Supplementary Information

Desymmetric Hydrolysis of Prochiral Imide for *S*-Pregabalin Synthesis by Rationally Designed D-Hydantoinase

Feixia Liu[a†], Jie Ren[b†], Lu Guo[c], Yang Liu[a,d], Defeng Li[c], Bo Yu[a*]

[a] *CAS Key Laboratory of Microbial Physiological & Metabolic Engineering, State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China*

[b] *State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100029, China*

[c] *State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, PR China*

[d] *University of Chinese Academy of Sciences, Beijing 100049, China*

† F.L. and J.R. contributed equally to this study.

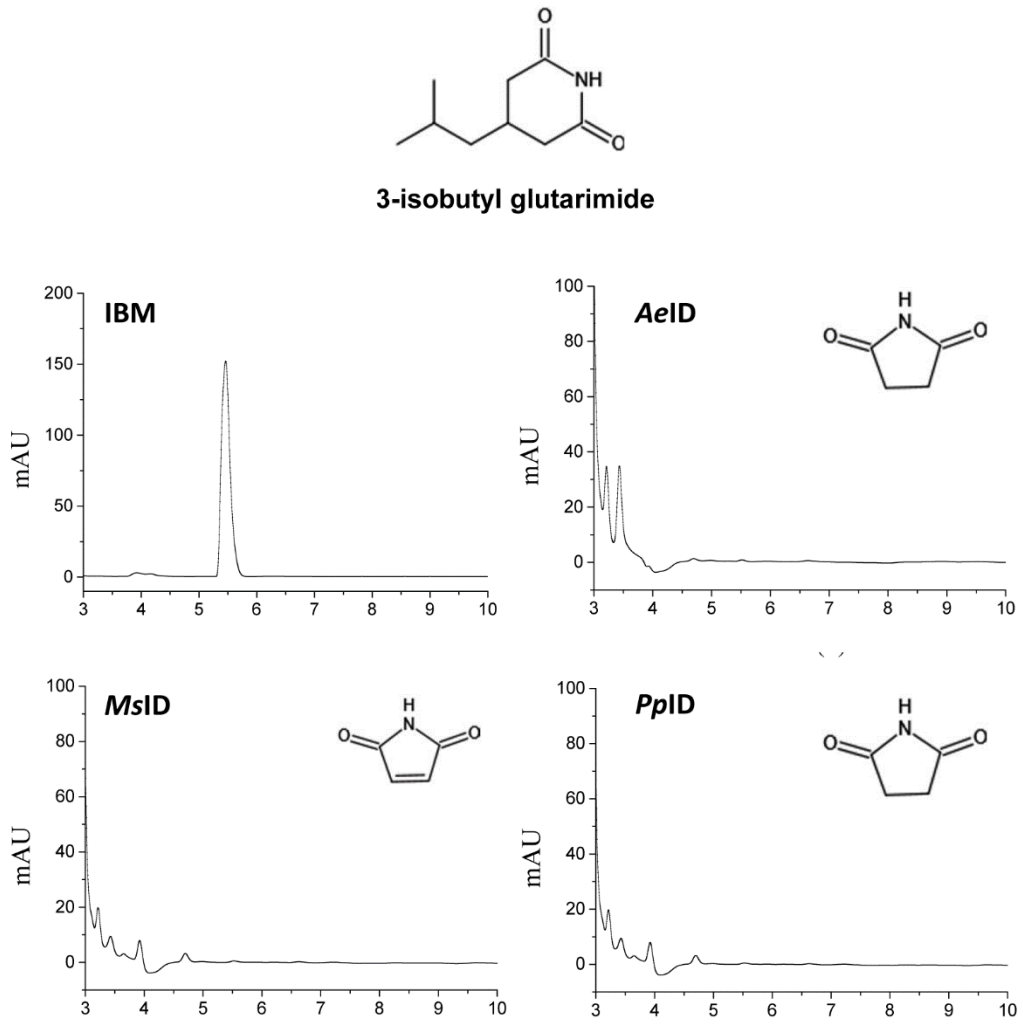
* Corresponding author.

E-mail: yub@im.ac.cn (B. Yu)

Phone/Fax: +86-10-64806132

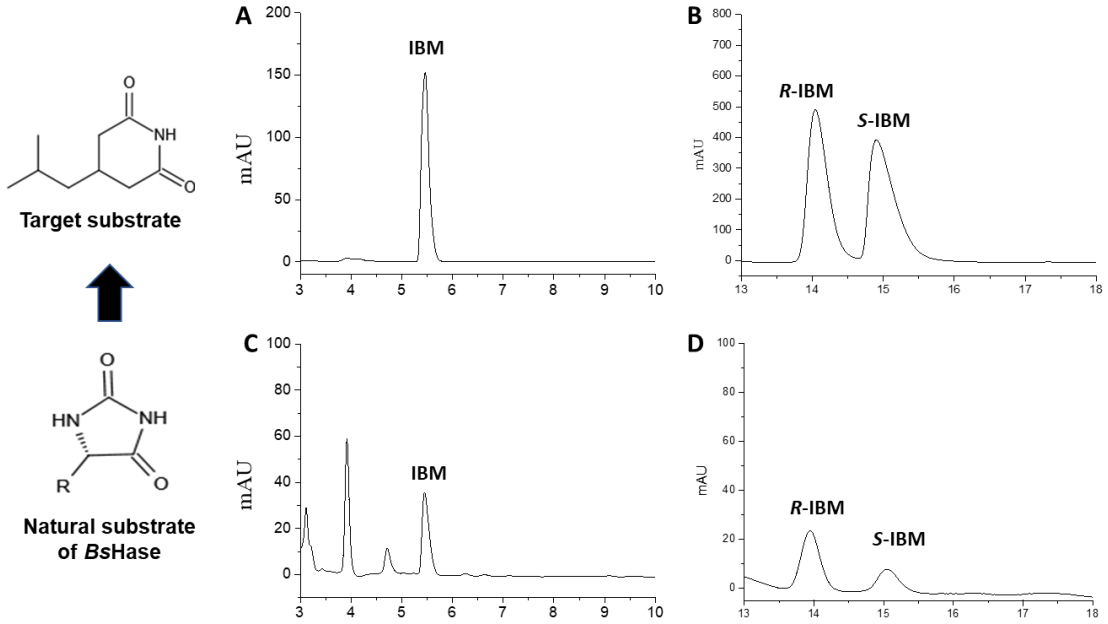
ORCID: 0000-0003-3148-0986

Fig. S1 Activity tests towards 3-isobutyl glutarimide by selected imidases.



IBM, the strandard of target product 3-isobutyl glutaric acid monoamide; *Ae*ID, the imidase gene from *Alcaligenes eutrophus* 112R4 (GenBank No. AAK53437.1); *Ms*ID, imidase gene from *Mycolicibacterium smegmatis* (GenBank No. WP_011727598.1) and *Pp*ID, imidase gene from *Pseudomonas putida* YZ-26 (GenBank No. AAY98498.1). The natural substrate structures of the selected enzymes were shown in the respective figures. The target substrate structure in this study, 3-isobutyl glutarimide, was also indicated.

Fig. S2 Activity test towards 3-isobutyl glutarimide by D-hydantoinase from *Bacillus stearothermophilus* SD-1.



(A) HPLC spectrum of standard of 3-isobutyl glutaric acid monoamide (IBM) analyzed by C18 column; (B) spectrum of standards of *R*-IBM and *S*-IBM analyzed by Chiralpak AD-H column; (C) product from 3-isobutyl glutarimide catalyzed by wide-type D-hydantoinase from *Bacillus stearothermophilus* SD-1 (*Bs*Hase) (GenBank No. QFU78355.1); (D) the chirality of produced IBM by wild-type *Bs*Hase.

Fig. S3 Michaelis-Menten-plots for the reaction of the wild-type and the mutants.

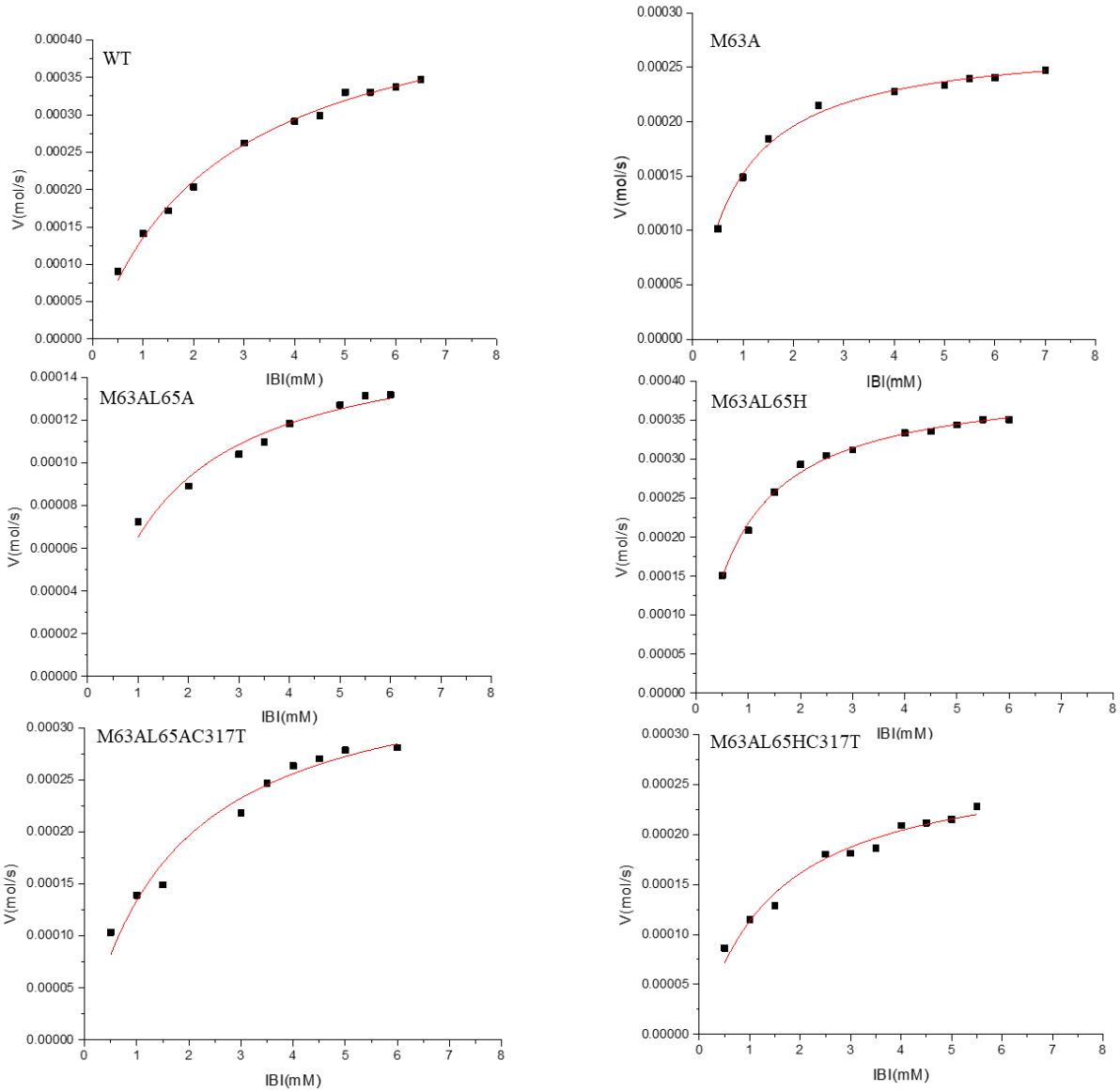
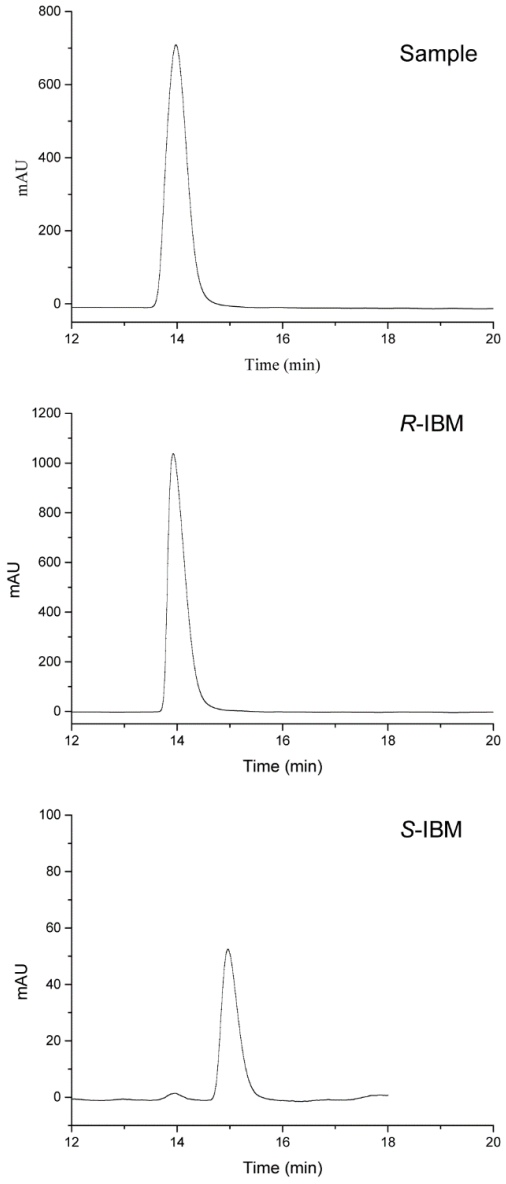


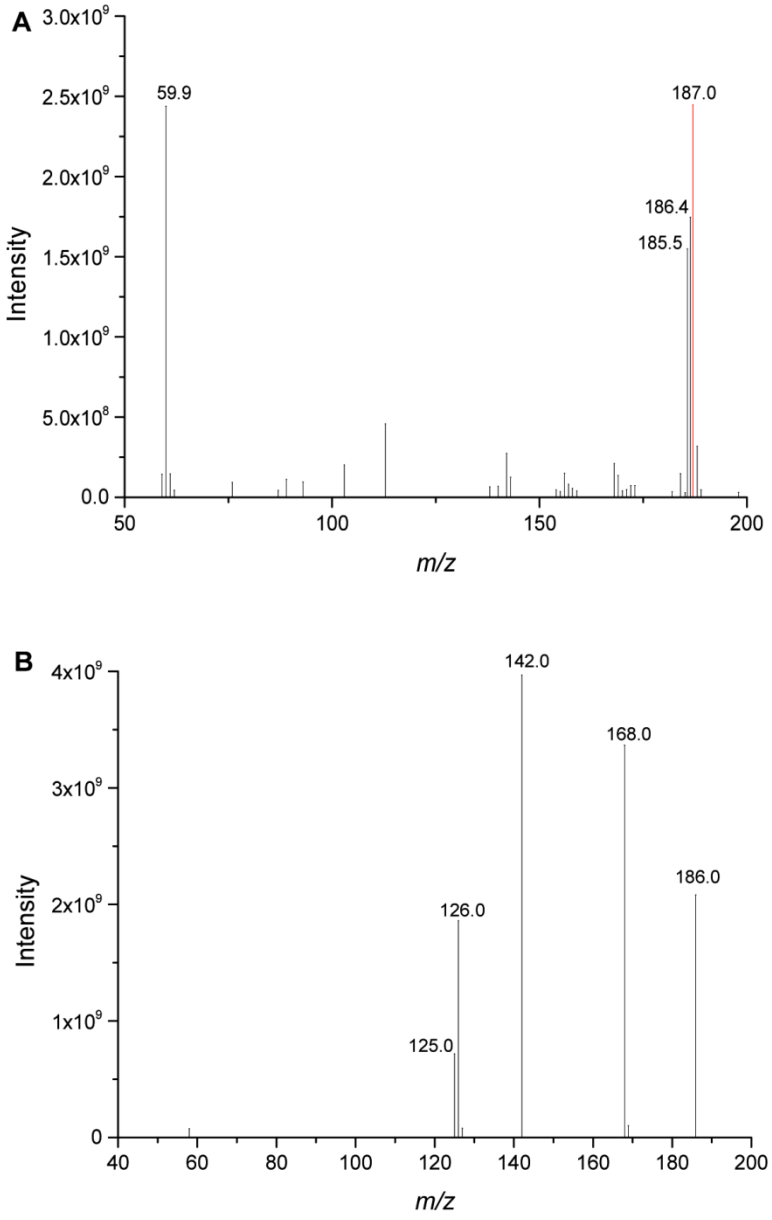
Fig. S4 The chirality of produced *R*-IBM by engineered *Bs*Hase with triple mutantion (M63AL65HC137T).



Sample, HPLC analysis spectrum of *R*-IBM product in the broth;

R-IBM and *S*-IBM, the standards.

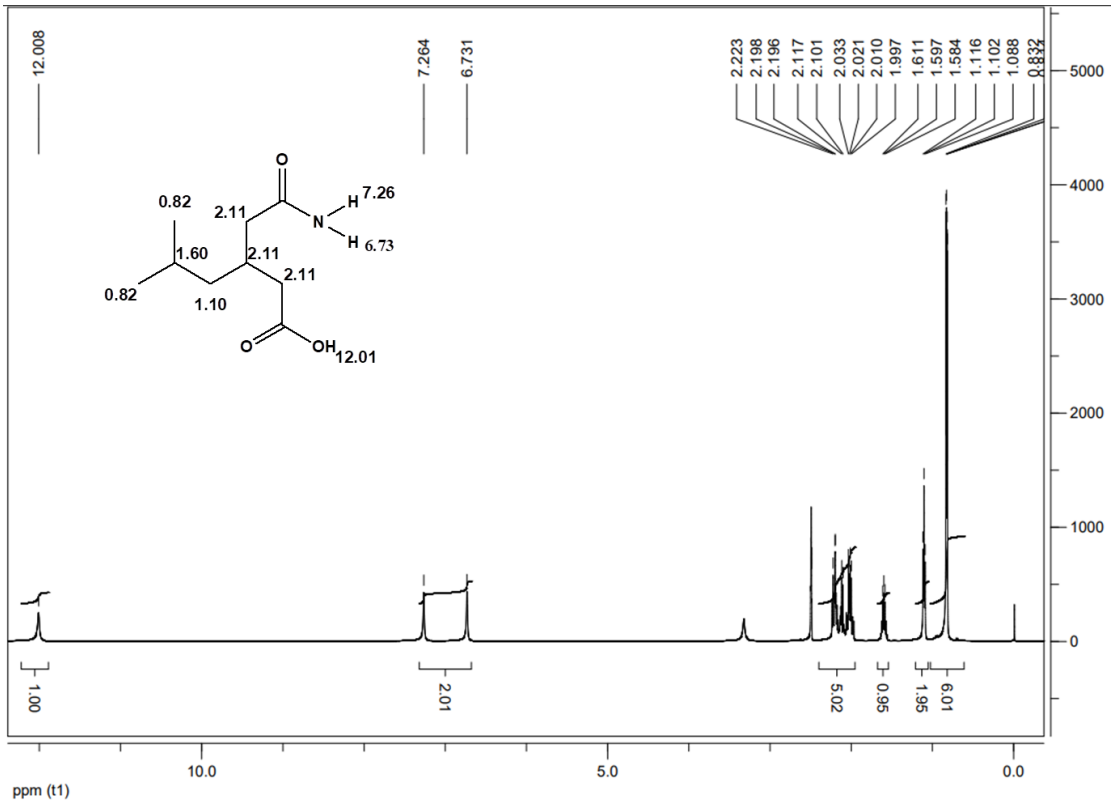
Fig. S5 Mass spectrometry analysis of produce *R*-IBM.



(A) Calculated value for *R*-IBM $[M-H]^-$ form (negative mode) was *m/z* 187.

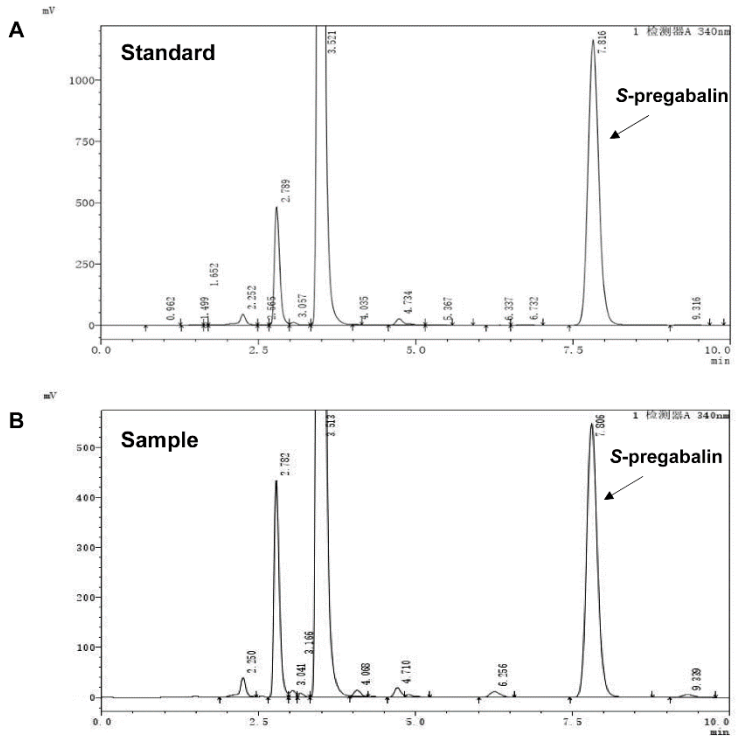
(B) MS-MS spectra for *R*-IBM.

Fig. S6 ^{1}H NMR spectrum of purified *R*-IBM.



(DMSO-D6), δ:. 0.82 (dd, 6H, $^{2}J_{H–H}$ =7.8 Hz, $^{3}J_{H–H}$ = 3.6 Hz), 1.10 (t, 2H, $^{2}J_{H–H}$ =8.4 Hz, $^{3}J_{H–H}$ =8.4 Hz), 1.60 (m, 1H, $^{3}J_{H–H}$ =8.4 Hz), 2.11 (m, 5H, $^{2}J_{H–H}$ =13.8 Hz, $^{3}J_{H–H}$ = 7.8 Hz), 6.73 (s, 1H), 7.26 (s, 1H), 12.01 (s, 1H).

Fig. S7 HPLC spectrum of *S*-pregabalin prepared from *R*-IBM obtained in this study.



(A) The authentic standard of commericial *S*-pregabalin;

(B) The *S*-pregabalin sample prepared from the *R*-IBM obtained in this study.

The peak at retention time of 7.81 min is *S*-pregabalin derivatized with 1-fluoro-2-4-dinitrophenyl-5-L-alanine amide (FDAA) and the peaks at 2.78 and 3.52 min are from the derivative reagent of FDAA.

Experimental

1. Genes and materials

Genes from *Bacillus stearothermophilus* SD-1 (GenBank No. QFU78355.1), *Alcaligenes eutrophus* 112R4 (GenBank No. AAK53437.1), *Mycolicibacterium smegmatis* (GenBank No. WP_011727598.1), *Pseudomonas putida* YZ-26 (GenBank No. AAY98498.1) were synthesized with optimized codons based on *E. coli* expression preference (GenScript Co., Ltd, China). *E. coli* BL21(DE3) was used as expression host and *E. coli* TOP10 was used as cloning hosts. The PCR and seamless cloning kit was purchased by Vazyme Biotech Co., Ltd., China. 3-Isobutyl glutarimide (IBI), *R*-3-isobutyl glutaric acid monoamide (*R*-IBM), *S*-IBM and the other reagents used were analytic grade and commercially available.

2. Screening for imide hydrolases

Four candidate genes were synthesized and subcloned into plasmid pET-28a and transformed into *E. coli* BL21(DE3). The recombinant strains were cultured at 37 °C and 200 rpm in Luria-Bertani (LB) medium supplemented with a final concentration of 40 μg·mL^{-1} kanamycin, then were induced by 0.5 mM IPTG at 24 °C for 12 h when OD 600 nm reached at 0.6 to 0.8. The whole cells were harvested by centrifugation at 5000 × g for 10 min at 4 °C and washed twice with 50 mM Tris-HCl buffer (pH 7.5). The whole cell reaction system which contains 10 mM IBI, 7.25 g/DCW (dry cell wight) in 50 mM Tris-HCl buffer (pH 7.5) was taken for initial screening process.

3. Molecular docking

The X-ray crystal structure of *B. stearothermophilus* D-hydantoinase was retrieved

from the protein data bank (PDB) with the entry code 1k1d. Superimposition was then carried out between 1k1d and the crystal structure of hydantoins from *Bacillus* sp. AR9 (PDB code: 1yny). The conserved water molecule of the di-zinc center that existed in 1yny was extracted to the corresponding active site of *B. stearothermophilus* D-hydantoinase to generate the protein model using PyMol. Wild-type and mutant models were pretreated by running Lepro, the zinc and conserved water atoms were retained. A three-dimensional (3D) structure of IBI was created by Chem3D software and free energy minimization was generated by the MMFF94 program. Docking IBI to D-hydantoinase and mutant were performed using Ledock. The di-zinc center in the crystal structure was set as the center of the enclosing box (25 Å × 25 Å × 24 Å) using the GetBox plugin in PyMOL.

4. Construction and selection of site-point mutagenesis libraries

The saturation mutagenesis libraries were incorporated by an overlapping PCR method using complementary oligonucleotides with plasmid pTrac99a contain target gene as template. The DNA fragments were digested with *Dpn*I and transformed into *E. coli* TOP10 after purification. The recombinant strains were cultured and reacted as mentioned above.

5. Purification of candidate protein

The wild type and the chosen mutaions were subcloned into plasmid pET-28a and transformed into *E. coli* BL21(DE3). The cells were cultivated and harvested as described above and disrupted by sonication. The supernatant was obtained after centrifugation at 13,000 × g for 15 min and purified through Ni-chelating affinity chromatography. Proteins were identified by sodium dodecyl sulfate-polyacrylamide

gel electrophoresis (SDS-PAGE), and quantified by Bradford method[1].

6. Activity and kinetic assay

The standard enzyme-assay mixture contained 50 mM Tris/HCl buffer (pH 7.5), 5 mmol IBI, 1 mM Mn^{2+} and 1-10 μg enzyme in 100 μL reaction system. The reaction was stopped with 2 μL HCl after incubation. One unit enzyme activity is defined as the amount of enzyme that catalyzed the production of 1 μmol IBM in 1 min under the appropriate assay conditions. The specific activity was defined as the amount of IBM produced per min per mg protein. Kinetic constants were determined from Lineweaver-Burk plot assuming Michaelis-Menten kinetics.

7. Fermention and transformation

A single colony of the choosen mutatant was inoculated in LB medium and cultivated at 37°C and 200 rpm for about 6-8 h, and then transferred to fresh LB medium at 2 % (v/v) inoculum dose and cultivated as the seed culture. The 5 L fermenter containing 3 L of sterilized CD (Chemical Determination) medium was sterilized and transferred with seed culture at an initial concentration of about 0.15-0.2 (OD_{600})[2]. The fermentation process was carried out at 37°C until OD_{600} reached 30, then temperature shifted to 30°C after 0.2 mM IPTG was added for about 24 h. During the whole protein induction process, the pH was controlled at 7.5 using ammonia and DO value at 30% with glucose supplment rate at 5 g/L/h[3].

The cells were harvested and resuspend into the reaction buffer include 500 mM IBI, 20 g/L glucose, 5 mM Mn^{2+} and 18 g/DCW whole cell. The reaction process was carried at 50 °C and pH was controlled at 7.5 using NaOH. Samples were taken every 2 hours for HPLC detection.

8. Synthesis of *S*-PGB from experimentally prepared *R*-IBM in this study

R-IBM (3.0 g) was first dissolved in 5.0 M NaOH (3.5 mL). In a separate flask, bromine (3.06 g) was added to 5.0 M NaOH (11.0 mL) with dropwise in ice bath. Then the bromine solution was added to the *R*-IBM solution dropwise, and heated to 80 ºC for 2 h. Then hydrochloric acid solution was added to adjust pH to 5.0, cooled in ice bath, filtered and the solid was wash with water, dried to give 2.4 g *S*-PGB with a final yield of 93.8%.

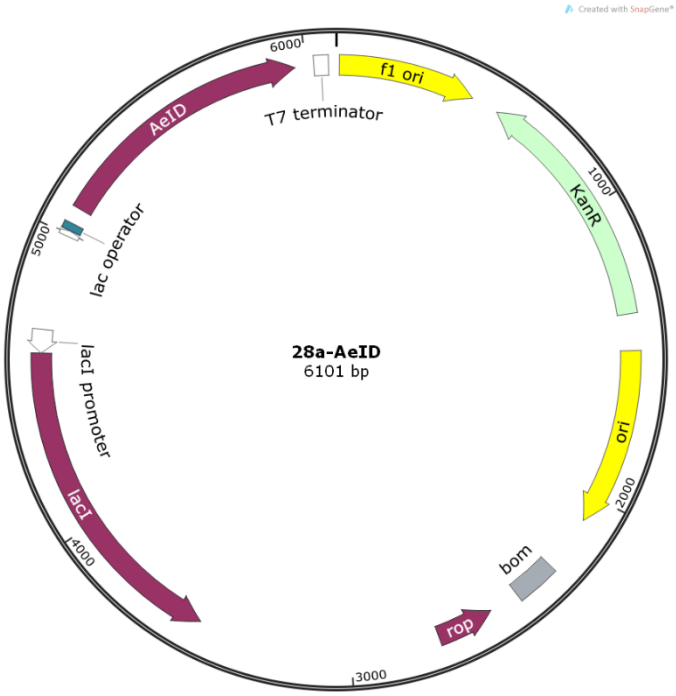
9. Analytic methods

For testing of IBI and IBM, the procedures are as follows: HPLC (Agilent 1260) equipped with a C18 column (Agilent XDB-C18, 4.6×250 mm, 5 μm) at flow rate of 1 mL/min, the column temperature was 30°C, the absorbance was set at 210 nm; solvent A was dipotassium phosphate (20 mM, pH 3.0) and solvent B was acetonitrile, the elution profile was that 0-20 min 20% B, 20-30 min 40% B, 30-35 min 40% B, 35-40 min 20% B, 40-45 min 20% B. The chirality of IBM was analysis with Chiralpak AD-H column (250 × 4.6 mm, 5 μm), at flow rate of 0.5 mL/min, the column temperature was 35 °C, the absorbance was set at 210 nm; TFA (0.1% (v/v)) in *n*-hexane and ethanol (supplied with 0.15% (v/v) TFA) were employed as mobile phases at ratio of 88:12. *S*-PGB derivatized with Marfey's reagent was detected by HPLC with the same C18 column at 340 nm. In brief, 400 μL of 1 mol/L 1-fluoro-2-4-dinitrophenyl-5-L-alanine amide (FDAA) acetone solution was mixed with 40 μL of 1 mol/L $NaHCO_3$ solution and 100 μL *S*-PGB sample. The mixture was reacted at 40 ℃ for 1 h, and then 10 μL of 2 mol/L hydrochloric acid solution was added. The derivatives were analyzed at a flow rate of 0.6 mL/min, with a mobile phase of 0.5% triethylamine solution (pH 3.0) and acetonitrile with a ratio of 55:45.

10. The investigated gene sequences and recombinant plasmid maps

*Ae*ID (GenBank No. AAK53437.1)

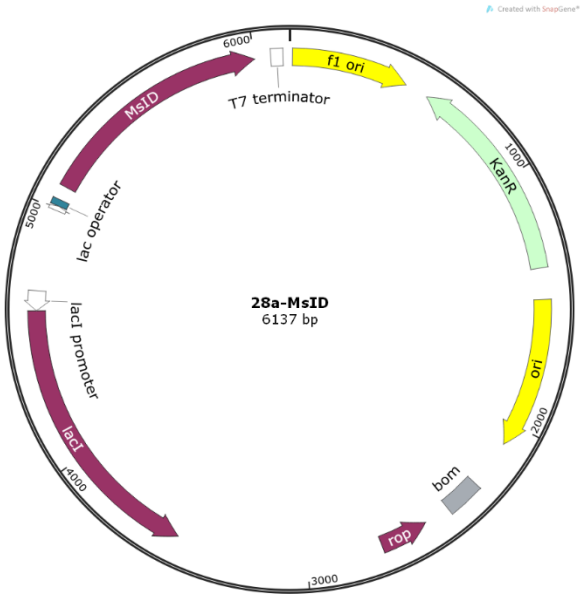
ATGAGCAAGGACATACAGATCACGTTTGGCGTGGACGTCGATGCCGTCGGC
GGCTGGCTCGGCTCGTATGGCGGGGAGGACTCGCCCGACGACATCTCGCG
CGGCATGTTCGCGGGCGAGGTCGGCTCGCTGCGGCTGCTCAGGCTGTTCG
ACAAATACGGCCTCAAGACCACCTGGTTCATCCCGGGCCACTCCGCCGAG
ACCTTTCCCGAGCAGATGAAGGCCGTGGTGGACGCGGGCCATGAGATCGG
CATCCACGGCTACAGCCACGAGAACCCGATTGCCATGACGCCCGAGCAGG
AGGAAGCGGTGCTCGACCGCAGCATCGAGGTCATCACGCGGCTGGCCGGC
CGCGTACCCACGGGCTATGTGGCGCCCTGGTGGGAGTTCAGCCCCGTCACC
AACGAGCTGCTGCTCAAGAAAGGCATCAAGTACGACCACAGCCTGATGCA
CAACGACTTCCACCCGTACTACGTGCGCGTGGGCGACCAGTGGACCCGCA
TCGACTACAACAAGCATCCCGACACCTGGATGAAGCCCCTGGTGCGCGGA
AAGGAAACGCGGCTGGTCGAGATCCCCGCCAACTGGTACCTGGACGACCT
GCCGCCCATGATGTTCATCAAGAAGTCGCCCAACAGCCACGGCTTCGTCAA
CCCGCGCGACATAGAGCAGATGTGGCGCGACCAGTTCGACTGGGTCTACC
GCGAGCACGACTACGCGGTCTTCCCCATCACCATCCACCCCGATGTCTCCG
GCCGCCCCCAGGTGCTGCTGATGCTGGAGCGCCTGATCGAGCATTTCCAGT
CGCACGACGGCGTGCGCTTCATGACGTGCGACCAGATCGCCGACGACTTC
CTGCGCCGCCAGCCGCGTTAA



*Ms*ID (GenBank No. WP_011727598.1)

ATGACCGACAAAAAAATCCAGGTTTGCGTTGGTATCGACGTTGACTCTTGC
GCTGGTTGGCTGGGTTCTTACGGTGGTCAGGACTCTCCGAACGACATGCAG
CGTGGTGTTTTCGCTGGTGAAGTTGGTGTTCCGCGTATGCTGAAACTGCTG

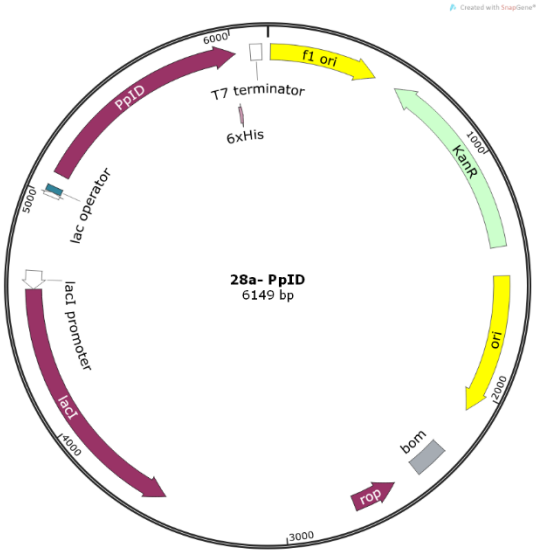
GAACGTCGTGGTATCGTTTCTACCTGGTTCTGGCCGGGTCACTCTATCGAAA
CCTTCCCGACCCAGGCTAAAATGTGCGTTGACGCTGGTCACGAAATCGGTG
CTCACGGTTACTCTCACGAAAACCCGCGTTCTCTGTCTCAGGAACAGGAAC
GTGCTGTTATGGCTAAATCTGTTGAACTGGTTGAACAGCTGACCGGTGCTC
GTCCGAAAGGTTACGTTGCTCCGTGGTGGGAAATGTCTGAACACACCGCTA
CCATCCTGAAAGAATACGACTTCTCTTACGACCACTCTCAGAACTACAACG
ACTTCGTTCCGTTCTACGCTCGTGTTGGTGACGAATGGACCCCGATCGACA
CCACCAAACCGGCTGAACACTGGATGAAACCGCTGAAACACGGTCACGAA
ATCGACCTGGTTGAATTCTGCGGTAACTGGTACGTTGACGACCTGCCGCCG
ATGATGTTCATCAAAGGTCACGCTAACTCTCACGGTTTCGTTAACCCGCGT
GACATCGAACAGATGTGGAACGACCAGTTCGACTGGGTTTACCGTGAACT
GGACTACGCTGTTATCCCGATGACCCTGCACCCGGACGTTTCTGGTCGTCC
GCAGGTTCTGCTGATGCTGGAACGTCTGCTGGACCGTTGGGCTGGTTTCGA
CGGTGTTGAATTCGTTACCATGGCTGACGCTGCTGAACAGTTCCGTTCTCG
TTACCCGCTGGACCAGCCGACCCGTCCGGAATGCGTTGGTCGTTAA



*Pp*ID (GenBank No. AAY98498.1).

ATGGCTAAAGAAATCCTGTGCTCTTTCGGTATCGACGTTGACGCTGTTGCTG
GTTGGCTGGGTTCTTACGGTGGTGAAGACTCTCCGGACGACATCTCTCGTG
GTCTGTTCGCTGGTGAAGTTGGTTCTCCGCGTCTGCTGAAACTGTTCGAAC
GTTTCGGTATCAAAACCACCTGGTTCATCCCGGGTCACTCTATCGAAACCTT
CCCGGAACAGATGCAGGCTGTTGCTGACGCTGGTCACGAAATCGGTATCCA
CGGTTACACCCACGAAAACCCGATCGCTATGACCCGTGAACAGGAAACCG
CTGTTCTGGACAAATGCATCGACCTGGTTACCAAACTGTCTGGTAAACGTC
CGACCGGTTACGTTGCTCCGTGGTGGGAATTCTCTAACGTTACCAACGAAC
TGCTGCTGGAACGTGGTATCAAATACGACCACTCTCTGATGCACAACGACT
TCACCCCGTACTACGTTCGTGTTGGTGACAAATGGACCAAAATCGACTACT
CTAAAAAACCGTCTGACTGGATGGTTCCGCTGACCCGTGGTAAAGAAACC
GACCTGATCGAAATCCCGGCTTCTTGGTACCTGGACGACCTGCCGCCGATG
ATGTTCATCAAAAAATCTCCGAACTCTCACGGTTTCGTTAACCCGCACGAC

ATCGAACAGATCTGGCGTGACCAGTTCGACTGGGTTTACCGTGAAATGGAC
TACGCTGTTTTCCCGATCACCATCCACCCGGACGTTGCTGGTCGTCCGCAG
GTTCTGATGATGCTGGAACGTCTGTACGCTCACATGATCAAACACCCGGGT
GTTAAATTCGTTACCATGAACGAAATCGCTGACGACTTCGCTAAACGTTTCC
CGCGTAAAAAATAA



*Bs*Hase (GenBank No. QFU78355.1)

ATGACCAAAATTATTAAAAATGGTACCATCGTTACCGCCACCGATACCTATG
AAGCCCATCTGCTGATTAAAGATGGTAAAATTGCAATGATTGGCCAGAATCT
GGAAGAAAAAGGTGCCGAAGTTATTGATGCGAAAGGCTGTTATGTGTTTCC
GGGCGGTATTGATCCGCATACCCATCTGGATATGCCTCTGGGTGGTACCGTG
ACCAAAGATGATTTTGAATCAGGTACCATTGCCGCCGCGTTTGGCGGTACC
ACCACCATTATTGATTTTTGCCTGACCAATAAAGGTGAACCGCTGAAAAAA
GCAATTGAAACCTGGCATAATAAAGCCAATGGCAAAGCGGTGATCGATTAT
GGTTTTCATCTGATGATTAGTGAAATTACTGATGATGTTTTAGAAGAACTGC
CGAAAGTTCTTGAAGAAGAAGGTATTACCTCACTGAAAGTGTTTATGGCGT
ATAAAAATGTTtTTCAGGCTGATGATGGTACCCTGTATTGCACCCTGCTGGCC
GCAAAAGAACTGGGTGCCCTGGTGATGGTGCATGCCGAAAACGGTGATGT
TATTGATTATCTGACGAAAAAAGCACTGGCAGATGGTAATACCGATCCGATC
TATCATGCGCTGACCCGTCCGCCGGAACTGGAAGGTGAAGCTACAGGTCG
CGCATGCCAGCTGACAGAACTGGCAGGCTCACAGCTGTATGTGGTGCATGT
GACATGTGCCCAGGCCGTGGAAAAAATTGCGGAAGCGCGCAATAAAGGTC
TGGATGTTTGGGGCGAAACCTGCCCGCAGTATTTAGTTCTGGATCAGAGCT
ATCTGGAAAAACCAAATTTTGAAGGTGCAAAATATGTGTGGAGTCCGCCGC
TGCGCGAAAAATGGCATCAGGAAGTTCTGTGGAATGCGCTGAAAAATGGT
CAGCTGCAGACCCTGGGTAGTGATCAGTGTAGTTTTGATTTTAAAGGTCAG
AAAGAACTGGGACGTGGTGATTTTACCAAAATTCCGAATGGTGGTCCGATT
ATTGAAGATCGTGTGAGTATTCTGTTTAGCGAAGGTGTTAAAAAAGGTCGC
ATTACCCTGAACCAGTTTGTTGATATTGTGAGCACCCGCATTGCAAAACTGT
TTGGCCTGTTTCCGAAAAAAGGCACCATTGTTGTGGGTAGCGATGCGGATC
TGGTGATTTTTGATCCGAATATTGAACGTGTGATTAGTGCCGAAACACATCA
TATGGCCGTTGATTATAATGCGTTTGAAGGCATGAAAGTGACCGGTGAACC

GGTGAGCGTTCTGTGTCGTGGTGAATTTGTGGTGCGTGATAAACAGTTTGT
TGGCAAACCGGGTTATGGTCAGTATCTGAAACGCGCGAAATATGGCACCAG
CACCATTAGTAAACAGAGCGAAGAACTGACCATTTAA

References

1. Kruger, N. J., The Bradford method for protein quantitation. *Methods Mol Biol. (Clifton, N.J.)* **1994,** *32*, 9-15.
2. Mu, Q. X., Zhang, S. S., Mao, X. J., Tao, Y., Yu, B., Highly efficient production of L-homoserine in *Escherichia coli* by engineering a redox balance route. *Metab Eng*. **2021,** *67*, 321-329.
3. Tao, Y.; Cheng, Q.; Kopatsis, A. D., Metabolic engineering for acetate control in large scale fermentation. *Methods Mol Biol. (Clifton, N.J.)* **2012,** *834*, 283-303.