Published in final edited form as:

Microb Pathog. 2009 June ; 46(6): 315–320. doi:10.1016/j.micpath.2009.03.003.

Genomic Analysis Identifies a Transcription Factor Binding Motif Regulating Expression of the Alpha C Protein in Group B *Streptococcus*

David C. Klinzing¹, Lawrence C. Madoff^{1,3,4,#}, and Karen M. Puopolo^{1,2,4,*,#}

¹Channing Laboratory, Brigham and Women's Hospital, 181 Longwood Avenue, Boston, MA 02115

²Department of Newborn Medicine, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115

³Division of Infectious Diseases, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115

⁴Harvard Medical School, Boston, MA 02115

Abstract

The virulence-associated alpha C protein (ACP) of Group B *Streptococcus* (GBS) facilitates the bacterial interaction with host epithelial cells. We previously demonstrated that phase-variable expression of ACP is controlled by variation in short-sequence repeat sequences present upstream of the promoter of *bca*, the gene encoding ACP. To determine if *trans*-acting transcriptional control also influences ACP expression, we developed an *in silico* prediction algorithm that identified a potential transcription-factor binding motif (TTT-N₆-ATAT) in the *bca* upstream region. *In vitro* reporter gene expression studies confirmed that this motif is required for full ACP expression, and DNA-binding assays with a GBS protein extract demonstrated that the predicted site is bound by a protein. This approach demonstrates the utility of *in silico* genomic predictive methods in the study of GBS regulatory mechanisms.

Keywords

Streptococcus agalactiae; group B *Streptococcus*; transcriptional regulation

1. Introduction

Group B *Streptococcus* (*Streptococcus agalactiae*, or GBS) is a Gram-positive human intestinal commensal organism that is a leading cause of infection in newborns, and in pregnant, elderly or immunocompromised adults [1]. Although multiple virulence factors have been identified as important to GBS disease pathogenesis [2], the regulation and expression of these

© 2009 Elsevier Ltd. All rights reserved

Corresponding author: Address correspondence and reprint requests to: Karen M. Puopolo, M.D., Ph.D. Channing Laboratory Brigham and Women's Hospital 181 Longwood Avenue Boston, MA 02115. Phone: 617-525-7737 Fax: 617-731-1541 Email: kpuopolo@partners.org.

[#]Dr. Madoff and Dr. Puopolo contributed equally to senior authorship of this manuscript.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

factors during different stages of infection remains poorly understood. The alpha C protein (ACP) is the prototype of a family of virulence-associated, repeat-containing surface proteins (the alpha-like proteins, or ALP's) that are expressed in most GBS strains examined to date [3,4]. ACP is a target of host immunity [5], and functions to promote uptake by and transcytosis across epithelial cells [6,7,8], interactions mediated by surface moieties on the host cell including glycosaminoglycans and integrins [6,7,8,9]. We have previously demonstrated that presentation of ACP to the host environment can be altered by two mechanisms. The antigenicity of ACP correlates with the number of 82-amino acid tandem repeat units within the protein [10,11]. Variation in repeat number occurs within the gene for the ACP (*bca*) by a *recA*-independent recombinatorial mechanism [12]. We have also demonstrated transcriptional regulation of ACP expression by phase-variable changes in short-sequence repeat sequences located upstream of the predicted -35 region [13]. This regulation likely occurs by a slip-strand mispairing mechanism that varies the promoter spacing. Both of these mechanisms involve the selection and expansion of isolates undergoing changes that proceed at a low baseline rate in a recombinatorial manner.

It remains unclear whether ACP expression is controlled in response to specific host environmental conditions by *trans*-acting, cognate transcriptional regulatory mechanisms, as has been demonstrated for other GBS virulence factors [12,14,15,16,17]. Deletion of the two-component regulatory system *csrR/S* (*covR/S*) in three different GBS strains results in either a slight increase in, or no observable change in, ALP expression [15,16]. We have constructed deletions in the two-component regulatory systems *ciaR/H*, *liaR/S* and *dltR/S* in GBS strain A909 with no effect on the baseline production of ACP (our laboratory, unpublished data). Analyses of the sequenced GBS genomes predict at least 17 two-component regulatory systems and over 100 genes with some proposed regulatory function, making a gene-by-gene search for regulators of ACP expression impractical [18,19]. Genomic computational analyses can be used to identify consensus transcriptional DNA binding sites once a set of regulated genes is identified by biologic means. Such an approach was used to predict a cognate regulatory site for CsrR/CovR [16]. We now describe the use of the converse approach - the *de novo* identification of potential transcription factor binding motifs within the upstream region of *bca* by genomic analyses followed by confirmation of the transcriptional role of this motif by biologic *in vitro* testing.

