The pneumococcal $\sigma^X$, activator, ComW, is a DNA-binding protein critical for natural transformation

Nicole L. Inniss‡, Gerd Prehna§, and Donald A. Morrison¶

From the ‡Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois 60607 and the §Department of Microbiology, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada

Edited by Chris Whitfield

Natural genetic transformation via horizontal gene transfer enables rapid adaptation to dynamic environments and contributes to both antibiotic resistance and vaccine evasion among bacterial populations. In *Streptococcus pneumoniae* (pneumococcus), transformation occurs when cells enter competence, a transient state in which cells express the competence master regulator, SigX ($\sigma^X$), an alternative $\sigma$ factor ($\sigma$), and a competence co-regulator, ComW. Together, ComW and $\sigma^X$ facilitate expression of the genes required for DNA uptake and genetic recombination. SigX activity depends on ComW, as $\Delta$comW cells transcribe late genes and transform at levels 10- and 10,000-fold below that of WT cells, respectively. Previous findings suggest that ComW functions during assembly of the RNA polymerase-$\sigma^X$ holoenzyme to help promote transcription from $\sigma^X$-targeted promoters. However, it remains unknown how ComW facilitates holoenzyme assembly. As ComW seems to be unique to Gram-positive cocci and has no sequence similarity with known transcriptional activators, here we used Rosetta to generate an *ab initio* model of pneumococcal ComW’s 3D-structure. Using this model as a basis for further biochemical, biophysical, and genetic investigations into the molecular features important for its function, we report that ComW is a predicted globular protein and that it interacts with DNA, independently of DNA sequence. We also identified conserved motifs in ComW and show that key residues in these motifs contribute to DNA binding. Lastly, we provide evidence that ComW’s DNA-binding activity is important for transformation in pneumococcus. Our findings begin to fill the gaps in understanding how ComW regulates $\sigma^X$ activity during bacterial natural transformation.

Bacterial natural genetic transformation is the uptake and incorporation of exogenous DNA into a cell’s genome. Transformation was discovered in *Streptococcus pneumoniae* (pneumococcus) (1, 2), and subsequently the genes required for transformation have been found in the genomes of streptococci from all species groups (3). As a result, streptococci have highly malleable genomes, making them well equipped for adaptation, as seen with frequent capsular switching, and the rapid spread of genes that mediate antibiotic resistance (4). Natural transformation has increased the diversity of the pneumococcal genome, directly contributing to the evolution of multi-drug resistant pneumococcal strains (4). As over 82 bacterial species are naturally transformable (5), and many of these are important human pathogens, a deep understanding of this horizontal gene transfer mechanism is important for continued progress in the fight against drug resistant pathogens.

Streptococci primed for transformation are described as competent. Competence is a transient state marked by a shift in both transcrip
tomic and proteomic profiles (6–9). Competence entry is controlled by production of the alternative $\sigma$ factor ($\sigma$)3, $\Sigma$X ($\sigma^X$), a member of the $\sigma^70$ family of proteins (10–12). $\sigma$ factors transiently associate with core RNA polymerase (E) to direct the holoenzyme (E$\sigma$) to specific promoters to initiate transcription (13, 14). Like all bacteria, streptococci contain a principal $\sigma$ factor, SigA ($\sigma^A$), responsible for most gene transcription. Many bacteria have multiple alternative $\sigma$ factors, mediating responses to diverse challenges. However, streptococci have a single alternative $\sigma$ factor, $\sigma^X$. Interestingly, some streptococcal species contain multiple copies of sigX (also known as comX) (3), and sigX expression is strictly linked to competence development. Streptococci utilize two tightly regulated quorum-sensing systems to coordinate $\sigma^X$-mediated competence and other group behaviors (15–18).

Activation of the ComRS or the ComCDE system coordinates uniform sigX expression. Species of the bovis, pyogenec, salivarius, and suis groups utilize the ComRS system to drive sigX production (15, 16). In this system, the pro-peptide ComS is cleaved to its mature form, XIP (SigX Inducing Peptide). XIP (ComS) is actively exported and reimported for binding to ComR, a member of the Rgg, Rap, NprR, PlcR, and PrgX (RRNPP) family of transcriptional regulators (19, 20). The ComR-XIP complex (or ComRS) interacts with DNA as a dimer (21, 22), and binds to a conserved inverted repeat upstream of sigX, termed the ComR-box (7, 23), to activate transcription of the competence regulon. In contrast, species of the anginosus and mitis groups use the ComCDE pathway (17), an auto-regulated Two-component Signal Transduction system (TCST).

3 The abbreviations used are: $\sigma$, $\sigma$ factor; $\sigma^X$, SigX; E, core RNA polymerase enzyme; E$\sigma$, core RNA polymerase-$\sigma$ factor holoenzyme; CSP, competence stimulating peptide; EMSA, electrophoretic mobility shift assay; $\beta$ME, $\beta$-mercaptoethanol.
Pneumococcal ComW binds to DNA

that responds to the Competence Stimulating Peptide (CSP) (24). Interestingly, Streptococcus mutans contains both pathways (15, 25), adding additional layers of complexity to its competence regulation. Therefore, although the competence master regulator, σX, is conserved, streptococci have evolved multiple molecular mechanisms to regulate its expression and consequently, competence.

Much of our understanding of streptococcal transformation comes from initial work done with the ComCDE system in pneumococcus. During exponential growth, the EσX basally transcribes comC leading to production of the pro-peptide, ComC (24). ComC is simultaneously cleaved and exported as CSP by the ABC-transporter, ComAB (comAB) (26–28). CSP is sensed by the histidine-kinase receptor, ComD, resulting in auto-phosphorylation and activation of its cognate response regulator, ComE, via a phospho-relay loop (29–31). Activated ComE promotes transcription of comCDE and other early competence genes (32). ComE-mediated transcription triggers robust competence among the cell population in an auto-positive feedback regulatory loop. This response culminates in ComE dependent production of σX (18, 32). During competence, EσX transcribes from the combox promoter, directly linking ComCDE quorum sensing to the expression of the transformation regulon (18). ComE also activates expression of a competence co-regulator, comW, encoding ComW, a 9.5-kDa protein of unknown function (33).

ComW homologs are only found in eleven species of the anginosus and mitis groups. Like σX, ComW is transient and tightly controlled by ComCDE, suggesting that ComW is unique to ComCDE competence induction. Although σX is the master regulator and sigX is expressed independently of ComW (33, 34), σX activity seems to be dependent on the presence of ComW (35). SigX’s requirement for ComW is supported by observations that mutants lacking ComW (ΔcomW) transcribe late genes and transform at levels 10- and 10,000-fold below that of WT cells, respectively (36). Furthermore, ΔcomW cells have decreased σX levels (35), and ComW and σX weakly interact (37), suggesting that they physically function together in the cell.

As multiple σ factor subunits are present in a cell simultaneously, bacteria have evolved many precise and rapid mechanisms to regulate σ activity. The most common regulatory mechanism of alternative σ activity is sequestration by an anti-σ, and subsequent release of σ upon sensing of the appropriate cellular or environmental cues. Release from sequestration allows alternative σ factors to interact with E resulting in specific Eσ formation (38, 39). Less common methods of σ factor activation include small protein activators, like the Gram-negative specific Crl (40) (an EσX assembly factor) and two-part σ factors, as seen in some bacteriophage systems (41–45) and in Bacillus subtilis (46). In an attempt to pinpoint which of these mechanisms may be applicable to ComW-mediated σX activation, a suppressor screen was conducted in ΔcomW pneumococcal background. Remarkably, only mutations in σX partially restored late gene expression and transformation efficiency (36, 37). These σX suppressor mutations were restricted to σ factor regions 2 and 4 (σ2, σ4) (36, 37), domains involved in E binding and promoter recognition (47). We interpret these results as indicators that ComW functions during the swap from reliance on EσX-mediated transcription to reliance on EσX-mediated transcription during competence and not as a component of a sequestration system.

SigX co-purifies with E from competent pneumococci; but ComW was not isolated with the EσX complex (48). Thus, it is unknown if binding to E is required for ComW function, or whether ComW specifically acts at combox promoters. Purified EσX can direct transcription from combox promoters in vitro without ComW (12). However, these assays were not done in competition with σX, leaving gaps in our understanding of how ComW promotes EσX formation over that of EσX at the onset of competence. As these previous attempts to define how ComW works have not involved direct biochemical or biophysical examination of the protein, our current understanding of ComW’s role during competence is based exclusively on indirect, genetic evidence.

To directly explore the biochemical, biophysical, and genetic properties of ComW to help decipher its function as an EσX assembly factor, we created a structural model of ComW using Rosetta (49). Using this model we identified ComW as structurally homologous to DNA binding proteins and show that ComW directly binds to DNA, independent of DNA sequence. Additionally, we show that specific residues on ComW’s conserved molecular surface are involved in DNA-binding activity and link this activity to efficient transformation in pneumococcus.

Results

ComW is a protein unique to Gram-positive cocci

ComW was initially identified as a CSP responsive competence regulator in multiple pneumococcal strains (18, 33, 50). An alignment of 19 ComW proteins from different pneumococcal strains showed that most alleles share a high percentage of residues that are identical to the R6 sequence. However, pneumococcal ComW alleles can vary by as much as 26% (Fig. 1A). This suggests that ComW is conserved within a species and likely plays an important role for the pneumococcal competence response. Furthermore, comparisons of sequences within and across species suggest that there is intra- and interspecies conservation of ComW.

BLAST searches with the ComW sequence from the R6 pneumococcal strain revealed ComW orthologs conserved in at least ten other streptococcal species. These species belong to the anginosus (S. anginosus, S. cristatus, S. oligofermentans, and S. sinensis) and mitis groups (S. dentisani, S. infantis, S. mitis, S. oralis, S. pseudopneumoniae, and S. tigurinus). Initial examination of such ComW orthologs showed that many retain a high percentage of residues that are identical to the R6 sequence. However, pneumococcal ComW alleles can vary by as much as 41% identity, like ComW from S. dentisani (Fig. 1B). Like pneumococcus, S. dentisani, S. mitis, and S. pseudopneumoniae are members of the mitis group, suggesting that ComW alleles vary across closely related species. Intra-species variation of ComW is common within S. mitis (Fig. S1). S. mitis ComW proteins that are most similar to pneumococcal ComW can
share 88% residue identity. However, some S. mitis ComW proteins share only 40% sequence identity with the ComW of the R6 pneumococcal strain. These relationships suggest that ComW is conserved across many members of the anginosus and mitis groups, but also highlights the divergence of ComW sequences, both at the inter- and intra-species level.

Figure 1. Multi-sequence alignment of ComW. A, alignment of 19 ComW proteins from different pneumococcal strains. A BLAST with the ComW sequence from an R6 strain (NP_357614.1) returned 63 different ComW alleles. Eighteen alleles, spanning the full range of diversity for the alleles identified, were selected at random, for alignment to R6 ComW. (National Center for Biotechnology Information accession number): R6 (NP_357614.1), SMRU2065 (WP_050233459.1), 27 (ETE01301.1), D39 (WP_0543900.1), 40803195 (WP_050253536.1). Note: Few pneumococcal sequences contain an alternative start codon (TTG) six bases upstream of the ATG start (not shown). As most comW sequences in pneumococci and other streptococci lack this TTG start codon, and it is in close proximity to the ribosome-binding site, the alternative start codon is not considered as part of the ComW sequence.

