

**Analyses of major malaria pre-erythrocytic stage antigens using worldwide isolates identify vaccine candidate variants and epitopes.**

### **Supplemental Information**

Amed Ouattara<sup>1\*</sup>, Dwivedi Ankit<sup>2</sup>, Matthew Adams<sup>1</sup>, Amadou Niangaly<sup>3</sup>, Matthew B. Laurens<sup>1</sup>, Myaing M. Nyunt<sup>4</sup>, Christopher V. Plowe<sup>3</sup>, Abdoulaye Djimde<sup>3</sup>, Shannon Takala-Harrison<sup>1</sup>, Joana C. Silva<sup>2</sup>

#### **Author affiliation**

<sup>1</sup>Malaria Research Program, Center for Vaccine Development and Global Health, University of Maryland School of Medicine, 685 West Baltimore St., Baltimore, MD 21201, USA

<sup>2</sup>Institute for Genome Sciences, University of Maryland School of Medicine, 670 West Baltimore St., Baltimore, MD, 21201, USA

<sup>3</sup>Malaria Research and Training Center, University of Sciences, Techniques and Technologies of Bamako, P.O. Box: 1805, Bamako, Mali

<sup>4</sup>Duke Global Health Institute, Duke University, NC 27708

\*Corresponding author

#### **Email addresses:**

AO: [amouattara@som.umaryland.edu](mailto:amouattara@som.umaryland.edu); AD: [ADwivedi@som.umaryland.edu](mailto:ADwivedi@som.umaryland.edu); MA: [madams@som.umaryland.edu](mailto:madams@som.umaryland.edu); AN: [niangaly@icermali.org](mailto:niangaly@icermali.org); MBL:

[mlaurens@som.umaryland.edu](mailto:mlaurens@som.umaryland.edu); MM: [myaing.nyunt@duke.edu](mailto:myaing.nyunt@duke.edu); CVP:

[chris.plowe@duke.edu](mailto:chris.plowe@duke.edu); AD: [adjimde@icermali.org](mailto:adjimde@icermali.org); STH: [stakala@som.umaryland.edu](mailto:stakala@som.umaryland.edu);

JCS: [jcsilva@som.umaryland.edu](mailto:jcsilva@som.umaryland.edu)

## A. Figures

**Supplemental Figure S1. Linear sequences of epitope fragments.** All individual epitopes have been concatenated into a single fragment and polymorphic residues are highlighted in yellow.

### - P36

EYIDKTNDVDIIATFNLICIKCKEPLLDKPMTNDKSYHLERIRREARKEGYITFTGFNF

### - UIS3

LKKGIEDDLYYNKYLEYNNIKYRHVNMMEYMVQGLQENNMY

### - UIS4

VNDSEELD DKDKDE

### - CELTOS

FRGNHSEQLESQVLPTVKKHAKSENLVAEPGA KVSDE

### - CSP

ALYQCYGS

SSTLNERSNKHKLKQPADGNPDNDPNEWSPCTCGNGIQVRIKPGSANKPKDEKKICKMEVN

### - SAP1

RHVMFLRDDMNFEMNEQGKTHNLDIQSYMPYDQQKGENQEPSYLNKKEDVKIPKTKSCED  
NQYLIISELVNYHYCFYITDLCKKLSSVKYFNKPISILGNEQIQNGYQKLTSNHQMNQKILCDK  
NKKCDIPQKCDNTQDDENSEQNQNNSNGKIQHIDNSDHHIDPHLNEPSSNDVPKIQKHYS

SIINVQKDISTQPTDDNNTAYLNRVLSHDILNEIQEIHPDILISNLVQELKQHDIRTTYNDNAI  
TAPFNIQQTQTDNKKDEYDKKNNVKNKHYNKAGNKFSSVQIDTQMDNPDKNNDIYNSIINKSQ  
DGTCTNFLQNSLWKHFQILSIYKKEKCDQAKPYSENTDVLCRWTPSYICCNALNNYHYFTSTL  
CPQQTFKFTYDYKNK

- LSA1

VNVSRIRKREDLEDLEQEEEEEDLEEGEEDLEEQEEEEEDLEEKEKQGQHSDEREHQSDEREHQSS  
LQDNRGNVEGRRKKIKPEQKEDDILETVNISDVDDDEFDNQDNDNDKSLYDYI

- LSA3

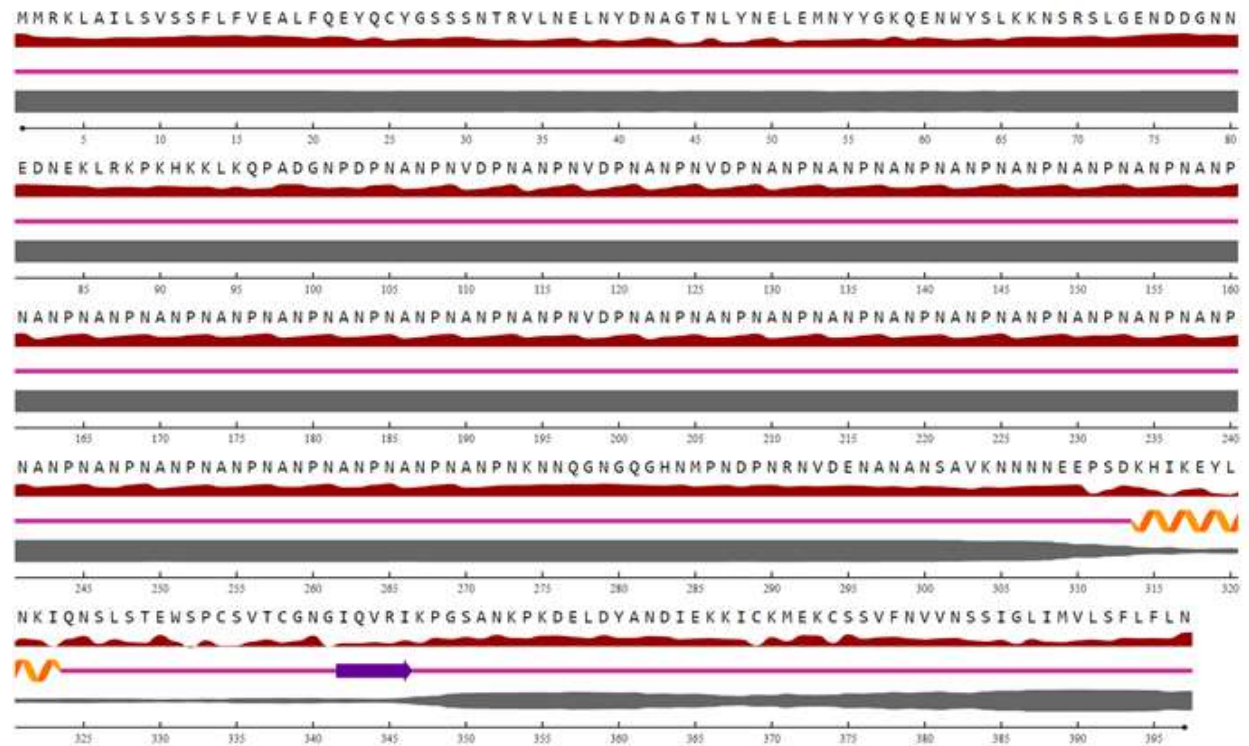
SDKLEGQLVKSTVLDKVENLLTMSIIVIGETTEVIKDDNAKLKELKLSSEDSKEIIDIEADILKEL  
VLKGGRLVDKKEVDKEVSKASKPFISTFSLIAEKK

