ACS SyntheticBiology

pubs.acs.org/synthbio Research Article

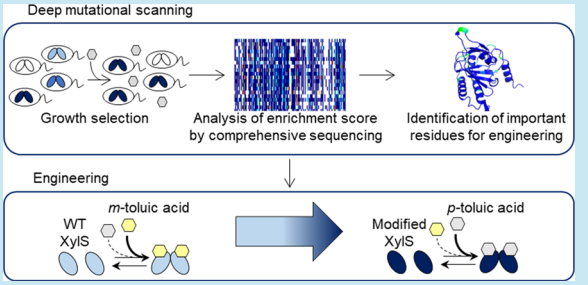
Engineering of the Ligand Specificity of Transcriptional Regulator XylS by Deep Mutational Scanning

Yuki Ogawa, Yohei Katsuyama,* and Yasuo Ohnishi

Cite This: *ACS Synth. Biol.* 2022, 11, 473−485 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Deep mutational scanning is a method for protein engineering. Here, we applied it to alter the ligand specificity of the transcriptional regulator XylS from *Pseudomonas putida* to recognize *p*-toluic acid instead of the native ligand *m*-toluic acid. For this purpose, we used an antibiotic resistance gene-based dual screening system, which was constructed for the directed evolution of XylS toward the above-mentioned ligand specificity. We constructed a *xylS* mutant library in which each codon for the amino acid residue of the putative ligand-binding domain (residues 1−213, except 7th residue) was randomized to generate all possible single amino acid-substituted XylS variants and introduced it into *Escherichia coli* harboring the selection plasmid for the screening system. The cells were cultured in the presence of appropriate antibiotics and *m*-toluic acid or *p*-toluic acid, and the frequency of each mutation present in the library was examined using a next-generation sequencer before and after cultivation. Heatmaps showing the enrichment score of each XylS variant were obtained. By searching for a *p*-toluic-acid-specific heatmap pattern, we focused on G71 and H77. Analysis of the ligand specificities of G71- or H77-substituted XylS variants revealed that several G71-substituted XylS variants responded specifically to *p*-toluic acid. Thus, the 71st residue was found to be an unprecedented residue that is important for switching ligand specificity. Our study demonstrated the usefulness of deep mutational scanning in engineering the ligand specificity of a transcriptional regulator without structural information. We also discussed the advantages and disadvantages of deep mutational scanning compared with directed evolution.



KEYWORDS: *XylS, deep mutational scanning, biosensor, ligand specificity*

INTRODUCTION

Protein engineering plays an increasingly important role in the field of chemical biotechnology.[1] The development of a high-throughput protein engineering method has been in great demand. Accumulation of protein structure and mechanism information combined with computational prediction technologies has driven rational protein engineering approaches,[2] and directed evolution is a powerful strategy for engineering proteins with unknown structures.[3−5] By designing an appropriate selection pressure or screening system for desired phenotypes, directed evolution enables protein engineering without structural information. It can solve some problems that rational engineering faces, such as difficulties in selecting a target site for engineering far from ligand-binding or active sites.[2]

Natural transcriptional regulators, which can detect endogenous or environmental compounds, are useful synthetic biological tools as biosensors.[6] Extensive efforts have been made to tune their functions, such as sensitivity and ligand specificity. In particular, when applied to metabolic engineering to detect target compounds, their specificities toward the compounds are important. Thus, altering ligand specificities is paramount for expanding their applications. We previously applied directed evolution to alter the ligand specificity of the transcriptional regulator XylS,[7] a transcriptional activator in *Pseudomonas putida* belonging to the AraC/XylS family.[8−10] XylS recognizes several aromatic compounds as ligands, especially *m*-toluic acid.[11] XylS forms a complex with its ligand and binds to the operator sequence of the *Pm* promoter (P_{Pm}) as a dimer to induce transcription.[12,13] Although directed evolution was an effective method for engineering the ligand specificity of XylS, we faced some difficulties in the previous study.[7] First, variants harboring beneficial functions (e.g., high ligand specificities) may be missed during the selection if their activities are relatively weak. Second, the number of variants that can be applied to screening is limited due to the difficulty in library construction[2] and the limitation of the amino acid substitutions generated by error-prone PCR,

Received: November 4, 2021
Published: December 29, 2021

https://doi.org/10.1021/acssynbio.1c00564
ACS Synth. Biol. 2022, 11, 473−485

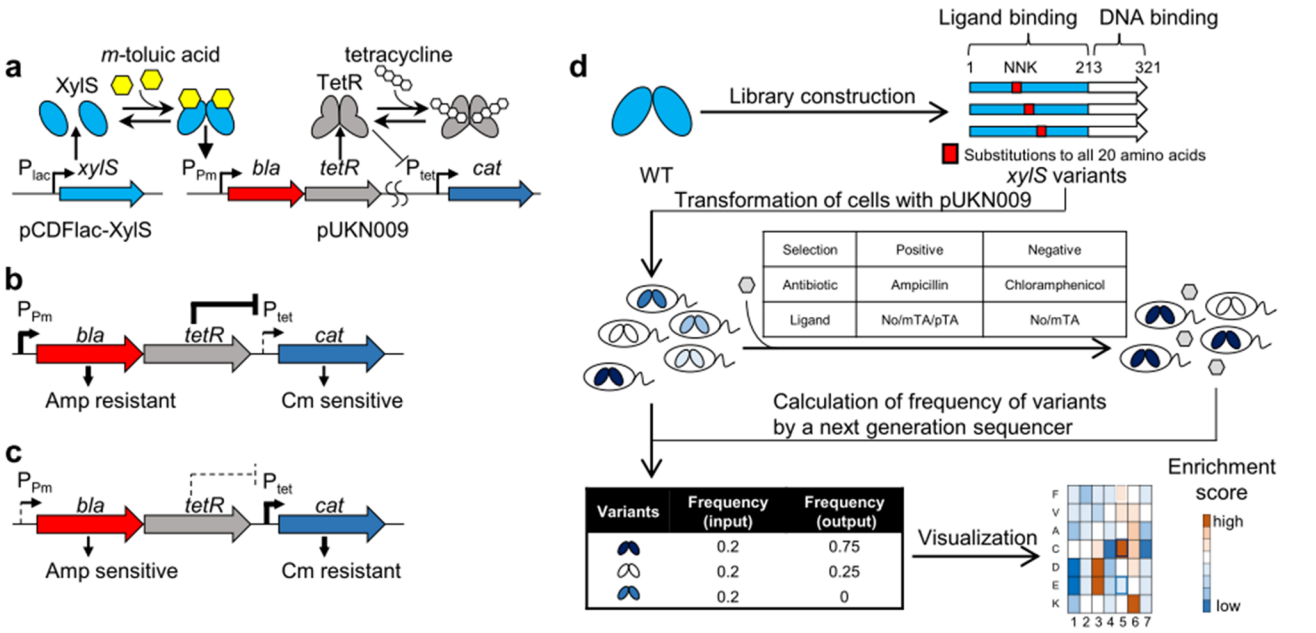


Figure 1. The dual selection system and overview of the methodology for deep mutational scanning of XylS. (a) XylS induces transcription from the *Pm* promoter (P_{Pm}) when it binds to the ligand. TetR represses transcription from the *tetR* promoter (P_{tet}). The repression can be suppressed by the addition of tetracycline. (b) When P_{Pm} is activated by XylS, the strain should become ampicillin (Amp)-resistant and chloramphenicol (Cm)-sensitive. (c) When P_{Pm} is not activated by XylS, the strain should become Amp-sensitive and Cm-resistant. (d) A *xylS* library harboring randomized codons at every residue of the putative ligand-binding domain (aa 1−213) was constructed and introduced into cells with pUKN009. In the positive selection, the cells were cultivated in the presence of Amp plus either no ligand (no), *m*-toluic acid (mTA), or *p*-toluic acid (pTA). In the negative selection, the cells were cultivated in the presence of Cm plus either no ligand (no) or *m*-toluic acid (mTA). The frequencies of variants before and after selection were analyzed by next generation sequencing. The enrichment scores of variants were estimated from their frequencies and visualized with heatmaps.

which usually results in a single nucleotide substitution in a codon.

Deep mutational scanning is a promising method to solve these problems because it enables estimating the effects of all possible single substitutions in a protein or domain.[2] This strategy is based on high-throughput screening and next generation sequencing.[14] The whole process of deep mutational scanning comprises the following steps: (i) construction of a protein variant library used for input, in which each residue has been substituted to every other amino acid to generate all possible single variants; (ii) selection of variants using an appropriate pressure, which links the enrichment of each variant to its function; (iii) comprehensive quantification of the frequency of each mutation present in the library before and after selection by a next generation sequencer, enabling us to estimate the effect of substitutions of certain residues on its function; and (iv) examination of the sequence-function map obtained by interpreting enrichment scores as indicators of functions.[15]

In the case of directed evolution, only a small portion of constructed variants that have passed the selective pressure are provided for further analysis (e.g., functions and sequences). In contrast, deep mutational scanning enables us to examine all substitutions even if their effects are negative or neutral by observing the effects of all the possible single substitutions in a protein (or a part of protein) in an unbiased way.[2,14,16] It can provide information on variants undetectable using other strategies because of their weak activities.[2] Furthermore, because the whole information obtained by deep mutational scanning reflects the sequence-function relationship of a whole protein in a quantitative way to some extent, interpretation of the results might lead to the discovery of important residues or motifs for desired functions. It might further be applied to the rational design of proteins with required functions.

