BIOCHEMISTRY
including biophysical chemistry & molecular biology

ACCELERATED PUBLICATION

pubs.acs.org/biochemistry

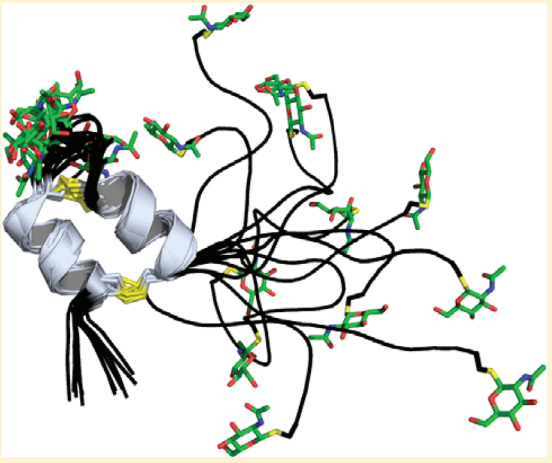
Structural, Dynamic, and Chemical Characterization of a Novel S-Glycosylated Bacteriocin

Hariprasad Venugopal,*,† Patrick J. B. Edwards,† Martin Schwalbe,† Jolyon K. Claridge,† David S. Libich,† Judith Stepper,‡ Trevor Loo,‡ Mark L. Patchett,‡ Gillian E. Norris,‡ and Steven M. Pascal*,†

†Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

‡Institute of Molecular Biosciences, Massey University, Palmerston North, New Zealand

ABSTRACT: Bacteriocins are bacterial peptides with specific activity against competing species. They hold great potential as natural preservatives and for their probiotic effects. We show here nuclear magnetic resonance-based evidence that glycocin F, a 43-amino acid bacteriocin from *Lactobacillus plantarum*, contains two β-linked *N*-acetylglucosamine moieties, attached via side chain linkages to a serine via oxygen, and to a cysteine via sulfur. The latter linkage is novel and has helped to establish a new type of post-translational modification, the S-linked sugar. The peptide conformation consists primarily of two α-helices held together by a pair of nested disulfide bonds. The serine-linked sugar is positioned on a short loop sequentially connecting the two helices, while the cysteine-linked sugar presents at the end of a long disordered C-terminal tail. The differing chemical and conformational stabilities of the two N-actetylglucosamine moieties provide clues about the possible mode of action of this bacteriostatic peptide.



Glycosylation is a common post-translational modification in eukaryotes but has been less well characterized in bacteria. The most common types of glycosylation involve linkage via the side chain N_δ atom of asparagine (N-linked) or to the side chain O_γ atom of serine or threonine (O-linked), and no linkages via side chain sulfur atoms had been previously described, other than two unverified reports in 1971.[1,2] This situation changed when *Bacillus subtilis* was shown to incorporate a glucose into the peptide sublancin via a side chain linkage to a cysteine residue through the sulfur atom,[3] and a similar S-linkage to a HexNAc-sized moiety was shown for the glycocin F peptide (GccF),[4] a bacteriocin from *Lactobacillus plantarum*. Thioether linkages are far more stable than N- or O-glycocidic linkages at both low and high pH[5−7] and may offer a mechanism for bypassing the threat of deglycosylating enzymes. The stability of synthetic S-linkages has already led to a recent associated thrust regarding therapeutic peptide development.[8−12] These discoveries, interesting in themselves, also suggest that there may be additional types of unidentified post-translational modifications awaiting discovery.

Bacteriocins such as GccF are ribosomally synthesized peptides secreted by bacteria to inhibit competing species, and they typically exhibit a phylogenetically narrow killing spectrum.[13] Bacteriocins have great potential for use as natural preservatives and are responsible for some probiotic effects of various lactic acid bacteria used in the production of fermented foods (e.g., ref 14). On the basis of their size, mode of action, and post-translational modifications, they have been categorized into four classes, with class IV comprised of "complex" bacteriocins containing carbohydrate and/or lipid moieties.[15] A lack of experimental evidence for class IV bacteriocins had led to the suggestion that this class be withdrawn from classification schemes,[16] but the recent characterizations of sublancin and GccF as glycopeptides conclusively re-established the existence of this class.

We report here a nuclear magnetic resonance-based chemical, structural, and dynamic analysis of GccF from the KW30 strain of *L. plantarum*, the first such analysis of any S-glycosylated peptide or protein. This class IV bacteriocin contains two GlcNAc moieties, one attached via a linkage to the O_γ atom from Ser18 and the second via a linkage to the S_γ atom from the C-terminal residue, Cys43. The peptide is conformationally comprised of a pair of two-turn helices tethered to each other by a pair of disulfide bonds. The helices are sequentially connected by a loop presenting the O-glycosylated Ser18. The second helix is followed by a highly dynamic C-terminal tail presenting the S-glycosylated Cys43 residue at its tip. Structural insights provide clues about the possible mode of action of this novel bacteriocin and may help to further rationalize the design of S-linked therapeutic agents.

EXPERIMENTAL PROCEDURES

Sample Preparation. Full-length glycocin F (43 amino acids) was produced with native post-translational modifications and at

Received: February 10, 2011
Published: March 11, 2011

natural isotopic abundance via direct purification from *L. plantarum*. This sample was used for all structural studies. In addition, for the purposes of assigning the sugar resonances, two fragments termed GccF-N (amino acids 1−32) and GccF-C (amino acids 33−43), containing one sugar each, were produced via tryptic digestion. To improve nuclear magnetic resonance (NMR) spectral quality, a mixed solvent system of 40% acetonitrile and 0.2% acetic acid in water was used. Sample preparation is described in detail elsewhere.[4]

NMR Spectroscopy. All experiments were performed at natural isotopic abundance. NMR experiments were conducted on a Bruker Avance 700 MHz spectrometer equipped with a cryoprobe, four radiofrequency channels, and gradient pulse capabilities, with glycocin F at a concentration of 3 mM and at 310 K unless otherwise noted. NMR spectra were processed using standard parameters with Topspin version 2.1 (Bruker-Biospin GmbH, Rheinstetten, Germany).

