E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory. Comput.* **11**, 1864-1874 (2015).

46. Olson,L.J. *et al.* Allosteric regulation of lysosomal enzyme recognition by the cation-independent mannose 6-phosphate receptor. *Commun. Biol.* **3**, 498 (2020).

47. Xie,B. & Sharp,J.S. Hydroxyl Radical Dosimetry for High Flux Hydroxyl Radical Protein Footprinting Applications Using a Simple Optical Detection Method. *Anal. Chem.* **87**, 10719-10723 (2015).

48. Misra,S.K., Orlando,R., Weinberger,S.R., & Sharp,J.S. Compensated Hydroxyl Radical Protein Footprinting Measures Buffer and Excipient Effects on Conformation and Aggregation in an Adalimumab Biosimilar. *AAPS. J.* **21**, 87 (2019).

49. Otto,T.D. *et al.* Evolutionary analysis of the most polymorphic gene family in falciparum malaria. *Wellcome. Open. Res.* **4**, 193 (2019).

50. Edgar,R.C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC. Bioinformatics.* **5**, 113 (2004).

51. Crooks,G.E., Hon,G., Chandonia,J.M., & Brenner,S.E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188-1190 (2004).

Figures

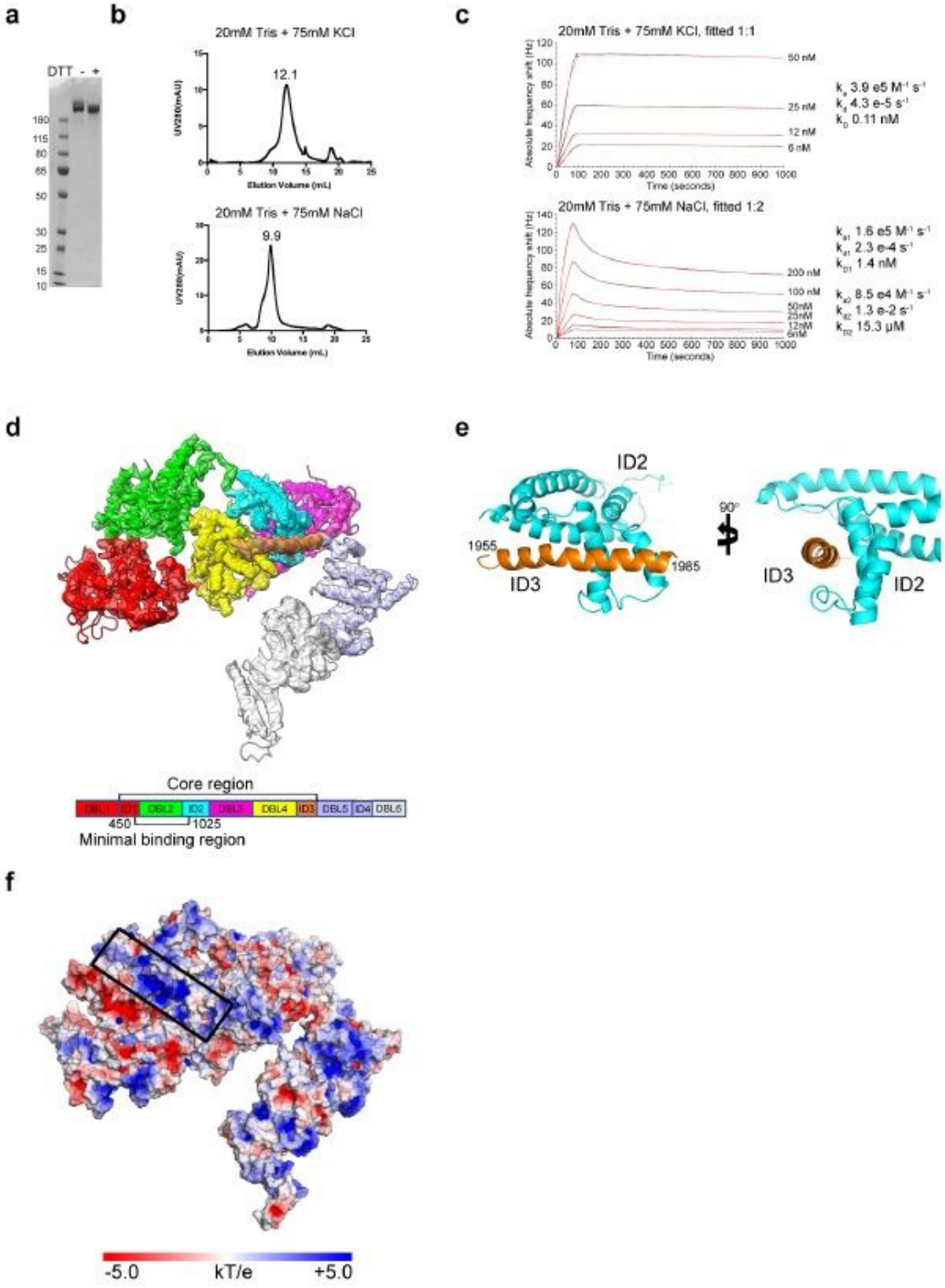

Figure 1

Apo-structure of the ectodomain of VAR2CSA. a) SDS-PAGE non-reduced/reduced of full-length VAR2CSA (307kDa). b) Gel filtration profile of VAR2CSA in different buffers, top panel KCl, bottom panel NaCl. c) QCM biosensor interaction between CSPG decorin and VAR2CSA, top panel KCl (kD 0.11nM), bottom

panel NaCl (kD1 1.4nM). Black curves represent recorded data, red curves fitted data. d) Overall structure of VAR2CSA. The separate domains are colored as shown in the overview below. The core region of the structure and minimal CS binding region (AA450-1025) are indicated. e) α-helix from ID3 (orange) AA1955-1985 interaction with C-terminal part of ID2 (blue). To the right the structure is turned 90° clockwise. f) Surface potential (kT/e) of VAR2CSA shown on apo-structure in a similar orientation as in panel d). The boxed region shows the positively charged binding groove in the minimal CS binding region.

