Supplementary Information

A plausible mechanism for the antimalarial activity of artemisinin: A computational approach

Ashutosh Shandilya[1], Sajeev Chacko[2], B. Jayaram[1, 3]*, Indira Ghosh[2]*

[1] Department of Chemistry & Supercomputing Facility for Bioinformatics & Computational Biology, Indian Institute of Technology, Hauz Khas, New Delhi 110016,
[2] School of Computational and Integrative Science, Jawaharlal Nehru University, New JNU Campus, New Delhi – 110 067, India
[3]Kusuma School of Biological Sciences, Indian Institute of Technology, Hauz Khas, New Delhi 110016, India

Website: www.scfbio-iitd.res.in
Email: B. Jayaram: bjayaram@chemistry.iitd.ac.in
Indira Ghosh: indira0654@gmail.com;

*Authors for correspondence

```
Alignment score=1956
Length of sequence 1:1228(1U5N)
Length of sequence 2:994 (1SU4)
Aligned length      :971
Identical length    :495
Sequence identity=0.510 495/971

Seq1     8 AHTYDVEDVLKFLDVNKDNGLKNEELDDRRLKYGLNELEVEKKKSIFELILNQFDDLLVK  67
           ::    :  :    :    ::          ::: :::  :  ::  ::   :: ::::
Seq2     4 AHSKSTEECLAYFGVSETTGLTPDQVKRHLEKYGHNELPAEEGKSLWELVIEQFEDLLVR  63

Seq1    68 ILLLAAFISFVLTLLDMKHKKIEICDFIEPLVIVLILILNAAVGVWQECNAEKSLEALKE 127
           :::::: :::::         :    : :: :: :::: :: :::::: :::   :::::
Seq2    64 ILLLAACISFVLAWFEEGEETITA--FVEPFVILLILIANAIVGVWQERNAENAIEALKE 121

Seq1   128 LQPTKAKVLRDGKWEI--IDSKYLYVGDIIELSVGNKTPADARIIKIYSTSLKVEQSMLT 185
             :   :: :        :       ::: :  :: : ::: ::  : :: : : :: ::
Seq2   122 YEPEMGKVYRADRKSVQRIKARDIVPGDIVEVAVGDKVPADIRILSIKSTTLRVDQSILT 181

Seq1   186 GESCSVDKYAEKMEDSYKNCEIQLKKNILFSSTAIVCGRCIAVVINIGMKTEIGHIQHAV 245
           ::: :: :  :   :       : ::: ::: : :  :     :   :  :::: :
Seq2   182 GESVSVIKHTEPVPDP--RAVNQDKKNMLFSGTNIAAGKALGIVATTGVSTEIGKIRDQM 239

Seq1   246 IESNSEDTQTPLQIKIDLFGQQLSKIIFVICVTVWIINFKHFSDPIHG-SFLYGCLYYFK 304
                :   :::: : : :: :::: :  ::: :: ::  :: :: :: :   :  ::::
Seq2   240 --AATEQDKTPLQQKLDEFGEQLSKVISLICVAVWLINIGHFNDPVHGGSWIRGAIYYFK 297

Seq1   305 ISVALAVAAIPEGLPAVITTCLALGTRRMVKKNAIVRKLQSVETLGCTTVICSDKTGTLT 364
           : :::::::::::::::::::::::::: ::::::: : :::::::: ::::::::::
Seq2   298 IAVALAVAAIPEGLPAVITTCLALGTRRMAKKNAIVRSLPSVETLGCTSVICSDKTGTLT 357

Seq1   365 TNQMTTTVFHLFRESDSLTEYQLCQKGDTYYFYESSNLTNDIYAGESSFFNKLKDEGNVE 424
           ::::           :          ::     : :  :   ::           :: :
Seq2   358 TNQMSVCKMFIIDKVD----------GDFCSLNEFS-ITGSTYA----------PEGEV- 395

Seq1   425 ALTDDGEEGSIDEADPYSDYFSSDSKKMKNDLNNNNNNNNNSSRSGAKRNIPLKEMKSNE 484
                                       :::            :::      : :
Seq2   396 --------------------------LKND---------KPIRSGQFDG--LVELAT-- 415

Seq1   485 NTIISRGSKILEDKINKYCYSEYDYNFYMCLVNCNEANIFCNDNSQIVKKFGDSTELALL 544
                                        :   ::      :       : :  :: ::
Seq2   416 ---------------------------ICAL-CNDSSLDFNETKGVYEKVGEATETAL- 445

Seq1   545 HFVHNFDILPTFSKNNKMPAEYEKNTTPVQSSNKKDKSPRGINKFFSSKNDNSHITSTLN 604
                                     :: :   :                          :
Seq2   446 ------------------------TTLVEKMN----------------------VFN 456

Seq1   605 ENDKNLKNANHSNYTTAQATTNGYEAIGENTFEHGTSFENCFHSKLGNKINTTSTHNNNN 664
               ::      :
Seq2   457 TEVRNLSKVERAN---------------------------------------------- 469

Seq1   665 NNNNNSNSVPSECISSWRNECKQIKIIEFTRERKLMSVIVENKKKEII-----LYCKGAP 719
                     : :  :   :     :: : :: :::     :            ::::
Seq2   470 -----------ACNSVIRQLMKKEFTLEFSRDRKSMSVYCSPAKSSRAAVGNKMFVKGAP 518
```

```
Seq1   720 ENIIKNCKYYLTKNDIRPLNETLKNEIHNKIQNMG--KRALRTLSFAYKKLSSK--DLNI 775
           :  :  : :         :     :  :   :   :     :: :  :       :
Seq2   519 EGVIDRCNYVRVGTTRVPMTGPVKEKILSVIKEWGTGRDTLRCLALATRDTPPKREEMVL 578

Seq1   776 KNTDDYYKLEQDLIYLGGLGIIDPPRKYVGRAIRLCHMAGIRVFMITGDNINTARAIAKE 835
                   : ::   :  :  ::::: :   : ::  ::::: :::::::  :: ::
Seq2   579 DDSSRFMEYETDLTFVGVVGMLDPPRKEVMGSIQLCRDAGIRVIMITGDNKGTAIAICRR 638

Seq1   836 INILNKNEGDDEKDNYTNNKNTQICCYNGREFEDFSLEKQKHILKNTPRIVFCRTEPKHK 895
           : :   ::                    : :::: :  :  :            : : :: ::
Seq2   639 IGIFGENEEVADR------------AYTGREFDDLPLAEQREACRRA--CCFARVEPSHK 684

Seq1   896 KQIVKVLKDLGETVAMTGDGVNDAPALKSADIGIAMGINGTEVAKEASDIVLADDNFNTI 955
             ::  :    :  :::::::::::::: : ::::::  :: ::: ::  :::::::: ::
Seq2   685 SKIVEYLQSYDEITAMTGDGVNDAPALKKAEIGIAMG-SGTAVAKTASEMVLADDNFSTI 743

Seq1   956 VEAIKEGRCIYNNMKAFIRYLISSNIGEVASIFITALLGIPDSLAPVQLLWVNLVTDGLP 015
           : :  ::: :::::: ::::::::: :::  :: :: :: : :  : ::::::::::::::
Seq2   744 VAAVEEGRAIYNNMKQFIRYLISSNVGEVVCIFLTAALGLPEALIPVQLLWVNLVTDGLP 803

Seq1   016 ATALGFNPPEHDVMKCKPRHKNDNLINGLTLLRYIIIGTYVGIATVSIFVYWFLFYPDSD 075
           :::::::::  : :   ::     :: :    ::  :: ::: :::     ::    :
Seq2   804 ATALGFNPPDLDIMDRPPRSPKEPLISGWLFFRYMAIGGYVGAATVGAAAWWFMYAEDGP 863

Seq1   076 MHTLINFYQLSHYNQCKAWNNFRVNKVYDMSEDH-------CSYFSAGKIKASTLSLSVL 128
                  :: :  ::               :::       :  : :      :  ::::
Seq2   864 G---VTYHQLTHFMQC--------------TEDHPHFEGLDCEIFEAPE--PMTMALSVL 904

Seq1   129 VLIEMFNALNALSEYNSLFEIPPWRNMYLVLATIGSLLLHVLILYIPPLARIFGVVPLSA 188
           : ::: :::: ::::  ::   ::: :  :      :  :: ::::  ::  ::    :
Seq2   905 VTIEMCNALNSLSENQSLMRMPPWVNIWLLGSICLSMSLHFLILYVDPLPMIFKLKALDL 964

Seq1   189 YDWFLVFLWSFPVIILDEIIKFYAKRKLK 217
             :  :   : ::: :::: :: :   :
Seq2   965 TQWLMVLKISLPVIGLDEILKFIARNYLE 993
```