B, alignment of ComW orthologs from S. pneumoniae (Sp, NP_357614.1), S. pseudopneumoniae (Sps, AEL09536.1), S. mitis (Sm, KEQ48646.1), S. anginosus (Sa, EJP26452.1), S. oralis (So, WP_000939510.1), S. dentisani (Sd, WP_038804352.1). For A and B, black text, residues identical to R6; red text, residues that differ from R6; asterisks, identical residues; colons, conserved residues; periods, semi-conserved residues. ComW orthologs from S. anginosus, S. mitis, S. oralis, and S. pseudopneumoniae were moved to pneumococcus, and tested for transformation efficiency. C, Espript3 Alignment of ComW orthologs from streptococci showing helical placement and solvent exposed residues based on a ComW model. Red shading, 100% conserved residues; red letters, similar within a group, residues framed in blue, similar across groups. Relative surface accessibility (acc) of each residue is indicated: dark blue box for accessible, cyan box for intermediate, and white box for buried.

Analysis of ComW sequences revealed four highly conserved regions. In the pneumococcus strain R6, these regions are EEEY, DWE, LIYLVR, and YHRVAYRL. Although ComW sequences exhibit variation, specific aromatic and charged residues within these regions are absolutely conserved in all available orthologs. These four motifs are:

- EEEY (17-20)
- DWE (29-31)
- LIYLVR (38-44)
- YHRVAYRL (56-65)
In addition, many of the residues in these motifs are predicted to be solvent exposed, based on a ComW model (Fig. 1B). ComW sequences also contain shorter variable regions. Of particular interest, the R6 C-terminal residues, 73RGFISC78, are varied or missing from some orthologs (Fig. 1B). These patterns suggest that residues in the conserved motifs are key for ComW function, whereas the C terminus may be dispensable in some species.

A BLAST search also identified an ortholog of ComW in the Gram-positive, anaerobic coccus (GPAC), *Peptoniphilus lacrimalis* (51). Only 47% of *P. lacrimalis*’ ComW residues are identical to those of the pneumococcal R6 strain, and the conserved motifs are retained. The full length ComW sequence was not found in any Gram-negative organisms. Thus, ComW seems to be unique to Gram-positive bacteria. ComW is critical for pneumococcal competence, but its function is unknown, and it lacks close relatives outside of Gram-positive cocci. This rarity impeded production of a homology model, for example by Phyre2D (52), and suggests that ComW may adopt a novel fold that is important to its function.

**A structural model of ComW**

We pursued a structural model to probe ComW’s biochemical and biophysical properties. Attempts to crystallize ComW were unsuccessful, and NMR studies proved cumbersome due to the buffer conditions required for long term stability of the protein above 4 °C (50 mm Tris, pH 8.0, 500 mm NaCl, 10% glycerol, 1 mm EDTA, 1 mm bME) (Fig. S2). As ComW has no close relatives with known structures for the creation of a homology model (Phyre2), we used Robetta (53) and Rosetta (49) to calculate an ab initio ComW model.

Initial models of pneumococcal ComW were built using the Robetta server (53). Robetta produced five models (models not shown). Each model was of a globular protein, composed of 3–4 α-helices connected by loops, with a disordered C-terminal tail (residues 73RGFISC78). Interestingly, Robetta predicted similar models for ComW from *S. pseudopneumoniae* (ComWSps, 86% residue identity to ComW from the pneumococcus R6 strain) and ComW from *S. anginosus* (ComWSa, 43% residue identity to ComW from pneumococcus R6 strain) (models not shown). These initial predictions suggested that, despite primary sequence variability, all ComW proteins share a similar fold.

To directly control model building parameters, as described in the methods, we also used a local installation of Rosetta (49) to model pneumococcal ComW. The Rosetta calculation produced 80,000 ComW models that were similar to those created by the Robetta server. To find a suitable model for ComW, we narrowed the model pool to the 8,000 lowest energy structures.
and clustered the models based on similarity. A summary of five Rosetta low energy clusters is shown in Fig. S3, A–E. Overall the models are globally similar 4-helix bundles with some variation in helical packing. The lowest energy models from each cluster are superimposed in Fig. S3F, and are similar in helical content but differ in their helical packing. This overall structural similarity suggests that the Rosetta calculation produced many models that converge on a similar fold.

Models in Clusters 0 and 1 were nearly identical in energy score (Fig. S3), making it difficult to distinguish which model is most representative of ComW’s native state. We used the DALI server (54) to determine which cluster likely contained models most similar to folds of known function. As detailed later, the lowest energy model from Cluster 1 yielded high quality alignments with proteins or domains of similar size, as determined by DALI Z-scores, indicating it was closer to known protein folds. This low energy structure was chosen as the representative ComW model, and is shown in Fig. 2A.

The Rosetta model of *S. pneumoniae*’s ComW is a globular protein, composed of four tightly packed α-helices that are connected by one short linker and two loops, with a disordered C-terminal tail (residues 73–78), placing the N and C termini at opposite sides of the molecule (Fig. 2A). Surface representations (Fig. 2, B and C) show that the tightly packed helices form a solid core with multiple exterior grooves and/or pockets, in suitable positions to act as binding sites for other biomolecules.

To visualize surface amino acid conservation the Consurf server (55) was used. From the conservation map, the model predicts that one face of ComW has a large area of highly conserved residues (Fig. 2B, right). This face is composed of elements from two conserved motifs; residues L38, Y40, Y41, L42, V43, and R44 (34LIYYLVR44) in helix α3, and Y56, H57, R59, Y62, and R63, (45HYRVAYRLY65) in helix α4 (Fig. 1B and 2B). Residues Y40, Y41, and R44 are conserved across species and are specifically predicted to be solvent exposed. Residues Y56, H57, R59, Y62, and R63 are also conserved across species and are also predicted to be solvent exposed on this face or the bottom of ComW. The rest of the residues from the 38LIYYLVR44 and 45HYRVAYRLY65 motifs are mostly buried in the ComW’s hydrophobic core.

On the opposite face of pneumococcal ComW, there is no contiguous stretch of conserved residues (Fig. 2B, left). However, the strictly conserved E17 residue of the 37EEEY20 motif in helix α1 is predicted to be partially solvent exposed and the center of a deep pocket. Residues E18 and E19 are fully exposed, Y20 is partially solvent exposed, and all three of these amino acids extend to ComW’s bottom.

To calculate the electrostatic potential the Adaptive Poisson Boltzmann Solver (APBS) (56) was used (Fig. 2C). Interestingly, ComW’s opposing faces have opposite electrostatic potential, as calculated by APBS (Fig. 1C). The nonconserved face is highly electronegative, especially the pocket centered on E17 (Fig. 1C, left), whereas the conserved face is electropositive (Fig. 1C, right). Based on these structural features, we hypothesized that these molecular surfaces are binding sites for a protein or other biomolecule and that specific residues in the conserved motifs are integral to such interactions.

**Pneumococcal ComW binds to DNA**

Although we have yet to obtain an X-ray crystal or NMR structure of ComW, a truncated version lacking C-terminal residues 73RGFISC78 (ComWΔ6, 12.4 kDa with V5H6 tag) was purified via affinity chromatography (Fig. S4) for biochemical and biophysical characterization. A circular dichroism (CD) spectrum indicated a predominately α-helical structure in agreement with the Rosetta model (Fig. 3).

Size exclusion chromatography (SEC) experiments suggested that ComWΔ6 oligomerizes in solution, as the pure protein eluted from the gel filtration column at a molecular weight of 25.5 kDa, indicative of a dimer (Fig. 4A). Analytical ultracentrifugation (AUC) sedimentation velocity experiments were performed to verify the oligomeric state of ComWΔ6 (Fig. 4B). The ability of ComWΔ6 to oligomerize at increasing protein concentrations (OD 280 of 0.5 (0.31 mg/ml), 1.0 (0.61 mg/ml), and 1.5 (0.92 mg/ml)) was examined. Analysis of sedimentation data revealed one primary peak for each concentration analyzed. At concentrations of 0.31 mg/ml and 0.61 mg/ml, over 60% of ComWΔ6 sedimented at 2.04S (MW ~16.6 kDa), and 2.12S (MW ~17.0 kDa), respectively. Although some higher molecular weight aggregates were observed, the above sedimentation values are largely indicative of a monomeric state. At a concentration of 0.92 mg/ml, 95% of ComWΔ6 sedimented at 2.69S, with a calculated molecular weight of 25.6 kDa in agreement with SEC experiments. This shows that ComWΔ6 has a propensity to form dimers in solution, and that dimerization is concentration dependent.

**ComWΔ6 oligomerizes in solution**

**ComW model is similar to α factor structures**

The ab initio ComW models produced by Rosetta were used to search the Protein Data Bank (PDB) (57) for similar protein folds, and to help validate our Rosetta calculation. We submitted one model from each Rosetta-generated cluster (Fig. S3), to the Distance Alignment Matrix (DALLI) (58) server. Searches with each model returned 600 hits as structurally similar proteins with known function. The combined search yielded 3,000 total hits. Some structural hits were repeated within and across model searches. To focus on pro-
Pneumococcal ComW binds to DNA

Figure 4. Quaternary structure of ComWΔ6 variants. A, size exclusion chromatography elution profiles of ComWΔ6 variants. The green curve is ComWΔ6, and the orange curves are the L42A and R44A mutants. Top table shows elution volumes and experimental molecular weights of four protein standards. Protein standards were used to calculate the experimental molecular weights of ComWΔ6 variants. Bottom table shows the theoretical and experimental molecular weights, plus the elution fraction (ml) of ComWΔ6 variants. B, sedimentation velocity profiles of ComWΔ6 at varied concentrations. Analytical ultracentrifugation was performed at A5 = 0.5 (gray peak, 0.31 mg/ml, 2.04S, 16.6 kDa), 1.0 (black peak, 0.61 mg/ml, 2.12S, 17.0 kDa), and 1.5 (green peak, 0.92 mg/ml, 2.75, 25.6 kDa). Vbar (0.736), buffer density (d, 1.02 g/L), and viscosity (Poise = 1.07 A10^−5) were determined using SEDNTERP. SEDFIT was used for continuous distribution and Bayesian analysis of the data.

Figure 5. Two structural alignments of the ComW model and selected σ factors. A, Chimera (Petersen, 2004) was used to superimpose the ComW model (coral) on the structurally similar σ factor, σE of E. coli. B, Chimera was used to superimpose the ComW model (coral) on the structurally similar σ factor, σW of B. subtilis. For both panels, the PDB entry and aligned residues, polypeptide chain, Z-score (calculated by the DALI server, and considers matched residues and domain size), root-mean-square-deviation (rmsd), and functional σ region are given in each table, for each structure.

Proteins that may offer clues to ComW’s biological function, DALI hits from all five searches were compared. Proteins that were less than 200 amino acids in length, and/or had DALI Z-scores above 4.0 were considered. Protein structures from the search displayed functions in protein binding, protein degradation, nucleic acid binding, and transcriptional regulation.

Notably, searches with each model returned the Esherichia coli (E. coli) primary σ, σ70 (Z-score = 5.5) (59), as structurally similar. In addition, a number of extra cytoplasmic (ECF) σ factors, including σW of Bacillus subtilis (B. subtilis) (Z-score = 5.7) (60), σE of E. coli (Z-score = 5.8) (61) (Fig. 5), and σK of Mycobacterium tuberculosis (M. tuberculosis) (Z-score = 5.6) (62), were structurally similar to the ComW model. Interestingly, all superimpositions of ComW with these σ factors identified by DALI showed that ComW is structurally homologous to the σ domain. This domain is known to directly interact with RNA polymerase, and function during −10 promoter element recognition and dsDNA melting (63, 64). Thus although ComW has no sequence homologs outside of Gram-positive cocci, it likely adopts a fold similar to some σ-factors that exist in many bacteria. Given that both σK-mediated transcription and pneumococcal transformation are dependent on ComW (37), these data strongly suggest that ComW acts as a DNA binding protein at the onset of competence.

ComWΔ6 binds to DNA, nonspecifically

To test for DNA-binding activity, ComWΔ6 was used in electrophoretic mobility shift assays (EMSA). Increasing amounts of ComWΔ6 were incubated with a fluorescently labeled probe containing the σK competence specific promoter, known as the combox, (65) (Fig. 6A). An increase in the amount of labeled DNA bound by protein was observed as increasing amounts of ComWΔ6 were added to EMSA reactions (Fig. 6C, left). Additionally, a control with cytochrome C showed that the results were likely not artificial protein-DNA interaction simply from the use of concentrated purified protein (Fig. S5). Thus ComWΔ6 can bind to DNA in the absence of any other protein partners.

To determine if the ComWΔ6-DNA interaction was specific for the combox promoter, a labeled probe containing the region upstream of sigX from Streptococcus mutans (Fig. 6B) was used in DNA binding assays. In S. mutans, ComRS, a member of the Rgg-like family of transcriptional regulators, controls expression of sigX from the ComR-box promoter, a 20-bp imperfect, inverted repeat (7, 15, 23). This promoter region differs greatly from the combox promoter that is targeted for transcription by σK (Fig. 6A). As ComW has only been identified in streptococcal species that utilize the ComCDE competence activation pathway, we predicted that ComW would not bind to ComR-box containing DNA. However, an increase in the amount of shifted S. mutans probe was observed when increasing amounts of ComWΔ6 were added to the reaction (Fig. 6C, right). A comparison of binding curves of these interactions (Fig. 6B) sug-
gested that ComWΔ6 binds to each probe with similar kinetics, and therefore binding is not dependent on DNA sequence. ComWΔ6 was also able to shift a plasmid derived DNA probe in EMSA (not shown), further supporting that ComWΔ6 can interact with DNA independently of sequence. In addition, ComWΔ6 interacted with a single stranded DNA (ssDNA) probe only at >64μM of protein (not shown), suggesting that ComWΔ6 may prefer dsDNA targets.

**Mutations to the 38LIYYLVR44 motif of ComW disrupt DNA binding**

Structural alignments of the ComW model with alternative σ factors showed that ComW likely adopts a similar fold to these transcriptional regulators (Fig. 5). Close examination of structural alignments between the ComW model and the crystal structure of σ70 of *E. coli* σ5 (61) shows that many of the residues that constitute ComW’s conserved surface align with residues that are important for σ5-DNA contact during promoter recognition and melting. To investigate if residues in this region of ComW played a role in DNA binding, we purified mutants in the 38LIYYLVR44 motif, and one mutant in the 17EEEY20 motif located on the opposite electronegative face. Only ComWΔ6E18A, ComWΔ6Y40A, ComWΔ6E18A, and ComWΔ6R44A proved soluble and stable 

![Image](image_url)

**Figure 6. Electrophoretic mobility shift of two DNA probes with ComWΔ6.** A, a schematic of DNA probes used in EMSA. Top, a schematic of pneumococcus’ combox promoter region, positioned upstream of the late competence gene, ssbB. Bottom, a schematic of *S. mutans* ComR-box region, upstream of the sigX gene. Pink residues are the conserved imperfect, inverted repeat recognized by the ComRS complex. B, binding curves of DNA probe shifts seen with increasing amounts of ComWΔ6 from C. Curves represent averages from three biological replicates. C, ComWΔ6 titration, up to 32μM.

Each soluble ComWΔ6 variant was tested for DNA binding by EMSA with combox and ComR-box DNA. As expected, the nonconserved E18A mutant was able to interact with both DNA probes, indicating that this residue was dispensable for DNA binding (Fig. 7B, left and middle). Comparisons of DNA binding curves also suggested ComWΔ6E18A mutant binds to different probes with similar kinetics; therefore binding was independent of sequence (Fig. 7B, right). ComWΔ6Y40A also interacted with both DNA probes, indicating that this residue was dispensable for DNA binding (Fig. 7C, left and middle). Again, DNA binding curves showed similar binding kinetics with pneumococcal and *S. mutans* probes (Fig. 7C, right), further supporting that DNA binding is not dependent on DNA sequence. At 4 μM protein, ComWΔ6E18A and ComWΔ6Y40A mutants did not differ in their ability to bind to the DNA probes when compared with ComWΔ6 (Fig. 7A, left). Interestingly, unlike ComWΔ6, ComWΔ6E18A and ComWΔ6Y40A point-mutants showed increased total DNA shift at 32 μM, irrespective of DNA sequence, and appeared to shift the DNA to a higher oligomeric form (Fig. 7, A, right and B and C). Mutation of residues E18 or Y40 to alanine may increase total binding of ComWΔ6 to DNA.

In contrast, ComWΔ6L42A and ComWΔ6R44A showed a decrease in ability to interact with combox or ComR-box DNA, when compared with ComWΔ6. ComWΔ6L42A binds to DNA at ≥4.0μM, compared with binding at 1.0 μM as seen with ComWΔ6 (Fig. 7, A and D, left and middle, and Fig. 6C). However, at 32 μM, ComWΔ6L42A appeared to bind the DNA...
Pneumococcal ComW binds to DNA

A

bound/unbound probe

0.3

0.2

0.1

0

ComW6 E18A

Y40A

L42A

R44A

20 nM probe + 4µM protein

bound/unbound probe

0.5

0.3

0.1

0

ComW6 E18A

Y40A

L42A

R44A

20 nM probe + 32µM protein

---

B

ComW6-E18A

+20nM combox

µM protein

0 0.5 1.0 2.0 4.0 8.0 16 32

+20nM ComR-box

µM protein

0 0.5 1.0 2.0 4.0 8.0 16 32

fraction of probe bound

0 0 0.2 0.4 0.6 0.8 1.0

0 0 0.2 0.4 0.6 0.8 1.0

---

C

ComW6-Y40A

+20nM combox

µM protein

0 0.5 1.0 2.0 4.0 8.0 16 32

+20nM ComR-box

µM protein

0 0.5 1.0 2.0 4.0 8.0 16 32

fraction of probe bound

0 0 0.2 0.4 0.6 0.8 1.0

0 0 0.2 0.4 0.6 0.8 1.0

---

D

ComW6-L42A

+20nM combox

µM protein

0 0.5 1.0 2.0 4.0 8.0 16 32

+20nM ComR-box

µM protein

0 0.5 1.0 2.0 4.0 8.0 16 32

fraction of probe bound

0 0 0.2 0.4 0.6 0.8 1.0

0 0 0.2 0.4 0.6 0.8 1.0

---

E

ComW6-R44A

+20nM combox

µM protein

0 0.5 1.0 2.0 4.0 8.0 16 32

+20nM ComR-box

µM protein

0 0.5 1.0 2.0 4.0 8.0 16 32

fraction of probe bound

0 0 0.2 0.4 0.6 0.8 1.0

0 0 0.2 0.4 0.6 0.8 1.0

---

** * 

Pneumococcal ComW binds to DNA

11108 J. Biol. Chem. (2019) 294(29) 11101–11118
probes at levels that were not significantly different than ComWΔ6 (Fig. 7A, right). In addition, a comparison of DNA binding curves with different probes, suggested that ComWΔ64-42A retains the ability to bind to DNA independently of sequence (Fig. 7D, right).

ComWΔ6444A showed the most dramatic decrease in DNA binding. At 4 μM, ComWΔ6444A did not bind to either DNA probe (Fig. 7E, left and middle, Fig. 7A, left). More than 8 μM of protein was required to shift small amounts of the probe in EMSA gels (Fig. 7E, left and middle). At 32 μM of protein, ComWΔ6444A shifted only 11–12% of either DNA probe (Fig. 7A, right), significantly less DNA than that shifted by ComWΔ6. This dramatic decrease in binding activity is evident in DNA binding curves. However, the DNA binding curves did show that there was no significant difference in binding to pneumococcal and S. mutans probes (Fig. 7E, right) by ComWΔ6444A. These results demonstrate that residues ComWΔ642 and ComWΔ644 participate in the ComWΔ6-DNA interaction. Thus, the 38LIYYLVR44 motif, present on the positively charged conserved face of ComW, is important for efficient DNA binding.

**Residues that participate in ComWΔ6 DNA binding are not required for oligomerization**

As ComWΔ6 oligomerizes (Fig. 4), which may contribute to DNA binding, we used SEC to determine the integrity of quaternary structure in the ComWΔ642A and ComWΔ644A mutants. Like ComWΔ6 (25.5 kDa), the ComWΔ642A and ComWΔ644A mutants eluted as dimers, with molecular weights of 19.4 and 24.4 kDa, respectively (Fig. 4). This demonstrated that mutations in the 38LIYYLVR44 motif of pneumococcal ComW did not disrupt oligomerization. Together, the SEC and EMSA results show that ComW’s 38LIYYLVR44 motif on the conserved, electropositive face is unlikely to participate in protein oligomerization, but instead is part of a direct binding site for DNA.

**ComW mutations alter protein levels and pneumococcal transformation efficiencies**

To determine levels of ComW production and transformation efficiency in comW mutant strains, we used Western blots and a pneumococcal natural transformation assay. A pneumococcal-specific antibody detected WT ComW as a 9.5-kDa band on blots, and WT cells transformed with an efficiency of 66%. In contrast, ΔcomW mutants produced no ComW and transformed at levels 10,000-fold below that of WT cells (Fig. 8, A and C, and Fig. S6B). We attribute the loss of transformability to complete loss of ComW protein in this strain. Following biochemical characterization of the ComWΔ6 variant, we also determined how removal of the disordered C-terminal domain altered ComW production and function in pneumococcus. ComWΔ6 was produced at levels that were 1/3 that of WT cells and comWΔ6 mutants transformed at 31% efficiency (Fig. 8, A and C, and Fig. S6B). Thus comWΔ6-expressing pneumococci exhibited a decrease in transformability that appears to be dependent on the level of ComWΔ6 production. Thus, the C-terminal residues, 78RGFISC78, which are not conserved across streptococci (Fig. 1D), are important for ComW stability and full function in vitro, but are dispensable for DNA binding in vivo. The specific functional importance of the disordered tail has yet to be determined.

We determined if changes to the 17EEE19 motif altered ComW levels and transformability of pneumococci. This motif is present on the electro-negative, nonconserved surface of ComW and does not appear to be required for DNA binding (Fig. 7, A and B). At 17 min post CSP induction, mutants ComWE17A, ComWE18A, or ComWE19A were produced at significantly lower levels than ComW (Fig. 8C, Fig. S6B). Interestingly, the comWE17A and comWE19A-expressing mutants transformed at only 24 and 33%, respectively, but the comWE18A expressing mutant transformed at 73% efficiency (Fig. 8A). Pneumococci did not produce detectable levels of ComWΔ17AAA19, and comWΔ17AAA19-expressing mutants transformed at levels similar to that of ΔcomW cells (Fig. 8, A and C, and Fig. S6B). Although decreases in ComWΔ17A, ComWΔ19A, and ComWΔ17AAA19 result in decreases in transformation efficiency, decreases in ComWΔ18A does not, and suggests that decreased protein production does not always result in a decrease in transformation efficiency. In addition, as these ComW variants are less stable than WT near peak competence, but mutation to this motif allows DNA binding, we hypothesize that the electro-negative surface of ComW serves an important functional role that may be separate from ComW DNA binding activity.

We also examined how changes to the 38LIYYLVR44 motif altered ComW levels and pneumococcal transformability. This motif is present on ComW’s electro-positive, conserved surface (Fig. 2) and is required for DNA binding (Fig. 7). ComWY40A, ComWY41A, ComWΔ42A, and ComWΔ44A were all detectable in pneumococci, but ComWY41A was the only mutant produced at significantly lower levels than ComW, and ComWΔ40AA41 was not detected (Figs. 8C, Fig. S6B). Interestingly, comWY40A, comWY41A, comWΔ42A, comWΔ44A, and comWΔ40AA44 expressing cells all showed significant decreases in transformation efficiency at 52%, 10%, 16%, 27%, and 0% efficiencies, respectively (Fig. 8A). As residues L42 and R44 are each important for DNA binding activity (Fig. 7), and point mutation of either residue disrupts transformation without depletion of the ComW variant protein, we interpret this as indication that DNA binding activity, facilitated by the electro-positive surface of ComW, plays a key role during competence development.

---

**Figure 7. Electrophoretic mobility shifts of two DNA probes with ComWΔ6 variants.** A comparison of DNA probe shifted by ComWΔ6 variants at 4 μM (left) and 32 μM (right). For all samples, n = 3 biological replicates. C. pneumo combox probe; R. S. mutans ComR-box probe. Asterisks (*) mark statistically significant differences between ComWΔ6 and mutants bound to probe (*, p < 0.05; **, p < 0.01). B–E. Representative gels of mutant ComWΔ6 titrations, up to 32 μM, with 20 nM labeled probe, combox (left) or ComR-box (middle), and binding curves (right). Both probes are 589 bp in length. The average curves from three biological replicates are shown.
Pneumococcal ComW binds to DNA

Figure 8. Function and expression of ComW variants in pneumococcus. A, transformation efficiency in pneumococcal cells expressing ComW mutant proteins. Residues mutated in the 17EEY20 motifs are blue, in the 38LIYYLVR44 motif are orange, and the C-terminal truncation mutation (Δ29RGFISC59) is green. For WT and ΔComW strains, n = 9; for ComW variant strains, n = 3. B, transformation efficiency in pneumococcal ComW chimer strains. Replacement of pneumococcal ComW with that of S. pseudopneumoniae (Sp), S. mitis (Sm), S. oralis (So), or S. anginosus (Sa). WT and ΔComW strains, n = 9 biological replicates; for ComW ortholog strains, n = 3 biological replicates. For both panels, after CSP induction of transformation and overnight growth on THY agar plates, supplemented with 10 μg/ml novobiocin, cells were counted and the #transformants/CFU determined. Error bars are standard deviation (S.D.); p ≤ 0.05, an asterisk (*) marks statistically significant differences in transformation efficiency compared with WT cells. C, representative images Western blotting detection of ComW variants in pneumococcal cells. 1 × 10^8 cells were added to each well. The RNA polymerase β subunit (∼137 kDa) was used as an endogenous loading control and to normalize ComW signal from each sample.

The ComW ortholog from S. anginosus does not complement pneumococcal ComW

Although all streptococci share competence specific genes, including the master regulator, ωC, they differ in competence activation pathways (3). ComW is only produced in streptococci that activate competence via the ComCDE quorum-sensing pathway (10, 33, 66). Therefore it is possible that ComW is unique to only the ComCDE system, and that streptococci have evolved multiple mechanisms to regulate competence via direct regulation of ωC activity. We explored how natural variation in ComW orthologues affected their ability to function in competent pneumococci. Pneumococcal ComW was replaced with that of S. anginosus (ComWsa), S. mitis (ComWsm), S. oralis (ComWso), or S. pseudopneumoniae (ComWsp), and we examined the production of ComW orthologues and the corresponding transformation efficiency in the resulting chimeric strains.

ComWsa and ComWsm are 73 and 78% identical to pneumococcal ComW respectively (with identical 17EEY20, 38LIYYLVR44, and 56YHYRAAYRLY65 motifs, and have varied sequences of C-terminal residues, Fig. 1B). Both ComWsm and ComWsp were detected in Western blots from competent chimeras at 17 min post CSP induction, but levels of ComWsp were significantly decreased compared with pneumococcal ComW, whereas ComWsm levels were not (Fig. 8C and Fig. S6B). The S. mitis and S. pseudopneumoniae chimeras transformed at 56% and 55%, respectively, and only the decrease in transformation with ComWsp was statistically significant (Fig. 8B). Although it is possible that differences in amino acid sequence contribute to the observed differences in protein levels, these data do demonstrate that the S. mitis and S. pseudopneumoniae orthologues are produced in pneumoccus and are functional in pneumoccus. Thus ComW functionality is retained across closely related species.

ComWsa shares only 43% identity with ComW (with 17EEY20, 38LIYYLIR44, and 56YHYRAAYR67Y65 motifs, and lacks the disordered C-terminal domain, Fig. 1D). Similarly, ComWso shares 40% identity with pneumococcal ComW (including 17EQEY20, 38LIYYLIR44, and 56YHFRAAYR65 motifs, and contains a disordered C-terminal domain, Fig. 1D).
ComW<sup>sa</sup> protein was not readily detected in Western blots and, in stark contrast to other orthologues, chimeras expressing comW<sup>sa</sup> transformed at levels 10,000-fold below WT cells, a phenotype like ΔcomW cells (Fig. 8, B and C). This suggests that ComW<sup>sa</sup> is not stably produced in pneumococcal cells.

In contrast ComW<sup>so</sup> was detectable, albeit at significantly decreased levels compared with WT, yet chimeras expressing comW<sup>so</sup> transformed at 41% efficiency (Fig. 8, B and C). This result is consistent with the fact that observed decreases in some ComW variants do not always predict equal decreases in transformability. It is possible that the specificity of the pneumococcal ComW antibody hinders detection of the more divergent ComW<sup>sa</sup> and ComW<sup>so</sup> alleles. However, combined with their decreased transformation efficiencies, these data suggest that ComW<sup>sa</sup> and ComW<sup>so</sup> are not stable and/or functionally in the pneumococcal competence system. Therefore, although ComW proteins are similar across species, the transformation and Western blotting data suggest that there are species-specific determinants that are important for ComW stability and function.

**Discussion**

Since ComW’s identification, there have been limited advances in our understanding of its function. Although a genetic link to the shift from σ<sup>+</sup> to σ<sup>+</sup> σ-dependent transcription has been established (36, 37), comW had not been probed for mutations that disrupt pneumococcal transformation. Furthermore, direct biochemical characterization of ComW was stalled by a lack of soluble protein. Bioinformatic tools, like Robetta, Rosetta, and the DALI server (49, 53, 58), have laid a foundation for deeper analysis of ComW function.

The 
<initio> ab model of ComW described here depicts a globular protein, α-helical in structure, with a DNA-binding-like fold. Combined predictive analysis of ComW’s surface characteristics using the APBS (56) and Consurf (55) servers identified two opposing faces of ComW: one conserved, electro-positively charged face, and one nonconserved, electro-negatively charged face. In vitro characterization of ComW<sub>Δ6</sub> (12.4 kDa, Δ<sup>73RGFISC78</sup>), a soluble variant with an α-helical structure as predicted by the model (Fig. 3), shows that the protein dimerizes in solution (Fig. 4).

Importantly, we report for the first time that ComW<sub>Δ6</sub> binds to DNA independently of DNA sequence (Fig. 6), an interaction that depends on residues that constitute the conserved, electro-positively charged face (Fig. 7). Additionally, conserved residues on both faces of ComW are important for pneumococcal transformation (Fig. 8).

Interestingly, point mutations to ComW’s nonconserved, electro-negative face appear to destabilize the protein more than point mutations to ComW’s conserved, electro-positive face. The presence of stable protein for poor DNA binding mutants ComW<sup>L42A</sup> and ComW<sup>R44A</sup> suggest that ComW-DNA interactions are important for transformation. Thus, we have begun to fill the gaps in understanding ComW’s function as a regulator of σ<sup>+</sup> activity during natural transformation.

Pneumococcal ComW and σ<sup>+</sup> factors are protein partners that promote transformation-specific gene transcription (33, 35–37). SigX is a member of the σ<sup>+</sup> family of proteins (10), most similar to Group 4 σ factors, the Extra-cytoplasmic Function (ECF) σ factor family. ECF σ factors are small, as they are composed of only the σ<sub>11</sub> and σ<sub>16</sub> domains essential for E binding, promoter recognition, and promoter melting (39). Our ComW model resembles the structure of the σ<sub>11</sub> domain of E. coli’s primary σ factor, σ<sup>+</sup>(59), and the structures of some alternative σ factors, like B. subtilis’ σ<sup>W</sup> (67), and E. coli’s σ<sup>+</sup> (61, 68). In addition, ComW<sub>Δ6</sub> interacts with DNA in EMSAs. This is an activity that has been observed with both Gram-negative and Gram-positive σ factors, including the structurally homologous σ<sub>II</sub> of M. tuberculosis (69), PGO162 of Porphyromonas gingivalis (70), σ<sub>II</sub> of Staphylococcus aureus (71), and σ<sup>+</sup> of Vibrio alginolyticus (72). In concert with the requirement of ComW for robust σ<sup>+</sup>-dependent transcription, we predict that ComW is an active member of the Eσ<sup>+</sup> complex during pneumococcal late-competence gene transcription.

Amino acids that are important for σ-factor-DNA contact have been identified in structures of multiple σ factors or σ factor domains in complex with DNA. Conserved aromatic and basic residues in the σ<sub>11</sub> domain act in promoter melting and nonsequence-specific DNA binding (63, 73–78). The ComW model has one conserved surface that is composed of motifs 38LIYYLVR<sup>44</sup> and 56YHYRVAYRLY<sup>65</sup>. Many of these aromatic and charged residues align with amino acids that are important for E. coli’s σ<sup>+</sup>-DNA contact during promoter recognition and melting (61). Furthermore, we determined that mutation to two pneumococcal ComW residues, L42 and R44 on the conserved face, are important for nonspecific DNA binding, and that ComW<sup>L42A</sup> and ComW<sup>R44A</sup> are stably produced in pneumococcus but disrupt transformation efficiency. These results are in agreement with current understanding of the types of amino acids that σ factors utilize for DNA interaction, and support a direct role for ComW in transcription activation at σ<sup>+</sup> promoters. We expect that additional residues within these motifs aid in promoter recognition and DNA binding.

Although E alone can bind to DNA and promote transcription (64, 79, 80), σ factors interact with E to specify transcriptional targets (13). SigX does not require ComW for interaction with E (unpublished data), and Eσ<sup>+</sup> can transcribe from combinatorial promoters in vitro, in the absence of ComW (12). However these assays did not determine Eσ<sup>+</sup> promoter specificity in the presence of σ<sup>+</sup> or σ<sup>+</sup> dependent promoters. Hence it will be important to determine if ComW alters the affinity of σ<sup>+</sup> for E, and if/how ComW affects Eσ<sup>+</sup> transcription in these more competitive contexts. ComW’s resemblance to σ<sub>11</sub> domains is peculiar because this domain also functions in DNA sequence-specific promoter recognition (63, 64). However, ComW interacts with DNA nonspecifically. Thus, it is possible that Eσ<sup>+</sup> promoter specificity is determined by σ<sup>+</sup> and that ComW mediates Eσ<sup>+</sup>-DNA interactions via binding to the DNA phosphate-backbone, or in another manner that does not depend on DNA sequence.

Sigma factor activity is often controlled via direct interaction by a regulatory protein. Canonically, anti-σ factors, many of which are membrane proteins, sequester ECF σ factors, inhibiting Eσ formation. Specific cues activate regulated intramembrane proteolysis (RIP) cascades, leading to the release of the ECF (81–84). An anti-σ<sup>+</sup> protein has not been identified (37),
making it unlikely that ComW-mediated $\sigma^X$ regulation occurs via alleviation of a sequestration mechanism.

Noncanonical $\sigma$ factor control via interaction with a small protein has been observed in both Gram-negative and Gram-positive organisms. In Gram-negative organisms, like E. coli, Crl (\sim 15.6 kDa) binds to the stress response specific $\sigma^D$ and to E, to promote formation of the E$\sigma^D$ complex (40, 85). In the absence of Crl, E$\sigma^D$ fails to form stable complexes, and transcription from $\sigma^D$ promoters is decreased (85). As we have observed E$\sigma^X$ formation in competent pneumococcal $\Delta$comW cells (48), the structure of Crl (compose of an $\alpha+\beta$ fold, (86)) vastly differs from the ComW model, and no Crl-DNA interactions have been reported, it is unlikely that ComW shares an equivalent $\sigma$ factor activating mechanism with Crl.

In the Gram-positive bacterium, B. subtilis, YvrI-YvrHa are two small proteins that promote transcription of genes required in acidic conditions (87). Activation of gene transcription by the $\sigma$ factor-like protein, YvrI, requires binding of its N terminus to the small co-regulator, YvrHa. In this system, YvrI is most similar to $\sigma_4$ domain, interacts with the $\beta$-flap domain of E, an interaction that is not dependent on YvrHa, and determines the promoter specificity of E$\sigma^{Xvr}$. YvrHa interacts with the $\beta$' subunit of RNAP, a conserved interaction for the $\sigma_2$ domain, and aids in open complex stabilization. As these independent proteins co-purify with RNAP from B. subtilis, YvrI and YvrHa likely function in vivo as a unified $\sigma$ factor (46).

ComW and $\sigma^X$ weakly interact in yeast-two-hybrid assays (37), but a direct interaction between these protein partners has not been demonstrated in competent pneumococcal cells or in vitro. Yet $\sigma^X$ is destabilized in $\Delta$comW cells, or cells with N- or C-terminally tagged ComW (35), suggesting that the $\sigma^X$-ComW interaction is important for stability of both proteins. ComW$\Delta$6 levels are decreased compared with WT, resulting in decreased transformation. In addition, point mutations to ComW’s electro-negative surface results in decreased transformation and decreased protein levels near peak competence. Thus it is possible that the C-terminal domain and/or the electro-negative surface of ComW mediate interaction with $\sigma^X$. Interestingly, point mutations to ComW’s DNA binding surface, specifically residues L42 and R44, disrupt transformation but produce stable protein. So it is likely that interaction with $\sigma^X$ is not mediated by this surface. Thus, it will be valuable to determine $\sigma^X$ protein levels in all ComW mutants in order to identify residues important for $\sigma^X$ stabilization.

The $\sigma^X$ and ComW pair shares some similarity to B. subtilis’ YvrI-YvrHa. At the protein level, $\sigma^X$ and YvrI are \sim 23 kDa, and ComW and YvrHa are \sim 10 kDa. SigX certainly functions as a $\sigma$ factor (12), and ComW’s predicted similarity to $\sigma_2$ domains suggests that it too can interact with E. Investigation into endogenous ComW protein-protein interactions are required to determine how it functions in the context of the E$\sigma^X$ complex at competence onset.

Orthologs of comW have been identified in anginosus and mitis group streptococci, and in the Gram-positive coccus, P. lacrinalis. All known ComWs are similar in primary structure and Robetta models of orthologs from S. anginosus and S. pseudopneumoniae are similar to that of pneumococcus. However, our findings that these orthologs differ in their ability to complement pneumococcal ComW during transformation and the observed differences in protein stability hint that species-specific determinants exist that dictate ComW function in vivo. As subunits of E are highly conserved among Gram-positive and Gram-negative organisms (47, 64), it is more likely that molecular interactions with species specific $\sigma^X$ are more important for ComW function. Like ComW, the $\sigma^X$ primary sequence is highly similar across species, but does vary. Furthermore, streptococci of different species groups utilize different quorum sensing systems to activate production of $\sigma^X$ and ComW seems to be unique to only one of these systems (ComCDE). If streptococci have evolved multiple mechanisms to regulate competence, the $\sigma^X$-ComW system provides a unique opportunity for deep study into the co-evolution of transcriptional regulators.

Lastly, although specific affinities of pneumococcal $\sigma^X$ and $\sigma^X$ for E have not yet been determined, $\sigma^X$ mutations that are predicted to disrupt E$\sigma^X$ formation (mutations in $\sigma_2$ and $\sigma_4$ domains), bypass the ComW requirement (36, 37). Thus, competition for E between $\sigma^X$ and $\sigma^X$ at the onset of competence is likely a key determinant of transformation efficiency. It is conceivable ComW helps to tip the competition in favor of E$\sigma^X$ formation over that of E$\sigma^X$. Previous work demonstrated that replacement of pneumococcal $\sigma^X$ with a chimeric $\sigma^X$, which included S. mutans’ regions $\sigma_2$ and $\sigma_4$, partially restored $\Delta$comW phenotypes (unpublished, Tovpeko Thesis, 2016). Recall that S. mutans are naturally transformable streptococci that do not produce ComW during competence. This is additional evidence that some species-specific determinants exist and are important factors in streptococcal transformation. More broadly, species of the bovis, pyogenic, salivarius, and suis streptococcal groups likely evolved ComW-independent mechanisms to regulate $\sigma^X$ activity, and consequently, natural transformation.

The pneumococcal competence specific ComW and $\sigma^X$ regulators provide a unique framework to study multiple molecular phenomena. ComW, predicted as structurally similar to $\sigma$-factors, and $\sigma^X$, the only known $\sigma$-factor in streptococci, work together to promote a robust change in pneumococcal transcription during competence. They have been shown to directly interact with DNA and/or RNA polymerase, and both are required for efficient pneumococcal transformation. Taken together, these observations suggest that ComW and $\sigma^X$ function together, perhaps, for example, as a competence specific two-part $\sigma$ factor. Therefore ComW and $\sigma^X$ provide an opportunity to study $\sigma$ factor function, regulation, and holoenzyme assembly. ComW and $\sigma^X$ also offer an unambiguous context to study how independent proteins function together to promote gene expression. Lastly, further investigation into the ComW-$\sigma^X$ system will add insight into a mechanism used to control natural genetic transformation in an important human pathogen. This knowledge will add to the growing number of mechanisms that bacteria use to promote horizontal gene transfer. A deeper understanding of the mechanisms bacteria use to scavenge evolutionarily favorable genes will increase ability to combat their rapid adaptation to antibiotics and bacterial specific vaccines.
Materials and methods

Bacterial strains and culture media

The bacterial strains used in this study are listed in Table S1. CP2137, a Δcps ΔcomA AssbB::pEVP3::ssbB derivative of strain Rx1 (10, 36, 88) was used as the wildtype (WT) standard for competence transformation assays. CP2137 does not secrete endogenous competence stimulating peptide (Csp) due to deletion of the CSP exporter, ComA. All comW mutations were placed in the CP2137 background (see below). CP2463, a ΔcomW::kan derivative of CP2137, was used as the ΔcomW standard for transformation assays (36). All pneumococcal strains were cultured in CAT medium and plated on CAT or THY supplemented with 1.5% agar and selective antibiotic, as needed. CAT medium was prepared from 5 g of tryptone (Difco Laboratories), 10 g of yeast extract (Difco), and 5 g of NaCl (Fisher Scientific) in 1 liter of H2O, sterilized for 40 min at 121 °C, and then supplemented to 0.2% glucose and 0.016 M KH2PO4 before use. THY was prepared by sterilizing for 40 min at 121 °C, and then supplemented to 0.2% glucose and 0.016 M KH2PO4 before use. THY was prepared by sterilizing for 40 min at 121 °C, and then supplemented to 0.2% glucose and 0.016 M KH2PO4 before use. THY was prepared by sterilizing for 40 min at 121 °C, and then supplemented to 0.2% glucose and 0.016 M KH2PO4 before use.

E. coli strains DH5α and BL21De3 were hosts for plasmid isolation and protein expression, respectively. For plasmid introduction, E. Coli strains were chemically transformed according to (89). E. coli strains were cultured in lysogeny broth (LB) (90). LB was prepared from 5 g bacto tryptone (Difco), 5 g NaCl (Fisher Scientific), 2.5 g of yeast extract (Difco) in 1 liter of H2O, sterilized for 40 min at 121 °C, and then supplemented to 0.2% glucose and 0.016 M KH2PO4 before use. THY was prepared from 10 g of yeast extract (Difco), 30 g of Todd Hewitt Broth (Difco) in 1 liter of H2O and sterilized for 20 min at 121 °C. Novobiocin was used in pneumococcal transformation assays at 2.5 μg/ml.

E. coli strains DH5α and BL21De3 were hosts for plasmid isolation and protein expression, respectively. For plasmid introduction, E. Coli strains were chemically transformed according to (89). E. coli strains were cultured in lysogeny broth (LB) (90). LB was prepared from 5 g bacto tryptone (Difco), 5 g NaCl (Fisher Scientific), 2.5 g of yeast extract (Difco) in 1 liter of H2O, sterilized for 40 min at 121 °C, and then supplemented to 0.2% glucose and 0.016 M KH2PO4 before use. THY was prepared from 10 g of yeast extract (Difco), 30 g of Todd Hewitt Broth (Difco) in 1 liter of H2O and sterilized for 20 min at 121 °C. Novobiocin was used in pneumococcal transformation assays at 2.5 μg/ml.

Computational modeling

The ab initio structural models of pneumococcal ComW were calculated using the Robetta server (53), and with a local installation of Rosetta (49). The Robetta server generated 3-residue and 9-residue fragment files that were used as input for the Rosetta calculation. For the Rosetta calculation, radius of gyration, contact-order, and sheet filters were used. Helix and loop structures were weighted equally and a fast relax protocol was performed.

Clustering of the 8,000 most energetically favorable models was done using Rosetta’s cluster application with automatic radius detection. The Global Distance Test (Rosetta specific gdtmmm) was used to cluster the models. Five clusters, each with nine models, were generated. One model from each of the five clusters was selected as representative of pneumococcal ComW’s structure, based on Rosetta energy score, and cluster agreement as determined by rmsd between superimposed structures.

Plasmids and primers used in this study are listed in Tables S2 and S3, respectively. The comW truncation and orthologous variants in pET22b+ vectors were synthesized by Genscript (New Jersey). Gene sequences for comW were taken from the genomes of streptococcal species (NCBI): S. anginosus (ALJO01000004.1), S. oralis (NZ_NCUJ01000012.1), S. mitis (NZ_RMNV01000010.1), S. pneumoniae R6 (NC_003098.1) and S. pseudopneumoniae (CP002925.1). Oligonucleotides were synthesized by IDT (Coralville, Iowa).

Site directed mutagenesis of pNL137 was used to generate point mutations in pneumococcal comW. pNL137 was used as a template for PCR amplification by Platinum Pfx Polymerase (Thermo Fisher Scientific). Mutation specific primers (Table S3) and PCR conditions optimized for each primer set (individual conditions not given) in 50 μl reactions. PCR products were digested with DpnI (New England Biolabs) for 2 h at 37 °C. To circularize plasmids, a modified Seamless Ligation Cloning Extract (SLICE) protocol (91) was used. Briefly, 8 μl of DpnI digested PCR product was mixed with 1.5 μl of SLICE reaction buffer and 1.5 μl of SLICE cell extract, and brought to 15 μl volume with H2O. Reaction mixtures were incubated at 37 °C for 15 min in a thermocycler. 3 μl SLICE reaction products were transformed into 50 μl of DH5α cells (Invitrogen), and then plated on LB agar with 100 μg/ml ampicillin. Single colonies were picked and cultured in LB medium overnight at 37 °C. Plasmids were isolated using the ZymoPURE Plasmid Miniprep Kit (Zymo Research) and concentration was measured using a ND-1000 Spectrophotometer (Nanodrop). Mutations were confirmed by Sanger sequencing (University of Illinois at Chicago Sequencing Core (UICSEQC)).

