CHAPTER III

Identification of the Integration Region of the Phage PIS136
3.1 INTRODUCTION

Bacteriophages depend on the host cell machinery for their survival and propagation. Lytic phages, after infection of the host cell, utilize the DNA replication enzymes of the host to synthesize multiple copies of its DNA. The host RNA Polymerase and ribosomes are used for the production of proteins required to form the coat of the virus as well as those that package the DNA into the phages. This leads not only to an explosion in the number of the progeny viruses but also to the death of the host bacterium. The phages finally lyse the cell and burst out.

Temperate phages on the other hand can recombine their genomes with the host genome and replicate as a part of the host genetic material. This mode is called as the lysogenic cycle and does not cause any harm to the host bacterium. Temperate phages also go through the lytic cycle for the production of a large number of progeny phages.

Phages use two distinct modes of recombination to insert their genomes into a host. Most of the retroviruses, phage Mu (E. coli, Chaconas et al., 1985); B3 (Pseudomonas aeruginosa, Braid et al., 2004) and Hin-Mu (Haemophilus influenzae, Saariaho et al., 2005) use transpositional recombination for insertion into the host genome. Transposition of the DNA sequences is mediated by a protein, transposase without the requirement of any host factors (Lampe et al., 1996; Vos et al., 1996). Transposase, in most instances, is the only requirement therefore the control of the process is often limited to the regulation and expression of transposase.

Transpositional recombination does not require any sequence homology between the two recombination partners and usually random with a very little preference in the target DNA.
sequence. Random transposition can thus cause mutations in the genome of the host and can sometimes be harmful.

The other mode of recombination used by most of the bacteriophages for insertion into the host genome is site-specific recombination (Frumerie et al., 2005). This recombination relies on sequences with a short stretches of homology and is thus limited in the targets of insertion. The process is catalyzed by two distinct families of recombinases which share an overall similarity in the process but differ in the mechanism of the reaction (Osborn and Boltner, 2002; Smith and Thorpe, 2002). Recombination occurs by the formation of a covalent DNA-Protein intermediate and does not cause any net synthesis or loss of DNA (Lee and Landy, 2004). Because of this the reaction is independent of any energy requirement earning the name conservative site-specific recombination.

The recombinases can be grouped as Serine and Tyrosine recombinases based on the nucleophile involved (Stark et al., 1992; Hallet and Sherratt, 1997). Phages have been found to use proteins belonging to either family for integration in the host genome. The targets of integration by site-specific recombination are very limited due to the requirement for sequence homology (Williams, 2002). In most cases the host bacterium does not have more than one site of phage integration though some secondary sites can also be used albeit with lesser efficiency. Site-specific recombination thus does not cause multiple insertions that lead to mutations in the host genome and thus differs from transpositional recombination.

The presence of a single preferred site of integration makes the identification of the target for integration in the host relatively easy. Recombination leads to the disruption of the parent sequences in both phage and host genome. Integration of the phage DNA into the host genome causes the formation of recombinant sequences that form the junction between the host and the phage (Landy, 1989). These recombinant sequences form the substrates for the reverse reaction known as excision that releases the phage DNA from the host genome (Frumerie et al., 2005; Warren et al., 2005). Thus the objective is to identify the region of phage PIS136 responsible for the integration into the host genome.
3.2 METHODS

3.2.1 Preparation of genomic DNA from *Saccharomonospora* sp. PA136 which is a lysogen for the phage PIS136

The natural lysogen of the phage PIS136, *Saccharomonospora* sp. PA136, was grown in STP medium for 6 days at 30 °C at 160 rpm. The cells were collected by centrifugation at 10,000 rpm for 15min at 4 °C in a Sorvall SLA-1500 rotor. Genomic DNA was prepared from the cells according to the method of Kieser *et al.* (2000). The cells were washed with 15ml of Streptomyces lysis buffer without lysozyme and collected by centrifugation as above. Finally, the cells (~50mg) were resuspended in a 1ml of *Streptomyces* lysis buffer with lysozyme (2mg/ml) and RNaseA (100μg) and were incubated at 37 °C in a water bath, with intermittent gentle agitation, for 30min or until the lysis was complete which was indicated by increase in viscosity of the solution.

To the cell lysate, 0.1 vol. 3M Sodium acetate pH 5.5, 50μg Proteinase K and SDS to a final concentration of 1% were added. This mixture was incubated at 53 °C for 2 hours to degrade the proteins and then shifted to 60 °C for 30min. After cooling to room temperature, the mixture was extracted twice with phenol: chloroform and then twice with chloroform. To the aqueous layer, 2 vol. of absolute ethanol (-20 °C) was added along the walls of the tube and mixed gently. This was incubated on ice with intermittent mixing. The DNA was precipitated at the interface and then spooled out into a fresh tube.

The precipitated DNA was washed with 70% ethanol, dried at 37 °C and then resuspended in 1 ml TE buffer. The DNA was extracted twice with phenol: chloroform and twice with chloroform before precipitating as above. The DNA was finally resuspended in 500 μl TE buffer and the concentration of DNA was estimated as given in materials and methods.

3.2.2 Identification of the integration region of the phage

To identify the region of the phage DNA that might be responsible for the integration of the phage, DNA from the phage and its natural lysogen, *Saccharomonospora* sp. PA136 was digested with the enzymes *Clal*, *PvuII*, *SnaBI+ PvuII*, *NheI*+ *NcoI*, *NcoI* and *SpeI*+ *NcoI*. The digests were resolved on a 0.7% agarose gel in 0.5X TBE buffer for 48 hours and then transferred onto a Hybond-N nylon membrane (Amersham Biosciences,
USA). The blot was hybridized with α-\(^{32}\)P dCTP labeled whole phage DNA as a probe at 55 °C in the presence of 40% formamide. The blot was washed twice at a high stringency using 0.1X SSC and 0.1% SDS at 65 °C.

**3.2.3.1 Identification of the junction fragments**

To identify and sequence the target of phage integration in the host, the genomic DNA from *Saccharomonospora* sp. PA136 and phage PIS136 was digested with restriction enzymes, *NheI*+ *NcoI*, *EcoRI*+ *NcoI*, *KpnI*+ *EcoRI*, *ClaI*, *KpnI*+ *ClaI*, *KpnI*, *KpnI*+ *NcoI* and *KpnI*+ *NheI*. The digests were resolved on a 0.7% agarose gel. The DNA was transferred onto Hybond-N membrane and then hybridized with the C5 fragment (a ~4.2Kb fragment of *ClaI* digest that may carry the gene responsible for integration) as a probe at 50 °C in presence of 35% formamide. The blot was washed with 2X SSC+ 0.1% SDS at 50 °C for 10min, then with 0.1X SSC+ 0.1% SDS at 37 °C and then at 34 °C for 10min each.