Since there is no experimentally determined 3D structure for PfATP6 and therefore, any question related to the study of protein-ligand interactions involving pfATP6 must rely solely on models. PfATP6 and mammalian SERCA share 51% identity which is considerably reasonable for homology modeling. Apart from the phospholomban domain (residue no 578-675) which is not present in the mammalian SERCA (this domain is far from the ligand binding site region), there is no major difference in the active site residues of the two enzymes. Alignment and superposition Of the structures of two proteins (1SU4 mammalian SERCA in green and 1U5N (model of *pf*ATP6) in red is shown below.

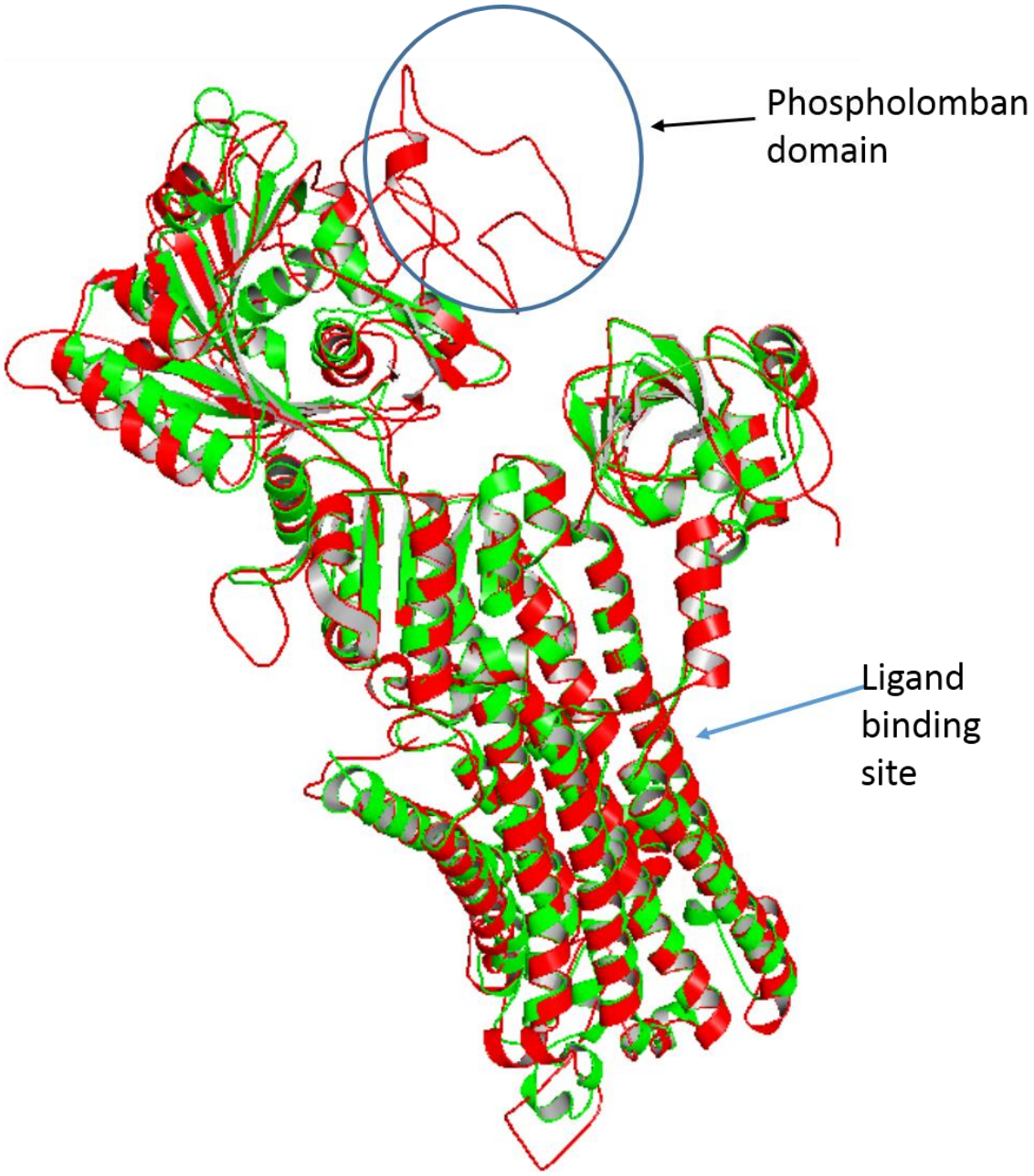
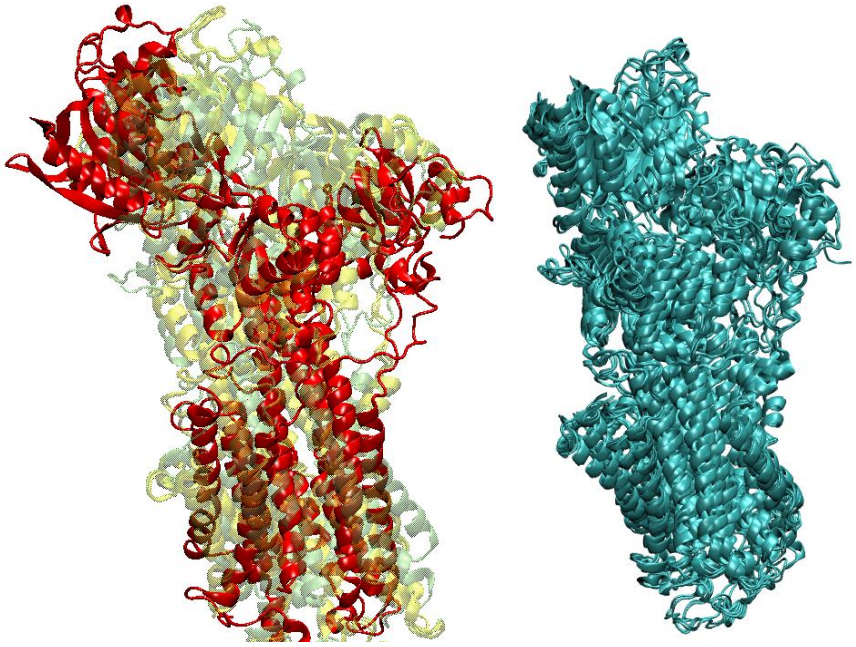