2. Results

2.1 Identification of potential transcription factor binding motif by genomic analysis

Comparative genomic methods have been developed for the prediction of potential transcription factor binding motifs [21,22]. Genomic sequences of eight GBS strains are now available; in each of these genomes, an ALP gene is present in an allelic position to that of *bca* in strain A909, with conserved upstream regions [20 and our own analysis], allowing us to utilize comparative genomic techniques to identify conserved regulatory sequences. We developed an algorithm based on that of Kellis, et. al. [21] and van Nimwegen, et. al. [22], similar to that utilized for the determination of a consensus binding site sequence for the CovR protein in GBS [16]. We began by analyzing the 250-bp intergenic region from the ATG start codon of *bca*, up to the ATG start codon of the *araC* gene (located immediately upstream of *bca*, in the opposite orientation) (Figure 1). Identified mini-motifs were expanded by one nucleotide at a time in either direction to generate full motifs. Conservation of the full motif was examined by comparing the upstream regulatory sequence of *bca* from six sequenced GBS strains. The mini-motif search resulted in the identification of the sequence TTT-N₆-ATA (where N₆ indicates any 6 nucleotides can be present) as one of the most frequently occurring motifs within the *bca* upstream regulatory region, occurring six times within the 250-bp intergenic region, four times on the plus strand and twice on the opposite (minus) strand. The

two occurrences on the minus strand overlap that of the plus strand, creating a palindrome-like motif of TTT-N₆-ATAT- N₆-AAA (Figure 1, Panel A). The six occurrences of the TTT-N₆-ATA mini-motif were perfectly conserved in all six genomes. In three of the four instances on the plus strand, we were able to expand the motif to the sequence TTT- N₆-ATAT while maintaining conservation between the six GBS genomes.

The full motif of TTT-N₆-ATAT occurs three times (on the plus strand) within the 250-bp intergenic region between *araC* and *bca*, but does not occur within the 3270 bp sequence of the *bca* gene itself, consistent with a non-random presence of this motif within the intergenic region ($p < 0.0001$ by two-tailed chi-square test comparing intergenic region to *bca* gene). We also analyzed the entire A909 genome for occurrences of the TTT-N₆-ATAT motif. We found a this motif occurred 2.2 times more frequently within predicted intergenic regions compared to predicted coding regions of the genome. Although the motif occurred at least once in the intergenic region of over 200 genes, it only occurred in clustered fashion (two or more occurrences) within the upstream region of 23 genes.

2.2 Biologic assessment of TTT- N₆-ATAT motif using reporter gene assay

We previously demonstrated that a reporter construct containing 455 bp of sequence upstream of the start codon for *bca* (pN2gfp) could drive expression of GFP *in vitro* [13]. This reporter plasmid consists of the GFP structural gene cloned into the promoterless multicloning site of plasmid pDL278, a gram-positive/gram-negative shuttle vector that is present in medium copy number in GBS [27]. The GFP gene cloned alone results in no GFP expression; the transcriptional activity of sequence cloned upstream of GFP is assayed by Western blot using a monoclonal antibody to GFP [13]. Although the assay is semi-quantitative, there is no background signal and relative levels of GFP expression are highly reproducible when compared to expression of a control protein. To determine the minimal sequence elements required for *bca* expression, a series of reporter constructs were generated containing progressively smaller stretches of *bca* upstream sequence fused to the gene for GFP (Figure 1A). The content of constructs containing the TTT-N₆-ATAT sites and the western blot analyses of GFP expression are shown in Figure 1, Panels A and B. Deletion of residues up to the start codon for the oppositely oriented *araC* (construct pN2Δ4) resulted in no change in expression of GFP compared to the full pN2gfp construct (data not shown). Construct pN2Δ5 does not contain the first three occurrences of the motif (two on the plus strand, one on the minus strand), but again no detectable change in expression of GFP is observed compared to pN2gfp. Construct pN2Δ6 does not contain the next occurrence of the motif and interrupts the palindrome resulting from the final occurrence on the minus strand; this construct results in no detectable GFP expression. To more precisely map the necessary elements for *bca* expression, five additional constructs (pN2Δ5A, pN2Δ5B, pN2Δ5C, pN2Δ5D and pN2Δ5E) were generated using PCR primers spaced approximately 6-10 bp apart in the region between pN2Δ5 and pN2Δ6 (Figure 1, Panel A). No detectable change in expression of GFP is observed using constructs pN2Δ5A, pN2Δ5B and pN2Δ5C compared to pN2gfp (data not shown). However, GFP expression was significantly diminished in construct pN2Δ5D (Figure 1, Panel B); construct pN2Δ5E was identical to pN2Δ5D (data not shown). We have previously mapped a predicted -35 and -10 region for *bca* as shown in Figure 1, Panel A [13]. These current results indicate that the sequence between the -124 and -103 positions is either required for proper structuring of the region to allow transcription to occur, and/or is the site of a required interaction with a DNA-binding transcription factor necessary for gene expression. Such a factor may bind in this region as a dimer, given the palindromic structure of the essential motif.

2.3 Site-directed mutagenesis of the TTT- N₆-ATAT motif

To determine if the spacing within the - 124 to -103 region, and/or the specific sequence in this region is important to *bca* transcription, mutations were made in the TTT- N₆-ATAT motif by

introducing nucleotide changes in the forward PCR primer used to generate the GFP reporter constructs. Four mutant constructs were generated in the TTT- N₆-ATAT motif by replacing nucleotides with guanosine residues; a final construct was generated containing an additional six nucleotides in between TTT and ATAT on the plus strand (Figure 1, Panel B.) Site-directed mutant 1 (SDM #1), containing a TTT → GGG change on the plus strand, showed a small decrease in expression compared to the wild-type represented by pN2δ5. Site directed mutant 2 (SDM #2), containing a TTT → GGG change on the minus strand, showed significantly diminished GFP expression. Neither combined mutations on the plus and minus strands (SDM #3 and #4), nor the addition of 6 extra residues (SDM #5), resulted in any observable additional decrease in GFP expression as compared to SDM #2. Given the inability to completely abolish expression by mutating the TTT- N₆-ATAT-N₆-AAA motif, we cannot rule out the possibility that other nucleotides within the motif are also involved in binding of a transcription factor. However, these results clearly indicate that both the specific sequence of the TTT- N₆-ATAT motif, as well as proper spacing within this motif, are important for full expression of *bca*.

2.4 Electrophoretic mobility shift assay (EMSA) assessment of the TTT-N₆-ATAT motif

To determine the potential for the TTT-N₆-ATAT motif to bind a specific transcription factor, EMSA assays were carried out using a biotinylated double-stranded oligonucleotide consisting of the DNA region at position -120 to -100 (CTTTTTTTAACCAAATATGATTCAAAAAAT). Competitor DNA consisted of non-biotinylated oligonucleotide of the same sequence. A GBS total protein extract was prepared and electrophoretic mobility shift assays (EMSA) were carried out (Figure 2A). Probe alone is unshifted (first lane); extract alone has no detectable signal (data not shown). Addition of increasing concentrations of GBS protein extract to a fixed volume of biotinylated oligonucleotide results in an increasing shift to higher molecular weight bands; the incomplete shift observed even at the highest protein concentration may indicate that the specific binding element is present in limiting quantities in this crude GBS extract. The specificity of the oligonucleotide-protein interaction is demonstrated by the significantly diminished binding observed in the presence of unbiotinylated competitor oligonucleotide. To demonstrate the importance of the TTT-N₆-ATAT-N₆-AAA motif in the binding EMSA was performed using the wild type non-biotinylated competitor oligonucleotide containing the TTT-N₆-ATAT-N₆-AAA motif and compared to competition with a mutated oligonucleotide containing the motif GGG-N₆-GGGG-N₆-GGG (Figure 2B). In the absence of competitor the biotinylated probe is almost completely shifted into an oligonucleotide-protein complex. In the presence of wild type competitor less probe is shifted and a significant amount of unbound probe is seen. However, in the presence of mutated competitor oligonucleotide, the probe is still almost completely bound and shifted suggesting that the TTT-N₆-ATAT-N₆-AAA motif is necessary for protein binding. Taken together, these results provide evidence that the TTT-N₆-ATAT motif can interact with a DNA-binding protein in a sequence-specific manner and may indeed function as a transcription-factor binding site.

3. Discussion

A number of mechanisms have been described for the regulation of bacterial surface-expressed virulence factors, including activation and repression via transcriptional regulators, as well as transcriptional and translational regulation by phase variation [17,24]. Detailed structural analyses of GBS promoter regulatory elements, however, are relatively few, including only our studies of the ACP promoter and EMSA-based study of the interaction of RovS with the fibrinogen receptor gene *fbsA* [24]. We previously demonstrated that expression of the alpha C protein in GBS was affected by phase variation-mediated changes in the DNA sequence within a short-sequence repeat region, just upstream from the canonical -10 and -35 promoter elements. In this study, we demonstrate the utility of genomic methods in the *de novo* prediction of gene regulatory elements in GBS. With the identification of a critical transcriptional motif

further upstream of the short-sequence repeats, we propose a working model for regulation of ACP expression in Figure 3. In this model, *bca* expression is regulated both by binding of a transcription factor at the TTT-N₆-ATAT motif; and by phase variation that alters the spacing between the transcription factor binding site and the canonical promoter site, affecting the binding of RNA polymerase. This speculative working model is supported by our work to date, and will provide a framework for determination of the direction of regulation at the TTT-N₆-ATAT site (positive or negative), and in the identification of specific regulatory molecules.

Our studies are limited by the fact they were performed in a reporter system, although we have previously demonstrated the consistency between chromosomal-based alterations in the *bca* upstream region and the same alterations in this reporter system [13]. The *in silico* method utilized in this study is based on the principles of clustering and conservation of motifs within regulatory regions that has primarily been utilized in the prediction of regulatory motifs in eukaryotes; a similar strategy has been used once before in GBS [16]. A more generalized application of this algorithm to accurately predict regulatory motifs *de novo*, which can then be confirmed by traditional promoter mapping techniques, will require further study. The identification of the proteins that bind to the predicted motif, TTT-N₆-ATAT, is currently under investigation.

The ability to regulate expression of ACP through the use of multiple transcriptional mechanisms underscores the importance of ACP in the pathogenesis of GBS infection. ACP is one of the main antigenic epitopes on the surface of GBS as well as one of the key components of GBS interaction with epithelial cells [6,7,25]. There is increasing evidence that epithelial-bacterial pathogen interactions involve a complex cross-talk that involves regulation of bacterial gene expression in response to changes in host cell environment [26]. The present study delineates on a structural level how GBS may regulate the level of expression of a virulence-associated surface molecule in response to host environment. By utilizing transcription factors, the organism can potentially transduce environmental signals to influence gene expression. The phase-variable on/off mechanism can additionally result in a population of bacteria readily available to respond to conditions requiring expression or non-expression of the surface molecule. Further structural study of virulence gene promoters will give insight into the strategies GBS utilize to exist as intestinal commensal organisms as well as identify the mechanisms involved in the transition to an invasive pathogen.

4. Conclusions

We demonstrate the utility of *in silico* analysis in the prediction of potential regulatory DNA protein binding sites in GBS and the confirmation of DNA-protein interactions *in vitro*. The ability to predict regulatory DNA binding sites will contribute to studies of environmental signals that are important to the GBS/host-epithelial interaction and GBS disease pathogenesis.

5. Material and methods

5.1 Bacterial strains and growth conditions

GBS strain A909 is a type Ia/C clinical isolate originally obtained from the Lancefield collection. GBS strain 2603V/R is a type V/rib strain clinical isolate from Italy. All GBS strains were grown in Todd-Hewitt broth (THB) or THY broth (THB supplemented with 5 g yeast extract per liter) or on THY-blood agar plates (Difco Laboratories). Spectinomycin was added to media at 100 µg/ml when appropriate. Cloning work was performed in *E. coli* strain DH5α, grown at 37°C in Luria-Bertani (LB) broth (Difco Laboratories) or on LB agar plates.

5.2 DNA manipulation, PCR amplification and DNA sequencing

Plasmid DNA was prepared using the Qiagen mini-prep kit (Qiagen). Cloning reactions were carried out at 16°C for 4-16 hours, using T4 DNA ligase (Invitrogen) and standard methods. PCR reactions were performed with protoplast DNA preparations (5) using a standard thermocycler and Platinum PCR Supermix (Invitrogen) with cycle: 95°C for 5 min; 35 cycles of 95°C for 30 seconds, 55°C for 30 seconds, 72°C for 30-60 seconds. For DNA sequencing, PCR products were purified with a PCR purification column (Qiagen), mixed with 0.2 µM oligonucleotide primer and sequenced by the Brigham and Women's Hospital Sequencing Facility.

5.3 Construction, transformation and detection of GFP reporter plasmids

Transcriptional reporter constructs were derived from plasmid pN2gfp [13]. This plasmid contains 455 bp of sequence immediately upstream of the start codon for *bca*, fused in frame to the gene for green fluorescent protein (*gfp*). Primers used to generate progressively shorter upstream sequences are given in Table 1; these were used with primer GFPR to generate reporter fusions. Amplified products were cloned into pCR2.1-TOPO cloning vector (Invitrogen) and the integrity of the fusion was verified by sequencing. Verified fusions were digested with EcoRI, gel-purified, and ligated into the gram-positive cloning vector, pDL278 [27]. Ligations were transformed into *E. coli* strain DH5α and spectinomycin-resistant colonies were screened for ligation of the insert DNA.

Electrocompetent GBS strain A909 were prepared as previously described [27]. Competent GBS were transformed by electroporation using a Bio-Rad MicroPulser at 1.5kV. Immediately following electroporation, cold THB broth was added and the bacteria were placed on ice for 20 minutes. The transformed bacteria were incubated for 1 hour at 37°C, and then grown on THY agar plates containing spectinomycin (spec) at 37°C for 16 hours. Spec-resistant colonies were analyzed for the presence of the reporter constructs by PCR: primers GFPF and GFPR were used to screen for *gfp* and pDL278F and pDL278R were used to confirm the presence of pDL278 and the size of the insert.

For detection of GFP, one milliliter overnight bacterial culture was pelleted, resuspended in 100 microliters of glucose-TE (pH 8.0) containing 50 µg/ml mutanolysin and 2.5 mg/ml lysozyme and incubated for 10 minutes at 37°C. The bacteria were then pelleted and resuspended in Laemmli sample buffer (+DTT). Gel electrophoresis, western transfer and antibody detection for GFP, ACP and beta C protein were as previously described [13].

5.4 Identification of putative transcription factor binding motifs

The algorithm developed was modeled on that of Kellis, et. al. [21] for the discovery of conserved motifs and van Nimwegen, et al. [22] for clustering of sequences. The algorithm is available as Appendix A. Short sequence patterns (or mini-motifs) consisting of the type XYZ-m-UVW, where XYZ and UVW are stretches of specific nucleotides exactly three base pairs long and separated by a fixed gap of m nucleotides, where m can range from 0 to 7 nucleotides, were used to search for conserved motifs. The sequence examined included the intergenic region containing 250 nucleotides from the ATG start codon of the divergently oriented *araC* gene, located immediately upstream of *bca*, to the ATG start codon of *bca* (Figure 1). Identified mini-motifs were expanded by one nucleotide at a time in either direction to generate full motifs. Conservation of the full motif was examined by comparing the upstream regulatory sequence of *bca* from six sequenced GBS strains (A909 (Genbank accession number CP000114), 2603V/R (Genbank accession number AE009948), NEM316 (Genbank accession number AL732656), COH1 (Genbank accession number NZ_AAJR000000000), H36B (Genbank accession number NZ_AAJR000000000), CJB111 (Genbank accession number NZ_AAJQ000000000)).

5.5 Electrophoretic Mobility Shift Assay (EMSA)

GBS protein extract was prepared from 500mls of culture grown in Todd Hewitt broth (THB) to OD₆₅₀~0.5. Cells were pelleted and resuspended in 25 mls of glucose-tris-EDTA (GTE) and treated with 1ml of 1mg/ml mutanolysin and 750ul of 100mg/ml lysozyme for 15 minutes at 37° to remove the cell wall. Protoplasts were pelleted and lysed in 50ml lysis buffer (20mM Tris-HCL, 50mM MgCl₂, 1mM DTT, 0.1 mM EDTA, 5% glycerol) containing 25ul protease inhibitor cocktail on ice for 3.5 hrs. 30mls of lysis buffer-1.3M KCl was added and incubation continued on ice for 30 minutes. Lysate was cleared by centrifugation and ammonium sulfate powder was slowly added to a final concentration 50% and incubated at 4° for 1 hour. Precipitated protein was pelleted and resuspended in 2 ml of phosphate buffered saline (PBS), added to a 3500 MW cutoff Slide-a-lyzer cassette (Pierce) and dialyzed overnight against 2L of 10mM HEPES, 1mM MgCl₂, 0.5mM DTT at 4°. Any precipitate from the dialysis was removed by centrifugation and the extract was concentrated by centrifugation using a Amicon Ultra-4 (5k) filter (Millipore) and frozen at -80°.

To remove any background proteins which might interfere with the nucleic acid detection system, 0.5ml of protein extract was incubated for 20 minutes at 4° with 0.5 ml Streptactin-agarose beads (Qiagen) previously washed with 1X binding buffer (10mM HEPES, 60mM KCl, 0.1 mM EDTA, 0.25mM DTT and 0.1 mg/ml BSA). The beads were removed by centrifugation and the supernatant was concentrated to ~100ul final volume by centrifugation using a Amicon Ultra-4 (5k) filter (Millipore).

The EMSA probe and competitor DNAs were generated by annealing single stranded DNA oligos in a thermocycler. 20uM of each oligo was mixed in annealing buffer (10mM Tris pH 8.0, 1mM EDTA, 50mM NaCl) in a volume of 50ul, heated to 95° and cooled by 1° every minute for 70 minutes. The probe DNA contained one oligo which was biotinylated on the 5' end. The competitor DNAs did not contain any biotin. A mutated competitor containing alterations in the proposed binding motif was generated by annealing the non-biotinylated D5D6M_EMSA_F and D5D6M_EMSA_R oligos containing nucleotide substitutions in the key motif residues.

2-fold serial dilutions of protein extract were set up in a final volume of 6ul. To this 2ul of 10X binding buffer, 1ul of 1ug/ul of poly (dI+dC), competitor DNA (where appropriate) and water were added to a final reaction volume of 19ul. For the "probe alone" reaction, water was added in place of protein extract. Reactions were incubated for 20 minutes at 37° to allow binding to the competitor DNA. 1ul of a 1:200 or 1:250 dilution of biotinylated probe DNA was added and incubation continued for 20 minutes at 37°. 5ul of 5X loading dye (1X TBE, 15% glycerol and bromophenol blue) was added and 10ul was loaded onto a 5% pre-cast acrylamide-TBE gel (Bio-Rad) and run for 42 minutes at 120 volts in 0.5X TBE buffer. The gel was transferred to positively charged nylon (Ambion) in 0.5X TBE and 20% methanol at 30 volts for 1 hour. DNA was crosslinked to the membrane using a UV Stratalinker and visualized using the Nucleic Acid Detection System from Pierce.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by Public Health Service grants K08-HD041534 from the National Institute of Child Health and Human Development (to K.M.P.) and R01-AI38424 from the National Institute of Allergy and Infectious Diseases (to L.C.M.) and by a Child Health Research grant from The Charles H. Hood Foundation (to K.M.P.). The authors would like to thank Dr. Stella Kourembanas, Dr. Michael Wessels, and Dr. Dennis Kasper for their ongoing support

of our work; Kelly Shields and Derek Yesucevitz for technical help; and Dr. Miriam Baron and Dr. Gilles Bolduc for helpful discussions.

References

- [1]. Puopolo, KM.; Madoff, LC.; Baker, CJ. Group B streptococcal infection in neonates and young infants. In: Rose, BD., editor. UpToDate. UpToDate; Waltham, MA: 2007.
- [2]. Liu GY, Nizet V. Extracellular virulence factors of group B Streptococci. *Front Biosci* 2004;9:1794–1802. [PubMed: 14977587]
- [3]. Ferrieri P, Baker CJ, Hillier SL, Flores AE. Diversity of surface protein expression in group B streptococcal colonizing & invasive isolates. *Indian J Med Res* 2004;119(Suppl):191–196. [PubMed: 15232193]
- [4]. Persson E, Berg S, Bevanger L, Bergh K, Valsö-Lyng R, Trollfors B. Characterisation of invasive group B streptococci based on investigation of surface proteins and genes encoding surface proteins. *Clin Microbiol Infect* 2008;14(1):66–73. [PubMed: 18034863]
- [5]. Li J, Kasper DL, Ausubel FM, Rosner B, Michel JL. Inactivation of the alpha C protein antigen gene, bca, by a novel shuttle/suicide vector results in attenuation of virulence and immunity in group B Streptococcus. *Proc Natl Acad Sci U S A* 1997;94:13251–13256. [PubMed: 9371832]
- [6]. Baron MJ, Filman DJ, Prophete GA, Hogle JM, Madoff LC. Identification of a glycosaminoglycan-binding region of the alpha C protein that mediates entry of group B streptococci into host cells. *J Biol Chem* 2007;282(14):10526–10536. [PubMed: 17259175]
- [7]. Baron MJ, Bolduc GR, Goldberg MB, Auperin TC, Madoff LC. Alpha C protein of group B Streptococcus binds host cell surface glycosaminoglycan and enters cells by an actin-dependent mechanism. *J Biol Chem* 2004;279:24714–24723. [PubMed: 15044471]
- [8]. Bolduc GR, Baron MJ, Gravecamp C, Lachenauer CS, Madoff LC. The alpha C protein mediates internalization of group B Streptococcus within human cervical epithelial cells. *Cell Microbiol* 2002;4:751–758. [PubMed: 12427097]
- [9]. Bolduc GR, Madoff LC. The group B streptococcal alpha C protein binds alpha1beta1-integrin through a novel KTD motif that promotes internalization of GBS within human epithelial cells. *Microbiology* 2007;153:4039–49. [PubMed: 18048918]
- [10]. Gravekamp C, Rosner B, Madoff LC. Deletion of repeats in the alpha C protein enhances the pathogenicity of group B streptococci in immune mice. *Infect Immun* 1998;66:4347–4354. [PubMed: 9712787]
- [11]. Madoff LC, Michel JL, Gong EW, Kling DE, Kasper DL. Group B streptococci escape host immunity by deletion of tandem repeat elements of the alpha C protein. *Proc Natl Acad Sci U S A* 1996;93:4131–4136. [PubMed: 8633028]
- [12]. Poyart C, Lamy MC, Boumaila C, Fiedler F, Trieu-Cuot P. Regulation of D-Alanyl-Lipoteichoic Acid Biosynthesis in Streptococcus agalactiae Involves a Novel Two-Component Regulatory System. *J. Bacteriol* 2001;183(21):6324–6334. [PubMed: 11591677]
- [13]. Puopolo KM, Madoff LC. Upstream short sequence repeats regulate expression of the alpha C protein of group B Streptococcus. *Mol Microbiol* 2003;50:977–991. [PubMed: 14617155]
- [14]. Gutekunst H, Eikmanns BJ, Reinscheid DJ. Analysis of RogB-controlled virulence mechanisms and gene repression in Streptococcus agalactiae. *Infect Immun* 2003;71:5056–5064. [PubMed: 12933848]
- [15]. Jiang SM, Cieslewicz MJ, Kasper DL, Wessels MR. Regulation of virulence by a two-component system in group B streptococcus. *J Bacteriol* 2005;187:1105–1113. [PubMed: 15659687]
- [16]. Lamy MC, Zouine M, Fert J, Vergassola M, Couve E, Pellegrini E, Glaser P, Kunst F, Msadek T, Trieu-Cuot P, Poyart C. CovS/CovR of group B streptococcus: a two-component global regulatory system involved in virulence. *Mol. Microbiol* 2004;54:1250–1268. [PubMed: 15554966]
- [17]. Spellerberg B, Rozdzinski E, Martin S, Weber-Heynemann J, Lutticken R. *rgf* Encodes a Novel Two-Component Signal Transduction System of *Streptococcus agalactiae*. *Infect. Immun* 2002;70:2434–2440. [PubMed: 11953380]
- [18]. Tettelin H, Massignani V, Cieslewicz MJ, Eisen JA, Peterson S, Wessels MR, Paulsen IT, Nelson KE, Margarit I, Read TD, Madoff LC, Wolf AM, Beanan MJ, Brinkac LM, Daugherty SC, DeBoy

