BIOSYNTHESIS, RESISTANCE AND REGULATION OF THE GLYCOPEPTIDE ANTIBIOTIC A47934

BIOSYNTHESIS, RESISTANCE AND RESISTANCE REGULATION OF THE GLYCOPEPTIDE ANTIBIOTIC A47934 IN *Streptomyces toyocaensis* NRRL 15009

BY

JEFFREY CHARLES POOTOOLAL, B. Sc.

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For The Degree

Master of Science

McMaster University

MASTER OF SCIENCE (2002)
(Biochemistry)

McMaster University
Hamilton, Ontario

TITLE: Biosynthesis, Resistance and Resistance Regulation of the Glycopeptide Antibiotic A47934 in *Streptomyces toyocaensis* NRRL 15009

AUTHOR: Jeffrey Charles Pootoolal, B. Sc. (McMaster University)

SUPERVISOR: Professor G.D. Wright

NUMBER OF PAGES:

Abstract

Multiple antibiotic resistant bacteria continue to be a threat to the health of the world's population. Glycopeptide antibiotics are one type of drug that are used to treat these serious pathogens. Increased usage over the years has led to the emergence of bacteria which are resistant to these glycopeptide antibiotics and now the need for altered antibiotics with an increased effectiveness has arisen. *Streptomyces toyocaensis* NRRL 15009 produces the glycopeptide antibiotic A47934. Here, the biosynthetic gene cluster for A47934 was sequenced in its entirety. All enzymes encoded by assigned open reading frames were analyzed and functions assigned where possible. The resulting biosynthesis cluster encodes all the enzymes necessary to produce A47934, as well as confer resistance and regulate the resistance response. In addition to sequencing the biosynthetic gene cluster, enzymatic studies were attempted on the two-component regulatory system (VanR and VanS) which confers resistance to A47934. Finally, inactivation of *staL* was attempted. Overall, the results presented here should help us to further understand how these chemically complex glycopeptide antibiotics are made and lend further insight into how we can attempt to produce new semi-synthetic versions.

Acknowledgements

I would first like to thank my family for all they have done for me throughout my entire education. All the encouragement, money and rides to and from Hamilton were greatly appreciated.

I would like to thank Dr. Gerry Wright for all the help, guidance and tolerance during my stay in grad school. I would like to thank everyone in the Wright lab for all the help and entertainment they provided. A special shout out goes to the Dawg Pound and Team Streptomyces. Special thanks to Dr. Gary Marshall and Dr. John Neu for all the specific help they gave me on this project.

I would like to thank the members of my committee, Dr. Justin Nodwell and Dr. Astrid Petrich for the suggestions they gave to me along the way. Finally I would like to thank Dr. Christopher Walsh, Dr. Michael Thomas and Dr. Brian Hubbard for the help they gave me in sequencing and annotating the A7934 biosynthesis cluster.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A domain	adenylation domain
ACP	acyl carrier protein
Ala	Alanine
Am	apramycin
aminoacyl-SNACs	aminoacyl-*N*-acetylcysteamine thioesters
AMP	adenosine monophosphate
Asn	asparagines
ATP	adenosine triphosphate
bp	base pairs
C domain	condensation domain
DF/HCC	Dana Farber/Harvard Cancer Center
DHPG	3,5-dihydroxyphenylglycine
E domain	epimerization domain
GPA	glycopeptide antibiotic
H-bonds	hydrogen bonds
His	histidine
HPG	*p*-hydroxyphenylglycine
kb	kilobase pairs
Lac	lactate
LB	Luria Broth
Leu	leucine
Lys	lysine
M domain	methylation domain
MBP	maltose binding protein
NRPS	non-ribosomal peptide synthetases
OH-Tyr	β-hydroxytyrosine
ORF	open reading frame
P-pant	phosphopantetheine
PAPS	3'-phosphoadenosine 5'-phosphosulfate
PBS	phosphate buffered saline
PVDF	hydrophobic polyvinylidene difluoride
SAM	*S*-adenosyl methionine
SAS	saturated ammonium sulfate
SVM	soygrit vegetative media
T domain	thiolation domain
Te domain	thioesterase domain
TEMED	N, N, N', N'-tetramethylethylenediamine
Thio	thiostrepton
Tyr	tyrosine
VISA	vancomycin intermediate *Staphylococci aureus*
VRE	vancomycin resistant *Enterococci*

1. Introduction:

In the continuing battle against antibiotic resistant bacteria, glycopeptide antibiotics (GPAs) such as vancomycin and teicoplanin are often used as the last line of defense against Gram-positive infections. Produced by actinomycetes as secondary metabolites, this large class of antibiotics has proved to be of great clinical relevance and the subject of much study. Unfortunately, their extensive clinical and agricultural use over the last two decades has selected for GPA resistance and consequently the need for alternative antibiotics. Due to their chemically complex structures, shown in Figure 1, altering the producing organisms' genetic machinery may prove to be the only feasible option to substantially modify this class of drugs.

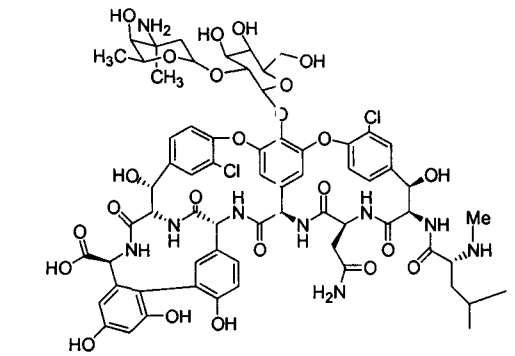
1.1 History of Glycopeptide Antibiotics

The first GPA to be discovered was vancomycin. Even though it had been discovered in 1956, vancomycin (and later teicoplanin) did not see much use until the 1980's due to toxicity issues as a result of impure preparations (1). Vancomycin and teicoplanin are still the only two clinically relevant GPAs and are used to treat infections caused by multiple antibiotic resistant Gram positive bacteria such as *Enterococci*, *Staphylococci*, *Streptococci* and others. As multiple antibiotic resistant bacteria became more prevalent during the 1980's, so to did the use of these GPAs. The increase in usage of vancomycin (over 500% between 1984 and 1994 in the United States alone) inevitably led to resistance which was first seen in *Enterococci* (2, 3). VISA (vancomycin

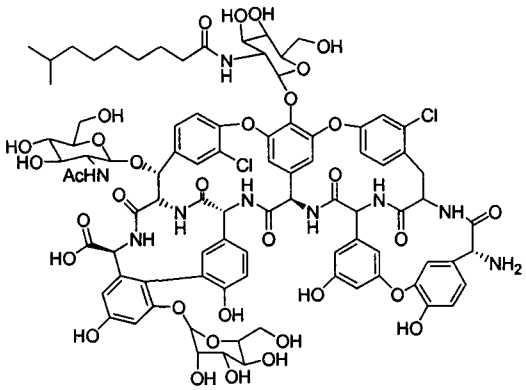
intermediate *Staphylococci aureus*) and most recently vancomycin resistant *Streptococci* have since followed the emergence of vancomycin resistant *Enterococci* (VRE) in the 1980's (4, 5). This spread of both resistance and tolerance to vancomycin increases the need for new GPAs with the ability to evade resistance mechanisms.

1.2 Structures of Glycopeptide Antibiotic

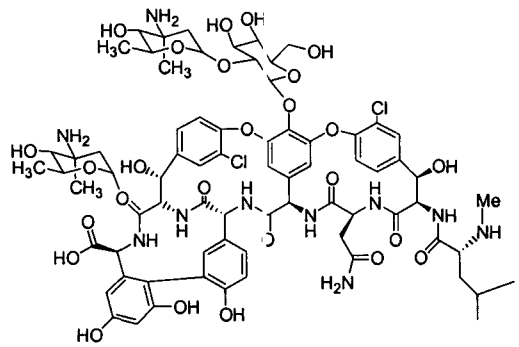
GPAs are a large class of antibiotics which have several common characteristics. GPAs are all composed of a heptapeptide backbone but can be subdivided into two subgroups, vancomycin-like and teicoplanin-like GPAs, based on amino acid sequence. Not only are vancomycin and teicoplanin the only two clinically used GPAs, they also represent the only two variants on the GPA backbone. The backbone itself is curious in that it is composed both of proteogenic amino acids such as asparagine (Asn), leucine (Leu) and tyrosine (Tyr) as well as the unusual amino acids 3,5-dihydroxyphenylglycine (DHPG), β-hydroxytyrosine (OH-Tyr) and *p*-hydroxyphenylglycine (HPG). Vancomycin is composed of NH_2 – Leu – OH-Tyr – Asn – DHPG – HPG – OH-Tyr – DHPG – COOH with the amino acids corresponding to numbers 1 – 7. Standard chemical depictions of GPAs, such as in Figure 1, are drawn from carboxy to amino terminus, thus in descending amino acid number (7 - 1). Teicoplanin is similar, but differs in the three N-terminal amino acids, NH_2 – HPG – Tyr – DHPG – COOH instead of NH_2 – Leu – OH-Tyr – Asn. The retrobiosynthesis of the teicoplanin-like GPA A47934 is shown in Figure 2 and illustrates the amino acid composition in the linear peptide form.



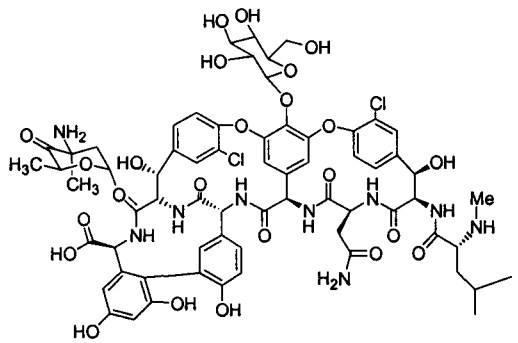
a) Vancomycin



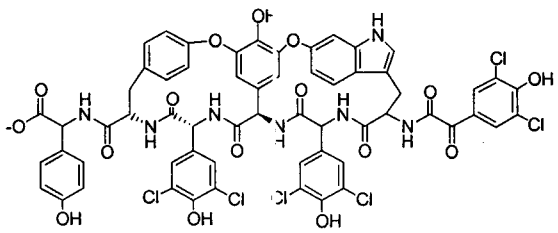
b) Teicoplanin



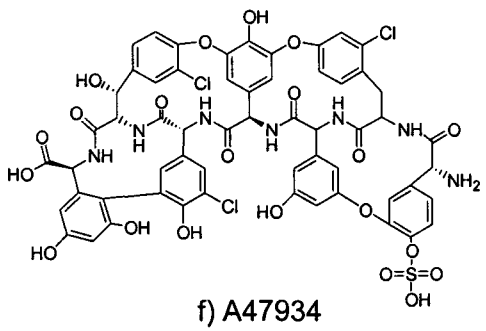
c) Chloroeremomycin



d) Balhimycin



e) Complestatin



f) A47934

Figure 1 – Structures of Representative Glycopeptide Antibiotics: The two clinically relevant GPAs are a) vancomycin and b) teicoplanin. The GPAs for which the biosynthetic gene clusters have been sequenced are the vancomycin-like c) chloroeremomycin and d) balhimycin along with the GPA-like secondary metabolite e) complestatin. The GPA studied in this thesis is f) A47934.

The extensive intramolecular crosslinks give the GPAs a rigid, cup-like structure. The heptapeptide is intramolecularly crosslinked through a series of aromatic ether and carbon-carbon bonds, as shown in Figure 2. For the teicoplanin class backbone, there are four sites of crosslinking: an ether linkage from HPG_1 (subscript referring to amino acid number) to $DHPG_3$, an ether linkage from Tyr_2 to HPG_4, a third ether linkage from HPG_4 to OH-Tyr_6 and a carbon-carbon bond between HPG_5 and $DHPG_7$. Vancomycin-like GPAs have the same crosslinks, but lack the ether linkage from amino acids 1 to 3 as the core amino acids at those positions do not include aromatic amino acids.

Along with crosslinks, there are numerous modifications that can be made to the constituent amino acids. Figure 3 outlines some of the possible modifications that the core heptapeptides (in this case the teicoplanin-like scaffold) undergo in various natural producers and does not even include semi-synthetic modifications. As can be seen there are a wide variety of antibiotics that are produced on these scaffolds. Modifications such as chlorination, sulfonylation, methylation, glycosylation and fatty acid acylation are all utilized to give a wide variety of natural antibiotics. Within these modifications occur subsets that can create an even greater diversity such as varying the pattern and site of modification or the specific type of modification (i.e. different sugar types).

In addition to these antibiotics, there are numerous examples of molecules that are structurally similar to GPAs. Peptide antibiotics with heptapeptide backbones that differ from the GPA class of backbones include complestatin (Figure 1) and kistamicine (Figure 4). Other molecules, such as novobiocin and coumermycin A_1, contain both amino acid

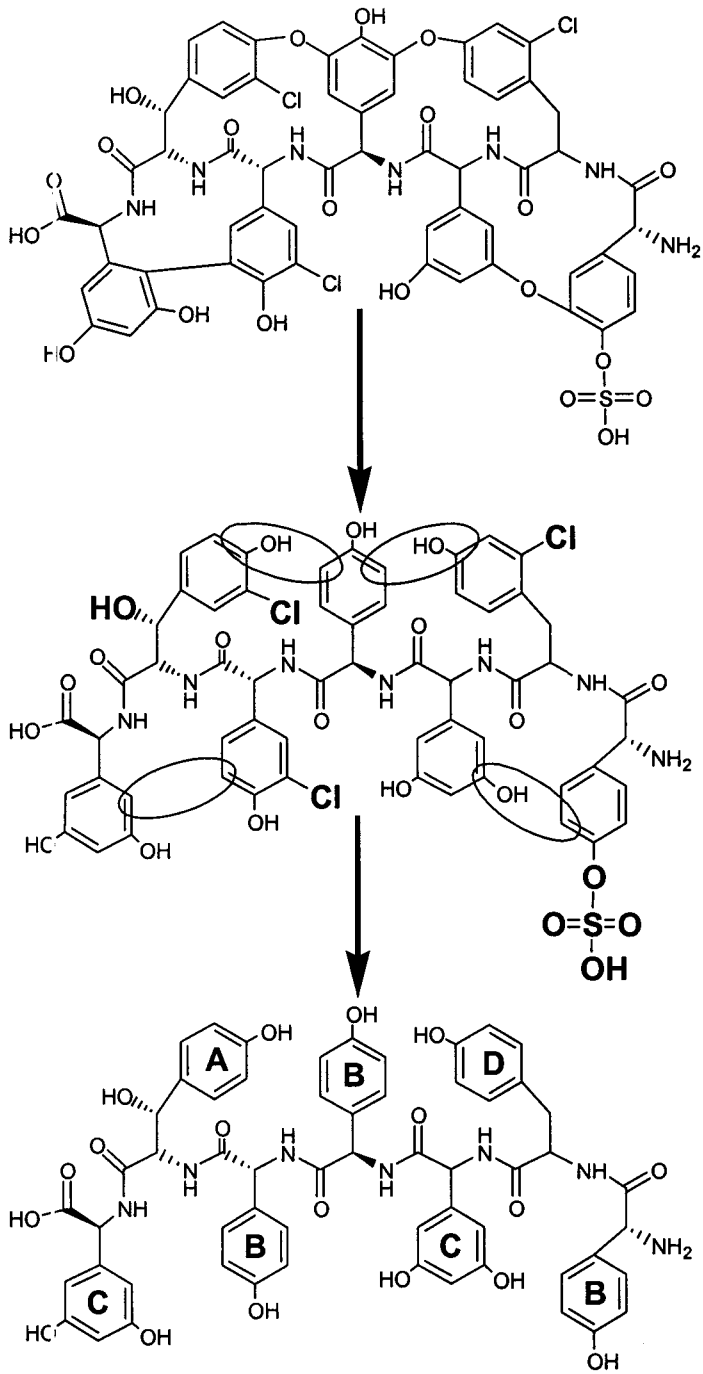


Figure 2 – Retrobiosynthesis of A47934: Retrobiosynthesis of a GPA (A47934 in this case) shows where the crosslinkings occur relative to the linear peptide. In bold are the sites of modification such as hydroxylation, chlorination and sulfonylation. Also shown are the individual amino acid components which compose the heptapeptide. Shown are the unusual amino acids A - β-hydroxytyrosine, B – *p*-hydroxyphenylglycine and C – 3,5-dihydroxyphenylglycine along with D – tyrosine.

and sugar residues, but are structurally quite different (Figure 4). Both types are of relevance to the study of GPAs as biosynthetic strategies utilized by the producing organisms show significant similarities.

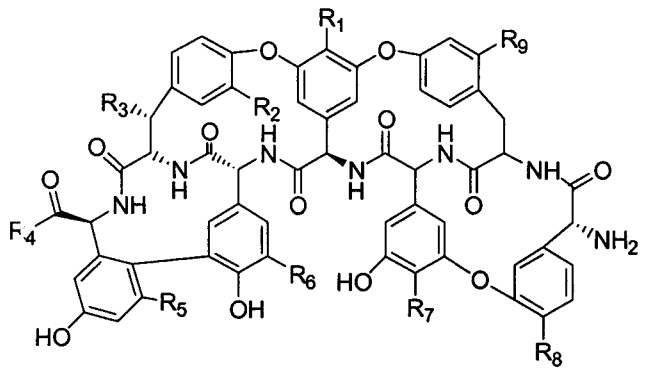


Figure 3 – Known Modifications On Teicoplanin-like Glycopeptide Antibiotics: Outlined above the possible modifications that have been found on teicoplanin-like GPAs. These modifications are the method by which nature generates structural diversity from a common scaffold.

1.3 Mechanism of Action of Glycopeptide Antibiotics

Crystallographic studies of GPAs have demonstrated that they adopt a cup-like structure (6, 7, 8, 9). This is the result of the crosslinks between the amino acid side chains of the heptapeptide backbone. This allows GPAs to stably bind to the growing cell wall in Gram-positive bacteria. Most eubacteria have a cell wall made of peptidoglycan that surrounds the cell membrane. Peptidoglycan consists of repeating units of *N*-acetylglucosamine and *N*-acetylmuramic acid, which are bound to the cell membrane via a lipid group. The repeating sugar units have an extending pentapeptide

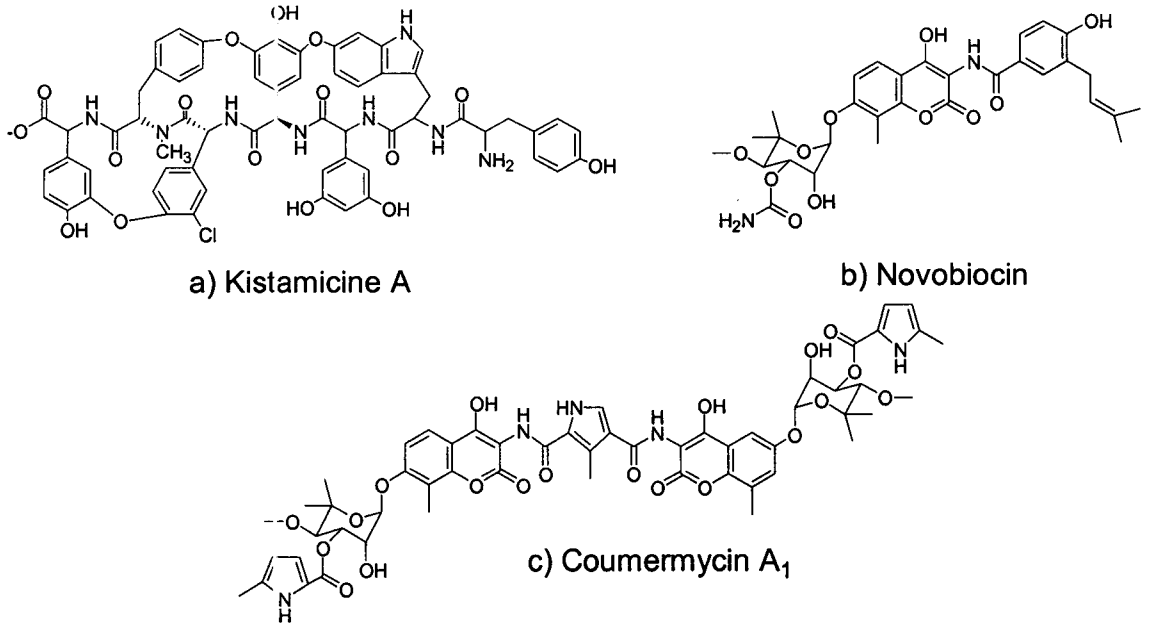


Figure 4 – Structures of GPA-like Molecules: Shown above are examples of natural products related to GPAs by chemical structure or biosynthetic strategy: a) kistamicine A, b) novobiocin and c) coumermycin A_1.

linker. The linkers, which extend from the *N*-acetylmuramic acid, consist of L-alanyl-γ-D-glutaryl-L-lysyl-D-alanyl-D-alanine (although this peptide sequence can vary). The peptide can be joined via a pentaglycine bridge from the terminal D-alanine (D-Ala) of one linker to the L-lysine (L-Lys) residue of another, which rigidifies this polymer, giving structural integrity to the cell (10).

GPAs bind tightly to the D-Ala-D-Ala residues in the peptidoglycan layer through a network of five hydrogen bonds (H-bonds), as seen in Figure 5 (10, 11). This interaction has been shown by NMR studies of vancomycin bound to diacetyl-L-Lys-D-Ala-D-Ala (11). The binding of vancomycin to the terminal D-Ala-D-Ala has the double effect of blocking both the extension of the *N*-acetylglucosamine and *N*-acetylmuramic acid units (transglycosylation) and the crosslinking of the peptidoglycan layers

(transpeptidation) (Figure 6). The net effect of inhibiting the two steps of peptidoglycan formation is a reduction in cell wall structural strength and ultimately cell lysis under slight changes in osmotic pressure.

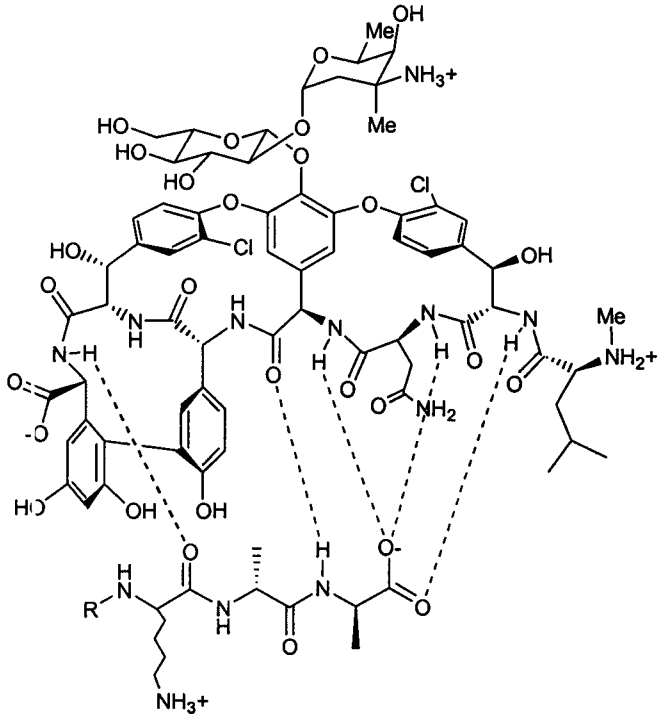


Figure 5 – Hydrogen Bonding of Vancomycin to Peptidoglycan in Gram-positive Bacteria: The GPA vancomycin's mechanism of action is binding to the growing cell wall in Gram-positive bacteria. Vancomycin's extensive crosslinking gives the drug a rigid, cup-like shape that allows it to stably bind to the D-Ala-D-Ala of the peptidoglycan layer via a network of five hydrogen bonds.

A second characteristic of the cup-like structure is the ability of GPAs to form back to back dimers via four hydrogen bonds, seen in Figure 7 (12, 13). Dimerization is thought to increase the potency of the GPAs as once a single unit is bound, the sequestered second unit is now placed closer to the desired site of activity and thus has an entropic binding advantage. A second theory holds that dimerized GPAs, once bound to

the peptidoglycan layer, create a greater steric hindrance of transglycosylation than a bound monomer unit (12).

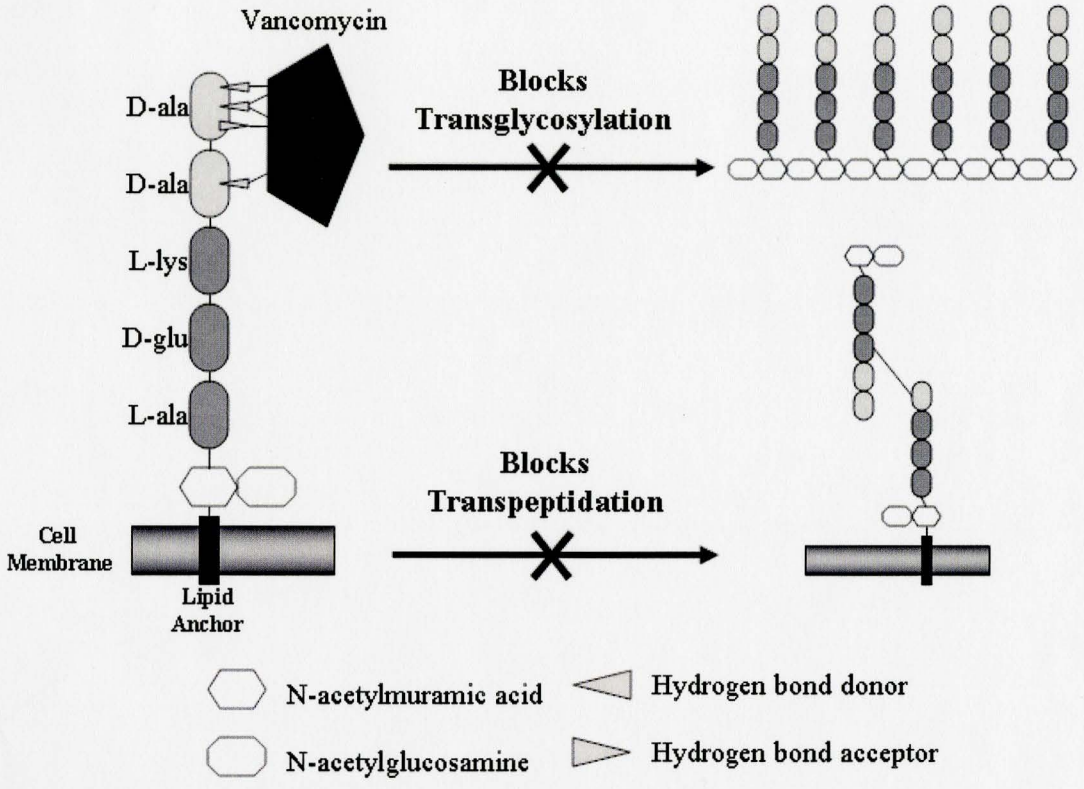


Figure 6 – Effects of Vancomycin on Gram-positive Bacteria: The bactericidal effects of vancomycin are to block peptidoglycan transglycosylation and transpeptidation. Blocking of transglycosylation inhibits the addition of the *N*-acetylglucosamine and *N*-acetylmuramic acid repeating units, while blocking transpeptidation inhibits crosslinking of the peptidoglycan layers. The overall result is a lowering of cell wall strength making the bacteria subject to lysis under minimal osmotic stress.

