

1 **Title:** Genetic diversity in the Plasmodium falciparum next-generation blood stage
2 vaccine candidate antigen PfCyRPA in Senegal

3

4 **Authors & contributions:** Aboubacar Ba¹, Laty Gaye Thiam¹, Mariama Nicole
5 Pouye¹, Yicheng Guo⁴, Saurabh D. Patel⁵, Seynabou Diouf Sene¹, Fatoumata
6 Diallo¹, Rebecca Li², Awa Cisse², Noemi Guerra², Safia Laqqa², Khadidjatou
7 Mangou¹, Adam J. Moore^{2, 6}, Bacary Djilocalisse Sadio¹, Jean Louis Abdourahim
8 Ndiaye³, Alassane Mbengue¹, Zizhang Sheng⁴, Lawrence Shapiro^{4, 7}, Amy K. Bei^{1, 2}

9

10 **Author affiliations:**

11 1- G4 Malaria Experimental Genetic Approaches & Vaccines, Pôle
12 Immunophysiopathologie et Maladies Infectieuses, Institut Pasteur de Dakar,
13 Dakar, Senegal.

14 2- Department of Epidemiology of Microbial Diseases, Yale School of Public
15 Health, New Haven, CT, USA.

16 3- University Iba Der THIAM of Thies and the University Cheikh Anta Diop of
17 Dakar, Senegal.

18 4- Aaron Diamond AIDS Research Center, Columbia University Vagelos College
19 of Physicians and Surgeons, New York, NY, USA.

20 5- Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY,
21 USA.

22 6- Department of Pathology, Microbiology, and Immunology, School of Veterinary
23 Medicine, University of California Davis, Davis, CA, USA.

24 7- Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY,
25 USA.

26

27

28

29

30

31

32

33

34

35 **Abstract**

36 The *Plasmodium falciparum* cysteine-rich protective antigen (PfCyRPA) is a
37 promising target as a next-generation blood-stage malaria vaccine and together with
38 PCRCR complex members, the reticulocyte binding-like homologous protein 5
39 (PfRh5) and the Rh5-interacting protein (PfRipr), are currently being evaluated in
40 clinical trials. PfCyRPA is essential for merozoite invasion and appears to be highly
41 conserved within the *P. falciparum* parasite populations. Here, we used a targeted
42 deep amplicon next-generation sequencing approach to assess the breadth of
43 PfCyRPA genetic diversity in 95 *P. falciparum* clinical isolates from Kédougou, an
44 area with a high seasonal malaria transmission in Senegal. Our data show the
45 dominant prevalence of PfCyRPA wild type reference allele, while we also identify a
46 total of 15 single nucleotide polymorphisms (SNPs). Of these, only five have
47 previously been reported, while the majority of the SNPs were present as singletons
48 within our sampled population. Moreover, the variant read frequency of the identified
49 SNPs varied from 2.6 to 100%, while the majority of the SNPs were present at
50 frequencies greater than 25% in polygenomic samples. We also applied a structure-
51 based modelling approach to thread these SNPs onto PfCyRPA crystal structures
52 and showed that these polymorphisms have different predicted functional impacts on
53 the interactions with binding partner PFRH5 or neutralizing antibodies. Our prediction
54 revealed that the majority of these SNPs have minor effects on PfCyRPA antibodies,
55 while others alter its structure, stability, or interaction with PFRH5. Altogether, our
56 present findings reveal conserved PfCyRPA epitopes which will inform downstream
57 investigations on next-generation structure-guided malaria vaccine design.

58

59 Introduction

60 Malaria is caused by parasites of the genus *Plasmodium spp.* and remains a major
61 cause of morbidity and mortality, especially in Africa which bears the brunt of over
62 90% of the disease burden. Despite being a preventable and curable disease,
63 malaria remains a global public health burden with an estimated 608,000 deaths and
64 an associated mortality rate of 14.3 deaths per 100,000 population at risk in 2022.
65 Combined efforts in both preventive and therapeutic measures have significantly
66 reduced the malaria burden over the last two decades. However, this fragile progress
67 has reversed in recent years with the emergence and spread of both insecticide-
68 resistant mosquitos and the antimalarial resistant parasites [1]. This emphasizes the
69 urgent need to accelerate the development of highly effective vaccines against the
70 human malaria parasites [2] which will further support current control measures to
71 reduce the incidence of this disease in endemic countries and strive towards malaria
72 elimination. Malaria vaccine development strategies have recently achieved a
73 milestone following the WHO's recommendation of the RTS, S/AS01 and R21/Matrix-
74 M (R21) malaria vaccines for the prevention of *P. falciparum* malaria in children living
75 in regions with moderate to high transmission [1, 3, 4]. Both RTS,S and R21 target
76 the circumsporozoite protein of the *P. falciparum*'s liver stage and have extensively
77 been evaluated in clinical trials and pilot implementation for RTS,S. Primary analysis
78 of R21 phase 3 clinical trial data showed protection of 67-75% against multiple
79 clinical malaria episodes after a 12-month follow-up of fully vaccinated children (5-36
80 months) [4], while that of RTS, S vaccine is limited by its modest efficacy, as
81 demonstrated in the large phase 3 clinical trial across eight African countries where
82 efficacy was 55.8% in children aged 5–17 months was observed over first year, and
83 waned to 18.3% and 28.2% in infants (6-12 weeks) and children (5-17 months),
84 respectively over 48 months of follow-up [5, 6]. There is an opportunity to
85 complement these first-generation with next-generation vaccines, preferably targeting
86 other stages of the parasite's life cycle, to complement the existing malaria vaccine
87 toolbox. Such vaccines need to consider genetic diversity at the very early stage of
88 development. Malaria vaccine development has tremendously benefited from the
89 publication of the genome of the *Plasmodium falciparum* [7], which has paved the
90 way for malaria reverse vaccinology [8]. This approach enabled the prioritization of
91 current lead blood-stage malaria vaccine *P. falciparum* reticulocyte binding homolog
92 5 (PfRh5), the terminal member of the PCRCR complex [9] that binds to erythrocyte
93 receptor Basigin [10]. In addition to PfRh5, members of this complex include the Rh5-
94 interacting protein (PfRipr), Cysteine-rich protective antigen (PfCyRPA), the
95 *Plasmodium* thrombospondin-related apical merozoite protein (PfPTRAMP), and the
96 cysteine-rich small secreted protein (PfCSS) [9, 11]. Of these, PfRh5 remains the
97 most advanced antigen of the complex in clinical development, having recently
98 completed Phase 2b clinical trials in Burkina Faso, while PfCyRPA and PfRipr are
99 currently being assessed in phase 1 clinical trials (NCT0538547) [12]. Our present
100 study evaluates the breadth of genetic diversity of PfCyRPA and uses structural

101 insights to predict the functional impact of such diversity, contributing to structure-
102 guided vaccine development.

