Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry.


# Supplemental Information

## A ThDP-dependent enzymatic carboligation reaction involved in Neocarazostatin A tricyclic carbazole formation

Li Su[1,#], Meinan Lv[1,#], Kwaku Kyeremeh[3], Zixin Deng[1], Hai Deng[2], Yi Yu[1,*]

[1]Key Laboratory of Combinatorial Biosynthesis and Drug Discovery (Ministry of Education), School of Pharmaceutical Sciences, Wuhan University, 185 East Lake Road, Wuhan 430071, P. R. China. Email: yu_yi@whu.edu.cn

[2]Marine Biodiscovery Centre, Department of Chemistry, University of Aberdeen, Aberdeen AB24 3UE, Scotland, U.K.

[3]Department of Chemistry, University of Ghana, P.O. Box LG56, Legon-Accra, Ghana

### Experimental Procedures

### Media and strains for cultivation and fermentation

*E. coli* DH10B*, E. coli* ET12567 (pUZ8002), and *E. coli* BL21(DE3) containing the expression plasmid were cultured in Luria-Bertani or Luria agar medium at 37 °C. ISP2 medium (yeast extract 4 g/L, glucose 4 g/L, malt extract 10 g/L, pH 7.2) was used for seed culture and fermentation of *Streptomyces* sp. MA37 and related strains. Modified ISP4 medium (10 g/L soluble starch, 2 g/L $(NH_4)_2SO_4$, 1 g/L$K_2HPO_4$, 1 g/L $MgSO_4$•$7H_2O$, 1 g/L NaCl, 1 g/L tryptone, 0.5 g/L yeast extract, 1 g/L peptone, 1 ml/L trace element solution, adjusted pH to 7.2 before sterilization) containing the final concentration of 30 mM $Mg^{2+}$ was used for conjugation.

*Streptomyces* sp. MA37 and its related mutant strains were inoculated into ISP2 medium and grown for 36 h at 28 °C, 220 *rpm*, and then transferred into ISP2 fermentation medium for continuous cultivation at 28 °C, 220 *rpm* for 5 days. The culture was then centrifuged and the supernatants were collected and further extracted by an equal volume of ethyl acetate. The organic phase was collected and evaporated

to dryness. Residues were re-dissolved in 1/100 of original culture volume of methanol for subsequent HPLC or LC-MS detection.

**Complementation of *nzsH* gene knockout strain**

The complementation plasmid was prepared by cloning the whole *nzsH* gene sequence, using the following forward and reverse primer pairs: 5′-TGACATGATTACGAATTCAGGCCGGGGTCATCGTTCCGGCG-3′ and 5′-GGTAGGATCCACATATGCGGATAGCCGACGCTTTAGTCA-3′, into the NdeI/EcoRI site of pIB139 to result in pWDY631 for complementation of *nzsH* gene knockout strain.[1] This pIB139 derived plasmid, which can be integrated into the *attB* site in *Streptomyces* chromosome via phage ΦC31 integrase-catalyzed site-specific recombination,[2,3] was transferred into *nzsH* gene knockout mutant WDY630 via *E. coli-Streptomyces* conjugation to give WDY631.

**RNA isolation and RT-PCR assay**

The expressions of *nzs* genes in *Streptomyces* sp.MA37 and WDY631 were measured at the transcriptional level. Both strains were harvested after 72 h of growth in ISP2 medium, and total RNA was prepared by TRIzol (Invitrogen). The cDNAs were synthesized from the total RNA using cDNA synthesis kit (Thermo scientific) and used for further amplification of the corresponding transcripts. The RT-PCR reaction was performed using 2xTaq mix (Novoprotein). Gene-specific primers are listed in Table S3. The 16S rRNA gene was used as an internal control to quantify the relative expression of target genes.

**Cloning and construction of $His_6$-tagged NzsH overexpression plasmid**

The *nzsH* gene was amplified from MA37 genomic DNA using a primer pair 5′-GTGCCGCGCGGCAGCCATATGCGGATAGCCGACGCTTTAGTCACGACC-3′ and 5′-TGGTGGTGGTGCTCGAGCTACTAGGCCGGGGTCATCGTTCCGGCGCTCCG-3′.

The product was digested with NdeI and HindIII, purified using a Omega PCR purification kit, and then inserted into the same restriction sites of the expression vector pET28a (Novagen). Chemically competent *E. coli* DH10B cells were transformed with

the ligation mixture and plated on LB agar containing kanamycin (50 μg $mL^{-1}$) to screen for positive clones, which were further confirmed by DNA sequencing.

**Overexpression and purification of NzsH in *E. coli***

*E. coli* BL21 (DE3) cells containing *nzsH* overexpression plasmid were grown in 500 mL SOB medium supplemented with 50 mg/ml kanamycin at 37 °C with shaking at 220 rpm until an $OD_{600}$~0.6-0.8 was reached. Protein expression was induced by addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) with further cultivation at 28 °C for 4 h. The cell pellet collected by centrifugation was re-suspended in 40 ml ice-cold lysis buffer (50 mM Tris-HCl, 500 mM NaCl, 10 mM imidazole, and 10%(V/V) glycerol, pH 8.0), and disrupted by a Nano Homogenize Machine (ATS Engineering Inc, AH100B). Cell debris was removed via centrifugation at 12000 rpm for 30 min at 4 °C. The supernatant was loaded onto 2 ml Ni-NTA-affinity column (GE Healthcare) pre-equilibrated with the lysis buffer. The desired elution fractions were combined and concentrated using an Centrifugal Filter Units (Milipore,10,000 MWCO), and the concentrated protein solution was desalted using a PD-10 Column (GE Healthcare) pre-equilibrated with the elution buffer (20 mM Tris-HCl, 100 mM NaCl, and 10% (v/v) glycerol, pH 8.0). The protein fraction was collected and concentrated, analyzed by SDS-PAGE (12% Tris-glycine gel), and was used directly for in vitro assay or stored at −80 °C upon further use. Protein concentration was determined using a Bradford assay kit (Promega) with bovine serum albumin (BSA) as the standard.

***In vitro* assay of NzsH**

NzsH activity assays were carried out on a 100 μL scale with 1 μM purified enzyme, 2 mM thiamin diphosphate, 5 mM $MgCl_2$, 1 mM indole-3-pyruvate, and 1mM pyruvate or 2-oxobutyrate in 100 mM Tris-HCl (pH 8.0). After incubated at 30 °C for 30 min, the reaction was quenched by the addition of equal volume methanol followed by vortex and centrifugation to remove protein. The supernatants were analyzed by HPLC and HR-ESI-MS. The detection condition was described as below.

**Kinetic analysis of NzsH**

The kinetic study of NzsH was performed in 50 µL reaction mixtures. For assays with pyruvate or 2-oxobutyrate, mixtures contained 1 µM NzsH, 2 mM thiamin diphosphate, 5 mM $MgCl_2$, 2 mM indol-3-pyruvate, and pyruvate (varying from 50 µM to 4.8 mM) or 2-oxobutyrate (varying from 0.4 mM to 16 mM) in 100 mM Tris-HCl (pH 8.0). The reactions were initiated by addition of NzsH, incubated at 30 °C for 15 min and terminated by the addition of 50 µL methanol. The initial velocity was determined by HPLC analysis of the product **2** or **8**, with purified compound **6** serving as the internal standard. Both the substrate quantification and the reaction were performed in triplicates. The resulting initial velocities were then fitted to the Michaelis-Menten equation by nonlinear regression analysis using GraphPad Prism 6.0 to extract $K_m$ and $k_{cat}$ parameters.

**Structural characterization of compound 6, 7, 9, and 10**

A 30 mL aqueous solution containing 100 mM Tris-HCl pH 8.0, 2 mM thiamin diphosphate, 1 mM indole-3-pyruvate, 1 mM pyruvate or 3 mM 2-oxobutyrate, and 1 µM NzsH was prepared, mixed with a pipet and divided into 300 µL per tube, incubated at 30 °C for 30 min and finally quenched with equal volume of methanol. The reactions were centrifuged (13,000 rpm x 10 min) and detected with HPLC described as followed. Because of the instability of compound **2** and **8**, over dose of $NaBH_4$ was treated to the NzsH catalyzed reaction. The final products (compound **6**, **7**, **9**, and **10)** were isolated by using HPLC as described below. The purified compounds were evaporated to dryness, resolved in $CD_3OD$ and characterized using NMR analysis.

The NMR data of the four compounds were recorded on a Bruker Avance 600 MHz NMR spectrometry.

Compound **6**: $^1$H-NMR, δH (600 MHz, $CD_3OD$), 7.56 (1H,d,7.8,H-7), 7.31 (1H,d,8.03,H-4), 7.09 (1H,s,H-2), 7.06 (1H,t,7.46,H-5), 6.97 (1H,t,7.32,H-6), 4.03 (1H,d,8.80,H-10), 2.99-2.82 (2H,m,H-8a,H-8b), 1.49 (3H,s, $CH_3$).

$^{13}$C-NMR, δC (500 MHz, $CD_3OD$), 178.28 (C, COOH), 136.76 (C, C-7a), 127.63 (C, C-3a), 122.81 (CH, C-2), 118.09 (CH, C-7), 117.96 (C, C-6), 111.85 (C,C-3), 110.67 (C, C-4), 77.40 (C, C-9), 75.37 (C, C-10), 27.60 ($CH_2$, C-8), 21.69 ($CH_3$, C-11).

Compound **7**: $^{1}$H-NMR, δH (600 MHz, DMSO-*d6*), 7.56 (1H,d,7.85,H-7), 7.32 (1H,d,8.08,H-4), 7.11 (1H,s,H-2), 7.07 (1H,t,7.31,H-5), 6.99 (1H,t,7.41,H-6), 4.03 (1H,d,8.94,H-10), 3.17 (1H,d,14.60,H-8a), 2.81-2.73 (1H,m,H-8b), 1.51 (3H,s, $CH_3$).

$^{13}$C-NMR, δC (500 MHz, $CD_3OD$), 179.09 (C, COOH), 138.64 (C, C-7a), 129.55 (C, C-3a), 124.75 (CH, C-2), 119.94 (CH, C-7), 119.87 (C, C-6), 113.53 (C,C-3), 112.63 (C, C-4), 79.37 (C, C-9), 77.33 (C, C-10), 28.32 ($CH_2$, C-8), 23.34 ($CH_3$, C-11).

Compound **9**: $^{1}$H-NMR, δH (600 MHz, $CD_3OD$), 7.57 (1H,d,7.90,H-7), 7.29 (1H,dd,12.00,4,46,H-4), 7.09 (1H,s,H-2), 7.05 (1H,t,7.50,H-5), 6.97 (1H,t,7.36,H-6), 4.03 (1H,t,4.04,H-10), 2.92-2.86 (2H,m,H-8a,H-8b), 1.95-1.83 (2H,m,$CH_2$), 0.94 (3H,t,7.33 $CH_3$).

Compound **10**: $^{1}$H-NMR, δH (600 MHz, DMSO-*d6*), 7.56 (1H,d,7.76,H-7), 7.32 (1H,d,8.08,H-4), 7.10 (1H,s,H-2), 7.06 (1H,t,7.31,H-5), 6.98 (1H,t,7.40,H-6), 4.01 (1H,s,H-10), 3.17 (1H,d,12.45,H-8a), 2.75 (1H,s,H-8b),2.04-1.69 (2H,m,$CH_2$) 0.96 (3H,s, $CH_3$).

**HPLC and LC-MS parameters**

A DIONEX System equipped with a P680 HPLC Pump and UVD 170U detector was used for HPLC analysis, which was performed on a DIKMA Diamonsil C18 column (5 μm, 250×4.6 mm).

HPLC conditions for analysis of the fermentation products of MA37 and its related mutants: C18 column was pre-equilibrated with 90%A/10%B and developed at a flow rate of 1 ml/min, 0-10 min, a linear gradient from 90%A/10%B to 60%A/40%B; 10-35 min, a linear gradient to 45%A/55%B; 35-40 min, a linear gradient to 0%A/100%B; 40-42 min, constant with 0%A/100%B; 42-44 min, a linear gradient to 90%A/10%B; UV monitored at 247 nm and 280 nm. Conditions for analysis of compound **6**, **7**, **9**, and **10**: C18 column was pre-equilibrated with 90%A/10%B and developed at a flow rate of 1 ml/min, 0-25 min a linear gradient from 95%A/5%B to 5%A/95%B; 25-26 min, a linear gradient to 0%A/100%B; 26-29 min, constant with 0%A/100%B;29-30min, a linear gradient to 95%A/5%B; UV monitored at 254 nm and 280 nm (both solvent A was 0.1% formic acid in $H_2O$ and solvent B was 0.1% formic acid in $CH_3CN$).

LC-MS analysis was carried out in positive mode by using a Thermo Scientific LTQ XL Orbitrap mass spectrometer equipped with a Thermo Scientific Accela 600 pump (Thermo Fisher Scientific Inc.). Each LC conditions were described as above. All MS analysis parameters were set as 45 V capillary voltage, 45 °C capillary temperature, auxiliary gas flow rate 10 arbitrary units, sheath gas flow rate 40 arbitrary units, 3.5 kV spray voltage, and 100-2000 Amu mass range (maximum resolution 30,000).

**Alignment and phylogenetic analysis**

Homologous sequences of NzsH were retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov/) with their accession numbers and the source organisms listed in **Table S2**. These sequences were aligned using MUSLE program embedded in MEGA6,[4] and the alignment was manually fine-tuned afterward. Bayesian Markov Chain Monte Carlo (MCMC) method was used to calculate posterior probability of various clades through the program MrBayes (version 3.2).[5] A mixed model of amino acid substitutions was used, and final analyses consisted of four chains (one cold and three heated) were run for about 2,800,000 generations by sampling parameters and trees every 1,000 generations. Analyses were run to reach a convergence with standard deviation of split frequencies <0.006873. Posterior probabilities were averaged over the final 75% of trees (25% burn in). The figure of the Bayesian phylogram was prepared by using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

**Figure S1.** ALS catalyzed reactions.[6,7] A) AHAS catalyzed carboligation of two molecules of pyruvate; B) YerE catalyzed carboligation of pyruvate and 2-oxobutyrate; C) ScyA catalyzed carboligation of indole-3-pyruvate and *p*-hydroxyphenylpyruate. All reactions produced (*S*)-enantiomer products.

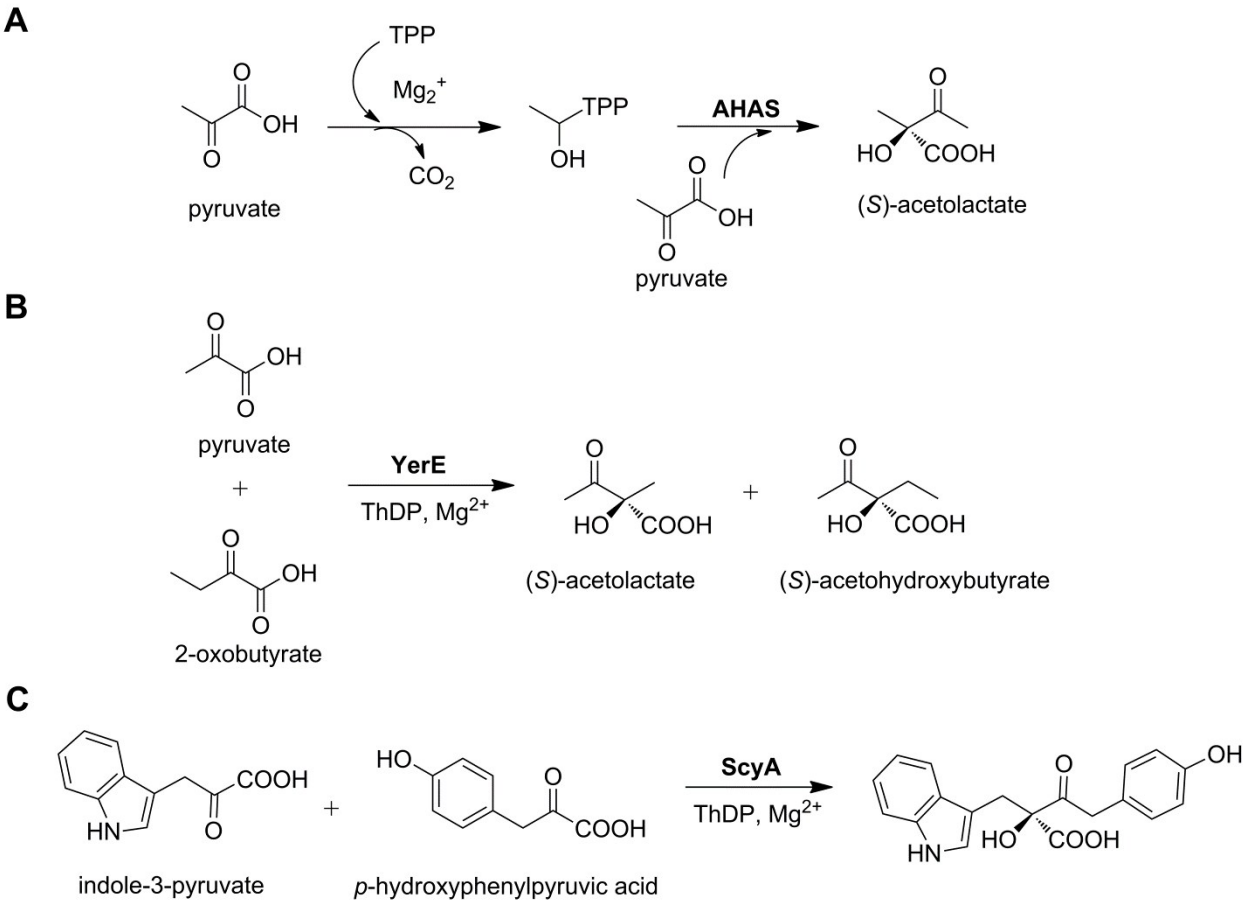



**Figure S2**. SDS-PAGE analysis of purified NzsH

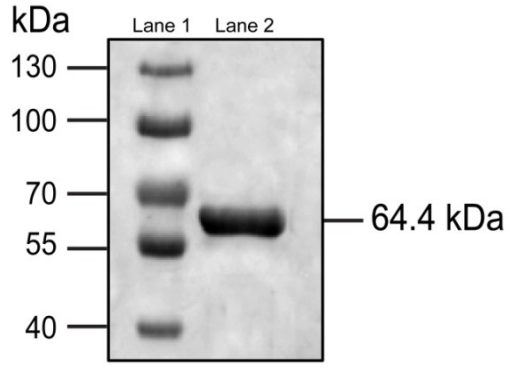
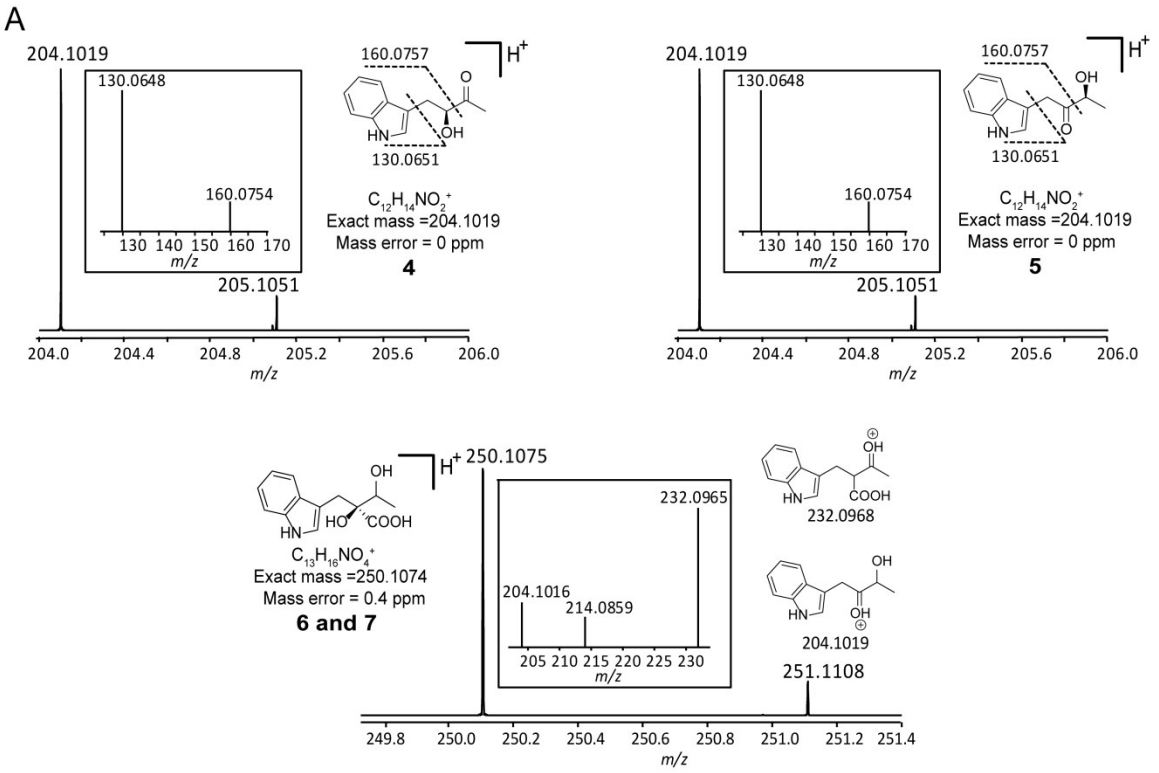


**Figure S3.** The HR-MS/MS spectrum of compounds **4**-**7** (A) and **9**-**10** (B), and their ion fragments upon collision induced dissociation (CID)



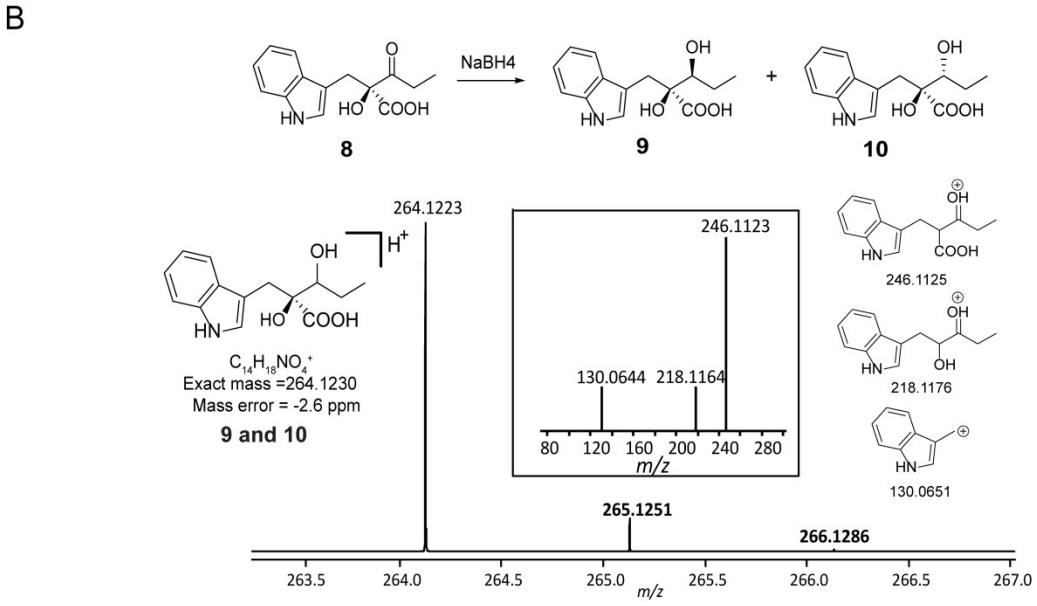



**Figure S4.** The $^{1}H$ NMR spectrum of compound **6** and **7** from NzsH catalyzed reaction with pyruvate as the donor substrate

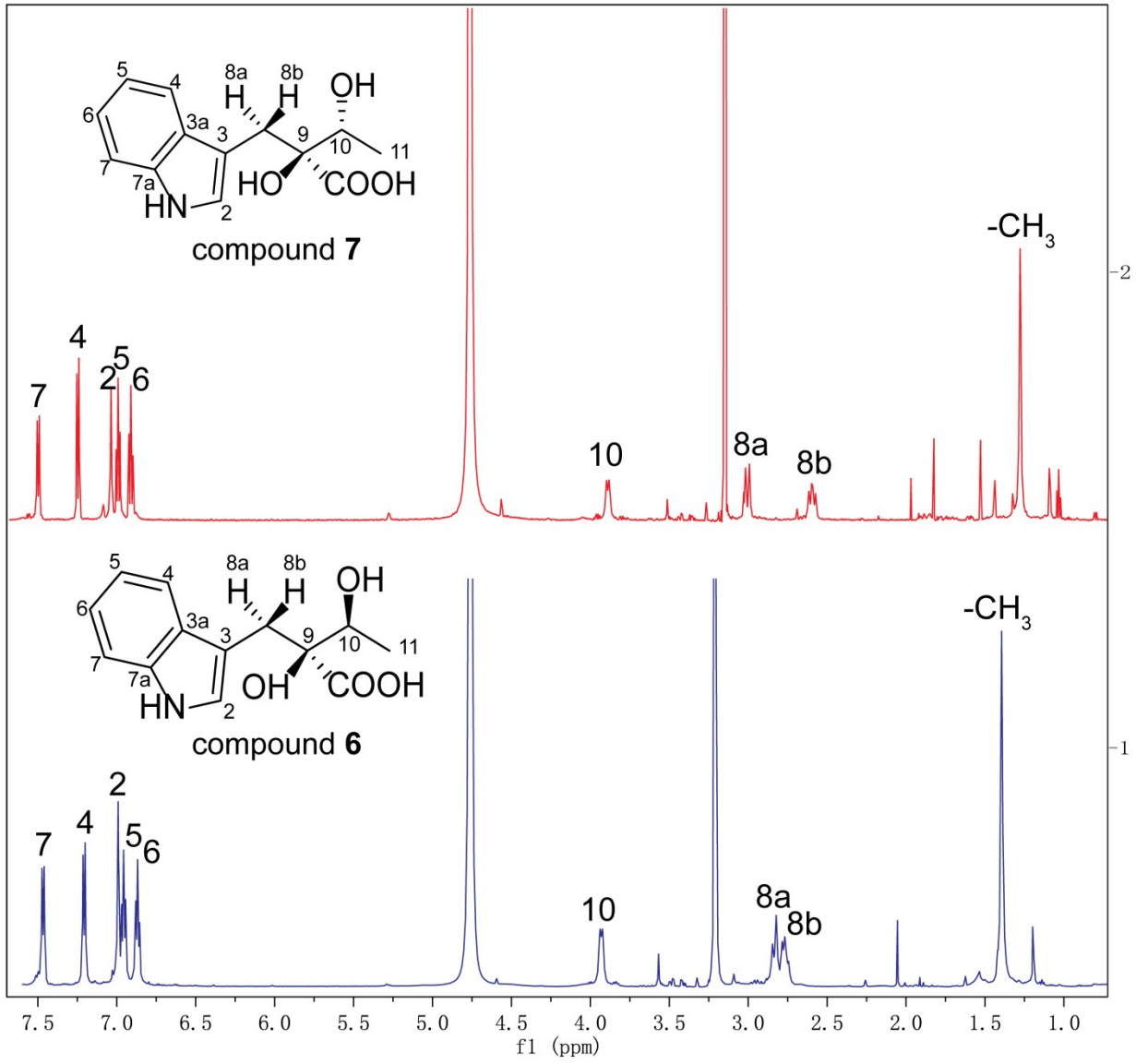


**Figure S5.** The $^{13}C$ NMR spectrum of compound **6**

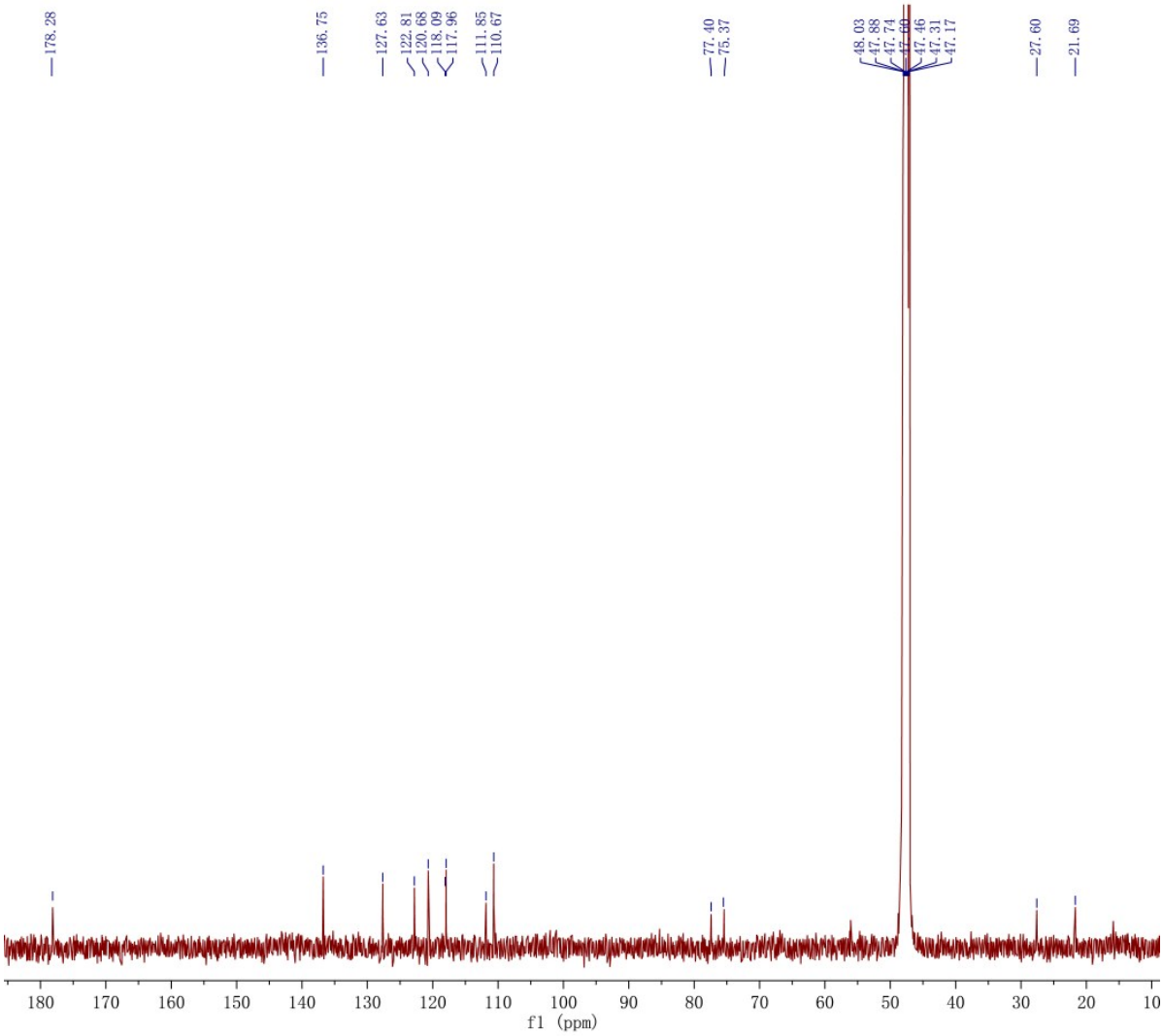


**Figure S6.** The $^{1}H$-$^{1}H$ COSY spectrum of compound **6**

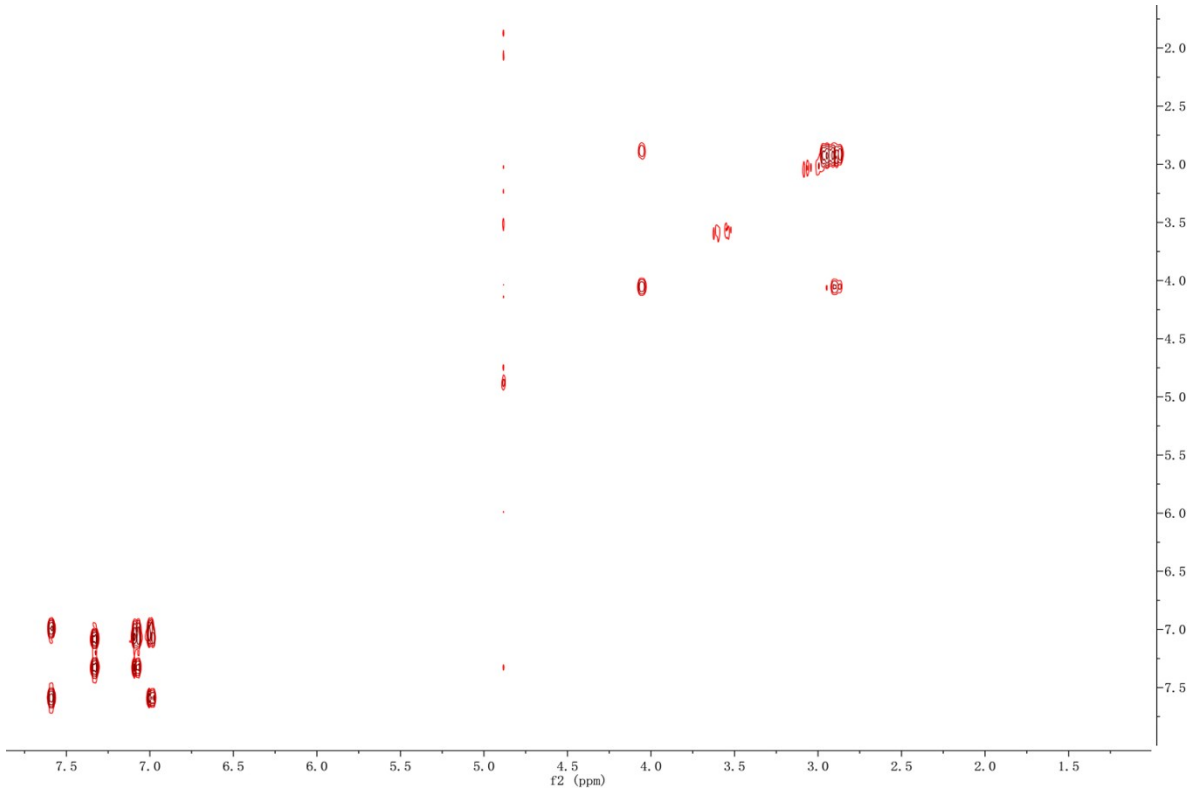



**Figure S7.** The HSQC spectrum of compound **6**

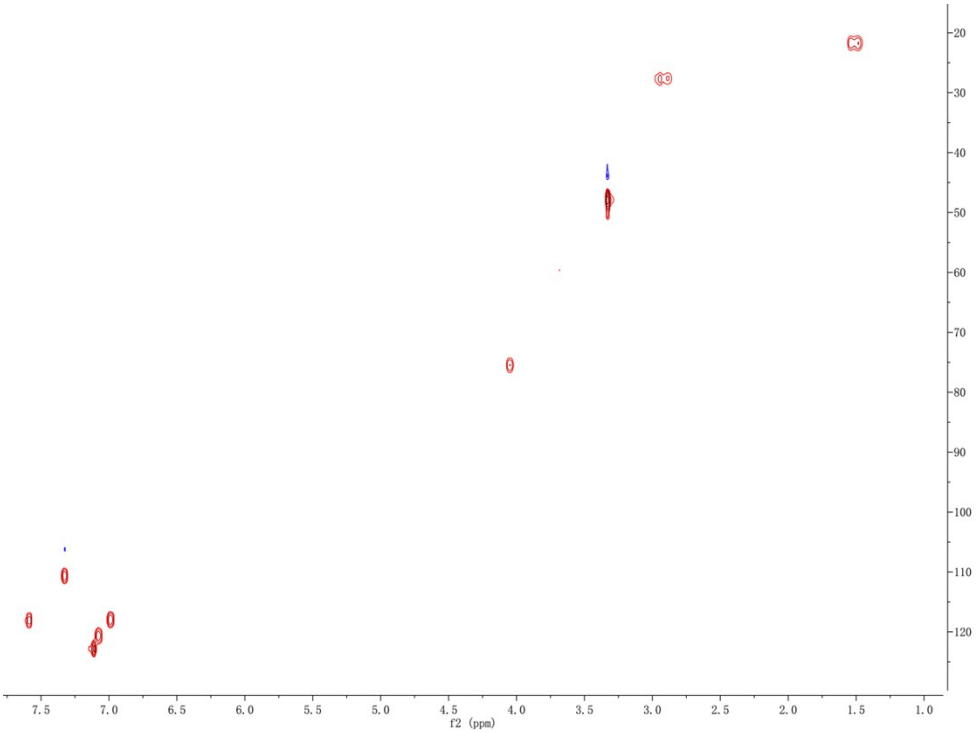


**Figure S8.** The HMBC spectrum of compound **6**

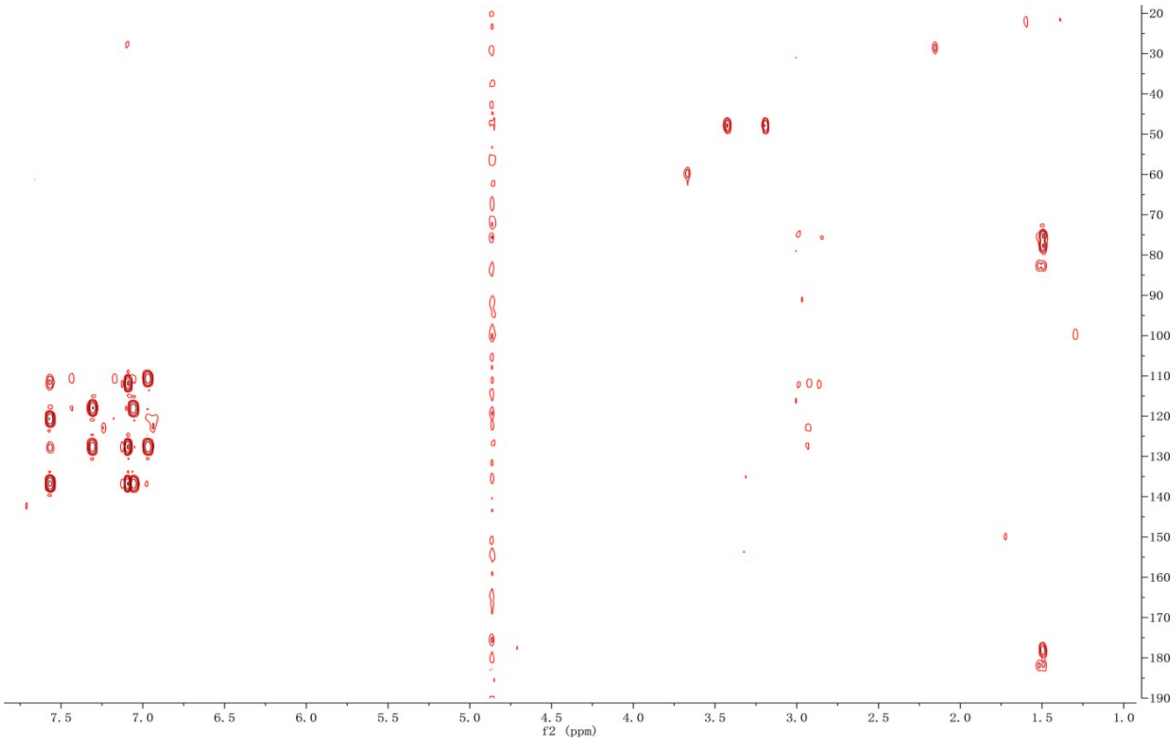


**Figure S9.** The $^{13}C$ NMR spectrum of compound **7**

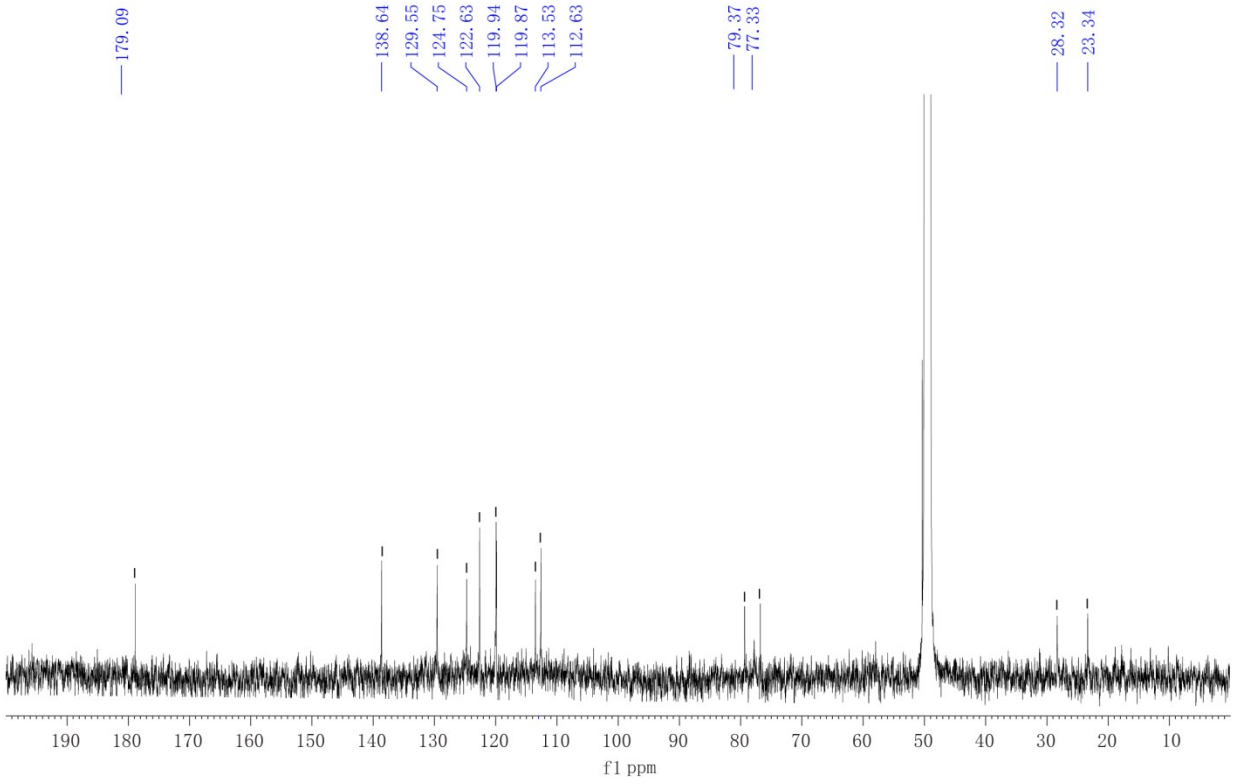


**Figure S10.** The[1]H NMR spectrum of compound **9** and **10** from NzsH catalyzed reaction with 2-oxobutyrate as the donor substrate

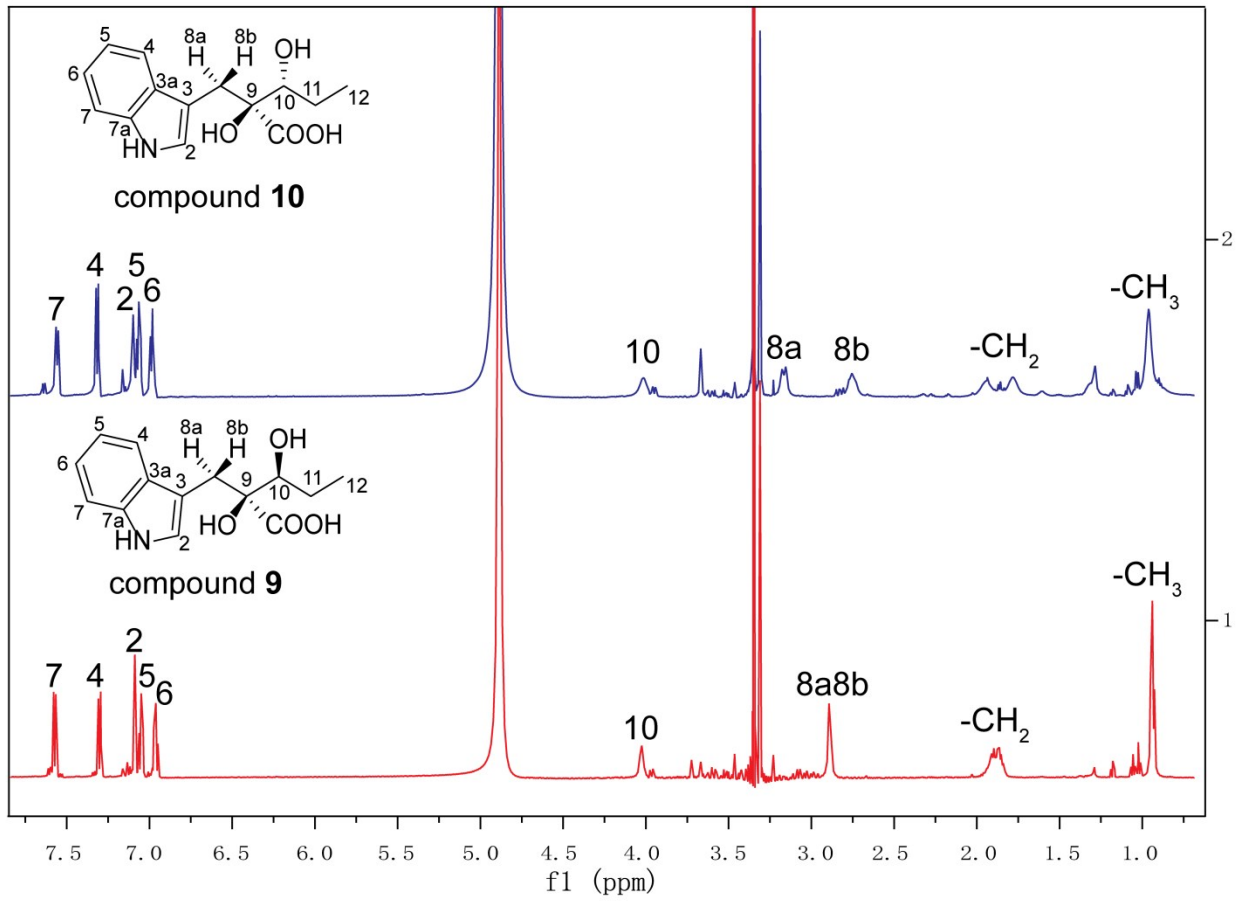


**Table S1.** Derivatives of indole-3-pyruvate and pyruvate used in this study to test the substrate tolerance of NzsH

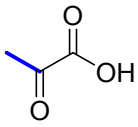
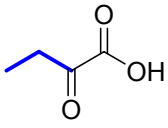
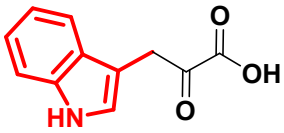
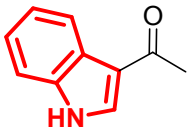
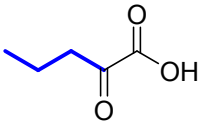
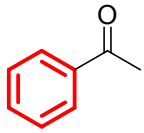
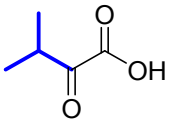
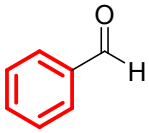
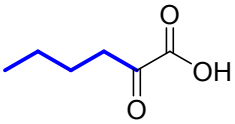
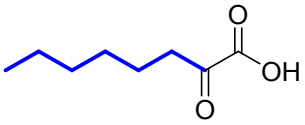
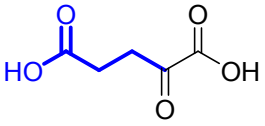
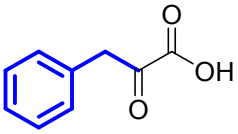
| donor | substitutes | acceptor | substitutes |
|---|---|---|---|
|  **pyruvate** |  **2-oxobutyrate** |  **indole-3-pyruvate** |  **3 - acetyl indole** |
| |  **2-oxovalerate** | |  **acetophenone** |
| |  **3-methyl-2-oxobutyrate** | |  **benzaldehyde** |
| |  **2-ketohexanoate** | | |
| |  **2-oxooctanoate** | | |
| |  **2-ketoglutarate** | | |
| |  **phenylpyruvate** | | |

**Table S2.** ThDP-dependent enzymes used for construction of the phylogenetic tree

| Accession number | Source | Enzymes |
|---|---|---|
| ALL53321 | *Streptomyces* sp. MA37 | NzsH |
| AEP25490.1 | *Yersinia pseudotuberculosis* | ALS (YerE) |
| AAC76694.1 | *Escherichia coli* K-12 | ALS (IlvB) |
| P00892 | *Escherichia coli* K-12 | ALS (IlvG) |
| ACC79997.1 | *Nostocpunctiforme* PCC 73102 | ALS (ScyA) |
| 2PGO-B | *Azoarcus* sp. strain 22lin | ALS (CDH) |
| P27696 | *Klebsiellapneumoniae* | ALS (BudB) |
| Q9LCV9 | *Streptomyces clavuligerus* | ALS (CeaS) |
| AMC91436.1 | *Mycobacterium tuberculosis* | ALS |
| 5AHK_A | *Pseudomonas Protegens* | ALS |
| EOT84436.1 | *Enterococcus faecalis* V583 | ALS |
| BAB38035.1 | *Escherichia coli* O157:H7 str. Sakai | ALS |
| AAG58963.1 | *Escherichia coli* O157:H7 str. EDL933 | ALS |
| BAB33504.1 | *Escherichia coli* O157:H7 str. Sakai | ALS |
| AAL22751.1 | *Salmonella Typhimurium* | ALS |
| AAZ90175.1 | *Shigellasonnei* Ss046 | ALS |
| AAV52901.1 | *Streptomyces cinnamonensis* | ALS |
| Q59498.1 | *Mycobacterium avium* | ALS |
| AAL99356.1 | *Bacillus stearothermophilus* | ALS |
| ZP_06876763.1 | *Pseudomonas aeruginosa* PAb1 | ALS |
| AAB53488.1 | *Methanococcusaeolicus* | ALS |
| AAX88283.1 | *Haemophilusinfluenzae*86-028NP | ALS |
| AAA22222.1 | *Bacillus subtilis* | ALS |
| ADH43113.1 | *Serratiamarcescens* | ALS |
| YP_004593689.1 | *Enterobacteraerogenes* KCTC 2190 | ALS |
| AAD02041.1 | *Leuconostoclactis* | ALS |
| P06672 | *Zymomonasmobilis* | PDC (ZmoPDC) |
| P06169 | *Saccharomyces cerevisiae* | PDC (ScePDC) |
| AAS49166 | *Lactococcuslactis* | PDC (KdcA) |
| BAA14242 | *Enterobacter cloacae* | InPDC |
| P51853 | *Pseudomonas fluorescens* | BAL (BznB) |
| P17109 | *Escherichia coli* K-12 | MenD |
| AAC15502 | *Pseudomonas putida* | BFD (MdlC) |

Abbreviations:ALS, acetolactate synthase; PDC, pyruvate decarboxylase; InPDC, indole-3-pyruvatedecarboxylase;BAL, benzaldehydelyase;BFD, benzoylformate decarboxylase.

**Table S3.** List of primers used for RT-PCR analyses

| Primers | Forward sequences | Reverse sequences |
|---|---|---|
| *NzsA* | 5′-AATGCTGAGTTCCCATTCGA -3′ | 5′- GGAATTCCGAATACTTCGCC-3′ |
| *NzsB* | 5′- CATCACCTTCTGCTTCGCCG-3′ | 5′- GGTTCGTCCGCCCTGGTCTT-3′ |
| *nzsC* | 5′- CTTCTCGGTCTTTCTCTTCG-3′ | 5′- GTGACCTACAACCACCCGGA-3′ |
| *NzsD* | 5′-ACTCGCCGAGCGCATCAACG-3′ | 5′- GCCACACCAACTTCTTCTGG-3′ |
| *NzsE* | 5′-AACAACCGGCTCGAGGCCCT-3′ | 5′- GGAACTCCTCACGTCCAAGG-3′ |
| *NzsF* | 5′-ACACCTTCTCGACGATCCTC-3′ | 5′-GACACGTACTTCCGGATGGA-3′ |
| *NzsG* | 5′-CATTCAGCACACAGGAAGAG-3′ | 5′-CACCTTCTTCTCGTTCGTCT-3′ |
| *NzsH* | 5′-AGACCTTCTTCGACTCCCGC-3′ | 5′-GTGCCGCCGTTCGCCAACTT-3′ |
| *NzsI* | 5′-CGGTCTCATGCAGAACTACC-3′ | 5′-CCGTACGACATCGTCTTCGA-3′ |
| *NzsJ* | 5′-GTTCAACGACGCCTTCTCCC-3′ | 5′-ACCAACTTCATCTCCCGACT-3′ |
| 16S rRNA | 5′-GCAGGGAAGAAGCGAGAGTG ACGG-3′ | 5′-ATCTAATCCTGTTCGCTCCCCA CG-3′ |

## References


1. S. Huang, Somayah S. Elsayed, M. Lv, J. Tabudravu, Mostafa E. Rateb, R. Gyampoh, K. Kyeremeh, R. Ebel, M. Jaspars, Z. Deng, Y. Yu and H. Deng, *Chem Biol*, 2015, 22, 1633-1642.
2. H. M. Thorpe and M. C. Smith, *Proc Natl Acad Sci U S A*, 1998, 95, 5505-5510.
3. M. J. B. Tobias Kieser, Mark J. Buttner, Keith F. Chater, David A. Hopwood, 2000.
4. K. Tamura, G. Stecher, D. Peterson, A. Filipski and S. Kumar, *Mol Biol Evol*, 2013, 30, 2725-2729.
5. F. Ronquist, M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Hohna, B. Larget, L. Liu, M. A. Suchard and J. P. Huelsenbeck, *Syst Biol*, 2012, 61, 539-542.
6. E. P. Balskus and C. T. Walsh, *J Am Chem Soc*, 2009, 131, 14648-14649.
7. M. Beigi, S. Loschonsky, P. Lehwald, V. Brecht, S. L. Andrade, F. J. Leeper, W. Hummel and M. Müller, *Org Biomol Chem*, 2013, 11, 252-256.