- P52

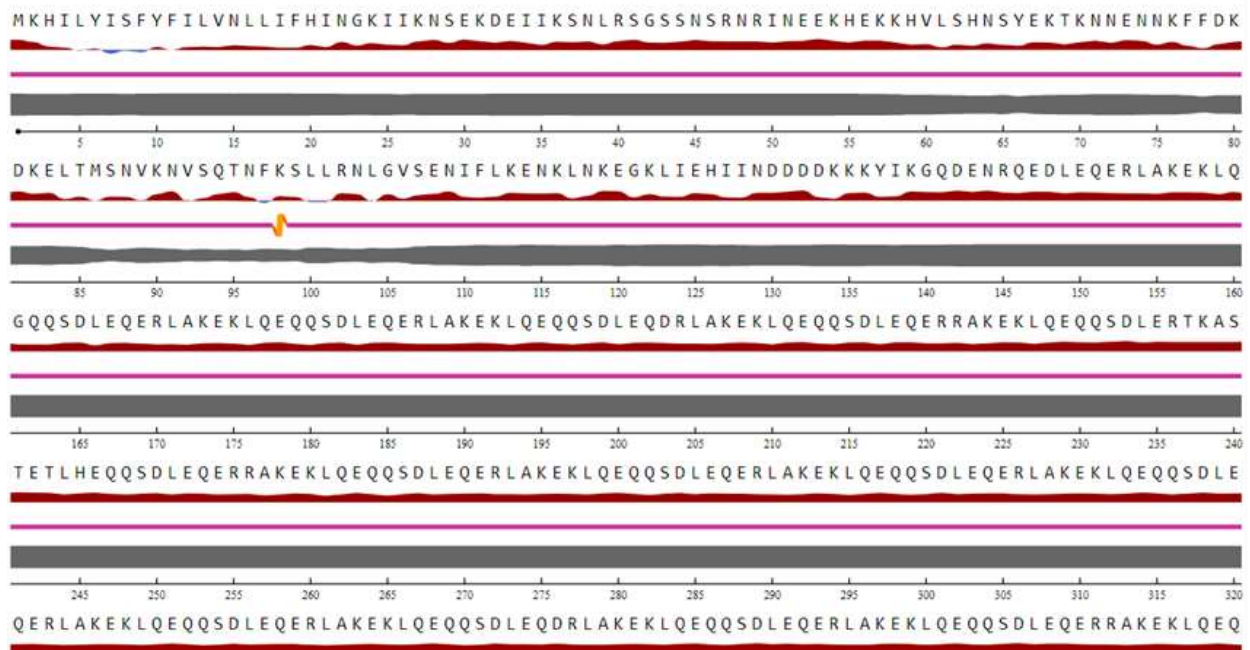
VILNNEFSTTAKTHNCPSRAPTPNFYSNMDIDFSCLCYGDQDKVNIPSKRDLHAYNDVLGI  
NCFKKVIIPDPSCCKREFVGTMIINMDYDHDRTIRTRTFWQNLL

- SPECT2

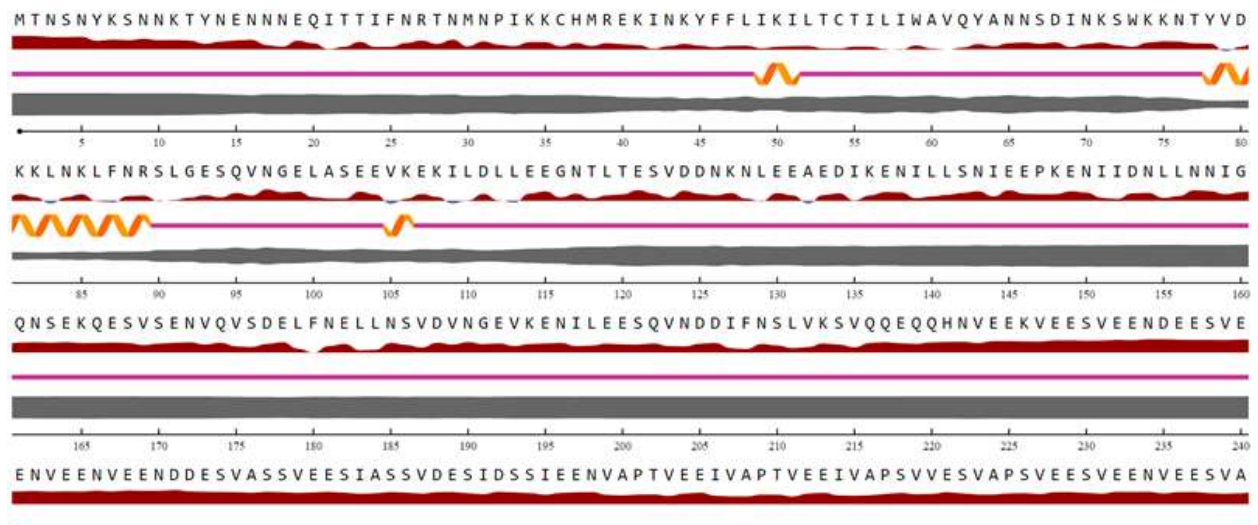
RSNCNSENDMDMLASFskHPLGEAILMLEGIANDLSTLQPVNGWIRKENACTLHVTCVIGLPP  
YPIGLPPYIPKTTAYKNAVELPAVFTGLDKESPSDVYNKSNLWMKFIINVSFSAGGSVDMNEQ  
LIVIGGNDSDIISPHDTDTKNSAPPQDTALKGYNIEVEAGSNSCTSKQSSNKYDTSYLY  
MECGDQPESTSTYPNYSILLGFGISSGVYSTPCIPGMKSCSLNMSYIYVLALDGYSDGELVGTC  
LDGLTGFKVEFHEKCAKSLKACGSAK



