Multivalency drives the neutralizing activity of antibodies against the *Plasmodium falciparum* circumsporozoite protein

Short Title: Antibody responses to the *Plasmodium* circumsporozoite protein

Camilla R. Fisher[1¶], Joe A. Kaczmarski[1¶], Henry J. Sutton[2¶], Ben Clifton[1], Joshua Mitchell[1], Yeping Cai[2], Johanna N. Dups[2], Nicholas J. D'Arcy[2], Mandeep Singh[2], Hayley A. McNamara[2], Aaron Chuah[2], Tom Peat[3], Colin J. Jackson[1]* and Ian A. Cockburn[2]*

1. Research School of Chemistry, The Australian National University, Canberra, ACT 2601, Australia

2. John Curtin School of Medical Research, The Australian National University, Canberra, ACT 2601, Australia

3. CSIRO Materials, Science and Engineering, Parkville, Victoria, Australia

* Corresponding authors

E-mail: colin.jackson@anu.edu.au

E-mail: ian.cockburn@anu.edu.au (lead contact).

[¶] These authors contributed equally to this work

Abstract

The repeat region of the *Plasmodium falciparum* circumsporozoite protein (CSP) is a major vaccine antigen because it can be targeted by parasite neutralizing antibodies; however, little is known about this interaction. We used isothermal calorimetry and X-ray crystallography to analyze the binding of the *Plasmodium*-neutralizing 2A10 antibody to CSP. Strikingly, we found that the repeat region of CSP is bound by multiple antibodies and that this multivalent interaction drives the affinity of this antibody. Because the CSP protein can cross-link multiple B cell receptors (BCRs) we hypothesized that the B cell response might be T-independent. However, by sequencing the BCRs of CSP-repeat specific cells we found that these cells underwent somatic hypermutation and affinity maturation indicative of a T-dependent response. Interestingly, the BCR repertoire of responding B cells was limited suggesting that the structural simplicity of the repeat may limit the breadth of the immune response.

Author Summary

Vaccines aim to protect by inducing the immune system to make molecules called antibodies that can recognize molecules on the surface of invading pathogens. In the case of malaria, our most advanced vaccine candidates aim to make antibodies that recognize the circumsporozoite protein molecule on the surface of the invasive parasite stage called the sporozoite. In this report we use X-ray crystallography to determine the structure of CSP-binding antibodies at the atomic level. We use other techniques such as isothermal titration calorimetry to examine how this antibody interacts with the CSP molecule. Strikingly, we found that each CSP molecule could bind 6 antibodies. This finding has implications for the immune response and may explain why high titers of antibody are needed for protection. Moreover because the structure of the CSP repeat is quite simple we determined that the number of different kinds of antibodies that could bind this molecule are quite small. However those antibodies can become quite high affinity as a result of a process called affinity maturation that allows the body to learn how to make improved antibodies specific for pathogen molecules. These data show that while it is challenging for the immune system to recognize and neutralize CSP, it should be possible to generate viable vaccines targeting this molecule.

Introduction

Malaria caused by *Plasmodium falciparum* causes the deaths of around 430,000 people each year [1]. The most advanced vaccine candidate is currently the RTS,S/AS01 vaccine which consists of a truncated version of the sporozoite-surface circumsporozoite protein (CSP), packaged in a Hepatitis C core virus-like particle delivered in AS01 - a proprietary liposome based vaccine [2]. Phase II and Phase III clinical trials have repeatedly demonstrated that the vaccine is capable of giving around 50% protection against clinical malaria in field settings for the first year following vaccination [3]. The bulk of protection is attributed to antibodies targeting the CSP repeat epitope included within the vaccine, with some contribution from CD4+ T cells [4]. It is still unclear why the antibody response to CSP is only partially protective. We lack structural information about how neutralizing antibodies bind to CSP and knowledge on the breadth and nature of the B cell response elicited.

Antibodies to CSP were first identified as potential mediators of protection following seminal studies that showed that immunization with irradiated sporozoites could induce sterile protection against live parasite challenge [5,6]. In the early 1980s, monoclonal antibodies (mAbs) isolated from mice immunized with sporozoites were found to be capable of inducing the neutralization of sporozoites (known as the circumsporozoite reaction) and were used to clone CSP, one of the first malaria antigens identified [7,8]. In all *Plasmodium* species CSP contains 3 domains: an N-terminal domain and a C-terminal domain, separated by a repeat region, which was the target of all the original mAbs identified [7,9,10]. In the 3D7 reference strain of *P. falciparum* the CSP repeat has 38 NANP-repeats interspersed with 4 NVDP repeats

87 especially towards the N-terminus [11] though different numbers of repeats have been
88 observed [12]. One of the most effective *P. falciparum* sporozoite neutralizing
89 antibodies identified in these early studies was 2A10 which can neutralize sporozoite
90 infectivity *in vitro* [13] and in *in vivo* mouse models utilizing rodent *P. berghei*
91 parasites expressing the *P. falciparum* CSP repeat region [14,15].
92
93 While CSP binding antibodies have been shown to be able to neutralize
94 sporozoites and block infection, it has also been proposed that CSP is an
95 immunological "decoy" that induces a suboptimal T-independent immune response
96 perhaps because of the CSP repeat cross-linking multiple B cell receptors (BCRs)
97 [16]. Nonetheless this hypothesis has not been tested and it remains unknown if the
98 repetitive regions of CSP can cross-link multiple BCRs as they are not as large as
99 typical type-II T-independent antigens [17]. Furthermore, the very little published
100 data on the sequences of CSP binding antibodies does not convincingly support
101 activation of a broad B cell repertoire: a small study of five *P. falciparum* CSP mouse
102 monoclonal antibodies (mAbs) identified some shared sequences [18]. In humans, a
103 study that generated mAbs from three individuals who received RTS,S found that the
104 three antibodies studied had distinct sequences though these all used similar heavy
105 chains [19].
106
107 Given these gaps in our understanding of the antibody response to CSP we
108 undertook a comprehensive biophysical characterization of the 2A10 sporozoite-
109 neutralizing antibody that binds to the CSP repeat. Using rigorous biophysical
110 methodology we found that this antibody binds with a higher affinity than expected,
111 in the nano-molar range. Previous studies using competition ELISAs with peptides

predicted a micro-molar affinity [20,21]. Strikingly, isothermal titration calorimetry and structural analyses revealed that the CSP repeat can be bound by around six antibodies suggesting that it may require large amounts of antibody for neutralization, and that the repeat may potentially crosslink multiple BCRs on the surface of a B cell. However, analysis of CSP-specific B cells revealed that CSP-specific B cells can enter germinal centers and undergo affinity maturation contradicting the notion that the response to CSP is largely T-independent. Moreover, we found that the BCR repertoire of CSP-binding B cells is quite limited which may restrict the size and effectiveness of the immune response.

Results

Characterization of the thermodynamics of 2A10-antigen binding

We began our analysis by performing isothermal titration calorimetry (ITC) to understand the interaction between 2A10 and CSP. For ease of expression we used a recombinant CSP (rCSP) construct described previously which was slightly truncated with 27 repeats [22]. ITC experiments were run on the purified 2A10 antibody and the purified 2A10 antigen-binding fragment (F_{AB}) fragment to test the thermodynamic basis of the affinity of 2A10 F_{AB} towards CSP. Experiments were also performed on the 2A10 F_{AB} fragment with the synthetic peptide antigen $(NANP)_6$, which is a short segment of the antigenic NANP-repeat region of CSP (**Table 1; Fig. 1**). The binding free energies (ΔG) and dissociation constants (K_D) were found to be -49.0 kJ/mol and 2.7 nM for the full 2A10 antibody with CSP, -40 kJ/mol and 94 nM for the 2A10 F_{AB} with CSP, and -36.4 kJ/mol and 420 nM for the 2A10 F_{AB} with the $(NANP)_6$ peptide.

Surprisingly, we did not observe a typical 1:1 antibody/F_{AB} domain:antigen binding stoichiometry (Table 1). We found that each $(NANP)_6$ peptide was bound to by ~2 F_{AB} fragments (2.8 repeats per F_{AB} domain). With the rCSP protein we observed that 10.8 ± 0.7 F_{AB} fragments were able to bind to each rCSP molecule, (2.5 repeats per F_{AB} domain. Finally, when the single-domain F_{AB} fragment is replaced by the full 2A10 antibody (which has two F_{AB} domains), we observe binding of 5.8 antibodies per rCSP molecule (4.7 repeats per antibody), i.e. all complexes exhibit approximately the same binding stoichiometry of two F_{AB} fragments/domains per ~5 repeat units. These results suggest that the antigenic region of CSP constitutes a multivalent antigen and that repeating, essentially identical, epitopes must be

146 available for the binding of multiple F_{AB} domains.

147 It is not possible to separate affinity from avidity in this system, although it is
148 apparent that there is a substantial benefit to the overall strength of binding between
149 the antibody and antigen through the binding of multiple F_{AB} domains. The F_{AB}:rCSP
150 complex and the 2A10:rCSP complex had similar enthalpy and entropy of binding
151 (**Table 1**), but the lower enthalpy of binding for the F_{AB} fragment:$(NANP)_6$ peptide
152 complex suggests that there are additional stabilizing interactions when larger
153 numbers of F_{AB} domains can bind to the rCSP protein. The observation that this
154 antibody-antigen (Ab-Ag) interaction is primarily enthalpically driven is consistent
155 with the general mechanism of Ab-Ag interactions [23]. Altogether, these data
156 suggest that the binding of multiple antibodies to the repeat region of CSP stabilizes
157 the interaction and the multivalent nature of this interaction increases the affinity of
158 binding.

159

160 **Table 1. Thermodynamic parameters for interactions between 2A10 F_{AB}, 2A10**
161 **and antigens.**

	$(NANP)_6$:F_{AB}	**rCSP:F_{AB}**	**rCSP:2A10**
K_a (M^{-1})	$(2.37 \pm 0.91) \times 10^6$	$(1.07 \pm 0.39) \times 10^7$	$(3.6 \pm 2.7) \times 10^8$
K_d (nM)	420 ± 160	94 ± 34	2.7 ± 2.1
ΔH (kJ/mol complex)	–113 ± 5	–1245 ± 112	–1175 ± 44
TΔS (kJ/mol complex)	–76.6 ± 4.9	–1205 ± 112	–1126 ± 44
ΔG (kJ/mol complex)	–36.4 ± 1.0	–40.0 ± 0.9	–49.0 ± 1.9
n (F_{AB}/2A10: Ag)	2.16 ± 0.06	10.8 ± 0.7	5.8 ± 0.1

162 Parameters were determined by ITC at 25 °C. Errors for *n* (Ag : F_{AB}), K_a and
163 ΔH (complex) are 95% confidence intervals estimated from a single titration; errors
164 for other parameters were propagated.

