Molecular Characterization and Directed Evolution of a Metagenome-Derived L-Cysteine Sulfinate Decarboxylase

Jie Deng1,2, Qiaofen Wu2, Hua Gao2, Qian Ou2, Bo Wu2, Bing Yan1* and Chengjian Jiang1,2*

1Guangxi Key Laboratory of Mangrove Conservation and Utilization, Guangxi Mangrove Research Center, Guangxi Academy of Sciences, 92 Changqing Rd., Beihai, Guangxi, PR China
2State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, College of Life Science and Technology, Guangxi University, 100 Daxue Rd., Nanning, Guangxi, PR China

Received: July 8, 2017
Accepted: February 8, 2018

SUMMARY
L-Cysteine sulfinate decarboxylase (CSD, EC 4.1.1.29), the rate-limiting enzyme in taurine synthesis pathway, catalyzes L-cysteine sulfinic acid to form hypotaurine. Identification of the novel CSD that could improve the biosynthetic efficiency of taurine is important. An unexplored decarboxylase gene named undec1A was identified in a previous work through sequence-based screening of uncultured soil microorganisms. Random mutagenesis through sequential error-prone polymerase chain reaction was used in Undec1A. A mutant Undec1A-1180, which was obtained from mutagenesis library, had 5.62-fold higher specific activity than Undec1A at 35 °C and pH=7.0. Molecular docking results indicated that amino acid residues Ala235, Val237, Asp239, Ile267, Ala268, and Lys298 in the Undec1A-1180 protein helped recognize and catalyze the substrate molecules of L-cysteine sulfinic acid. These results could serve as a basis for elucidating the characteristics of the Undec1A-1180. Directed evolution technology is a convenient way to improve the biotechnological applications of metagenome-derived genes.

Key words: taurine, L-cysteine sulfinate decarboxylase, metagenomic library, sequential error-prone PCR

INTRODUCTION
Taurine (2-amino ethanesulfonic acid), an essential amino acid, is abundant in the cells of humans and other eukaryotes (1). It serves multiple physiological activities, including anti-inflammatory, analgesic, antipyretic, and hypoglycaemic; it also regulates nerve conduction and lipid digestion and absorption, participates in endocrine activity, increases cardiac contractility, and improves immunity, among others. This molecule is widely used in the food, feed, and medical industries (2).

Currently, taurine production mainly depends on the following two methods: chemical synthesis and direct biological extraction (3). Biosynthesis is attractive due to its many advantages, such as moderate production conditions and environmental friendliness. The main pathway of taurine synthesis in living cells involves cysteine sulfinate decarboxylase (CSD), which is a rate-limiting enzyme that determines taurine biosynthetic capability (4), and some enzymes which can oxidize the hypotaurine to form taurine (5). Therefore, isolation and identification of the novel CSD that could improve the biosynthetic efficiency of taurine are important.

L-Cysteine sulfinate decarboxylase, whose activity directly restricts taurine synthesis, had been previously isolated from liver tissue (6). CSD from brain and liver tissues exerts the same activity towards cysteine sulfinic acid and cysteic acid, respectively. However, CSD from brain tissue is not completely dependent on 5’-pyridoxal phosphate (7). Furthermore, CSD isolated from brain tissue serves as glutamic acid decarboxylase to synthesize γ-aminobutyric acid. Subsequently, CSD from brain tissues of buffalo, dog or mouse could be divided into CSDI and CSDII, of which the latter serves as glutamic acid decarboxylase (8,9).

Current research on CSD is limited to eukaryotic cells (10). As it is known, microorganisms isolated from the environment using pure culture techniques make less than 1 % of
the total microbial population; others are uncultured microorganisms that may have more extensive diversity in physiological and biochemical characteristics (11-13). Metagenomic technology, an effective strategy to study uncultured microorganisms, includes genomic DNA extraction from environmental samples, construction of metagenomics libraries, and screening the libraries to find some interesting genes and active substances (14). Although a wide variety of novel enzymes have been isolated and identified from environmental samples, little data is available concerning CSD from uncultured microbes.

