RESULTS & DISCUSSION
The work described in this chapter is initiated as a consequence of a novel observation made during heterologous expression of *opd* gene. When expression plasmids were constructed to express *opd* gene to code organophosphate hydrolase (OPH) with and without signal peptide, the amount of protein coded by the plasmid expressing mOPH (without signal peptide) was several folds high when compared to the expression levels obtained from plasmid coding for preOPH (OPH with signal peptide). Such unusual increase in expression level, despite of having identical vector backbone indicates existence of regulatory elements in the *opd* region specifying signal sequence (Fig. 3.1). Therefore in the present study a detailed investigation is made to understand the regulation of *opd* gene expression. Before describing experimental strategies for studying the regulation of *opd* gene expression, a thorough bioinformatic analysis was made to identify regulatory elements that contribute for expression of *opd* gene in *B. diminuta*. Using Bioinformatic tools developed by Munch et al., 2005 (www.prodoric.de/vfp) two putative promoter elements were identified 33bp upstream of start codon ATG. One of them has shown resemblance to the classical sigma 70 ($\sigma^{70}$) dependent promoters (Fig. 3.1). The second one has shown identity to $\sigma^{54}$ dependent promoter element. Though existence of these two promoter elements was predicted by earlier investigators their functional status was not examined till to date (Harper et al., 1988; Mulbry and Karns, 1989). Further, on careful examination we have also identified an inverted repeat sequence at 32 bp downstream of start codon ATG. This IR sequence is found within the signal peptide coding sequence of *opd* gene. Removal of signal peptide coding sequence has resulted in elevated expression levels of OPH. Therefore, the IR sequence is expected to play a role in downregulation of *opd* gene expression. This chapter describes the experiments designed to assess the functional status of the predicted regulatory
elements. Based on the inference drawn from the experiments a logical conclusion is made explaining the regulation of opd gene expression in B. diminuta.

**Fig. 3. 1**

Identification of putative promoter elements upstream of start codon ATG of opd gene. The putative $\sigma^{70}$ and $\sigma^{54}$ dependent promoters are highlighted with red and pink colors. The putative ribosomal binding site is shown with dotted lines. Important restriction sites used to generate opd-lacZ fusions are mapped. The inverted repeat sequence found 32 bp downstream of start codon ATG is shown with bold case.

### 3. 1. Prediction of putative promoter elements of opd gene

The sequence of opd gene was analyzed using motif search programme (www.prodoric.de/vfp). The search revealed existence of two putative promoter elements upstream of start codon ATG. The conserved hexamers of one of the promoters have shown strong similarity to the consensus $\sigma^{70}$ dependent promoter. To gain better understanding, the predicted promoter motif was aligned to the consensus $\sigma^{70}$, $\sigma^{32}$, $\sigma^{E}$ and $\sigma^{S}$ dependent promoters. As shown in Table. 3. 1 the putative promoter motif has shown more resemblance to the consensus $\sigma^{70}$ motif. The -35 hexameric sequence has shown more conservation than the hexameric sequence found at -10 region. Downstream of this promoter motif, existence of
dinucleotides GG and GC with a gap of 11 bp, points towards existence of putative \( \sigma^{54} \) dependent promoter element upstream of the start codon of \( opd \) gene. Further, to validate the functional status of these putative promoters, they were independently cloned upstream of the promoter less \( lacZ \) gene of the promoter probe vector pMP220 (Spaink et al., 1987). The \( opd \)-\( lacZ \) fusions thus constructed were mobilized in \( B. \ diminuta \) and promoter activity was assessed through \( \beta \)-galactosidase activity.

### Table. 3. 1

<table>
<thead>
<tr>
<th>( \sigma ) Factor</th>
<th>Promoter consensus sequence</th>
<th>Space between -35 and -10 / -24 and -12 regions</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B. \ diminuta )</td>
<td>( opd ) putative ( \sigma^{70} )</td>
<td>TTGACA TAAAAG</td>
<td>16</td>
</tr>
<tr>
<td>( \sigma^{70} )</td>
<td>TTGACA TATAAT</td>
<td>16 ± 17</td>
<td>Gruber and Gross, 2003.</td>
</tr>
<tr>
<td>( \sigma^{32} )</td>
<td>TTGACA TATAAT</td>
<td>17</td>
<td>Typas and Hengge, 2006.</td>
</tr>
<tr>
<td>( \sigma^{28} )</td>
<td>CCTGAA NCCCATNT</td>
<td>17</td>
<td>Wang and deHaseth, 2003.</td>
</tr>
<tr>
<td>( B. \ diminuta )</td>
<td>( opd ) -putative ( \sigma^{54} )</td>
<td>CGGC TGCA</td>
<td>11</td>
</tr>
<tr>
<td>( \sigma^{54} )</td>
<td>TGGC TGCA</td>
<td>10</td>
<td>Barrios et al., 1999.</td>
</tr>
</tbody>
</table>

Comparison of putative \( \sigma^{70} \) and \( \sigma^{54} \) dependent promoter sequences of \( opd \) gene of \( B. \ diminuta \) with consensus \( \sigma^{70} \), \( \sigma^{5} \), \( \sigma^{32} \), \( \sigma^{28} \), \( \sigma^{54} \) promoters of \( E. \ coli \).

### 3. 2. Construction of \( opd \)-LacZ fusions

In order to determine functional status of putative \( opd \) promoter element predicted upstream of the start codon of \( opd \) gene, a series of promoter \( lacZ \) fusions were generated by cloning upstream region of \( opd \) gene in a broad host range mobilizable vector pMP220 (Spaink et al., 1987). The \( opd \) gene cloned from \( B. \ diminuta \) was taken as a source plasmid and the
5’region of opd gene was taken as various restriction fragments and cloned in pMP220. A detailed cloning strategy used to generate opd-lacZ fusions is shown in Fig. 3.2.

Fig. 3.2

A diagrammatic representation showing generation of opd-lacZ fusions. Plasmids pSM9 and pSM8 contain either $\sigma^{70}$ or $\sigma^{54}$ dependent promoters. Both $\sigma^{70}$ and $\sigma^{54}$ dependent promoters are fused to lacZ gene in plasmid pSM8. The IR sequence along with $\sigma^{70}$ and $\sigma^{54}$ promoters are fused to lacZ gene in pSM10. In pSM11 an omega (Ω) fragment is inserted between putative $\sigma^{70}$ and $\sigma^{54}$ dependent promoters. Abbreviations B, P, S, and X represent recognition sequences for restriction endonucleases BamHI, BglII, PstI, SphI, and XbaI, respectively.

3.3. Mobilization of opd-LacZ fusions and assay of promoter activity in B. diminuta

The opd-LacZ fusions generated were transformed into E. coli S17-1 and were used as helper strains to mobilize them into B. diminuta by following conjugation protocols described in materials and methods section. E. coli S17-1 cells containing opd-LacZ fusions were used as
donors while *B. diminuta* served as recipient. The exconjugants were selected on LB plates containing polymixinB and tetracycline.

The promoter activity for various *opd-lacZ* fusions was determined by measuring $\beta$-galactosidase activity (Miller, 1972) and the values obtained are shown in Table. 3. 2. A significant $\beta$-galactosidase activity (2040 miller units) was found in *B. diminuta* cells containing $\sigma^{70}$ dependant promoter-*lacZ* fusion (pSM7) indicating the predicted $\sigma^{70}$ dependant promoter is functional and responsible for transcription of *opd* gene in *B. diminuta*. In an attempt to assess the functional status of predicted $\sigma^{54}$ dependent promoter, an *opd-lacZ* fusion was constructed (pSM9) by including only $\sigma^{54}$ dependent promoter (Fig. 3. 2). When *B. diminuta* cells containing pSM9 were used for monitoring LacZ activity, only a negligible amount of LacZ activity is seen (Table. 3. 2). Such reduction in $\beta$-galactosidase activity points towards existence of a sequence motif that has no functional significance. However in literature, the dependency of $\sigma^{54}$ on upstream activating sequences (UAS) is well documented (Su et al., 1990). The UAS is normally present several base pairs upstream of the promoter element. In plasmid pSM9, the *opd* region cloned upstream of *lacZ* gene does not represent much of the upstream sequence of *opd* gene. In this construct there is no scope for existence of UAS. In order to avoid such situation another *opd-lacZ* fusion (pSM8) was generated by including the entire upstream region of *opd* gene. Hence in plasmid pSM8, an unique *PstI* site is there in between $\sigma^{70}$ and $\sigma^{54}$ dependent promoters. This unique *PstI* site was exploited to introduce an omega (Ω) fragment separating these putative promoters. The omega fragment is designed by introducing transcription terminator sequences flanking to the kanamycin gene. Existence of such transcription terminator sequence blocks transcription for $\sigma^{70}$ dependent promoter and hence the LacZ activation will be seen only when transcription initiation is made from predicted $\sigma^{54}$ dependent promoter. Further this situation
does not deprive promoter-UAS interactions, if there exists one upstream of putative $\sigma^{54}$ dependent promoter. Surprisingly, in *B. diminuta* cells, as shown in Fig. 3.3, the *lacZ* fusions having both $\sigma^{70}$ and $\sigma^{54}$ dependent promoters (pSM8) and the construct generated by introducing an omega fragment between putative $\sigma^{70}$ and $\sigma^{54}$ promoters (pSM11) have shown comparable promoter activities. Though the activity levels were very low when compared to the activity levels obtained for a construct (pSM7) having only $\sigma^{70}$ dependent promoter (2040 miller units), the functional status of $\sigma^{54}$ cannot be ignored. The $\sigma^{54}$ dependent genes are induced in response to a number of physiological conditions. Unless suitable physiological conditions are created maximal promoter activity is not achieved from these promoters (Reitzer and Schneider., 2001).

Though such physiological condition is yet to be identified, usage of $\sigma^{54}$ dependent promoter for expression of *opd* gene in *B. diminuta* cannot be ruled out.

