**Electronic Supplementary Information (ESI)**

# Efficient biosynthesis of enantiopure tolvaptan by utilizing alcohol dehydrogenase-catalyzed enantioselective reduction

Hong-Lei Zhang, Chao Zhang, Meng-Nan Han, Chao-Hong Pei, Zhi-Dong Xu and Wei Li*

Laboratory of Medicinal Chemistry and Molecular Diagnosis of the Ministry of Education, College of Chemistry and Environmental Science, Hebei University, Baoding 071002, China.

## 1. Experimental section

### 1.1 Biological and chemical materials

*Yarrowia lipolytica* CICC 32187, *Saccharomyces cerevisiae* CICC1002, *Pichia stipitis* CICC1960, *Candida parapsilosis* CICC1973, and *Rhodosporidium toruloides* CICC32489 were purchased from the China Center of Industrial Culture Collection (CICC). NADH and NADPH were purchased from Aladdin (Shanghai, China). Restriction enzymes were purchased from Takara (Kyoto, Japan). All commercially available chemicals were of analytical grade. Prochiral ketone, 7-chloro-1-[2-methyl-4-[(2-methylbenzoyl) amino] benzoyl]-5-oxo-2,3,4,5-tetrahydro-1H-1-benzazepine (PK1), was prepared in our lab following the process of Reddy. [1] The reference standards of (S)-tolvaptan and (R)-tolvaptan were supplied by Kewelchem (Shanghai, China).

### 1.2 Screening of transformed strains

Four yeast species obtained were cultivated in 4% (w/v) malt extract medium at 28 ºC for 48 h with shaking at 150 rpm. Wet cells were obtained after centrifugation (7104 ×*g*, 10 min) of the fermentation suspension. The bioconversion mixture (10 mL) consisted of 0.1 M potassium phosphate buffer (pH 7.2), 2 mg/mL PK1, 10.0 g/L glucose, and 1.0 g washed wet cells. The reactions were performed at 28 ºC with shaking at 150 rpm for 54 h. After removal of the cells by centrifugation, 100 μL of supernatant was dissolved in 900 μL acetonitrile, and the solution was filtered through a 0.22-μm PVDF syringe filter (Troody Technology, Shanghai, China) for HPLC analysis.

### 1.3 Cloning, expression, and purification of alcohol dehydrogenase

The genomic DNA of *P. stipitis* was extracted and purified using a TaKaRa Genomic DNA Extraction Kit (TaKaRa, Beijing, China). Primers were designed using Primer Premier version 5.5 (Premier, Palo Alto, CA, USA) based on the known or putative sequence of alcohol dehydrogenase from the registered *Scheffersomyces stipitis* CBS 6054 genome (NC_009068) in the NCBI database (Table S1). The DNA fragment of the alcohol dehydrogenase gene was amplified and then inserted into the expression vector pET-28a. The resulting plasmid was transformed into *E. coli* BL21 (DE3). The activity toward PK1 using

whole cells of recombinant *E. coli* was tested for choosing the effective alcohol dehydrogenase. Then, the purified amplification product and the plasmid were verified by DNA sequencing. Finally, the *psadh* gene, encoding alcohol dehydrogenase PsADH, was obtained. The recombinant *E. coli* BL21 (DE3)/pET28a-*psadh* was cultured at 28 °C in LB containing 0.5 mM kanamycin. When the optical density of the culture at 600 nm reached 0.6–0.8, IPTG was added to a final concentration of 0.1 mM, and cultivation was continued for another 14 h at 28 °C. The cells were harvested by centrifugation (8000 ×*g*, 10 min) at 4 °C, washed twice with 20 mM Tris-HCl buffer (pH 7.5), and lysed by sonication using an Ultrasonic Oscillator (Sonic Materials, Newton, CT, USA). The cell debris were removed by centrifugation (9000 ×*g*, 15 min) at 4 °C, and the supernatant was filtered through 0.45-μm filters and applied to an affinity column filled with precharged Ni Sepharose ™ (HisTrap HP; GE Healthcare, Piscataway, NJ, USA) using a purifier system (ÄKTA pure; GE Healthcare). The unfixed proteins were eluted out with 20 mM Tris-HCl (pH 7.5) and 0.5 M NaCl. Subsequently the column was eluted with a solution of 20 mM Tris-HCl (pH 7.5), 0.5 M NaCl, and 0.5 M imidazole. The collected enzyme solution was desalted and concentrated by an ultrafiltration membrane (Ultracel-10; Millipore, Billerica, MA, USA). The purified enzyme was evaluated by Coomassie brilliant blue staining of SDS-polyacrylamide gel electrophoresis (PAGE) gels.

