Endophytic Bacillus sp. CY22 from a Balloon Flower (Platycodon grandiflorum) Produces Surfactin Isoforms

CHO, SOO JEONG\(^2\), SU YOUNG HONG\(^1\), JIN YOUNG KIM\(^1\), SANG RYEOL PARK\(^1\), MIN KEUN KIM\(^1\), WOO JIN LIM\(^1\), EUN CHULE SHIN\(^1\), EUN JU KIM\(^1\), YONG UN CHO\(^1\), AND HAN DAE YUN\(^1,2\)

\(^1\)Division of Applied Life Science and \(^2\)Research Institute of Life Science, Gyeongsang National University, Chinju 660-701, Korea

Received: April 4, 2003
Accepted: July 21, 2003

Abstract Surfactin is a mixture of cyclic lipopeptides built from variants of a heptapeptide and a \(\beta\)-hydroxy fatty acid produced by several strains of Bacillus sp. Surfactin isoforms produced by endophytic Bacillus sp. CY22 from a balloon flower were isolated and characterized. It was found that the purified surfactin had three isoforms with protonated masses of m/z 1,008, 1,022, and 1,036, and different structures in combination with Na, K, Ca ions using MALDI-TOF MS, ESI-MS/MS, and ICP MS, respectively. In the MS/MS analysis, the isolated surfactin had the identical amino acid sequence (LLVDLL) and hydroxy fatty acids (with 13 to 15 carbons in length), even though isolated from different Bacillus strains. The \(sfp\) gene, required for producing the surfactin, consisted of an open reading frame (ORF) of 675 bp encoding 224 amino acid residues with a signal peptide of 20 amino acids. The predicted amino acid sequence of \(sfp\) was very similar to that of \(lpa-8\).

Key words: Endophytic Bacillus sp. CY22, surfactin, ESI-MS/MS, MALDI-TOF-MS, ICP MS

Surfactin, a lipopeptide produced by several strains of Bacillus, was first discovered in 1968 and named surfactin, because of its exceptional surfactant activity [1]. It is well known to be one of the most powerful biosurfactants. In addition to its high surface properties, surfactin displays an array of amazing activities such as fibrin clotting [1, 7], erythrocytes lysis inhibition [26], emulsification, foaming [20, 29], inhibition of starfish oocyte maturation [35], antibiotic [3, 21, 25], antitumoral [17], antiviral, antmycoplasmic [37], hypocholesterolemic [13], and anti-HIV activities [37]. When associated with an antibiotic and antifungal iturin A, another lipopeptide coproduced by Bacillus subtilis, the surfactin has strong synergistic effects [12, 18].

The most general structural characteristic of surfactin is the presence of a heptapeptide with an LLDLLDL chiral sequence linked, via a lactone bond, to a \(\beta\)-hydroxy fatty acid with 13 to 15 atoms. Natural surfactin is a mixture of isoforms [15, 16], which differ slightly in their physicochemical properties due to variations in chain length and branching of their hydroxy fatty acid components [30] as well as due to substitutions of the amino acid components of the peptide ring [2, 4, 5, 8, 14, 27]. These variations, rather than being genetically determined, depend on the specific B. subtilis strain and the nutritional and environmental conditions [19, 28]. Hbrid [11] demonstrated that the antifungal and hemolytic activities of these agents are enhanced with the increase of the number of carbon atoms in their fatty acid chains, presumably due to stronger interactions with biomembranes. On the other hand, Kowall et al. [22] suggested that the biological activities of both surfactin and the lipopeptides of the iturin family (from B. subtilis) depend on their amino acid compositions and the sequence of their peptide rings as well as on the nature of their lipid moiety.