Peptide Chemical Shift Assignment and NOEs. All peptide assignment and analysis was done using CCPNMR Analysis version 1.0.[24] Chemical shifts were assigned by means of $^{1}H-^{15}N$ HSQC spectra recorded with a spectral width of 13.95 ppm (2048 points) and 44 ppm (256 points) in the ^{1}H and ^{15}N dimensions, respectively. $^{1}H-^{15}N$ HSQC-TOCSY [25−27] (60 ms TOCSY mixing time) spectra were recorded with sweep widths of 13.95 ppm (2048 points) and 21.90 ppm (256 points) in the ^{1}H and ^{15}N dimensions, respectively. The $^{1}H-^{13}C$ HSQC spectra were recorded with sweep widths of 13.95 ppm - (2048 points) and 166.05 ppm (256 points) in the ^{1}H and ^{13}C dimensions, respectively. The $^{1}H-^{13}C$ HSQC-TOCSY[28] (100 ms TOCSY mixing time) and $^{1}H-^{13}C$ H2BC spectra were recorded with sweep widths of 16.08 ppm (2048 points) and 146.00 ppm (256 points) in the ^{1}H and ^{13}C dimensions, respectively. The $^{1}H-^{1}H$ TOCSY and $^{1}H-^{1}H$ NOESY spectra were recorded with sweep widths of 13.95 ppm in each dimension, with 620 and 4096 points in the indirect and direct dimensions, respectively.

The $^{1}H-^{15}N$ HSQC, $^{1}H-^{15}N$ HSQC-TOCSY, ^{13}C HSQC, ^{13}C HSQC-TOCSY, ^{13}C H2BC,[7] and $^{1}H-^{1}H$ TOCSY (with TOCSY mixing times of 30, 60, and 90 ms) spectra were used to identify spin systems followed by $^{1}H-^{1}H$ NOESY walks for sequential assignment. $^{1}H-^{15}N$ HSQC and $^{1}H-^{1}H$ TOCSY spectra recorded at 298, 301, 310, and 318 K were used to resolve overlapped resonances. Spectra recorded at 310 K produced the least overlap, and the chemical shifts reported here are based on that temperature. $^{1}H-^{1}H$ NOESY spectra were recorded with mixing times of 100, 150, and 200 ms and established that spin diffusion was not a significant factor. The 200 ms NOESY data set was then used for chemical shift assignment and to derive distance constraints.

Sugar Chemical Shift Assignment. The identities and linkages of the two HexNAcs were determined using 600 μL samples of ~1.5 mM $GccF_{1-32}$ or $GccF_{33-43}$ in a 59.8% D_2O/40% acetonitrile-d_3/0.2% acetic acid-d_4 mixture. DQF-COSY, $^{1}H-^{13}C$ HSQC, $^{1}H-^{13}C$ H2BC, $^{1}H-^{13}C$ HSQC-TOCSY, $^{1}H-^{13}C$ HMBC, and one-dimensional (1D) ^{1}H selective-TOCSY spectra were recorded at 305 K. DQF-COSY spectra were recorded with a spectral widths of 5.25 kHz in each dimension and with 512 and 4096 points in the indirect and direct dimensions, respectively. Phase discrimination in the indirect dimension was achieved using the echo−antiecho method. Multiplicity-edited $^{1}H-^{13}C$ HSQC and unedited $^{1}H-^{13}C$ H2BC echo−antiecho spectra were recorded with spectral widths of 25.71 kHz (256 points) and 11.26 kHz (2048 points) in the indirect and direct dimensions, respectively. Broadband ^{13}C decoupling was applied during the acquisition time using the GARP decoupling scheme. $^{1}H-^{13}C$ HMBC echo−antiecho spectra without broadband decoupling were recorded with spectral widths of 38.74 kHz (256 points) and 9.76 kHz (4096 points) in the ^{13}C and ^{1}H dimensions, respectively. 1D ^{1}H selective TOCSY spectra with mixing times between 30 and 300 ms were recorded using a spectral width of 14.0 kHz and 32768 points. Selective excitation of the anomeric resonances was obtained using a 50 ms 180° sinc pulse placed between two 1 ms field gradient pulses of 7.5 G/cm.

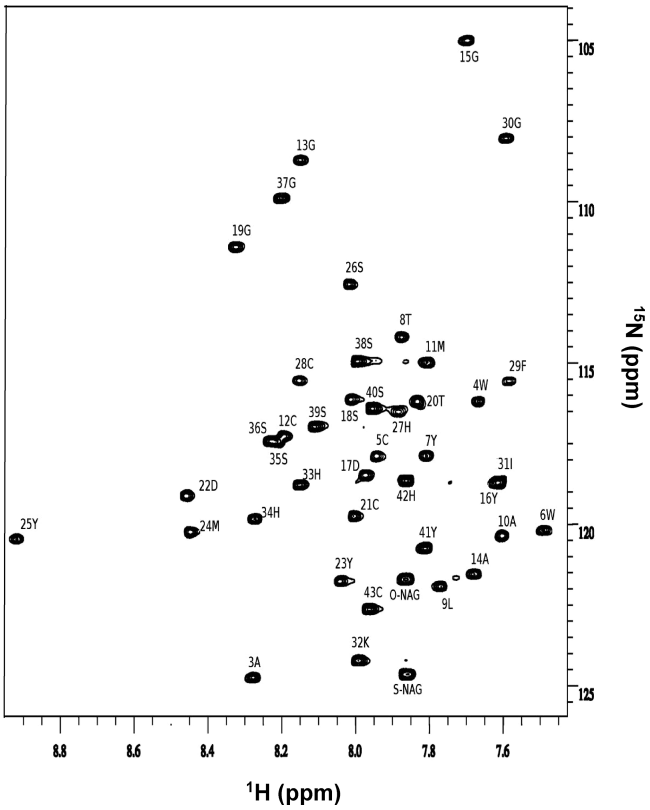


Figure 1. Backbone amide region of a $^{1}H-^{15}N$ HSQC spectrum of glycocin F at natural abundance (310 K). Sequence-specific assignments are indicated, and sugar correlations are labeled O-NAG and S-NAG.