1.4 Protein structure homology modelling

Homology-built models of PsADH were generated using a three-dimensional model based on the crystal structures of CPCR2 (PDB ID: 4C4O), using SWISS-MODEL workspace (http://swissmodel.expasy.org//SWISS-MODEL.html), and visualized using PyMOL.[2]

1.5 Assay of enzyme activity and optimization of catalytic conditions

(1) Assay of enzyme activity

The activity of PsADH was assayed by measuring the change of NADH absorbance at 340 nm for 1 min on a Multi-Detection Microplate Reader (Synergy HT; BioTek, Winooski, VT,

USA). The reaction mixture (0.1 mL) of the enzyme assay was comprised of 20 mM Tris-HCl buffer (pH 7.5), 0.5 mM NADH (NADPH), 2 mg/mL PK1, and the appropriate enzyme. One unit (U) of enzyme activity was defined as the amount of enzyme that catalyses the formation and oxidation of 1 μmol of NADH (NADPH)/min. Each experiment was performed in triplicate.

(2) Optimization of catalytic conditions

The catalytic conversion system was composed of 0.5 mM NADH, 2 mg/mL PK1, and the appropriate enzyme with concentration of 1 mg/mL. The operation was conducted at the temperature of 30 °C. The pH was determined by adjusting the pH value. The pH was controlled by adopting a solution of 100 mM NaAC-HAC buffer (pH 4.0–6.0), 100 mM potassium phosphate buffer (pH 6.0–7.0), or 100 mM Tris-HCl buffer (pH 7.0–9.0). The optimal pH was obtained based on the highest enzyme activity.

The catalytic conversion system was composed of 0.5 mM NADH, 2 mg/mL PK1, and the appropriate enzyme with concentration of 1 mg/mL. The pH of 100 mM potassium phosphate buffer was maintained at 7.0. The optimal temperature was determined by changing the temperature from 20°C to 70°C. The optimal temperature corresponded to the highest enzyme activity.

1.6 Construction of the co-expression system of PsADH and CpFDH in *E. coli*

The formate dehydrogenase CpFDH gene was obtained through the amplification of *Candida parapsilosis* genomic DNA. The co-expression system containing PsADH and CpFDH was constructed using a Shine-Dalgarno sequence, with an aligned spacing sequence (GAAGGAGATATACC) as a linker. The forward primers were adh'_F: 5′-CGCGGATCCATGCCTGTTCTCGAACCCAAGGCT-3′ and CoII_F 5′-AAGCCATGAGAAGGAGATATACCATGTATCCGGATTTAAAAGGAAAAGTC-3′, and the reverse primers were fdh_R: 5′-CCGGAATTCACCGCGGCCTGCCTG-3′ and CoII_R: 5′- CGGATACATGGTATATCTCCTTCTCATGGCTTGAACACGACTCTACCTTC-3′. The fusion gene PsADH-SD-AS-CpFDH was cloned using a modified overlap-extension

technique on plasmid pGEX4T-1.[3] The recombinant plasmid pG-Ps-SD-Cp-G was transformed into the competent *E. coli* BL21 (DE3), and the positive *E. coli* BL21/pG-Ps-SD-Cp-G clones were verified by DNA determination. The co-expression of PsADH and CpFDH proteins in the recombinant strains was analyzed by SDS-PAGE.

1.7 HPLC analytical methods

The retention time for (S)-tolvaptan and (R)-tolvaptan was identified by HPLC, equipped with a Chiralpak column OD-H (4.6 × 250 mm, 5.0 μM, Daicel, Osaka, Japan) and a UV detector at 254 nm. The flow phase was consisted of 0.05% trifluoroacetic acid aqueous solution (40%) and acetonitrile (60%). Its flow rate was adjusted to 1.0 mL/min with a linear gradient. All experiments were operated in triplicate. Statistical analysis was performed by one-way ANOVA (with LSD post-hoc analysis), using SPSS version 19.0. The content of (S)-tolvaptan was determined by HPLC external standard.

