Computational approach and functional analysis of the *Pectobacterium carotovorum* subsp. *carotovorum* low-molecular weight bacteriocin Carocin S2

Jyun-Wei Wang^{1#}, Ruchi Briam James S. Lagitnay^{2#}, Reymund C. Derilo², Jian-Li Wu², Duen-Yau Chuang^{2*}

¹Department of Gastroenterology, Chang Bing Show Chwan Memorial Hospital 6 Lukon Road, Lukong Town, Changhua 505, Taiwan

²Department of Chemistry, National Chung-Hsing University, Taichung City, Taiwan

J.W.W and R.B.J.S.L. contributed equally to this work.

*Corresponding author: Duen-Yau Chuang

Address: 250, Kuokuang Rd., Taichung 402, Taiwan Fax: +886-4-22862547; TEL: +886-4-22853448
E-mail: chuang@dragon.nchu.edu.tw

Running title: Computational and functional analysis of Carocin S2

Keywords: antibacterial, bacteriocin, Carocin S2K, Carocin S2I, computational approach, functional analysis, immunity protein, killer protein, *Pectobacterium carotovorum* subsp. *carotovorum*,

Abstract

Pectobacterium carotovorum subsp. *Carotovorum* 3F3 is a gram-negative phyto-parasitic enterobacterium. This strain is a producer of Carocin S2 bacteriocin, which comprises of two proteins of different sizes. Carocin S2K (killer protein) which is responsible for antibiotic resistance and Carocin S2I (immunity protein) which inhibits the antibiotic activity.

The present study aimed to predict the structure and functional properties of Carocin S2. Computational approaches utilizing various bioinformatic tools predicted that Carocin S2 is a putative membrane protein having the N-terminal at the extracellular side and the central domain at the coiled-coil region. Carocin S2 was predicted to have three domains, the translocation domains, receptor binding domain and the killer domain. Moreover, the killer domain was calculated to have the catalytic cleft. The in-vivo assays confirmed that for Carocin S2K, bound immunity protein was not a pre-requisite for cell attachment or translocation. The site-directed mutagenesis experiment led us to hypothesized the hydrolysis mechanism of Carocin S2.

The predicted structure of Carocin S2K provided a system of understanding on the biochemical and structural function which led to the mechanism of Carocin S2. It revealed that the role of immunity protein to Carocin S2 is not a pre-requisite for translocation pathway. Furthermore, this research led to hypothesized a hydrolytic mechanism of Carocin S2 to target the tRNA.

Introduction

Bacteriocins are antimicrobial peptides produced by a bacterium that's capable of inhibiting the growth of another microorganism. they're classified as proteinaceous toxins, bactericidal, and are produced by both Gram-positive and Gram-negative bacteria (1). they vary from small lanthipeptides produced by lactic acid bacteria to much larger multi-domain proteins of gram-negative bacteria like the colicins from *E. coli*. Despite huge differences in chemical structures and post-translational processing of bacteriocins between the tiny (un)modified lanthipeptides produced by lactic acid bacteria and therefore the much larger multi-domain polypeptides produced by gram-negative bacteria like the colicins, all bacteriocins are ribosomally synthesized proteinaceous toxins that share a standard biosynthetic pathway. Bacteriocins are generally narrow-spectrum antibacterial agents with biological activity against closely related species and their genes are localized on transposable elements, plasmids, or on the chromosome of the producer's genome (2). Bacteriocin production generally occurs following a stress signal, like DNA damage, inducing expression, and release of bacteriocin from the autolyzed cell (3).

Pectobacterium carotovorum subsp. *carotovorum* (Pcc) is a phytopathogenic enterobacterium accountable for the soft-rot disease of the many plant species. Some *Pectobacterium carotovorum* species produce one or more bacteriocin which boosts their competitiveness with other related bacterial species (4).

Previous studies revealed that Pcc produces Carocin S1, S2, and S3, each of which possessed different characteristics (5,6). They contain two proteins, one liable for the antimicrobial activity (the killing protein) and therefore the other for immunity (the immunity protein). The killing proteins are organized in functional domains with receptor binding, translocation, and DNase or RNase activity (7).

Proteins is easily understood in terms of its structure. Its three-dimensional structure is closely associated with its biological function. Proteins that perform similar functions tend to indicate a big degree of structural homology. amino acid sequence analysis provides important insight into the structure of proteins, which successively greatly facilitates the understanding of its biochemical and cellular function (8).

Several Bioinformatics tools are developed to predict the structure and functional properties of biomolecules (9). It provides a various array of computational tools for functional annotation and

characterization of the expression of genome products. The computational genomics and proteomics analysis combined with predictive functional analysis represents another way for the rapid identification of latest putative bacteriocins similarly as new potential antimicrobial drugs (10).

This work aims to predict the structure and functional properties of the Carocin S2 from Pcc to the best possible accuracy via a *in-vivo* study. Understanding these characteristics would enable us to deduce the hydrolytic mechanism, resulting in an improved understanding of the bacteriocin regulation mechanism.

Results

Computational analysis approach

An amino acid sequence was deduced from the nucleotide sequence of the *carocin S2* gene using DNASIS-Max software (Hitachi, Japan). The CaroS2K was observed to be highly conserved among the different strains (*Pectobacterium carotovorum subsp. brasiliensis* PBR1692, *Serratia proteamaculans* (strain 568), *Pseudomonas aeruginosa* (strain PA7), *Escherichia coli* (ColD-CA23), *Pseudomonas aeruginosa* (strain LESB58), *Klebsiella oxytoca*,). Even though there were several variations in the nucleotide sequences as observed in the multiple sequence alignment of CaroS2K to other bacteriocins (Figure 1) it demonstrated good conservation at the C-terminal amino acid level. Multiple sequence alignment and phylogenetic analysis of both nucleotide sequences and protein sequences (UniprotKB_F8RN7, UniprotKB_P17998, UniprotKB_Q5QGN1, UniprotKB_A6V5R2, UniprotKB_Q06584) from different strains showed that they were identical.

Several prediction programs yielded widely discrepant results on the topology analysis of the CaroS2K sequence. Using the TMpred program, the predicted topology of the CaroS2K is IN-OUT with putative transmembrane segments (from amino acid positions 55-76, 93-115, 151-172, 373-392, 457-482, 470-499) (Figure 2). The N-terminal is predicted as present in the extracellular side of the membrane while the central domain is predicted as a coiled-coil region for receptor binding. Our consensus topology also passed the positive-inside rule and charge bias test, with a charge bias of +1 towards the inside of the membrane.

We sought the presence of domains, patterns, and motifs within the CaroS2K protein sequence to realize an insight into its functions and structure. The SMART results showed the presence of all the structural domains that we earlier identified using topology prediction programs. Our pattern search within the CaroS2K protein sequence, using different programs against the PROSITE database, gave similar results showing the presence of the many putative phosphorylation sites. All of the anticipated sites were found within the exposed regions of the CaroS2K.

Secondary structure identification of protein follows the three states, namely helix (H), extended (β -sheet) (E), and coil (C). However, to get a more consistent set of secondary structure, we've utilized the reduced version of DSSP (Database of Secondary Structure in Proteins) classification that uses eight kinds of secondary structure assignment: H (α -helix), E (extended β -strand), G (310 helix), I (π -helix), B (bridge, one residue β -strand), T (β -turn), S (bend), and C (coil).

In Figure 3A, Carocin S2K secondary structures were predicted by GOR, HNN and SOPMA protein programs. The ninety amino acid sequence showed that among the three programs, SOPMA predicted the highest alpha-helix structure (77.7%) while GOR predicted 43.33%. The extended strand was predicted 15.56% by GOR, 12.22% by HNN and 5.56% by SOPMA. Lastly, a random coil predicted by HNN tends to be 42.22%, while SOPMA predicted the least.

For translation, we follow the strategy of CASP (Critical Assessment of Techniques for Protein Structure Prediction) and reduce the DSSP alphabet within the following manner: Helices (H, G, and I) within the

DSSP code are assigned the letter H within the three-letter secondary structure code; whereas strands (E) and bridges (B) within the DSSP code are translated into sheets (E). Other elements of the DSSP structure (T, S, C) are translated into a coil (C).

PSIpred program was also utilized to predict the 9 amino acid sequence of Carocin S2K (Figure 3B). Diagrams annotate the query sequence with secondary structure cartoons and confidence value at each position within the alignment. Confidence is given as a series of blue bars.

Homology based tertiary structure prediction for the CaroS2K protein was present in Figure 4A, using Colicin D as the template.

The CaroS2K killer domain adopts a crescent-shaped globular α/β fold and consists of a central four-stranded antiparallel β sheet ($\beta 1$, $\beta 2$, $\beta 3$ and $\beta 4$) sandwiched between two long helices ($\alpha 1$ and $\alpha 2$) on one side and also the C-terminal helix ($\alpha 3$) on the opposite. As per the modeling prediction, the catalytic could be formed at the kinked helix ($\alpha 1$) lies within the plane of the β sheet, interacting with strand $\beta 4$, parallel to $\beta 1^*$ -loop- $\beta 2^*$.

We also used fold recognition-based structure prediction server Phyre to model the tertiary structure of the CaroS2K protein, using Colicin D, Colicin E3, and Colicin E7 as the template (Figure 4B). We refined the expected structure by fixing side chains, fixing problematic loops, removal of amino acid clashes (bumps) and energy minimization. The refinements didn't yield any drastic change within the initial predicted structure. This was confirmed by visually inspecting the structure and verifying the backbone structure using the Ramachandran plot and computing the whole energy difference between the initial model and also the refined model.

The translocation domain formed a jelly-roll structure composed of the three β sheets flanked by two helical stretches (Figure 4B). The C-terminal element of secondary structure in the translocation domain was 30 Å long helix. At the end of this helix, there was a kink at Pro314, which led directly into the receptor-binding domain.

The receptor-binding domain consisted of 100 Å long antiparallel helical hairpins that extended from residue 315 to 455. The primary helical arm of this coiled-coil encompassed residues 315-380, with the last 3 residues of this stretch in an exceedingly 310-helical conformation. The second helical arm extended from residue 388-455. The antiparallel coiled-coil was “Ala coil” of the ROP type, during which every seventh residue was an alanine located within the core of the coiled-coil. The alanine residues occurred on alternate helices in every turn. There have been 3.5 residues per turn within the “Ala coil”, as distinct from a value of 3.6 for the canonical α helix. This was achieved by a smooth bending of the two helices around one another. There have been eight layers of alanine core residues within the receptor-binding domain. The receptor-binding domain was possibly the longest Alacoil ever observed. Such an antiparallel “Ala coil” is far tighter than the parallel leucine zipper.

The killer domain, consists of a four-stranded antiparallel β -sheet, with two helices on one side and a brief helix on the opposite side, as described above.

In the killer domain of Carocin S2, the triangle formation showed a similar arrangement, though these triangles are rather smaller than those of RNase A and T1. The bottom of the groove is rather hydrophobic (Phe760 and Trp764 from the $\beta 4$ sheet). His696 is further surrounded by two polar residues: Ser762 (strand $\beta 4$) and Tyr734 ($\beta 1^*$ -loop- $\beta 2^*$) (Figure 4C and 4D). Helix $\alpha 1$ further carries three residues (Lys691, Lys692, and Lys695) creating a positive charge cluster that could play an important role for substrate binding by neutralizing the negatively charged phosphate groups of the substrate tRNA backbone.

In vivo biological activity of CaroS2I

To confirm the activity of the predicted structure, in vivo biological activity of CaroS2I was done. The biological efficacy of overproduced CaroS2I was tested by establishing the level of immunity conferred to a sensitive strain of *Pectobacterium carotovorum* SP33. SP33 containing the expression vector pET30b was very sensitive to the bacterial activity of CaroS2I complex (Fig. 5A). This level of sensitivity was identical for SP33 without pET30b (data not shown). Constitutive expression of CaroS2I in SP33 resulted in complete resistance to CaroS2KI complex (Fig. 5B). Thus, expression of the gene for the CaroS2I protein conferred full biological immunity towards the action of its target Carocin, CaroS2K.

Isolation and Characterization of unbound killer protein (CaroS2K)

Evidence from the electrophoresis from the previous study indicated that immunity protein may interact specifically with CaroS2K. Most purified Carocin S2 complex has a band that runs at the same position as purified immunity protein. Since the purification procedure for Carocin favors the separation of molecules with widely different ionic properties, it is unlikely that the two would co-purify to this extent if they were not interacting.

In figure 6A, it can be noted that spots 1, 2 and 3 demonstrated that Carocin S2K possessed a little activity against the indicator strain SP33 while spot 4 showed no activity.

In general, the interaction between bacteriocin and its immunity protein is sufficiently strong that the separation can only be achieved under denaturing conditions. Denaturation by urea or guanidine hydrochloride was chosen as buffers to disrupt the interaction from Carocin S2KI-(His)6 complex. Moreover, spots 5 and 6 exhibited a clear zone of inhibition, which is a good indication that the unbound killer protein Carocin S2K was obtained successfully.

Furthermore, Carocin S2KI was also purified. The inhibitory activity of Carocin S2KI was shown in Figure 6B. Based on the figure, it may be observed that the expression of Carocin S2K without taq-protein at N-terminal (N2K) and Carocin S2K without taq-protein at C-terminal (C2K) has little or no activity against the indicator strain.

It has been shown that the Carocin S2K protein contains a catalytically active fragment capable of cleaving tRNA. Aside from Colicin E5, His residues play an important role within the catalytic mechanism of RNase. A minimum of one His, consistently present within the active site, is directly involved in proton abstraction of the 2'-OH ribose proton of the RNA substrate. Mutation of His696 abolishes its cleavage and toxicity activities. This His residue is found at the C-terminal end of the $\alpha 1$ helix, which forms the rim of the central groove of the protein. We next explored around His696 to find other candidate residues. Within the well-studied cases of RNase A and T1, the triangles formed by the catalytic triads (His12, Lys41, and His119 for RNaseA, and His40, Glu58, and His92 for RNase T1) display almost the identical geometric arrangement.

To clarify their functional involvement in tRNA hydrolysis, most of the residues mentioned above were turned into alanine and all of the mutants were compared to those of the wild-type bacteriocin-producing strain (Figure 7). The result revealed that Carocin molecules with mutated His696 or lysine residues forming the positive charge cluster (Lys692) are inactive. Mutant tRNAses affecting the most residues forming the central groove (Phe760, Ser762, and Trp764) or its rim (Tyr734) are not active. These structure-based mutagenesis experiments indicate that the central groove and also the positive charge cluster may form the active site of the Carocin S2.

Discussion

Computational approach analysis for CaroS2K

Bioinformatics tools use a large type of algorithms to predict the properties of proteins at different levels. The accuracy of those Bioinformatics tools has been improving; however, each tool has its own advantages and downsides. A selected algorithm has its own characteristic specificity, sensitivity, robustness, computational cost, etc. These characteristics are often tested against benchmarks of known datasets (9). With the arrival of assorted sequencing techniques, amino acid sequences for a variety of proteins are determined. However, three-dimensional structural information obtained through X-ray crystallography, nuclear magnetic resonance, and other experimental methods are available just for around 10% of those protein sequences (10).

It was reported from the previous study that the sequence of Carocin S2 was compared to other analogous proteins using BLAST and FASTA search tools. It had been revealed that carocin S2 gene contains two open reading frames; one containing the 2352 bp caroS2K gene (translated to 85-kDa CaroS2K) and the other containing 273-bp caroS2I gene (translated to 10 kDa CaroS2I). The stop codon (TGA) of caroS2K overlaps the start codon of caroS2I by 5-bp (ATGA) (5).

Our prediction results showed that the CaroS2K has three probable domains – translocation domain, receptor binding domain, and killer domain. We also predicted that the CaroS2K contains many phosphorylation sites, ATP/GTP-binding site, and cAMP- and cGMP-dependent protein kinase phosphorylation site. The 3-D structure analysis of the CaroS2K also predicted the possible catalytic cleft within the killer domain. Thus, supported this computational analysis, and former experimental data, we hypothesized the hydrolytic mechanism of the CaroS2K. Additionally, CaroS2K contains only one cysteine residue per protein molecule. The relatively low proportions of hydrophobic amino acids and also the highly restricted number of cysteine residues are common properties of a variety of other extracellular proteins, including amylases, penicillinases, and flagella of varied species. These common properties of primary structure are also an element within the process by which the extracellular proteins pass through the cell membrane and cell wall.

The role of immunity protein and unbound killer protein

Generally, bacteriocin operon consists of a killer gene, which encodes a killer protein that inhibits the growth of susceptible cell and an immunity gene, which encodes a protein which offers specific immunity against the killer protein (11,12).

Extracellular killer protein performs a biological action utilizing its receptor-binding domain which initially recognizes and connects the specific receptor on the membrane surface and further performs the mode of importation and also the translocation domain delivers the killer protein into the particular target within the susceptible cell (13).

In particular, the C terminal domain of killer protein determines the attacking mode while the killer protein gains entry into the infected cell. The bacteriocin producer is capable of specific immunity to the damage of its bacteriocin since the simultaneous production of the cognate immunity protein that ordinarily interacts with the C-terminal domain of the killer protein (14,15). It is notable that the immunity protein and the killer proteins interact at very high affinity due to charge attraction and are separated at the cell surface through the energy generated from the proton motif force (16).

This study represents the first of a series of investigations into the biochemical basis for immunity towards the RNase CaroS2K from species of *Pectobacterium*. We have described here the purification and properties of the unbound killer protein and also the protein liable for immunity to CaroS2K.

Immunity protein was probably interacting with the CaroS2K. As stated within the previous study, the presence of purified Carocin S2 complex band that co-electrophoreses with immunity protein on SDS acrylamide gels strongly suggested that the two interacted. Our results demonstrated clearly that, for CaroS2K, bound immunity protein was not a pre-requisite for cell attachment or translocation, since the free Carocin has precisely the same bacteriocidal activity (when analyzed by the zone spot test shown in Figure) as Carocin complexed to its immunity protein.

The *in vivo* activity of the overproduced CaroS2I was clearly observed by the whole protection it affords *Pectobacterium carotovorum* SP33, a strain which is generally sensitive towards the cytotoxic effects of CaroS2K. However, immunity protein incubated with CaroS2K so assayed *in vitro* for its protection of sensitive cells appears to be fully inactive. Therefore, whatever its mechanism of action, immunity protein cannot pass into Carocin-sensitive cells or perhaps prevent the attachment of the Carocin to its cell surface receptors. Otherwise, the immunity protein treated *in vitro* with CaroS2K would move *in vivo*. These properties are not expected for a molecule whose role was to shield Carocin-producing cells, but not the cells which are the targets of the Carocin.

Hydrolysis mechanism prediction of CaroS2KI

The general catalytic mechanism by which ribonucleases hydrolyzed RNA has been thoroughly studied by protein engineering and crystallographic analysis over several decades. RNase A, for example, has two active sites His residues that cooperate during the catalytic cycle. One acts as a general base and abstracts a proton from the ribose 2'-OH, thereby catalyzing the nucleophilic attack of this radical on the 3'-phosphate group, resulting in a cyclic intermediate, while the opposite serves as catalytic acid during the first cyclization step. Their catalytic roles are reversed during the next hydrolysis of the cyclic intermediate. Other ribonucleases, like barnase and colicin E3, proceed probably through the same mechanism, but with a His and a Glu as catalytic residues.

We probed the potential active site region by systematic site-directed mutagenesis of residues on the catalytic surface of CaroS2K (Figure 8). Taken together, the results from the site-directed mutagenesis experiments show a clear picture of the catalytic pocket centered around His696. The substitution of this His by Ala ends up in the loss of the catalytic activity, supporting the expected structure that His696 is the general base of the cyclization step of the reaction. The nucleophilic attack of the 2'-OH of the ribose on the 3'phosphate group creates a charged pentacovalent transition state that may be stabilized by nearby charged residues. In CaroS2K, the amine groups of Lys692 well-positioned to play such a task. Their mutation to Ala abolishes both cytotoxic and RNase activities of CaroS2K, confirming their role in tRNA substrate recognition or more directly within the catalytic activity. Mutation of Ser762, whose radical points towards His696, abolishes both cytotoxic and RNase activity, suggests a task for Ser762 in substrate binding and/or catalysis. Tyr734, Phe760, and Trp764 form a hydrophobic platform near the His696 imidazole side chain that exposes one face of its indole ring to the solvent, and will, therefore, stacks on a tRNA base. The lack of activity for the Tyr734Ala, Phe760Ala, and Trp764Ala mutant is in agreement with this proposal. Site-directed mutagenesis allowed us to map the CaroS2K active site and to spot residues critical for cytotoxicity (Lys692, His696, Tyr734, Phe760, Ser762 and Trp764).

The substrate specificity of the tRNAse activity of CaroS2K is remarkably narrow. This activity was conferred by a small domain carrying a modestly sized active site cleft, and subtle recognition mechanisms must be at work to achieve this specificity. How this can result in a highly specific cleavage will have to await the structural information by crystallography.

Experimental procedures

Bacterial strains and media. The bacterial strains, plasmids, and primers used in this study are listed in Table 1 and Table 2. Pcc strains were propagated at 28°C in a modified Luria-Bertani (LB) medium with 5g of NaCl per liter substituted for 10g. E. coli strains were grown in LB broth at 37°C with rotary agitation at 125 rpm. The antibiotic concentrations used for the selection of E. coli and Pcc strains were as follows: ampicillin, 50 µg/ml; kanamycin, 50 µg/ml; rifampicin, 50 µg/ml.

Recombinant DNA techniques. Standard procedures for restriction endonuclease digestions, agarose gel electrophoresis, purification of DNA from agarose gels, DNA ligation, and other cloning related techniques were followed (17) with the use of the various primers listed in Table 2. The BLAST program was used to find the protein-coding regions.

Plasmid construction and DNA manipulation. The immunity protein of Carocin S2 was amplified from the genomic DNA of strain F-rif-18 with primers CaroS2I_C-taq_for and CaroS2I_C-taq_rev (Table 2). The construct pET30b-S2I (pEC2I) contains 273-bp amplicon was purified from the agarose gel, digested with *NdeI* and *XhoI*, and ligated onto *NdeI* - *XhoI*-linearized pET30b (Novagen). For further investigation, the CaroS2I without (His)₆-taq protein was obtained by the method described previously with the following primers: X2I_forT, X2I_forS, X2I_revT, and X2I_revS (Table I). The pEX2I construct was then isolated from the kanamycin-resistant bacterial transformants, sequenced on both strands, and transformed into *Escherichia coli* BL21 (DE3) cells (Novagen).

PCR site-directed mutagenesis. A mutagenic primer was used in a PCR reaction together with either the universal primer or the reverse primer, whichever was appropriate. After isolation of the plasmid product with a Wizard Plus SV Minipreps (Promega), the PCR reaction was performed using the method described previously (18).

Protein expression and purification. Transformant of BL21, harboring protein expression from the plasmids pES2K or pES2I were grown in 500 mL LB medium (OD₅₉₅ 0.4). It was induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and were incubated at 25°C for 12 hours. Subsequently, the pellets were sonicated for 10 cycles for 9 seconds with 9-sec intervals. BL21/pES2KI pellet was harvested by ammonium sulfate precipitation (30%-40%) and was resuspended in buffer A (30 mM NaCl and 20 mM Tris-Cl; pH 8.0). The extract was equilibrated with buffer A and eluted with NaCl gradient (30 mM-1.4 M) was run with a fractogel column (Merck, USA). Purification was done with P-100 gel-filtration column (BioRad, USA) and the CaroS2K fraction were subjected for concentration using Amicon centrprep-50 column (Millipore, USA) and was dissolved in buffer A. BL21/pES2I pellets were precipitated by ammonium sulfate (70-100%), and were resuspended with buffer A. Similar purification process was followed for CaroS2I, however, it was concentrated using Amicon centrprep-3 column (Millipore, USA).

Antibacterial activity assay. The antibacterial activity of the bacteriocin producer strain was determined using the soft agar overlay method (19). Overnight cultures of the indicator strain Pcc SP33 were inoculated in prewarmed LB with an optical density of 0.5 at 600 nm, followed by aliquoting 100 µL into tubes with various Carocin concentrations. Sodium phosphate buffer (30 mM NaCl, 20 mM NaH₂PO₄ (pH 8.0) was used to prepare constant volumes for this inhibitory assay. Subsequently, the tubes harboring the indicator strains and Carocin were incubated with shaking at 28°C for one hour. The culture was plated on the LB plate and then incubated at 28°C for 16 hours. The production of an antibacterial substance was indicated by the presence of growth-free inhibition zones (clear zones) around the spotted area (20).

Computational approaches to analysis

Primary sequence analysis. The protein sequence was obtained by conceptual translation of the caroS2 open reading frame from the P.c.c genome. The primary sequence analysis was performed using ProtParam and ProtScale (21). ProtScale was used to predict the CaroS2 profile based on several amino

acid scales. ProtParam computed various properties like molecular weight, theoretical pI, instability index and grand average of hydropathicity.

Similarity searching. Different programs were used from NCBI like BLASTP and PSI-BLAST (22) and compared against several different databases like SWISSPROT and PDB. Multiple sequence alignment and phylogenetic analysis were done using Florence Corpet.

Topology prediction. The Topology of CaroS2 protein was predicted using TMpred (23), TMHMM (24), DAS (25), Coils, and MultiCoil. TMpred algorithm was based on the statistical analysis of TMbase, a database of naturally occurring transmembrane proteins, using a combination of several weight-matrices for scoring. TMHMM predicts transmembrane regions based on the hidden Markov model. DAS uses dense alignment surface method to predict transmembrane regions.

Secondary structure prediction. Programs GOR, HNN, SOPMA, and PSIPred in the NPS (Network Protein Sequence Analysis) consensus secondary structure server (26). This server runs the input sequence against several different secondary structure prediction tools and generates a consensus secondary structure out of them.

Domains/patterns/motif prediction. SMART (Simple Modular Architecture Research Tool) was used to identify the presence of any domains in the CaroS2 protein. Different pattern searching applications (ELM, Motif Scan, and PROSITE scan (27) were utilized to predict functionally relevant patterns in CaroS2 protein.

3-D Structure prediction and analysis. Initial attempts to predict the tertiary structure of CaroS2 were done using different approaches like homology modeling and threading. Automated homology modeling servers, Swiss-Model and different programs (CPHmodels, ESyPred3D, and 3DJIGSAW) were used for homology modeling. Predictions described in this study were done using fold recognition tools LOOPP, and a new version of 3-D PSSM (Phyre). The quality of the predicted structure was examined using an online version of the WHATIF program. Structure refinement was done using both WHATIF and Swiss-PDB Viewer. Structure visualization was done using PyMOL, RasWin and VMD. The figures were produced using a combination of the programs' PyMOL.

Data availability

The datasets used and analyzed during the current study are available from the corresponding authors on reasonable request.

Acknowledgements

Not applicable

Author contributions

JW, RBJS, RCD and JLW participated in the discovery and characterization of Carocin S2 and wrote the manuscript. DYC conceived the study, participated in its design and corrected the manuscript. All authors read and approved the final version of the manuscript.

Funding and additional information

The support of this work by grants from Chang Bing Show Chwan Memorial Hospital, Taiwan, is gratefully acknowledged.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Daw, M. A., and Falkner, F. R. (1999) Bacteriocins: Nature, function and structure. *Micron*. **27**, 467-479
2. Kim, Y. C., Tarr, A. W., and Penfold, C. N. (2014) Colicin import into *E. coli* cells: A model system for insights into the import mechanisms of bacteriocins. *BBA-Mol. Cell. Res.* **1843**, 1717-1731.
3. Kleanthous, C. (2010). Swimming against the tide: progress and challenges of colicin translocation. *Nat. Rev. Microbiol.* **8**, 843-848.
4. [preprint] Derilo, R. C., Wang, J. W., Lagitnay, R. B. J., Wu, H. P., Chen, K. I. and Chuang, D. Y. (2020) Identification and characterization of the bacteriocin Carocin S3 from the multiple bacteriocin strain of *Pectobacterium carotovorum subsp. carotovorum*. 10.21203/rs.3.rs-23597/v1
5. Chuang, D. Y., Chien, Y. C. and Wu, H. P. (2007) Cloning and expression of the *Erwinia carotovora subsp. carotovora* gene encoding low-molecular weight bacteriocin Carocin S1. *J Bacteriol.* **189**, 620-626.
6. Chan, Y. C., Wu, J. L., Wu, H. P., Tzeng, K. C. and Chuang, D. Y. (2011) Cloning, purification and functional characterization of Carocin S2, a ribonuclease bacteriocin produced by *Pectobacterium carotovorum*. *BMC Microbiology* **11**, 99.
7. Roh, E., Park, T. H., Kim, M., Lee, S., Ryu, S., Oh, C. S., Rhee, S., Kim, D., Park, B. S., Kim, D.H., Park, B. S. and Hue, S. (2010) Characterization of new bacteriocin, Carocin D, from *Pectobacterium carotovorum subsp. carotovorum* Pcc21. *AEM*. **76**, 7541-7549.
8. Toomula, N., Kumar, S. D. and Kumar, VVL. (2001) Computational methods for protein structure predictions and its application in drug design. *J. Proteomics. Bioinform.* **4**, 12.
9. Nagarajan, V. and Elasri, O. (2007) Structure and function predictions of the Msa protein in *Staphylococcus aureus*. *BMC Bioinformatics*. **8**, S5.
10. Solanki, D and George, J. (2020) Computational screening and characterization of novel bacteriocins through peptidome mining of pathogenic bacteria. *SSRN Electronic Journal* 10.2139/ssrn.3574482
11. Riley, M. A., and Weltz, J. E. (2002) Bacteriocin diversity: ecological and evolutionary perspectives. *Biochimie*. **84**, 367-364
12. Riley, M. A., and Weltz, J. E. (2002) Bacteriocins: evolution, ecology and application. *Annu. Rev. Microbiol.* **56**, 117-137
13. Walker, D., Moshabi, K., Vankemmelbeke, M., James, R., and Klenhous, C. (2007) The role of electrostatics in colicin nuclease domain translocation into bacterial cells. *J. Biol. Chem.* **282**, 31389-31397.

14. Jakes, K. S., and Zinder, N. D. (1974) Highly purified colicin E3 contains immunity protein. *Proc. Natl. Acad. Sci.* **71**, 3380-3384
15. Jakes, K., Zinder, N. D., and Boon, T. (1974) Purification and properties of colicin E3 immunity protein. *J. Biol. Chem.* **249**, 438-444.
16. Vankemmelbeke, M., Zhang, Y., Moore, G. R., Klenthou, C., Penfold, C. N., and James, R. (2009) Energy-dependent immunity protein release during tol-dependent nuclease colicin translocation. *J. Biol. Chem.* **284**, 18932-18941.
17. Sambrook, J., and Russel, D. W. (2001) *Molecular Cloning: A laboratory manual*. 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring, NY
18. Chiu, J., March, P. E., Lee, R., and Tillet, D. (2004) Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Res.*, **32**, 174-180.
19. Fyfe, J. M., and Harris, G. (1984) Revised Pyocin Typing Method for *Pseudomonas aeruginosa*, *J. Clin. Microbiol.* **20**, 47-50.
20. Morgan, S., Ross, R. P., and Hill, C. (1995). Bacteriolytic activity caused by the presence of a novel lactococcal plasmid encoding lactococcins A, B, and M. *Appl. Environ. Microbiol.* **16**, 57-67
21. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy Server. In: Walker J. M. (eds) *The Proteomics Protocol Handbook*. Springer Protocol Handbooks. Humana Press
22. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.
23. Hoffman, K., and Stoffel, W. (1993) TMbase – A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **374**, 166.
24. Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567-580.
25. Cserzo, M., Wallin, E., Simon, I., von Heijne, G., Elofsson, A. (1997) Prediction of transmembrane alpha helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng.* **10**, 673-676.
26. Deleage, G., Blanchet, C., and Geourjon, C. (1997) Protein structure prediction. Implications for the biologist. *Biochimie.* **79**, 681-686.
27. Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., Pagni, M., and Sigrist, C. J. (2006) The PROSITE database. *Nucleic Acids Res.* **34**, 227-230.

Abbreviations and nomenclature

The following abbreviations were used in this study: PCC- *Pectobacterium carotovorum* subsp. *carotovorum*; SMART- Simple Modular Architecture Research Tools; DSSP- Database of Screening Structure in Proteins; GOR- Garnier-Osguthorpe-Robson; HNN- Hierarchical Neural Network; SOPMA- Self-Optimized Prediction Method with Alignment; CASP- Critical Assessment of Techniques for Protein Structure Prediction.

Tables

Table 1. Bacteria and Plasmids used in this study

Strain or Plasmid	Relevant characteristics	Source of reference
<i>Pectobacterium carotovorum</i> subsp. <i>carotovorum</i>		
F-rif-18	Pcc, Rif ^r , wild-type	Laboratory stock
SP33	Wild type, indicator	Laboratory stock
<i>Escherichia coli</i>		
DH5α	supE44ΔlacU169(ψ80 lacZΔM15)hsdR17 recA1 endA1 gyrA96 thi-1 relA1	Laboratory stock
BL21 (DE3)	hsdS gal(λcIts857 ind1 Sam7 nin5 lacUV5-T7 gene 1)	Laboratory stock
Plasmid		
pGEM T-Easy	Amp ^r ; lacZ cloning vector	Promega
pET30b	Kan ^r ; expression vector with the C-terminal His-tag	Novagen
pET32a	Amp ^r ; expression vector with the N-terminal His-tag	Novagen
pEN2K	<i>caroS2K</i> subcloned into pET32a	This study
pEX2K	Derived from pEN2K; deleted series of Tag-element in front of expressed <i>caroS2K</i>	This study
pEC2I	<i>CaroS2I</i> subcloned into pET30b	This study
pEX2I	Derived from pEC2I; deleted series of Tag-element in front of expressed <i>caroS2I</i>	This study
pEH2K	Derived from pEC2I; adding (His) ₆ -Tag adjacent to <i>caroS2I</i>	This study

Kan^r: Kanamycin; Rif^r: Rifampicin; Amp^r: Ampicillin.

Table 2. Primers used in this study.

Primers	Sequence (5' -3')
CaroS2I_C-taq_for	CATATGATGAGTAATAAACT
CaroS2I_C-taq_rev	CTCGAGAAGAAGTTTGAA
X2I_forT	GGAAAAATTCAAACCTTCTTTGAGATCCGGCTGCT
X2I_forS	TGAGATCCGGCTGCT
X2I_revT	AAGAAGTTTGAATTTTCCAACGTGGCTTTTATTTC
X2I_revS	AACGTGGCTTTTATTTC
CarocinS2K_for2	CGGTCAGGATCCATGATTAAGTAC
CarocinS2I_rev2	GCGCCAAAGCTTCAAGAGATATCA
5IHT32a2KI_forT	GAAGGAGATATACATATGATTAAGTACCGTTTATA
5IHTGT2KI_forS	ATGATTAAGTACCGTTTATA
5IHT32a3KI_revT	ATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTC
5IHT32a4KI_revS	TTAAAGTTAAACAAAATTATTTC
C2KIH_forT	CACCACCACCACCACCTGATATCTCAAGCTTGCG

C2KIH_forS	TGATATCTCAAGCTTGCG
C2KIH_revT	GTGGTGGTGGTGGTGGTGAAGAAGTTTGAATTTTCC
C2KIH_revS	AAGAAGTTTGAATTTTCC
S2_H696A_forT	GTATCTTAAAGCTGCCAAAGATTTTGGTATTGTTGATACC
S2_H696A_forS	TTTTGGTATTGTTGATACC
S2_H696A_revT	TCTTTGGCAGCTTTAAGATACTTTTGTCCAGTTGCTTACG
S2_H696A_revS	TTTTTGTCCAGTTGCTTACG
S2_K698A_forT	AACATGCCGCAGATTTTGGTATTGTTGATACCAGAA
S2_K698A_forS	TATTGTTGATACCAGAA
S2_K698A_revT	CCAAAATCTGCGGCATGTTTAAGATACTTTTGTCCAG
S2_K698A_revS	TAAGATACTTTTGTCCAG
S2_R320A_forT	GCTGCGGAGGCAAAATTAAGTAAATCCACCAAGTGATC
S2_R320A_forS	TGAAATCCACCAAGTGATC
S2_R320A_revT	GTTAATTTTGCCTCCGCAGCGGTAATTGGATGTCTCAAC
S2_R320A_revS	GGTAATTGGATGTCTCAAC
S2_K321A_forT	GCGGAGCGAGCATTAAGTAAATCCACCAAGTGATCTC
S2_K321A_forS	AATCCACCAAGTGATCTC
S2_K321A_revT	TCAGTTAATGCTCGCTCCGCAGCGGTAATTGGATGTCTC
S2_K321A_revS	AGCGGTAATTGGATGTCTC
S2_T323A_forT	CGAAAATTAGCTGAAATCCACCAAGTGATCTCTTTTGC
S2_T323A_forS	ACCAAGTGATCTCTTTTGC
S2_T323A_revT	GGATTTTCAGCTAATTTTCGCTCCGCAGCGGTAATTGG
S2_T323A_revS	CTCCGCAGCGGTAATTGG
S2_K346A_forT	CCATTTTCAGGAGCAACTAAACTGTTGCGGTTTACCAAG
S2_K346A_forS	TGTTGCGGTTTACCAAG
S2_K346A_revT	GTTTTAGTTGCTCCTGAAATGGTTGCGGCAGAGCTTTCTT
S2_K346A_revS	TTGCGGCAGAGCTTTCTT
S2_T347A_forT	TCAGGAAAAGCTAAACTGTTGCGGTTTACCAAGAACAG
S2_T347A_forS	TGCGGTTTACCAAGAACAG
S2_T347A_revT	ACAGTTTTAGCTTTTCTGAAATGGTTGCGGCAGAGCT
S2_T347A_revS	AATGGTTGCGGCAGAGCT
S2_R550A_forT	CTTCCCATTTGCTCTGGCATTTTCTGATGAGAATGGAG
S2_R550A_forS	TTTCTGATGAGAATGGAG
S2_R550A_revT	ATGCCAGAGCAATGGGAAGATCCAGACTTCCATTTTTC
S2_R550A_revS	ATCCAGACTTCCATTTTTC
S2_T685A_forT	GATCGGTTTGTCTCGTAAGCAACTGGACAAAAAGTATCT
S2_T685A_forS	AACTGGACAAAAAGTATCT
S2_T685A_revT	GCTTACGAGCAAACCGATCTGAATCCAAGGGATCGTT
S2_T685A_revS	TGAATCCAAGGGATCGTT
S2_K687A_forT	GTTTACTCGTGCGCAACTGGACAAAAAGTATCTTAAAC
S2_K687A_forS	ACAAAAAGTATCTTAAAC
S2_K687A_revT	CCAGTTGCGCACGAGTAAACCGATCTGAATCCAAGGGA
S2_K687A_revS	CGATCTGAATCCAAGGGA
S2_Q688A_forT	ACTCGTAAGGCACTGGACAAAAAGTATCTTAAACATGC
S2_Q688A_forS	AAAGTATCTTAAACATGC
S2_Q688A_revT	TTGTCCAGTGCTTACGAGTAAACCGATCTGAATCCAAGG
S2_Q688A_revS	AAACCGATCTGAATCCAAGG
S2_K691A_forT	CAACTGGACGCAAAGTATCTTAAACATGCCAAAGATTTTGG
S2_K691A_forS	TAAACATGCCAAAGATTTTGG
S2_K691A_revT	AGATACTTTGCGTCCAGTTGCTTACGAGTAAACCGATC
S2_K691A_revS	CTTACGAGTAAACCGATC

S2_K692A_forT	CTGGACAAAGCGTATCTTAAACATGCCAAAGATTTTGGTA
S2_K692A_forS	ATGCCAAAGATTTTGGTA
S2_K692A_revT	GTTTAAGATACGCTTTGTCCAGTTGCTTACGAGTAAACCG
S2_K692A_revS	TTGCTTACGAGTAAACCG
S2_D699A_forT	CATGCCAAAGCTTTTGGTATTGTTGATACCAGAAAAAATAG
S2_D699A_forS	TTGATACCAGAAAAAATAG
S2_D699A_revT	CAATACCAAAAGCTTTGGCATGTTTAAGATACTTTTTGTC
S2_D699A_revS	TTTAAGATACTTTTTGTC
S2_R706A_forT	GTTGATACCGCAAAAAATAGTGAAACACTGACTAAATTTAG
S2_R706A_forS	AAACACTGACTAAATTTAG
S2_R706A_revT	CACATTTTTTGC GG TATCAACAATACCAAAATCTTTGG
S2_R706A_revS	AATACCAAAATCTTTGG
S2_K707A_forT	GATACCAGAGCAAATAGTGAAACACTGACTAAATTTAG
S2_K707A_forS	AACACTGACTAAATTTAG
S2_K707A_revT	TCACTATTTGCTCTGGTATCAACAATACCAAAATCTTTGG
S2_K707A_revS	AACAATACCAAAATCTTTGG
S2_S709A_forT	CAGAAAAAATGCTGAAACACTGACTAAATTTAGAGACGC
S2_S709A_forS	TGACTAAATTTAGAGACGC
S2_S709A_revT	GTGTTTCAGCATTTTTTCTGGTATCAACAATACCAA
S2_Y734A_forT	AAGGGACAGCTCTACTTGTGAAGGATTCAAAGGTTTTTC
S2_Y734A_forS	AAGGATTCAAAGGTTTTTC
S2_Y734A_revT	CACAAGTAGAGCTGTCCCTTTTTTCAAAGTTTCTTTTTTCC
S2_Y734A_revS	TTTTCAAAGTTTCTTTTTTCC
S2_S740A_forT	GTGAAGGATGCAAAGGTTTTCTTTAACCCGAAGACGAA
S2_S740A_forS	TTTAACCCGAAGACGAA
S2_S740A_revT	GAAAACCTTTGCATCCTTCACAAGTAGATATGTCCCT
S2_S740A_revS	AAGTAGATATGTCCCT
S2_D755A_forT	GTTGTCATGGCTAAGGATAATAAATTTATTTCCGGTTGG
S2_D755A_forS	AAATTTATTTCCGGTTGG
S2_D755A_revT	ATTATCCTTAGCCATGACAACAACATTATTCGTCTTCGG
S2_D755A_revS	AACATTATTCGTCTTCGG
S2_D757A_forT	CATGGATAAGGCTAATAAATTTATTTCCGGTTGGAAGTTGG
S2_D757A_forS	ATTTCCGGTTGGAAGTTGG
S2_D757A_revT	AAATTTATTAGCCTTATCCATGACAACAACATTATTCGT
S2_D757A_revS	ACAACAACATTATTCGT
S2_F760A_forT	GATAATAAAGCTATTTCCGGTTGGAAGTTGGATGTTGATTC
S2_F760A_forS	TGGAAGTTGGATGTTGATTC
S2_F760A_revT	ACCGGAAATAGCTTTATTATCCTTATCCATGACAACAAC
S2_F760A_revS	CTTATCCATGACAACAAC
S2_S762A_forT	AAATTTATTGCCGGTTGGAAGTTGGATGTTGATTCTCAG
S2_S762A_forS	TTGGATGTTGATTCTCAG
S2_S762A_revT	CTTCCAACCGGCAATAAATTTATTATCCTTATCCATGAC
S2_S762A_revS	ATTATCCTTATCCATGAC
S2_W764A_forT	ATTTCCGGTGCGAAGTTGGATGTTGATTCTCAGCAGTAT
S2_W764A_forS	GTTGATTCTCAGCAGTAT
S2_W764A_revT	ATCCAACCTTCGCACCGGAAATAAATTTATTATCCTTATCC
S2_W764A_revS	AAATTTATTATCCTTATCC
S2_D767A_forT	TGGAAGTTGGCTGTTGATTCTCAGCAGTATAAAAACTACG
S2_D767A_forS	AGCAGTATAAAAACTACG
S2_D767A_revT	GAGAATCAACAGCCAACCTTCCAACCGGAAATAAATTTATTATC
S2_D767A_revS	ACCGGAAATAAATTTATTATC

Figures and figure legends

Figure 1. The multiple sequence alignment of CaroS2K with other bacteriocins.

Figure 2. The topology prediction of Carocin S2K by TMpred program.

D

Figure 4. The CaroS2K modeling. (A) Homology modeling of CaroS2K by CPH models. (B) The threading method of CaroS2K by Phyre. (C) The catalytic clef of CaroS2K. (D) The functional residue of CaroS2K

Figure 5. Overproduction of CaroS2I confers resistance towards the action of Carocin S2. A lawn of *Pectobacterium carotovorum* SP33 (A) containing the overexpression vector pET30b (without insert), and (B) the same strain and plasmid containing the CaroS2I. Zones, indicating cell death, were clearly apparent in the graph.