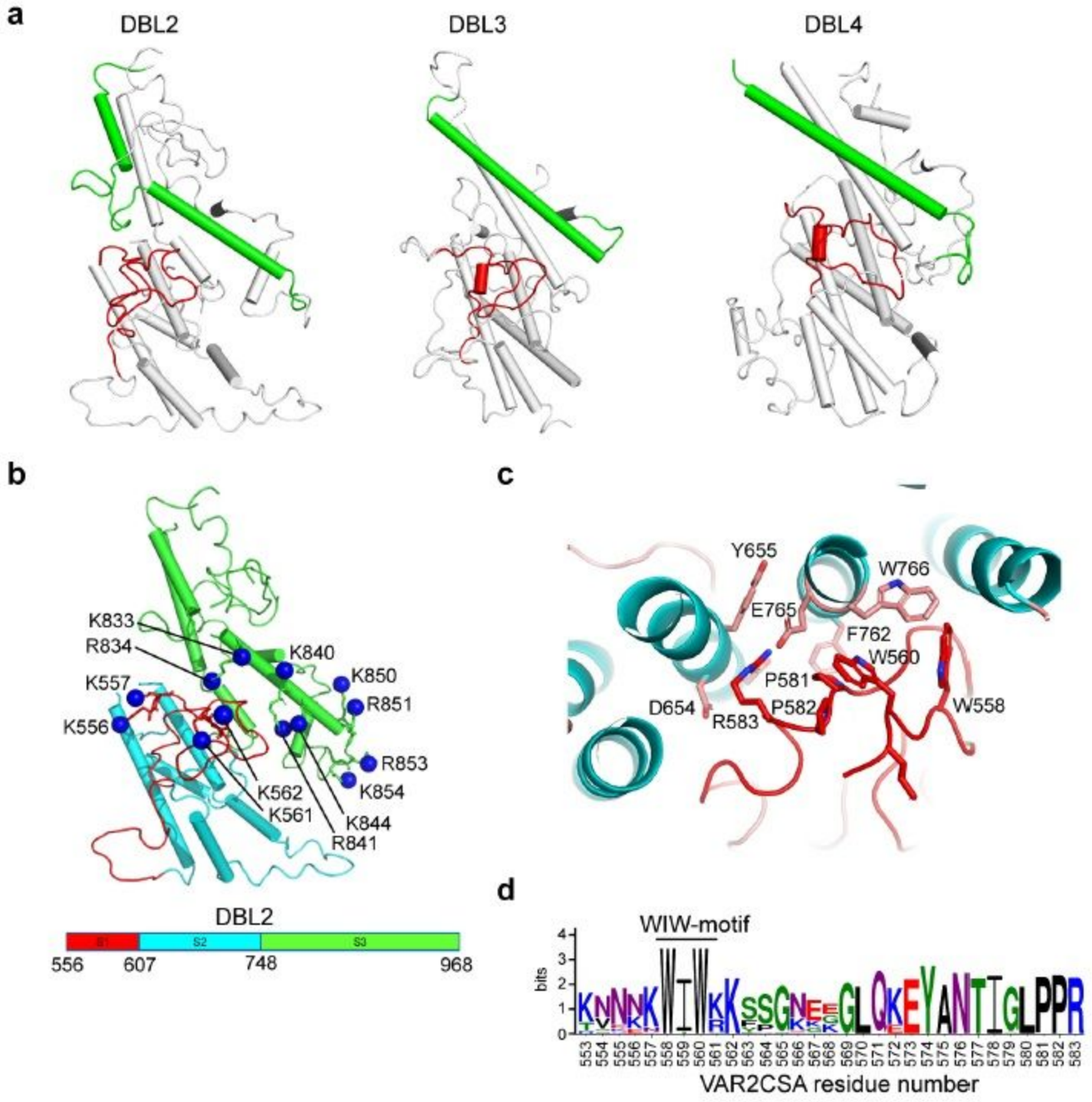

Figure 2

Structural analysis of DBL2 and CS binding a) Comparison of domains DBL2, DBL3 and The red loop shows the “WIW-motif”-loop from each domain, the green α-helix shows In DBL2, an additional loop in HB1 introduces a kink in the α-helix. b) DBL2 domain shown as cartoon, the red loop shows S1 (HB4 and “WIW-motif”-loop), blue α-helices show S2 (HB3-HB5) and green α-helices show S3 (HB2-HB1). S1-S3 residue boundaries and colors are shown in legend The tip nitrogen atom of Lys/Arg residues in binding groove are shown as blue spheres and labelled according to residues c) Structural details of the interaction between “WIW-motif” and HB2/HB3 in Interacting side chains are shown as sticks and are labelled with residue numbers. d) LOGO conservation analyses of “WIW-motif”-loop AA553-583 in var2csa Residue number from VAR2CSA sequence shown below.

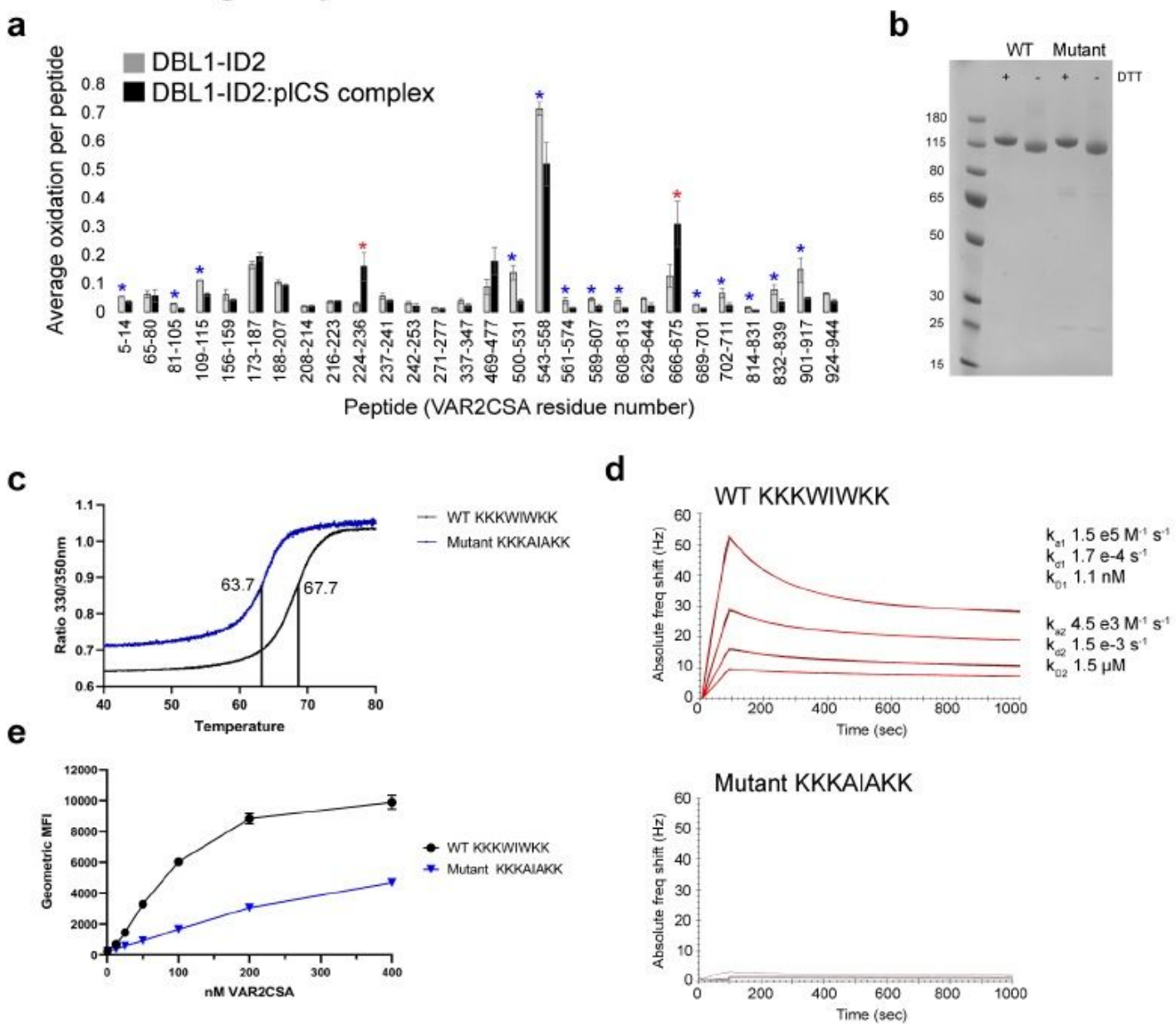

Figure 3

Hydroxyl radical protein footprinting and CSA binding analysis of DBL1-ID2 a) FPOP analysis of placental CS binding to wild-type DBL1-ID2. 13 chymotryptic peptides in VAR2CSA were found to exhibit significant protection from oxidation upon binding to placental CS (blue asterisks). Two peptides (red asterisks) exhibit significant exposure (p < 05, two-tailed Student's t-test). Peptides not detected as oxidized in either condition are not Error bars represent one standard deviation from a triplicate measure. b) SDS-PAGE non-reduced/reduced of DBL1-ID2 wild-type (KKKWIWKK) and mutant (KKKAIAKK). c) Tm determination by nanoDSF melting curves of wild-type (Tm 7°C) and mutant (Tm 7°C) DBL1-ID2. Region between 40°C and 80°C is shown d) QCM biosensor kinetic measurements of wild-type (top, kD1 1nM) and mutant DBL1-ID2 (bottom, no binding) to decorin e) Flow cytometry binding analyses of lung cancer cell line A549 to wild-type and mutant DBL1-ID2.

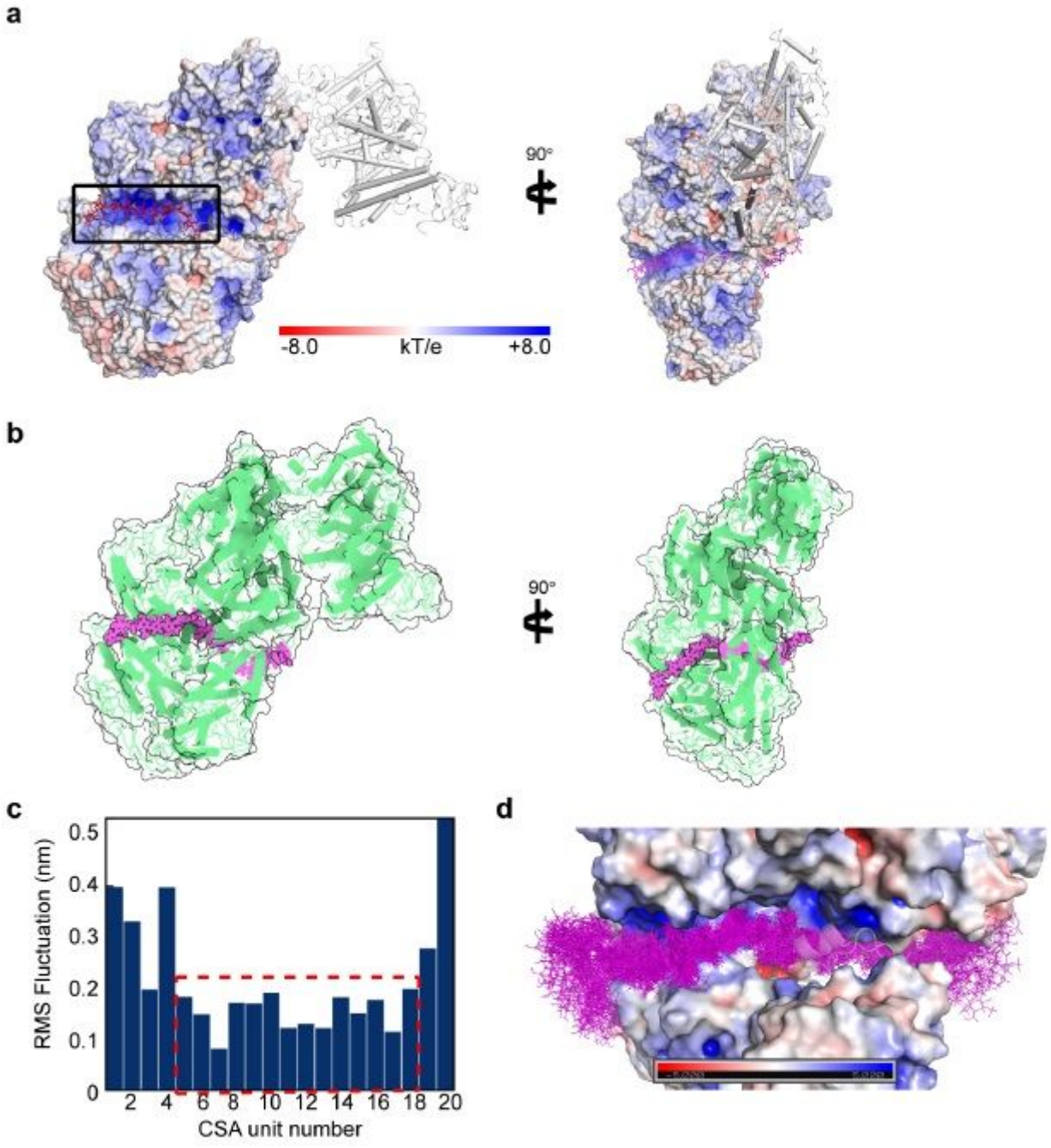

Figure 4

Docking model of VAR2CSA in complex with a CSA 20-mer. a) The docking structure of VAR2CSA in complex with CSA 20-mer. The electrostatic potential surfaces of VAR2CSA with the minimal binding region and binding groove, which is highlighted by a black The surface potential (kT/e) color bar is shown below. The CSA molecule is represented by sticks in magenta. The DBL5/6 domains are shown in white cartoon. The right panel shows the structure turned 90° b) The same model is shown as transparent surface in white and secondary elements in The CSA molecule is shown as magenta c) Conformational flexibility of CSA ligand in the binding groove of VAR2CSA. Boxed area shows the core part of the CSA chain with relatively low RMSF values, corresponding to roughly seven disaccharide units. d) Binding groove of VAR2CSA with an ensemble of CSA 20-mer with 70 conformers sampled in a 70ns MD simulation with an interval of 1ns.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- finalTable1.pdf
- FigureS1.pdf
- FigureS2.pdf
- FigureS3.pdf
- FigureS4.pdf
- FigureS5.pdf
- FigureS6.pdf
- FigureS7.pdf
- FigureS8.pdf
- FigureS9.pdf