Electronic Supplementary Material (ESI) for Catalysis Science & Technology.

Production of Novel Rieske Dioxygenase Metabolites Enabled by Enzyme Engineering

Elizabeth A. Osifalujo,[a] Bailey N. Rutkowski,[a] Louis R. Satterwhite,[a] Phillip C. Betts,[a] Angel K. Nkosi,[a] and Jordan T. Froese*[a]

a) Department of Chemistry, Ball State University, 1600 W Ashland Ave, Muncie, IN, 47303.

Table of Contents

General Experimental

E. coli BL21 (DE3) competent cells were obtained from ThermoFisher. Plasmid isolation/purification was performed using New England Biolabs Monarch® miniprep kit. Transformations of electrocompetent cells were performed on an Eppendorf Eporator®. Whole-cell assay cultures were grown in Greiner Bio-One polystyrene clear, round-bottom 96-well plates. All cultures were incubated in a Barnstead MaxQ 4000 Digital Orbital Incubator Shaker equipped with an Enzyscreen universal clamp system unless otherwise stated. Fluorescence analyses were performed using a Biotek® Synergy™ H1 monochromator-based multi-mode plate reader, using Corning® polystyrene black, opaque, flat-bottom 96-well plates. All reagents were obtained from MilliporeSigma unless otherwise stated. Media were made at pH 7.2 and streptomycin was added at 50 μg mL^{-1}. All *E. coli* cultures were maintained at 37 °C unless otherwise stated. NMR analyses were performed using a Jeol ECZ 400S (400 MHz) instrument. IR spectra were collected using a PerkinElmer Frontier FT-IR spectrometer. High-resolution mass spectrometry data were collected by the University of Illinois Urbana-Champaign Mass Spectrometry lab. Homology modelling was performed using Alphafold2.[1] Docking analyses were performed using AutoDock Vina.[2] Molecular dynamics simulations were performed using GROMACS.[3] Mapping of active site cavities was performed using PyMOL.[4]

Vector Construction[5]

A 2.1-kbp DNA fragment including todC2BA was PCR amplified from pDTG601A[6] using primers TODC2BAF (GTATAAGAAGGAGATATACAATGATTGATTCAGCCAAC) and TODC2BAR (TATCCAATTGAGATCTGCCATCACGTTAGGTCTCCTTC) and cloned into the NdeI site of pCDF-Duet-1 using NEBuilder® HiFi DNA Assembly (New England Biolabs) to yield plasmid pCP-01. todC1 (1.4-kbp) was then amplified from pDTG601A[6] using primers TODC1F (ACTTTAATAAGGAGATATACATGAATCAGACCGACACATCAC) and TODC1R (TGATGGTGATGGCTGCTGCCTCAGCGTGTCGCCTTCAG) and cloned into the NcoI site of pCP-01 using NEBuilder® HiFi DNA Assembly (New England Biolabs) to yield plasmid pCP-02. Full vector sequences and vector maps are provided below (**Figures S1 and S2**).

Mutant library generation[7]

The pCP-02 expression system was used as the template for toluene dioxygenase mutant library generation.[5] Saturation mutagenesis was performed following the procedure of Liu and Naismith.[7] Amplification was performed using an ABI GeneAmp® 9700 Thermal Cycler. Mutagenic primers were designed according to the procedure of Liu and Naismith[7] (toluene dioxygenase M220 forward primer – TGCAGCGACNNKTACCATGCCGGGACGACCTCGCATCTGTCTGGC; M220 reverse primer – GGCATGGTAMNNGTCGCTGCAAAACTGCTCTGCGGCGAATTTCC; A223 forward primer – CATGTACCATNNKGGGACGACCTCGCATCTGTCTGGCATCCTGGC; A223 reverse primer – GGTCGTCCCMNNATGGTACATGTCGCTGCAAAACTGCTCTGCGG; L272 forward primer – GACCCCAATNNKATGCTTGCCATCATGGGGCCAAAGGTCACCAGC; L272 reverse primer – GGCAAGCATMNNATTGGGGTCGCCGACATAGAAGCCACTTCCATG;

I276 forward primer – GATGCTTGCCNNKATGGGGCCAAAGGTCACCAGCTACTGGACCG; I276 reverse primer – TGGCCCCATMNNGGCAAGCATCAGATTGGGGTCGCCGACATAGAAG; V309 forward primer – GAAACTCATGNNKGAGCACATGACCGTCTTCCCCACGTGTTCCTTC; V309 reverse primer – CATGTGCTCMNNCATGAGTTTCGAGCCGCGCTCCACGCTACCCAG; L321 forward primer – GTGTTCCTTCNNKCCAGGTATCAATACGGTCCGGACATGGCATCC; L321 reverse primer – GATACCTGGMNNGAAGGAACACGTGGGGAAGACGGTCATGTGCTC; I324 forward primer – CTCCCAGGTNNKAATACGGTCCGGACATGGCATCCGCGCGGGCCG; I324 reverse primer – GACCGTATTMNNACCTGGGAGGAAGGAACACGTGGGGAAGACGG; F366 forward primer – CTGCGCACCNNKTCTGCCGGTGGCGTGTTCGAGCAGGACGACGGG; F366 reverse primer - ACCGGCAGAMNNGGTGCGCAGCGTCTGGCGCCGGAACTCTTCCTTG). Sequencing analyses were performed by Eurofins Genomics© (Louisville, KY).

Generation of combined (double and triple) mutants[7]

Plasmids bearing active site mutations generated as part of the study were used as the templates for the introduction of additional mutations. The introduction of point mutations was performed following the procedure of Liu and Naismith.[7] Amplification was performed using an ABI GeneAmp® 9700 Thermal Cycler. Mutagenic primers were designed according to the procedure of Liu and Naismith[7] (L272W/I276V forward primer – GATGCTTGCCGTGATGGGGCCAAAGGTCACCAGCTACTGGACCG; L272W/I276V reverse primer – TGGCCCCATCACGGCAAGCATCCAATTGGGGTCGCCGACATAGAAG; V309G forward primer – GAAACTCATGGGCGAGCACATGACCGTCTTCCCCACGTGTTCCTTC; V309G reverse primer – CATGTGCTCGCCCATGAGTTTCGAGCCGCGCTCCACGCTACCCAG). Sequencing analyses were performed by Eurofins Genomics© (Louisville, KY).

Whole-cell fermentation 96 well-plate assay protocol[5,8,9]

E. coli (BL21 (DE3)) electrocompetent cells were transformed with isolated pCP-02 plasmids expressing toluene dioxygenase (parent and/or mutant libraries), and with isolated pCP-01 plasmids as negative controls.[5] The transformation cultures were selected on LB + streptomycin plates overnight. Single colonies were inoculated into 160 µL LB + streptomycin media with 0.3% glucose in a 96-well round bottom seed plate and incubated with shaking overnight. All plates included 3 or more wells containing *E. coli* (BL21 (DE3)) pCP-02 cells expressing the parent toluene dioxygenase enzyme, and 3 or more wells containing *E. coli* (BL21 (DE3)) pCP-01 (negative control).[5] Seed plates were used to inoculate 5 µL into 155 µL LB media containing streptomycin in a fresh 96-well round bottom assay plate, and the cultures were incubated with shaking for 2.75 h. The assay plates were then pelleted, and the supernatant discarded. Cultures were resuspended in 150 µL minimal media (KH_2PO_4 – 7.5 g L^{-1}; citric acid – 2 g L^{-1}; $MgSO_4 \cdot 7H_2O$ – 5 g L^{-1}; trace metal solution – 2 mL L^{-1} [Na_2SO_4 – 1 g L^{-1}; $MnSO_4$ – 2 g L^{-1}; $ZnCl_2$ – 2 g L^{-1}; $CoCl_2 \cdot 6H_2O$ – 2 g L^{-1}; $CuSO_4 \cdot 5H_2O$ – 0.3 g L^{-1}; $FeSO_4 \cdot 7H_2O$ – 10 g L^{-1}; pH 1.0]; conc. H_2SO_4 – 1.2 mL L^{-1}; ferric ammonium citrate – 0.3 g L^{-1}; glucose – 4 g L^{-1}; thiamine – 0.034 g L^{-1}; pH 7.2)[8] containing streptomycin and incubated for a 1 h recovery period.

Following this, the cultures were induced to a final concentration of 0.5 mM IPTG and the incubation temperature was reduced to 30 °C. After a 2 h induction period, aromatic substrates were added as 68 mM stock solutions in DMSO to a final concentration of 2 mM. Cultures were incubated with aromatic substrates for 1.5 h at 30 °C, after which the cultures were pelleted. A 100 µL portion of supernatant from each well was transferred to 96-well black opaque assay plates. The reaction was initiated by adding a 50 µL of $NaIO_4$ stock solution to each well to a final concentration of 10 mM, and the assay plates were incubated with shaking at room temperature for 30 min. Cleaved diols were detected by adding 50 µL of fluoresceinamine stock solution (prepared with 3 µL conc. HCl (11.65 M)/1 mL fluoresceinamine solution) to each well to a final concentration of 0.1 mM.[5] Assay plates were incubated with shaking at room temperature for 5 h. The fluorescence response from each well was analyzed at 485 nm (ex), 520 nm (em), and normalized to the mean fluorescence response of the negative controls ($[I - I_0]/I_0$).

