DOCKING OF POLYKETIDE SYNTHASE FROM PGPR2 AGAINST SELECTED TARGETS FROM PHYTOPATHOGENS.

Illakkiam D1 and Usha Raja Nanthini A2.
1. Assistant Professor, Department of Biotechnology, Mother Teresa Women’s University, Kodaikanal.
2. Head, Department of Biotechnology, Mother Teresa Women’s University, Kodaikanal.

Abstract

Pseudomonads possess many traits that make them well suited as biocontrol and growth-promoting agents. PGPR influences plant growth either by direct or indirect mechanism. One of the most effective mechanisms that Plant Growth Promoting Rhizobacteria employs to prevent proliferation of phytopathogens is the synthesis of antibiotics. A wide range of different antibiotics synthesized is the PGPR trait and their specificity and mode of action have been studied detail and only some of these strains have been commercialized. Among the various groups of antibiotics produced by the PGPR, the polyketides such as 2,4 Diacetyl phloroglucinol, Pyoluteorin and Mupirocin are highly effective in suppression of plant pathogens. In this present study, a strain isolated from rhizosphere was found that Polyketides 3, 4-dihydroxy-N-methyl-4-(4-oxochroman-2-yl) butanamide, a novel secondary metabolite synthesised by the strain PGPR2 showed antagonism against various phytopathogens. Comparative analyses of bacterial genomes associated with fungal viability and virulence has led to the identification of many putative targets for novel antifungal agents. Genetic determinants that could contribute its plant growth promotion and biocontrol abilities have been identified by whole genome sequencing and comparative analyses with other strains of P. aeruginosa. Whole genome analysis further enhances our ability to study the secondary metabolite production and its regulation in this rhizosphere bacterial strain.

Introduction:

Plant diseases caused by soil-borne root pathogens account for major crop losses worldwide. To control the fungal infection demands the use of fungicides, but the residues of which remained in soil and ground water are a cause for concern to human beings. The use of biocontrol agents against soil-borne phytopathogens is an attractive alternative to chemical fungicides considering the potential health and environmental hazards in the usage of chemicals in plant disease management. Bacteria from soil can be used as a biological control of soil borne plant diseases (Kloeper and Beauchamp, 1992). Rhizobacteria have also been studied as plant growth promoters for increasing agricultural production and as biocontrol agents against plant diseases (Burris, 1998; Chen et al., 1996). An array of bacterial genera including but not limited to Pseudomonas, Azospirillum, Azotobacter, Bacillus, Klebsiella, Enterobacter, and others have been tested for biocontrol activity.
Xanthomonas and Serratia have been shown to promote plant growth. Among them fluorescent Pseudomonads are known to exhibit antifungal activity with varying degrees of antagonism (de Weger et al., 1986).

Fluorescent pseudomonads have been reported as potent biocontrol agents against soil-borne plant pathogens. They suppress the growth of fungal pathogens by producing secondary metabolites (antibiotics, iron-chelating siderophores and cyanide) and fungal cellular component degrading enzymes (chitinases, proteases and lipases) (Hazem et al., 2012). The antifungal metabolites that are characterized till date include, 2,4 Diacetyl pholoroglucinol, Pyoluteorin, Pyrrolnitrin, Phenazine-1-Carboxamide, Tensin, Viscosinamide, Aerugine, Amphisin, Furanone, 2-acetamidophenol, Rhamnolipids, Butyro lactones, Cepaciamide A, and Oomycin A. (Bhattacharyya et al., 2012; Lee et al., 2003; Sorensen et al., 2001; Paulitz et al., 2000; Slininger et al., 2000; Kim et al., 2000).

Among the various groups of antibiotics produced by the PGPR, the polyketides such as 2, 4 Diacetyl pholoroglucinol, Pyoluteorin and Mupirocin are highly effective in suppression of plant pathogens. Naturally occurring polyketides are produced by the successive condensation of small carboxylic acids in a process that resembles fatty acid synthesis. The genes encoding polyketide synthases (PKSs) show similarities to one other as well as to genes for fatty acid synthases (Hopwood et al., 1990; Hutchinson et al., 1995).

