CHARACTERIZING AND EXPLORING THE DYNAMIC PROPERTY OF A NOVEL TRANSCRIPTIONAL FACTOR FROM *BACILLUS AMYLOLIQUEFACIENS*

by

YUYANG ZHOU

(Under the Direction of Yajun Yan)

ABSTRACT

Transcriptional factor-based biosensors have been widely utilized for inducible control of gene expression in metabolic engineering and relative applications in pathway regulation and high-throughput screening. Two major ways to expand the sensor pool are engineering existing biosensors and exploring unknown transcriptional factors. In this study, we characterized and engineered a phenolic acid-responsive regulator PadR from *Bacillus Amyloliquefaciens* (BaPadR). The BaPadR sensor system possessed a unique inducer preference and showed a high output sstrength compared with other commonly used inducible biosensors. The DNA binding region of BaPadR was engineered to enhance the dynamic range of the biosensor system. By promoter truncation and the construction of hybrid promoters, the DNA binding sequence of BaPadR was identified. To further explore the tunability of the sensor system, base substitutions were performed on the BaPadR binding region of the P_{padC} promoter. This novel biosensor system can serve as a valuable tool in future synthetic biology applications.

INDEX WORDS: Transcriptional factors, PadR, engineering of biosensor

CHARACTERIZING AND EXPLORING THE DYNAMIC PROPERTY OF A NOVEL TRANSCRIPTIONAL FACTOR FROM *BACILLUS AMYLOLIQUEFACIENS*

by

YUYANG ZHOU

B.S., Beijing University of Chemical Technology, China, 2021

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2023

CHARACTERIZING AND EXPLORING THE DYNAMIC PROPERTY OF A NOVEL TRANSCRIPTIONAL FACTOR FROM *BACILLUS AMYLOLIQUEFACIENS*

by

YUYANG ZHOU

Major Professor:	Yajun Yan
Committee:	James N. Warnock
	William Kisaalita

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2023

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest appreciation to my instructor, Professor Yajun Yan, who continually conveyed a spirit of adventure throughout the research and excitement in regard to guiding me. Without your confidence and guidance on me, I cannot finish my master's research successfully.

Then, I would like to thank you for the assistance and support from all of my lab mates, especially my lab supervisor, Dr. Chenyi Li. Thank you for your patient guidance in my experiments and your care in my life. In addition, I would also like to say thanks to Xinyu Gong, Jianli Zhang, Tian Jiang, Yusong Zou, Qi Gan, and Yuxi Teng for their advice.

I really appreciate my committee, Dr. James N. Warnock, and Dr. William Kisaalita, who enlarge my reservoir of specialized knowledge. Your help is indispensable for my master's research.

In the end, I would like to say thank you to my mother and father. They raise me up and always have my back. Your deep love and trust in me are solid support for me on my way to pursuing my dream.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Metabolic engineering aims to use microorganisms to produce valuable chemicals and pharmaceuticals, such as 1-butanol*(1)*, 1-propanol*(1, 2)*, pyrogallol*(3)*, artemisinic acid*(4)*, etc. To generate chemicals and pharmaceuticals, researchers designed and introduced efficient biosynthetic pathways into the host cell and regulated its metabolism to improve the final yield, titer, or productivity*(5)*. Static pathway regulation is a common strategy used to redirect flux toward desired products by overexpressing genes within the pathway and disrupting genes from competitive pathways*(6-8)*. Although static regulation can increase the final titer, this approach still has drawbacks, such as bringing metabolic imbalance, generating toxic intermediates and finally reducing host cell fitness due to the conflict between cell growth and product generation*(5, 7, 9)*. Dynamic pathway regulation has been proven as a successful approach to achieving high productivity by sensing signals from the environment or intracellular and extracellular metabolites and maintaining the balance between cell growth and production*(10, 11)*.

Genetically-encoded biosensors play an important role in fine-tuning the metabolic pathway in dynamic regulation. For example, a dual global regulator factor Cra (catabolite repressor/activator) has been developed and applied to dynamically control the carbon flux distribution between glycolysis and the production of N-acetylglucosamine and mevalonate without generating glycolysis overflow*(12)*. Further, a photosensitive *lac* operon (OptoLAC) was developed in *E. coli* and further applied to improve the mevalonate and isobutanol

production*(13)*. Researchers also created a tunable growth phase-dependent autonomous bifunctional genetic switch to decouple the cell growth and γ -aminobutyric acid production from glycerol in *Corynebacterium glutamicum(14)*. The T-switch, a temperature-sensitive bifunctional bioswitch, yielded an ordered assembly of poly(3-hydroxybutyrate)-block-poly(4-hydroxybutyrate)*(15)*. Genetically encoded biosensors can improve production by screening high-yield strains and high-activity enzyme mutants. For instance, utilizing a Leu3p-based biosensor screens high-activity α-isopropylmalate synthases mutants and yeast strains with high productivity of isobutanol and isopentanol*(16)*.

Transcriptional factors (TFs) are a protein that controls the transcription process in the microorganism. TFs include transcriptional repressors and transcriptional activators. Transcriptional repressor will identify specific DNA sequence and occupy the -35 and -10 regions, thereby preventing RNA polymerase from accessing the promoter. With the existence of the inducer, the inducer will bind with the repressor and lead to a conformational change of the repressor, making the repressor lose the ability to bind the DNA sequence and then release the expression inhibition*(17)*. For the transcriptional activator, the activator binds to DNA. When the inducer is present, the activator binds to the inducer and changes its conformation, thus the regulator protein interacts and aggregates the RNA polymerase, and the expression of the downstream gene is activated*(17)*. Based on the mechanism, transcriptional factor-based biosensors are developed and widely applied in metabolic engineering and synthetic biology for dynamic pathway regulation, high-throughput screening, and real-time monitoring of cellular status due to the abundant presence of TFs in nature, as well as their ease of handling and modification*(18-20)*. However, the main complication is the low responsive activity of natural transcriptional factors. Since present biosensors possess limited substrate scope, engineering

existing biosensors and identifying unknown transcriptional factors are the focus of much research.

Phenolic acid responsive transcriptional regulator (PadR) was discovered in *Bacillus subtilis* 168 (BsPadR) in 2011*(21)*, while the BsPadR structure, DNA binding sequence, and regulator-DNA binding model were characterized in 2017*(22)*. PadR can inhibit the promoter P_{padC} and repress the expression of the phenolic acid decarboxylase (padC) gene. With the existence of phenolic acid in the environment, the phenolic acid induces a conformational change in the PadR when phenolic acid binds to the repressor and releases the inhibition on *padC* transcription activating the phenolic acid decarboxylation. Due to the capacity of sensing phenolic acids, such as *p*-coumaric acid and salicylic acid, which are the precursors of many valuable flavonoids and coumarins*(23-25)*, the PadR-based sensor system has been extensively studied, developed, and applied in metabolic engineering and synthetic biology*(26-28)*. Our previous study further increased the sensitivity of PadR, broader dynamic ranges, and expanded operational ranges *(26)*. Dynamic range is the ratio between the largest and smallest output of the sensor system. The operational range means the minimum and maximum input ligand limits within which a biosensor system will operate to specifications. The variants from this study were further applied to dynamically control the production of *p*-coumaric acid*(28)* and naringenin*(29)*.

With the rapid development of bioinformatic tools and fast-sequencing techniques, databases with vast quantities of genomic sequences have been established, which provides us with a huge reservoir of putative TFs waiting to be discovered. BLAST (Basic Local Alignment Search Tool) is a powerful tool for discovering novel TFs by comparing the protein and DNA sequences of known and well-characterized TFs with the sequence databases and searching similar sequences in the reservoir*(30)*. In this study, we discovered a novel phenolic acid-

responsive regulator PadR from *Bacillus Amyloliquefaciens* (BaPadR) throughout the protein sequence BLAST. Based on the previous study on BsPadR*(21, 22, 26)*, we were able to identify promoter BaP_{padC} controlled by BaPadR and the DNA binding sequence in the promoter that was recognized by BaPadR. Further characterization of the BaPadR regulator revealed a unique ligand profile. To improve its usability as a biosensor system in metabolic engineering and synthetic biology, we expanded the dynamic ranges and increased its sensitivity through site-directed mutagenesis of the BaPadR regulator and the base alteration on the BaPadR binding box. The BaPadR-based biosensor developed in this study expands the current repertoire of the small molecule-sensing transcriptional factors and can be a useful addition to the biosensor toolbox.

CHAPTER 2

METHOD AND MATERIALS

2.1 Strains, plasmid, and reagents

All strains and plasmids used in this study were listed in **Table 1**. *E. coli* strain XLI-Blue was used for plasmid construction. *E. coli* strain Bw25113(F') was used for biosensor characterization and fluorescence assay.

Lysogeny broth (LB) containing 10 g/L NaCl, 5 g/L yeast extract, and 10 g/L tryptone medium was utilized for E. coli culture. Ampicillin and kanamycin were supplemented in the medium as needed with the final concentration of 100, and 50 mg/mL respectively. Different concentration of IPTG was added into the medium if needed. For the inducer preparation, 100 mg of *p*-coumaric acid, caffeic acid, trans-cinnamic acid, ferulic acid, 4-hydeoxybenzoic acid, anthranilic acid, salicylic acid, and vanillic acid were dissolved in 1mL methanol to make the master stock with a concentration of 100 g/L. *p*-coumaric acid and ferulic acid were purchased from MP Biomedicals; caffeic acid and anthranilic acid were purchased from SIGMA; trans-cinnamic acid was purchased from Aldaich; salicylic acid and vanillic acid were purchased from Alfa Aesar.

2.2 DNA Manipulation.

All genetic sequences and components were listed in **Table 2**. High-copy plasmid pHA-egfp-MCS was constructed in this study by inserting the DNA sequence of egfp into the MCS of pHA-MCS*(29)* constructed in the previous study.

BaPadR and PadC promoter were amplified from the genome DNA of *Bacillus amyloliquefaciens* (ATCC23350). The promoter BaP_{padC} was flanked by XhoI and KpnI and cloned into pHA-egfp-MCS to construct pHA-BaPpadC-WT-egfp. A strong engineered ribosome-binding site (RBS, with a sequence of AAAGAGGAGAAA) and an EcoRI site were added to the BaP_{padC} promoter. The new promoter, BaP_{padC}-RBS, flanked by XhoI and EcoRI was cloned into plasmid pHA-egfp-MCS to construct pHA- BaP_{padC}-RBS-egfp. The BaPadR was constructed on medium-copy plasmid pCS27*(1)* by using KpnI and SalI, resulting in the plasmid pCS-BaPadR.

Site-directed mutagenesis on residue H38 and S39 of the BaPadR was amplified by overlap-extension PCR. The DNA fragments of the BaPadR variants, (BaPadR-H38A, BaPadR-H38S, BaPadR-H38D, BaPadR-H38N, BaPadR-H38E, BaPadR-H38I and BaPadR-S39A), were integrated into plasmid pCS27*(1)* by KpnI and SalI, generating plasmids pCS-BaPadR-H38A, pCS-BaPadR-H38S, pCS-BaPadR-H38D, pCS-BaPadR-H38N, pCS-BaPadR-H38E, pCS-BaPadR-H38I and pCS-BaPadR-S39A respectively.

pHA-BaP_{padC}-RBS-egfp was used as a template to amplify the truncated promoters BaP_{padC}-T1, BaP_{padC}-T2, BaP_{padC}-T3, and BaP_{padC}-T4. Those truncated promoter fragments were flanked by XhoI and EcoRI and were integrated into pHA-egfp-MCS to construct pHA-BaPpadC-T1-egfp, pHA-BaPpadC-T2-egfp, pHA-BaPpadC-T3-egfp, and pHA-BaPpadC-T4-egfp, respectively. The DNA fragments of hybrid promoter Phy11-egfp, Phy12-egfp, Phy21-egfp, Phy22-egfp, Phy35-egfp, and Phy10-egfp were amplified by using the pHA-egfp-MCS as the template. All hybrid promoter DNA fragments were integrated into plasmid pHA-MCS by XhoI and AvrII forming pHA-Phy11-egfp, pHA-Phy12-egfp, pHA-Phy21-egfp, pHA-Phy22-egfp, pHA-Phy35-egfp, and pHA-Phy10-egfp, respectively.