1.4 Resistance to Glycopeptide Antibiotics

A set of resistance genes, termed the van genes, has been shown to be both necessary and sufficient for GPA resistance in VRE. The overall effect of these five van genes (*vanR*, *vanS*, *vanH*, *vanA* and *vanX*) is to convert the D-Ala-D-Ala residues in the

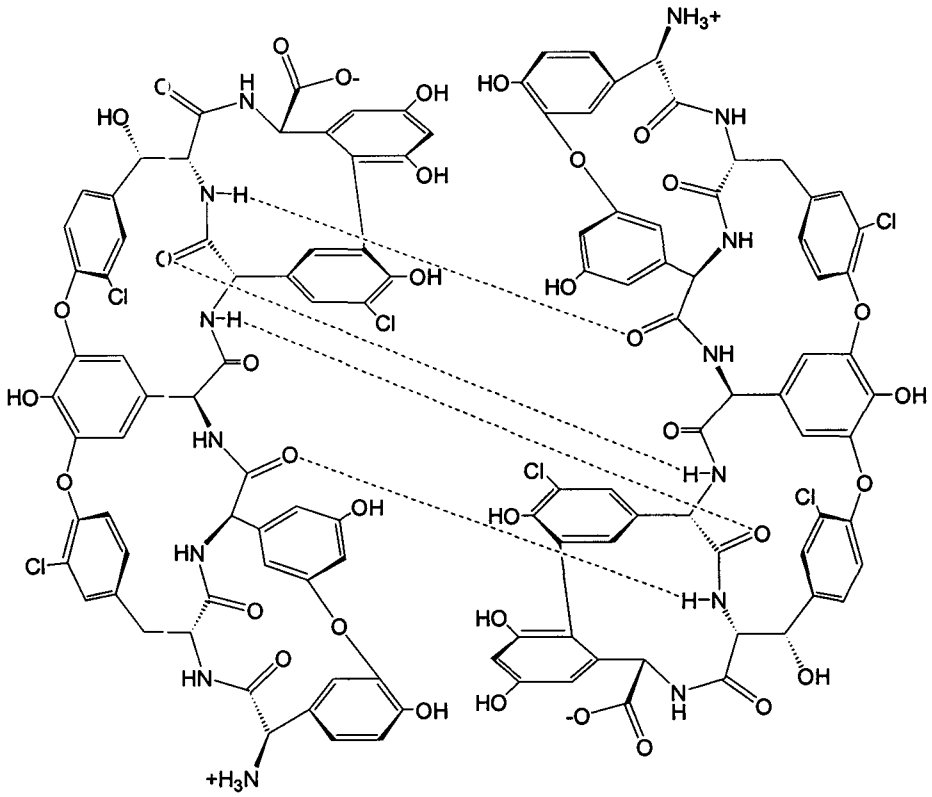


Figure 7 – Dimerization of Glycopeptide Antibiotics: The cup-like structure of GPAs result in an exposed peptide backbone. The backbone can form both homo and hetro-dimers using four hydrogen bonds resulting in an increased efficiency, which is illustrated here for the teicoplanin aglycone.

pentapeptide linker of peptidoglycan into D-Ala-D-lactate (D-Lac). The van gene cluster was first found as a set of plasmid borne genes conferring resistance in *Enterococci* (2). Eventually this set of genes was found to be present in other resistant bacteria and in glycopeptide producing organisms such as *Streptomyces toyocaensis* (14). Along with being located on autonomously replicating plasmids, the *van* gene cluster has since been shown to reside on transposable elements that have inserted into the bacteria's genome (15). In the case of GPA producing bacteria, the resistance genes are found within the genome.

The relatively conservative change of D-Ala to D-Lac results in the conversion of an amide bond to an ester bond and the net loss of one hydrogen bond donor group (Figure 8). Vancomycin is still able to bind D-Ala-D-Lac, but with only four H-bonds and the dissociation constant drops by 1000 fold making the drug no longer clinically useful (16).

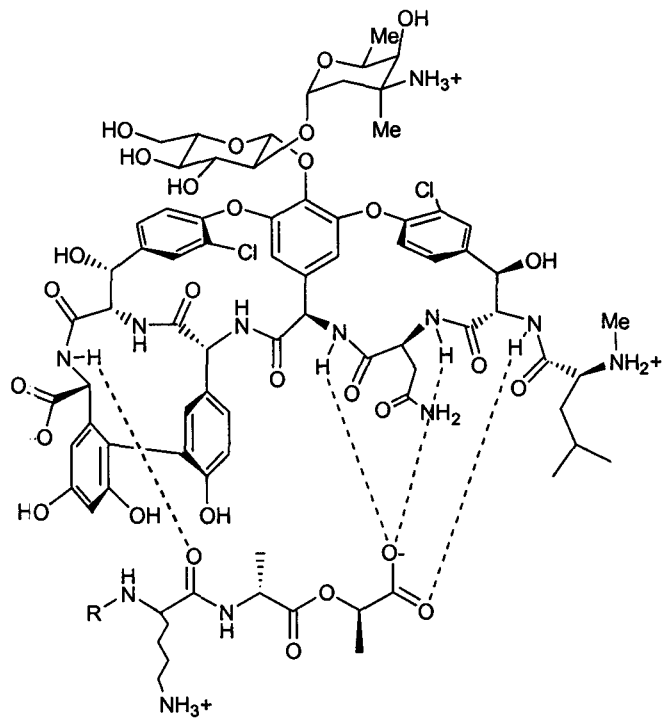


Figure 8 – Hydrogen Bonding of Vancomycin to the Altered Cell Wall in Vancomycin Resistant Gram-positive Bacteria: A simple change of D-Ala-D-Ala to D-Ala-D-Lac in the peptidoglycan results in a loss of one hydrogen bond in the binding of vancomycin to the altered cell wall. Compared to Figure 2, the change of D-Ala to D-Lac in the terminal position results in the loss of a hydrogen donor group as the amide bond is replaced with an ether bond.

The *vanHAX* genes together are responsible for the structural changes to the cell wall outlined in Figure 8. VanH is an α-keto acid reductase that serves to specifically produce D-Lac. Enzymatic studies of VanH showed that it could reduce pyruvate to give D-Lac, as well as produce D-hydroxybutyrate from α-ketobutyrate *in vitro* (16). VanA

serves a depsipeptide ligase that generates D-Ala-D-Lac from D-Ala and D-Lac to add to the pentapeptide linker (16, 17). VanA shows high homology to D-Ala-D-Ala ligases, commonly found in both Gram-positive and negative bacteria, which link together D-Ala-D-Ala via a peptide bond for further incorporation into the pentapeptide linker. VanA, however, has acquired a different specificity. While maintaining a low D-Ala-D-Ala ligase activity, it specifically produces D-Ala-D-Lac, joining the two via an ester bond. VanX is a Zn^{2+} dependent dipeptidase that cleaves any free D-Ala-D-Ala produced by the cell (18). It has been shown that VanX will cleave D-Ala-D-Ala almost exclusively over D-Ala-D-Lac thus increasing the proportion of D-Ala-D-Lac available for incorporation into the cell wall. The three enzyme system combines for the overall effect of efficiently replacing D-Ala-D-Ala with D-Ala-D-Lac in the pentapeptide linker.

Also necessary for inducable resistance to vancomycin are the *vanR* and *vanS* genes (19). Encoding a two component regulatory system consisting of a sensor kinase (VanS) and a response regulator (VanR), the *vanRS* system enables a regulated response in the presence of vancomycin (Figure 9). VanS is a membrane bound histidine kinase, which becomes autophosphorylated on a conserved histidine residue in response to an unknown signal. Autophosphorylation activity was first shown with a soluble maltose binding protein (MBP) fusion protein to the cytosolic domain of VanS (20). The phosphorylated MBP-VanS was then shown to be able to phosphorylate the response regulator VanR on a conserved aspartic acid residue *in vitro*. Reporter enzyme studies have shown that the system acts as a positive regulator on the promoter region of the *vanHAX* genes (21). In fact, phosphorylated VanR shows a greater than 500 fold increase

in affinity for the proposed promoter region over unphosphorylated VanR (21). Thus activated VanR serves as a transcriptional activator of the *vanHAX* genes.

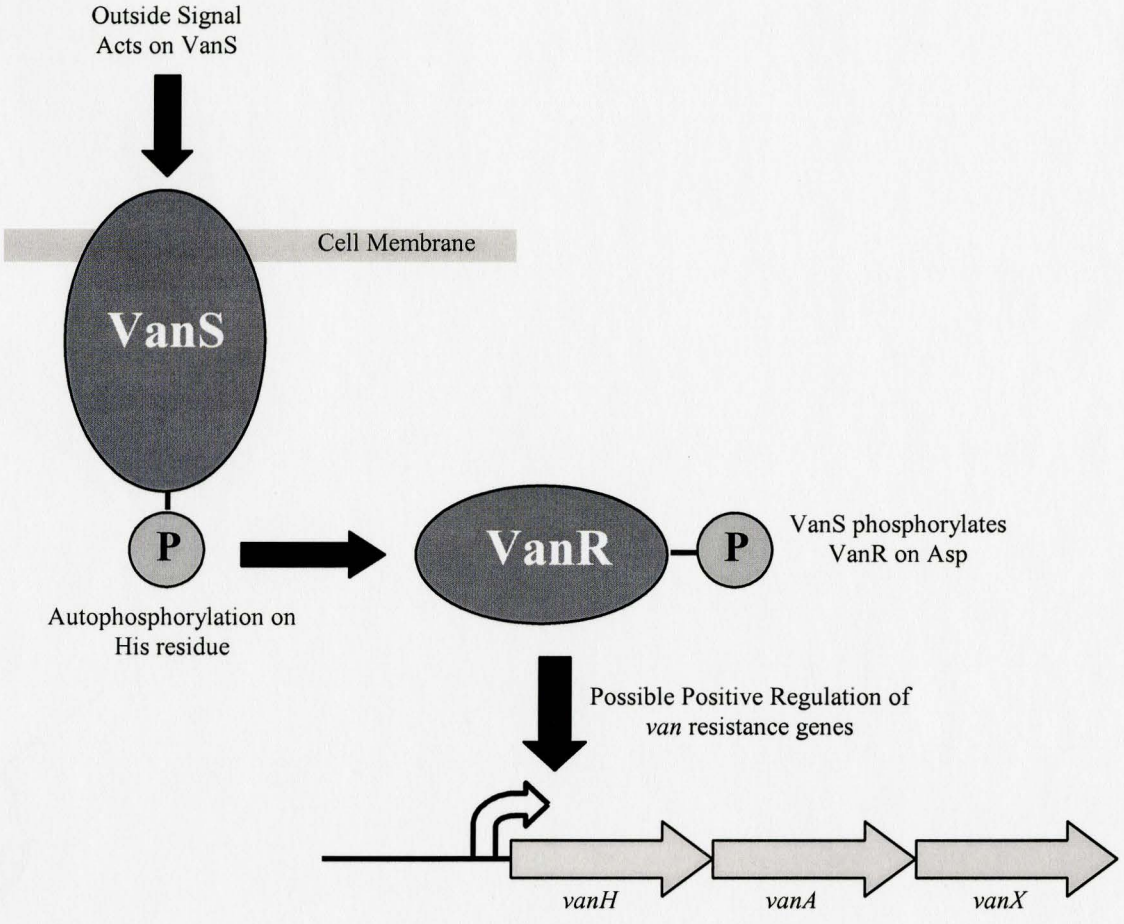


Figure 9 – The VanRS Two-Component Regulatory System: The *vanRS* genes encode a two-component regulatory system responsible for inducable resistance to vancomycin. VanS is a membrane bound sensor kinase which is autophosphorylated on a conserved histidine in response to vancomycin. Once phosphorylated, VanS can phosphorylate VanR on a conserved aspartic acid residue, thus activating it as a transcriptional activator for the *vanHAX* genes.

This mechanism of resistance is also seen in clinical isolates and in GPA producing organisms as a means of conferring resistance to the antibiotics which they produce themselves. Depending on the level of resistance, the acquired phenotype is

termed either VanA (high resistance to vancomycin and teicoplanin) or VanB (low resistance to vancomycin and teicoplanin susceptible) (1). An additional phenotype is VanD, which has intrinsic low level resistance to both vancomycin and teicoplanin due to frame shift mutations in both the D-Ala-D-Ala ligase gene *ddl* (thus only making D-Ala-D-Lac with $VanA_D$) and the $vanS_D$ gene (no normal activity with the *vanRS* two-component regulatory system) (22, 23). The VanC, VanE and VanG resistance phenotypes act in a slightly different manner, producing D-Ala-D-serine with an analogous set of enzymes to those outlined above (24, 25, 26). All provide low resistance to vancomycin, but only VanC has been fully characterized.

In addition to altering the cell wall composition, novel resistance and tolerance mechanisms to vancomycin have begun to emerge. Most recently tolerance to vancomycin was seen in *Streptococcus pneumoniae*. Normally, vancomycin will induce cell lysis in *S. pneumoniae* through action of the two-component regulatory system VncRS. VncS is the sensor kinase that serves to phosphorylate the response regulator VncR. Phosphorylated VncR acts as a repressor for a set of genes that cause autolysis. The presence of vancomycin causes VncS to dephosphorylate VncR, triggering cell death through lysis. Tolerant strains of *S. pneumoniae* have a mutated version of VncS, which no longer have phosphatase activity and therefore can no longer induce lysis as seen in Figure 10a (5).

A novel resistance mechanism differing from the Van phenotypes is present in *Staphylococci*. Several studies on *S. aureus* with an intermediate level of resistance to vancomycin have led to the hypothesis that a thickened peptidoglycan layer is responsible

for this phenotype (Figure 10b) (4). An increase in cell wall thickness results in an ability to "absorb" the GPA molecules without drastically harming the strength of the cell wall.

1.5 Assembling a Glycopeptide Antibiotic

In order to assemble a GPA there are several basic classes of enzymes that are necessary in all cases, along with others that are associated with biosynthesis when necessary. Sequencing studies of two vancomycin class GPA biosynthesis clusters (chloroeremomycin and balhimycin) and one GPA-like biosynthesis cluster (complestatin) have shown that there are indeed a specific set of genes necessary to produce a GPA (27, 28, 29). The basic sets of genes necessary are those coding for unusual amino acid production, non-ribosomal peptide synthetases (NRPS) for assembling the backbone and oxygenases for crosslinking the GPA. In addition to those, there are many modifying enzymes necessary for production of a specific antibiotic.

1.5.1 Unusual Amino Acid Biosynthesis Genes

Closer inspection of Figures 1 and 2 shows that GPAs are not composed simply of the 20 amino acids typically found in proteins. Instead we see the presence of unusual amino acids not normally produced by the bacteria for protein synthesis, thus there is a

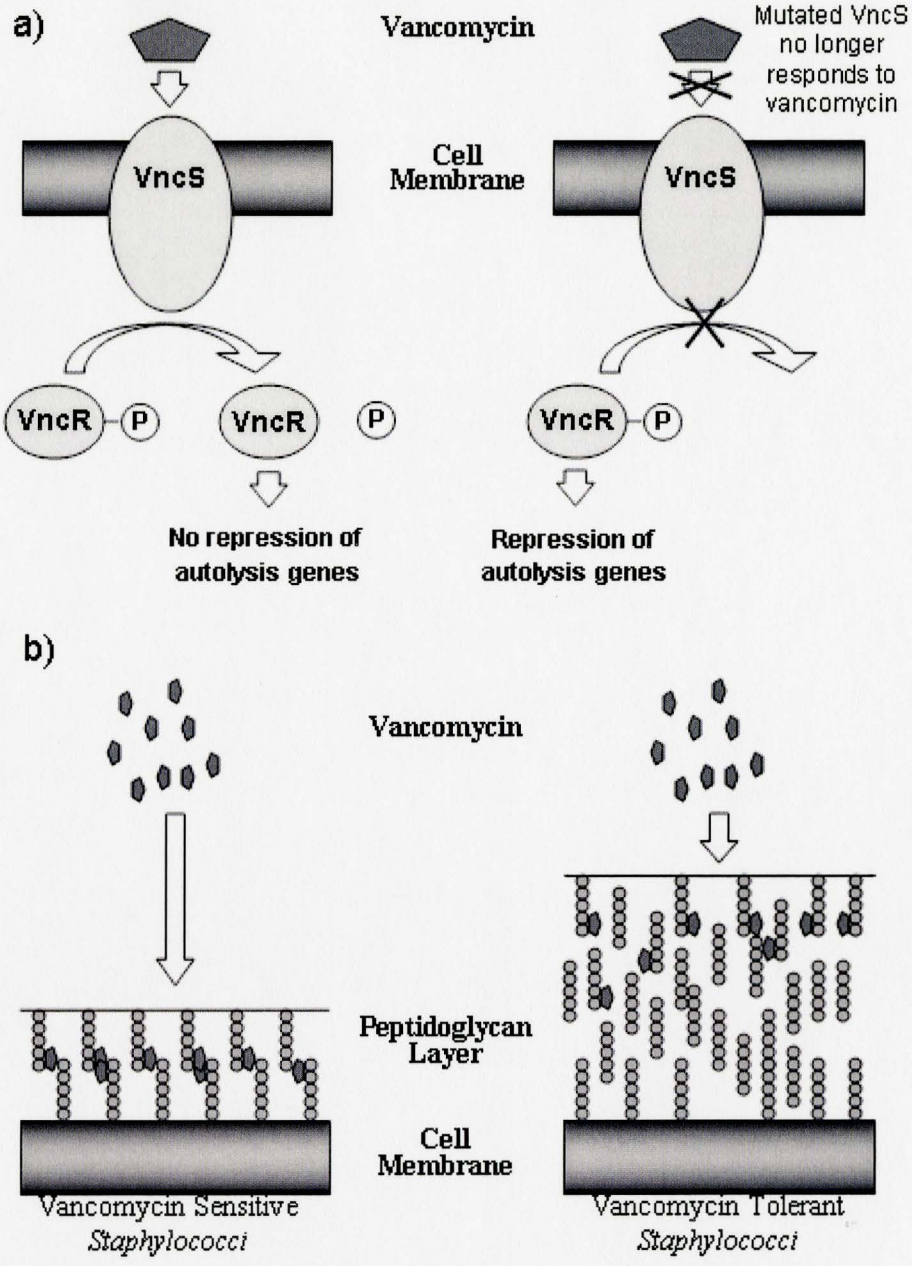


Figure 10 – Novel Resistance Mechanisms to Glycopeptide Antibiotics: Two new types of vancomycin resistance, differing from the classical Van resistance phenotypes, have emerged in recent years. Tolerance in *S. pneumoniae* (a) comes about through a natural mutation in the two component regulatory system which signals autolysis. Normally turned on in the presence of vancomycin, resistant strains no longer respond to the signal. A second novel resistance mechanism in *Staphylococci* (b) appears to be due a thickened peptidoglycan layer which can effectively absorb all the GPA without affecting the overall cell wall strength.

need to produce these amino acids specifically for incorporation into GPAs. The presence of these unusual amino acids is crucial for tailoring the GPAs as they are the sites for both the internal crosslinks and the secondary modifications. Sequencing of GPA biosynthesis clusters shows that there are a series of genes within the GPA biosynthesis clusters dedicated to the production of the unusual amino acids HPG, DHPG and OH-Tyr.

HPG is produced by a set of genes found in each of the four sequenced GPA or GPA-like biosynthesis clusters. While the arrangement and clustering of these genes can vary, the homology between them is high at the amino acid level. The four enzyme system has been worked out in the chloroeremomycin producer *Amycolatopsis orientalis* NRRL 18098 and confirmed in the complestatin producer *Streptomyces lavendulae* (30, 29). Figure 11 illustrates the strategy that is used to produce HPG. The first step, which is the only step that remains uncharacterized, is the priming of the HPG biosynthetic cycle with *p*-hydroxyphenylpyruvate from prephenate by the enzyme PD, a proposed prephenate dehydrogenase. From there the iron dependent dioxygenase HmaS converts *p*-hydroxyphenylpyruvate to L-*p*-hydroxymandelate. Hmo serves as an oxidase, using FMN as a cofactor, catalyzing the stereospecific conversion of L-*p*-hydroxymandelate to *p*-hydroxybenzoylformate. In the final step, L-tyrosine acts as an amino donor group for the aminotransferase HpgT, producing L-HPG from *p*-hydroxybenzoylformate. This final step has also been demonstrated in balhimycin production through inactivation of the orthologous aminotransferase gene *pgat* (31). The clever part of this biosynthetic cycle is

that *p*-hydroxyphenylpyruvate is regenerated from the deamination of L-tyrosine in the final step of HPG production so that the cycle can start again.

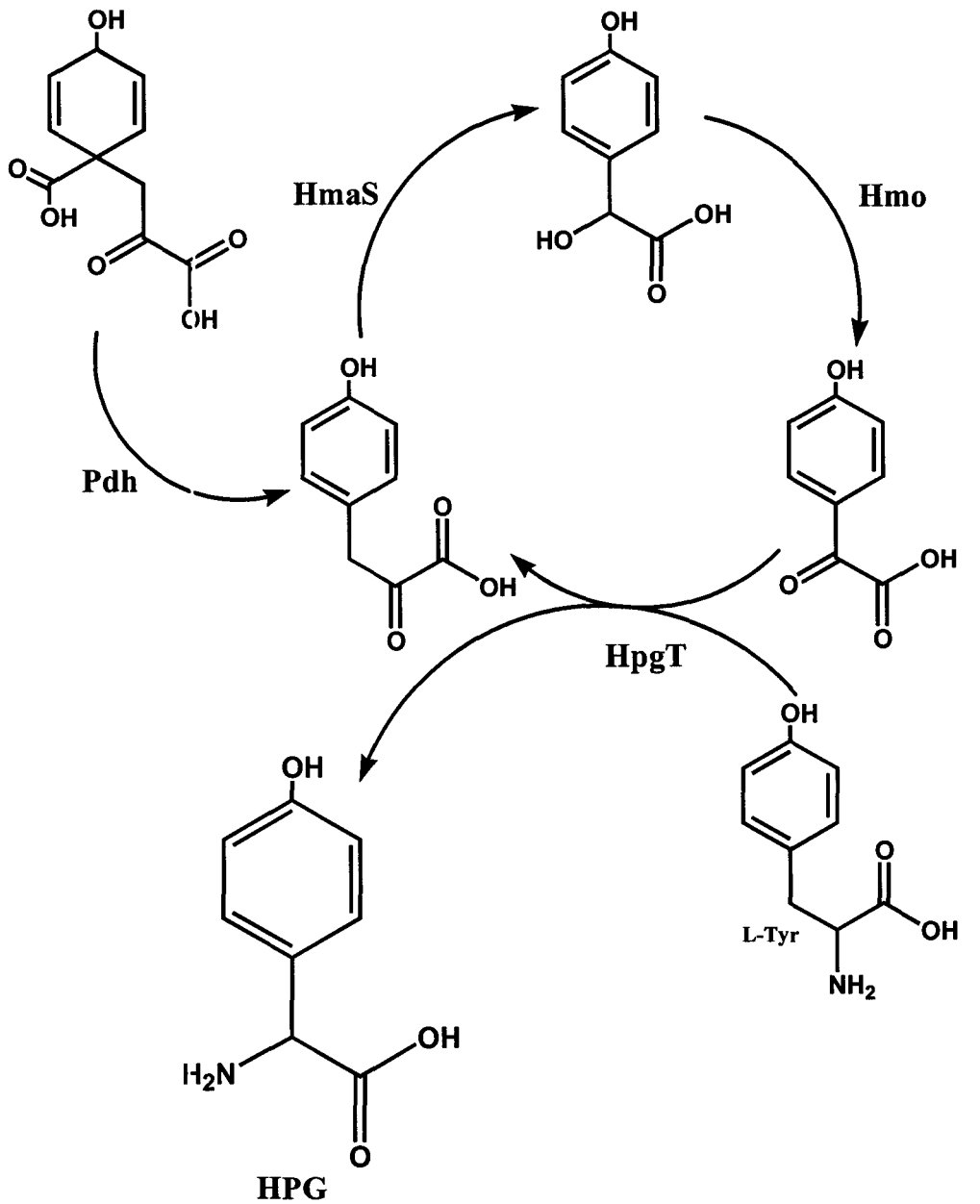


Figure 11 – Biosynthesis of HPG: The biosynthesis of HPG involves three enzymes: HmaS, Hmo and HpgT. These serve to produce L-HPG from *p*-hydroxyphenylpyruvate in a cycle which must first be primed by Pdh, which produces *p*- hydroxyphenylpyruvate from prephenate. In producing L-HPG, L-Tyr serves as the amino donor group and can then be used to generate more L-HPG in the cycle.

Despite similar structures, DHPG production takes on a completely different approach from that of HPG and is produced much like a polyketide. Characterization of this pathway through enzymatic studies has been performed with *A. orientalis* NRRL 18098. *dpgA*, *dpgB*, *dpgC* and *dpgD* are found co-linear in all biosynthetic clusters that incorporate DHPG (32). DpgA has been shown to be a polyketide synthetase, linking four malonyl-CoA molecules end to end to give three CoASH molecules, which is followed by intramolecular cyclization to give 3,5-dihydroxyphenylacetyl-CoA (DPA-CoA) as outlined in Figure 12. The rate of formation of DPA-CoA was increased 17-fold by addition of the DpgB and a further 2-fold by addition of DpgD, although the exact action of each of these proteins is unclear. One suggestion is that they contribute a dehydratase activity on the DPA-CoA intermediate. Also, whether or not the intermediate remains bound to DpgA is in question. The final step of the reaction showed DpgC to be a novel metal-free oxidase converting DPA-CoA into 3,5-dihydroxyphenylglyoxylate. From there, the DHPG biosynthesis mechanism borrows the necessary transamination step from HPG biosynthesis. Inactivation of *pgat* (the *hpgT* orthologue) from balhimycin biosynthesis has shown that removal of its activity removes both HPG and DHPG production (31).

The mechanism of OH-Tyr production was first illustrated in the biosynthesis of the coumarin group antibiotic novobiocin, which is an inhibitor of DNA gyrase (33). Since then, many analogous systems for hydroxylations of various amino acids have been identified within the biosynthetic clusters of various secondary metabolites, including Tyr

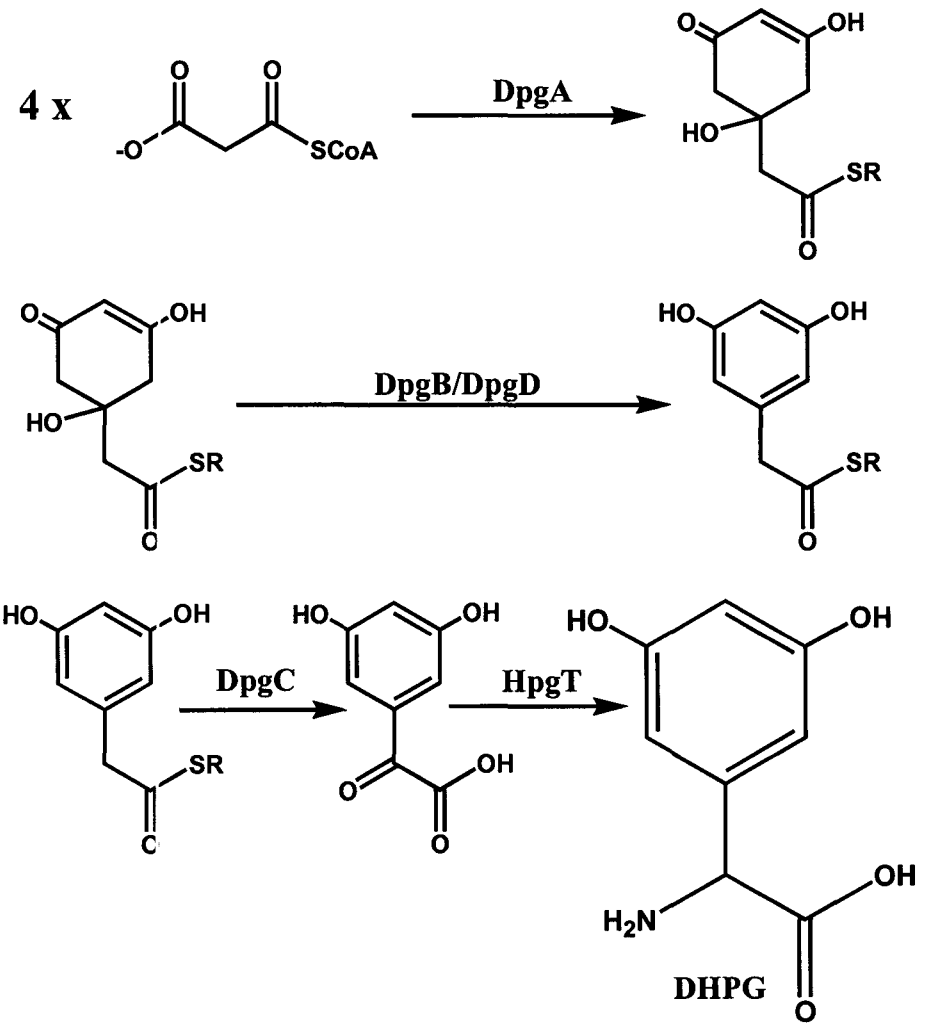


Figure 12 – Biosynthesis of DHPG: DpgA acts much like a polyketide synthetase, generating DPA-CoA from four units of malonyl-CoA. The addition of DpgB and DpgD enhance this reaction significantly, although precisely how is unclear. DpgC produces 3,5-dihydroxyphenylglyoxylate, which can then be converted to DHPG by HpgT from the HPG biosynthesis cluster (see Figure 11).

hydroxylation in the two GPA biosynthetic clusters sequenced (34). A three gene cluster composed of a small NRPS-like gene, a P450 hydroxylase and a thioesterase (not always present) are responsible for hydroxylation of the particular amino acid. In the case of OH-Tyr biosynthesis (Figure 13) the NRPS serves to specifically activate L-Tyr by adenylation then tethering it to the thiolation domain via a phosphopantetheine linker (the

specific steps of NRPS action are outlined in 1.5.2). Once linked, the cytochrome P450 hydroxylase can specifically hydroxylate the Tyr using molecular oxygen and generating water as a bi-product. Finally, the thioesterase releases the OH-Tyr product via acid hydrolysis with water.

