ACS SyntheticBiology

pubs.acs.org/synthbio

Research Article

Development of a Biosensor Platform for Phenolic Compounds Using a Transition Ligand Strategy

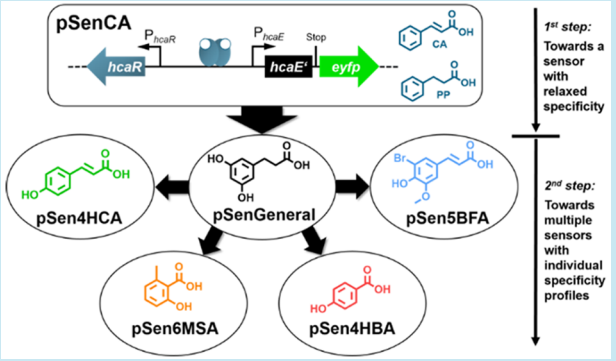
Lion Konstantin Flachbart, Christoph Gerhard Wilhelm Gertzen, Holger Gohlke, and Jan Marienhagen*

Cite This: *ACS Synth. Biol.* 2021, 10, 2002–2014

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The time-consuming and laborious characterization of protein or microbial strain designs limits the development of high-performance biocatalysts for biotechnological applications. Here, transcriptional biosensors emerged as valuable tools as they allow for rapid characterization of several thousand variants within a very short time. However, for many molecules of interest, no specific transcriptional regulator determining a biosensor's specificity is available. We present an approach for rapidly engineering biosensor specificities using a semirational transition ligand approach combined with fluorescence-activated cell sorting. In this two-step approach, a biosensor is first evolved toward a more relaxed-ligand specificity before using the resulting variant as the starting point in a second round of directed evolution toward high specificity for several chemically different ligands. By following this strategy, highly specific biosensors for 4-hydroxybenzoic acid, *p*-coumaric acid, 5-bromoferulic acid, and 6-methyl salicylic acid were developed, starting from a biosensor for the intracellular detection of *trans*-cinnamic acid.



KEYWORDS: *biosensors, directed evolution, fluorescence-activated cell sorting, protein engineering, molecular dynamics simulations*

Directed evolution approaches are widely used in academia and industry to engineer proteins and microbial production strains toward increased productivity. Typically, the large and genetically diverse libraries of proteins or microorganisms generated prior to screening for variants with the desired properties in such directed evolution campaigns harbor only a few variants with advantageous mutations. Hence, isolating the desired mutants is either very costly when massive multiplexing approaches are used, or the screened variants represent only a small fraction of the respective library.[1,2] In this context, transcriptional biosensors enabling the conversion of an otherwise inconspicuous production phenotype into a readily detectable output signal emerged as valuable tools as they allow for rapid characterization of several thousand protein or strain variants within a concise time frame.[3] As part of such biosensors, a transcriptional regulator acts as a ligand-detecting component by specifically binding the molecule of interest or a key intermediate of a pathway. In response, ideally in a dose-responsive manner, expression of a fluorescent reporter gene (*e.g.*, *eyfp*, enhanced yellow fluorescent protein) acting as an output component is induced (Figure 1A). In combination with fluorescence-activated cell sorting (FACS), individual cells can be isolated from cultures without the need for individual cultivation of each variant, which allows for a screening throughput orders of magnitude higher than other well-established microtiter plate-based screening formats.

If no biosensor for the molecule of interest is available, transcriptome analysis can identify suitable promoter regions to construct biosensors.[4] Furthermore, global and species-specific databases allow for data mining to identify suitable transcription factor (TF)/promoter pairs.[5−10] Another strategy is altering TF specificity toward the compound of interest by engineering the TF's signal sensing domain. Among others, successful examples for this approach describe altering ligand specificity of AraC and LacI from *Escherichia coli*. AraC, naturally accepting L-arabinose, was engineered to accept D-arabinose, mevalonate, ectoine, triacetic acid lactone, vanillin, orsellinic acid, or salicylic acid.[11−16] LacI, known to repress the *lac* operon in the absence of allolactose or isopropyl-*β*-D-thiogalactopyranoside, was engineered to also accept D-fucose, lactitol, sucralose, or gentiobiose.[17] Recently, the ligand spectrum of LysG, a transcriptional regulator of *Corynebacterium glutamicum* accepting all three basic amino acids could be focused using a semirational directed evolution approach. With the engineered L-lysine-insensitive LysG variant as metabolite-

Received: April 16, 2021
Published: August 9, 2021

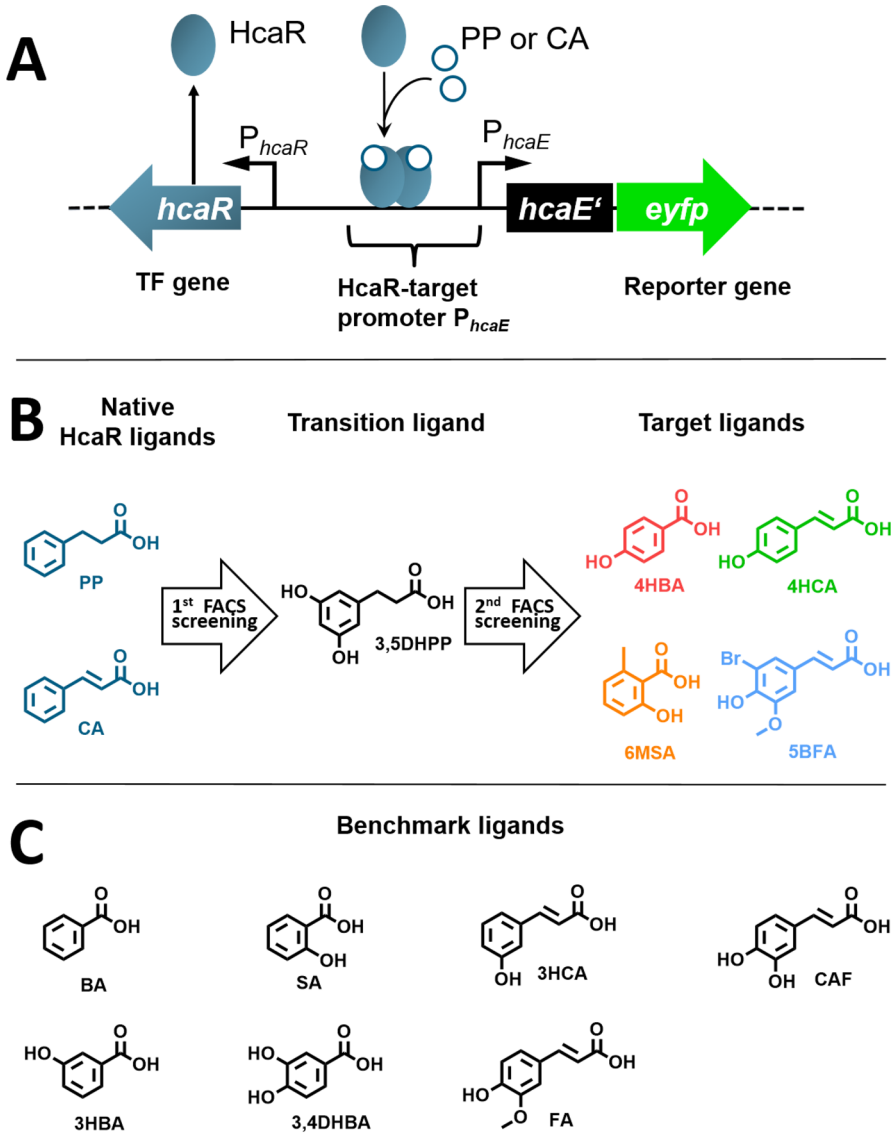


Figure 1. The pSenCA biosensor. (A) Schematic representation of the pSenCA biosensor. The regulator gene *hcaR* is constitutively expressed. Upon binding of phenylpropionic acid (PP) or cinnamic acid (CA), HcaR activates transcription of a truncated *hcaE′* gene and the reporter gene *eyfp*. (B) Native HcaR ligands phenylpropionic acid and cinnamic acid and all ligands used for FACS screening: 3,5DHPP, 3,5-dihydroxyphenylpropionic acid; 4HCA, p-coumaric acid; 5BFA, 5-bromoferulic acid; 6MSA, 6-methylsalicylic acid; 4HBA, 4-hydroxybenzoic acid. (C) Ligands used for the characterization of biosensor constructs: BA, benzoic acid; SA, salicylic acid; 3HBA, 3-hydroxybenzoic acid; 3,4DHBA, 3,4-dihydroxybenzoic acid; 3HCA, 3-hydroxycinnamic acid; CAF, caffeic acid; FA, ferulic acid.

detecting biosensor component, L-histidine producing *C. glutamicum* variants could be successfully isolated from a chemically mutagenized culture using FACS.[18] In all of these cases, the crystal structure of the target protein in complex with its native inducer molecule was available, enabling the semirational engineering of the TF toward accepting the small molecule of interest.