Earlier we reported a strain isolated from rhizosphere was found that Polyketides namely 3, 4 dihydroxy-N-methyl-4-(4-oxochroman-2-yl) butanamide, a novel secondary metabolite synthesized by the strain Pseudomonas sp. PGPR2 showed antagonism against various phytopathogens (illakiam et al. 2013). In this study, we report protein-protein interaction between antifungal metabolite protein namely polyketide synthase identified from whole genome sequence of Pseudomonas sp. PGPR2 and the selected targets such as Ergosterol, β-(1,3)-D-Glucanase , DNA gyrase, RNA polymerase and Chitinase from various phytopathogens.

Materials and Methods:-
Sequence data retrieval:-

Pseudomonas sp. PGPR2 has been deposited in the Genbank database with accession number ASQO0000000. Polyketide synthase amino acid sequences were retrieved in FASTA format from GenBank (http://www.ncbi.nlm.nih.gov/Genbank/) and Swiss-Port (http://www.expasy.ch/sprot/). Subsequently, a search was done using the BLASTp program (Altschul et al., 1990) at NCBI to identify and retrieve potentially related sequences. A homology threshold (E-value) of 0.01 was set to get significant hits to be included in the alignment comprising subfamilies. Target enzyme sequences such as ergosterol (2A1B), β-(1,3)-D-Glucanase (4M80) , DNA Gyrase (1BGW), RNA polymerase (3M3Y)and Chitinase (1O6I) were retrieved in PDB format from PDB database (http://www.rcsb.org/pdb).

Alignment process:-
Polyketide synthase amino acid sequences were aligned using the multiple sequence alignment program ClustalW2. (Thompson et al., 1997) Default gap opening and extension penalty were used. A careful manual examination was done to avoid misalignment. Gaps in the alignment were treated as missing data. Proteins were named using an abbreviation followed by the family name in the alignment.

Sequence and structural analysis:-
Amino acid composition of selected Polyketide synthase proteins were analyzed by ProtParam, an online tool available on ExPASy proteomics server (Gasteiger et al., 2005). Secondary structures were predicted by SOPMA on EMBOSS server. InterProScan was used to identify conserved domains in selected sequences. PSIPRED, a highly accurate secondary structure prediction method available at http://globin.bio.warwick.ac.uk/psipred. SWISS-MODEL (http://swissmodel.expasy.org) is a server for automated comparative modeling of three-dimensional (3D) protein structures. It pioneered the field of automated modeling starting in 1993 and is the most widely-used free web-based automated modeling facility today. LOMETS (Local Meta-Threading-Server) (http://zhanglab.ccb.med.umich.edu/LOMETS/) is an on-line web service for protein structure prediction. It generates 3D models by collecting high-scoring target-to-template alignments from 9 locally-installed threading programs (FFAS-3D, HHsearch, MUSTER, pGenTHREADER, PPAS, PRC, PROSPECT2, SP3, and SPARKS-X). ClusPro: an automated docking and discrimination method for the prediction of protein complexes. ClusPro (http://nrc.bu.edu/cluster) represents the first fully automated, web-based program for the computational docking of protein structures.
Characterization of secondary metabolites using Preparative HPLC:
Extraction of secondary metabolites was carried out following the method described earlier (Raaji makers et al., 2001) with modifications. Briefly, overnight culture of Pseudomonas sp. PGPR2 was inoculated with resorcinol and phloroglucinol to confirm the production of bioactive compounds. One liter of the culture supernatant was acidified to pH 2.0 with HCl and then extracted with equal volumes of ethyl acetate. The ethyl acetate extracts were dried in vacuum and dissolved in methanol. Aliquots were filtered (0.2 µm, Millipore Corp. USA) and the antifungal compound was purified by HPLC method (Keel et al., 1992). Briefly, extracts were resuspended in 1 ml of acetonitrile and fractionated using Shimadzu HPLC system with C18 (250 x 10000 mm) reverse-phase column (Phenomenex Luna, Torrance, CA, USA) with 40% Acetonitrile containing 0.1% TFA in water as solvent system with a flow rate of 4.5 ml min⁻¹. HPLC profiles were monitored at 254 nm using PDA. Appropriate solvent was used as negative controls.