103 **Results**

104 **Characteristics of study participants**

105 This study was conducted in Kedougou, a Southeastern region of Senegal, with a
106 seasonal malaria transmission from May to November. The study protocol was
107 approved by National Ethics Committee of Senegal (CNER) (SEN19/36 and
108 SEN23/09), the regulatory board of the Senegalese Ministry of Health and the
109 Institutional Review Board of the Yale School of Public Health (2000025417).
110 Informed consent was obtained from study participants and/or their legal guardians.
111 A total of 94 patients presenting confirmed cases of symptomatic *P. falciparum*
112 infection and recruited in 2019 and 2022 from five healthcare centres in Kédougou,
113 Bandafassi (N = 21), Camp militaire (N = 23), Dalaba (N = 33), Mako (N = 13) and
114 Tomboronkoto (N = 4). Table 1 summarizes the demographic and parasitological
115 characteristics of the study participants. Participants enrolled for this study were aged
116 2 to 67 years (Median 21.75; SD = 13.11), and there were 59 males and 35 females.
117 We observed significantly different sex ratios, with an overall sex ratio of 1.68 in
118 favour of the males. While no significant difference was observed in the median age
119 across sampling sites, we observed a lower proportion of children (≤ 10 years) across
120 sites. Moreover, we observed an overall complexity of infection (COI) of 3.65 our
121 study population (**Table 1**).

122 **Prevalence of SNPs**

123 To determine the degree of PfCyRPA-associated genetic diversity within the
124 population, we employed targeted deep amplicon sequencing using Illumina short-
125 read next-generation sequencing on a NovaSeq6000 platform. Genetic diversity was
126 assessed using a very sensitive threshold (2% variant allele frequency) and applied
127 both qualitative and quantitative metrics to enable accurate SNP discovery and
128 validity. A total of 93 isolates were included in this analysis and the resulting
129 sequences were compared to that of the reference strain (3D7). Overall, 26/93 (28%)
130 of the isolates carried at least one SNP in the PfCyRPA gene relative to the 3D7
131 reference, which represented the dominant allele 67/93 (72%) within our sampled
132 population (**Fig. 1A**). We identified 15 individual SNPs, of which only five (F41L,
133 V165I, D236N, N270T and V292F) have previously been reported. Additionally, of
134 the novel SNPs reported here, 2 were previously described at the same position but
135 we observed different amino acid substitutions (D236N and N338D). The majority of
136 the novel SNPs were rare and only identified in a single isolate, except for R50C,
137 I196F and K211Q, which were all found in two isolates each. Overall, most of the
138 isolates with a mutant allele carried only a single SNP at a time, while the highest
139 number of SNPs carried by a single isolate (320697) was 4 (R31H; F41L; D236N;
140 V292F). Moreover, the majority of the SNPs reported in this study were rare variants,
141 as only one (V292F) was detected at a prevalence greater than 5%. Of the novel

142 SNPs detected in the study, only two reached a prevalence of at least 2% in the total
143 population I196F (2.2%) and K211Q (2.2%), while the remaining SNPs were all
144 detected at a prevalence of 1.1%, corresponding to a single sample (**Fig. 1A**).

145 **Variant frequency of individual SNPs**

146 Given the likelihood of mixed-genotypes infections associated with natural malaria
147 infections in high transmission settings, we measured the complexity of infection
148 (COI) in our sample using *msp1* & 2 genotyping [13]. Overall, only 10 out of the 93
149 isolates reported here were from monogenomic infections, while the number of
150 polygenomic infections ranges from 1 to 11, with a mean COI of 4 genotypes per
151 isolate. Consequently, we next sought to assess range of variant allele frequencies of
152 these SNPs within these complex infections. This analysis showed that the identified
153 SNPs were distributed at varying frequencies ranging from 2.6 to 100% within the
154 individual isolates; hence the defined classification as low frequency (<5%),
155 intermediate frequency (5–25%) and high frequency (> 25%) SNPs (**Fig. 1B**).
156 Interestingly, 10 out of the 15 SNPs reported here were present at high variant
157 frequencies within the patient sample, while only two SNPs were present at low
158 frequencies. Of these SNPs with high variant frequencies, five were novel (D236V,
159 N338D, D110N, I196F and R31H), while all the previously reported SNPs were
160 present at high frequencies. This increases the confidence that even though rare in
161 the population, these represent SNPs. Two of the novel SNPs (R50C and T37A)
162 were present at low frequencies, while F187L, K211Q and I114V were present at
163 intermediate frequencies (**Fig. 1B**). Out of the 10 isolates with monogenomic
164 infections, only three carried a mutation on *pfcyrpa*, which is present as a single SNP
165 per isolate. Of these SNPs, only I114V (400133) was present at low frequency, while
166 V292F (400116) and N338D (400115) were both present at a high frequency. The
167 majority of the novel SNPs (8/10) described here are present within polygenomic
168 infections.

169 **Structural modelling of SNPs**

170 We assessed the predicted functional impact of identified polymorphisms by
171 threading the SNPs onto the crystal structure of PfCyRPA, a six-bladed β -propeller
172 protein [14, 15]. These blades, interconnected by loop regions, are each constructed
173 by a four-stranded anti-parallel β -sheet[14][15]. The detailed interactions between
174 PfCyRPA and its binding partners (RH5 and Ripr) or monoclonal antibodies have
175 more recently been reported [11, 16]. Consequently, we used a structure-guided
176 approach to thread the identified SNPs onto the PfCyRPA crystal structure in
177 complex with PfRH5 and earlier characterized mAbs (**Fig. 2A-B**). The threading
178 complex was built by superimposing the structure complexes of PfRH5 bound to its
179 receptor Basigin (PDB id: 4U0Q) or PfCyRPA bound to PfRH5 (PDB id: 6MPV), or
180 mAb Fab fragments PfCyRPA-Cy.003 (PBD:7PI2), PfCyRPA-Cy.004 (PBD:7PHW),
181 PfCyRPA-Cy.007 (PBD:7PHV) and PfCyRPA-8A7 (PBD: 5TIH). The structural
182 threading analysis revealed an even distribution of the identified SNPs between the