Gibson assembly (92) of pET22b+ containing comW point mutants was achieved by use of the Hifi Assembly Mix (New England Biolabs). Briefly, comW point mutants from plasmids pNL180, pNL187, pNL188, pNL189, pNL190, pNL191, and pNL100 were PCR amplified using primer pair NL226 and NL233 which deleted residues 73RGFISCF78 of comW. The PCR products contained 5’ and 3’, 12 bases overlap with NLI60 vector backbone that contained a C-terminal V5 epitope-6His tag (V5H6), followed by a stop codon. The vector backbone was PCR amplified using primer pair NL190 and NL232. PCR products were purified using the DNA Clean & Concentrator-5 kit (Zymo Research) and DNA concentrations measured using a Nanodrop spectrophotometer. DNA size and concentration was used to estimate picomolar concentration of each clean PCR product. Vectors and inserts were mixed at a 1:2 molar ratio, with up to 0.2 pmol of DNA in 10-μl reactions, and incubated for 15 min at 50 °C in a thermocycler. Assembly products (3 μl) were transformed into DH5α cells, and plated on LB-agar plates with ampicillin overnight at 37 °C. Single colonies were picked and cultured in LB medium overnight at 37 °C. Plasmids were isolated and sequences confirmed with Sanger sequencing (UICSEQC). The plasmids were transformed into BL21De3 (Invitrogen) cells and plated on LB-agar supplemented with 100 μg/ml ampicillin, overnight at 30 °C.

Design and construction of DNA donors for pneumococcal comW gene replacement, and of the novobiocin resistance gene cassettes

Strains, plasmids, and primers are listed in supporting information (Tables S1–S3). Strains CP2800, CP2803, and CP2805 were generated using restriction digestion and ligation. The flanking upstream (primers NL106 and NL107) and downstream (NL102 and NL103) regions of pneumococcal comW

Pneumococcal ComW binds to DNA

J. Biol. Chem. (2019) 294(29) 11101–11118

11113
Pneumococcal ComW binds to DNA

were PCR amplified with primers containing 5’ BtsI restriction sites using 5 ng of CP2137 genomic DNA, 1 μl of Phire Hot Start II DNA Polymerase (Thermo Fisher Scientific), and 0.2 mM dNTP mix (Thermo Fisher Scientific) in a 50-μl reaction in a thermocycler. The \( \text{comW}^\Delta \) gene variant was PCR amplified using primer pair NL104 and NL105, Phire Hot Start II DNA Polymerase, 5 ng of pNL160, and 0.2 mM dNTP mix in a 50-μl reaction, generating a fragment with 5’ and 3’ BtsI cut sites. The genomic upstream flank that was ligated to \( \text{comW}^\Delta \) was generated using primer pair NL101 and NL106, 50 ng of CP2137 genomic DNA, and the same PCR conditions as primer pair NL106 and NL107. The E19A and Y41A mutations in \( \text{comW} \) were PCR amplified from plasmids pNLI88 and pNLI80, respectively, using primer pair NL105 and NL108, in the same PCR conditions as \( \text{comW}^\Delta \) variant.

To link cassette pieces, 10 μl of the flanking arms and \( \text{comW} \) variant PCR reactions were digested with BtsI (New England Biolabs) overnight at 37 °C in 30-μl reactions, to generate 2-bp overhangs on the ends of each molecule. Digested DNA products were purified using Zymo research kits and the DNA concentration was measured using a nanodrop. A 1:1 flank to gene fragment ratio was used in 30-μl ligation reactions overnight at 16 °C in a thermocycler. Ligation products were transformed into pneumococcus for homologous recombination (see below).

Strains CP2801-CP2802, CP2804, CP2806-CP2807, and CP2808-CP2813 were obtained using Gibson assembly (92) with NEBuilder HiFi Assembly Mix (New England Biolabs). For assembly of pneumococcal \( \text{comW} \) point mutant cassettes, 5 ng of CP2137 genomic DNA was used in PCR with 0.1 units Phusion High-fidelity DNA polymerase (Thermo Fisher Scientific), 1 mM MgCl₂, 0.4 mM dNTP mix, and primers NL103 and NL177 to amplify a 3,327-bp \( \text{comW} \) downstream flank. Pneumococcal \( \text{comW} \) mutants were PCR amplified using 5 ng of specific plasmids with primer pair NL175 and NL176 containing 12-bp overlapping sequence with the upstream and downstream \( \text{comW} \) flanks in the PCR conditions mentioned above. 50 ng of CP2137 genomic DNA with primers NL179 and NL180 was used to amplify a 2,301-bp \( \text{comW} \) upstream flank, in the PCR conditions mentioned above without the use of MgCl₂.

For assembly of \( \text{comW} \) orthologousogenous cassettes, the flanking upstream (primers NL153 and NL154) and downstream (primers NL155 and NL156) regions of pneumococcal \( \text{comW} \) were PCR amplified using 50 ng of CP2137 genomic DNA, 0.02 units of Phusion High-fidelity DNA Polymerase, 0.4 mM dNTP mix, and 10 mM MgCl₂ in a 50-μl reaction in a thermocycler. Orthologous \( \text{comW} \) genes were amplified from plasmids (Table S2) with gene specific primers (Table S3) using 5 ng of plasmid DNA, 0.02 units of Phusion High-Fidelity DNA polymerase, 0.4 mM dNTP mix in 50-μl reactions, in a thermocycler. All PCR products were purified using Zymo kits. A Nanodrop was used to determine the concentration. Concentrations and DNA fragment sizes were used to calculate the number of picomole ends. Flanking arms and \( \text{comW} \) variant fragments were mixed in a 6:1 molar ratio and incubated in 10-μl reactions with HiFi assembly mix at 50 °C for 20 min in a thermocycler. 1 μl of each assembly product was used in subsequent PCR with primers DAM497 and DAM500 for amplification to generate gene cassette products of 2,200 bp. PCR products were purified and sequenced at the UICSQC.

The 7.4-kb gyrB novobiocin resistance marker was prepared by amplification with primers YT76 and YT77, and 10 ng of either CM6 or CP1500 genomic DNA as the template. PCR amplification was performed in 50 μl reactions with 1 μl Phire HotStart II polymerase and Phire reaction buffer, and 0.2 mM of dNTP. The concentration of the purified product was measured in a ND-1000 spectrophotometer.

Construction of \( \text{comW} \) mutant strains

CP2137 cells were cultured in 10 ml of CAT medium to an OD₅₅₀ of 0.05. For transformation, 100 μl of cells were mixed with 100 ng of CSP, 0.04% BSA, 0.001 M CaCl₂, and 100 ng of gene replacement cassettes made by Gibson assembly, or 300 ng of gene replacement cassettes made by restriction digestion and ligation. Transformation reactions were brought to a total volume of 1 ml with CAT, and incubated with CSP and DNA for 2 h at 37 °C, and then chilled. Serial dilutions of each reaction were plated on CAT agar and grown overnight at 37 °C. Single colonies were picked and cultured overnight in CAT supplemented with 0.016 M K₂HPO₄ at 37 °C. Overnight cultures were diluted into 10 ml CAT supplemented with 0.016 M K₂HPO₄ and 0.2% glucose, and grown at 37 °C to an OD₅₅₀ of 0.2. Cells were diluted 1:5 in H₂O and used in PCR reactions to amplify the \( \text{comW} \) region with primers DAM145 and DAM146. PCR products were cleaned and sequenced at the UICSQC.

Pneumococcal transformation assays

Aliquots of frozen pneumococcal strain stocks were diluted 1:100 in 10 ml of CAT medium and cultured at 37 °C for five hours until their OD₅₅₀ measured between 0.2–0.4. Each culture was diluted in CAT to an OD₅₅₀ 0.05, and 10 μl of cells were incubated at 37 °C for 80 min with 100 ng CSP, 0.04% BSA, 0.001 M CaCl₂, and 100 ng novobiocin resistance cassette in 1 ml with CAT medium. Transformation reactions were then diluted 1:100 in THY medium and incubated at 37 °C for an additional 80 min, and then on ice prior to serial dilution. Diluted cells were plated on THY agar with 2.5 μg/ml novobiocin and incubated overnight at 37 °C. Colonies were counted and the transformation efficiency determined as a ratio of transformants/total CFU.

\( \text{ComW}^\Delta \) purification

\( E. \ coli \) transformed with a truncated variant of \( \text{comW} \) (\( \text{ComW}^\Delta \)) in the pET22b+ vector were cultured in 3 L of LB medium at 37 °C, 200 rpm to an OD₅₅₀ 0.5; then protein expression was induced by addition of [1 mM] isopropyl β-D-1-thiogalactopyranoside (IPTG) (Gold Biotechnology). Induced cultures were incubated overnight at 20 °C, 200 rpm. The next day, the OD of each culture was taken using a spectrophotometer, and cells collected at 5,000 rpm, 4 °C, for 30 min. Cells were resuspended in 50 ml of \( E. \ coli \) resuspension buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl) supplemented with 25 mM imidazole, 50 mM MgCl₂, 100 μg/ml DNase, and a protease inhibitor tablet (Roche Applied Science). The cell resuspension was lysed using an emulsiflex. The soluble fraction was separated by centrifugation at 30,000 × g for 30 min at 4 °C.
The soluble supernatant was passed through a 5-ml nickel-nitritotriacetic acid resin (Thermo Fisher Scientific) on a column equilibrated with Column Buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% glycerol) via gravity. The column was washed with 500 ml of Column Buffer supplemented with 5 mM β-mercaptoethanol (βME) and 40 mM imidazole. ComWΔ6 was eluted from the column with 10 ml of column buffer supplemented with 5 mM βME and 275 mM imidazole. The ComWΔ6 eluate was dialyzed into 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1 mM EDTA, 10% glycerol, and 1 mM βME. Protein concentration was measured using a Nanodrop.

**Nano differential scanning fluorimetry**

After purification, ComWΔ6 was diluted to 0.5 mg/ml in ten different biological buffers (Fig. S2). Diluted ComWΔ6 was loaded into 10-μl capillaries in duplicate. Each capillary was loaded into a Prometheus NT.48 instrument (NanoTemper Technologies Inc). Samples were heated from 20 °C to 95 °C with a temperature ramp of 1.0 °C/min. Data were analyzed with the PR.ThermControl software (NanoTemper Technologies Inc).

**Analytical ultra-centrifugation**

Purified ComWΔ6 was diluted to an absorbance (280 nm) of 1.5 (0.92 mg/ml), 1.0 (0.61 mg/ml), and 0.5 (0.31 mg/ml) in cold 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1 mM EDTA, and 1 mM βME. Adjacent channels of analytical ultracentrifugation cells were separately loaded with 400 μl of ComWΔ6 sample or 420 μl of buffer and sealed under pressure. The samples were centrifuged at 40,000 rpm in a Beckman Coulter XL-I at 10 °C for 24 h. Scans were taken every minute. The raw sedimentation data were analyzed using a continuous distribution model with SEDFIT and then analyzed with a Bayesian model.

**Size-exclusion chromatography**

ComWΔ6 variants were concentrated in 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1 mM EDTA, 10% glycerol, and 1 mM βME and applied to a pre-packed Superdex 200 column (GE Healthcare) on an AKTA (Amersham Biosciences). The proteins were eluted at 0.5 ml/min, and collected in 1 ml fractions for analysis on SDS-polyacrylamide gels. Small standards RNase (13.7 kDa), chymotrypsinogen A (25.6 kDa), Ovalbumin (43.0 kDa), and Conalbumin (75.0 kDa) were applied over the same column in the same conditions.

**Circular dichroism (CD)**

Purified ComWΔ6 variants were diluted to 0.02 mg/ml in 30 mM sodium fluoride (NaF) in a 0.2 cm path quartz cell. Their structure was analyzed at 25 °C from 260 nm to 190 nm using a JASCO J-710 machine. Deconvolution of the data were performed using Dichroweb using the Contin-LL method (93–95).

**Electrophoretic mobility shift assay**

EMSA probes were generated using PCR amplification. For amplification of S. pneumoniae’s combox promoter region, upstream of sbbB, 50 ng of CP2137 genomic DNA was used as a template for amplification with primers NL223 (6’FAM) and NL224 for labeled probe, and NL228 and NL224 for unlabeled probe. For amplification of S. mutans ComR-box, 1 μl of S. mutans cells diluted 1:5 in H2O were used as a template for amplification with primers NE 25 (6’FAM) and NE27 for labeled probe, and NE26 and NE27 for unlabeled probe. Probes were PCR purified using Zymo kits and concentration measured using a nanodrop.

Purified protein at varied concentrations, and 20 nM of DNA probe were mixed together in 5× EMSA reaction buffer (250 mM Tris-HCl, pH 8.0, 1.5 mM NaCl, 5 mM MgCl2, 5 mM EDTA), supplemented with 50 mM βME, and the volume of each reaction was brought to 20 μl with ComWΔ6 dialysis buffer plus 30% glycerol. After incubation at 25 °C for 30 min, in a thermostyler, wells of a 5% polyacrylamide, nondenaturing gel were loaded with 10 μl of each EMSA reaction and run at 90 V for 4 h at 4 °C in 1X TBE. Gels were removed from casting plates and imaged on Typhoon 3000 (GE Healthcare). Band intensities were measured using Image Studio Lite software (Li-COR biosciences), and the amount of probe shifted by protein was determined as (bound probe/total probe) for each lane. The gels shown are representative of three biological replicates, and statistical significance was determined using an unpaired student’s t test.

**Western blotting**

Pneumococcal cells were cultured in 10 ml of CAT medium for 4 h, diluted to an OD550 = 0.2, and induced to competence by addition of 100 ng of CSP, 0.04% BSA, and 0.001 mM CaCl2, for 17 min, and then chilled. A 10-μl sample of each culture was removed for plating on CAT agar and incubated overnight at 37 °C to determine CFU. The remaining cells were collected by centrifugation at 4,000 rpm for 30 min at 4 °C. After removal of the supernatant, cell pellets were resuspended in pneumococcal wash saline (0.01 M Tris-HCl, pH 8.0, 0.15 M NaCl, and 0.01 M EDTA), and 0.1 mm beads were added to each tube for lysis by bead beating in the cold for one 5-min interval, followed by one 4-min interval. The lysates were centrifuged for 10 s and 180 μl of clarified cell lysates were moved to a fresh 1.5-ml microcentrifuge tube with 60 μl of 4X Laemmli buffer (Bio Rad). Cell samples were incubated at 95 °C for 3 min and stored at −20 °C.

For protein detection, the wells of a precast 4–20% SDS-PAGE (Bio Rad) gel (8.6 X 6.7 cm, 0.1 cm thickness) were filled with 1 × 106 cells of each sample. Gels were run at 200 V for 35 min. Proteins were then transferred for 2 h at 36 V to a PVDF membrane activated with cold methanol at 30 V at 4 °C in 1X transfer buffer (25 mM Tris-HCl, pH 8.0, 192 mM glycine, 10% MeOH). Membranes were then quickly rinsed with 1× TBST (20 mM Tris-HCl, pH 7.6, 137 mM NaCl, 0.1% Tween 20). Membranes were blocked with 5% BSA in 0.1% TBST for 1 h at room temperature with rocking. Blots were then incubated with a primary rabbit antibody, raised against pneumococcal ComW (48) (1:1000 in 1.0% BSA in 0.1% TBST), plus a primary mouse antibody raised against E. coli RNA polymerase β-subunit (Thermo Fisher Scientific) (1:3000 in 0.5% BSA in 0.1% TBST) overnight at 4 °C, with rocking. After three washes with 0.1% TBST for 5 min at room temperature, blots were simultaneously incubated with goat anti-rabbit and goat anti-mouse secondary antibodies (Sigma) diluted 1:10,000 in 1% BSA in 0.1% TBST, for 1 h and 20 min at room temperature, with rocking.

Pneumococcal ComW binds to DNA
**Pneumococcal ComW binds to DNA**

After one 10-min wash and two 5-min washes in 0.1% TBST at room temperature, blots were incubated with 10 ml of ECL substrate (Bio Rad) for 5 min at room temperature, with rocking. The wet blots were imaged using the LiCOR system. Image Studio Lite was used to quantify the signal intensity from endogenous ComW bands (9.5 kDa) and RNA polymerase (137 kDa). Signal intensities were normalized to the RNA polymerase signal. The blots shown are representative of three biological replicates. Statistical significance was determined using an unpaired Student’s *t* test.

**Author contributions**—N. L. I., G. P., and D. A. M. conceptualization; N. L. I., G. P., and D. A. M. resources; N. L. I. and G. P. data curation; N. L. I., G. P., and D. A. M. formal analysis; N. L. I., G. P., and D. A. M. funding acquisition; N. L. I., G. P., and D. A. M. investigation; N. L. I., G. P., and D. A. M. methodology; N. L. I. and G. P. writing—original draft; N. L. I., G. P., and D. A. M. writing—review and editing; G. P. and D. A. M. supervision; G. P. and D. A. M. project administration.

**Acknowledgments**—We thank The Center for Structural Biology at the University of Illinois at Chicago for use of the lab space equipment. DNA Sanger Sequencing was performed at the University of Illinois at Chicago Sequencing Core (UICSQC).

**References**

1. Griffith, F. (1928) The significance of pneumococcal types. *J. Hyg (Lond)*. 27, 113–159
2. Avery, O. T., Macleod, C. M., and McCarty, M. (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus Type III. *J. Exp. Med.* 79, 137–158
3. Hävarstein, L. S. (2010) Increasing competence in the genus *Streptococcus*. *Mol. Microbiol.* 78, 541–544
4. Andam, C. P., and Hanage, W. P. (2015) Mechanisms of genome evolution of *Streptococcus*. *Infect. Genet. Evol.* 33, 334–342
5. Johnston, C., Martin, B., Fichant, G., Polard, P., and Claverys, J.-P. (2014) Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat. Rev. Microbiol.* 12, 181–196
6. Khan, R., Rukke, H. V., Havik, H., Ámdal, H. A., Chen, T., Morrison, D. A., and Petersen, F. C. (2016) Comprehensive transcriptome profiles of *Streptococcus mutans* UA159 map core streptococcal competence genes. *mSystems* 1, e00038-15
7. Fontaine, L., Boutry, C., de Frahan, M. H., Delplace, B., Fremaux, C., Horvath, P., Boyaval, P., and Hols, P. (2010) A novel pheromone quorum-sensing system controls the development of natural competence in *Streptococcus thermophilus* and *Streptococcus salivarius*. *J. Bacteriol.* 192, 1444–1454
8. Khan, R., Junge, R., Ámdal, H. A., Chen, T., Morrison, D. A., and Petersen, F. C. (2017) A positive feedback loop mediated by Sigma X enhances expression of the streptococcal regulator ComR. *Sci. Rep.* 7, 5984
9. Gao, X.-Y., Zhi, X.-Y., Li, H.-W., Klenk, H.-P., and Li, W.-J. (2014) Comparative genomics of the bacterial genus *Streptococcus* illuminates evolutionary implications of species groups. *PLoS ONE* 9, e101229
10. Lee, M. S., and Morrison, D. A. (1999) Identification of a new regulator in *Streptococcus pneumoniae* linking quorum sensing to competence for genetic transformation. *J. Bacteriol.* 181, 5004–5016
11. Morrison, D. A., and Lee, M. S. (2000) Regulation of competence for genetic transformation in *Streptococcus pneumoniae*: a link between quorum sensing and DNA processing genes. *Res. Microbiol.* 151, 445–451
12. Luo, P., and Morrison, D. A. (2003) Transient association of an alternative sigma factor, ComX, with RNA polymerase during the period of competence for genetic transformation in *Streptococcus pneumoniae*. *J. Bacteriol.* 185, 349–358
13. Burgess, R. R., Travers, A. A., Dunn, I. J., and Bautz, E. K. (1969) Factor stimulating transcription by RNA polymerase. *Nature* 221, 43–46
14. Travers, A. A., and Burgess (1969) Cyclic re-use of the RNA polymerase sigma factor. *Nature* 222, 537–540
15. Mashburn-Warren, L., Morrison, D. A., and Federle, M. J. (2010) A novel double-tryptophan peptide pheromone controls competence in *Streptococcus* spp. via an Rgg regulator. *Mol. Microbiol.* 78, 589–606
16. Mashburn-Warren, L., Morrison, D. A., and Federle, M. J. (2012) The cryptic competence pathway in *Streptococcus pyogenes* is controlled by a peptide pheromone. *J. Bacteriol.* 194, 4589–4600
17. Cheng, Q., Campbell, E. A., Naughton, A. M., Johnson, S., and Masure, H. R. (1997) The Comc locus controls genetic transformation in *Streptococcus pneumoniae*. *Mol. Microbiol.* 23, 683–692
18. Peterson, S. N., Sung, C. K., Cline, R., Desai, B. V., Snesrud, E. C., Luo, P., Walling, J., Li, H., Mintz, M., Tsegay, G., Burr, P. C., Do, Y., Ahn, S., Gilbert, J., Fleischmann, R. D., and Morrison, D. A. (2004) Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. *Mol. Microbiol.* 51, 1051–1070
19. Parashar, V., Aggarwal, C., Federle, M. J., and Neiditch, M. B. (2015) Rgp protein structure—function and inhibition by cyclic peptide compounds. *Proc. Natl. Acad. Sci. U.S.A.* 112, 5177–5182
20. Neiditch, M. B., Capodagli, G. C., Prehna, G., and Federle, M. J. (2017) Genetic and structural analyses of RRNP intercellular peptide signaling of gram-positive bacteria. *Annu. Rev. Genet.* 51, 311–333
21. Shanker, E., Morrison, D. A., Talagás, A., Nessler, S., Federle, M. J., and Prehna, G. (2016) Pheromone recognition and selectivity by ComR proteins among *Streptococcus* species. *PLoS Pathog.* 12, e1005979
22. Talagas, A., Fontaine, L., Legedma-Garcia, L., Mignolet, J., Li de la Sierra-Galay, I., Lazar, N., Aumont-Nicaise, M., Federle, M. J., Prehna, G., Hols, P., and Nessler, S. (2016) Structural insights into streptococcal competence regulation by the cell-to-cell communication system ComR. *PLoS Pathog.* 12, e1005980
23. Fontaine, L., Goffin, P., Dubout, H., Delplace, B., Baulard, A., Lecat-Guillet, N., Chambellon, E., Gardan, R., and Hols, P. (2013) Mechanism of competence activation by the ComR signalling system in streptococci. *Mol. Microbiol.* 87, 1113–1132
24. Hävarstein, L. S., Coomaraswamy, G., and Morrison, D. A. (1995) An unmodified heptadecapeptide pheromone. *Proc. Natl. Acad. Sci. U.S.A.* 92, 11140–11144
25. Li, Y. H., Lai, P. C., Lee, J. H., Ellen, R. P., and Cvtikovitch, D. G. (2001) Natural genetic transformation of *Streptococcus mutans* growing in biofilms. *J. Bacteriol.* 183, 897–908
26. Hui, F. M., Zhou, L., and Morrison, D. A. (1995) Competence for genetic transformation in *Streptococcus pneumoniae*: organization of regulatory locus with homology to two lactococcin A secretion genes. *Gene* 153, 25–31
27. Hui, F. M., and Morrison, D. A. (1991) Genetic transformation in *Streptococcus pneumoniae*: nucleotide sequence analysis shows comA, a gene required for competence induction, to be a member of the bacterial ATP-dependent transport protein family. *J. Bacteriol.* 173, 372–381
28. Chandler, M. S., and Morrison, D. A. (1988) Identification of two proteins encoded by com, a competence control locus of *Streptococcus pneumoniae*. *J. Bacteriol.* 170, 3136–3141
29. Hävarstein, L. S., Gaustad, P., Nes, I. F., and Morrison, D. A. (1996) Iden- tification of the streptococcal competence- pheromone receptor. *Mol. Microbiol.* 21, 863–869
30. Martin, B., Granadel, C., Campo, N., Hénard, V., Prudhomme, M., and Claverys, J.-P. (2010) Expression and maintenance of Dcm-ComE, the two-component signal-transduction system that controls competence of
50. Bartilson, M., Marra, A., Christine, J., Asundi, J. S., Schneider, W. P., and Brazier, J., and Thomas, D. W. (2002) Heterogeneity within the gram-positive anaerobic cocci demonstrated by analysis of 16S–23S intergenic ribosomal RNA polymorphisms. J. Med. Microbiol. 51, 949–957 CrossRef Medline