### Table. 3.2

<table>
<thead>
<tr>
<th>S.NO</th>
<th><em>B. diminuta</em> containing <em>opd-lacZ</em> fusion</th>
<th>$\beta$-galactosidase activity (Miller Units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td><em>B. diminuta</em></td>
<td>19.4</td>
</tr>
<tr>
<td>2.</td>
<td><em>B. diminuta</em> (pMP220)</td>
<td>16.6</td>
</tr>
<tr>
<td>3.</td>
<td><em>B. diminuta</em> (pSM7)</td>
<td>2040.0</td>
</tr>
<tr>
<td>4.</td>
<td><em>B. diminuta</em> (pSM8)</td>
<td>943.61</td>
</tr>
<tr>
<td>5.</td>
<td><em>B. diminuta</em> (pSM9)</td>
<td>30</td>
</tr>
<tr>
<td>6.</td>
<td><em>B. diminuta</em> (pSM10)</td>
<td>549.44</td>
</tr>
<tr>
<td>7.</td>
<td><em>B. diminuta</em> (pSM11)</td>
<td>514.44</td>
</tr>
</tbody>
</table>

$\beta$-galactosidase assay for *B. diminuta* cells containing *opd-lacZ* fusions.

### Fig. 3.3

$\beta$-galactosidase activity of *opd-lacZ* fusions. Bar 1 and 2 represent *B. diminuta* without and with pMP220 vector and serves as negative controls. Bar 3, 4, 5, 6 and 7 represent levels of $\beta$-galactosidase activity obtained for *B. diminuta* cells having *opd-lacZ* fusions, pSM7, pSM8, pSM9, pSM10 and pSM11 respectively.
3.4. Existence of IR sequences in the signal peptide coding region of \textit{opd} gene

The $\beta$-galactosidase activities obtained for various \textit{opd-lacZ} fusions pSM7, pSM10, besides indicating the functional status of putative promoter elements, are suggestive of existing regulatory sequences in the coding region of \textit{opd} gene. In the \textit{opd-lacZ} fusion (pSM10) constructed by including the downstream region (downstream of start codon ATG), especially the signal peptide coding region, there was a significant reduction in $\beta$-galactosidase activity (Table 3.2). Such reduction in $\beta$-galactosidase activity is indeed supportive of elevated levels of OPH expression from expression plasmids coding for mOPH (mature form of OPH), where the signal peptide coding region of \textit{opd} gene is omitted while construction of expression plasmid. Such observation prompted to examine for presence of regulatory sequence in this region. When the sequence found downstream of translational start site of \textit{opd} gene was examined, a 22bp long inverted repeat (IR) was found in the \textit{opd} region specifying signal peptide (Fig. 3.4). Existence of such inverted repeat is expected to form a strong secondary structure at its 5' end of \textit{opd} specific mRNA.

\textbf{Fig. 3.4}

Nucleotide sequence of upstream region of \textit{opd} gene showing the existence of IR sequences (indicated in blue colored closed boxes) located exactly 29 base pairs downstream of the initiation codon ATG.
Therefore, the *opd* mRNA sequence of *B. diminuta* was used (using mFOLD analysis program) to predict existence of a secondary structure (www.bioinfo.rpi.edu/applications/mfold). As expected, existence of a stable stem loop structure requiring a free energy (ΔG) of -23.3 kCal / mol for destabilization was predicted (Fig. 3. 5).

**Fig. 3. 5**

Secondary structure prediction and calculation of free energy requirement of *opd* mRNA using mFOLD programme (www.rna.ccbb.utexas.edu). (ΔG= -23.3 kCal/mol)

On perusal of literature, such stable stem loops, especially at the 5’ end of mRNA, are shown to modulate the expression levels of cognate proteins (Punginelli et al., 2004). If the identified IR sequence has any regulatory role, it should have shown conservation among *opd* sequences found in other prokaryotic organisms. In order to gain information on existence of such IR sequences and associated secondary structures, the *opd* mRNA sequences of *B. diminuta* (Serdar et al., 1982), *Flavobacterium* sp. ATCC27551 (Sethunathan and Yoshida, 1973), *Flavobacterium balustinum* (Somara and Siddavattam, 1995), *Pseudomonas* sp. (Chaudhary and Wheeler, 1988) *Agrobacterium tumefaciens* (Horne et al., 2002) and *Mycobacterium tuberculosis* (Fleischmann et al., 2004) were downloaded from the data base and analyzed using mFOLD programme. In some of them the IR sequence was seen at an identical positions. The predicted secondary structures and ΔG values have shown close resemblance with the secondary
structure and $\Delta G$ value of $B. \textit{diminuta}$ opd mRNA (Fig. 3. 6), suggesting that the structural conservation might be having functional significance. To gain further insights into the IR sequence mediated modulation of OPH expression, a series of expression plasmids were constructed either by destabilizing or eliminating the IR sequence. Expression pattern obtained from these expression plasmids are presented below.

**Fig. 3. 6**

**Panel A** alignment of opd mRNA sequences from closely related bacteria. The percent homology is shown against each molecule. **Panel B** indicates the prediction of secondary structures found in opd mRNA sequences cloned from taxonomically diverse group of bacteria. $B. \dim$: $B. \textit{diminuta}$; $F. \textit{sp.}$: $\textit{Flavobacterium}$ sp. ATCC27551; $F. \textit{bal.}$: $\textit{Flavobacterium}$ balustinum; $P. \textit{sp.}$: $\textit{Pliseomonas}$ sp.; $A. \textit{tum.}$: $\textit{Agrobacterium}$ tumifaciens; $M. \textit{tub.}$: $\textit{Mycobacterium}$ tuberculosis.