183 PfCyRPA internal loops and individual blades (**Table 2**). Of these mutations, four
184 (K211Q, D236N, D236V and N270T) were located within the blade 4. Likewise, blade
185 6 also harboured four SNPs (R31H, T37A, F41L and N338D) while, three (V165I,
186 F187L and I196F) and two SNPs (D110N and I114V) were respectively located on
187 blade 3 and blade 2. However, the blades 1, and 5 each carries a single mutation
188 (**Fig. 2A-B**). Furthermore, our analysis showed four groups of SNPs with different
189 predicted functional outcomes. Interestingly, 7 out of the 15 SNPs reported here are
190 predicted to have minor effect on PfCyRPA structure or binding with PfRH5 or
191 potentially on monoclonal antibody interactions, while another group of SNPs (V165I,
192 I196F, D236N, N270T and V292F) is predicted to partially alter the structure of
193 PfCyRPA. While both D236 and N270 form hydrogen bonds with S233 and N218
194 residues, respectively, the D236V and N270T mutations alter the PfCyRPA structure
195 flexibility or stability, respectively through steric clashes or interruption of hydrogen
196 bonds. Another group of SNPs, predicted to affect PfCyRPA interaction with PfRH5
197 and included R50C, F187L and I196F, the former improving binding to PfRH5 by
198 removing the repulsion between R50 and K504 residues (**Table 2**). Intriguingly, 8
199 mutations may destabilize PfCyRPA (Table 1). One key implication of structure-
200 guided vaccine design relative to the next-generation blood-stage malaria vaccines is
201 to decipher the functional implication of the vaccine candidate-associated genetic
202 diversity to epitope-paratope interactions. We identified a subset of SNPs with a
203 potential impact on inhibitory monoclonal antibody binding, although these
204 interactions are predicted to be very mild and they are not directly in the epitope
205 bound by Cy.003, Cy.004, and Cy.007. Of these SNPs, only three (T37A, F41L,
206 N338D) were mapped to blade 6, which along with blade 2 have been shown to
207 trigger the most inhibitory antibodies [16], while R50C, D110N and I114V bound to
208 loop regions within blade 1 and 2, it is possible that these SNPs impact antibody
209 recognition through structural changes (**Fig. 2C-F**), but such predictions should be
210 functionally validated.

211 **Discussion:**

212 The *P. falciparum* cysteine-rich protective antigen (PfCyRPA) plays a crucial role in
213 merozoite invasion of the human erythrocyte. This antigen has attracted a particular
214 attention as a promising vaccine candidate, as it is essential[17, 18] and accessible
215 to naturally derived human antibodies[19]. Preclinical studies have shown that
216 PfCyRPA induces broadly neutralizing antibody response [15, 16, 20, 21], with a
217 relatively conserved sequence in the various malaria parasites[22, 23], suggesting
218 that a vaccine based on this protein may offer broader protection. However, despite
219 its general conservation, PfCyRPA has some genetic variability, which could limit the
220 effectiveness of a vaccine, as genetic variations in PfCyRPA may allow the parasite
221 to mutate and evade the vaccine-induced immune response. Together with PfRipr,
222 PfCyRPA has recently entered Phase 1 clinical testing (NCT05385471), thus a better
223 understanding of the breadth and functional impact of PfCyRPA-associated
224 polymorphisms in vaccine-induced immune response is needed to prioritize, design
225 and optimize PfCyRPA-based vaccine alleles.

226 This study was undertaken to assess the extent of PfCyRPA genetic diversity in *P.*
227 *falciparum* clinical isolates from naturally infected individuals in high malaria
228 transmission settings. Samples reported here were collected from patients diagnosed
229 with *P. falciparum* infections visiting healthcare centres in Kédougou, a Southeastern
230 region of Senegal with high seasonal malaria transmission[24]. A previous study by
231 Ndigwa and colleagues reported an excess of rare variants in proteins within the
232 PfRH5 complex, including PfCyRPA[23]. In this study, the authors used two
233 sequencing strategies, namely capillary Sanger sequencing and whole genome
234 sequencing (WGS), which respectively identified 4 and 10 PfCyRPA-associated
235 SNPs, while only a single SNP was concomitantly discovered by both strategies[23].

236 Long read sequencing strategies such as Sanger sequencing enable the manual
237 identification of genetic variation and the haplotype calling; however, they are limited
238 by both their overall low throughput and their inability to accurately identify and
239 segregate SNPs in the context of polygenomic infections such as those common in
240 high transmission settings like Kédougou. We previously reported on the high
241 prevalence of polygenomic infections in Kédougou [25]. This trend was confirmed in
242 this current study, with isolates harboring 1 to 11 genotypes, while the mean COI
243 reported here is 4 genotypes per isolate.

244 To increase our chance of discovering newly emerging and rare PfCyRPA-associated
245 variants, we opted for a targeted deep amplicon sequencing using the Illumina
246 Novaseq 6000 sequencing technology and used a sensitive discovery threshold of
247 2% for variant calling. We successfully sequenced *pfcyrpa* amplicons from 93
248 isolates and reported a total of 15 SNPs, of which 10 were novel, while only 5 were
249 reported in previous studies [23, 26]. Interestingly, our current data showed the
250 PfCyRPA reference allele being the most prevalent allele, while the opposite trend
251 was observed for PfRH5[25, 27]. This observation aligns with previous report
252 suggesting a stronger balancing selection pressure on PfRH5 than that on PfCyRPA
253 [19]. Additionally, our findings matched previous reports on the occurrence of an
254 excess of rare variants[23], as the majority of the SNPs reported here were present
255 as singletons (occurring in single isolates), while only three SNPs (V292F, D236N
256 and N270T) were present in more than two isolates. This result emphasizes the
257 power of deep amplicon sequencing strategies in identifying rare genomic variants in
258 polyclonal infections and agrees with previous reports involving the lead blood-stage
259 malaria vaccine candidate, PfRH5[25, 27].

260 One limitation of the deep amplicon sequencing strategy used here is its inability to
261 resolve individual parasite haplotypes due to the short reads but also to the high
262 complexity of infection in this population. Consequently, for each identified SNP, we
263 assessed the variant read frequency, defined as the percentage of variant reads in
264 the total reads mapped to a given position in the PfCyRPA reference. Given the
265 varying number of genotypes as well as their respective parasitemia in a given
266 isolate, this analysis of the variant reads frequencies enabled us to quantitatively
267 calculate the number of reads with the SNP relative to the total number of reads at its

268 position and therefore classify the SNPs as low (<5%), intermediate (5-25%) and
269 high (>25%) in each given sample. Interestingly, 10 out of the 15 SNPs reported here
270 were present at high frequencies, while 1 and 4 SNPs were respectively present at
271 intermediate and low frequencies. A similar observation was made in our previous
272 study on PfrH5[25], which resulted in a number of SNPs present at low frequencies.
273 While all previously reported SNPs, the most prevalent in the population, were also
274 present at high frequencies, we also showed 4 novel SNPs (R31H, D110N, D236V
275 and N338D) present at high frequencies, each of which were identified as singletons.
276 Of these novel SNPs present at high frequencies, three (R31H, D110N and D236V)
277 occurred in polygenomic infections. Interestingly, the D236V SNP, present as a
278 singleton with a frequency of 99.8% emerged from an isolate with a COI of 4, which
279 emphasizes a hypothesis to further test that the functional implication of this SNP
280 could increase the parasite's fitness.

