# Muconic Acid Production Using Gene-Level Fusion Proteins in Escherichia coli

Fujiwara, Ryosuke
Noda, Shuhei
Tanaka, Tsutomu
Kondo, Akihiko

(Citation)
ACS Synthetic Biology,7(11):2698-2705

(Issue Date)
2018-11

(Resource Type)
journal article

(Version)
Accepted Manuscript

# Muconic Acid Production Using Gene-level Fusion Proteins in *Escherichia coli*

Ryosuke Fujiwara[†], Shuhei Noda[*‡], Tsutomu Tanaka[*†], Akihiko Kondo[§]

[†] Department of Chemical Science and Engineering, Graduate School of Engineering, Kobe University

[‡] Center for Sustainable Resource Science, RIKEN

[§] Graduate School of Science, Technology and Innovation, Kobe University

ABSTRACT

This study reports on the improving of Muconic acid (MA) production by using metabolically engineered *Escherichia coli*. Three MA synthesis pathways separately were introduced into *E. coli*. After 72 h of cultivation, two of the three strains i.e., one carrying the Pathway 1 (1.00 g/L) and the other carrying the Pathway 3 (1.34 g/L) produced MA. To increase MA production, the enzymes of the shikimate pathway (AroC and AroD) were overexpressed in these strains. Although the overexpression of AroC increased the MA production (1.59 g/L) by the Pathway 1, AroD overexpression decreased it by the Pathway 3. The metabolic channeling using gene-level fusion proteins additionally increased the MA production. In the pathway 1 and pH-controlled cultures, the overexpression of a fusion protein (AroC and MenF) increased the MA production from 20 g/L glucose to >3 and 4.45 g/L, respectively. These results suggest that the metabolic channeling approach is promising strategy to increase the yield of the target compound.

Keywords: Muconic acid, Shikimate pathway, Fusion protein, *Escherichia coli*, Metabolic engineering

For Table of Contents Use Only

Manuscript title: Muconic Acid Production Using Gene-level Fusion Proteins in *Escherichia coli*

Authors: Ryosuke Fujiwara, Shuhei Noda, Tsutomu Tanaka, Akihiko Kondo

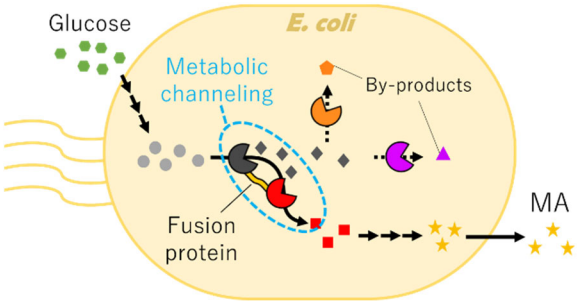



Table of Contents/Abstract Graphics

The global market for adipic acid, an important aliphatic dicarboxylic acid, is estimated to reach £8 billion by 2022.[1,2] More than half of the total adipic acid production is from cyclohexane and the remainder from cyclohexene, uncoupled cyclohexanol/cyclohexanone (KA oil), and phenol.[1] The production of nylon 66 fibers and engineering resins accounted for approximately 57% of the total amount of adipic acid consumed in 2016.[1] Recently, microbial production of adipic acid has attracted attention as a solution to the exhaustion of finite fossil resources along with the increasing global demand of adipic acid. The production of bio-adipic acid, wherein carbons are derived from a renewable feedstock, may solve these problems.[3] Yu et al. prepared 639 ± 34 μg/L of adipic acid from 10 g/L glucose via an artificial pathway with acetyl-CoA and succinyl-CoA as starting units (ACSC1 pathway) in *Escherichia coli.*[4] Cheong et al. created an *E. coli* strain producing adipic acid at higher amount by introducing another artificial pathway with same starting units of ACSC1 pathway (ACSC2 pathway). Adipyl-CoA, an intermediate of these two artificial pathways, occurs via three steps from succinyl-CoA and acetyl-CoA in the same way in these two pathways. In ACSC2 pathway, adipyl-CoA was converted to adipic acid by acyl-CoA thioesterases, whereas by phosphate butyryltransferase and butyryl kinase in ACSC1 pathway. This strain produced 2.5 g/L adipic acid from glycerol at a molar yield of

4.1%.[5] However, these findings showed that these pathways is not suitable for large-scale production of adipic acid from renewable resources, such as glucose.

In this study, we investigated the bio-production of muconic acid (*cis,cis*-muconic acid (MA) using the gene-level fusion proteins method. In nature, MA is an intermediate in the degradation pathway of various aromatic compounds in some bacteria, including *Pseudomonas putida* and *Klebsiella pneumoniae*.[6,7] Normally, adipic acid is easily synthesized from MA through double-bond hydrogenation.[8] MA is produced through microbial catalysis and is successfully converted to adipic acid by chemical hydrogenation in the presence of suitable catalysts.[9] MA is also used as a precursor in the production of terephthalic acid (TPA), the main monomer in the production of polyethylene terephthalate, an important component of industrial resins and thermoplastic synthetic fibers (Figure 1).[10,11] After MA isomerization from *cis,cis*-muconic acid to *trans,trans*-muconic acid, it follows a Diels–Alder reaction with ethylene and is further dehydrogenated to produce TPA. Given its versatility as a precursor, several laboratories investigated the fermentative production of MA using microbes.[12–15]When using *E. coli* as a host, 3-dehydroshikimate (DHS) and chorismate (intermediates or derivatives of the shikimate pathway) are used as starting materials for MA production (Figure 2A).[9,12,14-16] Therefore, increasing the activity of the shikimate

pathway is an important strategy to enhance MA yield. The first reaction involved in the shikimate pathway is the stereospecific condensation of erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP) to generate 3-deoxy-D-heptulosonate-7-phosphate (DAHP), catalyzed by DAHP synthase. DHS is synthesized via a 2-step reaction from DAHP. Chorismate, a key intermediate in the shikimate pathway, is synthesized downstream of DHS and can be converted to various aromatic compounds, including aromatic amino acids.[16]

Several synthetic pathways for MA production have been reported using *E. coli* as a host strain (Figure 2A). Lin et al. created an MA synthetic pathway via salicylate, which is converted to MA via a sequential reaction pathway involving catechol transformation in reactions catalyzed by salicylate 1-monoxygenase and catechol 1,2-dioxygenase (Figure 2A, Pathway 1).[12] Sengupta et al. used another pathway for MA production via 4-hydroxybenzoate (PHB) and protocatechuate (PCA) intermediates.[14] After hydroxylation of 4-hydroxybenzoate, a derivative of chorismate, the PCA product, was converted into MA via catechol by both PCA decarboxylase and catechol 1,2-dioxygenase (Figure 2A, Pathway 2). Niu et al. engineered *E. coli* by expressing the DHS dehydratase, PCA decarboxylase, and catechol 1,2-dioxygenase,[9] and creating a branching pathway from the intermediate DHS in the shikimate pathway

(Figure 2A, Pathway 3).

Previously, we engineered a metabolic pathway in *E. coli* to increase the activity of the shikimate pathway by accumulating intracellular PEP and to produce chorismate derivatives in high yields[15]. To prevent PEP consumption during glucose uptake, we generated a recombinant strain by replacing the endogenous phosphotransferase system (PTS) with the galactose permease (GalP)/glucokinase (Glk) system (GGS) ($PTS^-$ $GGS^+$ strain). The PTS consumes 1 mol of PEP for 1 mol of glucose while the GGS needs ATP as the cofactor for glucose phosphorylation without PEP consumption. Two genes encoding pyruvate kinase (*pykA, pykF*) from the $PTS^-GGS^+$ strain were disrupted to prevent the introduction of carbon-based compounds to the TCA cycle. As pyruvate kinase catalyzes the conversion of PEP to pyruvate in the glycolytic pathway, the disruption of the genes encoding these enzymes will deprive the engineered *E. coli* of pyruvate. In our platform strains, the lack of pyruvate was compensated by introducing an artificial pathway that produces chorismate derivatives, such as salicylate, *p*-hydroxybenzoate, 3-hydroxybenzoate, 2-aminobenzoate, *p*-aminobenzoate, phenol, and MA along with pyruvate.[15] For example, using the salicylate biosynthesis pathway, the chorismate is converted to isochorismate by isochorismate synthase (MenF). Then, the isochorismate is cleaved by isochorismate

pyruvate lyase to release pyruvate and salicylate (Figure 1).

Fusion protein enzymes are attracting attention as an efficient strategy to selectively increase the carbon flux to the target pathway while decreasing that to the competitive pathway without completely inactivating the genes. One strategy used to express fusion proteins is to bind the enzymes after translation by transpeptidase. Matsumoto et al. increased acetate production in recombinant *E.coli* synthesizing a fusion protein of pyruvate-formate lyase and phosphate acetyltransferase ligated by *Staphylococcal* sortase A.[17] Engineering of the fusion protein at the gene level is a simpler method of protein expression. Wang et al. increased the production of α-farnesene in recombinant *E. coli* by expressing a fusion protein of farnesyl diphosphate synthase and α-farnesene synthase.[18] In these studies, the metabolic channeling effectively increased the target compound production. Fusion enzymes behave in the same way as the native multi-domain enzymes (MDE). MDEs catalyze a series of reactions while preventing the diffusion of intermediate metabolites, thereby improving the efficiency of a metabolic reaction cascade by generating a substrate channeling from an active site to another.[17,19] As the MDE, reaction intermediates produced by fusion enzymes do not diffuse, increasing the efficiency of the subsequent reaction. Hence, the developing of the artificial metabolic channeling by fusion proteins is a powerful tool

for increasing the carbon flux to a specific pathway.

This study investigates the effect of overexpressing fusion proteins in the shikimate pathway using the gene-level fusion method on MA yield. Hence, the MA synthesis pathway was introduced into a strain with increased shikimate pathway flux, as described above. After selecting the optimal MA synthesis pathways from three candidates, the production of MA by the selected pathways was increased by overexpressing chorismate synthase (AroC). Finally, the effect of the gene-level fusion method was investigated. The production of MA was further increased by overexpressing gene-level fusion proteins of AroC and MenF.

## RESULTS AND DISCUSSION

### Selection of the candidate pathways for the MA production

Several previous reports have described methods for increasing MA production among them, the introduction of several MA synthesis pathways.[9,12,14] In this study, we tested three pathways for MA production using the CFT5 strain.[15] Two of the three pathways start from chorismate, a key compound in the shikimate pathway. The Pathway 1 produces MA via salicylate, whereas the Pathway 2 produces MA via PHB (Figure 2A). The Pathway 3 starts with DHS, another intermediate in the shikimate pathway, and produces MA via PCA (Figure 2A). In all three pathways, the catechol is produced from salicylate or PCA, followed by cleavage of the catechol benzene ring to form MA. In Pathway 1, the order of the genes was tested, and choose suited to CFT5 (Supporting information Table S3).

We engineered three *E. coli* transformants, CFT51a, CFT52a, and CFT53a (carrying Pathway 1, Pathway 2, and Pathway 3, respectively). Figure 3 shows the amount of MA produced by each transformant after 72 h cultivation. CFT51a and CFT53a produced 1.00 ± 0.11 and 1.34 ± 0.07 g/L of MA, respectively. Although CFT52a produced 0.40 ± 0.02 g/L of MA after 24 h cultivation, the amount produced decreased to the detection limit after 72 h cultivation. No difference in cell growth was

observed between strains, and all strains consumed glucose within 48 h of cultivation (Supplementary Figure S1, Figure S2). The intermediates of the MA synthesis pathway (salicylate [Pathway 1], PCA [Pathway 2], and catechol [Pathways 1, 2]) were not detected in the culture supernatants of CFT51a or CFT52a after 72 h of cultivation.

Although high PHB production was previously observed in CFT5-derived strain (1820 mg/L after 72 h of cultivation),[15] the CFT52a produced only a small amount of MA. CFT52a cultures contained no detectable PCA or catechol, and only a small amount of PHB was detected (0.108 ± 0.00 g/L after 72 h of cultivation). The Pathways 2 and 3 both include the reactions catalyzed by AroY and CatA. Because CFT53a, carrying Pathway3, produced the highest amount of MA among the three strains, these reactions were considered as optimum in CFT5-derived strains. In the SDS-PAGE analysis of CFT52a strain, the significant overexpression of PobA and UbiC was not detected (Supplementary Figure S3). In order to produce MA at high yield in pathway 2, it would be necessary to solve these expression problems. Hence, we conclude that the Pathway 2 is not suitable for MA production in CFT5-derived strains.

**Optimizing the production pathway of the MA**

To further increase the MA production we approached the intracellular

174 availability of the compound at the branch point between the endogenous shikimate and
175 synthetic MA pathway. The branch point compounds are chorismate in CFT51a
176 (Pathway 1) and DHS in CFT53a (Pathway 3). Since no accumulation of intermediates
177 was detected in the culture supernatants of CFT51a or CFT53a, the Pathways 1 and 3
178 were not bottlenecks in these strains. Improved availability of chorismate or DHS is
179 essential for further increasing MA production. Overexpression of AroC in CFT51a and
180 AroD in CFT53a was used to increase the carbon flux to chorismate and DHS
181 respectively. Figure 3 shows the amount of produced MA by each transformant after 72
182 h cultivation. CFT51b produced 1.59 ± 0.10 g/L of MA, which is 1.6-fold higher than
183 that produced by CFT51a. However, the amount of MA produced by CFT53b was
184 below the quantification limit. Lütke-Eversloh et al. investigated the key enzymes for
185 tyrosine production, reporting that overexpression of AroC is an effective way to
186 increase the production of compounds in the shikimate pathway.[20] This report is
187 consistent with our finding that showed that the MA production was effectively
188 increased by the overexpression of AroC. However, contrary to our expectations, the
189 overexpression of AroD had a negative effect on the MA production in CFT53b
190 (Pathway 3). The overexpression of AroD might increase the carbon flux downstream of
191 the shikimate pathway rather than that to PCA, the precursor of MA.

192

193 **Expression of fusion proteins to further increase MA yield**

194 As already stated, the metabolic channeling is a powerful strategy to enhance
195 the carbon flux to specific compounds of interest, by promoting the efficient transfer of
196 an intermediate product from an enzyme to another.[21,22] Hence, this strategy was applied
197 in our study for the MA biosynthesis aiming to increase its production (Figure 4). In
198 Pathways 1 and 3, the two enzymes at the branch point between the shikimate and MA
199 biosynthesis pathways were overexpressed as a fusion protein, AroC and MenF (AroC-
200 MenF) in Pathway 1 and AroD and AroZ (AroD-AroZ) in Pathway 3 (Figure 2A). First
201 we examined the effect of linker type on MA production. In AroC-MenF and AroD-
202 AroZ, MA production was compared in the case of using without linker, flexible linker,
203 and rigid linker (Supplementary Figure S4), respectively, and flexible linker was most
204 suitable for AroC-MenF and AroD-AroZ (Supplementary Table S4). Figure 3 displays
205 the amount of MA produced by each transformant after 72 h cultivation. CFT51c
206 (AroC-MenF) and CFT53c (AroD-AroZ) produced 3.45 ± 0.04 and 1.20 ± 0.10 g/L of
207 MA, respectively. The production of MA in CFT51c was 2.16-fold higher than that
208 produced by CFT51b, which expresses the unfused AroC and MenF proteins. We
209 confirmed that CFT51b and CFT51c expressed MenF and fused AroC/MenF fusion

protein respectively by western blotting analysis (supplementary Figure S5). In the AroC-MenF fusion protein, the two enzymes are placed in close proximity, thereby increasing the apparent concentration of the intermediate (chorismate) around MenF relative to that of the non-fusion enzyme, resulting in higher carbon flux to the MA biosynthesis pathway (Figure 4). If the competing pathway is branched from the synthetic pathway of the target compound, metabolic channeling by the fusion protein has the potential to significantly increase production. However, for the AroD-AroZ fusion protein, no improvement of the MA production was noticed. For CFT53b, in which AroD and AroZ were overexpressed separately, the MA production was drastically decreased than that of CFT53a (not overexpressing AroD) (Figure 3). In SDS-PAGE analysis of CFT53b, free AroD was significantly overexpressed in this strain (Supplementary Figure S6). One explanation could be related to the fact that the overexpression of AroD shifted the carbon flux in the shikimate pathway more effectively than to PCA. The expression of AroZ, AroY and CatA were not distinctively confirmed in CFT53b. The imbalance of expression levels of these enzymes might have a negative effect on MA production. On the other hand, in SDS-PAGE analysis of CFT53c, it was not distinctively confirmed to be overexpressed the free AroD and the fused AroD/AroZ (Supplementary Figure S6). Since the expression level of the

AroD/AroZ was greatly reduced as compared with AroD, the imbalance would have been eliminated and the production of MA was restored. However, the production titer of CFT53c was lower than CFT53a. It seems that metabolic channeling between AroD and AroZ would have not work well because of the inactivation of the AroD domain in the fusion protein. The native AroD protein is expressed as a homodimer in *E.coli.*[23] In CFT53c, the AroD may not dimerize properly, because it is fused to AroZ. In this case, only the endogenous AroD would function as a 3-dehydroquinate dehydratase, as in CFT53a. This hypothesis may explain why MA production was nearly the same in CFT53a and CFT53c. Thus, it is difficult to preserve the function of proteins with more complex structures, such as dimeric AroD, as fusion proteins. Another strategy for metabolic channeling involves the protein-level fusion after translation using transpeptidases.[17]

The protein fusion method has been used to increase the production of target compounds, as well. Matsumoto et al. created recombinant *E. coli* expressing a fusion protein of pyruvate-formate lyase (PFL) and phosphate acetyltransferase (PTA) ligated by *Staphylococcal* sortase A.[17] This strain produced a significantly higher amount of acetate as compared to a control strain expressing PFL and PTA separately. The Sortase-A–mediated protein fusion method was effective for *in vivo* metabolic channeling.

Hence, this approach seems to be effective to produce AroZ-AroD fusion proteins.

**pH-controlled tube-scale culture using $CaCO_3$**

We attempted increasing the MA production titer by cotrolling pH. The decrease in pH during culturing often has a bad influence on bioproduction. Although *E. coli* grows at pH 5-8, the environment inside the bacteria also changes due to the change in external pH[24]. On the acidic side, when the pH falls below 6, the intracellular pH decreases[25]. After culturing CFT51c for 72 hours, the pH in the medium decreased to 5.51 ± 0.0. Thompson et al. introduced multiple MA synthesis pathways into *E. coli* that produced 3153 ± 149 mg/L of MA.[26] This is the highest titer so far using microbes in batch culture, without pH-controlled conditions. The *E. coli* transformant constructed in the present study also produced 3.45 g/L of MA. According to these results, it is possible that the MA produced in CFT51c reached the limit in non-pH-control. Thus, we hypothesized that further MA production can be increased by pH control. To control the pH during cultivation, $CaCO_3$ (10 g/L) was added to the culture medium after 24 h of cultivation. Figure 5C illustrates the amount of MA produced by CFT51c cultured in a medium containing $CaCO_3$. CFT51c produced 4.45 ± 0.12 g/L of MA after 72 h of cultivation, meaning a 1.29-fold increase over that without $CaCO_3$, and the maximum

production rate reached 2.77 g/L/day (24 h). After culturing CFT51c for 72 hours with $CaCO_3$, the pH in the medium was maintained at 7.05 ± 1.4.

**Jar fermentor cultivation**

In order to perform cultivation more precisely while controlling the pH, CFT51c was cultured with a Jar fermentor. Figure 6 indicates the cell growth, MA production, and glucose consumption in jar fermentor cultivation. CFT51c produced 4.55 g/L of MA after 72 h of cultivation and maximum production rate was 2.20g/L/day at 48h cultivation. These results correspond with the culture profiles of pH-control test tube cultivation with $CaCO_3$. The results of test tube culture (with $CaCO_3$) and jar fermentor culture indicate that pH adjustment is important for MA production at this stage. In order to further increase the production, it will be necessary to examine the appropriate dissolved oxygen concentration and to handle the increase of osmotic pressure due to MA and by-products accumulation.

## CONCLUSIONS

In summary, three MA synthesis pathways were compared aiming to select the most suitable for MA production in CFT5-derived strains. The three initial strains were greatly different from each other in terms of MA production. Several possible pathways are available to produce a compound of interest by using microorganisms. However, as our results showed, the selection of the optimal pathway among the available ones is essential in the perspective of the industrial applications that need the optimization of the yield of a target compound.

To obtain a high yield of MA, we applied the fusion protein method. Hence, the expressed fusion protein directed the carbon flux to the MA synthesis in the shikimate pathway, resulting in a production of MA of 4.45 g/L with a maximum production rate of 2.77 g/L/day. The MA yield from glucose was 207 mg/g, which is an increase of 4.44-fold as compared to that of CFT51a, used as a control strain. The production titer and rate of MA production using glucose as the carbon source in CFT51c were higher than those in all previous studies using batch cultures in flasks or test tubes. This result suggests that metabolic channeling using the gene-level fusion proteins is a powerful tool to direct the carbon flux to the specific pathways selectively leading to the desired product. However, the gene-level fusion method is difficult to

298 apply to proteins with more complex structures, such as multimeric proteins. To further
299 increase the production titer and yield would require to combine multiple approaches,
300 such as gene disruption, metabolite funneling, and fusion proteins.
301

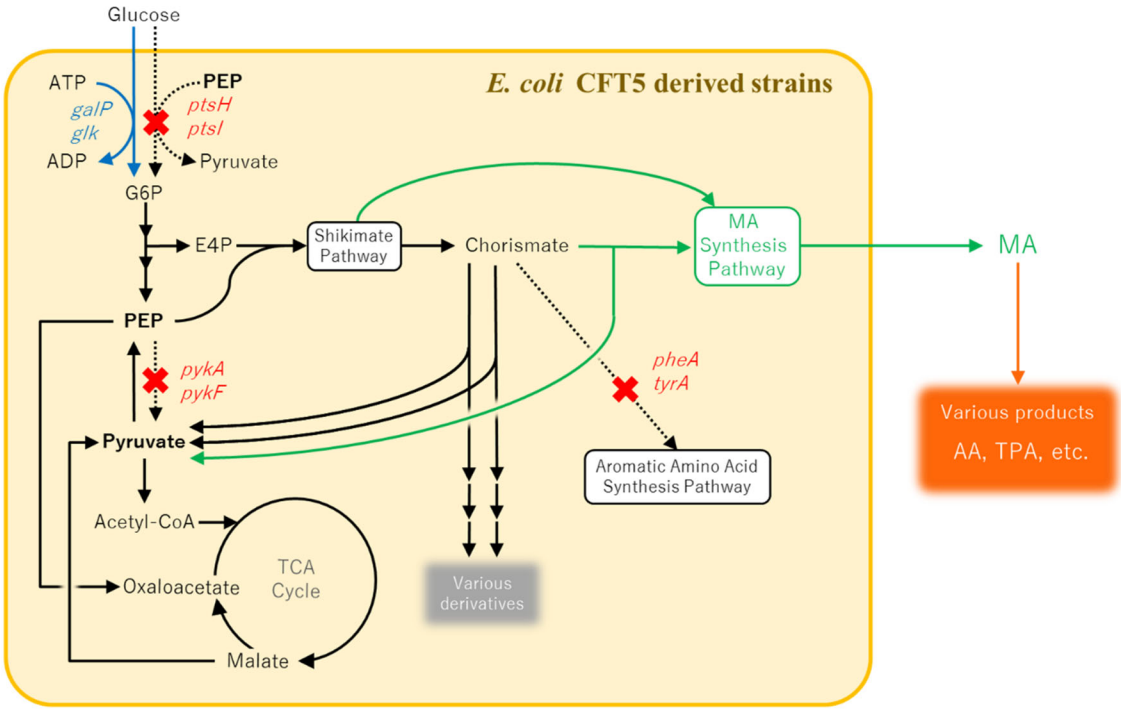
302 Table. 1 Bacterial strains and plasmids used in this study