Preparative scale production of [(5S,6R)-5,6-Dihydroxy-1,3-cyclohexadien-1-yl]methylacetamide (1) and [(5S,6R)-5,6-Dihydroxy-1,3-cyclohexadien-1-yl]ethylacetamide (2)[5,8,9]

E. coli (BL21 (DE3)) electrocompetent cells were transformed with isolated plasmid expressing the TDO L272W/I276V variant and selected on LB agar containing streptomycin overnight. Single colonies were inoculated into 5 mL LB medium containing streptomycin and 0.3% glucose and incubated with shaking overnight. Cultures (2 × 500 mL LB with streptomycin in 2000 mL Erlenmeyer flasks) were inoculated with 5 mL overnight culture each and incubated with shaking. Growth of the cultures was monitored via optical density measurement at 600 nm. Upon reaching an OD_{600} of 0.5–0.6 AU, 500 mL cultures were pelleted and resuspended in minimal media (KH_2PO_4 – 7.5 g L^{-1}; citric acid – 2 g L^{-1}; $MgSO_4 \cdot 7H_2O$ – 5 g L^{-1}; trace metal solution – 2 mL L^{-1} [Na_2SO_4 – 1 g L^{-1}; $MnSO_4$ – 2 g L^{-1}; $ZnCl_2$ – 2 g L^{-1}; $CoCl_2 \cdot 6H_2O$ – 2 g L^{-1}; $CuSO_4 \cdot 5H_2O$ – 0.3 g L^{-1}; $FeSO_4 \cdot 7H_2O$ – 10 g L^{-1}; pH 1.0]; conc. H_2SO_4 – 1.2 mL L^{-1}; ferric ammonium citrate – 0.3 g L^{-1}; glucose – 4 g L^{-1}; thiamine – 0.034 g L^{-1}; pH 7.2) containing streptomycin.[8] After 1 h of recovery in minimal media, the cultures were induced to a final concentration of 0.5 mM IPTG and the incubation temperature decreased to 30 °C. After a 2 h induction period, substrates prepared as solutions in DMSO were added directly via pipette to the cultures, in two portions, to a final concentration of 2 mM. The cultures were incubated with the substrates for 3 h, and subsequently pelleted and the supernatant decanted. The combined supernatant was then extracted with 3 × 1 L EtOAc, and the combined extracts were dried over anhydrous $MgSO_4$. The dried extract was concentrated and the compounds purified through column chromatography using deactivated silica as the stationary phase (95 : 5; dichloromethane (DCM) : methanol (MeOH)).

Computational Visualization of the TDO/TDO variant active sites[1,4]

Homology models were generated using Alphafold2[1] for analysis of TDO variant active sites. The active site cavity in each case was mapped using the surface feature of PyMOL (cavity detection radius – 3 solvent radii; cavity detection cutoff – 1 solvent radius).[4]

Enzyme-Substrate Docking Analysis and Molecular Dynamics Simulation[1-3,11]

Homology models generated with Alphafold2[1] were utilized for docking analysis with TDO variants. Enzyme structures were prepared for docking analysis by removing all heteroatoms, repairing all missing atoms, adding polar hydrogens and Kollman charges, and setting grid box parameters to limit binding to the active site (x = 30 Å, y = 30 Å, z = 30 Å) using AutoDock Tools.[9] Ligand structures were similarly prepared using AutoDock Tools.[10] Docking Analysis was performed using AutoDock Vina.[2] The resultant binding predictions were filtered for binding modes that could result in the successful metabolism of the substrate. All reported affinities represent the highest affinity levels among binding modes that could result in the successful metabolism of the substrate. The docking predictions produced were subsequently subjected to molecular dynamics simulations using GROMACS.[3] Topologies were generated for both the ligand and enzyme using the CHARMM36 force field (July 2022).[11]

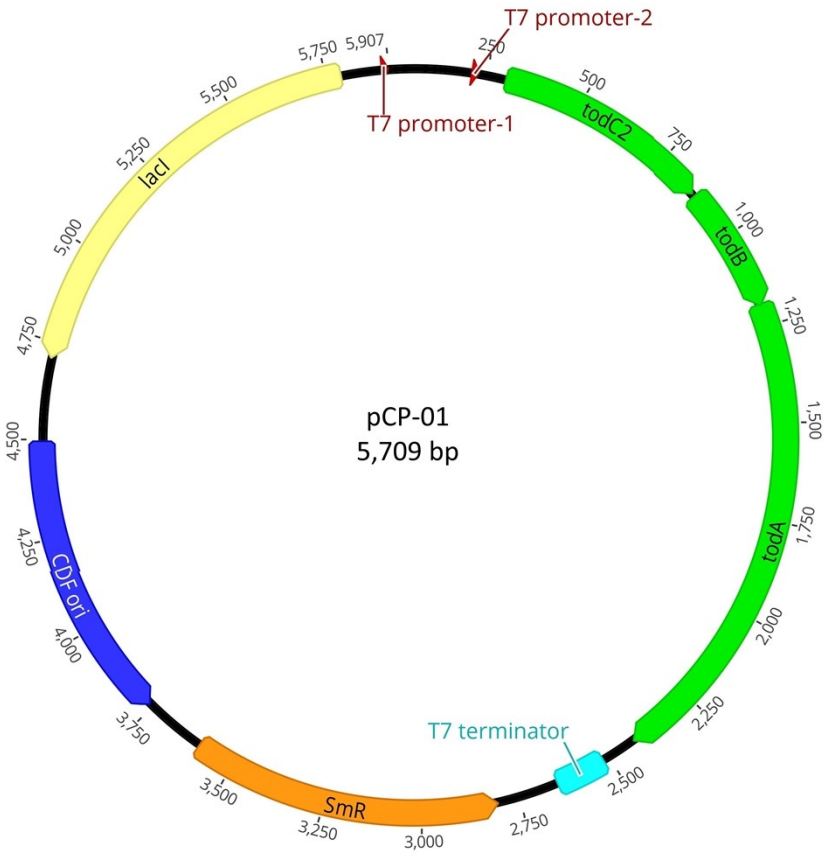


Figure S1: Vector map of pCP-01 harboring TDO genes todC2 (oxygenase – structural), todB (ferredoxin), todA (reductase). Vector map generated with Geneious version 2023.1 (Biomatters).

pCP-01 vector sequence:

ggggaattgtgagcggataacaattcccctgtagaaataattttgtttaactttaataaggagatataccatgggca
gcagccatcaccatcatcaccacagccaggatccgaattcgagctcggcgcgcctgcaggtcgacaagcttgcggcc
gcataatgcttaagtcgaacagaaagtaatcgtattgtacacggccgcataatcgaaattaatacgactcactatag
gggaattgtgagcggataacaattccccatcttagtatattagttaagtataagaaggagatatacaatgattgatt
cagccaacagagccgacgtctttctccgcaagccggcacccgtagcgcccgaactgcagcacgaagtcgagcagttc
tactattgggaggccaagcttctcaacgatcgccgcttcgaggagtggttcgcgctgctcgcggaagacattcacta
cttcatgcccattcgcaccacgcggatcatgcgggactcgcgccttgaatactcaggctcccgagagtacgcgcact
tcgatgacgacgccacgatgatgaagggacgcttgcgcaagatcacgtccgacgtgagctggtccgagaaccccgca
tcgcggacccggcatctcgtgagcaacgtgatgatcgtcggcgcagaggcagaaggggagtacgaaatctcaagcgc
cttcattgtgtaccgcaatcgtctggagcggcagctcgacatctttgccggtgagcgtcgcgatacgttgcgccgta
acacgagcgaggccgggttcgagatcgtcaatcggaccatcctgatcgaccagagcaccatcctggccaataacctc
agtttcttcttctaggtgatgtcatgacttggacatacatattgcggcagggtgacctgccacccggtgagatgcag
cgctacgaaggcggcccggaacctgtgatggtctgcaacgtcgatggcgagttcttcgcggtgcaggatacctgcac
gcatggggactgggcgttgtcggatggttacctggacggtgatattgtcgaatgcacgttgcatttcggcaagttct
gcgtgcggaccgggaaggtgaaggcgctgcctgcttgcaaacctatcaaggtattcccaatcaaggtcgaaggcgat
gaagtgcacgtcgatctcgacaacggggagttgaagtgatggctacccatgtggcgatcatcggcaatggcgtgggt
ggcttcacgaccgcgcaggccctacgtgccgagggcttcgaggggagaatctcgctgattggggacgaaccgcatct
cccctatgaccgaccatccttgtccaaggcggttctcgacggcagccttgagcggccgcccatactggccgaggccg
attggtacggcgaggcccgcatcgacatgctgaccggcccggaagtcactgcccttgatgtgcagacaaggacgatc
agtctggatgatggcaccacgctctctgcggacgccatcgtcatcgcgacgggcagtcgagcgcggacgatggcgtt
gcccggcagccaactgcccggcgtcgtaacgctgcgcacctacggtgacgtgcaggtattgcgcgatagttggactt
ccgcgacgcggctgctgattgtgggtggcggattgatcggctgcgaggtcgcgacgacggcgcgcaagctcggcctg
tcggtcacgatcctggaggcaggtgatgaactgctggtccgagtacttgggcggcgtatcggtgcctggctgcgcgg
cctgctgacagaacttggtgtgcaggtcgagttgggaacgggtgtcgtaggtttttctggtgagggccagctcgaac
aagtcatggccagcgatgggcgcagcttcgtagccgatagcgcactcatttgcgtcggcgcggagcccgcggatcaa
cttgcgcgtcaagcgggcttggcatgtgaccgcggcgtcattgtcgatcactgcggtgcgacgcttgccaaaggcgt
attcgccgtcggagatgtggccagttggccgctgcgcgccggcggccggcgttcgctcgaaacctatatgaacgcgc
agcgccaagccgccgcggtggctgcggccattctggggaaaaacgtatcggcaccgcaactgcccgtgtcctggacg
gagatcgctgggcatcgcatgcagatggcgggcgatatcgaaggacctggtgatttcgtctcgcgcggcatgcccgg
tagtggcgctgccctgttgttccgcctgcaggagcgaaggattcaggcggtcgtcgcggtcgatgcaccccgtgact
tcgcgcttgcaacccgattggtagaagcccgcgcggcaatcgagccagcacggctggcagatctttcaaacagtatg
cgcgattttgttcgtgcgaatgaaggagacctaacgtgatggcagatctcaattggatatcggccggccacgcgatc
gctgacgtcggtaccctcgagtctggtaaagaaaccgctgctgcgaaatttgaacgccagcacatggactcgtctac
tagcgcagcttaattaacctaggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacggg
tcttgaggggttttttgctgaaacctcaggcatttgagaagcacacggtcacactgcttccggtagtcaataaaccg
gtaaaccagcaatagacataagcggctatttaacgaccctgccctgaaccgacgaccgggtcatcgtggccggatct
tgcggcccctcggcttgaacgaattgttagacattatttgccgactaccttggtgatctcgcctttcacgtagtgga
caaattcttccaactgatctgcgcgcgaggccaagcgatcttcttcttgtccaagataagcctgtctagcttcaagt
atgacgggctgatactgggccggcaggcgctccattgcccagtcggcagcgacatccttcggcgcgattttgccggt
tactgcgctgtaccaaatgcgggacaacgtaagcactacatttcgctcatcgccagcccagtcgggcggcgagttcc
atagcgttaaggtttcatttagcgcctcaaatagatcctgttcaggaaccggatcaaagagttcctccgccgctgga
cctaccaaggcaacgctatgttctcttgcttttgtcagcaagatagccagatcaatgtcgatcgtggctggctcgaa
gatacctgcaagaatgtcattgcgctgccattctccaaattgcagttcgcgcttagctggataacgccacggaatga
tgtcgtcgtgcacaacaatggtgacttctacagcgcggagaatctcgctctctccaggggaagccgaagtttccaaa
aggtcgttgatcaaagctcgccgcgttgtttcatcaagccttacggtcaccgtaaccagcaaatcaatatcactgtg
tggcttcaggccgccatccactgcggagccgtacaaatgtacggccagcaacgtcggttcgagatggcgctcgatga
cgccaactacctctgatagttgagtcgatacttcggcgatcaccgcttccctcatactcttccttttttcaatattat
tgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagctag
ctcactcggtcgctacgctccgggcgtgagactgcggcgggcgctgcggacacatacaaagttacccacagattccg
tggataagcaggggactaacatgtgaggcaaaacagcagggccgcgccggtggcgtttttccataggctccgccctc
ctgccagagttcacataaacagacgcttttccggtgcatctgtgggagccgtgaggctcaaccatgaatctgacagt
acgggcgaaacccgacaggacttaaagatccccaccgtttccggcgggtcgctccctcttgcgctctcctgttccga
ccctgccgtttaccggatacctgttccgcctttctcccttacgggaagtgtggcgctttctcatagctcacacactg
gtatctcggctcggtgtaggtcgttcgctccaagctgggctgtaagcaagaactccccgttcagcccgactgctgcg

ccttatccggtaactgttcacttgagtccaacccggaaaagcacggtaaaacgccactggcagcagccattggtaac
tgggagttcgcagaggatttgtttagctaaacacgcggttgctcttgaagtgtgcgccaaagtccggctacactgga
aggacagatttggttgctgtgctctgcgaaagccagttaccacggttaagcagttccccaactgacttaaccttcga
tcaaaccacctccccaggtggttttttcgtttacagggcaaaagattacgcgcagaaaaaaaggatctcaagaagat
cctttgatcttttctactgaaccgctctagatttcagtgcaatttatctcttcaaatgtagcacctgaagtcagccc
catacgatataagttgtaattctcatgttagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctct
caagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctt
tccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattggg
cgccagggtggtttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagtt
gcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacat
gagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacgcgcagcccggactcggtaatggcgcg
cattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatgg
tttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatat
ttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacc
caatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggt
cagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcgga
tagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcg
ttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcg
cgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttg
ggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggttcac
cacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcacca
ccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatc
tcgacgctctcccttatgcgactcctgcattaggaaattaatacgactcactata

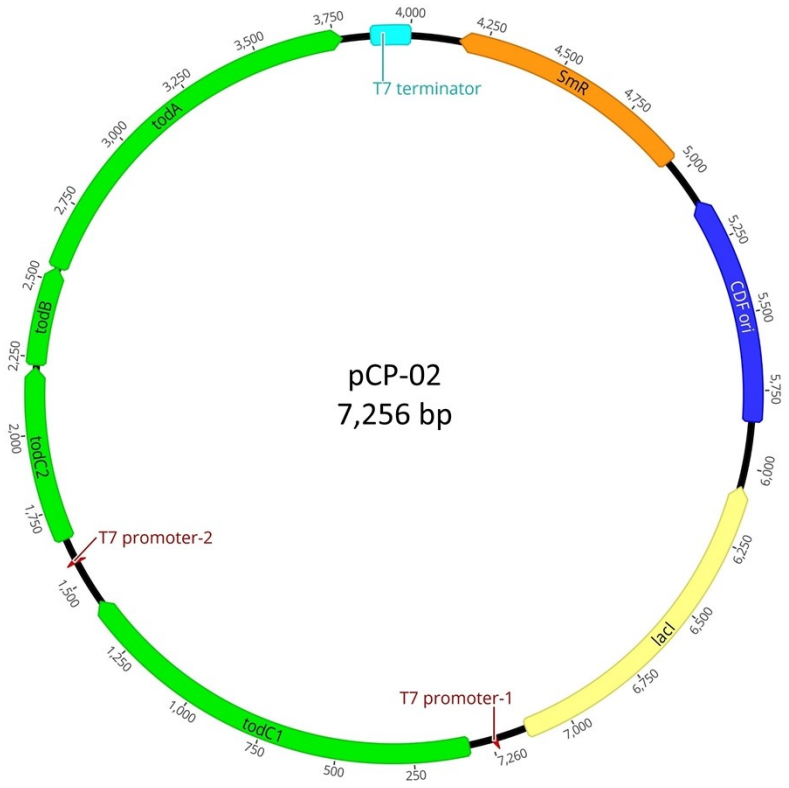


Figure S2: Vector map of pCP-02 harboring TDO genes todC1 (oxygenase – catalytic), todC2 (oxygenase – structural), todB (ferredoxin), todA (reductase). Vector map generated with Geneious version 2023.1 (Biomatters).

pCP-02 vector sequence:

ggggaattgtgagcggataacaattcccctgtagaaataattttgtttaactttaataaggagatatacatgaatca
gaccgacacatcacctatcaggctgcgcaggagctggaacaccagcgagatagaagcgctctttgacgagcatgccg
gacgtatcgatccgcgcatttataccgatgaggatctgtaccaactcgaactggagcgtgtcttcgcccggtcctgg
ctgctgttggggcatgaaacccagattcgcaagccgggcgattacatcacgacctacatgggtgaagaccctgtcgt
ggtcgtccggcagaaagacgccagcattgccgtgttcctgaaccagtgccgccaccgtggcatgcgcatctgccgcg
cggatgccggaaacgcgaaggcgttcacttgcagctaccacgggtgggcttacgacaccgccggcaatcttgtcaat
gtgccttacgaggccgaatccttcgcgtgcctgaacaagaaggaatggagcccgctgaaggcccgggtagaaaccta
caagggcctgattttcgccaactgggatgagaacgctgtagacctcgacacgtatctgggcgaggcgaagttctaca
tggaccacatgctcgaccgcaccgaggccggcaccgaagcgatcccgggcgtgcagaagtgggtcattccctgtaac
tggaaattcgccgcagagcagttttgcagcgacatgtaccatgccgggacgacctcgcatctgtctggcatcctggc
aggcctgccagaagaccttgaaatggccgaccttgctccgccgacagttggcaagcagtaccgtgcgtcatggggcg
gacatggaagtggcttctatgtcggcgaccccaatctgatgcttgccatcatggggccaaaggtcaccagctactgg
accgaaggccccgcgtcggaaaaggcggccgaacgtctgggtagcgtggagcgcggctcgaaactcatggtcgagca
catgaccgtcttccccacgtgttccttcctcccaggtatcaatacggtccggacatggcatccgcgcgggccgaacg
aggtcgaggtatgggcgtttacggtggtcgatgctgatgctcctgacgatatcaaggaagagttccggcgccagacg
ctgcgcaccttctctgccggtggcgtgttcgagcaggacgacggggagaactgggtcgagatccagcacatcctgcg
aggccacaaggcgcggagccgccctttcaatgccgagatgagcatggaccagaccgtcgacaacgacccggtttacc
ccgggcggatcagcaacaacgtctacagcgaggaagctgcccgcgggctctatgcccattggctgcggatgatgaca
tcccccgactgggacgcgctgaaggcgacacgctgaggcagcagccatcaccatcatcaccacagccaggatccgaa
ttcgagctcggcgcgcctgcaggtcgacaagcttgcggccgcataatgcttaagtcgaacagaaagtaatcgtattg
tacacggccgcataatcgaaattaatacgactcactataggggaattgtgagcggataacaattccccatcttagta
tattagttaagtataagaaggagatatacaatgattgattcagccaacagagccgacgtctttctccgcaagccggc
acccgtagcgcccgaactgcagcacgaagtcgagcagttctactattgggaggccaagcttctcaacgatcgccgct
tcgaggagtggttcgcgctgctcgcggaagacattcactacttcatgcccattcgcaccacgcggatcatgcgggac
tcgcgccttgaatactcaggctcccgagagtacgcgcacttcgatgacgacgccacgatgatgaagggacgcttgcg
caagatcacgtccgacgtgagctggtccgagaaccccgcatcgcggacccggcatctcgtgagcaacgtgatgatcg
tcggcgcagaggcagaaggggagtacgaaatctcaagcgccttcattgtgtaccgcaatcgtctggagcggcagctc
gacatctttgccggtgagcgtcgcgatacgttgcgccgtaacacgagcgaggccgggttcgagatcgtcaatcggac
catcctgatcgaccagagcaccatcctggccaataacctcagtttcttcttctaggtgatgtcatgacttggacata
catattgcggcagggtgacctgccacccggtgagatgcagcgctacgaaggcggcccggaacctgtgatggtctgca
acgtcgatggcgagttcttcgcggtgcaggatacctgcacgcatggggactgggcgttgtcggatggttacctggac
ggtgatattgtcgaatgcacgttgcatttcggcaagttctgcgtgcggaccgggaaggtgaaggcgctgcctgcttg
caaacctatcaaggtattcccaatcaaggtcgaaggcgatgaagtgcacgtcgatctcgacaacggggagttgaagt
gatggctacccatgtggcgatcatcggcaatggcgtgggtggcttcacgaccgcgcaggccctacgtgccgagggct
tcgaggggagaatctcgctgattggggacgaaccgcatctcccctatgaccgaccatccttgtccaaggcggttctc
gacggcagccttgagcggccgcccatactggccgaggccgattggtacggcgaggcccgcatcgacatgctgaccgg
cccggaagtcactgcccttgatgtgcagacaaggacgatcagtctggatgatggcaccacgctctctgcggacgcca
tcgtcatcgcgacgggcagtcgagcgcggacgatggcgttgcccggcagccaactgcccggcgtcgtaacgctgcgc
acctacggtgacgtgcaggtattgcgcgatagttggacttccgcgacgcggctgctgattgtgggtggcggattgat
cggctgcgaggtcgcgacgacggcgcgcaagctcggcctgtcggtcacgatcctggaggcaggtgatgaactgctgg
tccgagtacttgggcggcgtatcggtgcctggctgcgcggcctgctgacagaacttggtgtgcaggtcgagttggga
acgggtgtcgtaggtttttctggtgagggccagctcgaacaagtcatggccagcgatgggcgcagcttcgtagccga
tagcgcactcatttgcgtcggcgcggagcccgcggatcaacttgcgcgtcaagcgggcttggcatgtgaccgcggcg
tcattgtcgatcactgcggtgcgacgcttgccaaaggcgtattcgccgtcggagatgtggccagttggccgctgcgc
gccggcggccggcgttcgctcgaaacctatatgaacgcgcagcgccaagccgccgcggtggctgcggccattctggg
gaaaaacgtatcggcaccgcaactgcccgtgtcctggacggagatcgctgggcatcgcatgcagatggcgggcgata
tcgaaggacctggtgatttcgtctcgcgcggcatgcccggtagtggcgctgccctgttgttccgcctgcaggagcga
aggattcaggcggtcgtcgcggtcgatgcaccccgtgacttcgcgcttgcaacccgattggtagaagcccgcgcggc
aatcgagccagcacggctggcagatctttcaaacagtatgcgcgattttgttcgtgcgaatgaaggagacctaacgt
gatggcagatctcaattggatatcggccggccacgcgatcgctgacgtcggtaccctcgagtctggtaaagaaaccg
ctgctgcgaaatttgaacgccagcacatggactcgtctactagcgcagcttaattaacctaggctgctgccaccgct
gagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttgctgaaacctcaggcatttga
gaagcacacggtcacactgcttccggtagtcaataaaccggtaaaccagcaatagacataagcggctatttaacgac
cctgccctgaaccgacgaccgggtcatcgtggccggatcttgcggcccctcggcttgaacgaattgttagacattat

ttgccgactaccttggtgatctcgcctttcacgtagtggacaaattcttccaactgatctgcgcgcgaggccaagcg
atcttcttcttgtccaagataagcctgtctagcttcaagtatgacgggctgatactgggccggcaggcgctccattg
cccagtcggcagcgacatccttcggcgcgattttgccggttactgcgctgtaccaaatgcgggacaacgtaagcact
acatttcgctcatcgccagcccagtcgggcggcgagttccatagcgttaaggtttcatttagcgcctcaaatagatc
ctgttcaggaaccggatcaaagagttcctccgccgctggacctaccaaggcaacgctatgttctcttgcttttgtca
gcaagatagccagatcaatgtcgatcgtggctggctcgaagatacctgcaagaatgtcattgcgctgccattctcca
aattgcagttcgcgcttagctggataacgccacggaatgatgtcgtcgtgcacaacaatggtgacttctacagcgcg
gagaatctcgctctctccaggggaagccgaagtttccaaaaggtcgttgatcaaagctcgccgcgttgtttcatcaa
gccttacggtcaccgtaaccagcaaatcaatatcactgtgtggcttcaggccgccatccactgcggagccgtacaaa
tgtacggccagcaacgtcggttcgagatggcgctcgatgacgccaactacctctgatagttgagtcgatacttcggc
gatcaccgcttccctcatactcttccttttttcaatattattgaagcatttatcagggttattgtctcatgagcggat
acatatttgaatgtatttagaaaaataaacaaatagctagctcactcggtcgctacgctccgggcgtgagactgcgg
cgggcgctgcggacacatacaaagttacccacagattccgtggataagcaggggactaacatgtgaggcaaaacagc
agggccgcgccggtggcgtttttccataggctccgccctcctgccagagttcacataaacagacgcttttccggtgc
atctgtgggagccgtgaggctcaaccatgaatctgacagtacgggcgaaacccgacaggacttaaagatccccaccg
tttccggcgggtcgctccctcttgcgctctcctgttccgaccctgccgtttaccggatacctgttccgcctttctcc
cttacgggaagtgtggcgctttctcatagctcacacactggtatctcggctcggtgtaggtcgttcgctccaagctg
ggctgtaagcaagaactccccgttcagcccgactgctgcgccttatccggtaactgttcacttgagtccaacccgga
aaagcacggtaaaacgccactggcagcagccattggtaactgggagttcgcagaggatttgtttagctaaacacgcg
gttgctcttgaagtgtgcgccaaagtccggctacactggaaggacagatttggttgctgtgctctgcgaaagccagt
taccacggttaagcagttccccaactgacttaaccttcgatcaaaccacctccccaggtggttttttcgtttacagg
gcaaaagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctactgaaccgctctagatttcag
tgcaatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgtaattctcatgttagtcatgc
cccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtg
agctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatg
aatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcaccagtgagacgggc
aacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcg
aaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgaga
tgtccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagc
atcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttc
ccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacag
aacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccg
tcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgca
ggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaa
gattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttga
tcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcag
caacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccactt
tttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatac
tctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccat
accgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaaa
ttaatacgactcactata

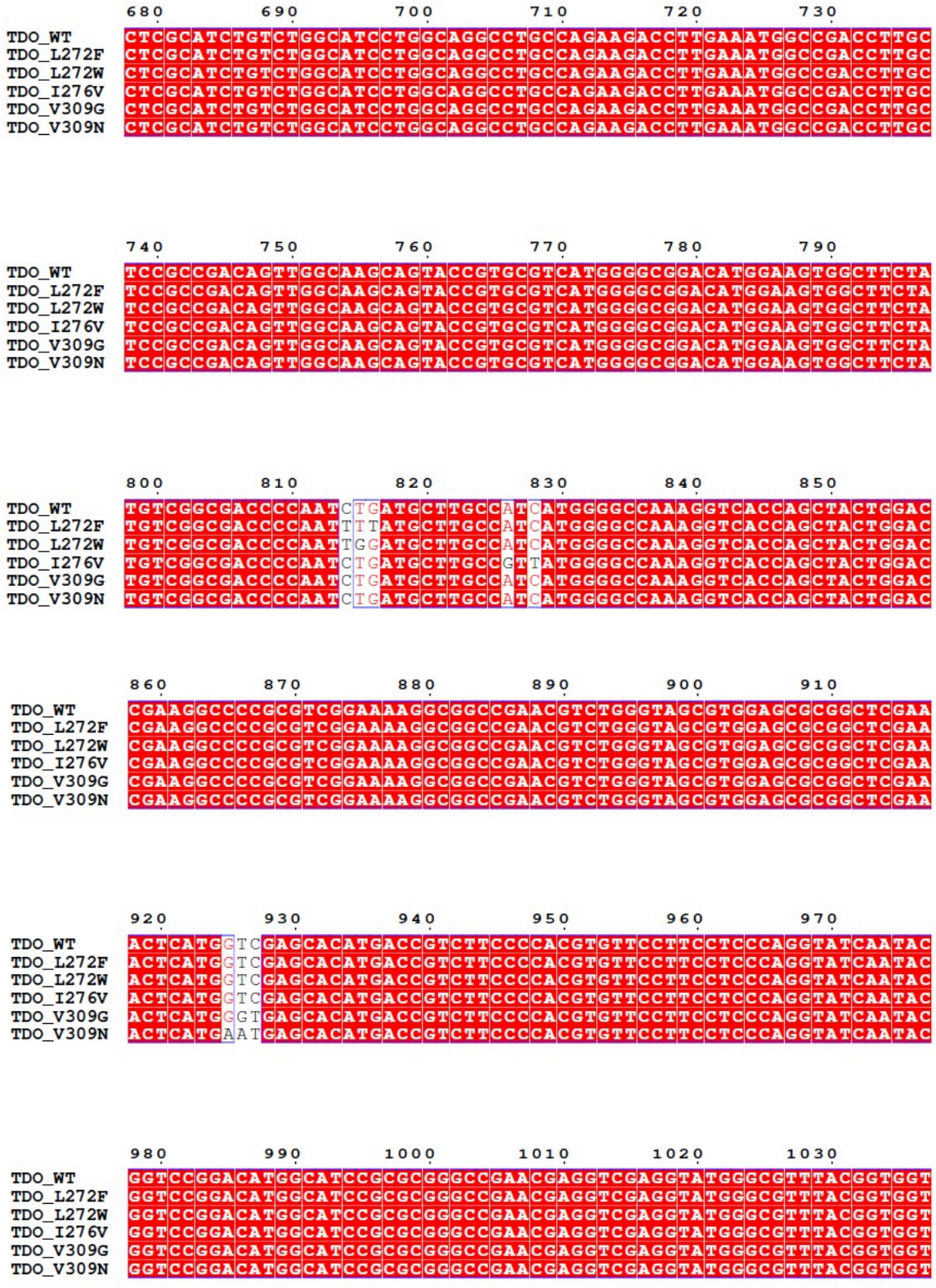


Figure S3: Multiple sequence alignment of sequencing data from TDO variants with single active site mutations. Alignment performed with M-Coffee.[12] Image generated with ESPript.[13]