281 Given the relationship between a protein's structure and its function, we sought to
282 investigate the impact of the identified SNPs in PfCyRPA structure and ultimately
283 predict their functional implication in binding with partner protein PfrH5 or its
284 recognition by neutralizing monoclonal antibodies with known binding epitopes. The
285 crystal structure of PfCyRPA has earlier been solved[14, 15], while more recent
286 studies have solved its structure in complex with neutralizing antibodies as well as
287 binding partners PfrH5 and PfrIpr[11, 16]. The functional implication of naturally
288 arising polymorphisms however might be very challenging to investigate within the
289 naturally circulating parasite populations, and mostly in the context of high malaria
290 transmission where individual isolates are often represented as polygenomic
291 infections.

292 As a primary investigation, we adopted an in-silico approach based on the threading
293 of the observed SNPs onto the crystal structure of PfCyRPA in complex with binding
294 partner PfrH5 or neutralizing antibodies. By superimposing the identified SNPs onto
295 the PfCyRPA structure, we were able to accurately map their distribution and predict
296 their impact on the antigen's functional structure. Interestingly, in addition to the even
297 distribution of the SNPs between the antigen's internal loops and blades, there was
298 at least one SNP present in each given blade. Moreover, out of the 15 SNPs located
299 within PfCyRPA blades, 7 were located within blades 3 and 4, which together form
300 much of the interface between PfCyRPA-PfrH5 interaction[11]. As these SNPs were
301 predicted to have either a minor effect on the antigen's structure or to impact its
302 binding with PfrH5, their functional impact remains a mystery to be solved
303 considering that their location might not be readily accessible to neutralizing
304 antibodies but also given that antibodies occurring in the binding interface between
305 the two antigens are not inhibitory[16, 28]. On the other hand, while a previous study
306 reported the predominance of conformational neutralizing epitopes within the
307 PfCyRPA structure [21], recent data have shown the most of the inhibitory antibody
308 binding epitopes to be located within the blades 1 and 2 of the PfCyRPA
309 structure[16]. Our structural threading analysis showed R50C to be located within
310 blade 1, while both D110N and V165I were located within the blade 2 of PfCyRPA,

311 with both predicted to have a minor effect on PfCyRPA structure.. Moreover, 4 out of
312 the 15 SNPs (R50C, F187L, I196F, N270T) reported here were predicted to have a
313 minor effect on antibody binding to PfCyRPA. However, even if there seems to be no
314 SNPs found within the most critical epitopes of the PfCyRPA so far reported, this
315 should not prevent further investigation on the potential impact of these SNPs, as
316 even though predicted with a minor effect, they could have an important contribution
317 in the parasite overall fitness. Finally, our predictions also reported on the presence
318 of SNPs that can impact the antigen's functionality in other ways, such as removing
319 repulsive interactions (R50C), disrupting hydrogen bonds (N270T) or causing steric
320 clashes (D236V), while another subset of SNPs has the propensity of either altering
321 (N338D) or introducing (D236N) new glycosylation sites. Given the importance of
322 their structural changes and their key roles in protein stability, folding and solubility,
323 these findings warrant further investigation in order to confirm the functional impact of
324 these SNPs in the context of malaria vaccinology as it relates to PfCyRPA.

325 Despite the relevance of the reported data from this study relative to the importance
326 of PfCyRPA as a malaria vaccine candidate, there is still room for improvement as it
327 relates to the strategy herein described. The cross-sectional approach described
328 here provides a snapshot of the *pfcyrpa* genetic diversity within the circulating
329 parasite populations. The passive recruitment strategy adopted here could prompt
330 the tendency to only focus on isolates driving the more prominent clinical disease that
331 influences the patient to seek care while overlooking the true natural genetic diversity
332 in isolates from the larger community. As we highlighted in our previous study, future
333 work can address this by sampling across the clinical presentation spectrum,
334 including active surveillance of asymptomatic cases[25]. Another limitation of this
335 study is the notable differences in the number of samples from the individual sites,
336 which also did not enable a site by site comparison of the *pfcyrpa* genetic diversity
337 across the sampling sites. This study was not powered to reflect a thorough
338 assessment of genetic diversity stratified by site, although this would be interesting to
339 explore in depth as there could be differences in the parasite populations circulating
340 within each site due to unique epidemiological characteristics. This could be as a
341 result of their geographical location related to the neighbouring countries with whom
342 the region shares borders (Mali and Guinea Conakry) due to the specific activities of
343 each site (mining and trading activities). Therefore, a larger and more thorough
344 sampling across the entire region could strengthen these preliminary data being
345 reported here. Furthermore, while our sequencing approach has advantages for deep
346 coverage and detection of rare SNPs, it also has its own limitations as the generated
347 short reads, combined with the high complexity of infection in this population, make it
348 difficult to accurately resolve individual parasite haplotypes. Finally, although our
349 structural modelling can imply potential functional impact of the reported SNPs, more
350 functional studies are needed to accurately decipher, if any, the true mechanistic
351 implication of these polymorphisms into the parasites' fitness and survival. Given the
352 working hypotheses generated from our investigations, this study, will inform

353 downstream biochemical and functional genetic approaches to evaluate the role of
354 each SNP in PfCyRPA function and survival strategies.

355

356 **Data Availability**

357 Sequencing Reads associated with this study have been deposited in the NCBI SRA
358 with the BioProject Accession: PRJNA1109877.

359

360 **Author Contributions**

361 A.K.B. conceived the experiments. A.K.B., L.S., and Z.S. supervised the research.
362 A.B., L.G.T., M.N.P., S.D.S., F.D., R.L., A.C., N.G., K.M., A.J.M, A.T., B.D.S., and
363 A.M. collected the samples. A.K.B and A.B. assisted with geolocation. A.B., L.G.T.,
364 M.N.P. S.D.S., and F.D. conducted the experiments. Y.G., Z.S., S.D.P. performed
365 structure modelling. A.B., L.G.T., S.L. N.G., and A.K.B. analysed the results. A.B.,
366 L.G.T., A.K.B. wrote the manuscript. A.B., L.G.T., A.K.B. J.L.A.N., Z.S., S.D.P. and
367 L.S. reviewed and edit the manuscript. All authors reviewed and approved the final
368 manuscript.

