Genetic analysis of indigenous plasmid, pPDL2 of *Flavobacterium* sp. ATCC 27551and its use in engineering benzoate degrading *Acinetobacter* sp. DS002

Thesis submitted for the Degree of
Doctor of Philosophy
In
Animal Sciences

Supervisor
Prof. S. Dayananda

By
P. Emmanuel Vijay Paul

Dept. of Animal Sciences
School of Life Sciences
University of Hyderabad
Hyderabad – 500 046, India

March, 2011

Enrolment No. 05LAPH06

Contents

Introduction

Materials & Methods

Contents

Contents

Results & Discussion

Chapter-1

Contents

Chapter-2

Chapter-3

University of Hyderabad

(Central University established in 1974 by an act of Parliament)

HYDERABAD-500 046, INDIA

CERTIFICATE

This is to certify that **Mr. P. Emmanuel Vijay Paul** has carried out the research work embodied in the present thesis under my supervision and guidance for a full period prescribed under the Ph.D. ordinance of this University. We recommend his thesis entitled **"Genetic analysis of indigenous plasmid, pPDL2 of *Flavobacterium* sp. ATCC 27551 and its use in engineering benzoate degrading *Acinetobacter* sp. DS002"** for submission for the degree of Doctor of Philosophy in Animal Sciences of this University.

Prof. S. Dayananda _____
Supervisor

Head _____
Department of Animal Sciences

Dean _____
School of Life Sciences

University of Hyderabad

(Central University established in 1974 by an act of Parliament)

HYDERABAD-500 046, INDIA

Declaration

I hereby declare that the work embodied in this thesis entitled **"Genetic analysis of indigenous plasmid, pPDL2 of *Flavobacterium* sp. ATCC 27551 and its use in engineering benzoate degrading *Acinetobacter* sp. DS002"** has been carried out by me under the supervision of Prof. S. Dayananda and this has not been submitted for any degree or diploma of any other university earlier.

Prof. S. Dayananda **P. Emmanuel Vijay Paul**

(Research Supervisor) **(Research Scholar)**

Acknowledgements

I would like to acknowledge my sincere thanks and heartfelt gratitude to my supervisor ***Prof. S. Dayananda*** *for his guidance, unwavering support throughout the course of my work.*

I thank the Dean ***Prof . Ramanadham*** *and the former Dean* ***Prof. A. S. Raghavendra*** *for allowing me to use the school facilities.*

My thanks to the Head, Department of Animal Sciences, ***Prof . Manjula Sritaran,*** *and former Head* ***Prof. S. Dayananda*** *and* ***Prof. Aparna Dutta Gupta*** *for allowing me to use the Department facilities.*

I would like to thank my Doctoral Committee members ***Prof . Manjula Sritaran*** *and* ***Prof. Appa Rao Podile*** *for their valuable suggestions.*

I would like to thank ***all the faculty members*** *of School of Life Sciences for their help whenever needed. I thank* ***Dr. Ch. Venkatramana*** *for his valuable suggestions. I also thank* ***Dr. O.H. Shetty*** *and* ***Prof. P. Reddanna*** *for allowing me to use their lab facilities. I thank* ***Gopi*** *for helping me in the oxygraph studies.*

I thank ***Mr. Anil, Mansi, Shailly****, and* ***Devilal*** *for their timely help during annotation of the sequences.*

I thank ***CSIR for*** *giving me financial assistance throughout my Doctoral studies. I thank the University for providing the financial assistance to visit TAMU, USA.*

I thank ***DST, DRDO, CSIR, UGC*** *for funding our laboratory .*

I thank ***Prof. James R. Wild****, Dept of Biochemistry and Biophysics, TAMU for inviting me to his lab as a visiting Research Scholar. Many thanks to* ***Dr. Melinda Wales, Dr. Boris, Dr. Janet Grimesly*** *for their useful discussions and valuble suggestions. I thank* ***Manoj, Gaurav, Piush, Nerraj, Subrithi*** *and* ***Dheeraj*** *for making my stay at TAMU enjoyable.*

*I thank **Mr.Lallan, Mr. Ankinedu** , **Mr.Jagan, Mr. Gopi** and **Mrs. Jythothi** who helped me in different endeavors during my work here.*

*I thank **Dr. Madhura Rekha, & Mrs. Leena Bhashyam** for their technical help whenever needed.*

*I thank my current **lab mates Purushotham, Aleem, Gopi, Prasanna, Toshi, Swetha, Sunil, Santosh** and **Ashok** for their constant support during my work.*

*I specially thank my senior colleagues **Dr. Pakala Suresh Babu , Dr. Syed Khajamohiddin, Dr. Pandey** and **Dr. Elisha** for their constant support during my work.*

*Thanks to **Manikanta (lab Asst) , Krishna (Lab Attendant)** and **Mallesh** for their assistance in administrative work.*

*I thank **Dr. Rajesh, Dr. Vijay Chiatanya, Girish, Ranjith, Vijay** all my friends at School of Life Sciences for all the help and support and making my stay at University of Hyderabad memorable.*

*Special thanks to **Naresh, Dinakar, Kiran, Vinay, Hari, Pavan, Madhu, Boney, Kauser, Christopher** and all my friends at CC Church for all their support and prayers. I greatly acknowledge the support of my Pastors. **Rev.Pascal Prakash** and **Rev. Promod**.*

I am grateful to my mother, wife, my daughter and son for their love, affection and cooperation. I also thank my brother, sister and her family for their love and support. I thank my grandfather and grandmother for their prayers and blessings.

Special thanks to all my relatives and friends for supporting me in time of need.

*I thank my **LORD God, Jesus** for his Grace, Constant presence and for help all through my life.*

P. Emmanuel Vijay Paul

As mentioned in the introduction chapter, OP compound degrading *Flavobacterium* sp. ATCC 27551 and *Brevundimonas diminuta* MG were isolated from agricultural soils collected from IRRI, Philippines (Sethunathan et al, 1973) and Texas, USA (Serdar et al, 1982) respectively. After these two reports a number of OP compound degrading bacteria belonging to different taxonomic groups were isolated from diverse geographical regions (Zhongli et al, 2001; Horne et al, 2002; Liu et al, 2005; Karpouzas and Singh, 2006; Singh, 2009). In all these OP degrading bacterial strains, a phosphotriesterase (PTE) is shown to be responsible for the hydrolysis of triester linkage found in structurally diverse groups of OP compounds (Benning et al, 1994; Cho et al, 2004). In *Brevundimonas diminuta* and *Flavobacterium* sp. ATCC 27551, the phosphotriesterase is coded by an identical organophosphate degrading (*opd*) gene present on large indigenous plasmids, pCMS1 and pPDL2 respectively (Serdar *et al*., 1982; Mulbry and Karns, 1986). Further studies from our laboratory have shown transposon-like *opd* gene cluster in *Flavobacterium* sp. ATCC 27551 (Siddavattam et al, 2003). In both pCMS1 and pPDL2 almost identical *opd* sequences were found. The DNA region identified 2.6 kb upstream and 1.7 kb downstream of the *opd* gene was found (Mulbry et al, 1988; Siddavattam et al, 2003). Beyond this region no detectable homology was found between these two indigenous plasmids. Such observation suggests existence of horizontal mobility of *opd* genes among soil bacteria. Horizontal mobility of genetic information occurs through various genetic elements such as plasmids, bacteriophages, genomic islands, Integrons, transposons, conjugative transposons and IS elements (Dobrindt et al, 2004). Existence of *opd* genes on large indigenous plasmids strengthens the hypothesis of spreading *opd* information through horizontal mobility. However, till date no experiments were conducted to validate if lateral gene transfer is contributing

for distribution of *opd* information among soil bacteria. Existence of *cis*-elements that contribute for horizontal mobility of plasmids, as well as structural information pertaining to the organization of *opd* information will be known from the primary sequence of the plasmids. Therefore, as a basic requirement to understand the HGT of *opd* gene, complete sequence of pPDL2 isolated from *Flavobacterium* sp. ATCC 27551 is determined.

3.1. Isolation and rescue cloning of Indigenous plasmid pPDL2 from *Flavobacterium* sp. ATCC 27551

The plasmid pPDL2 of *Flavobacterium* sp. ATCC 27551 is a low copy number, large indigenous plasmid of 39.75 kb in size. In order to obtain sequence of pPDL2 it has to be isolated and sub-cloned in multipurpose vectors. Further, in *Flavobacterium* sp. ATCC 27551 there are more than one plasmid. While sub-cloning pPDL2 it has to be isolated in a pure form. Therefore, plasmids pPDL2 was rescue cloned into *E. coli pir*-116 cells following procedures described in materials and methods. Initially all plasmids from *Flavobacterium* sp ATCC 27551 were isolated using a modified protocol of Courier and Nester method described in materials and methods section (Currier and Nester, 1976) and were analysed on agarose gels. As reported by Mulbry and his associates, plasmid preparations made from *Flavobacterium* sp. ATCC 27551 have revealed existence of four plasmids (Fig. 3.1) (Mulbry et al, 1986). Out of these four plasmids existence of *opd* was reported only in plasmid pPDL2. The *opd* plasmid pPDL2 was rescued from the rest of the three indigenous plasmids by rescue cloning technique.

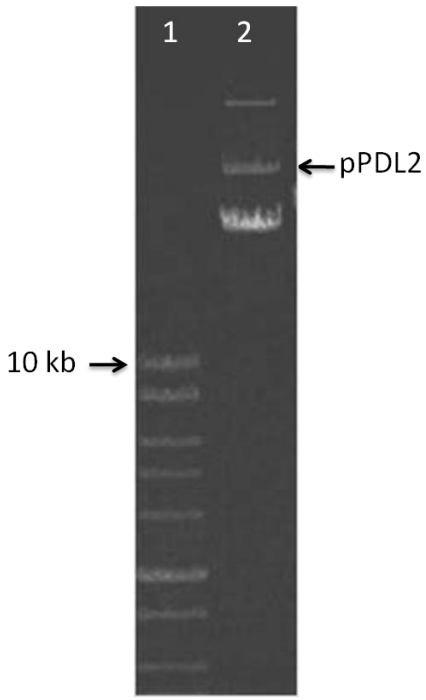


Fig. 3.1. Isolation of indigenous plasmids from *Flavobacterium* sp. ATCC 27551. Lane 1 represents 1 kb DNA ladder and lane 2 represents plasmid preparations from *Flavobacterium* sp. ATCC 27551. Plasmid pPDL2 is shown with an arrow mark.

3.2. Rescue cloning of plasmid pPDL2

The plasmid preparation containing mixture of four plasmids were tagged with R6Kγori replication origin containing mini-transposon EZ-Tn5<R6Kγori/Kan2>. These plasmids were then transformed into *E. coli pir*-116 cells and the kanamycin resistant colonies were then used for doing colony PCR using *opd* specific primers. Out of 100 colonies screened only 30 colonies gave amplification of *opd* gene indicating existence of pPDL2 in these plasmids. The authenticity and purity of pPDL2 was established by performing restriction analysis and by amplification of *orf306*, an ORF adjacently located to the *opd* gene (Siddavattam et al, 2003). Amplicons of *opd* and *orf306* obtained from the rescued clones coincided with the amplicon size obtained from *Flavobacterium* sp ATCC 27551 used as positive control, suggesting successful rescuing of pPDL2 into *E. coli*

pir-116 cells (Fig. 3.2). Further the restriction profile generated to the rescued plasmid has perfectly coincided with the similar profile reported by Mulbry and his associates (Mulbry et al, 1986),

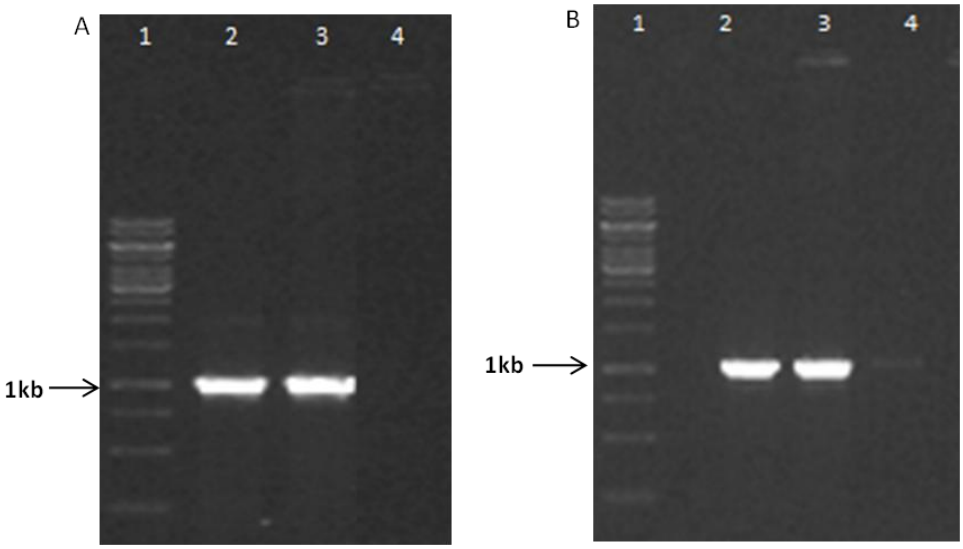


Fig. 3.2. Amplification of plasmid pPDL2 borne *opd* and *orf306* from *Flavobacterium* sp ATCC 27551 and *E. coli pir*-116 containing rescue cloned pPDL2:: Tn*5*<R6Kγori/Kan2>. Panel A, Lane 1 represents 1 kb DNA ladder. Lane2 and 3 represents amplicons of *opd* obtained from *E. coli pir*-116 (pPDL2::Tn*5*<R6Kγori/Kan2>) and *Flavobacterium* sp. ATCC 27551 respectively. Lane 4 represents negative control where a colony of *E.coli pir*-116 cells was used while performing colony PCR. Panel B indicated similar loading pattern except that *orf306* specific primers were used while performing colony PCR.

except that the size of the 14.7 kb large *Eco*RI fragment increased by 2kb. Further, a 5.7 kb *PstI* fragment found on plasmid pPDL2 has disappeared in pPDL2::Tn5<R6Kγori/Kan-2>. *In lieu* of that, two new bands with a size of 3 kb and 2.5 kb were seen after digestion with *PstI.* Obviously, this is due to existence of an internal *PstI* site in mini-transposon Ez-Tn5<R6Kγori-Kan2> (Fig. 3.3). With the exception of the increase in the size of large *Eco*RI fragment by 2 kb (A-14.7kb) and disappearance of the 3rd largest *Pst*I fragment (C-5.7kb), the restriction profile of pPDL2 perfectly matched with the restriction profile of pPDL2:: Tn*5*<R6Kγori/Kan2> (Fig. 3.3). The *Pst*I and *Eco*RI restriction profiles of

pPDL2::Tn5<R6Kγori/Kan2 gave 10 and 5 fragments respectively (Fig. 3. 3). Rescuing of pPDL2 was found to be advantageous in number of ways. One of them was apparent

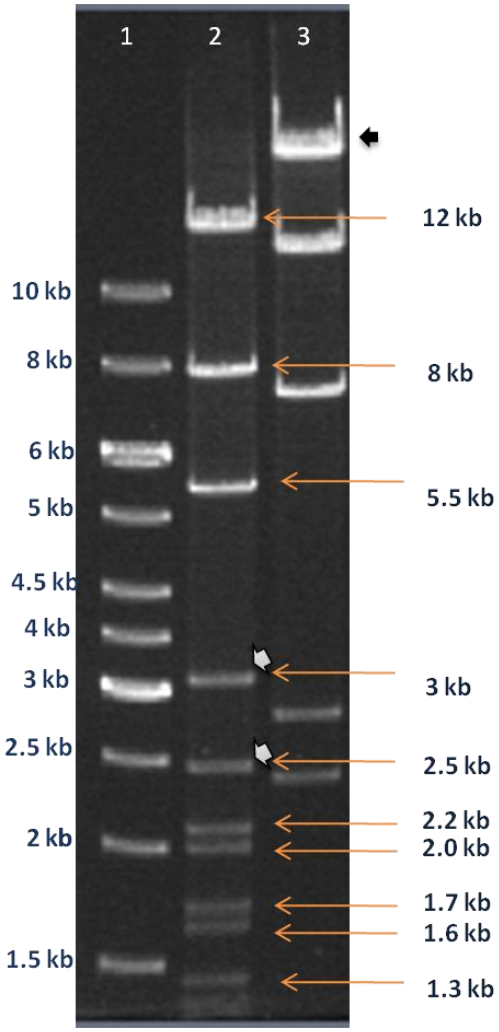


Fig. 3.3. Restriction profile of plasmid pPDL2::Tn5<R6Kγori/Kan2>.Lane 1 represents 1 kb DNA ladder. Lanes 2 and3 represents restriction profile of pPDL2::Tn5<R6Kγori/Kan2> generated by digesting with *PstI* and *Eco*RI respectively. Increase in size of *Eco*RI is indicated with a black arrow. Additional *PstI* fragments generated due to existence on mini-transposon specific *PstI* are indicated with open arrows

increase in copy number. Due to increase in copy number, pPDL2::Tn5<R6Kγori/Kan2> could be isolated from *E. coli pir*-116 cells by using mini-prep protocols optimized for isolation of high copy number plasmids from *E. coli* (Fig. 3.4). Such easy isolation of

plasmid pPDL2::Tn5<R6Kγori/Kan2> facilitated for easy sub-cloning of its fragments in multipurpose vectors.

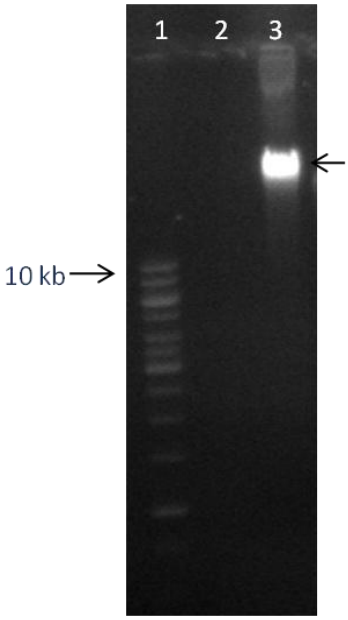
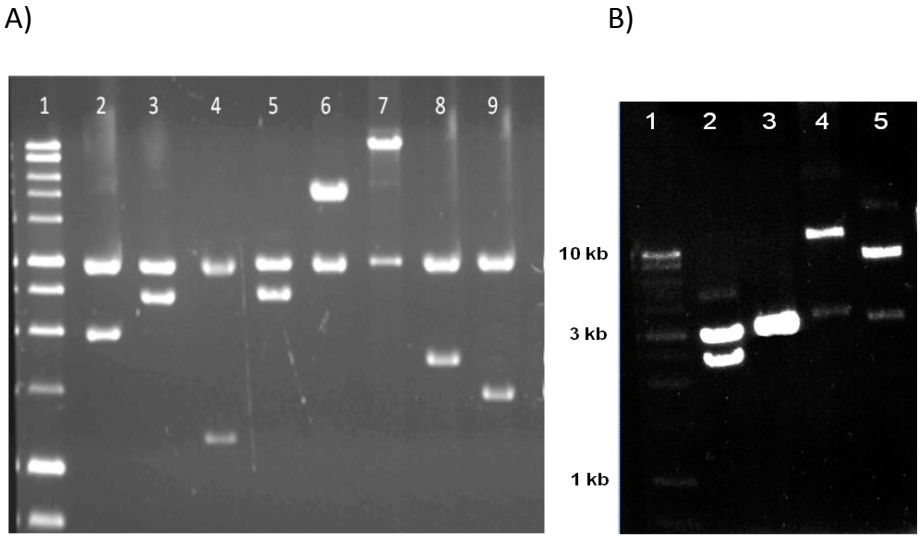


Fig. 3.4. Isolation of rescued plasmid pPDL2::Tn5<R6Kγori/Kan2> from *E.coli pir* 116 cells. Lane 1. Represents 1kb DNA ladder. Lanes 2 and 3 represent plasmid profile of *E.coli pir*-116 and *E.coli pir-116* (pPDL2::Tn5<R6Kγori/Kan2>). Plasmid pPDL2::Tn5<R6Kγori/Kan2> is shown with an arrow mark.

3.3. Sequencing of pPDL2 and annotation:

As stated before, 10 fragments were generated when plasmid pPDL2::Tn5<R6Kγori/Kan2> was digested with *Pst*I. All the 10 fragments with the following sizes 12 kb, 8 kb, 5.5 kb, 3 kb, 2.5 kb, 2.2 kb, 2.0 kb, 1.7 kb, 1.6 kb and 1.3kb were sub-cloned in pBluescript-II KS digested with similar enzymes (Fig. 3. 5). Similarly, 4 *Eco*RI fragments, with the sizes of 11 kb, 6.7 kb, 3.0 kb and 2.5 kb were cloned in pBluescript-II KS digested with *Eco*RI. Description of the recombinant plasmids containing different pPDL2 fragments used for sequencing is shown in Table 3.1.



3. 5. Sub-cloning of pPDL2. Lane 1 represents 1kb DNA ladder. Panel A shows shows sub-clones of pPDL2 generated by ligating *PstI* fragments in pBLuescriptII vector. Panel B shows similar sub-clones of *EcoRI* fragments. Refer table 3.1 for size description.

3.4. Sequence strategy

The detailed strategy used to obtain complete sequence of pPDL2 was shown in Fig. 3.6. Initially the sub-clones were directly used to generate sequence using universal forward and reverse primers. After obtaining plasmid pPDL2 specific internal sequence, fragment specific primers were designed to obtain entire sequence of the fragment through gene walking strategy. A detailed strategy used for obtaining the complete sequence of pPDL2 is shown in Fig. 3.6. Clones pE5 (2.2kb), pE12 (6.5kb), pP33II (12kb), pP4I (8kb), pP4II (5.5kb) and pP3II (5.7kb) were initially sequenced using vector specific primers and then were sequenced using primer walking strategy. Clone pP33II is 12kb in size. In order to reduce the size of the insert plasmid pP33II was further digested with both *PstI*, *SalI* and *Eco*RI and sub-cloned in pBluescript KSII as *PstI*-*Eco*RI (p33EP, 8kb) and *PstI*-*Sal*I (p33SP, 3kb) fragments. While obtaining the sequence of junction regions appropriate primers were designed and plasmids with overlapping fragments were used

Sub-clone of pPDL2	Description
pE10	11.0 kb *EcoRI* fragment of pPDL2 cloned in pBluescript II-KS
pE9	3.0 kb *EcoRI* fragment of pPDL2 cloned in pBluescript II-KS
pE5*	2.5 kb *EcoRI* fragment of pPDL2 cloned in pBluescript II-KS
pE12*	6.7 kb *EcoRI* fragment of pPDL2 cloned in pBluescript II-KS
pP33II*	12 kb *PstI* fragment of pPDL2 cloned in pBluescript II-KS
pP4I*	8 kb *PstI* fragment of pPDL2 cloned in pBluescript II-KS
pP3II*	5.5 kb *PstI* fragment of pPDL2 cloned in pBluescript II-KS
p33EP*	8 kb *Eco*RI-*PstI* fragment of pP33II cloned in pBluescript II-KS
p33SP*	4.5 kb *Sal*I-*PstI* fragment of pP33II cloned in pBluescript II-KS
pP1I*	1.5 kb *PstI* fragment of pPDL2 cloned in pBluescript II-KS
pP3I*	2.5 kb *PstI* fragment of pPDL2 cloned in pBluescript II-KS
pP7I*	2.2 kb *PstI* fragment of pPDL2 cloned in pBluescript II-KS

Table 3. 1. Details of *EcoRI* and *PstI* fragments of pPDL2::Tn5<R6Kγori-Kan-2> sub-cloned in pBluescript. Sub-clones used for sequencing are indicated with * mark.

as template to generate sequence reactions. The strategy followed to get the complete sequence was shown in Fig. 3.6.

3.5. Sequence assembly and analysis

Sequence assembly of pPDL2 was done using the contigexpress software of VectorNTI. The chromatograms were compared using contigexpress and contigs were created with chromatograms having good quality bases. All the contigs were then aligned using

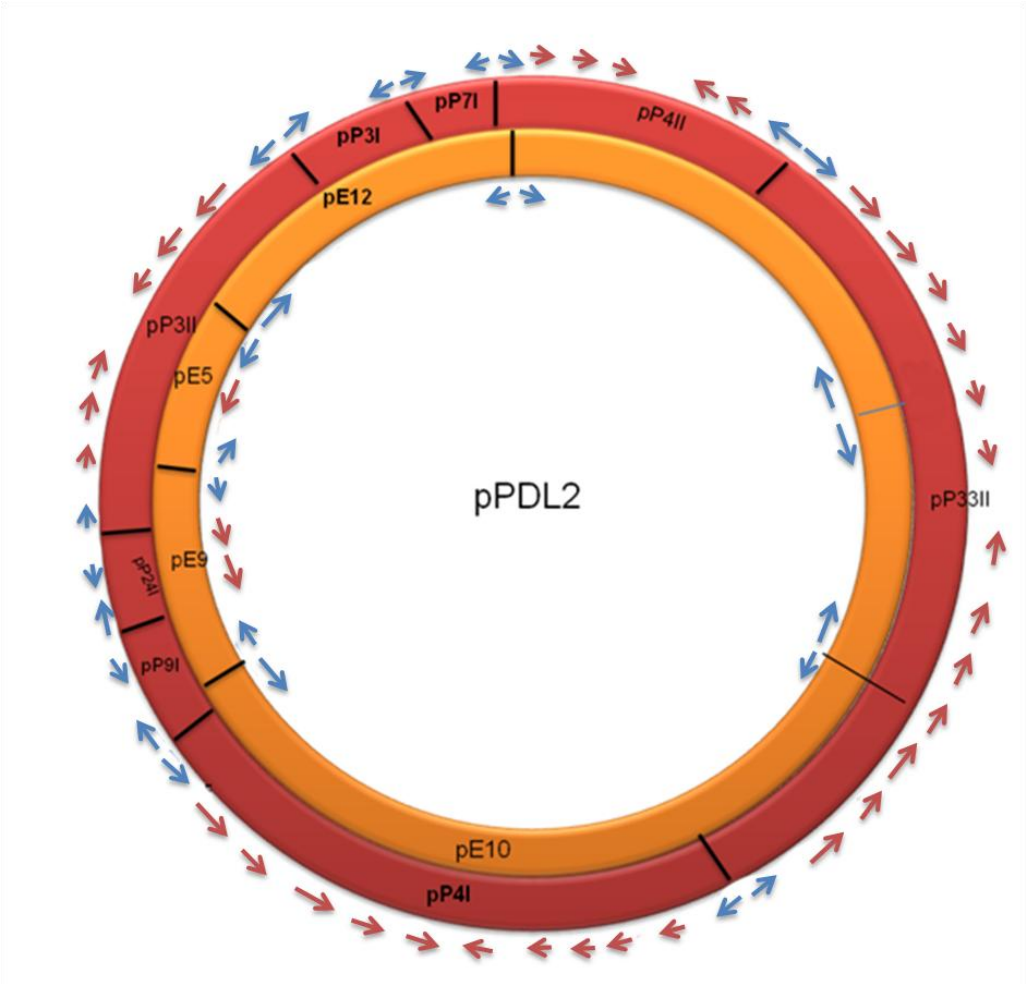


Fig. 3.6. Strategy used for sequencing of plasmid pPDL2 of *Flavobacterium* sp. ATCC 27551. The outer and inner circles represent *Pst*I and *Eco*RI fragments of pPDL2. Arrows indicate position of primers used for sequencing of pPDL2. Blue coloured arrows show vector specific primers. Orange and maroon coloured arrows indicate the primer positions used to sequence *Pst*I and *Eco*RI fragments cloned in pBluescript vector.

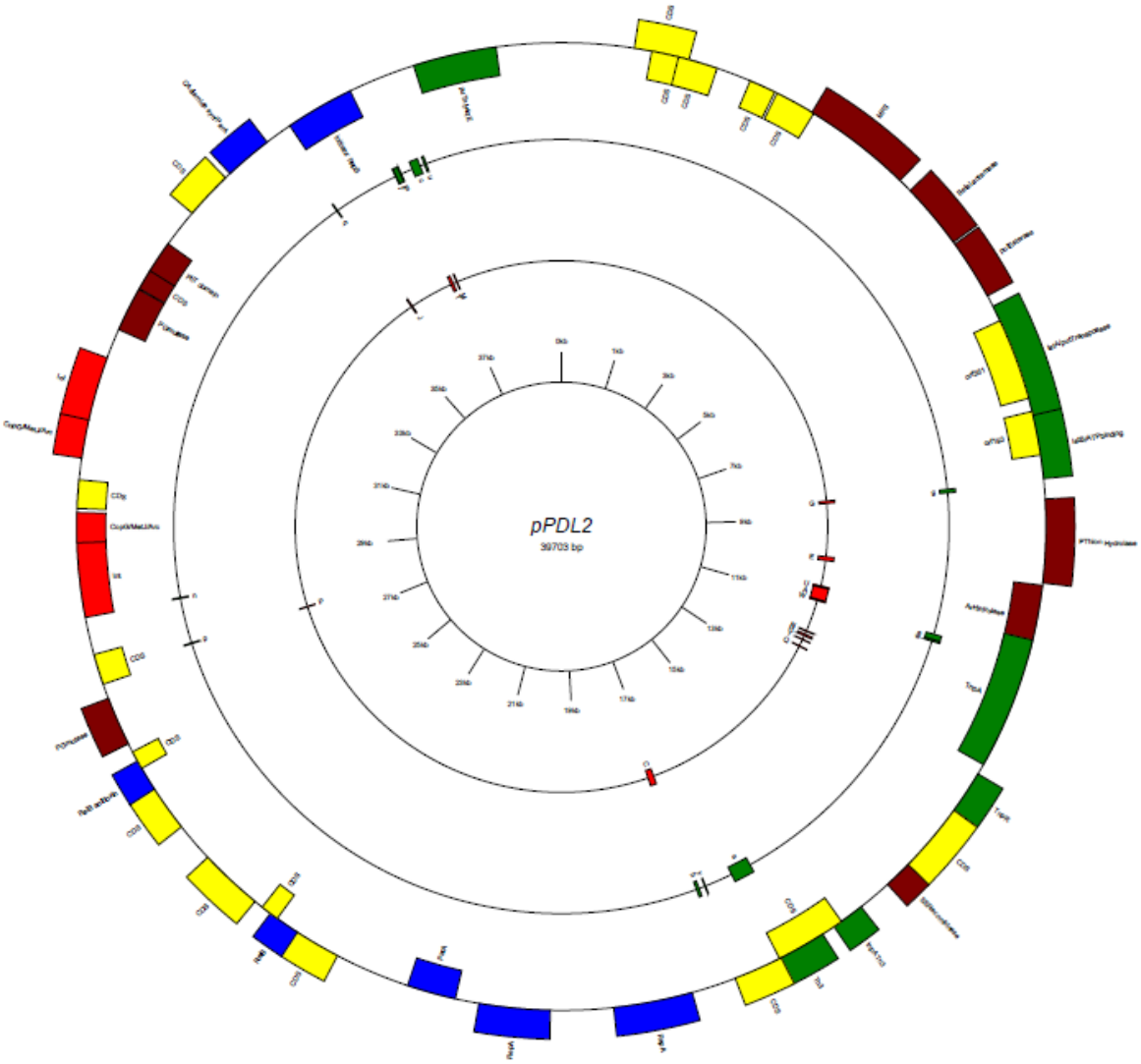
Primer Name	Sequence of the primer	Primer position
E5F	5'-TAAGGATAGTGGGACGTCGC-3'	3838 – 3819
E5R	5'-GTCGTGGGCGTCGTTAAGCTG-3'	2452 -2472
P3R	5'-TATCCTTGATGCCGAAGACC-3'	6509 – 6490
E13R	5'-GCCGGCAGGATATAGGTT-3'	28028 -28045
E13F	5'-CACCTCACCAGCAATTCGTA-3'	38306 -38287
E12F	5'-AGCCTGTTGACGCAGAAAGT-3'	4498 -4517
P33F	5'-GTTCGCGATCGTCAAGAACT-3'	29779 – 29760
P4I-Ext4	5'-GTTCTACAACACGCTGAAC-3'	38999 -39017
P4II-M	5'- GTCCGGGCTGATGAAATATG-3'	13915 -13934
P4IR4	5'-ACTCGCTGGCCTATGTGTTC-3'	32593-32574
P33-R4	5'-ATCCCGATCTGTTCATTTGC-3'	18240-18259
P4F	5'-ACAATTTCCAGGTCGTCACC-3'	15599 -15580
E12F-internal2	5-'GCCTCAATCTGGTGTTCGAT-3'	5157 -5176
E13F-internal2	5'-CGAATTGGTGGGATTTGTCT-3'	37795 -37776
E13R-internal2	5'-CCATCGCCTAGATCAACACC-3'	28576 -28595
P3RII-internal	5'-ACGATGTCGTCGATGTGTGT-3'	6248 -6229
p33FII-internal2	5'-CTGTTCCGCACGATCGCG-3'	31865 -31882
P33R-Internal	5'-TATCTCGCTCACCGGCGACT-3'	11523 11504
P4FII internal	5'-TGGCGTTGATCGGCTATG-3'	15341 -15324
P4RI external	5'-CACCAGGCCAACAAGAAATC-3'	30585 -30566
P4RI internal	5'-ATGGCCGATTGGCTGCTGGC-3'	29841 -29822
P4FI-Internal	5'-GCACTTTCGTGTAGTGACCCC-3'	16962 -16981
P4RI-Internal2	5'-ATCGCGCGCGGCACTAAGC-3'	31877 -31895
P4RII-Internal2	5'-AGATCCACCATCTATCGCGA-3'	13832 -13851
P3FII-Internal	5'-GAAGTCACCGAGGAGCACTT-3'	2209 -2228
H10_P33RII(External)	5'-CGGTGTAGTGCTCCTCGATT-3'	16219 -16200
D07_P4FI(External)	5'-CTTTGACTTCATCCGGCAGT-3'	37893 -37912
G09_P4FII(External)	5'- AGTTGTCGCATTCTCGATCC-3'	15852 -15871
E09_P3FII(External)	5'- GCACGATGTTCTTCGACCTT-3	1886 -1867
P33SPF-E	5'- CGCTGAATCTGAACTGACGA-3'	21983 -22002
p33F2-g	5'-CGGCTTTCGGCATCCAACCT-3'	28414 -28395

Table 3. 2. Primers used for sequencing of plasmid pPDL2 and their sequences

Pairwise alignments to get the complete sequence of pPDL2. The complete sequence of pPDL2 is available in web page (www.uohyd.ernet.in/uploads). After generating the complete sequence it was analyzed to indentify GC ratio, number of ORFs, inverted and direct repeats, promoter elements and other *cis*-elements that play a predominant role in integration excision and mobilization of plasmids.

3.6. The GC composition

The 39.75 kb sequence of pPDL2 has shown high similarity to either chromosomal or plasmid DNA sequences of *Sphingobium* and *Sphingomonas* sp. Since, plasmid pPDL2 has shown sequence similarity to the genome sequence of *Sphingomonas* and *Sphingobium,* the GC content of the total genome sequences were obtained from genome database. Based on the total genome sequence found in database (www.**ncbi**.nlm.nih.gov/) the GC content of *Sphingomonas wittichi* RW1 is 68.4%. It contains two indigenous plasmids designated as pSWIT01 and pSWIT02 which have a GC content of 64.1% and 61.2% respectively. Similarly there are two chromosomes 1 and 2 in *Sphingobium japonicum* UT26S and each of them have a GC content of 64.8 % and 65.9 % respectively. In addition to these two chromosomes there are three circular plasmids in *Sphingobium japonicum* UT26S (NBRC101211). These three plasmids designated as, pCHQ1, pUT1 and pUT2 have a GC % of 63.0%, 63.7% and 61.0% respectively (NBRC101211). The G+C content of pPDL2 of *Flavobacterium* sp. ATCC 27551 is 61.76% and found to be very close to plasmids pSWIT02 of *Sphingomonas wittichi* RW1 and pUT2 of *Sphingobium japonicum* UT26S. Consistent with GC content, the proteins coded by the open reading frames (Table 3. 3) of pPDL2 have shown homology to the proteins coded by the *Sphingomonas*. The 39.75kb plasmid pPDL2 codes for 41 open reading frames. Out of 42 predicted ORFs 18 of them are hypothetical proteins that show (31-95%) homology to the hypothetical proteins of *Sphingomonas* or *Sphingobium* sps. (Table 3.3, Fig. 3.7).



.3. 7. Circular map of plasmid pPDL2. The ORFs specified by sense or positive and antisense or negative strand are shown in outer and inner circles respectively. Colour boxes represent ORFs. Very distinct colours are used to indicate possible function of ORFs. Yellow, for all hypothetical proteins; dark green for transposons and IS elements; Dark blue, for proteins involved in plasmid replication and partition; Red to show phage integrases and CopG family transcription factors and brown colour for all ORFs which have accessory functions.

The ORFs identified in plasmid pPDL2 are given generic, conventional and functional names. The generic name indicates the serial order of the ORF on plasmid pPDL2. In the conventional names the ORF is followed by a number which is equivalent to the number

amino acids encoded by the *orf*. If any *orf* coded protein has significant homology to the functionally characterized proteins it is also given a functional name (Table 3. 3). While describing ORFs coded by plasmid pPDL2, they are divided into functional modules such as Replication and partition module, Mobilization module, Integration module, Degradation module and mobile genetic elements. Each functional module is independently described to facilitate easy description and understanding of plasmid pPDL2 sequence.

3.7. Replication and partition module

In general, replication module includes a well defined *oriV*, replication initiator protein, RepA, proteins involved in partitioning (Par) and plasmid maintenance (toxin-antitoxin modules).

3.7.1. Replicative origin (*oriV*)

The *oriV* generally contains sequence motifs (*cis*-elements) that interact with the replication initiator protein, RepA and other accessory proteins. In plasmids replicated through *theta* mode (Bramhill and Kornberg, 1988).They include (i) AT-rich region containing sequence repeats, often found to be located adjacent to RepA binding sites. At this AT-rich region of *oriV* the host initiation factors assemble immediately after its conversion into an open complex, and (ii) the next important *cis*-element found at origins of *theta*-replicating plasmids is DnaA binding box (Bramhill and Kornberg, 1988; Kornberg and Baker, 1992). Generally one or more *dnaA* boxes are found at the replicative origin of plasmids. Interaction of host encoded DnaA initiator protein is an essential event in initiation of plasmid replication process (Bramhill and Kornberg, 1988; Kornberg and

Baker, 1992). In pPDL2, two ORFs designated as *orf20* (18125-19168) and *orf21* (19995-20933) have shown very high homology (96%) to the RepA-like protein coding sequences of *Sphingobium japonicum* UT26S (Fig. 3.8). These two RepA sequences designated as

```
A)  pPDL    MKPRIEPDRQRNPELPLAVNETPNGAEGRPPLLPVRYPDPDLFICDVLDAIPKDDMASME 60
    UT26    MKPRIEPDRQRNPELPLAVNETPNGAEGRPPLLPVRYPDPDLFICDVLDAIPKDDMASME 60
            ************************************************************

    pPDL    HPIFSLATKPDRRVFRYEHNGNKLEIVPSVKGLATIHDKDILIYCISQLIGKMNQGERPS 120
    UT26    HPIFSLATKPDRRVFRYEHNGNKLEIVPSVKGLATIHDKDILIYCISQLIGKMNQGERPS 120
            ************************************************************

    pPDL    RTLHLTARDLLVWTNRQTDGDGYDRLRSAFERLSGTRITTNIKADGEEITEGFGLINEWR 180
    UT26    RTLHLTARDLLVWTNRQTDGDGYDRLRSAFERLSGTRITTNIKADGEEITEGFGLINEWR 180
            ************************************************************

    pPDL    IVRQTRSGQMSEIKVTLSDWLFKMVEGRSVLTLHRDYFRLRKPLERRIYELARKHCGAQE 240
    UT26    IVRQTRSGQMSEIKVTLSDWLFKMVEGRSVLTLHRDYFRLRKPLERRIYELARKHCGAQE 240
            ************************************************************

    pPDL    KWSISVETLQKKTGASSHLRVFRSMLRDLVAHDHLPDYAVEMNGDTVTFRNREALDTVEA 300
    UT26    KWSISVETLQKKTGASSHLRVFRSMLRDLVAHDHLPDYAVEMNGDTVTFRNREALDTVEA 300
            ************************************************************

    pPDL    IEAEPERPYIDPESFHDAQSVAPGYDVYALYDQWVSW----------------VGSTAGG 344
    UT26    IEAEPERPYIDPEGFHDAKSVAPGYDVYALYDQWVSWWIDSGRPELKSPRAAFIGFCRNK 360
            *************.****:******************                :*   .

    pPDL    RET---- 347
    UT26    HKTAPLR 367
            ::*

B)  pPDL2   AACAAAAAACGTGCCGGCGAACCCGGCTCATCATGGGGGATGCCGGGTGGGAGGATCGCT 60
    pUT1    AACAAAAAGCGTGCCGGCGAAGCCGGCTCATCATGGGGGATGCCGGGTGGGAGGATCGCA 60
            ******** ************ ************************************

    pPDL2   TTAGCGATCCGGGGCGGCCGAAGGCCGCGAGCCGGCGGGGGCGAAGCCCCCAAGAGGCGG 120
    pUT1    A-AGCGATCCGGGGCGGCCGCAGGCCGCGAGCCGGCGGGGGCGAAGCCCCCAAGAGGCGG 119
              ****************** ****************************************

    pPDL2   CGCTTTTCTTCCTTCTGCCTGTGCCTGAAGCGCGGCGCCAGTGACTTGCGCTGCCTCTGT 180
    pUT1    CGCTTTTCTTCCCTCTGACCGTGCTTCAAGCACTGCGCCAGTGAGTCGCGCTGCCTCTGT 179
            ************ **** * **** * **** * ********** * *************

    pPDL2   CCCGATTTTCGGGACAAGCTGTCCGATCTGGCGGGACGCTGGTGGGTGAAACCACGACGC 240
    pUT1    CCCGATTTCCGGGACAGGTTGTCCGATCTGGCGGGATGCTGGTGGGTGAAACCACGTCGC 239
            ******** ******* * ***************** ******************* ***

    pPDL2   GGCAAAAGTTATCCACAGGTCGGTCTAGTGTTAATATATGGGTTAAAGAGTCGTGTTACG 300
    pUT1    GGCAAAAGTTATCCACAGGGCGGCCTAGTGTTAATATATGGGTTAAAGAGTCGTGTTACG 299
            ******************* *** ************************************

    pPDL2   GGATTTTGGCGGTTCCGTAAGTTACTGAAAATAGGTATAGAATCGGGCGTTTTTGGGCCT 360
    pUT1    GGATTTTGGCGTCTCCGTAAGATACTGAATATAATAGTGAAATCGGCCATTTTTTGGCCT 359
            ***********  ******** ******* ***    *  ****** * ***** *****

    pPDL2   TTGGCGTCAACGGAACCCCGAATCTGGTGACTCCAGAACCCCGAATCTGCGTCAACGAAA 420
    pUT1    TGGGCGTCAACGGAACCCCGAATCCGGTGACTCCAGAACCCCGAATCTGCGTCAACGAAA 419
            * ********************** ************************************

    pPDL2   CCCCGAATCGTGGGCAGGCGTCAATGAAACCCCGAACTTTTCCGTTGCGTCAATGAAACC 480
    pUT1    CCCCGAATCGTCGGCGAGCGTCAATGAAACCCCGAACTTTTTCGTTGCGTCAATGAAACC 479
            *********** ***  ************************ ******************
```

Fig. 3. 8. Pairwise alignment of A) RepAa of pPDL2 of *Flavobacterium* sp. ATCC 27551 with RepA of pUT1 of *Sphingobium japonicum* UT26S and B) alignment of *oriV* sequences of pPDL2 and pUT1 of *Sphingobium japonicum* UT26S. Identical residues and gaps are shown with an (*) and (-) respectively. Similar residues are shown with (:) and (.) symbols.

RepAa and RepAb share 100% similarity. Though the reason for presence of two identical *repA* coding genes are unknown, considering its homology with well characterized RepAa protein of pUT1 of *Sphingobium japonicum* UT26S, the possible role of these two proteins in replication of plasmid pPDL2 is quite apparent. Since existence of two RepA proteins appeared unusual, a complete literature search was undertaken to gain more insights into this unusual phenomenon. As revealed by literature search, existence of more than one RepA is not uncommon. In an 184 kb indigenous catabolic plasmid pNL1 of *Sphingomonas aromaticivorans* F199 more than one *repA* genes were identified (Romine et al, 1999). However, no homology was seen between RepA of plasmid pPDL2 and RepA of pNL1 of *S. aromaticivorans.* When the upstream region of *repA* sequence was analyzed to indentify existence of promoter elements, a σ70 promoter was seen upstream of each *repA* gene, suggesting existence of two functional *repA* genes on plasmid pPDL2.

In order to assign incompatibility group to plasmid pPDL2, a two way approach was followed. Initially, the RepAa sequence was blasted to know its homologues from well characterized plasmids. The second approach was to identify similarity between the *oriV* sequence of pPDL2 and well characterized plasmid replicative origins available in the database. The RepAa sequence of pPDL2 has shown 96% identity to the RepA sequence of *Sphingobium japonicum* UT26S plasmid, pUT1 (BAI99177). Further, it has also shown about 85% identity to the plasmid, pAPA01-030 coded RepA of *Acetobacter pasteurians* IFO3283-01 and 62% identity with RepA of *Nitrospira multiformis* ATCC25196 (Fig. 3.9). As seen in phylogenetic tree constructed using blast output, no significant homology was found between RepA of pPDL2 to other bacterial RepA proteins (Fig. 3.9). However, the

plasmids of *Sphingobium* sp. to which *repA* of pPDL2 has shown strong identity are yet to be assigned with a distinct incompatibility group (AP010806).

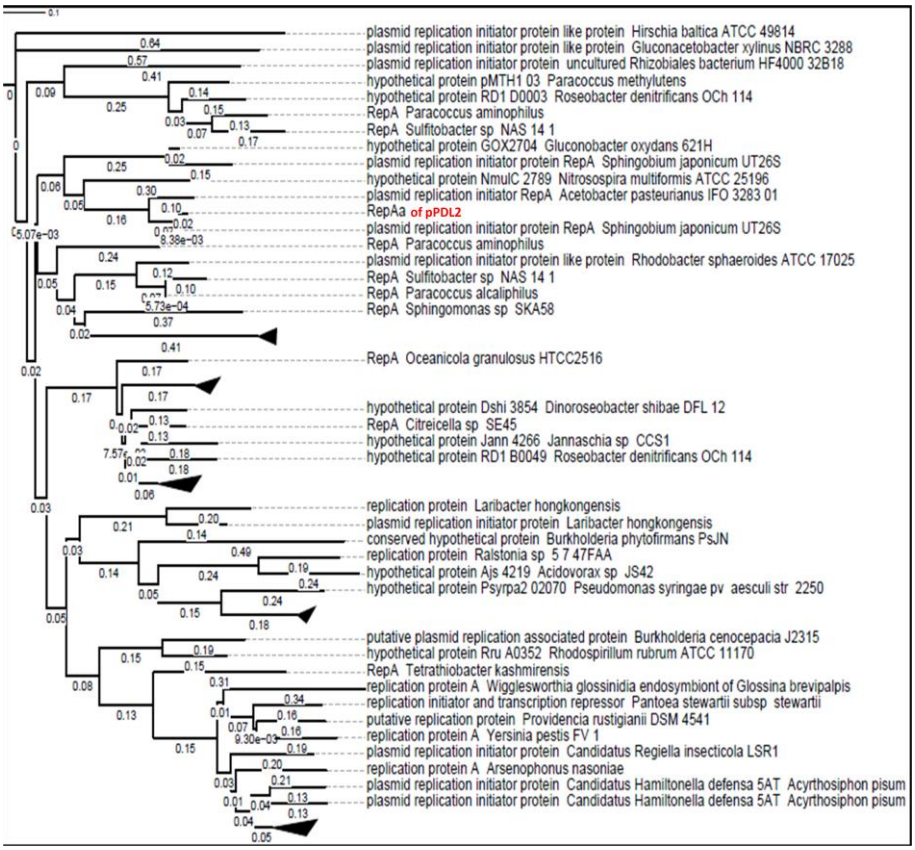


Fig. 3. 9. Phylogenetic analysis of RepAa of pPDL2 of *Flavobacterium* sp. ATCC 27551.

3.7.2. Origin of replication (*oriV*)

A 776 bp long putative *oriV* like sequence was seen immediately upstream of *repAa, spanning from sequence position* 17238 to 18114. In general the replicative origin is predicted based on the GC-skew analysis as described in materials and methods section (Grigoriev, 1998). As RepA protein present on plasmid pUT1 of *Spingobium japonicum* was 96 % identical with RepA protein of pPDL2, a thorough search was done to find out a sequence that has similarity with the replicative origin of plasmid pUT1 of *Sphingobium japonicum* UT26S. As shown in Fig. 3.8B considerable similarity was found between *oriV* sequences of pPDL2 and pUT1. Only very minor differences were noticed

between these two *oriV* sequences. The *oriV* predicted in pPDL2 has two imperfect tandem repeats. The first tandem repeat consists of 24bp sequence designated as A-24 and A'-27bp. There is a gap of 107 bp sequence between A-24 and A'27. The second tandem repeat is C-48 (48 bp) and C'-54 (54 bp) in size and are separated by a gap of 26 bp. In between these two tandem repeats a 17bp long palindromic sequence was identified (Fig. 3.10B). Interestingly, sequence found downstream of palindromic sequence is AT rich, while the upstream region is highly GC rich. In the upstream region of the palindromic sequence four typical DnaA binding boxes were identified with a consensus sequence of 5'-TTN4ACA-3' (Fig. 3.10C). The replicative origins of plasmids are shown to have configuration conservation (spatial arrangement of repeats) rather than showing strict sequence conservation (Gloria del Solar et al, 1988). When searched to find such configuration, homologues to *oriV* of plasmid pPDL2 the *oriV* of plasmid pSC01 isolated from *Pseudomonas* has revealed to have similar spatial arrangements of repeats and palindromic sequences (Fig. 3.10D). However, neither plasmid pUT1 nor pSC01 are assigned any compatibility group. Therefore, with the present data no clear incompatibility group can be assigned to the plasmid pPDL2. As shown in Fig. 3. 10A, the other important proteins that contribute for plasmid replication and maintenance are RepB, ParA and RelB. The *orf40* has shown 83% homology to replication initiator protein, RepB of *Gluconobacter diazotrophicus* PA15 (Fig. 3. 11A). RepB proteins are only seen in plasmids that replicate through rolling circle (RC) mode. The RC mode of DNA replication

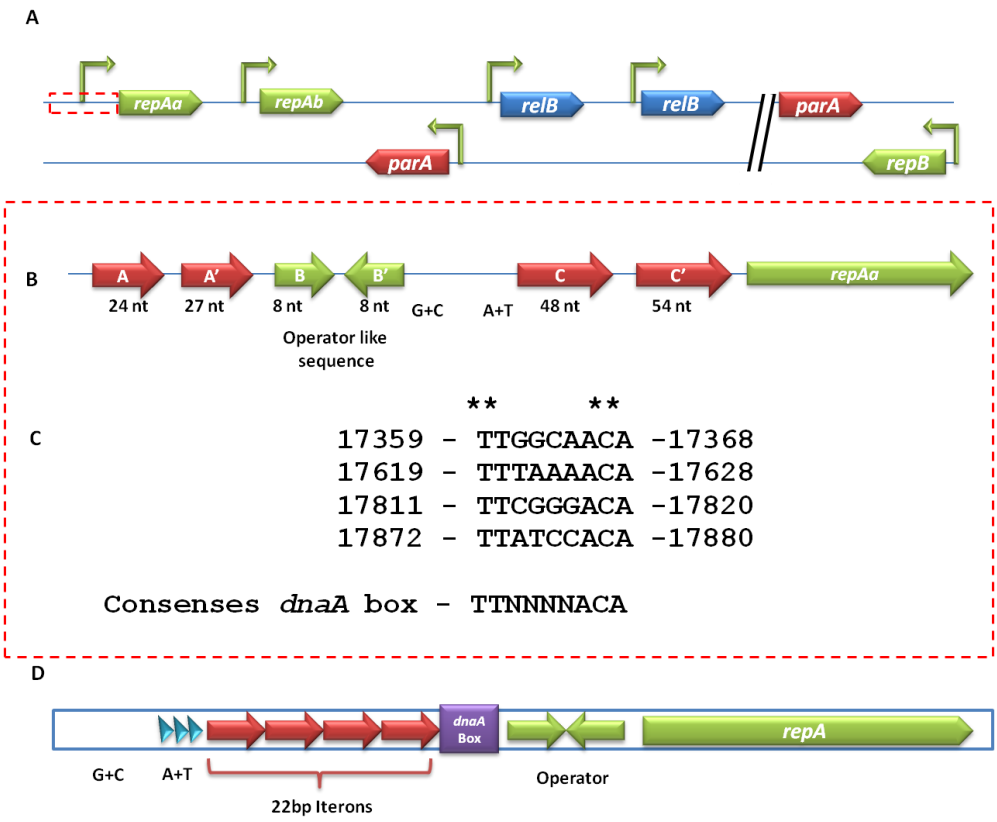


Fig. 3.10.Panel A shows physical map of plasmid pPDL2 DNA region showing the organization of ORFs coding for proteins involved in plasmid replication (*repA*, *repB*), partition (*parA*) and maintenance (*relB*). Panel B shows structural configuration of *oriV* found upstream of *repAa* in pPDL2 of *Flavobacterium* sp. ATCC 27551. In Panel B, the organization of tandem repeats A and A' found upstream of palindromic sequence (shown in green) also designated as operator sequence. The second tandem repeat sequence, C and C' found downstream of operator sequence is shown with maroon coloured arrows. The *repAa* is shown in green. Panel C. represents the alignment of predicted *dnaA* boxes found in the replicative origin *oriV* of plasmid pPDL2. The consensus *dnaA* box is shown separately. The configuration of *oriV* of pPS10 isolated from *Pseudomonas savastanoi* that show similarity to *oriV* of pPDL2 is shown in panel D

is generally seen in small plasmids (less than 10kb). The second place where such origin of replication is seen during conjugation process. In such plasmids, the replication initiators sequences are designated as the Mob class of initiators (Ilyina and Koonin, 1992).

A)

```
pPDL        MNHATSPVNGGKAKVALDGDTALTLAQKGRGNPFDPANYGEIVKPGELVDIVELSPLTLA 60
GPA1        ------------------------------------MNFG-------------------- 4
```

```
SPCl        MRVAAALQAKGGDEFAKPGSIVEVKFVKGQSLSLTASRLLALMILTAG------------ 48
                                                 .

pPDL        DRRIYNLLIANAWERIGEPVIHRIPKSALKGTHQGNERIESSLLRLMGTIAIVTIRKG-G 119
GPAl        ---------------FNEPIVS---------YAGGNERIEASLLRLMGTIAIVTIRKD-G 39
SPCl        -------------GDAWEDRPHKMRKADIRRGHKGNERISDMLEELHRTLFAVDDKSWRG 95
                                *                *****.  * .*  *:  *  :.  *

pPDL        KSFKRRVQLLGPSDESLEKD-----GFLHYRIPEELIEILRNSEVYARLKTQVMYCFESK 174
GPAl        KSYKRRVQLLGPSDESLEKD-----GFLHYRIPEELIEILQNSEVYARLKTQVMYCFESK 94
SPCl        KKATLRFSLISSSREEAEDEEGADAGWIEWEFTPEARKLIQESETYAVLNRQAVLGFRST 155
            *. . *..*:..* *. *.:     *::.:.:. *  :::::**.** *: *.:  *.*.

pPDL        YALCLYEMIERRIGLEYKQSEEFTIAELRGLLNVPEGKLERFADFNKYCLKVAQEEINKL 234
GPAl        YTLCLYEMTERRIGLEYKQTEEFTIEELRGLLNVPDGKLERFADLNKYCLKVATEEINKL 154
SPCl        YALKLYEIGALRLHRRQ-SLWKGDMTALRALLGIAPDVYKDFAQLRRKVLEKAKAEIDQL 214
            *:* ***:   *:  .  .  :  :  **.**.:. .  : **::.:  *: *  **::*

pPDL        CPFWVEFTPIKKGRKVERVSMMWLPKTMSGRRDAQNLIDQHSIVRRAKLRGDIPEMPVLV 294
GPAl        CPFYVDFSPIKKGRKVERIAFHWFPKTSSGKRDAQILIDQHRIVRRAKLRGLAAELPLLL 214
SPCl        AHFRVEWREIRQGRTVTEIEFRFEPKDAPAQIATVDEIGRHSAGRKARREDEVETVAVEA 274
            . * *::  *::**.* .: : : **  ..:  :   *.:*   *:*: ..    :.:

pPDL        DFSAPAAQR---------------------------------------------------- 303
GPAl        DFGTEPPEK---------------------------------------------------- 223
SPCl        VTQAAVAALVSKDKAGAGEVTFPNGTIRFGSDTLAAIGRSAGGGWDIDLIADAYRAQMGE 334
               :  .
```

B)

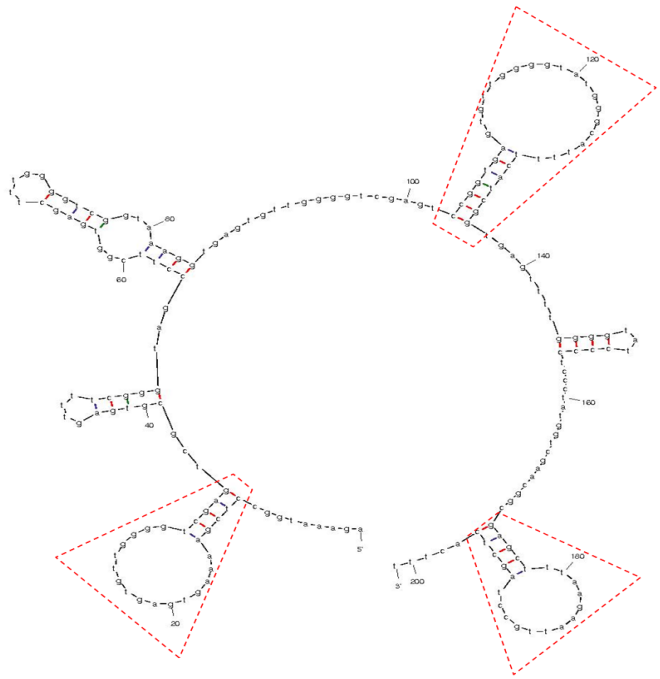


Fig. 3.11. Panel A shows multiple alignment of RepB of pPDL2 of *Flavobacterium* sp. ATCC 27551 with RepB of *Gluconacetobacter diazotrophicus* and *Sphingobium chlorophenolicum.* Identical residues and gaps are shown with an (*) and (-) respectively. Similar residues are shown with (:) and (.) symbols. Panel B shows Secondary structure of putative rolling circle replication origin found in upstream of *repB* sequence of pPDL2. The hair-pin and stem loop structures found to be typical structural features of rolling circle replication origin are highlighted with red colour.

In general, the plasmids replicating through rolling circle replication usually have two origins, the double stranded origin of replication and a single stranded origin of replication with a potential to form a cruciform and a hair-pin like structures respectively (del Solar et al, 1998). In well characterized streptococcal rolling circle replication (RCR) plasmid pMV158, the RepB is a homo hexamer and usually binds to a region in the origin known as 'Bind region' which is characterized by presence of direct repeats (de la Campa et al, 1990; Ruiz-Maso´ et al, 2004). In pPDL2, a sequence that shows strong structural similarity to the consensus RC replication origin was identified immediately upstream of RepB. Like in typical RC replication origin this region contains 4 direct repeats each measuring a length of 22 bp with a consensus sequence of 5'-ACCCCAACACTCACCGGACTCG-3' and a sequence with a potential to form a secondary structure (Fig. 3.11B). Its typical structural features and strategic location upstream of RepB suggest a role in replication of plasmid pPDL2. RC mode of replication is only seen in small plasmids (del Solar et al, 1998). In the background of such reports, further experimentation is required to validate its involvement in replication pPDL2. RC mode of replication is also seen during horizontal mobility of plasmids through conjugation (Lanka and Wilkins, 1995). Therefore, its involvement along with RepB in generating a relaxosome required during conjugation process cannot be ruled out. Mobilizable nature of plasmid pPDL2 (described in chapter-2) adds strength to such proposal.

3.7.3. ParA locus

In addition to the replication initiator proteins, pPDL2 has an ORF designated as *orf*22 (21246-21893c) immediately downstream of *repAb* gene. The 215 amino acids long protein coded by *orf22* has shown 99% homology to the partitioning protein ParA of

Sphingobium japonicum UT26S (BAI99179) (Fig. 3.12). Therefore, *orf22* is henceforth designated as *parA* gene. In the light of such high degree homology with the well characterized ParA protein it appears to involve in partitioning of pPDL2 soon after its replication. Interestingly, there exists another *parA* homologue, *parAa* immediately upstream of *repB*. However, it has also shown 58% identity with cobyrinic acid ac-diamide synthase of *Thauera* sp. M21T (ACK55109) and 57% identity to ParA of *Laribacter hongkongensis* (ABC70161). With the existing information it is not possible to assign a confirmed role of this gene in the maintenance of pPDL2.

```
pPDL            MKVLAILSQKGGVGKRAPGTTLATCLAVAAEQAGKVAAIIDLDPQATASFWKDVRQLDTP 60
SJUT            MKVLAILSQKGGVGK----TTLATCLAVAAEQAGKVAAIIDLDPQATASFWKDVRQLDTP 56
                ***************    ****************************************

pPDL            AVASIQPVRLPAMLKACEDAGTDLVVIDGAAVARDVAYEAARQADFILIPTKTAVFDTMS 120
SJUT            AVASIQPVRLPAMLKACEDAGTDLVVIDGAAVARDVAYEAARQADFILIPTKTAVFDTMS 116
                ************************************************************

pPDL            MTHTLDVVRQLDRAFAVVLTFVPPQGQETGDAIQAVAELGATVCPVTIGNRKAFFRAQAA 180
SJUT            MTHTLDVVRQLDRAFAVVLTFVPPQGQETGDAIQAVAELGATVCPVTIGNRKAFFRAQAA 176
                ************************************************************

pPDL            GQAVQEFEPHGPAADEIHRLYEYTTIRLYNEAEAA 215
SJUT            GQAVQEFEPHGPAADEIHRLYEYTTIRLYNEAEAA 211
                ***********************************
```

Fig. 3.12 .Pairwise alignment of ParA of pPDL2 of *Flavobacterium* sp. ATCC 27551 (pPDL) with ParA of *Sphingomonas japonicum* UT26S (SJUT). Identical residues and gaps are shown with an (*) and (-) respectively. Similar residues are shown with (:) and (.) symbols.

3.7.4. Toxin antitoxin module

Toxin-antitoxin systems (TA systems) increase the plasmid prevalence (number of plasmid containing cells/total number of cells) in growing bacteria populations by selectively eliminating daughter cells that did not inherit a plasmid copy during the process of cell division (Gerdes et al, 1986; Jaffe et al, 1985). This post-segregational killing mechanism relies on the differential stability of the toxin and antitoxin. Usually the toxin is stable which is rendered inactive by an unstable anti-toxin (Tsuchimoto et al,

1992; Van Melderen et al, 1994). In daughter bacteria devoid of a plasmid copy, because TA proteins are not replenished, the antitoxin pool rapidly decreases, freeing the stable toxin. These plasmid-free bacteria will eventually be killed by the deleterious activity of the toxin. Plasmid-encoded TA systems are also called addiction modules (Yarmolinsky, 1995) since this property renders the cell addicted to antitoxin production and therefore to the TA genes.

RelE and RelB is a toxin and antitoxin pair coded by the *relBE* toxin-antitoxin gene family which are discovered in *Escherichia coli* (Gotfredsen and Gerdes, 1998). The *relE* gene encodes a small (11-kDa) protein that is extremely toxic to bacterial cells, and the *relB* gene encodes an antitoxin of similar size that counteracts the cell killing activity of the RelE toxin (Gotfredsen and Gerdes, 1998; Grøndlund and Gerdes, 1999). Many *relBE* homologues have been identified in a broad range of both gram-negative and gram-positive bacteria and in archaea (Gotfredsen and Gerdes, 1998; Grøndlund and Gerdes, 1999).

3.7.5. RelB of plasmid pPDL2

Upstream of the *parA* gene, an ORF designated as *pilT* (33297-33722c) is identified. This ORF codes for a protein having 141 amino acid long PilT domain. The pilT domain of Orf38 shows 59% homology to PilT domain containing protein of *Sphingomonas wittichi* RW1. Monomers of PilT domain containing proteins have the ability to polymerise while forming Pilus fibre (Wall and Kaiser, 1999). Similarly, the PilT protein in its hexameric conformation is required for ATP-dependent retraction of the type IV pilus in gram-negative bacteria (Aukema et al, 2005). Retraction of type IV pili mediates intimate attachment and signalling to the host cells, surface motility, biofilm

formation (Chiang and Burrows, 2003; O'Toole and Kolter, 1998), natural transformation (Wolfgang et al, 1998; Whitchurch et al, 1994), and phage sensitivity. In the hexameric state, the ATPase activity of PilT, could actively promote dissociation of pilin monomers from the base of the pilus filament and thus has been shown to contribute to the pilin monomers pool observed within the cytoplasmic membrane (Morand et al, 2004). PilT could remove or inactivate a capping protein that prevents an energetically favourable retraction reaction. Alternatively, PilT could reverse the direction of the PilB motor, whose ATPase activity is required *in vivo* for the assembly of pilus filaments from pilin monomers (Turner et al, 1993). The exact function of the PilT domain in some proteins is unknown but this domain is present in some toxin proteins involved in bacterial plasmid stability such as the VapC (Francuski and Saenger, 2009; Robson et al, 2009). The exact nature of PilT domain containing protein of pPDL2 is not clear. However, in plasmid pPDL2, two ORFs, *orf20* (23492-23922) and *orf30* (26316-26582) encode for an antitoxin protein which shows 96% similarity to RelB, anti-toxin of *Sphingobium japonicum* UT26S. In the toxin - anti-toxin pair on pPDL2, absence of the toxin and the presence of pilT domain containing protein in plasmid pPDL2 suggests possible role of PilT domain containing protein as a toxin.

3.8. Mobilization module

3.8.1. Origin of Transfer

In Plasmid pPDL2, genes responsible for mating pair formation are not noticed. However on careful examination, genes responsible for initiation of its mobilization are identified on the plasmid. The initiation of transfer occurs through formation of relaxosome involving *oriT* and relaxase. The relaxase creates a nick at the *oriT*. In the

preceding sections existence of *repB* gene on plasmid pPDL2 is mentioned. As discussed before, the RepB creates a single stranded nick to initiate replication through RC mode of replication. In view of existence of RC replicative origin in its upstream region, it was implicated either in replication or in horizontal gene transfer (HGT) of pPDL2. RC mode of replication is seen in mobilizable plasmids and it is initiated from a sequence designated as *oriT* (Pansegrau and Lanka, 1991). The well characterized *oriTs* are characterized by a GC rich region with a stem loop structure (Lee and Grossman, 2007). Such GC rich region is present in the intergenic region of *repAa* and *repAb* (Fig. 3.13) This GC rich sequence (5'-AAAAAA**CCCCCCC**ACAAGAGA**GGGGGGG**GGGGG-3') has a potential to form a stem loop structure and possesses typical features of an *oriT* sequence (Fig. 3.13). Existence of RepB coding sequence along with putative *oriT* and RC replicative origin adds strength to the prediction of pPDL2 as a mobilizable plasmid. Such predictions have in fact formed basis for conducting experiments described in chapter II to show that plasmid pPDL2 is a mobilizable plasmid.

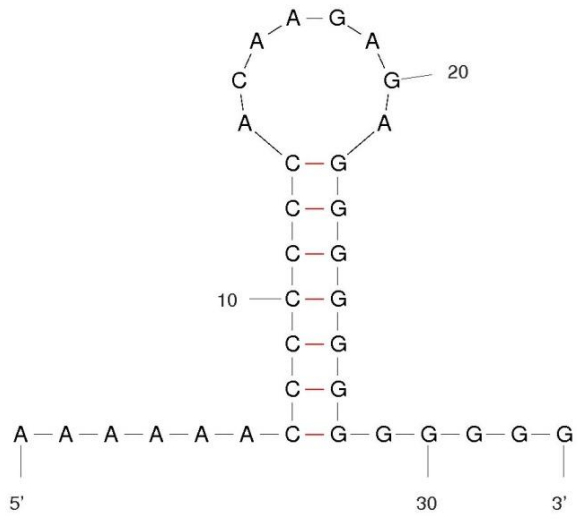


Fig. 3.13. Secondary structure of predicted *oriT* present in the intergeneic region of *repAa* and *repAb* in plasmid pPDL2.

3.9. Integrase module

An interesting observation in the entire sequence of pPDL2 is organization of integrase module. Integrase module contains two units, each consisting of *copG, int* and *pgm* genes. These units designated as CIP-I and CIP-II stretches from nucleotide position 26837 to 33015. Out of these three genes, both *copG* and *int* appears to be co-transcribed. The *copG* of CIP-I unit starts from 29962 nucleotide with a start codon ATG and ends with the presence of a stop codon present between nucleotide position 29564-29567. The stop codon of *copG* overlaps with the start codon of integrase. Integrase of CIP-I module is a 328 amino acid long protein and shows strong homology (86%) to phage integrase protein of *Sphingobium japonicum* UT26 and 70% homology to *Pseudomonas syringae* DC 3000 (Fig. 3.14).To gain further insights into the function of integrase, the protein coded by integrase

```
Int1            MNELAPLPPPPSSSALALPALVASADEAARLRFLEFFAVTIRNPHTRRAYMRAAGEFLAW 60
UT26            MNQLAPLPSP----ALVLPALIAAADERARLRFLEFFAVTIRNPHTRRAYARAAGDFLAW 56
SRW1            --------------------MIAAADDATRLRFLEFFAVTIRNPHTRRAYARAAGDFLAW 40
Bps             ------MNQIVRSSSATMPALVTAAGERAGVRFLEFFASAIRNPHTRRAYARAAEDFFAW 54
Psyr            ------------------------------MRFLEFFTANIRNPNTRRAYARATQEFLTW 30
                                              :******:  ****:***** **: :*::*

Tnt1            CEARGVASLAGVQPLHVAAWIEAQGGELAPPSVKQQLAGVRSLFDWLVMGQVVPANPAAS 120
UT26            CEARGVASLAGVQPLHVAAWVEALGRELAAPSVKQQLAGVRHLFDWLVTGHIVPVNPAGS 116
SRW1            CEARGVASLAGVQPLHVAAWVEALGRELAAPSVKQQLAGVRHLFDWLVTGHIVPVNPAGS 100
Bps             CANMGVTSIVAVQPLHVAAWVELQTQTLSAPTVKQRLAAIRHLFDWLVTGQVVPVNPAAS 114
Psyr            CQVVGVPSLTEVSPLHVATWIELQMQTLAAPSVKQRLAAIRHLFDWLVVGQVVPHNPAAS 90
                *   **.*:. *.*****:*:*     *:.*:***:**.:* ****** *::** ***.*

Int1            VRGPAYSQRRGKTPVLVPDEARHLLDTIDVATHAGLRDRALIGLMVYSFARIGAALAMRV 180
UT26            VRGPAHSQRRGKTPVLAPDEARRLLDSIDVITHAGLRDRALIGLMVYSFARIGAALAMRV 176
SRW1            VRGPAHSQRRGKTPVLAPDEARRLLDTIDVTTPAGLRDRALIGLMVYSFARIGAALAMRV 160
Bps             VRGPSHSSKVGKTPVLDATEARHLLDAIDVSTPAGLRDRALIALMVFSFARIGAALAMRV 174
Psyr            VRGPSHTSRTGKTPVLEPLEARQLLDSIDICTPAGLRDRALIALMVFSFARIGAALAMKV 150
                ****:::.: ****** . ***:***:**: * *********.***:************:*

Int1            EDVFMQNRRLWVRLHEKGGKRHEMPCHHNLEDYLTAYIDGAACARIARGPLFRTIARGTG 240
UT26            EDVFMQNRRLWVRLHEKGGKRHEMPCHHNLEDYLSAYIDGCELREDRKGPLFRTIARGTK 236
SRW1            EDVFVQNRRLWVRLHEKGGKRHEMPCHHNLEHYLAEYLDGCELREDRKGPLFRTIARGTK 220
Bps             DDVYVQNRRLWVRLREKGGKRHEMPCHHTLEAYLHAYLDGTGLANESKGPLFRTIARGTG 234
Psyr            EDVYIQNQRLWVRLKEKGGKQHVMPCQHSLEAYLHAYLVETGIDNDPKGPLFRTIGRGTE 210
                :**::**:******:*****:* ***:*.** **  *:      .  :*******.***

Int1           QLSETPLPKPMLSRCAFAMVRRRAAAAGIGTAIGNHSFRATGITTYLKNGGTLETAATMA 300
```

```
UT26            RLSETPLPQAN----AFAMVRRRAGAAEIGTAIGNHSFRATGITTYLKNGGTLETAATMA 292
SRW1            RLSDTPLPQAN----AFAMVRRRAGAAEIGTAIGNHSFRATGITTYLKNGGTLETAATMA 276
Bps             QLSTTPLPQAN----AYAMVRRRAAAAGIATKIGNHTFRATGITAYLKNGGTIENAAAMA 290
Psyr            QLSVNALPQAN----AHAMVRRRALAAGIKTSIGNHTFRATGITAYLKNGGTLENAAAMA 266
                :** ..**:.     *.******* ** * * ****:*******:*******:*.**:**

Int1            NHSSTRTTQLYDRRPDDVTLDEVERVLIAPSQCPSRWCGGARRRPRSARRSAITVSARPG 360
UT26            NHSSTRTTQLYDRRPDDVTLDEVERVLI-------------------------------- 320
SRW1            NHSSTRTTQLYDRRPDDVTLDEVERVLI-------------------------------- 304
Bps             NHASTRTTQLYDRRRDDISLDEVERIRV-------------------------------- 318
Psyr            NHASTRTTQLYDRRRDEISLDEVERIRLDR------------------------------ 296
                **:*********** *:::******: :
```

Fig. 3.14. Comparison of integrase1 of CIP-I module with the integrases present in NCBI database. Integrase of pPDL2 (Int1) is compared with integrases of *Sphingobium japonicum* UT26S (UT26), *Sphingomonas wittichi* (SRW1), *Burkholderia pseudomallei* (Bps) and *Pseudomonas syrigyae* (Psyr). Identical residues and gaps are shown with an (*) and (-) respectively. Similar residues are shown with (:) and (.) symbols.

gene of CIP-I module was aligned with similar sequences found in databases. Though the integrase has shown similarity to the entire stretch of tyrosine integrases, the C-terminus of the protein was found to be highly conserved than the central and N-terminal region (Fig. 3.15).

```
Prevotella        ISPHTLRHSFATELLKGGADLRAIQEMLGHESIGTTEIYTHIDISTLREEILNHHPRNIM 305
Alistipes         ISPHTFRHSFATHLLEGGASIRQVQEMLGHESILTTEIYTHLEGDHLRDTVEKYLPL--- 298
Paenibacillus     ITPHTLRHSFAVHMLEGGADLRSVQEMLGHADLSTTQVYAQTARRNMKEVYEKHHPHGGN 307
Moorella          ITPHTLRHSFATHLLENGADLRSVQELLGHADIGTTQIYTHLTRKKIREIYDHTHPRA-- 295
Desulfococcus     IKPHTLRHSFASHLLEGGADLRSVQIMLGHSDISTTQIYTHVTYRHLKDAHEKFHPR--- 297
Lawsonia          ISPHTFRHTFATHLLEGGADLRSVQLLLGHVDMSATELYTHVQSDRLKYIHSMFHPRSNY 299
Chlorobium        ISPHTFRHTFATHLLEGGADLRAVQEMLGHSSISTTQIYTHIDRSFVKEVHKTFHPRG-- 304
Rhodopirellula    ISPHSLRHSFATHLLAGGADLRQVQEMLGHASIQTTQIYTHVEHSRLQRVHRDFHPRA-- 325
Bacillus          ISPHVLRHTFATHLLDAGADLRAVQELLGHASLRSTQIYTHTTRERLLQVYLHAHPRA-- 304
Brevibacillus     VSPHTFRHTFATHMLNGGADLRTVQELLGHVNVSTTQVYTHVTKERLRHVYDTAHPRANP 304
Candidatus        VSPHTFRHSFATHLLDNGADLRAVQEFLGHSSLSTTQIYTHVTTERLKQVYDKTHPRA-- 298
Olsenella         ITPHAMRHTYATELLSGGADLRSVQELLGHSSLSTTQIYTHLSVDRLKAAARQAHPRGE- 308
Atopobium         LSPHAMRHTYATELLGGGADLRIVQELLGHESLSTTQVYTHLSVDRLKEAAKAAHPRSK- 305
Mycobacterium     IGPHGLRHSAATHLLEGGADLRIVQELLGHSTLATTQLYTHVTVARLRAVHDQAHPRA-- 300
Alkalilimnicola   VHPHMLRHSFASHLLESSGDLRAVQELLGHADIATTQVYTHLDFQHLARVYDQAHPRARK 298
Congregibacter    VHPHMLRHSFASHLLESSGDLRAVQELLGHSDISTTQIYTHLDFQHLAKVYDGSHPRARK 305
Conexibacter      VSPHALRHSFATHLLENGADLRSIQELLGHASISTTQVYTRVESARLRSAYANSHPRA-- 313
IntP of pPDL2     IGNHSFRATGITTYLKNGGTLETAATMANHSSTRTTQLYDRRPDDVTLDEVERVLI---- 328
Azoarcus          IGNHSFRATGITEYLRNGGKLEIAQQMANHESARTTGLYDRRNDQLTLDEVERIVV---- 328
Polaromonas       MGPHALRATAATNALEHQADIAKVQEWLGHASISTTRVYDRRGSRPEDSPTFKVAY---- 331
Pseudomonas       LGVHGLRATAATNA---------------------------------------------- 279
Desulfobacterium  ITPHSARATFITQALENNCPIEAVQKTVGHAQIKTTQMYDKRTAKYRESASFAVRY---- 301
Frankia           LSPHSLRATSVTLLLDAGASLRDAQDHARHADPRTTRAYDRARGSLDRAGTYQLVAYLDA 310
                  :  *  *  :
```

Fig. 3.15. Multiple alignment of C terminal region of integrase of pPDL2 with other integrases. The conserved motif H-X-X-R is boxed. Identical residues and gaps are

shown with an (*) and (-) respectively. Similar residues are shown with (:) and (.) symbols.

Considering the similarity of integrase coded by pPDL2 and its near identity at the C-terminus especially in the region containing the H-X-X-R domain (Fig. 3.15), a proposal is made in this study to assign the status of integration mobilizable element (IME) to plasmid pPDL2. Existence of *oriT* like sequence and *repB*, involved in site-specific single stranded nick at *oriT* adds strength to this proposition.

The CopG/MetJ/Arc family regulatory protein (further referred as CopG) coded by pPDL2 shows highest similarity to the CopG regulatory protein coded by plasmid pUT1 of *Sphingobium japonicum* UT26S. The 131 amino acid long protein shows homology to a number of other proteins belonging to CopG/MetJ/Arc family of transcriptional regulators. The proteins of this family act both as transcriptional repressors and activators (del Solar and Espinosa, 1992; del Solar et al, 1995). In a well characterized streptococcal plasmid, pMV185, the CopG is shown to act as a repressor by regulating the expression of *repB* by binding to a pseudosymmetirc region present overlapping the -35 hexameric sequence of σ70 dependent promoter. Such binding is shown to prevent transcription from the *copG* promoter of the *copG –repB* genes which are co-transcribed in plasmid pMV158 (Farris et al, 2008).

In addition to these two transcriptionally coupled integrase and *copG* an additional *orf* that codes for a protein showing homology (98%) to *Sphingobium japonicum* UT26S phosphoglycerate mutase is identified in each unit of the integrase modules. The phosphoglycerate mutase gene in ICP-I module is found 1.1 kb away from the stop codon of integrase. This 216 amino acid long protein shows high homology (98%) to phosphoglycerate mutase present in *Sphingomonas japonicum* UT26S. When

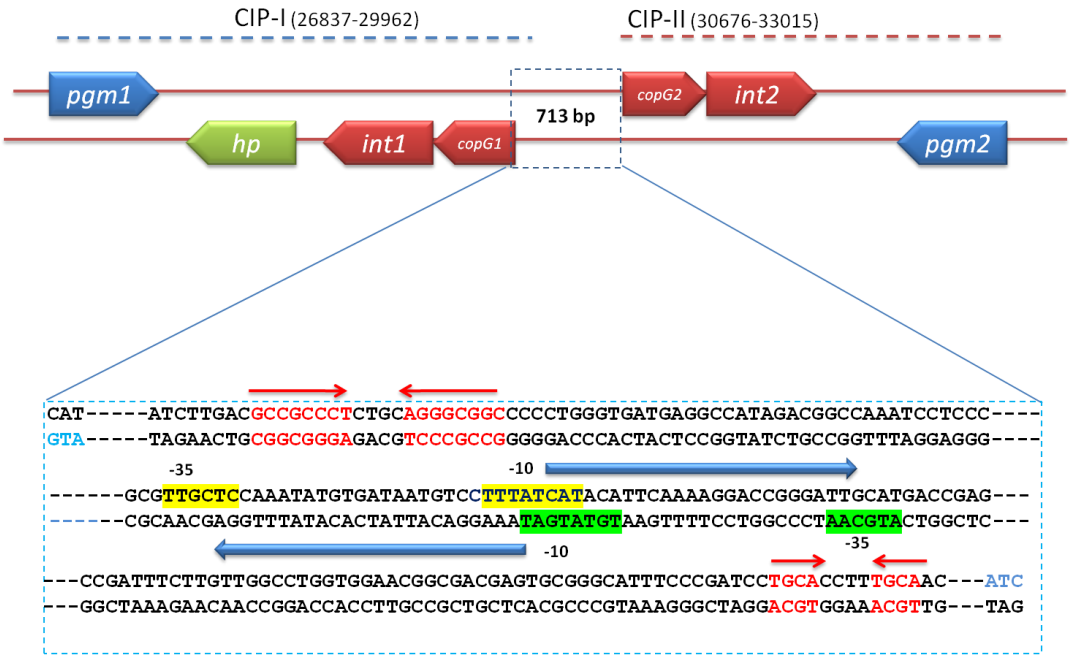
the sequence of phosphoglycerate mutase was analysed, it has shown existence of histidine phosphatase phosphoglycerate like domain (HP-PGM like domain) at the N-terminus of the protein. When examined the domain region has shown identity with similar a domain found in alpha-ribazole phosphatase (Fig. 3.16). Similarly, it has also shown similarity with the HP domain found in *Acinetobacter*.

```
PD1             -RGIRIDDAHAGPYSSATAKAPGNAGVPCHDLGAIELTELGQEQAREVAASWIEAPALIV 59
pUT1            MRAIFIRHGQS----------TGNAGVPCDDLGAIELTELGQEQAREVAASWTQAPALIV 50
                 *.* * ..::           .*******.********************** :******

PD1             TSPYTRTQQTAAPTIARFPGVPVEVWPIEEFTYLQPARWNGTRSAERMPHLERYWSAADP 119
pUT1            TSPYTRTRQTAAPTIARFPGVPVEVWPIEEFTYLQPARWNGTRSAERMPHLERYWSAADP 110
                *******:*****************************************************

PD1             DYCDGEGAESFATLLRRCEAALARLAAMPVASPVYVFGHGQFIQAARAIVADAHMDDRAK 179
pUT1            DYCDGEGAESFATLLRRCEAALARLAAMPAASLVYVFGHGQFIQAARAIVADAHLDERAK 170
                ******************************.** **********************:*:***

PD1             MRAFWRKGEPPAIANAQRVGFHWEGGPLVLCAGAGGVMSAAPLIPTEDEIREHQAWDEFA 239
pUT1            MRAFWRKGEPPAIANAQRVGFHWEGG----------RWSCAP------------------ 202
                **************************            *.**

PD1             GGRPGAPRKGKIPLCDSPPPPQCGVVHLGAKDRWGAVGGKAE 281
pUT1            -----------------------------------ALAGTA- 208
                                                  *:.*.*
```

Fig. 3.16. Shows alignment of Phosphoglycerate mutase (PGM) coded by plasmid pPDL2 (PD1) and PGM (pUT1) coded by plasmid pUT1 of *Sphingobium japonicum* UT26S.

Phosphoglycerate mutases (PGMs) are very well characterized group of enzymes. Their existence is seen both in prokaryotes and eukaryotes. PGMs are basically transferases and are involved in transfer of phosphate group (Parkinson and Kofoid, 1992). Transfer of phosphates from 3rd position to 2nd position converting glyceraldehyde 3-phosphate to glyceraldehyde 2-phosphate is the classical biochemical reaction catalyzed by this group of enzymes. However, the PGMs containing HP-PGM like domain are also known to involve in signal transduction process (Matsubara and Mizuno, 2000). The histidine present in the catalytic site of this group of enzymes undergoes phosphorylation during the signal relay process (Parkinson and Kofoid, 1992). Strong link

between PGM and CopG and integrase genes suggests functional relevance among these proteins. Such organization is also seen in pCHQ1 (NC014007) and pUT1 (NC014005) of *Sphingobium japonicum* UT26S and pSWIT01 A)



B)

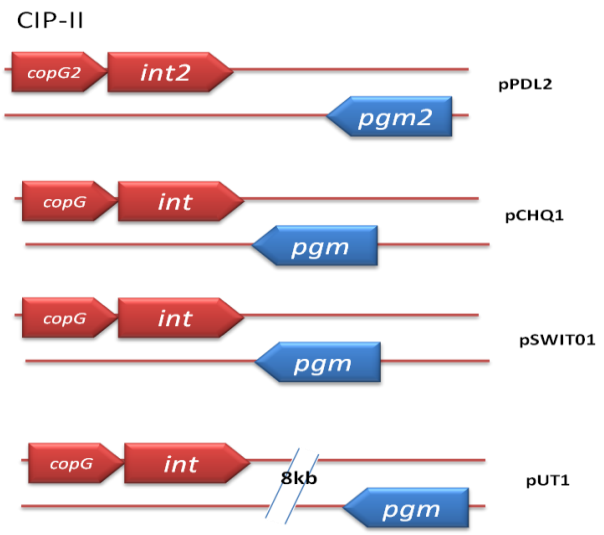


Fig. 3.17. Organization of Integrase modules in plasmid pPDL2. Panel A. The CIP units, CIPI and CIP-II are shown with dotted lines. The predicted regulatory region between CIP-I and CIP-II is boxed. The putative promoter elements identified upstream of *copGI* and *copGII* are highlighted with green and yellow colours respectively. The putative CopG

binding sites are shown with inverted arrows. Closed arrows indicate direction of transcription.
Panel B shows comparison of the organization of *cop*, *int* and *pgm* genes found in plasmids pPDL2, pCHQ1, pSWIT01 and pUTI.

of *Sphingomonas wittichi* RW1 (Fig. 3.17-B). Linkage of these three genes in plasmids isolated from different bacterial strains points towards having a functional relevance behind this conserved genetic organization. Further, studies are required to elicit the role of PGM in regulation of CopG-integrase operon expression.

As shown in Fig. 3.17-A, the integrase module contains a second copy of *copG*, *int* and *pgm*. We designate this region as CIP-II. This region also codes CopG, integrase and phosphoglycerate mutase as in CIP-I. As shown in unit-1, the *copG* and *int* are shown to be organized as one transcriptional unit. The stop codon of *copG* overlaps with the start codon of *int* gene. Such strong translational coupling suggests co-transcription of *copG* and *int* genes. CopG coded by CIP-II appears to have alternate start codon, ATC. Interestingly, the *copG* present in plasmid pCHQ1 of *Sphingobium japonicum* UT26S to which *copG* of pPDL2 shows high homology is also having an alternate start codon, ATC (AP010805). Like other proteins of plasmid pPDL2, CopG shows high homology to CopG coded by plasmids pCHQ1 and pUT1 of *Sphingobium japonicum* UT26S. CopGs coded by CIP-I and CIP-II are similar except that the CopG coded by CIP-II unit is using an alternate start codon, ATC (Fig. 3.18).

```
1.CopG1         -------------------------------------------MHKDDDTAFADNYAER 16
2.CopG2         IWRRSRPPTPTHGAIATRCTTFGCGIAAILPRAFKRAGDPRPKGSPMTDDELPDDNPAER 60
                                                            **    ** ***

1.CopG1         DQARALREQARTGGLRFEAYLPGDMADWLLAQVEQGHFVDPSEAVFAIVKNFIEMEPHRD 76
2.CopG2         AQAKALREQARAGGLRFEAYLPGSMADWLLAQIERGRFADPSEAVFLIVQNFIEMEPHRD 120
                 **:*******:***********.********:*:*:*.******* **:**********

1.CopG1         LRDELLRRILDASVVRGLEDVKAGRVRPAEEVFDELRRKMAEPRPEPARWAKIAR 131
2.CopG2         LQDELLRRILQA---R-IDDPRPG--IPHDEACARIDRLLAEPRPDPARWEKIAR 169
                *:********:*   * ::* :.*   * :*.  .: * :*****:**** ****
```

Fig. 3.18. Alignment of CopG1 and CopG2 sequences of integrase module of plasmid pPDL2. Pairwise alignment CopG1 and CopG2 coded by CIP-I and CIP-II units of integrase module are shown. Identical residues and gaps are shown with an (*) and (-) respectively. Similar residues are shown with (:) and (.) symbols.

After analysing the ORFs of integrase module, the proteins coded by CIP-I were aligned to similar proteins coded by CIP-II. This comparison is done mainly to know if CIP-I is duplicated to give rise to CIP-II. If CIP-II has evolved through an event of gene duplication, the CopG, Int and PGM coded by CIP-I should have 100% sequence identity with their counterparts found in CIP-II. When CopG1 encoded by unit CIP-I is compared with CopG2 coded by CIP-II sequence, significant differences were noticed with respect to the primary sequence (Fig. 3.18). There exists only 67% identity between these two proteins. The N-terminal region of CopG2 encoded by CIP-II unit was found to be much longer (Fig. 3.18). Similarly, the C-terminus was not highly conserved. However, the central region of CopG1 and CopG2 proteins were found to be almost identical. Such diversity in the primary sequence indicates divergent origin of the CopG1 and CopG2 sequences. Similar situation was seen when *int1* and *pgm1* were aligned with *int2* and *pgm2*. Between *int1* and *int2* only 80% identity was seen (Fig. 3.19). However, the identity continued throughout the sequence. Interestingly,

```
1.Int-1_pPDL2        MNELAPLPPPPSSSALALPALVASADEAARLRFLEFFAVTIRNPHTRRAYMRAAGEFLAW
60
2.Int-2_pPDL2        MNQLAPLPSPASS--PALPALIAAADDDTRRRFLEFFAVTIRNPHTRRAYARAAGDFLAW
58
                   **:*****.*.**   *****:*:**: :* ******************* ****:****

1.Int-1_pPDL2        CEARGVASLAGVQPLHVAAWIEAQGGELAPPSVKQQLAGVRSLFDWLVMGQVVPANPAAS
120
2.Int-2_pPDL2        CEARGVASLAGVQPLHVAAWVEALGRELAAPSVKQQLAGVRHLFDWLVTGHIVPVNPAGS
118
                   ********************:** * ***.*********** ****** *::**.***.*

1.Int-1_pPDL2        VRGPAYSQRRGKTPVLVPDEARHLLDTIDVATHAGLRDRALIGLMVYSFARIGAALAMRV
180
2.Int-2_pPDL2        VRGPAHSQRRGKTPVLAPDEARRLLDSIDVTTHAGLRDRALIGLMVYSFARIGAALSMRV
178
                   *****:**********.*****:***:***:*************************:***
```

```
1.Int-1_pPDL2        EDVFMQNRRLWVRLHEKGGKRHEMPCHHNLEDYLTAYIDGAACARIARGPLFRTIARGTG
240
2.Int-2_pPDL2        EDVFVQNRRLWVRLHEKGGKRHEMPCHHNLEHYLVEYIDGCGLREDRKGPLFRTIARGTK
238
                    ****:***************************.**. ****..  .  :***********

1.Int-1_pPDL2        QLSETPLPKPMLSRCAFAMVRRRAAAAGIGTAIGNHSFRATGITTYLKNGGTLETAATMA
300
2.Int-2_pPDL2        RLSDTPPAPSQCPSRWCGGARRRPRSA-----------RRSAITVSARPGSPPI------
281
                    :**:** . .  .    . .***. :*           * :.**.  : *..

1.Int-1_pPDL2       NHSSTRTTQLYDRRPDDVTLDEVERVLI 328
2.Int-2_pPDL2       ----------------------------
```

Fig. 3.19. Comparison of integrases coded by *int*-I and *int*-II genes of plasmid pPDL2. Pairwise alignment integrase-I of CIP-I and integrase-II of CIP-II modules is shown. Identical residues and gaps are shown with an (*) and (-) respectively. Similar residues are shown with (:) and (.) symbols.

the percent identity between PGM1 and PGM2 is rather high. About 91% sequence identity was seen between these two proteins. The difference is only seen at C-terminus and N-terminus regions. About 10 amino acids gap was seen in PGM1 and PGM2 at both the N and C-terminal regions. PGM2 was found to be shorter by 21 amino acids due to existence of a 10 and 11 amino acids long gap at the N terminus and extreme C-terminus regions (Fig. 3.20). Considering the diversity in the primary sequence of the protein coded by CIP-I and CIP-II of the integrase module, the plasmid pPDL2 is proposed to have acquired the CIP-I and CIP-II integrase modules by independent sources possibly through unique recombination process.

```
1PGM1            -RGIRIDDAHAGPYSSATAKAPGNAGVPCHDLGAIELTELGQEQAREVAASWIEAPALIV 59
2PGM2            MRAIFIRHGES----------TGNAGVPCHDLATIELTERGQEQARAVAASWTEAPALIV 50
                  *.* * ...:           .**********.:***** ****** ***** *******

1PGM1                TSPYTRTQQTAAPTIARFPGVPVEVWPIEEFTYLQPARWNGTRSAERMPHLERYWSAADP
119
2PGM2                TSPYTRTRQTAAPTIARFPGVPVETWPIEEFTYLQPSRWNGTRSAERMPHLERYWSAADP
110
                 *******:****************.***********:**********************

1PGM1                DYCDGEGAESFATLLRRCEAALARLAAMPVASPVYVFGHGQFIQAARAIVADAHMDDRAK
179
2PGM2                DYCDGEGAESFGTLLRRCEAALARLAAMPADSLAYVFGHGQFIQAARAIVADAHMDDRAK
170
                 ***********.*****************. * .**************************

1PGM1            MRAFWRKGEPPAIANAQRVGFHWEGGRWSCAPALAA 215
```

```
2PGM2            MRAFWRKGEPPAIGNAQRVGFHWQGD---------- 196
                 *************.*********:*.
```

Fig. 3.20. Pairwise alignment phosphoglycerate mutase -I and phosphoglycerate mutases-II coded by of CIP-I and CIP-II units of integrase module. Identical residues and gaps are shown with an (*) and (-) respectively. Similar residues are shown with (:) and (.) symbols.

The transcriptional organization of *copG* and *int* genes of CIP-I and CIP-II have revealed an interesting observation. As indicated before, in both CIP-I and CIP-II, these two genes are found to be co-transcriptional. In fact, in both the cases the translational stop codon overlaps of CopG with translational start codon of integrase. As shown in Fig. 3.17, *copG1* and *int1* of CIP-I and *copG2* and *int2* of CIP-II have opposite transcription orientation. Between these two transcriptional units there is a gap of 713 bp. When this gap region was analysed for identification of promoter and other *cis*-elements involved in regulation of gene expression, quite a few interesting features were noticed. As shown in Fig. 3.17 the gap region contained two independent putative promoter elements. The promoter element proposed to be involved in transcription of *copG1* and *Int1* operons has shown sufficient similarity to the consensus σ70 dependent promoter (Fig. 3.17 promoters are highlighted in green). Similarly, the promoter element predicted to be involved in transcription of *copG2* and *int2* operon was found to be similar to the consensus σ70 dependent promoter. There was an overlap of 5 bp between these putative *copG1* and *copG2* promoters. As indicated in earlier section CopG1 proteins are shown to be transcription repressors. As *copG* and *int* genes are co-trancriptional, the intergenic region found between the *copG1* and *copG2* operons were further analyzed to identify *cis*-elements that can serve as target sites for binding CopG proteins. In studies conducted by del Solar and his associates, CopG is shown to bind to a pseudosymmetric

cis-element present overlapping -35 region (del Solar et al, 2002). Inverted repeats are also shown to be putative binding sites of CopG in *Sulfolobus neozealandicus* (Greve et al, 2004). Interestingly, in CIP-I unit, a typical CopG binding site was predicted 276 bp upsteam of start codon of *copG1* (Fig. 3.17). Likewise in CIP-II an inverted repeat that can act as a potential CopG binding site was observed 45 bp upstream of translational start codon of *copG2* (Fig. 3.17). Existence of such CopG binding sites, if seen together with the CopG role as transcription repressor, the organization seen in integrase module of pPDL2 appears to be tailor made for regulation of integrase expression in *Flavobacterium* sp. ATCC 27551. Further work has to be done to gain better insights into the regulation of integrase expression.

3.9.1. The attachment (*attP*/*attB*) sites

Integrase is known to mediate integration of site-specific recombination between two conserved specific sites. In well studied bacteriophages λ, the phage integrase mediates site specific recombination between phage specific P site (*attP*) and bacterial chromosome specific B site (*attB*). These sites are called attachment sites. Conventionally, they are named as *attP* (P-phage) and *attB* (B-bacteria) sites. The site specific recombinase encoded by plasmid pPDL2 has high homology to tyrosine recombinases (Fig. 3.14 and 3.15). Tyrosine recombinases integrate target sequence at a specific site (*attB*), usually present at the 3' end of t-RNA genes (Williams, 2002). In the background of this information, the pPDL2 sequence was thoroughly analyzed to identify *attP* homologs. The bioinformatic searches have predicted existence of two such sites in plasmid pPDL2 and were designated as *attP*-I and *attP*-II. The *attP*-I is located between nucleotide position 15197 to 15186 and *attP*-II is found between nucleotide positions

37889 to 37878. The predicted *att* sites are of 12 bp long. When these two sites were aligned there was absolute identity between last 6bp of *attP*-I and first 6 bp of *attP*-II (Fig. 3.21). After identifying putative *attP* sites an attempt was made to identity *attB* sites. In order to identify *attB* in *Flavobacterium* sp. ATCC 27551 from where pPDL2 was isolated it is required to know complete genome sequence of the host organism. However, such information is not available for *Flavobacterium* sp. ATCC 27551. *Flavobacterium* sp. ATCC 27551 has recently been reclassified as *Sphingobium fuliginis* (Kawahara et al, 2010). A number of *Sphingobium* total genome sequences are available in the public domain as reported in the earlier sections. Plasmid pPDL2 has been shown to have considerable homology to plasmid pUT1 isolated from *Sphingobium japonicum* UT26S. The total genome sequence of *Sphingobium japonicum* UT26S is available in the public domain (www.**ncbi**.nlm.nih.gov/). The total genome sequence was used as input to identify putative *attB* sites that have homology to *attP* sites predicted in plasmid pPDL2.

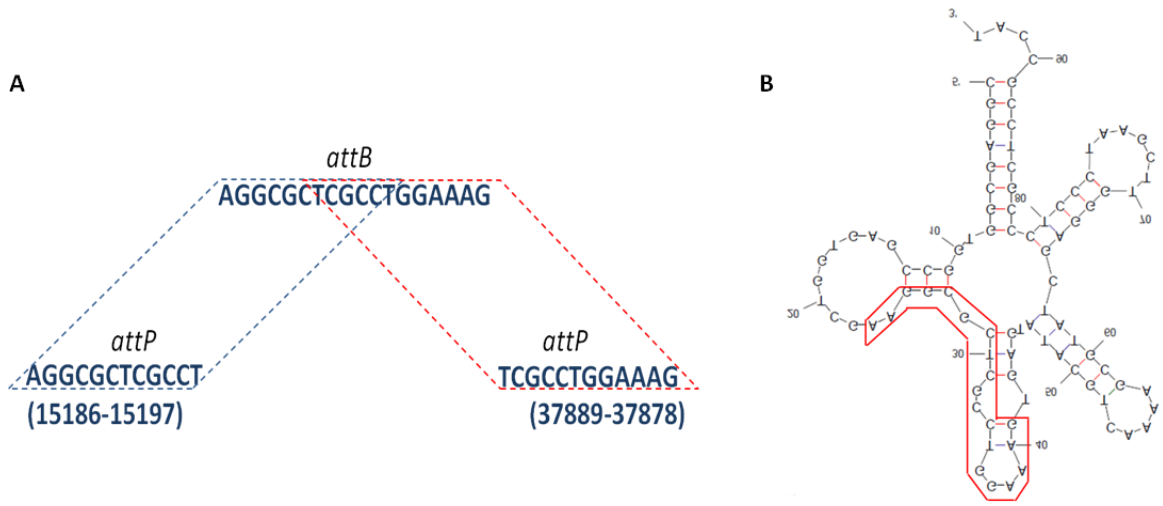


Fig. 3.21. Organization of *attP* sites on pPDL2 of *Flavobacterium* sp. ATCC 27551. Panel A shows the positions of *attP* sites on plasmid pPDL2. In Panel B the putative *attB* site on the tRNA-serine gene is outlined.

Interestingly, identical sequences were identified in the t-RNA gene sequences of *Sphingobium japonicum* UT26S (Fig. 3.21). If existence of *attP* site is taken together with the *attB* sequence in the *Sphingobium* genome, there exists an ample scope for plasmid pPDL2 to integrate into the genome.

3.10. Degradative module

As mentioned in aforementioned section, plasmid pPDL2 is associated with degradation of organophosphorus compounds. A well conserved parathion hydrolase coding *opd* gene has been shown as part of transposon-like element (Siddavattam et al, 2003). The sequence determined in this study agreed in total with sequence information reported from our laboratory. As indicated before the *opd* sequence is given a generic name as *orf11* (Table 3.3). A fine diagrammatic representation indicating the gene involved in degradation of organophosphorus compounds is made (Fig. 3.22). The degradation module includes parathion hydrolase or organophosphorus hydrolase coded by the *opd* gene, *meta* fission product hydrolase (*mfhA*), protocatechuate hydrolase (*pcaH*) and β-ketoadipate enol lactone hydrolase. Ample literature is available on catalytic properties, substrate range and membrane targeting of organophosphorus hydrolase (Karpouzas and Singh, 2006, Gorla et al, 2009).

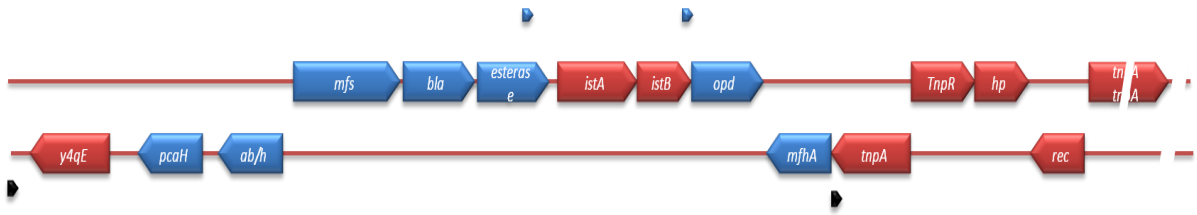


Fig. 3.22. Organization of degradative module in pPDL2. Transposases and resolvases are shown in red colour. Degradative genes are shown in blue coloured arrows. The

repeats of Tn*3* and IS*21* element are shown as black arrows and blue arrows respectively. Degradative reactions catalyzed by OPH, MfhA, P45O, b-keto adipate lactonase are shown using methyl parathion as model compound.

In our previous studies, we have we have reported existence of a *meta*-fission product hydrolase gene, *mfhA*, immediately downstream of *opd* gene (Khajamohiddin et al, 2006). The *mfhA* gene is also shown to be transcriptionally linked to the *tnpA* coding gene of transposon Tn*3* (Siddavattam et al, 2003). In this study, two more important genes that contribute to mineralization of organophosphorus compounds have been identified. Previous studies conducted by our lab has identified IS element belonging to IS*21* class upstream of *opd* gene (Siddavattam et al, 2003). When the complete sequence of pPDL2 was analyzed upstream of the IS element IS*Flsp1*, two more ORFs designated as *orf2* and *orf3* have been identified. These two ORFs were shown to code proteins that have significant homology to β- subunit of Protocatechuate 4, 5 dioxygenase and β-ketoadipate enol lactonase. Therefore these two ORFs were designated as *pca1* and *pca2*.

Organophosphorus hydrolase has been shown to hydrolase a variety of well known insecticides that are found to have aromatic moieties are linked to central phosphoric acid residue through an ester-linkage (Karpouzas and Singh, 2006). Upon OPH mediated hydrolytic cleavage most of OP-insecticides release aromatic compounds like 4-nitrophenol, 3-methyl 4-nitrophenol. These aromatic compounds are found to be much more toxic to the microbes than the parent OP compounds (Crbella et al, 2001). As shown in figure. 3. 22, existence of protocatechuate 4, 5 dioxygenase, β-ketoadipate enol lactonase in the upstream region of *opd* gene, flanked by transposase y4qE and transposase of Tn*3* suggests evolution of a well conserved degradation module to

mineralize OP insecticides. The dimethyl thiophosphoric acid generated from OP insecticides like methyl parathion, paraxon and parathion can be quickly utilized as source of carbon and sulphur. However, the aromatic compounds thus generated through OPH activity, require dioxygenase, β-ketoadipate enol lactonase and *meta* fission product hydrolase, to convert them as an intermediates of TCA cycle. Existence of β-ketoadipate enol lactonase and *meta*-fission product hydrolase and a dioxygenase coding region might contribute for such conversion. Though experimental evidence has been shown on the functions of *opd* gene and *mfhA* (Mulbry and Karns, 1989; Khajamohiddin et al, 2006), the role of these two genes *pca1* and *pca2* in mineralization of aromatic compounds need to be established by further experimentation. However, if degradation module is carefully examined, evolution of degradation traits in the form of a mobilizable element is very apparent.

3.10.1. **Protocatechuate 4, 5 dioxygenase (P45O)**

The *pca* designated with a generic name as *orf27* is present between the nucleotide positions 1185 to2067. The 277 amino acids long protein coded by *pca* has a domain belonging to extradiol dioxygenase 3B-like super-family and the complete sequence shows 79% identitiy to the Protocatechuate 4,5 dioxygenase identified in *Xanthomonas campestris* pv. *campestris.* Therefore, the *pca1* encoded protein is designated with a functional name as Protocatechuate 4,5 dioxygenase (P45O). The Protocatechuate 4,5 dioxygenase is the key enzyme in the benzoate degradation pathway (hydroxylation pathway) and 2,4 dichlorobenzoate degradation (Adriaens et al, 1989;). In addition to these two compounds the P45O has been shown to cleave many aromatic compounds (Fig. 3.23).

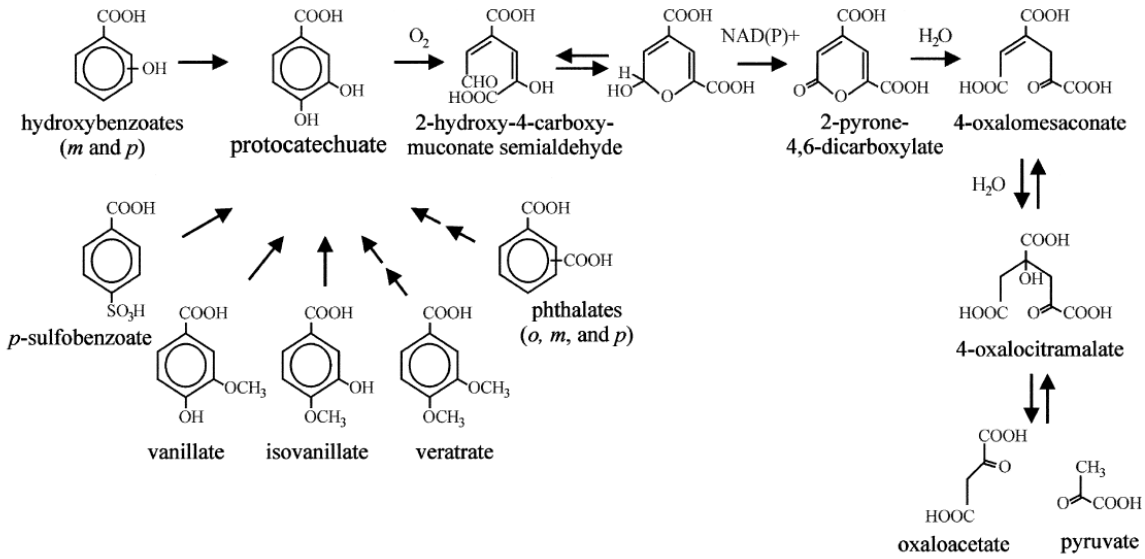


Fig. 3.23. Channelling of aromatic compounds into the TCA cycle through the Protocatechuate degradation pathway (Providenti et al, 2001).

Existence of such ring cleavage enzyme might contribute for conversion of 4-nitrophenol into intermediates of TCA cycle through a well known pathway known as protocatechuate degradation pathway. Degradation of aromatic compounds *via* protocatechuate pathway generates ring cleavage products with lactone ring. These lactones are further channelled into the TCA cycle through β-ketoadipate pathway (Khajamohiddin et al, 2008).The β-ketoadipate enol lactone hydrolase and Mfha coding *meta* fission product hydrolase contribute for such channelization. In degradation module along with OPH and P45O these two ORFs are present. The *orf311* coding protein shows 43% homology to the 3-oxo-adipate enol lactonase of *Xanthomonas campestris*. Its existence in the degradation module is yet another evidence to claim plasmid pPDL2 has evolved to mineralize OP compounds used as insecticides.

3.10.2. Major facilitator super-family protein

In the degradation module the next prominent ORF is *orf5* (3433-4887), designated as *orf484* encodes for a protein of 484 amino acids. The sequence of Orf484 shows homology to many of the aromatic acid transporters (Fig. 3.24). It shows 42% similarity with vanillate transporter of *Xanthomonas campesteris* pv. *campesteris.* Vanillate is shown to metabolize via protocatechuate degradation pathway (Fig. 3.23). As shown in Fig. 3.24 the transporter found immediately downstream of protocatechuate dioxygenase, is highly similar to number of aromatic acid transporters. Presence of similar transporters in close association of other genes that contribute for mineralization of organophosphates provides *prima facie* evidence to show existence of complete information on plasmid pPDL2 for mineralization of OP insecticides.

In Pseudomonads and in many Gram negative strains, genes are usually organized into operons and are co-ordinately regulated. Such organization is also seen in genes coding for enzymatic machinery involved in degradation of a number of aromatic compounds (Harayama and Rekik, 1990; Harwood and Parales, 1996; Yen and Serdar, 1988).

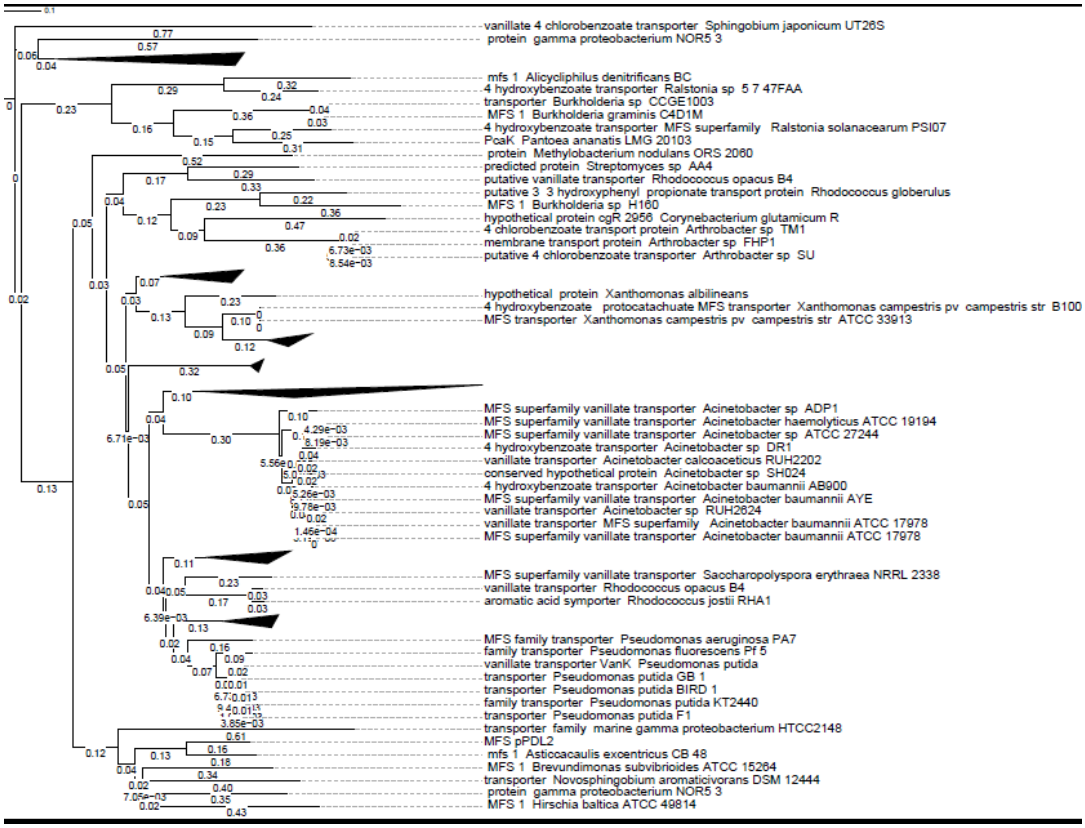


Fig. 3.24. Phylogenetic tree of Major Facilitator Superfamily (MFS) proteins. MFS of pPDL2 is seen in a clade that shows similarity with vanillate and other aromatic compounds

In the degradation module found in pPDL2 is compared with similar modules found in TOL and NAH plasmids the degradative genes on pPDL2 are dispersed. This kind of organization might be due to presence of transposons and recombinases causing rearrangement of genes. *Flavobacterium* sp. ATCC 27551 was reclassified as *Sphingobium fuliginis* based on chemotaxonomic and phylogenetic evidences (Kawahara, 2010). Dispersed organization of otherwise clustered genes is a common feature in Sphingomonads. Especially the genes coding for catabolic pathway enzymes in *Sphingomonas* strains are often found localized away from one another creating an unusual organization, where having a co-ordinately regulated operons are seen as a rare phenomenon (Basta, 2004). This has been described for the genes involved in the degradation of γ-hexachlorocyclohexane (lindane) by *S. paucimobilis* UT26 (Miyauchi et

al, 1998; Nagata, 1999), pentachlorophenol by *S. chlorophenolica* (Cai and Xun, 2002), protocatechuate by *S. paucimobilis* SYK-6 (Masai, 1999), naphthalene, biphenyl, and toluene by *S. yanoikuyae* B1 and *S. aromaticivorans* F199 (Romine et al, 1999; Zylstra and Kim, 1997), and dibenzo-*p*-dioxin by *S. wittichii* RW1 (Armengaud et al, 1998). Such dispersed organization, as seen in *opd* element of pPDL2 (Fig. 3.22.), may be a typical characteristic feature in Sphingomonads in which *Flavobacterium* sp. ATCC 27551 is placed according to the new classification (Kawahara et al, 2010).

3.11. Mobile genetic elements

In total 1 IS element and 3 transposons are present in the sequence of pPDL2. Existence of an IS element that shows homology to IS*21* class of Insertion elements is reported in our previous studies (Siddavattam et al, 2003) and this IS element, designated as ISFlsp*1* and was deposited in the IS database (http://www-is.biotoul.fr). This IS element is present in the upstream region of *opd* gene, spanning the nucleotide positions from 6941 to 9300. This IS element consists of two ORFs designated as *istA* and *istB* which encode for proteins of 507 and 279 amino acids, respectively. The stop codon of *istA* overlaps with start codon of *istB* which imply that the two genes, *istA* and *istB* are translationally coupled. The 507 amino acid protein designated as IstA, shows 59% identity with transposase ISMdi7 (IS*21* family) of *Methylobacterium extorquens* DM4 (YP_003065654) and 55% identity with IS*21* family transposase of *Agrobacterium tumefaciens* str. C58 (NP_355800) (Fig. 3. 26).

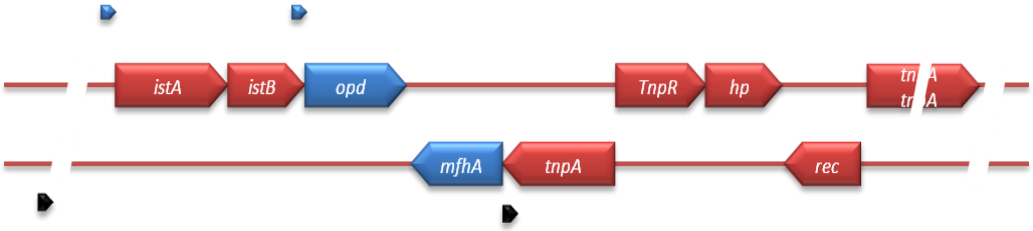


Fig. 3.25. Organization of mobile genetic elements on pPDL2. Transposases and resolvases are shown in red colour. Degradative genes are shown in blue coloured arrows. The repeats of Tn*3* and IS*21* element are shown with black and blue arrows respectively.

```
PD              MKSVEIYAKVRRAVLVEGMTRREAARYFGVHRNTITKMLQYAEPPGYRRAVPRVSEKLAP 60
Auma            MFAVEVYAAVRHFVLIERNSQREAARVFGLSRETVSKMCRFSLPPGYTRVKPVARPKLGA 60
Meex            MFVVEVYAAVRQFVFIEGQSRREAARVFGLSRETIAKMCRFSLPPGYTRSKPVEKPKLGP 60
                *  **:** **: *::*  ::***** **: *:*::** ::: **** *  *    **..

PD              FETLIDEILRSDKGAPPKQRHTCKRIYERLRTEHGYTGGLTILSDYVRSQRLRSREVFIP 120
Auma            LLPVIDWILEADGTAPVKQQHTAKRIFERLRDEHGYGGGLTVVKDYVRIARGRLRETFVP 120
Meex            LLPVIAAILEADRTAPLKQRHTAKRIFERLRDEHGYAGGYTVVKDHVRICRARGQETFVP 120
                : .:*  **.:*  ** **:**.***:**** **** ** *::.*:**  * * :*.*:*

PD              LSHRPGHAQVDFGEADAIIAGKRVRLHYFCMDLPQSDGCFVKAYPAEVAEAFCDGHVSAF 180
Auma            LAHSPGHAQVDFGEAIGVIGGVRQKIHFFCMDVPQSDAPFVKAYPRETTEAFLDGHVSAF 180
Meex            LAHPPGHAQVDFGEAVATIAGVRRKIHFFCMDLPHSDACFVKAYPRETTEAFLDGHVAAF 180
                *:* *********** . *.* * ::*:****:*:**. ****** *.:*** ****:**

PD              AFFGGVPTRILYDNTRLAVARILGDGRRERSRMFAGLQSHYLFDDRFGRPGKGNDKGKVE 240
Auma            DFFGKVPLSILYDNTTIAVARICGDGRRERTRAFTELQSHYLFADRFGRPGKGNDKGKVE 240
Meex            AFFGGVPLSILYDNTKIAVAKICGDGQRERTRAFTELVSHCLFRDRFGRPGRGNDKGKVE 240
                 *** **  ****** :***:* ***:***:* *: * ** ** *******:********

PD              GLVGYVRRNFMVPIPAAASIEELNARFADQCRRRGAAVLRGQSQSITARMEADSAAFMPL 300
Auma            GLVKYARSNFMTPIPQAASFDDLNAMLAERCRQRQGEVAGRHSETIGERLVADLEAFKDL 300
Meex            GLVKFARSHFMTPAPEAASFEALNADLERRCRARQNECAGRHPESIGTRLMADRVVLRAL 300
                *** :.* :**.* * ***:: *** :  :** *       :.::*  *: **  .:  *

PD              PEVAFDPCHIDSGCASSMALVRYRTNDYSVPTAFAHQQVVIKGYVDRVDIVCRGTCIASH 360
Auma            PATPLEPCEKRAARVSSTALVRYRCNDYSVPTSFGFRDVLVKGFVDEVVILCAGEEIARH 360
Meex            PAVPLEPCEKRAGRVSSTALVRYRGNDYSVPTTYGFRDVLVKGFVEEVVILCAGVEIARH 360
                * ..::**.  :. .** ****** *******::..::*::**:*:.* *:* *  ** *

PD              VRRYEREDFIANPLHYLALLEHKPGALDQAAPLDGWHLSEPVHRLRRLMEARSGKEGRRA 420
Auma            RRSYATGTFVFDPLHYLMLLEMKPNALDQAAPLQGWDLPETFQHLRHLLEARMGNRGKRE 420
Meex            PRSYGSGVFVAEPLHYLALIETKPNALDQAAALQGWDLPEAFQHLRHLLEARMGNRGKRE 420
                 * *    *: :***** *:* **.******.*:**.*.*..::**:*:*** *:.*:*

PD              FIQVLRLCEHYEQSLVEWAVARALELGAISFDAVKMILLARLEHRPARLDMSLYPYLPRA 480
Auma            FIQVLRLMEAMPMGIVAAAVTEAIRLGAIGFDAVKLIALSRIERRPLRLDLSRYPHLPKM 480
Meex            FIQVLRLMEAMPKDLVAWAVTEAIRLGAIGFDAVKLIALARLERRPPRLDLSAYPHLPRP 480
                ******* *    .:*  **:.*:.****.*****:* *:*:*:** ***:* **:**:

PD              NVGVTDTRAYLGLIPDAHRVTMKGASA 507
Auma            DVRTTAAADYAVLVPGKAA-------- 499
Meex            AVRATMAADYTVLVPEVAA-------- 499
                 * .* :  *  *:*
```

Fig. 3.26. Multiple alignment of IstA of pPDL2 of *Flavobacterium* sp. ATCC 27551 (PD) with similar proteins present in *Aurantimonas manganoxydans SI85-9A1*(Auma) and *Methylobacterium extorquens* DM4 (Meex). Identical residues and gaps are shown with an (*) and (-) respectively. Similar residues are shown with (:) and (.) symbols.

Similarly IstB, shows 71% identity with resolvase IstB of *Mesorhizobium* sp. F28 (ABY59054) and *Sphingomonas wittichii* RW1 (YP_001260016). The IS*21* family transposons have a length of 2 kb to 2.5 kb and are found to be among the largest bacterial IS elements. They carry related terminal IRs whose lengths may vary between 11 to 50 bp generally terminating with a dinucleotide 5'-CA-3' (Mahillon and Chandler, 1998). The *istA* and *istB* of pPDL2 are flanked IR sequences of 15 bp present 100 bp upstream of *istA* (IRL) and 150 bp downstream of *istB* (IRR).

3.11.1. Tn*3* transposon

The sequence of pPDL2 found immediately downstream of *mfhA* gene showed existence of two copies of Tn*3* elements separated by an open reading frame coding for a hypothetical protein. These two copies of Tn*3* elements are designated as Tn*3*-I and Tn*3*-II (Fig. 3.25). These two Tn*3* like elements span from nucleotide position 11424 to 16077. The transposon Tn*3 is a* well characterized mobile genetic element found in a number of taxonomic groups. The Tn*3* transposon contains two genes, *tnpA* and *tnpR* and a 38-bp terminal inverted repeat at the left (IRL) and right terminus (IRR) (Heffron et al, 1979). The *tnpA* and *tnpR* code for transposase and resolvase. Transposition event takes place in two steps (Heffron et al, 1979; Shapiro, 1979; Grindley, 1983). The first step is formation

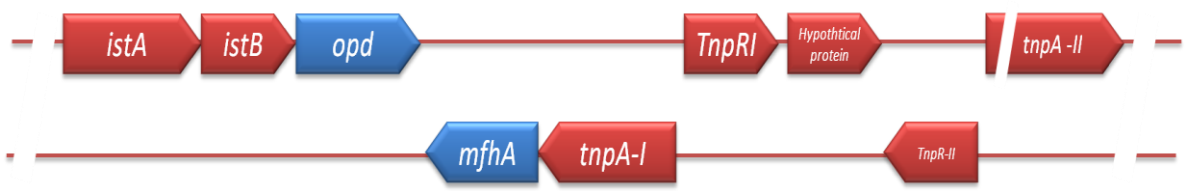


Fig. 3.27. Organization of transposon Tn*3* copies in plasmid pPDL2. The transpoase, resolvase coded by Tn*3*-I and Tn*3*-II are designated as *tnpa*-I *tnpA*-II, *tnpR*-I and *tnpR*-II respectively.

of cointegrate by the transposase, TnpA and the second step is resolution of the cointegrate into two separate replicons (Heffron et al, 1979; Gill et al, 1979; McCormick, et al, 1981). The resolvase catalyzes a site-specific recombination at the internal resolution site (IRS or *res*).

During the resolution process the cointegrate molecule is resolved to give recipient and donor replicons (McCormick, et al, 1981; Reed and Griendly, 1981). The TnpR protein is also a repressor that inhibits synthesis of both transposase and itself at the level of transcription (Gill et al, 1979; Chou et al, 1979).

3.11.1.1. Transposon Tn*3-I*

The Tn3-I spanning from nucleotide position 11424 to 138885 contains two oppositely transcribed open reading frames showing high homology to Tn3 family of transpoase (TnpA) and resolvase (TnpR) Therefore these two *orfs* are designated as *tnpA1* and *tnpR1,* which code for 583 amino acid long protein, TnpA1 and 189 amino acid long protein, TnpR1 respectively. TnpA1 of pPDL2 shows high similarity with TnpAs present in *Sphingomonas* and *Sphingobium,* especially to the TnpA sequence of (BAF03245) of *Sphingomonas* sp. KA1 (Fig. 3.27). Usually, full-length TnpAs coded by transposon Tn*3* are 985 amino acids. On comparison of TnpA1 of pPDL2 with that of the full-length transposases, a deletion of more than 400 amino acids is seen in the central region. The portion from amino acid 162 to 644 are found missing in TnpA-1 of pPDL2 (Fig. 3.27). The Tn*3* family transposases have a conserved domain which are typically seen in Transposase 7 super-family and is conserved in transposases coded by Tn*21*, Tn*1721*, Tn*2501*, Tn*3926* transposons. The domain contains an invariant triad, Asp689, Asp765, Glu895 (numbering as in Tn3) also reffered as D-D-35-E motif implicated in the catalysis of

transposition reactions by numerous transposases. Though there is deletion of more than 400 amino acids, the catalytic motif (D-D-35-E) is undisturbed as it is present at the C-terminus of Tn*3* transposase coded by *tnpA*-I of pPDL2 (Fig. 3.27). As substantial portion of TnpA is deleted, its functional status, despite of existing catalytic domain remains to be examined. In chapter-II, where transposition event is described an attempt is made to establish the functionality of TnpA.

```
TnpA1_PD      MTKRKHQLLTESERDQILAIPTDRDHLARLYTFEPSDIEIIGARRERRNQLGVALQLALL
TnpA_KA1      ------------MLAEHFDPSLDEREIARHFTLTRDDLELIASRRGDATRLGYAMLLLYL
TnpA_UT       --MARRRLVSLEIWAGHYDAPLDEREIARHYTLTSDDLEIVGRRRGDATRLGFAMLLLTM
TnpA_LB1      --MARRRLVSAEIWAGHYGAPLDEREIARHYTLTGDDLEIVGRRRGDATRLGYAMLLLYM
                                  . *. .:** :*:  .*:*::. **   .:** *: *  :

TnpA1_PD      RHPGITLAQLIQDRGAIPHDLAAFVAEQLGLHVTELANYAARDQTMTDHVRELAARLGLR
TnpA_KA1      RWP----GRVLEAGEAPPMPILAFVARQLNVSPAAWRDYARRDETRRTHLADLSRRFGHG
TnpA_UT       RWP----GRALEAGEVPPAPVLGYVARQLGVAPDAFADYARRDQTRREHLVEIRRSHGFR
TnpA_LB1      RWP----GRALEAGEVPPAPVLAYVAQQLGVAPEAFADYAHRDQTRREHLVEIRRSHGFR
              * *    .: ::   . *  : .:**.**.:      :** **:*   *: ::    *

TnpA1_PD      GPTRADIPFMVEAAARTAWATDKGMTIAMGVVTALREARILLPS----------------
TnpA_KA1      AFGRADFHTLVAFAMPIAQTVTQPSRLAGIIMDEMRRRRLLLPPVTIIEAIVRRARQQAG
TnpA_UT       IFDRDAFREVVAFSIPIAQTIIHPGQMAGVIVDELRRRQILLPSSSILEAVLRRARQQAE
TnpA_LB1      IFDRKAFHEVVAFSIPIAQTIVHPGQMAGVIVDELRRRQILLPSSSVLEAVLRRARQQAE
                 *  :  :*  :   * :  :   :*  ::  :*. ::***.

TnpA1_PD      ------------------------------------------------------------
TnpA_KA1      DMIHDVLAGDLGEPERTRLDALLSRRDDKSATWLSWLRNPPLSPAPRNILRLIERLDHVR
TnpA_UT       QLTYEVLTNGLRPDTLQDLDDLLARRTGQAATWLSWLRNASQSPAARNILRLIERLAYVR
TnpA_LB1      QLTYEVLTNGLRPDTLQGLDDLLARRTGQAATWLSWLRNAPQSPAARNILRLIERLTHIR

TnpA1_PD      -IGIEPP-----------------------------------------------------
TnpA_KA1      TLGIAASRAATIPQAAFDRIADEAARITPQHLAELPDKRRHAILAAAGIRLEESLTDAVL
TnpA_UT       ALGLDRGRADMIPASTFDRLADEGSRITPQHLGELNALRRHATLAATGIRLEEDLTDATL
TnpA_LB1      ALDLDRARADMIPALTFDRLADEGSRITPQHLGELNALRRHATLAAQGIRLEESLTDATL
               :.:

TnpA1_PD      ------------------------------------------------------------
TnpA_KA1      TMMDKFLGSMMRRAENRTKEKAIGTIRSLQAQLRLITGSCRTLLDARARGVDSLAAIGSI
TnpA_UT       TMFDKLLGSMVRRAENRTRDKALKTVRELQGHLRTLTGSCRILIDARTNGVDSLAQIEAL
TnpA_LB1      TMFDKLLGSMSRRAENRTRDKALKTVRELQGHLRTLTGSCRILIEARTNGVDSLAQIEAL

TnpA1_PD      ------------------------------------------------------------
TnpA_KA1      DWERLGTAVVNAELLIAPETIDRTAELIERQRSLRSVIGPFLNAFEFRGAGAVQGLLDAA
TnpA_UT       DWQRFAVSVEQAEVLSRPETVDRTAELIERHRTVKLFAGAFLNTFEFRGAGAVQGLLSAL
TnpA_LB1      DWQRFAVAVARAEVLGRPETVDRTAELIERHRTVKLFAGAFLNTFEFRGAGAVQGLLSAL

TnpA1_PD      -----------------LPLAARPSIFR--------------------------------
TnpA_KA1      RLVADIYRTGRRRFPDKPPLRFVPPSWRPFVLRDGEVVRAAYELCVLTQLRDRLRGGDIW
TnpA_UT       AIIAELYRTGKRRLPDRVPLRFVPSAWRPFILRDGIVDRAAYELCALSQLRERLRAGDIW
TnpA_LB1      TIIAELYRTGKRRLPDRVPLRFVPSAWRPFVLRDGIVDRAAYELCALSQLRERLRAGDIW
                                **   *. :*

TnpA1_PD      ------------------------------------------------------------
TnpA_KA1      VAESRQYRAFDSYLLPPATFEAMRARGPLPLAIETDFDKFIAGRRASLDTALERVTILAR
TnpA_UT       VAGSRQFRDFDSYLIPPATYAALREKGPLPLAIETDFERHIEERRTRLDTAIEQVTVLAR
TnpA_LB1      VSGSRQFRDFDSYLIPPATFDALREKGPLPLAIETDFDRHIEERRARLDTAIEQVTVLAR

TnpA1_PD      -------------------RTAAPP-----------------------------------
TnpA_KA1      QGELPQVRLDGNGLVISPLKAITPPDAEDMRRVAYDRLPRVKITDLLLEVDSWTGFSECF
TnpA_UT       QGELPQVRLDENGLIISPLKAATPPATEIARRAAYDRLPRVKITDLLLEVDAWTGFSECF
```

```
TnpA_LB1      QGELPQVRLDESGLIISPLKAATPPATEIARRAAYDRLPRVKITDLLLEVDAWTGFSECF
                                  :: :**

TnpA1_PD      --KRS-----------SVVLLK------------------RCLAYLPRY----ALYGEAL
TnpA_KA1      THRRSGRVADDRNALLTVILADGINLGLTRMAETCQGATLRQLAHLHDWHISEAAYGEAL
TnpA_UT       IHRRSGREADDRNALLTVILADGINLGLTRMAETCRGASLRQLAHLHDWHISEAAYGEAL
TnpA_LB1      IHRRSGREADDRNALLTVILADGINLGLTRMAETCRGASLRQLAHLHDWHISEAAYGEAL
                :**           :*:* .                  * **:*  :    * *****

TnpA1_PD      GRLIDVHRTVPLSALWGDGTTSSSDGQLFHAGGRGAAIGDINARNGNEPGVSFYTHVSDQ
TnpA_KA1      GRLIDVHRTVPLSALWGDGTTSSSDGQLFHSGGRGASIGDINARNGNEPGVSFYTHVSDQ
TnpA_UT       GRLINAHRTMPLAALWGDGTTSSSDGQQFHAGGRGAAIGDINARSGNEPGVAFYTHVSDR
TnpA_LB1      GRLIDAHRAMPLAALWGDGTTSSSDGQQFHAGGRGAAIGDINARSGNEPGVAFYTHVSDR
              ****:.**::**:************** **:*****:*******.******:*******:

TnpA1_PD      YDPFASRVIAATAGEAPYVLDGLLYHATGLSIEEHYTDTGGASDHVFGLMPFFGYRFAPR
TnpA_KA1      YDPFASRVIAATAGEAPYVLDGLLYHATGLSIEEHYTDTGGASDHVFGLMPFFGYRFAPR
TnpA_UT       YDPFASRVIAATAGEAPYVLDGLLYHQTGMTIEEHYTDTGGASDHVFGLMPFFGYRFAPR
TnpA_LB1      YDPFATRVIAATAGEAPYVLDGLLYQQTGLTIEEHYTDTGGASDHVFGLMPFFGYRFAPR
              *****:********************: **::******************************

TnpA1_PD      LRDLKDRRLHLLPGQEAGPLLAGMTGDPVAIGHVAAHWNELLRLTTSIRSGTTTASAMLR
TnpA_KA1      LRDLKDRRLHLLPGQEAGPLLAGMTGDAVAIGHVADHWDELLRLTTSIRSGTTTASAILR
TnpA_UT       LRDIKERRLHLLPGQESGPLLAGMTTEPIALGHVAAHWDELLRFATSIRTGTVTASAMLR
TnpA_LB1      LRDIKQRRLHLLPGQEAGPLLAGMTAEPIALGHVAAHWDELLRFATSIRTGTATASAMLR
              ***:*:**********:******** :.:*:**** **:****::****:**.****:**

TnpA1_PD      RLSAYPRQNGLALALREVGRIERSIFMLDWLRDLDLRRRTQAGLNKGEARNALARALFFN
TnpA_KA1      RLSAYPRQNGLALALREVGRVERSIFMLDWLRDLDLRRRTQAGLNKGEARNALARALFFN
TnpA_UT       RLSGYPRQNGLALALRELGRLERSIFMLDWLRDIDLRRRTQAGLNKGEARNALARALFFN
TnpA_LB1      RLSAYPRQNGLALAMRELGRLERSIFMLDWLRDIDLRRRTQAGLNKGEARNALARALFFN
              ***.**********:**:**:************:**************************

TnpA1_PD      QLGELRDRRPCGRHDAIWILPASGPLFPVKIPNDEVAHFNQLGELRDRRFENQTYRASGL
TnpA_KA1      QLGELRDRR---------------------------------------FENQAYRASGL
TnpA_UT       QLGELRDRR--------------------------------------FENQTYRASGL
TnpA_LB1      QLGELRDRR--------------------------------------FENQTYRASGL
              *********                                       ****:******

TnpA1_PD      NLLVAAIILWNTRYLERAVGALAIPDDVARHIAPLGWEHISLTGDYRWNVESRPDPGQLR
TnpA_KA1      NLLVAAIILWNTRYLEQAVGTLSIPGNIARHIAPLGWEHISLTGDYRWNVESRPDPGKLR
TnpA_UT       NLLVAAIILWNTRYLEMALADIGTPDEIARHVAPLGWEHISLTGDYSWNVEDRPDPDALR
TnpA_LB1      NLLVAAIILWNTRYLEVALADIGTPDEIARHVAPLGWEHISLTGDYSWNVEDRPDPDVLR
              **************** *:. :. *.::***:************** ****.****. **

TnpA1_PD      PLRTPSSLLAA
TnpA_KA1      PLRTPSSLLAA
TnpA_UT       PLRAVSSLLAA
TnpA_LB1      PLRAISSLLAA
               ***: ******
```

Fig. 3.27. Multiple alignment TnpA1 of pDPL2 of *Flavobacterium* with TnpAs reported from *Sphingomonas* (TnpA KA1, TnpA LB1) and *Sphingobium japonicum* UT26S (TnpA UT1). Identical residues and gaps are shown with an (*) and (-) respectively. Similar residues are shown with (:) and (.) symbols. The D-D-35-E motif is highlighted in yellow.

3.11.1.2. TnpR-I

Unlike TnpA-I the TnpR-I encoded by first copy of Tn3 (Tn3-1) shows high homology (99%) to resolvase of *Sphingomonas japonicum* UT26S (Fig. 3.29). Both these resolvases of Tn*3* show conserved features of the serine recombinase family such as the C-terminal DNA binding HTH motif and a conserved N-terminal catalytic domain having a serine residue in its active site (Fig. 3.30). As mentioned

before, resolvases are involved in resolution of co-integrates formed during transposition. However, in TnpA-I substantial portion is deleted formation of such co-integrate is questionable. In such a scenario the role of TnpR in transposition of *opd* cluster remains to be established.

3.11.2. Transposon Tn*3-II*

3.11.2.1. TnpA-II

As shown in figure the second copy to Tn*3* transposon starts at nucleotide position 15639 and ends at 16973nt. In between these two transposons, Tn*3*-1 and Tn*3*-2 a sequence that codes for an open reading frame of 315 amino acids is identified. The sequence of Orf315 has shown no homology to any other protein found in database. The TnpA coded by transposon copy II is designated as TnpA-II which has an opposite transcription orientation when compared to TnpR-II. Translated sequence of *tnpAII* when compared with full-length Tn*3* transposase has shown large deletions in the central region. The conserved catalytic motif having catalytic triad (D-D-E) is also not seen in TnpA-II. Therefore, the second transposase of Tn*3* family, TnpAII, present in pPDL2 is assumed as an inactive transposase (Fig. 3.28).

TnpRII (*orf16*) present in the second copy of Tn*3* spanning nucleotide postions14744 to 15325 is designated as *tnpR*-2 as it shows high homology (99%) to resolvase of *Sphingomonas japonicum* UT26S (Fig. 3.29). The resolvase designated as TnpR-II of Tn*3* show conserved features of the serine recombinase family such as the C-terminal DNA binding HTH motif and a conserved N-terminal catalytic domain having a serine residue in its active site (Fig. 3.30). Further, the sequence of TnpR-I and TnpR-II have shown high similarity to TnpR sequence found in transposon Tn3 elements

identified in *Sphingobium japonicum* UT26S. Considering the extensive homology found throughout the protein and

```
Tn3S_KA1        MLAEHFDPSLDEREIARHFTLTRDDLELIASRRGDATRLGYAMLLLYLRWPGRVLEAGEA 60
Tn3PD2          ----------------------------------------MLLLYMRWPGRALEAGEV 18
                                                         *****:*****.*****.

Tn3S_KA1        PPMPILAFVARQLNVSPAAWRDYARRDETRRTHLADLSRRFGHGAFGRADFHTLVAFAMP 120
Tn3PD2          PPAPVLAYVAQQLGVAPEAFADYAHRDQTRREHLVEIRRSHGFRIFDRKAFHEVVAFSIP 78
                ** *:**:**:**.*:* *: ***:**:*** **.:: * .*.  *.*  ** :***::*

Tn3S_KA1        IAQTVTQPSRLAGIIMDEMRRRRLLLPPVTIIEAIVRRARQQAGDMIHDVLAGDLGEPER 180
Tn3PD2          IAQTIVHPGQMAGVIVDELRRRQILLPSSSVLEAVLRRARQQAEQLTYEVLTNGLR---- 134
                ****:.:*.::**:*:**:***::***. :::**::******* :: ::**:..*

Tn3S_KA1        TRLDALLSRRDDKSATWLSWLRNPPLSPAPRNILRLIERLDHVRTLGIAASRAATIPQAA 240
Tn3PD2          ------------------------------------------------------------

Tn3S_KA1        FDRIADEAARITPQHLAELPDKRRHAILAAAGIRLEESLTDAVLTMMDKFLGSMMRRAEN 300
Tn3PD2          ------------------------------------------------------------

Tn3S_KA1        RTKEKAIGTIRSLQAQLRLITGSCRTLLDARARGVDSLAAIGSIDWERLGTAVVNAELLI 360
Tn3PD2          ------------------------------------------------------------

Tn3S_KA1        APETIDRTAELIERQRSLRSVIGPFLNAFEFRGAGAVQGLLDAARLVADIYRTGRRRFPD 420
Tn3PD2          ------------------------------------------------------------

Tn3S_KA1        KPPLRFVPPSWRPFVLRDGEVVRAAYELCVLTQLRDRLRGGDIWVAESRQYRAFDSYLLP 480
Tn3PD2          ------------------------------------------------------------

Tn3S_KA1        PATFEAMRARGPLPLAIETDFDKFIAGRRASLDTALERVTILARQGELPQVRLDGNGLVI 540
Tn3PD2          PDTLQACKLQG--------------------------------------------------- 145
                * *::* : :*

Tn3S_KA1        SPLKAITPPDAEDMRRVAYDRLPRVKITDLLLEVDSWTGFSECFTHRRSGRVADDRNALL 600
Tn3PD2          ------------------------------------------------------------

Tn3S_KA1        TVILADGINLGLTRMAETCQGATLRQLAHLHDWHISEAAYGEALGRLIDVHRTVPLSALW 660
Tn3PD2          ------------------------------------------------------------

Tn3S_KA1        GDGTTSSSDGQLFHSGGRGASIGDINARNGNEPGVSFYTHVSDQYDPFASRVIAATAGEA 720
Tn3PD2          ------------------------------------------------------------

Tn3S_KA1        PYVLDGLLYHATGLSIEEHYTDTGGASDHVFGLMPFFGYRFAPRLRDLKDRRLHLLPGQE 780
Tn3PD2          ---------------------------------MPFFGYRFAPRLRDIKQRRLHLLPGQE 172
                                                  **************:*:**********

Tn3S_KA1        AGPLLAGMTGDAVAIGHVADHWDELLRLTTSIRSGTTTASAILRRLSAYPRQNGLALALR 840
Tn3PD2          AGPLLAGMTAEPIALGHVAAHWDELLRFATSIRTGTATASAMLRRLSAYPRQNGLALAMR 232
                *********.:.:*:**** *******::****:**:****:****************:*

Tn3S_KA1        EVGRVERSIFMLDWLRDLDLRRRTQAGLNKGEARNALARALFFNQLGELRDRRFENQAYR 900
Tn3PD2          ELGRLERSIFMLDWLRDIDLRRRTQAGLNKGEARNALARALFFNQLGELRDRRFENQTYR 292
                *:**:************:****************************************:**

Tn3S_KA1        ASGLNLLVAAIILWNTRYLEQAVGTLSIPGNIARHIAPLGWEHISLTGDYRWNVESRPDP 960
Tn3PD2          ASGLNLLVAAIILWNTRYLEVALADIGTPDEIARHVAPLGWEHISLTGDYSWNVEDRPDP 352
                ******************** *:. :. *.:****:************** ****.****

Tn3S_KA1        GK-----------LRPLRTPSSLLAA----- 975
Tn3PD2          GCPAATARHQFVARRVTFTIRSRLACGCVTF 383
                *             *   *  * **.
```

Fig. 3.28. Pairwise alignment of Transposase TnpA-2of pPDL2 of *Flavobacterium* sp. ATCC 27551 is aligned with similar ssequences coded by *Sphingomonas* sp. _KA1. Identical residues and gaps are shown with an (*) and (-) respectively. Similar residues are shown with (:) and (.) symbols.

important catalytic site residues (Fig 3.29 and 3.30), the TnpR sequences coded by these two copies of Tn*3* elements can certainly be considered functional. However, in the absence of functional transposases, their presence in plasmid pPDL2 makes no functional significance. However, if transposition event occurs due to existence of a functional TnpA coding Tn*3* element located either in the chromosome or another indigenous plasmid, the resolvases found on plasmid pPDL2 can act to resolve the generated cointegrate.

```
TnpR2          --MWRSAMRASRRQTKKALTPARRAERGPGATRIFDDHASGAKADRPGLAEALAYLRSGD 58
TnpR_Sj_UT     -MALIGYARVSTADQKLSLQLD--ALNAAGCDRIFDDHASGAKADRPGLAEALAYLRSGD 57
TnpR1          MGGILGYARVSTGDQDVAGQTMR--LENAGAIKVFTDVISGKSMERPGLAELIAYARKGD 58
                    .  *.*  : . :         . .*. ::* *  ** . :****** :** *.**

TnpR2          TLVVWKLDRLGRSMSHLIEKVGELATRGIGFRSLTENIDTTTSGGMLVFNIFGSLAQFER 118
TnpR_Sj_UT     TLVVWKLDRLGRSMSHLIEKVGELATRGIGFRSLTENIDTTTSGGMLVFNIFGSLAQFER 117
TnpR1          TLAVVRLDRLGRSLTELLATVETLRSQGIALLSLEEKIDTSSAAGELIFHVFGAIAHFER 118
               **.* :*******::.*: .*  * ::**.: ** *:***:::.* *:*::**::*:***

TnpR2          DLIRERTHAGLKAARERGRPGGRRPVVTPDKLRKAREHIASGLTVREAAARLKIGKTALY 178
TnpR_Sj_UT     DLIRERTHAGLKAARERGRPGGRRPVVTPDKLRKAREHIASGLTVREAAARLKIGKTALY 177
TnpR1          RLISERTRDGIAAARAKGKQPGRQPLDMS-KVDAAIKLVEARISPTEAARQLGIGRSTIY 177
                ** ***: *: *** :*:  **:*:  . *:  * : : : ::  *** :* **::::*

TnpR2          KALEATEKNTKSQRSRSVRSRADK 202
TnpR_Sj_UT     KALEATEKNTKSQRSRSVRS---- 197
TnpR1          REMRRLGVERPA------------ 189
```

Fig. 3.29. Multiple alignments of resolvases of Tn*3* transposon. The resolvase of TnpRI and TnpRII of plasmid pPDL2 of *Flavobacterium* sp. ATCC 27551 are compared with TnpR found in *Sphingobium japonicum* UT26S (TnpR Sj UT). Identical residues and gaps are shown with an (*) and (-) respectively. Similar residues are shown with (:) and (.) symbols.

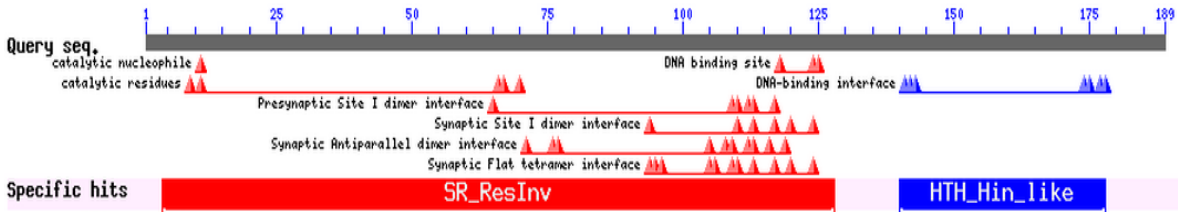


Fig. 3.30. Resolvases of pPDL2 showing N-terminal catalytic domain and C-terminal DNA binding HTH motif.

3.11.3. Transposon Tn3 specific terminal repeats

In pPDL2, two direct repeats containing characteristic features of Tn*3* family of transposons are present at positions 11354-1138 and 37419-37453. As mentioned earlier, transposon Tn*3* is present in many strains of the *Sphingomonas* sps. Analysis of their terminal repeats has showed minor variation in their length (Fig. 3.31). In plasmid pSY3

```
pPDL2RT       GGGGTCACTACACGAAAGTGCATTTTACGTACGCT---------------
35
pPDL2LT       GGGGTCACTACACGAAAGTGCATTTTACGTACGCT---------------
35
pCAR3         GGGGTCACTACACGAAAGTGCATTTTACGTACGCT---------------
35
PCAR3.2       ---------ACACGAAAGTGCATTTTACGTACGCT---------------
26
pSY3          --------------GGGGT-CACTACACGAAAGTGCATTTTACGTACGCT
35
Ibu-2         --------------GGGGT-CACTACACGAAAGTGCATTTTACGTACG--
33
UT26s         --------------GGGGT-CACTACACGAAAGTG---------------
20
pCHQ1         --------------GGGGT-CACTACACGAAAGTG---------------
20
                                   ** ** *  *** * *
```

Fig. 3.31. Terminal repeats of Tn*3* transposon found on plasmid pPDL2 (pPDL2RT and pPDL2LT) of *Flavobacterium* sp. ATCC 27551, plasmid of *Sphingomonas* sp KA1 (pCAR3), *Sphingomonas chungbukense* strain DJ77 (pSY3), *Sphingomonas* sp. Ibu-2 and *Sphingobium japonicum* chromosome UT26S (UT26S) and plasmid (pCHQ1).

isolated form *Sphingomonas chengbukense, a* 35bp terminal repeat was seen with 4 mismatches. The Tn*3* is also present both on the chromosome and plasmid pLB1 of *Sphingobium japonicum* UT26S. These are only 20 bp long. Terminal repeats in the Tn*3* element found in *Sphingomonas* sp., and *Sphingobium japonicum* UT26Ss suggesting that the first twenty base pairs of the repeats are enough for successful transposition of transposon Tn*3*.

3.11.4. The y4qE element

The ORF42 present at the region spanning the nucleotide positions 37745 to nucleotide position 38847 is designated as *y4qE* as it codes for a 374 amino acid long protein that shows 59% identity with the putative transposase y4qE (EFO28627) of *Roseibium* sp. *Trich*SKD4 and 57% identity with IS*116*/IS*110*/IS*902* family protein (ABS70259) of *Xanthobacter autotrophicus*. The IS*110* family elements are usually flanked by 51 bp terminal repeats (García-Trigueros et al, 2007). However, no repeats of 51 bp are noticed in the flanking region of y4qE transposase. In the absence of such terminal repeats the functional status of y4qE element is questionable. If the mobile elements found on pPDL2 are considered, Tn*3* alone appears to have TnpA with proper catalytic domain. The terminal repeats that are specific to Tn*3* are only seen flanking the degradative module. Experimentlly the transposition event needs to be determined to assess its status in transposition.

3.12. Discussion

The sequence of plasmid pPDL2 contains both features of a typical plasmid and Integration Mobilizable Element (IME). Existence of a double stranded replication origin (*oriV*) and a well defined *par* locus that is involved in maintenance of a plasmid together with a toxin anti-toxin domain perfectly justify assigning pPDL2 the status of a well defined plasmid. Historically pPDL2 was isolated as part of investigation aimed at understanding the molecular basis for degradation of organophosphate insecticides used in modern agricultural practices. Subsequent studies conducted on this plasmid have shown organophosphate degrading (*opd*) gene as part of a complex transposon-like element (Siddavattam et al, 2003). The present study is designed to understand horizontal mobility of *opd* gene, which is considered to be all most certain due to

existence of identical *opd* genes in bacterial strains isolated from bacterial strains belonging to diverse taxonomic groups. The complete sequence has indeed added an interesting dimension to genetics and biology of organophosphate degradation. In addition to the well defined replication and maintenance modules, the plasmid pPDL2 contained a well defined degradative module with a typical features of a complex transposon. The degradative module included organophosphate degrading (*opd*) gene, dioxygenase gene, and a *meta* fission product hydrolase gene and a β-ketoadipate hydrolase coding sequence. As shown in figure 3.25 transposon Tn*3* element and y4qE element are found flanking this degradative module. The perfect IR sequences that serve as target sites to the Tn*3* coded transposase have identified at the extreme ends. If the arrangement is seen it clearly indicates that *opd* gene is part of a well structured mobile element designed to contribute for the lateral transfer of *opd* information among soil bacteria. If the plasmid maintenance and degradative modules are alone taken into consideration it clearly assigns pPDL2 the status of a typical bacterial plasmid. However, it also contains a well defined integration module, which includes an integrase, CopG and phosphoglycerate mutase. The transcriptional arrangement of these genes and presence of *attP* site that show similarity to the chromosomally located *attB* site found at the 3' end of serine tRNA gene perfectly justifies the capability of pPDL2 to integrate into and excise from the chromosome. If these features are seen together with existence of *oriT*, the pPDL2 should be given a status of an integrative mobilizable element (IME). Before presenting the structural status of pPDL2 a brief discussion is given on the Integrative Conjugative Elements (ICEs) so that a structural comparison can be made between ICEs and plasmid pPDL2.

The conjugation systems encoded by chromosome-borne mobile genetic elements (MGEs) were recently identified. Such elements are often referred to as integrative and conjugative elements (ICEs)(Burrus et al, 2002; Wozniak and Waldor, 2010). ICEs are self-transmissible MGEs that encode for conjugation machinery as well as intricate regulatory systems to control their excision from the chromosome and their conjugative transfer (Salyers et al, 1995; Osborn and Boltner, 2002; Burrus and Waldor, 2004). The ICEs encompass all self-transmissible integrative and conjugative mobile elements regardless of their mechanisms of integration or conjugation (Wozniak and Waldor, 2010). These include elements that are commonly characterized as conjugative transposons, which often integrate into the host chromosome with minimal sequence specificity and, consequently, are capable of both intracellular and intercellular transfer (Burrus and Waldor, 2004). The Tn*916* of *Enterococcus faecalis* (Velikonja et al, 1994) and CTnDOT in *Bacteriodes thetaiotaomicron* are the first known MGEs with ICE-like properties (Shoemaker et al, 2001). Certain chromosomal elements which are previously classified as genomic islands also have properties of ICEs. Xenobiotic island, ICE*clc*^{B13} of *Pseudomonas knackmussii* sp. strain B13 and symbiotic island (Ravatn et al, 1998), ICE*MlSym*^{R7A} of *Mesorhizobium loti* are examples of genomic islands which have ICE properties (Ramsay et al, 2006).

3.12.1. Structure and Function of ICEs

Integrative and conjugative elements (ICEs) typically have modular structures in which genes with related functions are clustered together (Mohd-Zain et al, 2004; Juhas et al, 2007; Roberts and Mullany, 2009; Wozniak and Waldor, 2009). All ICEs have three

simple, distinct functional modules designated as i) maintenance, ii) dissemination and iii) regulation modules.

i) Maintenance modules

All ICEs encode an integrase (*Int*) that enables their integration into the host chromosome. The process of integration requires integrase and no additional factors are required. But for excision to occur, additional factors, known as recombination directional factors are required along with integrase. Integrases determine the site and frequency of ICE excision. Moreover, regulation of *int* expression is one of the key means of controlling ICE transmission. Most of the known ICE integrases are members of the tyrosine recombinase family (Argos et al, 1986). The best studied recombinase of this family is the Int encoded by phage λ. The phage Integrase uses a topoisomerase I-like mechanism to promote site-specific recombination between identical or near identical sequences in the host chromosome (referred to as *attB* sites) and the phage chromosome (the *attP* site). The strand exchange reactions catalysed by Integrase do not require a high-energy cofactor such as ATP, and no sequence duplication or deletion results from recombination. Integration of ICEs usually occurs into specific sites known as primary sites of integration (Table 3. 3). However, integration may also occur at secondary sites in absence of the primary sites (Burrus and Waldor, 2003; Lee et al, 2007).

Table 3. 4. ICEs of various sources and their sites of insertion

ICE	Host	Size (kb)	Site of insertion	Phenotype	Reference
SXT	*Vibrio cholera*	99.5	*prfC*	Cm^R, Sm^R, SXT^R	Beaber et al, 2002

R391	*Providencia rettgeri*	89	*prfC*	Hg^R, Kn^R	Boltner et al, 2002
ICE*Bs1*	*Bacillus subtilis*	20	$tRNA^{Leu}$ gene	None Known	Burrus et al, 2002
PAPI-1	*Pseudomonas aeruginosa*	108	$tRNA^{Lys}$ gene	Virulence factors and regulation of biofilm formation	
ICE*clc*^{B13}	*Pseudomonas knackmussii*	105	$tRNA^{Gly}$ gene	3-chlorobenzoic acid degradation	Ravatn et al, 1998
ICE*MISym*^{R7A}	*Mesorizobium loti*	502	$tRNA^{Phe}$ gene	Symbiosis with Lotus corniculatus involving nodulation and nitrogen fixation	Sullivan et al,2002
ICE*Hin1056*	*Haemophilus influenza*	49.4	$tRNA^{Leu}$ gene	Tet^R, Cm^R, Amp^R	Mohd-Zain et al, 2004
pSAM2	*Streptomyces ambofaciens*	10.9	$tRNA^{Pro}$ gene	None Known	Pernodet et al,1984
Tn*916*	*Enterococcus faecalis*	18	AT-rich regions	Tet^R	Clewell et al, 1995
CtnDoT	*Bacteroides spp.*	65	GTANNTTTTGC	Tet^R, Erm^R	Cheng et al, 2000
TnGBS2	*Streptococcus agalactiae*	33.5	Intergenic regions upstream of RNA polymerase sigma A promoters	None known	Brochet et al, 2009

Tn*4371*	*Ralstonia sp. A5*	55	TTTTTCAT	Biphenyl degradation	Toussaint et al, 2003
Tn*5397*	*Clostridium difficile*	21	Single site	Tc^R	Wang et al, 2000
Tn5252	*Lactococcus lactis*	70	TTTTTG	Sucrose utilization and nisin synthesis	Vijayakumar et al, 1993
pRS01/sex factor	*Lactococcus lactis*	48.4	Single site	Tellurium resistance	Gasson et al, 1995
Tn5801	*Staphlyococcus aureus*	25.8	3' end of a gene encoding GMP synthase	Tc^R	Kuroda et al, 2001

ii) Dissemination modules

Like conjugative plasmids, ICEs disseminate *via* conjugation. ICEs contain the genes that specify the synthesis of the 'mating machinery' that enables intimate contact between donor and recipient cells and that delivers DNA to the recipient cell. The dissemination modules of ICEs are also diverse. In most described cases, ICEs are thought to transfer as single-stranded DNA with few exceptions (pSAM2 from *Streptomyces mbofaciens)* (Grohmann et al, 2003). In some cases, such as in the *V. cholerae*-derived ICE SXT, the transfer genes bear similarity to those found in well-characterized conjugative plasmids such as the F plasmid (Beaber et al, 2003). In other cases, such as in Tn*1549* from *Enterococcus* spp., ICE transfer genes appear to be distantly related to those found in Gram-positive conjugative plasmids (Garnier et al, 2000). Finally, the genes required for the transfer of some ICEs, such as the Bacteroides-derived CTnDOT, are for the most part unrelated to previously characterized transfer genes (Bonheyo, 2001).

iii) Regulation modules

The genes and the mechanisms that regulate ICE transfer are just beginning to be defined; however, it is already apparent that ICE regulation modules are very diverse (Wozniak and Waldor, 2010). Transfer of both Tn*916* and CTnDOT is induced by sub-inhibitory concentrations of tetracycline (Slayers et al, 1995). The expression of the *int* gene of the *clc* element from *Pseudomonas* sp. was recently found to be stimulated by growing the bacterium on 3-chlorobenzoate-containing medium, but not by high cell density, heat shock, osmotic shock, UV irradiation or ethanol stress (Sentchilo et al, 2003). Interestingly, aromatic chlorinated compounds, such as 3-chlorobenzoate, are substrates for the degradation pathway encoded by *clc*. It was hypothesized that a metabolite of 3-chlorobenzoate modulates interactions of a putative activator and repressor with the integrase-encoding gene promoter (Sentchilo et al, 2003). Thus for three ICEs, Tn*916*, CTnDOT and *clc*, specific compounds in the environment seem to trigger their dissemination thereby conferring upon new hosts the ability to resist to or metabolize these compound. *V. cholerae* derived ICE, SXT encodes SetR, an orthologue of the phage lambda repressor CI and transfer of this ICE is regulated by the SOS response.

Integrative mobilizable elements (IMEs) are non-self-transmissible elements which can excise from chromosome, and transferred to a new host with the help of conjugative functions provided in *trans* by other elements, through the formation of a covalent closed circular molecule and integrate into the chromosome of the host. IMEs are reported from *Salmonella* and *Bacteroides* species (Doublet et al, 2005; Shoemaker et al, 1996). In *Salmonella*, the genomic island, SGI1 is not self-transmissible but can be transferred from a donor strain of *Salmonella* enteric to non-SGI1 containing S. *enteric*

and *Escherichia coli* recipient strains (Doublet et al, 2005). The functions necessary for transfer are provided in *trans* by the donor strains having the conjugative *IncC* plasmid R55 (Doublet et al, 2005). *Bacteroides* species have mobilizable insertion elements known as non-replicative Bacteroides units (NBUs). Four members of the NBU element family (NBU1, NBU2, NBU3, and Tn4555) are shown to be mobilizable *via* a covalently closed circle intermediate (Li et al, 1993, Li et al, 1995). This circle does not replicate but can be transferred by conjugation, starting from an internal transfer origin (*oriT*). NBUs possess *mob* gene but are not self-transmissible, nor can they excise on their own. Both excision and mobilization of NBUs require transacting functions provided by a Bacteroides conjugative transposon (CT) (Shoemaker and Slayers, 1988; Steven et al, 1990; Stevens et al, 1992).

As mentioned in earlier sections all modules required for maintenance and distribution of pPDL2 as an ICE are present in its primary sequence. However, the genes responsible for the formation of mating pair are not present. Instead, genes responsible for its mobilization and integration are present along with the *attP* site required to mediate site specific recombination. The replication initiator, RepB found on pPDL2 can initiate the replication process in rolling circle mode. In addition to *repB* two potential *oriT* sites needed to generate relaxed pPDL2 required to mediate mobilization process are identified on plasmid pPDL2.

The plasmid, pPDL2 of *Flavobacterium* sp. ATCC 27551 has an Integrase coding sequence which can perform its integration and excision into and out of the chromosome. As typically seen in ICE/IMEs,a a putative regulator *copG* is found translationally linked to integrase coding sequence. Further, the Integration Host Factor

(IHF) binding sites have been identified in the upstream region of *copG*, the protein shown to play a major role in excision of the integrative elements (Nash and Robertson, 1981; Bushman et al, 1985). Presence of such *cis*-element in fact strengthens the proposal of existence of an integration and excision event between chromosome and pPDL2. Further, *copG, int* and *pgm* genes are found clustered in pPDL2. Existence of linkage between *copG*, *int* and *pgm* genes, which are named in the present study as CIP unit suggests functional relevance (Fig. 3.17). In pPDL2 there exists, two sets of CIP units, without having an absolute identity. The PGM is shown to have a HP-PEM domain involved in dephosphorylation of histidine phosphates. Presence of such domain indicated regulation of integration /excision event through means of signal transduction. However, further studies are needed to prove existence of such regulation. It is not yet known if growth of *Flavobacterium* sp. ATCC 27551 in presence of OP compounds or its intermediates stimulates the transfer of pPDL2.

What is pPDL2?

After examining the complete sequence of plasmid pPDL2 and looking at the various modules present in the sequence an attempt is made to give a correct name to pPDL2. Historically, it's identified as a plasmid containing organophosphate degrading (*opd*) gene. The replication origin and well conserved replication initiation proteins RepA, maintenance module containing toxin and anti-toxin protein coding sequences justify calling it as a typical bacterial plasmid. However, it also contains a number of features that justify grouping it under Integrative Mobilizable Element. Existence of an integration module (Fig. 3.17 and 3.21), attachment site *attP* and a module that contribute for the mobilization of pPDL2 are typical features of an IME (Juhas et at, 2009). In the light of

these facts how should pPDL2 be called? A plasmid or an IME? It reminds mythological situation, a Chimera and Ganesha….. though considered powerful Gods, for a common man, they neither have complete human nor animal features. Most of the simplest things are hard to explain.

3.13. Conclusions

1. *Flavobacterium* sp. ATCC 27551 contains four indigenous plasmids designated as pPDL1, pPDL2, pPDL3 and pPDL4.

2. The indigenous pPDL2 alone contains *opd* gene and it is rescue cloned into *E.coli pir*-116 cells.

3. Plasmid library was constructed by cloning *Eco*RI and *PstI* fragments of pPDL2 in pBluescript KS II.

4. Complete sequence was determined for the 39.75 kb pPDL2.

5. The GC content of pPDL2 (61.76%) is found to be very close to the GC content of plasmids pSWIT02 and pUT2 of *Sphingomonas wittichi* RW1 and *Sphingobium japonicum* UT26S respectively.

6. Nucleotide sequence of plasmid pPDL2 of *Flavobacterium* sp. ATCC 27551 showed maximum homology to the indigenous plasmids pUT1, pCHQ1 of *Sphingobium japonicum* UT26S and pSWIT01of *Sphingomonas* sp.

7. Sequence analysis of pPDL2 has revealed presence of 42 ORFs out of which only 18 are hypothetical and the remaining 24 ORFs have shown strong homology to the well characterized proteins found in NCBI database

8. Open reading frames present on pPDL2 are organized into modules, such as (i) the replication and partition module, (ii) mobilization module, (iii) integrase module, (iv) degradation module and (v) mobile genetic elements.

Table 3.3 Open reading frames present on plasmid pPDL2

Generic Name	Position	Conventional Name	Putative function	Homologous protein	Amino acid identity (%)	GenBank accession no.
ORF1	954-1682	*orf242*	Unknown	Hypothetical protein	31	ZP_01623203
ORF2	1185-2067 C	*orf294*	P45O	β- component of Protocatechuate dioxygenase of *Xanthomonas axonopodis* pv. citri str. 306	78	gb|AAM35766.1
ORF3	2527-2880	*orf117*	Hydrolase	alpha-beta hydrolase fold protein of *Commamonas testosteroni* S44		gb|EFF44050.1|
ORF4	2915-3460	*orf181*	Hydrolase	Hydrolase of *Xanthomonas fuscans* subsp. aurantifolii str. ICPB		
ORF5	3433..4887	*orf484*	Transporter	MFS of *Asticcacaulis excentricus* CB 48	72	gb|ADU14055.1
ORF6	5075..5956	*orf293*	Antibiotic resistance	Beta-lactamase domain protein of *Methylobacterium nodulans* ORS 2060	48	gb|ACL58162.1
ORF7	5986-6786	*orf266*	Unknown	Hypothetical protein Swit_1907 *Sphingomonas wittichii* RW1	41	gb|ABQ68267.1
ORF8	6941-8464	*orf507*	Transposase(*istA*)	Transposase *Aurantimonas manganoxydans* SI85-9A1	59	gb|EAS49905
ORF9	8407-8988 C	*orf193*	Acetyl transferase	dihydrolipoamide acetyltransferase *Legionella pneumophila* str. Paris	22	emb|CAH12611.1
ORF10	8461-9300	*orf279*	Resolvase (*istB*)	IstB domain protein ATP-binding protein Nitrosomonas sp. AL212	78	gb|EET30478
ORF11	9567-10664	*orf365*	OP Hydrolase (*opd*)	Parathion hydrolase of P. diminuta	99	gb|AAA24930

ORF12	10696-11613	*orf305*	Aromatic hydrolase	Putative aromatic hydrolase		
ORF13	11424-13175 C	*orf583*	Transpoase (*tnpA*)	TnpA transposase Sphingomonas sp. KA1	86	dbj|BAF03245
ORF14	13316-13885	*orf189*	Resolvase (*tnpR*)	Resolvaseof *Sphingomonas* sp. KA1	96	dbj|BAE75870
ORF15	13886..14833	*orf315*	Unknown	Hypothetical protein SphchDRAFT_3708 *Sphingobium chlorophenolicum* L-1	85	gb|EFN09107
ORF16	14812-15186	*orf124*	Recombinase (*tnpR*)	Putative site-specific recombinase	63	gb|ACJ63562
ORF17	15640-16077	*orf145*	Transposase (*tnpA*)	Tn3 family transposase of plasmid pLB1	98	YP_740316
ORF18	15958-16902C	*orf314*	Unknown	Hypothetical protein of *Acinetobacter baumannii* AYE	40	emb|CAM88370
ORF19	16255-16974	*orf239*	Transposase (*tnpA*)	Tn3 family transposase of plasmid pLB1	98	YP_740316
ORF20	16878-17534	*orf218*	Plasmid replication initiation (*repAa*)	RepA of *Sphingobium japonicum*	96	dbj|BAI99177
ORF21	18125-19168	*orf347*	Plasmid replication initiation (*repAb*)	RepA of *Sphingobium japonicum*	97	dbj|BAI99177
ORF22	21246-21893 C	*orf215*	Partitioning (*parA*)	ParA of *Sphinogobium japonicum*	99	dbj|BAI99179
ORF23	22887-23507	*orf206*	Hypothetical protein	Hypothetical protein of *Sphingomonas wittichi* RW1	35	gb|ABQ66545
ORF24	23494-23922	*orf142*	Plasmid stability (*relB*)	RelB antitoxin of *Sphingobium japonicum*		dbj|BAI98982
ORF25	23810-24058 C	*orf82*	Unknown	Hypothetical protein		dbj|BAI99186
ORF26	25546-26166	*orf206*	Unknown	Hypothetical protein	95	dbj|BAI99186
ORF27	26469-26717 C	*orf82*	Hypothetical protein	Hypotheitical protein XAUC 31650	No significant	

					homology	
ORF28	26836-27483	*orf215*	Enzyme of glycolytic pathway (*pgm*)	Phosphoglycerate mutase family protein	95	dbj|BAI99186
ORF29	27679-28086 C	*orf135*	Unknown	Hypothetical protein	70	dbj|BAI99187
ORF 30	26153-26581	*orf142*	Antitoxin (*relB*)	RelB antitoxin of *S. japonicum* UT 26	96	dbj|BAI98982
ORF31	28583-29569 C	*orf328*	Invertase/ recombinase like protein (*int*)	Phage integrase family protein of *Sphingomonas wittichi* RW1	86	dbj|BAI98979
ORF32	29566-29961 C	*orf131*	Transcriptional factor (*copG*)	CopG/MetJ/Arc family protein	65	gb|ABQ71225
ORF33	30013-30390 C	*orf125*	Unknown	Hypotehtical protein	-	-
ORF34	30675-31184	*orf169*	Transcriptional factor (*copG*)	CopG/MetJ/Arc family protein	65	gb|ABQ71225
ORF35	31181-32027	*orf282*	Invertase/recombinase like protein (*int*)	Phage integrase family protein of *Sphingomonas wittichi* RW1	94	dbj|BAI98979
ORF36	32426-33014	*orf196*	Enzyme of glycolysis (*pgm*)	Phosphoglycerate mutase family protein	92	dbj|BAI99186
ORF37	33023-33305	*orf94*	Unknown	Conserved hypothetical protein of *Methylosinus trichosporium* OB3b	73	gb|EFH01001
ORF38	33296-33721	*orf141*	Signal transduction protein (*pilT*)	PilT domain containg protein of *Sphingomonas wittichi* RW1	59	gb|ABQ68989
ORF39	34228-34863	*orf211*	Unknown	hypothetical protein SJA_P1-00320 *Sphingobium japonicu* UT26	91	dbj|BAI99180
ORF40	34944-35579	*orf211*	Catalyzed conversion of cobyrinic acid to cobyrinic acid diamide	Cobyrinic acid ac-diamide synthase *Thauera* sp. MZ1T	58	gb|ACK55109
ORF41	35944-36854 C	*orf303*	Replication (*repB*)	RepB of *Gluconobacter diazotrophicus* PA15	83	gb|ACI53275
ORF42	37745-38847	*orf367*	Transposase (*y4qE*)	Putative transposase y4qE of *Roseibium* sp. TrichSKD4	59	gb|EFO28627

The complete sequence information of plasmid pPDL2 is presented in chapter-I. In the sequence of pPDL2, *oriT* and *repB* that contribute for horizontal mobility of plasmid pPDL2 were identified. Further, the literature available on the structure and organization of *opd* genes coding organophosphate hydrolase clearly suggest existence of horizontal transfer of *opd* genes among soil bacteria (Siddavattam et al, 2003; Horne et al, 2003; Wei et al, 2009). However no experimental evidence is available to support the hypothesis. This chapter describes horizontal mobility of *opd* genes through experimental design described in materials and methods section. The well characterized *opd* plasmids pPDL2 of *Flavobacterium* sp. ATCC 27551 and pCMS1 of *Brevundimonas diminuta* were taken as experimental tools while gaining the experimental evidence on horizontal gene transfer (HGT) of *opd* gene.

4.1. Horizontal mobility of pPDL2 of *Flavobacterium* sp. ATCC 27551

Horizontal transfer of genes among bacterial strains is a well defined subject. In order to accomplish HGT among bacterial strains plasmids should have genes that code for mating-pair formation (MPF) between the donor and the recipient cells. In addition to the mating pair formation module, an *oriT* sequence and a *mob* gene coding for a relaxase is required to achieve successful transfer of plasmid DNA (Holmes and Jobling, 1996; Russi et al, 2008). In plasmid pPDL2 no MPF machinery is present. However, existence of *oriT* like structures and a replicase, RepB, the rolling circle replication initiator suggests mobilizable nature of pPDL2. In order to know the mobilizable nature of pPDL2, biparental and triparental mating experiments were performed as described in materials and methods section. While performing biparental mating experiments, the pPDL2 derivative, pPDL2::Tn5<R6Kγori-Kan-2> containing *E. coli pir*-116 was used as donor. Usage of pPDL2 derivative generated by inserting mini-transposon, EZ-Tn5<R6Kγori-Kan-

2>, confirms kanamycin resistance on pPDL2:: Tn5<R6Kγori-Kan-2>. Usage of an *E. coli* cell having pPDL2::Tn5<R6Kγori-Kan-2> as donor facilitates easy monitoring of pPDL2 mobility by selecting on a kanamycin antibiotic resistance plate. In an experiment conducted using *E. coli pir*-116 (pPDL2::Tn5<R6Kγori-Kan-2>) as donor and *Acinetobacter* sp. DS002 as recipient no exconjugants of *Acinetobacter* sp. DS002 were identified on a selective plate having kanamycin and benzoate (Fig. 4. 1). Logically on a minimal agar plates supplemented with 5mM benzoate and kanamycin colonies should appear if pPDL2 is a self transmissible plasmid. *Acinetobacter* sp. DS002 strains are kanamycin sensitive and grow on benzoate using as sole source of carbon. If pPDL2::Tn5<R6Kγori-Kan-2> is mobilized into recipient strain colonies of *Acinetobacter* sp. DS002 exconjugants would have grown on benzoate + kanamycin plates. However, no exconjugants were observed on the selection plates, which are also in agreement of the sequence information of pPDL2 (Fig. 4. 1).

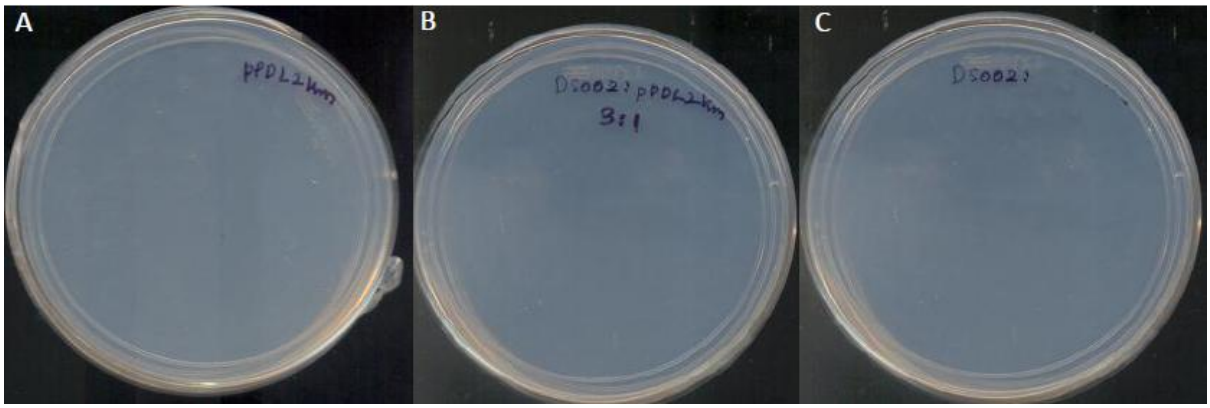


Fig. 4.1. Biparental mating using *E. coli pir*-116 (pPDL2::<R6Kγori-Kan-2>) and *Acinetobacter* sp. DS002 as donor and recipients. Panel A represents selection plate spread with *E. coli pir*-116 (pPDL2<R6Kγori-Kan>), B represents selection plate spread with conjugation mixture having *E. coli pir*-116 (pPDL2<R6Kγori-Kan>) (donor) and *Acinetobacter* sp. DS002 (recipient). Panel C represents selection plate with *Acinetobacter* sp. DS002 (recipient). No exconjugants were seen in selection plates having mating mixture.

4.1.1. Triparental mating

The above experiment clearly rules out the possibility of pPDL2 as a self-transmissible plasmid. However, if sequence information taken as basis to assess lateral transfer of pPDL2 it clearly suggests that it can only be mobilized if donor strain can supplement genetic machinery to code for a MPF. Therefore, such situation is created by performing triparental mating experiments, where the helper *E. coli* HB101 harbouring plasmid pRK2013 (Figurski and Helinski, 1979), *E.coli pir*-116 harbouring pPDL2::Tn5<R6Kγori-Kan-2> and *Acinetobacter* sp. DS002 strains served as helper, donor

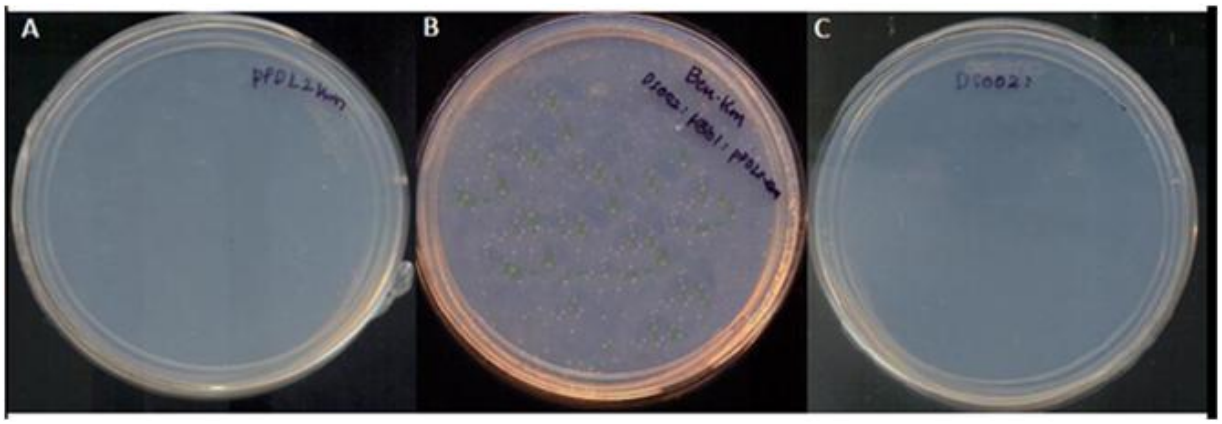


Fig. 4. 2. Triparental mating using *E. coli pir*-116 (pPDL2<R6Kγori-Kan>), *E. coli* HB101 (pRK2013) and *Acinetobacter* sp. DS002. Panel A represents selection plate spread with *E. coli pir*-116 (pPDL2<R6Kγori-Kan>), B represents selection plate spread with mating mixture having *E. coli pir*-116 (pPDL2<R6Kγori-Kan>) (donor), *E. coli* HB101 (pRK2013) (helper) and *Acinetobacter* sp. DS002 (recipient). Panel C represents selection plate with *Acinetobacter* sp. DS002. Appearance of exconjugants is seen in panel B.

and recipients respectively. The helper plasmid pRK2013 provides the genetic machinery for formation of mating pair formation and mobilization (Figurski and Helinski, 1979). Exconjugants appeared on the selection plates indicating the mobilization of plasmid pPDL2::Tn5<R6Kγori-Kan-2> into *Acinetobacter* sp. DS002 (Fig. 4.2). In the previous chapter, based on sequence information, a well conserved *oriT* was identified along with

RC replication origin and *repB* gene. The experiment described above has shown existence of functional mobilization module plasmid pPDL2.

4.1.2. Characterization of exconjugants

4.1.2.1. Detection of *opd* gene

After mobilization of plasmid pPDL2::Tn<R6Kγori-Kan-2> from *E.coli pir*-116 into *Acinetobacter* sp. DS002 through the helper strain *E.coli* HB101 (pRK2013), the presence of plasmid in the recipient strain was analyzed by performing PCR using the plasmid borne *opd* specific primers. When PCR was performed using *opd* domain specific primers, a specific amplicon of 500 bp was seen only in PCR mix having a colony of exconjugant and donor. No such amplicon was seen in recipient cells (Fig. 4.3).

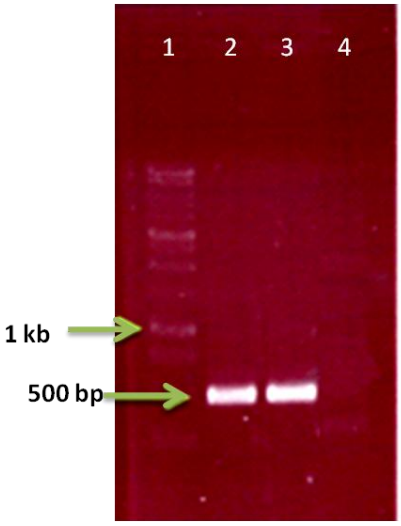


Fig. 4. 3. Confirmation of pPDL2 mobilization into *Acinetobacter* sp. DS002 by colony PCR using *opd* specific primers. Lane 1 represents 1 kb DNA ladder. Lanes 2-3 represent amplicons obtained from exconjugant and donor colony containing PCR mix. The PCR mix of recipient is shown in lane 4.

4.1.1.2. OPH assay

After establishing stable maintenance of plasmid pPDL2::Tn5<R6Kγori/Kan-2> in *Acinetobacter* sp. DS002, experiments were conducted to test their ability to degrade OP compound, paraoxon. Our lab has recently shown presence of OPH in the inner

membrane of *B. diminuta* and its dependence on Twin Arginine Transport (Tat) pathway for membrane targeting (Gorla et al, 2009). Therefore, the *Acinetobacter* cells harboring plasmid pPDL2::Tn*5*<R6Kγori/KAN-2>were fractionated in to cytoplasmic and membrane fractions and were assayed for OPH activity as described in methods. Most of the OPH activity was found in membrane fraction and very little activity was seen in cytoplasmic fraction (Fig. 4. 4). This result clearly suggests expression of active OPH from plasmid pPDL2::Tn*5*<R6Kγori/KAN-2>.

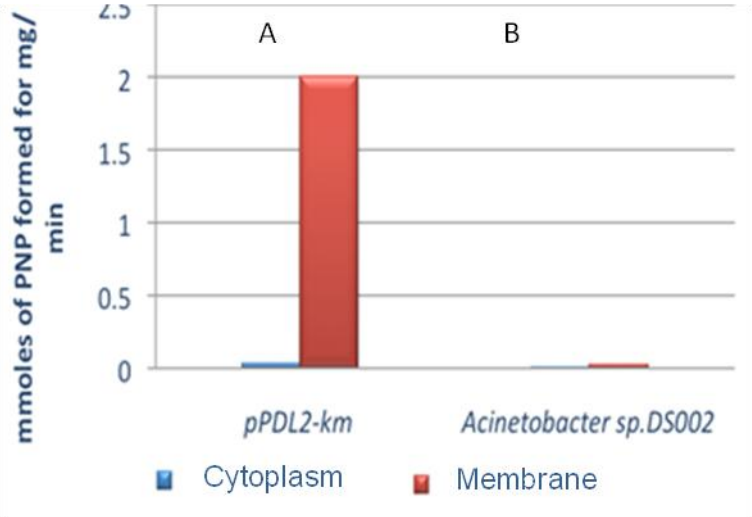


Fig. 4.4. Assay of organophosphorus hydrolase activity in cytoplasmic and membrane fractions of A) exconjugant (*Acinetobacter* sp. DS002 having plasmid pPDL2::Tn5<R6Kγ/Kan-2>) and B) wild type *Acinetobacter* sp. DS002.

4.2. *In vivo* transposition assay

As shown in Fig. 3.22 and described in chapter-1 the *opd* gene cluster of pPDL2 has shown a transposon-like module. Transposon Tn*3* and y4qE flank all the degradative traits that code for enzymatic machinery with a possible involvement of mineralizing OP insecticides like methyl parathion and fenitrothion. The organophosphorus hydrolase is a well characterized triesterase involved in hydrolysis of tri-ester linkages present in diverse group of OP compounds (Benning et al, 1994; Cho et al, 2004). Similarly, the

mfhA gene present immediately downstream of *opd* is a *meta* fission product hydrolase. Downstream of *mfhA* and *opd,* transposon Tn*3* consisting of *tnpA* and *tnpR* which codes for transposase

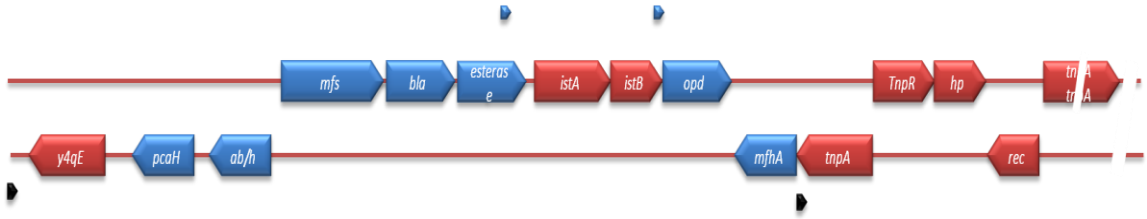


Fig. 3.22 of Chapter 1. Organization of degradative genes and mobile elements on pPDL2. Transposases and resolvases are shown in red colour. Degradative genes are shown in blue coloured arrows. The repeats of Tn*3* and IS*21* element are shown as black arrows and blue arrows respectively.

and resolvase was identified. In the upstream region of *opd* gene an IS element designated as IS*Flsp1* was identified (Siddavattam et al, 2003). In the sequence information generated in the present study two more open reading frames designated as ORF2 and ORF3 are identified upstream of IS*Flsp1*. One of them codes for β-sub-unit of protocatechuate 4,5 dioxygenase and the second ORF has homology to β-ketoadipate lactonase. Existence of another transposable element y4qE downstream of these genes indicates transposition of *opd* cluster. Existence of Tn*3* transposase specific inverted repeats flanking downstream of y4qE and upstream of transposase Tn*3* supports possible transposition event of *opd* element found in plasmid pPDL2.

In order to gain experimental evidence for such transposition event an *in vivo* transposition assay was performed by following procedures described in materials and methods section. The *in vivo* transposition assay was performed using three independent compatible plasmids. The plasmid pTras::*tet* contains the entire pPDL2 cluster cloned in

pUC18 vector. Its construction is described in one of the earlier reports published from our laboratory (Siddavattam et al, 2003). This plasmid serves as donor of *opd* cluster. The second plasmid, pJQ210SK (Quandt and Hynes, 1993) contains *sacB* gene. Presence of *sacB* gene is lethal for gram negative bacteria in presence of sucrose. The third plasmid is pMMBTnpA, derivative of pMMB206 generated by cloning *tnpA* under the control of inducible promoter. All the three plasmids were transformed into *E. coli* (pTrans::tet + pJQ210 + pMMB-TnpA) and were induced for expression of transposase by adding low concentrations of IPTG. After induction the cells were then plated on LB plates containing sucrose + gentamycin +tetracycline. A schematic representation is given in Fig. 4.5 to explain the functioning of the *in vivo* transposition assay.

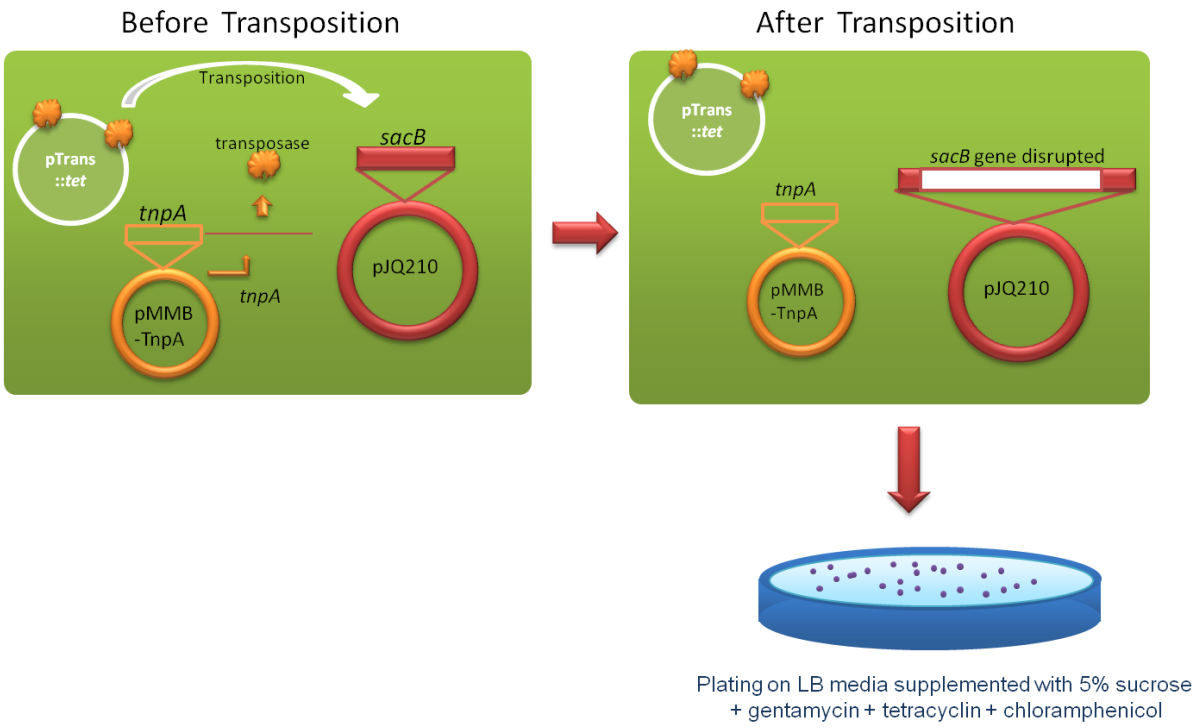


Fig. 4.5. Schematic representation of *In vivo* transposition assay

E.coli cells having either pTrans::*tet* (donor of *opd* cluster) or reporter plasmid (pJQ210SK) served as negative control. As shown in Fig. 4.6 a number of colonies are

found in selection plates containing sucrose + gentamycin + tetracycline + chloramphenicol. Such appearance of colonies on sucrose plates is only possible in the event of disruption of *sacB* gene due to transposition.

A) B)

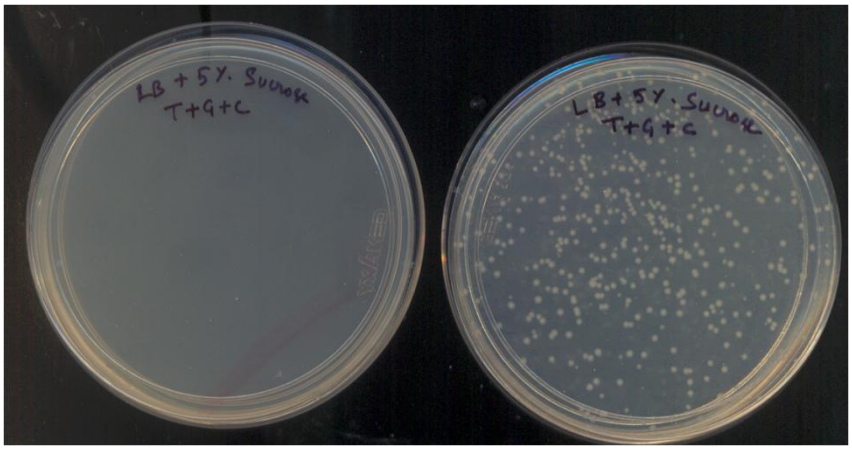


Fig. 4.6. *In vivo* transposition assay for demonstrating transposition of *opd* gene cluster.Panel A and B represent LB + 5% sucrose (Kanamycin + gentamycin + tetracycline) plates spread with *E.coli* having (pTrans::*tet* + pJQ210 + pMMB206) and *E.coli* having (pTrans::*tet* + pJQ210 + pMMB-TnpA). Sucrose resistant colonies are only seen in cells expressing TnpA.

4.2.1. Analysis of sucrose resistant colonies

In order to gain further insights into the nature of DNA fragment causing disruption of *sacB* gene, the pJQ210 derivatives were isolated from sucrose resistant colonies. The plasmids prepared from the resistant colonies were used as templates for performing PCR using *sacB* specific primers. PCR amplification using these primers gave 1.2 kb *sacB* amplicon, if it is not disrupted. However, if *sacB* is disrupted due to insertion of *opd* element found on the donor plasmid pTrasn::*tet*, there will be increase in size of the amplicon. Figure 4.7 shows that amplicons obtained from a number of colonies have

a size greater than the *sacB* gene. Colony PCR performed using *opd* specific primers from these colonies gave

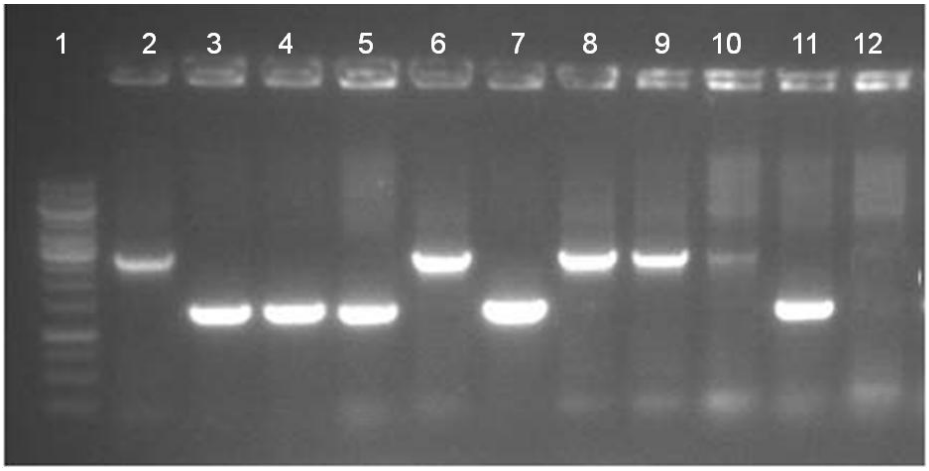


Fig. 4.7. Screening of plasmids isolated from sucrose resistant colonies using *sacB* specific primers. Increase in size indicates disruption of sacB gene due to transposition. Lane 1 represents 1kb DNA Ladder. Lanes 2 -10 represent amplicons from sucrose resistant colonies. Lane 11 and 12 represent positive control (pJQ210SK) and negative control (pTrans::*tet*, donor plasmid).

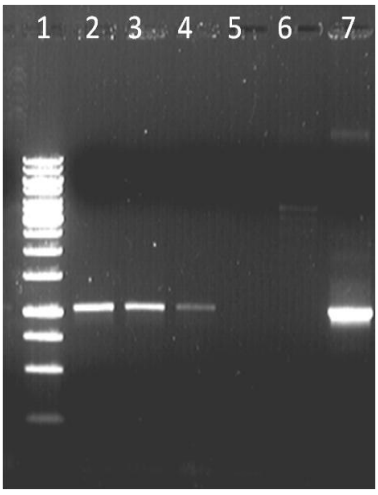


Fig. 4.8. Screening of sucrose resistant colonies using *opd* specific primers. Lane 1 represents 1kb DNA Ladder. Lanes 2 -6 represent amplicons of *opd* from sucrose resistant colonies. Lane 7represent positive control (pTrans::*tet*).

amplicons corresponding to the size of *opd* gene (Fig. 4.8). These two experiments provide a clear evidence to claim the transposable nature of the *opd* element found on plasmid pPDL2. However, these pJQ210 derivatives containing *opd* element insertion in

sacB gene needed to be sequenced to identify boarder sequences of the *opd* element. Due to paucity of time such experiments were not conducted in this study. However, our laboratory is undertaking such studies to identify precise junctions of *sacB* and *opd* element to determine *opd* specific sequences.

The studies described in chapter-2 document horizontal nature of *opd* genes found on pPDL2 of *Flavobacterium* sp. ATCC 27551 and clarify to large extent for the reasons behind the existence of identical *opd* elements on dissimilar plasmids and chromosomes of soil bacteria isolated from diverse geographical regions.

4.3. Horizontal transfer of plasmid pCMS1

Brevundimonas diminuta is one of the first microorganisms reported to have the capability to degrade organophosphorus pesticides (Serdar et al, 1982). It was further demonstrated that the OP compound degradation capability of *Brevundimonas diminuta* was due to existence of a large indigenous plasmid, pCMS1 (Mulbry et al, 1986; McDaniel and Wild, 1988). The comparison of restriction profile of pCMS1 and pPDL2 revealed existence of sequence similarity only in 5.1 kb DNA region containing *opd* gene. As described before plasmid pPDL2 is clearly a mobilizable plasmid. It can constantly contribute for HGT of *opd* gene under certain conditions, especially when it gains access to mating pair formation coded by another plasmids/ Integrative conjugative element.

4.3.1. Random sequencing of pCMS1

Unlike in pPDL2, no sequence information is available for pCMS1. In order to test the HGT of pCMS1 two things need to be generated. Initially the plasmid pCMS1 need to be tagged with an antibiotic marker so that its mobility can be monitored by performing a typical conjugation experiment. Secondly the plasmid pCMS1 need to be sequenced to

identify genetic modules coding for mating pair formation (MPF) and *oriT* and *mob* genes. In order to obtain quick information on the sequence of pCMS1 two fosmid clones having entire pCMS1 DNA were taken and *in vitro* transposon tagging experiments were performed to insert mini-transposon EZ-Tn5<R6Kγori-Kan-2> randomly into fosmid clones of pCMSA and pCMSB. After *in vitro* transposon tagging experiments, the kanamycin resistant *E. coli pir*-116 cells were taken and the fosmids having mini-transposon insertions were independently sequenced using transposon specific primers. Such sequence stretches were then analyzed to identify presence of *tra* genes that contribute for mating pair formation (MPF). The sequence information has shown the presence of *tra* genes on plasmid pCMS1 providing *prima face* evidence for horizontal mobility of plasmid pCMS1 (Table 4.1). As

Table 4.1. List of *tra* genes identified in plasmid pCMS1.

S. No.	Name of the recombinant fosmid.	Name of the Tra protein deduced from nucleotide sequence & % homology	Role in horizontal gene transfer (HGT)	*Reference strain/* GeneBank Accession No.
1	pCMSB314	TraI ; 54	DNA relaxase	*Comamonas sp.* CNB-1 / ABM06255
2	pCMSB32R	TraE ; 72	DNA topoisomerase	*Achromobacter xyloxidans sub-sp. denitrificans* / AAS49467
3	pCMSB32F	TraE ; 64	DNA topoisomerase	*Achromobacter xyloxidans sub-sp. denitrificans* / AAS49467
4	pCMSB31R	TraE ; 70	DNA topoisomerase	*Achromobacter xyloxidans sub-sp. denitrificans* / AAS49467
5	pCMSB31F	TraM ; 56	Conjugal transfer protein	*Achromobacter xyloxidans sub-sp. denitrificans* /AAS49476

shown in Table 4. 1, *traI*, *traE* and *traM* genes were identified on plasmid pCMS1. All the *tra* sequences share considerable similarity (54-72%) with *tra* genes present on plasmid

pEST4011 of *Achromobacter xyloxidans* sub-sps. *denitrificans*. The TraI is a relaxase which performs strand scission at the transfer origin (*oriT*) and TraM acts a topoisomerase which enhances TraI activity. TraE is one of the proteins present in the pilus assembly (Karl et al, 2003). As existence of the *tra* genes was apparent on pCMS1, further experiments were done to obtain experimental evidence on its horizontal mobility.

As no markers are available to track mobility of pCMS1, the *B.diminuta* strains having its pCMS1: *tet* generated previously by replacing the *opd* with *opd::tet* (Gorla et al, 2009) was used as donor strain. In a typical conjugation experiment, performed by using *B. diminuta* (pCMS1::*tet*) and *P. putida* KT2440 as donor and recipient, respectively, plasmid pCMS1::*tet* was successfully transferred with a frequency of 0.72 x 10^{-6} . This result clearly suggests that plasmid pCMS1 is a self transmissible plasmid and supports the sequence information that gave clear indication about the presence of *tra* genes.

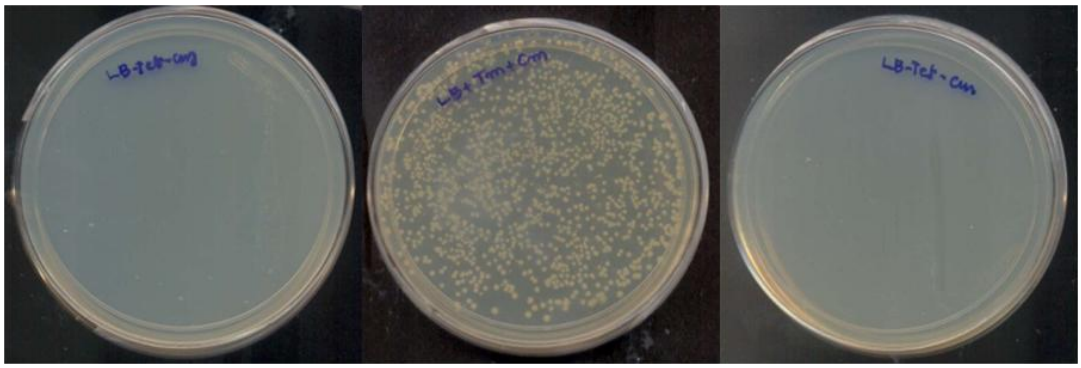


Fig. 4.9. Panel A, B and C indicate LB (Km+Cm) plates spread with donor (*B.diminuta* with pCMS1::*tet*), mating mixture and recipient strains (*Pseudomonas putida* KT2240). Colonies were only seen on plates spread with mating mixture.

4.3.2. Analysis of exconjugants

The exconjugants generated in this study were analyzed to see presence of *pCMS1*::*tet*. Initially, the presence of the plasmid was detected in exconjugant and donor by isolating plasmids following protocol described in materials and methods section

(Currier and Nester, 1976). As shown in figure 4.10-A a clear plasmid band was seen in exconjugant and donor. No such plasmid band appeared in recipient cells. The plasmid preparations were then analyzed for presence e of *opd*::*tet* by performing PCR using *opd* specific primers. Agarose gel analysis of PCR samples has shown amplification of *opd::tet* from an exconjugant. The *tet* gene has an internal *BamHI* site. The amplicon when digested with *BamHI* generated two fragments of 1.1 kb and 0.8 kb respectively, confirming existence of *opd*::*tet* in plasmid preparations made from exconjugant (Fig. 4.10).

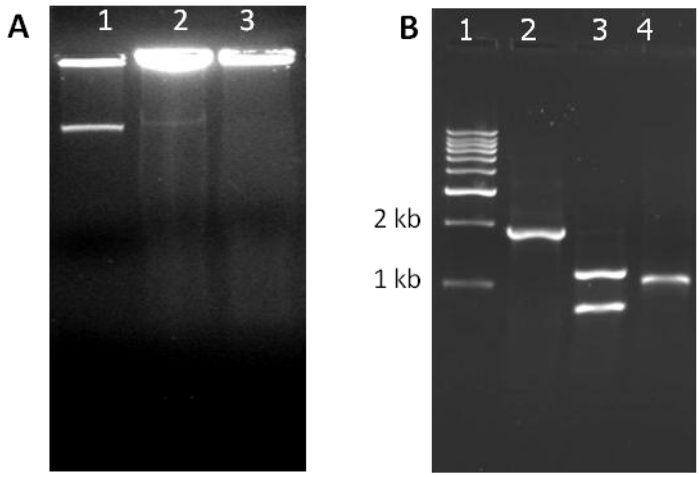


Fig. 4.10 Panel A. Agarose gel electrophoresis showing existence of pCMS1::*tet* in donor (lane1) and exconjugant cells (lane 2). No plasmid was seen in *Pseudomonas putida* KT4220 used as recipient (lane 3). Panel B. Agarose gel showing PCR amplification of *opd::tet* from exconjugants of *B. diminuta.* Lane 1, kb ladder, lane 2 PCR amplification of *opd::tet* from exconjugants. Lane 3 shows digestion of *opd::tet* with *Bam*HI, an unique site found in *tet* gene. Lane 4 represents amplification of 1.2 kb *opd* gene from wild type *B. diminuta* cells.

4.4. Discussion

The work presented in this chapter clearly suggests possible Horizontal Gene Transfer (HGT) of the *opd* gene among soil bacteria, in the light of these results a

thorough literature search was done to have an understanding of the origin, evolution and degradation of organophosphorus degrading traits. The organophosphorus hydrolyzing enzymes were originally named as phosphotriesterases due to their ability to hydrolyze tri-ester linkage found in organophosphates and nerve agents. The tri-esterases found in prokaryotes can be divided into three independent groups. There exists no sequence identity among these three groups. They are organophosphorus hydrolases encoded by *opd* gene identified in *Flavobacterium* sp. ATCC 27551 and *Brevundimonas diminuta*. The second group belongs to methyl parathion hydrolases (MPH) group. These are mainly isolated from Chinese agricultural soils. The third group belongs to organophosphorus acid anhydrolase (OPAA) group. A The OPAAs were later identified as prolidases, the dipeptidases found in variety of bacterial strains. The structural similarity between natural substrate, dipeptide and nerve agent strain is shown to be responsible for the triesterase activity (Cheng and DeFrank, 2000). Therefore, in the subsequent sections, a brief description is given only on the structure and function of *opd* and *mpd* genes, which are considered to be evolved from lactonases and β-lactams respectively. In the preceding sections a detailed description is given on organization and HGT of *opd* elements found on plasmid pPDL2 and pCMS1. Therefore further description on *opd* elements are avoided and a brief mention is made in the following sections on structural organization of other phosphotriesterase sequences found in taxonomically diverse group of organisms.

4.4.1. The Tn*opdA* element

The *opd* homologue of *Agrobacterium radiobacter* P230 is *opdA* (Horne et al, 2002). The chromosomally located *opdA* gene was later shown to be part of a

transposable element, which contained three further ORFs in addition to *opdA* along with inverted repeats typically seen in transposon Tn*610* of *Mycobacterium fortuitum* (Horne et al, 2003). The transposase TnpA is identical with the TnpA sequence of Tn*610*. The other two ORFs found between *tnpA* and *opdA* were predicted to code for a truncated transposase (*orfA*) and an ATP binding protein (*orfB*). Transposition was successfully shown in *E. coli* confirming the horizontal mobility of *opdA* sequences among soil bacteria (Horne et al, 2003).

4.4.2. The *mpd* elements

All methyl parathion-degrading (*mpd*) genes reported to date have been isolated from Chinese agricultural soils or from the activated sludge collected from a Chinese pesticide manufacturing unit. The first *mpd* gene was cloned from a *Plesiomonas* sp. strain M6 (Zhongli et al, 2001). This chromosomally located *mpd* gene surprisingly has shown no homology to any of the known *pte* genes. Following this discovery, a number of *mpd* sequences were cloned from Chinese agricultural soils (Liu et al, 2005; Zhang et al, 2006). However, the horizontal mobility of *mpd* genes among soil microbes gained acceptance only with the discovery of plasmid-borne *mpd* gene in *Pseudomonas* sp. strain WBC-3. In this soil isolate an indigenous plasmid of 70 kb designated pZWL0 contained both a *mpd* gene and genes responsible for degradation of *p*-nitrophenol. This strain uses methyl parathion and its degradation product *p*-nitrophenol as sole source of carbon, nitrogen and energy (Liu et al, 2005). Further investigations into the genetics of methyl parathion degradation revealed that the organization of the *mpd* gene in this strain was like that of a functional transposon (Fig. 4.11).

4.4.2.1. The Tn*mpd* element is a typical class I Transposon

Sequence analysis of the *mpd* region of pZWLO of *Pseudomonas* sp. strain WBC-3 revealed the existence of a functional *mpd* element. When the 6.5 kb *Kpn*I-*BamH*I fragment was sequenced, it revealed existence of the IS*6100* class of IS elements that flanking the *mpd* gene (Fig. 4.11). Further studies conducted by Wei et al (2009) have elegantly demonstrated the transposition event of the *mpd* gene in *Pseudomonas* sp. strain WBC-3 (Wei et al, 2009). This is the first functional transposable element with a *mpd* gene and the second one in the entire *pte* family of genes that codes for a phosphotriesterase.

4.4.2.2. Distribution of *mpd* elements

After establishing the existence of a functional *mpd* element, Tn*mph,* in *Pseudomonas* sp. strain WBC-3, Zhang et al, have isolated seven bacterial strains capable of degrading methyl parathion from different locations of Chinese soils with a history of using methyl parathion (Zhang et al, 2006). In all of them a *mpd* gene containing a 4.7 Kb region is highly conserved. In this conserved DNA region a total of five ORFs were identified. One of them shows similarity to the TnpA-coding sequence of an IS element, IS*6100*, and contains a perfect 14 base pair inverted repeat in its flanking sequences. The second ORF, designated as *orf463*, found immediately downstream of the IS element codes for a protein that shows considerable homology to a house-keeping sigma factor. Significance of its existence as part of *mpd* element is unclear. However, presence of other ORFs as part of the *mpd* cluster appears to have a strong functional relevance. The ORFs found upstream of *mpd* sequences, *orf232* and *orf259*, code for a permease of an ABC-transport system and an ExeB found to be the part of general protein secretion pathway. Although there is no experimental evidence to support, due to the presence of

these two ORFs in association with the *mpd* gene, which codes for a MPH precursor with a 35 amino acid-long signal peptide, it is proposed that these two proteins are involved transport and maturation (Zhang et al, 2006). Of the seven *mph* sequences known, two of them are identical, whereas and the rest of the five sequences code for MPH proteins with amino acid substitutions at 9 positions. Nevertheless, these variations have been shown to have positive effect on the catalytic properties of the MPH (Dong et al, 2005). The existence of such highly conserved *mpd* clusters in seven different bacterial strains that show a weak taxonomic relationship strongly supports horizontal mobility among *mpd* genes in soil microbes. The presence of IS*6100* in all these clusters adds strength to this proposal.

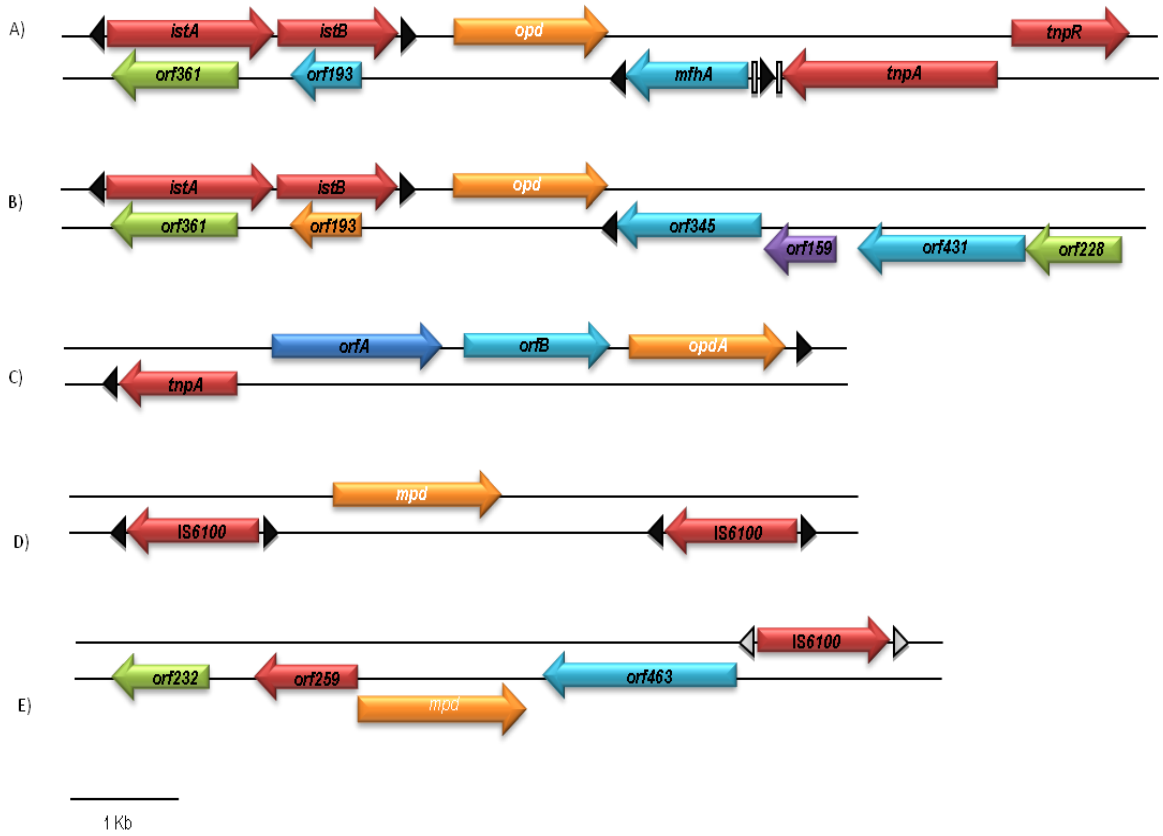


Fig. 4.11. Physical maps showing organization of the *opd* clusters of pPDL2 from *Flavobacterium* sp. ATCC 27551 (A) and pCMS1 from *Brevundimonas diminuta* (B), *opdA* of *Agrobacterium radiobacter* P230 (C), *mpd* of *Pseudomonas* sp. WBC-3 (D) and *Plesiomonas* (W). Arrows indicate the direction of transcription.

4.4.3. The *opaA* genes

After purification of organophosphate acid anhydrolase (OPAA) enzymes from various halophilic and *Alteromonas* species (DeFrank et al, 1993) and demonstration of their ability to degrade G-class nerve agents (Cheng et al, 1999), the *opaA* gene was cloned from the *Alteromonas* sp. strain JD6.5. The *opaA* gene product has more than 50% amino acid similarity to *E. coli* PepQ. Further investigations into the physiological role of OpaA have established that it has prolidase activity. There are no indications that the gene is organized as a mobile genetic element or of its presence on plasmid. The *opaA* gene does not appear to have evolved to code for a phosphotriesterase. The activity of its product, prolidase, on G-class nerve agents, is seen as an ancillary activity of these enzymes due to structural similarity of their substrates (Merone et al, 2005).

4.5. Evolutionary link between phosphotriesterases and lactonases

Promiscuous activities play a key role in the evolution of enzymes. They actually serve as starting point for acquiring a new function through gene duplication (Kolalowski et al, 1997; Lai et al, 1995; Rastogi et al, 1997; Benning et al 1994; Harper et al, 1988). In fact, these promiscuous activities are considered to be the vestiges of the function of their ancestral protein (Kolalowski et al, 1997; Lai et al, 1995). The phophotriesterases have been shown to have promiscuous phosphodiesterase, carboxyl esterase, and lactonase activities (McDaniel et al, 1988; DeFrank et al, 1993; Cheng et al, 1997). In general, the family members that have presumably diverged from a common ancestor often share promiscuous activities (Poelarends et al, 2005, Roodveltdt et al, 2005, Yew et al, 2005, Elias et al, 2008). Afriat et al have elegantly shown the existence of reciprocal promiscuities between lactonases and triesterases (Afriat et al, 2006). Based on the

structural differences, especially in the loops 1, 7 and 8 that comprise substrate binding sites, they have classified OPH homologues into three groups (Fig. 4.12). In the first group, designated as phosphotriesterases, with more than 86% identity to *bd*-OPH, they have shown existence of promiscuous lactonase. The *ec*-OPH is kept in the second group of OPH homologues, as it contained relatively shorter substrate binding loops. The third group of enzymes have only loop 7 (Fig. 4.12). These proteins annotated in the database as putative parathion hydrolases including AhlA from *R. erythropolis*, PPH from *Mycobacterium tuberculosis* and SsoPox of *Sulfolobus solfataricus* and are all re-classified as phophotriesterase like lactonases (PLLs). All of them proficiently hydrolyzed lactone with distictively low Km (10 -230uM) values and a very weak phosphotristerase activity (102 to 106 fold). The arylesterase activity was shown only by SsoxPox but not by any of the other PLLs. In principle, the promiscuous activity shown by a family member is not seen with other members of the family. If activity is shown by all the members, it is considered to be indicative of native function (Khersonsky et al, 2006). If this analogy is taken into consideration the PLLs are primarily lactonases with promiscuous phophotriesterase activity and probably the phophotriesterases are evolved from PLLs in the recent past. The substrate-binding loops contribute the main structural difference between PTEs and PLLS.

Indeed insertions, deletions and loop-swapping are believed to be a primary mechanism for creating enzyme diversity (Twafik, 2006, Park et al, 2006, Soskine and Tawfik, 2010). A number of studies have used PLLs as templates for directed evolution and succeeded in either enhancing the substrate range, catalytic efficiency (Chow et al, 2009) or converting PLLs to catalyze altogether new reactions (Mandrich and Manco,

2009). Among the proposed PLLs the archaeal triesterase alone is shown to have arylesterase activity (Afriat et al, 2009). As *ec*-OPH has also shown arylestrase activity the PLL, SsoxPox, due to the existence of similar activity and structural similarities, is proposed to be a "generalist" molecule that served as a template for evolution of phosphotriesterases found in mesophilic organisms (Afriat et al, 2006, Merone et al, 2008). In fact, the recent discovery of phosphotriesterase-like carboxyesterase, (*Mlo*PLC) from *Mesorhizobium loti* and its transformation into a diesterase through *in vitro* evolution supports the proposed hypothesis by Afriat et al, 2006 (Mandrich and Manco, 2009).

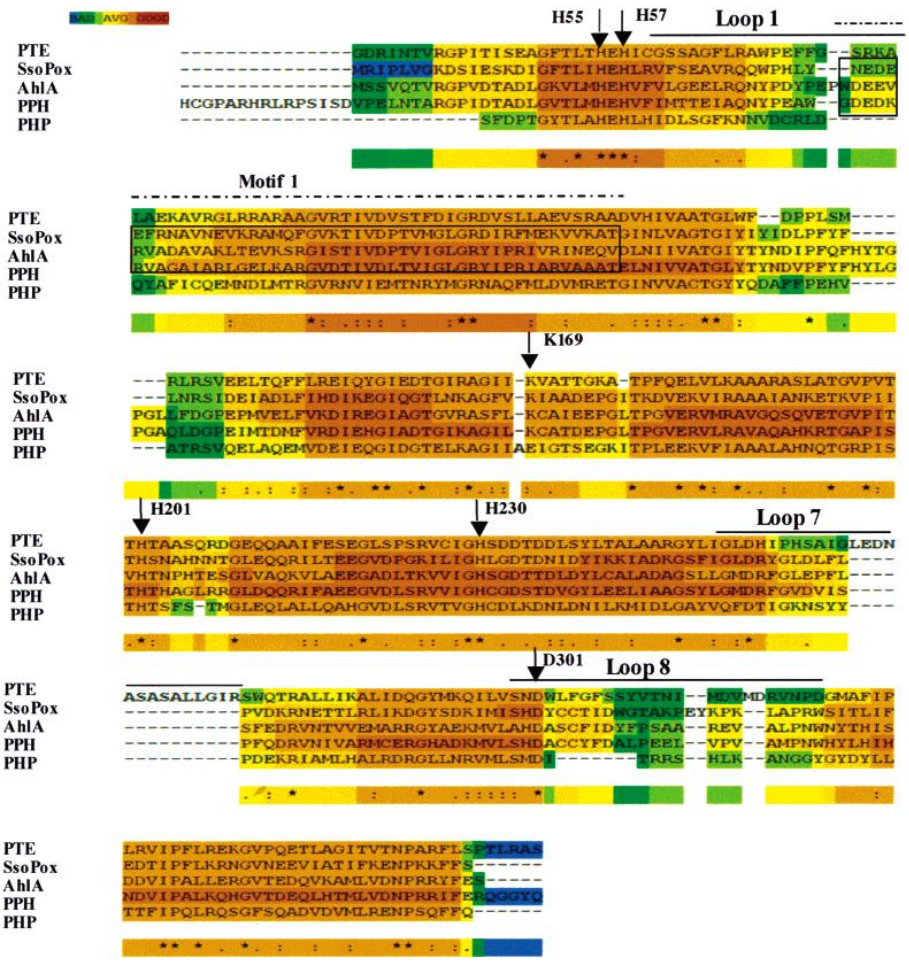


Fig. 4.12. Multiple Alignment of Phosphotriesterases of *B.diminuta* (PTE), *Sulfolobus sulfatricus* (SsoPox), *R. erythropolis* (AhlA), *Mycobacterium tuberculosis* (PPH) are shown using T-Coffee program.

4.6. MPH Scenario

The scenario with the evolution of methyl parathion hydrolases appears to be in no way different from the evolution of the OPHs. They appear to have evolved from ß-lactamases with which they share considerable structural homology. The N-acyl-L-homoserine lactone (AHL) lactonases are members of the metallo-ß-lactamase superfamily and contain two zinc ions in their catalytic center (Aravind, 1999; Daiyasu et al, 2001; Crowder et al, 2006). The recently solved crystal structure of *Bacillus thuringiensis* AHL lactonase (Liu et al, 2008) has shown striking similarity with the crystal structure of MPH (Dong et al, 2005) (Fig. 4.13). The MPH is also shown to have promiscuous lactonase activity (Afriat et al, 2006). If these findings are seen with the aforementioned experimental evidence gathered to show the structural relationship between lactonases and phosphotriesterases, the proposal that metallo ß-lactamases were the progenitors of MPH is worthy of consideration.

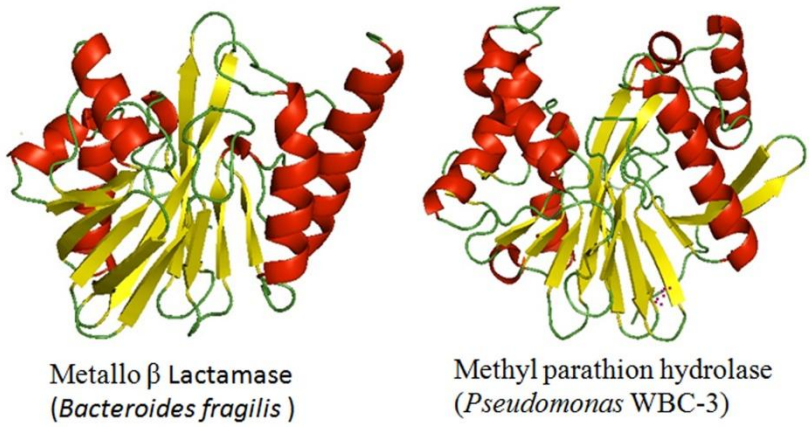


Fig. 4.13. Ribbon diagram of Metallo-β Lactamase and methyl parathion hydrolase showing similarities in their structure

Evolution of such new traits coding for biodegradation of recalcitrant xenobiotics and recalcitrant aromatic compounds is not uncommon and has been reported frequently in

the literature (Betsy et al, 1987; Dick et al, 2005). A number of studies conducted on dehalogenases (Mariel and Dick, 2002; Dick et al, 2005) have indicated that the evolution of dehalogenases. Information available on these catabolic enzymes serves as an illustration of many key concepts in enzymology including parallel evolution, convergent evolution, gene transfer, determination of reaction mechanisms and structure-activity relationships (Reviewed in Allpress and Gowland, 2010 and references therein). Considering the ample evidence available on the possible evolution of the phosphotriesterases from quorum 'quenching' hydrolases and the existence of these traits on transposable elements suggests their recent 'evolution' is a consequence of OP-induced gain-of-function. The increased presence of these OP compounds in agricultural soils due to repeated and excessive use would have created the necessary positive selection pressure to distribute the newly acquired functionality among microbial populations. The existence of such traits on self-transmissible plasmids like pCMS1 would greatly facilitate lateral gene transfer.

Conclusions

1. Plasmid pCMS1 of *B.diminuta* is self transmissible and showed horizontal mobility into *Pseudomonas putida*.

2. The *tra* genes were identified in pCMS1 in support of its self-transmissibility.

3. Plasmid pPDL2 is only a mobilizable plasmid. It can only be horizontal transferred into other bacterial strains in presence of helper plasmids.

4. The transposase of *Tn3* causes transposition of *opd* gene cluster present on plasmid pPDL2 of *Flavobacterium* sp. ATCC 27551.

Organophosphates, as stated in the introduction chapter, are widely used as insecticides to control various insect pests that affect economically important crops. The organophosphorus hydrolase (OPH) encoded by plasmid pPDL2 borne organophosphate (*opd*) gene hydrolyses triester linkage found in variety of OP insecticides (Mulbry and Karns, 1989; Dumas et al, 1989). This hydrolytic step inactivates OP compounds and thus reduces their toxicity towards mammals and other non-target organisms having well developed nervous system. However, this hydrolytic cleavage generates a number of aromatic compounds, especially nitrophenols (Fig. 5.1), which are highly toxic to soil microflora and are shown to adversely influence soil ecosystem (Camper et al, 1991). Complete mineralization of OP compounds means not only OPH mediated hydrolysis, the products

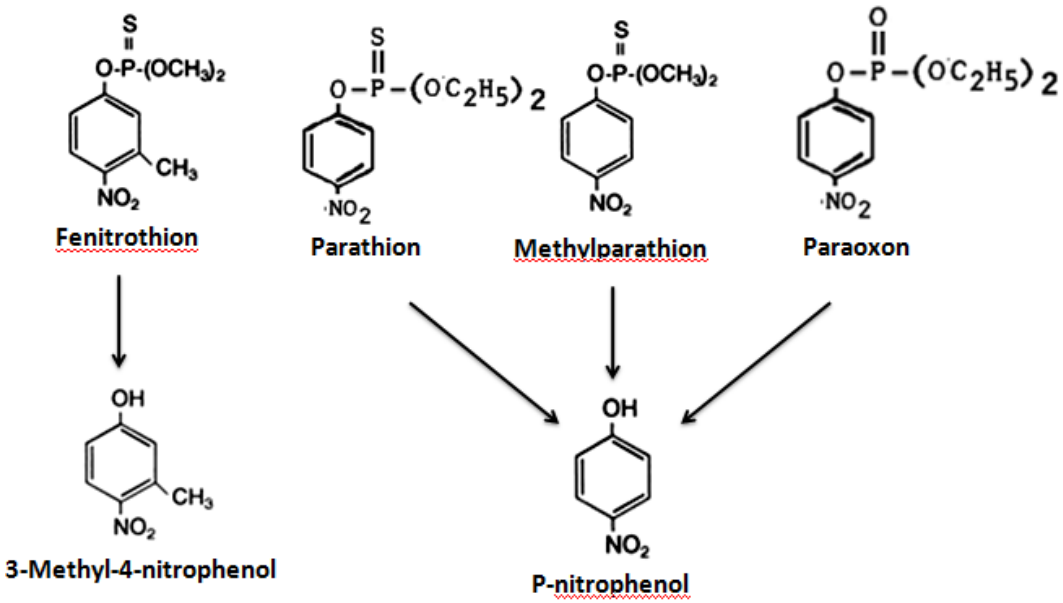


Fig. 5.1. Generation of nitrophenols through OPH mediated hydrolytic cleavage of certain OP_compounds.

generated through hydrolytic cleavage must be mineralized. In a very few cases the microbes are shown to possess genetic capability to mineralize OP compounds, due to existence of *opd* gene along with genes coding for enzymatic machinery needed for

degradation of aromatic compounds (Ou and Sharma, 1989; Rani and Lalithakumari, 1994; Keprasertsup et al, 2001). The alternative strategy is to mobilize *opd* containing native plasmids into native isolates having innate ability to mineralized aromatics and nirophenols. In our attempt to search for such native soil bacterial strain our laboratory has isolated an *Acinetobacter* sp. DS002 from a methyl parathion contaminated soil. When tested no *opd* gene was found in *Acinetobacter* sp. DS002. However, it has grown on a variety of aromatic compounds. Biodegradation of aromatic compounds is a well-studied aspect of biocatalysis. The degradation pathways operational in both gram-positive and gram-negative bacteria are well known (Harwood and Parales, 1996). Understanding of aromatic compound degradation pathway operational in *Acinetobacter* sp. DS002 is expected to generate basic information required for its manipulation towards achieving the mineralization of phenolic and nitrophenolic compounds. On careful examination, the isolate was shown to use a number of aromatic compounds and dicarboxylic acids as sources of carbon. Through comparative growth studies, the benzoate has been shown to serve as a better carbon source for *Acinetobacter* sp. DS002. Therefore aromatic degradation pathway found in *Acinetobacter* sp. DS002 was elucidated by growing it using benzoate as sole source of carbon. A combinatorial approach involving both metabolomic and proteomic tools were followed while elucidating the degradation pathway. Before actually analyzing the degradation pathways, the growth conditions were optimized and the cells grown in a physiological condition where the pathway enzymes were maximally induced.

5.1 Growth behavior of *Acinetobacter* sp. DS002 in benzoate

Initially, the optimal benzoate concentration for optimal growth of *Acinetobacter* sp. DS002 was determined. Growth was observed in all concentrations ranging from

5mM to 100mM benzoate. However, in high benzoate concentrations growth was seen only after 20 hours. A typical growth curve with lag, log and stationary phases were observed when grown in 5mM benzoate (Fig. 5.2). At higher concentrations of benzoate (50 mM) a typical diauxic growth curve with two exponential phases was observed. About 97% decrease in concentration of sodium benzoate was seen by the end of the logarithmic phase (Fig. 5.3).

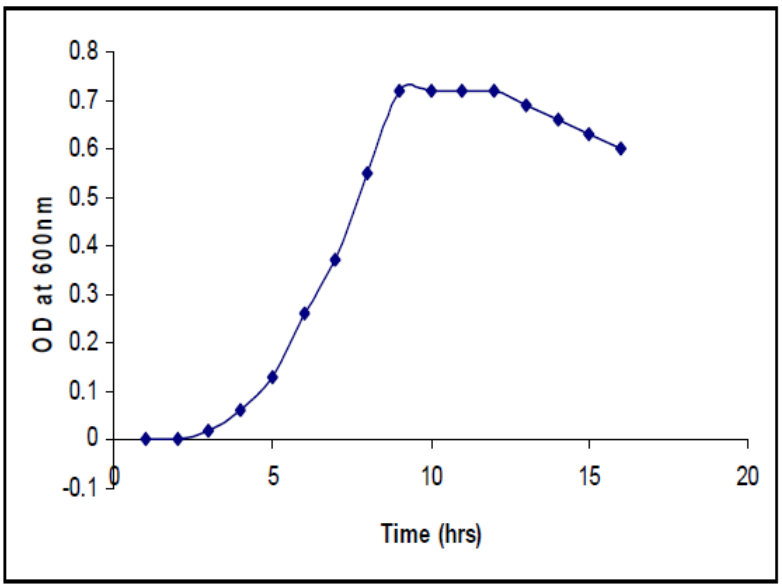


Fig. 5.2. Growth curve of *Acinetobacter* sp. DS002 in benzoate

When spent medium collected after eight hours of growth was analyzed on HPLC there was drastic reduction in benzoate peak with a concomitant appearance of additional peaks with retention times less than 4.4 min. The metabolites associated with these two new peaks were identified using LC/MS.

5.2 LC/MS analysis of catabolites

In order to identify the catabolic intermediates of benzoate, the spent medium collected from *Acinetobacter* sp. DS002 culture was extracted at different time intervals

following the procedures described in materials and methods section. When these extracts were separated on HPLC, three major peaks with retention times of (Fig. 5. 4) 1, 2.1 min (Fig. 5.5) and 3.9 min1.6 min (Fig. 5. 6) were observed. Though there were considerable differences in the individual peak intensities, they were constantly found from the extracts

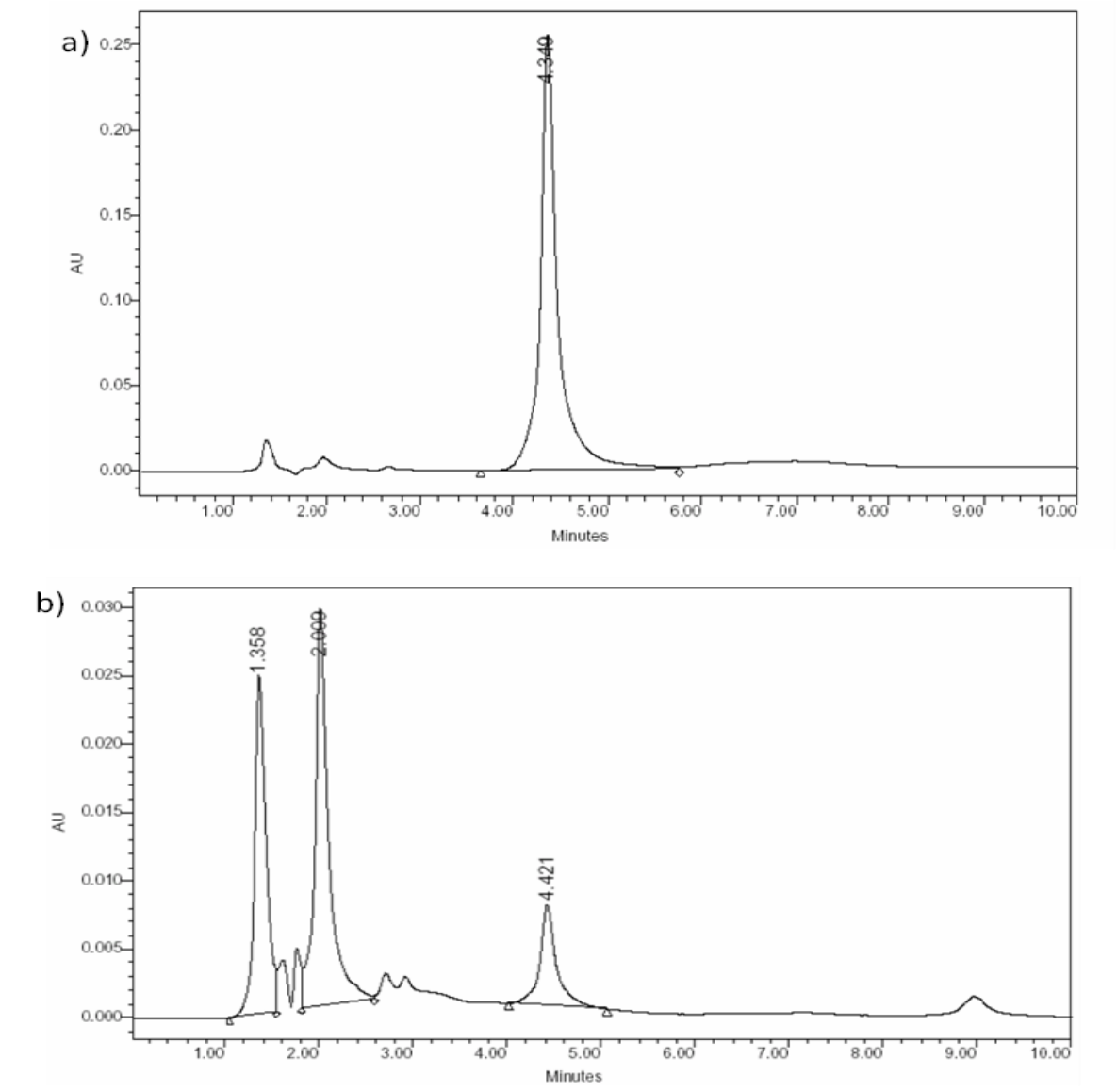


Fig. 5.3. HPLC analysis of catabolites at a) 0 hrs and b) 8 hrs. Peak with a retention time of 4.4 min represents benzoate.

prepared both at 8 h and 16 h time periods. These base peaks were then used for doing MS/MS in the negative mode to obtain molecular ion [M-] m/z values. The retention

times and molecular ions [M-] having the m/z values 141, 109 and 121 matched with m/z values of *cis, cis* -muconate, catechol and benzoate. As evidenced in the peak intensities, the concentration of the catabolic intermediates in the spent medium varied with time due to utilization of these compounds by *Acinetobacter* sp. DS002. The catabolite *cis,cis* muconate is the product formed from catechol by the action of catechol 1,2 dioxygenase, the first enzyme in the *ortho* cleavage pathway of catechol (Stanier et al, 1970).

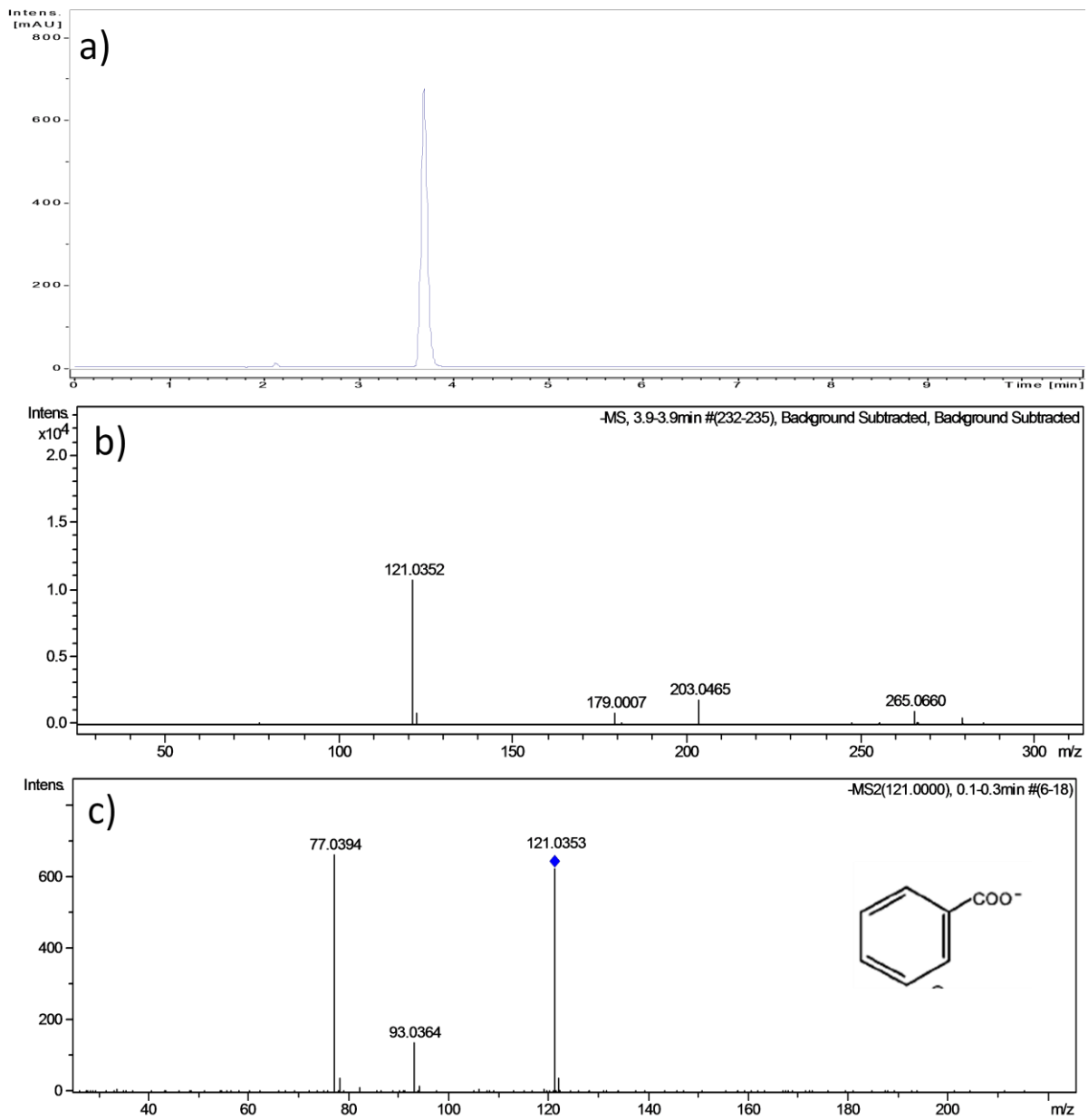


Fig. 5.4. LC-MS analysis of metabolites collected at 0 hrs time. Panels a, b and c represent LC profile, MS pattern and MS/MS pattern of a compound with retention time 3.9 min.

The initial step in the aerobic biodegradation of benzoate is incorporation of molecular oxygen into the aromatic nucleus by the enzyme benzoate 1, 2 dioxygenase to form a non-aromatic *cis-diol,* 2-hydro-1, 2-dihydroxybenzoate (DHB) which is further converted to catechol by the action of DHB dehydrogenase (Fig. 5)(Reiner, 1972; Reiner and

Hegeman, 1971). Catechol is cleaved either through the *ortho* pathway or *meta* pathway by the

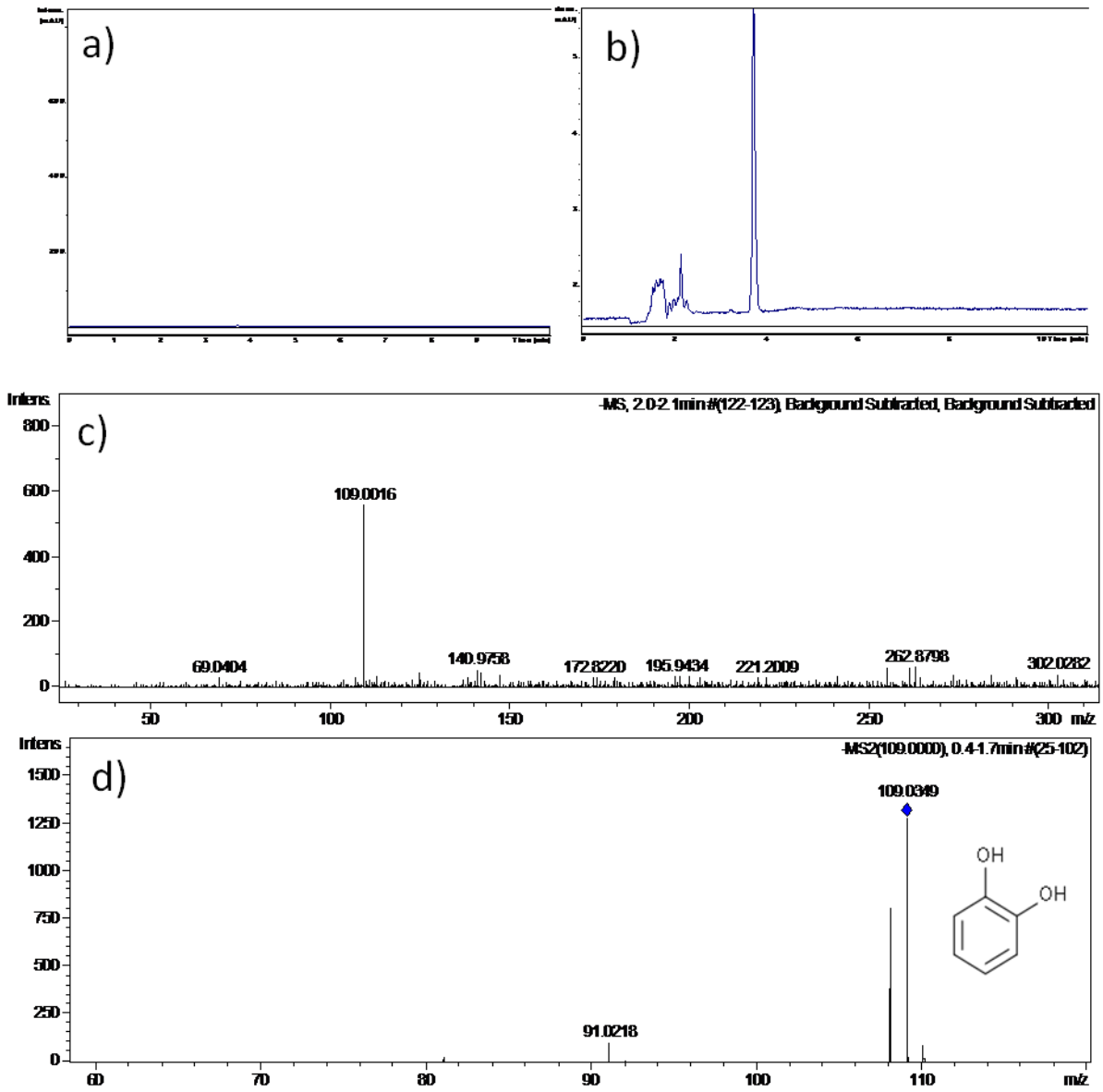


Fig. 5.5. LC-MS analysis of metabolites collected at 8 hrs time. Panels a, b, c and d represent LC profile, enlarged LC profile, MS pattern and MS/MS pattern of a compound with retention time 2.0 min.

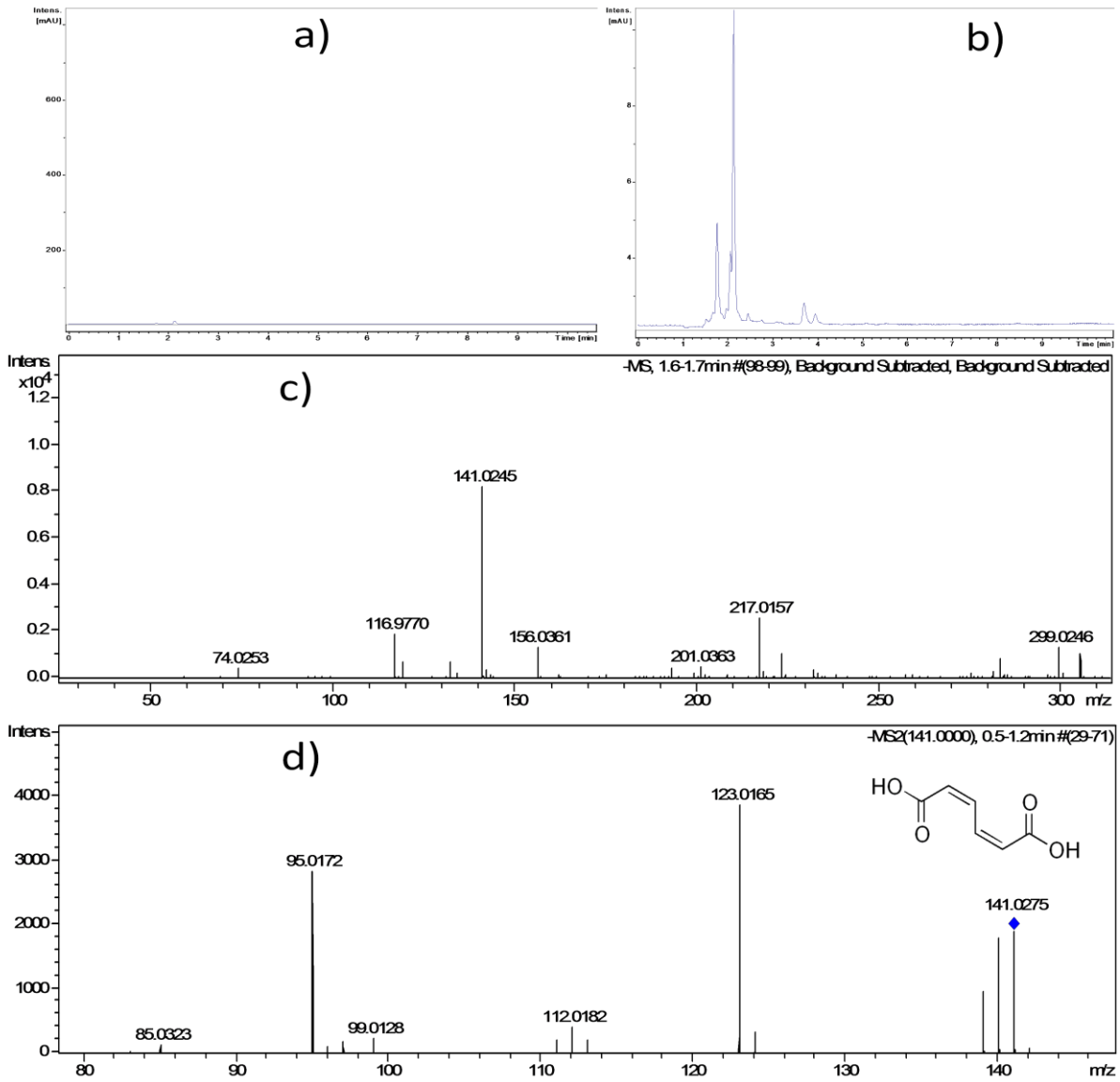


Fig 5.6. LC-MS analysis of metabolites collected at 8 hrs time. Panels a, b, c and d represent LC profile, enlarged LC profile, MS pattern and MS/MS pattern of a compound with retention time 1.6 min.

enzymes catechol 1,2 dioxygenase or catechol 2,3 dioxygenase leading to the formation of *cis-cis* muconate respectively and 2-hydroxymuconic semialdehyde (HMSA) respectively (Fig. 5. 7) (Loh and Chua, 2002). In the *ortho* pathway of catechol degradation, intradiol cleavage occurs leading to the formation of *cis,cis*- muconic acid (Fig. 5. 6) which is converted to β-ketoadipate-enol-lactone and finally to acetyl-CoA and

succinyl-CoA. In the *meta* cleavage pathway, extradiol cleavage of catechol occurs yielding 2-hydroxymuconic semialdehyde (HMSA) which is further converted to pyruvate and acetyl-CoA. Therefore, the existence of *cis, cis* –muconate and absence of HMSA in the metabolites indicate that degradation of benzoate in *Acinetobacter* sp. DS002 occurs through the *ortho* pathway.

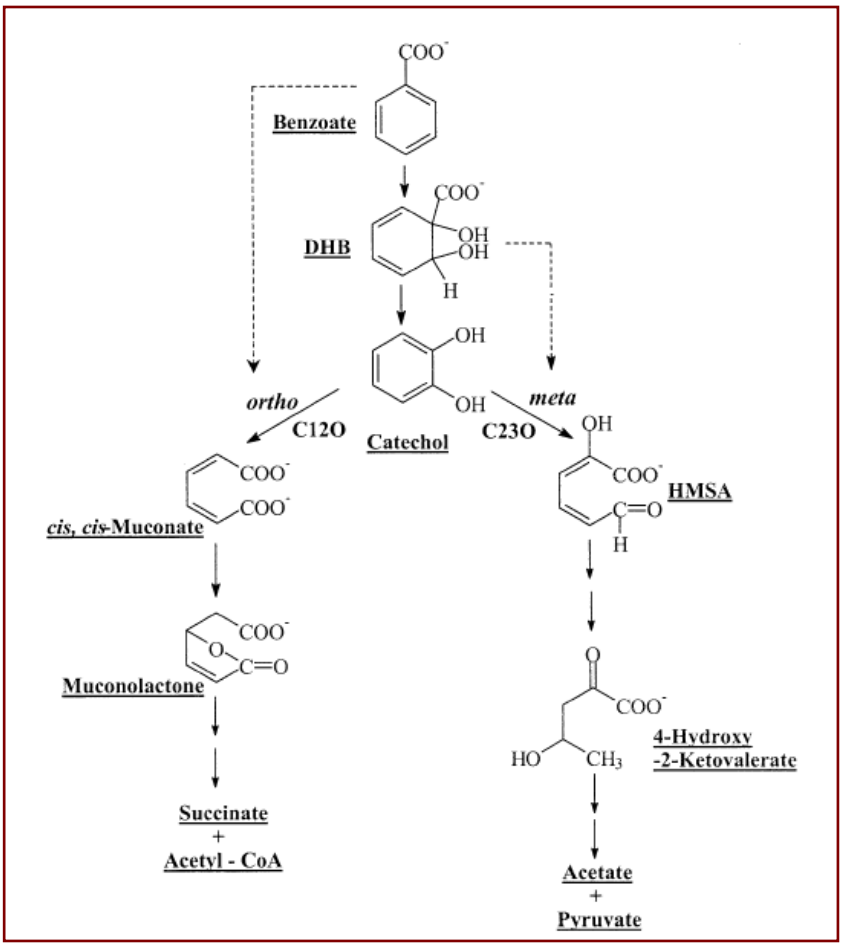


Fig. 5. 7. Schematic representation of o*rtho* and *meta* degradation pathways of benzoate.

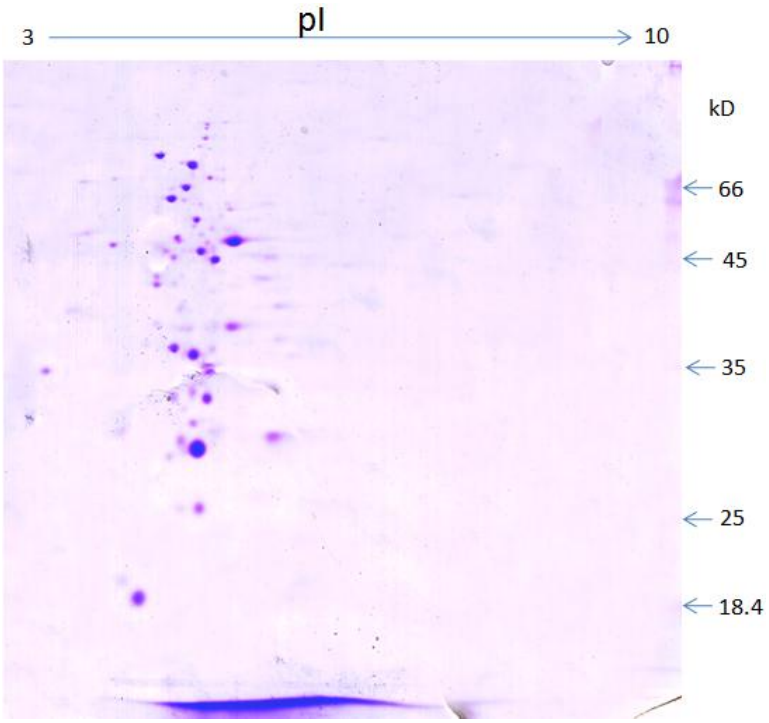
The metabolites generated in this study provide a clear understanding on pathway operational for aromatic compounds degradation in *Acinetobacter* sp. DS002.

5.3 Proteome analysis of *Acinetobacter* sp. DS002

In order to substantiate the data generated through identification metabolites parallel experiments were done to identify corresponding enzymes and to know their

regulation. Here in this study a proteomics approach was followed to gain information pertaining to i) upregulation of enzymes involved in degradation of benzoate ii) generate genome-wide expression profiling in response to shift in carbon source. In order to identify differently and differentially expressed proteins, initially the cells were grown in succinate where the enzymes involved in benzoate degradation are in highly repressed state. In this repressed state the proteins were extracted and basic proteome map was established by performing 2D electrophoresis as described in materials and methods section (Fig. 5. 8). The basic proteome map thus established was then compared with similar maps generated for the soluble proteins extracted from benzoate (5 and 50 mM) grown cultures (Fig.5.9, 5.10). Image analysis was performed using ImageMaster2D platinum software for normalization, spot detection, spot quantification, comparison of gels and for identification of differently and differentially expressed proteins due to shift in carbon source. Upon comparison of the proteome profiles, nearly 75 protein spots were found to be common both in benzoate (5 mM and 50mM) and succinate grown cultures (Fig. 5. 11). Further, the proteome profiles of *Acinetobacter* sp. DS002 grown in 5 mM and 50mM were almost identical matching more than 98.5%, except that the concentration dependent increase was seen in certain spots (Table. 5.1). Further, the protein spots that were showing significant intensities were picked for MALDI-TOF analysis. The 13 proteins that have shown significant score and sequence coverage with the proteins found in database were presented in table (Table 5. 1). Most of the proteins identified through MALDI-MS were essentially involved in degradation of benzoate via *ortho* pathway. Identification of electron transport component of benzoate 1,2 dioxygenase (spot 4344) and catechol 1,2 dioxygenase (spot 4364) through MS/MS, provided primary evidence to show that benzoate is degraded through *ortho* pathway in

Acinetobacter sp. DS002. Studies pertaining to benzoate degradation occasionally



resulted

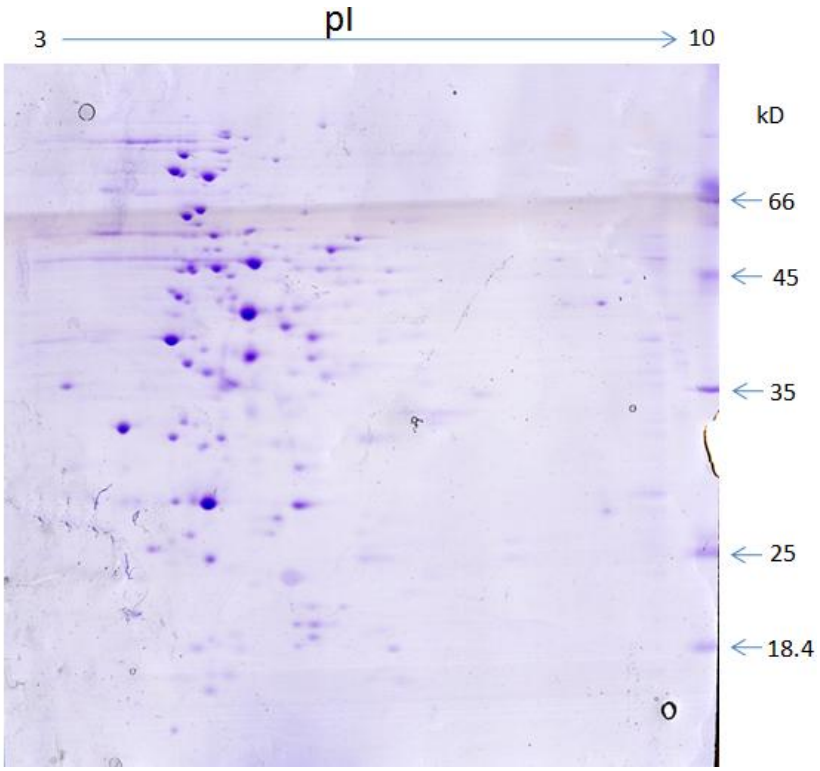


Fig. 5.8. 2D Proteome map generated for soluble proteins of *Acinetobacter* sp. DS002 grown in 10mM succinate

Fig. 5.9. 2D Proteome map generated for soluble proteins of *Acinetobacter* sp. DS002 grown in 5 mM benzoate

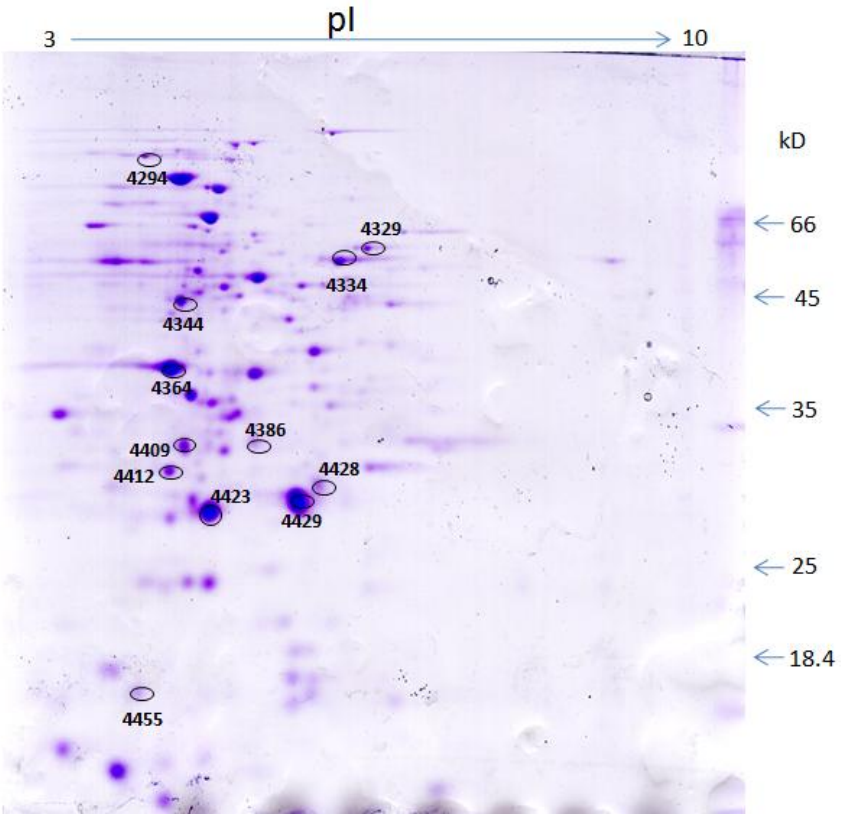


Fig. 5.10 2D Proteome map generated for soluble proteins of *Acinetobacter* sp. DS002 grown in 50 mM benzoate. Differently and differentially expressed spots that were subjected to MALDI-TOF were shown with an open circle. The number adjacent to the spot indicate the spot ID

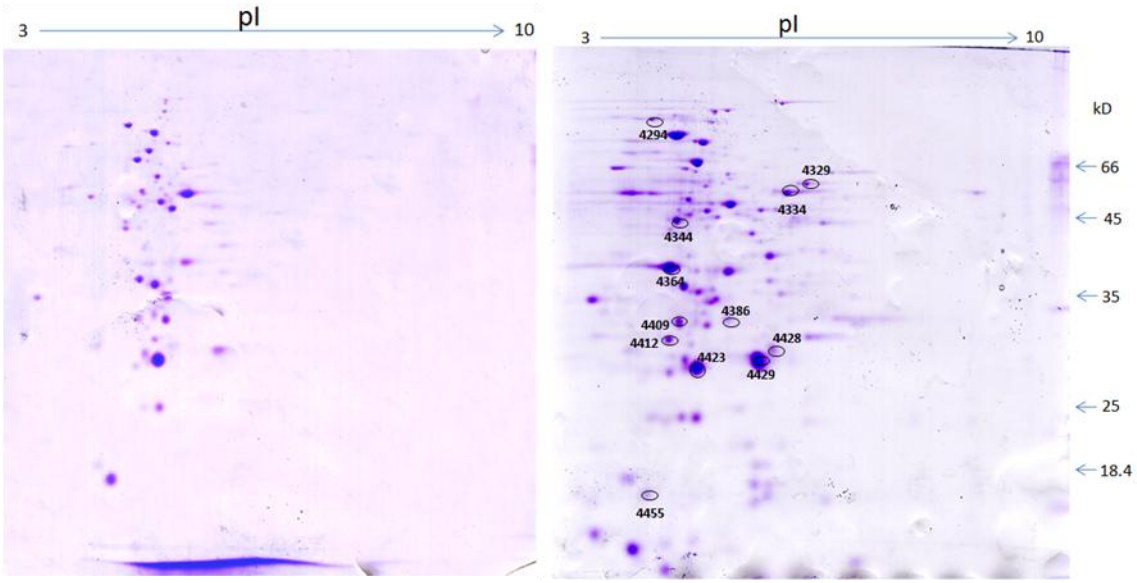
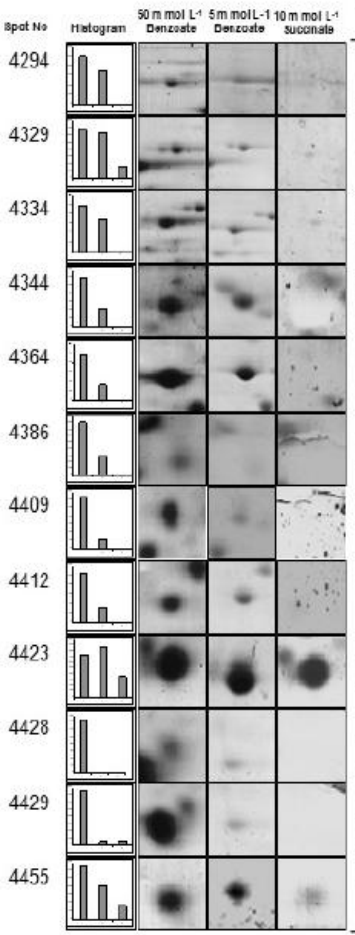


Fig. 5.11. Comparison of 2D proteome profiles of succinate (10 mM) and benzoate (50 mM) grown cultures of *Acinetobacter* sp. DS002. Differently and differentially expressed spots that were subjected to MALDI-TOF were shown with an open circle. The number adjacent to the spot indicate the spot ID.

in identification of isoforms of catechol 1,2 dioxygenase (Nakai et al, 1990). When searches were made at different pI points in the molecular mass range of 34kD no protein spots matched with mass fingerprint pattern of catechol 1,2 dioxygenase suggesting absence of its isoforms in *Acinetobacter* sp. DS002. Usually, benzoate and catechol degrading genes are organized as operons. Expression of these operons is strictly dependent on transcriptional regulator proteins such as BenM and CatM which belong to LysR family of transcriptional regulators (Collier et al, 1998; Romero-Arroyo et al, 1995). Interestingly the MS data of spot 4412 showed considerable similarity to LysR protein suggesting transcriptional regulation of benzoate degrading genes in *Acinetobacter* sp. DS002. Further, spots 4386 and 4428 have shown significant homology to ABC transport related proteins involved in unidirectional movement of a solute across the membranes (Saurin et al, 1999). Benzoate dependent induction of ABC transporters is rather unusual. The involvement of such transporters in unidirectional transport of

aromatic compounds such as phenols is well established (Kurbatov et al, 2006). Instead involvement of novel porin like proteins in benzoate transport is reported (Clark et al, 2002). Significant increase in the concentrations of ABC transporters only in benzoate grown cultures suggests existence of novel transport mechanism in *Acinetobacter* sp. DS002. However, further studies are required to validate this observation.

In addition to the aforementioned protein spots that are directly involved in catabolism of benzoate we have also seen a number of other protein spots in benzoate grown cultures that have no obvious link to benzoate catabolism. One of them is catalase (4294) and its induction very well correlates with the reports of catechol induced oxidative stress in microbes involved in biodegradation of aromatic compounds (Benndorf et al, 2001). The other two protein spots whose identity was established are



Protein ID	Mowse Score	% Sequence coverage*	Accession No
Catalase *Acinetobacter baumannii* ACICU	76	20	gi|184157746
Dihydrolipoamide dehydrogenase *Acinetobacter baumannii*	48	18	gi|169794948
Hypothetical protein R2601_23970 *Roseovarius* sp. HTCC2601	77	23	gi|114766962
Benzoate 1,2-dioxygenase electron transfer component *Acinetobacter baumannii* AYE	50	20	gi|169796581
Catechol 1,2-dioxygenase *Acinetobacter baumannii*	89	36	gi|90018513
ABC transporter related *Anaeromyxobacter* sp. Fw109-5	61	42	gi|153004317
Putative adenylate/guanylate cyclase *Silicibacter* sp. TM1040	61	14	gi|99082352
Transcriptional regulator, LysR family protein *Burkholderia mallei* PRL-20	44	80	gi|167004065
Alkyl hydroperoxide reductase, C22 subunit, *Acinetobacter* sp.	83	48	gi|169633181
ABC transporter related [*Burkholderia phymatum* STM815]	62	17	gi|186471060
Major outer membrane protein P44-4 *Anaplasma phagocytophilum*	62	48	gi|19223945
Hydroxylase for synthesis of 2-methylthio-cis-ribozeatin in tRNA *Klebsiella pneumoniae*	57	18	gi|152973133

* Percentage of protein sequence covered by the matching peptides

Table 5. 1. Magnified regions of 2D gel images with their histogram are represented in panel A. In the histogram the 1st, 2nd and 3rd bars represent protein intensities of a spot in 50 mM benzoate, 5 mM benzoate and 10 mM succinate grown cultures respectively. Panel B represent the hits obtained from MALDI TOF analyses and their accession numbers.

adenylate/guanylate cyclase (4409) and dihydrolipoamide dehydrogenase (4329). The role of cAMP in signaling mechanism is well known. In fact, it is part of the global switch that turns on several operons involved in metabolism of alternate carbon sources (Harman, 2001). Induction of adenylate/guanylate cyclase signifies synthesis of elevated cAMP/cGMP which might be needed for inducing genes required for benzoate catabolism. Dihydrolipoamide dehydrogenase (4329) and hydroxylase (4455) both are needed for operation of TCA cyles. One of them, dihydrolipoamide dehydrogenase plays a key role in formation of succinyl co-A (Kornfeld et al, 1977) and the enzyme hydroxylase has shown high homology to the product of *miaE* involved in conversion of of 2-methylthio-N-6 isopentenyl adenosine (ms2i6A) in to 2-methylthio-N-6 (cis-hydroxyl) isopentenyl adenosine ms2io6A). The ms2io6A is a modified base found adjacent to the anticodon of tRNAs that read codons beginning with "U". In *Salmonella typhimurium* presence of ms2io6A is a prerequisite to facilitate growth on TCA cycle intermediates such as succinate, fumerate or

Spot No	MW of Peptides	Sequence	Protein	Accession No.
4344	1044.494	FPWFEYR	Benzoate 1, 2-dioxygenase electron transfer component	gi|126641261
4344	2011.007	RSPGSGGLFSLAVNPYTCK	Flavodoxin/ Ferredoxin Oxidoreductase domain protein	gi|121996876
4364	1934.2730	RTIEGPLYVAGAPESVGFARM	Catechol 1,2 dioxygenase	Q43984
4364	2012.3720	KVEVWHANSLGNYSFFDKS	Catechol 1,2 dioxygenase	Q43984
4364	2081.3300	RHGNRPSHVHYFVSAPGYR.K	Catechol 1,2 dioxygenase	Q43984
4364	2780.7030	RKLTTQFNIEGDEYLWDDFAFATRD	Catechol 1,2 dioxygenase	Q43984

Table 5. 2. MS/MS hits obtained for protein spots 4344 and 4364. Sequence and MW of peptides that matched with either benzoate 1,2 dioxygenase or catechol 1,2 dioxygenase are provided.

malate (Persson et al, 1998). Upregulation of MaiE homologue in benzoate grown culture points towards enhanced biosynthesis of tRNA molecules having ms2io6A at 37th position. As end product of benzoate catabolism is a TCA intermediate, induction of MaiE during benzoate catabolism might be to enhance tRNA population with ms2io6A at 37th position which might be required for optimal operation of TCA cycle in *Acinetobacter* sp. DS002.

5.4 Cloning of *cat* operon

Organization of genes responsible for conversion of catechol to TCA intermediates was reported in several microbes (Harwood and Parales, 1996). A reverse genetic approach was employed in order to identify genes responsible for catechol degradation. As mentioned earlier, peptide mass fingerprint (PMF) of protein spot 4364 has shown significant similarity to catechol 1, 2 dioxygenase. The PMF hits were further confirmed by MS/MS analysis. The PMF of protein spot 4364 and its MS/MS data matched with catechol 1, 2 dioxygenase of *Acinetobacter* sp. (Q43984). A blast analysis was performed taking the *de novo* sequence of the protein spot 4364 and the catechol 1, 2 dioxygenase sequences of *Acinetobacter* genus found in the database. The alignment has shown high degree of sequence conservation throughout the length protein; except that the per cent homology was found to be little lower at the C-terminus of these proteins (Fig. 5. 12). Among these conserved regions two blocks that showed absolute sequence identity were identified. These two conserved sequence blocks (TPRTIEGPLYVAGA and DDFAFATRD) were taken to generate degenerate primers COF2 (5'-ACNCCNMGNACNATHGARGG-3') and COR1 (5'-CKNG TNGCRAANGCRAARTCRTC -3') for amplification of catechol 1, 2 dioxygenase gene from *Acinetobacter* sp. DS002 (Fig. 5. 12). When a typical PCR reaction was performed using genomic DNA of *Acinetobacter* sp. DS002 as template and COF2 and COR1 as primers an

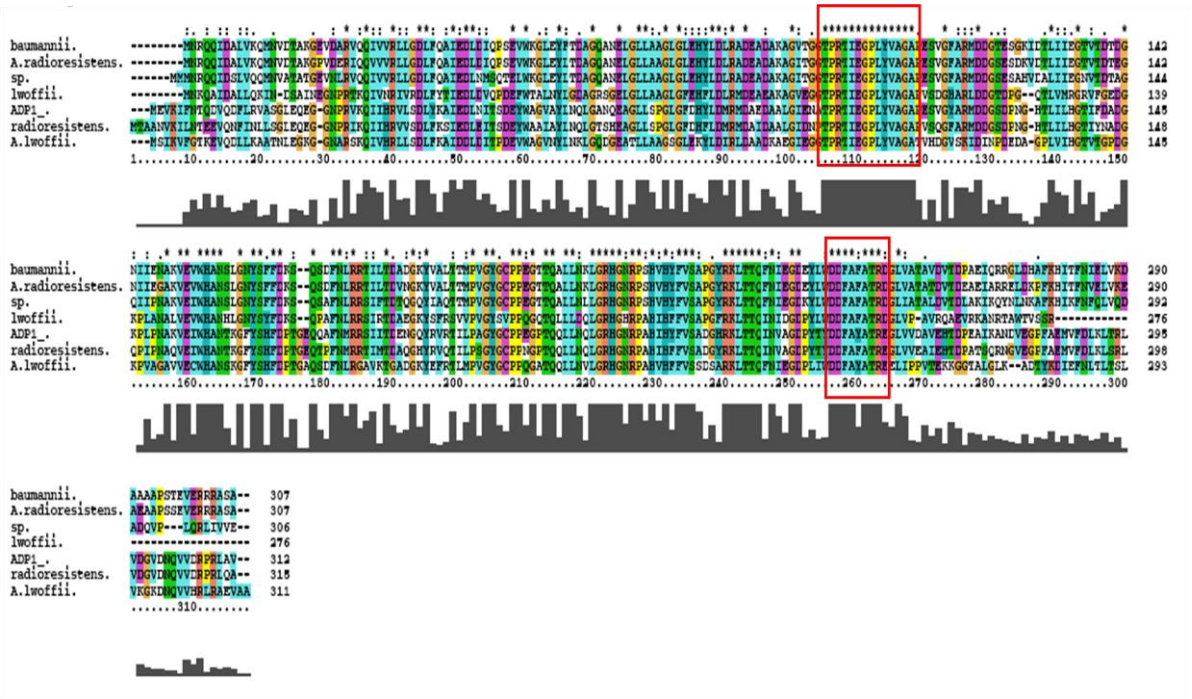


Fig 5. 12. Multiple alignment of various catechol 1,2 dioxygenases of the genus *Acinetobacter*. Conserved blocks of amino acids marked with red boxes were used to design degenerate primers.

amplicon of 500 bp was generated (Fig. 5. 13A). No amplicon was observed in the control reactions where genomic DNA of *E. coli* DH5α was used as template. Finally the identity of the 500 bp amplicon was confirmed by generating complete sequence. The sequence showed 98% identity with catechol 1,2 dioxygenase of *Acinetobacter baumannii* ATCC 17978 (Fig. 5. 14). The partial *catA* gene was then used as a probe to identify a fosmid clone containing complete *cat* / *ben* operons. When colony hybridization was performed five independent clones have given positive signals (Fig. 5.13B). Among these five positives, clone 912 was selected for further studies. In order to identify a fragment having the catechol 1,2 dioxygenase gene, a restriction profile of 912 clone was generated by digesting with restriction enzymes *Bam*HI, *Eco*RI, *Hind*III and *Sal*I . These fragments were again hybridized with partial *catA* gene to identify the profile that gives *cat* operon as a single restriction fragment. An 8kb *Sal*I fragment has been shown to have *cat* operon and was then

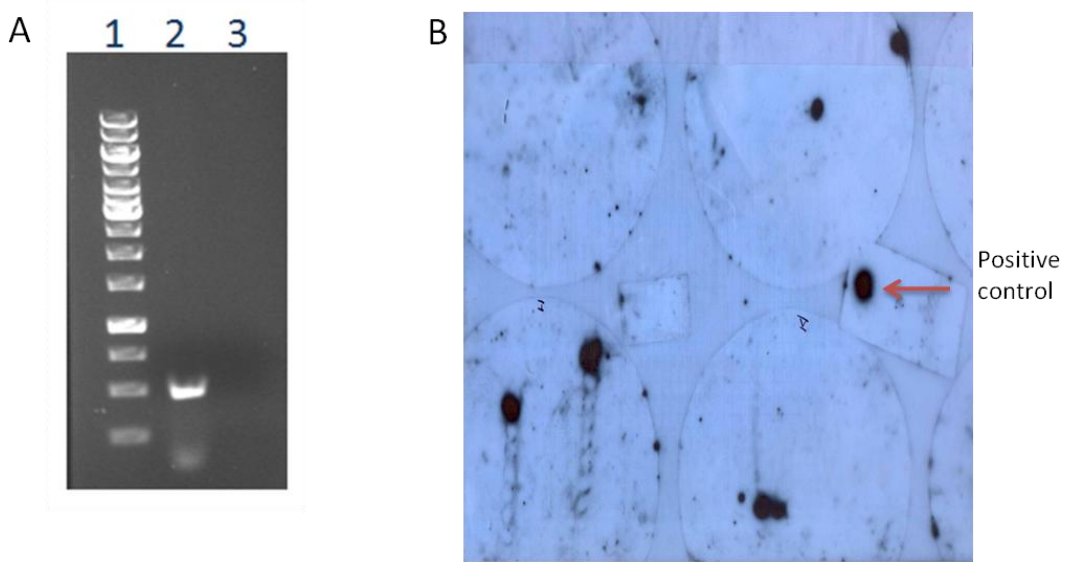


Fig. 5. 13. A) Amplification of *catA* gene using degenerate primers. Lane 1 represents 1 kb DNA ladder and lanes 2 and 3 represent amplicons obtained using genomic DNA of *Acinetobacter* sp. DS002 and *E.coli.* B*)* Identification of *catA* containing genomic clones by colony hybridization. The positive control is shown with an arrow mark.

```
>gb|CP000521.1| Acinetobacter baumannii ATCC 17978, complete genome
Length=3976747

 Features in this part of subject sequence:
   CatA3

 Score =  789 bits (427),  Expect = 0.0
 Identities = 455/469 (98%), Gaps = 0/469 (0%)
 Strand=Plus/Plus

Query  71       TACGCCGCGGACGATCGAGGGTCCACTTTATGTTGCTGGCGCACCTGAATCAGTTGGCTT  130
                ||| || || || ||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  2147156  TACACCACGTACTATCGAAGGTCCACTTTATGTTGCTGGCGCACCTGAATCAGTTGGCTT  2147215

Query  131      TGCACGTATGGATGACGGAACCGAGACTGGCAAAATCGATACCTTAATTATTGAAGGTAC  190
                ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Sbjct  2147216  TGCACGTATGGATGACGGAACTGAGACTGGCAAAATCGATACCTTAATTATTGAAGGTAC  2147275

Query  191      GGTAACCGACACTGATGGCAATATTATTGAAAATGCCAAAGTTGAAGTATGGCATGCCAA  250
                ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2147276  GGTAACCGACACTAATGGCAATATTATTGAAAATGCCAAAGTTGAAGTATGGCATGCCAA  2147335

Query  251      CAGTTTAGGTAACTATTCATTCTTTGATAAGTCACAATCTGACTTTAACTTACGCCGTAC  310
                ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Sbjct  2147336  CAGTTTAGGTAACTATTCATTCTTTGATAAGTCACAATCTGACTTTAACTTACGTCGTAC  2147395

Query  311      CATTTTCACTGATGCAGATGGTAAATATGTAGCGTTAACCACTATGCCAGTTGGTTATGG  370
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2147396  CATTTTCACTGATGCAGATGGTAAATATGTAGCGTTAACCACTATGCCAGTTGGTTATGG  2147455

Query  371      TTGCCCTCCTGAAGGTACAACACAGGCTCTTCTTAACAAGTTAGGCCGTCATGGTAACCG  430
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2147456  ATGCCCTCCTGAAGGTACAACACAGGCTCTTCTTAACAAGTTAGGCCGTCATGGTAACCG  2147515

Query  431      TCCATCTCACGTTCACTACTTTGTATCTGCACCGGGTTACCGCAAGCTGACTACTCAATT  490
                |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Sbjct  2147516  TCCATCTCACGTTCACTACTTTGTATCTGCACCAGGTTACCGCAAGCTGACTACTCAATT  2147575

Query  491      CAACATTGAGGGTGATGAGTATTTATGGGACGACTTTGCGTTCGCAACT  539
                |||||||||||||||||||||| |||||||| |||||||| ||||| |||
Sbjct  2147576  CAACATTGAGGGTGATGAGTACTTATGGGATGACTTTGCCTTCGCTACT  2147624
```

Fig. 5. 14. BLASTN analysis of *catA* of *Acinetobacter* sp. DS002. Pairwise alignment of *catA* of *Acinetobacter* sp. DS002 with *catA3* of *Acinetobacter baumannii* 17978 is shown.

used for further studies by sub-cloning it in pBluescript vector (Fig. 5.15). Sequencing of the 8kb *Sal*I fragment of the clone 912 showed presence of genes which code for catechol 1, 2 dioxygenase, 3-oxo-acid-CoA transferase and β-ketoadipate thiolase (Fig. 5. 16).

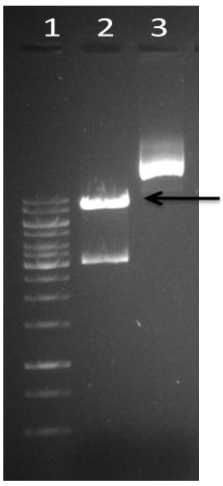


Fig. 5. 15. Sub-cloning of *catA* containing fragment into pBluescript II KS vector.Lane 1. represents 1kb DNA ladder. Lane 2 represents recombinant pBluescript plasmid having catA gene as a *Sal*I fragment. Lane 3 represents uncut recombinant plasmid.

Based on the identification of catabolic intermediates, PMF and MS/MS data and sequence information a pathway has been constructed for degradation of benzoate through the ortho pathway (Fig. 5. 16)

5.6 Purification of Catechol 1,2 dioxygenase

Catechol and substituted catechols occupy central position in biodegradation of aromatic compounds. Degradation of catechols further proceeds through either *ortho* or *meta* cleavage pathways depending on the availability of microbial dioxygenases. Catechol 1,2 dioxygenase, responsible for the intradiol cleavage of catechol, channels catechol degradation through ortho pathway, whereas catechol 2,3 dioxygenase diverts it towards *meta* cleavage pathway. The catachol 1, 2 dioxygenase is therefore a key enzyme in

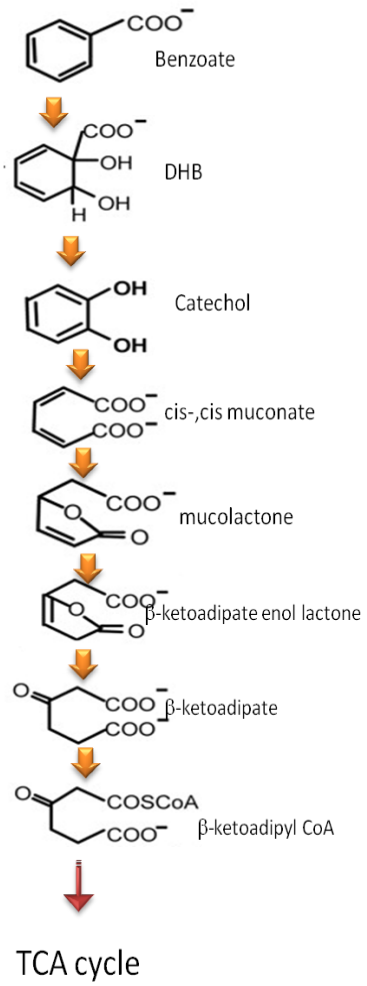


Fig 5. 16. Degradation pathway of benzoate in *Acinetobacter* sp. DS002

mineralization of catachols and substituted catechols. Phenolic and nirophenolic compounds generated during biodegradation of OP compounds will be converted to either catechols or nitrocatechols through the action of *p*-nitrophenol monoxygenase (Spain et al., 1979; Spain, 1994; Spain and Gibson, 1991; Zeyer and Kocher,1988). The nitrocatechol generated will be converted to benzentriol which is then serves as substrate for ring fission oxygneases (Hanne et al, 1993; Jain et al, 1994; Kadiyala and Spain, 1998; Chauhan et al, 2000). Alternatively, in certain cases the *p*-nitrophenol is converted to benzene triol *via* generation of hydroquinone. A detailed figure showing details of 4-nitrophenol degradation is given for quick reference (Fig. 5. 18). If fair assessment has to be made for further

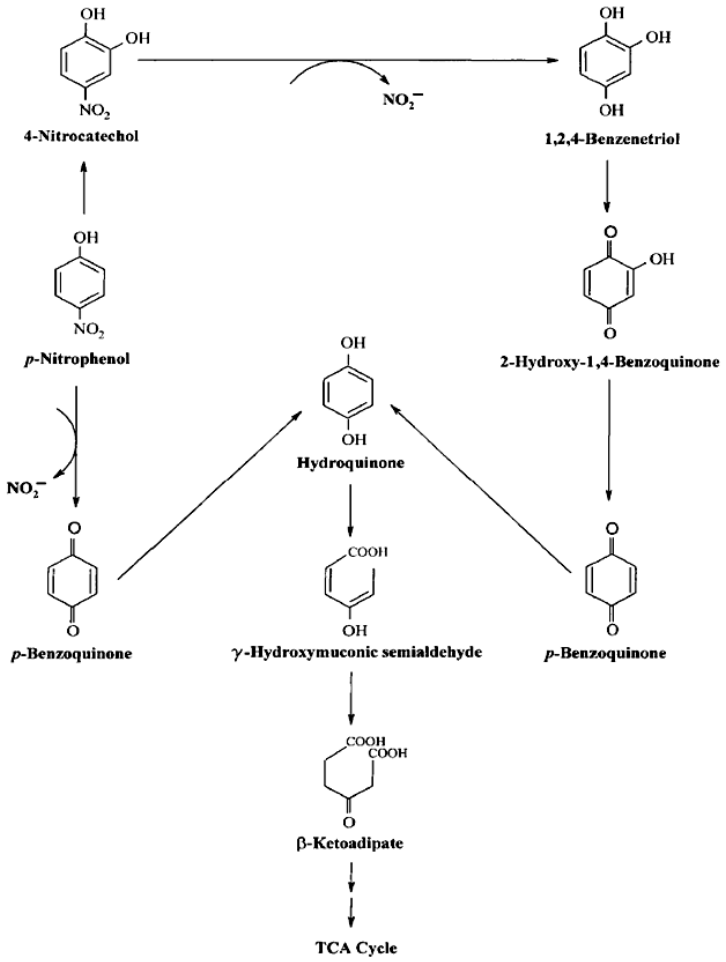


Fig. 5. 18. Degradation pathways of *p*-nitrophenol via 1,24 benzenetriol and benzoquinone

manipulation of *Acinetobacter* sp. DS002 it is necessary to assess the substrate specificity of catechol 1,2 dioxygenase using the intermediates of PNP degradation pathway as substrates. While attempting to do such experiments under *in vitro* conditions pure enzyme is necessary. Therefore an attempt was made to purify catechol 1,2 dioxygenase to electrophoretic homogeneity by following conventional protein purification techniques described in materials and methods section. The purification procedure gave 1 mg of protein from 20 g of cell pellet. Fractions containing more than 50% catechol 1,2 dioxygenase was taken at each stage of purification and the specific activity and fold purification was determined. Though the fold purification has increased with every stage of

purification, the gel permeation chromatography has given virtually homogenous catechol 1, 2 dioxygenase (Fig. 5. 19). Further, the protein fraction having activity showed a native mass of 66 kDa. The very same protein when analyzed on SDS-PAGE showed a molecular mass of 34 kDa (Fig. 5. 19d). Most of the C12Os with few exceptions are dimmers of identical or non-identical subunits (Aoki et al, 1984; Nakai et al, 1990). Gel permeation results indicate that C12O of *Acinetobacter* sp. DS002 is a homodimer with a molecular mass of 66 kDa.

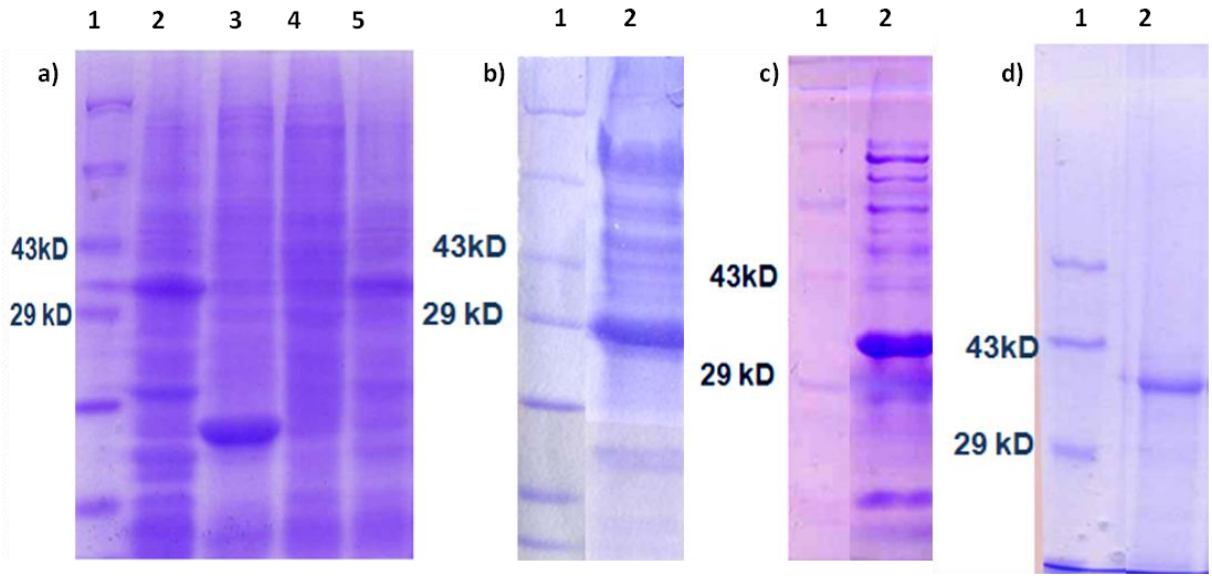


Fig. 5. 19. Purification of catechol 1, 2 dioxygenase of *Acinetobacter* sp. DS002. Panel a) represents ammonium sulphate fractionation. Lane 1 represents protein molecular weight marker. Lane 2 represents cytoplasmic fraction. Lanes 3-5 represent proteins obtained through 0-20% (Lane 3), 20-40% (Lane 4) and 40-60% (lane 5) saturation of ammonium sulphate. Panels b, c, d represent C12O purified using anion exchange (b), hydrophobic interaction (c) and gel permeation chromatography, respectively. In all panels lane 1 represents molecular weight markers, Lane 2 represents C12O. The C12O purified to electrophoretic homogeneity is seen in panel d.

5.7 Catechol 1, 2 dioxygenase assay

Substrate specificity was determined for C12O of *Acinetobacter* sp. DS002 using catechol and substituted catechols generated during OPH mediate hydrolysis of methyl parathion/ parathion degradation pathway by following procedures described in methods

section. In general the C12Os have showed relaxed specificity for catechols and methyl catechols but they have not shown any activity on nitro-catechols (Patel et al, 1976). However, C12O of *Acinetobacter* sp. DS002 has shown considerable activity when 1, 2, 4-benzenetriol and 4-nitrocatechol were used as substrate (Fig. 5. 20). Existence of C12Os with relaxed substrate specificity is not uncommon in literature. C12Os showing activity on methyl and halocatechols were seen in *Rhodococcus*, *Ralstonia* and *Pseudomonas arvilla* (Cha, 2006; Patel et al, 1997; Briganti et al, 1976; Briganti et al, 2000; Wang et al, 2006). The substituted phenols like 4-nitrophenol, 4-nitrocatechol and 1,2,4 benzenetriol are degradation products of methyl parathion and parathion (Pakala et al, 2007; Chauhan et al, 2000; Jain et al, 1994; Kadiyala and Spain, 1998). As benzenetriol and 4-nitrocatechol served

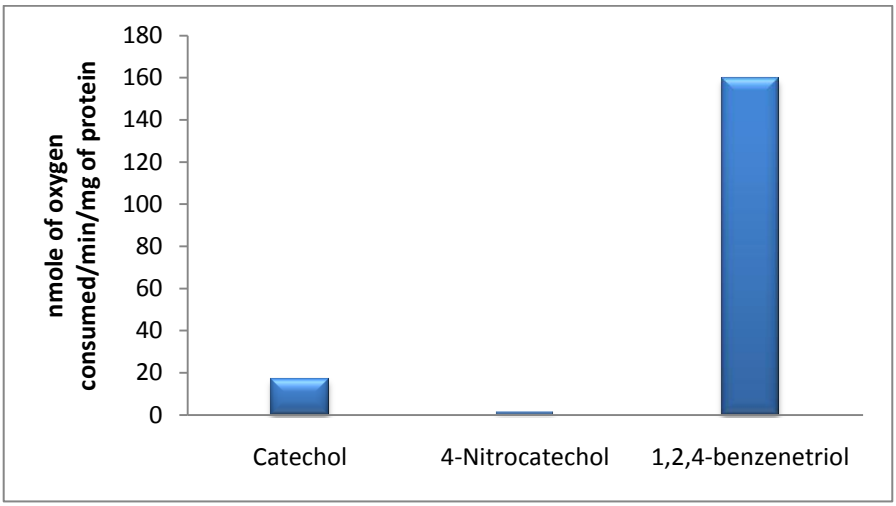


Fig. 5. 20. Activity of catechol 1,2 dioxygenase on catechol, 4-nitrocatechol and 1,2,4-benzenetriol respectively.

as substrate for C12O of *Acinetobacter* sp. DS002 channeling of these PNP intermediates appears to be possible through the *ortho* pathway.

5.8 Manipulation of *Acinetobacter* sp. DS002

Though *Acinetobacter* sp. DS002 was isolated from OP-polluted soils existence of *opd* gene, which codes for organophosphate hydrolase (OPH) involved hydrolytic cleavage of structurally diverse group of OP compounds was not apparent. The *Acinetobacter* strains were not tested positive for OPH activity. In agreement of this observation no amplification was seen when PCR was performed using *opd* specific primers. If the relaxed substrate specificity of C12O has to be exploited for complete mineralization of OP-compounds, expression of *opd* gene in *Acinetobacter* sp. DS002 is inevitable (Cha, 2006; Patel et al, 1997; Briganti et al, 1976; Briganti et al, 2000; Wang et al, 2006). The OPH activity alone can generate nitrophenols from OP-compounds, which serve as substrate for C12O. Therefore in the present study an attempt was made to mobilize a derivative of organophosphate degrading (*opd*) plasmid, pPDL2 of *Flavobacterium* sp. ATCC 27551 into *Acinetobacter* sp. DS002.

5.9 Mobilization of pPDL2 Tn*5*<R6Kγori/KAN-2>into *Acinetobacter* sp. DS002

As described in the previous chapter plasmid pPDL2 is a mobilizable plasmid. The derivative of pPDL2 generated by inserting mini-transposon Tn*5*<R6Kγori/KAN-2> was shown to be mobilizable. Therefore a triparental mating experiment was performed to transfer pPDL2:: Tn*5*<R6Kγori/KAN-2> into the aromatic compound degrading *Acinetobacter* sp. DS002. After mobilization of pPDL2- Tn*5*<R6Kγori/KAN-2> into *Acinetobacter* sp. DS002, its stability was frequently tested by monitoring the presence of PCR using *opd* specific primers (Fig. 5. 21). Interestingly plasmid pPDL2::Tn5<R6Kγori/Kan2> was found to be highly stable. Even in the absence of selection on kanamycin the plasmid was found to be highly stable. In consistence of its presence high amounts of OPH activity was found in all

Acinetobacter sp. DS002 (pPDL2::Tn*5*<R6Kγ*ori*/KAN-2>) clones (Fig. 5. 22).After establishing stable maintenance of plasmid pPDL2::Tn5<R6Kγ*ori*/Kan2> in *Acinetobacter* sp. DS002,

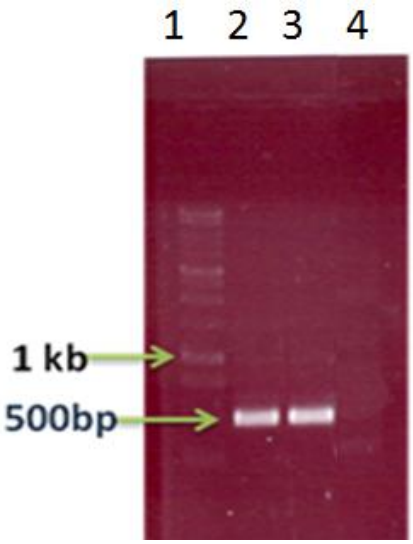


Fig. 5. 21. Confirmation of pPDL2 mobilization into *Acinetobacter* sp. DS002 by colony PCR using *opd* specific primers.Lane 1 represents 1 kb DNA ladder. Lanes 2-3 represent amplicons obtained from pPDL2::Tn5<R6Kγori/Kan2> harbouring *Acinetobacter* sp. DS002 and *E.coli pir*116 respectively. Lane 4 represents *E.coli pir*-116.

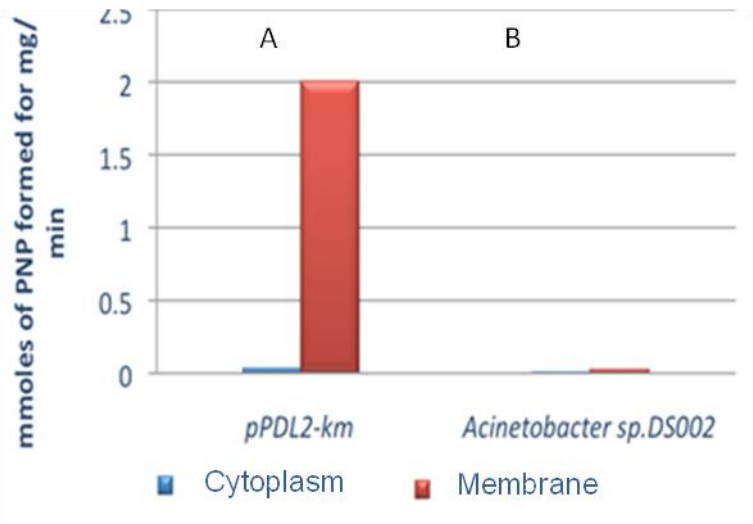


Fig. 5. 22. Assay of Phosphotriesterase activity in cytoplasmic and membrane fractions of A) *Acinetobacter* sp. DS002 having plasmid pPDL2::Tn5<R6Kγ/Kan2> and B) wild type *Acinetobacter* sp. DS002.

experiments were conducted to test their ability to degrade OP compound, paraoxon. Our lab has recently shown presence of OPH in the inner membrane of *B. diminuta* and its

dependence on Twin Arginine Transport (Tat) pathway for membrane targeting (Gorla et al, 2009). Therefore the *Acinetobacter* cells harboring plasmid pPDL2:: Tn*5*<R6Kγori/KAN-2>were fractionated to cytoplasmic and particulate fractions and were assayed for OPH activity as described in methods. Most of the OPH activity was found in membrane fraction and very little activity was seen cytoplasmic fraction, (Fig. 5. 22).

5.10 Degradation of methyl parathion

After establishing the stability and expression of OPH in *Acinetobacter* sp. DS002 (pPDL2:: Tn*5*<R6Kγ*ori*/KAN-2>), the manipulated strain was tested for its ability to degrade organophosphates like methyl parathion and parathion and their catabolic intermediates such as 4-nitrocatechol and 1, 2, 4-benzenetriol generated during their biodegradation. Immediately after adding parathion to the culture medium containing *Acinetobacter* sp. DS002 (pPDL2:: Tn*5*<R6Kγ*ori*/KAN-2>), it quickly turned into yellow colour indicating OPH mediated hydrolysis of methyl parathion. Such change of colour was not observed in control cultures having wild type strains of *Acinetobacter* sp. DS002. Subsequently the yellow colour of the medium generated due to the formation of *p*-nitrophenol from methyl parathion got slowly disappeared from the culture medium. When the culture medium was extracted for identification of *p*-nitrophenol metabolites both 4-nitrocatechol and 1,2,4-benzenetriol were identified indicating that the *p*-nitrophenol generated due to OPH mediated hydrolytic cleavage is further metabolized to generate nitrocatechol and 1,2,4-benzenetriol . As shown in the aforementioned sections benzenetriol has served as one of the substrates for the ring cleavage enzyme C12O purified from *Acinetobacter* sp. DS002. Such observation supports channelization of benzenetriol into TCA cycle indicating the possibility of *p*-nitrophenol serving as carbon source in *Acinetobacter* sp. DS002 (pPDL2:: Tn*5*<R6Kγori/KAN-2>). As shown in figure 5.18 PNP monooxygenase plays a critical role in conversion of PNP into 4-

nitrocatechol / 1, 2, 4-benzenetriol. The total sequence of pPDL2 presented in the first chapter gave no indication of PNP monooxygenase. As formation of nitrocatechol and 1, 2, 4-benzenetriol were found in the culture medium of *Acinetobacter* sp. DS002 (pPDL2:: Tn*5*<R6Kγori/KAN-2>) it suggests existence of such monooxygenases on the chromosome of *Acinetobacter* DS002. *Acinetobacter* sp. DS002 is a soil isolate and shown to grow on number of aromatic compounds. Though PNP at higher concentrations is shown to be toxic to *Acinetobacter* sp. DS002, it has supported for the growth of the strain at low concentrations. In the light of these observations degradation of PNP generated from OP-compounds is an understandable consequence. A schematic degradation pathway found to be operational in *Acinetobacter* sp. DS002 has been presented in Fig. 5. 23.

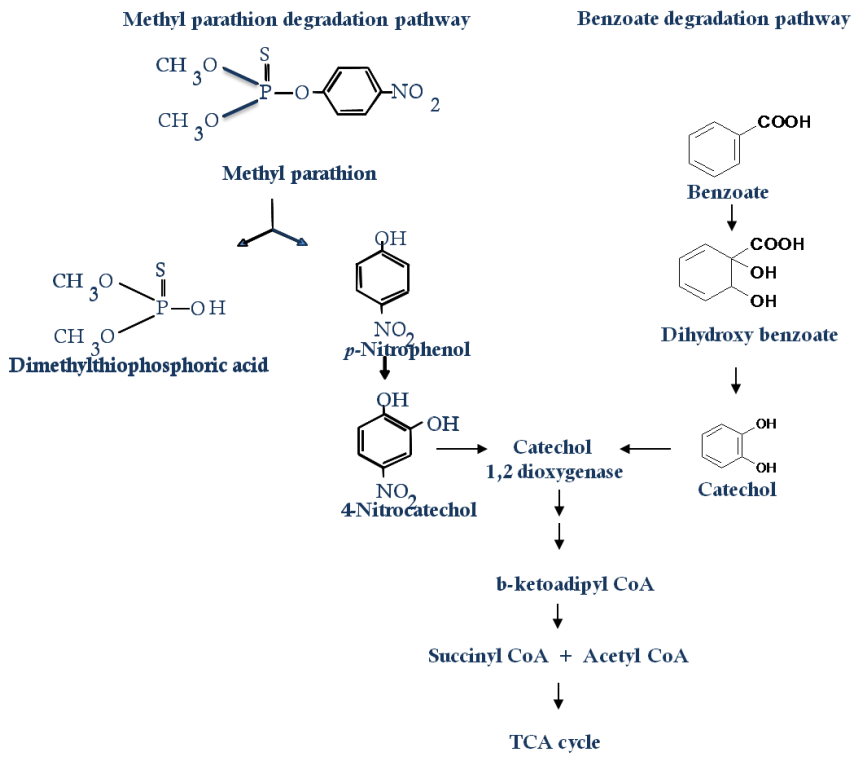


Fig. 5. 23. Schematic representation of methyl parathion and benzoate degradation pathways in *Acinetobacter* sp. DS002 (pPDL2:: Tn*5*<R6Kγ*ori*/KAN-2>)

Conclusions

1. *Acinetobacter* sp. DS002 utilizes benzoate as sole source of carbon. The optimum concentration of benzoate for growth is 5 mM.
2. Catechol and *cis,cis*-muconic acid were identified as intermediates of benzoate catabolism in *Acinetobacter* sp. DS002.
3. Basic proteome maps of soluble proteins of *Acinetobacter* sp. DS002 grown in 10 mM succinate, 5 mM and 50 mM benzoate were generated.
4. Proteome maps of 5 mM and 50 mM benzoate grown cultures were identical whereas the proteome map of benzoate grown cultures sowed significant differences with the proteome map of succinate grown cultures.
5. Benzoate 1,2 dioxygenase and catechol 1,2 dioxygenase, key enzymes of benzoate degradation were identified through MALDI-TOF and MS/MS analysis.
6. Degenerate primers were designed for *catA* gene and *catA* gene amplified from *Acinetobacter* sp. DS002 showed 98% identity with *catA* gene of *Acinetobacter baumannii* 17978.
7. Genomic library clones of *Acinetobacter* sp. DS002 with *cat* operon were identified. The *cat* operon was sub-cloned and sequenced.
8. Catechol 1,2 dioxygenase (C12O) of *Acinetobacter* sp. DS002 was purified to electrophoretic homogeneity. C12O shows significant enzyme activity towards PNP intermediates 4-nitrocatechol and 1, 2, 4- benzenetriol.
9. Plasmid pPDL2:: Tn*5*<*R6Kγori*/KAN-2> was successfully mobilized into *Acinetobacter* sp. DS002. *Acinetobacter* (pPDL2::Tn*5*<*R6Kγori*/KAN-2>) has successfully degraded OP compounds Paraoxon and methyl parathion.