However, often such straightforward TF engineering campaigns do not yield biosensor variants with the desired ligand specificities. In such cases, the ligand of interest may be chemically too different from the natural ligand(s) of the TF. As a result, directed evolution toward the binding of the ligand of interest would require multiple amino acid substitutions in the ligand binding site of the TF. This is often either impossible to predict or unattainable by available directed evolution strategies. Noteworthy, the structural impact of the necessary modifications might also be too drastic, either disrupting communication between the signal sensing domain and DNA-binding domain or, in the worst case, pushing the TF protein below the stability threshold.[19]

With the aim to tackle this challenge, we propose a two-step FACS-assisted directed evolution strategy to engineer TF ligand specificities directly in the biosensor context. In the first step, the TF is evolved toward a more relaxed ligand specificity using a small molecule as a transition ligand, which shares overall similarities to the native ligand and the desired ligand(s). The isolated TF variant(s) then serve(s) as a starting point for the second directed evolution cycle to generate highly specific biosensors to detect chemically different compounds. As proof of concept, we selected the previously constructed transcriptional biosensor pSenCA that detects *trans*-cinnamic acid (CA) and phenylpropionic acid (PP). pSenCA proved to be very useful in a directed evolution campaign yielding highly active aromatic amino acid ammonia

https://doi.org/10.1021/acssynbio.1c00165
ACS Synth. Biol. 2021, 10, 2002−2014

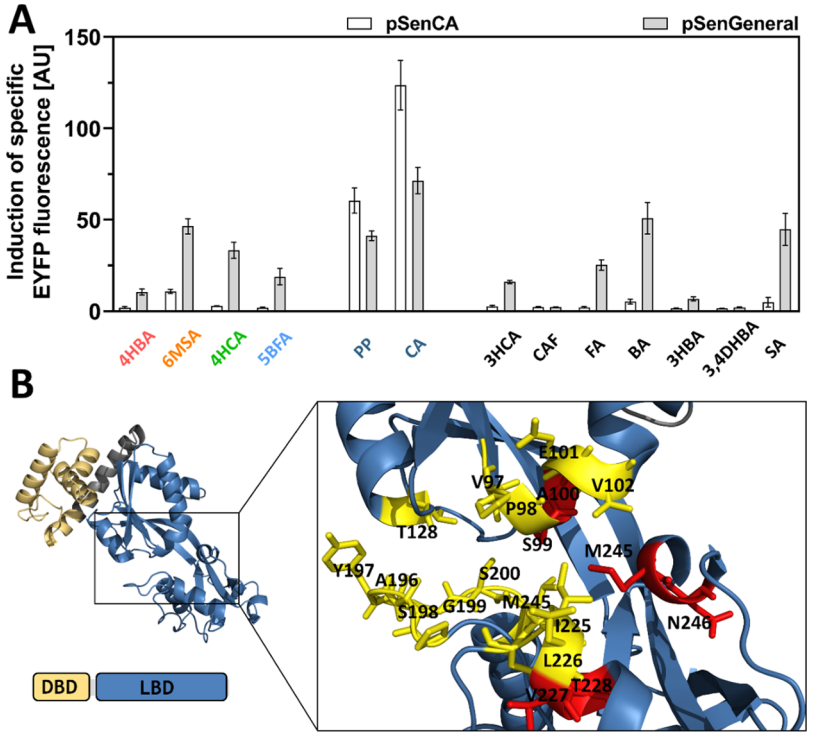


Figure 2. Specificity profiles of biosensors pSenCA and pSenGeneral. (A) Specificity profile of pSenCA and pSenGeneral for 13 different aromatic compounds. Mean fold induction in specific EYFP fluorescence (three biological replicates), error bars depict standard deviation. (B) Predicted structure of *E. coli* HcaR in cartoon representation. The DNA binding domain (DBD) is shown in yellow and the ligand binding domain (LBD) in blue. Overview of the complete structure (left) and a close-up of the proposed ligand binding cleft (right). Residues targeted for site-saturation mutagenesis (SSM) and multisite-directed saturation mutagenesis (MSSM) are highlighted and labeled. Yellow color depicts the initial target set used to broaden the specificity profile of pSenCA; red color depicts additional amino acid residues targeted during the second biosensor-based FACS screening.

lyases for applications in *E. coli*.[20] To obtain specific biosensor variants for 4-hydroxybenzoic acid (4HBA), *p*-coumaric acid (4HCA), 5-bromoferulic acid (5BFA), and 6-methylsalicylic acid (6MSA), we followed the described two-step semirational FACS-based engineering approach using 3,5-dihydroxyphenylpropionic acid (3,5DHPP) as the transition ligand. Obtained biosensor variants were characterized in detail as to their ligand spectrum, and the structural consequences of the identified amino acid substitutions were investigated using molecular modeling and simulation studies.

RESULTS AND DISCUSSION

The transcriptional biosensor pSenCA for *E. coli* comprises the regulatory elements of the *hca* cluster of *E. coli*, involved in the catabolization of PP and CA.[20,21] In this regulatory circuit, the transcriptional activator HcaR induces transcription from the P_{hcaE} promoter in the presence of intracellular PP or CA. As part of the biosensor, the HcaR/P_{hcaE} promoter combination induces expression of the reporter gene *eyfp* encoding the enhanced yellow fluorescent protein in response to the same compounds (Figure 1A). However, specific biosensors for the biotechnologically interesting aromatic compounds 4HCA, 4HBA, and 6MSA are highly desired. The same is true for 5BFA carrying bromine in *ortho* position as it offers the possibility to carry out palladium-catalyzed cross-coupling of the bromine group (Figure 1B).[22] The first biosensor for 4HCA and 4HBA have been constructed, but no sensors for 6MSA or 5BFA are available.[23−25] Hence, we aimed at expanding the detection repertoire of pSenCA, by semirationally engineering the ligand specificity of the wild-type HcaR regulator ($HcaR^{WT}$).

First, the specificity of $HcaR^{WT}$ in the biosensor context of pSenCA was characterized in detail using the two native ligands PP and CA, the four target ligands (4HCA, 4HBA, 6MSA, 5BFA), and a set of seven chemically similar aromatic compounds (benzoic acid (BA), salicylic acid (SA), 3-hydroxycinnamic acid (3HCA), caffeic acid (CAF), 3-hydroxybenzoic acid (3HBA), 3,4-dihydroxybenzoic acid (3,4DHBA), and ferulic acid (FA)), which served as benchmark ligands (Figure 1B and 1C). As host strain, the previously constructed *E. coli* DH10BΔ*hcaREFCBD* strain (*E. coli* Δ*hca*) was used, as it does not carry the *hca* operon responsible for the first steps of PP and CA degradation.[20,26,27] $HcaR^{WT}$ exhibits a clear preference for PP and CA with 120- and 60-fold induction of specific EYFP fluorescence, respectively (Figure 2A). The only other ligand tested triggering a response higher than 10-fold was 6MSA.

Regulation and function of the *hca* operon in *E. coli* has been studied in detail. However, to the best of our knowledge, no structural information regarding the ligand binding site or the ligand binding mode of HcaR is available, which is indispensable for a semirational protein engineering campaign.[26−28] To identify the ligand binding site, a homology model of HcaR was generated using TopModel and scored by TopScore.[29,30] As judged by a secondary structure analysis using the blastp suite (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and the PROSITE database (https://prosite.expasy.org/prosite.html), HcaR belongs to the LysR type transcriptional regulator (LTTR) family.[31−37] The N-terminal DNA-binding

https://doi.org/10.1021/acssynbio.1c00165
ACS Synth. Biol. 2021, 10, 2002−2014

Table 1. Amino Acid Substitutions in the Transcriptional Regulator HcaR Observed in the Isolated Biosensor Variants[a]

Biosensor variant	Ligand	Residues targeted by SSM										Additional AS substitutions
		97	98	99	100	101	102	227	228	245	246	
pSenCA	-	V	P	S	A	E	V	V	T	M	N	
pSenGeneral	-					H	Y					
pSen4HCA1	4HCA	T	*			H	Y					L105F
pSen4HCA2	4HCA	W	W			H	Y					
pSen4HCA3	4HCA	H	W			H	Y					A13T
pSen4HCA4	4HCA			A	*	H	Y					
pSen4HCA5	4HCA			*	V	H	Y					
pSen4HCA6	4HCA					H	Y			I		
pSen4HCA7	4HCA					H	Y			I	D	
pSen4HCA8	4HCA					H	Y	*	C			
pSen4HCA9	**4HCA**							**T**	**H**			**G138D**
pSen4HBA1	4HBA					H	Y			S		A175V
pSen4HBA2	4HBA					H	Y			S	N	
pSen4HBA3	4HBA					H	Y			S	L	
pSen4HBA4	4HBA					H	Y			S	L	A70V
pSen4HBA5	4HBA					H	Y			S	H	A259V
pSen4HBA6	4HBA					H	Y			G	F	
pSen4HBA7	4HBA					H	Y			G	S	
pSen4HBA8	4HBA					H	Y			G	Y	
pSen4HBA9	**4HBA**					**H**	**Y**	**S**	**E**			
pSen5BFA1	5BFA					H	Y	F	L			
pSen5BFA2	5BFA					H	Y	M	V			
pSen5BFA3	5BFA					H	Y	W	*			
pSen5BFA4	**5BFA**					**H**	**Y**	**W**	**C**			
pSen5BFA5	5BFA					H	Y	W	V			
pSen5BFA6	5BFA					H	Y	W	V			A210V
pSen5BFA7	5BFA					H	Y	Y	L			
pSen5BFA8	5BFA	W	W			H	Y					
pSen6MSA1	6MSA	F	L			H	Y					
pSen6MSA2	6MSA	H	V			H	Y					
pSen6MSA3	6MSA	H	I			H	Y					
pSen6MSA4	**6MSA**	**W**	**L**			**H**	**Y**					
pSen6MSA5	6MSA							S	P			
pSen6MSA6	6MSA							S	P			T129A

[a]The column "additional amino acid substitutions" lists additional amino acid substitutions found. Sensors in bold letters were subsequently characterized in more detail. * = Synonymous mutation.

domain comprises residues 1−86, and a large ligand-binding domain is spanning from residue 90 to 292. One of the top 10 threading templates for the HcaR homology model construction was a mutant (R156H) of the homotetrameric ligand binding domain of the regulator CatM (PDB ID: 3GLB) controlling benzoate consumption in *Acinetobacter baylyi* ADP1.[38] The structure of a CatM monomer with bound muconic acid (PDB ID: 3GLB) was aligned with the HcaR homology model. Assuming that the ligand binding site of CA of HcaR is in the same location as the one of muconic acid in CatM, we identified all residues protruding in the direction of the presumed binding site, namely, T128, P195, V196, Y197, S198, G199, S200, L201, V97, and P98 (Figure 2B). Additionally, residues E101, V102, N224, I225, and L226 were selected as they form the bottom of the surmised binding cleft. Subsequently, up to three neighboring codons were mutagenized in parallel using multi site-directed saturation mutagenesis (MSSM). The resulting seven *hcaR*-libraries were cloned directly into the genetic context of the biosensor for a FACS-based screening (Table S3).

