Solution Structure of LCI, a Novel Antimicrobial Peptide from *Bacillus subtilis*

Weibin Gong,† Jinfeng Wang,§ Zhangliang Chen,† Bin Xia,*‡† and Guangying Lu,*†

†The National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing, China 100871
‡Institute of Chemistry and Molecular Engineering, Peking University, Beijing, China 100871
§Institute of Biophysics, Chinese Academy of Sciences, Beijing, China 100101

**ABSTRACT**: LCI, a 47-residue cationic antimicrobial peptide (AMP) found in *Bacillus subtilis*, is one of the main effective components that have strong antimicrobial activity against *Xanthomonas campestris pv Oryzae* and *Pseudomonas solanacearum* PE1, etc. To provide insight into the activity of the peptide, we used nuclear magnetic resonance spectroscopy to determine the structure of recombinant LCI. The solution structure of LCI has a novel topology, containing a four-strand antiparallel β-sheet as the dominant secondary structure. It is the first structure of the LCI protein family. Different from any known β-structure AMPs, LCI contains no disulfide bridge or circular structure, suggesting that LCI is also a novel β-structure AMP.

Antimicrobial peptides (AMPs) exist in different organisms throughout bacteria and eukaryotic cells. Today, with the emergence of antibiotic-resistant pathogenic bacteria, AMPs have been regarded as important and promising candidates for novel antibiotics.¹–³ The amino acid sequence and structure of AMPs show considerable diversity, which can be classified into different types: α-helical structure forming amphipathic surface, β-structure, α/β-structure, and extended structure enriched for one or two specific residues, like proline, arginine, glycine, etc.⁴–⁸ Most β- and α/β-structured AMPs contain disulfide bridges, which are thought to be important for stability. The mechanisms of function of AMPs have also been intensively studied, and different models have been proposed.³,⁸,⁹

The antimicrobial peptide LCI was first screened and isolated by Liu et al. from a *Bacillus subtilis* strain named A014 that possesses very strong antagonistic activity against the Gram-negative pathogen *Xanthomonas campestris pv Oryzae*¹⁰ of rice leaf-blight disease, which is a serious threat to rice production and causes great losses in yields in most rice fields annually. Until now, there has been no efficient method for controlling this disease. In addition, LCI also has antagonistic activity against Gram-negative bacterium *Pseudomonas solanacearum* PE1, but it does not inhibit *Erwinia carotovora* subsp. *Carotovora* or *Escherichia coli*. The antimicrobial peptide LCI contains 47 residues (Figure 1). Its theoretical molecular mass is 5460 Da, and its isoelectric point is 10.25, indicating that LCI is a cationic antimicrobial peptide. The amino acid sequence of LCI contains no Cys, His, or Met residues, while there are 10 aromatic residues (three Phe, four Tyr, and three Trp residues) in the sequence. Heated for 20 min at 60, 80, and 100 °C, it maintains 100, 85.3, and 12.5% of its antimicrobial activity, respectively. This indicates that its three-dimensional structure should be highly thermally stable. After treatment with trypsin, pepsin, and lysozyme, it maintains 81.5, 90.5, and 100% of its activity, respectively. However, it lost activity after being treated with pronase E and proteinase K.¹⁰ With regard to the secondary structure of LCI,¹¹ different conformation contents calculated from circular dichroism (CD) are as follows: 4.7–6.0% α-helix, 67.2–72.8% β-sheet, and 21.2–28.1% coil and β-turn, consistent with the predicted secondary structures of four β-strands.

Because of the low yield of LCI secreted by wild *B. subtilis* A014, it is difficult to conduct structural and functional studies. Recently, an LCI DNA fragment was chemically synthesized using plant-preferred codons, according to the native LCI amino acid sequence, and then used to construct plasmid pBVAB16 in our laboratory. LCI was expressed in large quantities with an expression system using DH5α. Its molecular mass is 5464 Da (pH 7.8) as measured by mass spectrometry,¹² indicating the recombinant LCI has had the methionine encoded by the start codon of the plasmid removed like native LCI. The molecular mass of the recombinant LCI is consistent with that of the wild type, according to sodium dodecyl sulfate–polyacrylamide gel
electrophoresis and gel filtration. In addition, the recombinant LCI was stable at 60 °C, showing thermal stability comparable to that of the native protein. It repressed the growth of X. campestris pv Oryzea G strains at 10 μM (unpublished data). Wild-type LCI repressed 50% of the growth of X. campestris pv Oryzea G strains at 0.81 μM.

