

Using structurally fungible biosensors to evolve improved alkaloid biosyntheses

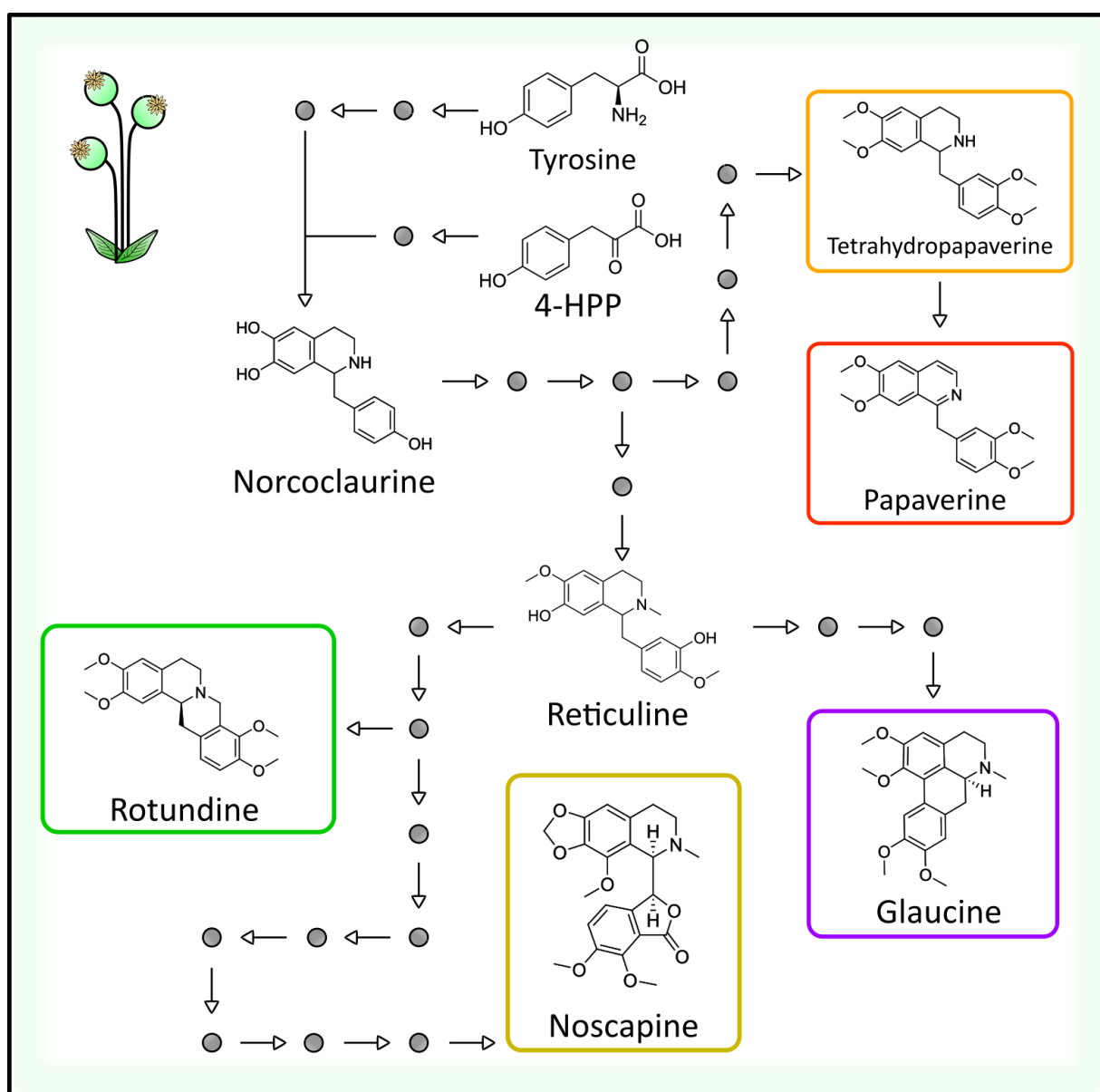
Simon d'Oelsnitz*, Wantae Kim, Nathaniel T. Burkholder, Kamyab Javanmardi, Ross Thyer, Yan Zhang, Hal Alper, Andrew D. Ellington*

Supplementary Information:

Supporting Figures 1-18

Supporting Tables 1-4

Supplementary Figures



Supplementary Figure 1. Benzylisoquinoline pathway map.

Arrows represent enzymatic steps and grey circles represent metabolites. Target alkaloids in this work are highlighted with a colored border.

a **Pbm3r1**
ggctgtcgacctttgaaaagttcg **TTGACA** gctagctcagtcctagg **TACAAT** CGGAATGAACGTTTCATTCCGttt

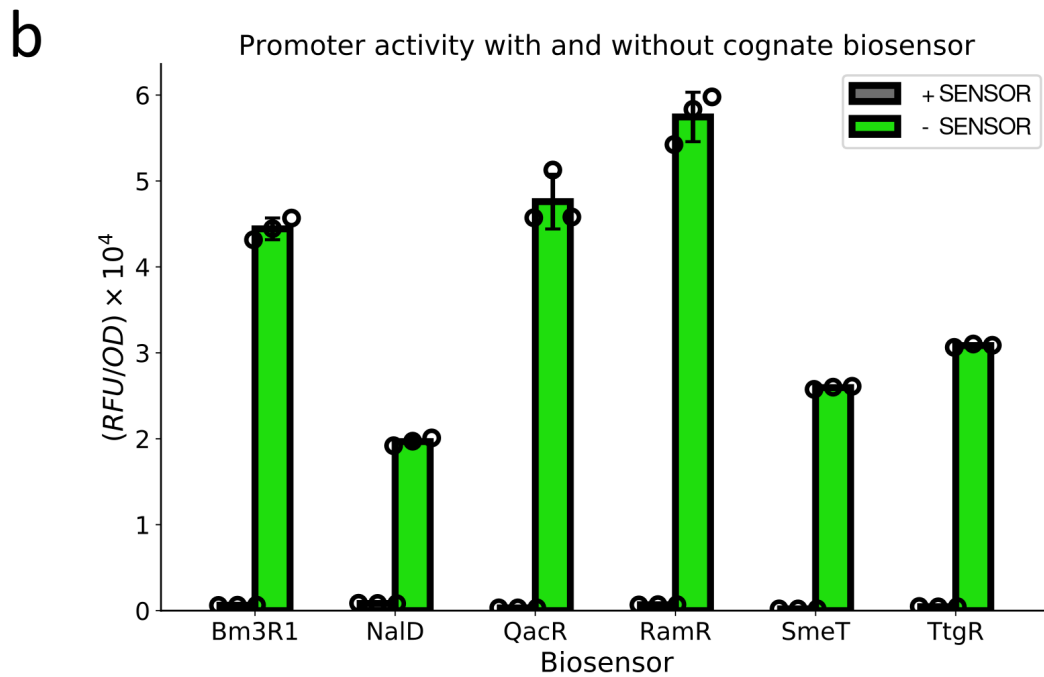
Pramr
ctttgaaaagtacc **TTGACG** gcgtatccttgcttc **TATAATG** AGTGCTTACTCACTCATA

Pqacr
attcgttacca **TTGACA** gctaGCTCAGTCCTACT **TTTAGTAT** AGAGACTGAGCGGTCGGTCTATAta

Pttgr
aacgcaccagcagT **ATTTACAA** ACAACCATGAATGTAA **GATAAAT** agttagcaat

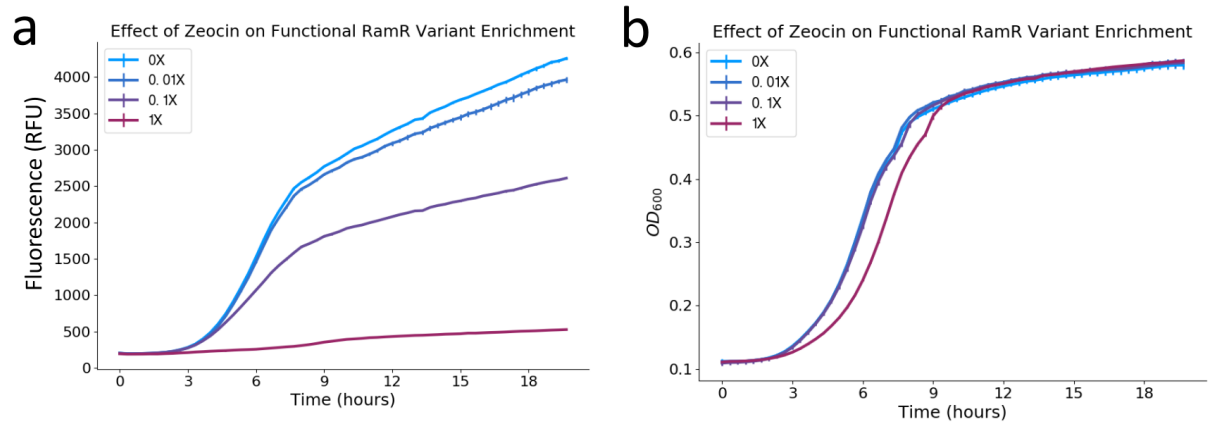
Pnald
caccttccggacg **TTTACAA** AACACCTATGAATGTAA **GATAAAT** agttagcaa

Psmet
caccgcagcagT **ATTTACAA** ACAAACAAGCATGTAT **GATAAAT** ctttagcaac



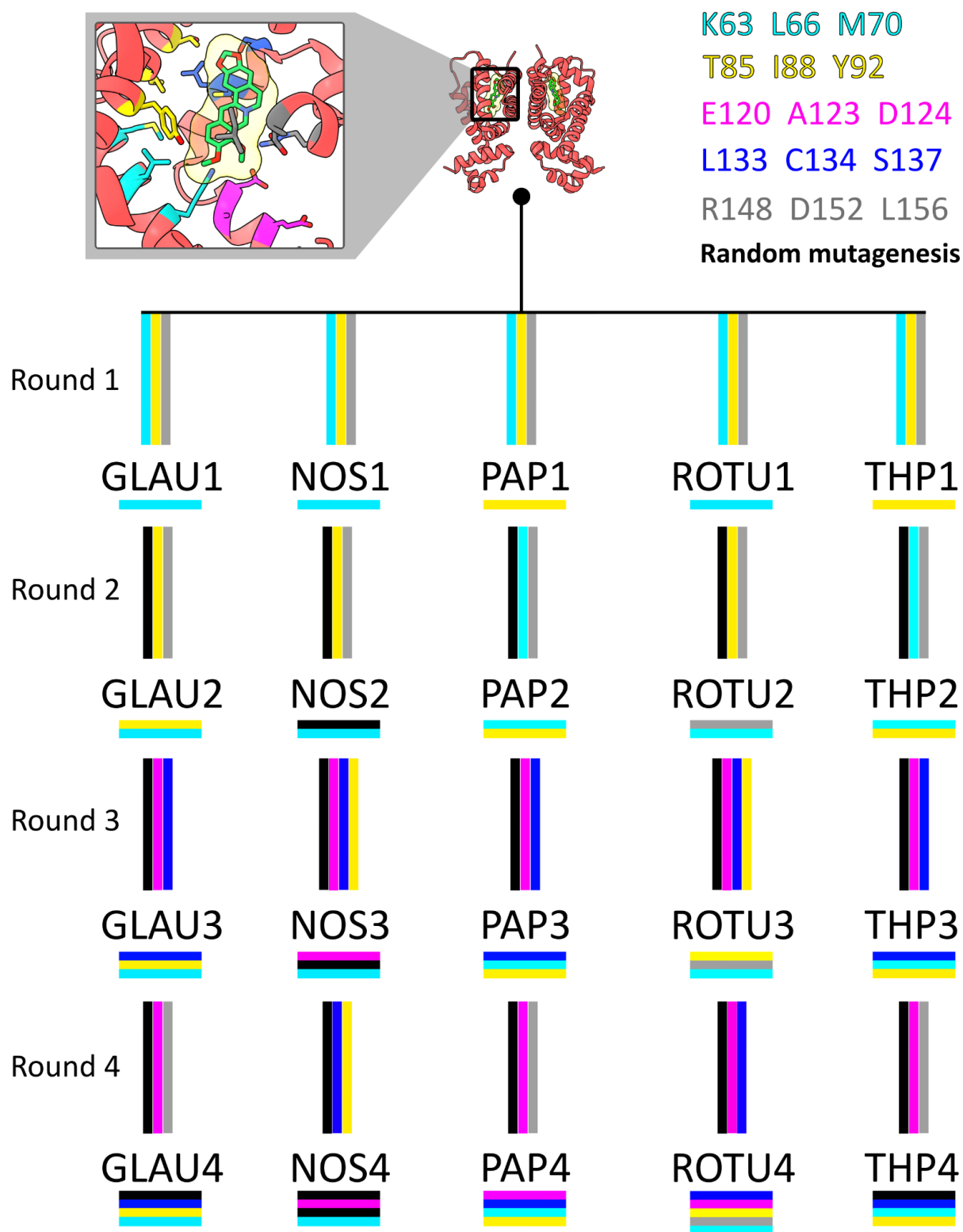
Supplementary Figure 2. Multidrug resistance regulator design and validation.

(a) Promoter design for each regulator. -35 and -10 promoter regions are highlighted with a red and yellow box, respectively. Operator sequences are underlined. All promoters are followed by the RiboJ riboregulator, a medium strength RBS, the sfGFP gene, and a strong terminator. (b) Validation of promoter activity and regulator repression in *E. coli*. Cells were co-transformed with the regulator's promoter and either an empty vector (- Sensor) or a vector expressing the cognate regulator (+ Sensor) and promoter activity was monitored via fluorescence.



