Identification of Regulatory Elements in the Untranslated Regions of Streptolysin S Associated Gene A Messenger RNA from Group A *Streptococcus*

Cameron R. Carroll,[1,3] Sara G. Nibar,[1,3] Alexis S. Brown,[1] Lauren R. Angello,[1,2] Gabriela C. Pérez-Alvarado[1*] and Brian M. Lee[1*]

[1]*Department of Chemistry, Coastal Carolina University, Conway, SC*
[2]*Department of Biology, Coastal Carolina University, Conway, SC*
[3]*Both authors contributed equally*

Streptococcus pyogenes, also known as group A *Streptococcus* (GAS), is a human pathogen associated with a variety of diseases such as strep throat, scarlet fever, toxic shock syndrome, and necrotizing fasciitis. One of the virulence factors released by GAS during an invasive infection is a cytotoxic peptide, streptolysin S (SLS), which inhibits the immune response to necrotizing fasciitis. The streptolysin S associated gene A product, SagA, is modified to produce SLS. The *sag* operon includes *sagA* and the genes required for enzyme-mediated post-translational modifications of SagA and the export of SLS. The *sagA* gene is contained within the pleiotropic effect locus (*pel*), which produces a small RNA (sRNA) that regulates the expression of other virulence factors. Potential mRNA interactions with the Pel sRNA have been mapped to the 5' and 3' untranslated regions (UTRs) of *sagA*. Our studies aim to identify and characterize RNA structural motifs in Pel/*sagA* that regulate the expression of *sagA* and other virulence factors. Several RNA constructs of Pel/*sagA* were designed to include regions predicted to contain secondary structure. The corresponding sequences were isolated by PCR from genomic DNA to create templates for *in vitro* transcription. After purification, the RNA constructs were analyzed by gel electrophoresis to verify size, and by RNase T_1 digestion to assay for secondary structure. Three-dimensional models were generated using the FARFAR algorithm in Rosetta in order to identify regions of Pel/*sagA* that may be involved in regulatory interactions. Differential scanning fluorimetry provided evidence that the 5' and 3' UTRs of Pel/*sagA* contain stable structural regions. It is expected that the identification of structural motifs necessary for the regulation of gene expression will aid in the design of therapeutic strategies to inhibit the production of streptolysin S and other virulence factors.

Introduction

Diseases ranging from a mild strep throat to more serious exotoxin-mediated diseases, such as scarlet fever, and the life-threating necrotizing fasciitis, are caused by the human pathogen, *Streptococcus pyogenes,* also known as group A *Streptococcus* (GAS).[1-3] During pathogenesis, the expression of virulence factors is closely regulated in response to changes in the host environment.[2, 4-8] One of these virulence factors is the exotoxin streptolysin S (SLS), which is a hemolytic peptide secreted by GAS.[9] Besides affecting erythrocytes, SLS also forms pores in cells from the innate and adaptive immune systems. SLS is the result of post-translational modifications of cysteine, serine and threonine residues in the SagA peptide to yield the thiazole, oxazole and methyloxazole heterocyclic residues that are essential for the cytolytic functions of the excreted SLS peptide.[10-13] SLS is part of a family of bacteriocins known as thiazole/oxazole-modified microcins (TOMMs).[14, 15] The SagA protein is encoded by the *sagA* gene, which is part of a nine -gene SLS-associated (*sag*) operon.[10, 16] SagA is a 53 amino acid precursor peptide with a 23 amino acid leader sequence followed by a 30 amino acid propeptide that is released by proteolytic cleavage after a Gly -Gly sequence.[13, 17] The enzymes and transport proteins necessary for the post-translational modifications of SagA and export of SLS, are all derived from genes found in the *sag* operon (Fig. 1a).[16] Specifically, the *sagB*, *sagC* and *sagD* genes code for a dehydrogenase, a cyclodehydratase, and a docking scaffold protein, respectively, that catalyze modifications to cysteine, serine and threonine residues. SagE protease has been proposed to cleave the modified precursor, releasing the mature SLS peptide.[17] Genes *sagG-sagI* code for components of the ATP-binding cassette transporter facilitating export of SLS.[10, 11]

In addition to cytolytic activity, SLS also elicits a series of host responses that promote the invasive phase of streptococcal infection. In epithelial keratinocytes, SLS induces attenuation of Akt-mediated signaling. SLS also activates the nuclear factor kappa B (NFκB)-dependent inflammatory signaling, which causes the release of an inflammatory cytokine, interleukin 1-beta (IL-1β), leading to programmed cell death and increased tissue damage during infection.[18, 19] During necrotizing fasciitis, SLS can stimulate cutaneous nociceptor neurons to release neuropeptides that inhibit the immune response.[20, 21] SLS activates the transient receptor potential vanilloid 1 (TRPV1) cation channel in peripheral nerve cells, leading to the release of calcitonin gene related peptide (CGRP), which inhibits the recruitment of neutrophils that are necessary to prevent invasive bacterial infections.[20]

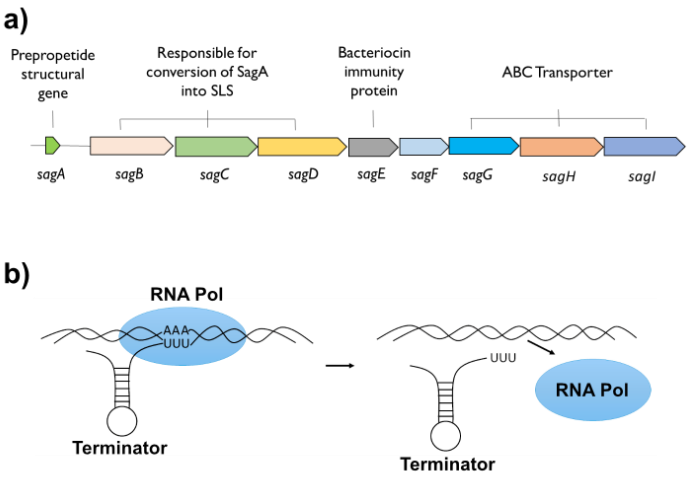

Fig. 1. **a)** The gene *sagA* in GAS codes for the protein streptolysin A. Streptolysin A is modified post-translationally to form the exotoxin streptolysin S (SLS). Each gene in the *sag* operon plays a role in the modification leading to SLS including export.
b) Rho-independent termination of transcription. Nascent formation of a stem-loop structure which is followed by an uracil-rich region of the mRNA transcript causes RNA polymerase to dissociate and terminate transcription.

The *sagA* gene is contained within the pleiotropic effect locus, *pel*, that produces a regulatory small RNA, Pel, which is known to influence the expression of multiple virulence factors, including the anti-phagocytic, M-protein (*emm*), streptococcal inhibitor of complement (*sic*), NAD-glycohydrolase (*nga*), and the cysteine protease, streptococcal pyrogenic exotoxin B (*speB*).[9, 12, 22-27] Interactions of Pel/*sagA* with other RNA transcripts or with proteins may influence the regulatory role of Pel/*sagA* either at the transcriptional and translational levels.[24, 25] The 5' untranslated region (UTR) of Pel/*sagA* contains a ribosome-binding site (RBS) just upstream of the start codon for *sagA*.[9, 16, 23] The 5' UTRs of mRNAs are often sites of translational regulation, which depends on the presence of RNA structural elements. Both the RBS and AUG start codon of *sagA* are contained within a predicted stem-loop structure that would block ribosome access and prevent the initiation of translation.[28]

Regions further upstream in the 5' UTR of Pel/*sagA* contain predicted secondary structures that are potential sites for either intramolecular mRNA interactions, or intermolecular sRNA and protein interactions that could destabilize this stem-loop and influence the availability of the RBS.[24]

Regions within the 3' UTR of bacterial mRNAs are also involved in post -transcriptional regulation of gene expression. The 3' UTR may promote the translation of *sagA* through intramolecular interactions with the 5' UTR that would disrupt any structural inhibitors of translation.[24] Alternatively, direct base-pairing between the 3' UTR and the Shine-Dalgarno sequence (AGGAGG) in the 5' UTR may promote cleavage by RNase III and inhibit translation through RNA degradation.[24, 29] Other regulatory mechanisms involving the instability of structural elements within the 3' UTR could affect translation by making the mRNA of *sagA* susceptible to cleavage by 3' to 5' exoribonucleases.[24, 29] The 3' UTR of the Pel/*sagA* transcript includes a predicted rho-independent terminator (Fig. 1b),[30-32] which is proposed to block the transcription of the remaining downstream genes of the *sag* operon from *sagB* to *sagI*.[16] This terminator is considered "leaky" since it allows some read-through transcription resulting in a differential abundance of the Pel/*sagA* transcript and the polycistronic *sag* mRNA transcript. Disruption of the terminator structure by inherent dynamic instability, or by intermolecular RNA-RNA interactions would allow for transcription of the entire *sag* operon that is required for production of SLS. Specific sequences and structural elements found in the 3' UTR may also be targeted by other sRNA transcripts that regulate the expression of related virulence factors.[24, 29]

The focus of our study has been on these predicted RNA structural elements: a) the 5' UTR that contains a branched stem-loop structure that could influence transcription of *sagA* after initiation, or interact with mRNA transcripts targeted by the Pel sRNA; b) both the RBS and AUG start codon are contained within a stem-loop structure that may prevent translation of the *sagA* gene prior to interactions with other regulatory factors, which would make this region accessible to the ribosome; and c) the 3' UTR of the Pel/*sagA* transcript that includes an unusual rho-independent terminator, which blocks transcription of the downstream genes in the *sag* operon.[16] The terminator structure includes a branched dual stem-loop structure that may mediate interactions with the 5' UTR of Pel/*sagA* or with other sRNAs. The terminator may be leaky due to the high proportion of AU base pairs. Disruption of the terminator structure through its inherent dynamic instability may allow for some low level of transcription of the full *sag* operon required for production of SLS. Interactions with the unusual branched, dual stem-loop would influence the stability of the terminator and modulate the expression of the downstream genes of the *sag* operon.

All these individual regions were characterized with the goal of identifying regulatory elements within Pel/*sagA* containing stable or transient structures that could be the sites for interactions affecting transcription, translation, RNA degradation, or other regulatory effects. Specific RNA constructs were designed using comparative sequence analysis, sequence covariance analysis, and the results of secondary structure predictions (Fig. 2). Selected sequences were isolated by PCR from the genomic DNA of GAS strain HSC5[33] and inserted into pUC18 plasmids. Purified plasmids were linearized and used as templates for *in vitro* transcription. RNA constructs were purified and analyzed by gel electrophoresis, RNase T_1 digestion, and differential scanning fluorimetry. The predicted secondary structures aided in the generation of three-dimensional models using the program FARFAR through the Rosetta webserver (ROSIE).[34-36] Specific conditions were identified that increased the stability of the RNA studied and will be used for structural studies of these RNA regulatory elements in the 5' and 3' UTR of Pel/*sagA*. Identification of structural motifs necessary for the regulation of gene expression will provide specific targets for the design of RNA-based therapeutic strategies that could inhibit the production of SLS and other virulence factors.

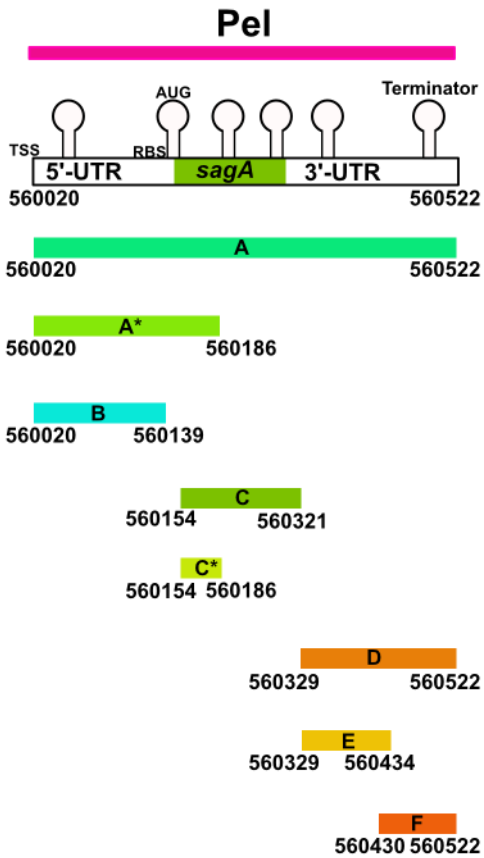

Fig. 2. RNA constructs containing regions with predicted secondary structure. Boundaries for RNA constructs were determined using information from secondary structure predictions. Each construct contains the regions of interest as described in Table 1. The cartoon diagram denotes presence of secondary structure, which could be a combination of one to three predicted stem-loops.