Here we report the nuclear magnetic resonance (NMR) solution structure of LCI expressed in E. coli. A DALI search indicates that the structure of LCI adopts a novel topology. It is a β-sheet structure without any disulfide bridge, representing a novel subgroup of β-structure AMP.

**RESULTS AND DISCUSSION**

**LCI-Homologous Proteins.** The genome sequences of *B. amyloliquefaciens* strains FZB42 and C31 have been determined, each of which also encodes a LCI-homologous sequence, sharing with the LCI sequence of *B. subtilis* 98% identity for 42 residues and 94% for 46 residues (Figure 1). In addition, the LCI genes in both FZB42 and C31 encode an additional N-terminal 48-residue fragment containing a 25-residue signal peptide according to SignalP. As LCI from *B. subtilis* strain A014 was found to be a secreted peptide, it seems that the LCI may also contain a similar signal peptide. Unexpectedly, the LCI genes from *B. subtilis* strain A014 and *B. amyloliquefaciens* strains FZB42 and C31 were not found in the *B. subtilis* genome. This fact means that the LCI gene may reside in the *B. subtilis* dissociative plasmids.

**Solution Structure of LCI.** Figure 2 exhibits 3J_HNHA coupling constant values (>8 or <5 Hz), chemical shift deviations of H atoms from random coils, and the NOE connectivity for the secondary structures. The results indicate that residues 3–7, 15–17, 22–24, 28–31, and 36–41 are in the β-conformation. These H-bond restraints were verified by the H-D exchange experiments, in which NH signals of Val5, Phe16, Leu18, Trp23, Tyr29, Asp31, and Tyr36 show slow exchange with D2O. A set of 100 structures of LCI was calculated in AMBER with 1096 unambiguous NOE restraints, 19 dihedral angle restraints from experimental data, and 19 H-bond restraints. Of the 100 final structures, 20 structures with the lowest energies were chosen to represent the solution structure of LCI, in which no distance violation is greater than 0.2 Å and no dihedral violation is larger than 5°. The structural statistics are listed in Table 1. Analysis using PROCHECK-NMR indicates that 71.2% of all the non-Pro/Gly residues were in the most favored regions of the Ramachandran plot.

Panels A and B of Figure 3 exhibit the ribbon diagram of the mean structure of LCI and the superposition of the 20 structures of LCI with the lowest energy, respectively. The rmsd values of the structure ensemble versus mean structure were 0.17 and 0.32 Å for the secondary structure backbone heavy atoms and all heavy atoms, respectively. The overall structure is an antiparallel β-sheet consisting of four β-strands, including Lys3–Ser (β1), Ser15–Val17 (β2), Lys18–Asp31 (β3), and Tyr29–Tyr41 (β4). In the structure, S1 of the Supporting Information exhibits some assigned cross-peaks of W23 Hr1 and V43 HN (Figure S1 of the Supporting Information). H-Bonds in the secondary structure were added as distance constraints in structure calculation. Dihedral angle constraints were derived from 3J_HNHA coupling constants using an E-COSY experiment and CSI analysis. Structure calculation was performed with CNS version 1.22 and further refined using AMBER7. Finally, the 20 lowest-energy structures were selected from all 100 structures to represent the solution structure of LCI. The final structures were evaluated using PROCHECK-NMR.

**Protein Data Bank Entry.** The three-dimensional structure of *B. subtilis* LCI has been deposited in the RCSB Protein Data Bank (PDB) as entry 2B9K.
strands $\beta_2$ and $\beta_3$ form a $\beta$-hairpin connected by residues Asp$^{19}$ and Gly$^{20}$, and $\beta_3$ and $\beta_4$ form another $\beta$-hairpin connected by Ser$^{32}$–Gly$^{35}$. Ser$^{32}$–Gly$^{35}$ form a type I $\beta$-turn as the Ser$^{32}$ C$\alpha$–Gly$^{35}$ C$\alpha$ distance in the mean structure was 4.7 Å. Strands $\beta_1$ and $\beta_2$ are connected by a well-defined seven-residue loop (Pro$^8$–Ala$^{14}$), and the backbone rmsd of this loop is 0.29 Å, meaning that it has a fairly rigid conformation. The large $^1$H–$^{15}$N heteronuclear NOE values of the residues in this loop also support the rigidity of this loop (Figure 4). In the final structure ensemble, Arg$^{46}$ and Lys$^{47}$ show large structural flexibility, as indicated by the low $^1$H–$^{15}$N heteronuclear NOE values (<0.4) of these two residues. The mobility of the positively charged Arg$^{46}$-Lys$^{47}$ motif at the end of C-terminus may help LCI to recognize negatively charged molecules such as lipopolysaccharide (LPS), in the target membrane. An ~20% decrease in antimicrobial activity due to trypsin digestion may be attributed to the removal of the C-terminal Lys$^{47}$ from the Arg$^{46}$-Lys$^{47}$ motif, which is the digestion site of trypsin and accessible for enzyme attack, while other fragments form considerably compact structure, possessing resistance to digestion by pepsin and trypsin.

