Supplementary Materials for

Using enantioselective biosensors to evolve asymmetric biocatalysts

This PDF file includes:

Figures S1 - S25

Tables S1, S2

Captions for Supplementary Data files

Other Supplementary Material for this manuscript includes the following:

Supplementary Data 1. DMS results.csv

Supplementary Data 2. Cytometry results.csv

Supplementary Data 3. GenBank files of all relevant plasmids

Supplementary Figures

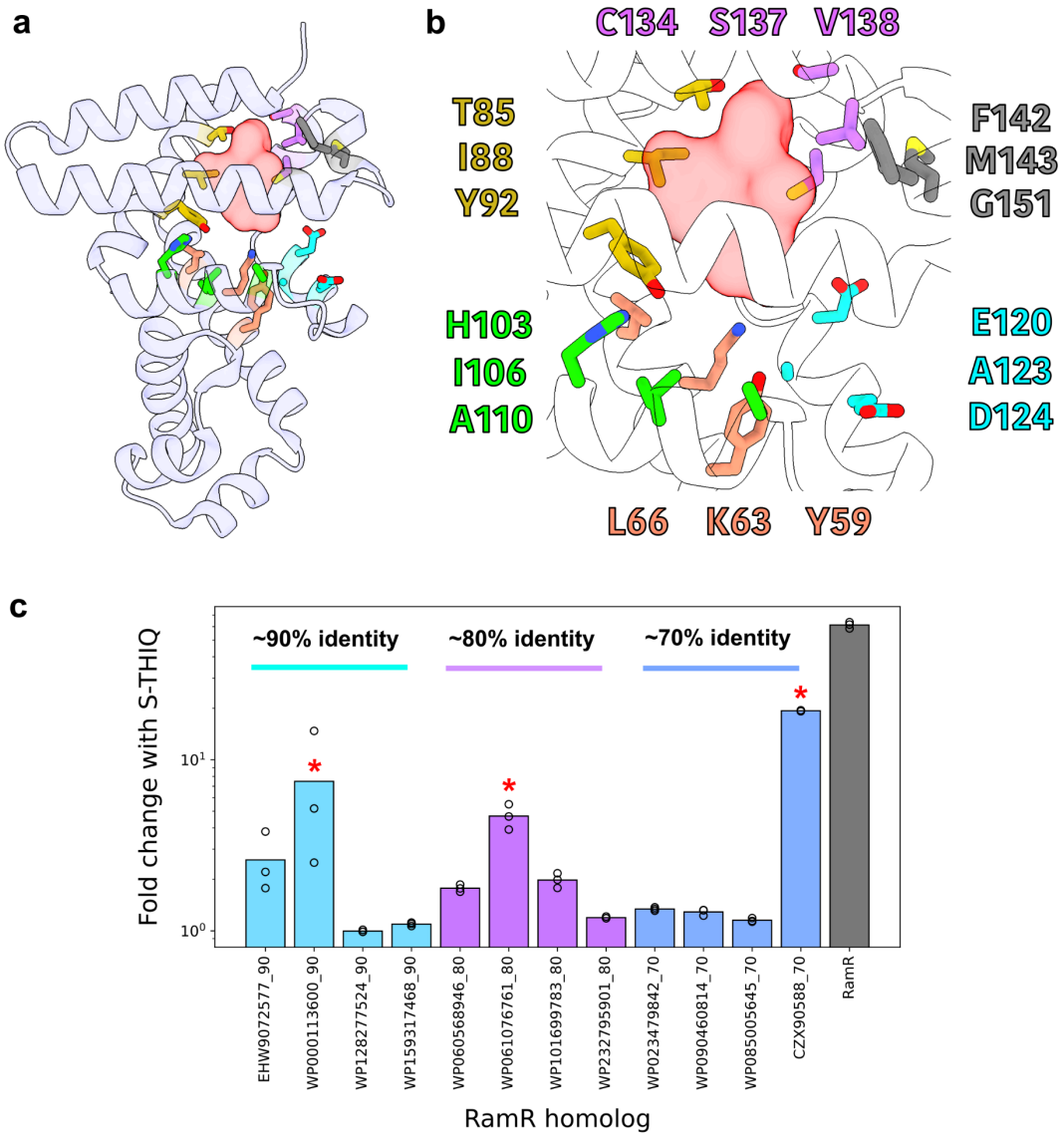


Figure S1. Site-saturation and shuffling library designs. (a) Global structure of a RamR (PDB: 3VVY) monomer with sets of residues targeted for combinatorial site-saturation mutagenesis shown as colored sticks. The ligand binding pocket is shown as a semi-transparent red surface. (b) Magnified view of the RamR ligand binding pocket. Residues subject to combinatorial site-saturation mutagenesis are colored in groups and labeled on the perimeter. (c) Response of RamR homologs to S-THIQ as measured by the fold change in fluorescence upon exposure to 1 mM of S-THIQ. Homologs are grouped and color-coded by their approximate percent identity to RamR. Asterisks indicate variants that were used to create the shuffled library. Measurements were performed in biological triplicate. Bars represent averages and circles represent individual measurements.

Growth Plate Layout	1	2	3	4	5	6	7	8	9	10	11	12
A	200 Zeo	0 Zeo	200 Zeo	125.0 DHIQ, 0 Zeo	125.0 R-THIQ, 0 Zeo	125.0 1S-TIQ, 0 Zeo	5.12 DHIQ, 200 Zeo	12.8 DHIQ, 200 Zeo	32.0 DHIQ, 200 Zeo	80.0 DHIQ, 200 Zeo	200.0 DHIQ, 200 Zeo	500.0 DHIQ, 200 Zeo
B	500.0 R-THIQ, 200 Zeo	200.0 R-THIQ, 200 Zeo	80.0 R-THIQ, 200 Zeo	32.0 R-THIQ, 200 Zeo	12.8 R-THIQ, 200 Zeo	5.12 R-THIQ, 200 Zeo	5.12 S-THIQ, 200 Zeo	12.8 S-THIQ, 200 Zeo	32.0 S-THIQ, 200 Zeo	80.0 S-THIQ, 200 Zeo	200.0 S-THIQ, 200 Zeo	500.0 S-THIQ, 200 Zeo

Figure S2. Plate map for the growth-based assay.
Culturing conditions for the pooled RamR biosensor libraries prior to measurement via barcode sequencing. This pattern was duplicated four times across eight columns in the 96-well plate for cell culturing; rows C, E, and G were the same as A; rows D, F, and H were the same as B. Concentrations are μmol/L for ligands and μg/mL for Zeocin. The ligand concentrations are a 2.5-fold serial dilution.

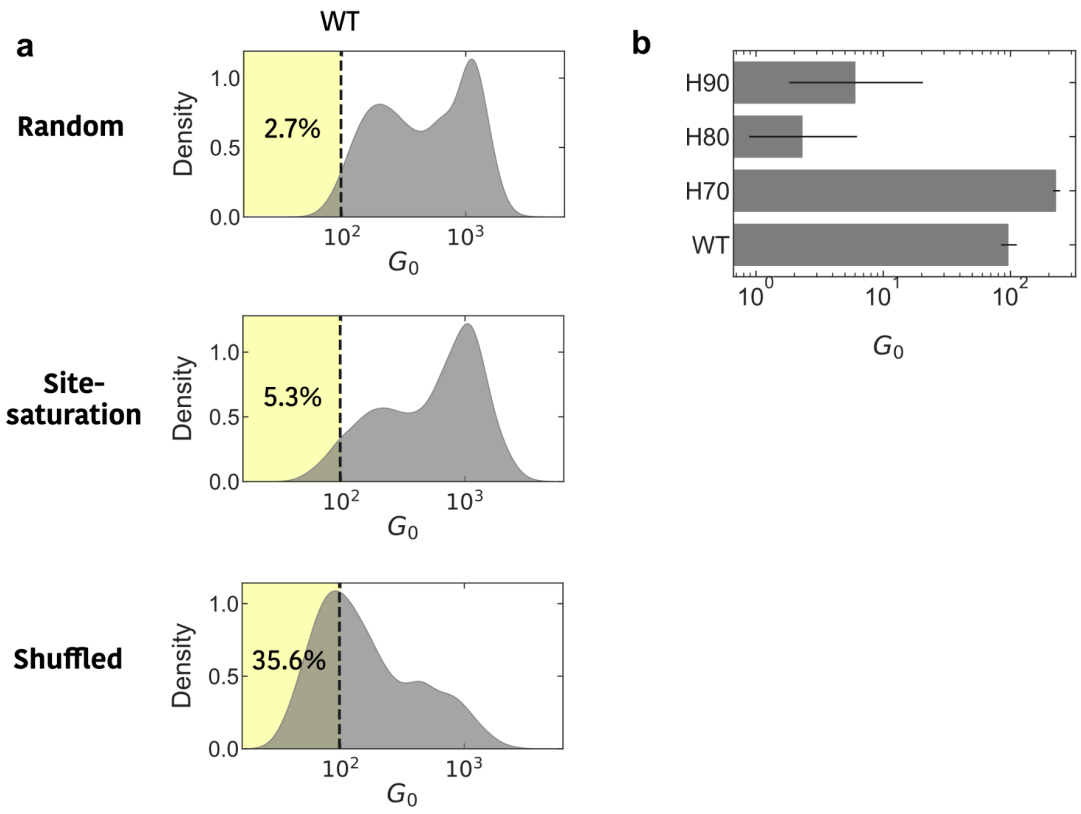


Figure S3. Background signal among library members. (a) Background signal distribution across three library designs. Dotted lines indicate the background (G_0) value for the template RamR protein. Yellow highlighted regions illustrate the fraction of variants with lower background than the template, which is displayed as a percentage within each plot. (b) Background signal of the wild-type RamR protein (WT) as well as the three homologs used for shuffling (H90, H80, H70), as measured using flow cytometry. Error bars represent the posterior one standard deviation for the log-transformed G_0 valuers.

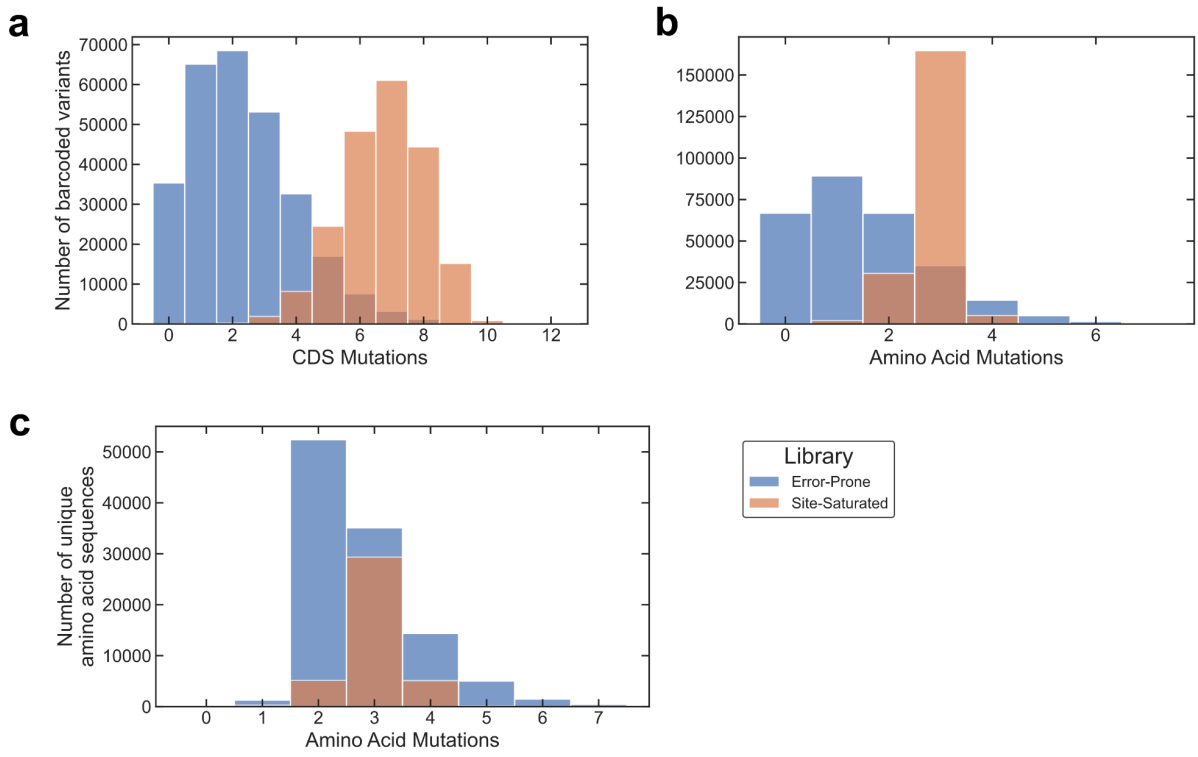


Figure S4. Distribution of mutations in the random and site-saturation libraries. Distribution of coding sequence mutations (a) and amino acid mutations (b) among all variants from the error-prone (blue) and site-saturation (orange) libraries with unique barcodes. Distribution of amino acid mutations among all variants with unique protein sequences from the error-prone (blue) and site-saturation (orange) libraries.

