

Clean synthesis of isoprene with six *E. coli* engineering strains

Yixuan Cui and Maurycy Daroch *

School of Environment and Energy, Peking University Shenzhen Graduate School, 2199 Lishui Rd., Shenzhen 518055, China

*Corresponding author e-mail: m.daroch@pkusz.edu.cn

Abstract. Isoprene is one of the most important platform chemicals containing conjugated double bonds and utilized as the precursor material of varieties of chemical raw materials. Currently, isoprene is mostly produced by petroleum smelting, which produces CO₂ and some pollution. In that case, the clean production method of isoprene is still the focus of current research. Isoprene synthase (IspS) catalyzes isoprene formation from a metabolic intermediate DMAPP. The expression level and enzyme activity of IspS from different strains affect the amount of isoprene production. In this study, six IspS protein sequences were obtained by searching literatures and NCBI database based on the phylogenetic analysis. We analysed the conservation of their catalytic structures and constructed pBAD expression frameworks for them. The isoprene synthesis systems in *E. coli* were built and worked in the best induction condition predicted in our study. We used toluene to collect the isoprene produced in the system detected and compared the productivity of each strain. Our results show that the production in cells were 1.55-5.51 µg/L/h.

Keywords: isoprene; biosynthesis; *E. coli*.

1. Introduction

Isoprene is the simplest isoprene family member, containing conjugated double bonds. It can copolymerize with a variety of compounds and be utilized as the precursor material of synthetic rubber, synthetic resin, isoprenoids medicines, spices binder [1-4]. It is also used in the production of isoprene drugs, spices and aviation fuels [5]. At present, most of the isoprene supplied in the market is from petroleum distillate [6]; the feedstock, crude oil, is non-renewable and ultimately unsustainable. This process may produce waste gas and large amounts of carbon dioxide resulting in the greenhouse effect or environmental pollution. Therefore, finding a sustainable way to produce isoprene has become an important research topic.

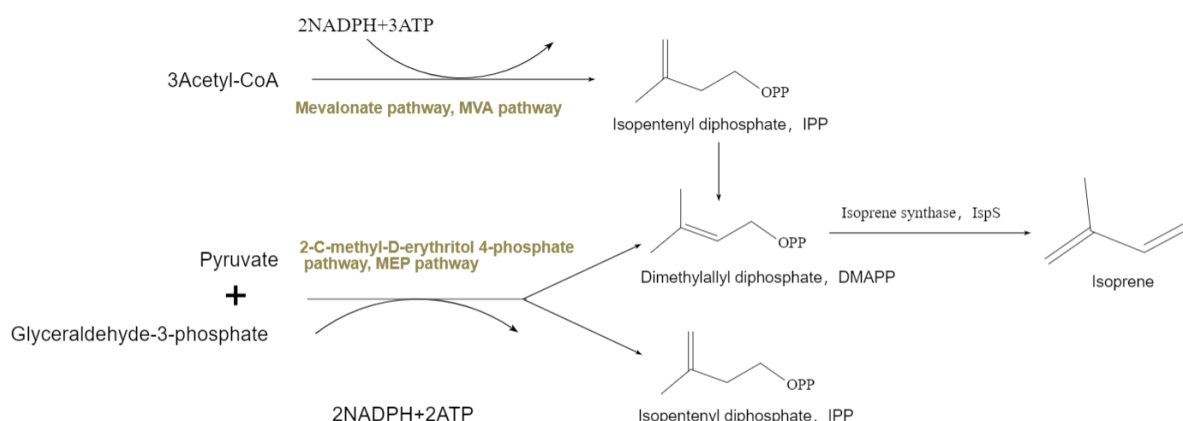


Fig. 1 Biosynthetic pathways of isoprene in microorganisms.

Biosynthesis of isoprene was first discovered in plants, which emit 1-2% fixed carbon in the form of isoprene [7]. There are two main pathways to synthesize isoprene precursors in organisms (Fig. 1). The first of these pathways is mevalonate pathway (MVA pathway), mainly found in eukaryotes, archaea and the cytoplasmic matrix of higher plants [8]. It starts from 3 molecules of acetyl-CoA and consumes 2 molecules of NADPH and 3 molecules of ATP catalysed by acetoacetyl-CoA thiolase, hydroxymethylglutaryl-CoA synthase, hydroxymethylglutaryl-CoA reductase and other four

enzymes. The other pathway is 2-C-methyl-D-erythritol 4-phosphate pathway (MEP pathway) [9]. It is present in most bacteria, green algae and the chloroplasts of higher plants, starting with glycolysis intermediates pyruvate and glyceraldehyde-3-phosphate. This reaction consumes 2 molecules of NADPH and 2 molecules of ATP and is catalysed by 1-deoxy-D-xylulose-5-phosphate synthase, 1-deoxy-D-xylulose-5-phosphate reductoisomerase and other six enzymes. Both pathways finally synthesize dimethylallyl diphosphate (DMAPP) or isopentenyl diphosphate (IPP), which are the precursor of isoprene and can be converted to isoprene with isoprene synthase (IspS).

Compared with traditional methods, biosynthesis of isoprene provides an attractive and potential alternative approach for the sustainable development of isoprene industries. At present, high purity isoprene synthesis systems have been successfully established in heterotrophic engineering bacteria such as *Escherichia coli* [10], *Saccharomyces cerevisiae* [11], *Bacillus subtilis* [12] and photosynthetic microorganisms such as cyanobacteria [13].

In this work, six IspS sequences from different species were selected based on the phylogenetic analysis. The structures of them were predicted with a homology-based approach. On this base, we conducted the expression of the six IspSs in *E. coli* BL21 using a low-copy, arabinose-inducible vector, pBAD_LIC_cloning vector (8A) and analysed the productivity of isoprene in *E. coli* cells.

2. Materials And methods

2.1 Protein sequence selected and structural analysis

Thirty-four amino acid sequences labelled as isoprene synthase were obtained from NCBI database (<https://www.ncbi.nlm.nih.gov/>) and literatures. A phylogenetic tree was constructed by Maximum Likelihood method.

According to the phylogenetic analysis results, six amino acid sequences from different strains were selected for further analysis. MEGA-X software (Available at <https://www.megasoftware.net/>) [14] was used to perform a bootstrap tree with Maximum Likelihood method and JTT matrix-based model [15] for phylogenetic analysis after aligning sequences using ClustalW with default parameters. The bootstrap consensus tree was inferred from 1000 replicates to represent the evolutionary relationship of IspS [16]. These sequences were also aligned in Geneious Prime (Available at <https://www.geneious.com/resources/#downloads>). Structural models were built by SWISS MODEL (<https://swissmodel.expasy.org/>) with a defined IspS structure of 3n0g.1.

2.2 Strains, plasmids and culture conditions

Plasmids were constructed in *E. coli* DH5 α and expressed in *E. coli* BL21 (ktsm-life, China). pBAD_LIC_cloning vector (8A) was used as the base of expression framework. The plasmid was a gift from Scott Gradia (Addgene plasmid #37501; RRID: Addgene_37501). The plasmid contained araBAD arabinose promoter and ampicillin resistance gene.

E. coli was grown on LB medium or agar plates at 37 °C in an incubator (HZQ-X300C, Yiheng, Shanghai, China). Further, 50-100 μ g/mL of ampicillin was added to maintain plasmids, while arabinose was used to induce the expression of IspS genes.

To express proteins and to produce isoprene, overnight cell culture was diluted into LB in a ratio of 1: 50, cultured at 37 °C for 180 rpm until it grew to the concentration of OD₆₀₀=0.4-0.6. Then, arabinose was added to the final concentration of 0.2% and the temperature was adjusted to 30 °C, shaking for 5 hours.

OD₆₀₀ of the culture was used for cell density measurement with an EPOCH micro-plate reader (BIOTEK, USA).

2.3 Construction of the recombinant plasmids

The protein sequences were obtained from NCBI and literature, reverse translated and synthesised by BGI Write (Guangdong, China). The recombinant plasmids and primers for construction used in

this study are shown in Table 1. A ClonExpressII One Step Cloning Kit (Vazyme, China) was used to combine *IspS* genes to EcoRV-predigested pBAD_LIC_cloning_vector. Primers for the assembly were designed with 15-20 bp overhangs (under-lined).

After constructing, the recombinant plasmids were transformed to *E. coli* DH5 α and ultimately to *E. coli* BL21 by the calcium ion transformation method.

Table 1. Recombinant plasmids and primers for them.

Plasmids	Orientation	Sequences of oligonucleotides
pBAD- <i>IspS</i> _PM	F	<u>ACTTTAAGAAGGAGAT</u> ATAGATATGTGTGCCACCAGTAGTCAATT
	R	TCCTTATGGAGTTGGGAT <u>CACATACATCAGTTGATT</u>
pBAD- <i>IspS</i> _PC	F	<u>ACTTTAAGAAGGAGAT</u> ATAGATATGTGTAGTGTGAGTACCGAAAAATG
	R	TCCTTATGGAGTTGGGAT <u>TGCGTTCAAAGGGCAG</u>
pBAD- <i>IspS</i> _MI	F	<u>ACTTTAAGAAGGAGAT</u> ATAGATATGTGTAGTGTGCCACCACCCAAAAT
	R	TCCTTATGGAGTTGGGATTTGACTGGCCACTTTGG
pBAD- <i>IspS</i> _IB	F	<u>ACTTTAAGAAGGAGAT</u> ATAGATATGTGTCTGGCCAGTAGTGCCC
	R	TCCTTATGGAGTTGGGATTTCCACGGGATTAATAATC
pBAD- <i>IspS</i> _EG	F	<u>ACTTTAAGAAGGAGAT</u> ATAGATATGTGTAGTGTGCCAGTACCCAAATTAG
	R	TCCTTATGGAGTTGGGATGGCGGCGGGACTAATG
pBAD- <i>IspS</i> _EP	F	<u>ACTTTAAGAAGGAGAT</u> ATAGATATGTGTCTGGCCAGTCGCCAAAC
	R	TCCTTATGGAGTTGGGATCAGACTAATGGGTTCC

2.4 Real-time RT-PCR analysis

1.5 mL *E. coli* cells were collected at 4 ° C to isolate RNA with Total RNA kit I (Omega bio-tec, USA). Prime-Script™ RT reagent Kit with gDNA Eraser (Takara, Japan) and TB Green Premix Ex Taq II (TaKaRa) was used for gDNA erasing, reverse transcription and RT-qPCR analysis in a QuantStudio 5 real-time system (ABI, USA). Primers for qPCR of *IspS* gene are: F: 5'-TTTCGCCTGCTGCGCCAACA-3' and R: 5'-CCAGGTGACTGGCTTCATAC-3'. Standard curves were finished with 10-fold serial dilution of the standard fragments (137 bp, amplified by qPCR primers) to verify the PCR efficiency and to calculate the RNA gene copies. A fragment of 16S RNA in *E. coli* was used as the reference gene for RT-qPCR. Its primers are F: 5'-ACTCCTACGG-GAGGCAGCAG-3' and R: 5'-ATTACCGCGGCTGCTGG-3'.

2.5 SDS-PAGE analysis and protein purification

Cells were collected by centrifugation and suspended again in phosphate buffer (PBS). Cells were broken in ice using an Ultrasonic Cell Disruptor (SCIENTZ-II D, China). Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse the protein in supernatant and pellet. The proportion of target protein in it was calculated by ImageJ with its integrated density. A protein MS Q-E test (BGI, China) was used to verify the results.

2.6 Isoprene production in cells

$$\text{Isoprene production activity} = W/(V_a \cdot T) \quad (1)$$

W: the content of isoprene in the bottle (μg); V_a : the volume of cell culture (mL); T: reaction time (h), 12 h;

Isoprene production in the gas phase was analysed in a headspace of a 60 mL headspace bottle with 15 mL cell culture. Toluene was injected into the bottle to absorb isoprene after incubating for 12 h and cooling overnight at 4 ° C to make isoprene present as liquid. Agilent Technology 6850N-5975 GC-MS with HP-5MS 5% Phenyl Methyl Siloxa column was used to measure the isoprene. Different gradient volumes of liquid isoprene were added into the bottle and detected with GC-MS to make a standard curve. Three repeat groups for each strain were prepared for measuring, while one

repeat group was for cell density measurement with OD600. The efficiency of isoprene production was calculated as follows:

3. Results and discussions

3.1 Protein sequence selected and structural analysis

Thirty-four amino acid sequences of IspSs from Magnoliopsida were extracted from NCBI and literatures. A ML-Bootstrap-phylogenetic tree representic phylogenetic relationships between the proteins was constructed and shown in Fig. 2. These sequences are source from Fabids, Vitals (orange branch) and Malvids (purple) of Rosids (rose-red branch) and Asterids (yellow branch).

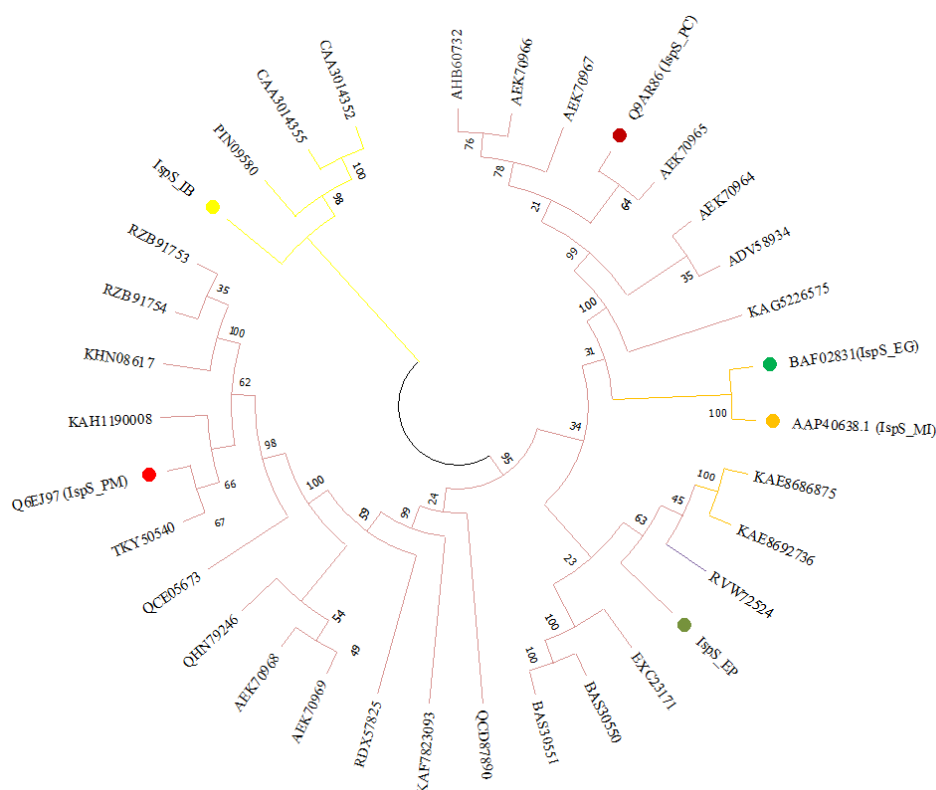


Fig. 2 ML-Boostrap-phylogenetic tree of IspS-homologous sequences (The sequences used in the next step are marked with different colours of points).

As representatives of the branches, we selected six sequences from different strains (marked with points): Red point: IspS_PM: *P. montana* var. *lobata* (Uniprot: Q6EJ97); Yellow point: IspS_IB: *I. batatas* [17]; Rose red point: IspS_PC: *P. canescens* (Uniprot: Q9AR86); Orange point: IspS_MI: *M. alternifolia* (GenBank: AAP40638); Green point: IspS_EP: *E. photiniifolius* [17]; Dark cyan point: IspS_EG: *E. globulus* (GenBank: BAF02831), for further analysis.

Sequence alignment and structural modelling were performed with 3n0g.1. SWISS-MODEL shows that these models have a high matching degree with parameters like GMQE, QMEANDisCo Globa and QMEAN.

Comparison results of six proteins is shown in Fig. 3. There were 30 groups of α -helix structure and 2 groups of β -strand structure in the model, and some small differences in α -helix structure and β -strand structure can be seen among them. The structure comparison results of SWISS-MODEL were shown in Fig. 4: the β -strand structures of IspS_PM, IspS_IB, IspS_EG and IspS_EP models were the same with the template, and only two groups of β -strand structures were found in the sites of D.528 and G.531 for IspS_MI. The α -helix structure of all the models was different from that of the template. The main change sites were Y.27, D.28, F.29, L.30, L.31, G.80, Y.81, E.82, F.83, P.130,

E.184, E.185, K.186, I.187, R.220, K.221. The models are mainly blue, indicating that the confidence level of the models is high and the changes are not significant. IspS structure 3N0G.1 mainly contains dimethylallyl S-thiolodiphosphate binding site and magnesium ion binding site, which are located in the site of R.268, F.298, D.305, D.309, F.380, E.383, S.405, S.406, F.445, R.446, N.449, E.457, A.464. There are some non-conserved changes in R. 268 (IspS_PC: K.272) and S.406 (IspS_IB: V.405, IspS_EP: V.405) at residues of the dimethylallyl S-thiolodiphosphate binding site. And there are non-conserved changes in A.464 (IspS_PM: T.468, IspS_EG: T.469) at residues of the magnesium ion binding site. Since all of them are in the residues, they may have little effect on the activity of the ligand.



Fig. 3. Amino acid sequence alignment (α -helix: green box, β -strand: blue arrow, activity sites: blue box) between the six IspSs.

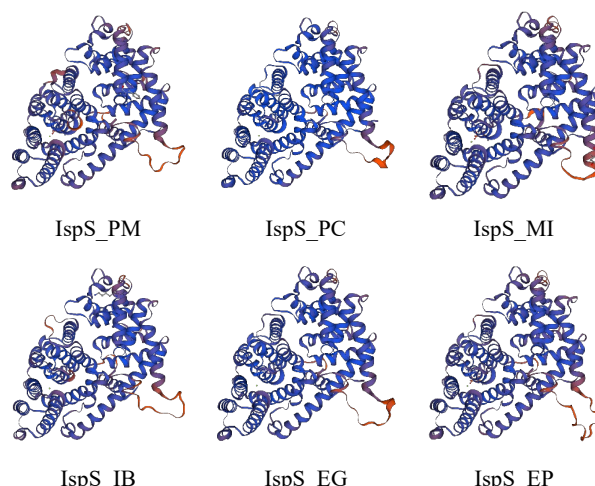


Fig. 4 The structural models of the six IspSs. (Structures are shown by confidence level: blue→red: high confidence→low confidence)

3.2 Expression of Isoprene synthetases in E. coli BL21

Six pBAD-IspS+ expression frameworks were constructed using the homology-based method in E. coli DH5 α and transformed to BL21 for expression. To trace the growth of each E. coli BL21_IspS+ strains OD600 were measured. As is shown in Fig. 5, the growth curves of BL21_IspS+ and wild type BL21 (WT) were similar. It indicated that the insertion of expression frameworks had no specific effect on the growth of E. coli cells.

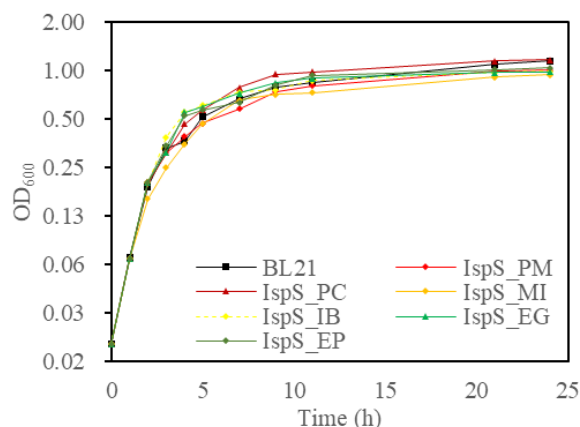


Fig. 5 Growth curve of the six engineering strains in LB with 2% arabinose and 50 µg/mL ampicillin.

3.3 Improvement of the induction environment

RT-qPCR and SDS-PAGE were used to determine the induction environment by using BL21_IspS_PC+ as the representative strain. Different concentrations of arabinose (0, 0.02%, 0.2%, 2%) and temperatures (20 °C, 30 °C) were set in cultivation system as the variables.

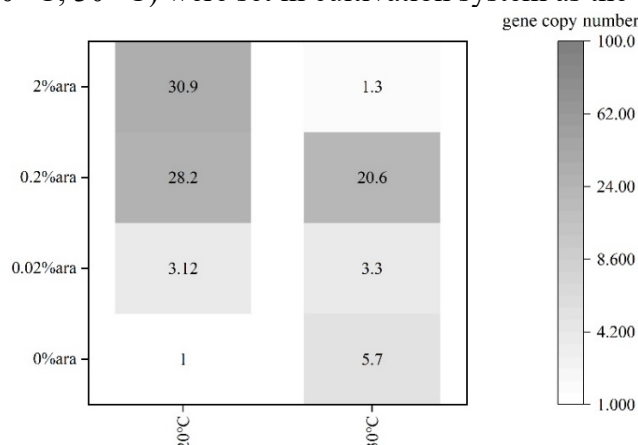


Fig. 6 RT-qPCR analysis of RNA level in different levels of arabinose and temperatures.

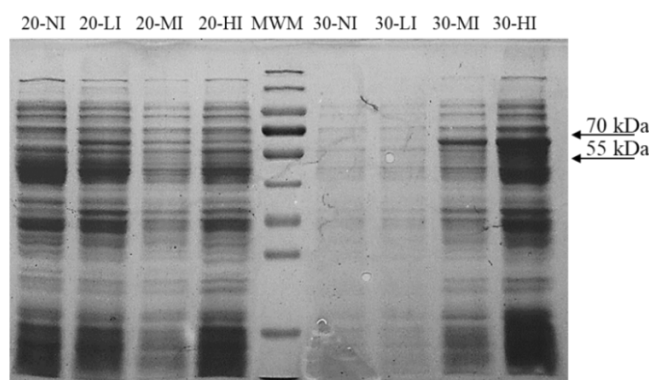


Fig. 7 SDS-PAGE of protein expression in different level of arabinose (0, 0.02%, 0.2%, 2%) and temperatures (20 °C, 30 °C): MWM: protein marker; 20-NI: 20 °C 0% ara: 3.8%; 20-LI: 20 °C 0.02% ara: 7.0%; 20-MI: 20 °C 0.2% ara: 6.3%; 20-HI: 20 °C 2% ara: 4.4%; 30-NI: 30 °C 0% ara: 4.3%; 30-LI: 30 °C 0.02% ara: 4.5%; 30-MI: 30 °C 0.2% ara: 9.8%; 30-HI: 30 °C 2% ara: 4.2%.

According to the results, mRNA gene copy number of IspS gene was the highest under the condition of 2% arabinose at 20 °C (the relative gene copy number was 24.51), while its protein level was low at this level, which may be partly affected by factors in the mRNA translation process.

However, better protein expression was shown at the concentration of 0.2% arabinose at 30°C, and the mRNA level was relatively high at this condition (the relative gene copy multiple was 20.6). Therefore, the concentration of 0.2% arabinose at 30 °C was set as the expression condition of *ispS* gene.

3.4 Heterologous expression of BL21_*IspS*⁺ strains

In this experiment, six BL21_*IspS*⁺ strains were induced in LB with 50 µg/mL ampicillin and 0.2% arabinose at 30 °C for 5 hours.

RT-qPCR was used to test the mRNA level. Standard curves were drawn to calculate the absolute gene copy numbers. Their PCR efficiencies were around 90%-110%, proving that primer and standard curves were reliable. Then, the ratio of the absolute gene copy number of *IspS* genes to the absolute gene copy number of the 16sRNA gene was calculated to get the relative gene copy number which is shown in Fig. 8. The results show that the mRNA level of *IspS*_MI was slightly lower than others (about 1/4 of that of other genes). But if it will influence the work of this strain should be adjusted in the following experiments.

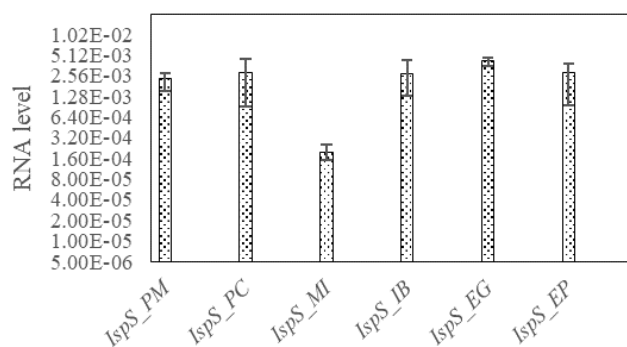


Fig. 8 Relative gene copy number of six *IspS* genes

SDS-PAGE was also used to analyse the protein production in the crude extract. The bands of *IspS* protein were around 64-66 kDa. These gel sections were cut and sent for a protein MS Q-E test to verify the results. SDS-PAGE analysis of the insoluble fraction showed the formation of inclusion bodies. It may be caused by excessive protein concentration or challenges in the folding process of eukaryotic proteins. These challenges may have influenced subsequent isoprene's productivity and need to be improved.

3.5 Isoprene production in cells

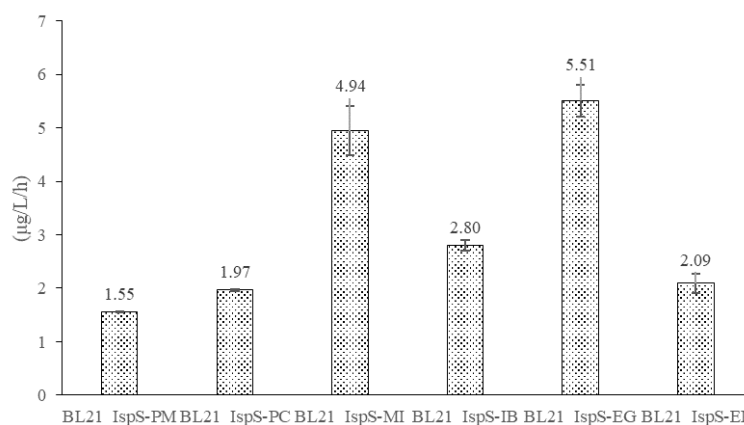


Fig. 9. Isoprene production of BL21_*IspS*⁺ engineering strains in LB with 0.2% arabinose under 20 °C

Because isoprene is insoluble in water, soluble in organic solvents, and has a low boiling point, we use toluene to collect the isoprene produced in isoprene-biosynthesis system. The isoprene production results of six isoprene synthase engineered strains for 12 hours are shown in Fig. 9.

BL21_IspS_MI+ and BL21_IspS_EG+ had higher productivity of isoprene (5.51 and 4.94 $\mu\text{g/L/h}$), which indicates that the lower mRNA level of IspS_MI may not influence the work of the IspS in cells or the enzyme activity is sufficient to compensate for this deficiency. The production efficiency of BL21_IspS_PM+ is 1.55 $\mu\text{g/L/h}$. Compared to the research of Zurbriggen et al. using high copy plasmid pGEX6P1 with T7 promoter, this efficiency was lower (18 h induction, efficiency of 22 $\mu\text{g/L/h}$) [10]. This may be due to the low copy plasmid pBAD used in this experiment or the loss of isoprene during toluene adsorption process.

4. Conclusion

In this experiment, we selected six isoprene synthases through phylogenetic analysis, modelling their structures and analysing the conservation among them. Six IspS genes were expressed in *E. coli* BL21 under the control of araC and araBAD promoters using the low-copy vector pBAD LIC. We found that some folded structures and ligand sites were not conserved in their model, but the main framework of these six proteins was roughly similar. RT-qPCR results showed that the mRNA level of IspS_MI was lower (about 1/4 of that of other genes), and the others were basically the same. However, isoprene production in BL21_IspS_MI+ cell was higher than others, indicating it may not influence the expression and isoprene production. Since pBAD is a low copy plasmid, it may be worth using high copy plasmids or adding precursors in medium to improve the expression level or improve the isoprene yield.

Acknowledgements

This research was funded by Shenzhen Fundamental Research Program (GXWD20201231165807007-20200806170221001).

References

- [1] J. R. King, B. M. Woolston and G. Stephanopoulos. *ACS Synth Biol* 6, 1416-1426 (2017).
- [2] L. Ye, X. Lv and H. Yu. *Metabolic Engineering* 38, 125-138 (2016).
- [3] V. D. Deepak, I. Mahmud and M. Gauthier. *Polymer Journal* 51, 327-335 (2019).
- [4] J. R. King, B. M. Woolston and G. Stephanopoulos. *ACS Synth Biol* 6, 1416-1426 (2017).
- [5] A. Alianell Gary, F. Derwitsch, D. Wells and T. Taylor. *Isoprene Compositions and Methods of Use* (2009).
- [6] O. S. Pavlov, S. A. Karsakov and S. Y. Pavlov. *Theoretical Foundations of Chemical Engineering* 45, 487 (2011).
- [7] J. Kesselmeier and M. Staudt. *Journal of Atmospheric Chemistry* 33, 23-88 (1999).
- [8] J. Yang, Q. Nie, H. Liu, M. Xian and H. Liu. *BMC Biotechnol* 16, 5 (2016).
- [9] T. Kuzuyama. *Bioscience, Biotechnology, and Biochemistry* 66, 1619-1627 (2002).
- [10] A. Zurbriggen, H. Kirst and A. Melis. *BioEnergy Research* 5, 814-828 (2012).
- [11] O. S. Pavlov, S. A. Karsakov and S. Y. Pavlov. *Theoretical Foundations of Chemical Engineering* 45, 487 (2011).
- [12] W. P. Wagner, D. Helmig and R. Fall. *Journal of Natural Products* 63, 37-40 (2000).
- [13] F. K. Bentley and A. Melis. *Biotechnology and Bioengineering* 109, 100-109 (2012).
- [14] S. Kumar, G. Stecher, M. Li, C. Knyaz and K. Tamura. *Mol Biol Evol* 35, 1547-1549 (2018).
- [15] D. T. Jones, W. R. Taylor and J. M. Thornton. *Comput Appl Biosci* 8, 275-82 (1992).
- [16] J. Felsenstein. *Evolution* 39, 783-791 (1985).
- [17] M. Ilmén, M. Oja, A. Huuskonen, S. Lee, L. Ruohonen and S. Jung. *Metab Eng* 31, 153-62 (2015).