3. 5. Expression of OPH with and without IR sequence

To gain better insights on IR mediated modulation of OPH expression, the opd gene was truncated by performing PCR and its variants generated with and without IR sequence were cloned under the control of vector driven transcriptional and translational signals. One of the expression plasmid pHYS400, was generated by cloning the entire ORF of opd gene in pET23b as NdeI and XhoI fragment (Fig. 3.7). These sites were engineered through PCR mutagenesis
The XhoI site was generated by modifying the stop codon of opd gene. Hence cloning of opd gene as NdeI-XhoI fragment generates inframe fusions of opd sequence with the vector specific sequence coding for six consecutive histidine residues.

**Fig. 3. 7**  
A diagrammatic representation showing strategy used to construct expression plasmids coding opd gene with and without IR sequence in pET23b.

The second expression plasmid was generated by cloning the opd gene without IR sequence. While achieving this target an artificial start codon was generated downstream of the IR sequence through PCR mutagenesis and cloned in pET23b as NdeI and XhoI fragment. The
detailed cloning strategy is shown in Fig. 3. 7 and the plasmid having opd gene without IR sequence was designated as pHNS400. As OPH, coded by pHYS400 and pHNS400 contain his tag, they can be detected through western blots using anti-his antibodies.

3. 6. Assessment of expression levels coded by pHYS400 and pHNS400

The E. coli cells having expression plasmids were induced and the induced proteins were analyzed on SDS-PAGE. The protein gel, if carefully analyzed, indicates existence of a thick band in lanes where protein extracts prepared from E. coli cells containing pHNS400 are loaded (Fig. 3. 8 Panel C, Lane 2). The size of this protein corresponds to the mature form of OPH (mOPH). In Fig. 3. 8 Lane 3 of panel C, where protein extracts prepared from E. coli cells having pHYS400 are loaded, there is no such protein band. However a faint band corresponding to the precursor form of OPH (preOPH) is seen in this lane. Western blots were performed to prove if these additional protein bands really represent OPH. As shown in panel C, WB, these two bands gave positive signals when probed with anti-his antibodies. The densitometry performed for these two signals indicated about 2.5 fold increase in expression levels of OPH in cultures expressing OPH from the expression plasmid pHNS400. This is rather a clear indication to show that the opd region coding signal peptide contains a cis element that contributes for modulation of OPH expression. Though the predicted IR sequence is seen in this region, the above described experiments do not justify its role in expression of OPH. Therefore, further experiments are designed to know if IR sequence is really responsible for down regulation of opd gene expression.
Expression of OPH with and without IR sequence in E. coli BL21 cells. **Panel A** shows existence of IR sequences in the signal peptide coding sequence of opd gene. **Panel B** shows extent of opd gene cloned in expression plasmids pHYS400 and pHNS400. **Panel C** indicates SDS-PAGE showing expression of OPH coded by pHNS400 (lane 2) and pHYS400 (lane 3). The corresponding western blot (WB) indicates expression levels of OPH coded by pHYS400 and pHNS400. **Panel D** shows the densitometry of OPH expression.

3.7 Site directed mutagenesis and destabilization of secondary structure in opd mRNA

In order to generate opd' variant having no potential to form the secondary structure in the mRNA sequence, a PCR mutagenesis was performed. While performing the PCR mutagenesis, sense 5’GTG CTC AAG TCT GCATGCA GCT GCA GGA ACT CTA CTA GGTGGA CTG GCT GGG TG3’ and anti-sense 5’CAC CCA GCC AGT CCA CCT AGT AGA
GTT CCT GCA GCT GCA GCA GAC TTG AGC AC3’ primers were designed by incorporating changes only at the inverted repeat region. These changes were found exclusively at the third base of the codon so that the protein coded by \( opd \) variant \( opd' \) is identical to the OPH coded by the wild type \( opd \) gene. Further, a non-complementary base is used to change the third base of the codon, the transcript generated from \( opd' \) is expected to loose its ability to form a secondary structure. Therefore, the \( opd \) variant \( opd' \) generated in this manner is expected to code for an identical OPH without forming a secondary structure at the 5’ end. Plasmid pHYS400 was used as template to perform PCR directed mutagenesis. The strategy used for performing site directed mutagenesis is shown in Fig. 3. 9. The generated PCR product was purified and treated with \( DpnI \) to eliminate methylated parent plasmid used as template. The unmethylated mutated amplicon will be rescued from the \( DpnI \) digestion. The amplicon was then transformed in to \textit{E. coli} DH5α cells which repairs the nicks of plasmid and maintains its integrity inside the host system. The plasmid containing mutant \( opd \) gene was then isolated and used for confirmation of introduced mutations by sequencing analysis.

**Fig. 3. 9**

Schematic representation showing PCR mediated site directed mutagenesis performed to destabilize IR sequence induced secondary structure in OPH mRNA.
Chapter III

After confirming the mutations, the mRNA coded by opd variant opd’ was subjected to mFOLD analysis programme to obtain secondary structure and ΔG values. Interestingly, as shown in Fig. 3.10, the mRNA transcript of opd variant opd’ showed no strong secondary structures. Even, the free energy (ΔG) calculated for destabilization of the weak secondary structure was just -13.3 kcal/mol. The expression plasmid containing the opd variant opd’ was designated as pHIR400 and transformed in to E. coli BL21. The expression levels of OPH coded by pHYS400, pHNS400 and pHIR400 were compared both by running SDS-PAGE for induced cultures and by performing western blots. The results obtained are shown in Fig. 3.10. Here, Lanes 7 and 9 are of having lot of significance. These two lanes contain protein extracts prepared from E. coli cells containing pHNS400 (lane 7) and pHIR400 (lane 9). As stated in earlier sections, the pHNS400 contains opd gene without inverted repeat sequence, whereas the plasmid pHIR400 has opd variant opd’ where IR sequence is modified in such a way that the transcript made from it is no longer having potential to form stable secondary structure (Fig. 3.10 panel A). If expression levels coded by these plasmids are seen, the opd’ coded OPH (Fig. 3.10, panel B, Lane 9) is equal to the OPH coded by pHNS400, the plasmid generated by removing signal peptide. Such observation is certainly an indication to show that IR sequence has a negative influence on the expression of OPH. While making such a bold statement, the results need to be supported through quantification of the opd transcript coded by all these plasmids. Such attempt was made by performing Real-Time PCR.
Expression of \textit{opd}' variants in \textit{E. coli}. \textbf{Panel A} shows the prediction of secondary structure and $\Delta G$ values of mRNA transcript made from \textit{opd} variant \textit{opd}'. The nucleotide changes are shown in green colored circles. \textbf{Panel B} SDS-PAGE and western blot analysis indicating expression profile of OPH coding from \textit{opd} gene and its variants. Lane 1 represents Molecular weight marker; lane 2 represents protein extracts prepared from \textit{E. coli} BL21 cells; lanes 3, 4, 5 and 6 represent protein extracts prepared from uninduced cultures of \textit{E. coli} BL21 having pET23b, pHNS400, pHYS400 and pHIR400 and lanes 7 to 9 similar extracts prepared from induced cultures.

3. 8. Quantification of mRNA transcripts of \textit{opd} variants

The OPH expression plasmids have identical vector backbone. These are all expressed using vector specific transcriptional and translational signals. Such a situation rules out involvement of transcriptional level regulation mediated by IR sequence. However to further justify this claim, a Real Time PCR was performed to quantify \textit{opd} specific RNA from \textit{E. coli} cells expressing OPH using expression plasmids pHYS400 (\textit{opd} gene with IR sequence),
pHNS400 (opd gene without IR sequence) and pHIR400 (opd variant where IR sequence is destabilized). The quantity of opd specific mRNA as determined through RT-PCR is shown in Fig. 3.11. As shown in Fig. 3.11, the level of mRNA transcript of opd gene in all three cultures are almost identical and there exist no variation. Therefore, the observed elevation in levels of OPH expression in cultures having plasmids pHIR400 and pHNS400 is due to IR mediated translation modulation of opd mRNA.

Fig. 3.11

Real Time PCR analysis. Histogram showing cyclic threshold values obtained for three independent experiments using total RNA isolated from E. coli cells containing expression plasmids pHYS400, pHNS400 and pHIR400. Bar lanes 1, 2 and 3 represent quantification of mRNA transcripts formed from expression plasmids pHYS400, pHNS400 and pHIR400 respectively.

3.9. Discussion

The transcription of all eubacterial genes or operons is initiated by recognizing the promoter regions and its upstream regulatory elements by RNA polymerases (Morett and Segovia, 1993). The eubacterial RNA polymerase is a heterotetrameric enzyme comprising one
β, one β’ and two α subunits (Helmann and Chamberlin, 1988). The ‘core’ enzyme interacts with the initiation factor σ to form the transcriptionally active enzyme (Eσ). The bacterial σ factors are divided into two different families. All the sigma factors that are evolutionarily related to the *E. coli* housekeeping factor σ\(^{70}\) are kept in the first group. The sigma factors that show considerable similarity to the alternative sigma factor σ\(^{54}\) are grouped in the second family (Gross et al., 1992). The Eσ\(^{70}\) does not form stable closed-promoter complex, therefore transcription can be initiated spontaneously in the absence of activator proteins. In contrast, the Eσ\(^{54}\) forms physically detectable closed-promoter complexes and fails to initiate transcription spontaneously (Gralla, 1990). Hence the σ\(^{54}\) dependent RNA polymerase is completely dependent on additional transcriptional factors such as enhancer binding proteins (EBPs) to initiate the formation of mRNA transcript (Morett and Segovia, 1993). The σ\(^{54}\) controls several auxillary processes in bacterial systems, which include degradation of xylene and toluene, transport of dicarboxylic acids, pilin synthesis, nitrogen fixation, hydrogen uptake (Fischer, 1994; Kustu et al., 1989), flagellar assembly (Arora et al., 1997), arginine catabolism (Gardan et al., 1995), alginate production (Zielinski et al., 1992), rhamnolipid production (Pearson et al., 1997), acetoin catabolism (Priefert et al., 1992), mannose uptake (Martin-Verstraete, 1995) and proline iminopeptidase activity (Albertson and Koomey, 1993).

The basic promoters recognized by the σ\(^{70}\) family are normally organized around two hexamers centered between -10 and -35 nucleotides upstream from the transcriptional start site though little diverse in their recognition sequence (Hawley and McClure, 1983). The initiation complex is further stabilized by the C-terminal domain of the core enzyme (aCTD), which can either interact directly with upstream DNA or with regulatory proteins (Benoff et al., 2002) and
its been reported that recognition of the -10 alone can also be sufficient for initiation to occur (Keilty and Rosenberg, 1987; Barne et al., 1997; Kumar et al., 1993).

Upstream sequence of *opd* gene has been used to identify promoter and other regulatory motifs using Bioinformatic tools. The results obtained in this study coincided with the predictions made by earlier investigators (Harper et al., 1988; Mulbry and Karns, 1989). The study conducted gathered fairly good evidence to show that a dual promoter system is involved in expression of *opd* gene. One of the promoters show highest similarity to the consensus $\sigma^{70}$ dependent promoter. Almost identical -35 hexameric sequence and optimal spacing between these conserved sequences (Hawley McClure, 1983) are some of the important features to believe that the predicted $\sigma^{70}$ dependent promoter is responsible for transcriptional activation of *opd* gene in *B. diminuta*. Upstream of *opd* gene existence of conserved dinucleotides, GG and GC with a gap of 11 nucleotides is the only reason to believe on existence of second $\sigma^{54}$ dependent promoter. In order to establish the functional status of these promoters, they were independently fused to the lacZ gene of promoter probe vector pMP220 (Spaink et al., 1987). This promoter test vector containing *opd-lacZ* fusions were subsequently mobilized into *B. diminuta* using *E. coli* S17-1 as helper strain. Significant $\beta$-galactosidase activity levels (2040 miller units) were noticed when consensus $\sigma^{70}$ dependent promoter is fused to the lacZ gene. This clearly indicates that the consensus $\sigma^{70}$ dependent promoter is responsible for transcriptional activation of *opd* gene in *B. diminuta*.

The lacZ fusions generated by including both $\sigma^{70}$ and $\sigma^{54}$ dependent promoters gave considerable $\beta$-galactosidase activity (943.61). However, the lacZ fusion (pSM9) obtained by fusing only putative $\sigma^{54}$ promoter motif failed to activate lacZ gene of promoter test vector.
Before concluding that $\sigma^{54}$ promoter motif non-functional, available literature on $\sigma^{54}$ dependent promoters was reviewed. In the $\sigma^{54}$ class of promoters the two conserved dinucleotides GG and GC at –12 and –24 regions are separated with a gap of 10 nucleotides (Bordes et al., 2004). The $\sigma^{54}$ has three functional domains, the carboxyl terminus is required for the binding of promoter DNA, the amino-terminal region is required for activation and the domain between these two regions is for binding core RNA polymerase (Sasse-Dwight and Gralla 1990). It has unique ability to bind promoter without the help of core polymerase. Upon binding to the promoter, recruits core polymerase and mediates interaction with transcription activator bound to UAS. This molecular events bring conformational changes in the promoter-DNA complex and facilitate opening of promoter to generate open promoter complex (Gralla, 1991; Wang et al., 1992; North et al., 1993). Yin and his co-workers have extensively studied regulation of $\sigma^{54}$ dependent glnAP2 promoter (Yin et al., 1995). The authors have demonstrated molecular mechanism involved in opening and closing of transcription complex formation by RNAP-$\sigma^{54}$ and showed that when sufficient nitrogen is present, $\sigma^{54}$-holoenzyme forms a closed complex that occupies the glnAp2 promoter in an inactive state (Sasse-Dwight and Gralla 1988). When nitrogen becomes insufficient, a cascade of reactions occur leading to the phosphorylation of enhancer-binding protein NtrC, which then binds to the upstream activating sequence (UAS) and activates transcription (Keener et al., 1988). The activation event occurs via DNA looping (Su et al. 1990) and involves the use of ATP to convert the closed complex to an open complex, which is active for transcription (Sasse-Dwight and Gralla 1988).

Open complex formation by RNAP-$\sigma^{54}$ requires ATP hydrolysis by activator proteins, which bind to enhancer like sequences located upstream of the promoter site (Sasse-Dwight and Gralla, 1988). The $\sigma^{54}$ activators are therefore also referred to as bacterial enhancer binding
proteins (bEBP). This is functionally analogous to enhancer dependent initiation of eukaryotic RNA polymerase II, which requires an input of energy from ATP hydrolysis provided by TFIIH (transcription factor IIH) (Lin et al., 2005). ATP hydrolysis by bEBPs provides energy for remodeling the $\sigma^{54}$–RNAP closed complex, resulting in further DNA melting and loading of the template strand of DNA into the RNAP active site (Cannon et al., 2003). Energy is transferred to the closed complex through a physical interaction between $\sigma^{54}$ and the AAA+ domain of the bEBP (Bordes et al., 2004; Lin et al., 2005). The process of closed to open promoter complexes is a multistep process (Davis et al., 2005). The bEBPs (bacterial enhancer-binding proteins) are AAA+ (ATPase associated with various cellular activities) family members involved in regulating bacterial gene expression (Ishihama, 2000; Browning and Busby, 2004). The bEBPs are essential for transcription initiation using the major variant $\sigma$ factor, $\sigma^{54}$, commonly involved in tightly regulated bacterial responses to stress (Buck et al., 2000; Bose et al., 2008).

On careful examination, in the putative $\sigma^{54}$ dependent promoter of opd gene, the conserved dinucleotides (GG and GC at –46 and –33 regions) are separated with a gap of 11 bp. A very weak promoter activity, obtained for lacZ fusion (pSM9) containing only $\sigma^{54}$ dependent promoter alone does not justify the putative $\sigma^{54}$ dependent promoter predicted upstream of opd gene is functional. As discussed in earlier sections, almost all consensus $\sigma^{54}$ dependent promoters contain upstream activating sequences (UAS) (Buck, et al. 1986) which participate in converting closed promoter complex into open promoter complex (Buck et al., 1986). Deletion of UAS from $\sigma^{54}$ dependent promoters significantly lowers the transcriptional activity (Reitzer and Magasanik, 1986).
The lacZ fusion pSM9, as shown in Fig. 3.2, is constructed by taking just basic promoter element. It does not contain much of the upstream region of opd gene which probably contain UAS. Therefore, the weak β-galactosidase activity cannot be taken as basis to dismiss the claim on existence of a σ\(^{54}\) dependent promoter upstream of opd gene. The UAS is found in all most all reported cases only upstream of the promoter motifs. They are located anywhere from 300 bp to 1 kb away from the promoter motif (Buck et al., 1986). In order to identify if such UAS like motif exists upstream of opd promoter, a lacZ fusion was constructed by cloning the entire upstream region of opd gene. Such lcaZ fusion (pSM11) includes both putative σ\(^{70}\) and σ\(^{54}\) dependent promoters. It is therefore difficult to predict actual promoter responsible for transcription activation of opd gene. Insertion of omega fragment between these two promoters terminates transcription initiated from upstreamly located σ\(^{70}\) dependent promoter and hence the lacZ activation as monitored by β-galactosidase activity will be exclusively from σ\(^{54}\) dependent promoter. The obtained β-galactosidase activity for this construct (514.44) suggests that σ\(^{54}\) dependent promoter is functional in B. diminuta. Such activity is not seen in the construct that was generated by including only σ\(^{54}\) dependent promoter (pSM9). This may be due to non availability of UAS in this construct. The omega fragment present in pSM11 only prevents transcription from σ\(^{70}\) dependent promoter and does not prevent possible interactions between σ\(^{54}\) dependent promoter and putative UAS by inducing a loop in the DNA region available between σ\(^{54}\) dependent promoter and UAS. A schematic model showing σ\(^{54}\) dependent transcriptional activation of opd gene is shown in fig. 3.12.
A schematic representation showing proposed model of $\sigma^{54}$ dependent transcriptional activation of *opd* gene.
The $\sigma^{54}$ dependent promoters are inducible promoter. They respond to a number of physiological conditions including that of nitrogen limiting condition. A phosphorylated transcription activator after bending to UAS influence the transcription activation of $\sigma^{54}$ dependent promoters. Phosphorylation of transcription activator protein is only done under defined physiological conditions. In the absence of phosphorylation transcription activator protein becomes inactive and only basal level of expression will be seen from $\sigma^{54}$ dependent promoter. In the present study, we have grown $B.\ diminuta$ cells in rich medium. There is no stress for the cells to show elevated $lacZ$ activation from $\sigma^{54}$ dependent promoter. In fact it is not really known under which circumstances the $opd$ gene is induced. As this study is mainly focused on membrane targeting of OPH no effort is made to identify a physiological situation that contributes for the induction of $opd$ gene in $B.\ diminuta$.

Existence of dual promoters upstream of genes or operons is not uncommon in prokaryotic world (Keener and Kustu, 1988). A classical example to cite for existence of both $\sigma^{70}$ and $\sigma^{54}$ dependent promoters is $ntr$ operon (Keener and Kustu, 1988). In $ntr$ regulon, the $\sigma^{70}$ is responsible for basal level expression, whereas $\sigma^{54}$ dependent promoter is induced in response to nitrogen limiting conditions (Hirschman et al., 1985). If $\beta$-galactosidase activities obtained from pSM7 and pSM11 are examined, it is a clear indication to suggest that both $\sigma^{70}$ and $\sigma^{54}$ dependent promoters found upstream of $opd$ gene are functional. A dual promoter system appears to be controlling the expression of $opd$ gene in $B.\ diminuta$.

