## 7. Summary

**Glycopeptide antibiotics** (GPAs) are glycosylated polycyclic non-ribosomal peptides which have been the drugs of choice for the treatment of infections caused by Gram-positive pathogenic bacteria, including penicillin, methicillin and oxacillin-resistant strains of *Staphylococcus* and *Streptococcus* for over three decades. These include: vancomycin, balhimycin, teicoplanin, avoparcin, nogabecin, telavancin, bleomycin and decaplanin etc.

The GPAs are remarkable in their chemical structures, consist of highly cross-linked and diversely modified core heptapeptides. The biosynthesis of core heptapeptide occurs via non ribosomal peptide synthestase (NRPSs) and modifications on this core heptapeptide backbone are added by a set of tailoring enzymes which include: glycosyltransferases (Gtfs), monooxygenases, acyltransferases (Atfs), halogenases, sulfotransferases (Stfs) and methyltransferases (Mtfs). These tailoring enzymes modify specific residues on the heptapeptide backbones which can alter chemical and biological properties of GPAs including solubility, dimerization, antimicrobial activity and stability etc.

However, the emergence and spread of vancomycin-resistant strains has raise the need for discovery of the next-generation durable antibiotics capable of addressing such bacterial infections. It is important to look for alternatives to vancomycin such as other glycopeptides or their analogs for the treatment of serious Gram-positive infections. New potent GPAs can be produced either by chemical modification or by synthesizing new hybrid GPAs. GPAs have crosslinked heptapeptide back bone which makes it very difficult to modify via chemical methods. However, the genes involved in biosynthesis of GPAs can be used to produce hybrid GPAs *in vitro*. Thus, in order to manipulate the biosynthesis of GPAs and to create new antibiotics by rational design, it is essential to identify the biosynthetic gene clusters (BSGCs) for GPAs biosynthesis.

**Decaplanin** is a GPA that has been reported 25 years ago, but still its biosynthetic gene cluster has not been identified yet. The heptapeptide backbone of decaplanin is similar to that of vancomycin but differs only in having one additional sugar moiety (rhamnose), attached to heptapeptide backbone and has one lesser chlorine atom as compared to vancomycin. Because of these minor structural changes activity pattern and stability of decaplanin varies from vancomycin. Identification and characterization of genes responsible for adding this structural diversity to natural compounds will help us to understand the remarkable complex

chemistry of their products and to play genetic engineering with them so as to derive a large variety of new compounds.

In the present study, the production of decaplanin was confirmed from *A. decaplanina* DSM 44594$^T$. The protocol for extraction and identification of GPAs has been standardized. As commercially available standard of decaplanin was not available, therefore, all experiments were validated using vancomycin as reference (Sigma). As methanol is miscible with culture broth, therefore, resin Diaion HP 20 was used and followed by extraction with different gradients of methanol. TLC, HPLC and LCMS results have shown that most of the impurities are washed off with 25 and 50% methanol and the compound extracted with 75% methanol is semi purified and could be directly used for further analysis. GPAs in LCMS forms doubly charged species and deconvoluted mass corresponds to the actual molecular weight of the GPAs. In positive and negative ionization modes of LCMS, doubly charged species [$(m/z)^{2+}$] of 780.7 and 778.7 were obtained and deconvulated mass from both of these is 1559.5 which corresponds to exact molecular weight of decaplanin as reported in the literature. Decaplanin has been isolated and identified from *A. decaplanina* DSM 44594$^T$.

For identifying decaplanin biosynthetic gene cluster which is beyond the range of classical cloning, whole genome sequencing of *A. decaplanina* DSM 44594$^T$ was performed. A bioinformatics tool, antiSMASH (Antibiotics Secondary Metabolites Analysis Shell) that can rapidly identify and annotate all types of known SM biosynthetic gene clusters was used (Weber et al., 2015). The characterization of decaplanin biosynthetic gene cluster identified by antiSMASH and bioinformatics analysis was further done by gene inactivation studies.