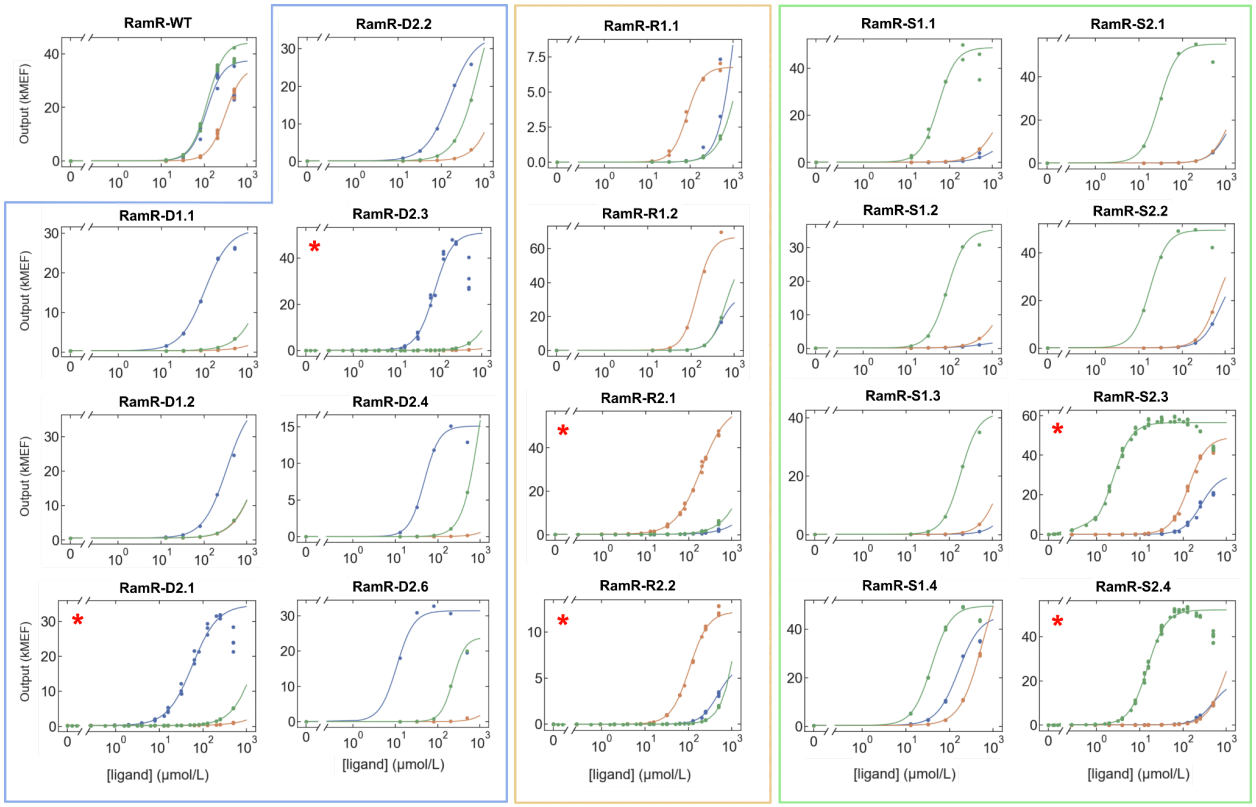


Figure S5. Dose-response plots of stereoselective variants. Variants were selected from the large scale library measurements on the basis of selectivity, dynamic range, and sensitivity. Red asterisks denote variants selected for crystallography. Plotted data is from flow cytometry. Measurements were performed in biological singlicate except for variants with red asterisks, which were measured in biological triplicate. Detailed phenotype and genotype information for each variant can be found in **Supplementary Data 2**.

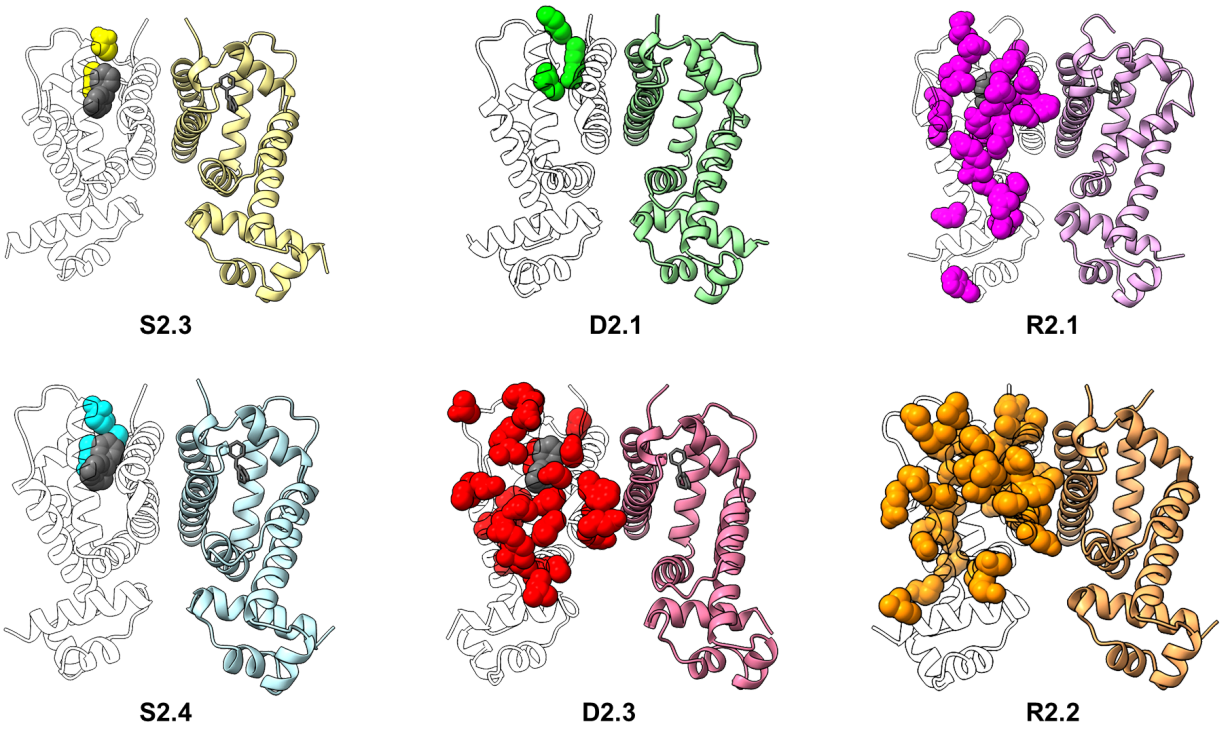


Figure S6. Global structure of solved RamR variants. Sites of mutations are shown as space-filling models on top of transparent cartoon models. For variants bound with the cognate ligand, ligands are shown together inside the binding cavity (gray stick / sphere).

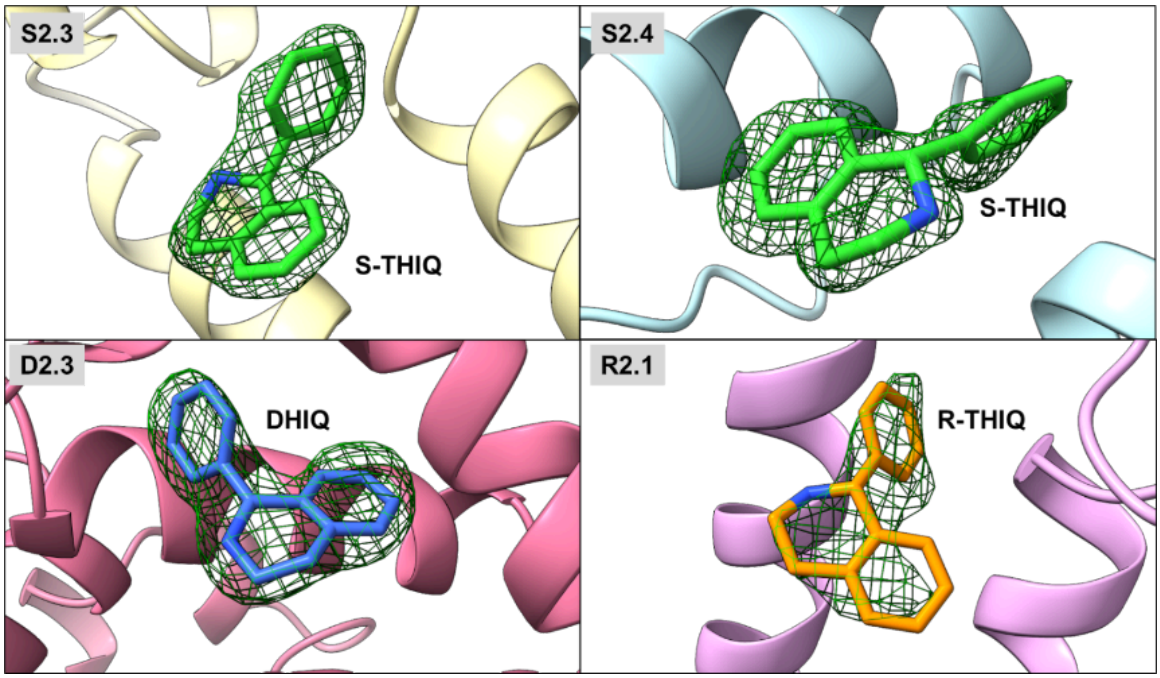


Figure S7. Ligand fitting into the X-ray crystallography data. F_o-F_c omit map (green mesh, 2.5σ) is shown for each complex structure. The ligands shown in stick models (S-THIQ - green, DHIQ - blue, R-THIQ - orange) are superimposed on top of the omit map to exhibit consistency.

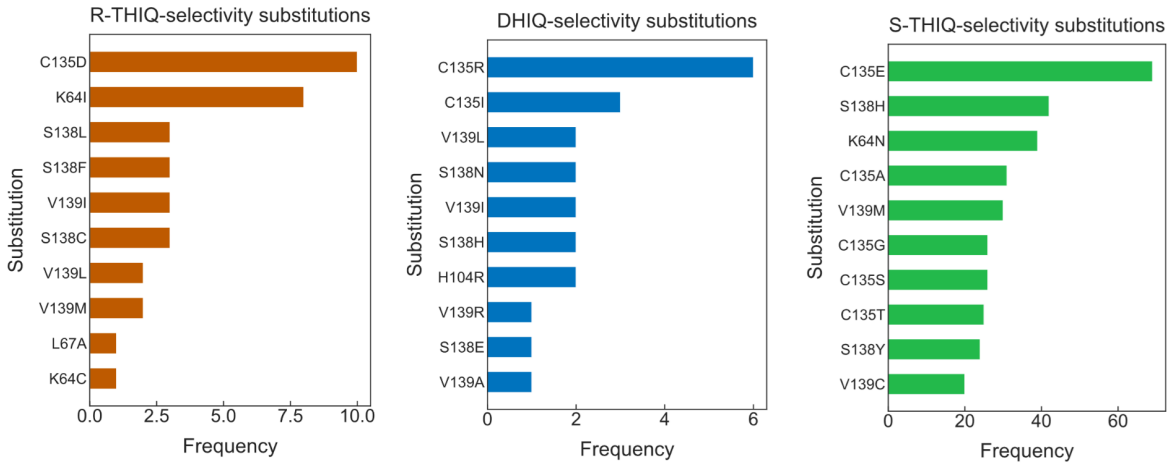


Figure S8. List of enriched substitutions that improve selectivity. Ligands are color-coded as follows: R-THIQ (orange), DHIQ (blue), and S-THIQ (green). The following filters were applied to the library dataset: (**R-THIQ**: Fold induction > 50; Selectivity > 40; EC_{50} < 200 µmol/L. **DHIQ**: Fold induction > 50; Selectivity > 40; EC_{50} < 200 µmol/L. **S-THIQ**: Fold induction > 50; Selectivity > 50; EC_{50} < 50 µmol/L).

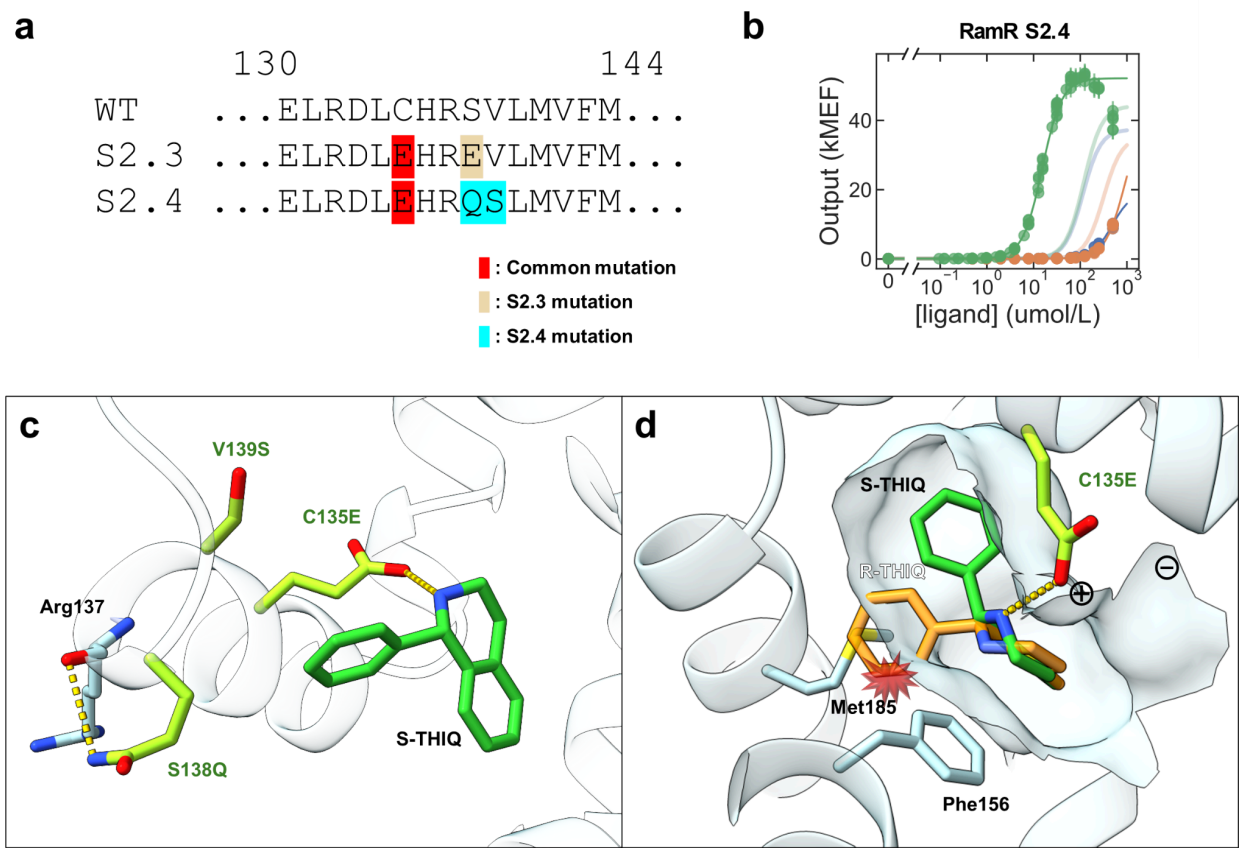


Figure S9. Sequence comparison of S-THIQ selective variants and structural characterization of S2.4 variant. (a) Sequence comparison of S2.3 and S2.4 variants. (b) Dose-response curves for S2.4 with all three target ligands, measured with flow cytometry. Transparent trendlines represent wild-type RamR ligand responses. Measurements were performed in biological triplicate. Individual data points are shown as filled circles. Error bars represent the standard deviation from the mean. (c) Ligand binding site configuration of S2.4. S-THIQ (green stick) and surrounding residues (sky blue - native residues, yellow - mutated residues) are shown. (d) Binding site analysis and modeling of different ligands into the S2.4 crystal structure. Modeled R-THIQ is shown as an orange stick model.

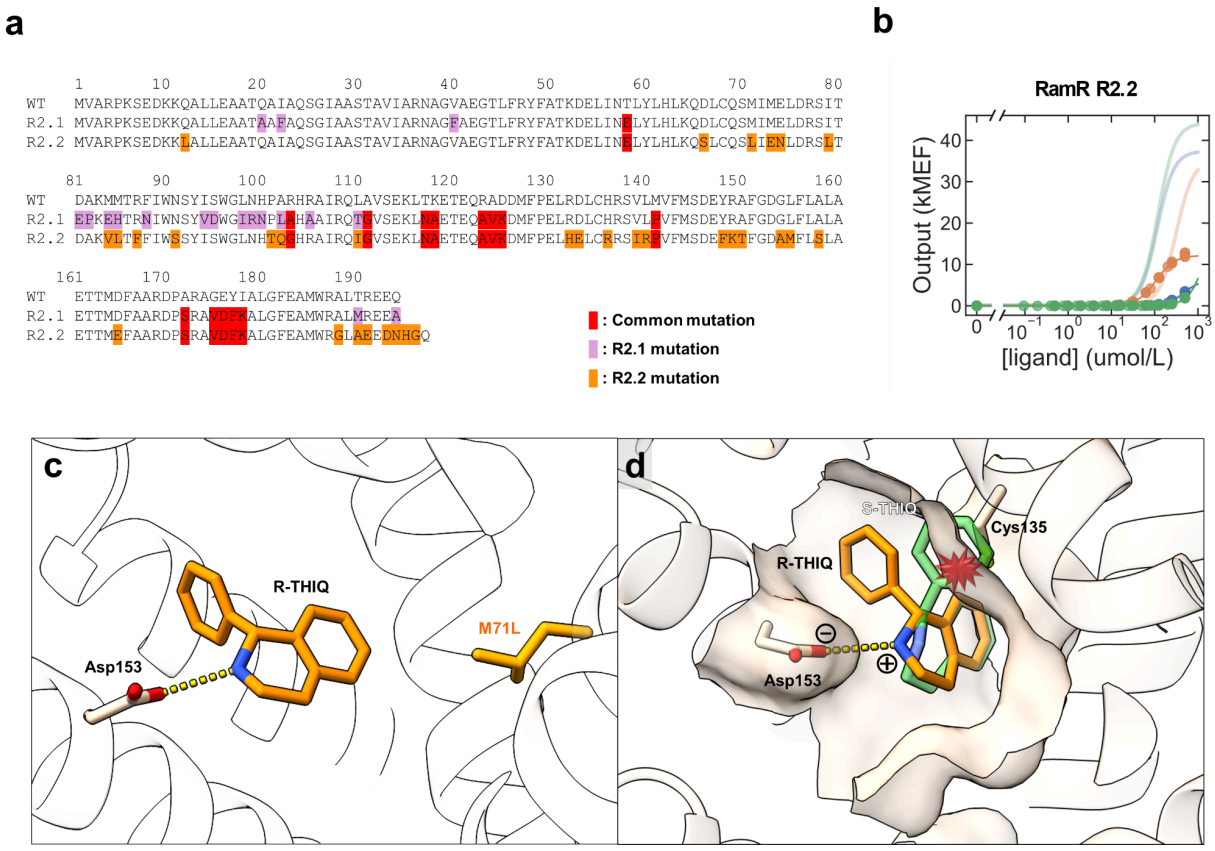


Figure S10. Sequence comparison of R-THIQ selective variants and R2.2 apo structure docked with R-THIQ. (a) Sequence comparison of R2.1 and R2.2. (b) Dose-response curves for R2.2 with all three target ligands, measured with flow cytometry. Transparent trendlines represent wild-type RamR ligand responses. Measurements were performed in biological triplicate. Individual data points are shown as filled circles. Error bars represent the standard deviation from the mean. (c) Ligand binding site of R2.2 docked with R-THIQ (orange stick). Surrounding residues (wheat - native residues, peach - mutated residues) are shown. (d) Binding site analysis and modeling of different ligands into the R2.2 apo structure. Modeled S-THIQ is shown as a green stick model.

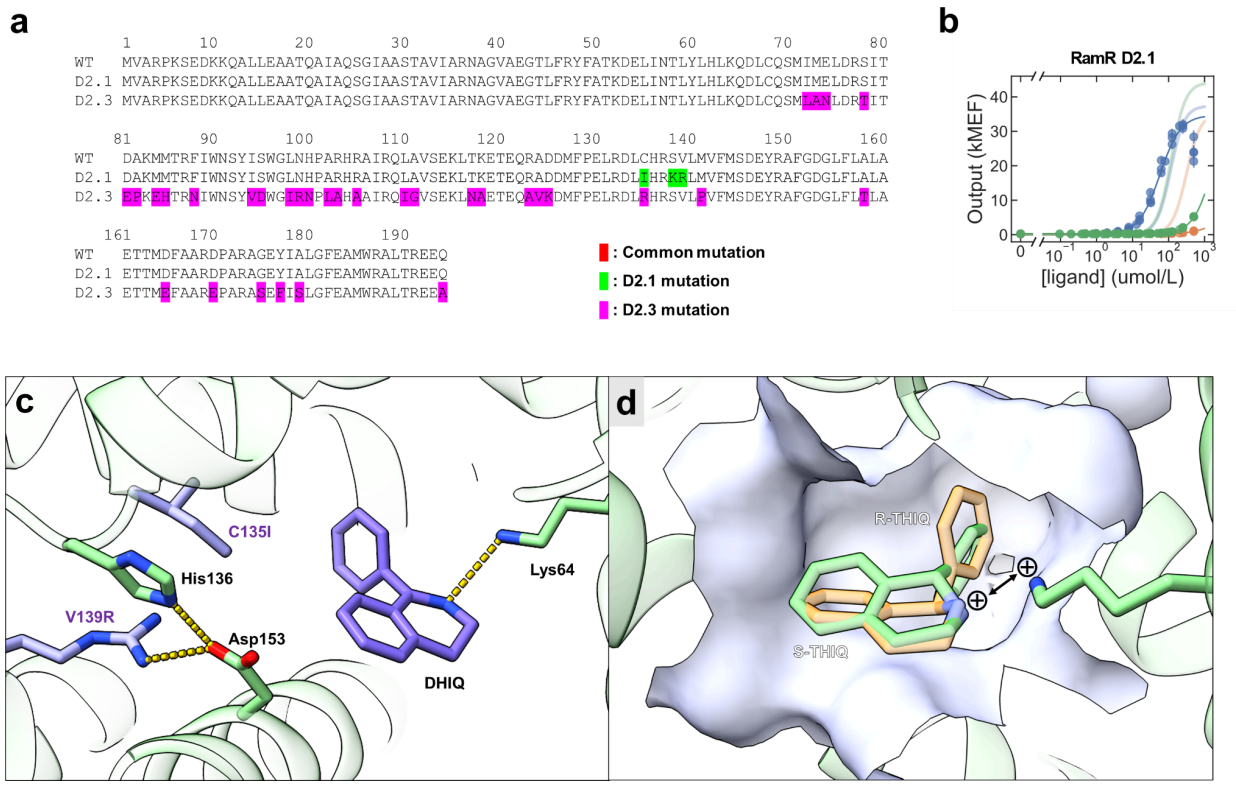


Figure S11. Sequence comparison of DHIQ selective variants and D2.1 apo structure docked with DHIQ. (a) Sequence comparison of D2.1 and D2.3. (b) Dose-response curves for D2.1 with all three target ligands, measured with flow cytometry. Transparent trendlines represent wild-type RamR ligand responses. Measurements were performed in biological triplicate. Individual data points are shown as filled circles. Error bars represent the standard deviation from the mean. (c) Ligand binding site of D2.1 docked with DHIQ (purple stick). Surrounding residues (green - native residues, violet - mutated residues) are shown. (d) Binding site analysis and modeling of different ligands into the D2.1 apo structure. Modeled S-THIQ and R-THIQ are shown as lime green and orange stick models, respectively.

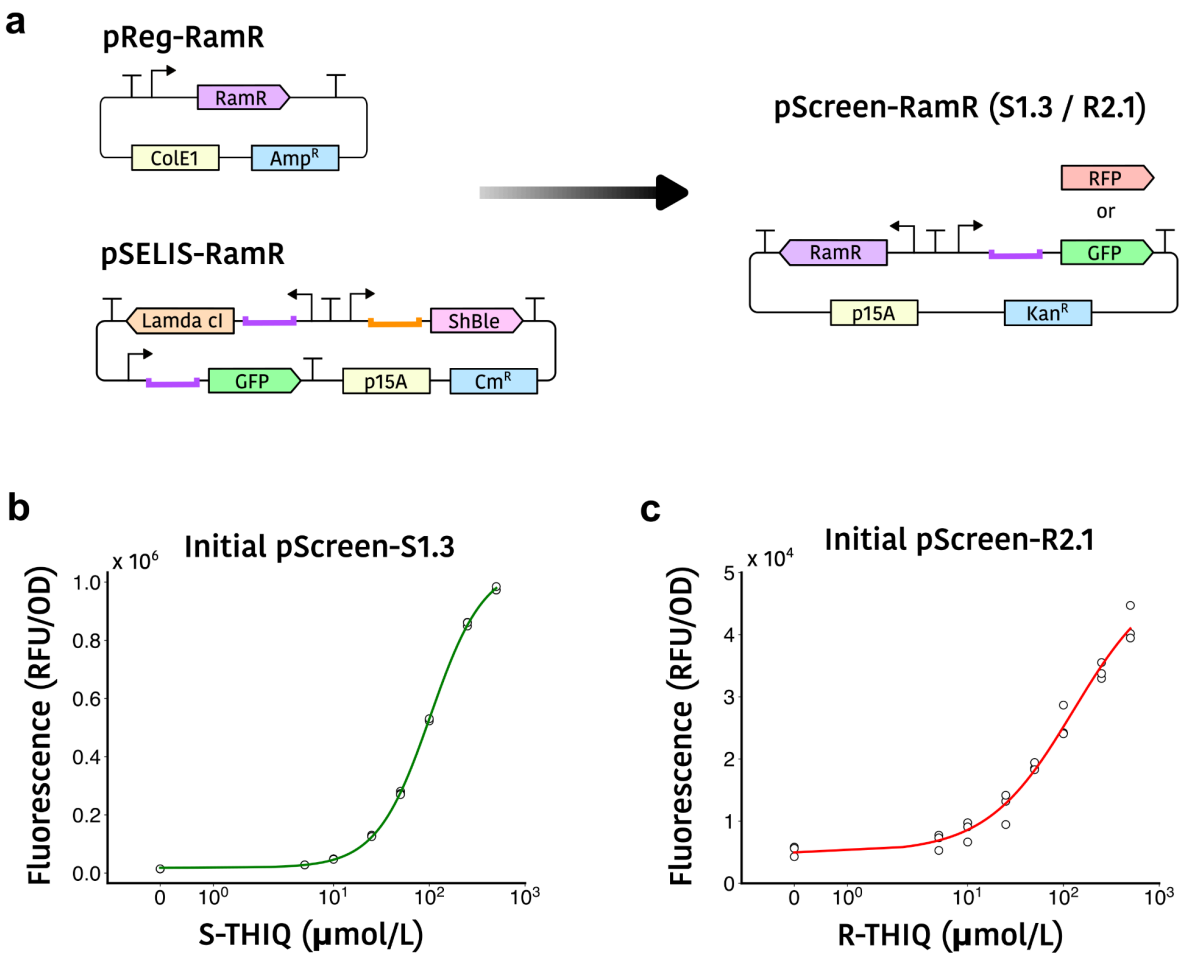


Figure S12. Design and characterization of initial screening plasmid. (a) Comparison of the two-plasmid system used for biosensor evolution (left) and the one-plasmid pScreen system used for biosensor-coupled fluorescent measurement of IRED activity and enantioselectivity (right). (b) Dose response of the original pScreenS1.3 plasmid design that uses RamR-S1.3 regulates the expression of GFP. (c) Dose response of the original pScreenR2.1 plasmid design that uses RamR-R2.1 to regulate RFP expression. Measurements were performed in biological triplicate and trendlines are fit to the mean. Individual data points are shown as unfilled circles.

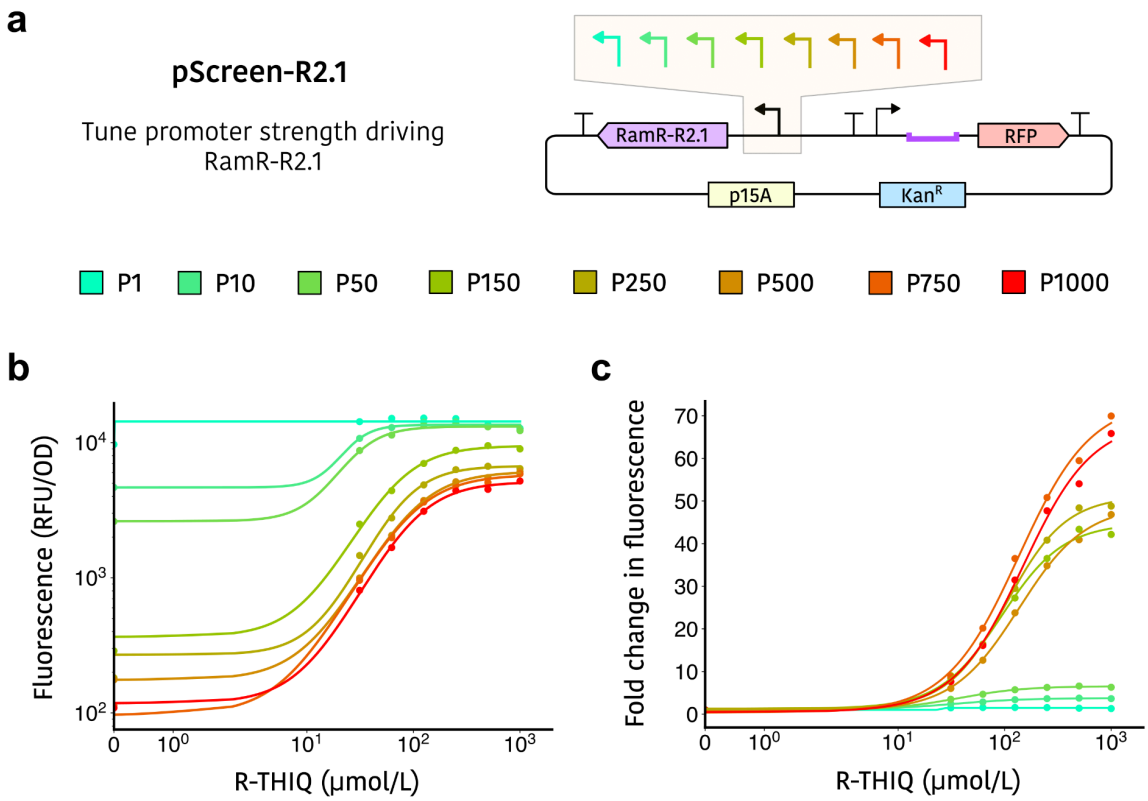


Figure S13. Tuning the pScreen-R2.1 plasmid. (a) Schematic of the pScreen-R2.1 biosensor plasmid, highlighting the promoter driving RamR-R2.1 expression, the strength of which is being tuned. (b,c) Dose responses of pScreen-R2.1 plasmids using different strength constitutive promoters (P1-P1000) to express RamR-R2.1. Measurements were performed in singlicate. Individual data points are shown as filled circles.

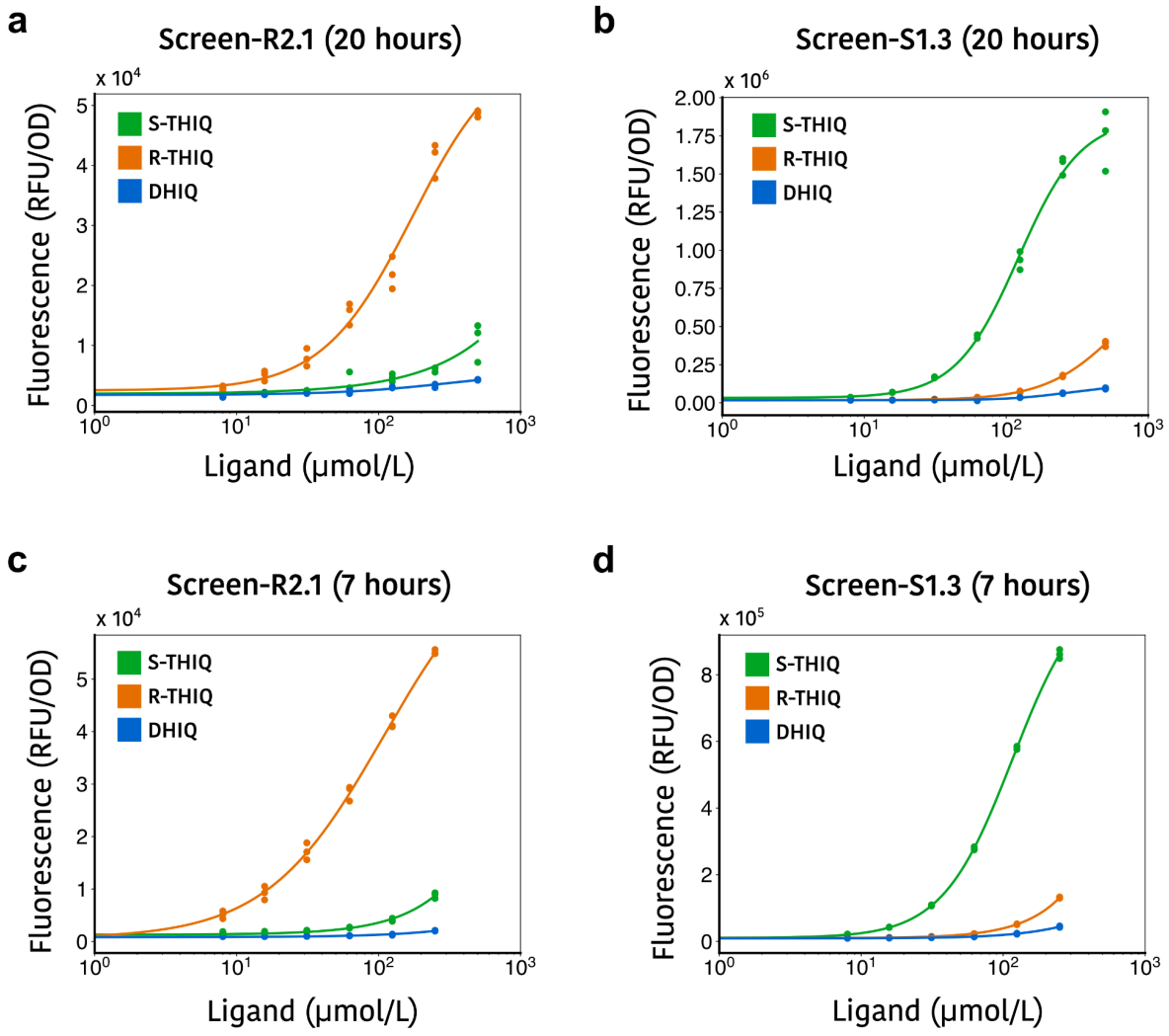


Figure S14. Characterization of final pScreen plasmids. The pScreen-R2.1 plasmid uses the P750 promoter to express the RamR-R2.1 gene and the pScreen-S1.3 uses the P50 promoter to express the RamR-S1.3 gene (**Supplementary Data 3**). Dose responses of the pScreen-R2.1 plasmid (a,c) and the pScreen-S1.3 plasmid (b,d) with S-THIQ, R-THIQ, and DHIQ when cultured for 20 hours (a,b) or 7 hours (c,d). Measurements were performed in biological triplicate and trendlines are fit to the mean. Individual data points are shown as filled circles.

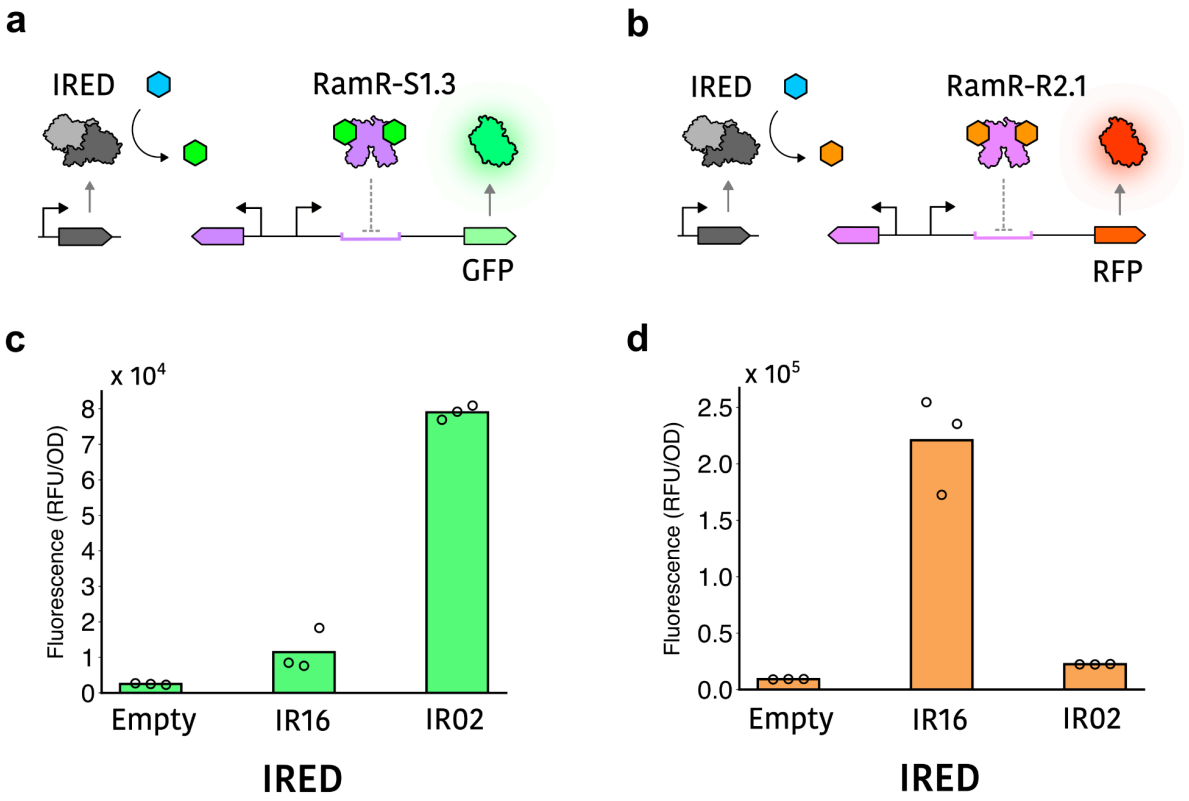


Figure S15. Preliminary screening of IREDs with the biosensor plasmids. Schematic of the pScreen-S1.3 plasmid co-expressed with an S-specific IRED (a) or the pScreen-R2.1 plasmid co-expressed with an R-specific IRED (b) when cultured in the presence of DHIQ. (c) Fluorescence of cells expressing the pScreen-S1.3 plasmid and either an empty plasmid, an R-specific IRED (IR16) or an S-specific IRED (IR02). (d) Same format as (c), but using the pScreen-R2.1 plasmid instead of the pScreen-S1.3 plasmid. Measurements were performed in biological triplicate and bars represent averages. Individual data points are shown as unfilled circles.

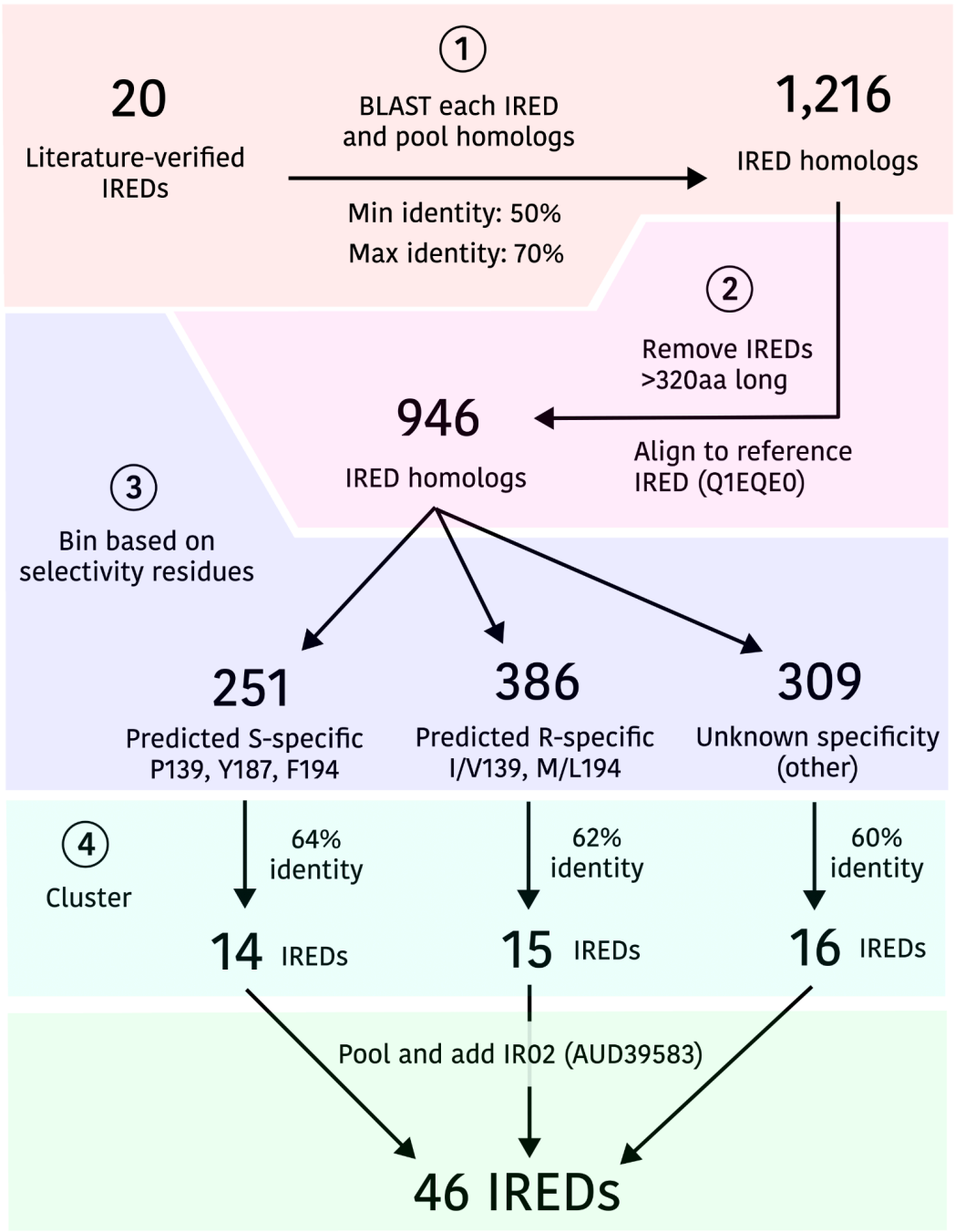
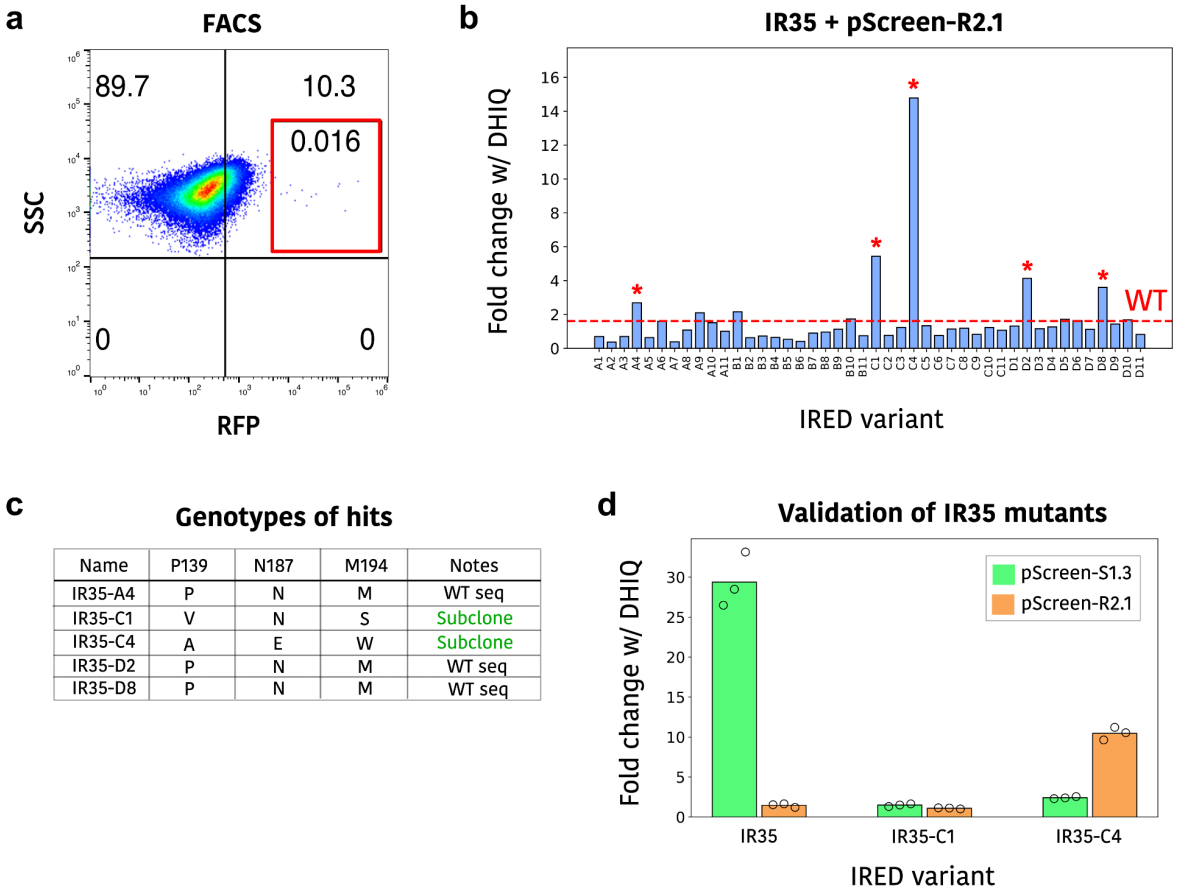


Figure S16. IRED curation workflow. See Methods section "IRED curation" for details.



Name	P139	N187	M194	Notes
IR35-A4	P	N	M	WT seq
IR35-C1	V	N	S	Subclone
IR35-C4	A	E	W	Subclone
IR35-D2	P	N	M	WT seq
IR35-D8	P	N	M	WT seq

Figure S17. FACS gating, secondary screen, and enantioselectivity validation of IR35 libraries. (a) Gating strategy used to sort the most fluorescent cells from the IR35 library co-transformed with the pScreen-R2.1 plasmid. Numbers in each quadrant indicate the percentage of the cell population that fit within that quadrant. (b) Secondary screen of IR35 mutants shown as the fold change in fluorescence of strains in the presence of 100 μmol/L of DHIQ versus the fluorescence in the absence of DHIQ. The fold change in fluorescence produced by IR35 is shown as a dotted red line. Measurements were performed in singlicate. Asterisks indicate mutants that were selected for sequencing. (c) Names and genotypes of selected IR35 mutants. (d) Fold change in fluorescence (with vs without DHIQ) of sub-cloned IR35 mutants and the template IR35 enzyme co-transformed with the pScreen-R2.1 (orange) or pScreen-S1.3 (green) plasmids. Measurements were performed in biological triplicate and bars represent averages. Individual data points are represented as unfilled circles.

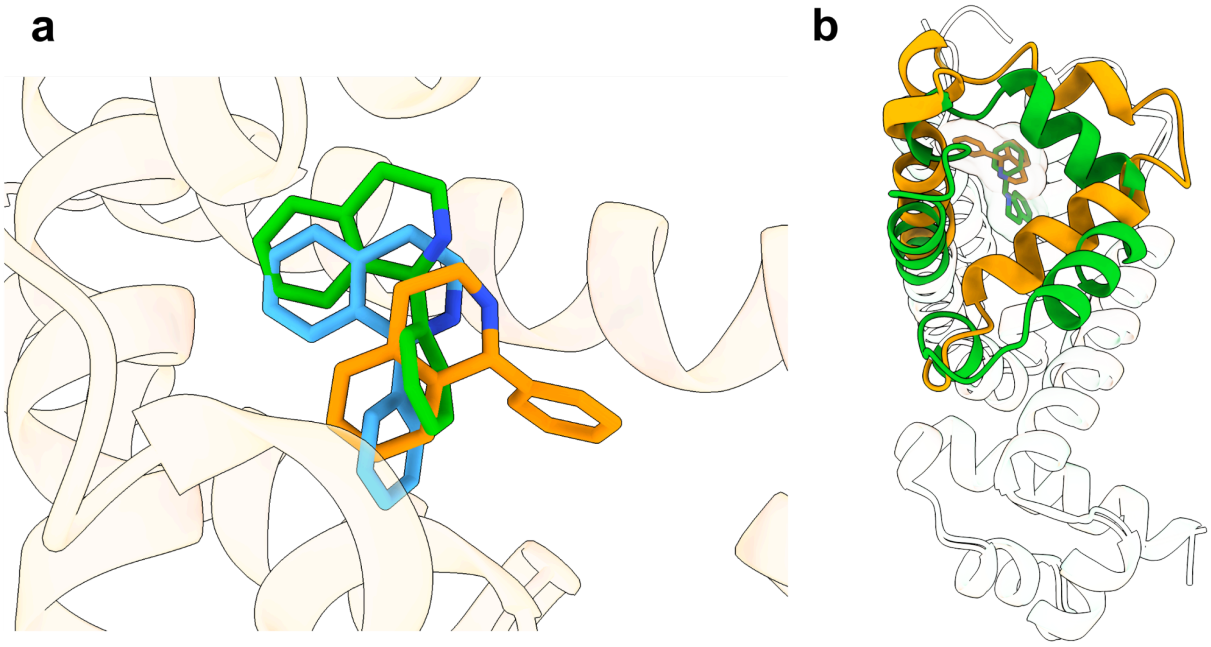


Figure S18. Structural comparison of ligand binding poses and binding pockets. (a) Superimposed variant structures of RamR-S2.3, RamR-R2.1, and RamR-D2.3, with the corresponding S-THIQ (green), R-THIQ (orange), and DHIQ (blue) ligands shown as sticks. (b) The superimposed structures of RamR-S2.3 (green) and RamR-R2.1 (orange). Regions of the structures that align almost perfectly are transparent while regions that do not align well are shown in solid color. The ligands S-THIQ (green) and R-THIQ (orange) are shown as sticks.

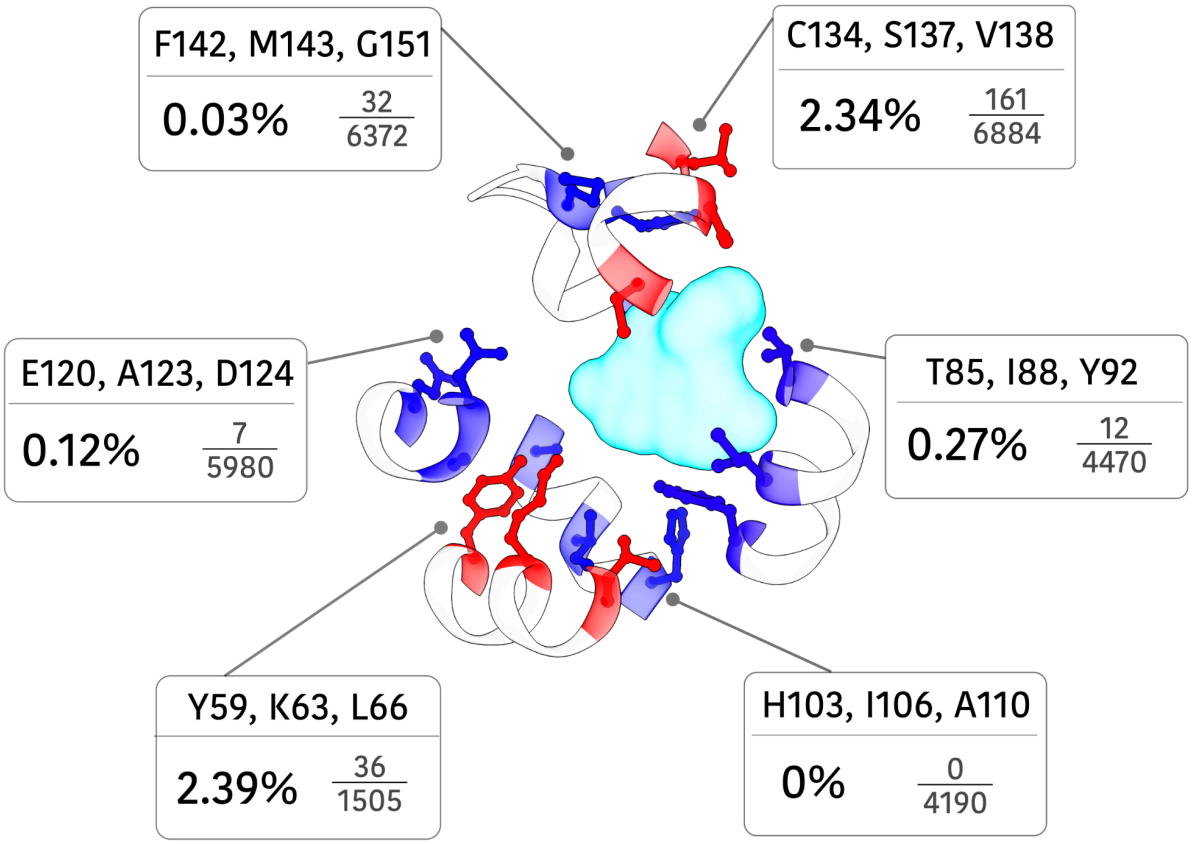


Figure S19. Comparison of selective variant frequencies among six site-saturation libraries. Percentages and fractions represent the number of variants within each library that had 2-fold higher selectivity for any ligand compared to both other ligands. Library residues are color-coded on a gradient scale with red representing a high frequency of selective library members and blue indicating a low frequency of selective library members.

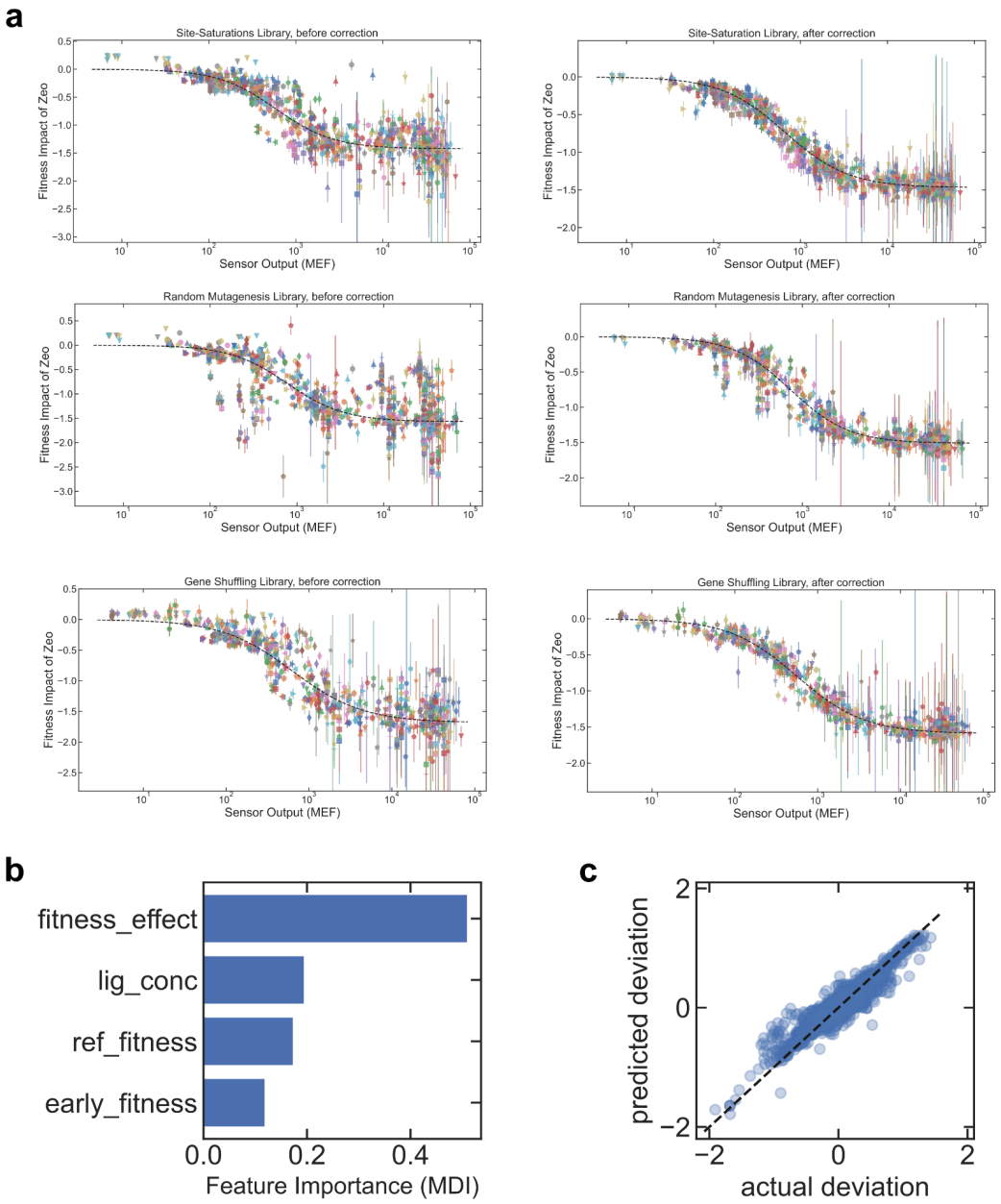


Figure S20. Calibration of the deep mutational scanning assays. (a) The plots show the calibration data (points with error bars) and the calibration fits (dashed black curves) used to calibrate from fitness to function for each library DMS measurement. The left-side plots show the raw data, before application of the gradient boosting regressor to correct for the calibration deviations. The right-side plots show the corrected data, after application of the gradient boosting regressor. Within each plot, the x-axis values are from flow cytometry measurements of the calibration variants, and the y-axis values are from the fitness derived from DNA barcode sequencing. (b) Estimated feature importance for the gradient boosting regressor: Mean Decrease in Impurity (MDI). (c) Plot of predicted deviations from the gradient boosting regressor vs. the actual deviations.

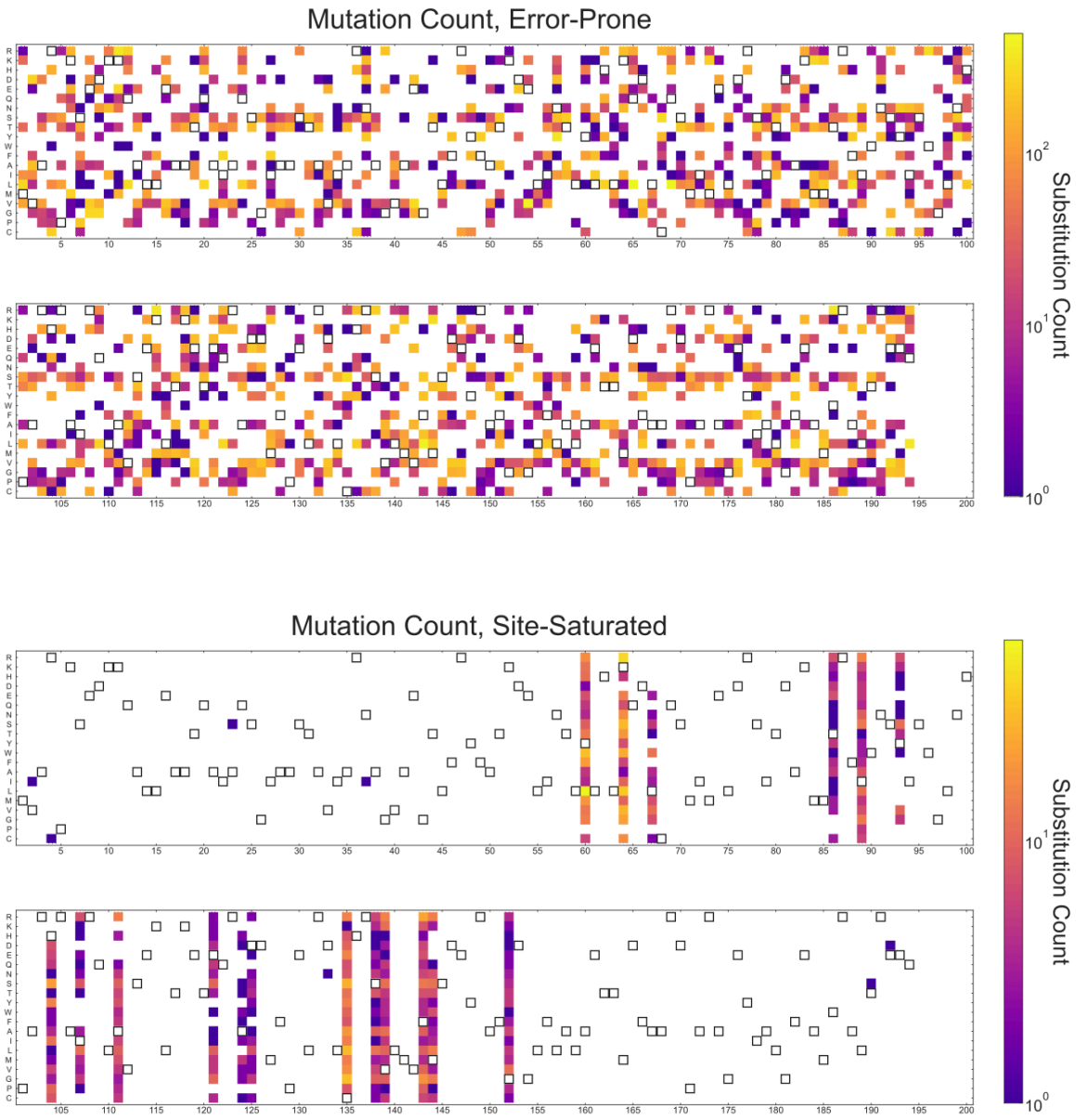


Figure S21. Substitution mutation counts across error-prone and site-saturation libraries. Outlined boxes represent the wild-type amino acid at each position. The heat map indicates the number of observations of each mutation across the full datasets (present as single or multiple substitutions). Blank (white) entries indicate non-observed substitutions.

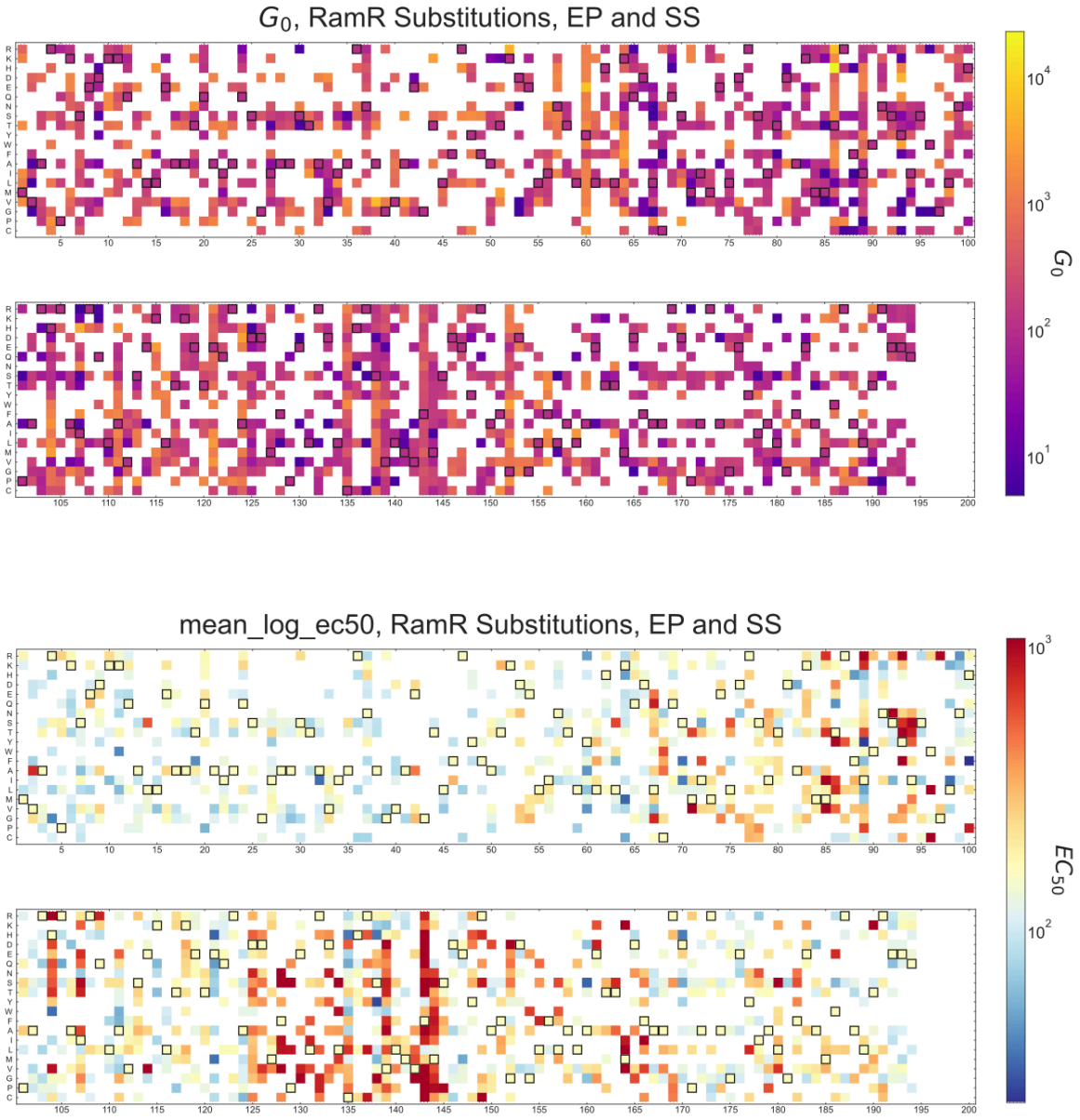


Figure S22. Background signal (G_0) and average sensitivity (geometric mean EC_{50}) of single substitutions.

Results are from both error-prone and site-saturation libraries. Outlined boxes represent the wild-type amino acid at each position. Blank (white) entries indicate non-observed substitutions.

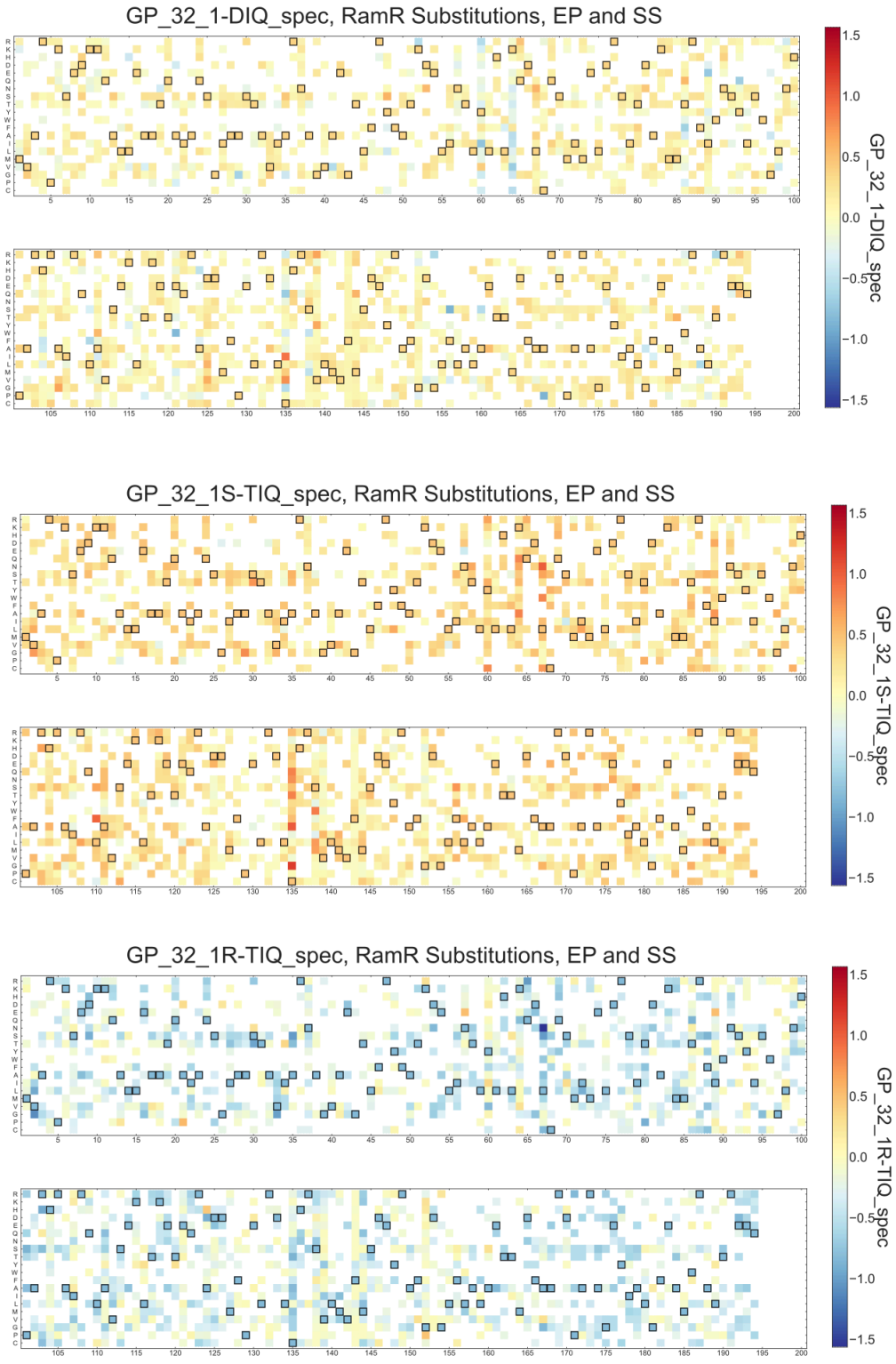


Figure S23. Specificity of single substitutions.

Outlined boxes represent the wild-type amino acid at each position. Results are from both error-prone and site-saturation libraries. Blank (white) entries indicate non-observed substitutions.

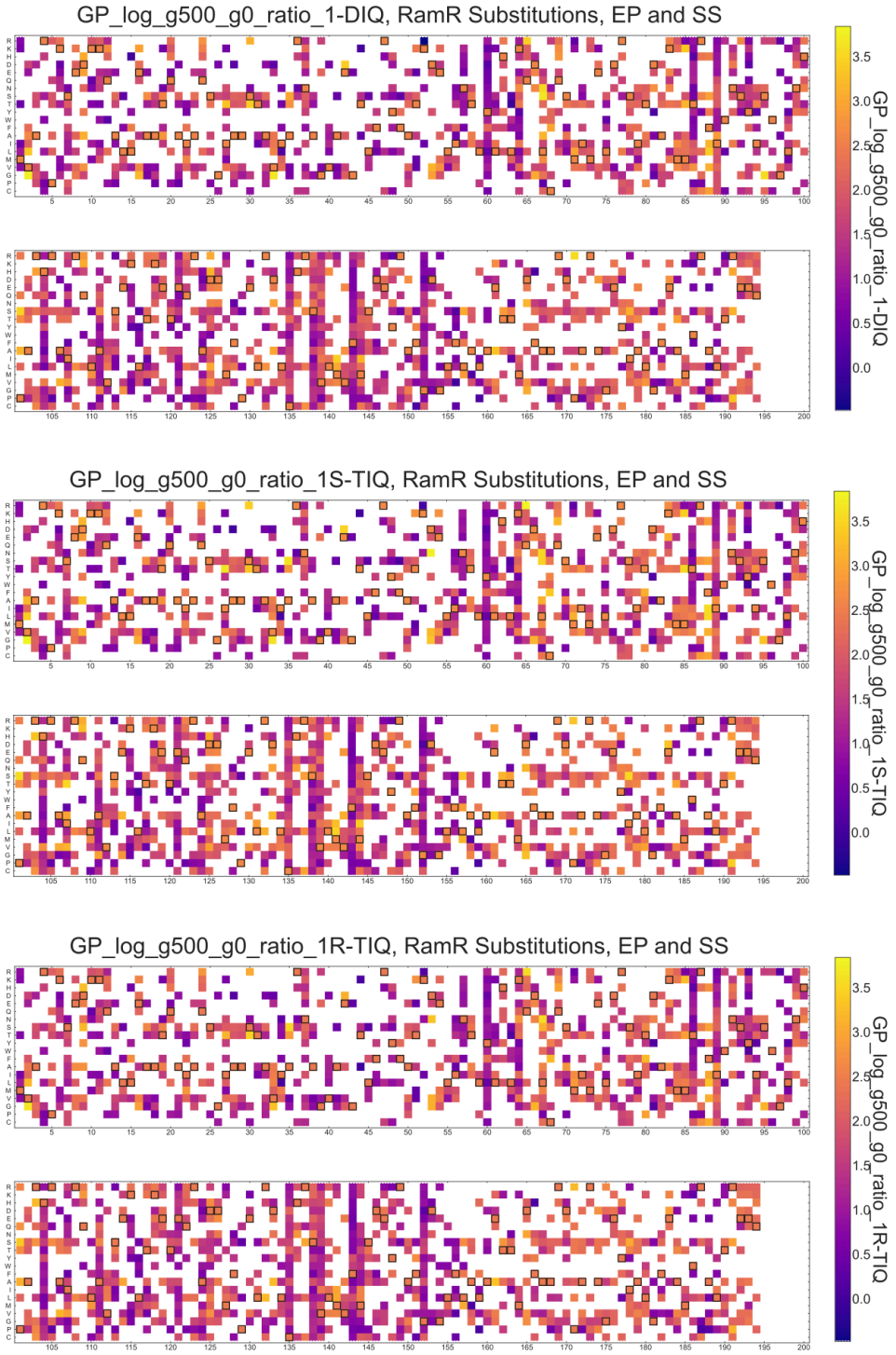


Figure S24. Fold change of single substitutions.

Outlined boxes represent the wild-type amino acid at each position. Results are from both error-prone and site-saturation libraries. Blank (white) entries indicate non-observed substitutions.

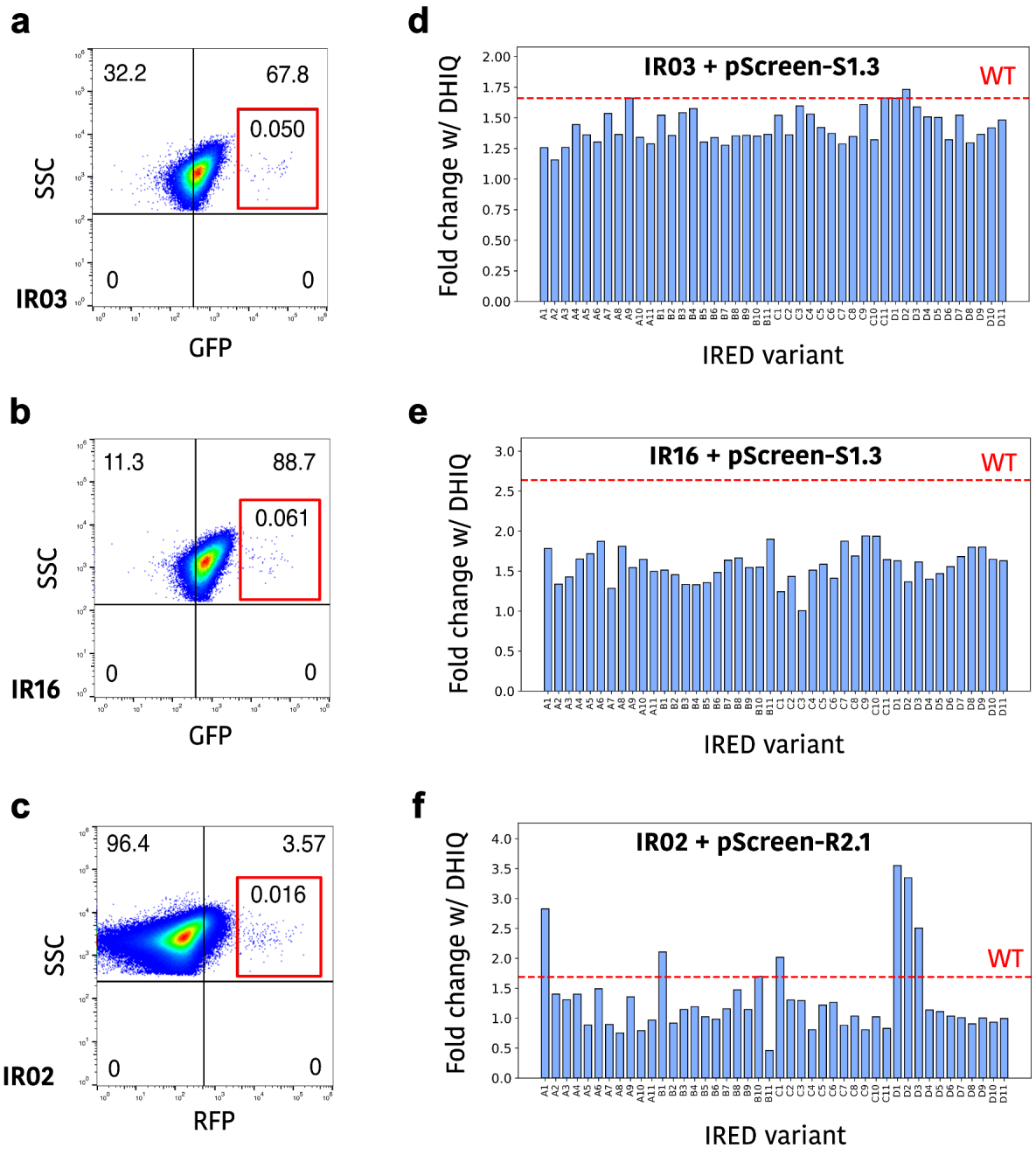


Figure S25. FACS gating and secondary screen of IR02, IR03, and IR17 templates. (a-c) Gating strategy used to sort the most fluorescent cells from the IR02, IR03, and IR17 libraries co-transformed with either pScreen-S1.3 (IR03 & IR16) or pScreen-R2.1 (IR02). (d-f) Secondary screen of IRED mutants shown as the fold change in fluorescence of strains in the presence of 100 μmol/L of DHIQ versus the fluorescence in the absence of DHIQ. The fold change in fluorescence produced by the template IRED enzyme is shown as a dotted red line. Measurements were performed in biological singlicate.

Supplementary Tables

Structure	S2.3 + S-THIQ	S2.4 + S-THIQ	D2.1	D2.3 + DHIQ	R2.1 +R-THIQ	R2.2
PDB Entry	**9DN8**	**9DN9**	**9DNC**	**9DNH**	**9DNK**	**9DNL**
Data collection						
Space group	$P4_12_12$	P1	$P2_12_12$	$P6_522$	C2	$P2_12_12_1$
Cell dimensions						
a, b, c (Å)	64.46, 64.46, 100.31	44.17, 54.36, 91.54	90.82, 150.40, 39.73	57.11, 57.11, 449.90	58.69, 93.39, 44.61	94.68, 95.22, 130.15
α, β, γ (°)	90.00, 90.00, 90.00	75.74, 81.98, 89.93	90.00, 90.00, 90.00	90.00, 90.00, 120.00	90.00, 93.76, 90.00	90.00, 90.00, 90.00
Resolution (Å)	50.15-2.12 (2.18-2.12)*	40.89-1.80 (1.85-1.80)	43.89-1.90 (1.95-1.90)	449.90-2.37 (2.41-2.37)	49.57-2.50 (2.54-2.50)	65.07-4.01 (4.11-4.01)
R_{sym} / R_{pim}	0.137 (1.496) / 0.038(0.418)	0.074 (0.650) / 0.067 (0.589)	0.069 (1.226) / 0.030 (0.516)	0.336 (3.174) / 0.077 (0.705)	0.096 (1.010) / 0.068 (0.762)	0.316 (1.308) / 0.165 (0.926)
CC ½ [Y]	0.999 (0.668)	0.994 (0.452)	0.998 (0.532)	0.976 (0.424)	0.990 (0.294)	0.967 (0.393)
I / σ	14.4 (1.9)	7.0 (1.5)	13.2 (1.5)	7.8 (0.9)	12.3 (1.3)	3.0 (0.8)
Completeness (%)	100.0 (100.0)	94.4 (94.7)	99.8 (100.0)	100.0 (99.7)	97.6 (96.5)	92.3 (92.0)
Redundancy	13.8 (13.5)	2.0 (2.0)	6.4 (6.6)	20.0 (21.0)	2.8 (2.8)	4.2 (2.5)
Refinement						
Resolution (Å)	45.58-2.12 (2.20-2.12)	40.89-1.80 (1.86-1.80)	43.89-1.90 (1.97-1.90)	49.46-2.37 (2.45-2.37)	49.62-2.50 (2.59-2.50)	65.08-4.01 (4.15-4.01)
No. reflections	12566 (1209)	71458 (7173)	43742 (4284)	19096 (1796)	7968 (724)	9361 (892)
R_{work}	0.1983 (0.2867)	0.2352 (0.3650)	0.1968 (0.3134)	0.2671 (0.3731)	0.2267 (0.3643)	0.2525 (0.3437)
R_{free} [±]	0.2362 (0.3466)	0.2636 (0.3900)	0.2157 (0.3486)	0.3261 (0.3952)	0.2674 (0.4143)	0.2947 (0.4006)
No. atoms	1574	6337	3178	2908	1489	5728
Protein	1483	5877	2988	2876	1473	5728
Ligand/ion	42	84	2	16	16	0
Water	49	376	188	42	0	0

B-factors (Å²)						
Protein	41.5	30.7	45.7	62.8	80.7	122.8
Ligand/ion	42	31.2	43.8	37.4	80.3	-
Water	46	37.2	49.5	55.7	-	-
R.m.s. deviations						
Bond lengths (Å)	0.0021	0.0046	0.0148	0.0057	0.0086	0.0039
Bond angles (°)	0.48	0.79	1.22	1.02	1.09	0.70
Ramachandran plot						
Favored	100.00%	99.59%	99.45%	99.17%	98.37%	99.44%
Allowed	0.00%	0.41%	0.55%	0.83%	1.63%	0.56%
Outliers	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Molprobity score^	1.51 / 98th percentile	1.73 / 83rd percentile	1.63 / 91st percentile	2.33 / 78th percentile	2.05 / 94th percentile	1.37 / 100th percentile

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

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

$^{\pm}R_{free}$ 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.

Table S1. X-ray Crystallography Data Collection and Refinement Statistics

Enantioselectivity	Alias	Accession ID	Enantioselectivity	Alias	Accession ID
S-THIQ	IR01	AUD39529.1	**S-THIQ**	IR24	WP_037254083.1
S-THIQ	IR02	AUD39583.1	**R-THIQ**	IR25	WP_111548309.1
R-THIQ	IR03	EMF01349.1	**R-THIQ**	IR26	WP_112256886.1
R-THIQ	IR04	NUT93668.1	**R-THIQ**	IR27	WP_116022473.1
S-THIQ	IR05	WP_009080771.1	**R-THIQ**	IR28	WP_220260546.1
No activity	IR06	WP_013019548.1	**R-THIQ**	IR29	WP_230421476.1
Mixed	IR07	WP_083466898.1	**R-THIQ**	IR30	WP_240598438.1
No activity	IR08	WP_123102485.1	**S-THIQ**	IR31	GHE75528.1
S-THIQ	IR09	WP_123993384.1	**R-THIQ**	IR32	MCP2166807.1
Mixed	IR10	WP_132480478.1	**No activity**	IR33	PFG99428.1
R-THIQ	IR11	WP_232847731.1	**S-THIQ**	IR34	WP_029384055.1
Mixed	IR12	WP_240033132.1	**S-THIQ**	IR35	WP_073484372.1
No activity	IR13	WP_253870842.1	**No activity**	IR36	WP_073486480.1
Mixed	IR14	WP_277212655.1	**R-THIQ**	IR37	WP_084468547.1
R-THIQ	IR15	MF540792.1	**Mixed**	IR38	WP_172381658.1
R-THIQ	IR16	AUD39487.1	**Mixed**	IR39	WP_184965005.1
R-THIQ	IR17	ATL31837.1	**No activity**	IR40	WP_187234818.1
No activity	IR18	ATW48861.1	**No activity**	IR41	WP_189878949.1
R-THIQ	IR19	AUD39517.1	**Mixed**	IR42	WP_191257941.1
R-THIQ	IR20	EME99395.1	**S-THIQ**	IR43	WP_229347914.1
R-THIQ	IR21	OMI34305.1	**Mixed**	IR44	WP_242906400.1
R-THIQ	IR22	PRX45020.1	**Mixed**	IR45	WP_253669224.1
No activity	IR23	WP_030629722.1	**S-THIQ**	IR46	WP_266533875.1

Table S2. Identifiers and measured enantioselectivity for all IREDs. Enantioselectivity is colored green or orange based on forming S-THIQ or R-THIQ as >70% of the product measured via HPLC, respectively. Otherwise, active IREDs that form <70% of either enantiomer are colored yellow and labeled “Mixed”. IREDs that generate no detectable product in our reaction conditions with our LC/MS system are labeled as “No activity” and colored gray.

Supplementary Data

Supplementary Data 1.DMS results.csv This CSV file contains the processed output data for the DMS results. Each row in the file corresponds to a RamR variant measured in the DMS experiments. The columns in the file contain the following data:

variant: the variant identifier for each row
library: the library in which the variant was measured, EP = random mutagenesis, SS = site-saturation mutagenesis, Shuffle = gene shuffling
mutation_codes: the amino acid substitution codes for each variant (where applicable), blank entry indicates wild-type RamR
total_counts: the total number of barcode read counts, across all time points
total_counts_plate_2: the number of barcode read count for the first time point (growth plate 2)
nanopore_count: the total number of nanopore read counts
pacbio_count: the total number of PacBio CCS read counts
amino_type_count_f: the number of distinct barcodes found for the variant
ramR_amino_seq: the amino acid sequence for the variant

log_g0: the base-10 logarithm of the basal output (G_0) level, inferred from the Hill equation model
log_g0_err: one-standard-deviation posterior uncertainty

log_ginf_1-DIQ: the base-10 logarithm of the saturated output in response to DHIQ ($G_{\infty,\mathrm{DHIQ}}$), inferred from the Hill equation model
log_ginf_1-DIQ_err: one-standard-deviation posterior uncertainty
log_ec50_1-DIQ: the base-10 logarithm of the sensitivity in response to DHIQ ($EC_{50,\mathrm{DHIQ}}$), inferred from the Hill equation model
log_ec50_1-DIQ_err: one-standard-deviation posterior uncertainty
n_1-DIQ: the effective Hill coefficient in response to DHIQ (n_{DHIQ}), inferred from the Hill equation model
n_1-DIQ_err: one-standard-deviation posterior uncertainty
log_ginf_g0_ratio_1-DIQ: the base-10 logarithm of the ratio of the saturated to basal output (G_∞/G_0) in response to DHIQ, inferred from the Hill equation model
log_ginf_g0_ratio_1-DIQ_err: one-standard-deviation posterior uncertainty
log_g32_g0_ratio_1-DIQ: the base-10 logarithm of the ratio of the output at 32 μmol/L to the basal output (G_{32}/G_0) in response to DHIQ, inferred from the Gaussian process model
log_g32_g0_ratio_1-DIQ_err: one-standard-deviation posterior uncertainty
g32_spec_1-DIQ: the specificity for DHIQ, calculated using the log_g32_g0_ratio_ values for the three different ligands
g32_spec_1-DIQ_err: one-standard-deviation posterior uncertainty
log_g500_g0_ratio_1-DIQ: the base-10 logarithm of the ratio of the output at 500 μmol/L DHIQ to the basal output ($G_{500,\mathrm{DHIQ}}/G_0$), inferred from the Gaussian process model
log_g500_g0_ratio_1-DIQ_err: one-standard-deviation posterior uncertainty
log_g500_1-DIQ: the base-10 logarithm of the the output at 500 μmol/L DHIQ ($G_{500,\mathrm{DHIQ}}$), inferred from the Gaussian process model

log_g500_1-DIQ_err: one-standard-deviation posterior uncertainty
hill_invert_prob_1-DIQ: the posterior probability that the variant has an inverted dose-response for DHIQ

log_ginf_1R-TIQ: the base-10 logarithm of the saturated output in response to R-THIQ ($G_{\infty,\text{R-THIQ}}$), inferred from the Hill equation model
log_ginf_1R-TIQ_err: one-standard-deviation posterior uncertainty
log_ec50_1R-TIQ: the base-10 logarithm of the sensitivity in response to R-THIQ ($EC_{50,\text{R-THIQ}}$), inferred from the Hill equation model
log_ec50_1R-TIQ_err: one-standard-deviation posterior uncertainty
n_1R-TIQ: the effective Hill coefficient in response to R-THIQ ($n_{\text{R-THIQ}}$), inferred from the Hill equation model
n_1R-TIQ_err: one-standard-deviation posterior uncertainty
log_ginf_g0_ratio_1R-TIQ: the base-10 logarithm of the ratio of the saturated to basal output (G_∞/G_0) in response to R-THIQ, inferred from the Hill equation model
log_ginf_g0_ratio_1R-TIQ_err: one-standard-deviation posterior uncertainty
log_g32_g0_ratio_1R-TIQ: the base-10 logarithm of the ratio of the output at 32 μmol/L to the basal output (G_{32}/G_0) in response to R-THIQ, inferred from the Gaussian process model
log_g32_g0_ratio_1R-TIQ_err: one-standard-deviation posterior uncertainty
g32_spec_1R-TIQ: the specificity for R-THIQ, calculated using the log_g32_g0_ratio_ values for the three different ligands
g32_spec_1R-TIQ_err: one-standard-deviation posterior uncertainty
log_g500_g0_ratio_1R-TIQ: the base-10 logarithm of the ratio of the output at 500 μmol/L R-THIQ to the basal output ($G_{500,\text{R-THIQ}}/G_0$), inferred from the Gaussian process model
log_g500_g0_ratio_1R-TIQ_err: one-standard-deviation posterior uncertainty
log_g500_1R-TIQ: the base-10 logarithm of the the output at 500 μmol/L R-THIQ ($G_{500,\text{R-THIQ}}$), inferred from the Gaussian process model
log_g500_1R-TIQ_err: one-standard-deviation posterior uncertainty
hill_invert_prob_1R-TIQ: the posterior probability that the variant has an inverted dose-response for R-THIQ

log_ginf_1S-TIQ: the base-10 logarithm of the saturated output in response to S-THIQ ($G_{\infty,\text{S-THIQ}}$), inferred from the Hill equation model
log_ginf_1S-TIQ_err: one-standard-deviation posterior uncertainty
log_ec50_1S-TIQ: the base-10 logarithm of the sensitivity in response to S-THIQ ($EC_{50,\text{S-THIQ}}$), inferred from the Hill equation model
log_ec50_1S-TIQ_err: one-standard-deviation posterior uncertainty
n_1S-TIQ: the effective Hill coefficient in response to S-THIQ ($n_{\text{S-THIQ}}$), inferred from the Hill equation model
n_1S-TIQ_err: one-standard-deviation posterior uncertainty
log_ginf_g0_ratio_1S-TIQ: the base-10 logarithm of the ratio of the saturated to basal output (G_∞/G_0) in response to S-THIQ, inferred from the Hill equation model
log_ginf_g0_ratio_1S-TIQ_err: one-standard-deviation posterior uncertainty
log_g32_g0_ratio_1S-TIQ: the base-10 logarithm of the ratio of the output at 32 μmol/L to the basal output (G_{32}/G_0) in response to S-THIQ, inferred from the Gaussian process model
log_g32_g0_ratio_1S-TIQ_err: one-standard-deviation posterior uncertainty
g32_spec_1S-TIQ: the specificity for S-THIQ, calculated using the log_g32_g0_ratio_ values for the three different ligands

g32_spec_1S-TIQ_err: one-standard-deviation posterior uncertainty
log_g500_g0_ratio_1S-TIQ: the base-10 logarithm of the ratio of the output at 500 μmol/L S-THIQ to the basal output ($G_{500,\text{S-THIQ}}/G_0$), inferred from the Gaussian process model
log_g500_g0_ratio_1S-TIQ_err: one-standard-deviation posterior uncertainty
log_g500_1S-TIQ: the base-10 logarithm of the the output at 500 μmol/L S-THIQ ($G_{500,\text{S-THIQ}}$), inferred from the Gaussian process model
log_g500_1S-TIQ_err: one-standard-deviation posterior uncertainty
hill_invert_prob_1S-TIQ: the posterior probability that the variant has an inverted dose-response for S-THIQ
spec_1-DIQ: an alternate specificity metric for DHIQ, calculated using the log_ec50_ values for the three different ligands
spec_1R-TIQ: an alternate specificity metric for R-THIQ, calculated using the log_ec50_ values for the three different ligands
spec_1S-TIQ: an alternate specificity metric for S-THIQ, calculated using the log_ec50_ values for the three different ligands

Supplementary Data 2.Cytometry results.csv This CSV file contains the processed output data for the Cytometry results used to calibrate and verify the DMS results. Each row in the file corresponds to a RamR variant and ligand measured flow cytometry. The columns in the file contain the following data:

variant: the identifier for the RamR variant
ligand: the ligand used in the cytometry measurement
log_g0: the base-10 logarithm of the basal output (G_0) level
log_g0_err: one-standard-deviation posterior uncertainty
log_ginf: the base-10 logarithm of the saturated output in response to the ligand ($G_{\infty,\text{lig}}$)
log_ginf_err: one-standard-deviation posterior uncertainty
log_ec50: the base-10 logarithm of the sensitivity in response to the ligand ($EC_{50,\text{lig}}$)
log_ec50_err: one-standard-deviation posterior uncertainty
n: the effective Hill coefficient in response to the ligand (n_{lig})
n_err: one-standard-deviation posterior uncertainty
log_ginf_g0_ratio: the base-10 logarithm of the ratio of the saturated to basal output (G_∞/G_0) in response to the ligand
log_ginf_g0_ratio_err: one-standard-deviation posterior uncertainty
ec50_spec: the specificity for the ligand, calculated using the log_ec50 values for the three different ligands
ec50_spec_err: one-standard-deviation posterior uncertainty
g32_spec: the specificity for the ligand, calculated using the log_g32_g0_ratio values for the three different ligands
g32_spec_err: one-standard-deviation posterior uncertainty
log_g32_g0_ratio: the base-10 logarithm of the ratio of the output at 32 μmol/L to the basal output ($G_{32,\text{lig}}/G_0$) in response to the ligand
log_g32_g0_ratio_err: one-standard-deviation posterior uncertainty
log_g500: the base-10 logarithm of the the output at 500 μmol/L ligand concentration ($G_{500,\text{lig}}$)
log_g500_err: one-standard-deviation posterior uncertainty

log_g500_g0_ratio: the base-10 logarithm of the ratio of the output at 500 μmol/L to the basal output ($G_{500,lig}/G_0$) in response to the ligand

log_g500_g0_ratio_err: one-standard-deviation posterior uncertainty