

***In silico* Structural Modeling and Analysis of Interactions of *Tremellomyces* Cytochrome P450 Monooxygenases CYP51s with Substrates and Azoles**

Olufunmilayo Olukemi Akapo ^{1,†}, Joanna M. Macnar ^{2,3,†}, Justyna D. Kryś ³, Puleng Rosinah Syed ⁴, Khajamohiddin Syed ^{1,†}, Dominik Gront ^{3,*}

¹ Department of Biochemistry and Microbiology, Faculty of Science and Agriculture, University of Zululand, KwaDlangezwa 3886, South Africa; akapoolufunmilayo@gmail.com (O.O.A); khajamohiddinsyed@gmail.com (K.S.)

² College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences, University of Warsaw, Stefana Banacha 2C, 02-097 Warsaw, Poland; joanna.macnar@student.uw.edu.pl (J.M.M.)

³ Faculty of Chemistry, Biological and Chemical Research Center, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland; juchxd@gmail.com (J.D.K); dgront@gmail.com (D.G);

⁴ Department of Pharmaceutical Chemistry, College of Health Sciences, University of KwaZulu-Natal, Durban 4000, South Africa; prosinah@gmail.com (P.R.S.)

† Authors contributed equally to the work

* Correspondence: khajamohiddinsyed@gmail.com (K.S.) and dgront@gmail.com (D.G.)

Supplementary Dataset 1: Structural alignment of *Tremellomyces* CYP51s with template 4LXJ. P450 characteristic notation for α -helices and β strands and substrate recognition sequences (SRS) were mapped based on the template. Residues interacting with ≥ 5 ligands highlighted in yellow. Residues playing a role in azole drug resistance are indicated with “X”.

amphipathic helix transmembrane helix (TM)

41xj 1 10 20 30 40

41xj NIGLSHFLALPLAQRISLII IIPFIYINVWQLLYSLRK
629036 QSILLGFLPTLFSSFGTIGSVLLLLSLPVLVAGNVFRQKVLPRDP
345805 MSSSRILSLPLEQLSWASIA AILVLGLPALLVVGNIIIRQKVLPKDP
809380 TEAQREFARNFREGKIQPIDPT IPLYILSSIAILLFTTVVGSNLMQQLAPRDP
627561 AS . VP . WAPQWLLSLPSLLVLLGIIGFPTLIIFLNVFRQLVLPDK
38441 MSI . VQ . GATEAALALPLAQKLAAVFIGLPVLAIVINVLSQLLLPKDP
1742 MSALPQVQQLLGQL . AQ . YIPPWFTALPTSLKIVIAVIGIPAFIIGLNVFQQLCLPRTK
KIR77383 MSALPQVQQLLGQL . AQ . FIPPWFAALPTSVKVVIAVIGIPALVICLNVFQQLCLPRRK
00040 QVQQLLGQL . AQ . FFPWFFAALPTSLKVAIAVVGIPALIIGLNVFQQLCLPRRK
EAL23379 MAMD . PS . AVLDDWAHALPLPVKIVGAVIGVPLLIVLNLVRLIVPTDP
271374 MAMD . PS . AVLDDWAHALPLPVKIVGAVIGVPLLIVLNLVRLIVPTDP
367271 MAMD . PS . AVLDDWAHALPLPVKIVGAVIGVPLLIVLNLVRLIVPTDP
523016 MAMD . PS . AVLDDWAHALPLPVKIVGAVIGVPLLIVLNLVRLIVPTDP