FACS-Based Screening in the Presence of a Transition Ligand Identified a Mutational Hotspot in the Ligand Binding Domain of HcaR. In a first step toward HcaR variants with altered ligand specificity, we conducted a FACS screening to isolate HcaR variants exhibiting a less stringent ligand profile than wild-type HcaR. The variants will be then engineered toward different aromatic ligands of interest in a second FACS screen. For this, 3,5DHPP was selected as a transition ligand to isolate such variants with a generally broadened ligand profile. The propane tail of 3,5DHPP allows for a rotation of the molecule around the α- and β-carbon bond, and the presence of the two hydroxy groups at the aromatic ring in *meta* position demands a larger binding pocket in comparison to the natural ligands PP and CA (Figure 1B).

The seven biosensor libraries were pooled, and three successive rounds of FACS-based screening were conducted to enrich HcaR-variants capable of binding 3,5DHPP and inducing *eyfp* expression. In each round, 3,5DHPP was supplemented, and 200 000 cells were characterized with regard to their EYFP-fluorescence. The top 5% of all fluorescent cells in each round were sorted. Following this step, 96 variants were isolated after the third enrichment step and cultivated without 3,5DHPP supplementation to identify possible biosensor variants giving a fluorescence response in the absence of any ligand (false-positives). The 22 biosensor variants exhibiting the lowest basal sensor response were individually assayed regarding their fluorescence response to 3,5DHPP in microtiter plates. Sequencing of the *hcaR* gene of these 22 variants revealed six different groups, all carrying mutations in two consecutive codons for E101 and V102, which are located in the ligand binding domain (Table S4).

https://doi.org/10.1021/acssynbio.1c00165
ACS Synth. Biol. 2021, 10, 2002−2014

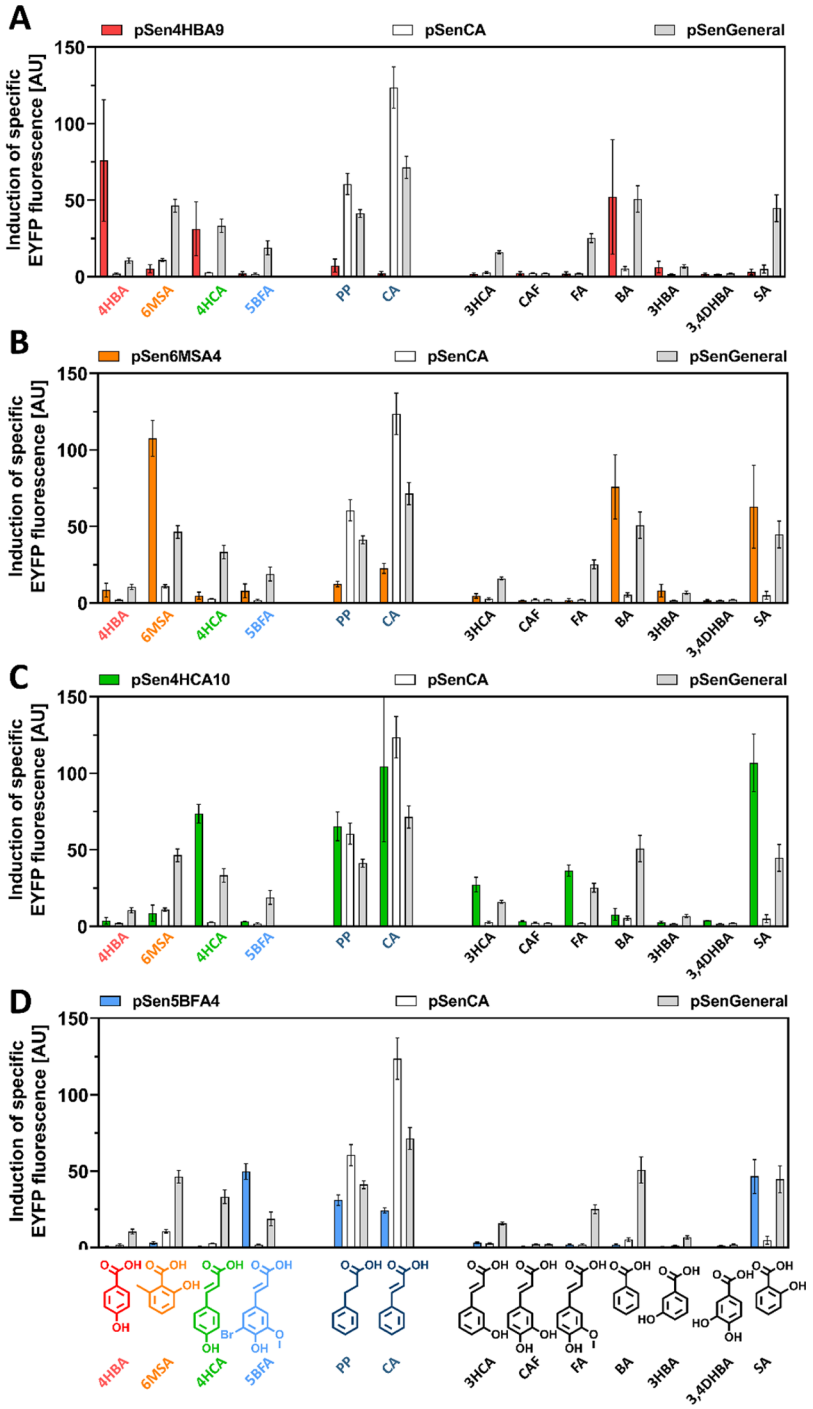


Figure 3. Specificity profiles of the evolved biosensor variants pSen4HBA9 (A), pSen6MSA4 (B), pSen4HCA10 (C), and pSen5BFA4 (D) in comparison to pSenCA and pSenGeneral. The mean fold induction in specific EYFP fluorescence (three biological replicates) in response to the presence of various ligands serves as a measure to compare the six biosensors. Error bars depict the standard deviation.

These mutations substitute E101 with relatively small, uncharged amino acids (C, S, A, or T) in five variants, whereas V102 was substituted with F in all cases. The sixth mutant group was the most frequent one, carrying E101H and V102Y. As these latter variants showed the highest fold induction in the presence of 3 mM 3,5DHPP, *hcaR*_E101H,V102Y was subcloned into a new pSenCA plasmid to exclude any possible mutations in the vector backbone and subsequently retransformed into *E. coli* Δ*hca*. This biosensor was then characterized as to the response to the selected ligands (PP, CA, 3HCA, 4HCA, CAF, FA, 5BFA, BA, 3HBA, 4HBA, 3,4DHBA, SA, and 6MSA) (Figure 1B). Interestingly, the biosensor exhibited a reduced fold induction in the presence of PP and CA, while the response to 4HCA, FA, BA, 3HBA, SA, and 6MSA was significantly increased (Figure 2C). Due to its relaxed ligand spectrum, this biosensor was named pSenGeneral and used as the starting variant for all subsequent screenings.

Both amino acid substitutions of the HcaR variant in pSenGeneral, E101H and V102Y, were incorporated into the

https://doi.org/10.1021/acssynbio.1c00165
ACS Synth. Biol. 2021, 10, 2002−2014

HcaR homology model, and their influence on the protein structure was assessed. In the wild-type HcaR protein, V102 and I225 constrain the binding cleft by pointing toward each other. In the E101H/V102Y variant, V102Y rotates away from I225 and toward M245, which widens the binding cleft and renders M245 at the bottom accessible (Figure S1).

The Second Round of FACS-Screening Yielded Specific Biosensor Variants. Motivated by identifying the mutagenic hot spots at positions E101 and V102 of HcaR, the two neighboring residues S99 and A100 were also targeted by MSSM as part of the second screening campaign using *hcaR* E101H/V102Y as a template to isolate specific biosensor variants. Moreover, V227 and T228 were also selected for MSSM to investigate secondary effects on HcaR ligand specificity by reorienting the preceding residues N224, I225, and L226 (Figure 2B). Additionally, M245 and N246 were targeted, as the structural changes in pSenGeneral suggest an influence of M245 on HcaR specificity. All 18 selected codons were targeted by site-directed mutagenesis (SSM) and MSSM in *hcaR* E101H/V102Y. Presuming that mutagenesis of pSenCA carrying the wild-type regulator could still yield variants with increased specificity toward any of the desired aromatic target ligands, the newly identified positions were also saturated in the wild-type HcaR protein as part of the original biosensor pSenCA. In addition, considering the enormous throughput capabilities of FACS, epPCR libraries of *hcaR* and *hcaR* E101H/V102Y were generated and cloned into the biosensor plasmid yielding 2×10^5 variants each. These additional libraries were also screened by FACS in the presence of the target ligands for identifying additional mutational hot spots with an impact on ligand specificity outside of the ligand binding site.

To make the FACS screening of the 21 resulting libraries (10 *hcaR* libraries, 9 *hcaR E101H/V102Y* libraries and 2 epPCR libraries) more feasible, all libraries were pooled. In a control experiment, all variants (2×10^5 SSM/MSSM variants and 4×10^5 epPCR variants) were incubated without supplementation of any target ligand to check for the presence of cells constitutively expressing *eyfp*. As only a few biosensor variants (<0.1%) with a strong fluorescence response in the absence of any ligand could be observed during FACS screening, the combined library was directly used to screen for biosensor responses to each of the target ligands (4HCA, 4HBA, 5BFA, 6MSA) individually. In total, five FACS enrichment steps (repetitive cultivation and FACS-screening in the presence of each target ligand) were performed, and 184 single clones from each recovery culture were subsequently characterized individually as to their response to their respective target ligand (Figures S2, S3). The sensor plasmids of 24 variants for each of the four target ligands (96 variants in total) showing the highest fluorescence response were sequenced to identify nonsynonymous mutations in the P_{hcaE} promoter and *hcaR* gene (Table 1). In total, 31 unique biosensors were isolated during the second round of directed evolution, 28 of these stemming from the HcaR E101H/V102Y variant of pSenGeneral. Only one HcaR variant carrying the two consecutive substitutions V97W and P98W in addition to E101H and V102Y was isolated twice in two independent FACS experiments when screening for 4HCA- and 5BFA-specific sensors (pSen4HCA2 and pSen5BFA8). The amino acid substitution pattern of all other isolated biosensor variants can be attributed to a specific target compound. Noteworthy, of the 20 codons targeted for SSM or MSSM, ten are highly conserved. In addition, no HcaR mutant stemming from epPCR library was isolated during the FACS screening campaign. One explanation is the low genetic variation typically introduced by epPCR, yielding only variants carrying a very limited number of single nucleotide polymorphisms.[39] Similar findings were reported in a study focusing on the modulation of the ligand specificity of an AraC-based biosensor in which variants specific to D-arabinose showed a significantly lower sensor response when they were obtained from epPCR libraries and not from semirational SSM libraries.[11]

However, all isolated biosensor variants were characterized concerning their ligand spectrum using all native, target, and a set of seven chemically similar aromatic benchmark ligands to detect unexpected ligand specificities and support subsequent molecular modeling studies (Table S5). For each of the four target ligands, the variant with the most specific response profile was compared to pSenCA and pSenGeneral in more detail (Figure 3). The best 4HBA sensor, pSen4HBA9, shows almost no residual sensor response to CA and PP but induces a 75-fold induction in specific EYFP fluorescence in the presence of 4HBA. Other, but weaker inducers of this biosensor are BA (50-fold) and 4HCA (27-fold). The 6MSA sensor (pSen6MSA4) is characterized by a low residual response to PP and CA (10-fold and 20-fold, respectively), whereas 6MSA triggers a 110-fold response of this biosensor. Other inducers are BA (75-fold) and SA (60-fold). The 4HCA sensor (pSen4HCA10) offers the same response to PP and CA as pSenCA, but shows a 75-fold fluorescence response to 4HCA, which cannot be recognized by pSenCA. Surprisingly, SA triggers a more than 100-fold response of pSen4HCA10, but 3HCA being very similar to the target compound 4HCA induces a rather low 25-fold sensor response. The 5BFA sensor (pSen5BFA4) offers a 50-fold induction of specific EYFP fluorescence upon the addition of 5BFA as desired but also shows biosensor response to SA of comparable intensity. The only other but weaker inducers of this biosensor are the native ligands PP and CA (31- and 24-fold, respectively).

Noteworthy, the importance of specificity for other substances depends on the desired application of the respective sensor. For example, 6MSA is formed by polyketide synthases *via* subsequent condensation of three malonyl-CoA moieties to an acetyl-CoA starter unit.[40−43] Throughout the condensation, no intermediate with structural similarity to 6MSA is formed, enabling the use of pSen6MSA despite its BA- and SA-sensitivity. Furthermore, if SA is the ligand of interest, pSen6MSA can be used for screening approaches as SA is derived from chorismate without the formation of other inducers of pSen6MSA.[44] However, when highly specific biosensor variants without response to any other ligand are desired, (*e.g.*, for improving simple transformations such as hydroxylations or halogenations), one could subject the engineered biosensor variants to a subsequent counterscreening step to isolate biosensors without any response to the presence of the undesired ligands. Such a FACS-based screening/counterscreening strategy was already successfully used during the development of an L-lysine insensitive L-histidine biosensor.[18]

Molecular Modeling and Simulation Studies to Assess the Observed Selectivity Profiles of the Sensors. To structurally assess the observed selectivity profiles of the individual biosensors, homology models of the HcaR variants of pSen4HBA9, pSen4HCA10, pSen5BFA4, and pSen6MSA4

https://doi.org/10.1021/acssynbio.1c00165
ACS Synth. Biol. 2021, 10, 2002−2014

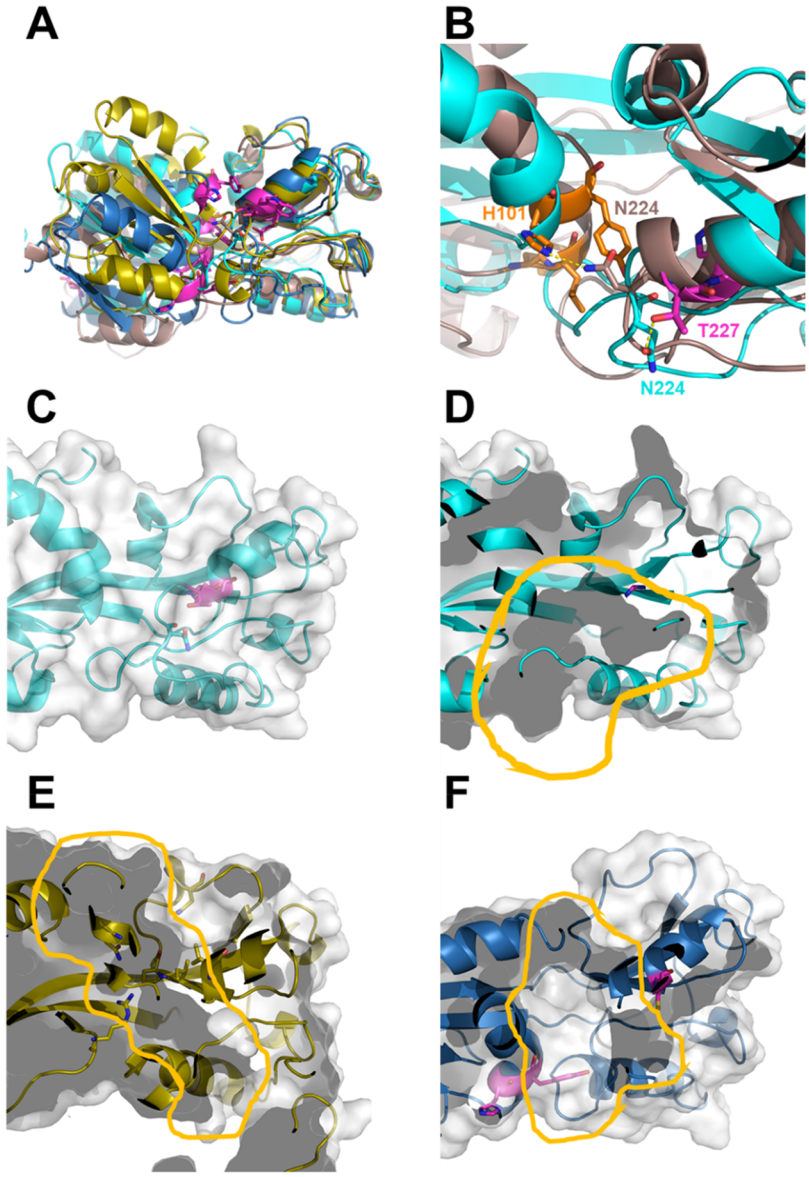


Figure 4. Differences in the HcaR overall structure and the structure of the binding pocket of the four isolated biosensor variants. (A) The cluster representatives from MD simulations of the HcaR variants of pSen4HCA10 (cyan), pSen4HBA9 (yellow), pSen5BFA4 (blue), and pSen6MSA4 (brown) in cartoon representation without the first 86 residues were aligned onto the portion of HcaR of pSen4HCA10 that forms the foot-like pocket of the binding site of the regulator protein (residues 161−266). Substituted residues are shown in magenta. (B) The binding pocket of the HcaR variant of pSen6MSA4 is narrower than the binding pocket of the HcaR variant of pSen4HCA10, presumably due to the hydrogen bond (yellow dashed line) of H101 with N224 in the HcaR variant of pSen6MSA4, as opposed to the hydrogen bond of T227 with N224 in the HcaR variant of pSen4HCA10. Here, the substituted residues in the HcaR variant of pSen6MSA4 are shown in orange. (C) Surface representation of the HcaR variant of pSen4HCA10 with a cut-away (D) revealing the binding pocket (gray), which is open to the bottom. (E) Cut-away, revealing the binding pocket of the HcaR variant of pSen4HBA9, which is closer to the back and open to the top and bottom. (F) Cut-away, revealing the binding pocket of the HcaR variant of pSen5BFA4, which is on the surface with pockets on the right and top. In panels (D) to (F), the yellow lines are based on the volumes of the binding sites tracked over the course of the MD simulations.

were generated as the HcaR variants of pSenCA and pSenGeneral before. The ligands CA, PP, 4HCA, 4HBA, 5BFA, and 6MSA were docked into the binding pockets of all regulator proteins, utilizing AutoDock as a docking engine and DrugScore as a scoring function.[45−47] Very similar valid docking poses for each ligand-regulator combination were obtained. Apparently, differences in local ligand−protein interactions in the static homology models cannot explain differential HcaR regulator ligand recognition spectra. These results point to the influence of structural dynamics on the HcaR regulator function. Hence, the homology models of each of the HcaR regulators were subjected to five replicas of molecular dynamics (MD) simulations of 2 μs length each. The structures of each of the replicas were clustered separately. The representative structure of the largest cluster of each of the replicas was used for docking. In the MD simulations, the HcaR regulators showed pronounced movements. Notably, the DNA binding domain comprising residues 1−86 (Figure 2B), which is connected to the rest of HcaR *via* a flexible loop, is mobile and does not adopt a preferred conformation (Figure S5). Furthermore, the two domains that form the binding pocket and are interconnected by two β-strands are rotating

https://doi.org/10.1021/acssynbio.1c00165
ACS Synth. Biol. 2021, 10, 2002−2014

against each other, which influences the shape and location of the binding pocket (Figure 4).

These motions lead to differences in shape and detailed location of the binding sites depending on the substitutions in the HcaR variants. Direct comparison of the HcaR variant of pSen4HCA10, sensing a broad range of ligands, and the variant of pSen6MSA4, sensing a narrower range of ligands, reveals that the binding pocket of the pSen6MSA4 regulator is smaller than that of the pSen4HCA10 regulator (Figure 4B). This could have a pronounced impact on the ligand scope and is caused by differential interactions of N224 in both HcaR variants. Further differences in shape and detailed location of binding pockets are illustrated in Figure 4C−F, where the respective HcaR variants of pSen4HCA10, pSen4HBA9, and pSen5BFA4 are shown in the same orientation with a molecular surface. The different shapes and localizations of the binding pockets of representative structures taken from MD simulations of the four HcaR sensors lead to differential docking poses. In contrast to docking into the homology models, the selectivity profiles of the HcaR-based sensors can now be fairly reproduced: The presence or absence of a response triggered by a ligand was correctly reproduced for 71% (17 out of 24) of the ligands. Additionally, the ligand to which the respective biosensor responds to most, always produced a valid docking pose in the respective HcaR variant and often specifically interacts with substituted amino acids. Particularly, valid docking poses of 5BFA were only found in the HcaR variant isolated in the biosensor-based FACS screening conducted to isolate biosensor variants for this ligand.

In the case of the HcaR variant isolated as part of the pSen4HCA10 sensor recognizing 4HCA, CA, and PP, valid docking poses were produced for these three ligands only, but other ligands were not accommodated by the binding pocket (Figure 5A, B). All the three ligands interact with H228, which results from the T228H substitution specific for the HcaR variant of pSen4HCA10 (Figure 5B).

As to the HcaR variant of pSen4HBA9 (Figure 5C, D), the predicted binding modes for 4HBA and 4HCA recognized by this sensor provide a possible explanation for the higher fold induction of the sensor by 4HBA than 4HCA. In addition to interactions formed by 4HCA, 4HBA forms a salt bridge to R146 with its carboxylic acid moiety (Figure S6). This additional salt bridge may confer a higher affinity of 4HBA, which may lead to the increased fold-induction of EYFP expression in the presence of 4HBA. In contrast, PP and CA only form a hydrogen bond with N249 in this HcaR variant, suggesting that their binding is weaker. Noteworthy, when considering that only a functional biosensor readout is available (Figure 3), it is also possible that PP and CA bind to the HcaR variant of pSen4HBA9 but do not activate the sensor.

In the HcaR variant of pSen5BFA4, valid docking poses were identified for PP, 5BFA, 6MSA, 4HCA, and CA (Figure S7A). Of these, PP, 5BFA, and CA lead to a detectable biosensor response of pSen5BFA4. FA, which is structurally similar to 5BFA and CA, does not lead to a detectable response; however, a docking analysis suggests that the binding modes of the phenyl ring in 5BFA and CA are structurally closer and buried deeper inside the binding pocket (Figure S7B) than that of FA (data not shown). Most of the ligands form a hydrogen bond with Y102, which results from the V102Y substitution in this HcaR variant (Figure S7B). Finally,

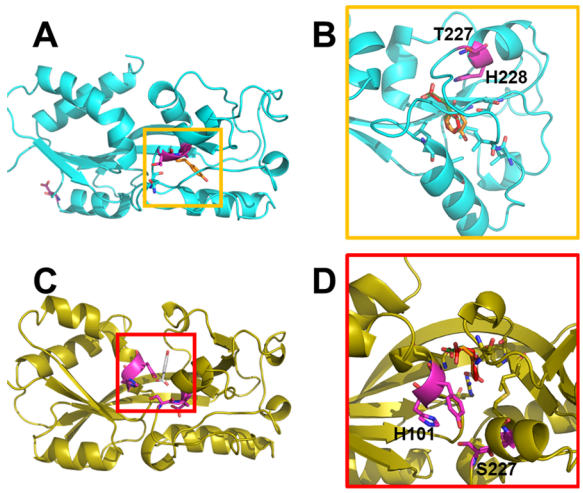


Figure 5. Docking poses of ligands in the HcaR variants of pSen4HCA10 and pSen4HBA9. HcaR variants of (A) pSen4HCA10 (cyan) and (C) pSen4HBA9 (yellow) highlighting the location of the binding pocket (square) and the most suitable ligands of the respective biosensor variant (4HCA, orange; 4HBA, white). Substituted residues in comparison to the wild-type HcaR are shown in magenta. (B) Docking poses of 4HCA (orange), CA (green), and PP (red) in HcaR variants of pSen4HCA10. All three ligands form hydrogen bonds to H228. (D) Docking poses of 4HBA (white), CA (green), 4HCA (orange), and PP (red) in the HcaR variant of pSen4HBA9. 4HBA is rotated by 180° compared to all other ligands, with its carboxylic acid moiety pointing into the binding pocket.

in the HcaR variant of pSen6MSA4, docking poses were identified for ligands 6MSA, 4HCA, and 4HBA. 6MSA eliciting the highest biosensor response is also the only ligand that forms hydrogen bonds with N230 of the HcaR variant (Figure S7D). The previously described interaction network involving N224 and the associated narrowing of the binding pocket leads to forming a shallower binding groove, which restricts the range of ligands able to bind to the HcaR variant of pSen6MSA4. The strengthening of interactions between the left and right domains, leading to the compaction and "upwards movement" of the binding pocket, is reflected in the comparatively low RMSD of this sensor (Figure S5G).

The results of the molecular modeling indicate that the substitutions have a pronounced influence on shape and detailed location of the respective binding pockets and their molecular recognition properties, which in turn dictates the spectrum of ligands recognized by each HcaR variant. Notably, the results reveal that substitutions outside of the binding pocket markedly influence the ligand recognition spectrum. For example, the HcaR variants of pSen4HCA8 and pSen5BFA4 only differ in the substitution V227W present in the latter variant. Although pSen4HCA8 recognizes most tested ligands (data not shown), pSen5BFA only responds to a small ligand subset (Figure 3). Although the substitution V227W points away from the binding pocket, it apparently modulates the structural dynamics of the binding pocket such that only a narrow ligand spectrum can be recognized. However, as the exact mechanism of ligand-induced transcriptional activation by HcaR is still unknown, predicting the effects of ligand binding to HcaR variants that lead to a fluorescence response is difficult, especially considering the pronounced flexibility of the binding pocket of HcaR.

https://doi.org/10.1021/acssynbio.1c00165
ACS Synth. Biol. 2021, 10, 2002−2014

CONCLUDING REMARKS

In summary, we engineered the ligand specificity of the transcriptional repressor HcaR, which acts as ligand-detecting component in the biosensor pSenCA, toward accepting the biotechnologically interesting aromatic compounds 4HCA, 4HBA, 5BFA, and 6MSA. Key to success was the strategy of first using a transition ligand to screen for an HcaR variant with a less stringent ligand profile before performing a second round of directed evolution to isolate several HcaR variants with new ligand specificities. With these semirational libraries available and the option to generate additional HcaR libraries, the repertoire of biosensor-detectable small aromatic compounds can be easily enlarged following the described principle. As a result, one can expect a large family of different HcaR-based biosensors and new insights into the structure−function relationships of this particular transcriptional regulator. Furthermore, this general strategy is of course not limited to HcaR and should be translatable to other transcriptional regulators. This offers the possibility to generate more custom-made biosensors to detect a broad range of biotechnologically interesting small molecules for a multitude of possible applications.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, Media, and Growth Conditions. All bacterial strains and plasmids used in this study and their relevant characteristics are listed in Supporting Table S1. For recombinant DNA work and library construction, *E. coli* TOP10 (Thermo Fisher Scientific, Waltham, MA, USA) was used. All chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA) or abcr GmbH (Karlsruhe, Germany). For recovery after electroporation, super optimal broth with catabolite repression (SOC) was used (20 g/L tryptone, 5 g/L yeast extract, 0.6 g/L NaCl, 0.2 g/L KCl, 10 mM MgCl/$MgSO_4$ and 20 mM glucose, pH 7). All strains were routinely cultivated at 37 °C in Lysogeny broth (LB) medium (10 g/L tryptone, 10 g/L NaCl and 5 g/L yeast extract).[48] Kanamycin (25 μg/mL) was used for selective propagation of biosensor plasmids. Kanamycin stocks were prepared as 1000× stocks in purified water and sterile filtered using a 200 nm filter. Target ligands were always prepared fresh as 100× stocks in DMSO.

Online monitoring of growth and formation of fluorescence was performed in 48 well microtiter FlowerPlates using the BioLector cultivation system (m2p-laboratories GmbH, Baesweiler, Germany).[49,50] The intensity of the backscatter light was used to follow biomass formation (wavelength 620 nm; signal gain factor 25). The enhanced yellow fluorescence protein (EYFP) fluorescence was determined as fluorescence emission at 532 nm (signal gain factor of 30) after excitation at a wavelength of 510 nm. Specific fluorescence was calculated as 532 nm fluorescence per 620 nm backscatter using Biolection software version 2.2.0.6 (m2p-laboratories GmbH, Baesweiler, Germany).

Molecular Biology. Techniques for DNA manipulation were performed according to standard protocols.[51] Enzymes were obtained from Thermo Fisher Scientific (Waltham, MA, USA) and used following the manufacturer's recommendations. Genes were amplified by PCR using the *Pfu Ultra* II DNA-polymerase (Agilent, Santa Clara, CA, USA) following the manufacturer's recommendation if not stated otherwise. Cloning of amplified PCR products was performed using Gibson assembly.[52] Oligonucleotide synthesis and Sanger DNA sequencing were performed at Eurofins MWG Operon (Ebersberg, Germany). All oligonucleotides used in this study are listed in Supporting Table S2.