Our previous research demonstrated that the Undec1A could catalyze the l-cysteine to form cysteamine using liquid chromatography-mass spectrometry, and detailed biochemical characterization was done (15). Compared with the amino acid sequence of CSDs from other sources, we found that Undec1A showed moderate similarity to them, and contained some conserved CSD domains. The previous study had revealed that the undec1A gene encodes a protein with CSD activity. The purpose of this study is to obtain more active mutants through protein engineering. One interesting variant, Undec1A-1180, showed increased decarboxylase activity. The identification and directed evolution of a metagenome-derived Undec1A also broadened our understanding of the mechanism of the metagenome-derived bifunctional CSD.

**MATERIALS AND METHODS**

**Chemicals, reagents and equipment**

T4 DNA ligases, Sall and Smal enzymes were obtained from Thermo Fisher (Waltham, MA, USA). Taq DNA polymerase and buffers were obtained from Takara Bio Inc (Kyoto, Japan). Primers, antibiotic, glycerol, sodium chloride and hydrochloric acid were obtained from Sangon Biotech Co Ltd (Shanghai, PR China). Cysteine sulfinic acid (CSA), pyridoxal 5’-phosphate (PLP), amino acids and substrate analogues were obtained from Merck KGaA (Darmstadt, Germany). Gel extraction kit and plasmid kit were obtained from Omega Bio-tek (Norcross, GA, USA). Tryptone and yeast extract were obtained from Oxoid (Basingstoke, UK). Nick-el-nitritotriacetic acid (Ni-NTA) agarose resin was obtained from Qiagen (Hilden, Germany). PCR instrument was model PTC-200 (Bio-Rad, Hercules, CA, USA). Sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) instrument was model JY-SCZ4+ (JunYi, Beijing, PR China). Automatic amino acid analyzer was model 6300 (Beckman Coulter, Pasadena, CA, USA). High-performance liquid chromatography (HPLC) was model 1100 (Agilent Technologies, Santa Clara, CA, USA).

**Plasmids and strains**

*Escherichia coli* DH5α and *E. coli* BL21 (DE3)pLysS (both from Novagen Merck KGaA, Darmstadt, Germany) were selected as the clone host and the expression host, respectively. Plasmid pGEM-3Zf(+) (Promega Corporation, Fitchburg, WI, USA) and pETBlue-2 (Novagen, Merck KGaA) were chosen as the clone vector and the expression vector, respectively. The *E. coli* strains were incubated in Luria-Bertani (LB; 1.0 % tryptone, 0.5 % yeast extract, 1.0 % sodium chloride, pH=7.0) medium at 37 °C.

**Screening of mutants from mutagenic library**

A mutagenic library of the l-cysteine sulfinate decarboxylase gene undec1A was constructed (16,17). Sall enzyme site (underlined) was designed in forward primer and Smal enzyme site (underlined) was designed in reverse primer (5’-ATAGTCGACATGATCACCCCCCTTACGCGGCAAC-3’/5’-CGTCTAGATTGGAAACCAGGTAAGTACCTTCCG-3’). The first error-prone polymerase chain reaction (PCR) was amplified in 100-µL medium consisting of forward primer (10 pM), reverse primer (10 pM), 1×Taq buffer, plasmid DNA (20 ng), Taq DNA polymerase (10 U), MgCl2 (4 mM) and MnCl2 (0.5 mM). The second round of PCR was performed with descending concentrations of dNTPs (20–200 µM) and the addition of corresponding concentrations of MnCl2. The error-prone PCR program was as follows: 94 °C for 10 min, 30 cycles at 94 °C for 30 s, 65 °C for 35 s, and 72 °C for 2 min, and final extension step at 72 °C for 10 min.

The error-prone PCR products (1177 bp) were purified and digested with Sall and Smal enzymes, ligated into the pGEM-3Zf(+) vector that had digested with same enzymes, and then transformed into competent *E. coli* DH5α strain (18). Colony PCR and restriction digestion were used to test the positive clones. The decarboxylase mutants were obtained by an automatic amino acid analyzer (19), and plasmids were sequenced.

The interesting genetic variants were ligated into pET-Blue-2 vector and then transformed into competent *E. coli* BL21(DE3)pLysS strain. The correct clones were overexpressed and purified as described by Jiang and Wu (15). The interesting variants were purified by Ni-NTA agarose resin. The molecular mass of Undec1A-1180 was measured by denaturing discontinuous SDS-PAGE (19-21).