Deuterium Exchange. A sample lyophilized from a 40% acetonitrile/0.2% acetic acid/D_2O mixture was dissolved into a 40% acetonitrile/0.2% acetic acid/H_2O mixture. A series of $^{1}H-^{15}N$ SOFAST-HMQC spectra with a 20 min duration each were recorded to quantitate arising peaks. The required experimental time of 20 min per spectrum was calibrated with a previous H_2O/acetonitrile/acetic acid-solvated sample and designed to provide the required sensitivity for detection of all heteronuclear correlations with the natural isotopic abundance sample.

Dynamics. The ^{15}N R_1 spectra were recorded with relaxation intervals of 51.1, 101.1, 251.1, 501.1, 751.1, and 1501.1 ms with spectral widths of 11.98 ppm (2048 points) and 26.00 ppm (128 points) in the ^{1}H and ^{15}N dimensions, respectively. The ^{15}N R_2 spectra were recorded with relaxation intervals of 16.3, 32.6, 48.9, 81.6, 114.2, 130.6, 146.9, and 179.6 ms with spectral widths of 11.98 ppm (1024 points) and 26.00 ppm (64 points) in ^{1}H and

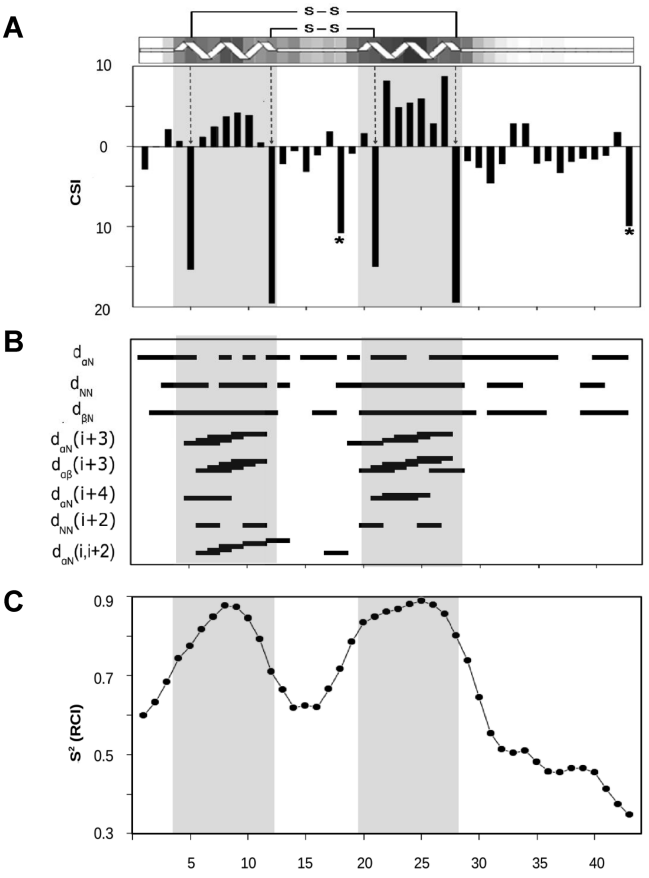


Figure 2. (A) Plot of the chemical shift index (CSI) using C_α and C_β resonances vs glycocin F residue number. Positive CSI values indicate helical propensities, while negative values indicate β-sheet propensity. Large negative CSI values (dashed lines) are observed for Cys5, Cys12, Cys21, and Cys28, because of the involvement in disulfide linkages. In addition, residues Ser18 and Cys43 (asterisks) also display large negative CSI values because of post-translational modification. (B) Plot of sequential and medium-range $^1H-^1H$ NOEs vs glycocin F residue number. Bars indicate the residues connected by the NOE. (C) Plot of the RCI-derived order parameter (S^2) vs glycocin F residue number. A PROCHECK-generated plot of secondary structure is shown above panel A, with solvent accessibility indicated via shading (dark is inaccessible), and disulfide bonds indicated schematically. Gray boxes extending through panels A–C highlight the PROCHECK-determined helical regions.

^{15}N dimensions, respectively. $^1H-^{15}N$ NOE spectra were recorded in an interleaved manner with a 6 s ^1H presaturation period and a control experiment in which the presaturation period was replaced by an equivalent delay without saturation and with spectral widths of 11.98 ppm (2048 points) and 22.73 ppm (64 points) in the ^1H and ^{15}N dimensions, respectively. The ^{15}N R_1 and R_2 values were determined using Analysis version 2.1.5 via a nonlinear least-squares fit of intensity decay as a function of relaxation period. The $^1H-^{15}N$ NOE was calculated as the ratio of the saturated versus unsaturated peak intensities. The residue-specific dynamics was also estimated via the Random Coil Indexing approach.

Structure Calculation. Distances were derived from NOE peak heights in the $^1H-^1H$ NOESY spectrum with a 200 ms mixing time and calibrated using isolated spin pair approximation in ARIA 2.2.[29] Hydrogen bond upper limit restraints for residues showing slow peak intensity buildup in the SOFAST-HMQC H–D exchange experiment were set to 2.2 Å for the HN(i)–O($i+4$) linkage and to 3.3 Å for the N(i)–O($i+4$) linkage. A patch was written into the topology file to define the O-linkage from Ser18 and the S-linkage from Cys43 to *N*-acetylglucosamine, and an otherwise standard CNS 1.1-based protocol was employed using the ARIA 2.2 interface. Structure visualization utilized PYMOL,[30] and structure evaluation was performed with PROCHECK.[31]