Methods

PCR-based isolation of DNA from genomic DNA

Using the previously identified promoter and terminator as boundaries,[9, 23] DNA sequences containing *pel/sagA* were isolated by PCR from the genomic DNA of *S. pyogenes* strain HSC5.[33] These were used as templates for further PCR methods to generate smaller DNA templates. The design of primers and constructs was aided by results from RNA secondary structure predictions using mfold and RNAfold.[37-42] The forward primers contained a restriction site for cloning and the T7 promoter sequence for *in vitro* transcription with T7 RNA polymerase. The reverse primers contained restriction sites for linearization in addition to restriction sites for cloning into a pUC18 vector (Novagen). PCR reactions contained 0.5 µM of each forward and reverse primer, 0.4 units of Phusion High-Fidelity DNA polymerase (Thermo Scientific), 0.2 mM of each NTP (NEB) and 0.5 µL of genomic DNA[43] or 0.5 µg of purified isolated *pel/sagA* DNA. The PCR reactions (20 µL) were carried in a MJ Mini thermal cycler (Bio-Rad) with the following parameters: 30 seconds melting time at 98 ºC, followed by 30 cycles of 10 seconds denaturation at 98 ºC, 30 seconds of annealing at 59 ºC and 45 seconds of elongation at 72 ºC, with a final step of 5 minutes at 72 ºC. PCR products were purified using the Wizard SV Gel and PCR Clean-Up System (Promega) before and after digestion at 37 ºC with 20 units of BamHI-HF (NEB) and HindIII-HF (NEB). The PCR products were characterized on a 1.5% agarose gel with 0.25 µg/mL ethidium bromide (Bio-Rad) with gel loading dye (0.04% bromophenol blue, 0.04% xylene cyanol FF and 5% glycerol) and run using TAE (Tris base, acetic acid and EDTA) solution as running buffer for approximately 45 minutes at 80 V.[44] Gel images were captured and

analyzed using a ChemiDoc XRS+ gel imaging system (Bio-Rad) by comparing gel bands to a DNA ladder, Quick-Load PCR Marker (NEB).

Preparation of plasmid DNA templates

After restriction digest, the purified PCR-derived DNA templates for A, B, C, D, E and F constructs were inserted into previously digested and purified pUC18 plasmids (Novagen) using 200 units of T4 DNA ligase (NEB). Twenty microliter reactions containing 4 µg of purified PCR digests and 1 µg of digested pUC18. Reactions were incubated 50 minutes at room temperature followed by overnight incubation at 16 ºC. Aliquots of each ligation reaction (4 µL) were used for transformation into 30 µL of XL1-Blue *Escherichia coli* cells (Agilent), plated in Luria Bertani (LB) media plates with 200 µg/mL ampicillin. Cells were incubated with the ligation reaction in ice for 30 minutes followed by a heat shock step of 45 seconds at 42 ºC and 2 minutes incubation in ice. After incubation in ice, super optimal broth with catabolite repression (SOC) media were added to the bacterial cells, and the culture was then incubated for approximately 1 hour in a shaker at 37 ºC.[44] Two hundred microliters of the culture were plated onto agar plates in LB media with 200 µg/mL ampicillin. After overnight incubation at 37 ºC, bacterial colonies were picked and screened by PCR to identify those that contained plasmids with the DNA template sequences. Briefly, each colony was picked with a pipette, an agar plate with LB media and ampicillin was touched with the tip, and the bacterial colony was placed in 50 µL of sterilized filtered ultrapure water. The tube was vortexed, heated for 5 minutes at 95-100 ºC, and centrifuged for 1 minute at 16,000 x *g*. An aliquot of the supernatant (10 µL) was used as template for PCR amplification while the plates were incubated at 37 ºC overnight.

Each 20 µL PCR reaction used for screening contained 10 µL of template, 0.25 µM of each forward and reverse primer, 0.5 units of Taq DNA polymerase with ThermoPol Buffer (NEB), 0.2 mM of each NTP (NEB). Thermal cycling was done using the following parameters: 2 minutes melting time at 94 ºC, followed by 35 cycles of 15 seconds denaturation at 94 ºC, 30 seconds of annealing at 59 ºC and 1 minute of elongation at 72 ºC with a final extension step of 5 minutes at 72 ºC. Aliquots of each reaction were characterized on 1.5 or 1.6% agarose-TAE gels with 0.25 µg/mL ethidium bromide (Bio-Rad). Gel loading dye (0.04% bromophenol blue, 0.04% xylene cyanol FF and 5% glycerol) was added to each reaction and the gels were run in TAE buffer at 80 V with running times of 40-50 minutes.[44] Gel images were captured and analyzed using the ChemiDoc XRS+ gel imaging system (Bio-Rad) by comparing gel bands to a DNA ladder (Quick-Load PCR Marker, NEB).

After overnight incubation of the plates, bacterial colonies that were determined to contain plasmids with the correct DNA template were selected and used to inoculate 5 mL LB media with 200 µg/mL ampicillin. Cultures were grown at 37 ºC overnight, the cells were harvested by centrifugation and plasmids were purified using a QIAprep Spin Miniprep kit (QIAGEN) with only minimal changes to the protocol. Aliquots of each selected plasmid (approximately 0.4 µg, 1 µL of eluted plasmid) were digested with 10 units of NdeI (NEB) and 10 units of EcoRI (NEB) to confirm if they contained the correct DNA template. Digests were analyzed by gel electrophoresis by comparing to a DNA ladder (Quick-Load PCR Marker, NEB). Colony screening by PCR amplification of the selected plasmids was also conducted similarly as described above, except that 0.25 µM M13 reverse primer replaced the forward primer on each reaction. Thermal cycling was done using the same parameters as above, except for 30 seconds of annealing at 55 ºC. Aliquots of each reaction were characterized on a 1.5% agarose gel with 0.25 µg/mL ethidium bromide (Bio-Rad) by comparing to a DNA ladder (Quick-Load PCR Marker, NEB).

In vitro transcription reactions with T7 RNA polymerase

For each construct, a plasmid with a confirmed DNA template insert was selected to be linearized by digesting with a restriction enzyme that would generate a blunt-end cleavage at the 3' end of the template. The labels given to each plasmid were: pA1, pA6, pB4, pC1, pC3, pD4, pE5 and pF1. Plasmids pA1 and pB4 were digested with 10 units of PmeI (NEB). Plasmids pA6 and pC3 were digested with 20 units of SspI (NEB) to generate the truncated DNA templates for the RNA constructs A* and C*. Plasmids pC1, pD4, pE5 and pF1 were digested with 20 units of DraI (NEB). The 50 µL reactions contained 25 µL of the previously purified plasmid (~10 µg) and were incubated overnight at 37 ºC. The digests were purified with a Wizard SV Gel and PCR Clean-Up System (Promega) kit. Linearized plasmids were characterized in a 1.2% agarose TAE gel with 0.25 µg/mL ethidium bromide (Bio-Rad) by comparing to 1 µg of the DNA ladder Phi-X174 HaeIII digest (NEB) and 0.5 µg of 100 bp DNA Ladder (NEB).

Initial 10 µL transcription reactions were carried at 37 ºC ranging from 6-7 hours in reaction buffer (NEB), varying concentrations of each NTP (NEB) depending on the nucleotide sequence of each RNA construct, with 0.1 µg/µL linearized template plasmid DNA, 25 units of T7 RNA polymerase (NEB), and varied concentrations of magnesium chloride ranging from 6 mM to 45 mM. On subsequent 12 µL reactions, conditions that ensured protection from ribonuclease activity were tested. Reactions were incubated for 6-7 hours at 37 ºC with 8 mM of each NTP (NEB), 0.1 µg/µL linearized template plasmid DNA, 25 units of T7 RNA polymerase (NEB), reaction buffer (NEB) and the optimal concentrations of magnesium chloride derived from the results of the initial reactions. A separate set of reactions under the same conditions also included 1 unit of RNasin (Promega), an inhibitor for the ribonucleases RNase A, B and C. After verifying the results, RNasin was included in reactions to synthesize RNA constructs B, C* and F.

After completion, all transcription reactions were brought to 100 µL total volume and treated at 37 ºC with 4 units of DNase I (NEB) for at least 30 minutes (30-80 minutes). Aliquots (4 µL) of initial reactions were screened before purification in 1.5 or 1.6% agarose-TAE gels with 0.25 µg/mL ethidium bromide and TAE running buffer. Samples were loaded with gel loading dye and were run for about 50-55 minutes at 80 V. The size of the RNA products was compared to 70 ng of RiboRuler Low Range RNA Ladder (Thermo Scientific). The RNA was purified using RNeasy Mini Elute and RNeasy Midi kits (QIAGEN) and by following their specific protocols. Transcription reactions and purification steps were also characterized for further optimization of *in vitro* transcription protocols using 4-20% polyacrylamide Tris-glycine gels (Invitrogen) that were run at 200 V for about 50 minutes using Tris-glycine running buffer (Invitrogen) and stained with SYBR Green II (Invitrogen). Samples were heated for 5 minutes at 37 ºC in denaturing RNA loading dye with formamide (Thermo Scientific) before loading onto the gel. RNA gel bands were compared against standards (70 ng per band) in the RiboRuler Low Range RNA Ladder (Thermo Scientific) and DNA bands were compared to 300 ng of the PCR Marker DNA ladder (NEB).

Ribonuclease T_1 digests

Ribonuclease T_1 digests (Thermo Scientific) were done to reveal areas of protection within the constructs. Each purified RNA construct (1 µg) was digested in separate 10 µL reactions with 0.2 and 1 units of RNase T_1 (Thermo Scientific). Samples were incubated for 5 minutes at 37 ºC after which the whole reactions were stopped with 50 mM dithiothreitol. Ten microliters of the denaturing Gel Loading Buffer II (Ambion) containing formamide and SDS was added and the solutions were incubated for 5 minutes at 65 ºC, placed in ice and half of the solution (10 µL) was loaded to a gel for analysis. Polyacrylamide 4-20% gradient Tris-glycine gels (Invitrogen) were run at 200 V for 45 minutes in Tris-glycine native gel running buffer (Invitrogen). Gels were stained in a 1:10,000 dilution of SYBR Green II dye (Invitrogen), rinsed with water and the RNA gel band migration and intensity were compared to the migration of 140 ng per band in the RiboRuler Low Range RNA Ladder (Thermo Scientific).

Differential scanning fluorimetry

RNA constructs were screened by differential scanning fluorimetry (DSF)[45] to assay thermal stability using the fluorescent dye SYBR Green II (Invitrogen) to detect structural changes that were monitored in a CFX96 Touch Real-Time PCR system (Bio-Rad). Data were processed and analyzed in Excel (Microsoft) and using the CFX Maestro Software (Bio-Rad). Initial screens for RNA constructs A, A*, B, C, C*, D and F were conducted in 20 mM Tris HCl, pH 7.5 with 10 mM potassium chloride and varying concentrations of magnesium chloride (0.5 mM and 2.5 mM). Sample volumes were 50 µL and contained approximately

Table 1 DNA templates for RNA constructs.