In such a small compact structure, LCI contains a hydrophobic core formed by Val$^5$, Tyr$^{41}$, and Trp$^{44}$. This hydrophobic core, as

Figure 2. $^3$J$_{HNHa}$ coupling constants, chemical shift deviations of H$\alpha$ from the random coil state, and NOE connectivities in the secondary structure elements. (A) $^3$J$_{HNHa}$ coupling constant values and chemical shift deviations of H$\alpha$ along the amino acid sequence of LCI. +1 and -1 denote the $^3$J$_{HNHa}$ values corresponding to >8.0 and <5.0 Hz, respectively. CSI labels $^+$, $^-$, and 0 indicate that the CSI value is more than, less than, and around 0.1 ppm, respectively. (B) NOE connectivities in the two-strand $\beta$-sheet and the three-strand $\beta$-sheet, together with the $\beta$-turn. Observed NOEs are shown by arrows, and dashed lines show the predicted hydrogen bonds. Two tight turns are also identified by NOE networks.
well as all 23 H-bonds (Figure 2) in the structure, may contribute to the considerable thermal stability of LCI. Further, aromatic residues (three Trp, three Phe, and four Tyr residues) hold a large weight in the protein sequence. These aromatic rings are well-defined in the three-dimensional structure (Figure 3c). There would be abundant aromatic-related interactions that help stabilize the three-dimensional structure, and such aromatic-aromatic interactions have been observed in the high-quality structure of human AMP LL-37. Figure 5 exhibits some of these interactions in the structure. We identified three amino-aromatic interactions between Lys3 and Trp23, Lys28 and Trp37, and Lys28 and Trp44 using CAPTURE31 and four edge-to-face aromatic-aromatic interactions between Phe12 and Phe25, Phe25 and Tyr41, Tyr30 and Trp37, and Tyr41 and Trp44. There are six pairs of aromatic-backbone amide interactions (Ar-HN interaction) in the structure,32 including Tyr36 (Ar) — Lys34 (HN), Tyr36 (Ar) — Tyr36 (HN), Phe25 (Ar) — Ser27 (HN), Trp44 (Ar) — Arg46 (HN), Tyr30 (Ar) — Tyr36 (HN), and Trp44 (Ar) — Asn11 (HN) interactions. These Ar-HN interactions were thought to help stabilize local structure.32 In conclusion, abundant interactions related to the 10 aromatic residues would attribute greatly to the stability of the LCI structure.

From the structure together with a molecular mass of 5464 Da, we further confirmed that the recombinant LCI peptide has exactly the same amino acid sequence as the native LCI peptide. This means that the recombinant LCI peptide may have undergone modification specific to the native LCI, at least the removal of the methionine encoded by the start codon of the plasmid.

A DALI14 search gives a series of structurally similar proteins with Z scores of 2.0–3.0 and rmsd values of 2.3–2.9 Å versus LCI structure. Further insight indicates that some of these structures are much larger than LCI, such as Menkes copper-transporting ATPase (PDB entry 1AW0), glucose-inhibited division protein B (PDB entry 1JSX), and hypothetical protein Ytmb (PDB entry 2NWA). Other structures contain structurally important helices, such as aminomethyltransferase (PDB entry 1VLO) and dodecin (PDB entry 2VKF). Thus, the three-dimensional structure of AMP LCI represents a novel topology, which mainly contains of ∼40 residues, forming a tight half-β-barrel structure consisting of a four-strand β-sheet. The solution structure of LCI is the first structure of the LCI family. Further, it is also identified as a new architecture of protein structure, as indicated in the Pfam database.33