So far, three isoforms in natural surfactin with substitutions of the L-Leu in position 7 of the main product by L-Val and L-Ile have been characterized by 2D NMR spectroscopy in combination with chemical analysis [2, 14, 27]. In addition, replacements of L-Val in position 4 by L-Ala, L-Leu, or L-Ile have been found when B. subtilis S499 was grown in a modified Landy medium containing these amino acids as a nitrogen source instead of glutamic acid [5]. Recently, Grangemard et al. [8] isolated and characterized (Ile4, Ile7)-surfactin and (Ile2, Ile4, Ile7)-surfactin, which were formed by the addition of L-isoleucine to the culture medium. The three-dimensional structure of the surfactin has been investigated on the basis

*Corresponding author Phone: 82-55-751-5469; Fax: 82-55-757-0178; E-mail: hdyun@nongae.gsnu.ac.kr
of NOE NMR data and molecular dynamics simulations in a vacuum [4, 14]. A horse saddle topology has been proposed for the peptide ring atoms. The two polar L-Glu and L-Asp side-chains form a claw structure, which helps to understand the cation binding and the ionophoric properties of surfactin [32, 34].

This paper reports on the preparative separation of the surfactin mixture isolated from endophytic Bacillus sp. CY22 [6, 11] and on the cloning of an essential gene, named as sfp22, for surfactin production. The surfactin isoforms were fractionated by reversed-phase HPLC and characterized by its MALDI-TOF and Q-TOF masses.

**MATERIALS AND METHODS**

**General Techniques and Materials**

Molecular cloning and PCR procedures were used as described by Sambrook et al. [32]. Media for growth of endophytic Bacillus sp. CY22 from the balloon flower (Platycodon grandiflorum) has previously been described [6]. The solvents, chemicals, and silica gel 60 plates were products of Merck (Darmstadt, FRG).

**Isolation of Surfactin**

Isolation of surfactin produced by strain CY22 was essentially performed as described by Nakano et al. [26]. Cells were grown in number 3 medium (No. 3: 10 g polypeptone, 10 g glucose, 1 g KH₂PO₄, 0.5 g MgSO₄·7H₂O, pH 6.8) at 28°C. After 4 days of cultivation, the supernatant was collected by centrifugation and adjusted to pH 2.0 using concentrated HCl. The precipitate was collected by centrifugation and extracted three times with methanol. The methanolic extracts were concentrated.

**TLC**

For further analysis, the methanolic extracts were separated using silica gel 60 plates (Merck). Chloroform/methanol/water (65:25:4, v/v/v) was used as a developing solvent. The various spots were visualized by charring after spraying with concentrated H₂SO₄. For preparative isolation of the surfactin fractions, the corresponding spots were scratched out from the thin-layer chromatograms, and the silica gel material was extracted with methanol. The extract was purified by reversed-phase HPLC.

**Mass Spectrometric Analysis of Surfactin Isoforms**

Surfactin was analyzed by using fast atom bombardment (FAB) mass spectrometry, matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF, Vg-instruments, Manchester) mass spectrometry, and inductively coupled plasma (ICP, Perkin-Elmer) mass spectrometry. FAB mass spectra were recorded on an MS 80 RF (Kratos Ltd.) spectrometer operating at a potential of 6–7 kV applied to the xenon gun used to form the fast-atom beam. Sample (4–5 µg) was introduced on a copper probe tip using a mixture of glycerol and triglycerol as matrix. The sample for the MALDI-TOF MS was mixed with an equal volume of saturated α-cyano-4-hydroxycinnamic acid in 70% acetonitrile/0.1% trifluoroacetic acid (v/v). One µl (2–3 pmol) was deposited on a sample plate and air-dried. Ions were accelerated with a voltage of 20 kV. The positive-ion and reflector mode was applied. ICP-MS was performed as described previously [38]. Samples were dissolved in 100% aqueous methanol before injection into the mass spectrometer.

**Amino Acid Analysis**

The amino acid composition analysis of the purified lipopeptides was carried out by the Pico-Tag method (Waters, Milford, U. S. A.) after hydrolysis in constantly boiling HCl (Sigma, St. Louis) at 110°C for 24 h. Derivatization of the hydrolyzed sample was accomplished using derivatizing solution (Ethanol/DW/Triethylamine/Phenylisothiocyanate: 7/1/1/1 by vol) for 15 min. PITC-derivatized free amino acids (PTH-derivatives) were applied to a 30 cm Pico-Tag Free Amino Acid Analysis column (3.9×300 mm), equilibrated with buffer A, and equipped with a Waters HPLC system (510 HPLC pump, 717 automatic sampler, 996 photodiode array detector, and Millennium 32 chromatography manager) and eluted with a linear gradient composed of buffer B (0, 14, 20, 46, 60, and 100%) at a flow rate of 1 ml/min at 46°C. Absorbance at 254 nm was measured. Buffer A was 140 mM sodium acetate (6% acetonitrile), and buffer B was 60% acetonitrile. The amino acid sequencing of purified lipopeptides was performed using ESI-MS/MS spectrometry. The sample was dissolved in 50% aqueous methanol containing 1% formic acid before injection into the mass spectrometer. ESI-MS/MS spectra were acquired by causing collisions between precursor ions and nitrogen collision gas at acceleration voltages of 50 V.