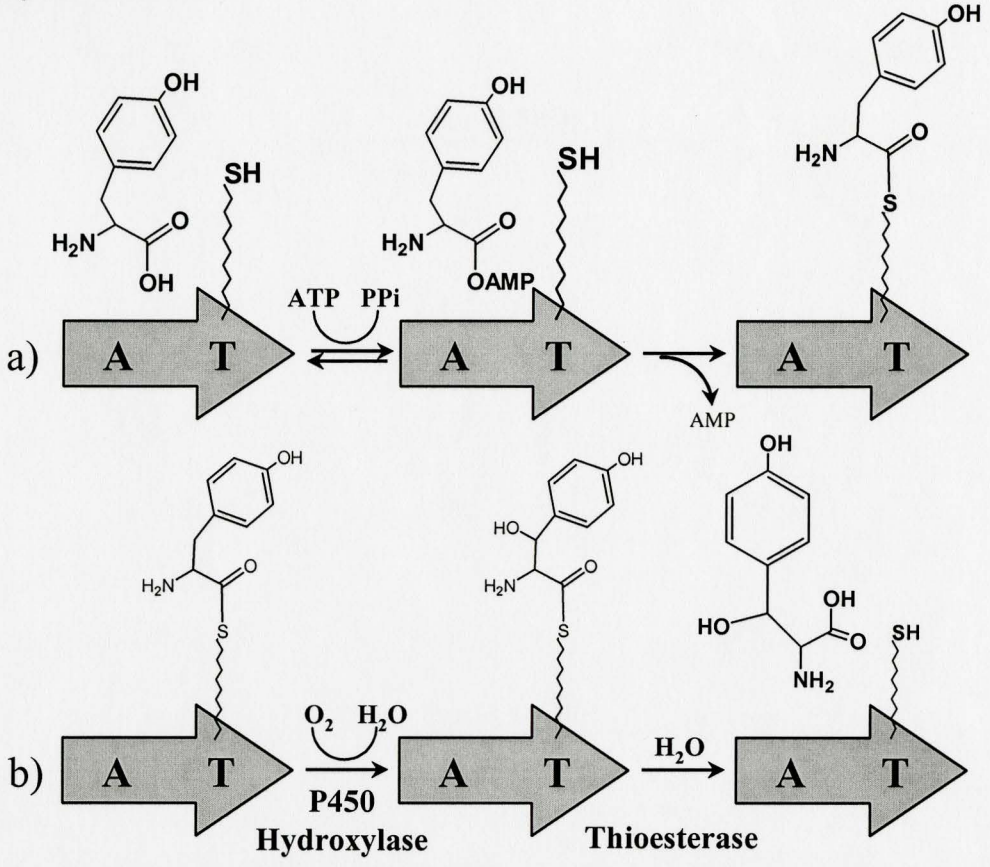


Figure 13 – Biosynthesis of OH-Tyr: OH-Tyr is produced via a three enzyme mechanism. a) First, L-Tyr is specifically recognized, activated and linked to a small NRPS-like enzyme. b) Once linked, a P-450 hydroxylase adds the hydroxyl group using O_2 and producing water as the disposal method for the remaining oxygen. From there the thioesterase releases the OH-Tyr via acid hydrolysis with water.

1.5.2 Non-Ribosomal Peptide Synthetases

With a peptide backbone containing unusual amino acids, a normal ribosomal production method involving tRNA, mRNA and ribosomes is unavailable. Production of these peptide backbones requires the action of NRPSs. Since initial work pioneered by

Fritz Lipmann in the 1950's much has been done to gain insight into how NRPSs function. NRPSs are extremely large, modular enzymes capable of specifically synthesizing peptides such as GPAs (35). Within an NRPS are modules or domains with specific functions. The adenylation (A) domain activates amino acids to form the acyl adenylate using ATP and releasing PP_i. The thiolation (T) domain attaches the activated amino acid to the NRPS via a phosphopantetheine linker with the release of AMP. The condensation (C) domain forms the peptide bonds between the amino acids. Finally the thioesterase (Te) domain releases the completed peptide. These four domains (A, T, C, and Te), outlined in Figure 14, are necessary for production of any peptide, with the A, T and C domains forming the basic repeating unit necessary for the addition of each individual amino acid. The exception is the first amino acid, which obviously requires no C domain. Epimerization (E) domains can also be present, responsible for the stereochemical conversion of L-amino acids to D-amino acids or methylation (Me) domains, which can specifically methylate amino acids.

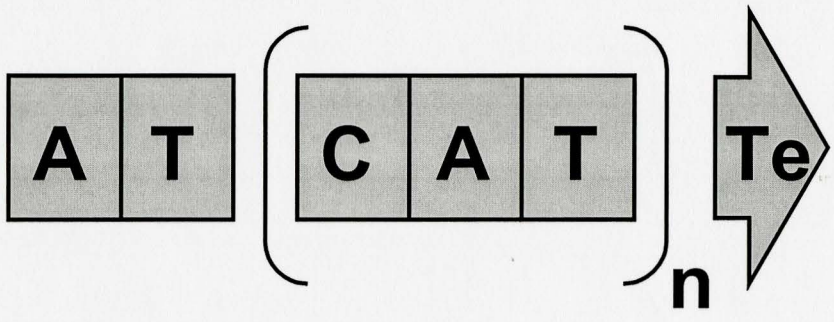


Figure 14 – Modular Nature of Non-Ribosomal Peptide Synthetases: In each NRPS gene there are four necessary domains. The adenylation (A), condensation (C) and thiolation (T) domains make up the basic repeating unit of an NRPS gene along with the terminal thioesterase domain (Te) domain which releases the completed peptide. Epimerisation and methylation domains are also present in some cases.

Organization of the NRPS is of utmost importance as the precise arrangement of the domains within the NRPS structure determines the exact nature of the peptide produced. Each repeating set of ATC domains represents the specific addition of one amino acid to the peptide. The arrangement can occur over several distinct enzymes (bacterial systems) or can be incorporated in one large multi-functional enzyme (fungal systems). The largest reported NRPS is involved in cyclosporin production and consists of 45.8 kb of DNA encoding a single protein of 1 689 243 Da (36).

Assembly of the peptide backbone for GPAs requires loading of the individual NRPS A domains with their cognate amino acids. Amino acids are both recognized and activated within the A domains of NRPSs. Activation involves adenylation of the amino acid at the carboxy group using ATP (Figure 15). Initially identified by sequence similarities to other adenylating enzymes such as amino acyl tRNA synthetases, much work has since been done to define a clear role for the A domain within the NRPS complex. The most revealing study was the crystallization and structural determination of the first A domain of GrsA (referred to as PheA), required for assembly of gramicidin, bound with the substrate phenylalanine and AMP (37). Comparison with the existing structure of firefly luciferase (also an adenylating enzyme) showed that while there was only 16% amino acid sequence identity, the 3D structure was well conserved. These enzymes consist of two subdomains, a large amino terminal domain (termed subdomain A) and a smaller carboxy terminal domain (B) (38). As the substrates bind in the region between subdomain A and B, the two subdomains move together to form the active site. The core conserved motifs, previously identified by sequence homology and confirmed

by several mutational and affinity labeling studies, line the active site (39, 40, 41). Domain A3 (see Figure 32 for all NRPS domains) contains a signature motif for a phosphate-binding loop, while A8 is conserved in various ATPases.

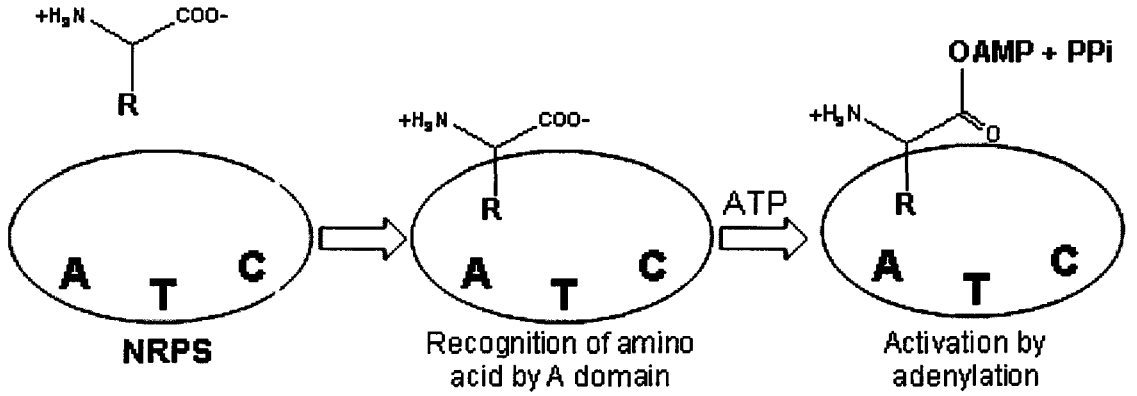


Figure 15 – Activation Through Adenylation of Amino Acids by NRPS A Domains: The A domain of NRPSs' first must recognize the correct amino acid. After recognition and binding into the active site, amino acids are activated using ATP through adenylation at the carboxy group generating an AMP-amino acid and PPi.

Mutational studies on these domains confirmed their roles which are now rationalized by the crystal structure. Domains A6 and A8 were shown through photoaffinity labeling to be involved in formation of the aminoacyl adenylate and were found lining the active site in the crystal structure (42). The crystal structure further illustrated that domains A5 and A7 were also important in aminoacyl binding. Finally, the crystal structure revealed that the role for domain A10 may be to stabilize the free phosphate on AMP in the transition state. Roles for A1 and A9 have yet to be identified. It is now believed that this general structure will prove to be applicable to all adenylating enzymes. Of even greater significance in the field of NRPSs was an exhaustive study of A domain residues that line the amino acid recognition portion of the active site for nearly

all known NRPSs by Townsend *et al* (43). Using the PheA crystal as a reference, the eight amino acid substrate recognition pocket for 80% of known A domains was determined and modeled to show how they bound their corresponding amino acid. This study also illustrated that A domains which recognized similar amino acids had similar binding pockets. In other words, the identity of the amino acid activated by a specific A domain can be determined by its amino acid sequence.

Once activated, the adenylated amino acid is covalently linked to the NRPS itself at the T domain. In the T domain, the apo-enzyme is converted to holo form by linking of phosphopantetheine (P-pant), donated by coenzyme A, to the NRPS itself at a conserved serine residue. In linking the activated amino acid, AMP is lost as the P-pant binds at the amino acid carboxy site. A model of T domain action is illustrated in Figure 16. The T domain consists of less than 100 amino acids and contains a highly conserved motif, highlighted by an invariable serine residue. Labeling experiments using [^{14}C] labeled amino acids showed that the serine was indeed the active site on the NRPS to which P-pant was bound (44). The apo form of the NRPS is phosphopantetheinylated by a phoshpopantetheinyl transferase, shown to be present in a wide variety of species (45). Confirmation of enzymatic activity by this class of transferases was completed for *Escherichia coli* and *Bacillus subtilis* and recently a 3-D structure was published showing a P-pant transferase with an acyl carrier protein (ACP) which is the polyketide/fatty acid synthetase equivalent of an NRPS (46, 47). Recombinant T domains could be used as a substrate by P-pant transferases (48, 49). The holo form of the recombinant T domains were able to link activated amino acids when provided with an A domain *in vitro*. The

purpose of joining the amino acid to the P-pant linker is to provide a flexible chain to pass the amino acid between the subdomains of the NRPS.

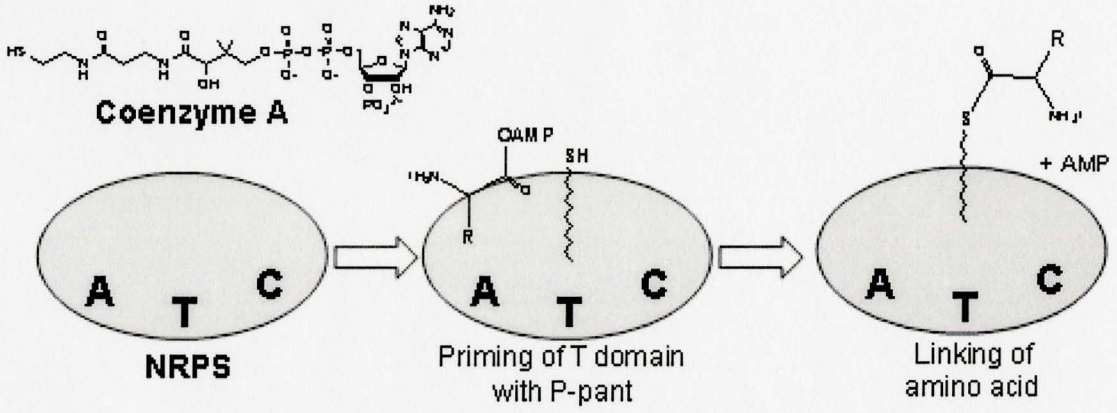


Figure 16 – Adding the P-pant Arm and Tethering the Amino Acid to T Domains: Before linking of the amino acid to the NRPS, the T domain must first be primed with the P-pant arm donated by Coenzyme A. Once the NRPS is in the holo form, the activated amino acid is linked via a thiol bond in place of the adenyl group at the carboxy site. AMP is released, while the amino acid remains tethered.

The final domain within the core NRPS repeating unit is the C domain. Once activated and then linked to the T domain, amino acids that are adjacent on the NRPS are linked via a peptide bond to the downstream amino acid remaining attached to the P-pant linker by the actions of the C domain. Initially viewed as simply the catalytic site for peptide bond formation, we now see that the C domain is much more complex than first thought. While A and T domains can be both individually and co-expressed, the C domain appears to lend much more specificity to the overall scheme of the NRPS and need to be specifically co-expressed with appropriate domains. Mutational analysis of the second histidine in the conserved catalytic motif HHxxxDG (C3 motif in Figure 36) first confirmed the role of the 500 amino acid C domain in peptide condensation (50). Using aminoacyl-*N*-acetylcysteamine thioesters (aminoacyl-SNACs), Walsh *et al* were able to

show the selectivity of C domains for their natural amino acid substrate as well as being able to distinguish between D and L enantiomers (51, 52). Phylogenic relationships between C domains of NRPSs show that those that catalyze similar condensations (i.e. L-amino acid with L-amino acid versus D-D or D-L, etc.) will cluster together. Closer study on C domains has shown that they have high specificity for the downstream amino acid to which the upstream peptide will be linked (Figure 17) (53). This acts as a proof-reading mechanism where incorrect peptide bonds are not made since an incorrect amino acid activated in the downstream site will not efficiently bind into the C domain. This has proven to be a stumbling block for creation of hybrid peptide antibiotics. While forced engineering of dipeptides has been accomplished *in vitro*; this takes place under conditions where the artificial NRPS system is fed with only the substrates for which the intended product is to have incorporated (53, 54, 55). In these cases a mixture of the various combinations of the two amino acid products is seen. Simple changes in an overall NRPS system, such as converting an A domain to recognize the methylated form of the original amino acid, have been shown to abolish condensation activity. So we see that the C domain not only acts to link together the peptide, but also to ensure the correct peptide is made each time.

While the ATC domains make up the basic repeating units, the final domain necessary for NRPS action in peptide synthesis is the Te domain. The Te domain

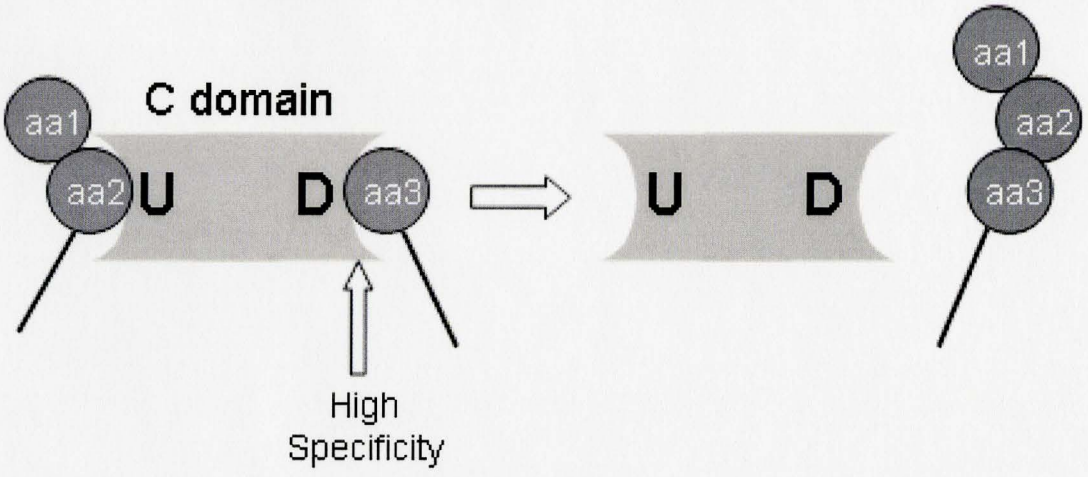


Figure 17 – Proof-reading Function of the Condensation Reaction in C Domains: In the above diagram shown are two recognition sites in the C domain (U – upstream and D – downstream). The U site has a low specificity for its particular amino acid/peptide, but the D site has a high specificity for the amino acid. An incorrectly activated amino acid will not be bound and thus no incorrect peptide bonds will be made. This acts as a sort of proof-reading mechanism. Once the peptide bond is made, the D site will no longer hold the peptide, releasing it for the next reaction.

contains the same signature GxSxG active site sequence as other thioesterases. Along with having overall amino acid sequence homology, they have been shown to act similarly through mutational studies (56). In the NRPS model, the completed peptide bound to the final T domain would be transferred onto the active site serine of the Te domain where it could then be hydrolyzed leaving the completed carboxy terminus. Te domains, much like T domains, show no specificity and domain fusions into the middle of NRPS genes will result in the expected truncations of the peptides. A series of experiments showed that various fusion of Te domains (including those in other peptide antibiotic systems which cyclize the peptide before release or leave an aldehyde) can be made to give a wide variety of novel products (57, 58, 59). The seemingly non-specific,

yet predictable nature of these domains will make them quite useful in the design of semi-synthetic peptide antibiotics.

Finally, often present are modifying domains such as the E (epimerization) domain and the M (methylation) domain. E domains have a great deal of homology with C domains and through mutational studies, have been shown to be responsible for epimerization of L-amino acids to the D enantiomer (60). The E2 domain has the same signature sequence as the C3 domain; both termed the His domain. Mutations of the second conserved histidine in both C and E His domains have eliminated activity and it is believed in both cases that the histidine acts as a base in the deprotonation of the NH_3^+ for either the condensation or racemization reaction. Studies on E domains have shown that the domain is specific for the immediate upstream amino acid (53). In other words it can not simply epimerize any amino acid in the peptide chain, only the most recently added. The high specificity of the C domain for the upstream amino acid means that once activated and linked to the T domain, the amino acid will be held in the C domain, not allowing for racemization to occur (53). Indeed, it is only after the condensation reaction has taken place and the peptide is released that the E domain can act. E domains do show some specificity, although to what degree is unclear. Interestingly, a simple scan of existing NRPSs and comparison to the product produced will quickly show that E domain presence does not determine the final L versus D configuration of amino acids in a peptide. Examples where there are D amino acids for which there are no corresponding E domains and vice versa exist in nature.

The M domain is much less complicated in its action. M domains act to N-methylate amino acids activated in the upstream A domain and exist in a wide variety of systems. For the M domain, SAM (*S*-adenosyl methionine) serves at the methyl donor (61). Methylation occurs on the amino acid prior to bond formation with the growing peptide chain (61). The difference between M and E domains is that the M domain acts before peptide bond formation, likely suggesting the C domain is specific for the methylated versus unmethylated amino acid.

1.5.3 Tailoring Enzymes

Once a linear GPA is released by the NRPS, there are still significant modifications that must occur. None are more structurally important than the crosslinking reaction catalyzed by the set of oxygenases encoded within the biosynthesis clusters of GPAs. For each GPA biosynthesis cluster sequenced thus far, there are the exact number of genes with homology to existing cytochrome P-450 monooxygenases as there are crosslinks within the GPA. Thus it appears that each is responsible for a single crosslinking reaction. Cytochrome P-450 monooxygenases are a large class of enzymes. They are iron dependent enzymes, which usually act to add oxygen (in the form of a hydroxyl group) from molecular oxygen with water as a byproduct (62). In this case they would serve to catalyze the crosslinking reaction. The nature of the crosslinking system for one set of oxygenases associated with GPA production has been worked out. Bischoff *et al* performed a series of inactivation experiments on the *oxyABC* genes in the balhimycin producer *A. mediterranei* (63, 64). These studies revealed the exact crosslinkages catalyzed by their gene products. OxyA catalyzes the ether crosslinkage of

OH-Tyr_2 and HPG_4. OxyB catalyzes the ether linkage between HPG_4 and OH-Tyr_6 and OxyC the carbon-carbon linkage between HPG_5 and $DHPG_7$. It is assumed that oxygenases for other GPA clusters will act similarly.

Analogous to the crosslinking P-450 monooxygenases, for each type of chlorination present in a GPA there is the same number of putative halogenases. The GPA halogenases appear to belong to the class of non-heme halogenases (also referred to as metal-free halogenases). In the case of the complestatin biosynthesis cluster there is a single halogenase present which is presumably responsible for the dichlorination of the HPG residues. For each of balhimycin and chloroeremomycin there is one halogenase present. In both cases there is only one type of halogenation on the GPA. Inactivation experiments on the halogenase in balhimycin (*bhaA*) has confirmed its role in chlorination events (65).

1.5.4 Modifying Enzymes

While the above set of genes is capable of generating a wide variety of distinct natural products, we see in the case of GPAs that for the most part they are structurally quite similar. As mentioned above only two different peptide backbones exist for which the crosslinking is the same, leaving only the chlorination reactions as the variants. Further structural diversity in these systems is achieved by three additional types of modifications: glycosylations, methylations and sulfonylations. In the case of glycosylations, the GPA biosynthesis cluster contains all the genes necessary for biosynthesis of the sugar(s) and for addition of the sugar group(s) onto the GPA backbone. Glycosylation has been shown to be important in homodimerization of GPAs

(through increased H-bonding) which leads to an increased effectiveness of the antibiotic. Methylation can also occur on sites on the GPA backbone. Differing from M domains within the NRPS, these reactions are catalyzed by separate methyltransferase encoded within the biosynthesis cluster. Sites of possible methylation are indicated in Figure 3 as well as on the amino terminal Leu of vancomycin-type GPAs (27).

For the A47934 model neither glycosylation nor methylation takes place. What does take place is the unique sulfonylation reaction. Sulfonylation events on metabolites are responsible for a wide number of activities. Addition of a sulfate can either activate or inactivate drugs, steroids, xenobiotics or hormones (66). Sulfate groups have been shown to be widely involved in signaling molecules where loss of the sulfate group eliminates the necessary protein-protein interaction and thus the signal as well (67). In all known cases where a specific sulfonylation reaction takes place, the sulfate donor is 3'-phosphoadenosine 5'-phosphosulfate (PAPS) (68). PAPS is generated by both eukaryotes and prokaryotes to serve as a source for activated sulfate. Sulfotransferases catalyze the addition of the 5' sulfate group onto a hydroxy group of the substrate (often a Tyr) generating the sulfonylated product as well as 3',5'-ADP as illustrated in Figure 18.

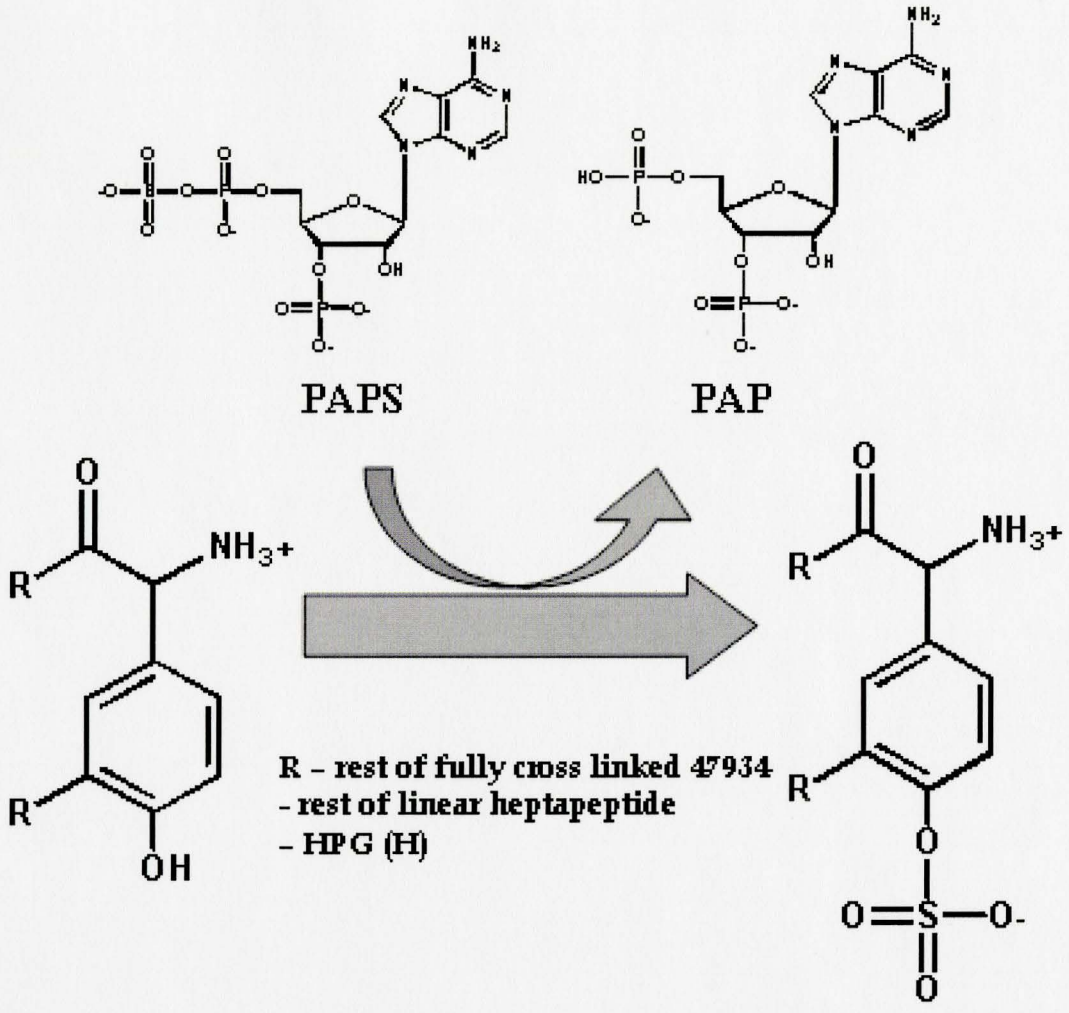


Figure 18 – General Action of Sulfotransferases: Sulfotransferases act to add a sulfate group to a hydroxyl group of their substrate. The sulfate group is donated by PAPS generating the sulfonylated product and PAP. The exact substrate for the sulfotransferase is not known, thus the R group could vary.

2. Objectives:

In order to generate secondary metabolites, such as GPAs, producing organisms harbor clusters of genes responsible for synthesizing the product. The objective of the Masters' thesis undertaken was to identify and sequence the biosynthesis cluster for the teicoplanin-like aglycone glycopeptide antibiotic A47934 in *S. toyocaensis* NRRL 15009 with at least two fold redundancy. As well as sequencing the biosynthesis cluster, we attempted to confirm that the cluster sequenced is indeed responsible for A47934 production and undertake biochemical studies on enzymes encoded within the biosynthesis cluster.

3. Materials and Methods:

3.1 Standard Materials and Techniques Used

Liquid media used for growing bacterial cultures included LB (Luria Broth - 10.0 g tryptone, 5.0 g yeast extract, 10.0 g NaCl per 1.0 L), SOC (10.0 g tryptone, 3.5 g yeast extract, 0.3 g NaCl, 0.1 g KCl, 1.0 g $MgCl_2 \cdot 6H_2O$, 0.6 g $MgSO_4$, 1.8 g glucose per 500 ml), TSB (15.0 g oxoid tryptone soya broth powder per 500 ml) and SVM (soygrit vegetative media – 1.5 g glucose, 2.0 g potato starch, 1.5 ground soygrits (Archer Daniels Midland Company), 1.0 g yeast extract, 0.2 g $CaCO_3$ and 1.0 ml corn steep liquor in 100 ml). Solid media used were LB-agar (LB with 15.0 g agar per 1.0 L), Bennet's agar (5.0 g potato starch, 1.0 g casaminoacids, 0.9 g yeast extract, 1.0 ml Czapek mineral mix (see below) and 7.5 g agar in 500 ml) and R2YE (51.5 g sucrose, 125.0 mg K_2SO_4, $MgCl_2 \cdot 6H_2O$, 5.0 g glucose, 50.0 mg casaminoacids, 11.0 g agar in 400 ml – at time of pouring plates add 50.0 ml 5.73% TES pH 7.2, 40.0 ml 3.68% $CaCl_2 \cdot 2H_2O$, 25.0 ml 10% yeast extract, 7.5 ml 20% L-proline, 5.0 ml 0.5% KH_2PO_4, 2.5 ml NaOH and 1.0 ml trace elements solution was added). Czapek mineral mix contains 10.0 g KCl, 10.0 g $MgSO_4 \cdot 7H_2O$, 12.0 g $NaNO_3$, 0.2 g $FeSO_4 \cdot 7H_2O$ and 200.0 µl of concentrated HCl in a final volume of 100 ml. Trace elements solution contains 12.0 mg of $ZnCl_2$, 60.0 mg $FeCl_3 \cdot 6H_2O$, 3.0 mg $CuCl_2 \cdot 2H_2O$, 3.0 mg $MnCl_2 \cdot 4H_2O$, 3.0 mg $Na_2B_4O_7 \cdot 10H_2O$ and 3.0 mg $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$ in 300 ml of water.