Code	Name of Construct	Forward Primer	Reverse Primer	Linearization Restriction Enzyme
A	Full Length Pel (TSS-terminator)[a] (560020..560522)	1	5	PmeI
A*	5' UTR and RBS (560020..560186)[b]	N/A	N/A	SspI
B	5' UTR (With TSS and no RBS) (560020..560139)	1	6	PmeI
C	RBS and *sagA* (560154..560321)	4	8	DraI
C*	RBS only[c] (560154..560186)	N/A	N/A	SspI
D	3' UTR with Terminator (560329..560522)	2	5	DraI
E	3' UTR (No Terminator) (560329..560434)	2	7	DraI
F	Terminator (560430..560522)	3	5	DraI
G	Alternate Full Length Pel (RBS-Terminator)[a] (560154..560522)	4	5	
H	Alternate Full Length Pel (TSS-Stop)[a] (560020..560321)	1	8	

[a] Isolated from genomic DNA library.
[b] Generated by cleaving the plasmid with the DNA template for RNA construct A with SspI.
[c] Generated by cleaving the plasmid with the DNA template for RNA construct C with SspI.

Table 2. PCR primers for isolation and amplification.

Code	Region/RNA Constructs	Primer Sequence[a,b,c,d]	Type of Primer	Restriction Enzyme for Cloning
1	Pel with TSS Constructs A 560020	CCG GAC *GGA TCC* TAA TAC GAC TCA CTA TAG TGA TAA GAA CTA GAT AGT TGT TGT GTT AC	5' UTR forward primer	BamHI
2	Pel and 3' UTR 560329	CCG GAC *GGA TCC* TAA TAC GAC TCA CTA TAC TAT TTA GCA TCT CTA TGT GAT AGT GA	3' UTR forward primer	BamHI
3	Pel and Terminator 560430	CCG GAC *GGA TCC* TAA TAC GAC TCA CTA TAG TTA GAA AAC ATG AGA CAA AAG TAA TCT TC	Terminator forward primer	BamHI
4	Pel and RBS 560154	CCG GAC *GGA TCC* TAA TAC GAC TCA CTA TAG GAG GTA AAC CTT ATG TTA AAA TTT ACT TC	Forward primer	BamHI
5	Pel 3' UTR and Terminator 560522	CCA CGG *AAG CTT* G**TT TAA AC**A TAT AGT AAT TAG CAG GTA CTG TCT AGT ACC	Reverse primer for end of 3' UTR	HindIII
6	Pel 5' UTR 560139	CCA CGG *AAG CTT* G**TT TAA AC**T GAT AAG AAC ACG AGT CGT TTA TTT TTA ACC	Reverse primer for 5' UTR	HindIII
7	3' UTR (without Terminator) 560434	CCA CGG *AAG CTT* G**TT TAA AC**T AAC ACC TCT ATG CCA TTC ATC TAT AAT CAG	Reverse primer for 3' UTR after stop codon	HindIII
8	Terminator 560321	CCA CGG *AAG CTT* G**TT TAA AC**C TGG CGT ATA ACT TCC GC	Reverse primer for *sagA* coding region	HindIII

[a] Single-lined underlined sequence corresponds to the T7 promoter sequence followed by a heterologous nucleotide.
[b] Italicized sequences correspond to recognition restriction sites for cloning.
[c] Double-lined underlined sequence corresponds to the recognition sequence for PmeI.
[d] Bold letters denote the recognition sequence for DraI.

1 µg of RNA. Each run started at 5 ºC and data points were collected every 0.5 ºC with a ramp time of 30 seconds between plate readings until reaching 95 ºC. Further solution conditions were assayed using fifteen selected reagents from the Nucleix Suite (QIAGEN). These solutions were diluted five-fold for the assays. Three solutions were selected based on optimal melting temperatures and shape of the melting curves for future structural studies. The first solution (Solution 1) contained 10 mM MES, pH 6.0, 20 mM ammonium acetate, 1 mM magnesium sulfate and 120 mM sodium chloride. The second solution (Solution 2) consisted of 10 mM MES, pH 6.0, 20 mM potassium chloride, 2 mM magnesium chloride and 2%(v/v) PEG 400. The third solution (Solution 3) was composed of 10 mM sodium cacodylate, pH 6.0, 2 mM magnesium chloride and 200 mM lithium sulfate.

RNA in silico characterization

Annealing parameters for PCR reactions were determined using ApE (v2.0.55)[46] and OligoCalc.[47] Secondary structure predictions were done with mfold[37-39] and RNAfold[40] using 2004 and 2007 RNA parameters. FARFAR[34-36] in Rosetta was used to calculate 3D models that were visualized with PyMOL.[48] Figures were made using Forna[49] and PyMOL.[48] TargetRNA2/3[50] and CopraRNA[51] were used to identify possible RNA targets for Pel sRNA interactions and to identify regions within Pel with a higher probability of been involved in RNA-RNA interactions.

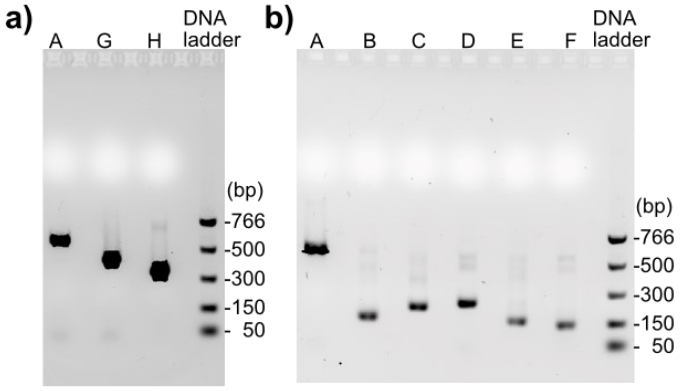

Fig. 3. DNA isolation from *S. pyogenes* HSC5 by PCR. **a)** The PCR-derived DNA templates for constructs A, G, H were isolated from genomic DNA. **b)** The purified PCR-derived DNA template for A, was amplified and used as PCR template for constructs B, C, D, E and F. The size of each band was confirmed by comparison to bands in the DNA ladder (50 to 766 bp) Quick-Load PCR marker (NEB). Characterization was done in a 1.5% agarose gels in TAE buffer. Gels were run at 80 V and 1 µL aliquots of each reaction before purification were loaded.

Results and Discussion

Isolation of the pel/sagA gene from streptococcal genomic DNA

DNA templates for various RNA constructs were designed based on the secondary structure predictions of Pel/*sagA* using mfold and RNAfold (Fig. 2).[37-42] The DNA sequence (Table 1) containing *pel*/*sagA* was isolated by PCR from the genomic DNA of *S. pyogenes* strain HSC5[33] using primers (Table 2) derived from the published DNA sequence. Three larger initial DNA constructs, constructs A, H and G in Table 1, were designed to ensure the isolation of the *pel/sagA* gene from genomic DNA. All three PCR reactions yielded positive results as determined from comparison to a DNA ladder (Fig. 3a). The DNA sequence for construct A was selected to be used as PCR-template for further PCR amplification and to generate the smaller DNA templates for constructs A-F (Fig. 3b). All PCR products contained the predicted size and were purified.

Design of DNA plasmid templates for the Pel/sagA RNA constructs

The PCR-derived DNA templates were inserted into pUC18 plasmids (Novagen). After amplification in *E. coli* bacterial cultures, the plasmids were isolated and were screened by PCR for the presence of the DNA templates using the same forward and reverse primers that were used for cloning (Fig. 4a-c). The isolated plasmids were referred as pA, pB, pC, pD, pE and pF followed by the colony number. Six colonies were picked for each construct. Five of the six colonies selected for construct A contained plasmids that were positive for the DNA template: plasmids pA1-2 and pA4-6 (Fig. 4a). All the colonies screened for the presence of plasmids with DNA templates for constructs B, C, D and E were positive: pB1-6, pC1-6, pD1-6, pE1-6 (Fig. 4a-c). Five of the six colonies selected for construct F contained plasmids positive for the DNA template: plasmids pF1-5 (Fig. 4c).

Selected plasmids that were positive for the presence of the DNA template construct sequences were further characterized by PCR amplification using a M13/pUC reverse primer and the reverse primer specific to each target construct in order to rule out concatenation of the insert during cloning (Fig. 4d). All were positive for the presence and size of the corresponding DNA template by PCR except for plasmids pB3, pE4 and pF3. Confirmation of positive plasmids was also carried by digestion with restriction enzymes that flanked the regions containing the target DNA sequences (Fig. 5). All digested plasmids were positive for their respective DNA template sequences except for plasmids pB3, pE4 and pF3. Plasmids that contained the correct DNA sequences were linearized with a blunt-end cutting restriction enzyme, which defined the 3' end of the RNA construct to be produced (Table 1). The linearized plasmids were characterized by gel electrophoresis (Fig. 6). The

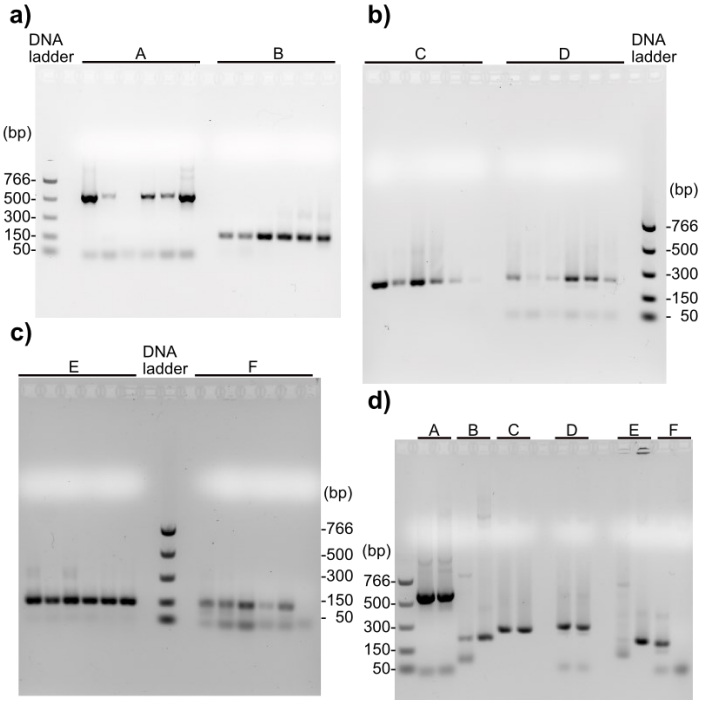

Fig. 4. Plasmids isolated from individual colonies were screened by PCR. Purified plasmids were screened for the presence of template DNA. **a)** Five out of six selected plasmids were positive for the DNA template for construct A (pA1-2 and pA4-6) and all the plasmids screened contained the DNA template for construct B (pB1-6). **b)** All the plasmids screened contained the DNA template for constructs C and D (pC1-6 and pD1-6, respectively). **c)** All the plasmids screened contained the DNA template for construct E (pE1-6). Only the first five plasmids screened contained the DNA template for construct F (pF1-5). **d)** PCR screen of purified selected plasmids for confirmation using a M13 reverse primer to replace each forward primer and the reverse primer specific to each construct. The plasmids screened from left to right in the gel were: pA1, pA6, pB3, pB4, pC1, pC3, pD4, pD5, pE4, pE5, pF1 and pF3. Only pB3, pE4 and pF3 appear negative for the presence of DNA template sequences for constructs E and F, respectively. All bands were compared against the Quick-Load PCR marker (NEB). Characterization was done in 1.5% (A and D) and 1.6% (B and C) agarose gels in TAE buffer. Gels were run at 80 V for 40-50 minutes.