**Cloning of an Essential Gene for Surfactin Production**

To isolate an essential gene for surfactin production, PCR amplifications were performed with strain CY22 genomic DNA using degenerate primers. The degenerate primers were designed based on a highly conserved region between the surfactin-producing genes of the Bacillus species. The forward primer was 5'-GGT GAR CTK TGC RTM RG-3' and the reverse primer was 5'-CHT CWG GCA CWA CHG GC-3'. Purified DNA was used as a template in a 50 µl PCR reaction containing 1x PCR buffer, 10 mM dNTPs, 20 pmol primer, and 0.5 U Super-Therm DNA polymerase (JMR, Kent, U.K.). Thermal cycling conditions were: 2 min of denaturation at 95°C followed by 30 cycles at 94°C for 20 sec, 54°C for 20 sec, and 72°C for 1 min. The PCR product was purified using the NucleoGen Gel Extraction Kit (NucleoGen, Seoul, Korea), sequenced, and...
confirmed by a BLAST search. From these initially sequenced DNA regions, the downstream (srfA) and upstream (orfE) outside DNA were amplified by primer walking using degenerate primers 5’-CGC YKA YGC AMA GYT CAC C-3’ and 5’-CTG GAC MAS TAG CDY CGT C-3’, respectively. The PCR fragments were sequenced and cloned into a pGEM-T Easy vector (Promega). The clone was digested by a Sau3AI treatment, subcloned into pBluescript II SK (+) vector (Stratagene), and sequenced. All nucleotide sequencing was done by the dideoxy-chain termination method using the PRISM Ready Reaction Dye terminator/primer cycle sequencing kit (Perkin-Elmer Corp., Norwalk, CN, U.S.A.). The samples were analyzed with an automated DNA sequencer (model 310; Applied Biosystems, Foster City, CA, U.S.A.). Assembly of the nucleotide sequences and the amino acid sequence analysis were performed with the DNAMAN analysis system (Lynnon Biosoft, Quebec, Canada).

RESULTS

Detection of Lipopeptide Production of Bacillus sp. CY22

The culture supernatant of Bacillus sp. CY22 was precipitated with concentrated HCl and then extracted with methanol. The methanolic extracts were resolved by thin layer chromatography (TLC) on silica gel 60 plates using chloroform/methanol/water 65:25:5 (v/v/v) as the mobile phase. Lipopeptide was detected by charring with concentrated 10% sulfuric acid and showed a broad spot with \( R_f \) value of 0.5–0.55. The spot was extracted with methanol and purified using reversed-phase HPLC for further analysis (Fig. 1). The purified lipopeptide was identified and analyzed by FAB mass and MALDI-TOF mass spectrometry. The FAB mass spectrum displayed \([M+H]^+\) peaks at \( m/z \) 1,008.8, 1,022.8, and 1,036.8, and three \([M+H]^+\) peaks were separated by \( m/z \) 14 (data not shown). Comparison with the mass database indicated that the peak at \( m/z \) 1,036.8 was similar to the known lipopeptide surfactin, and at \( m/z \) 1,008.8 and 1,022.8 isoforms were deduced according to the respective chain lengths of their \( \beta \)-hydroxy fatty acids. By MALDI-TOF mass spectrometry measurements, three series of ions were found. A series of ions was found at \( m/z \) 1,030.68, 1,044.69, and 1,058.71

Fig. 1. Thin-layer chromatography (A) and HPLC spectrogram (B) of compound isolated from Bacillus sp. CY22. (A) Developing solvent: CHCl\(_3\)/MeOH/H\(_2\)O=65:25:4; Visualization: 10% H\(_2\)SO\(_4\) in water. (B) Developing solvent: CH\(_3\)CN/H\(_2\)O=1:1; Flow rate: 2.5 ml/min; Absorbance: 250 nm.