**Fig 1:** Identification of Polyketide synthase from Whole genome sequence Pseudomonas sp. PGPR2.
Fig 2: Sequence similarity search using BLAST algorithm

Fig 3: Conserved Domains predicted by InterPro database
Fig 4: Prediction of Cysteine’s disulfide bonding state

```
MPQSRHLALPELSRLLSGAGGLDRPRVWAHV\CELDDPTEWLGEGLLMTGLGIP5DAERQDRYVRRLAEAG
LAGMMIGENMEAPSDDLAIERTAEALGFPILIIHYGVPFAAVTRAIDAGQKDEYERRNALARYESARISMQGIGGL
LPAQLQRLGKDIRSLLYLVPTLAAWQPGLPGLPVEALADALRNRRREGAQPQPVMVLHIPLAEYALSGISIGPSDRN
\CELJALGERLDPHYLHIILAVLGLIERLURVSEHRHLRLGSELLDDLLQRRLLYREGAERLEEELPPGLDLDDAVSVAS
SEGLATDPGLFRGRDSRLVRAGQGEI3LHQAGLSTQFEALDSSLGYVSAPILRRAERCEAEIAREIALAHHSRR
PI\GRYAEILDEAPWLPRLAEAGERFILVGLHDYDAANGSALLHTQVELEENRSVLANAACRGLHIHKQSLVYRI
RRIEISGRRSLDSTADVATLFWFQATRILSPEGIRGRTAKSILQEASAHGGDM
```

Total number of cysteines: 4
Predicted number of bonds: 2
Predicted disulfide bonds (cysteine pairs) ordered by probability in descending order:

<table>
<thead>
<tr>
<th>Bond_Index</th>
<th>Cys1_Position</th>
<th>Cys2_Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>367</td>
<td>388</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>228</td>
</tr>
</tbody>
</table>

Fig 5: HPLC analysis of Pseudomonas PGPR2 extracts with Phloroglucinol and Resorcinol

HPLC analysis of Pseudomonas PGPR2 culture supernatant extracts with Phloroglucinol and Resorcinol

<table>
<thead>
<tr>
<th>Retention time (min)</th>
<th>PGPR2 (mAU)</th>
<th>PHL (control)</th>
<th>Resorcinol (control)</th>
<th>PGPR2-PHL (mAU)</th>
<th>Resorcinol (mAU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.7 min</td>
<td>600</td>
<td>1760</td>
<td>–</td>
<td>600</td>
<td>560</td>
</tr>
<tr>
<td>3.7 min</td>
<td>200</td>
<td>–</td>
<td>1150</td>
<td>270</td>
<td>450</td>
</tr>
<tr>
<td>4.6 min</td>
<td>180</td>
<td>–</td>
<td>–</td>
<td>240</td>
<td>420</td>
</tr>
<tr>
<td>6.5 min</td>
<td>120</td>
<td>–</td>
<td>180</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>10.1 min</td>
<td>1300</td>
<td>8</td>
<td>2960</td>
<td>1760</td>
<td></td>
</tr>
<tr>
<td>13.5 min</td>
<td>480</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

1381
Fig 6: PSIPRED- Protein secondary structure prediction

Fig 7: Typical view of a SWISS-MODEL workspace result. A model for Polyketide synthase of *Pseudomonas* sp. PGPR2 has been generated in automated mode.
Fig 8: Threading modeling Image of the predicted model using Lomets. The confidence score for estimating the model quality is reported as C-score.

- Download Model 1
- C-score = 1.76 (Read more about C-score)
- Estimated TM-score = 0.50 ± 0.15
- Estimated RMSD = 11.6 ± 4.5 Å

Fig 9: *Pseudomonas* PGPR2 – target enzyme complexes of ClusPro. The *Pseudomonas* PGPR2 structure is colored in blue and the enzyme is colored in Red.