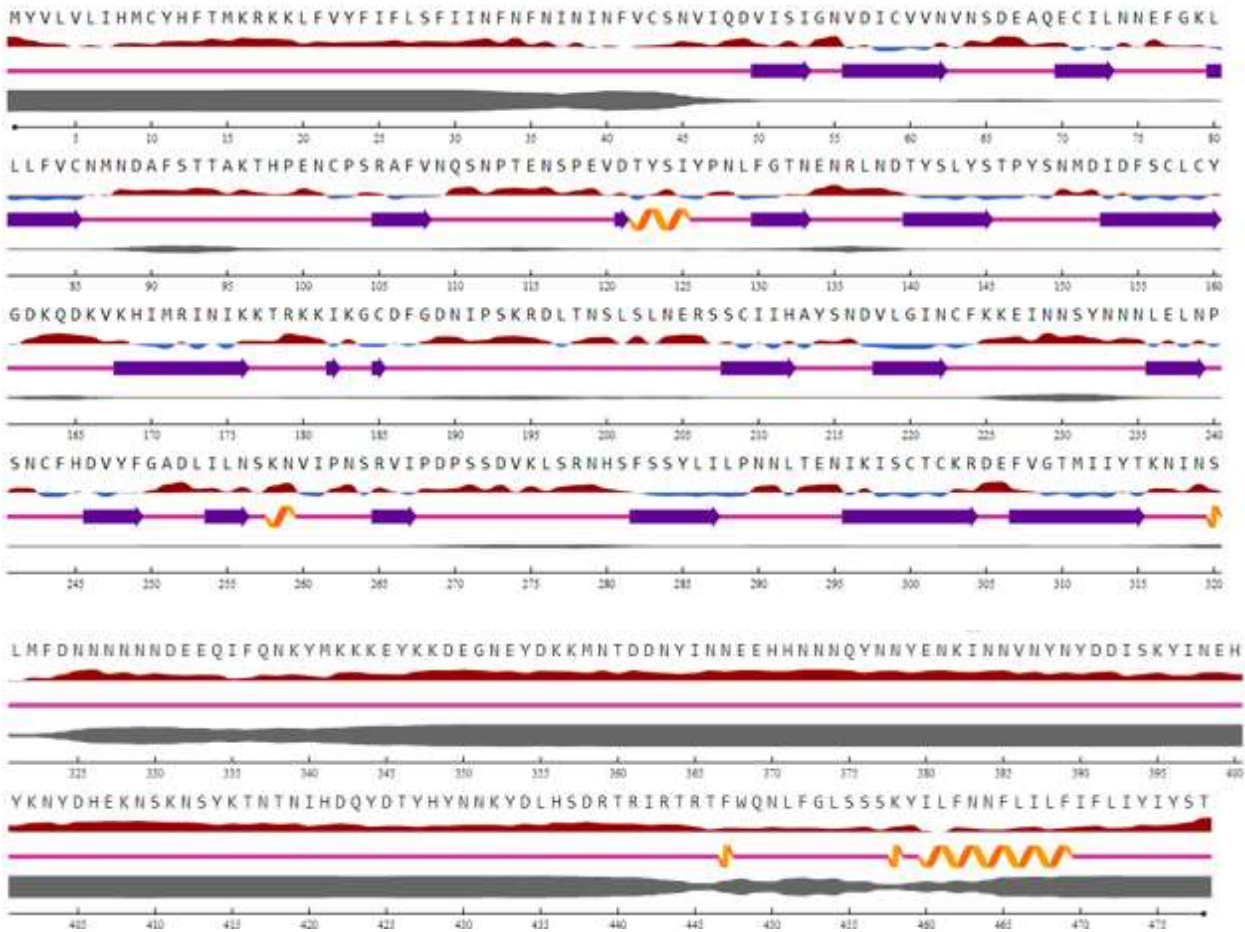
## LSA1



## LSA 3



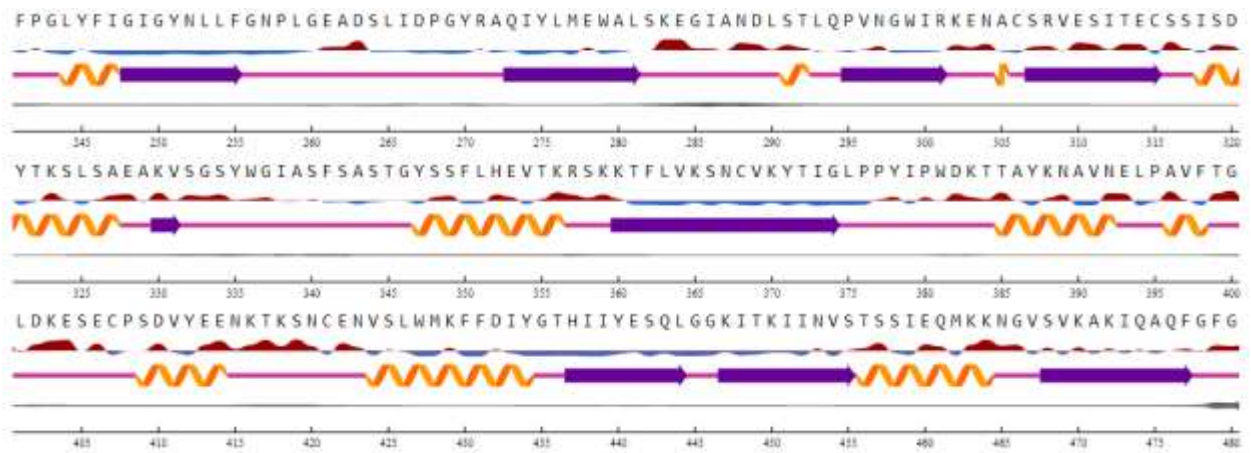
P52



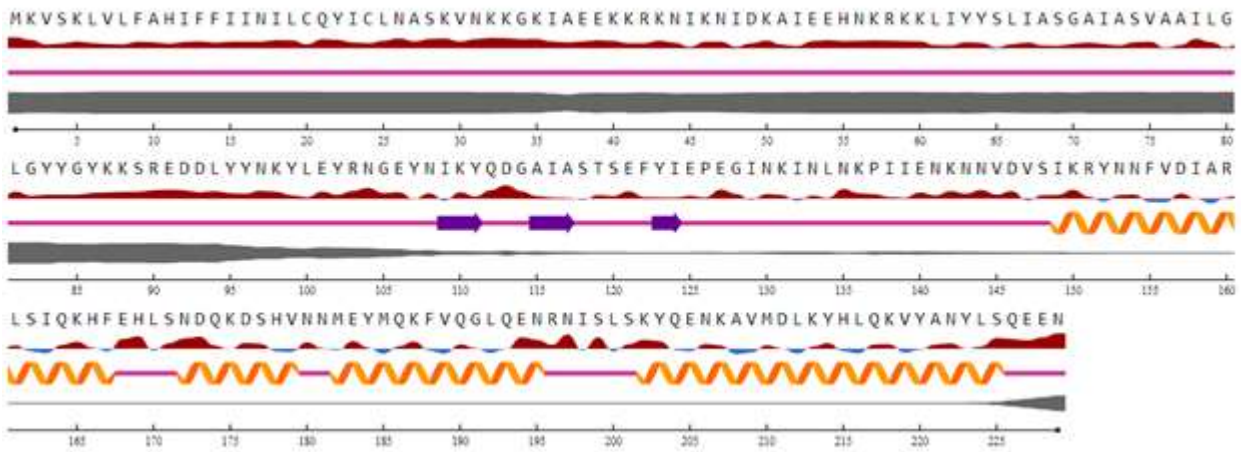
SPECT2



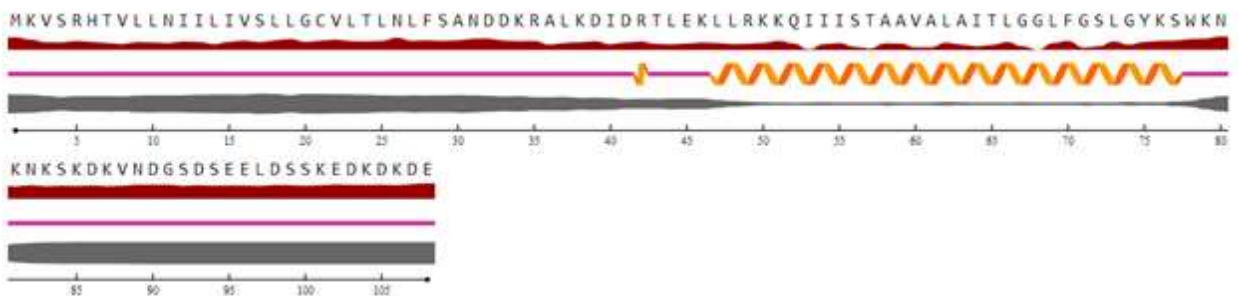




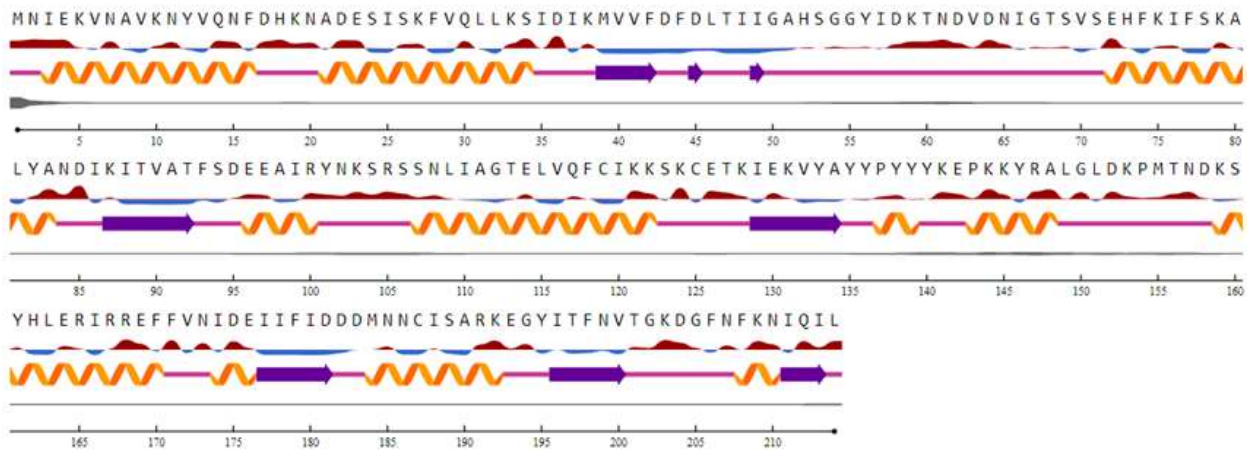
### UIS3



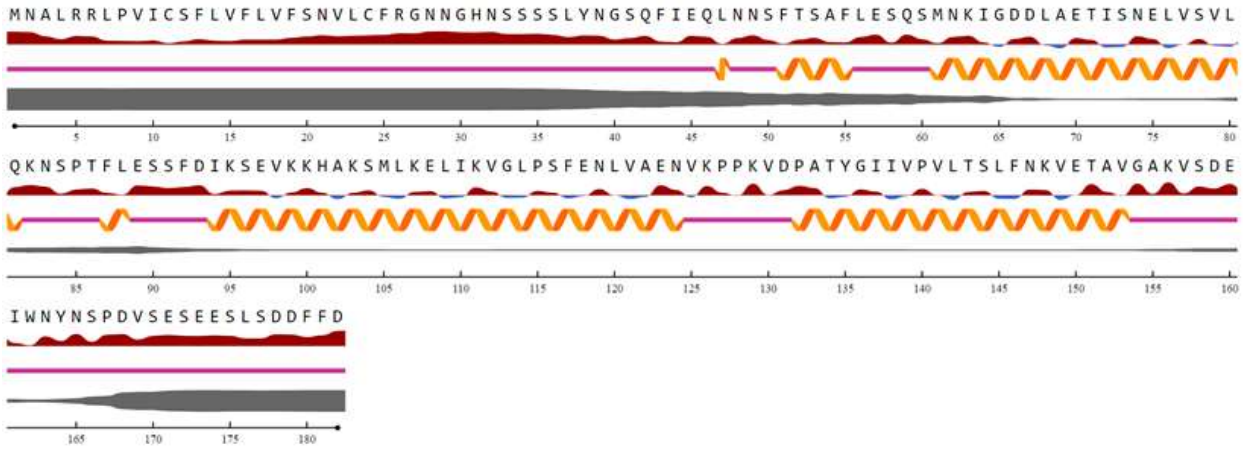
