

Supporting Information

Fragment Molecular Orbital Study of the Interaction between Sarco/Endoplasmic Reticulum Ca^{2+} -ATPase and its Inhibitor Thapsigargin toward Anti-Malarial Development.

Takeshi Ishikawa^{1,2,*}, Satoshi Mizuta¹, Osamu Kaneko^{2,3}, and Kazuhide Yahata³

¹*Department of Molecular Microbiology and Immunology, Graduate School of Biomedical Sciences, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan*

²*Leading Program, Graduate School of Biomedical Sciences, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan*

³*Department of Protozoology, Institute of Tropical Medicine (NEKKEN), Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan*

Table of Contents

Scheme S1. Organic synthesis of TG-acetate

Table S1. PDB codes used for the distance analysis

Figure S1. Sequence alignment

Figure S2. RMSD values between the template and models

Figure S3. Interaction energies for eight species of malaria parasite

Scheme S1. Organic synthesis of TG-acetate

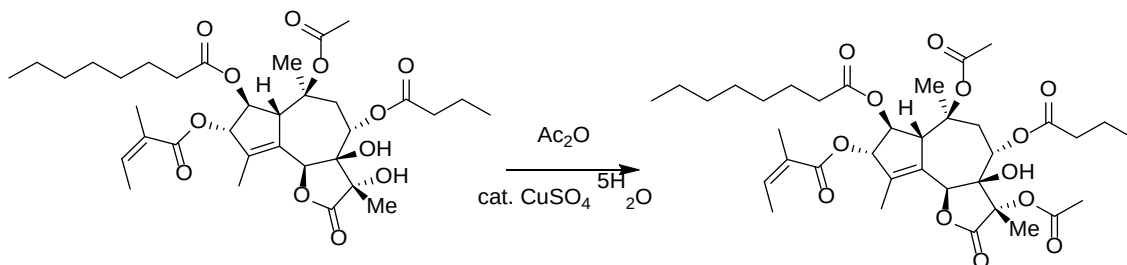
S1-1. General Information

All NMR spectra were recorded on Varian 500PS spectrometers. ^1H and ^{13}C NMR spectra are reported as chemical shifts (δ) in parts per million (ppm) relative to the solvent peak using tetramethylsilane and (2,2,3,3-D $_4$)-trimethylsilyl-3-propanoic acid, sodium salt (^1H and ^{13}C), as an internal standard. δ is quoted in ppm and coupling constants (J) are measured in hertz (Hz). The following abbreviations are used to describe multiplicities: s = singlet; d = doublet; t = triplet; q = quartet; quint. = quintet; sext. = sextet; sept. = septet; br = broad; and m = multiplet. NMR spectra were processed in ACD/SpecManager. High-resolution mass spectra (HRMS, m/z) were obtained on JEOL JMS-700N for FAB using m-nitrobenzylalcohol as a matrix or on JEOL JMS-T100TD for electrospray ionization. All reactions were performed in apparatuses with magnetic stirring under an inert atmosphere. Flash column chromatography was performed over Fuji Silysia Chemical Ltd. silica gel C60 (50–200 μm) and CHROMATOREX DIOL (MB 100-40/75) using an eluent system as described for each experiment. Thin-layer chromatography was performed using TLC Silica gel 60 F $_{254}$ aluminum sheets (Merck) and silica gel F $_{254}$ glass plates (Merck).

Materials

Unless otherwise stated, all starting materials and reagents were obtained from commercial suppliers and were used without further purification. All chemicals were purchased from Sigma-Aldrich, Nacalai Tesque, Tokyo Chemical Industry or Wako Pure Chemical Industries and used as received. Thapsigargin was purchased from Funakoshi. All solvents were purchased from Wako Pure Chemical Industries.