**3.2.3.2 Cloning and sequencing of the junction fragments**

The fragments of the host genomic DNA identified by hybridization with the C5 fragment were gel purified and end-filled and then cloned either into the *EcoRV* site of plasmid pGEM-5fZ(+) or in the *SmaI* site of pUC19. The ligated plasmids were transformed into *E. coli* DH5α and then selected on LB plates containing Ampicillin, X-gal and IPTG. The white clones were patched onto fresh LB Amp plates and the clones with the correct insert were identified by colony hybridization using C5 fragment as a probe as described in section 2.2.4.

Plasmid was prepared from the hybridized clones by alkaline lysis method and then digested with *NcoI*+ *SpeI* to check the size and integrity of the insert. Whenever necessary, the digests were transferred onto nylon membrane and then hybridized with C5 probe to confirm the presence of the correct insert as described in section 2.2.3. Two independent clones were transformed into *E.coli* DH5α and the plasmid was purified using the ABI PRISM™ Miniprep Kit (Perkin Elmer, USA). Sequencing of the insert was done with M13 forward and reverse universal primers using ABI PRISM™ Big Dye Terminator Cycle Sequencing Kit on ABI310 sequencer (Perkin Elmer, USA).
3.2.4 Sequencing the C5 fragment

In order to sequence the C5 fragment, multiple overlapping sub-fragments were identified by hybridization with C5. Phage DNA was digested with the enzymes BamHI, SacI, SalI, SmaI, MluI, KpnI+ PvuII, EcoRI+ PvuII and BamHI+ PvuII. The digests were resolved on a 0.7% agarose gel and the DNA was transferred onto a Hybond-N Nylon membrane. The DNA on the membrane was hybridized with the C5 fragment as in section 2.2.3.

The fragments obtained were end-filled and cloned in the EcoRV site of pGEM-5fZ(+). All the ligated plasmids were transformed into E.coli DH5α and selected on LB agar with Ampicillin, IPTG and X-gal. Clones having the correct insert were identified according to the method described in section 3.2.3.2. The fragments were also cloned into the vector pGEM-5fZ(-) to sequence both the strands as described earlier.

In order to fill the gaps, further sub-clones were generated from the above mentioned fragments. Wherever needed, the enzymes SalI, MluI, SacI, EagI, AatII, HindIII, ScaI and NarI were used to digest the clones. All the plasmids were identified and sequenced as described earlier.

While sequencing the plasmid pGS16, the sequence in the reaction was abruptly terminated at a particular position. It appeared that the polymerase used in the ABI PRISM™ Big Dye Terminator Cycle Sequencing method was falling off at that position. In order to sequence this region from the other end, the plasmid pGS16 was digested with MluI to get a fragment of ~330bps. This fragment was cloned after end-filling to get the plasmid pGS16M. Attempts made to sequence the insert including addition of DMSO in the reaction, did not yield any sequence.

3.2.5 Sequencing using a thermo stable polymerase

In order to have an insight into the sequence structure of pGS16M, the insert was PCR amplified using M13 forward and reverse primers. PCR resulted into a fragment of the expected size of ~330bp suggesting that at a higher temperature, the sequence structure can be melted.

A Thermosequenase Cycle Sequencing Kit (USB, Ohio, USA #US78500) uses a dideoxy chain termination method at a high extension temperature. Therefore, the kit was used to sequence pGS16M, with minor modifications to the manufacturer’s protocol.
Identification of Integrase

The method consists of an initial primer extension step which incorporates the radiolabel in the primer, followed by a termination step which generates the dideoxy terminated sequences. In the second step, the sequencing primers are labeled to a high specific radioactivity by limited extension of the primer and was done in the presence of two cold dNTPs and a radiolabeled dNTP.

The reaction was carried out with 0.5 pmol of M13 universal forward primer, 100ng of template plasmid pGC5Sma16-Mlu, 1μl each of 7-deaza-dGTP cycle mix and dTTP mix. 10μCi of α-32P dATP was added as the radiolabel and the primer was labeled by partial extension with Thermo Sequenase DNA Polymerase in a thermal cycler (Biometra, Germany) with the following cycling parameters:

95 °C, 15sec

60 °C, 30sec for 50 cycles.

When the labeling step was nearing completion, four tubes were prepared and 4μl of the appropriate dideoxy termination mix was added to each tube (ddGTP in G tube, and so on). After completion of the labeling reaction, 3.5μl of the labeling reaction mix was added to each tube containing the dideoxy termination mix. The reaction components were mixed well and the termination step was carried out using the following cycling parameters:

95 °C, 30sec

72 °C, 120sec for 50 cycles.

The reaction was stopped by adding 4μl of the stop solution. The samples were heated to 72 °C for 2min and resolved on a 6% (19:1), 1X TTE, 7M urea and 40% formamide polyacrylamide gel in 1X TTE at 45W. In order to maintain the denatured condition of DNA, the gel was pre-run for 30min to attain a gel temperature of 55 °C.

After the run, the gel was soaked in 5% Acetic Acid: 20% Methanol solution for 15min. The gel was lifted onto a Whatman 3MM paper, covered with cling film and dried at 80 °C for 2 hours in a Bio-Rad gel dryer under vacuum. The dried gel was exposed to Kodak X-Ray film for 34 hours without an intensifying screen at room temperature.
3.2.6 Sequencing of the regions flanking C5 fragment

In an effort to get the gene organization around the region of C5, attempts were made to sequence the regions flanking the C5. The C-terminal region of the assembled sequence was found to have a Clai site within it. The sequence of the fragment beyond the Clai site matched with another Clai fragment that was already sequenced (C6 fragment; Soni, V., 2001). Thus, the region downstream to the C5 fragment was already sequenced.

To identify the region present upstream to the C5 fragment, the phage DNA was digested with NarI, SalI+ BamHI, SmaI+ SacI, SmaI+ NarI, SmaI+ SalI, SalI+ SacI, SalI+ NarI and NarI+ SacI. The digests were resolved on a 0.7% agarose gel and then transferred onto a Hybond-N nylon membrane. The DNA was hybridized with C5 fragment as a probe as described earlier.

3.2.7.1 Assembly and analysis of the C5 fragment

The sequences obtained were assembled by using the programs ClustalW, Gene Runner and by manually looking for overlapping sequences among the various sub-clones.