P buffer used in protoplast production/transformation was composed of 51.5 g sucrose, 125.0 mg K_2SO_4, 1.01 g $MgCl_2 \cdot 6H_2O$ and 1.0 ml of trace elements solution in 400.0 ml. Immediately before use 0.5 ml of 0.5% KH_2PO_4, 5.0 ml of 3.68% $CaCl_2 \cdot 2H_2O$ and 5.0 ml of 5.73% TES pH 7.2 was added.

Antibiotics used for plasmid resistance marker selections were ampicillin (final concentration 100 μg/ml), kanamycin (final concentration 50 μg/ml), apramycin (final concentration 50 μg/μl) and thiostrepton (10 μg/ml) unless otherwise noted. All bacterial cultures were grown at 37 °C and 255 rpm unless specified.

Plasmid DNA samples were isolated from cultures grown overnight in 1.5 ml of either LB or SOC. The QIAprep Spin Miniprep Kit (Qiagen) was used to isolate plasmid DNA yielding approximately 45 μl of DNA (concentration from 0.1 μg/μl to 2.0 μg/μl between preparations). Restriction digests were carried out in 10.0 μl reactions at 37 °C for at least one hour using the recommended MBI Fermentas or NEB buffer, 10 U of the restriction enzyme(s) and approximately 1.0 μg of DNA. For restriction digests on vectors to be used in ligations, 1.0 U of Calf Intestine Alkaline Phosphatase (MBI Fermentas) was added to the restriction digests. DNA samples were visualized by electrophoresis on 1% agarose gels and stained with ethidium bromide. When necessary, DNA bands were excised from the agarose gels and DNA was isolated using the QIAEX II Gel Extraction Kit (Qiagen). Ligation reactions contained 3.0 μl of vector backbone, 9.5 μl of insert DNA (both at similar concentrations), 1.5 μl of 10X T4 Ligase Buffer (NEB) and 1.0 μl of T4 DNA ligase (5 U/μl). Ligation reactions were carried out either

at room temperature for one hour or overnight at 16 °C. Standard transformations (either 10.0 µl of ligation reactions or 100.0 ng of supercoiled DNA samples) were performed by incubating the DNA with the competent cells on ice for 30 min, heat shocking at 42 °C for 30 sec, incubating on ice for 2 min and then outgrowth of transformed cells at 37 °C for 1 hour. For *Escherichia coli* SURE2 competent cells (Stratagene), an initial incubation with TEMED (N, N, N', N'-tetramethylethylenediamine) on ice for 10 minutes was done. Ligation and transformation of blunt end inserts that were ligated in TOPO series vectors were done according to the associated TOPO Cloning Manual (Invitrogen).

Western blots were done following the procedure outlined in Marshall (70). Transfer of the protein to the PVDF (hydrophobic polyvinylidene difluoride) membrane was done for 1 hour at 100 volts in transfer buffer (25 mM Tris-Cl pH 7.5, 192 mM glycine and 20% methanol). Blocking of the membrane was done for 1 hour at room temperature (with agitation) with 1% No Fat Milk in TBST (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 0.1% polyoxyethylenesorbitan monolaurate) and the primary antibody (1:5000) was incubated overnight at 4 °C in 0.5% No Fat Milk in TBST at the dilution indicated. After washing the membrane, the secondary antibody was incubated at room temperature for 1 hour with agitation in 0.5% No Fat Milk in TBST (1:5000 dilution, Jackson ImmunoResearch). Visualization was done using alkaline phosphatase conjuagated to the secondary antibody. In 10 ml of AP buffer (100mM Tris pH 9.5, 100 mM NaCl, 5mM $MgCl_2$), 86.0 µl of 5-bromo-4-chloro-3-indolylphosphate (0.56mM

final) (Sigma) and 86.0 µl of nitroblue tetrazolium chloride (0.48mM final) (Sigma) was added and the Western Blot was immersed until colouring reaction was complete.

3.2 Sequencing the A47934 Biosynthesis Cluster

Prior to initiating the project, a *Streptomyces toyocaensis* NRRL 15009 cosmid library was prepared by Dr. Gary Marshall using the Gigapack II Packaging Extract Kit (Stratagene) with pWE15 (Stratagene) as the vector backbone. Two cosmids, pCepC1 and pCepC4, were identified as containing part of the A47934 biosynthesis cluster using a *cepC* probe. A shotgun subclone library was created by partial *Sau*3AI digestion (optimized for ~1kb insert) of the cosmids, followed by ligation into pUC19 at the *Bam*HI site.

3.2.1 Preparation of DNA for Sequencing

The shotgun subclone library was plated on LB-agar plates containing ampicillin for antibiotic selection. The plates were prepared by spreading 100 µl of a 10^{-6} dilution of the frozen stock ($1x10^9$ CFUs/ml) and growing overnight at 37 °C (store for up to two weeks). Colonies were completely removed from the plate upon inoculation to avoid rescreening the same colony. Plasmid DNA recovered was screened by *Eco*RI restriction digestions (*Eco*RI sites flank the subcloning site) of pUC19. Plasmids determined to contain an insert and that were of sufficient concentration for sequencing (determined by visual inspection following gel electrophoresis) were numbered and stored at –20 °C. A 10.0 µl sample of each was sequenced at the MOBIX Central Facility at McMaster

University (Hamilton, Canada) using the M13 Forward and Reverse Sequencing Primers which are complementary to sequences flanking the multiple cloning site of pUC19.

3.2.2 DNA Sequence Analysis

Sequencing of the subclone library was carried out at the MOBIX Central Facility by Brian Allore, using an ABI automated sequencer. Individual sequences obtained were edited by direct inspection of the chromatograms for any nucleotide assignment errors. The 5' and 3' ends were also manually edited, removing plasmid backbone DNA and downstream DNA which was unusable. As well, each sequence was checked for vector or *E. coli* DNA contamination.

Edited DNA sequences were compiled and aligned using Seqman (DNAStar) software. As increasing numbers of DNA sequences were obtained, Seqman parameters were optimized for aligning high GC content DNA. Each contig (contiguous DNA sequence) was checked to ensure no misalignment had occurred as a result of high homology between genes within the cluster. Contigs were manually assembled according to content to arrange the gaps (stretches of unsequenced DNA) for further sequencing.

3.2.3 Screening, Isolation and Sequencing of pCepC5

The *S. toyocaensis* cosmid library was screened for a third cosmid, pCepC5, to complete the sequencing of the biosynthetic cluster. LB-agar plates containing the cosmid library were prepared as described in 3.2.1. pCepC4 shotgun subclone plasmid pUC4226 (containing *staS*) was used to probe the library. This DNA sequence was selected as it has the lowest sequence homology when it is aligned with the partially completed *S. coelicolor* genome, thus decreasing the chance of identification of false

positive clones. The partial *staS* insert from pUC4226 was removed by *Eco*RI digestion and the insert was labeled with [α-^{32}P] dATP using the Random Primers DNA Labeling System (Gibco BRL). Six plates of the *S. toyocaensis* cosmid library (~1000 colonies) were transferred to Hybond-N+ (Amersham Pharmacia Biotech) nylon membranes by overlaying these on the plates for 10 min. Filters were then submerged sequentially in solutions of 0.5 M NaOH, 1.0 M Tris-Cl (pH 7.5) and 0.5 M Tris-Cl (pH 7.5)/1.25 M NaCl for 5 min each then air dried. Auto crosslinking of DNA to the filters was performed in a TekStar hybridizing oven (Bio/Can Scientific) using the autocrosslinking function. Filters were incubated in 20 ml of hybridization buffer (20.0 µl 0.5 M EDTA, 3.5 ml 20% SDS, 5.0 ml 1M Na_2HPO_4 pH 7.2, 0.1 g BSA fraction V per 10.0 ml) for one hour at 65 °C. The radiolabelled probe was boiled for 10 min, added to 200 µl of salmon sperm DNA (5.0 mg/ml) (final counts of >5.0x10^6 cpm/ml) and incubated in 20.0 ml of hybridization buffer overnight with the filters at 65 °C with constant shaking. Filters were washed twice with 50.0 ml of low stringency buffer (0.1 ml 0.5 M EDTA, 2.0 ml 1M Na_2HPO_4, 2.5 ml 20% SDS per 50.0 ml) for 30 min at room temperature and twice with high stringency buffer (0.1 ml 0.5 M EDTA, 2.0 ml 1M Na_2HPO_4, 12.5 ml 20% SDS, 0.25 g BSA fraction V per 50.0 ml) for 20 min at 65 °C. Filters were then exposed to Kodak X-OMAT K X-ray films for visualization. Positive colonies were identified and cosmid DNA was isolated using the same protocol as for plasmid DNA (section 3.1), except 4.5 ml of culture was prepared. Cosmid DNA was further screened by Polymerase Chain Reaction (PCR) with the primers AB9363 and AB9364 complementary to the

vanAst glycopeptide resistance gene. The PCR reaction conditions were 2.5 µl 10x PCR Buffer (MBI Fermentas), 3.5 µl of 25 mM $MgCl_2$, 2.0 µl of 40 µM AB9363, 2.0 µl of 40 µM AB9364, 1.0 µl of cosmid DNA prep, 2.0 µl of 10mM dNTP mix, 1.25 µl of DMSO and 0.6 U of *Taq* DNA polymerase (MBI Fermentas) in a 25.0 µl total reaction. The PCR reaction was carried out with a 3 min initial incubation at 94 °C, followed by 30 cycles of 94 °C (1 min), 50 °C cycling down to 47 °C by 0.1 °C/cycle (1 min) and 72 °C (1 min). A positive control of pBluKStoyddl3.0(+) (plasmid containing *vanAst* created by Dr. Gary Marshall) was used to confirm amplification conditions. Cosmid DNA from colonies positive for *vanAst* were digested with *Bam*HI and *Xho*I individually (2 µg DNA, 20 U of restriction enzyme, 3.0 µl of 10x buffer in 30.0 µl total, digested overnight at 37 °C). Cosmids that displayed unique digestion patterns (thus did not contain the same insert) were sequenced using the T3 and T7 promoter primers as these promoters flank the pWE15 vector cloning site.

One cosmid, pCepC5, was chosen to complete the sequencing. A subclone library was created by mechanical shearing the cosmid DNA using the TOPO Shotgun Cloning Kit (Invitrogen). Large quantities of cosmid DNA for creation of the subclone library were prepared using the EndoFree Plamid Mega Kit (Qiagen) for low-copy number plasmid conditions. Cosmid DNA (5.0 µg) was treated under N_2 pressure for 120 seconds to obtain sheared DNA in the range of ~1.8 – 2.5 kb range. The sheared DNA (1.8 – 2.5 kb) was isolated from an agarose gel, end-treated and subcloned into the pCR4Blunt-TOPO vector as prescribed in the TOPO Shotgun Cloning Kit. Twelve

colonies were screened for correct size inserts (using *Eco*RI digestion) to ensure subcloning was successful. The ligation mix was sent to Dr. Michael Thomas in Dr. Christopher Walsh's lab at Harvard University for fresh transformation and plating of colonies. Fresh colonies were picked, plasmid prepped and sequenced at the Dana Farber/Harvard Cancer Center (DF/HCC) High-Throughput Sequencing Facility (Boston, MA) using automated technology. Sequencing results obtained from the facility was treated as in Section 3.2.2.

3.2.4 Completing the Sequencing of the A47934 Biosynthesis Cluster

Two different methods were used to close the gaps left from shotgun sequencing of the three cosmids as well as to resequence areas that had only been sequenced in one direction. For pCepC1 and pCepC4 gap specific sequencing primers were used for sequencing directly off of the cosmids. Primers were manually designed using PrimerSelect software (DNAStar) keeping in mind the high homology between genes in the biosynthesis cluster and potential primer difficulties related to high GC DNA content (higher occurrence of primer dimers and hairpin loops). Cosmid DNA was prepared as described in Section 3.2.3. Primers, along with pCepC1 and pCepC4 cosmid DNA, were sent to the MOBIX Central Facility for direct sequencing from the cosmid using their standard conditions as well as altered conditions that included higher annealing temperatures and the use of DMSO. Cosmid DNA and sequencing primers were subsequently sent to the John P. Robarts Research Institute (London, Canada) where the facility utilized dGTP Dye Terminator chemistry (Amersham Pharmacia Biosciences) on ABI Prism 377 XL DNA Sequencers. A third facility, the DF/HCC High-Throughput

Sequencing Facility was tested where they used various conditions similar to those used at MOBIX.

The remaining gaps from pCepC1 and pCepC4, as well as the gaps from pCepC5 were PCR amplified using specifically designed primers. The same was done for regions of the cluster that had only been sequenced once or where multiple coverage was ambiguous. Purified PCR products were blunt cloned using the Zero Blunt TOPO PCR Cloning Kit for Sequencing (Invitrogen) and sent for sequencing at the DF/HCC High-Throughput Sequencing Facility. Sequencing obtained from both the direct cosmid sequencing method and the gap PCR cloning method were analyzed as in Section 3.2.2.

Annotation of open reading frames and gene functions was performed manually using Frameplot 2.3.2 (National Institute of Health, Japan) and BLAST (National Institute for Biotechnology Information, USA). The entire, complete A47934 biosynthesis cluster has been deposited in the NCBI GenBank (accession number U82965

3.3 VanR and VanS Two-Component Regulatory System

3.3.1 Subcloning of *vanRst* and *vanSst* Together

In order to insertionally inactivate *vanRst* and *vanSst* attempts were made to subclone the two genes together. Primers AB18785 and AB18786 were designed to amplify both genes as well as add restriction sites to facilitate cloning. PCR amplification was carried out using 5.0 µl of 10x PCR Buffer, 5.0 µl of 25 mM $MgCl_2$, 2.5 µl of AB18785, 2.5 µl of AB18786, 2.5 µl of DMSO, 0.1 µg pCepC4, 1.5 µl 10 mM dNTP mix and 5.0 U of *Taq* polymerase in a 50.0 µl total reaction. PCR reactions

conditions were a 3 min initial incubation at 94 °C followed by 30 cycles of 94 °C (1 min), 60 °C (1 min) and 72 °C (2 min). Amplified DNA corresponding to the correct size (~2 kb) was isolated and digested with *Xho*I and *Eco*RI. pSTBlue-1 vector DNA was similarly restriction digested. The putative *vanRSst* genes were ligated into pSTBlue-1 and transformed into SURE2 cells. Transformants were screened for inserts using the restriction enzymes *Xho*I and *Pst*I. All six colonies screened gave the same restriction pattern which was incorrect so one was chosen at random and sent for end sequencing into the insert from the vector backbone. As well, restriction digests with *Xho*I and *Eco*RI were done on the six samples.

The experiment was repeated using new primers that did not contain the promoter region. Using primer AB18810 in place of primer AB18785 the PCR conditions now used were 10.0 µl of 10x PCR Buffer, 10.0 µl of 25 mM $MgCl_2$, 5.0 µl of AB18810, 5.0 µl of AB18786, 5.0 µl of DMSO, 0.2 µg pCepC4, 3.0 µl 10mM dNTP mix and 5.0 U of *Taq* polymerase in a 100.0 µl total reaction. PCR reactions conditions were a 3 min initial incubation at 94 °C followed by 30 cycles of 94 °C (1 min), 60 °C (1 min) and 72 °C (2 min). The subcloning of the PCR amplified *vanRSst* genes, along with transformation and screening, was performed as described above, except the *Xho*I restriction site used was replaced with an *Xba*I restriction site and the vector used was pUC18. Screening for correct insertion of *vanRSst* into pUC18 was done by both restriction mapping with *Xba*I and PCR reamplification of the *vanRSst* genes using the conditions above.

3.3.2 Subcloning of *vanR* and *vanS* from *S. toyocaensis* and *S. coelicolor*

PCR amplification of *vanRst* and *vanSst* from *S. toyocaensis* and *vanRsc* and *vanSsc* from *S. coelicolor* was done individually using specific primers. Primers AB22119 and AB22120 were used for *vanRsc*; AB22121 and AB22122 for *vanSsc*; AB22123 and AB22124 for *vanRst*; and AB22125 and AB22126 for *vanSst*. PCR conditions for all four genes were 5.0 µl 10x ThermoPol Buffer (New England Biolabs), 5.0 µl 100 mM $MgSO_4$, 0.1 µg pCepC4, 2.5 µl of each primer,1.5 µl of 10mM dNTP mix, 5.0 µl of DMSO and 1.0 U of Vent DNA polymerase (New England Biolabs) in a 50.0 µl total reaction. The one exception was for *vanRsc* where only 2.5 µl of DMSO was used. The PCR reaction was 94 °C incubation for 3 min, then 25 cycles of 94 °C (1 min), 58 °C (1 min) and 72 °C (1.5 min). *vanRst* and *vanRsc* were subcloned into pET22b using *Nde*I and *Hind*III and *vanSst* and *vanSsc* were subcloned into pMal-C2 using *Hind*III and *Eco*RI. Subclones which were sequenced and shown to be mutation free were then transformed into *E. coli* BL21(pLys) competent expression cells (Stratagene).

3.3.3 Overexpression, Purification and Activity of VanS

1 L of LB with ampicillin was inoculated with 0.5% of an overnight culture of each of the pMal-VanS strains created from *S. toyocaensis*, *S. coelicolor* and the positive control pMAL-VanSen from Dr. Gerry Wright. Cultures were grown to an OD_{600nm} of approximately 0.6 and then induced for overexpression by adding IPTG (isopropyl-β-D-thiogalactopyranoside) (final concentration 1 mM) followed by continued growth at 37 °C

for 3 hours. Cells were harvested by centrifugation, resuspended in 20 ml of column buffer (20 mM Tris-Cl pH 7.4, 200 mM NaCl, 1 mM EDTA and 1 mM DTT) and lysed by passage through a French press three times. All three pMal-VanS fusion proteins were isolated on an amylose resin according to the pMal Protein Fusion and Purification System protocol (New England Biolabs). Fractions were separated on 11% SDS polyacrylamide gels to determine purity and which fractions (3 ml each) to pool.

Autophosphorylation activity was assayed using conditions optimized for the positive control pMal-VanSen by Wright *et al* (20). Approximately 7.0 mg of each of the MBP-VanS enzymes were incubated in 10.0 µl reactions with 1.0 µl of 10x VanS Reaction Buffer (500 mM Tris, 500 mM KCl and 10 mM $MgCl_2$, pH 7.4), 1.5 µl of [γ-^{32}P] ATP and 0.5 µl of 100 mM ATP for 1 hour at room temperature. The reactions were visualized by separating the samples on 11% SDS polyacrylamide gels and exposure overnight to X-ray film.

Unsuccessful trials under conditions optimized by Wright *et al* for VanSen led to attempting altered conditions (20). Reactions were carried out using 7 mg of enzyme in 10.0 µl reactions with 1.0 µl okadeic acid, 1.0 µl 10 mM ATP, 1.0 µl [γ-^{32}P] ATP and 1.0 µl of 10x Reaction Buffer at room temperature for 20 min. Two different buffer systems used were the original 10x VanS Reaction Buffer and 10x MOPS Reaction Buffer (500 mM MOPS, 500 mM KCl and 50 mM $MgCl_2$, pH 7.0).

4. Results:

4.1 DNA Sequencing Results

4.1.1 Cloning the A47934 Biosynthesis Cluster

The A47934 biosynthetic gene cluster was contained on three overlapping cosmids: pCepC1, pCepC4 and pCepC5. Detailed results and discussion on the creation of the cosmid library and isolation of the cosmids pCepC1 and pCepC4 are given in Dr. Gary Marshall's Ph.D. thesis (70). As shotgun sequencing of pCepC1 and pCepC4 progressed, it became apparent that some predicted genes necessary to produce A47934 were missing. The *S. toyocaensis* cosmid library was screened using a radiolabelled probe specific for an upstream region of the biosynthetic cluster so far sequenced. Figure 19 shows the results of a secondary screen on 18 putative positive cosmids, which consisted of PCR amplification of the *vanAst* gene. Our hypothesis in applying this double screen was that the resistance genes and the biosynthesis genes were linked. Five cosmids (labeled 3A, 3B, 4C, 5B and 6C) gave strong positive results in the PCR screen. Three cosmids (4C, 5B and 6C) were chosen to establish if the positive cosmids contained the same, or mostly the same, *S. toyocaensis* insert. Each cosmid was digested with *Bam*HI and *Xho*I individually (Figure 20) and showed a unique restriction pattern. All three cosmids were sequenced using the T3 and T7 promoter primers complimentary to the cosmid DNA flanking the multiple cloning site which would reveal the DNA sequence on the 5' and 3' end of the *S. toyocaensis* insert. Sequencing of cosmid 6C

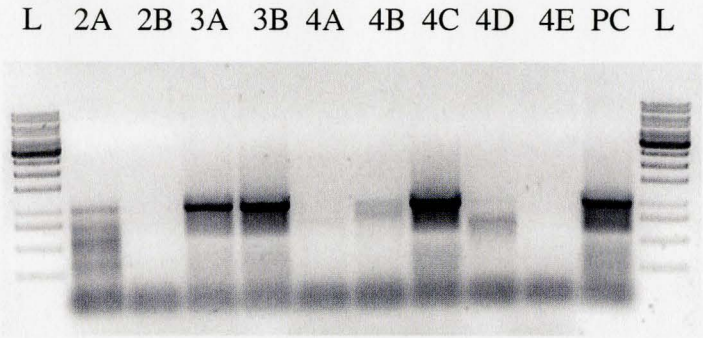


Figure 19 – PCR Results on Putative pCepC4 Overlap Cosmids: A *S. toyocaensis* cosmid library was screened for the pCepC4 upstream region using a *staS* probe. The 18 positive cosmids were rescreened to see if they contained *vanAst* using PCR, which is shown above. The cosmids 3A, 3B, 4C, 5B and 6C all gave positive results (5B and 6C are not shown). A positive control (PC) of pBluKStoyddl3.0(+) was used. L – 1 kb DNA ladder.

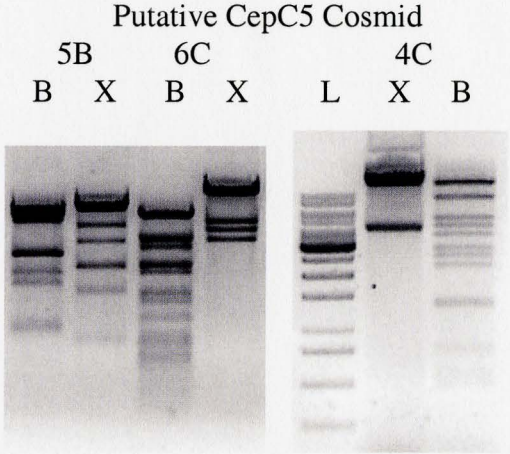


Figure 20 – Restriction Mapping of *vanAst/staS* Positive Cosmids: Three of the cosmids with contained both *vanAst* and *staS* (4C, 5B and 6C) were restriction enzyme digested and mapped using the enzymes B – *Bam*HI and X – *Xho*I. Each showed a distinct restriction pattern indicating that none contained the same *S. toyocaensis* insert. L – 1 kb DNA ladder.

from the T7 promoter primer contained the 5' end of *staC*, while sequencing from the T3 promoter primer contained the 3' end of *vanXst* in addition to approximately 300 bp of upstream sequence. Thus cosmid 6C (renamed pCepC5) contained the entire *vanHAXst*

resistance cluster as well as the smallest possible region between it and the biosynthesis genes (illustrated in Figure 21).

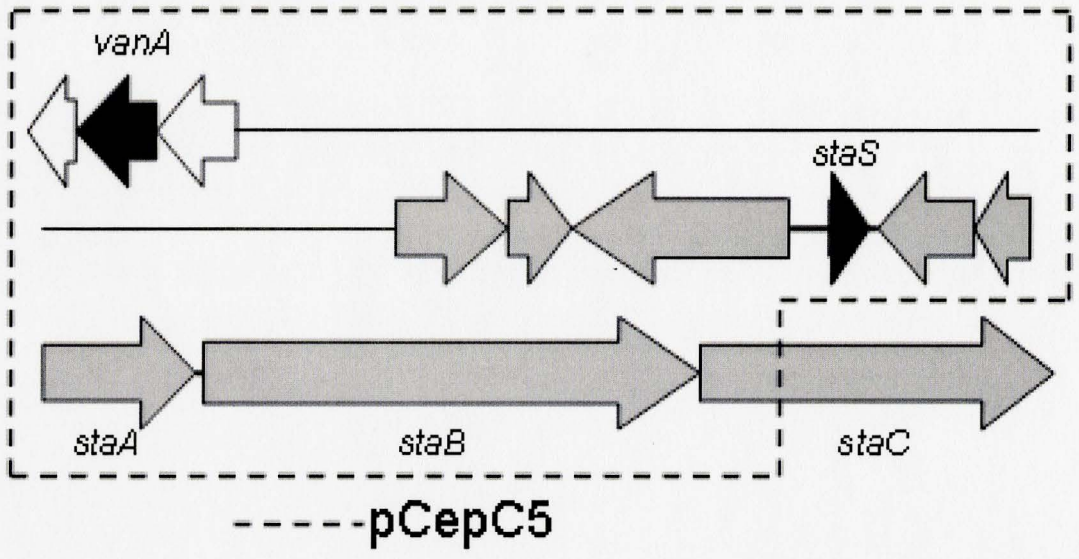


Figure 21 – Characteristics of the Third A47934 Biosynthesis Containing Cosmid pCepC5: Shown above are the characteristics of pCepC5 determined before shotgun sequencing of the cosmid began. Indicated is the area of pCepC4 (*staS*) used to probe for pCepC5 and the *vanAst* gene which was used as a PCR probe. pCepC5 shows a great deal of homology with pCepC4 (grey arrows part of pCepC4) while still containing the *vanHAX* resistance genes.

A shotgun subclone library for pCepC1 and pCepC4 was prepared by partial *Sau*3AI digestion followed by purification and subcloning of DNA fragments of approximately 1 kb. As a result of problems associated with this approach (see Discussion) the shotgun subclone library for pCepC5 was prepared in a different manner, using DNA shearing instead of restriction enzyme digests. Cosmid shearing conditions were optimized for fragments of 1.8 - 2.5 kb (Figure 22).

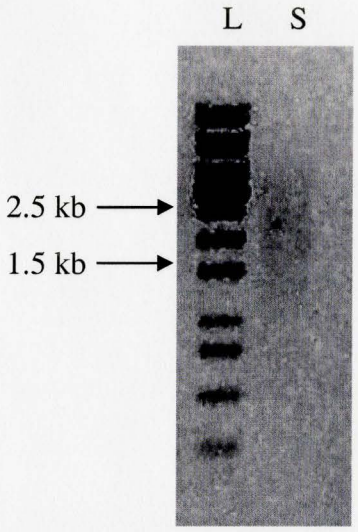


Figure 22 – Shearing of pCepC5 Cosmid DNA for Creation of Subclone Library: The above figure shows a 1% agarose gel shows a sample of the pCepC5 sheared DNA which was used to create the pCepC5 subclone library. Shearing of DNA under N_2 pressure was optimized for DNA in the range of 1.8 to 2.5 kb. L – 1 kb DNA ladder; S – sheared pCepC5 DNA

4.1.2 Sequencing the A47934 Biosynthesis Cluster

Shotgun subcloning sequencing results for the two methods used to create the subclone libraries, *Sau*3AI partial digestion and DNA shearing, are compared in Table 1. A total of 578 reactions were done during sequencing of pCepC1 and pCepC4, with 60% being *S. toyocaensis* DNA. Problems were encountered with high amounts of vector backbone sequenced, as well as *E. coli* contamination and a high number of sequencing reactions which did not yield good data. A total 768 sequencing reactions performed using pCepC5 were completed with a success rate of 44%. Unfortunately the facility was new at performing high-throughput DNA sequencing so there were cases where entire 96 well plates were ruined. Once the facility had worked out the correct conditions, the success rate for *S. toyocaensis* DNA jumped to 74% (out of 384 sequencing reactions).

Table 1 - Comparison of shotgun sequencing results using the restriction enzyme digested subclone library (pCepC1 and pCepC4) and the DNA sheared subclone library (pCepC5):

pCepC1 and pCepC4 Results	**pCepC5 Results**
Shotgun Sequencing Results	***Shotgun Sequencing Results***
Total number of sequencing reactions done = 578	Total number of sequencing reactions done = 768
Number that contain	Number of successful reaction = 340 (two plates of 96 each were ruined)
- *S. toyocaensis* – 346	
- vector DNA – 131	
- unreadable – 88	
- *E. coli* contamination – 13	
	Percent of reactions with usable data – 44%
Percent of reaction with usable data - 60%	**Percent of reactions with usable data (excluding failed plates) – 74%**

Shotgun sequencing of the three cosmids left several gaps in the cluster that required sequencing, as well as regions which had only been sequenced once over. Two different methods were used to sequence these regions. The first technique attempted was direct sequencing off of pCepC1 and pCepC4 with specifically designed primers. This was attempted using 93 different primers with only 52 successful sequencing reactions (56%). These 93 sequencing reactions were attempted using various sequencing techniques at three different sequencing facilities, all of which gave the same results. Conditions altered included DMSO concentration, annealing temperature and concentration of template and primers used. Subsequently a second technique was used; sequencing of subclone PCR amplified fragments. This was done with the remaining gaps/single coverage regions of pCepC1 and pCepC4, along with those on pCepC5. This strategy worked 94% of the time and allowed us to sequence those regions three times over. A comparison of the results is shown in Table 2.

Table 2 - Comparing the direct cosmid sequencing method (pCepC1 and pCepC4) and the PCR amplified subclone sequencing method (pCepC5):

pCepC1 and pCepC4 Results	**pCepC5 Results**
Cosmid Sequencing Results	***PCR Subclone Sequencing Results***
Total number of sequencing reactions done = 93	Total number of sequencing reactions done = 100
Number of successful reaction = 52	Number of successful reaction = 94
Percent of reactions successful – 56%	**Percent of reactions successful – 94%**

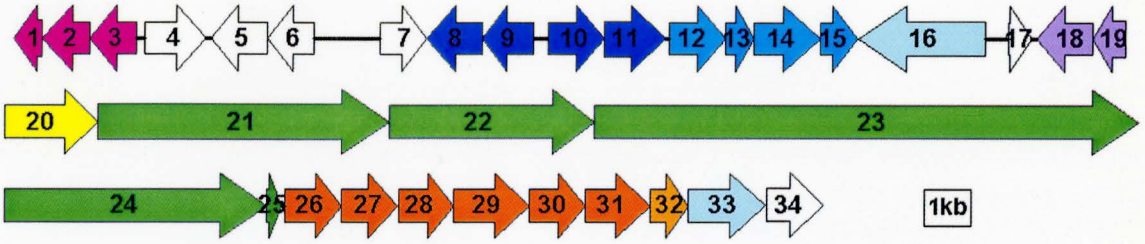
4.1.3 Editing, Compiling and Annotating the A47934 Biosynthesis Cluster

All sequences received were first submitted to a nucleotide BLAST search to screen for vector DNA and *E. coli* contamination. As well, all sequences were hand edited to ensure no computer errors in assigning nucleotides and removing vector DNA at the 5' end. The resulting sequences contained only *S. toyocaensis* DNA and were compiled using Seqman. As the amount of sequencing increased, the parameters were optimized for assembling the DNA. Due to the high GC content of the DNA, sequences were often incorrectly added to the ends of contigs if the Percent Match parameter was set too low. If the Percent Match was set too high, regions which should have aligned over small stretches at the ends of contigs would no longer align if there were any errors in the sequencing. Final parameters were set for a Percent Match of 85% and a Minimum Match Size of 12 bp to give accurate alignment of the DNA.

In order to arrange the contigs such that the gaps in sequencing could be determined for PCR amplifications, translational BLAST searches against proteins (BLASTX) for the contigs were done. Based on the content of the contig, sequences were arranged by comparison to the available chloroeremomycin cluster. The regions

containing the NRPS genes were aligned based on the domain content of the contigs. The amino acid structure of A47934 was used to estimate the approximate content and order of the NRPS domains.

Annotation of the cluster was initiated once sequencing was complete. BLAST searches of the cluster revealed stretches of DNA which putatively encoded genes of assigned function within the A47934 biosynthesis cluster. The exact open reading frame (*orf*) was determined using Frameplot 2.3.2 (National Institute of Health, Japan). The program allows for use of the unusual start codons in *Streptomyces*, ATG, GTG and TTG in assigning *orfs*. *orfs* are assigned based on the %GC content of the third base pair in the codon (wobble position). The degeneracy of the amino acid code combined with the high GC content of *Streptomyces* DNA has resulted in the wobble position containing approximately 90% GC. A full discussion of this program is available in Kieser *et al* (71). In total the A47934 biosynthesis cluster contains 34 *orfs* over ~68 kb of DNA. A detailed description of the cluster contents can be found in Figure 23. The completed A47934 biosynthesis cluster has been deposited in the NCBI GenBank (accession number U82965).



orf	A47934	Start-Stop (bp)	Cl-E	Bal	Com	Proposed Function
1	VanXst	914-288	-	-	-	D-ala-D-ala dipeptidase
2	VanAst	1951-911	-	-	-	D-ala-D-lactate ligase
3	VanHst	2931-1939	-	-	-	Lactate dehydrogenase
4	MurXst	3059-4399	-	-	-	D-Ala-D-Ala Adding Enzyme
5	StaO	5578-4412	-	-	-	FemABX homologue
6	StaP	6509-5667	-	-	-	Putative membrane protein
7	StaQ	8101-9102	-	-	ComG	Transcriptional Regulator
8	Hmo	10324-9224	Hmo	ORF6	Hmo	*p*-hydroxymandelate oxidase
9	HmaS	11417-10308	HmaS	ORF5	HmaS	*p*-hydroxymandelate synthetase
10	Pdh	11592-12746	Pdh	-	Pdh	Prephenate dehydrogenase
11	HpgT	12908-14266	HpgT	PgaT	HpgT	HPG and DHPG aminotransferase
12	DpgA	14412-15584	DpgA	DpgA	-	3,5-dihydroxyphenylacetyl-CoA synthase
13	DpgB	15596-16273	DpgB	DpgB	-	Enhances DpgA activity
14	DpgC	16270-17586	DpgC	DpgC	-	3,5-dihydroxyphenylacetyl-CoA oxygenase
15	DpgD	17583-18386	DpgD	DpgD	-	Enhances DpgA activity
16	StaR	21475-18767	-	-	-	Putative flavoprotein
17	StaS	21820-22284	-	-	-	Putative DNA binding protein
18	VanSst	23466-22363	-	-	-	Transmembrane histidine kinase
19	VanRst	24148-23453	-	-	-	Two-domain response regulator
20	StaT	24407-26350	CepM	-	ComL	ABC Transporter
21	StaA	26347-32559	CepA	BpsA	ComA	Peptide synthetase (Modules 1-2)
22	StaB	32562-37037	CepA	BpsA	ComB	Peptide synthetase (Module 3)
23	StaC	37067-49267	CepB	BpsB	ComC	Peptide synthetase (Modules 4-6)
24	StaD	49287-54854	CepC	BpsC	ComD	Peptide synthetase (Module 7)
25	StaE	55192-56367	CepD	ORF1	ComE	Hypothetical Protein
26	StaF	55192-56367	CepE	OxyA	ComI	P450-related oxidase
27	StaG	56388-57542	-	-	-	P450-related oxidase
28	StaH	57532-58728	CepF	OxyB	ComJ	P450-related oxidase
29	StaI	58733-60271	CepH	BhaA	-	Non-heme halogenase
30	StaJ	60315-61493	CepG	OxyC	-	P450-related oxidase
31	StaK	61538-62845	-	-	ComH	Non-heme halogenase
32	StaL	62842-63654	-	-	-	Sulfotransferase
33	StaM	63734-65320	-	-	-	Putative non-heme iron dioxygenase
34	StaN	65783-67039	CZA382.28	-	ComF	Integral membrane ion transporter

Figure 23 - A47934 Biosynthesis Cluster Results: The figure above shows the relative size, position and orientation of each of the genes involved in A47934 biosynthesis, resistance and regulation, with the corresponding table below outlining the names and functions of the enzymes encoded by each of the genes. Cl-E – chloroeremomycin orthologue, Bal – balhimycin orthologue, Com – complestatin orthologue

4.2 VanR and VanS Two Component Regulatory System

4.2.1 Cloning of *vanRst* and *vanSst* in Tandem

In order to attempt inactivation experiments of *vanRst* and *vanSst*, cloning of those genes together was attempted. Figure 24a shows that PCR amplification was possible for the *vanRSst* genes together off of the pCepC4 cosmid and restriction sites were engineered at the ends by the primers. Cloning into pSTBlue-1 at the *Eco*RI and *Xho*I sites was done. Restriction mapping of the six subclones with *Pst*I and *Xho*I showed that each contained the same insert size. The inserts were not the expected ~2 kb, but rather ~1.3 kb (shown in Figure 24b). Sequencing results proved that the *vanRst* gene and a small portion of the *vanSst* gene had been rearranged out of the plasmid. Attempts to reclone the two genes using DNA amplified with out the promoter region met with similar results.

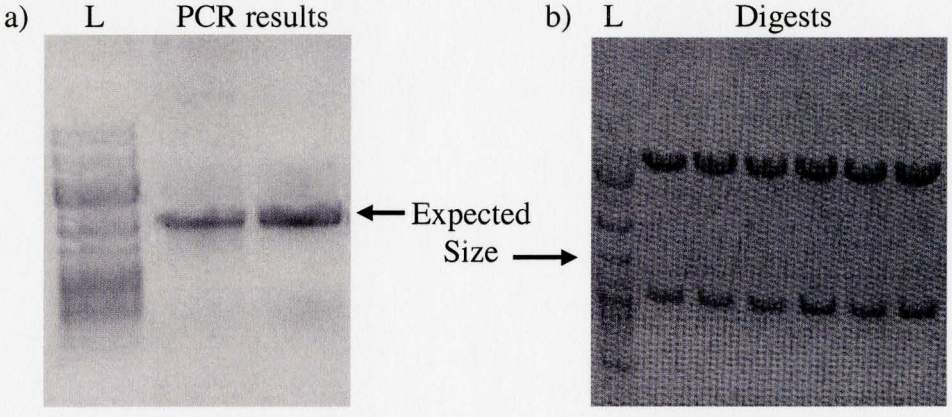


Figure 24 – Rearrangement of *S. toyocaensis* DNA in *E. coli*: A) The PCR results for amplification of the *vanRSst* genes from pCepC4. The PCR amplified genes were subcloned into pSTBlue-1. B) Restriction enzyme digestions of the plasmids thought to contain the *vanRSst* genes in pSTBlue-1. The restriction enzymes *Pst*I and *Xho*I were thought to give an approximately 2 kb fragment not the 1.3 kb fragment consistently seen. L – 1 kb ladder

4.2.2 Cloning and Purification of *vanRst, vanRsc, vanSst* and *vanSsc*

PCR amplification of the *vanR* and *vanS* genes from both *S. toyocaensis* and *S. coelicolor* individually were successful as shown in Figure 25. The *vanR* genes were cloned into pET22b (using engineered *Nde*I and *Hind*III sites in the primers) successfully, while the *vanS* genes were cloned into pMal-C2 (*Nde*I and *Hind*III). For *vanS*, the primers amplified only the putative cytosolic domain based on alignments with VanSen (Figure 26). It had previously been shown that MBP fusions (maltose binding protein fusion obtained when using pMal) with this domain yielded active VanSen, which was carried along in this case as a positive control (20). Purification on an amylose resin yielded 15.6 mg of VanSst, 18.0 mg of VanSsc and 39.0 mg of VanSen, which was >90% pure assessed by visual inspection of the Coomassie Blue stained SDS-polyacrylamide gels.

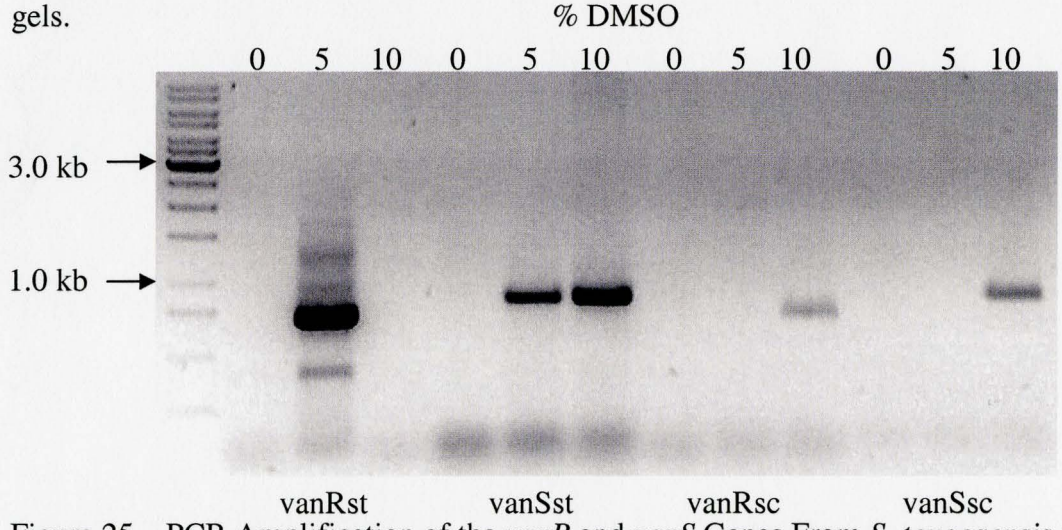


Figure 25 – PCR Amplification of the *vanR* and *vanS* Genes From *S. toyocaensis* and *S. coelicolor*: Amplification of the *vanR* and *vanS* genes from *S. toyocaensis* and *S. coelicolor* individually is shown above. It was possible to subclone these PCR products without rearrangement in *E. coli*. PCR amplification of *vanS* was a truncated version of the gene where the predicted transmembrane domain was not amplified.

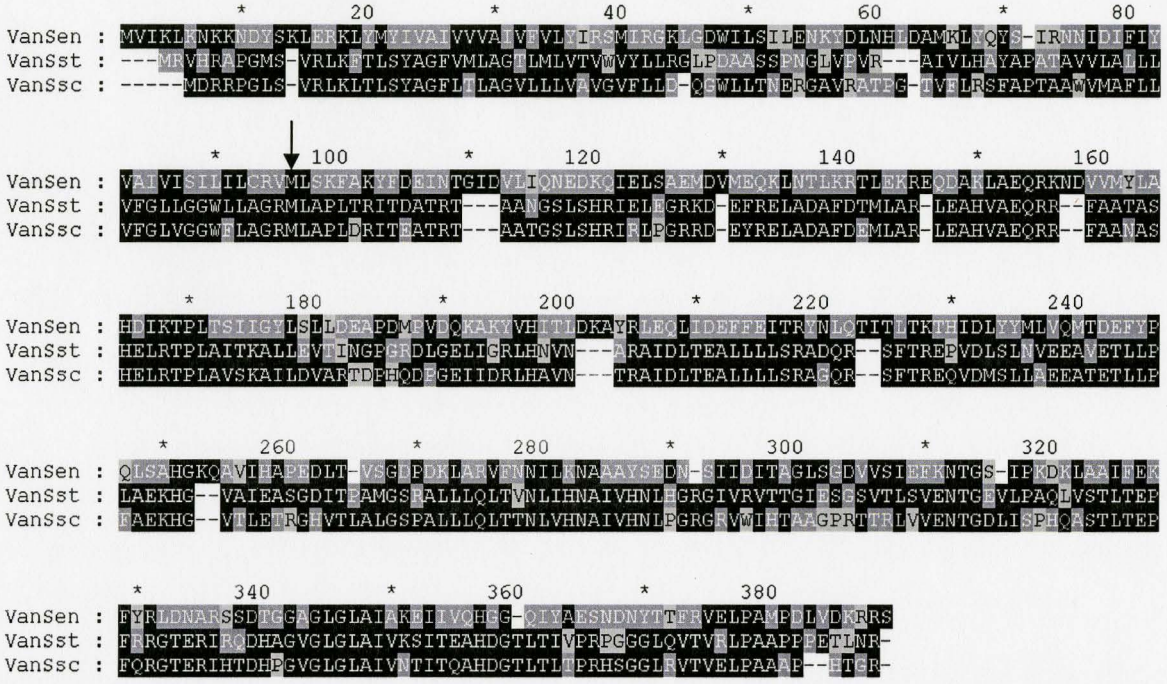


Figure 26 – Alignment of VanS Enzymes: The alignment of the VanS enzymes for *S. toyocaensis*, *S. coelicolor* and *E. faecium* is shown. Indicated is the site of the truncation (arrow), downstream to which is the putative cytosolic domain to which maltose-binding protein fusions were made.

4.2.3. Activity of VanS

Autophosphorylation activity of the purified VanS proteins was assayed using [γ ^{32}P]-ATP as the phosphate donor. VanS is predicted to autophosphorylate on a conserved His residue. In this assay, the γ phosphate is radiolabelled and thus will radiolabel VanS if it has autophosphorylation activity. Repeating conditions used by Wright *et al* yielded low activity for VanSst and no activity for VanSsc, while being able to maintain the high autophosphorylation activity for VanSen (Figure 27) (20). Repetition of the experiment using a different buffer system, MOPS, removed VanSst activity, while slightly lowering VanSen activity. VanSsc still had no activity.

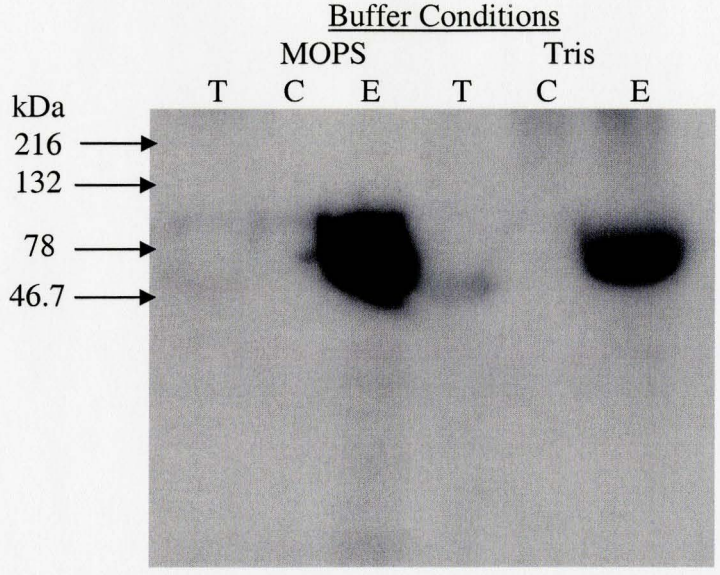


Figure 27 - Autophosphorylation Activity Present for VanS-MBP Enzymes: Using two different buffer conditions, MOPS buffer and Tris-Cl buffer, autophosphorylation activity was only clearly exhibited by the VanSen-MBP from *E. faecium* (E). VanSst-MBP from *S. toyocaensis* (T) showed low activity in Tris-Cl buffer and no activity in MOPS buffer, while no activity was seen from the *S. coelicolor* VanSsc-MBP (C).

5. Discussion:

The research described here for A47934 biosynthesis represents the first completely sequenced biosynthesis cluster of a teicoplanin-type GPA. It is also the most complete biosynthesis cluster analyzed to date, encoding 34 genes necessary for GPA production, resistance and regulation. With the sequencing of two vancomycin-type GPAs (balhimycin and chloroeremomycin) and one GPA-like metabolite (complestatin) biosynthetic cluster already complete, the additional data provided by the A47934 cluster now allows for meaningful and useful comparisons of exactly how GPAs are assembled. Variations and similarities can be highlighted among GPAs, thus suggesting definitive roles for nearly all of the gene products in the biosynthetic clusters.

5.1 Sequencing the A47934 Biosynthesis Cluster

In order to locate the biosynthetic gene cluster for A47934 within the *S. toyocaensis* NRRL 15009 genome, a cosmid library was prepared and screened using the *cepC* gene from the chloroeremomycin producer *A. orientalis* (70). This gene encodes the NRPS that adds the amino terminal DHPG and subsequently cleaves the linear GPA via the Te domain. Recent genome sequencing efforts have revealed several NRPS genes in the *S. coelicolor* genome, an organism with high homology to *S. toyocaensis*. Thus it is highly likely that the *S. toyocaensis* genome encodes sets of NRPS genes other than those involved in A47934 biosynthesis. However we predicted that it would be unlikely that there is more than one NRPS gene that encodes an A domain which recognizes the

unusual amino acid DHPG coupled with a Te domain. It is this unique characteristic that made the *cepC* a suitable probe for screening the *S. toyocaensis* cosmid library. Screening of the library resulted in four candidate cosmids (pCepC1, pCepC2, pCepC3 and pCepC4), two of which (pCepC1 and pCepC4) were chosen for further investigation based on amount of overlap. Shotgun sequencing of the two cosmids was undertaken using a subclone library created by partial *Sau*3AI digestion and ligation into pUC19. It was at this point that the project was turned over from Dr. Gary Marshall to myself.

While sequencing the pCepC1 and pCepC4 cosmids, it became apparent that a third cosmid would be needed in order to complete the A47934 biosynthesis cluster. Partial sequencing of the original two cosmids revealed that the genes encoding the enzymes thought to be responsible for unusual amino acid biosynthesis were missing. Closer inspection of the partially assembled sequences of the two cosmids (Figure 21) showed that the upstream end of pCepC4 contained what putatively was an unusual amino acid biosynthesis cluster. Unfortunately, at that time, no information on HPG, DHPG or OH-Tyr biosynthesis genes was available. The downstream region of pCepC1 appeared to contain the end of the biosynthesis cluster as well as approximately 20 kb of unrelated DNA. Based on these findings, a probe was designed to find an additional *S. toyocaensis* cosmid which would overlap with the upstream region of pCepC4 and contain the missing unusual amino acid biosynthesis genes. The first 10 kb of upstream pCepC4 DNA was submitted for a nucleotide BLAST search against the partially completed *Streptomyces coelicolor* genome. Based on the high genetic homology between *S. coelicolor* and *S. toyocaensis*, the region with the lowest homology to *S.*

coelicolor was expected to give the lowest number of false positives when used as a probe on *S. toyocaensis*. In this manner, the *orf* encoding *staS* was chosen as a new probe to screen the cosmid library. Cosmids positive for *staS* were further analyzed by PCR screening using primers specific for *vanAst*. To date, in all *van* resistance clusters, the resistance conferring genes, *vanHAX*, are located immediately upstream or downstream to the corresponding two-component regulatory system, *vanRS*. For *S. toyocaensis*, the *vanHAXst* genes had previously been identified. Sequencing 2 kb upstream and 5 kb downstream of this region revealed no corresponding *vanRSst* genes. Upon sequencing pCepC4, a two-component regulatory system showing very high homology to the *S. coelicolor vanRSsc* genes was identified (73% amino acid identity for VanS and 59% for VanR). PCR screening of *staS*-positive cosmids with *vanAst* primers showed that both genes were contained within the same cosmid in several cases. These cosmids were then subjected to restriction enzyme mapping for originality followed by sequencing of the 5' and 3' ends of the insert. The cosmid pCepC5 was found to contain the 5' end of the *vanHAXst* genes on one end and the 3' end of *staC* at the other. This cosmid was ideal as it contained the entire region of DNA between the resistance and biosynthesis genes and showed the highest amount of overlap, thereby minimizing the amount of sequencing necessary. A subclone library of this cosmid was created by mechanical shearing of the cosmid DNA and blunt end subcloning into the pCR4Blunt-TOPO vector.

Sequencing the A47934 biosynthesis cluster was a long and arduous journey, but not without its lessons. Tables 1 and 2 give a summary of the sequencing results,

comparing the original methods used for sequencing pCepC1 and pCepC4 and those implemented for pCepC5.

The first step in sequencing a cosmid is creating a subclone library. Initially, Dr. Gary Marshall created the pCepC1 and pCepC4 subclone libraries using a *Sau*3AI partial digest (70). *Sau*3AI was chosen as it is a four base pair cutter (recognizing ↓GATC), and thus it should cut frequently, DNA bands isolated from this digest could be subcloned into the pUC19 plasmid which has been predigested with *Bam*HI and dephosphorylated. The digest was optimized to give a DNA smear of greatest intensity at approximately 1.0 kb. This procedure led to two major problems. The first was the randomness of the digest. As the sequencing proceeded, it became evident that the digestion had not been completely random. In several cases, subclones carrying the exact same *S. toyocaensis* DNA inserts were seen up to one dozen times. Even more alarming was an inordinate amount of pWE15 vector backbone being sequenced. The *S. toyocaensis* DNA insert in each cosmid was approximately 40-45 kb whereas the pWE15 vector was only 8.2 kb. This gives an expected ratio of approximately 5:1 for sequenced *S. toyocaensis* DNA versus pWE15 DNA. The observed ratio was actually 3:1.

One possible explanation for this is that the *S. toyocaensis* DNA, with respect to vector DNA, is toxic to the host *E. coli* cells. Upon sequencing pCepC5 using different methods however, this was deemed not to be the case. The DNA shearing method used for pCepC5 actually showed a higher ratio of *S. toyocaensis* DNA sequenced versus vector.

sequence the cosmids. The unfortunate result was that ~15% of reactions sent for sequencing were unreadable.

In order to rectify the problems encountered in the shotgun sequencing of pCepC1 and pCep4 several changes were made. In order to circumvent the lack of randomness generated by *Sau*3AI digestion, a shotgun subcloning kit available through Invitrogen was used. This strategy permitted the mechanical shearing of DNA into random pieces. Shearing was achieved by passing the DNA through a small hole under high pressure (nitrogen gas). The size of the fragment obtained was controlled by varying the amount of time DNA was exposed to pressure. For pCepC5 an insert of approximately 2.0 kb was chosen. With a larger insert size, the entire length of DNA obtained from end sequencing of the subclone would be original. Also, any gaps spanned by a subclone could be closed by internal sequencing.

Along with changing the subcloning method and insert size, a new sequencing facility was used (DF/HCC High-Throughput Sequencing Center). Excluding the two 96-well plates that were ruined, the new sequencing center produced 70% more sequences than MOBIX that were not ruined.

Following shotgun sequencing, there still remained gaps of unsequenced DNA on the cosmids. Gaps on cosmids pCepC1 and pCepC4 were in the process of being completed as the search for pCepC5 was underway. Sequencing directly off of the cosmids with specifically designed sequencing primers was used to try and sequence the missing stretches of DNA. Similar to the problems encountered during sequencing of the subclones; the facility could not make this procedure work sufficiently for this technique

to be of practical use. Attempts were also made at the John P. Robarts Research Institute and at the DF/HCC High-Throughput Center, but were met with similar results. The three sequencing facilities could only make the direct sequencing reactions work approximately half of the time. Gaps that remained to be filled from pCepC1 and pCepC4, as well as all the gaps from pCepC5 were PCR amplified and blunt end subcloned using the Invitrogen Blunt End Subcloning Kit. This approach was initially not considered since it entailed much more work, however the lack of success in direct cosmid sequencing offered no alternative. Although time consuming, this method proved to be highly successful (>95% success rate). Overall, these combined techniques permitted the complete sequencing of the A47934 biosynthesis cluster.

For sequencing large stretches of DNA with the greatest ease, a few simple steps can now be followed. The first step would be to ensure the quality and content of your DNA to be sequenced. Visually inspecting all samples to be sequenced for correct insert (in the case of vector subcloned sequencing) and quantity immediately ensures the sample you are sending has no preparation error before you attempt to sequence. Creation of the subclone library correctly is also key. Shearing the DNA gives random samples which will enable a greater amount of original sequencing data to be generated. As well, the subclone size appears to be critical. The larger the stretch of high GC DNA in the vector, the more problems that seem to arise with sequencing. Generally subclone sizes less than 3 kb seem to work better. Sequencing with vector homologous sequencing primers works better than custom designed primers specific for subcloned stretches. This fact becomes of greater importance with longer stretches of high GC DNA subcloned. If internal

sequencing is required, there appears to be no real guide as to choosing primer sites. It appears to be random as to when a certain primer will work well or not. The final step is to carefully check all sequencing data. Sequencing high GC DNA often results is less than ideal sequencing data, thus manual inspection of all data is necessary to ensure accurate results.

5.2 Assembling the Unusual Amino Acids

A47934, shown in Figure 1, contains three unusual amino acids: HPG, DHPG and OH-Tyr. As outlined in the introduction, these unusual amino acids are assembled by enzymes encoded by sets of genes found within the biosynthesis cluster. The individual set of four enzymes necessary to produce HPG and DHPG are located within the A47934 biosynthesis cluster. The gene products Hmo, HmaS, Pdh and HpgT are orthologous to the chloroeremomycin HPG production enzymes which were characterized by Hubbard *et al* (66%, 74%, 51% and 56% similarity respectively) (30). As can be seen the homology and placement of the genes makes them the obvious candidate for HPG biosynthesis in *S. toyocaensis*. HPG biosynthesis genes differ from other unusual amino acid biosynthesis genes in that they are not always linearly arranged. While in the A47934 biosynthesis cluster the HPG production genes are all found co-linear (similar to complestatin), in both chloroeremomycin and complestatin they are scattered through the cluster; though in an analogous arrangement between the two.

Similarly, the genes necessary for DHPG biosynthesis are all found co-linearly arranged within the A47934 biosynthesis cluster. The arrangement of these genes is the same as that observed for both the complestatin and balhimycin clusters which are the

only other two known set of DHPG biosynthesis genes. Their role in DHPG biosynthesis has been illustrated for chloroeremomycin production and confirmed in balhimycin production (31, 32). Thus, it is fairly safe to assume that these four enzymes are responsible for DHPG production in *S. toyocaensis* (76%, 58%, 74% and 87% similarity between *S. toyocaensis* and complestatin DHPG genes).

Production of OH-Tyr for both of the previously sequenced GPAs utilizes a mechanism in which Tyr is hydroxylated prior to its incorporation into the peptide backbone. This mechanism, as outlined in the introduction, requires three enzymes, an NRPS, a hydroxylase and a thioesterase, all of which are coded for within the GPA biosynthetic clusters (33). Similar three enzyme mechanisms of amino acid hydroxylation are known to exist within other metabolite biosynthesis (34). Surprisingly, the A47934 biosynthetic cluster revealed no orthologues to this three enzyme system, not even in the extra 2 kb upstream and 20 kb downstream of DNA.

Inspection of the A47934 structure shows that it contains only one OH-Tyr residue within the heptapeptide backbone, as is the case with all teicoplanin-like GPAs. Teicoplanin-like GPAs also contain an unhydroxylated Tyr at position 2. This differs from vancomycin-like GPAs which incorporate only hydroxylated Tyr. This distinction, coupled with the fact that the GPA-like complestatin contains Tyr as well, meant that A domains of NRPSs corresponding to the OH-Tyr and Tyr incorporation sites for A47934, complestatin, chloroeremomycin and balhimycin could be compared. The second A domain for A47934 on StaA, as well as the sixth A domain for complestatin both specifically recognize Tyr. A domains number two and six for both chloroeremomycin

and balhimycin should recognize OH-Tyr. Figure 28 shows the amino acid alignment of the A_6 domain of the A47934 recognition pocket with the other Tyr and OH-Tyr recognition pockets from similar systems. While there is high homology between recognition pockets of OH-Tyr and Tyr, the A_6 domain would likely recognize Tyr.

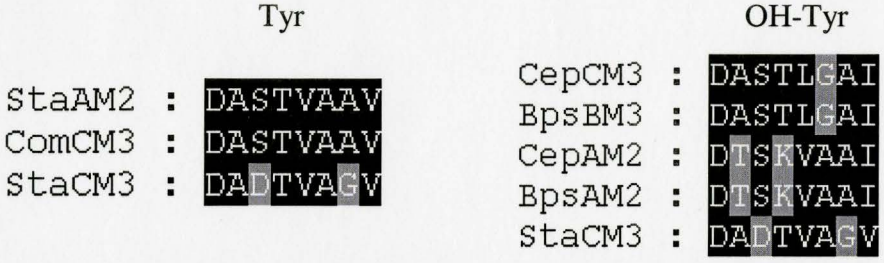


Figure 28 – Alignment of A_6 Domain Recognition Sites for A47934: The A_6 domain recognition site is shown aligned with other GPA A domain recognition sites of both Tyr and OH-Tyr. While there is high homology between all, it appears to be a closer match to those that activate Tyr. StaCM3 – StaC 3rd A module; Com – complestatin, Bps – balhimycin, Cep – chloroeremomycin NRPSs.

orf16, termed *staR*, encodes a putative flavoprotein. Identified by the conserved flavin-binding domain, $GxGxxGx_{18}E$, it is hypothesized that this large enzyme is a flavin dependent monooxygenase (72). If this is the case, then StaR could specifically recognize and hydroxylate the Tyr_6 incorporated in the peptide backbone. There is precedence for this type of system in pristinamycin biosynthesis (73, 74). In converting pristinamycin IIB to pristinamycin IIA, a D-proline is oxidized to give a double bond within its ring structure. A flavin dependent monooxygenase was shown to be involved in this oxidation, presumably by hydroxylating the proline (Figure 29) which then undergoes dehydration (74). The putative flavin monooxygenase StaR in A47934 biosynthesis may act in an analogous manner. This step would occur after incorporation of the Tyr, but whether it takes place while the peptide is still linked to the NRPS, released but still

linear or fully cross-linked is unclear. Given the specificity of hydroxylation, on Tyr_6 but not Tyr_2, it appears more logical for hydroxylation to occur on the crosslinked heptapeptide.

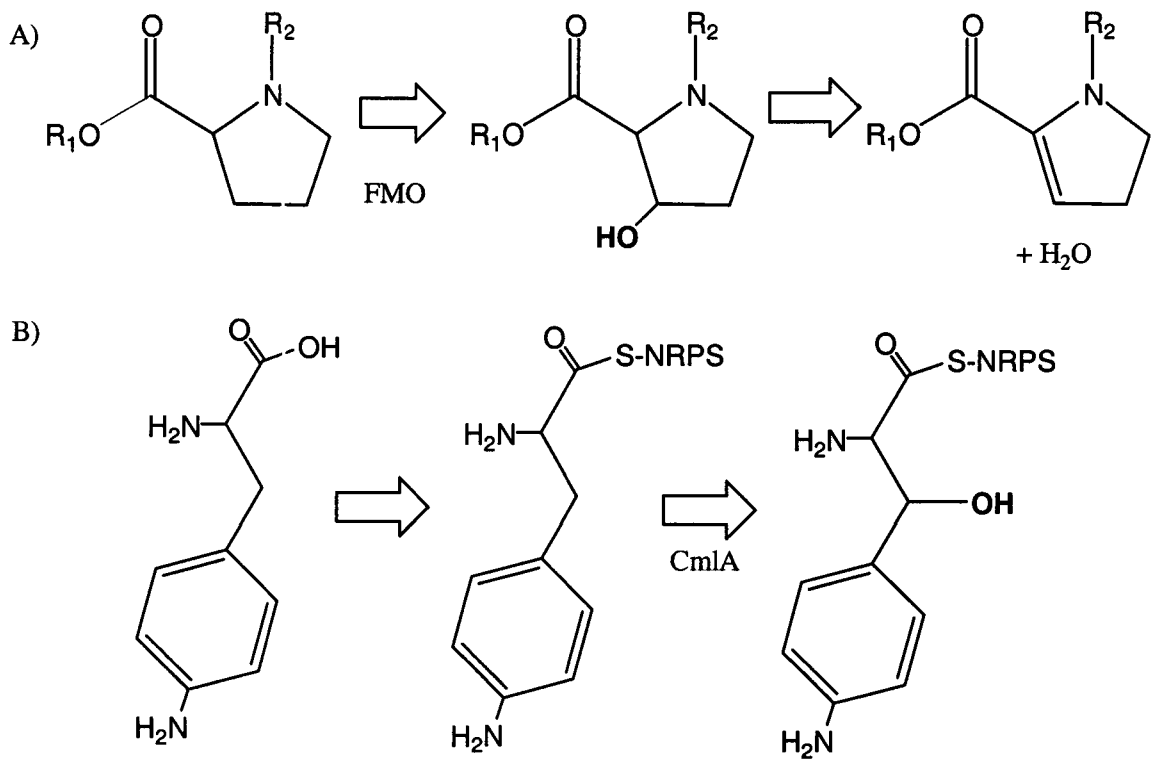


Figure 29 – Alternative Modes of Amino Acid Hydroxylation: A) In pristinamycin biosynthesis (R1 and R2 – rest of pristinamycin), the proline becomes hydroxylated by the direct action of a flavin dependent monooxygenase (FMO). B) For chloramphenicol biosynthesis, a non-heme iron dependent dioxygenase hydroxylates the *p*-aminophenylalanine, instead of the hydroxylase used in GPA biosynthesis clusters.

A second mode of hydroxylation may involve *orf33*, or *staM*. CmlA is involved in chloramphenicol biosynthesis. It is coupled with the NRPS-like CmlP and CmlH and is believed to hydroxylate *p*-aminophenylalanine to give *p*-aminophenylserine (Figure 29) in a manner analogous to the OH-Tyr biosynthesis outlined in the introduction (75). No corresponding NRPS-like gene was found in the A47934 biosynthesis cluster. Both StaM and CmlA contain the same conserved $HxDx_{60}H$ iron-binding motif seen in non-heme

iron-dependent dioxygenases. It is possible that even in absence of the corresponding NRPS, StaM may hydroxylate Tyr. Possible substrates for StaM include the growing peptide which is still attached to the NRPS, the completed heptapeptide or free Tyr (analogous to HmaS in HPG production).

5.3 Creating the Heptapeptide Backbone

Assembly of the amino acids to form the heptapeptide backbone for A47934 requires the action of four NRPSs encoded by the A47934 biosynthesis cluster. Stretching a total of 28.5 kb of DNA, StaA, StaB, StaC and StaD are large, modular NRPS enzymes having predicted molecular weights of 219.1 kDa, 158.6 kDa, 433.8 kDa and 199.6 kDa respectively. The arrangement of the A, T, C, E and Te domains for the Sta enzymes on which they are located is diagrammed in Figure 30.

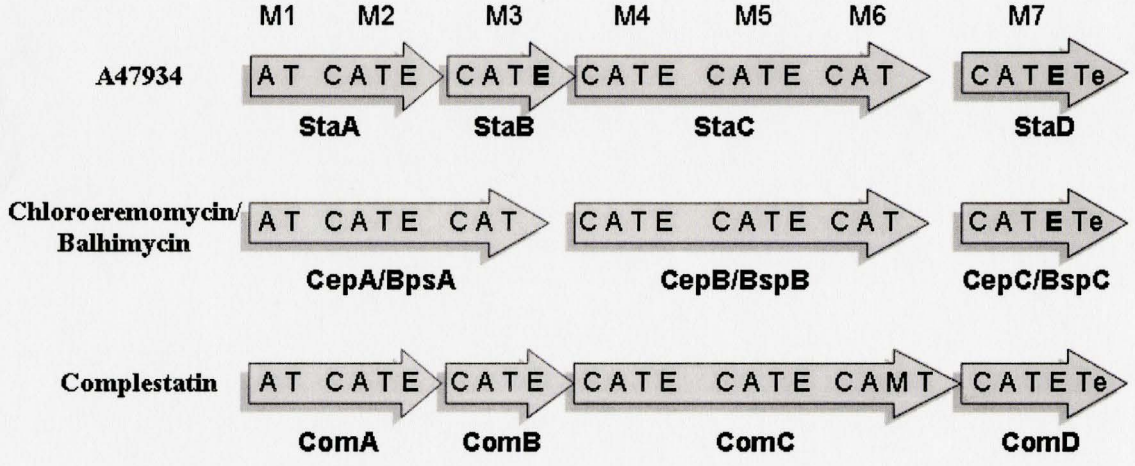


Figure 30 – Arrangement of Domains for StaABCD: The linear arrangement of the various domains for the A47934 biosynthesis cluster over the StaABCD enzymes is compared to the other GPA and GPA-like biosynthetic clusters. The E domains marked in bold are those which have likely lost their function as a result of natural mutations.

StaABCD contain a total of seven A domains, consistent with the production of a heptapeptide product. Using the binding pocket active site model established by Challis *et al*, the A domains were analyzed to give an indication of which amino acids they would specifically recognized (43). Figure 31 details the alignments of six of the A domains within the A47934 NRPS enzymes (for the A_6 domain see Figure 28). For each A domain, there is extremely high homology between the Sta domain and other established A domains. With one exception, these homologies are consistent with the expected order of amino acid incorporation for A47934. The one anomaly, activation of Tyr_6 by the third A domain on StaC, is outlined in the section 5.2.

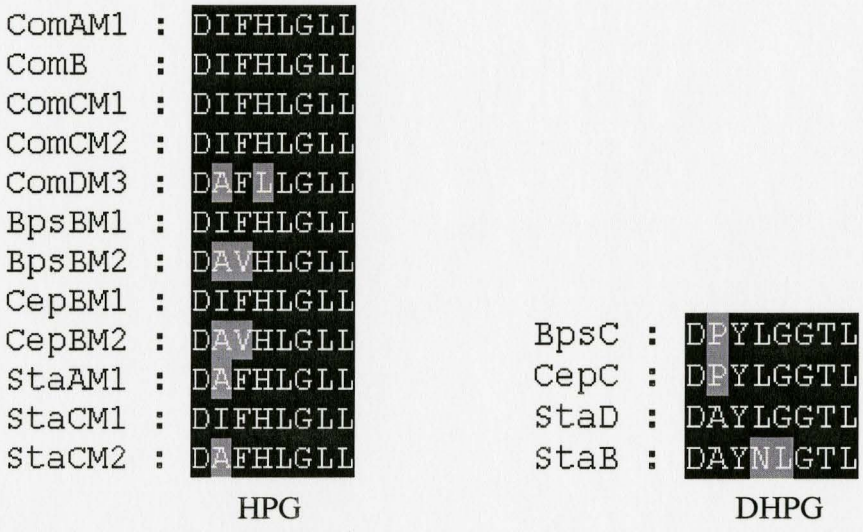


Figure 31 – Alignments of A Domain Recognition Sites: The A domain recognition sites are shown aligned with other corresponding GPA A domain recognition sites. There is high homology between all, thus all should activate the amino acid indicated. StaAM1 – StaA 1st A module; Com – complestatin, Bps – balhimycin, Cep – chloroeremomycin NRPSs. StaAM1 – amino acid 1; StaCM1 – amino acid 4; StaCM2 – amino acid 5; StaB – amino acid 2; StaD – amino acid 7.

The placement (or localization) of the A domains is distributed over four enzymes (Figure 30) giving a 2-1-3-1 arrangement. StaA initiates the heptapeptide with HPG_1, and then adds Tyr_2. StaB incorporates $DHPG_3$, while StaC incorporates HPG_4, HPG_5 and Tyr_6. Finally, StaD incorporates $DHPG_7$ and cleaves the linear heptapeptide. The 2-1-3-1 arrangement is analogous to the GPA-like complestatin arrangement, but varies from the 3-3-1 arrangement seen for both chloroeremomycin and balhimycin.

In addition to recognition of the appropriate amino acid, the A domain must also activate it by adenylation using ATP. The roles of the ten signature amino acid sequences for A domains are outlined in section 1.5.2. Figure 32 shows the alignment of the seven A47934 A domain signature sequences with the corresponding signature sequences. From this alignment, it is evident that all ten signature sequences are conserved in each of the A47934 NRPS A domains. Thus, the A47934 A domains act to specifically recognize an amino acid, bind the ATP substrate and possess the ATPase activity required for activation of the amino acid.

The C domains of the A47934 NRPS are found, ideally, located directly upstream of the A domains for peptide bond formation. There are six C domains, corresponding to the six peptide bonds formed in A47934 production. Alignment of the ~500 amino acid A47934 C domains revealed the presence of typical signature amino acids sequences of condensation domains (Figure 32). The main catalytic sequence (called the C3 motif or His motif) has the invariable His present in all cases. Of note is the lack of the C7 domain. Although thought to be conserved in C domains, the function of the C7 motif is unknown. The significance of its absence in the A47934 NRPSs is unknown. In the case

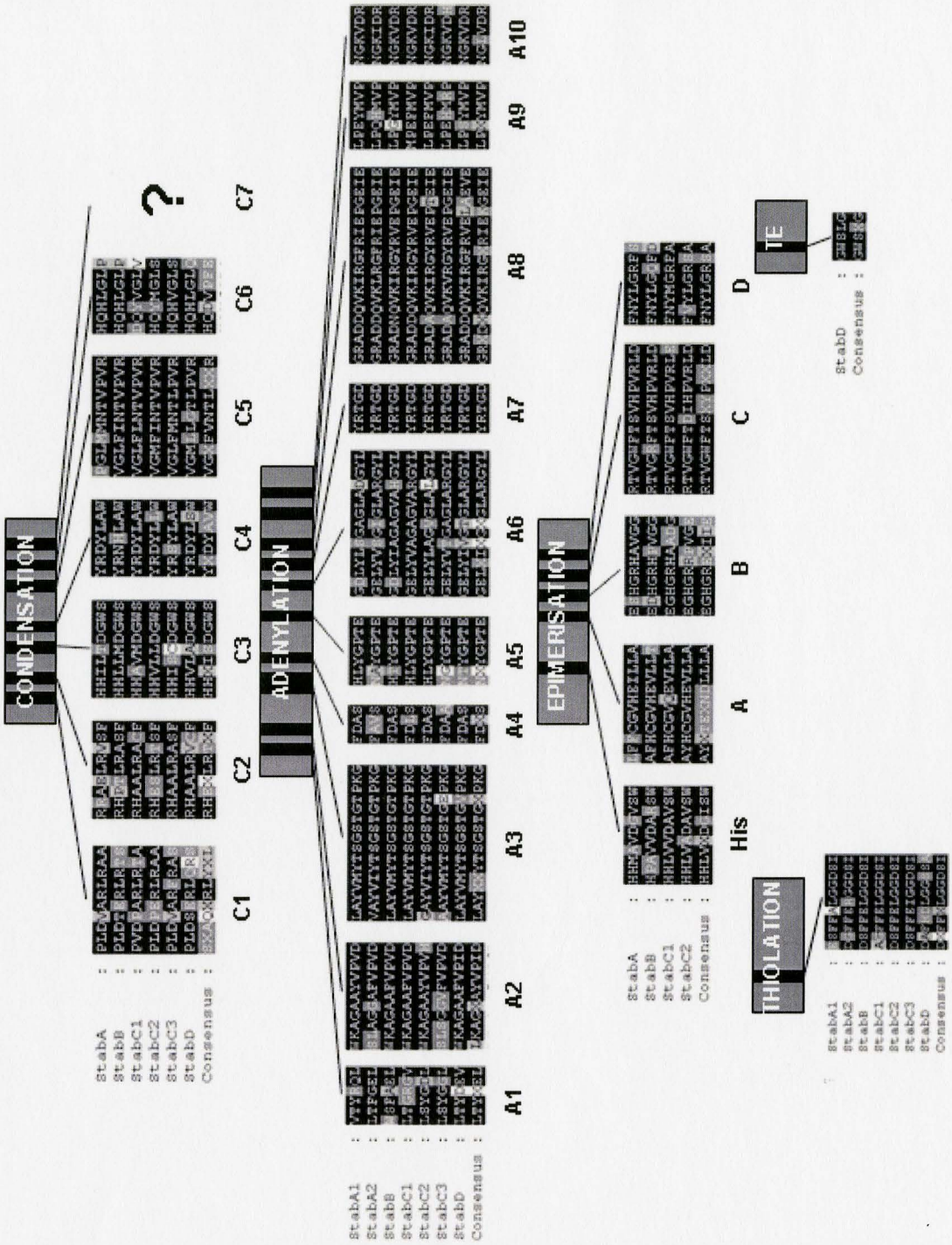


Figure 32 – Alignment of Consensus Sequences for NRPS Domains (previous page): The consensus sequences for each domain type is shown aligned with the corresponding Sta NRPS domain. Of note is the high homology in all cases except the lack of a C7 consensus sequence for the Sta enzymes and the mutation of the His domain in StaD. has not been established.

of A47934, the amino acid sequence of the C7 motif may be degenerate and was not recognized.

As seen in Figure 30, the T domains are appropriately arranged to link activated amino acids to the Sta enzymes. Amino acid composition of signature sequences for the T domains (Figure 32) is essentially invariable and in all cases the attachment site Ser is present.

The Te domain also demonstrates conservation of residues known to be important in this essential activity. The GxSxG serves as the site of attachment for the finished heptapeptide, where it is hydrolyzed for release.

Epimerisation domains are the final type of domain present in the A47934 NRPSs. They are responsible for stereochemical conversion of incorporated L-amino acids into the corresponding D enantiomer. A handful of structures for GPAs are known, including vancomycin, the teicoplanin-like ristocetin A and the teicoplanin aglycone A40946 (6, 7, 8). In all cases, the stereochemistry of the amino acids of the heptapeptide, regardless of the GPA type, is D-D-L-D-D-L-L from the N-terminus to the C-terminus.

Figure 2 shows the intricate H-bond network that is formed between the peptide backbone and the glycopeptide. In light of this, it is not surprising that stereochemistry of all GPAs is conserved. A change to the stereochemistry of even one amino acid could disrupt the interaction by moving the GPA backbone H-bond donor/acceptor out of H-bonding range with the peptidoglycan terminal D-Ala-D-Ala. Since A47934 shows similar activity to the known GPAs, it can be inferred that it too, along with all other GPAs, share the same stereochemistry. The A47934 NRPS enzymes contain four

putative E domains as outlined in Figure 32. High amino acid homology can be seen in the core Race A, B, C and D motifs, along with the catalytic His motif. Table 3 shows the location of the NRPS E domains of the three GPA biosynthesis clusters sequenced relative to the racemized amino acids. E domains are present to account for the epimerization of the amino acids D-(OH)-Tyr_2, D-HPG_4 and D-HPG_5 in the cases of balhimycin, chloroeremomycin and A47934. Missing in all cases, is an E domain corresponding to the amino terminal D-amino acid. Assumptions in the literature were originally made that the A_1 domain in these clusters would specifically recognize the previously racemized D-amino acid. Recently, *in vitro* activity of the recombinant chloroeremomycin CepA A_1 domain was shown to favorably incorporate L-Leu over D-Leu (76). The activity was quite low for both substrates, but these results indicate that L-Leu is actually incorporated then racemized downstream. Similar to the vancomycin class systems, the E domain for racemization of the amino terminal D-HPG is absent in StaA. It is unclear whether the StaA A_1 domain recognizes L-HPG or D-HPG. The other HPG residues found in GPAs are incorporated first as L-HPG then isomerized via the corresponding E domain. Alignment of the StaA A_1 binding pocket shows it has high homology to all other GPA HPG binding pockets and is identical to the StaC A_5 HPG binding pocket. This would indicate it preferentially incorporates L-HPG. Also of note, is the fact that *in vitro* characterization of HPG biosynthesis showed that L-HPG is the stereospecific product (30). Coupled with the results showing L-Leu incorporation by CepA, it appears that L-HPG is specifically incorporated by StaA and subsequently racemized.

Table 3 – Presence of E domains relative to racemized amino acids:

GPA	D	D	L	D	D	L	L
Chloroeremomycin		✓		✓	✓		
Balhimycin		✓		✓	✓		
A47934		✓	✓	✓	✓		

✓ - E domain present

A unique issue to the A47934 NRPSs is the extra E domain present which, hypothetically, would racemize $DHPG_3$. As there is no known example of a D-amino acid at the third position, the presence of this E domain seems unusual. Closer inspection of the E domain alignment (Figure 32) reveals that the mechanism of action of E domains is similar to that of the C domains (see Introduction 1.5.2). Conserved, in both cases, is the same His motif. Mutational studies involving the second His for both domain types, show that this residue is essential for catalytic activity. Careful inspection of the alignments of conserved regions shows that the E domain corresponding to the third amino acid position has a natural mutation of His to Pro. The region containing the mutation has been sequenced five times, on different subclones, and can be considered very reliable. The presence of a Pro at this position could abolish activity of this E domain, leaving A47934 with the same distribution of active E domains as the two vancomycin-like NRPS clusters. Speculative questions can be raised about the presence of such a mutated domain. Was A47934 originally produced under different stereochemistry? If so, was it active as a GPA in that state? Did A47934 originally serve a different function other than a GPA? And does it still?

The final domain within the A47934 NRPSs of note, is an uncharacterized domain displaying homology to an E or C domain. Located between the final T domain and the Te domain (Figure 30) of StaD, this domain is found in all four sequenced GPA and GPA-like clusters. The region contains motifs with low homology to (in order) His, C3, C4, E4, E5 and C5. The His domain is mutated such that the catalytic HH residues are now HR. Based on the high level of degeneracy found, this domain could be an artifact of the evolutionary process. Normally, there is not a large gap before the Te domain, hence it is likely that this region has been rendered inactive over time. Based on positioning, it most likely served as an E domain rather than a C domain although, whether it ever served any function remains unknown.

5.4 Crosslinking the A47934 Backbone

Once a linear heptapeptide is produced, extensive modifications take place to produce a complete GPA. One such set of modifications involves the crosslinking of the amino acid side chains of the GPA backbone. For each of the three GPA and the GPA-like biosynthesis clusters sequenced there are enzymes with high homology to cytochrome P450 monooxygenases. These would be capable of performing crosslinking reactions. Found clustered together, there are always the same number of oxygenases present as there are crosslinkages in the corresponding GPA. In the A47934 biosynthesis cluster, four such oxygenases are seen clustered together in *orfs* 26, 27, 28 and 30. Presumably, each associated enzyme is responsible for one of the four crosslinks seen in A47934. A series of gene inactivation experiments by Bischoff *et al* on the *oxyABC* genes in the balhimycin producer *A. mediterranei* has revealed the exact cross-linkages

catalyzed by each gene products (63, 64). Based on their findings, OxyA catalyzes the ether cross-linkage of OH-Tyr_2 and HPG_4 , OxyB catalyzes the ether linkage between HPG_4 and OH-Tyr_6 and OxyC performs the carbon-carbon linkage between HPG_5 and $DHPG_7$. Table 4 outlines the amino acid homologies between the oxygenases present in the four sequenced GPA and GPA-like biosynthesis clusters. Each oxygenase characterized in *A. mediterranei* has one orthologue in each of the other clusters for which there is a similar crosslinkage. It is reasonable to conclude that each set of orthologous oxygenases performs the same crosslinkage reaction. A47934 has one unique crosslinkage found only in teicoplanin-like GPAs. Therefore, the one oxygenase without a corresponding orthologue, StaG, is predicted to perform this unique ether cross-linkage between amino acids 1 and 3. Thus, for A47934 biosynthesis, StaF should cross-link the amino acids Tyr_2 and HPG_4, StaG should crosslink HPG_1 and $DHPG_3$, StaH cross-links HPG_4 and OH-Tyr_6 and StaJ cross-links HPG_5 and $DHPG_7$ (Figure 33).

Table 4 – Percent amino acid similarity between crosslinking enzymes of GPA and GPA-like biosynthesis clusters:

	A47934				Chloroeremomycin			Balhimycin			Complestatin	
	StaF	**StaG**	**StaH**	**StaJ**	**CepE**	**CepF**	**CepG**	**OxyA**	**OxyB**	**OxyC**	**ComI**	**ComJ**
StaF	100	62	54	52	**82**	53	53	**82**	52	53	57	52
StaG		100	56	53	60	56	54	59	57	54	64	52
StaH			100	**67**	55	**84**	64	53	**85**	63	60	**71**
StaJ				100	53	65	**84**	55	65	**82**	58	68
CepE					100	57	54	**92**	56	54	58	53
CepF						100	**63**	56	**92**	63	62	**69**
CepG							100	54	63	**94**	58	68
OxyA								100	55	55	58	53
OxyB									100	**63**	61	**69**
OxyC										100	58	68
ComI											100	59
ComJ												100

5.5 Enzymatic Modifications to A47934

GPAs are also modified by the covalent modification of the amino acid constituents of the heptapeptide backbone. A47934 contains two such modifications: chlorinations and a unique sulfonylation. There are no sites of methylation or glycosylation in the case of A47934.

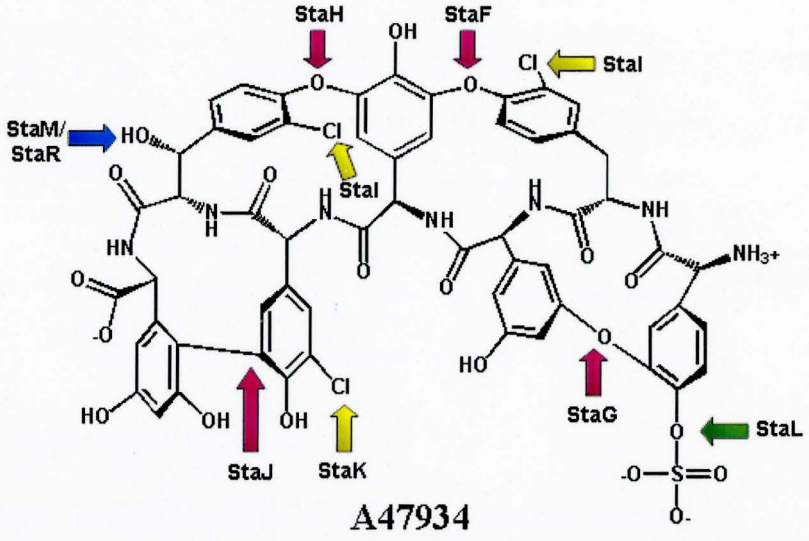


Figure 33 – Sites of Tailoring and Modifications on A47934: A47934 undergoes several crosslinking reactions, proposed to be carried out by StaF, StaG, StaH and StaJ. Halogenation reactions are catalyzed by StaI and StaK. StaL is the putative sulfotransferase with sulfonylates the N-terminal HPG.

As with the crosslinking enzymes, each previously sequenced cluster contains one enzyme with high homology to known non-heme halogenases for each type of chlorination present. Similar to that proposed in Section 5.4, the functions of the two halogenase enzymes in the A47934 biosynthesis cluster (StaI and StaK) can be predicted. StaI is highly homologous to CepH from chloroeremomycin (88% amino acid similarity) and BhaA from balhimycin (88%), which are also 97% homologous to each other. These

similar halogenases likely all perform a similar halogenation reaction, namely the chlorination of the Tyr/OH-Tyr (Figure 33). This function has been characterized by gene inactivation of *bhaA* (65). StaK is then left to chlorinate the HPG. Interestingly ComH, the putative halogenase from the complestatin producer, has higher homology to StaI than StaK, which is the opposite of what one would expect. ComH appears to be responsible for two halogenations as all HPG residues are 3,5-dichlorinated.

Unique to A47934 is the presence of a sulfate group on HPG_1. This would putatively be catalyzed by the sulfotransferase StaL (Figure 33). StaL shows homology to known sulfotransferases and, interestingly, has much higher homology to eukaryotic sulfotransferases than prokaryotic ones. Alignments of StaL with known eukaryotic sulfotransferases and with *Macaca fascicularis* for which the structure has been solved (77), show that conserved are the binding sites for both of the 3' and 5' phosphate groups of PAPS, as well as, the invariant Lys (K12) and His (H67) residues that act as the catalytic general acid and base respectively.

5.6 Resistance to A47934 in *S. toyocaensis*

Another unique feature of the A47934 biosynthesis cluster is the presence of genes involved in GPA to resistance and its regulation. The occurrence of the genes *vanHst*, *vanAst* and *vanXst* represent the first time resistance genes have been coupled to the biosynthesis genes. These genes are both necessary and sufficient for host resistance to A47934 by conversion of D-Ala-D-Ala in the peptidoglycan layer to D-Ala-D-Lac. Insertional inactivation of the *vanAst* gene (completed by Dr. John Neu) showed sensitivity to A47934. Resistance could be restored by supplying the *vanHAX* genes from

A. orientalis (chloroeremomycin producer) in *trans*. This shows that the *vanHAXst* genes clustered with the biosynthesis genes are involved in GPA resistance in a manner analogous to the VanA phenotype.

The two component regulatory system VanR, response regulator, and VanS, sensor kinase, are responsible for regulation of the resistance genes. VanRst and VanSst from *S. toyocaensis* show high amino acid homology to the two component regulatory system from *S. coelicolor* (73% identity to VanRsc and 59% identity to VanSsc). The *vanRSsc* genes were shown, through gene inactivation studies, to be involved in resistance regulation (personal communication - Dr. John Neu). Furthermore, a consensus promoter region has been identified upstream of *vanHst* and *vanHsc* homologous to that identified in *Enterococcus faecium* (Figure 34) (78).

Consensus	T	T/C	T	T	A	G/A	G	A	A	A	T/A	T
vanHsc	G	A	T	A	T	C	G	A	A	A	A	G
vanHst	C	A	T	A	T	G	G	A	A	A	A	A
staT	A	A	G	C	G	A	C	A	A	A	A	G

Figure 34 – Promoter Homologies for *sta* Genes: The putative promoter regions in front of the *vanH* genes from *S. coelicolor* and *S. toyocaensis* show high homology with the consensus van promoter sequenced determined for *E. faecium*. The promoter region in front of *staT* shows lower homology.

This is also the first example in which the v*anRS* two component system is not found directly coupled to the *vanHAX* genes. These sets of genes are separated by approximately 20kb. Comparisons of the locations of the *vanRS* genes in various clusters is illustrated in Figure 35. Of interest, is the appearance of two membrane proteins, StaO and StaP, which are also found clustered with *van* genes in *S. coelicolor*. StaO shows

high homology to the FemABX family of enzymes that have been implicated in the formation of the pentaglycine bridge in peptidoglycan formation (79). There is the possibility that conserved membrane proteins clustered directly with the resistance genes may play a part in the signaling mechanism.

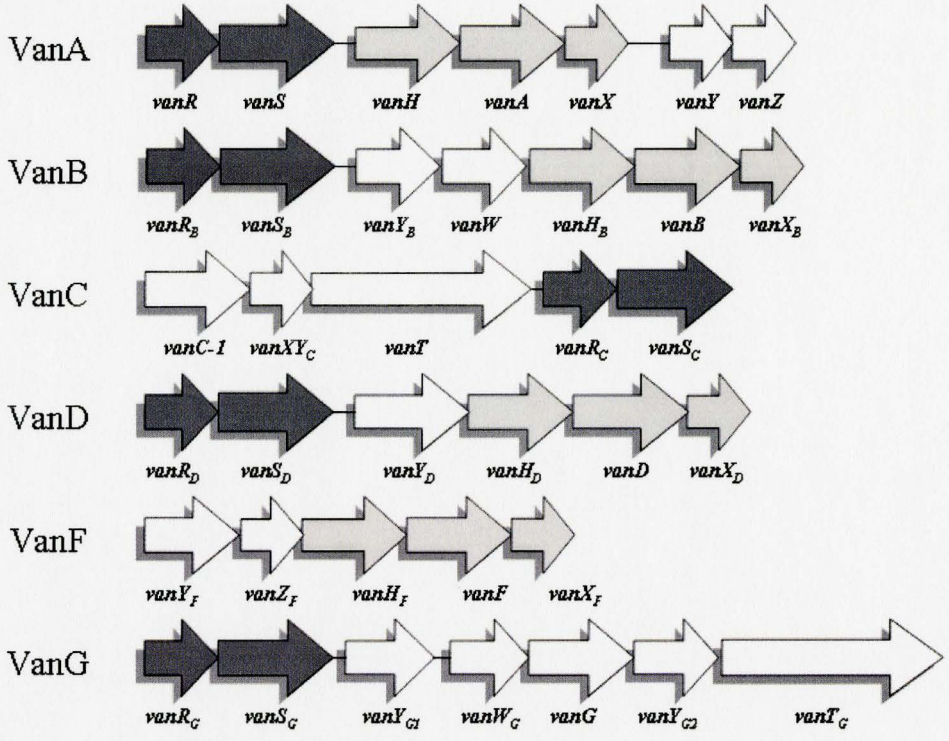


Figure 35 – Van Phenotypes: For each of the van phenotypes, there is close spacing of the *vanRS* (regulatory genes) and the resistance genes. The A47934 biosynthesis cluster is the first case where the regulatory genes are not located directly with the resistance genes (Figure 23).

Results of enzymolgy research on VanR and VanS from both *S. toyocaensis* and *S. coelicolor* are not encouraging. Failure to clone both *vanRst* and *vanSst* together was a result of unfortunate rearrangements by the host strain. This was determined through restriction enzyme mapping of the construct and confirmed by sequencing the insert. Despite using the rec$^-$ *E. coli* SURE2 host strain, recombination was unavoidable. There

were toxicity problems when the two genes were cloned together since individual cloning of the genes was possible. Such a problem is a good example of obstacles faced in all aspects of this project. Enzymatic activity of VanSst and VanSsc has yet to be characterized, as fully activate enzyme was never obtained. Extensive work on VanSst and VanSsc was done using reaction conditions shown to work for the positive control VanSen. Changing buffer systems did not improve the activity. In the end, only minimal autophosphorylation activity was shown for VanSst, while none was able to show for VanSsc. Further investigation being carried out in our lab by Dr. Xiao-Dong Wang on these two enzymes including using different over expression conditions, purification protocols, fusion proteins and truncations had yet to yield active enzyme. Since active VanSst was unavailable, no enzymology was carried out on VanRst or VanRsc.

5.7 Other Genes Encoded in the A47934 Biosynthesis Cluster

The final genes of note are *staU*, *staN*, *staQ* and *staS*. *staU* encodes a putative ABC-type transporter protein and *staN* encodes a putative ion antiporter, both of which are found in each of the biosynthesis clusters sequenced to date. Presumably, their purpose is to export the completed antibiotic into the extracellular environment. This might require both, or one of these enzymes. StaQ is a predicted transcriptional regulator found upstream of the biosynthesis genes. There is an orthologous protein to StaQ upstream of the complestatin cluster. Since neither the balhimycin nor chloroeremomycin clusters have any upstream sequencing, this transcriptional regulator could very well be present in those two systems as well. It is possible that this transcriptional regulator may

play a part in regulation of antibiotic production. Finally, *staU* encodes a putative DNA-binding protein of unknown function.

5.8 Conclusion

The sequencing of the A47934 biosynthesis cluster from *S. toyocaensis* NRRL 15009 represents the most complete glycopeptide biosynthesis cluster to date. The A47934 cluster is unique in two aspects; it is the first teicoplanin-like GPA biosynthetic cluster sequenced and it is the first case where an entire biosynthesis cluster has been linked to the associated resistance genes. Using genetic and enzymatic information present for orthologues of the enzymes encoded within the A47934 biosynthetic cluster, putative functions for the vast majority of the enzymes encoded can be assigned. This information allows us to see how A47934 is assembled and how self-resistance is gained and regulated.

The information provided by the A47934 biosynthesis cluster allows us further insight into the overall mechanism of GPA production. Comparisons and contrasts can be drawn to delve into *in vivo* characteristics of the producing organisms, which can be exploited in trying to create novel GPAs with increased efficiency. The A47934 biosynthesis cluster shows us the variations involved in the GPA classes and how modification changes, such as the lack of sugar residues or the presence of sulfate groups, are addressed by the producing bacteria. In the continuing struggle against antibiotic resistant bacteria, information on antibiotic production, resistance and regulation has been, and will continue to be, of the utmost importance in finding a way to combat these harmful bacteria.

6. References:

1) Pootoolal J, Neu J, Wright GD. Glycopeptide antibiotic resistance. Annu Rev Pharmacol Toxicol. 2002;42:381-408.

2) Leclercq R, Derlot E, Duval J, Courvalin P. Plasmid-mediated resistance to vancomycin and teicoplanin in *Enterococcus faecium*. N Engl J Med. 1988 Jul 21; 319(3):157-61.

3) Kirst HA, Thompson DG, Nicas TI. Historical yearly usage of vancomycin. Antimicrob Agents Chemother. 1998 May;42(5):1303-4.

4) Sieradzki K, Tomasz A. Gradual alterations in cell wall structure and metabolism in vancomycin-resistant mutants of *Staphylococcus aureus*. J Bacteriol. 1999 Dec;181(24): 7566-70.

5) Novak R, Henriques B, Charpentier E, Normark S, Tuomanen E. Emergence of vancomycin tolerance in *Streptococcus pneumoniae*. Nature. 1999 Jun 10;399(6736): 590-3.

6) Schafer M, Schneider TR, Sheldrick GM. Crystal structure of vancomycin. Structure. 1996 Dec 15;4(12):1509-15.

7) Barna JC, Williams DH. The structure and mode of action of glycopeptide antibiotics of the vancomycin group. Annu Rev Microbiol. 1984;38:339-57.

8) Schafer M, Pohl E, Schmidt-Base K, Sheldrick GM, Hermann R, Malabarba A, Nebuloni M, Pelizzi C. The molecular and crystal structure of the glycopeptide A-40926 aglycone. Helvetica Chima Acta. 1996;76:1916-23.

9) Schafer M, Sheldrick GM, Schneider TR, Vertesy L. Structure of balhimycin and its complex with solvent molecules. Acta Crystallogr D Biol Crystallogr. 1998 Mar 1;54 (Pt 2):175-83.

10) Walsh CT, Fisher SL, Park IS, Prahalad M, Wu Z. Bacterial resistance to vancomycin: five genes and one missing hydrogen bond tell the story. Chem Biol. 1996 Jan;3(1):21-8.

11) McPhail D, Cooper A, Freer A. Crystallization and preliminary X-ray crystallographic analysis of a vancomycin-N-acetyl-D-Ala-D-Ala complex. Acta

Crystallogr D Biol Crystallogr. 1999 Feb;55 (Pt 2):534-5.

12) Williams DH, Maguire AJ, Tsuzuki W, Westwell MS. An analysis of the origins of a cooperative binding energy of dimerization. Science. 1998 May 1;280(5364):711-4.

13) Staroske T, O'Brien DP, Jorgensen TJ, Roepstorff P, Williams DH, Heck AJ. The formation of heterodimers by vancomycin group antibiotics. Chemistry. 2000 Feb 4;6(3): 504-9.

14) Marshall CG, Lessard IA, Park I, Wright GD. Glycopeptide antibiotic resistance genes in glycopeptide-producing organisms. Antimicrob Agents Chemother. 1998 Sep; 42(9):2215-20.

15) Arthur M, Molinas C, Depardieu F, Courvalin P. Characterization of Tn1546, a Tn3-related transposon conferring glycopeptide resistance by synthesis of depsipeptide peptidoglycan precursors in *Enterococcus faecium* BM4147. J Bacteriol. 1993 Jan; 175(1):117-27.

16) Bugg TD, Wright GD, Dutka-Malen S, Arthur M, Courvalin P, Walsh CT. Molecular basis for vancomycin resistance in *Enterococcus faecium* BM4147: biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA. Biochemistry. 1991 Oct 29;30(43):10408-15.

17) Arthur M, Molinas C, Bugg TD, Wright GD, Walsh CT, Courvalin P. Evidence for in vivo incorporation of D-lactate into peptidoglycan precursors of vancomycin-resistant *Enterococci*. Antimicrob Agents Chemother. 1992 Apr;36(4):867-9.

18) Wu Z, Wright GD, Walsh CT. Overexpression, purification, and characterization of VanX, a D-, D-dipeptidase which is essential for vancomycin resistance in *Enterococcus faecium* BM4147. Biochemistry. 1995 Feb 28;34(8):2455-63.

19) Arthur M, Molinas C, Courvalin P. The VanS-VanR two-component regulatory system controls synthesis of depsipeptide peptidoglycan precursors in *Enterococcus faecium* BM4147. J Bacteriol. 1992 Apr;174(8):2582-91.

20) Wright GD, Holman TR, Walsh CT. Purification and characterization of VanR and the cytosolic domain of VanS: a two-component regulatory system required for vancomycin resistance in *Enterococcus faecium* BM4147. Biochemistry. 1993 May 18; 32(19):5057-63.

21) Haldimann A, Fisher SL, Daniels LL, Walsh CT, Wanner BL. Transcriptional regulation of the *Enterococcus faecium* BM4147 vancomycin resistance gene cluster by

the VanS-VanR two-component regulatory system in Escherichia coli K-12. J Bacteriol. 1997 Sep;179(18):5903-13.

22) Perichon B, Casadewall B, Reynolds P, Courvalin P. Glycopeptide-resistant *Enterococcus faecium* BM4416 is a VanD-type strain with an impaired D-Alanine:D-Alanine ligase. Antimicrob Agents Chemother. 2000 May;44(5):1346-8.

23) Boyd DA, Conly J, Dedier H, Peters G, Robertson L, Slater E, Mulvey MR. Molecular characterization of the vanD gene cluster and a novel insertion element in a vancomycin-resistant *Enterococcus* isolated in Canada. J Clin Microbiol. 2000 Jun;38(6): 2392-4.

24) Arias CA, Courvalin P, Reynolds PE. vanC cluster of vancomycin-resistant *Enterococcus gallinarum* BM4174. Antimicrob Agents Chemother. 2000 Jun;44(6): 1660-6.

25) Fines M, Perichon B, Reynolds P, Sahm DF, Courvalin P. VanE, a new type of acquired glycopeptide resistance in *Enterococcus faecalis* BM4405. Antimicrob Agents Chemother. 1999 Sep;43(9):2161-4.

26) McKessar SJ, Berry AM, Bell JM, Turnidge JD, Paton JC. Genetic characterization of vanG, a novel vancomycin resistance locus of *Enterococcus faecalis*. Antimicrob Agents Chemother. 2000 Nov;44(11):3224-8.

27) van Wageningen AM, Kirkpatrick PN, Williams DH, Harris BR, Kershaw JK, Lennard NJ, Jones M, Jones SJ, Solenberg PJ. Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. Chem Biol. 1998 Mar;5(3):155-62.

28) Recktenwald J, Shawky R, Puk O, Pfennig F, Keller U, Wohlleben W, Pelzer S. Nonribosomal biosynthesis of vancomycin-type antibiotics: a heptapeptide backbone and eight peptide synthetase modules. Microbiology. 2002 Apr;148(Pt 4):1105-18.

29) Chiu HT, Hubbard BK, Shah AN, Eide J, Fredenburg RA, Walsh CT, Khosla C. Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. Proc Natl Acad Sci U S A. 2001 Jul 17;98(15):8548-53.

30) Hubbard BK, Thomas MG, Walsh CT. Biosynthesis of L-p-hydroxyphenylglycine, a non-proteinogenic amino acid constituent of peptide antibiotics. Chem Biol. 2000 Dec;7(12):931-42.

31) Pfeifer V, Nicholson GJ, Ries J, Recktenwald J, Schefer AB, Shawky RM, Schroder J, Wohlleben W, Pelzer S. A polyketide synthase in glycopeptide biosynthesis: the

biosynthesis of the non-proteinogenic amino acid (S)-3,5-dihydroxyphenylglycine. J Biol Chem. 2001 Oct 19;276(42):38370-7.

32) Chen H, Tseng CC, Hubbard BK, Walsh CT. Glycopeptide antibiotic biosynthesis: enzymatic assembly of the dedicated amino acid monomer (S)-3,5-dihydroxyphenylglycine. Proc Natl Acad Sci U S A. 2001 Dec 18;98(26):14901-6.

33) Chen H, Walsh CT. Coumarin formation in novobiocin biosynthesis: beta-hydroxylation of the aminoacyl enzyme tyrosyl-S-NovH by a cytochrome P450 NovI. Chem Biol. 2001 Apr;8(4):301-12.

34) Chen H, Thomas MG, O'Connor SE, Hubbard BK, Burkart MD, Walsh CT. Aminoacyl-S-enzyme intermediates in beta-hydroxylations and alpha,beta-desaturations of amino acids in peptide antibiotics. Biochemistry. 2001 Oct 2;40(39):11651-9.

35) Lipmann F. Attempts to map a process evolution of peptide biosynthesis. Science. 1971 Sep 3;173(4000):875-84.

36) Marahiel MA. Protein templates for the biosynthesis of peptide antibiotics. Chem Biol. 1997 Aug;4(8):561-7.

37) Conti E, Stachelhaus T, Marahiel MA, Brick P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. EMBO J. 1997 Jul 16;16(14):4174-83.

38) Weber T, Marahiel MA. Exploring the domain structure of modular nonribosomal peptide synthetases. Structure (Camb). 2001 Jan 10;9(1):R3-9.

39) Stachelhaus T, Mootz HD, Marahiel MA. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem Biol. 1999 Aug;6(8) 493-505.

40) Kleinkauf H, Von Dohren H. A nonribosomal system of peptide biosynthesis. Eur J Biochem. 1996 Mar 1;236(2):335-51.

41) Stachelhaus T, Marahiel MA. Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. FEMS Microbiol Lett. 1995 Jan 1;125(1):3-14.

42) Pavela-Vrancic M, Pfeifer E, van Liempt H, Schafer HJ, von Dohren H, Kleinkauf H. ATP binding in peptide synthetases: determination of contact sites of the adenine moiety by photoaffinity labeling of tyrocidine synthetase 1 with 2-azidoadenosine triphosphate.

Biochemistry. 1994 May 24;33(20):6276-83.

43) Challis GL, Ravel J, Townsend CA. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem Biol. 2000 Mar;7(3):211-24.

44) Schlumbohm W, Stein T, Ullrich C, Vater J, Krause M, Marahiel MA, Kruft V, Wittmann-Liebold B. An active serine is involved in covalent substrate amino acid binding at each reaction center of gramicidin S synthetase. J Biol Chem. 1991 Dec 5;266(34):23135-41.

45) Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, Reid R, Khosla C, Walsh CT. A new enzyme superfamily - the phosphopantetheinyl transferases. Chem Biol. 1996 Nov;3(11):923-36.

46) Parris KD, Lin L, Tam A, Mathew R, Hixon J, Stahl M, Fritz CC, Seehra J, Somers WS. Crystal structures of substrate binding to *Bacillus subtilis* holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. Structure Fold Des. 2000 Aug 15;8(8):883-95.

47) Price AC, Zhang YM, Rock CO, White SW. Structure of beta-ketoacyl-[acyl carrier protein] reductase from Escherichia coli: negative cooperativity and its structural basis. Biochemistry. 2001 Oct 30;40(43):12772-81.

48) Stachelhaus T, Huser A, Marahiel MA. Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. Chem Biol. 1996 Nov;3(11):913-21.

49) Ku J, Mirmira RG, Liu L, Santi DV. Expression of a functional non-ribosomal peptide synthetase module in Escherichia coli by coexpression with a phosphopantetheinyl transferase. Chem Biol. 1997 Mar;4(3):203-7.

50) Stachelhaus T, Mootz HD, Bergendahl V, Marahiel MA. Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. J Biol Chem. 1998 Aug 28;273(35):22773-81.

51) Ehmann DE, Trauger JW, Stachelhaus T, Walsh CT. Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. Chem Biol. 2000 Oct;7(10):765-72.

52) Belshaw PJ, Walsh CT, Stachelhaus T. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. Science. 1999 Apr 16;284(5413):486-9.

53) Linne U, Marahiel MA. Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. Biochemistry. 2000 Aug 29;39(34):10439-47.

54) Doekel S, Marahiel MA. Dipeptide formation on engineered hybrid peptide synthetases. Chem Biol. 2000 Jun;7(6):373-84.

55) Mootz HD, Schwarzer D, Marahiel MA. Construction of hybrid peptide synthetases by module and domain fusions. Proc Natl Acad Sci U S A. 2000 May 23;97(11):5848-53.

56) Shaw-Reid CA, Kelleher NL, Losey HC, Gehring AM, Berg C, Walsh CT. Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of E. coli EntF catalyzes both elongation and cyclolactonization. Chem Biol. 1999 Jun;6(6):385-400.

57) Kohli RM, Trauger JW, Schwarzer D, Marahiel MA, Walsh CT. Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. Biochemistry. 2001 Jun 19;40(24):7099-108.

58) Trauger JW, Kohli RM, Walsh CT. Cyclization of backbone-substituted peptides catalyzed by the thioesterase domain from the tyrocidine nonribosomal peptide synthetase. Biochemistry. 2001 Jun 19;40(24):7092-8.

59) Schwarzer D, Mootz HD, Marahiel MA. Exploring the impact of different thioesterase domains for the design of hybrid peptide synthetases. Chem Biol. 2001 Oct;8(10):997-1010.

60) Stachelhaus T, Walsh CT. Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. Biochemistry. 2000 May 16;39(19):5775-87.

61) Patel HM, Walsh CT. In vitro reconstitution of the Pseudomonas aeruginosa nonribosomal peptide synthesis of pyochelin: characterization of backbone tailoring thiazoline reductase and N-methyltransferase activities. Biochemistry. 2001 Jul 31;40(30):9023-31.

62) Walsh CT. Enzymatic Reaction Mechanisms. W. H. Freeman and Company. New York, New York. 1979.

63) Bischoff D, Pelzer S, Holtzel A, Nicholson GJ, Stockert S, Wohlleben W, Jung G, Sussmuth RD. The Biosynthesis of Vancomycin-Type Glycopeptide Antibiotics-New

Insights into the Cyclization Steps. Angew Chem Int Ed Engl. 2001 May 4;40(9):1693-1696.

64) Bischoff D, Pelzer S, Bister B, Nicholson GJ, Stockert S, Schirle S, Wohlleben W, Jung G, Sussmuth RD. The Biosynthesis of Vancomycin-Type Glycopeptide Antibiotics-The Order of the Cyclization Steps. Angew Chem Int Ed Engl. 2001 May 4;40(24):4688-4691.

65) Puk O, Huber P, Bischoff D, Recktenwald J, Jung G, Sussmuth RD, van Pee KH, Wohlleben W, Pelzer S. Glycopeptide Biosynthesis in *Amycolatopsis mediterranei* DSM5908. Function of a Halogenase and a Haloperoxidase/Perhydrolase. Chem Biol. 2002 Feb;9(2):225-35.

66) Falany CN. Sulfation and sulfotransferases. Introduction: changing view of sulfation and the cytosolic sulfotransferases. FASEB J. 1997 Jan;11(1):1-2.

67) Kehoe JW, Bertozzi CR. Tyrosine sulfation: a modulator of extracellular protein-protein interactions. Chem Biol. 2000 Mar;7(3):R57-61.

68) Klaassen CD, Boles JW. Sulfation and sulfotransferases 5: the importance of 3'-phosphoadenosine 5'-phosphosulfate (PAPS) in the regulation of sulfation. FASEB J. 1997 May;11(6):404-18.

69) Matsushima P, Baltz RH. A gene cloning system for *'Streptomyces toyocaensis'*. Microbiology. 1996 Feb;142 (Pt 2):261-7.

70) Marshall CG. Glycopeptide Antibiotic Biosynthesis and Resistance in *Streptomyces toyocaensis* NRRL 15009. Ph. D. Thesis. McMaster University. 1999.

71) Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. Practical Streptomyces Genetics. The John Innes Foundation. Norwich, England. 2000.

72) Dym O, Eisenberg D. Sequence-structure analysis of FAD-containing proteins. Protein Sci. 2001 Sep;10(9):1712-28.

73) Blanc V, Lagneaux D, Didier P, Gil P, Lacroix P, Crouzet J. Cloning and analysis of structural genes from *Streptomyces pristinaespiralis* encoding enzymes involved in the conversion of pristinamycin IIB to pristinamycin IIA (PIIA): PIIA synthase and NADH:riboflavin 5'-phosphate oxidoreductase. J Bacteriol. 1995 Sep;177(18):5206-14.

74) Thibaut D, Ratet N, Bisch D, Faucher D, Debussche L, Blanche F. Purification of the two-enzyme system catalyzing the oxidation of the D-proline residue of pristinamycin IIB

during the last step of pristinamycin IIA biosynthesis. J Bacteriol. 1995 Sep;177(18): 5199-205.

75) He J, Magarvey N, Piraee M, Vining LC. The gene cluster for chloramphenicol biosynthesis in *Streptomyces venezuelae* ISP5230 includes novel shikimate pathway homologues and a monomodular non-ribosomal peptide synthetase gene. Microbiology. 2001 Oct;147(Pt 10):2817-29.

76) Trauger JW, Walsh CT. Heterologous expression in *Escherichia coli* of the first module of the nonribosomal peptide synthetase for chloroeremomycin, a vancomycin-type glycopeptide antibiotic. Proc Natl Acad Sci U S A. 2000 Mar 28;97(7):3112-7.

77) Yoshinari K, Petrotchenko EV, Pedersen LC, Negishi M. Crystal structure-based studies of cytosolic sulfotransferase. J Biochem Mol Toxicol. 2001;15(2):67-75.

78) Holman TR, Wu Z, Wanner BL, Walsh CT. Identification of the DNA-binding site for the phosphorylated VanR protein required for vancomycin resistance in *Enterococcus faecium*. Biochemistry. 1994 Apr 19;33(15):4625-31.

79) Hegde SS, Shrader TE. FemABX family members are novel nonribosomal peptidyltransferases and important pathogen-specific drug targets. J Biol Chem. 2001 Mar 9;276(10):6998-7003.

Appendix A – Strains Which Produce GPA or GPA-like Molecules

Strain	GPA or GPA-like Molecule Produced	Biosynthesis Cluster Sequenced
Streptomyces toyocaensis NRRL 15009	A47934	Yes
Amycolatopsis orientalis NRRL 18098	Chloroeremomycin	Yes
Amycolatopsis mediterranei	Balhimycin	Yes
Streptomyces lavenulae	Complestatin	Yes
Amycolatopsis orientalis C329.2	Vancomycin	No
Amycolatopsis coloradensis	Teicoplanin	Partial

Appendix B – Primers Referenced Within Thesis

Primer Reference	Sequence	Use
M13 Forward	5'-ACT GGC CGT CTT TTA C-3'	Sequencing primer
M13 Reverse	5'-CAG GAA ACA GCT ATG AC-3'	Sequencing primer
AB9363	5'-GAG ATA TAC ATA TGG CCA GAC TGA AGA TCG G-3'	*vanAst* amplification
AB9364	5'-TGA CAT AGC TTC AGA GCG AGG AGA CGG AGA-3'	*vanAst* amplification
T3 Promoter Primer	5'-ATT AA CCC TCA CTA AAG GG-3'	Sequencing primer
T7 Promoter Primer	5'-CCC TAT AGT GAG TCG TAT TA-3'	Sequencing primer
AB18785	5'-GGA GCT CTA GAT CAC CTG TTG AGG GGT TTC G-3'	*vanRSst* amplification
AB18786	5'-GAT TGC AAG CTT GAA TTC CCC TCT CCC ACG ACT CTG AT-3'	*vanRSst* amplification
AB18810	5'-GCT CTA GAG GCT TAG GAG GTG GCG CAT GCG TGT CTT GAT CGT C-3'	*vanRSst* amplification (no promoter)
AB22119	5'-GCG AAT TCC ATA TGC GTG TGC TGA TTG TCG AGG AC-3'	*vanRsc* amplification
AB22120	5'-GCG AAT TCA AGC TTG ACC GGG GCG CCT ATC CAC C-3'	*vanRsc* amplification
AB22121	5'-GCG AAT TCC ATA TGC TCG CCC CCC TGG ACC G-3'	*vanSsc* amplification
AB22122	5'-GCG AAT TCA AGC TTT CAC CTG CCG GTG TGC GGA G-3'	*vanSsc* amplification
AB22123	5'-GCG AAT TCC ATA TGC GTG TCT TGA TCG TCG AGG AC-3'	*vanRst* amplification
AB22124	5'-GCG AAT TCA CGA TTC CCT ATG CAC CCG CAT CGC C-3'	*vanRst* amplification
AB22125	5'-GCG AAT TCC ATA TGC TCG CCC CAC TGA CCC GCA T-3'	*vanSst* amplification
AB22126	5'-GCG AAT TCA AGC TTT CAC CTG TTG AGG GTT TCC GGC G-3'	*vanSst* amplification

Appendix C - VanA Antibody Production

Materials and Methods

Subcloning, Overexpression and Purification of VanAsc

Purification of VanAsc using a His_6-tag on a nickel column resulted in lack of purification (likely due to buried His_6-tag). The *vanAsc* gene was subcloned from pET28-vanAsc (made previously by Dr. Gary Marshall) into pET22b using the *Hind*III and *Nde*I sites and transformed into *E. coli* BL21(DE3) overexpression cells.

pET22b-vanAsc BL21(DE3) cells were grown, overexpressed and harvested as before (Section 3.3.3). Resuspension was in 20 ml of VanA lysis buffer (25 mM HEPES pH 7.5, 2 mM EDTA, 1 mM PMSF, 0.1 mM DTT and 100 mM NaCl). Resuspended cells were lysed by passage through a French Press three times and the lysed cells were loaded on a 50 ml Q Sepharose (anion exchange) column. The column was washed with 180 ml of VanA column buffer (25 mM HEPES pH7.5, 2 mM EDTA) and protein was eluted with a linear gradient from 0 mM NaCl to 500 mM NaCl over five column volumes into 6 ml fractions.

Assaying for Activity of VanAsc

Two different assays were used to test for VanAsc ligase activity. First was the malachite green assay which measures free phosphate. Fractions eluted from the Q Sepharose column were assayed using 10.0 µl 100mM HEPES pH 8.0, 4.0 µl 500mM $MgCl_2$, 4.0 µl 500mM KCl, 6.0 µl 100mM ATP and 40.0 µl 25mM D-Ala. To this

mixture was added 20.0 µl of the fraction or 20.0 µl of dH_2O (baseline control) to a total volume of 100.0 µl. Reactions were incubated at 37°C for 20 min. 50.0 µl of the reaction was added to 800 µl of the colour reagent (3 parts 0.045% Malachite Green hydrochloride, 1 part 4.2% ammonium molybdate) and incubated for 10 min at room temperature after which time 100.0 µl of 34% trisodium citrate was added. Absorbance at 660 nm was read for the enzymatic reaction which was then subtracted from the value obtained for the water control, thus obtaining the net reading.

A second assay was used for VanAsc ligase activity. Thin Layer Chromatography (TLC) was used with radiolabelled substrate to assay specific D-Ala-D-Ala or D-Ala-D-Lac activity. Reactions for activity were run as in the malachite green assay, except 0.4 µCi [^{14}C] labeled D-Ala or D-Lac was added as well to the reaction. 5.0 µl of the reaction mix was spotted on a Chromagram Thin-Layer Chromatography Sheet (Kodak) and developed until the solvent (n-butanol:acetic acid:water, 12:3:5) front reached one inch from the top. TLC plates were then exposed overnight to X-ray film for visualization of the results.

Purification of $VanA_{Aor}$

The *$vanA_{Aor}$* gene was previously cloned in pET28 with a His_6-tag and was transformed into *E. coli* BL21(pLys) cells by Dr. Gary Marshall. Cells were grown, overexpressed and harvested as before (Section 3.3.3). Cells were resuspended in 20 ml Buffer A (20mM imidazole, 300mM NaCl, 20mM HEPES pH 7.5) with 1 mM PMSF and 0.1 mM DTT added. Cells were then lysed by French Pressing three times and were

loaded on 2 ml of Ni^+-agarose beads. $VanA_{Aor}$ was eluted over 15 column volumes with a linear gradient from 20mM imidazole to 250mM imidazole.

$VanA_{Aor}$ Antibody Production

Antibodies for $VanA_{Aor}$ were raised in two rabbits. Enzyme for injection was purified by running approximately 1.25 mg of $VanA_{Aor}$ on large scale 15% SDS-PAGE acrylamide gels. Bands corresponding to $VanA_{Aor}$ were excised and emulsified. Emulsifications for $VanA_{Aor}$ were prepared by pulverizing the acrylamide bands with a glass rod then sonicating until gel-like in a minimal volume of PBS (phosphate buffered saline pH 7.3 – 137 mM NaCl, 2.7 mM KCl, 4.3 mM $Na_2HPO_4 7H_2O$, 1.4 mM KH_2PO_4) as possible (less than 500 µl). An equal amount of Freund's adjuvant (complete Freund's for the first injection and incomplete Freund's for the subsequent injections) was added and mixed by sonication until emulsified (i.e. does not disperse in water). Injections of approximately 125 µl were done subcutaneously in four places per rabbit by staff at Central Animal Facilities at McMaster University (Hamilton, ON). Ear bleed samples of 1 ml were obtained 10-14 days after injection to check the titre of the antibody (see below). Subsequent injections were done 3-4 weeks after the previous injection. Once the titre was deemed high enough, the rabbits were bled.

The titre of the antibody was monitored using the crude antiserum sample. Ear bleed samples were incubated at 37°C for one hour then left overnight at 4°C to clot and the antiserum was removed. Western blots were done using lysed pET28-$vanA_{Aor}$ BL21(pLys) cells mixed with purified $VanA_{Aor}$ as the control sample. The antibody was diluted 1:10, 1:100 and 1:1000. Once the final bleed was done, the antiserum sample was

collected as above and the IgG was purified by precipitation using ammonium sulfate. One part 33% saturated ammonium sulfate (SAS) was added dropwise to two parts antiserum slowly under constant stirring at 4°C. The solution was mixed continuously for 4 hours at 4°C and the precipitate was isolated by centrifugation. The pellet was washed with cold 33% SAS and resuspended in water to 5% its original volume. The precipitated antibody was then dialyzed against 4 L of water (three changes) over the next 48 hours to fully remove residual SAS.

Purification and Activity of VanAsc

VanAsc was purified on a Q-Sepharose column and fractions were assayed for enzymatic activity before continuing with the purification protocol (Figure C1). Two separate assays were used to determine if the VanAsc was active, the Malachite Green phosphate release assay and the [^{14}C]-D-Ala radiolabelling assay. Both would show any D-Ala-D-Ala or D-Ala-D-Lac ligase activity. The Malachite Green assay measures free phosphate which would be generated as a bi-product by the ATP dependent ligase. Comparison to the positive control DdlB (*E. coli* D-Ala-D-Ala ligase) did not identify a fraction with significant activity. Figure C2 shows the ligase assay in which radiolabelled D-Ala or D-Lac is added to the enzyme with a pool of D-Ala. Should the D-Ala be joined to either D-Ala or D-Lac, a band shift on the TLC plate would be seen (70). This assay was done on fraction 24, which contained the most enzyme (determined by visual examination of SDS-polyacrylamide gel). Figure C2b shows the reaction carried out, compared to negative controls, and no shifts were seen. Thus this fraction did not have any ligase activity.

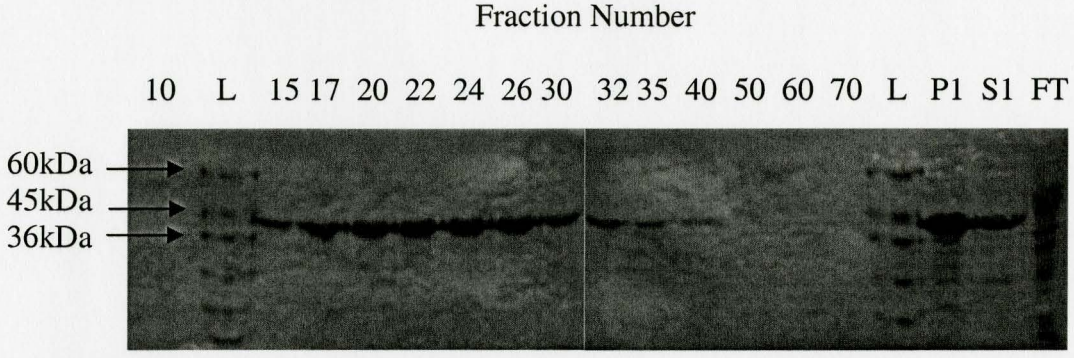


Figure C1 - Purification of VanAsc: The SDS polyacrylamide gels of the fractions from the purification of VanAsc show that VanAsc was over expressed and contained mainly in fractions 24 to 40.

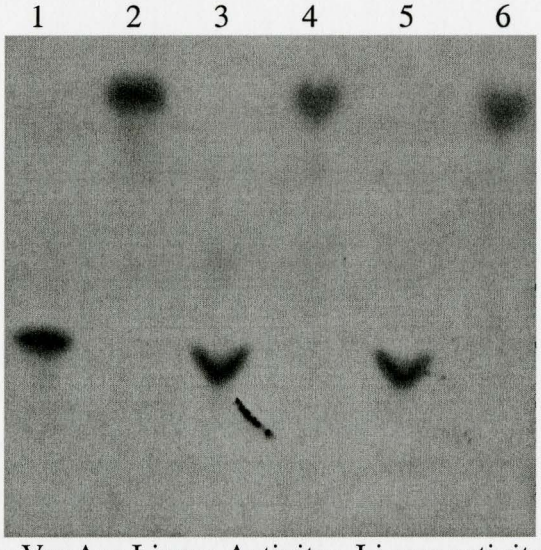


Figure C2 - Assaying for VanAsc Ligase Activity: Ligase activity was not detected when using radiolabelled D-Ala or D-Lac in a reaction containing VanAsc and D-Ala. Lanes 1 and 2 show the radiolabelled D-Ala and D-Lac respectively (control), Lanes 3 and 4 show the ligase assay on fraction 27 and lanes 5 and 6 on fraction 45 with radiolabelled D-Ala (lanes 3 and 5) and D-Lac (4 and 6).

Purification of $VanA_{Aor}$ and Antibody Production

Fractions were separated on 11% SDS-polyacrylamide gels and fractions 15 to 32 were pooled giving 36 ml of 1.24 mg/ml protein. The enzyme was deemed to be greater than 95% pure by visual inspection of the SDS-polyacrylamide gel. Enzyme purified from large scale SDS-PAGE acrylamide gels was injected into rabbits and the resulting

antibody titre was tested by Western blots on serum samples obtained from ear bleeds. Once the titre was deemed high enough, the rabbits were bled and the antibody was purified by precipitation of the IgG. Figure C3 shows the final $VanA_{Aor}$ antibody tested against the E. coli BL21 strain with the $VanA_{Aor}$ expression vector which was induced for $VanA_{Aor}$ overexpression. The antibody binds specifically to the $VanA_{Aor}$, but also has low specificity for a few E. coli proteins.

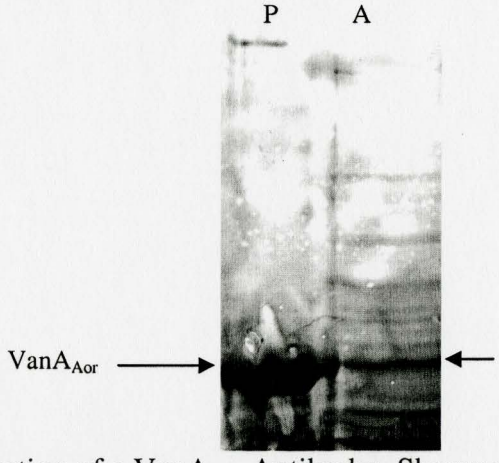


Figure C3 - Production of a $VanA_{Aor}$ Antibody: Shown above is a Westen blot using the $VanA_{Aor}$ antibody produced against a lysate of the E. coli BL21(DE3) strain carrying the pET22b-$VanA_{Aor}$ overexpression plasmid (A). The primary antibody was diluted 1:5000. The positive control (P) was 5.0 ug of purified $VanA_{Aor}$. $VanA_{Aor}$ (arrow) is detected along with some non-specific proteins.

Discussion

In order to proceed with further work on the resistance system present in S. toyocaensis, an antibody against VanAst was needed. Much like with VanS, active VanA in S. toyocaensis and S. coelicolor was never purified. Dr. Gary Marshall previously attempted this and was unable to purify active VanAst using His_6-tagged constructs. Using the same His_6-tagged system, as well as native expression, VanAsc, with the

highest amino acid homology to VanAst, was purified. Unfortunately, its role as a D-Ala-D-Lac ligase was never confirmed. $VanA_{Aor}$ activity had been previously shown and shows high homology to VanAst (73% amino acid identity). As such $VanA_{Aor}$ was chosen as the next best candidate for antibody production.

Antibodies to $VanA_{Aor}$ were raised in rabbits and were found to be specific for $VanA_{Aor}$ when probed with a lysate from E. coli. Some unspecified labeling by the antibody was seen, but was not as strongly labeled as $VanA_{Aor}$. If this proves to be a problem in further experiments the antibody can be further purified using Affi-Gel linked to $VanA_{Aor}$.

Appendix D - Insertional Inactivation of *staL*

Materials and Methods -

Creation of staL Inactivation Plasmids

The sulfotransferase gene *staL* was first PCR amplified and subcloned into pUC18 by Tejal Patel. The pSET-tsr vector (created by Dr. John Neu) was digested with *Age*I and the 1217 bp band containing the apramycin resistance gene *aac(3)IV* was cloned into the unique *Bsp*EI site within the *staL* gene of the pUC18-sulfo vector, which had been passaged through *E. coli* GM48 cells (*dcm*$^-$, *dam*$^-$). Screening by restriction digestion with *Hind*III and *Xho*I to isolate a vector in which the apramycin resistance gene was cloned in the complementary direction of the *staL* gene. The *staL:: aac(3)IV* insertional inactivation cassette was removed with *Eco*RI digestion and subcloned into pUC19-tsr (created by Dr. John Neu) and transformed onto LB-agar apramycin plates to screen for functional pUC19-tsr-sulfo-am colonies. Confirmation of insert was done by restriction mapping with *Eco*RI and *Pst*I. The direction of cloning of the *staL:: aac(3)IV* cassette was confirmed by sequencing using the M13 forward primer.

A second inactivation plasmid for removing activity of *staL* was designed to improve on the original strategy. A DNA region of the biosynthesis cluster upstream of the *staL* gene was ligated into the pUC19-tsr-sulfo-am vector. pCepC1 DNA was digested with *Sph*I (2.0 µg of pCepC1, 20 U *Sph*I, 3.0 µl NEBuffer 3 in 30.0 µl total reaction left overnight at 37 °C) and the 1969 bp fragment was isolated and cloned into

the unique *Sph*I site at the 5' end of *staL* in pUC19-tsr-sulfo-am giving pUC19-tsr-sulfo-am-CepC1. Screening to ensure the cosmid fragment had inserted in the correct orientation was done using *Xho*I restriction mapping.

Preparation and Transformation of S. toyocaensis Protoplasts

S. toyocaensis protoplasts were prepared using a modified procedure developed by Matsushima *et al* (69). A 50 ml TSB + 0.5% glycine culture of *S. toyocaensis* seeded from 200.0 µl of a spore suspension was grown for 48 hours in a baffled flask at 30 °C. Mycelium were harvested by centrifugation and washed once with 20 ml of P buffer, resuspended in 20 ml of P buffer, then homogenized manually in a glass homogenizer. 100 mg of lysozyme was added to the homogenized mycelium and incubated at room temperature for up to 3 hours (until the mixture is cloudy) with occasional inverting. The protoplasts were filtered through cotton and resuspended in 500.0 µl of P buffer and aliquoted in two Eppendorf tubes for immediate use.

Prior to transformation of protoplasts, plasmids were passaged through GM48 cells (*dcm*$^-$, *dam*$^-$) to increase transformation efficiency. Plasmid DNA (9.0 µl of 1.0 µg/µl) was denatured with 2.0 µl 1 M NaOH at 37 °C for 30 min and immediately cooled on ice. The mixture was neutralized with 2.0 µl 1 M HCl and 20.0 µl P buffer was added. The denatured plasmid was then added to one 250 µl aliquot of protoplasts and immediately 500 µl of 25% PEG 1000 (1.0 g PEG dissolved in 3.0 ml P buffer) was added. Protoplasts were then centrifuged and resuspended in 200 µl of fresh P buffer and plated on two R2YE plates (100 µl/plate) and incubated at 30 °C for 24 hours. Plates

were then flooded with 1.0 ml of SOC containing 40.0 μg of apramycin and further incubated for one week.

Testing For Inactivation of staL

Individual colonies which showed apramycin resistance were picked and resuspended in 40.0 μl of TSB and maintained in microtitre plates at 4 °C. Resuspended colonies (5.0 μl) were individually spotted on Bennet's agar-thiostrepton plates. Subsequent passages were done by preparing a spore suspension from individual apramycin resistant colonies (70, 71). Spore suspensions were then plated on Bennet's agar-apramycin plates and individual colonies were rescreened as above. Those colonies that show double recombination phenotypes should regenerate on apramycin plates, but not thiostrepton plates – Am^R, $Thio^S$. For the pUC19-tsr-sulfo-am inactivation plasmid, four passages screening for Am^R were done. For pUC19-tsr-sulfo-am-CepC1, no Am^R colonies were obtained, thus no subsequent passages were performed. The transformation was repeated by varying the amount of DNA transformed (0.9 μg or 9.0 μg) and the concentration of protoplasts transformed (four times original concentration, original concentration and 1/8 original concentration). The transformation was also repeated using the control plasmids pJN7 (37.2 μg transformed, selected for $Thio^R$), pSET-tsr (16.7 μg, $Thio^R$), pIJ8600 (0.12 μg, Am^R) and pOJ446 (3.0 μg, Am^R).

Results

Insertional inactivation of *staL* in *S. toyocaensis* was attempted using two different inactivation plasmids. Transformation of pUC19-tsr-sulfo-am into *S. toyocaensis* resulted in apramycin resistance (Am^R). Thus the plasmid, which was a

suicide vector, was homologously recombined into the genome. Am^R colonies were screened for thiostrepton sensitivity ($Thio^S$) which would be the phenotype of a double recombination (Figure D1). After multiple passages of Am^R *S. toyocaensis* screening for $Thio^S$, as outlined previously, no $Thio^S$ colonies were seen. The same procedure was repeated using a new inactivation plasmid, pUC19-tsr-sulfo-am-CepC1. A greater amount of homologous DNA to the region around *staL* meant it should recombine with a greater efficiency. When using this plasmid there were no cases of single recombinants (Am^R). Repetition of the experiment under various conditions met with the same results (no Am^R). Table D1 shows the results when carrying along four positive controls. There still were no recombination events for the inactivation plasmid pUC19-tsr-sulfo-am, but three different positive controls (pJN7, pSET-tsr and pIJ8600) did work. One control, pOJ446, did not work. This was likely due to the low amount of protoplasts used and the fact that the plasmid sample obtained from Dr. John Neu was quite old. The fact that the *S. toyocaensis* protoplasts were transformable meant that the problem was in the recombination of the plasmid with the DNA.

Table D1 - Results of insertional inactivation attempts of *staL* along with positive controls testing protoplast competency:

Plasmid	Type of Plasmid	Resistance Used for Selection	Amount of DNA used (ug)	Amount of Protoplasts Used	Worked
pJN7	integrating	tsr	37.2	high	Yes
pSET-tsr	integrating	tsr	16.7	high	Yes
pIJ8600	integrating	am	0.12	high	Yes
pOJ446	replicating	am	3	low	No
pUC19-tsr-sulfo-am-CepC1	recombination	am	10	high	No

Discussion

In order to confirm the role of the StaL in A47934 sulfonylation, insertional inactivation of the *staL* gene was attempted. A first trial utilized the plasmid pUC19-am-tsr-sulfo which contained the cloned *staL* gene interrupted by an apramycin resistance cassette. Only single homologous crossover events were observed using this plasmid. With small regions of homology on either side of the apramycin resistance cassette, 266 bp upstream and 546 bp downstream, it appeared that homologous recombination only occurred within the 546 bp downstream region. A double homologous recombination event was not seen after three further passages, possibly due to insufficient homology in the upstream region. A second construct was then created, pUC19-am-tsr-sulfo-CepC1, in which a region of DNA upstream of *staL* was also subcloned into the plasmid. This increased the amount of homologous DNA upstream of the apramycin resistance cassette to 2176 bp. When this plasmid was used no homologous recombination was seen. Numerous attempts were made, under a variety of conditions, to obtain recombinants without any success. Three positive controls for the transformation, however, were successful, thus indicating that the lack of recombination was due to the properties of the pUC19-am-tsr-sulfo-CepC1 construct.

The results of the attempts to inactivate the *staL* gene may indicate that removing A47934 sulfonylation is lethal to *S. toyocaensis*. In the case of the first inactivation attempts, a single homologous event is seen likely occurring downstream of the *staL* gene where there is higher homology. This would maintain *staL* in a functional in terms of the A47934 biosynthesis cluster. In the second attempt, the homologous recombination

would occur in the region upstream of *staL*. Supposing *staL* is co-transcribed with the rest of the biosynthesis cluster, such an event would inactivate the enzyme (Figure D1). Using the first plasmid, double recombination events would not be seen as they would prove to be lethal. The same logic applies for even single recombination events with the second plasmid.

staL mutants may be lethal as A47934 may act as a signaling molecule. The sulfate group could serve as the site of recognition by a yet to be determined receptor involved in upregulating production of the GPA resistance genes. Phenotypic studies of *S. toyocaensis* show some differences from other GPA resistant strains. Table D2 outlines the differences in resistance between the VanA phenotype in *Enterococci*, *S. coelicolor* and *S. toyocaensis*. Both VanA and *S. coelicolor* show resistance to vancomycin and A47934, while *S. toyocaensis* only shows resistance to A47934. Resistance to vancomycin in *S. toyocaensis* can be achieved by adding A47934 to the growth media. This apparently "turns on" the GPA resistance genes.

Inspection of Figure 1 points to a key difference between A47934 and other GPAs: the sulfate group. If *S. toyocaensis* truly does differentiate A47934 from other GPAs, this would be an excellent place for it to do so. It is unlikely that the aglycone aspect of A47934 is recognized, since this would require a large receptor capable of binding to the majority of the GPA. Nonetheless, if the presence of sugars is inhibiting recognition, this could easily be tested by checking the resistance pattern of *S. toyocaensis* to other aglycone GPAs. On the other hand, sulfate groups have been shown to be responsible for a wide variety of signaling events. Sulfate groups are advantageous

because of their ability to interact specifically with protein receptors. The fact that A47934 is the only GPA with a sulfate group may indicate that it functions as more than just a GPA. It is important to note that a role for the sulfate group in GPA function has yet to be identified. In order to study the role of *staL*, its inactivation would have to be characterized in an *S. toyocaensis* strain in which the *vanHAXst* genes are constitutively expressed.

Table D2 – Resistance phenotype trends:

Source	Vancomycin Resistance	A47934 Resistance	Inducible Vancomycin Resistance
VanA phenotype	✓	✓	✓
S. coelicolor	✓	✓	✓
S. toyocaensis		✓	✓

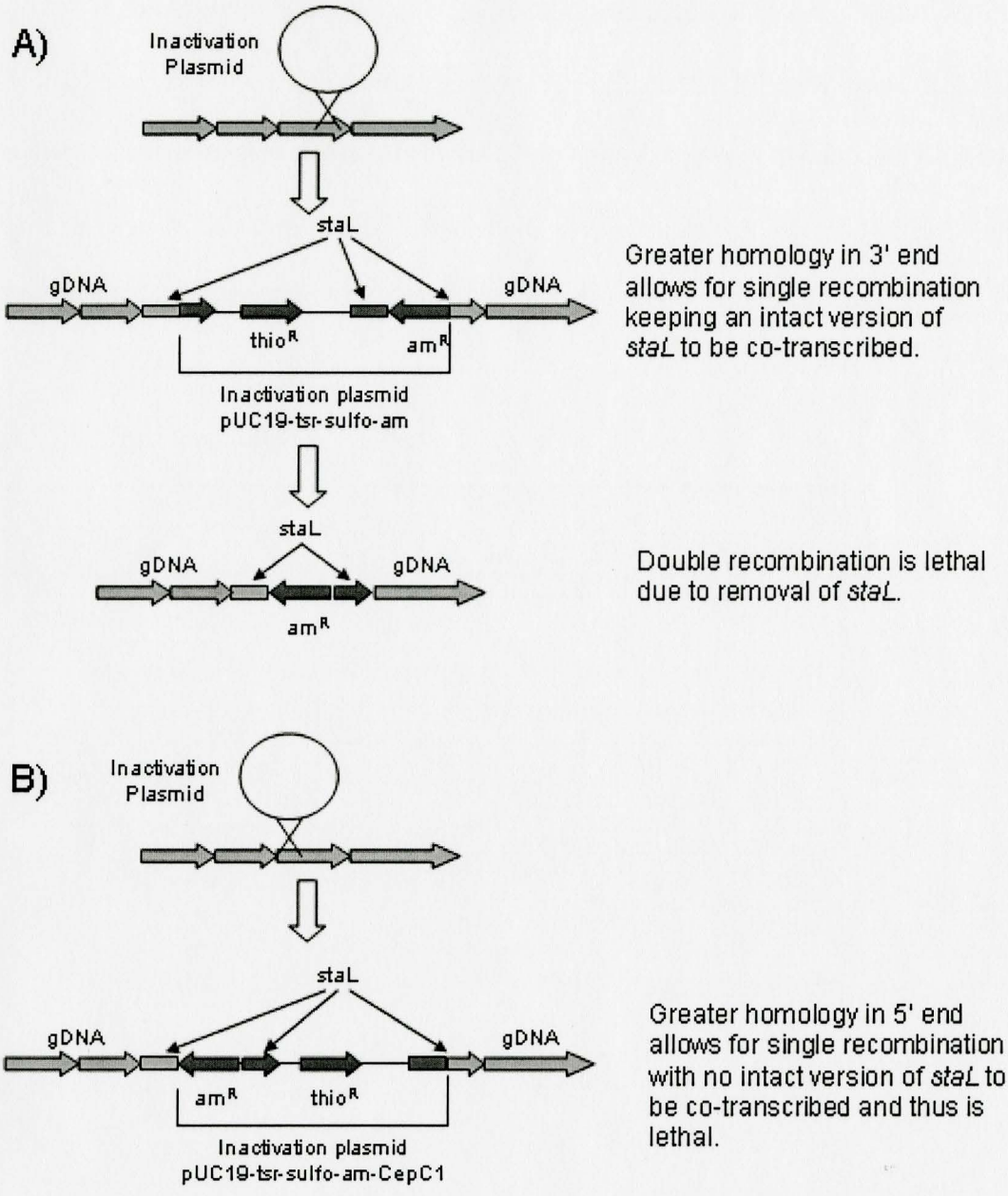


Figure D1 – Inactivation of *staL* by Two Different Plasmids: By assuming the site of recombination for the inactivation plasmid with *staL* is the region with the greatest homology we can see why the first inactivation plasmid (A) gave single recombinations, while the second (B) did not. It is proposed that it is necessary to keep a co-transcribed intact *staL* gene in order for the cells to survive.