African Journal of Biotechnology Vol. 10(38), pp. 7381-7388, 25 July, 2011
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.609
ISSN 1684–5315 © 2011 Academic Journals

Full Length Research Paper

De novo structural modeling and computational sequence analysis of a bacteriocin protein isolated from *Rhizobium leguminosarum bv. viciae* strain LC-31

Azeem Mehmood Butt[1], Ishaque Badshah Khan[2#], Farhan Haq[3#] and Yigang Tong[4*]

[1]National Centre of Excellence in Molecular Biology (CEMB), University of the Punjab, Lahore, Pakistan.
[2]School of Health and Medical Sciences, Orebro University, Orebro, Sweden.
[3]Quaid-i-Azam University, Islamabad, Pakistan.
[4]State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, People's Republic of China.

Accepted 6 June, 2011

Bacteriocins produced by different groups of bacteria are ribosomally synthesized peptides or proteins with antimicrobial and specific antagonistic bacterial interaction activity. *Rhizobium leguminosarum* is a Gram-negative soil bacterium which plays an important role in nitrogen fixation in leguminose plants. Bacteriocins produced by different strains of *R. leguminosarum* are known to impart antagonistic effects on other closely related strains. Recently, a bacteriocin gene was isolated from *R. leguminosarum bv. viceae* strain LC-31. Our study was aimed towards computational proteomic analysis and 3D structural modeling of this novel bacteriocin protein encoded by the earlier aforementioned gene. Different bioinformatics tools and machine learning techniques were used for protein structural classification. *De novo* protein modeling was performed by using I-TASSER server. The final model obtained was accessed by PROCHECK and DFIRE2, which confirmed that the final model is reliable. Until complete biochemical and structural data of bacteriocin protein produced by *R. leguminosarum bv. viceae* strain LC-31 are determined by experimental means, this model can serve as a valuable reference for characterizing this multifunctional protein.

Key words: Bacteriocin, rhizobium, protein modelling, nodulation, symbiosis, nitrogen fixation.

INTRODUCTION

Bacteriocins are proteinaceous toxins secreted by Gram-positive and Gram-negative bacteria. They have a narrow inhibitory spectrum against bacteria that are closely related to the producing bacterium. However, many of the bacteriocins produced by lactic acid bacteria (LAB) have inhibitory spectra spanning beyond the genus level and can potentially defend unwanted microflora (Klaenhammer, 1993; Riley, 1998; Shelburne et al., 2007). Bacteriocins were first identified almost 100 years ago as a heat labile product present in cultures of *Escherichia coli* V and were toxic to *E. coli* S. These were given the name of colicin to identify the producing species (Gratia, 1925). Since then, bacteriocins have been found in all major lineages of bacteria and, more recently, have been described as universally produced by some members of the Archaea (Riley and Wertz, 2002a, b). Bacteriocins are usually ribosomally synthesized. The genes encoding bacteriocin production and immunity are organized in epichromosomal operon clusters but some are also chromosomally encoded, such as *Lactobacillus sakei 5*, which produces two chromosomally encoded bacteriocins (Chassy et al., 2005; Nes et al., 1996; Sahl and Bierbaum, 1998). These polypeptides have attracted much attention due to their potential use as antibacterial agents for the treatment of infections, as well as preservation of food and animal feed. The bacteriocin

*Corresponding author. E-mail: tong62035@gmail.com. Tel: +86(10)6386-9835

#These authors contributed equally to this work.

family includes a diverse number of proteins in terms of size, microbial target, mode of action, release and immunity mechanisms and can be divided into two main groups: those produced by Gram-negative and those produced by Gram-positive bacteria (Gordon et al., 2006).

The symbiosis between legumes and N_2-fixing bacteria (rhizobia) is of huge agronomic benefit, allowing many crops to be grown without nitrogenous fertilizers. It is a sophisticated example of coupled development between bacteria and higher plants, culminating in the organogenesis of root nodules (Young et al., 2006).

Rhizobium leguminosarum is a Gram-negative bacterium living in symbiosis with leguminous plants in which it induces nitrogen-fixing root nodules (Smit et al., 1992). These strains have been shown to produce bacteriocins that have been characterized as small, medium or large based on their assumed sizes and diffusion characteristics. Large bacteriocins have been shown to resemble defective bacteriophages (Lotz and Mayer, 1972; Oresnik et al., 1999). Small bacteriocins were found to be chloroform soluble and heat labile and to have molecular masses of less than 2,000 daltons (van Brussel et al., 1985). Small bacteriocins were shown to be acylated homoserine lactone compounds related to quorumsensing molecules (Gray et al., 1996; Schripsema et al., 1996). Very little is known about medium bacteriocins produced by *R. leguminosarum*. The ability of soil bacteria to produce bacteriocins, defined as specific, nonself-propagating inhibitory agents causing antagonism between closely related strains, and bacteriocinogenic activity has been described in almost all rhizobial species (Triplett and Sadowsky, 1992). As bacteriocins act as pivotal substance in specific antagonistic bacterial interaction, they can be potentially used to control bacterial plant diseases by exerting their lethal effects on bacteria of the same or related groups. Thus, bacteriocins have most of the properties considered desirable for microbial control (Gray et al., 2006; Riley and Wertz, 2002a). Later on, it has been identified that rhizobial species are not only involved in symbiotic nitrogen fixation but also exploit range of mechanisms in direct or indirect manner to compete in nodulation and plants growth stimulation (Hafeez et al., 2005). Despite of bacteriocins antibacterial activity, the exact mechanism of their action is still vaguely understood. However, protein models of bacteriocin can be created for the deeper insights into its structure and function. In recent years, protein modelling became a promising tool with which we can predict structure of those proteins which are normally difficult to solve.