1.8 Optimization of whole-cell catalytic system

The reaction mixture (100 mL) was consisted of 0.1 M potassium phosphate buffer (pH7.0), which consisted of 100 g/L of washed wet cells and 0.1M sodium formate, and an organic solvent (dimethylbenzene, petroleum ether, cyclohexane, n-hexane, n-heptane, dimethicone, dibutylphthalate, ethyl acetate, or soybean oil). The concentration of PK1 in organic solvent was 100 mg/mL. The ratio of the organic phase to water phase (bioconversion mixture) was 1 to 2.3. The mixture was stirred at 30 °C for 48 h. The organic layer was sampled (100 μL each) and dissolved in 900 μL of acetonitrile. Tolvaptan content was determined by HPLC. The bioconversion of PK1 was also analysed in different ratios of organic phase: water phase (1:9, 1:4, 1:2.3, 1:1.5, and 1:1). Then, the optimal temperature (28°C, 30 °C, 35°C, 37°C, 40°C), pH of buffer solution (5.0, 6.0, 7.0, 8.0), and initial concentration of PK1 in soybean oil (20 mg/mL, 50 mg/mL, 100 mg/mL, 200 mg/mL) in the catalytic system were tested. The concentration of PK1 and tolvaptan were analysed by HPLC after every 6 h of reaction and the reaction mixture was sampled for 54 h. The operation was repeated 3 times.

1.9 Bioconversion and preparation of products

A typical bioconversion system for preparing enantiopure tolvaptan in the optimized two-phase system was performed as follows. 1 g of PK1 (2.2 mmol) was dissolved in a 10 mL of soybean oil and heated to dissolve. The solution was added to 90 mL of 0.1M of potassium phosphate buffer with pH of 6.0, which consisted of 10 g of washed wet cells and 1 g of sodium formate. The emulsion was stirred at 37 °C for 36 h. After the reaction was complete, the oil phase was separated and extracted with dichloromethane (10 mL x2). The combined dichloromethane phase was evaporated in vacuum to dryness. The residue was purified by silica gel chromatography to give a white solid, 0.79 g (yield 80.4%). $^{1}$H NMR (600MHz, $CD_3OD$-$d_4$) δ: 7.57 (1H, d, *J*=2.4), 7.49 (1H, s), 7.39 (1H, d, *J*=7.8Hz), 7.32-7.35 (2H, m), 7.28-7.30 (1H, m), 7.22-7.27 (3H, m), 6.70-6.98 (1H, m), 6.80-6.81 (1H, d, *J*=8.4 Hz), 6.70-6.72 (1H, d, *J*=8.4 Hz), 5.00-5.02 (1H, d, *J*=9.0 Hz), 4.71-4.73 (1H, d, *J*=13.8 Hz), 2.82-2.86 (1H, m), 2.45 (2H, s), 2.41 (3H, s), 2.21-2.30 (1H, m), 2.08-2.12 (1H, m), 1.82-1.84 (1H, m), 1.72-1.75 (1H, m), 1.58-1.65 (1H, m); $^{13}$C NMR (150 MHz, $CD_3OD$-$d_4$): δ: 171.77, 171.37, 145.58, 140.73, 139.55, 138.12, 137.83, 136.91, 134.66, 133.11, 131.83, 131.14, 130.27, 128.28, 128.10, 127.88, 126.85, 126.32, 122.84, 117.88, 71.52, 47.83, 36.55, 26.87, 20.06, 19.66. HRMS (m/z) calcd. for $C_{26}H_{25}ClN_2O_3$ $[M+H]^+$: 449.1624, found: 449.1626.

1.10 Nucleotide sequence accession number

The nucleotide sequences of *psadh* and *cpfdh* have been deposited in the GenBank database with accession numbers MG181955 and MG181956.

**References:**


[S1] B.P. Reddy, K.R. Reddy, D.M. Reddy, M.J. Reddy and B.V. Krishna, US Patent, **2015**, 9024015.