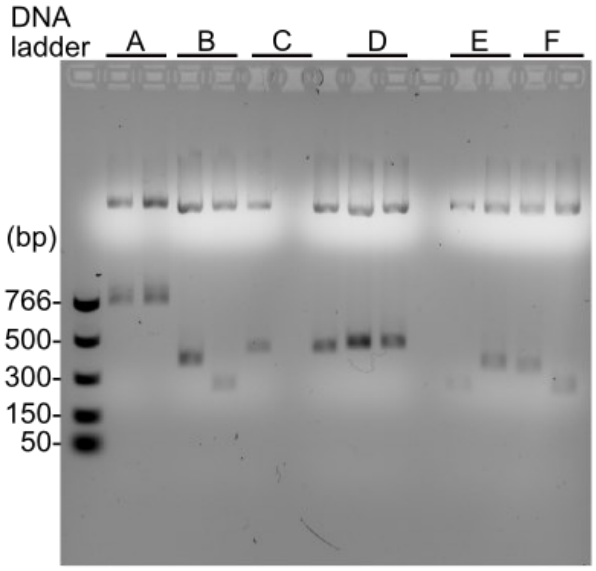

Fig. 5. Enzymatic digestion of plasmids isolated from different bacterial colonies. Aliquots of each plasmid were digested with 10 units of NdeI (NEB) and 10 units of EcoRI (NEB) to confirm that they contained the correct size DNA template. Double digests of pA1, pA6, pB4, pB3, pC1, pC3, pD4, pD5, pE4, pE5, pF1 and pF3 were loaded from left to right onto the gel leaving a space between pC1 and pC3 and between pD5 and pE4. Plasmids pB3, pE4 and pF3 appear negative for the presence of DNA template for constructs B, E and F, respectively. All bands were compared against the Quick-Load PCR marker (NEB).

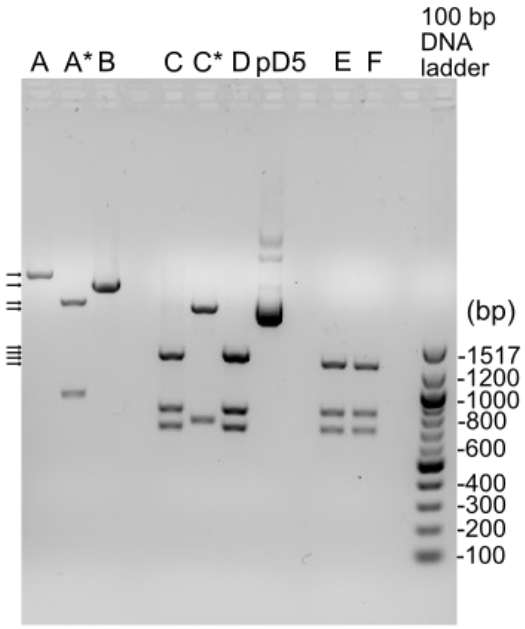

Fig. 6. Linearized plasmids with DNA templates for *in vitro* transcription. Selected plasmids were digested and then purified to be used as DNA templates in transcription reactions with T7 polymerase. Loading dye was added to aliquots of each reaction, which were then loaded onto a 1.2% agarose gel in TAE buffer that was run for 40 minutes at 80 V. Loaded from left to right digests of: pA1, pA6, pB4, pC1, pC3, pD4, pE5 and pF1. pA1 and pB4 were digested with PmeI, pA6 and pC3 were digested with SspI to generate the DNA templates for the RNA constructs A* and C*. DraI was used to digest pC1, pD4, pE5 and pF1. One of the positive plasmids for construct D, pD5, was used as an uncut plasmid control. Denoted with arrows are the bands that correspond to the DNA templates to produce the RNA constructs**.** The bands that correspond to fragments that contain the DNA templates were identified by comparing to bands in the 100 bp DNA Ladder (NEB).

plasmids were purified prior to using them as DNA templates for transcription reactions. Digestion of plasmids pA1 and pB4 with PmeI resulted in linearized plasmids to be used to produce the RNA constructs A and B. The restriction enzyme SspI was used to cleave plasmids pA6 and pC3, which created the truncated DNA templates for the synthesis of RNA constructs A* and C* (Fig. 6). Upon digestion two fragments were generated with the larger fragments containing the T7 promoter and DNA template sequences to be used for transcription. Digestion of plasmids pC1, pD4, pE5 and pF1 with DraI resulted in several fragments with the larger fragments containing the T7 promoter and DNA template sequences used to produce RNA fragments C, D, E and F, respectively.

Synthesis of the Pel/sagA RNA constructs

The linearized templates were used to synthesize the RNA constructs by *in vitro* transcription with T7 RNA polymerase. The RNA product sequence was used to calculate the optimal proportion of each NTP for each transcription reaction. Various concentrations of magnesium chloride were tested to identify conditions that resulted in higher yields of RNA. Initially, reactions included 6, 18, and 36 mM magnesium chloride (Fig. 7a-c). In order to identify the most favorable concentration of magnesium, the reaction mixtures were analyzed by agarose gel electrophoresis (Fig. 7). All three concentrations of magnesium chloride produce similar quantities for RNA constructs A, A*, B, and C (Fig. 7a-b). More significant effects on transcript yield were observed for RNA constructs C*, D, E and F (Fig. 7b-c). Increasing concentrations of magnesium enhanced the transcription of C*, D and E with the highest yield at 36 mM magnesium chloride. In the case of RNA construct F, the highest yield was obtained with 18 mM magnesium chloride (Fig. 7c). Due to the low yield for RNA construct E, additional concentrations of magnesium chloride were tested and the best yield was found with 18 mM magnesium chloride (Fig. 7d).

The most favorable conditions for the transcription reactions of each RNA construct were used to ensure reproducibility. At each step from transcription through purification, the RNA products were further characterized using the higher resolution and sensitivity of polyacrylamide gels stained with SYBR Green II (Fig. 8) to detect either aborted RNA transcripts or runoff transcription. The staining also revealed the DNA template in the transcription reactions allowing for observation of contamination due to uncut plasmid after linearization, which would reduce the yield of the RNA product. Due to inconsistent yields in some reactions, the protocols were further optimized by addition of RNase inhibitors to decrease the possibility of nuclease degradation (Fig. 8a-b). Aliquots of reactions treated with DNase I with (Fig. 8b) and without inhibitor (Fig. 8a) demonstrated no reduction in the production of transcripts with RNase inhibitors.

Secondary structure of Pel/sagA assayed by RNase T_1 digests

Secondary structure can protect RNA from digestion with RNase T_1, which cleaves phosphodiester bonds after guanine residues in single-stranded RNA. Paired regions are protected but external loops, internal loops and bulges make these regions accessible. The pattern of cleavage product sizes can be used to map the RNA secondary structure. RNase T_1 digests were carried out to assess the secondary structure and structural stability of RNA constructs A, A*, B, C, C*, D, E and F (Fig. 9a-b). Each of the purified RNA constructs was digested using increasing concentrations of RNase T_1. Based on secondary structure predictions, the number of expected fragments larger than 10 nucleotides would be: 18 for construct A, 8 for construct A*, 5 for construct B, 8 for construct C, 2 for construct C*, 6 for construct D, 5 for construct E, and 2 for construct F.[52]

After limited digestion with 0.2 units of RNase T_1, most RNA constructs remained intact with some slightly lower molecular weight bands forming. After incubation in 1 unit of RNase T_1, lower molecular bands form while some remaining intact constructs remain in the case of RNA constructs A, A*, C, D, E and F. RNA construct B showed the highest protection to digestion. Digestion of RNA construct C* could not be assessed due to the resolution of the gel, however construct C showed moderate protection upon digestion (Fig. 9a). Construct F contains a Rho -independent terminator, which is proposed to regulate transcription of the downstream genes in the *sag* operon. Based on this proposed function and the low GC content of the paired region, the terminator

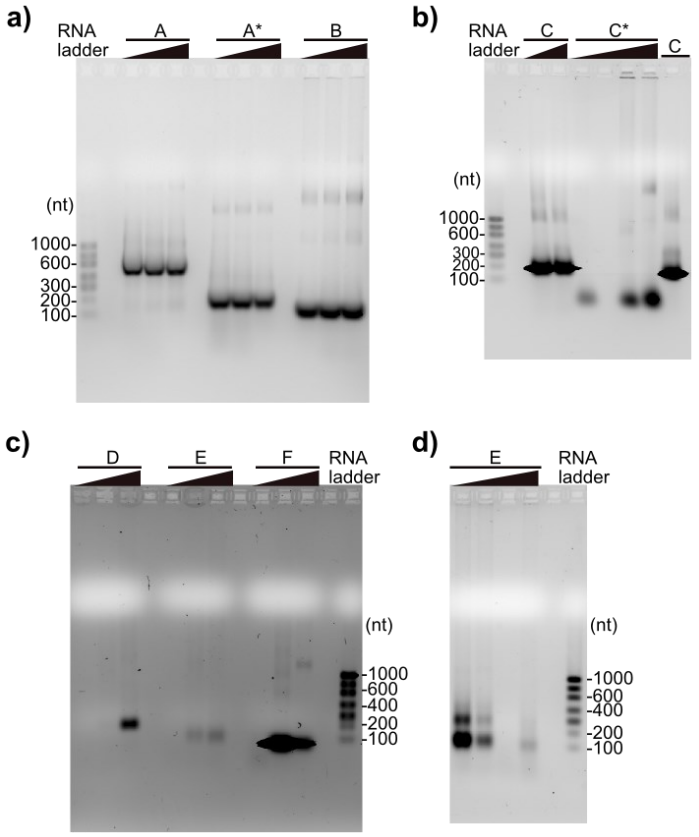

Fig. 7. Optimization of initial transcription reactions to determine the ideal concentration of magnesium chloride. **a)** Reactions with increasing concentration of magnesium chloride ($MgCl_2$): 6, 18, and 36 mM producing A, A*, and B RNA in a 1.5% agarose TAE gel. **b)** Reactions with increasing concentration of $MgCl_2$ producing C and C* RNA. From left to right in a 1.6% agarose TAE gel, 6 and 18 mM $MgCl_2$ for transcription of C, followed by reactions with 6, 18, and 36 mM $MgCl_2$ producing C* RNA. An aliquot of a reaction with 36 mM $MgCl_2$ for C RNA was loaded on the last lane. **c)** Reactions with increasing concentration of $MgCl_2$: 6, 18, and 36 mM producing D, E, and F RNA in a 1.6% agarose TAE gel. **d)** Reactions with increasing concentration of $MgCl_2$: 18, 27, 36, and 45 mM producing E RNA in a 1.5% agarose TAE gel. The size of the RNA products in all gels was compared to 70 ng of a RiboRuler Low Range RNA Ladder (Thermo Scientific).

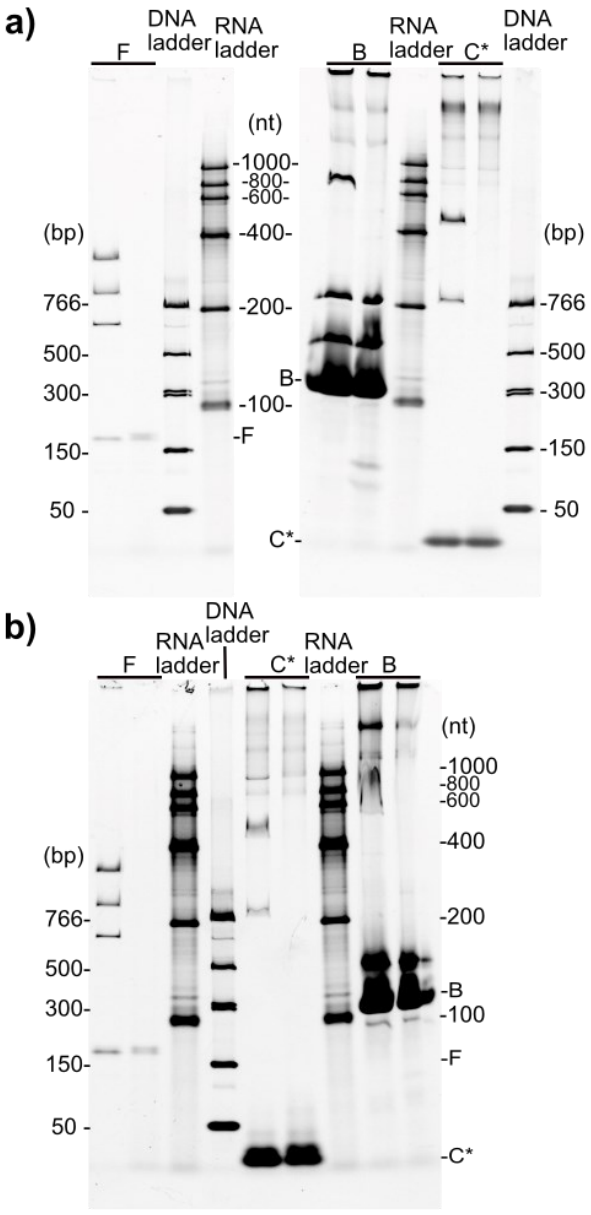

Fig. 8. Addition of RNasin inhibitor to transcription reactions did not affect the synthesis of RNA constructs. From left to right, before and after treatment with DNase I: **a)** reactions for F, B and C* in 18.5, 36 and 18.5 mM $MgCl_2$, respectively; and **b)** reactions including RNasin inhibitor for F, C* and B 18.5 mM, 18.5 and 31 mM $MgCl_2$, respectively. RiboRuler Low Range RNA Ladder (Thermo Scientific) and Quick-Load PCR marker (NEB) were used as size markers.

would likely contain a highly unstable structure. However, construct F showed a certain level of protection with the presence of medium size bands consistent with incomplete digestion due to transient formation of the terminator structure (Fig. 9b).[30-32]

RNA structure of Pel/sagA assayed by differential scanning fluorimetry

Differential scanning fluorimetry (DSF) can be used to evaluate the presence of structure in RNA using a fluorescent indicator dye to determine the melting temperatures associated with thermal denaturation of local secondary structure and long range tertiary interactions.[45] Initial results with increasing concentration of magnesium provided an estimate of the stability of secondary structures in each construct (Fig. 10 and Fig. 11). From the differential scanning fluorimetry assays of RNA construct A that contains the open reading frame for *sagA* and both the 5' and 3' UTR (Fig. 2), it is evident that the full-length transcript of Pel/*sagA* contains structured regions (Fig. 10). The melting temperature of construct A increased with increasing concentration of magnesium. A similar pattern was observed for construct A* that contains the 5' UTR with the ribosome binding site (RBS) and the AUG start codon. Both constructs A and A* have melting temperatures of 52.0, 55.0, and 59.5 ºC in 0.0, 0.5, and 2.5 mM magnesium chloride, respectively. Constructs A and A* showed multiple inflection points in the first derivative versus increasing temperature graphs in the absence of magnesium with a lower melting temperature of 52.0 ºC, which is consistent with the stabilization of tertiary structure in the presence of magnesium. Construct D comprising the 3' UTR with a rho-independent terminator (Fig. 2), had a melting temperature of 55.0 ºC in 0.5 mM magnesium chloride. The melting curves of constructs A, A*, and D indicate the presence of structural elements in all regions of the full-length Pel/*sagA* transcript.

RNA construct B encompasses the 5' UTR and is predicted to include a branched stem-loop structure that could influence the transcription of *sagA* or interact with other mRNA transcripts associated with virulence that are targeted by the Pel sRNA. From the differential scanning fluorimetry assays of RNA construct B that contains the 5' UTR (Fig. 11), it is evident that structural elements are present in this region with melting temperatures of 52.0, 55.0 and 60.0 ºC in the absence of magnesium and at 0.5 and 2.5 mM magnesium chloride, respectively. The melting temperatures measured are consistent with stable regions. The presence of structured regions in the 5' UTR could be associated

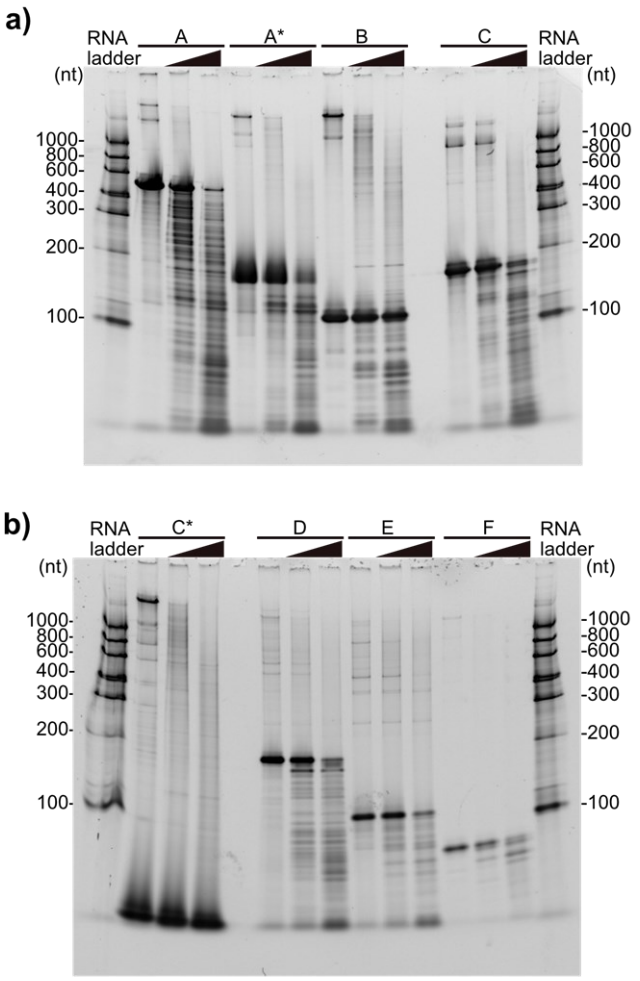

Fig. 9. RNase T_1 digests to probe RNA secondary structure. The secondary structure of RNA constructs of Pel/sagA **a)** A, A*, B and C, and **b)** C*, D, E and F was assayed by RNase T_1 digestion with 0, 0.2 and 1 units of enzyme added to each reaction. The gels were stained with SYBR Green II and RNA migration was compared to a RiboRuler Low Range RNA Ladder (Thermo Scientific).

with increased stability of Pel/*sagA* and affect translation. Construct C contains both the ribosome binding site (RBS) and AUG start codon within a predicted stem-loop structure (RNA construct C*) that could be involved in interactions that would inhibit the translation of *sagA* by blocking access to the Shine-Dalgarno sequence of the RBS. The graphs of the first derivative were broad with multiple peaks for construct C that contained the ribosome binding site, and the open reading frame for *sagA*. The melting temperature of construct C appeared to remain the same in the absence of magnesium and with 0.5 and 2.5 mM magnesium chloride with the maximum melting temperature peak consistently at 58.5 ºC (Fig. 12). For comparison, thermal melt of construct C* containing only stem-loop P1 in C with the RBS sequence and AUG start codon (Fig. 11) showed a similar broad first derivative curve in 0.5 mM magnesium chloride with a shoulder peak at 58.0 ºC, which is consistent with both constructs C and C* containing the same predicted secondary structural elements (Fig. 12).

Another region of interest includes the rho-independent terminator and predicted upstream structural regions within the 3' UTR that are contained within RNA construct F (Fig. 11). The first derivative DSF curves for construct F were broad in 0.5 and 2.5 mM magnesium chloride. Inflection points were discerned primarily in the absence of magnesium. Construct F contains the rho-independent terminator (P3), a predicted stem-loop upstream of the terminator (P2), and a predicted stem region (P1) that joins these two elements together (Fig. 11). Results from RNase T_1 digestion showed protection from cleavage with two strong bands forming after digestion of this RNA construct (Fig. 9b), which is consistent with the presence of these two structural elements. However, results from DSF demonstrate instability in these structures, which is consistent with the role of a transient rho-independent terminator that allows for the read-through transcription of all the genes in the *sag* operon necessary for production of SLS.

Additional thermal stability assays for buffer and salt conditions that could be used for structural studies were conducted for RNA constructs B, C and F. The relative fluorescence changes with increasing temperature and the first derivative with respect to increasing temperature were used to identify the melting temperatures for these RNA constructs when using different solution conditions (Fig. 13a-b). Data for RNA construct A was collected as a control for the presence of structure under the same conditions (Fig. 13a). Three conditions were identified that could be further refined for structural studies. The melting temperatures for RNA constructs A, B, C and F in Solution 1 (black curves) are 57.5, 58.0, 53.5 and 57.0 ºC, respectively (Fig. 13a-b). The melting temperature for RNA constructs A, B and C, were determined to be 62.0, 62.5 and 59.0 ºC, respectively, when data were collected in Solution 2 (purple curves). Data collected in Solution 3 (blue curves) resulted in melting temperatures of 60.0, 61.0 and 51.0 ºC for RNA constructs A, B and C, respectively (Fig. 13a). The determination of melting temperatures for construct F using Solutions 2 and 3 was ambiguous and led to inconclusive results. All three of the optimal solution conditions from the Nucleix Suite (QIAGEN) are buffered at pH 6.0 and contain 1-2 mM magnesium concentrations (1 mM for Solution 1 and 2 mM for Solutions 2 and 3). As observed from initial DSF results (Fig. 10 and Fig. 12), increasing the concentration of magnesium resulted in higher melting temperatures for these constructs, with sharper transitions and measurable inflection points at higher magnesium concentrations, which is indicative of tertiary interactions.

Three-dimensional models of the Pel/sagA RNA

Three-dimensional modeling helped to identify potential sites for RNA-RNA and protein-RNA interactions within the constructs. The FARFAR algorithm[34-36] in the Rosetta suite was used to calculate 3D models that were visualized with PyMOL (Fig. 14).[48] These models were used to identify regions of Pel/*sagA* that may be involved in regulatory interactions. RNA construct B begins at the transcription start site and contains five discernible paired regions with two stem-loops (P4 and P5) based on secondary structure predictions (Fig. 11). Loop residues in the P4 stem-loop and bulges between the P3 and P4 stem regions could provide access to nucleotides for interactions. The P4 loop contains a six residue palindromic sequence (5'-GCUAGC-3') with the potential to be involved in inter-molecular interactions (Fig. 14a). Interactions through this loop would affect the structure of the 5' UTR of Pel/*sagA*. Target predictions using TargetRNA3 identified possible interactions between stem-loop P4 in Pel sRNA and the 5' UTR of *prtS* that codes for the CXC chemokine-degrading serine protease, SpyCep, also known as PrtS and ScpC. This is a cell envelope serine protease that degrades the chemokine interleukin 8 and inhibits neutrophil recruitment during necrotizing fasciitis.[53-55]

A stem-loop structure was predicted for the region harboring the Shine-Dalgarno sequence (Fig. 14b). Construct C contains the RBS and AUG start codon within the P1 stem-loop followed by structural elements (P2-P6) that span the coding region of *sagA*. The truncated RNA construct C* only contains the predicted P1 stem-loop. Results from DSF demonstrate that this region has structure, but is not very stable, which would be expected considering that access to the RBS and the initiation codon is necessary for expression of *sagA*. Regulation of the structural stability may be accomplished through interactions with proteins or RNA. Activation leading to protein synthesis could be regulated by base-pairing to an sRNA causing the dissolution of the stem-loop and allowing binding of the ribosome as it has been described in other systems.[56] However, activation could also occur as a result of endonuclease cleavage to remove inhibitory structural elements after conformational changes induced as a result of base pairing between the 5' UTR and an sRNA.[56] The predicted structural elements within the coding region of *sagA* may play a role in the regulatory sRNA function of Pel. Interactions with Pel sRNA were predicted using TargetRNA,