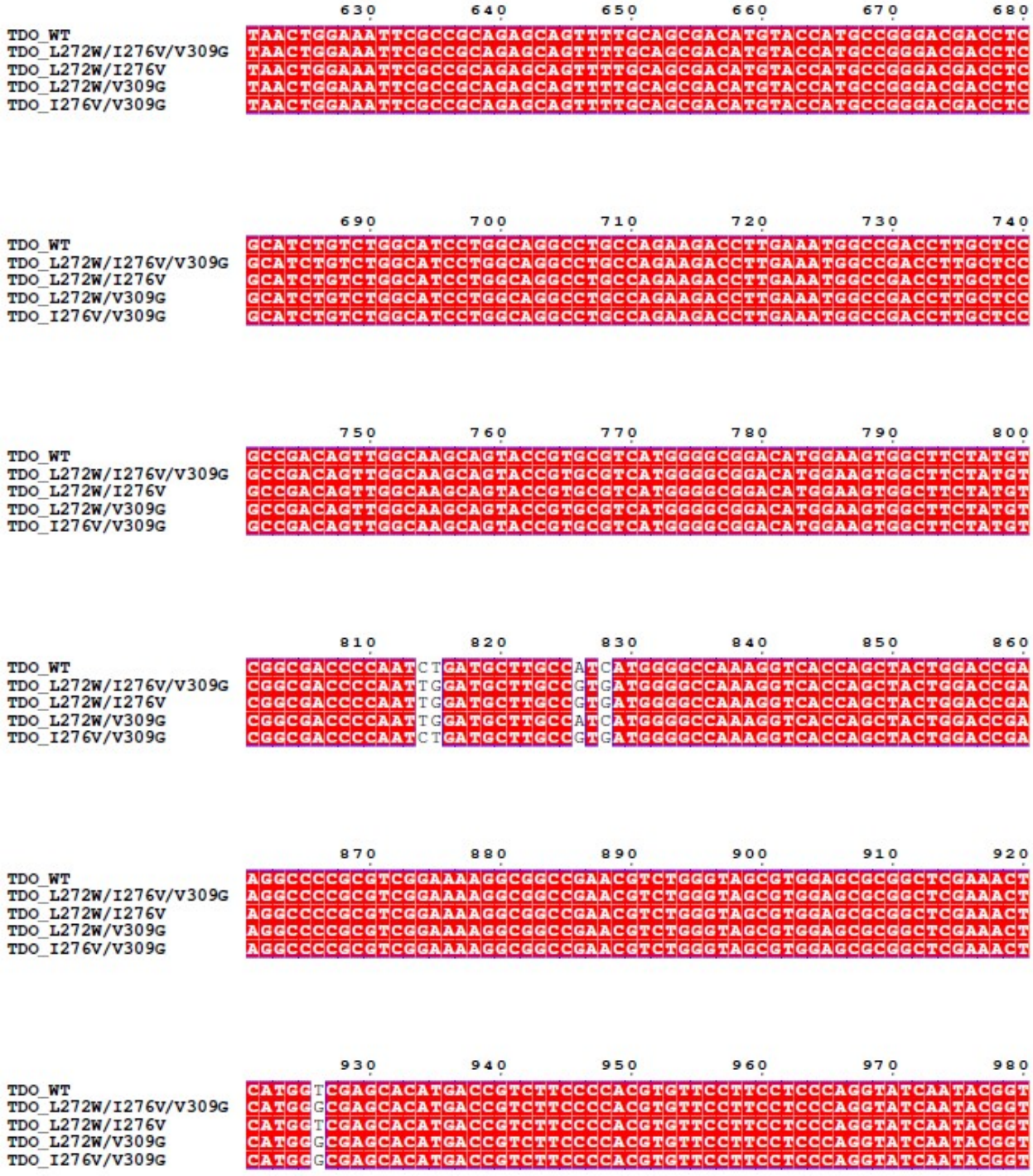


Figure S4: Multiple sequence alignment of sequencing data from TDO variants with double and triple active site mutations. Alignment performed with M-Coffee.[12] Image generated with ESPript.[13]

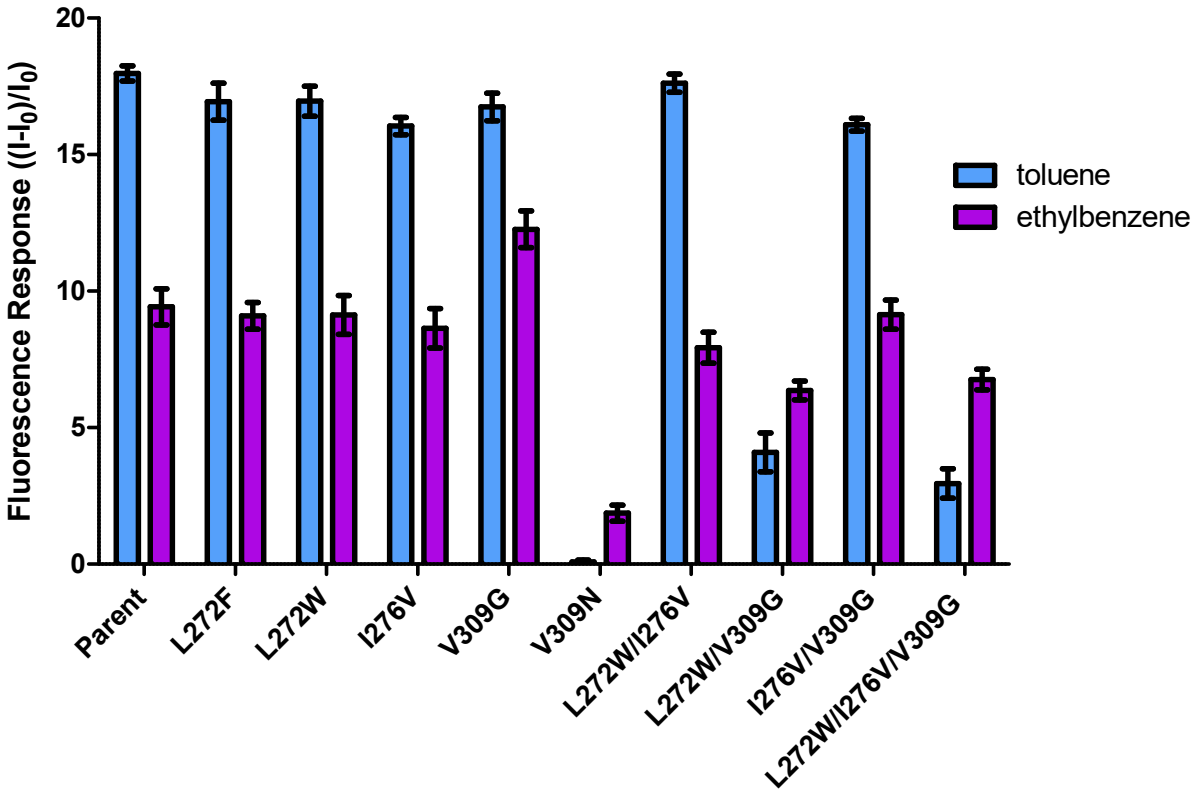
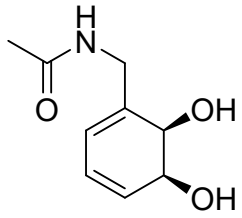


Figure S5: Relative activity of TDO variants for the native substrate (toluene) and ethylbenzene compared to the parent enzyme (n = 6). Fluorescence response of each variant was normalized to the mean fluorescence response of the negative controls (*E. coli* BL21 (DE3) pCP-01)[5] ([I – I_0]/I_0) (n = 6).

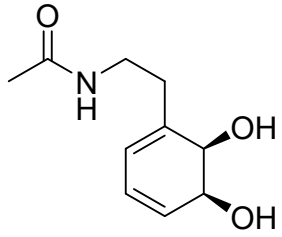
[(5S,6R)-5,6-Dihydroxy-1,3-cyclohexadien-1-yl]methylacetamide (**1**)



20

R_f = 0.45 [4:1 (DCM:MeOH)]; $[\alpha]^{D}$ = +12.1 (c = 0.51, MeOH); IR (film) ν 3288, 2917, 1637, 1551, 1422, 1373, 1290, 1151, 1075, 1036, 994, 928, 866, 810; ^{1}H NMR (400 MHz, MeOD) δ 5.91 (m, 1H), 5.80 (m, 2H), 4.21 (s, 1H), 4.05-3.99 (m, 2H), 3.83 (d, J = 16.3, 1H), 1.96 (s, 3H); ^{13}C NMR (100 MHz, MeOD) δ 176.1, 141.4, 133.1, 127.5, 124.0, 72.9, 72.2, 45.5, 25.1; HRMS (ESI) calcd. for $C_9H_{13}NO_3Na^+$: 206.0793, found: 206.0793.

[(5S,6R)-5,6-Dihydroxy-1,3-cyclohexadien-1-yl]ethylacetamide (**2**)



20

R_f = 0.52 [4:1 (DCM:MeOH)]; $[\alpha]^{D}$ = +22.0 (c = 0.58, MeOH); IR (film) ν 3288, 3104, 1631, 1555, 1436, 1372, 1297, 1157, 1081, 1062, 1018, 987, 805; ^{1}H NMR (400 MHz, MeOD) δ 5.89 (m, 1H), 5.73 (m, 2H), 4.19 (s, 1H), 3.99 (d, J = 5.9, 1H), 3.39-3.28 (m, 2H), 2.43-2.31 (m, 2H), 1.89 (s, 3H); ^{13}C NMR (100 MHz, MeOD) δ 171.9, 138.5, 128.2, 124.0, 121.1, 69.4, 69.0, 38.0, 33.5, 21.3; HRMS (ESI) calcd. for $C_{10}H_{15}NO_3Na^+$: 220.0950, found: 220.0950.