**DNA sequence analysis and gene structure characterization**

The Entrez server (22) was used to search relevant sequences and conserved domains. The deduced amino acid sequences were identified using ExPaSy translation tool (23). Align X in Vector NTI software (24) was used for sequence alignment analysis. GalaxyWEB server (25) and MolProbity server (26) were used for protein homology modelling analysis and modelling evaluation analysis, respectively. Protein Data Bank (27) was used for searching the appropriate template sequences and structures. The suitable template (PDB code: 5int) was selected for Undec1A-1180 and sequence identity between them was 47.12 %.

**Decarboxylase activity assay and biochemical characterization**

l-Cysteine sulfinate decarboxylase activity was performed as described by Agnello et al. (28). Enzymatic assays were conducted in 0.5 mL of 50 mM phosphate buffer (pH=6.8) containing 10 µM PLP, 0.5 µM purified CSD, and different concentrations of,...

CSA (0–15 mM). The mixture was reacted for 10 min at 35 °C, and then ended by adding 50 µL of 1 M hydrochloric acid. After centrifugation, the reaction products were detected by HPLC. The sample was separated by Supelcosil C18 column (150 mm×4.6 mm, 3 µm; Merck KGaA) eluted with 10 mM potassium phosphate buffer (pH=6.8) including 2 % acetonitrile and monitored at 230 nm. One unit of CSD activity (U) was described as the amount of the enzyme required to generate 1 µmol of CSA per min under the above-mentioned conditions. All reactions were repeated in three independent experiments.

The optimal temperature for CSD activity was measured at pH=7.0 (phosphate buffer, 50 mM) with 10 mM CSA at different temperatures (20–50 °C). The optimal pH for CSD activity was tested in Na₂HPO₄-citrate buffer (pH=6.0–8.0), Tris-HCl buffer (pH=7.5–9.0), and glycine-NaOH buffer (pH=9.0–10.6), with 10 mM CSA as the substrate at 35 °C for 10 min. For the thermostability of Undec1A and its variant Undec1A-1180 protein, the corresponding purified enzyme had been pre-incubated up to 30 min at different temperatures (10–50 °C) in 50 mM phosphate buffer (pH=7.0) and then the residual activity was analyzed with 10 mM CSA as the substrate at 35 °C and pH=7.0 for 10 min and ended by adding hydrochloric acid. The enzyme assays were carried out in triplicate and the results were presented as the mean value±standard deviation.

Enzyme kinetic assays

The kinetic parameters (Kᵢ, V₀ and kₗ) of purified variants were tested by Lineweaver–Burk plots with CSA as substrate (29). The concentration of CSA ranged within 0.1–5.0 mM. The product was measured after the mixture reacted at 35 °C and pH=7.0 for 10 min and ended by adding hydrochloric acid. The enzyme assays were carried out in triplicate and the results were presented as the mean value±standard deviation.

RESULTS AND DISCUSSION

Construction and isolation of a mutant library of undec1A

A metagenome-derived decarboxylase Undec1A (GenBank accession number: AB870896) (15) acted as a new member of HFCD protein family, isolated and identified from uncultured microorganisms (30). The undec1A gene shares no identity with the known CSD gene at the DNA level (30). Amino acid sequence analysis showed that Undec1A contains some conserved residues that are similar to CSD from Danio rerio (GenBank accession number: NP_001007349.1; 25.2 % identical and 26.1 % similar) (30,31). Undec1A also shared the PLP-binding motif and the substrate recognition motif with other CSD proteins (29,30,32). In addition, Undec1A has the DOPA decarboxylase conserved domain to catalyze the decarboxylation (29,30).

In order to get undec1A variants with higher decarboxylase activity, error-prone PCR was used in this study. Mutation efficiency analysis of different combinations of Mg²⁺ and Mn²⁺ is shown in Table 1. An automatic amino acid analyzer was used to initially screen the improved mutants from 10 000 mutants. Undec1A-1180 mutant showed the highest decarboxylase activity. Through the sequence alignment analysis of Undec1A and Undec1A-1180, we found that the amino acid substitutions were Val81Leu, Phe240Ser, Ile250Ser and Asp266Leu (Fig. 1).