Library Construction. For the generation of error-prone libraries, the *hcaR* gene was amplified from plasmid pSenCA using the Clontech Diversify kit (Takara Bio Europe, Saint-Germain-en-Laye, France) incorporating 2.3 mutations/kb and integrated into the pSenCA plasmid devoid of the wild-type *hcaR* gene using Gibson assembly. For (multi)site-saturation mutagenesis ((M)SSM), the pSenCA plasmid (6076 bp) was amplified using oligonucleotides with degenerated codons in a PCR reaction using the *PfuUltra* II DNA-polymerase (75 ng template DNA, 50 μL total volume, 10 pmol of each primer, 20 cycles, 300 s initial denaturation, 20 s denaturation, 20 s annealing, 372 s elongation, 420 s final elongation). Following all PCR amplification, template plasmid removal was performed using 1 μL of Anza 10 *Dpn*I (overnight, 37 °C). Plasmid libraries were purified using the Nucleospin Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany) and transformed into One Shot TOP10 electrocompetent *E. coli* cells following the manufacturer's recommendations (Thermo Fisher Scientific, Waltham, MA, USA). After 1 h of regeneration, cell suspensions were transferred into 50 mL LB medium and incubated for 20 h (37 °C, 120 rpm). The respective culture broth was subsequently used to prepare concentrated (5-fold) glycerol stocks that were snap-frozen in 100 μL aliquots for single use. Serial dilutions of the regenerated cultures were spread on selective agar plates to determine the size of the individual libraries. Library completeness, typically ranging from 95% to 100%, could be confirmed using the GLUE-IT algorithm.[53−55]

Cytometry and Fluorescence Activated Cell Sorting (FACS). Fifteen (15) mL LB medium were inoculated with 100 μL glycerol stock and cultivated (16 h, 37 °C, 120 rpm). This preculture was diluted 1/100 in 15 mL LB medium and incubated (1 h, 37 °C, 120 rpm). Afterward, 890 μL portions of the culture were transferred into 48 well flower plate (m2p-laboratories GmbH, Baesweiler, Germany) wells containing 9 μL 100× stock of the target ligand of choice at different concentrations depending on the ligand (phenylpropionic acid, 1 mM; *trans*-cinnamic acid, 1 mM; 3,5-dihydroxyphenylpropionic acid, 3 mM; *p*-coumaric acid, 1 mM; 5-bromoferulic acid, 2 mM; 6-methylsalicylic acid, 2 mM). The plates were sealed and incubated in a Multitron Pro HT Incubator (Infors AG, Bottmingen, Switzerland) for 20 h (37 °C, 900 rpm, 75% humidity, 3 mm throw).[49,50] Single-cell size- and fluorescence characteristics were determined and cell sorting was performed using a BD FACSAria II cell sorter (BD Biosciences, Franklin Lakes, NJ, USA) equipped with a 70 μm nozzle and using a sheath pressure of 70 psi. For excitation, a 488 nm blue solid laser was used. Forward-scatter characteristics (FSC) were recorded as small-angle scatter, and side-scatter characteristics (SSC) were recorded as orthogonal scatter of the 488 nm laser. A 502 nm long-pass, 530/30 nm band-pass filter combination enabled EYFP fluorescence detection. Debris and electronic noise were excluded from the analysis by electronic gating in the FSC-H against SSC-H plot. Additional gating was performed on the resulting population in the FSC-H against FSC-W plot to exclude doublets. Fluorescence analysis was always performed with the population resulting from this two-step gating procedure. Prior to measuring/sorting, cells were diluted to an OD_{600} below 0.1 where necessary using YNB base

https://doi.org/10.1021/acssynbio.1c00165
ACS Synth. Biol. 2021, 10, 2002−2014

buffer (6 g/L K_2HPO_4, 3 g/L KH_2PO_4 and 10 g/L 3-(*N*-morpholino)propanesulfonic acid (MOPS), pH 7), filtered using 50 μm Cup-type filters (BD Biosciences, Franklin Lakes, NJ, USA) and analyzed. During positive sorting, the upper 5% of most fluorescent cells were selected. In contrast, cells with a fluorescence equivalent to the bottom 96% of an *E. coli* TOP10 pSenCA culture without ligand supplementation were isolated during negative sorting. Routinely, 200 000 cells were sorted into 5 mL reaction tubes (Eppendorf AG, Hamburg, Germany), prefilled with 3 mL SOC medium using an in-house built adapter for 5 mL reaction tubes previously described.[56] To minimize residual sheath fluid in the recovery tube, sorted cells were centrifuged (10 min, 3000*g*, 4 °C). After removal of 3 mL supernatant, 4.5 mL fresh SOC medium was added. To reduce postsorting stress for the isolated variants, cells were regenerated for 30 min (37 °C, 170 rpm) without antibiotics. Afterward, the cells were transferred into 10 mL LB medium containing the appropriate antibiotic. This cell suspension was incubated for 20 h at 37 °C and 170 rpm). The following day, the culture was used to prepare glycerol stocks and to inoculate precultures for subsequent FACS steps. Sorting precision was always set to purity setting, and the total event rate while sorting never exceeded 16 000 events per second. FACSDiva 7.0.1 (BD Biosciences, San Jose, USA) was used for FACS control and data analysis. FlowJo for Windows 10.4.2 (FlowJo, LLC, Ashland, OR, USA) was used to generate high-resolution graphics of the obtained FACS data for visualization.

Rescreening and Characterization of Isolated Biosensor Variants. Following FACS screening, the regenerated cultures were spread on selective agar plates. Single colonies were picked to inoculate 200 μL LB medium in 96 well V-bottom plates (Brand GmbH + Co. KG, Wertheim, Germany). Precultures were cultivated in a Multitron Pro HT Incubator (Infors AG, Bottmingen, Switzerland) for 18 h (37 °C, 900 rpm, 75% humidity, 3 mm throw). From these precultures, 10 μL were used to inoculate 990 μL LB medium, followed by 20 h of cultivation with parameters described above. Subsequently, 10 μL of the preculture were used to inoculate 1090 μL LB medium in a Deepwell plate (96/2000 μL, PCR clean, Eppendorf AG, Hamburg, Germany). After 2 h of cultivation using the same parameters as described above, 495 μL were transferred to two fresh deep-well plates containing 5 μL of a 100× stock solution of the respective target ligand (phenylpropionic acid, 1 mM; *trans*-cinnamic acid, 1 mM; 3,5 dihydroxyphenylpropionic acid, 3 mM; *p*-coumaric acid, 1 mM; 5-bromoferulic acid, 2 mM; benzoic acid, 5 mM; salicylic acid 2 mM; 6 methylsalicylic acid, 2 mM; 3-hydroxybenzoic acid, 1 mM; 4-hydroxybenzoic acid, 5 mM; 3,4 dihydroxybenzoic acid, 5 mM; 3-hydroxycinnamic acid, 1 mM; caffeic acid, 1 mM; ferulic acid, 1 mM). After 20 h of cultivation (same parameters as above), 100 μL of the culture broth was transferred to a 96 well Flat-bottom plates (Brand GmbH + Co. KG, Wertheim, Germany), and the absorbance and fluorescence were determined using a M100 plate reader (Tecan Group, Maennedorf, Switzerland; $absorbance_{600\ nm}$; $fluorescence_{530\ nm}$ after excitation at 500 nm, gain 65). Absorbance and fluorescence of the cell-free medium were used for background subtraction of all samples.

Homology Modeling and Docking. For generating structural models of the isolated HcaR variants, the respective DNA sequences obtained from Eurofins MWG Operon were transcribed into their respective AA sequence and processed with TopModel and TopScore.[29,30] Using default program parameters, the comparative modeling included template identification, sequence alignment, modeling, refinement, and scoring.

For the molecular docking, ligands PP, 5BFA, 6MSA, 4HCA, 4HBA, and CA were drawn and converted into a 3D structure with Maestro Release 2017 (Schrödinger LLC, New York, NY, USA).[57] The ligands were subsequently docked into the binding pockets of the respective sensors utilizing a combination of AutoDock as a docking engine and the DrugScore2018 distance-dependent pair-potentials as an objective function. During docking, default parameters were used, except for the clustering RMSD cutoff, which was set to 2.0 Å.[45−47] Binding modes were considered valid if they were located in the binding pocket and contained in the largest cluster, which comprised at least 20% of all docking poses.

Identification of mutagenesis targets and preparation of protein structures were performed using PyMOL Molecular Graphics System 1.7 (DeLano Scientific LLC, South San Francisco, CA, USA).[58,59] Prism 8.02 (GraphPad Software, San Diego, CA, USA) was used for the preparation of diagrams.

Molecular Dynamics Simulations. The homology models of pSen4HCA10, pSen4HBA9, pSen5BFA4, and pSen6MSA4 were subjected to all-atom MD simulations. The variants were protonated with PROPKA according to pH 7.4, neutralized by adding counterions, and solvated in an octahedral box of TIP3P water with a minimal water shell of 12 Å around the solute.[60,61] The Amber package of molecular simulation software and the ff14SB force field were used to perform the MD simulations.[62,63] To cope with long-range interactions, the "Particle Mesh Ewald" method was used; the SHAKE algorithm was applied to bonds involving hydrogen atoms.[64,65] As hydrogen mass repartitioning was utilized, the time step for all MD simulations was 4 fs with a direct-space, nonbonded cutoff of 8 Å.[66] At the beginning, 17 500 steps of steepest decent and conjugate gradient minimization were performed; during 2500, 10 000, and 5000 steps positional harmonic restraints with force constants of 25 kcal mol^{-1} $Å^{-2}$, 5 kcal mol^{-1} $Å^{-2}$, and zero, respectively, were applied to the solute atoms. Thereafter, 50 ps of NVT-MD (constant number of particles, volume, and temperature) simulations were conducted to heat up the system to 100 K, followed by 300 ps of NPT-MD (constant number of particles, pressure, and temperature) simulations to adjust the density of the simulation box to a pressure of 1 atm and to heat the system to 300 K. During these steps, a harmonic potential with a force constant of 10 kcal mol^{-1} $Å^{-2}$ was applied to the solute atoms. As the final step in thermalization, 300 ps of NVT-MD simulations were performed while gradually reducing the restraint forces on the solute atoms to zero within the first 100 ps of this step. Afterward, five independent production runs of NVT-MD simulations with 2000 ns length each were performed. For this, the starting temperatures of the MD simulations at the beginning of the thermalization were varied by a fraction of a Kelvin. The location of the binding pocket was tracked using VMD.[66] Conformations of all the five independent production runs for each variant were subsequently clustered by a hierarchical agglomerative clustering algorithm as implemented in CPPTRAJ using the root-mean-square deviation of heavy atoms after superimposition to residues 87−296 as a measure and aiming for five clusters.[67] The cluster representative of the largest cluster for each variant was then used for docking.

https://doi.org/10.1021/acssynbio.1c00165
ACS Synth. Biol. 2021, 10, 2002−2014

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.1c00165.

Details regarding strains, plasmids, and oligonucleotides used, more information on the isolated mutants, and comparison of the predicted binding modes (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Jan Marienhagen − *Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, D-52425 Jülich, Germany; Institute of Biotechnology, RWTH Aachen University, D-52074 Aachen, Germany;* orcid.org/0000-0001-5513-3730; Phone: +49 2461 61 2843; Email: j.marienhagen@fz-juelich.de; Fax: +49 2461 61 2710

Authors

Lion Konstantin Flachbart − *Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, D-52425 Jülich, Germany*

Christoph Gerhard Wilhelm Gertzen − *Institute for Pharmaceutical and Medicinal Chemistry and Center for Structural Studies (CSS), Heinrich Heine University Düsseldorf, D-40225 Düsseldorf, Germany; John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC) and Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany*

Holger Gohlke − *Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf, D-40225 Düsseldorf, Germany; John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC) and Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany;* orcid.org/0000-0001-8613-1447

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.1c00165

Author Contributions

L.K.F. and J.M. conceived and designed this study. L.K.F. conducted all wet-lab experiments (mutagenesis, cultivations, FACS screening, plate-based screenings, and biosensor characterizations). C.G.W.G. performed the homology modeling, MD simulations, and molecular docking, and C.G.W.G. and H.G. evaluated the molecular modeling and simulation work. L.K.F., C.G.W.G., H.G., and J.M. wrote the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No 638718). The Center for Structural Studies is funded by the Deutsche Forschungsgemeinschaft (DFG Grant Number 417919780). We are grateful for computational support by the "Zentrum für Informations und Medientechnologie" at the Heinrich-Heine-Universität Düsseldorf and the computing time provided by the John von Neumann Institute for Computing (NIC) to HG on the supercomputer JURECA at Jülich Supercomputing Centre (JSC) (user ID: HKF7). Funding by Deutsche Forschungsgemeinschaft (DFG) (INST 208/704-1 FUGG) to purchase the hybrid computer cluster used in this study is gratefully acknowledged.

■ ABBREVIATIONS

TF, transcription factor; PP, phenylpropionic acid; CA, *trans*-cinnamic acid; 3,5DHPP, 3,5-dihydroxyphenylpropionic acid; 4HCA, *p*-coumaric acid; 5BFA, 5-bromoferulic acid; BA, benzoic acid; SA, salicylic acid; 6MSA, 6-methylsalicylic acid; 3HBA, 3-hydroxybenzoic acid; 4HBA, 4-hydroxybenzoic acid; 3,4DHBA, 3,4-dihydroxybenzoic acid; 3HCA, 3-hydroxycinnamic acid; CAF, caffeic acid; FA, ferulic acid DBD

■ REFERENCES

(1) Zeymer, C., and Hilvert, D. (2018) Directed Evolution of Protein Catalysts. *Annu. Rev. Biochem. 87* (1), 131−157.

(2) Arnold, F. H. (2018) Directed Evolution: Bringing New Chemistry to Life. *Angew. Chem., Int. Ed. 57* (16), 4143−4148.

(3) Schallmey, M., Frunzke, J., Eggeling, L., and Marienhagen, J. (2014) Looking for the Pick of the Bunch: High-Throughput Screening of Producing Microorganisms with Biosensors. *Curr. Opin. Biotechnol. 26*, 148−154.

(4) Shi, S., Choi, Y. W., Zhao, H., Tan, M. H., and Ang, E. L. (2017) Discovery and Engineering of a 1-Butanol Biosensor in Saccharomyces Cerevisiae. *Bioresour. Technol. 245* (B), 1343−1351.

(5) Abreu, V. A. C., Almeida, S., Tiwari, S., Hassan, S. S., Mariano, D., Silva, A., Baumbach, J., Azevedo, V., and Röttger, R. (2015) CMRegNet−An Interspecies Reference Database for Corynebacterial and Mycobacterial Regulatory Networks. *BMC Genomics 16* (1), 452.

(6) Novichkov, P. S., Kazakov, A. E., Ravcheev, D. A., Leyn, S. A., Kovaleva, G. Y., Sutormin, R. A., Kazanov, M. D., Riehl, W., Arkin, A. P., Dubchak, I., et al. (2013) RegPrecise 3.0 − A Resource for Genome-Scale Exploration of Transcriptional Regulation in Bacteria. *BMC Genomics 14* (1), 745.

(7) Salgado, H. (2006) RegulonDB (Version 5.0): *Escherichia coli* K-12 Transcriptional Regulatory Network, Operon Organization, and Growth Conditions. *Nucleic Acids Res. 34* (90001), D394−D397.

(8) Munch, R., Hiller, K., Barg, H., Heldt, D., Linz, S., Wingender, E., and Jahn, D. (2003) PRODORIC: Prokaryotic Database of Gene Regulation. *Nucleic Acids Res. 31* (1), 266−269.

(9) Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., Fulcher, C., Huerta, A. M., Kothari, A., Krummenacker, M., et al. (2013) EcoCyc: Fusing Model Organism Databases with Systems Biology. *Nucleic Acids Res. 41* (D1), D605−D612.

(10) Wilson, D., Charoensawan, V., Kummerfeld, S. K., and Teichmann, S. A. (2008) DBD−Taxonomically Broad Transcription Factor Predictions: New Content and Functionality. *Nucleic Acids Res. 36* (suppl_1), D88−D92.

(11) Tang, S.-Y., Fazelinia, H., and Cirino, P. C. (2008) AraC Regulatory Protein Mutants with Altered Effector Specificity. *J. Am. Chem. Soc. 130* (15), 5267−5271.

(12) Tang, S.-Y., and Cirino, P. C. (2011) Design and Application of a Mevalonate-Responsive Regulatory Protein. *Angew. Chem., Int. Ed. 50* (5), 1084−1086.

(13) Frei, C. S., Qian, S., and Cirino, P. C. (2018) New Engineered Phenolic Biosensors Based on the AraC Regulatory Protein. *Protein Eng., Des. Sel. 31* (6), 213−220.

(14) Tang, S.-Y., Qian, S., Akinterinwa, O., Frei, C. S., Gredell, J. A., and Cirino, P. C. (2013) Screening for Enhanced Triacetic Acid Lactone Production by Recombinant *Escherichia coli* Expressing a Designed Triacetic Acid Lactone Reporter. *J. Am. Chem. Soc. 135* (27), 10099−10103.

(15) Chen, W., Zhang, S., Jiang, P., Yao, J., He, Y., Chen, L., Gui, X., Dong, Z., and Tang, S. (2015) Design of an Ectoine-Responsive AraC

Mutant and Its Application in Metabolic Engineering of Ectoine Biosynthesis. *Metab. Eng. 30*, 149−155.
(16) Wang, Z., Doshi, A., Chowdhury, R., Wang, Y., Maranas, C. D., and Cirino, P. C. (2020) Engineering Sensitivity and Specificity of AraC-Based Biosensors Responsive to Triacetic Acid Lactone and Orsellinic Acid. *Protein Eng., Des. Sel. 33*, 1−9.
(17) 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., et al. (2016) Engineering an Allosteric Transcription Factor to Respond to New Ligands. *Nat. Methods 13* (2), 177−183.
(18) Della Corte, D., van Beek, H. L., Syberg, F., Schallmey, M., Tobola, F., Cormann, K. U., Schlicker, C., Baumann, P. T., Krumbach, K., Sokolowsky, S., et al. (2020) Engineering and Application of a Biosensor with Focused Ligand Specificity. *Nat. Commun. 11*, 1.
(19) Socha, R. D., and Tokuriki, N. (2013) Modulating Protein Stability - Directed Evolution Strategies for Improved Protein Function. *FEBS J. 280* (22), 5582−5595.
(20) Flachbart, L. K., Sokolowsky, S., and Marienhagen, J. (2019) Displaced by Deceivers - Prevention of Biosensor Cross-Talk Is Pivotal for Successful Biosensor-Based High-Throughput Screening Campaigns. *ACS Synth. Biol. 8* (8), 1847.
(21) Díaz, E., Ferrández, A., Prieto, M. A., and García, J. L. (2001) Biodegradation of Aromatic Compounds by *Escherichia coli*. *Microbiol. Mol. Biol. Rev. 65* (4), 523−569.
(22) Aschenbrenner, J., Marx, P., Pietruszka, J., and Marienhagen, J. (2019) Microbial Production of Natural and Non-natural Monolignols with *Escherichia coli*. *ChemBioChem 20*, 949.
(23) Siedler, S., Khatri, N. K., Zsohár, A., Kjærbølling, I., Vogt, M., Hammar, P., Nielsen, C. F., Marienhagen, J., Sommer, M. O. A., and Joensson, H. N. (2017) Development of a Bacterial Biosensor for Rapid Screening of Yeast P-Coumaric Acid Production. *ACS Synth. Biol. 6* (10), 1860−1869.
(24) Jha, R. K., Chakraborti, S., Kern, T. L., Fox, D. T., and Strauss, C. E. M. (2015) Rosetta Comparative Modeling for Library Design: Engineering Alternative Inducer Specificity in a Transcription Factor. *Proteins: Struct., Funct., Genet. 83* (7), 1327−1340.
(25) Jha, R. K., Kern, T. L., Kim, Y., Tesar, C., Jedrzejczak, R., Joachimiak, A., and Strauss, C. E. M. (2016) A Microbial Sensor for Organophosphate Hydrolysis Exploiting an Engineered Specificity Switch in a Transcription Factor. *Nucleic Acids Res. 44* (17), 8490−8500.
(26) Díaz, E., Ferrández, A., and García, J. L. (1998) Characterization of the Hca Cluster Encoding the Dioxygenolytic Pathway for Initial Catabolism of 3-Phenylpropionic Acid in *Escherichia coli* K-12. *J. Bacteriol. 180* (11), 2915−2923.
(27) Turlin, E., Perrotte-Piquemal, M., Danchin, A., and Biville, F. (2001) Regulation of the Early Steps of 3-Phenylpropionate Catabolism in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol. 3* (1), 127−133.
(28) Turlin, E., Sismeiro, O., Le Caer, J. P., Labas, V., Danchin, A., and Biville, F. (2005) 3-Phenylpropionate Catabolism and the *Escherichia coli* Oxidative Stress Response. *Res. Microbiol. 156* (3), 312−321.
(29) Mulnaes, D., and Gohlke, H. (2018) TopScore: Using Deep Neural Networks and Large Diverse Data Sets for Accurate Protein Model Quality Assessment. *J. Chem. Theory Comput. 14* (11), 6117−6126.
(30) Widderich, N., Pittelkow, M., Höppner, A., Mulnaes, D., Buckel, W., Gohlke, H., Smits, S. H. J., and Bremer, E. (2014) Molecular Dynamics Simulations and Structure-Guided Mutagenesis Provide Insight into the Architecture of the Catalytic Core of the Ectoine Hydroxylase. *J. Mol. Biol. 426* (3), 586−600.
(31) Marchler-Bauer, A., and Bryant, S. H. (2004) CD-Search: Protein Domain Annotations on the Fly. *Nucleic Acids Res. 32* (Web Server), W327−W331.
(32) Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., et al. (2011) CDD: A Conserved Domain Database for the Functional Annotation of Proteins. *Nucleic Acids Res. 39* (Database), D225−D229.
(33) Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., et al. (2015) CDD: NCBI's Conserved Domain Database. *Nucleic Acids Res. 43* (D1), D222−D226.
(34) Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2017) CDD/SPARCLE: Functional Classification of Proteins *via* Subfamily Domain Architectures. *Nucleic Acids Res. 45* (D1), D200−D203.
(35) Sigrist, C. J. A., De Castro, E., Langendijk-Genevaux, P. S., Le Saux, V., Bairoch, A., and Hulo, N. (2005) ProRule: A New Database Containing Functional and Structural Information on PROSITE Profiles. *Bioinformatics 21* (21), 4060−4066.
(36) Sigrist, C. J. A., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A., and Bucher, P. (2002) PROSITE: A Documented Database Using Patterns and Profiles as Motif Descriptors. *Briefings Bioinf. 3* (3), 265−274.
(37) Sigrist, C. J. A., De Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2012) New and Continuing Developments at PROSITE. *Nucleic Acids Res. 41* (Database), D344−D347.
(38) Craven, S. H., Ezezika, O. C., Haddad, S., Hall, R. A., Momany, C., and Neidle, E. L. (2009) Inducer Responses of BenM, a LysR-Type Transcriptional Regulator from Acinetobacter Baylyi ADP1. *Mol. Microbiol. 72* (4), 881−894.
(39) Mundhada, H., Marienhagen, J., Scacioc, A., Schenk, A., Roccatano, D., and Schwaneberg, U. (2011) SeSaM-Tv-II Generates a Protein Sequence Space That Is Unobtainable by EpPCR. *ChemBioChem 12* (10), 1595−1601.
(40) Fujii, I., Ono, Y., Tada, H., Gomi, K., Ebizuka, Y., and Sankawa, U. (1996) Cloning of the Polyketide Synthase Gene AtX from Aspergillus Terreus and Its Identification as the 6-Methylsalicylic Acid Synthase Gene by Heterologous Expression. *Mol. Gen. Genet. 253* (1−2), 1−10.
(41) Dimroth, P., Walter, H., and Lynen, F. (1970) [Biosynthesis of 6-Methylsalicylic Acid]. *Eur. J. Biochem. 13* (1), 98−110.
(42) Kealey, J. T., Liu, L., Santi, D. V., Betlach, M. C., and Barr, P. J. (1998) Production of a Polyketide Natural Product in Non-polyketide-Producing Prokaryotic and Eukaryotic Hosts. *Proc. Natl. Acad. Sci. U. S. A. 95* (2), 505−509.
(43) Cox, R. J., Skellam, E., and Williams, K. (2018) Biosynthesis of Fungal Polyketides. In *Physiology and Genetics*, pp 385−412, Springer International Publishing, Cham. DOI: 10.1007/978-3-319-71740-1_13.
(44) Kallscheuer, N., and Marienhagen, J. (2018) Corynebacterium Glutamicum as Platform for the Production of Hydroxybenzoic Acids. *Microb. Cell Fact. 17* (1), 70.
(45) Sotriffer, C. A., Gohlke, H., and Klebe, G. (2002) Docking into Knowledge-Based Potential Fields: A Comparative Evaluation of DrugScore. *J. Med. Chem. 45* (10), 1967−1970.
(46) Dittrich, J., Schmidt, D., Pfleger, C., and Gohlke, H. (2019) Converging a Knowledge-Based Scoring Function: DrugScore2018. *J. Chem. Inf. Model. 59* (1), 509−521.
(47) Goodsell, D. S., Morris, G. M., and Olson, A. J. (1996) Automated Docking of Flexible Ligands: Applications of Autodock. *J. Mol. Recognit. 9* (1), 1−5.
(48) Bertani, G. (1951) Studies on Lysogenesis. I. The Mode of Phage Liberation by Lysogenic *Escherichia coli*. *J. Bacteriol. 62* (3), 293−300.
(49) Kensy, F., Zang, E., Faulhammer, C., Tan, R.-K., and Büchs, J. (2009) Validation of a High-Throughput Fermentation System Based on Online Monitoring of Biomass and Fluorescence in Continuously Shaken Microtiter Plates. *Microb. Cell Fact. 8*, 31.
(50) Funke, M., Diederichs, S., Kensy, F., Müller, C., and Büchs, J. (2009) The Baffled Microtiter Plate: Increased Oxygen Transfer and Improved Online Monitoring in Small Scale Fermentations. *Biotechnol. Bioeng. 103* (6), 1118−1128.

(51) Sambrook, J., and Russel, D. W. (2001) *Molecular Cloning A Lab Manual*, Cold Spring Harbor Laboratory Press, New York.
(52) Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods 6* (5), 343−345.
(53) Patrick, W. M., Firth, A. E., and Blackburn, J. M. (2003) User-Friendly Algorithms for Estimating Completeness and Diversity in Randomized Protein-Encoding Libraries. *Protein Eng., Des. Sel. 16* (6), 451−457.
(54) Firth, A. E., and Patrick, W. M. (2008) GLUE-IT and PEDEL-AA: New Programmes for Analyzing Protein Diversity in Randomized Libraries. *Nucleic Acids Res. 36*, W281.
(55) Firth, A. E., and Patrick, W. M. (2005) Statistics of Protein Library Construction. *Bioinformatics 21*, 3314−3315.
(56) Freiherr von Boeselager, R., Pfeifer, E., and Frunzke, J. (2018) Cytometry Meets Next-Generation Sequencing − RNA-Seq of Sorted Subpopulations Reveals Regional Replication and Iron-Triggered Prophage Induction in Corynebacterium Glutamicum. *Sci. Rep. 8* (1), 14856.
(57) *Maestro Release* (2017) Schrödinger, LLC., New York, NY.
(58) DeLano, W. L. (2002) Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsl. Protein Crystallogr. 40* (1), 82−92.
(59) DeLano, W. L. (2005) The Case for Open-Source Software in Drug Discovery. *Drug Discovery Today 10* (3), 213−217.
(60) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys. 79* (2), 926−935.
(61) Bas, D. C., Rogers, D. M., and Jensen, J. H. (2008) Very Fast Prediction and Rationalization of PKa Values for Protein-Ligand Complexes. *Proteins: Struct., Funct., Genet. 73* (3), 765−783.
(62) Case, D. A., Berryman, J. T., Betz, R. M., Cerutti, D. S., Cheatham, T. E., III, Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., and Goetz, A. W., et al. (2015) *AMBER 2015*, Univ. California, San Francisco.
(63) Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput. 11* (8), 3696−3713.
(64) Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald: An *N*·log(*N*) Method for Ewald Sums in Large Systems. *J. Chem. Phys. 98* (12), 10089−10092.
(65) Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. (1977) Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys. 23* (3), 327−341.
(66) Hopkins, C. W., Le Grand, S., Walker, R. C., and Roitberg, A. E. (2015) Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput. 11* (4), 1864−1874.
(67) Roe, D. R., and Cheatham, T. E. (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput. 9* (7), 3084−3095.

https://doi.org/10.1021/acssynbio.1c00165
ACS Synth. Biol. 2021, 10, 2002−2014