S1-2. Experimental Procedures and Characterization Data

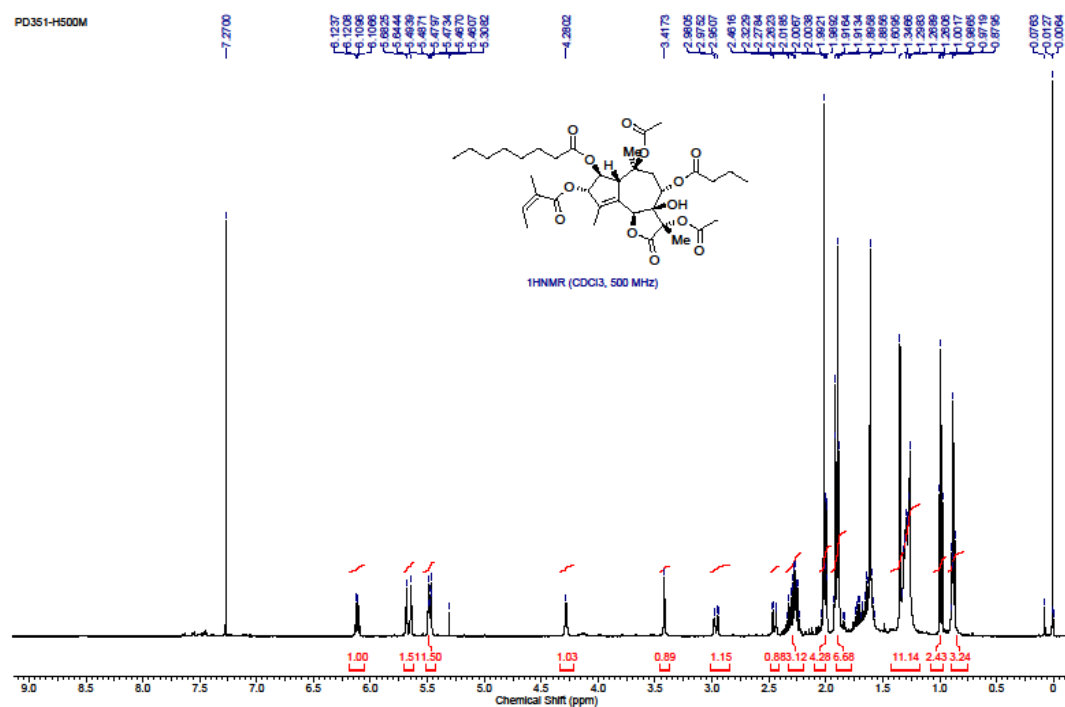


A mixture of thapsigargin (10 mg, 0.015 mmol), acetic anhydride (2.0 mL), and $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (1.8 mg, 7.5 μmol) was stirred at room temperature for 24 h. Upon completion of the reaction (monitored by TLC), the mixture was diluted with distilled water and the precipitate was filtered and dried in vacuum, affording the desired product as a white solid (8.0 mg, 77% yield).

**(3*S*,3*aR*,4*S*,6*S*,6*aR*,7*S*,8*S*,9*bS*)-3,6-(diacetyloxy)-4-(butyryloxy)-3*a*-hydroxy-3,6,9-trimethyl-8-
 {[(2*Z*)-2-methylbut-2-enoyl]oxy}-2-oxo-2,3,3*a*,4,5,6,6*a*,7,8,9*b*-decahydroazuleno[4,5-*b*]furan-7-
 yl octanoate**

¹H NMR (500 MHz, CDCl₃) δ 0.88 (t, *J* = 7.1 Hz, 3H), 0.99 (t, *J* = 7.3 Hz, 3H), 1.23–1.33 (m, 8H), 1.35 (s, 3H), 1.57–1.74 (m 4H), 1.61 (s, 3H), 1.89 (s, 3H), 1.90 (s, 3H), 1.94–1.93 (m, 3H), 2.00 (dq, *J* = 1.5, 7.3 Hz, 3H), 2.02 (s, 3H), 2.24–2.31 (m, 4H), 2.45 (dd, *J* = 3.7, 14.7 Hz, 1H), 2.96 (dd, *J* = 2.7, 14.7 Hz, 1H), 3.42 (s, 1H), 4.28 (s, 1H), 5.47 (t, *J* = 3.2 Hz, 1H), 5.49 (d, *J* = 3.4 Hz, 1H), 5.64 (s, 1H), 5.68 (s, 1H), 6.12 (qq, *J* = 1.5, 7.4 Hz, 1H); **¹³C NMR** (125 MHz, CDCl₃) δ 13.0, 13.7, 14.1, 15.8, 16.6, 18.0, 20.6, 21.1, 22.5, 22.6, 23.0, 24.8, 28.9, 29.0, 29.7, 31.6, 34.2, 36.3, 37.1, 57.7, 67.7, 77.4, 77.7, 78.8, 82.3, 84.0, 84.1, 127.4, 129.6, 138.8, 142.0, 167.0, 168.7, 170.7, 171.2, 172.3, 172.5; **HRMS** (FAB) *m/z* Calcd for C₃₆H₅₂NaO₁₃ [M+Na]⁺ 715.3306, found 715.3305.

