Science Advances
AAAS

Supplementary Materials for

Epistasis between promoter activity and coding mutations shapes gene evolvability

Angel F. Cisneros *et al.*

Corresponding author: Christian R. Landry, christian.landry@bio.ulaval.ca

Sci. Adv. **9**, eadd9109 (2023)
DOI: 10.1126/sciadv.add9109

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

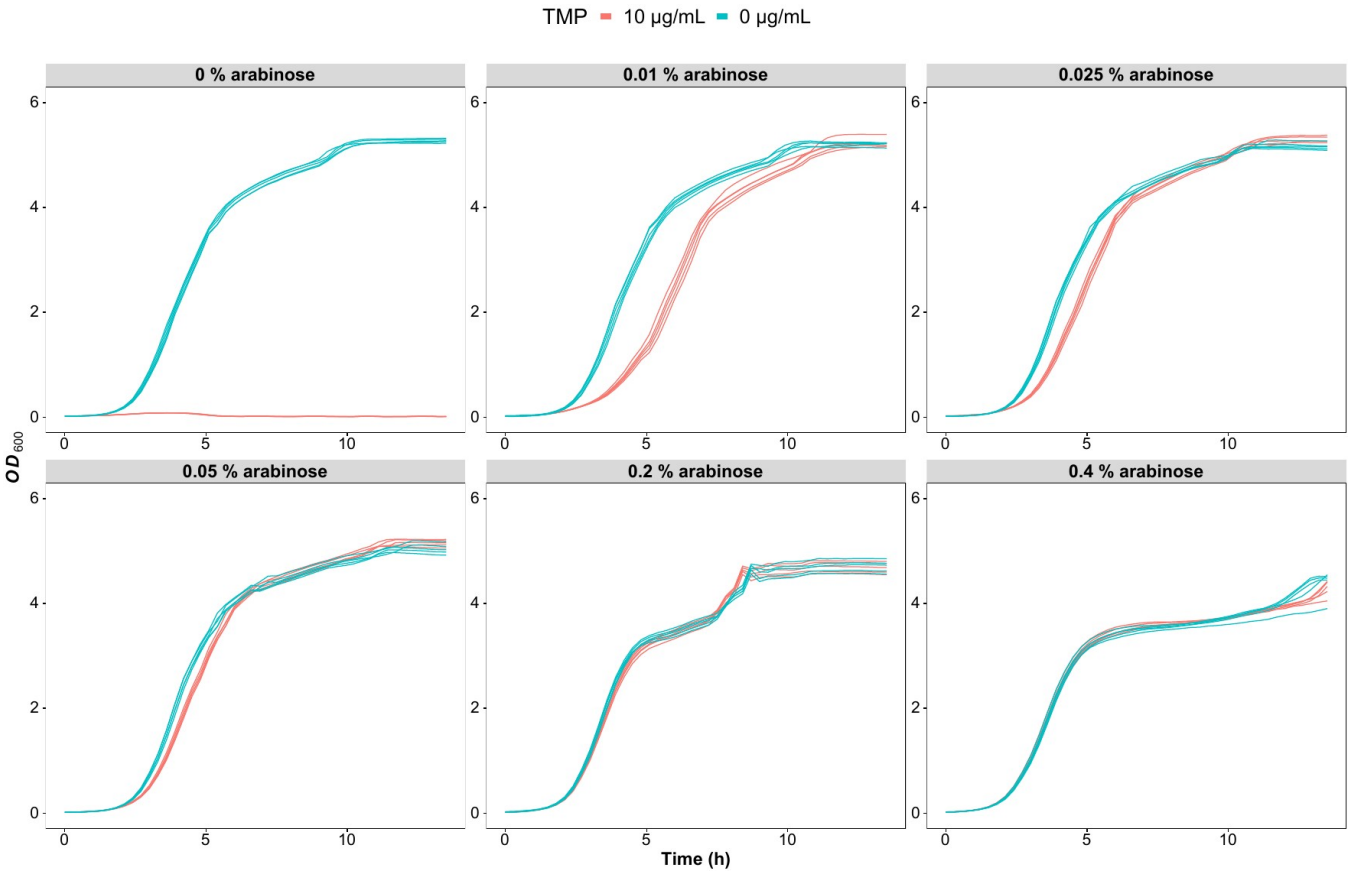


Fig. S1. Dose-dependent growth recovery upon ecDHFR inhibition and DfrB1 expression. Cells expressing WT DfrB1 under the control of the arabinose-inducible promoter were grown in LB medium with and without trimethoprim (TMP) and at different concentrations of arabinose as indicated above each panel. Same data as in Fig. 1C. Growth recovery of the inhibition of the endogenous ecDHFR by TMP through the expression of the DfrB1 can be calculated as the difference in the area under the curve for a given arabinose concentration.

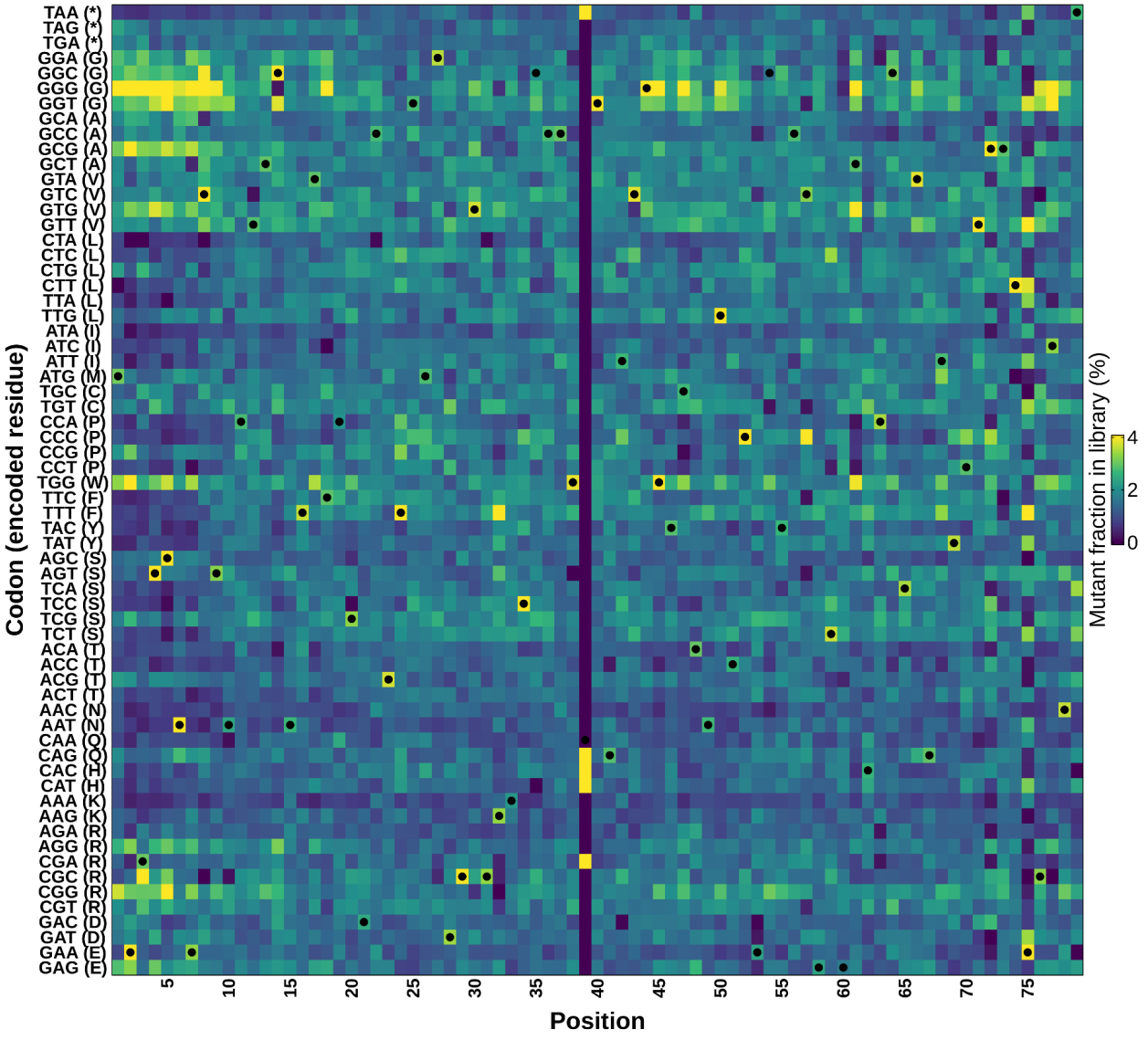


Fig. S2. Quality control of the DfrB1 Deep Mutational Scanning (DMS) library. The libraries corresponding to each position (NNN codons) were pooled and sequenced on a MiSeq PE300 (268 671 reads) to verify the coverage of the various codons across positions. A single position was not covered and was repeated and added to the final library. The color scale represents the percentage of reads that mapped to a mutant out of the library for that position. WT codons are labeled with dots.

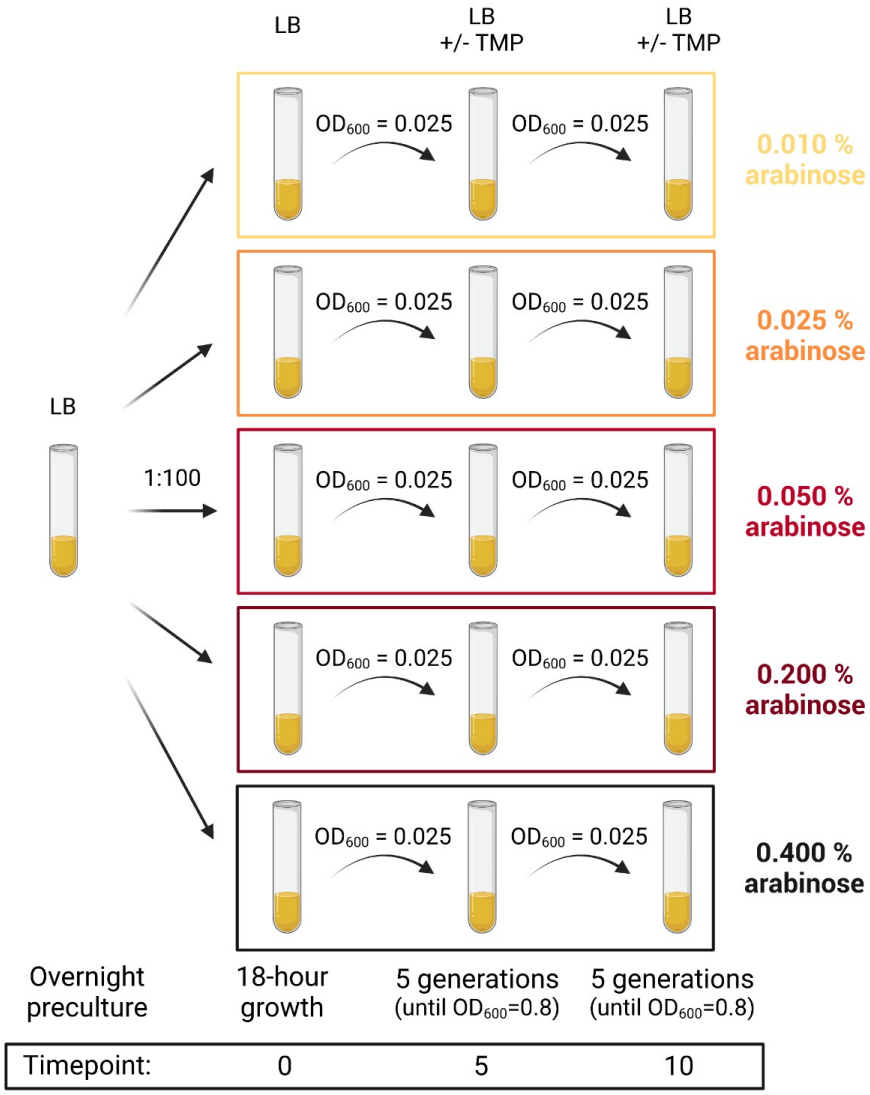
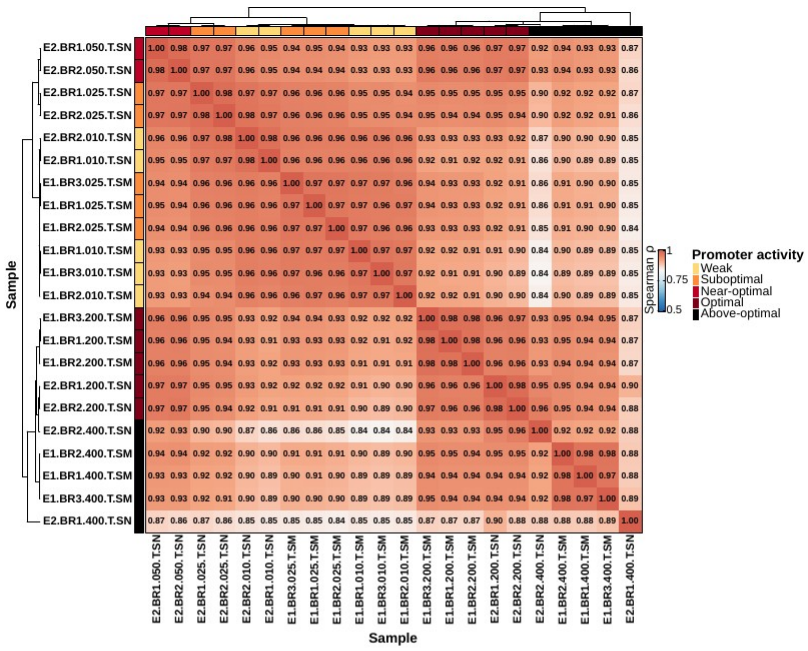


Fig. S3. Experimental setup for the bulk competition assays. Samples for the bulk competition experiments were taken from an overnight preculture of cells transformed with the DMS plasmid library. Cells were grown for five generations in LB medium with different concentrations of arabinose and with or without trimethoprim (TMP) until they reached an optical density of 0.8. They were then diluted again and grown for five more generations. DNA was extracted and sequenced at t = 0 and t = 10 generations. This experiment was repeated twice (see Table S1 for details).



Experiment with TMP	ID	Biological replicates	Arabinose concentration (%)	Sequencing platform
1	E1.BR1.010.T.SM	1	0.01	MiSeq
	E1.BR2.010.T.SM	2		
	E1.BR3.010.T.SM	3		
	E1.BR1.025.T.SM	1	0.025	MiSeq
	E1.BR2.025.T.SM	2		
	E1.BR3.025.T.SM	3		
	E1.BR1.200.T.SM	1	0.2	MiSeq
	E1.BR2.200.T.SM	2		
	E1.BR3.200.T.SM	3		
	E1.BR1.400.T.SM	1	0.4	MiSeq
	E1.BR2.400.T.SM	2		
	E1.BR3.400.T.SM	3		
2	E2.BR1.010.T.SN	1	0.01	NovaSeq
	E2.BR2.010.T.SN	2		
	E2.BR1.025.T.SN	1	0.025	NovaSeq
	E2.BR2.025.T.SN	2		
	E2.BR1.050.T.SN	1	0.05	NovaSeq
	E2.BR2.050.T.SN	2		
	E2.BR1.200.T.SN	1	0.2	NovaSeq
	E2.BR2.200.T.SN	2		
	E2.BR1.400.T.SN	1	0.4	NovaSeq
	E2.BR2.400.T.SN	2		

Fig. S4. Replicates of the bulk competition experiment with selection for DfrB1 activity (with TMP) correlate strongly. Spearman correlation between selection coefficients was estimated for different replicates of the experiment with selection for DfrB1 activity using the antibiotic TMP. Samples were named according to the table (bottom) extracted from Table S1.

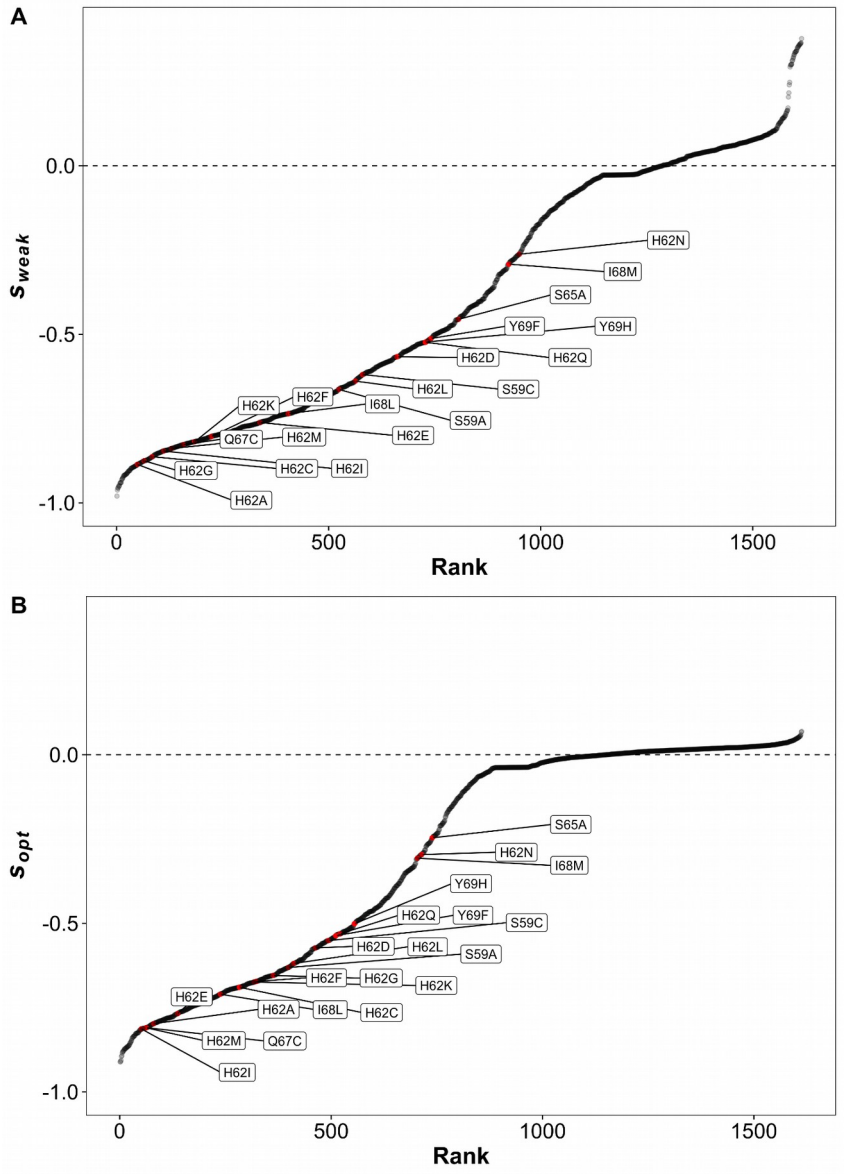


Fig. S5. Selection coefficients recapitulate previously characterized mutants with poor catalytic activity. Ranks and selection coefficients measured by bulk competition assay and sequencing for mutants constructed by Dam et al. (*78*) and Strader et al. (*24*) at a weak promoter activity level for WT DfrB1 (**A**) and at the optimal promoter activity level for WT DfrB1 (**B**). All mutants had lower activity than the WT in the experimental papers and have negative *s* here, as expected. Dashed horizontal lines indicate selection coefficients for the WT (s = 0).

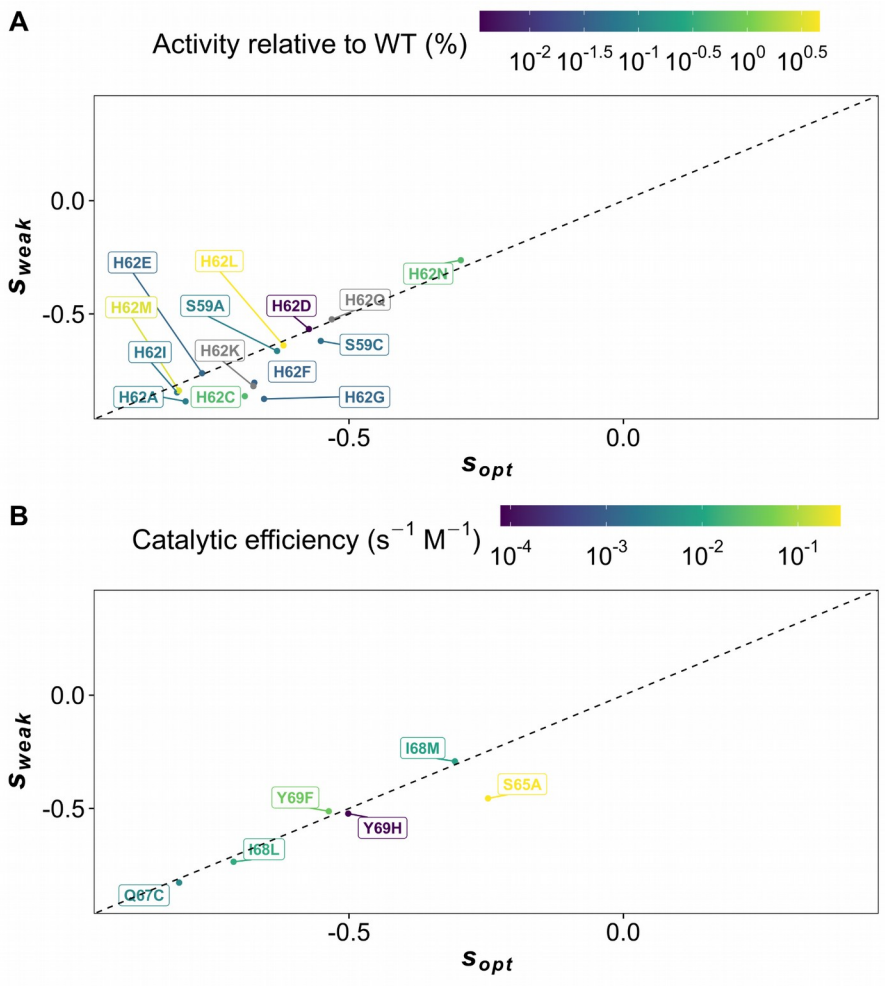


Fig. S6. The fitness effects of only some of the low activity mutants are improved at optimal promoter activity. (**A**) Differences in selection coefficients observed at weak and optimal promoter activity on mutants at the tetramerization interface from Dam et al. (*78*) with respect to their measured activity relative to WT. (**B**) Differences in selection coefficients observed at weak and optimal promoter activity on active site mutants from Strader et al. (*24*) with respect to their measured catalytic efficiency.

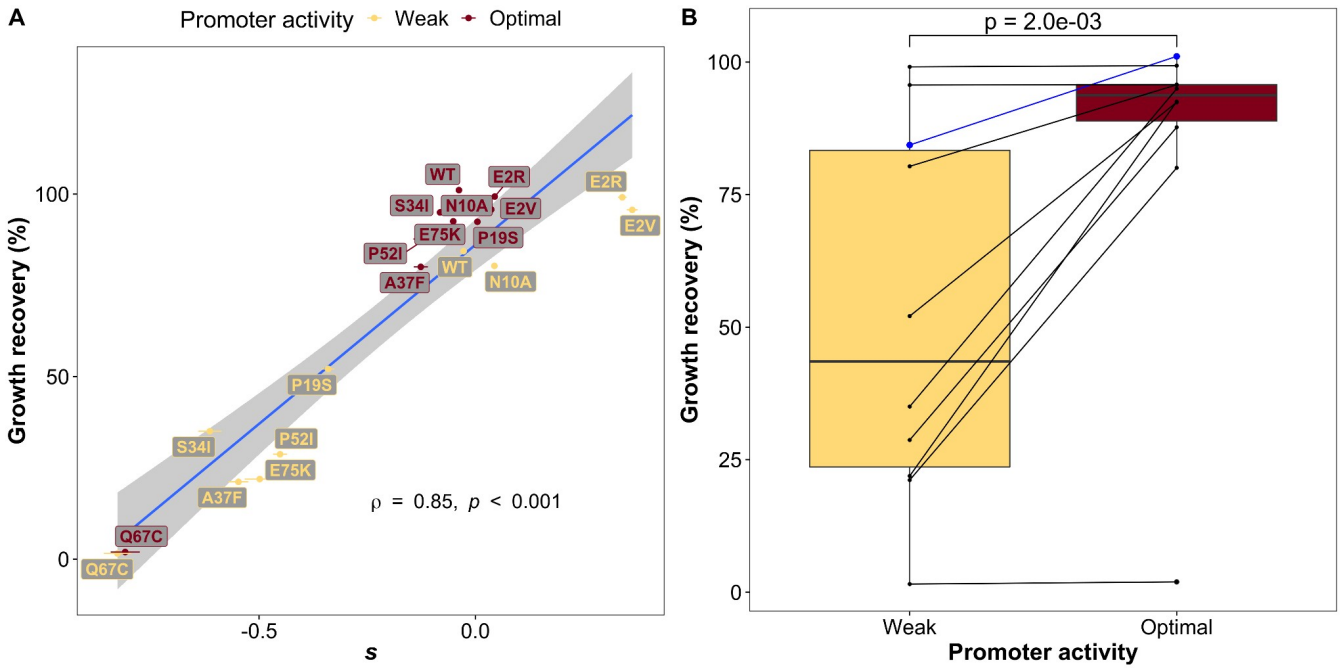


Fig. S7. The effects of mutations measured by bulk competition are recapitulated with individual growth measurements. Mutants with different magnitudes of changes in fitness effects across concentrations of arabinose were selected for validation in individual growth experiments. Growth was measured as the average of the area under the curve for two replicates per mutant at each of two promoter activity levels and subtracted to measure the % of growth recovered upon expression of the DfrB1. (**A**) Growth recovery of individual mutants correlates strongly with selection coefficients. The corresponding mutants and the WT are indicated next to each data point and the colors indicate promoter activities. (**B**) Promoter activity-level dependent differences in growth rate for the validated mutants on the left. WT is indicated in blue. P-values were calculated using Wilcoxon's test for differences in means of paired samples.

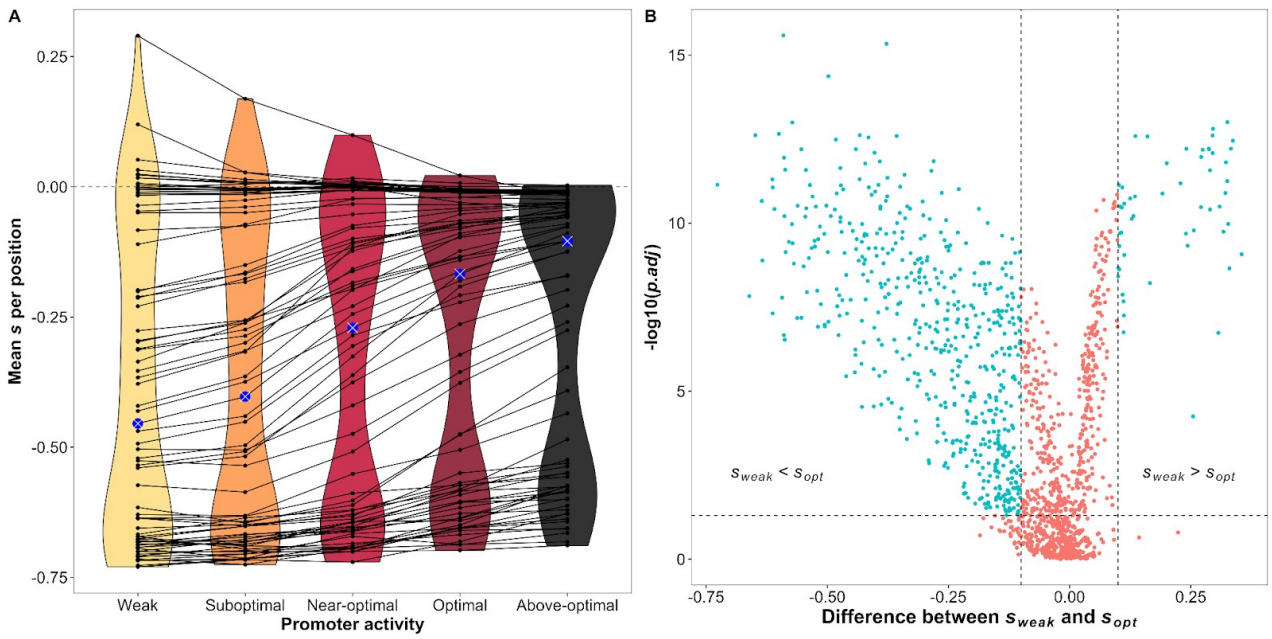


Fig. S8. Significance and magnitude of effects of promoter activity level on the fitness of mutants. (**A**) Averages of selection coefficients observed at each position of the DfrB1 coding sequence. The larger dot indicates the position with the median effect for each promoter activity level. Medians of each distribution are represented with the blue dots with white crosses. The dashed horizontal line indicates the selection coefficient for the WT (s = 0). (**B**) For each mutant, we performed an ANOVA testing for significant effects of promoter activity on fitness effects and corrected for multiple hypothesis testing by applying the Benjamini-Hochberg correction with a false discovery rate of 0.05 using the FDR estimation package (*74*). The x-axis shows the difference between s_{weak} and s_{opt} (s_{weak} - s_{opt}) and the y-axis shows the adjusted p-values from the statistical analysis. The horizontal dashed line indicates the cutoff for significance of adjusted p-values ($p < 0.05$) and the vertical lines indicate an arbitrary threshold for magnitude of changes in selection coefficients (greater than 0.1) due to promoter activity level.

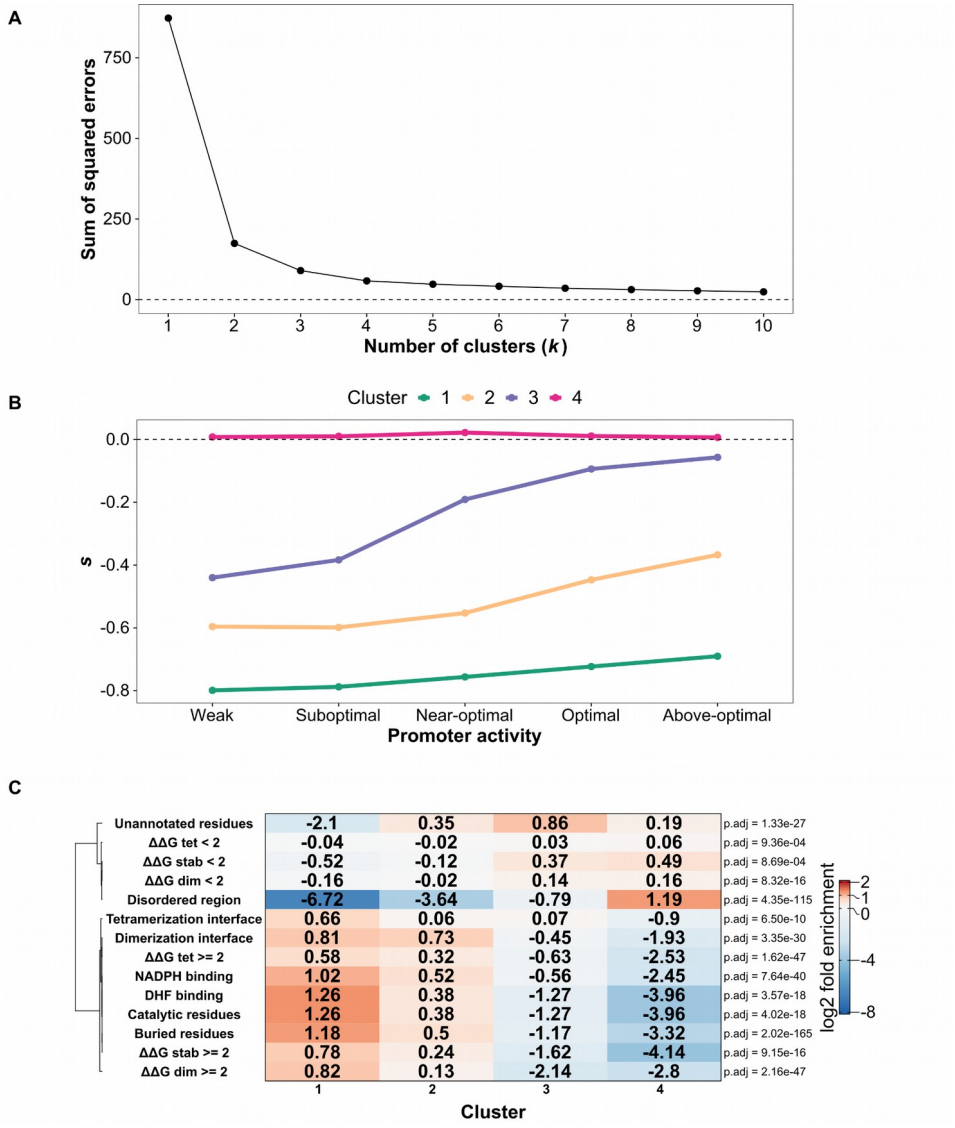


Fig. S9. k-means clusters capture the main patterns of promoter activity-dependent changes in fitness effects. (**A**) Sum of squared errors as a function of the numbers used in k-means clustering. Adding more clusters results in more accurate clustering but increases the probability of overfitting the model. We selected k = 4 because adding more than 4 clusters does not lead to significant decreases in the sum of squared errors. (**B**) Selection coefficients at each promoter activity level for the centroids of the four clusters. The dashed horizontal line indicates the selection coefficient for the WT (s = 0). (**C**) Relative enrichment of each of the four clusters with mutants at particular protein sites and with specific destabilizing effects. Stars indicate significant deviations from expectations as measured with a chi-squared test, adjusted for FDR ($p < 0.05$) with the Benjamini-Hochberg correction.

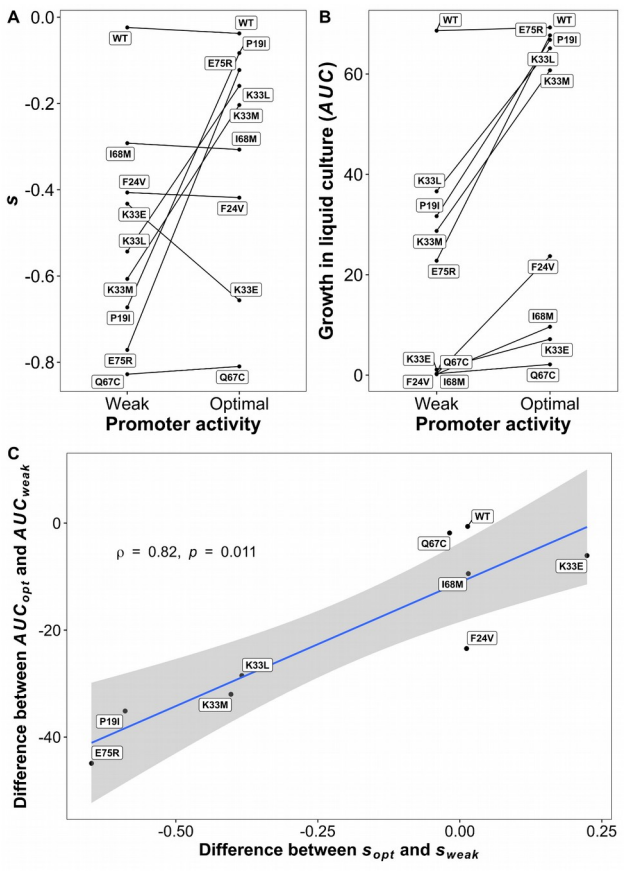


Fig. S10. Rank epistasis is not completely recapitulated in validation cultures. (**A**) Selection coefficients calculated from the bulk competition experiment for a selected group of mutations at weak and optimal promoter activity. Rank epistasis is observed as crossing lines, indicating that a mutation that is more deleterious in a particular condition becomes more beneficial in a different condition. (**B**) Area under the curve observed for validation cultures of the same selected group of mutations. Some patterns of the rank epistasis observed in the bulk competition experiment are recapitulated but not all of them. (**C**) The difference in area under the curve in the validation cultures generally correlates with the difference in selection coefficients from the bulk competition experiment. ΔAUC_{weak} was calculated as the differences in the area under the curve observed for growth in liquid cultures at weak and optimal promoter activity levels (AUC_{weak} - AUC_{opt}). Δs_{weak} was calculated as the differences in selection coefficients observed for each mutant in the bulk competition experiments at weak and optimal promoter activity levels (s_{weak} - s_{opt}).