TM αA β1_1 β1_2 αB

41xj 50 60 70 80 90 100

41xj DRPPLVFYWPVWGSAAVYGMKPYEFFECCQKKYGDIFSFVLLGRVMTVYLGPKGHEFVF
629036 TLPPPLVFHYIPWFGSAATYGMDDPYKFLFDCREQYGDIFTFILLGRKMTVALGPKGNLIL
345805 SLPPVVFHLLIPWFGSAATYGMDDPYKFLFDCREQYGDIFTFVLLGRKMTVALGPKGNLIL
809380 DPYKFMFGCRAKYGDLFTFILMGRMTVALGPKGNLIL
627561 SHPPVVLHYFPWFGSAYSIGTDPYKFLFKCRAKYGDVFTFIMFGKMTVALGPKGNLIL
38441 NLPPVVFHYIPWFGSAAAYGEDPYKFFYFECREKYGDVFTFVLLGRNITVALGPKGNLIL
1742 TLPPVVFHYIPWFGSAAAYGMDDPYKFLFDCREKYGDVFTFILLGRMTVALGPKGNLIL
KIR77383 DLPPVVFHYIPWFGSAAAYGEDPYKFLFECRDKYGDIFTFILMGRRTVALGPKGNLIL
00040 DLPPVVFHYIPWFGSAAAYGEDPYKFLFECRDKYGDIFTFILMGRRTVALGPKGNLIL
EAL23379 DLPPVVFHYIPWFGSAAAYGMDDPYKFLFECRDKYGDIFTFILMGRRTVALGPKGNLIL
271374 TAPPVVFHYIPWFGSAAAYGMDDPYKFMFECRDKYGDVFTFILLGRMTVALGPKGNLIL
367271 TAPPVVFHYIPWFGSAAAYGMDDPYKFMFECRDKYGDVFTFILLGRMTVALGPKGNLIL
523016 MTVALGPKGNLIL

β1_5 αB_1 TT αC

41xj 110 120 130 140

41xj NAKLADVSAEEAYTHLTTPVFGKGVYIDCPNSRLMEQKKF
629036 GGKLSQVSAEDAYTHLTTPVFGKGVYIDCPNEMLMQKKRM
345805 GGRLNQVSAEEAYTHLTTPVFGKGVYIDCPNDLLMQKKRM
809380 GGKLSQVSAEEAYTHLTTPVFGKGVYIDCPNDMLMQKKRM
627561 GGKLSQVSAEDAYTHLTTPVFGKGVYIDCPNEMLMQKKF
38441 GGKTSQVSAEEAYTHLTTPVFGKGVYIDCPNEMLMQKKF
1742 GGKITHVSAEEAYTHLTTPVFGKGVYIDCPNDMLMQKKFSRSRVCCFCRSASGSPMPRS
KIR77383 GGKISQVSAEEAYTHLTTPVFGKGVYIDCPNEMLMQKKF
00040 GGKISQVSAEEAYTHLTTPVFGKGVYIDCPNEMLMQKKF
EAL23379 GGKISQVSAEEAYTHLTTPVFGKGVYIDCPNEMLMQKKF
271374 GGKLSQVSAEDAYTHLTTPVFGKGVYIDCPNDMLMQKKF
367271 GGKLSQVSAEDAYTHLTTPVFGKGVYIDCPNDMLMQKKF
523016 GGKLSQVSAEDAYTHLTTPVFGKGVYIDCPNDMLMQKKF

SRS1 X

αD β3_1

41xj 150 160 170 180 190

41xj VKGALTKEAFKSYVPLIAEEVYKYFRDSKNFRLNERTTGTIDVMVT
629036 VKFGLSTENFRVYVPLIKQETLDFLGKDLKPS . NE . . WKGFNALHA
345805 VKFGLSTENLRAVGMITEETLGLFTNELKASPTA . . WQGFDAALNA
809380 VKFGLSTENLKSYPKITRETLDLFLSHGLHTSATK . . WQSFNALDK
627561 IKHGLTPQAFQTYAPLIWQETHQFFEQEKGFN . VYPGPKSYEAIKF
38441 IKHGLTALASYPPLINAECEHYFEKELGFSPTSPGPKTIEVFKI
1742 IKHGLTALSSYAEALMRGETRQYFRDHVVT GKPFVLELM
KIR77383 IKHGLTALSSYAEALMRGETRQYFRDHVVT GKPFVLELM
00040 IKHGLTALSSYAEALMRGETRQYFRDHVVT GKPFVLELM
EAL23379 IKHGLTALSSYAEALMRGETRQYFRDHVVT GKPFVLELM
271374 IKHGLTALSSYAEALMRGETRQYFRDHVVT GKPFVLELM
367271 IKHGLTALSSYAEALMRGETRQYFRDHVVT GKPFVLELM
523016 IKHGLTALSSYAEALMRGETRQYFRDHVVT GKPFVLELM

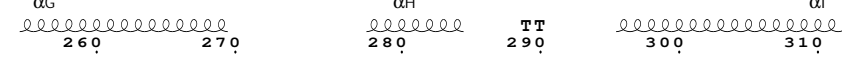
αE αF' αF α310F''

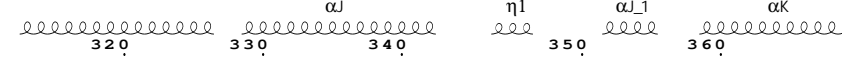
41xj 200 210 220 230 240 250


41xj QPMTIFTASRSLLGKEMRAKLDTFAYLYSDLDKGFPTINFVFPNLPLEHYKRDRDHAQK
629036 MAELTILTASATLQGGKEVRGKLDKTFKRYEALDGGFTPTINFVFPNLPPLSYKRRDKAQK
345805 MSELTILTASRTLQGGKEVRAGLDKTFKRYEALDGGFTPTINFVFPNLPPLSYKRRDKAQK
809380 LAELTILTASDCLQGGKEVRAGLDKTFKRYEALDGGFTPTMNFVFPNLPPLPAYWNRDKAQK
627561 FSELTILTASRTLQGGKEVRAGLTSSAAKHFEHLDKGGFTPTINFVFPNLPPLSYKRRDKAQK
38441 MSQILILTASRTLQGGKEVRSLDSKFAEYFHDLDGGFTPTINFVFPNLPPLSYKRRDKAQK
1742 MQQLILTASRTLQGGKEVRSLDSKFAEYFHDLDGGFTPTINFVFPNLPPLSYKRRDKAQK
KIR77383 MSELTILTASRTLQGGKEVRSLNGQFAKYEDLDGGFTPTINFVFPNLPPLSYKRRDKAQK
00040 MSELTILTASRTLQGGKEVRSLNGQFAKYEDLDGGFTPTINFVFPNLPPLSYKRRDKAQK
EAL23379 MSELTILTASRTLQGGKEVRSLNGQFAKYEDLDGGFTPTINFVFPNLPPLSYKRRDKAQK
271374 MAELIVLTASRTLQGGKEVRNMSIRFAKLELDKGGFTPTINFVFPNLPPLSYKRRDLAQK
367271 MAELIVLTASRTLQGGKEVRNMSIRFAKLELDKGGFTPTINFVFPNLPPLSYKRRDLAQK
523016 MSELTILTASRTLQGGKEVRSLNARFASLYEDLDGGFTPTINFVFPNLPPLSYKRRDLAQK

SRS2


SRS3

α G α H α I
 41xj 
 41xj AISGTYMSLIKERRKNNDIQ...DRDLIDSLMKNSTYKDGVKMTDQEIANLLIGVLMGGQH
 629036 EMSEFYQSIIRGRREGGASEH...DHDMIAA.LMASTYRDGTPLSDSDIAHMMIALLMAGQH
 345805 AMSEFYQGIIRKRREGTHDH...EYDMISA.LQSSKYKDGTPPLSDRDIAHMMIALLMAGQH
 809380 EMSEFYQEIIRKRREGTHDH...EHDMLEA.LMGSSYKNGTPLTSDIAHMMIALLMAGQH
 627561 AMSDFYLEIMRKRREGESDV...DEHDMIAA.LTGNEYKDGTPPLTDRDVAHMMIALLMAGQH
 38441 SMSDFYLNIMQKRREGATDVNGSFDMLAA.LQGCYVRNGVPLTDRDIAHMMIALLMAGQH
 1742 EMSDFYMSIIEKRRSGEHDH...ENDMIAA.LQGSVYKNGVPLSDRDISHIMIALLMAGSH
 KIR77383 AMSDFYLKIMENRRKGESDH...EHDMIEN.LQGCKYRNGVPLSDRDVAHIMIALLMAGQH
 00040 AMSDFYLKIMENRRKGESDH...EHDMIEN.LQSCYRNGVPLSDRDIAHIMIALLMAGQH
 EAL23379 AMSDFYLKIMENRRKGESDH...EHDMIEN.LQSCYRNGVPLSDRDIAHIMIALLMAGQH
 271374 EMSDFYMSIMAKRRTGEHDH...EPDMIQA.LQGSVYRNGTPLSDRDIAHMMIALTMAGQH
 367271 EMSDFYMSIMAKRRTGEHDH...EPDMIQA.LQGSVYRNGTPLSDRDIAHMMIALTMAGQH
 523016 EMSDFYLDIIRKRREGETEP...EHDMISV.LSGCVYRDGTPLTDRDVSHMMIALLMAGQH


α J η 1 α 1 α K
 41xj 
 41xj TSAATSAWILLHLAERP DVQQELYEEQMRVL...DGGKKELTYDLLQEMPLLNQTIKETL
 629036 TSSATSSWTLHLADRTDVYAEELYEEQKEKFGNPDGTFRDMTYEDLKDLPKLDVSVIRETL
 345805 TSSATSSWTLHLAERTDIWEELYREQKNKFGNPDGTFRDLTYEELKELPVLHDVIRETL
 809380 TSSATSSWTLHLAQRHDIYEALYQE QKEKFGNADGTFRDLTYEELKELPLDSCIRETL
 627561 TSSATSSWTLHLAQKPLIFKALFEEHRLDFONEDGTWEDVTYESTKEMPLMAVIRETL
 38441 TSSATSSWALLHLAHRPDVADALYEEQKTLFGNPDGTFRPFVEYGEDEKRM TLMNSIIRETL
 1742 TSSATSSWFLHLAYDQELQQALYDEQVQLFGNADGTFREMTLED TRELP LMTACIRETL
 KIR77383 TSSATSSWTLHLADRPDIYEALYQE QKEKLGNDGTFRDYKYEDLKELPIMDSIIRETL
 00040 TSSATSSWTLHLADRPDVVEALYQE QKQLGNPDGTFRDYKYEDLKELPIMDSIIRETL
 EAL23379 TSSATSSWTLHLADRPDVVEALYQE QKQLGNPDGTFRDYRYEDLKELPIMDSIIRETL
 271374 TSSATSSWFLHLAHDQDIQKRLYEEQVEYFGNPDGTFRPMTYEDTKKLP LMDSCIRETL
 367271 TSSATSSWFLHLAHDQDIQKRLYEEQVEYFGNPDGTFRPMTYEDTKKLP LMDSCIRETL
 523016 TSSATSSWILLHLAHRPDVVEALYEEQKLFNGDGTFRREVTYESTKQMTLM EAIIRETL

β 1_4 β 2_1 β 2_2 β 1_3 α K_1
 41xj 
 41xj RMHHPLHSLFRKVMKDMHVPN...TSYVIPAGYHVLVSPGYTHLRDEYFPNAH
 629036 RMHAP IHSIMRKVMTDMPVFNALAAAPTANEKSSYIVPKGHFVLASPGVAQMDPLIWSDAS
 345805 RLHAP IHSIMRKVISDIPVFNLTSSPKQHENATYVIPKGHFVLASPGVAQMDPLIWQNAE
 809380 RLHAPLH SIMRKVISDIPVFNLSPPSQTNDA SYVIPKGHFVLACPGVQMDPLIWQDST
 627561 RMHAP IHSIYRKVLQDLEVPALAAAGGN.GSRPYVVPKGHFVVAAPGVSAMD PKVWPDAD
 38441 RIHAP IHSIYRKVISPMPPVPA SLSAPS...ESATYIIPKNYIIVAAPGVSAMDPR IWKDAN
 1742 RLHAP IHSIYRKVMDIAVPPSLAAPS...KDGTYVIPKGHFIVAAPGVSQMDPR IWKDAK
 KIR77383 RMHAP IHSIYRKVLSDIPVPPSLAAPS...ENGQYIIPKGHYIMAAPGVSQMDPR I WQDAK
 00040 RMHAP IHSIYRKVLSDIPVPPSLAAPS...ENGQYIIPKGHYIMAAPGVSQMDPR I WQDAK
 EAL23379 RMHAP IHSIYRKVLSDIPVPPSLAAPS...ENGQYIIPKGHYIMAAPGVSQMDPR I WQDAK
 271374 RLHAP IHSIYRKVLQPIVPPQSLAAPS...EDKAYVIPKGNFIVAAPGVSQMDPKI WDDAP
 367271 RLHAP IHSIYRKVLQPIVPPQSLAAPS...EDKAYVIPKGNFIVAAPGVSQMDPKI WDDAP
 523016 RLHAP IHSIYRKVLSDIPVPA SLSAPS...ESSAYVIPEGHFIVAAPGVSQMDPL I WPDAX