### UIS4



P36

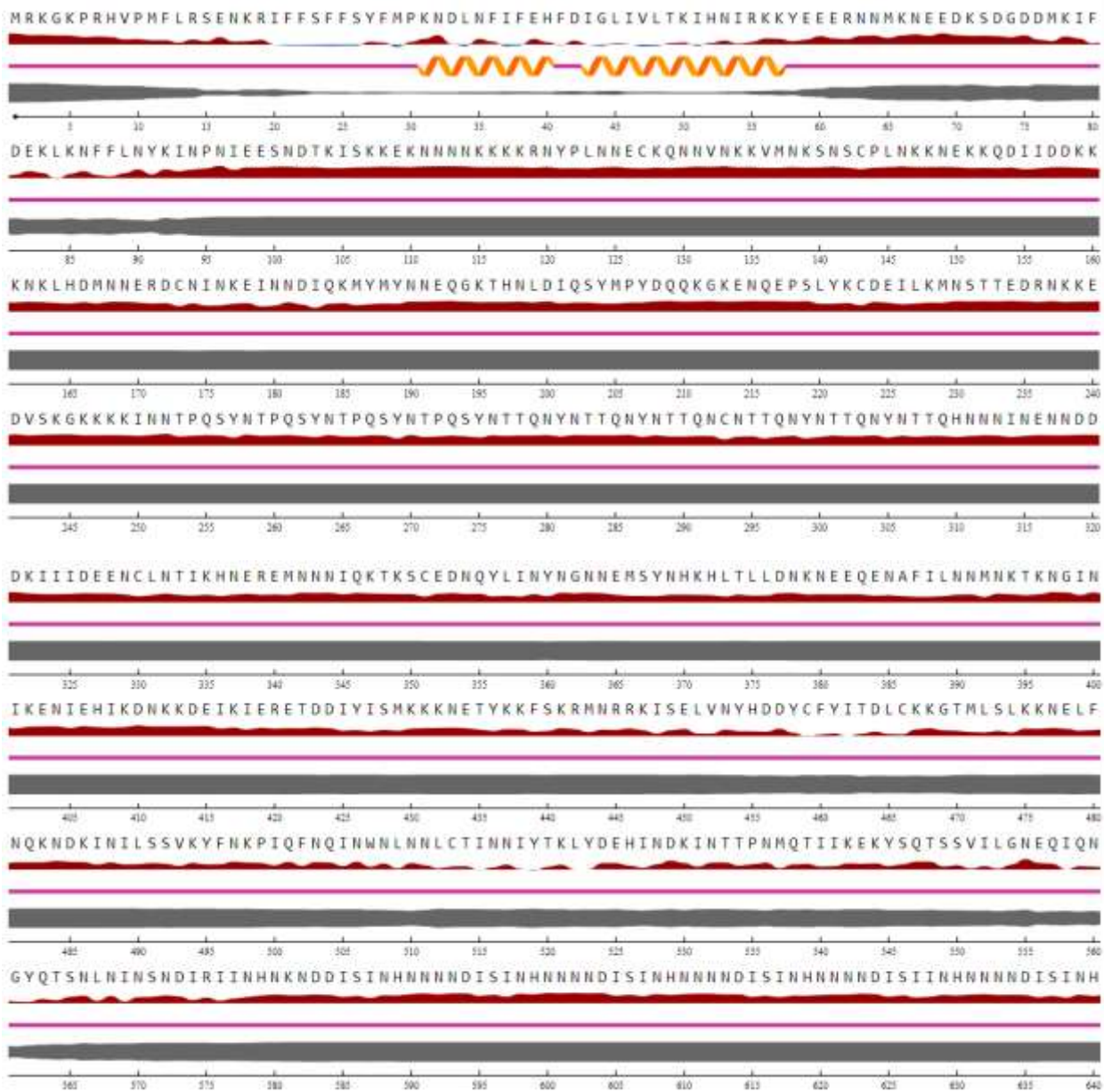


CELTOS

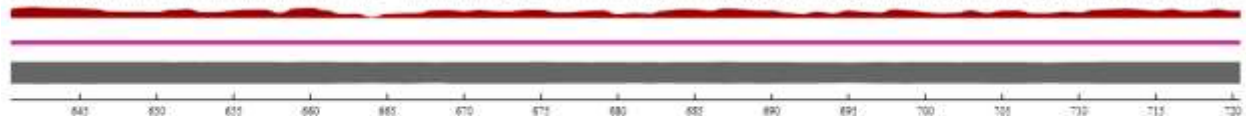




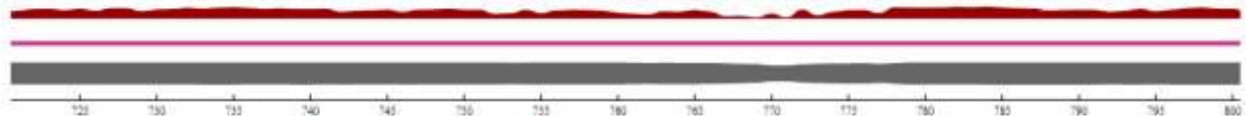
SAP 1



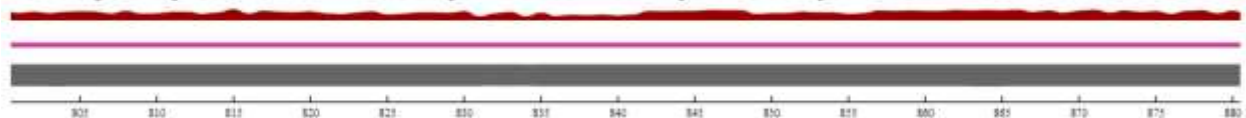
NNNNNNYYFMNNYMHNNIINNYYCYMMNNTNHVNNYYNNIYIQNNHDQNNAPILQPINNHHLAHINDLCYIHSEKNEYTKI



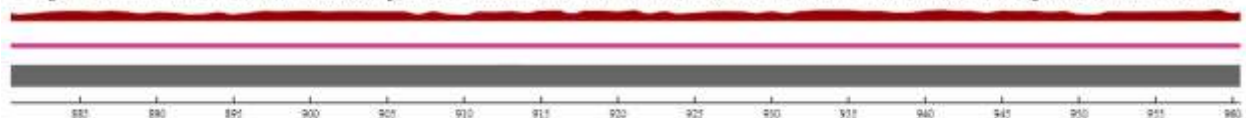
SKNHQNNINPQQSNGKNQNDISNNINKNDEYYNLNEQKILCDKNKSYIKCDIPQKCDNTQDDENSEQNQNYITNPSNGH



YKIKEQMNYVQHIPDYEHDNTTNEHINTQNYTNINLDPYIMNNQNDNNVYLNQNYFDTEKKNKEEDINITEANTHYNNVH



IYQNNSNHMKHINLNNNTNSKSTQDYDLYHNNMENFNNTNYNIMKEKIHINDOTSSVINNSSINSQNNLCNNKKTNDY



QNRNIKNDNSIPDSSINMNELKNNIQMNDYYASHIYNNNNNNNNNNFIISNNHISNYNNYMNESNVYPQINSNNYIPYID



PHLNNPEYKHTINNQLNKNCINNINISMDDNVYDSHTISIHISVNDNEYNNNSHFTNDMYNNNNNNNNNNNNCNSNSSNSN



SNQVCLYMPNHELYNSNYNITQHMPSSNDVPKIQKHIANNIIMNGHKEEHIEKKSKEETNKTNEQVYRSINQNNTLILK



ENEIDENDINTLNQNLNIKNDMMNMDHIINLKNLNNINNIYTPYQNNILKNNEIQFLNNNKEVITKHAYTHSSNEININV

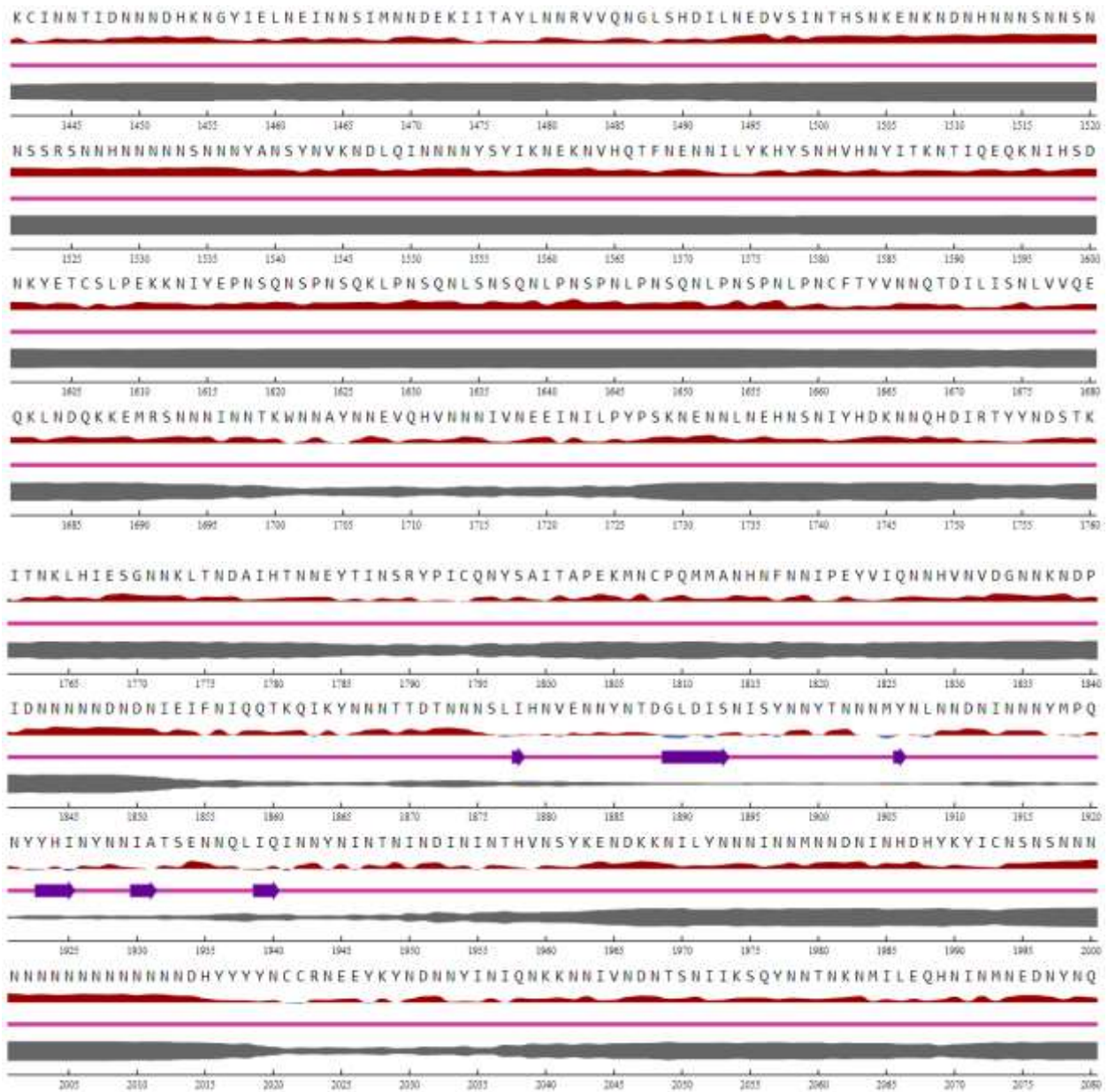


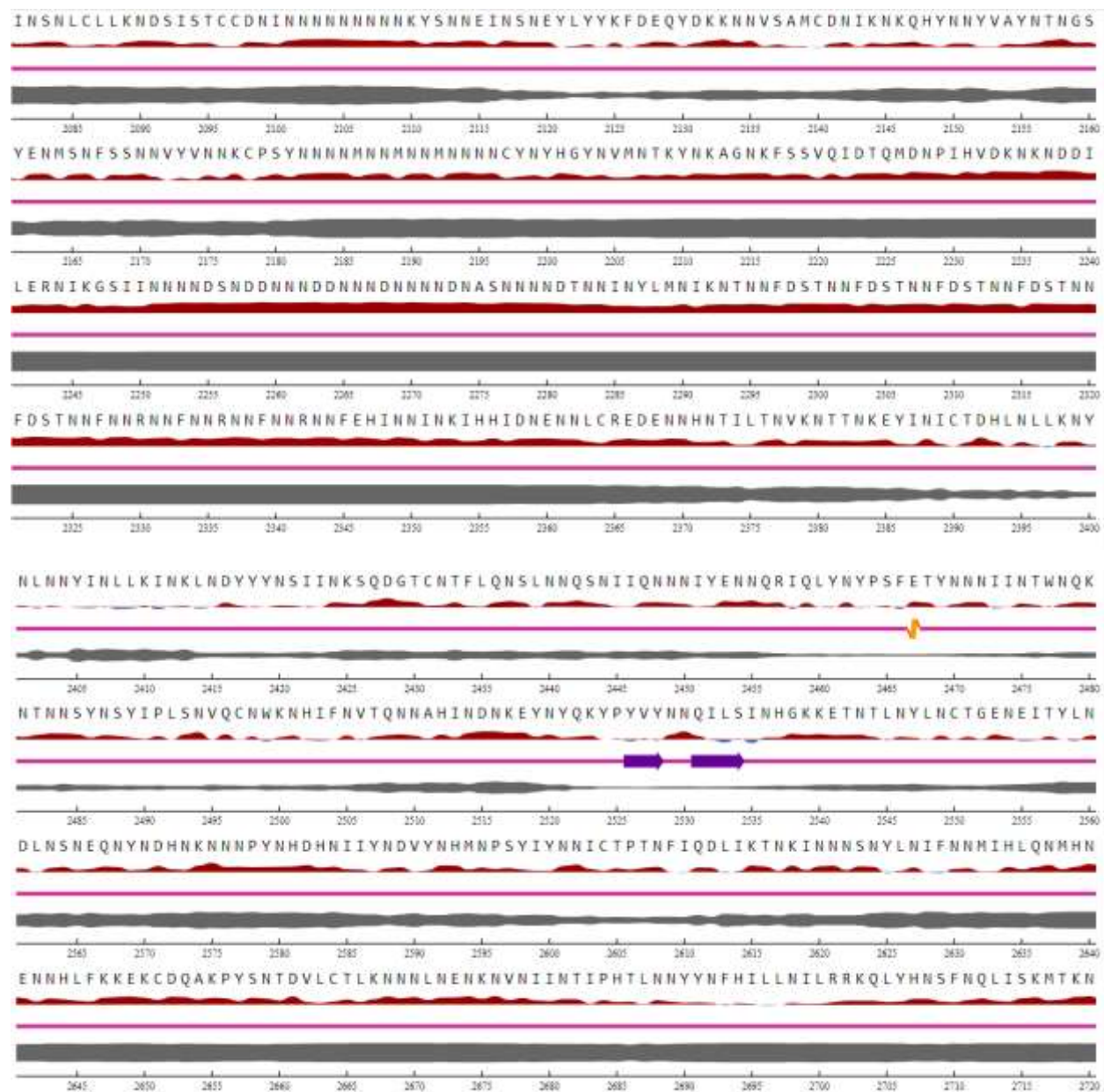
YKNIDTQINIRKNENNNNDIINKKEQKDISTNNEQNQTYNYITSSVKNEYSILDNSPTDKNYKKNLNLNLISSSYNTNDHIN



NYHDDIDKTNKDNKYEEEEKHKDINIYINNINSNGKNKNNFINSYFDLNEHEKKKKINIFYHDNKS LKNISNEELTNTNPY



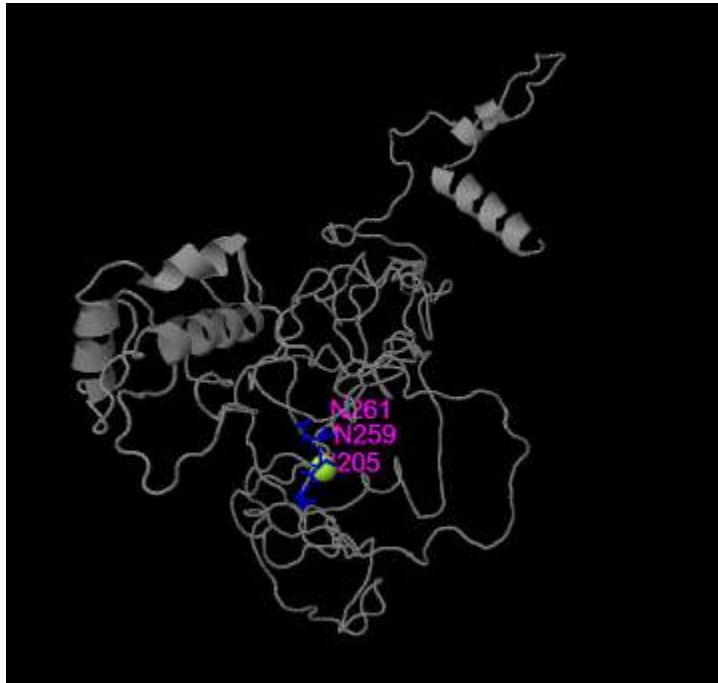




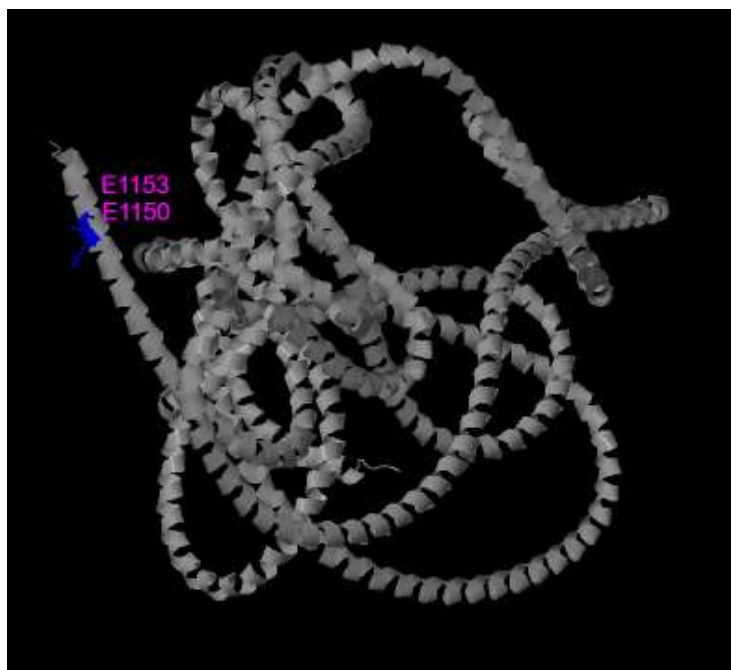


**Supplemental Figure S3.** Predicted 3D structure of malaria vaccine candidates. Hypothetical protein-ligand binding sites are shown in light purple.

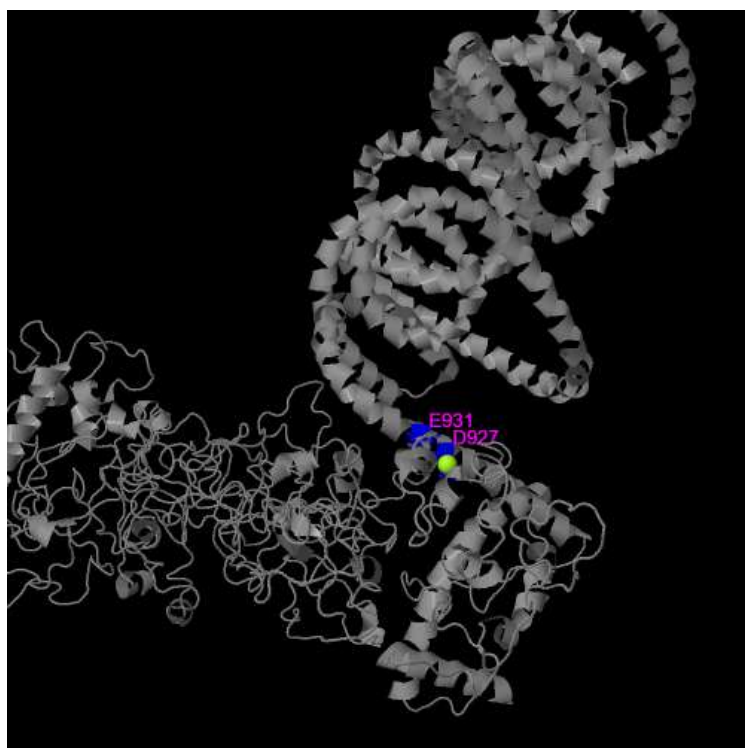
**CSP**



**LSA1**

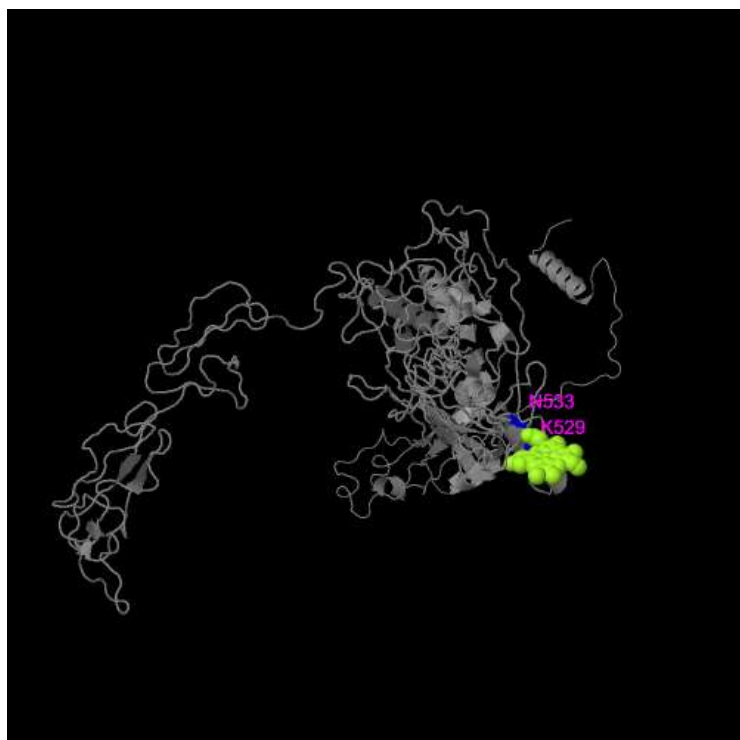


LSA3



SPECT2





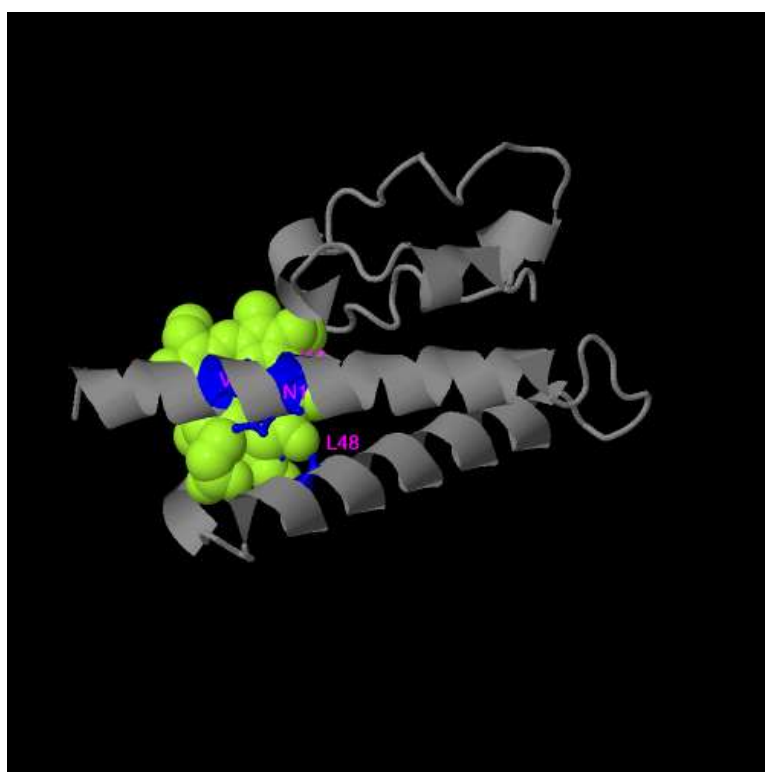
P52



UIS3



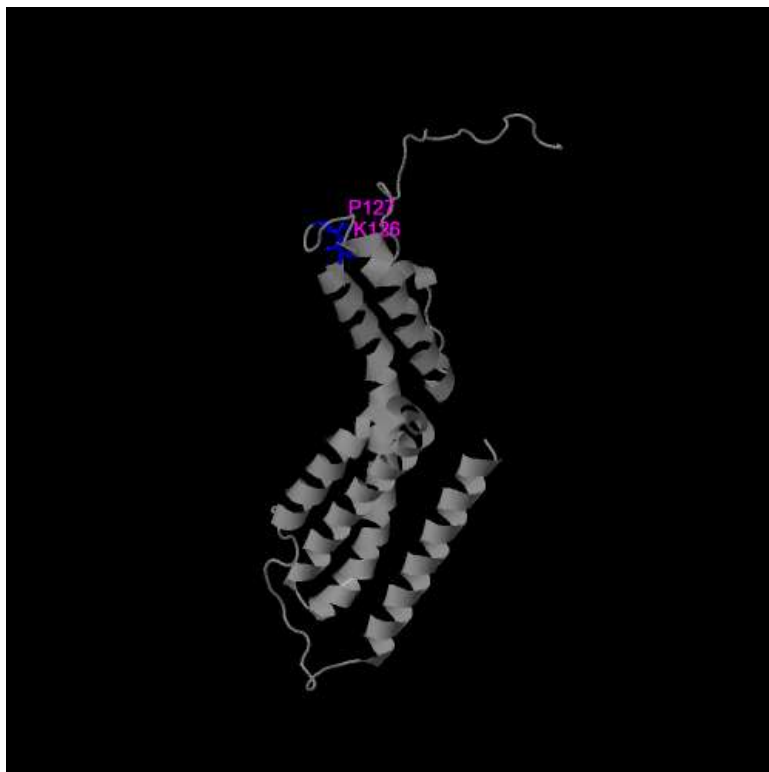
UIS4

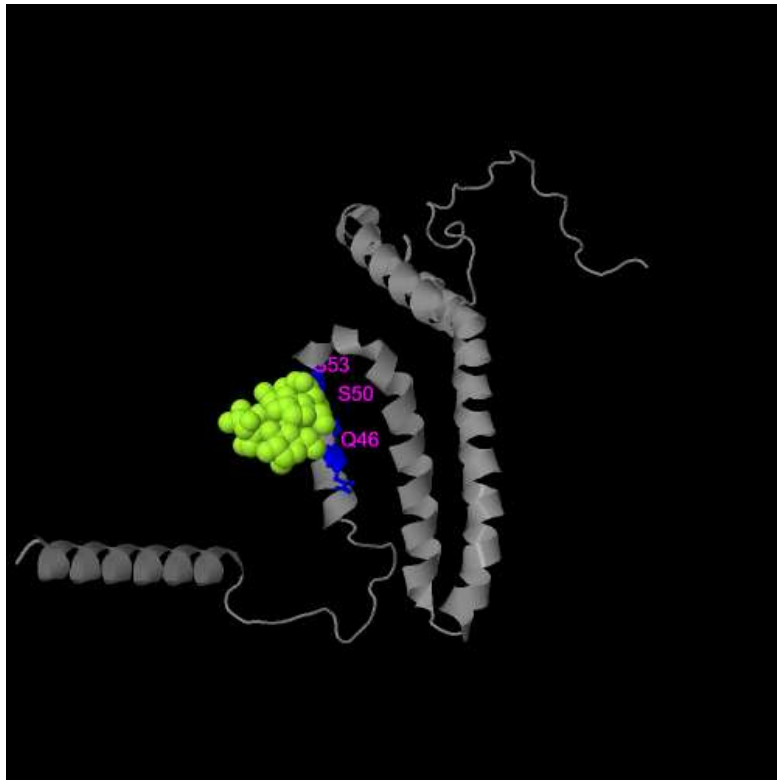


P36

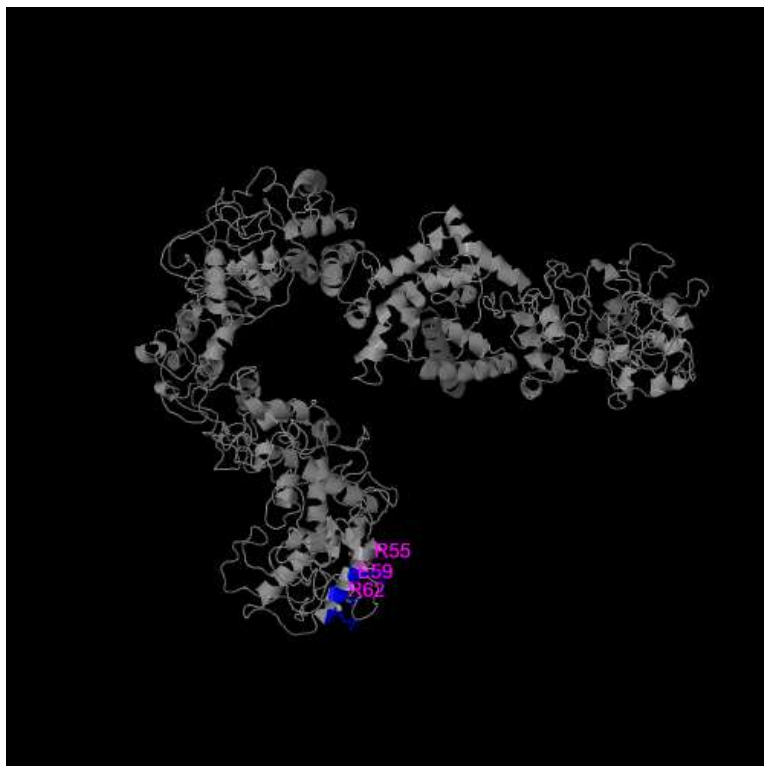


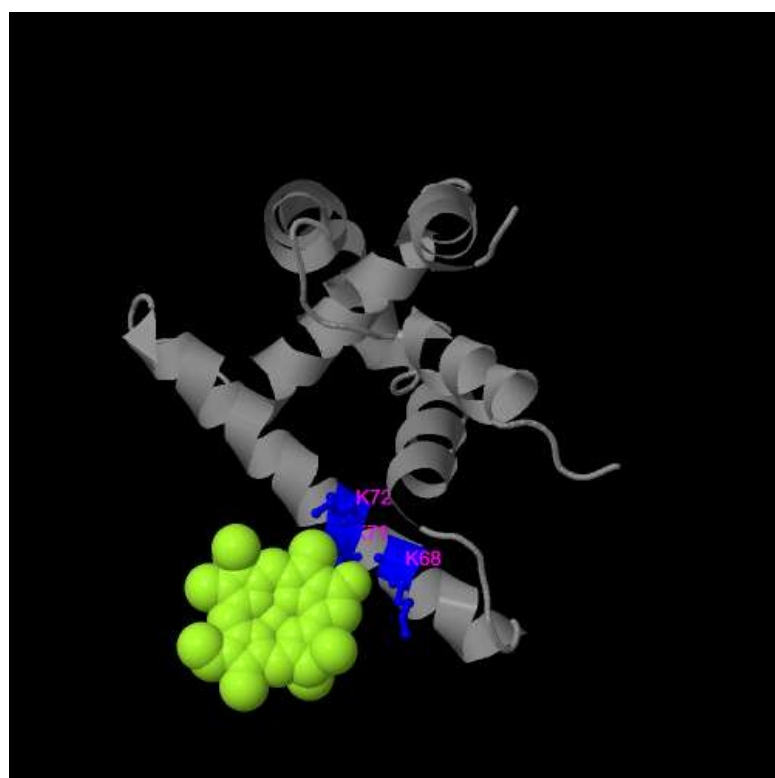
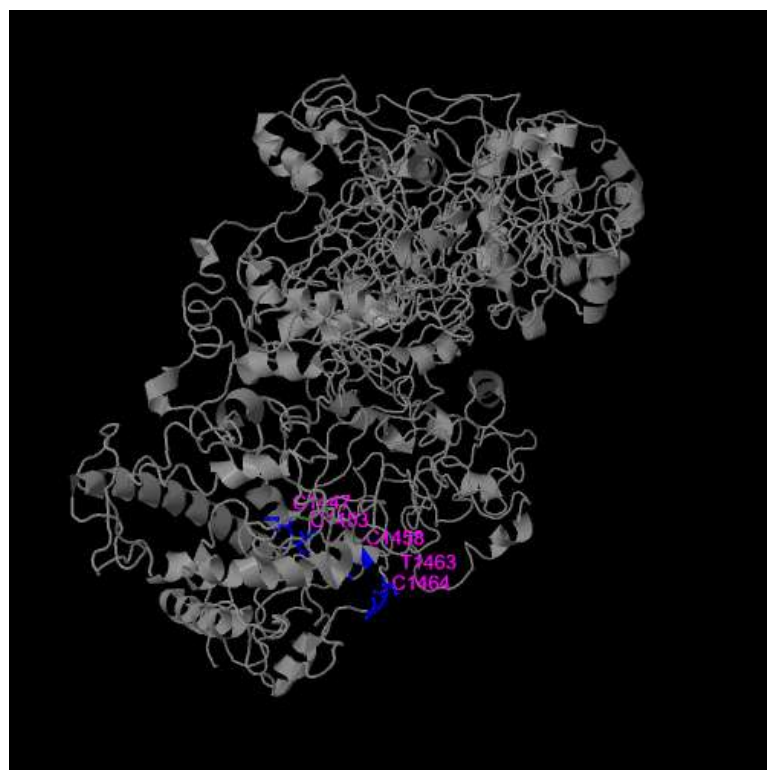
CELTOS





Three fragments of SAP1





## B. Tables

**Supplemental Table S1.** List of antigens selected and their putative function.

Antigen	Antigen Full name	Gene ID	Putative evidence	Reference
CELTOS	Cell-traversal protein for ookinetes and sporozoites	PF3D7_1216600	Associated with protection from clinical malaria	21
CSP	Circumsporozoite protein	PF3D7_0304600	Associated with protection from clinical malaria.	22
LSA1	Liver stage antigen 1	PF3D7_1036400	Associated with protection from clinical malaria.	23
LSA3	Liver stage antigen 3	PF3D7_0220000	Associated with protection from clinical malaria.	24
P36	6-cysteine protein (P36)	PF3D7_0404400	Gene knockout prevents parasite egress from liver	25
P52	6-cysteine protein (P52)	PF3D7_0404500	Gene knockout prevents parasite egress from liver	26
UIS3	Upregulated in infectious sporozoite	PF3D7_1302200	Gene knockout prevents parasite egress from liver	27
UIS4	Early transcribed membrane protein	PF3D7_1016900	Gene knockout prevents parasite egress from liver	28
SAP1	Sporozoite and liver stage asparagine-rich protein		Gene knockout abolish liver stage	29
SPECT2	sporozoite microneme protein essential for cell transversal	PF3D7_1147000	Essential for cell transversal	
		PF3D7_040870		30



**Supplemental Table S2.** Samples origin .

Continent	Country	Number of Sequences	<i>P. falciparum</i> endemicity	Collection year
Africa	Burkina Faso	56	High	2008
	Cameroon	130	High	2013
	Guinea	123	High	2011
	Kenya	57	High	2007
	Malawi	150	High	2007-08, 2014-16
	Mali	99	High	2002, 2010
	Tanzania	68	High	2010, 2013
Asia	Cambodia	129	High	2009-11
	Myanmar	18	Low	2013
	Thailand	36	Low	2013-14
Oceania	Papua New Guinea	59	Low	2008-10
South America	Brazil	22	Low	2016
	Colombia	16	Low	2011
	French Guiana	36	Low	2009-2013
	Peru	11	Low	2011

**Supplemental Table S3. FST values of the malaria vaccine candidates by continents.** Continents are in columns and each Excel sheet represent a different malaria vaccine candidate.

**Supplemental Table S4. FST values of the malaria vaccine candidates by countries.** Countries are in columns and each Excel sheet represent a different malaria vaccine candidate.

**Supplemental Table S5. Haplotypes distribution by country.** Prevalence of ten of the most frequent haplotypes of each of the studied malaria vaccine candidate by country. Each column represents a country and haplotypes, and their respective frequencies are in row. Each Excel sheet represents a different malaria vaccine candidate.

**Supplement Table S6.** Putative, strong binding CD4+ and CD8+ T-cell epitopes in malaria vaccine candidates.

Antigen	Strong binding CD4+ epitope	Strong binding CD8+ epitope	Number of strong and medium binders	Average number of epitopes per 100 amino acid residues (AA) (Protein length, in AA)
CeTOS				
	MNALRRLPVICS	LPVICSFLVF	9	4.9 (182)
CSP				
	KLAILSVSSFLF LAILSVSSFLFV ENWYSLKKNSRS	SSFLFVEALF*	10	2.5 (397)
LSA1				
	TNFKSLLRNLGV NFKSLLRNLGVS QTNFKSLLRNLG FKSLLRNLGVSE NFKSLLRNLGVS TNFKSLLRNLGV ISFYFILVNLLI SFYFILVNLLIF	KFIKSLFHIF*	16	1.4 (1162)
LSA3				
	None	ASYVVGFFTF* SYVVGFFTFS* PFYSFVFDIF* KVKNFVKKYK KVDKNNKVPK* KTRKKAQRPK* KVFAAPFISA* KINKYFFLIK	25	1.6 (1558)
P36				
	IRYNKSRSSNLI AIRYNKSRSSNL KFVQLLSIDIK RYNKSRSSNLIA FVQLLSIDIKM AIRYNKSRSSNL EAIRYNKSRSSN EEAIRYNKSRSS	KSKCETKIEK	20	5.3 (379)

	IRYNKSRSSNLI SKFVQLLSIDI			
P52	MCYHFTMKRKKL HMCYHFTMKRK K NLFGLSSSKYIL QNLFGLSSSKYI NININFVCSNVI ININFVCSNVIQ CYHFTMKRKKLF YHFTMKRKKLFV LFGLSSSKYILF GTMIIYTKNINS MIIYTKNINSLM TMIIYTKNINSL VGTMIYTKNIN FGLSSSKYILFN	KYILFNNFLI  ILFNNFLILF* VYFIFLSFII* KVKHIMRINI RTRTFWQNLF KLSRNHSFSS NPSNCFHDVY NPSNCFHDVY NPSNCFHDVY NPSNCFHDVY	39	8.15 (478)
Antigen	Strong binder CD4+ epitope	Strong binder CD8+ epitope	Number of Strong and medium binders	Average number of epitopes per 100 amino acid residues (AA) (Protein length, in AA)
SAP1	None	VKYFNKPIQF YKYIQNIILF YFMPKNDLNF KYIQNIILFL NYMPQNYHHI RIFFSFFSYF RFKLTCNFKF KLKNFFLNYK YTRAVWLLKK MPKNDLNFIF MPQNYHHINY	43	1.5 (2940)
SPECT2	KLRILKKHYYVV* LRILKKHYYVVF* MKLRILKKHYYV* MKLRILKKHYYV* KLRILKKHYYVV* LRILKKHYYVVF*	LYFIGIGYNL IYVLCVDTTI KRSKKTFLVK KVVMFGFSLK RSKKTFLVKS KKIKHSFNLA YIPWDKTTAY	34	4 (842)
UIS3	KYHLQKVYANYL * YHLQKVYANYLS*			

	MEYMQKFVQGL			
	Q*			
	NMEYMQKFVQG			
	L*			
	<u>NNMEYMQKFVQ</u>	None	9	3.9 (229)
	<u>G*</u>			
	<u>VNNMEYMQKFV</u>			
	<u>Q*</u>			
	<u>LIYYSLIASGAI*</u>			
	<u>IYYSLIASGAIA*</u>			
UIS4				
	<u>KQIIISTA AVAL*</u>			
	<u>QIIISTA AVALA*</u>			
	<u>RTLEKLLRKKQI*</u>	None	4	
	<u>DRTLEKLLRKKQ</u>			3.7 (108)
	<u>LEKLLRKKQII*</u>			

Bold: Mutated AA residue relative to 3D7 allele.

\* Have a mutation

Underline: Mutations likely influence peptide function