### Table 1. Structural Statistics of AMP LCI

<table>
<thead>
<tr>
<th>no. of distance constraints</th>
<th>intraresidue (dij = 0)</th>
<th>sequential (dij = 1)</th>
<th>medium-range (2 ≤ dij ≤ 4)</th>
<th>long-range (dij ≥ 5)</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1096</td>
</tr>
<tr>
<td>no. of H-bond constraints</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>no. of violations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOE violation &gt;0.3 Å</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>torsion angle violation &gt;5°</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>PROCHECK statistics (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>most favored regions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71.2</td>
</tr>
<tr>
<td>additionally allowed regions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23.0</td>
</tr>
<tr>
<td>generously allowed regions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>disallowed regions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>rmsd from mean structure (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>backbone heavy atoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all residues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.33 ± 0.05</td>
</tr>
<tr>
<td>regular secondary structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.17 ± 0.04</td>
</tr>
<tr>
<td>all heavy atoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all residues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.59 ± 0.07</td>
</tr>
<tr>
<td>regular secondary structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.39 ± 0.05</td>
</tr>
</tbody>
</table>

Figure 3. NMR structure of LCI. (A) Ribbon representation of the average conformer. (B) Backbone atom superposition of 20 LCI conformers. (C) Side chains (green) in LCI, in which all aromatic rings are colored magenta.

![Figure 4](image.png)

Figure 4. Plot of 1H—15N NOE as a function of amino acid residue number.

![Figure 5](image.png)

Figure 5. Aromatic interaction in the LCI structure. (A) Amino—aromatic interaction between Lys3 and Trp23. (B) Aromatic—aromatic interaction between Phe12 and Phe25. (C) Aromatic—HN interaction between Phe25 and Ser27.
APD database (http://aps.unmc.edu/AP/main.php), containing α-helical structure, β-structure, both α-helix and β-structure, and extended structure that is rich in specific residues such as Pro and Arg. AMPs with β-structure are further classified into subgroups according to the number and connective pattern of disulfide bridges, and these disulfide bridges are thought to be important for the stability of β-structure AMPs. In the APD database, all present β-structure AMPs have at least one disulfide bridge except microcin J25. Microcin J25 contains a cyclic structure from side chain to backbone and exhibits remarkable stability without any disulfide bond. Figure 6 shows the ribbon diagrams of the β-structure AMPs belonging to eight subgroups and LCI. LCI has neither disulfide bridges nor ringed structure, representing a novel subgroup of β-structure AMP. Without any disulfide bridge, this novel subgroup also exhibits high thermostability due to abundant aromatic-mediated interactions, together with numerous H-bonds, and maybe also due to hydrophobic interaction, as stated above.

The amino acid sequence of LCI contains cationic residues Lys3, Lys22, Lys26, Lys34, Arg46, and Lys47 and anionic residues Asp19, Asp31, Glu42, and Asp45. Unlike the distinct amphipathic surfaces of most α-helical AMPs and many β-structure AMPs, the surface of LCI is mixed with hydrophobic, negatively charged, and positively charged regions. Figure 7 exhibits the electrostatic potential surfaces of different β-structure AMPs. We find that PDB entries 1BK8, 1EWS, and 1MYN also exhibit nonamphipathic surfaces, as LCI does.

Figure 6. Ribbon diagrams of different β-structure AMPs and LCI. PDB entry 1PG1 belongs to the protegrin AMP family, PDB entry 1MA2 to the tachyplesin family, PDB entry 1ICA to the insect defensin family, PDB entry 1BNB to the β-defensin family, PDB entry 1HVZ to the θ-defensin family, PDB entry 1BK8 to the plant defensin family, PDB entry 1EWS to the α-defensin family, PDB entry 1MYN to the drosmysin family, and PDB entry 1Q71 to the microcin family. Disulfide bridges are represented by yellow bonds, and cyclic structure (PDB entry 1Q71) is also labeled in the diagram. The structures were from the RCSB PDB.

Figure 7. Electronic potential surfaces of LCI and other β-structure AMPs. Positively charged regions are colored blue, negatively charged regions red, and neutral regions gray. Generated with MOLMOL.
Functional Implications. To exert antimicrobial activity, AMPs first interact with the target membranes and then disrupt the membranes or function at intracellular targets. Three main models of AMP—membrane interaction have been presented, including the barrel-stave model, the toroidal pore model, and the carpet model.\textsuperscript{3,4,8,34} Recently, lipid clustering has been proposed as a new mechanism of action for some AMPs versus Gram-negative bacteria.\textsuperscript{37} LCI exhibits a nonamphipathic surface, similar to those of PDB entries 1BK8, 1EWS, and 1MYN (Figure 7); 1BK8 and 1MYN are antifungal AMPs from plants and insects, respectively, and 1EWS is an antibacterial AMP from animals. Except for the fact that no mechanism for 1MYN was proposed, 1BK8 is believed to interact with its target membrane receptor by electrostatic interaction,\textsuperscript{38} and 1EWS (the rabbit kidney AMP, RK-1) is proposed to cause the formation of short-lived rather than long-lived pores in the membrane of \textit{E. coli}, as rabbit neutrophil defensins do.\textsuperscript{39,40} The formation of short-lived pores is characteristic of the toroidal pore model.\textsuperscript{3,4,11} As the fungal cell membrane differs from the bacterial cell membrane, we suggested that LCI may adopt a membrane interaction mode similar to that of 1EWS, which belongs to the toroidal pore model. For example, LCI, secreted by \textit{B. subtilis},\textsuperscript{*} first recognizes and interacts with negatively charged LPS or other molecules on the cell membrane of Gram-negative bacteria by its positively charged residues, especially the C-terminal charged residue Arg.\textsuperscript{46} Lys \textsuperscript{47} motif. Then it causes a short-lived channel in the bacterial membrane because of the formation of toroidal pores, disrupts the integrity of the membrane or penetrates the membrane, and acts on an intracellular target to fulfill its antimicrobial activity. To investigate the interaction of LCI with the membrane proposed here, we recorded \textsuperscript{1}H—\textsuperscript{15}N HSQC spectra in solutions containing 20 mM sodium dodecyl sulfate (SDS) and 20 mM sodium octyl sulfate. LCI was precipitated in the both solutions, with a small amount of LCI remaining in the soluble state from which we obtained significantly weakened NMR signals (Figure S2B of the Supporting Information). Some HSQC peaks, including Lys\textsuperscript{1}, Val\textsuperscript{3}, Asp\textsuperscript{55}, and Arg\textsuperscript{46}, exhibited obvious shifts. However, it is uncertain whether these shifts were caused by interaction of LCI with the membrane-mimetic SDS micelle, by interaction of LCI with a single SDS molecule, or by other means. We also tried the DOPC micelle, and it also precipitated LCI. Thus, further investigation is needed for the hypothesized LCI activity model that we have mentioned above.

In conclusion, our study provides the solution structure of AMP LCI from \textit{B. subtilis}. This is the first structure of the LCI protein family. The unique structural character shows that it has a novel topology. Further, without any disulide bridge or circular structure, it is also a novel subgroup of \(\beta\)-structure AMP with considerable thermal stability. The structure of AMP LCI gives new insight into the diversity of AMP structures.

ASSOCIATED CONTENT

Supporting Information. Assignments of NOESY cross-peaks (Figure S1) and \textsuperscript{1}H—\textsuperscript{15}N HSQC spectra of LCI with membrane-mimetic detergents (Figure S2). This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

The three-dimensional structure of \textit{B. subtilis} LCI has been deposited in the RCSB Protein Data Bank as entry 2B9K.

AUTHOR INFORMATION

Corresponding Author

*G.L.: The National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing, China 100871; phone, 86-731-82355302; fax, 86-731-4497661; e-mail, gylu37@yahoo.com. Co-corresponding author.

B.X.: College of Chemistry and Molecular Engineering, Peking University, Beijing, China 100871; phone, 86-10-62758127; fax, 86-10-62758307; e-mail, binxiao@pku.edu.cn.

Funding Sources

This work was supported by the National Natural Science Foundation of China (Grants 39670160 and 30125009) and the Open Fund of the National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences.

ABBREVIATIONS

AMP, antimicrobial peptide; CD, circular dichroism; FPLC, fast protein liquid chromatography; DQF-COSY, double-quantum-filtered correlation spectroscopy; TQF-COSY, triple-quantum-filtered correlation spectroscopy; E-COSY, exclusive correlation spectroscopy; TOCSY, total correlation spectroscopy; NOESY, nuclear Overhauser effect spectroscopy; rmsd, root-mean-square deviation; H-bond, hydrogen bond; LPS, lipopolysaccharide. The amino acids and nomenclature of the peptide structure are in accordance with the recommendations of the IUPAC-IUB.

REFERENCES


