Identification of *Streptococcus gallolyticus* subsp. *gallolyticus* (biotype I) competence stimulating peptide pheromone

Anthony Harrington, Yftah Tal-Gan*

Department of Chemistry, University of Nevada, Reno

*Address correspondence to Yftah Tal-Gan, ytalgan@unr.edu.

Supporting information Table of Contents

Additional experimental details

Polymerase chain reaction (PCR) amplification of RNA polymerase B and *comC* gene. A modified whole cell polymerase chain reaction (PCR) direct amplification was performed to obtain template DNA for PCR amplification (1). A total volume of 1.5 mL of pure culture broth was placed in a sterile 1.5 mL microfuge tube. The cells were pelleted by centrifugation at 6000 x g for 1 min. The supernatants were discarded and 0.5 mL of sterile phosphate buffered saline (PBS) with the following composition: 8 g L^{-1} NaCl, 0.2 g L^{-1} KCl, 1.44 g L^{-1} Na_2HPO_4, and 0.24 g L^{-1} KH_2PO_4, with a final pH between 7.2-7.4, was added to the bacterial cell pellet. The pellet was resuspended using 5 sec pulses with a vortex mixer. The cells were pelleted again using centrifugation. The supernatants were discarded, and the process was repeated once more. After the second PBS wash, 0.5 mL of sterile dH_2O was added to the bacterial pellet and it was resuspended by pulse vortex mixing. The centrifugation was repeated once more. The supernatants were discarded and 200-250 µL of sterile ddH_2O was added to the tube. The pellet was again resuspended by vortex mixing, incubated in a standard heat block set at 95° C for 5 min, and immediately placed in a -80° C freezer for 10 min. After 10 min, the tubes were left at room temperature to thaw completely and gently vortexed for 5 sec followed by centrifugation at 9,500 x g for 60 sec. The supernatants were used as template DNA for PCR amplification. The following primers were obtained from Integrated DNA Technologies, StreptoF [5'-AARYTIGGMCCTGAAGAAAT-3', Tm = 52.9° C](2), StreptoR [5'-TGIARTTTRTCATCAACCATGTG-3', Tm = 53.3° C](2), boviscsp Fwd [5'-TGCTACAATTAATCACTAAGGAGGT-3', Tm = 53.8° C], and boviscsp Rev [5'-

TGCGATTGAAGCGGCTATT-3', Tm = 55.0° C]. The *rpoB* gene amplification PCR mixture contained: 25 µL of 2X Hot Start Taq Master Mix (VWR), 1 µM of StreptoR primer, 1uM of StreptoF primer, 5 µl of template DNA, and 10 µL of nuclease-free water (VWR) to make a final volume of 50 µL. The *comC* & partial *comD* gene amplification PCR mixture contained: 25 µL of 2X Hot Start Taq Master Mix (VWR), 1 µM of boviscsp FWD primer, 1uM of boviscsp Rev primer, 5 µl of template DNA, and 10 µL of nuclease-free water (VWR) to make a final volume of 50 µL. The PCR amplification was performed in an Eppendorf Mastercycler gradient 5331 PCR machine. The following conditions were used for amplification of the partial *rpoB* gene to confirm the identity of various streptococci isolates to the species level: initial melt cycle (94° C for 3 min) followed by 30 cycles of melting (94° C for 30 sec), annealing (50° C for 45 sec), extension (72° C for 60 sec), and a final extension (72° C for 5 min). The following conditions werer used for amplification of the *comC* & partial *comD* gene specifically found in *Sgg*: initial melt cycle (94° C for 3 min) followed by 30 cycles of melting (94° C for 30 sec), annealing (50° C for 45 sec), extension (72° C for 60 sec), and a final extension (72° C for 5 min). A total volume of 10 µL of each PCR reaction mixture was run on a 1% agarose gel containing 1X GelRed™ (PhenixResearch) along with 5 µL of 1kb DNA ladder (Promega) to verify proper amplification. Running conditions were 90 V for 1.5-2.0 h. The gel was visualized in a UVP Gel-Doc Imager using 302 nm excitation wavelength and amplified products were confirmed if they had a molecular weight around 780 bp for the partial *rpoB* gene and 490 for the *comC* & partial *comD* gene. Amplified PCR products were purified using an E.Z.N.A Cycle Pure Kit (Omega Bio-Tek) according to the manufacturer's instructions and collected into labeled, sterile 1.5 mL

microfuge tubes. The purified gene products were quantified for their purity and concentration using a NanoDrop™ spectrophotometer (GE Healthcare Bio-Sciences). PCR amplified gene products with an A_{260}/A_{280} ratio between 1.8 - 2.0 and a DNA concentration above 15 µg/µL were used for sequencing.

Sequencing and verification of PCR amplified *rpoB* and *comC* genes. PCR amplified gene products were submitted to the Nevada Genomics Center (NGC) located on the University of Nevada, Reno campus for Sanger sequencing using an ABI Prism 3730 DNA Analyzer (3). Samples were submitted according to the NGC guidelines (less than 30 ng of amplified DNA containing 2 µM of primer in a volume less than 25 µL. Sequencing data were obtained using the same primers used for PCR amplification. The NGC analyzes the sequencing data with Phred and provides trimmed reads containing bases with Q > 20 in addition to raw sequence data (4). The trimmed sequencing data provided by NGC were saved as fasta files and contigs of each gene were generated in MEGA 7.0 using the forward and reverse reads for each gene. The contigs were checked against the National Center of Biotechnology Information database using Blastn to verify species identity and the *comC* gene. The *comC &* partial *comD* contig contained non-coding DNA bases upstream of *comC* gene and part of the 3' coding strand region of the *comD* gene based on the designed primers. The non-coding DNA and 3' coding strand region of the *comD* gene were removed from the contig to reveal the correct full length *comC* gene DNA sequence. This nucleotide sequence was deposited in GenBank with the following accession number: MF964227.

Plasmid DNA Extraction. Plasmid pALH122 was provided transformed into *Escherichia coli* XL10 gold competent cells by Dr. Dennis G. Cvitkovitch from the University of Toronto. A volume of 20 µL from a frozen stock culture was inoculated into 5 mL of sterile THY and incubated at 37° C with constant shaking for 24 h. A volume of 1.5 mL was transferred to a sterile 1.5 mL microfuge tube and centrifuged at 6,000 x g for 1 min. The supernatant was discarded, and another 1.5 mL of overnight *E. coli* was added to the microfuge tube. The whole process was repeated until all 5 mL of *E. coli* culture was pelleted. Plasmid DNA was extracted from the pelleted cells using a Qiagen MiniPrep kit according to the manufacturer specifications (Qiagen Inc.). Plasmid DNA was eluted from the spin-column using 50 µL of nuclease-free water (VWR) and quantified for purity and concentration using a NanoDrop™ spectrophotometer (GE Healthcare Bio-Sciences). Isolated plasmid DNA was stored at -20° C until used for transformation assays.

Peptide Purification. Crude peptides were purified with RP-HPLC. Standard RP-HPLC conditions were as follows: flow rates = 5 mL min^{-1} for semipreparative separations and 1 mL min^{-1} for analytical separations; mobile phase A = ddH_2O + 0.1% TFA; mobile phase B = ACN + 0.1% TFA. Purities were determined by integration of peaks with UV detection at 220 nm. Preparative HPLC methods were used to separate the crude peptide mixture to different chemical components using the following linear gradient: 5% B → 15% B over 10 min, followed by 15% B → 45% B over 23 min, and 45% B → 95% B over 13 minutes. MALDI-TOF MS was used to identify the fractions containing the desired peptides. The purity of the peptides was quantified using an analytical HPLC method with the following linear gradient: 5% B → 95% B over 27 min.

Only peptide fractions that were purified to homogeneity (>95%) were used for further characterization. The identity of the isolated peptides (naturally isolated or synthesized) was confirmed by ESI+ HRMS comparing the observed mass-to-charge (m/z) ratio to the expected m/z ratio.

Loading of Resin for Solid Phase Peptide Synthesis. The synthetic CSP was constructed using standard 9-fluorenylmethyloxycarbonyl (Fmoc)-based solid-phase peptide synthesis (SPPS) procedures on Cl-MPA ProTide™ Resin. Synthesis was conducted on a 0.02 mmol scale. Loading of the resin was performed on a discover microwave (CEM Corp.) using 5 mL of 0.04 mM Fmoc-L-Ala-OH * $2H_2O$ dissolved in dimethylformamide (DMF) with 2 mL of 1 M diisopropylethylamine (DIPEA) and 0.125 M KI solution in DMF. The resin was loaded using a 2-stage microwave method with the 1st stage using 75 W to reach a temperature of 80° C for 1 min and the second stage using 20 W to reach and maintain a temperature of 90° C for 9 min. The resin was drained and washed twice with DMF followed by dichloromethane (DCM). The loading of the resin was assumed to be 0.16 mmol/gram based on the manufacturer's specifications.