*Pseudomonas* PGPR2- RNA Polymerase *Pseudomonas* PGPR2- DNA Topoisomerase

*Pseudomonas* PGPR2- Beta Glucanase

*Pseudomonas* PGPR2- Ergosterol *Pseudomonas* PGPR2- Chitinase
Table 1:- Ranking of the best cluster for the benchmark set of 1000 docked conformations using ClusPro tool

<table>
<thead>
<tr>
<th>S. No</th>
<th>Name of Sequences</th>
<th>Cluster-member size</th>
<th>Representative</th>
<th>Weighted Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>Pseudomonas</em> PGPR2-RNA Polymerase</td>
<td>47</td>
<td>Lowest Energy</td>
<td>-1646.5</td>
</tr>
<tr>
<td>2</td>
<td><em>Pseudomonas</em> PGPR2-DNA Topoisomerase</td>
<td>61</td>
<td>Lowest Energy</td>
<td>-1267.0</td>
</tr>
<tr>
<td>3</td>
<td><em>Pseudomonas</em> PGPR2-Ergosterol</td>
<td>64</td>
<td>Lowest Energy</td>
<td>-1101.4</td>
</tr>
<tr>
<td>4</td>
<td><em>Pseudomonas</em> PGPR2-Beta glucanase</td>
<td>68</td>
<td>Lowest Energy</td>
<td>-1059.7</td>
</tr>
<tr>
<td>5</td>
<td><em>Pseudomonas</em> PGPR2-Chitinase</td>
<td>42</td>
<td>Lowest Energy</td>
<td>-1037.8</td>
</tr>
</tbody>
</table>

Results and Discussion:-

Sequence Analysis on Polyketide synthase:-
The sequenced *Pseudomonas* sp. PGPR2 genome was 6772433 bp long and comprise of 198 contigs. It has been deposited in the Genbank database with accession number ASQ00000000. Earlier we demonstrate that 3, 4-dihydroxy-N-methyl-4-(4-oxochroman-2-yl) butanamide, a secondary metabolite synthesized by *Pseudomonas* sp. PGPR2 showed antagonism against *Macrophomina phaseolina*. Based on the survey of the PGPR2 genome putative biosynthetic pathway was identified. One molecule of acetyl CoA can condense with five molecules of malonyl CoA, which in turn can be obtained by the action of acetyl CoA carboxylase with biotin as the cofactor, to give a hexaketo compound, which can then be cyclised to give different chromone intermediates. The basic chromanone skeleton of 3, 4-dihydroxy-N-methyl-4-(4-oxochroman-2-yl) butanamide could have been generated by a similar reaction (Illakkiamb et al. 2013). A sequence length of 515 amino acid was retrieved from the whole genome sequence of *Pseudomonas* sp. PGPR2 and it showed 100% identity with Polyketide synthase of *Pseudomonas aeruginosa* with accession number WP_0001000104.1 from NCBI databases. Around 99% similarity was observed among Polyketide synthase among *Pseudomonas* sp., and molecular weight is approximately 57110.1 KDa (Fig.1). This suggested a high conservation among Polyketide synthase sequence from bacterial species (Fig.2). Domain distribution and organization of PKS protein was analyzed by InterProScan. The protein sequence contain two functional domain namely Purine catabolism PurC-like domain and PUCR C-terminal helix-turn-helix domain. The HTH domain is situated near the C-terminal end of the protein. The regions from 65-495 showed DNA-binding transcriptional regulator, PucR family domain with the E-value 2.50e-57 and the 432-487 regions showed PucR C-terminal helix-turn-helix domain with the E-value 2.36e-15 in the protein sequence (Fig.3). Neil et al., (2011) stated that disulfide bridges play a major role in the stabilization of the folding process also regulate redox-dependent functions. In Polyketide synthase sequence showed presence of 4 cysteines and two bonds were predicted. The position of first cysteines is 367th and 388 amino acid and second position on 35th and 228th amino acid (Fig.4). To study the structural and functional properties of specific proteins, prediction of disulphide bridges from the sequence is an important tool. In addition, knowledge about the disulfide bonding state of cysteines may help the experimental structure determination process and may be useful in other genomic annotation tasks.

HPLC analysis of *Pseudomonas* PGPR2 extracts with Phloroglucinol and Resorcinol:-
Ethyl acetate extract of *Pseudomonas* sp. PGPR2 culture supernatant was co-inoculated with 250ng of phloroglucinol and resorcinol was analyzed by Shimadzu HPLC system with a C18 reverse-phase column. This ethyl extract of *Pseudomonas* sp. PGPR2 showed six fractions with different retention time. Earlier we reported that the partially purified fraction with 10.1min retention time exhibited antifungal activity was identified as chromanone compound (Illakkiamb et al. 2013). In this, we examined the addition of phloroglucinol and resorcinol to the
**Pseudomonas** sp. PGPR2 culture, showed to enhance the production of chromanone compound about 35% with 10.1 min retention time (Fig. 5).

