Green acetyl modification of puerarin to form puerarin 6″-*O*-acetate using engineered *Escherichia coli* with favorable pathways and elevated acetyl-CoA supply

Xue-Ning Wang, Li Sheng, Xin-Yi Guo, Hua Sun, Jian-Qiang Kong*

Institute of Materia Medica, Chinese Academy of Medical Sciences & Peking Union Medical College (State Key Laboratory of Bioactive Substance and Function of Natural Medicines, NHC Key Laboratory of Biosynthesis of Natural Products & CAMS Key Laboratory of Enzyme and Biocatalysis of Natural Drugs), Beijing, 100050, China

*, Author to whom correspondence should be addressed.

E-mail: jianqiangk@imm.ac.cn

Figure legends

Figure S1 The conversion of P6″A from puerarin through the biological and chemical processes.

Figure S2 The PCR product of *E.coli* genes. 1: *acs*; 2: *pank*; 3: a*ceE*; 4: *aceF*;5: *lpdA*; M: DNA markers indicated with bp.

Figure S3 The co-expression analysis of *acs* and *pank* genes.1-4, the induced expression of pACYCDuet-acs-pank, pCDFDuet-acs-pank, pETDuet-acs-pank and pRSFDuet-acs-pank, respectively; CK, the uninduced strains; M: protein markers indicated as kDa; Red arrows indicated the target proteins.

Figure S4 The co-expression analysis of *aceE*, *aceF* and *lpdA* genes. 1-4, the induced expression of pACYCDuet-pdh, pCDFDuet-pdh, pETDuet-pdh and pRSFDuet-pdh, respectively; CK: the uninduced strains; M: protein mass standards indicated with kDa. Red arrows indicated the target proteins.

Figure S5 The co-expression analysis of *acs*, *pank*, *aceE*, *aceF* and *lpdA* genes. 1-2, the induced expression of pACYCDuet-pdh-pank-acs and pCDFDuet-pdh-pank-acs, respectively; CK, the uninduced strains; M, protein mass standards indicated with kDa; Red arrows showed the expressed proteins.

Figure S6 MAT-catalyzed acetylation of puerarin in the engineered strains.

Figure S7 The condition optimizations to improve P6″A yield in the engineered cells. A, the effect of casamino acid on P6″A production; B, the effect of concentrations of puerarin (Pur) and glucose (Glc) on P6″A production; C, the effect of IPTG concentration on P6″A production.

Figure S8 Acetyltransferase-catalyzed acetylation of puerarin. A, HPLC profiles of acetylated metabolites. a, acetylation assay without acetyltransferases; b, GAT-catalyzed acetylation; c, MAT-catalyzed acetylation; d, the co-elution of the reaction

mixture and the authentical P6″A standard; e, P6″A standard; B, the MS spectrum of the metabolite 2. **1** and **2** represented puerarin and P6″A, respectively.

Figure S9 The effect of pH (A) and temperature (B) on the acetylation activity of GAT. Values are means ± SD, n = 3.

Figure S10 Kinetic parameters of GAT for acetyl-CoA (A) or puerarin (B). Values are means ± SD, n = 3.

Figure S11 The whole-cell-catalyzed acetylation of puerarin to form P6″A.

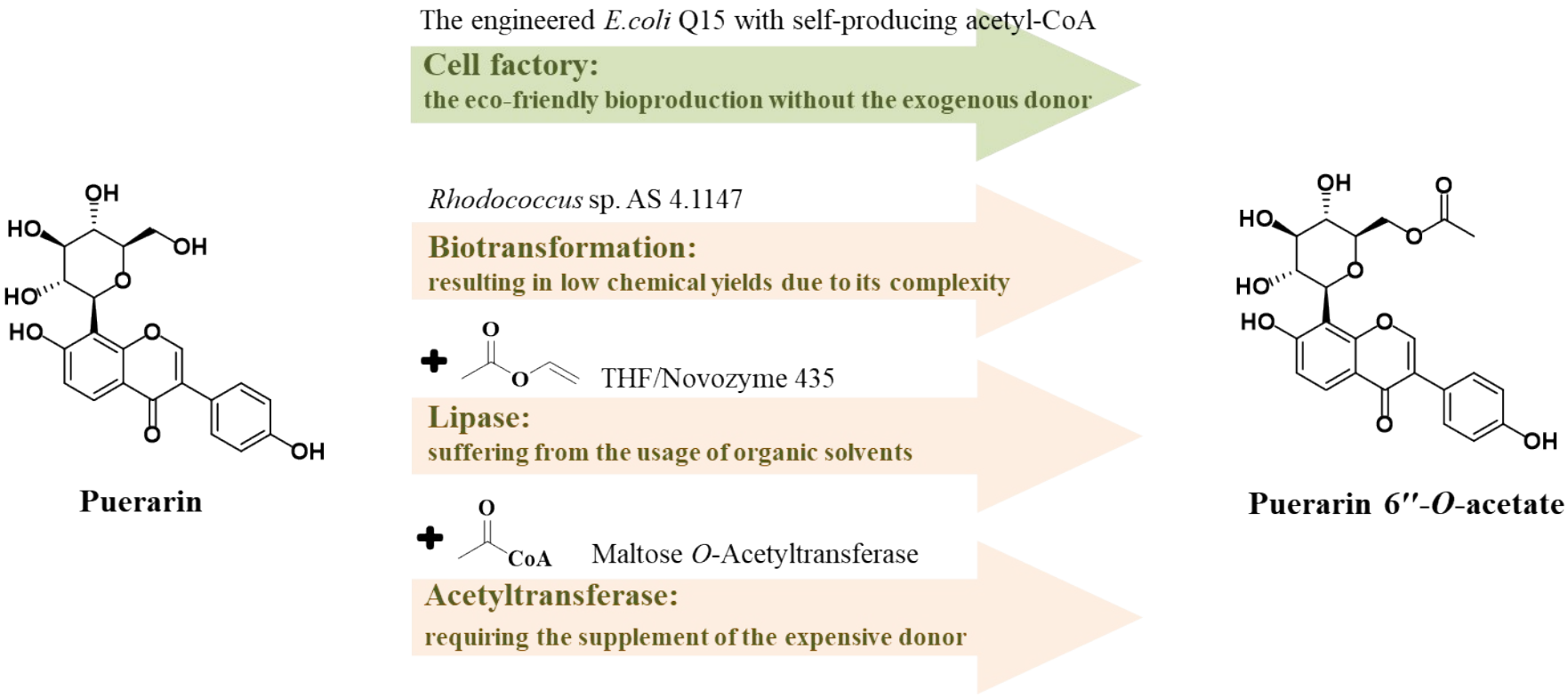

Figure S1 The conversion of P6″A from puerarin through the biological and chemical processes.

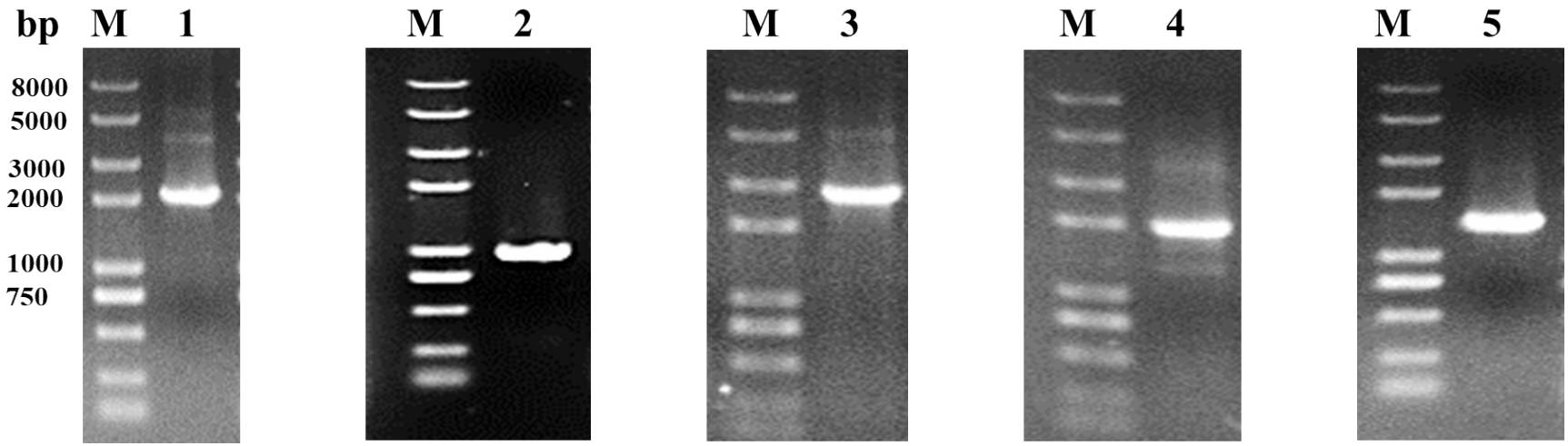

Figure S2 The PCR product of *E.coli* genes.

1:*acs*; 2:*pank*; 3:*aceE*; 4:*aceF*;5:*lpdA*; M: DNA markers indicated with bp.

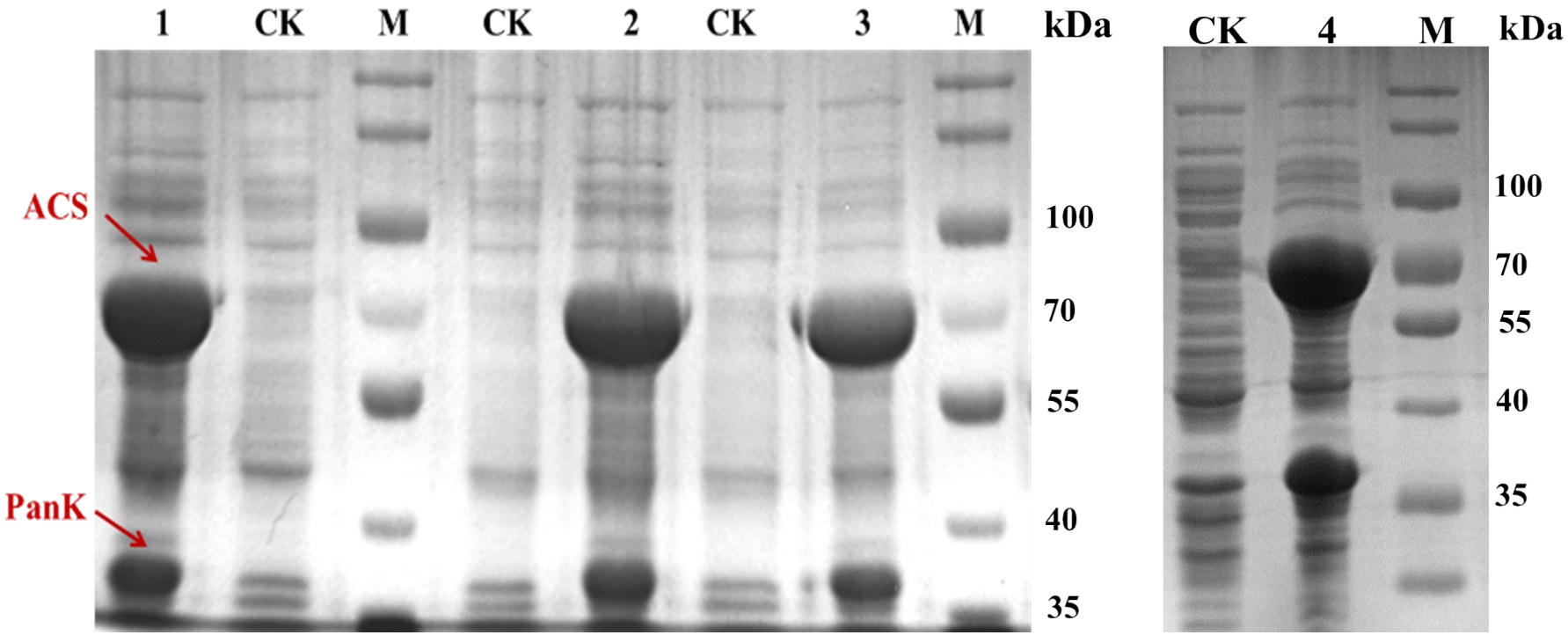

Figure S3 The co-expression analysis of *acs* and *pank* genes.1-4, the induced expression of pACYCDuet-acs-pank, pCDFDuet-acs-pank, pETDuet-acs-pank and pRSFDuet-acs-pank, respectively; CK, the uninduced strains; M: protein markers indicated as kDa; Red arrows indicated the target proteins.

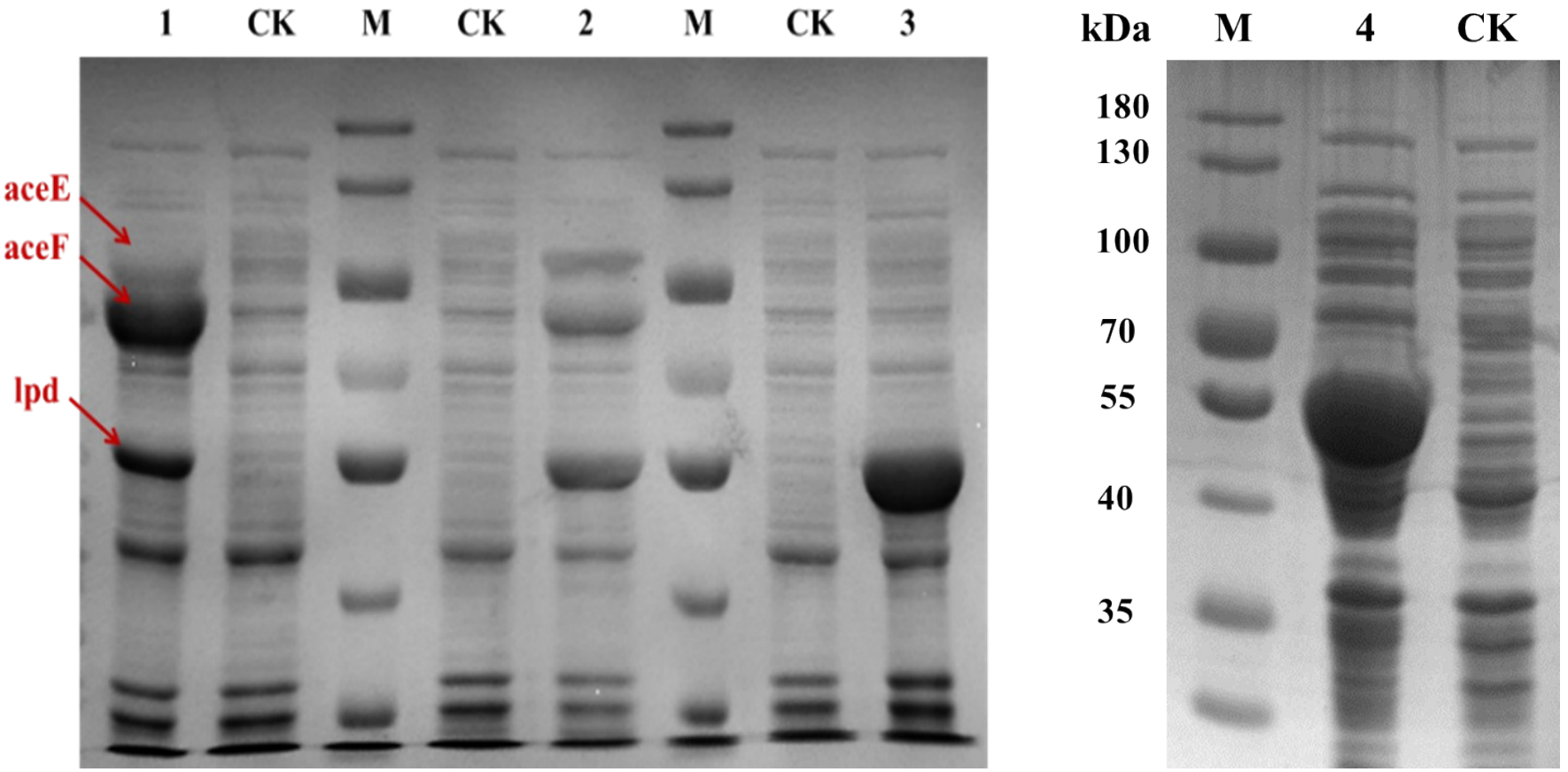

Figure S4 The co-expression analysis of *aceE*, *aceF* and *lpdA* genes.1-4, the induced expression of pACYCDuet-pdh, pCDFDuet-pdh, pETDuet-pdh and pRSFDuet-pdh, respectively; CK: the uninduced strains; M: protein mass standards indicated with kDa. Red arrows indicated the target proteins.

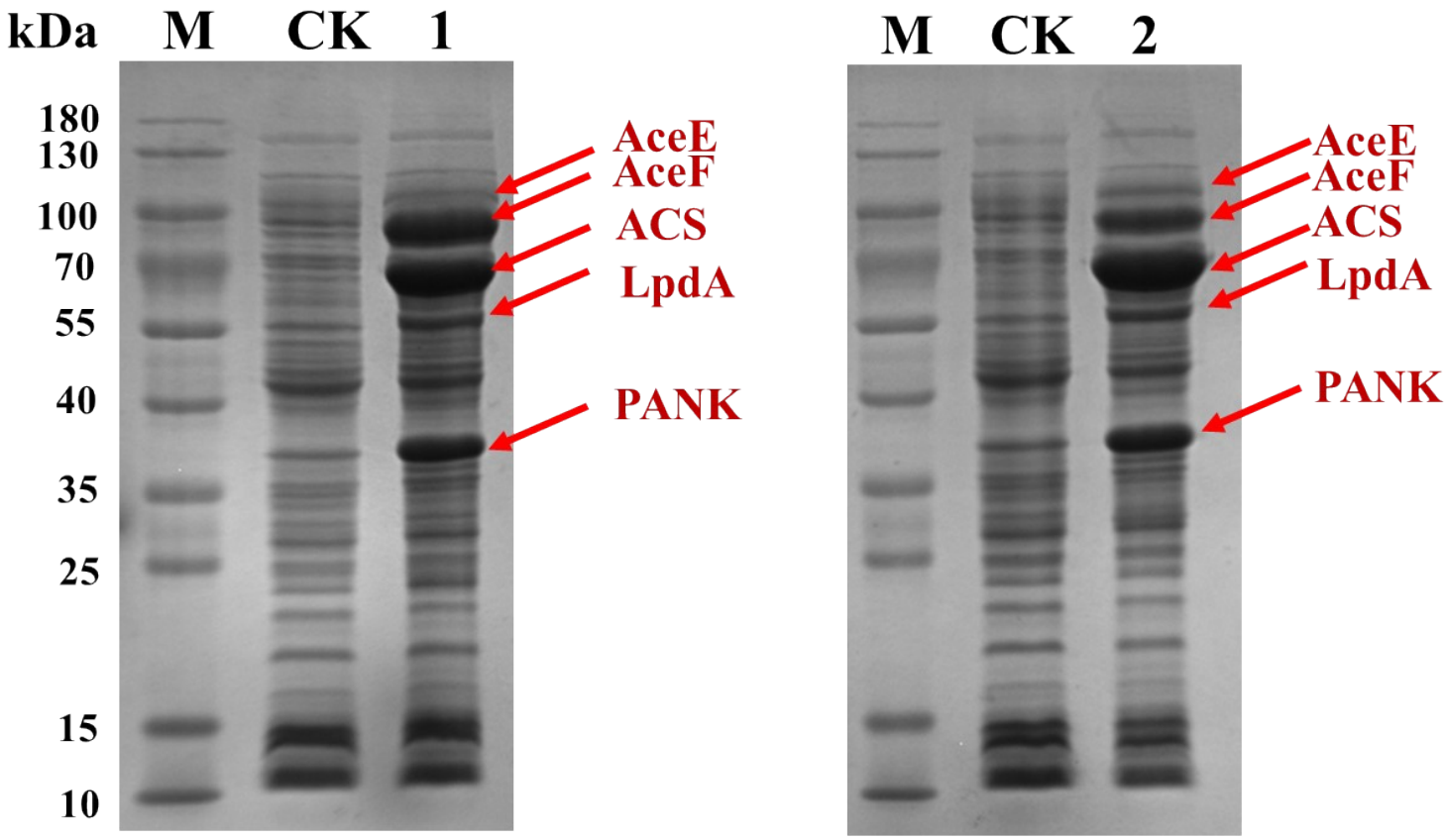

Figure S5 The co-expression analysis of *acs*, *pank*, *aceE*, *aceF* and *lpdA* genes.1-2, the induced expression of pACYCDuet-pdh-pank-acs and pCDFDuet-pdh-pank-acs, respectively; CK, the uninduced strains; M, protein mass standards indicated with kDa; Red arrows showed the expressed proteins.

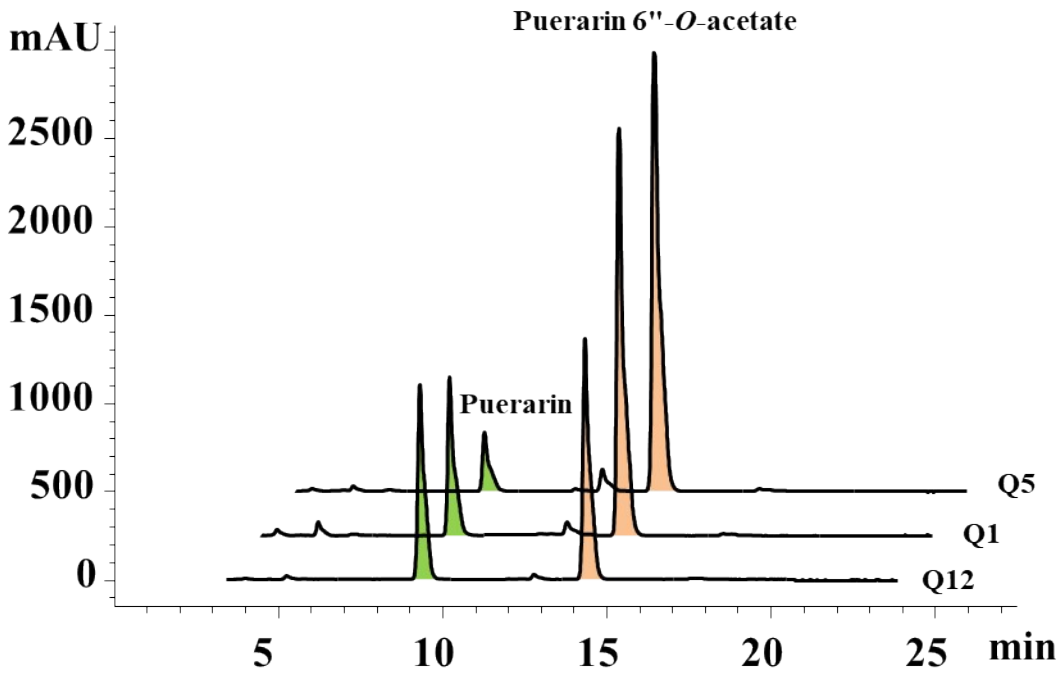

Figure S6 MAT-catalyzed acetylation of puerarin in the engineered strains.

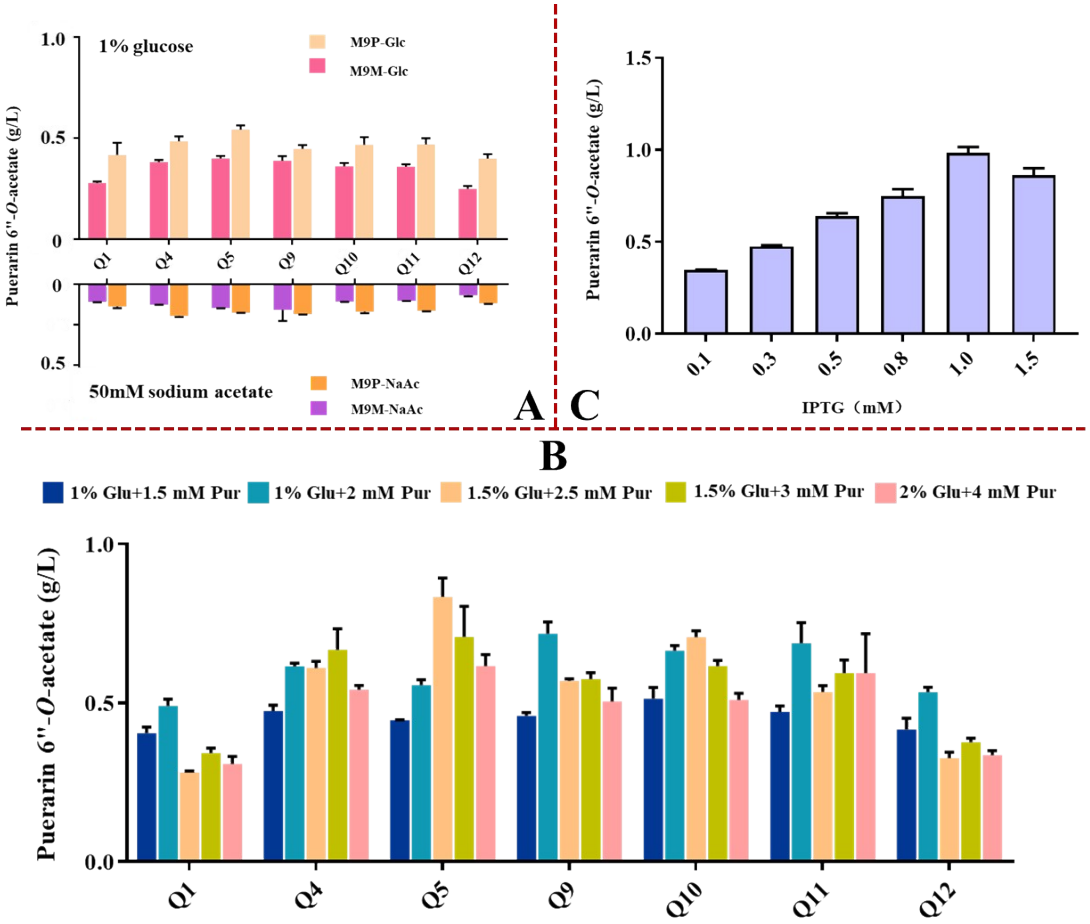

Figure S7 The condition optimizations to improve P6″A yield in the engineered cells. A, the effect of casamino acid on P6″A production; B, the effect of concentrations of puerarin (Pur) and glucose (Glc) on P6″A production; C, the effect of IPTG concentration on P6″A production.

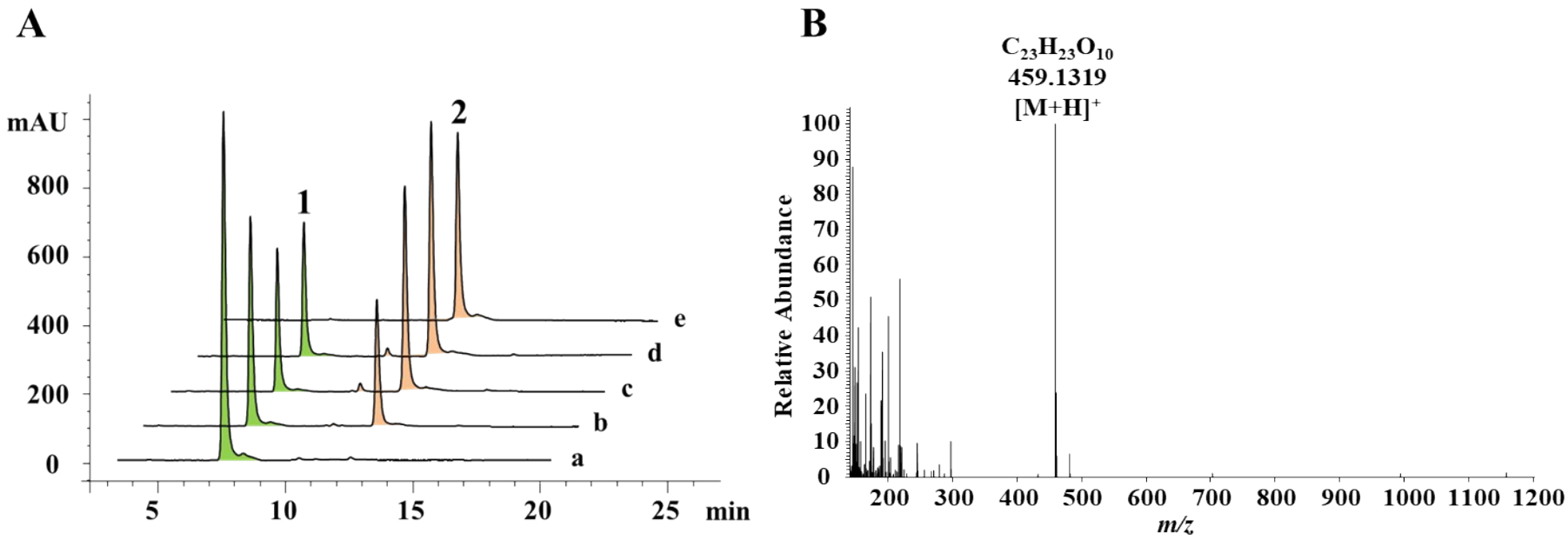

Figure S8 Acetyltransferase-catalyzed acetylation of puerarin. A, HPLC profiles of acetylated metabolites. a, acetylation assay without acetyltransferases; b, GAT-catalyzed acetylation; c, MAT-catalyzed acetylation; d, the co-elution of the reaction mixture and the authentical P6″A standard; e, P6″A standard; B, the MS spectrum of the metabolite 2. **1** and **2** represented puerarin and P6″A, respectively.

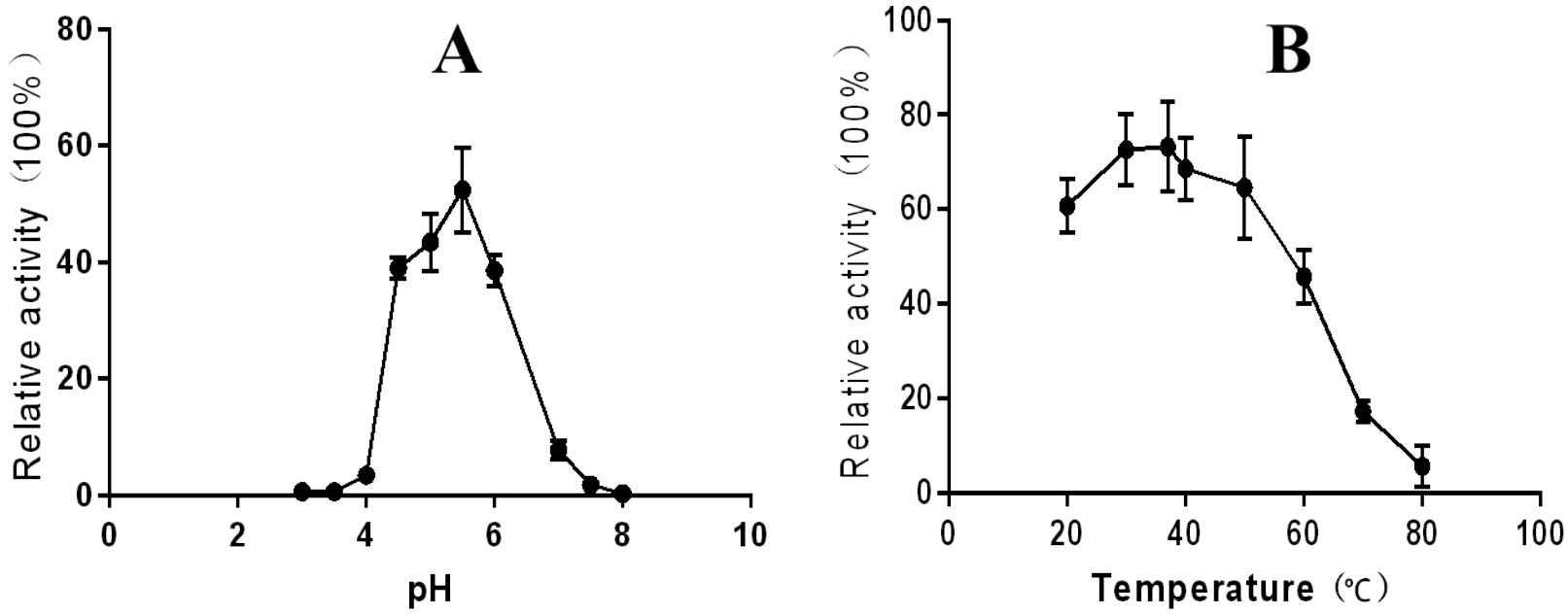

Figure S9 The effect of pH (A) and temperature (B) on the acetylation activity of GAT. Values are means ± SD, n = 3.

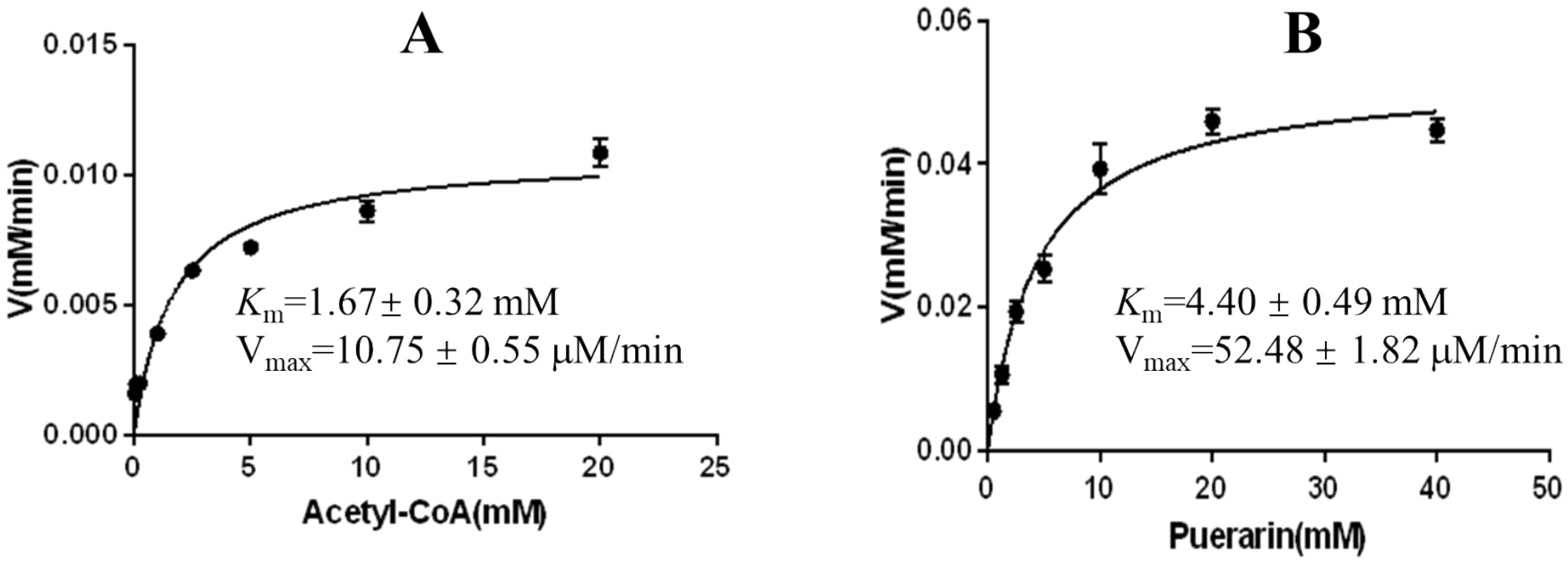

Figure S10 Kinetic parameters of GAT for acetyl-CoA (A) or puerarin (B). Values are means ± SD, n = 3.

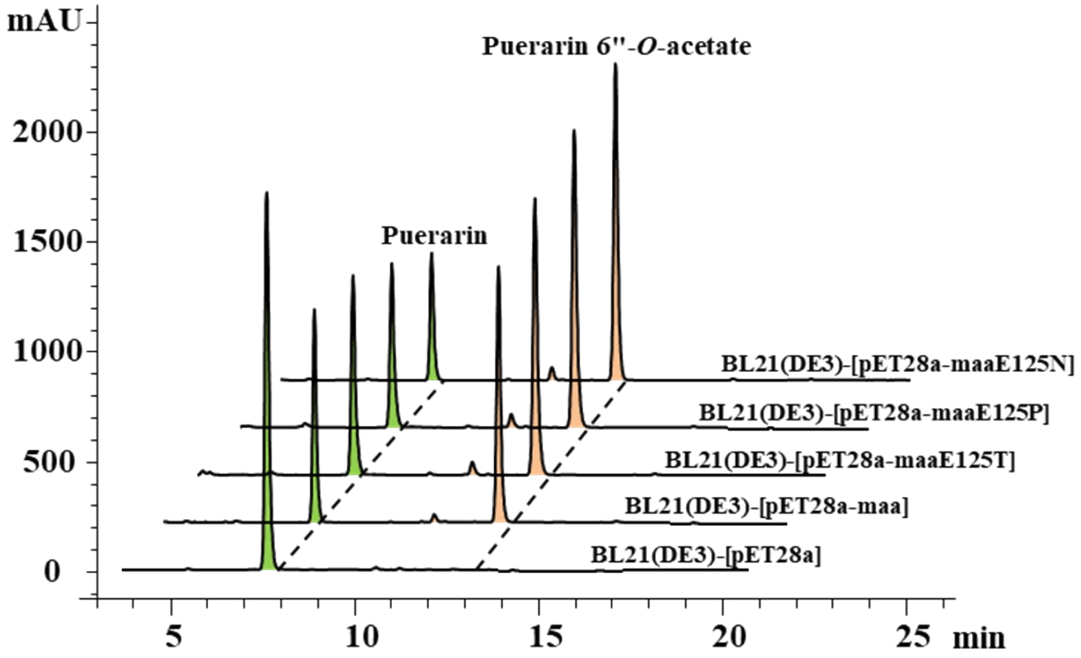

Figure S11 The whole-cell-catalyzed acetylation of puerarin to form P6″A.

Table caption

Table S1 Strains used in this study

Table S2 Plasmids used in this study

Table S3 Genes used in this study

Table S4 Primers used in this study

Table S1 Strains used in this study

Strain	Genotype	Source
*Trans*1-T1	F-φ80(lacZ)ΔM15ΔlacX74hsdR(rk^-, mk^+) ΔrecA1398 endA1tonA	Transgen
BL21(DE3)	F^- *omp*T *hsdS* ($r_B^-m_B^-$) *gal dcm* (DE3)	Transgen
BL21(DE3)-CoA	BL21(DE3)[pRSFDuet-acs + pCDFDuet-pdh]	This study
Q1	BL21(DE3)[pETDuet-maa]	This study
Q2	BL21(DE3)[pETDuet-maa+pRSFDuet-acs]	This study
Q3	BL21(DE3)[pETDuet-maa+pRSFDuet-acs-pank]	This study
Q4	BL21(DE3)[pETDuet-maa+pCDFDuet-pdh]	This study
Q5	BL21(DE3)[pETDuet-maa+pRSFDuet-acs+pCDFDuet-pdh]	This study
Q6	BL21(DE3)[pETDuet-maa+pRSFDuet-acs-pank+pCDFDuet-pdh]	This study
Q7	BL21(DE3)[pRSFDuet-maa+pACYCDuet-acs]	This study
Q8	BL21(DE3)[pRSFDuet-maa+pACYCDuet-acs-pank]	This study
Q9	BL21(DE3)[pRSFDuet-maa+pCDFDuet-pdh]	This study
Q10	BL21(DE3)[pRSFDuet-maa+pACYCDuet-acs+pCDFDuet-pdh]	This study
Q11	BL21(DE3)[pRSFDuet-maa+pACYCDuet-acs-pank+pCDFDuet-pdh]	This study
Q12	BL21(DE3)[pRSFDuet-maa]	This study
Q13	BL21(DE3)[pRSFDuet-maa+pACYCDuet-pdh-pank-acs]	This study
Q14	BL21(DE3) [pRSFDuet-maa+pCDFDuet-pdh-pank-acs]	This study
Q15	BL21(DE3)[pETDuet-maaE125N+pRSFDuet-acs+pCDFDuet-pdh]	This study

Table S2 Plasmids used in this study

Plasmids	Relevant properties or genetic marker	Source
pEASY-Blunt	A cloning vector	TransGen
pET-28a(+)	An expression plasmid developed for the protein expression in *E. coli*	Novagen
pACYCDuet™-1; pCDFDuet™-1; pETDuet™-1; pRSFDuet™-1	Duet plasmids, containing two multiple cloning sites (MCS), are designed for the co-expression of two target genes.	Novagen
pEASY-aceF; pEASY-aceE; pEASY-lpd; pEASY-pank; pEASY-acs	pEASY-Blunt derived plasmids, each containing an *E.coli* gene	This study
pET28a-lpd	A pET-28a(+) derived plasmid containing *lpd* gene	This study
pACYCDuet-acs; pCDFDuet-acs; pETDuet-acs; pRSFDuet-acs	Duet-derived plasmids containing *acs* gene	This study
pACYCDuet-acs-pank; pCDFDuet-acs-pank; pETDuet-acs-pank; pRSFDuet-acs-pank	Duet-derived plasmids, each containing both *acs* and *pank* genes	This study
pACYCDuet-aceF; pCDFDuet-aceF; pETDuet-aceF; pRSFDuet-aceF	Duet-derived plasmids containing *aceF* gene	This study
pACYCDuet-aceE/F; pCDFDuet- aceE/F; pETDuet- aceE/F; pRSFDuet- aceE/F	Duet-derived plasmids, each containing both *aceE* and *aceF* genes	This study
pACYCDuet-pdh; pCDFDuet-pdh; pETDuet-pdh; pRSFDuet-pdh	Duet-derived plasmids, each containing *aceE*, *aceF* and *lpd* genes	This study
pACYCDuet-pdh-pank-acs and pCDFDuet-pdh-pank-acs	Duet-derived plasmids containing pdh, pank and acs genes	This study
pET28a-lacA	A pET-28a(+) derived plasmid containing *lacA* gene	[1]
pET28a-maa	A pET-28a(+) derived plasmid containing *maa* gene	[1,2]
pACYCDuet-maa, pCDFDuet-maa, pETDuet-maa and pRSFDuet-maa	Duet-derived plasmids containing *maa* gene	
pET28a-maaE125N	A pET-28a(+) derived plasmid containing *maaE125N* gene	[2]
pET28a-maaE125T	A pET-28a(+) derived plasmid containing *maaE125T* gene	[2]
pET28a-maaE125P	A pET-28a(+) derived plasmid containing *maaE125P* gene	[2]

Table S3 Genes used in this study

Protein	CDS (aa)	gene	ORF (bp)	GenBank
ACS (acetyl-CoA synthetase)	652	*acs*	1959	NP_418493.1
PanK (type I pantothenate kinase)	316	*panK*	951	QNG34802.1
AceE (pyruvate dehydrogenase subunit E1)	887	*aceE*	2664	QNG31207.1
AceF (pyruvate dehydrogenase subunit E2)	630	*aceF*	1893	QNG31208.1
Lpd (dihydrolipoyl dehydrogenase)	474	*lpdA*	1425	QNG31209.1

Table S4 Primers used in this study

Primer	Sequence（5′-3′）	Description
acs-TF	ATGAGCCAAATTCACAAACACACC	A forward primer used for *acs* gene isolation
acs-TR	TTACGATGGCATCGCGATAGCCTG	A forward primer used for *acs* gene isolation
*ace*E-TF	ATGTCAGAACGTTTCCCAAATGACGTGG	A forward primer used for *aceE* gene isolation
*ace*E-TR	TTACGCCAGACGCGGGTTAACTTTATCT	A reverse primer used for *aceE* gene isolation
*ace*F-TF	ATGGCTATCGAAATCAAAGTACCGGACATCGGG	A forward primer used for *aceF* gene isolation
*ace*F-TR	TTACATCACCAGACGGCGAATGTCA	A reverse primer used for *aceF* gene isolation
lpd-TF	ATGAGTACTGAAATCAAAACTCAGGTCGTGG	A forward primer used for *lpdA* gene isolation
lpd-TR	ATGAGTACTGAAATCAAAACTCAGGTCGTGG	A reverse primer used for *lpdA* gene isolation
pank-TF	ATGAGTATAAAAGAGCAAACGTTAATGACGCC	A forward primer used for *pank* gene isolation
pank-TR	TTATTTGCGTAGTCTGACCTCTTCTACCGC	A reverse primer used for *pank* gene isolation
Duet-MAT-F	CACCACAGCCAGGATCCGAATTCGATGAGCACAGAAAAA	A forward primer to construct Duet-derived plasmids containing *maa* gene
Duet-MAT-R	TAAGCATTATGCGGCCGCAAGCTTTTACAATTTTTTAAT	A reverse primer to construct Duet-derived plasmids containing *maa* gene
acs-DuetF	GTATAAGAAGGAGATATACATATGATGAGCCAAATTCACAAACAC	A forward primer to construct Duet-derived plasmids containing *acs* gene
acs-DuetR	GCCGATATCCAATTGAGATCTTTACGATGGCATC	A reverse primer to construct Duet-derived plasmids containing *acs* gene
pank-DuetF	CCATCATCACCACAGCCAGGATCCGATGAGTATAAAAGAGCAA	A forward primer to construct Duet-derived plasmids containing *pank* gene
pank -DuetR	CAGGCGCGCCGAGCTCGAATTCTTATTTGCGTAGTCT	A reverse primer to construct Duet-derived plasmids containing *pank* gene
lpd-28aF	GCAAATGGGTCGCGGATCCGAATTCATGAGTACTGAAATC	A forward primer to construct pET28a-lpd
lpd-28aR	CTCGAGTGCGGCCGCAAGCTTTTACTTCTTCTTCGC	A reverse primer to construct pET28a-lpd
T7+*lpd*-DuetF	GCTGACGTCGGTACCCTCGAGAGATCTCGATCCCGC	A forward primer to construct Duet-derived plasmids containing *lpd* gene
T7+*lpd*-DuetR	CGGTGGCAGCAGCCTAGGTTAATTAATTACTTCTTCTTCGCTTTCGGG	A reverse primer to construct Duet-derived plasmids containing *lpd* gene
*ace*E DuetF	CATCATCACCACAGCCAGGATCCGATGTCAGAACGTTTCCC	A forward primer to construct Duet-derived plasmids containing *aceE* gene
*ace*E DuetR	AGGCGCGCCGAGCTCGAATTCTTACGCCAGACGCGG	A reverse primer to construct Duet-derived plasmids containing *aceE* gene

*ace*F DuetF	ATAAGAAGGAGATATACATATGATGGCTATCGAAATCAAAGTACC	A forward primer to construct Duet-derived plasmids containing *aceF* gene
*ace*F DuetR	GCCGATATCCAATTGAGATCTTTACATCACCAGACG	A reverse primer to construct Duet-derived plasmids containing *aceF* gene
pdh-acs-pank-DuetF	AAAGCGAAGAAGAAGTAATTAATTAAGCCATACCGCGAAAGGTT	A forward primer to construct pACYCDuet-pdh-pank-acs
pdh-acs-pank-DuetR	TCAGCGGTGGCAGCAGCCTAGGTTACGATGGCATCGC	A reverse primer to construct pCDFDuet-pdh-pank-acs

Reference

1 M. Liu, and J.-Q.Kong, *Acta Pharm Sin B*. **2018**, 8, 981-994.
2 X. N.Wang, L.L.Hong and J.Q.Kong, *J Agric. Food Chem*. **2021**, 69, 6623-6635.