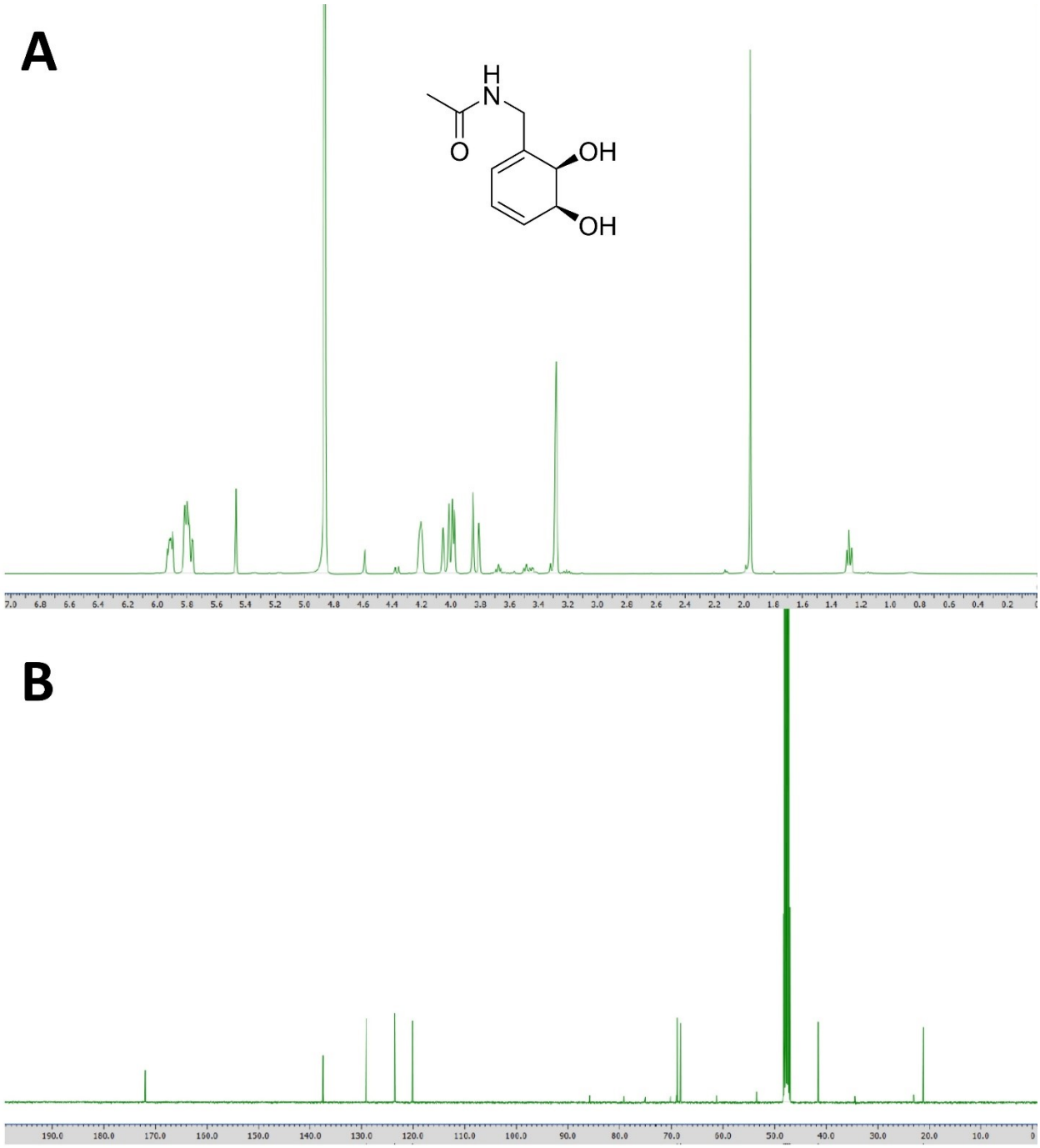


Figure S6: (A) ^{1}H NMR spectrum of [(5S,6R)-5,6-Dihydroxy-1,3-cyclohexadien-1-yl]methylacetamide (**1**). NMR analysis performed in MeOD; (B) ^{13}C NMR spectrum of [(5S,6R)-5,6-Dihydroxy-1,3-cyclohexadien-1-yl]methylacetamide (**1**). NMR analysis performed in MeOD.

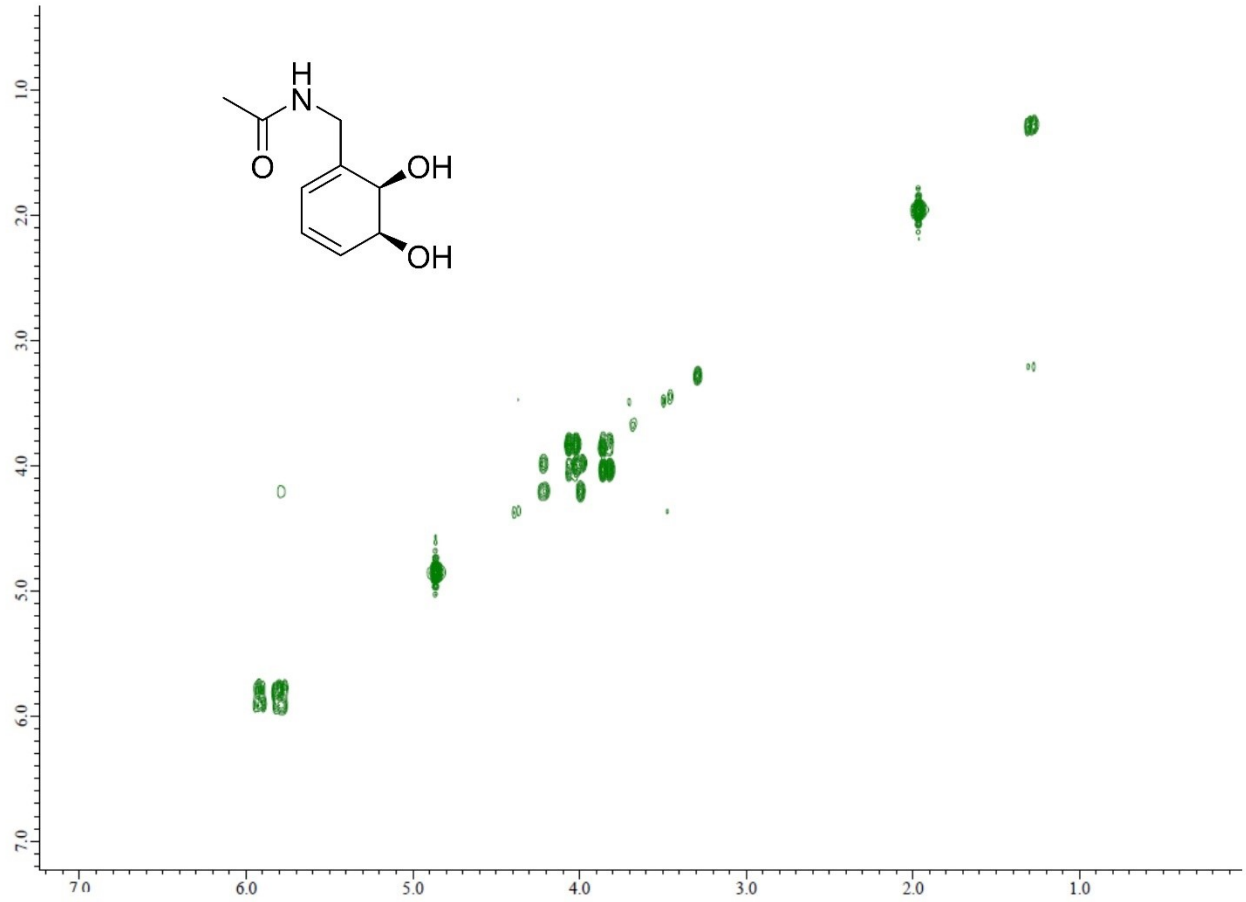


Figure S7: Correlation NMR (COSY) spectrum of [(5S,6R)-5,6-Dihydroxy-1,3-cyclohexadien-1-yl]methylacetamide (**1**). NMR analysis performed in MeOD.

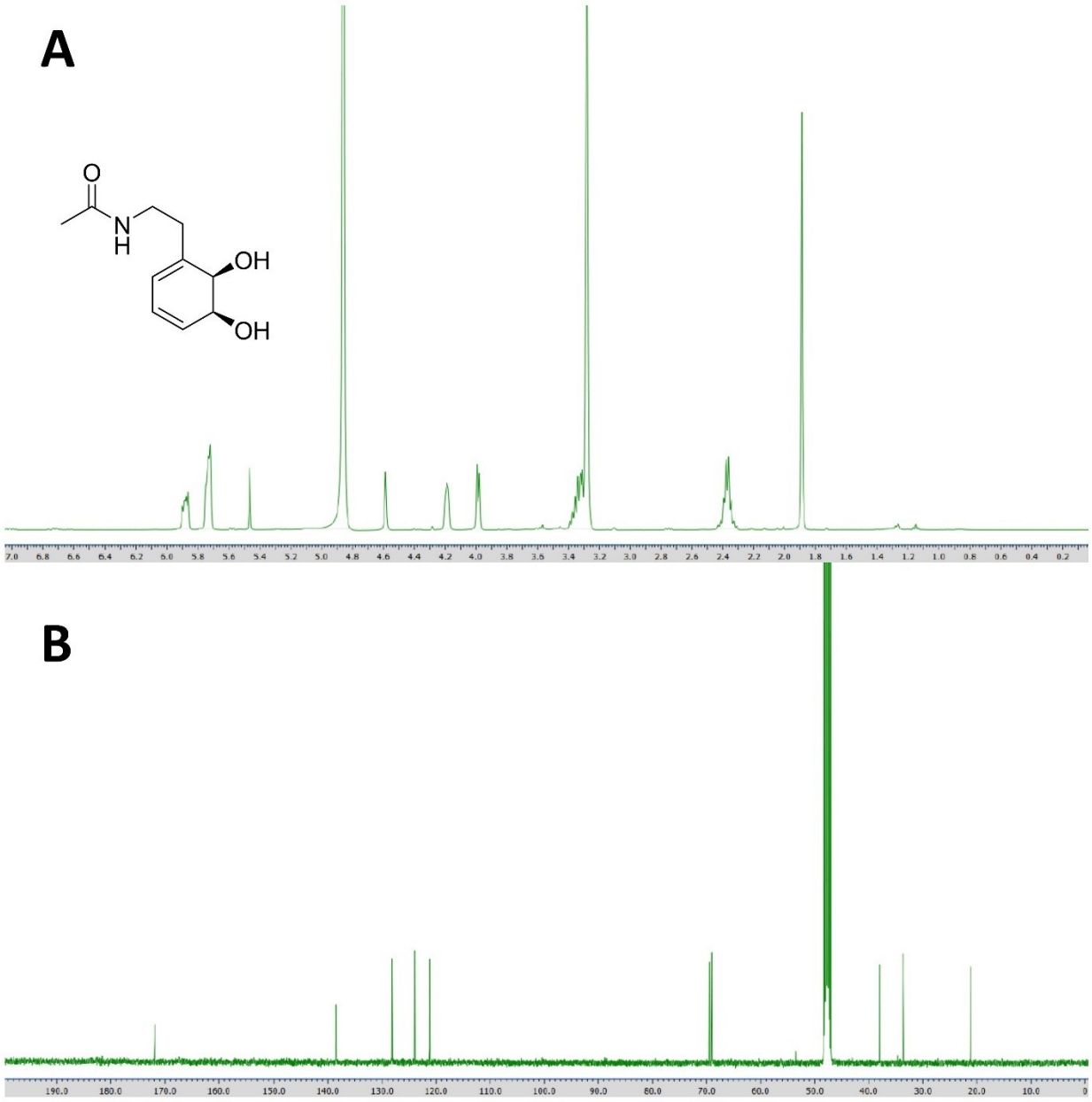


Figure S8: (A) ^{1}H NMR spectrum of [(5S,6R)-5,6-Dihydroxy-1,3-cyclohexadien-1-yl]ethylacetamide (**2**). NMR analysis performed in MeOD; (B) ^{13}C NMR spectrum of [(5S,6R)-5,6-Dihydroxy-1,3-cyclohexadien-1-yl]ethylacetamide (**2**). NMR analysis performed in MeOD.

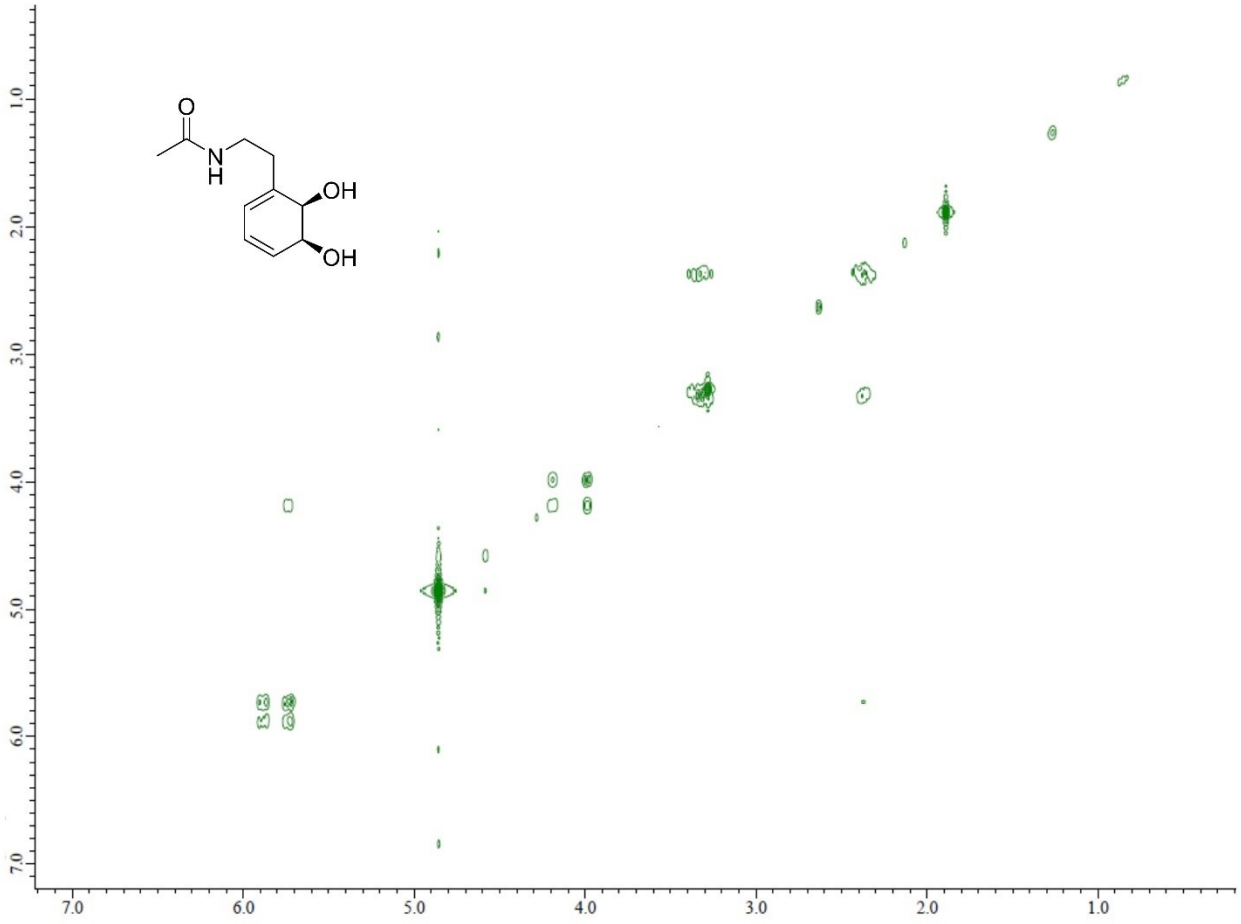


Figure S9: Correlation NMR (COSY) spectrum of [(5S,6R)-5,6-Dihydroxy-1,3-cyclohexadien-1-yl]methylacetamide (**1**). NMR analysis performed in MeOD.

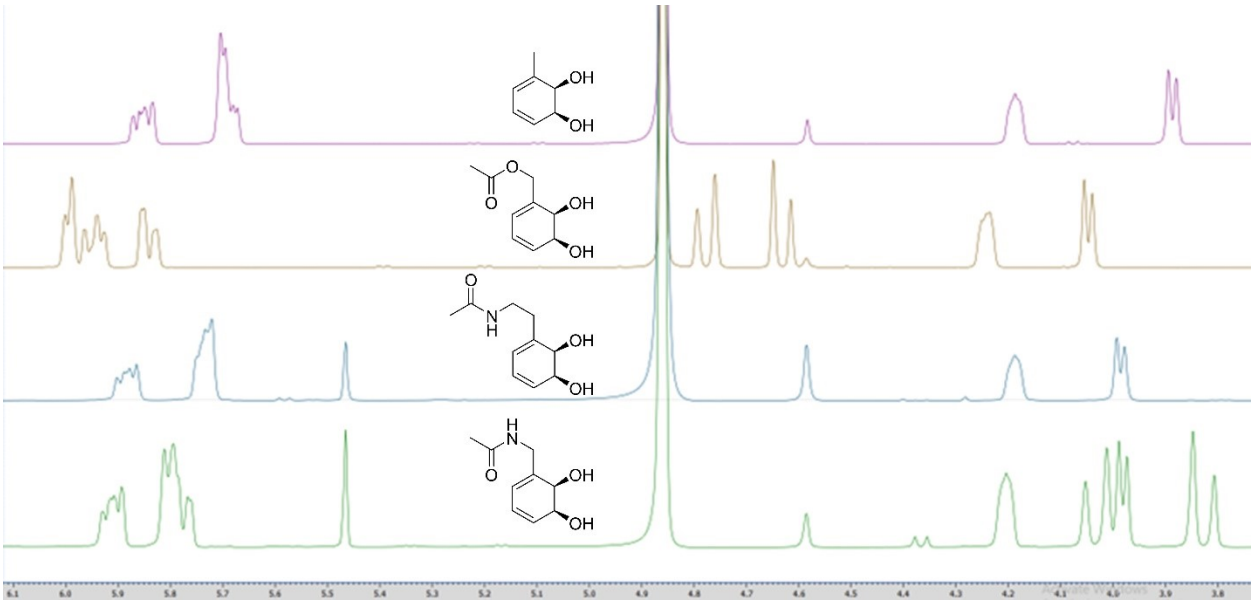


Figure S10: Comparison of the ^{1}H NMR spectra of [(5S,6R)-5,6-Dihydroxy-1,3-cyclohexadien-1-yl]methylacetamide (**1**) and [(5S,6R)-5,6-Dihydroxy-1,3-cyclohexadien-1-yl]ethylacetamide (**2**) with other previously characterized *cis*-diol metabolites produced by the wild-type toluene dioxygenase (TDO). Overlay isolates alkene signals and C<u>H</u>OH signals. All NMR analyses performed in MeOD.

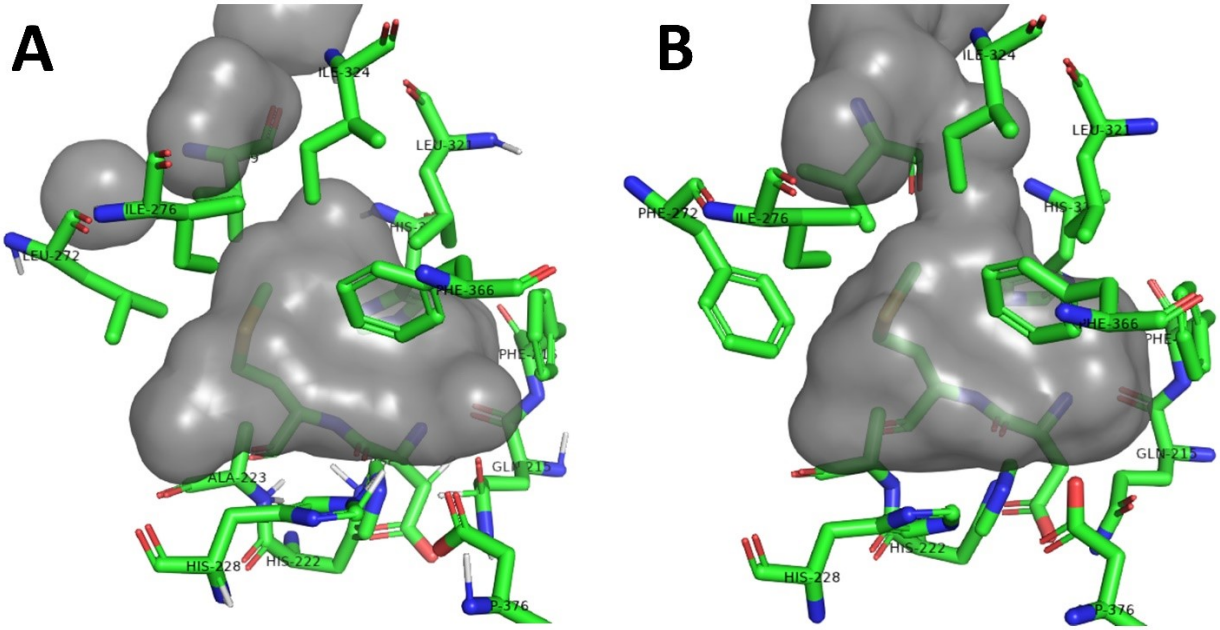


Figure S11: (A) Visualization of the active site cavity of TDO (wild-type). Image generated using the reported crystal structure of TDO[14] and with PyMOL.[4] (B) Visualization of the active site cavity of TDO L272F. Image generated using a homology model of TDO L272F and with PyMOL.[4]