The assembled sequence was analyzed for putative ORFs using the web based programs; ORF finder (www.ncbi.nlm.nih.gov/gorf/), Frameplot 2.3.2 (Ishikawa and Hotta, 1999) and GeneMark (Lukashin and Borodovsky, 1998). The detected putative ORFs were then checked for the presence of conserved regions and putative functional proteins by aligning with non-redundant protein database using the program BLASTP (Altschul et al., 1997) and by aligning with the phage database at FASTA (fasta@ebi.ac.uk). Motifs and domains in the proteins were detected using PROSITE (Hulo et al., 2004) and ProDom servers (Servant et al., 2002) and by searches against the Pfam database (Bateman et al., 2004).

3.2.7.2 Multiple sequence alignment and phylogenetic relationship of Integrase

Multiple sequence alignment was constructed using a combination of methods. Fourteen sequences were identified by BLAST and Psi-BLAST searches using the translated protein sequence of ORF1 in C5. For each sequence, structural homology to all proteins in the PDB database (Sussman et al., 1998) was evaluated with the mGenTHREADER fold-recognition algorithm (Jones, 1999). All the searches, except Mx8
<table>
<thead>
<tr>
<th>Organism</th>
<th>GeneBank id</th>
<th>Position of Starting aa</th>
<th>la0p Score</th>
<th>E-value</th>
<th>Score</th>
<th>E-value</th>
<th>Score</th>
<th>E-value</th>
<th>Score</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIS136</td>
<td>Saccharomonospora sp.</td>
<td>AF459434</td>
<td>161</td>
<td>0.878</td>
<td>2e-04</td>
<td>0.949</td>
<td>4e-05</td>
<td>0.934</td>
<td>5e-05</td>
<td>0.880</td>
</tr>
<tr>
<td>Lambda</td>
<td>Escherichia coli</td>
<td>P03700</td>
<td>175</td>
<td>0.958</td>
<td>3e-05</td>
<td>0.961</td>
<td>3e-05</td>
<td>0.960</td>
<td>3e-05</td>
<td>0.971</td>
</tr>
<tr>
<td>Ms6</td>
<td>Mycobacterium sp.</td>
<td>AAD03774</td>
<td>177</td>
<td>0.805</td>
<td>8e-04</td>
<td>0.803</td>
<td>8e-04</td>
<td>0.950</td>
<td>4e-05</td>
<td>0.896</td>
</tr>
<tr>
<td>Mv4</td>
<td>Lactobacillus delbrueckii</td>
<td>AAG31327</td>
<td>229</td>
<td>0.962</td>
<td>3e-05</td>
<td>0.954</td>
<td>3e-05</td>
<td>0.963</td>
<td>3e-05</td>
<td>0.941</td>
</tr>
<tr>
<td>Mx8</td>
<td>Myxococcus xanthus</td>
<td>BAA13089</td>
<td>171</td>
<td>0.584</td>
<td>0.081</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.529</td>
</tr>
<tr>
<td>L5</td>
<td>Mycobacterium tuberculosis</td>
<td>NP 039697</td>
<td>171</td>
<td>0.967</td>
<td>3e-05</td>
<td>0.963</td>
<td>3e-05</td>
<td>0.957</td>
<td>3e-05</td>
<td>0.938</td>
</tr>
<tr>
<td>pMLP1</td>
<td>Micromonospora sp. ATCC 39149</td>
<td>AAO46045</td>
<td>170</td>
<td>0.958</td>
<td>3e-05</td>
<td>0.957</td>
<td>3e-05</td>
<td>0.957</td>
<td>3e-05</td>
<td>0.947</td>
</tr>
<tr>
<td>VWB</td>
<td>Streptomyces venezuelae</td>
<td>YP 115515</td>
<td>188</td>
<td>0.928</td>
<td>6e-05</td>
<td>0.914</td>
<td>8e-05</td>
<td>0.897</td>
<td>1e-04</td>
<td>0.897</td>
</tr>
</tbody>
</table>

### Table 3.1 Similarity of IntPIS136 with PDB structures
1a0p- Structure file for *E. coli* XerD, 4crx- Structure file for Phage P1 Cre recombinase, 1aih- Catalytic domain of HP1 (*Haemophilus influenzae*) Int, 1ae9- Catalytic domain of phage Lambda Int.
Identification of Integrase

generated alignments to all the four structures in the database, 1a0p (XerD), 4crx (Cre), 1aih (HP1 Int catalytic domain) and 1ae9 (Lambda Int catalytic domain).

Only the C-terminal regions of the proteins were used in the alignments (Table 3.1). Since all of the pairwise alignments were consistent with each other, the entire collection was assembled into a composite multiple sequence alignment. ClustalX (Thompson et al., 1994) was used to evaluate the protein sequence conservation. Secondary structure of Int was analyzed using PSIPRED protein structure prediction server (McGuffin et al., 2000). Phylogenetic relationship of Int^{P136} with the proteins picked by BLAST searches (Table 3.1) was determined by multiple alignments using ClustalX (Thompson et al., 1994). The rooted tree was generated using the Draw N-J Tree option and viewed using NJPLOT.

3.3 RESULTS

3.3.1 Identification of the integration region

The integration of the phage DNA into the host genomic DNA involves a specific sequence in the phage DNA called attachment site (attP) and the process is mediated by a site specific recombinase called integrase. In most of the phages, it has been observed that the integrase gene is present in the vicinity of the attP. The recombinase recognizes a sequence in the genome (named attB) of the host that is identical with that of the attP. The recombination between these two sites leads to the disruption of the parent DNA and forms two new products termed the left (attL) and the right (attR) arms.

To identify the region the DNA of the phage and the lysogen PA136 was digested with different enzymes. Upon hybridization with the whole phage DNA (Figure 3.1), the host (PA136) DNA digests Clal, NcoI, Nhel+ NcoI and SpeI+ NcoI were missing one fragment each when compared with similar digests of the phage (PIS136) DNA. The missing fragments: ~4.2Kb fragment of Clal (C5), ~2.1Kb fragment of Nhel+ NcoI (NN9), ~2.7Kb fragment of NcoI digest (N6) and ~2.1Kb fragment of NcoI+ SpeI (NS7) are thus likely to harbour the region responsible for the attachment of the phage DNA with the cognate host DNA.
Figure 3.1 Identification of the integration region of PIS136

Phage DNA and genomic DNA from the lysogen *Saccharomonospora* sp. PA136 were digested with restriction enzymes (Lane 1: *HindIII* digest of λ DNA, 2&3: *Clal*, 4&5: *PvuII*, 6&7: SnaBI+ *PvuII*, 8&9: *Nhel*+ *NcoI*, 10&11: *NcoI* and 12&13: *Spel*+ *NcoI*) and resolved on a 0.7% agarose gel. The DNA was hybridized with the whole phage DNA as probe at 50 °C in presence of 35% formamide for 24 hr. The blot was washed with 2X SSC+ 0.1% SDS at 50 °C for 10 min and then in 0.1X SSC+ 0.1% SDS at 35 °C and at 32 °C for 10 min each. Lanes 2,4,6,8,10 and 12 represent the lysogen DNA while 3,5,7,9,11 and 13 represent Phage DNA. The fragments of the phage harbouring the probable region for integration are shown by arrows.
Figure 3.2 Identification of the junction fragments

Genomic DNA of the lysogen *Saccharomonospora* sp. PA136 and the phage DNA were digested with restriction enzymes (Lane 1& 2: *NheI*+ *NcoI*, 3& 4: *EcoRI*+ *NcoI*, 5& 6: *KpnI*+ *EcoRI*, 7& 8: *ClaI*, 9& 10: *KpnI*+ *ClaI*, 11& 12: *KpnI*, 13& 14: *KpnI*+ *NcoI*, 15& 16: *KpnI*+ *NheI*) and separated on a 0.8% agarose gel. The DNA was hybridized with the C5 fragment as a probe at 50 °C in presence of 35% formamide for 24 hr. The blot was washed with 2X SSC+ 0.1% SDS at 50 °C for 10 min and then in 0.1X SSC+ 0.1% SDS at 35 °C and at 32 °C for 10 min each. Lanes 1,3,5,7,9,11,13 and 15 are phage and Lanes 2,4,6,8,10,12,14 and 16 are lysogen DNA. The probable junction fragments are denoted by dots.
3.3.2.1 Identification of the junction fragments

To further confirm that the above fragments have homology to one another and to identify the probable junction regions between the host and the phage, both the phage and the host DNA were digested with the enzymes Nhel+ NcoI, EcoRI+ NcoI, KpnI+ EcoRI, Clal, KpnI+ Clal, KpnI, KpnI+ NcoI and KpnI+ Nhel. The DNA was hybridized with C5 as a probe.

Surprisingly, C5 did not hybridize with NN9 as expected, but hybridized with a much bigger fragment (Figure 3.2). It also did not hybridize with the fragments N6 and NS7 suggesting that these fragments may not have integration region of the phage but likely to harbour the region responsible for packaging (discussed below).

The KpnI+ Clal and KpnI digest of the host DNA showed two extra bands hybridizing with C5 compared to the phage DNA (Figure 3.2). In the other digests: Nhel+ NcoI, EcoRI+ NcoI, KpnI+ EcoRI, Clal, KpnI+ NcoI and KpnI+ Nhel, only one hybridizing band was seen. These hybridizing fragments may contain the junction between the phage and the host DNA.

3.3.2.2 Cloning and Sequencing of the Junction Fragments

The fragments in the host DNA that were identified by hybridization with C5 were used to sequence the region of crossover between the phage and host DNA. After failing to sequence the junction fragments, two different enzyme digested fragments: KpnI+ Clal (~10Kb and ~2Kb, Lane 10 Figure 3.2) and NcoI+ Nhel (~2.5Kb, Lane 2 Figure 3.2) were selected. The fragments were cloned separately into pGEM-5fZ(+) and pUC19 to facilitate sequencing. Presence of inserts was confirmed by restriction digestion as described in methodology. Surprisingly, all the clones which hybridized strongly with C5 fragment had inserts that were less than 1Kb. All attempts to obtain correct sized insert by cloning failed and all the time the inserts were smaller than the original fragments used for cloning. Although these clones were sequenced, a probable junction between the phage DNA and the host DNA was not identified.

3.3.3 Sequencing the C5 fragment

Integration of the phage DNA into the host is mediated by the integrase gene at the attP site. To identify the genes present in the C5 fragment and to locate the probable attP, the entire C5 fragment was sequenced. To complete the entire sequence, multiple sub-
<table>
<thead>
<tr>
<th>Source of DNA</th>
<th>Enzyme Used</th>
<th>Name of the Fragment</th>
<th>Approx. Size</th>
<th>Name of the Plasmid</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIS136</td>
<td>Smal</td>
<td>Sma15</td>
<td>1.6Kb</td>
<td>pGS15</td>
</tr>
<tr>
<td>PIS136</td>
<td>Smal</td>
<td>Sma16</td>
<td>1.4Kb</td>
<td>pGS16</td>
</tr>
<tr>
<td>PIS136</td>
<td>Smal</td>
<td>Sma20</td>
<td>700bp</td>
<td>pGS20</td>
</tr>
<tr>
<td>PIS136</td>
<td>Mlu1</td>
<td>Mlu24</td>
<td>2.4Kb</td>
<td>pGM24</td>
</tr>
<tr>
<td>PIS136</td>
<td>Mlu1</td>
<td>Mlu28</td>
<td>1.2Kb</td>
<td>pGM28</td>
</tr>
<tr>
<td>PIS136</td>
<td>Sal1</td>
<td>Sal5</td>
<td>4.2Kb</td>
<td>pGS5</td>
</tr>
<tr>
<td>pGS15</td>
<td>SacI+ NarI</td>
<td>S15-SN</td>
<td>1.0Kb</td>
<td>pGS15-SN</td>
</tr>
<tr>
<td>pGS5</td>
<td>SacI</td>
<td>S5-S</td>
<td>1.6Kb</td>
<td>pGS5-S</td>
</tr>
<tr>
<td>pGM24</td>
<td>EagI+ Scal</td>
<td>M24-ES1</td>
<td>700bp</td>
<td>pGM24-ES1</td>
</tr>
<tr>
<td>pGM24</td>
<td>EagI+ Scal</td>
<td>M24-ES2</td>
<td>1.4Kb</td>
<td>pGM24-ES2</td>
</tr>
<tr>
<td>pGM28</td>
<td>HindIII+ AalI</td>
<td>M28-HA</td>
<td>1.0Kb</td>
<td>pGM28-HA</td>
</tr>
<tr>
<td>pGS16</td>
<td>Mlu1</td>
<td>S16-M</td>
<td>330bp</td>
<td>pGS16-M</td>
</tr>
</tbody>
</table>

**Table 3.2 Fragments Used for Sequencing C5**

PIS136 DNA was digested with restriction enzymes Smal, SalI and MluI and the fragments were cloned in the plasmid pGEM-5fZ(+) and digested with the same enzymes. The recombinants were sequenced and for subcloning they were further digested with restriction enzymes identified in the sequenced region. The sub-fragments were end-filled and then cloned in EcoRV site of pGEM-5fZ(+).
**Figure 3.3 Strategy used for sequencing the C5 fragment**

The Phage DNA was digested with different enzymes to identify fragments that hybridize with the C5 fragment. The fragment Sal5 was cloned into the SalI site, fragments Mlu24 and Mlu28 into MluI site and rest of the fragments were end filled and were cloned into the EcoRV site of pGEM-5Zf(+) to sequence both the strands. The enzymes used to obtain the fragments are listed in Table 3.2.
Identification of Integrase

fragments were sequenced. The fragments that were chosen for sequencing are listed in Table 3.2 and the positions of the fragments in C5 are given in Figure 3.3. The entire C5 fragment was of 4410bp.