Fig. S1: Superposition of crystal structure of mammalian SERCA in green and model structure of Plasmodium SERCA (*pf*ATP6) in red.



(a) (b)

Fig S2: (a) Superimposition of all 41 structures along with 1SU4 and 2C9M (open conformation) shown in red and orange color respectively. Remaining mammalian SERCA structures are shown in transparent colors to avoid confusion. (b) Superimposition of all 39 closed conformation structures in one color.

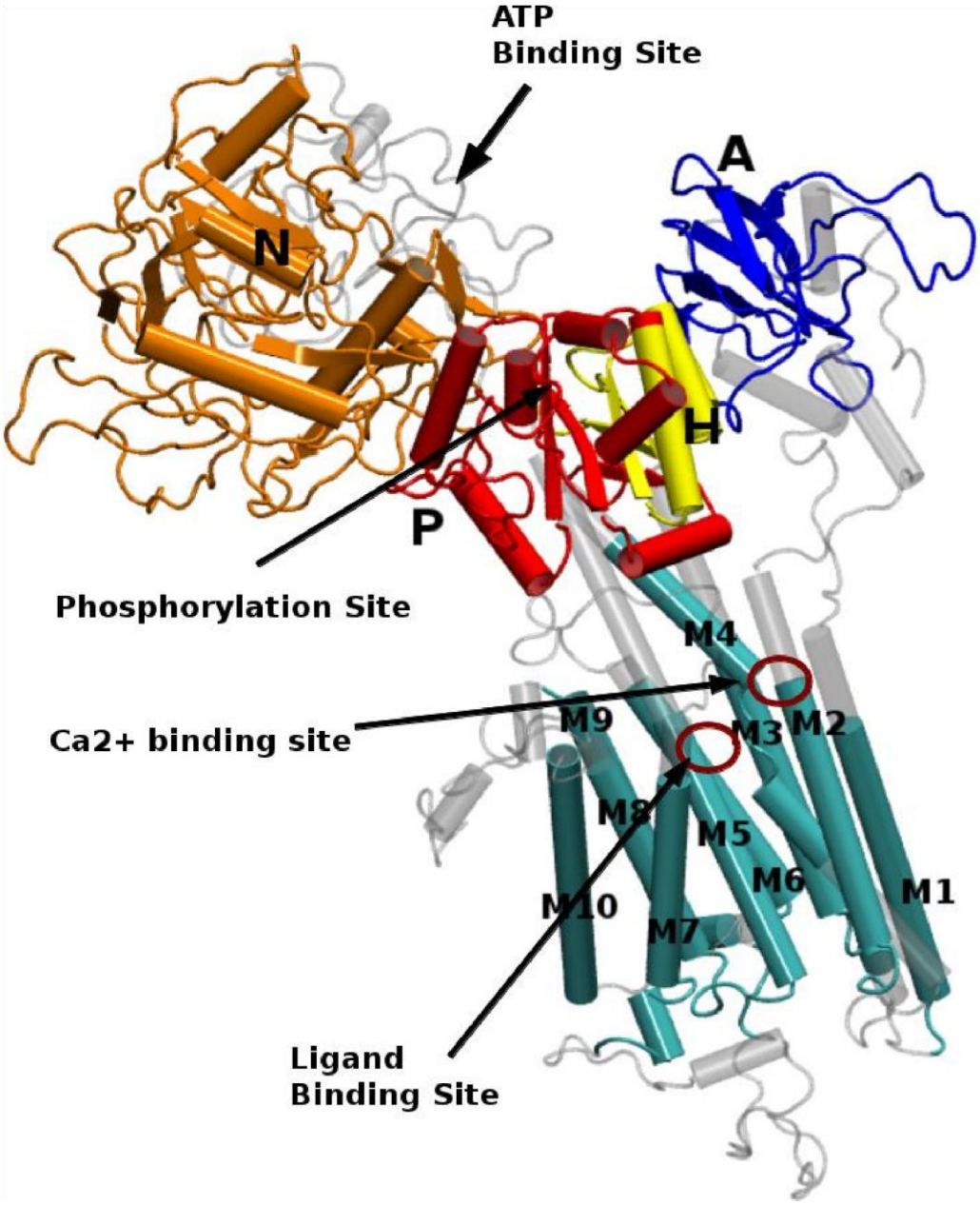


Fig. S3. Overview of the structure of the Ca^{2+} bound PfATP6. Domains depicted are nucleotide domain (N) (orange), phosphorylation domain (red), actuator domain (blue), hinge domain (H) (yellow) and the transmembrane region including 10 helices (M1-M10) (cyan). Marked circle between M3, M5 and M7 is the ligand binding site and the circle between M2, M3 and M4 is the calcium binding site.

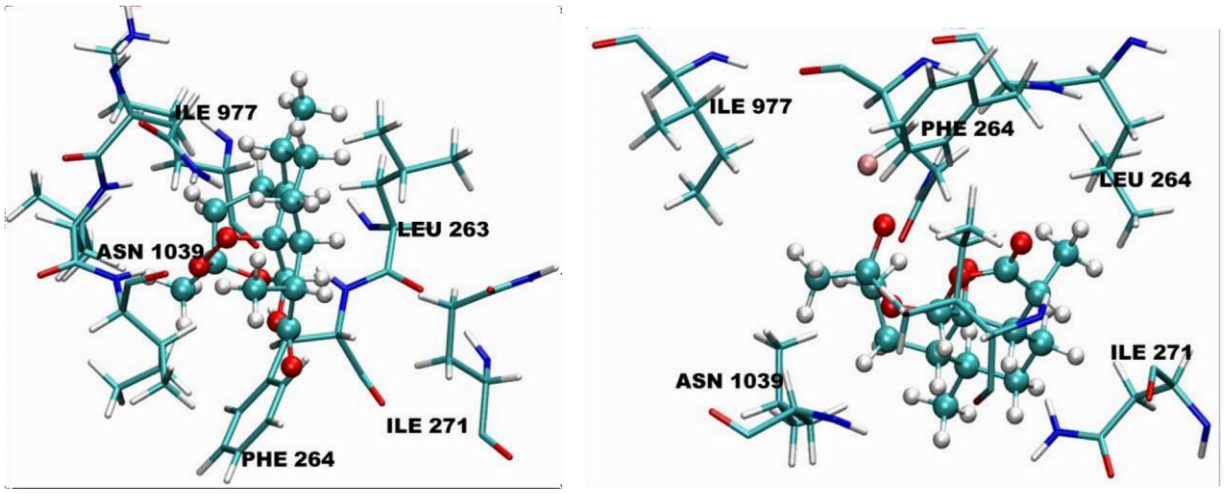


Fig. S4. Initial docked structure of Artemisinin with SERCA on the left; Fe-artemisinin adduct docked to SERCA on the right.

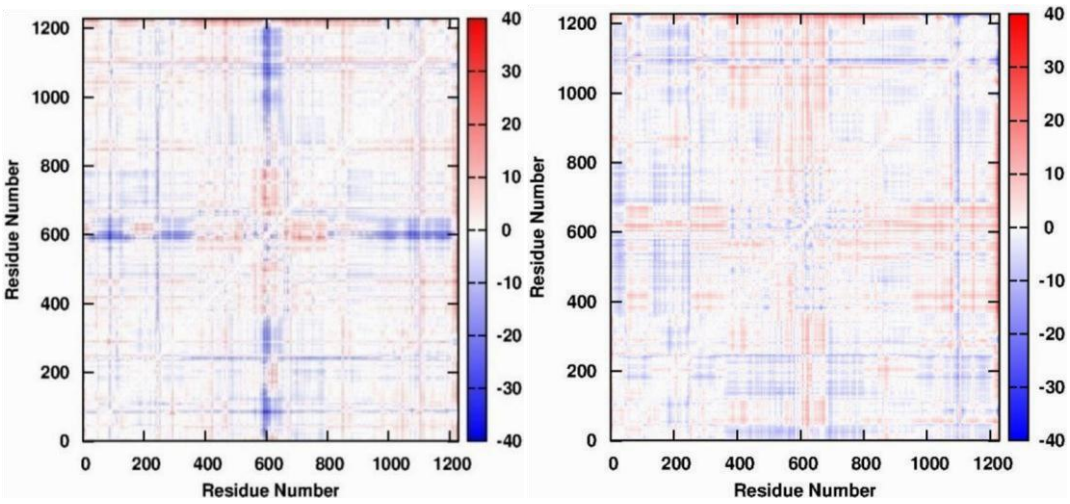


Fig. S5. Differential contact map artemisinin and Fe-activated artemisinin with Free SERCA showing Nucleotide domain and actuator domain coming closer whereas with Fe-activated artemisinin they are moving farther. Residues 580-620 is part of Nucleotide domain and residue 1-50 and 220-270 is part of actuator domain. Blue region shows decrease in distances whereas red region shows increase in distance throughout the trajectory.

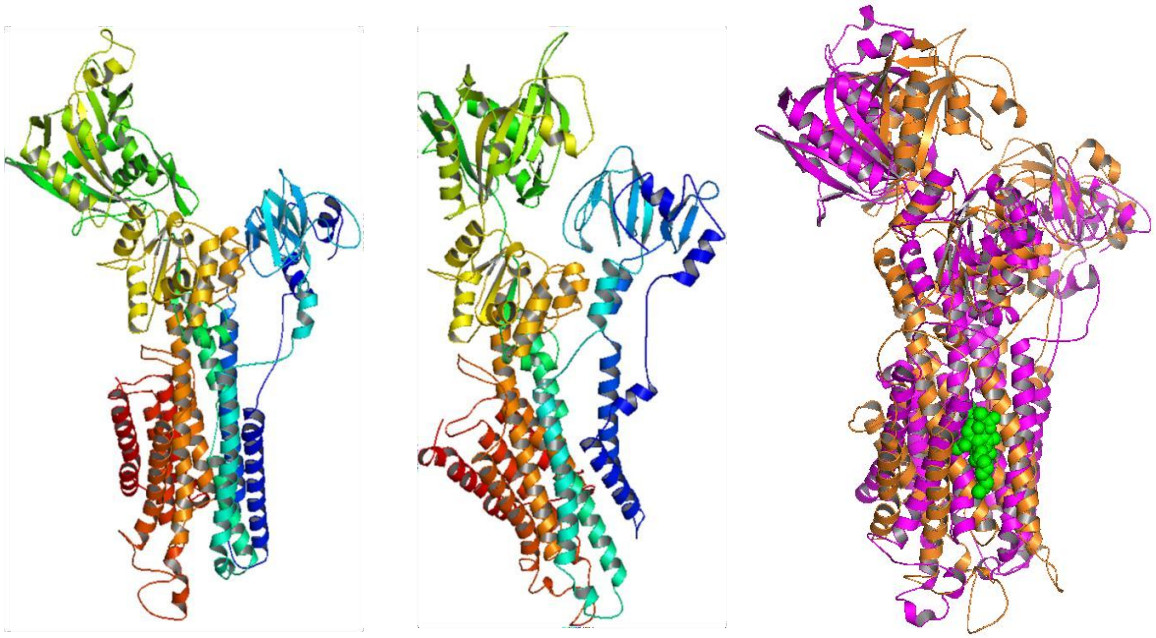


Fig S6. Crystal structure of mammalian SERCA (1SU4) Cytosolic region including Nucleotide binding domain and Actuator domain getting closed after 40 ns of simulation in presence of Thapsigargin. We have also shown the superimposed image of crystal structure of mammalian SERCA prior (pink) and after (orange) the MD simulation. RMSD between these two structures is 11 Å.

For further clarity we have also incorporated a distance plot between the interatomic distances between side chain atoms of ASP 133 and LYS464 from nucleotide binding domain and actuator domain respectively of the mammalian SERCA.

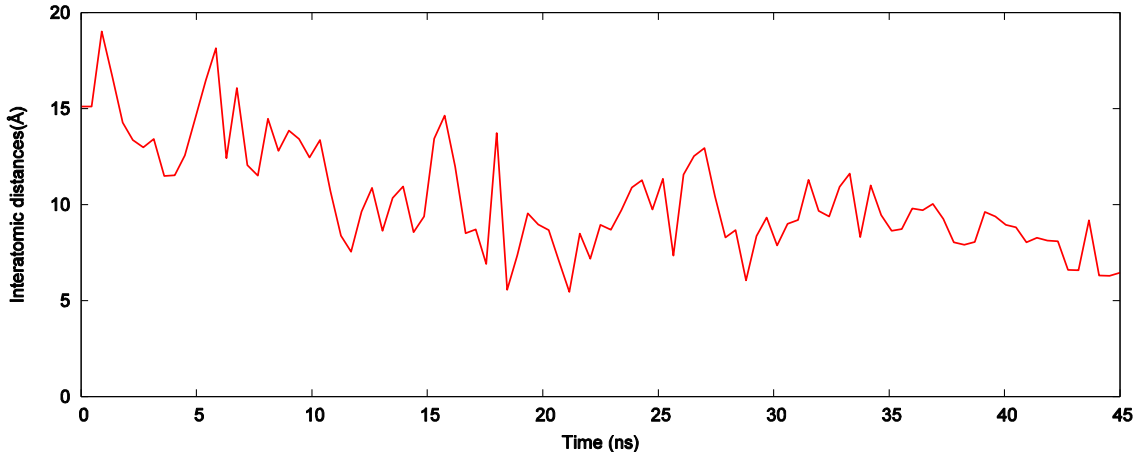


Fig S7. Interatomic distances between side chains of ASP 133 and side chain of LYS 464 from the cytoplasmic region (nucleotide binding domain and actuator domain respectively)

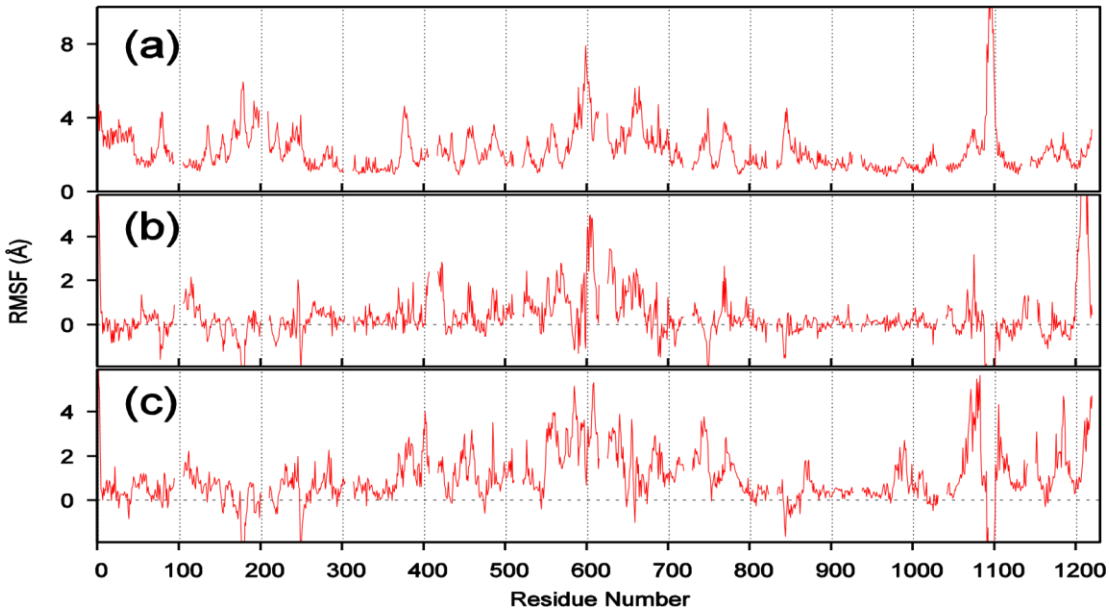


Fig. S8. RMSF plot for (a) *pf*ATP6 enzyme, (b) artemisinin- *pf*ATP6 system and (c) Fe-artemisinin adduct *pf*ATP6 system. For artemisinin and Fe-artemisinin bound system rmsf difference with respect to *pf*ATP6 enzyme is plotted. Residue number 580 to 700 is the loop region of Nucleotide binding domain. Residue 256-282 are part of M1 helix.

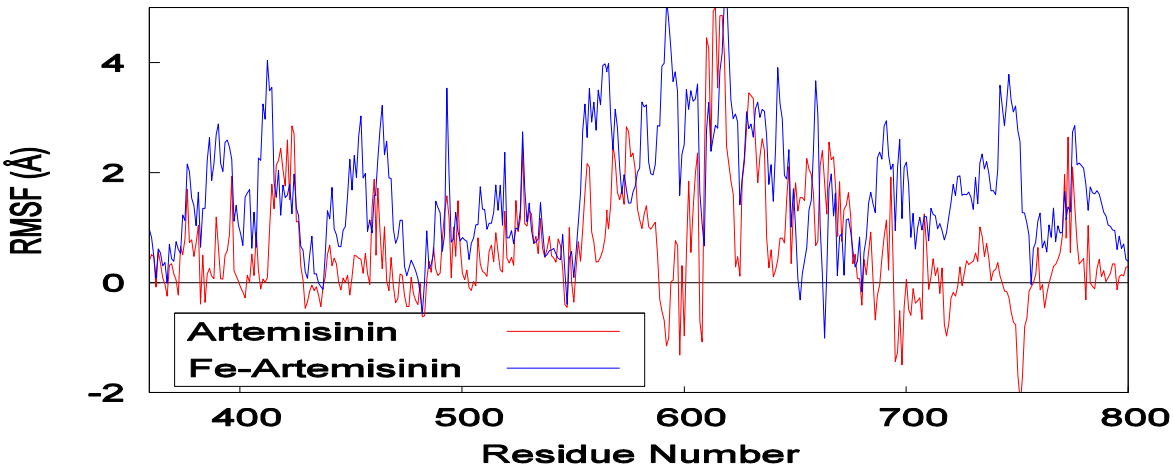


Fig. S9. Differential root mean square fluctuation map of Artemisinin (blue) and Fe-artemisnin adduct (red) bound SERCA with Ca^{2+} bound SERCA.

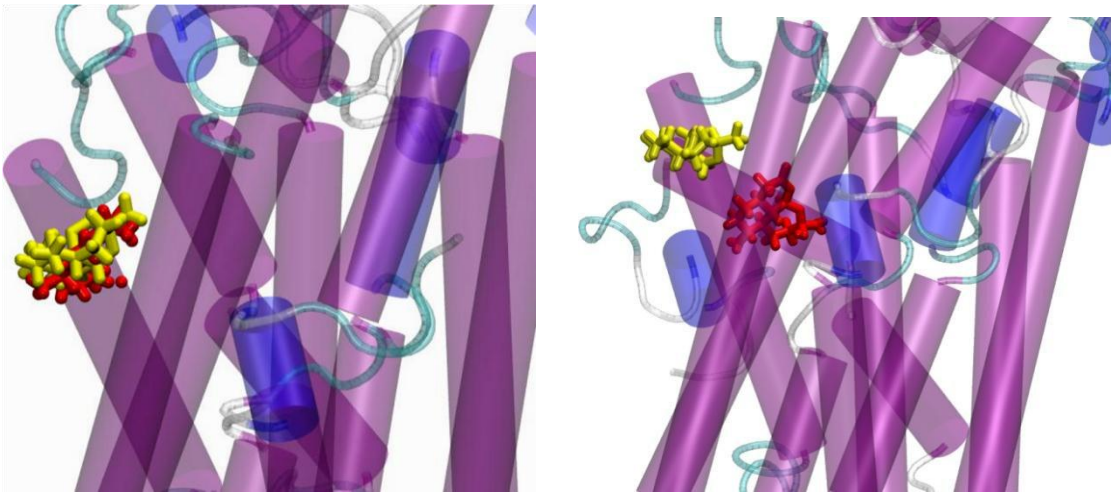


Fig. 5a Fig. 5b

Fig. S10. (a) Artemisinin (yellow) and Fe-artemisinin (red) adduct docked to *pf*ATP6 at the start of the MD simulation initial coordinates. (b) Fe-artemisinin adduct (red) moves towards the ligand binding site disrupting the helical region after 40ns.

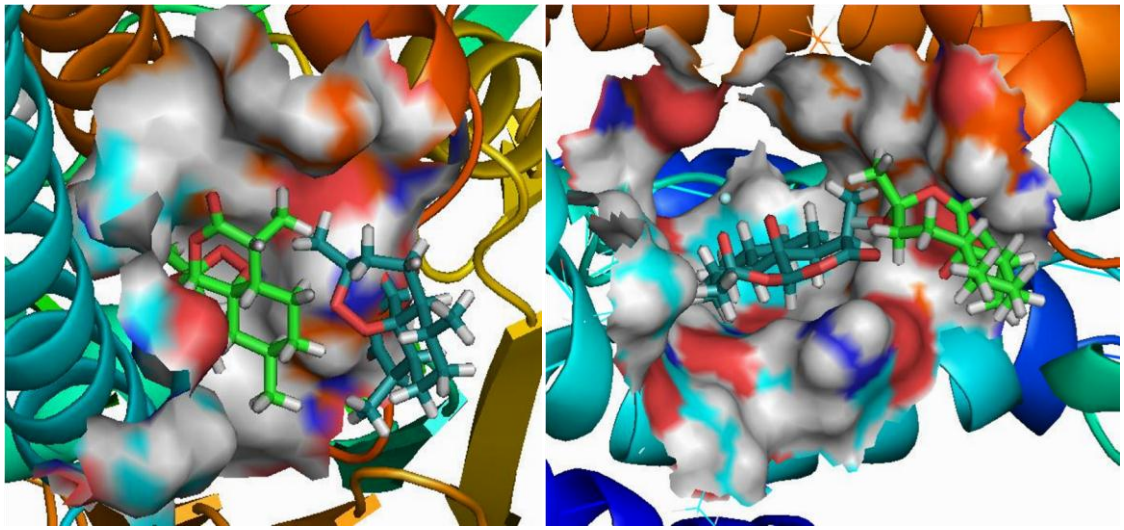


Fig. S11. Artemisinin (cyan) coming out of the cavity after simulation (see left panel) and Fe-activated artemisinin (cyan) moves inside the membrane region or the defined cavity (right panel).

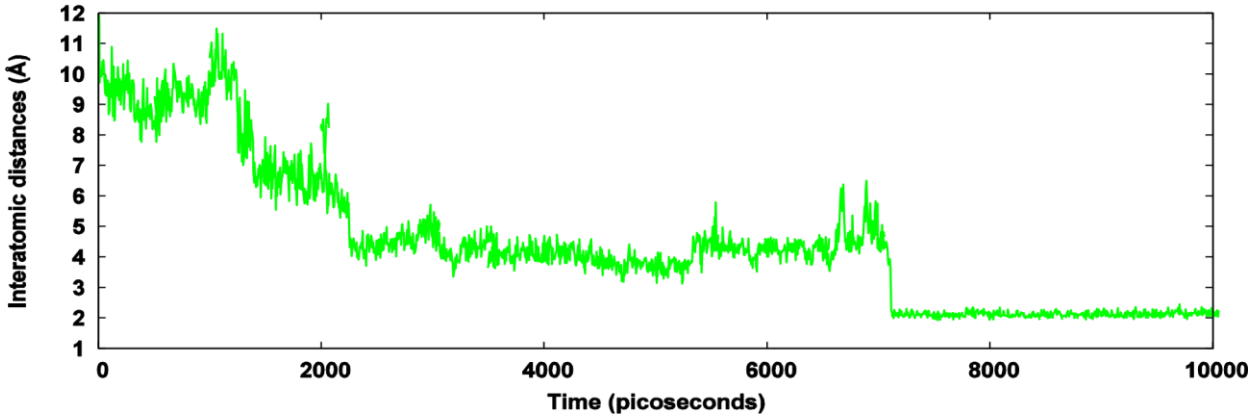


Fig. S12. Inter-atomic distance between Fe-artemisinin adduct and side chain of residue ASN 980

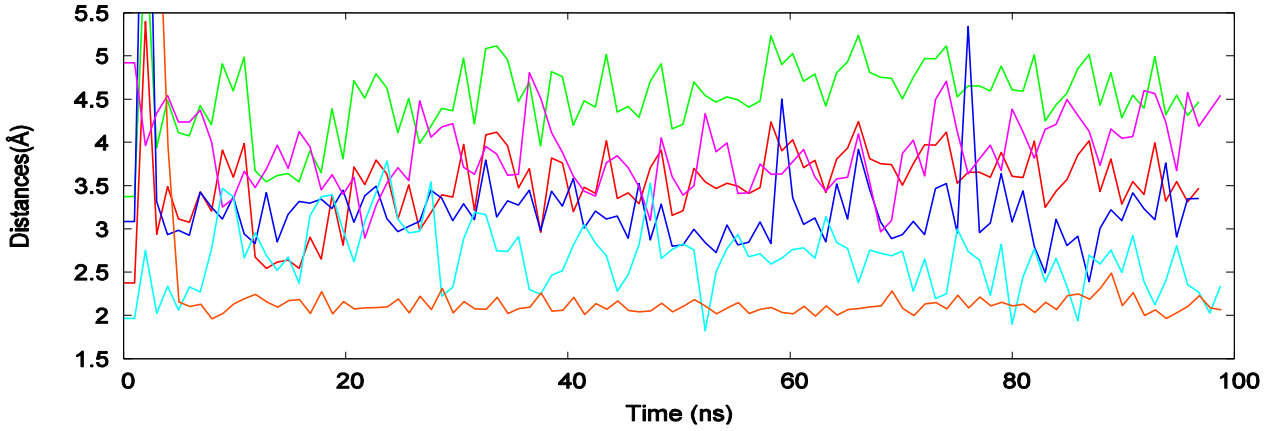


Fig. S13. Inter atomic distances between Artemisinin carbonyl oxygen atom and N-H of the main chain LEU 263 (green), ILE 977 (pink), ASN 1039 (red) which is more than the hydrogen bond cutoff of 3.2 Å however inter atomic distances between atoms of Fe-artemisinin adduct and LEU 263 (blue), ILE 977 (cyan), ASN 1039 (orange) shown is less than the cutoff distance which is modulating the motion of two domains.

Supplementary Table 1. Partial atomic charges

Molecule	Selective Atoms (Charge e^{-})
Artemisinin	O_1 (-0.323), O_2 (-0.321), O_{11}(-0.528), O_{13}(-0.591), O_{14}(-0.590)
Deoxy-Artemisinin	O_1 (-0.586), O_{10}(-0.543), O_{12}(-0.577), O_{13}(-0.587)
Heme-Artemisinin (triplet)	O_1 (-0.561), O_2 (-0.585), O_{11}(-0.548), O_{13}(-0.596), O_{14}(-0.665), Fe(0.710)

Supplementary Table 2.

Comparative RMSDs of three dimensional structures of Ca^{2+}-ATPase representative and there transport intermediates

Functional State	PDB id. Organism *Oryctolygus cunniculus*	Seq. length 994 Resolution (Å)	Ligands/ Identifiers	RMSD (Å) A-domain	RMSD (Å) N-domain	RMSD (Å) All atom	Reference
Ca^{2+}E1	1SU4	2.40		0.0	0.0	0.0	*Nature* **405** 647 (2000)
HnE2	1IWO	3.10	TG1	1.088	1.179	13.97	*Nature* **418** 605 (2002)
Ca^{2+}E2	1KJU	6.0*	ACP	2.811	1.58	13.29	*Jour. of Mol. Bio* **316** 201 (2002)
Ca^{2+}-E1-AMPPCP	1T5S	2.6	ADP	1.842	1.279	13.55	*Science* **304** 1672 (2004)
Ca^{2+}-E1-ADP:AlF_4-	1T5T	2.9	ADP, ALF	1.842	1.323	13.58	*Science* **304** 1672 (2004)
E1-AMPPCP	1VFP	2.9	ACP	1.763	1.768	13.75	*Nature* **430** 529 (2004)
E2-AlF_4-	1XP5	3.0	ALF	0.906	1.401	13.46	*Science* **306** 2251 2004
E2·MgF_4^{2-}	1WPG	2.3	ADP, MF4	1.117	0.738	13.79	*Nature* **432** 361 2004
E2-(TG) BHQ	2AGV	2.4	TG, BHQ	0.962	1.046	14.02	*PNAS* **102** 14589 2005
E2(TG)	2C8K	2.8	ACP	0.993	1.547	13.93	*Embo. J.* **25** 2305 2006
E2(TG)	2C8L	3.1	TG1	0.974	1.654	13.56	*Embo. J.* **25** 2305 2006
E2(TG):AMPPCP	2C88	3.1	ACP	1.003	1.554	13.96	*Embo. J.* **25** 2305 2006
Ca^{2+}E1	2C9M	3.0		0.367	1.096	2.447	*Embo. J.* **25** 2305 2006

	2DQS	2.5	ACP, TG1	0.993	1.547	13.92	To be published
E2(TG)P2	2EAR	3.10	TG1	0.864	1.014	13.94	*PNAS* **104** 5800 (2007)
E2CPA	2EAS	3.40	CPA	1.215	1.667	14.01	*PNAS* **104** 5800 (2007)
E2(CPATTG)	2EAT	2.90	CZA,TG1	1.189	1.601	13.84	*PNAS* **104** 5800 (2007
E2(CPATCC)	2EAU	2.80	CZP, PT4	1.211	1.658	13.93	*PNAS* **104** 5800 2007)
E2-P_1	2ZBD	2.40	ADP	0.92	1.016	13.79	*Nature* **432** 361 (2004)
E2-BF_3	2ZBE	3.80	PC1	1.166	1.766	13.57	*PNAS* **104** 19831 (2007)
E2-BF_3.TG	2ZBF	2.40	BEF	0.871	1.339	12.39	*PNAS* **104** 19831 (2007)
E2-ALF.TG	2ZBG	2.55	ALF,TG1	0.941	0.797	13.76	PNAS **104** 19831 (2007
E1-AMPPCP	3AR2	2.50	ACP	1.197	1.772	13.58	*PNAS* **108** 1833 (2011)
E2-ADP(TG)	3AR3	2.30	ADP	0.852	1.038	13.96	*PNAS* **108** 1833 (2011)
E2-ATP(TG)	3AR4	2.15	ATP,TG1	0.861	1.042	13.93	*PNAS* **108** 1833 (2011)
E2-TNP-AMP (TG)	3AR5	2.20	TNP, AMP, TG1	0.868	1.039	13.96	*PNAS* **108** 1833 (2011)
E2-TNP-ADP (TG)	3AR6	2.20	TNP, ADG, TG1	0.887	1.019	13.94	*PNAS* **108** 1833 (2011)
E2-TNP-ATP (TG)	3AR7	2.15	128, PT4, TG1	0.897	1.037	13.96	*PNAS* **108** 1833 (2011)
E2-ALF_4-TNP-AMPCC	3AR8	2.60	ALF, TG1, TM1	0.95	0.781	13.74	*PNAS* **108** 1833 (2011)

E2-BeF_3-TNP-AMP(TG)	3AR9	2.60	BEF, TG1, TM1	0.893	1.343	12.33	*PNAS* **108** 1833 (2011)
Ca^{2+} E1P	3BA6	2.80	AN2	1.146	1.297	12.14	*Nature* **450** 1036 (2007)
E2-AlF_4^-	3N5K	2.20	ACT ALF	1.228	0.894	13.41	*JBC* **288** 10759 (2013)
E2-BeF_3	3B9B	2.65	BEF	1.79	1.32	13.55	*Nature* **450** 1036 (2007)
E2-AlF_4-	3B9R	3.00	ACP, ALF	0.918	1.626	13.57	*Nature* **450** 1036 (2007)
Ca^{2+}-E_1-CaAMPPCP	3N8G	2.59	ACP	1.835	1.336	13.57	*JBC* **288** 10759 (2013)
E1.Ca^{2+}.ACP	3TLM#	2.95	ACP	1.937	2.778	13.76	*Jour. Str. Biol.* **178** 38 (2012)
E1.Mg^{2+}	3W5A	3.01	PTY, TM1	0.765	1.133	9.67	*Nature* **495** 260 (2013)
E1.Mg^{2+}	3W5B	3.20	PTY, TM1	0.89	1.157	9.57	*Nature* **495** 260 (2013)
E2	3W5C	2.50	PTY	0.959	1.041	13.85	*Nature* **495** 260 (2013)
E2 + Pi	3W5D	2.45	PTY, SO4	0.875	0.985	13.83	*Nature* **495** 260 (2013)
E1-SLN	4H1W	3.10	ACP	0.792	1.093	10.60	*Nature* **495** 265 (2013)

*Resolution of 1KJU is with Electron microscopy

#3TLM is from Bos Tauras (cattle) as rest of them are from the organism *Oryctolagus Cunniculus*

Details of docking protocol:

Identification of best possible grid/translational points in radius of 3Å around the reference points. b) Generation of protein grid and preparation of energy grid in and around the active site of protein to re calculate the energy of each atom in the candidate ligand c) Monte Carlo docking and intensive configurational search of the ligand in the active site and d) Identification of the best docked structures on an energy criterion and prediction of the binding free energy of the complex. Protein grid of length 10Å is generated in order to identifying the protein residues in a specified range which are interacting with the atoms of the candidate ligand. The Monte Carlo configurations are generated around six degrees of freedom which result in many combinations of ligand configuration, energy points are selected and pre calculated energy of each atom is added. The scoring function employed considers the non-bonded energy of a protein-ligand complex as a sum of three energy terms electrostatic, van der Waals and hydrophobic.