**Structural Analysis on Polyketide synthase:-**
Evaluating the structure of unknown protein is a key factor for identification of protein function. Koswatta et al. 2011 reported a comparative evaluation of five widely used protein secondary structure prediction programs available in World Wide Web was carried out. Secondary structure prediction tools such as GOR, PSIPRED, HNN, PROF, and YASPIN were used to evaluate the amino acid sequences of the proteins. Among the most frequently used software in the World Wide Web, PSIPRED is the best program for secondary structure prediction.

In present study, secondary structure of a protein was analyzed using PSIPRED and predicts the percentage of alpha helix, beta sheet and coil structure. The purple colour indicates the presence of helix region and magenta colour indicates the sheet region. In this, we set 70 as output width, so it shows 70 amino acids and corresponding predicted structures in each line. The sequence length is also displayed in the output (515 amino acids in this case). The percentage of each structure is also listed as Alpha helix is 56.31%, extended strand as 10.29%, and random coil is 23.50% (Fig. 6). Homology modeling aims to build three-dimensional protein structure models using experimentally determined structures of related family members as templates. SWISS-MODEL workspace is an integrated Web-based modeling expert system (Lorenza et al. 2009). For a given target protein, a library of experimental protein structures is searched to identify suitable templates. On the basis of a sequence alignment between the target protein and the template structure, a three-dimensional model for the target protein is generated. Based on template 3onqB, a model for Polyketide synthase of *Pseudomonas* sp. PGPR2 has been generated in automated mode and the identity was 18.5% with E value 6.2E-31 (Fig. 7). Since the homology between template and target sequence is very low, threading modeling was generated using LOMETS. LOMETS (Local Meta-Threading-Server) is an on-line web service for protein structure prediction. 3D models were generated by collecting high-scoring target-to-template alignments from Eight (8) locally-installed threading programs (FUGUE, HHsearch, MUSTER, PPA, PROSPECT2, SAM-T02, SPARKS, and SP3). For a given target, 160 models are generated by the eight (8) component servers where each server generates 20 models, as sorted by their Z-scores in each algorithm. The best model was selected from the 160 models based on a scoring function (Fig. 8). From the generated model, the PDB file was downloaded for the docking analysis. Simultaneously the target enzymes such as Ergosterol (2A1B), β-(1,3)-D-Glucanase (4M80), DNA gyrase (1BGW), RNA polymerase (3M3Y) and Chitinase (1O6I) were retrieved from PDB database in PDB format. ClusPro represents the first fully automated, web-based program for the computational docking of protein structures. The coordinate files of two protein structures were uploaded through ClusPro’s web interface (Comeau et al. 2004). The docking algorithms evaluate billions of putative complexes, retaining a preset number with favorable surface complementarities. A set of structures were selected by a filtering method, then those structures with good electrostatic values were selected and de-solvation free energies for further clustering. The program output is a short list of putative complexes ranked according to their clustering properties (Fig. 9). Docking results revealed the weighted score for the putative complexes of PGPR2- RNA Polymerase, PGPR2- DNA Topoisomerase, PGPR2- Ergosterol, PGPR2- Beta glucanase and PGPR2- Chitinase (Table 1). *Pseudomonas* sp. PGPR2 template protein Polyketide synthase might be target the fungus by inhibiting the DNA synthesis. This was proved by weighted score of PGPR2- RNA Polymerase complex which has low score compared to other complexes and this might be due to the presence of conserved domains responsible for transcription in *Pseudomonas* sp. PGPR2.

**Summary:-**
In silico sequence and structural analysis of Polyketide synthase protein showed structurally related regions in a compilation of amino acid sequences. Structural analysis showed that the protein’s secondary structures were mainly composed of alpha helices and random coils. Presence of conserved cysteine residues plays a major role in the stabilization of the folding process of protein. The docking and discrimination method also provides good results for a number of complexes that were used as targets in the Critical Assessment of Predictions of Interactions experiment. Hence in future, in silico analyses of bacterial genomes associated with fungal viability and virulence has led to the identification of many putative targets for novel antifungal agents.
References: