Mapping the transcription start point of the *Staphylococcus aureus* *eap*, *emp* and *vwb* promoters reveals a conserved octanucleotide sequence that is essential for expression of these genes.

Niamh Harraghy¹*, Dagmar Homerova², Mathias Herrmann¹, Jan Kormanec²

¹Institute of Medical Microbiology and Hygiene, University of Saarland, 66421 Homburg/Saar, Germany
²Institute of Molecular Biology, Slovak Academy of Sciences, 84551 Bratislava, Slovakia

**Short title:** Role of the COS in the *eap*, *emp* and *vwb* promoters

*Corresponding author and present address:

Institute of Biotechnology UNIL
EPFL-FSB LBTM (Bâtiment CH)
Station 6
CH-1015 Lausanne
Switzerland
Tel: +41 21 693 76 16
Fax: +41 21 693 76 10
E-mail: niamh.harraghy@unil.ch
Abstract

Mapping the transcription start point of the *eap*, *emp* and *vwb* promoters revealed a conserved octanucleotide sequence (COS). Deleting this sequence abolished expression of *eap*, *emp* and *vwb*. However, electrophoretic mobility shift assays (EMSA) gave no evidence that this sequence was a binding site for SarA or SaeR, known regulators of *eap* and *emp*. 
The ability of *Staphylococcus aureus* to cause such diverse infections as endocarditis, pneumonia, skin infections and biofilms is linked to its great repertoire of virulence factors, including adhesins, immunomodulatory molecules and toxins (31). The *S. aureus* cell surface adhesins belong to one of two groups, the MSCRAMMs (microbial surface components recognizing adhesive matrix molecules), which include Protein A, Clumping Factors A and B and the Fibronectin binding proteins; and the SERAMs (secretable expanded repertoire adhesive molecules), which include the Extracellular adherence protein (Eap), the Extracellular matrix binding protein (Emp) and the Extracellular fibrinogen-binding protein (Efb) (reviewed in (5, 7)). The MSCRAMMs contain an LPXTG motif, which is involved in anchoring them to the staphylococcal cell surface (10, 32, 45), while the SERAMs lack this motif and may bind to the staphylococcal cell surface either covalently or via specific cell-surface receptors (e.g. (11, 25, 35)). Together the MSCRAMMs and SERAMs facilitate attachment of *S. aureus* to eukaryotic cells, platelets, extracellular matrix proteins and inert surfaces (reviewed in (43)) and may aid the survival and persistence of the staphylococci in the host due to their ability to interfere with the host’s immune response (1, 4, 15, 22, 28, 30, 41, 44, 47).

Eap, Emp and von Willebrand factor binding protein (vWbp) are members of the SERAM family (5). While Eap is functionally well characterized (18), less is known about Emp and vWbp. Emp was described as an extracellular matrix binding protein but additional functional roles have not yet been described (23). vWbp was identified during a screen for factors binding von Willebrand factor and was subsequently shown to be a coagulase (2, 3). Although the members of the SERAM family do not share significant sequence homology, they are recognized as sharing similar functional properties, such as being important in adhesion and modulation of the host
immune response to staphylococcal infections (1-3, 5, 12, 16, 17, 23, 29, 30, 42, 44, 47). What is not yet known, however, is whether the regulation of the SERAMs at the molecular level is governed by a common mechanism or factor.

We became interested in studying the regulation of emp and vwb as these two genes together with clfA are located adjacent to each other (SACOL 0856-0858), with 223 bp separating clfA from vwb and 353 bp separating vwb from emp. As ClfA is also an important S. aureus virulence factor, an interesting scenario would be the co-transcription of clfA, vwb, and emp. We mapped the transcription start points of emp and vwb in S. aureus strain Newman using primer extension analysis, as described in (19) and found that both genes had their own promoter (Fig. 1A). These findings suggest the genes are not co-transcribed and fit with the observations by us and others that the expression profiles of the three genes are different (3, 19, 34, 46). eap was also found to be transcribed from a single promoter (Fig. 1A). Putative promoter elements were identified by analysis of the region upstream of the transcription start point. All three promoters have a conserved -10 box (Fig. 1B), but homology to the consensus -35 box, TTGACA (20, 33, 36, 37), is less conserved, particularly in the vwb promoter. However, we have found a conserved octanucleotide sequence (COS), AGTTAATT, that is just 5’ to a putative -35 box in each promoter (Fig. 1B). Moreover, searching the S. aureus COL genome for this COS revealed a COS in the same position (i.e. immediately upstream of a putative -35 box) in the promoter of several important virulence factors (Table 1). A common feature of these virulence factors is that they are involved in modulating the immune response to S. aureus infections or antibiotic resistance (5, 9, 24, 48). Taken together, these data suggested that the COS could be important in the regulation of these genes.
To investigate the importance of the COS we deleted it in the eap, emp and vwb promoters in a two-step PCR reaction. For deleting the COS in the emp promoter, two primer pairs were used. Primers emp-cs_R (5’-GTTTACTTCAATTACTGAATAATTC-3’) and emp-cs_F (5’-GAATTTCAGTATAATTGAAATGAAAC-3’) are complementary and lack the COS. Primers empPF1 and empPR1 were described previously (19). In the first PCR, primers empPF1 and emp-cs_F were used to amplify the region 5’ of the COS while primers empPR1 and emp-cs_R amplified the region 3’ of the COS. In the second PCR, the two PCR products were joined together using primers empPF1 and empPR1. For deleting the COS in the vwb promoter primers vwbPF1 (5’-TTCGAATTCTAGCTCCTGGACTC-3’) and vwbPR1 (5’-CCTAAGCTTAAATTTCCCCTAATTAAAC-3’) amplified the entire promoter region, while primers vwb-cs_F (5’-CTACCTTTTAAATAAGGTGAATGAA-3’) and vwb-cs_R (5’-ATTCATCAACATTTTAAAAAGGTAG-3’) were the complementary internal primers lacking the COS. The COS in the eap promoter was deleted using the QuikChange Mutagenesis Kit (Stratagene) using primers QCF1 (5’-GATAATTTATTATTAATATCCCAGAATATAGAAAGTCTGAC-3’) and QCR1 (5’-GCCAGACTTTTCTCTATTGGATATATATAAATTATC-3’). All clones were sequenced to confirm deletion of the COS and that no additional mutations had been introduced during cloning. The mutated promoters were cloned in their respective reporter gene vectors and transduced into strain Newman as described in (19). Expression of eap was analysed using a bioluminescence assay, while emp and vwb were analysed using a β-galactosidase assay. As shown in Fig. 2, deleting the COS in all three promoters severely repressed expression of the reporter gene. To exclude the possibility that deletion of the COS per se was responsible for the
decrease in expression, we mutated the COS in the *emp* promoter, changing the sequence from AGTTAATT to TCATAATT (thereby changing the first three nucleotides of the COS while leaving the putative -35 box intact) using the QuikChange Mutagenesis Kit (Stratagene) with primers QCF3_emp (5’-GACAACGTTTACTTCATCATAATTATTATATCTGAAAATTCTGG-3’) and QCF3_emp-r (5’-CCAGAATTTTCAGTATAATAATTATGATGAAGTAAACGTTGTC-3’). As shown in Fig. 3, mutagenesis of the COS in the *emp* promoter resulted in a more than 50% decrease in *emp* expression but did not completely abrogate expression. This is likely due to the partial homology of the region to the COS. Taken together our findings suggested that the COS could be the binding site for a regulator of *eap*, *emp* and *vwb*.

In our previous study, we showed that *sarA* and RNAIII are involved in the regulation of *eap* and *emp* and that *sae* is essential for expression of both genes (19). Six of the eleven genes in Table 1 are also regulated by *sae* (39). To investigate if SaeR was binding to the *eap*, *emp* and *vwb* promoters, SaeR was amplified from *S. aureus* Newman using primers saeR_F2 (5’-GGCATACATATGACCCACTTACTGATC-3’) saeR_R3 (5’-CCCCCAAGCTTATCGGCTCCTTTCAAATTTATATCC-3’) and cloned in the pET28a vector (Novagen) and overexpressed in *E. coli*. The purified protein was subsequently assessed for binding to each promoter (see supplementary data for the DNA sequences used) using electrophoretic mobility shift assays (EMSA). However, no binding of SaeR to the promoters was found (data not shown).

As it is possible that SaeR needs to be phosphorylated to bind to its target promoters (13), we decided to purify the DNA binding domain of SaeR and looked for binding of this to the promoters using EMSA. However, we did not observe any binding of the
SaeR DNA binding domain to the three promoters. These findings as well as the observation that \( vwb \) is not regulated by \( sae \) (39) and that some \( sae \)-regulated genes, e.g. \( scn \) and \( chp \) (40), do not have a COS in their promoter, suggests the COS is not the binding site for SaeR.

The COS is similar to a proposed binding site for SarA (AGTTAAG) (38). As SarA is known to be involved in the regulation of \( eap \) and \( emp \), and SarA binding to different promoters has been demonstrated (6, 38), we investigated whether the COS could be a binding site for SarA. Although SarA binds to the \( eap \), \( emp \) and \( vwb \) promoters (Harraghy and Kormanec, unpublished data), deleting the COS did not have any effect on SarA binding to the three promoters (data not shown), indicating that the COS is not essential for SarA binding.

In summary, we have identified a conserved octanucleotide sequence (COS) in the \( eap \), \( emp \) and \( vwb \) promoters as well as in the promoters of several genes recognised as being involved in modulation of the immune response to staphylococcal infection. The nature of the relationship between the SERAMs (5) and leukocidins is intriguing as it was recently shown that in some strains expression of the Panton-Valentine leukocidin interferes with the regulation of the other major group of staphylococcal adhesins, the MSCRAMMs (27). Although it is unlikely that the leukocidins described here affect regulation of the SERAMs, it is possible that they share a common regulator.

Our findings suggest that the COS has an important functional role because deletion of the COS in the \( eap \), \( emp \) and \( vwb \) promoters as well as mutation of the COS in the \( emp \) promoter affected expression of the reporter gene. Although deleting the COS only partially disrupted the proposed -35 box (in the case of the \( eap \) promoter there is only one mismatch when compared with the original promoter) the deletion
dramatically affected promoter activity. Moreover, the mutation of the COS in the *emp* promoter, which preserved the -35 box and maintained homology to the COS (only three bases were different), also affected promoter activity, although not to the same extent as when the COS was deleted. Thus the changes in expression of the reporter genes appear to be the result of modifications to the COS and the possible loss of a transcription factor -binding site. Our findings suggest that the COS is the binding site for an as yet unidentified regulator of *eap*, *emp* and *vwb* that may function together with SaeR. The existence of such a factor was postulated by Goerke et al., (14) and is supported by work in our laboratory as well as the recent findings of Kuroda et al., (26). Emerging data from microarray studies and on-going work in our laboratory will help reveal such candidates.

Acknowledgements

We thank Sylvain Kerudou and Markus Bischoff for critical reading of the manuscript and Karin Hilgert for excellent technical assistance. The work in our laboratories is funded by grants from the University of Saarland HOMFOR to N.H., the Deutsche Forschungsgemeinschaft He 1850/8-1 to M. H. and the VEGA grant 2/6010/26 from Slovak Academy of Sciences to J. K.
Figure and Table legends

**Figure 1:** Mapping the transcription start points (TSP) of the *emp*, *vwb* and *eap* promoters (modified from Harraghy et al. (19), with permission). (A) The TSP of the *vwb*, *emp*, and *eap* promoters were mapped by primer extension analysis as described in (19). (B) Putative -35 and -10 elements were identified by visual inspection of the region upstream of the TSP and are shown in bold. The conserved octanucleotide sequence (COS) is highlighted.

**Figure 2:** The effect of deleting the COS in the *emp* (A), *vwb* (B) and *eap* (C) promoters. The β-galactosidase assay was used for measuring *emp* and *vwb* promoter activity and the bioluminescence assay described in Harraghy et al. (19) was used for measuring *eap* expression. The data shown are the mean ± SEM of at least two independent experiments.

**Figure 3:** Effect of mutating the COS on *emp* expression. The COS was changed from AGTTAATT to TCATAATT, and the effect on *emp* expression was assayed using the β-galactosidase assay as described in (19). The data shown are the mean ± SEM of two independent experiments.

**Table 1:** Mapping the conserved octanucleotide sequence (COS) in selected promoters relative to their putative -35 and -10 boxes. The COS is shown in bold

- putative – 35 and -10 boxes identified by visual inspection of the respective promoters. Putative -35 boxes are underlined and in italics; putative -10 boxes are underlined and experimentally mapped transcription start points are in bold and underlined.

- SaeRS regulated genes (39)
lukE and lukD form a bi-cistronic operon. The COS in lukD is located just upstream of a RBS.

The COS is also found on the complementary strand of the abcA (21) and icaR promoters (8).

Reference List


properties of the multifunctional *Staphylococcus aureus* protein Eap. Microbiology 149:2701-2707.


Table: 1 Mapping the conserved octanucleotide sequence (COS) relative to the -35 and -10 boxes in selected promoters

<table>
<thead>
<tr>
<th>ORFID (COL and N315)</th>
<th>Gene</th>
<th>Mapping the COS relative to the -35 and -10 boxes in the selected promotersa</th>
</tr>
</thead>
<tbody>
<tr>
<td>SACOL2002 SACOL2002</td>
<td>SA1751</td>
<td>eap\textsuperscript{b} taatatc\textsuperscript{AGTTAATT}tgaagaatagctggaatattatat N\textsubscript{36} ATG</td>
</tr>
<tr>
<td>SACOL0858 SACOL0858</td>
<td>SA0744</td>
<td>emp\textsuperscript{b} ttacttca\textsuperscript{AGTTAATT}atatattcgaagaattctgtataaaatat N\textsubscript{63} ATG</td>
</tr>
<tr>
<td>SACOL0857 SACOL0857</td>
<td>SA0743</td>
<td>vwb\textsuperscript{b} tcataca\textsuperscript{AGTTAATT}attaatgagatattatatatat N\textsubscript{66} ATG</td>
</tr>
<tr>
<td>SACOL1168 SACOL1168</td>
<td>SA1003</td>
<td>efb\textsuperscript{b} gtgttta\textsuperscript{AGTTAATT}taatatagttgtaaaaatgataactt N\textsubscript{66} ATG</td>
</tr>
<tr>
<td>SACOL2419 SACOL2419</td>
<td>SA2207</td>
<td>hlgA\textsuperscript{b} taacgaat\textsuperscript{AGTTAATT}gaaacgctataatgatattatat N\textsubscript{27} ATG</td>
</tr>
<tr>
<td>SACOL2421 SACOL2421</td>
<td>SA2208</td>
<td>hlgC\textsuperscript{b} ttaatga\textsuperscript{AGTTAATT}ataatacggccaaatatattatat N\textsubscript{41} ATG</td>
</tr>
<tr>
<td>SACOL2418 SACOL2418</td>
<td>SA2206</td>
<td>sbi\textsuperscript{b} taatattg\textsuperscript{AGTTAATT}taatgttctcatgtgatatata N\textsubscript{36} ATG</td>
</tr>
<tr>
<td>SACOL1881 SACOL1881</td>
<td>SA1638</td>
<td>lukE\textsuperscript{c} ttaatga\textsuperscript{AGTTAATT}ataataatcgaatatataat N\textsubscript{34} ATG</td>
</tr>
<tr>
<td>SACOL1880 SACOL1880</td>
<td>SA1637</td>
<td>lukD\textsuperscript{c} agata\textsuperscript{AGTTAATT}ataatgaaagaatgactaatat N\textsubscript{34} ATG</td>
</tr>
<tr>
<td>SACOL1389 SACOL1389</td>
<td>SA1300</td>
<td>parE\textsuperscript{c} atgaa\textsuperscript{AGTTAATT}gataatgctgacttataat N\textsubscript{63} ATG</td>
</tr>
<tr>
<td>SACOL2136 SACOL2136</td>
<td>SA1946</td>
<td>Hypothetical protein of czr operon</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Numbers in brackets indicate the position of the octanucleotide sequence relative to the -10 and -35 boxes.
Fig 2
Fig 3

![Graph showing RLU over time for emp and emp(mut) conditions.]

- **RLU**
  - emp
  - emp(mut)

- **Time**
  - 3h
  - 6h

- emp bars are higher than emp(mut) bars at both 3h and 6h.