[S2] D. Seeliger, B.L. de Groot. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J Comput Aided Mol Des.* **2010**, 24, 417–422.

[S3] R.H. Peng, A.S. Xiong, Q.H. Yao, *Appl Microbiol Biotechnol.* **2006**, 73, 234–240.

**Table S1.** Oligonucleotides used in this study

| Primer | Oligonucleotide sequences (5′-3′) | Primers | Oligonucleotide sequences (5′-3′) |
|---|---|---|---|
| F1 | ATGAARWSNACNGGNYTNTGYCAY | F11 | ATGWSNGTNCCNACNACNCARA |
| R1 | RTANACNARDATNGGYTTNACNAC | R11 | YTTNSWNGTRTCNACNACRTANC |
| F2 | ATGGCNGARATHAARGTNGCNATH | F12 | ATGWSNACNGCNCCNCCNAAYA |
| R2 | DATNCKRAANACNACYTTYTCNCC | R12 | YTTDATNACNGTYTTRTARTANG |
| F3 | ATGWSNACNWSNAARACNGTNCCN | F13 | ATGWSNCAYGARAARCCNAARG |
| R3 | NARYTTNSWRTCRAANGCYTTRTCR | R13 | NARRTAYTTYTCNACYTGYTTRT |
| F4 | ATGACNACNWSNMGNACNGTNCCN | F14 | ATGWSNGCNWSNCCNGTNCARA |
| R4 | CCAYTTNSWRTCRAANGCYTGRTCR | R14 | RTCRAANARNCCRTANACYTTYT |
| F5 | ATGWSNATHCCNGCNACNCARTAY | F15 | ATGACNACNGGNAARACNTAYT |
| R5 | NGGYTTRAANACNACNCKNCCYTC | R15 | RTANGGNARNGTYTTNCCRTCRT |
| F6 | ATGACNGTNCARGARACNACNATH | F16 | ATGACNACNGGNAARACNTAYT |
| R6 | NGCNGTRAANGCYTTRTCRAANCC | R16 | NGGNARNACYTGNCCRTCRTART |
| F7 | ATGGGNTAYCCNGAYACNTTYCAR | F17 | ATGGTNGCNYTNACNAAYGAYW |
| R7 | YTCRAAYTGYTTYTCRTARTCNGTN | R17 | NGTNSWRAARTCRTTCCANSWN |
| F8 | ATGWSNATHCCNAARACNCARGTN | F18 | ATGACNAARWSNGCNWSNAARC |
| R8 | NARRTTNSWYTTNCKRAARTTNGG | R18 | RTGRTCRTAYTCYTCNACRAARA |
| F9 | ATGWSNAARWSNACNWSNACNAC | F19 | ATGWSNWSNATHATHWSNWSNG |
| R9 | RTTRTTRAANGCYTTRTCRTARTCN | R19 | NGCNACNGGRTCYTTRTANARNG |
| F10 | ATGWSNATHCCNACNACNCARAAR | F20 | ATGCCNGCNACNAARGCNACNA |
| R10 | YTTNSWNGTRTCNACNACRTANCK | R20 | NARYTTNSWRTCNGCRTCRAANA |

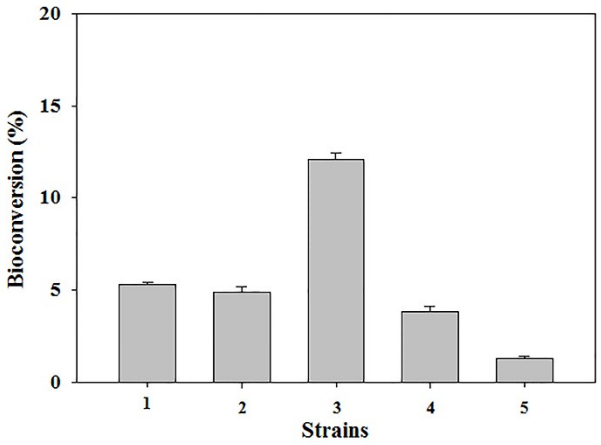



**Figure S1 Biocatalytic reduction of PK1 in various strains.** Commercially available microorganisms were tested for their potential for the preparation of tolvaptan via asymmetric reduction of PK1 in the single water phase. Numbers 1-5 represent the microorganisms *Yarrowia lipolytica* CICC 32187, *Saccharomyces cerevisiae* CICC1002, *Pichia stipitis* CICC1960, *Candida parapsilosis* CICC1973, and *Rhodosporidium toruloides* CICC32489, respectively. Error bars represent the standard deviation (n=3).