369 **Funding Statement**

370 This work was supported by the by the Fogarty International Center of the NIH (K01
371 TW010496), National Institute of Allergy and Infectious Diseases of the NIH (R01
372 AI168238), and G4 group funding (G45267, Malaria Experimental Genetic
373 Approaches & Vaccines) from the Institut Pasteur de Paris and Agence Universitaire
374 de la Francophonie (AUF) to AKB. This work has been produced with the financial
375 assistance of the European Union (Grant no. DCI-PANAF/2020/420-028), through the
376 African Research Initiative for Scientific Excellence (ARISE), pilot program. ARISE is
377 implemented by the African Academy of Sciences with support from the European
378 Commission and the African Union Commission. The contents of this document are
379 the sole responsibility of the authors and can under no circumstances be regarded as
380 reflecting the position of the European Union, the African Academy of Sciences, and
381 the African Union Commission. AB and LGT are partially supported by an ARISE
382 grant from the African Academy of Sciences (ARISE-PP-FA-056).

383

384 **Acknowledgements**

385 We would like to thank Souleymane Ngom from Dalaba, Lt. Dr Charles Latyr Diagne
386 and Lamine Kane from Camp Militaire, Moctar Mansaly and Gerald Keita from
387 Bandafassi, Safietou Sane and Astou Ndiaye from Mako, Adama Gueye from
388 Bantaco, Der Ciss from Tomboronkoto and all the healthcare workers at these sites
389 for their partnership with Institut Pasteur Dakar. We would also like to thank the
390 people of Kédougou for their invaluable contributions to this work. We would like to

391 thank the Yale Center for Genome Analysis (YCGA). We would like Dr Ines VIGAN-
392 WOMAS for her help and assistance in Immunophysiopathology and Infectious
393 diseases Unit at Institut Pasteur Dakar.

394

395 **Methods**

396 **Study Sites and sample collection**

397 This study was conducted in Kedougou, a Southeastern region of Senegal, with a
398 seasonal malaria transmission from May to November. Informed consents were
399 obtained from the study participants or their legal guardians and samples were
400 collected following the approved ethical protocol by the National Ethics Committee of
401 Senegal (CNER) (SEN19/36), the regulatory board of the Senegalese Ministry of
402 Health and the Institutional Review Board of the Yale School of Public Health
403 (2000025417). Samples used in this study were collected through passive case
404 detection from patients visiting healthcare facilities Bandafassi, Bantaco, Camp
405 militaire, Dalaba, Mako, and Tomboronkoto in 2019 and 2022, during the peak of the
406 malaria transmission season (July and August) with malaria-like symptoms. If
407 participants met the enrollment criteria of fever in the past 24 h, an axillary
408 temperature $\geq 38^{\circ}\text{C}$ and/or a positive *P. falciparum* malaria diagnosis from a rapid
409 diagnostic test (RDT) and microscopy, they were offered the opportunity to enroll in
410 the study. After informed consent was obtained, a venous blood sample was drawn
411 into EDTA vacutainers and samples were transported at room temperature to the
412 laboratory for processing; no more than 6 hours between draw and processing.

413 **DNA extraction, PCR amplification and NGS Library & Sequencing**

414 DNA was extracted from infected erythrocyte pellets using the ZYMO Quick-DNA
415 Miniprep Kit (D3024) following the manufacturer's instructions. The extracted DNA
416 samples were eluted in 30ul of nuclease free water and stored at -20°C prior to PCR
417 amplification. For PCR amplification, *PfCyRPA*-specific primers were designed using
418 the Geneious Prime software version 23.1.1. The *PfCyRPA* 3D7 reference sequence
419 (PF3D7_0423800, PlasmoDB [29]) was used as template for primer designing and
420 the amplification was performed using a classic PCR protocol. The PCR was done by
421 using Phusion® High-Fidelity DNA Polymerase (Catalog: M0530L, 50X higher fidelity
422 than Taq). **Supplemental table 1** shows the primer pairs and the **Supplemental**
423 **Table 2 and 3** the PCR conditions and PCR program respectively used for *PfCyRPA*
424 amplification. Following successful amplification, *PfCyRPA* sequences were bead-
425 purified (Omega) and quantified using a Qubit 2.0 fluorometer; and subsequently
426 adjusted to equivalent concentration. Sequencing library preparation was performed
427 with the Nextera XT using unique dual indexes (UDIs) and subjected to a subsequent
428 bead-purification. DNA libraries were quantified by qPCR using Roche KAPA Library
429 Quantification Kit. All samples were normalized to a final concentration of 4 nM. The
430 96 samples quantified and normalized were pooled into 8-sub-pools, which were

431 further bead-purified and quantified using a KAPA qPCR. The 8 sub-pools were
432 further normalized and combined in equal quantities to form one final pool. This final
433 pool was sent to the Yale Center for Genome Analysis (YCGA) for sequencing on an
434 Illumina NovaSeq 6000 platform with targeted coverage of 500,000 reads per
435 sample.

436 **Data processing and polymorphism analysis**

437 De-multiplexed forward and reverse sequencing reads obtained for each sample
438 were individually imported to Geneious Prime and paired sequences were obtained
439 using the Illumina paired end setting. Paired sequences were subsequently trimmed
440 using BBDuk plugin. A minimum quality score (Q) of 30 was set for the trimming with
441 a minimum length of 75 base pairs, as we were expecting reads around 150 base
442 pairs. Trimmed sequences were aligned with the 3D7 reference sequence that had
443 been annotated with all known non-synonymous mutations. Single nucleotide
444 polymorphisms (SNPs) annotation was performed using five iterations the criteria for
445 SNP calling was set to a minimum frequency of 0.02 (2%) and 1000 read coverage.
446 Sequence data and SNP analysis was performed by at least 3 individuals for each
447 sample.

448 **Structural modelling of PfCyRPA-associated SNPs**

449 The structures of PfCyRPA and PfRH5 were downloaded from the Protein Data Bank
450 (PDB, [https:// www. rcsb. org/](https://www.rcsb.org/)). PfRh5-PfCyRPA complex was constructed using
451 Pymol (PDB ID: 4U0Q and 6MPV) and structural predictions with mAb binding was
452 performed using PDB IDs 5TIH, 7PI2, and 7PHW). Individual FASTA files containing
453 amino acid sequences of PfCyRPA and individual novel SNPs were generated.
454 These amino acid sequence files were threaded onto the crystal structure of
455 PfCyRPA in complex with binding partner PfRH5 and/or known monoclonal
456 antibodies. Pymol version 2.3.2 was used to predict the effect and to plot the
457 structural location of each SNPs. The structural effect of the mutant versions of the
458 protein were evaluated in terms of biochemical properties such as hydrogen bonding
459 patterns, steric interactions, and predicted binding affinity between the mutant version
460 of the protein and the Basigin receptor the binding energy alternation for SNPs were
461 predicted by FoldX.