Peptide synthesis. The following conditions were used for coupling and deprotection cycles in the Liberty1 automated peptide synthesizer (CEM Corp.). Deprotection of the Fmoc group was performed first using 5 mL of 2% piperidine with 2% 1, 8-Diazabicyclo [5.4.0] undec-7-ene (DBU) in DMF (90 sec, 75° C) followed by another 5 mL of 2% piperidine with 2% DBU in DMF (90 sec, 75° C). The resin was washed with DMF (3 x 5 mL) after each deprotection cycle. Coupling reactions were performed using 2.5 mL solution containing Fmoc-protected amino acid (5 equiv.) *N*, *N*′-Diisopropylcarbodiimide

(5 equiv.) and Ethyl cyano (hydroxyimino) acetate (5 equiv.). All amino acids were coupled for 20 min (30 W, 90°C), except His. His was coupled for 10 min (0 W, 25°C) then for 40 min (20 W, 50° C). After the synthesis was completed the resin was washed with DMF (3 x 5 mL).

Cleavage from the resin. Following synthesis, the resin was washed with diethyl ether (2 mL) and then kept in open air for 15 min before it was transferred into a 15 mL falcon tube. The peptide was cleaved with 3 mL cleavage cocktail consisting of 95% TFA, 2.5% triisopropylsilane (TIPS), and 2.5% dH_2O for 3 h with agitation. The cleaved peptide was separated from the resin by filtration, transferred to a 50 mL falcon tube and precipitated with a cold solution of diethyl ether:hexane (1:1, 45 mL, 0 °C). The precipitated crude peptide was centrifuged for 5 min at 4,600 x g, redissolved in 10 mL ACN:ddH_2O (1:1) and lyophilized before HPLC purification.

Crystal violet biofilm quantification. A 1:10 dilution of overnight culture was made in 10 mL of sterile THB containing 1% glucose in a 50 mL tube and incubated aerobically for 30 min at 37° C with shaking at 50 rpm. A volume of 10 µL was added from a 1 mM stock solution of synthetic CSP in DMSO to a well in a clear, flat-bottom, 96-well polystyrene microtiter plate (Costar). Serial dilutions of synthetic CSP were used to afford three final concentrations (10 µM, 1 µM and 100 nM). A volume of 2 µL CSP in DMSO was added to an individual well with DMSO containing no CSP used as a negative control. This process was performed in triplicate. The wells were then filled with 198 uL of the fresh culture and mixed by pipetting. Three individual wells received 200 µL of sterile THY as a “no-bacteria” negative control used for non-specific crystal violet staining. The 96-well plate was incubated statically for 24 h in a CO_2 incubator

(37° C with 5% CO_2). After the 24 h incubation, OD_{600} values were measured for each well prior to decanting the liquid culture. The wells were washed with PBS (3 x 200 µL) gently prior to being placed in an oven (50° C, 1 h) to fix any biofilm bacteria. A volume of 200 µL of crystal violet (0.1% w/v) was added to each well, incubated for 5 min, and decanted. Each well was washed (3 x 200 µL) with ddH_2O and treated with 30% acetic acid for 15 minutes with gentle agitation. The 595 nm absorbance of each well was recorded on a plate reader. The normalized biofilm values were calculated using the following formula:

$$\frac{[595\ \text{absorbance} - \text{average sterile THY 595 absorbance}]}{\text{OD600}}$$

Circular Dichroism (CD) Spectroscopy. CD spectra were recorded with an Aviv Biomedical CD spectrometer (model 202−01). All the measurements were performed using 400 µM peptide concentration in PBS buffer (137 mM NaCl, 2.7 mM KCl,10 mM Na_2HPO_4, 1.8 mM KH_2PO_4; pH was adjusted to 7.4) containing either 0% or 20% trifluoroethanol (TFE) at 25° C with a quartz cuvette (Science Outlet) with a path length of 0.1 cm. Spectra were obtained by scanning samples one time at 3 nm min^{-1} with a bandwidth of 1 nm and a response time of 20 sec over a wavelength range (195 to 260 nm). Percent helicity (f_H) was calculated for peptides that displayed a significant helical pattern using the following equation:

$$\frac{[\Theta]_{222}}{[\Theta_\infty]_{222}\,(1-\frac{x}{n})}$$

$[\Theta]_{222}$ is the mean residue ellipticity of the sample peptide at 222 nm, $[\Theta_\infty]$ is the mean residue ellipticity of an ideal peptide with 100% helicity (−44000 deg cm^2 $dmol^{-1}$), x is an empirical correction for end effects (2.5), and n is the number of residues in the potential helical region (5).

Multiple sequence alignment

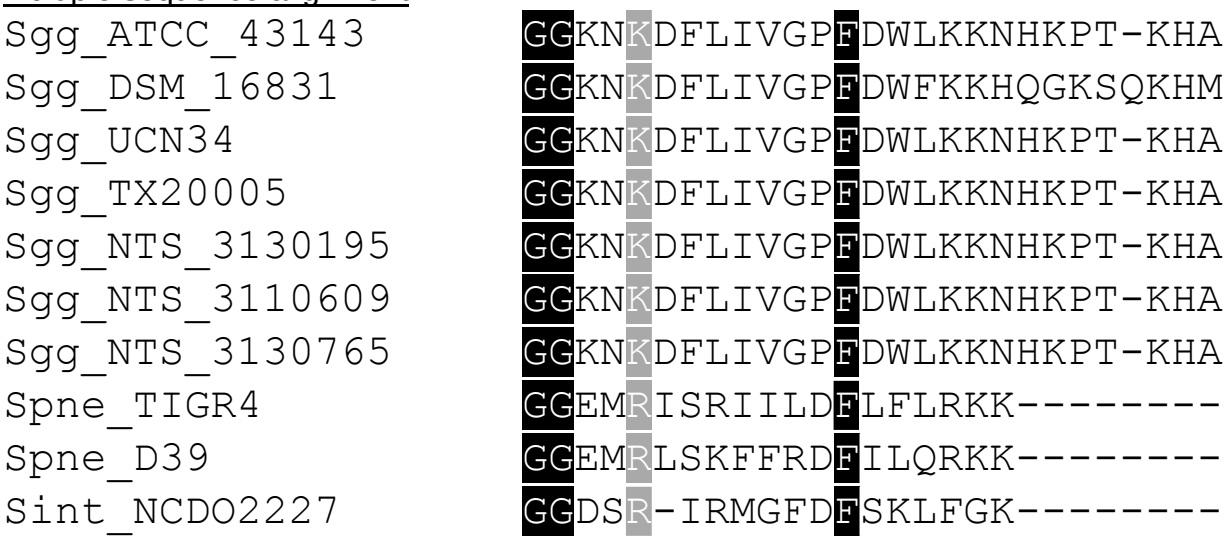


Figure S1. Multiple sequence alignment (MSA) of various *Sgg* strains, *S. pneumoniae* (TIGR4 and D39) and *S. intermedius* NCDO2227 with residues from the *N*-terminus up to the double glycine signal sequence removed.

HPLC traces of purified natural and synthetic CSP (220 nm)

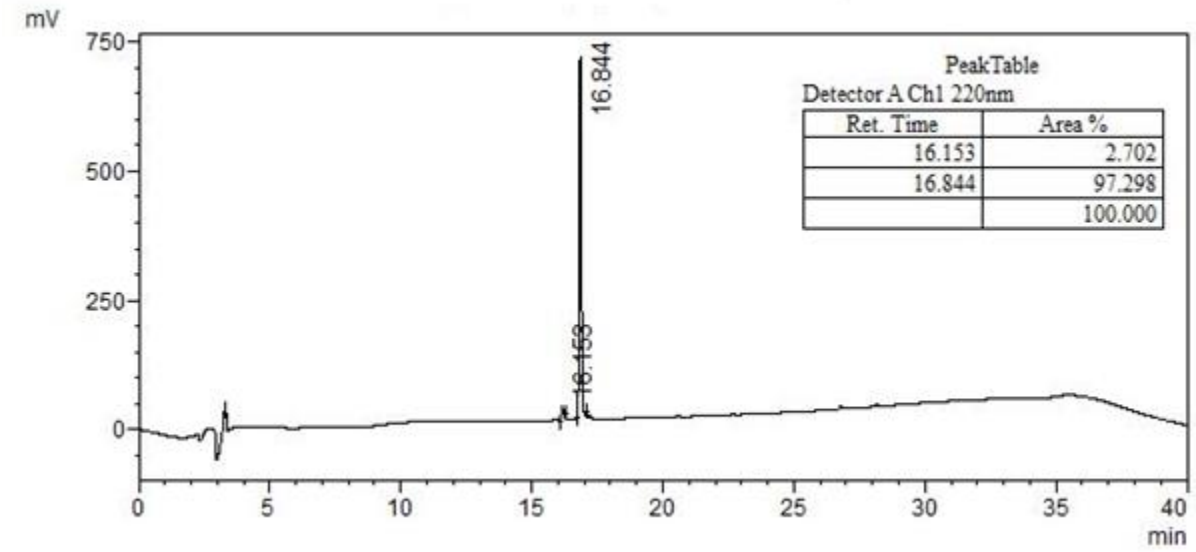


Figure S2. *Sgg* natural CSP analytical RP-HPLC chromatogram.

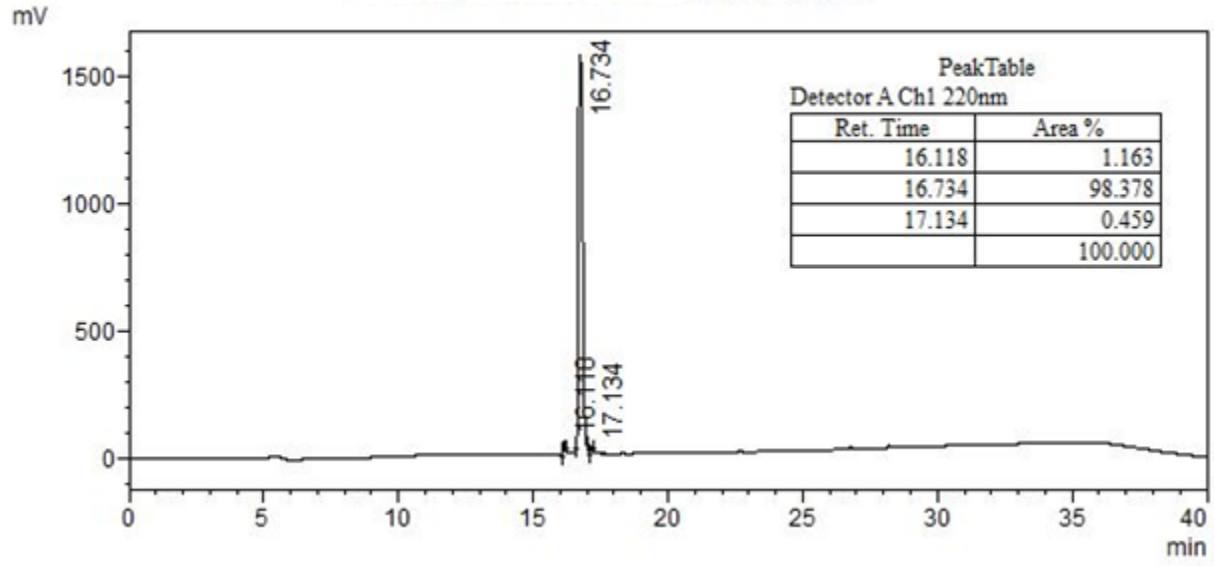


Figure S3. *Sgg* synthetic CSP analytical RP-HPLC chromatogram.

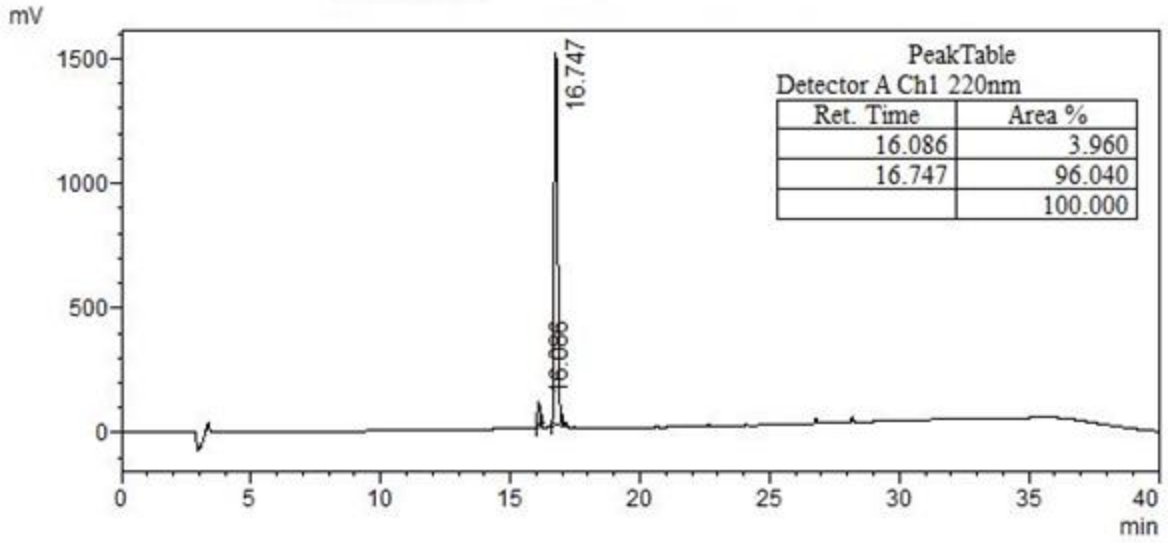


Figure S4. *Sgg* natural + synthetic CSP analytical RP-HPLC chromatogram.

HPLC traces of natural and synthetic CSP following chymotrypsin digestion (220 nm)

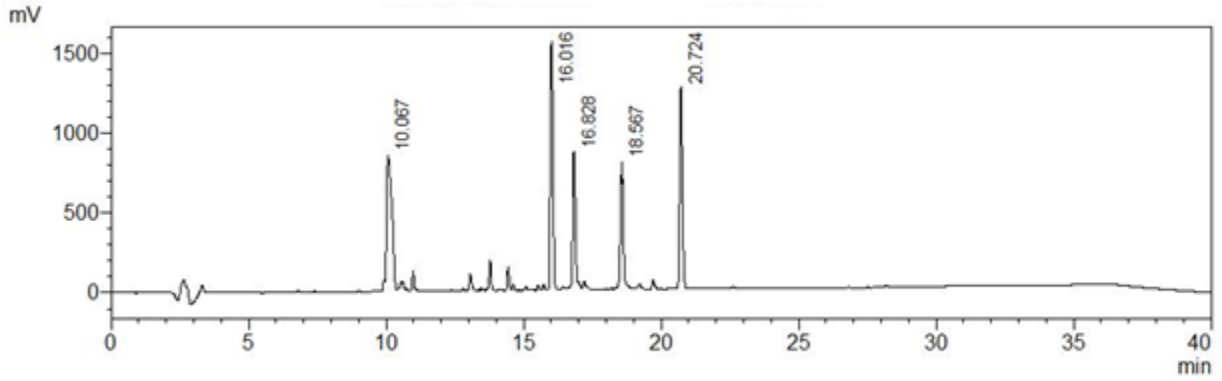


Figure S5. *Sgg* natural CSP analytical RP-HPLC chromatogram following chymotrypsin digestion.

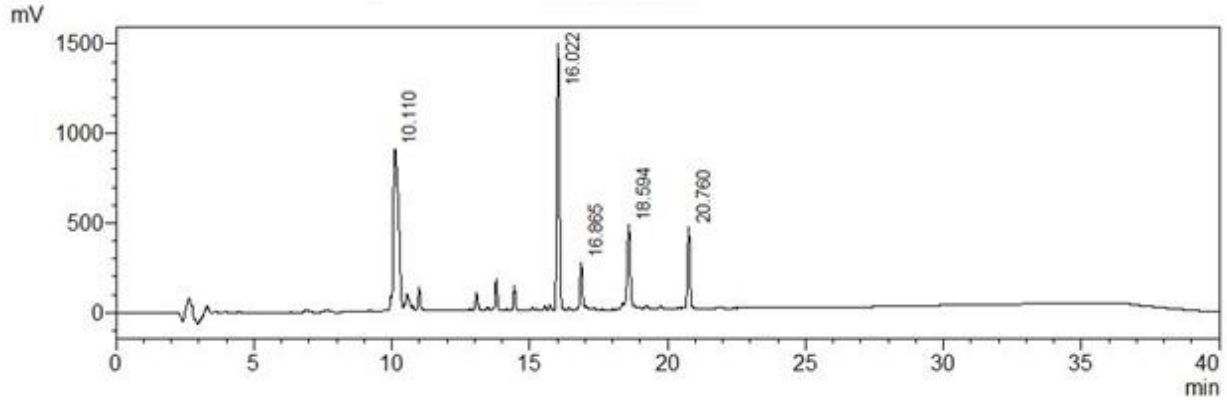


Figure S6. *Sgg* synthetic CSP analytical RP-HPLC chromatogram following chymotrypsin digestion.

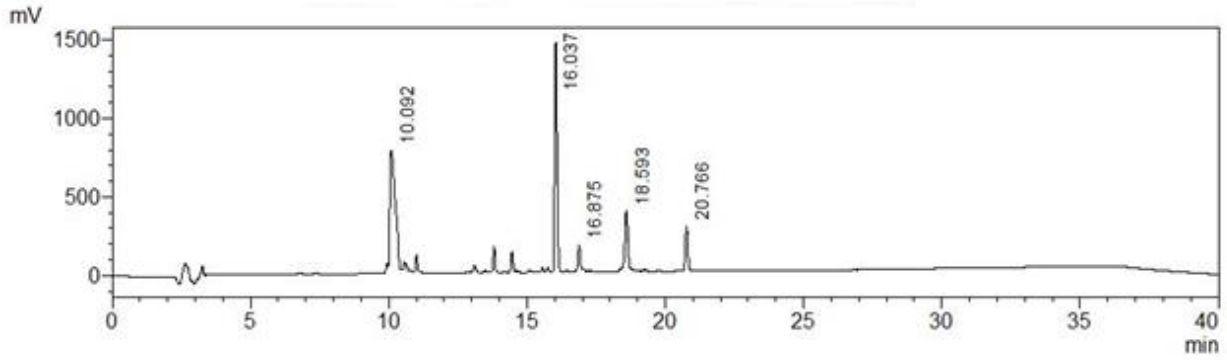
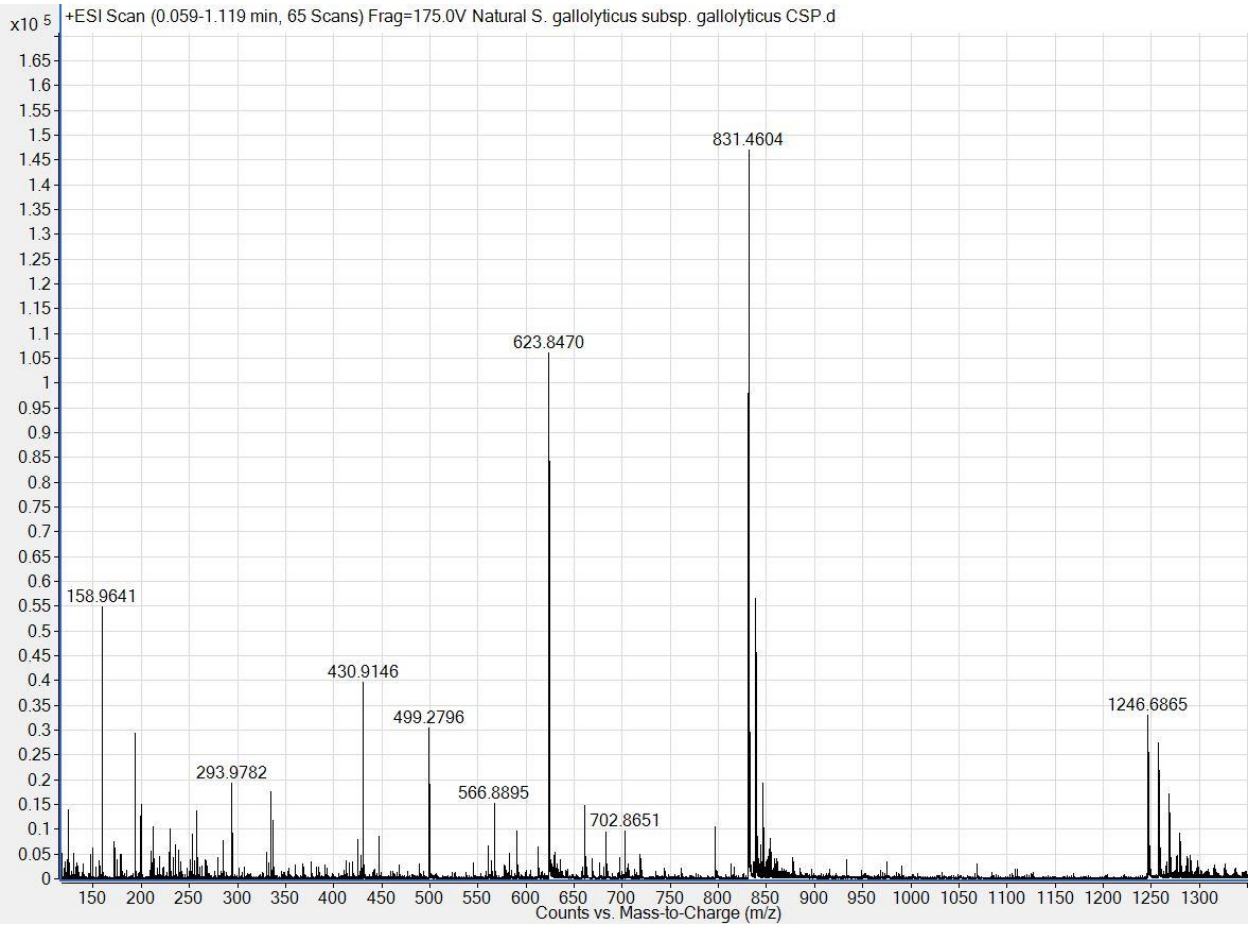


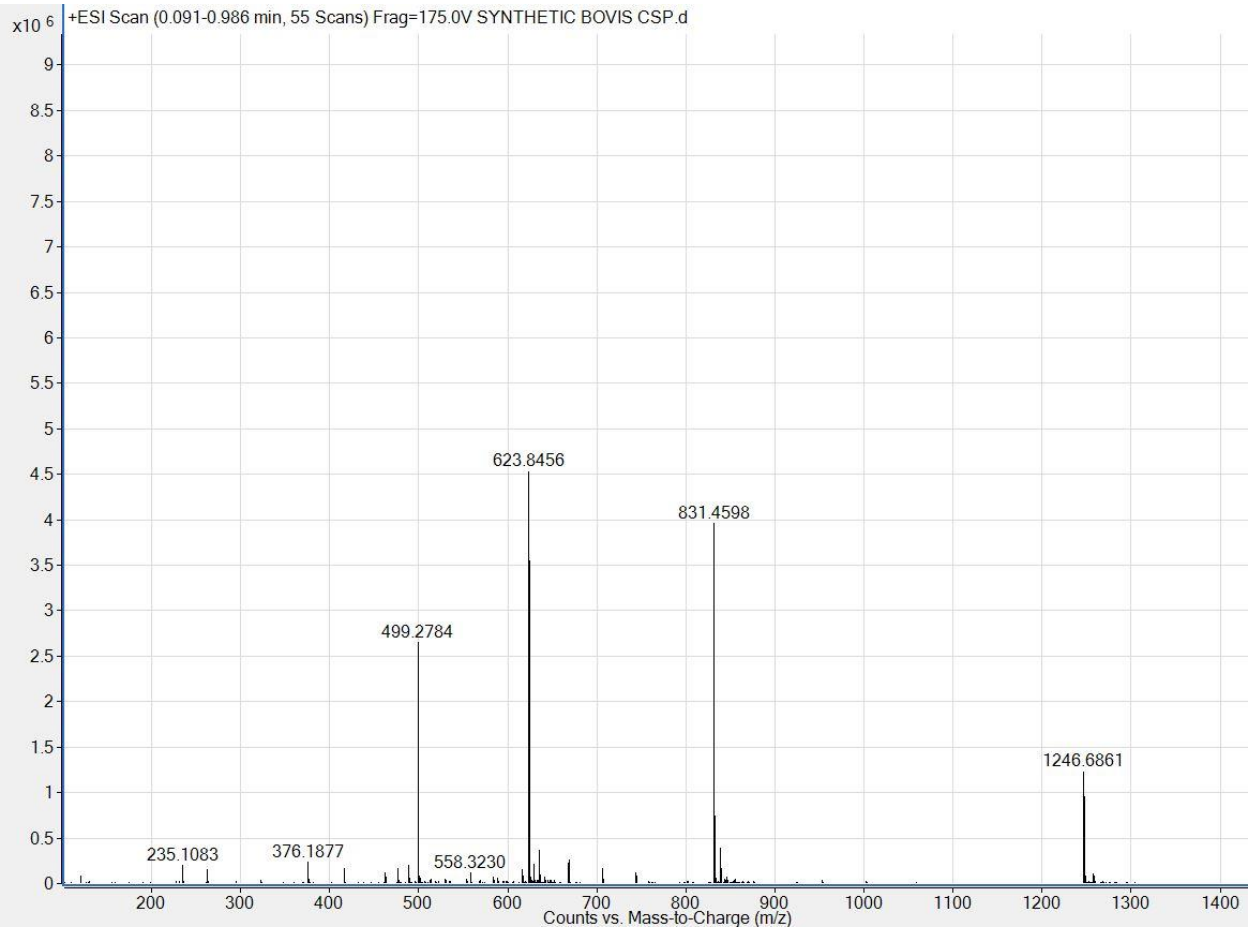
Figure S7. *Sgg* natural + synthetic CSP analytical RP-HPLC chromatogram following chymotrypsin digestion.

HRMS spectra of natural and synthetic CSP



Charged Ion	Expected Mass (m/z)	Observed Mass (m/z)	Ppm
+2H	1246.6855	1246.6865	0.802
+3H	831.4594	831.4604	1.203
+4H	623.8464	623.8470	0.962
+5H	499.2786	499.2796	2.02

Figure S8. *Sgg* natural CSP ESI+ HRMS spectrum with table of multiple charged states detected.



Charged Ion	Expected Mass (m/z)	Observed Mass (m/z)	Ppm
+2H	1246.6855	1246.6861	0.481
+3H	831.4594	831.4598	0.481
+4H	623.8464	623.8456	1.282
+5H	499.2786	499.2784	0.401

Figure S9. *Sgg* synthetic CSP ESI+ HRMS spectrum with table of multiple charged states detected.

MS/MS peptide mapping

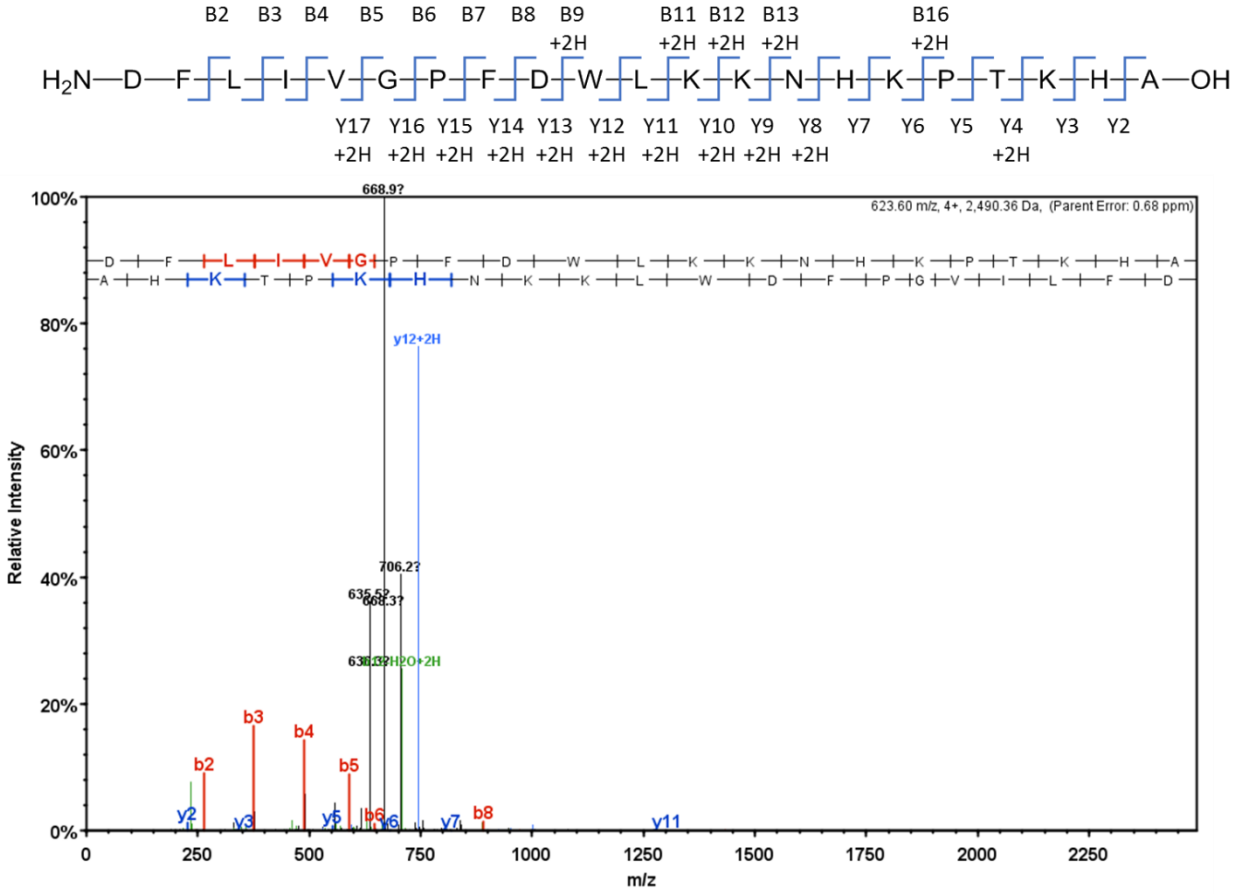
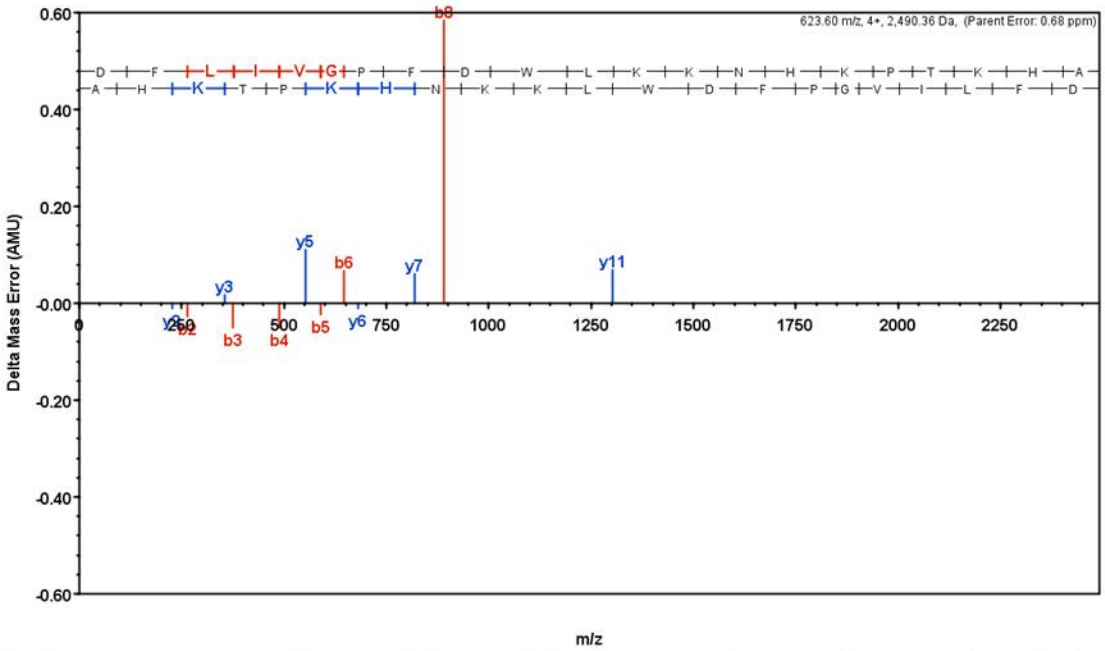


Figure S10. *Sgg* natural CSP peptide mapping spectrum of +4 charged state with various ion peaks and sequence showing various ions detected in the raw spectrum.



B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	116.0	58.5		98.0	D	2,491.4	1,246.2	2,474.3	2,473.4	21
2	263.1	132.1		245.1	F	2,376.3	1,188.7	2,359.3	2,358.3	20
3	376.2	188.6		358.2	L	2,229.3	1,115.1	2,212.2	2,211.3	19
4	489.3	245.1		471.3	I	2,116.2	1,058.6	2,099.2	2,098.2	18
5	588.3	294.7		570.3	V	2,003.1	1,002.1	1,986.1	1,985.1	17
6	645.4	323.2		627.4	G	1,904.0	952.5	1,887.0	1,886.0	16
7	742.4	371.7		724.4	P	1,847.0	924.0	1,830.0	1,829.0	15
8	889.5	445.2		871.5	F	1,750.0	875.5	1,732.9	1,731.9	14
9	1,004.5	502.8		986.5	D	1,602.9	801.9	1,585.9	1,584.9	13
10	1,190.6	595.8		1,172.6	W	1,487.9	744.4	1,470.8	1,469.8	12
11	1,303.7	652.3		1,285.7	L	1,301.8	651.4	1,284.8	1,283.8	11
12	1,431.8	716.4	1,414.7	1,413.8	K	1,188.7	594.9	1,171.7	1,170.7	10
13	1,559.9	780.4	1,542.8	1,541.9	K	1,060.6	530.8	1,043.6	1,042.6	9
14	1,673.9	837.5	1,656.9	1,655.9	N	932.5	466.8	915.5	914.5	8
15	1,811.0	906.0	1,793.9	1,793.0	H	818.5	409.7	801.4	800.5	7
16	1,939.1	970.0	1,922.0	1,921.0	K	681.4	341.2	664.4	663.4	6
17	2,036.1	1,018.6	2,019.1	2,018.1	P	553.3	277.2	536.3	535.3	5
18	2,137.2	1,069.1	2,120.1	2,119.1	T	456.3	228.6	439.2	438.2	4
19	2,265.3	1,133.1	2,248.2	2,247.2	K	355.2	178.1	338.2		3
20	2,402.3	1,201.7	2,385.3	2,384.3	H	227.1	114.1			2
21	2,491.4	1,246.2	2,474.3	2,473.4	A	90.1	45.5			1

Figure S11. *Sgg* natural CSP peptide mapping mass error with table of identified peaks from +4 charged state raw spectrum. Highlighted masses indicate corresponding ion peaks detected in raw spectrum.

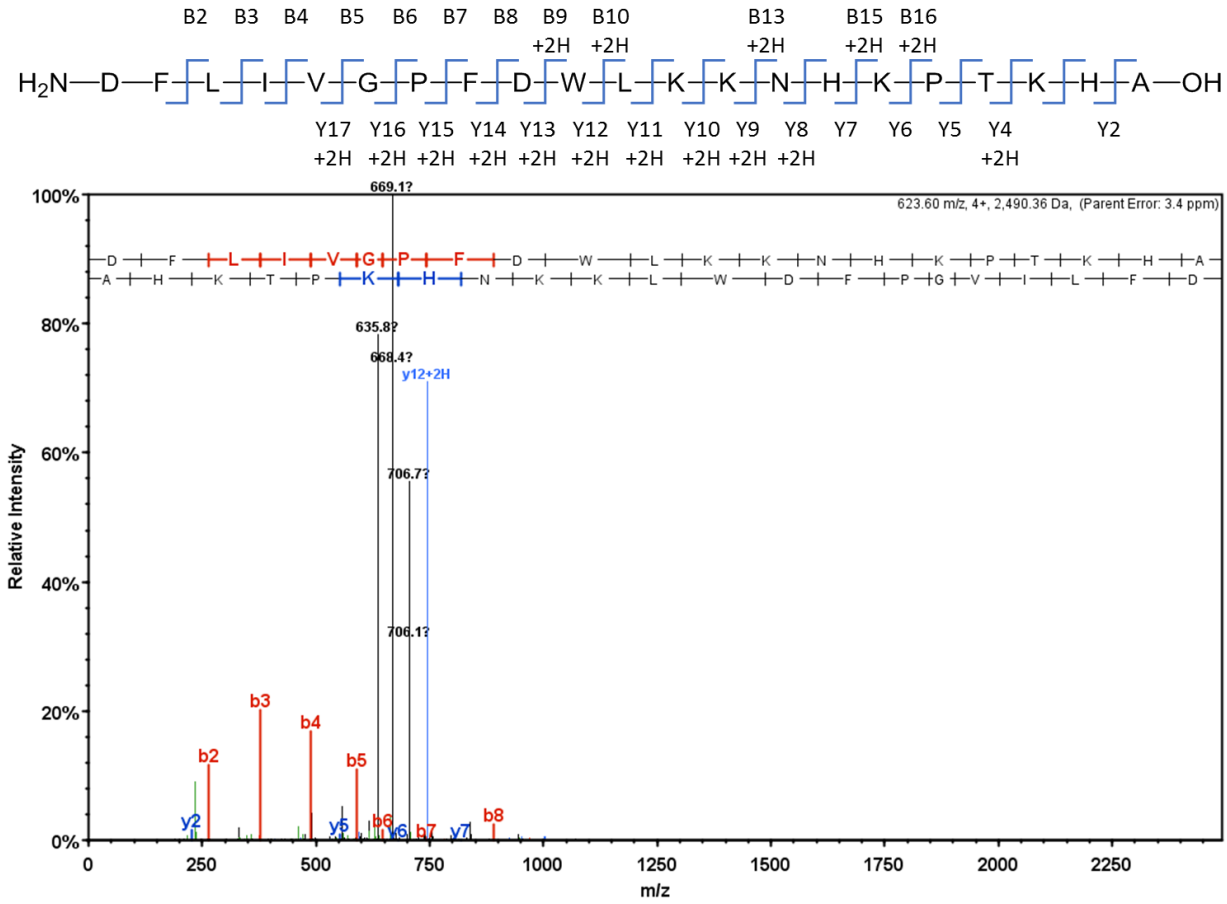
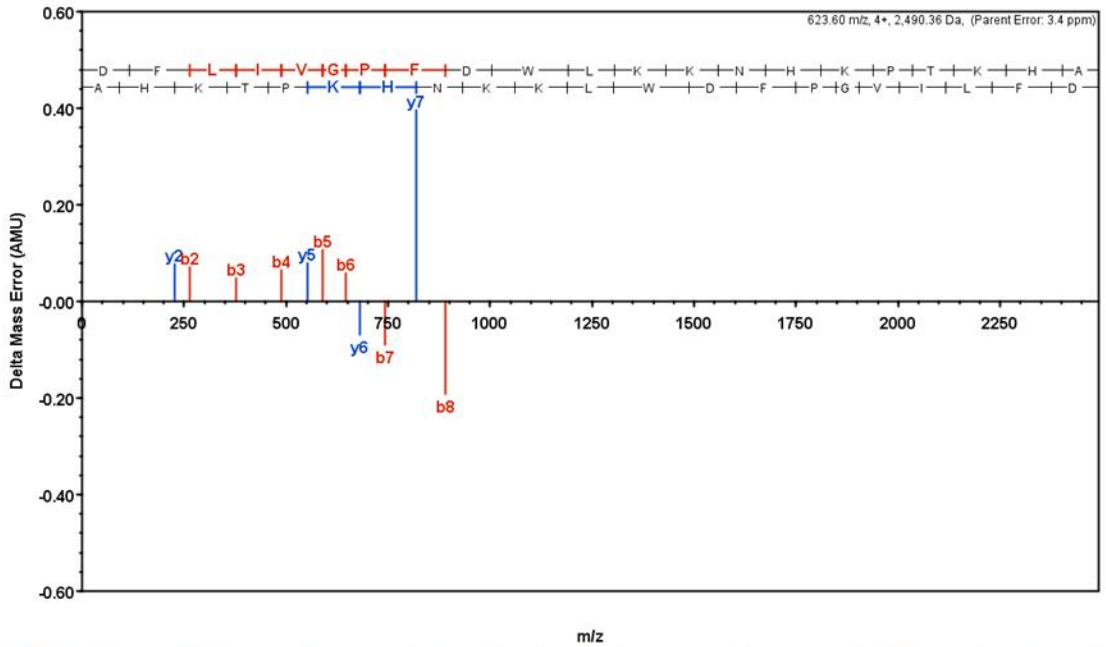


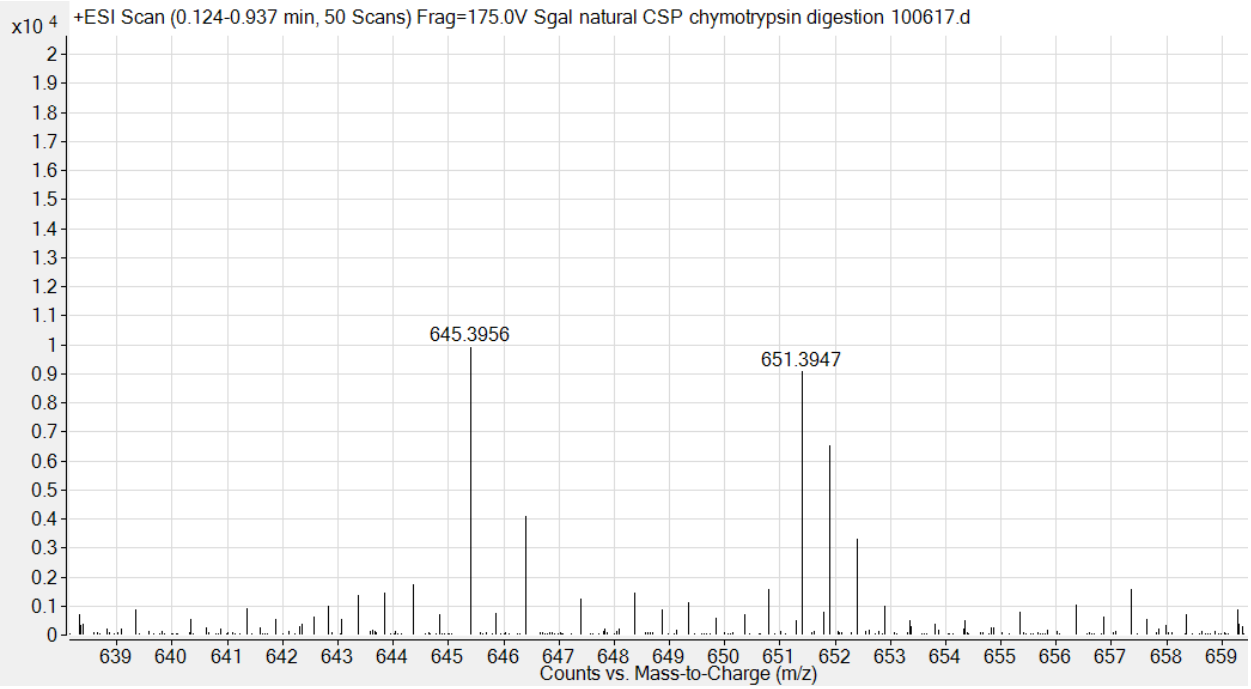
Figure S12. *Sgg* Synthetic CSP peptide mapping spectrum of the +4 charge state with various ion peaks and sequence showing various ions detected in the raw spectrum.



B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	116.0	58.5		98.0	D	2,491.4	1,246.2	2,474.3	2,473.4	21
2	263.1	132.1		245.1	F	2,376.3	1,188.7	2,359.3	2,358.3	20
3	376.2	188.6		358.2	L	2,229.3	1,115.1	2,212.2	2,211.3	19
4	489.3	245.1		471.3	I	2,116.2	1,058.6	2,099.2	2,098.2	18
5	588.3	294.7		570.3	V	2,003.1	1,002.1	1,986.1	1,985.1	17
6	645.4	323.2		627.4	G	1,904.0	952.5	1,887.0	1,886.0	16
7	742.4	371.7		724.4	P	1,847.0	924.0	1,830.0	1,829.0	15
8	889.5	445.2		871.5	F	1,750.0	875.5	1,732.9	1,731.9	14
9	1,004.5	502.8		986.5	D	1,602.9	801.9	1,585.9	1,584.9	13
10	1,190.6	595.8		1,172.6	W	1,487.9	744.4	1,470.8	1,469.8	12
11	1,303.7	652.3		1,285.7	L	1,301.8	651.4	1,284.8	1,283.8	11
12	1,431.8	716.4	1,414.7	1,413.8	K	1,188.7	594.9	1,171.7	1,170.7	10
13	1,559.9	780.4	1,542.8	1,541.9	K	1,060.6	530.8	1,043.6	1,042.6	9
14	1,673.9	837.5	1,656.9	1,655.9	N	932.5	466.8	915.5	914.5	8
15	1,811.0	906.0	1,793.9	1,793.0	H	818.5	409.7	801.4	800.5	7
16	1,939.1	970.0	1,922.0	1,921.0	K	681.4	341.2	664.4	663.4	6
17	2,036.1	1,018.6	2,019.1	2,018.1	P	553.3	277.2	536.3	535.3	5
18	2,137.2	1,069.1	2,120.1	2,119.1	T	456.3	228.6	439.2	438.2	4
19	2,265.3	1,133.1	2,248.2	2,247.2	K	355.2	178.1	338.2		3
20	2,402.3	1,201.7	2,385.3	2,384.3	H	227.1	114.1			2
21	2,491.4	1,246.2	2,474.3	2,473.4	A	90.1	45.5			1

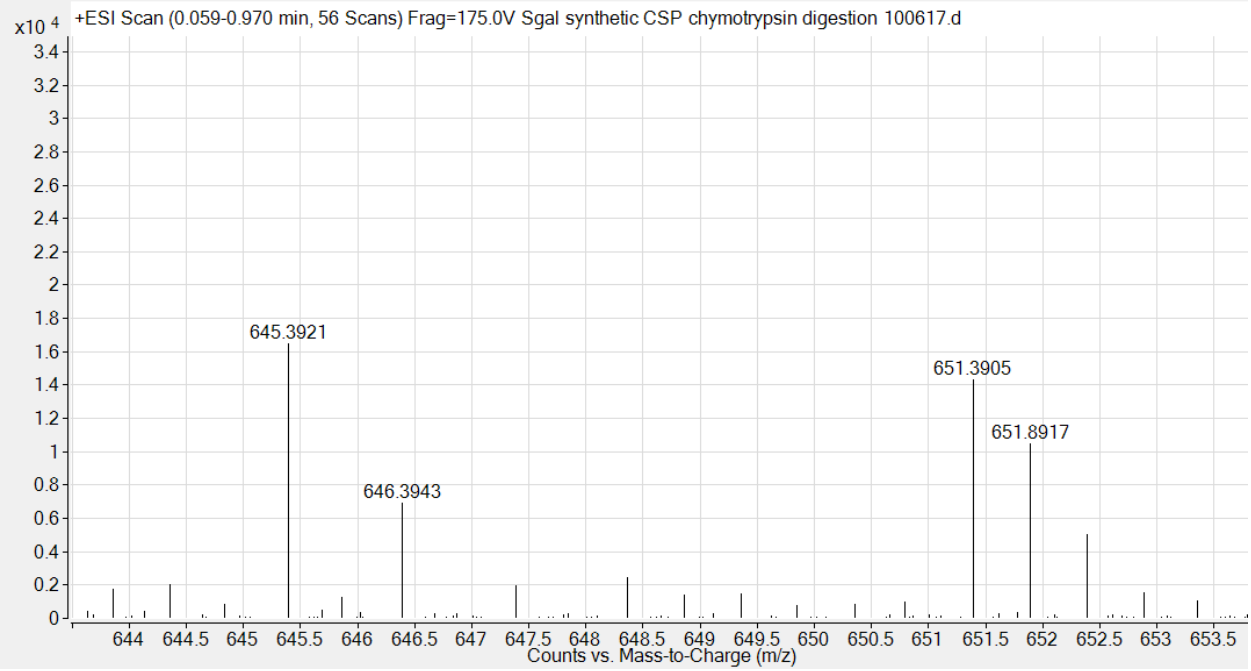
Figure S13. *Sgg* synthetic CSP peptide mapping mass error with table of identified peaks from +4 charged state raw spectrum. Highlighted masses indicate corresponding ion peaks detected in raw spectrum.

HRMS spectra of natural and synthetic CSP chymotrypsin digestion



Fragment	Charged Ion	Expected Mass (m/z)	Observed Mass (m/z)	ppm
LIVGPF	**+1H**	**645.3970**	**645.3956**	**2.17**
LKKNHKPTKHA	**+2H**	**651.3937**	**651.3947**	**1.5**

Figure S14. *Sgg* natural CSP chymotrypsin digestion ESI+ HRMS spectrum with table of multiple charged states detected.



Fragment	Charged Ion	Expected Mass (m/z)	Observed Mass (m/z)	ppm
LIVGPF	**+1H**	**645.3970**	**645.3921**	**7.59**
LKKNHKPTKHA	**+2H**	**651.3937**	**651.3905**	**4.91**

Figure S15. *Sgg* synthetic CSP chymotrypsin digestion ESI+ HRMS spectrum with table of multiple charged states detected.

Circular Dichroism spectra of natural and synthetic CSP

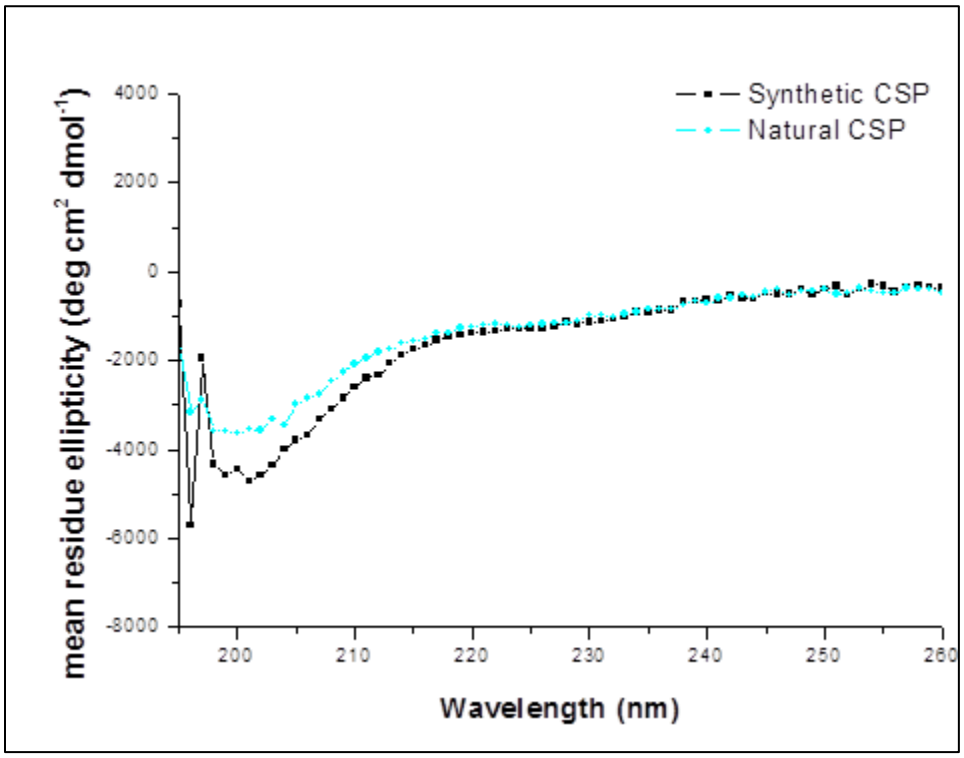


Figure S16. CD spectra of natural and synthetic CSP in PBS.

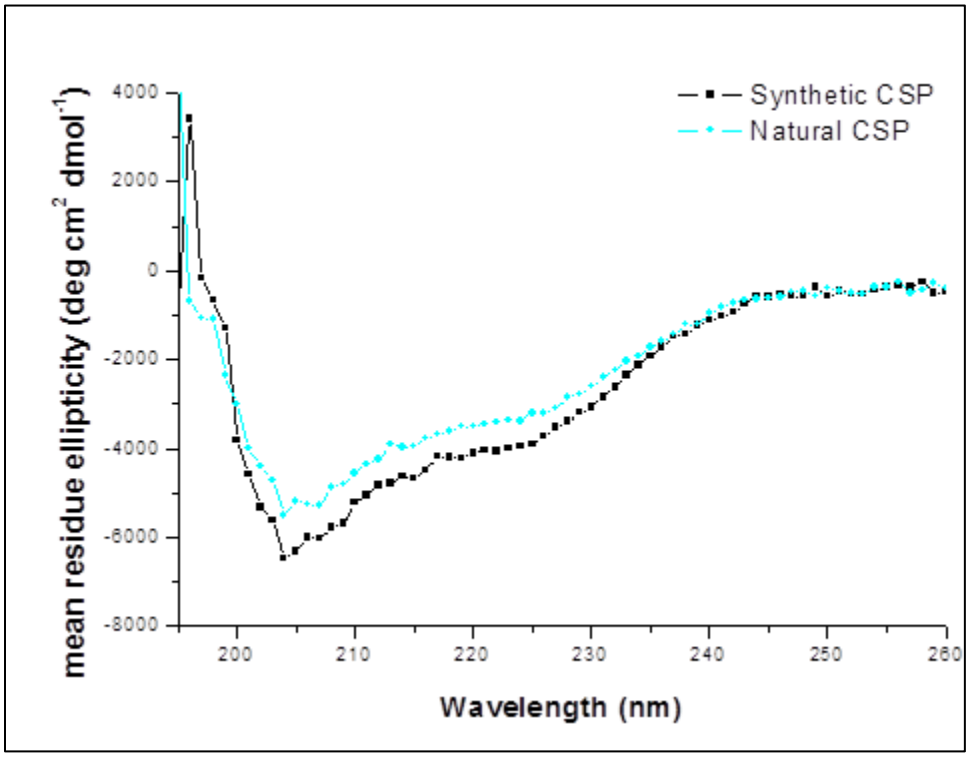


Figure S17. CD spectra of natural and synthetic CSP in PBS with 20% TFE.

Inhibition Assays

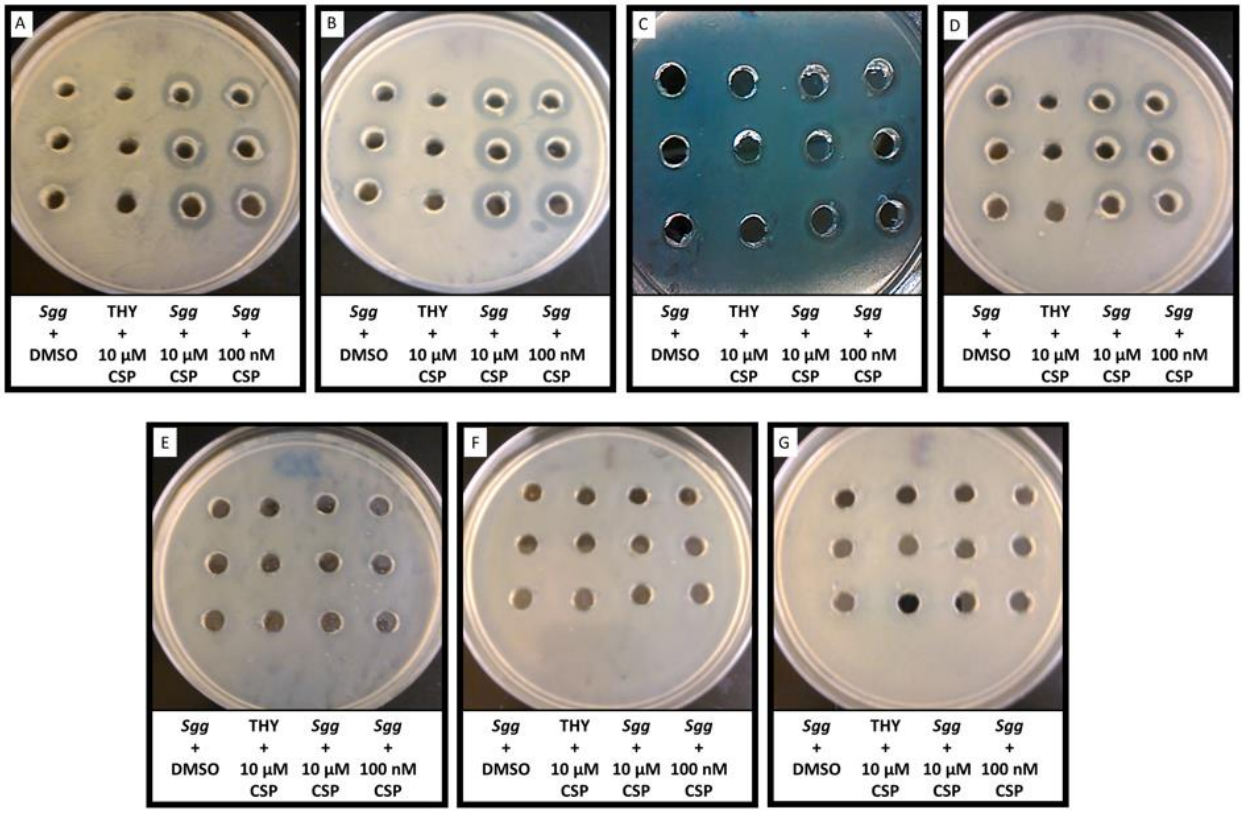


Figure S18. Images comparing results from *Sgg* interspecies inhibition assay, where *Sgg* is incubated with CSP (10 µM or 100 nM) or DMSO and the supernatants are tested against A) *S. anginosus* ATCC 33397 B) *S. constellatus* ATCC 27823 C) *S. vestibularius* F0396 D) *S. intermedius* F0413 E) *S. mutans* ATCC 25175 F) *S. agalactiae* MNZ938 G) *Sgg* TX20005. Note that part C looks different because the image was taken against a different background for better visualization.

Bacteriocin-like inhibitory substance partial characterization

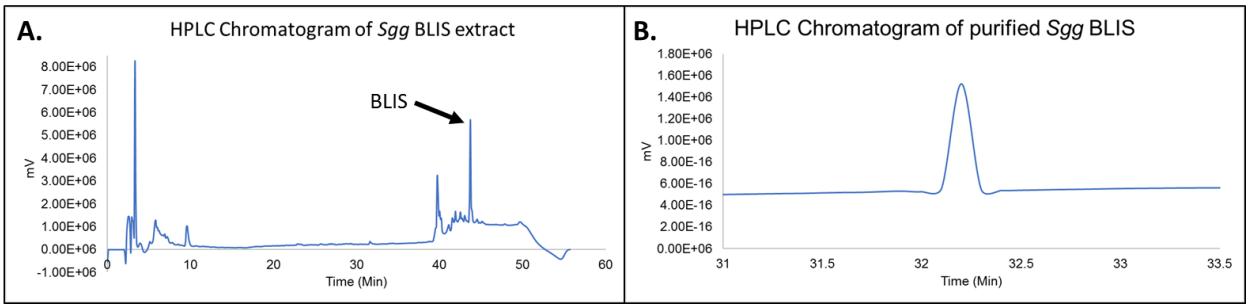


Figure S19. Isolation of *Sgg* BLIS. A) HPLC trace of the PBS with Tween-20 extract; B) Zoomed-in analytical HPLC trace of the active fraction (>95% purity).

References

1. Saris PEJ, Paulin LG, Uhlén M. 1990. Direct amplication of DNA from colonies of Bacillus subtilis and Escherichia coli by the polymerase chain reaction. Journal of Microbiological Methods 11:121-126.
2. Drancourt M, Roux V, Fournier P-E, Raoult D. 2004. rpoB gene sequence-based identification of aerobic Gram-positive cocci of the genera Streptococcus, Enterococcus, Gemella, Abiotrophia, and Granulicatella. Journal of clinical microbiology 42:497-504.
3. Kheterpal I, Mathies RA. 1999. Peer Reviewed: Capillary Array Electrophoresis DNA Sequencing. Analytical chemistry 71:31A-37A.
4. Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces usingPhred. I. Accuracy assessment. Genome research 8:175-185.
5. Luo P, Baldwin RL. 1997. Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. Biochemistry 36:8413-8421.