Since Whitehead *et al.* provided the first example of the application of deep mutational scanning to protein engineering,[17,18] it has been extended to various types of protein engineering, including switching specificities of protein−protein interactions,[19] enhancing affinities of protein−protein interactions,[20,21] and enhancing certain features of enzymes.[17,22,23] Fields *et al.* applied this method to alter the ligand specificity of the transcriptional regulator LacI.[24] They adopted both a structural information-based approach and a high-throughput screening-based approach to identify influential residues not only around the ligand-binding pocket but also apart from it. They succeeded in altering the ligand specificity of LacI from its conventional ligand isopropyl *β*-D-1-thiogalactopyranoside to each of four other ligands (fucose, gentiobiose, lactitol, and sucralose) by combining these methods.[24]

In this study, we applied deep mutational scanning to engineer the ligand specificity of XylS, whose structure has not been reported. We aimed to alter the ligand specificity of XylS so that it recognizes *p*-toluic acid instead of its native ligand *m*-toluic acid. We previously constructed an antibiotic resistance gene-based dual screening system for the directed evolution of XylS and successfully obtained the XylS-N7R-T74P variant that recognizes *p*-toluic acid instead of *m*-toluic acid.[7] In the current deep mutational scanning, we used the same dual-selection system for generating XylS variants with the same ligand specificity as XylS-N7R-T74P to compare the utility of deep mutational scanning to that of directed evolution. As a result, we succeeded in identifying new residues important for *p*-toluic acid sensitivity and obtained several XylS variants with

https://doi.org/10.1021/acssynbio.1c00564
ACS Synth. Biol. 2022, 11, 473−485

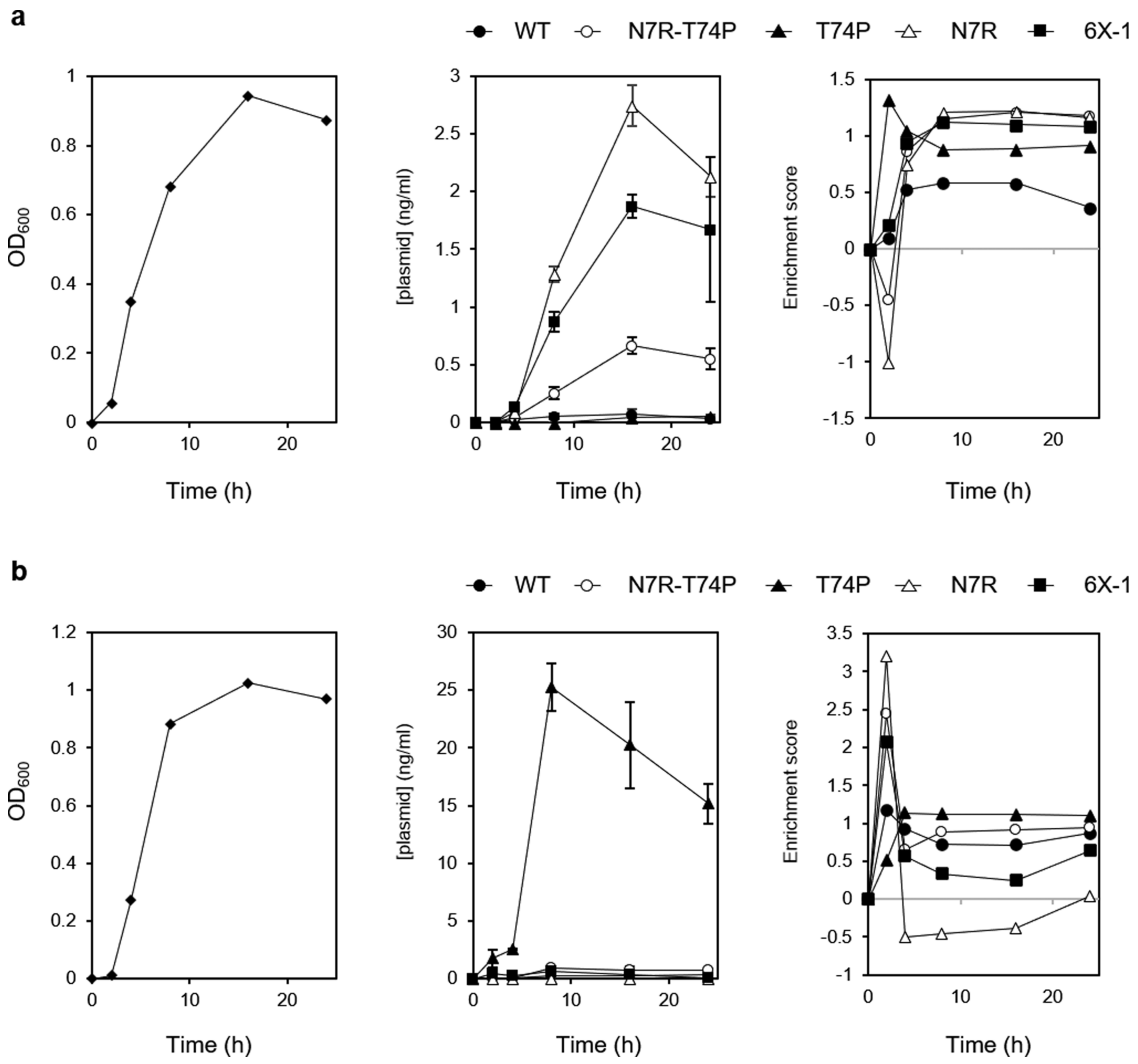


Figure 2. Model experiment for deep mutational scanning using XylS variants with different degrees of sensitivity to *p*-toluic acid and *m*-toluic acid. (a) Growth of the mixed cells (left), the quantity of plasmids encoding WT-XylS, XylS-N7R-T74P, XylS-T74P, XylS-N7R, and XylS-6X-1 (center), and the enrichment scores of each XylS variant (right) in the positive selection (in the presence of ampicillin and *p*-toluic acid). (b) Growth of mixed cells (left), the quantity of plasmids encoding WT-XylS, XylS-N7R-T74P, XylS-T74P, XylS-N7R, and XylS-6X-1 (center), and the enrichment scores of each XylS variant (right) in the negative selection (in the presence of chloramphenicol and *m*-toluic acid). The data (quantity of plasmids) are means, and the error bars indicate standard errors obtained with three technical replicates. Each data point represents the result obtained from a single biological replicate. We carried out the same experiment several times and obtained results showing a similar tendency.

altered ligand specificity toward *p*-toluic acid. Thus, we showed that deep mutational scanning is a powerful strategy to engineer the ligand specificity of a transcriptional regulator with no structural information.

To the best of our knowledge, this study is the first to alter the ligand specificity of a structure-unknown transcriptional regulator by deep mutational scanning.

Furthermore, comprehensive analysis of the sequence-function relationship of XylS over the whole putative ligand-binding domain (213 aa) provided landscapes of the relationship between amino acid substitution and ligand specificity. Our study demonstrated the feasibility of engineering the function of a transcriptional regulator without structural information using deep mutational scanning.

RESULTS

Model Experiment Using a Dual Selection System. In a previous study, we designed a dual selection system, which enabled us not only to enrich XylS variants with desired ligand specificities but also to remove those with non-desired ligand specificities.[7] This system used antibiotic resistance genes as selection markers; we placed an artificial operon composed of an ampicillin resistance gene (*bla*) and the *tetR* gene under the control of the Pm promoter (P_{Pm}) and a chloramphenicol resistance gene (*cat*) under the control of the *tet* promoter (P_{tet}) (Figure 1a) on the same plasmid, pUKN009. The *xylS* gene was placed under the control of the *lac* promoter (P_{lac}) on another plasmid, pCDFlac-XylS. *Escherichia coli* harboring these two plasmids was ampicillin resistant in the presence of a desired ligand and chloramphenicol resistant in the presence of a non-desired ligand (Figure 1b,c). Using this system for directed evolution, we succeeded in altering the ligand specificity of XylS, as described in the Introduction section.[7]

We applied this system to the deep mutational scanning of XylS (Figure 1d). To correlate the abundance of each variant after cultivation under a selection pressure with its ligand sensitivity, we defined "enrichment score" as the value that reflects the ratio of the relative growth rate of the strain

https://doi.org/10.1021/acssynbio.1c00564
ACS Synth. Biol. 2022, 11, 473−485

Table 1. Selective Conditions, Sampling Conditions, and Statistics of Read Counts from Each Selection Culture

	before selection	after selection				
		negative selection		positive selection		
antibiotic		chloramphenicol	chloramphenicol	ampicillin	ampicillin	ampicillin
ligand		no ligand	*m*-toluic acid	no ligand	*m*-toluic acid	*p*-toluic acid
[antibiotic] (mg/L)		34	34	250	500	500
sampling time (h)	0	16	16	20	16	16
cell density (Abs)	0.0002	1.15	1.139	0.231	1.014	0.583
raw read count	1,237,844	1,231,705	998,272	914,656	1,124,153	1,372,421
filtered read count	712,739	695,489	632,498	605,113	753,296	428,680
sequence depth	85	86	79	70	91	52
variant type count[a]	1950	1678	1715	266	1270	357

[a]Variant type count indicates the number of different types of XylS variants that could be produced in the sampled cells.

producing a certain variant to that of all cells in the culture. The relative growth rate was estimated by comparing the number of cells estimated from the concentration of plasmids quantified by quantitative PCR (qPCR) before and after selection (see Materials and Methods for details). Because the host of an XylS variant that responds to a ligand would be enriched during cultivation in the presence of the ligand and ampicillin (positive selection), the enrichment score in this positive selection should reflect its ability to recognize the ligand. In addition, because the host of an XylS variant that does not respond to a ligand would be enriched during cultivation in the presence of the ligand and chloramphenicol (negative selection), the enrichment score in this negative selection should reflect its ability to not respond to the ligand. Thus, we would be able to find XylS variants with desired ligand specificities by using the enrichment scores in both the positive and negative selections as indicators of ligand specificity.

We first examined the feasibility of the selection system for deep mutational scanning using a model experiment. The purposes of this experiment were (i) to test whether the enrichment scores of XylS variants would reflect their ligand specificities and (ii) to optimize cultivation and sampling conditions.

For the model experiment, we used WT-XylS and four XylS variants, which we have already analyzed in the previous study and responded differently to *m*-toluic acid and *p*-toluic acid:[7] XylS-N7R-T74P, XylS-T74P, XylS-N7R, and XylS-6X-1 (N7S-T74A-K83E-I205V-N227D) (Figure S1). While WT-XylS specifically responded to *m*-toluic acid, XylS-N7R-T74P responded to *p*-toluic acid. XylS-T74P also responded specifically to *p*-toluic acid but with a much lower efficiency, which resulted in the complete loss of the response to *m*-toluic acid. XylS-N7R and XylS-6X-1 responded to both ligands. We expected that the cells harboring the selection plasmids pUKN009 and XylS variants with high *p*-toluic acid sensitivity (XylS-N7R-T74P, XylS-N7R, and XylS-6X-1) would be enriched in the presence of *p*-toluic acid and ampicillin, whereas the cells harboring pUKN009 and XylS variants with no *m*-toluic acid sensitivity (XylS-T74P) would be enriched in the presence of *m*-toluic acid and chloramphenicol. Thus, the combination of enrichment scores of the host cells in the positive and negative selections would reflect the *p*-toluic acid and *m*-toluic acid specificities of the XylS variants.

We measured the enrichment scores of the four XylS variants and WT-XylS using q-PCR. First, plasmids encoding WT-XylS and respective XylS variants were labeled with DNA barcodes.[25] Next, labeled plasmids were used to transform *E. coli* harboring pUKN009 and the transformants were mixed at the same ratio. The cells were cultivated in the presence of *p*-toluic acid and ampicillin or in the presence of *m*-toluic acid and chloramphenicol. Plasmids were extracted from the culture at several time points, and their quantities were measured by q-PCR, in which primers binding to respective DNA barcodes were used. The ratio of the concentration of each plasmid (for each XylS variant) before and after the selection was calculated (Figure 2). In addition, colonies were formed from the cell mixture of the 24 h culture, and eight *xylS* clones were used for DNA sequencing. As expected, XylS-6X-1, XylS-N7R, and XylS-N7R-T74P were enriched in the positive selection (Figure 2a, Table S1), and XylS-T74P was predominantly enriched in the negative selection (Figure 2b, Table S1). They showed high enrichment scores in the corresponding selection. Thus, we expected to identify XylS variants with desired ligand specificities by using the enrichment scores as an indicator.

Deep Mutational Scanning of XylS. Since it was reported that the N-terminal region of XylS is important for ligand binding and dimerization,[26−30] we constructed an *xylS* mutant library in which each codon for the amino acid residue in the putative ligand-binding domain (residues 1−213) was randomized to alter the amino acid residue to all the possible 20 amino acid residues (Figure 1d). This library presumably directs the production of 4047 (19 × 213) single amino acid-substituted XylS variants and WT-XylS. This *xylS* library was introduced into *E. coli* harboring the selection plasmid pUKN009. The transformants were cultivated under five different conditions (Figure 1d, Table 1). Two conditions were used for the negative selection: culture containing chloramphenicol with (i) no ligand or (ii) *m*-toluic acid. In addition, three conditions were used for the positive selection: culture containing ampicillin with (iii) no ligand, (iv) *m*-toluic acid, or (v) *p*-toluic acid (Table 1). After 16−20 h, plasmids were extracted from each culture. To check whether the selection system worked properly, eight clones each of *xylS* mutants from (iii) and (v) were used as representatives for DNA sequencing, and the ligand specificity of each XylS variant was analyzed. Although we had expected that different kinds of XylS variants should be enriched in these two conditions, we noticed that XylS variants harboring substitutions at N7 were enriched in both conditions (Table S2). Our previous study indicated that substitution at N7 increased the basal activity of XylS in the absence of ligands. Thus, to avoid the dominant growth of the N7-substituted XylS variant-producing strains, they should be eliminated from the experiment.

Therefore, we constructed a similar *xylS* mutant library in which each codon of the amino acid residue in the putative

https://doi.org/10.1021/acssynbio.1c00564
ACS Synth. Biol. 2022, 11, 473−485

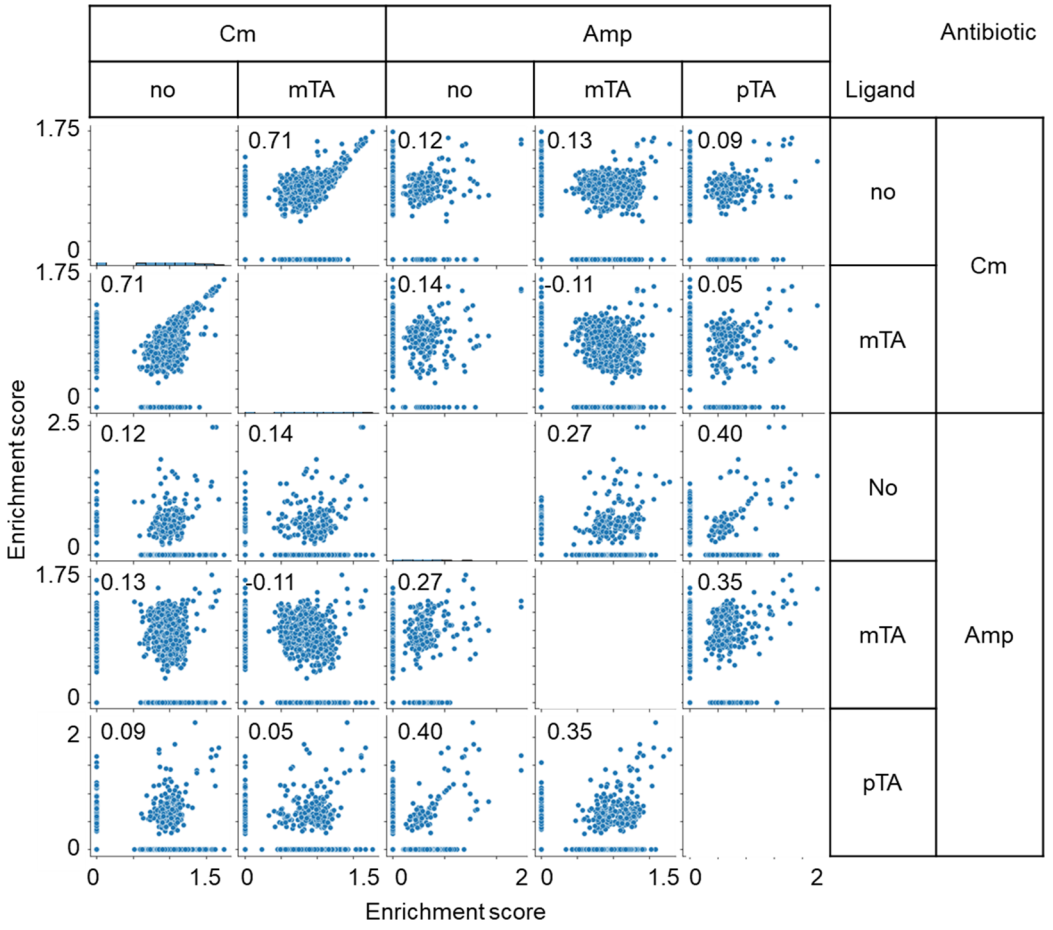


Figure 3. Pairwise scatter plots between different conditions. The enrichment scores of XylS variants in the five selection conditions described in the table on the right were plotted against those in the table above the plots (Cm, chloramphenicol; Amp, ampicillin; mTA, *m*-toluic acid; no, no ligand; pTA, *p*-toluic acid). Each correlation coefficient is displayed in the upper left of each plot.

ligand-binding domain, except that of N7, was randomized and conducted the same selection. As the growth rate of cells from (iii) was very low compared to the other conditions, the concentration of ampicillin was lowered, and cultivation time was extended to 20 h (Table 1). The reference library, which corresponded to the culture before selection, was taken at time = 0. Namely, the same amount of the mixture of transformants harboring the XylS variant library was inoculated in LB medium and plasmids were immediately extracted from the cells in the LB medium at time = 0 without cultivation. The plasmids extracted from the culture before and after selection were used as templates for PCR amplification of the 5′ region of *xylS*, which corresponded to the putative ligand-binding domain. To obtain the landscape of an amino acid substitution-ligand specificity relationship, *xylS* sequences in the input and output libraries (before and after the selection, respectively) were determined by high-throughput sequencing using Nova-seq. In each library, the average number of read pairs was 1.1 million. Among them, 56% of the sequenced read pairs passed quality filtering and were translated to align with the amino acid sequence of XylS (Table 1).

Validation and Interpretation of Data Obtained by Next-Generation Sequencing. To correlate the abundance of each variant with its ligand sensitivity, we also used "enrichment score" here. The relative growth rate was estimated by comparing the number of cells estimated from the ratio between read counts and coverages before and after selection (see Materials and Methods for details). We obtained pairwise scatter plots (Figure 3) and heatmaps (Figure 4) for each library using the enrichment scores indicating the ligand sensitivity of the XylS variants. We then used these pairwise scatter plots and heatmaps to screen the amino acid residues important for switching ligand specificity.

We first used pairwise scatter plots to analyze the correlation of the enrichment scores of XylS variants among the different cultures. Our analysis indicated that enrichment scores of variants in the negative selections hardly correlated with those in the positive selections (Figure 3, correlation coefficients −0.11−0.14), indicating that each selective pressure worked properly. Enrichment scores of variants in the positive selections showed weak positive correlations (Figure 3, correlation coefficients 0.27−0.40), which might reflect that XylS variants with high basal activities were predominantly enriched in these selections. Finally, the enrichment scores of variants in the two negative selections showed a strong positive correlation (Figure 3, correlation coefficient 0.71), from which we expected XylS variants not responding to *m*-toluic acid because of their low basal activities to be predominantly enriched.

We next analyzed heatmaps to capture the rough landscape of the relationship between amino acid substitution and ligand specificity. Two heatmaps of the negative selections with *m*-toluic acid and with no ligand showed similar landscapes (Figure 4a,b, Figure S2a,b). In contrast, the heatmap of the *m*-

https://doi.org/10.1021/acssynbio.1c00564
ACS Synth. Biol. 2022, 11, 473−485

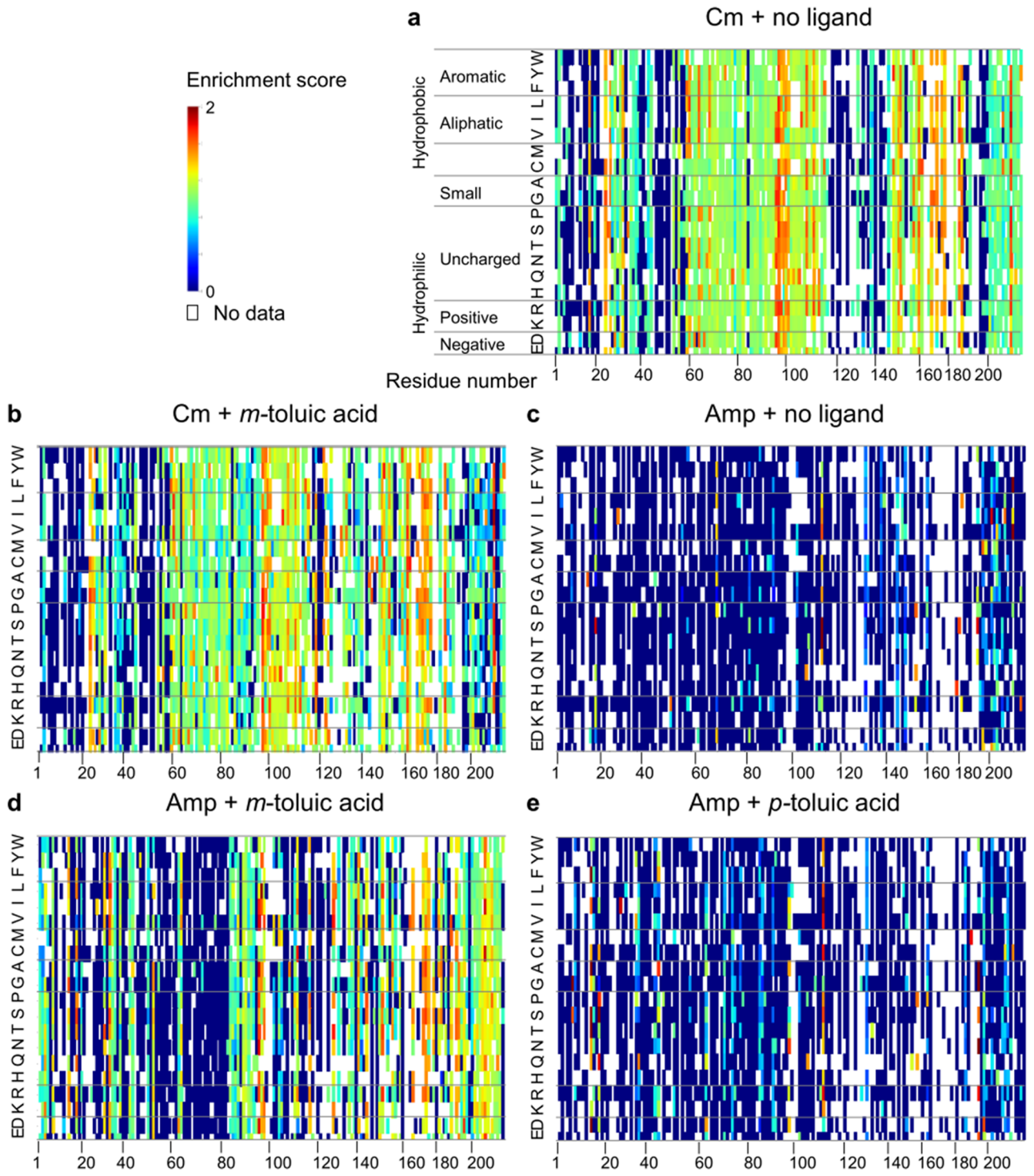


Figure 4. Landscape of amino acid substitution-ligand specificity relationship of XylS. The colors in the heatmaps indicate the enrichment score (from 0 to 2; the scale is shown on the upper left corner), which is defined as the relative growth rate of a strain producing a certain XylS variant divided by the relative growth rate of the whole cells. Enrichment scores of XylS variants that were not observed either before or after selection are not shown on the heatmaps or are shown in white as "no data". The chemical properties of amino acids are shown on the left of panel (a). The position of substitution is shown on the horizontal axis, and the amino acids used for substitution are shown on the vertical axis. Cm, chloramphenicol; Amp, ampicillin.

toluic acid-added positive selection showed an almost inverted landscape compared to the heatmaps of the negative selections (Figure 4a,b,d, Figure S2a,b,d). Landscapes in the positive selection with no ligand or with *p*-toluic acid indicated that few variants were predominantly enriched (Figure 4c,e, Figure S2c,e). This observation indicates that XylS hardly acquires a *p*-toluic-acid-recognizing ability by a single amino acid substitution. Looking into the heatmaps, we noticed that replacement of A111 with any kind of amino acid resulted in high basal expression in the absence of ligand (Figure 4c,e, Figure S2c,e).

https://doi.org/10.1021/acssynbio.1c00564
ACS Synth. Biol. 2022, 11, 473−485

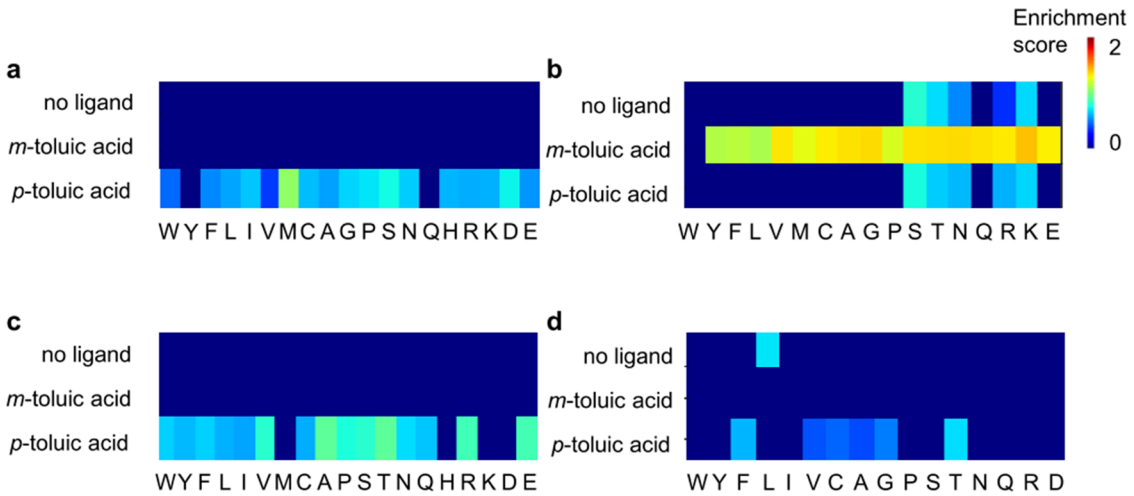


Figure 5. Single residue enrichment score heatmaps at residues T74 (a), I205 (b), G71 (c), and H77 (d). The colors in the heatmaps correspond to those in the scale shown on the upper right corner. The amino acids used for substitution are shown on the horizontal axis, and the ligands added in the positive selection are shown on the vertical axis.

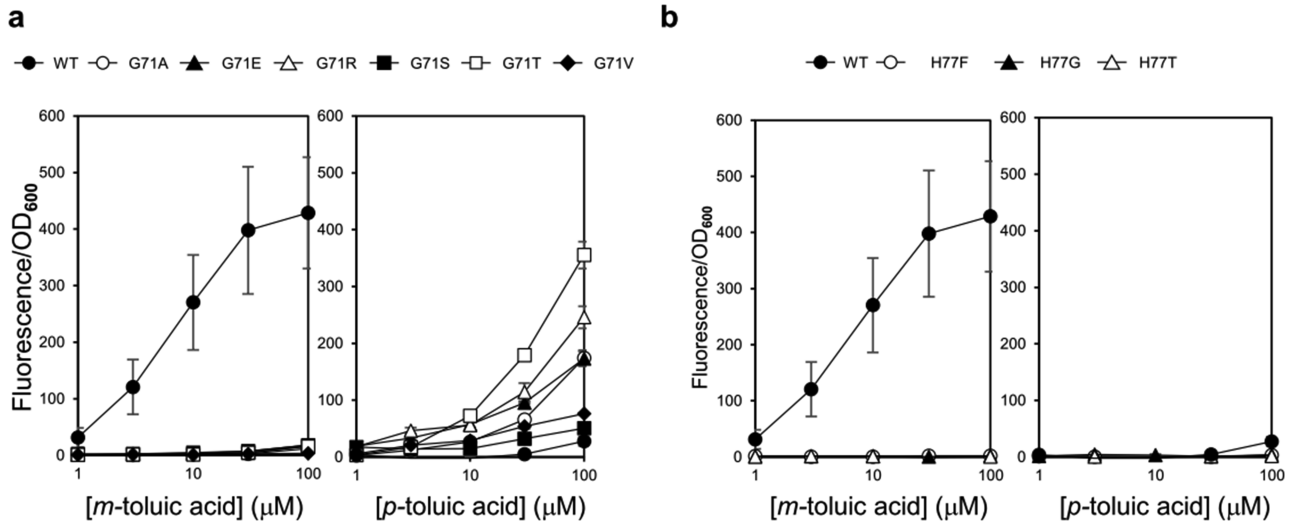


Figure 6. Sensitivity of WT-XylS and XylS variants to *m*-toluic acid and *p*-toluic acid. *E. coli* strains harboring both pCOLAPmmCherry and a pCDFlac-XylS-series plasmid were cultivated for 10 h in the presence of various concentrations of *m*-toluic acid or *p*-toluic acid. The fluorescence intensity of each culture was measured and normalized by the OD_{600} value of the culture. pCDFlac-XylS-WT, G71A, G71E, G71R, G71S, G71T, and G71V were used in (a) and pCDFlac-XylS-WT, H77F, H77G, and H77T were used in (b).

A111-substituted XylS variants have also been reported by another group to acquire relaxed ligand selectivity.[31] As it was reported that the truncated C-terminal domain of XylS induces transcription[27] and that the N-terminal domain of XylS represses DNA binding in the absence of a ligand,[32] substitution at A111 might repress the function of the N-terminal domain.

To estimate the relationship between the function and structure of XylS, we used a homology-based 3D model of XylS built by Phyre2[33] based on the structure of the N-terminal domain of CuxR[34] (PDB ID: 5NLA, 11% identity) (Figure S3). As mentioned in the previous study,[7] this model consists of a β-barrel domain connected via a loop to a hair-pin domain composed of two helices, α-4 and α-5. We visualized the averaged enrichment score of each residue in the presence of ampicillin and *m*-toluic acid or *p*-toluic acid by coloring the model (Figure S3a,b). The heatmap from the culture with ampicillin and *m*-toluic acid showed that residues 200−213 had high enrichment scores. The 3D model of XylS showed that residues 200−213 were located on a loop connecting α-5 and α-6. Notably, A111 is located at the end of another loop, which is close to the loop connecting α-5 and α-6, although A111 is far from residues 200−213 on the primary sequence. α-4 and α-5 are part of the dimerization interface in CuxR;[34] therefore, we assumed that residues 200−213 are also important for the dimerization of XylS.[8,10,35] This suggests that the mutations introduced in the region might affect the ability to induce XylS by altering the dimerization efficiency. Fields *et al.* reported in their study of deep mutational scanning that many LacI variants responsive to a new ligand had substitutions at the dimerization interface of LacI.[24] Therefore, introducing substitutions into the dimerization interface might be effective for engineering transcriptional regulators that form a dimer to enhance the induction ability. Most of the residues with a high enrichment score obtained from the culture with ampicillin and *m*-toluic acid or *p*-toluic acid were located on the loops (Figure S3a,b). We assumed that loops would be tolerant to amino acid substitutions because they are

https://doi.org/10.1021/acssynbio.1c00564
ACS Synth. Biol. 2022, 11, 473−485

intrinsically flexible and that amino acid substitutions at the ligand-binding pocket might result in loss of function.

Predicting *p*-Toluic Acid-Specific XylS Variants Using Single Residue Enrichment Score Heatmaps. Next, we obtained heatmaps for every single amino acid residue to visualize the influence of single amino acid substitution on ligand specificity and attempted to identify variants whose ligand specificity was switched from *m*-toluic acid to *p*-toluic acid. In these "single residue enrichment score heatmaps", substitute amino acids were lined up horizontally and ligands used for the positive selection (in the presence of ampicillin) were lined up vertically (Figure 5, Figure S4). First, we focused on the heatmaps of the 74th and 205th residues (originally T74 and I205) because they were indicated to be important residues for *p*-toluic acid recognition in our previous study[7] (Figure 5a,b); in particular, the 74th residue was found to be important for *p*-toluic acid specificity. The heatmap of the 74th residue clearly shows that T74-substituted XylS variants were enriched only in the presence of *p*-toluic acid. Therefore, we considered this as a *p*-toluic acid-specific enrichment pattern and searched the single residue enrichment score heatmaps of other residues for a similar pattern in order to find other *p*-toluic-acid-specific XylS variants. As a result, we found that the heatmaps of the 71st and 77th residues (originally G71 and H77) have similar patterns (Figure 5c,d, Figure S4) and focused on these two residues for further experiments.

Validation of Ligand Specificities of XylS Variants. We validated the ligand specificities of five G71-substituted, five T74-substituted, and three H77-substituted XylS variants by the previously developed system using an *E. coli* strain harboring a plasmid carrying the mCherry gene under the control of P_{Pm}.[7] As a result, all the G71-substituted XylS variants hardly recognized *m*-toluic acid and recognized *p*-toluic acid twice to 12 times stronger than WT-XylS (Figure 6a). In particular, the G71T variant showed remarkable *p*-toluic acid specificity; it recognized *p*-toluic acid almost as strongly as WT-XylS recognized *m*-toluic acid (Figure 6a). Contrary to our expectation, the H77-substituted XylS variants did not recognize *m*-toluic acid or *p*-toluic acid (Figure 6b). This may be attributed to the low enrichment scores of H77-substituted XylS variants in the *p*-toluic acid-supplemented condition; they were much lower than those of G71-substituted XylS variants and were assumed to be "false-positive" results. The T74-substituted XylS variants showed weaker *m*-toluic acid recognition than WT-XylS, while they showed stronger *p*-toluic acid recognition to varying extents (Figure S5). Among the five T74-substituted XylS variants analyzed, XylS-T74M and XylS-T74K recognized *p*-toluic acid more strongly than *m*-toluic acid, while XylS-T74I and XylS-T74S recognized *m*-toluic acid more strongly than *p*-toluic acid. XylS-T74D recognized both ligands to the same extent. In most cases, the enrichment scores of G71- and T74-substituted XylS variants in the presence of *p*-toluic acid reflected the relative level of *p*-toluic acid sensitivity. Thus, there was a positive correlation between the enrichment score and ligand specificity of XylS variants (Figure S6).

Combining the Effects of Amino Acid Substitutions Important for *p*-Toluic Acid-Specific Recognition. To obtain a XylS variant further specific to *p*-toluic acid, we decided to combine amino acid substitutions important for *p*-toluic acid-specific recognition. We constructed a G71-site saturation library based on the XylS-N7R-T74P variant. This library was introduced into *E. coli* harboring pUKN009 and selected by the strategy described previously.[7] First, the cells were cultivated in the presence of chloramphenicol and *m*-toluic acid for negative selection. Next, the survived cells were cultivated in the presence of ampicillin and *p*-toluic acid for positive selection. Then, the eight *xylS* clones of the survived cells were sequenced by Sanger sequencing. As a result, six of them were XylS-N7R-G71M-T74P, and the others were XylS-N7R-G71L-T74P and XylS-N7R-G71P-T74P (Table S3). We also analyzed the XylS-N7R-G71T-T74P variant because the G71T variant showed the best specificity toward *p*-toluic acid (Figure 6). Analysis of ligand specificities of these XylS variants showed that all of them hardly responded to *m*-toluic acid (Figure S7). XylS-N7R-G71M-T74P, XylS-N7R-G71L-T74P, and XylS-N7R-G71P-T74P recognized *p*-toluic acid stronger than WT-XylS but their responses were weaker than XylS-G71T (Figure S7). Furthermore, XylS-N7R-G71T-T74P lost the *p*-toluic acid-recognizing ability (Figure S7). These results indicated that combining N7R-T74P substitutions with a G71 substitution does not improve the *p*-toluic acid specificity. Since G71 and T74 were predicted to locate on the same loop, substituting both amino acids would result in a structural change, which would weaken or lose the ability to recognize *p*-toluic acid.

DISCUSSION

In this study, we applied deep mutational scanning to alter the ligand specificity of XylS, the structure of which has not been elucidated. Using comprehensive information derived from deep mutational scanning, we obtained several G71-substituted XylS variants that showed a *p*-toluic acid-specific ligand response. Our results also indicated that residues 71st and 74th are the only ones critical to gain *p*-toluic acid specificity in XylS by a single amino acid substitution. It cannot be achieved by conventional directed evolution or rational engineering to prove that there are no other candidate residues useful for switching ligand specificity from *m*- to *p*-toluic acid other than 71st and 74th. Furthermore, we could assume that residues 60−80 and residues 100−120 are essential for maintaining the ligand recognizing ability from Figure 4d. Thus, deep mutational scanning can provide useful information for engineering the function of XylS not only by suggesting important residues for acquiring desired function but also by identifying residues with no or negative effect for engineering.

The XylS variants that acquired good selectivity to *p*-toluic acid were G71T, G71R, G71V, and G71A variants. The codon for G71 in WT-XylS was GGT in our synthetic gene, and it can be substituted with Asp (G<u>A</u>T), Ala (G<u>C</u>T), Val (G<u>T</u>T), Cys (<u>U</u>GT), Arg (<u>C</u>GT), and Ser (<u>A</u>GT) by a single nucleotide substitution. Therefore, the best variant (G71T) was not accessible by error-prone PCR. In addition, G71S and G71V variants, which can be established by single nucleotide substitution, responded to *p*-toluic acid only weakly. Taken together with the limitation of library size, we expected that these are the reasons why these substitutions were overlooked in our previous study using directed evolution, which tends to enrich only variants with strong activity. Furthermore, this study also proved that directed evolution using the dual-selection system is an effective way to identify important mutations for ligand specificity to some extent, as T74, another important residue, was discovered by directed evolution in our previous study.[7] Although a relatively small size of the XylS library was screened for each selection round, the repetitive dual-selection strategy succeeded in obtaining *p*-toluic-acid-

https://doi.org/10.1021/acssynbio.1c00564
ACS Synth. Biol. 2022, 11, 473−485

specific XylS variants containing T74 substitution. In addition, T74 substitution did not demonstrate high specificity and sensitivity until it was combined with the N7 substitution. It would be difficult to discover this combination by deep mutational scanning because it only focuses on a single amino acid substitution. Thus, our study clarified the advantages and disadvantages of deep mutational scanning and directed evolution by comparing the results obtained by both methods using the same selection system.

Interestingly, all the G71-substituted variants lost the ability to recognize *m*-toluic acid while obtaining the ability to recognize *p*-toluic acid. We assumed the role of the 71st residue from the homology-based 3D model of XylS built by Phyre2[33] based on the structure of CuxR[34] (Figure S3b). The β-barrel forms a possible ligand-binding pocket, and G71, as well as T74, is expected to be located on a loop of the β-barrel. Glycine is a small amino acid, and thus, changing it to a larger amino acid residue might bring steric hindrance causing the structural change of its surrounding region and resulting in a dynamic change in the shape of the whole ligand-binding pocket. As reported in several previous studies by other groups,[36−38] the flexible nature of loops might be related to acquiring new functions, and flexible loops should be promising targets for mutagenesis in protein engineering.[39]

CONCLUSIONS

In this study, deep mutational scanning was applied to engineer the ligand specificity of XylS so that it recognizes its non-native ligand *p*-toluic acid instead of its native ligand *m*-toluic acid. By combining the dual-selection system constructed in our previous study with comprehensive sequencing using a next generation sequencer, the landscapes of the amino acid substitution-ligand specificity relationship were obtained. Based on "single residue enrichment score heatmaps," we discovered an unprecedented important residue, G71, for switching the ligand specificity of XylS in addition to T74, which was previously found in our directed evolution analysis.[7] Thus, we succeeded in obtaining a series of *p*-toluic-acid-specific XylS variants. Because any kind of compound could be added as a possible ligand during selection, the strategy reported in this study is applicable to the engineering of the ligand specificities of other transcriptional regulators to generate a series of biosensors with variable ligand specificities. Considering the development of high-throughput sequencing technologies in recent years, deep mutational scanning is a promising approach for high-throughput engineering of biosensors, regardless of structural information.

MATERIALS AND METHODS

Strains, Media, Chemicals, and Other Materials. *E. coli* strain JM109 (Takara Biochemicals, Shiga, Japan) was used as the host for DNA manipulation. Luria-Bertani (LB) broth was purchased from Difco (Franklin Lakes, New Jersey). All chemicals were purchased from FUJIFILM Wako Pure Chemical Corporation (Osaka, Japan), Sigma-Aldrich (St. Louis, Missouri), and Tokyo Chemical Industry Co., Ltd. (Tokyo, Japan). Enzymes used for DNA manipulation were purchased from Takara Bio Inc. (Shiga, Japan). Primers were purchased from Thermo Fisher Scientific (Waltham, Massachusetts), and their sequences are shown in Tables S4 and S5. The DNA sequences were analyzed by Fasmac Co. Ltd. (Kanagawa, Japan).

Construction of Plasmids Encoding XylS Variants with DNA Barcodes. The first DNA barcodes were inserted into XylS-encoding plasmids (e.g., pCDFlac-XylS, -N7R-T74P, -T74P, -N7R, and -6X-1) by inverse PCR using a pair of primers (each of primers 1−5 containing respective barcodes/primer 6, Table S4) to obtain the barcode-inserted XylS-encoding plasmids (barcode-pCDFlac-XylS, -N7R-T74P, -T74P, -N7R, and -6X-1). The template plasmids were then digested with *Dpn*I. The PCR products were phosphorylated with T4 DNA polynucleotide kinase (Takara Bio Inc.) and cyclized with T4 DNA ligase (Takara Bio Inc.). Unintentional mutations introduced during PCR were corrected by additional inverse PCR for barcode-pCDFlac-XylS-N7R, -T74P, and -6X-1 using primers 5 and 7−11 in the same procedure as above. The second DNA-barcode inserted XylS-encoding plasmid series (e.g., pCDFlac-barcoded-XylS, -N7R-T74P, -T74P, -N7R, and -6X-1) was constructed by inverse PCR using a pair of primers (each of primers 13−17 containing respective barcodes/primer 12, all of which with a *Hin*dIII restriction site, Table S4). The amplified DNA fragments were digested with *Hin*dIII and cyclized with T4 DNA ligase.

Model Experiments for the Selection System Used for Deep Mutational Scanning. Five *E. coli* strains harboring pUKN009 and either pCDFlac-barcoded-XylS, -N7R-T74P, -T74P, -N7R, or -6X-1 were pre-cultivated in LB medium at 37 °C overnight. In the model experiment for positive selection, overnight cultures were mixed at the same ratio and subsequently inoculated (0.2% v/v) into 100 mL of fresh LB medium containing 50 μM IPTG, 50 mg/L kanamycin, 50 mg/L streptomycin, 350 mg/L ampicillin, and 100 μM *p*-toluic acid. In the model experiment for negative selection, these cultures were mixed at the same ratio and subsequently inoculated (0.2% v/v) into 100 mL of fresh LB medium containing 50 μM IPTG, 50 mg/L kanamycin, 50 mg/L streptomycin, 34 mg/L chloramphenicol, 100 μM *m*-toluic acid, and 3 μg/L tetracycline. The cells were incubated for 24 h at 37 °C. To analyze the competitive growth of these *E. coli* strains, the growth was measured, and the plasmids were extracted from the culture and analyzed by qPCR using pairs of primers specific to respective barcode sequences (primers 18−27, Table S4). Based on the concentration of plasmids in the culture, the frequency (f_v) of each variant from each library was calculated as f_v = [the concentration of a plasmid encoding a XylS variant at the sampling time]/[the sum of the concentration of plasmids encoding all XylS variants]. The enrichment score of the XylS variant ($E_{variant}$) was calculated as the specific growth rate of each variant. The specific growth rate (g_v) of each variant was calculated as $g_v = \log_2(X_t/X_0)/t$, where X_t is a parameter to estimate the number of cells harboring this variant at time t, and X_0 is that at time 0. X_t was calculated as $X_t = OD_t \times f_v$, where OD_t is the OD_{600} value at time t. Based on the specific growth rate of cells harboring each variant, the enrichment score (E) was calculated as $E = g_v/g_w$, where g_v is the specific growth rate of the cells harboring each variant and g_w is that of the whole-cell culture. g_w was estimated from OD_{600} at time 0 and t.

Construction of a Site-Saturation Mutagenesis *xylS* Library. A site-saturation mutagenesis *xylS* sub-library harboring mutations at the codon for a certain amino acid was constructed by inverse PCR using a pair of primers harboring NNK at the codon to be changed (Table S5). The PCR product was usually self-cyclized during PCR, and the template plasmid was digested with *Dpn*I. For the construction

https://doi.org/10.1021/acssynbio.1c00564
ACS Synth. Biol. 2022, 11, 473−485

of the *xylS* sub-library harboring mutations at the codon for D2, F3, L5, L6, Q11, I12, Y19, D23, Y28, T31, I34, P37, K38, P42, H47, D55, C57, R58, Y61, and V65, the PCR product was phosphorylated with T4 DNA polynucleotide kinase, and cyclized with T4 DNA ligase. We constructed a sub-library for every residue from 1 to 213. Since saturation mutations could not be introduced at the codon for G209 in spite of repeated trials of inverse PCR, they were introduced as follows. A linear form of pCDFlac-XylS, in which the codon for G209 was eliminated, was amplified by PCR using a pair of primers 30−31. The PCR product and primer 32, which contains NNK at the position of the codon for G209, were connected using the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs Inc., Ipswich, Massachusetts). The template plasmids were then digested with *Dpn*I.

The cyclized plasmid of each sub-library was used to transform *E. coli* harboring the selection plasmid pUKN009 and inoculated onto LB agar supplemented with appropriate antibiotics (50 mg/L kanamycin [the marker for pUKN009] and 50 mg/L streptomycin [the marker for pCDFlacXylS]). The plates were then incubated overnight at 37 °C. The generated *E. coli* library, composed of more than 120 colonies so that the probability of harboring all the possible 19 amino acid substitutions would be high enough, was suspended in LB medium and an appropriate amount of the suspension of the *E. coli* cells was prepared. The suspensions of all 213 sub-libraries were mixed together to generate an *E. coli* library harboring a site-saturation mutagenesis *xylS* library that would direct the production of 4047 (19 × 213) single amino acid-substituted XylS variants and WT-XylS. For the construction of the library lacking in the production of N7-substituted XylS variants, the *E. coli* cell suspensions of the 212 sub-libraries other than those for N7 were used.

Selection from the Site-Saturation Mutagenesis *xylS* Library. An appropriate amount of the *E. coli* library (the mixture of the *E. coli* cells harboring respective sub-libraries) was inoculated into fresh LB medium (100 mL) to obtain a population density of 4×10^4 cells/mL. The LB medium was supplemented with 50 μM IPTG, 100 μM ligand (*m*-toluic acid or *p*-toluic acid), and appropriate antibiotics (50 mg/L kanamycin, 50 mg/L streptomycin, and 350 mg/L ampicillin) for positive selection. For negative selection, the LB medium (100 mL) was supplemented with 50 μM IPTG, 100 μM *m*-toluic acid, and appropriate antibiotics (50 mg/L kanamycin, 50 mg/L streptomycin, 34 mg/L chloramphenicol, and 3 μg/L tetracycline). The cells were cultivated at 37 °C for 16−20 h, and the cell density of each culture was estimated by measuring the absorbance at 600 nm (5×10^8 cells/mL for OD_{600} = 1.0). To track the mutations introduced into *xylS*, plasmids were extracted from the culture, and *E. coli* JM109 was transformed with the plasmids. Colonies were grown on LB agar containing 50 mg/L streptomycin at 37 °C overnight. Eight colonies were inoculated into 2 mL of LB medium and cultivated at 37 °C overnight. Finally, each plasmid was extracted, and the *xylS* sequence was determined using primers 50 and 51.

Deep Sequencing. Plasmids extracted from *E. coli* cells were amplified by PCR using primers 28 and 29 (Table S4). The PCR products were purified using a gel extraction kit (Nippon Genetics Co., Ltd.). Fragmentation and sequencing were outsourced to Novogene (Beijing, China). They were randomly fragmented to a size of approximately 250 bp using Bioruptor and sequenced with NovaSeq6000/PE250. The average Phred quality score was calculated for each read in a given set of paired-end reads; if either of reads had an average quality score lower than 5, the read pair was discarded. The remaining read pairs were merged into a single sequence using CLC Genome Workbench software (CLC bio, Aarhus, Denmark). DNA sequences and their amino acid translations were aligned to the sequences of *xylS* and XylS, respectively. To examine the reproducibility of the deep sequencing experiment, we selected the culture condition containing *m*-toluic acid and ampicillin, and carried out technical replicates from the amplification. The result confirmed the reproducibility of the experiment (R_2 = 0.96, Figure S8).

Computational Analysis of High-Throughput Sequencing Results. The CLC genomics workbench (CLC bio) and Python scripts were used for the analysis. First, the number of reads harboring each amino acid substitution from each library was counted. Next, the frequency (f_v) of each variant from each library was calculated as $f_v = \mathrm{Read}_{sv}/\mathrm{coverage}$, where Read_{sv} is the number of times that this variant appeared in the library and coverage is the number of read counts at the codon for the corresponding amino acid residue. Since the coverage was not uniform over the whole *xylS* sequence, we calculated the frequency based on the coverage at each codon to avoid the bias arising from the different sequence depths of the different codons. Based on the frequency, the specific growth rate (g_v) of each variant was calculated as $g_v = \log_2(X_t/X_0)/t$, where X_t is a parameter to estimate the number of cells harboring this variant at time t and X_0 is that at time 0. X_t was calculated as $X_t = \mathrm{OD}_t \times f_v$, where OD_t is the OD_{600} value at time t. Based on the specific growth rate of cells harboring each variant, the enrichment score (E) was calculated as $E = g_v/g_w$, where g_v is the specific growth rate of the cells harboring each variant and g_w is that of the whole-cell culture. g_w was estimated from OD_{600} at times 0 and t. When X_0 = 0, X was replaced by the minimum value of X_0, as it was considered to be under detection threshold. When both X_0 and X_t = 0, E was not calculated because we assumed it was a missing value ("no data" in Figure 4). Finally, paired-plots and heatmaps were obtained based on enrichment scores. Sequence coverages, read counts, and enrichment scores are provided in independent csv files as Supporting Information. Variants with asterisk (*) indicate the nonsense mutation and a variant with question mark (?) indicates the mutation in the start codon. "Minus-infinite" in enrichment scores indicates that the variants had no read after selection. These data were shown in the same color (dark blue) as 0 in Figures 4 and 5, and Figures S2 and S4.

Site-Directed Mutagenesis of *xylS*. Single amino acid-substituted XylS-encoding plasmid series (e.g., pCDFlac-XylS-G71T and -H77T) were constructed by inverse PCR using a pair of primers (primers 33−49, 52, Table S4) containing the corresponding mutation. The PCR products were self-cyclized during PCR, and the template plasmid was digested with *Dpn*I.

Screening from G71-Substituted XylS-N7R-T74P Library. A site-saturation mutagenesis *N7R-T74P-xylS* library with mutations at G71 was constructed by inverse PCR using a pair of primers containing NNK at the codon for G71 and pCDFlac-XylS-N7R-T74P as a template (Table S4). The PCR products were self-cyclized during PCR, and the template plasmid was digested with *Dpn*I. The cyclized plasmid of each sub-library was used to transform *E. coli* harboring the selection plasmid pUKN009 and inoculated onto LB agar supplemented with appropriate antibiotics (50 mg/L kanamycin [the marker for pUKN009] and 50 mg/L streptomycin [the marker for

https://doi.org/10.1021/acssynbio.1c00564
ACS Synth. Biol. 2022, 11, 473−485

pCDFlacXylS]). The plates were then incubated overnight at 37 °C. The generated *E. coli* library, composed of more than 120 colonies so that the probability of harboring all the possible 19 amino acid substitutions would be high enough, was suspended in LB medium and an appropriate amount of the suspension of the *E. coli* cells was prepared. An appropriate amount of the mixture of the *E. coli* cells was inoculated into fresh LB medium to give a population density of 3×10^4 cells/mL. The LB medium was supplemented with 50 μM IPTG, different concentrations of *m*-toluic acid (1, 3, 10, 30, or 100 μM), and appropriate antibiotics (50 mg/L kanamycin, 50 mg/L streptomycin, 34 mg/L chloramphenicol, and 3 μg/L tetracycline). The cells were cultivated at 37 °C for 8 h and the OD_{600} of each culture was measured. Then, we selected a culture that showed a distinct growth under the highest concentration of *m*-toluic acid (100 μM) as the inoculant for the positive selection as follows. The selected culture was inoculated into each fresh LB medium to a population density of 3×10^4 cells/mL. The LB medium was supplemented with 50 μM IPTG, different concentrations of *p*-toluic acid (1, 3, 10, 30, or 100 μM), and appropriate antibiotics (50 mg/L kanamycin, 50 mg/L streptomycin, and 350 mg/L ampicillin). The cells were cultivated at 37 °C for 8 h, and the OD_{600} of each culture was measured. Then, we selected a culture that showed a distinct growth under the lowest concentration of *p*-toluic acid (100 μM). To check the mutations concentrated during the selections, plasmids were extracted from the selected culture after positive selection, and *E. coli* JM109 was transformed with the plasmids. Colonies were grown on LB agar containing 50 mg/L streptomycin at 37 °C overnight. Then, eight colonies were inoculated into 2 mL LB medium and cultivated at 37 °C overnight. Finally, each plasmid was extracted and the *xylS* sequence was determined using primers 50 and 51 (Table S4).

Fluorescence Assay. *E. coli* strains for the fluorescence assay were constructed by transforming *E. coli* JM109 with both pCOLAPmmCherry[7] and pCDFlacXylS (or its derivatives encoding XylS variants). The cells were pre-cultivated overnight in LB medium at 37 °C. The cultures were inoculated (1% v/v) into 200 μL of fresh LB medium supplemented with 50 μM IPTG, appropriate antibiotics (50 mg/L kanamycin and 50 mg/L streptomycin), and different concentrations of *m*-toluic acid or *p*-toluic acid (1, 3, 10, 30, or 100 μM) in a microtube. Cells were incubated at 37 °C with vigorous shaking at 500 rpm for 10 h. Cells were harvested by centrifugation for 15 min at 1100 × *g* and resuspended in 200 μL PBS buffer (pH 7.4−7.6). Then, the OD_{600} and fluorescence (excitation wavelength of 587 nm and emission wavelength of 610 nm) values were measured in triplicate using the microplate reader Spectramax M2 (Molecular Devices, San Jose, California). The average ± standard error was plotted against the ligand concentration.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.1c00564.

m-Toluic acid and *p*-toluic acid sensitivity of WT-XylS and XylS variants derived from barcoded plasmids; enlarged landscapes of amino acid substitution-ligand specificity relationship of XylS; visualization of averaged enrichment scores on the model structure of XylS; representatives of single residue enrichment score heatmaps; *m*-toluic acid and *p*-toluic acid sensitivity of WT XylS and T74-substituted XylS variants; correlation between enrichment score and *p*-toluic acid sensitivity; *m*-toluic acid and *p*-toluic acid sensitivity of WT XylS and N7, G71, and T74-substituted XylS variants; reproducibility of the sequencing results for the library from the culture in the presence of ampicillin and *m*-toluic acid; number of XylS variants observed in the model experiment; amino acid substitutions examined by a canonical DNA sequencer in the eight selected XylS variants after the positive selection with no ligand or with *p*-toluic acid; number of XylS variants observed in the screening from the XylS-N7R-G71NNK-T74P library; and primers used in this study except those for the construction of the deep mutational scanning library (PDF)

Sequence coverages for each base position in the deep sequencing experiments, read counts for each variant in the deep sequencing experiments, and enrichment scores for each variant in the deep sequencing experiments (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Yohei Katsuyama − *Department of Biotechnology, The Graduate School of Agricultural and Life Sciences and Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, Tokyo 113-8657, Japan;* orcid.org/0000-0003-4905-5320; Email: aykatsuhko@g.ecc.u-tokyo.ac.jp

Authors

Yuki Ogawa − *Department of Biotechnology, The Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan*

Yasuo Ohnishi − *Department of Biotechnology, The Graduate School of Agricultural and Life Sciences and Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, Tokyo 113-8657, Japan;* orcid.org/0000-0001-7633-9236

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.1c00564

Author Contributions

Y.K. and Y.Og. designed the study. Y.Og. performed the experiments. Y.Og. analyzed the data. Y.Og., Y.K., and Y.Oh. wrote the paper. All authors read and approved the final manuscript. Y.K. and Y.Oh. supervised the study.

Notes

The authors declare no competing financial interest.

The raw sequence data generated during this study are available at the DDBJ Sequence Read Archive database under the accession number PRJDB12536, DRA012999, DRA013000, DRA013001, DRA013002, DRA013003, DRA013004, and DRA013005.

■ ACKNOWLEDGMENTS

This research was supported, in part, by Grants-in-Aid for Challenging Exploratory Research (16K14880 to Y.K.) and JSPS Fellows (20J12216 to Y.Og.) from the Japan Society for the Promotion of Science (JSPS), and JSPS A3 Foresight

Program (to Y.Oh.). We thank Dr. Hikaru Nakazawa and Dr. Mitsuo Umetsu in Tohoku University for useful discussions on this study.

REFERENCES

(1) Chen, Z.; Zeng, A.-P. Protein engineering approaches to chemical biotechnology. *Curr. Opin. Biotechnol.* **2016**, *42*, 198−205.
(2) Shin, H.; Cho, B.-K. Rational protein engineering guided by deep mutational scanning. *Int. J. Mol. Sci.* **2015**, *16*, 23094−23110.
(3) Cobb, R. E.; Si, T.; Zhao, H. Directed evolution: an evolving and enabling synthetic biology tool. *Curr. Opin. Chem. Biol.* **2012**, *16*, 285−291.
(4) Bassalo, M. C.; Liu, R.; Gill, R. T. Directed evolution and synthetic biology applications to microbial systems. *Curr. Opin. Biotechnol.* **2016**, *39*, 126−133.
(5) Dietrich, J. A.; McKee, A. E.; Keasling, J. D. High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu. Rev. Biochem.* **2010**, *79*, 563−590.
(6) De Paepe, B.; Peters, G.; Coussement, P.; Maertens, J.; De Mey, M. Tailor-made transcriptional biosensors for optimizing microbial cell factories. *J. Ind. Microbiol. Biotechnol.* **2017**, *44*, 623−645.
(7) Ogawa, Y.; Katsuyama, Y.; Ueno, K.; Ohnishi, Y. Switching the ligand specificity of the biosensor XylS from *meta* to *para*-toluic acid through directed evolution exploiting a dual selection system. *ACS Synth. Biol.* **2019**, *8*, 2679−2689.
(8) Gawin, A.; Valla, S.; Brautaset, T. The XylS/ *Pm* regulator/ promoter system and its use in fundamental studies of bacterial gene expression, recombinant protein production and metabolic engineering. *Microb. Biotechnol.* **2017**, *10*, 702−718.
(9) Ramos, J.-L. *Pseudomonas. Volume 2*; Springer: New York, 2004.
(10) Gallegos, M. T.; Schleif, R.; Bairoch, A.; Hofmann, K.; Ramos, J. L. AraC/XylS family of transcriptional regulators. *Microbiol. Mol. Biol. Rev.* **1997**, *61*, 393−410.
(11) Xue, H.; Shi, H.; Yu, Z.; He, S.; Liu, S.; Hou, Y.; Pan, X.; Wang, H.; Zheng, P.; Cui, C.; Viets, H.; Liang, J.; Zhang, Y.; Chen, S.; Zhang, H. M.; Ouyang, Q. Design, construction, and characterization of a set of biosensors for aromatic compounds. *ACS Synth. Biol.* **2014**, *3*, 1011−1014.
(12) Kessler, B.; Herrero, M.; Timmis, K. N.; de Lorenzo, V. Genetic evidence that the XylS regulator of the *Pseudomonas* TOL meta operon controls the Pm promoter through weak DNA-protein interactions. *J. Bacteriol.* **1994**, *176*, 3171−3176.
(13) Zwick, F.; Lale, R.; Valla, S. Regulation of the expression level of transcription factor XylS reveals new functional insight into its induction mechanism at the Pm promoter. *BMC Microbiol.* **2013**, *13*, 262.
(14) Badenhorst, C. P. S.; Bornscheuer, U. T. Getting momentum: from biocatalysis to advanced synthetic biology. *Trends Biochem. Sci.* **2018**, *43*, 180−198.
(15) Araya, C. L.; Fowler, D. M. Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol.* **2011**, *29*, 435−442.
(16) Fowler, D. M.; Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **2014**, *11*, 801−807.
(17) Kowalsky, C. A.; Klesmith, J. R.; Stapleton, J. A.; Kelly, V.; Reichkitzer, N.; Whitehead, T. A. High-resolution sequence-function mapping of full-length proteins. *PLoS One* **2015**, *10*, No. e0118193.
(18) Whitehead, T. A.; Chevalier, A.; Song, Y.; Dreyfus, C.; Fleishman, S. J.; De Mattos, C.; Myers, C. A.; Kamisetty, H.; Blair, P.; Wilson, I. A.; Baker, D. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.* **2012**, *30*, 543−548.
(19) Naftaly, S.; Cohen, I.; Shahar, A.; Hockla, A.; Radisky, E. S.; Papo, N. Mapping protein selectivity landscapes using multi-target selective screening and next-generation sequencing of combinatorial libraries. *Nat. Commun.* **2018**, *9*, 3935.
(20) Fujino, Y.; Fujita, R.; Wada, K.; Fujishige, K.; Kanamori, T.; Hunt, L.; Shimizu, Y.; Ueda, T. Robust in vitro affinity maturation strategy based on interface-focused high-throughput mutational scanning. *Biochem. Biophys. Res. Commun.* **2012**, *428*, 395−400.
(21) Ernst, A.; Gfeller, D.; Kan, Z.; Seshagiri, S.; Kim, P. M.; Bader, G. D.; Sidhu, S. S. Coevolution of PDZ domain−ligand interactions analyzed by high-throughput phage display and deep sequencing. *Mol. BioSyst.* **2010**, *6*, 1782−1790.
(22) Starita, L. M.; Pruneda, J. N.; Lo, R. S.; Fowler, D. M.; Kim, H. J.; Hiatt, J. B.; Shendure, J.; Brzovic, P. S.; Fields, S.; Klevit, R. E. Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. *Proc. Natl. Acad. Sci.* **2013**, *110*, E1263−E1272.
(23) Wrenbeck, E. E.; Faber, M. S.; Whitehead, T. A. Deep sequencing methods for protein engineering and design. *Curr. Opin. Struct. Biol.* **2017**, *45*, 36−44.
(24) Taylor, N. D.; Garruss, A. S.; Moretti, R.; Chan, S.; Arbing, M. A.; Cascio, D.; Rogers, J. K.; Isaacs, F. J.; Kosuri, S.; Baker, D.; Fields, S.; Church, G. M.; Raman, S. Engineering an allosteric transcription factor to respond to new ligands. *Nat. Methods* **2016**, *13*, 177−183.
(25) Somervuo, P.; Koskinen, P.; Mei, P.; Holm, L.; Auvinen, P.; Paulin, L. BARCOSEL: a tool for selecting an optimal barcode set for high-throughput sequencing. *BMC Bioinformatics* **2018**, *19*, 257.
(26) Ramos, J. L.; Rojo, F.; Zhou, L.; Timmis, K. N. A family of positive regulators related to the *Pseudomonas putida* TOL plasmid XylS and the *Escherichia coli* AraC activators. *Nucleic Acids Res.* **1990**, *18*, 2149−2152.
(27) Kaldalu, N.; Toots, U.; de Lorenzo, V.; Ustav, M. Functional domains of the TOL plasmid transcription factor XylS. *J. Bacteriol.* **2000**, *182*, 1118−1126.
(28) Ibarra, J. A.; Pérez-Rueda, E.; Segovia, L.; Puente, J. L. The DNA-binding domain as a functional indicator: the case of the AraC/ XylS family of transcription factors. *Genetica* **2008**, *133*, 65−76.
(29) Michán, C.; Kessler, B.; de Lorenzo, V.; Timmis, K. N.; Ramos, J. L. XylS domain interactions can be deduced from intraallelic dominance in double mutants of *Pseudomonas putida*. *Mol. Gen. Genet. MGG* **1992**, *235*, 406−412.
(30) Manzanera, M.; Marqués, S.; Ramos, J. L. Mutational analysis of the highly conserved C-terminal residues of the XylS protein, a member of the AraC family of transcriptional regulators. *FEBS Lett.* **2000**, *476*, 312−317.
(31) Monteiro, L. M. O.; Arruda, L. M.; Sanches-Medeiros, A.; Martins-Santana, L.; Alves, L. D. F.; Defelipe, L.; Turjanski, A. G.; Guazzaroni, M.-E.; de Lorenzo, V.; Silva-Rocha, R. Reverse engineering of an aspirin-responsive transcriptional regulator in *Escherichia coli*. *ACS Synth. Biol.* **2019**, *8*, 1890−1900.
(32) Domínguez-Cuevas, P.; Ramos, J.-L.; Marqués, S. Sequential XylS-CTD binding to the Pm promoter induces DNA bending prior to activation. *J. Bacteriol.* **2010**, *192*, 2682−2690.
(33) Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternpberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845−858.
(34) Schäper, S.; Steinchen, W.; Krol, E.; Altegoer, F.; Skotnicka, D.; Søgaard-Andersen, L.; Bange, G.; Becker, A. AraC-like transcriptional activator CuxR binds c-di-GMP by a PilZ-like mechanism to regulate extracellular polysaccharide production. *Proc. Natl. Acad. Sci.* **2017**, *114*, E4822−E4831.
(35) Ruiz, R.; Marques, S.; Ramos, J. L. Leucines 193 and 194 at the N-terminal domain of the XylS protein, the positive transcriptional regulator of the TOL meta-cleavage pathway, are involved in dimerization. *J. Bacteriol.* **2003**, *185*, 3036−3041.
(36) Jacquet, P.; Hiblot, J.; Daudé, D.; Bergonzi, C.; Gotthard, G.; Armstrong, N.; Chabrière, E.; Elias, M. Rational engineering of a native hyperthermostable lactonase into a broad spectrum phosphotriesterase. *Sci. Rep.* **2017**, *7*, 16745.
(37) McIntosh, J. A.; Heel, T.; Buller, A. R.; Chio, L.; Arnold, F. H. Structural adaptability facilitates histidine heme ligation in a cytochrome P450. *J. Am. Chem. Soc.* **2015**, *137*, 13861−13865.
(38) Horita, S.; Kataoka, M.; Kitamura, N.; Nakagawa, T.; Miyakawa, T.; Ohtsuka, J.; Nagata, K.; Shimizu, S.; Tanokura, M. An engineered old yellow enzyme that enables efficient synthesis of

https://doi.org/10.1021/acssynbio.1c00564
ACS Synth. Biol. 2022, 11, 473−485

(4*R*,6*R*)-actinol in a one-pot reduction system. *ChemBioChem* **2015**, *16*, 440−445.
(39) Kreß, N.; Halder, J. M.; Rapp, L. R.; Hauer, B. Unlocked potential of dynamic elements in protein structures: channels and loops. *Curr. Opin. Chem. Biol.* **2018**, *47*, 109−116.

https://doi.org/10.1021/acssynbio.1c00564
ACS Synth. Biol. 2022, 11, 473−485