The aim of this study was to perform computational sequence analysis and 3D structural modelling of a bacteriocin protein produced by *R. leguminosarum bv. viciae* strain LC-31. Understanding the bacteriocin 3D structure could help us to understand how these extracellular proteins may contribute to nodulation, inhibition or suppression of other pathogenic plant bacteria and related processes that are known to be influenced by *R. leguminosarum* strains.

MATERIALS AND METHODS

Sequence data

Recently, isolation and characterization of the novel bacteriocin gene produced by *R. leguminosarum bv. viceae* strain LC-31 was performed (Naeem et al., 2009). Work performed by that group showed that the bacteriocin gene has three components; *RzcA, RzcB* and *RzcD*. While *RzcB* and *RzcD* are required for bacteriocin secretion, *RzcA* was found to actually encode the bacteriocin protein. By using recombination and cloning techniques, the nucleotide sequence of the *RzcA* fragment from *R. leguminosarum bv. viceae* strain LC-31 was determined to be 5'-TACGAAACTCTGGACGGCTCACCAATGCCGAAGCATCTCGTTG CCGA CGCATCACTTATTTATCGGCCCACCAATGCCACAT-3'.

In this study, we used this nucleotide sequence as a query for homology searching and computational modelling of the bacteriocin protein from *R. leguminosarum bv. viceae* strain LC-31.

Protein sequence and structure analysis

Nucleotide sequence translation

For the prediction of structural properties and the 3D structure of any protein, we first require its amino acid sequence. Up until now, the protein sequence of this specific bacteriocin gene has not been uploaded to any database; therefore, we used Translate (Wilkins et al., 1999) from Expasy to translate the query nucleotide sequence into its protein sequence.

Primary and secondary structures

ProtParam (Wilkins et al., 1999) was used to predict the physiochemical properties of the translated protein sequence. The parameters computed by ProtParam included the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY). Information regarding the secondary structure of proteins allows us to predict fold recognition and *ab initio* protein structures, classification of structural motifs and refinement of sequence alignments. Secondary structure predictions (helix, sheets and coils) were made by using different types of neural networks. In comparison to other prediction methods, machine learning approaches such as neural networks have a major advantage, as these methods use training sets of solved structures to identify common sequence motifs associated with particular arrangements of secondary structures. The hierarchical neural network (HNN) secondary structure prediction method used in this study was based on artificial neural networks (Combet et al., 2000). Two networks have been implemented in this program; these were the sequence to structure network and the structure to sequence network. JPred3 (Cole et al., 2008) is another secondary structure prediction server that uses a double neural networks approach. The recently updated Jnet algorithm provides a three-state (α-helix, β-strand and coil) prediction of secondary structure at an accuracy of 81.5%. Another server used for secondary structure predictions is PSIPRED(McGuffin et al., 2000). It incorporates two feed-forward

Table 1. *R. leguminosarum bv. viceae* strain LC-31 nucleotide sequence translation.

Number	Translated frame
1	5' - Y E T L D G S P **Met** P K H L V A D A S L I Y R P T N A T – 3'
2	5' - T K L W T A H Q C R S I S L P T H H L F I G P P **Met** P H – 3'
3	5' - R N S G R L T N A E A S R C R R I T Y L S A H Q C H – 3'

Table 2. Primary structure analysis of the bacteriocin protein's three translated frames.

Translated frame	Amino acid	Molecular weight (kDa)	Theoretical pI	Formula	Total number of atom
Reading frame 1	27	2.9	5.38	$C_{131}H_{206}N_{34}O_{42}S_1$	414
Reading frame 2	27	3.1	9.50	$C_{142}H_{217}N_{41}O_{34}S_2$	436
Reading frame 3	26	3.0	10.66	$C_{121}H_{202}N_{48}O_{38}S_2$	411

neural networks which perform an analysis on output obtained from Position Specific Iterated – BLAST (PSI-BLAST). Using a very stringent cross validation method to evaluate the method's performance, PSIPRED achieves an average accuracy of 80.7%.

Subcellular localization prediction

Determining subcellular localization is important for understanding protein function and is a critical step in genome annotation. PSORTb v3.0.2 (Yu et al., 2010) used for the study is the most precise bacterial localization prediction tool. It can make localization predictions for both Gram-positive and negative bacterial sequences and Archaea sequences.

3D structural modelling and assessment

The 3D structure is the final shape that a functional protein assumes. Various bonding interactions between the side chains on the amino acid residues determine the tertiary structure of the protein. These interactions include salt bridges, disulfide bonds, hydrophobic interactions and hydrogen bonds. No high resolution x-ray or NMR structure is available for the bacteriocin produced by *R. leguminosarum bv. viceae* strain LC-31. Therefore, we modelled the 3D structure using two approaches: homology modelling and *de novo* structural modelling. Homology modelling works best when the query matches an already present high resolution structure from the database with more than 60% sequence similarity. In cases where no good template is available, threading is done to predict the 3D structure of the target protein. For homology modelling, we used an academic version of MODELLER v 9.2 (Eswar et al., 2007). In the case of *de novo* structural modelling, I-TASSER (Roy et al., 2010) was used. Furthermore, the predicted 3D structures were evaluated by PROCHECK (Laskowski et al., 1996) and DFIRE2 (Yang and Zhou, 2008) and the calculation of disulfide bond formation was checked by DiANNA (Ferre and Clote, 2005) and DISULFIND (Ceroni et al., 2006). Structures visualization was performed by UCSF Chimera 1.5 (Pettersen et al., 2004).

RESULTS AND DISCUSSION

Sequence translation and homology searching

The nucleotide sequence of *R. leguminosarum bv. viceae* strain LC-31 RzcA was obtained (Naeem et al., 2009) and then subjected to nucleotide sequence translation tools for determination of the bacteriocin protein sequence. A total of six reading frames were generated. Stop codons were observed in all of the three 3'-5' reading frames (data not shown) and they were discarded. The remaining 5-3' frames which are given in Table 1, were then subjected to blastp analysis for the purpose of similarity searching, determining the level of conservation among other bacteriocin proteins and determination of possible templates for 3D structure prediction by homology modelling. The search was performed against all non-redundant GenBank CDS translations, PDB, Swiss-Prot, PIR, and PRF databases using default para-meters. A total of 100 targets were obtained. However, the overall percentage of sequence homology was not satisfactory (data not shown). This explains the level of diversity that bacteriocin proteins have among different bacterial species and strains.

Primary and secondary structure analysis

ProtParam was used to analyze different properties of the translated reading frames. Frames 1 and 2 were found to be composed of 27 amino acids, whereas frame 3 had 26 amino acids. The molecular weight for frames 1, 2 and 3 were calculated to be 2.96, 3.11 and 3 kDa, respectively. Detailed physiochemical results for translated frames are given in Table 2. The molecular weight and small protein

Table 3. Secondary structure analysis of the bacteriocin protein's three translated frames.

Tools	Frame 1	Frame 2	Frame 3
HNN	YETLDGSPMPKHLVADASLIYRPTNAT CCCCCCCCCCCC**HHH**CCCC**EEE**CCCCCC	TKLWTAHQCRSISLPTHHLFIGPPMPH CC**EEE**CCCC**EEE**CCCCC**EEE**CCCCCCC	RNSGRLTNAEASRCRRITYLSAHQCH CCCCCCCC**HHHHHHHHEEHE**CCCCCC
Jpred3	YETLDGSPMPKHLVADASLIYRPTNAT ----------**HHH**-----**EEE**------	TKLWTAHQCRSISLPTHHLFIGPPMPH ----------**EEE**----**EEE**-------	RNSGRLTNAEASRCRRITYLSAHQCH ----------**EEEE**-------
PsiPred	YETLDGSPMPKHLVADASLIYRPTNAT CCCCCCCCCCCC**EEE**CC**EEEEEE**CCCC	TKLWTAHQCRSISLPTHHLFIGPPMPH CC**EE**CC**EEEEEE**CCCCC**EEEE**CCCCCC	RNSGRLTNAEASRCRRITYLSAHQCH CCCCCCC**HHHHHHEEEEEEEE**CCCCC

H, Alpha helices; E, extended strands; C, coils.

length of bacteriocin produced by *R. leguminosarum bv. viceae* strain LC-31 suggests that it is biologically active and therefore, may possess a wide range of antimicrobial activity.

Different machine learning and neural network based approaches were used to analyze the secondary structures and predict the presence of alpha helices, coils and extended strands for each frame. Prediction results from different tools are summarized in Table 3. Overall, little variation was observed in the results from different prediction tools and servers. Combining the results from each approach, it was observed that reading frame 1 can form two types of secondary structures: alpha helices and beta sheets. Reading frame 2 was predicted to have only beta sheets, whereas reading frame 3 can also form both alpha helices and beta sheets. However, frame 3 was predicted to have more secondary structures as compared to frame 1.

Subcellular localization predictions

Subcellular localization is a key functional attribute of a protein. Since cellular functions are often localized in specific compartments, predicting the subcellular localization of unknown proteins may be used to obtain useful information about their functions and to select proteins for further study. Moreover, studying the subcellular localization of proteins is also helpful in understanding disease mechanisms and for developing novel drugs (Wang et al., 2005). All bacterial proteins are synthesized in the cytoplasm and most remain there to carry out their unique functions. Other proteins, however, contain export signals that direct them to other cellular locations. In Gram-positive bacteria, these include the cytoplasmic membrane, cell wall and extracellular space and in Gram-negative bacteria, they include the cytoplasmic membrane, the periplasm, the outer membrane and the extracellular space. In most cases, the whole protein is located in a single compartment; however, proteins can also span multiple localization sites (Gardy et al., 2006). Bacterial cell surface and secreted proteins are of interest for their potential as vaccine candidates or as diagnostic targets. It is also known that bacteriocins are proteins secreted by bacteria to kill other closely related bacterial species. We analyzed all the three (5'-3') reading frames for their localization potential by PSORTb. Based on prediction results, reading frame 1 was found to be an unknown protein whereas, reading frames 2 and 3 were predicted to be extracellular proteins.

Tertiary structure prediction, evaluation and assessment

Protein 3D structures can provide us with precise information of how proteins interact and localize in their stable conformation. Homology or comparative modelling is one of the most common protein structure prediction methods in structural genomics and proteomics. Therefore, we tried to model bacteriocin 3D structure using homology modelling. Numerous online servers and tools are available for homology modelling or comparative modelling of proteins. Despite minimal modifications, one initial step that was common in all modelling tools and servers was to find the best matching template. This was done by performing a sequence homology search by BLASTP. Templates are experimentally determined 3D structures of other proteins which share certain levels of sequence similarity with the query sequence. In the next step, template sequence and the protein sequence whose structure has to be deter-mined are aligned using ClustalW2 (Larkin et al., 2007). A well-defined alignment is very important for the prediction of a reliable 3D

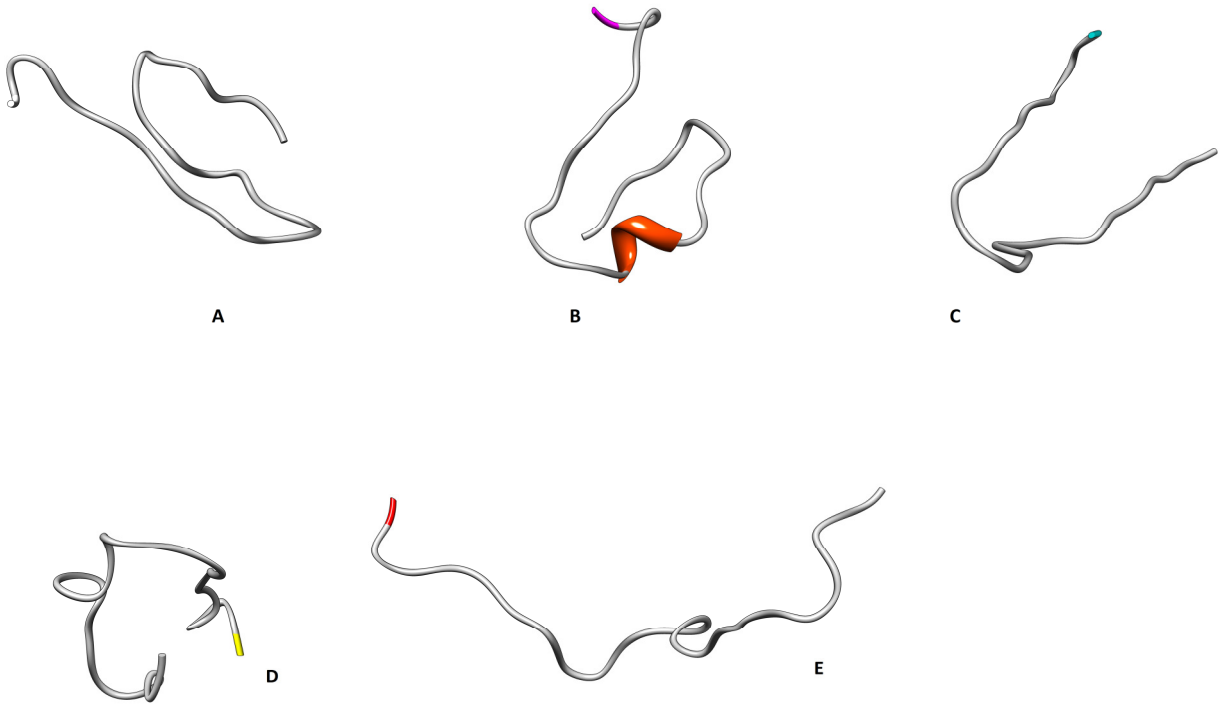

Figure 1. D*e novo* 3D models of frame 1 of the bacteriocin protein as determined by I-TASSER. Five models were generated for bacteriocin frame 1 by I-TASSER. The alpha helices and loops are shown in red and light gray, respectively.

structure. Swissmodel and Geno3D are two different servers that were used to model the 3D structure of bacteriocin. However, neither of these servers was able to model the structure for any of the three reading frames, because of the absence of a suitable template. We were also unable to model the 3D structure by MODELLER due to the absence of any suitable template. These findings are in parallel to the earlier mentioned blast homology search results where the query does not share more than 30% identity with any other protein in the protein databases at the NCBI, PDB and Uniprot. Due to template dependent limitations of homology modelling, another computational biology approach, known as *de novo* protein structure prediction, was undertaken. *Ab initio* or *de novo* protein modelling works on the principle that all the information for a protein structure lies in its amino acid sequence. This method builds a 3D structure based on physical principles rather than on previously solved structures. Several online servers, grid services and offline standalone software applications have been developed for *de novo* protein modelling. Amongst them, I-TASSER is one of the most widely used online servers for protein structure and func-tion predictions. It works by using a combination of *ab initio* folding and threading methods. In this study, I-TASSER was used for the prediction of the bacteriocin 3D structure. Each reading frame was separately mo-delled in I-TASSER and five models were generated for each frame. Models generated for frames 1, 2 and 3 are shown in Figures 1, 2 and 3 respectively.

Once the models were generated, they were subjected to structural assessment and validation using PROCHECK, DFIRE2 and the C-Score values from the I-TASSER. Ramachandran plots were generated by PROCHECK. Additionally, the stereochemical qualities were assessed for each predicted model. The assess-ment results from PROCHECK are summarized in Table 4.

A total of 15 structural models from three reading frames were analyzed in DFIRE2 and protein conformation free energy scores were calculated. Free energy calculations made by DFIRE2 are provided in Table 4.

The final assessment and validation conclusion of protein structures were made on the basis of combined results from PROCHECK, DFIRE2 and I-TASSER's C-Score. In the case of frame 1, models 2 and 3 contained no residues in the disallowed region, one residue in the generously allowed region and more than 57% of residues were in the most favoured regions. By using DFIRE2, predicted energy values for models 2 and 3 were found to be -23.55 and -22.48, respectively, which are comparable to the energy values of models 1, 4 and 5. For frame 2, poor ramachandran plots were obtained. In the models generated for reading frame 3, models 1 and 3 had no residues in the disallowed region and one

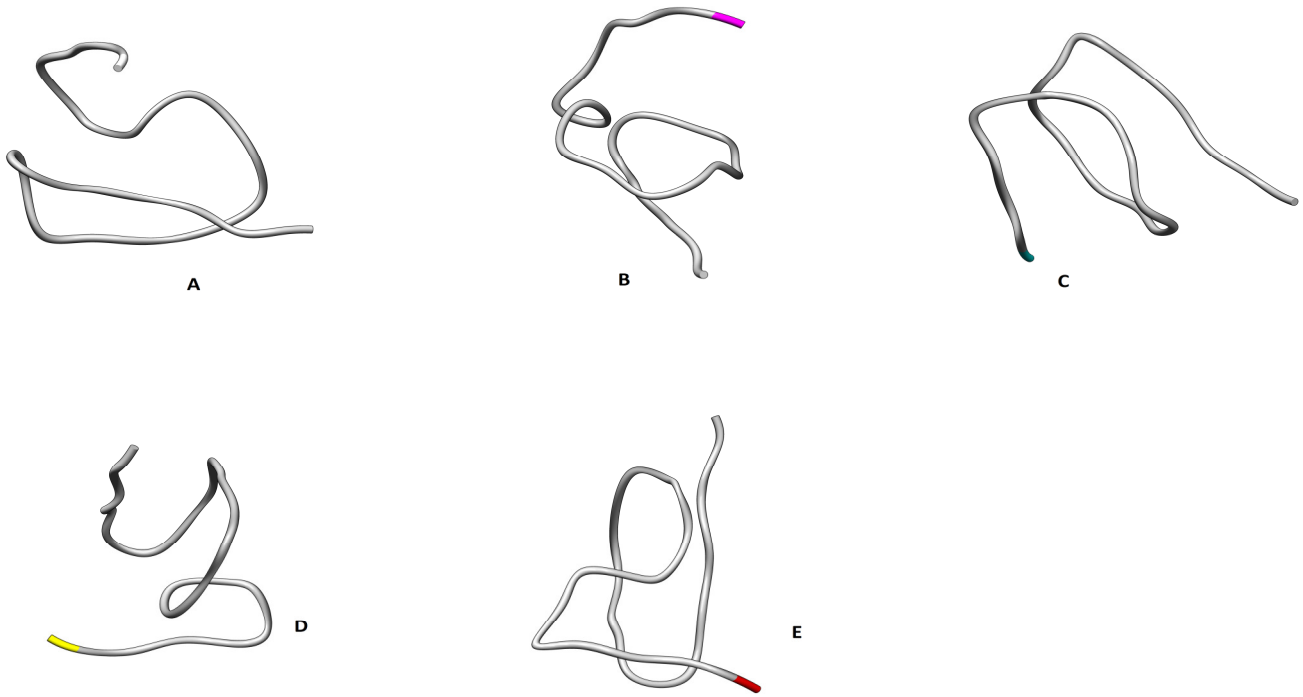

Figure 2. D*e novo* 3D models of frame 2 of the bacteriocin protein as determined by I-TASSER. Five models were generated for bacteriocin frame 2 by I-TASSER. The alpha helices and loops are shown in red and light gray, respectively.