Table 1. Mutation efficiency analysis of different combinations of Mg²⁺ and Mn²⁺

<table>
<thead>
<tr>
<th>Combination number</th>
<th>Base pair conversion</th>
<th>Base pair transversion</th>
<th>Total mutations of base pairs</th>
<th>Total base pairs</th>
<th>Mutation rate of base pairs/%</th>
<th>Total mutation of amino acids</th>
<th>Total amino acids</th>
<th>Mutation rate of amino acids/%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AT→GC</td>
<td>AT→TA</td>
<td>10</td>
<td>10</td>
<td>0.46</td>
<td>20</td>
<td>1795</td>
<td>1.11</td>
</tr>
<tr>
<td>2</td>
<td>AT→GC</td>
<td>GC→AT</td>
<td>4</td>
<td>4</td>
<td>0.24</td>
<td>11</td>
<td>1795</td>
<td>0.61</td>
</tr>
<tr>
<td>3</td>
<td>AT→TA</td>
<td>GC→CG</td>
<td>7</td>
<td>7</td>
<td>0.39</td>
<td>14</td>
<td>1795</td>
<td>0.78</td>
</tr>
<tr>
<td>4</td>
<td>AT→TA</td>
<td>GC→TG</td>
<td>15</td>
<td>15</td>
<td>0.58</td>
<td>22</td>
<td>1795</td>
<td>1.22</td>
</tr>
</tbody>
</table>

The three-dimensional structure of Undec1A-1180 used Sint (plant chloroplast protease) as a template for homology modelling (Fig. 2a). The identity of Undec1A and Undec1A-1180 with the template was 38 %. Recently, a substrate recognition motif with three-residues in the active centre of human CSD has been identified to affect its catalytic efficiency (32). Sequence alignment of CSDs from other sources showed that this motif was not conserved, especially the latter two residues. The motif was also found in fish species such as Japanese flounder, yellowtail, large yellow croaker and medaka (29,32). Among them, the Japanese flounder and yellowtail
have a limited biosynthetic capability to taurine, but other fish species display unknown taurine biosynthetic capabilities (29,33).

Molecular docking results indicated that amino acid residues Ala235, Val237, Asp239, Ile267, Ala268 and Lys298 in the Undec1A-1180 protein contribute to the recognition and combination of the substrate molecules of CSA (Fig. 2b). These amino acid residues could also form hydrogen bonds with CSA. The results revealed that the Undec1A-1180 protein has a similar substrate recognition and catalysis model as the known CSDs. The optimal reaction pH can improve the interaction capacity between the binding sites of the enzyme and substrate molecule in a microenvironment (34,35). We speculated that the pK_a of Undec1A-1180, which occurred in amino acid substitutions in the active centre, was not affected.

Physicochemical characterization of Undec1A-1180

Undec1A-1180 was expressed in E. coli strain with pET-Blue-2 vector (28), and recombinant protein was purified to homogeneity. The enzymatic reaction products were detected by HPLC (Fig. 3). The final product derivatives consisted of two peaks. The peak at 2.569 min was hypotaurine, which had the same retention time as hypotaurine standard. The other peak at 2.768 min was the substrate CSA that had the same retention time as CSA standard. Therefore, the Undec1A-1180 has the ability to catalyze CSA to hypotaurine.

The optimum pH of recombinant Undec1A-1180 protein was measured at pH=4.5–10.0. The result showed that the maximum activity of Undec1A-1180 was achieved at pH=7.0. Compared with Undec1A (30), we found that the optimum pH of Undec1A-1180 had not changed. The observed pH range of Undec1A-1180 was different from the reported characteristics of CSDs from eukaryotes (36).

The recombinant Undec1A-1180 protein had high activity at 30–40 °C; the highest activity was observed at 35 °C, which is the same as the wild type protein (30). Compared with the other CSDs, Undec1A-1180 exhibited similar optimal temperature (35 °C) (37). The thermostability of the purified Undec1A and Undec1A-1180 is shown in Fig. 4. In the absence of substrate, the activity of purified decarboxylase decreased dramatically if the temperature was above 30 °C. These results also revealed that both Undec1A and Undec1A-1180 proteins are similarly sensitive to the change of temperatures.