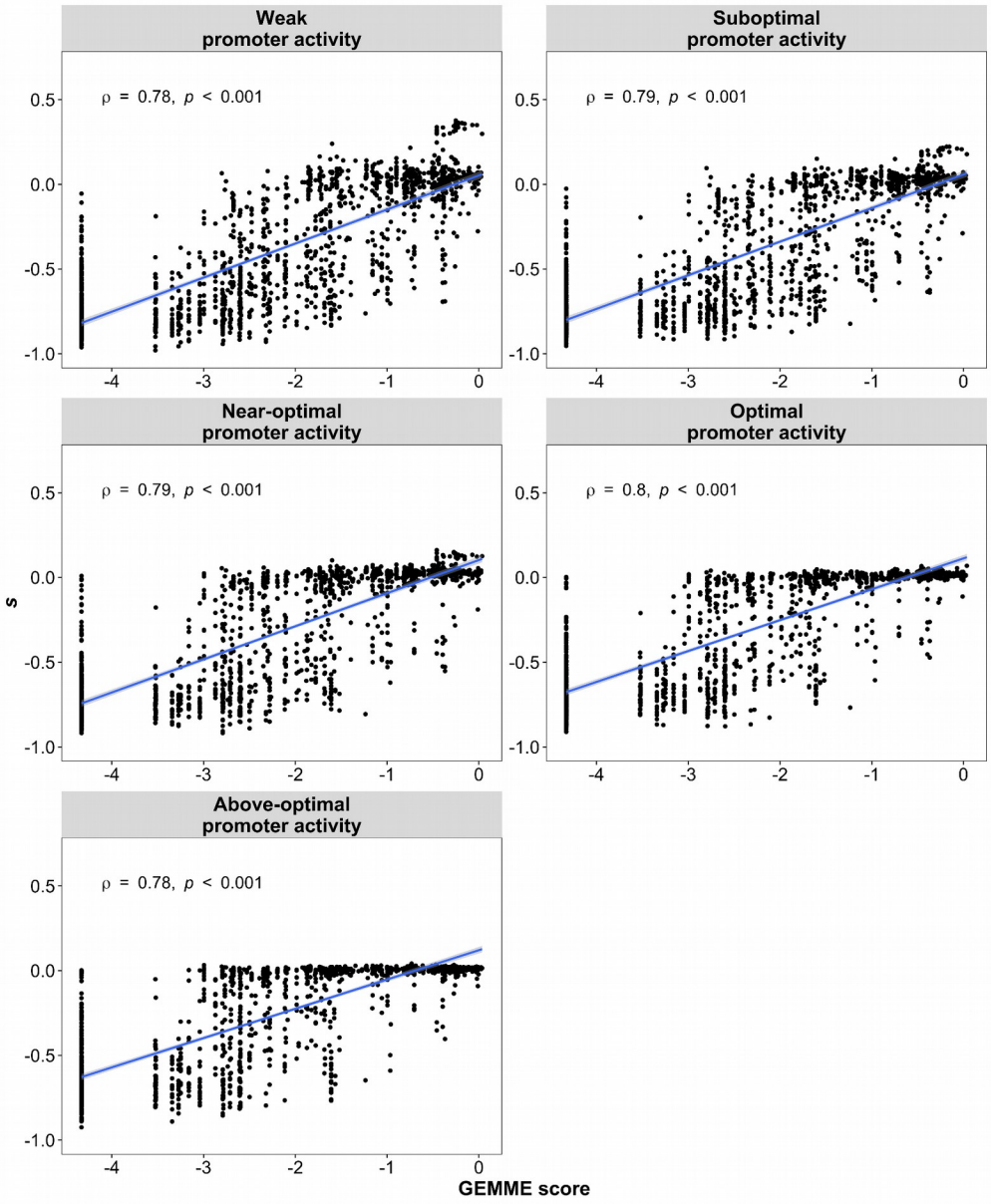


Fig. S11. Selection coefficients measured in the DMS bulk competition experiment with TMP correlate well with fitness effects deduced variation observed in natural sequences. We used GEMME (*20*) to predict mutational effects based on patterns of conservation and substitution. GEMME scores (0 for residues observed as often as the WT DfrB1 residue at that position and negative for residues observed less often) correlate well with the effects measured in the bulk competition assay with the mutant library. These analyses show that the mutational effects measured experimentally reflect the selective pressure these homologous sequences experience in nature.

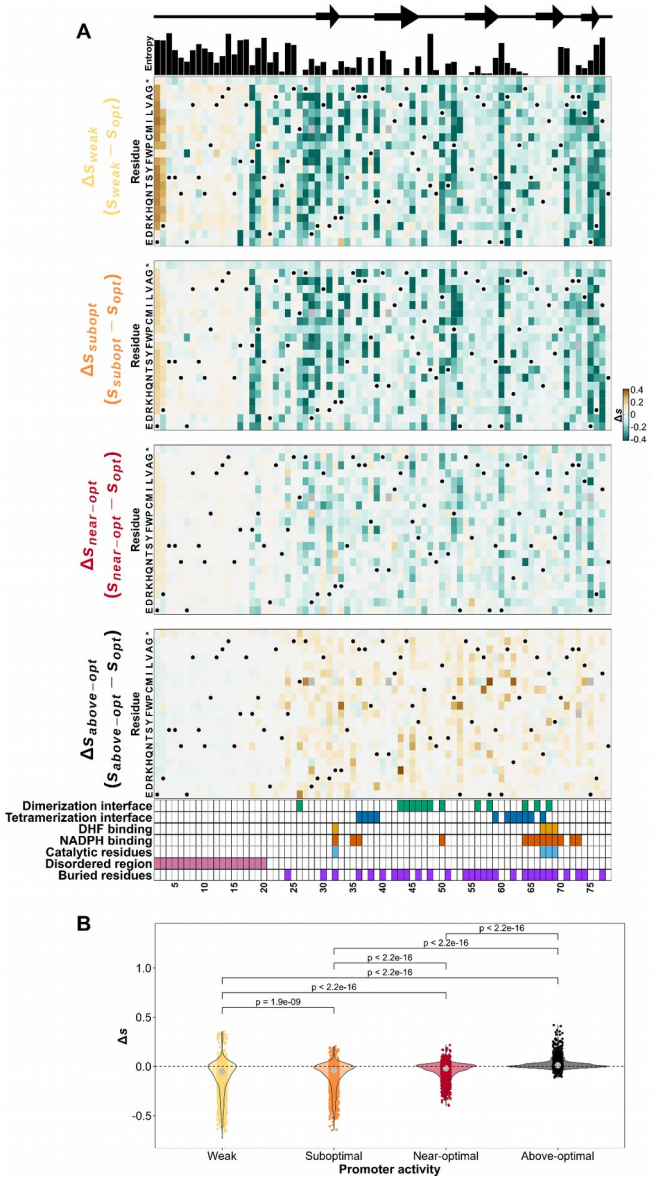


Fig. S12. Promoter activity-dependent differences in fitness effects become weaker as promoter activity level approaches the optimum. (**A**) Difference between the selection coefficients of an amino acid substitution at one of the non-optimal promoter activity levels and at the optimal promoter activity for the WT ($\Delta s_{non\text{-}opt} = s_{non\text{-}opt} - s_{opt}$). WT residues in panel A are labeled with dots. Entropy as a metric of variation in homologous sequences and secondary structures are shown at the top. Annotations at the bottom indicate residues that participate in the interfaces of the DfrB1 homotetramer, that come into contact with either the substrate (DHF) or the cofactor (NADPH), key catalytic residues, the disordered region, and buried sites. Interfaces are labeled as dimerization and tetramerization interfaces following (*19*). (**B**) Distributions of Δs for the different comparisons against s_{opt}. P-values were calculated using Wilcoxon's test for differences in means of paired samples.

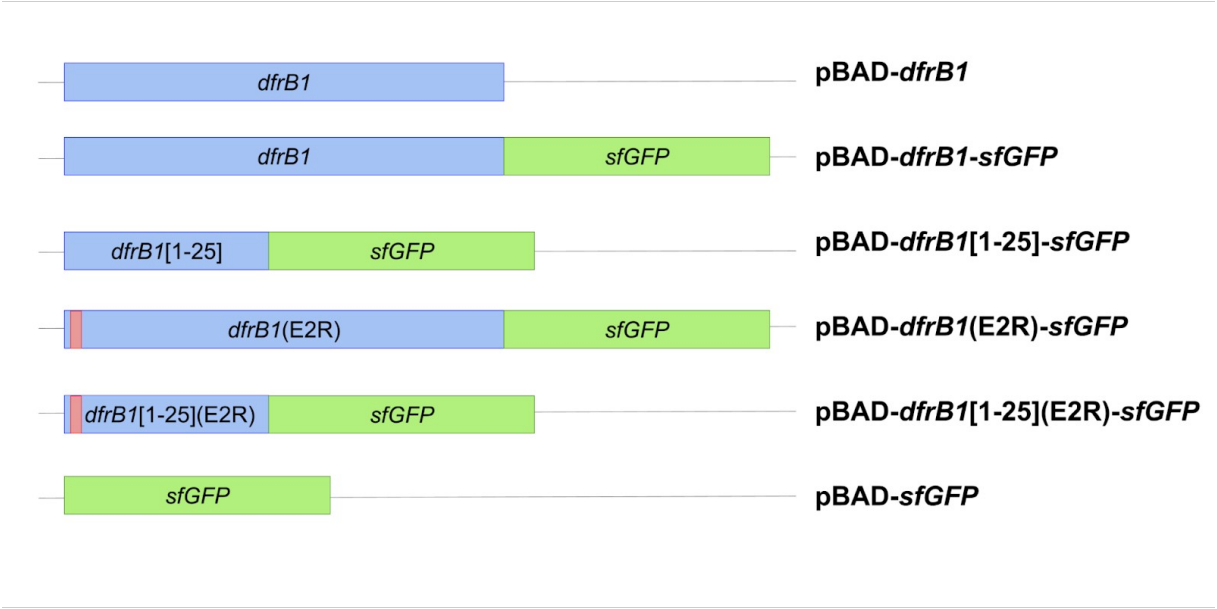


Fig. S13. Constructs used to evaluate the effect of E2R on protein abundance. The fluorescence of GFP fused to several constructs of DfrB1 was measured by flow cytometry (Fig. 4C). Constructs used include: 1) DfrB1 alone, 2) DfrB1 fused to GFP, 3) truncated DfrB1 (positions 1-25) fused to DfrB1, 4) E2R DfrB1 mutant fused to GFP, 5) E2R truncated DfrB1 mutant (positions 1-25) fused to GFP, 6) GFP alone.