51. Hill, K. E., Davies, C. E., Wilson, M. J., Stephens, P., Lewis, M. A., Hall, V., Brazier, J., and Thomas, D. W. (2002) Heterogeneity within the gram-positive anaerobic cocci demonstrated by analysis of 16S–23S intergenic ribosomal RNA polymorphisms. J. Med. Microbiol. 51, 949–957 CrossRef Medline

52. Bennett-Lovsey, R. M., Herbert, A. D., Sternberg, M. J., and Kelley, L. A. (2008) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. Proteins 70, 611–625 Medline

53. Chivian, D., Kim, D. E., Malmström, L., Bradley, P., Robertson, T., Murphy, P., Strauss, C. E. M., Bonneau, R., Rohl, C. A., and Baker, D. (2003) Automated prediction of CAP-5 structures using the Robetta server. Proteins 53, Suppl. 6, 524–533 CrossRef Medline

54. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. J. Mol. Biol. 233, 123–138 CrossRef Medline

55. Ashkenazy, H., Ezrez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 38, W329–W353 CrossRef Medline

56. Luo, P., Li, H., and Morrison, D. A. (2004) Identification of ComW as a new component in the regulation of genetic transformation in Streptococcus pneumoniae. Mol. Microbiol. 54, 172–183 CrossRef Medline

57. Piotrowski, A., Luo, P., and Morrison, D. A. (2009) Competence for genetic transformation in Streptococcus pneumoniae: termination of activity of the alternative σ factor ComX is independent of proteolysis of ComX and ComW. J. Bacteriol. 191, 3359–3366 CrossRef Medline

58. Sung, C. K., and Morrison, D. A. (2005) Two distinct functions of ComW in stabilization and activation of the alternative σ factor ComX in Streptococcus pneumoniae. J. Bacteriol. 187, 3052–3061 CrossRef Medline

59. Tovpеко, Y., and Morrison, D. A. (2014) Competence for genetic transformation in Streptococcus pneumoniae: mutations in A bypass the comW requirement. J. Bacteriol. 196, 3724–3734 CrossRef Medline

60. Tovpеко, Y., Bai, I., and Morrison, D. A. (2016) Competence for genetic transformation in Streptococcus pneumoniae: mutations in σA bypass the ComW requirement for late gene expression. J. Bacteriol. 198, 2370–2378 CrossRef Medline

61. Gruber, T. M., and Gross, C. A. (2003) Multiple σ subunits and the partitioning of bacterial transcription space. Annu. Rev. Microbiol. 57, 441–466 CrossRef Medline

62. Paget, M. (2015) Bacterial σ factors and anti-σ factors: structure, function and distribution. Biomolecules 5, 1245–1265 CrossRef Medline

63. Cavaliere, P., and Norel, F. (2016) Recent advances in the characterization of Crl the unconventional activator of the stress σ factor σS/RpoS. Biol. Mol. Concepts 7, 197–204 Medline

64. Fox, T. D., Losick, R., and Pero, J. (1976) Regulatory gene 28 of bacteriophage SP01 codes for a phage-induced subunit of RNA polymerase. J. Mol. Biol. 101, 427–433 CrossRef Medline

65. Talkington, C., and Pero, J. (1978) Promoter recognition by Phage SP01-modified RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 75, 1185–1189 CrossRef Medline

66. Pero, J., Nelson, J., and Fox, T. D. (1975) Highly asymmetric transcription by RNA polymerase containing Phage SP01-induced polypeptides and a new host protein. Proc. Natl. Acad. Sci. U.S.A. 72, 1589–1593 CrossRef Medline

67. Nechaev, S., Kamali-Moghaddam, M., André, E., Léonetti, J.-P., and Geiduschek, E. P. (2004) The bacteriophage T4 late-transcription coactivator torgp33 binds the flap domain of Escherichia coli RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 101, 17365–17370 CrossRef Medline

68. Williams, K. P., Müller, R., Rüger, W., and Geiduschek, E. P. (1989) Over-produced bacteriophage T4 gene 33 protein binds RNA polymerase. J. Bacteriol. 171, 3579–3582 CrossRef Medline

69. MacLellan, S. R., Guariglia-Oropesa, V., Gaballa, A., and Helmann, J. D. (2009) A two-subunit bacterial -factor activates transcription in Bacillus subtilis. Proc. Natl. Acad. Sci. U.S.A. 106, 21233–21238 CrossRef Medline

70. Murakami, K. S., and Darst, S. A. (2003) Bacterial RNA polymerses: the whole story. Curr. Opin. Struct. Biol. 13, 31–39 CrossRef Medline

71. Piotrowski, A. (2010) Regulation of the activities of ComX and ComW during competence development in Streptococcus pneumoniae. P. J. R. 66, 623–633 CrossRef Medline

72. Raman, S., Vernon, R., Thompson, I., Tyka, M., Sadreyev, R., Pei, J., Kim, D., Kellogg, E., DiMaio, F., Lange, O., Kinc, L., Sheffler, W., Kim, B.-H., Das, R., Grishin, N. V., and Baker, D. (2009) Structure prediction for CAP8 with all-atom refinement using Rosetta. Proteins 77, 89–99 CrossRef Medline

73. Bartelson, M., Marra, A., Christine, J., Asundi, J. S., Schneider, W. P., and Hromocký, A. E. (2001) Differential fluorescence induction reveals Streptococcus pneumoniae loci regulated by competence stimulatory peptide. Mol. Microbiol. 39, 126–135 CrossRef Medline
Pneumococcal ComW binds to DNA

71. Liu, H., Shang, W., Hu, Z., Zheng, Y., Yuan, J., Hu, Q., Peng, H., Cai, X., Tan, L., Li, S., Zhu, J., Li, M., Hu, X., Zhou, R., Rao, X., and Yang, Y. (2018) A novel SigB(Q225P) mutation in Staphylococcus aureus retains virulence but promotes biofilm formation. Emerg. Microbes Infect. 7, 72 10.1038/s41426-018-0078-1 Medline

72. Gu, D., Guo, M., Yang, M., Zhang, Y., Zhou, X., and Wang, Q. (2016) A σE-mediated temperature gauge controls a switch from LuxR-mediated virulence gene expression to thermal stress adaptation in Vibrio alginolyticus. PLoS Pathog. 12, e1005645 CrossRef Medline

73. Jones, C. H., and Moran, C. P., Jr. (1992) Mutant σ factor blocks transition between promoter binding and initiation of transcription. Proc. Natl. Acad. Sci. U.S.A. 89, 1958–1962 CrossRef Medline

74. Jones, C. H., Tatti, K. M., and Moran, C. P., Jr. (1992) Effects of amino acid substitutions in the −10 binding region of σE from Bacillus subtilis. J. Bacteriol. 174, 6815–6821 CrossRef Medline

75. Juang, Y.-L., and Helmann, J. D. (1994) A promoter melting region in the primary σ factor of Bacillus subtilis: identification of functionally important aromatic amino acids. J. Mol. Biol. 235, 1470–1488 CrossRef Medline

76. Juang, Y.-L., and Helmann, J. D. (1995) Pathway of promoter melting by Bacillus subtilis RNA polymerase at a stable RNA promoter: effects of temperature, .delta. protein, and . σ. factor mutations. Biochemistry 34, 8465–8473 CrossRef Medline

77. deHaseth, P. L., and Helmann, J. D. (1995) Open complex formation by Escherichia coli RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA. Mol. Microbiol. 16, 817–924 CrossRef Medline

78. Tomsic, M., Tsujikawa, L., Panaghi, G., Wang, Y., Azok, J., and deHaseth, P. L. (2001) Different roles for basic and aromatic amino acids in conserved region 2 of Escherichia coli σ70 in the nucleation and maintenance of the single-stranded DNA bubble in open RNA polymerase-promoter complexes. J. Biol. Chem. 276, 31891–31896 CrossRef Medline

79. Basu, R. S., Warner, B. A., Molodtsov, V., Pupov, D., Esyunina, D., Fernández-Tornero, C., Kulbachinskiy, A., and Murakami, K. S. (2014) Structural basis of transcription initiation by bacterial RNA polymerase holozyme. J. Biol. Chem. 289, 24549–24559 CrossRef Medline

80. Chamberlin, M., and Berg, P. (1962) Deoxyribonucleic acid-directed synthesis of ribonucleic acid by an enzyme from Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 48, 81–94 CrossRef Medline

81. Missiakas, D., Mayer, M. P., Lemaire, M., Georgopoulos, C., and Raina, S. (1997) Modulation of the Escherichia coli σE (RpoE) heat-shock transcription-factor activity by the RseA, RseB, and RseC proteins. Mol. Microbiol. 24, 355–371 CrossRef Medline

82. Brown, M. S., Ye, J., Rawson, R. B., and Goldstein, J. L. (2000) Regulated intramembrane proteolysis: review a control mechanism conserved from bacteria to humans. Cell 100, 391–398 CrossRef Medline

83. Heinrich, J., and Wiegent, T. (2009) Regulated intramembrane proteolysis in the control of extracytoplasmic function σ factors. Res. Microbiol. 160, 696–703 CrossRef Medline

84. Konovalova, A., Seggaard-Andersen, L., and Kroos, L. (2014) Regulated proteolysis in bacterial development. FEMS Microbiol. Rev. 38, 493–522 CrossRef Medline

85. England, P., Westblade, L. F., Karimova, G., Robbe-Saule, V., Norel, F., and Kolb, A. (2008) Binding of the unorthodox transcription activator, Crl, to the components of the transcription machinery. J. Biol. Chem. 283, 33455–33464 CrossRef Medline

86. Banta, A. B., Cuff, M. E., Lin, H., Myers, A. R., Ross, W., Joachimiak, A., and Gourse, R. L. (2014) Structure of the RNA polymerase assembly factor Crl and identification of its interaction surface with Sigma S. J. Bacteriol. 196, 3279–3288 CrossRef Medline

87. MacLellan, S. R., Wecke, T., and Helmann, J. D. (2008) A previously unidentified σ factor and two accessory proteins regulate oxalate decarboxylase expression in Bacillus subtilis. Mol. Microbiol. 69, 954–967 CrossRef Medline

88. Weng, L., Biswas, I., and Morrison, D. A. (2009) A self-deleting Cre-lox-ermAM cassette, Cheshire, for marker-less gene deletion in Streptococcus pneumoniae. J. Microbiol. Methods 79, 353–357 CrossRef Medline

89. Mandel, M., and Higa, A. (1970) Calcium-dependent bacteriophage DNA infection. J. Mol. Biol. 69, 159–162 CrossRef Medline

90. Bertani, G. (1951) Studies on Lysogenesis. J. Bacteriol. 62, 293–300 Medline

91. Zhang, Y., Werling, U., and Edelmann, W. (2012) SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res. 40, e55 CrossRef Medline

92. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., Smith, J., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345 CrossRef Medline

93. Whitmore, L., and Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res. 32, W668–W673 CrossRef Medline

94. Provencher, S. W., and Göckner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. Biochemistry 20, 33–37 CrossRef Medline

95. van Stokkum, I. H., Spoelder, H. J., Bloemendal, M., van Grondelle, R., and Groen, F. C. (1992) Estimation of protein secondary structure from circular dichroism spectra. Anal. Biochem. 191, 110–118 CrossRef Medline