| | Genotype | Source or reference |
|---|---|---|
| Strains | | |
| Nova Blue | *endA1 hsdR17(rK12-mK12$^{+}$) supE44 thi-I gyrA96 relA1 lac recA1/F' [proAB$^{+}$ lacIq ZΔM15::Tn10(Tet r)];* used for gene cloning. | Novagen |
| ATCC31882 | L-Phenylalanine-overproducing strain | ATCC |
| CFT5 | ATCC31882 *ptsHI::P $_{A1lacO-1}$ -glk-galP* Δ*pykF*Δ*pykA* Δ*pheA* Δ*tyr* | Noda et al.[15] |
| CFT51a | CFT5 harboring pZA23-ncmI | This study |
| CFT51b | CFT5 harboring pZA23-ncamI | This study |
| CFT51c | CFT5 harboring pZA23-nca/mI | This study |
| CFT52a | CFT5 harboring pZA23-UpYc | This study |
| CFT53a | CFT5 harboring pZA23-ZYc | This study |
| CFT53b | CFT5 harboring pZA23-DZYc | This study |
| CFT53c | CFT5 harboring pZA23-D/ZYc | This study |
| Plasmids | | |

| | | |
|---|---|---|
| pZE12MCS | $P_{LlacO1}$, *colE ori, Amp*$^r$ | Expressys |
| pZA23MCS | $P_{A1lacO-1}$, *p15A ori, Km*$^r$ | Expressys |
| pZA23-ncmI | pZA23MCS containing *nahG, catA, menF,* and *pchB* | This study |
| ZA23-ncamI | pZA23-ncmI containing *aroC* | This study |
| pZA23-nca/mI | pZA23MCS containing *nahG, catA, pchB,* and *aroC/menF,* which is a gene expressing fusion protein composed of AroC and MenF. | This study |
| pZA23-UpYc | pZA23MCS containing *ubiC* from *E. coli*, *pobA* from *P. putida* KT2440(codon optimized for *E.coli*), *aroY* from *K. pneumoniae* (codon optimized for *E.coli*), and *catA* | This study |
| pZA23-ZYc | pZA23MCS containing *aroZ* from *Bacillus thuringiensis* (codon optimized for *E.coli*), *aroY,* and *catA* | This study |
| pZA23-DZYc | pZA23-ZYc containing *aroD* from *E.coli* | This study |
| pZA23-D/ZYc | pZA23MCS containing *aroY*, catA, and AroD-AroZ, which is a gene expressing fusion protein composed of | This study |

AroD and AroZ.

303




304

305 Figure 1. The synthetic metabolic pathway for the production of chorismate derivatives,

306 including MA, in *E. coli* CFT5-derived strains. The red X indicates deletion of the *ptsH*,

307 *ptsI*, *pykA*, *pykF*, *pheA*, and *tyrA* genes. The blue arrows indicate the GalP/Glk system

308 (GGS) revised by introducing *galP* and *glk* genes. The green arrows indicate the MA

309 synthesis pathways introduced. The orange arrow indicates the conversion of MA to

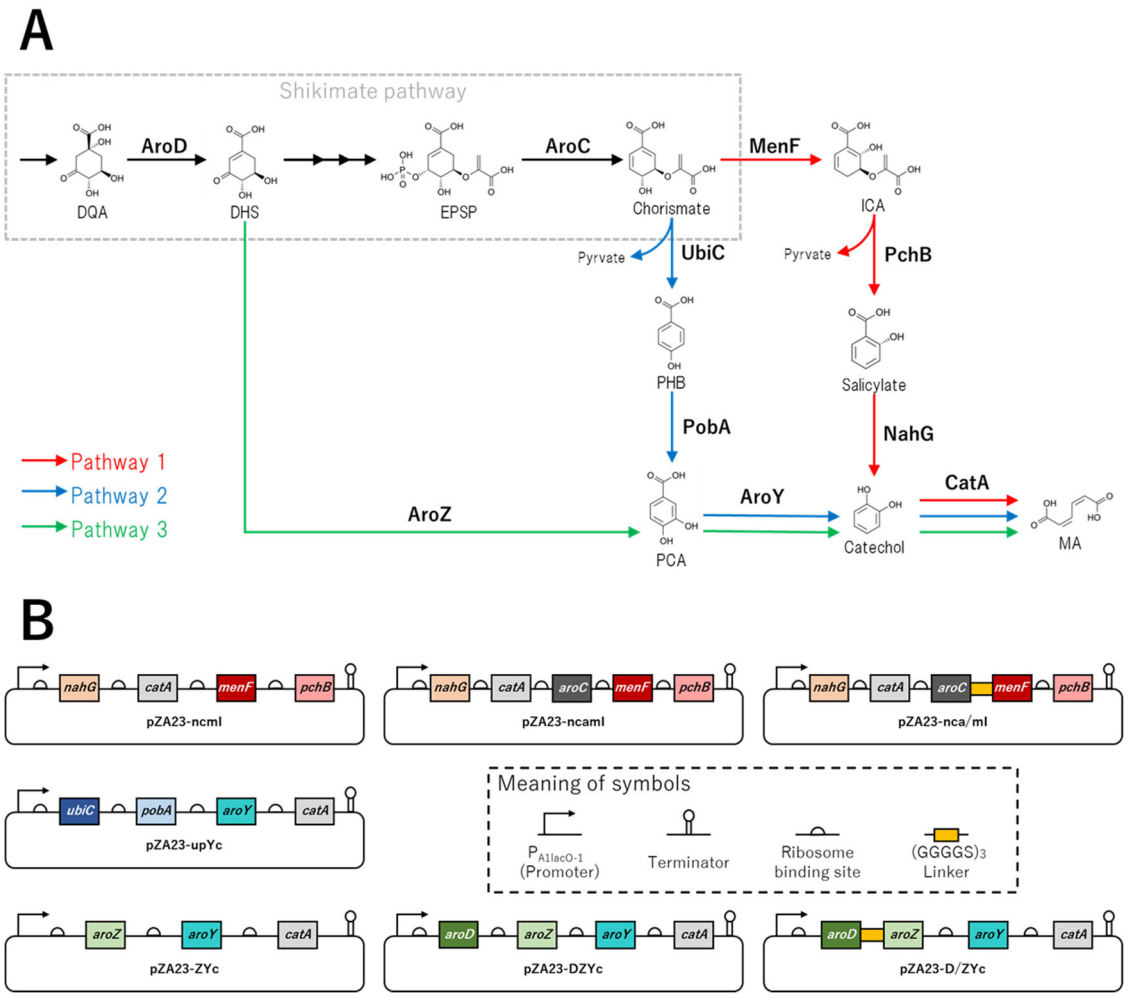
310 various products by chemical catalysis. GalP - galactose permease, Glk - glucokinase,

311 G6P - glucose 6-phosphate, ATP - adenosine triphosphate, ADP - adenosine

312 diphosphate, G6P - glucose-6-phosphate, E4P - erythrose-4-phosphate, PEP -

313 phosphoenolpyruvate, Acetyl-CoA - acetyl-coenzyme A, AA - adipic acid, TPA-

314 terephthalic acid.




315

316 Figure 2. (A) Diagram of MA synthesis pathways in CFT5-derived strains. DQA -

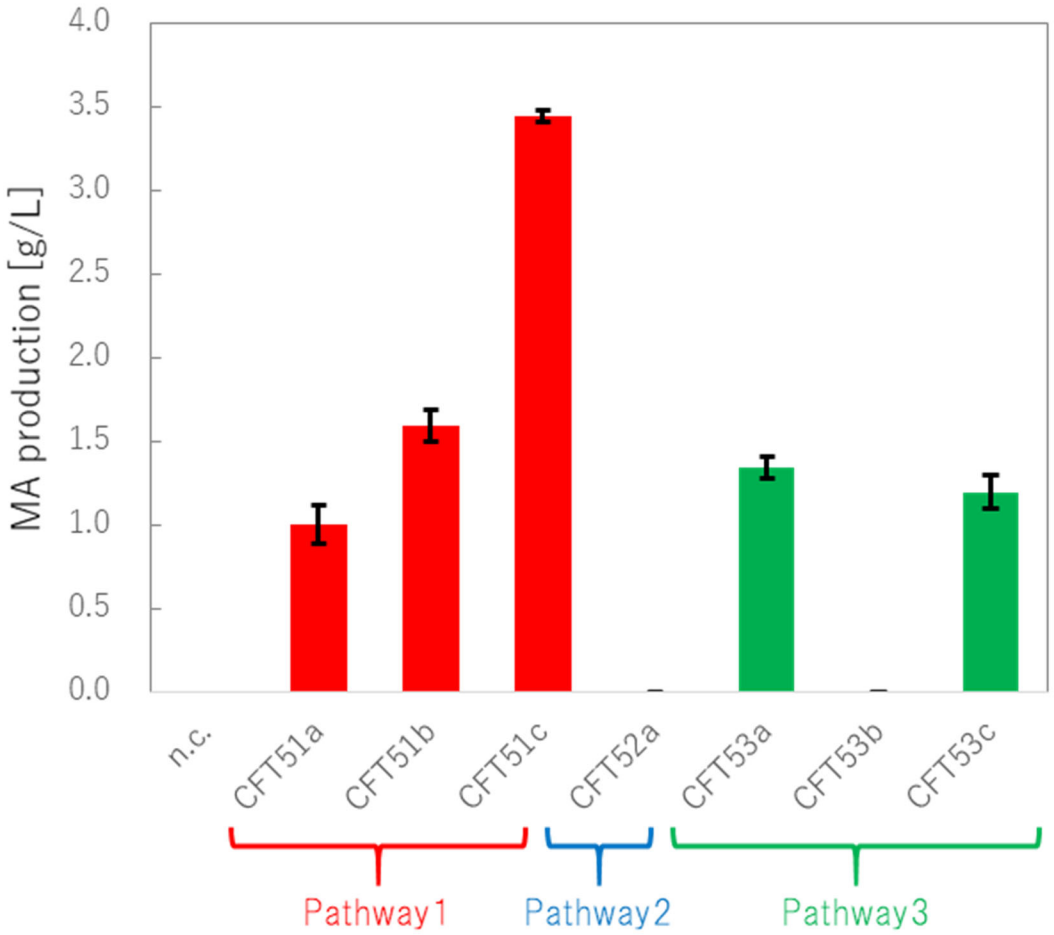
317 dehydroquinate, DHS - dehydroshikimate, EPSP - 5-O-(1-Carboxyvinyl)-3-

318 phosphoshikimate, PHB - *p*-hydroxbenzoate, ICA - isochorismate, PCA - protocatechuate,

319 Red arrows - Pathway 1, blue arrows - Pathway 2, and green arrows - Pathway 3. (B)

320 Diagram of MA synthesis gene cassette in each plasmid.

321




322

323 Figure 3. MA production titer in CFT5-derived strains. Red, blue, and green bars

324 indicate MA production via Pathway 1, Pathway 2, and Pathway3, respectively. n.c. is

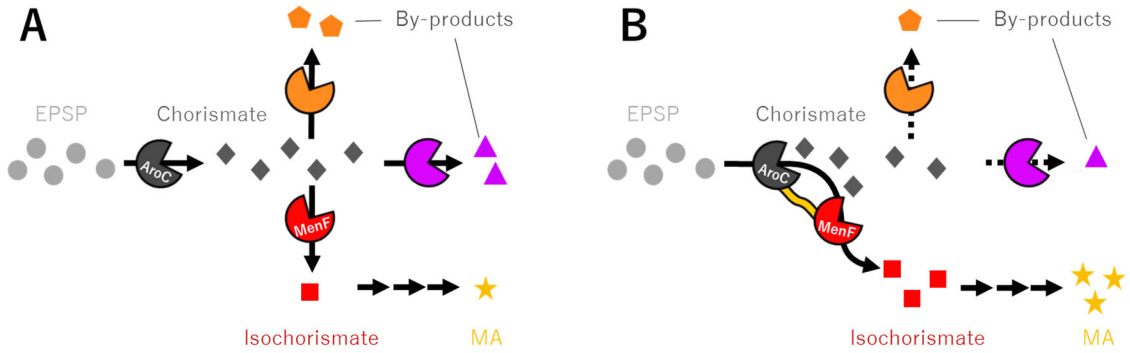
325 empty vector control (CFT5 containing pZA23MCS). The data are presented as the

326 average of three independent experiments, and error bars indicate the standard error.

327

328

329




330

331 Figure 4. Effect of AroC-MenF fusion protein expression on MA production. (A)

332 Diagram of the shikimate pathway in CFT51b. (B) Diagram of shikimate pathway in

333 CFT51c. Black and red Pac-man shapes indicate the AroC and MenF, respectively.

334 Orange and purple Pac-man shapes indicate the enzymes consuming chorismate as

335 substrate. Yellow band indicates the flexible $(GGGGS)_3$ linker. Light gray circles - 5-O-

336 (1-Carboxyvinyl)-3-phosphoshikimate (EPSP), dark gray diamonds - chorismate, red

337 squares - isochorismate, and yellow stars - MA. Orange pentagons and purple triangles

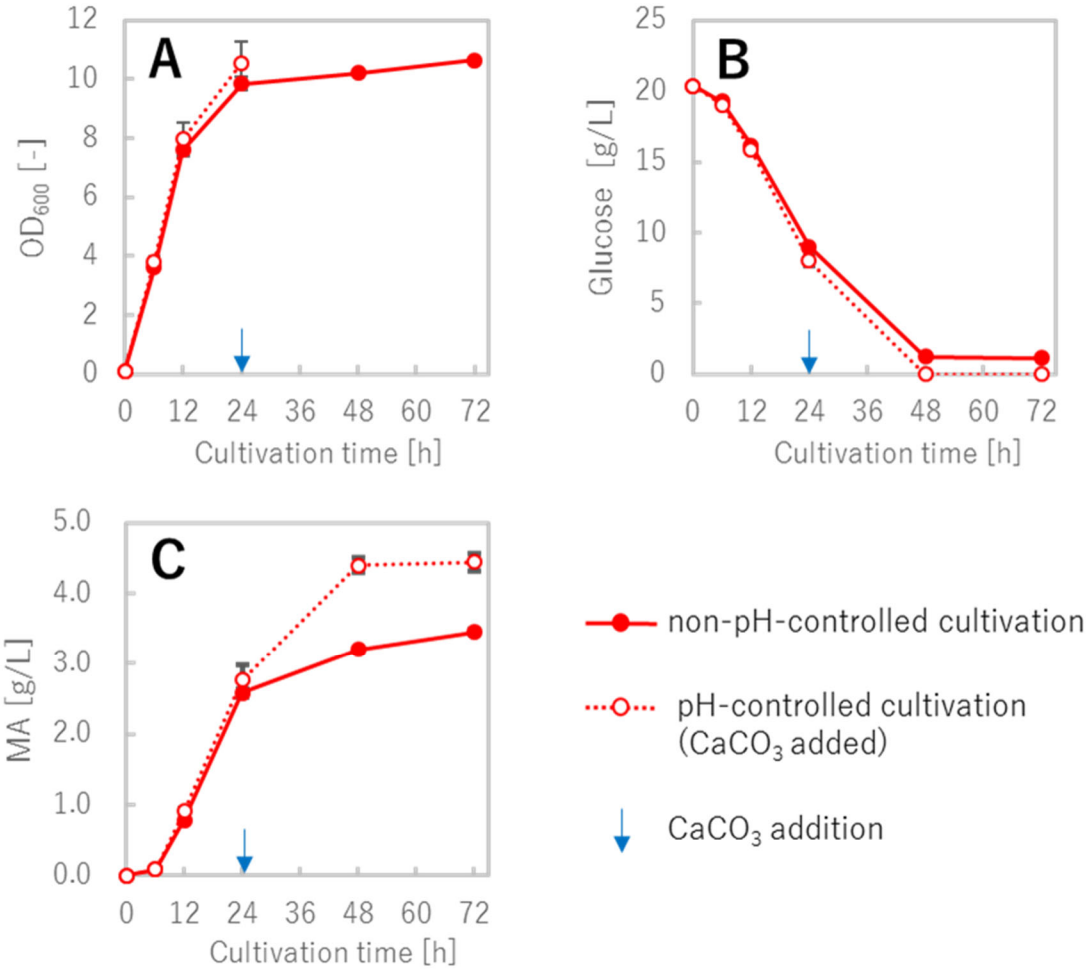
338 indicate the byproducts derived from chorismate.

339

340

341

342




343

344 Figure 5. Culture profiles of CFT51c. Red closed circles - non–pH-controlled

345 cultivation, red open circles - pH-controlled cultivation ($CaCO_3$ added). (A) Bacterial

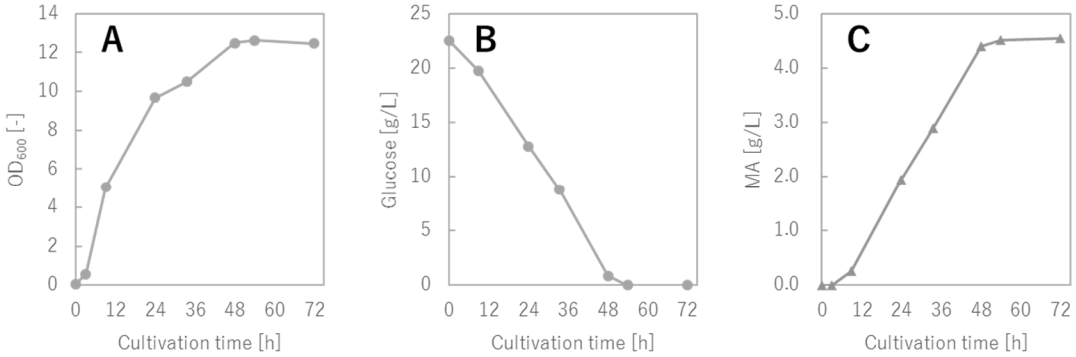
346 cell growth (CFT51c). (B) Glucose consumption by CFT51c. (C) MA production by

347 CFT51c. Blue arrows indicate when $CaCO_3$ was added in to the culture medium. The

348 data are presented as the average of three independent experiments, and error bars

349 indicate the standard error.

350




351

352 Figure 6. Culture profiles of CFT51c with jar fermentor. (A) Bacterial cell growth

353 (CFT51c). (B) Glucose consumption by CFT51c. (C) MA production by CFT51c.

354

355

356

357

358

359

360 METHODS

361 **Strains and plasmid construction**

362 The strains and plasmids used in this study are listed in Table 1. *E. coli* NovaBlue

363 competent cells (Novagen, Cambridge, MA, USA) were used for gene cloning.

364 Polymerase chain reaction was performed using Quick Taq HS (TOYOBO, Osaka,

365 Japan). Custom DNA oligonucleotide primers were synthesized by Invitrogen Custom

366 DNA Oligos (Thermo Fisher Scientific, Tokyo, Japan) (Supplementary Table S2).

367 Codon optimized exogenous genes fragments (*pobA* from *P. putida* KT2440, *aroY* from

368 *K. pneumoniae*, and *aroZ* from *Bacillus thuringiensis*) were synthesized by the

369 Invitrogen GeneArt Gene Synthesis service (Thermo Fisher Scientific, Tokyo, Japan).

370 Sequences of primers and exogenous genes fragments and detailed plasmid construction

371 methods are described in Supporting Information.

372

373 **Fusion protein construction**

374 The two proteins were expressed as one fusion protein using the following method. A

375 flexible $(GGGGS)_3$ linker was inserted between the C-terminal of the upstream protein

376 (AroC or AroD) and the N-terminal of the downstream protein (MenF or AroZ) i.e.,

377 (upstream protein C-terminal)-GGGGSGGGGSGGGGS-(downstream protein N-

terminal). A plasmid expressing the fusion protein was constructed, as follows. The linker sequence was inserted between the 3'end of the upstream gene (*aroC* or *aroD*) and the 5' end of the downstream gene (*menF* or *aroZ*), and the stop codon of the upstream gene was deleted, i.e., (3'end of upstream gene)-GGTGGTGGTGGTAGTGGTGGCGGTGGTTCAGGCGGTGGTGGTTCC-(5' end of downstream gene). The detailed method used to construct plasmids expressing the fusion proteins is described in Supporting Information.

**Transformation of CFT5-derived strains**

The transformation of CFT5-derived strains was carried out by electroporation with a 1350 kV, 600 Ω, 10 μF electric pulse in a 0.1 cm cuvette using a Gene Pulser (Bio-Rad Laboratories, Hercules, CA, USA).

CFT5 harboring pZA23-ncmI, pZA23-ncamI, pZA23-nca/mI, pZA23-UpYc, pZA23-ZYc, pZA23-DZYc, and pZA23-D/ZYc were designated CFT51a, CFT51b, CFT51c, CFT52a, CFT53a, CFT53b, and CFT53c, respectively. CFT51c harboring pSAK-tktA was designated CFT51ct.

**Media**

M9YP medium was used for MA production in 5-mL test tube-scale cultures. M9YP medium comprised M9 minimal medium supplemented with 0.5% yeast extract and 10 mM sodium pyruvate. M9 minimal medium contains (per liter) 20 g glucose, 0.5 g NaCl, 17.1 g $Na_2HPO_4 \cdot 12H_2O$, 3 g $KH_2PO_4$, 1g $NH_4Cl$, 246 mg $MgSO_4 \cdot 7H_2O$, 14.7 mg $CaCl_2 \cdot 2H_2O$, 2.78 mg $FeSO_4 \cdot 7H_2O$, 10 mg thiamine hydrochloride, 40 mg L-tyrosine, 40 mg L-tryptophan, and 100 mg L-phenylalanine (Tyr and Trp were included because ATCC31882 is auxotrophic for these amino acids, and the CFT5 derivative strains are auxotrophic for Phe). When needed, kanamycin was added to the initial medium to final concentrations of 20 mg/L. In pH-controlled tube-scale cultures, autoclaved 100 g/L $CaCO_3$ (dispersed in distilled water) was added to the cultures to a final concentration of 10 g/L. LB medium was used for MA production in batch-scale cultures in a 2.0 L jar fermentor. LB medium contains (per liter) 20 g glucose, 10 g tryptone, 5 g yeast extract, 5 g NaCl, 20mg kanamycin.

**Culture conditions**

For MA fermentation from glucose, metabolically engineered strains were pre-cultured in 4 mL M9YP medium for 1 day at 37 °C with shaking at 220 rpm in a 15-mL test tube.

414 Each pre-culture medium was seeded to 5 mL M9YP medium (supplemented with 0.1
415 mM isopropyl β-D-1-thiogalactopyranoside) in a 15 mL test tube at an initial optical
416 density ($OD_{600}$) of 0.1. The tube-scale cultures were incubated at 37 °C with shaking at
417 220 rpm. In the pH-controlled tube-scale cultures, $CaCO_3$ was added to the cultures at
418 24 h after seeding. Batch-scale cultures were performed in a 2.0 L jar fermentor with a
419 400-mL working volume. LB medium was used for MA production at this scale.
420 The 400mL of culture medium in the jar fermentor was inoculated with preculture
421 while adjusting the initial OD to 0.05. In order to maintain the pH at 7.0 during
422 cultivation, 1M HCl and 7% $NH_4OH$ was automatically added to the medium. The
423 dissolved oxygen (DO) was maintained at over 0.69 ppm by automatically controlling
424 the agitation speed from 200 to 600 rpm and supplementing and supplementing with air
425 at 400m L/min. In this experiment, 0.1mM IPTG was added to the medium at 3 h after
426 the initiation of cultivation.
427
428 **Analytical methods**
429 Cell growth was determined by measuring the OD at 600 nm on a UVmini-
430 1240 spectrophotometer (Shimadzu Corporation, Kyoto, Japan). Glucose was analyzed
431 using a Prominence HPLC System (Shimadzu) equipped with a Shodex SUGAR KS-

801 column (6 μm, 300 mm × 8.0 mm, L × I.D., Shodex). Water was used as the mobile phase with a flow rate of 0.8 mL/min, and the column was maintained at 50 °C. The HPLC profile was monitored using a refractive index detector.

MA was analyzed using an HPLC equipped with a SCR-102H column (7 μm, 8.0 mm × 300 mm, I.D. × L, respectively, Shimadzu). *p*-Toluenesulfonic acid (5 mM) was used as the mobile phase with a flow rate of 2.0 mL/min, and the column was maintained at 40 °C. The HPLC profile was monitored using a conductivity detector.

PHB, PCA, and catechol ware analyzed using an HPLC equipped with a PBr column (5 μm, 4.6 mm × 250 mm, I.D. × L, respectively, Nacalai Tesque). A dual-solvent system was used. Solvent A was 0.2 % phosphate buffer and solvent B was methanol. The flow rate of the mobile phase was 1.0 mL/min, and the column was maintained at 40 °C. The gradient was initiated as an 80:20 mixture of A and B (0–15 min), shifted to a 50:50 mixture of A and B (15–20 min), and subsequently shifted to an 80:20 mixture of A and B (20–25 min). The HPLC profile was monitored using a UV-VIS detector at 240 nm.

## SUPPORTING INFORMATION

Supporting information is containing cultivation profiles, detailed methods of

plasmids and strains construction, and sequences of the synthetic genes and custom DNA oligonucleotide primers.

## AUTHOR INFORMATION


### Corresponding Authors

*E-mail: shuhei.noda@riken.jp; tanaka@kitty.kobe-u.ac.jp


### Author Contribution

R.F., N.S. and T.T. proposed the idea, designed the experiments, and wrote the paper. R.F. performed the experiments. A.K. supervised the research. All authors read and approved the final manuscript.

### Conflict of Interest Disclosure

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS


This work was supported by JST-Mirai Program Grant Number JPMJMI17EI, Japan. R.F. was supported by a Research Fellowship of the Japan Society for the Promotion of Science (JSPS) for Young Scientists. This work was supported by the RIKEN Center for Sustainable Resource Science, Special Postdoctoral Researcher Program.

469

## REFERENCES


1. IHS Markit. (2017) Adipic Acid - Chemical Economics Handbook (CEH). https://www.ihs.com/products/adipic-acid-chemical-economics-handbook.html. Accessed November 16, 2017.

2. Global Industry Analysts Inc. (2016) Stable Demand from End-Use Sectors to Sustain Growth in the Global Adipic Acid Market. http://www.strategyr.com/MarketResearch/Adipic_Acid_Market_Trends.asp. Accessed November 16, 2017.

3. Kruyer, N. S., and Peralta-Yahya, P. (2017) Metabolic engineering strategies to bio-adipic acid production. *Curr. Opin. Biotechnol. 45*, 136-143. DOI: 10.1016/j.copbio.2017.03.006

4. Yu, J. Le, Xia, X. X., Zhong, J. J., and Qian, Z. G. (2014) Direct biosynthesis of adipic acid from a synthetic pathway in recombinant *Escherichia coli*. *Biotechnol. Bioeng. 111*, 2580-2586. DOI: 10.1002/bit.25293

5. Cheong, S., Clomburg, J. M., and Gonzalez, R. (2016) Energy- and carbon-efficient synthesis of functionalized small molecules in bacteria using non-decarboxylative Claisen condensation reactions. *Nat. Biotechnol. 34*, 556-561. DOI: 10.1038/nbt.3505

6. Chobchuenchom, W., Mongkolsuk, S., and Bhumiratana, A. (1996) Biodegradation of 3-chlorobenzoate by *Pseudomonas putida* 10.2. *World J. Microbiol. Biotechnol. 12*, 607-614. DOI: 10.1007/BF00327723

7. Jung, H. M., Jung, M. Y., and Oh, M. K. (2015) Metabolic engineering of *Klebsiella pneumoniae* for the production of *cis,cis*-muconic acid. *Appl. Microbiol. Biotechnol. 99*, 5217-5225. DOI: 10.1007/s00253-015-6442-3

8. Xie, N.-Z., Liang, H., Huang, R.-B., and Xu, P. (2014) Biotechnological production of muconic acid: current status and future prospects. *Biotechnol. Adv. 32*, 615-622. DOI: 10.1016/j.biotechadv.2014.04.001

9. Niu, W., Draths, K. M., and Frost, J. W. (2002) Benzene-free synthesis of adipic acid. *Biotechnol. Prog. 18*, 201-211. DOI: 10.1021/bp010179x

10. Barton, N., Horbal, L., Starck, S., Kohlstedt, M., Luzhetskyy, A., and Wittmann, C. (2018) Enabling the valorization of guaiacol-based lignin: Integrated chemical and biochemical production of *cis,cis*-muconic acid using metabolically engineered Amycolatopsis sp ATCC 39116. DOI: 10.1016/j.ymben.2017.12.001

11. Shen, L., Worrell, E., and Patel, M. (2010) Present and future development in plastics from biomass. *Biofuels, Bioprod. Biorefining*. *4*, 25-40. DOI: 10.1002/bbb.189

12. Lin, Y., Sun, X., Yuan, Q., and Yan, Y. (2014) Extending shikimate pathway for the production of muconic acid and its precursor salicylic acid in *Escherichia coli*. *Metab. Eng. 23*, 62-69. DOI: 10.1016/j.ymben.2014.02.009

13. Zhang, H., Li, Z., Pereira, B., and Stephanopoulos, G. (2015) Engineering *E. coli–E. coli* cocultures for production of muconic acid from glycerol. *Microb. Cell Fact. 14*, 134. DOI: 10.1186/s12934-015-0319-0

14. Sengupta, S., Jonnalagadda, S., Goonewardena, L., and Juturu, V. (2015) Metabolic engineering of a novel muconic acid biosynthesis pathway via 4-hydroxybenzoic acid in *Escherichia coli*. *Appl. Environ. Microbiol. 81*, 8037-8043. DOI: 10.1128/AEM.01386-15

15. Noda, S., Shirai, T., Oyama, S., and Kondo, A. (2016) Metabolic design of a platform *Escherichia coli* strain producing various chorismate derivatives. *Metab. Eng. 33*, 119-129. DOI: 10.1016/j.ymben.2015.11.007

16. Noda, S., and Kondo, A. (2017) Recent Advances in Microbial Production of Aromatic Chemicals and Derivatives. DOI: 10.1016/j.tibtech.2017.05.006

17. Matsumoto, T., Furuta, K., Tanaka, T., and Kondo, A. Sortase A□Mediated Metabolic Enzyme Ligation in *Escherichia coli*. DOI: 10.1021/acssynbio.6b00194

18. Wang, C., Yoon, S.-H., Jang, H.-J., et al. (2011) Metabolic engineering of *Escherichia coli* for a-farnesene production. *Metab. Eng. 13*, 648-655. DOI: 10.1016/j.ymben.2011.08.001

19. Singh, S. A., and Christendat, D. Structure of *Arabidopsis* Dehydroquinate Dehydratase-Shikimate Dehydrogenase and Implications for Metabolic Channeling in the Shikimate Pathway. DOI: 10.1021/bi060366+

20. Lütke-Eversloh, T., and Stephanopoulos, G. (2007) L-Tyrosine production by deregulated strains of *Escherichia coli*. *Appl. Microbiol. Biotechnol. 75*, 103-110. DOI: 10.1007/s00253-006-0792-9

21. Miles, E. W., Rhee, S., and Davies, D. R. (1999) The molecular basis of substrate channeling. *J. Biol. Chem. 274*, 12193-12196. DOI: 10.1074/JBC.274.18.12193

22. Zhang, Y.-H. P. (2011) Substrate channeling and enzyme complexes for biotechnological applications. *Biotechnol. Adv. 29*, 715-725. DOI: 10.1016/j.biotechadv.2011.05.020

23. Duncan, K., Chaudhuri, S., Campbell, M. S., and Coggins, J. R. (1986) The overexpression and complete amino acid sequence of *Escherichia coli* 3-dehydroquinase. *Biochem. J. 238*, 475-483. DOI: 10.1042/bj2380475

24. Small, P., Blankenhorn, D., Welty, D., Zinser, E., and Slonczewski, J. L. (1994)

Acid and base resistance in *Escherichia coli* and *Shigella flexneri*: role of *rpoS* and growth pH. *J. Bacteriol. 176*, 1729-1737. DOI: 10.1128/JB.176.6.1729-1737.1994

25. Mugikura, S., Nishikawa, M., Igarashi, K., and Kobayashi, H. (1990) Maintenance of a Neutral Cytoplasmic pH Is Not Obligatory for Growth of *Escherichia coli* and *Streptococcus faecalis* at an Alkaline pH. *J. Biochem. 108*, 86-91. DOI: 10.1093/oxfordjournals.jbchem.a123168

26. Thompson, B., Pugh, S., Machas, M., and Nielsen, D. R. (2018) Muconic Acid Production *via* Alternative Pathways and a Synthetic "Metabolic Funnel." *ACS Synth. Biol. 7*, 565-575. DOI: 10.1021/acssynbio.7b00331

Supporting Information

# Muconic Acid Production Using Gene-level Fusion Proteins in *Escherichia coli*

Ryosuke Fujiwara[†], Shuhei Noda[*‡], Tsutomu Tanaka[*†], Akihiko Kondo[§]

[†] Department of Chemical Science and Engineering, Graduate School of Engineering, Kobe University

[‡] Center for Sustainable Resource Science, RIKEN

[§] Graduate School of Science, Technology and Innovation, Kobe University

Supplementary methods

Plasmid construction

pZA23-ncmI, pZA23-ncamI, pZA23-nca/mI, pZA23-UpYc, pZA23-ZYc, pZA23-DZYc, pZA23-D/ZYc were constructed as follows. Synthetic genes corresponding to *Pseudomonas aeruginosa pchB*, *Pseudomonas putida* KT2440 *nahG*, and *P. putida* DOT-T1E *catA* were obtained as previously described.[1] Synthetic genes corresponding to *P. putida* KT2440 *pobA*, *Klebsiella pneumonia aroY*, and *Bacillus thuringiensis aroZ* optimized for the *Escherichia coli* codon usage were obtained from a commercial source (Invitrogen). The *pchB* gene was amplified by PCR using the synthetic gene fragment as a template with the primer pair pchB_f and pchB_r. The amplified fragment was cloned into the *Hind*III site of pZA23MCS. The resulting plasmid was designated as pZA23-I. The *menF* gene fragments were amplified by PCR using *E. coli* MG1655 genomic DNA as a template with the primer pair menF_f and menF_r, respectively. The amplified fragment was cloned into the *Kpn*I site of pZA23-I, and the resulting plasmid was designated as pZA23-mI. The *menF-pchB* gene fragment was amplified by PCR using the pZE12-mI as a template with the primer pair menF_f2 and pchB_r2. The amplified fragment was cloned into the KpnI site of pZA23-nGcA, and the resulting plasmid was designated as pZA23-mInc.The *nahG-catA* gene fragment

was amplified by PCR using the pZE12-nGcA as a template with the primer pair nahG_f and catA_r. The amplified fragment was cloned into the *Kpn*I site of pZA23-mI, and the resulting plasmid was designated as pZA23-ncmI. The *aroC* gene fragment (Contain a flexible $(GGGGS)_3$ linker fragment on the 3'end side) was amplified by PCR using *E. coli* MG1655 genomic DNA as a template with the primer pair aroC_f and aroC_r. The amplified fragment was cloned into the *Kpn*I site of pZA23 mI, and the resulting plasmid was designated as pZA23-a/mI. The *nahG-catA* gene fragment for pZA-nca/mI was amplified by PCR using the pZE12-nGcA as a template with the primer pair nahG_f and catA_r2. The amplified fragment was cloned into the *Kpn*I site of pZA23-a/mI, and the resulting plasmid was designated as pZA23-nca/mI. The linearized fragment (for pZA23-ncamI) was amplified by inverse PCR using the pZA23-nca/mI as a template with the primer pair inv_f and inv_r. The linker sequence was replaced by the stop codon and RBS in this inverse PCR. The amplified fragment was circularized using In-Fusion HD Cloning Kit (Takara), and the resulting plasmid was designated as pZA23-ncamI. The *aroY* gene fragment was amplified by PCR using the *aroY* synthetic gene as a template with the primer pair aroY_f and aroY_r. The amplified fragment was cloned between the *Kpn*I site and *EcoR*V of pZA23MCS, and the resulting plasmid was designated as pZA23-Y. The *catA* gene fragment was

amplified by PCR using the pZE12-nGcA as a template with the primer pair catA_f2 and catA_r2. The amplified fragment was cloned into the *EcoR*V of pZA23-Y, and the resulting plasmid was designated as pZA23-Yc. The *pobA* gene fragment was amplified by PCR using the *pobA* synthetic gene as a template with the primer pair pobA_f and pobA_r. The amplified fragment was cloned into the *Kpn*I of pZA23-Yc, and the resulting plasmid was designated as pZA23-pYc. The linearized fragment (for pZA23-UpYc) was amplified by inverse PCR using the pZA23-pYc as a template with the primer pair inv_f2 and inv_r2. The *ubiC* gene fragment was amplified by PCR using *E. coli* MG1655 genomic DNA as a template with the primer pair ubiC_f and ubiC_r. The amplified fragment was cloned into the linearized fragment (for pZA23-UpYc), and the resulting plasmid was designated as pZA23-UpYc. The *aroZ* gene fragment was amplified by PCR using the *aroZ* synthetic gene as a template with the primer pair aroZ_f and aroZ_r. The amplified fragment was cloned into the *Kpn*I of pZA23-Yc, and the resulting plasmid was designated as pZA23-ZYc. The linearized fragment (for pZA23-DZYc, -D/ZYc) was amplified by inverse PCR using the pZA23-ZYc as a template with the primer pair inv_f3 and inv_r3. The *aroD* gene fragment was amplified by PCR using *E. coli* MG1655 genomic DNA as a template with the primer pair aroD_f and aroD_r. The amplified fragment was cloned into the linearized fragment (for

pZA23-DZYc, -D/ZYc), and the resulting plasmid was designated as pZA23-DZYc. The *aroD* gene fragment (Contain the flexible $(GGGGS)_3$ linker fragment on the 3'end side) was amplified by PCR using *E. coli* MG1655 genomic DNA as a template with the primer pair aroD_f and aroD_r2. The amplified fragment was cloned into the linearized fragment (for pZA23-DZYc, -D/ZYc), and the resulting plasmid was designated as pZA23-DZYc and pZA23-D/ZYc. The linearized fragment (for pZA23-nca·mI, -nca//mI) was amplified by inverse PCR using the pZA23-nca/mI as a template with the primer pair inv_f4 and inv_r4 (for pZA23-nca·mI) or inv_f5 and inv_r5 (pZA23-nca//mI). The amplified linearized fragment (for pZA23-nca·mI, - nca//mI) was self-cloned, and the resulting plasmid was designated as pZA23- nca·mI and pZA23-nca//mI. The linearized fragment (for pZA23-D·ZYc, -D//ZYc) was amplified by inverse PCR using the pZA23-D/ZYc as a template with the primer pair inv_f6 and inv_r6 (for pZA23-D·ZYc) or inv_f7 and inv_r7 (pZA23-D//ZYc). The amplified linearized fragment (for pZA23-D·ZYc, -D//ZYc) was self-cloned, and the resulting plasmid was designated as pZA23-D·ZYc and pZA23-D//ZYc

Supplementary figures

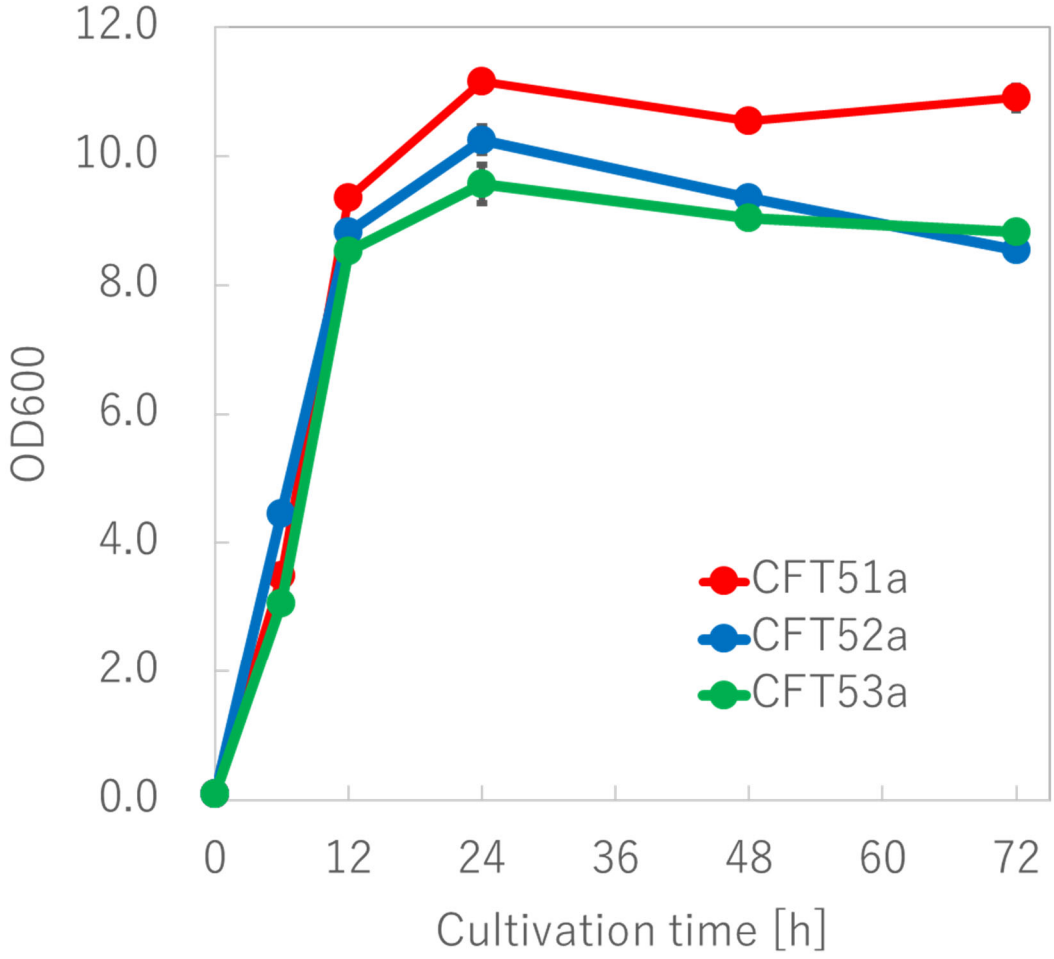



Figure S1

Bacterial cell growth in CFT51a, CFT52a, and CFT53a. The data are presented as the average of three independent experiments, and error bars indicate the standard error.

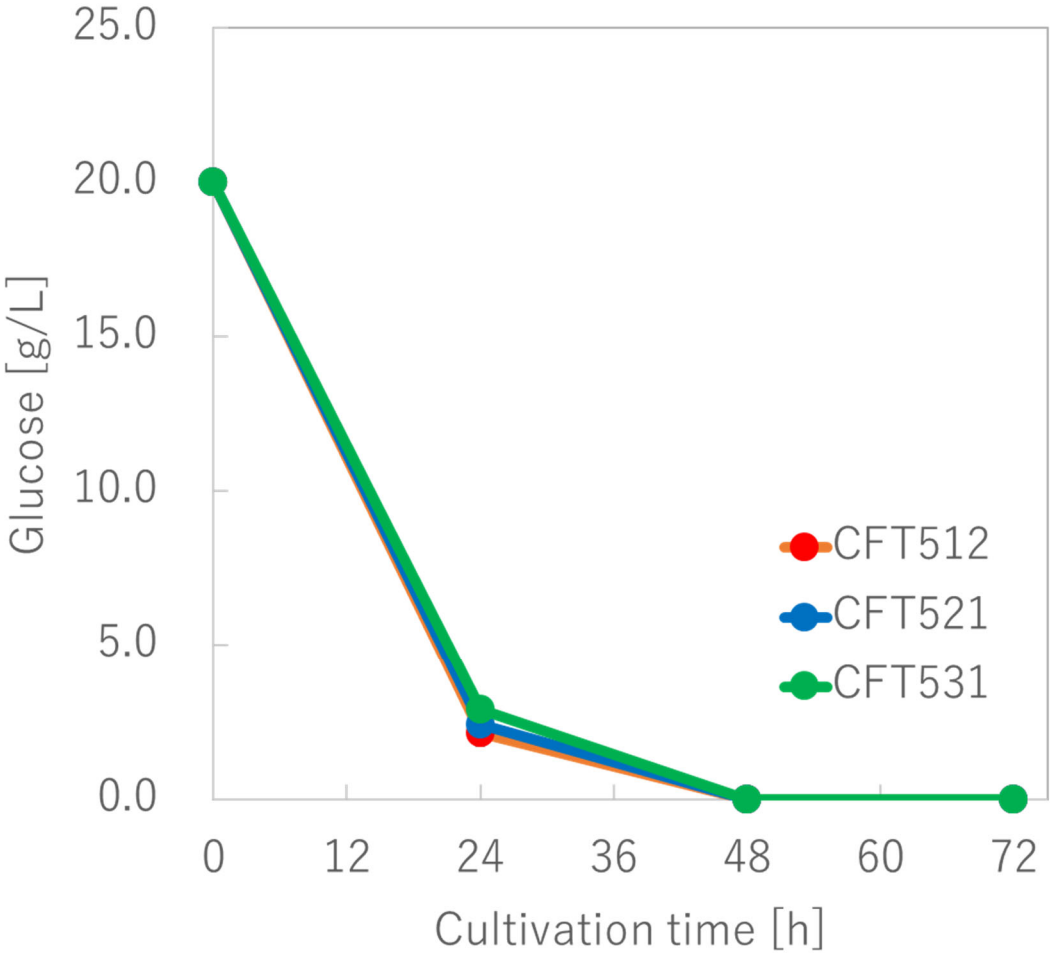



Figure S2

Glucose consumption by CFT51a, CFT52a, and CFT53a. The data are presented as the average of three independent experiments, and error bars indicate the standard error.

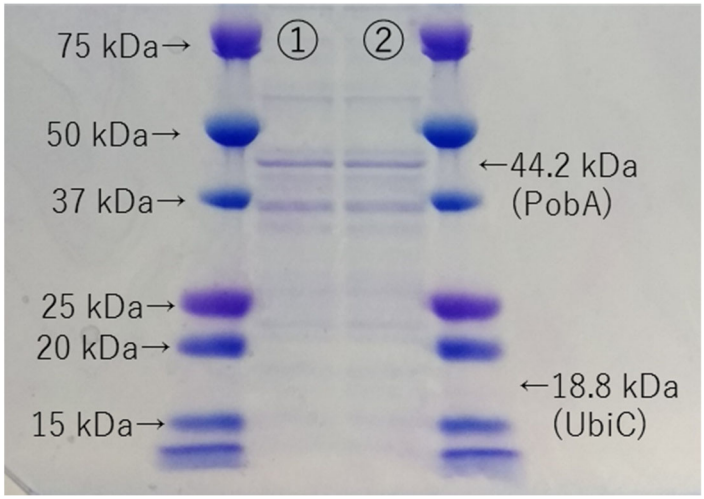
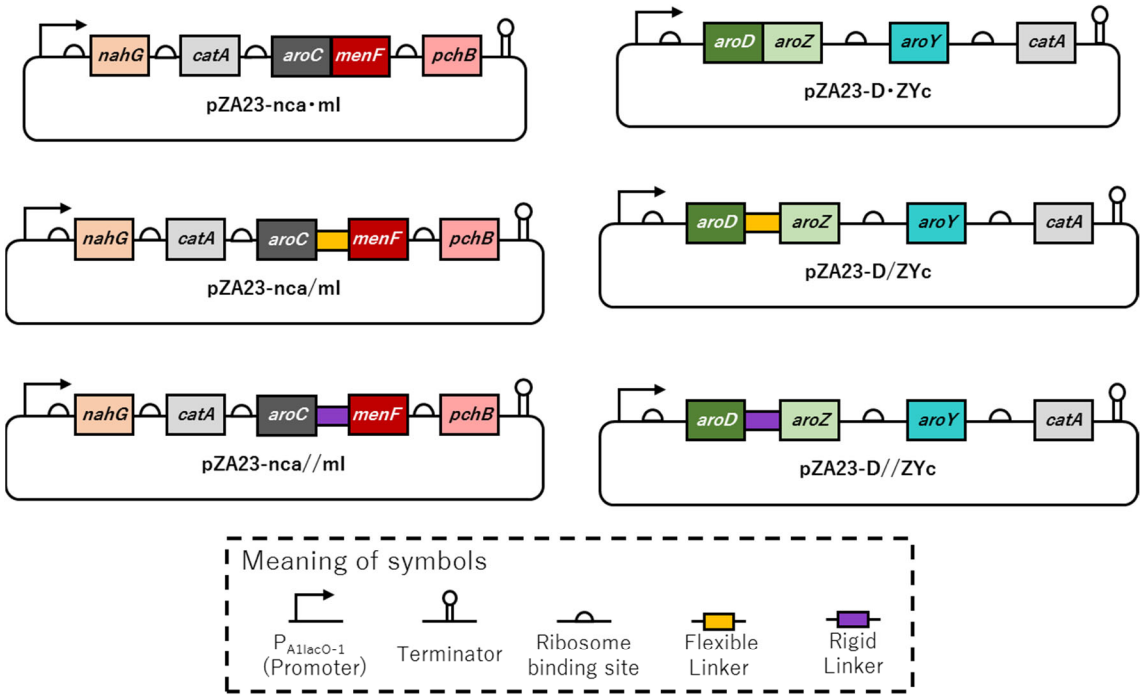



Figure S3

SDS-PAGE analysis of CFT52a. Both outer lanes: protein marker, Lane1: control (CFT5 horboring pZA23MCS), Lane2: CFT52a




| | DNA sequences | Protein sequences |
|---|---|---|
| without Linker | (3'end of upstream gene)-(5' end of downstream gene) | (upstream protein C-terminal)-(downstream protein N-terminal). |
| Flexible Linker | (3'end of upstream gene)-GGTGGTGGTGGTAGTGGTGGCGGTGGTTCAGGCGGTGGTGGTTCC-(5' end of downstream gene) | (upstream protein C-terminal)-GGGGSGGGGSGGGGS-(downstream protein N-terminal). |
| Rigid Linker | (3'end of upstream gene)-GAAGCTGCAGCAAAAGAAGCCGCTGCCAAAGAAGCGGCAGCGAAA-(5' end of downstream gene) | (upstream protein C-terminal)-EAAAKEAAAKEAAAK-(downstream protein N-terminal). |

Figure S4

Diagram of MA synthesis gene cassette in each plasmid for examining linkers.

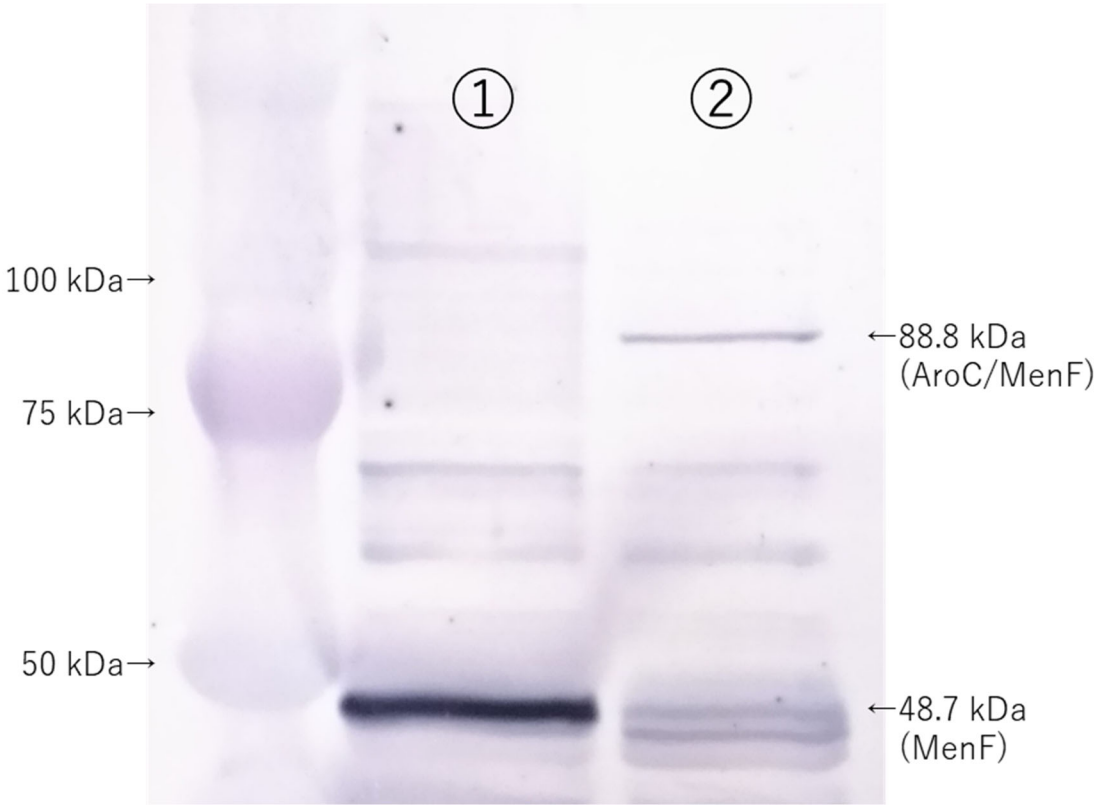



Figure S5

Western blotting analysis of CFT51b and CFT51c. Left side lane: protein marker, Lane1: CFT51b, Lane2: CFT51c. An ANTI-FLAG M2 Monoclonal Antibody (Sigma-Aldrich) and an anti-Mouse IgG (H+L), HRP Conjugate (Promega Corp.) were used to detect MenF and fused AroC/MenF.

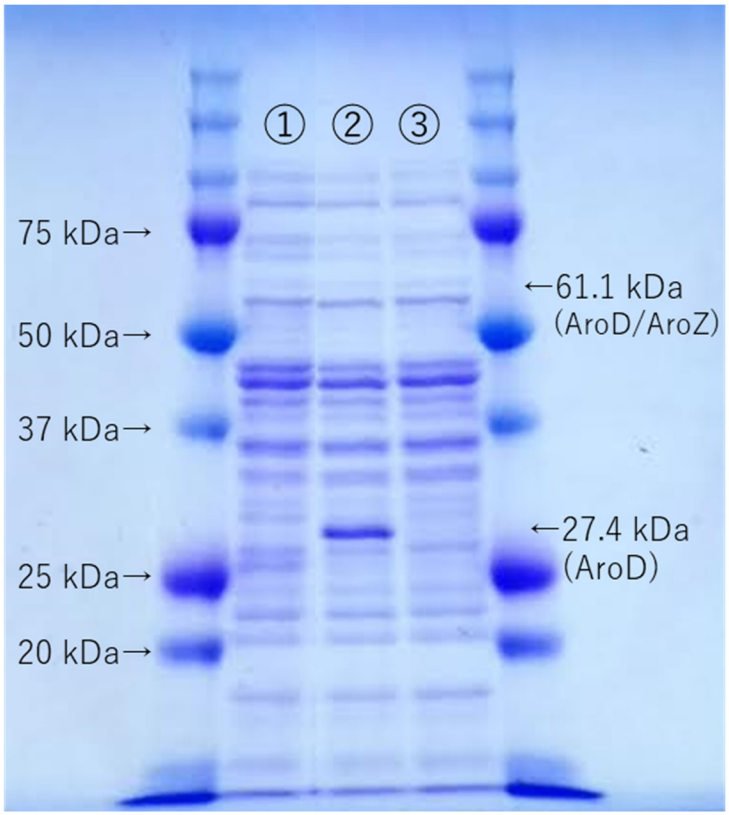



Figure S6

SDS-PAGE analysis of CFT53b and CFT53c. Both outer lanes: protein marker, Lane1: control (CFT5 horboring pZA23MCS), Lane2: CFT53b, Lane2: CFT53c.

121 Supplementary tables

122

123 Table S1. Plasmids used in this study.

| Plasmids | Genotype | Source or reference |
|---|---|---|
| pZE12MCS | $P_{LlacO1}$, *colE ori, Amp*$^r$ | Expressys |
| pZA23MCS | $P_{A1lacO-1}$, *p15A ori, Km*$^r$ | Expressys |
| pZE12-mI | pZE12MCS containing *menF* from *E. coli* and *pchB* from *P. aeruginosa* | Noda et al.[1] |
| pZA23-nGcA | pZA23MCS containing *nahG* from *P. putida* KT2440 and *catA* from *P. putida* DOT-T1E | Noda et al.[1] |
| pZA23-I | pZA23MCS containing *pchB* from *P. aeruginosa* | This study |
| pZA23-mI | pZA23-I containing *menF* from *E. coli* | This study |
| pZA23-mInc | pZA23MCS containing *menF, pchB, nahG and catA,* | This study |
| pZA23-ncmI | pZA23MCS containing *nahG, catA, menF,* and *pchB* | This study |
| ZA23-ncamI | pZA23-ncmI containing *aroC* | This study |
| pZA23-nca/mI | pZA23MCS containing *nahG, catA, pchB,* and *aroC/menF,* which is a gene expressing fusion protein | This study |

| | | |
|---|---|---|
| | composed of AroC and MenF connected by flexible linker | |
| pZA23-nca・mI | pZA23MCS containing *nahG, catA, pchB,* and *aroC/menF,* which is a gene expressing fusion protein composed of AroC and MenF connected directly (without linker) | This study |
| pZA23-nca//mI | pZA23MCS containing *nahG, catA, pchB,* and *aroC/menF,* which is a gene expressing fusion protein composed of AroC and MenF connected by rigid linker | This study |
| pZA23-Y | pZA23MCS containing *aroY* from *K. pneumoniae* (codon optimized for *E.coli*) | This study |
| pZA23-Yc | pZA23-Y containing *catA* | This study |
| pZA23-pYc | pZA23-Yc containing *ubiC* from *E. coli* | This study |
| pZA23-UpYc | pZA23-pYc containing *pobA* from *P. putida* KT2440(codon optimized for *E.coli*) | This study |
| pZA23-ZYc | pZA23-Yc containing *aroZ* from *Bacillus thuringiensis* (codon optimized for *E.coli*) | This study |
| pZA23-DZYc | pZA23-ZYc containing *aroD* from *E.coli* | This study |

| | | |
|---|---|---|
| pZA23-D/ZYc | pZA23-Yc containing AroD-AroZ, which is a gene expressing fusion protein composed of AroD and AroZ connected by flexible linker | This study |
| pZA23-D･ZYc | pZA23-Yc containing AroD-AroZ, which is a gene expressing fusion protein composed of AroD and AroZ connected directly (without linker) | This study |
| pZA23-D//ZYc | pZA23-Yc containing AroD-AroZ, which is a gene expressing fusion protein composed of AroD and AroZ connected by rigid linker | This study |

125 Table S2 Oligonucleotide primers used in this study.

| Oligonucleotide primers | Sequence |
|---|---|
| pchB_f | 5'- GTCGACGGTATCGATAAAGAGGAGAAAAAGCTTATGAAAACCCCTGAAGATTG -3' |
| pchB_r | 5'- CAGGAATTCGATATCTAAATGATGATGATGATGATGGGCTGCACCACGGGTCTGAC -3' |
| menF_f | 5'- ATTAAAGAGGAGAAAGGTACCATGCAATCACTTACTACGGC -3' |
| menF_r | 5'- CTCGAGGGGGGGCCCTTACTTGTCATCGTCATCCTTGTAGTCTTCCATTTGTAATAAAGTAC -3' |
| menF_f2 | 5'- TTAAAGAGGAGAAAggtaccATGCAATCACTTACTACGGC -3' |
| pchB_r2 | 5'- GTACTGTTCTGCATgtaccAAGCTTTTTCTCCTCTTTttaggctgcaccacgggtctgac -3' |
| nahG_f | 5'- TTAAAGAGGAGAAAggtaccATGCAGAACAGTACCAGCGC-3' |
| catA_r | 5'- gtaagtgattgcatggtaccAAGCTTTTTCTCCTCTTTctattaACCTTCTTGCAGTG -3' |
| aroC_f | 5'- TTAAAGAGGAGAAAGGTACCatggctggaaacacaattggacaac -3' |

| | |
|---|---|
| aroC_r | 5'- gtaagtgattgcat GGAACCACCACCGCCTGAACCACCGCCACCACTACCACCACCACCccagcgtggaatatcagtcttcaca -3' |
| catA_r2 | 5'- gtgtttccagccat ggtaccAAGCTTTTTCTCCTCTTT ctattaACCTTCTTGCAGTG -3' |
| inv_f | 5'- cattaaagaggagaaaggatgcaatcacttactacggcgctgg -3' |
| inv_r | 5'- ttctcctctttaatgctattaccagcgtggaatatcagtcttcac -3' |
| aroY_f | 5’- TTAAAGAGGAG AAAggtaccATGACCGCACCGATTCAGGATCTGC -3’ |
| aroY_r | 5’- CTGCAGGAATTCgatatcCTATTA TTTTGCGCTACCCTGATTTTTTTCC -3’ |
| catA_f2 | 5’- AGCGCAAAATAATAG gatatc AAAGAGGAGAAAAAGCTT ATGACCGTGAAAATTAGCCATACCG -3’ |
| catA_r2 | 5’- CTGCAGGAATTCgatCTATTAACCTTCTTGCAGTGCACGC -3’ |
| pobA_f | 5’- TTAAAGAGGAGAAAggtacc ATGAAAACCCAGGTTGCAATTATTG -3’ |
| pobA_r | 5’- ATCGGTGCGGTCATggtaccAAGCTTTTTCTCCTCTTTctattaGGCAACTTCTTCAAACGGCAGACCA -3’ |

| | |
|---|---|
| inv_f2 | 5’- AAAGAGGAGAAAAAGCTTATGAAAACCCAGGTTGCAATTATTGGTGC -3’ |
| inv_r2 | 5’- ggtaccTTTCTCCTCTTTAATGAATTCTGTGTGAAATTG -3’ |
| ubiC_f | 5’- GAGGAGAAAggtacc ATGTCACACCCCGCGTTAACGC -3’ |
| ubiC_r | 5’- CTTTTTCTCCTCTTTctattaGTACAACGGTGACGCCGGTAAAAACA -3’ |
| aroZ_f | 5’- TTAAAGAGGAGAAAggtaccATGAAATATAGCCTGTGCACCATTAG -3’ |
| aroZ_r | 5’- ATCGGTGCGGTCATggtaccAAGCTTTTTCTCCTCTTTctattaGCTGGTAACAACTTCCAGTTTACGG -3’ |
| inv_f3 | 5’- ATGAAATATAGCCTGTGCACCATTA -3’ |
| inv_r3 | 5’- GGTACCTTTCTCCTCTTTAATGAATTC -3’ |
| aroD_f | 5’- GAGGAGAAAggtacc atgaaaaccgtaactgtaaaagatc -3’ |
| aroD_r | 5’- CAGGCTATATTTCATAAGCTTTTTCTCCTCTTTctattatgcctggtgtaaaatagttaatacc -3’ |

| | |
|---|---|
| | 5’- |
| aroD_r2 | CAGGCTATATTTCATTGAACCACCACCGCCTGAACCACCGCCACCACTACCACCACCACCtgcctggtgtaaaat |
| | agttaatacc -3’ |
| inv_f4 | 5’- CTGGggtaccATGCAATCACTTACTACGGCGCTGGAAAAT -3’ |
| inv_r4 | 5’- TGCATggtaccCCAGCGTGGAATATCAGTCTTCACATCG -3’ |
| inv_f5 | 5’- GCCGCTGCCAAAGAAGCGGCAGCGAAAATGCAATCACTTACTACGGCGCTGGAAAAT -3’ |
| inv_r5 | 5’- TTCTTTGGCAGCGGCTTCTTTTGCTGCAGCTTCCCAGCGTGGAATATCAGTCTTCACATCG -3’ |
| inv_f6 | 5’- GCAggatccATGAAATATAGCCTGTGCACCATTAGCTTTCGT -3’ |
| inv_r6 | 5’- TTTCATggatccTGCCTGGTGTAAAATAGTTAATACCGTGCG -3’ |
| inv_f7 | 5’- GCCGCTGCCAAAGAAGCGGCAGCGAAAATGAAATATAGCCTGTGCACCATTAGCTTTCGT -3’ |
| inv_r7 | 5’- TTCTTTGGCAGCGGCTTCTTTTGCTGCAGCTTCTGCCTGGTGTAAAATAGTTAATACCGTGCG -3’ |

Table S3 Results of the order of the gene examination in Pathway 1.

| Parelental strain | Plasmid | Production titer (72h) [g/L] |
|---|---|---|
| CFT5 | pZA23-mInc | 0.14 ± 0.05 |
| CFT5 | pZA23-ncmI | 1.00 ± 0.11 |

132 Table S4 Results of linkers examination.

| Parelental strain | Plasmid | Production titer (72h) [g/L] |
|---|---|---|
| CFT5 | pZA23-nca·mI | n.d. |
| CFT5 | pZA23-nca/mI | 3.45 ± 0.04 |
| CFT5 | pZA23-nca//mI | 2.98 ± 0.21 |
| CFT5 | pZA23-D·ZYc | n.d. |
| CFT5 | pZA23-D/ZYc | 1.20 ± 0.10 |
| CFT5 | pZA23-D//ZYc | n.d. |

133

134

Supplementary materials

Sequence of *pobA* from *Pseudomonas putida* KT2440.

ATGAAAACCCAGGTTGCAATTATTGGTGCAGGTCCGAGCGGTCTGCTGCTGG

GTCAGCTGCTGCATAAAGCAGGTATTGATAACATTATTGTGGAACGTCAGACC

GCAGAATATGTTCTGGGTCGTATTCGTGCCGGTGTTCTGGAACAGGGCACCG

TTGATCTGCTGCGTGAAGCCGGTGTTGCAGAACGTATGGATCGTGAAGGTCT

GGTTCATGAAGGTGTTGAACTGCTGGTTGGTGGTCGTCGTCAGCGTCTGGAT

CTGAAAGCACTGACCGGTGGTAAAACCGTTATGGTTTATGGTCAGACCGAAG

TTACCCGTGATCTGATGCAGGCACGTGAAGCAAGTGGTGCACCGATTATCTAT

AGCGCAGCAAATGTTCAGCCGCATGAACTGAAAGGTGAAAAACCGTATCTG

ACCTTTGAAAAAGATGGTCGTGTTCAGCGTATCGATTGTGATTATATTGCAGG

TTGTGATGGCTTTCATGGTATTAGCCGTCAGAGCATTCCGGAAGGTGTGCTGA

AACAGTATGAACGTGTTTATCCGTTTGGTTGGCTGGGTCTGCTGAGCGATACC

CCTCCGGTTAATCACGAACTGATTTATGCACATCATGAACGTGGTTTTGCACT

GTGTAGCCAGCGTAGTCAGACCCGTAGCCGTTATTATCTGCAGGTTCCGCTGC

AGGATCGTGTTGAAGAATGGTCAGATGAACGTTTTTGGGATGAGCTGAAAGC

CCGTCTGCCTGCAGAAGTTGCAGCCGATCTGGTTACCGGTCCGGCACTGGAA

153 AAAAGCATTGCACCGCTGCGTAGCCTGGTTGTTGAACCGATGCAGTATGGTC

154 ACCTGTTTCTGGTGGGTGATGCAGCACATATTGTTCCGCCTACCGGTGCAAAA

155 GGTCTGAATCTGGCAGCAAGTGATGTGAATTATCTGTATCGTATTCTGGTGAA

156 AGTGTATCATGAAGGCCGTGTGGATCTGCTGGCACAGTATAGTCCGCTGGCA

157 CTGCGTCGTGTTTGGAAAGGCGAACGTTTTAGCTGGTTTATGACCCAACTGC

158 TGCATGATTTTGGTAGCCATAAAGATGCATGGGACCAGAAAATGCAAGAAGC

159 AGATCGCGAATATTTTCTGACCAGTCCGGCAGGTCTGGTGAATATTGCAGAA

160 AATTATGTTGGTCTGCCGTTTGAAGAAGTTGCCtaa

161

Sequence of *aroY* from *Klebsiella pneumonia*.

ATGACCGCACCGATTCAGGATCTGCGTGATGCAATTGCCCTGCTGCAACAGC

ATGATAATCAGTATCTGGAAACCGATCATCCGGTTGATCCGAATGCAGAACTG

GCAGGCGTTTATCGTCATATTGGTGCCGGTGGCACCGTTAAACGTCCGACAC

GTATTGGTCCGGCAATGATGTTTAATAACATTAAAGGTTATCCGCACAGCCGT

ATTCTGGTTGGTATGCATGCAAGCCGTCAGCGTGCAGCACTGCTGCTGGGTT

GTGAAGCAAGTCAGCTGGCACTGGAAGTTGGTAAAGCAGTTAAAAAACCGG

TTGCACCGGTGGTTGTTCCGGCAAGCAGCGCACCGTGTCAAGAGCAGATTTT

TCTGGCAGATGATCCGGATTTTGATCTGCGTACCCTGCTGCCTGCACCGACCA

ATACACCGATTGATGCAGGTCCGTTTTTTTGTCTGGGTTTAGCACTGGCAAGC

GATCCGGTGGATGCAAGCCTGACCGATGTTACCATTCATCGTCTGTGTGTTCA

GGGTCGTGATGAACTGAGCATGTTCCTGGCAGCAGGTCGCCATATTGAAGTT

TTTCGTCAGAAAGCAGAAGCAGCAGGTAAACCGCTGCCGATTACCATTAATA

TGGGTTTAGATCCGGCAATCTATATCGGTGCATGTTTTGAAGCACCGACAACA

CCGTTTGGTTATAATGAACTGGGTGTTGCGGGTGCACTGCGTCAGCGTCCGG

TTGAACTGGTTCAGGGTGTTAGCGTTCCGGAAAAAGCAATTGCACGTGCCGA

AATTGTTATTGAAGGTGAACTGTTACCGGGTGTTCGTGTTCGTGAAGATCAGC

ATACCAATAGCGGTCATGCAATGCCGGAATTTCCGGGTTATTGTGGTGGTGCA

AATCCGAGCCTGCCGGTTATTAAAGTTAAAGCCGTTACCATGCGCAATAATGC

AATTCTGCAGACCCTGGTTGGTCCGGGTGAAGAACATACCACACTGGCAGGT

CTGCCGACCGAAGCAAGCATTTGGAATGCAGTTGAAGCAGCAATTCCGGGTT

TCCTGCAGAATGTTTATGCACATACCGCAGGCGGTGGTAAATTTCTGGGTATC

CTGCAGGTAAAAAAACGTCAGCCTGCAGATGAAGGTCGTCAGGGTCAAGCA

GCCCTGCTGGCCCTGGCAACCTATAGCGAACTGAAAAATATCATTCTGGTGG

ATGAGGATGTGGACATTTTTGATAGTGATGATATTCTGTGGGCAATGACCACA

CGTATGCAGGGTGATGTTAGCATTACCACCATTCCGGGTATTCGCGGTCATCA

GCTGGATCCGAGCCAGACACCGGAATATTCACCGAGCATTCGTGGTAATGGT

ATTAGCTGCAAAACCATCTTTGATTGTACCGTTCCGTGGGCACTGAAAAGCC

ATTTTGAACGTGCACCGTTTGCAGATGTTGATCCGCGTCCGTTTGCACCTGAA

TATTTTGCACGTCTGGAAAAAAATCAGGGTAGCGCAAAA<u>taa</u>

193 Sequence of *aroZ* from *Bacillus thuringiensis*.

194 ATGAAATATAGCCTGTGCACCATTAGCTTTCGTCACCAGCTGATTAGCTTTACC

195 GATATTGTTCAGTTTGCCTATGAAAACGGCTTTGAAGGTATTGAACTGTGGGG

196 CACCCATGCACAGAATCTGTATATGCAAGAACGTGAAACCACCGAACGTGAA

197 CTGAATTTCCTGAAAGATAAGAACCTGGAAATCACCATGATCAGCGATTATCT

198 GGATATTAGCCTGAGCGCAGATTTTGAAAAAACCATCGAAAAAAGCGAACA

199 GCTGGTTGTTCTGGCCAATTGGTTTAACACCAACAAAATTCGTACCTTCGCAG

200 GTCAGAAAGGCAGCAAAGATTTTAGCGAACAAGAACGCAAAGAATACGTGA

201 AACGCATTCGCAAAATCTGTGATGTTTTTGCCCAGCACAATATGTATGTTCTG

202 CTGGAAACCCATCCGAATACACTGACCGATACACTGCCGAGCACCATTGAAC

203 TGCTGGAAGAAGTTAATCATCCGAACCTGAAAATTAACCTGGATTTTCTGCAT

204 ATCTGGGAAAGCGGTGCAAATCCGATTGATAGCTTTCATCGTCTGAAACCGT

205 GGACACTGCATTATCACTTTAAAAACATTAGCAGCGCAGACTATCTGCATGTG

206 TTTGAACCGAATAATGTTTATGCAGCAGCAGGTAGCCGTATTGGTATGGTTCC

207 GCTGTTTGAAGGCATTGTGAACTATGATGAAATCATCCAAGAAGTTCGCGGT

208 ACAGACCTGTTTGCAAGCCTGGAATGGTTTGGTCATAACAGCAAAGAGATTC

209 TGAAAGAAGAGATGAAAGTTCTGATCAACCGTAAACTGGAAGTTGTTACCA

210 GCtaa

211 Underline: - Initiation and termination codons

## References

1. Noda, S., Shirai, T., Oyama, S., and Kondo, A. (2016) Metabolic design of a platform *Escherichia coli* strain producing various chorismate derivatives. *Metab Eng*. *33*, 119-129. DOI: 10.1016/j.ymben.2015.11.007