S1-3. NMR spectra



PD351CS00M

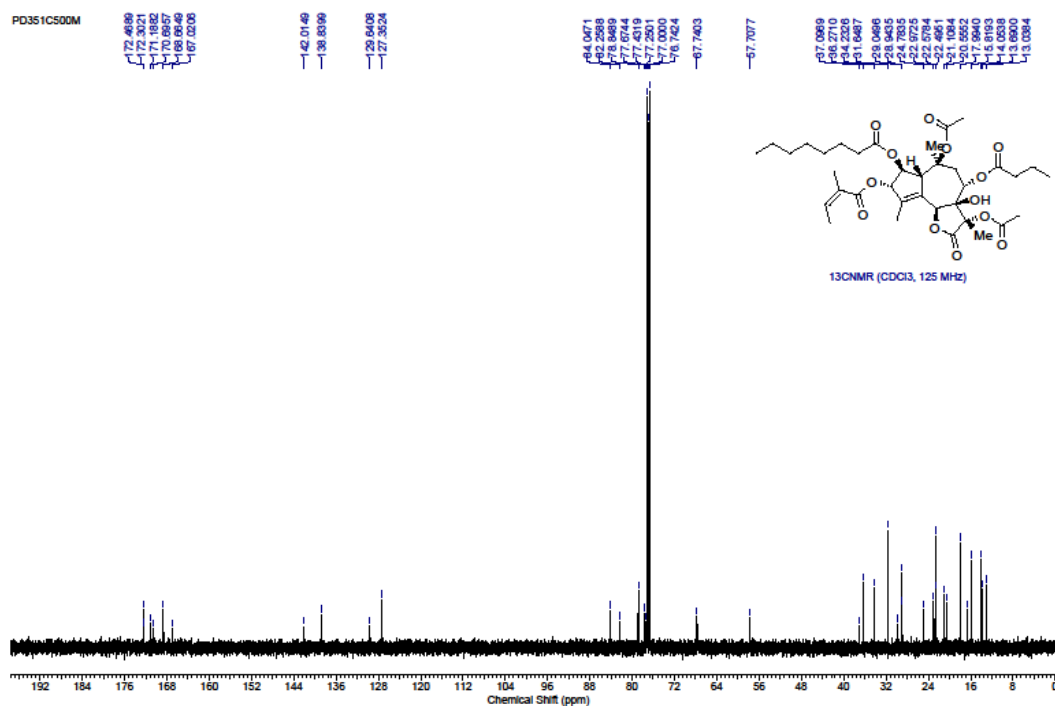


Table S1. *PDB codes used for the distance analysis*

PDB codes of X-ray structures used in the analysis of the distance between TG and amino acid residues of SERCA.

PDB code	Literature
1IWO	C. Toyoshima et al, Nature 418 (2002) 605
1WPG	C. Toyoshima et al, Nature 432 (2004) 361
1XP5	C. Olesen, et al, Science 306 (2004) 2251
2AGV	K. Obara, et al, Proc. Natl. Acad. Sci. USA 102 (2005) 14489
2C88	A. M. Jensen, et al, Embo J. 25 (2006) 2305
2C8K	A. M. Jensen, et al, Embo J. 25 (2006) 2305
2C8L	A. M. Jensen, et al, Embo J. 25 (2006) 2305
2EAR	M. Takahashi, et al, Proc. Natl. Acad. Sci. USA 104 (2007) 5800
2EAT	M. Takahashi, et al, Proc. Natl. Acad. Sci. USA 104 (2007) 5800
2ZBF	C. Toyoshima, et al, Proc. Natl. Acad. Sci. USA 104 (2007) 19831
2ZBG	C. Toyoshima, et al, Proc. Natl. Acad. Sci. USA 104 (2007) 19831
3AR3	C. Toyoshima, et al, Proc. Natl. Acad. Sci. USA 108 (2011) 1833
3AR4	C. Toyoshima, et al, Proc. Natl. Acad. Sci. USA 108 (2011) 1833
3AR5	C. Toyoshima, et al, Proc. Natl. Acad. Sci. USA 108 (2011) 1833
3AR6	C. Toyoshima, et al, Proc. Natl. Acad. Sci. USA 108 (2011) 1833
3AR7	C. Toyoshima, et al, Proc. Natl. Acad. Sci. USA 108 (2011) 1833
3AR8	C. Toyoshima, et al, Proc. Natl. Acad. Sci. USA 108 (2011) 1833
3N5K	M. Bublitz, et al, J. Biol. Chem. 288 (2013) 10759
4UU0	N. D. Drachmann, et al, FEBS J. 281 (2014) 4249
4UU1	N. D. Drachmann, et al, FEBS J. 281 (2014) 4249
2DQS	Not published (Deposited: 2006-05-29)

Figure S1. Sequence alignment

Sequence alignment of huSERCA (A) and *Pf*SERCA (B) with raSERCA. The sequence identity was 96.7% and 40.8%, respectively.