Reference:

1. WHO, *World malaria report 2023*, W.H. Organization, Editor. 2023, Geneva: World Health Organization: <https://apps.who.int/iris>. p. 356.
2. Draper, S.J., et al., *Malaria Vaccines: Recent Advances and New Horizons*. Cell Host Microbe, 2018. **24**(1): p. 43-56.
3. Rts, S.C.T.P., *Efficacy and safety of the RTS,S/AS01 malaria vaccine during 18 months after vaccination: a phase 3 randomized, controlled trial in children and young infants at 11 African sites*. PLoS Med, 2014. **11**(7): p. e1001685.
4. Dattoo, M.S., et al., *Safety and efficacy of malaria vaccine candidate R21/Matrix-M in African children: a multicentre, double-blind, randomised, phase 3 trial*. Lancet, 2024. **403**(10426): p. 533-544.
5. Rts, S.C.T.P., et al., *A phase 3 trial of RTS,S/AS01 malaria vaccine in African infants*. N Engl J Med, 2012. **367**(24): p. 2284-95.
6. Rts, S.C.T.P., et al., *First results of phase 3 trial of RTS,S/AS01 malaria vaccine in African children*. N Engl J Med, 2011. **365**(20): p. 1863-75.
7. Gardner, S.N., *Cell cycle phase-specific chemotherapy: computational methods for guiding treatment*. Cell Cycle, 2002. **1**(6): p. 369-74.
8. Tuju, J., et al., *Vaccine candidate discovery for the next generation of malaria vaccines*. Immunology, 2017. **152**(2): p. 195-206.
9. Scally, S.W., et al., *PCRCR complex is essential for invasion of human erythrocytes by Plasmodium falciparum*. Nat Microbiol, 2022. **7**(12): p. 2039-2053.
10. Crosnier, C., et al., *Basigin is a receptor essential for erythrocyte invasion by Plasmodium falciparum*. Nature, 2011. **480**(7378): p. 534-7.
11. Farrell, B., et al., *The PfRRCR complex bridges malaria parasite and erythrocyte during invasion*. Nature, 2023.
12. Williams, B.G., et al., *Development of an improved blood-stage malaria vaccine targeting the essential RH5-CyRPA-RIPR invasion complex*. Nat Commun, 2024. **15**(1): p. 4857.
13. Snounou, G., et al., *Biased distribution of msp1 and msp2 allelic variants in Plasmodium falciparum populations in Thailand*. Trans R Soc Trop Med Hyg, 1999. **93**(4): p. 369-74.
14. Chen, L., et al., *Structural basis for inhibition of erythrocyte invasion by antibodies to Plasmodium falciparum protein CyRPA*. Elife, 2017. **6**.
15. Favuzza, P., et al., *Structure of the malaria vaccine candidate antigen CyRPA and its complex with a parasite invasion inhibitory antibody*. Elife, 2017. **6**.
16. Ragotte, R.J., et al., *Heterotypic interactions drive antibody synergy against a malaria vaccine candidate*. Nat Commun, 2022. **13**(1): p. 933.
17. K. Sony Reddy, E.A., Alok K. Pandey, Pallabi Mitra, Virander S. Chauhan, and Deepak Gaur, *Multiprotein complex between the GPI-anchored CyRPA with PfRH5 and PfRipr is crucial for Plasmodium falciparum erythrocyte invasion*. PNAS, 2015. **112**: p. 1179–1184.
18. Volz, J.C., et al., *Essential Role of the PfRh5/PfRipr/CyRPA Complex during Plasmodium falciparum Invasion of Erythrocytes*. Cell Host Microbe, 2016. **20**(1): p. 60-71.
19. Mian, S.Y., et al., *Plasmodium falciparum Cysteine-Rich Protective Antigen (CyRPA) Elicits Detectable Levels of Invasion-Inhibitory Antibodies during Natural Infection in Humans*. Infect Immun, 2022. **90**(1): p. e0037721.

20. Healer, J., et al., *RH5.1-CyRPA-Ripr antigen combination vaccine shows little improvement over RH5.1 in a preclinical setting*. *Front Cell Infect Microbiol*, 2022. **12**: p. 1049065.
21. Knudsen, A.S., et al., *Strain-Dependent Inhibition of Erythrocyte Invasion by Monoclonal Antibodies Against Plasmodium falciparum CyRPA*. *Front Immunol*, 2021. **12**: p. 716305.
22. Dreyer, A.M., et al., *Passive immunoprotection of Plasmodium falciparum-infected mice designates the CyRPA as candidate malaria vaccine antigen*. *J Immunol*, 2012. **188**(12): p. 6225-37.
23. Ndwiga, L., et al., *The Plasmodium falciparum Rh5 invasion protein complex reveals an excess of rare variant mutations*. *Malar J*, 2021. **20**(1): p. 278.
24. Paludisme, P.N.d.L.c.l., *BULLETIN EPIDEMIOLOGIQUE ANNUEL 2023 DU PALUDISME AU SENEGAL*. 2024.
25. Mangou, K., et al., *Structure-guided insights into potential function of novel genetic variants in the malaria vaccine candidate PfRh5*. *Sci Rep*, 2022. **12**(1): p. 19403.
26. Waweru, H., et al., *Limited genetic variations of the Rh5-CyRPA-Ripr invasion complex in Plasmodium falciparum parasite population in selected malaria-endemic regions, Kenya*. *Front Trop Dis*, 2023. **4**: p. 1102265.
27. Laty G. Thiam, K.M., Aboubacar Ba, Rebecca Li, Yicheng Guo, Mariama N. Pouye, Awa Cisse, Dimitra Pipini, Fatoumata Diallo, Seynabou D. Sene, Saurabh D. Patel, Alassane Thiam, Bacary D. Sadio, Alassane Mbengue, Inés Vigan-Womas, Zizhang Sheng, Lawrence Shapiro, Simon J. Draper, and Amy K. Bei, *Vaccine-induced human monoclonal antibodies to PfRH5 show broadly neutralizing activity against P. falciparum clinical isolates*. medRxiv, 2024.
28. Alanine, D.G.W., et al., *Human Antibodies that Slow Erythrocyte Invasion Potentiate Malaria-Neutralizing Antibodies*. *Cell*, 2019. **178**(1): p. 216-228 e21.
29. Alvarez-Jarreta, J., et al., *VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center in 2023*. *Nucleic Acids Res*, 2024. **52**(D1): p. D808-D816.

Figure Legends

Figure 1. Population prevalence of PfCyRPA SNPs. (A) The prevalence of PfCyRPA-associated SNPs was calculated as the percentage of SNPs detected within the total number of clinical samples in the population (N=93) using a variant allele frequency (VAF) threshold of 2%. PfCyRPA sequencing was performed from pfcyrpa amplicons using the Illumina NovaSeq 6000 sequencing platform and variant analysis was performed using the Geneious Prime software version 23.1.1. The graphs were plotted using the GraphPad Prism version 1.0.2 software. (B) Variant read frequency of PfCyRPA SNPs. Variant read frequency was determined from the sequencing data outputs and calculated as the percentage of the variant reads relative to the coverage at the variant position. The data are presented as bar graphs showing the number of isolates (black dots) with the error bars presenting the minimum and the maximum frequencies for each SNP in complex clinical samples. The SNPs are categorized as low <20% (golden), intermediate 2-25% (lavender) or high frequency SNPs >25% (maroon), based on their respective frequency within the individual complex sample. The dotted line depicts the 2% VAF threshold. The graphs were plotted using the GraphPad Prism version 1.0.2 software.

Figure 2: Structure-function predictions for the novel SNPs identified in CyRPA.

The complete structure was obtained by superimposing the structure of PfCyRPA in complex with PfrH5 (PDB id: 6MPV), PfrH5 bound to its ligand Basigin (PDB id: 4U0Q) and of known monoclonal antibodies Fab regions 8A7(PBD:5TIH), Cy.003 (PBD:7PI2), Cy.004 (PBD:7PHW) and Cy007/c12 (PBD:7PHV). Both CyRPA blades and individual antibodies are color coded. (A) The location of SNPs within the BSG–RH5–CyRPA complex. The complex construction was achieved by superimposing the RH5 of the RH5–BSG complex (PDB ID: 4U0Q) onto the RH5–CyRPA complex (PDB ID: 6MPV). BSG and RH5 are depicted in light blue and grey, respectively. CyRPA is represented in a wheat color, while blade 1 and blade 2 are indicated in dark orange. (B) shows the distribution of the SNPs across PfCyRPA blades and their position relative to the monoclonal antibody binding epitopes. BSG and PfrH5 ribbons are shown by light blue and grey, while the different blades of PfCyRPA are depicted in different colors. Ribbons of the 8A7, Cy.003, Cy.004 and Cy.007/c12 R5.011 are respectively shown dark green, light green, indigo and light marron. Pymol version 2.3.2 was used to predict the effect and to plot the structural location of each SNPs. (C) The positions of D110N and I114V relative to monoclonal antibodies (mAbs). Antibodies Cy.003, Cy.004, Cy.007, and 8A7 are represented in green, beige, magenta, and dark turquoise, respectively. (D) Structural modelling revealed that SNPs V165Y, N270T, and V292F influence the conformation of CyRPA. Small red plates signify the potential steric hindrances. (E) This panel highlights the SNPs that might confer resistance to antibodies, including F41L, D110N, and I114V. (F) SNPs R50C and F187L are shown to directly interact with PfrH5.

Table 1: Socio-demographic and parasitological characteristics of study participants

Sites	BF	CM	DB	MK	TM	Total	p-value
Patients, No	21	23	33	13	4	94	
Sex ratio (M/F)	1.62	10.5	1.06	0.86	1	1.68	0.02*
Age (median, years)	19.61	20.09	25.31	18.3	25.25	21.75	ns
[min-max]	[2 – 40]	[2 – 50]	[4 – 67]	[10 – 35]	[3 – 8]	[2 – 67]	
≤10 years (%)	4.76	12.04	12.12	7.69	0	9.57	
>10 years (%)	95.24	86.96	87.88	92.31	100	90.43	
COI (mean)	2.71	4.83	2.76	4	5.75	3.65	0.01**
[min-max]	[1 – 5]	[1 – 11]	[1 – 7]	[2 – 9]	[3 – 8]	[1 – 11]	
≤10 years (mean)	2	8.5	3	3	0		
>10 years (mean)	3.45	4.37	2.72	4.09	5.75		

* Statistical differences were calculated using the Chi-square test

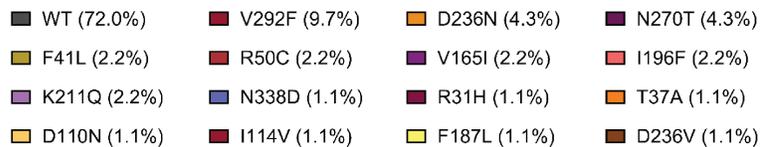
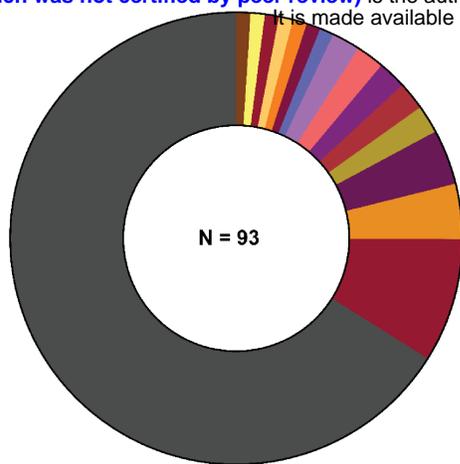
** Statistical differences were calculated using the Kruskal Wallis test

BF – Bandafassi, CM – Camp Militaire, DB – Dalaba, MK – Mako, TM - Tomboronkoto

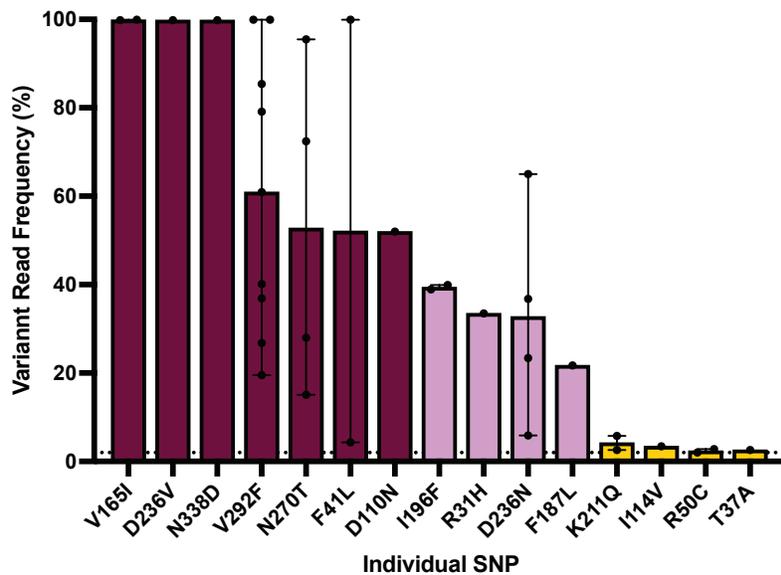
Table 2. Predicted functional characteristics of PfCyRPA SNPs. Individual FASTA files of PfCyRPA alleles were threaded through the crystal structure and the impact of the mutant versions of the protein was evaluated for predicted binding affinity of PfCyRPA to its binding partner PfRH5 or neutralizing human mAbs. The positions of the SNPs were determined using structural data from Chen et al., 2017. The binding energy alternation from the SNPs was predicted by FoldX version 5.0. Predicted binding energies are shown for reference and mutant alleles of the protein in Kcal/Mol for each SNP. Changes between the two are shown as $\Delta\Delta G$ (Kcal/mol). A negative $\Delta\Delta G$ indicates a predicted increase in PfCyRPA stability while a positive $\Delta\Delta G$ is associated with a predicted decrease in PfCyRPA stability.

SNP	Blade	Predicted impact on interaction with RH5 and antibodies Cy.003, Cy.004, and Cy.007	$\Delta\Delta G$ of CyRPA stability (Kcal/mol)
R31H	6	Minor effect	1.91
T37A	6	Minor effect	0.45
F41L	6	Minor effect	1.23
R50C	1	Minor effect on antibody binding, but may improve RH5 binding by removing repulsion between CyRPA R50 and RH5 K504	1.11
D110N	2	Minor effect	-2.02
I114V	2	Minor effect	0.57
V165I	3	Minor effect; may alter CyRPA structure	-0.99
F187L	3	Minor effect on antibody binding, but may affect binding to RH5	0.2
I196F	3	Minor effect on antibody binding, may affect binding to RH5; may alter CyRPA structure	9.33
K211Q	4	Minor effect	0.95
D236N	4	Add a new N-glycosylation site, minor effect	0.05
D236V	4	Alter local structure through steric clash with S233, minor effect on antibody and RH5 binding	0.09
N270T	4	Minor effect on antibody and RH5 binding; alter CyRPA structure, but abolishes a hydrogen bond between CyRPA N218 and N270	-0.03
V292F	5	Minor effect; may alter CyRPA structure	8.54
N338D	6	Minor effect	1.75

A



B



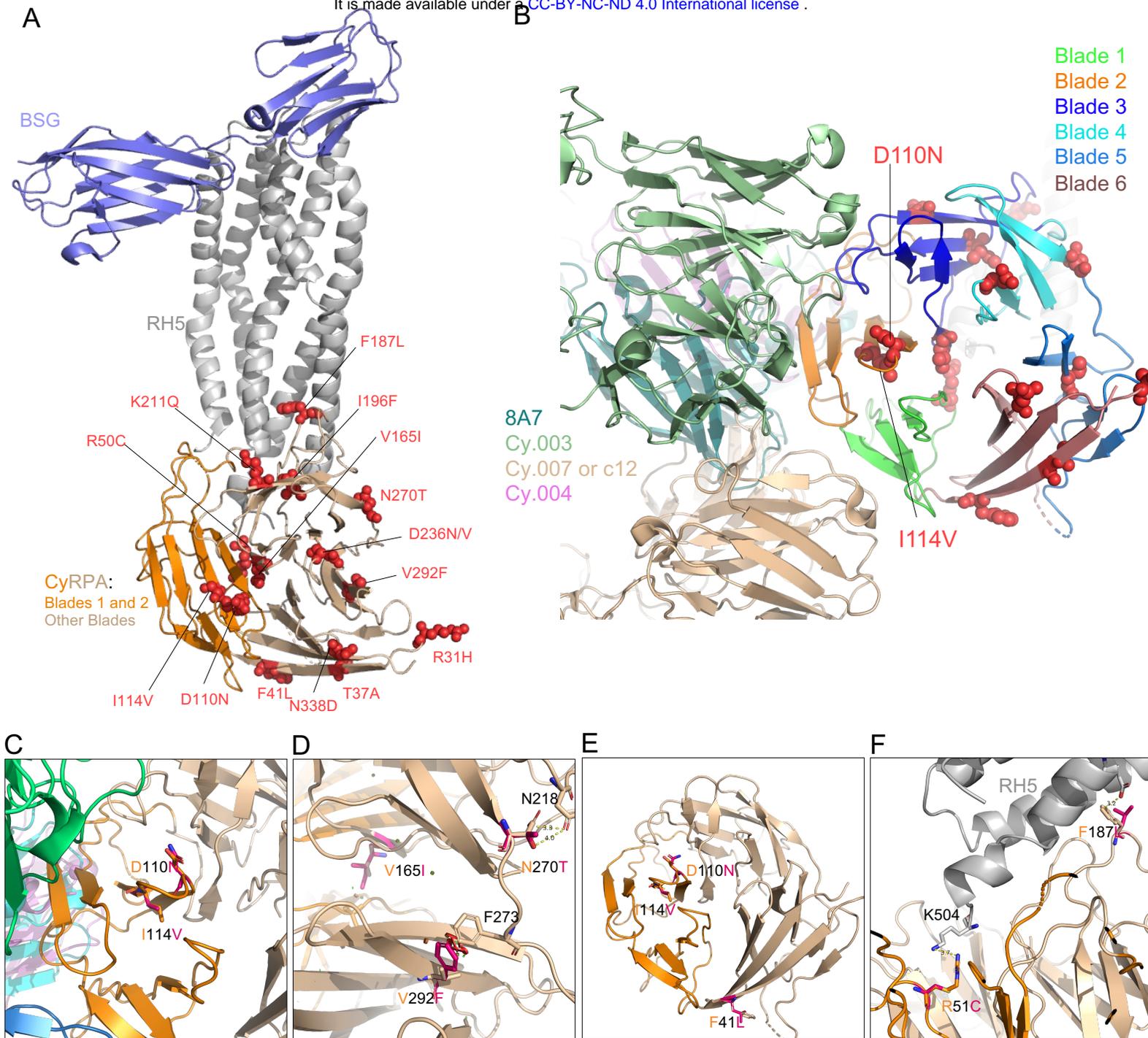


Figure 2. Structure-function predictions for the novel SNPs identified in CyRPA.

(A) The location of SNPs within the BSG–RH5–CyRPA complex. The complex construction was achieved by superimposing the RH5 of the RH5–BSG complex (PDB ID: 4U0Q) onto the RH5–CyRPA complex (PDB ID: 6MPV). BSG and RH5 are depicted in light blue and grey, respectively. CyRPA is represented in a wheat color, while blade 1 and blade 2 are indicated in dark orange. (B) The positions of D110N and I114V relative to monoclonal antibodies (mAbs) with CyRPA blades color coded. Antibodies Cy.003, Cy.004, Cy.007, and 8A7 are represented in green, magenta, blue, and cyan, respectively. (C) The positions of D110N and I114V relative to monoclonal antibodies (mAbs). Antibodies Cy.003, Cy.004, Cy.007, and 8A7 are represented in green, beige, magenta, and dark turquoise, respectively. (D) Structural modeling revealed that SNPs V165Y, N270T, and V292F influence the conformation of CyRPA. Small red plates signify the potential steric hindrances. (E) This panel highlights the SNPs that might confer resistance to antibodies, including F41L, D110N, and I114V. (F) SNPs R51C and F187L are shown to directly interact with RH5.