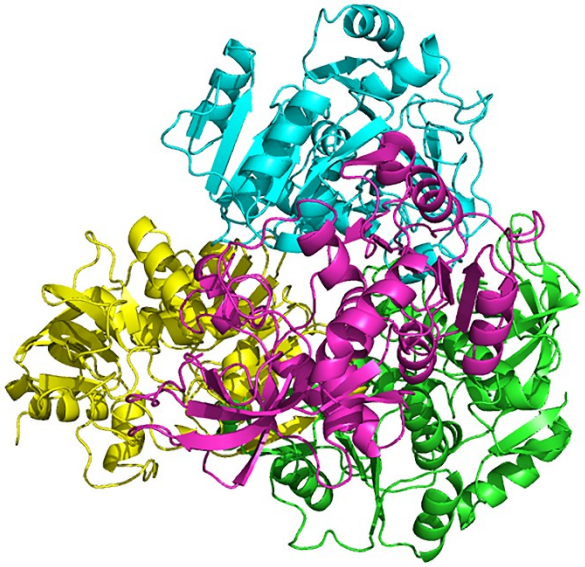


**Figure S2** Homologous modeling of PsADH. Three-dimensional model was constructed based on the template structure of CPCR2 (PDB ID: 4C4O), and the structure was prepared using the program PyMOL (DeLano Scientific).

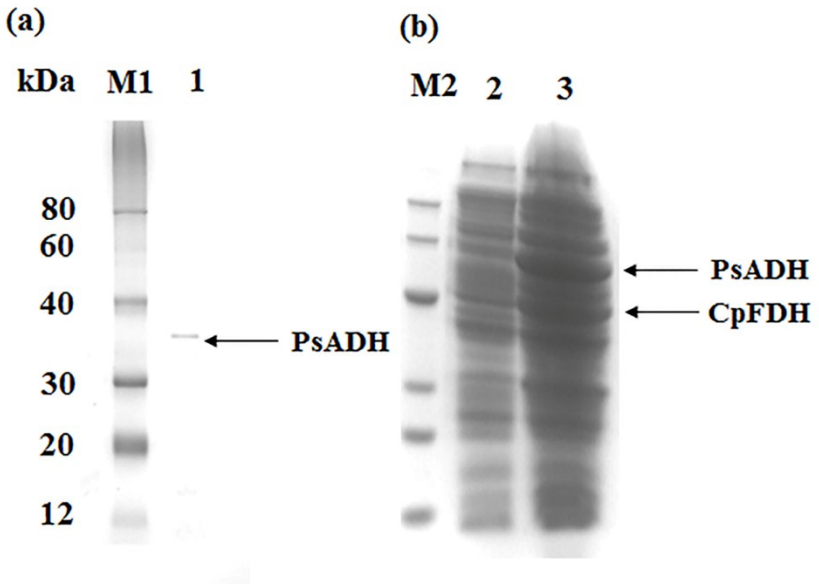



**Figure S3** SDS-PAGE analysis of purified PsADH (a) and the co-expression of PsADH and CpFDH (b). lane M1, molecular weight markers. lane 1, purified PsADH. lane M2, molecular weight markers; lane 2, recombinant *E. coli* BL21/ pG-Ps-SD-Cp-G without IPTG introduction; lane 3, recombinant *E. coli* BL21/ pG-Ps-SD-Cp-G with IPTG introduction.

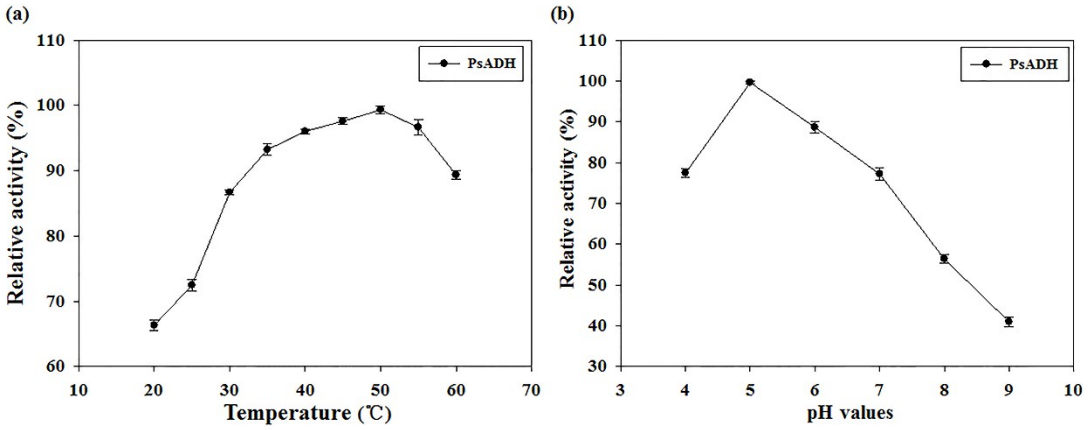



**Figure S4** Screening of the optimal temperature (a) and pH (b) for assay of PsADH activity. The relative activity of PsADH at each optimal temperature and pH for the reduction reactions was defined as 100%.

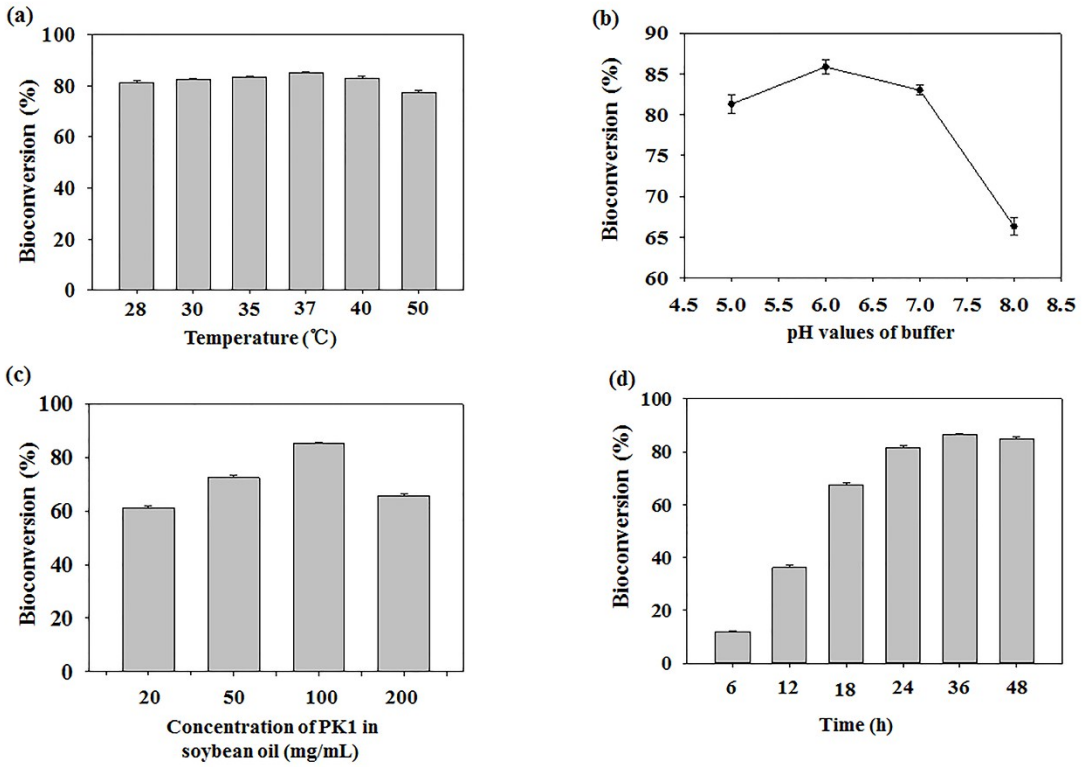
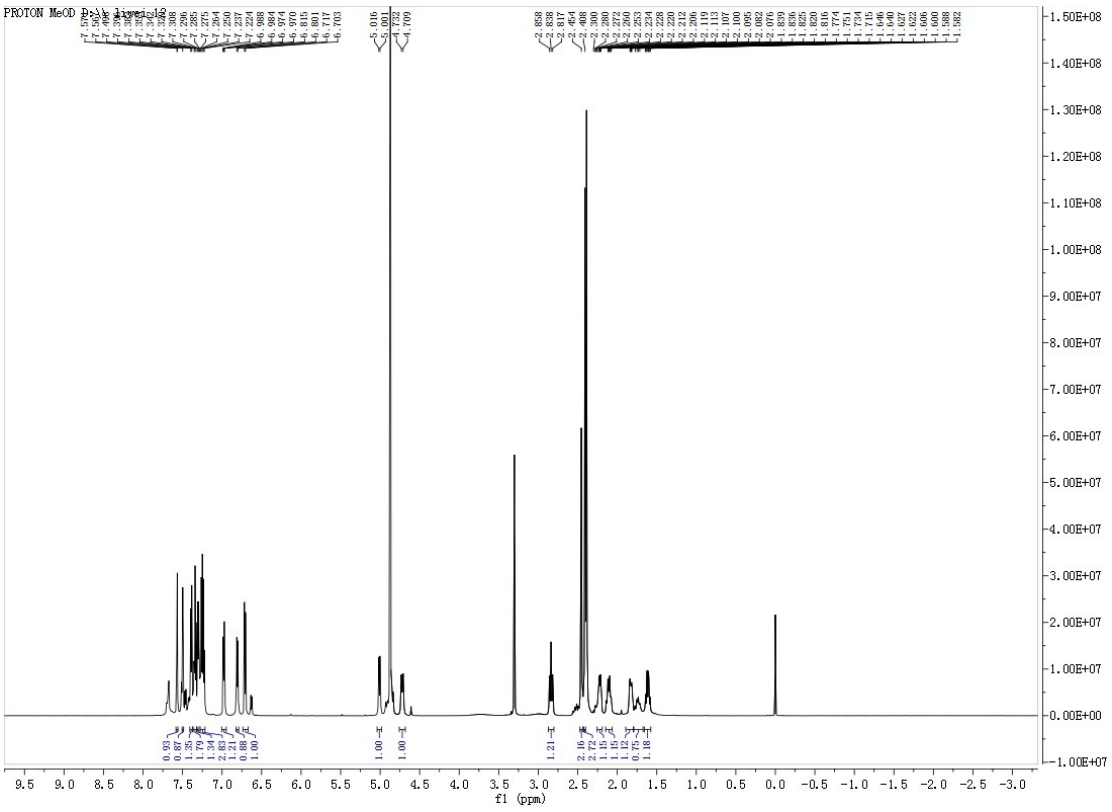
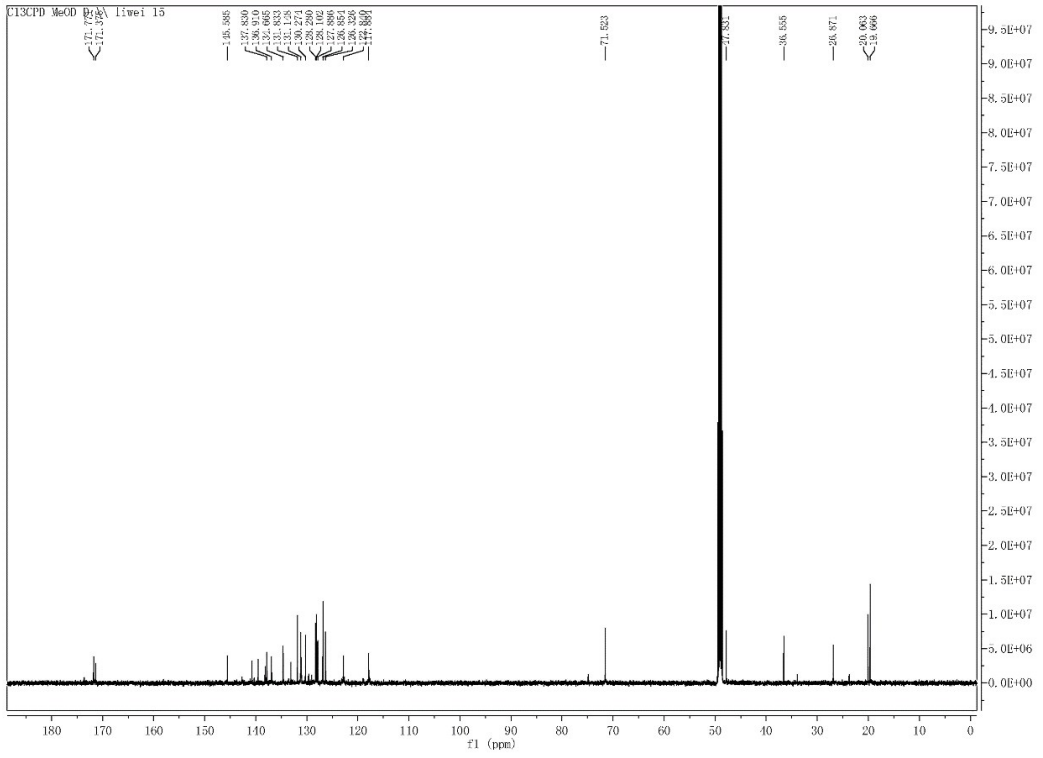