A

```

      10      20      30      40      50      60      70      80      90     100
raSERCA MEAAHKSSTEECLAYFGVSETTGLTPDQVKRHLKEYGHNELPAEEGKSLWELVIEQFEDLLVRILLAAACISFVLAWFEEGEETITAFVEPFVILLILIA
huSERCA MEAAHAKTTEECLAYFGVSETTGLTPDQVKRNLEKYGLNELPAEEGKTLWELVIEQFEDLLVRILLAAACISFVLAWFEEGEETITAFVEPFVILLILIA
*****

      110     120     130     140     150     160     170     180     190     200
raSERCA NAIVGVWQERNAENAI EALKEYEPEMGKVYRADRKSVQRIKARDIVPGDIVEVAVGDKVPADIRILSIKSTTLRVDQSILTGESVSVIKHTEPVPDPRAV
huSERCA NAIVGVWQERNAENAI EALKEYEPEMGKVYRADRKSVQRIKARDIVPGDIVEVAVGDKVPADIRILAISKSTTLRVDQSILTGESVSVIKHTEPVPDPRAV
*****

      210     220     230     240     250     260     270     280     290     300
raSERCA NQDKKNMFLSGTNI AAGKALGI VATTGVSTEIGKIRDQMAATEQDKTLPQQKLEDFGEQLSKVISLICVAVWLINIGHFNDPVGGSWIRGAIYYFKIAV
huSERCA NQDKKNMFLSGTNI AAGKALGI VATTGVSTEIGKIRDQMAATEQDKTLPQQKLEDFGEQLSKVISLICVAVWLINIGHFNDPVGGSWFRGAIYYFKIAV
*****

      310     320     330     340     350     360     370     380     390     400
raSERCA ALAVAAIPEGLPAVITTCALGTRRMMAKNNAIVRSLPSVETLGCTSVICSDKTGTLTTNQMSVCKMFIIDKVDGDFCSLNEFSITGSTYAPEGEVLKNDK
huSERCA ALAVAAIPEGLPAVITTCALGTRRMMAKNNAIVRSLPSVETLGCTSVICSDKTGTLTTNQMSVCKMFIIDKVDGDI CLLNEFSITGSTYAPEGEVLKNDK
*****

      410     420     430     440     450     460     470     480     490     500
raSERCA PIRSGQFDGLVELATICALCNDSSLDNETKGYYEKVGEATELTTLVEKMNVFNTDVRSLSKVERANACNSVIRQLMKKEFTLEFSRDRKSMSSVYCSP
huSERCA PVRPGQYDGLVELATICALCNDSSLDNEAKGVYKVGGEATELTTLVEKMNVFNTDVRSLSKVERANACNSVIRQLMKKEFTLEFSRDRKSMSSVYCSP
* * *

      510     520     530     540     550     560     570     580     590     600
raSERCA AKSSRAAVGNKMFVKGAPEGVIDRCNYVRVGTTRVPMTGPVKEKILSVIKEWGTGRDTRLCLALATRDTPPKREEMVLDSSRFMEYETDLTFVGVVGM
huSERCA AKSSRAAVGNKMFVKGAPEGVIDRCNYVRVGTTRVPLTGPVKEKIMAVIKEWGTGRDTRLCLALATRDTPPKREEMVLDSSARFLEYETDLTFVGVVGM
*****

      610     620     630     640     650     660     670     680     690     700
raSERCA DPPRKEVMSGIQLCRDAGIRVIMITGDNKGTAIAICRRIGIFGENEEVADRAYTGREFDDLPLAEQREACRRACCFARVEPSHKSKIYEYLSYDEITAM
huSERCA DPPRKEVTGSIQLCRDAGIRVIMITGDNKGTAIAICRRIGIFGENEEVADRAYTGREFDDLPLAEQREACRRACCFARVEPSHKSKIYEYLSYDEITAM
*****

      710     720     730     740     750     760     770     780     790     800
raSERCA TGDGVNDAPALKKAEIGIAMSGGTAVAKTASEMVLADDFNSTIVAAVEEGRAIYNNMKQFIRYLISSNVGEVVCIFLTAALGLPEALIPVQLLWVNLVTD
huSERCA TGDGVNDAPALKKAEIGIAMSGGTAVAKTASEMVLADDFNSTIVAAVEEGRAIYNNMKQFIRYLISSNVGEVVCIFLTAALGLPEALIPVQLLWVNLVTD
*****

      810     820     830     840     850     860     870     880     890     900
raSERCA GLPATALGFNPDDLIMDRPPRSPKEPLISGWLFFRYMAIGGYVGAATVGAAAWFMYAEDGPGVTYHQLTHFMQCTEDHPHFEGLDCEIFEAPEPMTMA
huSERCA GLPATALGFNPDDLIMDRPPRSPKEPLISGWLFFRYMAIGGYVGAATVGAAAWFLYAEDGPHVNYSQLTHFMQCTEDNTHFEGIDCEVFEAPEPMTMA
*****

      910     920     930     940     950     960     970     980     990
raSERCA LSVLVTIEMCNALNSLSENQSLMRMPWVNIWLLGSI CLSMSLHFLILYVDPLPMIFKLKALDLTQWLMVLKISLPVIGLDEILKFVARNYLEG
huSERCA LSVLVTIEMCNALNSLSENQSLMRMPWVNIWLLGSI CLSMSLHFLILYVDPLPMIFKLKALDLTQWLMVLKISLPVIGLDEILKFVARNYLEG
*****