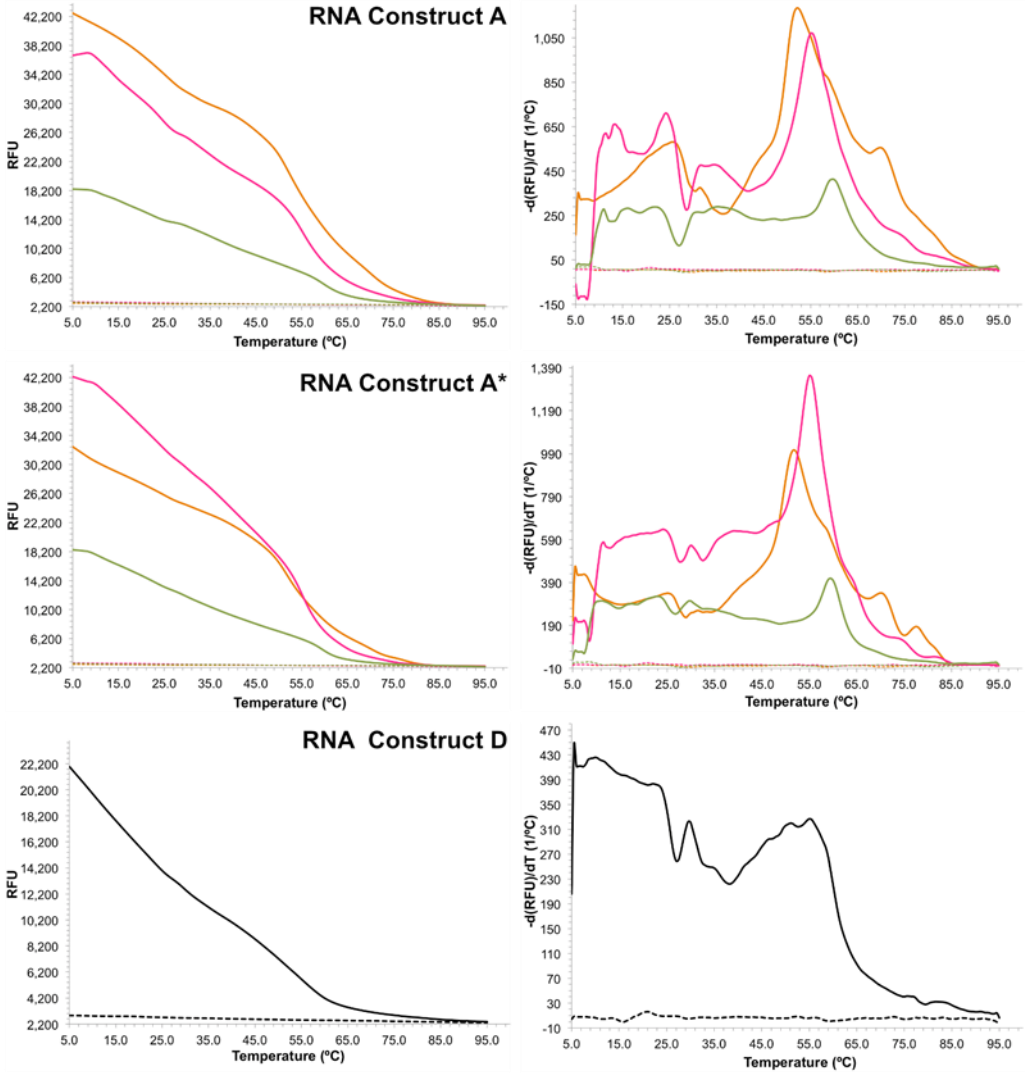

Fig. 10. Differential Scanning Fluorimetry (DSF) analysis probing structural stability of Pel/*sagA*, and the 5' and 3' UTR of Pel/*sagA*. Curves on the left correspond to the relative fluorescence changes with increasing temperature. Curves to the right correspond to the first derivative with respect to increasing temperature. Scans were done with increasing concentration of magnesium. For A and A*, orange, pink and green correspond to 0, 0.5 and 2.5 mM $MgCl_2$, respectively. Dotted lines correspond to control scans using the same conditions without RNA present. Data was only collected using 0.5 mM $MgCl_2$ for D. The melting temperature (T_M) of construct A is 52.0, 55.0 and 59.5 ºC in 0, 0.5 and 2.5 mM $MgCl_2$, respectively. The T_M of construct A* is 52.0, 55.0 and 59.5 ºC in 0, 0.5 and 2.5 mM $MgCl_2$, respectively. The T_M of construct D is 55.0 ºC in 0.5 mM $MgCl_2$.

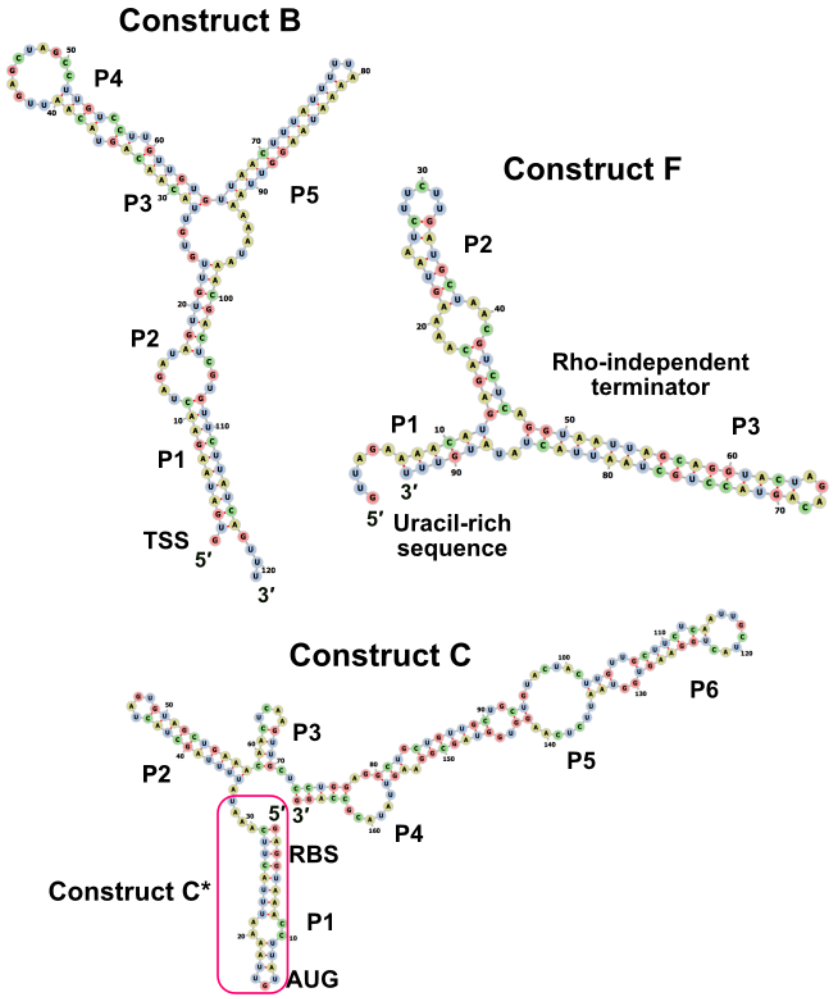

Fig. 11. Results from RNA secondary structure predictions using mfold were used as input to generate three-dimensional models of discrete regions with predicted secondary structural elements. The RNA construct B corresponds to the 5' UTR and begins with the transcription start site (TSS). Construct C includes the ribosome binding site (RBS) and the region coding for SagA beginning at the AUG start codon. RNA construct F includes the rho-independent terminator. C* consists of P1 stem-loop in C and is denoted with a box. Figures were generated using Forna.

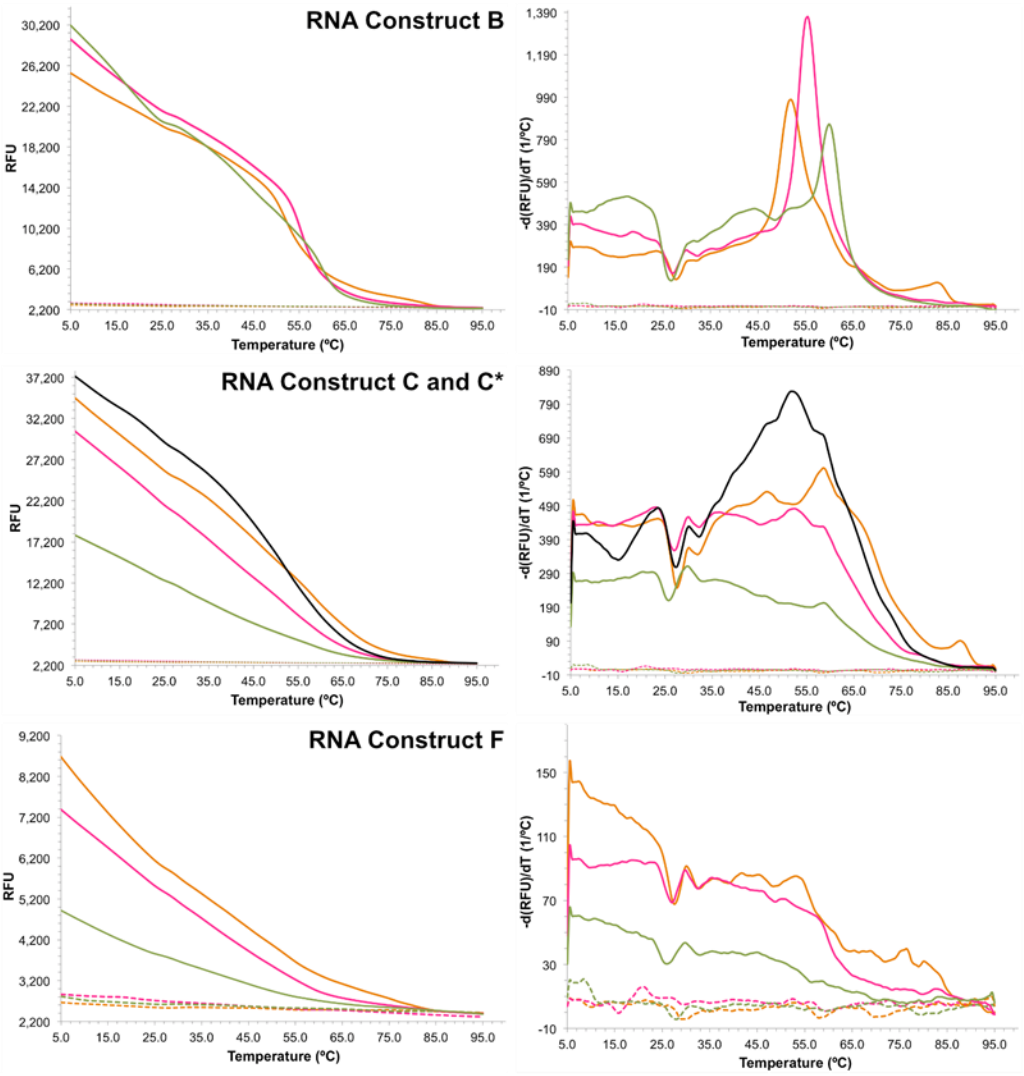

Fig. 12. DSF analysis probing structural stability of RNA constructs B, C, C* and F. Scans were done with increasing concentration of magnesium. For B, C and F, orange, pink and green correspond to 0, 0.5 and 2.5 mM $MgCl_2$, respectively. Dotted lines correspond to control scans using the same conditions without RNA present. Data was only collected using 0.5 mM $MgCl_2$ for C* (shown in black). The melting temperature (T_M) of construct B is 52.0, 55.0 and 60.0 ºC in 0, 0.5 and 2.5 mM $MgCl_2$, respectively. The T_M of construct C is 58.5, 52.0/58.5 and 58.5 ºC in 0, 0.5 and 2.5 mM $MgCl_2$, respectively. The T_M of construct C* is 52.0 ºC in 0.5 mM $MgCl_2$ with a minor peak at 58.0 ºC. The T_M of construct F is 53.0 and 50.5 ºC in 0 and 0.5 mM $MgCl_2$, respectively. The T_M of construct F in 2.5 mM $MgCl_2$ was inconclusive.