×

meander_loop η 3
 41xj 
 41xj QFNIRHRWNDSAS...SYS.VGEEVDYGFGAISKGVS SPYLPFGGGRHRCIGEHFA
 629036 IWNPHRWTDEKGMASQALEEYT...SGDKVDYGYGVS VKGTESPYQPF GAGRHR CIGETFA
 345805 EWNPHRWDD ESGVAAMAAEQYT...EGDKVDYGYGVS VKGTESPYQPF GAGRHR CIGESFA
 809380 TWNPYRWSDERGVAQA VEEYQGEKGEKVDYGF GSVSKGTESPYQPF GAGRHR CIGESFA
 627561 KWDPPFRWLEDKGM AQEALDSYSGATSEQIDYGF GQVSKGTESPYMPF GAGRHR CVGEQFA
 38441 VWEPFRWSDEKGTAAAMALNEYTT.GGEKMDYGF GQVSKGTESPYQPF GAGRHR CVGEAFA
 1742 TWRPRRWLEEGGVANAANE EYTS...GERVDYGF GAVSKGTESPYQPF GAGRHR CVGEQFA
 KIR77383 VWNPARWHDEKGF AAAAMAQYTK...AEQVDYGF GSVSKGTESPYQPF GAGRHR CVGEQFA
 00040 VWNPARWHDEKGF AAAAMAQYSK...AEQVDYGF GSVSKGTESPYQPF GAGRHR CVGEQFA
 EAL23379 VWNPARWHDEKGF AAAAMVQYTK...AEQVDYGF GSVSKGTESPYQPF GAGRHR CVGEQFA
 271374 RWNPLRWLVDGGIAKTANEQYQA...GEKVDYGF GAVSKGTESPYQPF GAGRHR CIGEQFA
 367271 RWNPLRWLVDGGIAKTANEQYQA...GEKVDYGF GAVSKGTESPYQPF GAGRHR CIGEQFA
 523016 TWNPYRWLDEKGVAA SAKSEQYSGATSEQVDYGF GQVSKGTESPYQPF GAGRHR.....

×

α L β 3_3 β 3_2
 41xj 
 41xj YCQLGVLMSIFIRTLKWHYPEG.KTVPPDPFTSMVLTLP TGPAKI IWEKRNPEQKIGGRHH
 629036 YLQLQTILATFIRNVEMKLD...GPFK.TNYQTMVVLPLPGTKIL...YRNRK.....
 345805 YVQLQTILATLCRRVVLKLE...GPPAPNYQTMIVLPLKGQTKI.MYKLRA.....
 809380 YVQLQTIIATVVRNLT LKFE...GDFPSPNYQTMVVLPLKGGEIL...FKSRQ.....
 627561 YLQLSVIMSYIVRN YDIQLL...GAFPPQNTYNTMIVLPLNGYVTF.KKRQSKWI.....
 38441 YVQLSVIIAYIIRNYS LRLTEPDGLFPKNTYRTMIVLPLNGLMSL.TKRRQE.....
 1742 YLQLTMLVSEIIRTFKVEPGAP...QFPETNYQTMIVLPLHGMIKL.EPRK.....
 KIR77383 YTQLSTIFTYVVRNFT LKLAVP...KFPETNYRTMIVQPTNPLVTF.TLRNAEVK.....
 00040 YTQLSTIFTYVVRNFT LKLAVP...KFPETNYRTMIVQPTNPLVTF.TLRNAEVK.....
 EAL23379 YTQLSTIFTYVVRNFT LKLAVP...KFPETNYRTMIVQPTNPLVTF.TLRNAEVKQEV...
 271374 YLQLTLVVGEVVRN YKLTAAQR...EFPKNTYQTMIVLPLDPKITF.EARV.....
 367271 YLQLTLVVGEVVRN YKLTAAQR...EFPKNTYQTMIVLPLDPKITF.EARV.....
 523016 YLQLTLVVGEVVRN YKLTAAQR...EFPKNTYQTMIVLPLDPKITF.EARV.....