```

B

```

      10      20      30      40      50      60      70      80      90     100
raSERCA  ME---AAHSKSTEECLAYFGVSETTGLTPDQVKRHLEKYGHNELPAEEGKSLWELVIEQFEDLLVRILLAAACISFVLAWFE-EGEE-TITAFVEPFVI
P/SERCA  MEEVIKNAHTYDVEDVLKFLDVNKDNLKNEELDDRRLLKYGLNELEVEKKKSI FELILNQFDDLLVKILLAAFI SFVLTLLDMKHKKI EICDFIEPLVI
      **      **      * *      *      **      *** **      * **      **      **      * * * * *      * * * *

      110     120     130     140     150     160     170     180     190     200
raSERCA  LLILIANAIVGVWQERNAENAI EALKEYEPEMGKVYRADRKSVQRIKARDIVPGDIVEVAVGDKVPADIRILSIKSTTLRVDQSILTGESVSVIKHTEPV
P/SERCA  VLILILNAAVGVWQECNAEKSLEALKELOPTKAKVLR-DGK-WEIIDS KYLYVGD I IELSVGNKTPADARI IKIYSTLSLKVEQSMLTGESCSVDKYAEKM
      **** * * * * * * *      * * * * *      * * * * *      * * * * *      * * * * *      * * * * *

      210     220     230     240     250     260     270     280     290     300
raSERCA  PDP-RAVN-ODKKNMLFSGTNIAAGKALGIVATTGVSTEIGKIRDQMAAT--EQDKTPLQQKLDEFGEQLSKVISL ICVAVWLINIGHFNDPVHGGSWIR
P/SERCA  EDSYKNCIEQLKKNILFSSTAIVCGRCIAVVINIGMKTEIGHIQHAVIESNSEDQTPLQIKIDLFGQQLSKI IFVICVTWII NFKHFSDP IHG-SFLY
      *      * * * * * * *      * * * * *      * * * * *      * * * * *      * * * * *      * * * * *

      310     320     330     340     350     360     370     380     390     400
raSERCA  GAIYYFKIAVALAVAAIPEGLPAVITTCALGTRRMMAKKNAIVRSLPSVETLGCTSVICSDKTGTLT TNQMVSCKMFI IDKVD-----GD-FCSL
P/SERCA  GC LY YFKI SVALAVAAIPEGLPAVITTCALGTRRMVKKNAIVRKLSVETLGCTTVICSDKTGTLT TNQMTTTFVHLFRESDSL TEYQLCQKGDYYFF
      *      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

      410     420     430     440     450     460     470     480     490     500
raSERCA  NEFSITGSTYAPE-----G-EVL-----KND-----KPIRSGQ--FDGLVELAT-----
P/SERCA  ESSNLTNDIYAGESSFFNKLKDEGNVEALTDGEEGSI DEADPYSDYFSSDSKKMKNDLNNNNNNNNSSRSAGAKRN I PLKEMKSNENT I ISRGSKILED
      *      * * *      * * *      ***      ***      * *

      510     520     530     540     550     560     570     580     590     600
raSERCA  -----ICAL-CNDSSLD FNETKGVEKVG EATETALTTLVEKMN VF-----NT-EV-----RNL-----SK-----
P/SERCA  KINKYCYSEYDYNFYMCLVNCNEANIFCNDNSQIVKKFGDSTELALLHFVHNFDILPTFSKNNKMPAEYKNTTPVQSSNKKDKSPRGINKFSSKNDNS
      *      * *      *      * * * * *      *      * * * * *      * * *      *      * *

      610     620     630     640     650     660     670     680     690     700
raSERCA  -----VERANACNSVIRQLMKK-----EFTLEF-----SR-----DRKSMSVYCSPA-----KS-----SR-----
P/SERCA  HITSTLNENDKNLKNANHSNYTTAQTNGYEAIGENTFEHGTSENFCHSKLGNKINTTSHNNNNNNNNNSNSVPSECISSWRNECKQIKI I EFTRR
      *      * *      *      * * *      *      *      *      * *      *      * *      *      *

      710     720     730     740     750     760     770     780     790     800
raSERCA  ---AA-VGNK---MFVKGAPEGVIDRCNYVRVGTTRVPMTPGVKKEILSVIKEWGTGRDTRLCLALATRDTPPKREEMVLDDSSRFMEYETDLTFVGV
P/SERCA  KLMSVIVENKKEI ILYCKGAPENI IKNCKYLLTKNDIRPLNETLKNEIHNKIQNMGK-R-ALRTL SFAYKKLSSK--DLN IKN TDDY YKLEQDL IYLG
      * *      * * * * *      * *      *      * * * * *      * * * * *      *      * * * *

      810     820     830     840     850     860     870     880     890     900
raSERCA  VGMLDPPRKEVMGSIQLCRDAGIRVIMITGDNKGTAIAICRRIGIFGENE---EVA---DRA-----YTGREFDDLPLAEQREACRRA--CCFARVEPS
P/SERCA  LGIIDPPRKYVGRAIRLCHMAGIRVFMITGDNINTARIAKEINILNKNEGDDEKDNYTNNKNTQICCYNGREFEDFSLEKQKHILKNTPRIVFCRTEPK
      *      * * * * *      * * * * * * * * * * * * * * *      * * * * *      * * * *

      910     920     930     940     950     960     970     980     990     1000
raSERCA  HKSKI VEYLQSYDEITAMTGDGVNDAPALKKAEIGIAMG-SGTAVAKTASEMVLADDNFSTIVA AVEEGRAIYNNMKQFI RYL I SSN VGEVVCIFLTAAL
P/SERCA  HKKQIVKVLKDLGETVAMTGDGVNDAPALKSADIGIAMGINGTEVAKEASD I VLADDNFNT IVEAIKEGRCIYNNMKAFIRYL I SSN IGEVASIFITALL
      ** * * *      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

      1010    1020    1030    1040    1050    1060    1070    1080    1090    1100
raSERCA  GLPEALIPVQLLWVNLVTDGLPATALGFNPDDLIMDRPPRSPK EPLISGWLFFRYMAIGGYVGAATVGA AAWFM YAEADGPGVT---YHQLTHFMQCTE
P/SERCA  GTPDSLAPVQLLWVNLVTDGLPATALGFNPPEHDVMCKKPRHKNDNL INGLTLLRYI I IGT YVG IATVS IFVYWFLFYPDSMDH LTFYQLSHYNQCKA
      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

      1110    1120    1130    1140    1150    1160    1170    1180    1190    1200
raSERCA  -DHPHFEGL-D-----CEIFEAP--EPMTMALSVLVT IEMCNALNSLSENGSLMRMPWPVNIWLLGSI CLSMLS HFL ILYVDPLPMIFKLKALDLTQWLML
P/SERCA  WNNFRVKNKYVDMSEDHCYSFSAGKIKASTLSLSVLVLIEMFNALNALSEYNSLFEIPPWNRNMYLVLATIGSLLHLVLIYIPPLARIFGVVPLSAYDWFL
      *      * * *      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

      1210    1220    1230
raSERCA  VLKISLPIGLDEILKFIARNYL-EG-----
P/SERCA  VFLWSFPV I I LDEI I KFYAKRKLKEEQRTKKIKID
      *      * * * * * * *      * *