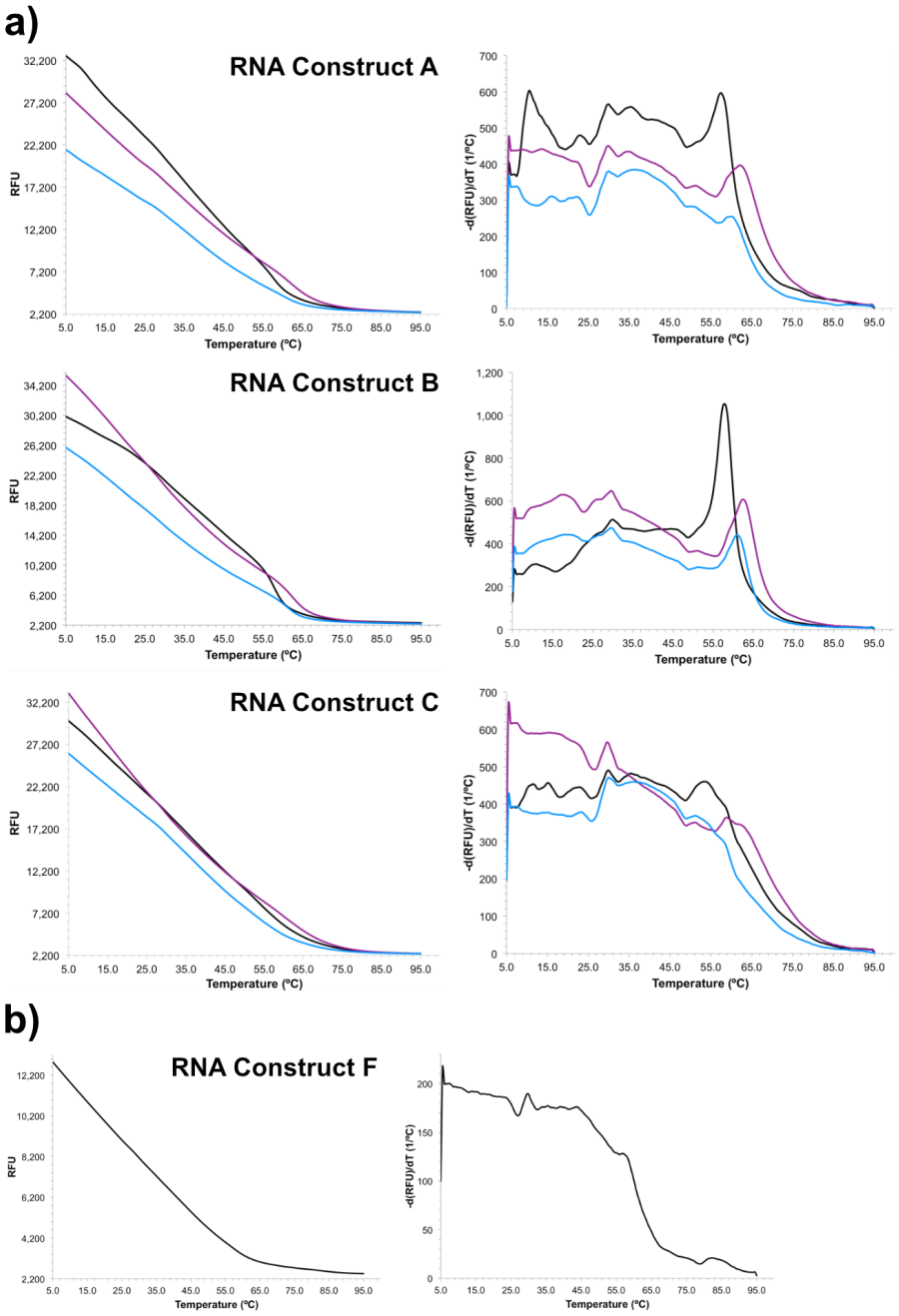

Fig. 13. Probing for conditions that could be used for structural studies and did not disrupt structural elements. Black, purple and blue curves correspond to solutions 1, 2 and 3, respectively. **a)** Data for RNA constructs A, B and C. **b)** Data for RNA construct F.

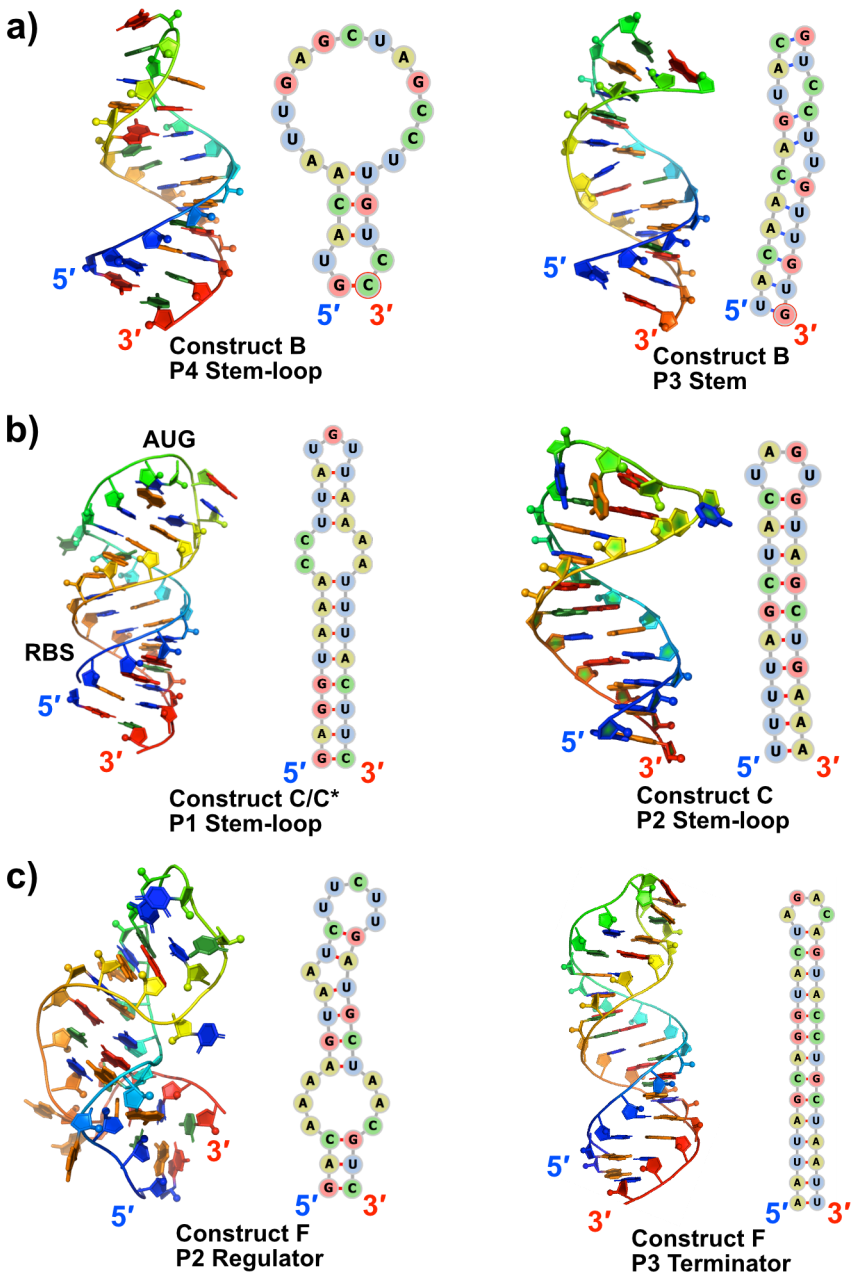

Fig. 14. Three-dimensional models of selected secondary structural elements in RNA constructs B, C, C* and F. Backbone colors use a rainbow gradient from blue (5') to red (3') for the backbone. The color for each nucleobase is: orange for adenosine, blue for uracil, red for guanosine and green for cytosine. Figures for the 3D models were generated using PyMol and the secondary structure representations used Forna. **a)** P3 stem and P4 stem-loop regions of RNA construct B. **b)** Stem-loops P1 and P2 of RNA constructs C and C*. **c)** Regulator and terminator structures within RNA construct F (P2 stem-loop and P3 stem-loop, respectively).

which predicted base-pairing between the internal loops in regions P4-P6 in the RNA construct C (Fig. 11) and the RBS region in SpyCep.

RNA construct F contains a stem region (P1) that branches into two stem-loops, P2 and P3 (Fig. 11). P3 is a Rho-independent terminator[16] and the adjacent P2 stem-loop may play a role in binding interactions with the 5' UTR of Pel/*sagA* or with other sRNA transcripts. The adjacent stem-loop P2 could act as a regulator affecting the stability of the P3 terminator structure (Fig. 14c). Relative stability was observed from limited digestion with RNase T_1 while thermal melting was consistent with dynamically unstable structures. This is consistent with a regulatory role for the terminator with a transient structure that allows for read-through transcription of downstream genes within the *sag* operon. Even though from the 3D model of P3 is a stable stem-loop structure, the presence of mostly A-U base pairs could account for its instability. Stem-loop P2 is also AU-rich and contains an internal loop and a bulge that would also increase the instability of this structure (Fig. 14c).

Predicted RNA-RNA interactions with Pel/sagA

In silico analysis using TargetRNA2 and 3,[50] and CopraRNA[51] were used to identify possible targets for Pel sRNA interactions. Potential mRNA interactions were present primarily within the 5' UTR of mRNA transcripts from other genes. Results from secondary structure predictions using mfold and RNAfold were correlated with regions predicted by TargetRNA2 to bind the mRNA transcripts of genes regulated by Pel. The 3D models of the structural elements reveal potential sites for interactions with the targeted mRNA transcripts, other regulatory sRNA transcripts, or RNA binding proteins that regulate gene expression. Some of the potential mRNA targets predicted by CopraRNA that are unique to these specific constructs were SAM-dependent RNA methyltransferase (class I), cell division protein FtsQ, iron ABC transporter permease, GlpR transcriptional regulator, and the translation initiation factor IF3. These potential targets could involve interactions between Pel/*sagA* and other virulence factors allowing *S. pyogenes* to adapt to various host environments during pathogenesis.

Conclusion

Results from these studies identified structural elements in Pel/*sagA* that could be involved in the regulation of transcription and translation. RNA constructs were designed using secondary structure predictions as a guide. PCR-derived DNA templates isolated from genomic DNA were used to synthesize RNA constructs by *in vitro* transcription containing different regions within Pel/*sagA*. Results from RNase T_1 digests and differential scanning fluorimetry support the presence of stable structural motifs within Pel/*sagA*. RNA constructs from the 5' UTR of Pel/*sagA* contained structural elements that were highly stable. RNA constructs comprising the RBS sequence and *sagA* coding region showed less stability. This inherent instability is consistent with the regulatory roles of these regions during translation. Similarly, RNA constructs from the 3' UTR were inherently unstable, which correlates with the role of an intrinsic rho-independent terminator in regulating transcription of the downstream genes of the sag operon. Characterizing the structure of the Pel/*sagA* sRNA and its interactions with other virulence factors will provide insight into virulence regulation by *S. pyogenes* that may contribute to life-threating diseases, such as necrotizing fasciitis. Further understanding of the role of Pel/*sagA* in virulence regulation is necessary to achieve the long-term goal of finding targets for the design of alternative therapeutics.

Acknowledgements

The authors thank Dr. Kyu Hong Cho at Indiana State University for the genomic DNA from *S. pyogenes* HSC5. The authors would also like to thank Kayla Calderon and Regan Finn for valuable discussions. The authors would like to thank South Carolina INBRE and Summer Coastal Research Experience (SCoRE) at CCU. This work was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under award number P20GM103499. Portion of the research was funded by the National Institute of General Medical Sciences of the National Institutes of Health under award number R15GM101603 (to B.M.L., K.H.C. and G.C.P.A), a CCU Professional Enhancement Grant 33-4844 (to B.M.L. and G.C.P.A.), start-up funds from the Gupta College of Science at CCU (to B.M.L.), a SAF Research Fellowship (to C.R.C), a CCU Women in STEM Research Fellowship (to S.G.N.) and funds from the Department of Chemistry at CCU (to C.R.C and S.G.N.). We thank Dr. Paul Richardson for access to a PCR thermocycler and the Department of Biology at CCU for access to a gel imaging system and qPCR instrumentation.

Notes and References

**Corresponding authors email: brianlee@coastal.edu and gperezal@coastal.edu*

1. Ralph AP, Carapetis JR. 2013. Group A streptococcal diseases and their global burden. In Chhatwal GS, eds, Host-Pathogen Interactions in Streptococcal Diseases, 2013 ed, Springer-Verlag, Germany, pp 1-27.
2. Cunningham MW. 2000. Pathogenesis of group A streptococcal infections. Clin Microbiol Rev 13: 470-511.
3. Walker MJ, Barnett T, C., McArthur JD, Cole JN, Gillen CM, Henningham A, Sriprakash KS, Sanderson-Smith ML, Nizet V. 2014. Disease manifestations and pathogenic mechanisms of group A Streptococcus. Clin Microbiol Rev 27: 264-301.
4. Vega LA, Malke H, McIver KS. 2016. Virulence-related transcriptional regulators of Streptococcus pyogenes. In Ferretti JJ, Stevens DL, Fischetti VA, eds, Streptococcus pyogenes Basic Biology to Clinical Manifestations, 2016 ed, The University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA, pp 271-322.
5. Hynes W, Sloan M. 2016. Secreted extracellular virulence factors. In Ferretti JJ, Stevens DL, Fischetti VA, eds, Streptococcus pyogenes Basic Biology to Clinical Manifestations, 2016 ed, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA, pp 323-352.
6. Leitch HA, Palepu A, Fernandes CMB. 2000. Necrotizing fasciitis secondary to group A streptococcus. Morbidity and mortality still high. Can Fam Physician 46: 1460-1466.
7. Kreikemeyer B, McIver KS, Podbielski A. 2003. Virulence factor regulation and regulatory networks in Streptococcus pyogenes and their impact on pathogen-host interactions. Trends Microbiol 11: 224-232.
8. Steiner K, Malke H. 2000. Life in protein-rich environments: the relA-independent response of Streptococcus pyogenes to amino acid starvation. Mol Microbiol 38: 1004-1016.
9. Betschel SD, Borgia SM, Barg NL, Low DE, De Azavedo JC. 1998. Reduced virulence of group A streptococcal Tn916 mutants that do not produce streptolysin S. Infect Immun 66: 1671-1679.
10. Lee SW, Mitchell DA, Markley AL, Hensler ME, Gonzalez D, Wohlrab A, Dorrestein PC, Nizet V, Dixon JE. 2008. Discovery of a widely distributed toxin biosynthetic gene cluster. Proc Natl Acad Sci 105: 5879-5884.
11. Datta V, Myskowski SM, Kwinn LA, Chiem DN, Varki N, Kansal RG, Kotb M, Nizet V. 2005. Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection. Mol Microbiol 56: 681-695.
12. Molloy EM, Cotter PD, Hill C, Mitchell DA, Ross RP. 2011. Streptolysin S-like virulence factors: the continuing sagA. Nat Rev Microbiol 9: 670-681.
13. Mitchell DA, Lee SW, Pence MA, Markley AL, Limm JD, Nizet V. 2009. Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S. J Biol Chem 284: 13004-13012.
14. Melby JO, Nard NJ, Mitchell DA. 2011. Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates. Curr Opin Chem Biol 15: 369-378.
15. Metelev MV, Ghilarov DA. 2014. Structure, function, and biosynthesis of thiazole/oxazole-modified microcins. Mol Biol 48: 29-45.
16. Nizet V, Beall B, Bast DJ, Datta V, Kilburn L, Low DE, De Azavedo JC. 2000. Genetic locus for streptolysin S production by group A streptococcus. Infect Immun 68: 4245-4254.
17. Maxson T, Deane CD, Molloy EM, Cox CL, Markley AL, Lee SW, Mitchell DA. 2015. HIV protease inhibitors block streptolysin S production. ACS Chem Biol 10: 1217-1226.
18. Flaherty RA, Puricelli JM, Higashi DL, Park CJ, Lee SW. 2015. Streptolysin S promotes programmed cell death and enhamces inflammatory signaling in epithelial keratinocytes during group A Streptococcus infection. Infect Immun 83: 4118-4133.
19. Flaherty RA, Donahue DL, Carothers KE, Ross JN, Ploplis VA, Castellino FJ, Lee SW. 2018. Neutralization of streptolysin S-dependent and independent inflammatory cytokine IL-1B activity reduces pathology during early group A Streptococcal skin infection. Front Cell Infect Microbiol 8: 211.
20. Pinho-Ribeiro FA, Baddal B, Haarsma R, O'Seaghdha M, Yang NJ, Blake KJ, Portley M, Verri WA, Dale JB, Wessels MR, Chiu IM. 2018. Blocking neuronal signaling to immune cells treats streptococcal invasive infection.

Cell 173: 1083-1097.
21. Tracey KJ. 2018. Neurons are the inflammatory problem. Cell 173: 1066-1068.
22. Kreikemeyer B, Boyle MD, Buttaro BA, Heinemann M, Podbielski A. 2001. Group A streptococcal growth phase-associated virulence factor regulation by a novel operon (Fas) with homologies to two-component-type regulators requires a small RNA molecule. Mol Microbiol 39: 392-406.
23. Li Z, Sledjeski DD, Kreikemeyer B, Podbielski A, Boyle MD. 1999. Identification of pel, a Streptococcus pyogenes locus that affects both surface and secreted proteins. J Bacteriol 181: 6019-6027.
24. Le Rhun A, Charpentier E. 2012. Small RNAs in streptococci. RNA Biol 9: 414-426.
25. Mangold M, Siller M, Roppenser B, Vlaminckx BJM, Penfound TA, Klein R, Novak R, Novick RP, Charpentier E. 2004. Synthesis of group A streptococcal virulence factors is controlled by a regulatory RNA molecule. Mol Microbiol 53: 1515-1527.
26. Biswas I, Germon P, McDade K, Scott JR. 2001. Generation and surface licalization of intact M protein in Streptococcus pyogenes are dependent on sagA. Infect Immun 69: 7029-7038.
27. Eberhard TH, Sledjeski DD, Boyle MDP. 2001. Mouse skin passage of Streptococcus pyogenes Tn917 mutant of sagA/pel restores virulence, beta-hemolysis and sagA/pel expression without altering the position or sequence of the transposon. BMC Microbiol 1: 1471-2180.
28. Mars RAT, Nicolas P, Denham EL, van Dijl JM. 2016. Regulatory RNAs in Bacillus subtilis: a Gram-positive perspective on bacterial RNA-mediated regulation of gene expression. Microbiol Mol Biol Rev 80: 1029-1057.
29. Ren G-X, Guo X-P, Sun Y-C. 2017. Regulatory 3' unstranslated regions of bacterial mRNAs. Front Microbiol 8: 1-6.
30. Ray-Soni A, Bellecourt MJ, Landick R. 2016. Mechanisms of bacterial transcription termination: all good things must end. Annu Rev Biochem 85: 319-347.
31. Chen J, Morita T, Gottesman S. 2019. Regulation of transcription termination of small RNAs and by small RNAs: molecular mechanisms and biological functions. Front Cell Infect Microbiol 9: 201.
32. Peters JM, Vangeloff AD, Landick R. 2011. Bacterial transcription terminators: the RNA 3'-end chronicles. J Mol Biol 412: 793-813.
33. Port GC, Paluscio E, Caparon MG. 2013. Complete genome sequence of emm Type 14 Streptococcus pyogenes Strain HSC5. Genome Announc 1: e00612-e00613.
34. Cheng CY, Chou F-C, Das R. 2015. Modeling complex RNA tertiary folds with Rosetta. Methods Enzymol 553: 35-64.
35. Das R, Karanicolas J, Baker D. 2010. Atomic accuracy in predicting and designing noncanonical RNA structure. Nat Methods 7: 291-294.
36. Lyskov S, Chou F-C, Conchúir SO, Der BS, Drew K, Kuroda D, Xu J, Weitzner BD, Renfrew PD, Sripakdeevong P, Borgo B, Havranek JJ, Kuhlman B, Kortemme T, Bonneau R, Gray JJ, Das R. 2013. Serverification of molecular modeling applications: the Rosetta online server that includes everyone (ROSIE). PLoS One 8: e63906.
37. Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406-3415.
38. Zuker M, Jacobson AB. 1998. Using reliability information to annotate RNA secondary structures. RNA 4: 669-679.
39. Waugh A, Gendron P, Altman R, Brown JW, Case D, Gautheret D, Harvey SC, Leontis N, Westbrook J, Westhof E, Zuker M, Major F. 2002. RNAML: a standard syntax for exchanging RNA information. RNA 8: 707-717.
40. Lorenz R, Bernhart SH, Siederdissen CHZ, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. Algorithms Mol Biol 6: 26.
41. Andronescu M, Condon A, Hoos HH, Mathews DH, Murphy KP. 2007. Efficient parameter estimation for RNA secondary structure prediction. Bioinformatics 23: i19-28.
42. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci USA 101: 7287-7292.
43. Kang SO, Caparon MG, Cho KH. 2010. Virulence gene regulation by CvfA, a putative RNase: the CvfA-enolase complex in Streptococcus pyogenes links nutritional stress, growth-phase control, and virulence gene expression. Infect Immun 78: 2754-2767.
44. Sambrook J, Russell DW. 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York, NY, USA.
45. Silvers R, Keller H, Schwalbe H, Hengesbach M. 2015. Differential scanning fluorimetry for monitoring RNA stability. ChemBioChem 16: 1109-1114.
46. Davis WM, Jorgensen EM. 2022. ApE, a plasmid editor: a freely available DNA manipulation and visualization program. Front Bioinform 2: 818619.
47. Kibbe WA. 2007. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res 35: W43-46.
48. The PyMOL Molecular Graphics System. Version 2.0 Schrödinger, LLC.
49. Kerpedjiev P, Hammer S, Hofacker IL. 2015. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. Bioinformatics 31: 3377-3379.
50. Kery MB, Feldman M, Livny J, Tjaden B. 2014. TargetRNA2: identifying targets of small regulatory RNAs in bacteria. Nucleic Acids Res 42: W124-W129.
51. Wright PR, Georg J, Mann M, Sorescu DA, Richter AS, Lott S, Kleinkauf R, Hess WR, Backofen R. 2014. CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. Nucleic Acids Res 42: W119-W123.
52. Madanecki P, Nozell S, Ochocka R, Collawn JF, Bartoszewski R. 2014. RNAdigest: A web-based tool for the analysis and prediction of structure-specific RNAse digestion results. PLoS ONE 9: e96759.
53. Kaur SJ, Nerlich A, Bergmann S, Rohde M, Fulde M, Zähner D, Hanski E, Zinkernagel A, Nizet V, Chhatwal GS, Talay SR. 2010. The CXC chemokine-degrading protease SpyCep of Streptococcus pyogenes promotes its uptake into endothelial cells. J Biol Chem 285: 27798-27805.
54. Hidalgo-Grass C, Mishalian I, Dan-Goor M, Belotserkovsky I, Eran Y, Nizet V, Peled A, Hanski E. 2006. A streptococcal protease that degrades CXC chemokines and impairs bacterial clearance from infected tissues. EMBO J 25: 4628-4637.
55. Edwards RJ, Taylor GW, Ferguson M, Murray S, Rendell N, Wrigley A, Bai Z, Boyle J, Finney SJ, Jones A, Russell HH, Turner C, Cohen J, Faulkner L, Sriskandan S. 2005. Specific C-terminal cleavage and inactivation of interleukin-8 by invasive disease isolates of Streptococcus pyogenes. J Infect Dis 192: 783-790.
56. Podkaminski D, Vogel J. 2010. Small RNAs promote mRNA stability to activate the synthesis of virulence factors. Mol Microbiol 78: 1327-1331.