Base-replacement of the BaP_{padC}-RBS promoters (BaP_{padC} -C6A-RBS, BaP_{padC} -C6T-RBS, BaP_{padC} -C6G-RBS, BaP_{padC} -T8A-RBS, BaP_{padC} -T8C-RBS, BaP_{padC} -T8G-RBS, BaP_{padC} -T18'A-RBS, BaP_{padC}- -T18'C RBS, and BaP_{padC} -T18'G-RBS) were amplified by using pHA-BaP_{padC}-RBS-egfp as the template, and were constructed by SLIM strategy*(31)* forming pHA-BaP_{padC} -C6A-RBS -egfp, pHA-BaP_{padC} -C6T-RBS -egfp, pHA-BaP_{padC} -C6G-RBS -egfp, pHA-BaP_{padC} -T8A-RBS-egfp, pHA-BaP_{padC} -T8C-RBS-egfp, pHA-BaP_{padC} -T8G-RBS-egfp, pHA-BaP_{padC} -T18'A-RBS-egfp, pHA-BaP_{padC} -T18'C-RBS-egfp, and pHA-BaP_{padC} -T18'G-RBS-egfp respectively.

2.3 Dynamic preformation characterization

After transforming the BaPadR-based biosensor system into BW25113/F', all transformants were cultured under 37 ℃ overnight. Three single colonies were randomly picked and inoculated into 3.5 mL LB medium with specific antibiotics, cultivating for 10 h at 37 ℃, 270 rpm. Then, 150 μL cultures were served as seeds and were transferred into 3.5 mL fresh LB medium containing appropriate antibiotics and specific concentration of IPTG. Gradient concentrations of inducers were added into the medium after 1 h of culture under 37 ℃, 270 rpm. Samples were collected after 12 h of cultivation after adding ligands and were diluted for the measurement of the cell densities (OD_{600}) and green fluorescence intensities.

2.4 Fluorescence Assay

40 μL of cell cultures were transformed into a black 96-well plate (Corning® 96-well Flat Clear Bottom Black Polystyrene TC-treated Microplates, Corning 3603) and diluted with 160μL of water. The Synergy HT reader from BioTek was used for the fluorescence assay. The

96-well plate carried with samples was scanned by Synergy HT reader. The egfp fluorescent intensity was detected by using an excitation filter of 485/20 nm and an emission filter of 528/20 nm. The OD600 of cell cultures was detected by 96-well plate reader. The egfp expression levels were represented by normalizing the fluorescence intensities with their corresponding cell densities OD600. The normalized fluorescence was calculated as the equation below:

$$\frac{\text{RFU}}{\text{OD}_{600}} = \frac{\text{Fluorescence} - \text{background}}{(\text{Cell density} - \text{background}) * 1.76}$$

The inhibition efficiency was defined as the equation below:

$$\text{Inhibition effiency} = \left[\frac{\text{normalized fluoresence without BaPadR} - \text{normalized fluoresence with BaPadR}}{\text{normalized fluoresence without BaPadR}}\right] \times 100\%$$

The P value was also calculated by two independent samples t-test. A P value of <.05 was considered statistically significant.

Table 1. Strains and Plasmids were used in this study.

Name	Genotype	Reference
E. coli XLI-Blue	*recA1 endA1gyrA96thi-1hsdR17supE44relA1lac [F' proAB lacIqZDM15Tn10 (TetR)]*	Stratagene
E. coli BW25113 F’	*rrnBT14 ΔlacZWJ16 hsdR514 ΔaraBADAH33 ΔrhaBADLD78 F' [traD36 proAB lacIqZΔM15 Tn10(Tetr)]*	*(1)*

Plasmid	Description	Source
pCS27	pL_{lacO1}; *p15A ori*; Kan^R	*(1)*
pHA-MCS	pL_{lacO1}; *ColE1 ori*, Amp^R	*(29)*
pHA-egfp-MCS	pHA-MCS containing pLlacO1-controlled *egfp* expression cassette	This study
pHA-BaP_{padC}-WT-egfp	pHA-egfp-MCS harboring the wild type PadC promoter (BaP_{padC}) and RBS region from *Bacillus amyloliquefaciens*	This study
pHA-BaP_{padC}-RBS-egfp	pHA-egfp-MCS harboring the wild type PadC promoter from *Bacillus amyloliquefaciens* with the strong engineered RBS inserted between the BaPpadC promoter and the egfp gene	This study
pCS-BaPadR	pCS27 with a pL_{lacO1}-controlled wild-type BaPadR expression cassette	This study
pCS-BaPadR-H38A	pCS27 carrying BaPadR mutant H38A	This study
pCS-BaPadR-H38S	pCS27 carrying PadR mutant H38S	This study
pCS-BaPadR-H38D	pCS27 carrying PadR mutant H38D	This study

pCS-BaPadR-H38N	pCS27 carrying PadR mutant H38N	This study
pCS-BaPadR-H38E	pCS27 carrying PadR mutant H38E	This study
pCS-BaPadR-H38I	pCS27 carrying PadR mutant H38I	This study
pCS-BaPadR-S39A	pCS27 carrying PadR mutant S39A	This study
pHA-BaP_{padC}-T1-egfp	pHA-egfp-MCS carrying the truncated BaP_{padC} promoter T1 and the strong RBS region	This study
pHA-BaP_{padC}-T2-egfp	pHA-egfp-MCS carrying the truncated BaP_{padC} promoter T2 and the strong RBS region	This study
pHA-BaP_{padC}-T3-egfp	pHA-egfp-MCS carrying the truncated BaP_{padC} promoter T3 and the strong RBS region	This study
pHA-BaP_{padC}-T4-egfp	pHA-egfp-MCS carrying the truncated BaP_{padC} promoter T4 and the strong RBS region	This study
pHA- Phy11-egfp	pHA-egfp-MCS harboring the hybrid promoter Phy11	This study
pHA- Phy12-egfp	pHA-egfp-MCS harboring the hybrid promoter Phy12	This study
pHA- Phy21-egfp	pHA-egfp-MCS harboring the hybrid promoter Phy21	This study
pHA- Phy22-egfp	pHA-egfp-MCS harboring the hybrid promoter Phy22	This study
pHA- Phy10-egfp	pHA-egfp-MCS harboring the hybrid promoter Phy10	This study
pHA- Phy35-egfp	pHA-egfp-MCS harboring the hybrid promoter Phy35	This study

pHA-BaPpadC-C6A-RBS-egfp	pHA-egfp-MCS harboring the BaP_{padC}-C6A promoter variant with the strong engineered RBS inserted between the BaPpadC promoter and the egfp gene	This study
pHA-BaPpadC-C6T-RBS-egfp	pHA-egfp-MCS harboring the BaP_{padC}-C6T promoter variant with the strong engineered RBS inserted between the BaPpadC promoter and the egfp gene	This study
pHA-BaPpadC-C6G-RBS-egfp	pHA-egfp-MCS harboring the BaP_{padC}-C6G promoter variant with the strong engineered RBS inserted between the BaPpadC promoter and the egfp gene	This study
pHA-BaPpadC-T8A-RBS-egfp	pHA-egfp-MCS harboring the BaP_{padC}-T8A promoter variant with the strong engineered RBS inserted between the BaPpadC promoter and the egfp gene	This study
pHA-BaPpadC-T8C-RBS-egfp	pHA-egfp-MCS harboring the BaP_{padC}-T8C promoter variant with the strong engineered RBS inserted between the BaPpadC promoter and the egfp gene	This study
pHA-BaPpadC-T8G-RBS-egfp	pHA-egfp-MCS harboring the BaPpadC-T8G promoter variant with the strong engineered RBS inserted between the BaPpadC promoter and the egfp gene	This study
pHA-BaPpadC-T18'A-RBS-egfp	pHA-egfp-MCS harboring the BaP_{padC}-T18'A promoter variant with the strong engineered RBS inserted between the BaPpadC promoter and the egfp gene	This study

pHA-BaPpadC-T18'C-RBS-egfp	pHA-egfp-MCS harboring the BaPpadC-T18'C promoter variant with the strong engineered RBS inserted between the BaPpadC promoter and the egfp gene	This study
pHA-BaPpadC-T18'G-RBS-egfp	pHA-egfp-MCS harboring the BaPpadC-T18'G promoter variant with the strong engineered RBS inserted between the BaPpadC promoter and the egfp gene	This study

Table 2. The sequence of Genes and Promoters in This Study

Name	DNA sequence (5'-3') [a]
BaP_{padC}-WT	GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAA TGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACCTTTTTAC CGGGTGAAGCATCTCATCATTTGACAGTCCATTTTAACAGCGTTACAATTAATCATGTAA ATAGTTACATGTATATATAAACATAATGTCTTGGGAGGTGATACGTCCTGAATGAACCTT CTTTAAAAAAAGTG
BaP_{padC}-RBS [b]	GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAA TGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACCTTTTTAC CGGGTGAAGCATCTCATCATTTGACAGTCCATTTTAACAGCGTTACAATTAATCATGTAA ATAGTTACATGTATATATAAACATAATGTCTTGGGAGGTGATACGTCCTGAATGAACCTT CTTTAAAAAAAGTGGAATTCATT**AAAGAGGAGAAA**
BaPadR	ATGAGAATTTTAAAGTACGCGATTTTAGGACTTTTGCGAAAAGGCGAATTGAGCGGATA TGATATATCGAGCTATTTTAAAGAAGAGCTAGGCCAGTTTTGGAGCGCAAAGCACAGCC AGATTTATCCGGAATTAAAAAAGCTGACGGCTGAGGGATTCATTACGTTCCGCACTGCG ATTCAGGGAACGAAGCTGGAGAAAAAAATGTACACGCTGACTGACAATGGAGAGCGGG AGCTTTGTGCATGGCTGACGAAAAAAGATCCGATTCCGGAAACGGTGAAGGATGAATTT

	ATGCTGAAGGCTTATTTTATCTCAGCTTTGACGAATGAAGAAGCGGATGAGCTATTCACC GATCAGCTCGTAAAGCGAAAGGAGAAGTTGTCCGATCTGGAAAACAGTTATCATGAACT GATGACATCCTCCGAGGAGGCGGATTCCTTTTCTTCTCCGGATTTCGGCCATTATCTCGT GCTGACAAAAGCGCTGGAGCGGGAAAGGAATTATATTTCCTGGCTTGAGCATATTTTGG CTCTCATCAAAAAAGCATAA
BaPadR-H38A	ATGAGAATTTTAAAGTACGCGATTTTAGGACTTTTGCGAAAAGGCGAATTGAGCGGATA TGATATATCGAGCTATTTTAAAGAAGAGCTAGGCCAGTTTTGGAGCGCAAAGGCGAGCC AGATTTATCCGGAATTAAAAAAGCTGACGGCTGAGGGATTCATTACGTTCCGCACTGCG ATTCAGGGAACGAAGCTGGAGAAAAAAATGTACACGCTGACTGACAATGGAGAGCGGG AGCTTTGTGCATGGCTGACGAAAAAAGATCCGATTCCGGAAACGGTGAAGGATGAATTT ATGCTGAAGGCTTATTTTATCTCAGCTTTGACGAATGAAGAAGCGGATGAGCTATTCACC GATCAGCTCGTAAAGCGAAAGGAGAAGTTGTCCGATCTGGAAAACAGTTATCATGAACT GATGACATCCTCCGAGGAGGCGGATTCCTTTTCTTCTCCGGATTTCGGCCATTATCTCGT GCTGACAAAAGCGCTGGAGCGGGAAAGGAATTATATTTCCTGGCTTGAGCATATTTTGG CTCTCATCAAAAAAGCATAA
BaPadR-H38S	ATGAGAATTTTAAAGTACGCGATTTTAGGACTTTTGCGAAAAGGCGAATTGAGCGGATA TGATATATCGAGCTATTTTAAAGAAGAGCTAGGCCAGTTTTGGAGCGCAAAGAGCAGCC AGATTTATCCGGAATTAAAAAAGCTGACGGCTGAGGGATTCATTACGTTCCGCACTGCG ATTCAGGGAACGAAGCTGGAGAAAAAAATGTACACGCTGACTGACAATGGAGAGCGGG AGCTTTGTGCATGGCTGACGAAAAAAGATCCGATTCCGGAAACGGTGAAGGATGAATTT ATGCTGAAGGCTTATTTTATCTCAGCTTTGACGAATGAAGAAGCGGATGAGCTATTCACC GATCAGCTCGTAAAGCGAAAGGAGAAGTTGTCCGATCTGGAAAACAGTTATCATGAACT GATGACATCCTCCGAGGAGGCGGATTCCTTTTCTTCTCCGGATTTCGGCCATTATCTCGT GCTGACAAAAGCGCTGGAGCGGGAAAGGAATTATATTTCCTGGCTTGAGCATATTTTGG CTCTCATCAAAAAAGCATAA
BaPadR-H38D	ATGAGAATTTTAAAGTACGCGATTTTAGGACTTTTGCGAAAAGGCGAATTGAGCGGATA TGATATATCGAGCTATTTTAAAGAAGAGCTAGGCCAGTTTTGGAGCGCAAAGGATAGCC AGATTTATCCGGAATTAAAAAAGCTGACGGCTGAGGGATTCATTACGTTCCGCACTGCG ATTCAGGGAACGAAGCTGGAGAAAAAAATGTACACGCTGACTGACAATGGAGAGCGGG AGCTTTGTGCATGGCTGACGAAAAAAGATCCGATTCCGGAAACGGTGAAGGATGAATTT ATGCTGAAGGCTTATTTTATCTCAGCTTTGACGAATGAAGAAGCGGATGAGCTATTCACC

	GATCAGCTCGTAAAGCGAAAGGAGAAGTTGTCCGATCTGGAAAACAGTTATCATGAACT
	GATGACATCCTCCGAGGAGGCGGATTCCTTTTCTTCTCCGGATTTCGGCCATTATCTCGT
	GCTGACAAAAGCGCTGGAGCGGGAAAGGAATTATATTTCCTGGCTTGAGCATATTTTGG
	CTCTCATCAAAAAAGCATAA
BaPadR-H38N	ATGAGAATTTTAAAGTACGCGATTTTAGGACTTTTGCGAAAAGGCGAATTGAGCGGATA
	TGATATATCGAGCTATTTTAAAGAAGAGCTAGGCCAGTTTTGGAGCGCAAAGAACAGCC
	AGATTTATCCGGAATTAAAAAAGCTGACGGCTGAGGGATTCATTACGTTCCGCACTGCG
	ATTCAGGGAACGAAGCTGGAGAAAAAAATGTACACGCTGACTGACAATGGAGAGCGGG
	AGCTTTGTGCATGGCTGACGAAAAAAGATCCGATTCCGGAAACGGTGAAGGATGAATTT
	ATGCTGAAGGCTTATTTTATCTCAGCTTTGACGAATGAAGAAGCGGATGAGCTATTCACC
	GATCAGCTCGTAAAGCGAAAGGAGAAGTTGTCCGATCTGGAAAACAGTTATCATGAACT
	GATGACATCCTCCGAGGAGGCGGATTCCTTTTCTTCTCCGGATTTCGGCCATTATCTCGT
	GCTGACAAAAGCGCTGGAGCGGGAAAGGAATTATATTTCCTGGCTTGAGCATATTTTGG
	CTCTCATCAAAAAAGCATAA
BaPadR-H38E	ATGAGAATTTTAAAGTACGCGATTTTAGGACTTTTGCGAAAAGGCGAATTGAGCGGATA
	TGATATATCGAGCTATTTTAAAGAAGAGCTAGGCCAGTTTTGGAGCGCAAAGGAAAGCC
	AGATTTATCCGGAATTAAAAAAGCTGACGGCTGAGGGATTCATTACGTTCCGCACTGCG
	ATTCAGGGAACGAAGCTGGAGAAAAAAATGTACACGCTGACTGACAATGGAGAGCGGG
	AGCTTTGTGCATGGCTGACGAAAAAAGATCCGATTCCGGAAACGGTGAAGGATGAATTT
	ATGCTGAAGGCTTATTTTATCTCAGCTTTGACGAATGAAGAAGCGGATGAGCTATTCACC
	GATCAGCTCGTAAAGCGAAAGGAGAAGTTGTCCGATCTGGAAAACAGTTATCATGAACT
	GATGACATCCTCCGAGGAGGCGGATTCCTTTTCTTCTCCGGATTTCGGCCATTATCTCGT
	GCTGACAAAAGCGCTGGAGCGGGAAAGGAATTATATTTCCTGGCTTGAGCATATTTTGG
	CTCTCATCAAAAAAGCATAA
BaPadR-H38I	ATGAGAATTTTAAAGTACGCGATTTTAGGACTTTTGCGAAAAGGCGAATTGAGCGGATA
	TGATATATCGAGCTATTTTAAAGAAGAGCTAGGCCAGTTTTGGAGCGCAAAGATTAGCC
	AGATTTATCCGGAATTAAAAAAGCTGACGGCTGAGGGATTCATTACGTTCCGCACTGCG
	ATTCAGGGAACGAAGCTGGAGAAAAAAATGTACACGCTGACTGACAATGGAGAGCGGG
	AGCTTTGTGCATGGCTGACGAAAAAAGATCCGATTCCGGAAACGGTGAAGGATGAATTT
	ATGCTGAAGGCTTATTTTATCTCAGCTTTGACGAATGAAGAAGCGGATGAGCTATTCACC
	GATCAGCTCGTAAAGCGAAAGGAGAAGTTGTCCGATCTGGAAAACAGTTATCATGAACT

	GATGACATCCTCCGAGGAGGCGGATTCCTTTTCTTCTCCGGATTTCGGCCATTATCTCGT GCTGACAAAAGCGCTGGAGCGGGAAAGGAATTATATTTCCTGGCTTGAGCATATTTTGG CTCTCATCAAAAAAGCATAA
BaPadR-S39A	ATGAGAATTTTAAAGTACGCGATTTTAGGACTTTTGCGAAAAGGCGAATTGAGCGGATA TGATATATCGAGCTATTTTAAAGAAGAGCTAGGCCAGTTTTGGAGCGCAAAGCACGCGC AGATTTATCCGGAATTAAAAAAGCTGACGGCTGAGGGATTCATTACGTTCCGCACTGCG ATTCAGGGAACGAAGCTGGAGAAAAAAATGTACACGCTGACTGACAATGGAGAGCGGG AGCTTTGTGCATGGCTGACGAAAAAAGATCCGATTCCGGAAACGGTGAAGGATGAATTT ATGCTGAAGGCTTATTTTATCTCAGCTTTGACGAATGAAGAAGCGGATGAGCTATTCACC GATCAGCTCGTAAAGCGAAAGGAGAAGTTGTCCGATCTGGAAAACAGTTATCATGAACT GATGACATCCTCCGAGGAGGCGGATTCCTTTTCTTCTCCGGATTTCGGCCATTATCTCGT GCTGACAAAAGCGCTGGAGCGGGAAAGGAATTATATTTCCTGGCTTGAGCATATTTTGG CTCTCATCAAAAAAGCATAA
P_{PadC}-RBS-T1	GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAA TGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACCTTTTTAC CGGGTGAAGCATCTCATCATTTGACAGTCCATTTTAACAGCGTTACAATTAATCATGTAA ATAGTTACATGTATATATAAACATAATGTCTTGGGAGGTGATAC
P_{PadC}-RBS-T2	GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAA TGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACCTTTTTAC CGGGTGAAGCATCTCATCATTTGACAGTCCATTTTAACAGCGTTACAATTAATCATGTAA ATAGTTACATGTAT
P_{PadC}-RBS-T3	GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAA TGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACCTTTTTAC CGGGTGAAGCATCTCATCATTTGACAGTCCATTTTAACAGCGTT
P_{PadC}-RBS-T4	GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAA TGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACCTTTTTAC CGGGTGAAGCATCT
Original BaPadR binding sequence	CATGTAAATAGTTACATGTATATATAAACATA

BaPadR-1	CATGTAAATAGTTACATG
BaPadR-2	CATGTAAATAGTTACATG
Phy11[c]	ACATGTAAATAGTTACATGTTGACACATGTAAATAGTTACATGATACTGAGCACATCAG CAGGACGCACTGACC
Phy12[c]	ACATGTAAATAGTTACATGTTGACACATGTATATATAAACATAATACTGAGCACATCAG CAGGACGCACTGACC
Phy21[c]	ACATGTATATATAAACATATTGACACATGTAAATAGTTACATGATACTGAGCACATCAG CAGGACGCACTGACC
Phy22[c]	ACATGTATATATAAACATATTGACACATGTATATATAAACATAATACTGAGCACATCAG CAGGACGCACTGACC
Phy35[d]	TAAATTCATGTAAATAGTTTTGACAATATATAAACATAACAAGATACTGAGCACATCAG CAGGACGCACTGACC
Phy10[d]	TAAATTATCTCTGGCGGTGTTGACATCATGTAAATAGTTACAGATACTTATAAACATAGC AGGACGCACTGACC
P_{PadC}-C6A-RBS[d,e]	GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAA TGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACCTTTTTAC CGGGTGAAGCATCTCATCATTTGACAGTCCATTTTAACAGCGTTACAATTAATAATGTAA ATAGTTACATGTATATATAAACATAATGTCTTGGGAGGTGATACGTCCTGAATGAACCTT CTTTAAAAAAAGTGGAATTCATT**AAAGAGGAGAAA**
P_{padC} -C6T-RBS[d,e]	GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAA TGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACCTTTTTAC CGGGTGAAGCATCTCATCATTTGACAGTCCATTTTAACAGCGTTACAATTAATTATGTAA ATAGTTACATGTATATATAAACATAATGTCTTGGGAGGTGATACGTCCTGAATGAACCTT CTTTAAAAAAAGTGGAATTCATT**AAAGAGGAGAAA**
P_{padC} -C6G-RBS[d,e]	GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAA TGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACCTTTTTAC CGGGTGAAGCATCTCATCATTTGACAGTCCATTTTAACAGCGTTACAATTAATGATGTAA ATAGTTACATGTATATATAAACATAATGTCTTGGGAGGTGATACGTCCTGAATGAACCTT CTTTAAAAAAAGTGGAATTCATT**AAAGAGGAGAAA**

P_{padC} -T8A-RBS[d,e]	GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAA TGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACCTTTTTAC CGGGTGAAGCATCTCATCATTTGACAGTCCATTTTAACAGCGTTACAATTAATCAAGTAA ATAGTTACATGTATATATAAACATAATGTCTTGGGAGGTGATACGTCCTGAATGAACCTT CTTTAAAAAAAGTGGAATTCATT**AAAGAGGAGAAA**
P_{padC} -T8C-RBS[d,e]	GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAA TGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACCTTTTTAC CGGGTGAAGCATCTCATCATTTGACAGTCCATTTTAACAGCGTTACAATTAATCACGTAA ATAGTTACATGTATATATAAACATAATGTCTTGGGAGGTGATACGTCCTGAATGAACCTT CTTTAAAAAAAGTGGAATTCATT**AAAGAGGAGAAA**
P_{padC} -T8G-RBS[d,e]	GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAA TGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACCTTTTTAC CGGGTGAAGCATCTCATCATTTGACAGTCCATTTTAACAGCGTTACAATTAATCAGGTAA ATAGTTACATGTATATATAAACATAATGTCTTGGGAGGTGATACGTCCTGAATGAACCTT CTTTAAAAAAAGTGGAATTCATT**AAAGAGGAGAAA**
P_{padC} -T18'A-RBS[d,f]	GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAA TGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACCTTTTTAC CGGGTGAAGCATCTCATCATTTGACAGTCCATTTTAACAGCGTTACAATTAATCATGTTA ATAGTTACATGTATATATAAACATAATGTCTTGGGAGGTGATACGTCCTGAATGAACCTT CTTTAAAAAAAGTGGAATTCATT**AAAGAGGAGAAA**
P_{padC} -T18'C-RBS[d,f]	GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAA TGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACCTTTTTAC CGGGTGAAGCATCTCATCATTTGACAGTCCATTTTAACAGCGTTACAATTAATCATGTGA ATAGTTACATGTATATATAAACATAATGTCTTGGGAGGTGATACGTCCTGAATGAACCTT CTTTAAAAAAAGTGGAATTCATT**AAAGAGGAGAAA**
P_{padC} -T18'G-RBS[d,f]	GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAA TGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACCTTTTTAC CGGGTGAAGCATCTCATCATTTGACAGTCCATTTTAACAGCGTTACAATTAATCATGTCA ATAGTTACATGTATATATAAACATAATGTCTTGGGAGGTGATACGTCCTGAATGAACCTT CTTTAAAAAAAGTGGAATTCATT**AAAGAGGAGAAA**

[a] The mutations of H38 and S39 residue in the BaPadR was highlighted in yellow.

[b] Green sequence in BaP_{PadC}-RBS represents the engineered RBS sequence (in bold) along with the inserted EcoRI sequence.

[c] The BaPadR binding box1 (BaPadR-1) was highlighted in blue. The BaPadR binding box 2 (BaPadR-2) was highlighted in green.

[d] Deep gray highlight represents shows the entire (original) BaPadR bind box, but when overlapping with the -35 or -10 region, the corresponding sequence was replaced with the -35 or -10 region of the pL_{lacO1} promoter, respectively.

[e] The base replacement on C6 and T8 sites of the BaPadR-1 binding box on the BaP_{PadC} promoter was highlighted in red.

[f] The base replacement on the T18' site of the BaPadR-1 binding box on the BaP_{PadC} promoter was highlighted in violet. The mutations occurred on the complementary chain.

CHAPTER 3

RESULT

3.1 Identifying a potential PadR in *Bacillus Amyloliquefaciens*.

The protein sequence of the well-characterized PadR from *Bacillus subtilis 168* (BsPadR) in the previous study*(26)* was used as the template for BLAST (https://blast.ncbi.nlm.nih.gov/). An 88% similarity amino acid sequence compared with BsPadR was found (BAMF_RS24485, NCBI reference ID: WP_013351422.1) in *Bacillus amyloliquefaciens* ATCC 23350 (**Figure 1a**). We hypothesized that the regulator found in *Bacillus amyloliquefaciens* has a similar function to BsPadR due to its high similarity in protein sequence. This protein is referred to as BaPadR and belongs to the PadR family of transcriptional regulators. However, despite its annotation, BaPadR's function has not yet been experimentally validated. Sequence alignment further highlighted the difference between BaPadR and BsPadR in the N75-D88 and A107-D145 sequence regions (**Figure 1a**). Since the A107-D145 region overlaps the substrate binding pocket, we hypothesized that BaPadR might possess a different inducer scope or even a varied dynamic property. Based on the previous study*(22)* on the crystal structure information of the PadR from *Bacillus subtilis subsp. spizizenii str. W23* (BssPadR, with a sequence similarity of 89.6% compared to BaPadR), the molecular model of BaPadR was built by employing Swiss-Model*(32)* (https://swissmodel.expasy.org/) using the BssPadR (PDB ID: 5Y8T) as the template. These two proteins can be nearly perfectly aligned (**Figure 1c**). The residues that interact with the natural substrate, *p*-coumaric acid, were conserved and identified to be H155 and R165 in BaPadR (**Figure 2**).

Having identified BaPadR, our next objective was to locate the promoter that it controls in *B. amyloliquefaciens*. Based on our understanding that BsPadR regulates the expression of phenolic acid decarboxylase (PadC) in *B. subtilis(26)*, we hypothesized that a similar mechanism might be at play in *B. amyloliquefaciens*. To locate PadC in *B. amyloliquefaciens*, we used the protein sequence of PadC from *B. subtilis* (BsPadC) as a template. A hit (BAMF_RS37310, NCBI reference ID: WP_013353702.1) with 97.5% similarity was found and was believed to be the PadC of *B. amyloliquefaciens* (BaPadC) (**Figure 1b**). We located the PadC promoter using the upstream DNA sequence (700 bp) of the *BapadC* gene. During this process, we identified a short (429 bp) open reading frame (ORF) within the speculated promoter region (**Figure 1c**). As there are two short pseudogenes, *yveF*, and *yveG*, as the components of the P_{padC} promoter in *B. subitilis* 168*(26)*, we hypostasized that the homologous protein of YveG in *B. amyloliquefaciens* was encoded in this ORF due to the length of this ORF. The protein sequence of YveG in *B. subitilis* 168 (BsYveG) was used as the template for BLAST. And a 55.6% similarity protein sequence was found (**Figure 1d**). However, this similarity was not enough to conclude that this was YveG homologous protein in *B. amyloliquefaciens*. Further observing the alignment results, we found that a 46-aa region at the N terminus of BaYveG did not match the BsYveG. After removing this sequence, a hit of 83.2% similarity in protein sequence was found, which confirmed our hypothesis that this was likely the YveG homologous protein in *B. amyloliquefaciens* (**Figure 1d**). Thus, the BaP_{pad}C promoter was hypothesized to be the upstream sequence of the Ba*yveG* gene (**Figure 1c**).

(a)

```
# Aligned_sequences: 2
# 1: BsPadR
# 2: BaPadR
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 183
# Identity:     145/183 (79.2%)
# Similarity:   161/183 (88.0%)
# Gaps:           1/183 ( 0.5%)
# Score: 743.0
#
#
#=======================================

BsPadR             1 MRVLKYAILGLLRKGELSGYDITSYFKEELGQFWSAKHSQIYPELKKLTD     50
                     ||:||||||||||||||||:|||||||||||||||||||||||||||||.
BaPadR             1 MRILKYAILGLLRKGELSGYDISSYFKEELGQFWSAKHSQIYPELKKLTA     50

BsPadR            51 EGFITFRTTIQGTKLEKKMYTLTDSGKQELHDWLIRHQPIPETVKDEFML    100
                     ||||||||.|||||||||||||||:::||..||.:...||||||||||||
BaPadR            51 EGFITFRTAIQGTKLEKKMYTLTDNGERELCAWLTKKDPIPETVKDEFML    100

BsPadR           101 KAYFISCLSRQEASDLFKDQLQKRQAKLSDLQGSYEKLMASAEPM-SFSS    149
                     ||||||.|:.:||.:||.|||.||:.|||||:.||.:||.|:|.. ||||
BaPadR           101 KAYFISALTNEEADELFTDQLVKRKEKLSDLENSYHELMTSSEEADSFSS    150

BsPadR           150 PDFGHYLVLTKALEREKNYVSWLESILAMIDKD    182
                     ||||||||||||||:||:||||.|||:|.|.
BaPadR           151 PDFGHYLVLTKALERERNYISWLEHILALIKKA    183
```

(b)

```
# Aligned_sequences: 2
# 1: BsPadC
# 2: BaPadC
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 161
# Identity:     147/161 (91.3%)
# Similarity:   157/161 (97.5%)
# Gaps:           0/161 ( 0.0%)
# Score: 827.0
#
#
#=======================================

BsPadC             1 MENFIGSHMIYTYENGWEYEIYIKNDHTIDYRIHSGMVAGRWVRDQEVNI     50
                     ||||||||||||||||||||||||||||||||||||||||.||||||||||
BaPadC             1 MENFIGSHMIYTYENGWEYEIYIKNDHTIDYRIHSGMVGGRWVRDQEVNI     50

BsPadC            51 VKLTEGVYKVSWTEPTGTDVSLNFMPNEKRMHGIIFFPKWVHEHPEITVC    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
BaPadC            51 VKLTEGVYKVSWTEPTGTDVSLNFMPNEKRMHGIIFFPKWVHEHPEITVC    100

BsPadC           101 YQNDHIDLMKESREKYETYPKYVVPEFAEITFLKNEGVDNEEVISKAPYE    150
                     ||||:||:|||||||:||||||||||:||:|.|.|::||.:||:||||
BaPadC           101 YQNDYIDVMKESREKYDTYPKYVVPEFADITYLNNAGINNEALISEAPYE    150

BsPadC           151 GMTDDIRAGRL    161
                     |||||||||:|
BaPadC           151 GMTDDIRAGKL    161
```

(c) BaP_{padC} **promoter**

GATTCCTTTCTAGTTCGCGAGCCGCATGTGAGCATGGACCTAAGGAGCGCCAGAGTAAATGAAAAAGACAAGGGTTTCGGCATTTGCCGTCCTTGTCTTTTTTATTAAACACC

BaPadR-1

TTTTTACCGGGTGAAGCATCTCATCATTTGACAGTCCATTTTAACAGCGTTACAATTAATCATGTAAATAGTTACATGTATATATAAACATAATGTCTTGGGAGGTGATACGTCCT

BayveG ?　　BaPadR-2

GAATGAACCTTCTTTAAAAAAAGTGATGTCTCTGCAAATTAAAGTGGATGAACCGATCGTTGCTGGGCAGACTGGATTAGGGAAAAGGCAGCTGATCCTGATTCGGGGCGG

AACGGTCAGCGGCGCCGTAAAGGGGCGTGTTCTCCCGGGCGGGGCAGACGCCCAGATCATCCGTCCGAACGGCCGGGTTGATCTTTCTGCTAGGTACGCGCTGGAAACC

GAAGAGCACGAAGTCATTTATGTTGAGAACAACGGCATCCGCCAAGTCAGTGAACCGTTCCGCCAGCAGGCTGCAGAGGGGCGCATCATTGATCACGAGCACGTGTATTTC

CGCACTGTGCCTGTGTTTAAAACAAGCAGTAAGGCATATCAGCATCTGCAGGACCGGATGTTTATCGGGGCTGCGGTCAGATTGCCGGATGACATTCGGTTGGATATTTATGA

BapadC

AGTTCAATAATAATAATATGGAGAGTGAGTTAGCATG

(d)

```
# Aligned_sequences: 2
# 1: BsYveG
# 2: BaYveG-core
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 142
# Identity:      66/142 (46.5%)
# Similarity:    79/142 (55.6%)
# Gaps:          47/142 (33.1%)
# Score: 352.0
#
#
#=======================================

BsYveG             1 ------------------------------------------------MIR      3
                                                                    :||
BaYveG             1 MSLQIKVDEPIVAGQTGLGKRQLILIRGGTVSGAVKGRVLPGGADAQIIR     50

BsYveG             4 ANGRTDLSARYVIETADHELIYIENNGIRQVSKPFRKQAAAGEIIEPEHV     53
                     .|||.||||||.:||.:||:||:|||||||:|||:|||.|.||:.|||
BaYveG            51 PNGRVDLSARYALETEEHEVIYVENNGIRQVSEPFRQQAAEGRIIDHEHV    100

BsYveG            54 YFRTVPTFETGSEVYQWLHDRLFIGSAERTPDYVLLDIYEVQ     95
                     ||||||.|:|.|:.||.|.||:|||:|.|.||.::|||||||
BaYveG           101 YFRTVPVFKTSSKAYQHLQDRMFIGAAVRLPDDIRLDIYEVQ    142
```

(e)

```
# Aligned_sequences: 2
# 1: BsYveG
# 2: BaYveG-core
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 96
# Identity:      66/96 (68.8%)
# Similarity:    79/96 (82.3%)
# Gaps:           1/96 ( 1.0%)
# Score: 352.0
#
#
#=======================================

BsYveG             1 -MIRANGRTDLSARYVIETADHELIYIENNGIRQVSKPFRKQAAAGEIIE     49
                      :||.|||.||||||.:||.:||:||:|||||||||:|||:|||.|.||:
BaYveG-core        1 QIIRPNGRVDLSARYALETEEHEVIYVENNGIRQVSEPFRQQAAEGRIID     50

BsYveG            50 PEHVYFRTVPTFETGSEVYQWLHDRLFIGSAERTPDYVLLDIYEVQ     95
                     .|||||||||.|:|.|:.||.|.||:|||:|.|.||.::|||||||
BaYveG-core       51 HEHVYFRTVPVFKTSSKAYQHLQDRMFIGAAVRLPDDIRLDIYEVQ     96
```

Figure 1. Sequence alignment results: (a) Protein sequence alignment of BaPadR and BsPadR. The red rectangle (rounded corners) highlighted the variation regions between BaPadR and BsPadR. **(b)** Protein sequence alignment of BaPadC and BsPadC. **(c)** The upstream sequence of BaPadC. The sequence in green represented the BaP_{padC} promoter used in this study. The sequence in yellow exhibited the putative binding box of the BaPadR. The blue sequence showed the gene of Ba*yveG*. **(d)** Protein sequence alignment of BaYveG and BsYveG. **(e)** Protein sequence alignment of truncated BaYveG and BsYveG.

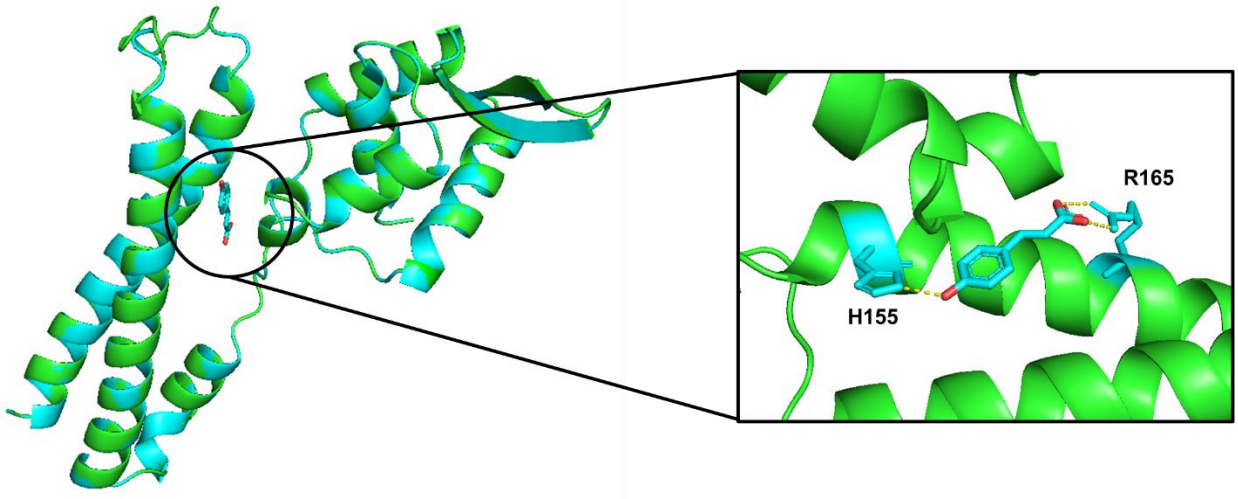

Figure 2: Simulation of BaPadR structure based on the crystal structure of BssPadR by SWISS-Model with substrate binding. The green protein structure represents the BaPadR structure. The blue protein structure represents the BssPadR structure.

3.2 Rebuilding the BaPadR-BaP_{PadC} biosensor system in *E. coli*.

After identifying the BaPadR and BaP_{padC} promoter, we aimed to reconstruct this biosensor system in *E. coli* to test its dynamic performance (**Figure 3a**). The BaPadR regulator and BaP_{padC} promoter were cloned from the genome of *B. Amyloliquefaciens*. The BaPadR was carried by medium-copy plasmid pCS27*(1)* and was under the control of the promoter pL_{lacO1}, a IPTG-inducible (IPTG was short for isopropyl ß-D-1-thiogalactopyranoside) promoter, resulting in the plasmid pCS-BaPadR. The BaP_{padC} promoter was applied to control the expression of the

reporter gene *egfp* in the high-copy plasmid pHA-egfp-MCS resulting in the plasmid pHA-BaP_{padC}-WT-egfp. The plasmid pHA-BaP_{padC}-WT-egfp was co-transferred with pCS-BaPadR into *E. coli* BW25113 F'. An empty plasmid pCS27 without the BaPadR expression was also co-transferred with pHA-BaP_{padC}-WT-egfp to serve as a positive control (PC). *p*-coumaric acid was selected as the induction ligand to test the BaPadR biosensor system since *p*-coumaric was reported as the natural substrate for BsPadR and BssPadR*(26)*. Gradient concentrations of *p*-coumaric acid (0, 300, 600 mg/L) were added into the cell culture to induce the sensor system, and 0.5 mM IPTG was added to induce the BaPadR expression. The green fluorescence intensity was normalized with cell density (OD_{600}) to represent the promoter strength. Unexpectedly, the promoter showed an extremely low output strength and did not exhibit any response toward the BaPadR or *p*-coumaric acid (**Figure 3b**). We suspected that the low activity of the original RBS of the BaPpadC promoter resulted in this situation. Thereby, an engineered strong RBS used in our previous studies*(28)* was added between the BaP_{padC} promoter and *egfp* gene. (**Figure 3c**). The new promoter configuration was constructed to the plasmid pHA-egfp-MCS, resulting in pHA-BaPpadC-RBS-egfp, and it was utilized to test the dynamic performance. The promoter with a strong RBS displayed a very strong output when BaPadR was absent (**Figure 3b**, PC of BaPpadC-RBS), with fluorescence intensity normalized with OD600 exceeding 12,000 a.u. Upon induction of BaPadR expression by 0.5 mM IPTG, the promoter activity drastically decreased, with nealy 96% of its activity being inhibited by BaPadR, resulting in only 539.16 a.u. being detectable. However, further addition of *p*-coumaric did not relieve the inhibition. Meanwhile, the promoter activity was further decreased (**Figure 3b**). In this case, we hypothesized that the high BaPadR expression level resulted in the strict inhibition of promoter activity, which was also observed in our previous study*(26)*. Thus, gradient concentrations of

IPTG were added to the medium to find out the optimized BaPadR expression level (**Figure 3d**). The biosensor system exhibited a response to *p*-coumaric when the IPTG concentration was reduced to 10 μM. A further decrease to 0.5 μM increased the responsiveness, but also increased the leaky activity of the promoter. These results demonstrated the successful reconstruction of the BaPadR biosensor system in *E. coli*. Moreover, the high inhibition efficiency of BaPadR towards the BaP_{padC} promoter was revealed, as it maintained a high inhibition ability under 2.5 μM IPTG, which is 200-fold lower than the normal induction concentration of 0.5 mM. Under 2.5 μM IPTG, the sensor system maintained a relatively low level of leakage, and 600 mg/L of *p*-coumaric acid could release 60% of promoter activity from BaPadR inhibition. Therefore, 2.5 μM of IPTG was selected to induce the BaPadR expression for the remainder of the study.

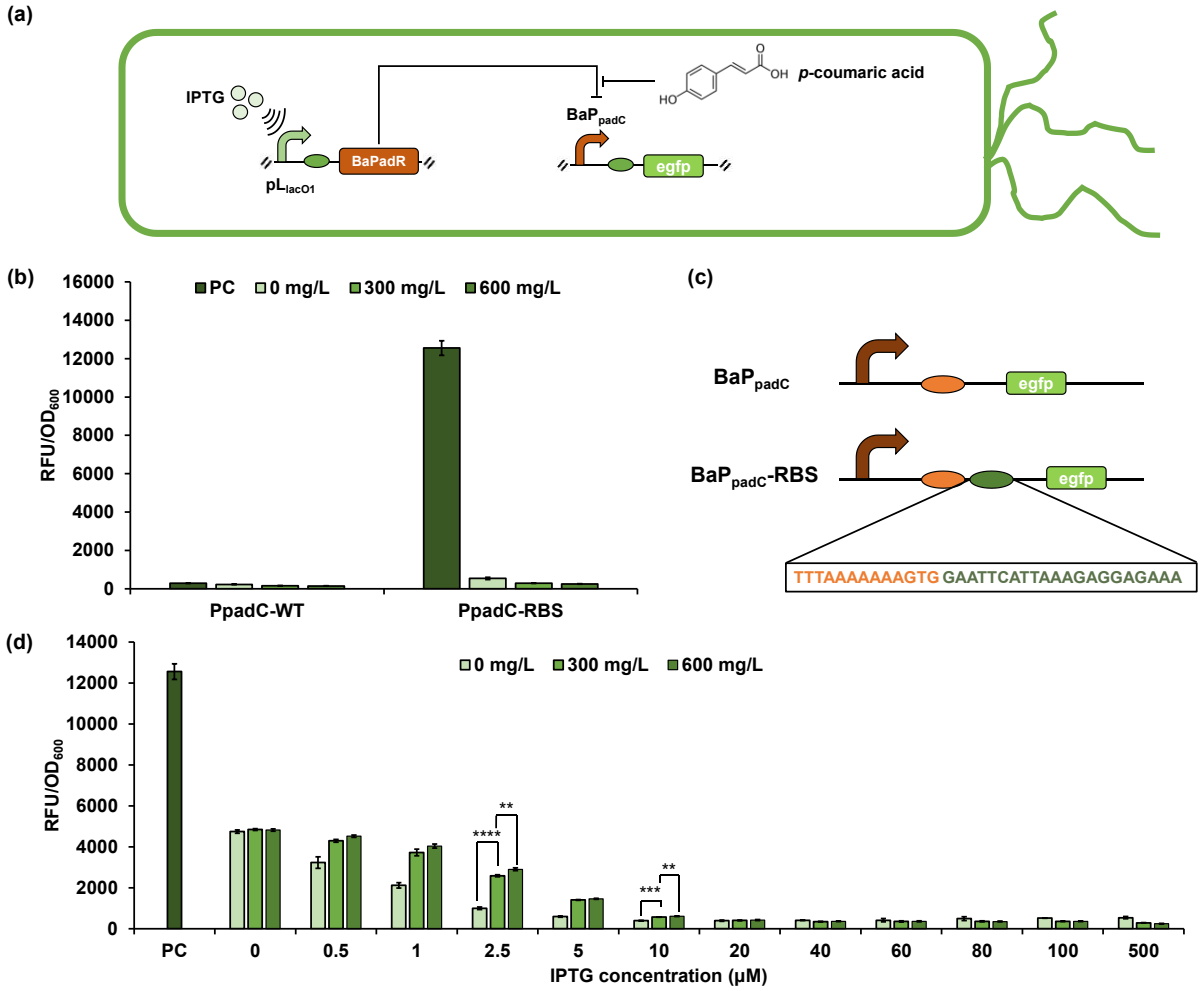

Figure 3. Reconstructing the BaPadR-BaPpadC sensor system in E. coli: (a) Illustration for the reconstructed BaPadR-BaP$_{padC}$ sensor system in *E. coli*. The BaPadR expression was carried by IPTG-inducible promoter pL$_{lacO1}$ and repressed BaP$_{padC}$ promoter, who controlled the expression of reporter protein egfp. The *p*-coumaric acid served as the inducer to release the inhibition from BaPadR. **(b)** The dynamic performance of wild-type BaP$_{padC}$ promoter and BaP$_{padC}$-RBS promoter. **(c)** Schematic illustration of how the strong RBS was added. The brown sequence represents the native promoter sequence. Between the *egfp* gene and promoter, the additional RBS (dark green oval) was inserted. **(d)** The dynamic performance of the BaPadR-BaPpadC sensor system in *E. coli* with gradient concentration of IPTG. BaPpadC-WT represents the wild-type BaP$_{padC}$ promoter-controlled *egfp* expression cassette. BaPpadC-RBS was the wild-type BaP$_{padC}$ promoter with the addition of an engineered RBS. PC represents the positive control without BaPadR expression. All data represent the mean of 3 biologically independent samples and

error bars show standard deviation. Statistics were calculated by two independent samples t-test: *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001; NS, not significant.

3.3 Substrate scope of the BaPadR-based sensor system.

In the previous sections, we successfully established the BaPadR-BaP_{padC} sensor system in *E. coli*, and we also gained the BaPadR-S39A variant with a good dynamic range. Our next goal is to investigate the inducer scope of the BaPadR-based sensor system. As the protein sequence in the substrate-binding domain (from A107-D145) of BaPadR differs from that of BsPadR (Figure 1a), we believe that BaPadR has substrate specificity distinct from that of BsPadR.

PadR is commonly known as a regulator responsive to phenolic acids. To investigate its responsiveness, we tested PadR's reaction to three phenolic acids as well as four smaller derivatives of benzoic acid, including caffeic acid, cinnamic acid, ferulic acid, 4-hydroxybenzoic acid, anthranilic acid, salicylic acid, and vanillic acid. Those aromatic compounds are important precursors for flavonoids*(33)*, alkaloids*(34)*, and coumarins*(35)*. Compared to the non-induced control group, 300 mg/L of caffeic acid, ferulic acid, and salicylic acid generated a noticeable increase in *egfp* expression, which indicated potential responsiveness of BaPadR towards these chemicals (**Figure 4a**). Based on the result, we further tested the dynamic ranges of the three additional substrates besides the *p*-coumaric acid (**Figure 4b**). The best factor was determined to be *p*-coumaric acid, which resulted in a 3.36-fold increase of *egfp* expressional level under the concentration of 600 mg/L, followed by ferulic acid (2.58-fold), salicylic acid (1.56-fold), and caffeic acid (1.40-fold), respectively. The induction by salicylic acid was unexpected as it is much smaller than the other inducers and lacks the hydroxy group at the *para* position, which we thought is critical to be recognized as the substrate of the PadR regulator*(29)*.

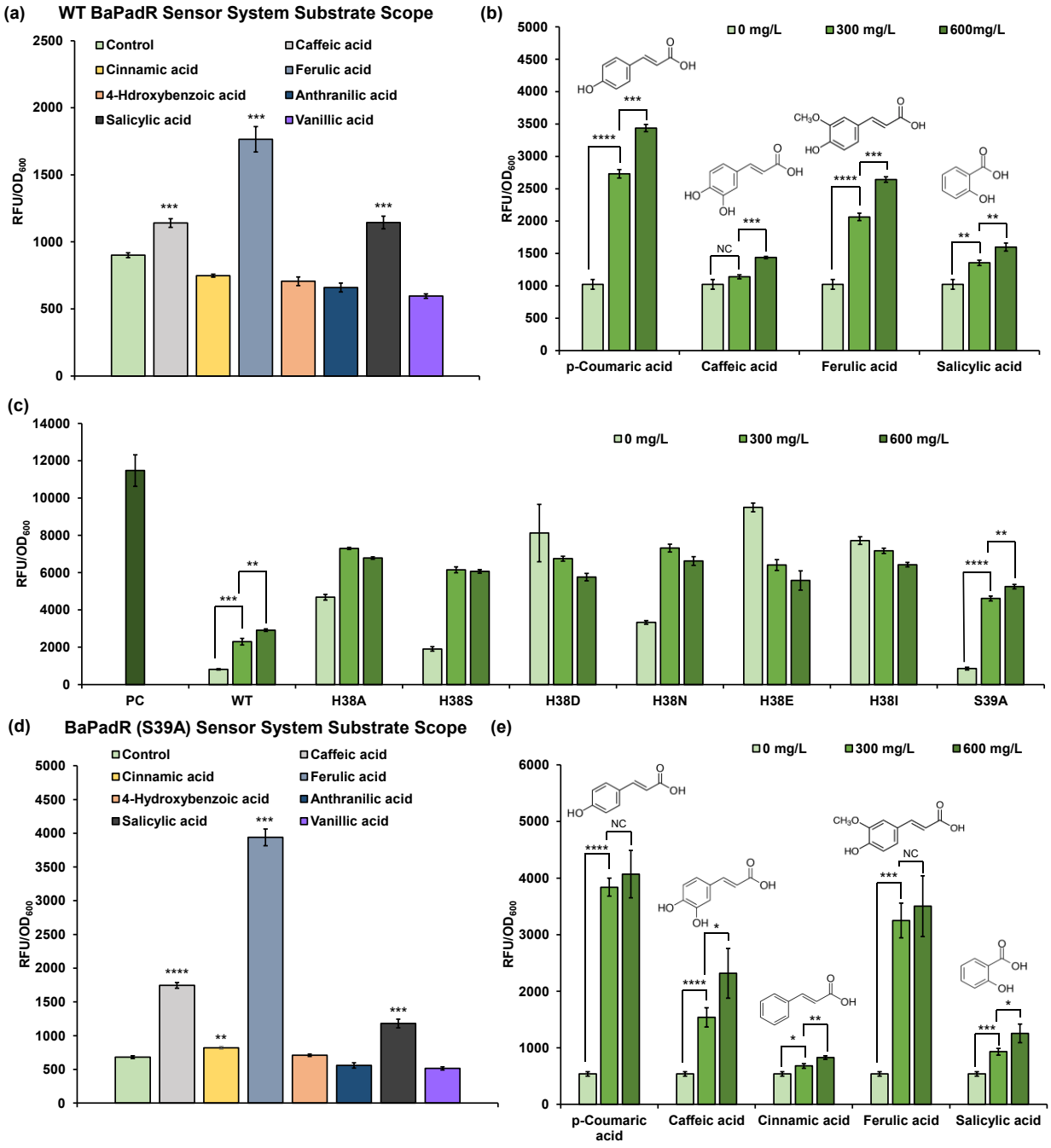

Figure 4. Exploring the substrate scope of the re-constructed BaPadR-based sensor system, and site-directed mutagenesis to weaken the BaPadR binding affinity: (a) Screening the potential ligands of wild-type BaPadR. 300 mg/L of the different substrates were added. **(b)** The dynamic performance of the wild-type BaPadR towards *p*-coumaric acid, caffeic acid, ferulic acid, and salicylic acid. **(c)** Screening the potential ligands of the BaPadR-S39A

variant. 300 mg/L of the different substrates were added. (d) The dynamic performance of the BaPadR-S39A variant towards *p*-coumaric acid, caffeic acid, cinnamic acid, ferulic acid, and salicylic acid. All data represent the mean of 3 biologically independent samples and error bars show standard deviation. Statistics were calculated by two independent samples t-test: *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001; NS, not significant.

3.4 Site-directed mutagenesis to weaken the BaPadR binding affinity.

The BaPadR-based sensor system exhibited a usable dynamic performance. However, the inhibition of the BaPadR toward the BaP_{padC} promoter cannot be fully released. We assumed that the high inhibition efficiency of BaPadR was due to the high binding affinity between BaPadR and the promoter BaP_{padC}. To further increase the dynamic range of the BaPadR sensor system, we aimed to weaken the binding affinity between BaPadR and the promoter BaP_{padC}.

Previous research on BsPadR*(26)* found that mutating either the H38 or S39 amino acids to alanine resulted in the two most significant increases in output strength, with H38 showing an increase of over 3-fold and S39 showing an increase of approximately 1.6-fold. This increase was caused by a decrease in binding affinity between BsPadR and its corresponding promoter, P_{padC}, resulting from the mutations. Since the BaPadR possesses the identical protein sequence in the promoter binding region (Figure 1a), we hypothesized that the same scenario will apply when mutating on the H38 and S39 sites of BaPadR. Since the prominent increase in the dynamic response was generated from the mutations on H38 sites of BsPadR, we believe the amino acid size would also be the dominant factor influencing the binding between the BaPadR and the corresponding promoter. We hypothesized that decreasing the size of the amino acid residue at the H38 and S39 sites would reduce steric hindrance, resulting in weaker binding affinity between the BaPadR regulator and the P_{padC} promoter. As a result, the activation strength of the BaP_{padC} promoter induced by *p*-coumaric acid would increase, but the expression leakage would

also enhance. We predicted that the response range would gradually increase as the size of the amino acid residues decreased. So, we designed several BaPadR mutants (H38I, H38E, H38N, H38D, H38S, H38A, and S39A) on H38 and S39 sites. The dynamic performance of all variants was exhibited in **Figure 4c**. Surprisingly, the H38D, H38E, and H38I variants abolished regulation by BaPadR. We suspected that the change in amino acid altered the protein structure of these variants, preventing the binding of BaPadR to the promoter. The results suggested that the binding affinity between the protein and DNA is not solely dependent on steric hindrance; polarity and charge status can also influence binding affinity. BaPadR-S39A variant performed the best dynamic range, with a 1.80-fold increase in the response under 600 mg/L *p*-coumaric acid and no leaking expression increase of the biosensor system compared with the wild-type BaPadR. The S39A variant of the BaPadR-BaP_{padC} biosensor system enabled a 1.71-fold increase in the dynamic range. The enhanced dynamic range of the biosensor would be more practical for metabolic engineering applications.

Since the modification of the amino acid may alter the regulator protein structure, it is likely that the inducer scope of the BaPadR variant would also be changed. Thus, we tested the substrate scope of the best-performed BaPadR (S39A) variant. By following the previous procedure for testing the inducer scope of the wild-type BaPadR, 300 mg/L of different ligands were added to test the initial responsiveness. We further used the gradient concentration of those chemicals plus *p*-coumaric acid to induce the BaPadR-S39A-BaP_{padC} sensor system (**Figure 4d**). *p*-coumaric acid still performed best to activate the *egfp* expression, which resulted in a 7.55-fold increase of *egfp* expression level under 600 mg/L inducer concentration, followed by ferulic acid (6.50-fold), caffeic acid (4.30-fold), salicylic acid (2.33-fold), and cinnamic acid (1.53-fold). Surprisingly, cinnamic acid, which also lacks the hydroxy group at the *para* position, can serve

as a ligand to release the expression inhibition from the BaPadR-S39A regulator (**Figure 4e**). Notably, compared to WT BaPadR, all responsive substrates showed increased dynamic ranges when inducing the new variant BaPadR (S39A).

3.5 Identifying the potential BaPadR binding box.

Through the above study, we successfully reconstructed BaPadR-BaP_{padC} sensor system in *E. coli* and explored the substrate scopes. Next, we aimed to identify the DNA binding sequence identified and bound by BaPadR. By comparing the BsPadR binding box with the BaP_{padC} promoter, two 18-bp sequences in the promoter were identified (**Figure 1c**): one is "-CATGTAAATAGTTACATG-" (labeled with BaPadR-1) that was completely identical to the binding sequence of BsPadR*(22, 26)*,the other one is "-CATGTATATATAAACATA-" (labeled with BaPadR-2) which contains 5 mismatches but is still highly similar to the BsPadR binding sequence. These two units overlapped in the BaP_{padC} promoter region (**Figure 1c**). A similar scenario also happened in BsPadR*(22, 26)*. Thus, we hypothesized that these two sequences are the potential binding box of BaPadR.

To verify our assumption, we first decided to truncate the promoter to eliminate the binding sequence and determine whether the promoter could still respond to BaPadR and *p*-coumaric acid. Thus, four truncated promoters (Trunc1-4) were designed (**Figure 5a**). The truncation process began upstream of the RBS region, with each truncated promoter shortening by 30 bp from the 3' end, resulting in Trunc1 to Trunc4 (**Figure 5a**). The Trunc1 still processed both BaPadR-1 and BaPadR-2 binding sequences. For the Trunc2, only the BaPadR-1 sequence ("-CATGTAAATAGTTACATG-") was kept, the BaPadR-2 ("-CATGTATATATAAACATA-") sequence was removed. For the promoter Trunc3 and Trunc4, the putative BaPadR binding

boxes were abolished in these two promoters. We tested the dynamic performance of those truncated promoters (**Figure 5b**). As expected, Trunc3 and Trunc4 showed almost no promoter activity and were not responsive to either BaPadR or *p*-coumaric acid. Contrarily, Trunc1 showed low promoter activity but could be still repressed by BaPadR and induced by *p*-coumaric acid. To our surprise, Trunc2 with only the BaPadR-1 binding sequence can still be controlled by BaPadR and *p*-coumaric acid. Meanwhile, the promoter strength of Trunc2 was higher than that of Trunc1. The promoter truncation experiment strongly suggested that there two DNA sequences (BaPadR-1 and BaPadR-2) are responsible for BaPadR recognition. Our findings indicate that the BaPadR-1 binding box is sufficient for the BaPadR to recognize and bind to the promoter sequence.

To further affirm the function of BaPadR-1 and BaPadR-2, We aim to investigate if the BaPadR binding boxes can be utilized in a "plug-and-play" approach. In the previous study*(36)*, the two LacO binding boxes were inserted into position 1 and position 2 of the pL promoter (**Figure 5c**) and converted this constitutive promoter to the commonly used inducible promoter pL_{lacO1}. Following this strategy, we designed several hybrid promoters (Phy11, Phy12, Phy21, Phy22, Phy35, and Phy10) by inserting the BaPadR binding boxes in various positions of the pLlacO1 promoter and with different orders (**Figure 5c**). Phy11, Phy12, Phy21, and Phy22 promoters were designed by replacing the LacO binding boxes with BaPadR binding boxes at different positions (**Figure 5c**). The Phy35 and Phy10 were designed by placing the original BaPadR binding box (“-CATGTAAATAGTTACATGTATATATAAACATA-”) to overlap with the -35 and -10 regions of the pL promoter (**Figure 5c**). When the original BaPadR binding sequence overlapped with the -35 or -10 region, we preserved the -35 or -10 region of the pL_{lacO1} promoter to keep the promoter function, and the corresponding sequence of the BaPadR binding

box was deleted. Those six hybrid promoters were constructed to pHA-egfp-MCS plasmid to control the expression of the reporter gene *egfp*. We tested the dynamic performance of the biosensor system harboring these six hybrid promoters among the WT PadR and PadR-S39A variants (**Figure 5d**). The original pL_{lacO1} promoter was also included as the control group. The dynamic performance of pL_{lacO1} indicated that the pL_{lacO1} promoter cannot be regulated by BaPadR and *p*-coumaric acid (**Figure 5d**). BaPpadC promoter exhibited greater activity compared to the pLlacO1 promoter (**Figure 5d**), highlighting its considerable potential for use in synthetic biology and metabolic engineering applications. Out of the six hybrid promoters, only Phy12 and Phy21 demonstrated the regulatory capacity of BaPadR and *p*-coumaric acid. Phy12 maintained a similar promoter activity to pL_{lacO1}, with BaPadR repressing about 82.8% of the promoter activity. The addition of 600 mg/L *p*-coumaric acid resulted in the recovery of approximately 29.7% of promoter activity (**Figure 5d**). The Phy21 promoter, which showed low expression strength, can also be repressed by the BaPadR with 91.7% of promoter activity. However, only 10.2% of the promoter activity can be released with 600 mg/L *p*-coumaric acid exist. Except for Phy12 and Phy21, the rest four hybrid promotes did not exhibit any inhibition when BaPadR was present. The BaPadR-S39A variant showed a better dynamic performance compared to WT BaPadR among the BaP_{padC}, Phy12, and Phy21 promoters (**Figure 5e**). The inhibition efficiency of BaPadR on Phy12 and Phy21 promoters was changed to 77.4% and 49.2%, respectively, but the response of these two promoters towards *p*-coumaric acid was also improved. Around 63.4% and 68.5% of the promoter activities can be recovered by inducing 600 mg/L *p*-coumaric acid for Phy12 and Phy21, respectively. With the results of the Phy12 and Phy21, the two sequences (BaPadR-1 and BaPadR-2) were verified to be the binding boxes recognized by BaPadR.

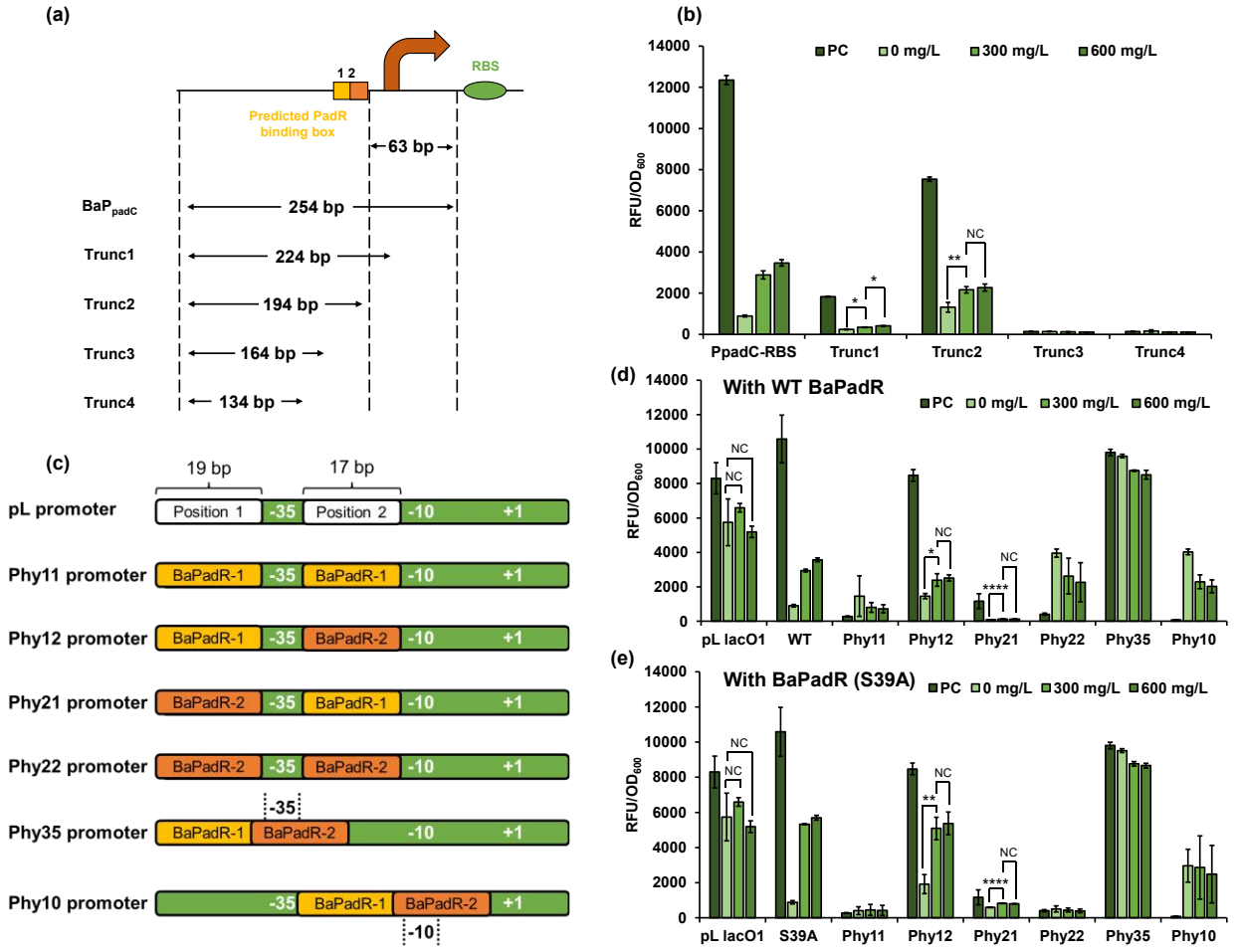

Figure 5. Identification of potential PadR binding box: (a) Illustrations for the truncation of BaPpadC promoter. **(b)** The dynamic performance of the truncated promoters. **(c)** Illustration of the hybrid promoter design strategy. **(d)** Dynamic performance of the hybrid promoters with WT BaPadR. **(e)** Dynamic performance of the hybrid promoters with BaPadR (S39A). All data represent the mean of 3 biologically independent samples and error bars show standard deviation. Statistics were calculated by two independent samples t-test: *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001; NS, not significant.

3.6 Base replacement on the BaPadR binding box to further weaken the binding affinity.

Given that the BaPadR binding boxes have already been characterized, our next objective was to further decrease the binding affinity between the regulator and the promoter, while

exploring the effects of nucleotide alterations in the binding box on the overall output of the sensor system.

Upon further investigation on the promoter truncation results, the Trunc2 promoter, which only possessed the BaPadR-1 sequence (-CATGTAAATAGTTACATG-), could be repressed by BaPadR and activated by *p*-coumaric acid (**Figure 5a and 5b**). We believed the BaPadR-1 sequence was the primary sequence recognized and bound by BaPadR. Thereby, we decided to replace the key bases in the BaPadR-1 binding box and investigate the impact on dynamic performance of the sensor system. In addition, as revealed in a previous study*(22)*, the BsPadR makes directly contacted with the bases on C6, T8, G9, T10, T17', and T18' sites (' indicate the bases are on the complementary strand) in the BsPadR-BsP_{paC} binding structure (**Figure 6a**). Base on the previous research*(22)*, nucleotides replacement on C6, T8, and T18' led to obvious reduction in the binding affinity of BsPadR-BsP_{paC} with the degree of reduction ranging from 1-fold to nearly 40-fold*(22)*. Since the BaPadR binding boxes is totally identical to the BsPadR binding box, we hypothesized that these three positions (C6, T8, and T18') of the nucleotides may also play a crucial role in the binding of BaPadR to the promoter. Therefor, we introduced base substitutions into the BaPadR-1 binding box at the C6, T8, and T18'. Based on the P_{padC} promoter, several base-replaced promoters were obtained, including P_{padC}-C6A-RBS, P_{padC}-C6T-RBS, P_{padC}-C6G-RBS, P_{padC}-T8A-RBS, P_{padC}-T8C-RBS, P_{padC}-T8G-RBS, P_{padC}-T18'A-RBS, P_{padC}-T18'C-RBS, P_{padC}-T18'G-RBS. All those promoter variants were constructed to pHA-egfp-MCS, carrying the expression of reporter gene *egfp*. For the Phy12 promoter, we did not induce base substitution into this hybrid promoter because the BaPadR-Phy12 sensor system severely impact the normal cell growth.

The dynamic performance of base-altered P_{padC} promoter variants was tested. The original P_{padC} promoter was included as a control. With the WT BaPadR, the change of C6 nucleotide slightly increased the leaking expression and the corresponding response to *p*-coumaric acid (**Figure 6b**). The binding affinity was further decreased when the base substitution happened on the T8 site (**Figure 6b**). With the WT BaPadR, P_{padC}-T8A exhibited the highest output level induced with 600 mg/L *p*-coumaric acid, which was near 4,810 a.u. Mutations at the T18' position resulted in a more complex dynamic range change scenario. We noticed that the alteration from thymine (T) to cytosine (C) on the T18’ site severely impacted the promoter activity, which resulted in a 39.0% promoter activity reduction (**Figure 6b PC group**). However, 61.0% promoter activity of P_{padC}-T18’C could be released with 600 mg/L *p*-coumaric acid (**Figure 6b**).

As we controlled the sensor system with the S39A variant, a similar trend by base substitution can be observed compared with that by WT BaPadR. However, the dynamic performance of each promoter was enhanced when BaPadR-S39A was performed. The activation activity of the P_{padC}-C6A, P_{padC}-C6T, and P_{padC}-C6G under 600 mg/L *p*-coumaric acid increased by 43.5%, 70.0%, and 55.0% compared to that controlled with WT BaPadR. The activation level of the P_{padC}-T8C promoter was further enhanced, and 83.8% of expression activity was recovered with the existence of 600 mg/L *p*-coumaric (**Figure 6c**). The P_{padC}-T18'A exhibited a similar dynamic performance compared to the original P_{padC} promoter. It is worth noting that expression strength of the P_{padC}-T18'C promoter can be recovered to 100% of the original strength under the control of the S39A variant after adding 600 mg/L inducer (**Figure 6c**).

In summary, through nucleotide substitutions, we obtained several engineered promoters that showed enhanced responsiveness while still maintaining relatively strict efficiencies, such as

P_{padC}-C6A, P_{padC}-T8A, and P_{padC}-T8G. These base-replacement promoters greatly expanded the reservoir of the phenolic acid-respond biosensor system, and they can be readily used in biosensor-enabled metabolic engineering applications.

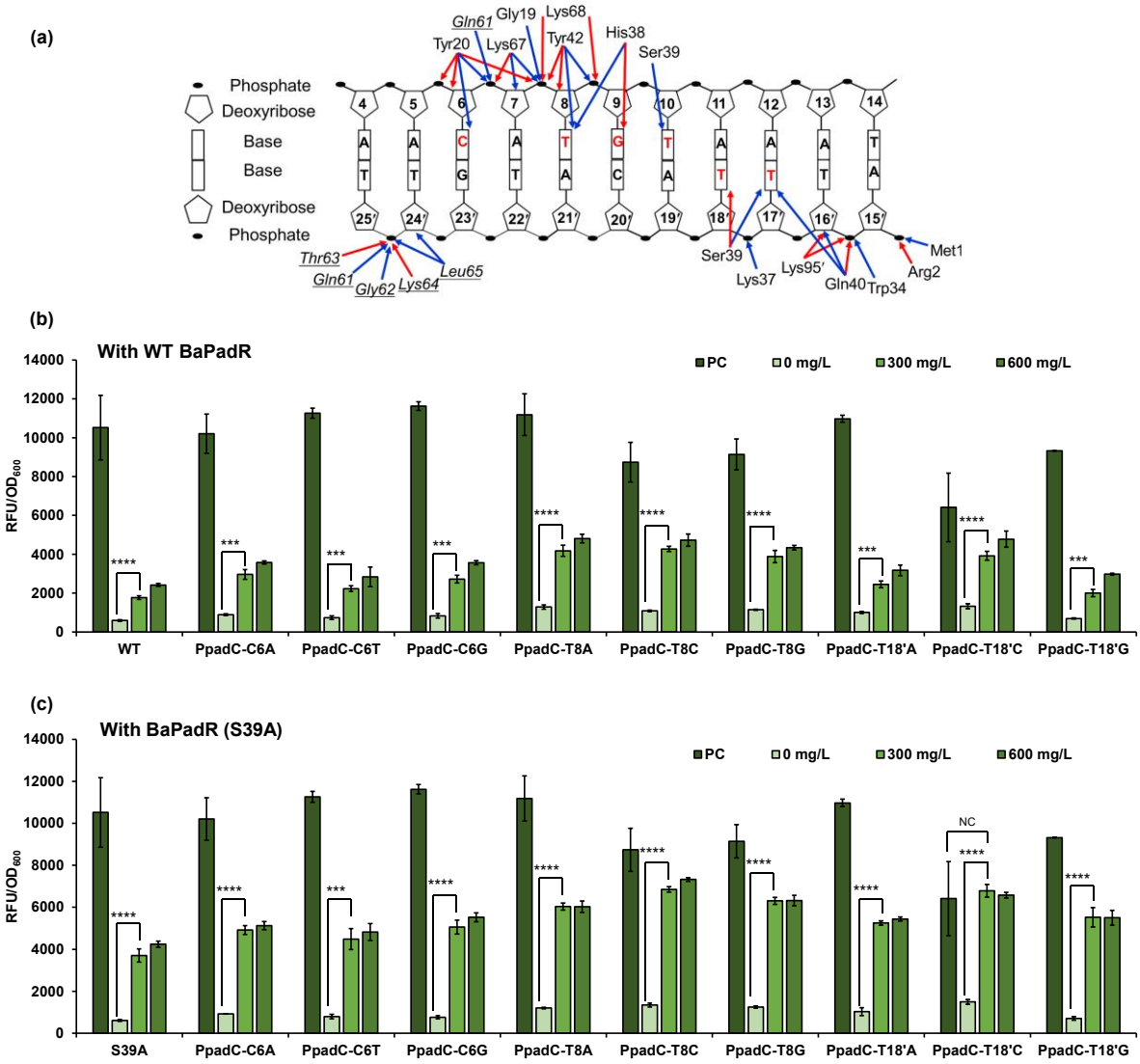

Figure 6. Exploring the influence of base substitution on the dynamic performance: (a) Illustration of the interaction between BsPadR amino acid residues and the P_{padC} promoter*(22)*. The red and blue arrows represent hydrogen bonds and van der Waals interactions, respectively. **(b)** Dynamic performance of the P_{padC} promoter and its variants controlled with WT BaPadR. **(c)** Dynamic performance of the P_{padC} promoter and its variants controlled

with BaPadR-S39A variant. All data represent the mean of 3 biologically independent samples and error bars show standard deviation. Statistics were calculated by two independent samples t-test: *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001; NS, not significant.

CHAPTER 4

DISCUSSION

The use of transcriptional factors-based biosensors has become increasingly important in constructing and regulating metabolic pathways for the biosynthesis of valuable chemicals. However, to expand the applications of this approach in metabolic engineering and synthetic biology, a current library of transcriptional factors-based biosensors needs to be expanded. In this study, we identified, characterized, and engineered a novel phenolic acid-respond regulator from *Bacillus amyloliquefaciens* (BaPadR). We successfully located the P_{padC} promoter regulated by BaPadR through sequence alignment. The function of BaPadR and the corresponding promoter BaPpadC was confirmed by re-establishing the sensor system in E. coli. The BaP_{padC} promoter showed a very strong output after the addition of a strong RBS sequence. During our study, we found the binding affinity between the BaPadR regulator and the BaP_{padC} promoter was high. To enhance the dynamic performance and dynamic range of the system, we performed site-directed mutagenesis and generated several variants, with the BaPadR-S39A variant performing the best. Compared to the previously well-characterized BsPadR regulator, BaPadR showed a unique response towards salicylic acid, and the BaPadR-S39A exhibited a response towards salicylic acid and cinnamic acid. Through sequence alignment and promoter truncation, we located the binding box in the promoter that is responsible for BaPadR recognition. The hybrid promoter construction using the binding box confirmed its function and demonstrated a "plug-and-play" feature. Further study on the change of base replacement in the binding box of the promoter expanded the dynamic performance scope of the phenolic acid-respond biosensor

system. This BaPadR-based biosensor system and its variants identified and developed in this study would be beneficial for further research involving biosensor-based applications.

REFERENCE

[1].Atsumi, S., Cann, A. F., Connor, M. R., Shen, C. R., Smith, K. M., Brynildsen, M. P., Chou, K. J. Y., Hanai, T., and Liao, J. C. Metabolic engineering of Escherichia coli for 1-butanol production. *Metab. Eng.* **2008**, *10* (6), 305-311. DOI: https://doi.org/10.1016/j.ymben.2007.08.003.

[2].Shen, C. R., and Liao, J. C. Metabolic engineering of Escherichia coli for 1-butanol and 1-propanol production via the keto-acid pathways. *Metab. Eng.* **2008**, *10* (6), 312-320. DOI: https://doi.org/10.1016/j.ymben.2008.08.001.

[3].Wang, J., Shen, X., Yuan, Q., and Yan, Y. Microbial synthesis of pyrogallol using genetically engineered Escherichia coli. *Metab. Eng.* **2018**, *45*, 134-141. DOI: https://doi.org/10.1016/j.ymben.2017.12.006 From NLM.

[4].Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R., and Keasling, J. D. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006**, *440* (7086), 940-943. DOI: https://doi.org/10.1038/nature04640.

[5].Lalwani, M. A., Zhao, E. M., and Avalos, J. L. Current and future modalities of dynamic control in metabolic engineering. *Curr. Opin. Biotechnol.* **2018**, *52*, 56-65. DOI: https://doi.org/10.1016/j.copbio.2018.02.007.

[6].Shen, X., Wang, J., Li, C., Yuan, Q., and Yan, Y. Dynamic gene expression engineering as a tool in pathway engineering. *Curr. Opin. Biotechnol.* **2019**, *59*, 122-129. DOI: https://doi.org/10.1016/j.copbio.2019.03.019.

[7].Holtz, W. J., and Keasling, J. D. Engineering Static and Dynamic Control of Synthetic Pathways. *Cell* **2010**, *140* (1), 19-23. DOI: https://doi.org/10.1016/j.cell.2009.12.029.

[8].Keasling, J. D. Manufacturing Molecules Through Metabolic Engineering. *Science* **2010**, *330* (6009), 1355-1358. DOI: https://doi.org/10.1126/science.1193990.

[9].Teng, Y., Zhang, J., Jiang, T., Zou, Y., Gong, X., and Yan, Y. Biosensor-enabled pathway optimization in metabolic engineering. *Curr. Opin. Biotechnol.* **2022**, *75*, 102696. DOI: https://doi.org/10.1016/j.copbio.2022.102696.

[10].Liu, D., Evans, T., and Zhang, F. Applications and advances of metabolite biosensors for metabolic engineering. *Metab. Eng.* **2015**, *31*, 35-43. DOI: https://doi.org/10.1016/j.ymben.2015.06.008.

[11].Tan, S. Z., and Prather, K. L. J. Dynamic pathway regulation: recent advances and methods of construction. *Curr. Opin. Biotechnol.* **2017**, *41*, 28-35. DOI: https://doi.org/10.1016/j.cbpa.2017.10.004.

[12].Zhu, Y., Li, Y., Xu, Y., Zhang, J., Ma, L., Qi, Q., and Wang, Q. Development of bifunctional biosensors for sensing and dynamic control of glycolysis flux in metabolic engineering. *Metab. Eng.* **2021**, *68*, 142-151. DOI: https://doi.org/10.1016/j.ymben.2021.09.011.

[13].Lalwani, M. A., Ip, S. S., Carrasco-López, C., Day, C., Zhao, E. M., Kawabe, H., and Avalos, J. L. Optogenetic control of the lac operon for bacterial chemical and protein

production. *Nat. Chem. Biol.* **2021**, *17* (1), 71-79. DOI: https://doi.org/10.1038/s41589-020-0639-1.

[14].Wei, L., Zhao, J., wang, Y., Gao, J., Du, M., zhang, Y., Xu, N., Du, H., Ju, J., Liu, Q., and Liu, J. Engineering of Corynebacterium glutamicum for high-level γ-aminobutyric acid production from glycerol by dynamic metabolic control. *Metab. Eng.* **2022**, *69*, 134-146. DOI: https://doi.org/10.1016/j.ymben.2021.11.010.

[15].Wang, X., Han, J.-N., Zhang, X., Ma, Y.-Y., Lin, Y., Wang, H., Li, D.-J., Zheng, T.-R., Wu, F.-Q., Ye, J.-W., and Chen, G.-Q. Reversible thermal regulation for bifunctional dynamic control of gene expression in Escherichia coli. *Nat. Commun.* **2021**, *12* (1), 1411. DOI: https://doi.org/10.1038/s41467-021-21654-x.

[16].Zhang, Y., Cortez, J. D., Hammer, S. K., Carrasco-López, C., García Echauri, S. Á., Wiggins, J. B., Wang, W., and Avalos, J. L. Biosensor for branched-chain amino acid metabolism in yeast and applications in isobutanol and isopentanol production. *Nat. Commun.* **2022**, *13* (1), 270. DOI: https://doi.org/10.1038/s41467-021-27852-x.

[17].Li, C., Jiang, T., Li, M., Zou, Y., and Yan, Y. Fine-tuning gene expression for improved biosynthesis of natural products: From transcriptional to post-translational regulation. *Biotechnol. Adv.* **2022**, *54*, 107853. DOI: https://doi.org/10.1016/j.biotechadv.2021.107853.

[18].Rogers, J. K., and Church, G. M. Genetically encoded sensors enable real-time observation of metabolite production. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113* (9), 2388-2393. DOI: https://doi.org/10.1073/pnas.1600375113.

[19].Ding, N., Zhou, S., and Deng, Y. Transcription-Factor-based Biosensor Engineering for Applications in Synthetic Biology. *ACS Synth. Biol.* **2021**, *10* (5), 911-922. DOI: https://doi.org/10.1021/acssynbio.0c00252.

[20].Mitchler, M. M., Garcia, J. M., Montero, N. E., and Williams, G. J. Transcription factor-based biosensors: a molecular-guided approach for natural product engineering. *Curr. Opin. Biotechnol.* **2021**, *69*, 172-181. DOI: https://doi.org/10.1016/j.copbio.2021.01.008.

[21].Nguyen, T. K., Tran, N. P., and Cavin, J. F. Genetic and biochemical analysis of PadR-padC promoter interactions during the phenolic acid stress response in Bacillus subtilis 168. *J. Bacteriol.* **2011**, *193* (16), 4180-4191. DOI: https://doi.org/10.1128/JB.00385-11.

[22].Park, S. C., Kwak, Y. M., Song, W. S., Hong, M., and Yoon, S.-i. Structural basis of effector and operator recognition by the phenolic acid-responsive transcriptional regulator PadR. *Nucleic Acids Res.* **2017**, *45* (22), 13080-13093. DOI: https://doi.org/10.1093/nar/gkx1055.

[23].Huang, Q., Lin, Y., and Yan, Y. Caffeic acid production enhancement by engineering a phenylalanine over-producing Escherichia coli strain. *Biotechnol. Bioeng.* **2013**, *110* (12), 3188-3196. DOI: https://doi.org/10.1002/bit.24988.

[24].Lin, Y., Sun, X., Yuan, Q., and Yan, Y. Combinatorial biosynthesis of plant-specific coumarins in bacteria. *Metab. Eng.* **2013**, *18*, 69-77. DOI: https://doi.org/10.1016/j.ymben.2013.04.004.

[25].Yang, Y., Lin, Y., Li, L., Linhardt, R. J., and Yan, Y. Regulating malonyl-CoA metabolism via synthetic antisense RNAs for enhanced biosynthesis of natural products. *Metab. Eng.* **2015**, *29*, 217-226. DOI: https://doi.org/10.1016/j.ymben.2015.03.018.

[26].Jiang, T., Li, C., and Yan, Y. Optimization of a p-Coumaric Acid Biosensor System for Versatile Dynamic Performance. *ACS Synth. Biol.* **2021**, *10* (1), 132-144. DOI: https://doi.org/10.1021/acssynbio.0c00500.

[27].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. Development of a Bacterial Biosensor for Rapid Screening of Yeast p-Coumaric Acid Production. *ACS Synth. Biol.* **2017**, *6* (10), 1860-1869. DOI: https://doi.org/10.1021/acssynbio.7b00009.

[28].Li, C., Zou, Y., Jiang, T., Zhang, J., and Yan, Y. Harnessing plasmid replication mechanism to enable dynamic control of gene copy in bacteria. *Metab. Eng.* **2022**, *70*, 67-78. DOI: https://doi.org/10.1016/j.ymben.2022.01.003.

[29].Jiang, T., Li, C., Zou, Y., Zhang, J., Gan, Q., and Yan, Y. Establishing an Autonomous Cascaded Artificial Dynamic (AutoCAD) regulation system for improved pathway performance. *Metab. Eng.* **2022**, *74*, 1-10. DOI: https://doi.org/10.1016/j.ymben.2022.08.009.

[30].Ye, J., McGinnis, S., and Madden, T. L. BLAST: improvements for better sequence analysis. *Nucleic Acids Res.* **2006**, *34* (Web Server issue), W6-9. DOI: https://doi.org/10.1093/nar/gkl164.

[31].Chiu, J., March, P. E., Lee, R., and Tillett, D. Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Res.* **2004**, *32* (21), e174-e174. DOI: https://doi.org/10.1093/nar/gnh172.

[32].Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A P., Rempfer, C., Bordoli, L., Lepore, R., and Schwede, T. SWISS-

MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296-W303. DOI: https://doi.org/10.1093/nar/gky427.

[33].Wang, J., Shen, X., Rey, J., Yuan, Q., and Yan, Y. Recent advances in microbial production of aromatic natural products and their derivatives. *Appl. Microbiol. Biotechnol.* **2018**, *102* (1), 47-61. DOI: https://doi.org/10.1007/s00253-017-8599-4.

[34].Zhang, R., Li, C., Wang, J., Yang, Y., and Yan, Y. Microbial production of small medicinal molecules and biologics: From nature to synthetic pathways. *Biotechnol. Adv.* **2018**, *36* (8), 2219-2231. DOI: https://doi.org/10.1016/j.biotechadv.2018.10.009.

[35].Lin, Y., Shen, X., Yuan, Q., and Yan, Y. Microbial biosynthesis of the anticoagulant precursor 4-hydroxycoumarin. *Nat. Commun.* **2013**, *4* (1), 2603. DOI: https://doi.org/10.1038/ncomms3603.

[36].Lutz, R., and Bujard, H. Independent and Tight Regulation of Transcriptional Units in Escherichia Coli Via the LacR/O, the TetR/O and AraC/I1-I2 Regulatory Elements. *Nucleic Acids Res.* **1997**, *25* (6), 1203-1210. DOI: https://doi.org/10.1093/nar/25.6.1203.