Fig. 2. MALDI-TOF mass spectrum of the lipopeptide surfactin fraction of Bacillus sp. CY22. The cyclic lipopeptidpeptide appeared as a complex mixture of several isoforms that show variations in the length of their \( \beta \)-hydroxy fatty acid moiety (n=7 to 9 CH\(_2\) groups). The \([M+H]^+\) peaks at \( m/z \) 1,045.7 and 1,059.7 were accompanied by the corresponding \([M+Na]^+\) peaks (\( m/z \) 1,030.7, 1,044.7, and 1,058.7).
and respective sodium adducts were deduced because of lower m/z. The signals at m/z 1,074.68 and 1,076.67 can be attributed to their potassium and calcium adducts, respectively (Fig. 2).

**Ion Detection by ICP-MS**
The purified lipopeptide was analyzed by ICP-mass spectrometry, and potassium and calcium ions were detected. This result agreed with that of the MALDI-TOF spectrometry analysis, which suggested the presence of ions.

**Amino Acid Analysis and Peptide Sequence**
Four amino acids were found in the purified HPLC fraction of strain CY22 after hydrolysis and derivatization with Marfey’s reagent: Asp, Glu, Val, and Leu in ratios of 1:1:1:4 (Table 1). The peptide sequence was deduced by interpreting the ESI-MS/MS spectrum of the precursor ion m/z 1,036.7, assuming preferential cleavage of the ring in the lactone bond in the collision chamber. Subtraction of 895.7 from the precursor ion m/z 1,008.7 results in a value of 113 (Leu), 895.7-782.7 results in a value of 113 (Leu), 782.7-667.8 leaves a value of 114.9 (Asp), 667.8-568.6 a value of 99.2 (Val), 568.6-455.6 a value of 113 (Leu), 455.6-342.6 a value of 113 (Leu), and 342.6-0 results in a value of 129+213.6 (Glu+C13) (Fig. 3A). Also, the peaks of 1,022.741 and 1,036.751 m/z appeared to be amino acid sequences identical to that of 1,008.7 (Figs. 3B and 3C). The ESI-MS/MS analysis proved that surfactin from strain CY22 is a lipopeptide, in which the carbon chain length of the β-hydroxy fatty acid isoforms ranged from 13 to 15.

**Amplification and Cloning of the srfD-sfp Region**
For the isolation of the gene required for the production of lipopeptide surfactin, a 186 bp fragment was amplified using degenerate primers derived from the highly conserved consensus sequences [HFNI-X(1)-SHS] and [IK-X(2)-GKG]. From the results of a sequenced and a BLAST search, the fragment showed homology with the sequence of the sfp gene, which is known to be the regulatory gene for surfactin production. From these initial sequences, the ‘outside’ regions of upstream and downstream were then amplified by primer walking using degenerate primers for each region. The PCR product of 3,500 bp in the downstream region was amplified and digested with Sau3AI.

---

**Table 1.** Amino acid composition of the surfactin from Bacillus sp. CY22.

<table>
<thead>
<tr>
<th>Amino acid/peptide</th>
<th>Val</th>
<th>Leu</th>
<th>Asp</th>
<th>Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surfactin mol/mol</td>
<td>1.1</td>
<td>4.4</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

---

**Fig. 3.** ESI-MS/MS fragmentation of the protonated ions, m/z of 1,030.7, 1,044.7, and 1,058.7, shown in Fig. 1.

The fragment values and the corresponding amino acid residues are shown with ESI-MS/MS fragmentation patterns.
Af534916, pCY100 carrying the phosphopantetheinyl transferase gene of Bacillus subtilis 168, BAA08991, A36931, the following proteins: D50562, to the right of that branch occurred among 100 trees. The sequences are of representative genomes from Bacillus subtilis 168/JH624, YB8 (P39135), and the surfactin synthetase operon (Fig. 4). The first ORF, sfp22, encodes a protein for the production of surfactin with a 72% similarity to phosphopantetheinyl transferase. Also, the deduced amino acid sequence of Sfp22 (AF534916) was shown to be similar to that of the following genes: Lpa-8 of Bacillus subtilis YB8 (BAA09125), Sfp of B. subtilis (A36931), Sfp of B. subtilis (CAB12151), Sfp of B. subtilis 168 trpC2 (BAA08991), Psf-1 of B. pumilus (P55810), Gsp of Brevibacillus brevis (P40683), and Sfp of Clostridium acetobutylicum ATCC824 (AAK79297). The phylogenetic analysis of the sfp gene based on a GenBank database indicated evolutionary relationships between sfp22 and the genes for lipopeptide production. The sfp22 was recovered in a monophyletic clade containing the genes of sfp of B. subtilis 168 (BAA08991), sfp of B. subtilis (A36931), and lpa-8 of B. subtilis YB8 (P39135), and all these relationships are supported by a high bootstrap value of 100% (Fig. 5).

**Fig. 4.** Comparison of the genomic structure of the surfactin operon and physical map of recombinant DNA.

(A) Comparison of the genomic structure of the surfactin operon in Bacillus subtilis 168 (top line) and that of the surfactin partial operon in Bacillus sp. CY22 (bottom line). (B) Physical map of recombinant DNA pCY100 carrying the phosphopantetheinyl transferase gene of Bacillus sp. CY22. The cleavage sites of restriction enzymes PstI, HindIII, EcoRV, and SacI are shown. pCY100 was constructed by cloning a 1,000 bp fragment of Bacillus sp. CY22 into pBluescript II SK+ vector.

Digested fragments were then cloned into pBluescript II SK+ and sequenced. From upstream, the PCR product of 700 bp was amplified and sequenced. The results of the sequence analysis revealed that amplified fragments were constructed from three ORFs, designated srf22D, aat22, and sfp22, respectively (AF534916). The multimodular structure of the srf22D, aat22, and sfp22 genes are shown in Fig. 4A.

**Nucleotide Sequence Analysis of the Gene Required for the Production of Surfactin**

The inserted 1,000 bp fragment in the pGEM-T Easy vector was sequenced using the deoxy chain-termination method. It contained one complete open reading frame (ORF). Figure 4 depicts the sfp22 structural gene with its flanking regions. The open reading frame contains 675 nucleotides and encodes a protein of 224 amino acid residues with a predicted molecular weight of 25,506 Da. The size of the sfp22 and the orientation of restriction cleavage sites were determined. The sfp22 contained the restriction sites for PstI, HindIII, EcoRV, and SacI (Fig. 4B).

**Amino Acid Sequence Similarities Between sfp22 and Other sfp Genes**

Sequence analysis revealed that the cluster of three ORFs, designated sfp22, aat22, and sfp22-D, belonged to a surfactin synthetase operon (Fig. 4). The first ORF, sfp22, encodes a protein for the production of surfactin with a 72% similarity to phosphopantetheinyl transferase. Also, the deduced amino acid sequence of Sfp22 (AF534916) was shown to be similar to that of the following genes: Lpa-8 of Bacillus subtilis YB8 (BAA09125), Sfp of B. subtilis (A36931), Sfp of B. subtilis (CAB12151), Sfp of B. subtilis 168 trpC2 (BAA08991), Psf-1 of B. pumilus (P55810), Gsp of Brevibacillus brevis (P40683), and Sfp of Clostridium acetobutylicum ATCC824 (AAK79297). The phylogenetic analysis of the sfp gene based on a GenBank database indicated evolutionary relationships between sfp22 and the genes for lipopeptide production. The sfp22 was recovered in a monophyletic clade containing the genes of sfp of B. subtilis 168 (BAA08991), sfp of B. subtilis (A36931), and lpa-8 of B. subtilis YB8 (P39135), and all these relationships are supported by a high bootstrap value of 100% (Fig. 5).

**Fig. 5.** Phylogenetic tree showing the evolutionary relationships of Sfp22 and other closely related amino acid sequences. The estimated genetic distance between sequences is proportional to the lengths of the horizontal lines connecting one sequence to another. This is a rooted tree. Bootstrap values of the major branch points are shown; they represent the number of times by which the group consisting of the species to the right of that branch occurred among 100 trees. The sequences are of the following proteins: D50562, Bacillus subtilis YB8, lpa-8; P99135, AF534916, Bacillus sp. CY22, sfp22; Bacillus subtilis 168JH624, sfp; A36931, Bacillus subtilis 168, sfp0; CAB12151, Bacillus subtilis 168, sfp; BAA08991, Bacillus subtilis 168/trpC2, sfp; P55810, Bacillus pumilus A-1, psf-1; AAK79297, Clostridium acetobutylicum ATCC824, CAC1329; P40683, Brevibacillus brevis ATCC9999, gsp.