- RT, Durkin AS, Kolonay JF, Madupu R, Lewis MR, Radune D, Fedorova NB, Scanlan D, Khouri H, Mulligan S, Carty HA, Cline RT, Van Aken SE, Gill J, Scarselli M, Mora M, Iacobini ET, Brettoni C, Galli G, Mariani M, Vegni F, Maione D, Rinaudo D, Rappuoli R, Telford JL, Kasper DL, Grandi G, Fraser CM. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. *Proc Natl Acad Sci U S A* 2002;99:12391–6. [PubMed: 12200547]
- [19]. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc Natl Acad Sci U S A* 2005;102:13950–13955. [PubMed: 16172379]
- [20]. Lachenauer CS, Creti R, Michel JL, Madoff LC. Mosaicism in the alpha-like protein genes of group B streptococci. *Proc Natl Acad Sci U S A* 2000;97:9630–9635. [PubMed: 10944228]
- [21]. Kellis M, Patterson N, Birren B, Berger B, Lander ES. Methods in comparative genomics: genome correspondence, gene identification and regulatory motif discovery. *J Comput Biol* 2004;11:319–355. [PubMed: 15285895]
- [22]. van Nimwegen E, Zavolan M, Rajewsky N, Siggia ED. Probabilistic clustering of sequences: Inferring new bacterial regulons by comparative genomics. *Proc Natl Acad Sci USA* 2002;99:7323–7328. [PubMed: 12032281]
- [23]. Puopolo KM, Hollingshead SK, Carey VJ, Madoff LC. Tandem repeat deletion in the alpha C protein of group B *Streptococcus* is recA-independent. *Infect Immun* 2001;69:5037–5045. [PubMed: 11447184]
- [24]. Samen UM, Eikmanns BJ, Reinscheid DJ. The Transcriptional Regulator RovS Controls the Attachment of *Streptococcus agalactiae* to Human Epithelial Cells and the Expression of Virulence Genes. *Infect. Immun* 2006;74:5625–5635. [PubMed: 16988238]
- [25]. Gravekamp C, Kasper DL, Michel JL, Kling DE, Carey V, Madoff LC. Immunogenicity and protective efficacy of the alpha C protein of group B streptococci are inversely related to the number of repeats. *Infect Immun* 1997;65:5216–5221. [PubMed: 9393818]
- [26]. Aldridge PD, Gray MA, Hirst BH, Anjam Khan CM. Who's talking to whom? Epithelial-bacterial pathogen interactions. *Mol. Micro* 2005;55(3):655–663.
- [27]. Dunny GM, Lee LN, Leblanc DJ. Improved electroporation and cloning vector system for gram-positive bacteria. *Appl Environ Microbiol* 1991;57:1194–1201. [PubMed: 1905518]

A.

|N2Δ4
 |
 CATatgagttaccaccTTTgatttaATATgttatctaTTTtagagaATATttctttAAAgtcaacagaataattaggtagtcgata
 araC start -198 -186 -177 -156
 |N2Δ5 |N2Δ5A |N2Δ5B
 |
 |N2Δ5C|N2Δ5D|N2Δ5E |N2Δ6 | -short repeat region- |
 | (4) | (5) |
 ctttTTTaaccaaATATgattcaAAAaatcgaaaaagtcaaattatatatAAAAAAAAAAAAgcAGATTAGATTagataaaaaagca
 -124 -103
 tagataTTCTAAtatttgtgtgttaagccTATAAtttactctGtatagagttatacagagtaaaggagaatattATG → GFP
 -35 -10 +1 bca start

B.

Mutant construct	-124 to -103 Sequence
N2Δ5	TTT - N6 - ATAT - N6 - AAA
N2Δ5D	--- - N6 - ATAT - N6 - AAA
N2Δ5E	-----ATAT - N6 - AAA
N2Δ6	----- AAA
N2Δ5 - SDM #1	GGG - N6 - ATAT - N6 - AAA
N2Δ5 - SDM #2	TTT - N6 - ATAT - N6 - GGG
N2Δ5 - SDM #3	GGG - N6 - ATAT - N6 - GGG
N2Δ5 - SDM #4	GGG - N6 - GGGG - N6 - GGG
N2Δ5 - SDM #5	TTT - N12- ATAT - N6 - AAA

Figure 1.