Structural analysis of the (NANP)-repeat region and the 2A10 F_{AB}

To better understand the molecular basis of the multivalent interaction between 2A10 and rCSP, we performed structural analysis of the components. Previous work has indicated that the NANP-repeat region of CSP adopts a flexible rod-like structure with a regular repeating helical motif that provides significant separation between the N-terminal and the C-terminal domains [24]. Here, we performed far-UV circular dichroism (CD) spectroscopy to investigate the structure of the $(NANP)_6$ peptide. These results were inconsistent with a disordered random coil structure (**S1 Fig**.). Rather, the absorption maximum around 185 nm, minimum around 202 nm and shoulder between 215 and 240 nm, is characteristic of intrinsically disordered proteins that can adopt a spectrum of states [25].

The lowest energy structures of the $(NANP)_6$ repeat were predicted using the PEP-FOLD *de novo* peptide structure prediction algorithm [26]. The only extended state among the lowest energy structures that was consistent with the reported spacing of the N-and C-terminal domains of CSP [24], and which presented multiple structurally similar epitopes was a linear, quasi-helical structure, which formed a regularly repeating arrangement of proline turns (**Fig. 2A**). The theoretical CD spectrum of this conformation was calculated (**S1 Fig.**), qualitatively matching the experimental spectra: the maximum was at 188 nm, the minimum at 203 nm and there was a broad shoulder between 215 and 240 nm. To investigate the stability of this conformation, we performed a molecular dynamics (MD) simulation on this peptide, which showed that this helical structure could unfold, and refold, on timescales of

tens of nanoseconds, supporting the idea that it is a low-energy, frequently sampled, configuration in solution (**S1 Mov.**, **S2 Fig**.). We also observed the same characteristic hydrogen bonds between a carbonyl following the proline and the amide nitrogen of the alanine, and the carbonyl group of an asparagine and a backbone amide of asparagine three residues earlier, that are observed in the crystal structure of the NPNA fragment [27]. Thus, this configuration, which is consistent with previously published experimental data, is a regular, repeating, extended conformation that would allow binding of multiple F_{AB} domains to several structurally similar epitopes.

To better understand the interaction between the 2A10 and the (NANP)-repeat region, we solved the crystal structure of the 2A10 F_{AB} fragment in two conditions (**S1 Table**), yielding structures that diffracted to 2.5 Å and 3.0 Å. All of the polypeptide chains were modeled in good quality electron density maps (**Fig. 2B**), except for residues 134-137 of the light chain. This loop is located at the opposite end of the F_{AB} fragment to the variable region and not directly relevant to antigen binding. The 2.5 Å structure contained a single polypeptide in the asymmetric unit, whereas the 3.0 Å structure contained three essentially identical chains. Superimposition of the four unique F_{AB} fragments from the two structures revealed that the variable antigen binding region is structurally homogeneous, suggesting that this region might be relatively pre-organized in the 2A10 F_{AB}. This is consistent with the observation that antibodies typically undergo relatively limited conformational change upon epitope binding [23]. Indeed, a recent survey of 49 antibody:antigen complexes revealed that within the binding site, CDR-H3 was the only element that showed significant conformational change upon antigen binding and even this was only observed in one

third of the antibodies [28].

Attempts to obtain a crystal structure of a complex between 2A10 F_{AB} and the $(NANP)_6$ peptide were unsuccessful; unlike binary Ab-Ag interactions, in which the Ab will bind to a single epitope on an antigen and produce a population of structurally homogeneous complexes that can be crystallized, in this interaction we are dealing with an intrinsically-disordered peptide, the presence of multiple binding sites (epitopes) on the peptide and the possibility that more than one 2A10 F_{AB} domain can bind the peptide, i.e. it is difficult to obtain a homogeneous population of complexes, which is a prerequisite for crystallization. Attempts to soak the $(NANP)_6$ peptide into the high-solvent form of 2A10 F_{AB}, in which there were no crystal packing interactions with the binding-loops, caused the crystals to dissolve, again suggesting that the heterogeneity of the peptide and the presence of multiple epitopes produces disorder that is incompatible with crystal formation.

Modeling the interaction of the 2A10 F_{AB} with the NANP-repeat region and testing the model through site-directed mutagenesis

Although it was not possible to obtain a crystal structure of the 2A10-$(NANP)_6$ peptide complex, the accurate structures of the 2A10 F_{AB} fragment, the $(NANP)_6$ peptide, and the knowledge that antibodies seldom undergo significant conformational changes upon antigen binding [28], allowed us to model the interaction, which we tested using site directed mutagenesis. Computational modeling of Ab-Ag interactions has advanced considerably in recent years and several examples of complexes with close to atomic accuracy have been reported in the

literature [29]. Using the SnugDock protein-protein docking algorithm [29], we obtained an initial model for binding of the peptide to the CDR region of the 2A10 F_{AB} fragment (**Fig. 2C**). We then performed, in triplicate, three 50 ns molecular dynamics simulations on this complex to investigate whether the interaction was stable over such a time period (**S2 Mov**., **S3 Fig.**). These simulations confirmed that the binding mode that was modeled is stable, suggesting that it is a reasonable approximation of the interaction between these molecules. To experimentally verify whether our model of the 2A10 F_{AB}:$(NANP)_6$ peptide interaction was plausible, we performed site directed mutagenesis of residues predicted to be important for binding. Our model predicted that the interaction with $(NANP)_6$ would be mainly between CDR2 and CDR3 of the light chain and CDR2 and CDR3 of the heavy chain (**Fig. 2C**).

In the light chain (**Fig. 3A,B**), Y38 is predicted to be one of the most important residues in the interaction; it contributes to the formation of a hydrophobic pocket that buries a proline residue and is within hydrogen bonding distance, *via* its hydroxyl group, to a number of backbone and side-chain groups of the peptide. Loss of this side-chain abolished binding. Y56 also forms part of the same proline-binding pocket as Y38, and loss of this side-chain also resulted in an almost complete loss of binding. R109 forms a hydrogen bond to an asparagine residue on the side of the helix; mutation of this residue to alanine results in a partial loss of binding. Y116 is located at the center of the second proline-binding pocket; since loss of the entire side-chain through an alanine mutation would lead to general structural disruption of the F_{AB} fragment, we mutated this to a phenylalanine (removing the hydroxyl group), which led to a significant reduction in binding. Finally, S36A was selected as a

control: the model indicated that it was outside the binding site, and the ELISA data indicated that had no effect on $(NANP)_n$ binding.

Within the heavy chain (**Fig. 3C,D**), mutation of N57 to alanine led to complete loss of binding, which is consistent with it forming a hydrogen bond to a side-chain asparagine but also being part of a relatively well packed region of the binding site that is mostly buried upon binding. T66 is located on the edge of the binding site and appears to provide hydrophobic contacts through its methyl group with the methyl side-chain of an alanine of the peptide; mutation of this residue resulted in a partial loss of binding. Interestingly, mutation of E64, which is location in an appropriate position to form some hydrogen bonds to the peptide resulted in a slight increase in binding, although charged residues on the edge of protein:protein interfaces are known to contribute primarily to specificity rather than affinity [30]. Specifically, the cost of desolvating charged residues such as glutamate is not compensated for by the hydrogen bonds that may be formed with the binding partner. Y37 is located outside the direct binding site in the apo-crystal structure; the loss of affinity could arise from long-range effects, such as destabilization of the position of nearby loops. In general, the effects of the mutations are consistent with the model of the interaction.

The multivalency of the CSP repeat region

As shown in **Fig. 3A** and **3C** the binding mode of the F_{AB} fragment to the $(NANP)_6$ peptide is centered on two proline residues from two non-adjacent NANP-repeats. These cyclic side-chains are hydrophobic in character and are buried deeply

in the core of the F_{AB} antigen binding site, into hydrophobic pockets formed by Tyr38 and Tyr56 of the light chain and the interface between the two chains. In contrast, the polar asparagine residues on the sides of the helix are involved in hydrogen binding interactions with a number of polar residues on the edge of the binding site, such as N57 of the heavy chain. Due to the twisting of the $(NANP)_6$ repeat, the binding epitope of the peptide is 2.5-3 alternate NANP repeats, with a symmetrical epitope available for binding on the opposite face (**Fig. 4A**). Thus, this binding mode is consistent with the stoichiometry of the binding observed in the ITC measurements, where we observed a stoichiometry of two 2A10 F_{AB} fragments per $(NANP)_6$ peptide. To investigate whether this binding mode was also compatible with the indication from ITC that ~10.7 2A10 F_{AB} fragments, or six antibodies (containing 12 F_{AB} domains) could bind the CSP protein (**Table 1**), we extended the peptide to its full length. It is notable that the slight twist in the NANP helix results in the epitope being offset along the length of the repeat region, thereby allowing binding of ten 2A10 F_{AB} fragments (**Fig. 4B**). Six 2A10 antibodies can bind if two antibodies interact by a single F_{AB} domain and the other four interact with both F_{AB} domains. The observation that the F_{AB} fragments bind sufficiently close to each other to form hydrogen bonds also explains the observation from the ITC that the complexes with rCSP, which allow adjacent F_{AB} fragment binding, have more favorable binding enthalpy, i.e. the additional bonds formed between adjacent F_{AB} fragments further stabilize the complex and lead to greater affinity (**Table 1**). Thus, the initially surprising stoichiometry that we observe through ITC appears to be quite feasible based in the structure of the NANP-repeat region of the rCSP protein and the nature of the rCSP-2A10 complex. It is also clear that the effect of antibody binding to this region would be to prevent the linker flexing between the N- and C-terminal domains and

maintaining normal physiological function, explaining the neutralizing effect of the antibodies.

Identification of endogenous $(NANP)_n$ specific B cells to determine the BCR repertoire

We next set out to determine the implications of our structure for the B cell response to CSP. In particular we wanted to know if the repeat structure drove a diverse T-independent response resulting in a broad antibody response or if B cells specific for CSP were able to enter the germinal center and undergo affinity maturation. To test this hypothesis we used $(NANP)_n$-based tetramers to identify antigen specific B cells in BALB/C mice immunized with *P. berghei* sporozoites expressing the repeat region of the *P. falciparum* CSP (*P. berghei* CS^{Pf}) [15]. We then used high throughput BCR sequencing to identify the BCR sequences of the $(NANP)_n$ specific cells. The tetramers are formed by the binding of 4 biotinylated $(NANP)_9$ repeats with streptavidin conjugated phycoerythrin (PE) or allophycocyanin (APC). We used BALB/C mice as this was the strain originally used to generate 2A10. To validate our tetramer approach, mice were immunized with either *P. berghei* CS^{Pf} or another line of *P. berghei* with a mutant CSP (*P. berghei* CS^{5M}) that contains the endogenous (*P. berghei*) repeat region, which has a distinct repeat sequence $(PPPPNPND)_n$. $(NANP)_n$-specific cells were identified with two tetramer probes bound to different conjugates to exclude B cells that are specific for the PE or APC components of the tetramers which are numerous in mice [31]. We found that mice immunized with *P. berghei* CS^{Pf} sporozoites developed large tetramer double positive populations, which had class switched (**Fig. 5A** and **B**). In contrast, the number of

tetramer double positive cells in mice receiving control parasites was the same as in unimmunized mice; moreover these cells were not class switched and appeared to be naïve precursors indicating that our tetramers are identifying bona-fide $(NANP)_n$-specific cells (**Fig. 5B** and **C**). Further analysis of the different populations of B cells showed that most B cells present at this time-point were $GL7^+$ $CD38^-$ indicating that they are germinal center B cells (**Fig. 5B** and **D**). Given that T cells are required to sustain germinal center formation beyond ~3 days these data indicate that a T-dependent response can develop to CSP following sporozoite immunization [32].

A restricted repertoire of BCRs can bind to the $(NANP)_n$ repeat

We next set out to determine the diversity of the B cell response to CSP. While the repeat structure of CSP is hypothesized to induce a broad polyclonal response [33], an alternative hypothesis is that the antigenically simple structure of the repeat epitope might only be recognized by a small number of naive B cells. To examine the BCR usage of $(NANP)_n$-specific B cells we sorted $(NANP)_n$-specific cells 35 days post immunization with sporozoites. We then prepared cDNA from the cells and amplified the heavy and kappa chain sequences using degenerate primers as described previously [34,35]. Heavy and light chain libraries were prepared from 4 immunized mice as well as from 3 naïve mice from which we bulk sorted B cells as controls. We obtained usable sequences from 3 of the 4 mice for both the heavy chain and kappa chain. Analysis of the heavy chain revealed that in each mouse 3 or 4 V regions dominated the immune response (**Fig. 6A**). The V regions identified (IGHV1-20; IGHV1-26; IGHV1-34 and IGHV5-9) were generally shared among the mice. As a formal measure of the diversity of our V region usage in the $(NANP)_n$ specific cells

and the bulk B cells from naïve mice we calculated the Shannon entropy for these populations. This analysis formally demonstrated that the diversity of the antigen specific B cells was significantly lower than the diversity of the repertoire in naïve mice (**Fig. 6B**). We further found that each V region was typically associated with the same D and J sequences even in different mice. For example, IGHV1-20 was typically associated with J4, IGHV5-9 with J4 while in different mice IGHV1-34 was variously paired with J1 or J4 (**Fig. 6C**). Similar results were obtained for the kappa chain with the response dominated by IGKV1-135; IGKV5-43/45; IGKV1-110; IGKV1-117 and IGKV14-111 (**Fig. 6D and E**). The V regions were typically paired with the same J regions even in different mice (**Fig. 6F**), for example IGKV5.43/45 was typically paired with IGKJ5 or IGKJ2 and IGKV1-110 was typically paired with IGKJ5, although IGKV1-135 was typically more promiscuous. One limitation of our high throughput sequencing approach is that the degenerate primers only amplified ~70% of the known IGHV and IGKV sequences in naive mice, suggesting that we may not capture the full diversity of the response. However, comparison with the 5 published antibody sequences (**S2** and **S3 Table**) that include IGHV-1-20, IGKV5-45 and IGKV1-110 reveals that we are likely capturing the bulk of the antibody diversity. Together these data suggest that the number of B cell clones responding to CSP may be limited, potentially reducing the ability of the immune system to generate effective neutralizing antibodies.

CSP-binding antibodies undergo somatic hypermutation to improve affinity

While it is clear that CSP is the target of neutralizing antibodies it has been suggested that CSP might induce large T-independent responses at the expense of

potentially more useful T-dependent germinal center responses that can result in robust B cell memory [33]. We therefore examined our deep sequencing data to determine if CSP-specific antibodies had undergone somatic hypermutation (SHM) that would be indicative of B cells specific for CSP entering the germinal center. Taking advantage of the fact that our kappa chain primers capture the entire V-J sequences of the antibodies we sequenced we asked: 1) if the kappa chains shared between immune animals differed from the germline (providing evidence of SHM) and 2) if the mutations were conserved between different mice indicative of directed selection. Analysis of the reads from the kappa chains of the three immune mice showed that these had a much higher degree of mutation than bulk B cells from naïve mice, demonstrating SHM in the CSP-specific antibodies (**Fig. 7A**). We further examined each specific common kappa chain in turn (IGVK1-110; IGKV1-135; IGVK5-53/54) comparing the sequences obtained from naive B cells and $(NANP)_n$ specific cells in immune mice. This analysis showed that while, as expected, sequences from naïve mice contained few mutations, the sequences from immune mice had much higher levels of SHM. Importantly mutations were found to be concentrated in the CDR loops, and were frequently shared by immunized mice providing strong circumstantial evidence for affinity maturation (**Fig. 7B**; data for IGVK1-110 only shown).

To directly test if CSP-binding antibodies undergo affinity maturation we expressed the predicted germline precursor to the 2A10 antibody (2A10 gAb) in HEK293T cells. We identified the predicted germline precursors of the 2A10 heavy and light chains using the program V-quest (**Figs. S4** and **S5**). This analysis identified the heavy chain as IGHV9-3; IGHD1-3; IGHJ4 and the light chain as IGKV10-

94;IGKJ2, with the monoclonal antibody carrying 6 mutations in the heavy chain and 7 in the light chain. The 2A10 gAb had considerably lower binding in ELISA assays compared to the 2A10 mAb itself (**Fig. 7C**) indicative that affinity maturation has taken place in this antibody. To determine the relative contribution of mutations in the heavy and light chain to enhancing binding we also made hybrid antibodies consisting of the mAb heavy chain and the gAb light chain and vice versa. Interestingly mutations in the light chain were almost entirely sufficient to explain the enhanced binding by the mAb compared to the gAb (**Fig. 7C**).

To identify the specific mutations that were important we introduced the mutations individually into the gAb light chain construct. We prioritized mutations that were shared with the 27E antibody which has previously been found to be clonally related to 2A10 having been isolated from the same mouse and which shares the same germline heavy and light chains as the 2A10 mAb [18]. We found that two mutations (L114F and T117V) in the CDR3 of the light chain appeared to account for most of the gain in binding (**Fig. 7C**). The effect of these antibodies appeared to be additive rather than synergistic as revealed by experiments in which we introduced these mutations simultaneously (**Fig. 7D**). A further mutation close to the light chain CDR2 (H68Y) also caused a modest increase in binding. As expected mutations in the heavy chains appeared generally less important for increasing binding though M39I, N59I and T67F all gave modest increases in binding (**Fig. 7E**). Collectively our data suggest that CSP repeat antibodies can undergo somatic hypermutation in germinal centers resulting in affinity maturation, however the antibody response may be limited by the number of naïve B cells that can recognize and respond to this antigen.

Discussion

Here we provide an analysis of the structure of a *Plasmodium falciparum* sporozoite-neutralizing antibody (2A10). We further model the binding of this antibody with its antigen target, the repeat region of CSP, and provide a thermodynamic characterization of this interaction. Finally, we used novel tetramer probes to identify and sort antigen specific B cells responding to sporozoite immunization in order to measure the diversity and maturation of the antibody response. We found that the avidity of 2A10 for the rCSP molecule was in the nanomolar range, which was much higher than the affinity previously predicted from competition ELISAs with small peptides [20,21]. This affinity is a consequence of the multivalent nature of the interaction, with up to 6 antibodies being able to bind to each rCSP molecule. To spatially accommodate this binding the antibodies must surround the CSP in an off-set manner, which is possible due to the slight twist in the helical structure that the CSP can adopt. It is notable that the twisted, repeating arrangement of the CSP linker is the only structure that would allow binding in the stoichiometry observed through the ITC. We further found that the diversity of the antibody repertoire to the CSP repeat was limited, perhaps due to the relative simplicity of the target epitope. However, these antibodies have undergone affinity maturation to improve affinity, potentially allowing protective immune responses to develop.

Using ITC we determined the affinity of 2A10 for rCSP to be 2.7 nM, which is not unusual for a mouse mAb. However it is a much higher affinity than that predicted from competition ELISAs that predicted a micro-molar affinity [20,21]. However, these competition ELISAs were performed with short peptides rather than

rCSP. Indeed, when we performed ITC with a short peptide and F_{AB} fragments we too obtained an affinity in the micro-molar range (0.42 μM). The difference in avidity between the F_{AB} binding to the peptide or full-length CSP and that of the antibody appears to be driven by a more favorable enthalpy of binding. It is likely that additional stabilizing interactions between adjacent F_{AB} domains, which is consistent with the structural model, contribute to this. One caveat of these data is that we used a slightly truncated repeat, however it is likely that longer repeats will have further stabilization of the interaction that could result in even higher affinity interaction between CSP and binding antibodies.

Our data provide important insights into the requirements for sporozoite-neutralization by CSP binding antibodies. The finding that each rCSP molecule can be bound by ~ 6 antibody molecules is consistent with the finding that relatively large amounts of antibody are required for protection against sporozoites [14,19,36]. Indeed, this may be an underestimate and it may be that full-length CSP can accommodate additional antibody binding. If on the one hand, the surface of the sporozoite provides multiple binding sites for CSP binding antibodies this may allow the parasites to be relatively easily opsonized and phagocytized. On the other hand, this may not be an important mechanism of action as anti-CSP F_{AB} fragments have previously been shown to be sufficient for blocking sporozoite infectivity in vivo [36]. Moreover CSP is readily shed from the surface of the sporozoite as the sporozoites undergo migration, which may act as a means of sloughing off bound antibody to evade this response [37]. It has also been suggested that the CSP repeat might act as a hinge allowing the N-terminal domain to mask the C-terminal domain that is believed to be important for binding to and invading hepatocytes [9]. Antibody

490 binding as observed here may disrupt this hinge perhaps resulting in the premature
491 exposure of the C-terminal domain and the loss of sporozoite infectivity.
492
493 Our results uncovering how neutralizing antibodies bind to CSP has several
494 implications for understanding the development of the immune response to CSP.
495 Notably the finding that the CSP molecule can be bound by multiple antibodies/B cell
496 receptors raises the possibility that this molecule can indeed crosslink multiple BCRs
497 and potentially act as a type-II T independent antigen [17]. Such antigens typically
498 induce large but relatively short-lived immune responses [17]. Interestingly, the
499 RTS,S/AS01 vaccine based on that contains 18 CSP repeats and does appear to induce
500 high, but relatively short-lived, titers of anti-CSP antibodies [4,38], which would be
501 consistent with it inducing a type-II T-independent response. Nonetheless, we also
502 detected extensive somatic hypermutation and affinity maturation, not only from
503 established monoclonal antibodies but also from CSP-specific B cells following a
504 single sporozoite immunization. Thus, there is a T-dependent germinal center
505 component to the antibody response too. The relative contributions of short-lived
506 antibody production and long-term B cell memory to protection is an area for future
507 investigation.
508
509 The finding of a limited repertoire in the BCR sequences specific for the
510 $(NANP)_n$ repeat contradicts previous suggestions that the response to CSP might be
511 broad and polyclonal [33]. One explanation for this limited antibody diversity is that
512 the antigenic simplicity of the CSP repeat region limits the range of antibodies that are
513 capable of responding. A prior example of this is the antibodies to the Rhesus (Rh) D
514 antigen. The RhD antigen differs from RhC by only 35-36 amino acids, resulting in

515 the creation of a minimal B cell epitope [39]. The repertoire of antibodies that can
516 bind this epitope are accordingly limited and mainly based on the VH3-33 gene
517 family [40]. Another potential explanation for a limited antibody repertoire could be
518 that the $(NANP)_n$ repeat shares structural similarity with a self-antigen as is
519 speculated to happen with meningococcus type B antigens [41], however it is not
520 clear what this self-antigen might be. One potential outcome of this finding is that if
521 each B cell clone has a finite burst size this may limit the magnitude of the overall B
522 cell response.
523
524 One area for future investigation is to determine the binding modes and
525 sporozoite neutralizing capacities of other antibodies in the response. It is clear that
526 not all CSP-repeat binding antibodies have the same capacity for sporozoite
527 neutralization [13]. As such the finding of a limited repertoire of responding B cells
528 may lead to the possibility that some people have holes in their antibody repertoires
529 limiting their ability to make neutralizing antibodies. This may explain why, while
530 there is a broad correlation between ELISA tires of antibodies to the CSP repeat and
531 protection following RTS,S vaccination, there is no clear threshold for protection [4].
532
533 While our work has been performed with mouse antibodies, there are major
534 similarities between mouse and human antibody loop structure [42]. The main
535 difference between the two species is the considerably more diverse heavy chain
536 CDR3 regions that are found in human antibodies [43]. In terms of our sequence data,
537 it may be that humans may have a more diverse antibody repertoire, not least because
538 as larger individuals they may have a greater diversity of naïve B cells specific CSP.
539 However, it is notable that all 4 human monoclonal antibodies described to date from

different volunteers share the use of the IGHV3-30 gene family [19,20], suggesting that in humans as well as mice there is a constrained repertoire of responding B cells.

Overall our data provide important insights into how the antibody response to CSP develops. Our results also help explain why relatively large amounts of antibodies are required for sporozoite neutralization and suggest that the ability to generate an effective B cell response may be limited by the very simplicity of the repeat epitope. These data support previous suggestions that CSP may be a suboptimal target for vaccination. However, we also find that CSP binding antibodies can undergo somatic hypermutation and reach high affinities. This suggests if we can develop vaccination strategies to diversify the repertoire of responding B cells and favor the germinal center response it may be possible to generate long-term protective immunity targeting this major vaccine candidate antigen.

Methods

Ethics statement

All animal procedures were approved by the Animal Experimentation Ethics Committee of the Australian National University (Protocol numbers: A2013/12; A2014/62 and A2015/76). All research involving animals was conducted in accordance with the National Health and Medical Research Council's (NHMRC) Australian Code for the Care and Use of Animals for Scientific Purposes and the Australian Capital Territory Animal Welfare Act 1992.

Mice and Immunizations

BALB/C mice (bred in house at the Australian National University) were immunized IV with 5 x 10^4 *P. berghei* CS^{5M} sporozoites expressing mCherry [44] or 5 x 10^4 *P. berghei* CS^{Pf} sporozoites dissected by hand from the salivary glands of *Anopheles stephensi* mosquitoes. Mice were then treated with 0.6mg choloroquine IP daily for 10 days to prevent the development of blood stage infection.

Flow Cytometry and sorting

Single cell preparations of lymphocytes were isolated from the spleen of immunized mice and were stained for flow cytometry or sorting by standard procedures. Cells were stained with lineage markers (anti-CD3, clone 17A2; anti-GR1, clone RB6-8C5 and anti-NKp46, clone 29A1.4) antibodies to B220 (clone RA3-6B2), IgM (clone II/41), IgD (11-26c2a) and $(NANP)_9$ tetramers conjugated to PE or APC. Antibodies were purchased from Biolegend while tetramers were prepared in house by mixing biotinylated $(NANP)_9$ peptide with streptavidin conjugated PE or APC (Invitrogen) in

the a 4:1 molar ratio. Flow-cytometric data was collected on a BD Fortessa flow cytometer (Becton Dickinson) and analyzed using FlowJo software (FlowJo). Where necessary cells were sorted on a BD FACs Aria I or II machine.

Sequencing of $(NANP)_n$ specific cells and BCR analysis

Single cell suspensions from the spleens of immunized mice were stained with (NANP)n tetramers and antibodies to B cell markers as described in the supplemental experimental procedures. Antigen specific cells were sorted on a FACS ARIA I or II instrument prior to RNA extraction with the Arturus Picopure RNA isolation kit (Invitrogen) and cDNA preparation using the iScript cDNA synthesis kit (BioRad). BCR sequences were amplified using previously described heavy and kappa chain primers including adaptor sequences allowing subsequent indexing using the Nextera indexing kit (Illumina). Analysis was performed in house using R-scripts and the program MiXCR as described in supplemental experimental procedures.

Binding of antibody variants

Variants of the 2A10 antibody were expressed in HEK293 T cells (a kind gift of Carola Vinuesa, Australian National University) as described in the supplemental experimental procedures. Binding to the CSP repeat was tested by ELISA and ITC using standard techniques as described in the supplemental experimental procedures.

Data Deposition

Sequence data generated in this paper is deposited at the NCBI BioProject database accession number PRJNA352758. Atomic coordinates and related experimental data

for structural analyses are deposited in the Protein Data Bank (PDB) with PDB codes 5ZSF and 5T0Y.

606 ## Acknowledgments

607

608 We thank the C3 Crystallisation Centre at CSIRO for help with crystal formation and
609 the Australian Synchrotron and beamline scientists for help with data collection. We
610 thank Michael Devoy, Harpreet Vohra and Catherine Gillespie of the Imaging and
611 Cytometry Facility at the John Curtin School of Medical Research for assistance with
612 flow cytometry and multi-photon microscopy.

613

614 **References**

615

616 1. World Health Organization (2016) World Malaria Report 2016. Geneva: World
617 Health Organization.
618 2. Casares S, Brumeanu TD, Richie TL (2010) The RTS,S malaria vaccine. Vaccine
619 28: 4880-4894.
620 3. RTS,S Clinical Trials Partnership (2015) Efficacy and safety of RTS,S/AS01
621 malaria vaccine with or without a booster dose in infants and children in
622 Africa: final results of a phase 3, individually randomised, controlled trial.
623 Lancet 386: 31-45.
624 4. White MT, Bejon P, Olotu A, Griffin JT, Riley EM, et al. (2013) The relationship
625 between RTS,S vaccine-induced antibodies, CD4(+) T cell responses and
626 protection against Plasmodium falciparum infection. PLoS One 8: e61395.
627 5. Nussenzweig RS, Vanderberg J, Most H, Orton C (1967) Protective immunity
628 produced by the injection of x-irradiated sporozoites of plasmodium berghei.
629 Nature 216: 160-162.
630 6. Seder RA, Chang LJ, Enama ME, Zephir KL, Sarwar UN, et al. (2013) Protection
631 against malaria by intravenous immunization with a nonreplicating sporozoite
632 vaccine. Science 341: 1359-1365.
633 7. Dame JB, Williams JL, McCutchan TF, Weber JL, Wirtz RA, et al. (1984)
634 Structure of the gene encoding the immunodominant surface antigen on the
635 sporozoite of the human malaria parasite Plasmodium falciparum. Science
636 225: 593-599.

637 8. Yoshida N, Nussenzweig RS, Potocnjak P, Nussenzweig V, Aikawa M (1980)
638 Hybridoma produces protective antibodies directed against the sporozoite
639 stage of malaria parasite. Science 207: 71-73.
640 9. Coppi A, Natarajan R, Pradel G, Bennett BL, James ER, et al. (2011) The malaria
641 circumsporozoite protein has two functional domains, each with distinct roles
642 as sporozoites journey from mosquito to mammalian host. J Exp Med 208:
643 341-356.
644 10. Zavala F, Cochrane AH, Nardin EH, Nussenzweig RS, Nussenzweig V (1983)
645 Circumsporozoite proteins of malaria parasites contain a single
646 immunodominant region with two or more identical epitopes. J Exp Med 157:
647 1947-1957.
648 11. Gardner MJ, Hall N, Fung E, White O, Berriman M, et al. (2002) Genome
649 sequence of the human malaria parasite Plasmodium falciparum. Nature 419:
650 498-511.
651 12. Zeeshan M, Alam MT, Vinayak S, Bora H, Tyagi RK, et al. (2012) Genetic
652 variation in the Plasmodium falciparum circumsporozoite protein in India and
653 its relevance to RTS,S malaria vaccine. PLoS One 7: e43430.
654 13. Hollingdale MR, Nardin EH, Tharavanij S, Schwartz AL, Nussenzweig RS (1984)
655 Inhibition of entry of Plasmodium falciparum and P. vivax sporozoites into
656 cultured cells; an in vitro assay of protective antibodies. J Immunol 132: 909-
657 913.
658 14. Espinosa DA, Gutierrez GM, Rojas-Lopez M, Noe AR, Shi L, et al. (2015)
659 Proteolytic Cleavage of the Plasmodium falciparum Circumsporozoite Protein
660 Is a Target of Protective Antibodies. J Infect Dis 212: 1111-1119.

15. Persson C, Oliveira GA, Sultan AA, Bhanot P, Nussenzweig V, et al. (2002) Cutting edge: a new tool to evaluate human pre-erythrocytic malaria vaccines: rodent parasites bearing a hybrid Plasmodium falciparum circumsporozoite protein. J Immunol 169: 6681-6685.

16. Schofield L, Uadia P (1990) Lack of Ir-Gene Control in the Immune-Response to Malaria .1. A Thymus-Independent Antibody-Response to the Repetitive Surface Protein of Sporozoites. Journal of Immunology 144: 2781-2788.

17. Defrance T, Taillardet M, Genestier L (2011) T cell-independent B cell memory. Current Opinion in Immunology 23: 330-336.

18. Anker R, Zavala F, Pollok BA (1990) VH and VL region structure of antibodies that recognize the (NANP)3 dodecapeptide sequence in the circumsporozoite protein of Plasmodium falciparum. Eur J Immunol 20: 2757-2761.

19. Foquet L, Hermsen CC, van Gemert GJ, Van Braeckel E, Weening KE, et al. (2014) Vaccine-induced monoclonal antibodies targeting circumsporozoite protein prevent Plasmodium falciparum infection. J Clin Invest 124: 140-144.

20. Chappel JA, Rogers WO, Hoffman SL, Kang AS (2004) Molecular dissection of the human antibody response to the structural repeat epitope of Plasmodium falciparum sporozoite from a protected donor. Malar J 3: 28.

21. Zavala F, Tam JP, Hollingdale MR, Cochrane AH, Quakyi I, et al. (1985) Rationale for development of a synthetic vaccine against Plasmodium falciparum malaria. Science 228: 1436-1440.

22. Cerami C, Frevert U, Sinnis P, Takacs B, Clavijo P, 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.

686 23. Braden BC, Poljak RJ (1995) Structural features of the reactions between
687 antibodies and protein antigens. FASEB J 9: 9-16.
688 24. Plassmeyer ML, Reiter K, Shimp RL, Jr., Kotova S, Smith PD, et al. (2009)
689 Structure of the Plasmodium falciparum circumsporozoite protein, a leading
690 malaria vaccine candidate. J Biol Chem 284: 26951-26963.
691 25. Kelly SM, Jess TJ, Price NC (2005) How to study proteins by circular dichroism.
692 Biochim Biophys Acta 1751: 119-139.
693 26. Shen Y, Maupetit J, Derreumaux P, Tuffery P (2014) Improved PEP-FOLD
694 Approach for Peptide and Miniprotein Structure Prediction. J Chem Theory
695 Comput 10: 4745-4758.
696 27. Ghasparian A, Moehle K, Linden A, Robinson JA (2006) Crystal structure of an
697 NPNA-repeat motif from the circumsporozoite protein of the malaria parasite
698 Plasmodium falciparum. Chem Commun (Camb): 174-176.
699 28. Sela-Culang I, Alon S, Ofran Y (2012) A systematic comparison of free and
700 bound antibodies reveals binding-related conformational changes. J Immunol
701 189: 4890-4899.
702 29. Sircar A, Gray JJ (2010) SnugDock: paratope structural optimization during
703 antibody-antigen docking compensates for errors in antibody homology
704 models. PLoS Comput Biol 6: e1000644.
705 30. Davis SJ, Davies EA, Tucknott MG, Jones EY, van der Merwe PA (1998) The
706 role of charged residues mediating low affinity protein-protein recognition at
707 the cell surface by CD2. Proc Natl Acad Sci U S A 95: 5490-5494.
708 31. Pape KA, Taylor JJ, Maul RW, Gearhart PJ, Jenkins MK (2011) Different B cell
709 populations mediate early and late memory during an endogenous immune
710 response. Science 331: 1203-1207.

711 32. de Vinuesa CG, Cook MC, Ball J, Drew M, Sunners Y, et al. (2000) Germinal
712 centers without T cells. J Exp Med 191: 485-494.
713 33. Schofield L (1990) The circumsporozoite protein of Plasmodium: a mechanism of
714 immune evasion by the malaria parasite? Bull World Health Organ 68 Suppl:
715 66-73.
716 34. Arnaout R, Lee W, Cahill P, Honan T, Sparrow T, et al. (2011) High-resolution
717 description of antibody heavy-chain repertoires in humans. PLoS One 6:
718 e22365.
719 35. Busse CE, Czogiel I, Braun P, Arndt PF, Wardemann H (2014) Single-cell based
720 high-throughput sequencing of full-length immunoglobulin heavy and light
721 chain genes. Eur J Immunol 44: 597-603.
722 36. Potocnjak P, Yoshida N, Nussenzweig RS, Nussenzweig V (1980) Monovalent
723 fragments (Fab) of monoclonal antibodies to a sporozoite surface antigen
724 (Pb44) protect mice against malarial infection. J Exp Med 151: 1504-1513.
725 37. Stewart MJ, Vanderberg JP (1988) Malaria Sporozoites Leave Behind Trails of
726 Circumsporozoite Protein during Gliding Motility. Journal of Protozoology
727 35: 389-393.
728 38. White MT, Bejon P, Olotu A, Griffin JT, Bojang K, et al. (2014) A combined
729 analysis of immunogenicity, antibody kinetics and vaccine efficacy from
730 phase 2 trials of the RTS,S malaria vaccine. BMC Med 12: 117.
731 39. Avent ND, Madgett TE, Lee ZE, Head DJ, Maddocks DG, et al. (2006) Molecular
732 biology of Rh proteins and relevance to molecular medicine. Expert Rev Mol
733 Med 8: 1-20.

734 40. Chang TY, Siegel DL (1998) Genetic and immunological properties of phage-
735 displayed human anti-Rh(D) antibodies: implications for Rh(D) epitope
736 topology. Blood 91: 3066-3078.
737 41. Finne J, Leinonen M, Makela PH (1983) Antigenic similarities between brain
738 components and bacteria causing meningitis. Implications for vaccine
739 development and pathogenesis. Lancet 2: 355-357.
740 42. North B, Lehmann A, Dunbrack RL (2011) A New Clustering of Antibody CDR
741 Loop Conformations. Journal of Molecular Biology 406: 228-256.
742 43. Stanfield RL, Wilson IA (2014) Antibody Structure. Microbiology spectrum 2.
743 44. Cockburn IA, Tse SW, Zavala F (2014) CD8+ T cells eliminate liver stage
744 Plasmodium parasites without detectable bystander effect. Infect Immun.
745
746
747

Supporting Information Legends

S1 Fig. Theoretical (A) and experimental (B) CD spectra of the $(NANP)_6$ peptide. The computational prediction of the spectra (A) was performed using DichroCalc [9], the experimental spectra was measured at 222 nm at 25 °C. A peak at 185 nm, minimum at 205 nm and shoulder between 215 and 240 nm are consistent with an intrinsically disordered, but not random coil, structure.

S2 Fig. Cluster analysis for MD simulations of $(NANP)_6$ peptide. Conformations were clustered by concatenating the trajectory and performing a Jarvis-Patrick analysis. The clusters are sorted by their RMSD from the first cluster (starting geometry). As shown, Run 2 is stable in the starting geometry for several ns, while Run 3 diverged, then reconverged to the starting geometry, where it was stable for several ns. These data suggest the quasi-helical structure observed from the ab initio calculations is stable, and can be spontaneously sampled, on a timescale of several ns.

S3 Fig. Cluster analysis for MD simulations of $(NANP)_6$ peptide. Molecular dynamics simulation of the $(NANP)_6$:F_{AB} complex. Root mean square deviation (RMSD) of the $(NANP)_6$:F_{AB} complex as a function of time, independent simulations are shown in green, black and red.

S4 Fig: Alignment of 2A10 heavy chain and the predicted germline sequence Residues that are mutated away from the predicted germline sequence in more one or more other antibody heavy chain (2E7 or 3D6) are highlighted in red, mutations that are predicted to be involved in binding to CSP are highlighted in blue.

S5 Fig: Alignment of 2A10 heavy chain and the predicted germline sequence

Residues that are mutated away from the predicted germline sequence in both 2A10 and the related 2E7 antibody are highlighted in red, mutations that are predicted to be involved in binding to CSP are highlighted in blue.

Movie S1: Molecular Dynamics simulation of the solution structure of the $(NANP)_6$ peptide

Excerpt from $(NANP)_6$ run 3. Trajectory was fitted to minimize alpha-carbon RMSD and then passed through a low-pass filter with filter length 8 frames to reduce temporal aliasing.

Movie S2: Molecular Dynamics simulation of the interaction of the (NANP)n repeat with the 2A10 FAB

Excerpt from 2A10:$(NANP)_6$ run 3. Trajectory was fitted to minimize alpha-carbon RMSD and then passed through a low-pass filter with filter length 8 frames to reduce temporal aliasing.

Figure Legends

Fig. 1. ITC data for interactions between 2A10 F_{AB} and antigens. (A) Titration of 2A10 F_{AB} with $(NANP)_6$. (B) Titration of 2A10 F_{AB} with rCSP. (C) Titration of 2A10 with rCSP. The upper panels represent baseline-corrected power traces. By convention, negative power corresponds to exothermic binding. The lower panels represent the integrated heat data fitted to the independent binding sites model.

Fig 2. Structures of the $(NANP)_6$ peptide (A), the 2A10 F_{AB} fragment (B) and the model of the F_{AB} fragment-$(NANP)_6$ complex (C). (A) The calculated structure of the $(NANP)_6$ peptide is a helical structure containing the same hydrogen bonds between a carbonyl following the proline and the amide nitrogen of the alanine, and the carbonyl group of an asparagine and a backbone amide of asparagine 3 residues earlier (highlighted in red) that are observed in [27]. (B) Electron density (blue mesh; $2mF_o$-dF_c at 1 σ) of the 2A10 F_{AB} fragment viewed from above the antigen-binding site. Light chain is shown as yellow sticks, heavy chain as cyan. (C) A calculated model of the $(NANP)_6$:2A10 F_{AB} fragment complex. The CDR2 regions of each chain are shown in red, the CDR3 regions of each chain are shown in blue.

Fig 3. Detailed view of the $(NANP)_6$:2A10 F_{AB} interface and site directed mutagenesis. (A) A model of the light chain:$(NANP)_6$ interface. (B) ELISA results showing the effect of mutating light chain interface residues; error bars are based on technical replicates from one of two independent experiments. (C) A model of the heavy chain:$(NANP)_6$ interface. (D) ELISA results showing the effect of mutating heavy chain interface residues; error bars are based on technical replicates from one

of two independent experiments.

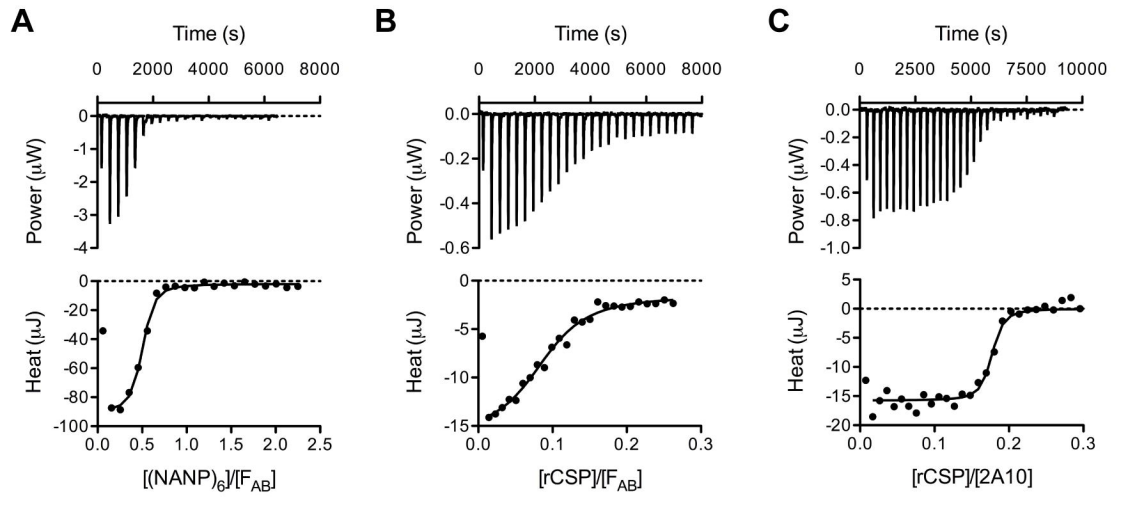
Fig 4. The multivalency of the NANP repeat region of the CSP protein. (A) An $(NANP)_6$ peptide results in the presentation of two symmetrical epitopes, formed by alternating repeats (cyan and magenta), allowing binding by two F_{AB} domains, in keeping with the stoichiometry observed by ITC. (B) The full 27-mer repeat region results in the presentation of at least 10 separate epitopes and the twist of the helix results in displacement along the length of the repeat region, which allows binding of up to 10 separate F_{AB} fragments, consistent with 4 antibodies bound by both F_{AB} domains, and two bound by a single F_{AB} domain.

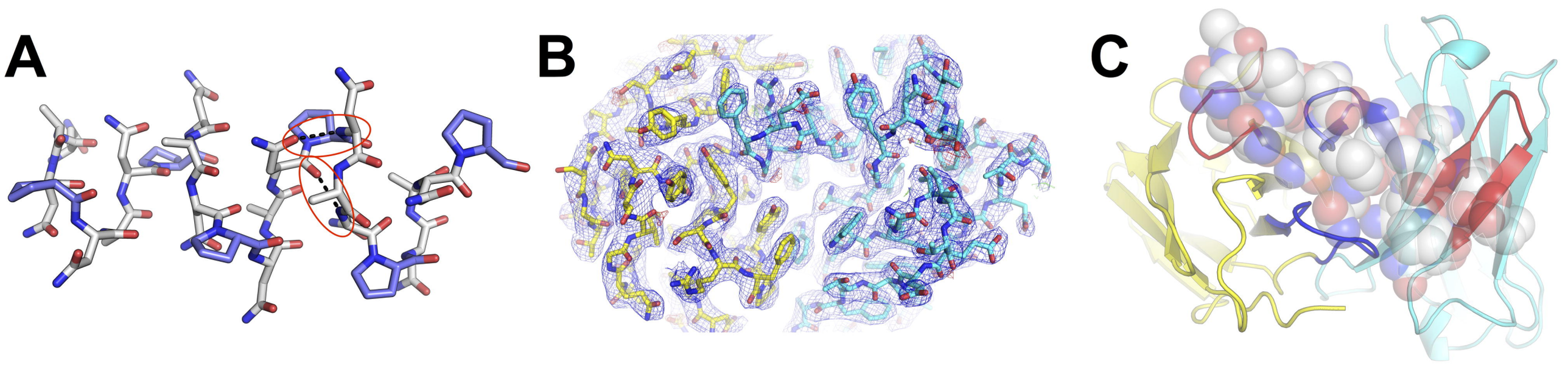
Fig 5. CSP-specific B cells enter the germinal center following sporozoite immunization. BALB/C Mice were immunized with either 5 x 10^4 *P. berghei* CS^{5M} (expressing the endogenous *P. berghei* CSP repeat) or 5 x 10^4 *P. berghei* CS^{Pf} (expressing the circumsporozoite protein from *P. falciparum*). 12 days later the B cell response was analyzed by flow cytometry and putative $(NANP)_n$-specific cells were identified using PE and APC conjugated tetramers. (A) Representative flow cytometry plots showing the identification of $(NANP)_n$-specific (Tetramer+) cells. (B) Representative flow cytometry plots showing the proportion of Tetramer+ cells that have class switched and entered a germinal center. (C) Quantification of the number of class switched Tetramer+ cells under different immunization conditions. (D) Quantification of the number of germinal center Tetramer+ cells under different immunization conditions.

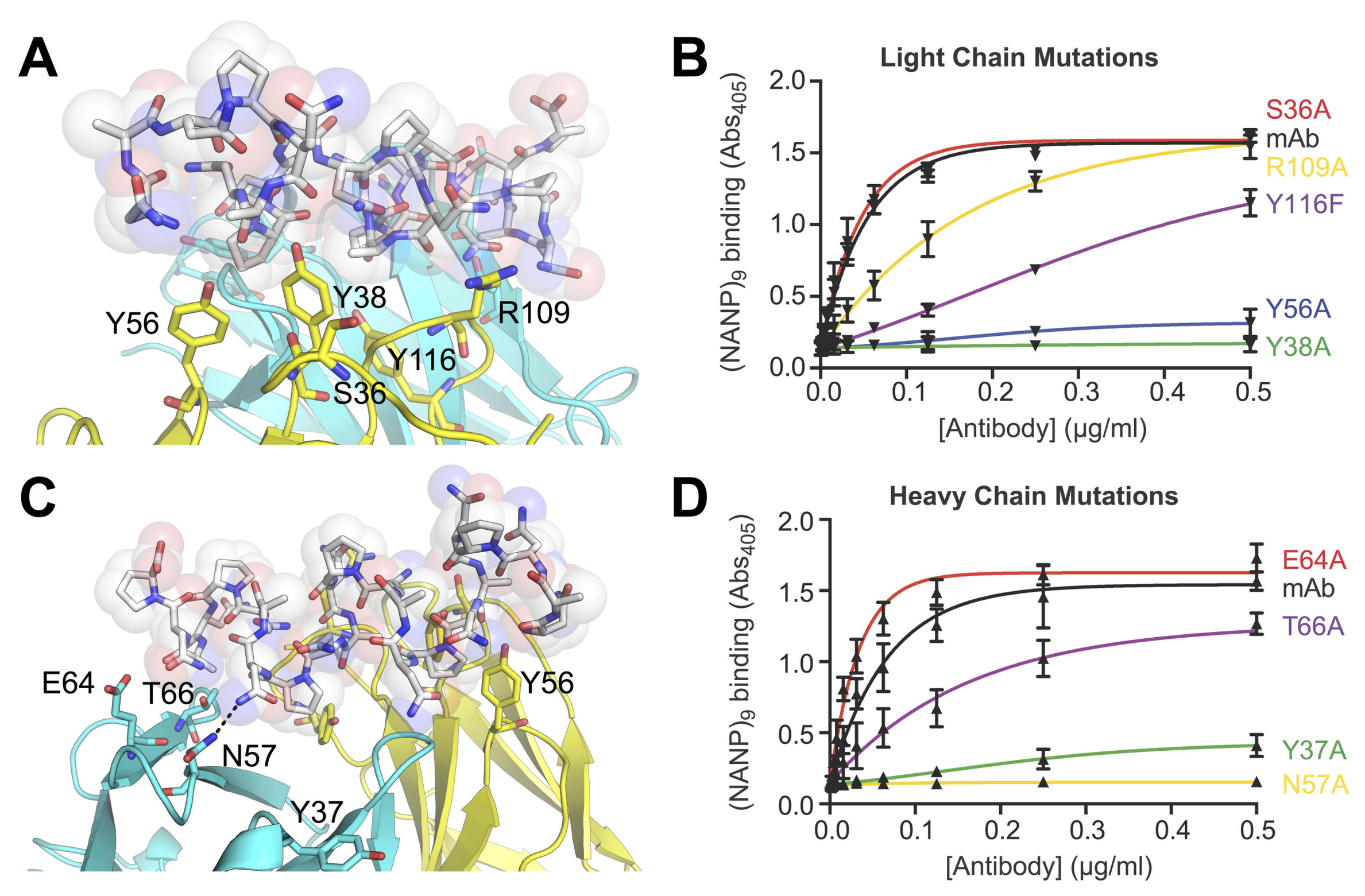
Fig 6. Limited diversity of $(NANP)_n$ specific antibodies. BCR sequences were amplified from Tetramer+ cells sorted from BALB/C mice 35 days after

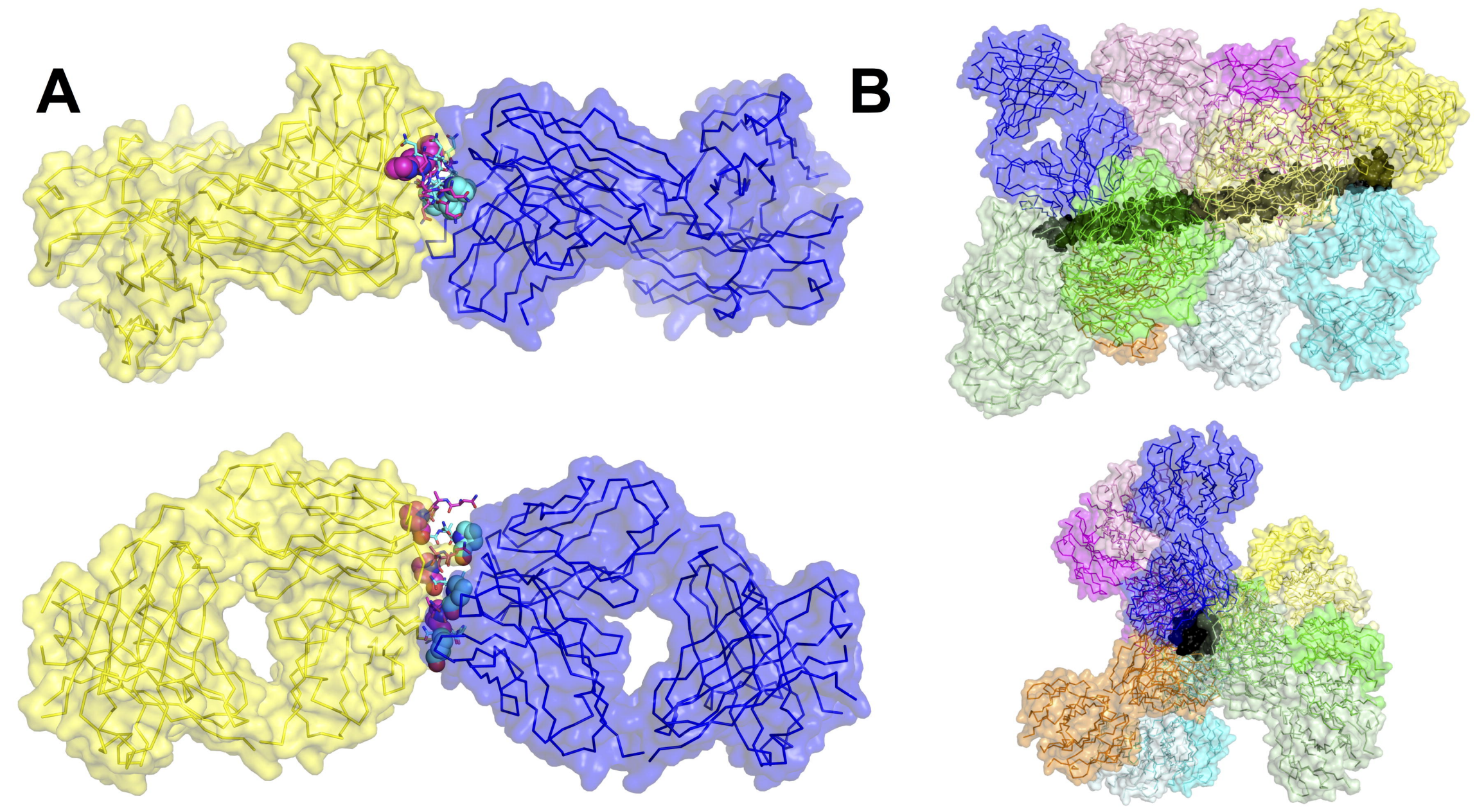
immunization with *P. berghei* CS^{Pf} as well as bulk B cells from naïve BALB/C mice (A) IGHV gene usage from among B cells from a representative naïve mouse (grey bars) and Tetramer+ cells from immune mice (red, blue and yellow bars). (B) Shannon's diversity calculated for the diversity of IGHV region usage among bulk B cells and Tetramer+ cells. (C) Circos plots showing the IGHV-IGHJ pairings in a representative naive mice and 3 immune mice. (D) IGKV gene usage from among B cells from a representative naïve mouse (grey bars) and Tetramer+ cells from immune mice (red, blue and green bars). (E) Shannon's diversity calculated for the diversity of IGKV region usage among bulk B cells and Tetramer+ cells. (F) Circos plots showing the IGKV-IGKJ pairings in a representative naïve mouse and 3 immune mice.

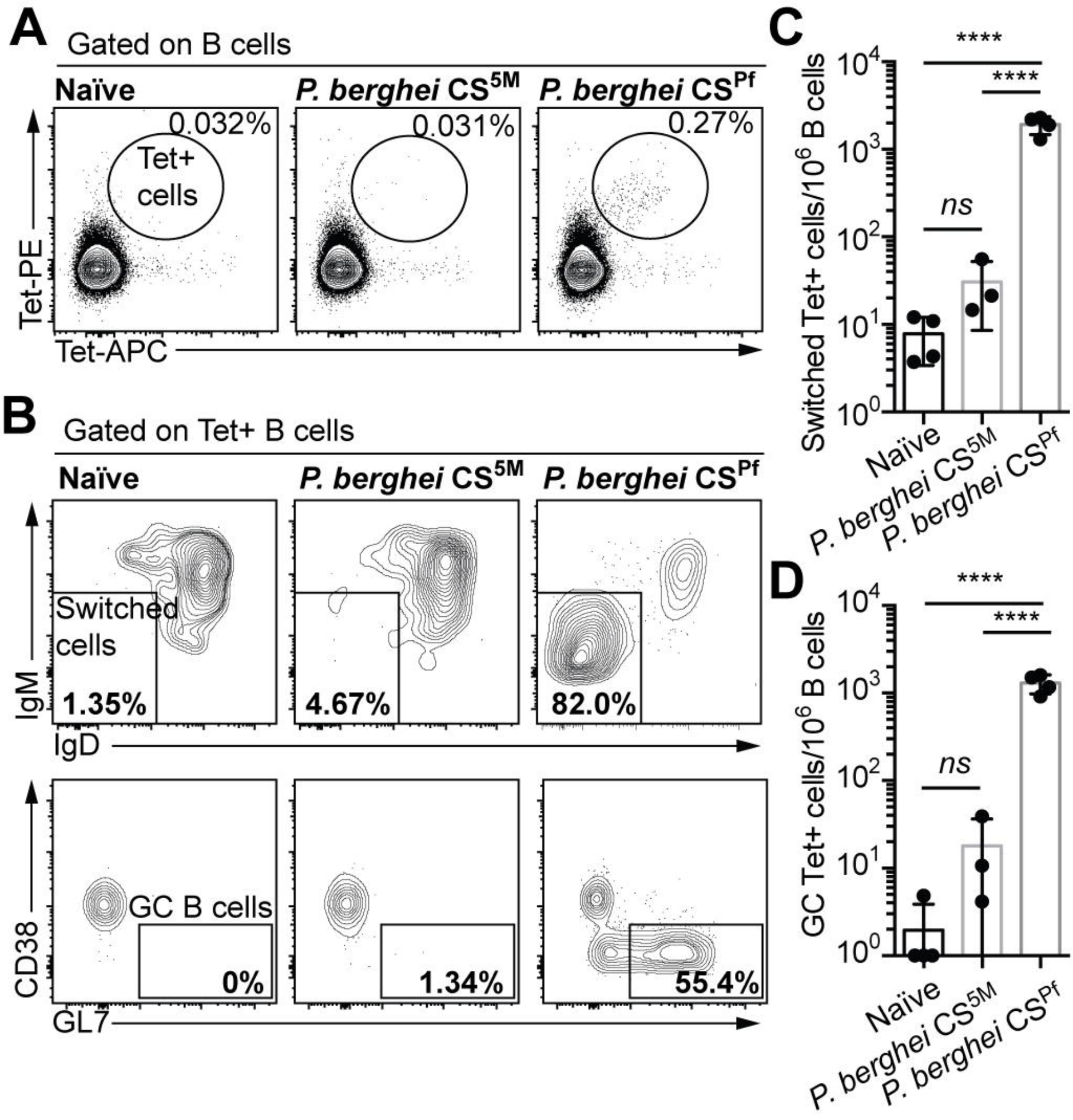
Fig 7: CSP-binding antibodies undergo somatic hypermutation and affinity maturation (A) Violin plots showing the number of mutations per read from bulk B cells from 3 individual naive mice and sorted $(NANP)_n$ specific B cells from sporozoite immunized mice **(**B) Skyscraper plots showing the location of mutations away from germline in the IGKV1-110 gene in a naïve mouse and in sorted $(NANP)_n$ specific cells in three sporozoite immunized mice. (C) ELISA binding to the $(NANP)_9$ peptide of recombinant antibodies corresponding to the 2A10 mAb, the predicted germline precursor, and hybrid antibodies containing the 2A10 heavy chain (mHC) paired with the germline light chain (gLC) and the 2010 light chain (mLC) paired with germline heavy chain (gHC). (D) Predicted mutations in the gLC were introduced to the germline precursor and their effect on binding to $(NANP)_9$ measured by ELISA (E) Predicted mutations in the gHC were introduced to hybrid antibodies consisting of the mLC and the gHC heavy chain and their effect on binding to $(NANP)_9$ measured by ELISA.

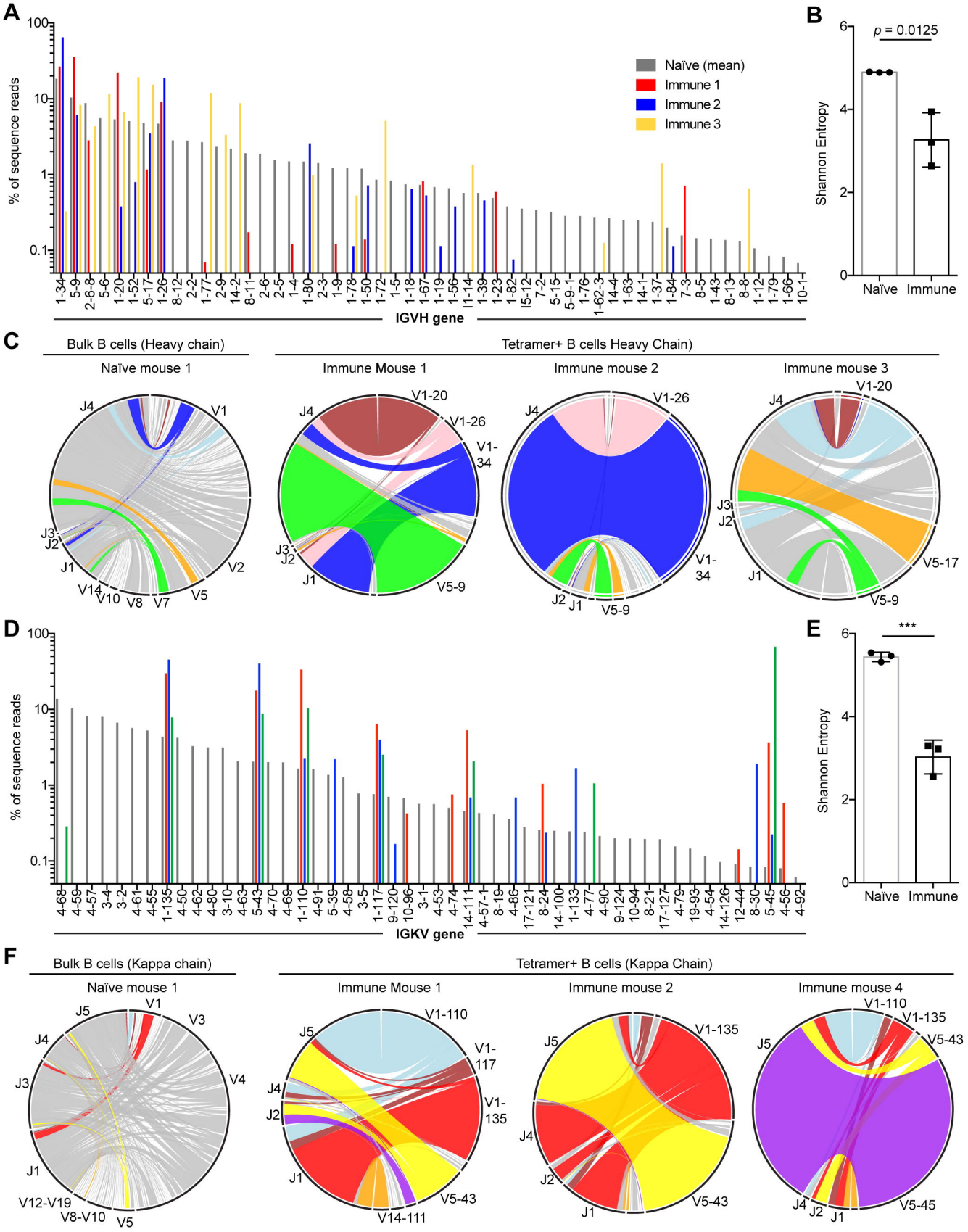

A
Time (s)
0 2000 4000 6000 8000
Power (µW)
0 -1 -2 -3 -4
Heat (µJ)
0 -20 -40 -60 -80 -100
0.0 0.5 1.0 1.5 2.0 2.5
[(NANP)6]/[FAB]
B
Time (s)
0 2000 4000 6000 8000
Power (µW)
0.0 -0.2 -0.4 -0.6
Heat (µJ)
0 -5 -10 -15
0.0 0.1 0.2 0.3
[rCSP]/[FAB]
C
Time (s)
0 2500 5000 7500 10000
Power (µW)
0.0 -0.2 -0.4 -0.6 -0.8 -1.0
Heat (µJ)
5 0 -5 -10 -15 -20
0.0 0.1 0.2 0.3
[rCSP]/[2A10]

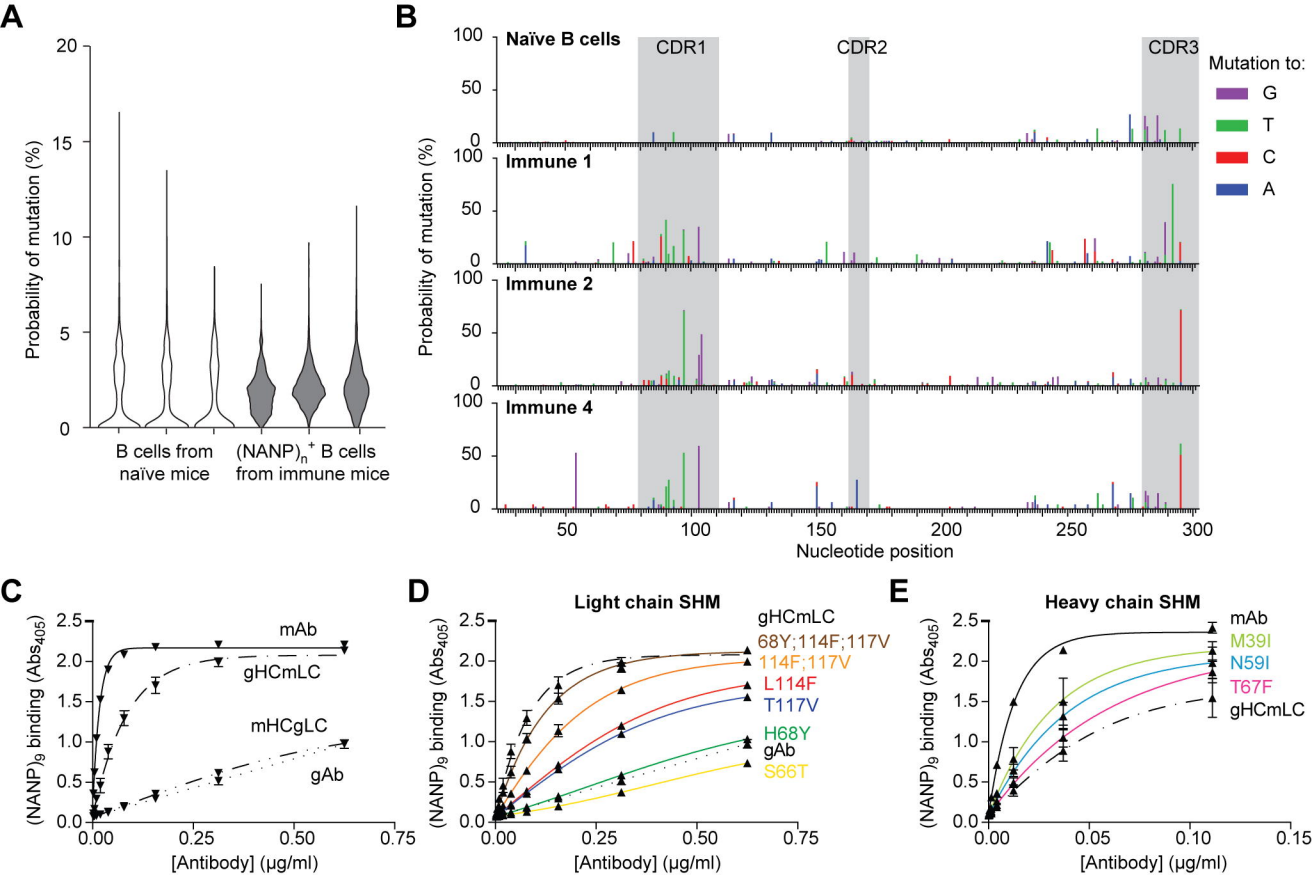

A
B
C


A
Y56
Y38
R109
Y116
S36
B
Light Chain Mutations
(NANP)9 binding (Abs405)
[Antibody] (µg/ml)
S36A
mAb
R109A
Y116F
Y56A
Y38A
C
E64
T66
N57
Y37
Y56
D
Heavy Chain Mutations
(NANP)9 binding (Abs405)
[Antibody] (µg/ml)
E64A
mAb
T66A
Y37A
N57A


A
B


A
Gated on B cells
Naïve
P. berghei CS5M
P. berghei CSPf
0.032%
0.031%
0.27%
Tet+ cells
Tet-PE
Tet-APC
B
Gated on Tet+ B cells
Naïve
P. berghei CS5M
P. berghei CSPf
Switched cells
1.35%
4.67%
82.0%
IgM
IgD
GC B cells
0%
1.34%
55.4%
CD38
GL7
C
Switched Tet+ cells/10^6 B cells
****
****
ns
Naïve
P. berghei CS5M
P. berghei CSPf
D
GC Tet+ cells/10^6 B cells
****
****
ns
Naïve
P. berghei CS5M
P. berghei CSPf


A
% of sequence reads
Naïve (mean)
Immune 1
Immune 2
Immune 3
IGVH gene
B
p = 0.0125
Shannon Entropy
Naïve
Immune
C
Bulk B cells (Heavy chain)
Tetramer+ B cells Heavy Chain)
Naïve mouse 1
Immune Mouse 1
Immune mouse 2
Immune mouse 3
D
% of sequence reads
IGKV gene
E
***
Shannon Entropy
Naïve
Immune
F
Bulk B cells (Kappa chain)
Tetramer+ B cells (Kappa Chain)
Naïve mouse 1
Immune Mouse 1
Immune mouse 2
Immune mouse 4


A
Probability of mutation (%)
B cells from naïve mice
(NANP)n+ B cells from immune mice
B
Naïve B cells
Immune 1
Immune 2
Immune 4
CDR1
CDR2
CDR3
Mutation to:
G
T
C
A
Probability of mutation (%)
Nucleotide position
C
mAb
gHCmLC
mHCgLC
gAb
(NANP)9 binding (Abs405)
[Antibody] (μg/ml)
D
Light chain SHM
gHCmLC
68Y;114F;117V
114F;117V
L114F
T117V
H68Y
gAb
S66T
(NANP)9 binding (Abs405)
[Antibody] (μg/ml)
E
Heavy chain SHM
mAb
M39I
N59I
T67F
gHCmLC
(NANP)9 binding (Abs405)
[Antibody] (μg/ml)