The relative decarboxylation rates of various substrates by Undec1A-1180 are shown in Table 2. The Undec1A (30) and Undec1A-1180 proteins can both effectively decarboxylate L-cysteine, L-asparaginate, L-glutamate, L-cysteine sulfinic acid and L-glutamate. However, the Undec1A-1180 cannot be activated by L-alanine.
The initial rate of Undec1A-1180 was tested under optimal conditions with different concentrations of CSA. The kinetic parameters of Undec1A-1180 were analysed by Lineweaver–Burk plots. Undec1A has an apparent K_m value of (1.56±0.02) mM, a v_max value of (48.5±2.0) µM/min and a k_cat value of (45.8±1.3) min⁻¹ (30), whereas Undec1A-1180 has an apparent K_m value of (1.10±0.02) mM, a v_max value of (108.7±3.6) µM/min, a k_cat value of (88.6±2.1) min⁻¹ and a k_cat/K_m value of 80.2 min/mM. The k_cat of the Undec1A-1180 protein was approximately 1.9-fold higher than that of the wild protein. The k_cat of the Undec1A protein was higher than those of the CSDs from the wild-type Japanese flounder (29).

Compared with the Undec1A (30), the Undec1A-1180 showed similar optimum reaction conditions, had an enhanced affinity, and could better decarboxylize the CSA. The specific activity of Undec1A-1180 was (24.1±1.6) U/mg, which was 5.62-fold higher than that of the Undec1A protein.

Recent research has indicated that CSD from Synchococcus sp. PCC 7335 could decarboxylize CSA to hypotaurine, and taurine can be accumulated in this strain (28). Considering that some CSD recognition motifs are present in some genomes of marine bacteria, the authors evaluated the decarboxylases found in bacteria including CSA recognition motifs, which could decarboxylize CSA to hypotaurine, that had been annotated as glutamate decarboxylase. CSD homologues existed in some bacteria and the genes were found in operons including cysteine dioxygenases, which had the ability to transform L-cysteine into CSA. This reaction may give a clue to support the idea that the bacterial taurine synthesis pathway may exist in prokaryotes. Current research demonstrates that Undec1A-1180 exerts CSD activity. Moreover, it shows its better catalytic ability than that of Undec1A protein to decarboxylize L-cysteine to cysteamine. This research also extends the knowledge on the novel genes from uncultured microorganisms and provides a new reference to solve the bottleneck problem in the biosynthesis of taurine in vitro.

### CONCLUSIONS

The detailed biochemical characteristics, including the molecular profile, pH-activity profile, temperature-activity profile, specific activity, and enzyme kinetics of Undec1A-1180 protein were analyzed. Compared with the wild-type Undec1A, the Undec1A-1180 exhibited 5.62-fold increase of CSD activity under the optimum reaction conditions. These results could serve as a basis for elucidating the properties of the Undec1A. Directed evolution technology with sequential error-prone PCR is a convenient way to improve the biotechnological applications of metagenome-derived genes.

### ACKNOWLEDGEMENTS

This research was supported by the Science and Technology Basic Resources Investigation Program of China (Grant no. 2017FY100704), the National Natural Science Foundation of China (Grant no. 21262003), the Natural Science Foundation of Guangxi Zhuang Autonomous Region of China (Grant no. 2017JJB130020), the Open Research Fund Program of Guangxi Key Lab of Mangrove Conservation and Utilization (Grant no. GKLMC-201702), and the Distinguished Employment Offered Unit of Guangxi for Conservation and Ecological Monitoring of Mangroves and Seagrasses.

### REFERENCES


### Table 2. Relative decarboxylase activities of Undec1A-1180 proteins towards different substrates

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Relative activity/%</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Cysteine</td>
<td>99.3±1.6</td>
</tr>
<tr>
<td>L-Cysteine sulfinate acid</td>
<td>89.3±3.2</td>
</tr>
<tr>
<td>L-Alanine</td>
<td>n.d.</td>
</tr>
<tr>
<td>L-Proline</td>
<td>21.7±2.7</td>
</tr>
<tr>
<td>L-Asparaginate</td>
<td>61.7±2.0</td>
</tr>
<tr>
<td>L-Glutamate</td>
<td>76.2±2.5</td>
</tr>
</tbody>
</table>

Values are expressed as mean±standard deviation of three independent tests; n.d.=not detected.