A. Upstream regulatory region of the *bca* gene. Shown is the 250-bp *araC-bca* intergenic region. The ATG start codon for the *bca* gene, the transcription start site (+1), predicted -10 and -35 motifs and regulatory regions are in bold and underlined, including the putative binding motif TTT-N₆-ATAT, the poly-A stretch and the AGATT pentanucleotide repeats. The 5' extent of sequence fused to *gfp* in each reporter construct is indicated by a labeled line. Each occurrence of the expanded motif TTT- N₆-ATAT is underlined and numbered (1)-(5); occurrences (1), (2) and (4) are on the plus strand; and occurrences (3) and (5) on the minus strand, creating palindrome-like motifs of TTT-N₆-ATAT- N₆-AAA. The sequence used in the EMSA assays is shaded in grey.

B. Mutational analysis of TTT-N₆-ATAT palindrome: Western blot of protein extracts from A909 containing specified reporter constructs. Reporter expression was determined using a monoclonal antibody against GFP. Blots were also probed with a polyclonal antibody against the beta-C protein as a protein loading control. The sequence of each mutant construct in the -156 to -135 region is as shown.

Figure 2.

A. EMSA analysis of TTT-N₆-ATAT palindrome: Electrophoretic mobility shift assay (EMSA) using double-stranded oligonucleotide DNA containing the TTT-N₆-ATAT-N₆-AAA motif and a GBS total protein extract. A fixed concentration of biotinylated oligonucleotide probe and increasing amounts of protein extract, was added to each reaction. The final protein concentration was increased by a factor of two in each lane, ranging from 0.07 mcg/ml to 4.5 mcg/ml. Identical reactions were performed with (right side) or without (left side) non-biotinylated competitor oligonucleotide DNA. The “free probe” reaction contained only biotinylated DNA and no protein or competitor. Binding reactions were electrophoresed under native conditions on an acrylamide gel, transferred to nylon and visualized using the LightShift Chemiluminescent Detection System (Pierce).

B. EMSA with competitor containing a mutated binding motif: Electrophoretic mobility shift assay (EMSA) with a fixed concentration of protein extract (~0.19 mcg/ml) and a non-biotinylated competitor oligonucleotide containing either the wild type TTT-N₆-ATAT-N₆-AAA motif or a mutated version (GGG-N₆-GGGG-N₆-GGG) of the motif.

Figure 3.
Model for transcriptional regulation of the *bca* gene.

TABLE 1

Primers Used in this Study

Name	Sequence (5' - 3')
N2Δ1	AAGTGAAATAGTTTCTCGGATTT
N2Δ2	CAGAAAAATTTTGATCACAAACA
N2Δ3	AGTTAGACCTAGAAGACTGTGTAA
N2Δ4	CATATGAGTTACCACCTTTGATTT
N2Δ5	GTCAACAGAATAATTAGGTAGTCGAT
N2Δ6	AAAAATCGAAAAAGTCAAATTA
N2Δ7	AAAGCAGATTAGATTAGATAAAAA
N2Δ5-A	TAATTAGGTAGTCGATACTTTTTTAA
N2Δ5-B	GTCGATACTTTTTTAACCAAATATG
N2Δ5-C	TTTTTTAACCAAATATGATTCAAAAAATC
N2Δ5-D	AACCAAATATGATTCAAAAAATCGAAA
N2Δ5-E	ATATGATTCAAAAAATCGAAAAAGTCAAAT
Δ5_mutant_1_F	TTAGGTAGTCGATACTTTGGGAACCAAATATGATTCAAAAAAT
Δ5_mutant_2_F	TTAGGTAGTCGATACTTTTTTAACCAAATATGATTCAAGGGATCGAAA
Δ5_mutant_3_F	TTAGGTAGTCGATACTTTGGGAACCAAATATGATTCAAGGGATCGAAA
Δ5_mutant_4_F	TTAGGTAGTCGATACTTTGGGAACCAAGGGGATTCAAGGGATCGAAA
Δ5_mutant_5_F	CGATACTTTTTTAACCAACATATGATATGATTCAAAAAATCGAAAAAGT
GFP-F	ATTATGGCTAGCAGTAAAGGAGAAGAACTTTCACTGG
GFP-R	TTTATCTAGATCATTATTTGTAGAGCTCATCCATGC
pDL278-F	CAAGGCGATTAAAGTTGGGTA
pDL278-R	TGTTGTGTGGAATTGTGAGC
22F	GGTGGACAAGAAAAAGTTCTC
EMSA oligo F	CTTTTTTAACCAAATATGATTCAAAAAAT
EMSA oligo R	ATTTTTTGAATCATATTTGGTTAAAAAG
D5D6M_EMSA_F	CTTTGGGAACCAAGGGGATTCAAGGAAT
D5D6M_EMSA_R	ATTCCCTGAATCCCCCTTGGTTCCCAAAG