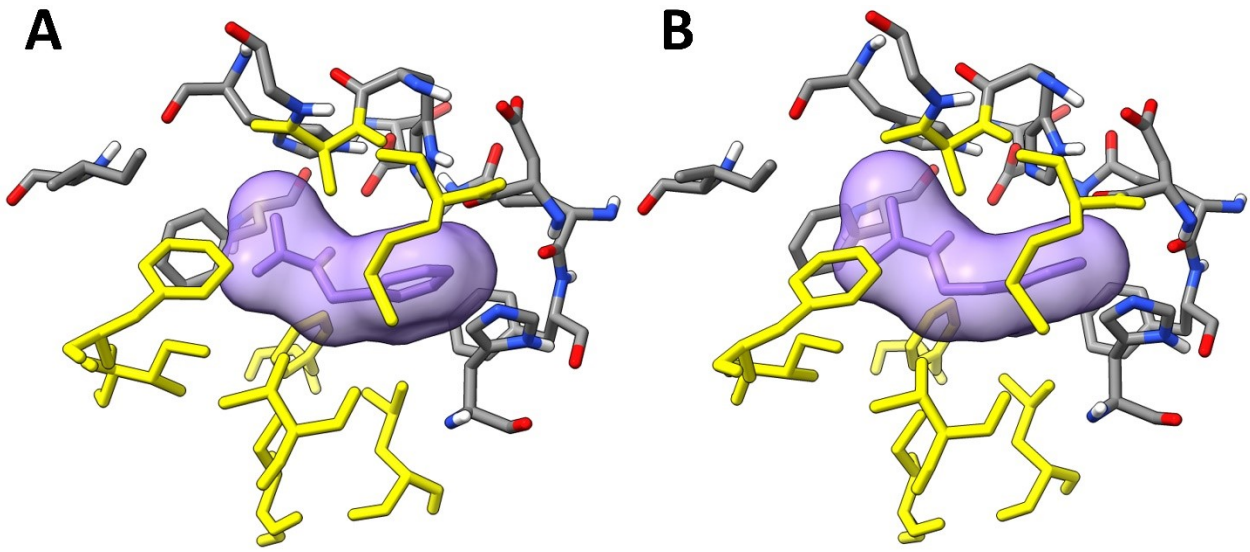


Figure S12: (A) Active site structure produced from molecular dynamics simulations performed on the docking prediction of TDO L272F with *N*-benzylacetamide (purple). Residues targeted for mutagenesis in this study are highlighted (yellow). Docking was performed with AutoDock Vina,[2] using a TDO L272F homology model. Molecular dynamics simulations were performed with GROMACS.[3] Image was generated using ChimeraX software.[15] (B) Active site structure produced from molecular dynamics simulations performed on the docking prediction of TDO L272F with *N*-(2-phenethyl)acetamide (purple). Residues targeted for mutagenesis in this study are highlighted (yellow). Docking was performed with AutoDock Vina,[2] using a TDO L272F homology model. Molecular dynamics simulations were performed with GROMACS.[3] Image was generated using ChimeraX software.[15]

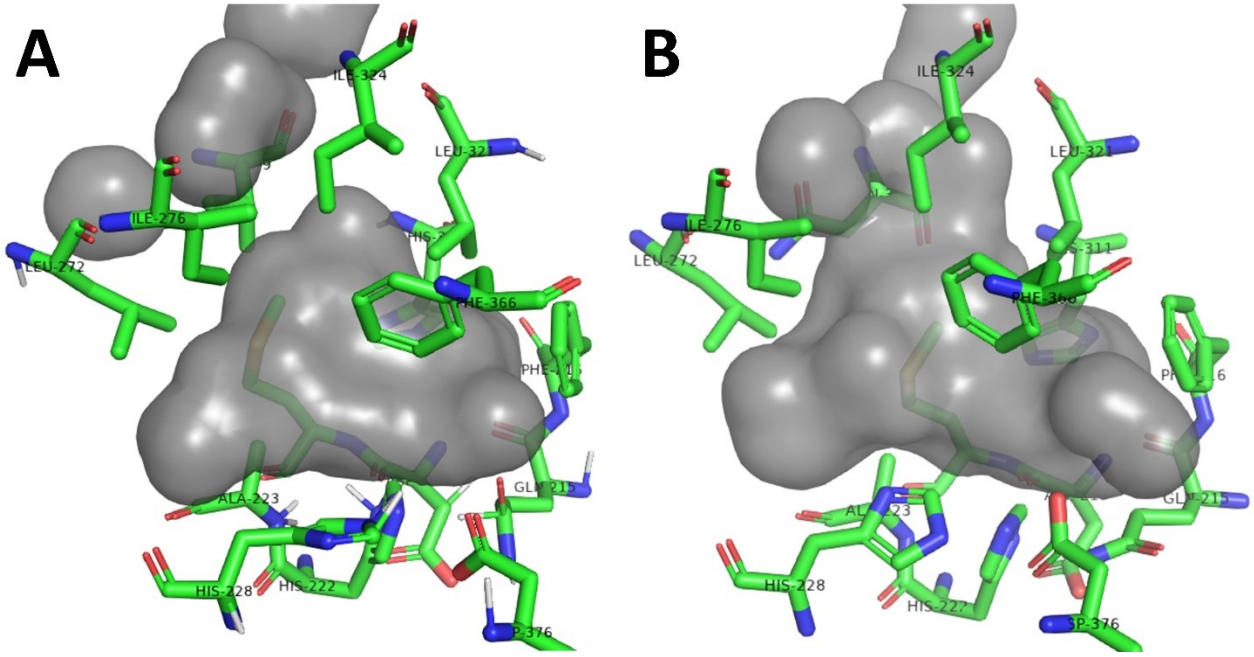


Figure S13: (A) Visualization of the active site cavity of TDO (wild-type). Image generated using the reported crystal structure of TDO[14] and with PyMOL.[4] (B) Visualization of the active site cavity of TDO V309N. Image generated using a homology model of TDO V309N and with PyMOL.[4]

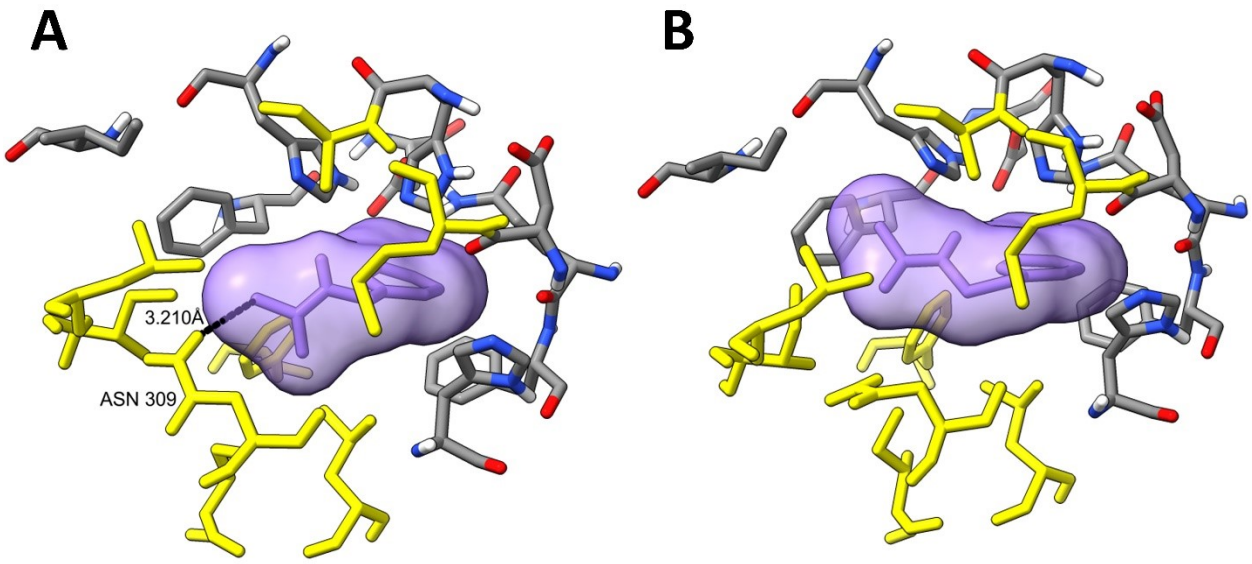


Figure S14: (A) Active site structure produced from molecular dynamics simulations performed on the docking prediction of TDO V309N with *N*-benzylacetamide (purple). The distance between the substrate and the N309 residue is shown. Residues targeted for mutagenesis in this study are highlighted (yellow). Docking was performed with AutoDock Vina,[2] using a TDO V309N homology model. Molecular dynamics simulations were performed with GROMACS.[3] Image was generated using ChimeraX software.[15] (B) Active site structure produced from molecular dynamics simulations performed on the docking prediction of TDO V309N with *N*-(2-phenethyl)acetamide (purple). Residues targeted for mutagenesis in this study are highlighted (yellow). Docking was performed with AutoDock Vina,[2] using a TDO V309N homology model. Molecular dynamics simulations were performed with GROMACS.[3] Image was generated using ChimeraX software.[15]

References

1) 16 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, *Nature* 2021, **596**, 583.
2) O. Trott and A. J. Olson, *J. Comput. Chem*. 2010, **31**, 455.

3) a) D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comp. Chem*. 2005, **26**, 1701; b) S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845; c) M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX* 2015, **1**, 19.

4) The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

5) C. Preston-Herrera, A. S. Jackson, B. O. Bachmann and J. T. Froese, *Org. Biomol. Chem*. 2021, **19**, 775.

6) G. J. Zylstra and D. T. Gibson, *J. Biol. Chem*. 1989, 264, 14940.

7) (a) L. Zheng, U. Baumann and J.-L. Reymond, *Nuc. Acids Res*. 2004, **32**, e115; (b) H. Liu and J. H. Naismith, *BMC Biotechnology* 2008, **8**, 91.

8) M. A. Endoma, V. P. Bui, J. Hansen and T. Hudlicky, *Org. Process Res. Dev*. 2002, **6**, 525.

9) J. Marley, M. Lu and C. Bracken, *J. Biomol. NMR* 2001, **20**, 71.

10) G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem*. 2009, **16**, 2785.

11) (a) R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig and A. D. MacKerell Jr., *J. Chem. Theory Comput*. 2009, **8**, 3257; (b) J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmuller and A. D. MacKerell Jr., *Nature Methods*, 2016, **14**, 71.

12) (a) C. Notredame, D. G. Higgins and J. Heringa, *J. Mol. Biol*. **2000**, *302*, 205; (b) I. M. Wallace, O. O'Sullivan, D. G. Higgins and C. Notredame, *Nuc. Acids Res*. **2006**, *34*, 1692; (c) S. Moretti, F. Armougom, I. M. Wallace, D. G. Higgins, C. V. Jongeneel and C. Notredame, *Nuc. Acids Res*. **2007**, *35*, W645; (d) P. Di Tommaso, S. Moretti, I. Xenarios, M. Orobitg, A. Montanyola, J.-M. Chang, J.-F. Taly and C. Notredame, *Nuc. Acids Res*. **2011**, *39*, W13.

13) X. Robert and P. Gouet, *Nuc. Acids Res*. **2014**, *42*, W320.

14) R. Friemann, K. Lee, E. N. Brown, D. T. Gibson, H. Eklund and S. Ramaswamy, *Acta Crys*. 2009, **D65**, 24.

15) T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris and T. E. Ferrin, *Protein Sci*. 2018, **27**, 14.