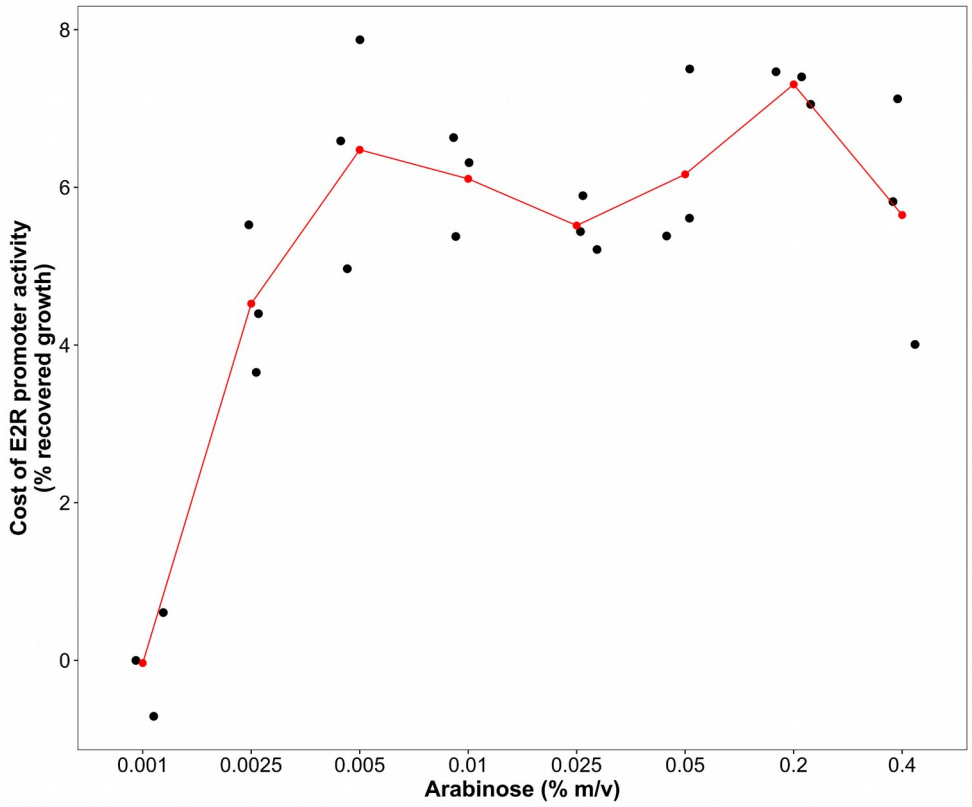


Fig. S14. Cost of promoter activity for WT DfrB1 and E2R mutant at different promoter activity levels. Cost of promoter activity for the E2R mutant in terms of percentage of recovered growth compared to the median recovery at 0.001 % arabinose, calculated using the same data from Fig. 4D. The red dots indicate the mean of three replicates for each arabinose concentration.

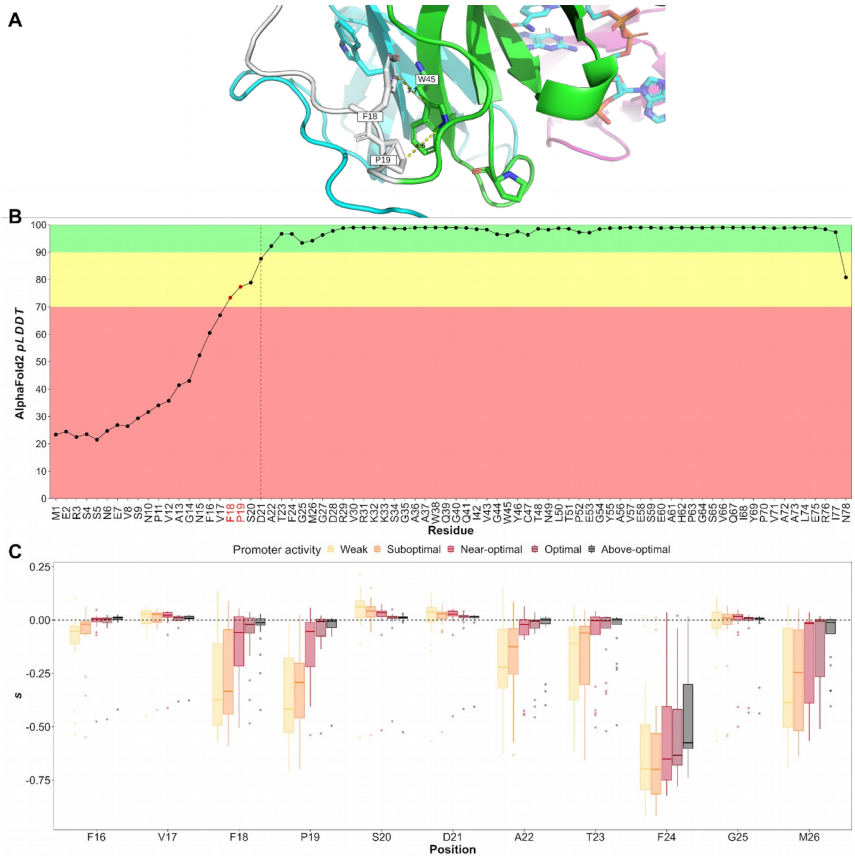


Fig. S15. Predicted interactions of W45 with F18 and P19. (**A**) Residues F18 and P19 are predicted to interact with W45 in an AlphaFold2 predicted model. Residues F18, P19, and W45, as well as measurements are labeled. The disordered region of subunit A is shown in white, the rest of subunit A is shown in green, subunit B is shown in magenta, and subunit C is shown in cyan. The structure was visualized with PyMOL (*79*). (**B**) *pLDDT* values assigned by AlphaFold2 to each of the predicted residues. Background colors indicate confidence levels: high confidence (*pLDDT* > 90, green), overall good backbone prediction (70 > *pLDDT* > 90, yellow), and low confidence (*pLDDT* < 70, red). Positions to the left of the dashed vertical line are not present in the crystal structure for DfrB1 (PDB: 2RK1). (**C**) Distributions of mutational effects for residues between positions 16 - 26 show that mutations at positions 18 and 19 are deleterious at low promoter activity levels but can be masked at the optimal promoter activity for the WT.

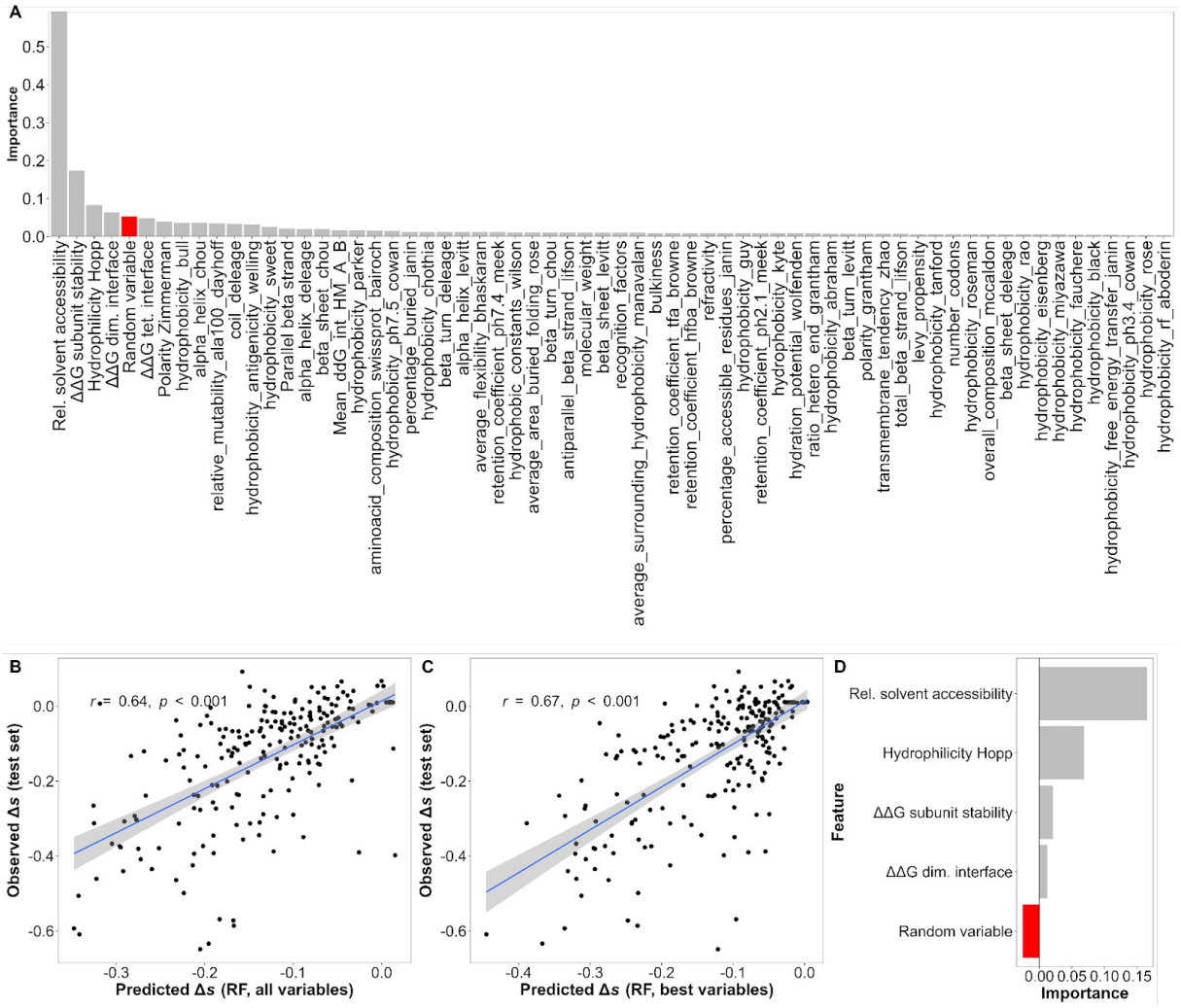


Fig. S16. Identifying the most relevant features used to predict promoter activity-dependent differences in selection coefficients. (**A**) Relative importance of all the features considered in the random forest model as calculated by the decrease in out-of-bag R^2 when permuting the values of that feature. A random variable was included to identify the features that contribute the most to the model. Table S9 contains a list of all variables used in this analysis. (**B**) Comparison of observed Δs_{weak} (s_{weak} - s_{opt}) versus predicted Δs_{weak} using a random forest (RF) regressor with all the features for the test set (20 % of the total dataset). (**C**) Comparison of observed Δs_{weak} (s_{weak} - s_{opt}) versus predicted Δs_{weak} using an RF regressor with all the top three features for the test set (20 % of the total dataset). (**D**) Relative importance of the top features in the final model as calculated by the decrease in out-of-bag R^2 when dropping one variable at a time and retraining the model.

