Computational analysis suggests that virulence of *Chromobacterium violaceum* might be linked to biofilm formation and poly-NAG biosynthesis

Sidnei Becker, Cíntia Soares and Luismar Marques Porto

*Laboratório de Tecnologias Integradas, Departamento de Engenharia Química e de Alimentos, Universidade Federal de Santa Catarina, Florianópolis, SC, Brazil.*

**Abstract**

Groups of genes that produce exopolysaccharide with a N-acetyl-D-glucosamine monomer are in the genome of several pathogenic bacteria. *Chromobacterium violaceum*, an opportunistic pathogen, has the operon hmsHFR-CV2940, whose proteins can synthesize such polysaccharide. In this work, multiple alignments among proteins from bacteria that synthesize such polysaccharide were used to verify the existence of amino acids that might be critical for pathogen activity. Three-dimensional models were generated for spatial visualization of these amino acid residues. The analysis carried out showed that the protein HmsR preserves the amino acids D135, D228, Q264 and R267, considered critical for the formation of biofilms and, furthermore, that these amino acids are close to each other. The protein HmsF of *C. violaceum* preserves the residues D86, D87, H156 and W115. It was also shown that these residues are also close to each other in their spatial arrangement. For the proteins HmsH and CV2940 there is evidence of conservation of the residues R104 and W94, respectively. Conservation and favorable spatial location of those critical amino acids that constitute the proteins of the operon indicates that they preserve the same enzymatic function in biofilm synthesis. This is an indicator that the operon hmsHFR-CV2940 is a possible target in *C. violaceum* pathogenicity.

**Key words:** biofilms, exopolysaccharide, *Chromobacterium violaceum* pathogenicity, comparative genomics.

Received: November 4, 2008; Accepted: May 25, 2009.

**Introduction**

Some microorganisms develop cooperative strategies in the formation of biofilms. Biofilm can be defined as interdependent communities of microorganisms, usually connected with a surface presenting high resistance to environmental stress. It consists of a complex symbiotic system of great importance to biotechnological and medical applications, since it is correlated with bacterial resistance to antibiotics and promotion of lethal infections.

The formation of biofilms in several microorganisms involves the presence of a complex matrix where the polymer poly(beta-1,6-N-acetyl-D-glucosamine), poly-NAG, can play an important role (Itoh et al., 2005), possibly related to resistance to antibiotics and promotion of bacterial infections. Poly-NAG has been implicated in the formation of biofilms in several pathogenic bacteria. The polymer, which is involved in cell adhesion in *Staphylococcus epidermidis* (Mack et al., 1996), is also involved in abiotic surface binding, and in intercellular adhesion formation of biofilm in *Escherichia coli* (Itoh et al., 2005).

The hmsHFRS operon has been involved in the formation of biofilms in vitro in *Yersinia pestis*. This operon was related to the blockage of the digestive system of fleas, and it is associated with the transmission of *Y. pestis* to mammals (Jarrett et al., 2004). The bacterium *Bordetella pertussis* has the operon bpsABCD related to a polymer that contributes to the formation of biofilms in the respiratory tract of rats (Sloan et al., 2007). An understanding of the relationship between the operons of those bacteria in the formation of biofilms could potentially lead to the development of a vaccine applicable to different bacterial species.

Recently, the sequencing of the complete genome of the *Chromobacterium violaceum*, strain ATCC 12472 (Brazilian National Genome Project Consortium, 2003; www.brgene.lncc.br; www.ncbi.nlm.nih.gov; access number NC_005085) was carried out to promote domestic genome projects. The investment was justified due to the great biotechnological and pharmaceutical potential of this microorganism. *Chromobacterium violaceum* is considered a possible pathogen for humans and animals, with several reported cases of infection in humans and animals in tropical and subtropical areas, where it is normally found.
Its potential pathogenicity in humans was first described in 1927 in Malaysia (Sneath et al., 1953). Although cases of infection with *C. violaceum* are rare, it has a mortality rate of more than 57%, as reported in a case of chronic granulomatous disease in children (Macher et al., 1982). *C. violaceum* reveals some potentially pathogenic genes (Brazilian National Genome Project Consortium, 2003); however, their extent is not yet known or how these genes are organized in their genome.

This work aims to link, through comparative analysis, the genes hmsH, hmsF, hmsR and the ORF CV2940 of *C. violaceum* with the genes of organisms known and related to the formation of biofilm in an effort to shed light on the pathogenic potential of *C. violaceum*. The analysis of the *C. violaceum* genome reveals the presence of possible genes involved in the formation of biofilms, and the sequence of the group of genes hmsHFR and CV2940 are potentially functional for it. Initially, the relationship among the proteins is carried out by the alignment of amino acid sequences and it is complemented by structural modeling in order to identify the importance of the amino acid through spatial viewing.

**Methods**

Comparisons were carried out using the amino acid sequences of the organisms listed in Table 1 and obtained from GenBank. These organisms with their genomic regions were selected for this study because of their confirmed connection with biofilm formation or production of polysaccharides such as poly(beta-1,6-N-acetyl-D-glucosamine).

### Table 1 - Bacteria with their genomic regions used in the comparative study.

<table>
<thead>
<tr>
<th>Organism and GenBank access number</th>
<th>GenBank access number</th>
<th>Gene/ORF</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Chromobacterium violaceum</em> ATCC 12472</td>
<td>NP_902610</td>
<td>CV2940</td>
</tr>
<tr>
<td>NC_005085</td>
<td>NP_902611</td>
<td>hmsR</td>
</tr>
<tr>
<td>NC_005085</td>
<td>NP_902612</td>
<td>hmsF</td>
</tr>
<tr>
<td>NC_005085</td>
<td>NP_902613</td>
<td>hmsH</td>
</tr>
<tr>
<td><em>Escherichia coli</em> K12</td>
<td>P69435</td>
<td>pgaA</td>
</tr>
<tr>
<td>NC_000913</td>
<td>P75906</td>
<td>pgaB</td>
</tr>
<tr>
<td>NC_000913</td>
<td>P75905</td>
<td>pgaC</td>
</tr>
<tr>
<td>NC_000913</td>
<td>P69433</td>
<td>pgaD</td>
</tr>
<tr>
<td><em>Staphylococcus epidermidis</em></td>
<td>AAZ78357</td>
<td>icaA</td>
</tr>
<tr>
<td>DQ149646</td>
<td>AAZ78358</td>
<td>icaD</td>
</tr>
<tr>
<td>DQ149646</td>
<td>AAZ78359</td>
<td>icaB</td>
</tr>
<tr>
<td>DQ149646</td>
<td>AAZ78360</td>
<td>icaC</td>
</tr>
<tr>
<td><em>Yersinia pestis</em> KIM6+</td>
<td>AAB66588</td>
<td>hmsH</td>
</tr>
<tr>
<td>YPU22837</td>
<td>AAB66589</td>
<td>hmsF</td>
</tr>
<tr>
<td>AAB66590</td>
<td>hmsR</td>
<td></td>
</tr>
<tr>
<td>AAB66591</td>
<td>hmsS</td>
<td></td>
</tr>
<tr>
<td><em>Bordetella pertussis</em></td>
<td>NP_880625</td>
<td>BP1941</td>
</tr>
<tr>
<td>NC_002929</td>
<td>NP_880626</td>
<td>hmsR</td>
</tr>
<tr>
<td>NC_002929</td>
<td>NP_880627</td>
<td>hmsF</td>
</tr>
<tr>
<td>NC_002929</td>
<td>NP_880628</td>
<td>hmsH</td>
</tr>
</tbody>
</table>

An initial approach among groups of proteins (Table 1) was performed using the Clustal W server (Jeanmougin et al., 1998) through multiple alignments in order to identify possible amino acids conserved among the proteins. The structural modeling was performed by 3D-JIGSAW (Bates et al., 2001) and PHYRE servers (Kelley et al., 2000) with the models with higher identity chosen as the template. The ProCheck program (Laskowski et al., 1996) was used to check the quality of the three-dimensional model generated. The quality of the model was analyzed through the Ramachandran diagram (Ramachandran et al., 1963). Analyses were carried out using the Verify3D program (Eisenberg et al., 1997) in order to ensure a more solid and comprehensive quality of the model.

**Results**

In a multiple alignment comparison generated by the Clustal W server with other glycosyltransferase proteins, the critical amino acids related to formation of biofilms are preserved also in the HmsR protein of *C. violaceum* (Figure 1). Conserved amino acids that are critical for function of HmsR of *Y. pestis* (D176, D269, Q305 and R308) correspond to residues D135, D228, Q264 and R267 in *C. violaceum*, respectively.

The structural model generated (Figure 2) comprises the critical amino acids for the function of this protein in the formation of biofilms arranged in a three-dimensional conformation with the possibility of fitting a substrate. The residue D228 is in exactly the same position as D191 (SpsA) with other conserved residues (D135, Q264 and R267) also positioned on the active site.

The multiple alignment among organisms in this study shows that the polypeptide HmsF of *C. violaceum* retains the amino acids that are crucial to the formation of extracellular matrix in biofilms (Figure 3). The D114 and D115 aspartate amino acids of the protein HmsF of *Y. pestis*, which have HmsF deacetylase activity functionality, are conserved on all related organisms in this study, and in *C. violaceum* correspond to aspartate D86 and D87. The important amino acids in the synthesis of extracellular matrix in *Y. pestis* (W143 and H184), which correspond in *C. violaceum* to W115 and H156, are also retained.

In the three-dimensional model of HmsF of *C. violaceum* obtained (Figure 4), the 3D-JIGSAW server made use of the protein SpPgdA as a template (PDB code: 2c1g) (Blair et al., 2005), which is a N-acetyl-glucosamine deacetylase. The model generated for HmsF of *C. violaceum* presenting the spatial location of critical residues in the biofilm formation to *Y. pestis* shows that the residues D86, D87, W115 and H156 are closely located in space (Figure 4). These residues are situated around the active site.

In the multiple alignment of HmsH protein, it was observed that the residue arginine R113, which had a poor performance in the formation of biofilms in *Y. pestis* (For-
man et al., 2006), and that in C. violaceum is the residue R104, was conserved in almost all organisms studied, but it was not observed in S. epidermidis, a Gram-positive bacterium (not shown). The multiple alignment performed among bacteria of this study for ORF CV2940 shows that the critical residue tryptophan W80 of Y. pestis is also conserved in Gram-negative bacteria E. coli and C. violaceum, corresponding to the residue W94 in C. violaceum.

The analysis of the three-dimensional model generated for HmsR of C. violaceum by the Ramachandran diagram shows that the model holds 99.1% of the residues in allowed regions, demonstrating a good quality. The analysis of the environment of each amino acid residue carried out by the Verify3D program presented two regions (residues 175 to 187 and 230 to 237) in the HmsR with a negative 3D-1D score, indicating a probable incorrect folding in this region. However, the critical residues for HmsR in the formation of biofilm do not appear in this region. The structure of HmsR of C. violaceum is shown in Figure 2, and the active site is indicated by the superimposed proteins closing up the active site of SpsA protein bound to Mn-UDP with the built model of HmsR protein C. violaceum. The residues of SpsA are green with HmsR residues in red and the UDP-Mn, Mg and Glycerol in yellow.

The critical residues aspartate (D), glutamine (Q) and arginine (R) are demarcated. The analysis of the three-dimensional model generated for HmsR of C. violaceum by the Ramachandran diagram shows that the model holds 99.1% of the residues in allowed regions, demonstrating a good quality. The analysis of the environment of each amino acid residue carried out by the Verify3D program presented two regions (residues 175 to 187 and 230 to 237) in the HmsR with a negative 3D-1D score, indicating a probable incorrect folding in this region. However, the critical residues for HmsR in the formation of biofilm do not appear in this region. The structure of HmsR of C. violaceum is shown in Figure 2, and the active site is indicated by the superimposed proteins closing up the active site of SpsA protein bound to Mn-UDP with the built model of HmsR protein C. violaceum. The residues of SpsA are green with HmsR residues in red and the UDP-Mn, Mg and Glycerol in yellow.

Y. pestis 121 IAINDGS8DD TAQUIDALLA EDPLRLVHL AHMPOSKAIA RMQAAEARSE YLVDGDBAL E. coli 108 IAVNDS8TDD TRAIDNDA QIPHRIRVHL AQMGPOSKAIL KTGAAKAASSE YLVDGDBAL C. violaceum 80 IANVDS8GRL TAALINQNOQ EMPHRIRVQ IAHGPOSKVGLW ATTALISLYR ELMCQGDBAL D. persius 83 IAINDS8SRD TQGALINHEL QVPFRVNLIV SRNQGPOSKN ITTAQLCLDS LRDLGQDBAL S. epidermidis 86 IIIINDS8GSD TQALINHEL QVPFRVNLIV SRNQGPOSKN ITTAQLCLDS LRDLGQDBAL

Y. pestis 181 LDKNAYVFLV ALFLAQNPTRG AVGMPOPRRT RSTLGQVROU GFPSSTIGLI KRTQRYGQV E. coli 168 LDRAAAYTV RFLNYPRVUG AVGMPOPRRT RSTLGQVROU GFPSSTIGLI KRTQRYGNV C. violaceum 140 LDPAAKWLM RHPGSGRQV AVGMPOPRRT RSTLGQVROU GFPSSTIGLI KRTQRYGQL B. persius 143 HFPSSTYML TPHLAQHNGU AVGMPOPRRT RSTLGQVROU GFPSSTIGLI KRTQRYGKL S. epidermidis 139 IIDDAPFPMT EEKHLNPRVQ AVGMPOPRRT RSTLGQVROU GFPSSTIGLI KRTQRYGKL

Y. pestis 241 PVTVGTVLAF RPRALAVGWW WSMDTNGH KDWQFPRG ALWMLFETLR E. coli 228 PVTVGTVLAF RPRALAVGWW WSMDTNGH KDWQFPRG ALWMLFETLR C. violaceum 200 PVTVGTVLAF RPRALAVGWW WSMDTNGH KDWQFPRG ALWMLFETLR B. persius 203 PSTGVTMTNF RPKALQHWW WSMDTNGH KDWQFPRG ALWMLFETLR S. epidermidis 199 NTGCTVPFLP KSALNVQRLG WSMDTNGH KDWQFPRG ALWMLFETLR

Y. pestis 301 GILMCRGKLA OGGAEYFVRK MFPLMNRRNN RRML-operativey REYSF DOLY4YLLGL E. coli 288 GIMCNGKLA OGGAEYFVRK MFPLMNRRNN RRML-operativey REYSF DOLY4YLLGL C. violaceum 260 GILMCRGKLA OGGAEYFVRK MFPLMNRRNN RRML-operativey REYSF DOLY4YLLGL B. persius 263 GILMCRGKLA OGGAEYFVRK MFPLMNRRNN RRML-operativey REYSF DOLY4YLLGL S. epidermidis 259 GILMCRGKLA OGGAEYFVRK MFPLMNRRNN RRML-operativey REYSF DOLY4YLLGL

Figure 1 - Partial multiple alignment of HmsR from C. violaceum compared with HmsR of Y. pestis; PgaC of E. coli; IcaA of S. epidermidis and HmsR of B. pertussis. The critical residues aspartate (D), glutamine (Q) and arginine (R) are demarcated.

Figure 2 - Superimposed proteins closing up the active site of SpsA protein bound to Mn-UDP with the built model of HmsR protein C. violaceum. The residues of SpsA are green with HmsR residues in red and the UDP-Mn, Mg and Glycerol in yellow.

Figure 3 - Multiple partial alignment of HmsF of C. violaceum compared with PgaB of E. coli; IcaB of S. epidermidis; HmsF of Y. pestis and HmsF of B. pertussis. The critical residue to extracellular matrix formation in Y. pestis: D (aspartate), W (tryptophan) and H (histidine) are demarcated.
The generated model with the spatial location of critical amino acids that are critical to the formation of biofilms. Therefore the conclusion is that the operon hmsHFR-CV2940 might be linked to C. violaceum pathogenicity.

References


**Guest Editor: José Carlos Merino Mombach**

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.