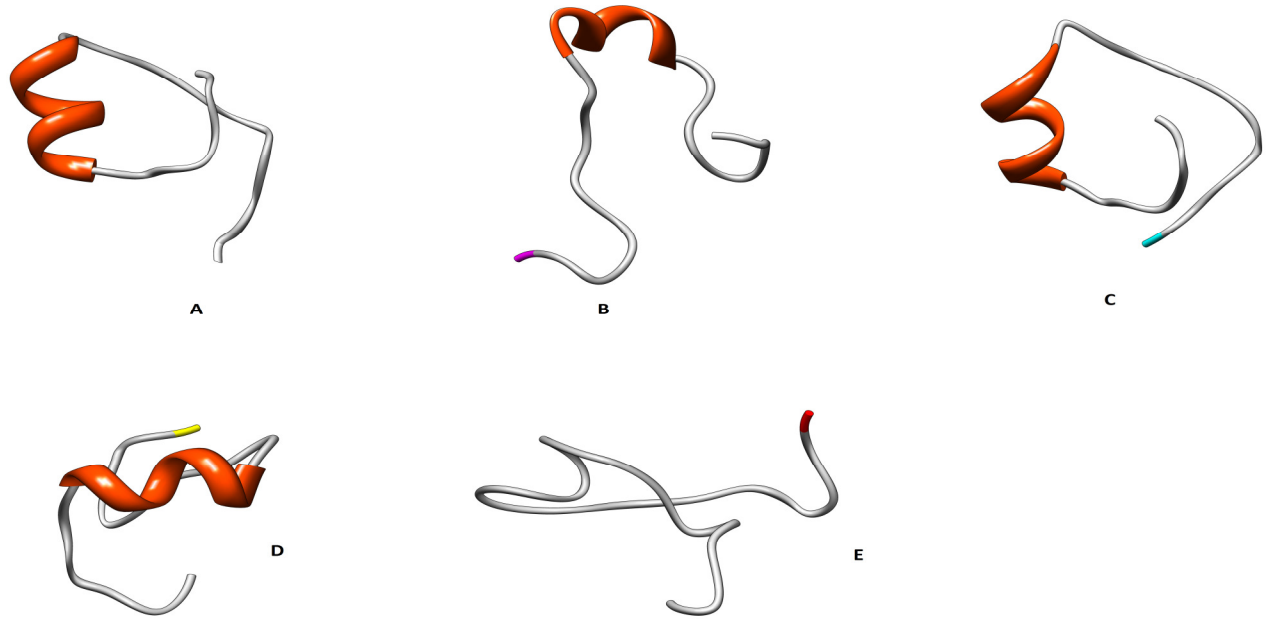

Figure 3. D*e novo* 3D models of frame 3 of the bacteriocin protein as determined by I-TASSER. Five models were generated for bacteriocin frame 3 by I-TASSER. The alpha helices and loops are shown in red and light gray, respectively.

residue in the generously allowed region. However, model 5 also had no residues in the disallowed region and only one residue in the generously allowed region. The energy value for model 1 and 3 were calculated to be -30.64 and -27.64, respectively which were the lowest among the five models. In addition, C-Score value for model 1(-1.86) and 3(-1.91) were found to be the highest among the five models.

Table 4. Evaluation and assessment of the predicted 3D models of frames 1, 2 and 3.

Frame		Model 1	Model 2	Model 3	Model 4	Model 5
1	I – TASSAR (C-Score)	-2.02	-2.07	-3.46	-2.47	-5
	DFIRE2	-17.97	-23.55	-22.48	-16.30	-16.10
	PROCHECK					
	Residues in most favored region	57.1%	57.1%	57.1%	52.4%	66.7%
	Residues in additional allowed regions	33.3%	38.1%	38.1%	38.1%	19%
	Residues in generously allowed regions	4.8%	4.8%	4.8%	4.8%	14.3%
	Residues in disallowed regions	4.8%	0.0%	0.0%	4.8%	0.0%
	Total	100%	100%	100%	100%	100%
2	I – TASSAR (C-Score)	-1.28	-2.84	-3.33	-4.75	-2.20
	DFIRE2	-25.06	-27.12	-26.99	-24.05	-25.45
	PROCHECK					
	Residues in most favored region	50.0%	55.0%	70.0%	50.0%	40.0%
	Residues in additional allowed regions	45.0%	30.0%	15.0%	45.0%	50.0%
	Residues in generously allowed regions	5.0%	5.0%	5.0%	5.0%	10.0%
	Residues in disallowed regions	0.0%	10.0%	10.0%	0.0%	0.0%
	Total	100%	100%	100%	100%	100%
3	I – TASSAR (C-Score)	-1.86	-2.61	-1.91	-3.04	-2.86
	DFIRE2	-30.64	-22.92	-27.64	-25.54	-22.23
	PROCHECK					
	Residues in most favored region	69.6%	69.6%	69.6%	47.8%	73.9%
	Residues in additional allowed regions	26.1%	13.0%	30.4%	26.1%	21.7%
	Residues in generously allowed regions	4.3%	8.7%	0.0%	13.0%	4.3%
	Residues in disallowed regions	0.0%	8.7%	0.0%	13.0%	0.0%
	Total	100%	100%	100%	100%	100%

The presence of two or more than two cysteine residues results in the formation of disulfide bonds which are known to play an important role in bacteriocin protein stabilization. Two cysteine residues were found in translated frame 3, one cysteine residue in frame 2 and no cysteine residue in frame 1. Therefore, reading frame 3 was inspected for potential disulfide bonding. Two servers were used for the prediction of disulfide bonding state and connectivity prediction: DiANNA (Ferre and Clote, 2005) and DISULFIND (Ceroni et al., 2006). DiANNA employs a novel diresidue neural network based approach. In the initial stage, PSIPRED is run to predict the protein's secondary structure. PSIBLAST is then run against the non-redundant SwissProt database to obtain a multiple alignment of the input sequence. Next, the cysteine oxidation state is predicted and then each pair of cysteines in the protein sequence is assigned a likelihood of forming a disulfide bond. Finally, Rothberg's implementation of Gabow's maximum weighted matching algorithm is applied to diresidue neural network scores in order to produce the final connectivity prediction. On the other hand, DISULFIND employs a support vector machines (SVM) binary classifier to predict the bonding state of each cysteine, followed by a refinement stage that classifies all the cysteines in a chain in a collective fashion. Almost similar results were obtained from both disulfide bonding and connectivity prediction servers. Two cysteine residues were found at positions 14 and 25 in the reading frame 3, separated by a distance of 11 amino acids. The presence of disulfide bond forming cysteine residues is a characteristic feature of bacteriocins. It can also be used as a basis for sub-grouping. It has been observed that the antibacterial efficiency of a bacteriocin increases with an increase in the number of disulfide bonds. For example, pediocin AcH with two disulfide bridges has a wider range of antimicrobial activity when compared with lactococcin B which has a single disulfide bridge (Ralph et al., 1995). Also, disulfide bonds are known to be important for the stability of the bacteriocin protein (Olivera et al., 2003; Rober, 2005). In agreement with the earlier mentioned structure assessment analysis, frame 3 contained two cysteine residues with a highly predicted potential for bond formation and may be a potential bacteriocin protein sequence.

Based upon the current knowledge regarding the activity and functionality of bacteriocins and compu-

tational assessment results, the only models selected as representatives of bacteriocin 3D structure met the following criteria: (1) predicted to be an extracellular protein with the maximum number of secondary structures in comparison to other predicted models; (2) presence of cysteines residues for disulfide bonding; (3) Ramachandran plots showing the maximum number of residues in allowed and the least number of residues in disallowed regions; (4) minimum free energy score of protein conformation and highest value from C-Score.

Therefore, we concluded that reading frame 1 is not likely to be the protein of the given bacteriocin, as it was not considered to be an extracellular protein by PSORTb, had less secondary structure predictions than reading frame 3 and contained no cysteine residues. Reading frame 2 was least likely to be the protein of the given bacteriocin, as it was predicted to have a lower level secondary structure, which is required for bacteriocin function.

We proposed that reading frame 3 was the desired protein sequence of the bacteriocin in question and models 1 and 3 were considered as the most probable 3D structure of the given bacteriocin. PSORTb predicted frame 3 to be an extracellular protein, with the maximum number of secondary structures compared with frames 1 and 2. The presence of cysteine residues and disulfide bonding was confirmed by DiANNA and DISULPHID. PROCHECK, DFIRE2 and C-Score assessments, provided the best tertiary structures for frame 3. Although, the bonding distance between the cysteine residues was found to be more than the allowed distance (data not shown), further structure refinements of models 1 and 3 may result in the decreased distance between two cysteine residues.

With the assistance of a well-defined structure of bacteriocin, one can predict its functional and binding sites, which can help in understanding the multi-functional role of bacteriocin for competition in nodulation. This knowledge can be further used in drug design to enhance or suppress the production of bacteriocin as required.

Acknowledgments

This work was supported by the Hi-Tech Research and Development (863) Program of China (No. 2009AA02Z111) and the National Natural Science Foundation of China (No. 30872223).

REFERENCES

Ceroni A, Passerini A, Vullo A, Frasconi P (2006). DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. Nucleic Acids Res. 34: W177-181.

Chassy B, Hlywka JJ, Kleter GA, Kok EJ, Kuiper HA, McGloughlin M, Munro IC, Phipps RH, Reid JE, Stein J, Zabik J (2005). Nutritional and safety assessments of foods and feeds nutritionally improved through biotechnology. Food Nutr. Bull. 26: 436-442.

Cole C, Barber JD, Barton GJ (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Res. 36: W197-201.

Combet C, Blanchet C, Geourjon C, Deleage G (2000). NPS@: network protein sequence analysis. Trends Biochem. Sci. 25: 147-150.

Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A (2007). Comparative protein structure modeling using MODELLER. Curr Protoc. Protein Sci. Chapter 2: Unit 2 9.

Ferre F, Clote P (2005). DiANNA: a web server for disulfide connectivity prediction. Nucleic Acids Res. 33: W230-232.

Gardy JL, Brinkman FS (2006). Methods for predicting bacterial protein subcellular localization. Nat. Rev. Microbiol. 4: 741-751.

Gordon DM, O'Brien CL (2006). Bacteriocin diversity and the frequency of multiple bacteriocin production in Escherichia coli. Microbiology, 152: 3239-3244.

Gratia A (1925). Sur un remarquable example d'antagonisme entre deux souches de colibacille. Compt. Rend. Soc. Biol. 93: 1040-1042.

Gray EJ, Lee KD, Souleimanov AM, Di Falco MR, Zhou X, Ly A, Charles TC, Driscoll BT, Smith DL (2006). A novel bacteriocin, thuricin 17, produced by plant growth promoting rhizobacteria strain Bacillus thuringiensis NEB17: isolation and classification. J. Appl. Microbiol. 100: 545-554.

Gray KM, Pearson JP, Downie JA, Boboye BE, Greenberg EP (1996). Cell-to-cell signaling in the symbiotic nitrogen-fixing bacterium *Rhizobium leguminosarum*: autoinduction of a stationary phase and rhizosphere-expressed genes. J. Bacteriol. 178: 372-376.

Hafeez FY, Naeem FI, Naeem R, Zaidi AH, Malik KA (2005). Symbiotic effectiveness and bacteriocin production by Rhizobium leguminosarum bv. viciae isolated from agriculture soils in Faisalabad Environ. Exp. Bot. 54: 142-147.

Klaenhammer TR (1993). Genetics of bacteriocins produced by lactic acid bacteria. FEMS Microbiol Rev. 12: 39-85.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007). Clustal W and Clustal X version 2.0. Bioinformatics, 23: 2947-2948.

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR, 8:477-486.

Lotz W, Mayer F (1972). Isolation and characterization of a bacteriophage tail-like bacteriocin from a strain of Rhizobium. J. Virol. 9: 160-173.

McGuffin LJ, Bryson K, Jones DT (2000). The PSIPRED protein structure prediction server. Bioinformatics, 16: 404-405.

Naeem F, Khan SA, Kiran A, Mukhtar Z, Hafeez FY (2009). Role of bacteriocin like substances in rhizosphere ecology, In: Hafeez FY, Malik KA, Zafar Y (Eds.). Microbial technologies for sustainable agriculture: exploring the hidden potentials of microbes. NCB/NIBGE Press, Faisalabad, Pakistan, pp. 15 - 18.

Nes IF, Diep DB, Havarstein LS, Brurberg MB, Eijsink V, Holo H (1996). Biosynthesis of bacteriocins in lactic acid bacteria. Antonie Van Leeuwenhoek, 70: 113-128.

Oresnik IJ, Twelker S, Hynes MF (1999). Cloning and characterization of a Rhizobium leguminosarum gene encoding a bacteriocin with similarities to RTX toxins. Appl Environ Microbiol. 65: 2833-2840.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. 25: 1605-1612.

Riley MA (1998). Molecular mechanisms of bacteriocin evolution. Annu. Rev. Genet. 32: 255-278.

Riley MA, Wertz JE (2002a). Bacteriocin diversity: ecological and evolutionary perspectives. Biochimie, 84: 357-364.

Riley MA, Wertz JE (2002b). Bacteriocins: evolution, ecology, and application. Annu. Rev. Microbiol. 56: 117-137.

Roy A, Kucukural A, Zhang Y (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5: 725-738.

Sahl HG, Bierbaum G (1998). Lantibiotics: biosynthesis and biological

activities of uniquely modified peptides from gram-positive bacteria. Annu. Rev. Microbiol. 52: 41-79.
Schripsema J, de Rudder KE, van Vliet TB, Lankhorst PP, de Vroom E, Kijne JW, van Brussel AA (1996). Bacteriocin small of Rhizobium leguminosarum belongs to the class of N-acyl-L-homoserine lactone molecules, known as autoinducers and as quorum sensing co-transcription factors. J. Bacteriol. 178: 366-371.
Shelburne CE, An FY, Dholpe V, Ramamoorthy A, Lopatin DE, Lantz MS (2007). The spectrum of antimicrobial activity of the bacteriocin subtilosin A. J. Antimicrob. Chemother. 59: 297-300.
Smit G, Swart S, Lugtenberg BJ, Kijne JW (1992). Molecular mechanisms of attachment of Rhizobium bacteria to plant roots. Mol. Microbiol. 6: 2897-2903.
Triplett EW, Sadowsky MJ (1992). Genetics of competition for nodulation of legumes. Annu. Rev. Microbiol. 46: 399-428.
van Brussel AA, Zaat SA, Wijffelman CA, Pees E, Lugtenberg BJ (1985). Bacteriocin small of fast-growing rhizobia is chloroform soluble and is not required for effective nodulation. J. Bacteriol. 162: 1079-1082.
Wang J, Sung WK, Krishnan A, Li KB (2005). Protein subcellular localization prediction for Gram-negative bacteria using amino acid subalphabets and a combination of multiple support vector machines. BMC Bioinformatics, 6: p. 174.
Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF (1999). Protein identification and analysis tools in the ExPASy server. Methods Mol. Biol. 112: 531-552.
Yang Y, Zhou Y (2008). Specific interactions for ab initio folding of protein terminal regions with secondary structures. Proteins, 72: 793-803.
Young JP, Crossman LC, Johnston AW, Thomson NR, Ghazoui ZF, Hull KH, Wexler M, Curson AR, Todd JD, Poole PS, Mauchline TH, East AK, Quail MA, Churcher C, Arrowsmith C, Cherevach I, Chillingworth T, Clarke K, Cronin A, Davis P, Fraser A, Hance Z, Hauser H, Jagels K, Moule S, Mungall K, Norbertczak H, Rabbinowitsch E, Sanders M, Simmonds M, Whitehead S, Parkhill J (2006). The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome Biol. 7: R34.
Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics, 26: 1608-1615.