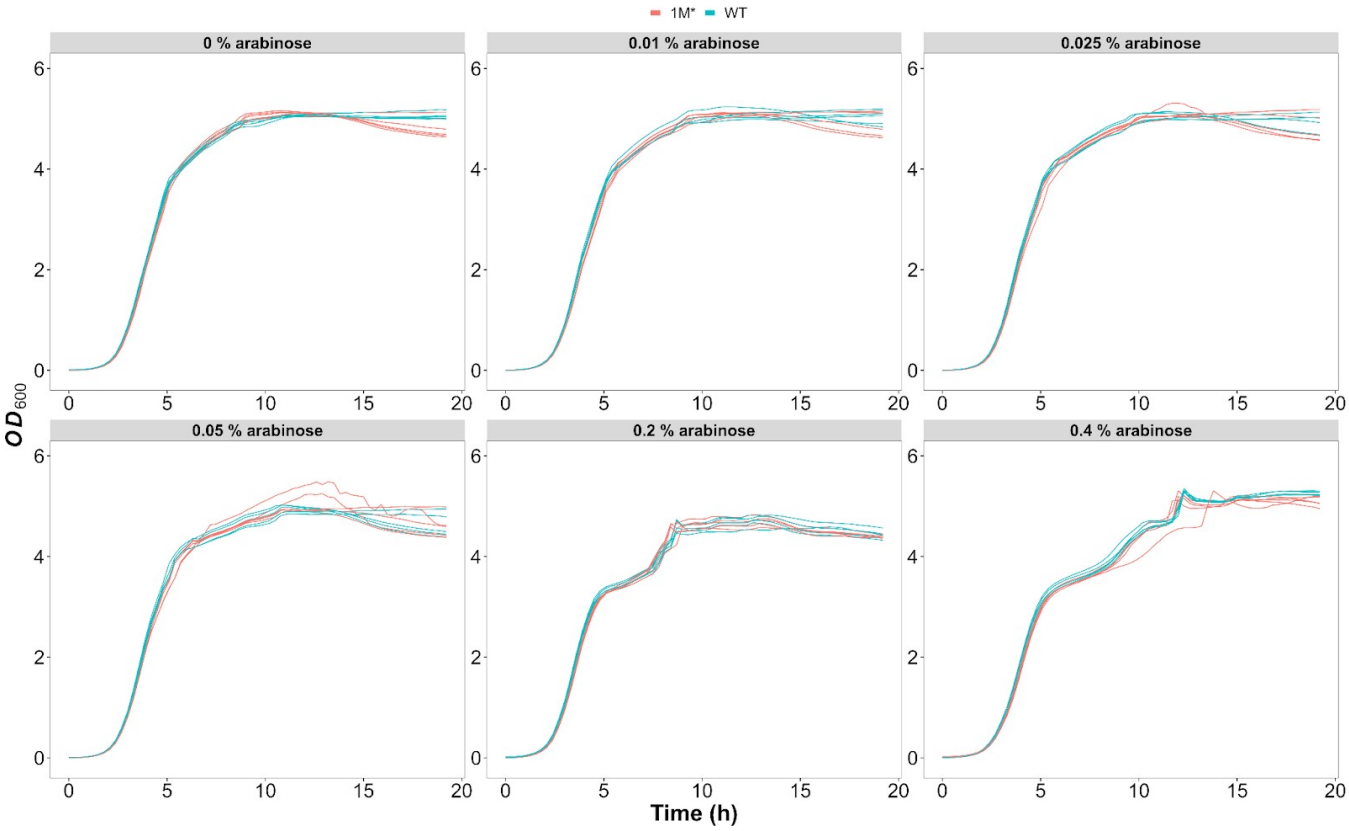
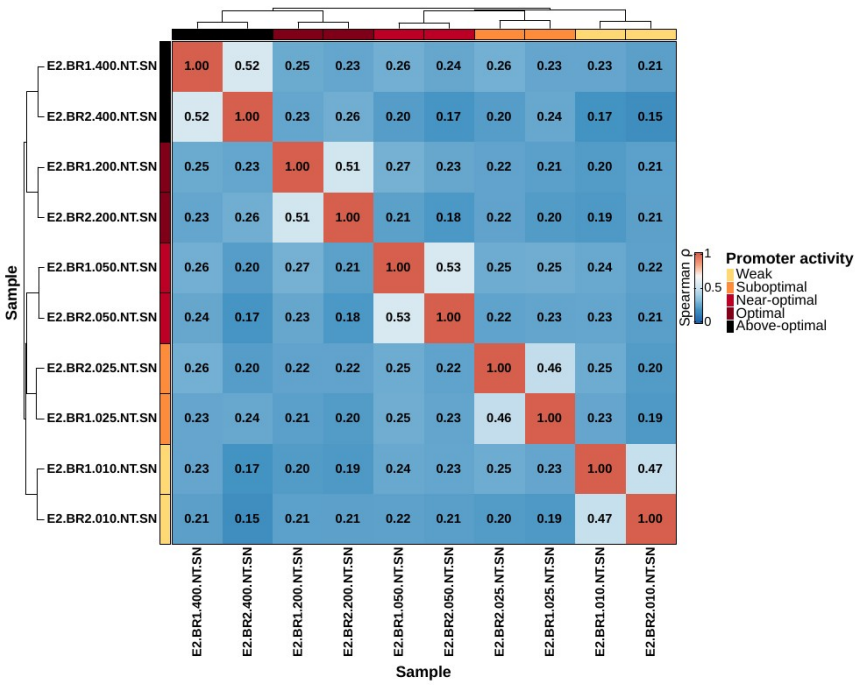


Fig. S17. No clear toxic effects of DfrB1 expression. Growth curves for cells transformed with WT DfrB1 or with a DfrB1 mutant with a stop codon at position 1 (AUG to TGA mutant - M1*) and grown in medium without TMP. Removal of the start codon should not prevent transcription but rather protein expression. A difference between the two alleles would therefore represent the cost of expressing the protein.



Experiment without TMP	ID	Biological replicates	Arabinose concentration (%)	Sequencing platform
2	E2.BR1.010.NT.SN	1	0.01	NovaSeq
	E2.BR2.010.NT.SN	2		
	E2.BR1.025.NT.SN	1	0.025	NovaSeq
	E2.BR2.025.NT.SN	2		
	E2.BR1.050.NT.SN	1	0.05	NovaSeq
	E2.BR2.050.NT.SN	2		
	E2.BR1.200.NT.SN	1	0.2	NovaSeq
	E2.BR2.200.NT.SN	2		
	E2.BR1.400.NT.SN	1	0.4	NovaSeq
	E2.BR2.400.NT.SN	2		

Fig. S18. Low signal-to-noise ratio in selection coefficients in the absence of selection for DfrB1 activity (no TMP). Spearman correlation between selection coefficients was estimated for different replicates of the experiment without TMP. Samples were named according to the table (bottom) extracted from Table S1.

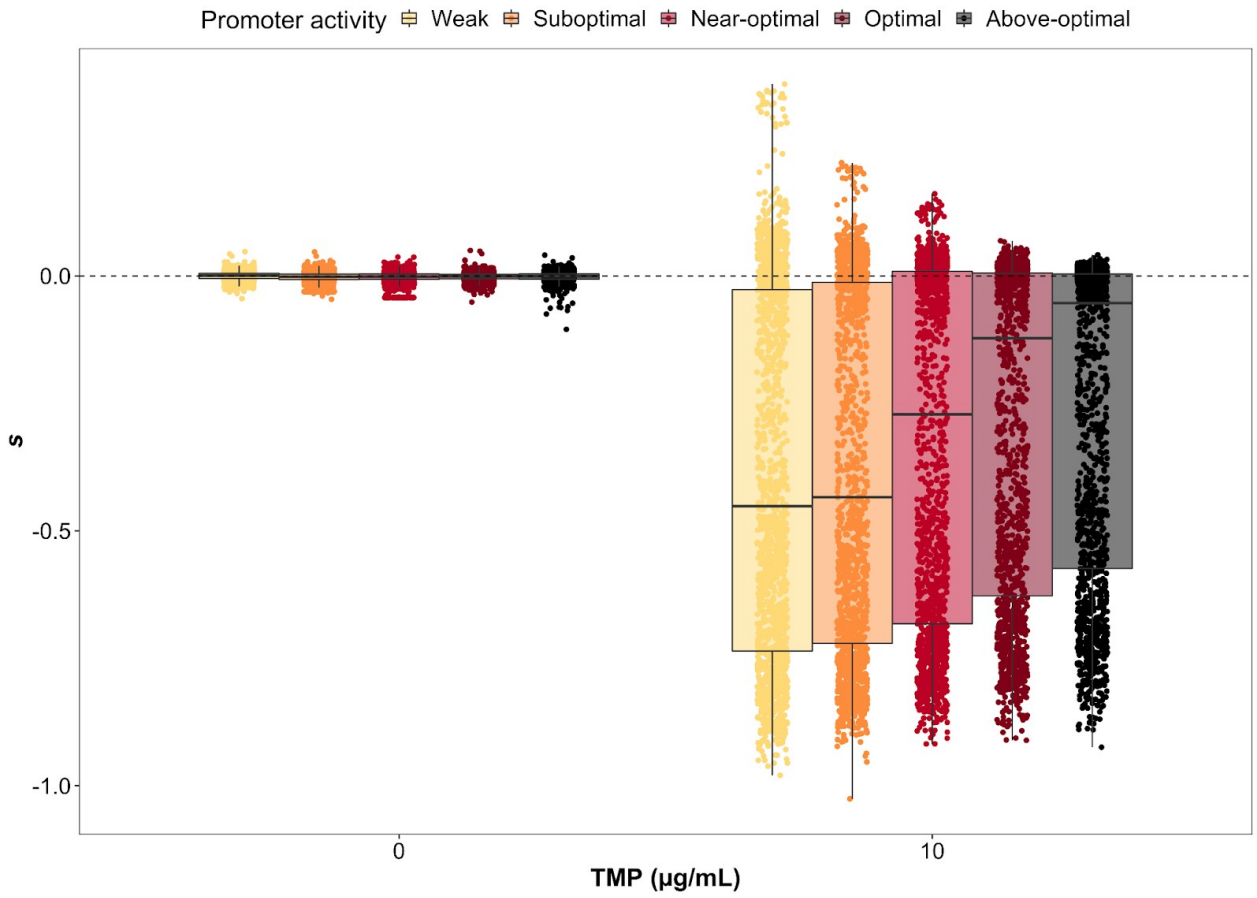


Fig. S19. Overall distributions of selection coefficients in the experiments with and without selection for DfrB1 activity (with and without TMP). The range of selection coefficients observed in the experiment with selection for DfrB1 activity is much broader than in the experiment without selection.

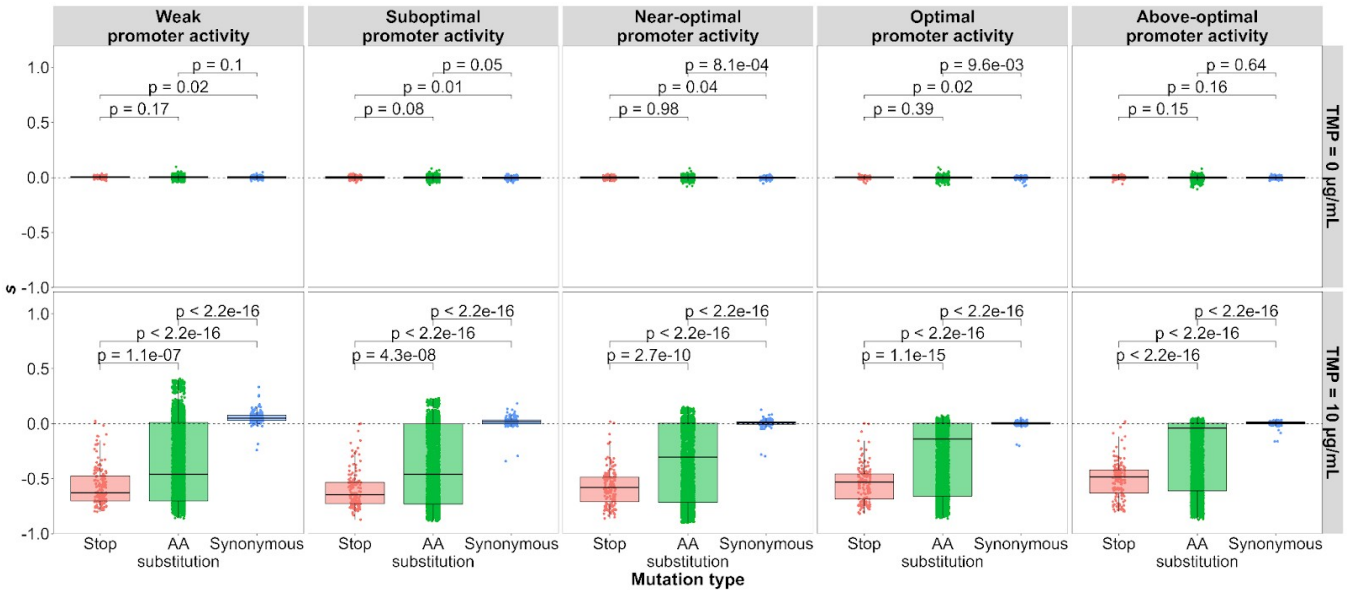


Fig. S20. Distributions of selection coefficients for stop codons, amino acid substitutions, and synonymous mutants. Mutations introducing amino acid substitutions or stop codons are deleterious in the experiment with selection for DfrB1 activity (with TMP), but not in the experiment without selection for DfrB1 activity (without TMP). P-values were calculated using Wilcoxon tests.

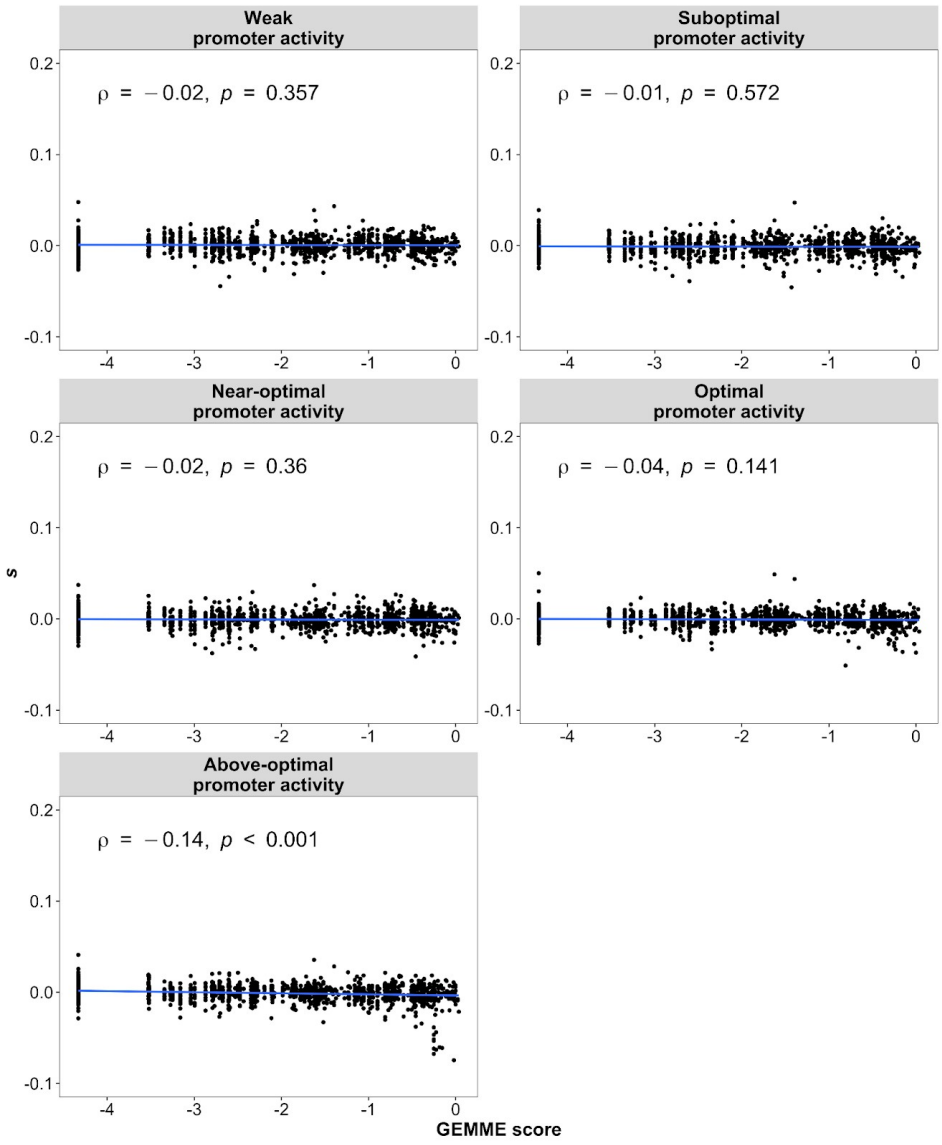


Fig. S21. Fitness effects measured in the DMS bulk competition experiment without TMP do not correlate well with fitness effects deduced variation observed in natural sequences. The GEMME scores indicating patterns of conservation and substitution (0 for residues observed as often as the WT DfrB1 residue at that position and negative for residues observed less often) are the same as in Fig. S11.

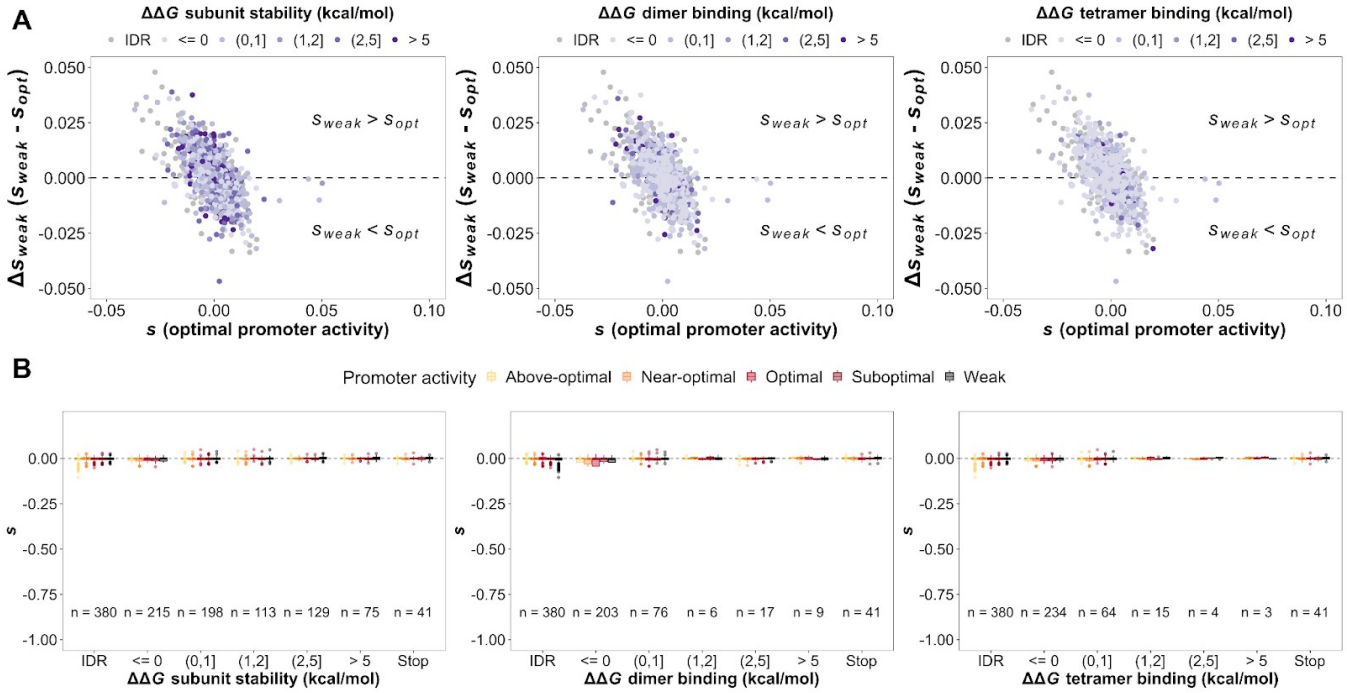


Fig. S22 Protein destabilization by itself does not result in deleterious effects in the absence of selection for DfrB1 activity. (**A**) Landscape of selection coefficients measured at optimal promoter activity (x-axis) and the promoter activity-dependent change in fitness effects (y-axis) labeled with respect to mutational effects on subunit stability and binding affinity. (**B**) Distributions of measured selection coefficients for mutants with different effects on protein stability or binding affinity at the interfaces. Stop codons from the functional protein core (residues 30 - 70) are shown for reference.

Supplementary Table Legends

Table S1. DMS experiment summary. List of samples used in this study and the conditions of each experiment.

Table S2. DMS sample description. Detailed description of conditions, read counts, and oligonucleotides used for each sample.

Table S3. Selection coefficients calculated for each mutation (averaged over biological replicates), estimated mutational effects, and changes in amino acid properties. Full dataset used for analysis with experimental conditions, selection coefficients, structural data, predicted mutational effects, and changes in amino acid properties.

Table S4. ANOVAs for individual mutants. Results of individual ANOVAs for each mutant with all replicates at all promoter activity levels. Benjamini-Hochberg correction was used to correct for multiple hypothesis testing.

Table S5. Centroids for k-means clusters. Coordinates and overall description of centroids identified from the k-means clustering.

Table S6. ANOVA on ranks. General ANOVA on ranks with data from all mutations in all replicates and all promoter activity levels.

Table S7. P-values for Tukey HSD test for fluorescence observed for different constructs (Fig. 4C). ANOVA for comparisons between samples shown in Figure 4C.

Table S8. P-values for Tukey HSD test for selection coefficients in each protein site (Fig. 5C). ANOVA for comparisons between samples shown in Figure 5C.

Table S9. Features used in the random forest regressor. List of features used to model promoter activity-dependent differences in selection coefficient. Sources for each feature are included.

Table S10. ANOVA of selection coefficients with respect to expression level and destabilizing effects. ANOVA on promoter activity levels and bins of mutational effects.

Table S11. Key resources table.

Table S12. Oligonucleotides. List of oligonucleotides used in this study.

Supplementary Data Legends

Data S1. MAFFT alignment of DfrB1 homologs.

REFERENCES AND NOTES

1. J. A. G. M. de Visser, J. Krug, Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* **15**, 480–490 (2014).

2. I. Fragata, A. Blanckaert, M. A. Dias Louro, D. A. Liberles, C. Bank, Evolution in the light of fitness landscape theory. *Trends Ecol. Evol.* **34**, 69–82 (2019).

3. B. Lehner, Molecular mechanisms of epistasis within and between genes. *Trends Genet.* **27**, 323–331 (2011).

4. M. C. King, A. C. Wilson, Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).

5. P. J. Wittkopp, G. Kalay, Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* **13**, 59–69 (2011).

6. D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).

7. K. M. Brown, M. A. Depristo, D. M. Weinreich, D. L. Hartl, Temporal constraints on the incorporation of regulatory mutants in evolutionary pathways. *Mol. Biol. Evol.* **26**, 2455–2462 (2009).

8. X. Li, J. Lalić, P. Baeza-Centurion, R. Dhar, B. Lehner, Changes in gene expression predictably shift and switch genetic interactions. *Nat. Commun.* **10**, 3886 (2019).

9. L. Jiang, P. Mishra, R. T. Hietpas, K. B. Zeldovich, D. N. A. Bolon, Latent effects of Hsp90 mutants revealed at reduced expression levels. *PLOS Genet.* **9**, e1003600 (2013).

10. S. E. Castel, A. Cervera, P. Mohammadi, F. Aguet, F. Reverter, A. Wolman, R. Guigo, I. Iossifov, A. Vasileva, T. Lappalainen, Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nat. Genet.* **50**, 1327–1334 (2018).

11. K. H. Pattishall, J. Acar, J. J. Burchall, F. W. Goldstein, R. J. Harvey, Two distinct types of trimethoprim-resistant dihydrofolate reductase specified by R-plasmids of different compatibility groups. *J. Biol. Chem.* **252**, 2319–2323 (1977).

12. E. E. Howell, Searching sequence space: Two different approaches to dihydrofolate reductase catalysis. *Chembiochem* **6**, 590–600 (2005).

13. M. Faltyn, B. Alcock, A. McArthur, Evolution and nomenclature of the trimethoprim resistant dihydrofolate (dfr) reductases (2019); www.preprints.org/manuscript/201905.0137.

14. N. Narayana, D. A. Matthews, E. E. Howell, N.-H. Xuong, A plasmid-encoded dihydrofolate reductase from trimethoprim-resistant bacteria has a novel D2-symmetric active site. *Nat. Struct. Mol. Biol.* **2**, 1018–1025 (1995).

15. J. M. Krahn, M. R. Jackson, E. F. DeRose, E. E. Howell, R. E. London, Crystal structure of a type II dihydrofolate reductase catalytic ternary complex. *Biochemistry* **46**, 14878–14888 (2007).

16. C. Lemay-St-Denis, S.-S. Diwan, J. N. Pelletier, The bacterial genomic context of highly trimethoprim-resistant DfrB dihydrofolate reductases highlights an emerging threat to public health. *Antibiotics (Basel).* **10**, 433 (2021).

17. H. Kacser, J. A. Burns, The molecular basis of dominance. *Genetics* **97**, 639–666 (1981).

18. S. Stefl, H. Nishi, M. Petukh, A. R. Panchenko, E. Alexov, Molecular mechanisms of disease-causing missense mutations. *J. Mol. Biol.* **425**, 3919–3936 (2013).

19. A. R. Schmitzer, F. Lépine, J. N. Pelletier, Combinatorial exploration of the catalytic site of a drug-resistant dihydrofolate reductase: Creating alternative functional configurations. *Protein Eng. Des. Sel.* **17**, 809–819 (2004).

20. E. Laine, Y. Karami, A. Carbone, GEMME: A simple and fast global epistatic model predicting mutational effects. *Mol. Biol. Evol.* **36**, 2604–2619 (2019).

21. J. Feng, J. Grubbs, A. Dave, S. Goswami, C. G. Horner, E. E. Howell, Radical redesign of a tandem array of four R67 dihydrofolate reductase genes yields a functional, folded protein possessing 45 substitutions. *Biochemistry* **49**, 7384–7392 (2010).

22. L. J. Reece, R. Nichols, R. C. Ogden, E. E. Howell, Construction of a synthetic gene for an R-plasmid-encoded dihydrofolate reductase and studies on the role of the N-terminus in the protein. *Biochemistry* **30**, 10895–10904 (1991).

23. M. Verma, J. Choi, K. A. Cottrell, Z. Lavagnino, E. N. Thomas, S. Pavlovic-Djuranovic, P. Szczesny, D. W. Piston, H. S. Zaher, J. D. Puglisi, S. Djuranovic, A short translational ramp determines the efficiency of protein synthesis. *Nat. Commun.* **10**, 5774 (2019).

24. M. B. Strader, R. D. Smiley, L. G. Stinnett, N. C. VerBerkmoes, E. E. Howell, Role of S65, Q67, I68, and Y69 residues in homotetrameric R67 dihydrofolate reductase. *Biochemistry* **40**, 11344–11352 (2001).

25. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakfool, 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, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

26. M. Mirdita, K. Schutze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

27. G. J. Fuente-Gomez, C. L. Kellum, A. C. Miranda, M. R. Duff, E. E. Howell, Differentiation of the binding of two ligands to a tetrameric protein with a single symmetric active site by 19F NMR. *Protein Sci.* **30**, 477–484 (2021).

28. L. Breiman, Random Forests. *Mach. Learn.* **45**, 5–32 (2001).

29. E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server, in *The Proteomics Protocols Handbook*, J. M. Walker, Ed. (Humana Press, 2005), pp. 571–607.

30. E. D. Levy, S. De, S. A. Teichmann, Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 20461–20466 (2012).

31. Z. Wu, X. Cai, X. Zhang, Y. Liu, G.-B. Tian, J.-R. Yang, X. Chen, Expression level is a major modifier of the fitness landscape of a protein coding gene. *Nat. Ecol. Evol.* **6**, 103–115 (2022).

32. P. C. Després, A. F. Cisneros, E. M. M. Alexander, R. Sonigara, C. Gagné-Thivierge, A. K. Dubé, C. R. Landry, Asymmetrical dose responses shape the evolutionary trade-off between antifungal resistance and nutrient use. *Nat. Ecol. Evol.* **6**, 1501–1515 (2022).

33. F. Duveau, D. C. Yuan, B. P. H. Metzger, A. Hodgins-Davis, P. J. Wittkopp, Effects of mutation and selection on plasticity of a promoter activity in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E11218–E11227 (2017).

34. R. Guerois, J. E. Nielsen, L. Serrano, Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. *J. Mol. Biol.* **320**, 369–387 (2002).

35. D. R. Usmanova, N. S. Bogatyreva, J. A. Bernad, A. A. Eremina, A. A. Gorshkova, G. M. Kanevskiy, L. R. Lonishin, A. V. Meister, A. G. Yakupova, F. A. Kondrashov, D. N. Ivankov, Self-consistency test reveals systematic bias in programs for prediction change of stability upon mutation. *Bioinformatics* **34**, 3653–3658 (2018).

36. J. Hausser, A. Mayo, L. Keren, U. Alon, Central dogma rates and the trade-off between precision and economy in gene expression. *Nat. Commun.* **10**, 68 (2019).

37. S. Karve, P. Dasmeh, J. Zheng, A. Wagner, Low protein expression enhances phenotypic evolvability by intensifying selection on folding stability. *Nat. Ecol. Evol.* **6**, 1155–1164 (2022).

38. D. A. Drummond, J. D. Bloom, C. Adami, C. O. Wilke, F. H. Arnold, Why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14338–14343 (2005).

39. C. Bédard, A. F. Cisneros, D. Jordan, C. R. Landry, Correlation between protein abundance and sequence conservation: What do recent experiments say? *Curr. Opin. Genet. Dev.* **77**, 101984 (2022).

40. J.-F. Gout, D. Kahn, L. Duret; Paramecium Post-Genomics Consortium, The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLOS Genet.* **6**, e1000944 (2010).

41. R. Green, E. J. Rogers, Transformation of chemically competent *E. coli*. *Methods Enzymol.* **529**, 329–336 (2013).

42. Cold Spring Harbor Protocols, *2× YT Medium* (Cold Spring Harbor Laboratory Press, 2014).

43. Cold Spring Harbor Protocols, *LB (Luria-Bertani) Liquid Medium* (Cold Spring Harbor Laboratory Press, 2006).

44. L. M. Guzman, D. Belin, M. J. Carson, J. Beckwith, Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).

45. E. C. Hagan, H. L. T. Mobley, Haem acquisition is facilitated by a novel receptor Hma and required by uropathogenic Escherichia coli for kidney infection. *Mol. Microbiol.* **71**, 79–91 (2009).

46. D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison III, H. O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

47. S. J. Miyake-Stoner, C. A. Refakis, J. T. Hammill, H. Lusic, J. L. Hazen, A. Deiters, R. A. Mehl, Generating permissive site-specific unnatural aminoacyl-tRNA synthetases. *Biochemistry* **49**, 1667–1677 (2010).

48. N. Yachie, E. Petsalaki, J. C. Mellor, J. Weile, Y. Jacob, M. Verby, S. B. Ozturk, S. Li, A. G. Cote, R. Mosca, J. J. Knapp, M. Ko, A. Yu, M. Gebbia, N. Sahni, S. Yi, T. Tyagi, D.

Sheykhkarimli, J. F. Roth, C. Wong, L. Musa, J. Snider, Y.-C. Liu, H. Yu, P. Braun, I. Stagljar, T. Hao, M. A. Calderwood, L. Pelletier, P. Aloy, D. E. Hill, M. Vidal, F. P. Roth, Pooled-matrix protein interaction screens using Barcode Fusion Genetics. *Mol. Syst. Biol.* **12**, 863 (2016).

49. U. Dionne, É. Bourgault, A. K. Dubé, D. Bradley, F. J. M. Chartier, R. Dandage, S. Dibyachintan, P. C. Després, G. D. Gish, N. T. H. Pham, M. Létourneau, J.-P. Lambert, N. Doucet, N. Bisson, C. R. Landry, Protein context shapes the specificity of SH3 domain-mediated interactions in vivo. *Nat. Commun.* **12**, 1597 (2021).

50. K. Sprouffske, A. Wagner, Growthcurver: An R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinform.* **17**, 172 (2016).

51. R. L. Blakley, Crystalline Dihydropteroylglutamic acid. *Nature* **188**, 231–232 (1960).

52. S. Andrews, FastQC A Quality Control tool for High Throughput Sequence Data (2010); www.bioinformatics.babraham.ac.uk/projects/fastqc/.

53. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

54. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

55. A. P. Masella, A. K. Bartram, J. M. Truszkowski, D. G. Brown, J. D. Neufeld, PANDAseq: Paired-end assembler for illumina sequences. *BMC Bioinform.* **13**, 31 (2012).

56. T. Rognes, T. Flouri, B. Nichols, C. Quince, F. Mahé, VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).

57. E. B. Kramer, P. J. Farabaugh, The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA* **13**, 87–96 (2007).

58. G. Korkmaz, M. Holm, T. Wiens, S. Sanyal, Comprehensive analysis of stop codon usage in bacteria and its correlation with release factor abundance. *J. Biol. Chem.* **289**, 30334–30342 (2014).

59. T. Benaglia, D. Chauveau, D. R. Hunter, D. S. Young, mixtools: An R package for analyzing mixture models. *J. Stat. Softw.* **32**, v032i06 (2009).

60. J. L. Toulouse, T. J. Edens, L. Alejaldre, A. R. Manges, J. N. Pelletier, Integron-associated DfrB4, a previously uncharacterized member of the trimethoprim-resistant dihydrofolate reductase B family, is a clinically identified emergent source of antibiotic resistance. *Antimicrob. Agents Chemother.* **61**, e02665–16 (2017).

61. L. S. Johnson, S. R. Eddy, E. Portugaly, Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinform.* **11**, 431 (2010).

62. K. Katoh, K. Misawa, K.-I. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

63. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

64. M. P. Berger, P. J. Munson, A novel randomized iterative strategy for aligning multiple protein sequences. *Comput. Appl. Biosci.* **7**, 479–484 (1991).

65. O. Gotoh, Optimal alignment between groups of sequences and its application to multiple sequence alignment. *Comput. Appl. Biosci.* **9**, 361–370 (1993).

66. A. Bakan, A. Dutta, W. Mao, Y. Liu, C. Chennubhotla, T. R. Lezon, I. Bahar, Evol and ProDy for bridging protein sequence evolution and structural dynamics. *Bioinformatics* **30**, 2681–2683 (2014).

67. R Core Team, R: A Language and Environment for Statistical Computing (2013); www.R-project.org/.

68. W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

69. S. Miller, J. Janin, A. M. Lesk, C. Chothia, Interior and surface of monomeric proteins. *J. Mol. Biol.* **196**, 641–656 (1987).

70. J. Delgado, L. G. Radusky, D. Cianferoni, L. Serrano, FoldX 5.0: Working with RNA, small molecules and a new graphical interface. *Bioinformatics* **35**, 4168–4169 (2019).

71. M. Tiberti, T. Terkelsen, K. Degn, L. Beltrame, T. C. Cremers, I. da Piedade, M. Di Marco, E. Maiani, E. Papaleo, MutateX: An automated pipeline for in silico saturation mutagenesis of protein structures and structural ensembles. *Brief. Bioinform.* **23**, bbac074 (2022).

72. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

73. G. Van Rossum, F. L. Drake, *Python 3 Reference Manual* (CreateSpace Independent Publishing Platform, 2009).

74. M. Murray, J. Blume, FDRestimation: Estimate, Plot, and Summarize False Discovery Rates (2020); https://CRAN.R-project.org/package=FDRestimation.

75. J. O. Wobbrock, L. Findlater, D. Gergle, J. J. Higgins, The aligned rank transform for nonparametric factorial analyses using only anova procedures, in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11)* (ACM, 2011), pp. 143–146.

76. M. Kay, L. A. Elkin, J. J. Higgins, J. O. Wobbrock, ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs (2021); https://zenodo.org/record/4721941#.Y7UsS3ZBwdU.

77. F. de Mendiburu, M. Yaseen, agricolae: Statistical Procedures for Agricultural Research (R package version 1.4.0, 2020); https://myaseen208.com/agricolae/authors.html.

78. J. Dam, T. Rose, M. E. Goldberg, A. Blondel, Complementation between dimeric mutants as a probe of dimer-dimer interactions in tetrameric dihydrofolate reductase encoded by R67 plasmid of *E. coli*. *J. Mol. Biol.* **302**, 235–250 (2000).

79. L. L. C. Schrödinger, The PyMOL Molecular Graphics System, version 1.8 (2015).