Based on antiSMASH results and similarity search, 66.3 kb genomic region 44594$^T$ starting from 2,76,857 bp to 3,43,203 bp in contig number 74 of *A. decaplanina* DSM 44594$^T$ is identified as decaplanin encoding gene cluster. Like vancomycin class of antibiotics gene cluster, decaplanin gene cluster consists of three genes encoding for NRPSs involved in heptapeptide backbone biosynthesis, three cytochrome P450 monoxygenases coding genes, required for cross-linking of amino acids, three genes for glycosyltransferases, genes for amino acid precursor and sugar biosynthesis, transport, resistance and regulation etc. From these findings, a **NRPS gene cluster** is identified as decaplanin coding gene cluster in the genome of *A. decaplanina* DSM 44594$^T$ and **which is designated as "*dec*".**

Decaplanin consists of three carbohydrate moieties: a disaccharide unit (L-rhamnose, D-glucose) attached to aa-4 and amino sugar, L-4-epivancosamine present at aa-6 whereas,

vancomycin consists of a disaccharide unit at aa-4 and lacks third sugar moiety. Accordingly, *dec* BSGC encodes three glycosyltransferase genes presumably responsible for addition of sugar moieties to a glycon backbone. In contrast with *vcm* cluster, the number and arrangement of ORFs coding for glycosyl transferases (gtfs) also differ in *dec* BSGC. An extra glycosyltransferase, gtfA presumed to add epi-vancosamine at βHt (aa6) is present in *dec* BSGC. As vancomycin consists of chlorine at aa-6 (βHt) and lacks carbohydrate moiety at this position, hence, this gene is missing in *vcm* but present *dec*. The module organization of NRPSs and other tailoring enzymes are identical to *vcm* cluster, except for the glycosyltransferase ORFs which differ both in arrangement and numbers as well. The genes adding structural variations to decaplanin as compared to vancomycin have been identified. These genetic level differences explain the structural diversification of decaplanin from vancomycin.

To validate the function of identified *dec* gene cluster in the genome of *A. decaplanina* DSM 44594$^T$ for the biosynthesis of decaplanin, the putative *decC* (*nrpsC*) gene was inactivated using homologous recombination method. A modified transformation protocol is developed for *A. decaplanina* DSM 44594$^T$. The cells harvested at stationary phase gave efficient transformation efficiency and also recombinants were obtained. In the expected positive clones, the production of decaplanin was abolished, hence did not produce zone of inhibitions against Gram positive test strains as compared to wild type *A. decaplanina* DSM 44594$^T$. The sequencing of clones confirmed single cross-over leading to gene disruption but unable to induce second crossover. In conclusion, gene inactivation of *decC* (*nrpsC*) presented in this work demonstrate the role of *dec* BSGC in the biosynthesis of decaplanin in *A. decaplanina* DSM 44594$^T$.

The members of the genus *Amycolatopsis* are well known for the production of GPAs. At the time of start of this project, there were eight species of the genus *Amycolatopsis* which were known for the production of GPAs and are grouped together in cluster A except for *A. balhimycina* DSM 44591$^T$. Hence, we have selected cluster A of this genus for studying GPA BSGCs. GPAs have conserved heptapeptide backbone and differ in glycosylation, methylation, sulfonation or chlorination patterns that lead to different levels of activity against the same organisms. Deciphering the identity of these genes responsible for providing structural diversity to these compounds could lead to synthesis of large variety of compounds through bio-combinatorial synthesis. Thus, in view of emerging importance of identification

of BSGCs, we aimed to identify GPA BSGCs from strains of cluster A (eleven strains) of the genus *Amycolatopsis*.

The genome of eight strains were sequenced, annotated, analyzed and evaluated for secondary metabolite coding biosynthetic potential: *A. coloradensis* MTCC 4020$^T$, *A. keratiniphila* subsp. *nogabecina* DSM 44586$^T$, *A. azurea* MTCC 11287$^T$, *A. alba* DSM 44262$^T$, *A. regifaucium* DSM 45072$^T$, *A. keratiniphila* subsp. *keratiniphila* DSM 44409$^T$, *A. thailandensis* JCM 16380$^T$ and *A. vancoresmycina* DSM 55592$^T$. For the remaining four strains: *A. orientalis* DSM 44040$^T$, *A. lurida* DSM 43134$^T$, *A. japonica* DSM 44213$^T$ and *A. balhimycina* DSM 44591$^T$, the genomes were available in the public domain, therefore downloaded from NCBI and analyzed. All the genomes possess much higher biosynthetic potential than reported so far. On an average ~ 20% of their genomes encode for genes responsible for synthesis of secondary metabolites.

In our study, we have identified and analyzed the GPA BSGCs from four species of cluster A of *Amycolatopsis* genus, *A. coloradensis* MTCC 4020$^T$, *A. keratiniphila* subsp. *nogabecina* DSM 44586$^T$, *A. azurea* MTCC 11287$^T$ and *A. alba* DSM 44262$^T$ which are already known to produce GPA (avoparcin, nogabecin, azureomycins and a unknown GPA) but their clusters are not identified, yet.

*Amycolatopsis coloradensis* MTCC 4020$^T$ and *A. keratiniphila* subsp. *nogabecina* DSM 44586$^T$ are known for the production of avoparcin and nogabecin which are type II glycopeptide antibiotics (GPAs) that consist of aromatic amino acid residues at position 1 and 3: [(hpg-βht-phe) + (hpg-hpg-βht) + dhpg], than the aliphatic one. The genome sequence of Type I (vancomycin and balhimycin), type III (ristocetin and ristomycin) and type IV (teicoplanin) producers and their biosynthetic gene clusters are reported but for type II GPAs (avoparcin, actinoidin and nogabecin), neither the genome sequence is published nor the gene cluster is identified so far. This is the first report of genome sequence and genome mining of type II GPA producing strains and also the GPA gene clusters encoded by them (avoparcin and nogabecin) are identified.

From the *insilco* analysis, the GPA produced by *A. alba* was predicted to be vancomycin. To validate *insilico* data, the compound was extracted and analysed using HPLC and LCMS which supported the *insilico* findings and the GPA produced by *A. alba* is identified as vancomycin.

Interestingly, we found that GPA cluster data of *A. azurea* MTCC 11287$^T$ was not supporting the previous literature which states the production of azureomycins of molecular weights 774 and 812 Da from this organism. From the bioinformatics analysis, a 74 kb genomic region was predicted to encode GPA biosynthesizing genes in the genome of *A. azurea*. It comprises of 42 ORFs which includes four non ribosomal peptide synthetases, one halogenase, five glycosyltransferases, one mannosyltransferase and four cytochrome oxygenases involved in cross linking of the backbone and is the largest GPA BSGC reported till date. This analysis suggested that the GPA produced by *A. azurea* belongs to type III GPA and is predicted as chlorinated heptapeptide with six carbohydrate moieties [(Hpg-(βHt-Dpg) + (Hpg-Hpg-βHt) + Dpg)] which could be of approx.2000 Da.

Three species of cluster A of the genus *Amycolatopsis*, *A. keratiniphila* subsp. *keratiniphila* DSM 44409$^T$, *A. thailandensis* JCM 16380$^T$ and *A. regifaucium* DSM 45072$^T$ were not known for GPA production. The genome mining results of these strains revealed their GPA biosynthesis potential which was then confirmed by wet lab experiments (antibiosis, TLC, HPLC and LCMS). From the genomic data GPA encoded by these strains were predicted as decaplanin, nogabecin and A35512B which were confirmed by HPLC and LCMS data. Thus, three new producers of already known GPAs are reported for the first time: *A. regifaucium* (decaplanin), *A. keratiniphila* subsp. *keratiniphila* (nogabecin) and *A. thailandensis* (A35512B).

We also checked the sequences (genomes available in public domain) of other strains from the genus *Amycolatopsis* for GPA BSGCs, but except *A. balhimycina*, GPA BSGCs are not found in any other strains of cluster B, C, D and E of the genus *Amycolatopsis*.

Parallel to this, a detailed comparative analysis of all these GPA BSGCs (eight identified in this study and four previously reported) was performed. The arrangement of NRPSs is conserved according to the type of GPA (type I, type II, type III) produced from these organisms, but as the different GPAs of same type differ in post assembly modification of their backbones, therefore, these all GPA BSGCs possess different numbers of tailoring enzymes such as gtfs, mtfs, halogenases etc that too vary in their arrangement in the respective gene cluster. Thus, comparative genomic analysis of GPA BGSCs help us to understand the role of genetic level differences in the diversity of biosynthesis and differences in the activity patterns of these antibiotics.

The emergence of drug resistance along with a declining pipeline of clinically useful antibiotics has made it vital to develop more effective antimicrobial therapeutics, particularly high throughput screening methods. Hence, in view of such importance of genome mining for discovery of secondary metabolites, comparative genome studies of cluster A of *Amycolatopsis* are performed with focus on the assessment of secondary metabolite biosynthetic potential. We have taken eleven strains of cluster A of *Amycolatopsis* and two strains from cluster C for comparative analysis.

The prediction of secondary metabolite production potential of these selected strains revealed the presence of many gene clusters although the later are reported to produce one or two secondary metabolites only. This study also revealed the presence of type II Polyketide Synthase (type II PKS or PKS II) gene clusters in these strains of *Amycolatopsis*, which encode for different types of Type II polyketide antibiotics such as tetracycline, anthracycline, angucycline etc which can lead to discovery of very potent antibiotics in near future. Apart from PKS II clusters, other conserved gene clusters (terpenes, siderophores and ectoine) from this cluster of *Amycolatopsis* also have been identified.

For the first time, we identified type II PKS gene clusters from these strains which are earlier known for GPA production only. These strains have huge genetic potential to produce different types of aromatic polyketides which can lead to discovery of very potent antibiotics in near future i.e. antitumour, anticancer and antimicrobial. Till now there is no report for screening of these strains with focus on aromatic polyketide, ectoine and terpene production.

Interestingly, *A. balhimycina* from cluster C share large number of SM gene clusters with strains of cluster A although it belongs to separate cluster phylogeneticaly. This analysis revealed that GPA, type II PKS, some terpene, ectoine and sidherophore clusters are conserved among these strains but differ in organization and numbers of tailoring enzymes which can lead to biosynthesis of structurally different compounds of same class from strains of this cluster of *Amycolatopsis*.

In summary, the genome analyses of these strains of the genus *Amycolatopsis* have provided new insights into the clustering of secondary metabolite gene clusters at the species or cluster level of this genus. Many gene clusters such as GPA, type II PKS, 2-MIB, albachelin are conserved among strains of cluster A of this genus with structural diversity of these compounds. These strains have huge genetic potential to produce different types of aromatic

polyketides i.e. antitumour, anticancer and antimicrobial. Till now there is no report for screening of these strains with focus on aromatic polyketide, ectoine and terpene production.

Present work has brought attraction towards this fascinating genus *Amycolatopsis* as an inexhaustible source of novel biologically active secondary metabolites. However, the presence of these genes does not necessarily mean that the strains will produce the antibiotic. The genes may not be expressed at all in the strain (silent genes) or may only be expressed under specific conditions (e.g. under particular environmental conditions, such as the type of media used for the antibacterial testing). The expression of these gene clusters under different conditions could lead to discovery of new interesting biological activities as well as new chemical scaffolds thus proving the concept of genome mining for the discovery of new secondary metabolites.

## Highlights of the study

- Purification and identification of decaplanin from *A. decaplanina* DSM 44594$^T$ has been done.
- The characterization of ***dec* BSGC** has been done by gene inactivation experiments.
- Analysis of cluster A of the genus *Amycolatopsis* for GPA BSGCs and production. In support of bioinformatics data, the identification of GPAs produced from these clusters was done by using HPLC and LCMS.
- Type II PKS gene clusters: First time reported and identified in the cluster A of *Amycolatopsis*.
- Three new producers of already known GPAs are reported.
- GPA BSGC encoded by *A. azurea* MTCC 11287$^T$ is the largest among the known one till date.
- Identification of kigamicin gene cluster in *A. regifaucium* DSM 45072$^T$.
- 2-MIB, albachelin and ectoine coding gene clusters have been identified.
- Comparative studies with respect to genetic potential for secondary metabolite genes of thirteen strains of the genus *Amycolatopsis* have been performed which revealed their high genetic capacity for the biosynthesis of secondary metabolites.