RESULTS

Peptide Chemical Shift-Based Analyses. A $^1H-^{15}N$ HSQC spectrum of GccF recorded at natural abundance produced 43 peaks in the backbone region (Figure 1), two more than the expected number for a 43-residue peptide containing one proline residue (the N-terminal amino group is not typically observed). This number is consistent with the existence of two N-containing post-translational modifications. These modifications were previously shown by mass spectrometry to be consistent with the attachment of HexNAcs to Ser18 and Cys43[4] and are discussed further below. Backbone resonance assignments (^1H_N, $^1H_\alpha$, $^{13}C_\alpha$, and ^{15}N) are 100% complete, with the exception of the N-terminal ammonium group. Side chain ^1H assignments are 83% complete, with most of the unassigned resonances due to overlap in the aromatic resonances. The secondary structure was probed via chemical shift indexing (CSI)[17] as presented in Figure 2A. Positive CSI values for residues 7–10 and 22–27 are consistent with helix formation, with residues 6 and 11 showing partial helical character. Residues 5–12 and 21–28 show medium- and long-range NOEs consistent with formation of two helices (Figure 2B), although the 6–10 and 7–11 Hα–HN(i, $i+4$) connectivities could not be assigned because of spectral overlap. Deuterium exchange results, discussed further below, are also consistent with stable helix formation in the region of residues 5–12 and 21–28. The significant helical content of the peptide is corroborated by CD spectroscopy under similar solvent conditions as shown by our laboratory previously.[4]

The large negative CSI values observed for Cys5, Cys12, Cys21, and Cys28 (Figure 2A) are consistent with the previously established disulfide bonding of these residues.[4] The large negative CSI values observed for residues Ser18 and Cys43 (labeled with asterisks in Figure 2A) are consistent with the previously established post-translational modification of these residues. The C-terminal tail (residues 30–43) lacks characteristic secondary structural and long-range NOEs (Figure 2B). Analysis of backbone chemical shifts via the random coil index procedure (RCI)[18] shows the presence of less internal dynamics in the helices than in the other regions of the peptide (Figure 2C).

Sugar Identification and Linkage. To simplify the chemical shift assignment of the two post-translational modifications, the tryptic peptides $GccF_{1-32}$ (amino acids 1–32) and $GccF_{33-43}$ (amino acids 33–43), each of which contains only one modification, were produced as described previously[4] and analyzed separately by NMR. The resonances of each modification were assigned using a combination of DQF-COSY, $^1H-^{13}C$ HSQC, $^1H-^{13}C$ H2BC,[19] and $^1H-^{13}C$ HMBC experiments. The identification of each as β-GlcNAc residues was first established tentatively by comparison of the assigned chemical shifts with those in the SUGABASE database (http://www.boc.chem.uu.nl/sugabase/sugabase.html). The stereochemistry was

Table 1. NMR Parameters for the Two GlcNAc Moieties of Glycocin F

	$GccF_{1-32}$ (contains Ser18 O-GlcNAc)			$GccF_{33-43}$ (contains Cys43 S-GlcNAc)		
	$\delta(^1H)$ (ppm)[a]	$\delta(^{13}C)$ (ppm)[a]	J (Hz)	$\delta(^1H)$ (ppm)[a]	$\delta(^{13}C)$ (ppm)[a]	J (Hz)
1	4.56	102.00	$J_{1,2}$ = 8.80	4.59	84.81	$J_{1,2}$ = 10.17
2	3.72	56.18	$J_{2,3}$ = 8.85	3.77	55.36	$J_{2,3}$ = 10.17
3	3.56	74.62	$J_{3,4}$ = 9.50	3.53	75.95	$J_{3,4}$ = 9.10
4	3.42	70.93		3.43	70.71	
5	3.47	76.85		3.43	80.92	$J_{5,6}$ = 5.23
6	3.74	61.77	$J_{6,6'}$ = 12.29	3.71	62.04	$J_{6,6'}$ = 12.10
6′	3.94	–		3.89	–	
Me	2.08	23.35		1.99	23.07	
C=O	–	174.80		–	174.4	

[a] Chemical shifts are referenced relative to acetone at a $\delta(^1H)$ of 2.225 ppm and a $\delta(^{13}C)$ of 31.08 ppm.

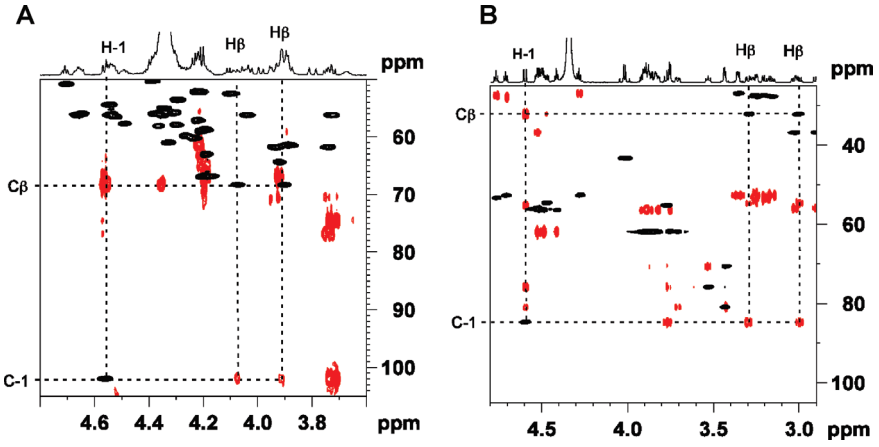


Figure 3. Sugar-containing region of $^1H-^{13}C$ HSQC (black) and $^1H-^{13}C$ HMBC (red) spectra for (A) the $GccF_{1-32}$ peptide showing connectivities between GlcNAc and Ser18 and (B) the $GccF_{33-43}$ peptide showing connectivities between GlcNAc and Cys43.

subsequently determined by analysis of the $^3J_{H,H}$ coupling constants extracted from 1D selective TOCSY experiments. The GlcNAc attached to $GccF_{33-43}$ exhibited two unusual NMR parameters: the C-1 chemical shift was anomalously low (84.81 ppm), and the $^3J_{H1,H2}$ coupling was large (10.17 Hz), compared to the more typical values found for the $GccF_{1-32}$ sugar [102.00 ppm and 8.80 Hz, respectively (see Table 1)]. These atypical data can be attributed to the sugar being linked to the peptide via sulfur rather than via a more electronegative oxygen atom. Estimation of the C-1 chemical shift assuming an S-linkage using ACD/CNMR Predictor 8.04 (Advanced Chemistry Developments Inc., Toronto, ON) produced a result very similar to that observed experimentally (84.34 ± 0.75 ppm estimated vs 84.81 ppm observed). The linkage to Ser18 was confirmed as GlcNAc(β-1-O)Ser18 via HMBC correlations from GlcNAc H-1 to the Ser18 C_β resonance and from GlcNAc C-1 to the Ser18 H_β resonance (Figure 3A). Analogous HMBC correlations from the second GlcNAc to Cys43 (Figure 3B) similarly confirm this linkage as GlcNAc(β-1-S)Cys43. These findings are in agreement with previously reported analyses in which Ser18 was specifically deglycosylated by *N*-acetyl-β-D-glucosaminidase.[4]

Three-Dimensional Structure. Initial structural calculations were performed using distance constraints derived from 443 unambiguous $^1H-^1H$ NOEs, without the use of disulfide-, hydrogen bond-, secondary structure-, or deuterium exchange-based constraints. GlcNAc moieties were computationally attached to Ser18 and Cys43. These initial structures showed two helical regions from residues 5−12 and 21−28. Consistent with the chemical shift-based analyses described above and previous reports, the initial structures placed the sulfide groups of Cys5 and Cys12 sufficiently close to Cys28 and Cys21, respectively, to suggest the formation of two disulfide bonds. The presence of interhelical NOEs between residues 5 and 28, residues 8 and 24, residues 8 and 25, and residues 9 and 20 is consistent with interactions predicted from a helical wheel model (Figure 4A) with residues 5−12 and 21−28 forming a pair of antiparallel helices.

To gain experimental evidence of the presence of hydrogen bonding, we performed H−D exchange experiments on a sample that was solvated in a D_2O/acetonitrile/acetic acid mixed solvent system, lyophilized, and redissolved in an H_2O/acetonitrile/acetic acid mixture. A series of sequentially recorded $^1H-^{15}N$ SOFAST-HMQC[20] experiments were performed to allow rapid examination of heteronuclear correlations arising when amide-bound deuterium exchanged with solvent protons. The resulting peak intensities plotted as a function of time showed slow build-up for residues Thr8, Leu9, Ala10, Met11, Met24, and Tyr25, all from the putative helical regions. The Cys12 amide peak could not be analyzed because of peak overlap, while residues 26−28 from the tail end of helix 2 exchange rapidly. The observed results are consistent with the residues from the first helix and the

dx.doi.org/10.1021/bi200217u |*Biochemistry* 2011, 50, 2748–2755

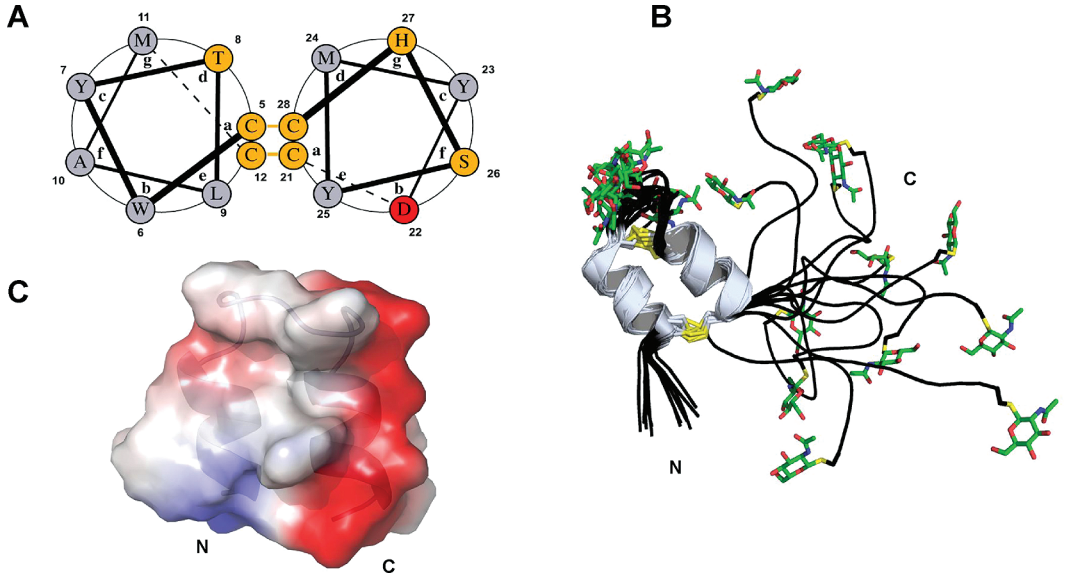


Figure 4. Glycocin F structure. (A) Helical wheel diagram (http://www.gevorggrigoryan.com/drawcoil/) of residues 5−12 and 21−28. The Cys5−Cys28 and Cys12−Cys21 disulfide bonds are represented by yellow lines. Polar residues are colored yellow and hydrophobic residues gray, and the lone charged residue in this region is colored red. The N- and C-termini would project above the plane of the page and the loop connecting the two helices below the plane of the page. (B) Superposition of the 12 calculated structures with the lowest energy, with the N- and C-termini labeled. All sulfur atoms are shown as yellow sticks, including the two disulfide bonds between the helices and the S-linked sugar at the C-terminus. The sugars are also shown as sticks. (C) Surface rendering of residues 4−28, colored according to electrostatic potential: red for negative, blue for positive, and white for neutral. A cartoon rendering, with the first and second helices labeled N and C, respectively, is visible under the partially transparent surface to assist with orientation.

Table 2. NMR Structural Statistics for Glycocin F

distance constraints	
NOE	443
hydrogen bond	12
disulfide bridge (Cys5−Cys28, Cys12−Cys21)	2
distance constraint violations of >0.3 Å	0
NOE r.m.s.d. (Å)	0.014 ± 0.004
deviations from idealized geometry	
bond lengths (Å)	0.0014 ± 0.0003
bond angles (°)	0.62 ± 0.44
improper angles (°)	0.60 ± 0.62
average pairwise r.m.s.d. (Å)	
heavy atoms, residues 5−12, 21−28	1.67 ± 0.40
backbone, residues 5−12, 21−28	0.65 ± 0.24
heavy atoms, all residues 1−43	7.57 ± 2.07
backbone, all residues 1−43	6.44 ± 1.95
Procheck-calculated Ramachandran plot analysis	
residues in most favoured regions	57.1%
residues in additionally allowed regions	37.9%
residues in generously allowed regions	4.0%
residues in disallowed allowed regions	1.0%

N-terminal half of the second helix participating in stable hydrogen bonds.

At this point, conservative hydrogen bond constraints, where indicated by the exchange data given above, were introduced into the α-helical region as upper limit distance constraints of 2.2 Å for the HN(i)−O($i+4$) connectivity and 3.3 Å for the N(i)−O($i+4$) connectivity, with the pattern of hydrogen bonding inferred from the initial calculated structures. Although hydrogen bonds were not introduced for residues 26−28 because of their exchange characteristics, the initial structures also placed these residues into an α-helical conformation. This suggests that the second helix forms but is somewhat unstable near its C-terminus. In addition, disulfide bonds were inserted via standard topological constraints in the form of covalent linkages. A new set of structures was then calculated. The resulting final structures (Figure 4B) contain an N-terminal α-helix of residues 5−12, a loop of residues 13−20, and a second α-helix of residues 21−28. The C-terminal residues 29−43 appear largely disordered. Each helix is ∼10.6 Å long, measured as the $C_\alpha-C_\alpha$ distance from residue 5 to residue 12 and from residue 21 to residue 28, incorporating two turns for each helix. The helical interface is bounded at either end by the Cys5−Cys28 and Cys12−Cys21 disulfide bonds. The center of the helical interface includes relatively close contacts from Thr8 and Leu9 to Met24 and Tyr25. A superposition of the 12 best structures shows the helices to be well-defined [backbone rmsd of 0.65 ± 0.24 Å (see Table 2)] relative to the intervening loop and the disordered tail. A Ramachandran analysis for the ensemble shows that the residues in the structured region fall within the standard α-helical region.

No NOEs were observed between the carbohydrate moieties and the peptide component of the molecule, likely because of flexibility within the sugars and the O- and S-linkages together with their presence in less structured regions of the peptide (see below).

Dynamics. The ^{15}N R_1 and R_2 profiles show a pattern of higher values for residues 5−12 and 21−28 (Figure 5A,B), indicating a more rigid conformation relative to residues 13−20 that produce lower relaxation rates. A steady decrease in R_1 and R_2 is observed from residue 29 to 34, and the rates remain low from this point to the C-terminus, indicating a largely flexible tail and explaining the

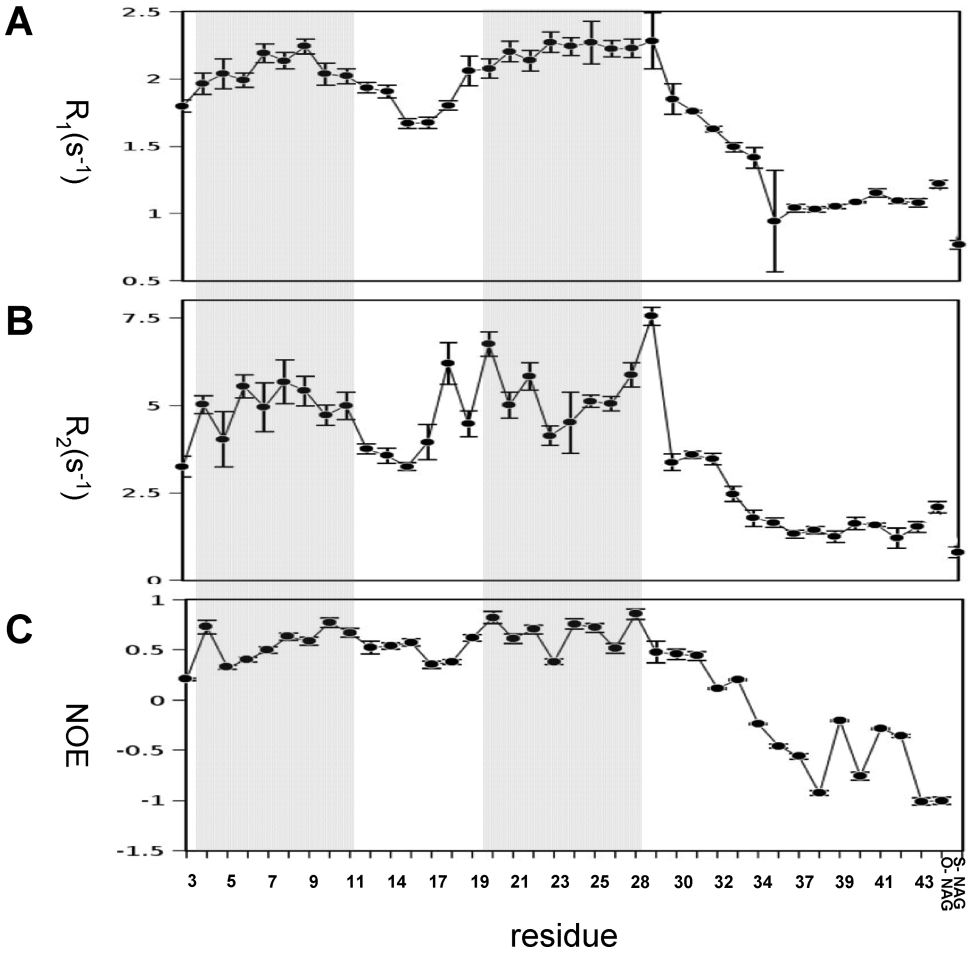


Figure 5. Plot of ^{15}N relaxation parameters for glycocin F. Consistent with the RCI result in Figure 2, relaxation parameters for the two helices (shaded regions) display less internal dynamics than the remainder of the molecule. Relaxation data from the sugar moieties, labeled O-NAG and S-NAG, show that they are highly mobile.

lack of long- and medium-range ^{1}H−^{1}H NOEs for the C-terminal tail. The ^{1}H−^{15}N NOE profile (Figure 5C) follows a trend similar to that of the ^{15}N R_1 and R_2 data, with larger NOEs indicating more rigidity in the helical regions. Residues 12, 16, and 36 were excluded from relaxation analysis because of peak overlap. These results are consistent with the random coil index-inferred order parameter (S^2) analysis presented in Figure 2C. The *N*-acetylglucosamine ^{15}N nuclei (labeled O-NAG and S-NAG in Figure 5) display low R_1 and R_2 values, and the O-NAG ^{1}H−^{15}N NOE value is low, consistent with the high flexibility of the carbohydrate moieties as suggested above by the lack of NOEs between the sugars and the peptide component. The S-NAG ^{1}H−^{15}N NOE could not be accurately determined because of low signal intensity.

Accession Numbers. The accession numbers of the atomic coordinates of the ensemble of 12 conformers representing the NMR structure have been deposited in the Brookhaven Protein Data Bank (entry 2KUY). The chemical shift list has been deposited in the BioMagRes Data Bank under accession code 16747.

DISCUSSION

The glycocin F structure consists of two α-helices connected by a short loop and held together by two nested disulfide bonds (Figure 4B). The two-helix bundle is preceded by a four-residue N-terminus and followed by a 15-residue C-terminal tail. Relaxation and RCI analyses (Figure 5) show that the helical regions are somewhat less dynamic than the intervening loop and that the C-terminal tail is highly disordered. Independent and conclusive NMR-based analyses of the post-translational modifications confirm the previous report[4] that Ser18 is linked to a β-GlcNAc, and that the linkage is through O_γ. Furthermore, our analyses show that the previously unidentified HexNAc attached to S_γ of Cys43 is also a β-GlcNAc. The latter modification is completely novel, though genomic analyses suggest that a significant number of related peptides may also contain S-glycosylation sites.[3,4] NMR-based structures of three bacteriocins that each contain a single disulfide bond have been reported,[21−23] and a nested pair of disulfide bonds have been proposed for sublancin 168,[2] suggesting a similarity of mechanism with GccF. However, this work represents the first detailed structural analysis of a bacteriocin with nested disulfide bonds, and the first high-resolution structure of any S-glycosylated protein.

Direct relaxation data from the two GccF sugar moieties show that they are highly mobile. The C-terminal helix and the serine-rich C-terminal tail are highly polar, while the N-terminal half of the molecule (N-terminus, N-terminal helix, and interhelical loop) is of mixed polar and nonpolar character. The N-terminal helix is amphipathic, presenting a largely hydrophobic surface to

solvent, and a more polar surface to the interface with the polar C-terminal helix. As a result, the surface of the two-helix bundle is amphipathic (Figure 4C). This amphipathicity, along with the bacteriostatic activity of GccF and its enhanced solubility in a mixed solvent system relative to water, suggests that glycocin F may function at least in part through association with the cellular envelope of the target bacterium.

Although the antimicrobial mechanism(s) of glycocin F is not yet well understood, it is known that the O-linked GlcNAc is essential for the bacteriostatic activity, while the S-glycosylated C-terminus significantly enhances bacteriostasis.[4] Furthermore, addition of free GlcNAc to the medium reverses GccF-mediated bacteriostasis.[4] The linked sugars may help to modify the physical properties of the peptide, including its solubility and self-association state. They may also help to protect the peptide from proteolysis at the otherwise exposed interhelical loop and C-terminus.

However, the sugars likely also play a more specific role in the as yet unidentified pathway(s) through which GccF effects bacteriostasis. GccF at high concentrations in the extracellular medium can initiate bacteriostasis within minutes (unpublished results), suggesting that the specific target of GccF may reside at the cell wall. Furthermore, GlcNAcs are common substrates for cell wall synthesis, where they are incorporated by bacterial phosphotransferases after transport across the cell membrane. Thus, cell wall remodelling would seem a likely mechanism for GccF activity. It is clear from the three-dimensional structure of GccF that the C-terminal tail is highly soluble and conformationally mobile. This mobility may allow the C-terminal S-linked sugar to serve as bait on a fishing line of sorts, hooking the peptide to the cell wall. The S-linkage would provide resistance from glycosidases and transglycosylases. Once so anchored, the amphipathic helical bundle would more efficiently scan the cell wall, where the O-linked GlcNAc could help it to associate with and inhibit key transglycosylases required for peptidoglycan synthesis. The flexible tail provides an approximate separation between the S-linked and O-linked sugars (C-1−C-1 distance) of 34 ± 8 Å as measured from the structural ensemble. This distance would allow a cell wall surface area of approximately 3500 Å^2 to be sampled by the O-linked sugar and the attached two-helix bundle, once the S-linked sugar is anchored. According to this proposed mechanism, free GlcNAc would rescue the cell by competing with GccF for the GlcNAc interaction sites on or within the cell wall. The reduction but not total loss of activity observed upon tryptic removal of the C-terminus of GccF is consistent with the S-linked sugar helping to localize the key active part of the peptide, the O-linked sugar, and the associated two-helix bundle. Additional studies identifying specific molecular receptors and targets of inhibitory action should help to clarify these issues.

AUTHOR INFORMATION

Corresponding Author

*Phone: +646-356-9099. Fax: +646-350-5682. E-mail: s.pascal@massey.ac.nz (S.M.P.) or h.venugopal@massey.ac.nz (H.V.).

Funding Sources

This work was supported in part by an RSNZ Marsden grant to S.M.P. and an RSNZ Marsden grant to G.E.N. and M.L.P.

ABBREVIATIONS

GccF, glycocin F; GccF$_{1-32}$, glycocin F fragment containing residues 1−32; GccF$_{33-43}$, glycocin F fragment containing residues 33−43; GlcNAc, *N*-acetylglucosamine; HexNAc, *N*-acetyl six-carbon sugar; O-NAG, O-linked *N*-acetylglucosamine; S-NAG, S-linked *N*-acetylglucosamine; rmsd, root-mean-square deviation.

REFERENCES

(1) Löte, C. J., and Weiss, J. B. (1971) Identification of digalactosylcysteine in a glycopeptide isolated from urine by a new preparative technique. *FEBS Lett. 16*, 81–85.

(2) Weiss, J. B., Löte, C. J., and Bobinski, H. (1971) New low molecular weight glycopeptide containing triglucosylcysteine in human erythrocyte membrane. *Nat. New Biol. 234*, 25–26.

(3) Oman, T. J., Boettcher, J. M., Wang, H., Okalibe, X. N., and van der Donk, W. A. (2011) Sublancin is not a lantibiotic but an *S*-linked glycopeptide. *Nat. Chem. Biol. 7* (2), 78–80.

(4) Stepper, J., Shastri, S., Loo, T. S., Preston, J. C., Novak, P., Man, P., Moore, C. H., Havlíček, V., Patchett, M. L., and Norris, G. E. (2011) Cysteine *S*-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Lett. 585*, 645–650.

(5) Baran, E., and Drabarek, S. (1978) *Pol. J. Chem. 52*, 941–946.

(6) Jahn, M., Marles, J., Warren, R. A., and Withers, S. G. (2003) Thioglycoligases: Mutant Glycosidases for Thioglycoside Synthesis. *Angew. Chem., Int. Ed. 42*, 352–354.

(7) Driguez, H. (2001) Thiooligosaccharides as Tools for Structural Biology. *ChemBioChem 2*, 311–318.

(8) Cohen, S. B., and Halcomb, R. L. (2002) Application of Serine- and Threonine-Derived Cyclic Sulfamidates for the Preparation of *S*-Linked Glycosyl Amino Acids in Solution- and Solid-Phase Peptide Synthesis. *J. Am. Chem. Soc. 124*, 2534–2543.

(9) Gerz, M., Matter, H., and Kessler, H. (1993) *S*-Glycosylated Cyclic Peptides. *Angew. Chem., Int. Ed. 32*, 269–271.

(10) Galonić, D. P., van der Donk, W. A., and Gin, D. Y. (2004) Site-Selective Conjugation of Thiols with Aziridine-2-Carboxylic Acid-Containing Peptides. *J. Am. Chem. Soc. 126*, 12712–12713.

(11) Galonić, D. P., Ide, N. D., van der Donk, W. A., and Gin, D. Y. (2005) Aziridine-2-carboxylic Acid-Containing Peptides: Application to Solution- and Solid-Phase Convergent Site-Selective Peptide Modification. *J. Am. Chem. Soc. 127*, 7359–7369.

(12) Hili, R., Rai, V., and Yudin, A. K. (2010) Macrocyclization of Linear Peptides Enabled by Amphoteric Molecules. *J. Am. Chem. Soc. 132*, 2889–2891.

(13) Gillor, O., Etzion, A., and Riley, M. A. (2008) The dual role of bacteriocins as anti- and probiotics. *Appl. Microbiol. Biotechnol. 81*, 591–606.

(14) Corr, S. C., Li, Y., Riedel, C. U., O'Toole, P. W., Hill, C., and Gahan, C. G. (2007) Bacteriocin production as a mechanism for the antiinfective activity of *Lactobacillus salivarius* UCC118. *Proc. Natl. Acad. Sci. U.S.A. 104*, 7617–7621.

(15) Klaenhammer, T. R. (1993) Genetics of bacteriocins produced by lactic acid bacteria. *FEMS Microbiol. Rev. 12*, 39–85.

(16) Cotter, P. D., Hill, C., and Ross, R. P. (2005) Bacteriocins: Developing innate immunity for food. *Nat. Rev. Microbiol. 3*, 777–788.

(17) Wishar, D. S., and Sykes, B. D. (1994) The ^{13}C chemical-shift index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR 4*, 171–180.

(18) Berjanskii, M. V., and Wishart, D. S. (2007) The RCI server: Rapid and accurate calculation of protein flexibility using chemical shifts. *Nucleic Acids Res. 35*, W531–W537.

(19) Nyberg, N. T., Duus, J. O., and Sørensen, O. W. (2005) Heteronuclear two-bond correlation: Suppressing heteronuclear three-bond or higher NMR correlations while enhancing two-bond correlations even for vanishing $^2J_{CH}$. *J. Am. Chem. Soc. 127*, 6154–6155.

(20) Schanda, P., Kupce, E., and Brutscher, B. (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J. Biomol. NMR 33*, 199–211.

(21) Fregeau Gallagher, N. L., Sailer, M., Niemczura, W. P., Nakashima, T. T., Stiles, M. E., and Vederas, J. C. (1997) Three-dimensional structure of leucocin A in trifluoroethanol and dodecylphosphocholine micelles: Spatial location of residues critical for biological activity in type IIa bacteriocins from lactic acid bacteria. *Biochemistry 36*, 15062–15072.

(22) Uteng, M., Huge, H. H., Markwick, P. R., Fimland, G., Mantzilas, D., Nissen-Meyer, J., and Muhle-Goll, C. (2003) Three-dimensional structure in lipid micelles of the pediocin-like antimicrobial peptide sakacin P and a sakacin P variant that is structurally stabilized by an inserted C-terminal disulfide bridge. *Biochemistry 42*, 11417–11426.

(23) Haugen, H. S., Fimland, G., Nissen-Meyer, J., and Kristiansen, P. E. (2005) Three-dimensional structure in lipid micelles of the pediocin-like antimicrobial peptide curvacin A. *Biochemistry 44*, 16149–16157.

(24) Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins 59*, 687–696.

(25) Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *J. Am. Chem. Soc. 111*, 1515–1517.

(26) Fesik, S. W., and Zuiderweg, E. R. (1990) Heteronuclear three-dimensional NMR spectroscopy of isotopically labelled biological macromolecules. *Q. Rev. Biophys. 23*, 97–131.

(27) Bax, A., and Grzesiek, S. (1993) Methodological advances in protein NMR. *Acc. Chem. Res. 26*, 131–138.

(28) Fairman, R., Beran-Steed, R. K., and Handel, T. M. (1997) Heteronuclear (^{1}H, ^{13}C, ^{15}N) NMR assignments and secondary structure of the basic region-helix-loop-helix domain of E47. *Protein Sci. 6*, 175–184.

(29) Habeck, M., Rieping, W., Linge, J. P., and Nilges, M. (2004) NOE assignment with ARIA 2.0: The nuts and bolts. *Methods Mol. Biol. 278*, 379–402.

(30) DeLano, W. L. (2008) *The PyMOL Molecular Graphics System*, DeLano Scientific LLC, Palo Alto, CA.

(31) Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR 8* (4), 477–486.