

Article

# Structural Characterization of the S-glycosylated Bacteriocin ASM1 from *Lactobacillus plantarum*

Alexander K. Goroncy <sup>\*,†</sup> , Trevor S. Loo, Adrian M. Koolaard, Mark L. Patchett and Gillian E. Norris

School of Fundamental Sciences, Massey University, Palmerston North 4442, New Zealand; T.S.Loo@massey.ac.nz (T.S.L.); A.M.Koolaard@massey.ac.nz (A.M.K.); M.L.Patchett@massey.ac.nz (M.L.P.); G.Norris@massey.ac.nz (G.E.N.)

\* Correspondence: agoroncy@uwyo.edu; Tel.: +1-307-766-2093

† Current address: Department of Chemistry, University of Wyoming, Laramie, WY 82071, USA.

Received: 12 October 2019; Accepted: 10 March 2020; Published: 22 March 2020



**Abstract:** In order to protect their environmental niche, most bacteria secrete antimicrobial substances designed to target specific bacterial strains that are often closely related to the producer strain. Bacteriocins, small, ribosomally synthesised antimicrobial peptides, comprise a class of such substances and can either inhibit (bacteriostatic) or kill (bactericidal) target cells. Glycocins are a class of bacteriocin that are post-translationally modified by one or more carbohydrate moieties that are either  $\beta$ -O-linked to either a serine or threonine and/or  $\beta$ -S-linked to a cysteine. The solution nuclear magnetic resonance structure (NMR) of the glycocin ASM1 (produced by *Lactobacillus plantarum* A-1), an orthologue of GccF, has been determined. In both structures, the disulfide bonds are essential for activity and restrict the mobility of the N-acetyl-glucosamine (GlcNAc) attached to Ser-18 (O-linked), compared to the much more flexible GlcNAc moiety on Cys-43 (S-linked). Interestingly, despite 88% sequence identity, the helical structure of ASM1 is less pronounced which appears to be consistent with the far ultra-violet circular dichroism (UV CD) spectra.

**Keywords:** NMR; circular dichroism; bacteriocin; post-translational modifications; S-linked glycosylation; O-linked glycosylation

## 1. Introduction

Most bacteria secrete ribosomally synthesized proteins/peptides, called bacteriocins, designed to enhance their access to shared resources. Bacteriocins form a functionally and structurally diverse group, and many have been structurally characterised. They have been divided in categories a number of times in an attempt to group them into families with similar characteristics. In 2006, Heng and Tagg [1] proposed a four-class scheme in which class I bacteriocins are modified peptides (lantibiotics), class II are unmodified peptides, class III are large proteins, and class IV are cyclic peptides. More recently, in 2016, the classification of lactic acid bacteriocins was modified to include three main groups: modified, unmodified and heat labile [2]. Ribosomally-synthesised post-translationally modified peptides (RiPPs) [3] include a group of bacteriocins known as the glycocins, so named because of the carbohydrate moieties that are covalently linked to specific serine, threonine and/or cysteine residues in the polypeptide chain [4].

The NMR solution structures of two glycocins, glycosin F (GccF) and sublancin 168, have been reported and show a remarkable similarity (a helix-loop-helix architecture stabilised by a pair of nested disulfide bonds), despite significant sequence differences [5,6]. ASM1, a 43 amino acid bacteriocin secreted by *Lactobacillus plantarum* A-1, is a homologue of GccF (with only 5 amino acid differences, all of which are in the C-terminal tail [7,8], see Figure 1). Apart from Ser34 (GccF His34), the remaining

four changes are conservative and appear to have only a small effect on inhibitory activity [9]. Like GccF, it has been shown that both sugars contribute to the antimicrobial activity, and that this activity is bacteriostatic rather than bactericidal. In common with most bacteriocins, ASM1 appears to have a relatively narrow phylogenetic spectrum of toxicity restricted to other *Lactobacillus* species and some *Enterococcus* species. As for GccF, its activity can be attenuated, by the addition of free GlcNAc [9]. It is therefore expected that the two proteins will have very similar structures. Circular dichroism spectra of both GccF [7] and synthetic sublancin [10] suggested substantial  $\alpha$ -helical content, and the NMR solution structures of the natural products showed that both have a helix-loop-helix architecture that is stabilised by nested disulfide bonds [5,6].

KPAWCWYTLAMCGAGYDSGTCDYMYSHCFGIKHHSSGSSSYHC GccF  
 KPAWCWYTLAMCGAGYDSGTCDYMYSHCFGV**KHSSGGGS**SYHC ASM1

**Figure 1.** Sequences of bacteriocins glycosin F (GccF) and ASM1 showing the 5 sequence changes in bold.

Variations of this helix-loop-helix architecture, known as the Cs  $\alpha/\alpha$  fold, are also found in several unrelated peptides, many of them acting as toxins and requiring both disulfide bonds to be intact for activity [11–14].

Here we present the NMR solution structure of ASM1 from *Lactobacillus plantarum* A-1.

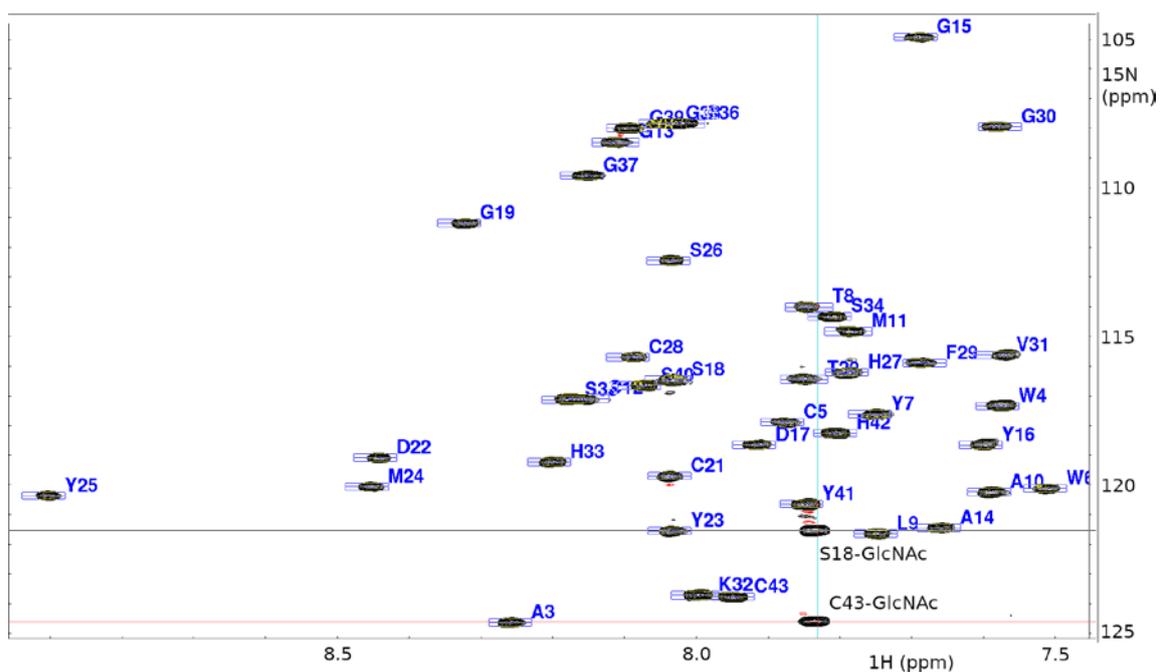
## 2. Results

### 2.1. NMR Assignment and Post-Translational Modifications of ASM1

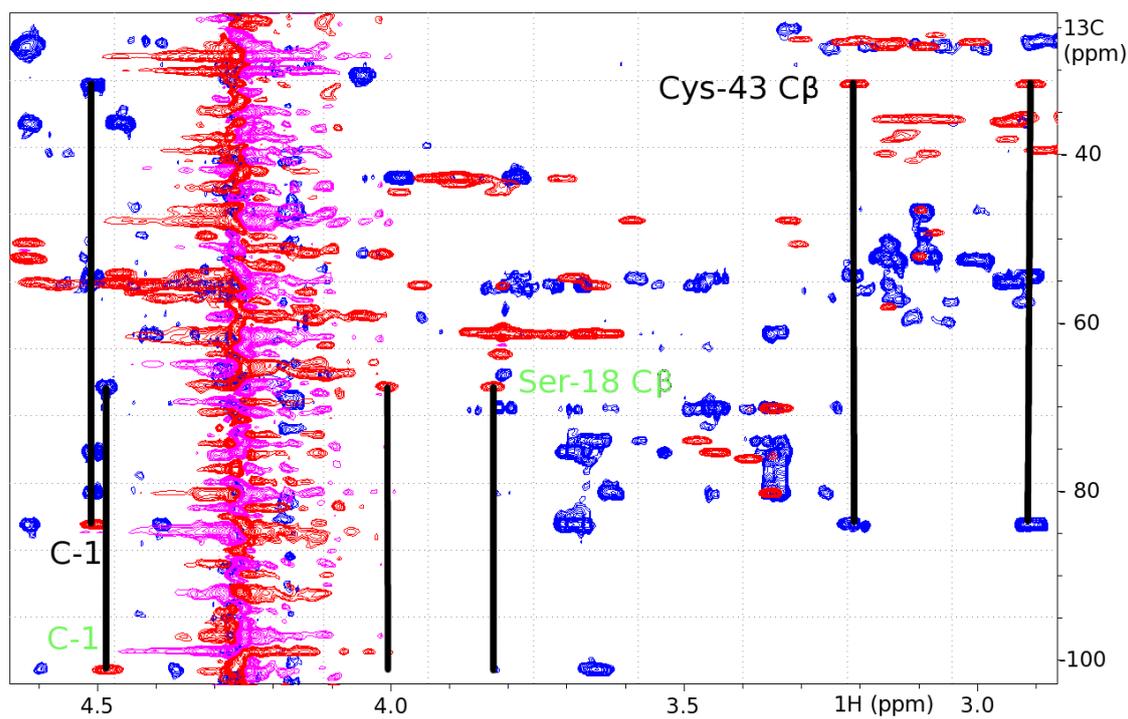
The bacteriocin ASM1 was monomeric and homogenous at the used solvent conditions of 40% acetonitrile, 0.2% acetic acid and 59.8% water. We empirically tested combinations of solvent/acid using dynamic light scattering (Malvern), analytical centrifugation and mass spectrometry [9]. No nuclear Overhauser effect (NOE) peaks between monomers were seen. Strong characteristic NOE cross peaks indicative of aggregation were not observed either. The  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum contained two extra peaks originating from neither the amino acid backbone nor from sidechain nitrogen atoms (Figure 2). The  $^1\text{H}$ - $^{13}\text{C}$  HSQC,  $^1\text{H}$ - $^{13}\text{C}$  HMBC and  $^1\text{H}$ - $^{13}\text{C}$  HSQC-TOCSY spectra established that they belonged to two separate GlcNAc entities (Figures 3–5) and that, as in GccF, one GlcNAc was linked to the  $\gamma\text{O}$  atom of Ser-18 (O-linked) and the other was linked to the  $\gamma\text{S}$  atom of Cys-43 (S-linked). This was confirmed by mass spectrometry [9]. This was also supported by both the chemical data (Table 1) and NOE cross peaks.

**Table 1.** Chemical shift data of the post-translational modifications of ASM1.

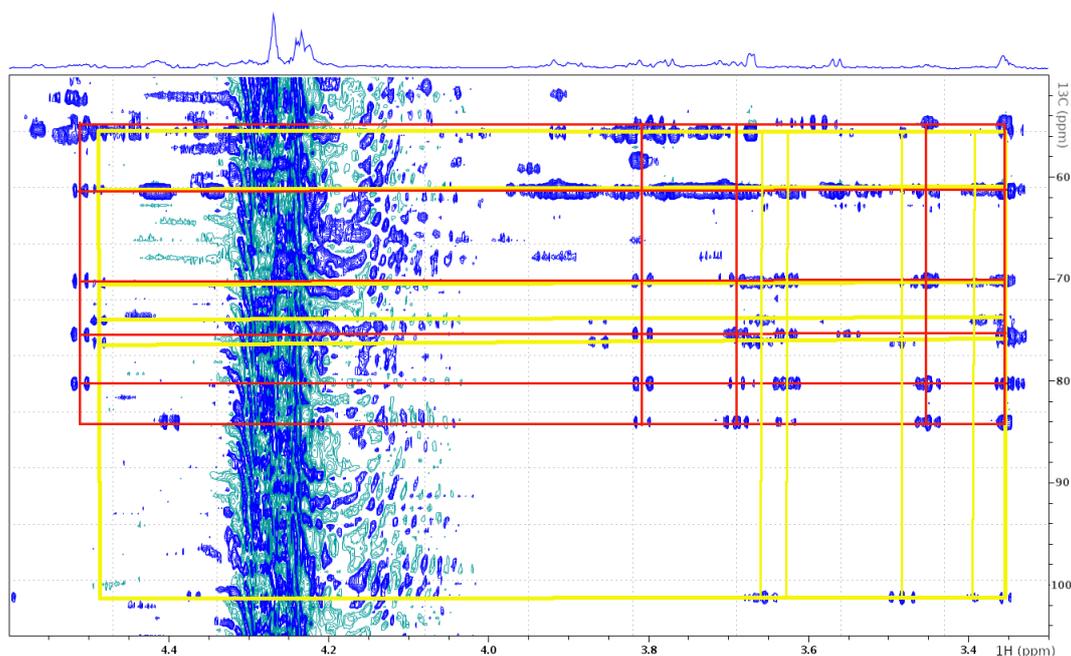
Constituent	GlcNAc (Attached to Ser-18)		GlcNAc (Attached to Cys-43)	
	$\delta(^1\text{H})/\text{ppm}$	$\delta(^{13}\text{C})/\text{ppm}$	$\delta(^1\text{H})/\text{ppm}$	$\delta(^{13}\text{C})/\text{ppm}$
1	4.479	100.952	4.508	83.723
2	3.653	55.404	3.691	54.452
3	3.485	73.788	3.447	75.184
4	3.345	69.747	3.350	69.939
5	3.388	75.898	3.352	80.108
6	3.627	61.050	3.808	61.226
6'	3.627		3.808	
Me	1.994	22.413	1.900	22.134



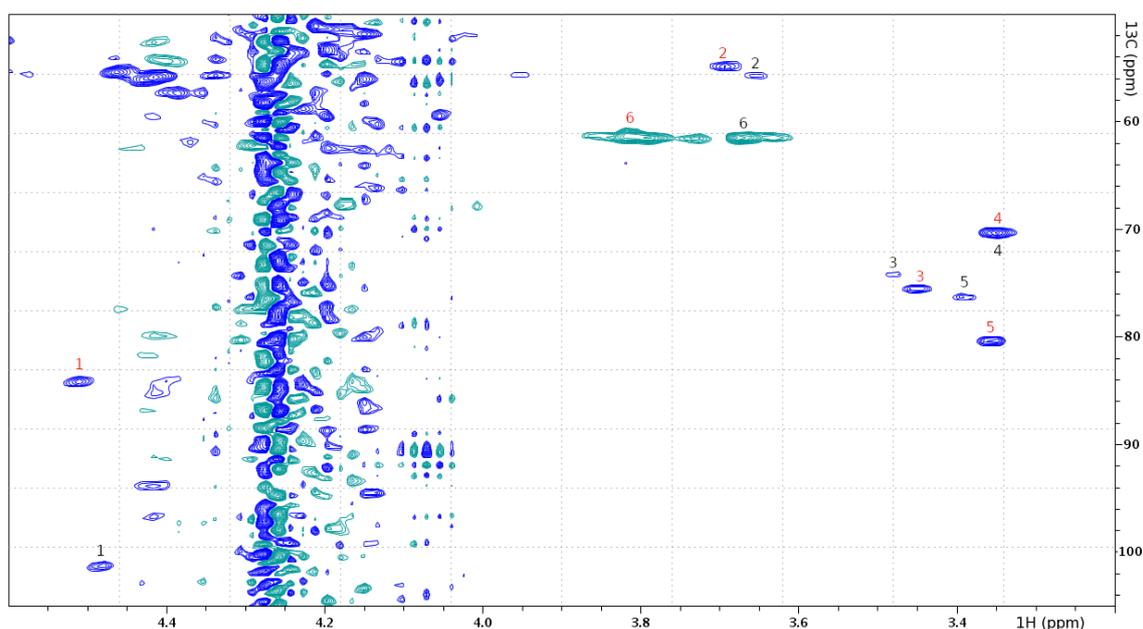
**Figure 2.**  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear single quantum coherence (HSQC) spectrum of ASM1 shows presence of  $^{15}\text{N}$ -containing post-translational modification (in cross hairs): S18-GlcNAc and C43-GlcNAc.



**Figure 3.** Overlay of  $^1\text{H}$ - $^{13}\text{C}$  HSQC (red) and  $^1\text{H}$ - $^{13}\text{C}$  HMBC (blue). Connectivities between the C-1 atom of a GlcNAc (83.723 ppm, 4.508 ppm) to the  $\beta$ -carbon of Cys-43 (31.471 ppm, 3.214 and 2.906 ppm) and of the C-1 atom of a second GlcNAc residue (100.952 ppm, 4.479 ppm) to the  $\beta$ -carbon of Ser-18 (67.330 ppm, 4.006 and 3.819 ppm) are indicated by solid lines.



**Figure 4.**  $^1\text{H}$ - $^{13}\text{C}$  HSQC-TOCSY spectrum (no decoupling) shows internal connectivities of the GlcNAc attached to Ser-18 (yellow) and the GlcNAc attached to Cys-43 (red).



**Figure 5.**  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum shows assignments of the GlcNAc attached to Ser-18 (black) and the GlcNAc attached to Cys-43 (red), labelling according to Table 1.

The rest of the NMR assignment procedure followed established methodology, overlaying two-dimensional spectra similar to Figure 3. There were 33 internal (i.e., within the same moiety, in this case within the same sugar) NOE peaks for the O-linked GlcNAc, 17 internal NOE peaks for the S-linked GlcNAc, and 21 and 17 NOE peaks linking the GlcNAc moieties to the polypeptide chain, respectively. The external (i.e., not within the same moiety, between different moieties, in this case between the sugar and the rest) NOE peaks were also consistent with a  $\beta$ -linkage between GlcNAc and Ser-18 and between another GlcNAc and Cys-43. For the GlcNAc attached to Ser-18,

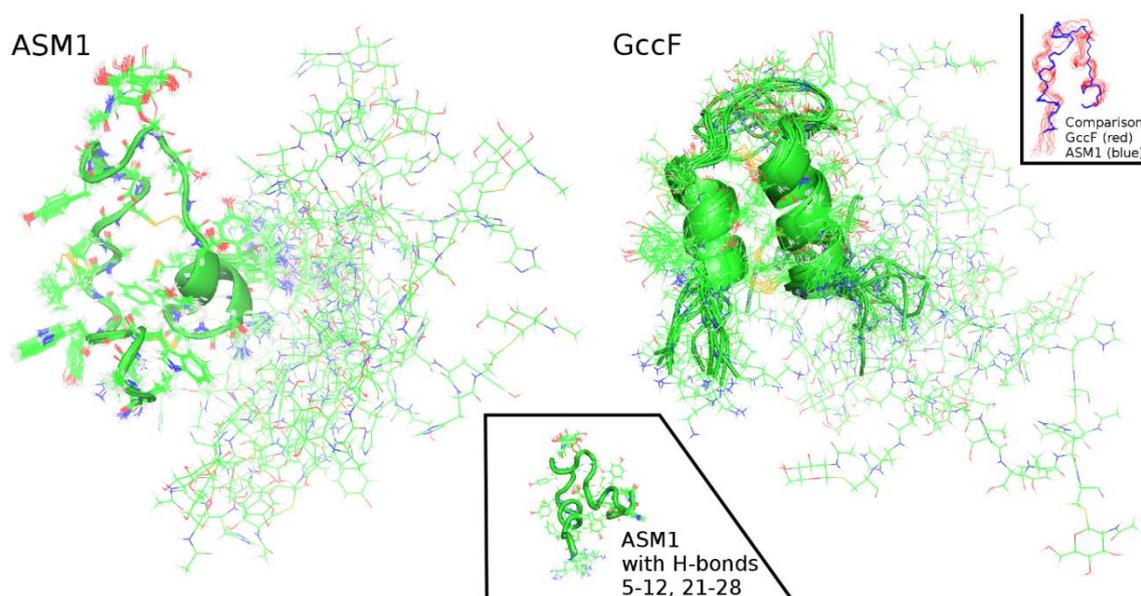
there were 20 inter-peptide-glycan NOE peaks, and for the GlcNAC attached to Cys-43, there were 11 inter-peptide-glycan NOE peaks.

When an  $\alpha$ -glycosidic linkage was modelled, it resulted in unassigned NOE peaks and violations in the structure calculations. As the intensity of the NOE peaks is strongly related to the distance between nuclei ( $r^{-6}$  dependence), NOEs attributed to specific nuclei restrain the relative spatial orientation of the surrounding atoms. Almost all the NMR peaks in the data set were assigned, with an average of 41 NOE peaks for each of the 43 amino acids as well as 54 and 34 NOE peaks for the O- and S-linked-GlcNAc moieties, respectively. Altogether, 1768 NOE peaks were used to produce a high-resolution structure.

## 2.2. Solution NMR Structure of ASM1

The structure is shown in Figure 6, and the statistics are summarized in Table 2.

The six criteria that define a high-resolution structure [15] were satisfied: the target function of cycle 1, a measure of how well the experimental distance restraints and torsion angle restraints are met and how much all non-bonded atom pairs satisfy a check for the absence of steric overlap, was  $15.86 \text{ \AA}^2$ , (well below the minimum value of  $250 \text{ \AA}^2$ ); the RMSD value in cycle 1 was  $2.71 \text{ \AA}$  (less than  $3 \text{ \AA}$ ); the average final CYANA target function was  $0.68 \text{ \AA}^2$ , (well below  $10 \text{ \AA}^2$ ); the RMSD between the mean structures of the first and last cycle was  $2.54 \text{ \AA}$  (below  $3 \text{ \AA}$ ); the total number of unassigned NOE peaks was 4.1% (less than 20%); the total number of discarded long-range NOE peaks ( $|i - j| > 5$ ) was 8.6% (below 20%).



**Figure 6.** Comparison of solution NMR structures of ASM1 and Glycocin F (GccF). Superpositions of the 20 model structures of ASM1 and the 12 model structures of GccF (PDB: 2KUY). A comparison of the backbones 1–27 is shown in the insert on the right-hand side. The bottom insert represents a model of the ASM1 structure calculated using hydrogen-bond restraints for  $\alpha$ -helices 5–12, 21–28; it resembles the GccF-model more closely.

The overall architecture of the molecule is similar to that of GccF, comprising two polypeptide segments (stems) linked by a pair of nested disulfide bonds (Cys-5–Cys-28 and Cys-12–Cys-21) that constrain the surrounding atoms. The structure is quite rigid from residues 1–28 as is shown by comparing the 20 model structures (Figure 6). The C-terminal residues from residue 29 on, appears to be largely unstructured and flexible. Although differences in the calculated model structures cannot be directly correlated with movement, they represent a collection of structures consistent with the

experimental NOE restraints that represent snapshots of movement. Just how much time the protein spends in each conformation is, however, not known.

**Table 2.** Statistics for the NMR structure solution of ASM1 (H-bonds: Hydrogen-bonds; a: for residues 1–27).

Input Data	Without H-Bond Restraints	With H-Bond Restraints
Completeness of NMR assignments (%)		
- of backbone amide protons and aliphatic protons	99.6%	99.6%
Total assigned atoms	438	438
Dihedral angle constraints ( $\phi$ , $\psi$ )	33 (17,16)	33 (17,16)
Total number of $^1\text{H}$ - $^1\text{H}$ -NOESY derived peaks	1768	1768
CYANA-version used	3.0	3.0
Restraints for H-bonds $\alpha$ -helices 5–12, 21–28	0	16
<b>Output Data</b>		
CYANA target function value ( $\text{\AA}^2$ )	0.68	12.00
NOE-derived distance constraints	577	569
- Intraresidual ( $ i - j  < 1$ )	323	295
- Medium-range ( $1 < i - j < 5$ )	138	168
- Long-range ( $ i - j  > 5$ )	116	106
Unassigned NOE peaks	72	98
RMS deviation from idealized geometry		
- Bond lengths ( $\text{\AA}$ )	0.004	0.005
- Bond angles (degrees)	0.7	0.8
RMS deviation from averaged coordinates ( $\text{\AA}$ ) <sup>a</sup>		
- Backbone	0.17	0.11
- Heavy atoms	0.38	0.42
Ramachandran analysis (%) <sup>a</sup>		
- Most favoured	52.8	48.4
- Additionally allowed	38.8	40.5
- Generously allowed	8.4	11.1
- Disallowed	0.0	0.0

Segments from residues 5–12 and 21–28 do not appear to be  $\alpha$ -helical in contrast to what is seen in GccF and sublancin. The GccF structure [5] is quite different (RMSD of 2.2  $\text{\AA}$  between residues 1–27) from the ASM1 model (see Figure 6), especially in the stem regions between residues 5–12 and 21–28. This was completely unexpected and prompted a thorough analysis of the data describing these segments to ensure a correct interpretation had been made. The Kabsch/Sander algorithm assigned only residues 24–26, to a helix, predicting they form a  $3_{10}$ -helix. This is at odds with the CD analysis that predicted the peptide contained 26% standard and 18% distorted helical content. Residues 13–20 form a loop between the disulfide bonded residues Cys-12 and Cys-21.

We next looked for the presence or absence of characteristic NOE cross peaks that are commonly found in helices:  $\text{H}_\text{N}\text{H}_\text{N}(i,i+1)$ ,  $\text{H}_\text{N}\text{H}_\text{N}(i,i+2)$ ,  $\text{H}_\alpha\text{H}_\text{N}(i,i+3)$ ,  $\text{H}_\alpha\text{H}_\text{N}(i,i+4)$ ,  $\text{H}_\alpha\text{H}_\beta(i,i+3)$  [16]. Some were present while others were not. The absence of a clear pattern reflects the absence of clear helical structure.

A chemical shift index (CSI) analysis was also carried out comparing the ASM1 data with the data collected from GccF to investigate possible differences in secondary structure formation. This was accomplished by inserting the chemical shift lists of ASM1 and GccF, in particular of HA, CA, CB, CO, into a computer program [15,17] that calculates the CSI. The chemical shift lists are available from the Biological Magnetic Resonance Bank (BMRB), ID 25269 for ASM1 and ID 16747 for GccF, corresponding to the structures deposited into the Protein Data Bank (PDB), accession codes 2mvi and

2KUY, respectively. By calculating the CSI [17], based on the difference between measured chemical shifts with respect to some predefined “random coil” values, regions of helical and sheet propensity can be probed as positive and negative values, respectively. Figure 7 shows that the regions of helical secondary structure encompass residues 4–11 and 20–27 for ASM1, and 5–10, and 20–27 for GccF. It should be noted however, that these analyses cannot distinguish distorted helical structures, such as  $3_{10}$  helices from  $\alpha$ -helices. The biggest differences in the CSI values are seen for residues 38 and 39, which are glycines in ASM1 and serines in GccF. There are, however, very few differences for the two structures around the loop region 13–20, including Ser-18 where GlcNAc is attached, or in the CSI values for residues 41–43, including Cys-43 linked to the second GlcNAc. This would suggest comparable chemical environments for the GlcNAc moieties of ASM1 and GccF. In both structures, the disulfide bonds give a stability restricting the movement of the GlcNAc attached to Ser-18 (O-linked), while the long flexible tail allows the GlcNAc moiety on Cys-43 (S-linked) to sample a large volume of space.

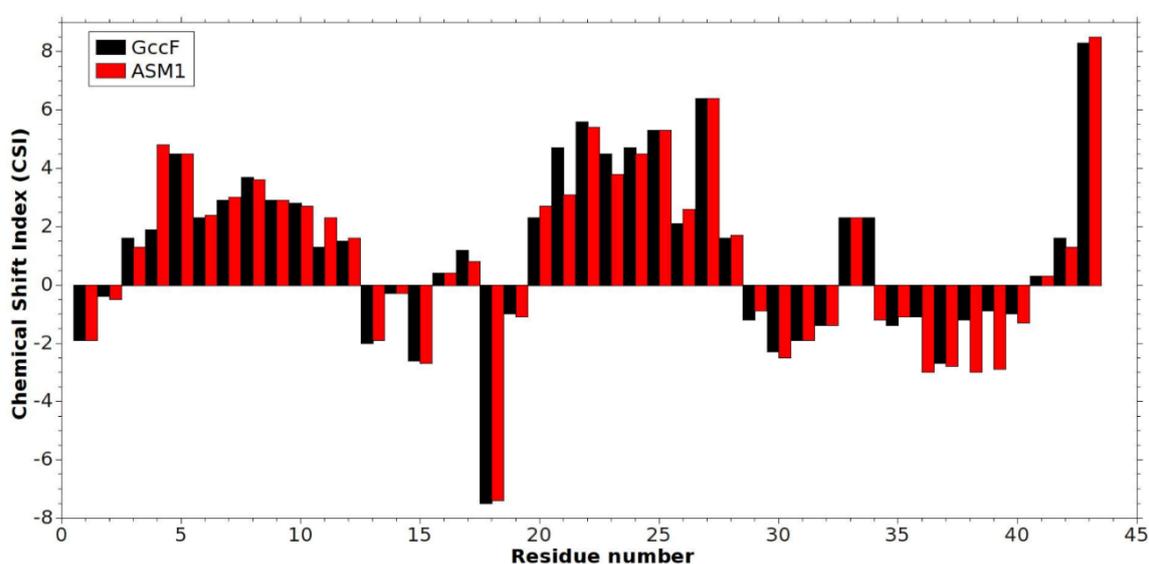
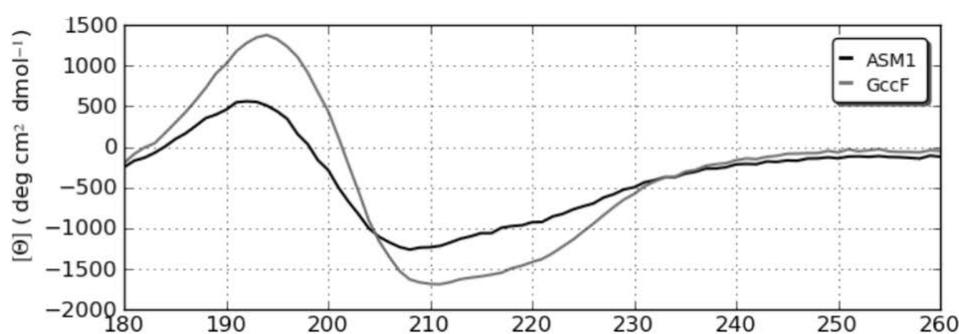


Figure 7. Chemical shift index (CSI).

In order to increase confidence in the structure observed between residues 5–12 and 21–28, hydrogen-bond restraints characteristic for both  $\alpha$ -helical (the secondary structure experimentally verified by H/D exchange for these residues in GccF) and for  $3_{10}$ -helical structure were imposed on residues 5–12 and 21–28 or residues 4–11 and 19–30. The goal of this exercise was to observe whether the resulting structures are consistent with the experimental NOE data. This is reflected by the value of the CYANA target function and an increase is reflected in a larger number of violations. The results are shown in Table 2. The structures resemble GccF more closely (Figure 6), but the target function increased from  $0.68 \text{ \AA}^2$  to  $12.00 \text{ \AA}^2$  when  $\alpha$ -helical restraints were imposed between residues 5–12, 21–28, and to  $22.11 \text{ \AA}^2$  when  $3_{10}$  restraints were imposed, i.e., 18 fold and 32 fold, respectively. Clearly, these secondary structures are not consistent with the ASM1 NMR data.

### 2.3. Circular dichroism (CD) Spectroscopy

The CD spectrum of ASM1 was slightly different to that of GccF (Figure 8) despite both spectra being collected at almost identical concentrations ( $198 \mu\text{M}$  for ASM1 and  $192 \mu\text{M}$  for GccF) and in the same solvent.



**Figure 8.** Far ultra-violet circular dichroism (UV CD) spectra of ASM1 and GccF.

The band at 192 nm in ASM1 was of a slightly shorter wavelength than that for GccF at 194 nm and was considerably less intense. A similar band shift was seen at 208 nm for ASM1 compared to 210 nm for GccF and again was much more intense for the latter. Comparing the spectra using the program CONTINLL [18] predicted that ASM1 contained 26.4% regular helix and 17.9% distorted helix compared to 35.1% regular and 13.4% distorted for GccF. The amount of  $\beta$  structure was predicted to be minimal (~9% distorted sheet), with the remainder of the structure predicted to be made up of turns and unordered segments. The confidence level as estimated by the root mean square deviation (NRMSD) of 0.092 for ASM1 and 0.082 for GccF was high. The reasons for the differences between the two peptides are not obvious, as the potentially structured parts of each molecule had identical sequence. One possible explanation is that short transient structural preferences could partially exist even within the disordered region.

### 3. Discussion

The ASM1 structure possesses some unexpected features in the segments between residues 5 and 12 and between residues 21 and 28. It might be assumed a priori that these segments should be  $\alpha$ -helical because the earlier NMR structures of the glycocins GccF and sublancin 168 have structures that are  $\alpha$ -helical between the stem regions of the molecule [5,6]. The bacteriocins ASM1 and GccF share 88% amino acid identity and all amino acid substitutions occur in the unstructured C-terminal “tail”: V31I, S34H, G36S, G38S, G39S. On the basis of sequence, their structures are, therefore, likely to be similar. Furthermore, on the basis of conservation of secondary structure, sublancin, which has only 5 residues that are identical to GccF and ASM1 on the basis of amino acid sequence has the same disulfide-stabilized Cs  $\alpha/\alpha$  fold. However, CD results indicate that there might be some difference in the helical content. Bearing in mind that the GlcNAcs have been shown to contribute to the CD signal at 193 and 210 nm [19], and the reference set used to analyse the data contains relatively large globular proteins, the CD-based secondary structure predictions for both GccF and ASM1 should be regarded with some caution. More compelling for this model is the quality and resolution of the NMR data. The structure of ASM1 is based on 1768 NOE peaks versus 443 for GccF. Furthermore, the experimentally derived data is less compatible with a model that restrains residues 5–12 and 21–28 to be  $\alpha$ -helices as in the GccF structure [5] or  $3_{10}$  helices. Further experimental evidence for helical formation can be supplied by deuterium exchange experiments as in the case of GccF and sublancin [5,6]; however, those authors were handicapped by a relative dearth of NOE cross peaks, and a higher number of NOE cross peaks is generally preferred [16]. Despite the high sequence identity, the best ASM1 model structure does indeed differ from that reported for GccF. While this may seem unlikely, small changes in sequence have previously been shown to have dramatic effects [20,21]. In the case of Streptococcus protein G, only one mutation (L45Y) radically changed the fold of the protein from one containing two helices to a structure containing four  $\beta$ -strands [20], a change of structure much more dramatic than the differences observed between ASM1 and GccF [5]. However, in these two structures the core sequences, residues 1–30, are identical. Thus, any difference in the final tertiary structure must

originate from difference in folding. It is possible that sequence differences in the cluster proteins responsible for the maturation of ASM1 [9] may result in changes to the folding pathway for ASM1, resulting in a different folded structure.

#### 4. Materials and Methods

Full-length ASM1 was purified from *Lactobacillus plantarum* A-1 [5], using methods similar to that used to purify GccF. Briefly, 50 mL of an overnight culture of *L. plantarum* A-1 in MRS media was used to inoculate 4 L of MRS which was incubated without shaking or aeration for four days at 25 °C. Cells were removed from the media by centrifugation at 6000 × *g* for 30 min after which the pH was reduced, if necessary, to approximately pH 3.8 using formic acid. 50 mL of SP-Sephadex (GE Healthcare) pre-equilibrated in 50 mM sodium formate pH 3.8 was mixed with the 4 L of clarified supernatant and stirred overnight at room temperature using an overhead stirrer. The unbound fraction was decanted until the resin formed a slurry that was packed into an XK30 (GE-Healthcare) column by gravity. The column was then washed with 1 L of equilibration buffer and bound protein was eluted with 1 L of 20 mM MOPS (pH 7.2), followed by 0.5 L 20 mM ammonium bicarbonate (AmBic) and finally 0.5 L 20 mM AmBic in 60% acetonitrile (MeCN). 50 mL fractions were collected during the elution and tested for activity using an indicator plate assay [6]. Fractions with ASM1 activity (eluted by AmBic + 60% MeCN) were concentrated by lyophilisation. The lyophilised powder was dissolved in distilled water, adjusted to pH 4.0 using dilute acetic acid and further purified by HPLC (2 mL injection volume, Jupiter C<sub>18</sub> 10 × 250 mm) using a linear gradient: 0.1% trifluoroacetic acid (TFA) in distilled water, to 0.08% TFA in 40 % MeCN, over 15 min at 4 mL/minute. Peak fractions were collected manually. The fractions were tested for activity using indicator plates and the active fractions were then combined and lyophilised for storage at −20 °C.

##### 4.1. NMR Spectroscopy

For NMR measurements, 3.7 mg of the purified protein was mixed in 700 µL of a solution containing 40% MeCN-d<sub>3</sub>, 59.8% water, and 0.2% acetic acid-d<sub>4</sub>, resulting in 1.05 mM ASM1. 5% D<sub>2</sub>O was added for deuterium frequency lock to compensate for slight magnetic drifts of the magnet.

All experiments were performed at 310.0 K (36.85 °C) on a Bruker AVANCE 700 MHz spectrometer, operating at 700.13 MHz, equipped with four radiofrequency channels, gradient pulse capabilities, and a cryoprobe (CPTCI <sup>1</sup>H-<sup>13</sup>C/<sup>15</sup>N/D Z-GRD), running Topspin version 2.1 (Bruker-Biospin GmbH, Rheinstetten, Germany). As the ASM1 sample was unlabelled, i.e., the NMR active isotopes <sup>15</sup>N and <sup>13</sup>C were at natural abundance, at just 0.366% and 1.1%, respectively, the NMR experiments used for structure determination were not the same as for an isotopically enriched sample [21]. The following experiments were carried out: <sup>1</sup>H-<sup>15</sup>N HSQC (spectral widths: 13.95 ppm (2048 points) for <sup>1</sup>H and 44 ppm (256 points) for <sup>15</sup>N, 256 scans), <sup>1</sup>H-<sup>13</sup>C HSQC (spectral widths: 16.08 ppm (2048 points) for <sup>1</sup>H and 166.05 ppm (512 points) for <sup>13</sup>C, 96 scans), <sup>1</sup>H-<sup>13</sup>C HMBC (spectral widths: 13.95 ppm (4096 points) for <sup>1</sup>H and 180 ppm (256 points) for <sup>13</sup>C, 336 scans), <sup>1</sup>H-<sup>1</sup>H-TOCSY (spectral widths: 13.95 ppm (4096 points) for first <sup>1</sup>H and 14 ppm (512 points) for second <sup>1</sup>H, 96 scans, 90 ms mixing time), <sup>1</sup>H-<sup>15</sup>N-HSQC-TOCSY (spectral widths: 13.95 ppm (2048 points) for <sup>1</sup>H and 21.9 ppm (128 points) for <sup>15</sup>N, 1408 scans, 60 ms mixing time), <sup>1</sup>H-<sup>13</sup>C-HSQC-TOCSY (spectral widths: 16.08 ppm (2048 points for decoupled and 16384 points for non-decoupled experiments) for <sup>1</sup>H and 146 ppm (192 points) for <sup>13</sup>C, 600 scans, 100 ms mixing time), <sup>1</sup>H-<sup>13</sup>C -H2BC (spectral widths: 16.08 ppm (1024 points) for <sup>1</sup>H and 146 ppm (160 points) for <sup>13</sup>C, 512 scans), <sup>1</sup>H-<sup>1</sup>H-NOESY (spectral widths: 13.95 ppm (4096 points) for first <sup>1</sup>H and 14 ppm (512 points) for second <sup>1</sup>H, 256 scans, 200 ms mixing time). The latter experiment was repeated at 277.0 K (3.85 °C) with 384 scans to account for the lower signal-to-noise ratio, to assist in chemical shift assignments in case of overlapped resonances. The spectra were processed with the program NMRPipe version 2011.118.08.55 [22] and were analysed with the programs Kujira version 0.9843 [23] and NMRView [24]. Sequence-specific backbone and sidechain assignments of the amino acids were mainly based on the HSQC, <sup>1</sup>H-<sup>13</sup>C -HSQC-TOCSY, and NOESY spectra. For the

post-translational modification assignments and their linkages to the amino acid chain, HSQC, TOCSY,  $^1\text{H}$ - $^{13}\text{C}$ -H2BC and NOESY spectra were found to be most useful. NOESY spectra, collected at 310.0 K (36.85 °C) were used to obtain distance restraints.

Structural calculations of the protein with its native post-translational modifications required an elaborate procedure. Specifically, the program HIC-Up [25] was used to retrieve the topology of the GlcNAc moieties, and hydrogen atoms were attached with MolProbity [26]. The resulting file was incorporated into the CYANA library file using a text editor, ensuring that appropriate rotational degrees of freedom were maintained. At this stage of model development, the two GlcNAcs were not covalently linked to the polypeptide chain. To establish these bonds, the “link” statement which eliminates the steric repulsion between two covalently bound atoms was used in CYANA. As this instruction will not enforce the formation of a covalent bond, additional explicit distance constraint conditions had to be written. The amino acid chain without post-translational modifications and the two GlcNAc molecules required for CYANA calculation were thus incorporated into a single sequence file and the molecules linked together using a stretch of “invisible” linker residues as follows: Twenty “invisible” linker residues of type LL2 (linker residue with a virtual bond length of 2 Å) were placed between residue 43 and the one GlcNAc followed by a further 20 residues to the second GlcNAc. To connect the real residues to the protein sequence one linker of type PL and one of type LP were used. PL stands for a link between the protein P and an invisible linker L. LL stands for an invisible linker linking to another invisible linker. Disulfide bond restraints were set when there was corroborating evidence from the chemical shifts of the cysteine residues involved, as well as strong NOE evidence. Dihedral angle restraints ( $\phi$ ,  $\psi$ ) were calculated using TALOS [27] with the HA, CA, CB, CO, and N chemical shifts, and used conservatively if there was no conflicting evidence from the NOE data. No hydrogen bond constraints were used as input for the structure calculation, due to the lack of direct experimental evidence [28,29]. Because of the differences observed between this model and the GccF structure [5], hydrogen-bond restraints were added to force a helical structure between residues 5–12 and 21–28, in order to test the compatibility of the experimental data with this model.

In order to obtain a better structure, refinement in explicit solvent is done often. However, the ASM1 structure was not refined in explicit solvent, as the solvent composition (40% MeCN- $d_3$ , 59.8% water, 0.2% acetic acid- $d_4$ ) is not constant within the protein solution due to the “salting out” of acetonitrile at the protein surface: Although d-acetic acid forms a zeotropic mixture [30] in this solvent system, acetonitrile and water form an azeotrope [31], implying there must be a change of free energy to account for the acetonitrile-protein interactions [32]. The free energy of transfer from water to aqueous acetonitrile is negative for most nonpolar sidechains of amino acids and positive for the peptide group.

Automated NOE (Nuclear Overhauser Effect) peak assignments and structure calculations with torsion angle dynamics were carried out using CYANA version 3.0 [33,34]. A total of 100 structures were independently calculated and the 20 best (according to the target function) were selected for further analysis. The structures were validated using PROCHECK-NMR [35]. PyMol [36] and Molmol [37] were used to analyse the calculated structures and to prepare models. The atomic coordinates are available from the Protein Data Bank (PDB) [38], accession code 2mvi. Further information is available from the Biological Magnetic Resonance Bank (BMRB), ID 25269.

#### 4.2. CD Spectroscopy

Circular dichroism (CD) spectra were collected on a Chirascan<sup>TM</sup> spectrometer (Applied Photophysics, U.K.). Samples of ASM1 and GccF were diluted in degassed 40% MeCN, 0.2% acetic acid, to final concentrations of about 1 mg mL<sup>-1</sup> (0.198 and 0.192 mM respectively) using extinction coefficients of 36 at 205 nm and 2.7 and 2.5 at 280 nm for 1 mg mL<sup>-1</sup> solutions of GccF and ASM1 respectively. Approximately 100  $\mu\text{L}$  of sample in a Quartz SUPRASIL precision cell (Hellma<sup>®</sup>, Germany) with a 0.1 mm path length was used for each measurement. Spectra were collected using the following parameters: wavelength range, 180–250 nm; time per point, 0.5 s; bandwidth, 1 nm; step

size, 0.5 nm; number of repeats, 20, temperature 20 °C. Data were analysed using CONTILL [39] with reference data set 3 on the Dichroweb [40] interactive web site.

**Author Contributions:** A.K.G. collected, analysed, and interpreted the NMR spectra and wrote the initial draft of the manuscript; T.S.L. collected CD data and prepared samples; A.M.K. prepared samples; M.L.P. and G.E.N. initiated and planned the research, and participated in data interpretation and manuscript writing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** We thank Tomomi Hata for the kind gift of *Lactobacillus plantarum* strain A-1.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Heng, N.C.K.; Tagg, J.R. What's in a name? Class distinction for bacteriocins. *Nat. Rev. Microbiol.* **2016**, *4*, 160. [[CrossRef](#)]
2. Alvarez-Sieiro, P.; Montalban-Lopez, M.; Mu, D.; Kuipers, O.P. Bacteriocins of lactic acid bacteria: Extending the family. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 2939–2951. [[CrossRef](#)] [[PubMed](#)]
3. Arnison, P.G.; Bibb, M.J.; Bierbaum, G.; Bowers, A.A.; Bugni, T.S.; Bulaj, G.; Camarero, J.A.; Campopiano, D.J.; Challis, G.L.; Clardy, J.; et al. Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **2013**, *30*, 108–160. [[CrossRef](#)] [[PubMed](#)]
4. Norris, G.E.; Patchett, M.L. The glycocins: In a class of their own. *Curr. Opin. Struct. Biol.* **2016**, *40*, 112–119. [[CrossRef](#)]
5. Venugopal, H.; Edwards, P.J.B.; Schwalbe, M.; Claridge, J.K.; Libich, D.S.; Stepper, J.; Loo, T.; Patchett, M.L.; Norris, G.E.; Pascal, S.M. Structural, dynamic, and chemical characterization of a novel S-glycosylated bacteriocin. *Biochemistry* **2011**, *50*, 2748–2755. [[CrossRef](#)]
6. Garcia De Gonzalo, C.V.; Zhu, L.; Oman, T.J.; van der Donk, W.A. NMR structure of the S-linked glycopeptide sublancin 168. *ACS Chem. Biol.* **2014**, *9*, 796–801. [[CrossRef](#)]
7. Stepper, J.; Shastri, S.; Loo, T.S.; Preston, J.C.; Novak, P.; Man, P.; Moore, C.H.; Havlicek, V.; Patchett, M.L.; Norris, G.E. Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Lett.* **2011**, *585*, 645–650. [[CrossRef](#)]
8. Hata, T.; Tanaka, R.; Ohmomo, S. Isolation and characterization of plantaricin ASM1: A new bacteriocin produced by *Lactobacillus plantarum* A-1. *Int. J. Food Microbiol.* **2010**, *137*, 94–99. [[CrossRef](#)]
9. Main, P.; Hata, T.; Loo, T.S.; Man, P.; Novak, P.; Havlíček, V.; Norris, G.E.; Patchett, M.L. Bacteriocin ASM1 is an O/S-diglycosylated plasmid-encoded homologue of glycocin F. *FEBS Lett.* **2019**. [[CrossRef](#)] [[PubMed](#)]
10. Hsieh, Y.S.; Wilkinson, B.L.; O'Connell, M.R.; Mackay, J.P.; Matthews, J.M.; Payne, R.J. Synthesis of the bacteriocin glycopeptide sublancin 168 and S-glycosylated variants. *Org. Lett.* **2012**, *14*, 1910–1913. [[CrossRef](#)] [[PubMed](#)]
11. Chagot, B.; Pimentel, C.; Dai, L.; Pil, J.; Tytgat, J.; Nakajima, T.; Corzo, G.; Darbon, H.; Ferrat, G. An unusual fold for potassium channel blockers: NMR structure of three toxins from the scorpion *Opisthacanthus madagascariensis*. *Biochem. J.* **2005**, *388*, 263–271. [[CrossRef](#)] [[PubMed](#)]
12. Moller, C.; Rahmankhah, S.; Lauer-Fields, J.; Bubis, J.; Fields, G.B.; Mari, F. A novel conotoxin framework with a helix-loop-helix (Cs alpha/alpha) fold. *Biochemistry* **2005**, *44*, 15986–15996. [[CrossRef](#)] [[PubMed](#)]
13. Nolde, S.B.; Vassilevski, A.A.; Rogozhin, E.A.; Barinov, N.A.; Balashova, T.A.; Samsonova, O.V.; Baranov, Y.V.; Feofanov, A.V.; Egorov, T.A.; Arseniev, A.S.; et al. Disulfide-stabilized helical hairpin structure and activity of a novel antifungal peptide EcAMP1 from seeds of barnyard grass (*Echinochloa crus-galli*). *J. Biol. Chem.* **2011**, *286*, 25145–25153. [[CrossRef](#)] [[PubMed](#)]
14. Saucedo, A.L.; Flores-Solis, D.; Rodriguez de la Vega, R.C.; Ramirez-Cordero, B.; Hernandez-Lopez, R.; Cano-Sanchez, P.; Noriega Navarro, R.; Garcia-Valdes, J.; Coronas-Valderrama, F.; de Roodt, A.; et al. New tricks of an old pattern: Structural versatility of scorpion toxins with common cysteine spacing. *J. Biol. Chem.* **2012**, *287*, 12321–12330. [[CrossRef](#)]

15. Güntert, P. Automatic NMR Structure Calculation with CYANA. *Methods Mol. Biol.* **2004**, *278*, 353–378. [[CrossRef](#)]
16. Dyson, H.J.; Wright, P.E. Defining solution conformations of small linear peptides. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *60*, 795–825. [[CrossRef](#)]
17. Wishart, D.S.; Sykes, B.D.; Richards, F.M. The chemical shift index: A fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **1992**, *31*, 1647–1651. [[CrossRef](#)]
18. Provencher, S.W.; Glockner, J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **1981**, *20*, 33–37. [[CrossRef](#)]
19. Domard, A. Circular dichroism study on N-acetylglucosamineoligomers. *Int. J. Biol. Macromol.* **1986**, *8*, 243–246. [[CrossRef](#)]
20. Alexander, P.A.; He, Y.; Chen, Y.; Orban, J.; Brayn, P.N. A minimal sequence code for switching protein structure and function. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21149–21154. [[CrossRef](#)]
21. Goroncy, A.K.; Koshiba, S.; Tochio, N.; Tomizawa, T.; Sato, M.; Inoue, M.; Watanabe, S.; Hayashizaki, Y.; Tanaka, A.; Kigawa, T.; et al. NMR solution structures of actin depolymerizing factor homology domains. *Protein Sci.* **2009**, *18*, 2384–2392. [[CrossRef](#)] [[PubMed](#)]
22. Delaglio, F.; Grzesiek, S.; Vuister, G.W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995**, *6*, 277–293. [[CrossRef](#)] [[PubMed](#)]
23. Kobayashi, N.; Iwahara, J.; Koshiba, S.; Tomizawa, T.; Tochio, N.; Güntert, P.; Kigawa, T.; Yokoyama, S. KIJIRA, a package of integrated modules for systematic and interactive analysis of NMR data directed to the high-throughput NMR structure studies. *J. Biomol. NMR* **2007**, *39*, 31–52. [[CrossRef](#)] [[PubMed](#)]
24. Johnson, B.; Blevins, R. NMRView: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **1994**, *4*, 603–614. [[CrossRef](#)]
25. Kleywegt, G.J. Crystallographic refinement of ligand complexes. *Acta Crystallogr. Sect. D* **2007**, *63*, 94–100. [[CrossRef](#)]
26. Chen, V.B.; Arendall, W.B., 3rd; Headd, J.J.; Keedy, D.A.; Immormino, R.M.; Karpral, G.J.; Murray, L.W.; Richardson, J.S.; Richardson, D.C. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D* **2010**, *66*, 12–21. [[CrossRef](#)]
27. Cornilescu, G.; Delaglio, F.; Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **1999**, *13*, 289–302. [[CrossRef](#)]
28. Cordier, F.; Nisius, L.; Dingley, A.J.; Grzesiek, S. Direct detection of N–H=O=C hydrogen bonds in biomolecules by NMR spectroscopy. *Nat. Protoc.* **2008**, *3*, 235–241. [[CrossRef](#)]
29. Blackledge, M. NMR provides evidence for dynamic hydrogen bonding in proteins. *Protein Sci.* **2007**, *16*, 1247–1248. [[CrossRef](#)]
30. Singh, S. Excess Gibbs for binary mixtures of acetonitrile with acetic acid, propionic acid, isobutyric acid, and trimethylacetic acid. *Can. J. Chem.* **1991**, *69*, 2117–2121. [[CrossRef](#)]
31. Pavlov, S.Y.; Pavlova, S.P.; Kirnos, A.B. Azeotropes of acetonitrile with C5 hydrocarbons and with water. *Zhurnal Prikladnoi Khimii* **1966**, *39*, 1555–1559.
32. Gekk, K.; Ohmae, E.; Kameyama, K.; Takagi, T. Acetonitrile-protein interactions: Amino acid solubility and preferential solvation. *Biochim. Biophys. Acta* **1998**, *1387*, 195–205. [[CrossRef](#)]
33. Güntert, P. Automated NMR structure calculation. *Prog. Nucl. Magn. Reson. Spectrosc.* **2003**, *43*, 105–125. [[CrossRef](#)]
34. Güntert, P.; Mumenthaler, C.; Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program CYANA. *J. Mol. Biol.* **1997**, *273*, 283–298. [[CrossRef](#)]
35. Laskowski, R.A.; Rullmann, J.A.; MacArthur, M.W.; Kaptein, R.; Thornton, J.M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **1996**, *8*, 477–486. [[CrossRef](#)] [[PubMed](#)]
36. DeLano, W.L. *The PyMol Molecular Graphics System*; DeLano Scientific: San Carlos, CA, USA, 2008.
37. Koradi, R.; Billeter, M.; Wüthrich, K. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **1996**, *14*, 51–55. [[CrossRef](#)]
38. Available online: <http://www.pdb> (accessed on 2 March 2020).

39. Whitmore, L.; Wallace, B.A. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* **2008**, *89*, 392–400. [[CrossRef](#)] [[PubMed](#)]
40. Lobley, A.; Whitmore, L.; Wallace, B.A. DICHROWEB: An interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* **2002**, *18*, 211–212. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).