```

Figure S2. RMSD values between the template and models

Red values are the RMSDs between the template and models obtained from the homology modeling.

Blue and red sticks are the three-dimensional structures of the template and models, respectively.

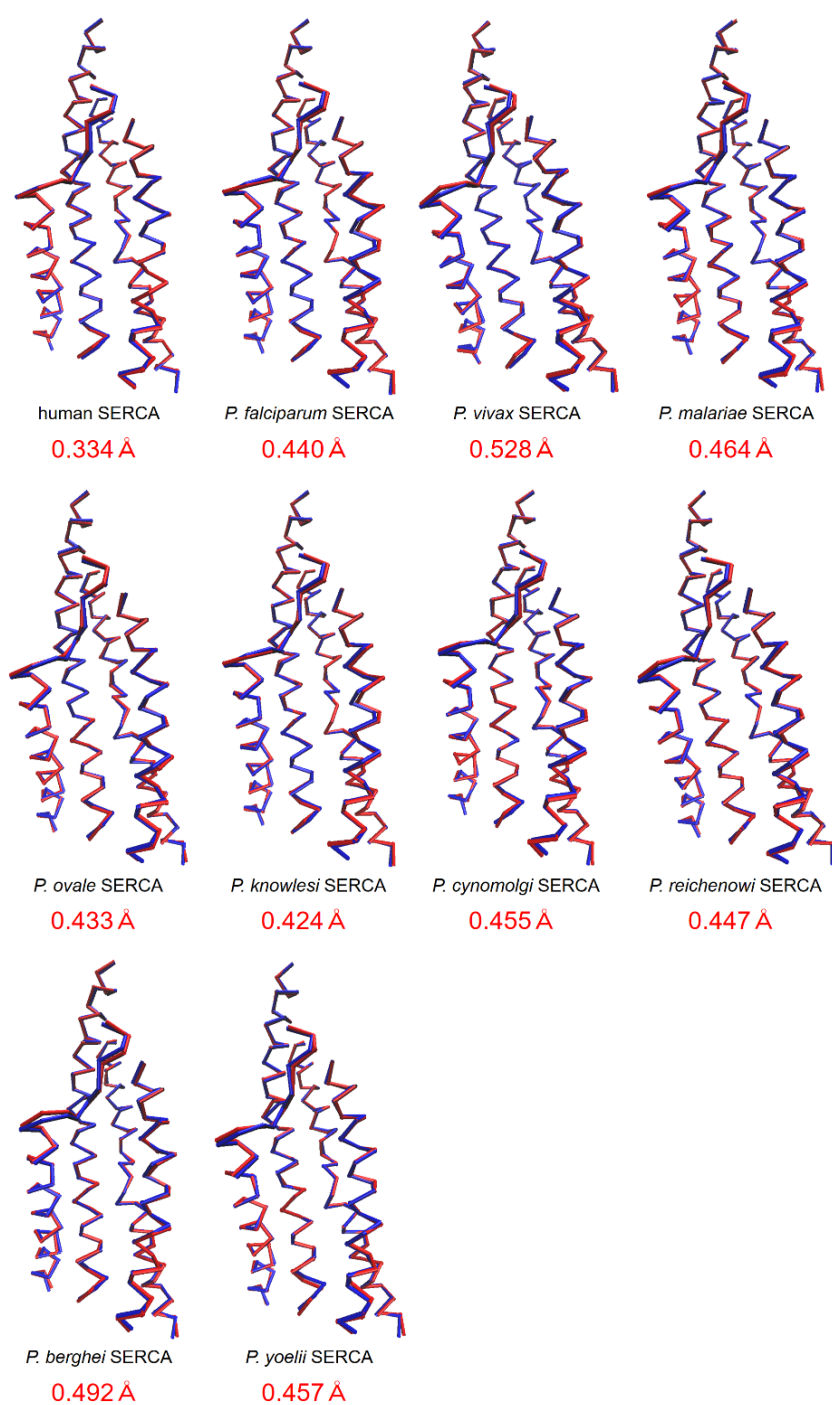
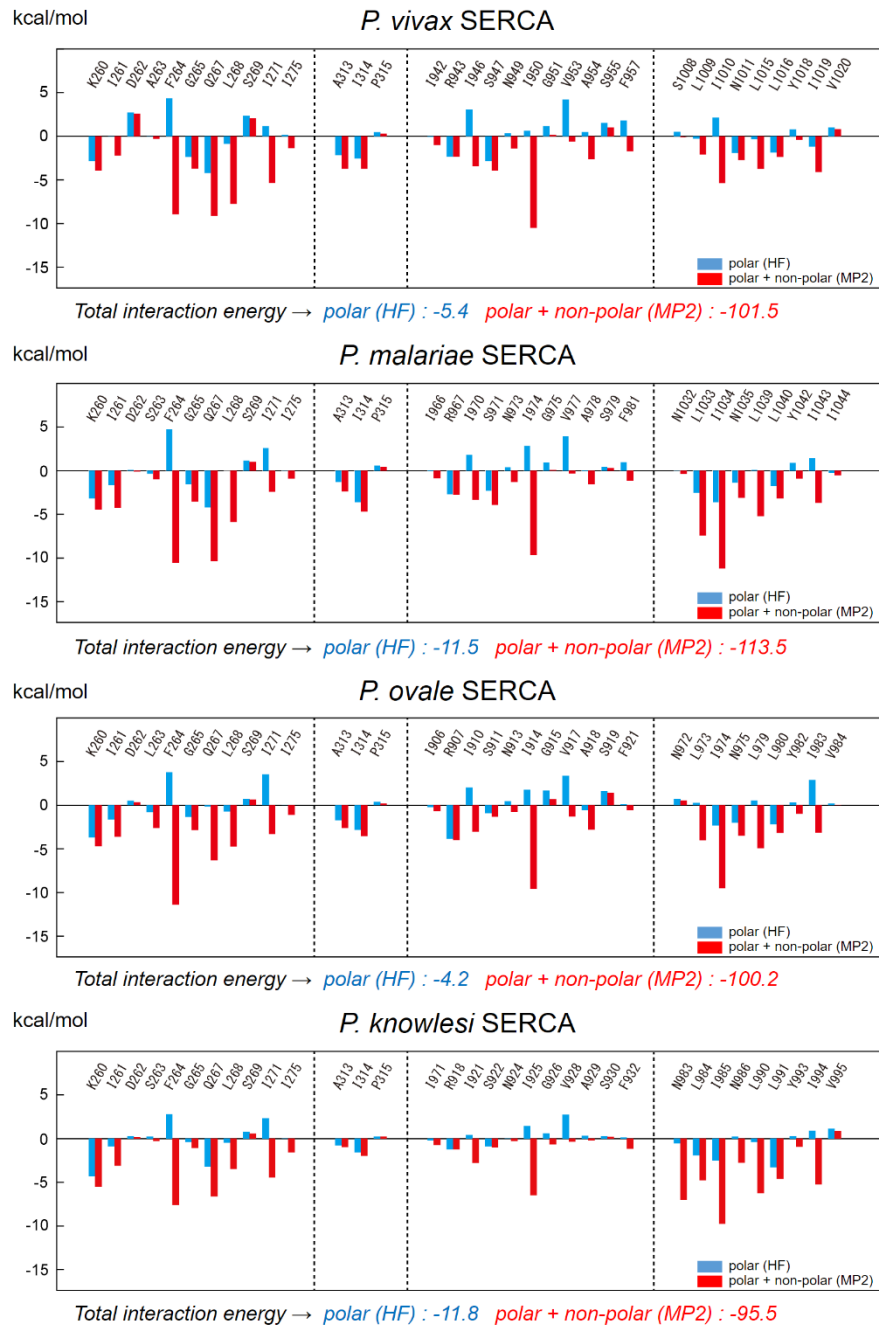


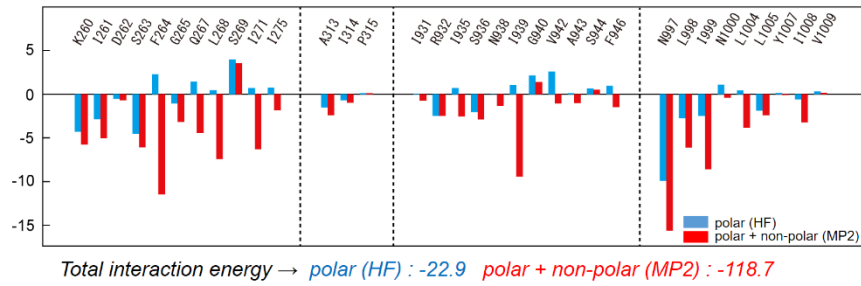
Figure S3. Interaction energies for eight species of malaria parasite

Interaction energies for five other human-infecting malaria parasites (*P. vivax*, *P. malariae*, *P. ovale*, *P. knowlesi*, and *P. cynomolgi*) and three malaria parasites infecting non-human mammals (*P. reichenowi*, *P. berghei*, and *P. yoelii*). Blue bars indicate HF energy, which is mainly polar interaction energy. Red bars indicate MP2 energy, which additionally includes non-polar interaction energy. The total interaction energy is the sum of the interaction energies of the 144 residues.



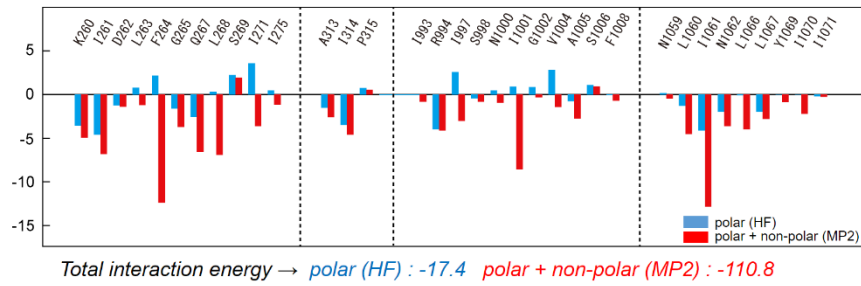
kcal/mol

P. cynomolgi SERCA



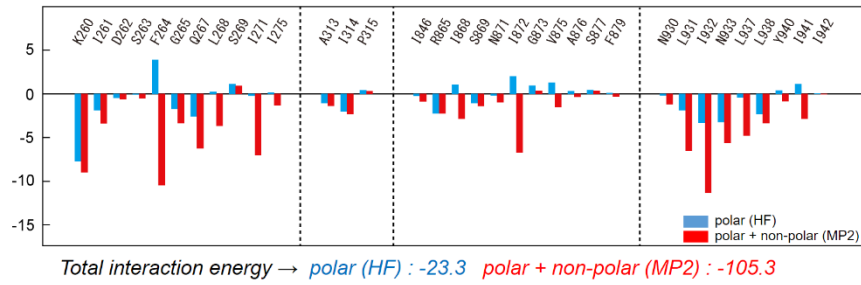
kcal/mol

P. reichenowi SERCA



kcal/mol

P. berghei SERCA



kcal/mol

P. yoelii SERCA