**Figure S5** The influences of temperature (a), pH (b), concentration of PK1 in soybean oil (c) and conversion time (d) on the bioconversion (%) in the catalytic system.


PROTON MeOD D:\A liwei.10
f1 (ppm)


C13CPD MeOD D:\A liwei 15
f1 (ppm)

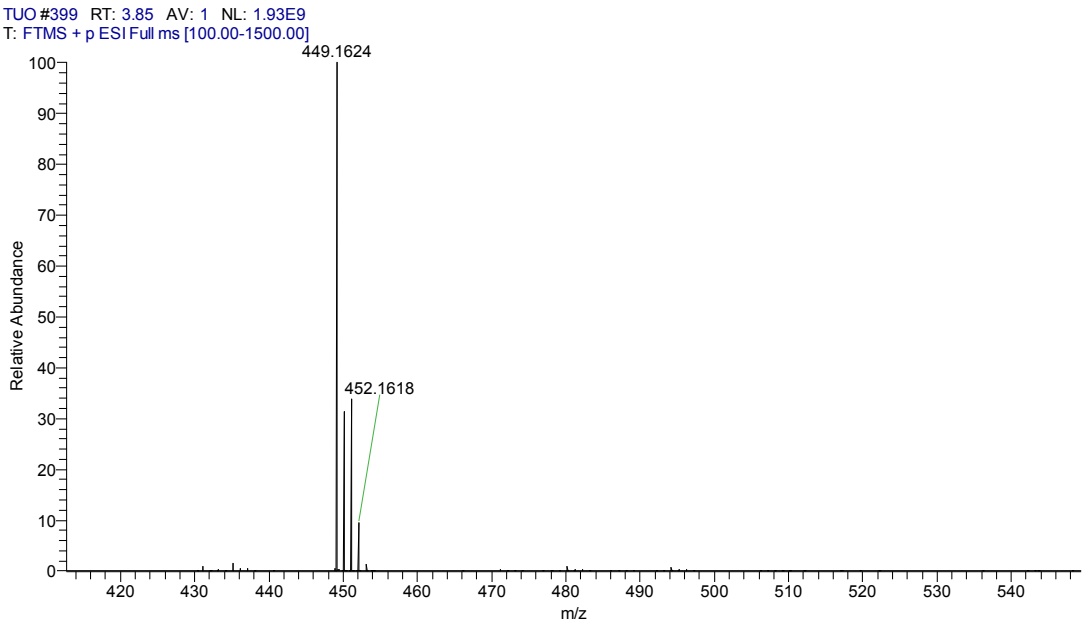



**Figure S6** $^{1}$H NMR, $^{13}$C NMR, HRMS spectra of the (S)-tolvaptan. $^{1}$H NMR (600 MHz, 25°C, $CD_3OD$-$d_4$), $^{13}$C NMR (150 MHz, 25°C, $CD_3OD$-$d_4$).