RESEARCH ARTICLE

A structural classification of the variant surface glycoproteins of the African trypanosome

Sara Đaković[1], Johan P. Zeelen[1], Anastasia Gkeka[2], Monica Chandra[1,2], Monique van Straaten[1], Konstantina Foti[1], Janet Zhong[1], Evi P. Vlachou[2], Francisco Aresta-Branco[1,2], Joseph P. Verdi[2], F. Nina Papavasiliou[2], C. Erec Stebbins[1]*

1 Division of Structural Biology of Infection and Immunity, German Cancer Research Center, Heidelberg, Germany, 2 Division of Immune Diversity, German Cancer Research Center, Heidelberg, Germany

* e.stebbins@dkfz-heidelberg.de

Abstract

Long-term immune evasion by the African trypanosome is achieved through repetitive cycles of surface protein replacement with antigenically distinct versions of the dense Variant Surface Glycoprotein (VSG) coat. Thousands of VSG genes and pseudo-genes exist in the parasite genome that, together with genetic recombination mechanisms, allow for essentially unlimited immune escape from the adaptive immune system of the host. The diversity space of the "VSGnome" at the protein level was thought to be limited to a few related folds whose structures were determined more than 30 years ago. However, recent progress has shown that the VSGs possess significantly more architectural variation than had been appreciated. Here we combine experimental X-ray crystallography (presenting structures of N-terminal domains of coat proteins VSG11, VSG21, VSG545, VSG558, and VSG615) with deep-learning prediction using Alphafold to produce models of hundreds of VSG proteins. We classify the VSGnome into groups based on protein architecture and oligomerization state, contextualize recent bioinformatics clustering schemes, and extensively map VSG-diversity space. We demonstrate that in addition to the structural variability and post-translational modifications observed thus far, VSGs are also characterized by variations in oligomerization state and possess inherent flexibility and alternative conformations, lending additional variability to what is exposed to the immune system. Finally, these additional experimental structures and the hundreds of Alphafold predictions confirm that the molecular surfaces of the VSGs remain distinct from variant to variant, supporting the hypothesis that protein surface diversity is central to the process of antigenic variation used by this organism during infection.

Author summary

The African trypanosome is a single-celled parasite that causes African Sleeping Sickness in humans and related diseases in many animals. It reproduces in the blood of an infected organism, completely exposed to the immune system, but survives due to its ability to

repeatedly disguise itself with new surface coat proteins that fool the antibody response. The African trypanosome has thousands of these Variant Surface Glycoproteins (VSGs) in its genome to switch onto the parasite surface. Therefore, trypanosomiasis is centered on the antibody-VSG interaction and how it varies over time. At the most basic level, understanding the possibilities and constraints of VSG protein architecture is a critical foundation to understand the interaction of the pathogen with the host immune system. The work in this paper thoroughly characterizes this genomic collection of VSGs at the molecular level, classifying the coat proteins into related families based on several important criteria, thereby providing the groundwork for developing an understanding of how this organism is able to evade immunity and thrive during infection.

Introduction

African trypanosomiasis is a human and animal infectious disease caused by several species of protozoan parasites of the genus *Trypanosoma* [1,2]. These single-celled, eukaryotic parasites can live and reproduce extracellularly in the bloodstream of the host, and are transmitted by the tsetse fly vector (*Glossina sp.*) [3]. The geographical range of the tsetse fly in Africa correlates to the distribution of human trypanosomiasis, which at present covers a region of 8 million km^2 between 14 and 20 degrees latitude [4]. African trypanosomiasis has hampered Central Africa's economic progress due to its impact on both human and livestock populations [5,6].

Within the blood, the trypanosome population is continuously exposed to the immune system of the host, yet is able to thrive and persist. This feat is made possible by a highly optimized system of antigenic variation, in which the ~10 million, monoallellically expressed molecules of the Variant Surface Glycoprotein (VSG) coat undergo repeated cycles of "switching", a process by which antigenically distinct VSGs are expressed at different times [7–9]. This creates a cyclic process of trypanosome growth to high parasitemia, immune response and clearance of the dominant VSG variants, and the subsequent growth of immune-escape variants expressing different VSGs. This process renders the host in a perpetual state of infection that cannot be cleared without pharmacological intervention, typically leading to long-term morbidity and mortality [10]. Central to this process are the thousands of VSG genes and pseudogenes in the parasite genome that serve as the parasite's extensive antigen repertoire.

The mature VSG proteins are attached to the cell surface by a GPI-anchor, and consist of two main regions: (1) a large, N-terminal domain (NTD) of roughly 300–400 amino acids that is most distal to the membrane and (2) a smaller, membrane-proximal C-terminal domain (CTD) spanning 80–120 amino acids beneath the NTD that harbors the GPI-anchor [11]. Multiple studies have shown that the CTD is minimally immunogenic [12] and quite possibly inaccessible to the immune system when it is part of the coat [12,13], which correlates well with it being much more highly conserved in sequence than the NTD [14,15]. The NTD is therefore presumed to be the most antigenic portion of the VSG. Consistent with this hypothesis, the NTD typically shows approximately 10–30% identity from variant to variant [14]. Most of the conserved residues occur in the architecturally common regions of the VSGs, whereas the molecular surfaces (as visualized by protein structures) show very little similarity [11].

Therefore, at the heart of antigenic variation in the African trypanosome are the NTDs of the VSGs. These large subdomains are elongated folds centered on a three-helix bundle scaffold. The three-helix bundle is a ubiquitous fold harbored in proteins with diverse activities, where the structural elements specific for differing functions are often coded by sequences

inserted between the helices or at the termini of the scaffold. All VSGs studied to date are so structured, possessing subdomains at spatially opposite ends of the bundle: a top lobe and bottom lobe. The recent elucidation of additional VSG structures upturned the notion that all VSGs would possess highly similar protein folds to the initial structures determined in the 1990s. What is now evident from recent work is that the VSGs possess much more architectural and topological variation than had initially been appreciated [16–18].

Furthermore, for the thirty years after the determination of the first VSG structures [19–23], it was assumed that all VSGs were homodimers. However, more recent structures and biochemical analyses have shown that two broad VSG classes can be distinguished by their oligomeric state: one class harboring exclusively dimers and the other class characterized by concentration-dependent trimers (existing in solution in either a stable monomeric or trimeric state, depending on the protein concentration [16,17,24]). At the extreme densities of packing in the two-dimensional trypanosome surface coat, it is possible that all VSGs of this class will exist in the trimeric state, although no high-resolution imaging of the parasite coat that could establish this has been reported.

The uniqueness of each VSG molecular surface led to the assumption that antigenic variation occurred exclusively through amino acid sequence divergence. Recent structures and mass spectrometric analyses of VSGs have shown that this is not solely the case. Many VSGs can be modified by post-translational O-linked glycosylation at the top of the molecule and this modification has been shown to be potently immunomodulatory [17]. Intriguingly, the same class that adopts the trimeric oligomerization is to-date the only class in which such O-linked glycosylation has been observed.

Thus, in the last few years, many notions regarding the VSGs have had to be reexamined. Concurrently, the diversity space of the VSGs has been broadened dramatically, raising the question of whether the thousands of possible VSG proteins in the genome could be organized into a coherent schema relating sequence, structure, and function. Along those lines, several groups have undertaken bioinformatic analysis of the "VSGnome" (the set of possible VSG proteins from the trypanosome genome). Most prominent have been two papers that have used sequence clustering algorithms to divide the VSGs into subclasses [25,26]. While these papers differed in aspects of methodology and resultant classification, there was broad agreement on the classification of VSGs that appeared to be much more highly related to each other. However, none of these efforts directly took protein architecture into account, likely due to the paucity of experimentally determined structures available at the time of the analyses.

To address this issue, we use here a collection of experimentally determined structures of VSGs (including many solved over the last few years by our group, with five previously unpublished structures included in this manuscript) as well as hundreds of predicted structures generated by the deep-learning system, AlphaFold, that has proven powerful at the creation of accurate three-dimension protein models from amino acid sequence alone [27,28]. The combined experimental and predicted data establish a structure-based classification scheme for the VSG proteins that is generally consistent with previous attempts to classify the VSGs, but that also places classification efforts on an architectural and functional foundation.

Materials and methods

Structural determination of VSG11

Two constructs of VSG11 NTD were used in the structural studies in this manuscript (all VSGs, unless otherwise noted, are from the strain Lister 427). The first was the wild type sequence (VSG11$_{WT}$) that (1) crystallized as a monomer in the asymmetric unit (with a

crystallographic trimer) in sodium-potassium tartrate with two structures determined: VSG11$_{WT}$-Iodine, a crystal soaked in Na/K iodine diffracting to 1.27Å, and VSG11$_{WT}$-Oil, a crystal diffracting to 1.23 Å resolution cryo-cooled in oil, and (2) crystallized as two monomers in the asymmetric unit (not a dimer and with a crystallographic trimer–see Results below) in ammonium sulfate (diffracting to 1.75 Å, denoted VSG11$_{WT}$-AS). The second construct of VSG11 was a chimeric form consisting of the VSG11$_{WT}$ NTD connected to the VSG2 CTD, denoted VSG11$_{N2C}$. This construct crystallized with 18 monomers of VSG11$_{N2C}$ in an asymmetric unit comprised of six trimeric assemblies (denoted VSG11$_{N2C}$-18mer). VSG11$_{WT}$-18mer crystals were also obtained that diffracted to nearly 3Å resolution, but the best diffracting crystals were produced by the VSG11$_{N2C}$ construct (2.6 Å resolution). Therefore, the presence of the VSG2 CTD was not required for the formation of the 18mer form, and since in both cases the CTDs were not present in the crystallized form, it is unlikely that the 18mer form is tied in any manner to the chimeric form.

Plasmids used to generate VSG11$_{WT}$ and VSG11$_{N2C}$ are described in [17]. Plasmids were first linearized by EcoRV (New England Biolabs), and then transfected into VSG2-expressing cells (2T1): 10ug of each plasmid were mixed with 100ul of cells (at a concentration of between 2.5x10^7 and 3x10^7) in Tb-BSF buffer (90mM Na$_2$HPO$_4$, pH 7.3, 5mM KCl, 0.15mM CaCl$_2$, 50mM HEPES, pH7.3), using an AMAXA nucleofector (Lonza) program X-001, as previously described [29]. Blasticidin at a concentration of 100ug/ml was added after 6h and single-cell clones were obtained by serial dilutions in 24-well plates and collected after 5 days.

Clones initially screened with flow cytometry for VSG2 loss of expression using a monoclonal VSG2$_{WT}$ antibody [30] and VSG11 gain of expression with anti-VSG11 antisera. To determine the binding of antisera to live trypanosomes, 1 x 10^6 parasites were collected and incubated with VSG2$_{WT}$ antisera (1:4000) or VSG11$_{WT}$ (1:1000) together with Fc block (1:200, BD Pharmingen) in cold HMI-9 without FBS for 10 min at 4˚C. Cells were washed once with cold HMI-9 and resuspended in 200µl cold HMI-9 with rat anti-mouse IgM-FITC (1:500, Biolegend). After one wash with cold HMI-9, cells were resuspended in 150µl HMI-9 and immediately analyzed with FACSCalibur (BD Biosciences) and FlowJo software (v10). For a second step, clones were sequenced by isolating RNA using the RNeasy Mini Kit (Qiagen), followed by DNAse treatment with the TURBO DNA-free kit (Invitrogen) and cDNA synthesis with ProtoScript II First Strand cDNA Synthesis (New England Biolabs). The sequences were then amplified, using Phusion High-Fidelity DNA Polymerase (New England Biolabs), a forward primer binding to the spliced leader sequence and a reverse binding to the VSG 3´untranslated region. The final products were purified by gel extraction from a 1% gel with the NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel) and sent for Sanger sequencing.

All VSG11 constructs were expressed in *T. b. brucei* cultured at 37˚C and 5% CO$_2$ in HMI-9 media (PAN Biotech) supplemented with 10% fetal calf serum (Gibco), L-cysteine and ß-mercaptoethanol. The cells from 3.6 liter culture were pelleted and the VSG11 proteins purified through modifications of previously published protocols [31]. The cells were lysed with 40 ml 0.4 mM ZnCl$_2$, after centrifugation (10.000 g, 10 min) the pellet was resuspended in 30 ml 20 mM Hepes/NaOH pH = 8.0, 150 mM NaCl (42˚C) and centrifuged (10.000 g, 10 min). The supernatant was passed through a 25 ml Q-sepharose Fast-flow column (GE Healthcare) equilibrated with 20 mM Hepes/NaOH pH = 8.0, 150 mM NaCl. The flow through was collected and concentrated. The protein was further purified on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) equilibrated in 10 mM HEPES/NaOH pH = 8.0, 150 mM NaCl. The fractions containing the VSG protein were pooled and concentrated.

VSG11$_{WT}$ crystals containing only the N-terminal domain appeared after several days at 22˚C using the hanging drop method with 6 mg/ml VSG11 (full-length) protein in a 1:1 volume ratio against 100 mM Tris/HCl pH = 7.5 1.6–1.75M sodium-potassium tartrate. CryoOil

(MiTeGen) was used as cryoprotectant and the VSG11$_{WT}$-Oil crystals flash-cooled in liquid nitrogen. The loss of the CTD is presumed to have occurred during the crystallization stage. VSG11$_{WT}$-Iodine crystals soaked in 200 mM KI, 100 mM Tris/HCl pH 8.0 and 1.7 M NaKTartrate were flash-cooled directly in liquid nitrogen.

VSG11$_{WT}$-AS crystals were grown at 22˚C by vapor diffusion using hanging drops with a 1:1 volume ratio of 6mg/ml protein to an equilibration buffer consisting of 0.1M sodium acetate pH 4.5 and 2M ammonium sulfate. For cryoprotection the crystals were transferred to the same buffer as that used for equilibration but supplemented with 25% v/v glycerol and were flash-cooled in liquid nitrogen.

VSG11$_{N2C}$-18mer crystals were grown at 22˚C by vapor diffusion using hanging drops with a 1:1 volume ratio of 6 mg/ml protein to equilibration buffer containing 19% (w/v) PEG 2000MME, 0.2 M NaCL 0.1 M MES pH 6.0. For cryoprotection the crystals were transferred to the same buffer as used for the equilibration with 25% (v/v) glycerol and were flash-cooled in liquid nitrogen.

Native VSG11$_{WT}$ datasets were collected at a wavelength of 1.0 Å at the Paul Scherrer Institut Villingen. For phasing, an iodine soaked crystal was collected at 1.54 Å on a Rigaku X-ray generator and DECTRIS PILATUS3 R detector. The structure was solved by single wavelength anomalous diffraction (SAD) using SHELX [32] and HKL3000 suite [33]. The initial model was built using Arp/wARP [34] with PHENIX [35], COOT [36] and PDB_REDO [37] for model optimization and refinement. That model was placed in a high-resolution VSG11$_{WT}$-Iodine dataset (collected at 1Å at the SLS synchrotron–see S1 Table) by molecular replacement with PHASER [38] and the model optimized and refined with PHENIX-REFINE [35] and with cycles of manual model building using COOT [36]. The structures of VSG11$_{WT}$-Oil, VSG11$_{WT}$-AS, and VSG11$_{N2C}$-18mer were solved by molecular replacement using the refined VSG11$_{WT}$-Iodine model with the PHASER package [38] of PHENIX, and the models optimized and refined with PHENIX-REFINE [35] and with cycles of manual model building using COOT [36]. The structure of VSG11$_{N2C}$-18mer is characterized by a high Wilson B factor (66.92Å^2) and disorder in many regions of the model. Final statistics are shown in S1 Table.

Structural determination of VSG615

T. brucei brucei expressing VSG615 was obtained as a kind gift of Dr. Hee-Sook Kim [39] and were cultured at 37˚C and 5% CO$_2$ in HMI-9 media (PAN Biotech) supplemented with 10% fetal calf serum (Gibco), L-cysteine and ß-mercaptoethanol. The cells from 4L culture were pelleted and the VSG615 protein purified through modifications of previously published protocols [31]. The cells were lysed with 40 ml 0.2 mM ZnCl$_2$, after centrifugation (10.000 g, 10 min) the pellet was resuspended in 30 ml 20 mM HEPES/NaOH pH = 7.5, 150 mM NaCl (42˚C) and centrifuged (10.000 g, 10 min). The supernatant was passed through a 25 ml Q-sepharose Fast-flow column (GE Healthcare) equilibrated with 20 mM HEPES/NaOH pH = 7.5, 150 mM NaCl. The flow-through was collected and concentrated to final concentration 1 mg/ml.

To generate the NTD of VSG615, the concentrated protein was subjected to limited proteolytic digestion using trypsin. The VSG at the concentration of 1 mg/ml was mixed with trypsin (5 mg/ml) (Sigma Aldrich) at 1:50 trypsin:VSG ratio and incubated for 3 hours on ice. The reaction was terminated by adding PMSF to 1 mM final concentration. The protein was further purified (separating the NTD from the CTD) by size exclusion chromatography on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) equilibrated in 20 mM HEPES/NaOH pH = 7.5, 150 mM NaCl. The fractions containing the NTD of VSG615 from size

exclusion chromatography were concentrated to 10 mg/ml in 500 µl of final volume. The lysine residues on the protein were subsequently methylated by reductive alkylation [40]. 10 µl of 1M borane dimethylamine complex (DMAB) and 20 µl of 1M formamide into the protein solution and mixed gently. The mixture was incubated for 2 hours in the dark at 4°C with rotation and the entire process repeated. Prior to overnight incubation, 5 µl of 1M DMAB was added. To stop the reaction, 1M Tris pH 7.5 was added to bring the reaction to a final volume of 1 ml. The buffer was exchanged to 20 mM HEPES pH 7.5, 150 mM NaCl by size exclusion chromatography with Superdex 200 10/300 GL column. The fractions containing the methylated protein was further concentrated to 10 mg/ml for crystallization.

The methylated VSG615 protein was crystallized at 22°C by vapour diffusion using hanging drops formed from mixing a 1:1 volume ratio of the protein with an equilibration buffer consisting of 23% (w/v) PEG 4000, 100 mM sodium cacodylate pH = 6.0, 10 mM $ZnCl_2$. For data collection, crystals were soaked in the same buffer augmented to 20% glycerol, flash-cooled in liquid nitrogen. A native dataset was collected at the Paul Scherrer Institut, Villingen. The structure was solved with molecular replacement using a model of a VSG615 trimer predicted by AlphaFold using the PHENIX package (loop regions with a pLDDT below 50 were removed for the search model). The structure is characterized by a very high Wilson B factor ($85.19Å^2$), high atomic temperature factors, and correspondingly high disorder in many regions of the model, although the electron density for the O-linked sugars (a principle reason for pursuing the VSG615 structure) was clear. Final model statistics are shown in S1 Table.

Structural determination of VSG558

VSG558 was amplified from genomic DNA of *T. brucei brucei* strain Lister 427 VSG2 expressing cells. The plasmid used to generate VSG558-expressing trypanosomes was a modification of previous plasmids designed for integration into the trypanosome genome [30]. The plasmid was first linearized by EcoRV (New England Biolabs), and then transfected into VSG2-expressing cells (2T1): 10ug of plasmid was mixed with 100ul of $4x10^7$ cells in Tb-BSF buffer (90mM Na_2HPO_4, pH 7.3, 5mM KCl, 0.15mM $CaCl_2$, 50mM HEPES, pH7.3), using an AMAXA nucleofector (Lonza) program X-001, as previously described [29]. After 6h hygromycin was added to a concentration of 25 ug/ml and single-cell clones were obtained by serial dilutions in 24-well plates and collected after 5 days.

Clones were initially screened with flow cytometry for VSG2 loss of expression using a monoclonal $VSG2_{WT}$ antibody [30]. To determine the binding of antisera to live trypanosomes, $2 x 10^6$ parasites were collected and incubated with 200ul FITC conjugated $VSG2_{WT}$ antisera (FITC conjugation kit, Abcam ab102884) (1:200) in cold HMI-9 without FBS for 10 min on ice. Cells were washed twice with cold HMI-9, resuspended in 200µl cold HMI-9 and immediately analyzed with a Guava EasyCyte 4HT Flow Cytometer (Luminex). For a second step, negative clones were sequenced by isolating RNA using the RNeasy Mini Kit with on-column DNAse digestion (Qiagen) and cDNA synthesis with Superscript IV first strand synthesis system (Thermo Fisher). The sequences were then amplified, using Phusion High-Fidelity DNA Polymerase (New England Biolabs), a forward primer binding to the spliced leader sequence and a reverse binding to the VSG 3′ untranslated region. The final products were purified by gel extraction from a 1% gel with the NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel) and verified by Sanger sequencing.

VSG558 was expressed and purified from trypanosomes in the same manner as VSG11. Purified full-length VSG558 was mixed with trypsin at 1:50 trypsin:VSG ratio and incubated 1 hour on ice and the NTD domain was isolated by size exclusion chromatography on a Superdex 200 10/300 GL column equilibrated in 10 mM HEPES/NaOH pH = 8.0, 150 mM NaCl.

After purification the VSG558 NTD was concentrated to 5 mg/ml and crystallized at 22˚C by vapour diffusion using hanging drops formed from mixing a 1:1 volume ratio of the protein with an equilibration buffer consisting of 100 mM Citric acid/NAOH pH = 5.5 and 17.5% PEG 3350. The crystals appeared after one week and the same condition supplemented with 25% PEG 400 or 25% MPD was used as a cryoprotectant for crystals flash-cooled in liquid nitrogen. The crystals diffracted to 1.74 Å and were collected at a wavelength of 1.0 Å at the Paul Scherrer Institut Villingen. The structure was solved by molecular replacement using a model of a dimer predicted with Alphafold [27], followed by model optimization and refinement using PHENIX and COOT. Final model statistics are shown in S1 Table.

Structural determination of VSG21

A VSG21 stably-expressing *T.b. brucei* strain was obtained as a kind gift of Dr. Hee-Sook Kim [39]. VSG21 was expressed and purified from trypanosomes according to the same protocol as VSG11. Full-length VSG21 at the concentration of 12.3 mg/ml was crystallized at 22˚C by vapour diffusion using sitting drops formed from mixing a 1:1 volume ratio of the protein with an equilibration buffer consisting of 100 mM Tris/HCl pH = 7.0, 200 mM CaCl2 and 20% (W/V) PEG 3335. For data collection, crystals were soaked in 100 mM Tris/HCl pH = 7.0, 25% (V/V) ethylene glycol and 20% (W/V) PEG 3335 and flash-cooled in liquid nitrogen. A native dataset was collected at the Paul Scherrer Institut, Villingen. The structure was solved with molecular replacement using a model of a VSG21 dimer predicted by AlphaFold using the PHENIX package, followed by model optimization and refinement using PHENIX and COOT. Final model statistics are shown in S1 Table. VSG21 was found to be truncated, with the far C-terminal sequence corresponding to the lower lobe removed.

Structural determination of VSG545

VSG545 cloning and verification were performed in the same manner as VSG558 (except that for selection a hygromycin concentration of 5μg/ml was used). For purification of VSG545, cells from 2.4 L culture were pelleted and lysed with 20 ml 0.4 mM ZnCl2, containing a protease inhibitor cocktail (Roche cOmplete), after centrifugation (10.000 g, 10 min) the pellet was resuspended in 20 ml 10 mM Sodium Phosphate buffer, pH8 containing protease inhibitors (42˚C) and centrifuged (10.000 g, 10 min). The supernatant was passed through a 20 ml Q-sepharose Fast-flow column (Cytiva) equilibrated with 10 mM Sodium Phosphate buffer, pH8. The flow through was collected and concentrated. To remove protease inhibitors, the protein was purified on a Superdex 200 Increase 10/300 GL column (GE Healthcare) equilibrated in 10 mM HEPES/NaOH pH = 8.0, 150 mM NaCl. The fractions containing the VSG protein were pooled and concentrated.

To generate N-terminal domain, the concentrated protein was subjected to limited proteolytic digestion. VSG545 at the concentration of 2 mg/ml was mixed with endoproteinase LysC (New England Biolabs) at 1:1200 LysC:VSG ratio and incubated for 1 hour at 37˚C. The reaction was terminated by adding TLCK to 50μg/ml final concentration. The protein was further purified (separating the NTD from the CTD) by size exclusion chromatography.

In order to obtain well-diffracting crystals, N-linked oligosaccharides were removed by treatment of the purified NTD with PNGase F. VSG545 NTD was mixed with PNGase F (New England Biolabs, P0704S) to final concentrations of 0.4 mg/ml and 15,000 U/ml respectively, incubated for 3 hours at 37˚C and purified again by size exclusion chromatography.

The protein was concentrated to 2 mg/ml and was crystallized at 22˚C by vapour diffusion using sitting drops formed from mixing a 1:1 volume ratio of the protein with an equilibration buffer consisting of 100 mM MES pH = 6.5, 22% PEG 8000 and 300mM Li_2SO_4. Crystals

appeared after about 8 weeks and the same condition supplemented with 25% PEG 400 was used as a cryoprotectant for crystals flash-cooled in liquid nitrogen.

The VSG545 data set was highly anisotropic and the Staraniso server [41] was used to determine the anisotropic diffraction cutoff and the output, modified data file was used to solve the structure. A dataset was collected at a wavelength of 1.0 Å at the Paul Scherrer Institut Villingen. The structure was solved by molecular replacement using a model of a dimer predicted with Alphafold [27]. After model building and refinement using PHENIX and COOT, the final model contains a truncated dimer in the asymmetric unit likely produced by further proteolysis in the crystallization drop. Final model statistics are shown in S1 Table. Like with VSG21, VSG545 was found to be truncated, with the far C-terminal sequence corresponding to the lower lobe removed.

Protein structural prediction with AlphaFold (ColabFold)

The *Trypanosoma Brucei Brucei* Lister 427 VSG sequences were obtained from a publically accessible database at the Rockefeller University (https://tryps.rockefeller.edu/). Structural prediction was performed only for the N-terminal domain sequences (spanning approximately 350 amino-acids, depending on the VSG). As the signal peptide is not present in the mature VSG protein, the signal sequence predicted by SignalP 6.0 (https://services.healthtech.dtu.dk/services/SignalP-6.0/) was removed. The end of the helix that connects the N-terminal bottom lobe with C-terminal domain terminates the N-terminal domain.

For the structure predictions, LocalColabFold [42] version 1.3.0 was installed on a local computer (using https://github.com/YoshitakaMo/localcolabfold). The LocalColabFold command line interface was used with arguments to specify an input FASTA file, an output directory, using the default options for structure predictions. For monomer prediction "colabfold_batch—amber—templates—use-gpu-relax—num-recycle 3 input.fasta output_directory" was used. For the oligomer prediction the input FASTA file contained the VSG NTD sequence 2 x for class A and 3 x for class B separated by a colon, and adding—model-type AlphaFold2-multimer-v2 to the colabfold_batch command line.

The multiple sequence alignment (MSA) was generated for each sequence by the MMseq2 web server implemented in Colabfold, using the Uniref, PDB70 and Environmental sequence database [42]. The server aligns the input sequence against the database and prepares the input files for the structure prediction that ran locally on our Nvidia P-4000 GPU. Improvement of the structural models was achieved through recycling three times (the software default). In the final step in Colabfold a relaxation/energy minimization is performed using AMBER (Assisted Model Building with Energy Refinement) [42]. At the end, Colabfold created 5 predictions for each sequence, ranked by pLDDT for a monomer and TM-score for the oligomers.

Model quality was assessed quantitively using calculations of root-mean-square-deviation (RMSD) and global distance test total score (GDT_TS) [43] as calculated by the AS2TS system (http://as2ts.proteinmodel.org/) [44]. Parameters for the structural alignments and scoring were taken from default suggestions to match those used for the scoring in CASP (https://proteopedia.org/wiki/index.php/Calculating_GDT_TS), where for the comparison only one chain of an oligomer was used.

Results

Additional experimental structures to map VSG diversity

Three novel VSG structures published between 2018 and 2021—VSG3, VSG13, and VSGsur [17,18]—have demonstrated that there is far more diversity in the VSG coat proteins in fold, oligomerization, topology, and post-translational modification than was appreciated. To

further map the diversity space of the VSGnome, we have undertaken X-ray crystallographic studies of numerous VSGs. Guided by recent bioinformatic analyses [25,26], we sought to examine VSGs from different clustering classes. In a separate paper [16], we determined structures of three metacyclic VSGs: VSG397, VSG531, and VSG1954, representing three clustering classes. In this paper we present five new experimental structures from *T. brucei brucei* strain Lister 427: VSG11, VSG21, VSG545, VSG558, and VSG615.

VSG11 and VSG615 are related to VSG3 (the topologically-similar class of monomer/trimers with *O*-linked glycan PTMs), denoted as class B [25] or N4 [26] in the literature. These were examined to confirm the fold and oligomerization of this class as well as the presence of *O*-linked glycosylation. Previously determined structures of VSG3 [17] and VSG1954 [16] showed that both possess highly similar trimers in the crystals. Further confirmation of the trimeric oligomerization state of members of this class came from a biochemical analysis of several VSGs that showed evidence of concentration dependent monomeric and trimeric forms of VSG9 in solution [24]. However, unlike VSG3, the structure of VSG1954 showed no evidence of *O*-linked glycans, despite possessing a conserved glycosylation sequence [16,17]. Whether this absence is due to unknown regulatory processes yet to be identified, to the sugar being labile and removed during purification/crystallization as has been observed for VSG3 [45], or to a difference between metacyclic and bloodstream VSGs, was unclear. Therefore, pursuing additional studies of related VSGs could aid in determining how many VSGs are so modified.

NTD structures of VSG11 and VSG615 were solved to 1.23Å and 3.2Å resolution, respectively (Methods, Figs 1A and 1B and S1 and S1 Table). Both reveal a fold similar to VSG3 (Fig 1C) as well as the presence of *O*-linked glycans on the top surface of the protein–one observed on VSG11 and two on VSG615 (Fig 1A and 1B). This crystal form of VSG11 (VSG11$_{WT}$-Oil, see Methods and S1 Table) crystallizes as a monomer in the asymmetric unit of the crystal, but forms a trimer along a crystallographic three-fold axis of symmetry that is nearly identical to the crystallographic trimers seen with VSG3 and VSG1954 (Fig 2A). Interestingly, VSG615 crystallized with two trimers in the asymmetric unit, and these trimers align well with those of VSG3, VSG11, and VSG1954 (Fig 2A), providing additional support that class B trimer formation is not a crystallization artifact, but could represent the relevant, biological assembly.

An additional crystal form of VSG11 shows non-crystallographic trimerization in the asymmetric unit, while another form also raises the possibility of an inherent flexibility that could be present in the VSG proteins. In a second crystal form (VSG11$_{N2C}$-18mer, determined to 2.6 Å resolution, Methods and S1 Table), the asymmetric unit contains six trimers of the same arrangement as observed previously for all class B VSGs, for a total of 18 independent molecules (hence "18mer", Fig 2B). In addition to helping establish that this specific trimeric arrangement is the preferred oligomerization of this class (at least at higher concentrations), the independent monomers in this asymmetric unit display a range of alternative conformations, particularly in the bottom lobe of the protein and even in some limited regions of the three-helix bundle (Fig 3A). An overall structural alignment of the monomers to themselves gives an average RMSD of 1.33 Å (RaptorX), a somewhat high value for aligning a protein to itself. Examination of the set of alignments from the monomers of the 18mer show that it appears that the divergence primarily occurs in the lower lobe (Fig 3A).

The third crystal form (VSG11$_{WT}$-AS grown in ammonium sulfate and solved to 1.75Å resolution, Methods, S1 Table) contains two molecules in the asymmetric unit (S2A Fig). However, these two monomers are not arranged in a standard class A type dimeric arrangement in which the three-helix bundles, top, and bottom lobes align alongside each other, but are instead "tail-to-tail" with the lower lobe of one monomer interacting with the lower lobe of the other (S2A Fig). In contrast, the crystal packing is characterized by a three-fold trimer identical
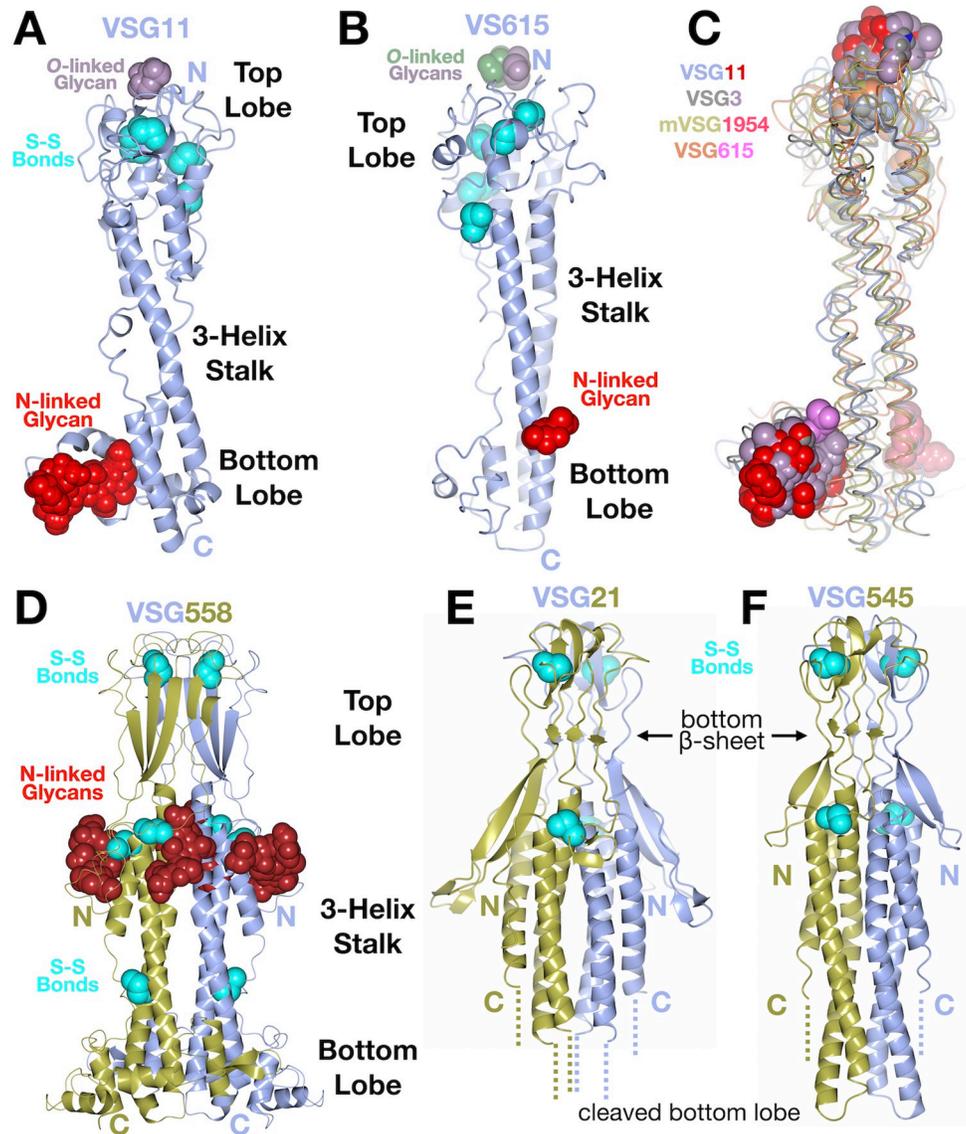
Fig 1. Crystal Structures of Five VSGs from Trypanosome brucei. (A) and (B) Structures of the VSG11 and VSG615 NTD monomers shown as a ribbon diagram colored light blue. The N-linked glycans attached to the bottom lobe are shown in a space-filling representation colored red. Disulfide bonds (labeled S-S bonds) are shown in a space-filling representation, colored cyan. The *O*-linked glycans on the top of the molecules are shown in a space-filling representation colored purple and green (for the second sugar in VSG615). (C) Alignment of four B-class VSG monomers (VSG3, VSG11, VSG615, and VSG1954 [16]) shown as thin "worm" drawings with each separate protein colored as indicated in the figure labels. The N- and *O*-linked sugars at the bottom and top of the molecules (respectively) are differentially colored as noted in the figure. (D) Structure of the VSG558 dimer illustrated as in panel (A), which each chain of the dimer colored differently. (E and F) Structures of the VSG21 and VSG545 dimers illustrated as in panel (A), the disordered regions at the C-terminal portion (e.g., missing bottom lobe) denoted by dotted lines. Illustrations were produced with CCP4mg [57].

https://doi.org/10.1371/journal.pntd.0011621.g001

to those seen in all the class B protein structures solved to date, including the alternative crystal forms of VSG11 (S2B Fig). Therefore, the assembly of VSG11$_{WT}$-AS is most accurately described as trimeric and not dimeric (and certainly not dimers in the sense of those seen in class A VSGs), providing an interesting example in which the contents of the asymmetric unit of the crystal do not present the likely biological assembly, whereas the crystallographic
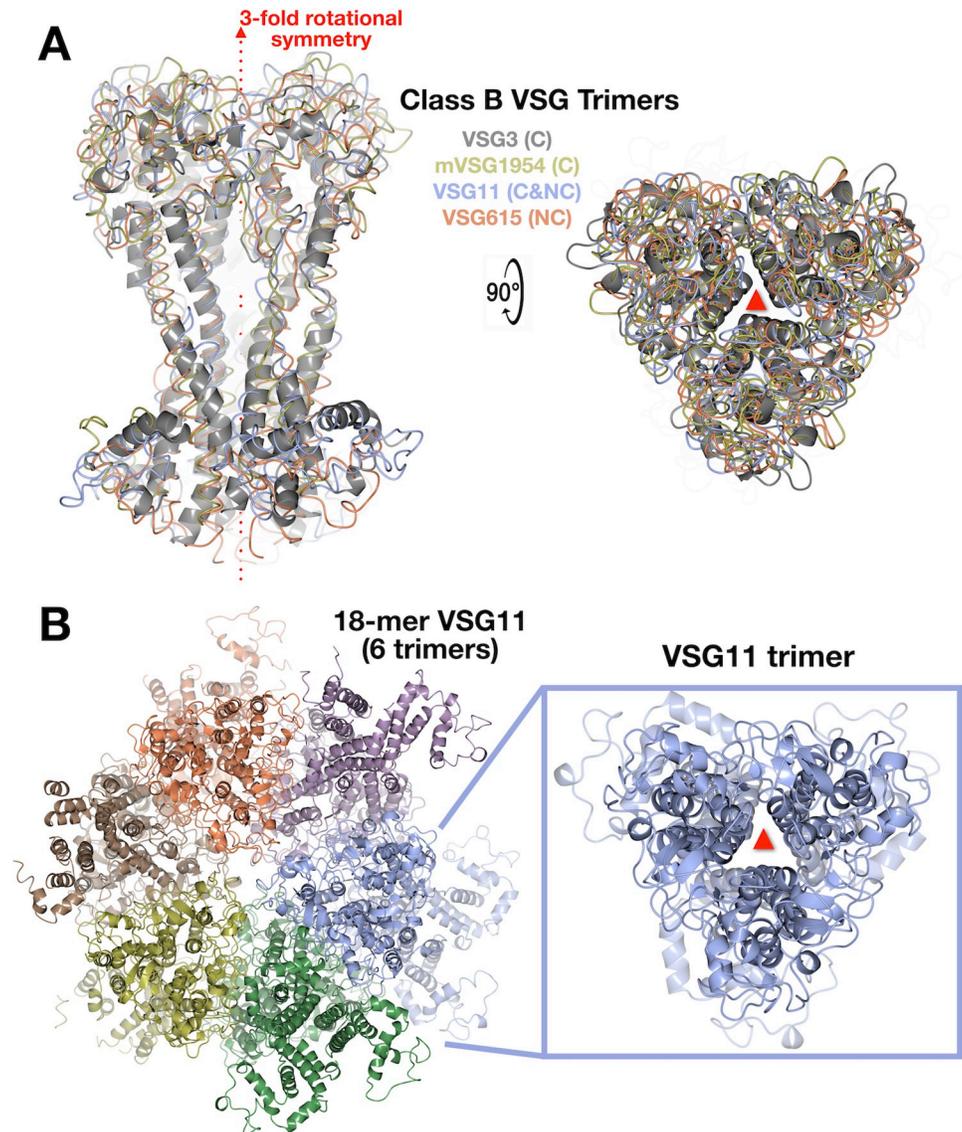
Fig 2. Class-B VSG Trimers. (A) Both crystallographic (VSG3 and VSG1954) and non-crystallographic (VSG11 and VSG615) trimeric arrangements aligned and illustrated as ribbon (gray, VSG3) or worm tracing of the mainchain (other VSGs). Two orientations are shown rotated 90 degrees. The three-fold axis of symmetry is shown in red. "C" indicated a crystallography three-fold symmetry for the trimer and "NC" indicates a non-crystallographic three-fold symmetric trimer in the asymmetric unit. **(B)** The full asymmetric unit of the VSG$_{WT}$-18mer is shown on the left with the proteins drawn as ribbon diagrams and each set of trimers colored differently. On the right is a single trimer isolated from the other six. Illustrations produced using CCP4mg.

https://doi.org/10.1371/journal.pntd.0011621.g002

arrangement does. Intriguingly, these two independent monomers are characterized by a significant conformational change in bottom lobe of the protein. This change consists of a 12-degree rotation of the bottom lobe from one monomer to the other (Fig 3B).

In addition, alignments of the VSG11$_{WT}$-Iodine crystal form with the VSG11$_{WT}$-Oil form (Methods and S1 Table) show that the second helix of the three-helix bundle displays a significant conformational change in the region near the middle of the bundle, the helix becoming disordered and forming a random coil in one monomer while keeping helical structure in the other (S2C Fig). Finally, when the different monomers of the 18mer are compared with the
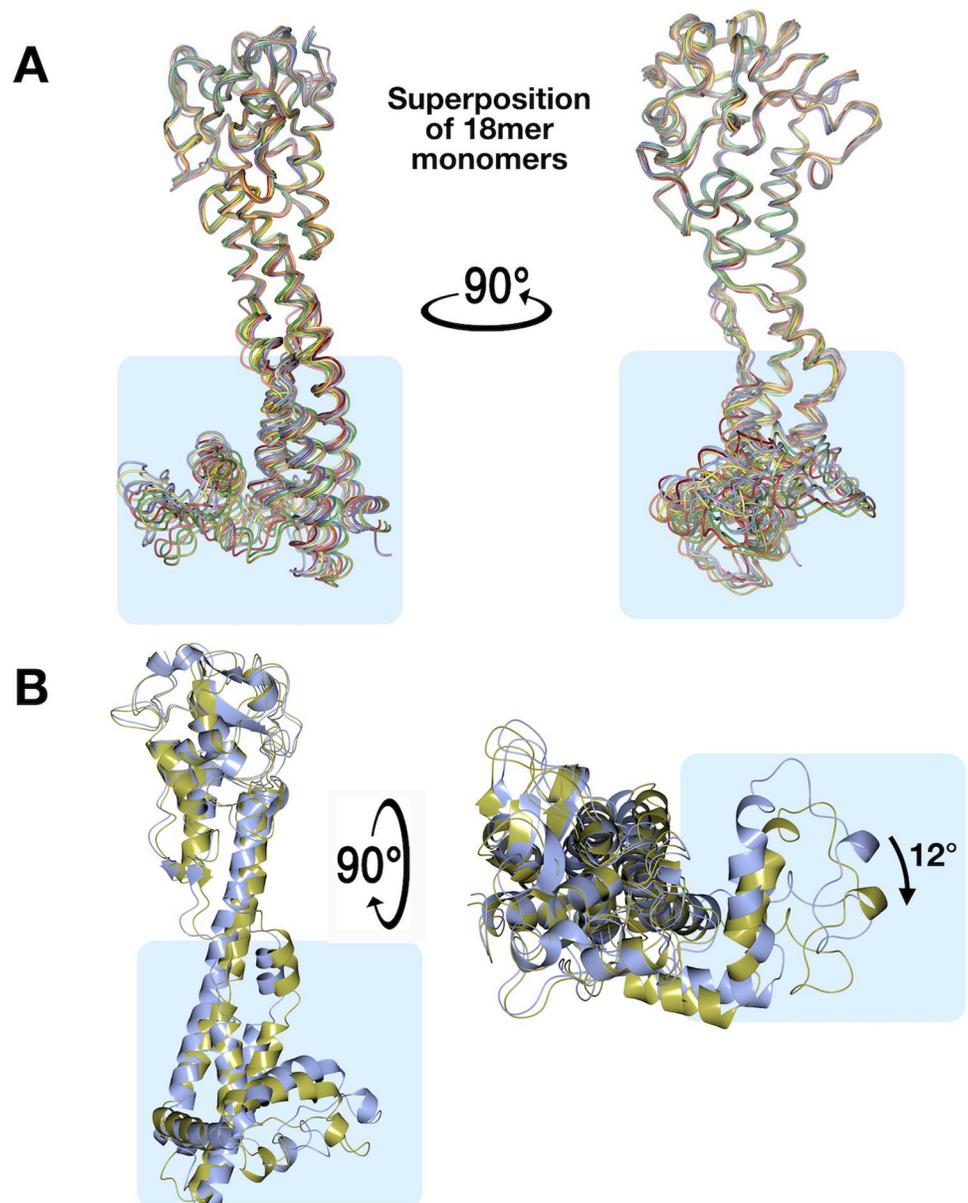
Fig 3. Lower Lobe Conformational Variability Observed in VSG11 Structures. (A) Alignment of the monomers in the asymmetric unit of VSG_{WT}-18mer, each monomer given a separate color, shown in two orientations related by a 90-degree rotation about the axis of the helical bundle. **(B)** Alignment of the two monomers in the asymmetric unit of the $VSG11_{WT}$-AS crystals. Two views are shown rotated by 90 degrees and the rotation of the bottom lobes relative to each other is indicated. Illustrations produced using CCP4mg.

https://doi.org/10.1371/journal.pntd.0011621.g003

other crystal forms described, a wide variety of conformational changes is seen, although these changes are limited to the bottom lobe and lower regions of the bundle (S3 Fig).

These alternative conformations, seen in several crystal forms where the protein is not highly constrained by the energetics of crystal packing (non-crystallographic symmetry), suggests that the VSGs on the membrane could potentially be characterized by a collection of conformers, each with molecular surfaces that differ in these regions. It is possible that such differing conformers could present distinct antigenic surfaces and impact the immunogenicity

of the VSGs, combining with other forms of variability such as surface protein features and post-translational modifications to augment the diversity space of the coat.

The structure of the VSG558 NTD was determined to 1.74 Å resolution (Fig 1D, Methods, S1A Fig, and S1 Table). It has a highly similar structure to VSG13 [18] with a 2.54 Å RMSD over 324 residues. This includes a broad and flat beta-sheet top-lobe that forms an intermolecular beta-sandwich in the dimeric assembly, a distribution of disulfides throughout the length of the rod-like structure, an N-terminus located in the "middle lobe" of the protein (instead of at the top, most membrane distal face of the VSG), and the presence of N-linked glycans directly below the top lobe. A notable difference with all the previously solved VSG NTD structures published is the presence of two N-linked glycans at the middle lobe of the VSG558 protein. To date, all VSG structures determined have been characterized by a single N-linked glycan (or none, as seen in the case of ILTat1.24). However, it has been established biochemically that some VSGs do possess more than one such sugar in the NTD, such as VSG5 [46]. The structure of VSG558 confirms the presence of multiple N-linked glycans for another VSG. This shows that not only are the locations of the N-linked glycans variable (found in both the bottom and middle lobes), but that the number of possible such carbohydrates is also variable.

The structures of VSG21 and VSG545 were also determined, both harboring closely related folds and biochemical properties that distinguish them from other VSGs (Fig 1E and 1F). Of note is the fact that both structures were obtained with proteolytic removal of the bottom lobe (contained in the C-terminal region of the NTD sequence). This cleavage occurred spontaneously during crystallization experiments performed with full-length VSG21 and the NTD of VSG545 that had been produced by limited proteolysis with Endoproteinase Lys-C (LysC) (see Methods). Presumably these cleavage events in the crystallization experiments occurred from residual contaminating endogenous protease(s) from *T. brucei* (and/or possibly residual LysC for VSG545). While such proteolysis has been seen before in the purification and crystallization of the VSG proteins [16,17,19,23], the cleavage typically occurs between the NTD and CTD of the protein in the linker sequence. In contrast, for VSG21 and VSG545, the cleavage in the crystallization drops produced a truncated NTD missing the bottom lobe sequence but maintaining the 3-helix bundle and top lobe of the protein. Superpositions with predicted models from AlphaFold (which were accurate enough to solve the experimental structures by molecular replacement) show possible conformations of the full NTD for each protein (S4 Fig). Finally, it should be noted that for VSG545, crystals were only obtained with PNGase treated material, removing the N-linked glycan (Methods). No sugar is seen in the VSG21 electron density, and no predicted site for N-linked glycosylation (Asn-Xaa-Ser/Thr) was found in the sequence of the NTD.

Both structures overall resemble the structures of VSG13, VSG397, VSG558, and VSGsur, possessing an N-terminus of the protein located approximately a third of the way down the long-axis of the VSG, a beta-sandwich fold for the top lobe, and disulfides both in the top lobe and spread elsewhere in the VSG. VSG21 and VSG545 are more similar to each other than these other VSGs, however, distinguished from them by the tightly twisted nature of the top lobe and, directly beneath it, the presence of a second three-stranded beta-sheet perpendicular to the axis of the 3-helix bundle (Fig 1E and 1F). VSG21 and VSG545 align well to each other in the top lobe but not to the other VSGs with beta-sheet top lobes (S5 Fig).

Overall classification of the VSGs based on structures

Altogether, as of the writing of this manuscript, there are fourteen published, experimentally determined protein structures of the VSG NTD from *Trypanosoma brucei* in hand: VSG1, VSG2, VSG3, VSG11, VSG13, VSG21, VSG397, VSG531, VSG545, VSG558, VSG615,
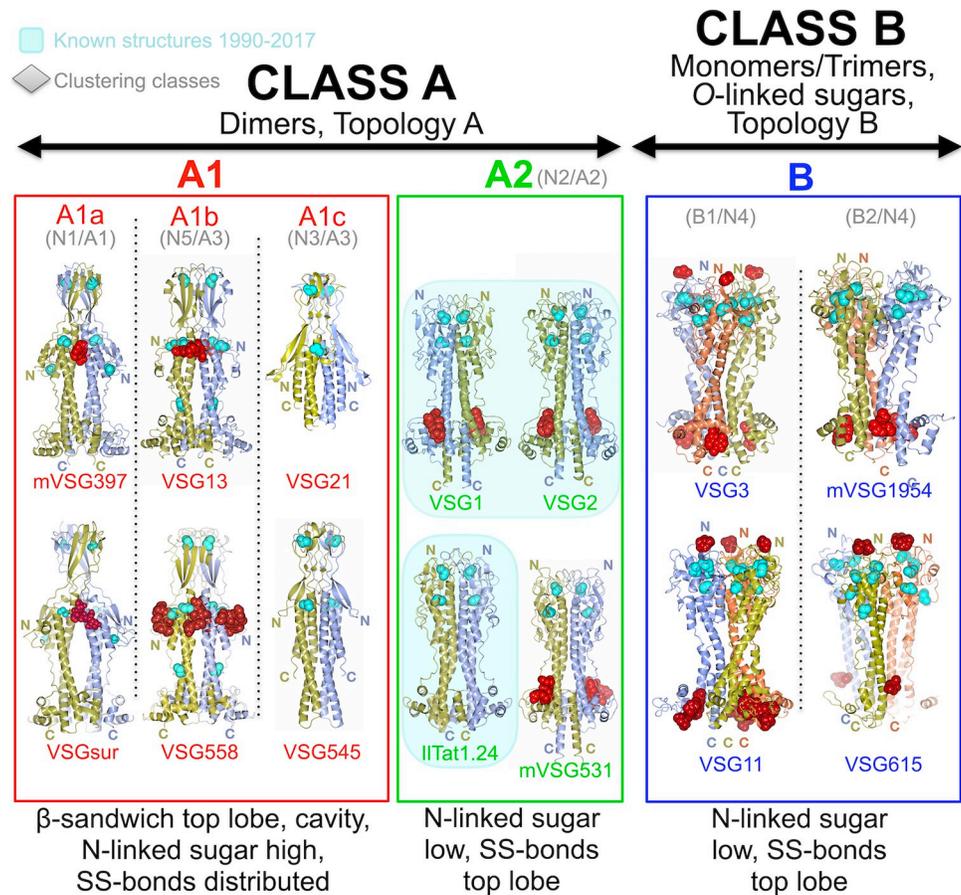
Fig 4. Structure-based VSG Classification Scheme. The two broad subclasses of VSG protein discussed in the main text are denoted at the top of the figure (class A and class B). Subclasses are shown in different colors with a collection of structures illustrating each class contained in a color-coordinated box matching the subclass name. Structures are drawn as in previous figures with the exception that *O*-linked glycans are colored red like N-linked glycans. Beneath each subclass name in gray are the clustering designations of the classes in previous papers discussed in the main text. Cyan backgrounds indicate structures solved prior to 2017. "N" and "C" refer to the N- and C-termini of the respective NTDs (colored by chain and placed nearby the terminal residue itself). As discussed in the text, VSG21 and VSG545 crystal structures are missing the bottom lobe sequence in the models due to truncation of the protein during crystallization.

VSG1954, VSGsur, and IlTat1.24. These fourteen structures not only map well to the bioinformatic clustering schemes published previously, but they better discriminate between them and put the classification schemes on an architectural foundation. We have sought to take these structures, with an eye to the bioinformatics work, and create what we conclude is a more explanatory organizational scheme for the VSG proteins (Fig 4). To provide continuity with the previous classification schemes of Hutchinson [15] and Cross [25], we have preserved the A/B designation for classes. However, this has required a change in the meaning of subdivisions within class A to reflect insight from the protein structures (discussed below). Finally, this schema was then tested by modeling hundreds of VSG proteins with the deep learning system AlphaFold and comparing the resultant folds to the classes we created.

To begin, there are two broad "super-families" of VSG structures based on the topological arrangement of the bottom lobe in the primary sequence (classes A and B, Figs 4 and S6). All VSG NTD structures determined have a bottom lobe subdomain structure (with the exceptions of VSG21 and VSG545 in this report that had this domain proteolytically removed). In

class A the bottom lobe residues are present at the C-terminal portion of the NTD sequence, directly following the final helix of the bundle. In contrast, in class B the bottom lobe residues are present at the N-terminal portion of the NTD sequence between the amino acids that form the first and second helices of the bundle (S6 Fig). Further cementing this broad division into two main super-families are two observations. The first is that all class A VSGs studied to-date are found as dimers in solution and in protein crystals (the latter either as non-crystallographic dimers in the asymmetric unit or as monomers in the asymmetric unit where a crystallographic two-fold symmetry produces the dimer, [18,21,45]), whereas all studied class B VSGs are characterized by the same trimeric arrangement in the crystals (either through crystallographic or non-crystallographic symmetry, with monomers or trimers in the asymmetric unit, respectively), with biochemical evidence of monomer to trimer transitions based on protein concentration [24]. The second fact is that many of the class B VSGs are post-translationally modified by *O*-linked carbohydrates, whereas none of the class A VSGs have been found so modified. When comparing with previous efforts to classify the VSGs, our class A would correspond to class "A" in Cross [25], classes N1-N3/N5 in Weirather [26], and the older class A of Carrington [47]. Class B would correspond in these sources to classes B, N4, and B, respectively.

Within our class A superfamily are two large structural subclasses, A1 and A2 (Fig 4). In class A2 are found all the VSG structures that were solved prior to 2018 (Iltat1.24, VSG1, and VSG2), illustrating that the set of conclusions about VSG structure and function accepted for over a quarter century were based on a very limited subset of VSGs containing highly related protein folds. This subclass is characterized by a top lobe that contains all the cysteine disulfides in the NTD, a top lobe fold that is a hodge-podge of alpha helices and beta-strands, and N-linked glycan chains located in the bottom lobe of the NTD. In sharp contrast, subclass A1 VSGs are significantly longer due to the presence of a large beta-sheet subdomain for the top lobe (forming a beta-sandwich in the dimer). Further distinguishing A1 from A2 is the distributed nature of the disulfide bonds (present throughout the length of the VSG in A1), the presence of a "middle lobe" of secondary structure straddling the beginning of the three-helix bundle, and the location of the N-linked sugar(s) just below the beta-sandwich top lobe. This dramatically different arrangement in structural elements is reflected by the differing positions of the N-terminus of the NTD, namely toward the middle of the VSG fold in A1 but located at the very top of the fold in A2.

Additionally, the folds of the A1 VSGs from experimental and predicted structures (see below) suggest that this subclass can be further subdivided into three groups based on the size, conformation, and twist of the top lobe beta-sheet and the width of the space between the three-helix bundles in the dimer. These subclasses of A1 were in previous sequence clustering classification systems denoted as the separate classes A1 and A3[25] and N1, N3, and N5 [26]. However, all these subclasses are structurally similar to each other in the manners described above while differing markedly from the A2/N2 classes. Therefore, we considered it better to combine A1/A3 and N1/N3/N5 into a single class, A1 (a family contrasted to A2), and then subdivide them within A1: A1a, A1b, A1c (Fig 4). In this subdivision, we split the A3 class from Cross into A1b and A1c (which correspond to classes N5 and N3 in Weirather, respectively) based on differences in the top lobe architecture (e.g., the smaller top lobe beta-sandwich in A1c compared to A1b and the presence of a second beta-sheet over the middle lobe in A1c that is not present in A1b).

Finally, Cross et al. [25] divide the class B VSGs into two subgroups, whereas Weirather et al. [26] do not subdivide their equivalent class, N4. In contrast to the marked structural divergences between classes A1 and A2, and even the differences within the distinct subclasses of A1, we find no broad structural differences within the class B VSGs (examining features

such as the protein fold, disulfides, N- or O-linked glycans, or oligomerization). However, three more subtle differences between members of the B class can be used to divide them into two subgroups. For example, two helical regions differ between the B1 and B2 subgroups. One helix that exists in the top lobe of the B2 class is not generally present in the B1 class (S7A Fig). Secondly, one of the bundle helices in class B1 is disordered in places relative to the same helix in many class B2 members (S7B Fig). Thirdly, in class B2 several VSGs possess an NTD with more amino acids. In the AlphaFold modelling discussed below, these additional residues are predicted to form both an extended loop in the disordered region of the helix and also longer loops in the top lobe (S7A Fig). Buttressing this subdivision of class B, when multiple B1 and B2 VSGs (from experimental and AlphaFold predicted models) are analyzed by structure (using the Dali Server [48]), these are divided into two classes consistent with the B1 and B2 groupings produced by sequence analysis (see the structural dendogram in S7C Fig). We have therefore denoted a split in Fig 4 between the B1 and B2 classes.

Mapping the VSGnome with AlphaFold

We sought to utilize the structural prediction system of AlphaFold [27,28,42] to quickly examine the sequences of hundreds of VSG proteins and compare the structural predictions both to experimentally determined VSGs as well as assess the fit of our classification scheme to the VSGnome as a whole. In the work that follows, all our predictive modeling was performed with AlphaFold2 [27,28,42] (see Methods), although we will simply refer to it as "AlphaFold". Models were assessed by two quantitative metrics: (1) the root mean square deviation in the Cα residues of the protein main chain between experimental models and AlphaFold predictions (measured in Å with a lower value indicating a more accurate model) and (2) the global distance test total score (GDT_TS, scored as a % from 0–100, with higher score indicating a more accurate model) that is also computed over the Cα positions between model and experimental structure but which is less sensitive to small, localized discrepancies that occur in flexible/disordered loops [43]. The GDT_TS score is a primary metric for evaluating protein structural accuracy in the Critical Assessment of Structure Prediction (CASP), a global competition between protein structure prediction algorithms [49]. In the CASP14 competition, AlphaFold2 performed better than other submitted software systems in fold prediction and is now widely utilized and available through online servers and as open source software. Because RMSD values are measured in Å, values below 2Å are approaching the length of a carbon-carbon bond (1.54Å). Models with a GDT_TS value of 80% or higher are considered to have modeled most of the protein accurately [50], and above 90% are considered to be as accurate as experimental structures [51]. Therefore models with a score of ~80% constitute an accuracy sufficient for our efforts in classifying a given VSG with respect to our scheme discussed above (Fig 4).

We began by assessing the accuracy of AlphaFold against known VSG structures that existed in the PDB database at the time of writing this manuscript (VSG1, VSG2, VSG3, IlTat1.24, VSG13, and VSGsur), and which would therefore be accessible to AlphaFold as structural templates for training the deep learning neural networks. In addition, as the VSGs are multimers (dimers and trimers) and AlphaFold can predict both isolated folds and folds as oligomers [42], we generated models for monomer VSG sequences as well as for the appropriate oligomer for a given VSG sequence. Using ColabFold (Methods) and an input sequence of these VSGs, we examined five ranked models output by the software from which the rank 1 model (highest pLDDT score, see Methods) was compared with structures in the PDB database. All VSG structures predicted by AlphaFold were of high accuracy (Fig 5A), with oligomer predictions showing better values compared to monomer predictions (structures were

predicted for the monomer and the oligomer using dimers for class A and trimers for class B VSG sequences, respectively). The predicted local distance test (pLDDT) of model confidence as well as the comparisons of model to experimental structures (as scored by the GDT_TS and RMSD) all showed statistically significant improvements in the predictions for oligomeric assemblies compared to monomers (S8A–S8C Fig).

Next, we repeated this analysis against structures not yet available in the protein database at the time of the analysis, namely VSG397, VSG531, and VSG1954 [16] and VSG11 and VSG615 from this manuscript. These recently determined structures were well-predicted by AlphaFold, although with slightly higher RMSD and lower GDT_TS values as compared to the predictions for VSG structures already in the database (Fig 5B). Frequently in these models, the three-helix bundle region showed the most accurate alignment, while discrepancies appeared mostly in loop regions connecting secondary structure. The predicted location of disulfide bonds matches very closely those observed in the experimental structures (e.g., S8D Fig). Finally, as noted in the Methods, the structures of VSG21, VSG545, VSG558, and VSG615 were solved by molecular replacement using an AlphaFold model of these VSGs as search models, indicating a very close fit of the model to the experimental data (a general idea for X-ray phasing implemented by several groups due to the accuracy of AlphaFold models
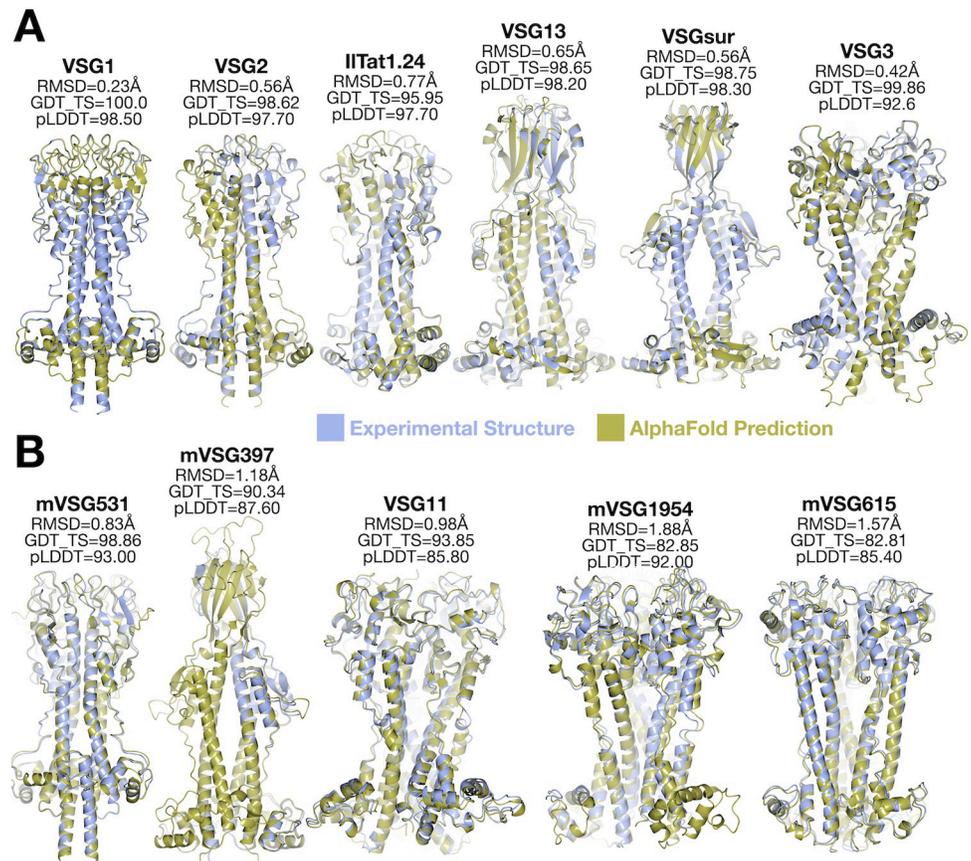
Fig 5. Comparison of Experimental VSG Structures with AlphaFold Models. **(A)** Alignment of experimentally determined structures (blue) and AlphaFold models (gold) for structures whose coordinates were available at the time of the writing of this manuscript for use by AlphaFold as a template. **(B)** Alignment of experimentally determined structure (blue) and AlphaFold model (gold) for structures whose coordinates were not yet available at the time of the writing of this manuscript for use by AlphaFold as a template. RMSD, GDT-TS, and pLDDT are defined in the main text and Methods. Illustrations of alignments produced by CCP4mg.

https://doi.org/10.1371/journal.pntd.0011621.g005

[52–55]). These robust results from AlphaFold provide some confidence that it can be used as a tool to rapidly and robustly map the structural diversity of the VSGnome.

Focusing on the Lister427 strain database (Methods), we examined 215 NTD sequences with AlphaFold and compared these predictions to the existing structures known. All of the predicted structures conform to our classification scheme and no structure was predicted that conflicted with our schema (Fig 6). We also examined 83 sequences from a second *Trypanosoma brucei* strain, *T. brucei brucei* 927. All VSGs examined from Tb 927 were predicted to adopt structures fitting our classification scheme (S9 Fig). Extending our analysis to more distally related species and VSGs like *T. vivax* showed less confidence in the modeling (S10 Fig), and therefore it is likely that template experimental structures (which currently do not exist for this species) may be required on which the AlphaFold system can train to produce more reliable models.

From this we hypothesize that the space of protein structure for the *T. brucei* VSGs is well-described by the classes we have presented in this manuscript, although it is quite possible that some variants will be discovered outside this scheme. As an overarching observation, both the experimental structures and the predicted AlphaFold models all display unique molecular surfaces in terms of the distribution of chemical properties exposed to the immune system (charge, hydrophobicity, topography, etc., S11 Fig). This is of course consistent with the notion that the VSGs are each antigenically distinct, presenting unique faces to the immune system.

Discussion

A mechanistic understanding of antigenic variation in the African trypanosome will require a complicated mixture of knowledge regarding the epitope space exposed to the host immune system and how the repertoires of antigen recognition in immune cells respond to this chemical space in the dynamic process of coat switching and the adaptive immune response [45]. At the most basic level, such an understanding necessitates a thorough knowledge of VSG antigenic diversity. Therefore, understanding the possibilities and constraints of VSG protein architecture is a critical step in this direction.

We have presented in this manuscript and in recently published work a collection of new experimental VSG structures. Combined with those already known, we have developed a structure-based classification scheme for the *T. brucei* VSGs that shows the breadth of possible protein folds for the major coat protein of the African trypanosome as well as constraining the underlying conformational possibilities inherent in the VSGnome. This classification scheme is enriched by including hundreds of VSG structural models by predicting folds using the deep learning system, AlphaFold. Our approach was first validated on the VSG proteins by predicting the folds of experimentally determined VSGs in the database as well as several novel VSG structures we have determined that had not been published (and thus could not serve as templates for AlphaFold). AlphaFold predicted all with high accuracy, always performing better (as determined by match to the experimental structures) when asked to predict an oligomer of the right number for a given VSG (dimer or trimer). Since protein folding can be partially dependent on higher-order protein-protein associations, this suggests that the deep learning training of the AlphaFold algorithms may have captured some aspect of this dependency. Once we were confident in the predictions of AlphaFold for the VSG proteins, we analyzed hundreds of sequences from *T. brucei*, establishing that all predicted structures fall into the groups of our proposed classification system.

This provides an architectural framework to understand the VSGs, integrating the key structural and biochemical features of each class (protein fold and topology, oligomerization, positions of disulfides and N-linked glycans, and the presence of surface *O*-linked
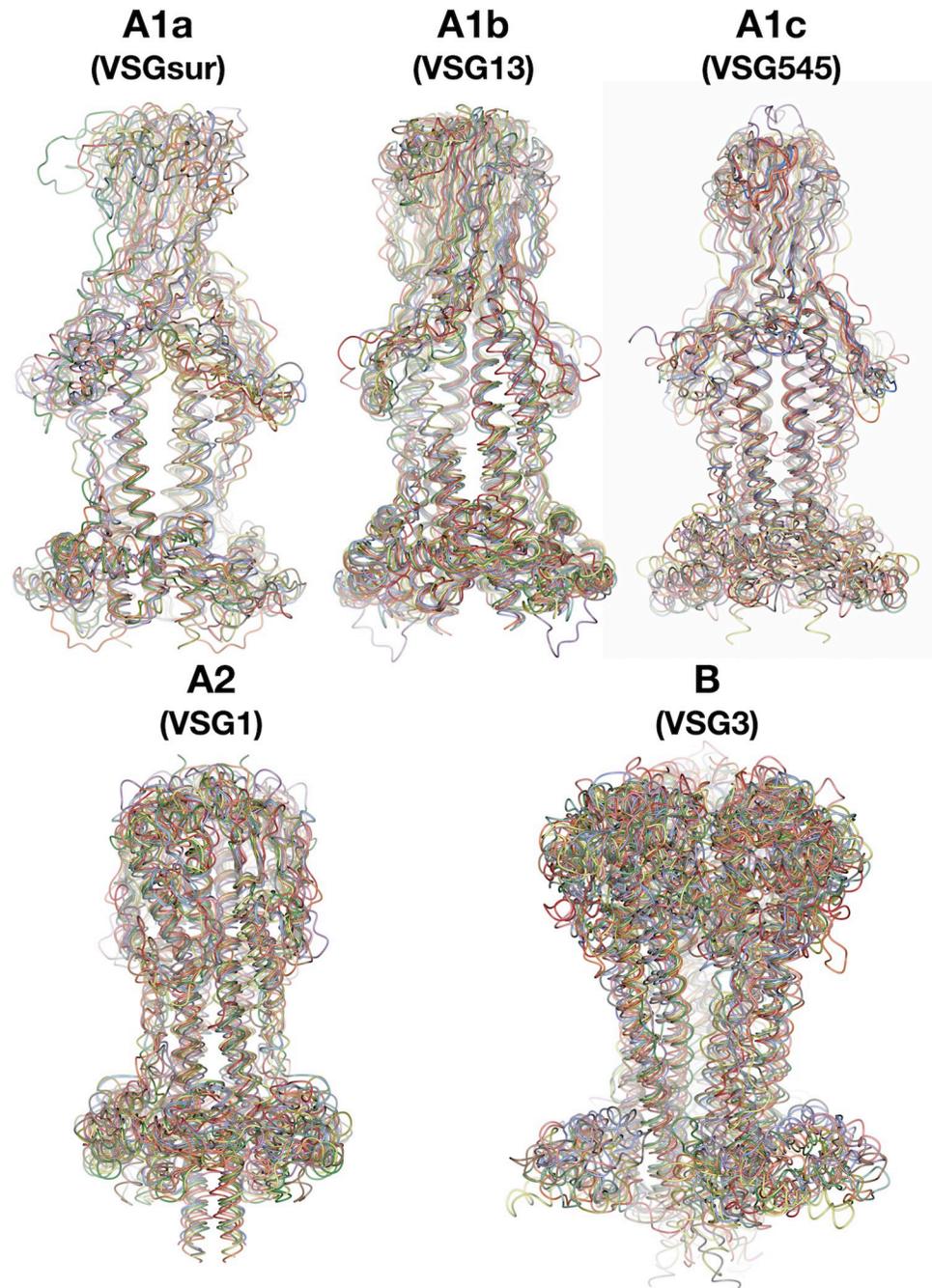
Fig 6. AlphaFold Models of Sequences from VSG Structural Classes. Structures are drawn in different colors with a thin worm style. Class A1a with models for the NTD sequences of VSGsur (experimental) and AlphaFold predictions of VSG478, VSG483, VSG485, VSG507, VSG520, VSG523, VSG529, VSG535, VSG537 Class A1b with models for the NTD sequences of VSG13 (experimental) and AlphaFold predictions of VSG510, VSG650, VSG660, VSG661, VSG662, VSG663, VSG671, VSG717, VSG719 Class A1c with models for the NTD experimental structures of VSG21 and VSG545 together with the AlphaFold predictions of sequences of VSG391, VSG479, VSG493, VSG526, VSG545, VSG584, VSG612, VSG659 Class A2 with models for the NTD sequences of VSG1(experimental) and AlphaFold predictions of, VSG472, VSG476, VSG494, VSG495, VSG504, VSG514, VSG517, VSG525, VSG527, VSG536 Class B with models for the NTD sequences of VSG3 (experimental) and AlphaFold predictions of VSG473, VSG475, VSG480, VSG486, VSG491, VSG506, VSG513, VSG516, VSG528, VSG534. Alignments generated and illustrated with CCP4mg, each structure aligned in a different color [57].

https://doi.org/10.1371/journal.pntd.0011621.g006

glycosylation). The importance of these constraints and the broadening of previous knowledge can be seen in the incorrect assumptions that were derived from the first three VSG structures determined over the period of 1990–2017 (VSG1, VSG2, and IlTat1.24). By happenchance, these three VSGs are not a random sampling of the VSGnome, but are clustered by structure into a single VSG class (A2) with specific oligomeric and structural properties that differ markedly from those in the other VSG classes. Our classification scheme and examination of hundreds of VSGs from multiple species suggests that the greater diversity that we have mapped out is likely close to the full space of variability in structure in *T brucei*, and likely also covers a substantial portion of the VSGs in related genomes like *T.b.* strain 927. It is quite possible that the VSGs in even distantly related species like *T. vivax* will find substantial fit with most of the *T. brucei* classifications, although currently some VSGs in *T. vivax* do not model well with AlphaFold, suggesting that there could exist novel structural classes in this and other species.

In this sense, while it was mistakenly assumed the "VSG-folding problem" was understood decades ago, it is now likely that this is indeed beginning to be realized, at least in *T. brucei*, and that further work across the breadth of trypanosome species will be able to fully map the antigenic space of these coat proteins. While AlphaFold is currently a strong choice for this structural mapping (as perhaps are newer systems like ESMFold and others [56]), it has to be noted that the system currently has limitations and does not predict perfectly all structures from sequence. Another caveat is that only experimental structures (or biochemical mapping) can presently establish the locations of post-translational modifications, modifications that have been shown to be immunomodulatory. Additionally, as shown with attempts to model VSG sequences from *T. vivax*, the low confidence scores of some of these predictions suggest that AlphaFold is more robust at present when there exist structural templates of sequences in specific families, e.g., *T. vivax* experimental structures are needed as templates rather than relying solely on *T. brucei* templates. Finally, our structural work has revealed that the VSGs, especially in the bottom lobe, can be characterized by extensive flexibility and alternative conformations which could have an impact on the molecular surfaces recognized by the adaptive immune system, adding an additional layer of variability to the coat.

The next steps in understanding antigenic variation in the African trypanosome should likely focus on mapping immunogenic epitopes on the VSG surfaces combined with antibody-VSG complex structures. For the time being, this will likely rest in the realm of experimental structural biology, as neither AlphaFold nor any other software has yet succeeded in robustly predicting the structures of protein-protein interactions. Once many of these have been determined, efforts to categorize the classes and natures of epitopes and the modes of interaction with antibodies will be the final piece in the puzzle to begin to develop a mechanistic framework for how this pathogen can continuously evade even the most potent immune responses.

Supporting information

S1 Fig. Purification, Crystallization, and Representative Electron Density of New VSG Crystal Structures. Summary of various steps in the crystallographic structural solution. Panels showing the gel filtration chromatogram (Superdex 200, Methods) of purified **(A)** VSG558, **(B)** VSG615, **(C)** VSG21 **(D)** VSG11 and **(E)** VSG545. A Coomassie stained SDS-PAGE gel of the final material used for crystallization, is shown, except for VSG21 and VSG545 which show an SDS-PAGE gel stained with Coomassie Blue of crystals of each VSG (crystals dissolved and run on the gel). The first lane shows molecular weight markers and the second the ~25kD band of the truncated protein in the crystals (highlighted with an arrow). Images of crystals grown in hanging drops, X-ray diffraction, and the final model 2Fo-Fc electron density

contoured at 1σ are added.
(TIFF)

S2 Fig. Alternative Crystal Forms of VSG11. (A) The asymmetric unit of the wild type VSG11 NTD structure from ammonium sulfate. Two molecules of VSG11 are packed "head-to-tail" and shown in green and salmon. **(B)** The full crystal packing in the form from (A) with the salmon monomer shown and the crystal-packing that produces the standard B class trimer highlighted with a red circle. **(C)** A structural alignment of two wild type VSG11 models showing the conformation change at bundle helix 2 (VSG11$_{WT}$-Iodine in blue, VSG11$_{WT}$-Oil in gold).
(TIFF)

S3 Fig. Alignment of the wild type VSG11 structure soaked in iodine with the 18 monomers in the chimeric VSG11 18mer. (A)-(R) show the VSG11$_{WT}$-Iodine in light blue and a different, individual VSG11$_{N2C}$ monomer in red. Due to the high flexibility of the bottom lobes in the 18mer crystals, not all amino acids in some monomers of the bottom lobe could be modeled.
(TIFF)

S4 Fig. Comparisons of VSG21 and VSG545 with AlphaFold models. VSG21 (left) and VSG545 are colored in blue and gold (for the two chains in the dimer) and the respective AlphaFold models are light gray.
(TIFF)

S5 Fig. Comparisons of Class A1c VSGs. On the left an alignment of the A1c defining dimeric structures of VSG21 and VSG545 (red and gray, respectively) with disulfides in cyan (VSG21) and green (VSG545). On the right is a monomer alignment of A1c class member (VSG545 blue with cyan disulfides) with A1b class member VSG13 (gold with green disulfides).
(TIFF)

S6 Fig. Topology of Class A and Class B (A) Two broad superfamily classes of VSGs identified through sequence analysis and defined by structural topology are shown here with representative structures. VSG monomers are shown as ribbon diagrams colored in a gradient from blue to red from N- to C-terminus. Structures reported in this manuscript are denoted in red and described in detail below. The lower, bottom lobe subdomains of the VSGs are highlighted in a green-tinted box. Structures drawn with CCP4*mg*. (B) Sequence and structurally-determined secondary structure of two representative VSGs from class A (VSG397) and class B (VSG11). The sequence region that forms the bottom lobe is indicated by green highlighting. Secondary structure illustrated with PDBSUM (https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/).
(TIFF)

S7 Fig. Subtle Structural Differences Between Class B1 and Class B2 VSGs. (A) Alignments of class B1 and B2 VSGs (experimental and AlphaFold models), highlighting the top lobe helices in B2 but not in B1 (red circle), the disordered bundle helix in B1 (yellow circle), and the inserted sequence in the bundle region (orange circle). The structural drawings match in color the names of the VSGs in panel (C). **(B)** Focus on the helixes in the 3-helix bundle differing between class B1 and B2. Orange is B2, showing that the helix is unbroken in most B2 VSGs. In B1 VSGs (light blue) the helix becomes disordered in the middle of the bundle and resumes a displaced helical structure afterward. **(C)** Structural dendrogram generated by the Dali Server comparing a number of experimental (VSG names with orange outline box) and AlphaFold

models of the class B VSGs, showing that they split into two groups. B1 and B2 subgroups are shown with light blue and light orange backgrounds, respectively.
(TIFF)

S8 Fig. Analysis of AlphaFold Predictions of VSG Structures. (A) Comparison of the predicted local distance test (pLDDT) for all predicted VSGs for the monomer compare to that for the oligomer. (B) Distribution of the predicted local distance test (pLDDT) for all predicted VSGs for the monomer and oligomer. The predicted local distance test (pLDDT) showed that the model confidence for all monomer predictions was 86.87 (standard deviation 4.831) and for the oligomer predictions 88.86 (standard deviation 4.438), evincing a statistically significant improvement in the pLDDT for the oligomer (mean difference of 1.992, P value <0.0001). (C) Comparison of the X-ray structures with the corresponding predictions (monomers and oligomers) showed a higher GDT_TS for the oligomer 93.52 (standard deviation 6.257) compared to the monomer 87.94 (standard deviation 8.101) with a mean of difference 5.583, P value 0.004, and a lower RMSD for the oligomer 0.9344 Å (standard deviation 0.4547 Å) compared to the monomer 1.283 Å (standard deviation 0.562 Å) with a mean of difference -0.3488 Å, P value 0.0027. (D) Comparison of experimental and AlphaFold predicted structures of VSG397. Ribbon diagrams of the VSG397 monomers are shown in light blue and gold for the experimental and predicted structures, respectively. Disulfides are shown in cyan and yellow for the experimental and predicted structures, respectively, marked by black arrows. For P values, P > 0.05 is considered not significant, "*" indicates a P $\leq$ 0.05, "**" indicates a P $\leq$ 0.01, "***" indicates P $\leq$ 0.001, and "****" indicates a P $\leq$ 0.0001.
(TIFF)

S9 Fig. AlphaFold Models of Sequences from VSG Structural Classes from *T. brucei brucei* 927. Structures of aligned VSG monomers are drawn in different colors with a thin worm style. **(A)** Class A1a with models for the UniProtKB/Swiss-Prot accession numbers Q580P4, Q580P5, Q57TR8, Q57TR3, Q38CQ8, Q57X40, Q583L8, Q583L5, Q57X41, Q22KU2, Q38G13, Q380W0, and Q38CP5 aligned to VSGsur. **(B)** Class A1b with models for the UniProtKB/Swiss-Prot accession numbers Q57Y76, Q57X39, Q4GY52, Q586M4, Q57XI7 aligned to VSG13. **(C)** Class A1c with models for the UniProtKB/Swiss-Prot accession numbers Q380U5, Q380V8, Q380U9, Q380Y0, Q580N8, Q38CP2 aligned to an AlphaFold model for VSG21. **(D)** Class A2 with models for the UniProtKB/Swiss-Prot accession numbers Q57X38, Q580N9, Q57Z50, Q4FKU3, Q380X8, Q57XH3, Q38G16, Q57TR9, Q380W1, Q38CQ1, Q583L4, Q38G20 aligned to VSG1. **(E)** Class B with models for the UniProtKB/Swiss-Prot accession numbers Q380U8, Q583L3, Q4FKE9, Q38CP0, Q57TR6, Q387P0, Q57TR7, Q4GY50 aligned to VSG11. Alignments generated and illustrated with CCP4mg, each structure aligned in a different color.
(TIFF)

S10 Fig. AlphaFold *T. vivax* VSG Predictions. (A) A collection of several *T. vivax* VSG protein structures predicted by AlphaFold and colored by pLDDT confidence score (as indicated). (B) Comparison of several *T. vivax* putative VSG structures (from Fam23 of the Fam23-26 VSG gene groups) and the prediction of *T. brucei brucei* VSG2, both colored by pLDDT confidence scores.
(TIFF)

S11 Fig. VSG Molecular Surfaces. Colored by charge distribution with blue indicating positive or basic, white neutral, and red acidic or negative (produced with Chimera using the "Coulombic Surface Coloring" option). The top portion shows surfaces from several experimental structures, the bottom the surfaces from models produced by AlphaFold (in either the

dimeric or trimeric assemblies).
(TIFF)

S1 Table. Crystallographic Statistics.
(PDF)

Acknowledgments

We acknowledge synchrotron time at the at the Paul Scherrer Institut, Villingen, Switzerland (SLS, beamline PX I and PX III).

Author Contributions

Conceptualization: Johan P. Zeelen, Monica Chandra, Monique van Straaten, C. Erec Stebbins.

Data curation: Sara Đaković, Johan P. Zeelen, Monica Chandra, Monique van Straaten, Janet Zhong, C. Erec Stebbins.

Formal analysis: Sara Đaković, Johan P. Zeelen, Monica Chandra, Monique van Straaten, Janet Zhong, Joseph P. Verdi, F. Nina Papavasiliou, C. Erec Stebbins.

Funding acquisition: F. Nina Papavasiliou, C. Erec Stebbins.

Investigation: Sara Đaković, Johan P. Zeelen, Anastasia Gkeka, Monica Chandra, Monique van Straaten, Konstantina Foti, Janet Zhong, Evi P. Vlachou, Francisco Aresta-Branco.

Methodology: Johan P. Zeelen, Monica Chandra, Monique van Straaten, Janet Zhong, C. Erec Stebbins.

Project administration: Johan P. Zeelen, Monique van Straaten, C. Erec Stebbins.

Resources: Sara Đaković, Anastasia Gkeka, Monica Chandra, Janet Zhong, Evi P. Vlachou, Francisco Aresta-Branco, C. Erec Stebbins.

Software: C. Erec Stebbins.

Supervision: Johan P. Zeelen, Monique van Straaten, C. Erec Stebbins.

Validation: Sara Đaković, Anastasia Gkeka, C. Erec Stebbins.

Visualization: C. Erec Stebbins.

Writing – original draft: Sara Đaković, Johan P. Zeelen, Monica Chandra, Monique van Straaten, C. Erec Stebbins.

Writing – review & editing: Sara Đaković, Johan P. Zeelen, Monica Chandra, Monique van Straaten, Joseph P. Verdi, F. Nina Papavasiliou, C. Erec Stebbins.

References

1. Keating J, Yukich JO, Sutherland CS, Woods G, Tediosi F. Human African trypanosomiasis prevention, treatment and control costs: A systematic review. Acta Trop. 2015; 150: 4–13. https://doi.org/10.1016/j.actatropica.2015.06.003 PMID: 26056739

2. Ponte-Sucre A. An Overview of Trypanosoma brucei Infections: An Intense Host–Parasite Interaction. Front Microbiol. 2016;7. https://doi.org/10.3389/fmicb.2016.02126 PMID: 28082973

3. Matthews KR, McCulloch R, Morrison LJ. The within-host dynamics of African trypanosome infections. Philos Trans R Soc Lond B Biol Sci. 2015;370. https://doi.org/10.1098/rstb.2014.0288 PMID: 26150654

4. Burri C. Sleeping Sickness at the Crossroads. Trop Med Infect Dis. 2020; 5: E57. https://doi.org/10.3390/tropicalmed5020057 PMID: 32276514

5. Abro Z, Kassie M, Muriithi B, Okal M, Masiga D, Wanda G, et al. The potential economic benefits of controlling trypanosomiasis using waterbuck repellent blend in sub-Saharan Africa. Simuunza MC, editor. PLOS ONE. 2021; 16: e0254558. https://doi.org/10.1371/journal.pone.0254558 PMID: 34283848

6. Holt HR, Selby R, Mumba C, Napier GB, Guitian J. Assessment of animal African trypanosomiasis (AAT) vulnerability in cattle-owning communities of sub-Saharan Africa. Parasit Vectors. 2016; 9: 53. https://doi.org/10.1186/s13071-016-1336-5 PMID: 26825496

7. Bangs JD. Evolution of Antigenic Variation in African Trypanosomes: Variant Surface Glycoprotein Expression, Structure, and Function. BioEssays. 2018; 40: 1800181. https://doi.org/10.1002/bies.201800181 PMID: 30370931

8. Horn D. Antigenic variation in African trypanosomes. Mol Biochem Parasitol. 2014; 195: 123–9. https://doi.org/10.1016/j.molbiopara.2014.05.001 PMID: 24859277

9. Manna PT, Boehm C, Leung KF, Natesan SK, Field MC. Life and times: synthesis, trafficking, and evolution of VSG. Trends Parasitol. 2014; 30: 251–258. https://doi.org/10.1016/j.pt.2014.03.004 PMID: 24731931

10. Mugnier MR, Stebbins CE, Papavasiliou FN. Masters of Disguise: Antigenic Variation and the VSG Coat in Trypanosoma brucei. PLOS Pathog. 2016; 12: e1005784. https://doi.org/10.1371/journal.ppat.1005784 PMID: 27583379

11. Aresta-Branco F, Erben E, Papavasiliou FN, Stebbins CE. Mechanistic Similarities between Antigenic Variation and Antibody Diversification during Trypanosoma brucei Infection. Trends Parasitol. 2019; 35: 302–315. https://doi.org/10.1016/j.pt.2019.01.011 PMID: 30826207

12. Schwede A, Jones N, Engstler M, Carrington M. The VSG C-terminal domain is inaccessible to antibodies on live trypanosomes. Mol Biochem Parasitol. 2011; 175: 201–204. https://doi.org/10.1016/j.molbiopara.2010.11.004 PMID: 21074579

13. Hempelmann A, Hartleb L, van Straaten M, Hashemi H, Zeelen JP, Bongers K, et al. Nanobody-mediated macromolecular crowding induces membrane fission and remodeling in the African trypanosome. Cell Rep. 2021; 37: 109923. https://doi.org/10.1016/j.celrep.2021.109923 PMID: 34731611

14. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The genome of the African trypanosome Trypanosoma brucei. Science. 2005; 309: 416–22. https://doi.org/10.1126/science.1112642 PMID: 16020726

15. Hutchinson OC, Smith W, Jones NG, Chattopadhyay A, Welburn SC, Carrington M. VSG structure: similar N-terminal domains can form functional VSGs with different types of C-terminal domain. Mol Biochem Parasitol. 2003; 130: 127–131. https://doi.org/10.1016/s0166-6851(03)00144-0 PMID: 12946849

16. Chandra M, Đaković S, Foti K, Zeelen JP, van Straaten M, Aresta-Branco F, et al. Structural similarities between the metacyclic and bloodstream form variant surface glycoproteins of the African trypanosome. Caljon G, editor. PLoS Negl Trop Dis. 2023; 17: e0011093. https://doi.org/10.1371/journal.pntd.0011093 PMID: 36780870

17. Pinger J, Nešić D, Ali L, Aresta-Branco F, Lilic M, Chowdhury S, et al. African trypanosomes evade immune clearance by O-glycosylation of the VSG surface coat. Nat Microbiol. 2018; 3: 932–938. https://doi.org/10.1038/s41564-018-0187-6 PMID: 29988048

18. Zeelen J, van Straaten M, Verdi J, Hempelmann A, Hashemi H, Perez K, et al. Structure of trypanosome coat protein VSGsur and function in suramin resistance. Nat Microbiol. 2021; 6: 392–400. https://doi.org/10.1038/s41564-020-00844-1 PMID: 33462435

19. Freymann D, Down J, Carrington M, Roditi I, Turner M, Wiley D. 2.9 A resolution structure of the N-terminal domain of a variant surface glycoprotein from Trypanosoma brucei. J Mol Biol. 1990; 216: 141–160. https://doi.org/10.1016/S0022-2836(05)80066-X PMID: 2231728

20. Metcalf P, Blum M, Freymann D, Turner M, Wiley DC. Two variant surface glycoproteins of Trypanosoma brucei of different sequence classes have similar 6 A resolution X-ray structures. Nature. 1987; 325: 84–6. https://doi.org/10.1038/325084a0 PMID: 2432433

21. Freymann DM, Metcalf P, Turner M, Wiley DC. 6 Å-Resolution X-ray structure of a variable surface glycoprotein from Trypanosoma brucei. Nature. 1984; 311: 167–169. https://doi.org/10.1038/311167a0 PMID: 6472475

22. Blum ML, Down JA, Gurnett AM, Carrington M, Turner MJ, Wiley DC. A structural motif in the variant surface glycoproteins of Trypanosoma brucei. Nature. 1993; 362: 603–9. https://doi.org/10.1038/362603a0 PMID: 8464512

23. Bartossek T, Jones NG, Schafer C, Cvitkovic M, Glogger M, Mott HR, et al. Structural basis for the shielding function of the dynamic trypanosome variant surface glycoprotein coat. Nat Microbiol. 2017; 2: 1523–1532. https://doi.org/10.1038/s41564-017-0013-6 PMID: 28894098

24. Umaer K, Aresta-Branco F, Chandra M, van Straaten M, Zeelen J, Lapouge K, et al. Dynamic, variable oligomerization and the trafficking of variant surface glycoproteins of Trypanosoma brucei. Traffic Cph Den. 2021; 22: 274–283. https://doi.org/10.1111/tra.12806 PMID: 34101314

25. Cross GAM, Kim H-S, Wickstead B. Capturing the variant surface glycoprotein repertoire (the VSGnome) of Trypanosoma brucei Lister 427. Mol Biochem Parasitol. 2014; 195: 59–73. https://doi.org/10.1016/j.molbiopara.2014.06.004 PMID: 24992042

26. Weirather JL, Wilson ME, Donelson JE. Mapping of VSG similarities in Trypanosoma brucei. Mol Biochem Parasitol. 2012; 181: 141–52. https://doi.org/10.1016/j.molbiopara.2011.10.011 PMID: 22079099

27. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596: 583–589. https://doi.org/10.1038/s41586-021-03819-2 PMID: 34265844

28. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022; 50: D439–D444. https://doi.org/10.1093/nar/gkab1061 PMID: 34791371

29. Burkard G, Fragoso CM, Roditi I. Highly efficient stable transformation of bloodstream forms of Trypanosoma brucei. Mol Biochem Parasitol. 2007; 153: 220–3. https://doi.org/10.1016/j.molbiopara.2007.02.008 PMID: 17408766

30. Pinger J, Chowdhury S, Papavasiliou FN. Variant surface glycoprotein density defines an immune evasion threshold for African trypanosomes undergoing antigenic variation. Nat Commun. 2017; 8: 828. https://doi.org/10.1038/s41467-017-00959-w PMID: 29018220

31. Cross GA. Release and purification of Trypanosoma brucei variant surface glycoprotein. J Cell Biochem. 1984; 24: 79–90. https://doi.org/10.1002/jcb.240240107 PMID: 6725422

32. Sheldrick GM. A short history of SHELX. Acta Crystallogr A. 2008; 64: 112–22. https://doi.org/10.1107/S0108767307043930 PMID: 18156677

33. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. Acta Crystallogr D Biol Crystallogr. 2006; 62: 859–866. https://doi.org/10.1107/S0907444906019949 PMID: 16855301

34. Langer G, Cohen SX, Lamzin VS, Perrakis A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc. 2008; 3: 1171–9. https://doi.org/10.1038/nprot.2008.91 PMID: 18600222

35. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr Biol Crystallogr. 2010; 66: 213–21. https://doi.org/10.1107/S0907444909052925 PMID: 20124702

36. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010; 66: 486–501. https://doi.org/10.1107/S0907444910007493 PMID: 20383002

37. Joosten RP, Long F, Murshudov GN, Perrakis A. The PDB_REDO server for macromolecular structure model optimization. IUCrJ. 2014; 1: 213–220. https://doi.org/10.1107/S2052252514009324 PMID: 25075342

38. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. *Phaser* crystallographic software. J Appl Crystallogr. 2007; 40: 658–674. https://doi.org/10.1107/S0021889807021206 PMID: 19461840

39. Kim H-S, Cross GAM. TOPO3α Influences Antigenic Variation by Monitoring Expression-Site-Associated VSG Switching in Trypanosoma brucei. Mansfield JM, editor. PLoS Pathog. 2010; 6: e1000992. https://doi.org/10.1371/journal.ppat.1000992 PMID: 20628569

40. Rypniewski WR, Holden HM, Rayment I. Structural consequences of reductive methylation of lysine residues in hen egg white lysozyme: An x-ray analysis at 1.8-.ANG. resolution. Biochemistry. 1993; 32: 9851–9858. https://doi.org/10.1021/bi00088a041 PMID: 8373783

41. Vonrhein C, Tickle IJ, Flensburg C, Keller P, Paciorek W, Sharff A, et al. Advances in automated data analysis and processing within *autoPROC*, combined with improved characterisation, mitigation and visualisation of the anisotropy of diffraction limits using *STARANISO*. Acta Crystallogr Sect Found Adv. 2018; 74: a360–a360. https://doi.org/10.1107/S010876731809640X

42. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods. 2022; 19: 679–682. https://doi.org/10.1038/s41592-022-01488-1 PMID: 35637307

43. Zemla A. LGA: a method for finding 3D similarities in protein structures. Nucleic Acids Res. 2003; 31: 3370–3374. https://doi.org/10.1093/nar/gkg571 PMID: 12824330

44. Zemla A, Zhou CE, Slezak T, Kuczmarski T, Rama D, Torres C, et al. AS2TS system for protein structure modeling and analysis. Nucleic Acids Res. 2005; 33: W111–W115. https://doi.org/10.1093/nar/gki457 PMID: 15980437

45. Gkeka A, Aresta-Branco F, Triller G, Vlachou EP, van Straaten M, Lilic M, et al. Immunodominant surface epitopes power immune evasion in the African trypanosome. Cell Rep. 2023; 42: 112262. https://doi.org/10.1016/j.celrep.2023.112262 PMID: 36943866

46. Mehlert A, Bond CS, Ferguson MAJ. The glycoforms of a Trypanosoma brucei variant surface glycoprotein and molecular modeling of a glycosylated surface coat. Glycobiology. 2002; 12: 607–612. https://doi.org/10.1093/glycob/cwf079 PMID: 12244073

47. Carrington M, Miller N, Blum M, Roditi I, Wiley D, Turner M. Variant specific glycoprotein of Trypanosoma brucei consists of two domains each having an independently conserved pattern of cysteine residues. J Mol Biol. 1991; 221: 823–35. https://doi.org/10.1016/0022-2836(91)80178-w PMID: 1942032

48. Holm L. Dali server: structural unification of protein families. Nucleic Acids Res. 2022; 50: W210–W215. https://doi.org/10.1093/nar/gkac387 PMID: 35610055

49. Moult J, Pedersen JT, Judson R, Fidelis K. A large-scale experiment to assess protein structure prediction methods. Proteins Struct Funct Genet. 1995; 23: ii–iv. https://doi.org/10.1002/prot.340230303 PMID: 8710822

50. Pereira J, Simpkin AJ, Hartmann MD, Rigden DJ, Keegan RM, Lupas AN. High-accuracy protein structure prediction in CASP14. Proteins Struct Funct Bioinforma. 2021; 89: 1687–1699. https://doi.org/10.1002/prot.26171 PMID: 34218458

51. Elofsson A. Progress at protein structure prediction, as seen in CASP15. Curr Opin Struct Biol. 2023; 80: 102594. https://doi.org/10.1016/j.sbi.2023.102594 PMID: 37060758

52. Bond PS, Cowtan KD. *ModelCraft*: an advanced automated model-building pipeline using *Buccaneer*. Acta Crystallogr Sect Struct Biol. 2022; 78: 1090–1098. https://doi.org/10.1107/S2059798322007732 PMID: 36048149

53. Barbarin-Bocahu I, Graille M. The X-ray crystallography phase problem solved thanks to *AlphaFold* and *RoseTTAFold* models: a case-study report. Corrigendum. Acta Crystallogr Sect Struct Biol. 2023; 79: 353–353. https://doi.org/10.1107/S2059798323002826 PMID: 36995234

54. McCoy AJ, Sammito MD, Read RJ. Implications of *AlphaFold* 2 for crystallographic phasing by molecular replacement. Acta Crystallogr Sect Struct Biol. 2022; 78: 1–13. https://doi.org/10.1107/S2059798321012122 PMID: 34981757

55. Terwilliger TC, Afonine PV, Liebschner D, Croll TI, McCoy AJ, Oeffner RD, et al. Accelerating crystal structure determination with iterative *AlphaFold* prediction. Acta Crystallogr Sect Struct Biol. 2023; 79: 234–244. https://doi.org/10.1107/S205979832300102X PMID: 36876433

56. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic level protein structure with a language model. Synthetic Biology; 2022 Jul. https://doi.org/10.1101/2022.07.20.500902

57. McNicholas S, Potterton E, Wilson KS, Noble MEM. Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallogr D Biol Crystallogr. 2011; 67: 386–394. https://doi.org/10.1107/S0907444911007281 PMID: 21460457