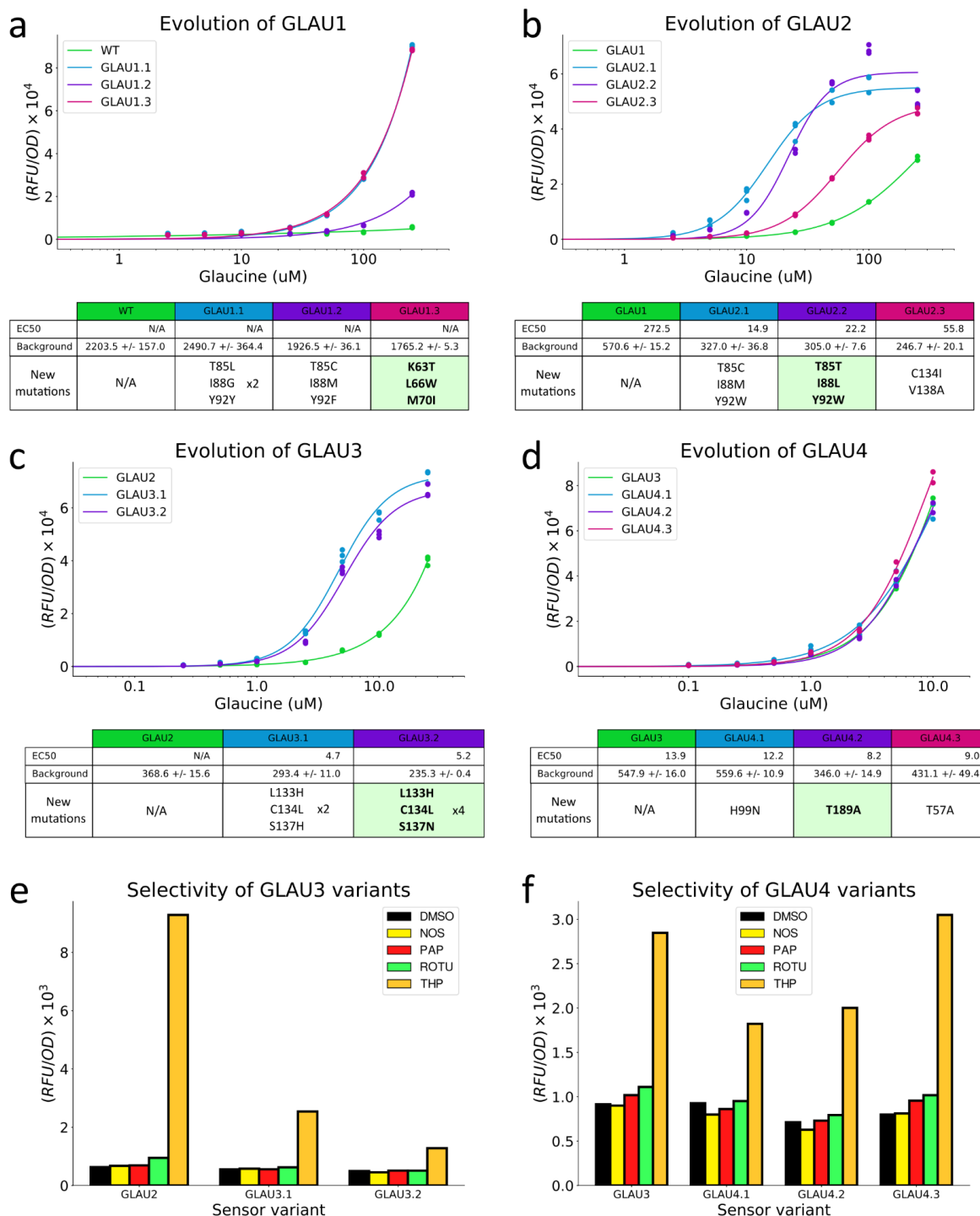
Supplementary Figure 3. Trial negative selections with the pZeoG plasmid.

Cells co-transformed with both pReg expressing a library of RamR variants and the pZeoG plasmid were grown for 20 hours in the presence of variable amounts of zeocin and fluorescence (**a**) and cell density (**b**) were monitored. The “1X” concentration represents 100 ug/mL of zeocin. Assays were performed in biological triplicate.



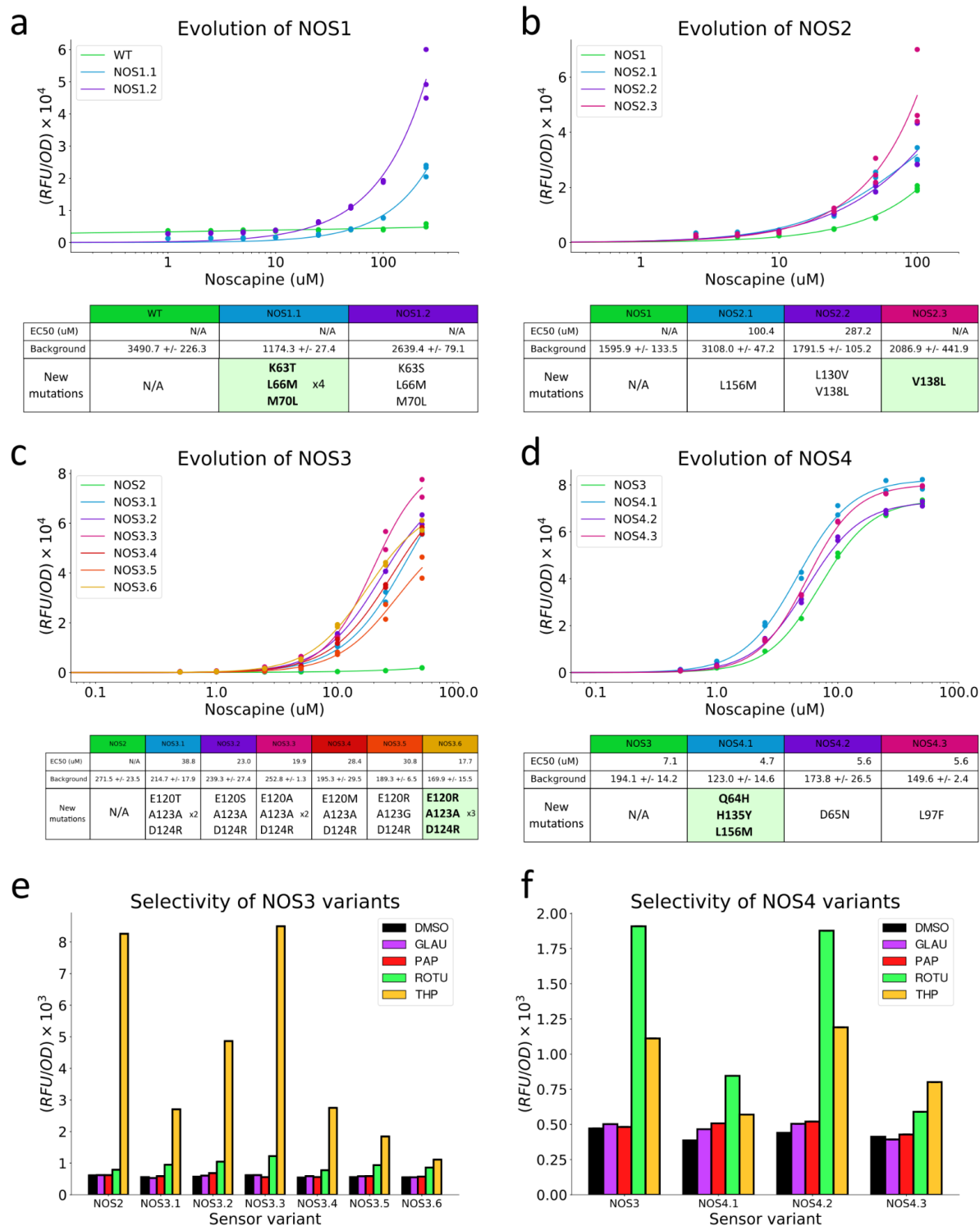
Supplementary Figure 4. Visual representation of libraries used throughout evolution.

(top left) A magnified structure of RamR bound to berberine (PDB: 3VW2) displays residues targeted for mutagenesis. These residues were chosen based on their proximity to berberine. (top middle) global structure of RamR. (top right) The mapping of library color code to the corresponding residues targeted for combinatorial site saturation mutagenesis. (bottom) Libraries used and fixed during evolution. Colored vertical lines represent libraries used to introduce diversity prior to selection. Colored horizontal lines represent library positions fixed.

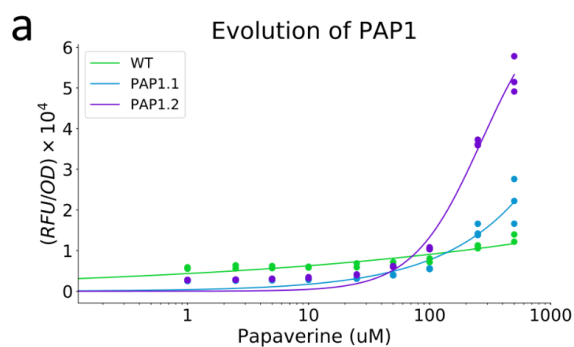


Supplementary Figure 5. Performance of all top GLAU variants recovered.

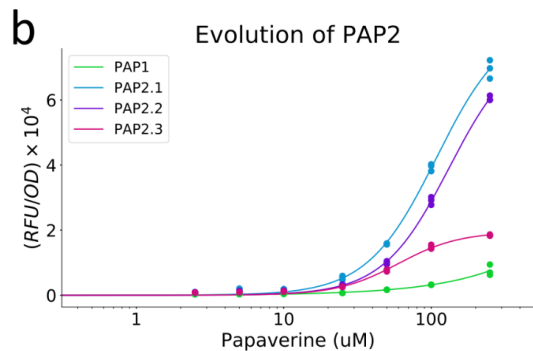
(a-d) Dose response functions of top unique GLAU variants. Variants were chosen based on their signal/noise ratio measured during evolution (See Figure 2c). All variants were subcloned into a new pReg backbone prior to characterization with the pGFP plasmid. The "x2" symbol denotes that this amino acid sequence was recovered twice following evolution. The variant genotype highlighted in green was chosen as the template for the following round of evolution. Dose response measurements were performed in biological triplicate. (e,f) Selectivity of generation three and four sensor variants. Cells were induced with 100 μ M of all non-target BIAs, separately.



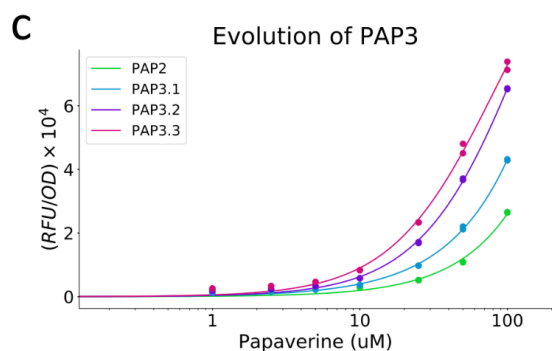
Supplementary Figure 6. Performance of all top NOS variants recovered.
See Supplementary Figure 5 legend for details.



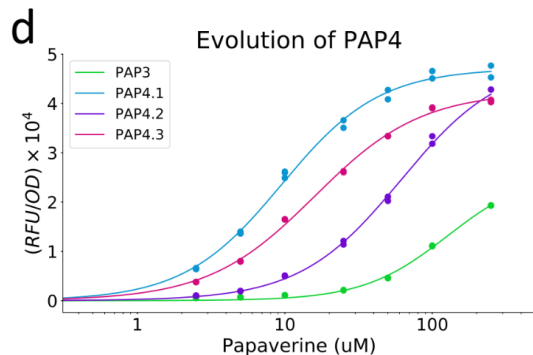
	WT	PAP1.1	PAP1.2
EC50 (uM)	N/A	N/A	265.6
Background	5728.3 +/- 309.4	2581.0 +/- 201.7	2626.8 +/- 31.9
New mutations	N/A	T85L I88F Y92G	T85M I88I Y92G x2



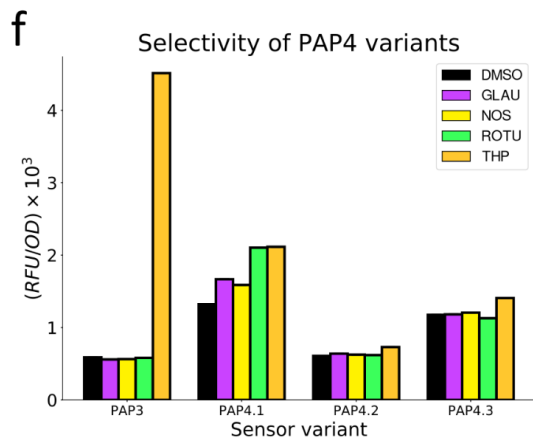
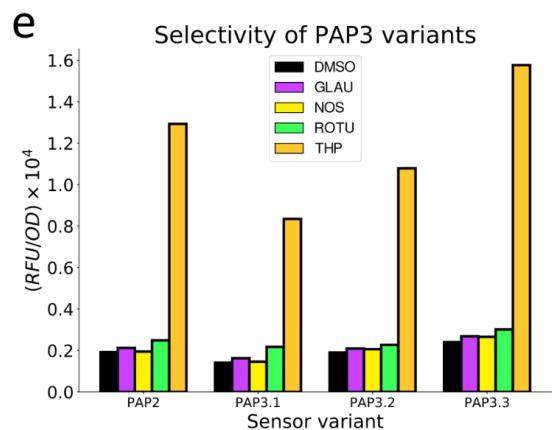
	PAP1	PAP2.1	PAP2.2	PAP2.3
EC50 (uM)	N/A	108.3	132.0	60.0
Background	237.1 +/- 13.8	697.5 +/- 98.9	673.4 +/- 10.4	621.5 +/- 16.4
New mutations	N/A	K63R L66L M70L	K63H L66L M70L	C134I x2



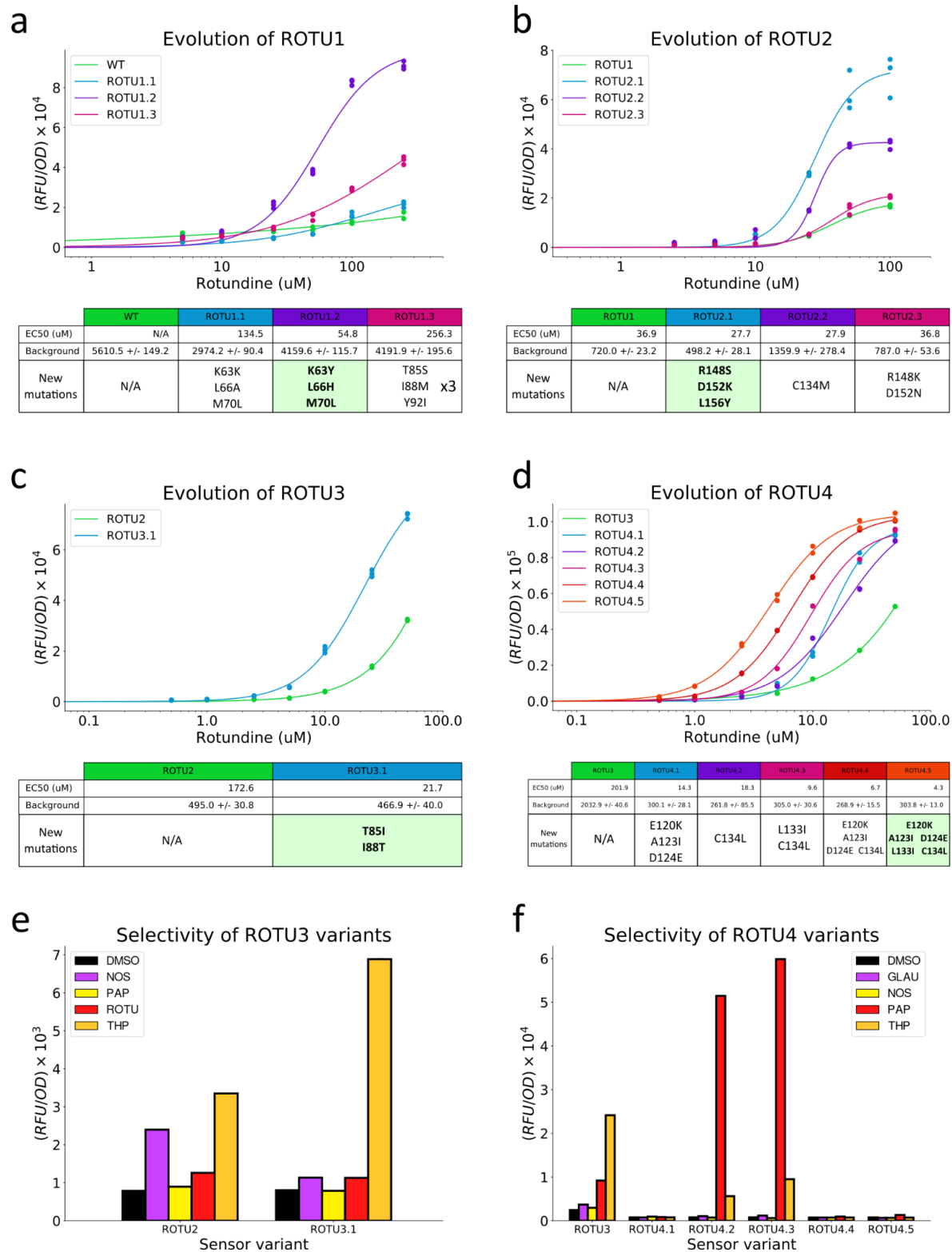
	PAP2	PAP3.1	PAP3.2	PAP3.3
EC50 (uM)	N/A	N/A	139.7	82.9
Background	1577.1 +/- 51.0	1108.0 +/- 38.2	1567.7 +/- 14.5	2037.1 +/- 63.4
New mutations	N/A	E120H A123G D124E	L133I C134L S137R	L133I C134I S137K



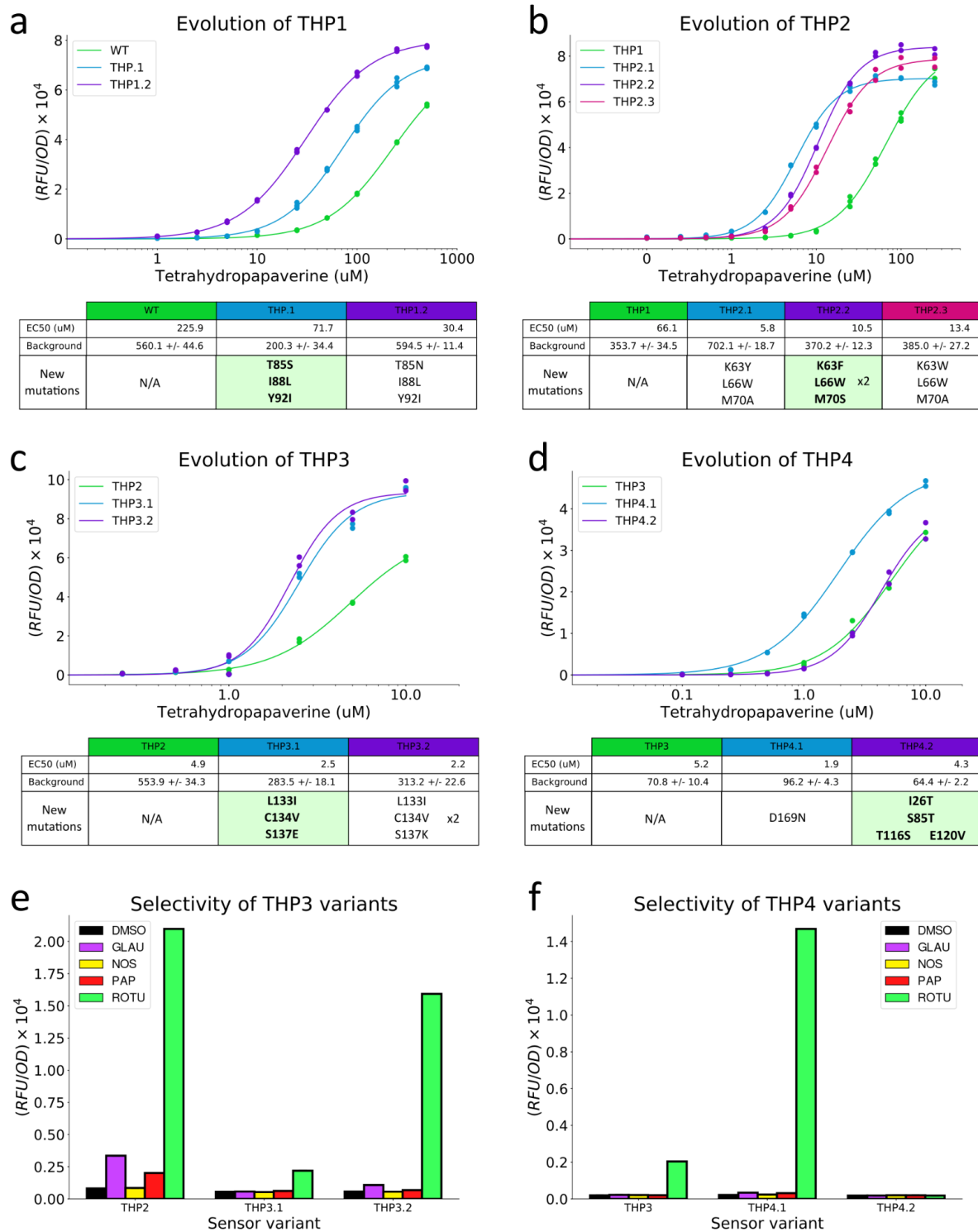
	PAP3	PAP4.1	PAP4.2	PAP4.3
EC50 (uM)	126.8	9.5	61.1	15.9
Background	417.1 +/- 19.6	1266.0 +/- 49.4	508.2 +/- 4.5	980.4 +/- 6.5
New mutations	N/A	E120H A123G x3 D124L	E120H A123D D124L	E120R A123D D124E



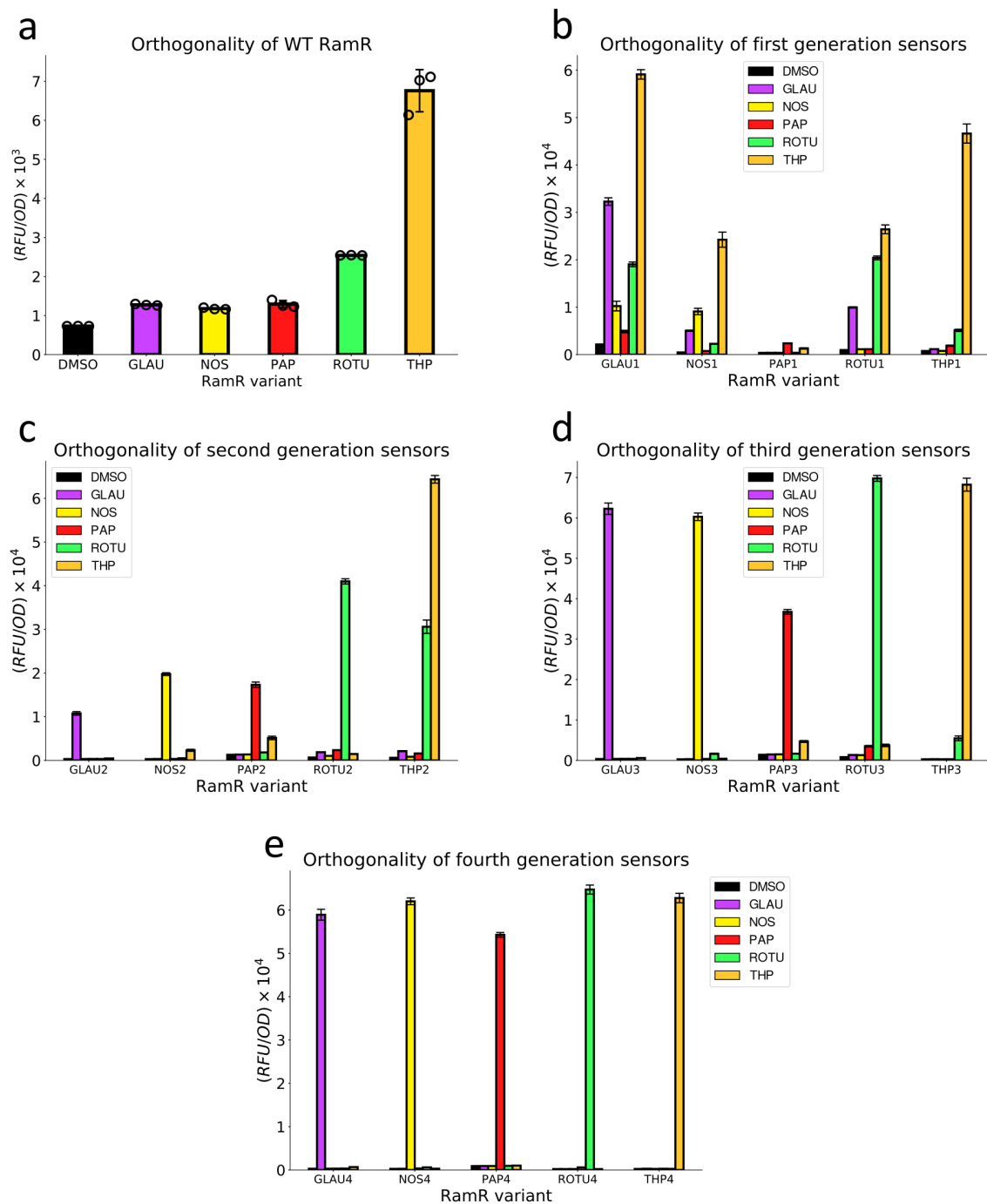
Supplementary Figure 7. Performance of all top PAP variants recovered.
See Supplementary Figure 5 legend for details.



Supplementary Figure 8. Performance of all top ROTU variants recovered.
See Supplementary Figure 5 legend for details.

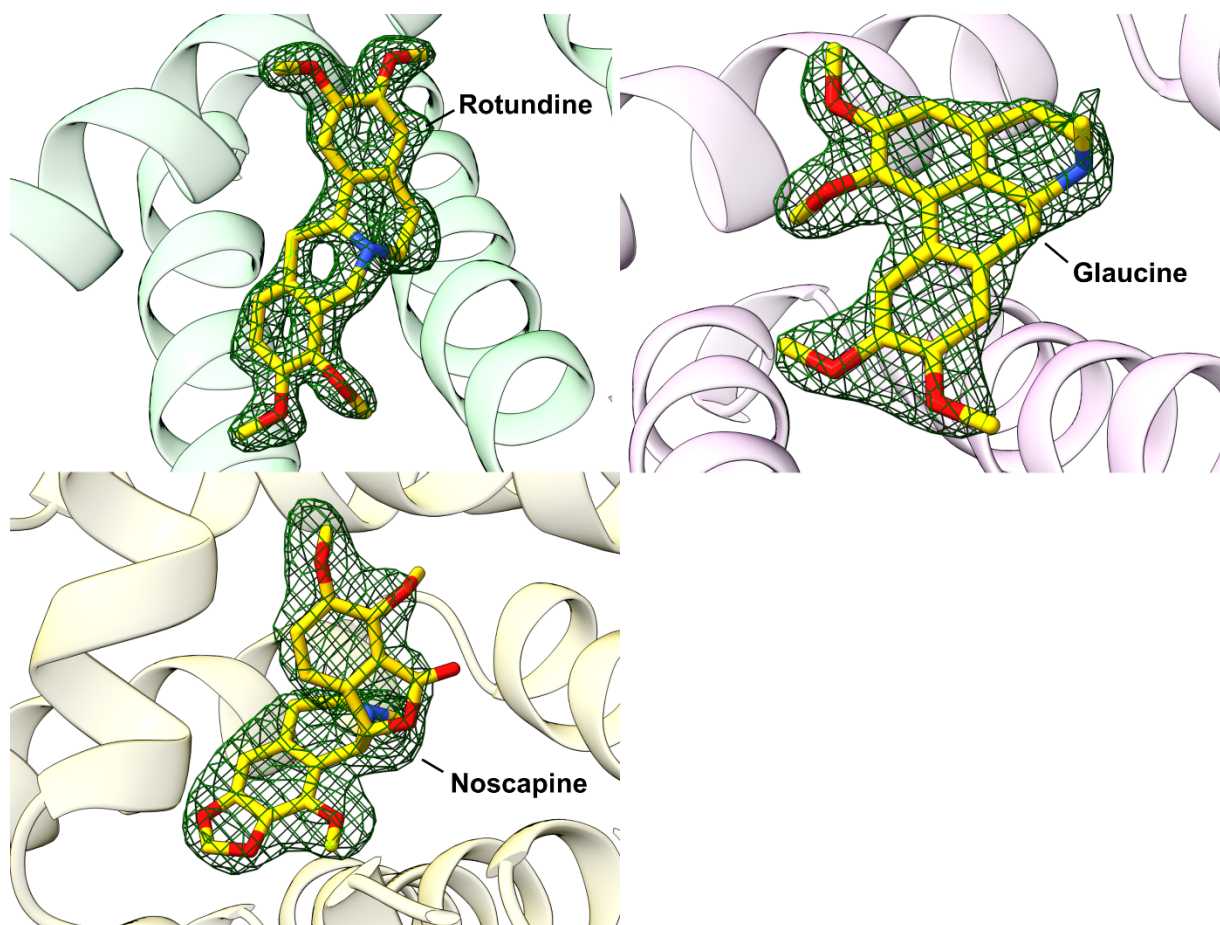


Supplementary Figure 9. Performance of all top THP variants recovered.
See Supplementary Figure 5 legend for details.

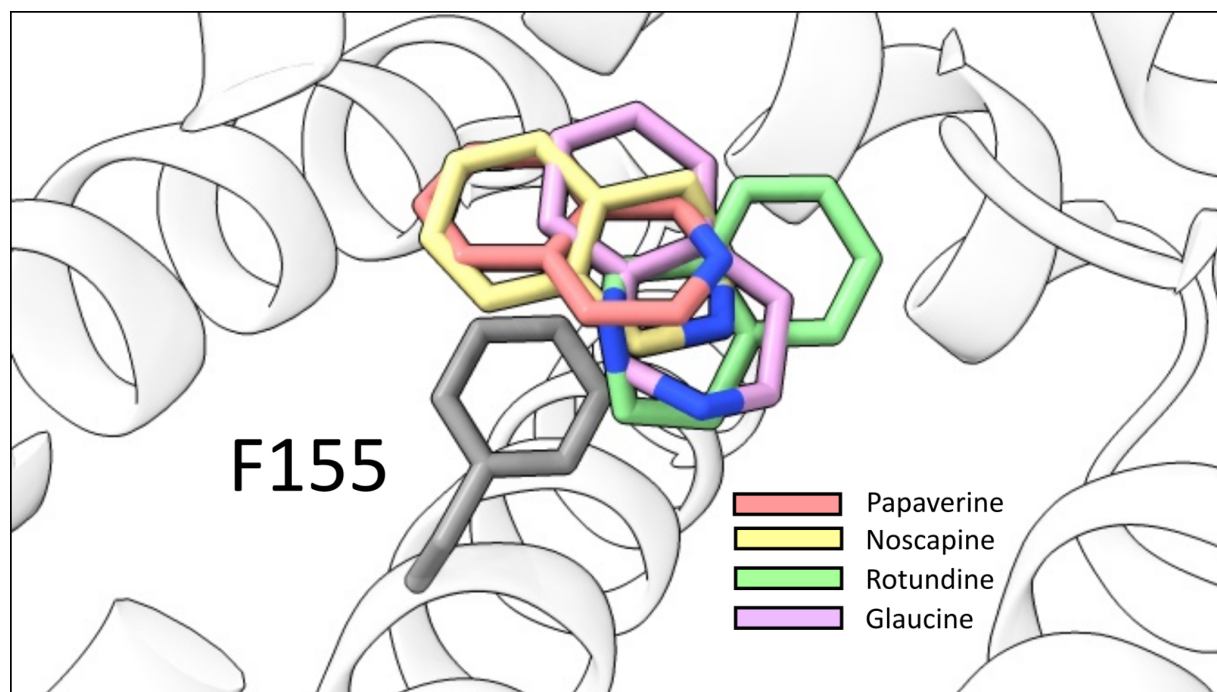


Supplementary Figure 10. Orthogonality of all final RamR variants.

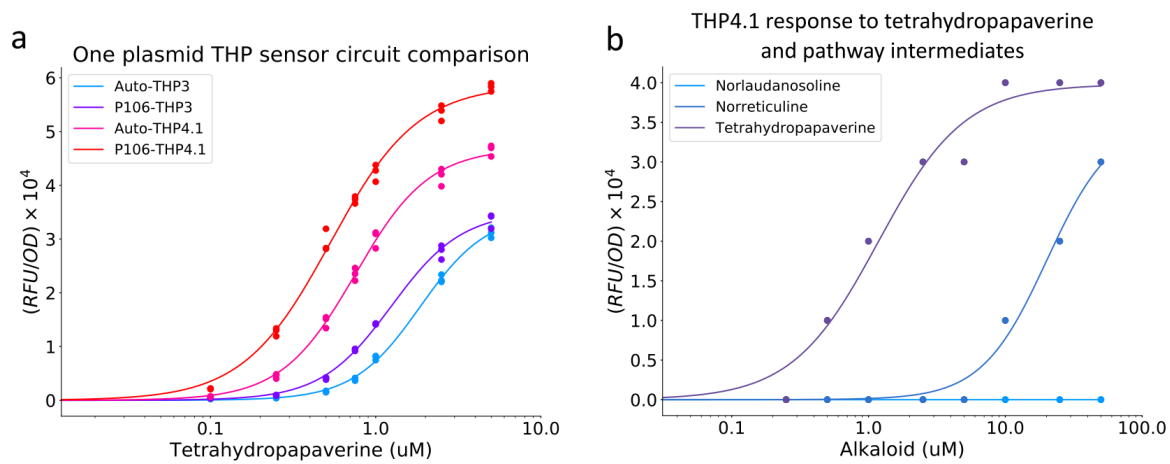
Fluorescent response of cells expressing pGFP and pReg with WT RamR (**a**), Gen1 variants (**b**), Gen2 variants (**c**), Gen3 variants (**d**), and Gen4 variants (**e**) that were induced with 100 μ M of each BIA, separately. Measurements were performed in biological triplicate. See “Methods - Orthogonality Assays” for the list of promoters used to express each variant.



Supplementary Figure 11. Electron density omit maps shown for each BIAs in sticks occupying the binding site of each RamR variant (ROTU4 - green, GLAU4 - purple, NOS4 - pale yellow). Fo-Fc maps (contoured at 3.0σ for Rotundine / Glaucine, and 2.5σ for Noscapine) are shown as forest green mesh superimposed on the model of each BIA.

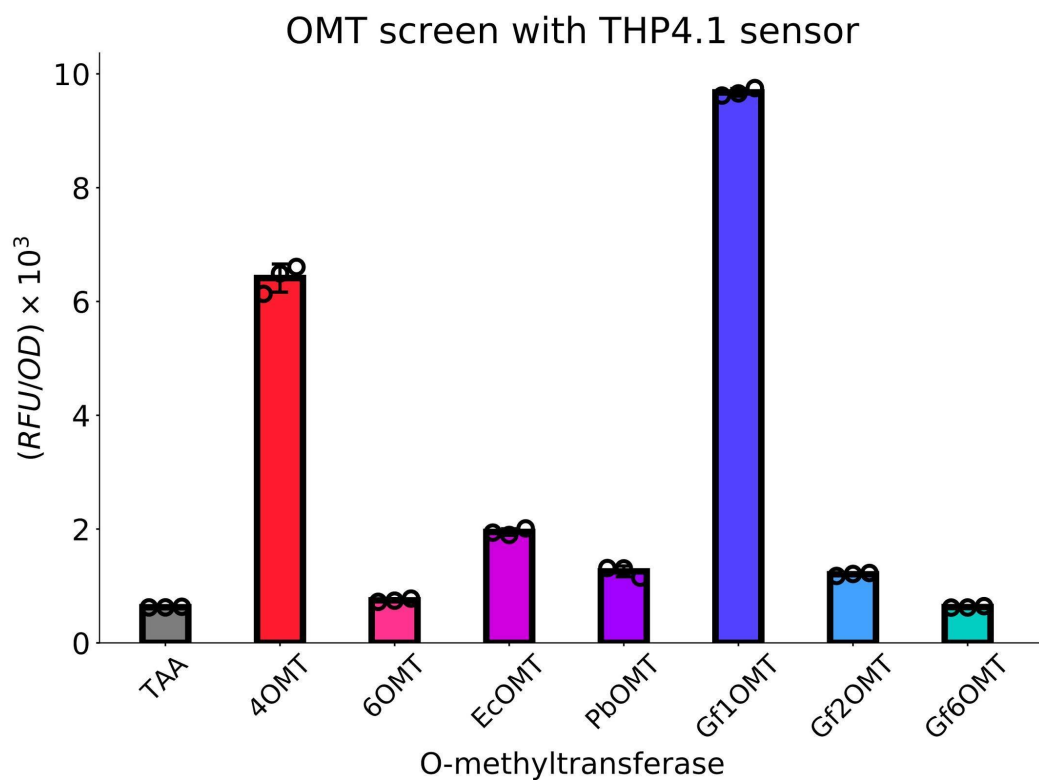


Supplementary Figure 12. Superimposed BIA structures from all BIA-specific RamR variants. The isoquinoline moiety of glaucine, papaverine, noscapine, and rotundine are shown in the context of the RamR ligand binding cavity relative to the F155 residue.



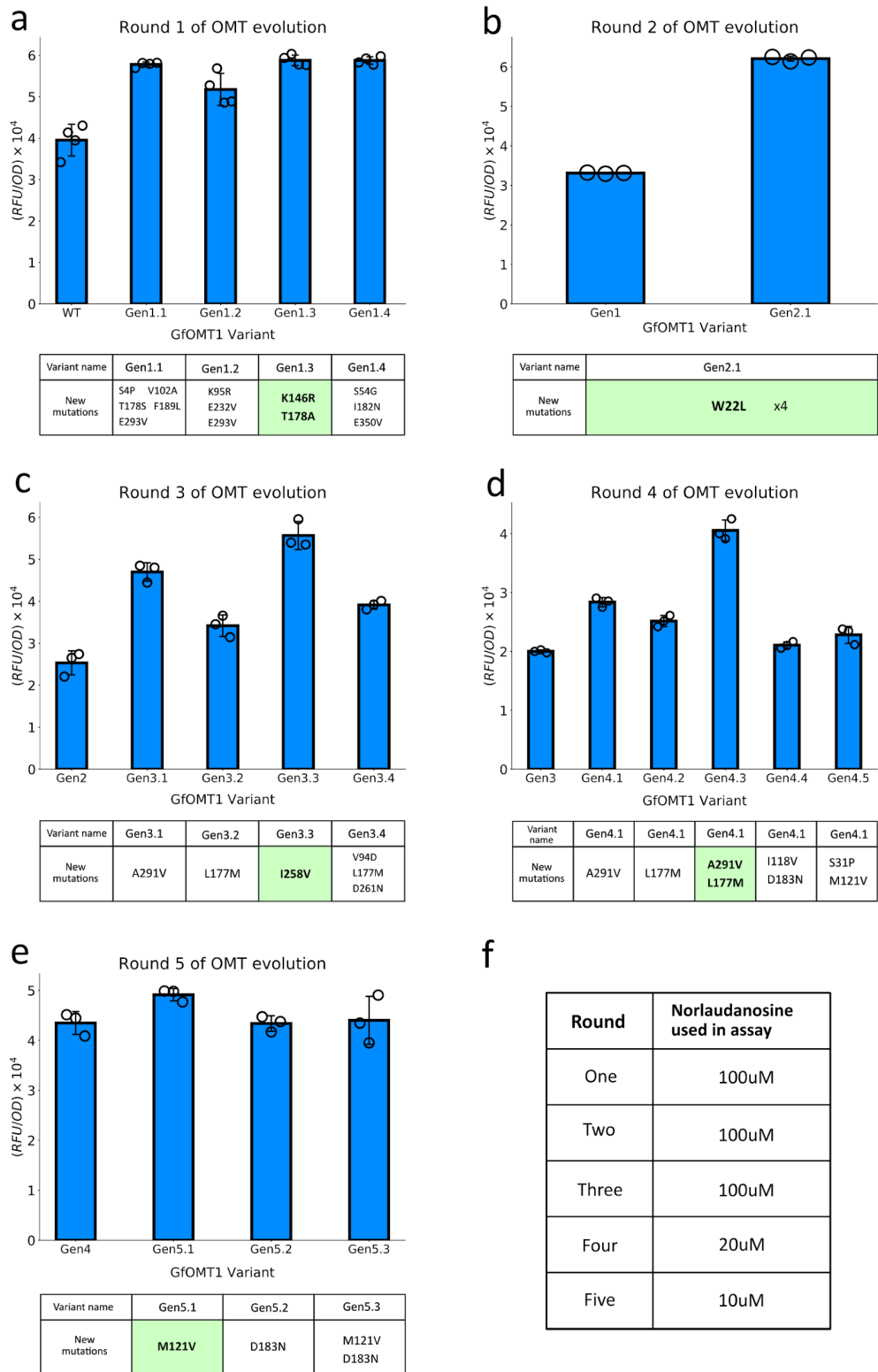
Supplemental Figure 13. Characterization of the THP reporter plasmid (pThpR).

(a) Dose response function of pThpR variants with different THP sensor variants and regulation types. “Auto” denotes that the THP sensor regulates its own expression whereas “P106” denotes that the THP sensor is constitutively expressed from the P106 Anderson promoter. Both THP3 and THP4.1 were compared. The P106-THP4.1 variant of pThpR was used for subsequent fluorescence-based THP measurement assays. (b) The dose response of pThpR (P106-THP4.1) to norlaudanosoline, norreticuline, and tetrahydropapaverine.



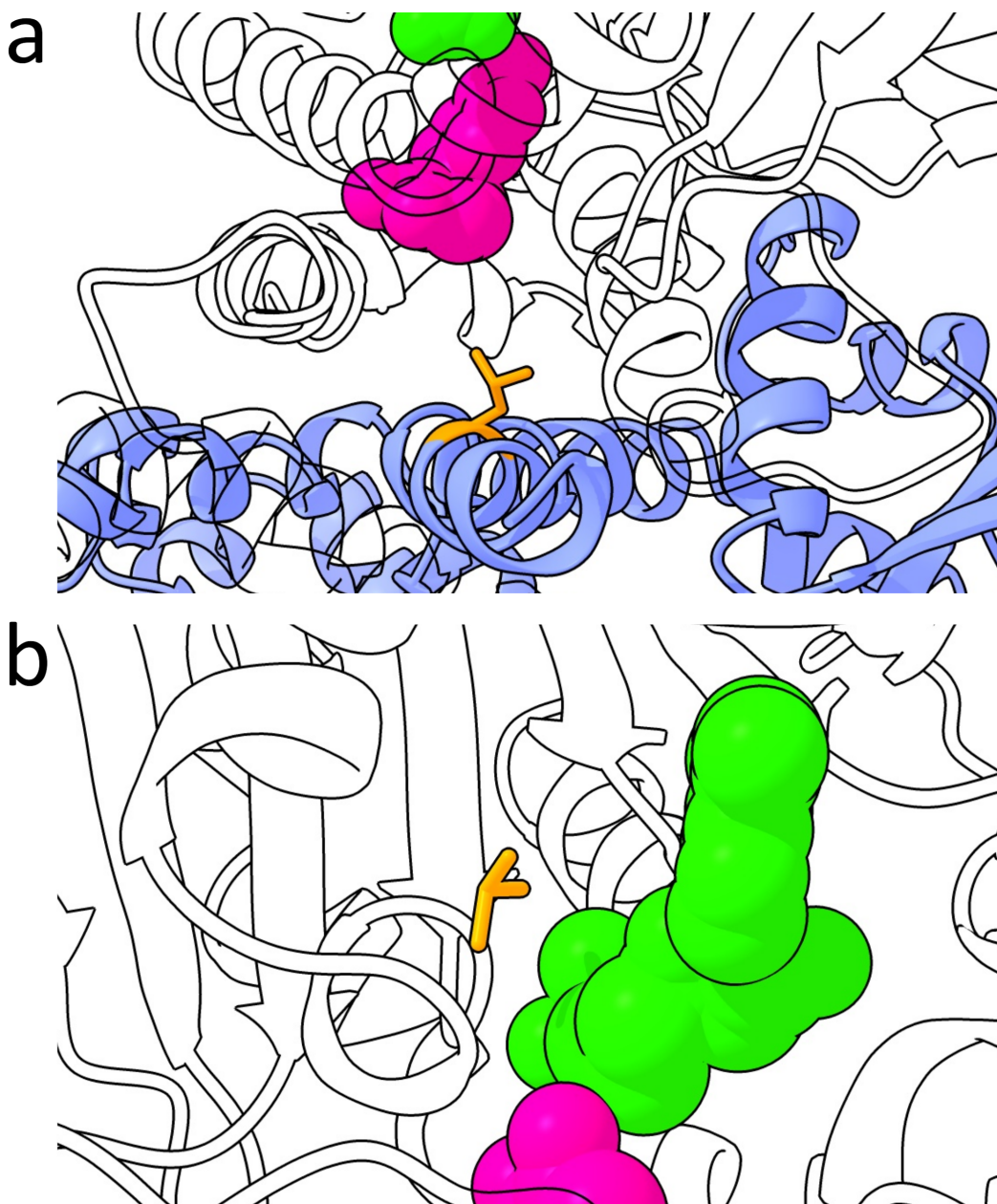
Supplementary Figure 14. Screening O-methyltransferases using pThpR.

Cells expressing either an empty plasmid (TAA) or a BIA methyltransferase were co-transformed with pThpR and grown in the presence of 100 μ M of norlaudanosoline and 1 mg/mL ascorbic acid for 18 hours at 30C and culture fluorescence was subsequently measured. Measurements were performed in biological triplicate.



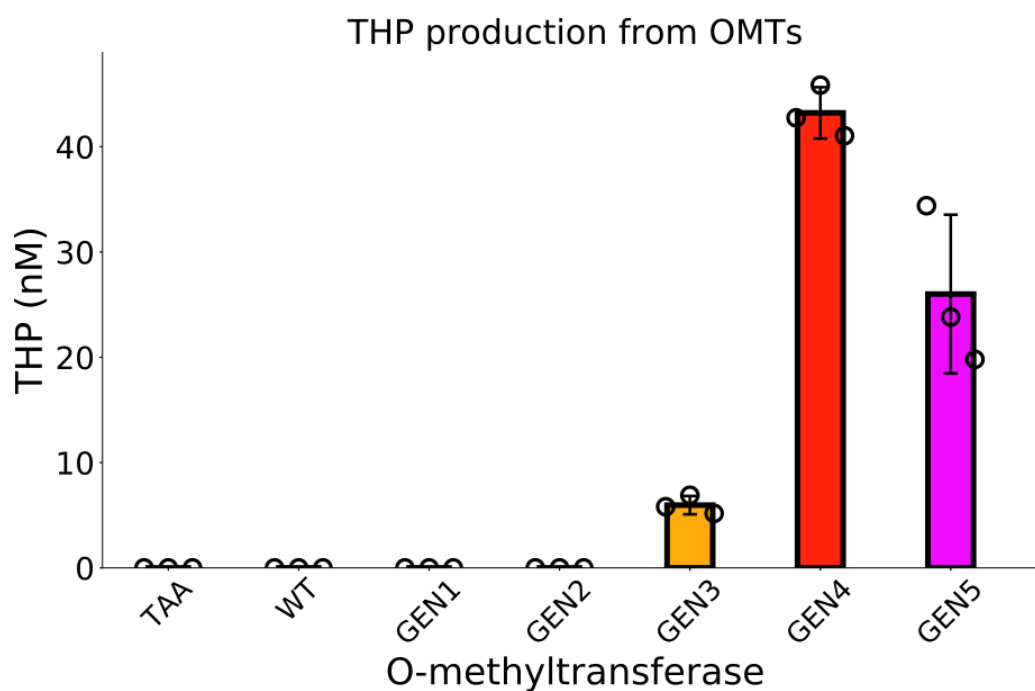
Supplemental Figure 15. Performance of all top OMT variants recovered from each round of evolution.

(a-e) Fluorescent response of top unique GLAU variants using the pThpR reporter. All variants were subcloned into a fresh pReg backbone prior to characterization with the pThpR plasmid. The “x4” symbol denotes that this amino acid sequence was recovered four times following evolution. The variant genotype highlighted in green was chosen as the template for the following round of evolution. Measurements were performed in biological triplicate or quadruplicate. (f) Concentration of the norlaudanosoline substrate used to characterize the performance of evolved OMTs in (a-e).



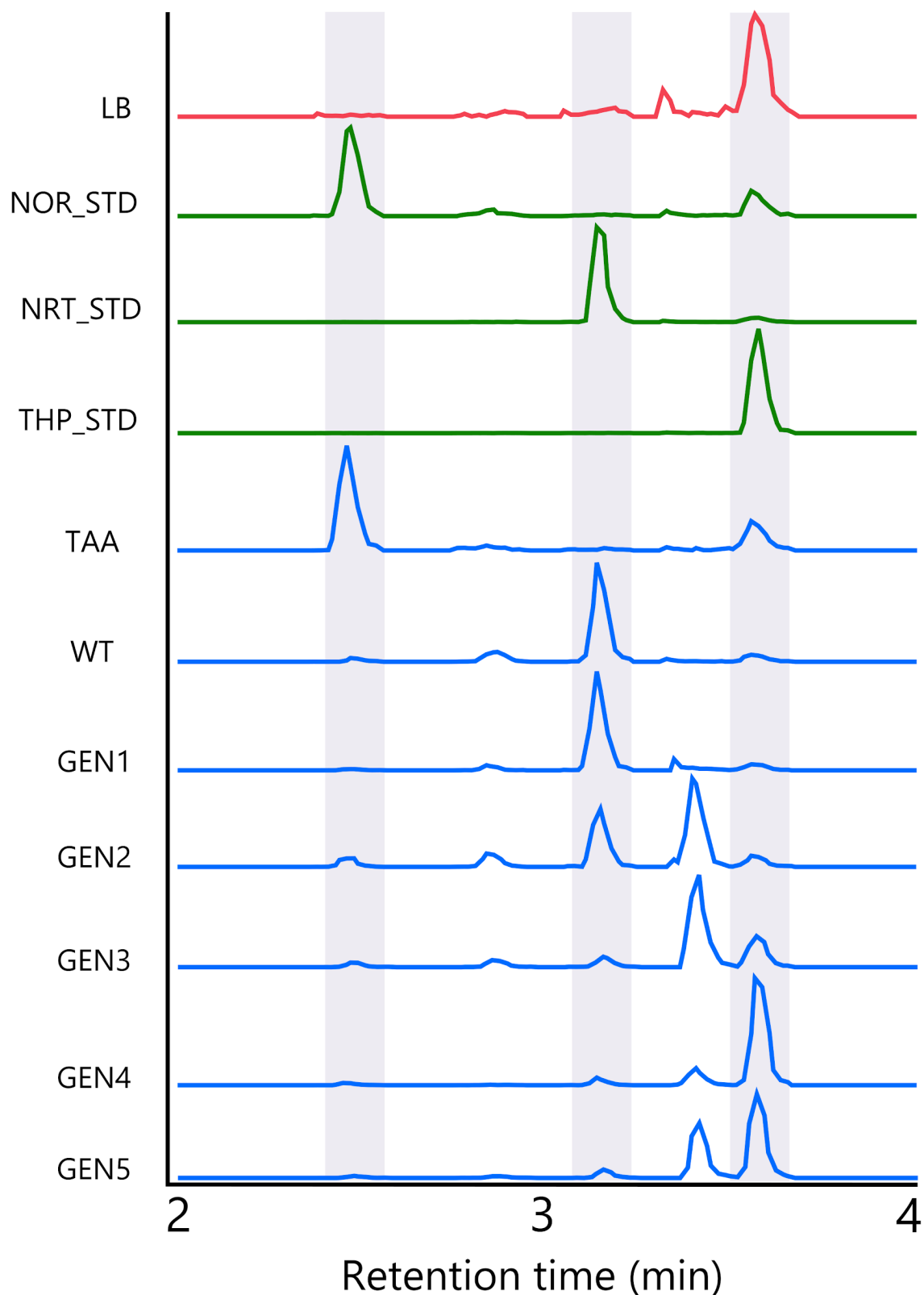
Supplemental Figure 16. Local environment of W22L and I258V OMT mutations.

A homology structure of the GEN5 OMT was constructed using SwissModel to infer the local environment of enzyme mutations. The substrate norlaudanosoline is shown in pink, the co-factor S-adenosyl-methionine is shown in green and mutations are shown in orange. One monomer is transparent while the other monomer is colored blue. **(a)** Environment of the W22L mutation that first appears in the GEN2 OMT variant. **(b)** Environment of the I258V mutations that first appears in the GEN3 OMT variant.



Supplemental Figure 17. Quantification of THP produced by each OMT variant.

Cells expressing an empty plasmid (TAA) or an OMT variant were cultured in the presence of 10 μ M NOR and 1 mg/mL ascorbic acid for 18 hours at 30 C and THP was quantified, after a 1:100 fold dilution, with LC/MS by fitting to a standard curve ($R^2 = 0.9999$). Measurements were made with filtered cell supernatant and were performed in biological triplicate.



Supplemental Figure 18. Representative ion extracted chromatograms of samples and standards. Standards are labelled as “X_STD” and samples are from strains expressing either an empty plasmid (TAA) or enzyme variants (WT, GEN1, GEN2, GEN3, GEN4, GEN5). All LC-MS chromatograms were selected for the theoretical m/z ratios of norlaudanosoline (NOR), 6-O-Methylnorlaudanosoline, norreticuline (NRT), norlaudanine, and tetrahydropapaverine (THP) (See **Supplementary Table 3** below). Grey bars provide a reference for the expected elution times of NOR, NRT, and THP. The LB control sample displays a significant background peak at the expected retention time for THP. This peak may be caused by the tripeptides Pro-Glu-Val or Asp-Leu-Pro, which have a similar exact mass compared to THP

Template → Sensor variant	Alkaloid used (uM)	Promoter expressing template	Libraries used
WT → GLAU1	100	P103-RBS(elvj) 	Y, C, G
GLAU1 → GLAU2	25	P103-RBS(elvj) 	Y, G, E
GLAU2 → GLAU3	5	P103-RBS(elvj) 	P, B, E
GLAU3 → GLAU4	1	P114-RBS5 	P, G, E
WT → NOS1	100	P103-RBS(elvj) 	Y, C, G
NOS1 → NOS2	50	P103-RBS(elvj) 	Y, G, E
NOS2 → NOS3	25	P106-RBS5 	P, Y, B, E
NOS3 → NOS4	2.5	P114-RBS5 	Y, B, E
WT → PAP1	200	P103-RBS(elvj) 	Y, C, G
PAP1 → PAP2	100	P103-RBS(elvj) 	C, G, E
PAP2 → PAP3	25	P103-RBS(elvj) 	P, B, E
PAP3 → PAP4	2.5	P103-RBS(elvj) 	P, G, E
WT → ROTU1	100	P103-RBS(elvj) 	Y, C, G
ROTU1 → ROTU2	25	P103-RBS(elvj) 	Y, E, G
ROTU2 → ROTU3	10	P106-RBS5 	P, Y, B, E
ROTU3 → ROTU4	2.5	P114-RBS5 	P, B, E
WT → THP1	50	P103-RBS(elvj) 	Y, C, G
THP1 → THP2	5	P103-RBS(elvj) 	C, G, E
THP3 → THP3	2	P106-RBS5 	P, B, E
THP3 → THP4	1	P114-RBS5 	P, G, E

Supplementary Table 1. Key parameters of each round of RamR evolution.

Each row indicates the round of evolution (“Template → Sensor variant”), the amount of the target BIA applied to the LB agar plate for screening (“Alkaloid used (uM)”), the promoter/RBS used to express the RamR variant template undergoing evolution (“promoter expressing template”), and the libraries used to introduce diversity (“Libraries used”). For column three, the colored box represents the relative expression level with red being strongest, orange being medium, and yellow being the weakest. For column four, letter codes represent the following (Y= yellow = T85, I88, Y92. C= cyan = K63, L66, M70. P= purple = E120, A123, D124. B = blue = L133, C134, S137. G = grey = R148, D152, L156. E = random mutagenesis)

	PAP4	ROTU4	NOS4	GLAU4
Data collection				
Space group	C2	P1	P1	P1
Cell dimensions				
a, b, c (Å)	106.76, 68.57, 69.57	46.14, 50.84, 50.83	41.63, 54.86, 92.64	43.18, 54.20, 91.52
α, β, γ (°)	90.00, 127.53, 90.00	120.05, 90.17, 89.96	74.14, 81.82, 89.96	104.90, 98.00, 89.99
Resolution (Å)	50.00-1.6 (1.63-1.60)*	50.00-1.73 (1.76-1.73)	50.00-2.21 (2.25-2.21)	50.00 - 2.00 (2.03-2.00)
$R_{\text{sym}}/R_{\text{pim}}$	0.055(0.474)/0.034(0.306)	0.060(0.285)/0.056(0.263)	0.072(0.354)/0.068(0.341)	0.061(0.320)/0.056(0.300)
CC $\frac{1}{2}$ γ	0.947 (0.751)	0.936 (0.830)	0.921 (0.76)	0.943 (0.813)
I / σ	21.3 (1.7)	16.6 (1.9)	12.1 (1.6)	14.7 (1.7)
Completeness (%)	99.8 (99.5)	96.0 (95.1)	95.0 (96.2)	94.3 (84.2)
Redundancy	3.5 (3.2)	1.9 (1.8)	1.8 (1.8)	1.8 (1.5)
Refinement				
Resolution (Å)	48.82-1.60 (1.66-1.60)	46.14-1.74 (1.80-1.74)	44.07 - 2.21 (2.29 - 2.21)	43.76 - 2.00 (2.07 - 2.00)
No. reflections	52242 (5094)	39481 (3649)	36872 (3591)	50140 (4375)
R_{work}	0.1868 (0.2380)	0.2162 (0.2855)	0.2418 (0.2837)	0.2022 (0.2552)
$R_{\text{free}}^{\pm}$	0.2108 (0.2635)	0.2577 (0.2714)	0.2849 (0.3413)	0.2334 (0.2938)
No. atoms	3239	3140	5950	6345
Protein	2935	2842	5657	5865
Ligand/ion	60	62	140	124
Water	244	236	153	356
B-factors (Å²)				
Protein	25.4	23.3	51.7	42.9
Ligand/ion	16.8	21.2	53.4	31.7
Water	38.6	31.7	48.4	45.1
R.m.s. deviations				

Bond lengths (Å)	0.006	0.001	0.01	0.008
Bond angles (°)	0.72	0.48	0.84	0.75
Ramachandra n plot				
Favored	99.45%	99.14%	99.15%	99.31%
Allowed	0.55%	0.86%	0.85%	0.69%
Outliers	0.00%	0.00%	0.00%	0.00%
Molprobity score[^]	1.22 / 98th percentile	1.50 / 92nd percentile	1.81 / 93rd percentile	1.53 / 96th percentile

*Values for the corresponding parameters in the outermost shell in parenthesis.

^γCC_{1/2} is the Pearson correlation coefficient for a random half of the data, the two numbers represent the lowest and highest resolution shell, respectively.

[±]R_{free} is the R_{work} calculated for about 10% of the reflections randomly selected and omitted from refinement.

[^]MolProbity score is calculated by combining clashscore with rotamer and Ramachandran percentage and scaled based on X-ray resolution. The percentage is calculated with 100th percentile as the best and 0th percentile as the worst among structures of comparable resolution.

Supplementary Table 2. X-ray Crystallography Data Collection and Refinement Statistics

Sample	Observed EIC Areas				
	NOR-4OH (288.1230 m/z)	3OH (302.1387 m/z)	NRT-2OH (316.1543 m/z)	1OH (330.1700 m/z)	THP-0OH (344.1856 m/z)
Retention Time (min.)	2.46	2.85	3.14	3.41	3.58
1(TAA_1)	1,900,714.32	0.00	0.00	0.00	732,062.09
2(TAA_2)	1,915,746.57	0.00	0.00	0.00	739,719.41
3(TAA_3)	2,429,528.49	0.00	0.00	0.00	683,951.95
4(WT_1)	292,886.91	838,007.19	5,016,094.31	0.00	715,848.03
5(WT_2)	183,019.07	678,233.01	4,455,448.19	0.00	726,021.76
6(WT_3)	553,463.54	956,624.78	5,512,157.20	0.00	724,935.11
7(Gen1_1)	127,968.53	262,284.74	5,028,466.49	0.00	662,206.48
8(Gen1_2)	189,597.68	357,195.05	5,328,381.15	0.00	685,462.09
9(Gen1_3)	69,624.59	194,528.60	4,578,048.19	0.00	627,833.84
10(Gen2_1)	641,245.71	730,076.27	2,721,068.90	4,013,082.04	768,144.56
11(Gen2_2)	251,793.95	485,465.06	1,499,029.94	4,748,195.68	825,357.07
12(Gen2_3)	457,481.05	582,826.88	2,227,163.03	4,367,581.54	796,201.94
13(Gen3_1)	304,453.82	474,632.34	603,324.03	6,336,545.51	2,220,655.54
14(Gen3_2)	259,191.32	327,542.61	565,407.61	6,114,959.98	2,469,776.82
15(Gen3_3)	388,680.89	444,919.64	622,087.57	6,641,956.94	2,075,163.88
16(Gen4_1)	242,876.32	0.00	639,571.41	1,994,412.63	10,694,641.10
17(Gen4_2)	139,427.97	0.00	488,824.56	1,353,523.79	11,406,507.42
18(Gen4_3)	249,690.78	0.00	684,700.15	2,054,059.80	10,304,271.54
19(Gen5_1)	33,523.35	55,140.14	377,855.13	3,057,004.77	8,779,382.47
20(Gen5_2)	159,548.08	235,055.43	565,294.52	4,882,357.32	6,352,723.48
21(Gen5_3)	385,426.52	238,451.26	608,492.27	5,854,903.52	5,431,738.76
22(NOR_50nM_STD)	1,674,119.68	0.00	0.00	0.00	650,085.33
23(NRT_50nM_STD)	0.00	0.00	6,704,519.34	0.00	659,703.80
24(THP_10nM_STD)	0.00	0.00	0.00	0.00	3,178,252.46
25(THP_50nM_STD)	0.00	0.00	0.00	0.00	11,969,036.43

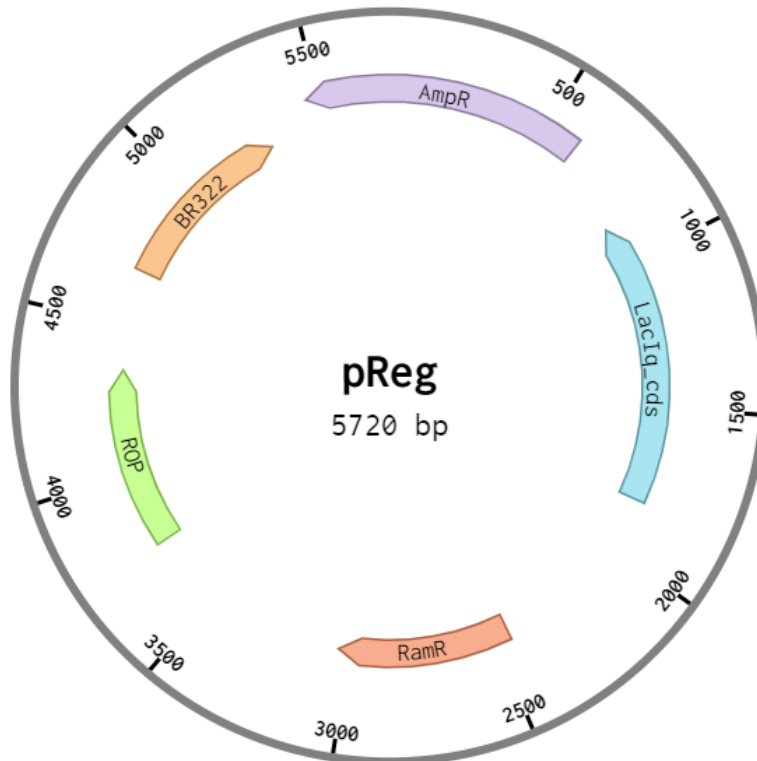
*THP overlaps with matrix peaks, so area never actually goes to zero

Supplemental Table 3. Observed extracted ion chromatograms areas for all controls, standards, and enzymatic reactions. The top two rows indicate the m/z ratio and retention time for compounds of interest (norlaudanosoline (NOR-4OH), 6-O-Methylnorlaudanosoline (3OH), norreticuline (NRT-2OH), norlaudanine (1OH), and tetrahydropapaverine (THP-0OH). All sample measurements were performed in biological triplicate.

> pReg

GAAGCTAGAGTAAGTAGTTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTGCAGGCATCG
TGGTGTACGCTCGTCGTTTGGTATGGCTTCATTACGCTCCGGTCCCAACGATCAAGGCGAGTTAC
ATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCCTCCGATCGTTGTGAGAAGTAAG
TTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGT
AAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCG
AGTTGCTCTTGCCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCA
TCATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTTCGAT
GTAACCCACTCGTGACCCAACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCA
AAAACAGGAAGGCAAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCAT
ACTCTTCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGA
ATGTATTTAGAAAAATATGCGCCTTGAGCGACACGAATTATGCAGTGATTACGACCTGCACAGCCAT
ACCACAGCTTCCGATGGCTGCCTGACGCCAGAAGCATTGGTGCACCGTGCAGTCGATGATAAGCTG
TCAAACATGAGAATTGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCCG
TTTCCAGTCGGGAAACCTGTCTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG
TTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTACCAGTGAGACGGGCAACAGCTGATTGCCCTT
CACCGCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAAT
CCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTATCTTCGGTATCGTCGTATCCCCTACC
GAGATATCCGCACCAACGCGCAGCCCCGACTCGGTAATGGCGCGCATTGCGCCAGCGCCATCTGA
TCGTTGGCAACCAGCATCGCAGTGCGGAACGATGCCCTCATTACGATTTGCATGGTTTGTGAAAAC
CGGACATGGCACTCCAGTCGCCTTCCCGTTCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTA
TGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCCGCTAACAGCGCGATTGCG
TGGTGACCCAATGCGACCAAGATGCTCCACGCCAGTCGCGTACCATCTTCATGGGAGAAAATAATAC
TGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGACAGGCAGCTTCCA
CAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAG
ATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCA
CCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTG
GAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATG
TAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTTCGAGAAACGTGGCTGGCCTGGT
TCACCACGCGGGAACCGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGG
TTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGC
ACCATTGATGGTGTGCGGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTG
TTGCGGCCGCGAAGACAGCGTTATCAGAGATGAGACACCGTGCGATAATGTGCGGCAATCAGGTGC
GACTCGGTACCAAATTCAGAAAAGAGGGGAGCGGGAAACCGCTCCCCCTTTTTTCGTTTTTGGTCCC
AATCTATCGATTGTATGGACTTCATCTTCATTACCTCGTATCATTGTACACCTGCCGAACGCAAGGGC
ATGGGCTGTGACCTTTGAAAAGTACCCTGATAGCTAGCTCAGTCCTAGGGATTATGCTAGCAATTAC
GAGCCCCATAGGGTGGTGTGTACCACCCCTGATGAGTCCAAAAGGACGAAATGGGGCCTCTACAAA
TAATTTTGTTTAACGGAACCACGTATCAGAAGGAGGTTAGTATATGGTTGCTCGCCCAAAGTCTGAG
GACAAAAGCAGGCATTGCTTGAAGCGGCAACTCAAGCCATCGCGCAATCAGGCATTGCCGCTAGT
ACCGCTGTAATTGCACGCAATGCGGGAGTTGCGGAAGGGACGTTGTTCCGCTATTTTCGCAACGAAA
GATGAGTTGATCAACACCCTTTACTTACATTTGAAACAGGACCTGTGCCAATCAATGATCATGGAATT
GGATCGTTCTATTACTGACGCTAAGATGATGACCCGTTTTATCTGGAACAGTTATATTAGCTGGGGATT
GAACCACCCAGCTCGCCATCGTGCCATTCTGTCAGTTGGCGGTTTTCTGAAAAGTTGACGAAGGAAAC
CGAACAACGCGCGGATGATATGTTCCCGGAGTTACGCGACTTGTGCCACCGTAGTGTTCTTATGGTG
TTTATGTCCGACGAGTACCGCGCCTTCGGCGACGGGTTGTTCTTGCGGCTTGCTGAGACGACTATGG
ATTTGCTGCGCGCGACCCGGCTCGCGCTGGTGAGTACATTGCGTTGGGCTTCGAGGCTATGTGGC
GCGCACTTACGCGCGAAGAGCAGTAAATATCCTAAGAATTGAGGAGTGACAGGCTCGGTAACATAC
GGTCTAGCTATCTGACTATCGCCGCTGTGAGCTCGGTACCAAATTCAGAAAAGAGGCCGCGAAAG
CGGCCTTTTTTCGTTTTTGGTCCAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTATC
ACTAGTCTTCGCGGCCGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG
ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGATTTATGCCG
CCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCC
CGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGC
TAACGGATTACCACTCCAAGAATTGGAGCCAATCAATTCTTTCGCGGAGAACTGTGAATGCGCAAAC
CAACCCTTGGCAGAACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGG
CAGCGTTGGGTCTTGCCACGGGTGCGCATGATCGTGCTCCTGTGTTGAGGACCCGGCTAGGCTG
GCGGGGTTGCCTTACTGGTTAGCAGAAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTG
CTGCAAAA

GGAAACGCGGAAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCT
 ACCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTCTCTGG
 TCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCATGTTTCATCATC
 AGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACCCCATGAACAGAAATCCCC
 CTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCCCTTAACATGGCCCGCTTTATCAGAA
 GCCAGACATTAACGCTTCTGGAGAACTCAACGAGCTGGACGCGGATGAACAGGCAGACATCTGTG
 AATCGCTTCACGACCACGCTGATGAGCTTTACCGCAGCTGCCTCGCGCGTTTCGGTGATGACGGTG
 AAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCA
 GACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCCAGTCA
 CGTAGCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCA
 CCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAAATACCGCATCAGGCGCTCTTCCGCTTCCTC
 GCTCACTGACTCGCTGCGCTCGGTTCGGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGT
 AATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAAGGCCAGCAAA
 AGGCCAGGAACCGTAAAAAGGC CGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGC
 ATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAACAGGCGT
 TTCCCCCTGGAAGCTCCCTCGTGCGCTCTCTGTTCGGACCCCTGCCGCTTACCGGATACCTGTCCGC
 CTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTA
 GTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTAGCCCCGACCGCTGCGCCTTATCC
 GGTAAGTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA
 ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACG
 GCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGT
 TGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCA
 GATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG
 TGGAACGAAAACCTCACGTAAAGGGATTTTGGTTCATGAGATTATCAAAAAGGATCTTCACCTAGATCC
 TTTTAAATTAATAATGAAGTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTAC
 CAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACT
 CCCCCTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
 CGCGATCCACGCTCACC GGCTCCAGATTTATCAGCAATAAACAGCCAGCCGGAAGGGCCGAGCGC
 AGAAGTGGTCTGCAACTTTATCCGCTCCATCCAGTCTATTAATTGTTGCCGG



> pGFP

CGAGCATACTATCACGTTCGGCGACCACTAGTCAGTTAACGCAAGGGCATGGGCTGTGACCTTTGAA

AAGTACCTTGACGGCGTATCTTTGCTTTCTATAATGAGTGCTTACTCACTCATACAATAGTCAGTCATA
AGTCTGGGCTAAGCCCACTGATGAGTCGCTGAAATGCGACGAAACTTATGACCTCTACAAATAATTT
TGTTTAAACGTAAACCTCCGGGTAAATAAGGAGTAATTATGGCATCCAAGGGCGAGGAGCTCTTTACT
GGCGTAGTACCAATTCCTGATAGAGCTCGATGGCGATGTAAATGGCCATAAGTTTTCCGTACGCGGCG
AGGGCGAGGGCGATGCAACTAACGGCAAGCTCACTCTCAAGTTTATTTGTACTACTGGCAAGCTCCC
AGTACCATGGCCAACTCTCGTAATACTCTGACCTATGGCGTACAATGTTTTCCCGCTATCCAGATC
ACATGAAGCAACATGATTTTTTTAAGTCCGCAATGCCAGAGGGCTATGTACAAGAGCGCACTATTAG
CTTTAAGGATGATGGCACCTATAAGACTCGCGCAGAGGTAAAGTTTGAGGGCGATACTCTCGTAAAT
CGCATTGAGCTCAAGGGCATTGATTTTAAGGAGGATGGCAATATTCTCGGCCATAAGCTGGAGTATA
ATTTCAATTCCCATAAATGTATATATTACCGCAGATAAGCAAAAAGAATGGCATTAAAGGCGAATTTAAG
ATTCGCCATAATGTGGAGGATGGCTCCGTACAACCTCGCAGATCATTATCAACAAAATACTCCAATTGG
CGATGGCCCACTACTCCTCCCAGATAATCATTATCTCTCCACTCAATCCGTGCTCTCCAAAGATCCAA
ATGAGAAGCGCGATCACATGGTACTCCTGGAGTTTGTAAGTGCAGCAGGCATTACTCATGGCATGGA
TGAGCTCTATAAGCTCGAGCACCACCACCACCACCCTGATAATCCAAACCTGTTATATGTTAGCTGA
GACTAGTTGGAAGTGTGGCTGTCTCAAGCGTTTTAGTTTCGTGCGTCAGTTTCACCTGATTTACGTA
AAAACCCGCTTCGGCGGGTTTTTGCTTTTGGAGGGGCGAGAAAGATGAATGACTGTGCGCCATTCTGA
TGGTGTGCGGTAGCATAACCCCTTGTGATAGTCTTCGCGGCCGCTCACACTGCTTCCGGTAGTCAAT
AAACCGGTAAACCAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGG
TCGAATTTGCTTTTCAATTTCTGCCATTATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGC
GTTTAAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTT
GTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAG
CGGCATCAGCACCTTGTGCGCTTGGCGTATAATTTTGCCCATAGTGAAAACGGGGGCGAAGAAGTTG
TCCATATTGGCCACGTTTAAATCAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAAC
ATATTCTCAATAAACCCCTTAGGGAAATAGGCCAGGTTTTACCCGTAACACGCCACATCTTGCGAATA
TATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGC
TCATGGAACCGGTGTAACAAGGGTGAACACTATCCCATATCACAGCTCACCGTCTTTCATTGCCA
TACGGAACCTCCGGATGAGCATTATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAACTTGT
GCTTATTTTTCTTTACGGTCTTTAAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT
GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATAT
CCAGTGATTTTTTTCTCCATTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCC
CGGTAGTGATCTTATTTTATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCACGTCTCATTTTCG
CCAAAGTTGGCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATNNTGCGAAGTGATC
TTCCGTACAGGTATTTATTTCGGCGCAAAGTGCGTCGGGTGATGCTGCCAACTTACTGATTTAGTGTA
TGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAGCTGTCCCTCCTGTTACGCTACTGA
CGGGTGGTGCGTAACGGCAAAAGCACCGCCGGACATCAGCGCTAGCGGAGTGATACTGGCTTAC
TATGTTGGCACTGATGAGGGTGTGAGTGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGG
TGCGTCAGCAGAATATGTGATACAGGATATATCCGCTTCTCGCTCACTGACTCGCTACGCTCGGTC
GTTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCTGGAAGATGCCAGGAA
GATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAAGCCGTTTTTCCATAGGCTCCGCCCCCTGAC
AAGCATCACGAAATCTGACGCTCAAATCAGTGGTGCGGAAACCCGACAGGACTATAAAGATACCAG
GCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCTCGCTTTCGGTTTACCGGTGTCATT
CGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTTCGCTCCA
AGCTGGACTGTATGCACGAACCCCCGTTTTCAGTCCGACCGCTGCGCCTTATCCGGTAAGTATCGTCT
TGAGTCCAACCCGGAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGG
AGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTC
CTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCA
AGGCGGTTTTTTTCGTTTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCT
TATTAATCAGATAAAATATTTCTAGATTTTCAAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAG
CCCCATACGATATAAGTTGTAATTTCTCATGTTAGTCATGCCCCGCGCCACCGGAAGGAGCTGACTG
GGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAAT
TGCGTTGCGCTCACTGCCCCGCTTTCCAGTCCGGGAAACCTGTGCTGCCAGCTGCATTAATGAATCGGC
CAACGCGCGGGGAGAGGGCGGTTTGGCTATTGGGCGCCAGGGTGGTTTTTCTTTTACCAGTGAGAC
GGGCAACAGCTGATTGCCCTTACCAGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGT
TTGCCCCAGCAGGCGAAAAATCCTGTTTGTGGTGGTTAACGGCGGGATATAACATGAGCTATCTTCG
GTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATAGCGCGCA
TTGCGCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTACAGCAT
TTGCATGGTTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCGCTATCGGCTGAATT
TGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCGAGACGCGCCGAGACAGAACTTAATGGG
CCCCGTAACAGCGCGATTGCTGGTGACCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCAT

CTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAA
 CATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCC
 ACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACC
 ATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGAC
 GGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTG
 TTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTC
 GCAGAAACGTGGCTGGCCTGGTTACCCACGCGGGAACGGTCTGATAAGAGACACCGGCATACTCT
 GCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCA
 TGCCATACCGCGAAAGGTTTTGCGCCATTTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGA
 CGCGGCCGCGAAGACACTGTCCTCAATGGTTTCGTTGTGATGGCGGTAGGAATGTAATCGTTAATCCG
 CAAATAACGTAAAAACCCGCTTCGGCGGGTTTTTTTATGGGGGGAGTTTAGGGAAAGAGCATTTGTC
 ATCCCGTTGAATATGGCTCGCATCTTAT

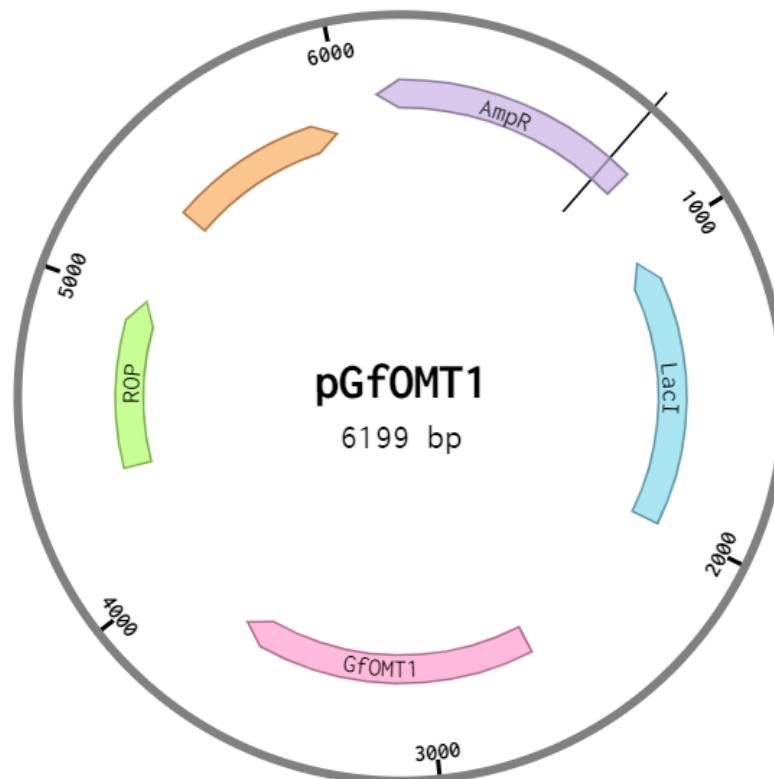


> pGfOMT1

TGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGCGATCC
 ACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGG
 TCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTTCGC
 CAGTTAATAGTTTTCGCAACGTTGTTGCCATTGCTGCAGGCATCGTGGTGTACGCTCGTCGTTTGG
 TATGGCTTCATTCAGCTCCGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAA
 AAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTCAGAAAGTAAGTTGGCCGCAGTGTTATCACTCA
 TGGTTATGGCAGCACTGCATAATTCTTTACTGTTCATGCCATCCGTAAGATGCTTTTTCTGTGACTGGT
 GAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAA
 CACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTTGGAACGTTCTTCGGG
 GCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAAC
 TGATCTTCAGCATCTTTTACTTTACCCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCC
 GCAAAAAGGGAATAAGGGCGACACGGAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATT
 GAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATATGCGCC
 TTGAGCGACACGAATTATGCAGTGATTACGACCTGCACAGCCATAACCACAGCTTCCGATGGCTGCC
 TGACGCCAGAAGCATTTGGTGCACCGTGCAGTCGATGATAAGCTGTCAAACATGAGAATTGTGCTTA
 ATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCTG
 TGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCCTATTGGGCGCCAGGGT
 GGTTTTTCTTTTACCCAGTGAGACGGGCAACAGCTGATTGCCCTTACCCGCTGGCCCTGAGAGAG
 TTGCAGCAAGCGGTCCACGCTGGTTTGGCCAGCAGGCGAAAATCCTGTTTGTGATGGTGGTTAACGG
 CGGGATATAACATGAGCTATCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCA

GCCCCGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAG
TGGGAACGATGCCCTCATTAGCATTTGCGATGGTTTGTGTTGAAAACCGGACATGGCACTCCAGTCGCC
TTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGA
CGCGCCGAGACAGAACTTAATGGGCCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGA
TGCTCCACGCCCAGTCGCGTACCATCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAG
AGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCAT
CCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACA
GGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGAT
TTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGC
AACGACTGTTTGGCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCG
CTTCCACTTTTTCCCGCGTTTTTCGCAGAAACGTGGCTGGCTGGTTTACCACGCGGGAAACGGTCT
GATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTACACCCTGAAT
TGACTCTCTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCACCATTGATGGTGTGCGGAC
GTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTGCGGCCGCGAAGACAGCGT
TATCAGAGATGAGACACCGTGCAGATAATGTGCGGCAATCAGGTGCGACTCGGTACCAAATCCAGA
AAAGAGGGGAGCGGGAAACCGCTCCCTTTTTTCGTTTTGGTCCCAATCTATCGATTGTATGGACTT
CATCTTCATTACCTCGTATCATTGTACACCTGCCGAACGCAAGGGCATGGGCTGTGCACCTTTGAAA
AGTACCTTTATGGCTAGCTCAGTCCTAGGTACAATGCTAGCAATCAAGATACTGAGCACAGCTGTCAC
CGGATGTGCTTTCCGGTCTGATGAGTCCGTGAGGACGAAACAGCCTCTACAAATAATTTGTTTAAA
CTAGTGAACCACGAGGCCTACATATGGGAGTTTCAGATAATAAACCAGAAAGCCAAGAGGTTGACAT
CAAAGCGCAGGCCCATCTGTGGAACATCATTACGGGTTTGCAGATTATTAGTATTACGCTGTGCTG
TTGAGATTGGCATTGCAGATATCATTAAAGAGTAATAATGGGAGTATTAGCGTAACGGAACCTTGCTTCT
AAGCTGCCCATCACTAACGTAAATAGCGACAACCTTTACCGCGTACTTCGCTACCTTGTGCATATGGG
AATTTTGAAGGAAGTCAGTGATTCAAATGAAGTCAAACCTGTATTCTTTCAGCCAGTCGCAACTCTG
CTGCTTCGTGATGCTGAGCGTTCCATGGTCCCAATTATCTGGGGATGACCCAGAAGGACTTTATGA
TTCCCTGGCATTTCATGAAGGAAGGTTTGGGGAATGACACAACGGCCTTCGAAAAGGGGATGGGCA
TGACAATCTGGCAATACTTGGAAGGACATCCCGAGCAAAGCAATTTGTTTAAACGAAGGAATGGCTG
GTGAGACTCGCCTTTTGACTAAGTCTTTAATCGATGGCTGTCTGTGATACCTTCGAGGGCCTGACAAG
CCTTTGCGATGTGCGCGGGGGTAATGGGACAACAATTAAGGGCATCTACGACGCATTCCCTCAGATC
AAGTGTTCCGTCTATGATTTGCCACATGTAATCGCTAGTTCCCCCGAGCACCCGAACATCGAACGCA
TTCCAGGTGACATGTTCAAGTCAGTTCCAAGTGCCAGGCCATCTTATTGAAACTGATTTTGCATGA
CTGGACGGACGAAGAATGCGTCAACATCTTGATCAAATGCCGCGAAGCTGTACCCAAGGATACCGG
TAAAGTCATTATTGTGGACGTGGCTCTTGAGGAAGAGAGTCAACATGAGTTGACCAAGACTCGTCTT
ATCCTTGATATTGATATGCTGGTTAATACTGGAGGTGCGGAACGTTCCGAAGATGACTGGGAGAAGC
TGTTAAAGCGTGCCGTTTTTCGTGGGCATAAGATTTCGTACATCGCGGCCATCCAAAGTGTAATCGA
AGCCTTTCCGTAAATATCCTAAGAATTCGAGGAGTGCAGGCTCGGTAACATACGGTCTAGCTATCTG
ACTATCGCCGCTGTGAGCTCGGTACCAAATTCAGAAAAGAGAGCCGCGAAAGCGGCCTTTTTTCGT
TTTGGTCCAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTATCACTAGTCTTCGCGG
CCGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTC
TTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGATTATGCCGCCTCGGCGAGCAC
ATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGC
GGTGCATGGAGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTACCA
CTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGA
ACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCTG
GCCACGGGTGCGCATGATCGTGCTCCTGTGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTAC
TGGTTAGCAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAA
CGTCTGC
GACCTGAGCAACAACATGAATGGTCATCGGTTTTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGTC
AGCGCCCTGCACCATATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCTGTGGAAACACCT
ACATCTGTATTAACGAAGCGCTGGCATTGACCTGAGTGATTTTCTCTGGTCCCGCCGCATCCATAC
CGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCATGTTTCATCATCAGTAACCCGTATCGTG
AGCATCCTCTCTCGTTTCATCGGTATCATTACCCCCATGAACAGAAATCCCCCTTACACGGAGGCATC
AGTGACCAAACAGGAAAAAACCGCCCTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAAACGCT
TCTGGAGAACTCAACGAGCTGGACGCGGATGAACAGGCAGACATCTGTGAATCGCTTCACGACCA
CGCTGATGAGCTTACCGCAGCTGCCTCGCGCGTTTTCCGGTGATGACGGTGAAAACCTCTGACACAT
GCAGCTCCCGGAGACGGTCACAGCTTGCTGTAAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGG
GCGCGTCAGCGGGTGTGGCGGGTGTGCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGA
GTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCGGTGTGAAATACC
GCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTG
CGCTCGGTCTGTTCCGGTGTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACA

GAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAA
AAAGGC**CGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACG**
CTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTC
CCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA
AGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGC
TGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGA
GTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGC
GAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACA
GTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCG
GCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTCGCAAGCAGCAGATTACGCGCAGAAAAA
AAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCAC
GTAAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATAAAAATGA
AGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGA
GGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCG

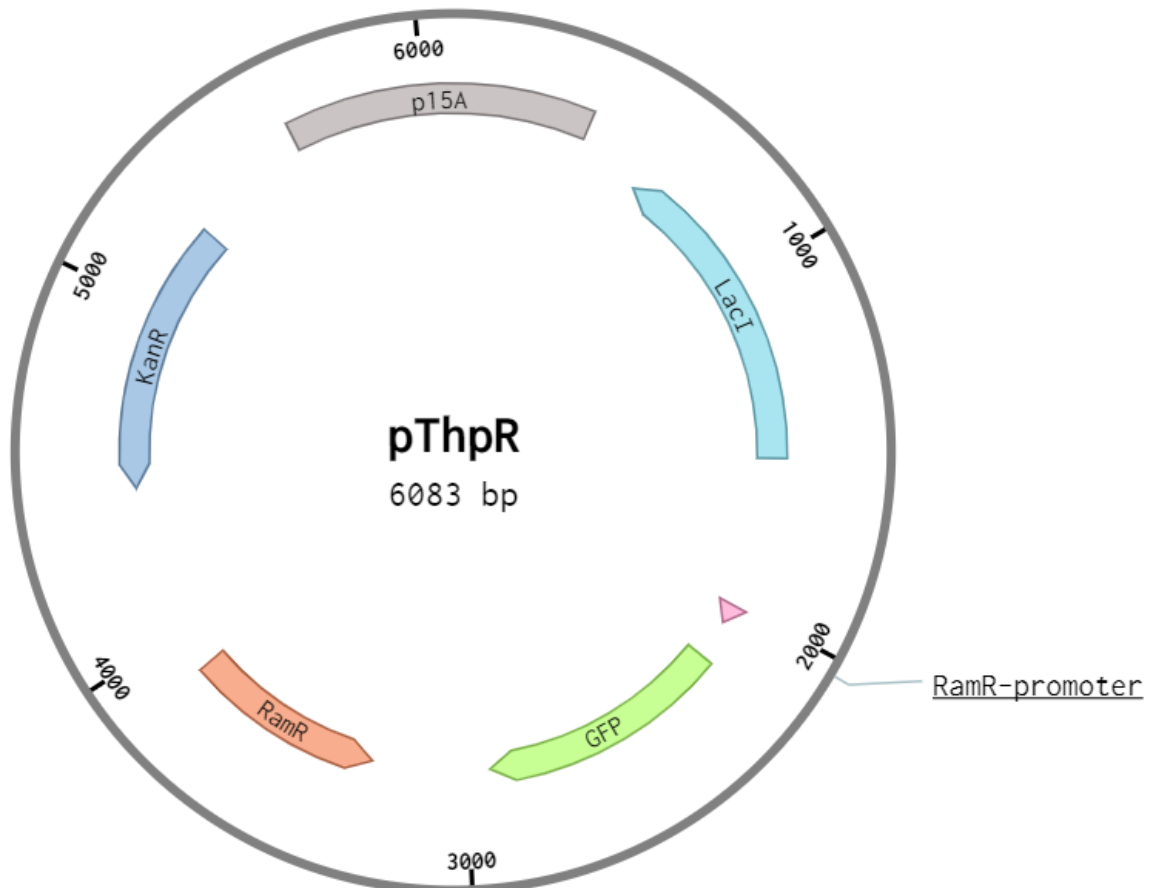


> pThpR

CCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTTCA
GTCCGACCGCTGCGCCTTATCCGGTAACATATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGC
ACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGG
CTAAACTGAAAGGACAAGTTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTT
GGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTTCAGAGCAAGAGAT
TACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTTAG
TGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGTT
AGTCATGCCCGCGCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAG
ATCCCGGTGCCTAATGAGTGAGCTAATTACATTAATTGCGTTGCGCT**CACTGCCCCGCTTTCCAGTCC**
GGAAACCTGTCTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATT
GGGCGCCAGGGTGGTTTTTCTTTTACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTG
GCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGAT
GGTGGTTAACGGCGGGATATAACATGAGCTATCTTCGGTATCGTCGTATCCCACTACCGAGATGTCCG
CACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCAGCGCCATCTGATCGTTGGCAA
CCAGCATCGCAGTGGGAACGATGCCCTCATTACAGATTTGCATGGTTTGTTGAAAACCGGACATGGC

ACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAG
CCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCCGCTAACAGCGCGATTGCTGGTGACCCA
ATGCGACCAGATGCTCCACGCCCAGTCGCGTACCATCTTCATGGGAGAAAATAACTGTTGATGGG
TGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGC
ATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACC
GCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGAT
CGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCA
ACGCCAATCAGCAACGACTGTTTGGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGC
TCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTTCGAGAAACGTGGCTGGCCTGGTTCACCACGC
GGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATT
CACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACC GCGAAAGGTTTTGCGCCATTCTG
ATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACGCGGCCGCGGCATCAGAGCAGATTGTACTG
TGTCTCAATGGTTCGTTGTGATGGCGGTAGGAATGTAATCGTTAATCCGCAAATAACGTAAAAACC
CGCTTCGGCGGGTTTTTTTATGGGGGGAGTTTAGGGAAAGAGCATTGTGCATCCCGTTGAATATGGC
TCGCATCTTATCGAGCATACTATCACGTCGGCGACCACTAGTCAGTTAACGCAAGGGCATGGGCTGT
CGACTTTTGAAAAGTACCTTGACGGCGTATCTTTGCTTTCTATAATGAGTGCTTACTACTCATACAA
TAGTCAGTCATAAGTCTGGGCTAAGCCCACTGATGAGTCGCTGAAATGCGACGAACTTATGACCTC
TACAAATAATTTGTTTAAACGTAAACCTCCGGGTTAATAAGGAGTAATTATGGCATCCAAGGGCGAG
GAGCTCTTTACTGGCGTAGTACCAATTCTCGTAGAGCTCGATGGCGATGTAAATGGCCATAAGTTTTC
CGTACGCGGCGAGGGCGAGGGCGATGCAACTAACGGCAAGCTCACTCTCAAGTTTATTTGTACTACT
GGCAAGCTCCCAGTACCATGGCCAACTCTCGTAACTACTCTGACCTATGGCGTACAATGTTTTTCCCG
CTATCCAGATCACATGAAGCAACATGATTTTTTTAAGTCCGCAATGCCAGAGGGCTATGTACAAGAG
CGCACTATTAGCTTTAAGGATGATGGCACCTATAAGACTCGCGCAGAGGTAAAGTTTGAGGGCGATA
CTCTCGTAAATCGCATTGAGCTCAAGGGCATTGATTTTAAGGAGGATGGCAATATTCTCGGCCATAA
GCTGGAGTATAATTTCAATTCCCATAAATGTATATATTACCGCAGATAAGCAAAAGAATGGCATTAAAG
CGAATTTTAAGATTCGCCATAATGTGGAGGATGGCTCCGTACAACTCGCAGATCATTATCAACAAAAT
ACTCCAATTGGCGATGGCCCACTACTCCTCCAGATAATCATTATCTCTCCACTCAATCCGTGCTCTC
CAAAGATCCAAATGAGAAGCGCGATCACATGGTACTCCTGGAGTTTGTAAGTGCAGCAGGCATTACT
CATGGCATGGATGAGCTCTATAAGCTCGAGCACCACCACCACCACCCTGATAATCCAAACCTGTTA
TATGTTAGCTGAGACTAGTTGGAAGTGTGGCTGTCTCAAGCGTTTTAGTTTCGTGGTCAGTTTCAC
CTGATTTACGTAAAAACCCGCTTCGGCGGGTTTTTGCTTTTGGAGGGGCAGAAAGATGAATGACTG
TCGGCCATTTCGATGGTGTCTGGGTAGCATAACCCCTTGTGATACCTTTGCCATGTTTCAGAAACAACT
CTGGCGCATCGGGCTTGGACCAAAACGAAAAAAGGCCGCTTTCGCGGCCCTCTTTCTGGAATTTGG
TACCGAGCTCACAGCGGCGATAGTCAGATAGCTAGACCGTATGTTACCGAGCCTGCACTCCTCGAAT
TCTTAGGATTAATTACTGCTCTTCGCGCGTAAGTGCAGCGCCACATAGCCTCGAAGCCCAACGCAATGT
ACTCACCAGCGCGAGCCGGGTTGCGCGCAGCGAAATCCATAGTCGTCTCAGCAAGCGCCAAGAAC
AACCCGTCGCCGAAGGCGCGGTACTCGTCGGACATAAACACCATAAGAACCTCACGGTGGACGATG
TCGCGTAACTCCGGGAACATATCATCCGCGCGTTGTTTCGGTCTCCTTCGTCAACTTTTCAGAAACCG
CCAACTGACGTATGGCACGATGGCGAGCTGGGTGGTTCAATCCCCAGCTAATGATACTGTTCCAGAG
AAAACGCGACATCATCTTAGCGTCAGTAATAGAACGATCCAATTCCATGATGGATGATTGGCACCAAG
TCCTGGAACAAATGTAAGTAAAGGGTGTGATCAACTCATCTTTCGTTGCGAAATAGCGGAACAAC
GTCCCTTCCGCAACTCCCGCATTGCGTGCAATTACAGCGGTACTAGCGGCAATGCCTGATTGCGCGA
TGGCTTGAGTTGCCGCTTCAAGCAATGCCTGCTTTTTGTCTCAGACTTTGGGCGAGCAACCATATA
CTAACCTCCTTCTGATACGTGGTTCGGTTAAACAAAATTTATTTGTAGAGGCCCATTTTCGTCTTTTG
GACTCATCAGGGGTGGTACACACCACCCTATGGGGCTCGTAATTGCTAGCATAATCCCTAGGACTGA
GCTAGCTATCAGGGTACTTTTCAAAGGTCGACAGCCCATGCCCTTGCCTTCGGCAGGTGTACAATGA
TACGAGGTAAATGAAGATGAAGTCCATACAATCGATAGATTGGGACCAAAACGAAAAAAGGGGAGCG
GTTTTCCCGCTCCCTCTTTTCTGGAATTTGGTACCGAGTCGCACCTGATTGCCCGACATTATCGCACG
GTGTCTCATCTCTGATAACGCATATTGTCTGTAGAACTCGGCGCGGCCGCTCACACTGCTTCCGGTA
GTCAATAAACCGGTAAACCAGCAATAGACATAAGCGGCTATTAAACGACCCTGCCCTGAACCGACGA
CCGGGTTCGAATTTGCTTTTCGAATTTCTGCCATTATCCGCTTATTATCACTTATTCAGGCGTAGCAAC
CAGGCGTTTAAAGGGCACCAATAACTGCCTTAAAAAAAATTAGAAAACTCATCGAGCATCAAATGAA
ACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAG
AAAATCACCAGGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCC
AACATCAATACAACCTATTAATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAG
TGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCA
GCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAAACCGTTATTCATTCGTGATTGCGCCTGAG
CGAGACGAAATACGCGGTGCTGTGTTAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCA
GGAACACTGCCAGCGCATCAACAATTTTTACCTGAATCAGGATATTCTTCTAATACCTGGAATGCT

GTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGG
 TCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAAC
 GCTACCTTTGCCATGTTTCAGAAACAACCTCTGGCGCATCGGGCTTCCCATAACAATCGATAGATTGTGCG
 CACCTGATTGCCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTA
 ATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATTTTAGCTTCCTTAGCTCCTGAAAAT
 CTCGATAACTCAAAAAATACGCCCCGGTAGTGATCTTATTTTCATTATGGTGAAAGTTGGAACCTCTTAC
 GTGCCGATCACGTCTCATTTTCGCCAAAGTTGGCCAGGGCTTCCCGGTATCAACAGGGACACCAGG
 ATTTATTTATNNTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGCGTCGGGTGATG
 CTGCCAACTTACTGATTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAG
 CTGTCCCTCCTGTTTCAGCTACTGACGGGGTGGTGCGTAACGGCAAAGCACC GCCGGACATCAGCG
 CTAGCGGAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTGAGTGAAGTGCTTCATGTGG
 CAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATATATTCCGCTTCCTCG
 CTCCTGACTCGCTACGCTCGGTCGTTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGA
 GATTTCTTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTT
 TCCATAGGCTCCGCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACC
 CGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCTGCTGC
 CTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACG



Supplementary Table 4. Sequences and graphical maps of representative plasmids.