**DISCUSSION**

Endophytic Bacillus sp. CY22 from the root of a balloon flower [31] produces surfactin, a lipopeptide with several isomers. In this report, the identification of a peptide synthetase partial operon is described, and the synthesis of this biosurfactant and three isoforms of surfactin are specified according to the carbon length of their fatty acid chains. Surfactin is known as a potent biosurfactant and may act as a surfactant with a higher critical micelle concentration than other surfactants. The biological activity of surfactin from B. subtilis depends both on the amino acid composition and the sequence of their peptide ring as well as on the nature of their lipid moiety. Hbrid [11] demonstrated that the hemolytic activity of surfactin is enhanced by an increase in the number of carbon atoms in their fatty acid side chains, and that this enhanced activity is most likely due to stronger interactions with biomembranes. The earlier studies used NMR analysis for...
identification and structural characterization of surfactin. However, this report is the first to describe the combined application of reversed-phase HPLC, MALDI-TOF MS (matrix-assisted laser desorption ionization with time of flight mass spectrometry) and Q-TOF MS (quadrupole orthogonal acceleration time-of-flight) in structural studies of surfactin. One of the advantages of this strategy is that the structure of surfactin is rapidly identified without NMR analysis. Thus, this method could reduce the time needed to identify and analyze structures of the surfactin. The surfactin produced by endophytic Bacillus sp. CY22 had a heptapeptide sequence LL VDLL linked to a β-hydroxy fatty acid with 13, 14, or 15 carbon atoms. This peptide sequence was the same as that of the natural surfactin [22].

Recently, Kowell et al. [22] has reported that the calcium-induced spectral changes of lipopeptide suggest a binding of the divalent ions to the surfactin COOH groups first (until the calcium-lipopeptide mole ratio reached 1), followed by bulk conformational changes (at higher mole ratios). The present study showed that monovalent ions as well as divalent ions could bind to the surfactin COOH groups using the MALDI-TOF MS.

In a related activity, a gene, sfp22 was cloned from Bacillus sp. CY22, and the nucleotides sequence was determined. The deduced amino acid sequence of sfp22 showed a sequence similar to that of lpe8 [36], sfp [24], sfp0 [9], sfp [33], sfp [22], and sfp [23]. The surfactin synthetase operon (srfA) was located in the downstream of the sfp22 gene. To obtain the complete sequence of the surfactin synthetase-encoding operon, an outside DNA amplification was performed by primer walking using degenerate primer. When the gene was inactivated in Bacillus sp. CY22 by integration mutation, surfactin was no longer produced (data not shown). From the result, sfp22 was identified as a gene involved in the production of the surfactin.

Since early human history, the root of a balloon flower has been used for medicinal purposes in Asian countries [31]. Specifically, the roots of more than ten-year-old balloon flower are generally considered to have the same medicinal effect as those of Korean ginseng in Korea. However, the medically effective ingredients still remain as “the secret workings of nature” in spite of many studies over a long time. Bacillus sp. CY22 was isolated from the interior of a ten-year-old balloon flower root, and the bacteria were found to be mainly localized within the aerenchyma and the intercellular spaces inside the balloon flower root, as had previously been found [6]. The ultimate goal of this investigation is to prove the hypothesis that surfactin may play some role in the medicinal effects of balloon flower root.

Further studies are required for cloning and analysis of the complete surfactin synthesis operon of endophytic Bacillus sp. CY22 from balloon flower. Identification of the medicinal effects of surfactin and improvements of surfactin production are also required.

Acknowledgments

This work was supported by a Biotechnology Research Foundation Grant of Gyeongnam Province, Korea. S.Y.H. is the recipient of a BK21 fellowship from the Ministry of Education, and S.J.C. is the recipient of a Post-doctoral Fellowship Program of Korea Science & Engineering Foundation (KOSEF).

References