As stated in the earlier sections of this chapter, the study was prompted by an unusual observation made during the heterologous expression of $opd$ gene. The construct made to express mature form of OPH (mOPH) produced 2.5 fold more protein when compared to the construct
generated for coding precursor form of OPH (preOPH). If this result is seen together with β-galactosidase activity levels obtained for opd-lacZ fusions pSM8 and pSM10 (Fig. 3.3), it clearly indicates existence of a ‘cis’ element modulating the expression of opd gene. Inverted repeats have been shown to play crucial roles in various cellular activities. Their involvement in regulation of DNA replication, transcriptional regulation, as well as translational control of gene expression is well documented (Scortecci et al. 1999). In fact, the secondary structures of leader sequences have also been identified as both transcriptional and translational down-regulators of gene expression (Wood et al. 1996; Curie and McCormick 1997; Hemmings-Mieszczak et al. 1998; Scortecci et al. 1999). Initially Real Time PCR was done to assess if the IR sequence is acting as transcription terminator. Though the opd variants produced variable amounts of OPH they all have identical vector back bone and are controlled by identical transcriptional and translational signals. One of the possible reasons for pHYS00 to produce less amount of OPH, is due to termination of transcript at IR sequence, initiated from the T7 promoter of the vector. Such termination of transcript should lead to variation in production of opd specific mRNA from these expression plasmids. The reduced transcription level of mRNA might be contributing for the low level of OPH synthesis in cells having pHYS400. In order to gain better insights on this hypothesis, mRNA produced from these plasmids pHYS400, pHNS400 and pHIR400 were quantified by Real Time PCR. As shown in Fig. 3.11 there was no difference in the concentration of opd specific mRNA made by these three constructs. In the event of finding identical opd specific mRNA concentration, for the reduced OPH concentration in cells having pHYS400, the event of transcription is not responsible. Therefore IR mediated regulatory event may be associated with translation process.
Predictions of mRNA secondary structure mediated modulation of gene expression is very common both in prokaryotic and eukaryotic world. Instead of taking very unsimilar situation for discussion, here mRNA secondary structure translation modulation of Tat-dependent Formate Dehydrogenase N (FDH-N) is taken for drawing comparison with the regulation of OPH. Like OPH, FDH-N is a metalloenzyme. It is associated with membrane (Punginelli et al., 2004). If *E. coli* cells are grown under anaerobic condition, FDH-N represents about 10% of membrane protein. The FDH-N, like OPH, is a Tat substrate and depended on Twin arginine transport machinery for membrane targeting. It is also associated to the inner membrane facing periplasm (Punginelli et al., 2004). The FDH-N is coded by an operon designated as *fdnGHI*. The high-resolution crystal structure of FDH-N has shown that it adopts an (αβγ)_3 “trimer-of-trimers” architecture, with the active site of the enzyme facing periplasm (Jormakka et al., 2002). Out of these three polypeptides only FdnG has signal peptide, which has twin arginine motif. The rest of the polypeptides do not contain any signal peptides. The FdnG leads them to the membrane in a prefolded form (Stanley et al., 2002). While experiments were conducted to show that FDH-N is a Tat substrate, mutations were generated by changing one of the invariant arginines (R) of Tat motif to lysine. This mutation has contributed for dramatic increase in FdnG synthesis (Punginelli et al., 2004). Further, it also contributed for the increase of partner subunits synthesis contributing for elevated synthesis of FDH-N (Punginelli et al., 2004). Punginelli and associates have done a number of experiments to show that *fdnG* mRNA has a secondary structure which required a free energy ΔG of -12.6 kcal/mol (Fig. 3. 13). Further they have shown that the secondary structure has modulated the expression of FDH-N in *E. coli*. The authors also suggested that this modulation facilitates interaction of partner subunits promoting the formation of a functional complex to be transported in a prefolded form.
Like FDH-N, OPH is a membrane protein and targets membrane via Tat pathway. It is not yet known if it is part of the multi subunit complex. As described earlier four types of proteins choose Tat route for membrane targeting. OPH does not require large cofactors and therefore it does not belong to those group of proteins that take Tat route along with large cofactors. However, having known that it interacts with Orf306, a product of orf306 gene located adjacent to the opd gene of transposon-like opd gene cluster (unpublished results), its existence as multishubunit complex in the membrane of B. diminuta cannot be ruled out. If that is true, the mRNA secondary structure mediated modulation of OPH expression gain lot of significance.
Conclusions

- A well defined Tat motif that shows high similarity to the consensus Tat motif involved in transport / membrane targeting of prefolded proteins in prokaryotes is discovered in *B. diminuta*.

- The 29 amino acid long signal peptide is not required for folding of OPH to gain active conformation.

- The invariant arginines found in Tat motif are required for membrane targeting of OPH in *B. diminuta*.

- The Tat motif inserts OPH at the periplasmic face of inner membrane.

- Upstream of *opd* gene two putative promoter elements were identified. One of them showed high similarity to the consensus \( \sigma^{70} \) dependent promoter. The second promoter element showing similarity to the consensus \( \sigma^{54} \) dependent promoter activated promoter less *lacZ* gene, indicating the possibility of existing a second promoter upstream of *opd* gene.

- An inverted repeat sequence found in the signal peptide coding region of *opd* gene is probably responsible for down regulation of OPH in *E. coli*. 


Research Publications:


