Unexpected fold in the circumsporozoite protein target of malaria vaccines

Michael B. Doud, Adem C. Koksal, Li-Zhi Mi, Gaojie Song, Chafen Lu, and Timothy A. Springer[1]

Immune Disease Institute, Children's Hospital Boston and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115

Contributed by Timothy A. Springer, April 4, 2012 (sent for review March 19, 2012)

Circumsporozoite (CS) protein is the major surface component of *Plasmodium falciparum* sporozoites and is essential for host cell invasion. A vaccine containing tandem repeats, region III, and thrombospondin type-I repeat (TSR) of CS is efficacious in phase III trials but gives only a 35% reduction in severe malaria in the first year postimmunization. We solved crystal structures showing that region III and TSR fold into a single unit, an "αTSR" domain. The αTSR domain possesses a hydrophobic pocket and core, missing in TSR domains. CS binds heparin, but αTSR does not. Interestingly, polymorphic T-cell epitopes map to specialized αTSR regions. The N and C termini are unexpectedly close, providing clues for sporozoite sheath organization. Elucidation of a unique structure of a domain within CS enables rational design of next-generation subunit vaccines and functional and medicinal chemical investigation of the conserved hydrophobic pocket.

During a blood meal, *Plasmodium* sporozoites are transmitted from an infected mosquito to a vertebrate host. Sporozoites are uniformly covered with circumsporozoite protein (CS), a highly expressed and immunodominant surface protein antigen necessary for sporozoite development and host cell targeting (1, 2). Gliding motility enables sporozoites to move directly through host cells to reach the vasculature and later to penetrate the liver. During gliding, CS moves from the anterior to the posterior end of the sporozoite, where it is shed and deposited in a trail (3). CS is composed of an N-terminal domain, a conserved pentapeptide protease cleavage site termed region I, a repeat region, a short conserved sequence termed region III, and a C-terminal region with sequence homology to the thrombospondin type-1 repeat superfamily (TSR) (Fig. 1*A*). The C terminus of the CS TSR contains a glycosylphosphatidylinositol (GPI) anchor attachment site responsible for the association of CS with the sporozoite plasma membrane (4). The RTS,S malaria vaccine currently in phase III trials is composed of a C-terminal CS fragment comprising a portion of the repeats, region III, and TSR (Fig. 1*A*) fused to hepatitis B surface antigen (5, 6). In the absence of a 3D structure, there is currently no structural basis for vaccine improvement.

Results

Region III and TSR are distinguished from the more N-terminal sequence containing the linker to the repeat region by their high sequence conservation across *Plasmodium* species (Fig. 1*B*). Therefore, we chose a CS fragment containing both region III and TSR as the unit for crystallization (marked above Fig. 1*B*). This fragment contains at its C terminus the small residue that is predicted on the sporozoite surface to be linked to the GPI anchor (Fig. 1*B*). The fragment was expressed with a signal peptide sequence and a C-terminal His tag as a protein secreted into the culture medium from the yeast *Pichia pastoris* or from mammalian cells. The *Pichia* protein additionally contained a 4-residue N-terminal Tyr-Val-Glu-Phe (YVEF) sequence derived from a restriction enzyme cloning site. Proteins were purified by His-tag affinity and gel filtration, with final yields of ~15 and 2 mg/L of culture medium from *Pichia* and HEK293T cells, respectively. Protein from *Pichia* crystallized in two different forms, and their structures were solved using anomalous dispersion to 1.7 and 2.04 Å (Fig. 2 and Table 1).

The CS structures reveal a unique "αTSR" domain that is related to TSR domains yet has unique features (Figs. 1 *C* and *D* and 2). Region III forms an integral part of the fold as a unique amphipathic α-helix running orthogonally to the classic three-stranded TSR homology region. This α1-helix, together with a "CS flap" (Fig. 1 *C* and *D*), decorates one face of the TSR β-sheet (Fig. 2*A*). By packing against the β-sheet, the region III α-helix gives the αTSR domain a hydrophobic core lacking in TSR domains. A short linker containing a conserved hydrophobic residue connects the helix to TSR strand 1 (Fig. 1*B*).

TSR domains contain three strands; strand 1 is not a β-strand, yet strands 1, 2, and 3 are positioned alongside one another as if in an antiparallel β-sheet (Figs. 1*D* and 2*C*) (7–10). TSR domains are termed a layered fold because they are stabilized by eight layers of stacking residues that form one face of the sheet (Fig. 2 *B* and *C*). In the αTSR domain, this stack is shortened to four layers, and they form the side of the sheet opposite the hydrophobic core (Fig. 2 *A* and *B*). Thus, the (Wxx)WxxW motif in strand 1 of TSR domains (Fig. 1*C*) is truncated in the CS αTSR domain to the third W of this motif, Trp-331. However, the layers of basic residues on either side of this Trp seen in other TSRs are preserved (Fig. 2*B*); the Trp-331 sidechain in strand 1 is sandwiched between Arg-345 in strand 2 and Lys-367 in strand 3, which form π-cation bonds to knit the three strands together. Two conserved disulfides further connect strands 1 and 3, and connect the turn between strands 1 and 2 to the penultimate C-terminal residue (Figs. 1 *C* and *D* and 2*A*). Strands 2 and 3 in TSR domains are halved in length in αTSR by the CS flap; however, the portion forming the β-sheet is conserved (Fig. 1 *C* and *D*).

Surprisingly, αTSR contains a hydrophobic pocket that is contiguous with the hydrophobic core. The core is formed by α1-helix residues Ile-316, Tyr-319, and Ile-323 and strand 2 residues Ile-342 and Val-344 on the face opposite that bearing layer residues (Fig. 2*D*). It is extended on one side by residues Pro-311 and Ile-368, which interact with one another and are invariant or conserved as hydrophobic, respectively. The other side of the hydrophobic core extends into an exposed hydrophobic pocket formed by the α1-helix residue Leu-320, loop residue Leu-327, flap residues Leu-358 and Tyr-360, and β3-strand residue Ile-364 (Fig. 2*E*). Although of unknown function, the importance of this hydrophobic pocket is emphasized by the invariantly hydrophobic character of the residues forming its surface among species of *Plasmodium* parasites that infect a range of hosts including humans, nonhuman primates, rodents, and birds (gold circles below alignment; Fig. 1*B*). On the opposite side of the β-sheet, an invariant salt bridge between Lys-353 and Asp-363 fastens the

Author contributions: M.B.D., A.C.K., L.-Z.M., G.S., C.L., and T.A.S. designed research; M.B.D., A.C.K., L.-Z.M., G.S., C.L., and T.A.S. performed research; M.B.D., A.C.K., L.-Z.M., G.S., C.L., and T.A.S. analyzed data; and M.B.D., A.C.K., and T.A.S. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The atomic coordinates reported in this paper have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 3VDJ, 3VDK, and 3VDL).

[1]To whom correspondence should be addressed. E-mail: springer@idi.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1205737109/-/DCSupplemental.

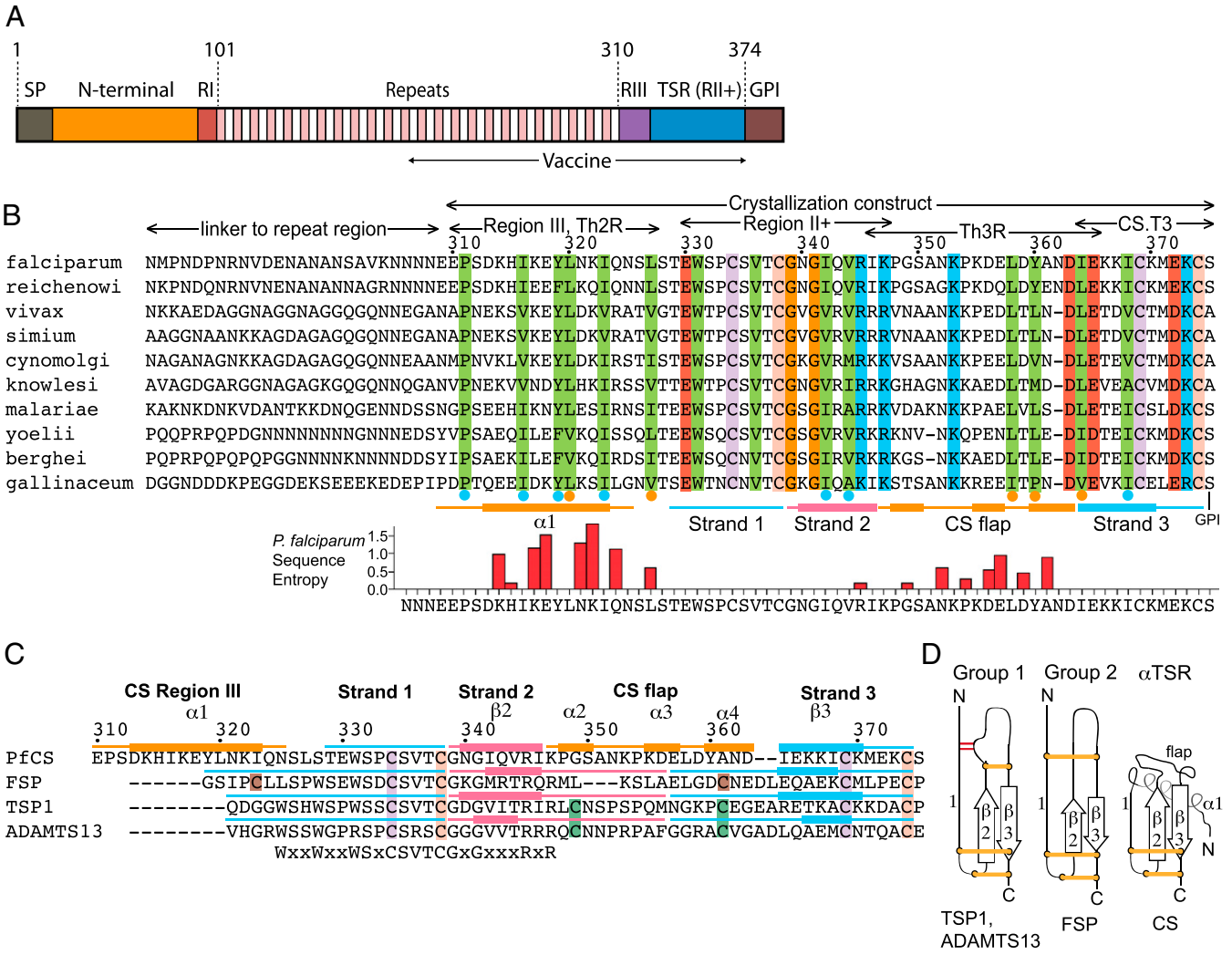

Fig. 1. Sequence alignments and fold schematics. (*A*) Gene architecture of *P. falciparum* circumsporozoite protein. The portion of CS fused to hepatitis B surface antigen in the RTS,S vaccine is indicated by "vaccine." (*B*) C-terminal portion of CS from different species. Conserved residues are colored according to chemical type. Disulfide-linked cysteines are colored according to connectivity. Dots below green (hydrophobic) columns mark residues in the hydrophobic core (blue) or exposed hydrophobic pocket (orange). Secondary structure is designated below the alignment by thick lines (helices and β-strands) colored by TSR strand direction (cyan and pink are antiparallel strand directions). Shannon sequence entropy was calculated from 42 distinct *P. falciparum* isolates from Freetown, Sierra Leone (27) using the Protein Variability Server (46). (*C*) Alignment with TSR domains from F-spondin (FSP) (9), thrombospondin-1 (TSP1) (7), and ADAMTS13 (10). Unique portions of PfCS αTSR are overlined in orange. Only sequences overlined in identical colors are structurally equivalent. Segments forming sheets or helices are shown with thickened lines. Disulfide connectivity is color coded. *P. falciparum* CS (3D7 strain) mature numbering is shown above the sequences. TSR consensus residues are shown below the alignment. (*D*) Diversity of TSR domain architectures, modified from ref. 7.

flap at the apex of the domain where the aliphatic portion of the Lys sidechain caps the hydrophobic core (Fig. 2*A*).

CS binds heparan sulfate proteoglycans (HSPGs) on hepatocytes (11–13). The highly sulfated HSPGs of hepatocytes are proposed to mediate targeting of sporozoites to the liver and trigger sporozoite cell invasion (14, 15). Heparin and HSPG binding sites have been identified in the CS ectodomain (11, 16) and in conserved N-terminal region I and αTSR region II$^+$ peptides (12, 13, 17, 18). The two basic residues in region II$^+$, Arg-345, and Lys-347, are neither highly exposed nor present in a basic patch, and Arg-345 in the β2-strand interacts with Trp-331 in the π-cation bond described above and is also neutralized by a salt bridge to Glu-365 at the adjacent position in the β3-strand (Fig. 2 *A* and *B*). The complete *Plasmodium falciparum* CS ectodomain bound to heparin-sepharose under stringent conditions, eluting in 0.35 M NaCl; however, the αTSR domain did not bind, even in absence of salt (Fig. 2*H*). These results are consistent with mapping of HSPG binding to the conserved basic residues clustered around region I (16, 17) (Fig. 1*A*). In vertebrate hosts, cleavage in region I unmasks the αTSR domain and is thought to permit it to play a role in liver cell invasion (19).

As sometimes seen with short tags that are appended to proteins used in crystallization, the N-terminal YVEF peptide encoded by a restriction enzyme site in the *Pichia* construct played a role in forming the crystallization lattice. Essentially, this tetrapeptide acted as a hydrophobic probe for the hydrophobic pocket of the αTSR domain. Similar trimers were formed in each of the two independent crystal forms, in which the YVEF peptide from one monomer bound to the hydrophobic pocket of a neighboring monomer around a threefold symmetry or pseudosymmetry axis (Fig. S1). The resulting trimeric, triangular prisms are provocatively similar in shape to that of *Plasmodium vivax* p25, which is proposed to tile the surface of the ookinete mosquito stage (20); however, we have no evidence that the trimeric form seen in crystals (Fig. S2) is physiologically relevant to packing on the sporozoite surface. Thus, most of the trimeric interface

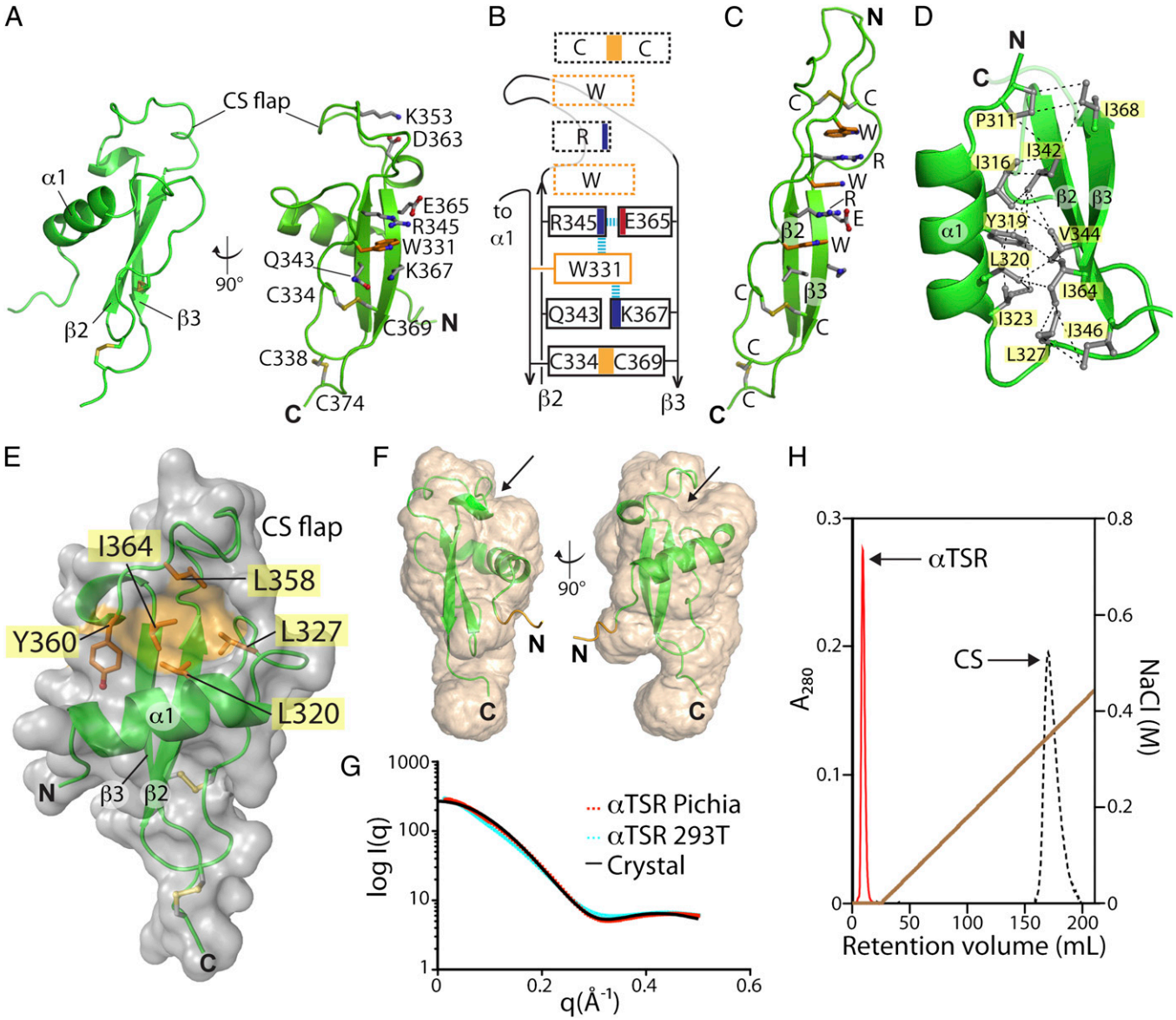

Fig. 2. Structure of αTSR. (*A*) CS αTSR showing layer residues and conserved flap fastening residues. (*B*) Layers of the TSR fold. Dashed boxes indicate layers lost in the αTSR. Dashed blue lines indicate π-cation and salt bridge interactions retained in αTSR, adapted from ref. 7. (*C*) Thrombospondin-1 TSR domain 1 (7) superimposed on the αTSR domain in *A*, *Right* and separated horizontally. Trp layers are colored orange and stacking layers are labeled by residue type. (*D*) Hydrophobic core of the αTSR domain. Conserved hydrophobic sidechains (with the exception of I346, in which many species have arginine or lysine sidechains) and their interactions are shown. A portion of the structure is omitted for clarity. (*E*) Transparent surface representation of αTSR with conserved hydrophobic pocket residues and their surface representation shaded orange. (*F*) Fitting of the crystal structure of αTSR to the final SWAXS envelope, shown as a transparent surface. The αTSR crystal structure is shown in cartoon in green, except the N-terminal YVEF is in orange. The YVEF sequence may be flexible in solution, because it is not included in the SWAXS envelope. Arrows indicate an observed indent on the SWAXS model corresponding to the approximate location of the hydrophobic pocket in the crystal structure. (*G*) Scattering curve calculated from the crystal structure and excluding the YVEF sequence (black) is overlaid to the raw experimental SWAXS scattering data for *Pichia* αTSR (red) and HEK293T αTSR (cyan). (*H*) Heparin-sepharose affinity assay of *Pichia* αTSR (red) and CS (black). Each construct was passed over a HiTrap heparin-sepharose column at a flow rate of 0.2 mL/min in 10 mM Tris pH 7.4 buffer and washed with a sodium chloride gradient (brown) over 40 column volumes.

involves burial of the tetrapeptide in the hydrophobic pocket, with only minor contributions from other portions of αTSR monomers. We did not obtain diffraction-quality crystals of the αTSR domain expressed in mammalian cells or expressed in *Pichia* after replacing the YVEF peptide with the native CS tetra- or non-apeptide N-terminal sequences.

αTSR is highly soluble and analytical ultracentrifugation and small angle X-ray scattering showed it to be monomeric, even at concentrations as high as 130 mg/mL (Fig. 2*F* and Fig. S3). The molecular envelope revealed by combined small and wide-angle X-ray scattering (SWAXS) is in excellent agreement with the crystal structure (Fig. 2 *F* and *G*), with a χ^2 of 0.93. An indent seen in the SWAXS structure (Fig. 2*F*, arrows) corresponds to the hydrophobic pocket; however, the YVEF peptide appears to be disordered in the monomer structure in solution because it did not contribute to the SWAXS envelope. Thus, the αTSR monomer in solution adopts essentially the same structure as in trimers in crystals. Furthermore, SWAXS from the mammalian material lacking the N-terminal YVEF peptide shows a scattering profile similar to that of the *Pichia* material with YVEF and

Table 1. X-ray diffraction data collection, phasing, and refinement statistics for αTSR structure

	Protein Data Bank deposition code*		
	3VDJ	3VDK	3VDL
Data collection			
Space group	H32	H32	$P4_32_12$
Molecules/asymmetric unit	1	1	3
Cell dimensions			
a, b, c (Å)	66.52, 66.52, 85.51	66.53, 66.53, 86.49	63.37, 63.37, 118.16
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 90
Wavelength, Å	0.97949	1.07195	0.97949
Resolution, Å	50–1.70 (1.73–1.70)†	50–1.85 (1.92–1.85)†	50–2.04 (2.11–2.04)†
Observations (unique)	40,160 (8,160)	41,336 (6,375)	176,960 (15,611)
I/σ overall	22.7 (3.6)	42.6 (2.9)	21.6 (1.44)
Completeness	99.1 (91.1)	98.1 (86.9)	97.1 (74.7)
R_{merge}‡	5.7 (35.1)	9.1 (64.3)	9.6 (69.4)
SAD phasing			
Figure of merit before/after density modification		0.34/0.58	
Structure refinement			
Resolution	50–1.70	50–1.85	50–2.04
R_{work}/R_{free}, %§	16.9/19.8	17.9/21.5	19.3/24.3
Protein/solvent atoms	568/103	572/53	1701/90
Bond length deviation from ideal geometry (Å)	0.006	0.006	0.007
Bond angle deviation from ideal geometry (°)	0.986	1.0646	1.040
Ramachandran favored/allowed/outliers (%)	98.6/1.4/0	100/0/0	98.6/1.4/0
Average B factor for protein/solvent	22.0/33.8	34.9/39.3	32.9/33.8

*Experimental data has been deposited with the indicated deposition ID codes.
†Values in parentheses are for highest-resolution shell.
‡$R_{merge} = \sum i,h|I(i,h) - <I(h)>| / \sum i,h|I(i,h)|$ where $I(i, h)$ and $<I(h)>$ are the ith and mean measurement of intensity of reflection h.
§The R_{free} value was calculated using 5% of the data.

to the profile predicted by the crystal structure (Fig. 2*G*). Thus, the hydrophobic pocket is unoccupied in solution. Furthermore, the unique αTSR fold is independent of the eukaryotic expression system used, and the N-terminal tetrapeptide extension (like the C-terminal His tag, which is also present in the crystallized construct) has little or no impact on the folded structure.

Discussion

The TSR superfamily is widespread among animals and protozoans, with 187 TSR domains in 41 human proteins (21). However, the αTSR domain in CS is strikingly divergent in sequence and structure from other TSR domains, including those in other *Plasmodium* surface proteins (8). Residues that form the unique αTSR hydrophobic core and pocket, α1-helix, and CS flap are highly conserved across 10 *Plasmodium* species infecting a range of vertebrate hosts (Fig. 1*B*). The structural adaptations that evolved specifically in the CS αTSR domain change its overall shape. The thin and elongated prototypical TSR domain is shortened on its long axis by 15 Å and widened 10 Å on the other axis by the α1-helix and flap (Figs. 1*D* and 2 *A*–*C*). Moreover, in TSR domains the N- and C termini are located on opposite ends of the long axis, enabling their tandem use in multidomain proteins (7) (Fig. 1*D*). In contrast, the N terminus of the CS αTSR domain is located on the same end of the domain as the C terminus (Figs. 1*D* and 2 *A* and *E*). This has important implications for how the αTSR is oriented on the *P. falciparum* sporozoite surface, between the tetrapeptide repeats and C-terminal GPI anchor.

CS is a major sporozoite protein, accounting for 5–15% of total sporozoite [^{35}S]-methionine incorporation (22), and uniformly coats the sporozoite surface (23). Before proteolytic cleavage by a parasite protease in region I, CS is sufficiently densely packed on sporozoites to protect αTSR from antibody recognition (19). How CS forms a protective sheath over the sporozoite plasma membrane and whether dense packing is facilitated by specific packing interactions between neighboring αTSR domains are important questions for further research. The atomic structure of αTSR defined here enables lower resolution investigation of whether the sporozoite sheath contains ordered assemblies of smaller units.

As the immunodominant sporozoite antigen, for many years CS has been the main target of exoerythrocytic vaccine research. During natural infections, young children in endemic regions and adults in nonendemic regions initially develop antibody responses to all three major regions of CS: the N terminus, the central repeat region, and the C terminus containing the αTSR, as shown by reaction with synthetic peptides (24, 25). Repeated infections in endemic regions skew the antibody response to the central repeat region, and reinfections continue to occur.

An important caveat to all previous studies on antibody responses to CS subregions is that the antibodies were interrogated using short synthetic peptides. However, we found here that the αTSR domain adopts a highly specific 3D shape, in which amino residues that are distant in sequence are close in structure. Hence the important epitopes on the αTSR domain are conformational and are unlikely to be well represented by synthetic peptide fragments. Our study now makes available to research a well-folded, highly characterized domain from CS. It will be important not only to test this domain as a vaccine, but also to characterize the antibody responses to the αTSR domain of individuals from endemic regions, and of individuals vaccinated with sporozoites and RTS,S.

The αTSR domain also contains several T-cell epitopes, one of which, CS.T3, is responsible for a $CD4^+$ T-cell response that correlates with protection (26) (Fig. 1*B*). The other T-cell epitopes, Th2R and Th3R, are polymorphic regions of the αTSR (27). These polymorphic T-cell epitopes and the highest sequence variation between *P. falciparum* isolates locate to the structurally unique elements of the αTSR: the hydrophilic faces of the α1-helix and the CS flap (Fig. 1*B*). Thus, the unique features of the αTSR domain, which endow it with a hydrophobic core, a hydrophobic pocket, and close proximity between its

N- and C-terminal ends with implications for orientation in the parasite sheath, evolved despite considerable immunological cost, and must be adaptive for the parasite.

The highly conserved hydrophobic pocket in the αTSR domain is particularly fascinating and completely unexpected. Most previous work on the role of the TSR domain of CS in host cell recognition and invasion has focused on putative heparin-binding peptides; however, whereas peptides derived from the TSR region have been shown to bind to heparin, we found that the intact, folded αTSR domain does not. The pocket may have a hydrophobic ligand that is important in parasite–host interactions. The most definitive way to test this hypothesis would be to introduce mutations in this pocket into the parasite genome and test their function in vivo.

A CS fragment containing 19 tetrapeptide repeats and the sequence of the αTSR region is in large phase III trials; however, to date infants and 1-y olds receiving three vaccine doses show only a 35% lower rate of infection in the first year after immunization (6). The current vaccine is produced in the yeast *Saccharomyces cerevisiae* (28, 29). The CS fragment is fused N terminal to the small hepatitis B virus surface antigen (s-HBsAg) to make the component termed RTS (repeat/TSR/s-HBSAG). There is no secretion signal sequence N terminal to the CS in this fusion protein. The same yeast cells coexpress intact s-HBsAg, which is termed S, and thus the vaccine product is termed RTS,S.

It is unclear whether the αTSR sequence present in RTS,S has the same disulfide-bonded structure or fold as in our crystal structure. Although both proteins are made in yeast, the s-HBsAg displays unusual characteristics. s-HBsAg has an N-terminal transmembrane domain 1 (TM-I) that is not removed, a cytoplasmic domain, a TM-II domain followed by an extracellular loop with a single N-linked glycosylation site, and two further TM domains near the C terminus. Expression of s-HBsAg in mammalian cells results in formation of virus-like, lipid-associated particles that are N-glycosylated and secreted (30). However, in yeast, s-HBsAg accumulates in the endoplasmic reticulum (ER), and virus-like particles form after the cell wall is broken and cells are treated with detergents (31). In further contrast to s-HBsAg made in mammalian cells, and to proteins secreted by yeast cells, s-HBsAg made in yeast cells is not N-glycosylated (28). Therefore, it is not known whether the CS fragment fused to s-HBsAg and expressed in yeast has been exposed to the lumen of the ER where disulfide isomerases, chaperonins, and other machinery regulate disulfide bond formation, folding, and quality control. Our αTSR construct has a signal sequence, which is removed as shown by mass spectrometry, and is secreted into the medium, and thus has passed ER quality control and is expected to be properly folded.

TSR domains can be fucosylated by the protein O-fucosyl transferase 2 (*POFUT2*). In TSR domain 1 of thrombospondin, the CSVTCGDG sequence is fucosylated on Thr (7, 32). The fucosylation motif is highly conserved in the CSVTCGNG sequence in CS αTSR. Fucose can be further modified by addition of β1–3 glucose (33, 34). PSI-BLAST searches suggest *POFUT2* (and not *POFUT1*, a homolog involved in Notch fucosylation) is conserved in *P. falciparum* (gene ID PFI0445c) and other apicomplexans. Mass spectrometry of our purified αTSR preparations showed absence of glycosylation of the *Pichia* protein (calculated 8880.0 Da with four half-cystines, found 8879.9 Da) and presence of both fucose and hexose (e.g., glucose) on the HEK293T protein (calculated 8649.4 Da with four half-cystines, found 8649.1 Da). Lack of fucosylation in *Pichia* is in agreement with lack of density for fucose in our crystal structure and lack of a *POFUT2* homolog in yeast. The possibility of fucosylation of CS has not previously been raised in the malaria literature. Whether CS is fucosylated on the sporozoite surface and how fucosylation affects vaccine efficacy are important questions for the future.

The elucidation of a unique structured domain in CS as the αTSR domain, comprising not just the TSR homology region but also region III, is a significant step forward for preerythrocytic malaria vaccines. This work enables subunit vaccine design to transition from sequence based to structure based. Furthermore, it will now be possible to correlate protective and broadly cross-reactive immune responses in vaccinated individuals with specific structural features of αTSR, much as is now being done in the analysis of protective antibodies to influenza and HIV (35, 36).

Materials and Methods

Plasmid Construction, Protein Expression, and Purification. A codon-optimized gene encoding E310-C374 of *P. falciparum* 3D7 CS followed by sequence PHHHHHHA (DNA2.0) was cloned in frame to the *S. cerevisiae* α-factor signal sequence of pPIC9K (Invitrogen) using EcoRI and NotI sites. The EcoRI site introduced a YVEF sequence before E310. *P. pastoris* strain GS115 was electroporated with SalI linearized vector. Standard protocols were used for culture growth in buffered complex glycerol media followed by protein expression for 72 h in buffered complex methanol media (Invitrogen; *Pichia* Expression kit, manual 25-0043).

Culture supernatant supplemented with 0.5 M NaCl, 10 mM imidazole, and 0.5 mM nickel chloride and sodium hydroxide to pH 7.8 was allowed to precipitate overnight at 4 °C. Supernatant (2 L) was applied to a 10-mL nickel-NTA agarose (Qiagen) at 5 mL/min. After wash with 0.3 M NaCl, 10 mM Tris pH 8.0, 20 mM imidazole, protein was eluted with 0.15 M NaCl, 10 mM Tris pH 8.0, 0.5 M imidazole. Fractions containing CS protein were concentrated and further purified by gel filtration on Superdex 75 (GE Healthcare).

The same E310-C374 CS sequence followed by PHHHHHHA was expressed in HEK293T cell transfectants using the pET8/pLEXm vectors. Material was purified by nickel-NTA agarose and gel filtration with the same protocol as the *Pichia* protein.

Recombinant CS was from Science Applications International Corporation (SAIC), the primary contractor for the Malaria Vaccine Production and Support Services (MVPSS) contract (AI-N01-054210) administered through the Division of Microbiology and Infectious Diseases/National Institute of Allergy and Infectious Diseases/National Institutes of Health. CS (residues 21–382) of *P. falciparum* 3D7 strain (GenBank accession no. CAB38998) was expressed in *Pseudomonas fluorescens* at Pfenex. CS was 96% pure and 95% monomeric. SDS/PAGE showed a single band at ∼55 kDa.

Crystallization and Structure Determination. Protein in 10 mM Tris pH 7.5, 0.25 M sodium chloride (20–40 mg/mL) formed tetragonal crystals in 1.4 M sodium citrate pH 6.1 and trigonal crystals in 0.1 M citrate pH 4.0, 1 M lithium chloride, 20% (wt/vol) PEG 6000. Trigonal crystals were soaked in 0.1 M citrate pH 4.0, 1.0 M lithium chloride, 25% (wt/vol) PEG 6000 with 2.4 mM potassium tetranitroplatinate (II) for 18 h (Hampton Research). Crystals were cryopreserved by incrementing PEG by 5% (in the case of the trigonal crystals) and then successive 10-s soaks with 5% (wt/vol) increments of glycerol up to 20% before freezing in liquid nitrogen. Diffraction data were collected on beamline 24ID at APS and indexed and scaled with HKL 2000 (37). The structure was determined by SAD using CCP4 and PHENIX software suites (38, 39). Structure refinement was with COOT and PHENIX (39, 40).

SWAXS Data Collection, Analysis, and Model Reconstruction. Small and wide-angle X-ray scattering was at beam line X9A at the National Synchrotron Light Source (Upton, NY). αTSR samples were passed through a flow capillary, collecting data at 20-s exposures in triplicate at concentrations from 12 to 130 mg/mL (*Pichia* protein) or 15–45 mg/mL (HEK293T protein). Scattering from 10 mM Tris pH 7.5, 0.25 M NaCl buffer was subtracted. Guinier analysis showed no signs of radiation damage or aggregation. Data were reduced by circular averaging of the images and scaled to obtain the scattering curve. GNOM (41) was used to calculate distance distributions. DAMMIN (42) was used to restore low resolution ab initio shapes. Twenty solutions from DAMMIN were averaged to obtain the final model using DAMAVER (43). This model was converted to a surface map using the SITUS program suite (44). The theoretical solution scattering of the crystal structure of αTSR was calculated using CRYSOL (45).

Mass Spectrometry. Intact protein samples were desalted by reversed-phase microbore HPLC and subjected to electrospray ionization mass spectrometry on a Bruker ESQ-LC ion trap mass spectrometer (mass accuracy within 1 Da).

ACKNOWLEDGMENTS. We thank beamlines X9 at National Synchotron Light Source and 24ID at Advanced Photon Source and David King (Howard Hughes Medical Institute, Department of Molecular and Cell Biology, University of California at Berkeley) for mass spectrometry. This work was supported by National Institutes of Health Grant A1095686.

1. Kappe SH, Buscaglia CA, Nussenzweig V (2004) Plasmodium sporozoite molecular cell biology. *Annu Rev Cell Dev Biol* 20:29–59.
2. Ménard R (2000) The journey of the malaria sporozoite through its hosts: Two parasite proteins lead the way. *Microbes Infect* 2:633–642.
3. Stewart MJ, Vanderberg JP (1991) Malaria sporozoites release circumsporozoite protein from their apical end and translocate it along their surface. *J Protozool* 38: 411–421.
4. Wang Q, Fujioka H, Nussenzweig V (2005) Mutational analysis of the GPI-anchor addition sequence from the circumsporozoite protein of Plasmodium. *Cell Microbiol* 7:1616–1626.
5. Vekemans J, Ballou WR (2008) *Plasmodium falciparum* malaria vaccines in development. *Expert Rev Vaccines* 7:223–240.
6. Agnandji ST, et al.; RTS,S Clinical Trials Partnership (2011) First results of phase 3 trial of RTS,S/AS01 malaria vaccine in African children. *N Engl J Med* 365:1863–1875.
7. Tan K, et al. (2002) Crystal structure of the TSP-1 type 1 repeats: A novel layered fold and its biological implication. *J Cell Biol* 159:373–382.
8. Tossavainen H, et al. (2006) The layered fold of the TSR domain of *P. falciparum* TRAP contains a heparin binding site. *Protein Sci* 15:1760–1768.
9. Pääkkönen K, et al. (2006) Solution structures of the first and fourth TSR domains of F-spondin. *Proteins* 64:665–672.
10. Akiyama M, Takeda S, Kokame K, Takagi J, Miyata T (2009) Crystal structures of the noncatalytic domains of ADAMTS13 reveal multiple discontinuous exosites for von Willebrand factor. *Proc Natl Acad Sci USA* 106:19274–19279.
11. Pancake SJ, Holt GD, Mellouk S, Hoffman SL (1992) Malaria sporozoites and circumsporozoite proteins bind specifically to sulfated glycoconjugates. *J Cell Biol* 117: 1351–1357.
12. Cerami C, et al. (1992) The basolateral domain of the hepatocyte plasma membrane bears receptors for the circumsporozoite protein of *Plasmodium falciparum* sporozoites. *Cell* 70:1021–1033.
13. Frevert U, et al. (1993) Malaria circumsporozoite protein binds to heparan sulfate proteoglycans associated with the surface membrane of hepatocytes. *J Exp Med* 177: 1287–1298.
14. Pradel G, Garapaty S, Frevert U (2002) Proteoglycans mediate malaria sporozoite targeting to the liver. *Mol Microbiol* 45:637–651.
15. Coppi A, et al. (2007) Heparan sulfate proteoglycans provide a signal to Plasmodium sporozoites to stop migrating and productively invade host cells. *Cell Host Microbe* 2: 316–327.
16. Rathore D, et al. (2001) Direct measurement of the interactions of glycosaminoglycans and a heparin decasaccharide with the malaria circumsporozoite protein. *Biochemistry* 40:11518–11524.
17. Ancsin JB, Kisilevsky R (2004) A binding site for highly sulfated heparan sulfate is identified in the N terminus of the circumsporozoite protein: Significance for malarial sporozoite attachment to hepatocytes. *J Biol Chem* 279:21824–21832.
18. Sinnis P, et al. (1994) Structural and functional properties of region II-plus of the malaria circumsporozoite protein. *J Exp Med* 180:297–306.
19. Coppi A, et al. (2011) The malaria circumsporozoite protein has two functional domains, each with distinct roles as sporozoites journey from mosquito to mammalian host. *J Exp Med* 208:341–356.
20. Saxena AK, et al. (2006) The essential mosquito-stage P25 and P28 proteins from Plasmodium form tile-like triangular prisms. *Nat Struct Mol Biol* 13:90–91.
21. Tucker RP (2004) The thrombospondin type 1 repeat superfamily. *Int J Biochem Cell Biol* 36:969–974.
22. Yoshida N, Potocnjak P, Nussenzweig V, Nussenzweig RS (1981) Biosynthesis of Pb44, the protective antigen of sporozoites of *Plasmodium berghei*. *J Exp Med* 154: 1225–1236.
23. Nussenzweig V, Nussenzweig RS (1985) Circumsporozoite proteins of malaria parasites. *Cell* 42:401–403.
24. Calle JM, et al. (1992) Recognition of different domains of the *Plasmodium falciparum* CS protein by the sera of naturally infected individuals compared with those of sporozoite-immunized volunteers. *J Immunol* 149:2695–2701.
25. Lopez JA, et al. (1996) Recognition of synthetic 104-mer and 102-mer peptides corresponding to N- and C-terminal nonrepeat regions of the *Plasmodium falciparum* circumsporozoite protein by sera from human donors. *Am J Trop Med Hyg* 55: 424–429.
26. Reece WH, et al. (2004) A CD4(+) T-cell immune response to a conserved epitope in the circumsporozoite protein correlates with protection from natural *Plasmodium falciparum* infection and disease. *Nat Med* 10:406–410.
27. Jalloh A, Jalloh M, Matsuoka H (2009) T-cell epitope polymorphisms of the *Plasmodium falciparum* circumsporozoite protein among field isolates from Sierra Leone: Age-dependent haplotype distribution? *Malar J* 8:120.
28. Jacobs E, et al. (1989) Simultaneous synthesis and assembly of various hepatitis B surface proteins in *Saccharomyces cerevisiae*. *Gene* 80:279–291.
29. Gordon DM, et al. (1995) Safety, immunogenicity, and efficacy of a recombinantly produced *Plasmodium falciparum* circumsporozoite protein-hepatitis B surface antigen subunit vaccine. *J Infect Dis* 171:1576–1585.
30. Patient R, Hourioux C, Roingeard P (2009) Morphogenesis of hepatitis B virus and its subviral envelope particles. *Cell Microbiol* 11:1561–1570.
31. Lünsdorf H, Gurramkonda C, Adnan A, Khanna N, Rinas U (2011) Virus-like particle production with yeast: ultrastructural and immunocytochemical insights into *Pichia pastoris* producing high levels of the hepatitis B surface antigen. *Microb Cell Fact* 10:48.
32. Hofsteenge J, et al. (2001) C-mannosylation and O-fucosylation of the thrombospondin type 1 module. *J Biol Chem* 276:6485–6498.
33. Sato T, et al. (2006) Molecular cloning and characterization of a novel human β1,3-glucosyltransferase, which is localized at the endoplasmic reticulum and glucosylates O-linked fucosylglycan on thrombospondin type 1 repeat domain. *Glycobiology* 16: 1194–1206.
34. Kozma K, et al. (2006) Identification and characterization of a β1,3-glucosyltransferase that synthesizes the Glc-β1,3-Fuc disaccharide on thrombospondin type 1 repeats. *J Biol Chem* 281:36742–36751.
35. Moir S, Malaspina A, Fauci AS (2011) Prospects for an HIV vaccine: Leading B cells down the right path. *Nat Struct Mol Biol* 18:1317–1321.
36. Rappuoli R, Aderem A (2011) A 2020 vision for vaccines against HIV, tuberculosis and malaria. *Nature* 473:463–469.
37. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
38. Bailey S; Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760–763.
39. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213–221.
40. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
41. Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Cryst* 25:495–503.
42. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76:2879–2886.
43. Volkov VV, Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Cryst* 36:860–864.
44. Wriggers W (2010) Using Situs for the integration of multi-resolution structures. *Biophys Rev* 2:21–27.
45. Svergun DI, Barberato C, Koch MH (1995) CRYSOL: A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Cryst* 28:768–773.
46. Garcia-Boronat M, Diez-Rivero CM, Reinherz EL, Reche PA (2008) PVS: A web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res* 36(Web Server issue):W35–W41.

Supporting Information

Doud et al. 10.1073/pnas.1205737109

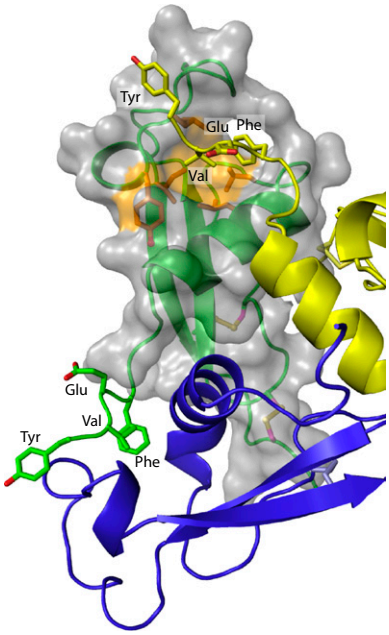

Fig. S1. Tyr-Val-Glu-Phe (YVEF) tetrapeptide binding to the hydrophobic pocket. Portion of a trimer is shown. Each monomer is shown in cartoon in a different color. Transparent surface is shown for the non-YVEF portion of the green αTSR monomer. Residues forming the pocket and its surface are colored orange as in Fig. 2*E*. Val and Phe are buried in the pocket. Each monomer's tetrapeptide is buried in a neighboring monomer's hydrophobic pocket.

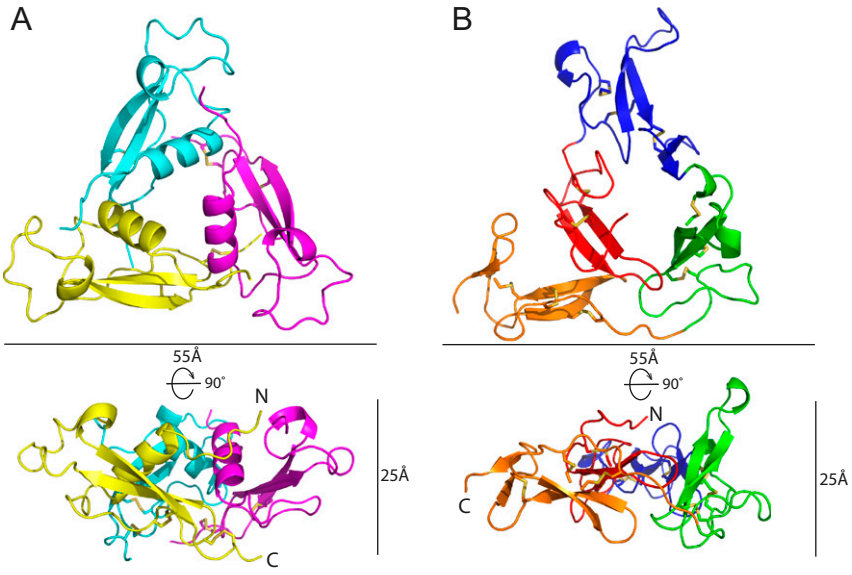

Fig. S2. Crystallographic trimer assembly of the αTSR domain and comparison with Pvs25. (*A*) Crystallographic αTSR trimer viewed parallel to (*Upper*) and orthogonal to (*Lower*) the threefold axis. Two nearly identical trimer assemblies were observed in different crystal lattices and space groups. (*B*) Similar views of the Pvs25 monomer. The Pvs25 monomer was fit to the αTSR trimer by cross-correlation after both structures were resolution filtered to 15 Å. "Superimposed" structures were then separated horizontally on the page to create common views in *A* and *B*. Four EGF-like domains are individually colored.

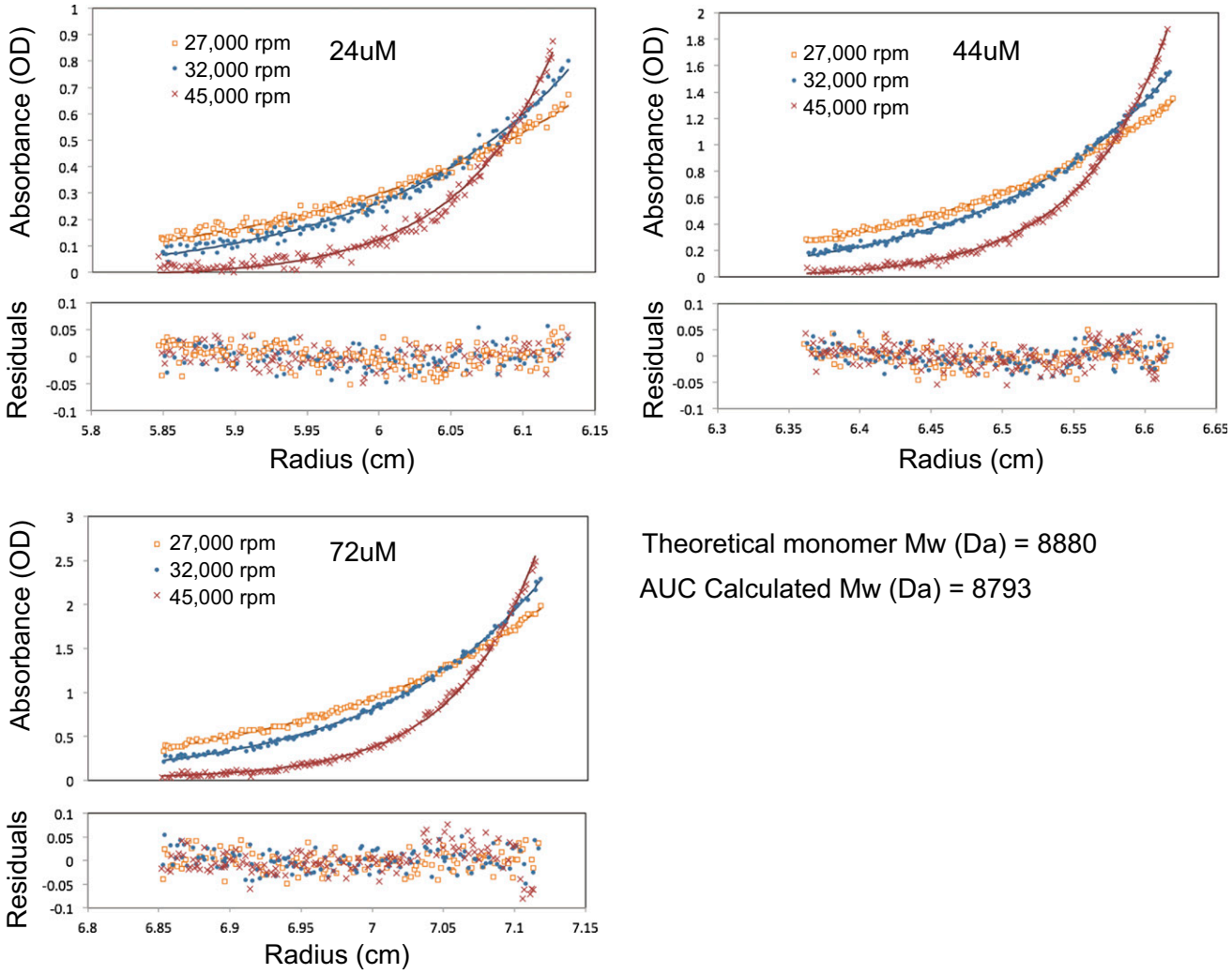

Fig. S3. Analytical ultracentrifugation sedimentation equilibrium. Symbols represent the recorded absorbance values at the given radius after the system reached equilibrium for three αTSR concentrations. Lines represent fit to a self-association model assuming partial specific volume 0.7309 mL/g, solvent density 1.00702 g/mL, and viscosity 1.5422 × 10^{-2} cm^2/s and no self-association (n = 1). Best fit was achieved for a monomeric model (Mw = 8.79 kDa, n = 1). Purified *Pichia* αTSR domain was diluted in 10 mM Tris pH 7.5, 0.25 M NaCl to final concentrations of 24, 44, and 72 μM and used in sedimentation equilibrium analysis on a Beckman XL-A analytical ultracentrifuge (Beckman Coulter). Samples were centrifuged at 27,000, 32,000, and 45,000 rpm using an An-60 Ti rotor. Sedimentation equilibrium was achieved as verified during data analysis. Sednterp was used to calculate v-bar and buffer density and viscosity. SEDFIT was used to process the raw data and SEDPHAT was further used to perform a global fitting of the data to self-association models (1).

1. Vistica J, et al. (2004) Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Anal Biochem* 326:234–256.