3.3.4 Identification of a Palindrome and its Sequencing

The sequence obtained with one of the primer using the plasmid pGS16 as a template was not complete and terminated before the full length sequence could be read. The Polymerase seemed to “fall off” in between. This suggested that the Sma16 fragment may be having a region with a strong secondary structure which the enzyme is unable to read through under normal sequencing conditions.

To shorten the fragment and try to sequence from a different end of this region, the Sma16 fragment was cut with MluI to get a fragment of ~330bp (S16-M; figure 3.3). The sequence of this fragment also terminated at the same position as that of Sma16. Moreover, the sequence obtained with the forward and reverse primer was identical at the region where the sequence terminated. The result suggested the presence of a long palindrome which may form a secondary structure.

In order to read through the region, a thermostable polymerase was used. The Thermosequenase cycle sequencing kit (Amersham Biosciences, USA) was able to read through the region of the S16-M fragment where the polymerase was falling off. The sequence obtained suggested that the region harbours an interrupted palindrome that can fold into a cruciform structure. The stem of the cruciform structure was made of 21bp with 8bp loop region (Figure 3.4a). The free energy of the stem-loop is calculated to be \(-39.6\) Kcal mol\(^{-1}\), using the program PCFOLD 4.0 (Zuker, 1989) which shows that the palindrome is a stable structure.

3.3.5 Sequencing of the Regions Flanking C5

The Sal5 fragment has a Clal site within its sequence (Figure 3.3) and the sequence extends to another 227bp beyond the Clal site. This part of the region has already been sequenced as part of a flanking fragment C6 (Soni, 2001).

To find the complete sequence of the putative ORF that may code for a dCTP deaminase, the region upstream to the C5 fragment was sequenced. Hybridization with the C5 fragment identified a ~3.0Kb NarI+ SacI fragment (NS4). The NS4 fragment was
Figure 3.4  
Panel a: The sequence of the stem-loop region was determined using the Thermosequenase kit from USB (USA). The secondary structure of the sequence was determined using the RNA Secondary structure prediction program PCFOLD 4.0.
Panel b: shows the different ORFs predicted in the C5 fragment. The position of the interrupted palindrome is also shown. The restriction enzymes and their positions that were used to sequence the entire fragment are given. The fragment contains two complete and two partial ORFs.
Figure 3.5 Nucleotide and protein sequence of the putative dCTP Deaminase

The nucleotide sequence of a 650bp region upstream to the integrase gene comprises of an ORF of 582bp. The ORF codes for a putative protein of 193aa that has high homology with dCTP deaminase. The amino acid sequence is given above the codon. The putative sequence of the Ribosome Binding Sequence is underlined.
digested with Smal to get the fragments; NS4Sma1 (~1.9Kb) and NS4Sma2 (~1.2Kb). The fragment NS4Sma2 was end-filled and cloned in the EcoRV site of pGEM-5fZ(+/-).

The sequence obtained with the NS4-S2 fragment had the N-terminal region of the putative dCTP deaminase gene. The entire ORF (Figure 3.5) is of 582bp (194aa) with a very high homology to the dCTP deaminase gene products of *M. leprae* (2e-73), *S. coelicolor* (3e-73) and *M. tuberculosi*s H37Rv (3e-73).

### 3.3.6.1 ORFs Prediction in the C5 Fragment

The sequences obtained from the overlapping clones were aligned using ClustalW and Generunner. The total length of the sequenced region was of 4410bp. The entire sequence has been submitted to GenBank with the sequence accession number AF459434. Two complete ORFs and two partial ORFs were detected in the sequenced region (Figure 3.4b). One complete ORF (ORF1) of 1221bp (Figure 3.6) was identified which had a very high homology with the integrase family of proteins. The other complete ORF (ORF2) is of 171bp and is present upstream to and in the same frame as that of ORF1 but a putative RBS could not be detected. As this work was not pursued further the status of the ORF2 expression is not known.

One of the partial ORF (ORF3) of 2214bp showed a very high nucleotide and aminoacid homology to a putative α-subunit of DNA Polymerase III and is in the same frame as that of ORFs 1 and 2. The ORF was completely sequenced from a neighbouring fragment, C6 and the complete sequence has been reported (Soni, 2001). The second partial ORF (ORF4) codes for a putative dCTP deaminase and only the C-terminal region is present within C5. The ORF4 is present upstream to ORF2 but in a different coding frame.

### 3.3.6.2 Recognition of Putative integrase gene by Sequence Analysis

Sequence analysis of the complete ORF1 suggested that it may code for a protein of 406aa (calculated molecular weight of 45421.4Da). BLASTP and Psi-BLAST searches detected homology with integrases from a number of phages and proteins belonging to the tyrosine recombinase family. Conserved domain search against the pfam database using NCBI Domain Architecture Retrieval Tool (NCBI DART) found homology to the phage integrase family protein motif (pfam00589) with a score of 3.6e-09 and a score of 1.6e-11 was found using hmmpfam at the Washington University School of Medicine (www.gwumc.edu). All the above results indicated that the ORF1 might code for an
**Figure 3.6 Nucleotide and Protein sequence of the putative integrase gene**

The nucleotide sequence of a 1890bp region in the C5 fragment comprises of two complete ORFs. The ORF1 is of 1221bp and codes for a putative protein that belongs to the family of phage integrases. The amino acid sequence is given above the codon. The putative promoter and RBS sequences are underlined.
Figure 3.7 Multiple sequence alignment of the integrase of PIS136 with other Integrases

Sequences used for the alignment were selected based on the BLASTP result. The amino acids that form the catalytic domain only were used for generating the alignment (Table 3.2). The multiple sequence alignment was constructed from individual pairwise alignments to Cre or XerD constructed by mGenTHREADER. The secondary structure elements are of PIS136 sequence as determined by using PSIPRED protein structure prediction server. The residues were coloured according to their chemical properties, using CLUSTALX. The different regions of conservation were assigned according to Nunes-Düby et al., 1998.
Integrase protein that may be responsible for the integration of the phage PIS136 into the host genome.

Further analysis of the protein product of ORF1 using the ProtParam Tool of ExPASy Proteomics server (http://au.expasy.org) classified the protein as hydrophobic with Grand average of hydropathicity of -0.566. The computed pl is 10.33 with a predominance of Arginine residues (13.5%). The protein has been classified as unstable with a half life of >10 hours in E. coli. The HTHScan tool of GCG detected the presence of a HTH motif spanning from 117-136 residues (WVDDLVRAGYSPSYVRSFA) with a score of 1.499E-02.

Multiple sequence alignment (Figure 3.7) with those proteins which showed a high homology with the protein encoded by ORF1, as described in section 3.2.7.2 clearly demonstrated the conservation of residues into different boxes and patches as described by Nunes-Duby et al. (1998). The conserved tetrad of R-H-R-Y was easily identifiable where the Histidine residue is replaced by a Tyrosine in the protein product of ORF1. Therefore, here onwards the ORF1 product will be referred as Integrase of PIS136 (Int\textsuperscript{PIS136}).

Int\textsuperscript{PIS136} Shows a very high homology with proteins identified as putative phage integrases in actinomycetes like Nocardia farcinica IFM 10152, Streptomyces avermitilis MA-4680 and Streptomyces coelicolor A3(2) (Table 3.3). Phylogenetic analysis showed that Int\textsuperscript{PIS136} groups into a cluster (Figure 3.8) comprising of Int from phages mv4, Mx8, Conjugative Transposon Tn916, Prophage Lp2 of Lactobacillus plantarum and the putative phage Integrase of S. coelicolor A3(2).

3.4 DISCUSSION

DNA of the temperate bacteriophages integrate into their host genome by site-specific recombination which is mediated by a phage encoded protein, Integrase. Integrases from most of the phages belong to the tyrosine recombinase family. Indeed, the family was initially called as the phage integrase family because of the well studied Lambda integrase that belongs to this family. The identification of new phages from diverse bacteria has however changed the scenario that the integration reaction is not exclusive to Tyrosine recombinases.

Lactococcal phage TP901-1, Actinophages φC31 and R4 use a recombinase that has a resemblance to the resolvase/invertase family (Kuhstoss and Rao, 1991; Matsuura et al.,
Figure 3.8 Phylogenetic relationship of Int\textsuperscript{PIS136}

The translated amino acid sequence of the putative integrase gene was searched for homologous sequences using BLASTP. The entire length of the protein was used in multiple alignments using ClustalX and the phylogenetic relationship with the other Integrases was determined. A rooted tree was generated using the Draw N-J Tree function and viewed using NJPLOT. The source of the proteins used for the phylogenetic analysis are as given in Table 3.1.
Identification of Integrase

1996). These proteins are described as large resolvases because of their higher molecular weight compared to the classical invertases/resolvases. Though phages have been shown to use the serine recombinase family proteins for integration into the bacterial genome, they remain to be a minority group.

Integration of phage DNA into the host genomic DNA results in the formation of two new recombination products, \textit{attL} and \textit{attR} in place of the phage attachment site \textit{attP}. The \textit{attL} and \textit{attR} fragments form the junction between the host and the phage DNA in the lysogen. These can be identified in the lysogen DNA by hybridization with whole Phage DNA. The lysogen DNA can be seen to have two new fragments in place of the intact \textit{attP} present in the phage DNA. When the DNA from the natural lysogen of PIS136, \textit{Saccharomonospora} sp. PA136, was hybridized with the whole phage DNA, a \textit{ClaI} fragment of \textasciitilde 4.2Kb was missing in the lysogen DNA lane (Figure 3.1). When the \textasciitilde 4.2Kb fragment was fully sequenced, it was found to harbour an \textit{integrase} gene. Amino acids sequence alignment showed that this protein belongs to the Tyrosine recombinase family.

Attempts were made to identify the junction fragments in the host PA136 genome using C5 as a probe. Possible junction fragments could clearly be identified by Southern hybridization (Figure 3.2). Surprisingly, the intact integrative fragments were also present in the lysogen DNA in all the different digests. This fragment was identical to that found in the phage DNA digests (compare the lanes of phage DNA with Lysogen DNA in Figure 3.2). This result is possible if an extrachromosomal form of the phage DNA co-exists with the integrated prophage form. A free form of the \textit{Lactobacillus delbrueckii} phage mv4 was also identified in lysogens along with a stably integrated prophage (Lahbib-Mansais \textit{et al.}, 1992). Spontaneous induction of the prophage in a part of the lysogen cell population could explain the presence of this probable extrachromosomal form of the phage PIS136. Another explanation for the presence of the intact integrative fragment signal in the lysogen DNA blots can be explained by tandem integration of the phage DNA at the same bacterial site as seen in Mx8 (Magrini \textit{et al.}, 1999b).

The region responsible for the integration of the phage DNA has retained a specific organization in almost all the phages studied till date. The \textit{integrase} gene, the site of the recombination (\textit{attP}) and the gene responsible for the excision of the phage DNA (\textit{excisionase}) are generally clustered. A putative 171bp ORF2 is present upstream to the \textit{int} gene in PIS136, which may code for Xis. Though a putative RBS is not prominent, the translated product of 56aa has a predicted pI of 10.67 which is similar to other Excisionases.
Identification of Integrase

Though most of the Xis do not share high homology at the protein level they can be identified by the presence of an N-terminal motif (Boccard et al., 1989a). This motif is common to lambdiod phages and the integrative plasmid pSAM2 of Streptomyces ambofaciens. In the putative ORF2 of PIS136 this conserved N-terminal motif could not be identified. Thus it is assumed that the putative ORF2 in the integration region of PIS136 might not code for an Excisionase. In this regard PIS136 shares its organization with the Haemophilus influenzae phage HP1, Mycobacteriophage Bxb1 ( Mediavilla et al., 2000) and phages of lactic acid bacteria that do not have a xis in front of int. In HP1, the excision reaction needs a protein called Cox that also acts as a transcriptional regulator for many other proteins (Lewis and Hatfull, 2001). It remains to be seen if such a Recombination Directionality factor (RDF) is present in PIS136 or not.

Phages that infect bacteria with a high G+C content in their genomes, like Mycobacteriophage D29 (Ford et al., 1998), φC31 (Smith et al., 1999) and φBT1 (Streptomyces lividans), are known to encode a dCTP (dCMP) deaminase upstream to the int. This enzyme is responsible for deaminating Cytosine to Uracil which may in turn be converted to Thymidine and ensure its proper supply during lytic cycle in a cytosine rich environment. The role of dCTP deaminase identified in PIS136 may thus be limited to ensure proper replication of the phage PIS136 DNA. Incidentally, unlike in other phages, PIS136 has an unusual gene organization where integrase gene is followed by α- subunit of DNA Polymerase III gene (dnaE).

Many of the temperate phages are known to carry a DNA Polymerase similar to the DNA Polymerase I. Actinophages like φC31 and Bxb1 ( Mediavilla et al., 2000) have a DNA Pol I subunit for processivity. Other phages like φ29 ( Escarmis and Salas, 1982) and RB69 ( Wang et al., 1997) have DNA Pol II. Both these polymerases have comparatively less processivity than that of DNA Polymerase III. So far just one phage; SPBc2 (Bacillus subtilis, GeneBank id NP_046685) has been shown to have DNA Polymerase III. Whatever be the type of polymerase present in these phages, the location of the gene is usually distant from the region responsible for integration. In phages like L5 ( Hatfull and Sarkis, 1993) and Bxb1 ( Mediavilla et al., 2000), the DNA Pol gene is present ~5.0Kb away from the int gene and also the direction of transcription is not same. The presence of the DNA Pol III gene immediately down stream and transcribing in the same direction as that of int in PIS136 is unique and probably significant for the survival or the wide host range of PIS136 (could be a matter of further research).
The DNA sequence of C5 was analyzed for the presence of putative IHF binding sites using the FINDPATTERNS algorithm of GCG (Genetics Computer Group). The consensus sequence (TATCAAN₄TTG) of binding sites for *E. coli* IHF (Goodrich *et al.*, 1990) was used for the analysis allowing 25% mis-matches. Six putative IHF binding sites were detected within a 2.5Kb region encompassing parts of *int* and *dnaE* genes, two of which are present in the non-coding strand.

The sequence and position (from the start of C5) of putative IHF binding sites are given below:

<table>
<thead>
<tr>
<th>Strand</th>
<th>Start</th>
<th>Start</th>
<th>Sequence</th>
<th>% Mis-match</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>838</td>
<td>850</td>
<td>TATCTACTCTTCG</td>
<td>23.07%</td>
</tr>
<tr>
<td>+</td>
<td>958</td>
<td>970</td>
<td>TATCCAGGAATGG</td>
<td>15.4%</td>
</tr>
<tr>
<td>-</td>
<td>978</td>
<td>965</td>
<td>CATCTACCATTC</td>
<td>23.07%</td>
</tr>
<tr>
<td>-</td>
<td>2203</td>
<td>2190</td>
<td>CATCAAAGTCCTGT</td>
<td>23.07%</td>
</tr>
<tr>
<td>+</td>
<td>3001</td>
<td>3014</td>
<td>CATCGAGCGTTC</td>
<td>23.07%</td>
</tr>
<tr>
<td>+</td>
<td>3370</td>
<td></td>
<td>CATCGAGCGGTTC</td>
<td>23.07%</td>
</tr>
</tbody>
</table>

Consensus sequence: TATCAANNNTTG

The distribution of the sites was very inconsistent compared to that found in the phage Lambda *attP* region, wherein 3 sites for IHF binding are present within a 240bp region. In PIS136, two sites are present on the coding strand within the *int* ORF and two sites within the *dnaE* ORF. One of the sites on the non-coding strand was found in the intergenic region between *int* and *dnaE* and the other within the *int* ORF. The significance of the presence of these sites can only be speculated till the exact *attP* region is identified. The IHF binding sites identified are based on the *E. coli* consensus sequence which has a relatively low G+ C content unlike in PIS136 which has a high G+ C content (71mol percent). Thus the sequence of the IHF binding site may actually be very different for PIS136.

The sequence of *attP* could not be identified by cloning the fragments in the vectors pGEM-5fZ(+) and pUC19. Attempts were also made to sequence the genomic DNA using primers intout-N and intout-C designed to sequence outward from the *int* gene. The whole genome of PA136 was used as a template to sequence the regions flanking the *int* gene in an attempt to identify the junction sequence. Some sequences were obtained but they were of free phage. It has been mentioned that existence of a free extrachromosomal copy of PIS136 cannot be ruled out. Thus, so far all attempts to identify a possible *attP* of PIS136 has failed.
The *integrase* gene identified in PIS136 shows very high homology to the Tyrosine Recombinase family of phage integrases. Pair wise alignment of Integrase sequence of PIS136 shows considerable homology with the secondary structures of the four members whose 3D crystal structures have been solved (Table 3.1). Comparisons against the pfam database identified a conserved phage Integrase motif common to all proteins of this family. The predicted pI of 10.33 is similar to that of other Integrases identified till date. Multiple sequence alignment of PIS136 Int with Integrases from different phages identified regions of marked sequence similarity to the Box I and Box II regions. The high degree of conservation and clustering of hydrophobic residues in PIS136 is evident from the alignment (Figure 3.6). Deviating from the accepted signature tetrad of residues R-H-R-Y that identifies the integrases, PIS136 Int has a Tyrosine in place of the Histidine. This replacement is present in *S. coelicolor* A3(2) (Figure 3.6) and in the SLP1 element of *Streptomyces lividans*. The alignment also identifies other conserved residues present in patches I, II and III that are essential for recombination.

Surprisingly, phylogenetic analysis shows that the Int<sup>PIS136</sup> clusters away from the Int of other Actinophages phages like VWB, Ms6 and L5 (Figure 3.7). The intimacy of its phylogeny with Integrases of Conjugative Transposon Tn916, phage Mx8 and *Lactobacillus* sp. (have much lower G+C content than *Saccharomonospora* sp. PA136) is in deviation to the usually observed scenario where proteins belonging to similar genera cluster together. Though phylogenetically, the genus *Saccharomonospora* is closer to the genus *Nocardia*, Int<sup>PIS136</sup> does not cluster with the putative phage Integrase of *Nocardia farcinica*. The clustering of the Integrases of Actinophages VWB, Ms6, pMLP1 and putative phage integrase of *Nocardia farcinica* IFM 10152 (Figure 3.7) supports the fact that these organisms belong to similar genera. One surprising observation is the high amount of similarity with the Integrase of the Conjugative Transposon Tn916, where Int<sup>PIS136</sup> shares a very high percentage of similar residues with Int<sup>Tn916</sup> in a short stretch (denoted by the total number of similar and identical residues, Table 3.3) used for the alignment. Such a similarity is next only to the Int from prophage Lp2 (*Lactobacillus plantarum* WCFS1) and the putative phage Integrases of *Rubrobacter xylanophilus* DSM 9941 and *Streptomyces avermitilis* MA-4680.

The secondary structure predicted using the PSIPRED protein structure prediction server follows the pattern of arrangement of helices and strands that has been identified from the crystal structures of Integrases of phages HP1 and Lambda, *E. coli* XerD and Cre
### Phages

<table>
<thead>
<tr>
<th>Source</th>
<th>Genbank id</th>
<th>Total residues</th>
<th>Residues used For alignment</th>
<th>% identity</th>
<th>% similarity</th>
<th>Score</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ms6</td>
<td>Mycobacterium sp.</td>
<td>AAD03774</td>
<td>372</td>
<td>344</td>
<td>26</td>
<td>73.6</td>
<td>2e-11</td>
</tr>
<tr>
<td>Mv4</td>
<td>Lactobacillus delbrueckii</td>
<td>AAG31327</td>
<td>427</td>
<td>222</td>
<td>23</td>
<td>42.4</td>
<td>0.038</td>
</tr>
<tr>
<td>Mx8</td>
<td>Myxococcus xanthus</td>
<td>BAA13089</td>
<td>533</td>
<td>376</td>
<td>23</td>
<td>45.1</td>
<td>0.006</td>
</tr>
<tr>
<td>L5</td>
<td>Mycobacterium tuberculosis</td>
<td>NP_039697</td>
<td>371</td>
<td>362</td>
<td>21</td>
<td>40.8</td>
<td>0.1</td>
</tr>
<tr>
<td>pMLP1</td>
<td>Micromonospora sp. ATCC 39149</td>
<td>AAO46045</td>
<td>392</td>
<td>308</td>
<td>22</td>
<td>62.8</td>
<td>3e-08</td>
</tr>
<tr>
<td>VWB</td>
<td>Streptomyces venezuelae</td>
<td>YP_115515</td>
<td>427</td>
<td>395</td>
<td>24</td>
<td>71.2</td>
<td>8e-11</td>
</tr>
<tr>
<td>Lambda</td>
<td>Escherichia coli</td>
<td>P03700</td>
<td>356</td>
<td>345</td>
<td>22</td>
<td>62.8</td>
<td>3e-08</td>
</tr>
</tbody>
</table>

### Prophages

<table>
<thead>
<tr>
<th>Source</th>
<th>Genbank id</th>
<th>Total residues</th>
<th>Residues used For alignment</th>
<th>% identity</th>
<th>% similarity</th>
<th>Score</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lp2</td>
<td>Lactobacillus plantarum WCFS1</td>
<td>NP_785910</td>
<td>373</td>
<td>267</td>
<td>25</td>
<td>64.3</td>
<td>9e-09</td>
</tr>
<tr>
<td>phiRv2</td>
<td>Mycobacterium tuberculosis H37Rv</td>
<td>NP_217175</td>
<td>375</td>
<td>375</td>
<td>23</td>
<td>60.1</td>
<td>2e-07</td>
</tr>
</tbody>
</table>

### Bacterial

<table>
<thead>
<tr>
<th>Source</th>
<th>Genbank id</th>
<th>Total residues</th>
<th>Residues used For alignment</th>
<th>% identity</th>
<th>% similarity</th>
<th>Score</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N.frac</td>
<td>Nocardia farcinica IFM 10152</td>
<td>YP_116920</td>
<td>403</td>
<td>341</td>
<td>25</td>
<td>78.2</td>
<td>6e-13</td>
</tr>
<tr>
<td>R.xyla</td>
<td>Rubrobacter xylanophilus DSM 9941</td>
<td>ZP_00186935</td>
<td>379</td>
<td>360</td>
<td>25</td>
<td>87.8</td>
<td>8e-16</td>
</tr>
<tr>
<td>S.aver</td>
<td>Streptomyces avermitilis MA-4680</td>
<td>NP_824885</td>
<td>393</td>
<td>361</td>
<td>24</td>
<td>95.5</td>
<td>4e-18</td>
</tr>
<tr>
<td>S.coeli</td>
<td>Streptomyces coelicolor A3(2)</td>
<td>NP_625296</td>
<td>370</td>
<td>338</td>
<td>25</td>
<td>77.0</td>
<td>1e-12</td>
</tr>
</tbody>
</table>

### Conjugative Transposons

<table>
<thead>
<tr>
<th>Source</th>
<th>Genbank id</th>
<th>Total residues</th>
<th>Residues used For alignment</th>
<th>% identity</th>
<th>% similarity</th>
<th>Score</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tn916</td>
<td>Enterococcus faecalis</td>
<td>P22886</td>
<td>405</td>
<td>273</td>
<td>26</td>
<td>52.4</td>
<td>4e-05</td>
</tr>
</tbody>
</table>

**Table 3.3 Similarity of Int^PIS136 with Integrases identified by BLAST search**

The sequences were aligned using the pair-wise alignment tool of NCBI BLAST (bl2seq) and the similarity to Int^PIS136 determined.
Identification of Integrase

of Phage P1. The presence of the Tyrosine nucleophile on a helix is consistent with the crystal structure for Cre of phage P1 (Guo et al., 1997) and differs from that of phage Lambda (Kwon et al., 1997).

The length of the region C-terminal to the Tyrosine residue is rather varied in different phages. Only 14 residues are present in Lambda beyond the Tyrosine while in the phage Mx8, 179 residues are present. The C-terminal region of Mx8 plays a more proactive role in that that the loss of the residues due to integration of the phage (attP lies within int) leads to the reduction in the specific activity of the protein (Magrini et al., 1999b). It is also presumed that excision of Mx8 prophage is catalyzed by a modified, prophage encoded IntX (Magrini et al., 1999a). The presence of a long C-terminus with 55 residues in PIS136 Int beyond the Tyrosine nucleophile may be speculated as to its role in the integration reaction.

In conclusion, an ORF coding for a putative 406aa Integrase was identified in the phage PIS136 which is related to the Integrases of phage Mx8, mv4 and putative prophage integrase of S. coelicolor. Though potential junction fragments were identified, neither a putative attachment site nor a target for integration could be identified. The product of the ORF1 which belongs to the Tyrosine Recombinase family may be responsible for the integration of phage PIS136 into the host genome. Putative E. coli IHF binding sites were also identified around the integrase gene. A putative dCTP deaminase gene and α-subunit of DNA Polymerase III were also identified in the sequenced region. A strong secondary structure present immediately downstream to the dCTP deaminase gene may be a potential transcriptional terminator for late genes or a potential attP site and thus responsible for the instability of all potential attP related clones.