Dissecting tunicamycin biosynthesis: A potent carbohydrate processing enzyme inhibitor

A thesis submitted in partial fulfilment for the degree of Doctor of Philosophy at the University of Oxford

By

Filip J. Wyszyński

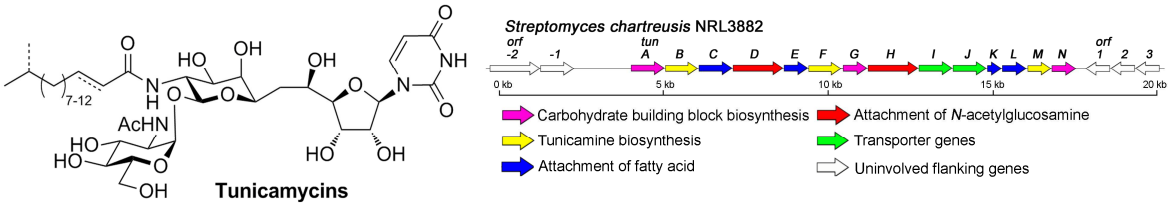
Jesus College, Oxford

Trinity Term 2010

Dissecting tunicamycin biosynthesis: A potent carbohydrate processing enzyme inhibitor

by Filip J. Wyszyński, Jesus College
Submitted towards the degree of D.Phil. in Trinity Term 2010

Tunicamycin nucleoside antibiotics were the first known to target the formation of peptidoglycan precursor lipid I in bacterial cell wall biosynthesis. They have also been used extensively as inhibitors of protein *N*-glycosylation in eukaryotes, blocking the biogenesis of early intermediate dolichyl-pyrophosphoryl-*N*-acetylglucosamine. Despite their unusual structures and useful biological properties, little is known about their biosynthesis. Elucidating the metabolic pathway of tunicamycins and gaining an understanding of the enzymes involved in key bond forming processes would not only be of great academic value in itself, it would also unlock a comprehensive toolbox of biosynthetic machinery for the production of tunicamycin analogues which have the potential to act as novel therapeutic antibiotics or as specific inhibitors of medicinally important NDP-dependent glycosyltransferases.



I – Cloning the tunicamycin biosynthetic gene cluster

We report identification of the tunicamycin biosynthetic genes in *Streptomyces chartreusis* following genome sequencing and a chemically-guided strategy for *in silico* genome mining that allowed rapid identification and unification of an operon fractured across contigs. Heterologous expression established a likely minimal gene set necessary for antibiotic production, from which a detailed metabolic pathway for tunicamycin biosynthesis is proposed.

II – Natural product isolation and degradation

We have developed efficient methods for the isolation of tunicamycins from liquid culture in preparative quantities. A subsequent relay synthesis furnished advanced biosynthetic intermediates for use as precursors in the production of tunicamycin analogues and as substrates for the *in vitro* characterisation of individual Tun enzymes.

III – Functional characterisation of *tun* gene products

Individual *tun* gene products were over-expressed and purified from recombinant *E. coli* hosts, allowing *in vitro* functional studies to take place. An NMR assay of biosynthetic enzyme TunF showed it acted as a UDP-GlcNAc-4-epimerase. Putative glycosyltransferase TunD showed hydrolytic activity towards substrate UDP-GlcNAc but failed to accept to the expected natural acceptor substrate, providing unexpected insights into the ordering of biosynthetic events in the tunicamycin pathway. Initial studies into the over-expression of the putative sugar *N*-deacetylase TunE were also described.

IV – Towards synthesis of tunicamycin fragments

Investigations into a novel synthesis of D-galactosamine – a structural motif within tunicamycin – led to the unexpected observation of inverted regioselectivity upon Rh^{II}-catalysed C–H insertion of a D-mannose-derived sulfamate. This was the first example of *N*-insertion at the β- rather than γ-C–H based on conformation alone and warranted further investigation. The X-ray structure of a key sulfamate precursor offered valuable insights as to the source of this unique selectivity.

Acknowledgements

I am extremely grateful for the loving support of my family throughout my DPhil: my parents, my brother Luke, his wife Vanessa and their newest addition beautiful baby Amélie.

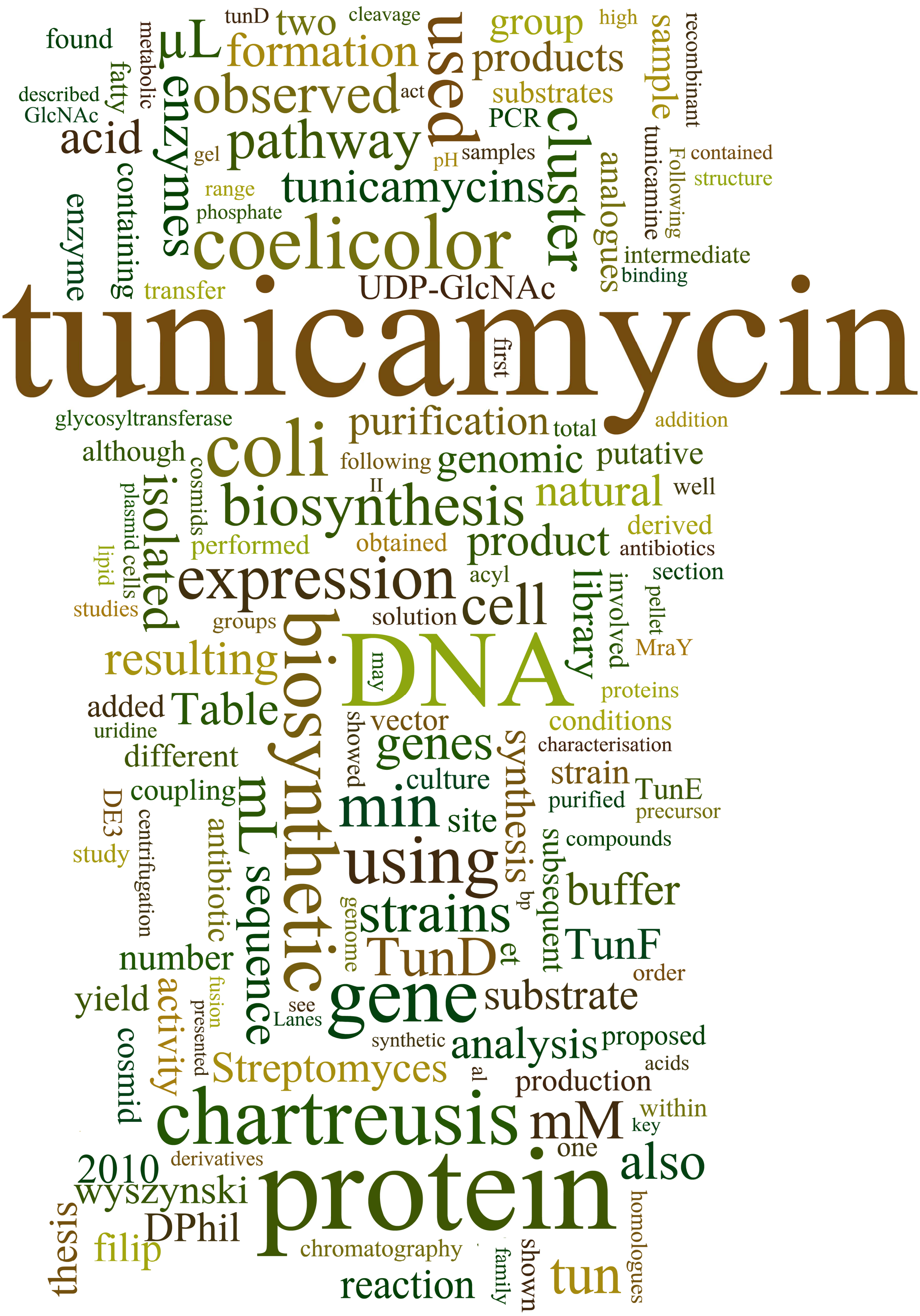
I am also very thankful to Prof. Ben Davis for the opportunity to experience cutting-edge research involving so many facets of chemistry and biology. I am grateful for being given the freedom to direct my own project, for being inspired to push myself and to fulfil my potential and finally for Ben's constructive criticisms and his receptiveness to open debate.

A huge thank you is also reserved to Prof. Mervyn Bibb and Andy Hesketh who agreed to host me at the John Innes Centre and without whom the majority of this thesis would simply not have been possible.

Dr Robert Whitam and Dr Jonathan Burton were inspirational teachers and I thank them for fuelling my interest in chemistry at school and as an undergraduate respectively. Without them I would certainly be doing something else.

Big thanks go to all BGD group members past and present, with special mentions to James, Katie, Alister, Marta, Keri, Hong, Sam, Ollie and Eoin for the early days, to Bala, Chris, Smita, Jonny, Inga, Alba, Matt, Conor, Jo, Julia, Seung and Tom for more recent times and to heavyweights Justin, Mitul and Lingbing for being here *in perpetuam*. Additional shout-outs go to any members of LG9, DPhils from the class of 07-08 (Justin, Lingbing, Isa and the Pipster) and anyone who kindly agreed to the delights of proofreading.

I am also thankful to the many great friends I have made over the years from Jesus College (Oxford) and Corpus Christi College (Cambridge), providing welcome respite from chemistry and supplying inordinate amounts of fun and happy memories. Special mention goes to my long suffering housemates George, Nicky and Melissa, to the institutions of JCRFC, Pelicans, Elizabethans and Eagles, and to all other individuals who have enriched my life.


found
described
GlcNAc
acid
enzyme
fatty
metabolic
containing
μL
enzymes
tunD
two
cleavage
formation
observed
act
gel
pathway
range
phosphate
used
group
high
products
substrates
PCR
pH
samples
sample
recombinant
tunicamine
Following
contained
structure
tunicamycins
coelicolor
cluster
analogues
intermediate
binding
transfer
UDP-GlcNAc
tunicamycin
first
glycosyltransferase
purification
total
addition
although
coli
following
genomic
putative
cosmids
plasmid
isolated
II
biosynthesis
natural
well
lipid
cells
performed
obtained
product
derived
antibiotics
studies
expression
acyl
section
groups
solution
cell
library
involved
pellet
resulting
biosynthetic
may
DNA
MraY
proteins
added
Table
uridine
showed
vector
genes
synthesis
conditions
characterisation
different
culture
strain
coupling
min
site
purified
TunE
precursor
DE3
centrifugation
study
antibiotic
mL
sequence
using
subsequent
compounds
genome
strains
bp
buffer
number
TunD
et
TunF
order
yield
activity
fusion
presented
see
Lanes
gene
substrate
synthetic
analysis
proposed
cosmid
Streptomyces
al
acids
production
chartreusis
mM
within
key
one
2010
derivatives
wyszynski
protein
also
thesis
DPhil
filip
chromatography
family
shown
homologues
reaction
tun

List of publications

Filip J. Wyszynski, Andrew R. Hesketh, Mervyn J. Bibb and Benjamin G. Davis

Dissecting tunicamycin biosynthesis by genome mining: cloning and heterologous expression of a minimal gene cluster.

Chemical Science, **2010**, 1 (5), 581-589

Filip J. Wyszynski, Amber L. Thompson and Benjamin G. Davis

Inverted regioselectivity of C–H amination: Unexpected oxidation at β- rather than γ-C–H.

Organic & Biomolecular Chemistry, **2010**, 8 (19), 4246-4248

Table of contents

Chapter 4: Functional characterisation of tun gene products

Chapter 5: Towards synthesis of tunicamycin analogue fragments

Chapter 6: Summary

Appendix

Detailed table of contents

Chapter 1: Introduction

Chapter 5: Towards synthesis of tunicamycin analogue fragments

Chapter 6: Summary

Appendix

Chapter 1: Introduction

1.1 The tunicamycin family of nucleoside antibiotics

1.1.1 Foreword

The tunicamycins represent a family of closely related nucleoside antibiotics with important biological properties – namely inhibition of bacterial peptidoglycan biosynthesis and eukaryotic *N*-linked protein glycosylation.[1,2] Since their discovery almost four decades ago they have received widespread attention across many research fields and have earned over 5,500 citations in the literature.[3] Tunicamycins will form the core topic of this DPhil thesis and will be the focus of all research presented herein. In this chapter I review the history, structure and biological function of the tunicamycin family and summarise existing investigations into their biogenesis and chemical synthesis, in addition providing more general background information in relevant areas. Finally, I set out a hypothesis for tunicamycin analogues acting as novel targeted glycosyltranferase inhibitors and outline the overall objectives of this research project.

1.1.2 Tunicamycins

Tunicamycin was first isolated in 1971 by Takatsuki *et al.* from *Streptomyces lysosuperificus* following the screening of over 4000 *Actinomycetes* strains, isolated from soil samples, for antibiotic activity.[4] In 1980 it was additionally isolated from strains of *Streptomyces chartreusis*.[5,6] Its structure was subsequently solved and it emerged that tunicamycin was present as a mixture of closely related nucleoside antibiotics (Fig. 1.1).[7] The core structure shared by all tunicamycin family members is highly unusual. It comprises a unique eleven-carbon dialdose termed tunicamine, resembling ribose and 6-deoxy-galactosamine residues linked tail-to-tail through a C–C bond. At one end of this pseudodisaccharide lies an *N*-glycosidically linked uracil group and at the other an *N*-acetyl-D-glucosamine (GlcNAc) residue attached through a non-reducing α,β-1,1 glycosidic bond – also referred to as an α,β-trehalose linkage.[8] The diversity of the tunicamycin family derives from the nature of the acyl chain attached to this common core through the tunicamine nitrogen.

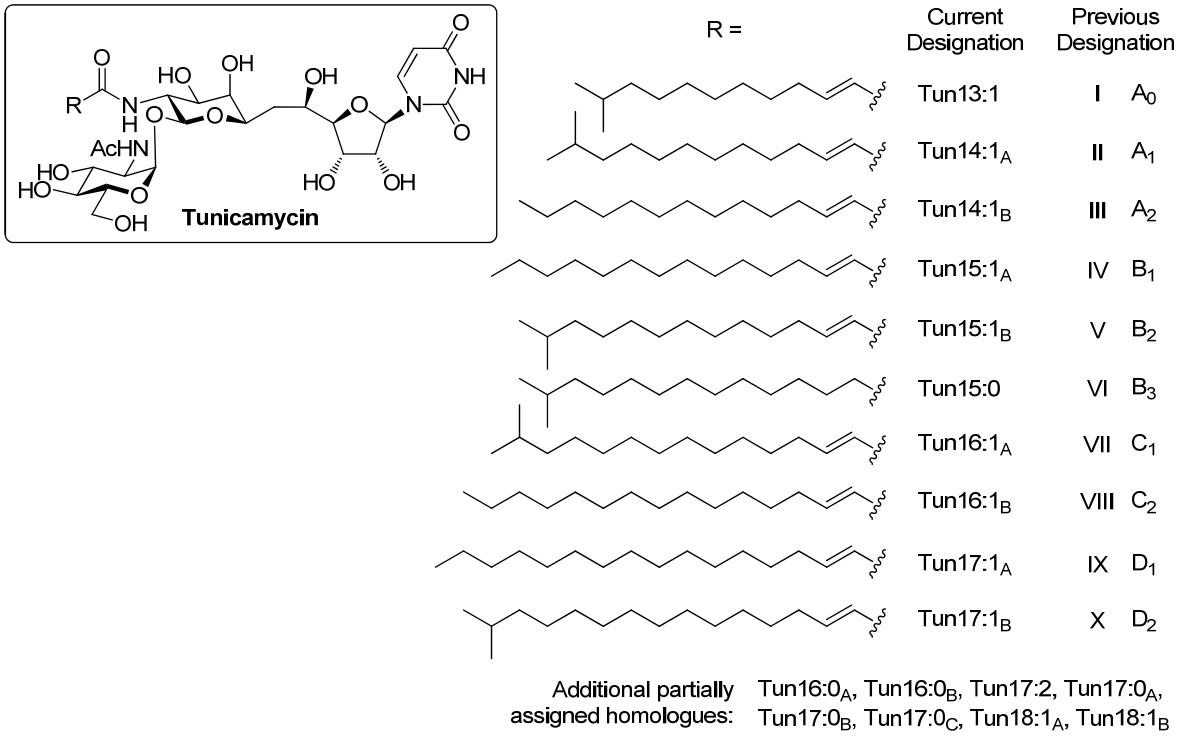


Fig. 1.1 Chemical structures of the tunicamycins

A number of investigations have succeeded in separating and characterising individual components of the tunicamycins, and it was found that the resulting homologues differ in length, unsaturation and terminal branching of their *N*-linked fatty acyl chains.[7,9-13] To date at least eighteen tunicamycin homologues have been resolved, of which ten have fully assigned structures (Fig. 1.1).[1,13] A number of different nomenclatures have emerged to describe these variants, typically based on their relative elution mobility under reverse-phase HPLC.[11,12] More recently Tsvetanova *et al*. proposed a more consistent system, incorporating structural features into the homologue names (Fig. 1.1). Taking the form Tun*x*:y_N, *x* refers to the number of carbons in the acyl chain (including C=O), *y* refers to the degree of unsaturation in the chain and N refers to the order of elution of regioisomers.[13] Based on the varying biological activity of individual purified homologues, the nature of the acyl group was shown to have a bearing in the function of the molecule. The biological activity of the tunicamycins will be discussed in Section 1.2.

1.1.3 Streptovirudins

Streptovirudins, first isolated from strains of *Streptomyces griseoflavus* in 1975, also belong to the tunicamycin family of nucleoside antibiotics.[14] Their structures have been elucidated and resolved into ten distinct homologues (Fig. 1.2).[15,16] They are almost identical to the tunicamycins, differing only in composition of the acyl chain and in some cases the uracil moiety; in fact streptovirudins B_{2a} and C_2 are identical to tunicamycins Tun13:1 and Tun14:1_A. The acyl chains are similarly unsaturated, but their length is typically shorter and in some cases *anteiso*- forms are present alongside their more typical *iso*- regioisomers. In addition, in five of the homologues isolated uracil is replaced with dihydrouracil. Once again, the antimicrobial activity observed was partially dependent on the length of the fatty acid side chain, although the form of the uracil moiety was shown to have little effect on biological activity. It could be argued that streptovirudins are no more than further homologues of the tunicamycins although they have been treated separately in the literature and hence the distinction remains.[1]

1.1.4 Corynetoxins

Corynetoxins constitute further members of the tunicamycin antibiotic family and were first reported as the causative agent for the toxicity of annual rye grass to grazing animals. They were isolated from *Corynebacterium rathayi* which infects the galled seed head of the grass, together with nematode *Anguina agrostis*.[17] Fourteen homologues were seperated and structurally characterised (Fig. 1.2);18-20 they were found to differ from the tunicamycins only in the nature of their fatty acid side chains. Indeed, corynetoxins $U16_i$ and $U17_i$ are identical to tunicamycins Tun16:1_A and Tun17:1_B – this is of interest since the producing organisms belong to relatively different taxonomic groups. Relative to the tunicamycins, the corynetoxin acyl chains are typically longer and include additional *anteiso*- branched forms. β-Hydroxylation of the fatty acids is also observed, alongside more typical unsaturated and fully-saturated forms. The nomenclature used to distinguish between corynetoxins is based on the length of the constituent fatty acids (C15 to C19), whether they are saturated (S), unsaturated (U) or β-hydroxylated (H), and whether they have a normal (n), *iso*- (i) or *anteiso*- (a) termination.[18] Like the tunicamycins, they have been shown to inhibit *N*-glycosylation of proteins,[21] although studies have predominantly focused on the physiology of their toxicity to animals.[22,23]

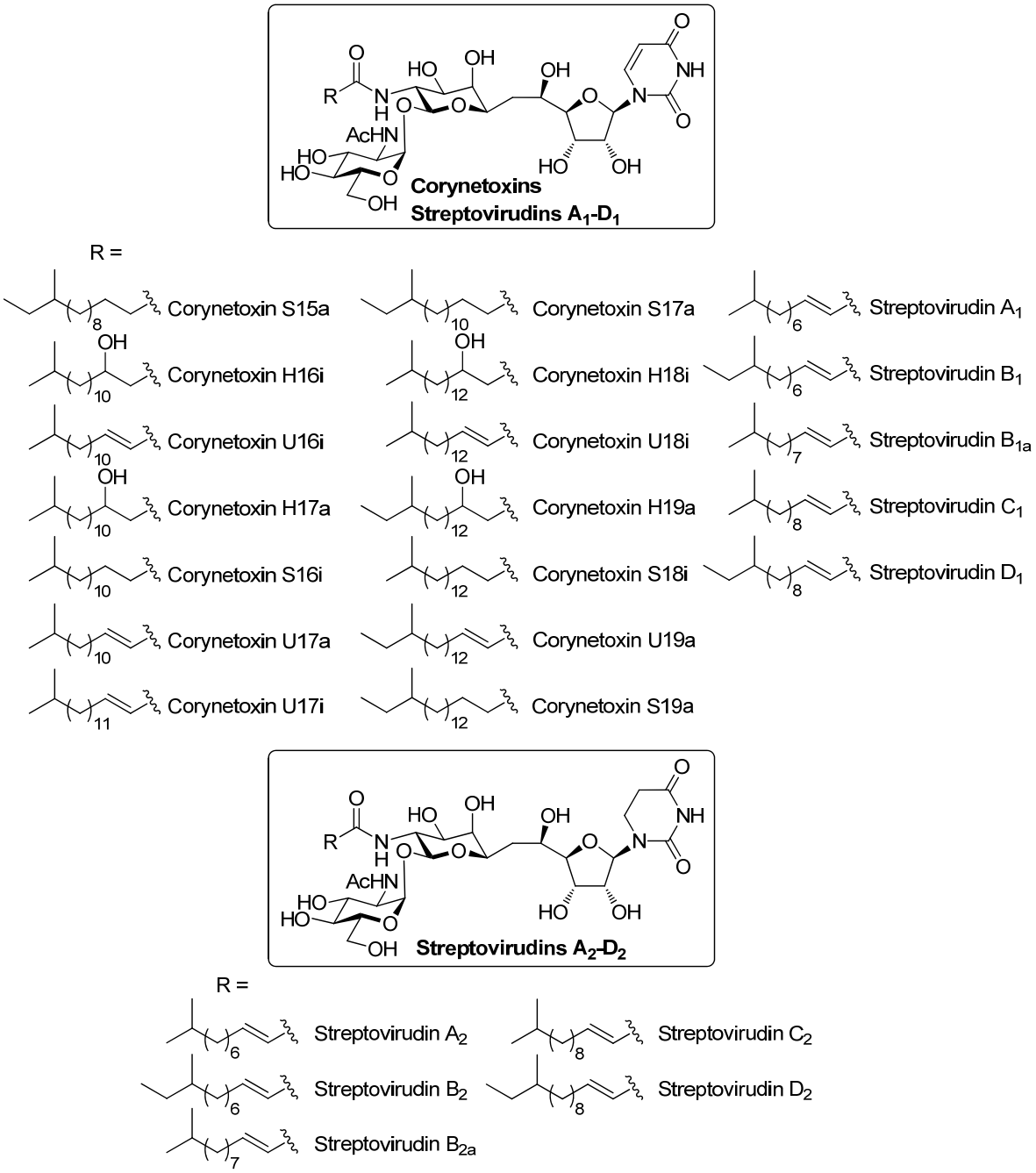


Fig. 1.2 Chemical structures of the corynetoxins and streptovirudins

1.1.5 Other members

In addition to streptovirudins and corynetoxins, three further natural products are suggested to belong to the tunicamycin family of nucleoside antibiotics – MM-19290,[24] antibiotic 24010[25] and mycospocidin.[26] These compounds have not been studied in detail and their structures have not been elucidated, but they do share key physiochemical properties with the tunicamycins.[24-28] In addition, antibiotic 24010 and mycospocidin have been shown to inhibit the formation of *N*-acetylglucosaminyl-pyrophosphoryl-dolichol involved in protein *N*-glycosylation, much like parent compound tunicamycin.[28-30]

1.2 Biological activity of the tunicamycins

1.2.1 Overview

The journey to determine the tunicamycins' biological activity and precise mode of action was comprehensively reviewed in G. Tamura's 1982 book 'Tunicamycin'.[2] These antibiotics were first isolated based on the criteria that they inhibit plaque formation caused by viruses yet show little cytotoxicity to a confluent layer of chick embryo fibroblast cell culture – establishing them as antiviral antibiotics.[4,31] Tunicamycins were shown to inhibit the multiplication of a large range of enveloped RNA and DNA viruses[2] and their mode of action was first studied using a virus-cell system comprising of Newcastle disease virus and chick embryo fibroblasts.[31,32] Different stages of the virus life cycle were sequentially ruled out as targets for the antibiotic, including DNA, RNA and protein synthesis. However, its specific inhibition of glycoprotein biosynthesis was deduced from the curtailment of hemagglutinin and neuraminidase viral glycoprotein activity upon addition of tunicamycin.[2] Further studies showed the tunicamycins retarded cell growth in growing cell culture, and also produced morphological changes in Gram-positive bacteria, yeasts and fungi.[33,34] Inhibition of sugar incorporation was attributed to these observations, and it was concluded that tunicamycin inhibited glycoconjugate biosynthesis in a range of organisms. The observation that viral envelope glycoproteins and bacterial cell wall biosynthesis were inhibition targets of this novel antibiotic gave rise to the name 'tunicamycin', derived from the Latin *tunica* meaning coat or covering.[4] The mode of action of the tunicamycins was established as inhibition of polyprenol-linked sugar formation, intermediates in glycoconjugates biosynthesis. Specifically, formation of dolichyl-pyrophosphoryl-*N*-acetylglucosamine (Dol-PP-GlcNAc) and undecaprenyl-pyrophosphoryl-*N*-acetylmuramoyl pentapeptide (lipid I), active in protein *N*-glycosylation

and peptidoglycan biosynthetic pathways respectively, were inhibited.[35-37] This dual activity against eukaryotic glycoprotein formation and bacterial cell wall biosynthesis has thus far prevented the tunicamycins from being used as antibacterial agents. The origins of tunicamycin's biological activity against these pathways and the resultant physiological implications are discussed in the following subsections.

1.2.2 Peptidoglycan biosynthesis

1.2.2.1 Peptidoglycan biosynthetic pathway

The cell walls of Gram-positive and Gram-negative bacteria are made up of peptidoglycan, a macromolecular polymer of linear glycan chains interlinked by short peptides (Fig. 1.3). It is found outside the cytoplasmic membrane and functions to provide structural strength to the cell and preserve cell integrity against internal osmotic pressure.[38,39] Its absence from eukaryotic cells has made it a key target for antimicrobial drugs for many decades.[40]

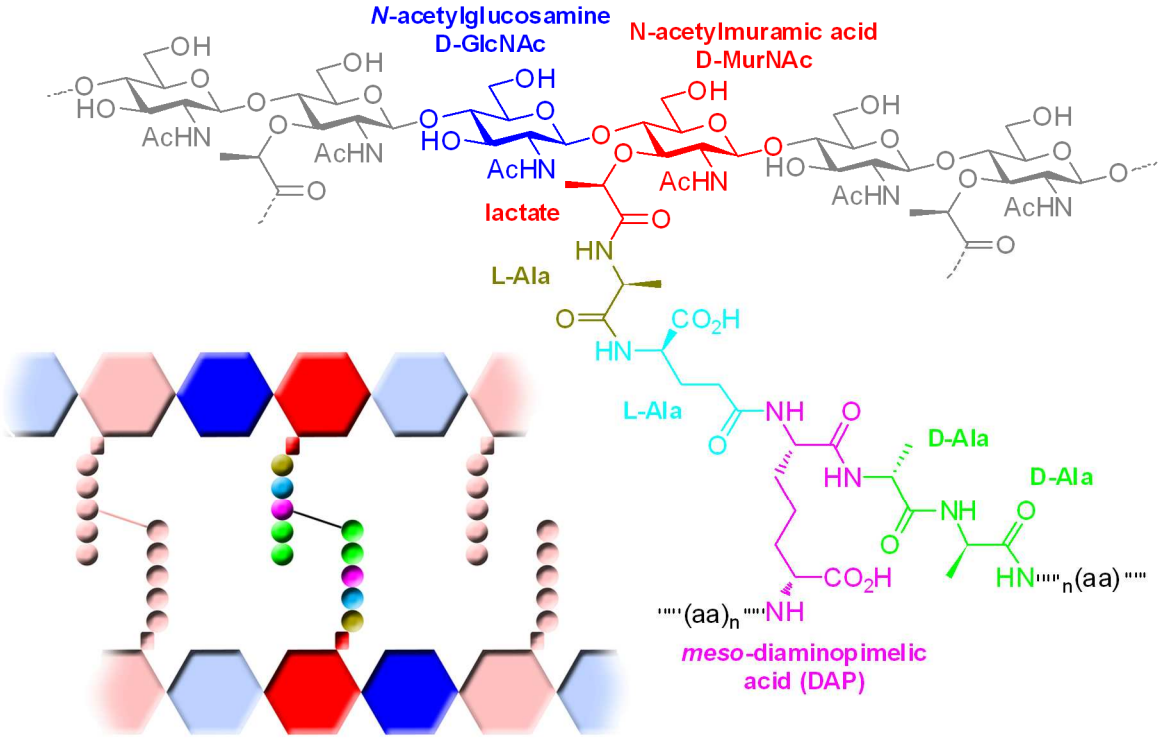


Fig. 1.3 Typical structure of peptidoglycan

Peptidoglycan is made up of a β-1,4-linked polysaccharide of alternating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) units, with MurNAc

consisting of a GlcNAc residue harbouring a 3-*O*-lactyl sidechain (Fig. 1.3). The lactyl group in turn forms an amide bond with a pentapeptide of general structure L-Ala-γ-D-Glu-Xaa-D-Ala-D-Ala, where Xaa is either L-Lys in most Gram-positive bacteria or *meso*-diaminopimelic acid (DAP) in most Gram-negative bacteria. These polymers are then cross-linked through transpeptidation of the ε-amino group of L-Lys or DAP with the D-Ala of another polysaccharide chain; sometimes a short peptide is inserted in this linkage to act as a spacer or bridge. This cross-linking confers rigidity to the macromolecular structure and forms a 3D mesh across the surface of the cell, providing the necessary properties to function as a protective cell wall. Finally, minor modifications to terminal D-Ala residues in the resulting lattice leads to the formation of mature peptidoglycan.[38,39]

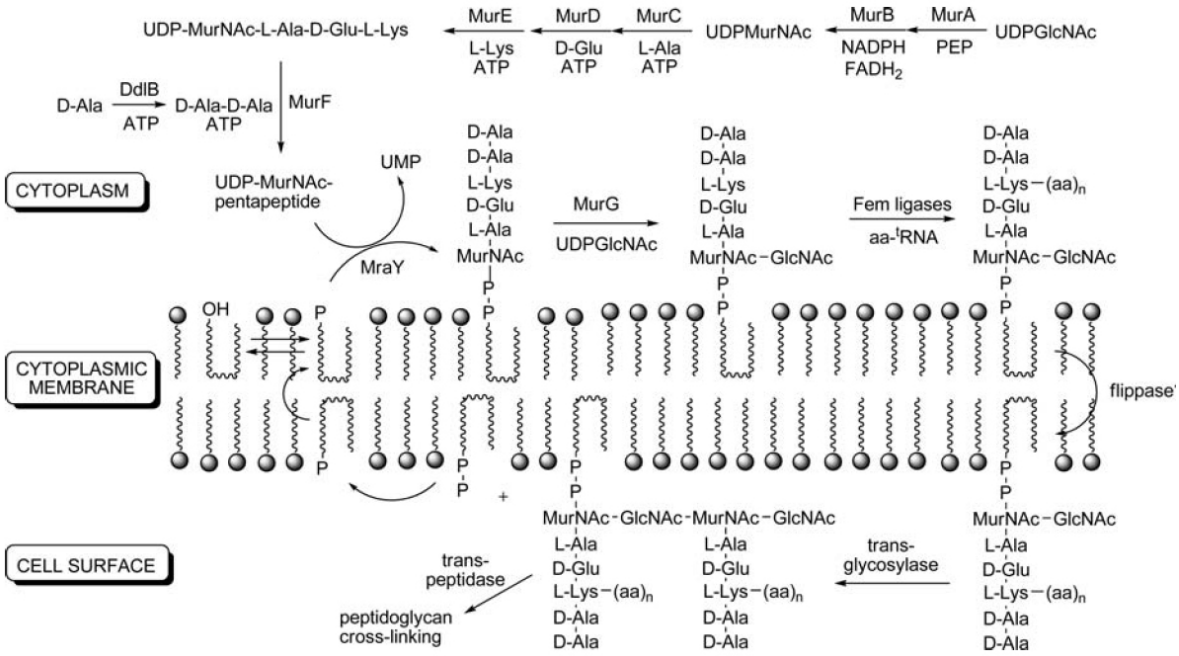


Fig. 1.4 The peptidoglycan biosynthetic pathway.
(from Winn *et al*.[41] – reproduced by permission of The Royal Society of Chemistry)

The biosynthesis of peptidoglycan has been studied for over fifty years and is well understood. The different stages of peptidoglycan biosynthesis have been the subject of extensive, detailed reviews.[42-47] The first stage involves construction of UDP-MurNAc-pentapeptide in the cytoplasm. Phosphoenolpyruvate (PEP) is ligated to metabolic precursor UDP-GlcNAc and reduced with NADPH, by the actions of MurA and MurB respectively, to yield UDP-MurNAc. Stepwise addition of specific amino acids follows, catalysed by a series of ATP-dependent ligase enzymes MurC-F. This soluble intermediate is then associated with the cytoplasmic membrane by the action of MraY (a.k.a. translocase I), which transfers the MurNAc-pentapeptide to a membrane-bound undecaprenyl

phosphate lipid carrier. This lipid I intermediate then undergoes a series of transformations on the cytoplasmic face of the cell membrane. A GlcNAc residue is attached to the MurNAc moiety within lipid I by the action of MurG, yielding lipid II – defined as GlcNAc β-(1→4)-linked to undecaprenyl-pyrophosphoryl-MurNAc pentapeptide. In some bacteria additional amino acids are transferred to this molecule. A flippase protein then shuttles this advanced intermediate to the other side of the cytoplasmic membrane, where it is polymerised by transglycosylase enzymes – part of the penicillin-binding protein family. The undecaprenol phosphate lipid carrier is released at this point and recycled by the cell. Cross-linking of the resulting polysaccharides and their binding to the cell wall occurs through the action of transpeptidases (also part of the penicillin-binding protein family). Some final tailoring reactions follow, resulting in the formation of mature peptidoglycan.

1.2.2.2 Existing inhibitors of peptidoglycan biosynthesis

Inhibitors of the peptidoglycan biosynthetic pathway are much sought after for use as clinical antibiotics and have been subject of intensive research for many decades, particularly in light of the constant battle against emerging antibiotic resistance in bacteria.[40,48] Several classes of natural products target different stages of this pathway, a number of which are well known and have seen much use as therapeutic antibacterial drugs.

Only a small number of antibiotic inhibitors are known which block the early steps of peptidoglycan biosynthesis, taking place in the cytoplasm. Fosfomycin targets pyruvyltransferase MurA preventing UDP-MurNAc formation,[49,50] whereas synthetic D-alanine analogues such as D-cycloserine inhibit alanine racemase DdlB, preventing the formation and hence incorporation of non-proteinogenic D-Ala into UDP-MurNAc.[51] The formation of cytoplasmic membrane-associated lipid I by translocase I (MraY) is the target of a number of related nucleoside antibiotic families, including liposidomycins, mureidomycins and tunicamycins. Their action will be discussed in more detail in the following section.

Lipid II is one of the most important targets for antibiotics, with a variety of natural product classes exhibiting their function against it. Amongst these is moenomycin, a phosphoglycolipid which acts as a lipid II substrate mimic, binding strongly to the transglycosylase active site and blocking polymerisation. Unfortunately it has so far eluded therapeutic use due to poor absorption and pharmacokinetic properties.[45,52]

Glycopeptide antibiotics are probably the most important class of natural products targeting lipid II, with vancomycin and teicoplanin being used as a last-resort antibiotic

against resistant Gram-positive pathogens such as methicillin-resistant *Staphylococcus aureus* (MRSA). The mode of action of these non-ribosomal peptides relies on non-covalent binding to the D-Ala-D-Ala terminus of lipid II, blocking subsequent transglycosylation and transpeptidation steps.[53-55] Notable synthetic variants dalbavancin, telavancin and oritavancin exhibit optimised binding characteristics and pharmacokinetic properties, showing improved activity against emerging vancomycin-resistant enterococci (VRE) and vancomycin-resistant *Staphylococcus aureus* (VRSA).[55,56]

Lantibiotics are post-translationally modified, ribosomally-synthesised peptides and they exhibit a different mode of action on lipid II.[57,58] They are classed into two major subgroups, type A and type B. Type B lantibiotics such as mersacidin and actagardine possess globular structures which bind lipid II and block transglycosylase and transpeptidase activities. Type A lantibiotics such as nisin and epidermin on the other hand possess straight-chain, amphiphilic structures which perform the dual action of binding lipid II and then using it as a docking molecule, allowing the binary complex to form a defined pore in the cellular membrane, thereby disrupting its integrity. Both type A and B compounds bind the carbohydrate-pyrophosphate portion of lipid II, thus differentiating them from glycopeptides. Recently-discovered plectasin is an unmodified fungal defensin peptide with a lipid II-binding mode of action similar to type B lantibiotics.[59]

Further lipid II-binding compounds are generally classed as cyclic depsipeptides or lipopeptides and include ramoplanin, mannopeptomycins, katanosin, plusbactin and daptomycin.[48,58] Their structures typically involve a cyclic peptide including non-proteinogenic amino acids and are adorned with acyl chains or hydrophobic residues essential for activity, as well as carbohydrate moieties in some cases. Although evidence suggests they bind lipid II, the precise mode of action in each case is still a matter of debate and active research.[48,58,60]

Other, related peptides are known to target different areas of peptidoglycan biosynthesis. Bacitracin forms a Zn^{2+}-dependent complex with undecaprenyl pyrophosphate, preventing the recycling of this essential lipid-carrier after release during transglycosylation.[48,61] Similarly, lipopeptides amphomycin and friulimicin form Ca^{2+}-dependent complexes with undecaprenyl phosphate.[48,62] It should be noted that these lipid carriers are utilised in other pathways such as teichoic acid and capsid biosynthesis, so their sequestration has more general effects outside the peptidoglycan pathway.

By far the best known and most widely used class of peptidoglycan biosynthesis inhibitors are the large group of β-lactam antibiotics, encompassing subgroups of

penicillins, cephalosporins, monobactams, carbapenems and their numerous semisynthetic variants.[63,64] They all possess a β-lactam ring which structurally mimics the D-Ala-D-Ala terminus of lipid II and are thus recognised and tightly bound by the active site of transglycosylases and transpeptidases, giving these enzymes the pseudonym of penicillin-binding proteins. These antibiotics form a stable binary complex or enzyme-bound intermediate which inactivates the protein.[65]

These inhibitors, their mode of action and in many cases mechanisms of resistance to them have been well studied and the area remains a very active topic in medicinal chemistry, microbiology and natural product research.

1.2.2.3 Structure and function of MraY

MraY, also known as translocase I (EC 2.7.8.13), is an integral membrane protein which catalyses the synthesis of lipid I on the interior surface of the cell membrane, a key precursor to polymerisation monomer lipid II. It mediates the transfer of phospho-MurNAc pentapeptide from UDP-MurNAc pentapeptide to the membrane-bound lipid carrier undecaprenol phosphate (Fig. 1.5). Its genetic and functional characterisation has been reviewed in detail.[43]

MraY is part of a larger family of enzymes, all catalysing the transfer of a hexosamine phosphate from a soluble UDP-hexosamine donor to a membrane-bound lipid phosphate acceptor. Members of this family are termed UDP-*N*-acetylhexosamine:polyprenol phosphate *N*-acetylhexosamine-1-phosphate transferases and orthologues include enzymes involved in prokaryotic cell envelope biosynthesis such as WecA, TagO, WbcO, WbpL (lipopolysaccharides) and RgpG (rhamnose-glucose polysaccharide), as well as eukaryotic enzyme GPT, involved in protein *N*-glycosylation.[66,67] These enzymes share a common global structure and mode of action, although differ in their selectivity for the UDP-hexosamine substrate, and to some extent the nature of the lipid phosphate acceptor – which is undecaprenyl phosphate in prokaryotes and closely related dolichyl phosphate in eukaryotes.

MraY consists of ten transmembrane α-helices and has three conserved aspartic acid residues located on the second and fourth of five cytoplasmic loops.[43,68] These are mandatory for activity, and two of these residues are proposed to bind an essential Mg^{2+} cofactor while the third is the postulated active site nucleophile – D115, D116 and D267 respectively in *E. coli* MraY.[69] Other cytoplasmic loops are proposed to be involved in substrate specificity and binding.

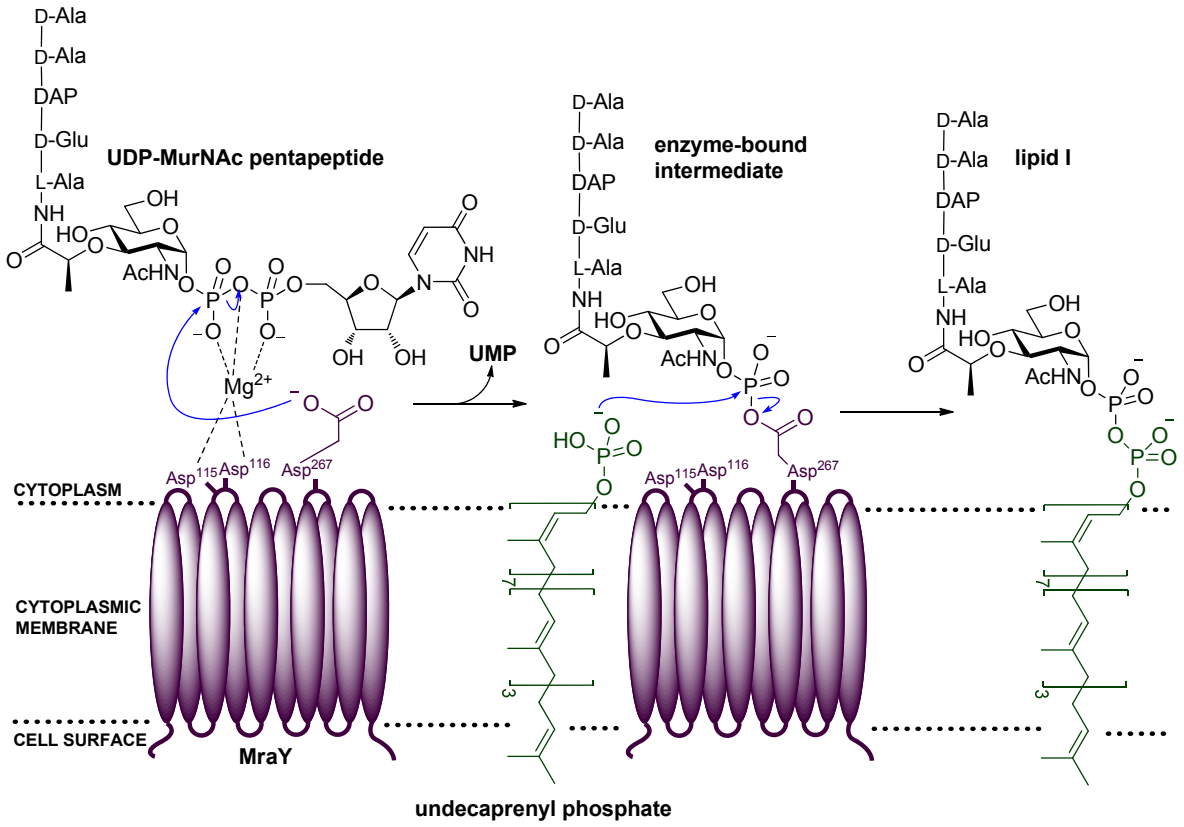


Fig. 1.5 Synthesis of lipid I catalysed by MraY

A two-step catalytic mechanism has been proposed, proceeding *via* an enzyme-bound intermediate and with attack at the phosphate α to the sugar residue in both stages (Fig. 1.5).[43,70] In the first step, the Mg^{2+} cationic cofactor interacts with the negatively charged pyrophosphate group of UDP-MurNAc pentapeptide and activates it as a leaving group, promoting nucleophilic attack at the phosphate group by the active site aspartate residue. Following UMP release, the second step involves a terminal hydroxyl on the membrane-bound undecaprenol phosphate substrate attacking the enzyme-phospho-MurNAc pentapeptide intermediate at the phosphorus atom. The resulting undecaprenol-pyrophosphoryl-MurNac pentapeptide product lipid I is subsequently released and the native enzyme regenerated. Although this mechanism has not been proved beyond doubt, two observations strongly support its validity: a) release of phospho-MurNAc pentapeptide has been observed, which would result from hydrolysis of the enzyme-bound intermediate;[43] b) the enzyme catalyses the exchange of [^{14}C]UMP with the UMP moiety of UDP-MurNAc pentapeptide with or without the presence of undecaprenyl phosphate.[71]

The study of MraY is hampered by its extremely low natural abundance and integral membrane localisation. A number of soluble, active preparations of MraY have been described from micrococci, *S. aureus* and *E. coli*, typically extracted with detergents and

activated by addition of phospholipids such as phosphatidylglycerol.[43] This means enzymatic assays have generally been performed with partially purified enzyme preparations. More recently, MraY from a number of sources has been overexpressed in *E. coli* which, together with optimisation of extraction detergents, has allowed it to be purified to homogeneity for the first time.[72] Substrate specificity experiments using polyprenol phosphate acceptor and UDP-MurNAc-pentapeptide donor substrates showed some plasticity as to the nature of the peptide side chain attached to MurNAc, reflecting the variations observed in these residues in peptidoglycan from different bacteria.[43] A number of assays using fluorescent or radiolabelled substrates have been developed, allowing enzyme kinetics to be analysed and inhibitors screened.[43]

1.2.2.4 Tunicamycin inhibits MraY and lipid I formation

Tunicamycins are reversible inhibitors of MraY, with K_i = 550 nM and IC_{50} = 44 µg/mL.[37,73-75] The inhibition was found to be competitive with respect to UDP-MurNAc-dansyl pentapeptide modified substrate, but non-competitive with respect to undecaprenyl phosphate.[74] Their structures act as bi-substrate analogues and mimic the transition state of the transfer reaction, blocking the active site. The tunicamycins have not seen use as antibacterial agents due to cytotoxicity in mammals associated with their inhibition of eukaryotic GPT involved in the first committed step of protein *N*-glycosylation (see section 1.2.3.4).[36]

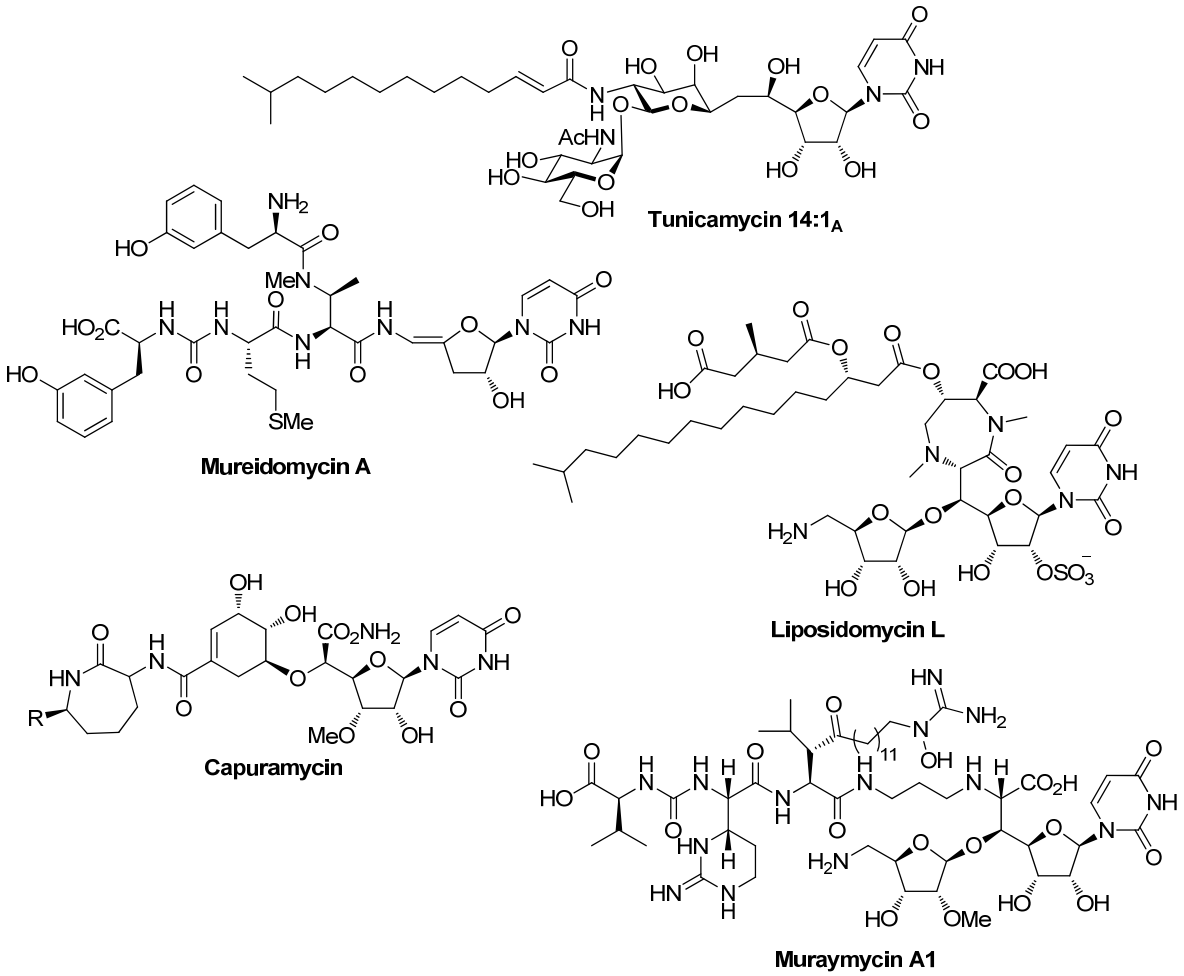


Fig. 1.6 Nucleoside antibiotic inhibitors of MraY

Other inhibitors of MraY typically consist of uridylpeptide and fatty acyl nucleoside antibiotics. The structure-activity relationships and inhibitory properties of these compounds and their semisynthetic analogues have been recently reviewed;[41,43,76] the main families of natural products involved are the mureidomycins, capuramycins, liposidomycins and muraymycins (Fig. 1.6). The mureidomycin family, which includes the pacidamycins and napsamycins, shows potent inhibition of MraY with no toxicity against mammalian cells. Mureidomycin A was shown to be a slow-binding inhibitor of *E. coli* MraY with K_i = 35 nM, K_i* = 2 nM and IC_{50} = 50 ng/µL. It is competitive against both enzyme substrates – UDP-MurNAc pentapeptide and undecaprenyl phosphate – as well as against cofactor Mg^{2+}. Aside from the mureidomycins mimicking the uridyl portion of the natural substrate, their terminal amine moiety has been proposed to bind to the Mg^{2+} cofactor binding site. The liposidomycin family also contains the caprazamycins, muraminocin and A-97065. They contain an additional acyl chain which targets the molecule to the membrane, where MraY is found, and which may mimic the undecaprenol phosphate

substrate. These compounds have been shown to act as slow-binding inhibitors of *E. coli* MraY with K_i* = 80 nM (K_i not determined) and are competitive with respect to undecaprenyl phosphate but non-competitive with respect to UDP-MurNAc pentapeptide. They were not cytotoxic when tested against mammalian cells. Members of the muraymycin and capuramycin families have also been shown to be potent inhibitors of MraY *in vitro*, with IC_{50} values of 27 and 17 ng/mL respectively. Unlike the tunicamycins, these nucleoside antibiotics all show greatly increased specificity for MraY over eukaryotic GPT and hence are not toxic to mammalian cells.

A further inhibitor of MraY has also been discovered which is not a nucleoside antibiotic. Bacteriolytic E protein from bacteriophage ΦX174 is a 91-amino acid protein containing a transmembrane domain.[77,78] It was proposed that E protein inhibits the formation of protein-protein interactions between MraY and other peptidoglycan assembly proteins in the cytoplasmic membrane. Inhibition was shown to be non-competitive with respect to both natural substrates, implying protein E binds outside of the MraY active site.[79]

1.2.3 Protein N-*glycosylation*

1.2.3.1 Structure of N-*linked glycans*

Carbohydrates, otherwise known as glycans, make up one of four essential building blocks in cells, alongside proteins, nucleic acids and lipids. They mediate a huge range of vital cellular processes, owing largely to their extraordinary structural diversity. Many glycans exist on the outside of cells as lipid and protein glycoconjugates and exert some of their most important functions throughout all stages of these secretory pathways. They provide structural and modulatory roles with wide-ranging effects including physical shielding and structural rigidity thanks to extracellular polysaccharide coats, stabilisation of proper protein folding controlled by nascent protein glycosylation and localisation of mature glycoprotein constructs to their final cellular destinations. They also function as recognition epitopes in a wide variety of cell-cell and protein-protein interactions, with mediation of vital events including innate immune system activation by glycan-binding lectins, nervous system modulation by gangliosides and embryonic development and differentiation regulation by glycosphingolipids. The major classes of glycoconjugates found in eukaryotic systems are *N*- and *O*-linked glycoproteins, glycosaminoglycans, glycolipids and glycosylphosphatidylinositol (GPI) anchors. There remains great structural and functional variation within each class, and each member has been extensively studied and comprehensively reviewed.[80,81] In the following sections we will focus on the structure, functions and biosynthesis of *N*-linked glycoproteins, list small molecule inhibitors of this pathway and highlight the important effect of tunicamycin in the perturbation and study of these processes.

The *N*-linked glycosylation of proteins is the most common covalent protein modification in all eukaryotes and many archaea, but is very rarely observed in bacteria. It involves the co-translational transfer of a conserved tetradecasaccharide precursor to an asparagine residue on targeted proteins, typically through recognition of a Asn-x-Ser/Thr motif. This conserved oligosaccharide is next trimmed and edited by a variety of glycosidases and glycosyltransferases, yielding an incredibly diverse array of protein-bound glycan structures. These constructs are classified into three main groups – oligomannose, complex and hybrid glycans (Fig. 1.7). All of these share a pentasaccharide core of $Man_3GlcNAc_2Asn$, derived from the common oligosaccharide biosynthetic precursor. The variations between these glycan groups stem from the different terminal branches of carbohydrates attached to this core. Oligomannose glycans typically contain 2-6 additional

mannose residues attached to the terminal mannose units of the pentasaccharide core. Complex glycans are typically more branched; between two and four additional branches stem from the core with a wide range of modifications observed. The most common of these are sialyl-lactosamine antennae, although others include a β-1→4-linked GlcNAc or an α-1→6-linked fucose attached to the branched mannose or the innermost GlcNAc of the core respectively, and sialic acid, fucose and galactose modifications to lactosamine branches. Hybrid glycans contain features from both oligomannose and complex structures. Typically these involve oligomannose structures on the α-1,6-Man arm and complex-type structures on the Manα1,3 arm, as well as frequent occurrence of a bisecting β-1→4-linked GlcNAc attached to the branching core mannose. There is significant structural overlap between these three groups and their functional roles are complex and under dynamic, temporal control. This has greatly complicated the study of their actions in living cells, although significant advances are being made in this multidisciplinary field of glycobiology. The structures and functions of specific *N*-linked glycans have been comprehensively reviewed.[81,82]

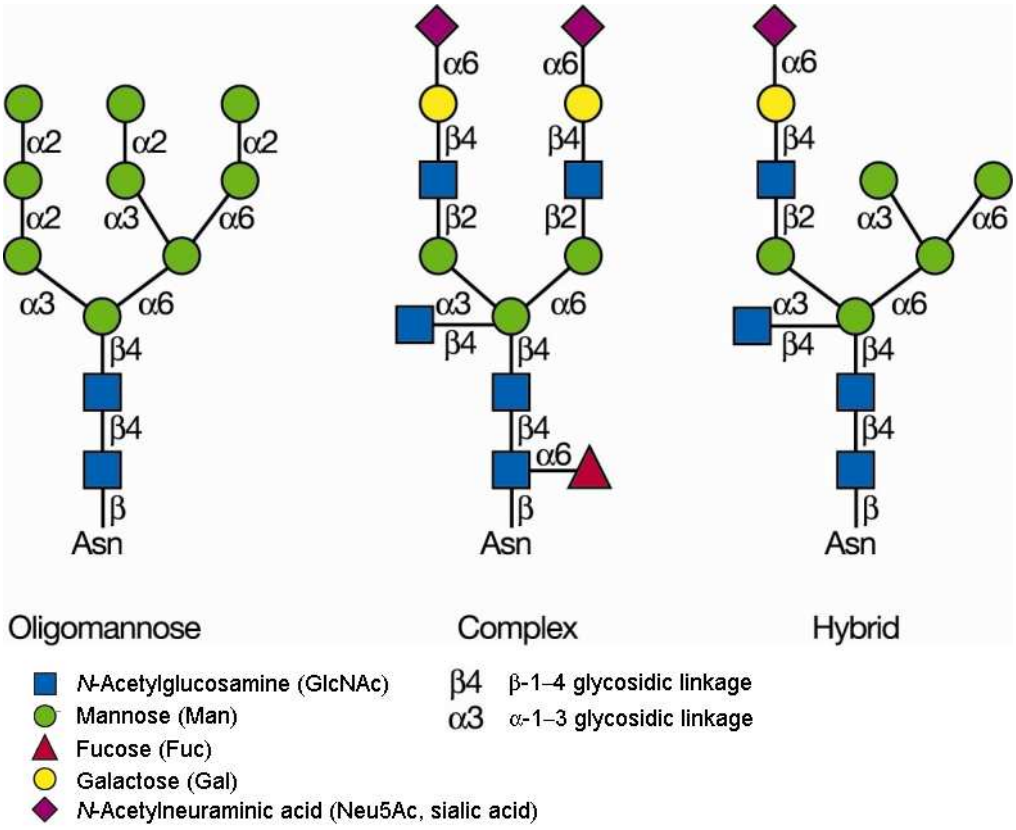


Fig. 1.7 Structures of major *N*-linked glycan classes.
(reprinted with permission from Varki *et al.*,[81] © 2008 Consortium of Glycobiology Editors, La Jolla)

1.2.3.2 Functions of N-linked glycans

One of the most important functions of *N*-linked glycans in eukaryotic cells is to promote the proper folding of growing polypeptide chains and newly synthesised proteins in the endoplasmic reticulum (ER), and accordingly this property has been well reviewed.[83,84] Inhibition of *N*-linked glycosylation is most often associated with the observation of misfolded and aggregated proteins, and of the resulting physiological effects. *N*-linked glycans have been shown to impose conformational preferences on the nascent polypeptide, assisting the folding process. One third of asparagine-linked glycans occur at compact β-turns in the folded protein, with the glycan helping induce this conformation during folding. That said, the reliance of successful protein folding on cotranslational *N*-glycosylation is entirely protein dependent; some proteins show no requirement for such modifications whereas others exhibit partial or even complete dependence. *N*-glycosylation can also maintain the solubility of intermediates in the protein folding process, mimicking molecular chaperones. However, once a protein has achieved its final, fully-folded state, the presence of an *N*-linked glycan is much less vital for stability and can often be safely removed without adverse effects.

An important associated role of asparagine-linked glycans is in the recruitment of proteins into the calnexin-calreticulin chaperone cycle in the ER.[81,83-86] Membrane-bound calnexin and soluble calreticulin are lectins found in the ER which transiently bind to nascent glycoproteins, assisting in the folding process. Knockouts of this pathway in transgenic mice have been shown to be lethal at the early embryonic stage. The lectins both bind specifically to monoglucosylated forms of the *N*-linked oligosaccharide with the form $Glc_1Man_{6\text{-}9}GlcNAc_2$, following the action of ER α-glucosidases I and II on the initial polypeptide-bound core tetradecasaccharide (Fig. 1.9). Binding of these lectins to appropriately tagged proteins can last between a few minutes and a number of hours, and the process partly functions to slow down the rate of protein folding with the result of improved folding efficiency. Calnexin and calreticulin also form a complex with ERp57, a protein disulfide isomerase which forms short-lived disulfide bridges between itself and the lectin-bound glycoproteins. It ensures disulfide bonds are formed in the appropriate order and with the correct partners on the way to a mature protein fold. Additionally, the complementary actions of soluble ER enzymes glucosidase II and UDP-glucose:glycoprotein glucosyltransferase (UGGT) control the calnexin-calreticulin binding cycle, further mediating the protein-folding process. Glucosidase II removes the final glucose residue required for lectin binding in the $Glc_1Man_{6\text{-}9}GlcNAc_2$ motif, causing steady

release of proteins from the calnexin-calreticulin chaperone system irrespective of whether they are correctly folded. Fortunately, UGGT recognises any misfolded proteins through their biophysical properties – such as exposed hydrophobic regions and excessive dynamic mobility – and reinstalls the necessary glucose residue, enabling re-entry into the calnexin-calreticulin cycle. This cycle embodies a quality control system for protein folding in the ER prior to transfer to the Golgi for further glycan processing, and is entirely dependent on the presence of a suitably processed *N*-linked glycan.

Specific *N*-linked glycan motifs also play a vital role in ER-associated protein degradation, where misfolded or mutant proteins are targeted to the cytosol for ubiquitination and 26S proteasome-mediated degradation.[81,83,84,87,88] The precise mechanisms of how this system operates are as yet unclear, although inhibition of ER mannosidase I significantly ablates this process, implying the product of this glycoside hydrolase – glycan motif $Glc_{0-3}Man_8GlcNAc_2$ – plays an important role as a sorting criteria for protein degradation.

Another important role of asparagine glycans is in transport and targeting of correctly folded proteins.[81,83,84] Several mannose-specific lectins have been identified which mediate glycan-dependent trafficking of glycoproteins from the ER to the Golgi complex. Additionally, the many processing enzymes resident in the ER, Golgi and lysosome membranes themselves contain characteristic *N*-glycan motifs, recognised by specific lectins and targeted to the relevant cellular compartments. Selective targeting of lysosomal hydrolases is particularly well studied.[81] Introduction of a 6-phosphate group onto particular terminal mannose residues in trimmed *N*-linked oligosaccharides targets these glycans for recognition by mannose-6-phosphate lectin receptors in the trans-Golgi network, which sequester these glycoproteins and escort them *via* vesicles to endosomes and lysosomes. Binding to the lectin is pH-dependent, and the glycoprotein cargo is selectively released in the low pH environment of endosomes and lysosomes.

The Golgi complex is responsible for introducing the great diversity of terminal glycan modifications found in different glycoproteins. These specifically tailored glycan structures help to correctly localise the mature protein to its relevant cellular compartment or onto the cell surface. Once in place, the *N*-glycans of these glycoproteins have been shown to be essential in a wide variety of physiological functions. The molecular basis for many of these functions as well as the roles of most terminal glycan modifications have not yet been elucidated, but the better studied examples have been subject to excellent reviews.[81,82,89,90] Some examples where *N*-linked glycans play important roles can be found

in cellular metabolic enzymes, hormones and cytokines, viral glycoproteins, plasma proteins and coagulation factors, and a wide variety of membrane proteins. These roles vary between structural, protective and stabilising functions, or as antigens for lectin receptors from numerous physiological pathways.

1.2.3.3 Biosynthesis of N-linked glycans

The biosynthetic pathway to *N*-linked glycans incorporates many of their most important functions related to mediation of protein folding. The pathway can be resolved into two stages taking place in a combination of the endoplasmic reticulum and Golgi apparatus, both of which have been comprehensively reviewed.[81,82,91]

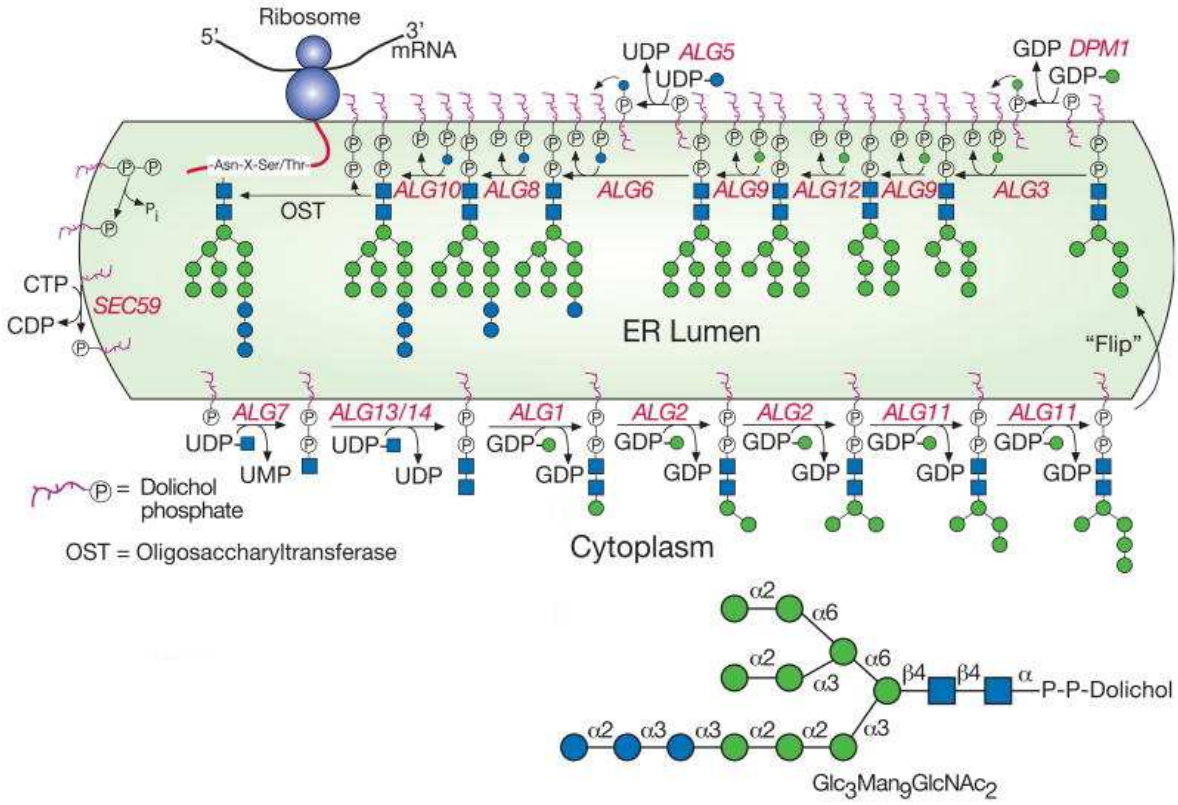


Fig. 1.8 The dolichol pathway – synthesis of precursor oligosaccharide $Glc_3Man_9GlcNAc_2$-PP-Dol. (reprinted with permission from Varki *et al*.,[81] © 2008 Consortium of Glycobiology Editors, La Jolla)

The first of these stages is called the dolichol pathway and involves the synthesis of a lipid-linked oligosaccharide with a conserved tetradecasaccharide structure (Fig. 1.8).[81,82,91] Construction of this $Glc_3Man_9GlcNAc_2$-PP-Dol precursor commences from dolichol phosphate (P-Dol), a membrane-bound polyisoprenol lipid (Fig. 1.8). The enzymes involved in glycosyl transfer to this lipid have been identified primarily from genetic mutants of the yeast *Saccharomyces cerevisiae*, and the genes involved have been named

ALG genes from 'altered in glycosylation'. A GlcNAc-1-phosphate residue is transferred from UDP-GlcNAc to P-Dol on the cytoplasmic face of the ER by the action of the gene product of *ALG7* – also known as GlcNAc-1-phosphate transferase (GPT) – to yield dolichyl-pyrophosphoryl-*N*-acetylglucosamine (GlcNAc-PP-Dol). A second GlcNAc residue is transferred from UDP-GlcNAc, followed by five mannose (Man) residues from GDP-mannose. The resulting $Man_5GlcNAc_2$-Dol-PP is next flipped from the cytoplasmic face to the ER lumen, where subsequent transfer reactions utilise membrane bound enzymes and lipid-linked monosaccharide donors. Four more mannose residues are added to the oligosaccharide from Man-P-Dol donors, followed by three glucose (Glc) residues from Glc-P-Dol. The resulting 14-sugar oligosaccharide is a common precursor to all mature *N*-linked glycan structures and undergoes *en bloc* transfer from Dol-PP to asparagine residues of growing polypeptide chains through recognition of a conserved Asn-X-Ser/Thr sequon. This event is catalysed by multisubunit enzyme oligosaccharyltransferase (OST) and occurs on the luminal face of the rough endoplasmic reticulum while the protein is being synthesised by ER-bound ribosomes and is threading through the translocon in the ER membrane.

The second stage in the biosynthetic pathway of *N*-linked glycans follows directly after the attachment of the conserved oligosaccharide to the nascent protein.[81,82,92] The 14-sugar glycan undergoes a variety of trimming reactions in the ER and Golgi apparatus, starting with the removal of the three glucose residues by ER α-glucosidases I and II. At this point the nascent glycoprotein is intimately associated with the calnexin-calreticulin cycle and the ER-associated degradation pathway as described previously. Successfully folded proteins resulting from these quality control measures have a terminal mannose residue from their glycan's central arm removed by ER α-mannosidase I, before being transferred to the *cis* compartment of the Golgi. Here, most glycoproteins have additional mannose residues removed by the action of α-1,2 mannosidases IA, IB and IC to yield a $Man_5GlcNAc_2$-Asn intermediate, with the exception of glycans destined to form oligomannose-type structures. Following transfer to the *medial*-Golgi, glycosyltransferase GlcNAcT-I transfers a GlcNAc residue from UDP-GlcNAc to the α-1,3-arm mannose. α-Mannosidase II can subsequently remove the two outer mannose residues generating $GlcNAcMan_3GlcNAc_2$-Asn and leading to complex-type glycans. Hybrid *N*-glycans are formed when α-mannosidase II does not act on this glycan.

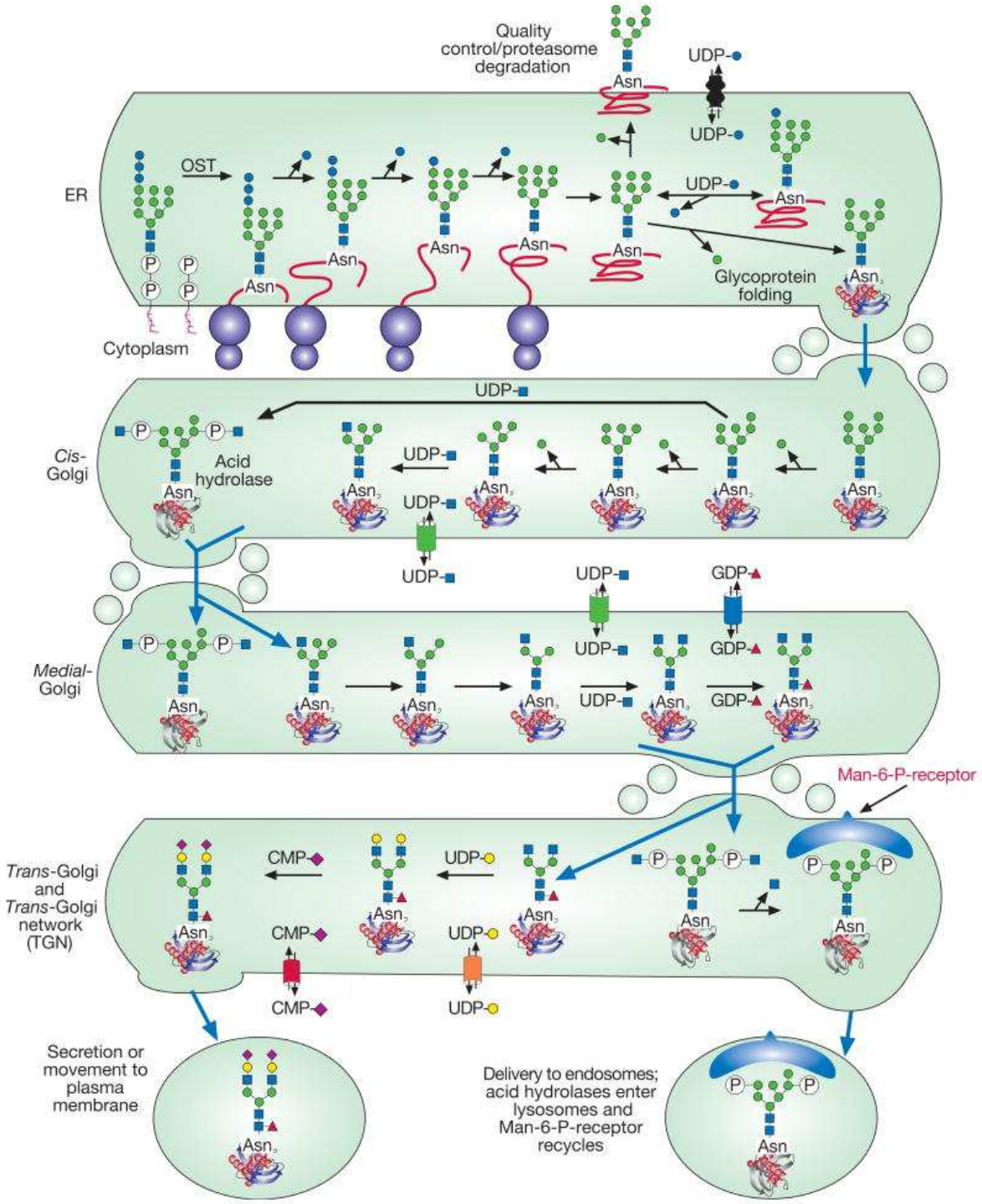


Fig. 1.9 Processing and maturation of N-linked glycans in the ER and Golgi (reprinted with permission from Varki *et al.*,[81] © 2008 Consortium of Glycobiology Editors, La Jolla)

Diversification of these intermediates follows, with addition of further GlcNAc, galactose, fucose, mannose and sialic acid residues to terminal and internal sugars in the oligosaccharides occurring in the *medial*- and *trans*-Golgi. This results in complex and hybrid glycans with a wide range of structural modifications, including a variety of terminal

sugar moieties, multiple branches and additional residues attached to the core. These glycoproteins are then trafficked to the plasma membrane, internal organelles or for secretion. Fig. 1.9 additionally shows the modification of lysosomal hydrolase glycans to form terminal Man-6-P residues facilitating the specific endosomal localisation discussed previously.

It should be noted that glycoproteins are predominantly heterogenous, which means that proteins sharing a polypeptide backbone can be decorated with a variety of different *N*-linked glycans. Additionally, if more than one Asn-X-Ser/Thr sequon exists in a protein sequence, populations with glycosylation on each different or multiple sequons can exist, alongside variation in the glycan structure which is attached. These different forms of the same DNA-encoded protein are termed glycoforms and arise from a number of dynamic factors, including local protein conformations, transport rates through different ER and Golgi compartments, the state of nucleotide sugar metabolism, competition between processing enzymes for the same glycoprotein substrate and the proximity of the asparagine sequon to the transmembrane domain.[81]

1.2.3.4 Tunicamycin inhibits N-linked glycoprotein biosynthesis

Tunicamycin has become a very important tool for the study of glycoproteins in a wide variety of biological systems. It inhibits the first committed step in the dolichol pathway of *N*-glycosylation, namely the formation of membrane-bound *N*-acetylglucosaminyl-pyrophosphoryl dolichol (GlcNAc-PP-Dol).[36,93,94] As such, it terminates protein *N*-glycosylation at its earliest stage, allowing the physiological effects of these important post-translational modifications to be studied. Additionally, tunicamycin is used in cell culture applications to ensure homogeneity of target proteins produced using mammalian cell lines.[95,96]

The enzyme it targets is GlcNAc-1-phosphate translocase (GPT), which catalyses the transfer of GlcNAc-1-P from UDP-GlcNAc to lipid carrier dolichol phosphate (Fig. 1.10).[93] GPT is the only eukaryotic member of a family of transmembrane proteins known as the UDP-D-*N*-acetylhexosamine:polyprenol phosphate D-*N*-acetylhexosamine-1-phosphate transferases.[66,67] All other members are from bacterial sources and are involved in cell envelope biosynthesis; they consist of MraY (peptidoglycan biosynthesis), WecA, TagO, WbcO, WbpL (lipopolysaccharide biosynthesis) and RgpG (rhamnose-glucose polysaccharide formation). The structural properties and mechanism of MraY were discussed in section 1.2.2.4 and GPT shares many of these features. Briefly, it has ten

transmembrane domains with five cytoplasmic loops, housing a DDxxD Mg^{2+}-binding motif and an active site nucleophilic aspartate residue. Catalysis proceeds through a double displacement mechanism, forming an enzyme-bound intermediate between the active site aspartate residue and the phosphorus of GlcNAc-1-phosphate, before displacement by incoming dolichol phosphate.

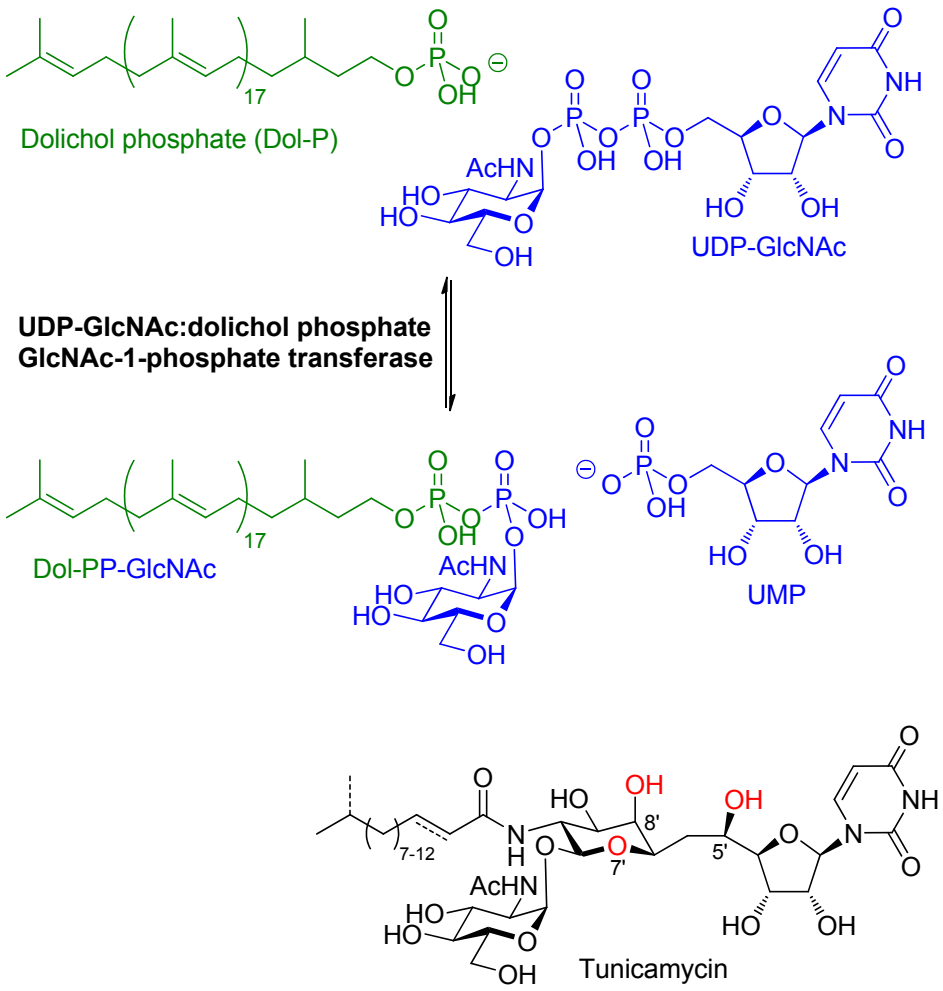


Fig. 1.10 Relation of tunicamycin structure to reaction catalysed by GPT

GPT, like most UDP-HexNAc:Dol-P HexNAc-1-P transferases, has resisted successful over-expression or purification to a homogenous form, owing to its properties as an intrinsic membrane protein. As such, detailed kinetic parameters for tunicamycin binding are not available, although inhibition of the enzyme have been studied using solubilised and partially purified preparations from pig aorta,[93] as well as membrane and microsomal fractions from yeast and calf liver.[36,94] Tunicamycin is a potent inhibitor of GPT with an IC_{50} of 0.05-0.1 µg/mL and inhibition has been reported to be irreversible and noncompetitive with respect to both substrates.[93] In contrast, its inhibition of MraY is

approximately 500-fold lower at 44 µg/mL and is competitive with respect to the nucleotide substrate but noncompetitive with regard to the lipid phosphate.[74] These differences are surprising given the similarity of the two enzymes and the equivalent mode of action of tunicamycin in both cases. The dissimilarities observed may be a result of tunicamycin mimicking the pentapeptide-containing UDP-MurNAc used by MraY less well than the UDP-GlcNAc substrate of GPT. It should be noted that the IC_{50} values reported for both enzymes use protein samples of very different purity; a homogenous preparation of GPT using modern protein purification techniques would likely provide a more comparable result.

The potency of tunicamycin derives from its close structural similarity to the UDP-GlcNAc substrate of GPT (Fig. 1.10). It binds to the active site of the enzyme much more strongly than the native substrate, and may act as a transition state mimic. Molecular modeling and ^{1}H and ^{13}C NMR studies – including rotating angle nuclear Overhauser effect (ROESY) analysis – of native and deuterated tunicamycins have established their preferred conformations, and offer insights into active site binding.[97] The uracil and GlcNAc portions of tunicamycin clearly mimic the same moieties in UDP-GlcNAc, but the central tunicamine residue is also proposed to mimic the pyrophospho-ribose of the native substrate. 5´-OH, 8´-OH and the 7´-O ring oxygen are predicted to coordinate with the active site Mg^{2+} cation.

1.2.3.5 Other inhibitors of protein N-glycosylation

A number of compounds other than tunicamycin have been found which inhibit different stages of the *N*-glycosylation pathway. These have been well reviewed, but the most important inhibitors will be summarised below.[81,96,98]

In the dolichol pathway, a number of natural product antibiotics have been shown to interfere with the formation of the conserved lipid-linked oligosaccharide. Peptide antibiotics amphomycin, tsushimycin and showdomycin inhibit the formation of dolichol-linked monosaccharide precursors, while bacitracin has shown a spectrum of activity against various steps in the pathway, derived from the formation of a tight complex with lipid-carrier dolichol. Diumycin inhibits the formation of GlcNAc-PP-Dol and Man-PP-Dol building blocks, but has also been shown to block the transfer of the second GlcNAc in the formation of $GlcNAc_2$-PP-Dol. A variety of sugar analogues have also been observed to interfere with the dolichol pathway. These are typically deoxy- or fluoro- analogues of D-glucose and D-mannose, and include 2-deoxy-D-glucose, 2-deoxy-2-fluoro-D-glucose, 2-

deoxy-2-fluoro-D-mannose, 4-deoxy-4-fluoro-D-mannose and 4-deoxy-D-mannose. All of these sugars are first metabolised to their respective nucleoside diphosphate esters in the cell, consequently acting as substrate mimics and thus competitive inhibitors of the enzymes involved in forming dolichol-linked monosaccharide precursors and the subsequent glycosyl transfer reactions in the dolichol pathway. An exception is D-glucosamine which exerts its action directly, showing inhibitory activity against mannosyltransferase enzymes involved in constructing the lipid-linked oligosaccharide. It should be noted that many of these sugar analogues can exhibit a myriad of cellular effects derived from metabolic conversion into other reactants not related to glycoprotein biosynthesis, and care should be taken when interpreting experimental data. As can be seen in Fig. 1.11, these compounds typically exhibit broad activity across a number of enzymes in the dolichol pathway. Tunicamycin is the only exception, specifically targeting GPT and the formation of GlcNAc-PP-Dol.

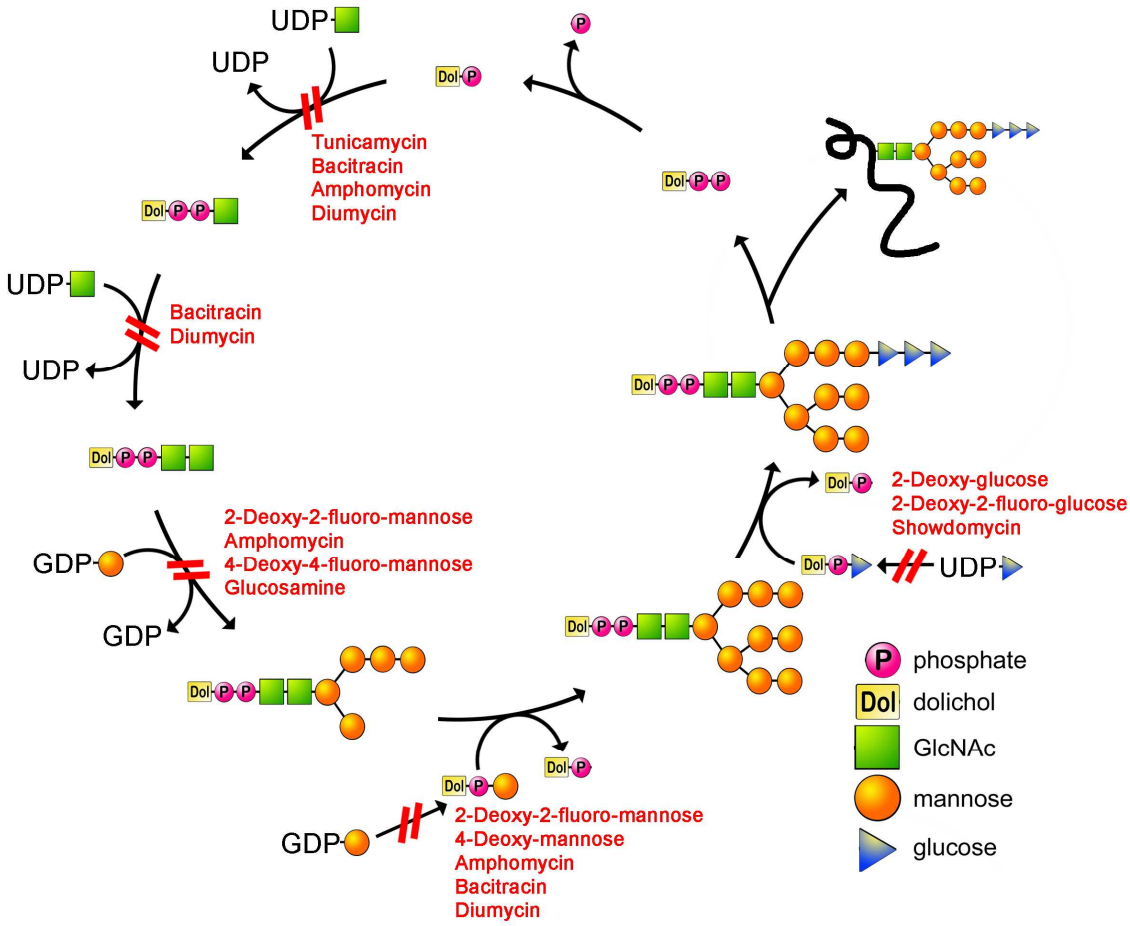


Fig. 1.11 Inhibitors of the dolichol pathway of N-linked glycosylation

A variety of inhibitors have been discovered targeting the trimming pathway of protein *N*-glycosylation, following transfer of the conserved oligosaccharide to the nascent polypeptide chain.[81,96,98] These compounds are typically polyhydroxylated alkaloids with structures resembling the transition states of trimming pathway glycosidases, mimicking the hydroxyl group orientations and ring skeleton conformations of the natural substrates. One class of compounds targets α-glucosidases involved in the initial processing of the 14-sugar precursor and associated protein folding chaperone systems. Notable examples from this class are deoxynokiromycin, castanospermine and australine, which target α-glucosidases I and II with differing, but not exclusive specifities. Treatment of cells with these compounds led to the accumulation of fully or partially glycosylated oligosaccharide chains, although this did not cease subsequent trimming by α-mannosidases in many cases. The second class of glycosidase inhibitors targets α-mannosidase trimming enzymes. Deoxymannojirimycin targets α-mannosidases I, whereas swainsonine and mannostatin A target α-mannosidase II; in both cases a build up of high-mannose and hybrid *N*-glycans is observed, at the expense of complex glycans. Some of these compounds have shown positive effects as drugs in the treatment of diabetes, lysosomal storage disorders, cancer metastasis and even HIV infection, and remain an area of active research.[81,92]

1.3 Chemical synthesis of the tunicamycins

1.3.1 Total synthesis of natural products

The chemical synthesis of natural products has been an important area of research ever since the total synthesis of urea in 1828.[99] Hundreds, if not thousands, of natural products of ever-increasing complexity have since been synthesised, many of which offering valuable contributions to the study and treatment of human disease.[100] Such synthetic studies play a number of functional roles. Firstly, they can be used to validate the precise structural assignment of a natural product, by comparing well defined synthetic products and intermediates with authentic samples and their degradation products. Furthermore, some natural products are isolated in minute quantities by unculturable organisms, and a synthetic route to their production can provide sufficient quantities for testing the myriad of biological activities they possess. Functional analogues can be accessed *via* these routes to further probe the biological roles of these compounds and assess structure-activity relationships. Finally, such studies offer great contributions to the field of organic chemistry, furthering our understanding of fundamental reactivity and chemical theory, and also providing ingenious synthetic methodologies that often extend beyond those employed by nature. Vast progress has been made in the field of organic chemistry, and armed with a diverse toolbox of synthetic reactions chemists can now access even the most challenging natural product scaffolds. This broad field remains a vital area of research, and continues to complement the incredible advances made in areas such as biochemistry, genomics and therapeutic medicine.

A number of synthetic studies towards the tunicamycins have been published, with two full syntheses reported. In this section we will summarise these efforts and discuss them in the context of producing tunicamycin analogues.

1.3.2 Suami (1984): Total synthesis of tunicamycin

Between 1981 and 1988, the laboratory of Tetsuo Suami in Japan published a series of papers on the synthesis of tunicamycins,[101-107] culminating in the first total synthesis of tunicamycin in 1984.[102] Native tunicamycin homologues III, V, VI and VIII – later renamed $14{:}1_B$, $15{:}1_B$, 15:0 and $16{:}1_B$ respectively – were synthesised, together with eight fatty acyl analogues. A subsequent publication presented the synthesis of an *N*-acetylgalactosamine (GalNAc) analogue.[108]

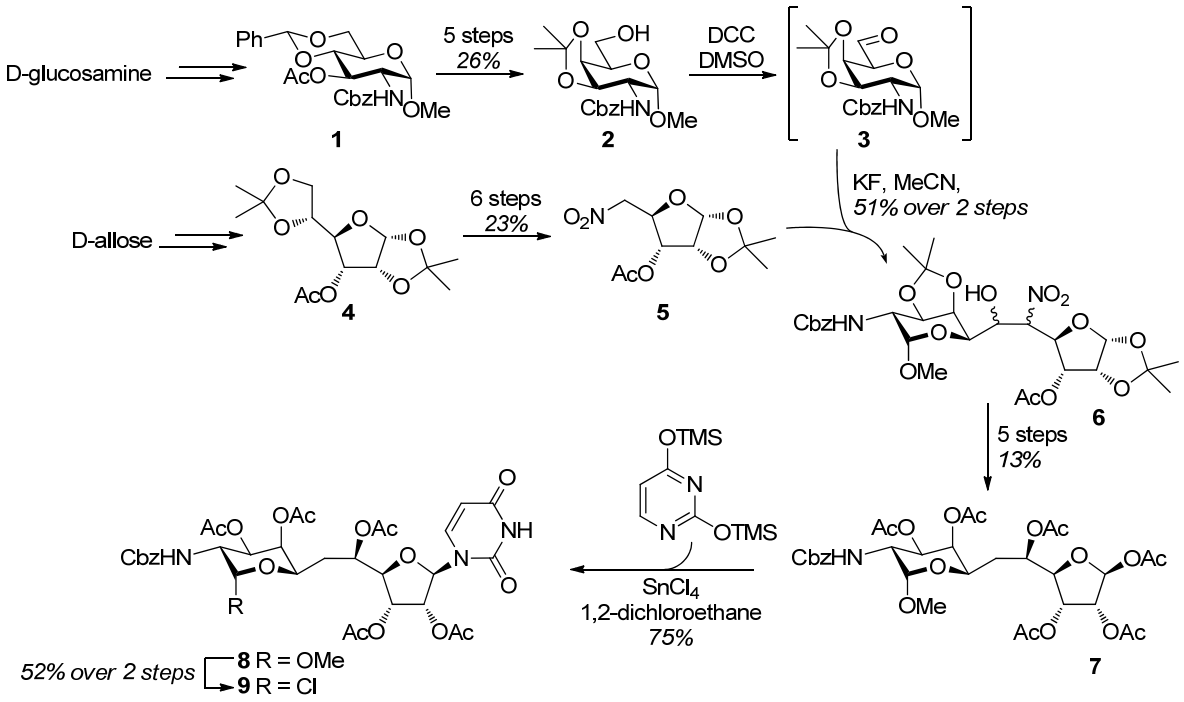


Fig. 1.12 Suami's synthesis of tunicaminyl-uracil precursor

The synthesis began with the construction of the tunicamine undecose skeleton, its salient feature being a Henry reaction between 5-nitro-ribose **5** and galactosamine aldehyde derivative **3** (Fig. 1.12).[107] The 5-nitro-ribose precursor was derived from known, protected D-allofuranose **4**[109] through periodic acid-mediated oxidative cleavage of the terminal CH_2OH followed by S_N2 displacement of a subsequent 5-iodo intermediate with sodium nitrite.[107] The aldehyde coupling partner was accessed from known, protected D-galactosamine derivative **1**[110] *via* S_N2 inversion of an activated 4-*O*-tosyl intermediate – to yield a *galacto*- conformation – together with accompanying protecting group manipulations, and followed by Pfitzner-Moffatt oxidation of the C–6 hydroxyl to the requisite aldehyde in **3**.[105] Following the fluoride-mediated coupling of these two partners, the resulting nitro undecose **6** was dehydrated and hydrolysed yielding a mixture of C-5 isomers. Further protecting group manipulations ensued and the desired isomer of tunicamine derivative **7** was isolated in 13% yield over the 5 steps – together with 25% of its epimer. A uracil moiety was next introduced into this sugar, yielding tunicaminyl-uracil **8**, which was subsequently activated as an anomeric chloride (**9**).[102,107]

The previously constructed tunicaminyl-uracil chloride **9** was used as a glycosyl donor for construction of the requisite α,β-1,1 glycosidic bond in tunicamycin (Fig. 1.13).[102,106] The glycosyl acceptor **11** was derived from allyl 2-acetamido-2-deoxy-

glucoside **10** through standard protecting group manipulations. The glycosyl transfer was catalysed by silver perchlorate and silver carbonate, yielding 18% of the desired α,β anomer and 15% of the β,β by-product. The α,β anomer was converted to free amine **13**, and acylated in moderate yield with fatty acid **16** and *N*,*N*´-dicyclohexylcarbodiimide (DCC). This fatty acid was accessed by addition of (2-methylbutyl)zinc chloride **14** to 9-ethoxycarbonylnonanoyl chloride **15**, followed by Clemmensen reduction and α,β-dehydrogenation of the resulting acyl ethyl ester intermediate. Final deprotection of the acylated intermediate gave compound **17**, analytically identical to an authentic sample of tunicamycin V (Fig. 1.13).[102,106]

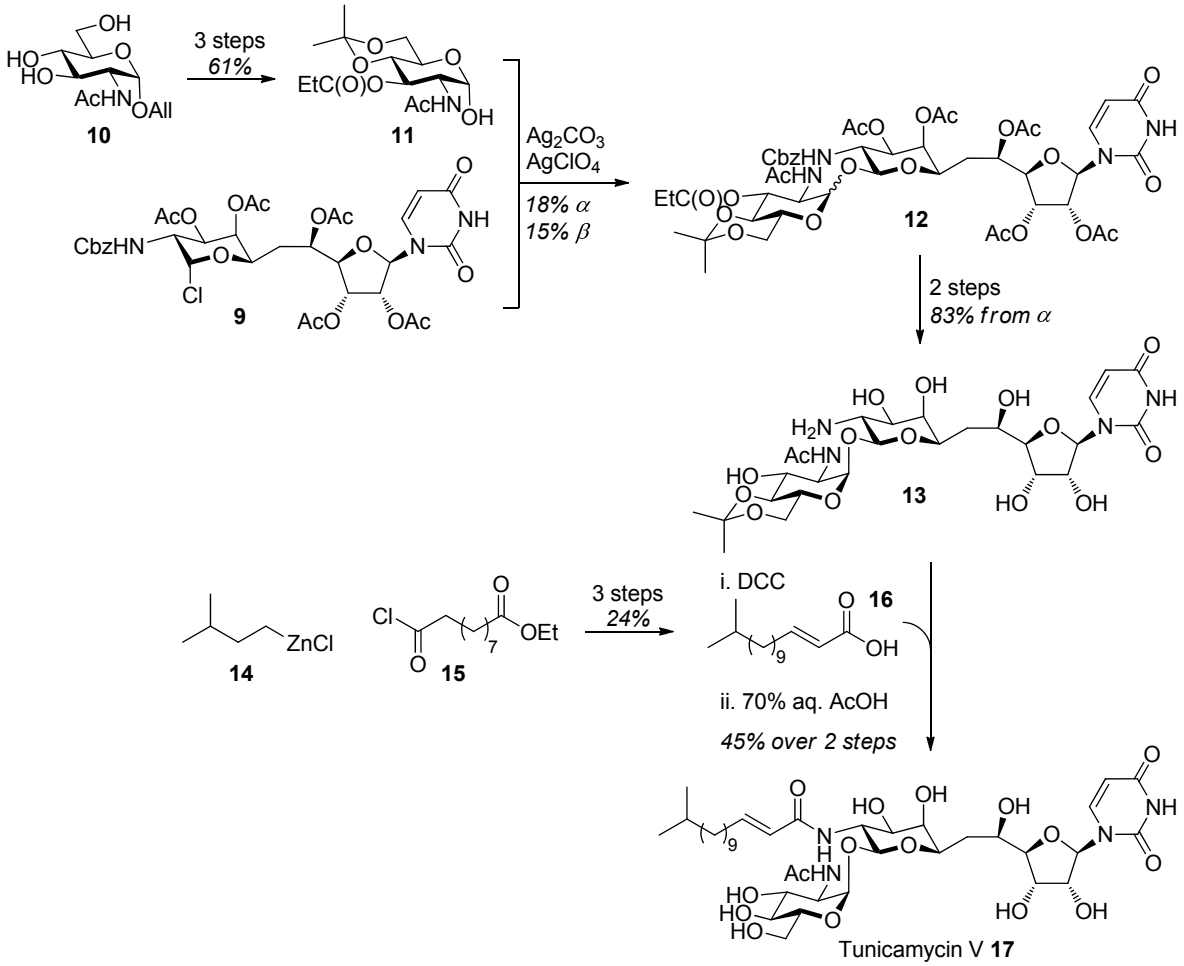


Fig. 1.13 The end stages of Suami's synthesis of tunicamycin V

The synthesis described by Suami *et al.* represents the first total synthesis of any tunicamycin homologues. It successfully overcomes the challenges presented by the tail-to-tail coupling of two monosaccharide units in tunicamine and also the rare α,β-trehalose linkage between sugars. However, the synthesis is lengthy and contains a large number of protecting group manipulations which collectively decrease the efficiency of the synthesis.

Other notable concerns are the particularly low yields and stereoselectivity observed during the nitrate hydrolysis sequence (13%) and the glycosyl transfer reaction (18% desired epimer). Additionally the moderate 45% yield observed during acylation and final deprotection is costly at such a late stage in the synthesis. There is much potential for improvement to this synthesis, although the work represents a vital contribution, enabling the construction of unnatural analogues of tunicamycin for the first time. The eight fatty acyl variants exhibited biological activity against Newcastle disease virus and the formation of glycoconjugates and lipid-linked intermediates.[102] The terminal GalNAc analogue showed no antiviral or antimicrobial activity and moderate inhibition of GlcNAc-PP-dolichol formation.[108]

1.3.3 Danishefsky (1989): De novo *synthesis of tunicaminyl-uracil*

Two publications from Samuel Danishefsky's laboratory at Yale University were published in the late 1980s, offering a very contrasting route to the tunicaminyl-uracil scaffold within tunicamycin.[111,112] Instead of using well known and commonly available monosaccharide precursors, the authors chose to undertake a *de novo* synthesis of these structures, utilising methodology developed in their laboratory involving cyclocondensation of aldehydes with siloxydienes.[113]

The synthesis began with commercially available uridine derivative **18** (Fig. 1.14). Selective protecting group introduction and oxidation to a 5´-aldehyde (**19**) was followed by strereoselective allylation and protection of the resulting homoallylic alcohol (**20**). The terminal alkene was converted to an aldehyde by dihydroxylation and subsequent oxidative diol cleavage. Aldehyde **21** then took part in the key coupling step with siloxydiene **22**,[114] consisting of a cyclocondensation with subsequent loss of methoxide to yield dihydropyrone **23**. This compound was a mixture of stereoisomers, although subsequent reduction of the ketone and protecting group manipulation allowed isolation of glycal **24** as the major isomer with the desired stereochemistry. Diversification of this glycal was afforded by cycloaddition with dibenzyl azodicarboxylate **25** under photochemical control, yielding **26**. This adduct was finally converted to differentially *N*-protected tunicaminyl-uracil derivates **27** and **28** by anomeric displacement with MeOH and Raney Ni-catalysed hydrogenolysis, followed by standard protecting group manipulations (Fig. 1.14). Analytical data for acetamide **27** exactly matched that of a naturally-derived degradation product, confirming the successful synthesis of the tunicaminyl-uracil scaffold.

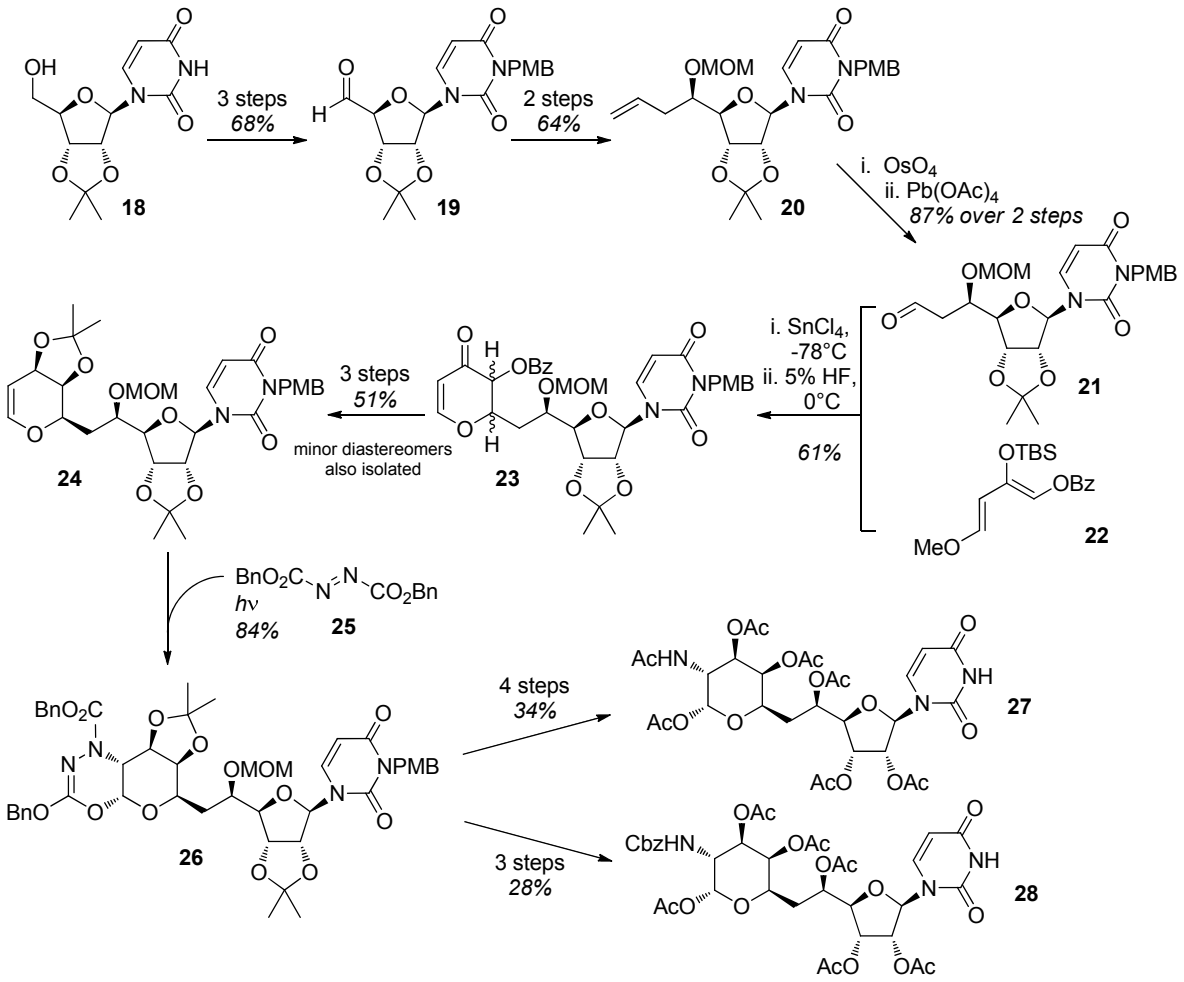


Fig. 1.14 Danishefsky's de novo route to tunicaminyl-uracil

The synthetic route presented by Danishefsky *et al.* is very different to all other published syntheses, and involves *de novo* synthesis of monosaccharide motifs within the molecule. The yields are moderate, but offer an improvement over Suami's synthesis. However, there are still many costly protecting group manipulations and losses associated with moderate stereoselectivities of key steps. The *de novo* construction of the galactosamine moiety in tunicamine may be a good exemplification of synthetic methodology developed within this laboratory, but it remains cumbersome considering the rich stereochemical information possessed by and the widespread availability of alternative monosaccharide precursors. An interesting note is that the authors described serious difficulties replicating the glycosyl coupling reaction described by Suami *et al.*.[102,106] They suggest that the original authors met success by performing the reaction on a large scale using relay synthesis from fermented tunicamycin, whereas the reaction is not amenable on a laboratory scale.

1.3.4 Zamojski (1992): Synthesis of protected tunicamine

The third set of studies towards synthesis of the tunicamycins was published in 1992 by the laboratory of Aleksander Zamojski in Poland.[115,116] In contrast to Danishefsky's synthesis where the pseudo-galactosamine moiety was constructed *de novo* onto a ribose precursor, the approach utilised here was to start with a D-galactosamine precursor and build the pseudo-ribose moiety of tunicamine onto it using a non-carbohydrate precursor (Fig. 1.15).

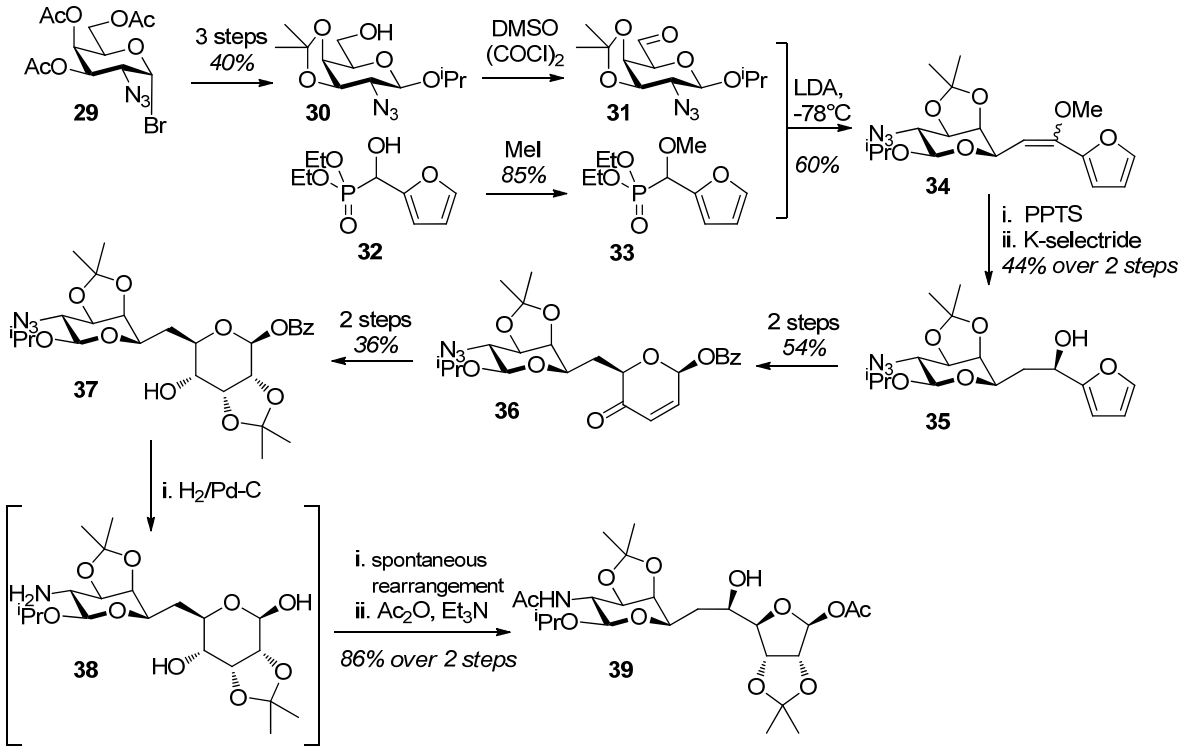


Fig. 1.15 Zamojski's synthesis of fully-protected tunicamine

The first key bond-forming step was a Horner-Wadsworth-Emmons condensation of furyl phosphonate **33** and D-galactosamine aldehyde **31**, to yield the 11-carbon chain adduct **34** (Fig. 1.15). The galactosamine aldehyde was accessed from known D-galactosamine derivative **29**[117] through a series of protecting group manipulations followed by a Swern oxidation of the C-6 hydroxyl group. The phosphonate ester was derived by methylation of known hydroxyl precursor **32**.[118] Following the coupling of these two units, enol ether **34** was hydrolysed and the subsequent ketone stereoselectively reduced with K-selectride to yield alcohol **35**. In another key transformation, this furfuryl alcohol was next converted to a 'ribose' moiety through the approach of Achmatowicz *et al.*.[119,120] Treatment of the furfuryl alcohol with Br_2 in MeCN/H_2O provoked rearrangement to an unstable pyran-4-

ulose intermediate which was benzoylated (**36**) and subsequently derivatised to a polyhydroxylated pyranose structure bearing the appropriate stereochemistry at each carbon centre (**37**). Deprotection of the anomeric hydroxyl by catalytic hydrogenation was followed by spontaneous isomerisation to the ribofuranose conformation. Partial acetylation led to the final compound **39**, a protected derivative of undecose tunicamine.

The synthetic strategy of Zamojski *et al.* resulted in the production of a fully protected tunicamine derivative from D-galactosamine and furan precursors. Unfortunately this intermediate was not embellished any further to more advanced tunicaminyl-uracil or tunicamycin scaffolds. As such it represents an incomplete story, but it does offer a new route to higher sugar construction *via* a Wittig-type condensation, plus an interesting rearrangement of a furfuryl alcohol to a 'ribose' moiety. Once again, the key chain extension event suffers from a moderate yield, compounded by the subsequent ketone reduction which has good (3.3:1) but incomplete stereoselectivity for formation of the desired C-5 epimer (**35**). Moreover, the formation of D-galactosamine precursor **29** itself from D-galactal is a far from efficient process[117] and is followed by further costly protecting group manipulations, reducing the overall efficacy of this synthesis.

1.3.5 Myers (1993): Total synthesis of tunicamycin

Between 1991 and 1994, Andrew P. Myers and then graduate student David Y. Gin published a series of papers documenting the synthesis of tunicaminyluracil[121] and a total synthesis of tunicamycin V[122] and a C5´-*epi*-tunicamycin V analogue.[123] The authors revert to using monosaccharide-derived precursors, and construct the tunicamine core through a silicon-mediated reductive coupling of an aldehyde and an allylic alcohol (Fig. 1.17). They also report a new method of constructing the α,β-1,1 glycosidic bond in tunicamycin, which proved much more efficient than Suami's Koenigs-Knorr route (Fig. 1.16).

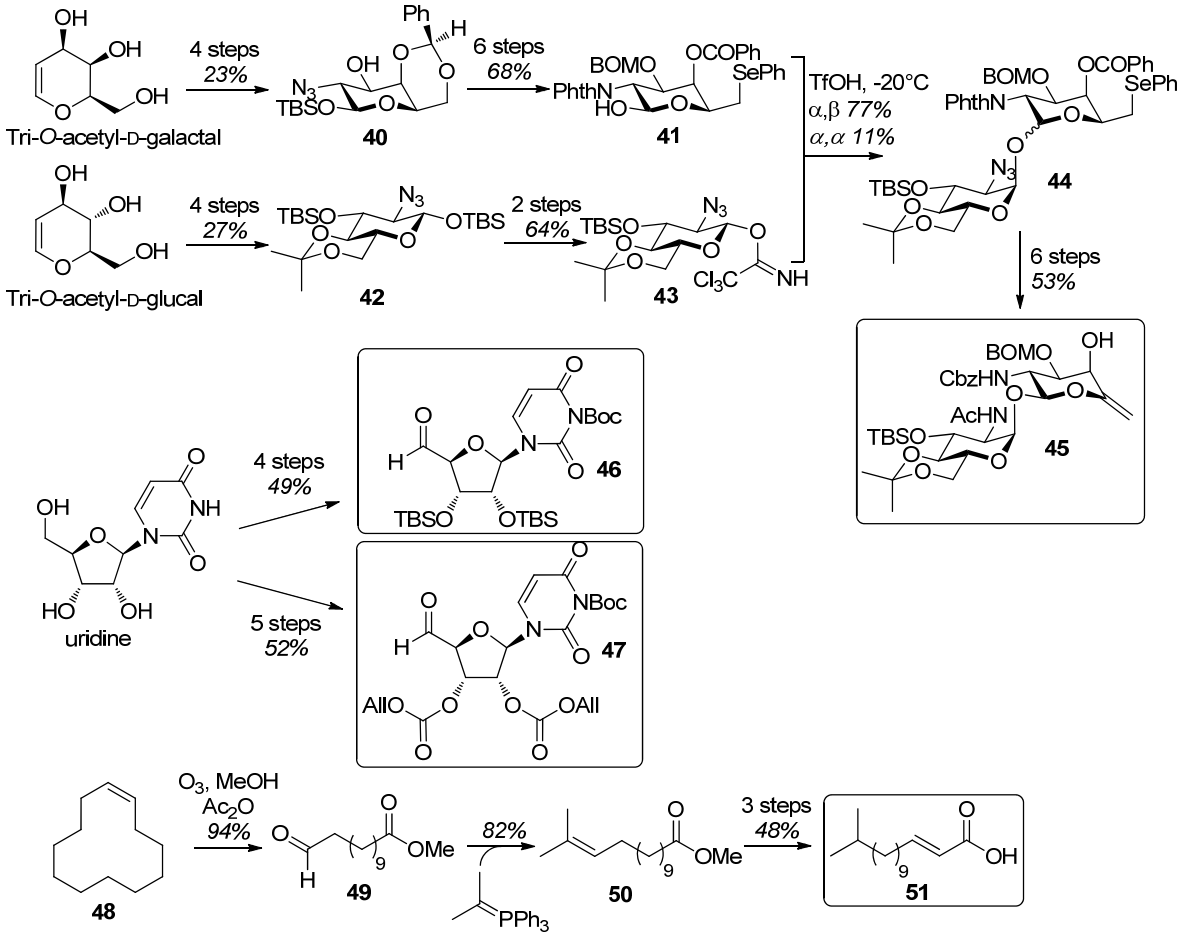


Fig. 1.16 Myers' synthesis of α,β-trehalose linkage and key intermediates en route to tunicamycin

The construction of all the precursor units and formation of the α,β-trehalose linkage is summarised in Fig. 1.16. The glycosidic bond in question was installed prior to construction of the tunicamine skeleton. As such, the coupling partners were derived from monosaccharide building blocks; glycosyl acceptor **41** was synthesised from tri-*O*-acetyl-D-galactal over 10 steps and glycosyl donor **43** was constructed from tri-*O*-acetyl-D-glucal over 6 steps. These building blocks contain not only the necessary stereochemistry and functional groups of the relevant motifs within tunicamycin, but also a phenylselenide required for the later reductive coupling event to form the tunicamine skeleton. Additionally, glycosyl donor **43** was activated as a trichloroacetimidate rather than an anomeric chloride for efficient glycosyl transfer, and glycosyl acceptor **41** contained an *N*-phthalimide to provide neighbouring group participation during the glycosylation event – thus maximising the proportion of the α,β-product. In this way, a 77% yield was obtained for the glycosylation reaction, in contrast to the 19% observed by Suami *et al.*. The resulting disaccharide **44** was next converted to allylic alcohol **45** *via* a number of protecting group manipulations, oxidation of the phenylselenide and subsequent elimination

of the selenoxide intermediate. The coupling partners for this allylic alcohol in the reductive coupling event were uridine-5'-aldehydes **46** and **47**. Both of these were obtained directly from uridine by simple protecting group manipulation and primary hydroxyl oxidation – they differ only in the 2´- and 3´-hydroxyl protecting groups. Concurrently with the preparation of these key reductive coupling substrates, fatty acid **51** was also synthesised for use during end stage *N*-acylation of the tunicamycin core skeleton; this acyl chain corresponds to the one seen in natural tunicamycin V. Commercial cyclododecene (**48**) was first cleaved by ozonolysis to yield aldehyde **49**, which was then olefinated with isopropylidenetriphenylphosphorane, hydrogenated, α,β-unsaturated and finally saponified to yield branched acid **51**.

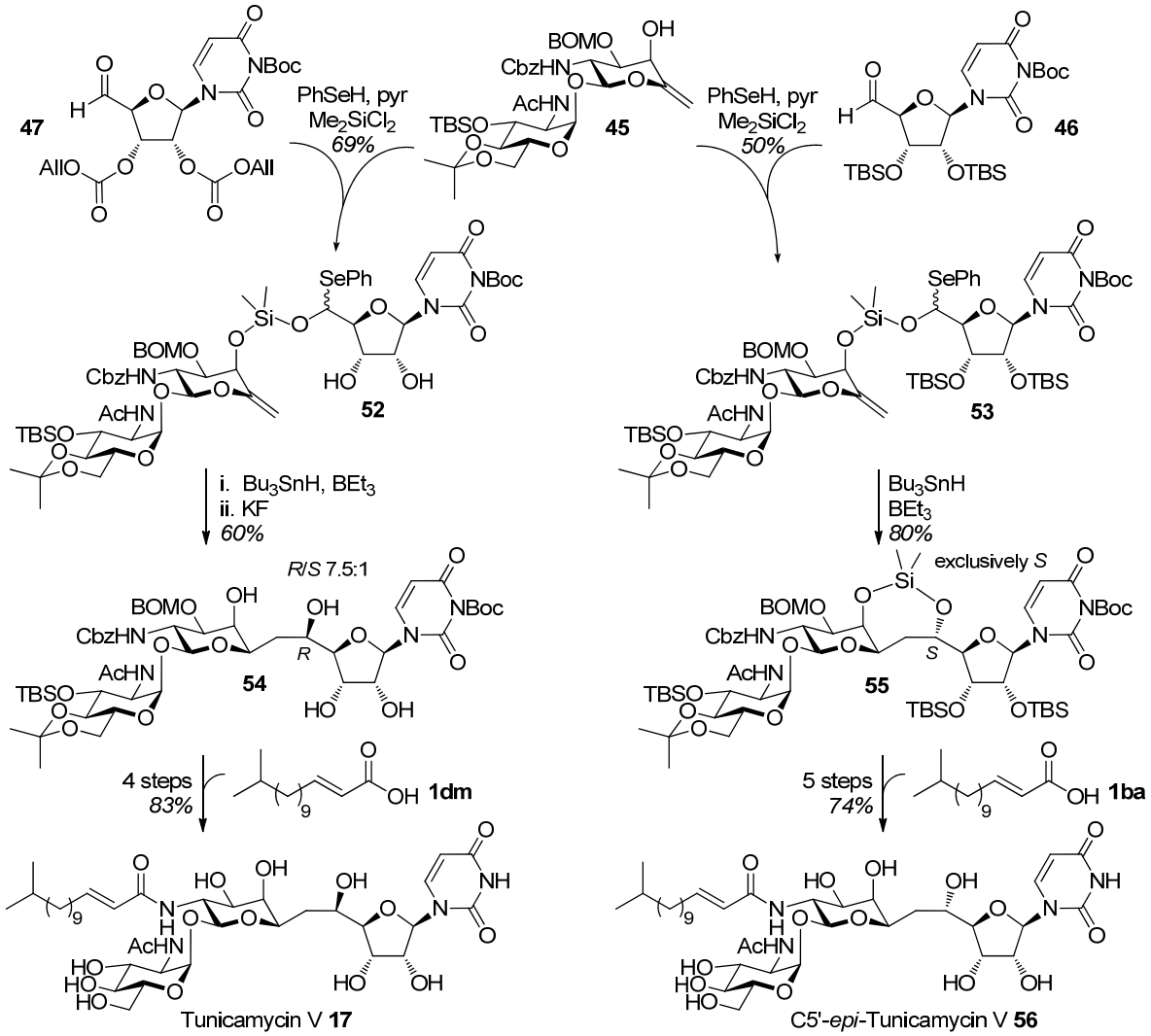


Fig. 1.17 Completion of Myers' route to tunicamycin V and its C5'-epimer

To install the undecose framework of the tunicamine moiety, uridine aldehydes **47** and **46** were first converted to hemiselenoacetals and then tethered to allylic alcohol **45** with dichlorodimethylsilane to yield adducts **52** and **53** respectively, as a mixture of C5´ epimers (Fig. 1.17). The allyloxy carbonate protecting groups in **47** were incorporated to allow their mild cleavage following tether formation without interfering with the sensitive hemiselenoacetal group. Both tethered intermediates were then subjected to free radical cyclisation catalysed by Bu_3SnH. The silicon tether facilitated the coupling step by bringing the two substrates into close proximity, although the stereochemical outcome of the reaction was found to be highly dependent on the 2´- and 3´-hydroxyl protecting groups on the ribose moiety. In fact, bulky TBS groups at these positions favoured formation of the unnatural C5´ *S* isomer, whereas the deprotected hydroxyls in **54** favoured the native *R* conformation. As such, following *N*-acylation analogous to the method of Suami *et al.*, tether removal and final sugar deprotection, the authors were able to access both tunicamycin V and its C5´ epimer **56**.

The synthetic scheme to the tunicamycins laid out by Myers *et al.* is certainly impressive, and not only offers a great improvement over that of Suami *et al.* but also exemplifies novel synthetic methodology in the Si-mediated reductive coupling reaction. The formation of the α,β-trehalose linkage is also greatly improved, applying modern carbohydrate methodology and judicious choice of participatory/non-participatory *N*-protecting groups to achieve efficient glycosidic bond formation. The synthesis offers the best solution yet to forming the challenging structural motifs present in tunicamycin. The synthesis remains somewhat lengthy with a great deal of protecting group manipulations, although the yields are good and thus losses at each step have been minimised. Some of the key steps are very delicate and show a great deal of dependence on the neighbouring steric and stereochemical environment. Production of tunicamycin analogues may require optimisation of specific conditions and protecting groups in each case, which would not be conducive with the formation of analogue libraries. Nonetheless, the synthesis presented here remains the most efficient to date, and also acts as a case study of thorough and logical natural product research.

1.3.6 Banaszek (1997): Synthesis of protected tunicamines

Between 1994 and 1997, the laboratory of Anna Banaszek in Poland published a series of papers reporting their studies towards synthesis of the tunicamycins.[124-128] They presented a novel synthesis of protected tunicamine subunits[126] and additionally demonstrated the elaboration of this scaffold to an α,β-1,1-linked disaccharide.[125,126,128] The authors constructed the undecose motif through a Wittig coupling using monosaccaride-derived aldehyde and phosphonium ylide coupling partners.

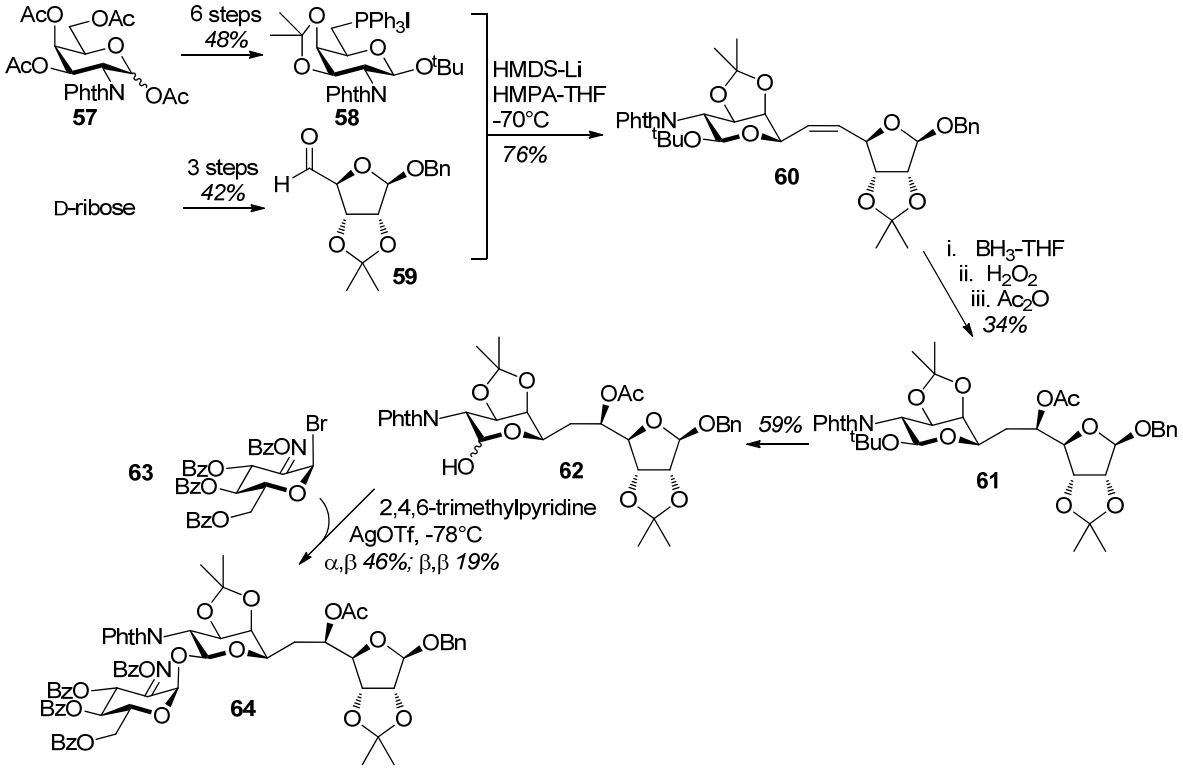


Fig. 1.18 Banaszek's synthesis of protected tunicamines and formation of an α,β-trehalose linkage

To construct the requisite coupling partners for a Wittig reaction, Banaszek *et al.* synthesised D-galactosamine-based phosphonium iodide **58** and ribosyl iodide **59** from known or commercially available monosaccharide precursors. Ylide formation and subsequent Wittig coupling was afforded by slow addition of Li-HMDS base to a mixture of both coupling partners at -70 °C; this method ensured rapid reaction of transient ylides with present aldehydes and thus minimised losses to competing β-elimination processes. Hydroboration-oxidation of the resulting alkene (**60**) with BH_3-THF was found to favour the desired 5*R* regio- and stereoisomer over the other three possibilities in a ratio of

61:10:21:8 (5*R*/5*S*/6*R*/6*S*). This fortunate regio- and stereoselectivity is postulated to result from initial coordination of boron to the furanose oxygen. Further derivatisation of this tunicamine scaffold was presented, with anomeric deprotection and glycosylation with D-glucosamine-derived bromide **63**[129] yielding the α,β-linked pseudo-trisaccharide **64** in 29% over 2 steps from *tert*-butyl glycoside **61**.

The synthesis of the tunicamine scaffold by Banaszek *et al.* represents a simple protocol utilising known methodology to afford the challenging tail-to-tail C–C linkage between galactosamine and ribose derivatives. The coupling step itself is surprisingly efficient considering the known risk of β-elimination from the generated anion. Unfortunately, the reported route suffers from poor yields in key areas, such as the stereoselective hydroboration-oxidation and the subsequent anomeric deprotection. The studies were not extended to a total synthesis of tunicamycin, nor were any analogues synthesised.

1.3.7 Sarabia (2003): Synthesis of protected tunicaminyl-uracil

Following from early studies into diazo derivatives of carbohydrates in the mid-1990s,[130,131] Sarabia and coworkers published a synthesis of differentially protected tunicaminyl-uracil derivatives in 2003, which included synthesis of analogues with a hydroxyl group at C-6´ rather than C-5´.[132] Their general approach consisted of coupling galactosamine and uridine moieties *via* diazo and aldehyde functionalities incorporated into the relevant coupling partners (Fig. 1.19).

Starting from known D-galactosamine derivative **65**, *N*-acetyl-*N*-nitroso sugars **66** and **67** were constructed through protecting group manipulation, tosylation and azide displacement at C-6, followed by azide reduction and *N*-nitrosoylation and *N*-acetylation (Fig. 1.19). These precursors acted as latent diazo compounds, which when exposed to KOH underwent transformation to reactive diazo compounds **68** and **69** respectively. Exposure to uridine aldehyde **70**[111] gave 5´-keto tunicamines **71** and **73** in good yield. It should be noted that with model substrates some formation of epoxides was observed, although with the true substrates only ketone products were isolated.[130,131] Reduction of the ketones to the requisite hydroxy groups of tunicamine proceeded in good yield, although in both cases a mixture of stereoisomers was observed. Although good stereoselectivity was achieved in the reduction, the stereochemistry of the C-5´ position was only tentatively assigned, if at all. As such, the true configuration of the final tunicaminyl-uracil derivatives **72** and **74** has not been confirmed, and a level of ambiguity remains over this crucial final

reduction step. In addition to synthesising these two tunicaminyl-uracil derivatives, the authors also switched the roles of the two coupling partners (**75** and **76**) and produced a 5´-deoxy-6´-hydroxy analogue of tunicaminyl-uracil using the same diazo coupling strategy. Once again, the products arising from reduction of the intermediate undecose ketone were not fully characterised and the stereochemistry at C-6´ in these compounds remains unassigned.

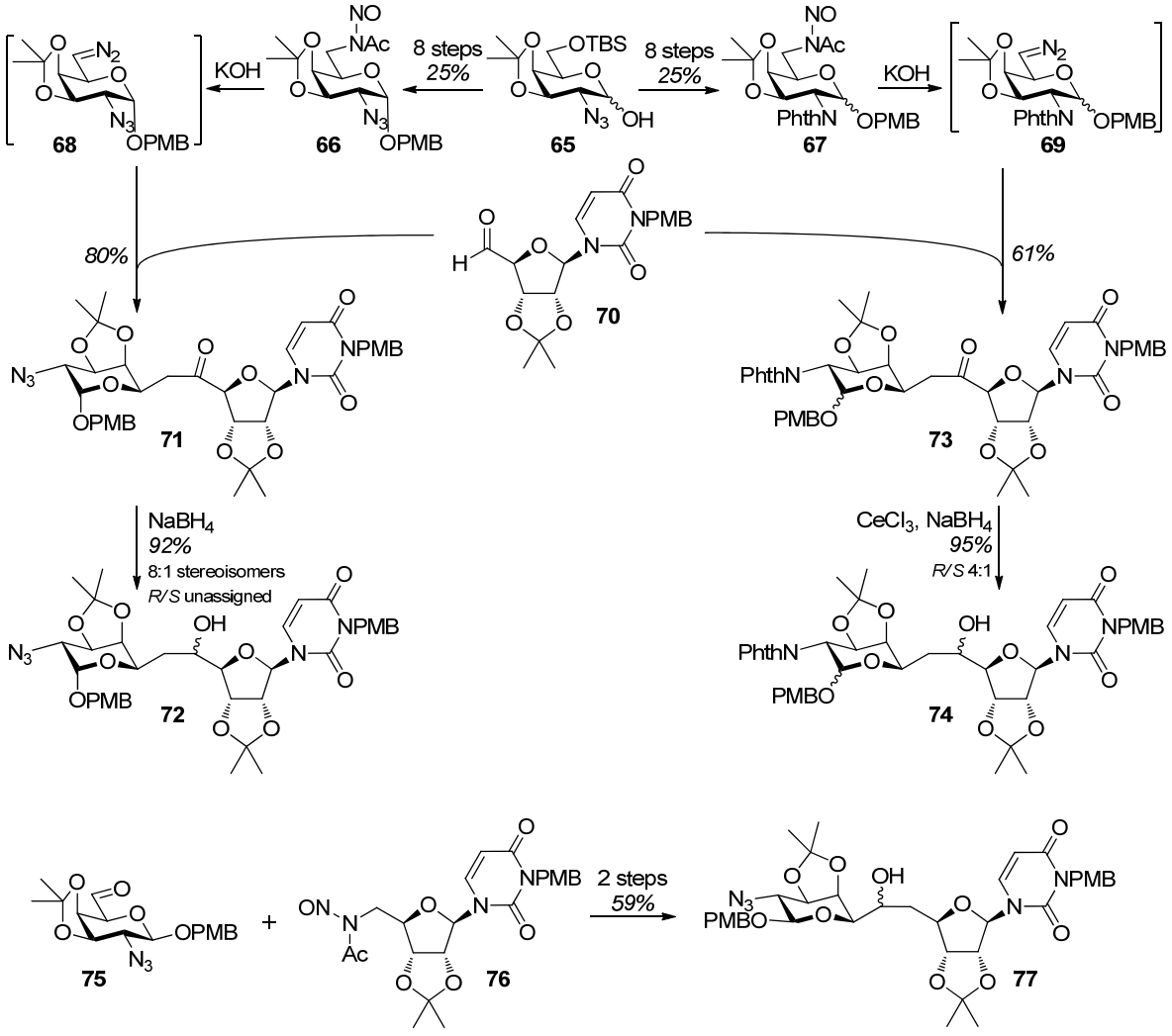


Fig. 1.19 Sarabia's synthesis of protected tunicaminyl-uracil derivatives

The work of Sarabia *et al.* provides an interesting application of the group's work developing diazo chemistry with carbohydrates. However, the studies have been left incomplete and a number of questions remain over the selectivity of the final reduction step. Additionally, preparation of the diazo precursors derived from D-galactosamine is lengthy and inefficient, somewhat diluting the appeal of this synthetic route. Nonetheless, there is scope for improvement during these early protecting group manipulations, and importantly

the coupling step itself is efficient, offering a further alternative to the various strategies already presented.

1.3.8 Matsuda (2004): Synthesis of protected tunicaminyl-uracil

The laboratory of Akira Matsuda in Japan published a single report in 2004, documenting a novel route to protected tunicaminyl-uracil derivatives.[133] The approach they used was a SmI_2-mediated aldol coupling of uridine and D-galactosamine-derived subunits followed by an intramolecular Pummerer reaction (Fig. 1.20).

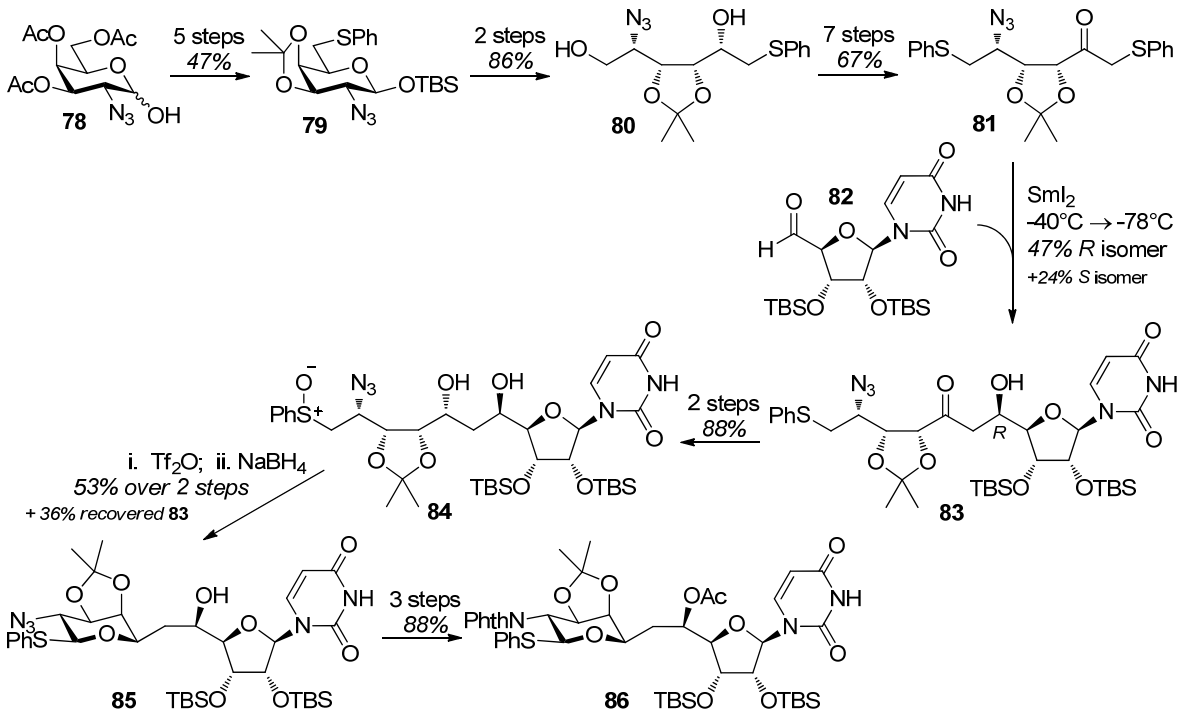


Fig. 1.20 Matsuda's synthesis of protected tunicaminyl-uracil

The key precursor to the SmI_2-mediated aldol reaction is α-phenylthio ketone **81**, which is derived from known D-galactosamine building block **78**. Protecting group manipulation and installation of a 6-thiophenyl group yielded pyranoside **79**. Anomeric deprotection followed by reductive ring-opening gave open chain diol **80**, which was further elaborated through installation of a C-5 ketone and a C-1 thiophenyl group to yield the aforementioned α-phenylthio ketone **81**. Uridine aldehyde **82** was obtained by simple protection and oxidation of uridine, and acted as the second coupling partner for the subsequent aldol reaction. Thus, exposure of aldehyde **82** and α-phenylthio ketone **81** to SmI_2 under carefully optimised conditions led to formation of undecose **83** as the major

diastereoisomer, *via* a two electron reduction of the α-phenylthio ketone leading to a samarium enolate which attacked the resident uridine aldehyde under mild conditions. Diastereoselective reduction of ketone **83** was afforded by intramolecular hydride delivery from $NaBH(OAc)_3$, followed by stereoselective sulfide oxidation to yield sulfoxide **84**. Activation of this sulfoxide with Tf_2O initiated an intramolecular Pummerer rearrangement to yield **85**, installing the pyranose moiety present within the tunicamine scaffold. Reduction of the crude product mixture allowed good recovery of unreacted starting materials whilst not affecting the product structure. Subsequent protecting group manipulation afforded a second tunicaminyl-uracil derivative **86**, containing a differentially protected C-10´ amine.

The synthetic scheme developed by Matsuda *et al.* provides mild access to the challenging undecose scaffold, and possesses a reasonable level of stereocontrol. The preparation of the D-galactosamine-based precursor is lengthy and the healthy levels of oxidation-reduction steps could be deemed inefficient, but overall the chemical logic is neat and the synthesis is a great improvement over a number of previous efforts. Whether the carefully optimised conditions for the SmI_2-mediated aldol coupling are amenable to modified coupling partners for the production of analogues remains to be determined.

1.4 Biosynthesis of the tunicamycins

1.4.1 Natural product biosynthesis

The study of natural product biosynthesis has held the attention of chemists and biologists for many years.[134] The importance of natural products is highlighted by the high proportion of pharmaceutical drugs based on or derived from natural product scaffolds.[135] The production of these biologically active compounds is predominantly controlled by specific sets of genes, which code for biosynthetic enzymes that convert common metabolic precursors into complex secondary metabolites. By understanding how these processes operate in cells, we can expand our use of the broad range of biological activities these compounds possess and access the extraordinary diversity of their chemical scaffolds in new and unprecedented ways. Heterologous expression of entire gene clusters in quick-growing bacteria has led to enhanced production of valuable therapeutic compounds originating from scarce and non-culturable organisms.[136-139] The study and manipulation of individual genes has led to the development of mutasynthesis[140,141] and combinatorial biosynthesis techniques[142-144] – involving knockout or hybrid genetic mutants producing biosynthetic precursors or natural product analogues with altered biological properties. Work with individual biosynthetic enzymes has not only enabled *in vitro* production of natural product analogues,[145] but has driven forward our fundamental understanding of chemical reaction mechanisms and protein structure, inspiring rational protein engineering and the development of completely new biotransformations.[146]

Research has historically focussed on bacteria, in particular soil actinomycetes such as *Streptomyces*, whose natural products are present in great abundance and diversity and as such have been the major source of natural product-derived therapeutics.[147,148] These organisms typically are relatively easy to culture and handle. Techniques for the genetic manipulation of these bacteria have also undergone extensive development. Following recent advances in DNA sequencing technology, access to entire bacterial genomes has become increasingly feasible and affordable, resulting in a rapid increase in the characterisation of specific biosynthetic pathways.[149,150] Unlike in plants and certain eukaryotes, these genes are typically found clustered together on the bacterial chromosome, allowing concomitant identification of all the genes necessary for a particular pathway.[149] The genetic homology of enzymes with similar functions has been exploited to rapidly identify novel biosynthetic clusters, despite these pathways producing structurally varied or even unrelated compounds. A great deal of progress has been made in understanding how

these biosynthetic machinery operate, allowing the application of pathway engineering to produce new and altered bioactive compounds. Below is a summary of the most significant and best studied classes of biosynthetic mechanisms involved in producing the observed cornucopia of natural product structures, together with references to extensive reviews for further reading.

1.4.1.1 Biosynthesis of polyketides and non-ribosomal peptides

In prior studies of biosynthetic pathways, the most attention has been focussed on two major classes of natural products: polyketides[151] and non-ribosomal peptides.[152] These compounds exhibit extraordinary levels of structural diversity yet are constructed from a very limited set of precursors. Both rely on the modular assembly of building blocks using different combinations of independently folded, conserved enzyme subunits in a template-driven process.[153,154]

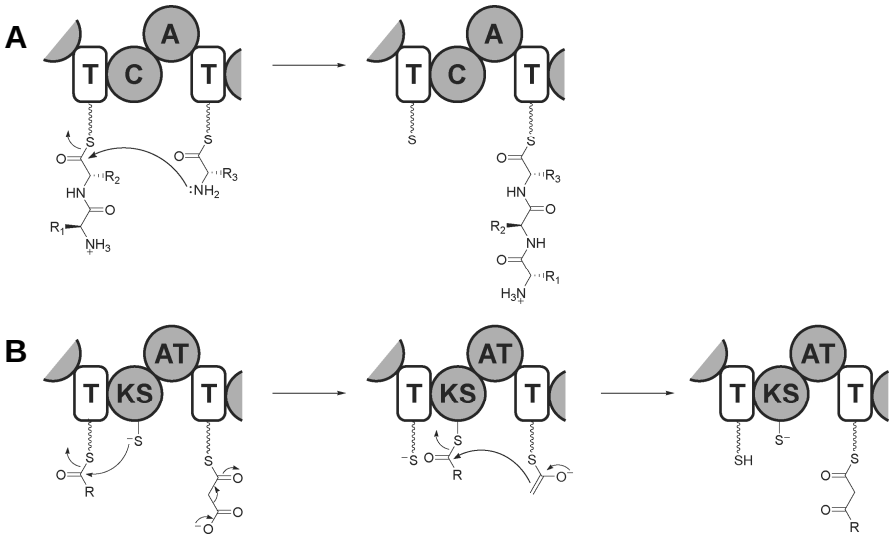


Fig. 1.21 Condensations during NRPS- and PKS-catalysed biosynthesis (reprinted with permission from Fischbach *et al.*,[153] © 2006 American Chemical Society); **A**: During non-ribosomal peptide biosynthesis, nucleophilic attack of a downstream, tethered amino acid amine on the upstream peptidyl thioester results in amide bond formation; **B**: During polyketide biosynthesis, formation of a downstream enolate anion precedes a Claisen condensation with the upstream acyl thioester to form a *C-C* bond.

The monomer building blocks used in these assembly lines are all derived from carboxylic acids, activated as thioesters by attachment to the thiolate anion of pantetheinyl-phospho carrier protein domains in the enzymatic assembly line. In polyketide synthetase (PKS) enzymes these monomers are typically restricted to extraordinarily simple acetate, malonate and methylmalonate moieties in the form of acyl-CoA thioesters. In non-

ribosomal peptide synthetases (NRPS), the monomers are predominantly proteinogenic or non-proteinogenic amino acids preactivated as aminoacyl-AMP mixed anhydrides (Fig. 1.21). Chain extension in both cases occurs through nucleophilic attack on the activated thioester by another monomer similarly tethered to a second thiolation domain. Inbetween these iterative extensions additional protein domains tailor the growing chain, installing further functional diversity prior to formation of the final product. In PKSs these include ketoreductases, dehydratases and enoylreductases, which iteratively reduce the β-carbon of the growing chain. In NRPSs additional domains include cyclases which can install thiazoline or oxazoline rings. Crucially, all substrates, intermediates and products are covalently tethered to the enzymatic assembly line in an analogous manner to solid-phase peptide synthesis, thus overcoming the inefficiency of multiple solution-phase couplings. Following final cleavage and/or cyclisation, further tailoring of the resulting scaffold often takes place including methylation, glycosylation and P450-mediated oxidation amongst many others (Fig. 1.22). Precursors formed by other biosynthetic pathways can often be coupled at this stage to create hybrid natural products.[153,154] The genes for hundreds of such assembly lines have been sequenced, with notable examples including polyketides erythromycin,[155,156] calicheamycin[157] and tetracycline,[158] the non-ribosomal peptides vancomycin[159] and the penicillin and cephalosporin families,[160,161] and PKS-NRPS-derived hybrids such as immunosuppressants FK506[162,163] and rapamycin[164,165].

1.4.1.2 Biosynthesis of isoprenoids

Another major class of natural products are the isoprenoids, otherwise known as terpenes or terpenoids. Together they comprise the largest family of compounds found in nature with over 55,000 known examples,[166] predominantly originating from plants and fungi, but sometimes from bacteria also.[167,168] Isoprenoids typically consist of carbon-rich scaffolds composed of multiple rings and many stereocentres but with relatively few heteroatoms.

It is well understood that isoprenoid biosynthesis begins with the condensation of two key five-carbon building blocks: isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP).[169,170] Multiple elongations lead to linear polyprenyl chains of varying length such as geranyl diphosphate (C_{10}) and farnesyl disphosphate (C_{15}), or non-linear polyprenol chains with branched, cyclobutyl or cyclopropyl functionalities. Preactivation of the polyprenol chains occurs through cleavage of the pyrophosphate or by protonation or epoxidation of a double bond in the linear isoprenoid chain.[166,171,172] These

activated chains are most often processed by a dedicated cyclase, providing a template that binds the flexible isoprenoid chain in such an orientation and conformation that cyclisation and C–C bond formation occurs in a highly regio- and stereoselective fashion, introducing extraordinary molecular complexity in a single step (Fig. 1.22).[172,173] Additional elaboration of the cyclised scaffold can follow through hydroxylation, methylation or attachment of products from other biosynthetic pathways such as polyketides or modified carbohydrates. Notable examples of isoprenoids whose biosynthetic pathways have been cloned and dissected include agricultural mycotoxin precursor trichodiene,[174] paclitaxel (Taxol) precursor taxadiene,[175,176] antifungal phytoalexin capsidiol[177-179] and cholesterol precursor lanosterol.[180,181]

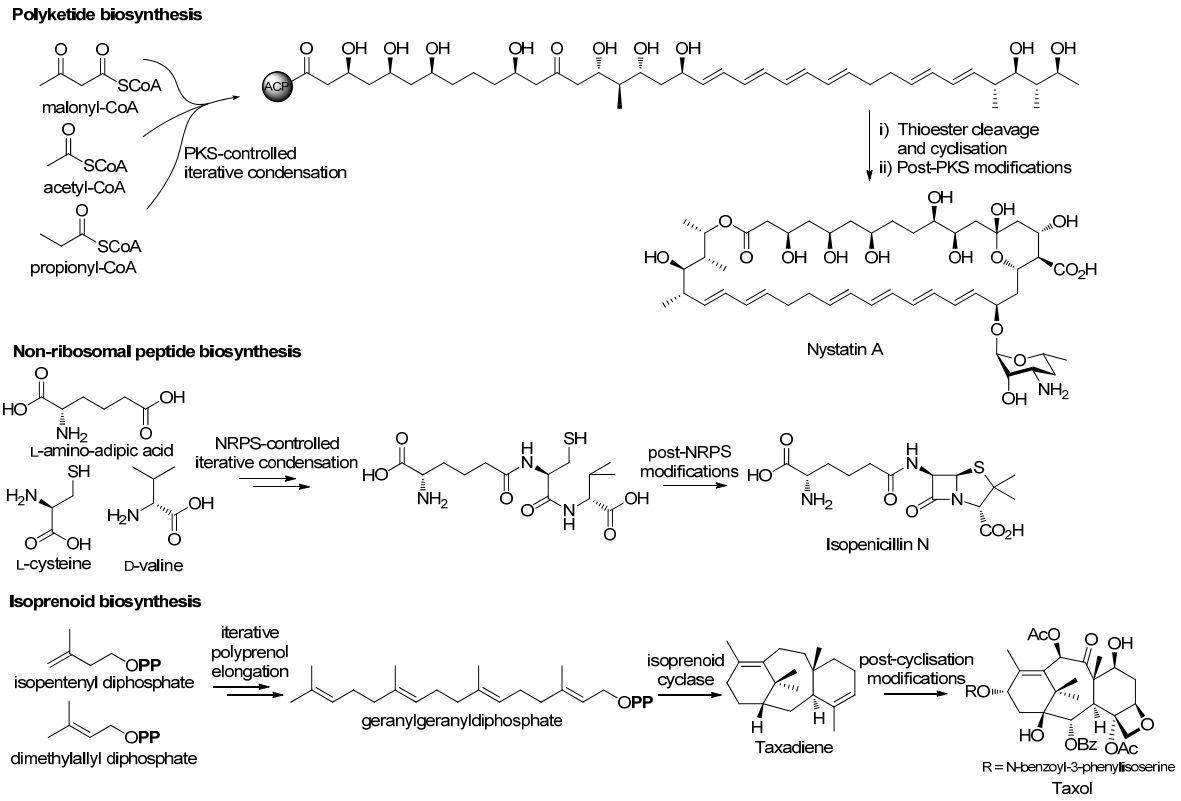


Fig. 1.22 Representative examples of polyketide, non-ribosomal peptide and isoprenoid biosynthesis

1.4.1.3 Alternative biosynthetic mechanisms to natural products

A number of other natural products classes exist, including oligosaccharide natural products, alkaloids, shikimate pathway metabolites and siderophores (Fig. 1.23). Most of these utilise non-templated assembly and are predominantly derived from primary metabolic pathways.

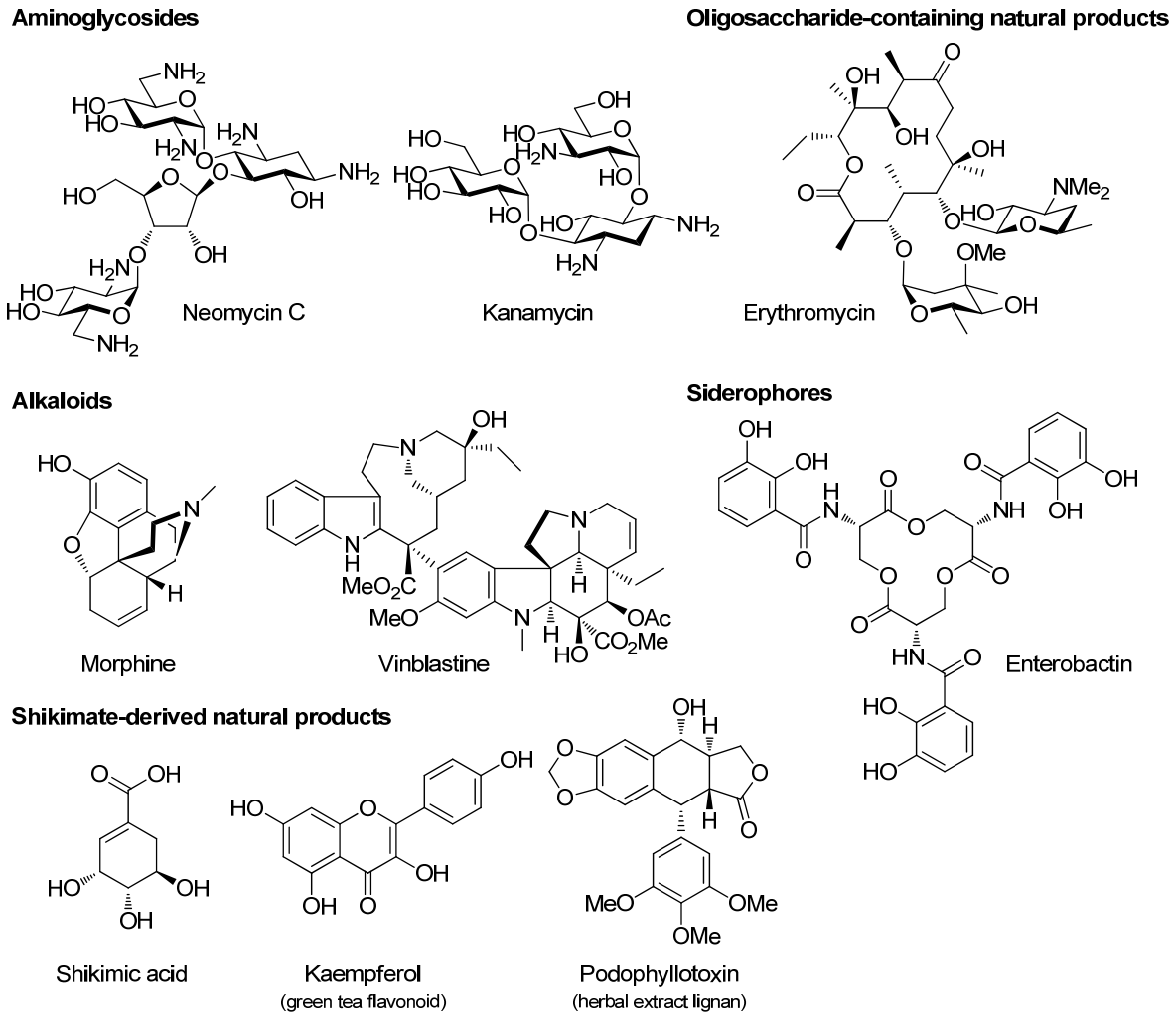


Fig. 1.23 Examples of natural products from alternative biosynthetic pathways

Aside from oligosaccharide natural products, such as aminoglycoside antibiotics kanamycin and neomycin,[182] many natural products from diverse pathways are functionalised with carbohydrates. In many cases biological activity is lost with their removal. Notable examples include enediyne calicheamycin, non-ribosomal peptide vancomycin, macrolide antibiotic erythromycin and polyketide nystatin.[183] The associated glycosyl transfer reactions predominantly involve hexose monomers anomerically activatated of as nucleotide diphosphates (NDP). The coupling of these sugar donors with

acceptor substrates is catalysed by glycosyltransferases,[184,185] driven by the loss of NDP. Despite a pool of only nine commonly occurring nucleotide sugar donors,[81] prokaryotic and plant polysaccharides and glycosylated natural products contain over a hundred different sugars. These can be epimerised, deoxygenated, aminated, methylated, acylated or sulfated versions of common NDP-sugars, amongst other diversifications.[186,187] These variations serve to alter the properties of the resulting glycoconjugates and as a result much work has gone in to understanding and manipulating these pathways.[143,182,186-188]

Alkaloids are a diverse group of nitrogen-containing secondary metabolites.[189] They are predominantly produced in plants. Notable examples include analgesics morphine and codeine, anticancer agents vinblastine and taxol, as well as well known molecules such as caffeine, nicotine and cocaine. The sources of nitrogen atoms in most cases are amino acids, in particular phenylalanine, tyrosine, tryptophan, lysine and arginine. The range of reactions converting such precursors into alkaloid products is extremely diverse and has been well studied.[190-193]

The shikimate pathway is employed by plants and microorganisms in order to synthesise aromatic amino acids phenylalanine, tyrosine and tryptophan.[191] The pathway is also the source of a great number of natural products, the majority of them derived from the aromatic amino acid primary metabolites. These include cinnamic acids, coumarins, lignans and flavonoids, as well as cyclohexane carboxylic acids which are used as polyketide starter units in important immunosuppressant antibiotics rapamycin, FK506 and FK520.[194] The central intermediate is almost always shikimic acid,[194] which is usually further diversified to yield the three aforementioned aromatic amino acids. A multitude of possible transformations to these amino acids or their biosynthetic precursors then leads to the observed plethora of natural products.[191,194-197]

Siderophores, such as catecholate siderophore enterobactin, are high-affinity iron-chelating molecules employed by various bacteria and fungi to tackle the problem of low iron bioavailability.[198] Siderophores are synthesised *via* NRPS[199] or NRPS-independent pathways, [200] depending on whether their structures comprise polypeptides or non-peptidic amide or ester bonded scaffolds respectively.

1.4.2 Biosynthesis of nucleoside antibiotics

Nucleoside antibiotics comprise a distinct class of natural products. They all share a nucleoside unit at their core, although their structures are often hybrids containing components with different biosynthetic origins. Owing to the numerous vital functions of

standard nucleosides in primary metabolism, the natural products derived from them exhibit widespread biological activity, including antibacterial, antiviral and antitumour properties.[201,202] One particularly important target is cell wall biogenesis; inhibitors of bacterial peptidoglycan and fungal chitin biosynthesis have the potential to act as antibiotics with novel modes of action, orthogonal to existing antibiotics and active against multiply-resistant bacteria.[41,203] The tunicamycins themselves fall into this category, and as with all compounds of this class its nucleoside component plays a central role in directing its biological activity.[97] The structures and biological functions of nucleoside antibiotics have been extensively reviewed, and numerous biosynthetic investigations have taken place.[41,76,201-203] These have predominantly taken the form of classic experiments with labelled precursors and genetic mutants, although the advent of genomic sequencing has led to the biosynthetic clusters of some of these compounds being cloned in the past few years. Progress in the characterisation of the genes involved and the implications for biosynthetic mechanisms for some of these compounds has been the subject of a recent review.[41] What follows is a summary of significant findings relevant to tunicamycin biosynthesis.

1.4.2.1 Liposidomycins and caprazamycins

Liposidomycins and caprazamycins are both liponucleoside antibiotics isolated from different *Streptomyces* strains in 1985 and 2003 respectively.[204,205] They share some biological activity with the tunicamycins, targeting the formation of lipid I by enzyme MraY during peptidoglycan biosynthesis.[204,205] They are structurally homologous, containing uridine, aminoribose, diazepanone and fatty acyl moieties; caprazamycin differs in the presence of a permethylated L-rhamnose and the absence of any sulfation at C-2´´ of the aminoribose (Fig. 1.24).[206,207] As with many nucleoside antibiotics inhibiting MraY, each compound is isolated as a mixture of homologues with different length acyl chains. The biosynthetic clusters of liposidomycin,[208] caprazamycin[209] and closely related A-90289[210] have very recently been cloned; their high levels of homology confirming closely related biosynthetic origins.

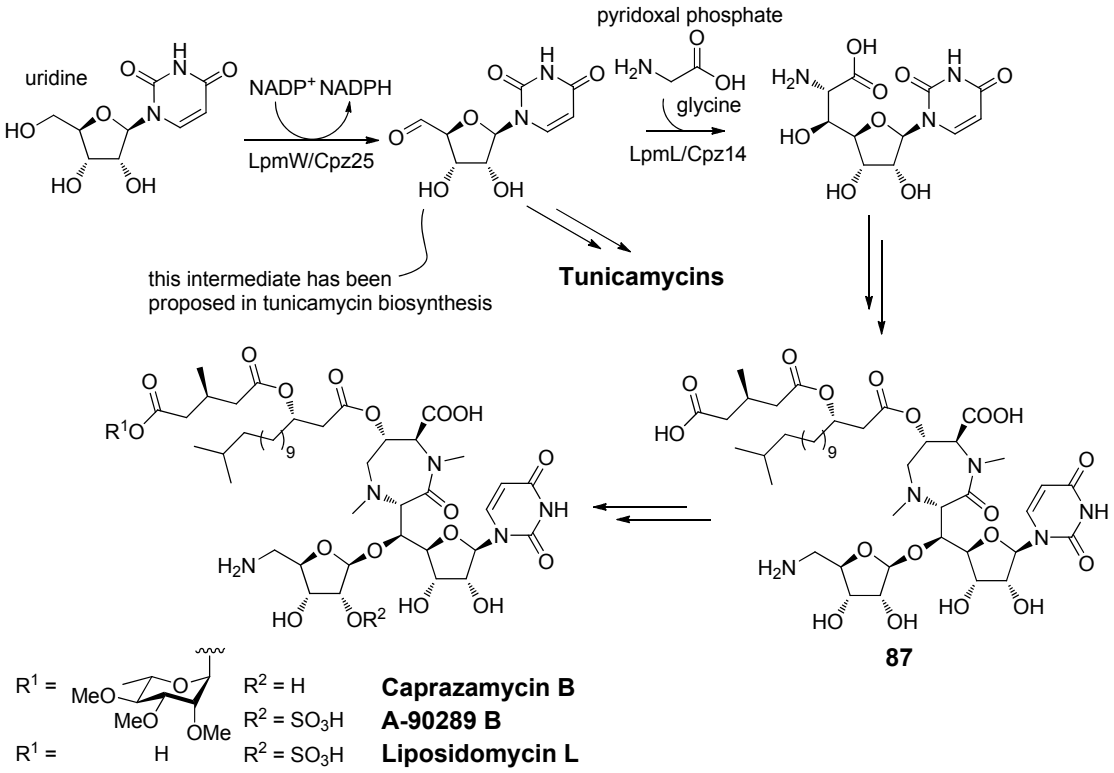


Fig. 1.24 Proposed biosynthetic pathway of the liposidomycins and caprazamycins

Some of the proposed biosynthetic pathway is outlined in Fig. 1.24.[208,209] Construction begins with oxidation of the 5´-OH in uridine by putative alcohol dehydrogenase Cpz25/LpmW – interestingly, formation of a uridine 5´-aldehyde intermediate has also been proposed in tunicamycin biosynthesis.[211] Oxidation is followed by the aldol addition of a pyridoxal phosphate-glycine adduct. The resulting uridine-glycine adduct is subsequently converted to advanced precursor **87**, to which a combination of sulfate and permethylated L-rhamnose moieties are introduced, yielding either liposidomycin, caprazamycin or A-90289 final products. The metabolic pathways detailing the sequential transformations to these final products were predicted from analysis of corresponding biosynthetic gene clusters.[208,209,212] Functional characterisation of individual genes has been limited[209,210] and much remains to be learnt about the exact role of certain enzymes and the precise ordering of the presented biosynthetic transformations.

1.4.2.2 Nikkomycins and polyoxins

The nikkomycins and polyoxins are closely related peptidyl nucleosides, first isolated from different *Streptomyces* strains in 1976 and 1965 respectively.[213,214] Interest in them lies in their specific inhibition of chitin synthase, arising from their structural similarity to UDP-GlcNAc – a substrate for this enzyme.[215,216] Both natural products have an aminohexuronic acid at their core and *N*-glycosidically attached bases: uracil and 4-formyl-4-imidazolin-2-one[41] in the nikkomycins and uracil or a number of thymine derivatives in the polyoxins (Fig. 1.25). To these nucleoside mimics are attached unusual amino acid residues such as hydroxypyridylhomothreonine (HPHT), carbamoyl polyoxamic acid (CPOAA) and polyoximic acid (POIA),[213,217] whose biogenesis from amino acid precursors has been reviewed.[218]

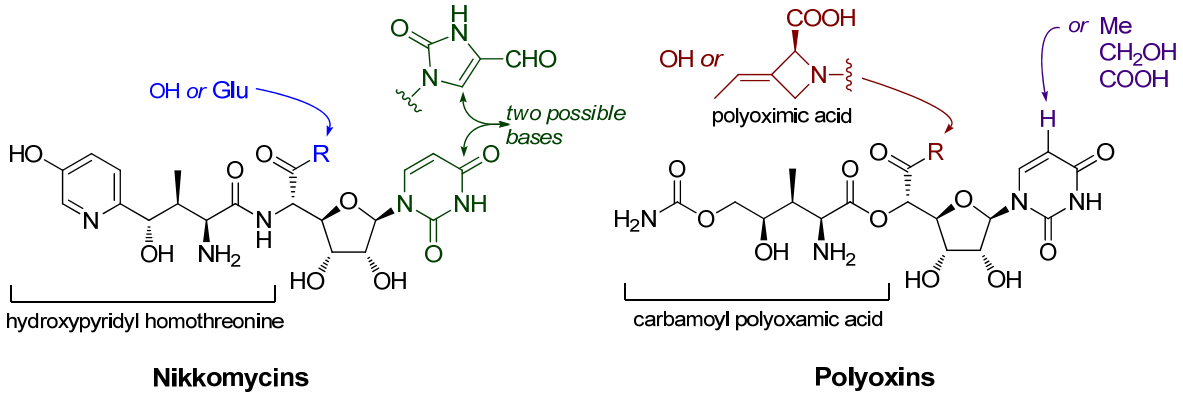


Fig. 1.25 General structures of nikkomycins and polyoxins

The biosynthetic gene clusters of both families of compounds[218-220] together with biochemical investigations into the pathways involved have recently been reviewed.[41] The construction of the aminohexuronic acid begins with the attachment of phosphoenolpyruvate to the 3´-OH of UMP (Fig. 1.26). *In vitro* experiments showed NikO and purified PolA homologue catalysed this reaction – interestingly uridine was not accepted as a substrate. The conversion to the octofuranuloseuronic acid intermediate is not well understood but is thought to involve an intramolecular aldol reaction of the 3´-enolpyruvyl moiety on a transient uridine 5´-aldehyde, together with the action of phosphatase PolJ/NikL. Subsequent cyclisation to the next bicyclic intermediate is predicted to be catalysed by radical SAM enzyme PolH/NikJ. Final conversion to the aminohexuronic acid is even more perplexing, but is postulated to involve enzymes PolK, PolD and PolI, analogous to NikM, NikI and NikK respectively. Clearly further characterisation is required to decipher the biosynthesis of nikkomycins and polyoxins, but

the results may offer clues about some aspects of tunicamycin biosynthesis. Crucially, all of these classes of compound require the formation of a C–C bond at C-5′ of a uracil-derivative.

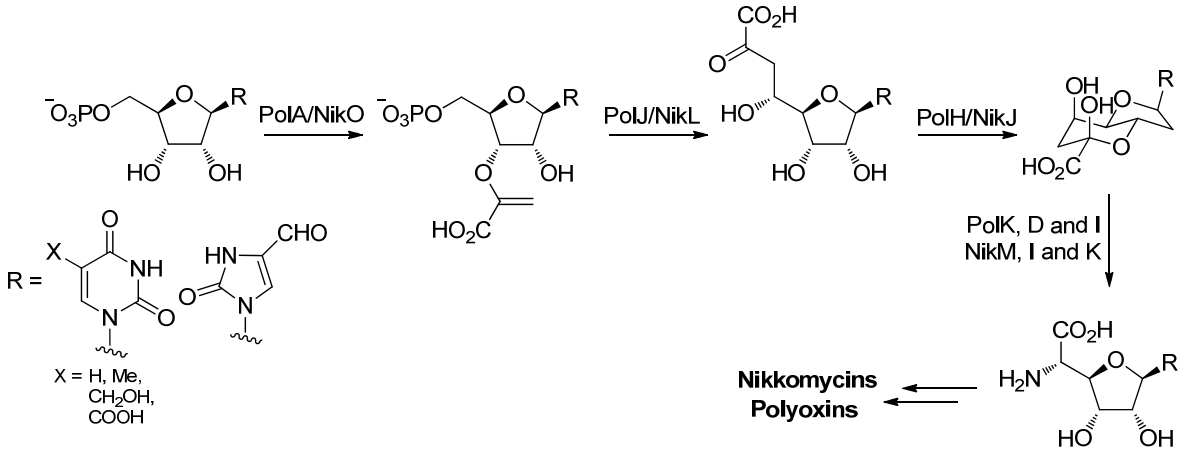


Fig. 1.26 Overview of polyoxin and nikkomycin biosynthesis

1.4.2.3 Puromycin

Puromycin is an aminonucleoside antibiotic produced by *Streptomyces alboniger*.[221] It inhibits protein synthesis by interfering with peptidyl transfer on both 70S and 80S ribosomes and has widely been used as a basic tool for studying protein synthesis. Its structure has a tyrosine residue attached to adenosine through an amide bond, together with further minor modifications. The biosynthesis of puromycin has been well studied and the biosynthetic genes of puromycin and closely related antibiotic A201A have been cloned.[221,222] A total of seven individual enzymes have been functionally characterised *in vitro*,[223-229] allowing an accurate biosynthetic pathway to be constructed (early stages summarised in Fig. 1.27).[230] Self-resistance is achieved through export of an *N*-acetylated precursor, which is converted to active puromycin by the action of *N*-deacetylase NapH outside the cell.[229] The early stages of puromycin biosynthesis involve processing of nucleotide triphosphates and monophosphates, and may be of relevance when considering the biosynthesis of the uridyl portion of tunicamycin.

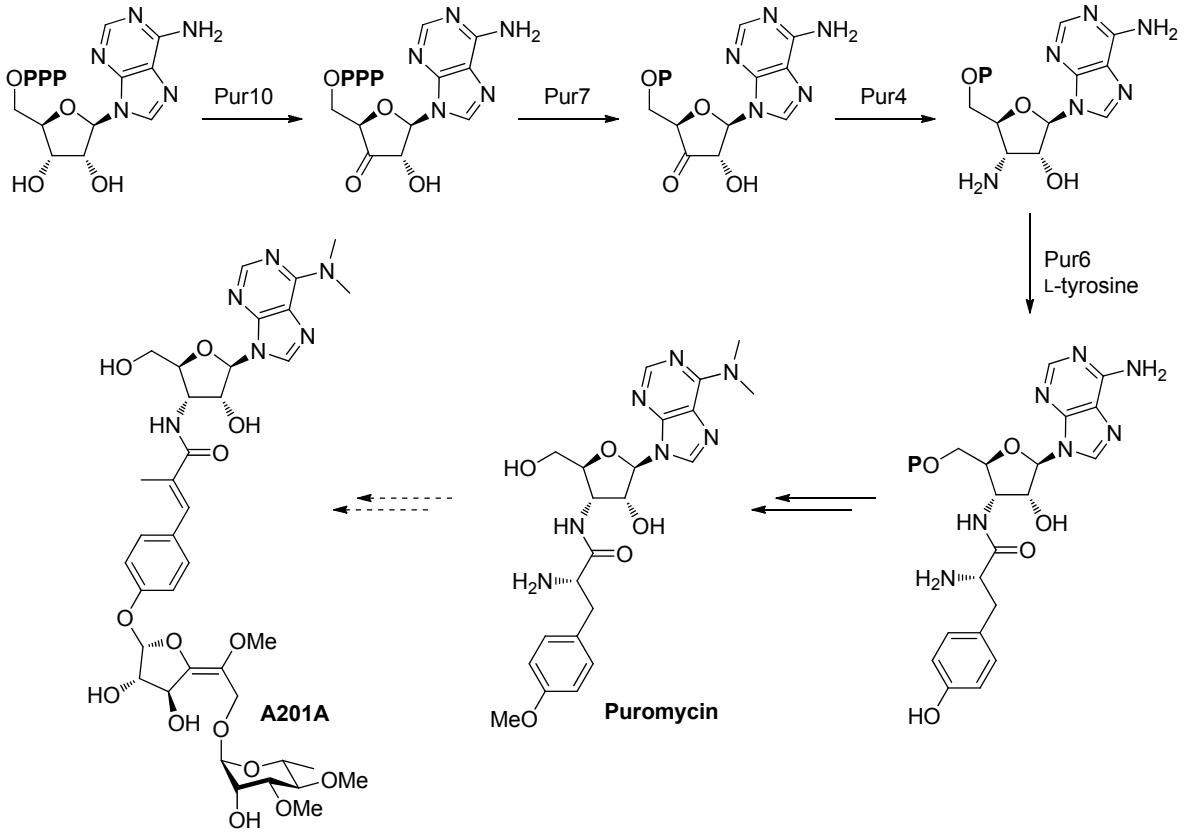


Fig. 1.27 Overview of puromycin biosynthesis

1.4.2.4 Capuramycin and A-500359

Capuramycin[231] and A-500359[232] (Fig. 1.28) were both first isolated from *S. griseus* and exert a strong inhibitory effect on MraY from peptidoglycan biosynthesis, much like the related uracil nucleoside antibiotic families of liposidomycins, mureidomycins, tunicamycins and muraymycins. Isotope feeding experiments confirmed uridine as a metabolic precursor, leading to the proposed involvement of a uridine 5´-aldehyde intermediate. Interestingly, this intermediate has also been postulated in a number of other nucleoside antibiotic pathways, including that of tunicamycin.[211] The biosynthetic gene cluster of A-500539 has recently been sequenced, although no further functional characterisation of the pathway has taken place and key transformations in the proposed biosynthetic pathway remain unexplained.[233]

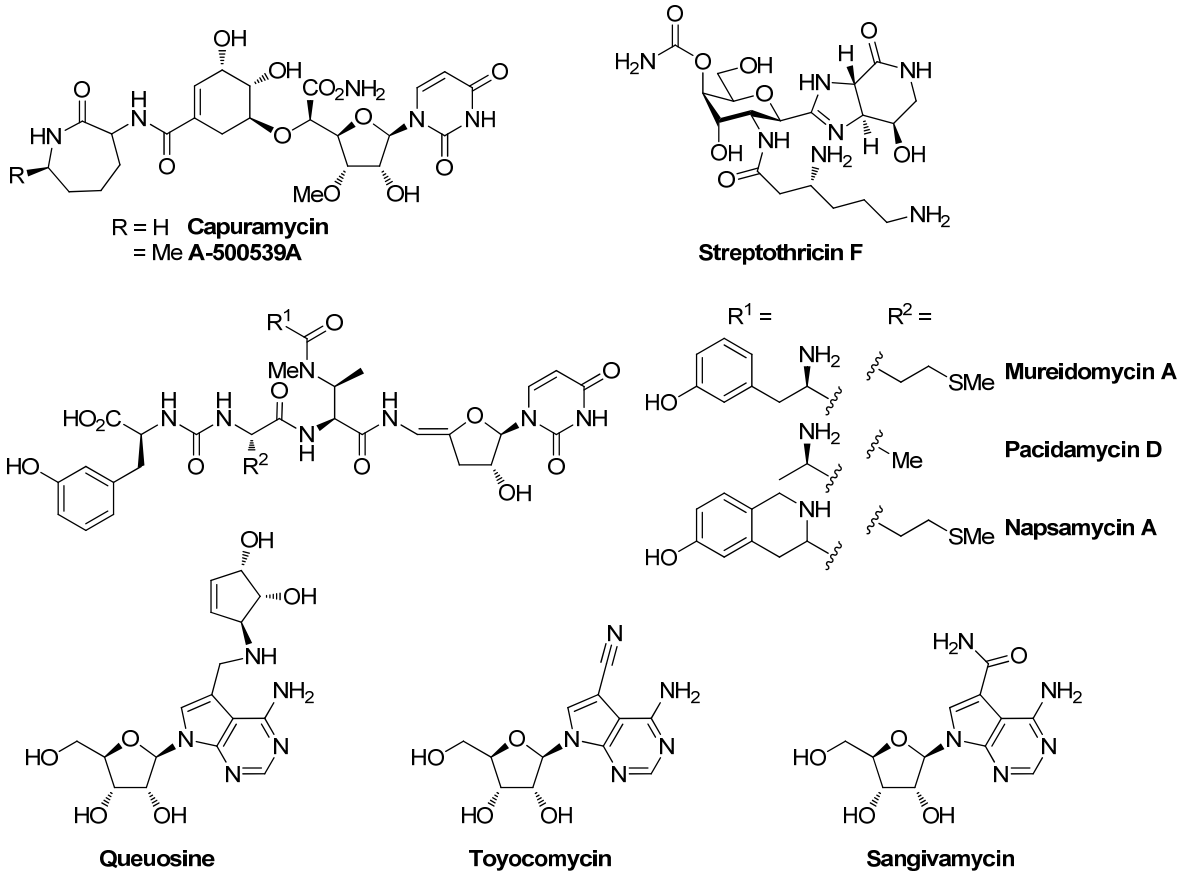


Fig. 1.28 Structures of other nucleoside antibiotics whose biosynthesis is described

1.4.2.5 Other nucleoside antibiotics

The biosynthesis of a number of other nucleoside antibiotics has been investigated. These include the uridyl-peptide families of mureidomycins[234] and pacidomycins[41,235] – both of which target MraY from peptidoglycan biosynthesis (Fig. 1.28). The metabolic pathways of streptothricin,[236,237] toyocamycin,[238] sangivamycin,[238] queuosine[239] and fungicides blasticidin S[240-245] and mildiomycin[246] have also been investigated. The structural differences between these compounds and tunicamycin suggest their biosynthetic pathways are unlikely to share any significant similarities with that of our compound of interest.

Following the submission of this thesis, the gene cluster for pacidamycin biosynthesis in *Streptomyces coeruleorubidus* was published.[247] Intriguingly, the gene cluster contained a putative oxidoreductase which was predicted to oxidise uridine to uridine 5′-aldehyde, an intermediate also predicted in tunicamycin biosynthesis.

1.4.3 Previous studies with the tunicamycins

In this section we summarise the existing work done towards elucidating tunicamycin biosynthesis. To date, the biosynthesis of the tunicamycins has not been studied in detail and no genes have been isolated or characterised. Some proposals have been made based on feeding experiments, using incorporation patterns of labelled precursors.[211,248] No genomes of any tunicamycin family producers had been studied until very recently, when the sequence of MM-19290 producer *S. clavuligerus* was deposited in GenBank;[249,250] no reports have yet appeared detailing the assignment of MM-19290 biosynthetic genes. Some tunicamycin resistance genes have also been characterised from a number of organisms (see section 1.4.3.2).[251,252]

1.4.3.1 Feeding experiments with S. chartreusis *using radiolabelled precursors*

In 2002, Tsvetanova *et al.* performed isotopic enrichment of tunicamycins through metabolic incorporation experiments with radiolabelled precursors.[211] The loci of isotopic enrichment were determined by TLC, MS and NMR characterisation. They had previously studied the fragmentation patterns of the tunicamycins by MS,[13] and applied this electrospray ionisation-collision induced dissociation-mass spectrometry (ESI-CDI-MS) technique in their subsequent experiments.

Feeding tunicamycin-producer *S. chartreusis* with [2-^{14}C]uridine led to successful incorporation into tunicamycin, suggesting uridine is a biosynthetic precursor. Upon feeding with D-[1-^{14}C]glucosamine, radiolabelled tunicamycin was once again isolated. Acid hydrolysis of this compound indicated [1-^{14}C]GlcN had been incorporated into both the 11-carbon tunicaminyl moiety and the terminal GlcNAc residue, suggesting the involvement of a 6-carbon hexose precursor in tunicamine biosynthesis and terminal glycosylation. Next, *S. chartreusis* was fed with D-[1-^{13}C]glucose. Quantitative isotopic enrichment with ^{13}C into tunicamycin was deduced by HSQC ^{13}C-^{1}H correlation NMR spectroscopy. The largest enrichments were of an equal extent at C-11´ and C-1´´, signifying direct incorporation of hexose units in each case from the same metabolic pool of GlcNAc or UDP-GlcNAc. Moderate enrichment was observed at C-6´´, C-6´, C-5´ and C-1´, derived from primary metabolism of D-[1-^{13}C]glucose to D-[6-^{13}C]glucose then UDP-[6-^{13}C]GlcNAc *via* glycolysis and gluconeogenesis[253-255] and to [1´,5´-^{13}C]uridine *via* the pentose phosphate pathway.[256-258] Further enrichment was observed in the acyl lipid chain at C-2´´ and C-4´´, alongside the *N*-acetyl methyl carbon. These incorporations are again

derived from metabolism of [1-^{13}C]glucose first to [3-^{13}C]phosphoenolpyruvate (PEP) by glycolysis, then by pyruvate dehydrogenase to [2-^{13}C]acetyl-CoA. These results support the hypothesis that tunicamycin is derived from uridine and GlcNAc building blocks and that its lipid is formed by a fatty acid synthase utilising acetyl-CoA. Construction of tunicamine could occur through sequential addition of smaller carbon units such as PEP to uridine-5´-aldehyde, as observed in the biosynthesis of polyoxin,[218] nikkomycin[41,219] and 7-, 8- and 9-carbon long-chain sugars.[259] However, if this were the case a different incorporation pattern would be observed resulting from the formation of [3-^{13}C]PEP by glycolysis.

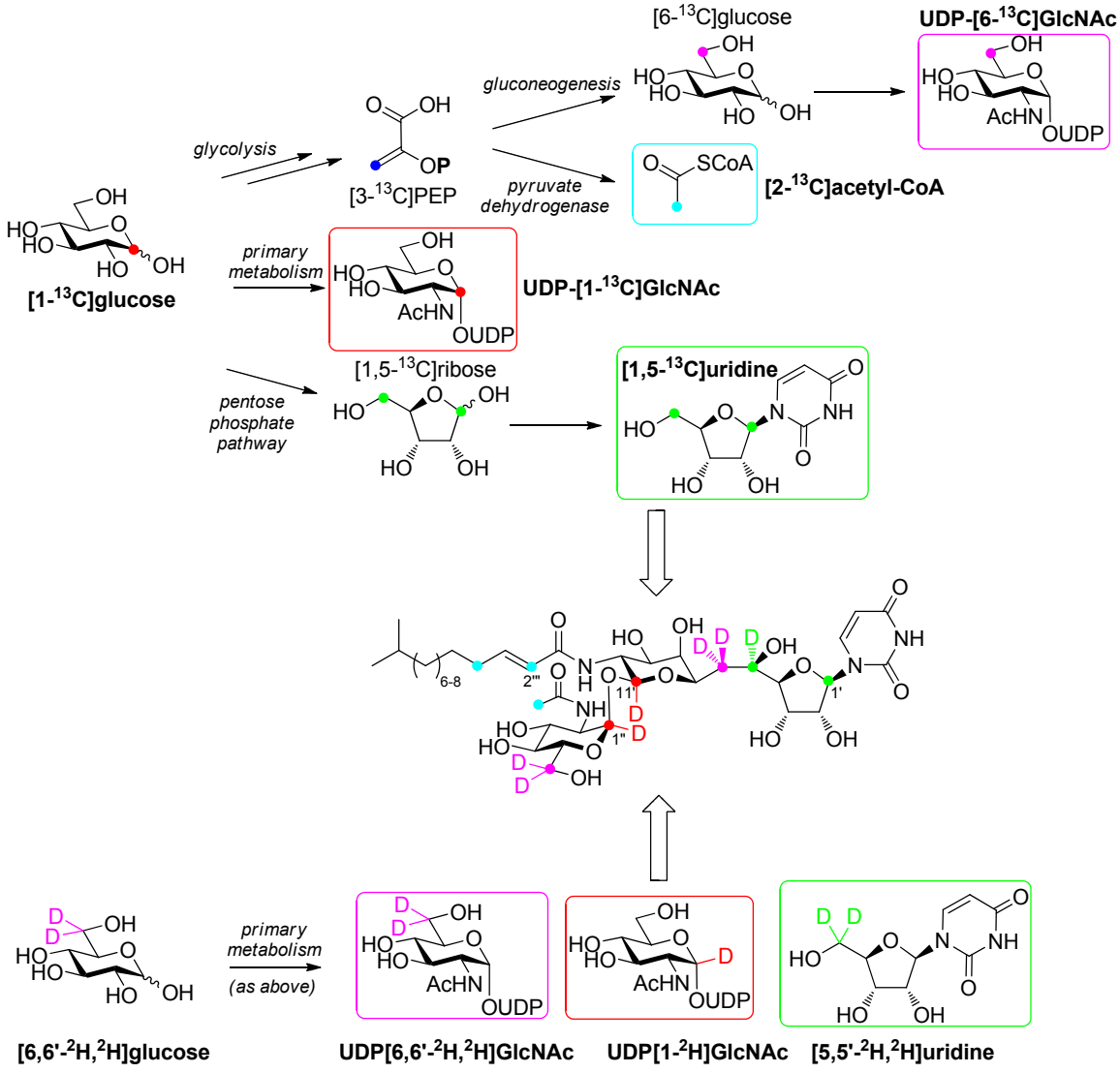


Fig. 1.29 Isotope enrichment in tunicamycin through feeding of labelled precursors

S. chartreusis was also fed with D-[6,6´-^{2}H,^{2}H]glucose and deuterium incorporation monitored by analysing ESI-CDI-MS fragments. A total of seven ^{2}H atoms were

incorporated into tunicamycin (Fig. 1.29). They were determined to be attached to C-5´, C-6´, C-11´, C-1´´ and C-6´´, showing a similar pattern of labelling to that with D-[1-^{13}C]glucose; presumably it followed the same primary metabolic shunt pathways to alternative biosynthetic precursors as those discussed previously. Assignments were supported by competitive metabolic experiments, where *S. chartreusis* cells were co-incubated with D-[6,6´-^{2}H, ^{2}H]glucose and one of a selection of unlabelled metabolites: GlcNAc, uridine, ribose, glycerol or succinate. Competition of these supplements with D-[6,6´-^{2}H, ^{2}H]glucose led to depletion of isotopic enrichment at certain positions in tunicamycin, which allowed the metabolic origin of enriched loci to be ascribed to the previously proposed shunt pathways.

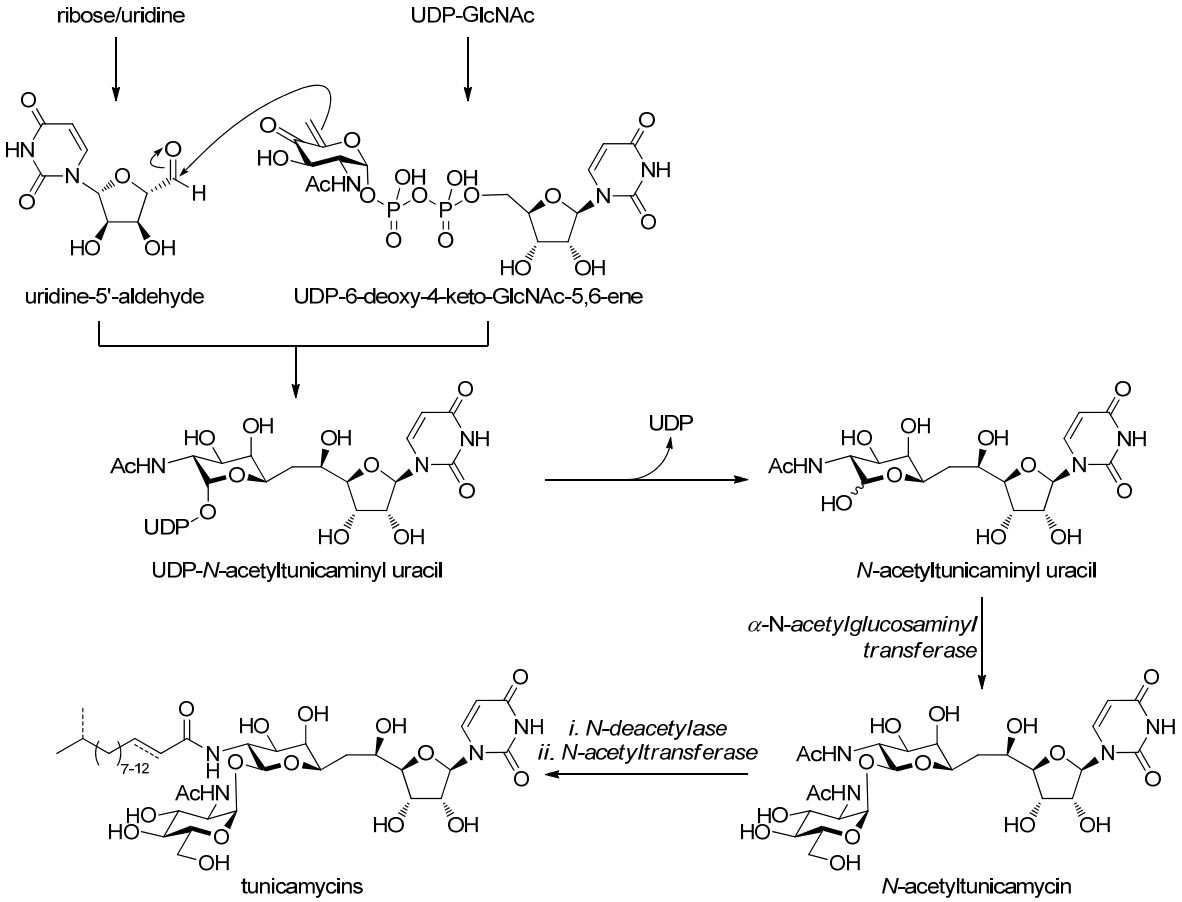


Fig. 1.30 Proposed biosynthetic pathway to the tunicamycins

Together, these experiments allowed a biosynthetic pathway for the tunicamycins to be rationalised (Fig. 1.30).[211,248] Uridine and UDP-GlcNAc are first converted to key intermediates uridine-5´-aldehyde and UDP-6-deoxy-4-keto-GlcNAc-5,6-ene. The former intermediate is proposed to feature in the biosynthesis of related nucleoside antibiotics lipisidomycins[208] and caprazamycins,[41] as well as the nikkomycins[219] and polyoxins.[218] The latter intermediate is homologous to those seen in 6-deoxysugar metabolism,[260] including FlaA1 from *Helicobacter pylori* which converts UDP-GlcNAc to UDP-6-deoxy-4-keto-GlcNAc[261] and RffG from *E. coli* which converts dTDP-D-glucose to dTDP-6-deoxy-4-keto-D-glucose.[262,263] The mechanism for coupling these two intermediates remains ambiguous and is only hypothesised, although in the biosynthesis of the nikkomycins and polyoxins, parallels can be drawn to the addition of an analogous enol-ether moiety to uridine-5´-aldehyde.[41] The ketone in the resulting adduct is then selectively reduced to give UDP-*N*-acetyltunicaminyl uracil, a reaction precedented in the same 6-deoxysugar biosynthetic pathways.[260] Following construction of the tunicamine core, hydrolysis of UDP yields a free anomeric hydroxyl to which GlcNAc is stereospecifically transferred from UDP-GlcNAc to form the requisite α,β-1´´,11´ glycosidic linkage. This completed tunicamycin skeleton is then selectively *N*-deacetylated and *N*-acylated with the range of unsaturated fatty acids observed in the final natural products. An alternative ordering exists whereby the α,β-trehalose linkage is formed first, followed by activation and coupling to uridine-5´-aldehyde. However, this would require two different sources of GlcNAc: UDP-GlcNAc and free *N*-acetylglucosamine. The equivalent enrichment of both C-1´´ and C-11´ strongly suggests the incorporated GlcNAc molecules come from the same metabolic pool, supporting the pathway in Fig. 1.30 which uses only UDP-GlcNAc.

In conclusion, these studies have helped establish the metabolic origin of different portions of the tunicamycin structure, allowing a rationalised biosynthetic pathway to be drawn. However, the proposals made remain largely hypothetical, and no real insights into some of the unique biosynthetic mechanisms have been attained. As such, any information on the genetic make up of the tunicamycin biosynthetic pathway or isolation and characterisation of any of the constituent enzymes would be of great value.

1.4.3.2 Tunicamycin resistance proteins

Tunicamycin-resistant strains of *Bacillus subtilis* were first isolated in 1978[264] and the gene conferring resistance, *tmrB*, has since been sequenced and studied.[251] Cloning this gene into *E. coli* conferred resistance to the antibiotic[265] and the TmrB protein has been shown to bind tunicamycin directly; a C-terminal α-helix was shown to be involved in this binding activity.[266] The crystal structure of a derivative protein TmrD from radiation-resistant bacterium *Deinococcus radiodurans* has recently been published.[252] The exact mechanism of inhibition by this protein is unknown although studies have shown the *B. subtilis* protein to be an ATP-binding membrane protein.[267] Indeed, it contains a GxxxxGKT/S ATP-binding signature[268] and is predicted to be an ATPase by sequence homology. The crystal structure of TmrD provides important information and shows strong similarity to nucleoside monophosphate kinases,[252] suggesting it inactivates tunicamycin by phosphorylation on one of the pseudoribose hydroxyls in an analogous manner to chloramphenicol phosphotransferase.[269,270]

In summary, it should be noted that the mechanisms used by these TmrB-like proteins may or may not be related to self-resistance mechanisms of the tunicamycin producing strains and so it is as yet unclear whether these studies will assist in locating genes for tunicamycin biosynthesis.

1.5 Tunicamycin analogues as glycosyltransferase inhibitors

1.5.1 Hypothesis

The tunicamycins function as non-hydrolysable analogues of UDP-GlcNAc, and together with their fatty acyl chains likely act as tightly-binding bi-substrate transition state mimics, allowing them to exhibit extraordinary properties as highly specific and potent glycosyltransferase inhibitors. We hypothesise that through modifying terminal portions of the tunicamycin structure, the inhibition target of the resulting analogues could be switched from UDP-HexNAc:Dol-P HexNAc-1-P transferases to a diverse range of physiologically important glycosyltransferase enzymes (Fig. 1.31). A library of rationally designed inhibitors could be created, targeting glycosyltransferase enzymes whose functions could be completely unrelated to tunicamycin's native target, but which share a dependence on nucleotide sugar substrates.

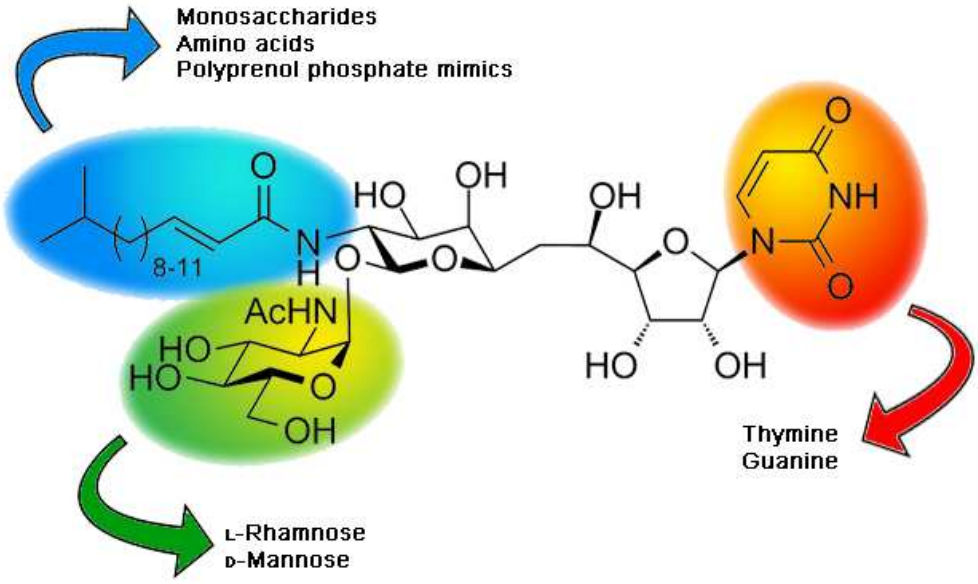


Fig. 1.31 Modification of tunicamycin structure to target glycosyltransferase inhibition

The central pseudo-GalNAc moiety of tunicamycin seemingly acts as an efficient mimic of a nucleotide-sugar pyrosphosphate and is a key motif to retain in any potential analogues. The peripheral portions of the molecule can be tailored so the overall structure mimics other NDP-sugars together with the desired acceptor substrate. For example, the uracil moiety can be replaced with thymine or guanine and the terminal D-GlcNAc can be replaced with L-rhamnose or D-mannose to yield TDP-L-rhamnose and GDP-D-mannose substrate analogues. Additionally, the amide-linked fatty acid chain can be replaced with

aglycone mimics from the target glycosyltransferase, ranging from particular monosaccharides to polyprenol or amino acid acceptors (Fig. 1.31). A variety of potentially therapeutically useful targets are listed below.

1.5.1.1 Peptidoglycan biosynthesis: MraY

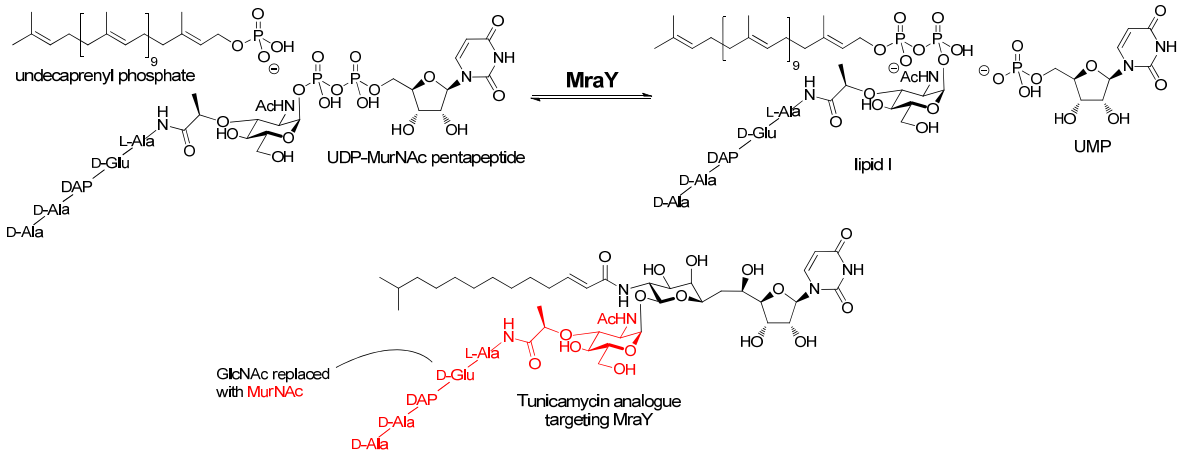


Fig. 1.32 Tunicamycin analogues as potential MraY inhibitors in peptidoglycan biosynthesis

As discussed in section 1.2.2.4, tunicamycin is already a potent inhibitor of MraY. However, it has eluded clinical use thanks to its concurrent inhibition of eukaryotic GPT (see section 1.2.3.4). By tailoring the terminal GlcNAc moiety of tunicamycin to more closely resemble the native MurNAc substrate of MraY, a greater selectivity for MraY over GPT might be achieved (Fig. 1.32). This has the potential to provide a novel therapeutic antibiotic with a mode of action orthogonal to those of existing drugs and thus helping to combat emerging antibiotic-resistance in bacteria. A library of tunicamycin variants could also be created possessing a wide variety of fatty acyl chains, in the hope of improving MraY/GPT selectivity. Alternatively, the uracil base could be modified to thymine in various oxidation states or 4-formyl-4-imidazolin-2-one – bases present in the nikkomycins and polyoxins – in the hope of observing improved levels of inhibition or biological selectivity.

1.5.1.2 Protein O-GlcNAc transferase: OGT

O-Linked β-*N*-acetylglucosamine (*O*-GlcNAc) is a dynamic post-translational modification of nuclear and cytosolic protein serine or threonine residues, analogous to phosphorylation.[271,272] Proteins modified by *O*-GlcNAc include transcription factors, signaling components and metabolic enzymes, and the modification has diverse functions

in regulating signaling cascades that mediate glucose homeostasis and stress responses through effects on transcription, translation, organelle targeting and protein-protein interactions. Elevated *O*-GlcNAc levels have been implicated in diabetes, and as such inhibitors of *O*-GlcNAcylation are in great demand for the study and possible control of this disease but also as probes of other biological functions of this abundant protein modification.[273] The addition and removal of *O*-GlcNAc is catalysed by two enzymes: UDP-GlcNAc:polypeptide *O*-β-*N*-acetylglucosamine transferase (OGT) and *O*-β-*N*-acetylglucosaminidase (*O*-GlcNAcase). Potent inhibitors of *O*-GlcNAcase have been developed,[274] but although OGT has been successfully overexpressed and purified and a high-throughput assay developed,[275] inhibitors with high potency and specificity remain to be found. Tunicamycin analogues could be designed to potentially target OGT by mimicking the transition state of the enzyme involving UDP-GlcNAc and a peptide-based serine/threonine acceptor (Fig. 1.33). Tunicamycin is already a UDP-GlcNAc mimic, but the fatty acyl chain could be replaced with a serine or threonine homologue to mimic OGT's native substrate.

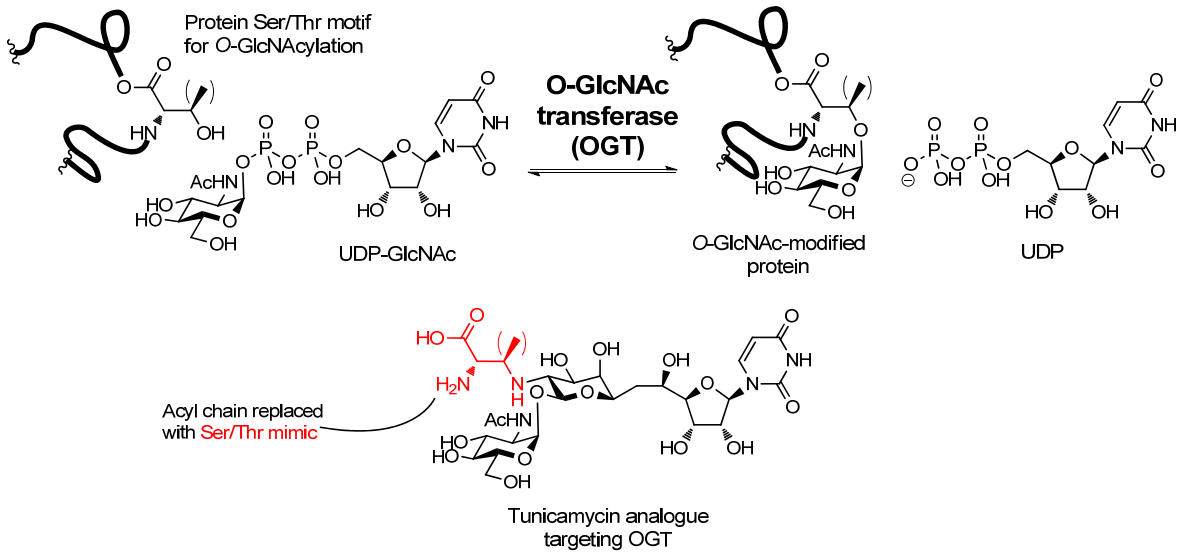


Fig. 1.33 Tunicamycin analogues as potential OGT inhibitors in protein *O*-GlcNAcylation

1.5.1.3 Mycobacterial cell wall biosynthesis: WbbI

Tuberculosis, caused by *Mycobacterium tuberculosis* is currently the leading cause of death from a single infectious agent, and increasing resistance to existing drugs is being observed in affected regions.[276,277] The mycobacterial cell envelope has a unique structure consisting of a cross-linked network of peptidoglycan where muramic acid (MurNAc)

residues are replaced with a complex polysaccharide arabinogalactan, which is in turn acylated with unique mycolic acids.[278] Targeting the biosynthesis of this cell envelope is an active area of research in the search for new drugs active against *M. tuberculosis*.[277] WbbL catalyses the transfer of an L-rhamnose residue from dTDP-L-rhamnose to decaprenol-PP-GlcNAc during the early stages of forming the arabinogalactan component of the cell wall. Deletion of the *wbbL* gene has been shown to be lethal to mycobacterial viability[279] and since L-rhamnose is absent from eukaryotic systems, WbbL represents an attractive target for antimycobacterial therapeutic drugs.[280] Tunicamycin analogues could be designed to potentially target WbbL if the terminal fatty acyl and GlcNAc moieties were replaced with D-GlcNAc and L-rhamnose residues respectively, and if the uracil moiety was converted to a thymine group. The resulting compound would mimic the dTDP-L-rhamnose substrate as well as the reactive end of the decaprenyl-PP-GlcNAc acceptor (Fig. 1.34). To further mimic the reaction transition state, the decaprenyl-PP-GlcNAc acceptor could be mimicked more closely by inclusion of a fatty acyl chain on the introduced GlcNAc moiety.

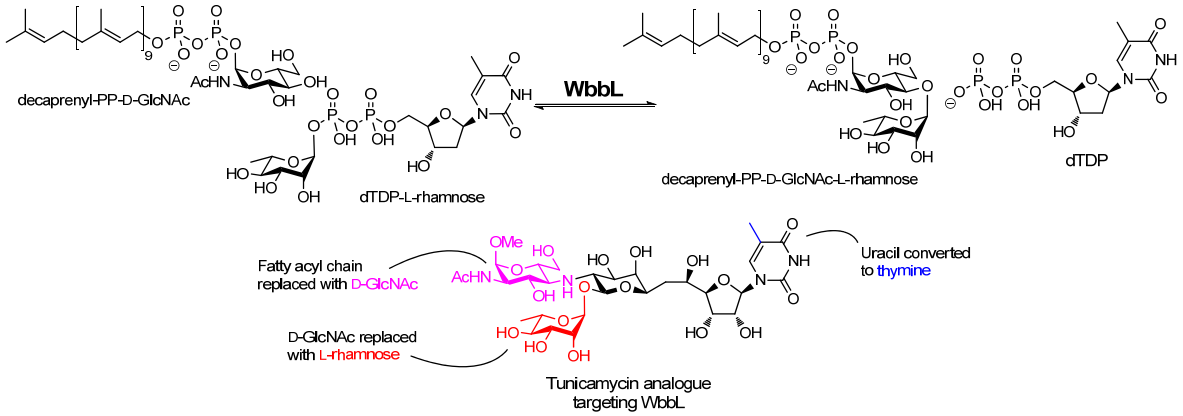


Fig. 1.34 Tunicamycin analogues as potential WbbL inhibitors and antituberculosis therapeutics

1.5.1.4 Dolichol-phosphate-mannose biosynthesis: DPM1

Dolichol-phosphate-mannose (DPM) is a membrane-bound mannose donor involved in *N*-linked glycosylation, GPI anchor biosynthesis and *O*- and *C*-protein mannosylation in eukaryotic cells.[281] It is synthesised from GDP-mannose and dolichol phosphate by a glycosyltransfer reaction catalysed by DPM synthase.[281] Inhibitors of these enzymes would act as useful tools in studying the biological function of these post-translational modifications, as well as in the structural and biochemical investigations of the specific enzymes involved. Tunicamycin analogues could be envisaged to inhibit DPM synthase if the terminal GlcNAc moiety were replaced with mannose and the uracil base were replaced

with guanine, resulting in a non-hydrolysable GDP-mannose mimic attached to a fatty acyl chain homologous to dolichol phosphate (Fig. 1.35).

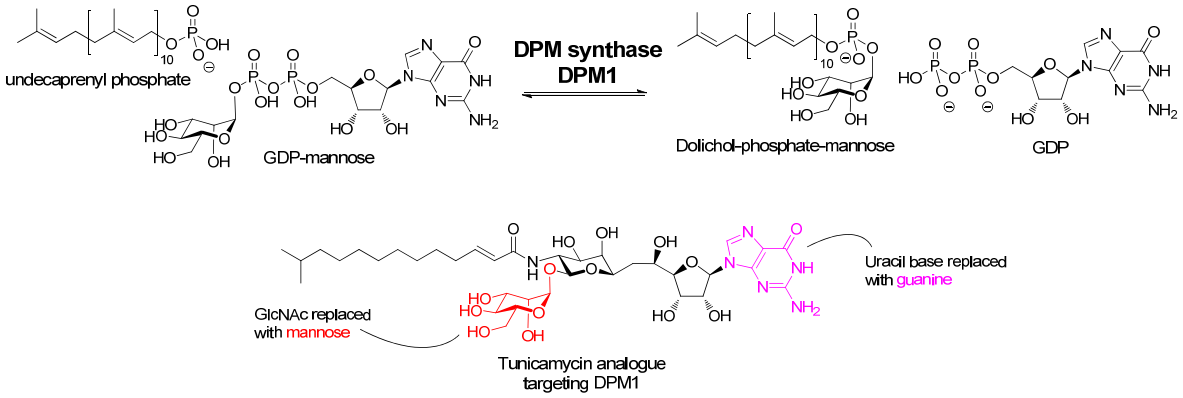


Fig. 1.35 Tunicamycin analogues as potential inhibitors of dolichol-phosphate-mannose biosynthesis

1.5.2 Strategy

A number of strategies could be adopted to access rationally designed tunicamycin analogues. As described in section 1.3, several synthetic routes to tunicamycin and its key intermediates have been developed and as such a purely synthetic strategy could be envisaged. However, these published protocols all suffer from low overall efficiency and involve a large number of chemical steps. Moreover, since the tunicamycins can be isolated in large quantities by fermentation of *Streptomyces chartreusis*, advanced precursors can be accessed through chemical degradation of isolated natural product and derivatised through relay synthesis, offering a more effective alternative to *de novo* synthesis. To complement such an approach, unlocking the toolbox encoded by the tunicamycin biosynthetic genes would be of great value. Following isolation of such a gene cluster, individual enzymes could be expressed, potentially enabling chemoenzymatic diversification of both natural intermediates isolated by chemical degradation of isolated tunicamycins and unnatural variants of these intermediates accessed through existing synthetic methods (Fig. 1.36).

As such, our initial focus is to identify the biosynthetic genes responsible for tunicamycin formation in producing organisms. Following experimental analysis of this gene cluster, individual genes will be expressed and the resulting enzymes biochemically characterised to ascertain their precise function in the biosynthetic pathway. The scope of their substrate specificity will also be determined, which is of direct relevance to the subsequent production of unnatural tunicamycin variants. Additional benefits of targeting

the tunicamycin biosynthetic pathway are numerous. First of all, a fundamental understanding can be gained of how nature produces some of the rare and even unique linkages present in tunicamycin. This could potentially lead to the discovery of novel enzymatic mechanisms or even contribute to novel synthetic methodologies. Secondly, knowledge of the genetic basis for tunicamycin production could lead to the generation of targeted mutants, with the deletion or upregulation of specific genes. These mutants could be used for *in vivo* production of advanced biosynthetic intermediates to tunicamycin, or semisynthetic production of fully formed analogues through the feeding of unnatural substrates.

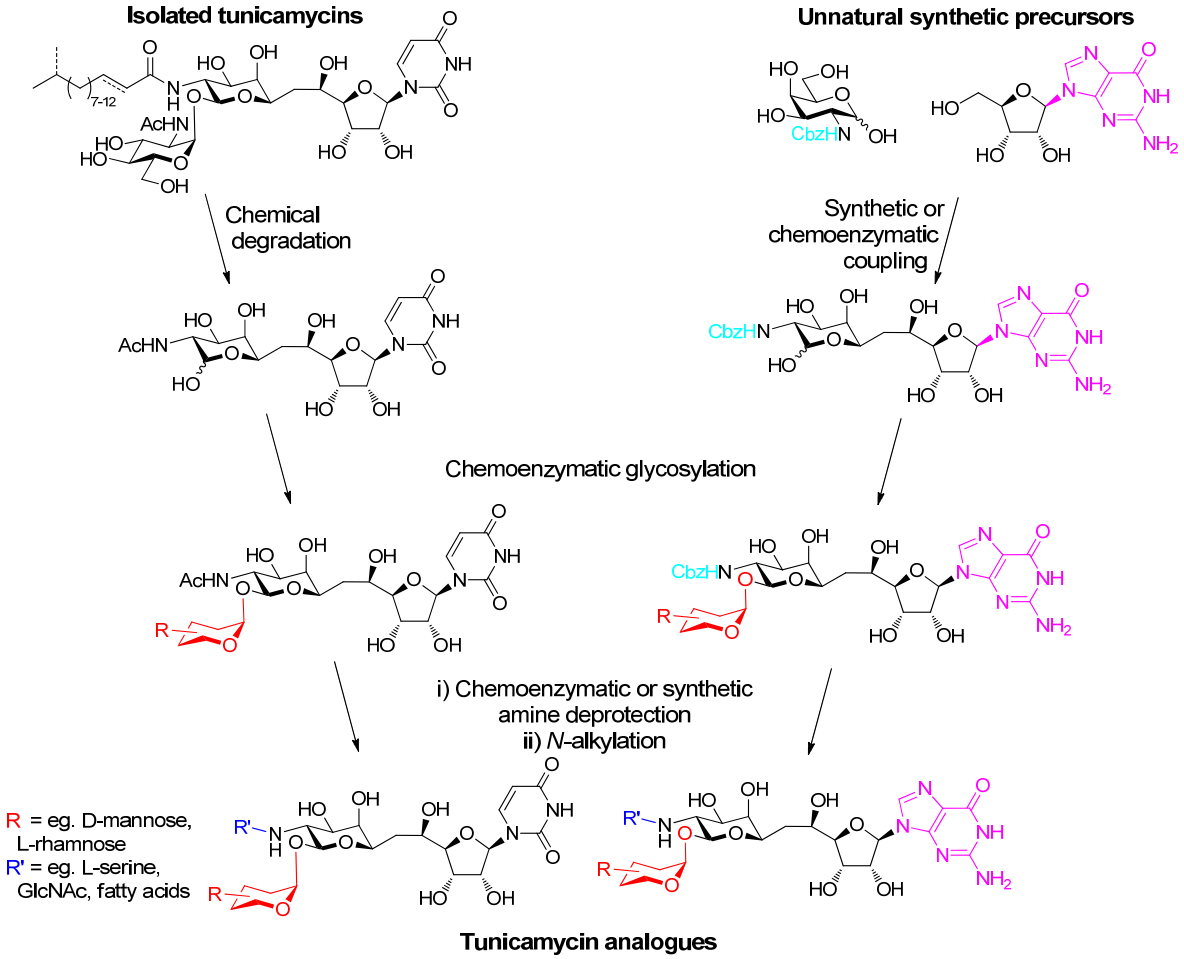


Fig. 1.36 A combined relay, chemoenzymatic and synthetic strategy to tunicamycin analogues

1.6 References

1. Eckardt K (1983): **Tunicamycins, Streptovirudins, and Corynetoxins, a Special Subclass of Nucleoside Antibiotics**. *J. Nat. Prod.* **46**: 544-550.
2. Tamura G (1982): *Tunicamycin.* Japan Scientific Societies Press: Tokyo.
3. A search performed on 19/04/2010 using ISI Web of Knowledge (http://wok.mimas.ac.uk) yielded 5,589 citations containing the topic 'tunicamycin'.
4. Takatsuki A, Arima K, Tamura G (1971): **Tunicamycin, a New Antibiotic: 1. Isolation and Characterization of Tunicamycin**. *J. Antibiot.* **24**: 215-223.
5. Hamill RG, Eli Lilly Co. (1980): **Process for preparing tunicamycin**. U. S. Patent 4237225
6. Hamill RG, Eli Lilly Co. (1982): **Process for preparing tunicamycin**. U. S. Patent 4336333
7. Takatsuki A, Kawamura K, Okina M, Kodama Y, Ito T, Tamura G (1977): **Structural Elucidation of Tunicamycin: 2. Structure of Tunicamycin**. *Agr. Biol. Chem.* **41**: 2307-2309.
8. Taga T, Miwa Y, Min Z (1997): **α,β-Trehalose monohydrate**. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* **53**: 234-236.
9. Keenan RW, Hamill RL, Occolowitz JL, Elbein AD (1981): **Biological Activities of Isolated Tunicamycin and Streptovirudin Fractions**. *Biochemistry* **20**: 2968-2973.
10. Mahoney WC, Duksin D (1979): **Biological Activities of the 2 Major Components of Tunicamycin**. *J. Biol. Chem.* **254**: 6572-6576.
11. Mahoney WC, Duksin D (1980): **Separation of Tunicamycin Homologs by Reversed-Phase High-Performance Liquid-Chromatography**. *J. Chromatogr.* **198**: 506-510.
12. Ito T, Takatsuki A, Kawamura K, Sato K, Tamura G (1980): **Structural Elucidation of Tunicamycin: 5. Isolation and Structures of Components of Tunicamycin**. *Agr. Biol. Chem.* **44**: 695-698.
13. Tsvetanova BC, Price NPJ (2001): **Liquid Chromatography-Electrospray Mass Spectrometry of Tunicamycin-Type Antibiotics**. *Anal. Biochem.* **289**: 147-156.
14. Thrum H, Eckardt K, Bradler G, Fugner R, Tonew E, Tonew M (1975): **Streptovirudins, New Antibiotics with Antibacterial and Antiviral Activity: 1. Culture Taxonomy, Fermentation and Production of Streptovirudin Complex**. *J. Antibiot.* **28**: 514-521.
15. Eckardt K, Thrum H, Bradler G, Tonew E, Tonew M (1975): **Streptovirudins, New Antibiotics with Antibacterial and Antiviral Activity: 2. Isolation, Chemical Characterization and Biological-Activity of Streptovirudins A1, A2, B1, B2, C1, C2, D1 and D2**. *J. Antibiot.* **28**: 274-279.
16. Eckardt K, Ihn W, Tresselt D, Krebs D (1981): **The Chemical Structures of Streptovirudins**. *J. Antibiot.* **34**: 1631-1632.
17. Vogel P, Petterson DS, Berry PH, Frahn JL, Anderton N, Cockrum PA, Edgar JA, Jago MV, Lanigan GW, Payne AL, et al. (1981): **Isolation of a Group of Glycolipid Toxins from Seedheads of Annual Ryegrass (*Lolium rigidum* Gaud.) Infected by *Corynebacterium rathayi***. *Aust. J. Exp. Biol. Med. Sci.* **59**: 455-467.
18. Edgar JA, Frahn JL, Cockrum PA, Anderton N, Jago MV, Culvenor CCJ, Jones AJ, Murray K, Shaw KJ (1982): **Corynetoxins, Causative Agents of Annual Ryegrass Toxicity - Their Identification as Tunicamycin Group Antibiotics**. *J. Chem. Soc., Chem. Commun.*: 222-224.
19. Cockrum PA, Edgar JA (1983): **Corynetoxin: A Chromatographic Study**. *Toxicon*: 65-68.

20. Frahn JL, Edgar JA, Jones AJ, Cockrum PA, Anderton N, Culvenor CCJ (1984): **Structure of the Corynetoxins, Metabolites of Corynebacterium-Rathayi Responsible for Toxicity of Annual Ryegrass (Lolium-Rigidum) Pastures**. *Aust. J. Chem.* **37**: 165-182.

21. Jago MV, Payne AL, Peterson JE, Bagust TJ (1983): **Inhibition of Glycosylation by Corynetoxin, the Causative Agent of Annual Ryegrass Toxicity - a Comparison with Tunicamycin**. *Chem.-Biol. Interact.* **45**: 223-234.

22. Finnie JW, Mukherjee TM (1986): **Ultrastructural-Changes in the Cerebellum of Nursling Rats Given Corynetoxin, the Etiologic Agent of Annual Ryegrass Toxicity**. *J. Comp. Pathol.* **96**: 205-216.

23. Jago MV, Culvenor CCJ (1987): **Tunicamycin and Corynetoxin Poisoning in Sheep**. *Aust. Vet. J.* **64**: 232-235.

24. Kenig M, Reading C (1979): **Holomycin and an Antibiotic (MM 19290) Related to Tunicamycin, Metabolites of *Streptomyces clavuligerus***. *J. Antibiot.* **32**: 549-554.

25. Mizuno M, Shimojima Y, Sugawara T, Takeda I (1971): **Antibiotic 24010**. *J. Antibiot.* **24**: 896-899.

26. Nakamura S, Arai M, Karasawa K, Yonehara H (1957): **On an Antibiotic, Mycospocidin**. *J. Antibiot.* **10**: 248-253.

27. Cockrum PA, Edgar JA (1983): **High-Performance Liquid-Chromatographic Comparison of the Tunicaminyluracil-Based Antibiotics Corynetoxin, Tunicamycin, Streptovirudin and MM-19290**. *J. Chromatogr.* **268**: 245-254.

28. Tkacz JS, Wong A (1978): **Comparison of Biological and Chemical Properties of Tunicamycin and Mycospocidin**. *Fed. Proc.* **37**: 1766-1766.

29. Yamamori S, Murazumi N, Araki Y, Ito E (1978): **Formation and Function of *N*-Acetylglucosamine-Linked Phosphorylundecaprenols and Pyrophosphorylundecaprenols in Membranes from *Bacillus cereus***. *J. Biol. Chem.* **253**: 6516-6522.

30. Elbein AD, Gafford J, Kang MS (1979): **Inhibition of Lipid-Linked Saccharide Synthesis - Comparison of Tunicamycin, Streptovirudin, and Antibiotic-24010**. *Arch. Biochem. Biophys.* **196**: 311-318.

31. Takatsuki A, Tamura G (1971): **Tunicamycin, a New Antibiotic: 2. Some Biological Properties of Antiviral Activity of Tunicamycin**. *J. Antibiot.* **24**: 224-231.

32. Takatsuki A, Tamura G (1971): **Effect of Tunicamycin on Synthesis of Macromolecules in Cultures of Chick Embryo Fibroblasts Infected with Newcastle Disease Virus**. *J. Antibiot.* **24**: 785-794.

33. Takatsuki A, Tamura G, Shimizu K (1971): **Effect of Tunicamycin on Microorganisms - Morphological Changes and Degradation of RNA and DNA Induced by Tunicamycin**. *J. Antibiot.* **25**: 75-85.

34. Katoh Y, Kuninaka A, Yoshino H, Takatsuki A, Yamasaki M, Tamura G (1976): **Formation of Fungal Multinuclear Giant-Cells by Tunicamycin**. *J. Gen. Appl. Microbiol.* **22**: 247-258.

35. Takatsuki A, Kohno K, Tamura G (1975): **Inhibition of Biosynthesis of Polyisoprenol Sugars in Chick-Embryo Microsomes by Tunicamycin**. *Agr. Biol. Chem.* **39**: 2089-2091.

36. Tkacz JS, Lampen JO (1975): **Tunicamycin Inhibition of Polyisoprenyl *N*-Acetylglucosaminyl Pyrophosphate Formation in Calf-Liver Microsomes**. *Biochem. Biophys. Res. Commun.* **65**: 248-257.

37. Tamura G, Sasaki T, Matsuhashi M, Takatsuki A, Yamasaki M (1976): **Tunicamycin Inhibits Formation of Lipid Intermediate in Cell-Free Peptidoglycan Synthesis of Bacteria**. *Agr. Biol. Chem.* **40**: 447-449.

38. Schleife KH, Kandler O (1972): **Peptidoglycan Types of Bacterial Cell-Walls and Their Taxonomic Implications**. *Bacteriol. Rev.* **36**: 407-477.

39. Vollmer W, Blanot D, de Pedro MA (2008): **Peptidoglycan structure and architecture**. *FEMS Microbiol. Rev.* **32**: 149-167.

40. Walsh C (2003): **Where will new antibiotics come from?** *Nat. Rev. Microbiol.* **1**: 65-70.

41. Winn M, Goss RJM, Kimura K-I, Bugg TDH (2010): **Antimicrobial nucleoside antibiotics targeting cell wall assembly: Recent advances in structure-function studies and nucleoside biosynthesis**. *Nat. Prod. Rep.* **27**: 279-304.

42. Bugg TDH, Walsh CT (1992): **Intracellular Steps of Bacterial-Cell Wall Peptidoglycan Biosynthesis - Enzymology, Antibiotics, and Antibiotic-Resistance**. *Nat. Prod. Rep.* **9**: 199-215.

43. Bouhss A, Trunkfield AE, Bugg TDH, Mengin-Lecreulx D (2008): **The biosynthesis of peptidoglycan lipid-linked intermediates**. *FEMS Microbiol. Rev.* **32**: 208-233.

44. Sauvage E, Kerff F, Terrak M, Ayala JA, Charlier P (2008): **The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis**. *FEMS Microbiol. Rev.* **32**: 234-258.

45. van Heijenoort J (2001): **Formation of the glycan chains in the synthesis of bacterial peptidoglycan**. *Glycobiology* **11**: 25r-36r.

46. van Heijenoort J (2001): **Recent advances in the formation of the bacterial peptidoglycan monomer unit**. *Nat. Prod. Rep.* **18**: 503-519.

47. van Heijenoort J (2007): **Lipid Intermediates in the Biosynthesis of Bacterial Peptidoglycan**. *Microbiol. Mol. Biol. Rev.* **71**: 620-635.

48. Schneider T, Sahl HG (2010): **An oldie but a goodie - cell wall biosynthesis as antibiotic target pathway**. *Int. J. Med. Microbiol.* **300**: 161-169.

49. Hendlin D, Stapley EO, Jackson M, Wallick H, Miller AK, Wolf FJ, Miller TW, Chaiet L, Kahan FM, Foltz EL, et al. (1969): **Phosphonomycin, a New Antibiotic Produced by Strains of Streptomyces**. *Science* **166**: 122-123.

50. Kahan FM, Kahan JS, Cassidy PJ, Kropp H (1974): **Mechanism of Action of Fosfomycin (Phosphonomycin)**. *Ann. N.Y. Acad. Sci.* **235**: 364-386.

51. Neuhaus FC, Hammes WP (1981): **Inhibition of Cell-Wall Biosynthesis by Analogs of Alanine**. *Pharmac. Ther.* **14**: 265-319.

52. Lovering AL, de Castro LH, Lim D, Strynadka NCJ (2007): **Structural Insight into the Transglycosylation Step of Bacterial Cell-Wall Biosynthesis**. *Science* **315**: 1402-1405.

53. Reynolds PE (1989): **Structure, Biochemistry and Mechanism of Action of Glycopeptide Antibiotics**. *Eur. J. Clin. Microbiol. Infect. Dis.* **8**: 943-950.

54. Williams DH, Bardsley B (1999): **The vancomycin group of antibiotics and the fight against resistant bacteria**. *Angew. Chem., Int. Ed.* **38**: 1173-1193.

55. Kahne D, Leimkuhler C, Wei L, Walsh C (2005): **Glycopeptide and lipoglycopeptide antibiotics**. *Chem. Rev.* **105**: 425-448.

56. Van Bambeke F, Mingeot-Leclercq MP, Struelens MJ, Tulkens PM (2008): **The bacterial envelope as a target for novel anti-MRSA antibiotics**. *Trends Pharmacol. Sci.* **29**: 124-134.

57. Bierbaum G, Sahl HG (2009): **Lantibiotics: Mode of Action, Biosynthesis and Bioengineering**. *Curr. Pharm. Biotechnol.* **10**: 2-18.

58. Breukink E, de Kruijff B (2006): **Lipid II as a target for antibiotics**. *Nat. Rev. Drug Discovery* **5**: 321-332.

59. Mygind PH, Fischer RL, Schnorr KM, Hansen MT, Sonksen CP, Ludvigsen S, Raventos D, Buskov S, Christensen B, De Maria L, et al. (2005): **Plectasin is a peptide antibiotic with therapeutic potential from a saprophytic fungus**. *Nature* **437**: 975-980.

60. Baltz RH (2009): **Daptomycin: mechanisms of action and resistance, and biosynthetic engineering**. *Curr. Opin. Chem. Biol.* **13**: 144-151.

61. Stone KJ, Strominger Jl (1971): **Mechanism of Action of Bacitracin - Complexation with Metal Ion and C55-Isoprenyl Pyrophosphate**. *Proc. Natl. Acad. Sci. U.S.A.* **68**: 3223-3227.

62. Schneider T, Gries K, Josten M, Wiedemann I, Pelzer S, Labischinski H, Sahl HG (2009): **The Lipopeptide Antibiotic Friulimicin B Inhibits Cell Wall Biosynthesis through Complex Formation with Bactoprenol Phosphate**. *Antimicrob. Agents Chemother.* **53**: 1610-1618.

63. Demain AL, Elander RP (1999): **The beta-lactam antibiotics: past, present, and future**. *Antonie Leeuwenhoek* **75**: 5-19.

64. Hou JP, Poole JW (1971): **Beta-Lactam Antibiotics - Their Physicochemical Properties and Biological Activities in Relation to Structure**. *J. Pharm. Sci.* **60**: 503-532.

65. Waxman DJ, Strominger JL (1983): **Penicillin-Binding Proteins and the Mechanism of Action of Beta-Lactam Antibiotics**. *Annu. Rev. Biochem.* **52**: 825-869.

66. Anderson MS, Eveland SS, Price NPJ (2000): **Conserved cytoplasmic motifs that distinguish sub-groups of the polyprenol phosphate: N-acetylhexosamine-1-phosphate transferase family**. *FEMS Microbiol. Lett.* **191**: 169-175.

67. Price NP, Momany FA (2005): **Modeling bacterial UDP-HexNAc: polyprenol-P HexNAc-1-P transferases**. *Glycobiology* **15**: 29R-42R.

68. Bouhss A, Mengin-Lecreulx D, Le Beller D, van Heijenoort J (1999): **Topological analysis of the MraY protein catalysing the first membrane step of peptidoglycan synthesis**. *Mol. Microbiol.* **34**: 576-585.

69. Lloyd AJ, Brandish PE, Gilbey AM, Bugg TDH (2004): **Phospho-N-acetyl-muramyl-pentapeptide translocase from *Escherichia coli*: Catalytic role of conserved aspartic acid residues**. *J. Bacteriol.* **186**: 1747-1757.

70. Heydanek MG, Struve WG, Neuhaus FC (1969): **On Initial Stage in Peptidoglycan Synthesis: 3. Kinetics and Uncoupling of Phospho-N-Acetylmuramyl-Pentapeptide Translocase (Uridine 5'-Phosphate)**. *Biochemistry* **8**: 1214-1221.

71. Pless DD, Neuhaus FC (1973): **Initial Membrane Reaction in Peptidoglycan Synthesis - Lipid Dependence of Phospho-N-Acetylmuramyl-Pentapeptide Translocase (Exchange-Reaction)**. *J. Biol. Chem.* **248**: 1568-1576.

72. Bouhss A, Crouvoisier M, Blanot D, Mengin-Lecreulx D (2004): **Purification and characterization of the bacterial MraY translocase catalyzing the first membrane step of peptidoglycan biosynthesis**. *J. Biol. Chem.* **279**: 29974-29980.

73. Inukai M, Isono F, Takatsuki A (1993): **Selective Inhibition of the Bacterial Translocase Reaction in Peptidoglycan Synthesis by Mureidomycins**. *Antimicrob. Agents Chemother.* **37**: 980-983.

74. Brandish PE, Kimura K, Inukai M, Southgate R, Lonsdale JT, Bugg TDH (1996): **Modes of action of tunicamycin, liposidomycin B, and mureidomycin A: Inhibition of phospho-N-acetylmuramyl-pentapeptide translocase from *Escherichia coli***. *Antimicrob. Agents Chemother.* **40**: 1640-1644.

75. Ward JB (1977): **Tunicamycin Inhibition of Bacterial Wall Polymer Synthesis**. *FEBS Lett.* **78**: 151-154.

76. Kimura K-I, Bugg TDH (2003): **Recent advances in antimicrobial nucleoside antibiotics targeting cell wall biosynthesis**. *Nat. Prod. Rep.* **20**: 252-273.

77. Bernhardt TG, Roof WD, Young R (2000): **Genetic evidence that the bacteriophage phi X174 lysis protein inhibits cell wall synthesis**. *Proc. Natl. Acad. Sci. U.S.A.* **97**: 4297-4302.

78. Bernhardt TG, Struck DK, Young R (2001): **The lysis protein E of phi X174 is a specific inhibitor of the MraY-catalyzed step in peptidoglycan synthesis**. *J. Biol. Chem.* **276**: 6093-6097.

79. Zheng Y, Struck DK, Young R (2009): **Purification and Functional Characterization of phi X174 Lysis Protein E**. *Biochemistry* **48**: 4999-5006.

80. Lowe JB, Marth JD (2003): **A Genetic Approach to Mammalian Glycan Function**. *Annu. Rev. Biochem.* **72**: 643-691.

81. Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (2008): *Essentials of Glycobiology* 2nd edn. Cold Spring Harbor Laboratory Press: New York.

82. Kornfeld R, Kornfeld S (1985): **Assembly of asparagine-linked oligosaccharides**. *Annu. Rev. Biochem.* **54**: 631-664.

83. Helenius A, Aebi M (2001): **Intracellular Functions of N-Linked Glycans**. *Science* **291**: 2364-2369.

84. Helenius A, Aebi M (2004): **Roles of N-linked glycans in the endoplasmic reticulum**. *Annu. Rev. Biochem.* **73**: 1019-1049.

85. Helenius A, Trombetta ES, Hebert DN, Simons JF (1997): **Calnexin, calreticulin and the folding of glycoproteins**. *Trends in Cell Biology* **7**: 193-200.

86. Zapun A, Jakob CA, Thomas DY, Bergeron JJM (1999): **Protein folding in a specialized compartment: the endoplasmic reticulum**. *Structure* **7**: R173-R182.

87. Klausner RD, Sitia R (1990): **Protein Degradation in the Endoplasmic Reticulum**. *Cell* **62**: 611-614.

88. Plemper RK, Wolf DH (1999): **Retrograde protein translocation: ERADication of secretory proteins in health and disease**. *Trends Biochem. Sci.* **24**: 266-270.

89. Varki A (1993): **Biological roles of oligosaccharides: all of the theories are correct**. *Glycobiology* **3**: 97-130.

90. Paulson JC (1989): **Glycoproteins - What Are the Sugar Chains For?** *Trends Biochem. Sci.* **14**: 272-276.

91. Burda P, Aebi M (1999): **The dolichol pathway of *N*-linked glycosylation**. *Biochim. Biophys. Acta, Gen. Subj.* **1426**: 239-257.

92. Dwek RA, Butters TD, Platt FM, Zitzmann N (2002): **Targeting glycosylation as a therapeutic approach**. *Nat. Rev. Drug Discovery* **1**: 65-75.

93. Heifetz A, Keenan RW, Elbein AD (1979): **Mechanism of action of tunicamycin on the UDP-GlcNAc:dolichyl-phosphate GlcNAc-1-phosphate transferase**. *Biochemistry* **18**: 2186-2192.

94. Lehle L, Tanner W (1976): **The specific site of tunicamycin inhibition in the formation of dolichol-bound *N*-acetylglucosamine derivatives**. *FEBS Lett.* **71**: 167-170.

95. Elbein AD (1981): **The tunicamycins - useful tools for studies on glycoproteins**. *Trends Biochem. Sci.* **6**: 219-221.

96. Elbein AD (1987): **Inhibitors of the Biosynthesis and Processing of N-Linked Oligosaccharide Chains**. *Annu. Rev. Biochem.* **56**: 497-534.

97. Xu L, Appell M, Kennedy S, Momany Frank A, Price Neil PJ (2004): **Conformational analysis of chirally deuterated tunicamycin as an active site probe of UDP-N-acetylhexosamine:polyprenol-P N-acetylhexosamine-1-P translocases**. *Biochemistry* **43**: 13248-13255.

98. McDowell W, Schwarz RT (1988): **Dissecting glycoprotein biosynthesis by the use of specific inhibitors**. *Biochimie* **70**: 1535-1549.

99. Wöhler F (1828): **Über künstliche Bildung des Harnstoffs.** *Ann. Phys. Chem.* **12**: 253-256.

100. Nicolaou KC, Vourloumis D, Winssinger N, Baran PS (2000): **The art and science of total synthesis at the dawn of the twenty-first century**. *Angew. Chem., Int. Ed.* **39**: 44-122.

101. Sasai H, Matsuno K, Suami T (1985): **Synthetic Approach toward Antibiotic Tunicamycins: 7. Synthesis of Tunicamine and Tunicaminyl Uracil Derivative**. *J. Carbohydr. Chem.* **4**: 99-112.

102. Suami T, Sasai H, Matsuno K, Suzuki N, Fukuda Y, Sakanaka O (1984): **Synthetic approach toward antibiotic tunicamycins. VI. total synthesis of tunicamycins**. *Tetrahedron Lett.* **25**: 4533-4536.

103. Fukuda Y, Sasai H, Suami T (1981): **Synthetic Approach toward Antibiotic Tunicamycins**. *Bull. Chem. Soc. Jpn.* **54**: 1830-1834.

104. Fukuda Y, Kitasato H, Sasai H, Suami T (1982): **Synthetic Approach toward Antibiotic Tunicamycins. 2.** *Bull. Chem. Soc. Jpn.* **55**: 880-886.

105. Fukuda Y, Sasai H, Suami T (1982): **Synthetic Approach toward Antibiotic Tunicamycins .3. Methyl 3,4,7,8-Tetra-O-Acetyl-10-O-Benzyl-2-Benzyloxycarbonylamino-2,6-Dideoxy-11,12-O-Isopropylidene-Beta-L-Dodecodialdo-(12r)-Furanose-(12,9)-Pyranosides-(1,5)**. *Bull. Chem. Soc. Jpn.* **55**: 1574-1578.

106. Suami T, Sasai H, Matsuno K, Suzuki N (1985): **Synthetic approaches toward antibiotic tunicamycins. Part VIII. Total synthesis of tunicamycin**. *Carbohydr. Res.* **143**: 85-96.

107. Suami T, Sasai H, Matsuno K (1983): **Synthetic approach toward antibiotic tunicamycins. V. Synthesis of methyl hexaacetyltunicaminyluracil**. *Chem. Lett.*: 819-822.

108. Kominato K, Ogawa S, Suami T (1988): **Synthesis of a 2-acetamido-2-deoxy-α-D-galactopyranosyl analogue of tunicamycin**. *Carbohydr. Res.* **174**: 360-368.

109. Haga M, Takano M, Tejima S (1970): **Diisopropylidene Acetals of D-Allose and D-Ribo-Hexulose (D-Psicose)**. *Carbohydr. Res.* **14**: 237-244.

110. Zyoong K (1956): *Sci. Sin. (Engl. Ed.)* **5**: 461.

111. Danishefsky SJ, Deninno SL, Chen S, Boisvert L, Barbachyn M (1989): **Fully Synthetic Stereoselective Routes to the Differentially Protected Subunits of the Tunicamycins**. *J. Am. Chem. Soc.* **111**: 5810-5818.

112. Danishefsky S, Barbachyn M (1985): **A fully synthetic route to tunicaminyluracil**. *J. Am. Chem. Soc.* **107**: 7761-7762.

113. Danishefsky SJ (1986): **Reflections on Organic Synthesis: The Evolution of a General Strategy for the Stereoselective Construction of Polyoxygenated Natural Products**. *Aldrichimica Acta* **19**: 59-69.

114. Danishefsky SJ, Maring CJ (1985): **A New Approach to the Synthesis of Hexoses - an Entry to (+/-)-Fucose and (+)-Daunosamine**. *J. Am. Chem. Soc.* **107**: 1269-1274.

115. Ramza J, Zamojski A (1992): **New convenient synthesis of tunicamine**. *Tetrahedron* **48**: 6123-6134.

116. Ramza J, Zamojski A (1992): **The synthesis of 5-*C*-(6-deoxy-1,2:3,4-di-*O*-isopropyidene-α-D-galactopyranos-6-yl)2,3-*O*-isopropylidene-D-*allo*-pentofuranose [deaminotri-*O*-isopropylidene tunicamine]**. *Carbohydr. Res.* **228**: 205-216.

117. Lemieux RU, Ratcliffe RM (1979): **Azidonitration of Tri-O-Acetyl-D-Galactal**. *Can. J. Chem.* **57**: 1244-1251.

118. Abramow VS, Kapustina AS (1957): *Zh. Obshch. Khim.*: 173-176.

119. Achmatowicz O, Bukowski P, Szechner B, Zwierzch.Z, Zamojski A (1971): **Synthesis of Methyl 2,3-Dideoxy-DL-alk-2-enopyranosides from Furan Compounds - General Approach to Total Synthesis of Monosaccharides**. *Tetrahedron* **27**: 1973-1996.

120. Achmatowicz O, Grynkiewicz G (1977): **Stereoselective Synthesis of 2,3-*O*-Isopropylidene-DL-Ribofuranose and Methyl β-DL-Ribopyranoside from Furfuryl Alcohol**. *Carbohydr. Res.* **54**: 193-198.

121. Myers AG, Gin DY, Widdowson KL (1991): **Silicon-mediated reductive coupling of aldehydes and allylic alcohols. A stereoselective synthesis of tunicaminyluracil**. *J. Am. Chem. Soc.* **113**: 9661-9663.

122. Myers AG, Gin DY, Rogers DH (1993): **A convergent synthetic route to the tunicamycin antibiotics. Synthesis of (+)-tunicamycin V**. *J. Am. Chem. Soc.* **115**: 2036-2038.

123. Myers AG, Gin DY, Rogers DH (1994): **Synthetic Studies of the Tunicamycin Antibiotics. Preparation of (+)-Tunicaminyluracil, (+)-Tunicamycin-V, and 5'-*epi*-Tunicamycin-V**. *J. Am. Chem. Soc.* **116**: 4697-4718.

124. Krajewski JW, Karpiesiuk W, Banaszek A (1994): **X-ray and conformational investigation of 6-deoxy-6-*C*-(1,5-di-*O*-acetyl-2,3-*O*-isopropylidene-β-D-*allo*-pentofuranos-5-yl)-1,2:3,4-di-*O*-isopropylidene-α-D-galactopyranose and 6-*C*-(benzyl 5-deoxy-2,3-*O*-isopropylidene-β-D-ribofuranosid-5-yl)-1,2:3,4-di-*O*-isopropylidene-D-*glycero*-α-D-*galacto*-hexopyranose**. *Carbohydr. Res.* **257**: 25-33.

125. Karpiesiuk W, Banaszek A (1994): **Highly stereoselective synthesis of α,β-linked, nonreducing disaccharides related to tunicamycin**. *Carbohydr. Res.* **261**: 243-253.

126. Karpiesiuk W, Banaszek A (1997): **Stereoselective syntheses of the *O*,*N*-protected subunits of the tunicamycins**. *Carbohydr. Res.* **299**: 245-252.

127. Karpiesiuk W, Banaszek A (1994): **Simple synthesis of *O*,*N*-protected tunicamine - the undecose part of tunicamycins**. *Bioorg. Med. Chem. Lett.* **4**: 879-882.

128. Karpiesiuk W, Banaszek A (1994): **Simple approach to *O*-protected deaminotunicaminyluracil**. *Tetrahedron* **50**: 2965-2974.

129. Lichtenthaler FW, Kaji E, Weprek S (1985): **Disaccharide-Derived 2-Oxo and 2-Oximinoglycosyl Bromides: Novel, Conveniently Accessible Building-Blocks for the Expedient Construction of Oligosaccharides with α-D-Glucosamine, β-D-Mannose, and β-D-Mannosamine as Constituent Sugars**. *J. Org. Chem.* **50**: 3505-3515.

130. Sarabia-García F, López-Herrera FJ, Pino-González MS (1995): **Unstabilized diazo derivatives from carbohydrates. Application to the synthesis of 2-deamino-tunicamine and products related to C-disaccharides**. *Tetrahedron* **51**: 5491-5500.

131. Sarabia-Garcia F, Lopez-Herrera FJ (1996): **Studies on the synthesis of tunicamycin. The preparation of 7-deoxy-2-deamino-6-hydroxy tunicamine and related products**. *Tetrahedron* **52**: 4757-4768.

132. Sarabia F, Martin-Ortiz L, Lopez-Herrera FJ (2003): **Synthetic studies towards the tunicamycins and analogues based on diazo chemistry. Total synthesis of tunicaminyl uracil**. *Org. Biomol. Chem.* **1**: 3716-3725.

133. Ichikawa S, Matsuda A (2004): **Synthesis of Tunicaminyluracil Derivatives**. *Nucleos. Nucleot. Nucleic Acids* **23**: 239-253.

134. Li JWH, Vederas JC (2009): **Drug Discovery and Natural Products: End of an Era or an Endless Frontier?** *Science* **325**: 161-165.

135. Newman DJ, Cragg GM (2007): **Natural Products as Sources of New Drugs over the Last 25 Years**. *J. Nat. Prod.* **70**: 461-477.

136. Watanabe K, Rude MA, Walsh CT, Khosla C (2003): **Engineered biosynthesis of an ansamycin polyketide precursor in *Escherichia coli***. *Proc. Natl. Acad. Sci. U.S.A.* **100**: 9774-9778.

137. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C (2001): **Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of *E. coli***. *Science* **291**: 1790-1792.

138. Julien B, Shah S (2002): **Heterologous Expression of Epothilone Biosynthetic Genes in *Myxococcus xanthus***. *Antimicrob. Agents Chemother.* **46**: 2772-2778.

139. Wenzel SC, Gross F, Zhang Y, Fu J, Stewart AF, Müller R (2005): **Heterologous Expression of a Myxobacterial Natural Products Assembly Line in Pseudomonads via Red/ET Recombineering**. *Chem. Biol.* **12**: 349-356.

140. Wohlleben W, Pelzer S (2002): **New Compounds By Combining "Modern" Genomics and "Old-Fashioned" Mutasynthesis**. *Chem. Biol.* **9**: 1163-1164.

141. Weist S, Süssmuth RD (2005): **Mutational biosynthesis—a tool for the generation of structural diversity in the biosynthesis of antibiotics**. *Appl. Microbiol. Biotechnol.* **68**: 141-150.

142. Staunton J, Wilkinson B (2001): **Combinatorial biosynthesis of polyketides and nonribosomal peptides**. *Curr. Opin. Chem. Biol.* **5**: 159-164.

143. Walsh CT (2002): **Combinatorial Biosynthesis of Antibiotics: Challenges and Opportunities**. *ChemBioChem* **3**: 124-134.

144. Walsh CT (2004): **Polyketide and Nonribosomal Peptide Antibiotics: Modularity and Versatility**. *Science* **303**: 1805-1810.

145. Bode HB, Müller R (2007): **Reversible Sugar Transfer by Glycosyltransferases as a Tool for Natural Product (Bio)synthesis**. *Angew. Chem., Int. Ed.* **46**: 2147-2150.

146. Corre C, Challis GL (2009): **New natural product biosynthetic chemistry discovered by genome mining**. *Nat. Prod. Rep.* **26**: 977-986.

147. Watve MG, Tickoo R, Jog MM, Bhole BD (2001): **How many antibiotics are produced by the genus Streptomyces?** *Arch. Microbiol.* **176**: 386-390.

148. Hopwood DA (2007): *Streptomyces in Nature and Medicine: The Antibiotic Makers.* Oxford University Press: New York.

149. Bode HB, Müller R (2005): **The Impact of Bacterial Genomics on Natural Product Research**. *Angew. Chem., Int. Ed.* **44**: 6828-6846.

150. Zerikly M, Challis GL (2009): **Strategies for the Discovery of New Natural Products by Genome Mining**. *ChemBioChem* **10**: 625-633.

151. Hertweck C (2009): **The Biosynthetic Logic of Polyketide Diversity**. *Angew. Chem., Int. Ed.* **48**: 4688-4716.

152. Mohamed AM (2009): **Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis**. *J. Pept. Sci.* **15**: 799-807.

153. Fischbach MA, Walsh CT (2006): **Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms**. *Chem. Rev.* **106**: 3468-3496.

154. Walsh CT, Fischbach MA (2010): **Natural Products Version 2.0: Connecting Genes to Molecules**. *J. Am. Chem. Soc.* **132**: 2469-2493.

155. Cortes J, Haydock SF, Roberts GA, Bevitt DJ, Leadlay PF (1990): **An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea***. *Nature* **348**: 176-178.

156. Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L (1991): **Modular organization of genes required for complex polyketide biosynthesis**. *Science* **252**: 675-679.

157. Ahlert J, Shepard E, Lomovskaya N, Zazopoulos E, Staffa A, Bachmann BO, Huang K, Fonstein L, Czisny A, Whitwam RE, et al. (2002): **The Calicheamicin Gene Cluster and Its Iterative Type I Enediyne PKS**. *Science* **297**: 1173-1176.

158. Kim E-S, Bibb MJ, Butler MJ, Hopwood DA, Sherman DH (1994): **Sequences of the oxytetracycline polyketide synthase-encoding *otc* genes from *Streptomyces rimosus***. *Gene* **141**: 141-142.

159. van Wageningen AA, Kirkpatrick PN, Williams DH, Harris BR, Kershaw JK, Lennard NJ, Jones M, Jones SJM, Solenberg PJ (1998): **Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic**. *Chem. Biol.* **5**: 155-162.

160. Aharonowitz Y, Cohen G, Martin JF (1992): **Penicillin and Cephalosporin Biosynthetic Genes: Structure, Organization, Regulation, and Evolution**. *Annu. Rev. Microbiol.* **46**: 461-495.

161. Martin JF (1998): **New aspects of genes and enzymes for beta-lactam antibiotic biosynthesis**. *Appl. Microbiol. Biotechnol.* **50**: 1-15.

162. Wu K, Chung L, Revill WP, Katz L, Reeves CD (2000): **The FK520 gene cluster of *Streptomyces hygroscopicus* var. *ascomyceticus* (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units**. *Gene* **251**: 81-90.

163. Haideh M, Ali S (1998): **The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506**. *Eur. J. Biochem.* **256**: 528-534.

164. Schwecke T, Aparicio JF, Molnar I, Konig A, Khaw LE, Haydock SF, Oliynyk M, Caffrey P, Cortes J, Lester JB (1995): **The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin**. *Proc. Natl. Acad. Sci. U.S.A.* **92**: 7839-7843.

165. Aparicio JF, Molnár I, Schwecke T, König A, Haydock SF, Ee Khaw L, Staunton J, Leadlay PF (1996): **Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: Analysis of the enzymatic domains in the modular polyketide synthase**. *Gene* **169**: 9-16.

166. Christianson DW (2008): **Unearthing the roots of the terpenome**. *Curr. Opin. Chem. Biol.* **12**: 141-150.

167. Daum M, Herrmann S, Wilkinson B, Bechthold A (2009): **Genes and enzymes involved in bacterial isoprenoid biosynthesis**. *Curr. Opin. Chem. Biol.* **13**: 180-188.

168. Dairi T (2005): **Studies on biosynthetic genes and enzymes of isoprenoids produced by actinomycetes**. *J. Antibiot.* **58**: 227-243.

169. Poulter CD, Rilling HC (1978): **The prenyl transfer reaction. Enzymic and mechanistic studies of the 1'-4 coupling reaction in the terpene biosynthetic pathway**. *Acc. Chem. Res.* **11**: 307-313.

170. Kellogg BA, Poulter CD (1997): **Chain elongation in the isoprenoid biosynthetic pathway**. *Curr. Opin. Chem. Biol.* **1**: 570-578.

171. Brown GD (1998): **The biosynthesis of steroids and triterpenoids**. *Nat. Prod. Rep.* **15**: 653-696.

172. Wendt KU, Georg ES, Elias JC, David RL (2000): **Enzyme Mechanisms for Polycyclic Triterpene Formation**. *Angew. Chem., Int. Ed.* **39**: 2812-2833.

173. Christianson DW (2006): **Structural Biology and Chemistry of the Terpenoid Cyclases**. *Chem. Rev.* **106**: 3412-3442.

174. Cane DE, Swanson S, Murthy PPN (1981): **Trichodiene biosynthesis and the enzymic cyclization of farnesyl pyrophosphate**. *J. Am. Chem. Soc.* **103**: 2136-2138.

175. Hezari M, Lewis NG, Croteau R (1995): **Purification and Characterization of Taxa-4(5),11(12)-diene Synthase from Pacific Yew (*Taxus Brevifolia*) that Catalyzes the First Committed Step of Taxol Biosynthesis**. *Arch. Biochem. Biophys.* **322**: 437-444.

176. Williams DC, Wildung MR, Jin AQ, Dalal D, Oliver JS, Coates RM, Croteau R (2000): **Heterologous Expression and Characterization of a "Pseudomature" Form of Taxadiene Synthase Involved in Paclitaxel (Taxol) Biosynthesis and Evaluation of a Potential Intermediate and Inhibitors of the Multistep Diterpene Cyclization Reaction**. *Arch. Biochem. Biophys.* **379**: 137-146.

177. Whitehead IM, Threlfall DR, Ewing DF (1989): **5-*epi*-aristolochene is a common precursor of the sesquiterpenoid phytoalexins capsidiol and debneyol**. *Phytochemistry* **28**: 775-779.

178. Back KW, Yin SH, Chappell J (1994): **Expression of a Plant Sesquiterpene Cyclase Gene in *Escherichia coli***. *Arch. Biochem. Biophys.* **315**: 527-532.

179. Starks CM, Back K, Chappell J, Noel JP (1997): **Structural Basis for Cyclic Terpene Biosynthesis by Tobacco 5-Epi-Aristolochene Synthase**. *Science* **277**: 1815-1820.

180. Thoma R, Schulz-Gasch T, D'Arcy B, Benz J, Aebi J, Dehmlow H, Hennig M, Stihle M, Ruf A (2004): **Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase**. *Nature* **432**: 118-122.

181. Ruf A, Müller F, D'Arcy B, Stihle M, Kusznir E, Handschin C, Morand OH, Thoma R (2004): **The monotopic membrane protein human oxidosqualene cyclase is active as monomer**. *Biochem. Biophys. Res. Commun.* **315**: 247-254.

182. Flatt PM, Mahmud T (2007): **Biosynthesis of aminocyclitol-aminoglycoside antibiotics and related compounds**. *Nat. Prod. Rep.* **24**: 358-392.

183. Wong C-H (2003): *Carbohydrate-based Drug Discovery.* Wiley VCH: Weinheim, Germany.

184. Hu YN, Walker S (2002): **Remarkable structural similarities between diverse glycosyltransferases**. *Chem. Biol.* **9**: 1287-1296.

185. Lairson LL, Henrissat B, Davies GJ, Withers SG (2008): **Glycosyltransferases: Structures, functions, and mechanisms**. *Annu. Rev. Biochem.* **77**: 521-555.

186. Salas JA, Méndez C (2007): **Engineering the glycosylation of natural products in actinomycetes**. *Trends Microbiol.* **15**: 219-232.

187. He XM, Liu H-w (2002): **Formation of Unusual Sugars: Mechanistic Studies and Biosynthetic Applications**. *Annu. Rev. Biochem.* **71**: 701-754.

188. Blanchard S, Thorson JS (2006): **Enzymatic tools for engineering natural product glycosylation**. *Curr. Opin. Chem. Biol.* **10**: 263-271.

189. *The Alkaloids (Book Series)* (1950 - present). Edited by G. A. Cordell. Natural Products Inc.: Evanston, IL.

190. Holger S, Michael W (2009): **Medicinally important secondary metabolites in recombinant microorganisms or plants: Progress in alkaloid biosynthesis**. *Biotechnol. J.* **4**: 1684-1703.

191. Dewick PM (2009): *Medicinal Natural Products: A Biosynthetic Approach* 3rd edn. John Wiley & Sons Ltd: Chichester.

192. Facchini PJ (2001): **Alkaloid Biosynthesis in Plants: Biochemistry, Cell Biology, Molecular Regulation, and Metabolic Engineering Applications**. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**: 29-66.

193. Kutchan TM (1995): **Alkaloid Biosynthesis - The Basis for Metabolic Engineering of Medicinal Plants**. *Plant Cell* **7**: 1059-1070.

194. Floss HG (1997): **Natural products derived from unusual variants of the shikimate pathway**. *Nat. Prod. Rep.* **14**: 433-452.

195. Knaggs AR (1999): **The biosynthesis of shikimate metabolites**. *Nat. Prod. Rep.* **16**: 525-560.

196. Knaggs AR (2001): **The biosynthesis of shikimate metabolites**. *Nat. Prod. Rep.* **18**: 334-355.

197. Knaggs AR (2003): **The biosynthesis of shikimate metabolites**. *Nat. Prod. Rep.* **20**: 119-136.

198. Drechsel H, Jung G (1998): **Peptide siderophores**. *J. Pept. Sci.* **4**: 147-181.

199. Crosa JH, Walsh CT (2002): **Genetics and Assembly Line Enzymology of Siderophore Biosynthesis in Bacteria**. *Microbiol. Mol. Biol. Rev.* **66**: 223-249.

200. Gregory LC (2005): **A Widely Distributed Bacterial Pathway for Siderophore Biosynthesis Independent of Nonribosomal Peptide Synthetases**. *ChemBioChem* **6**: 601-611.

201. Perigaud C, Gosselin G, Imbach JL (1992): **Nucleoside Analogues as Chemotherapeutic Agents: A Review**. *Nucleosides and Nucleotides* **11**: 903 - 945.

202. Isono K (1988): **Nucleoside Antibiotics - Structure, Biological-Activity, and Biosynthesis**. *J. Antibiot.* **41**: 1711-1739.

203. Ichikawa S, Matsuda A (2007): **Nucleoside natural products and related analogs with potential therapeutic properties as antibacterial and antiviral agents**. *Expert Opin. Ther. Pat.* **17**: 487-498.

204. Isono K, Uramoto M, Kusakabe H, Kimura KI, Izaki K, Nelson CC, Mccloskey JA (1985): **Liposidomycins - Novel Nucleoside Antibiotics Which Inhibit Bacterial Peptidoglycan Synthesis**. *J. Antibiot.* **38**: 1617-1621.

205. Igarashi M, Nakagawa N, Doi N, Hattori S, Naganawa H, Hamada M (2003): **Caprazamycin B, a novel anti-tuberculosis antibiotic, from Streptomyces sp.** *J. Antibiot.* **56**: 580-583.

206. Ubukata M, Isono K, Kimura K, Nelson CC, McCloskey JA (1988): **The structure of liposidomycin B, an inhibitor of bacterial peptidoglycan synthesis**. *J. Am. Chem. Soc.* **110**: 4416-4417.

207. Igarashi M, Takahashi Y, Shitara T, Nakamura H, Naganawa H, Miyake T, Akamatsu Y (2005): **Caprazamycins, Novel Lipo-nucleoside Antibiotics, from Streptomyces sp**. *J. Antibiot.* **58**: 327-337.

208. Kaysser L, Siebenberg S, Kammerer B, Gust B (2010): **Analysis of the Liposidomycin Gene Cluster Leads to the Identification of New Caprazamycin Derivatives**. *ChemBioChem* **11**: 191-196.

209. Kaysser L, Lutsch L, Siebenberg S, Wemakor E, Kammerer B, Gust B (2009): **Identification and Manipulation of the Caprazamycin Gene Cluster Lead to New Simplified Liponucleoside Antibiotics and Give Insights into the Biosynthetic Pathway**. *J. Biol. Chem.* **284**: 14987-14996.

210. Masanori F, Satoshi B, Koichi N, Masahiko H, Yoko F, Tomoyuki S, Steven GVL (2010): **The Biosynthesis of Liposidomycin-like A-90289 Antibiotics Featuring a New Type of Sulfotransferase**. *ChemBioChem* **11**: 184-190.

211. Tsvetanova BC, Kiemle DJ, Price NPJ (2002): **Biosynthesis of tunicamycin and metabolic origin of the 11-carbon dialdose sugar, tunicamine**. *J. Biol. Chem.* **277**: 35289-35296.

212. Kagami S, Esumi Y, Nakakoshi M, Yoshihama M, Kimura K (2003): **Control of liposidomycin production through precursor-directed biosynthesis**. *J. Antibiot.* **56**: 552-556.

213. Dähn U, Hagenmaier H, Höhne H, König WA, Wolf G, Zähner H (1976): **Stoffwechselprodukte von Mikroorganismen**. *Arch. Microbiol.* **107**: 143-160.

214. Isono K, Nagatsu J, Kawashim.Y, Suzuki S (1965): **Studies on Polyoxins Antifungal Antibiotics. I. Isolation and Characterization of Polyoxins A and B**. *Agr. Biol. Chem.* **29**: 848-854.

215. Min-Kyoung K, Hyun-Sook P, Chi-Hwa K, Hee-Moon P, Wonja C (2002): **Inhibitory effect of nikkomycin Z on chitin synthases in *Candida albicans***. *Yeast* **19**: 341-349.

216. Cabib E (1991): **Differential Inhibition of Chitin Synthetases 1 and 2 from *Saccharomyces cerevisiae* by Polyoxin D and Nikkomycins**. *Antimicrob. Agents Chemother.* **35**: 170-173.

217. Isono K, Asahi K, Suzuki S (1969): **Polyoxins, antifungal antibiotics. XIII. Structure of polyoxins**. *J. Am. Chem. Soc.* **91**: 7490-7505.

218. Chen W, Huang T, He X, Meng Q, You D, Bai L, Li J, Wu M, Li R, Xie Z, et al. (2009): **Characterization of the Polyoxin Biosynthetic Gene Cluster from *Streptomyces cacaoi* and Engineered Production of Polyoxin H**. *J. Biol. Chem.* **284**: 10627-10638.

219. Bormann C, Mohrle V, Bruntner C (1996): **Cloning and heterologous expression of the entire set of structural genes for nikkomycin synthesis from *Streptomyces tendae* Tu901 in *Streptomyces lividans***. *J. Bacteriol.* **178**: 1216-1218.

220. Chen W, Zeng H, Tan H (2000): **Cloning, Sequencing, and Function of *sanF* : A Gene Involved in Nikkomycin Biosynthesis of *Streptomyces ansochromogenes***. *Current Microbiology* **41**: 312-316.

221. Lacalle RA, Tercero JA, Jimenez A (1992): **Cloning of the Complete Biosynthetic Gene-Cluster for an Aminonucleoside Antibiotic, Puromycin, and Its Regulated Expression in Heterologous Hosts**. *Embo Journal* **11**: 785-792.

222. Irene S, Eloísa S, Miguel Ángel R, Juan Carlos E, Antonio J (2002): **Identification of a set of genes involved in the biosynthesis of the aminonucleoside moiety of antibiotic A201A from *Streptomyces capreolus***. *Eur. J. Biochem.* **269**: 5527-5535.

223. Rubio MA, Barrado P, Espinosa JC, Jimenez A, Lobato MF (2004): **The *pur6* gene of the puromycin biosynthetic gene cluster from *Streptomyces alboniger* encodes a tyrosinyl-aminonucleoside synthetase**. *FEBS Lett.* **577**: 371-375.

224. Espinosa JC, Tercero JA, Rubio MA, Jimenez A (1999): **The *pur7* gene from the puromycin biosynthetic *pur* cluster of *Streptomyces alboniger* encodes a nudix hydrolase**. *J. Bacteriol.* **181**: 4914-4918.

225. Tercero JA, Lacalle RA, Jimenez A (1993): **The *pur8* Gene from the *pur* Cluster of *Streptomyces alboniger* Encodes a Highly Hydrophobic Polypeptide Which Confers Resistance to Puromycin**. *Eur. J. Biochem.* **218**: 963-971.

226. Rubio MA, Espinosa JC, Tercero JA, Jimenez A (1998): **The Pur10 protein encoded in the gene cluster for puromycin biosynthesis of *Streptomyces alboniger* is an NAD-dependent ATP dehydrogenase**. *FEBS Lett.* **437**: 197-200.

227. Vara J, Perezgonzalez JA, Jimenez A (1985): **Biosynthesis of Puromycin by *Streptomyces alboniger* - Characterization of Puromycin *N*-Acetyltransferase**. *Biochemistry* **24**: 8074-8081.

228. Rao MM, Rebello PF, Pogell BM (1969): **Biosynthesis of Puromycin in *Streptomyces alboniger* - Enzymatic Methylation of *O*-Demethylpuromycin**. *J. Biol. Chem.* **244**: 112-&.

229. Lacalle RA, Tercero JA, Vara J, Jimenez A (1993): **Identification of the Gene Encoding an *N*-Acetylpuromycin *N*-Acetylhydrolase in the Puromycin Biosynthetic Gene Cluster from *Streptomyces alboniger***. *J. Bacteriol.* **175**: 7474-7478.

230. Tercero JA, Espinosa JC, Lacalle RA, Jimenez A (1996): **The biosynthetic pathway of the aminonucleoside antibiotic puromycin, as deduced from the molecular analysis of the pur cluster of *Streptomyces alboniger***. *J. Biol. Chem.* **271**: 1579-1590.

231. Yamaguchi H, Sato S, Yoshida S, Takada K, Itoh M, Seto H, Otake N (1986): **Capuramycin, a New Nucleoside Antibiotic - Taxonomy, Fermentation, Isolation and Characterization**. *J. Antibiot.* **39**: 1047-1053.

232. Muramatsu Y, Muramatsu A, Ohnuki T, Ishii MM, Kizuka M, Enokita R, Tsutsumi S, Arai M, Ogawa Y, Suzuki T, et al. (2003): **Studies on novel bacterial translocase I inhibitors, A-500359s I. Taxonomy, fermentation, isolation, physico-chemical properties and structure elucidation of A-500359 A, C, D and G**. *J. Antibiot.* **56**: 243-252.

233. Funabashi M, Nonaka K, Yada C, Hosobuchi M, Masuda N, Shibata T, Van Lanen SG (2009): **Identification of the biosynthetic gene cluster of A-500359s in *Streptomyces griseus* SANK60196**. *J. Antibiot.* **62**: 325-332.

234. Lam WH, Rychli K, Bugg TDH (2008): **Identification of a novel beta-replacement reaction in the biosynthesis of 2,3-diaminobutyric acid in peptidylnucleoside mureidomycin A**. *Org. Biomol. Chem.* **6**: 1912-1917.

235. Gruschow S, Rackham EJ, Elkins B, Newilll PLA, Hill LM, Goss RJM (2009): **New Pacidamycin Antibiotics Through Precursor-Directed Biosynthesis**. *ChemBioChem* **10**: 355-360.

236. Fernandez-Moreno MA, Vallin C, Malpartida F (1997): **Streptothricin biosynthesis is catalyzed by enzymes related to nonribosomal peptide bond formation**. *J. Bacteriol.* **179**: 6929-6936.

237. Jackson MD, Gould SJ, Zabriskie TM (2002): **Studies on the formation and incorporation of streptolidine in the biosynthesis of the peptidyl nucleoside antibiotic streptothricin F**. *J. Org. Chem.* **67**: 2934-2941.

238. McCarty RM, Bandarian V (2008): **Deciphering deazapurine biosynthesis: Pathway for pyrrolopyrimidine nucleosides toyocamycin and sangivamycin**. *Chem. Biol.* **15**: 790-798.

239. Reader JS, Metzgar D, Schimmel P, de Crecy-Lagard V (2004): **Identification of four genes necessary for biosynthesis of the modified nucleoside queuosine**. *J. Biol. Chem.* **279**: 6280-6285.

240. Cone MC, Yin X, Grochowski LL, Parker MR, Zabriskie TM (2003): **The Blasticidin S Biosynthesis Gene Cluster from *Streptomyces griseochromogenes*: Sequence Analysis, Organization, and Initial Characterization**. *ChemBioChem* **4**: 821-828.

241. Cone MC, Petrich AK, Gould SJ, Zabriskie TM (1998): **Cloning and Heterologous Expression of Blasticidin S Biosynthetic Genes from *Streptomyces griseochromogenes***. *J. Antibiot.* **51**: 570-578.

242. Zhang Q, Cone MC, Gould SJ, Zabriskie TM (2000): **Reevaluation of the Final Steps in the Biosynthesis of Blasticidin S by *Streptomyces griseochromogenes* and Identification of a Novel Self-Resistance Mechanism**. *Tetrahedron* **56**: 693-701.

243. Gould SJ, Zhang QB (1995): **Cytosinine - Pyridoxal-Phosphate Tautomerase, a New Enzyme in the Blasticidin-S Biosynthetic-Pathway**. *J. Antibiot.* **48**: 652-656.

244. Gould SJ, Guo J (1994): **Cytosylglucuronic acid synthase (cytosine: UDP-glucuronosyltransferase) from *Streptomyces griseochromogenes*, the first prokaryotic UDP-glucuronosyltransferase**. *J. Bacteriol.* **176**: 1282-1286.

245. Laura LG, Zabriskie TM (2006): **Characterization of BlsM, a Nucleotide Hydrolase Involved in Cytosine Production for the Biosynthesis of Blasticidin S**. *ChemBioChem* **7**: 957-964.

246. Li L, Xu Z, Xu X, Wu J, Zhang Y, He X, Zabriskie TM, Deng Z (2008): **The Mildiomycin Biosynthesis: Initial Steps for Sequential Generation of 5-Hydroxymethylcytidine 5'-Monophosphate and 5-Hydroxymethylcytosine in *Streptoverticillium rimofaciens* ZJU5119**. *ChemBioChem* **9**: 1286-1294.

247. Rackham EJ, Gruschow S, Ragab AE, Dickens S, Goss RJM (2010): **Pacidamycin Biosynthesis: Identification and Heterologous Expression of the First Uridyl Peptide Antibiotic Gene Cluster**. *ChemBioChem* **11**: 1700-1709.

248. Price NPJ, Tsvetanova B (2007): **Biosynthesis of the tunicamycins: A review**. *J. Antibiot.* **60**: 485-491.

249. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2009): **GenBank**. *Nucl. Acids Res.* **37**: D26-D31.

250. Genomic sequence data for *S.clavuligerus* ATCC 27064 is found under GenBank accession number: ABJH00000000.

251. Harada S, Yoda K, Mori M, Tanimoto A, Furusato T, Yamane K, Takatsuki A, Yamasaki M, Tamura G (1988): **The Gene Responsible for Tunicamycin Resistance, *tmrb*, in *Bacillus subtilis***. *Agr. Biol. Chem.* **52**: 1785-1789.

252. Kapp U, Macedo S, Hall DR, Leiros I, McSweeney SM, Mitchell E (2008): **Structure of Deinococcus radiodurans tunicamycin-resistance protein (TmrD), a phosphotransferase**. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **64**: 479-486.

253. Beale JM, Foster JL (1996): **Carbohydrate fluxes into alginate biosynthesis in *Azotobacter vinelandii* NCIB 8789: NMR investigations of the triose pools**. *Biochemistry* **35**: 4492-4501.

254. Follstad BD, Stephanopoulos G (1998): **Effect of reversible reactions on isotope label redistribution - Analysis of the pentose phosphate pathway**. *Eur. J. Biochem.* **252**: 360-371.

255. Portais JC, Tavernier P, Gosselin I, Barbotin JN (1999): **Cyclic organization of the carbohydrate metabolism in *Sinorhizobium meliloti***. *Eur. J. Biochem.* **265**: 473-480.

256. Isono K, Sato T, Hirasawa K, Funayama S, Suzuki S (1978): **Biosynthesis of Polyoxins, Nucleoside Peptide Antibiotics - Biosynthesis of Nucleoside Skeleton of Polyoxins**. *J. Am. Chem. Soc.* **100**: 3937-3939.

257. Katz J, Rognstad R (1967): **Labeling of Pentose Phosphate from Glucose-^{14}C and Estimation of Rates of Transaldolase Transketolase Contribution of Pentose Cycle and Ribose Phosphate Synthesis**. *Biochemistry* **6**: 2227-2247.

258. Lee WNP, Boros LG, Puigjaner J, Bassilian S, Lim S, Cascante M (1998): **Mass isotopomer study of the nonoxidative pathways of the pentose cycle with [1,2-$^{13}C_2$]glucose**. *Am. J. Physiol. Endocrinol. Metabol.* **37**: E843-E851.

259. Sheflyan GY, Sundaram AK, Taylor WP, Woodard RW (2000): **Substrate ambiguity of 3-deoxy-D-manno-octulosonate 8-phosphate synthase from *Neisseria gonorrhoeae* revisited**. *J. Bacteriol.* **182**: 5005-5008.

260. Trefzer A, Salas JA, Bechthold A (1999): **Genes and enzymes involved in deoxysugar biosynthesis in bacteria**. *Nat. Prod. Rep.* **16**: 283-299.

261. Creuzenet C, Schur MJ, Lin JJ, Wakarchuk WW, Lam JS (2000): **FlaA1, a new bifunctional UDP-GlcNAc C-6 dehydratase C-4 reductase from *Helicobacter pylori***. *J. Biol. Chem.* **275**: 34873-34880.

262. Gross JW, Hegeman AD, Vestling MM, Frey PA (2000): **Characterization of enzymatic processes by rapid mix-quench mass spectrometry: The case of dTDP-glucose 4,6-dehydratase**. *Biochemistry* **39**: 13633-13640.

263. Hegeman AD, Gross JW, Frey PA (2002): **Concerted and stepwise dehydration mechanisms observed in wild-type and mutated *Escherichia coli* dTDP-glucose 4,6-dehydratase**. *Biochemistry* **41**: 2797-2804.

264. Nomura S, Yamane K, Sasaki T, Yamasaki M, Tamura G, Maruo B (1978): **Tunicamycin-Resistant Mutants and Chromosomal Locations of Mutational Sites in *Bacillus subtilis***. *J. Bacteriol.* **136**: 818-821.

265. Noda Y, Urakawa I, Yoda K, Yamasaki M (1996): ***Bacillus subtilis* TmrB protein confers resistance to tunicamycin on *Escherichia coli***. *J. Gen. Appl. Microbiol.* **42**: 343-347.

266. Noda Y, Takatsuki A, Yoda K, Yamasaki M (1995): **Tmrb Protein, Which Confers Resistance to Tunicamycin on *Bacillus subtilis*, Binds Tunicamycin**. *Biosci., Biotechnol., Biochem.* **59**: 321-322.

267. Noda Y, Yoda K, Takatsuki A, Yamasaki M (1992): **Tmrb Protein, Responsible for Tunicamycin Resistance of *Bacillus subtilis*, Is a Novel ATP-Binding Membrane-Protein**. *J. Bacteriol.* **174**: 4302-4307.

268. Walker JE, Saraste M, Runswick MJ, Gay NJ (1982): **Distantly Related Sequences in the Alpha-Subunits and Beta-Subunits of Atp Synthase, Myosin, Kinases and Other Atp-Requiring Enzymes and a Common Nucleotide Binding Fold**. *Embo Journal* **1**: 945-951.

269. Izard T (2001): **Structural basis for chloramphenicol tolerance in *Streptomyces venezuelae* by chloramphenicol phosphotransferase activity**. *Protein Sci.* **10**: 1508-1513.

270. Izard T, Ellis J (2000): **The crystal structures of chloramphenicol phosphotransferase reveal a novel inactivation mechanism**. *Embo Journal* **19**: 2690-2700.

271. Zachara NE, Hart GW (2002): **The emerging significance of *O*-GlcNAc in cellular regulation**. *Chem. Rev.* **102**: 431-438.

272. Wells L, Hart GW (2003): ***O*-GlcNAc turns twenty: functional implications for post-translational modification of nuclear and cytosolic proteins with a sugar**. *FEBS Lett.* **546**: 154-158.

273. Lefebvre T, Dehennaut V, Guinez C, Olivier S, Drougat L, Mir AM, Mortuaire M, Vercoutter-Edouart AS, Michalski JC (2010): **Dysregulation of the nutrient/stress sensor *O*-GlcNAcylation is involved in the etiology of cardiovascular disorders, type-2 diabetes and Alzheimer's disease**. *Biochim. Biophys. Acta, Gen. Subj.* **1800**: 67-79.

274. Macauley MS, Vocadlo DJ (2010): **Increasing *O*-GlcNAc levels: An overview of small-molecule inhibitors of *O*-GlcNAcase**. *Biochim. Biophys. Acta, Gen. Subj.* **1800**: 107-121.

275. Gross BJ, Kraybill BC, Walker S (2005): **Discovery of O-GlcNAc transferase inhibitors**. *J. Am. Chem. Soc.* **127**: 14588-14589.

276. Moran N (1996): **WHO issues another gloomy tuberculosis report**. *Nat. Med.* **2**: 377-377.

277. Umesiri FE, Sanki AK, Boucau J, Ronning DR, Sucheck SJ (2010): **Recent Advances Toward the Inhibition of mAG and LAM Synthesis in *Mycobacterium tuberculosis***. *Med. Res. Rev.* **30**: 290-326.

278. Alderwick LJ, Birch HL, Mishra A, Eggeling L, Besra GS (2007): **Structure, function and biosynthesis of the *Mycobacterium tuberculosis* cell wall: arabinogalactan and lipoarabinomannan assembly with a view to discovering new drug targets**. *Biochem. Soc. Trans.* **35**: 1325-1328.

279. Mills JA, Motichka K, Jucker M, Wu HP, Uhlik BC, Stern RJ, Scherman MS, Vissa VD, Pan F, Kundu M, et al. (2004): **Inactivation of the mycobacterial rhamnosyltransferase, which is needed for the formation of the arabinogalactan-peptidoglycan linker, leads to irreversible loss of viability**. *J. Biol. Chem.* **279**: 43540-43546.

280. Grzegorzewicz AE, Ma YF, Jones V, Crick D, Liav A, McNeil MR (2008): **Development of a microtitre plate-based assay for lipid-linked glycosyltransferase products using the mycobacterial cell wall rhamnosyltransferase WbbL**. *Microbiology* **154**: 3724-3730.

281. Maeda Y, Kinoshita T (2008): **Dolichol-phosphate mannose synthase: Structure, function and regulation**. *Biochim. Biophys. Acta, Gen. Subj.* **1780**: 861-868.

Chapter 2: Cloning the tunicamycin biosynthetic gene cluster

2.1 Introduction

As discussed in section 1.5, there are many benefits to dissecting the biosynthesis of tunicamycin on a genetic level. Crucially, doing so would unlock an enzymatic toolbox for the long-term goal of producing targeted, functional analogues of this important natural product. In this chapter we will discuss our approach to locating the genes responsible for tunicamycin production in *Streptomyces chartreusis*, their subsequent molecular cloning and their expression in a heterologous host. Additionally, comprehensive bioinformatic analysis of constituent genes in this biosynthetic pathway will be presented, culminating in the proposal of a detailed metabolic route to tunicamycin production from common cellular precursors.

2.2 Identifying gene cluster encoding tunicamycin biosynthesis

2.2.1 Strategy

Some of the typical biosynthetic mechanisms observed in the formation of a wide range of natural products were outlined in section 1.5. Although great diversity is seen in the compounds created, many conserved features and patterns exist. Polyketides and non-ribosomal peptides are particularly common and their biosynthetic pathways have received a great deal of attention.[1,2] The gene clusters in bacteria encoding these natural products have often been identified through PCR amplification of highly conserved signature genes using degenerate primers, followed by screening of genomic libraries for the presence of these sequences.[282] The highly conserved genetic motifs include characteristic acyl carrier protein, ketoreductase, dehydratase and enoylreductase modules in polyketide synthase (PKS) systems and adenylation, thiolation and condensation domains and peptide carrier proteins in non-ribosomal peptide synthetase (NRPS) systems.[283] More specific genetic motifs can also be targeted conserved between closely related natural product families or structural subunits.[282,284,285] In the wake of recent advances in high-throughput DNA sequencing, there has been a great increase in the number of whole genome projects, where entire genomes are fully assembled and carefully annotated through detailed bioinformatic

analysis. A consequence of such projects has been the discovery of new biosynthetic gene clusters, not only of existing bioactive compounds but also of previously unobserved secondary metabolites encoded by so-called 'cryptic' gene clusters.[286,287] Such a predominantly *in silico* approach is relatively new and has been made possible by the well-organised and expansive genetic databases available to researchers, such as the International Nucleotide Sequence Database Collaboration.[288]

Unfortunately, tunicamycin does not fall into the major classes of natural products such as polyketides, non-ribosomal peptides or isoprenoids. It is unlikely to require a large number of genes for its production and few of its biosynthetic genes can be predicted with enough precision to confidently assign a particular genetic homologue as a highly conserved probe sequence for degenerate primer design. For this reason, we elected to rapidly sequence the entire bacterial genome of a known tunicamycin producer and perform direct *in silico* genome scanning in order to find the gene cluster responsible for tunicamycin biosynthesis. Such an approach was expected not only to avoid the dangers of false positive results derived from speculative PCR-amplified probe sequences, but also to expediate the discovery process through use of efficient and rapid bioinformatic tools rather than time-consuming experimental techniques. What follows is an account of our high-throughput sequencing of the *S. chartreusis* genome and the subsequent *in silico* genome mining based on chemical logic, which led to the identification of a putative tunicamycin biosynthetic gene cluster.

2.2.2 Isolation of genomic DNA

In order to sequence the genome of a tunicamycin-producing organism, a sample of high molecular weight (Mw) genomic DNA first needed to be isolated. The sample would additionally function as the source of DNA for constructing a genomic library of a tunicamycin-producing organism, so it needed to be of sufficient purity and at least 30 kb in size. The process of isolating such DNA was non-trivial since each molecule was extremely large – fragments over 200 kb are 70 µm long – and thus they were sensitive to DNAse degradation and mechanical shear from narrow-bore pipette tips and other forms of physical handling. Fortunately, a number of protocols have been developed by Kieser *et al*. for specific use with *Streptomycetes* and these were screened and optimised for use in our system.[289]

Streptomyces chartreusis NRRL 3882[290] and *Streptomyces lysosuperificus* ATCC 31396[291] have both been documented to produce tunicamycin. These strains were duly

sourced and processed in parallel. The two *Streptomyces* strains were grown in a variety of liquid media – including TSB/F and 1:1 TSB: superYEME 34 % sucrose – under standard conditions which maximise aeration and dispersed growth.[289] The choice of media and length of incubation were important factors in the protocol, influencing the thickness of and polysaccharides levels in the cell wall which in turn affected the efficiency of the DNA isolation process. As such, a number of mycelial preparations were subjected to the subsequent lysis and extraction protocols as it could not be predicted at this stage which conditions would lead to the best sample of high Mw DNA.

Cell lysis was typically achieved with lysozyme treatment, followed by the release of cell contents with SDS detergent and the digestion of intracellular proteins with proteinase K.[289] Low isolated yields of DNA further down the line led to the conclusion that lysozyme was not sufficient to fully digest the cell wall of these strains, even under prolonged incubation. As a result, mutanolysin was used in conjunction with the lysozyme. Acting on a different portion of the cell wall, the use of mutanolysin proved to be an essential modification to the protocol, leading to more complete cell wall lysis and release of cell contents. An additional modification was included at this stage for *S. lysosuperificus* preparations, after existing samples were found to predominantly contain DNA fragments below 30 kb. Treatment with RNA protect solution (Qiagen) inhibited cellular DNAse enzymes enough to maintain a high Mw distribution throughout the isolation process (see Fig. 2.1).

The next stage following cell lysis and protein digestion involved organic extraction of unwanted cell contents from the aqueous solution.[289] Unfortunately, the two routinely used extraction solvents, chloroform and 25:24:1 phenol/chloroform/isoamyl alcohol, both resulted in poor quality DNA isolated in low yields. It was noted that the incredibly viscous and sticky aqueous layer could be due to a high polysaccharide content of the cell wall, whose release upon lysis could adversely affect subsequent extraction by organic solvent. Solubilisation of these polysaccharides was facilitated by incubation with hexadecyltrimethylammonium bromide (CTAB), an amphiphilic molecule which bound to the polar polysaccharides and transferred them into the organic phase.[289] This modification was particularly beneficial with *S. lysosuperificus* samples, although its use with *S. chartreusis* preparations actually decreased the yield of isolated DNA.

Following implementation of cell lysis and solvent extraction protocols optimised for individual samples, the resulting aqueous layers were collected with wide bore pipette tips (to prevent DNA shearing). The genomic DNA from each sample was then precipitated

with isopropanol[289] and subjected to spectrophotometric nucleic acid quantitation *via* Nanodrop and size estimation by pulsed-field gel electrophoresis (Fig. 2.1).

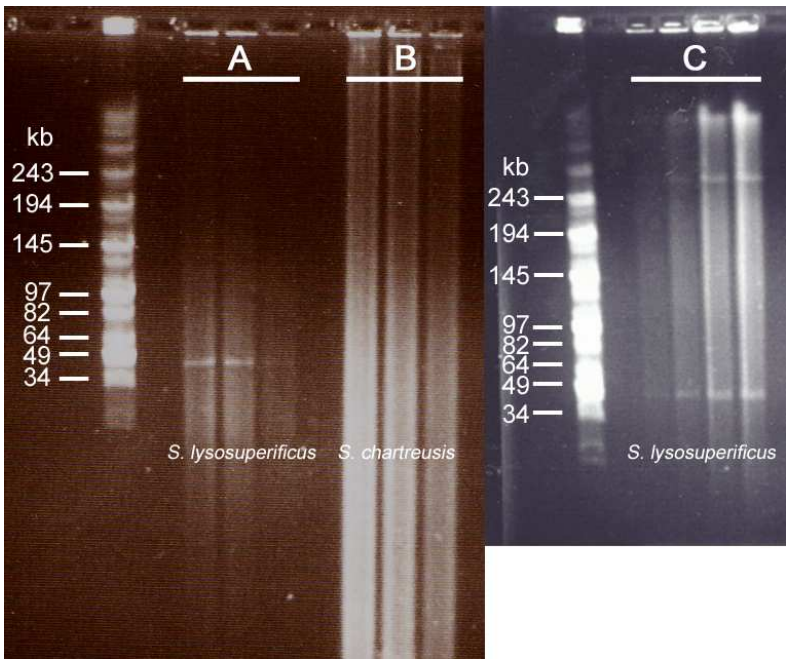


Fig. 2.1 Pulsed-field gel electrophoresis (PFG) of isolated genomic DNA; **A**: *S. lysosuperificus* genomic DNA isolated using no protocol modifications; **B**: *S. chartreusis* genomic DNA isolated using no protocol modifications; **C**: *S. lysosuperificus* genomic DNA isolated using an optimised protocol, incorporating mutanolysin, RNA Protect (Qiagen) and CTAB treatment.

The effect of modifications to the literature protocol can be clearly seen when comparing *S. lysosuperificus* samples in Fig. 2.1 (A and C). Sample C was obtained using the optimised protocol and the DNA clearly shows a markedly increased size distribution over that of sample A, which was obtained using the unmodified protocol. Spectrophotometric analysis of the isolated DNA (Nanodrop) showed an OD_{260}/OD_{280} ratio of 1.8-2.1 in all samples, indicating minimal protein contamination.

Notably, both *S. lysosuperificus* samples showed strong bands within the smear of genomic DNA (Fig. 2.1, A and C). This was likely due to multiple copy numbers of up to two non-chromosomal plasmids present as part of the genome, at approximately 40 and 260 kb; such plasmids are not uncommon amongst *Streptomyces* species. That said, the gel of *S. chartreusis* genomic DNA did not exhibit any unusual bands and does appear to harbour any non-chromosomal plasmids (Fig. 2.1, B). This is a significant observation, since the presence of plasmids in high copy number risks distorting any subsequent cosmid library with multiple copies of identical plasmid sequences. Additionally, raw genome sequencing data would be biased towards plasmid-derived sequences, reducing overall sequencing

coverage. For this reason, we elected to search for the tunicamycin biosynthetic cluster solely in the *S. chartreusis* NRRL 3882 strain, using its genomic DNA for subsequent whole genome sequencing and genomic library construction experiments.

2.2.3 Whole genome sequencing of S. chartreusis

A fresh sample of high Mw genomic DNA from *S. chartreusis* NRRL 3882 was prepared and sequenced by the University of Liverpool Advanced Genomics Facility with a ½ plate run using the Roche 454 Titanium pyrosequencing platform. The data was supplied as a draft assembly constructed using Roche Newbler assembler software; the associated metrics data is summarised in Table 2.1.

***Streptomyces chartreusis* NRRL 3882**	
From 286 Mb of 454 sequence data (½ plate run):	
largeContigMetrics	
numberOfContigs	= 2,638
numberOfBases	= 7,773,698 bp
avgContigSize	= 2,946 bp
N50ContigSize	= 4,582 bp
largestContigSize	= 53,916 bp
Q40PlusBases	= 7,736,953 bp, 99.53 %
Q39MinusBases	= 36,745 bp, 0.47 %
allContigMetrics	
numberOfContigs	= 3,112
numberOfBases	= 7,952,445 bp

Table 2.1 Metrics data from S. chartreusis genome sequencing

A total of 286 Mb data was collected, giving an genome coverage of 36.0× based on large contig data. The genome was assembled into a total of 3,112 contigs,[292] 2,638 of which were above 500 bases and considered large. The average contig size was 2,946 bases and the largest contig was 53,916 bases. The N50ContigSize and Q40PlusBases/Q39MinusBases parameters represent an alternative average contig size calculation and a measure of sequence quality respectively.[293,294] The contigs covered 7.95 Mb of the *S. chartreusis* chromosome with a G+C content of 70 %, both consistent with data for existing *Streptomyces* genomes (Table 2.2).

Streptomyces strain[a]	State of assembly	Chromosome size (Mb)	G+C content (%)
S. chartreusis NRRL 3882[b]	Single Pass	7.95	70
S. coelicolor A3(2)[c]	Complete	8.67	72
S. griseus subsp. *griseus* NBRC 13350[c]	Complete	8.55	72
S. avermitilis MA-4680[c]	Complete	9.03	70
S. scabies[c]	Complete	10.15	72
S. viridochromogenes DSM 40736	Draft	8.55	71
S. sviceus ATCC 29083	Draft	9.06	70
Streptomyces sp. e14	Draft	7.15	68
Streptomyces sp. SPB78	Draft	6.90	72
Streptomyces sp. SPB74	Draft	6.51	70
Streptomyces sp. Mg1	Draft	7.11	70
Streptomyces sp. C	Draft	7.92	72
Streptomyces sp. ACTE	Draft	7.37	71
Streptomyces sp. ACT-1	Draft	8.57	72
Streptomyces sp. AA4	Draft	9.15	69
S. roseosporus NRRL 15998	Draft	7.56	70
S. roseosporus NRRL 11379	Draft	7.76	71
S. pristinaespiralis ATCC 25486	Draft	7.63	71
S. lividans TK24	Draft	8.19	71
S. hygroscopicus ATCC 53653	Draft	10.47	71
S. griseoflavus Tu4000	Draft	7.36	71
S. ghanaensis ATCC 14672	Draft	8.22	70
S. flavogriseus ATCC 33331	Draft	7.59	71
S. clavuligerus ATCC 27064	Draft	8.53	72
S. albus J1074	Draft	6.62	72

[a] Data obtained from NCBI Genome database (http://www.ncbi.nlm.nih.gov/sites/genome) in May 2010; [b] Data from our *S. chartreusis* sequencing project; [c] Data from the Streptomyces Annotation Server (http://strepdb.streptomyces.org.uk/);

Table 2.2 Chromosome sizes and G+C contents from existing Streptomyces genome projects

It should be noted that due to the high G+C content of *Streptomyces* genomes, larger contigs show a bias towards a lower G+C content (Fig. 2.2). This is a consequence of the assembly algorithm, which is most efficient for G+C levels close to 50 %. For this reason, the actual G+C content for *S. chartreusis* may be slightly higher than the 70 % observed across all combined contigs.

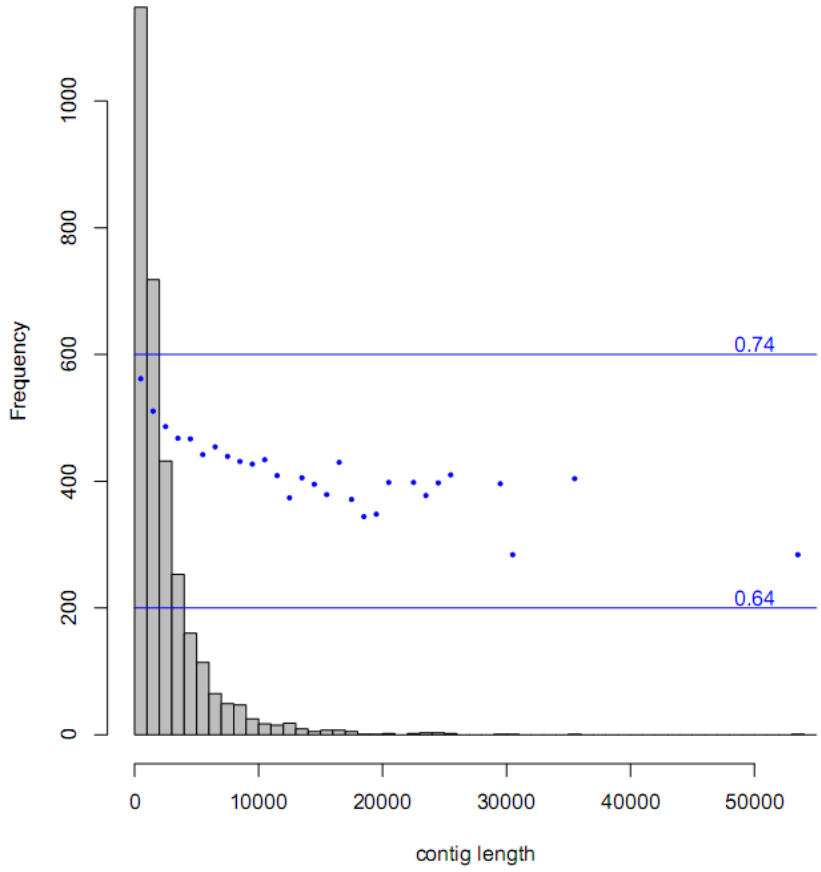


Fig. 2.2 Chart showing the distribution of contigs according to their length (grey bars) and the relationship between contig length and G+C content (blue dots)

2.2.4 Genome mining of an S. chartreusis BLASTable database

2.2.4.1 Query sequences used as in silico probes

Using no more than a draft assembly of the *S. chartreusis* genome, DNA sequences contained within the multiple contig fragments were screened for the presence of putative tunicamycin biosynthetic genes. A wide range of candidate genes from other organisms were selected as query sequences and screened against an *S. chartreusis* BLASTable database, kindly constructed by Dr. Govind Chandra (JIC). Typically, tBLASTn searches were performed, whereby protein sequences of candidate genes were compared to the six-frame translations (both strands) of the nucleotide database of *S. chartreusis* contigs.[295] The *in silico* probes used were chosen based on analysis of the chemical reactivity proposed to feature in the construction of tunicamycin from metabolic precursors and also on enzymes in the putative biosynthetic pathway of tunicamycin proposed by Tsvetanova *et al.* (see section 1.4.3).[296] This generated a series of 'logic filters' for the genome scanning process, iteratively homing in on candidate contigs containing genes of interest. Moreover, since genes for secondary metabolite biosynthesis in bacteria are predominantly found in clusters,[286] multiple hits against different probes found on a single contig would identify potential locations of the tunicamycin biosynthetic gene set.

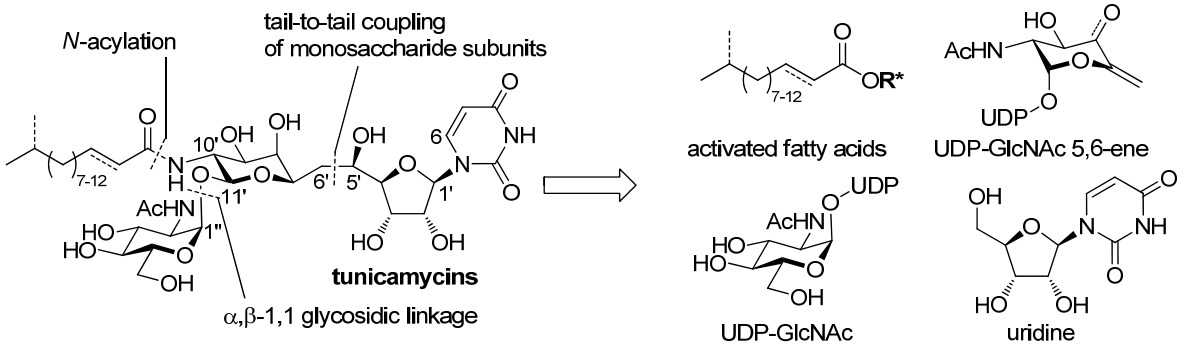


Fig. 2.3 Proposed disconnections for tunicamycin biosynthesis

Some of the more notable query sequences used in this genome mining process and the logic behind their selection are presented below.

2.2.4.1.1 1,1-Linking glycosyltransferases

Tunicamycin contains a rare α,β-1,1 linkage between tunicamine and a terminal GlcNAc residue (Fig. 2.3). Such non-reducing glycosidic linkages are typified by trehalose, a naturally occurring disaccharide consisting of an 1,1-glycosidic bond between two α-glucose units. There are three distinct pathways for trehalose biosynthesis conserved across trehalose-producing prokaryotes – the OtsA-OtsB, TreY-TreZ and TreS pathways.[297] 3,3´-Neotrehalosediamine from *Bacillus subtilis* contains an α,β-trehalose linkage between two aminosugars, a motif shared with tunicamycin.[298] The biosynthetic cluster of trehalase inhibitor salbostatin from *Streptomyces albus* contains homologues of trehalose forming enzymes although the antibiotic does not itself contain a trehalose linkage.[299] Sequences of glycosyltransferase enzymes (GTs) in all of these pathways and from a wide range of sources were used as probes against our BLAST database to identify potential GTs of interest in *S. chartreusis*. It should be noted that most of the enzymes queried are involved in – or homologous to those involved in – the formation of α,α-1,1 glycosidic bonds. It is possible these enzymes operate *via* a different mechanism to the tunicamycin cluster glycosyltransferase, which forms an α,β-1,1 glycosidic linkage, and as such might belong to different GT families with unrelated primary sequences and protein folds.

2.2.4.1.2 Hexose epimerases and dehydratases

Tsvetanova *et al*. showed that the terminal GlcNAc and the pseudo-GalNAc moieties within tunicamycin are derived from the same metabolic pool of UDP-GlcNAc.[296] As such tunicamycin's biosynthetic cluster is likely to contain a GlcNAc 4-epimerase to form the GalNAc motif with its axial C–4 hydroxyl. In addition, the C–C linkage between C–6´ of the GalNAc moiety and C–5´ of the uridine portion in tunicamycin is proposed to be formed *via* a GlcNAc-5,6-ene intermediate (Fig. 2.3), whose formation would require a hexose 4,6-dehydratase. The two proposed enzymes are highly related, and the two functions could even be combined into one enzyme. As such, sequences from epimerases, dehydratases and epimerase/dehydratases from a wide range of metabolic pathways and acting on NDP-glucose and NDP-GlcNAc substrates were queried against the *S. chartreusis* BLAST database.

2.2.4.1.3 *N*-acetyl-hexosamine deacetylases

UDP-GlcNAc has been implicated as the common metabolic precursor to both aminosugar moieties within tunicamycin.[296] As such, it is highly likely that the fatty acid chain at C–10´ in tunicamycin is introduced by *N*-deacetylation at C–2 of UDP-GlcNAc – or the equivalent position in a more advanced intermediate – followed by *N*-acylation of the resulting free amine (Fig. 2.3). A number of enzymes exhibiting this function have been characterised – these include *N*-acetyl-hexosamine deacetylases involved in teicoplanin[300] and neomycin[301] biosynthesis, as well as GlcNAc-inositol deacetylase MshB from mycothiol biosynthesis.[302] Sequences from all of these enzymes were probed against the *S. chartreusis* BLAST database.

2.2.4.1.4 Fatty acid and polyketide synthase subunits

The fact that tunicamycin exists as a mixture of up to 18 homologues differing in the composition of their acyl chains suggests that its fatty acids are derived from the common cellular pool rather than biosynthesised *de novo*. However, it was hoped that the tunicamycin gene cluster would contain homologues of certain polyketide synthase or fatty acid synthase subunits for use in activating these cellular lipids (Fig. 2.3), in particular homologues of highly conserved acyl carrier proteins.[303,304] As such, a number of sequences from the actinorhodin cluster of *S. coelicolor* were queried against the *S. chartreusis* BLAST database.

2.2.4.2 Bioinformatic analysis of candidate contigs

Genome mining of the *S. chartreusis* BLASTable database yielded candidate contigs possessing open reading frames (ORFs) homologous to some of the query sequences discussed above. In order to home in on a putative tunicamycin gene cluster, these candidate contigs were analysed in greater detail using Artemis, a DNA sequence visualisation and annotation tool (a screenshot can be seen in Fig. 2.4).[305] Artemis displays any DNA sequence along with its amino acid translation in all six frames. Many features can then be applied, such as the visualisation of all possible stop codons, a GC frame plot and G+C content analysis. All of these aid the discovery of ORFs and the localisation of gene boundaries within a stretch of DNA sequence. Each of these ORFs can then be appropriately labelled and annotated following BLAST comparison with the compiled sequences of international genetic databases.

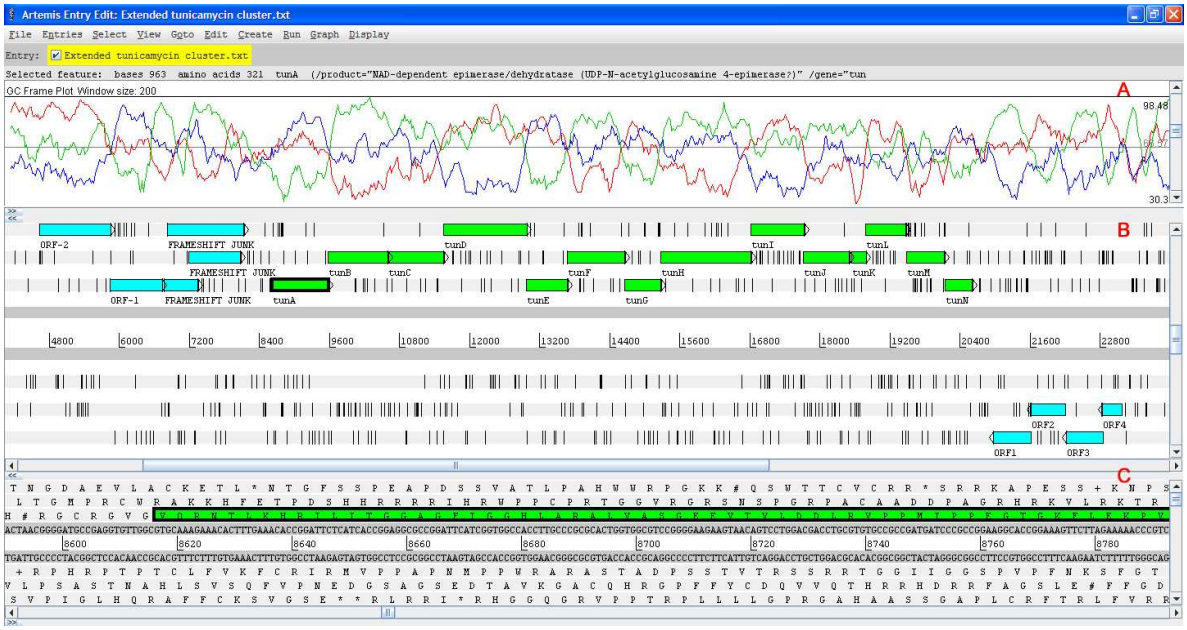


Fig. 2.4 Bioinformatic analysis in Artemis; **A**: GC frame plot and average G+C content (grey line); **B**: Overview of DNA sequence in 6 reading frames with stop codons (vertical lines), annotated genes (green) and other open reading frames (blue); **C**: Zoomed window with amino acid sequence translation in 6 reading frames.

Some of the strategies we used for identifying ORFs and determining the precise boundaries of key genes are summarised briefly below.

2.2.4.2.1 GC frame plot

In this algorithm, the G+C content of every third base pair in three frames is plotted as green, blue and red lines respectively. For example, the G+C content of the 1st, 4th, 7th, 10th, etc. base pairs will be plotted for the whole sequence with a green line. Similarly, the G+C content of the 2nd, 5th, 8th, 11th, etc. base pairs will be plotted as a blue line and the 3rd, 6th, 9th, 12th, etc. as a red line. The resulting graphic is termed a GC frame plot.[306,307] *Streptomyces* bacteria exhibit a high G+C content – typically greater than 70 %. Due to the degeneracy of the genetic code – particularly in position 3 of a codon – there is a highly biased codon usage in Streptomycetes, with 92 % of codons ending in G or C.[308] As such, coding regions of DNA can be identified from the patterns of the GC frame plot, as seen for the genes in Fig. 2.4. The reading frame of each coding region can also be determined in this way.[306,307]

2.2.4.2.2 Start codons

Initiation codons at the start of all genes code for *N*-formylmethionine or methionine, which are used to initiate protein synthesis.[309] The most common start (or initiation) codon is ATG, coding for methionine. In some cases however, GTG (valine) or in very rare cases TTG (leucine) can be used. Due to the GC-rich nature of Streptomyces genomes, GTG initiation codons are much more common, and care must be taken when assigning the start point of any gene.[289]

2.2.4.2.3 Stop codons

In messenger RNA (mRNA), certain nucleotide triplets signal the termination of translation. In the corresponding DNA templates these stop (or termination) codons are TAA, TGA and TAG.[310,311] In GC-rich organisms such as *Streptomyces*, TGA stop codons predominate.[289] By visualising possible locations of stop codons in all six reading frames within Artemis, regions of DNA showing a conspicuous lack of these codons often indicate the presence of an ORF.

2.2.4.2.4 Ribosome binding sites

When there are a number of potential initiation sites for a particular gene, identifying ribosome binding sites can assist in correctly assigning the true transcriptional start site. Ribosome binding sites of mRNAs usually occur 5-9 bp upstream of the translational start codon and show complementarity to sequences close to the 3´ end of 16S rRNA.[289,312] This Shine-Delgarno sequence is generally centred around the mRNA sequence GGAGG and complementarity to the 16S rRNA can extend 3-7 bp in either direction.[313] Although this is a good guide for determining true initiation codons, not all genes exhibit ribosome binding sites and instead rely on polycistronic transcription following on from a preceding gene.

2.2.5 Identification of a putative tunicamycin biosynthetic gene cluster

2.2.5.1 Bioinformatic analysis of candidate contigs

Following extensive genome mining, a single contig was identified containing multiple hits. Analysis of contig 00341 with Artemis and BLAST identified homologues of inosityl-GlcNAc deacetylase, hexose epimerase/dehydratase and acyl carrier protein, in a cluster of eleven open reading frames (ORFs). The 3´-end of this operon coincided with the edge of the contig and terminated with the partial sequence of a putative GT-1 family glycosyltransferase (termed *tunD*). The full length sequences of glycosyltransferases homologous to *tunD* were screened against the *S. chartreusis* genome, identifying contig 02796. This additional contig contained the remaining portion of *tunD* and three further clustered ORFs. The precise gene boundaries and functions of the 14 ORFs in this operon were determined and careful analysis concluded that these genes spanning two non-overlapping contigs are likely to constitute a tunicamycin biosynthetic gene cluster (*tun*). The constituent genes were consequently labelled *tunA*-*N*.

2.2.5.2 Genetic organisation of putative tun gene cluster

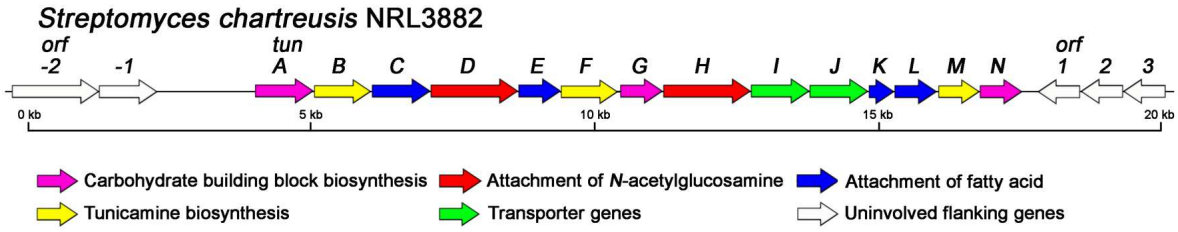


Fig. 2.5 Genetic organisation of putative *tun* gene cluster

The genetic organisation of the putative *tun* gene cluster is shown in Fig. 2.5 and the predicted functions of *tunA* through *tunN* along with those of flanking ORFs are presented in Table 2.3. More detailed analysis of individual biosynthetic enzymes and their proposed roles in the tunicamycin biosynthetic pathway are discussed in section 2.4. The contiguous 12.0 kb stretch contains a total of 14 ORFs, all of which are oriented in the same sense and with minimal gaps. This suggests the entire cluster is contained within a single polycistronic transcript, which is feasible given its small overall size. The overall G+C content of this region is 65.0 %, well below that of a typical *Streptomyces* genome or indeed the rest of the *S. chartreusis* genome (see Table 2.2). This suggests this gene cluster

may have been acquired from another lower G+C organism during the course of its evolution.

	Amino acids	Proposed biosynthetic function	Closest protein homologue[a], Origin	Identity/Similarity, Accession no.
orf-2	406		Integrase, *Streptomyces sviceus* ATCC 29083	52% / 65%, ZP_05020140
orf-1	297		Transposase, *Rhodococcus jostii* RHA1	50% / 64%, YP_700005
tunA	321	UDP-GlcNAc epimerase/dehydratase	NAD-dependent epimerase/dehydratase, *Streptomyces* sp. Mg1	42% / 58%, ZP_04996782
tunB	338	Fe-S oxidoreductase	Radical SAM domain protein, *Haloterrigena turkmenica* DSM5511	32% / 48%, YP_003405396
tunC	318	*N*-acyltransferase	GCN5-related N-acetyltransferase, *Fervidobacterium nodosum*	33% / 50%, YP_001410548
tunD	474	Glycosyltransferase	Group 1 family glycosyltransferase, *Thermococcus barophilus*	28% / 45%, ZP_04876510
tunE	234	*N*-deacetylase	N-acetylglucosaminyl-phosphatidylinositol de-N-acetylase, *Cylindrospermopsis raciborskii*	43% / 57%, ZP_06309433
tunF	327	UDP-GlcNAc-4-epimerase	UDP-glucose 4-epimerase, *Paenibacillus* sp. oral taxon 786	46% / 63%, ZP_04852226
tunG	203	UMP phosphatase	Phosphoglycerate mutase, *Frankia* sp. CcI3	29% / 47%, YP_481446
tunH	515	Nucleotide pyrophosphatase	Type I nucleotide pyrophosphatase, *Burkholderia* sp. 383	35% / 50%, YP_370731
tunI	304	ABC transporter ATP-binding subunit	Putative ABC transporter ATP-binding subunit, *Streptomyces scabiei* 87.22	41% / 61%, YP_003492364
tunJ	262	ABC transporter permease subunit	ABC-2 type transporter, *Thermobaculum terrenum*	32% / 51%, YP_003322218
tunK	81	Acyl carrier protein	Phosphopantetheine-binding protein *Catenulispora acidiphila* DSM44928	32% / 61%, YP_003117493
tunL	229	Phospholipid phosphatase	Phosphoesterase PA-phosphatase, *Micromonospora aurantiaca*	33% / 42%, ZP_06217896
tunM	216	Radical SAM protein	Methyltransferase family protein, *Saccharomonospora viridis*	48% / 63%, YP_003133112
tunN	152	UTP pyrophosphatase	NUDIX hydrolase, *Nakamurella multipartita*	36% / 55%, YP_003200035
orf1	213		Secreted protein, *Streptomyces viridochromogenes*	81% / 90%, ZP_05533938
orf2	573		Secreted protein, *Streptomyces viridochromogenes*	90% / 94%, ZP_05533937
orf3	606		Secreted protein, *Streptomyces viridochromogenes*	82% / 89%, ZP_05533936

[a] Homologues from *S. clavuligerus* ATCC 27064 and *A. mirum* DSM 43827 were omitted.

Table 2.3 Deduced functions of genes in the *tun* operon

The three flanking genes downstream of the putative *tun* cluster (*orf1-3*) have close homologues in almost every sequenced *Streptomyces* genome and as such are likely to encode conserved housekeeping genes rather than be involved in tunicamycin biosynthesis. The two upstream flanking genes (*orf-1* and *orf-2*) encode putative transposase and integrase enzymes respectively, which are not likely to be involved in tunicamycin biosynthesis but which lends support to the idea that *S. chartreusis* acquired the *tun* gene cluster from another organism. Between *orf-1* and *tunA* there is a 1900 bp putative ORF with multiple frameshifts - "junk" DNA.

2.2.5.3 PCR verification of contigs spanning the putative tun operon

PCR of *S. chartreusis* genomic DNA was performed with primers matching the 5´ and 3´ ends of *tunD* (primers TunDF and TunDR), in order to bridge the gap between contigs 00341 and 02796 (Fig. 2.6). Sequencing of this PCR product showed 100 % identity to a 762 bp area spanning the gap between the two contigs and confirmed that the sequence in contig 02796 was directly flanked by the sequence in contig 00341 on the *S. chartreusis* chromosome. This showed that there were no missing base pairs between the two contigs and as such the full sequence of *tunD* could be unambiguously reported.

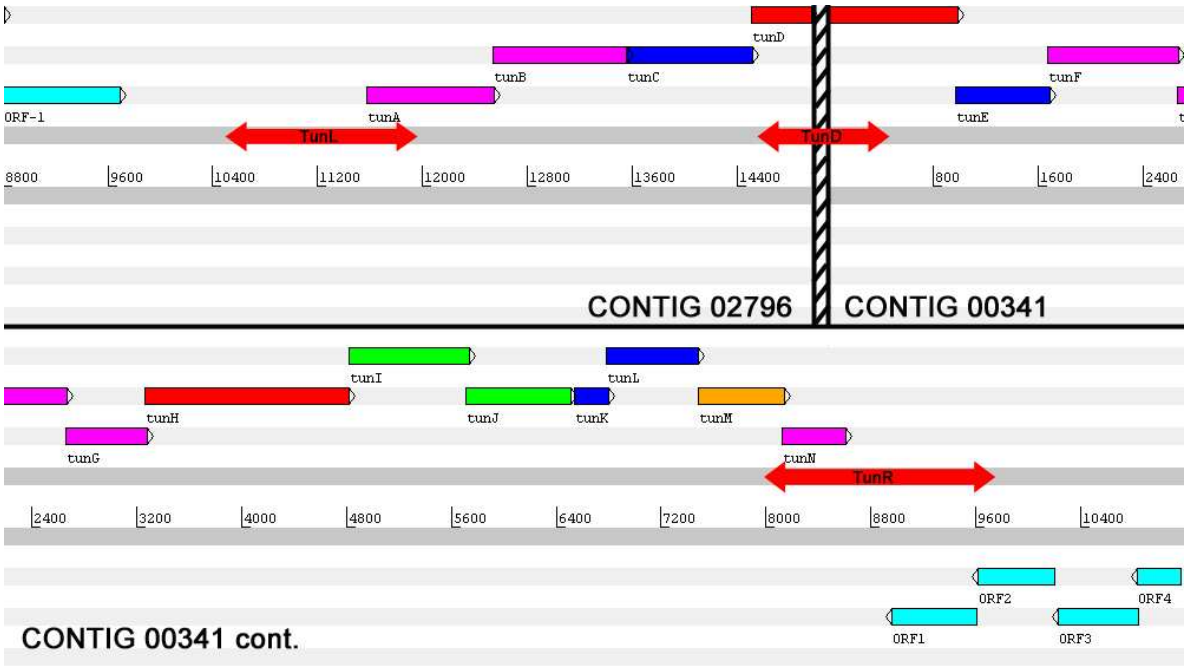


Fig. 2.6 PCR verification of *tun* gene cluster boundaries and of the *tunD* sequence

Two further PCR products were obtained against *S. chartreusis* genomic DNA, using pairs of primers designed to amplify regions of sequence bordering either side of the *tun*

gene cluster (primers TunLF/TunLR and TunRF/TunRR). The resulting TunL and TunR PCR products both showed 100 % identity to 762 and 2032 bp regions on the left and right boundaries of the *tun* cluster respectively (Fig. 2.6). This confirmed that no misassembly had occurred during genome sequencing and hence that the predicted boundaries of the *tun* cluster were indeed correct.

2.2.6 *Homologous* tun *gene clusters in S. clavuligerus and A. mirum*

Bioinformatic analysis of the *tun* cluster allowed identification of other potential tunicamycin producers. Homologous sets of genes were found in *Actinosynnema mirum* DSM 43827[314] and *Streptomyces clavuligerus* ATCC 27064; the latter has been reported to produce the closely related antibiotic MM19290.[315] Only minor differences were observed between the genetic organisation of these gene clusters, suggesting a recently shared evolutionary heritage (Fig. 2.7). The homologies of genes from both clusters were compared with the parent *tun* genes from *S. chartreusis* (Table 2.4).

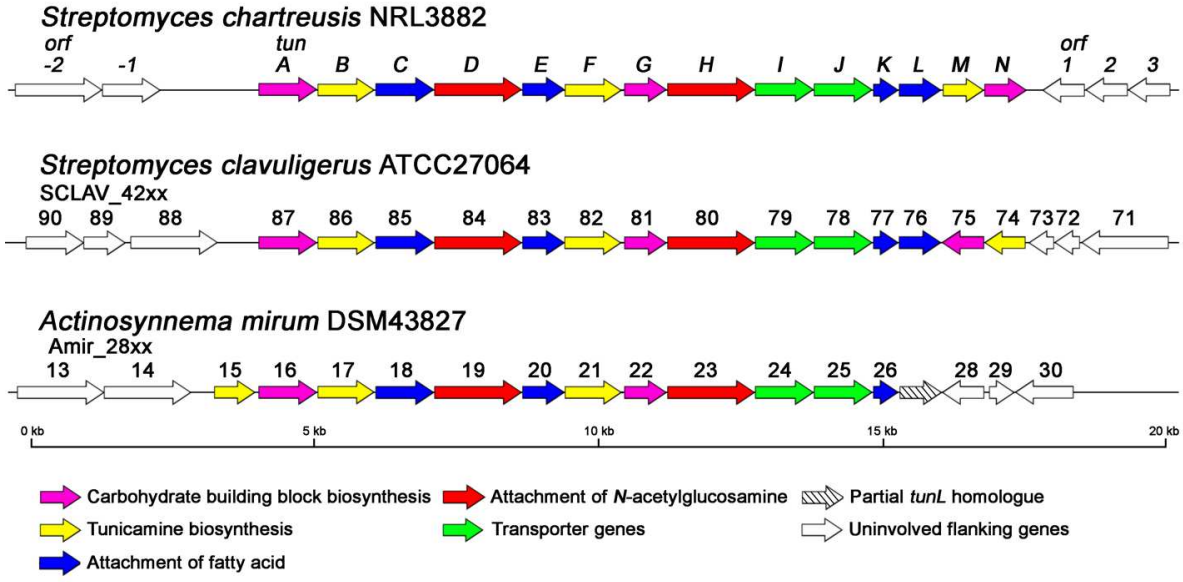


Fig. 2.7 Genetic organisation of *tun* gene cluster homologues from *S. clavuligerus* and *A. mirium*

The *S. clavuligerus* cluster exhibited greatest similarity to the genes from *S. chartreusis*, with identities ranging from 52 to 90 % and each *tun* gene accounted for. The genes are almost all present in the same order, the only difference being that *tunM* and *tunN* have reversed orientation at the downstream end of the cluster. It is highly likely that the genes with annotated loci SCLAV_4274 to SCLAV_4287 are responsible for MM19290 biosynthesis in *Streptomyces clavuligerus* ATCC27064. Additionally, although the structure of MM19290 has not been disclosed, the large degree of homology with *tun* genes

from *S. chartreusis* strongly suggests that – like the streptovirudins and corynetoxins – this compound shares its core carbohydrate skeleton with tunicamycin.

The *A. mirum* cluster exhibits identities with *tun* genes from *S. chartreusis* ranging from 30 to 78 %. The order of *tunA* to *tunK* homologues is conserved, although *tunM* has moved from the end to the beginning of the cluster. Interestingly no homologue of *tunN* or *tunL* was found, although upon closer inspection of this stretch of *A. mirum* sequence a truncated version *tunL* was discovered, containing a number of frameshift mutations. Since this organism has not been reported to produce any antibiotics structurally related to the tunicamycins, it is probable that we have uncovered a silent gene cluster which has lost the ability to produce its tunicamycin-like metabolite.

***S. chartreusis* NRRL 3882**		**Proposed biosynthetic function**	***S. clavuligerus* ATCC 27064 homologue locus, aa, (%Id/Si)[b], Accession no.**	***A. mirum* DSM 43827 homologue locus, aa, (%Id/Si), Accession no.**
ORF	**aa[a]**			
TunA	321	UDP-GlcNAc epimerase/dehydratase	SCLAV_4287, 276, (72/80), ZP_06773762	Amir_2816, 322, (54/65), YP_003100592
TunB	338	Fe-S oxidoreductase	SCLAV_4286, 338, (90/95), ZP_06773761	Amir_2817, 340, (78/87), YP_003100593
TunC	318	*N*-acyltransferase	SCLAV_4285, 322, (60/72), ZP_06773760	Amir_2818, 318, (43/57), YP_003100594
TunD	474	Glycosyltransferase	SCLAV_4284, 461, (63/77), ZP_06773759	Amir_2819, 451, (47/58), YP_003100595
TunE	234	*N*-deacetylase	SCLAV_4283, 236, (77/85), ZP_06773758	Amir_2820, 230, (63/76), YP_003100596
TunF	327	UDP-GlcNAc-4-epimerase	SCLAV_4282, 327, (76/83), ZP_06773757	Amir_2821, 332, (58/70), YP_003100597
TunG	203	UMP phosphatase	SCLAV_4281, 208, (65/73), ZP_06773756	Amir_2822, 223, (50/60), YP_003100598
TunH	515	Nucleotide pyrophosphatase	SCLAV_4280, 518, (66/76), ZP_06773755	Amir_2823, 510, (53/65), YP_003100599
TunI	304	ABC transporter ATP-binding subunit	SCLAV_4279, 302, (77/88), ZP_06773754	Amir_2824, 302, (60/73), YP_003100600
TunJ	262	ABC transporter permease subunit	SCLAV_4278, 261, (76/83), ZP_06773753	Amir_2825, 253, (61/78), YP_003100601
TunK	81	Acyl carrier protein	SCLAV_4277, 81, (65/87), ZP_06773752	Amir_2826, 79, (34/54), YP_003100602
TunL	229	Phospholipid phosphatase	SCLAV_4276, 223, (52/67), ZP_06773751	–
TunM	216	Radical SAM protein	SCLAV_4274, 212, (54/67), ZP_06773749	Amir_2815, 232, (30/54), YP_003100591
TunN	152	UTP pyrophosphatase	SCLAV_4275, 170, (68/77), ZP_06773750	–

[a] Number of amino acids; [b] (% Identity / Similarity)

Table 2.4 Comparison of *S. chartreusis tun* cluster with *S. clavuligerus* and *A. mirium* homologues

2.3 *Heterologous expression of* tun *genes*

2.3.1 *Construction of* S. chartreusis *genomic library*

2.3.1.1 Cosmid library overview

A cosmid library is a collection of DNA fragments stored and propagated in *E. coli* within cosmid vectors. In this case the DNA fragments are random 35-45 kb fragments of the *S. chartreusis* genome, with each individual recombinant *E. coli* colony containing a single such fragment. Across 3000 *E. coli* colonies, every part of the *S. chartreusis* chromosome is mapped many times over and as such, a genomic library of the *S. chartreusis* genome can be created in an *E. coli* host.

The SuperCos I cosmid vector used in these libraries are hybrids of plasmid and bacteriophage lambda DNA.[316] Derived from a plasmid are an origin of replication (*ori*), antibiotic resistance genes and suitable restriction sites for cloning. In addition they contain a COS recognition sequence from phage λ DNA, allowing cosmid recombinants to be packaged into viral particles *in vitro* and transduced into bacterial cells. Since most of the phage λ DNA has been discarded, this sequence can be replaced by the DNA of interest. Once in the cell, the cosmids do not form viral particles as the recombinant DNA no longer codes for any λ proteins. Instead, large circular plasmids are formed and maintained within the host cell, conferring antibiotic resistance which can be used as a selection marker. Cosmids can hold much larger inserts than standard plasmids, in the order of 35-45 kb, the size being limited by the amount of DNA which can be packaged into the viral head prior to transduction. In addition, infection of the *E. coli* host by λ virions occurs at a thousandfold greater frequency than transformation by plasmids so this method is an extremely efficient way of cloning DNA into *E. coli*.

There are a number of benefits to producing a cosmid library. It allows the *S. chartreusis* genome to be efficiently screened with labelled DNA probes for the presence of genes of interest – in our case these probes were based on precise sequences from a putative tunicamycin gene cluster previously identified by genome mining. Once hits have been identified, individual cosmids can be isolated which contain only a small proportion of the entire genome yet contain the desired genes. Since *E. coli* is much more tractable than most source organisms, subsequent genetic manipulation of gene clusters contained in specific cosmids is greatly simplified.

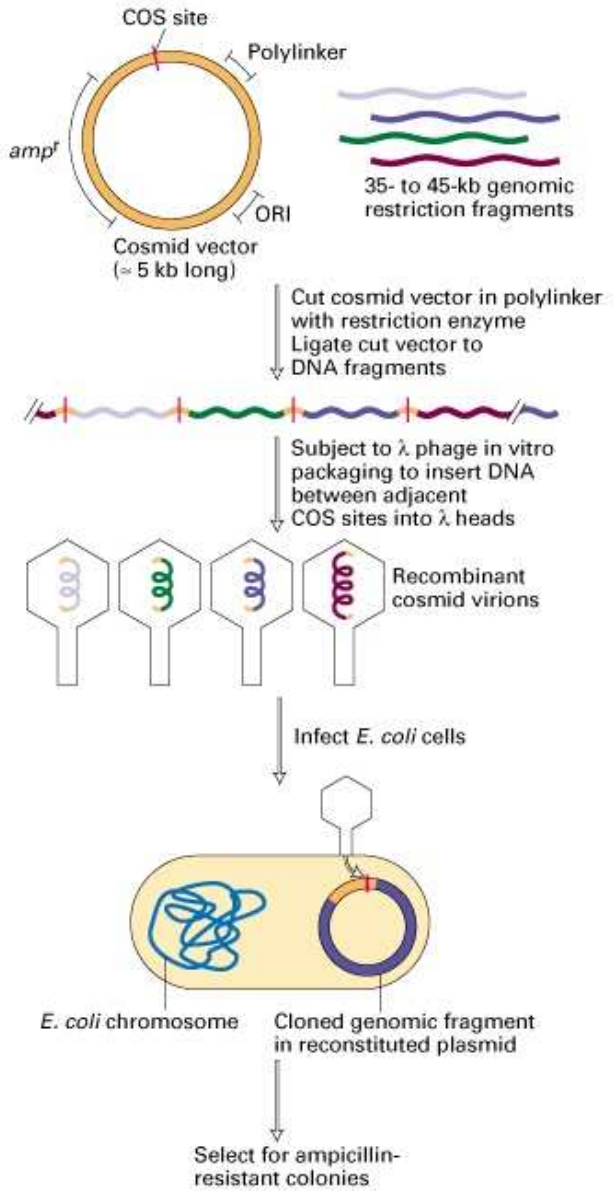


Fig. 2.8 Overview of constructing cosmid library of *S. chartreusis* in *E. coli*.
(reprinted with permission from Lodish *et al*.,[317] © 2008 W.H.Freeman & Co Ltd, New York)

Fig. 2.8 shows a general schematic for the formation of a cosmid library in *E. coli*. First of all, the genomic DNA is partially digested to yield restriction fragments 35-45 kb in size. These fragments, representing all parts of the target genome, are ligated into the polylinker site of a cosmid vector. The products are concatomeric pieces of DNA with multiple copies of cosmids linked end to end, each with a different insert from the genome of interest. The concatomers contain equally spaced COS sites derived from the cosmid vectors. These protein-binding nucleotide sequences are recognised by λ-head proteins, so when preassembled phage λ heads are added to the mixture, DNA segments between two adjacent COS sites are efficiently packaged into an empty viral head. Following addition of

a preassembled phage λ tail, the recombinant virions are ready to infect *E. coli* cells. Once in the cell, the recombinant viral DNA forms a reconstituted circular plasmid which is stably maintained within the bacteria. Successful transduction can be selected for with ampicillin and kanamycin, resistance to which is encoded in the cosmid backbone.

2.3.1.2 Partial Sau3A I digest of genomic DNA

In order to clone genomic DNA fragments into the *Bam*H I site of the SuperCos I cosmid vector, the chromosomal DNA was first partially digested with *Sau*3A I – which cuts at the same site as *Bam*H I. This was done to reduce the size range of the genomic DNA to that of the 30-50 kb inserts accepted by the SuperCos vector, as well as to create the requisite sticky ends for ligation. Conditions for the partial digest were screened on a test scale to determine the optimum restriction enzyme concentration and length of incubation. The results of this screen can be seen in Fig. 2.9.

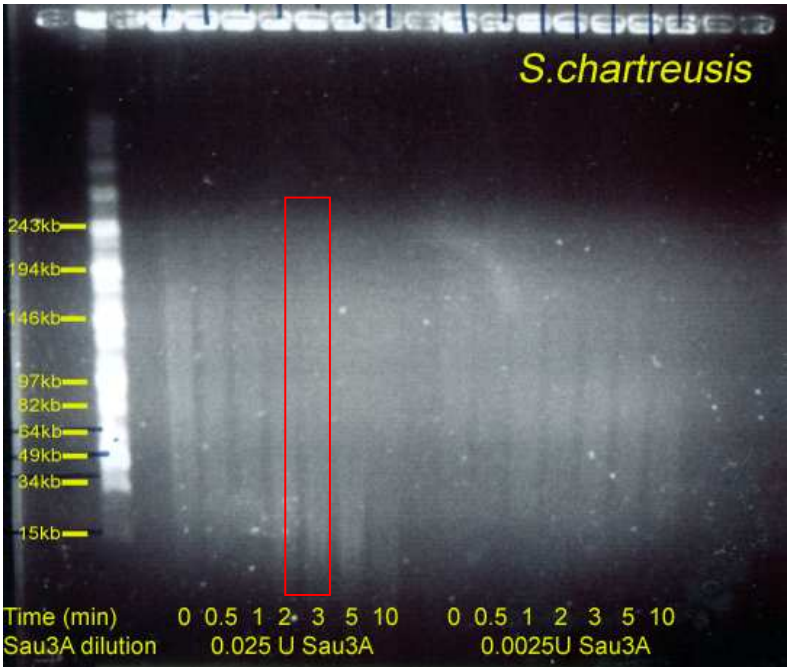


Fig. 2.9 Partial *Sau*3A I digest of *S. chartreusis* genomic DNA

Upon scaling these optimised conditions to a preparatory scale, it was found a slightly shorter incubation time was sufficient to obtain the desired DNA distribution. In order to remove any DNA fragments below 30 kb – which pose the risk of inserting twice into one cosmid vector – the partially digested DNA was separated by gel electrophoresis and fragments of approximately 25 kb and above were excised from the gel (Fig. 2.10, A). DNA was then electroeluted into dialysis tubing, precipitated and resuspended; pulsed-field

gel electrophoresis analysis showed the partially digested *S. chartreusis* genomic DNA to range between 15 and 60 kb in size (Fig. 2.10, B). This distribution was slightly on the low side, but the sample was carried forward for cosmid library construction.

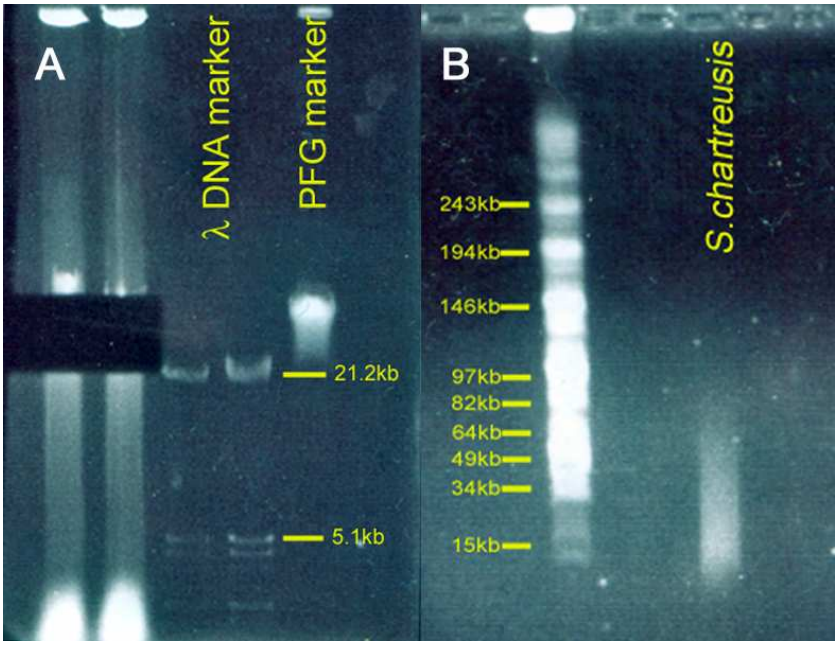


Fig. 2.10 Electroelution of partially digested genomic DNA; **A**: Gel excision of size-fractionated, partially *Sau*3A I-digested *S. chartreusis* DNA; **B**: Pulsed-field gel electrophoresis of the electroeluted and precipitated DNA sample from the gel slice in A.

2.3.1.3 Ligation, packaging and transduction of genomic DNA

The size-fractionated, *Sau*3A I-digested *S. chartreusis* genomic DNA was next cloned into SuperCos I vectors and packaged into recombinant λ-phage according to manufacturer's instructions using a SuperCos I Cosmid Vector Kit (Stratagene) and a Gigapack III Gold Packaging Extract Kit (Stratagene). Briefly, dephosphorylation of the *S. chartreusis* insert DNA was followed by incubation with T4 DNA ligase and SuperCos I vector, previously treated with *Xba* I, calf-intestinal alkaline phosphatase and *Bam*H I. The resulting concatameric recombinant DNA was then packaged into λ phage *in vitro* and used to infect *E. coli* XL1-Blue MR cells. Successful transduction of cosmid DNA into *E. coli* cells was selected for by incubation with ampicillin and kanamycin. These reactions were performed by Andrew Hesketh (JIC). Using the size-fractionated *S. chartreusis* DNA samples described above, a total of 3072 recombinant *E. coli* colonies were individually picked by robot and arrayed onto 3 × 384-well microtiter plates at the Genome Analysis Centre (Norwich, UK). The entire library of 3072 clones contained a total of ~123 Mb *S. chartreusis* DNA, giving 15× theoretical coverage of the genome. DNA from each colony

in this library was then fixed onto nylon membranes according to standard procedures,[318] ready for hybridisation screening with carefully designed PCR probes.

2.3.2 Isolation of cosmids containing tun *operon*

2.3.2.1 Formation of ^{32}P-labelled PCR probes

PCR primers were designed which amplified ~260 bp regions from each end of the *tun* gene cluster, using *S. chartreusis* genomic DNA as template. These amplified regions represented internal fragments of genes *tunA* and *tunN* respectively (Fig. 2.5). The resulting TunA and TunN products were passed through a QIAquick PCR Purification Kit (Qiagen) and analysed on an agarose gel (Fig. 2.11). The products were then labelled with [α-^{32}P]-dCTP, using a Rediprime II Random Prime Labelling Kit (Amersham) in a designated hot lab.

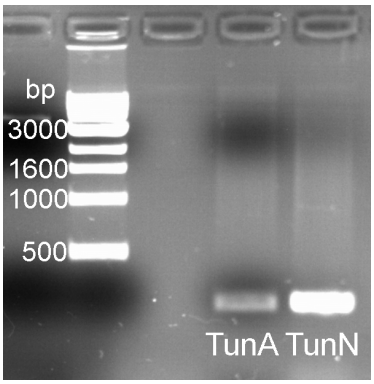


Fig. 2.11 Gel confirmation of TunA and TunN PCR products used to probe cosmid library

2.3.2.2 Hybridisation screening of S. chartreusis cosmid library

Two nylon membranes containing *S. chartreusis* genomic library DNA were washed to remove cell debris and pre-hybridised with calf thymus DNA to reduce non-specific binding of the probe sequence. Next, the filters were hybridised in a rotisserie oven with a solution of the radioactive probe at 60 °C overnight; one was incubated with TunA probe, the other with TunN. After washing the filters, binding events were visualised by exposure to a phosphoimager plate. The resulting image can be seen in Fig. 2.12, overlaid with an array grid. The image shows a number of specific binding events with very little background activity.

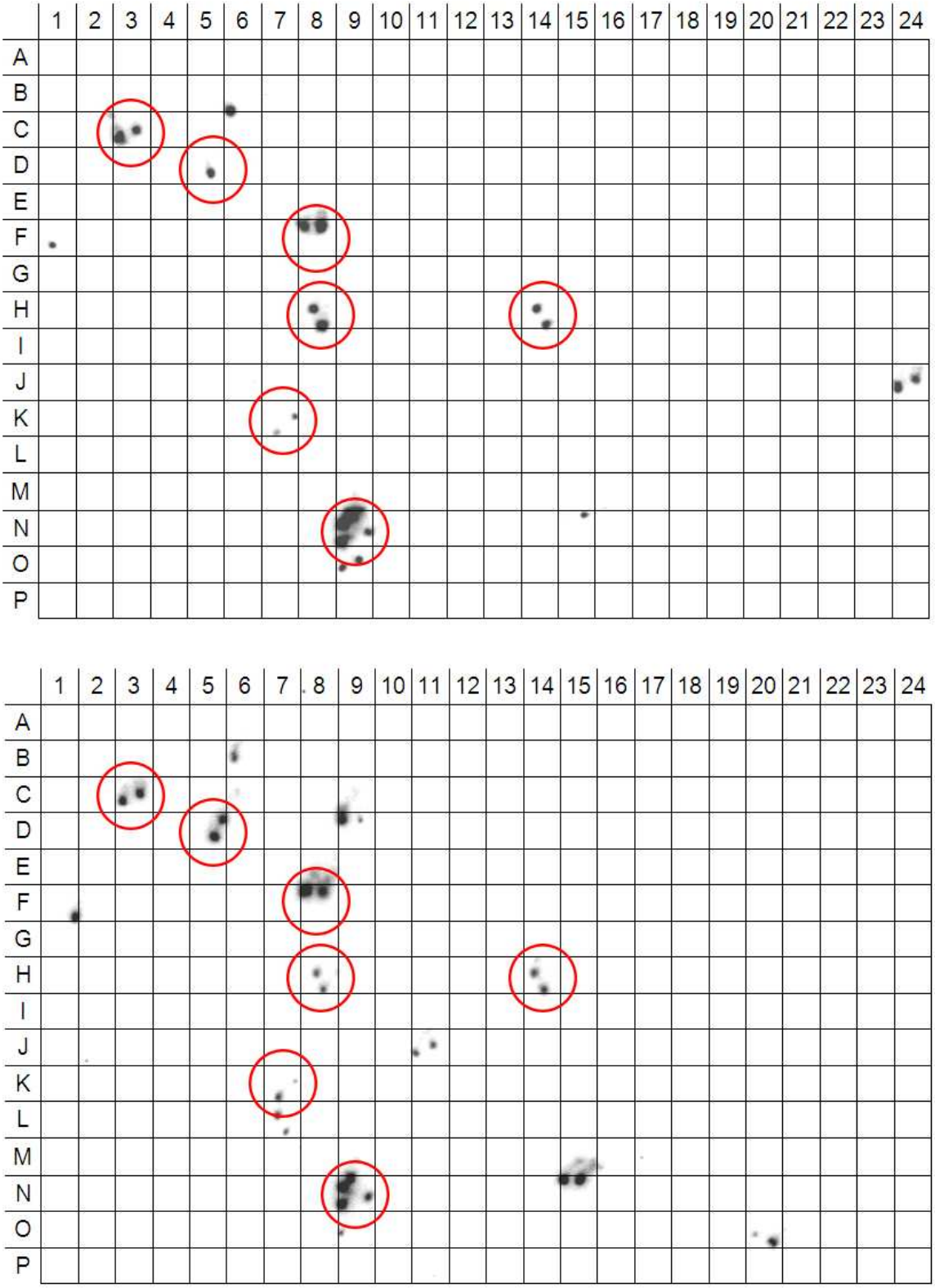


Fig. 2.12 Hybridisation of ^{32}P-labelled TunA (top) and TunN (bottom) probes to cosmid library filter. Colonies binding to both probes are circled in red

All 3072 colonies from the 8 × 386-well plates in the *S. chartreusis* library were printed onto each individual nylon membrane. Since these membranes were arrayed in 386-well format, each grid position on the membrane contains DNA from eight separate *E. coli* colonies. To distinguish between them each grid square has colonies printed in two positions of a 4 × 4 matrix (Fig. 2.13). The pattern of these two dots allows the precise origin of each hit to be located within the 8 plate library. For example in Fig. 2.13, the circled hit farthest to the left clearly corresponds to position C3 of plate 7, abbreviated to 7C3. Noting that cosmids containing the entire *tun* cluster will bind both TunA and TunN probes, hits appearing on both arrays were selected (circled in Fig. 2.12). The locations of these cosmids were determined to be 1N9, 2F8, 3D5, 4H8, 4N14, 5K7, 6N9 and 7C3. It should be noted that grid position N9 showed four dots and thus yielded two hits, from plates 1 and 6 respectively.

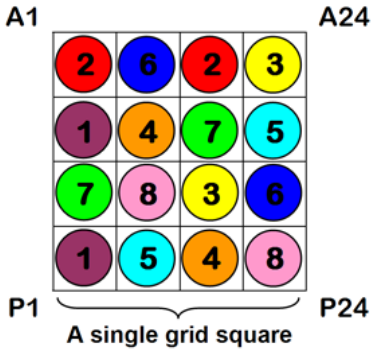
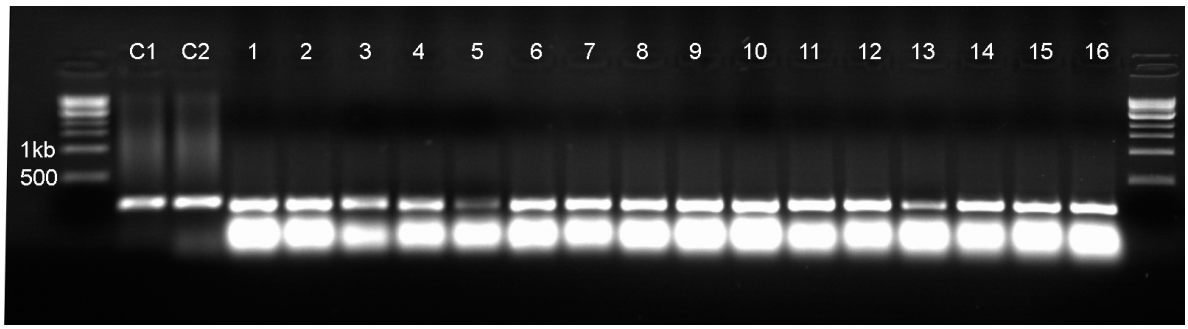


Fig. 2.13 Key for specifying the "address" of each colony within a given grid square

2.3.2.3 Restriction digest of isolated cosmids

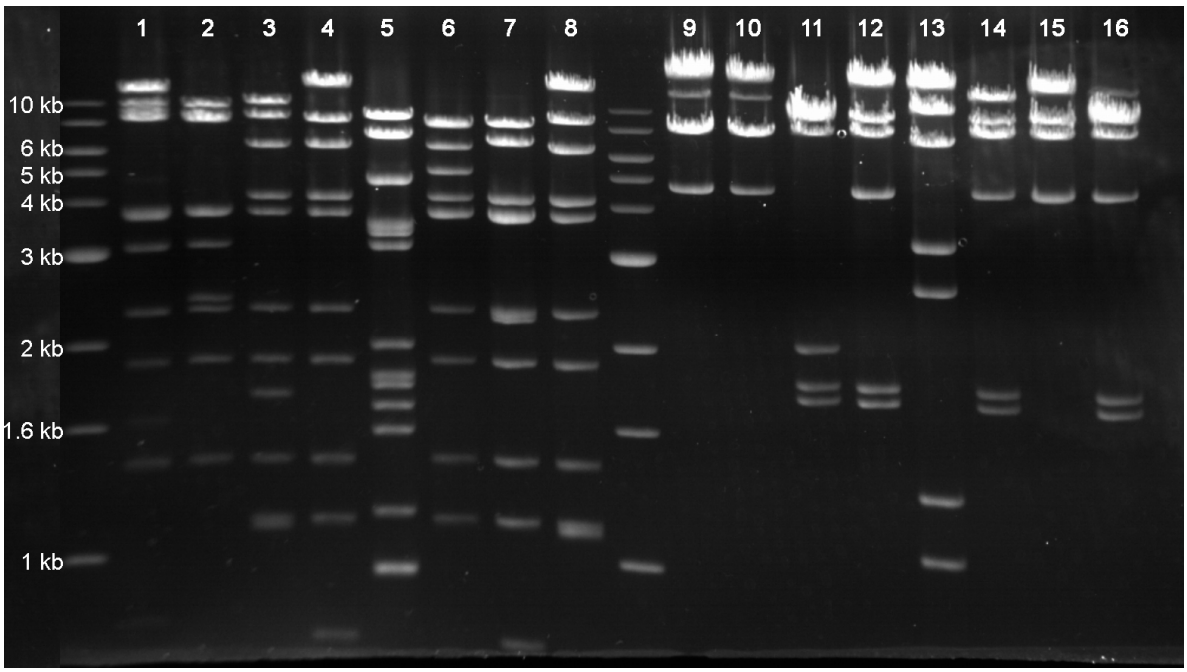
E. coli colonies from each of these positions in the genomic library were picked, propagated and the eight different cosmids isolated from them. These cosmids were then used as the template for the formation of TunA and TunN probes by PCR (Fig. 2.14). Each PCR reaction gave a product of equivalent size to the control products TunA and TunN (Fig. 2.14, lanes C1 and C2). As such all eight isolated cosmids provided a successful template for the formation of TunA and TunN, confirming that the colonies specified by the radioactive probes were correctly identified and isolated from the genomic library. The lower diffuse bands present correspond to RNA contaminants from the cosmid preparation.



Lane C1: TunA primers, *S. chartreusis* genomic DNA template;
Lane C2: TunN primers, *S. chartreusis* genomic DNA template;
Lanes 1-8: TunA primers, cosmid DNA template (1N9, 2F8, 3D5, 4H8, 4N14, 5K7, 6N9, 7C3);
Lanes 9-16: TunN primers, cosmid DNA template (1N9, 2F8, 3D5, 4H8, 4N14, 5K7, 6N9, 7C3);

Fig. 2.14 PCR amplification of TunA and TunN probes using isolated cosmids as template

The eight cosmids isolated from the *S. chartreusis* genomic library were also subjected to restriction analysis with *Bam*H I and *Xho*I to further verify the presence of the *tun* cluster (Fig. 2.15). The restriction pattern was compared to that from contigs 02796 and 00341, which when combined together contain the *tun* cluster and 14 kb of flanking sequence (Fig. 2.16). Cosmid 4N14 (Fig. 2.15, lanes 5 and 13) immediately looked anomalous and was not considered further. The remaining seven cosmids contained all of the characteristic bands originating from within the *tun* cluster, except for the 367 bp band which was too small to see on the agarose gel. The presence of restriction fragments flanking the *tun* cluster were then considered, to identify cosmids whose inserts have the *tun* genes located as centrally as possible. Cosmids 1N9 and 2F8 did not show a 4.2 kb band in the *Bam*H I digest originating from the left hand end of the *tun* cluster and thus had these genes very close to the edge of their cosmid insert (Fig. 2.15, lanes 1-2). Similarly, cosmid 3D5 did not have the 4.7 kb band in its *Xho*I digest originating from the right hand end of the *tun* cluster (Fig. 2.15, lane 11). As such, cosmids 4H8, 5K7, 6N9 and 7C3 remained as the best candidates for subsequent genetic manipulation. It should be noted that without any further sequence information from this region of the *S. chartreusis* chromosome, the precise origins of the other bands within the restriction digests could not be determined. In addition a number of the extra bands present are > 8 kb and are poorly resolved. It should also be noted that *Bam*H I cuts only once and *Xho*I does not cut at all in the 6852 bp SuperCos I backbone.



Lanes 1-8: *Bam*H I digestion; Cosmids 1N9, 2F8, 3D5, 4H8, 4N14, 5K7, 6N9, 7C3.
Lanes 9-16: *Xho*I digestion; Cosmids 1N9, 2F8, 3D5, 4H8, 4N14, 5K7, 6N9, 7C3.

Fig. 2.15 Restriction digest of isolated cosmids with *Bam*H I and *Xho*I

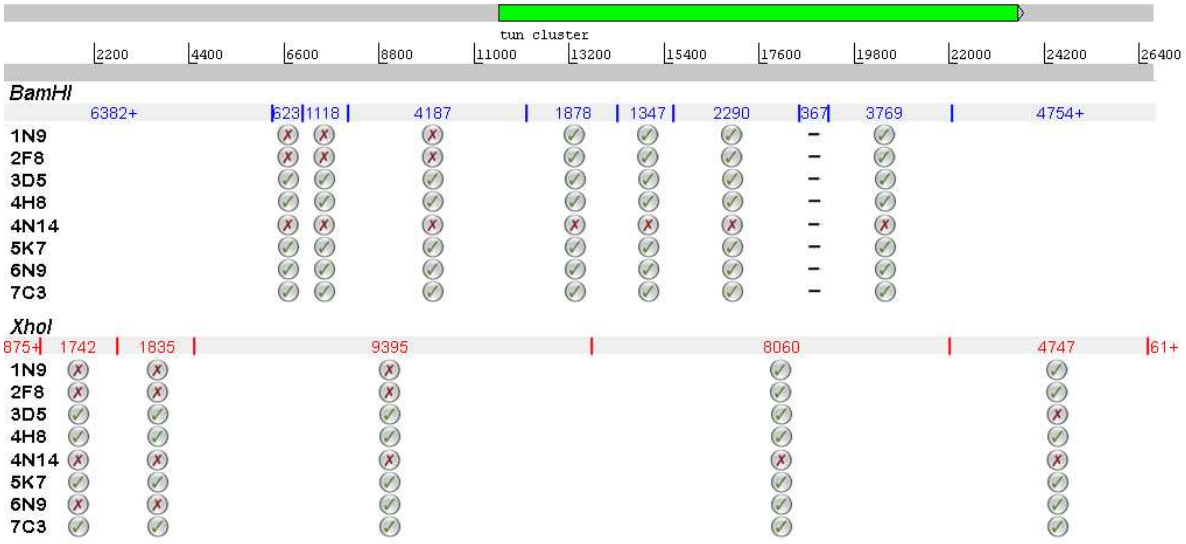
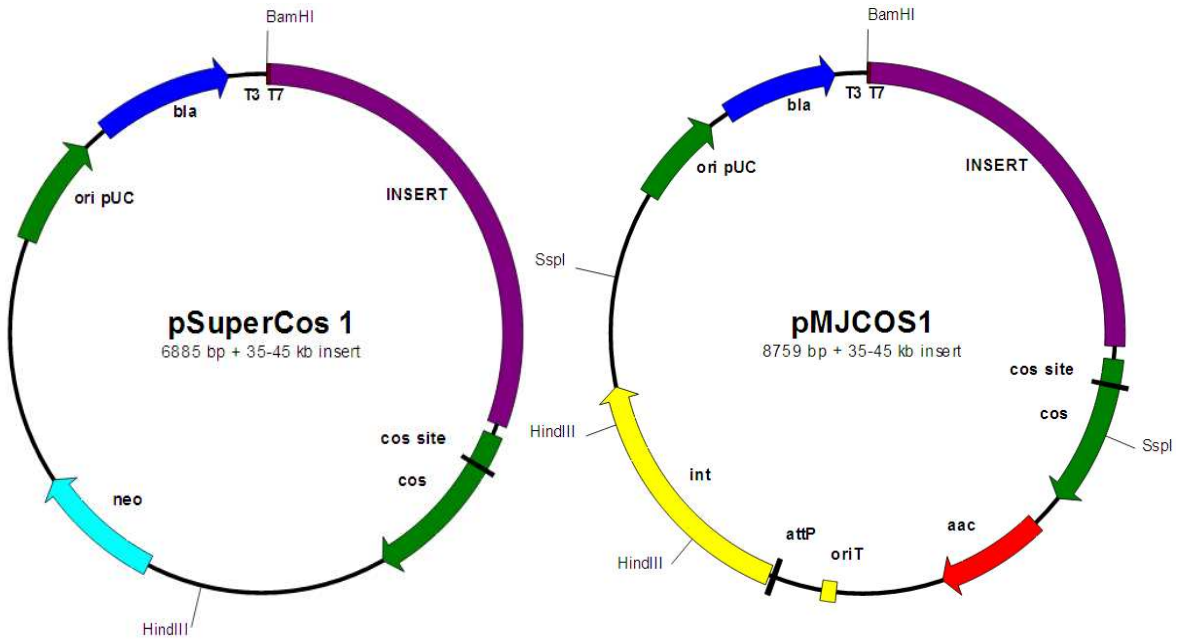


Fig. 2.16 Comparison of cosmid digest to theoretical restriction digest of contigs 02796 and 00341

2.3.3 Modification of cosmid backbones and cloning into S. coelicolor hosts

In order to screen for heterologous expression of the *tun* gene cluster in a non-producing *Streptomyces* host, the backbone of the cosmid vector must first be modified to make it conjugative and integrative. The cosmids isolated from the *S. chartreusis* genomic library do not contain a *Streptomyces* origin of replication or machinery to integrate their sequence into the host's chromosome. If they were conjugated from *E. coli* into *Streptomyces*, they would not be able to replicate or survive in the host strain. To overcome this, the SuperCos I vector backbone was converted to that of pMJCOS1.[319]



neo	kanamycin resistance gene (neomycin phosphotransferase II)
bla	carbenicillin resistance gene (β-lactamase)
aac	apramycin resistance gene (aminoglycoside acetyltransferase type IV – *aac(3)-IV*)
oriT	origin of transfer from IncP-group plasmid RK2
attP	φC31 phage attachment site
int	φC31 integrase
INSERT	35-45 kb insert of *S. chartreusis* genomic DNA (not drawn to scale)
T3	T3 priming site
T7	T7 priming site
***ori* pUC**	ColE1 origin of replication derived from pUC vectors
***cos* site**	λ phage cos packaging site
cos	λ phage cos recognition domain

Fig. 2.17 Maps of SuperCos 1 and pMJCOS1 vectors

Cosmid pMJCOS1 shares a lot of sequence with SuperCos I (Fig. 2.17), but the kanamycin resistance cassette (*neo*) has been replaced with an apramycin resistance gene (*aac(3)-IV*) to allow for differential antibiotic selection between the two vectors. The most important additions however are *oriT*, *attP* and *int*. Successful conjugative transfer from *E. coli* to *Streptomyces* is dependent on the presence of both an origin of transfer (*oriT*) – from IncP-group plasmid RK2 – and additional genes mediating the transfer process termed 'transfer functions'.[320] Typically, only *oriT* must be present in the plasmid destined for conjugative transfer; the transfer functions are supplied *in trans* by an additional driver plasmid in the *E. coli* donor strain, such as self-transmissible pUB307 or the non-transmissible pUZ8002.[289] Once in *Streptomyces*, recombinase machinery from the temperate phage φC31 is used to irreversibly integrate the cosmid sequence into the chromosome. Expression of the φC31-derived integrase (*int*) catalyses site-specific recombination between the vector (*attP*) and chromosome (*attB*) attachment sites.[321,322]

To convert the SuperCos I backbone to that of pMJCOS1, a λ-Red mediated recombination strategy was employed (see Fig. 2.18).[323,324] First of all, the four suitable cosmids (pIJ12311-4, see Table 2.5) isolated from the *S. chartreusis* genomic library were transformed into *E. coli* BW25113/pIJ790 by electroporation. After a round of antibiotic selection, individual colonies of successful transformants were cultured and further transformed with a 5.2 kb *Ssp*I fragment from pMJCOS1 that contained the key genes mentioned above – *aac(3)-IV*, *oriT*, *attB* and *int* from φC31 – as well as flanking sequences with identity to corresponding regions of the SuperCos I backbone (see Fig. 2.17). Prior to electroporating the cells, expression of λ-Red genes contained within the pIJ790 plasmid was induced with L-arabinose. These λ-Red (*gam*, *bet*, *exo*) genes promote a greatly enhanced rate of double-crossover recombination using linear DNA.[325] Successful transformation of the *Ssp*I DNA fragment into *E. coli* BW25113/pIJ790/pIJ12311-4 and subsequent recombination with the SuperCos I backbone of pIJ12311-4 was selected for using apramycin. In addition, due to pIJ790 containing a temperature-sensitive replication domain (*rep101*^{ts}),[325] growing the apramycin resistant recombinants at 37 °C rather than 30 °C promoted the loss of pIJ790 which left an *E. coli* BW25113 strain containing only recombined cosmids pIJ12315-8 (see Table 2.5).

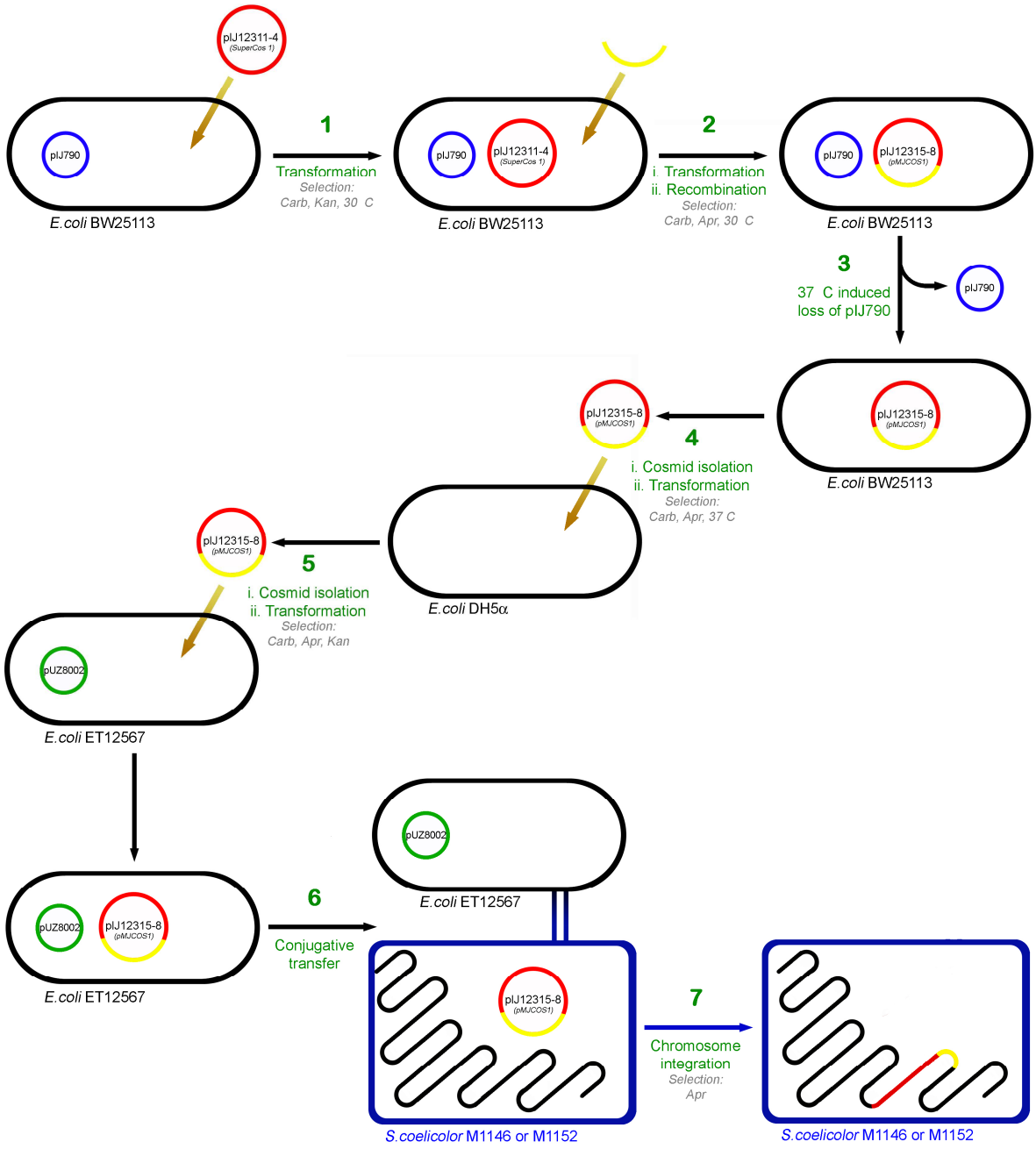


Fig. 2.18 Overview of λ-Red mediated recombination of cosmid backbone

Unfortunately, the *E. coli* BW25113 strain used can maintain multiple copies of plasmids and cosmids, so each cell may contain a mixture of original SuperCos I-derived cosmids (pIJ12311-4) and their corresponding pMJCOS1-based recombinants (pIJ12315-8). For this reason, total cosmid preparations were performed on combined samples of all the apramycin resistant colonies and these cosmid mixtures were transformed directly into *E. coli* DH5α. As this strain only accepts one cosmid per cell, apramycin selection was now entirely selective for colonies harbouring pMJCOS1-based cosmids pIJ12315-8.

Name	Cosmid Description
pIJ12311	Cosmid library clone 4H8 with SuperCos I backbone
pIJ12312	Cosmid library clone 5K7 with SuperCos I backbone
pIJ12313	Cosmid library clone 6N9 with SuperCos I backbone
pIJ12314	Cosmid library clone 7C3 with SuperCos I backbone
pIJ12315	pMJCOS1 packaged with a *S. chartreusis* genomic DNA fragment from cosmid library clone 4H8
pIJ12316	pMJCOS1 packaged with a *S. chartreusis* genomic DNA fragment from cosmid library clone 5K7
pIJ12317	pMJCOS1 packaged with a *S. chartreusis* genomic DNA fragment from cosmid library clone 6N9
pIJ12318	pMJCOS1 packaged with a *S. chartreusis* genomic DNA fragment from cosmid library clone 7C3

Table 2.5 Description and names of novel SuperCos I- and pMJCOS1-based DNA constructs

Following this step, uncontaminated samples of cosmids pIJ12315-8 were isolated and further transformed into *E. coli* ET12567/pUZ8002 by electroporation. The pUZ8002 helper plasmid in this strain contains transfer functions necessary for *in trans* mobilisation of *oriT*-containing vectors from *E. coli* to *Streptomyces via* intergeneric conjugation[320] – this plasmid has had its own *oriT* sequence modified so it itself is not transferred.[326] The ET12567 strain of *E. coli* used is methylation deficient, so any DNA transferred into *Streptomyces* is not prone to methyl-specific restriction mechanisms present in some host strains.[327] Successful *E. coli* ET12567/pUZ8002/pIJ12315-8 transformants were isolated and grown overnight on mannitol-soyaflour (MS) agar in the presence of heat-shocked spores from *S. coelicolor* M1146 and M1152 strains (see Table 2.8). The resulting plates were then overlaid with nalidixic acid to selectively kill the *E. coli* cells and with apramycin to select for successful cosmid transfer and integration into the host chromosome.[328,329] This yielded recombinant *S. coelicolor* strains harbouring *S. chartreusis* genomic fragments derived from cosmid library clones 4H9, 5K7, 6N9 and 7C3 respectively (see Table 2.6). As such, each of these recombinant strains possessed a copy of the putative *tun* gene cluster. Spores from these strains were collected and used in subsequent heterologous expression experiments.

The two strains used in these experiments were developed by Dr Gomez-Escribano (JIC) to act as super-hosts. They are based on the fully sequenced model organism *Streptomyces coelicolor* A3(2)[330] and have had four antibiotic clusters knocked out together with mutations to RNA polymerase B in order to boost expression and production of any heterologously introduced gene clusters (see Table 2.8).

2.3.4 Cloning minimal tun *cluster directly from* S. chartreusis *genomic DNA*

In the previous section, the putative *tun* cluster was cloned by manipulating the backbones of *S. chartreusis* genomic library cosmids, converting them to pMJCOS1-derived integrative and conjugative vectors which can enter heterologous *Streptomyces* hosts and integrate into their chromosomes. In parallel with these experiments, attempts were made by Andrew Hesketh (JIC) to clone the putative *tun* operon from a genomic library cosmid directly into a similarly conjugative and integrative vector, using conventional DNA restriction and ligation techniques. The benefit of this approach was that it avoided cloning significant flanking sequences either side of the putative *tun* gene cluster, as was the case with the λ-Red recombination approach. Here, prudent choice of restriction enzyme would result in exclusive cloning of the 14 ORFs in the putative *tun* operon, together with only minimal regions of flanking sequences. An added benefit of this method was the potential for improved simplicity and overall efficiency of the cloning process, all of which was made possible by the relatively small size of the putative *tun* gene cluster – at only 12.0 kb any resulting restriction fragments could be easily handled *in vitro* by conventional methods.

As such, a 12.9 kb *Sac*I fragment from genomic library cosmid 4H8 (pIJ12311) was cloned directly into the *Sac*I site of pRT802,[331] yielding construct pIJ12003. The *Sac*I fragment contained the complete predicted *tun* cluster plus 427 bp upstream of *tunA* and 500 bp downstream of *tunN*; neither of these flanking DNA sequences was predicted to possess an entire ORF. Subsequently, pIJ12003a was conjugated into *S. coelicolor* M1146 *via* triparental mating using *E. coli* S17-1[332] and *E. coli* ET12567/pUZ8002. The resulting recombinant *S. coelicolor* strain was consequently subjected to heterologous expression experiments (see section 2.3.5). If heterologous expression were to be observed from these clones, the boundaries of the minimal *tun* cluster required for tunicamycin production would be delineated, ruling out the presence of any essential yet overlooked genes outside of the putative *tun* gene cluster.

2.3.5 Heterologous expression in S. coelicolor *hosts*

2.3.5.1 Summary of recombinant S. coelicolor *strains*

Two parallel approaches to cloning the putative *tun* gene cluster from *S. chartreusis* were reported in the previous two sections. These led to the construction of nine recombinant *S. coelicolor* strains each harbouring a copy of the putative *tun* operon. Additionally, three further *S. coelicolor* recombinant strains were obtained as controls, produced by analogous methods and containing the relevant vectors lacking any DNA insert. All of these new *S. coelicolor* strains were assigned new names for clarity (Table 2.6) and are described further in the experimental section (Table 2.8 and Table 2.10).

Name	Description
***S. coelicolor* host strains**	
S. coelicolor M1146	Model strain *S. coelicolor* M145[330] with deleted actinorhodin (*act*), undecylprodigiosin (*red*), calcium-dependent antibiotic (*cda)* and type I PKS cluster CPK (*cpk*) gene clusters
S. coelicolor M1152	*S. coelicolor* M1146 carrying a point mutation in *rpsL* for enhanced expression of antibiotic gene clusters
pMJCOS1-based clones – obtained by λ-Red mediated recombination	
S. coelicolor M1021	*S. coelicolor* M1146 with pIJ12315 (derived from library clone 4H8)
S. coelicolor M1022	*S. coelicolor* M1146 with pIJ12316 (derived from library clone 5K7)
S. coelicolor M1023	*S. coelicolor* M1146 with pIJ12317 (derived from library clone 6N9)
S. coelicolor M1024	*S. coelicolor* M1146 with pIJ12318 (derived from library clone 7C3)
S. coelicolor M1025	*S. coelicolor* M1152 with pIJ12315 (derived from library clone 4H8)
S. coelicolor M1026	*S. coelicolor* M1152 with pIJ12316 (derived from library clone 5K7)
S. coelicolor M1027	*S. coelicolor* M1152 with pIJ12317 (derived from library clone 6N9)
S. coelicolor M1028	*S. coelicolor* M1152 with pIJ12318 (derived from library clone 7C3)
S. coelicolor M1029	CONTROL: *S. coelicolor* M1146 with empty pMJCOS1 vector (vector-only control strain)
S. coelicolor M1030	CONTROL: *S. coelicolor* M1152 with empty pMJCOS1 vector (vector-only control strain)
pRT802-based clones – obtained by ligation	
S. coelicolor M1031	*S. coelicolor* M1146 with pIJ12003a (minimal *tun* gene cluster in pRT802 plasmid)
S. coelicolor M1035	CONTROL: *S. coelicolor* M1146 with pRT802 (vector-only control strain)

Table 2.6 Descriptions of Streptomyces coelicolor recombinant strains

These recombinant *S. coelicolor* strains were screened for evidence of tunicamycin production in order to verify the function of the putative *tun* gene cluster in tunicamycin biosynthesis.

2.3.5.2 Agar-diffusion bioassays

The basic procedure for the tunicamycin bioassay involves growing recombinant or control *Streptomyces* strains on suitable agar until the mycelial forming stage. Cylindrical bores are transferred to an empty plate and flooded with low melting point agar pre-infected with a reporter strain. After incubation for 24 h, the plates are screened for zones of inhibition surrounding the *Streptomyces* colonies. Two common reporter strains were screened for their susceptibility to tunicamycin, *Micrococcus luteus* and *Bacillus subtilis* EC1524. It was found that only *B. subtilis* EC1524 was efficiently killed upon exposure to tunicamycin and so this strain was subsequently used as a reporter strain for all bioassay experiments (Fig. 2.19).

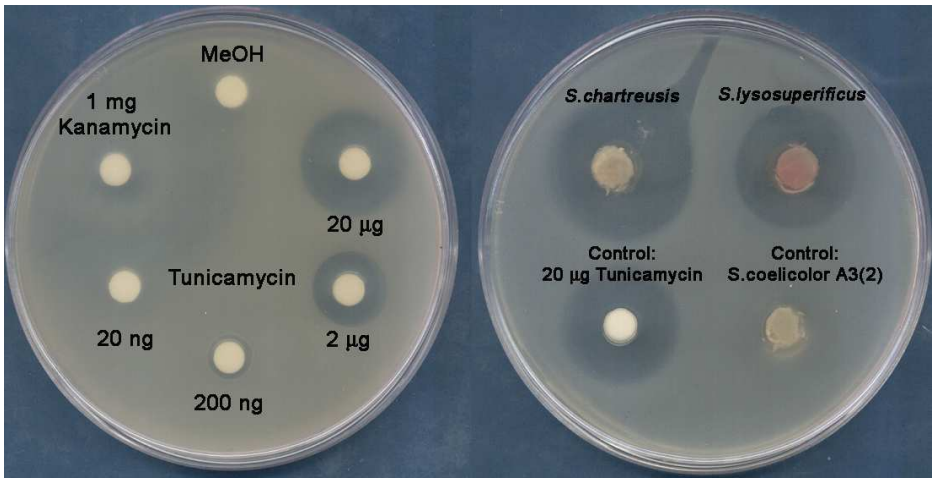


Fig. 2.19 Tunicamycin susceptibility bioassay using *B. subtilis* EC1524 as reporter strain.

Fig. 2.19 shows *B. subtilis* was efficiently killed by tunicamycin, with 2 µg enough to produce a halo of inhibition. The kanamycin control produced a very broad and diffuse halo, suggesting it was not highly toxic to *B. subtilis* and hence not an ideal control antibiotic. The MeOH control simply showed that the MeOH in which the tunicamycin dilutions were dissolved had no effect on the growth of the reporter strain. On the right hand side of Fig. 2.19, tunicamycin producing strains *S. chartreusis* NRRL3882 and *S. lysosuperificus* ATCC31396 clearly produced an antibiotic toxic to *B. subtilis*, whereas the *S. coelicolor* A3(2) model organism showed no zone of inhibition. These test bioassays

identified *B. subtilis* as a suitable reporter strain, and provided optimised growth conditions and antibiotic concentrations for accurate and reproducible tunicamycin bioassays.

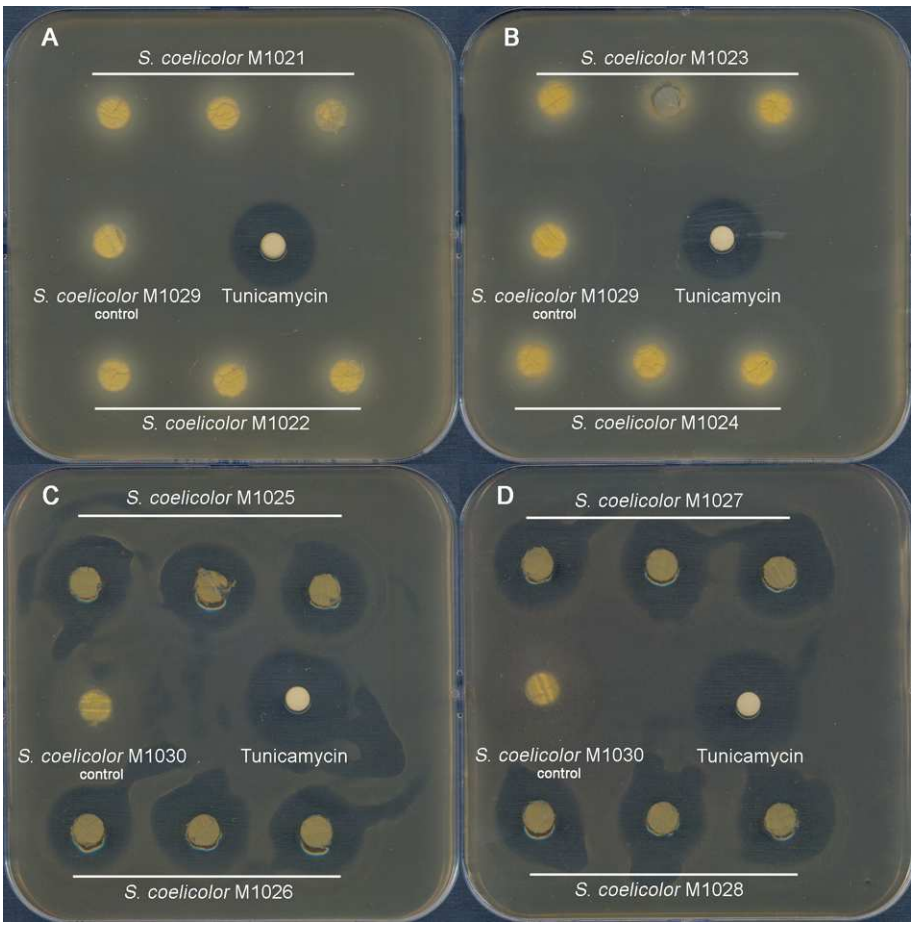


Fig. 2.20 Agar-diffusion bioassays of recombinant *S. coelicolor* strains with cosmids pIJ12315-8.

Agar-diffusion bioassays of pMJCOS1-containing *S. coelicolor* recombinants are shown in Fig. 2.20. The results clearly showed that *S. coelicolor* strains M1025-8 all produced an antibiotic toxic to *B. subtilis* (Fig. 2.20, A and B). All four clones produced an equal response and were consistent when repeated in triplicate. These strains contained one of the four cosmids derived from genomic library clones 4H8, 5K7, 6N9 and 7C3 (pIJ12315-8) within host *S. coelicolor* M1152 (see Table 2.6). When compared with the lack of bioactivity of vector-only control strains, production of the antibiotic toxic to *B. subtilis* could be assigned with confidence to the *S. chartreusis*-derived sequences in cosmids pIJ12315-8.

It was curious, however, to observe no heterologous production in *S. coelicolor* strains M1021-4 (Fig. 2.20, C and D). These strains contained exactly the same plasmids in alternative host *S. coelicolor* M1146, which is identical to the M1152 strain except that it lacks specific point mutations to the resident *rpoB* gene (see Table 2.6). This mutation has

been shown to boost levels of antibiotic production and may explain the differences observed.[333]

Agar-diffusion bioassays of pRT802-containing *S. coelicolor* recombinants are shown in Fig. 2.21. The result clearly showed that *S. coelicolor* M1031 produced an antibiotic toxic to the *B. subtilis* reporter strain. This strain contained the minimal putative *tun* cluster in plasmid pRT802 within host *S. coelicolor* M1146 (see Table 2.6). The control strain *S. coelicolor* M1035 showed no zones of inhibition yet differed from *S. coelicolor* M1031 only by the lack of the 12.0 kb putative *tun* gene cluster plus ~500 bp of flanking sequences (see Table 2.6). This was a highly encouraging result, confirming that the putative *tun* operon codes for the antibiotic observed in the bioassay.

Fig. 2.21 Agar-diffusion bioassay of the recombinant *S. coelicolor* strain with cosmid pIJ12003a

Both approaches to cloning the putative *tun* cluster from *S. chartreusis* yielded recombinant *S. coelicolor* strains which showed heterologous expression of an antibiotic toxic to *B. subtilis*. The positive result from the cloning of the minimal putative *tun* cluster confirmed that this antibiotic can only be coded for by the 14 genes contained within this operon. In the subsequent sections, the isolation and chemical characterisation of this unknown antibiotic responsible for the halos of inhibition is reported.

It is interesting to note that when using the host *S. coelicolor* M1146, clones derived from pMJCOS1-based cosmids pIJ12315-8 (strains *S. coelicolor* M1021-4, see Table 2.6) showed no heterologous production, yet clones with cosmid pIJ12003a containing the minimal putative *tun* gene cluster successfully underwent heterologous expression (strain *S. coelicolor* M1031, see Table 2.6). The reasons for this are unclear, but it is possible that there is a sequence outside of the cloned minimal *tun* cluster which mediates transcriptional regulation in some way. Further studies are required to investigate this suggestion.

2.3.5.3 MS and ^{1}H NMR verification of heterologous tunicamycin production

To confirm the identity of the bactericidal metabolite produced by our recombinant bacteria, *S. coelicolor* strains M1027 and M1031 and their respective controls – *S. coelicolor* M1030 and M1035 (see Table 2.6) – were grown in liquid culture for five days, after which the pelleted mycelium was extracted with MeOH. These crude extracts were then characterised by liquid-chromatography/mass spectrometry (LC/MS) (Fig. 2.22).

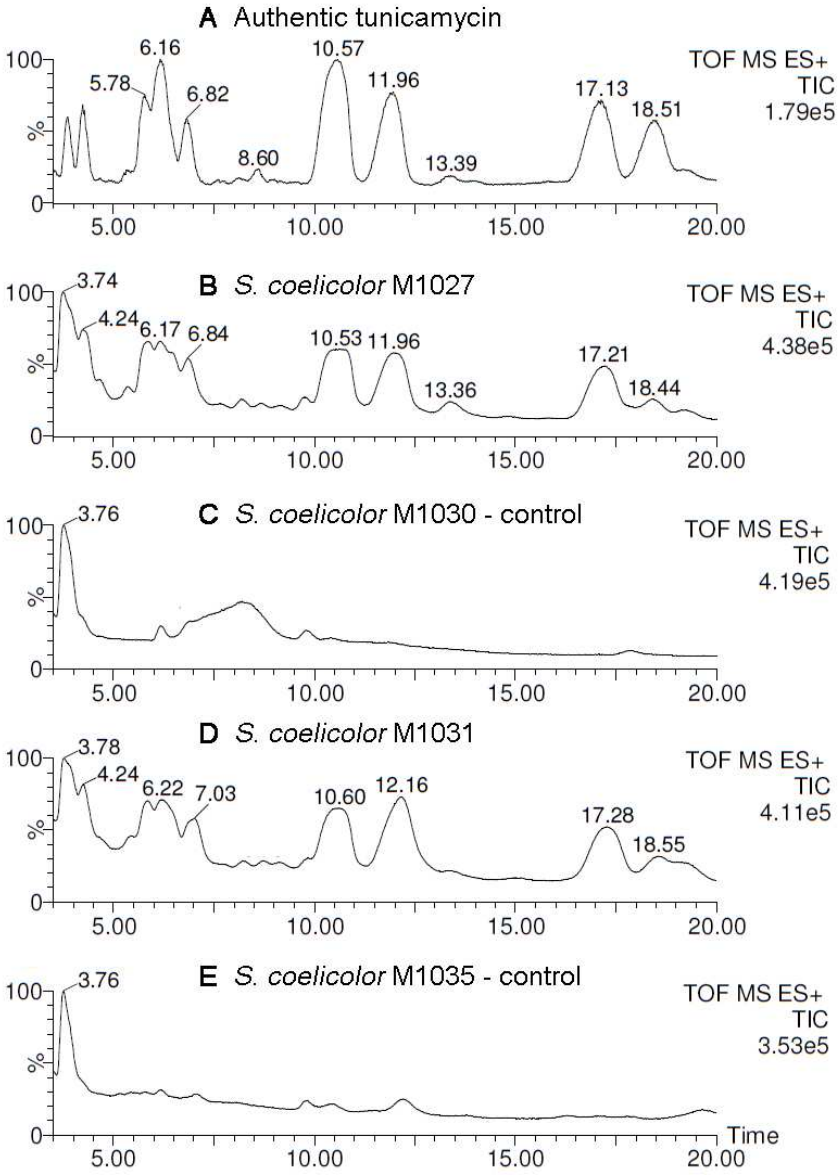


Fig. 2.22 Total ion chromatograms from LC/MS of extracts from recombinant *S. coelicolor* strains; For more detailed strain descriptions see Table 2.6.
A: Authentic tunicamycin sample;
B: *S. coelicolor* M1027 (strain with pMJCOS1-based cosmid pIJ12317 in host *S. coelicolor* M1146);
C: *S. coelicolor* M1027 (vector-only control of strain M1027);
D: *S. coelicolor* M1030 (minimal *tun* cluster in plasmid pRT802 in host *S. coelicolor* M1146);
E: *S. coelicolor* M1035 (vector-only control of strain M1031);

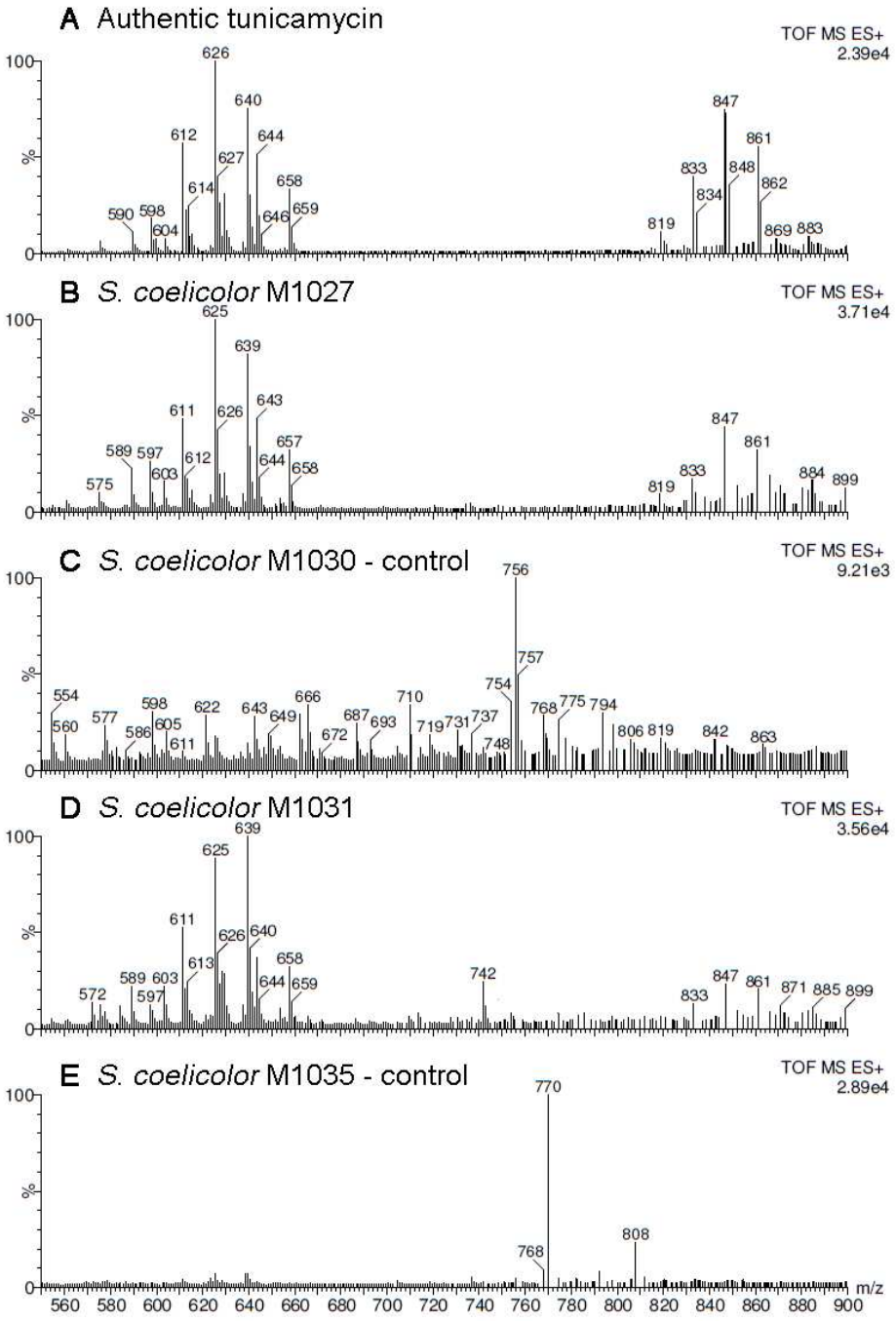


Fig. 2.23 Mass spectra from LC/MS analysis of extracts from recombinant *S. coelicolor* strains; (Individual mass spectra were extracted from LC/MS total ion chromatograms (see Fig. 2.22) for retention times between 3.5 and 20 min, then combined to yield a single summed spectrum for each extract sample); For more detailed strain descriptions see Table 2.6.

A: Authentic tunicamycin sample;

B: *S. coelicolor* M1027 (strain with pMJCOS1-based cosmid pIJ12317 in host *S. coelicolor* M1146);

C: *S. coelicolor* M1027 (vector-only control of strain M1027);

D: *S. coelicolor* M1030 (minimal *tun* cluster in plasmid pRT802 in host *S. coelicolor* M1146);

E: *S. coelicolor* M1035 (vector-only control of strain M1031);

For both *S. coelicolor* M1027 and M1031, the resulting total ion chromatograms (TICs) showed sets of peaks between 3.5 and 20 min with retention times identical to those found in an authentic tunicamycin sample; these peaks were absent from control strains (Fig. 2.22). Additionally, mass spectra obtained from these samples were found to contain mass distributions and fragmentation patterns precisely matching those of authentic tunicamycin (Fig. 2.23). Mass spectra of individual peaks within each TIC – corresponding to tunicamycin homologues running at different retention times – revealed that the heterologously expressed tunicamycins and the authentic sample shared an almost identical fatty acid distribution (Appendix A3). It can be concluded that either the fatty acid composition of the host *S. coelicolor* strain is very similar to that of the *S. lysosuperificus* which produced the commercial standard, or that the *tun* genes show a high level of specificity for particular acyl chain lengths when sequestering cellular lipids during tunicamycin biosynthesis.

The presence of tunicamycin product was further confirmed by ^{1}H NMR spectroscopy. Crude extracts from *S. coelicolor* M1031 were purified by flash chromatography and subjected to ^{1}H NMR analysis alongside an authentic tunicamycin sample. The resulting spectra were identical and can be seen in appendix A4.

Overall, these results show beyond doubt that heterologous production of tunicamycin has been achieved and that all the genes necessary are contained within the fourteen-gene *tun* cluster.

2.3.5.4 Yield of heterologous tunicamycin production

A number of recombinant *S. coelicolor* strains were selected and grown in different liquid media. The tunicamycin extraction/purification procedure described in section 3.2.3 was performed in each case after 7 days of growth and the yields of tunicamycin compared (Table 2.7). Initially, R5 production medium was used, which is a rich medium designed to maintain a complete nutrient supply, promote dispersed growth and maximise secondary metabolite production. Unfortunately, R5 contains 100 g sucrose per litre of medium, and this caused significant contamination during tunicamycin purification. As such, despite giving a slightly higher yield than with tryptone-yeast extract-dextrose (TYD) medium in mg per litre of culture, the yield with respect to cell pellet weight was lower. The medium was not used any further due to its high relative cost and the significant problems it provides during purification.

The most significant observation, however, was that both recombinant strains failed to produce more tunicamycin than native *S. charteusis*. This is surprising given the *S. coelicolor* strains M1146 and M1152 used are designed to act as super hosts; they boost antibiotic production by having existing antibiotic gene clusters knocked out and having advantageous mutations to the gene coding for RNA polymerase B. The reasons for this drop in relative production levels is unclear, although may be related to as-yet-unknown regulatory elements or self-resistance mechanisms additional to the putative ABC transporter encoded within the *tun* gene cluster.

Strain	Media	Tunicamycin yield (mg/L culture)	Tunicamycin yield (mg/g cell pellet)
S. coelicolor M1031	R5	25	1.5
S. coelicolor M1027	R5	8	0.4
S. chartreusis	R5	41	2.3
S. coelicolor M1031	TYD	7	2.7
S. coelicolor M1027	TYD	8	0.5
S. chartreusis	TYD	30	2.8

Table 2.7 Comparison of tunicamycin production yields

2.4 Proposed biosynthetic pathway to the tunicamycins

Thus far we have reported the identification and verification of the gene cluster responsible for tunicamycin biosynthesis in *S. chartreusis*. In this section we will present analysis of individual genes in *tun* operon together with an inspection of conserved active-site residues, using a combination of BLAST,[295] Artemis,[305] Conserved Domain Database[334] and HHpred[335] bioinformatic tools. This will culminate in the proposal of a propose a detailed metabolic pathway for the biosynthesis of the tunicamycins, reconciling insights gained through previous metabolic feeding experiments.[296] We will also comment on possible resistance mechanisms and how expression of this biosynthetic cluster might be regulated in *S. chartreusis*.

2.4.1 Carbohydrate building block biosynthesis

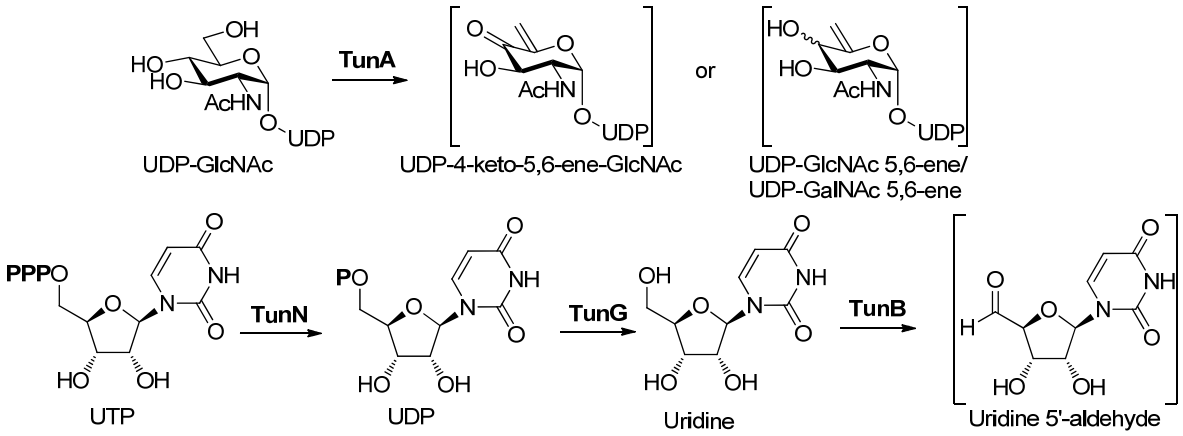


Fig. 2.24 Proposed role of Tun gene products in carbohydrate building block biosynthesis

The two suggested coupling partners involved in the formation of the first eleven-carbon dialdose intermediate on the way to tunicaminyl-uracil are uridine 5´-aldehyde and UDP-4-keto-5,6-ene-*N*-acetylglucosamine.[296] The closest homologues of the TunA gene product are NAD-dependent epimerase/dehydratases from various bacteria (42% identity and 58% similarity with one from *Streptomyces* sp. Mg1). The protein contains a TGxxGxxG signature at the N-terminus denoting the presence of a nucleotide-binding Rossmann fold,[336] as well as a $Sx_{24}Yx_3K$ catalytic triad.[337] These features are characteristic of short-chain dehydrogenase/reductase superfamily members, which includes the epimerase/dehydratase family of enzymes.[338] Examples sharing these motifs include FlaA1, a UDP-GlcNAc C_6-dehydratase/C_4-oxidoreductase from *Helicobacter pylori* producing UDP-6-deoxy-GlcNAc[339] and dTDP-glucose 4,6-dehydratase from *Streptomyces*

venezuelae yielding dTDP-6-deoxy-4-keto-glucose.[340] We propose that TunA converts UDP-GlcNAc – which has been shown to be a precursor in tunicamycin biosynthesis – to the UDP-4-keto-5,6-ene-GlcNAc intermediate involved in undecose construction, although reduction to the allylic alcohol is possible, resulting in UDP-5,6-ene-GlcNAc (Fig. 2.24). Recent work with TunA performed within our research group supports the latter proposal. This enzyme was found to act as a 5,6-dehydratase, accepting UDP-GlcNAc and converting it to UDP-5,6-ene-GlcNAc – which was successfully characterised by NMR. Crucially no 4-keto species was detected, nor was UDP-GalNAc accepted as a substrate.[341]

The suggested uridine 5´-aldehyde coupling partner has not only been implicated in tunicamycin biosynthesis,[296] but also in that of nikkomycin, polyoxin, liposidomycin and capuramycin nucleoside antibiotic families.[285,342-344] Its formation and mechanistic role has not been studied in detail, although in our system its formation is likely to begin with the action of TunN (Fig. 2.24). The closest homologues of the *tunN* gene encode NUDIX hydrolases from various bacteria (36% identity and 55% similarity with one from *Nakamurella multipartita* DSM 44233). It contains a 23-residue $Gx_5Ex_7RExxEExGV$ motif highly conserved amongst members of the NUDIX-hydrolase superfamily, forming a structural motif involved in Mg^{2+} or Mn^{2+} binding and subsequent catalysis.[345] Substrates of this family of enzymes have the general structure of a nucleotide diphosphate linked to another moiety X (NUDIX), and include nucleoside triphosphates, dinucleotide polyphosphates and nucleotide sugars. Enzymes acting on uridine 5´-triphosphate (UTP) have been identified.[346] In our case, TunN is likely to act as a pyrophosphatase on UTP to give uridine 5´-monophosphate (UMP), which acts as a substrate for further phosphate hydrolysis by TunG (Fig. 2.24). The amino acid sequence of TunG shows homology to phosphoglycerate mutases (29% identity and 47% similarity with one from *Frankia* sp. CcI3) and α-ribazole phosphatases from a number of sources (27% identity and 42% similarity with one from *Clostridium botulinum* Bf). α-Ribazole 5´-phosphate is an intermediate in cobalamin (vitamin B_{12}) biosynthesis, and its structure is closely related to that of UMP – our proposed substrate for TunG.[347,348] Analysis using the Conserved Domains Database (CDD)[334] showed TunG to belong to the phosphoglycerate mutase-like subgroup of the histidine phosphatase superfamily, to which both of the aforementioned enzyme classes belong.[349] The action of TunN and TunG on UTP is suggested to be similar to the actions of Pur7 and Pur3 in the biosynthesis of nucleoside antibiotic puromycin, where adenosine 5´-triphosphate is the initial substrate.[350,351]

The uridine which results from these transformations we suggest is next oxidised to uridine 5´-aldehyde by TunB (Fig. 2.24). All of the closest homologues of the gene product of *tunB* are radical SAM domain proteins (32% identity and 48% similarity with one from *Haloterrigena turkmenica* DSM 5511) with examples of homology to coenzyme pyrroloquinoline quinone (PQQ) synthesis protein E (*pqqE*) being the most common (33% identity and 51% similarity with one from *Mycobacterium abscessus* ATCC 19977). Members of the radical SAM superfamily all bind a 4Fe4S cluster and *S*-adenosyl-L-methionine (SAM); indeed TunB contains a characteristic CxxxCxxC motif for binding the iron-sulfur cluster.[352] Enzymes of this class generate an enzyme bound 5´-deoxyadenosyl radical from SAM, a powerful oxidant which catalyses a range of radical-mediated reactions at specific positions of bound substrates.[353] In our case, we propose a hydrogen is abstracted from C–5´ of uridine to generate a substrate radical, which could undergo further oxidation to uridine 5´-aldehyde or feature directly in the subsequent coupling with UDP-4-keto-5,6-ene-GlcNAc.

2.4.2 Tunicamine Biosynthesis

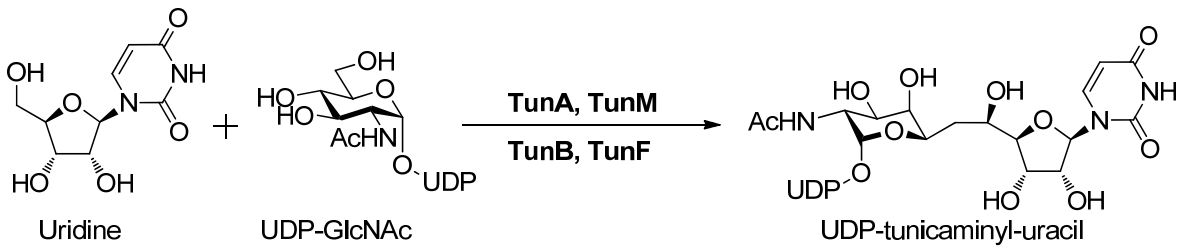


Fig. 2.25 Proposed role of *tun* gene products in forming the undecose scaffold of tunicaminyl-uracil

The undecose skeleton of tunicaminyl-uracil is proposed to be formed by the tail-to-tail coupling of uridine 5´-aldehyde and UDP-4-keto-5,6-ene-GlcNAc to install the necessary C–C bond.[296] We have suggested that these intermediates are the products of UTP and UDP-GlcNAc through the actions of TunN, TunG, TunB and TunA (Fig. 2.24). The gene product TunM shows greatest homology to a methyltransferase family protein from *Saccharomonospora viridis* DSM 43017 (48% identity and 63% similarity). The vast majority of homologues are from this family, an assignment supported by CDD analysis which shows TunM to belong to the *S*-adenosylmethionine dependent methyltransferase superfamily. Since methyl transfer is not envisaged during tunicamycin biosynthesis, the precise role of TunM remains uncertain. Although it contains a SAM-binding motif of ExGxGxG, it lacks any CxxxCxxC motif for the binding of 4Fe4S, thus explaining its

classification as a methyltransferase family protein rather than a radical SAM superfamily protein.[354] However, in the tunicamycin biosynthetic pathway it likely acts to mediate a radical reaction between the aforementioned coupling partners, *via* a 5´-deoxyadenosyl radical. One possibility is that TunB and TunM are closely associated and act together to bring both coupling partners in close proximity of an enzyme-bound reactive radical species, which is first created from the SAM cofactor of TunB by the action of its 4Fe4S cluster and then transferred to the SAM cofactor of its mediating partner TunM, where the C–C bond forming event takes place. The coupling of the two activated carbohydrate intermediates may proceed *via* a radical mechanism either by addition to an α,β-unsaturated ketone or through a Barbier-type mechanism. Clearly further work is necessary to obtain direct experimental evidence in support of this hypothesis, but there remains the exciting prospect of deciphering a unique and highly unusual biosynthetic mechanism, possibly uncovering novel chemistry unprecedented for an enzymatic reaction (Fig. 2.25).

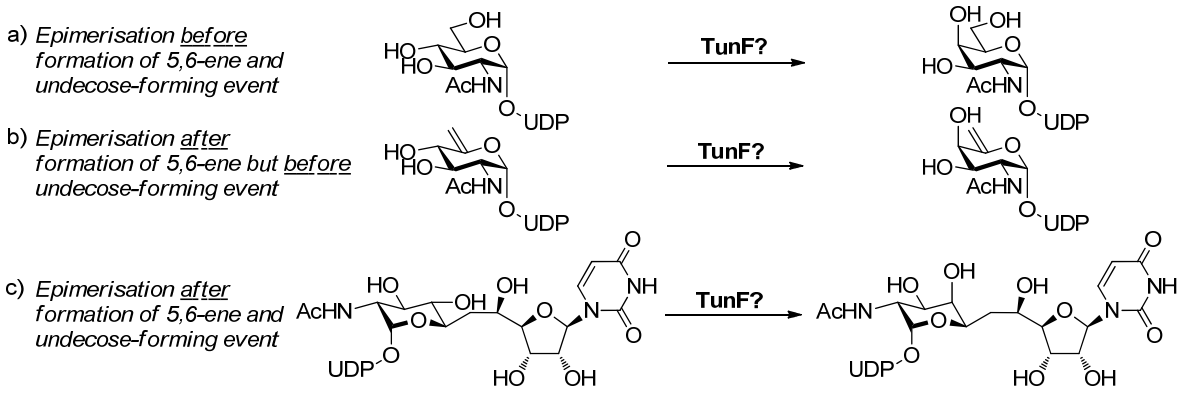


Fig. 2.26 Possible substrates for epimerase TunF

The product of gene *tunF* shows homology to a large number of UDP-glucose-4-epimerase proteins (46% identity and 63% similarity with one from *Paenibacillus* sp. oral taxon 786 str. D14). It contains an N-terminal Rossmann fold as denoted by CDD analysis and the presence of a characteristic TGxxGxxG nucleotide-binding motif. It also contains the $Sx_{24}Yx_3K$ catalytic triad.[337] Since TunF is not predicted to have 4,6-dehydratase activity, it may serve to function purely as a C-4 epimerase. The precise position of TunF in the biosynthetic pathway of tunicamycin is unclear – three potential substrates can be rationalised (Fig. 2.26). Firstly, TunF could act early in the pathway on UDP-GlcNAc, epimerising it to UDP-GalNAc (Fig. 2.26, entry a). Subsequent action by dehydratase TunA would erode the stereochemistry at C-4 *via* a UDP-4-keto-5,6-ene-GlcNAc intermediate, meaning TunA would have to mediate stereospecific reduction of the 4-keto group – which could occur before or after the C–C coupling event – yielding only *galacto*-

configured products. The second and third postulated substrates involve the action of TunF after that of TunA. After oxidising C-4 to a ketone and catalysing dehydration to a 5,6-ene, the hydride abstracted by TunA is returned to the substrate from NADH to complete the catalytic cycle. This hydride return may occur before or after the undecose-forming event postulated to be controlled by TunB and TunM. In the case of tunicamycin biosynthesis, reduction of the 4-keto moiety in this intermediate is more likely than that of the 5,6-ene – as observed in 6-deoxysugar biosynthesis[355] – since this alkene moiety is likely to play an important role in the subsequent undecose-forming event. Since the hydride at C-4 was initially abstracted from the axial position it may return from the same face, reducing the 4-keto functionality to an equatorial hydroxyl and yielding a *gluco*- configured product. It is at this point that TunF may catalyse an epimerisation to the *galacto*- conformer. Depending on the timing of the hydride return from NADH in TunA, this TunF-mediated epimerisation may occur before or after the C–C coupling event (Fig. 2.26, entries b and c). Recent studies in our research group revealed UDP-GlcNAc to be a substrate for TunA, forming UDP-5,6-ene-GlcNAc. UDP-GalNAc was not accepted as a substrate, suggesting entry a in Fig. 2.26 did not represent the true ordering of biosynthetic events within the tunicamycin pathway. Moreover, co-incubation of UDP-GlcNAc with TunA and TunF led to the detection of UDP-5,6,ene-GalNAc, suggesting UDP-5,6-ene-GlcNAc is accepted by epimerase TunF. This lends support to the sequence of events outlined in entry b of Fig. 2.26.[341]

2.4.3 Attachment of N-acetylglucosamine

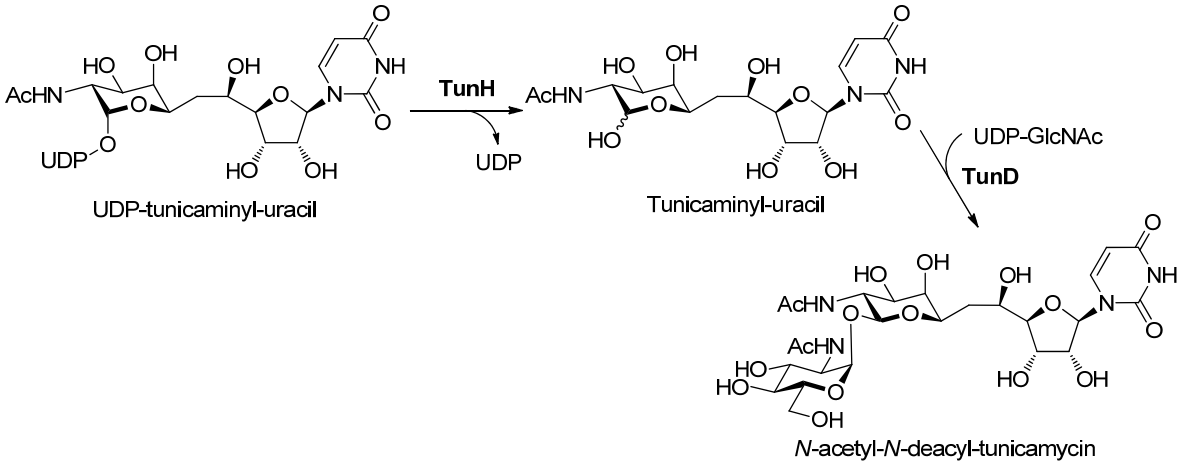


Fig. 2.27 Proposed role of Tun gene products in attachment of the α,β-1,1-linked GlcNAc moiety

The gene product of *tunH* shows homology to type I phosphodiesterases/nucleotide pyrophosphatases (35% identity and 50% similarity with one from *Burkholderia* sp. 383). CDD analysis reveals N- and C-terminal domains characteristic of the sulfatase superfamily, with particular homology to pfam01663 family members, namely Type I phosphodiesterases/nucleotide pyrophosphatases. The likely function of TunH is in the hydrolysis of UDP-tunicaminyl-uracil, releasing the anomeric 11´-OH for subsequent glycosylation with GlcNAc (Fig. 2.27). It is possible that the UDP liberated in this step is subsequently used as a source of uridine for early biosynthetic transformations described in section 2.4.1. Formation of such an unusual α,β-1,1-trehalose linkage is extremely rare and is most probably catalysed by the only putative glycosyltransferase present in the *tun* gene cluster TunD, using UDP-GlcNAc as the glycosyl donor. TunD is homologous to group 1 glycosyl transferases from various bacteria (26% identity and 42% similarity with one from *Magnetococcus* sp. MC-1). CDD analysis predicts TunD is a member of the GT-1 Yqgm-like family, named after glycosyltransferase Yqgm from *Bacillus lichenformis* about which little is known. It contains a motif present in GT-1 family glycosyltransferases (pfam00534) and is part of the larger glycosyltransferase GT-B superfamily possessing the GT-B fold. Through analysis of substrates and products it can be assigned as a retaining glycosyltransferase.[356]

2.4.4 Attachment of Fatty Acid

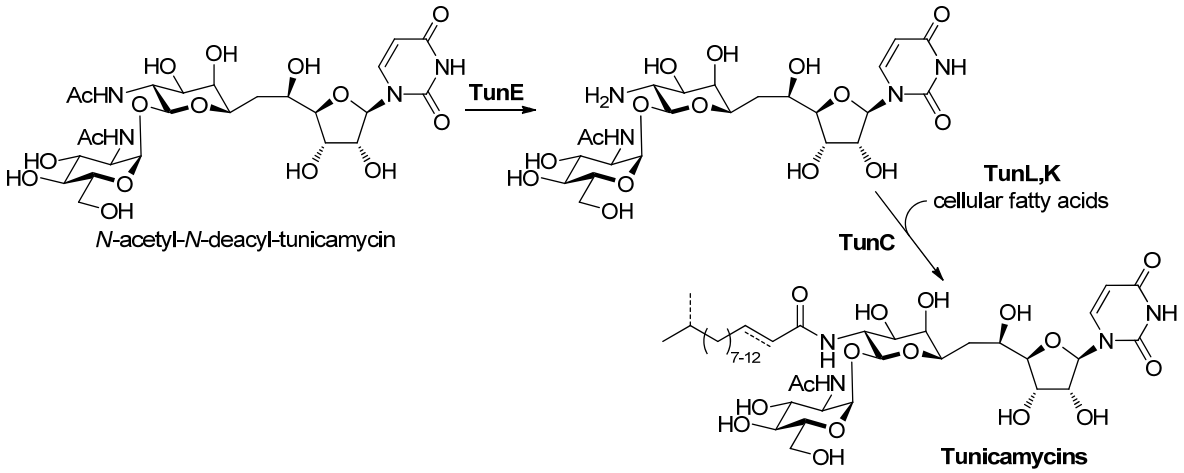


Fig. 2.28 Proposed role of Tun gene products in *N*-acylation of the tunicamycin core skeleton

TunE shows homology to LmbE family proteins from a variety of organisms (43% identity and 57% similarity with one from *Syntrophomonas wolfei* subsp. *wolfei* str. Goettingen). This family is part of the larger PIG-L superfamily and also contains *N*-acetyl-glucosaminyl-phosphatidylinositol de-*N*-acetylases from glycosyl-phosphatidyl-inositol (GPI) anchor biosynthesis[357] and 1-D-*myo*-inosityl-2-acetamido-2-deoxy-α-D-glucopyranoside deacetylases from mycothiol biosynthesis.[302] TunE contains an unusual HPDD zinc-binding motif conserved across both of the latter two enzyme families[358] and CDD analysis also places it in the PIG-L superfamily (COG2120). Based on the similarities between *N*-acetyl-*N*-deacyl-tunicamycin and the substrates of mycothiol and GPI-anchor biosynthesis, we propose that TunE catalyses Zn^{2+}-mediated cleavage of the 10´-acetamido group of the pseudotrisaccharide precursor to reveal the free amine, in preparation for subsequent fatty acid attachment.

Since heterologously produced tunicamycins in *S. coelicolor* were fully acylated even though there are no fatty acid synthase genes in the *tun* cluster, the range of fatty acids observed amongst tunicamycin homologues are likely to be derived from the cellular pool of fatty acids; this has previously been observed in teicoplanin biosynthesis.[359] The gene product of *tunL* shows homology to type 2 phosphatidic acid phosphatases (PAP2) from various bacteria (33% identity and 42% similarity with one from *Micromonospora aurantiaca* ATCC 27029). CDD analysis identifies it as a member of the PAP2-like protein subfamily specific to bacteria (cd03392) and its sequence contains a three-part catalytic motif characteristic of this Mg^{2+}-independent phosphatase family – KxxxxxxRP (domain 1), PSGH (domain 2) and SRxxxxxHxxxD (domain 3).[360] A possible function of

TunL is in the regulation of lipid synthesis in the producing bacterium. By down-regulating the levels of cellular phosphatidic acid and up-regulating levels of its cleavage product diacylglycerol, phospholipid biosynthesis is repressed and cellular pools of fatty acids can be diverted for use in tunicamycin biosynthesis *via* β-oxidative degradation pathways.[361] It is fascinating that tunicamycin-producing organisms appear to have evolved an efficient way of perturbing the complex and subtle signalling pathways of lipid metabolism regulation, allowing increased tunicamycin biosynthesis without negatively affecting vital cellular processes.

Intriguingly, the *tun* cluster contains an acyl carrier protein (ACP) in TunK – which exhibits homology to ACPs from numerous bacterial sources (32% identity and 61% similarity with one from *Catenulispora acidiphila* DSM 44928) – but does not encode a fatty acyl ACP ligase. Presumably, fatty acids are activated as ACP-thioesters for use as substrates in the subsequent *N*-acylation of the tunicamycin skeleton through the action of existing fatty acyl ACP ligases from primary metabolism.

The gene product of *tunC* shows homology to GCN5-related *N*-acetyltransferases (GNAT) from a variety of bacteria (33% identity and 50% similarity with one from *Fervidobacterium nodosum* Rt17-B1). Analysis by HHPred shows homology to mycothiol synthase MshD across the full length of TunC (PDB 1P0H), but suggests the presence of two separate GNAT domains making up the N-terminal and C-terminal halves of the protein respectively.[335] Indeed, MshB has been shown to similarly contain two GNAT domains, one of which appears to be catalytically non-functional.[362] We propose that TunC catalyses the acylation of the tunicamycin core skeleton 10´-NH_2, using fatty acids sequestered from primary metabolism and activated as ACP-thioesters. There is, however, a possibility that TunC uses fatty acyl-coenzyme A (acyl-CoA) substrates directly from the β-oxidative degradation pathway as substrates and first catalyses the *S*-acyl transfer from CoA to an associated ACP (in this case TunK), before also catalysing subsequent *N*-acyl transfer to tunicamycin. This activity has been observed in one of the polyketide synthase modules of curacin A,[363] although further experimental evidence is clearly required to distinguish between these proposals.

2.4.5 Resistance genes

The gene products of *tunI* and *tunJ* comprise components of the ATP-binding cassette (ABC) transporter systems from a variety of bacteria. TunI is homologous to ABC transporter ATP-binding subunits (41% identity and 61% similarity with one from *Streptomyces scabiei* 87.22) and TunJ is homologous to ABC-2 transporter permeases (32% identity and 51% similarity with one from *Thermobaculum terrenum* ATCC BAA-798). Such transporter systems are well studied and have been observed in many bacteria producing antibiotics, functioning as a self-resistance mechanism by coupling ATP hydrolysis to metabolite efflux through the cell membrane.[364,365] As such, TunI and TunJ together probably provide the self-resistance mechanism necessary for the viability of tunicamycin-producing organisms.

2.4.6 Regulation of expression

No regulatory genes were found in the tun gene cluster, suggesting that tunicamycin production may be subject to global control associated with growth rate reduction. The presence of rare TTA leucine codons (only 2% of *S. coelicolor* genes contain a TTA codon) in *tunA* and *tunM* may well reflect an element of translational regulation. In *S. coelicolor*, the accumulation of $_{Leu}tRNA^{UUA}$ is temporally regulated, and translation of mRNAs containing this codon may be largely confined to later stages of growth.[366,367]

2.4.7 Proposed metabolic pathway for tunicamycin biosynthesis

Over the previous sections, the biosynthesis of tunicamycin has been broken down into different stages and rationalised biosynthetic proposals have been made based on detailed bioinformatic analysis of the enzymes encoded by the *tun* gene cluster. The proposals from these different stages are now brought together to form a unified and detailed putative biosynthetic pathway to the tunicamycins from metabolic precursors (Fig. 2.29). It should be noted that alternative possibilities exist to the intermediates shown in brackets and further work is required to decipher their precise structures. For further discussion of these alternatives please refer to the relevant preceding sections (sections 2.4.1 and 2.4.2).

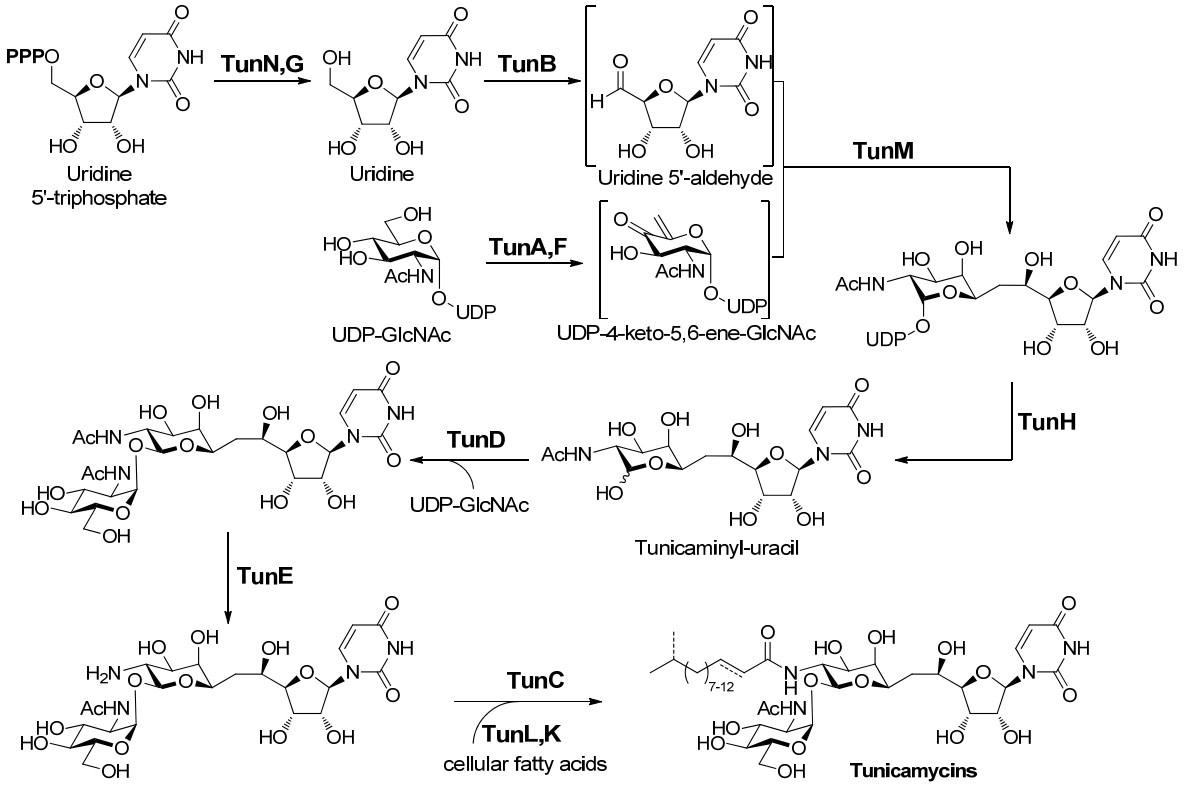


Fig. 2.29 Proposed biosynthetic pathway for the tunicamycins

The *tun* gene cluster described is relatively small in size, although a previous suggestion that as few as five genes would be necessary for the biosynthesis of tunicamycin has proved too conservative.[368] Of the nine additional genes not originally predicted, two are involved in the generation of free uridine from UTP, contrary to suggestions that uridine would be obtained directly from primary metabolism.[296] Two further genes are implicated in formation of UDP-tunicaminyl-uracil – one coding for a sugar epimerase supplementary to the dehydratase catalysing UDP-4-keto-5,6-ene-GlcNAc formation and one which mediates the radical coupling event alongside the gene responsible for uridine oxidation. Hydrolysis of UDP from the undecose intermediate has also been shown to require enzyme catalysis. Although the acyl side chains are likely to originate from cellular pools of fatty acids – consistent with the lack of a fatty acid synthase – the *tun* gene cluster still encodes two enzymes that provide sufficient fatty acid flux and are involved in sequestering lipids and processing them prior to attachment. Finally, the last two additional *tun* genes are not directly involved in tunicamycin biosynthesis, but are likely to be crucial in conferring self-resistance to the producing organism. *tunI* and *tunJ* together encode for an ABC transporter, homologues of which are responsible for rapid ATP-driven efflux of antibiotics from cells in a large number of antibiotic-producing organisms.[365]

2.5 Conclusions

In this chapter we have successfully identified and verified the biosynthetic genes of a tunicamycin-family antibiotic for the first time, offering rich insights into the poorly understood biosynthetic pathway of this fascinating family of nucleoside antibiotics. Through molecular cloning and heterologous expression of this *tun* gene cluster in a *S. coelicolor* host, we have precisely defined the minimal set of genes required for tunicamycin production. Additionally, we have identified close homologues of the *tun* cluster in *A. mirum* DSM43827 and *S. clavuligerus* ATCC27064. The latter organism is known to produce MM19290[315] – an antibiotic closely related to tunicamycin – and based on the close similarity of its homologous cluster with the *tun* genes, we claim this cluster must be responsible for MM19290 biosynthesis in *S. clavuligerus* ATCC27064. Furthermore, we propose MM19290 shares the core structure of the tunicamycins and differs only in the nature of its acyl side chains.

The genome scanning approach to identifying the *tun* gene cluster described here represents a rapid and efficient way of identifying natural product gene clusters, without the need for labourious biochemical probes and avoiding frustrating false-positive results. It utilised rationalised predictions of the chemical reactivity that is likely to be involved in the construction of tunicamycin from metabolic precursors, acting as a 'logic filter' that was applied during the genome mining process. The exponential increase in publically available gene sequences in recent years has dramatically expanded the possibilities afforded by bioinformatic analysis and the work presented herein highlights *in silico* mining of partially assembled genome sequences as an increasingly effective tool during the early stages of dissecting a bacterial biosynthetic pathway.

The findings presented here act as a gateway, allowing more detailed investigations into the tunicamycin biosynthetic pathway to take place. Functional characterisation of individual enzymes will provide confirmation of their proposed functions and ordering within the metabolic pathway and will allow further insights into how some of the unique linkages in tunicamycin are constructed.

2.6 Experimental

2.6.1 Bacterial strains

2.6.1.1 Streptomyces

Strain	Description	Reference/Source
S. chartreusis NRRL 3882	Tunicamycin producing strain	DSMZ culture collection
S. lysosuperificus ATCC 31396	Tunicamycin producing strain	ATCC culture collection
S. coelicolor M1146	*S. coelicolor* M145[330] with deleted actinorhodin (*act*), undecylprodigiosin (*red*), calcium-dependent antibiotic (*cda)* and type I PKS cluster CPK (*cpk*) gene clusters	Juan-Pablo Gomez-Escribano (JIC) unpublished data
S. coelicolor M1152	*S. coelicolor* M1146 carrying a point mutation in *rpsL* for enhanced expression of antibiotic gene clusters	Juan-Pablo Gomez-Escribano (JIC) unpublished data
S. coelicolor M1021	*S. coelicolor* M1146 carrying pIJ12315	This study
S. coelicolor M1022	*S. coelicolor* M1146 carrying pIJ12316	This study
S. coelicolor M1023	*S. coelicolor* M1146 carrying pIJ12317	This study
S. coelicolor M1024	*S. coelicolor* M1146 carrying pIJ12318	This study
S. coelicolor M1025	*S. coelicolor* M1152 carrying pIJ12315	This study
S. coelicolor M1026	*S. coelicolor* M1152 carrying pIJ12316	This study
S. coelicolor M1027	*S. coelicolor* M1152 carrying pIJ12317	This study
S. coelicolor M1028	*S. coelicolor* M1152 carrying pIJ12318	This study
S. coelicolor M1030	*S. coelicolor* M1152 carrying pMJCOS1 (vector-only control)	This study
S. coelicolor M1031	*S. coelicolor* M1146 carrying pIJ12003a	This study
S. coelicolor M1035	*S. coelicolor* M1146 carrying pRT802 (vector-only control)	This study

Table 2.8 Details of *Streptomyces* strains used in this study

2.6.1.2 Escherichia coli and other bacteria

Strain	Description	Reference/Source
E. coli XL1-Blue MR	Restriction-deficient cloning strain for cosmid-based cloning. Used for construction of genomic library	Stratagene
E. coli DH5α	Cloning strain for routine subcloning and high-quality plasmid preparations	Invitrogen
E. coli BW25113	Host strain for λ Red recombination	[324]
E. coli ET12567	Methylation-deficient strain used for conjugation with *Streptomyces*	[327]
E. coli S17-1	Donor strain for conjugative transfer of mobilisable plasmids to *Streptomyces*	[332]
Bacillus subtilis EC1524	Tunicamycin-sensitive reporter strain	JIC, Norwich
Micrococcus luteus	Commonly used reporter strain found not to be sensitive to tunicamycin	JIC, Norwich

Table 2.9 Details of *E. coli* strains and other bacteria used in this study

2.6.2 Plasmids

Name	Description	Reference/Source
SuperCos I	Cosmid vector used in *S. chartreusis* genomic library	Stratagene
pMJCOS1	Conjugative derivative of Supercos1 that integrates in a single copy at the φC31 attachment site in the *Streptomyces* host	[319]
pRT802	Conjugative vector that integrates in a single copy at the ϕBT1 attachment site in the Streptomyces host	[331]
pIJ790	λ-Red recombination helper plasmid	[328]
pUZ8002	Non-transmissible *oriT*⁻ helper plasmid containing transfer functions for intergenic conjugation	[326]
pIJ12311	Cosmid library clone 4H8 with SuperCos I backbone	This study
pIJ12312	Cosmid library clone 5K7 with SuperCos I backbone	This study
pIJ12313	Cosmid library clone 6N9 with SuperCos I backbone	This study
pIJ12314	Cosmid library clone 7C3 with SuperCos I backbone	This study
pIJ12315	pMJCOS1 packaged with a *S. chartreusis* genomic DNA fragment from cosmid library clone 4H8	This study
pIJ12316	pMJCOS1 packaged with a *S. chartreusis* genomic DNA fragment from cosmid library clone 5K7	This study
pIJ12317	pMJCOS1 packaged with a *S. chartreusis* genomic DNA fragment from cosmid library clone 6N9	This study
pIJ12318	pMJCOS1 packaged with a *S. chartreusis* genomic DNA fragment from cosmid library clone 7C3	This study

Table 2.10 Details of plasmids used in this study

2.6.3 PCR primers

All DNA oligonucleotides were purchased from Sigma-Aldrich.

Name	Sequence	Use
TunLF	5´-TCACCTACCCCTACGACGAC	To verify the upstream boundary of putative *tun* gene cluster
TunLR	5´-TTGAAGGAGCGAGGAACACT	
TunRF	5´-CGTCCCTGTGAACGCTATTT	To verify the downstream boundary of putative *tun* gene cluster
TunRR	5´-GGTGGGAACTTGAAGAACGA	
TunDF	5´-ACCGCTACATGCACGACAT	To span the gap between two contigs containing putative *tun* gene cluster
TunDR	5´-CTCGTCCTCGACGAACTCC	
TunA:20-40	5´-AACACCGGATTCTCATCACC	To generate a hybridisation probe for *tunA*
TunA:240-260	5´-TTGAAGGAGCGAGGAACACT	
TunN probe for	5´-GGTGGGTGTCTACCAGCATC	To generate a hybridisation probe for *tunN*
TunN prove rev	5´-GTGGCCTCGAAGAGTGAGAG	

Table 2.11 Details of PCR primers used in this study

2.6.4 Culture media, buffers, solutions, reagents and enzymes

2.6.4.1 Media

Unless otherwise stated, the recipes for all media can be found in Kieser *et al*.[289] Additional recipes are listed below:

sYEME
(Super yeast extract-malt extract medium)

Difco yeast extract	3 g
Difco bacto-peptone	5 g
Oxoid malt extract	3 g
Glucose	10 g
Sucrose	340 g (34% final)
Deionised water	to 1 L

After autoclaving, the following sterile supplements were added:

$MgCl_2.6H_2O$ (2.5 M)	2 mL (5 mM final)
Glycine (20%)	25 mL (0.5% final)

MYM agar

Maltose	4 g
Yeast Extract	4 g
Malt Extract	10 g
Agar	20 g
Deionised/tap water (1:1)	to 1L

After autoclaving add:

R2 Trace Elements solution	2 mL

OB agar

Porridge Oats (Tesco, UK)	40 g
Agar	20 g
Tap water	to 1 L

Autoclave twice

2.6.4.2 Buffers and solutions

Unless otherwise stated, the recipes for all buffers and solutions can be found in Kieser *et al*.[289] and Sambrook *et al*.[318]

Hybridisation buffer	Hybridisation wash buffer
6× SSC	0.1× SSC
1× Denhard's solution	1% SDS
0.5% SDS	
50 µg/mL denatured calf thymus DNA	

2.6.4.3 Sources of chemicals, enzymes and miscellaneous materials

Unless otherwise stated, all chemicals, enzymes and other reagents were purchased from the suppliers listed below. Chemical reagents were purchased from Sigma-Aldrich and used without further purification. Media components were purchased from Sigma-Aldrich or BD Biosciences. Restriction endonucleases were purchased from New England Biolabs and remaining enzymes from Invitrogen or Sigma-Aldrich.

2.6.5 Propagation of bacterial strains

2.6.5.1 Streptomyces strains

Streptomyces strains were propagated on solid media as described by Kieser *et al*.[289] To collect spores, agar plates were inoculated with *Streptomyces* spores, streaked out and grown at 30 °C. A single sporulating colony was resuspended in 40 µL distilled water placed in the middle of a new agar plate, and the spores spread evenly over the plate. The plates were incubated at 30 °C for 6-10 days, until confluent lawns of spores were visible. Plates were left for no more than two weeks to prevent significant loss of spore viability. Spores were harvested as described by Kieser *et al*.[289] and stored in 20% (v/v) glycerol at -20 °C. All *S. coelicolor* strains were propagated on MS agar, and typically produced yellow spores. *S. chartreusis* NRRL3882 was propagated on OB agar and *S. lysosuperificus* ATCC31396 on MYM agar, producing green and red spores respectively.

Unless otherwise stated, *Streptomyces* strains were grown in liquid culture as described by Kieser *et al*.[289]

2.6.5.2 *Escherichia coli strains*

E. coli strains were routinely grown at 37 °C in Luria-Bertani broth (LB) or on 1.5% LB agar plates supplemented with appropriate antibiotics, as described by Sambrook *et al*.[318] For recombinant strain selection, antibiotics were used in the following concentrations: carbenicillin (100 µg/mL), kanamycin (50 µg/mL), apramycin (50 µg/mL), chloramphenicol (25 µg/mL) or nalidixic acid (25 µg/mL). Glycerol stocks were prepared by collecting a cell pellet from a sample of fresh overnight culture (1 mL), resuspending it in 20% (v/v) glycerol (800 µL) and storing at -80 °C. New stocks were prepared from single colonies every 4-6 months.

2.6.5.3 *Other bacterial strains*

B. subtilis EC1524 and *M. luteus* were grown in LB or on 1.5% LB agar plates at 30 °C.

2.6.6 Standard protocols for the manipulation of DNA

2.6.6.1 *Pulsed-field gel electrophoresis*

Pulsed-field gel electrophoresis was performed as described by Kieser *et al*[289] using a Bio-Rad CHEF-DR III system. Gels contained 1% agarose in 1 × TBE buffer and electrophoresis was performed at 6 V/cm, 15 °C for 18 h, with switch times ramped from 1-25 s and using 1 × TBE buffer. Midrange PFG Marker I (New England Biolabs) was used as an internal DNA marker.

2.6.6.2 *Polymerase chain reaction (PCR)*

The reaction mixture for PCR contained *Taq* DNA polymerase (1 µL, Invitrogen), 10× PCR reaction buffer (5 µL, Invitrogen), 5 mM dNTP mix (2 µL), 40% DMSO (6 µL), 100 pg of each primer (1 µL of 100 pg/µL stock) and 100 ng template DNA (1 µL of ~100 ng/µL stock) made up to a total volume of 50 µL with sterilised deionised water. When using genomic DNA as template, 200-300 ng was used. Reactions were performed on a Bio-Rad DNA Engine Peltier Thermal Cycler with cycling parameters as described in Table 2.12. PCR products were analysed by agarose gel electrophoresis (0.7 – 1 % w/v, depending on fragment size).

Segment	No. cycles	Temperature	Duration
1	1	95 ℃	4 min
2	26	95 ℃ 55 ℃ 72 ℃	45 s 45 s 45 s
3	1	72 ℃	7 min

Table 2.12 PCR cycling parameters

2.6.6.3 Isolation of plasmid DNA

Plasmid DNA from *E. coli* was typically isolated using a QIAprep Miniprep Kit (Qiagen). The isolation of cosmid DNA from *E. coli* was performed as follows: relevant *E.coli* strains were grown overnight in LB (10 mL) at 37 °C. Cells were collected by centrifugation and resuspended by vortexing in solution I (200 µL of 50 mM Tris/HCl, pH 8, 10 mM EDTA). Solution II (400 µl of 200 mM NaOH, 1% SDS) was immediately added and the tubes mixed by inversion (10 ×). Solution III (300 µL, 3 M potassium acetate, pH 5.5) was then added and the tubes inverted (10 ×). Following centrifugation at room temperature (14,000 × *g*, 5 min), the supernatant was extracted with phenol/chloroform (600 µL), vortexed for 2 min and further centrifuged (14,000 × *g*, 5 min). The upper phase was then transferred to a fresh tube containing 750 µL isopropanol and left on ice for 20 min. Following centrifugation (14,000 × *g*, 5 min), the resulting pellet was washed with 70% EtOH (200 µL) and centrifuged once again (14,000 × *g*, 5 min). The supernatant was removed with a pipette and the pellet air dried for 5 min at room temperature before being resuspended in 10 mM Tris/HCl (60 µL, pH 8).

2.6.7 Isolation of genomic DNA from tunicamycin-producing Streptomyces

2.6.7.1 Isolation of total DNA from S. chartreusis NRRL 3882

Total DNA was isolated using procedures described by Kieser *et al*, suitably optimised for *S. chartreusis* NRRL 3882.[289] 50 mL of 1:1 TSB/sYEME in a conical flask was inoculated with 10 µL of a dense *S. chartreusis* NRRL 3882 spore suspension and grown for 24 h at 30 °C, 250 rpm. Cells were harvested by centrifugation for 10 min at 3000 rpm and then washed twice with 10.3 % sucrose. Half the resultant cell pellet was resuspended in 5 mL SET buffer. 100 µL lysozyme (50 mg/mL) and 90 µL mutanolysin (1 mg/mL in 0.1M K_2PO_4 buffer pH7) were added and the suspension incubated at 37 °C for

2 h. 280 µL proteinase K (10 mg/mL) and 600 µL 10% SDS were added, the suspension was then mixed gently by inversion and incubated at 55 °C for 2 h yielding an opaque viscous solution. From this stage, to ensure a high Mw product, any mixing of the DNA was done by gentle inversion to prevent shearing. Similarly only wide bore tips were used when pipetting solutions containing DNA. 5M NaCl (2 mL) was added and the solution allowed to cool to room temperature. 5 mL 25:24:1 phenol/chloroform/isoamyl alcohol was added and the solution mixed for 30 min, before centrifugation for 10 min at 8000 rpm. As much upper aqueous phase as possible was removed to a fresh tube and 0.6 volumes of ice-cold isopropanol were added. After 15 min at 4 °C, white strings of DNA appeared, which were spooled onto a sealed glass Pasteur pipette. The spooled DNA was rinsed with 70% EtOH, air dried and resuspended in 0.5-1 mL TE buffer (depending on the yield of DNA) overnight at 4 °C.

2.6.7.2 Isolation of total DNA from S. lysosuperificus ATCC 31396

Total DNA was isolated using procedures described by Kieser *et al*, suitably optimised for *S. lysosuperificus* ATCC 31396.[289] 10 mL of 1:1 TSB/sYEME in a conical flask was inoculated with 10 µL of a dense *S. chartreusis* NRRL 3882 spore suspension and grown for 24 h at 30 °C, 250 rpm. 5 mL of cell culture was added to 10 mL of RNA protect solution (Qiagen), vortexed and left to stand for 15 min. Cells were then harvested by centrifugation for 10 min at 3000 rpm. The resultant cell pellet was resuspended in 5 mL SET buffer and 100 µL lysozyme (50 mg/mL) and 90 µL mutanolysin (1 mg/mL in 0.1M K_2PO_4 buffer pH7) were added. The resulting suspension incubated at 37 °C for 90 min. 280 µL proteinase K (10 mg/mL) and 600 µL 10% SDS were added, after which the suspension was mixed gently by inversion and incubated at 55 °C for 2 h to give an opaque viscous solution. From this stage onwards, any mixing of the DNA was done by gentle inversion to ensure a high Mw product. Similarly only wide bore tips were used when pipetting solutions containing DNA. 2 mL 5M NaCl was added and the solution allowed to cool to room temperature. 0.65 mL of sterile 10% hexadecyltrimethylammonium bromide solution (in 0.7M aq. NaCl) was added, mixed, and incubated at 55 °C for 15 min. After cooling to room temperature, 5 mL 25:24:1 phenol/chloroform/isoamyl alcohol was added and the solution mixed gently for 30 min, before centrifugation for 10 min at 4000 rpm. As much of the upper aqueous phase as possible was removed, to a fresh tube, and 0.6 volumes of ice-cold isopropanol were added. After 15 min at 4 °C white strings of DNA appeared, which were spooled onto a sealed glass Pasteur pipette. The spooled DNA was rinsed with

70% EtOH, air dried and resuspended in 0.5-1 mL TE (depending on the yield of DNA) overnight at 4 °C.

2.6.8 Identification of putative tunicamycin biosynthetic gene cluster

2.6.8.1 Whole genome sequencing of S. chartreusis NRRL 3882

High Mw genomic DNA from *S. chartreusis* NRRL 3882 was sequenced and assembled by the University of Liverpool Advanced Genomics Facility using a Roche 454 Titanium pyrosequencing platform (½ plate run) and the Roche Newbler (v2.0.00.20) assembler software.

2.6.8.2 Genome mining of S. chartreusis NRRL 3882

A BLAST database of *S. chartreusis* contigs was constructed by Govind Chandra (JIC). Subsequent sequence data analysis and annotation was performed with Artemis v12.0 software[305]. Nucleotide and amino acid sequence homology searches were performed using BLAST search tools,[295] the Conserved Domains Database (CDD)[334] and HHPred.[335]

2.6.8.3 PCR verification of contigs spanning the tun operon

The sequences of PCR primers are listed in Table 2.11. To validate the upstream and downstream boundaries of the putative *tun* cluster, primer pairs TunLF/TunLR and TunRF/TunRR were used. The connectivity between contig 00341 and 02796 was confirmed by PCR with primers TunDF and TunDR. *S. chartreusis* genomic DNA was used as template in each case. DNA sequencing of the PCR products was performed using an Applied Biosystems BigDye Terminator v3.1 Cycle Sequencing Kit and an AbiPrism 3730XL DNA Analyzer.

2.6.9 Construction of S. chartreusis cosmid library

2.6.9.1 Test scale partial Sau3A I digest of S. chartreusis genomic DNA

To a microcentrifuge tube was added high Mw genomic DNA from *S. chartreusis* (30 µg), 10× NEBuffer 1 (32 µL) and sterilised water to a final volume of 320 µL. The solution was split into two tubes of 150 and 170 µL respectively and these were pre-incubated at 37 °C for 5 min. To the 150 µL tube was added 0.025 U of *Sau*3A I (0.5 µL of

5×10^{-2} U/µL stock solution) and to the 170 µL tube was added 0.0025 U of *Sau*3A I (0.5 µL of 5×10^{-3} U/µL stock solution). The tubes were mixed by pipetting with a wide bore tip and incubated at 37 °C. 20 µL aliquots were removed at the following time points, added to tubes containing 0.5M EDTA (10 µL) and immediately frozen in liquid N_2: a) tube with 0.025U Sau3A I – 0 , 0.5, 1, 2, 3, 5 and 10 min; b) tube with 0.0025U Sau3A I – 0, 0.5, 1, 2, 3, 5, 10 and 20 min. Once thawed, 3.3 µL of 10× DNA gel loading buffer was added to each sample, and the results analysed by: a) electrophoresis in 1× TBE on a 0.3 % (w/v) agarose gel containing 4 µL ethidium bromide and b) pulsed-field gel electrophoresis.

2.6.9.2 Partial Sau3A I digest of S. chartreusis genomic DNA

To a microcentrifuge tube was added high Mw genomic DNA from *S. chartreusis* (100 µg), 10× NEBuffer 1 (100 µL) and sterilised water to a final volume of 1 mL. The tube was pre-incubated at 37 °C for 5 min, then 0.08 U of *Sau*3A I (1.6 µL of 5×10^{-2} U/µL stock solution) added and the solution mixed and incubated at 37 °C. After 90 s, 0.5 M EDTA was added (15 µL) and the solution mixed by pipetting with a wide bore tip. A 10 µL aliquot was taken and both tubes were then frozen in liquid N_2. The 10 µL sample was thawed and analysed by pulsed-field gel electrophoresis to confirm its size distribution. The main sample was thawed then 3M NaOAc buffer (100 µL, pH 6) and EtOH (2.2 mL) were added. The solution was stored for 16 h at -20 °C, after which the DNA was harvested by centrifugation (8000 rpm, 10 min). The supernatant was discarded and the white pellet was washed with 70% EtOH, air dried and dissolved in TE buffer (50 µL).

10× DNA loading buffer (10 µL) was added to this solution, and the sample separated by electrophoresis using a 0.3 % agarose gel containing 4 µL ethidium bromide. DNA markers used were: Midrange PFG Marker I (New England Biolabs) and 0.1× λ EcoR1/HindIII Marker (CamBio). Bands were visualised using long wavelength UV light (310 nm) to minimise nicking of DNA molecules. A gel slice was cut which contained DNA >20 kb and <100 kb.

Spectra/por 3 membrane dialysis tubing (MWCO 3500, Spectrum) was boiled in distilled water for 10 min in a microwave and left soaking in water overnight. The gel slice was placed inside and the tubing filled with 0.1× TBE (~1 mL). The DNA was electroeluted at 100V for 2 h using 0.1× TBE. The polarity was reversed for 30 s to dislodge the DNA from the dialysis tubing wall. The buffer contained within the dialysis tubing was added to 500 µL chloroform and the resultant solution was mixed by inversion for 5 min, separated by centrifugation and the upper aqueous layer collected. 3M NaOAc

(100 µL, pH 6) buffer and EtOH (2.2 mL) were added and the solution stored for 16 h at -20 °C. The DNA was harvested by centrifugation (8000 rpm, 10 min). The white pellet was washed with 70% EtOH, air dried and then dissolved in TE buffer (50 µL).

2.6.9.3 Ligation, packaging and transduction of S. chartreusis DNA fragments

The size-fractionated, *Sau*3A I-digested *S. chartreusis* genomic DNA was cloned into SuperCos I vectors using a SuperCos I Cosmid Vector Kit (Stratagene) according to the manufacturer's protocol. Using a Gigapack III Gold Packaging Extract Kit (Stratagene) and following the manufacturer's instructions, these cosmids were then packaged into recombinant λ-phage which were used to infect *E. coli* XL1-Blue MR cells. A total of 3072 recombinant *E. coli* colonies were individually picked by robot and arrayed onto 3 × 384-well microtiter plates at the Genome Analysis Centre (Norwich, UK). DNA from this library was fixed onto nylon membranes according to standard procedures.[318]

2.6.10 Hybridisation screening of S. chartreusis cosmid library

2.6.10.1 Synthesis of ^{32}P-labelled PCR probes

Internal fragments of genes *tunA* and *tunN* were amplified from *S. chartreusis* genomic DNA as template using primer pairs TunA:20-40/TunA:240-260 and TunN probe for/TunN probe rev. The resulting PCR products were purified using a QIAQuick PCR Purification Kit (Qiagen) and labelled with [α-^{32}P]-dCTP using the Rediprime II DNA Labelling System (Amersham). When ready for use in hybridisation experiments, the probes were denatured at 95 °C for 5 mins and cooled rapidly on ice.

2.6.10.2 Hybridisation to cosmid library

Two pre-baked nylon filters arrayed with *S. chartreusis* genomic library DNA were soaked for >2 h at 42 °C in 5× SSC with 0.5% SDS and 1 mM EDTA pH8, before cell debris was scraped from them with a moistened paper towel. The filters were then rinsed with 2× SSC (2 × 10 min) and either hybridised immediately or stored short-term at 4 °C in fresh 2× SSC. The freshly-washed filters were pre-hybridised in a Techne Hybridiser HB-1 oven at 60 °C inside two hybridisation tubes loaded with hybridisation buffer (50 mL). After 4 h, pre-hybridisation solutions were replaced with fresh hybridisation buffer (20 mL) and ^{32}P-labelled TunA probe (30 µL) was added to one tube and ^{32}P-labelled TunN probe (30 µL) to the other. Following incubation at 60 °C for 18 h, the hybridisation solution was

replaced with hybridisation wash buffer (50 mL) and further incubated at 60 °C for 30 min. After two such wash steps, the filters were wrapped in clingfilm and exposed to a phosphoimager plate (Fujifilm) for 5 h. The plates were visualised in a Fujifilm FLA-7000 bio-imaging analyzer and later blanked for re-use in a Fujix BAS-1000. Hybridised filters were additionally visualised by exposure to X-Ray film (4 h) and development in a Konica Minoltra SRX-101A developer. Visualised hybridisation images were overlaid with a 386-grid array and the location of hybridising cosmids within the *S. chartreusis* genomic library located using the grid-square key shown in Fig. 2.13.

2.6.10.3 PCR verification of isolated cosmids

Eight cosmids hybridising to both probes TunA and TunN were isolated from the *S. chartreusis* genomic library. Together with primer pairs TunA:20-40/TunA:240-260 and TunN probe for/TunN probe rev, they were used as DNA templates in a series of PCR reactions to confirm the result of the hybridisation screen and to confirm that all eight cosmids had been correctly isolated from the arrayed *S. chartreusis* genomic library.

2.6.10.4 Restriction digest of isolated cosmids

Samples of the eight cosmids isolated from genomic library clones were digested with either *Bam*H I or *Xho*I. Each reaction mixtures contained either *Bam*H I or *Xho*I (1 µL), one of the eight cosmid preparations (8 µL, derived from library clones 1N9, 2F8, 3D5, 4H8, 4N14, 5K7, 6N9, 7C3), 10× NEBuffer 2 (2 µL), RNAse I (0.2 µL) and sterilised water (10 µL). Restriction digests were visualised by agarose gel electrophoresis.

2.6.11 Cloning tun genes into S. coelicolor via pMJCOS1

2.6.11.1 λ Red-mediated recombination of SuperCos1 backbone

The SuperCos1 backbone of genomic library-derived cosmids 4H8, 5K7, 6N9 and 7C3 were converted to that of conjugative and integrative vector pMJCOS1 by λ Red-mediated recombination according to published procedures.[328] Bacterial strains and plasmids involved in these protocols are listed in Table 2.9 and Table 2.10 respectively. The DNA fragment used for λ Red-mediated recombination was 5.2 kb *Ssp*I fragment from pMJCOS1. Following antibiotic selection for the recombination event, total cosmid preparations were produced for each of the four recombined cosmids. Agar plates containing *E. coli* BW25113/pIJ790/pIJ12315-8 recombinants were flooded with water

(1 mL), then the dislodged cells were transferred to a microcentrifuge tube and collected by centrifugation. Cosmids were isolated from the resulting cell pellet as previously described (section 2.6.6.3) but omitting the phenol/chloroform extraction step. These total cosmid preparations were next transformed into Subcloning Efficiency *E. coli* DH5α Competent Cells (Invitrogen) according to the manufacturer's instructions. Individual colonies were picked following antibiotic selection with carbenicillin and apramycin and cosmids pIJ12315-8 were subsequently isolated and confirmed by restriction mapping with *Bam*H I and *Hind*III.

2.6.11.2 Conjugative cosmid transfer into S. coelicolor

Recombinant cosmids pIJ12315-8 were next transformed into *E. coli* ET12567/pUZ8002 and conjugated with *S. coelicolor* strains M1146 and M1152 according the published procedures.[328] Bacterial strains and plasmids involved in these protocols as well as details of recombinant *S. coelicolor* strains produced are listed in Table 2.8, Table 2.9 and Table 2.10.

2.6.12 Cloning minimal tun genes into S. coelicolor via pRT802

A 12.9 kb *Sac*I restriction fragment was isolated from genomic library-derived cosmid 4H8 and ligated into conjugative and integrative vector pRT802 at its unique *Sac*I site, yielding construct pIJ12003a. This *Sac*I fragment contains the complete *tun* cluster plus 427 bp upstream of *tunA* and 500 bp downstream of *tunN*; neither of these additional DNA sequences is predicted to possess an entire ORF. Subsequently, pIJ12003a was conjugated into *S. coelicolor* M1146 *via* triparental mating using *E. coli* S17-1 and *E. coli* ET12567/pUZ8002. Bacterial strains and plasmids involved in these protocols as well as details of recombinant *S. coelicolor* strains produced are listed in Table 2.8, Table 2.9 and Table 2.10.

2.6.13 Heterologous expression in S. coelicolor hosts

2.6.13.1 Agar-diffusion bioassay

Dense spore suspensions (10 µL) of recombinant *S. coelicolor* M1025-1028, M1031 and their vector-only control strains were spread on R5 agar and incubated at 30 °C for 48 h

until confluent lawns of mycelia were visible. Meanwhile, LB (5 mL) was inoculated with a single colony of tunicamycin-sensitive *B. subtilis* EC 1524 and grown overnight at 30 °C. This reporter strain was then subcultured (1 in 25 dilution) in 50 mL LB media for 3-4 h (30 °C, 250 rpm). 50 mL of soft nutrient agar was melted and cooled to 50 °C, to which 5 mL of the *B. subtilis* subculture was added. This pre-infected agar was then used to flood empty Petri dishes (25 mL per dish) containing agar cores from the recombinant *S. coelicolor* plates. Additionally, sterile filter disks spotted with 20 µL of tunicamycin stock solution (1 mg/mL) were placed atop the solidified agar. These plates were then grown at 30 °C for 18 h and examined for growth inhibition of the reporter strain.

2.6.13.2 LC/MS analysis of crude extracts

Recombinant *S. coelicolor* strains and their vector-only control strains were grown in liquid TYD medium for 5 days at 30 °C. The suspected tunicamycin metabolites were isolated and characterised by methanolic mycelium extraction and LC/MS analysis, as previously described.[296] In this case a Micromass LCT (ESI-TOF-MS) coupled to an Agilent 1200 Series LC System was used.

2.6.13.3 ^{1}H NMR analysis of purified extracts

Previously isolated crude extracts from *S. coelicolor* M1031 were purified by flash chromatography using Fluka Kieselgel 60 220-440 mesh silica gel (mobile phase: water/isopropanol/ethyl acetate 1:3:6), dissolved in CD_3OD and subjected to ^{1}H NMR spectroscopy along with an authentic sample to further confirm the presence of tunicamycin.

2.6.13.4 Calculation of tunicamycin production yields

S. chartreusis and *S. coelicolor* M1027 and M1031 were grown in R5 and TYD production media for 5 days (28 °C, 200 rpm). Mycelia were collected and the wet cell pellet weighed, prior to extraction of tunicamycins through the use of protocols described in section 3.5.1.4. Yields of tunicamycin production were calculated in terms of milligrams of tunicamycin per litre of culture and also in terms of milligrams of tunicamycin per gram of wet cell pellet.

2.7 References

282. Weber T, Welzel K, Pelzer S, Vente A, Wohlleben W (2003): **Exploiting the genetic potential of polyketide producing streptomycetes**. *J. Biotechnol.* **106**: 221-232.

283. Walsh CT, Fischbach MA (2010): **Natural Products Version 2.0: Connecting Genes to Molecules**. *J. Am. Chem. Soc.* **132**: 2469-2493.

284. Ostash B, Saghatelian A, Walker S (2007): **A Streamlined Metabolic Pathway for the Biosynthesis of Moenomycin A**. *Chem. Biol.* **14**: 257-267.

285. Kaysser L, Lutsch L, Siebenberg S, Wemakor E, Kammerer B, Gust B (2009): **Identification and Manipulation of the Caprazamycin Gene Cluster Lead to New Simplified Liponucleoside Antibiotics and Give Insights into the Biosynthetic Pathway**. *J. Biol. Chem.* **284**: 14987-14996.

286. Bode HB, Müller R (2005): **The Impact of Bacterial Genomics on Natural Product Research**. *Angew. Chem., Int. Ed.* **44**: 6828-6846.

287. Corre C, Challis GL (2009): **New natural product biosynthetic chemistry discovered by genome mining**. *Nat. Prod. Rep.* **26**: 977-986.

288. International Nucleotide Sequence Database Collaboration - http://www.insdc.org/index.html.

289. Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA (2000): *Practical Streptomyces Genetics.* John Innes Foundation: Norwich.

290. Hamill RG, Eli Lilly Co. (1980): **Process for preparing tunicamycin**. U. S. Patent 4237225

291. Takatsuki A, Arima K, Tamura G (1971): **Tunicamycin, a New Antibiotic: 1. Isolation and Characterization of Tunicamycin**. *J. Antibiot.* **24**: 215-223.

292. Staden R (1980): **A New Computer Method for the Storage and Manipulation of DNA Gel Reading Data**. *Nucl. Acids Res.* **8**: 3673-3694.

293. *Genome Sequencer Data Analysis Software Manual* (2008). Roche Diagnostics GmbH: Mannheim.

294. The N50 average contig size represents the largest contig E such that at least half of the total size of the contigs is contained in contigs larger than E. For example in a collection of contigs with sizes 7, 4, 3, 2, 2, 1, and 1 kb (total size = 20 kb), the N50 length is 4 because we can cover 10 kb with contigs bigger than 4kb.

295. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997): **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucl. Acids Res.* **25**: 3389-3402.

296. Tsvetanova BC, Kiemle DJ, Price NPJ (2002): **Biosynthesis of tunicamycin and metabolic origin of the 11-carbon dialdose sugar, tunicamine**. *J. Biol. Chem.* **277**: 35289-35296.

297. De Smet KAL, Weston A, Brown IN, Young DB, Robertson BD (2000): **Three pathways for trehalose biosynthesis in mycobacteria**. *Microbiology* **146**: 199-208.

298. Inaoka T, Ochi K (2007): **Glucose uptake pathway-specific regulation of synthesis of neotrehalosadiamine, a novel autoinducer produced in *Bacillus subtilis***. *J. Bacteriol.* **189**: 65-75.

299. Choeng YH, Yang JY, Delcroix G, Kim YJ, Chang YK, Hong SK (2007): **Expression and characterization of trehalose biosynthetic modules in the adjacent locus of the salbostatin gene cluster**. *J. Microbiol. Biotechnol.* **17**: 1675-1681.

300. Truman AW, Robinson L, Spencer JB (2006): **Identification of a deacetylase involved in the maturation of teicoplanin**. *ChemBioChem* **7**: 1670-1675.

301. Yokoyama K, Yamamoto Y, Kudo F, Eguchi T (2008): **Involvement of two distinct *N*-acetylglucosaminyltransferases and a dual-function deacetylase in neomycin biosynthesis**. *ChemBioChem* **9**: 865-869.

302. Newton GL, Av-Gay Y, Fahey RC (2000): ***N*-acetyl-1-D-*myo*-inosityl-2-amino-2-deoxy-alpha-D-glucopyranoside deacetylase (MshB) is a key enzyme in mycothiol biosynthesis**. *J. Bacteriol.* **182**: 6958-6963.

303. Staunton J, Weissman KJ (2001): **Polyketide biosynthesis: a millennium review**. *Nat. Prod. Rep.* **18**: 380-416.

304. Crump MP, Crosby J, Dempsey CE, Parkinson JA, Murray M, Hopwood DA, Simpson TJ (1997): **Solution structure of the actinorhodin polyketide synthase acyl carrier protein from *Streptomyces coelicolor* A3(2)**. *Biochemistry* **36**: 6000-6008.

305. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, Barrell B (2000): **Artemis: sequence visualization and annotation**. *Bioinformatics* **16**: 944-945.

306. Bibb MJ, Findlay PR, Johnson MW (1984): **The Relationship between Base Composition and Codon Usage in Bacterial Genes and Its Use for the Simple and Reliable Identification of Protein-Coding Sequences**. *Gene* **30**: 157-166.

307. Ishikawa J, Hotta K (1999): **FramePlot: a new implementation of the Frame analysis for predicting protein-coding regions in bacterial DNA with a high G plus C content**. *FEMS Microbiol. Lett.* **174**: 251-253.

308. Wright F, Bibb MJ (1992): **Codon Usage in the G+C-Rich *Streptomyces* Genome**. *Gene* **113**: 55-65.

309. Benbassat A, Bauer K, Chang SY, Myambo K, Boosman A, Chang S (1987): **Processing of the Initiation Methionine from Proteins - Properties of the *Escherichia coli* Methionine Aminopeptidase and Its Gene Structure**. *J. Bacteriol.* **169**: 751-757.

310. Brenner S, Barnett L, Katz ER, Crick FHC (1967): **UGA: A Third Nonsense Triplet in Genetic Code**. *Nature* **213**: 449-450.

311. Brenner S, Stretton AO, Kaplan S (1965): **Genetic Code - Nonsense Triplets for Chain Termination and Their Suppression**. *Nature* **206**: 994-998.

312. Shine J, Dalgarno L (1975): **Determinant of Cistron Specificity in Bacterial Ribosomes**. *Nature* **254**: 34-38.

313. Strohl WR (1992): **Compilation and Analysis of DNA-Sequences Associated with Apparent Streptomycete Promoters**. *Nucl. Acids Res.* **20**: 961-974.

314. Land M, Lapidus A, Mayilraj S, Chen F, Copeland A, Rio TGD, Nolan M, Lucas S, Tice H, Cheng J-F, et al. (2009): **Complete genome sequence of *Actinosynnema mirum* type strain (101^T)**. *Standards in Genomic Science* **1**: 46-53.

315. Kenig M, Reading C (1979): **Holomycin and an Antibiotic (MM 19290) Related to Tunicamycin, Metabolites of *Streptomyces clavuligerus***. *J. Antibiot.* **32**: 549-554.

316. *SuperCos 1 Cosmid Vector Kit Instruction Manual* (2008). Agilent Technologies Inc.: Santa Clara.

317. Lodish H, Darnell JE, Berk A, Zipursky L, Matsudaira P, Baltimore D (1999): *Molecular Cell Biology* 4th edn. W.H.Freeman & Co Ltd: New York.

318. Sambrook J, Russell D (2000): *Molecular Cloning: A Laboratory Manual* 3rd edn. Cold Spring Harbor Laboratory Press: New York.

319. Yanai K, Murakami T, Bibb MJ (2006): **Amplification of the entire kanamycin biosynthetic gene cluster during empirical strain improvement of *Streptomyces kanamyceticus***. *Proc. Natl. Acad. Sci. U.S.A.* **103**: 9661-9666.

320. Mazodier P, Petter R, Thompson C (1989): **Intergeneric Conjugation between *Escherichia coli* and *Streptomyces* Species**. *J. Bacteriol.* **171**: 3583-3585.

321. Thorpe HM, Smith MCM (1998): ***In vitro*** **site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family**. *Proc. Natl. Acad. Sci. U.S.A.* **95**: 5505-5510.

322. Thorpe HM, Wilson SE, Smith MCM (2000): **Control of directionality in the site-specific recombination system of the *Streptomyces* phage ϕC31**. *Mol. Microbiol.* **38**: 232-241.

323. Gust B, Chandra G, Jakimowicz D, Tian YQ, Bruton CJ, Chater KF (2004): **lambda red-mediated genetic manipulation of antibiotic-producing *Streptomyces***. *Adv. Appl. Microbiol.* **54**: 107-128.

324. Datsenko KA, Wanner BL (2000): **One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products**. *Proc. Natl. Acad. Sci. U.S.A.* **97**: 6640-6645.

325. Gust B, Challis GL, Fowler K, Kieser T, Chater KF (2003): **PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin**. *Proc. Natl. Acad. Sci. U.S.A.* **100**: 1541-1546.

326. Paget MSB, Chamberlin L, Atrih A, Foster SJ, Buttner MJ (1999): **Evidence that the Extracytoplasmic Function Sigma Factor sigma E Is Required for Normal Cell Wall Structure in *Streptomyces coelicolor* A3(2)**. *J. Bacteriol.* **181**: 204-211.

327. Macneil DJ, Gewain KM, Ruby CL, Dezeny G, Gibbons PH, Macneil T (1992): **Analysis of *Streptomyces avermitilis* Genes Required for Avermectin Biosynthesis Utilizing a Novel Integration Vector**. *Gene* **111**: 61-68.

328. Gust B, Kieser T, Chater KF (2002): *REDIRECT technology: PCR-targeting system in Streptomyces coelicolor.* John Innes Foundation: Norwich.

329. Zhang YM, Buchholz F, Muyrers JPP, Stewart AF (1998): **A new logic for DNA engineering using recombination in *Escherichia coli***. *Nature Genetics* **20**: 123-128.

330. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, et al. (2002): **Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2)**. *Nature* **417**: 141-147.

331. Gregory MA, Till R, Smith MCM (2003): **Integration site for *Streptomyces* phage ϕBT1 and development of site-specific integrating vectors**. *J. Bacteriol.* **185**: 5320-5323.

332. Simon R, Priefer U, Puhler A (1983): **A Broad Host Range Mobilization System for *In vivo* Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria**. *Nat. Biotechnol.* **1**: 784-791.

333. Hu HF, Zhang Q, Ochi K (2002): **Activation of antibiotic biosynthesis by specified mutations in the *rpoB* gene (encoding the RNA polymerase beta subunit) of *Streptomyces lividans***. *J. Bacteriol.* **184**: 3984-3991.

334. Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, et al. (2009): **CDD: specific functional annotation with the Conserved Domain Database**. *Nucl. Acids Res.* **37**: D205-D210.

335. Soding J (2005): **Protein homology detection by HMM-HMM comparison**. *Bioinformatics* **21**: 951-960.

336. Rossmann MG, Moras D, Olsen KW (1974): **Chemical and Biological Evolution of a Nucleotide-Binding Protein**. *Nature* **250**: 194-199.

337. Filling C, Berndt KD, Benach J, Knapp S, Prozorovski T, Nordling E, Ladenstein R, Jornvall H, Oppermann U (2002): **Critical residues for structure and catalysis in short-chain dehydrogenases/reductases**. *J. Biol. Chem.* **277**: 25677-25684.

338. Oppermann U, Filling C, Hult M, Shafqat N, Wu XQ, Lindh M, Shafqat J, Nordling E, Kallberg Y, Persson B, et al. (2003): **Short-chain dehydrogenases/reductases (SDR): the 2002 update**. *Chem.-Biol. Interact.* **143**: 247-253.

339. Creuzenet C, Schur MJ, Lin JJ, Wakarchuk WW, Lam JS (2000): **FlaA1, a new bifunctional UDP-GlcNAc C-6 dehydratase C-4 reductase from *Helicobacter pylori***. *J. Biol. Chem.* **275**: 34873-34880.

340. Allard STM, Cleland WW, Holden HM (2004): **High resolution x-ray structure of dTDP-glucose 4,6-dehydratase from *Streptomyces venezuelae***. *J. Biol. Chem.* **279**: 2211-2220.

341. Yabe T (2010): *Functional characterizations of the tun genes in the biosynthetic cluster of tunicamycin.* Part II Thesis: University of Oxford.

342. Kaysser L, Siebenberg S, Kammerer B, Gust B (2010): **Analysis of the Liposidomycin Gene Cluster Leads to the Identification of New Caprazamycin Derivatives**. *ChemBioChem* **11**: 191-196.

343. Ohnuki T, Muramatsu Y, Miyakoshi S, Takatsu T, Inukai M (2003): **Studies on novel bacterial translocase I inhibitors, A-500359s: IV. Biosynthesis of A-500359s**. *J. Antibiot.* **56**: 268-279.

344. Winn M, Goss RJM, Kimura K-I, Bugg TDH (2010): **Antimicrobial nucleoside antibiotics targeting cell wall assembly: Recent advances in structure-function studies and nucleoside biosynthesis**. *Nat. Prod. Rep.* **27**: 279-304.

345. McLennan AG (2006): **The Nudix hydrolase superfamily**. *Cell. Mol. Life Sci.* **63**: 123-143.

346. Xu WL, Shen JY, Dunn CA, Bessman MJ (2003): **A new subfamily of the nudix hydrolase superfamily active on 5-methyl-UTP (ribo-TTP) and UTP**. *J. Biol. Chem.* **278**: 37492-37496.

347. Roth JR, Lawrence JG, Rubenfield M, Kiefferhiggins S, Church GM (1993): **Characterization of the Cobalamin (Vitamin-B12) Biosynthetic Genes of *Salmonella typhimurium***. *J. Bacteriol.* **175**: 3303-3316.

348. Otoole GA, Trzebiatowski JR, Escalantesemerena JC (1994): **The *cobC* Gene of *Salmonella typhimurium* Codes for a Novel Phosphatase Involved in the Assembly of the Nucleotide Loop of Cobalamin**. *J. Biol. Chem.* **269**: 26503-26511.

349. Rigden DJ (2008): **The histidine phosphatase superfamily: structure and function**. *Biochem. J.* **409**: 333-348.

350. Sanchez MB, Barrado P, Jimenez A, Lobato MF (2006): **The *pur3* gene from the *pur* cluster encodes a monophosphatase essential for puromycin biosynthesis in *Streptomyces***. *FEBS Lett.* **580**: 1807-1811.

351. Espinosa JC, Tercero JA, Rubio MA, Jimenez A (1999): **The *pur7* gene from the puromycin biosynthetic *pur* cluster of *Streptomyces alboniger* encodes a nudix hydrolase**. *J. Bacteriol.* **181**: 4914-4918.

352. Wang SC, Frey PA (2007): ***S*-adenosylmethionine as an oxidant: the radical SAM superfamily**. *Trends Biochem. Sci.* **32**: 101-110.

353. Jarrett JT (2003): **The generation of 5'-deoxyadenosyl radicals by adenosylmethionine-dependent radical enzymes**. *Curr. Opin. Chem. Biol.* **7**: 174-182.

354. Martin JL, McMillan FM (2002): **SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold**. *Curr. Opin. Struct. Biol.* **12**: 783-793.

355. Trefzer A, Salas JA, Bechthold A (1999): **Genes and enzymes involved in deoxysugar biosynthesis in bacteria**. *Nat. Prod. Rep.* **16**: 283-299.

356. Lairson LL, Henrissat B, Davies GJ, Withers SG (2008): **Glycosyltransferases: Structures, functions, and mechanisms**. *Annu. Rev. Biochem.* **77**: 521-555.

357. Nakamura N, Inoue N, Watanabe R, Takahaski M, Takeda J, Stevens VL, Kinoshita T (1997): **Expression cloning of PIG-L, a candidate *N*-acetylglucosaminylphosphatidylinositol deacetylase**. *J. Biol. Chem.* **272**: 15834-15840.

358. Urbaniak MD, Crossman A, Chang TH, Smith TK, van Aalten DMF, Ferguson MAJ (2005): **The *N*-acetyl-D-glucosaminylphosphatidylinositol De-*N*-acetylase of glycosylphosphatidylinositol biosynthesis is a zinc metalloenzyme**. *J. Biol. Chem.* **280**: 22831-22838.

359. Borghi A, Edwards D, Zerilli LF, Lancini GC (1991): **Factors affecting the normal and branched-chain acyl moieties of teicoplanin components produced by *Actinoplanes teichomyceticus***. *J. Gen. Microbiol.* **137**: 587-592.

360. Carman GM, Han GS (2006): **Roles of phosphatidate phosphatase enzymes in lipid metabolism**. *Trends Biochem. Sci.* **31**: 694-699.

361. Carman GM, Han GS (2009): **Phosphatidic Acid Phosphatase, a Key Enzyme in the Regulation of Lipid Synthesis**. *J. Biol. Chem.* **284**: 2593-2597.

362. Vetting MW, de Carvalho LPS, Yu M, Hegde SS, Magnet S, Roderick SL, Blanchard JS (2005): **Structure and functions of the GNAT superfamily of acetyltransferases**. *Arch. Biochem. Biophys.* **433**: 212-226.

363. Gu LC, Geders TW, Wang B, Gerwick WH, Hakansson K, Smith JL, Sherman DH (2007): **GNAT-like strategy for polyketide chain initiation**. *Science* **318**: 970-974.

364. Davidson AL, Dassa E, Orelle C, Chen J (2008): **Structure, function, and evolution of bacterial ATP-binding cassette systems**. *Microbiol. Mol. Biol. Rev.* **72**: 317-364.

365. Mendez C, Salas JA (2001): **The role of ABC transporters in antibiotic-producing organisms: drug secretion and resistance mechanisms**. *Res. Microbiol.* **152**: 341-350.

366. Chater KF (2006): ***Streptomyces* inside-out: a new perspective on the bacteria that provide us with antibiotics**. *Philos. Trans. R. Soc., B* **361**: 761-768.

367. Leskiw BK, Mah R, Lawlor EJ, Chater KF (1993): **Accumulation of *bldA*-Specified tRNA Is Temporally Regulated in *Streptomyces coelicolor* A3(2)**. *J. Bacteriol.* **175**: 1995-2005.

368. Price NPJ, Tsvetanova B (2007): **Biosynthesis of the tunicamycins: A review**. *J. Antibiot.* **60**: 485-491.

Chapter 3: Natural product isolation and degradation

3.1 Introduction

In section 1.5 we hypothesised that through producing libraries of tunicamycin analogues the selectivity for bacterial MraY over human GPT might be greatly improved, thereby potentially yielding a therapeutically vital antibiotic with orthogonal function to existing drugs. Additionally, we suggested that analogues of tunicamycin could be rationally designed to mimic the nucleotide-sugar substrate or even the bi-substrate transition state of a range of important glycosyltransferases unrelated to the natural target – UDP-HexNAc:Dol-P HexNAc-1-P transferases. These analogues would all be based around the tunicamine core of tunicamycin, which has been proposed to act as a pyrophosphate mimic.[369] This core would then be decorated with a range of peripheral functionalities, carefully tailored to mimic specific glycosyltransferase substrates (section 1.5.2).

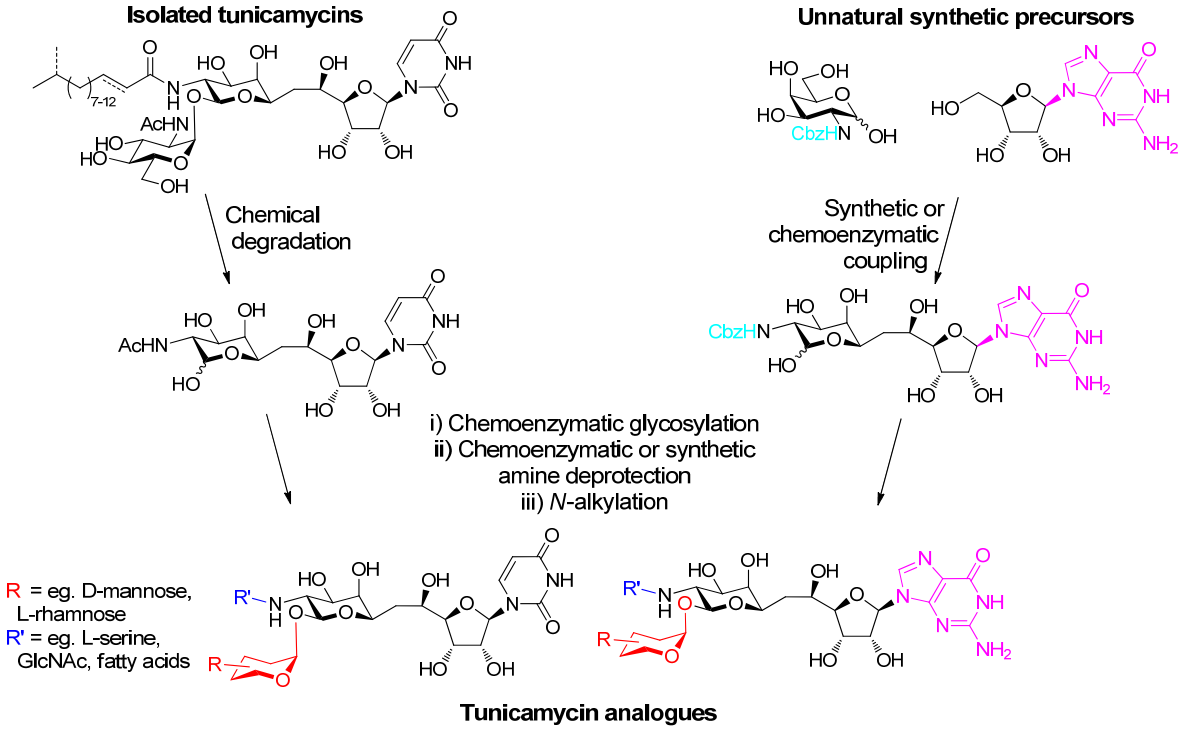


Fig. 3.1 A combined relay, chemoenzymatic and synthetic strategy to tunicamycin analogues

Access to the tunicamine core and its variants was envisaged through both chemical degradation of naturally isolated tunicamycins and chemoenzymatic formation from unnatural precursors – utilising enzymes from tunicamycin biosynthesis. Further functionalisation of these core intermediates at the C-11´ anomeric hydroxyl and the C-10´ amine was then to be realised by a combination of synthetic and chemoenzymatic methods (Fig. 3.1).

In this chapter I report the optimisation of procedures for the isolation and purification of native tunicamycins in preparative quantities from fermentation broths of *Streptomyces chartreusis*. I then discuss the development of efficient routes to the chemical degradation of these isolates, yielding advanced biosynthetic intermediates containing the key tunicamine scaffold. These intermediates are valuable not only as precursors in the construction of tunicamycin analogues, but also as putative native enzyme substrates for the study of individual enzymes encoded by genes in the tunicamycin biosynthetic gene cluster (*tun*).

3.2 Isolation of tunicamycins from S. chartreusis

3.2.1 Analytical scale tunicamycin isolation

When tunicamycin was first discovered in 1971, the method used by Takatsuki *et al.* for its isolation from liquid culture of *S. lysosuperificus* involved multiple rounds of solvent extraction, acidification and precipitation of both the cell pellet and the media supernatant.[370] The method was very labour intensive although it yielded crystalline samples of tunicamycin. The isolation of tunicamycin from *S. chartreusis* was additionally described by Hamill *et al.* in a series of patents a decade later.[371,372] They utilised a similar approach but on a large scale (up to 200 L cultures) and made use of chromatography on a range of solid absorbents. When Tsvetanova *et al*. performed their initial investigations of tunicamycin biosynthesis with metabolic feeding experiments, they used a method similar to that of Takatsuki *et al*., involving methanolic extraction of acidified *S. chartreusis* cultures.[373] Working on a small scale, they purified their tunicamycin isolates by preparative thin layer chromatography. They additionally isolated tunicamycins at low levels from solid media.[374]

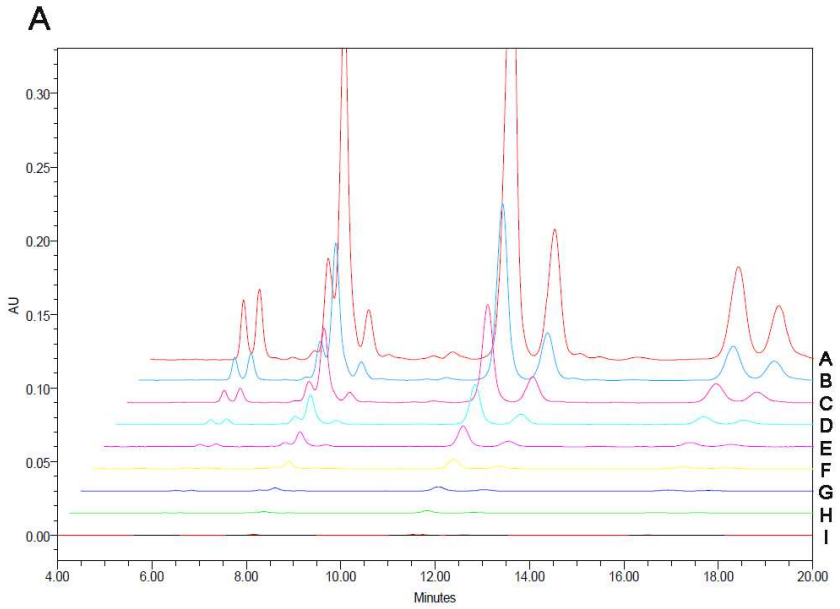
We attempted to reproduce these conditions in order to access preparatory quantities of the tunicamycins for degradation studies. Despite *S. lysosuperificus* being the source organism when the discovery of tunicamycin was first reported,[370] Tsvetanova *et al.* could only detect production on agar plates using this strain and reported far superior production from *S. chartreusis* when liquid culture was used.[374] Since this was our preferred method of fermentation in terms of facile scalability, we elected to work with *S. chartreusis* NRRL 3882, sourced from the German Collection of Microorganisms and Cell Cultures (DSMZ). Following the protocol of Tsvetanova *et al.*, liquid tryptone/yeast extract/dextrose (TYD) medium was inoculated with *S. chartreusis* and grown for 28 °C with agitation. After 7 days, sporulation caused the media to adopt a characteristic red colour. Interestingly, it was found that unbaffled culture flasks were required to maximise tunicamycin production; it is possible that the reduction in culture aeration associated with these conditions stimulates a stress response resulting in sporulation and initiation of concomitant secondary metabolite production. Since tunicamycins are entirely insoluble at low pH, the culture media was then acidified with conc. HCl – to a final concentration of 0.2 M – in order to precipitate any tunicamycins. The pelleted mycelium was subsequently extracted multiple times with MeOH and the combined extracts concentrated. The following section documents the

characterisation of these crude preparations, alongside an authentic sample, in order to detect tunicamycin production.

3.2.2 Characterisation of isolated tunicamycins

3.2.2.1 High-performance liquid chromatography (HPLC)

A



B

Standard	Dilution	Conc. (µg/mL)[a]	HPLC total area
A	1	1,320	15,800,000
B	2½	528	5,860,000
C	5	264	2,990,000
D	12½	106	1,250,000
E	25	52.9	641,000
F	50	26.8	329,000
G	125	11.6	134,000
H	250	5.6	76,900
I	500	2.8	52,800
		0	0

[a]based on mass used to prepare standard A

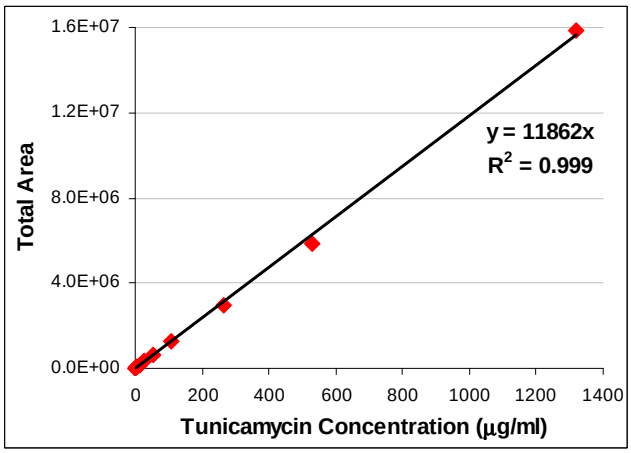


Fig. 3.2 HPLC of tunicamycin standards A to I; **A**: UV chromatograms at 260 nm; **B**: Calibration curve relating total HPLC area to tunicamycin concentration;

Serial dilutions of an authentic tunicamycin sample (Fluorochem Ltd) were prepared in MeOH. Their concentrations showed good correlation with those calculated from absorbance values at 260 nm and quoted extinction coefficients (ε%),[375] although these absorbance profiles became saturated at high tunicamycin concentrations. The dilutions were then analysed by C-18 reverse phase chromatography with detection at 260

nm (Fig. 3.2, A). The resulting spectra show peaks at a range of retention times between 6 and 20 minutes, corresponding to the differentially *N*-acylated tunicamycin homologues and resulting from a positive correlation between acyl chain length and interaction with the C-18 sorbent. Such separation of tunicamycin homologues using reversed-phase chromatography is consistent with published reports.[374,376] The total area underneath peaks in each spectra were used to construct a calibration curve (Fig. 3.2, B), allowing the quantification of tunicamycin concentration in unknown samples.

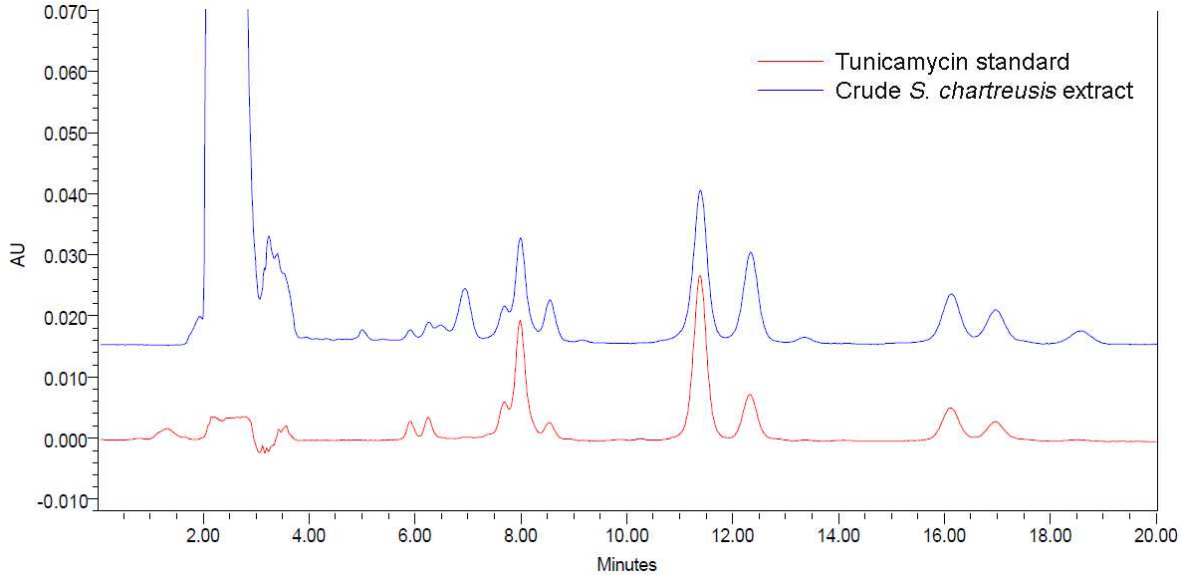


Fig. 3.3 HPLC chromatograms (260 nm) of authentic and isolated tunicamycin samples

A sample of the previously obtained crude extract from *S. chartreusis* fermentation broth in MeOH was analysed by HPLC (Fig. 3.3). Its chromatogram clearly exhibited the characteristic peak distribution of tunicamycin homologues observed in the chromatograms of tunicamycin standards and confirmed the production of tunicamycin in our fermentation experiments. A number of additional peaks were present in the chromatogram, particularly near the void volume of the column (2 to 4 min), although the identity of these impurities was not determined due to the crude nature of the sample tested. Integration of the peaks purportedly deriving from tunicamycin homologues were next referenced against a calibration curve. This gave an estimated yield of tunicamycin production by *S. chartreusis* of 7.1 mg per L of fermentation broth.

3.2.2.2 Liquid chromatography-mass spectrometry (LC/MS)

Serial dilutions of a tunicamycin standard and our crude *S. chartreusis* isolate were further subjected to LC/MS analysis, employed to couple HPLC separation of individual tunicamycins with accurate mass detection. The retention time distribution of differentially *N*-acylated homologues in the 260 nm absorption spectrum was again observed and this result was mirrored in the total ion chromatogram (TIC)(Fig. 3.5, A: i and ii). The resulting mass spectrum showed a series of peaks at *m/z* 817, 831, 833, 845 and 860 corresponding to $[M + H]^+$ ions of the tunicamycins (Fig. 3.5, A: iii). The separation of 14 Da between these peaks corresponded to variations in the number of methylene (CH_2) groups in the fatty acid side chains of different tunicamycin homologues. Additionally, the spectrum showed a series of peaks at *m/z* 596, 610, 612, 624 and 638, corresponding to $[M + H - 221]^+$ ions of the tunicamycins resulting from their fragmentation within the spectrometer and the consequent loss of their terminal GlcNAc moiety (Fig. 3.4). This fragmentation was also observed by Tsvetanova *et al.* and is characteristic of the tunicamycins.[374] Pleasingly, the UV and TIC chromatograms of our crude *S. chartreusis* extract showed the same characteristic peaks as the authentic sample, with retention times ranging from 5 to 13 minutes. The resulting mass spectrum was almost identical to that of the authentic sample and clearly showed the fatty acyl distribution of tunicamycin homologues as well as the characteristic $[M + H - 221]^+$ fragmentation (Fig. 3.5, B). Minor $[M + H - 203]^+$ fragmentations were also observed.

Mw of tunicamycin homologue Tun16:1$_{A/B}$ = 844
$[M + H]^+$ = 845
major fragmentation $[M + H - 221]^+$ = 624
minor fragmentation $[M + H - 203]^+$ = 642

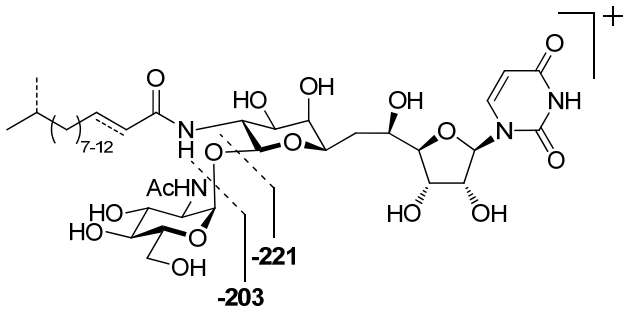


Fig. 3.4 Characteristic fragmentations of tunicamycins across α,β-glycosidic bond under ESI+ MS

Aside from confirming the production of tunicamycin in our fermentation experiments, the spectrum showed the fatty acid side chains in our mixture of tunicamycin homologues was very similar to that of the commercial sample, isolated from *S. lysosuperificus*. The relative proportions of these homologues varied between the two samples; this may be a consequence of different growth conditions used to produce the two

samples, or alternatively of a different fatty acid distribution within the cellular lipid pools of *S. chartreusis* and *S. lysosuperificus*.

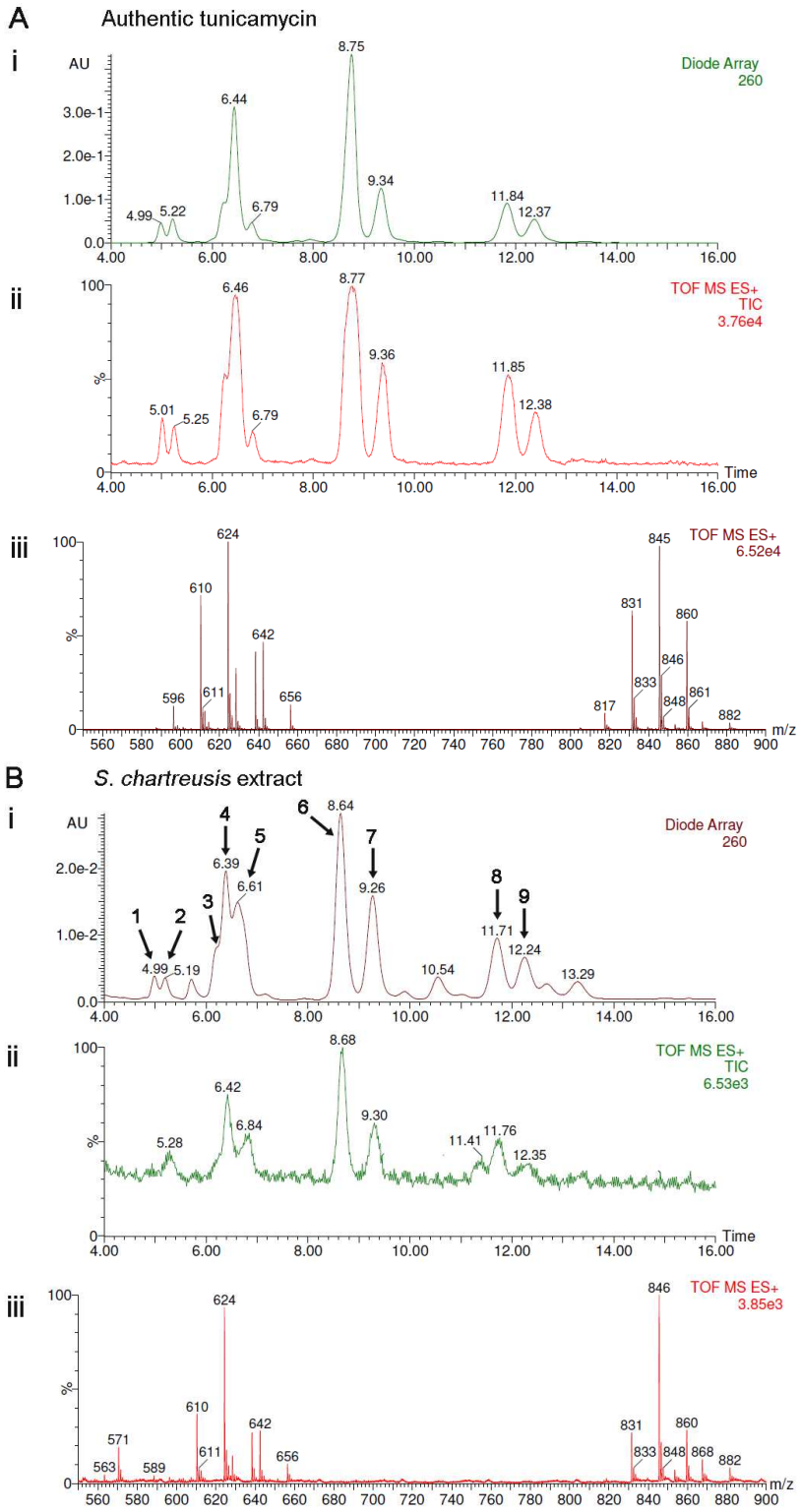


Fig. 3.5 LC/MS analysis of tunicamycin extracts; **A**: Authentic tunicamycin sample: (i) chromatogram at 260 nm, (ii) TIC and (iii) extracted mass spectrum of 4-14 min; **B**: Crude *S. chartreusis* extract: (i) chromatogram at 260 nm, (ii) TIC and (iii) extracted mass spectrum 4-14 min;

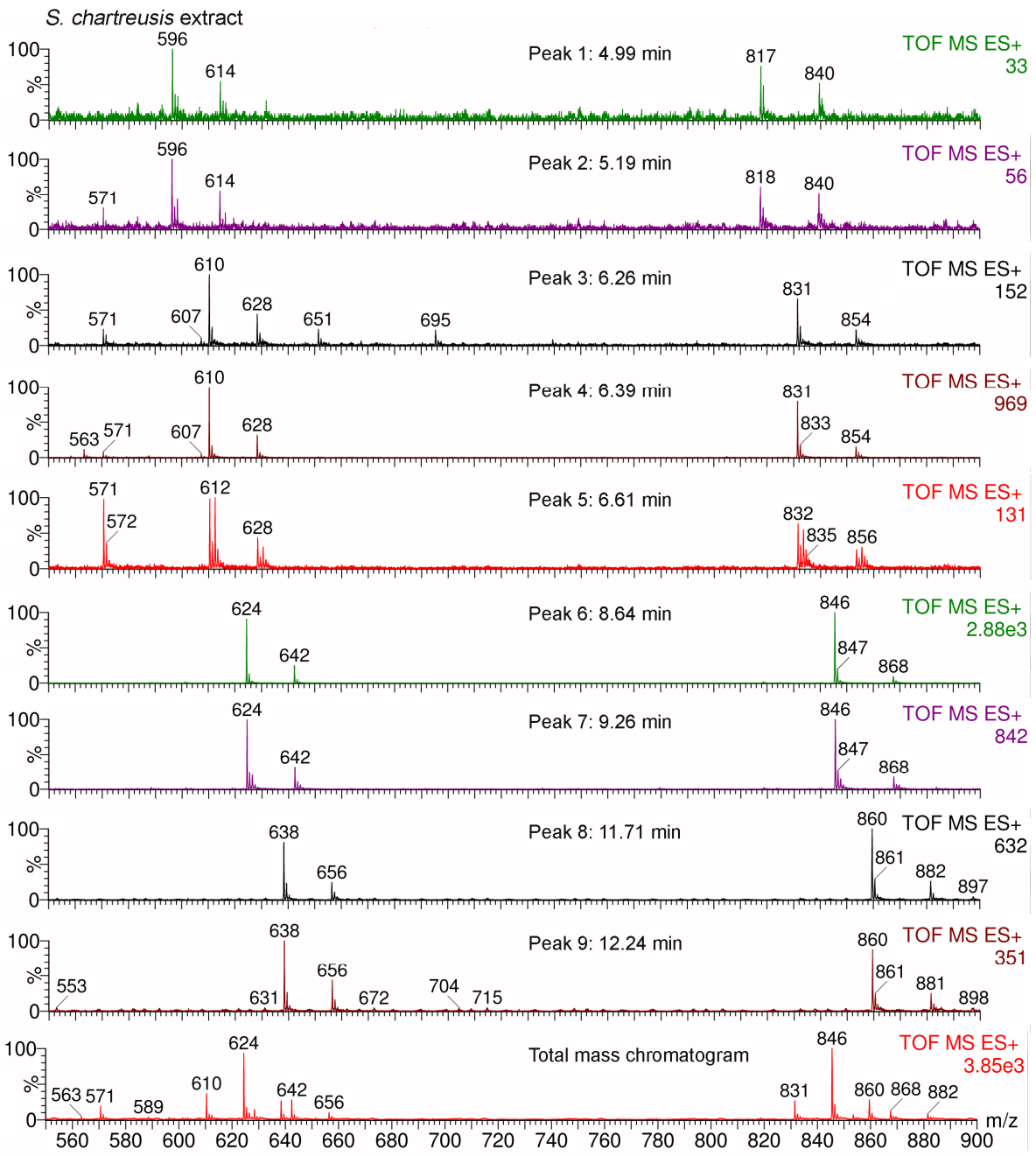


Fig. 3.6 Individual mass spectra extracted from peaks 1-9 of a crude tunicamycin extract (see Fig. 3.5, B) and a total mass chromatogram for reference

The sample of crude tunicamycin extract showed a total of nine distinguishable homologues in its UV chromatogram, each possessing a different retention time. Separate mass chromatograms were extracted for each individual HPLC peak, allowing the nature of each homologue to be assessed in greater detail (Fig. 3.6). As expected, it was found that increasing retention time correlated with increasing acyl chain length. Certain sets of peaks contained tunicamycins with the same *m/z* despite having different retention times, suggesting these homologous are branched and straight-chain isomers of one another. The

data collected closely matched that of an authentic tunicamycin sample and were also consistent with previously reported observations.[374]

3.2.3 Preparative scale tunicamycin isolation

Thus far we have successfully detected tunicamycin production in fermentation broths of *S. chartreusis*. However, the experiments were performed only on an analytical scale and characterisation took place on extracts which were not purified. Additionally, although data is limited on the precise yields obtained by previous research groups, our value of 7.1 mg/L determined from an unpurified sample was much less than the 31 mg/L of purified tunicamycin reported by Hamill *et al*.[371] In order to access the preparative quantities of tunicamycin required for degradation studies, the fermentation scale was increased and improvements to the extraction procedure introduced. Additionally, efficient purification protocols were developed. These modifications are described below.

3.2.3.1 Extraction from cell pellet

Typically, a scaled-up fermentation was performed using twelve flasks each containing 1 L of culture media. Following centrifugation the majority of the tunicamycins were found in the cell pellet, owing to the aggregation of cultured cells and also to the moderate aqueous solubility of tunicamycins. Since tunicamycin is entirely insoluble at low pH, washing the mycelial cake with 1M aq. HCl was found to selectively remove a large proportion of water-soluble contaminants. Subsequent extraction of the cell pellet with multiple portions of MeOH led to increased levels of isolated tunicamycins.

3.2.3.2 Extraction from culture supernatant

A large volume of culture media remained following removal of *S. chartreusis* cells; this supernatant was found to contain appreciable quantities of tunicamycin. Solid-phase extraction (SPE) methods were thus adopted to selectively extract tunicamycins from this solution, providing the dual benefits of introducing an efficient purification step whilst rapidly concentrating the tunicamycin solution to a much smaller and more manageable volume. Noting that tunicamycins exhibit good retention profiles on C-18 reversed-phase HPLC columns, the culture supernatant was passed through pre-equilibrated C-18 SPE cartridges (Waters). The majority of contaminants in the solution were polar carbohydrates or amino acids – derived from the TYD medium – which would not bind to the C-18 silica sorbent, resulting in selective retention of tunicamycin along with minor contaminants.

These non-polar analytes were eluted from the SPE cartridges with a small volume of MeOH, yielding a concentrated solution of highly enriched in tunicamycin. The use of C-18 SPE cartridges was eventually superceded by reversed-phase XAD-16 resin (Amberlite) which could: (i) be incubated for longer periods of time with the media supernatant, (ii) be rapidly isolated by simple filtration and (iii) avoid problems associated with column blockage by residual insoluble cell debris. Both sorbents also proved cost-effective since they could be recycled and used multiple times.

3.2.3.3 Purification by flash chromatography

Methanolic extracts of the cell pellet were highly impure and although supernatant extraction with C-18 resin afforded highly enriched tunicamycin samples, these also required further purification. The partially purified extracts were thus combined and dry loaded on silica. A suitable polar elution solvent system was devised such that the tunicamycins could be efficiently purified in a single pass by normal phase flash chromatography. The resulting tunicamycins were shown to contain <5% impurities by ^{1}H NMR and HPLC (see experimental section 3.5.3.2).

3.2.3.4 Significance of optimised procedure

The result of these modifications to the tunicamycin extraction procedure was that the isolated yield of pure tunicamycin increased to 35 mg/L. This was a significant improvement over initial efforts and was high enough to provide sufficient supplies of tunicamycin for subsequent degradation experiments. It should be noted that there was some variation in the yields observed; this is likely due to differences in the precise growth conditions and levels of *tun* gene expression between individual fermentation batches.

3.3 Relay synthesis: Tunicamycin degradation

3.3.1 Putative TunD substrate: N-Acetyl-tunicaminyl-uracil

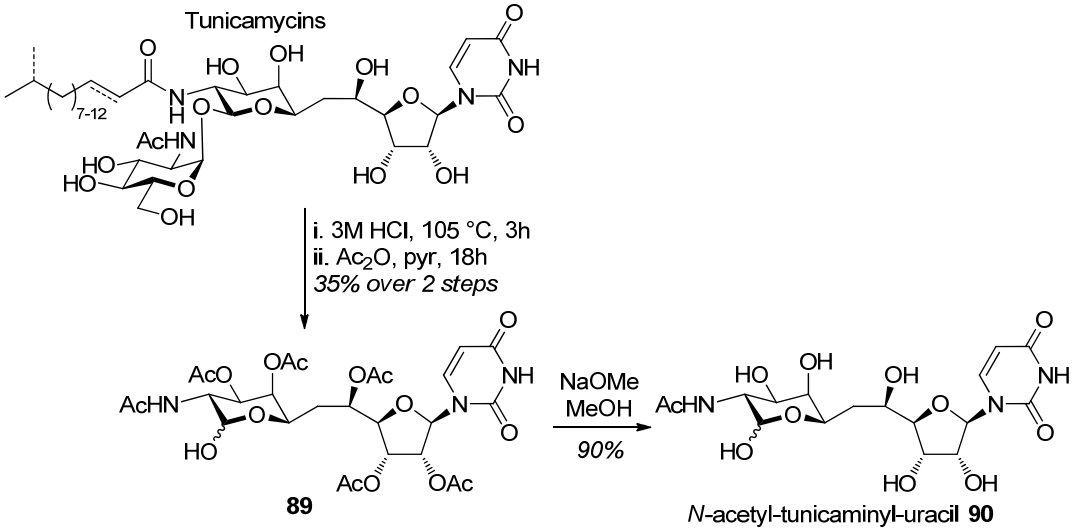
Tunicaminyl-uracil is a vital intermediate in the formation of targeted tunicamycin analogues (see section 1.5.2). It possesses the key tunicamine core of the parent natural product and can potentially be functionalised with a wide variety of subunits at both its free C-11´ anomeric hydroxyl and C-10´ amine groups. *N*-acetyl-tunicaminyl-uracil **90** also features in the putative biosynthetic pathway of the tunicamycins and is proposed to act as a natural substrate for glycosyltransferase TunD (see section 2.4.3).

We have shown that tunicamycin can be isolated in pure form and in preparative quantities from fermentation broths of *S. chartreusis*. This is in contrast to other natural products which are produced by organisms that are extremely difficult to culture or grow very slowly. As such, although a number of synthetic routes to tunicaminyl-uracil derivatives have been developed (see section 1.3), a far more economical route to this intermediate is through degradation of native tunicamycins.

In their 1994 paper on the total synthesis of tunicamycin, Gin *et al*. heated a sample of tunicamycin at reflux in 3 M HCl for 3 h and acetylated the crude mixture to give heptaacetyl-tunicaminyl uracil in 30% yield.[377] Similar cleavage of the tunicamycins' *N*-acyl and terminal GlcNAc moieties were also described by Tamura in his 1982 book, although no yields were reported.[378] As such, tunicamycins were subjected to acidic hydrolysis and subsequent purification and deprotection steps were screened in order to access useful quantities of pure *N*-acetyl-tunicaminyl-uracil **90** in relatively good yields.

Tunicamycins isolated from *S. chartreusis* cultures were first refluxed in 3 M HCl for 3 hours (Scheme 3.1). An aqueous acidic work up was found to be necessary at this stage to remove a large proportion of liberated fatty acids – contaminants difficult to remove later on. The resulting tunicaminyl-uracil shared the same R_f as by-product D-glucosamine by TLC. It was found that acetylation of the crude reaction mixture followed by careful flash chromatography using an optimised solvent gradient was necessary for the separation of these compounds, thus affording heptaacetyl-tunicaminyl uracil **89** in 35% yield. Subsequent deacetylation with freshly prepared sodium methoxide yielded *N*-acetyl-tunicaminyl uracil **90** in 90% yield. The resulting sugar was fully assigned using a combination of NMR experiments, in particular 1D TOCSY with specific irradiation of a

number of isolated spin systems. Informative spectra following irradiation of H-1´$^{\alpha+\beta}$, H-11´$^{\alpha}$, H-11´$^{\beta}$ and H-6´b$^{\alpha+\beta}$ resonances can be found in Appendix A5.



Scheme 3.1 Acidic hydrolysis of isolated tunicamycins

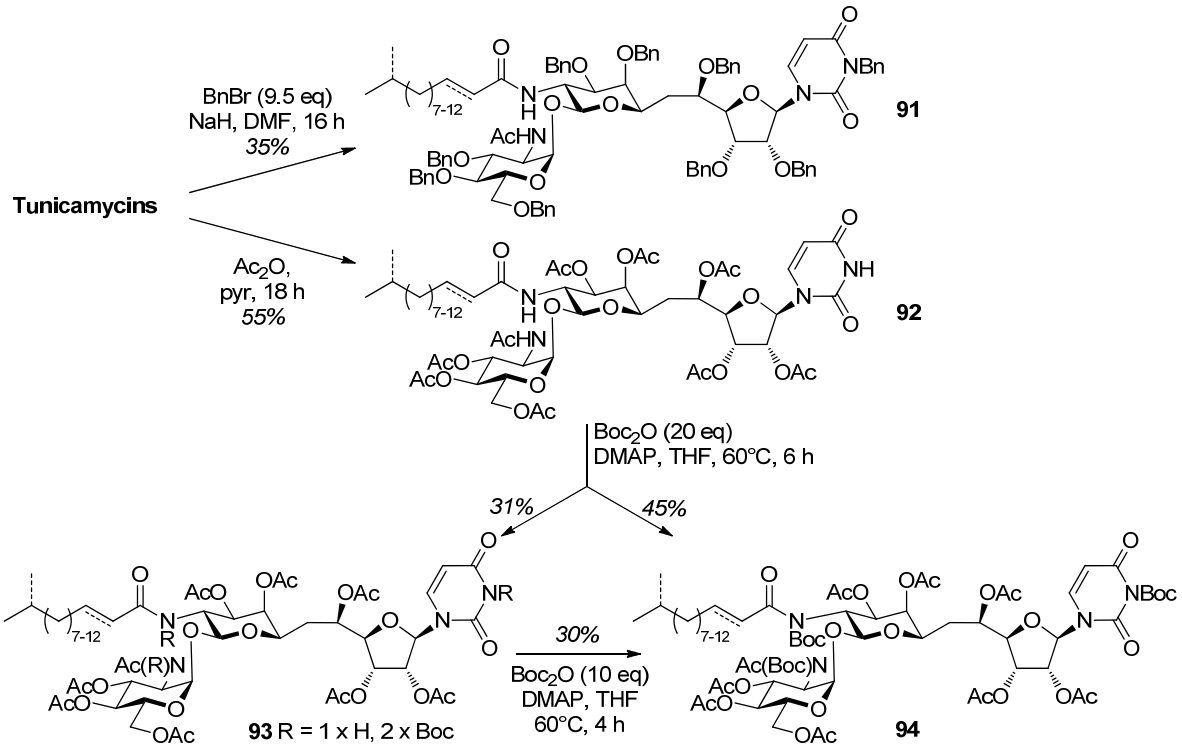
3.3.2 Putative TunE substrate: N-*Acetyl*-N-*deacyl-tunicamycin*

Selective removal of the *N*-acyl chain from tunicamycin would eliminate the heterogeneity arising from different acyl chain lengths present in the isolated natural product. The resulting intermediate would be a valuable intermediate for subsequent diversification into potential GT inhibitors (see section 1.5.2). In $N^{10´}$-acetyl-$N^{10´}$-deacyl-tunicamycin **99**, the fatty acid sidechains of different tunicamycin homologues have all been replaced with a simple *N*-acetyl group. This compound is a valuable intermediate for the *in vitro* characterisation of tunicamycin biosynthetic enzymes since it is proposed to act as a natural substrate for *N*-deacetylase TunE (see section 2.4.4).

The necessary cleavage of an amide bond to afford this transformation is not trivial, owing to the high stability and low reactivity of this functional group. The classic method involves acidic or basic hydrolysis at high temperature, which results in low yields due to the harsh conditions involved. Fortunately, a number of milder strategies have been developed,[11-14] of particular note being Kunieda's *N*-Bocylation methodology.[379] Protecting the amide nitrogen with a Boc group removes electron density from the amide carbonyl, allowing it to be readily cleaved at room temperature with catalytic amounts of base. This strategy has been applied to carbohydrate chemistry, with the successful

cleavage of acetamides in a range *N*-acetyl-hexosamine derivatives.[380,381] The application of this methodology to the tunicamycins is presented below.

The first stage involved selective protection of tunicamycin's eight hydroxyl groups. Since the primary goal was to introduce Boc groups onto the amide nitrogens, a one pot protection of all hydroxyl and amide groups was attempted. Unfortunately, despite a large excess of Boc anhydride and addition of nucleophilic catalysts, the reaction could not be driven to completion and only incompletely protected derivatives were isolated. Silyl ethers were next considered, since they are known to selectively silylate alcohols over amines/amides.[382,383] After subsequent Boc protection of the amides and deacylation, the silyl ethers would then be conveniently orthogonal to the resulting NHBoc groups. A wide range of conditions were screened using a variety of silylating agents based around TES and TBS protecting groups, but in all cases incomplete hydroxyl protection was once again observed. Since the tunicamycin molecule contains a total of eight hydroxyl groups, their complete protection with bulky silyl or Boc protecting groups is likely to lead to significant steric repulsion. For this reason, the incomplete protection observed with these reagents may be attributed to steric crowding. In order to circumvent this problem, less hindered protecting groups were chosen. A variety of benzylation conditions were screened, but satisfactory results were only observed with NaH and BnBr in DMF. Under these conditions all eight hydroxyl groups were protected, although tunicamycin's three amide bonds were also partially benzylated. Careful optimisation of conditions led to the isolation of nonabenzylated tunicamycin **91**, albeit in a low 35% yield (Scheme 3.2). Unfortunately this compound was unreactive to subsequent *N*-Bocylation, presumably because steric crowding remained a problem, preventing access to the amide nitrogens. The best results were obtained with acetyl protection of tunicamycin; octaacetyl-tunicamycin **92** was readily accessed in 55% yield (Scheme 3.2). The acetyl groups would not be stable to methanolysis during later amide deprotection, but this was outweighed by the ease of their installation and superior yields to all other protecting group strategies attempted.



Scheme 3.2 Boc protection of amides within tunicamycin

Exposure of purified octaacetyl-tunicamycin to Boc_2O produced a mixture of di-*N*-Boc and tri-*N*-Boc derivatives **93** and **94** – isolated in 31 and 45% yields respectively (Scheme 3.2). Full conversion to the tri-*N*-Boc species could not be achieved, despite optimising numerous parameters such as temperature, concentration, number of equivalents and solvent, as well as attempting enhancement through microwave irradiation. However, di-*N*-Boc derivative **93** was readily isolated and partially converted to the desired species in 30% yield, providing a salvage pathway for these partially protected by-products.

At this point it was observed that the ^{1}H NMR spectrum of tri-*N*-Boc **94** showed two distinct compounds, suggesting an unidentified impurity might be present. An EXSY NMR experiment was performed, which showed cross peaks between H_α and $H_\alpha{}'$ and between H_β and $H_\beta{}'$ at room temperature (Fig. 3.7). This indicated that the resonances for H_α and H_β were interchangeable with those of $H_\alpha{}'$ and $H_\beta{}'$ respectively and confirmed that the two species observed by NMR were in fact two rotamers of the same compound. It was suspected that the steric bulk around the tertiary amide at N-10′ resulted in restricted rotation about one of the N–C bonds, leading to two distinct isomers by NMR.

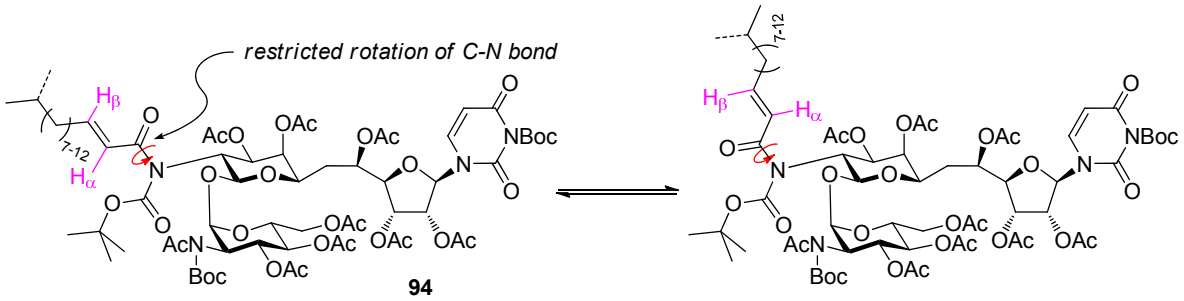
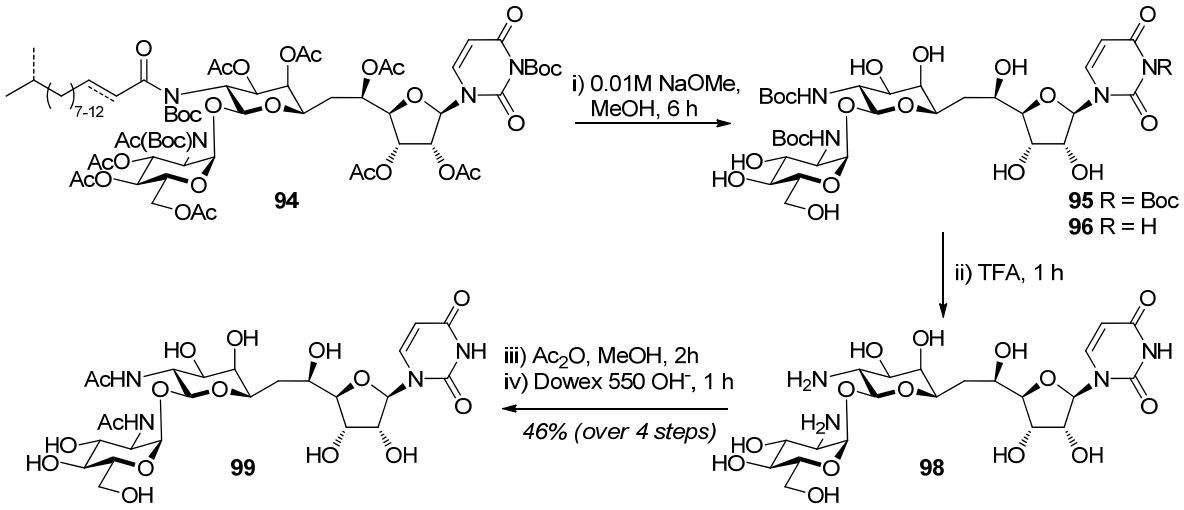


Fig. 3.7 Restricted rotation about $C^{10'}$–N bond accounts for two species observed in ^{1}H NMR of **94**

The tri-*N*-Boc derivative **94** was next subjected to methanolysis with catalytic sodium methoxide. Successful cleavage of both the *N*-acyl and *N*-acetyl moieties was observed alongside the expected deprotection of pendant *O*-acetyl groups, to give deacylated intermediate **95** (Scheme 3.3). Upon prolonged exposure to the basic conditions by-products resulting from methoxide cleavage of the uracil Boc group (**96**) and from amide cleavage within the uracil moiety were detected. Reaction times were reduced to minimise the formation of these by-products. However, under these conditions reactions could not be driven to completion and appreciable quantities of partially acetylated and acylated tunicamycins were observed. It appeared the *O*-acetyl, activated *N*-acetyl and certain *N*-Boc groups harboured very similar reactivities to methanolysis, resulting in a subtle balance between insufficient acyl cleavage leading to partially acylated tunicamycin derivatives and excessive cleavage resulting in the removal of a Boc group or even uracil cleavage.



Scheme 3.3 Selective *N*-deacylation of tunicamycin

Following these experiments, the strategy adopted involved homogenising this mixture of products by removing all Boc groups and then selectively acetylating the resulting primary amines to give the desired deacylated tunicamycin derivative **99**. As such, a multi-step procedure was developed, requiring only a single purification step (Scheme 3.3). Tri-*N*-Boc derivative **94** was exposed to freshly prepared 0.01M NaOMe in MeOH. After 6 h, TLC showed formation of a major product tri-*N*-Boc **95**, together with di-*N*-Boc **96** and partially acetylated/acylated byproducts. The quenched reaction mixture was then treated with TFA, yielding diamine intermediate **98**. Following evaporation of TFA, selective acetylation of the free amine groups was achieved with Ac_2O in MeOH. Initially, only a single acetylation event was observed despite prolonged reaction times and an excess of Ac_2O. The diamine intermediate likely existed as a TFA salt, thereby precluding one of the amines from acetylation. Upon adjusting the pH to 6-7 with Dowex 550 OH^- resin, the TFA counterion was removed and the second amine group successfully acetylated. After this sequence of reactions, and following careful chromatography to remove fatty acid contaminants, proposed TunE substrate **99** was isolated in 46% yield over 4 steps. The resulting sugar was fully assigned using a combination of NMR experiments, in particular 1D TOCSY with specific irradiation of anomeric, H-5´´ and H-6´b protons. Informative spectra can be found in Appendix A5.

3.4 Conclusions

In this chapter we have successfully optimised methods for the preparative scale isolation, purification and characterisation of tunicamycins from the fermentation broths of *S. chartreusis*. Isolated yields up to 35 mg/L culture were obtained, providing sufficient natural product for degradation studies. Subsequently, advanced intermediates **90** and **99** were obtained through a combination of acidic hydrolysis, mild methanolysis and effective protecting group strategies. Both compounds are valuable intermediates for the relay synthesis of tunicamycin analogues. Additionally, *N*-acetyl-tunicaminyl-uracil **90** and *N*-acetyl-*N*-deacyl-tunicamycin **99** are proposed to act as natural substrates for enzymes TunD and TunE respectively from the tunicamycin biosynthetic pathway. They will be extremely valuable for use as predicted native substrates in the *in vitro* functional characterisation of these enzymes. Overall, results presented in this chapter complement the characterisation of the tunicamycin biosynthetic pathway described in chapters 2 and 4, as well as the eventual formation of tunicamycin analogues based around the core tunicamine scaffold – outlined in section 1.5.

3.5 Experimental

3.5.1 Biological protocols

3.5.1.1 General considerations

All solutions and equipment used in the handling of microbial cultures were autoclaved or sterilised. Microbial manipulations of *Streptomyces* were performed in a HERAsafe KS-12 Class II Laminar Flow Cabinet. Media, plastic and glassware were autoclaved at 121 °C for 20 min. Bacterial cultures were incubated in a New Brunswick Series 25 shaker. Centrifugation was performed in Beckman Coulter Avanti J-25, Beckman GS-6R or Mistral 1000 centrifuges. Microcentrifugation was performed in an Eppendorf Centrifuge 5415R or a Beckman Coulter Microfuge 18. Water was deionised and passed through a Millipore 0.22 μm filter prior to use. HPLC grade MeCN and MeOH were obtained from Rathburn Chemicals Ltd and used as supplied.

3.5.1.2 Bacterial strains

S. chartreusis NRRL 3882 was obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ). The strain was maintained on solid TYD agar and grown aerobically in liquid TYD medium.[373]

3.5.1.3 Analytical scale isolation of tunicamycins

Dense *S. chartreusis* spore suspension (10 μL) and 1% Antifoam 204 solution (100 μL) were added to 50 mL of TYD media in a 250 mL spring coiled flask and incubated at 28 °C and 200 rpm for 6-8 days. After this time, the cultures adopted a characteristic red colour which was used as a qualitative indicator of tunicamycin production. Cultures were acidified with conc. HCl to a final concentration of 0.2 M and cells were collected by centrifugation (8000 × *g*, 10 min). The cell pellet was washed with 0.2 M aq. HCl, resuspended and again collected by centrifugation. The cell pellet was then vortexed with MeOH, collected by centrifugation and the supernatant discarded. This step was repeated three times and the combined MeOH extracts were concentrated *in vacuo* before a final centrifugation step to remove any remaining cell debris.

3.5.1.4 Preparative scale isolation of tunicamycins

50 mL of TYD media in a 250 mL spring coiled flask was inoculated with dense *S. chartreusis* spore suspension (10 µL) and incubated at 28 °C and 200 rpm. After 24 h, aliquots of this culture (12 × 1 mL) were used to inoculate 12 × 1 L of TYD media in unbaffled 2 L conical flasks, which were subsequently incubated at 28°C and 200 rpm for 6-8 days. After this time the cultures adopted a characteristic red colour which was used as a qualitative indicator of tunicamycin production. Cells were collected by centrifugation (8000 × *g*, 10 min).

Extraction from cell culture supernatant: Amberlite XAD-16 was preconditioned by washing with MeOH (× 3), then distilled water (× 2). This preconditioned resin (15 g/L) was added to the cell culture supernatant and stirred for 2 h. The magnetic stirrer was then turned off and the XAD-16 resin allowed to settle to the bottom of the flask, after which the majority of the supernatant was decanted along with any remaining cell debris. The remaining supernatant was removed by filtration through cotton wool. The combined resin was next stirred in water (200 mL, 10 min) and filtered through cotton wool. This process was repeated with water (2 × 200 mL), MeOH (2 × 200 mL) and ^{i}PrOH (2 × 200 mL) with a final wash in MeOH (200 mL, 60 min). TLC analysis (H_2O/^{i}PrOH/EtOAc 1:2:4) identified tunicamycin containing fractions (R_f = 0.4), which were combined and concentrated *in vacuo*. The residue was washed with 1 M aq. HCl (2 × 50 mL) and the collected by centrifugation (3500 × *g*, 5 min).

Extraction from cell pellet: The cell pellet was stirred in 1M aq. HCl (800 mL, 20 min) and the cells collected by centrifugation (8000 × *g*, 10 min). This process was repeated, after which the cell pellet was stirred in MeOH (400 mL, 2 h). The cells were collected by centrifugation (3500 × *g*, 5 min) and the supernatant decanted. The cells were then resuspended in MeOH (400 mL) and vortexed for 5 min, prior to centrifugation (3500 × *g*, 5 min) and collection of the supernatant. The resulting MeOH fractions were combined and subjected to further purification.

Purification of combined extracts: The cell pellet extract and culture supernatant extract were combined and dissolved in MeOH, then dried onto silica *in vacuo*. Flash chromatography (H_2O/*i*PrOH/EtOAc 1:3:6 → 1:2:4) afforded tunicamycin (~ 400 mg) as a off-white powder, which was freezdried from water prior to use.

3.5.2 Analytical Methods

3.5.2.1 Analytical scale HPLC

The HPLC system consisted of a Waters 2795 separations module coupled to a Waters 2996 photodiode array detector, controlled by Empower software. Reversed-phase chromatography was achieved on a Phenomenex Gemini C-18 column (250 mm, 5 mm, 5 μm particle size), protected with an in-line Phenomenex C-18 guard column. Solvents used in the mobile phase were as follows: Solvent A – H_2O, 5% MeCN, 0.1% formic acid; Solvent B – MeCN, 0.1% formic acid. An isocratic elution was performed (1 mL/min; 55% A, 25 min), with a typical injection volume of 20 μl and detection at 260 nm.

3.5.2.2 LC/MS methods

The LC/MS system consisted of a Micromass LCT (ESI-TOF-MS) coupled to a Waters Alliance 2790 HPLC separations module and Waters 996 photodiode array. Reversed-phase chromatography was achieved using a Phenomenex Gemini C-18 column (250 mm, 5 mm, 5 μm particle size) protected with an in-line Phenomenex C-18 guard column. Solvents used in the mobile phase were as follows: Solvent A – H_2O, 5% MeCN, 0.1% formic acid; Solvent B – MeCN, 0.1% formic acid. An isocratic elution was performed (1 mL/min; 55% A, 25 min), with a typical injection volume of 20 μl and detection at 260 nm. The electrospray source of the LCT was operated with a capillary voltage of 3 kV and a cone voltage of 10-65 V. Nitrogen was used as a nebuliser and desolvation gas at a total flow of 400 L/h. Spectra were processed using MassLynx 4.1 software. Myoglobin (horse heart) was used as a standard to construct calibration curves.

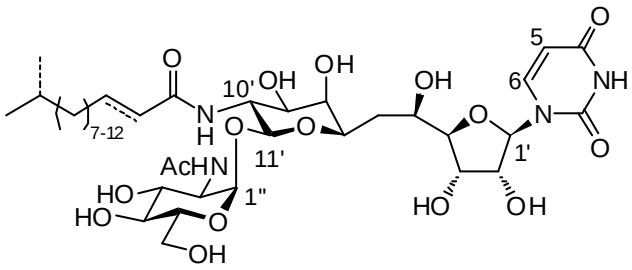
3.5.3 Chemical synthesis

3.5.3.1 General Considerations

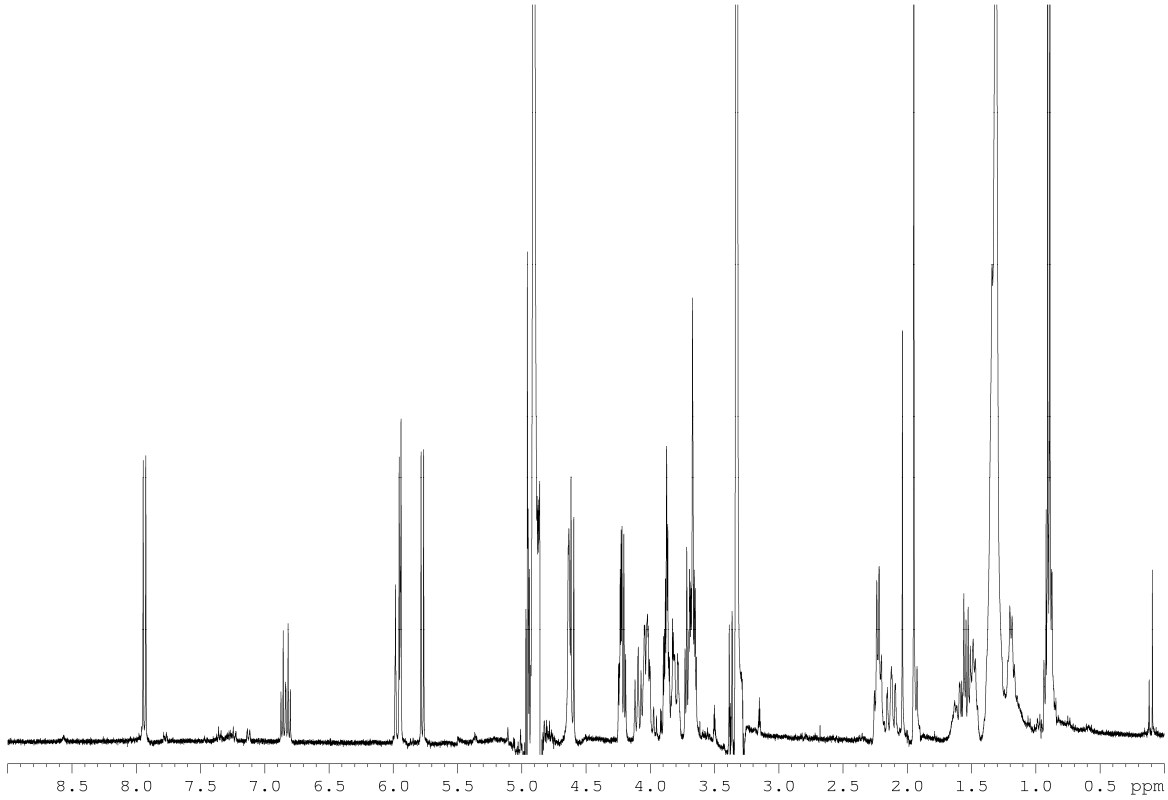
Melting points were recorded on a Kofler hot block and are uncorrected. Proton nuclear magnetic resonance (δ_H) spectra were recorded on a Bruker DPX 200 (200 MHz), Bruker DPX 400 (400 MHz), Bruker DQX 400 (400 MHz) or Bruker AC 500 (500 MHz) spectrometer. Carbon nuclear magnetic resonance spectra were recorded on a Bruker DPX 200 (50 MHz), Bruker DQX 400 (100 MHz) spectrometer or on a Bruker AC 500 (125 MHz) with a ^{13}C cryoprobe (125 MHz). Spectra were fully assigned using a combination of COSY, HMQC, and DEPT 135. All chemical shifts were quoted on the δ- scale in ppm using residual solvent as the internal standard. Coupling constants (*J*) are reported in hertz (Hz). Infrared spectra were recorded on a Bruker Tensor 27 Fourier Transform spectrophotometer with absorption maxima recorded in wavenumbers (cm^{-1}). Oils were analysed as thin films and solids as potassium bromide disks. Low resolution mass spectra were recorded on an LCT Premier XE using electrospray ionisation (ES). High resolution mass spectra were recorded on a Bruker microTOF. Specific rotations were measured on a Perkin Elmer 241 polarimeter with a pathlength of 1.0 dm with concentrations (*c*) given in g/100 mL. Thin layer chromatography (TLC) was carried out on Merk Kieselgel $60F_{254}$ precoated aluminium backed plates and visualised with a combination of the following: 254 nm UV lamp; aqueous $KMnO_4$ (5% $KMnO_4$ in 1 M NaOH); aqueous phosphomolybdic acid and Ce(IV) (2.5% phosphomolybdic acid hydrate, 1% cerium(IV) sulfate hydrate, and 6% H_2SO_4); ammonium molybdate (5% in 2M H_2SO_4); and/or sulfuric acid (2 M in ethanol/water 1:1). Flash column chromatography was carried out with Fluka Kiegselgel 60 220-440 mesh silica gel. THF and DCM (HPLC grade) were dried through a column of Al_2O_3. Remaining anhydrous solvents were purchased from Aldrich or Fluka and stored over molecular sieves (<0.005% H_2O). All other solvents were used as supplied (analytical or HPLC grade), without further purification. Petrol refers to the fraction of petroleum ether boiling in the range 40-60 °C. Brine refers to a saturated solution of sodium chloride. Imidazole was recrystallised from ethanol.

3.5.3.2 Synthetic procedures

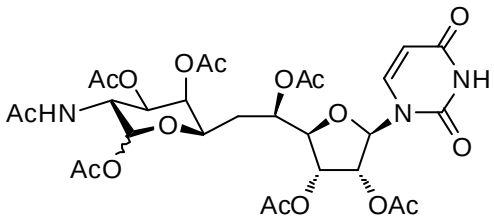
Tunicamycins[378]



Tunicamycin was isolated from *S. chartreusis* fermentation broth as previously described. Apart from HPLC and LC/MS analysis, samples were also subjected to characterisation by TLC and ^{1}H NMR; **TLC** : R_f 0.3 in water/isopropanol/ethyl acetate (W.IP.E) 1:3:6; **^{1}H NMR** (500 MHz, CD_3OD) δ 0.89, 0.90 (2 × s, 2 × 3 H, -CH(C*H*$_3$)$_2$), 1.16 - 1.61 (m, n × $CH_2^{\text{fatty acid}}$), 1.95 (s, 3 H, -CH_3^{NHAc}), 3.35 - 4.05 (m, -CH_2^{sugar}, -CH^{sugar}), 4.10 (t, J = 9.3 Hz, 1 H, H-10´), 4.21 (t, $J_{2',1'}$ = 5.7 Hz, J = 5.7 Hz, 1 H, H-2´), 4.59 - 4.63 (m, H-11´), 4.95 (d, J = 3.2 Hz, 1 H, H-1´´), 5.77 (d, $J_{5,6}$ = 8.2 Hz, 1 H, H-5$^{\text{uracil}}$), 5.95 (d, $J_{1',2'}$ = 5.7 Hz, 1 H, H-1´), 5.96 (d, $^{3}J_{\text{HC=CH trans}}$ = 15.4 Hz, 1 H, =C*H*C(O)-), 6.84 (dt, $^{3}J_{\text{HC=CH trans}}$ = 15.4 Hz, J = 6.9 Hz, 1 H, -CH_2*H*C=), 7.93 (d, $J_{6,5}$ = 8.2 Hz, 1 H, H-6$^{\text{uracil}}$), assignments were assisted by data from COSY, HSQC and 3D TOCSY spectra, the presence of at least 9 homologues prevented full assignment, resulting in broad multiplets in the δ3 - 4 (sugar C-H) and δ1.1 - 1.6 (fatty acid CH_2) regions;



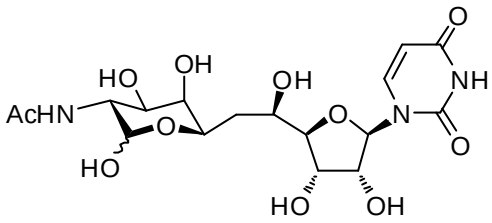
Heptaacetyl-tunicaminyl-uracil 89[377]



Tunicamycin (108 mg, 0.220 mmol) was suspended in 3M aq. HCl (15 mL) and refluxed at 105 °C. After 3 h the solvent was removed *in vacuo*, and the residue dissolved in dry pyridine (3 mL) and acetic anhydride (3 mL). After 24 h at room temperature, TLC (MeOH/EtOAc 1:19) showed two products at R_f 0.3 (**89**) and 0.35 (pentaacetyl-D-glucosamine). After removal of solvent *in vacuo*, flash chromatography (MeOH/EtOAc 1:19), afforded heptaacetyltunicaminyluracil **89** (54 mg, 0.077 mmol, 35%, α/β ratio 56:44) as a white foam; **^{1}H NMR** (500 MHz, $CDCl_3$) δ ppm 8.82 (d, J = 1.9 Hz, 1 H, N-H$^{\text{uracil, }\beta}$), 8.70 (d, J = 2.0 Hz, 1 H, N-H$^{\text{uracil, }\alpha}$), 7.22 (d, $J_{6,5}$ = 8.2 Hz, 1 H, H-6$^{\text{uracil, }\alpha}$), 7.19 (d, $J_{6,5}$ = 8.2 Hz, 1 H, H-6 $^{\text{uracil, }\beta}$), 6.14 (d, $J_{11',10'}$ = 3.8 Hz, 1 H, H-11′$^{\alpha}$), 5.91 (d, $J_{1',2'}$ = 5.7 Hz, 1 H, H-1′$^{\alpha}$), 5.82 (d, $J_{1',2'}$ = 5.0 Hz, 1 H, H-1′$^{\beta}$), 5.81 (m, J = 2.0 Hz, $J_{5,6}$ = 8.2 Hz, 1 H, H-5$^{\text{uracil, }\alpha}$), 5.78 (dd, J = 1.9 Hz, $J_{5,6}$ = 8.2 Hz, 1 H, H-5$^{\text{uracil, }\beta}$), 5.71 (d, $J_{\text{N-H},10'}$ = 9.8 Hz, 1 H, N-H$^{\text{Ac, }\beta}$), 5.64 (d, $J_{11',10'}$ = 8.8 Hz, 1 H, H-11′$^{\beta}$), 5.50 (d, $J_{\text{N-H, }10'}$ = 9.5 Hz, 1 H, N-H$^{\text{Ac, }\alpha}$), 5.43 (dd, $J_{3',4'}$ = 4.9 Hz, $J_{3',2'}$ = 6.1 Hz, 1 H, H-3′$^{\beta}$), 5.37 (dd, $J_{3',4'}$ = 4.7 Hz, $J_{3',2'}$ = 5.8 Hz, 1 H, H-3′$^{\alpha}$), 5.34 (app t, $J_{2',1'}$ = 5.7 Hz, $J_{2',3'}$ = 5.8 Hz, 1 H, H-2′$^{\alpha}$), 5.30 (app t, $J_{2',1'}$ = 5.0, $J_{2',3'}$ = 6.1 Hz, 1 H, H-2′$^{\beta}$), 5.24 (dd, $J_{9',8'}$ = 2.8 Hz, $J_{9',10'}$ = 11.5 Hz, 1 H, H-9′$^{\alpha}$), 5.22 (dd, $J_{8',7'}$ = 1.0 Hz, $J_{8',9'}$ = 3.2 Hz, 1 H, H-8′$^{\beta}$), 5.20 (m, 1 H, H-5′$^{\beta}$), 5.19 (dd, $J_{8',7'}$ = 1.0 Hz, $J_{8',9'}$ = 2.8 Hz, 1 H, H-8′$^{\alpha}$), 5.11 (ddd, J = 2.8 Hz, $J_{5',4'}$ = 4.7 Hz, $J_{5',6b'}$ = 8.3 Hz, 1 H, H-5′$^{\alpha}$), 5.08 (dd, $J_{9',8'}$ = 3.2 Hz, $J_{9',10'}$ = 11.0 Hz, 1 H, H-9′$^{\beta}$), 4.71 (ddd, $J_{10',11'}$ = 3.8 Hz, $J_{10',\text{N-H}}$ = 9.5 Hz, $J_{10',9'}$ = 11.5 Hz, 1 H, H-10′$^{\alpha}$), 4.42 (ddd, $J_{10',9'}$ = 8.8 Hz, $J_{10',\text{N-H}'}$ = 9.8 Hz, $J_{10',9'}$ = 11.0 Hz, 1 H, H-10′$^{\beta}$), 4.13 (app t, $J_{4',5'}$ = $J_{4',3'}$ = 4.9 Hz, 1 H, H-4′$^{\beta}$), 4.08 (app t, $J_{4',5'}$ = $J_{4',3'}$ = 4.7 Hz, 1 H, H-4′$^{\alpha}$), 4.07 (m, 1 H, H-7′$^{\alpha}$), 3.92 (ddd, $J_{7',8'}$ = 1.0 Hz, $J_{7',6b'}$ = 3.2 Hz, J = 9.6 Hz, 1 H, H-7′$^{\beta}$), 2.20 (s, 3 H, -CH$_3^{\text{Ac, }\alpha}$), 2.20 (s, 3 H, -CH$_3^{\text{Ac, }\beta}$), 2.18, 2.13 (2 × s, 2 × 3 H, 2 × -CH$_3^{\text{Ac, }\alpha}$), 2.12, 2.12 (2 × s, 2 × 3 H, 2 × -CH$_3^{\text{Ac, }\beta}$), 2.11 (s, 6 H, 2 × -CH$_3^{\text{Ac, }\beta}$), 2.09 (s, 6 H, 2 × -CH$_3^{\text{Ac, }\alpha}$), 2.03 (s, 3 H, -CH$_3^{\text{Ac, }\alpha}$), 2.01 (s, 3 H, -CH$_3^{\text{Ac, }\beta}$), 2.00 - 2.02 (m, 1 H, H-6a′$^{\beta}$), 1.97 - 1.99 (m, 1 H, H-6a′$^{\alpha}$), 1.96 (s, 3 H, -CH$_3^{\text{NHAc, }\alpha}$), 1.94 (s, 3 H, -CH$_3^{\text{NHAc, }\beta}$), 1.75 (ddd, $J_{6b',7'}$ = 3.2 Hz, J = 8.2 Hz, $^2J_{6b',6a'}$ = 14.8 Hz, 1 H, H-6b′$^{\beta}$), 1.57 (ddd, J = 1.7 Hz, $J_{6b',5'}$ = 8.3 Hz, $^2J_{6b',6a'}$ = 14.9 Hz, 1 H, H-6b′$^{\alpha}$); **^{13}C NMR** (126 MHz, MeOD-d_4) δ ppm 171.2, 170.7, 170.6, 170.6, 170.4, 170.2, 170.1, 170.0, 169.7, 169.6, 169.5, 169.3, 169.3, 169.1 (14 × s, 14 × C, C=O$^{\text{NHAc, }\alpha}$, C=O$^{\text{NHAc, }\beta}$, 6 × C=O$^{\text{Ac, }\alpha}$, 6 × C=O$^{\text{Ac, }\beta}$), 162.5 (C-4 C=O$^{\beta}$), 162.5 (C-4 C=O$^{\alpha}$), 149.9 (C-2 C=O$^{\alpha}$), 149.9 (C-2 C=O$^{\beta}$),

140.2 (C-6$^{uracil,\ \beta}$), 139.8 (C-6$^{uracil,\ \alpha}$), 103.4 (C-5$^{uracil,\ \alpha}$), 103.3 (C-5$^{uracil,\ \beta}$), 93.0 (C-11´$^{\beta}$), 91.1 (C-11´$^{\alpha}$), 88.7 (C-1´$^{\beta}$), 88.1 (C-1´$^{\alpha}$), 82.5 (C-4´$^{\beta}$), 82.4 (C-4´$^{\alpha}$), 72.4 (C-2´$^{\alpha}$), 72.4 (C-2´$^{\beta}$), 70.9 (C-7´$^{\beta}$), 70.5 (C-9´$^{\beta}$), 69.6, 69.6, 69.3, 68.7, 68.1 (5 × s, 7 × C, C-3´$^{\alpha}$, C-3´$^{\beta}$, C-5´$^{\alpha}$, C-5´$^{\beta}$, C-8´$^{\alpha}$, C-8´$^{\beta}$, C-9´$^{\alpha}$)67.6 (C-7´$^{\alpha}$), 49.5 (C-10´$^{\beta}$), 46.7 (C-10´$^{\alpha}$), 32.5 (C-6´$^{\alpha}$), 31.6 (C-6´$^{\beta}$), 23.3 (-$CH_3^{NHAc,\ \beta}$), 23.2 (-$CH_3^{NHAc,\ \alpha}$), 20.4 – 21.0 (12 × C, 6 × -$CH_3^{Ac,\ \alpha}$, 6 × -$CH_3^{Ac,\ \beta}$); **IR** (thin film) v : 3500-3700 (multiple w, N-H), 1747 (br s., C=O^{Ac}), 1717 (s, unsat. C=O$^{2°amide\ I}$), 1698 (s, C=O$^{2°amide\ I}$), 1684 (s, C=O$^{2°amide\ I}$), 1653 (s, C=C), 1636 (s, C=C), 1558 (m, unsat. C=O$^{2°amide\ II}$), 1540 (m, C=O$^{2°amide\ II}$), 1521 (m, C=O$^{2°amide\ II}$), 1374 (s), 1224 (br. s); **MS** m/z (ESI$^+$) : 717 [(M+ NH_4)$^+$, 80%], 722 [(M+ Na)$^+$, 100%]; (ESI$^-$) : 698 [(M-H)$^-$, 100%]; **HRMS** m/z (ESI$^+$) : calc. for $C_{29}H_{37}N_3NaO_{17}$ (M + Na)$^+$ = 722.2015. Found 722.2019;

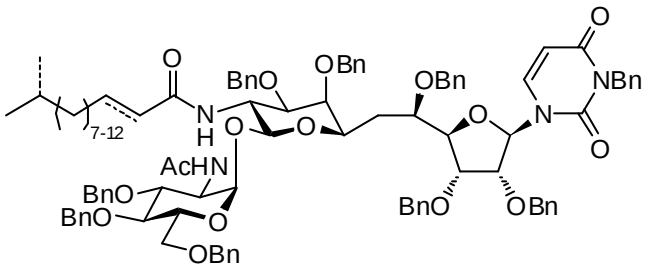
***N*-acetyl-tunicaminyl-uracil 90**



Heptaacetyltunicaminyl-uracil **89** (54 mg, 0.077 mmol) was dissolved in dry MeOH (5 mL) and the solution cooled to 0 °C. 1M NaOMe in MeOH (50 μL) was added, resulting in a 0.01M concentration of base. After 3 h, TLC (W.IP.E. 1:2:2) showed complete conversion to two spots at R_f 0.4 (one anomer of **90**), and 0.3 (the other anomer of **90**). The reaction was neutralised carefully with DOWEX 50WX8 H^+ resin and diluted with MeOH. After filtration through cotton the solvent was removed *in vacuo*. Flash chromatography (W.IP.E. 1:2:2) afforded *N*-acetyl-tunicaminyl-uracil **90** (31 mg, 0.069 mmol, 90%) as a white amorphous solid; $\mathbf{[\alpha]_D^{18}}$ = -7, c = 1, H_2O; **^{1}H NMR** (500 MHz, D_2O) δ ppm 7.82 (d, $J_{6,5}$ = 8.0 Hz, 1 H$^{\alpha}$, H-6$^{uracil,\ \alpha}$), 7.81 (d, $J_{6,5}$ = 8.0 Hz, 1 H$^{\beta}$, H-6$^{uracil,\ \beta}$), 5.90 (m, 1 H$^{\alpha}$ + 1 H$^{\beta}$, H-1´$^{\alpha}$, H-1´$^{\beta}$), 5.87 (d, $J_{5,6}$ = 8.0 Hz, 1 H$^{\alpha}$ + 1 H$^{\beta}$, H-5$^{uracil,\ \alpha}$, H-5$^{uracil,\ \beta}$), 5.18 (d, $J_{11´,10´}$ = 3.7 Hz, 1 H$^{\alpha}$, H-11´$^{\alpha}$), 4.62 (d, $J_{11´,10´}$ = 8.5 Hz, 1 H$^{\beta}$, H-11´$^{\beta}$), 4.27 - 4.32 (m, 2 H$^{\alpha}$ + 2 H$^{\beta}$, H-2´$^{\alpha}$, H-2´$^{\beta}$, H-3´$^{\alpha}$, H-3´$^{\beta}$), 4.24 (d, J = 9.1 Hz, 1 H$^{\alpha}$, H-7´$^{\alpha}$), 4.08 (dd, $J_{10´,9´}$ = 11.3 Hz, $J_{10´,11´}$ = 3.7 Hz, 1 H$^{\alpha}$, H-10´$^{\alpha}$), 4.04 (dt, J = 10.7 Hz, J = 2.8 Hz, 1 H$^{\beta}$, H-5´$^{\beta}$), 3.96 - 4.01 (m, 2 H$^{\alpha}$ + 1 H$^{\beta}$, H-4´$^{\alpha}$, H-4´$^{\beta}$, H-5´$^{\alpha}$), 3.92 (dd, $J_{9´,10´}$ = 11.3 Hz, $J_{9´,8´}$ = 3.2 Hz, 1 H$^{\alpha}$, H-9´$^{\alpha}$), 3.84 (d, $J_{8´,9´}$ = 3.2 Hz, 1 H$^{\alpha}$, H-8´$^{\alpha}$), 3.82 (dd, $J_{10´,9´}$ = 10.8 Hz, $J_{10´,11´}$ = 8.5 Hz, 1 H$^{\beta}$, H-10´$^{\beta}$), 3.79 (d, J = 9.3 Hz, 1 H$^{\beta}$, H-7´$^{\beta}$), 3.77 (d, $J_{8´,9´}$ = 3.5 Hz, 1 H$^{\beta}$, H-8´$^{\beta}$), 3.72 (dd, $J_{9´,10´}$ = 10.8 Hz, $J_{9´,8´}$ = 3.5 Hz, 1 H$^{\beta}$, H-9´$^{\beta}$), 2.02 (s, 3 H$^{\alpha}$ + 3 H$^{\beta}$, -$CH_3^{NHAc,\ \beta}$, -$CH_3^{NHAc,\ \alpha}$), 1.90 - 2.00 (m, 1 H$^{\alpha}$ + 1 H$^{\beta}$, H-6a´$^{\alpha}$, H-6a´$^{\beta}$), 1.59 - 1.67 (m, 1 H$^{\alpha}$ + 1 H$^{\beta}$, H-6b´$^{\alpha}$, H-6b´$^{\beta}$), please see appendix for 1D TOCSY spectra of irradiated anomeric protons

and H-6´b which were used to assign overlapping peaks; **^{13}C NMR** (126 MHz, D_2O) δ ppm 176.1, 175.8 ($C{=}O^{NHAc,\ \beta}$, $C{=}O^{NHAc,\ \alpha}$), 167.3 (C-4 $C{=}O^{\alpha+\beta}$), 153.0 (C-2 $C{=}O^{\alpha+\beta}$), 142.9 ($C\text{-}6^{uracil,\ \alpha+\beta}$), 103.6 ($C\text{-}5^{uracil,\ \alpha+\beta}$),96.5 (C-11´$^{\beta}$), 92.0 (C-11´$^{\alpha}$), 89.2 (C-1´$^{\alpha+\beta}$), 88.2, 88.2 (C-4´$^{\alpha}$, C-4´$^{\beta}$), 74.6, 74.5 (C-2´$^{\alpha}$, C-2´$^{\beta}$), 72.4, 72.2, 72.0, 71.5 (C-7´$^{\beta}$, C-8´$^{\beta}$, C-9´$^{\beta}$, C-8´$^{\alpha}$), 70.0, 70.0 (C-3´$^{\alpha}$, C-3´$^{\beta}$), 68.7 (C-9´$^{\alpha}$), 68.2, 68.2 (C-5´$^{\alpha}$, C-5´$^{\beta}$), 67.4 (C-7´$^{\alpha}$), 54.7 (C-10´$^{\beta}$), 51.3 (C-10´$^{\alpha}$), 34.8, 34.7 (C-6´$^{\alpha}$, C-6´$^{\beta}$), 23.3, 23.1 ($\text{-}CH_3^{NHAc,\ \alpha}$, $\text{-}CH_3^{NHAc,\ \beta}$); **IR** (thin film) v : 3383 (b, O-H), 2928 (w, C-H), 1688 (s, $C{=}O^{amide}$), 1558 (m, $N\text{-}H^{amide}$), 1468 (m), 1417 (m), 1383 (m); **MS** m/z; (ESI$^-$) : 446 [$(M\text{-}H)^-$, 100%]; **HRMS** m/z (ESI$^-$) : calc. for $C_{17}H_{24}N_3O_{11}$ $(M - H)^-$ = 446.1416. Found 446.1415;

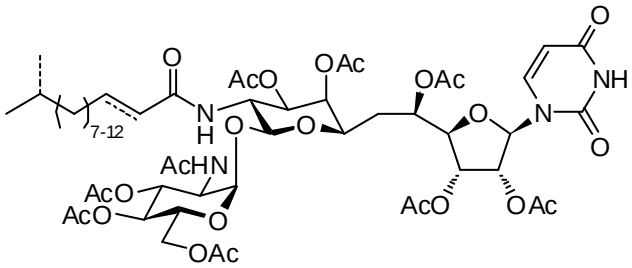
***N*-Benzyl-octa-*O*-benzyl-tunicamycin 91**



Tunicamycin (34 mg, 0.040 mmol) was dissolved in dry DMF (680 µL) and the solution cooled to 4 °C. A 60% dispersion of sodium hydride in mineral oil (16 mg, 0.400 mmol) was added and the effervescing solution stirred for 10 min. Benzyl bromide (45 µL, 0.380 mmol) was then added. After stirring for 16 h at room temperature, TLC (EtOAc/Petrol 1:1) showed a spots at R_f 0.8 (BnBr), 0.6 (BnOH) and 0.3 (**91**). The reaction was quenched with TFA (1 drop) and the solvent removed *in vacuo*. Flash chromatography (EtOAc/Petrol 3:17→3:2→1:0), afforded *N*-benzyl-octa-*O*-benzyl-tunicamycin **91** (23 mg, 0.014 mmol, 35%) as a clear glass containing a mixture of compounds with different aliphatic chain lengths; **^{1}H NMR** (500 MHz, $CDCl_3$) δ ppm 7.49 (m, 3 H, 3 × $C\text{-}H^{Ar}$), 7.09 - 7.40 (m, $C\text{-}H^{Ar}$, $H\text{-}6^{uracil}$), 6.74 (dt, J = 6.9 Hz, $^3J_{HC=CH\ trans}$ = 15.1 Hz, 1 H, C=C*H*-CH_2), 6.03 (d, $J_{1´,2´}$ = 1.9 Hz, 1 H, H-1´), 5.75 (d, J = 9.8 Hz, 1 H, 2´´-NH), 5.54 (d, $^3J_{HC=CH\ trans}$ = 15.1 Hz, 1 H, C=C*H*-CO), 5.08 (m, 2 H, PhCH_2-N) HMBC couplings show this benzyl group to be on $N\text{-}3^{uracil}$, 5.05 (d, $J_{1´´,2´´}$ = 3.5 Hz, 1 H, H-1´´), 4.91 (d, J = 8.2 Hz, 1 H, $H\text{-}5^{uracil}$), 4.88 (d, ^2J = 11.7 Hz, 1 H, PhCH_2-O), 4.82 (d, ^2J = 11.0 Hz, 1 H, PhCH_2-O), 4.80 (d, ^2J = 11.7 Hz, 1 H, PhCH_2-O), 4.77 (d, $J_{11´,10´}$ = 8.5 Hz, 1 H, H-11´), 4.73 (s, 2 H, PhCH_2-O), 4.66 (d, ^2J = 11.3 Hz, 1 H, PhCH_2-O), 4.59 (m, 3 H, PhCH_2-O), 4.50 (m, 3 H, PhCH_2-O), 4.44 (d, ^2J = 13.6 Hz, 1 H, PhCH_2-O), 4.41 (d, ^2J = 11.7 Hz, 1 H, PhCH_2-O), 4.37 (d, ^2J = 12.0 Hz, 1 H, PhCH_2-O), 4.29 (d, ^2J = 11.7 Hz, 1 H, PhCH_2-O), 4.26 (dd, $J_{2´´,1´´}$ = 3.5 Hz, $J_{2´´,2´´\text{-}NH}$ = 9.8 Hz, 1 H, H-2´´), 4.19 (dd, $J_{2´,1´}$ = 1.9 Hz, $J_{2´,3´}$ = 7.6 Hz, 1 H, H-2´), 4.13 (dd, J = 5.7 Hz, $J_{5´´,6´´a}$ = 10.1 Hz, 1 H, H-5´´), 4.05 (dd, $J_{3´,4´}$ = 5.4 Hz, $J_{3´,2´}$ = 7.6 Hz, 1 H, H-3´), 4.01 (d, J = 9.5 Hz, 1 H, H-7´), 3.90 -

3.98 (m, 2 H, H-8´, H-10´), 3.78 (dd, $J_{4´,5´}$ = 2.2 Hz, $J_{4´,3´}$ = 5.4 Hz, 1 H, H-4´), 3.75 (m, 1 H, H-3´´), 3.73 (d, $J_{6´´a,5´´}$ = 10.1 Hz, 1 H, H-6´´a), 3.58 - 3.64 (m, 3 H, H-4´´, H-9´, H-6´´b), 3.55 (dd, $J_{5´,4´}$ = 2.2 Hz, J = 10.3 Hz, 1 H, H-5´), 2.01 - 2.15 (m, 3 H, H-6´a, -CH_2CH=C), 1.80 (m, 1 H, H-6´b), 1.77 (s, 3 H, CH_3^{NHAc}), 1.53 (spt, J = 6.6 Hz, 1 H, CH(CH_3)$_2$), 1.40 (q, J = 7.0 Hz, 2 H, -CH_2CH_2CH=C), 1.20 - 1.34 (m, 19 H, -CH_2^{acyl}), 1.16 (m, 2 H, -CH_2CH(CH_3)$_2$), 0.88, 0.87 (2 × s, 2× 3 H, -CH(CH_3)$_2$); **^{13}C NMR** (126 MHz, $CDCl_3$) δ ppm 170.0 (C=O^{NHAc}), 167.2 (C=O^{Acyl}), 162.3 (C-4 C=O), 150.5 (C-2 C=O), 145.6 (C=*C*H-CH_2), 138.4, 138.2, 138.0, 137.9, 137.7, 137.5, 137.4, 137.3, 136.8 (9 × C, 9 × C^{Ar}), 126.9 - 129.2 (46 × C, 45 × *C*H^{Ar}, C-6^{uracil}), 123.3 (C=*C*H-CO), 101.5 (C-5^{uracil}), 99.6 (C-11´), 96.7 (C-1´´), 88.6 (C-1´), 83.5 (C-2´), 80.9 (C-3´´), 78.4 (C-4´), 78.3 (C-4´´), 78.1 (C-8´), 75.3, 74.9 (2 × Ph*C*H_2-), 74.8 (C-7´), 74.6 (Ph*C*H_2-), 74.6 (C-9´), 74.0 (C-3´), 73.8, 73.5, 71.8 (3 × Ph*C*H_2-), 71.7 (2 × C, 2 × Ph*C*H_2-), 71.4 (C-5´´), 71.2 (C-5´), 69.4 (C-6´´), 53.2 (C-10´´), 52.1 (C-2´´), 44.0 (Ph*C*H_2N-), 39.0 (-*C*H_2CH(CH_3)$_2$), 33.1 (C-6´), 32.1 (-*C*H_2CH=C), 29.6 (5 × C, 5 × -*C*H_2^{acyl}), 28.3 (-*C*H_2CH_2CH=C), 27.9 (-*C*H(CH_3)$_2$), 27.4 (-*C*H_2^{acyl}), 23.4 (-*C*H_3^{NHAc}), 22.7 (2 × C, -CH(*C*H_3)$_2$); **IR** (thin film) ν : 3309 (br. w, N-H), 3088 (w), 3063 (w, C-H$^{C=C}$), 3031 (w, C-H$^{C=C}$), 2925 (s), 2854 (m), 1709 (m, unsat. C=O$^{2°amide\ I}$), 1669 (s, C=O$^{2°amide\ I}$), 1606 (s, C=C^{aromatic}), 1585 (w, C=C^{aromatic}), 1540 (m, C=O$^{2°amide\ II}$), 1496 (s, C=C^{aromatic}), 1389 (m), 1358 (m); **MS** *m/z* (ESI$^-$) : 1641 [(M-H)$^-$, 20%], 1701 [(M+AcOH-H)$^-$, 20%]; (ESI$^+$) : 1701 [(M+MeCN+NH_4)$^+$, 20%]. Flanking peaks with mass ± 14 corresponding to 8 × CH_2 and 10 × CH_2 in the aliphatic chain were seen in ESI$^+$, at 5% and 10% intensity respectively; **HRMS** *m/z* (ESI$^+$) : peaks observed (M + Na)$^+$ = 1663.8 (89%), 1664.8 (100%), 1665.8 (59%), 1666.8 (24%), 1667.8 (8%), 1668.9 (2%); peaks calc. for $C_{101}H_{116}N_4O_{16}Na$ (M + Na)$^+$ = 1663.8 (79%), 1664.8 (100%), 1665.8 (57%), 1666.8 (19%), 1667.8 (6%), 1668.9 (2%);

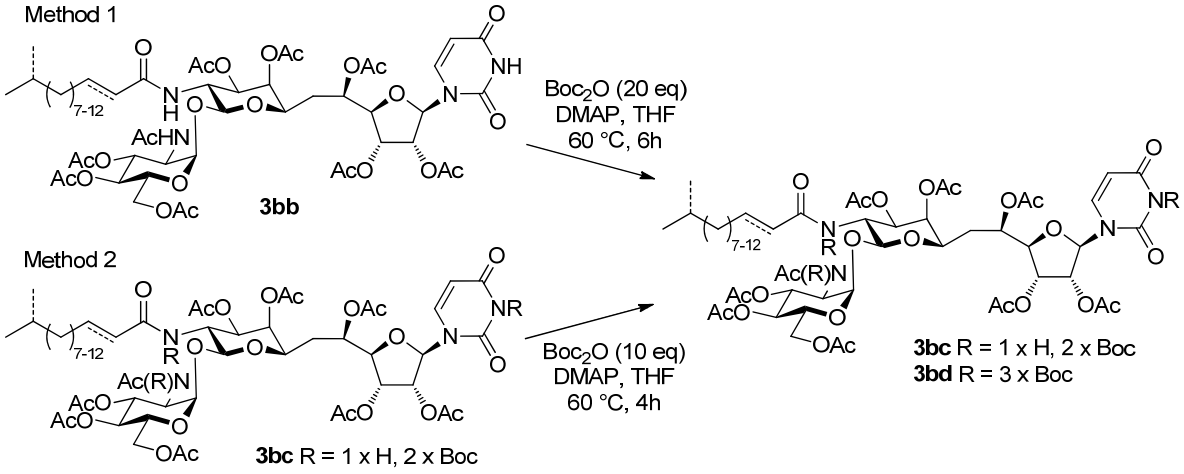
Octa-*O*-acetyl-tunicamycin 92



Tunicamycin # (88 mg, 0.103 mmol) was dissolved in dry pyridine (500 µL) and acetic anhydride (500 µL). After stirring for 18 h at room temperature, TLC (EtOAc) showed a major product at R_f 0.3 (**92**). After the solvent removed in vacuo, flash chromatography (MeOH/DCM 3:97), afforded octa-*O*-acetyl-tunicamycin **92** (68 mg, 0.058 mmol, 55%) as a clear glass containing a mixture of compounds with different aliphatic chain lengths; **^{1}H NMR** (700 MHz, MeOD-d_4) δ ppm

7.42 (d, $J_{6,5}$ = 8.2 Hz, 1 H, H-6^{uracil}), 6.78 (dt, ^{3}J = 7.0 Hz, $^{3}J_{HC=CH\ trans}$ = 15.3 Hz, 1 H, C=C*H*-CH$_2$), 5.81 (dt, $^{4}J_{allylic}$ = 1.6 Hz, $^{3}J_{HC=CH\ trans}$ = 15.3 Hz, 1 H, C=C*H*-CO), 5.76 (d, $J_{1´,2´}$ = 5.1 Hz, 1 H, H-1´), 5.69 (d, $J_{5,6}$ = 8.2 Hz, 1 H, H-5^{uracil}), 5.50 (dd, $J_{3´,4´}$ = 5.5 Hz, $J_{3´,2´}$ = 6.3 Hz, 1 H, H-3´), 5.45 (dd, $J_{2´,1´}$ = 5.1 Hz, $J_{2´,3´}$ = 6.3 Hz, 1 H, H-2´), 5.24 (dd, $J_{3´´,4´´}$ = 9.6 Hz, $J_{3´´,2´´}$ 10.9 Hz, 1 H, H-3´´), 5.19 (ddd, J = 1.0 Hz, $J_{5´,4´}$ = 3.7 Hz, $J_{5´,6b´}$ = 8.8 Hz, 1 H, H-5´), 5.19 (dd, $J_{8´,7´}$ = 1.0 Hz, $J_{8´,9´}$ = 3.2 Hz, 1 H, H-8´), 5.05 (app. t, $J_{4´´,3´´}$ = 9.6 Hz, $J_{4´´,5´´}$ = 10.0 Hz, 1 H, H-4´´), 5.02 (dd, $J_{9´,8´}$ = 3.2 Hz, $J_{9´,10´}$ = 11.3 Hz, 1 H, H-9´), 4.98 (d, $J_{1´´,2´´}$ = 3.6 Hz, 1 H, H-1´´), 4.70 (d, $J_{11´,10´}$ = 8.5 Hz, 1 H, H-11´), 4.30 (dd, $J_{6a´´,5´´}$ = 3.0 Hz, $J_{6a´´,6b´´}$ = 12.2 Hz, 1 H, H-6a´´), 4.28 (dd, $J_{10´,11´}$ = 8.5 Hz, $J_{10´,9´}$ = 11.3 Hz, 1 H, H-10´), 4.27 (ddd, $J_{5´´,6b´´}$ = 2.4 Hz, $J_{5´´,6b´´}$ = 3.0 Hz, $J_{5´´,4´´}$ = 10.0 Hz, 1 H, H-5´´), 4.15 (dd, $J_{4´,5´}$ = 3.7 Hz, $J_{4´,3´}$ = 5.5 Hz, 1 H, H-4´), 4.16 (dd, $J_{2´´,1´´}$ = 3.6 Hz, $J_{2´´,3´´}$ = 10.9 Hz, 1 H, H-2´´), 4.13 (dd, $J_{6b´´,5´´}$ = 2.4 Hz, $J_{6b´´,6a´´}$ = 12.2 Hz, 1 H, H-6b´´), 3.87 (ddd, $J_{7´,8´}$ = 1.0 Hz, $J_{7´,6b´}$ = 3.2 Hz, J = 9.3 Hz, 1 H, H-7´), 2.17 (s, 3 H, C*H*$_3^{Ac}$), 2.15 (m, 2 H, -C*H*$_2$CH=C), 2.09 (s, 6 H, 2 × C*H*$_3^{Ac}$), 2.04 (s, 3 H, C*H*$_3^{Ac}$), 2.02 (m, 2 H, H-6a´), 1.99, 1.97, 1.93, 1.89, 1.83 (5 × s, 5 × 3 H, 5 × C*H*$_3^{Ac}$), 1.73 (ddd, $J_{6b´,7´}$ = 3.2 Hz, $J_{6b´,5´}$ = 8.8 Hz, $J_{6b´,6a´}$ = 14.9 Hz, 1 H, H-6b´), 1.48 (spt, J = 6.7 Hz, 1 H, -C*H*(CH$_3$)$_2$), 1.41 (quin, J = 6.9 Hz, 2 H, -C*H*$_2$CH$_2$CH=C), 1.22 - 1.31 (m, 14 H, -C*H*$_2^{acyl}$), 1.13 (dt, J = 6.7, 7.8 Hz, 2 H, -C*H*$_2$CH(CH$_3$)$_2$), 0.85, 0.84 (2 × s, 2× 3 H, -CH(C*H*$_3$)$_2$); **^{13}C NMR** (126 MHz, MeOD-*d*$_4$) δ ppm 173.3, 172.5, 172.5, 172.4, 172.1, 171.8, 171.6, 171.4, 171.3 (9 × s, 9 × C, 8 × C=O^{Ac} , C=O^{NHAc}), 169.6 (C=O^{Acyl}), 166.0 (C-4 C=O), 151.9 (C-2 C=O), 147.8 (C=*C*H-CH$_2$), 143.5 (C-6^{uracil}), 124.3 (C=*C*H-CO), 103.7 (C-5^{uracil}), 101.7 (C-11´), 100.1 (C-1´´), 91.2 (C-1´), 84.2 (C-4´), 73.7 (C-2´), 72.4 (C-3´´), 72.3 (C-9´), 71.8 (C-7´), 71.0 (C-3´), 71.0, 70.4 (C-5´, C-8´), 69.9 (C-5´´), 69.9 (C-4´´), 63.1 (C-6´´), 52.7 (C-2´´), 51.9 (C-10´), 40.4 (-*C*H$_2$CH(CH$_3$)$_2$), 33.4 (-*C*H$_2$CH=C), 33.2 (C-6´), 30.4 – 31.2 (5 × C, 5 × -*C*H$_2^{acyl}$), 29.5 (-*C*H$_2$CH$_2$CH=C), 29.3 (-*C*H(CH$_3$)$_2$), 28.7 (-*C*H$_2^{acyl}$), 23.2 (2 × C, -CH(*C*H$_3$)$_2$), 23.1 (-*C*H$_3^{NHAc}$), 21.2 (-*C*H$_3^{Ac}$), 20.9 (s, 2 × C, 2 × -*C*H$_3^{Ac}$), 20.8, 20.7, 20.7, 20.7, 20.5 (5× s, 5 × C, 5 × -*C*H$_3^{Ac}$); **IR** (thin film) *v* : 3500-3700 (multiple w, N-H), 2926 (w), 2880 (w), 1749 (br s., C=O^{Ac}), 1717 (s, unsat. C=O$^{2°amide\ I}$), 1699 (s, C=O$^{2°amide\ I}$), 1684 (s, C=O$^{2°amide\ I}$), 1653 (s, C=C), 1635 (s, C=C), 1558 (m, unsat. C=O$^{2°amide\ II}$), 1540 (m, C=O$^{2°amide\ II}$), 1521 (m, C=O$^{2°amide\ II}$), 1374 (m), 1233 (br. s); **MS** *m/z* (ESI$^-$) : 1165 [(M-H)$^-$, 100%]; (ESI$^+$) : 1225 [(M+MeCN+NH$_4$)$^+$, 100%]. Flanking peaks with mass ± 14 corresponding to 8 × CH$_2$ and 10 × CH$_2$ in the aliphatic chain were seen, at 30% and 40% intensity respectively; **HRMS** *m/z* (ESI$^+$) : peaks observed (M + Na)$^+$ = 1189.4 (57%), 1190.5 (18%), 1191.5 (21%), 1192.5 (4%); peaks calc. for $C_{54}H_{78}N_4O_{24}Na$ (M + Na)$^+$ = 1189.5 (52%), 1190.5 (32%), 1191.5 (12%), 1192.5 (3%), 1193.5 (1%);

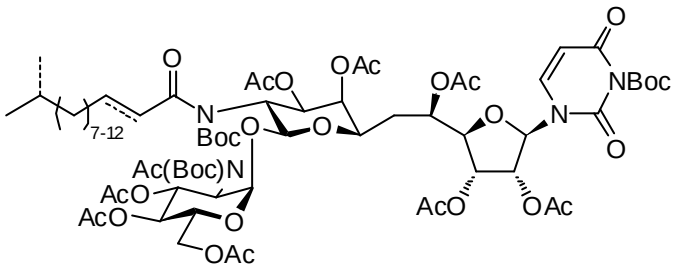
Boc protection of octa-*O*-acetyl-tunicamycin



Method 1: Octa-*O*-acetyl-tunicamycin **92** (75 mg, 0.064 mmol) was dissolved in dry THF (1 mL) and di-*tert*-butyl dicarbonate (140 mg, 0.643 mmol) and 4-(dimethylamino)pyridine (1.6 mg, 0.013 mmol) were added. The flask was fitted with an oven dried condenser and the solution heated to 60 °C. After 2 h, further di-*tert*-butyl dicarbonate (140 mg, 0.643 mmol) was added. After 4 h, TLC (EtOAc/Petrol 6:4) showed spots at R_f 0.3 (**94**) and 0.2 (**93**). The reaction was cooled and the solvent removed *in vacuo*. Flash chromatography (EtOAc/Petrol 6:4→7:3), afforded tri-*N*-(*tert*-butoxycarbonyl)-octa-*O*-acetyl-tunicamycin **94** (42 mg, 0.029 mmol, 45%) as a yellow foam and di-*N*-(*tert*-butoxycarbonyl)-octa-*O*-acetyl-tunicamycin **93** (27 mg, 0.020 mmol, 31%) as a pale yellow foam;

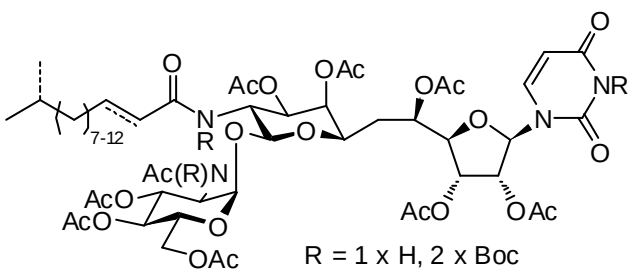
Method 2: Di-*N*-(*tert*-butoxycarbonyl)-octa-*O*-acetyl-tunicamycin **93** (27 mg, 0.020 mmol) was dissolved in dry THF (1 mL) and di-*tert*-butyl dicarbonate (22 mg, 0.099 mmol) and 4-(dimethylamino)pyridine (0.5 mg, 4 µmol) were added. The flask was fitted with an oven dried condenser and the solution heated to 60 °C. After 2 h, further di-*tert*-butyl dicarbonate (22 mg, 0.099 mmol) was added. After 2 h, TLC (EtOAc/Petrol 6:4) showed spots at R_f 0.3 (**94**) and 0.2 (**93**). The reaction was cooled and the solvent removed *in vacuo*. Flash chromatography (EtOAc/Petrol 6:4→7:3), afforded tri-*N*-(*tert*-butoxycarbonyl)-octa-*O*-acetyl-tunicamycin **94** (8 mg, 5.5 µmol, 28%) as a yellow foam and di-*N*-(*tert*-butoxycarbonyl)-octa-*O*-acetyl-tunicamycin **93** (15 mg, 11.0 µmol, 55%) as a pale yellow foam;

Tri-*N*-(tert-butoxycarbonyl)-octa-*O*-acetyl-tunicamycin 94



^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 7.24 (d, $J_{6,5}$ = 8.2 Hz, 1 H, H-6^{uracil}), 6.89 (dt, *J* = 7.0 Hz, $^3J_{HC=CH\ trans}$ = 15.2 Hz, 0.7 H, C=C*H*-CH_2), 6.82 (dt, *J* = 7.0 Hz, $^3J_{HC=CH\ trans}$ = 15.4 Hz, 0.3 H, C=C*H*-CH_2), 6.38 (d, $^3J_{HC=CH\ trans}$ = 15.4 Hz, 0.3 H, C=C*H*-CO), 6.27 (d, $^3J_{HC=CH\ trans}$ = 15.2 Hz, 0.7 H, C=C*H*-CO), 6.15 (m, 1 H, C-H^{anomeric}), 5.89 (m, 1.7 H, C-H^{anomeric}, H-5^{uracil}), 5.83 (d, *J* = 8.2 Hz, 0.3 H, H-5^{uracil}), 5.60 (dd, *J* = 11.4 Hz, *J* = 3.3 Hz, 0.3 H), 5.53 (dd, *J* = 11.4 Hz, *J* = 3.5 Hz, 0.7 H), 5.48 (d, *J* = 8.1 Hz, 0.3 H), 5.36 - 5.44 (m, 1.7 H), 5.30 - 5.35 (m, 1.3 H), 5.13 - 5.25 (m, 2.7 H), 4.95 - 5.05 (m, 2 H), 4.92 (dd, *J* = 11.7 Hz, *J* = 8.3 Hz, 0.7 H), 4.76 (m, 0.3 H), 4.56 (dd, *J* = 11.5 Hz, *J* = 8.2 Hz, 0.3 H), 4.27 - 4.42 (m, 2.3 H), 4.16 (dd, *J* = 13.4 Hz, *J* = 2.5 Hz, 0.7 H), 4.11 (d, *J* = 7.1 Hz, 0.3 H), 4.04 - 4.10 (m, 1.7 H), 3.76 (dd, *J* = 8.8 Hz, *J* = 3.8 Hz, 0.3 H), 3.68 (dd, *J* = 10.0 Hz, *J* = 2.2 Hz, 0.7 H), 2.28 (s, 2 H, C*H*$_3^{NHAc}$), 2.27 (s, 1 H, C*H*$_3^{NHAc}$), 1.85 – 2.20 (m, 29 H, 8 × C*H*$_3^{Ac}$, 2 × H-6´, C*H*^{acyl}, C*H*$_2$CH=C), 1.59, 1.56, 1.55, 1.52, 1.51 (5 × s, 27 H, 9 × C*H*$_3^{Boc}$), 1.43 (m, 4 H, 2 × C*H*$_2^{acyl}$), 1.19 - 1.33 (m, 20 H, C*H*$_2^{acyl}$), 1.09 - 1.18 (m, 2 H, C*H*$_2^{acyl}$), 0.86, 0.84 (2 × s, 2 × 3 H, C*H*$_3^{acyl}$). An EXSY spectrum showed cross peaks between resonances at δ 6.38 and 6.27 ppm and between resonances at δ 6.82 and 6.89 ppm, confirming that the two species observed by NMR are rotamers of one another arising from restricted rotation at the N-10´ amide bond; **IR** (thin film) *v* : 2923 (m), 2852 (m), 1787 (w), 1747 (br s., C=O^{Ac}, C=O^{Boc}), 1685 (s, C=O$^{2°amide\ I}$), 1635 (w, C=C), 1447 (m), 1370 (s); **MS** *m/z* (ESI$^+$) : 1540 [(M+MeCN+NH_4)$^+$, 100%]. Flanking peaks with mass ± 14 corresponding to 8 × CH_2 and 10 × CH_2 in the aliphatic chain were seen in ESI$^+$, at 80% and 20% intensity respectively;

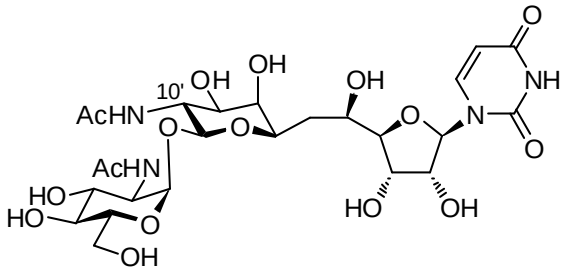
Di-*N*-(*tert*-butoxycarbonyl)-octa-*O*-acetyl-tunicamycin 93



^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 8.75 (s, 0.1 H, N-H^{uracil}), 8.21 (d, *J* = 8.5 Hz, 0.5 H, N-H), 7.45 (d, *J* = 8.5 Hz, 0.5 H, N-H), 7.26 (d, *J* = 8.3 Hz, 0.5 H, H-6^{uracil}), 7.18 (d, *J* = 8.3 Hz, 0.5 H, H-6^{uracil}), 7.00 (dt, *J* = 7.2 Hz, $^3J_{HC=CH\ trans}$ = 14.9 Hz, 0.5 H, C=C*H*-CH_2), 6.94 (dt, *J* = 7.1 Hz, $^3J_{HC=CH\ trans}$ = 15.1 Hz, 0.5 H, C=C*H*-CH_2), 6.29 (d, $^3J_{HC=CH\ trans}$ = 15.4 Hz, 0.5 H, C=C*H*-CO), 6.25 (d, $^3J_{HC=CH\ trans}$ = 15.1 Hz, 0.5

H, C=C*H*-CO), 5.86 (m, 1 H, C-H^{anomeric}), 5.82 (m, 1 H, H-5^{uracil}), 5.77 (d, *J* = 9.9 Hz, 7.0 Hz, 1 H), 5.71 (dd, *J* = 3.7 Hz, 11.5 Hz, 0.5 H), 5.66 (dd, *J* = 3.5 Hz, 11.6 Hz, 0.5 H), 5.49 (d, *J* = 8.3 Hz, 0.5 H), 5.10 - 5.45 (m), 4.98 (d, *J* = 3.3 Hz, 0.5 H), 4.88 (d, *J* = 3.5 Hz, 0.5 H), 4.70 (dd, *J* = 11.5 Hz, 8.2 Hz, 0.5 H), 4.12 (m), 3.77 (dd, *J* = 2.8 Hz, 9.9 Hz, 0.5 H), 3.71 (dd, *J* = 2.3 Hz, 9.9 Hz, 0.5 H), 2.23, 2.22 (2 × s, 2 × 1.5 H, 2 × C*H*$_3^{NHAc}$), 1.81 - 2.14 (m, 8 × C*H*$_3^{Ac}$, 2 × H-6´, C*H*^{acyl}, C*H*$_2$CH=C), 1.63 (s, 3 H, C*H*$_3^{Boc}$), 1.59 (br. s, 9 H, 3 × C*H*$_3^{Boc}$), 1.50 (s, 6 H, 2 × C*H*$_3^{Boc}$), 1.42 (m, 4 H, 2 × C*H*$_2^{acyl}$), 1.20 - 1.33 (m, C*H*$_2^{acyl}$), 1.16 (m, 2 H, C*H*$_2^{acyl}$), 0.87, 0.85 (2 × s, 2 × 3 H, C*H*$_3^{acyl}$). The compound was isolated as a complex mixture of 3 isomers, each with 2 × NBoc and 1 × NH, so more detailed NMR study was not performed; **IR** (thin film) *v* : 3400 (br. w, N-H), 2926 (s), 2854 (m), 1787 (w), 1749 (br s., C=O^{Ac}, C=O^{Boc}), 1685 (s, C=O$^{2°amide\ I}$), 1634 (w, C=C), 1516 (w, C=O$^{2°amide\ II}$), 1447 (m), 1370 (s); **MS** *m/z* (ESI$^+$) : 1440 [(M+MeCN+NH$_4$)$^+$, 100%]. Flanking peaks with mass ± 14 corresponding to 8 × CH_2 and 10 × CH_2 in the aliphatic chain were seen in ESI$^+$, at 60% and 15% intensity respectively;

$N^{10'}$-Acetyl-$N^{10'}$-deacyl-tunicamycin 99



Tri-*N*-(*tert*-butoxycarbonyl)-octa-*O*-acetyl-tunicamycin **94** (127 mg, 0.087 mmol) was dissolved in dry MeOH (5 mL) and cooled to 0 °C. 1M NaOMe in MeOH (50 μL) was added, resulting in a 0.01M concentration of base. After 6 h, the reaction was neutralised carefully with DOWEX 50WX8 H^+ resin and filtered through cotton. The solvent was removed *in vacuo* and the residue redissolved in trifluoroacetic acid (2 mL). After stirring at room temperature for 1 h, the solvent was removed *in vacuo* and the residue coevaporated with toluene. The product was dissolved in MeOH (5 mL) and acetic anhydride (1 mL). After stirring for two hours at room temperature, the pH was adjusted to 6-7 with DOWEX 550 OH^- resin. After a further hour, the solution was filtered and concentrated *in vacuo*. Flash chromatography (W.IP.E. 1:2:2) afforded $N^{10'}$-acetyl-$N^{10'}$-deacyl-tunicamycin **99** (26 mg, 0.040 mmol, 46%) as a white powder; $[\alpha]_D^{18}$ = -24, c = 1, H_2O; **^{1}H NMR** (500MHz, D_2O): δ ppm 7.81 (d, $J_{6,5}$ = 8.2 Hz, 1 H, H-6^{uracil}), 5.88 (d, $J_{1',2'}$ = 5.4 Hz, 1 H, H-1′), 5.89 (d, $J_{5,6}$ = 8.2 Hz, 1 H, H-5^{uracil}), 5.03 (d, $J_{1'',2''}$ = 3.5 Hz, 1 H, H-1′′), 4.63 (d, $J_{11',10'}$ = 8.5 Hz, 1 H, H-11′), 4.30 (dd, $J_{2',1'}$ = 5.4 Hz, $J_{2',3'}$ = 5.9 Hz, 1 H, H-2′), 4.26 (dd, $J_{3',4'}$ = 3.5 Hz, $J_{3',2'}$ = 5.9 Hz, 1 H, H-3′), 3.94 - 4.00 (m, 2 H, H-4′, H-5′), 3.91 (dd, $J_{10',9'}$ = 9.5 Hz, $J_{10',11'}$ = 8.5 Hz, 1 H, H-10′), 3.88 (m, 1 H, H-4′), 3.85 (dd, $J_{2'',1''}$ = 3.5 Hz, $J_{2'',3''}$ = 10.7 Hz, 1 H, H-2′′), 3.81 (d, *J* = 10.5 Hz, 1 H, H-7′), 3.74 - 3.78 (m, 2 × 1 H, H-8′, H-6b′′), 3.74 (dd, $J_{3'',2''}$ = 10.7 Hz, $J_{3'',4''}$ = 8.8 Hz, 1 H, H-3′′), 3.75 (app t, *J* = 10.0 Hz, $J_{6a'',5''}$ = 9.8 Hz, 1 H, H-6′′a), 3.74 (d, $J_{9',10'}$ = 9.5 Hz, 1 H, H-9′), 3.48 (app. t, $J_{5'',6a''}$ = 9.8 Hz, *J* = 9.8 Hz, 1 H, H-5′′), 2.02, 1.99 (2 × s, 2 × 3 H, 2 × -CH_3^{NHAc}), 1.98 (dd, $J_{6b'',6a''}$ = 14.4 Hz, $J_{6b'',5''}$ = 10.5 Hz, 1 H, H-6b′), 1.62 ppm (dd, $J_{6a'',6b''}$ = 14.4 Hz, $J_{6a'',5''}$ = 11.0 Hz, 1 H, H-6a′), please see appendix for 1D TOCSY spectra of irradiated anomeric protons, H-5′′ and H-6′b which were used to assign overlapping peaks; **^{13}C NMR** (126 MHz, D_2O) δ ppm 175.6, 175.2 (2 × C=O^{NHAc}), 167.3 (C-4 C=O), 152.9 (C-2 C=O), 143.0 (C-6^{uracil}), 103.6 (C-5^{uracil}), 100.9 (C-11′), 99.4 (C-1′′), 89.6 (C-1′), 88.3 (C-4′), 74.6 (C-2′), 73.7 (C-4′′), 72.4 (C-7′), 72.2, 71.6, 71.0 (C-3′′, C-8′, C-9′), 70.7 (C-5′′), 70.1 (C-3′), 68.2 (C-5′), 61.6 (C-6′′), 54.5 (C-2′′), 53.9 (C-10′), 34.6 (C-6′), 23.4, 23.2 (2 × -CH_3^{NHAc}); **IR** (thin film) *v* : 3409 (b, O-H), 2923 (w, C-H), 1667 (s, C=O^{amide}), 1559 (m, N-H^{amide}), 1466 (m), 1425 (m), 1381 (m); **MS** *m/z*; (ESI^+) : 673 [$(M+Na)^+$, 100%]; **HRMS** *m/z* (ESI^+) : calc. for $C_{25}H_{38}N_4NaO_{16}$ $(M + H)^+$ = 673.2175. Found 673.2176;

3.6 References

369. Price NP, Momany FA (2005): **Modeling bacterial UDP-HexNAc: polyprenol-P HexNAc-1-P transferases**. *Glycobiology* **15**: 29R-42R.

370. Takatsuki A, Arima K, Tamura G (1971): **Tunicamycin, a New Antibiotic: 1. Isolation and Characterization of Tunicamycin**. *J. Antibiot.* **24**: 215-223.

371. Hamill RG, Eli Lilly Co. (1980): **Process for preparing tunicamycin**. U. S. Patent 4237225

372. Hamill RG, Eli Lilly Co. (1982): **Process for preparing tunicamycin**. U. S. Patent 4336333

373. Tsvetanova BC, Kiemle DJ, Price NPJ (2002): **Biosynthesis of tunicamycin and metabolic origin of the 11-carbon dialdose sugar, tunicamine**. *J. Biol. Chem.* **277**: 35289-35296.

374. Tsvetanova BC, Price NPJ (2001): **Liquid Chromatography-Electrospray Mass Spectrometry of Tunicamycin-Type Antibiotics**. *Anal. Biochem.* **289**: 147-156.

375. *Product Information: Tunicamycin from Streptomyces sp.* (2004). Sigma-Aldrich: St. Louis.

376. Mahoney WC, Duksin D (1980): **Separation of Tunicamycin Homologs by Reversed-Phase High-Performance Liquid-Chromatography**. *J. Chromatogr.* **198**: 506-510.

377. Myers AG, Gin DY, Rogers DH (1994): **Synthetic Studies of the Tunicamycin Antibiotics. Preparation of (+)-Tunicaminyluracil, (+)-Tunicamycin-V, and 5'-*epi*-Tunicamycin-V**. *J. Am. Chem. Soc.* **116**: 4697-4718.

378. Tamura G (1982): *Tunicamycin.* Japan Scientific Societies Press: Tokyo.

379. Ishizuka T, Kunieda T (1987): **Mild and selective ring-cleavage of cyclic carbamates to amino alcohols**. *Tetrahedron Lett.* **28**: 4185-4188.

380. Benakli K, Zha C, Kerns RJ (2001): **Oxazolidinone Protected 2-Amino-2-deoxy-D-glucose Derivatives as Versatile Intermediates in Stereoselective Oligosaccharide Synthesis and the Formation of α-Linked Glycosides**. *J. Am. Chem. Soc.* **123**: 9461-9462.

381. Lee BC, Sung HJ, Kim JS, Jung KH, Choe YS, Lee KH, Chi DY (2007): **Synthesis of Tc-99m labeled glucosamino-Asp-cyclic(Arg-Gly-Asp-D-Phe-Lys) as a potential angiogenesis imaging agent**. *Bioorganic & Medicinal Chemistry* **15**: 7755-7764.

382. Greene TW, Wuts PGM (1999): *Protective Groups in Organic Synthesis* 3rd edn. John Wiley & Sons, Inc.: Chichester.

383. Kocienski PJ (2005): *Protecting Groups* 3rd (revised) edn. Thieme Publishing Group: Stuttgart.

Chapter 4: Functional characterisation of tun gene products

4.1 Introduction

In chapter 2 the genes responsible for tunicamycin biosynthesis in *S. chartreusis* were successfully identified. Cloning and heterologous expression of this gene cluster in *S. coelicolor* verified their biological function and delineated the minimal gene set necessary for tunicamycin production. In depth bioinformatic analysis of *tun* gene sequences allowed putative assignment of individual enzyme function, culminating in the proposal of a detailed metabolic pathway of tunicamycin biosynthesis from common metabolic precursors. In order to lend weight to these proposals, in this chapter the functional characterisation of individual *tun* genes is presented. Genes were cloned and overexpressed in *E. coli* hosts and following purification, these gene products were assayed for functional activity. Putative biosynthetic intermediates prepared from native tunicamycins in chapter 3 were also utilised in these studies.

4.2 TunF: A UDP-GlcNAc-4-epimerase

4.2.1 Protein sequence analysis of TunF

Bioinformatic analysis of TunF and its putative role in tunicamycin biosynthesis was presented in section 2.4.2. The amino acid sequence showed homology to UDP-glucose-4-epimerases from a wide range of sources. Reference against the Conserved Domains Database (CDD) highlighted the presence of an N-terminal Rossmann fold;[384] a TGxxGxxG signature motif characteristic of nucleotide binding proteins was found (Table 4.1).[385] TunF also contained an $Sx_{24}Yx_3K$ catalytic triad, where proton transfer from the tyrosine to the substrate carbonyl is supported by a proton relay involving the lysine and serine residues.[385] In the tunicamycin biosynthetic pathway TunF was proposed to install the pseudo-GalNAc functionality within tunicamine from a GlcNAc-configured precursor and was therefore assigned the putative function of a UDP-GlcNAc-4-epimerase.

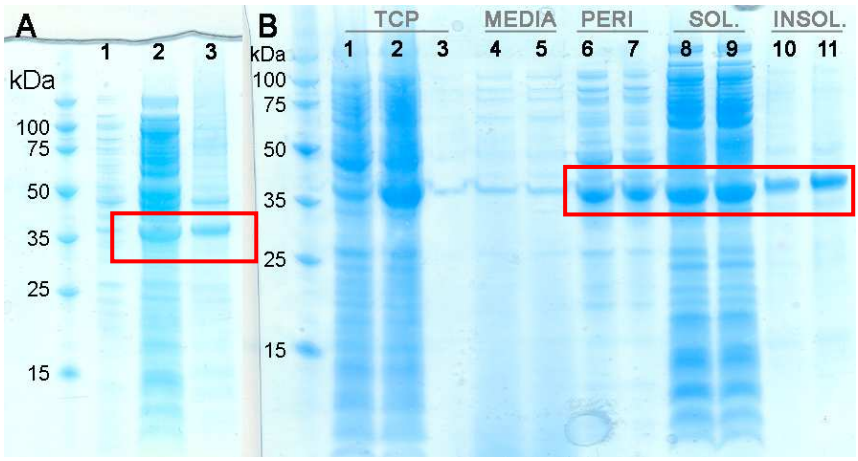
Protein	Amino acids	Amino acid sequence[a]	Mw[b] / pI
Native TunF	327	MRVLVTGGAGYVGSFTVRRLLAAGHEVVVVDNLSTGRREAV RGCRLHVVDILDIASMGTVFEEFHPEAVIHFAALKSSEESLRDI NTYFSVNLTGTQNVLALCARTGVERFVFSSSCAVYGTPQICP VDETAPVRPESPYGETKYLCERVIASYAQATGMRYANLRYFN AAGAADDASLGEYAGPASTQLLPVAIRSTLGLAPTLRIFGNDY PTTDGTALRDYIHVEDLAQAHVRVLEGLQKGSQCGPVNLGR GQPVSVRELIDVLHDVSGKETPVAMAPRRPGDPGLSWADPS LALSRFGWRAERDLDHIVRTAWNWHTSTPASLE	Mw = 35428 Da pI = 5.87

[a] Nucleotide binding motif highlighted in magenta, catalytic triad highlighted in yellow; [b] Average molecular weight;

Table 4.1 Amino acid sequence and selected properties of TunF

4.2.2 Expression and purification of TunF

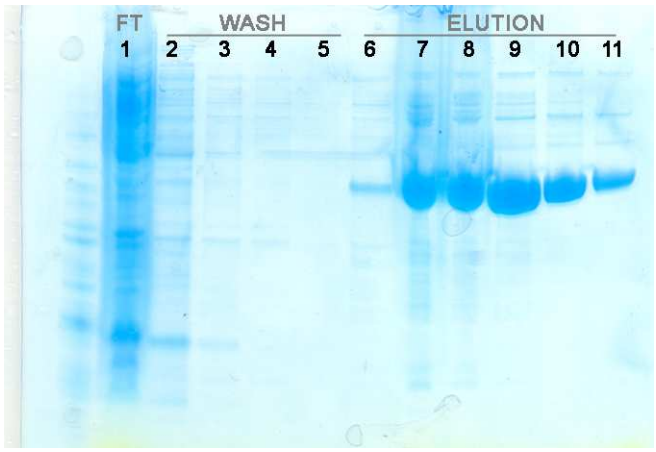
The DNA sequence of *tunF* was codon optimised for *E. coli* and synthesised by Genscript. The gene was cloned into expression vector pET-16b (Novagen), which encodes an N-terminal His_{10} tag and Factor Xa cleavage site. The resulting gene product His_{10}-TunF is predicted to have 347 residues, an average Mw of 37,949 Da and a pI of 6.39. The plasmid, which we named pET-16b:*tunF*, was transformed into *E. coli* BL21(DE3) according to the supplier's protocol (Novagen) and subjected to protein expression trials (Fig. 4.1).



A: Induction for 3 h at 37 ℃. Lane 1: Total cell pro tein (TCP); Lane 2: Soluble cytosolic fraction; Lane 3: Insoluble cytosolic fraction;
B: Induction for 3-6 h at 30 ℃. Lane 1-3: TCP fract ion after 0, 3 and 6 h; Lane 4-5: Media fraction after 3 and 6 h; Lane 6-7: Periplasmic fraction after 3 and 6 h; Lane 8-9: Soluble cytosolic fraction after 3 and 6 h; Lane 10-11: Insoluble cytosolic fraction after 3 and 6 h;

Fig. 4.1 Expression trials of TunF from *E. coli* BL21(DE3)/pET16b:*tunF*

Expression of *tunF* in recombinant *E. coli* cultures was induced with 1mM IPTG at 37 °C for 3 h. Analysis of the resulting cell pellet showed bands corresponding to TunF (~38 kDa) in the total cell protein, soluble cytosolic and insoluble cytosolic protein fractions (Fig. 4.1, A). Levels of gene expression appeared high – as judged by SDS-PAGE analysis – although a good proportion of the over-expressed protein was being stored as insoluble inclusion bodies. In order to increase the proportion of soluble and hence active TunF, the temperature following induction was reduced to 30 °C. Once again good levels of *tunF* expression were obtained and crucially an increase in the proportion of soluble protein was observed (Fig. 4.1, B).



Lane 1: Column flow-through; Lanes 2-5: Wash fractions containing 20 mM imidazole; Lanes 6-11: Elution fractions containing 250 mM imidazole; NB. Protein ladder marker to the left of lane 1 showed signs of degradation.

Fig. 4.2 Purification of His_{10}-TunF by Ni^{2+} affinity chromatography

In order to obtain preparative quantities of TunF for protein purification and subsequent enzymatic assays, the optimised expression conditions were scaled up. A cell pellet was collected from 1 L of induced culture and lysed with sequential lysozyme treatment and sonication. The N-terminally His_{10}-tagged construct was purified by immobilised metal affinity chromatography (IMAC).[386,387] The cleared crude extract following cell lysis was batch bound with Ni-NTA Resin (Novagen). Following a wash step to remove unbound proteins and other soluble contents of the cell, bound proteins were eluted with a linear gradient of imidazole. This single chromatographic step resulted in a protein with >90% homogeneity, as shown in Fig. 4.2. Fractions containing purified His_{10}-TunF were pooled, dialysed and concentrated prior to functional characterisation. A typical yield of 25 mg TunF was obtained per L of culture.

4.2.3 Functional NMR assay of TunF

UDP-glucose/GlcNAc epimerases are typically assayed with a coupled assay using an additional carbohydrate acceptor and a glycosyltransferase specific for the epimerised product – such as β-1,4-*N*-acetylgalactosaminyltransferase.[388] Capillary electrophoresis has also been used to resolve reaction products and in the case of UDP-GlcNAc a spectrophotometric assay has been developed, relying on 4,6-*O*-benzylidenes of *gluco*- and *galacto*- configured products having different absorbance properties at specific wavelengths.[389,390] We envisaged the development of a greatly simplified and rapid enzymatic assay that would also allow unambiguous confirmation of reaction products. With access to a 700 MHz NMR spectrometer, product formation could be detected with high sensitivity. Additionally, the high field of the instrument would resolve chemical shifts originating from the nucleotide-sugar substrate and any TunF-catalysed product, allowing accurate characterisation of newly formed compounds without the need for laborious chromatographic separations. Since the predicted substrate for TunF was commercially available UDP-GlcNAc, enzyme assays could readily be performed on a scale compatible with sensitive ^{1}H NMR detection.

A sample of TunF was thus buffer exchanged into 30 mM sodium phosphate in D_2O pD 7.8. Using this stock solution, 25 µg of enzyme was added to a solution containing sodium phosphate in D_2O (30 mM, pD 7.8), UDP-GlcNAc (25 mM) and NAD^+ (0.1 mM). An NMR timecourse experiment was performed at 37 °C with a Bruker AV700 MHz machine, collecting a ^{1}H spectrum every four minutes. The results clearly showed formation of a new product with chemical shifts indistinguishable from those of a commercial sample of UDP-GalNAc (Fig. 4.3). By comparison to an internal standard, the decrease in integration of UDP-GlcNAc-derived peaks was shown to precisely mirror the increase in integration of novel shifts assigned to UDP-GalNAc. A control experiment lacking TunF was conducted in parallel; no product formation was observed. As such the functional role of TunF has been experimentally confirmed and shown to be that of a UDP-GlcNAc-4-epimerase, as previously predicted by bioinformatic analysis of the amino acid sequence.

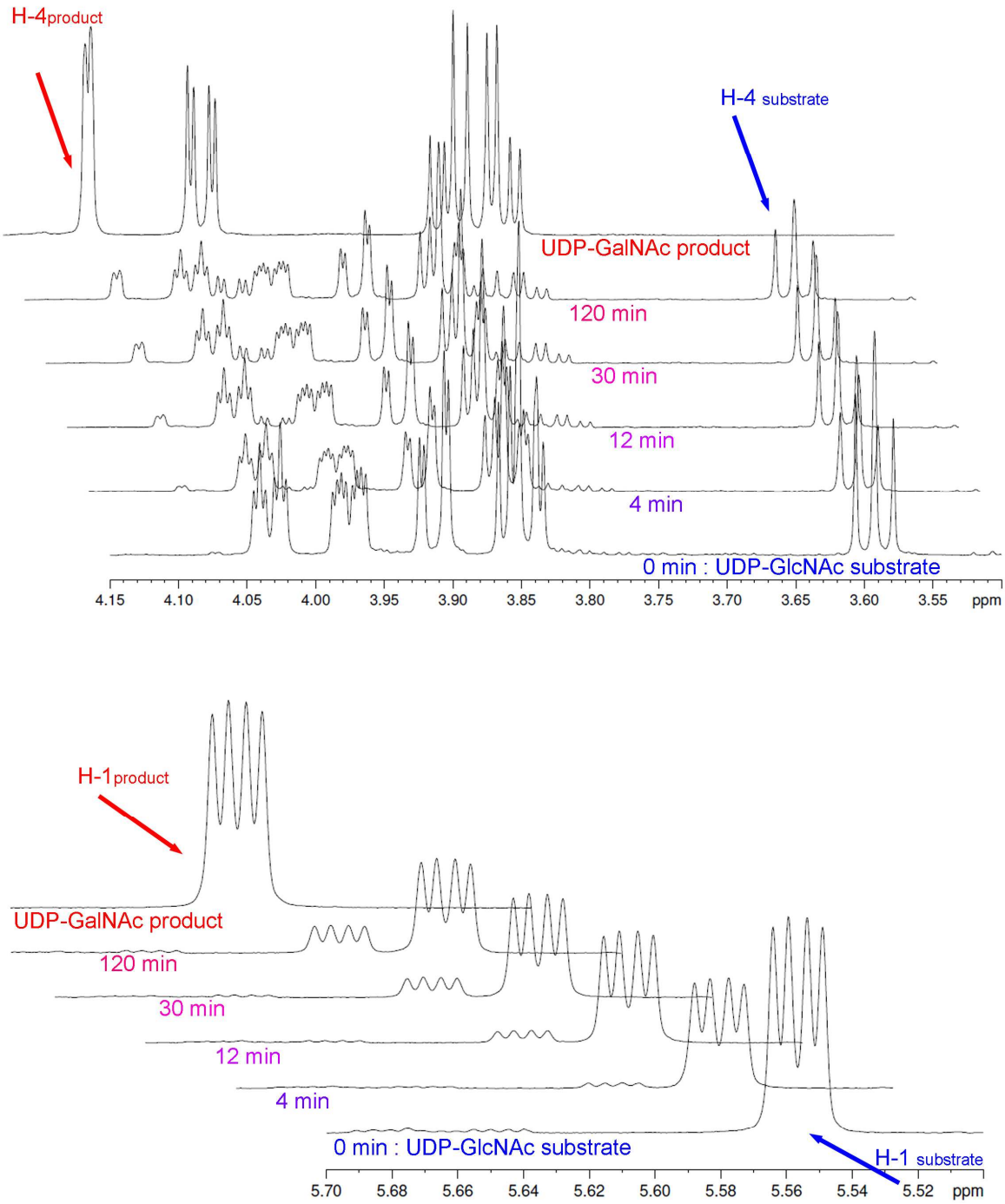


Fig. 4.3 Timecourse NMR assay of TunF activity

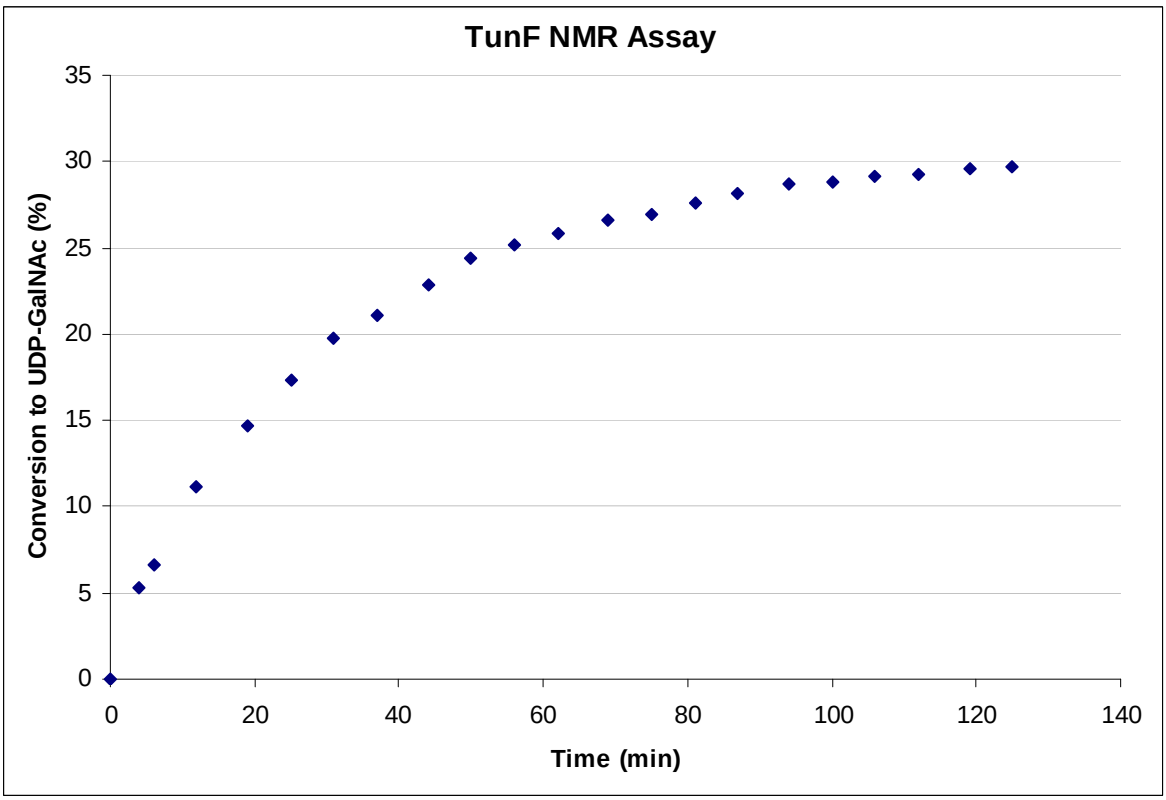


Fig. 4.4 Timecourse of TunF activity and establishment of equilibrium conversion

The ratio of anomeric peaks in spectra taken at different time points on the reaction coordinate were compared, allowing a timecourse of the TunF-catalysed reaction to be plotted (Fig. 4.4). The resulting graph clearly showed that an equilibrium was reached at 30% conversion to UDP-GalNAc. This has been previously observed and is not unexpected given the UDP-GalNAc is also a substrate for the enzyme.[390] Its axial H-4 makes it less thermodynamically stable than UDP-GlcNAc, suggesting an explanation why the equilibrium is shifted in favour of UDP-GlcNAc. It is likely that *in vivo* products from such enzyme reactions are coupled closely with subsequent enzymes in the pathway, which specifically utilise these products as substrates, thereby drawing the equilibrium of the epimerase reaction over to product. The moles of product formed at a given time point was determined by comparing the integrals of H-1 peaks from both the substrate and the product. These values were used to estimate the initial rate of reaction (V_0) for this experiment in mol of product formed per second. Dividing this by the initial enzyme concentration ($[E]_0$) produced a pseudo k_{cat} constant with a value of approximately 4.2 s^{-1}, indicating that TunF was a catalytically competent enzyme. Further experiments are required to determine true k_{cat} and K_M parameters.

4.2.4 Oligomerisation state of TunF

His_{10}-TunF was subjected to size exclusion chromatography[391] and compared to protein standards of defined size to determine its oligomerisation status (Fig. 4.5 and Table 4.2). Its retention time of 72 min corresponds to an estimated Mw of 103,000 kDa. Since the monomer of His_{10}-TunF is 37.8 kDa, the retention time observed suggests TunF eluted as a trimer. This is in contrast with results obtained with other UDP-hexose-4-epimerases, where dimers have typically been observed.[5] The shape of the peak corresponding to the suspected TunF trimer is symmetrical, indicative of a stable protein complex with minimal aggregation caused by misfolded protein. The additional peak seen at 115 min could possibly be attributed to the NAD^+ cofactor necessary for the activity of TunF. Indeed, samples purified by gel filtration chromatography were found to be inactive.

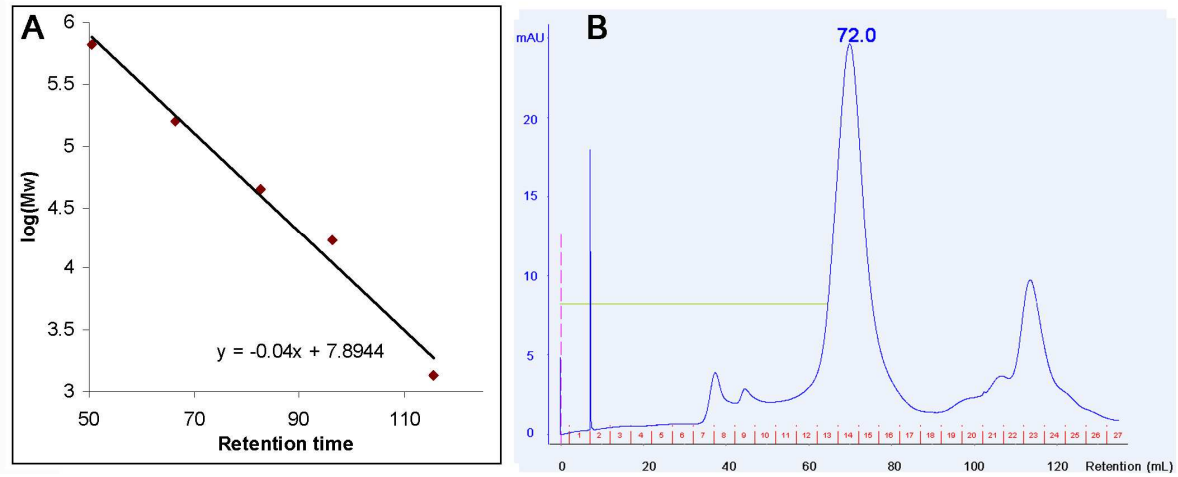


Fig. 4.5 Size exclusion chromatography of TunF;
A: Calibration plot of gel filtration standards (Bio-Rad); **B**: Elution profile of His_{10}-TunF

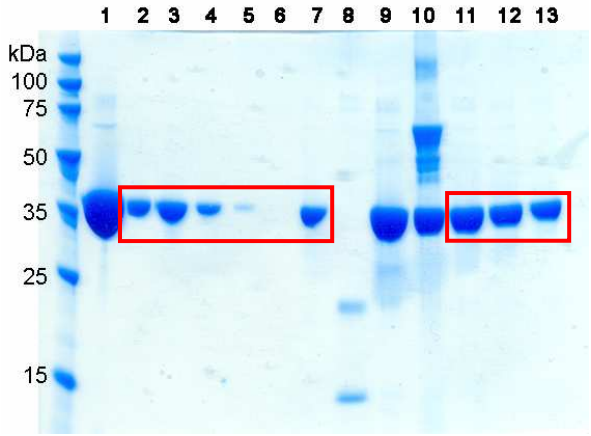
Component	Molecular weight (Da)	Retention (mL)
Thyroglobulin (bovine)	670,000	50.5
γ-globulin (bovine)	158,000	66.5
Ovalbumin (chicken)	44,000	82.5
Myoglobin (horse)	17,000	96.3
Vitamin B12	1,350	115.6

Table 4.2 Gel filtration standards (Bio-Rad) used calibrate size exclusion column

4.2.5 TunF crystallisation screen

A 3D crystal structure of TunF would allow the active site and substrate-binding pockets of the enzyme to be accurately visualised. Comparison with existing crystal structures of related enzymes would allow conclusions to be drawn regarding the natural substrate for TunF and the extent of substrate specificity. It would also provide valuable further information about this important class of enzymes.

A high-throughput strategy was envisaged for the crystallisation of TunF, screening a large number of crystallisation conditions in parallel using a Rigaku Crystallisation Robot. In order to maximise chances of success, a highly pure, concentrated sample of TunF was sought. Previous expression protocols and subsequent IMAC purification was first performed, followed by a further round of purification by high resolution anion exchange chromatography using a MonoQ column (Fig. 4.6). The result was a highly homogenous sample of TunF, which was desalted and concentrated to 57.8 mg/mL.



Lane 1: Sample prior to anion-exchange purification; Lanes 2-7: Wash fractions eluting unbound protein; Lanes 8-13: Fractions from elution across an ionic gradient of 0→1 M NaCl

Fig. 4.6 MonoQ anion-exchange purification of His_{10}-TunF

With a highly concentrated pure sample of His_{10}-TunF in hand, two protein solutions were prepared for crystallisation studies (Table 4.3). Both contained 30 mg/mL TunF in 50 mM TEA pH7.8, diluted from the 57.8 mg/mL stock. In addition to this, one sample contained cofactor NAD^+ and substrate UDP-GlcNAc, while the other contained just NAD^+. Apart from the possibility of these cofactors/substrates stabilising the protein conformation and favouring crystallisation, the formation of tertiary or binary complexes would yield important information about the enzyme active site.

Sample	Contents
1	30 mg/mL His_{10}-TunF 50 mM TEA pH 7.8 3 mM NAD^+ 3 mM UDP-GlcNAc
2	30 mg/mL His_{10}-TunF 50 mM TEA pH 7.8 3 mM NAD^+

Table 4.3 TunF solutions used in crystallisation trial

These samples were each arrayed on seven commercial crystallisation screens at both 4 °C and 20 °C. Each screen consists of 96 different crystallant solutions on a single plate and each crystallant solution is screened against three different protein concentrations, thus each plate assays 288 crystallisation conditions. Since 28 plates were set up in total, over 8000 individual conditions were screened simultaneously for the crystallisation of TunF.

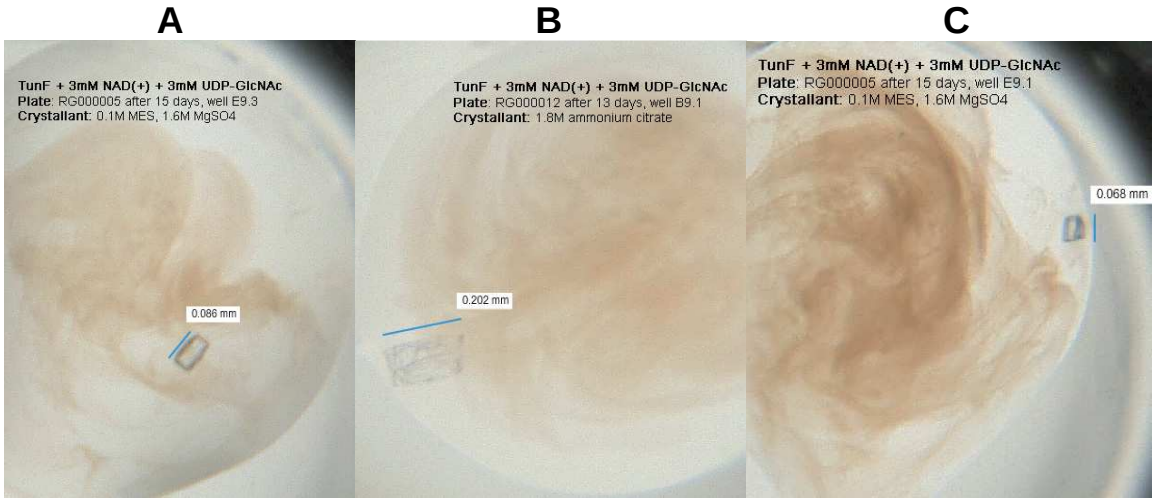


Fig. 4.7 Possible crystals of His_{10}-TunF from high-throughput Rigaku screen

A large proportion of the crystallisation conditions led to protein precipitation. This may suggest the protein was not as stable in the minimal TEA buffer as originally thought. Indeed, after 4 days of storage at 4 °C and 57.8 mg/mL, approximately half of this protein sample had precipitated. Overall, three crystals were obtained from this screen (Fig. 4.7). Two were <100 nm while the third showed cross-shaped nucleation. Upon initial diffraction analysis, crystals A and C were unfortunately lost during harvesting, and crystal B failed to show satisfactory diffraction. Further insight as to whether these crystals derived from protein or crystallised salts was not possible. Work is ongoing in our research group to undertake an optimisation screen using conditions from crystals A and C as a starting point (0.1 M MES pH6.5 and 1.6 M $MgSO_4$) and a frozen sample of TunF.

4.3 TunD: A putative N-acetylglucosaminyltransferase

4.3.1 Protein sequence analysis of TunD

Bioinformatic analysis of TunD and its putative role in tunicamycin biosynthesis was presented in section 2.4.3. Through a range of homology comparisons, the protein was predicted to be a GT-1 family glycosyltransferase possessing a GT-B fold. As the only predicted glycosyltransferase in the *tun* gene cluster TunD is likely to catalyse the formation of the an unusual α,β-1,1-glycosidic linkage between GlcNAc and tunicamine. Classification of glycosyltransferases is typically dependent on tertiary structure rather than primary sequence, so the assessment of associated substrates and reaction mechanisms by bioinformatic analysis alone is somewhat limited. *In vitro* characterisation of purified TunD would thus provide valuable information on the formation of such an unusual trehalose-type linkage and a more detailed assignment of glycosyltransferase family classification.

Protein	Amino acids	Amino acid sequence	Mw[a] / pI
Native TunD	472	MEIIFTVSDQVWGGKHRYMHDMALGLAQAGHTVTVLAEEGG AMLQQCRAAGVTTVSVPAFASDDAAEAVRKALRHRRPHIVCV SGRADAAAVHHAQSQGVTDAAVCLFRHSAFPLGTTDEVRDL FTGVNLVFTTSLEQRQRQFEPLINAGVLKDEQVEILTSGVGEP LLAALDAADRGAARKELRAESDQFVFLVLARLAWEKGIDQVID AFADLELPPDAAPPLLVVAGEGPLEAELRGQTIERGVAERVQF LGHQDHVAPVIKASDAVVLTSTVPETGPLALKEAMAAGRPVIA SVQGGIPEFVEDERHGLLVIDDEDLRQAMQRLLSDREAAETM GAAGSESVRGGHRAVRRVEYLAHRLDLLALEQLAPDTVLHEV VWDDVRLREETQGGFVFVPRTSHIMELDSATYAVVRTAVEAG DPLQLIKLPEETLGVIAHRLYAMGALVRQDGQATPAARTADGE RPVTEPA	Mw = 50832 Da pI = 5.01

[a] Average molecular weight;

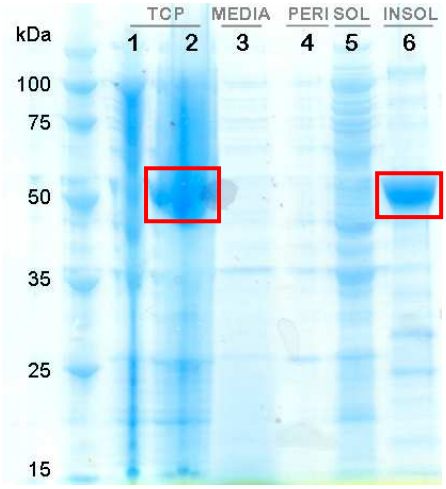
Table 4.4 Amino acid sequence and selected properties of TunD

4.3.2 Expression and purification of TunD

4.3.2.1 Expression of TunD as an N-terminal His_{10} fusion

The DNA sequence of *tunD* was codon optimised for *E. coli* and synthesised by Genscript. The gene was cloned into expression vector pET-16b (Novagen). The resulting gene product His_{10}-TunF is predicted to have 494 residues, an average Mw of 53,215 Da and a pI of 5.43. The plasmid, which we named pET-16b:*tunD*, was transformed into

E. coli BL21(DE3) according to the supplier's protocol (Novagen) and subjected to protein expression trials.

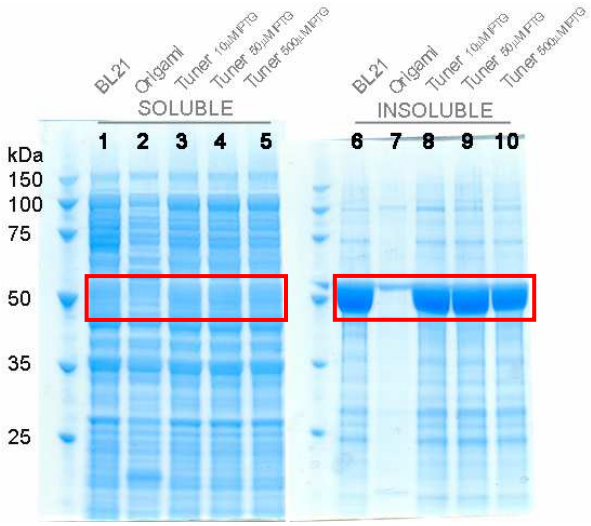


Lanes 1-2: Total cell protein (TCP) fraction after 0 and 16 h; Lane 3: Media fraction; Lane 4: Periplasmic fraction; Lane 5: Soluble cytosolic fraction; Lane 6: Insoluble cytosolic fraction;

Fig. 4.8 TunD expression in *E. coli* BL21(DE3)/pET16b:*tunD*, induced with IPTG at 20 ℃ for 16h

Expression trials were performed on an analytical scale and expression of TunD monitored by SDS-PAGE analysis of cell lysates, produced by BugBuster Mastermix (Novagen) treatment of collected cell pellets. Initial conditions involved induction with IPTG at 30 °C for 6 h, similar to conditions used with TunF. Good levels of expression were observed, as indicated by the presence of a strong band corresponding to TunD (~55 kDa) in the total cellular protein fraction. Unfortunately, all of the expressed protein appeared to aggregate as insoluble inclusion bodies and no soluble protein was observed. A range of expression conditions were thus screened in order to increase the proportion of soluble protein. Cultures following IPTG induction were grown at reduced temperatures in order to slow down the rate of protein production and allow more time for newly formed proteins to fold correctly.[392,393] Incubation times were increased in tandem to allow sufficient quantities of protein to be produced. As such, IPTG induction of TunD expression was additionally performed at 20 °C for 16 h and 14 °C for 20 h – a representative SDS-PAGE gel from these trials is shown in Fig. 4.8. Autoinduction protocols were also tested at 20 °C as these have been shown to offer a more gradual and controlled induction of protein expression[394] and detergents such as Triton X-100 were included in lysis buffers to assist protein solubilisation. Unfortunately, all the conditions tested yielded only insoluble TunD, with almost undetectable levels of soluble protein.

In pursuit of soluble protein, pET16b:*tunD* was transformed into alternative *E. coli* hosts: BL21(DE3)pLysS, Origami(DE3), Tuner(DE3), OrigamiB(DE3)pLysS and Rosetta-gamiB(DE3)pLysS (Novagen). These modified *E. coli* strains confer specific properties during protein expression.[392] The pLysS strains yield more stringent expression of toxic or difficult proteins due to basal expression of T7 lysozyme, a natural inhibitor of T7 RNA polymerase. Under pre-induction conditions, any background expression of T7 RNA polymerase and thus expression of the protein of interest contained in the pET vector is kept to a minimum. Origami strains contain glutathione reductase (*gor*) and thioredoxin reductase (*trxB*) mutations which greatly enhance the level of disulfide bonds during protein over-expression. B strains are termed 'Tuner' strains due to their expression of lac permease (*lacY1*) which allows a concentration dependant entry of IPTG into all cells, producing a concentration-dependant and homogenous level of protein induction. Finally, Rosetta strains express rare tRNAs allowing foreign genes containing uncommon codons to be efficiently translated. In our case the *tunD* gene has been codon optimised for *E. coli* so any effects in using these strains were expected to be moderate.

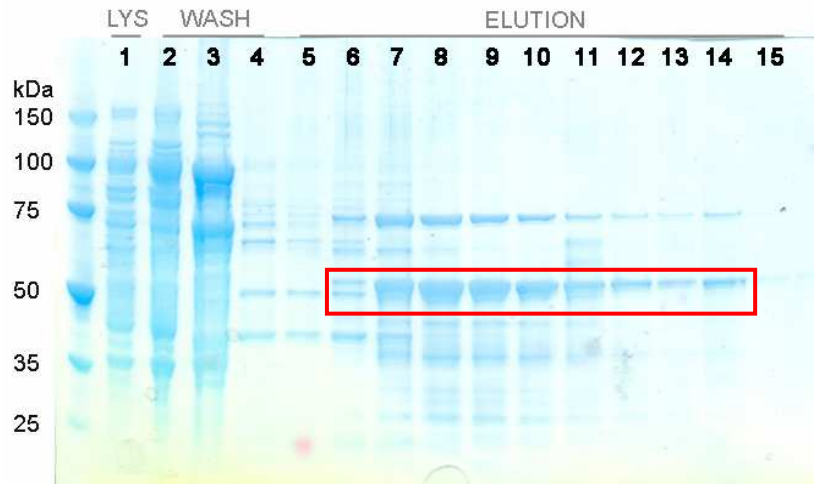


Lanes 1-5: Soluble cytosolic fractions for *E. coli* strains listed; Lanes 6-10: Insoluble cytosolic fractions for *E. coli* strains listed; Tuner strains induced at three different IPTG concentrations

Fig. 4.9 Expression of pET16b:*tunD* in various *E. coli* strains, induced with IPTG at 14℃ for 20 h

Under standard conditions involving IPTG induction at either 20 °C for 16 h or 14 °C for 20 h, it was observed that expression of *tunD* was severely repressed in any *E. coli* host carrying a pLysS mutation. One of the two remaining strains – Origami(DE3) – exhibited high expression levels but unfortunately offered no improvements to the proportion of soluble TunD. The other remaining strain – Tuner(DE3) – was induced with varying IPTG

concentrations. Although no noticeable solubility improvements were observed over the original BL21(DE3) strain good levels of expression were effected by IPTG concentrations as low as 10 µM, in contrast to the 500 µM used with other strains (Fig. 4.9).



Lane 1: Cell lysate from *E. coli* BL21(DE3)/pET16b:*tunD* expression; Lanes 2-4: Wash fractions; Lanes 5-15: Elution fractions with an imidazole gradient;

Fig. 4.10 Purification of His_{10}-TunD by Ni^{2+} affinity chromatography

Despite failing to achieve acceptable levels of soluble protein, it was decided to scale up the expression protocol and attempt Ni^{2+}-affinity purification of TunD; the original *E. coli* BL21(DE3) strain harbouring pET16b:*tunD* was selected for these experiments. Following IPTG induction and incubation for 24 h at 12 °C, cells were collected and lysed by a combination of lysozyme treatment and sonication. Cleared cell lysates were subjected to Ni^{2+}-affinity chromatography in order to specifically target the N-terminal His_{10} tag of the expressed TunD fusion protein. The UV trace obtained by the ÄKTA FPLC system (GE Healthcare) during the purification run indicated almost no TunD. SDS-PAGE analysis revealed protein bands corresponding to the expected Mw of His_{10}-TunD (~53 kDa), albeit in low concentration (Fig. 4.10). These fractions were pooled, dialysed and concentrated. Unfortunately, the protein was found to be unstable to protein concentration – using either an Amicon filter or Vivaspin – with extremely low concentrations being detected. It is likely that TunD was sufficiently unstable that it aggregated upon concentration, causing it to stick to dialysis membranes and resulting in an observed decrease in total protein content of these samples. Despite extensive efforts involving numerous purifications and alterations to the protocol, no further improvements could be achieved. These alterations involved the inclusion of additives in lysis and dialysis buffers such as 10% glycerol and 20 mM $MgCl_2$ to preserve stability and potential

activity respectively, 0.5% Triton X-100 detergent to solubilise the unstable protein and 1 mM TCEP to reduce any intermolecular disulfide bonds which could cause aggregation (TunD contains three cysteine residues).

4.3.2.2 Cloning TunD into alternative expression vectors

Following the mixed success in obtaining sufficient quantities of soluble TunD for functional studies, the *tunD* gene was ligated into several alternative expression vectors. These vectors encoded different TunD fusion proteins, each having the potential to encourage improved solubility of expressed recombinant protein from *E. coli* hosts. The features of each new fusion are summarised in Fig. 4.11. One vector encoding maltose-binding protein (MBP) fusions was selected, since this highly soluble protein has been shown to promote the solubility of poorly soluble proteins attached to them.[395,396] The vector used – pDB-His.MBP – additionally encoded an N-terminal His_6 tag for facile Ni^{2+}-affinity purification. Two further vectors encoded glutathione *S*-transferase (GST) fusions; again these have been shown to promote protein solubility.[14-16] The second of these vectors incorporated an N-terminal His_6 tag through judicious choice of PCR primers and restriction enzymes during the cloning process. The final vector used encoded a C-terminal His_6 tag. Moving the poly-histidine tag from one terminus of the target protein to the other can alter the physio-chemical properties of the protein and in some cases lead to improved stability or solubility.[392]

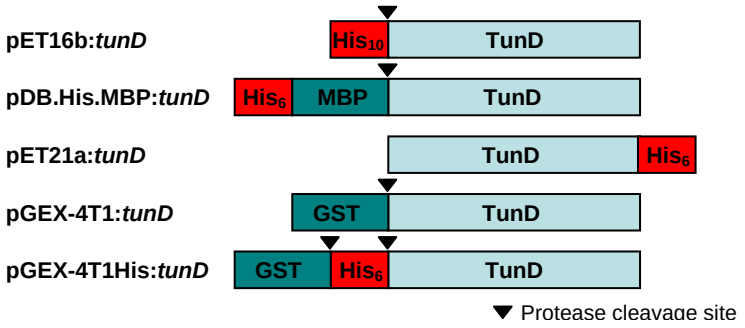
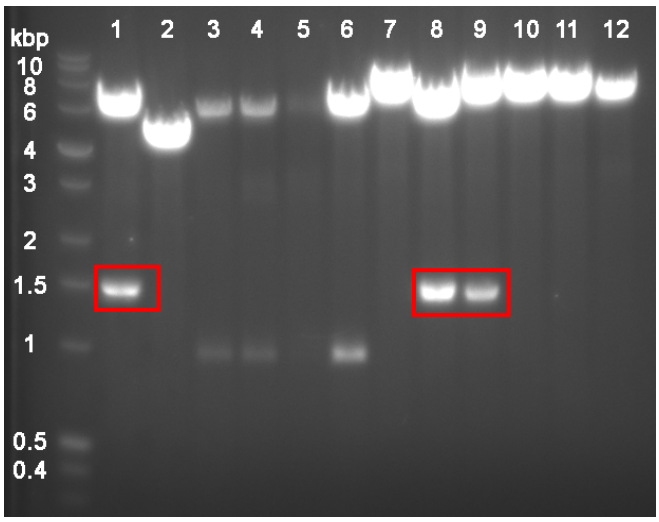


Fig. 4.11 TunD fusions encoded by alternative expression vectors trialled

pDB.His.MBP was propagated in *E. coli* DH5α and linearised with *Bam*HI and *Nde*I restriction endonucleoases. A sample of pET16b:*tunD* was similarly digested and the resulting *tunD* fragment was isolated and ligated into the pDB.His.MBP vector. Restriction digest analysis of clones indicated successful ligation (Fig. 4.12), with the positive clones further verified by sequence analysis. This novel plasmid was transformed into BL21(DE3),

BL21(DE3)pLysS, Origami(DE3) and Tuner(DE3) *E. coli* strains for protein expression trials.

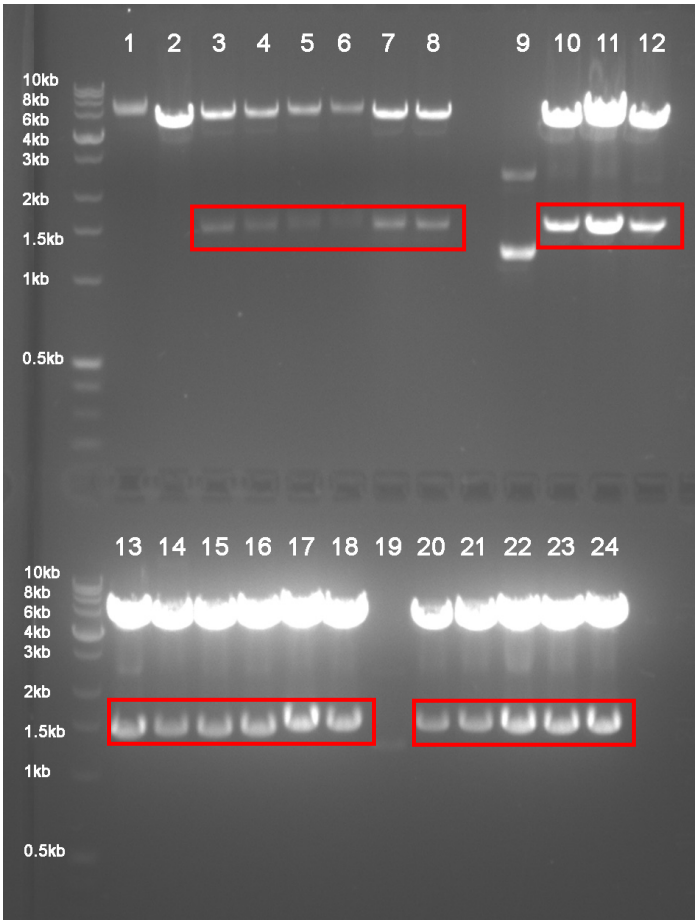


Lanes 1, 8 and 9 show plasmids containing a ~1.4 kb band corresponding to tunD (in red boxes);

Fig. 4.12 Confirmation digest of putative pDB.His.MBP:*tunD* plasmids following ligation

When cloning *tunD* inserts from pET16b:*tunD* for ligation into pET21a or pGEX-4T1, careful design of appropriate PCR primers was required. 5´ overhangs and single base changes were included so as to introduce additional restriction sites at either end of the *tunD* gene. For pET21a, an *Nco*I restriction site was introduced before the start of *tunD* and the stop codon (TAA) was mutated to a glutamic acid residue (GAA) at the 3´ end of the gene. For pGEX-4T1, a *Bam*HI restriction site was introduced before the start of *tunD*. To create pGEX-4T1His, a *Bam*HI restriction site was introduced before the start of the His_6 tag present in pET16b:*tunD*, allowing the gene to be cloned into pGEX-4T1 along with its N-terminal His_6 tag. This created a novel vector and protein fusion, offering potentially increased protein solubility thanks to the GST tag together with facile purification *via* the His_6 tag.

The gene was successfully amplified by PCR in each case and following digestion samples of each vector and corresponding amplified *tunD* genes with the relevant restriction enzymes, ligation afforded positive clones in each case. These novel DNA constructs were verified by restriction digest analysis of clones (Fig. 4.13) and sequence analysis. The resulting plasmids were all transformed into *E. coli* BL21(DE3) for protein expression trials.

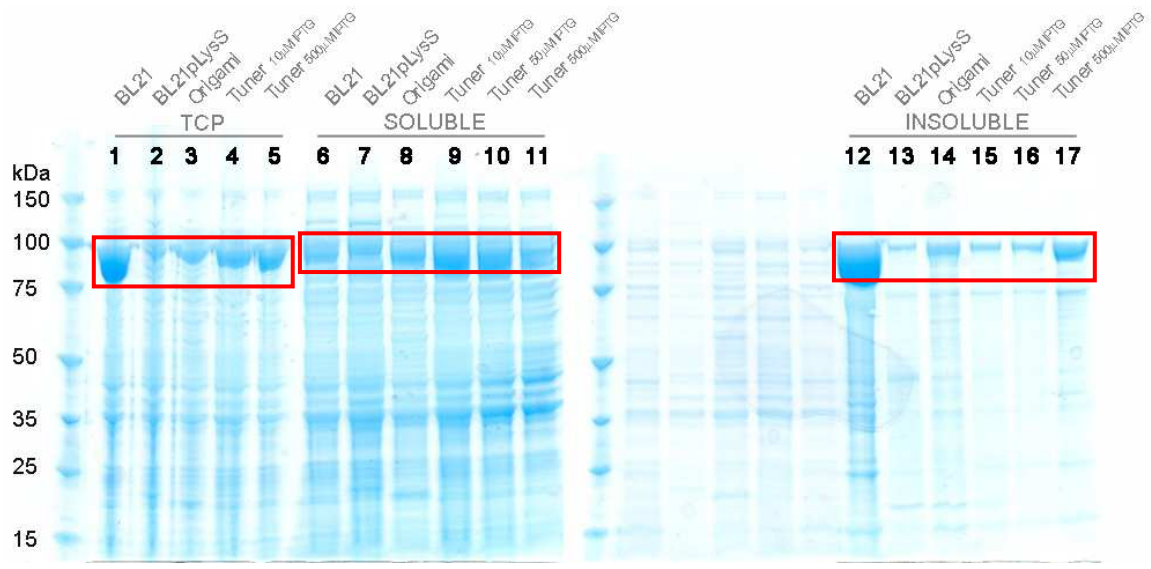


Lanes 1-8: Eight plasmids from pET21a:*tunD* ligation, digested with *Nde*I/*Xho*I;
Lanes 9-16: Eight plasmids from pGEX-4T1:*tunD* ligation, digested with *BamH*I/*Xho*I;
Lanes 17-24: Eight plasmids from pGEX-4T1His:*tunD* ligation, digested with *BamH*I/*Xho*I;
Bands corresponding to *tunD* (~1.4 kb) are boxed in red.
The identity of plasmids in lanes 7, 10 and 17 was confirmed by sequencing.

Fig. 4.13 Confirmation digest following *tunD* ligation into pET21a, pGEX-4T1 and pGEX-4T1His

4.3.2.3 Expression of TunD fusions from alternative vectors

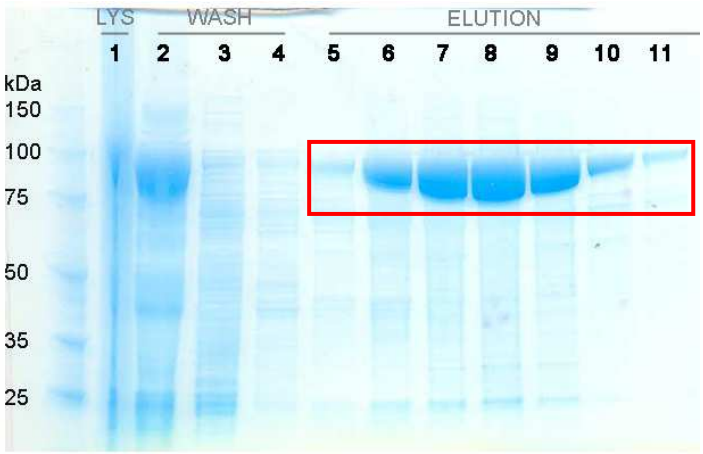
E. coli strains BL21(DE3), BL21(DE3)pLysS, Origami(DE3) and Tuner(DE3) harbouring plasmid pDB.His.MBP:*tunD* were all subjected to analytical scale expression trials. A number of temperatures and IPTG concentrations – for lac permease-expressing Tuner cells – were screened. At 30 °C only inclusion bodies were observed, but upon lowering the temperature to 14 °C weak bands corresponding to His_6-MBP-TunD (~93 kDa) were observed (Fig. 4.14). The highest level of soluble fusion protein appeared to occur using the *E. coli* Tuner(DE3) strain with a low IPTG induction concentration of 10 µM.



Lanes 1-5: Total cell protein for *E. coli* strains listed; Lanes 6-11: Soluble cytosolic fractions for *E. coli* strains listed; Lanes 12-17: Insoluble cytosolic fractions for *E. coli* strains listed;

Fig. 4.14 Expression screen of pDB.His.MBP:*tunD* at 14 ℃ for 20 h in different *E. coli* hosts

Following these expression trials, the *E. coli* Tuner(DE3)/pDB.His.MBP:*tunD* strain with a 10 µM IPTG induction was subjected to a scaled up expression and Ni^{2+}-affinity purification on an ÄKTA FPLC system (Fig. 4.15). 0.5% Triton X-100 and 10% glycerol were included in the lysis buffer to maximise the stability of any soluble protein.



Lane 1: Cell lysate from *E. coli* Tuner(DE3)/pDB.His.MBP:*tunD* expression; Lanes 2-4: Wash fractions; Lanes 5-11: Elution fractions with an imidazole gradient;

Fig. 4.15 Purification of His_6-MBP-TunD by Ni^{2+} affinity chromatography

A reasonable level of soluble protein was isolated in good purity as judged by SDS-PAGE. Unfortunately, no activity was observed under later activity screens trialled. A common cause of this is the MBP fusion itself. MBP is extremely soluble, but the target protein may not be correctly folded even though the overall fusion appears to be soluble. It is therefore advisable to cleave the fusion after initial purification in order to ascertain the state of the target protein. The vector used encodes a TEV (tobacco etch virus) protease cleavage site between the MBP and TunD. Attempted cleavage was thus performed with commercial AcTEV protease (Invitrogen) at various temperatures. Unfortunately, the level of cleavage was always very low and uncleaved fusion protein remained in all cases, even after extended protease incubation. Attention was thus focused on expression of the remaining *tunD* constructs previously created, in our quest for a soluble, active sample of TunD glycosyltransferase.

Lanes 1-3: Total cell protein for *E. coli* strains harbouring plasmids listed; Lanes 4-6: Soluble cytosolic fractions; Lanes 7-9: Insoluble cytosolic fractions; Predicted sizes of TunD fusions: TunD-His_6 = 53 kDa, GST-TunD = 75 kDa, GST-His_6-TunD = 76 kDa;

Fig. 4.16 Test expression of *E. coli* BL21(DE3) harbouring pET21a:*tunD*, pGEX-4T1:*tunD* and pGEX-4T1His:*tunD* at 18 ℃ for 18 h

A test expression was performed with IPTG induction at 18 °C for 18 h using *E. coli* BL21(DE3) strains each harbouring one of the new plasmid constructs: pET21a:*tunD*, pGEX-41T:*tunD* and pGEX-41This:*tunD*. As shown in Fig. 4.16, expression levels were high for all three fusions, although in each case the majority of the protein resided in the insoluble fraction as inclusion bodies. However, the soluble fraction of the pET21a:*tunD* expression contained a clear band corresponding to TunD-His_6 (~53 kDa). Expression of this *E. coli* strain was thus scaled up ready for Ni^{2+} affinity purification of this C-terminally His_6-tagged TunD fusion. The resulting cell pellet was lysed by standard methods and applied to a HisTrap HP column (GE Healthcare). SDS-PAGE analysis of chromatography fractions indicated protein elution during the imidazole gradient of a protein with a size corresponding to expected Mw of TunD-His_6 (~53 kDa) and the levels of soluble protein appeared greater than with N-terminally His-tagged TunD from vector pET16b (Fig. 4.17). The relevant fractions were pooled, dialysed and concentrated to 9.1 mg/mL. Little or no protein precipitation or degradation was observed upon protein concentration and the sample was stable to storage at 4 °C – unlike with previously expressed His_{10}-TunD.

Lane 1: Cell lysate from *E. coli* BL21(DE3)/pET21a:*tunD* expression; Lane 2: Insoluble fraction from this expression; Lanes 3-4: Wash fractions from Ni^{2+} purification; Lanes 5-9: Elution fractions;

Fig. 4.17 Purification of TunD-His_6 by Ni^{2+} affinity chromatography

4.3.3 TunD activity assays

4.3.3.1 Spectrophotometric coupled assay for glycosyltransferases

In 1994 Gosselin *et al.* reported a continous spectrophotometric assay for nucleotide-diphospho-sugar dependent glycosyltransferases (Fig. 4.18).[397] In the glycosyl transfer reactions catalysed by these enzymes, a nucleotide disphosphate is released from the NDP-sugar donor. This product is then coupled to the activity of pyruvate kinase, which transfers a phosphate group from phosphoenolpyruvate (PEP) to NDP, generating NTP and pyruvate. The pyruvate molecule in turn acts as a substrate for lactate dehydrogenase, which uses NADH to reduce this compound to lactate and liberates NAD^+. Since NADH has a specific absorption at 340 nm absent from NAD^+, progression of the glycosyl transfer can be monitored spectrophotometrically by supplying exogenous PEP, NADH, pyruvate kinase and lactate dehydrogenase to the glycosyltransferase reaction mixture. A decrease in A_{340} absorbance indicates consumption of NDP-sugar catalysed by the assayed glycosyltransferase.

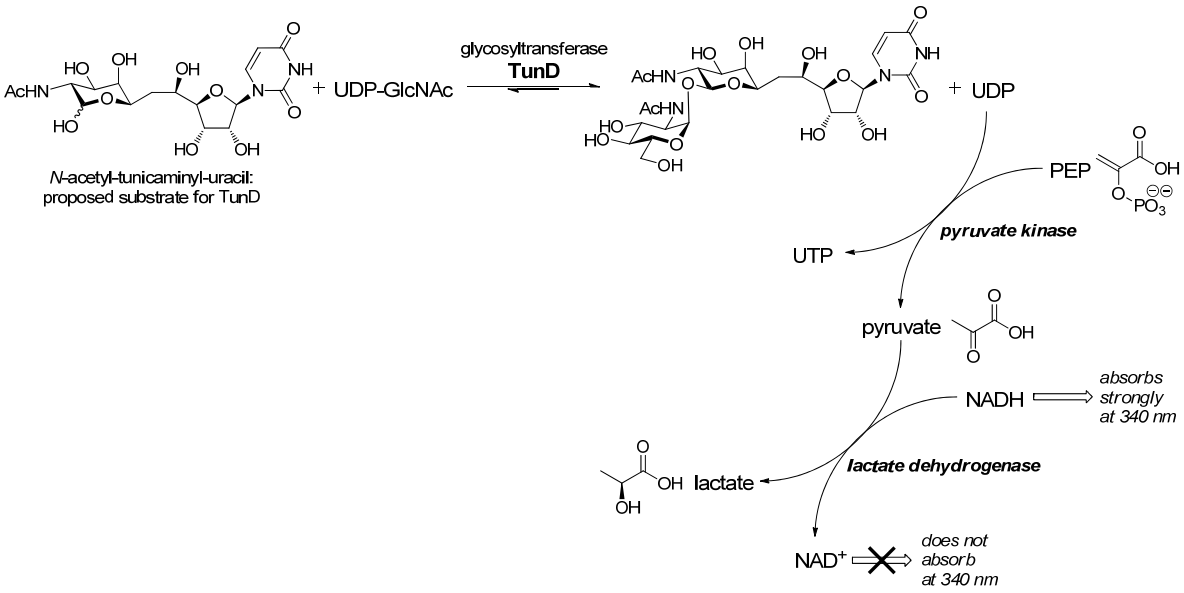


Fig. 4.18 Overview of spectrophotometric coupled assay for TunD activity

The previously purified sample of TunD-His_6 was thus incubated with 1mM UDP-GlcNAc and 3mM *N*-acetyl-tunicaminyl-uracil – the two suspected substrates for TunD – in a 96-well microplate. Remaining components of the coupled assay were added and the plate incubated at 30 °C within a spectrophotometer; absorbance readings at 340 nm were taken every 10 s. Reactions were all performed in 50 mM MOPS buffer across a range of

pH values, in order to detect any potential pH dependence on enzyme activity. In control reactions the TunD-His_6 sample was additionally inactivated by heating to 90 °C for 10 min prior to use; all other parameters were unchanged.

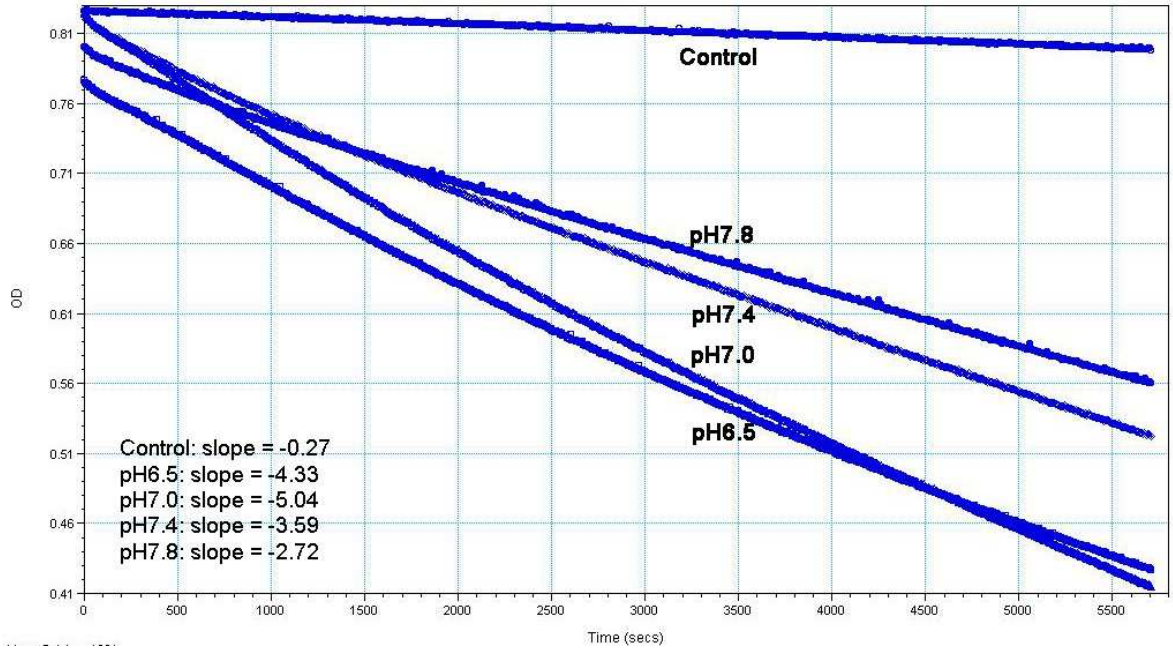


Fig. 4.19 Spectrophotometric coupled assay of TunD-His_6 activity across different pH values

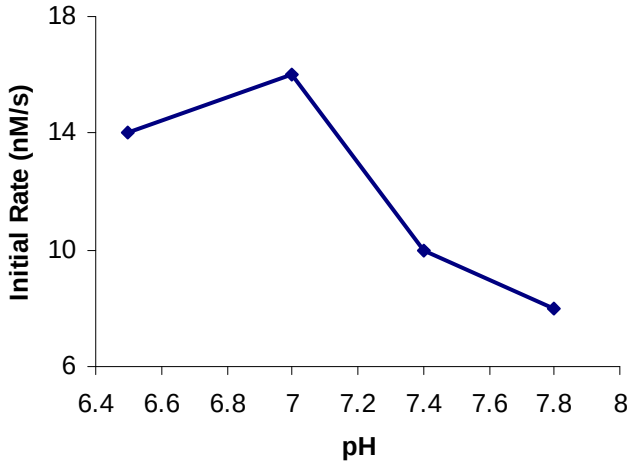


Fig. 4.20 pH profile of TunD activity

A clear difference was observed in the slopes of the control and the enzyme-containing assays (Fig. 4.19). This indicated that UDP was being released by the action of our TunD-His_6 sample, causing NADH to be consumed which translated to a decrease in A_{340}. These results suggested that TunD-His_6 is indeed acting as a glycosyltransferase with the substrates provided in the assay (UDP-GlcNAc and *N*-acetyl-tunicaminyl-uracil).

Using the NADH extinction coefficient at 340 nm (ε_{340} = 6,220 $M^{-1}cm^{-1}$) and the Beer-Lambert Law, the concentration of NADH at given time points was calculated. Initial rates of reaction (V_0) were thus calculated and plotted against pH (Fig. 4.20). In the resulting pH profile, the greatest relative activity was observed at pH 7.0. It should be noted, however, that the overall TunD-His_6 concentration in the assay was very high (~230 µg/mL), meaning the observed activity was in fact very low. The assay was repeated, this time screening a range of buffers at pH 7.0, namely: MOPS, Tris, HEPES and sodium phosphate. All buffers showed similar levels of activity, but HEPES was marginally the most active.

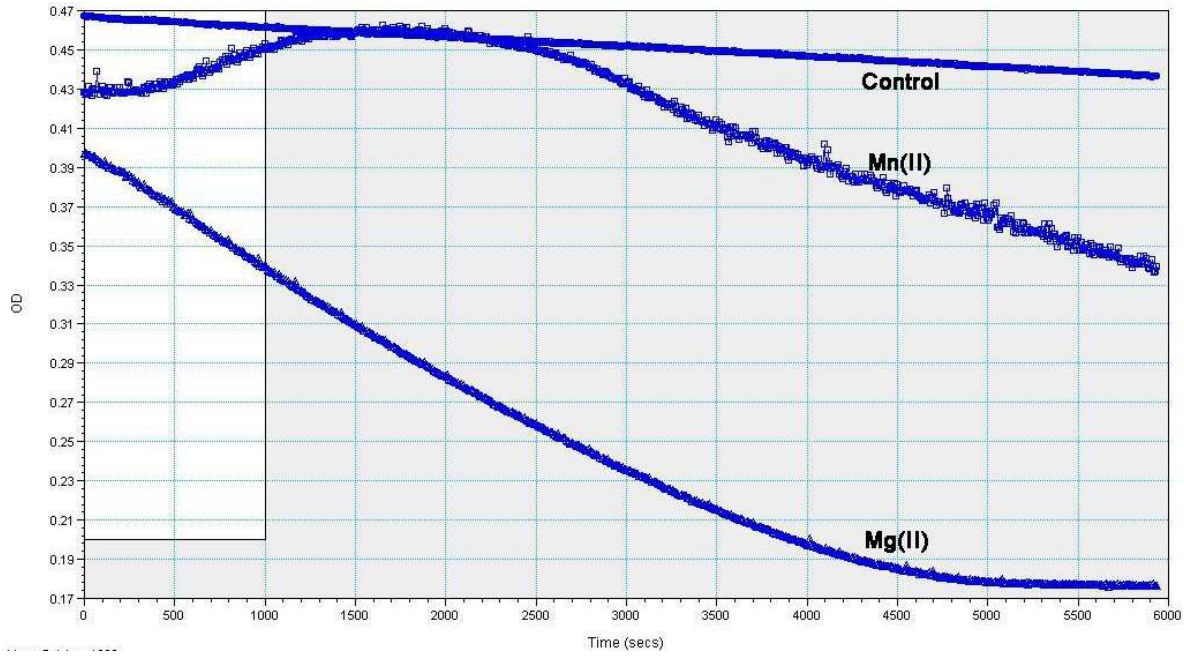


Fig. 4.21 Spectrophotometric coupled assay of TunD-His_6 activity with Mg^{2+} and Mn^{2+} cofactors

Glycosyltransferases are often dependent on the presence of an active site divalent metal ion – typically Mg^{2+} or Mn^{2+} – which stabilises the negatively charged pyrophosphate group of the bound NDP-sugar donor and the produced NDP by-product. Metal ion-independent glycosyltransferases have also been documented, utilising alternative strategies to stabilising these active site negative charges.[398] In the previous spectrophotometric assays, reaction mixtures were supplemented with $MgCl_2$ as a source of Mg^{2+} ions. Now this was replaced with $MnCl_2$ to test whether any changes in reaction rate could be observed and gain a deeper insight into the metal-ion dependence of TunD. A curious result was observed whereby upon Mn^{2+} addition, a long lag occurred before any activity was detected (Fig. 4.21). This suggested that in this time the tightly bound metal cation in the TunD active site may have been exchanging with the high Mn^{2+} ion concentration in the surrounding solution. This observation suggested Mg^{2+} was the native metal cofactor in

TunD, although there appeared to be little difference between Mn^{2+} or Mg^{2+} cofactors with respect to the initial rates of reaction. It should be noted that these results are far from conclusive and a more thorough investigation is required to support this hypothesis.

Reaction mixtures from TunD-His_6 enzymatic assays – lacking the additional components from the previous spectrophotometric coupled assay – were analysed by HPLC, with monitoring by UV or evaporative light-scattering detection. Surprisingly, no product formation was observed even though a very high TunD-His_6 concentration was used (480 µg/mL). The possibility was considered that *N*-acetyl-tunicaminyl-uracil was not the natural substrate for TunD and that the observed activity arose purely from the enzyme-catalysed hydrolysis of UDP-GlcNAc. To test this hypothesis, a spectrophotometric coupled assay was devised to compare the activity of TunD with both substrates present against the activity with only UDP-GlcNAc present. As shown in Fig. 4.22, the rate of NADH consumption using solely UDP-GlcNAc was the same as with both substrates present, or in the case where GlcNAc replaced TunD substrate as an alternative acceptor. This result suggested that the activity observed in previous assays derives from hydrolysis of UDP-GlcNAc but not concomitant transfer to the proposed glycosyl acceptor.

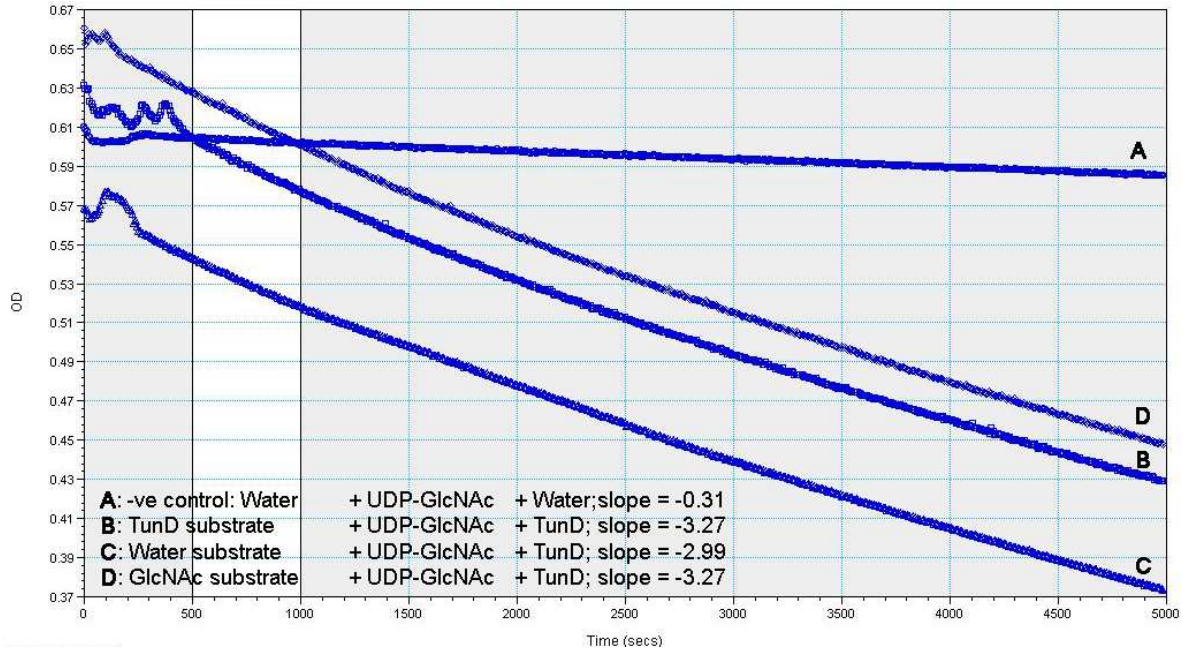


Fig. 4.22 TunD-His_6 coupled assay testing relative rates of UDP-GlcNAc hydrolysis

4.4 *TunE: A putative* N-*deacetylase*

4.4.1 *Protein sequence analysis of TunE*

Bioinformatic analysis of TunE and its putative role in tunicamycin biosynthesis was presented in section 2.4.4. Through a range of homology comparisons, the protein was predicted to belong to the PIG-L superfamily, which contains *N*-deacetylases from GPI anchor[399] and mycothiol biosynthesis.[400] The substrates for both of these enzyme groups share key similarities with likely intermediates in the tunicamycin biosynthetic pathway and typically involve the deacetylation of an *N*-acetylglucosamine moiety. Indeed TunE contains an unusual HPDD zinc-binding motif conserved across both of these enzyme families.[401] TunE was proposed to act on substrate *N*-acetyl-*N*-deacyl-tunicamycin, catalysing a Zn^{2+}-mediated cleavage of this compound's 10´-acetamido group to reveal a free amine, in preparation for subsequent fatty acid attachment. *In vitro* characterisation of purified TunE would not only verify this proposed function and the predicted natural substrate, but also give access to selectively *N*-deacetylated intermediates which could be diversified into tunicamycin analogues through facile amine alkylation or acylation.

Protein	Amino acids	Amino acid sequence[a]	Mw[b] / pI
Native TunE	234	VKVLVIAAHPDDEVLGAGATVSALSQQGAVVQIHILAEGISLRH SGVKIEEARERCQVAAKDLGAAEVSFGGLAADGRLLADLPQR QVVDAVSHALREAEPEVVFTHHPGDIHADHRSVAHAVGYGT RILGSGSVRQVLHFEVLSSTEQQTGLVAPFTPNLFYDVTGHV EAKCRALAAYPYELYDPPHPRSLAAVRTLASYRGTQVGVEAA EAFMIGRELRGPMAAVPSGVDVR	Mw = 24799 Da pI = 5.93

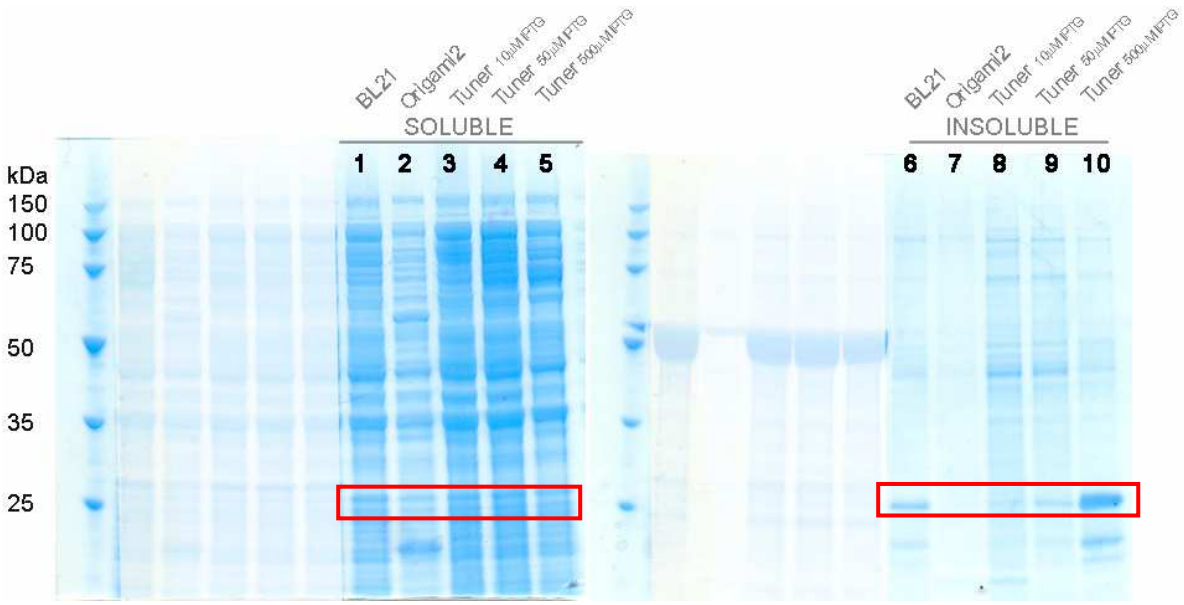
[a] Conserved Zn^{2+}-binding motif highlighted in yellow; [b] Average molecular weight;

Table 4.5 Amino acid sequence and selected properties of TunE

4.4.2 *Expression and purification of TunE*

4.4.2.1 *Expression of TunE as an N-terminal His_6 fusion*

The DNA sequence of *tunE* was codon optimised for *E. coli* and synthesised by Genscript. The gene was cloned into expression vector pET-28a(+) (Novagen). The resulting gene product His_6-TunE is predicted to have 254 residues, an average Mw of 26,994 Da and a pI of 6.33. The plasmid, which we named pET28a:*tunE*, was transformed into *E. coli* BL21(DE3), Origami2(DE3) and Tuner(DE3) strains according to the supplier's protocols (Novagen) and subjected to protein expression trials.



Lanes 1-5: Soluble cytosolic fractions for *E. coli* strains listed; Lanes 6-10: Insoluble cytosolic fractions for *E. coli* strains listed; Tuner strains induced at three different IPTG concentrations

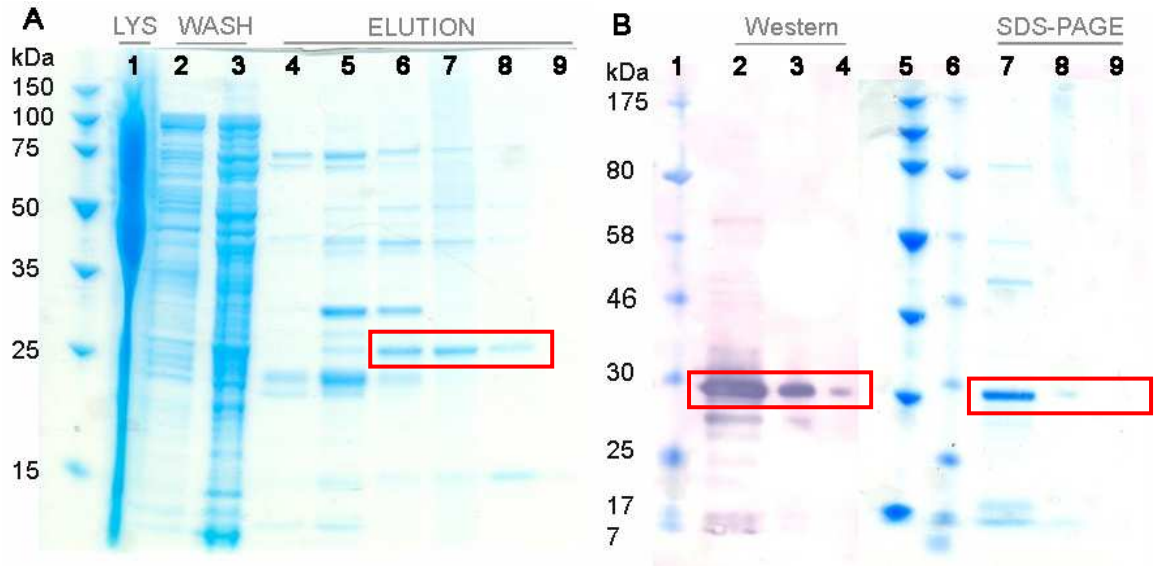
Fig. 4.23 Expression of pET28a:*tunE* in various *E. coli* strains, induced with IPTG at 14 ℃ for 20 h

Analytical scale expression of TunE was monitored by SDS-PAGE analysis of cell lysates, produced by BugBuster Mastermix (Novagen) treatment of collected cell pellets. Initial conditions involved induction with IPTG at 30 °C for 6 h, similar to conditions used with TunF. Good levels of expression were observed in all recombinant *E. coli* strains, as indicated by the presence of a strong band corresponding to His_6-TunE (~27 kDa) in the total cellular protein fraction. Unfortunately, all of the expressed protein appeared to aggregate as insoluble inclusion bodies and no soluble protein was observed. The use of autoinduction conditions with *E. coli* BL21(DE3)/pET28a:*tunE* offered no improvements, so IPTG induced expression was repeated at 14 °C and allowed a longer incubation time of 20 h. Under these conditions a small proportion of TunE was detected in the soluble cytoplasmic fraction (Fig. 4.23) with both BL21(DE3) and Tuner(DE3) *E. coli* strains. The former strain was selected for scaled up expression and protein purification.

4.4.2.2 Purification of His_6-TunE

E. coli BL21(DE3)/pET28a:*tunE* strain was grown in 6 × 800 mL cultures and cell pellets collected. A high ionic strength lysis buffer used supplemented with 0.5% Triton X-100 and 10% glycerol to maximise protein solubility. The resulting lysate was passed through a Ni^{2+} affinity column on an ÄKTA FPLC system (Fig. 4.24). A small amount of

protein corresponding in size to His_6-TunE (~26 kDa) was observed. Due to the low abundance, a Western blot with monoclonal anti-polyhistidine-alkaline phosphatase antibody was performed, which verified that the observed band was indeed His_6-TunE and not a contaminating protein from the *E. coli* host.



A: Ni^{2+}-affinity purification of His_6-TunE; Lane 1: Cell lysate from *E. coli* BL21(DE3)/pET28a:*tunE* expression; Lanes 2-3: Wash fractions from Ni^{2+} purification; Lanes 4-9: Elution fractions;
B: Western blot of His_6-TunE; Lanes 1 and 6: Prestained Markers (NEB); Lanes 2-4: Anti-polyhistidine Western blot against 1×, 0.1× and 0.01× dilutions of His_6-TunE; Lane 5: 0.1-12kb Perfect Protein Markers (Novagen); Lanes 7-9: SDS-PAGE gel of 1×, 0.1× and 0.01× dilutions of His_6-TunE ran in parallel with Western Blot for comparison;

Fig. 4.24 Ni^{2+} affinity purification and Western blot of His_6-TunE

4.4.3 Attempted assay of TunE activity

The previously purified sample of His_6-TunE was incubated with *N*-acetyl-*N*-deacyl-tunicamycin, previously synthesised from tunicamycin in section 3.3.2. Reaction mixtures were analysed by HPLC and progress monitored by UV absorbance at 260 nm. Unfortunately, despite trialling a number of reaction buffers and using fresh enzyme preparations, no activity was observed in these assays. The low levels of soluble His_6-TunE obtained suggests that this fusion is unstable in solution and may undergo facile precipitation or inactivation.

4.5 Conclusions

In this chapter we have initiated studies pertaining to the *in vitro* characterisation of individual Tun enzymes, focussing on epimerase TunF, glycosyltransferase TunD and *N*-deacetylase TunE. We have developed greatly optimised protocols for their over-expression and purification from recombinant *E. coli* hosts, allowing subsequent *in vitro* studies which have offered valuable insights into the functions of these enzymes. The work presented in this chapter directly complements proposals made in chapter 2 as to the pathway to tunicamycin biosynthesis and lays the groundwork for future investigation of these and other enzymes encoded by the *tun* gene cluster.

4.5.1 Epimerase TunF

First of all, we successfully cloned and over-expressed TunF as an N-terminal His_{10} fusion protein and experimentally showed it to function as a UDP-GlcNAc-4-epimerase. This is in line with previously reported predictions and those made in section 2.4.2, where UDP-GlcNAc rather than UDP-GalNAc was proposed to act as a metabolic precursor in tunicamine biosynthesis.[402] The TunF is necessary to install the pseudo-GalNAc moiety within tunicamine by epimerisation of a *gluco*- configured precursor at some stage during the biosynthetic pathway. Unfortunately, despite verifying that TunF acts as this crucial epimerase, the precise identity of its true *in vivo* substrate remains unclear. As discussed in section 2.4.2, TunF could act at a number of stages in the biosynthetic pathway. Depending on whether it acts before, during or after the undecose C–C bond-forming event a different UDP-*N*-acetylglucosamine derivative can be rationalised as a potential substrate in each case (Fig. 4.25). Work in our research group has recently established UDP-GlcNAc as a substrate for dehydratase TunA, yielding UDP-5,6-ene-GlcNAc.[403] Following co-incubation of UDP-GlcNAc with TunA and TunF, a further product was detected – suggested to be UDP-GalNAc-5,6-ene. These results have implications on the findings presented here, since it would appear that UDP-GlcNAc-5,6-ene can also act as a substrate for epimerase TunF (Fig. 4.25, entry b). Together, these findings go some way to deciphering the true order of biosynthetic events in the construction of tunicamine. However, further work is still required in order to determine the true substrate for TunF, not least by gaining a deeper understanding of the carbohydrate tail-to-tail coupling event likely catalysed by TunA, TunB and TunM. Access to potential substrates UDP-GlcNAc-5,6-ene and UDP-8´-*epi*-tunicamine is also required; these compounds are unfortunately much more difficult to obtain through standard synthetic methods.

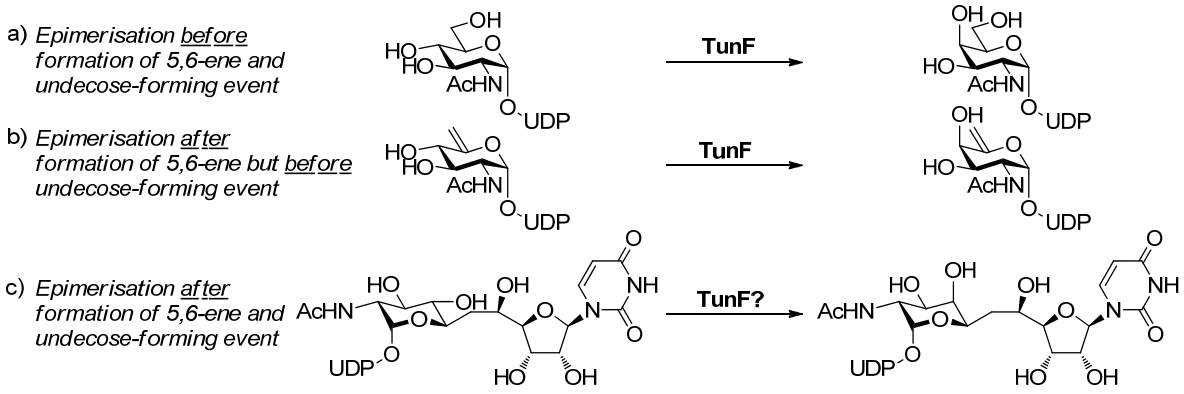


Fig. 4.25 Possible substrates for epimerase TunF

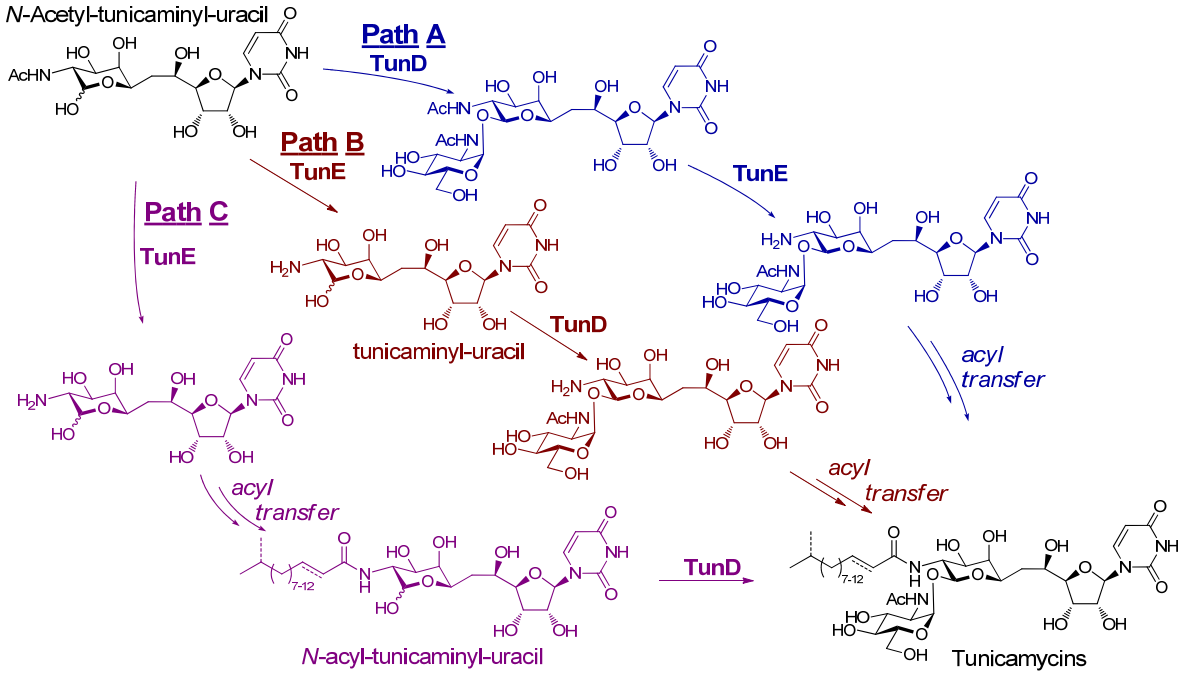
In addition to demonstrating functional activity of TunF, we have developed a novel enzymatic assay based on ^{1}H NMR. The method offers a simple and rapid method for detecting enzymatic activity and has the added benefit of allowing accurate elucidation of product structures directly from reaction mixtures, eliminating the need for laborious chromatographic separation and purification of reaction products.

4.5.2 Glycosyltransferase TunD

The second enzyme discussed herein was putative glycosyltransferase TunD. Over-expression and purification of glycosyltransferase TunD from *E. coli* proved to be much less facile than that of TunF. Significant hurdles resulting from poor protein solubility and stability were overcome through the systematic screening of a wide range of conditions for heterologous protein expression and through the production of a number of fusion protein constructs with different properties. In the end an optimised procedure for TunD expression was developed, involving production of a C-terminal His_6 fusion and allowing subsequent efficient purification by Ni^{2+}-affinity chromatography.

This TunD-His_6 construct was subjected to extensive activity assays, providing unexpected insights into TunD activity and its role in tunicamycin biosynthesis. Biological activity was detected with a coupled spectrophotometric assay, although the results showed that the observed UDP-GlcNAc hydrolysis catalysed by TunD was not coupled to glycosyl transfer to the *N*-acetyl-tunicaminyl-uracil acceptor. This was surprising and suggested that *N*-acetyl-tunicaminyl-uracil was not the natural substrate for glycosyltransferase TunD (Scheme 4.1, Path A). It is possible that the proposed order of biosynthetic events is incorrect, such that *N*-acetyl-tunicaminyl uracil may first undergo deacetylation through the action of TunE and then *N*-acylation. This would mean that the natural substrate for TunD

is in fact *N*-acyl-tunicaminyl-uracil (Scheme 4.1, Path C). Alternatively, GlcNAc transfer to tunicaminyl-uracil may occur directly after *N*-deacetylation and before *N*-acylation, requiring tunicaminyl-uracil – with a free amine group – as the preferred substrate for TunD (Scheme 4.1, Path B). Work is currently underway in our research group to further test these hypotheses, building on the important initial findings presented herein. The optimised conditions for TunD expression and purification will be of great use in these studies and in themselves provide a valuable contribution to the study of this key glycosyltransferase.



Scheme 4.1 Implications of observed TunD-His_6 activity on tunicamycin biosynthetic pathway

4.5.3 N-deacetylase TunE

Initial investigations into the over-expression and purification of putative *N*-deacetylase TunE have also been presented. Soluble TunE was obtained in low levels as an N-terminal His_6 fusion and unfortunately failed to show activity in a HPLC assay utilising *N*-acetyl-*N*-deacyl-tunicamycin as the proposed natural substrate for this enzyme (Scheme 4.1, Path A). The results suggest that this fusion is unstable in solution and may undergo facile precipitation or inactivation following expression in *E. coli*. Indeed, the enzyme is predicted to bind Zn^{2+} at its N-terminus through a conserved HPDD motif. The presence of an N-terminal His_6-tag and subsequent purification by Ni^{2+} affinity purification is likely to bring nickel ions into close proximity of the enzyme bound Zn^{2+} ion, potentially perturbing its binding and adversely affecting its catalytic activity. It has been previously shown that charging metal-chelating resins with Zn^{2+} rather than Ni^{2+} can still afford selective affinity chromatography of His-tagged proteins whilst avoiding the replacement of active site Zn^{2+} with Ni^{2+} ions.[404] Additionally, transferring the His-tag to the C-terminus is likely to reduce any perturbation of the N-terminal Zn^{2+}-binding site.

Another possible explanation for the lack of activity observed is that *N*-acetyl-*N*-deacyl-tunicamycin is not the native substrate for TunE. We have already hypothesised a number of possible substrates for glycosyltransferase TunD following analysis of experimental data. Each of the scenarios summarised in Scheme 4.1 alter the biosynthetic pathway of tunicamycin and suggest alternative substrates not only for TunD but also for *N*-deacetylase TunE. Although only preliminary investigations into the expression, purification and functional verification of TunE have been presented herein, they act as a solid foundation for the future construction of alternative TunE constructs and functional enzyme assays.

4.6 Experimental

4.6.1 Bacterial strains

***E. coli* strain**	**Description**	**Antibiotic resistance**	**Source/ reference**
BL21(DE3)	General purpose expression host	–	Novagen[a]
BL21(DE3)pLysS	High-stringency expression host derived from *E. coli* BL21(DE3)	Chloramphenicol^R	Novagen[a]
Origami(DE3)	General expression host derived from *E. coli* K-12. Two mutations in cytoplasmic disulfide reduction pathway – to genes *trxB* and *gor* – enhance disulfide formation in *E. coli* cytoplasm	Kanamycin^R, Streptomycin^R, Tetracycline^R	Novagen[a]
Origami2(DE3)	The same properties as Origami(DE3) but is additionally kanamycin sensitive	Streptomycin^R, Tetracycline^R	Novagen[a]
Tuner(DE3)	General expression host derived from *E. coli* BL21(DE3). Mutation to lac permease (*lacY1*) allows adjustable levels of protein expression dependent on IPTG concentration	–	Novagen[a]
Rosetta(DE3)	General expression host derived from *E. coli* BL21(DE3). Carries plasmid providing six rare codon tRNAs	Chloramphenicol^R	Novagen[a]
OrigamiB(DE3) pLysS	Expression host derived from *E. coli* BL21(DE3)pLysS. Additionally contains mutations and associated properties of Origami(DE3) and Tuner(DE3) strains	Chloramphenicol^R Kanamycin^R, Tetracycline^R	Novagen[a]
Rosetta-gamiB (DE3)pLysS	Expression host derived from *E. coli* BL21(DE3)pLysS. Additionally contains mutations and associated properties of Origami(DE3), Tuner(DE3) and Rosetta (DE3) strains	Chloramphenicol^R Kanamycin^R, Tetracycline^R	Novagen[a]

[a] For comprehensive descriptions of these strains see ref. [392]

Table 4.6 Details of *E. coli* strains used in this chapter

4.6.2 Plasmids

Name	Description	Size, Antibiotic resistance	Source/ reference
pET16b	*E. coli* expression vector encoding (5´ → 3´): N-terminal His_{10} tag Factor Xa protease site (for His-tag cleavage) Multiple cloning site for gene of interest	5711 bp, $Carbenicillin^R$	Novagen
pDB.His.MBP	*E. coli* expression vector encoding (5´ → 3´): N-terminal His_6 tag Maltose-binding protein (MBP) Tobacco Etch Virus (TEV) protease site site (for His_6-MBP cleavage) Multiple cloning site for gene of interest	6494 bp, $Kanamycin^R$	Structural Genomics Centre, Berkeley; (*via* PlasmID Repository)
pET21a	*E. coli* expression vector encoding (5´ → 3´): Multiple cloning site for gene of interest C-terminal His_6 tag	5443 bp, $Carbenicillin^R$	Novagen
pGEX-4T1	*E. coli* expression vector encoding (5´ → 3´): N-terminal glutathione *S*-transferase (GST) fusion Thrombin protease site (for GST-tag cleavage) Multiple cloning site for gene of interest	~4900 bp, $Carbenicillin^R$	GE Healthcare
pET28a(+)	*E. coli* expression vector encoding (5´ → 3´): N-terminal His_6 tag Thrombin protease site (for His-tag cleavage) Multiple cloning site for gene of interest	5369 bp, $Kanamycin^R$	Novagen
pGEX-4T1His	*E. coli* expression vector encoding (5´ → 3´): N-terminal glutathione *S*-transferase (GST) fusion Thrombin protease site (for GST cleavage) Internal His_{10} tag Factor Xa protease site (for His_{10}+GST cleavage) Multiple cloning site for gene of interest	~4930 bp, $Carbenicillin^R$	This study
pET16b:*tunF*	*tunF* from *S. chartreusis* cloned into vector pET16b at *Nde*I-*Xho*I sites. The *tunF* sequence was codon optimised for *E. coli* and synthesised by Genscript.	6704 bp, $Carbenicillin^R$	This study
pET16b:*tunD*	*tunD* from *S. chartreusis* cloned into vector pET16b at *Nde*I-*Xho*I sites. The *tunD* sequence was codon optimised for *E. coli* and synthesised by Genscript.	7139 bp, $Carbenicillin^R$	This study
pDB.His.MBP:*tunD*	*tunD* from pET16b:*tunF* cloned into vector pDB.His.MBP at *Nde*I-*Bam*HI sites.	7922 bp, $Kanamycin^R$	This study
pET21a:*tunD*	*tunD* from pET16b:*tunF* cloned into vector pET21a at *Nde*I-*Xho*I sites.	6871 bp, $Carbenicillin^R$	This study
pGEX-4T1:*tunD*	*tunD* from pET16b:*tunF* cloned into vector pGEX-4T1 at *BamH*I-*Xho*I sites.	~6330 bp, $Carbenicillin^R$	This study
pGEX-4T1His:*tunD*	*tunD* from pET16b:*tunF* cloned into vector pGEX-4T1His at *BamH*I-*Xho*I sites.	~6360 bp, $Carbenicillin^R$	This study
pET28a:*tunE*	*tunE* from *S. chartreusis* cloned into vector pET28a(+) at *Nde*I-*Bam*HI sites. The *tunE* sequence was codon optimised for *E. coli* and synthesised by Genscript.	6083 bp, $Kanamycin^R$	This study

Table 4.7 Details of plasmids used in this chapter

4.6.3 PCR primers

All DNA oligonucleotides were purchased from Sigma-Aldrich.

Name	Sequence	Use
TunD NdeI Forward	5´-AAGGTCGTCATATGGAAATTAT	Forward primer to clone *tunD* into vector pET21a, incorporating an *Nde*I restriction site (underlined)
TunD BamHI Forward	5´-AAAAAAAAGGATCCATGGAAAT	Forward primer to clone *tunD* into vector pGEX-4T1, incorporating an *Bam*HI restriction site (underlined)
TunD His BamHI Forward	5´-AAAAAAAAGGATCCATGGGCCA	Forward primer to clone *tunD* into vector pGEX-4T1, incorporating an *Bam*HI restriction site (underlined) and creating novel vector pGEX-4T1His
TunD No Stop Reverse	5´-GGATCCTCGAGTTCCGCCGGTT	Reverse primer to clone *tunD* into vector pET21a, incorporating a mutation converting stop (TAA) to Glu (GAA)
TunD Stop Reverse	5´-GGATCCTCGAGTTACGCCGGTT	Reverse primer to clone *tunD* into vectors pGEX-4T1 and pGEX-4T1His
pGEX 5´	5´-GGGCTGGCAAGCCACGTTTGGTG	To sequence constructs pGEX-4T1:*tunD* and pGEX-4T1His:*tunD*
pGEX 3´	5´-CCGGGAGCTGCATGTGTCAGAGG	
T7 promoter (F)	5´-TAATACGACTCACTATAGGG	To sequence constructs pDB.His.MBP:*tunD* and pET21a:*tunD*
T7 terminator (R)	5´-GCTAGTTATTGCTCAGCGG	

Table 4.8 Details of PCR primers used in this chapter

4.6.4 General considerations

All solutions and equipment used in the handling of microbial cultures were autoclaved or sterilised. Microbial manipulations were performed in a HERAsafe KS-12 Class II Laminar Flow Cabinet. Media, plastic and glassware were autoclaved at 121 °C for 20 min prior to use. Bacterial cultures were incubated in a New Brunswick Series 25 shaker. Centrifugation was performed in Beckman Coulter Avanti J-25, Beckman GS-6R or Mistral 1000 centrifuges. Microcentrifugation was performed in an Eppendorf Centrifuge 5415R or a Beckman Coulter Microfuge 18. Sonication of cells was performed using a Sonics Vibra-Cell VC505. Proteins were purified using an ÄKTA FPLC system (GE Healthcare) with monitoring by UV absorption at 254 nm or temperature compensated conductivity. Protein samples requiring denaturation were heated in a Techne Dri-Block DB-2A at 90 °C. Water was deionised and passed through a Millipore 0.22 μm filter prior to use.

Unless otherwise stated, the recipes for all media, buffers and solutions can be found in Sambrook *et al*[405] and all chemicals, enzymes and other reagents were purchased from the suppliers listed below. Chemical reagents were purchased from Sigma-Aldrich and used without further purification. Media components were purchased from Sigma-Aldrich or Melford Laboratories Ltd. Restriction endonucleases were purchased from New England Biolabs and remaining enzymes from Invitrogen or Sigma-Aldrich.

4.6.5 Propagation of Escherichia coli strains

E. coli strains were routinely grown at 37 °C in Luria-Bertani broth (LB) or on 1.5% LB agar plates supplemented with appropriate antibiotics, as described by Sambrook *et al*.[405] For recombinant strain selection, antibiotics were used in the following concentrations: carbenicillin (100 µg/mL), kanamycin (50 µg/mL), tetracycline (13 µg/mL), chloramphenicol (34 µg/mL) or streptomycin (50 µg/mL). Recombinant plasmid DNA was transformed into *E. coli* strains using chemically competent cells (Novagen) according to supplier's protocols and including appropriate antibiotic selection. Glycerol stocks were prepared by collecting a cell pellet from a sample of fresh overnight culture (1 mL), resuspending it in 20% (v/v) glycerol (800 µL) and storing at -80 °C. New stocks were prepared from single colonies every 4-6 months.

4.6.6 Standard protocols for the manipulation of DNA

4.6.6.1 Agarose gel electrophoresis of DNA

Agarose gel electrophoresis was performed as described by Sambrook *et al*.[405] Typically 0.8% agarose gels containing 0.01% ethidium bromide (v/v) were used, but concentrations down to 0.7% and up to 1.2% were used for the analysis of large or small fragments respectively. Gels were run submerged in 1× TAE buffer at 10-15 V/cm using a Bio-Rad PowerPac Basic. 6× Orange Loading Dye (Fermentas) was used to load samples and 0.1-12 kbp Perfect DNA Markers (Novagen) were used as DNA size markers. Gels were photographed using UV illumination at 254 nm.

4.6.6.2 Polymerase chain reaction (PCR) amplification of DNA

The reaction mixture for PCR contained *PfuTurbo* Hotstart 2× PCR Mastermix (25 µL, Stratagene), 100 pg of each primer (1 µL of 100 pg/µL stock) and 100 ng template DNA (1 µL of ~100 ng/µL stock) made up to a total volume of 50 µL with sterilised

deionised water. Reactions were performed on an Applied Biosystems GeneAmp PCR System 9700 with cycling parameters as described in Table 2.12. PCR products were analysed by agarose gel electrophoresis (0.7 – 1 % w/v, depending on fragment size).

Segment	No. cycles	Temperature	Duration
1	1	95 ℃	4 min
2	26	94 ℃	45 s
		55 ℃	45 s
		72 ℃	2 min
3	1	72 ℃	7 min

Table 4.9 PCR cycling parameters

4.6.6.3 Restriction endonuclease digestion of DNA

Digestion of DNA with restriction endonuclease enzymes was carried out using conditions recommended by the supplier.

4.6.6.4 Isolation of DNA fragments from agarose gels

DNA restriction fragments contained within agarose gels were visualised using long-wavelength UV light at 330 nm and physically excised with a razor blade. DNA was purified using a QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's protocol.

4.6.6.5 Ligation of DNA

Ligation of DNA was performed with a Quick Ligation Kit (NEB) using the protocol recommended by the supplier. *E. coli* Novablue Gigasingles competent cells (Novagen) were used for transformations together with 2 µL of the ligation mixture. Plasmid DNA was isolated from colonies following antibiotic selection and successful ligation was confirmed by test-scale restriction endonuclease digestion and DNA sequencing with appropriate primers.

4.6.6.6 Isolation of plasmid DNA

Plasmid DNA was isolated from *E. coli* using a QIAprep Miniprep Kit (Qiagen) according to the manufacturer's instructions.

4.6.6.7 Spectrophotometry of nucleic acids

Quality analysis and quantification of DNA was achieved using a Nanodrop 1000 Spectrophotometer according to manufacturer's instructions.

4.6.6.8 Nucleotide sequencing

Isolated samples of plasmid DNA (~100 ng/µL) were sequenced by Source BioScience UK Ltd. based at the University of Oxford Department of Biochemistry and using appropriate primers listed in Table 2.11. Sequence files were viewed using FinchTV v1.4 software (Geospiza).

4.6.7 Standard protocols for the manipulation of proteins

4.6.7.1 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

A NuPAGE Pre-Cast Gel System (Invitrogen) and a Bio-Rad PowerPac Basic was used for all SDS-PAGE, together with 10% or 4-12% NuPAGE Bis-Tris Pre-Cast gels in either 12- or 17-well format. Samples containing 20% 5× SDS loading buffer (v/v) were loaded onto gels along with Perfect Protein Markers 15-150 kDa (Novagen). Electrophoresis was performed at 200 V for 50 min in 1× MES or 1× MOPS buffer (800 mL). Gels were removed from their pre-cast casings and stained with Instant Blue (Gentaur, 25 mL, 15 min).

4.6.7.2 Western immunoblots of protein samples

1×, 0.1× and 0.01× dilutions of each protein samples were seperated by SDS-PAGE, running two identical gels in parallel. One of these gels was stained as usual with Instant Blue stain whereas the other was placed in an iBlot Dry Blotting System (Invitrogen). Proteins were transferred to a transfer membrane following the manufacturer's protocol. Membranes were soaked with blocking buffer at 4 °C (50 mL, 3 h) on a rocking table, then washed with TBST buffer (3 × 50 mL, 15 min). Membranes were then washed with 50 mL blocking buffer containing 10 µL monoclonal anti-polyhistidine-alkaline phosphatase antibody at 4 °C for 16 h, before being washed with TBST buffer (50 mL, 1 h) and stained with dropwise addition of NBT-BCIP solution. Excess stain was washed off with deionised water.

10× TBS buffer

Trizma	24.23 g
NaCl	80.06 g

TBST buffer

10× TBS buffer	100 mL
H_2O	900 mL
Tween-20	1 mL

Blocking buffer

Bovine serum albumin, fraction V	2.5 g
TBST buffer	50 mL

Transfer buffer

Glycine	2.9 g
Tris	5.8 g
SDS	0.37 g
MeOH	200 mL
H_2O	700 mL

Adjust to pH 8.3, make up to 1 L with H_2O

4.6.7.3 Buffer exchange

Protein sample buffers were exchanged using dialysis tubing (12,000-14,000 MWCO, Medicell International Ltd) according to protocols supplied by the manufacturer and those described by Sambrook *et al*.[405] For rapid buffer exchange of small-volume samples PD-10 Desalting Columns (GE Healthcare) were used according to manufacturer's protocols.

4.6.7.4 Concentration of protein samples

Protein samples were concentrated using Vivaspin 6 and Vivaspin 500 Centrifugal Concentrators (Sartorius) according to manufacturer's instructions. Large volume protein samples were concentrated using Series 8000 Stirred Cells (10 or 50 mL cell capacity, Amicon) fitted with Millipore YM10 or YM30 Ultrafiltration Membranes (10,000 or 30,000 MWCO, 25 or 44.5 mm diameter, Amicon).

4.6.7.5 Determination of protein concentration

Protein concentrations were calculated with a bicinchoninic acid (BCA) protein assay (Pierce) according to manufacturer's instructions and using bovine serum albumin as the protein standard.

4.6.7.6 Storage of protein samples

Protein samples were stored at 4 °C at >3 mg/mL. For extended storage glycerol was added to a final concentration of 20% (v/v) and aliquoted samples were stored at -80 °C.

4.6.8 Standard protocols for analytical scale protein expression

4.6.8.1 Analytical scale protein expression with IPTG induction

LB broth (10 mL) and relevant antibiotics were added to a 50 mL falcon tube and inoculated with 20 µL glycerol stock of appropriate *E. coli* cells. Cultures were grown at 37 °C to an A_{600} of between 0.6 and 0.8. After cooling to the final induction temperature, protein expression was induced by the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Cells were left to grow for a specified length of time depending on the induction temperature.

4.6.8.2 Analytical scale protein expression with autoinduction

ZYM-5052 autoinduction media[394] (10 mL) and relevant antibiotics were added to a 50 mL falcon tube and inoculated with 10 µL glycerol stock of appropriate *E. coli* cells. Cultures were grown at 37 °C until turbid. After cooling to the final induction temperature, cells were left to grow overnight or until the A_{600} had plateaued, indicating cells were close to saturation.

4.6.8.3 Analytical scale screening for protein expression

Total Cell Protein Fraction: Samples of cell culture (1 mL) were collected both immediately prior to and at the end of the induction period. Cell pellets were collected by centrifugation (10,000 × *g*, 1 min) and frozen at -20 °C until required. After thawing, pellets were resuspended in 1× phosphate-buffered saline (PBS, 50 µL). 4× SDS sample buffer (50 µL) was added and the solutions were passed through a 27-gauge needle ten times to reduce viscosity. The samples were denatured for 20 min at 90 °C on a heat block and diluted with water (100 µL). The samples were stored at -20 °C until SDS-PAGE analysis whereupon 4 µL was loaded when using a 12-well gel.

Soluble cytoplasmic fraction (BugBuster Method): A sample of cell culture (1.5 mL) was collected at the end of the induction period. Cell pellets were collected by centrifugation (10,000 × *g*, 1 min) and frozen at -20 °C until required. After thawing, pellets were resuspended in BugBuster Master Mix (Novagen, 300 µL). 25× protease inhibitor cocktail stock solution (Roche Complete Protease Inhibitor Cocktail Tablets EDTA-free, 12 µL) was added and the solution incubated on a rotating mixer at rt for 20 min. The insoluble cell debris was collected by centrifugation (16,000 × *g*, 20 min) and

saved for analysis of the insoluble cytosolic fraction. The supernatant (25 µL) was transferred to a fresh tube and 5× SDS sample buffer (25 µL) added. The sample was denatured for 5 min at 90 °C on a heat block and stored at -20 °C until SDS-PAGE analysis, whereupon 18 µL was loaded when using a 12-well gel.

Insoluble cytoplasmic fraction (BugBuster Method): The insoluble pellet from the soluble cytoplasmic fraction was resuspended in BugBuster Master Mix (300 µL). 1:10 diluted BugBuster Master Mix (1 mL) was added and the solution vortexed for 1 min prior to centrifugation (5,000 × *g*, 15 min, 4 °C). The pellet was next resuspended in 1:10 diluted BugBuster Master Mix (750 µL), vortexed for 1 min and centrifuged (5,000 × *g*, 15 min, 4 °C). This step was repeated. The pellet was then resuspended in 1:10 diluted BugBuster Master Mix (750 µL), vortexed for 1 min and centrifuged a final time (16,000 × *g*, 15 min, 4 °C). The resulting pellet was resuspended in 1 % SDS solution (300 µL) and denatured for 10 min at 90 °C on a heat block. The suspension (25 µL) was transferred to a fresh tube and 5× SDS sample buffer (25 µL) added. The sample was denatured for 5 min at 90 °C on a heat block and stored at -20 °C until SDS-PAGE analysis, whereupon 18 µL was loaded when using a 12-well gel.

4.6.9 TunF expression, purification and characterisation

4.6.9.1 Preparative scale expression of tunF

20 µL of *E. coli* BL21(DE3)/pET16b:*tunF* glycerol stock was used to inoculate LB broth (20 mL) containing carbenicillin (100 µg/mL). Following overnight growth at 37 °C this starter culture was used to inoculate 6 × 800 mL of LB medium in 2 L baffled conical flasks. These cultures were incubated at 37 °C and 200 rpm to an A_{600} of 0.6-0.8, cooled to 30 °C for 30 min and then protein expression was induced by the addition of 0.5 mM IPTG. The cultures were left to grow for 6 h (30 °C, 200 rpm) before cell pellets were collected by centrifugation (8,000 × *g*, 12 min, 4 °C) and stored at -80 °C.

4.6.9.2 His_{10}-TunF purification

Method A: All protein purification steps were carried out at 4 °C. The frozen cell pellet from 1 L of induced culture (4.5 g) was thawed and resuspended in 1× Bind Buffer A (20 mL). The cells were incubated with lysozyme (10 mg) at rt for 30 min and disrupted by sonication (4 × 30 s). DNAse I (100 µg) was added and the suspension incubated on ice for 15 min prior to centrifugation (10,000 × *g*, 30 min, 4 °C). The supernatant (20 mL) was

added to pre-equilibrated Ni-NTA HisBind slurry (5 mL) and stirred for 60 min at 4 °C. The mixture was transferred to an empty column and the flow-through collected. The column was washed with 1× Wash Buffer B (4 × 10 mL) and eluted with 1× Ni-NTA Elution Buffer C (6 × 2.5 mL) under gravity. Elution fractions containing ~38 kDa protein – as determined by SDS-PAGE analysis – were pooled and dialysed against dialysis buffer E (2 × 1 L, 12,000-14,000 MWCO).

Bind buffer A	50 mM sodium phosphate pH8.0, 300 mM NaCl, 10 mM imidazole
Wash buffer B	50 mM sodium phosphate pH8.0, 300 mM NaCl, 20 mM imidazole
Elution buffer C	50 mM sodium phosphate pH8.0, 300 mM NaCl, 250 mM imidazole
Dialysis buffer D	30 mM sodium phosphate pH7.8

Method B: All protein purification steps were carried out at 4 °C. The frozen cell pellet from 4.8 L of culture (14.6 g) was suspended in binding buffer E (50 mL) and stirred for 15 min at rt. Lysozyme (15 mg), DNAseI (0.2 mg) and one tablet of Complete EDTA-free protease inhibitor cocktail (Roche) were added and the suspension stirred for 30 min at 4 °C. Cells were further broken by pulsed sonication (5 × 70 s of 0.5 s on/0.2 s off, 60 s cooling cycle, 55% amplitude) and the resulting lysate was cleared by centrifugation (65,000 × *g*, 40 min, 4 °C) and filtered (0.2 µm filter). The protein was purified by Ni^{2+}-affinity chromatography (HisTrap HP 1 mL column, GE Healthcare) using binding buffer E as the mobile phase. Elution was achieved with a linear gradient of imidazole (20 → 500 mM) whilst monitoring UV absorption at 280 nm. Elution fractions containing ~38 kDa protein – as determined by SDS-PAGE analysis – were pooled and dialysed against dialysis buffer F (2 × 1 L, 12,000-14,000 MWCO). Following filtration (0.2 µm filter) the sample was further purified by anion exchange chromatography (MonoQ 5/50 GL, GE Healthcare) using dialysis buffer F as the mobile phase. Elution was achieved with an ionic strength gradient (0 → 1 M NaCl), monitored by UV absorption at 280 nm. Elution fractions containing ~ 38 kDa protein were pooled and dialysed against dialysis buffer F (2 × 1 L, 12,000-14,000 MWCO) prior to concentration (Vivaspin, 10,000 MWCO) and storage at 4 °C. A typical yield of 25 mg His_{10}-TunF was obtained per L of culture.

Binding buffer E	50 mM TEA pH 8.0, 300 mM NaCl, 20 mM imidazole
Dialysis buffer F	50 mM TEA pH 8.0

4.6.9.3 LC/MS analysis of His_{10}-TunF

Liquid chromatography/mass spectrometry (LC/MS) was performed on a Micromass LCT (ESI-TOF-MS) coupled to an Agilent 1100 Series LC System. A Phenomenex Jupiter C-4 column was used (250 mm, 5 mm, 5 µm particle size). A gradient was programmed (1 mL min^{-1}, Step 1 : 45% B, 25 min; Step 2 : 45% D → 100% D, 5 min), with a typical injection volume of 20 µl. The electrospray source of the LCT was operated with a capillary voltage of 3.2 kV and a cone voltage of 25 V. Nitrogen was used as a nebuliser and desolvation gas at a total flow of 600 L/h. Spectra were calibrated using a calibration curve constructed from a minimum of 17 matched peaks from the multiply charged ion series of equine myoglobin obtained at a cone voltage of 25 V. Total mass spectra were obtained from the ion series using the MaxEnt algorithm preinstalled on Masslynx v4.1 software.

Protein	Amino acids	Amino acid sequence[a]	Mw[b] / pI
His_{10}-TunF fusion[c]	347	MGHHHHHHHHHHSSGHIEGR▲HMRVLVTGGAGYVGSFTV RRLLAAGHEVVVVDNLSTGRREAVRGCRLHVVDILDIASM GTVFEEFHPEAVIHFAALKSSEESLRDINTYFSVNLTGTQNV LALCARTGVERFVFSSSCAVYGTPQICPVDETAPVRPESPY GETKYLCERVIASYAQATGMRYANLRYFNAAGAADDASLG EYAGPASTQLLPVAIRSTLGLAPTLRIFGNDYPTTDGTALRD YIHVEDLAQAHVRVLEGLQKGSQCGPVNLGRGQPVSVREL IDVLHDVSGKETPVAMAPRRPGDPGLSWADPSLALSRFG WRAERDLDHIVRTAWNWHTSTPASLE	Mw = 37812 Da pI = 6.39

[a] Vector-encoded fusion sequence highlighted in red. Factor Xa protease recognition sequence is underlined and cleavage site denoted by a chevron; [b] Average molecular weight; [c] Data for native TunF – 327 residues, 35428 Da, pI 5.87;

Table 4.10 Amino acid sequence and selected properties of His_{10}-TunF

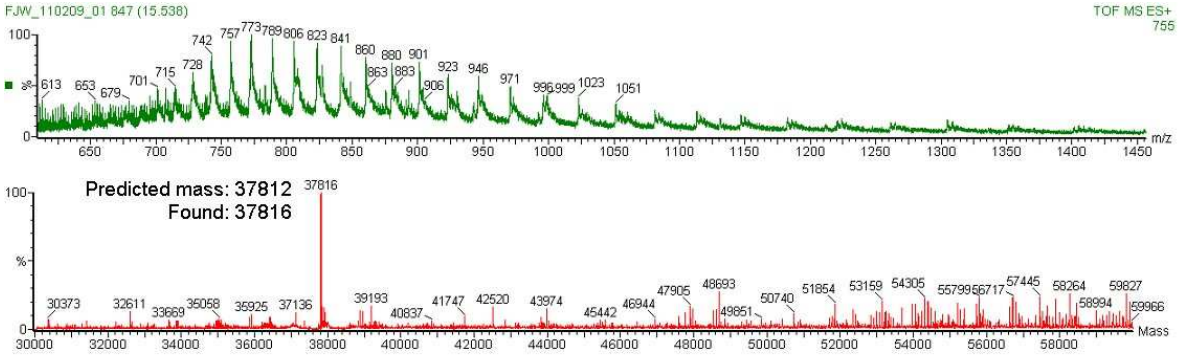


Fig. 4.26 Native and deconvoluted mass spectra of His_{10}-TunF

4.6.9.4 Gel filtration analysis of His_{10}-TunF

All gel filtration steps were performed at 4 °C. Purified His_{10}-TunF protein was loaded onto a HiLoad Superdex 200 16/60 pg column (GE Healthcare) equilibrated with 1.2 CV of buffer G. The chromatographic run was carried out in the same buffer at a flow rate of 0.5 mL/min for 1.5 CV using an ÄKTA FPLC system (GE Healthcare). The column was calibrated by running a set of gel filtration standards (Bio-Rad).

Buffer G: 50 mM HEPES pH7.8, 300 mM NaCl, 10% (v/v) glycerol

4.6.9.5 TunF NMR enzyme assay

A sample of purified His_{10}-TunF was buffer exchanged into sodium phosphate in D_2O (30 mM, pD 7.8) using a PD Minitrap G-25 Sephadex column (GE Healthcare). Stock solutions of freezedried NAD^+ (10 mM) and UDP-GlcNAc (100 mg/mL) were prepared in D_2O. An NMR tube was charged with sodium phosphate in D_2O (30 mM, pD 7.8), UDP-GlcNAc (25 mM), NAD^+ (0.1 mM) and enzyme preparation (25 µg) upto a total volume of 600 µL. Prior to addition of enzyme the tube was preheated to 37 °C, then enzyme was added and the tube span in a manual centrifuge to remove any air bubbles. An NMR timecourse experiment was performed at 37 °C with a Bruker AV700 MHz machine, collecting a ^1H spectrum every four minutes.

4.6.9.6 TunF crystallisation screen

Fresh His_{10}-TunF fusion protein was expressed and purified according to Method B described in section 4.6.9.2. Stock solutions of NAD^+ (10 mM) and UDP-GlcNAc (100 mg/mL) were also prepared. These solutions were used to prepare two different crystallisation samples, described in Table 4.3. Any remaining protein was flash frozen in 100 µL aliquots and stored at -20 °C. Twenty eight crystallisation plates were arrayed using an Art Robbins Instruments Phoenix RE liquid handling robot and monitored with a Rigaku CrystalMation incubator and autoimager. The details of each plate can be seen in Table 4.11.

Identification code	Crystallisation screen (manufacturer)	Temperature (℃)	Protein sample
RG000001	MD Structure Screen I and II (Molecular Dimensions)	20	1
RG000002		4	
RG000003	Emerald Wizard I and II (Emerald Biostructures)	20	
RG000004		4	
RG000005	Easy Xtal JCSG+ (Qiagen)	20	
RG000006		4	
RG000007	PACT premier (MD)	20	
RG000008		4	
RG000009	PEG/Ion HT Screen (Hampton Research)	20	
RG000010		4	
RG000011	Index Screen (Hampton Research)	20	
RG000012		4	
RG000013	SaltRX Screen (Hampton Research)	20	
RG000014		4	
RG000015	MD Structure Screen I and II (Molecular Dimensions)	20	2
RG000016		4	
RG000017	Emerald Wizard I and II (Emerald Biostructures)	20	
RG000018		4	
RG000019	Easy Xtal JCSG+ (Qiagen)	20	
RG000020		4	
RG000021	PACT premier (MD)	20	
RG000022		4	
RG000023	PEG/Ion HT Screen (Hampton Research)	20	
RG000024		4	
RG000025	Index Screen (Hampton Research)	20	
RG000026		4	
RG000027	SaltRX Screen (Hampton Research)	20	
RG000028		4	

Table 4.11 Commercial crystallisation screens used in TunF crystallisation trial

4.6.10 TunD expression, purification and characterisation

4.6.10.1 Construction of expression vectors containing tunD *gene*

The sequences of PCR primers are listed in Table 2.11.

Construction of pET16b:*tunD*

The DNA sequence of *tunD* from *S. chartreusis* NRL 3882 was codon optimised for *E. coli* and synthesised by Genscript USA Inc. The gene was supplied in a pUC57 cloning vector and in pET16b expression vector, cloned between *Nde*I and *Xho*I restriction sites.

Construction of pET21a:*tunD*

Gene *tunD* was amplified by PCR with primers TunD NdeI Forward and TunD No Stop Reverse from plasmid template pET16b:*tunD*, incorporating an *Nde*I restriction endonuclease site and a mutation converting stop codon TAA to Glu codon GAA (underlined in Table 2.11). The amplified DNA product was purified and digested with *NdeI* and *XhoI*. A sample of plasmid pET21a:*otsB* was digested with the same enzymes. The purified insert was then ligated into the linearised pET21a vector and transformed into *E. coli* Novablue Gigasingles competent cells. Plasmid DNA from positive clones was isolated and the formation of pET21a:*tunD* confirmed by sequencing with primers T7 promoter (F) and T7 terminator (R).

Construction of pGEX-4T1:*tunD*

Gene *tunD* was amplified by PCR with primers TunD BamHI Forward and TunD Stop Reverse from plasmid template pET16b:*tunD*, incorporating a *Bam*HI restriction endonuclease site (underlined in Table 2.11). The amplified DNA product was purified and digested with *Bam*HI and *Xho*I. A sample of plasmid pGEX-4T1:*luxP* was digested with the same enzymes. The purified insert was then ligated into the linearised pGEX-4T1 vector and transformed into *E. coli* Novablue Gigasingles competent cells. Plasmid DNA from positive clones was isolated and the formation of pGEX-4T1:*tunD* confirmed by sequencing with primers pGEX 5´ and pGEX 3´.

Construction of pGEX-4T1His:*tunD*

Gene *tunD* together with a vector derived N-terminal His_{10}-tag was amplified by PCR with primers TunD His BamHI Forward and TunD Stop Reverse from plasmid template pET16b:*tunD*, incorporating a *Bam*HI restriction endonuclease site (underlined in Table 2.11). The amplified DNA product was purified and digested with *Bam*HI and *Xho*I.

A sample of plasmid pGEX-4T1:*luxP* was digested with the same enzymes. The purified insert was then ligated into the linearised pGEX-4T1 vector and transformed into *E. coli* Novablue Gigasingles competent cells. Plasmid DNA from positive clones was isolated and the formation of pGEX-4T1His:*tunD* confirmed by sequencing with primers pGEX 5´ and pGEX 3´.

4.6.10.2 Preparative scale expression and purification of His_{10}-TunD

Expression: 20 µL of *E. coli* BL21(DE3)/pET16b:*tunD* glycerol stock was used to inoculate LB broth (20 mL) containing carbenicillin (100 µg/mL). Following overnight growth at 37 °C this starter culture was used to inoculate 6 × 800 mL of LB medium in 2 L baffled conical flasks. These cultures were incubated at 37 °C and 200 rpm to an A_{600} of 0.6-0.8, cooled to 12 °C over 60 min and then protein expression was induced by the addition of 0.5 mM IPTG. The cultures were left to grow for 24 h (12 °C, 200 rpm) before cell pellets were collected by centrifugation (8,000 × *g*, 12 min, 4 °C) and stored at -20 °C.

Purification: All protein purification steps were carried out at 4 °C. The frozen cell pellet from ~2.4 L induced culture (10.3 g) was suspended in binding buffer H (25 mL) and stirred for 15 min at rt. Lysozyme (10 mg), DNAseI (0.2 mg) and ½ a tablet of Complete EDTA-free protease inhibitor cocktail (Roche) were added and the suspension stirred for 40 min at 4 °C. Cells were further broken by sonication (6 × 60 s, 60 s cooling cycle, 55% amplitude) and the resulting lysate was cleared by centrifugation (65,000 × g, 40 min, 4 °C) and filtered (0.2 µm filter). The protein was purified by Ni^{2+}-affinity chromatography (HisTrap HP 1 mL column, GE Healthcare) using binding buffer H as the mobile phase. Elution was achieved with a linear gradient of imidazole (10 → 500 mM) whilst monitoring UV absorption at 280 nm. Elution fractions containing ~53 kDa protein – as determined by SDS-PAGE analysis – were pooled and dialysed against dialysis buffer I (2 × 1 L, 12,000-14,000 MWCO), prior to concentration (Vivaspin, 10,000 MWCO) and storage at 4 °C. A typical yield of <1.5 mg His_6-MBP-TunD was obtained per L of culture.

Binding buffer H	50 mM HEPES pH 7.5, 100 mM NaCl, 10% glycerol, 0.5% Triton X-100, 10 mM imidazole
Dialysis buffer I	50 mM HEPES pH 7.5, 100 mM NaCl, 10% glycerol, 0.5% Triton X-100

Protein	Amino acids	Amino acid sequence[a]	Mw[b] / pI
His_{10}-TunD fusion[c]	493	MGHHHHHHHHHHSSGHIEGR▲HMEIIFTVSDQVWGGKHR YMHDMALGLAQAGHTVTVLAEEGGAMLQQCRAAGVTTVS VPAFASDDAAEAVRKALRHRRPHIVCVSGRADAAAVHHAQ SQGVTDAAVCLFRHSAFPLGTTDEVRDLFTGVNLVFTTSLE QRQRQFEPLINAGVLKDEQVEILTSGVGEPLLAALDAADRG AARKELRAESDQFVFLVLARLAWEKGIDQVIDAFADLELPP DAAPPLLVVAGEGPLEAELRGQTIERGVAERVQFLGHQDH VAPVIKASDAVVLTSTVPETGPLALKEAMAAGRPVIASVQG GIPEFVEDERHGLLVIDDEDLRQAMQRLLSDREAAETMGA AGSESVRGGHRAVRRVEYLAHRLDLLALEQLAPDTVLHEV VWDDVRLREETQGGFVFVPRTSHIMELDSATYAVVRTAVE AGDPLQLIKLPEETLGVIAHRLYAMGALVRQDGQATPAART ADGERPVTEPA	Mw = 53352 Da pI = 5.47

[a] Vector-encoded fusion sequence highlighted in red. Factor Xa protease recognition sequence is underlined and cleavage site denoted by a chevron; [b] Average molecular weight; [c] Data for native TunD – 472 residues, 50832 Da, pI 5.01;

Table 4.12 Amino acid sequence and selected properties of His_{10}-TunD

4.6.10.3 Preparative scale expression and purification of His_6-MBP-TunD

Expression: 20 µL of *E. coli* Tuner(DE3)/pDB.His.MBP:*tunD* glycerol stock was used to inoculate LB broth (20 mL) containing kanamycin (50 µg/mL). Following overnight growth at 37 °C this starter culture was used to inoculate 6 × 800 mL of LB medium in 2 L baffled conical flasks. These cultures were incubated at 37 °C and 200 rpm to an A_{600} of 0.6-0.8, cooled to 12 °C over 60 min and then protein expression was induced by the addition of 10 µM IPTG. The cultures were left to grow for 22 h (12 °C, 200 rpm) before cell pellets were collected by centrifugation (8,000 × *g*, 12 min, 4 °C) and stored at -20 °C.

Purification: All protein purification steps were carried out at 4 °C. The frozen cell pellet from ~1.6 L induced culture (5.6 g) was suspended in binding buffer J (20 mL) and stirred for 15 min at rt. Lysozyme (5 mg), DNAseI (0.1 mg) and ½ a tablet of Complete EDTA-free protease inhibitor cocktail (Roche) were added and the suspension stirred for 40 min at 4 °C. Cells were further broken by pulsed sonication (5 × 70 s of 0.5 s on/0.2 s off, 60 s cooling cycle, 55% amplitude) and the resulting lysate was cleared by centrifugation (65,000 × g, 40 min, 4 °C) and filtered (0.2 µm filter). The protein was purified by Ni^{2+}-affinity chromatography (HisTrap HP 1 mL column, GE Healthcare) using binding buffer J as the mobile phase. Elution was achieved with a linear gradient of imidazole (10 → 500 mM) whilst monitoring UV absorption at 280 nm. Elution fractions containing ~93 kDa

protein – as determined by SDS-PAGE analysis – were pooled and dialysed against dialysis buffer K (2 × 1 L, 12,000-14,000 MWCO), prior to concentration (Vivaspin, 30,000 MWCO) and storage at 4 °C. A typical yield of 4 mg His_6-MBP-TunD was obtained per L of culture.

Binding buffer J	50 mM HEPES pH 7.5, 200 mM NaCl, 10% glycerol, 20 mM $MgCl_2$, 10 mM imidazole
Dialysis buffer K	50 mM HEPES pH 7.5, 200 mM NaCl, 10% glycerol, 20 mM $MgCl_2$

Protein	Amino acids	Amino acid sequence[a]	Mw[b] / pI
His_6-MBP-TunD fusion[c]	867	MGSSHHHHHHGTKTEEGKLVIWINGDKGYNGLAEVGKKFE KDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGG YAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVE ALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQE PYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLT FLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWS NIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKE LAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDP RIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQT VDEALKDAQTGTDYDIPTTENLYFQ▲GHMEIIFTVSDQVWG GKHRYMHDMALGLAQAGHTVTVLAEEGGAMLQQCRAAG VTTVSVPAFASDDAAEAVRKALRHRRPHIVCVSGRADAAA VHHAQSQGVTDAAVCLFRHSAFPLGTTDEVRDLFTGVNLV FTTSLEQRQRQFEPLINAGVLKDEQVEILTSGVGEPLLAALD AADRGAARKELRAESDQFVFLVLARLAWEKGIDQVIDAFAD LELPPDAAPPLLVVAGEGPLEAELRGQTIERGVAERVQFLG HQDHVAPVIKASDAVVLTSTVPETGPLALKEAMAAGRPVIA SVQGGIPEFVEDERHGLLVIDDEDLRQAMQRLLSDREAAE TMGAAGSESVRGGHRAVRRVEYLAHRLDLLALEQLAPDTV LHEVVWDDVRLREETQGGFVFVPRTSHIMELDSATYAVVR TAVEAGDPLQLIKLPEETLGVIAHRLYAMGALVRQDGQATP AARTADGERPVTEPA	Mw = 94307 Da pI = 5.12

[a] Vector-encoded His_6-maltose-binding protein fusion sequence highlighted in red. Tobacco etch virus (TEV) protease recognition sequence is underlined and cleavage site denoted by a chevron; [b] Average molecular weight; [c] Data for native TunD – 472 residues, 50832 Da, pI 5.01;

Table 4.13 Amino acid sequence and selected properties of His_6-MBP-TunD

4.6.10.4 Attempted TEV protease cleavage of His_6-MBP-TunD

Samples of His_6-MBP-TunD were subjected to protease cleavage by AcTEV Protease (Invitrogen) following protocols supplied by the manufacturer. Incubation temperatures of 16, 23 and 30 °C were screened. Extended incubation times of up to 48 h were tested as well as a 5-fold increase in protease concentration over that suggested by the manufacturer. Levels of protease cleavage were monitored by SDS-PAGE.

4.6.10.5 Preparative scale expression and purification of TunD-His$_6$

Expression: 20 µL of *E. coli* BL21(DE3)/pET21a:*tunD* glycerol stock was used to inoculate LB broth (20 mL) containing carbenicillin (100 µg/mL). Following overnight growth at 37 °C this starter culture was used to inoculate 6 × 800 mL of LB medium in 2 L baffled conical flasks. These cultures were incubated at 37 °C and 200 rpm to an A_{600} of 0.6-0.8, cooled to 15 °C for 30 min and then protein expression was induced by the addition of 0.5 mM IPTG. The cultures were left to grow for 24 h (15 °C, 200 rpm) before cell pellets were collected by centrifugation (8,000 × *g*, 12 min, 4 °C) and stored at -20 °C.

Purification: All protein purification steps were carried out at 4 °C. The frozen cell pellet from 0.8 L induced culture (4.0 g) was suspended in binding buffer L (25 mL) and stirred for 15 min at rt. Lysozyme (5 mg), DNAseI (0.1 mg) and ½ a tablet of Complete EDTA-free protease inhibitor cocktail (Roche) were added and the suspension stirred for 30 min at 4 °C. Cells were further broken by pulsed sonication (5 × 70 s of 0.5 s on/0.2 s off, 60 s cooling cycle, 55% amplitude) and the resulting lysate was cleared by centrifugation (65,000 × g, 40 min, 4 °C) and filtered (0.2 µm filter). The protein was purified by Ni^{2+}-affinity chromatography (HisTrap HP 1 mL column, GE Healthcare) using binding buffer L as the mobile phase. Elution was achieved with a linear gradient of imidazole (20 → 500 mM) whilst monitoring UV absorption at 280 nm. Elution fractions containing ~53 kDa protein – as determined by SDS-PAGE analysis – were pooled and dialysed against dialysis buffer M (2 × 1 L, 12,000-14,000 MWCO), prior to concentration (Vivaspin, 10,000 MWCO) and storage at 4 °C. A typical yield of 2.2 mg TunD-His_6 was obtained per L of culture.

Binding buffer L	50 mM HEPES pH 7.4, 300 mM NaCl, 10% glycerol, 20 mM imidazole
Dialysis buffer M	50 mM HEPES pH 7.4, 300 mM NaCl, 10% glycerol, 10 mM $MgCl_2$

Protein	Amino acids	Amino acid sequence[a]	Mw[b] / pI
TunD-His_6 fusion[c]	481	MEIIFTVSDQVWGGKHRYMHDMALGLAQAGHTVTVLAEEG GAMLQQCRAAGVTTVSVPAFASDDAAEAVRKALRHRRPHI VCVSGRADAAAVHHAQSQGVTDAAVCLFRHSAFPLGTTD EVRDLFTGVNLVFTTSLEQRQRQFEPLINAGVLKDEQVEILT SGVGEPLLAALDAADRGAARKELRAESDQFVFLVLARLAW EKGIDQVIDAFADLELPPDAAPPLLVVAGEGPLEAELRGQTI ERGVAERVQFLGHQDHVAPVIKASDAVVLTSTVPETGPLAL KEAMAAGRPVIASVQGGIPEFVEDERHGLLVIDDEDLRQAM QRLLSDREAAETMGAAGSESVRGGHRAVRRVEYLAHRLD LLALEQLAPDTVLHEVVWDDVRLREETQGGFVFVPRTSHI MELDSATYAVVRTAVEAGDPLQLIKLPEETLGVIAHRLYAM GALVRQDGQATPAARTADGERPVTEPAELEHHHHHH	Mw = 52026 Da pI = 5.16

[a] Vector-encoded C-terminal His_6 fusion sequence highlighted in red; [b] Average molecular weight; [c] Data for native TunD – 472 residues, 50832 Da, pI 5.01;

Table 4.14 Amino acid sequence and selected properties of TunD-His_6

4.6.10.6 TunD enzyme assay: Spectrophotometric coupled assay

Assays for TunD fusions were performed according to methods described by Gosselin *et al*.[397] Reactions were set up in 96-well microplates containing the following components in a total volume of 200 µL: 50 mM MOPS pH 6.5 - 8.0, 200 mM NaCl, 10 mM $MgCl_2$/$MnCl_2$, 0.01 % w/v BSA, 1 mM phospho(enol)pyruvic acid, 300 µM NADH, 17 units pyruvate kinase, 24 units lactate dehydrogenase, 1 mM NDP-sugar donor, 3 mM glycosyl acceptor and various concentrations of enzyme preparation. Plates were incubated at 30 °C in a SpectraMax Plus spectrophotometer (Molecular Devices), with absorbance readings measured every 10 s at 340 nm for 2 h.

4.6.10.7 TunD enzyme assay: HPLC assay

A typical reaction was incubated at 30 ºC and contained the following components in a total volume of 50 µL: 10 mM HEPES pH 7.0, 100 mM NaCl, 5 mM $MgCl_2$, 5 mM NDP-sugar donor, 5 mM glycosyl acceptor, 0.1 % BSA and various concentrations of enzyme preparation. The reaction was quenched by heating to 85 °C for 3 min followed by centrifugation to remove precipitated protein (16,000 × *g*, 3 min). Control reactions contained heat-denatured enzyme or water in the place of the TunD preparation.

HPLC Method A: The HPLC system consisted of a Waters 2795 separations module coupled to a Waters 2996 photodiode array detector, controlled by Empower software. Reversed-phase chromatography was achieved on a Phenomenex Luna NH_2 column

(250 mm × 4.6 mm × 5 µm). Solvents used in the mobile phase were as follows: solvent A - H_2O; solvent B – MeCN. An isocratic gradient of 70 % B was used, with a typical injection volume of 20 µl and detection at 260 nm.

HPLC Method B: The HPLC system consisted of a Dionex Ultimate 3000 pump and Varian PL-ELS 2100 Ice detector, controlled by Chromeleon v6.8 software. Reversed-phase chromatography was achieved on a Phenomenex Luna NH_2 column (250 mm × 4.6 mm × 5 µm). Solvents used in the mobile phase were as follows: solvent A - H_2O; solvent B – MeCN. An isocratic gradient of 70 % B was used, with a typical injection volume of 20 µl and ELS detection (evaporation = 80 ºC, nebulisation = 50 ºC, N_2 flow = 1.6 SLM).

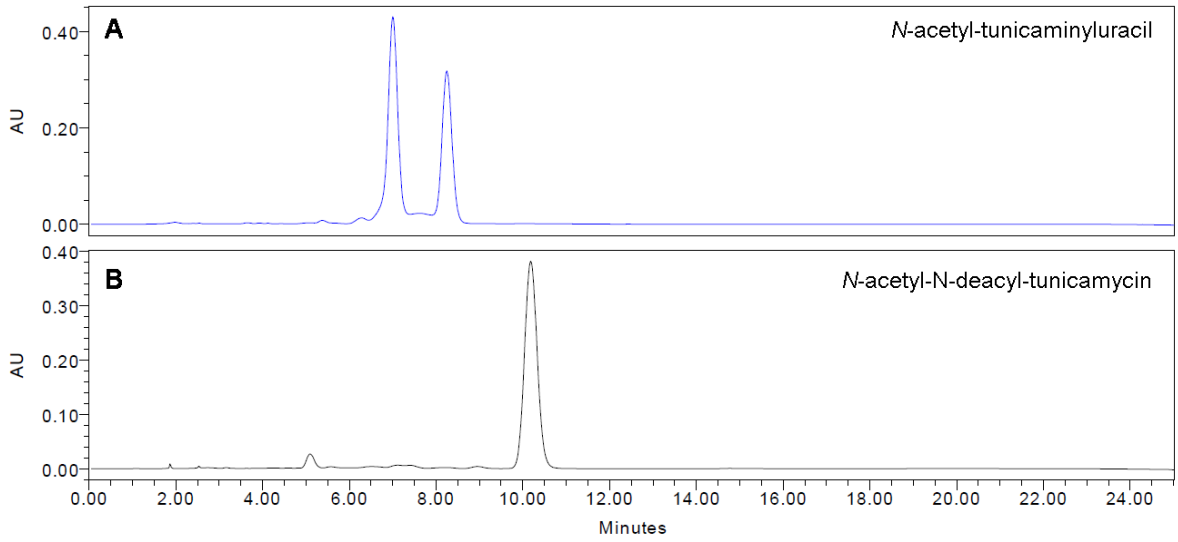


Fig. 4.27 Select UV chromatograms from TunD HPLC assays; **A**: Proposed natural substrate for TunD; **B**: Proposed product following TunD-catalysed reaction of *N*-acetyl-tunicaminyl-uracil with UDP-GlcNAc

4.6.11 TunE expression, purification and characterisation

4.6.11.1 Preparative scale expression and purification of His_6-TunD

Expression: 20 µL of *E. coli* BL21(DE3)/pET28a:*tunE* glycerol stock was used to inoculate LB broth (20 mL) containing kanamycin (50 µg/mL). Following overnight growth at 37 °C this starter culture was used to inoculate 6 × 800 mL of LB medium in 2 L baffled conical flasks. These cultures were incubated at 37 °C and 200 rpm to an A_{600} of 0.6-0.8, cooled to 12 °C for 60 min and then protein expression was induced by the addition of 0.5 mM IPTG. The cultures were left to grow for 24 h (12 °C, 200 rpm) before cell pellets were collected by centrifugation (8,000 × *g*, 12 min, 4 °C) and stored at -20 °C.

Purification: All protein purification steps were carried out at 4 °C. The frozen cell pellet from 1.6 L induced culture (6.1 g) was suspended in binding buffer N (25 mL) and stirred for 15 min at rt. Lysozyme (6 mg), DNAseI (0.1 mg) and ½ a tablet of Complete EDTA-free protease inhibitor cocktail (Roche) were added and the suspension stirred for 30 min at 4 °C. Cells were further broken by sonication (5 × 30 s, 15 s cooling cycle, 55% amplitude) and the resulting lysate was cleared by centrifugation (65,000 × g, 40 min, 4 °C) and filtered (0.2 µm filter). The protein was purified by Ni^{2+}-affinity chromatography (HisTrap HP 1 mL column, GE Healthcare) using binding buffer N as the mobile phase. Elution was achieved with a linear gradient of imidazole (10 → 500 mM) whilst monitoring UV absorption at 254 nm. Elution fractions containing ~26 kDa protein – as determined by SDS-PAGE analysis – were pooled and dialysed against dialysis buffer P (2 × 1 L, 12,000-14,000 MWCO), prior to concentration (Vivaspin, 10,000 MWCO) and storage at 4 °C. A typical yield of <1.5 mg His_6-TunD was obtained per L of culture.

Binding buffer N	50 mM HEPES pH 7.5, 300 mM NaCl, 10% glycerol, 1 mM TCEP 0.5% Triton X-100, 10 mM imidazole
Dialysis buffer P	50 mM HEPES pH 7.5, 100 mM NaCl, 10% glycerol, 1 mM TCEP 0.05% Triton X-100, 10 mM imidazole

Protein	Amino acids	Amino acid sequence[a]	Mw[b] / pI
His_6-TunE[c]	254	MGSSHHHHHHSSGLVPR▲GSHMKVLVIAAHPDDEVLGAGAT VSALSQQGAVVQIHILAEGISLRHSGVKIEEARERCQVAAKDL GAAEVSFGGLAADGRLLADLPQRQVVDAVSHALREAEPEVVF THHPGDIHADHRSVAHAVGYGTRILGSGSVRQVLHFEVLSST EQQTGLVAPFTPNLFYDVTGHVEAKCRALAAYPYELYDPPHP RSLAAVRTLASYRGTQVGVEAAEAFMIGRELRGPMAAVPSGV DVR	Mw = 26994 Da pI = 6.33

[a] Vector-encoded His_6 fusion sequence highlighted in red. Thrombin recognition sequence is underlined and cleavage site denoted by a chevron. The initiating valine in native TunE has been replaced with a methionine in His_6-TunE – highlighted in blue; [b] Average molecular weight; [c] Data for native TunE – 234 residues, 24799 Da, pI 5.93;

Table 4.15 Amino acid sequence and selected properties of His_6-TunE

4.6.11.2 TunE enzyme assay: HPLC assay

A typical reaction was incubated at 30 ºC and contained the following components in a total volume of 100 µL: 50 mM HEPES pH 7.5, 50 mM NaCl, 10% glycerol, 1 mM TCEP and various concentrations of enzyme preparation. The reaction was quenched by heating to 85 °C for 3 min followed by centrifugation to remove precipitated protein (16,000 × *g*, 3 min). Control reactions contained heat-denatured enzyme or water in the place of the TunE preparation. Reaction mixtures were analysed using HPLC Method B as used with TunD enzymatic assays (see section 4.6.10.7).

4.7 References

384. Rossmann MG, Moras D, Olsen KW (1974): **Chemical and Biological Evolution of a Nucleotide-Binding Protein**. *Nature* **250**: 194-199.

385. Filling C, Berndt KD, Benach J, Knapp S, Prozorovski T, Nordling E, Ladenstein R, Jornvall H, Oppermann U (2002): **Critical residues for structure and catalysis in short-chain dehydrogenases/reductases**. *J. Biol. Chem.* **277**: 25677-25684.

386. Hochuli E, Bannwarth W, Dobeli H, Gentz R, Stuber D (1988): **Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent**. *Bio/Technology* **6**: 1321-1325.

387. Scopes R (1994): *Protein Purification: Principles and Practice.* Springer-Verlag: New York.

388. Guo HJ, Li L, Wang PG (2006): **Biochemical characterization of UDP-GlcNAc/Glc 4-epimerase from *Escherichia coli* O86:B7**. *Biochemistry* **45**: 13760-13768.

389. Bernatchez S, Szymanski CM, Ishiyama N, Li JJ, Jarrell HC, Lau PC, Berghuis AM, Young NM, Wakarchuk WW (2005): **Single bifunctional UDP-GlcNAc/Glc 4-epimerase supports the synthesis of three cell surface glycoconjugates in *Campylobacter jejuni***. *J. Biol. Chem.* **280**: 4792-4802.

390. Creuzenet C, Belanger M, Wakarchuk WW, Lam JS (2000): **Expression, purification, and biochemical characterization of WbpP, a new UDP-GlcNAc C4 epimerase from *Pseudomonas aeruginosa* serotype O6**. *J. Biol. Chem.* **275**: 19060-19067.

391. *Gel Filtration: Principles and Methods* (2006). GE Healthcare Bio-Sciences AB: Uppsala.

392. *Novagen pET System Manual* (2006). EMD Biosciences Inc.: Darmstadt.

393. Schein CH (1989): **Production of Soluble Recombinant Proteins in Bacteria**. *Bio-Technology* **7**: 1141-1147.

394. Studier FW (2005): **Protein production by auto-induction in high-density shaking cultures**. *Protein Expr. Purif.* **41**: 207-234.

395. Kapust RB, Waugh DS (1999): ***Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused**. *Protein Sci.* **8**: 1668-1674.

396. Fox JD, Routzahn KM, Bucher MH, Waugh DS (2003): **Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers**. *FEBS Lett.* **537**: 53-57.

397. Gosselin S, Alhussaini M, Streiff MB, Takabayashi K, Palcic MM (1994): **A Continuous Spectrophotometric Assay for Glycosyltransferases**. *Anal. Biochem.* **220**: 92-97.

398. Lairson LL, Henrissat B, Davies GJ, Withers SG (2008): **Glycosyltransferases: Structures, functions, and mechanisms**. *Annu. Rev. Biochem.* **77**: 521-555.

399. Nakamura N, Inoue N, Watanabe R, Takahaski M, Takeda J, Stevens VL, Kinoshita T (1997): **Expression cloning of PIG-L, a candidate *N*-acetylglucosaminylphosphatidylinositol deacetylase**. *J. Biol. Chem.* **272**: 15834-15840.

400. Newton GL, Av-Gay Y, Fahey RC (2000): ***N*-acetyl-1-D-*myo*-inosityl-2-amino-2-deoxy-alpha-D-glucopyranoside deacetylase (MshB) is a key enzyme in mycothiol biosynthesis**. *J. Bacteriol.* **182**: 6958-6963.

401. Urbaniak MD, Crossman A, Chang TH, Smith TK, van Aalten DMF, Ferguson MAJ (2005): **The *N*-acetyl-D-glucosaminylphosphatidylinositol De-*N*-acetylase of glycosylphosphatidylinositol biosynthesis is a zinc metalloenzyme**. *J. Biol. Chem.* **280**: 22831-22838.

402. Tsvetanova BC, Kiemle DJ, Price NPJ (2002): **Biosynthesis of tunicamycin and metabolic origin of the 11-carbon dialdose sugar, tunicamine**. *J. Biol. Chem.* **277**: 35289-35296.

403. Yabe T (2010): *Functional characterizations of the tun genes in the biosynthetic cluster of tunicamycin.* Part II Thesis: University of Oxford.

404. Newton GL, Ko M, Ta PL, Av-Gay Y, Fahey RC (2006): **Purification and characterization of *Mycobacterium tuberculosis* 1D-*myo*-inosityl-2-acetamido-2-deoxy-α-D-glucopyranoside deacetylase, MshB, a mycothiol biosynthetic enzyme**. *Protein Expr. Purif.* **47**: 542-550.

405. Sambrook J, Russell D (2000): *Molecular Cloning: A Laboratory Manual* 3rd edn. Cold Spring Harbor Laboratory Press: New York.

Chapter 5: Towards synthesis of tunicamycin analogue fragments

5.1 Introduction

Rationally designed analogues of tunicamycin have the potential to inhibit a diverse set of clinically important glycosyltranferases. As discussed in section 1.5, tailoring key portions of the molecule will yield a library of transition state mimics containing components from both the acceptor and the sugar donor substrates, together targeting a wide range of nucleotide diphospho-sugar dependent glycosyltransferases. Moreover, since these compounds are based closely around the unique molecular scaffold of the parent tunicamycin structure, they are predicted to retain the high potency and specificity observed with tunicamycin towards its natural targets MraY and GPT.[406,407]

Ideally, the majority of such analogues would be accessed from simple building blocks, utilising the natural biosynthetic machinery to assemble the tunicamycin analogues. In previous chapters, attention has focused on deciphering the biosynthetic pathway of tunicamycin in detail and functionally characterising individual enzymes. It is likely that substrate specificities will vary, so where one enzyme may tolerate broad functional group modifications to its native substrate, the next enzyme in the pathway may not. A synthetic supply of analogue fragments would help overcome these pitfalls by providing non-natural substrates for enzymes whose substrates are not readily available or reliant on successful transformations by preceding enzymes. This way, the enzymology and substrate specificity of the full range of biosynthetic enzymes can be studied in detail irrespective of their position in the pathway.

Adopting such a purely chemical strategy will provide an alternative pathway to key non-natural intermediates which will complement the chemoenzymatic route by overcoming inherent limitations of using the natural enzymes alone, thereby broadening the chemical space available for a tunicamycin analogue library.

The fragments targeted in this study are based upon the tunicaminyl-uracil core of the natural product, consisting of an eleven carbon dialdose linked to uracil. This skeleton is derived from the tail-to-tail coupling of *N*-acetyl-galactosamine and uridine moieties to form the necessary C–C linkage (see sections 1.4.3 and 2.4.2). Through replicating this unusual coupling synthetically, analogues of the tunicaminyl-uracil core could be accessed

incorporating any of the naturally occurring nucleosides (Fig. 5.1). Additionally, protection of the amine group in galactosamine could be altered depending on the downstream application. Incorporation of an *N*-acetyl group would allow for enzymatic *N*-deacetylation at a late stage – with biosynthetic enzyme TunE – followed by alkylation of the resulting free amine. More labile protecting groups could be incorporated if chemical deprotection was envisaged. Further synthetic/chemoenzymatic alkylation and glycosylation of this core skeleton would then lead to a diverse set of tunicamycin analogues.

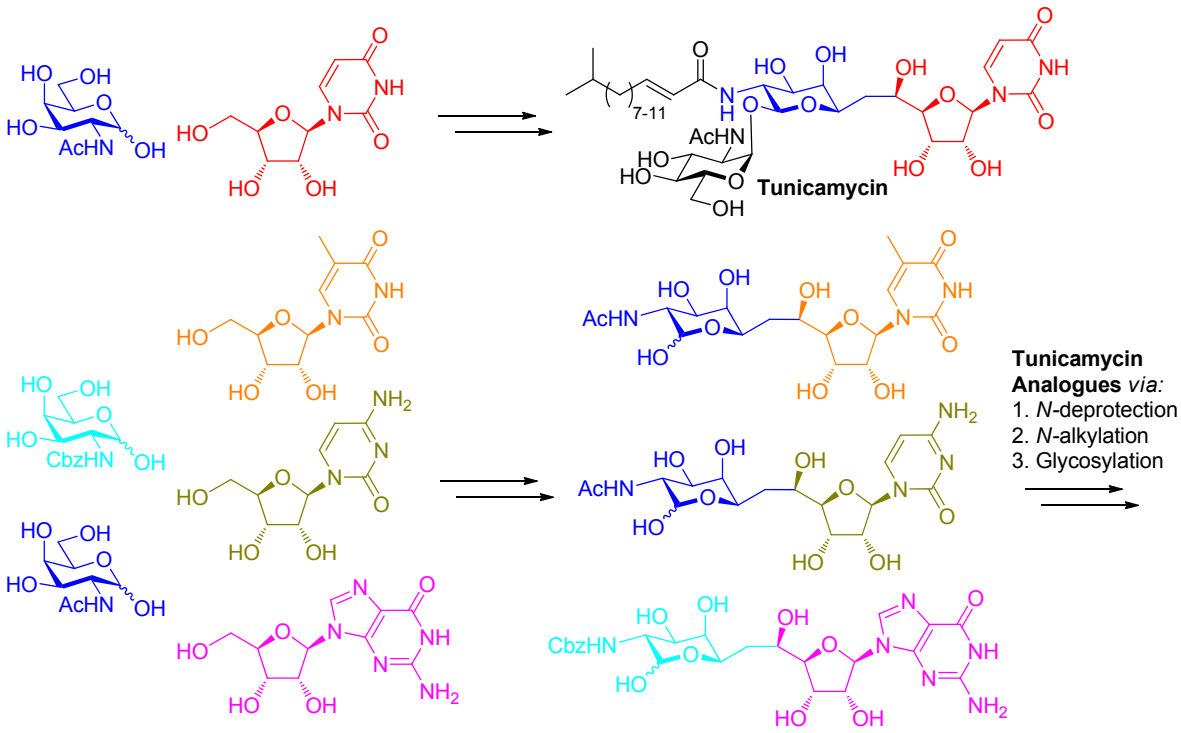


Fig. 5.1 Targetting synthetic analogues of tunicaminyl-uracil

To this end, work presented in this chapter focuses on studies towards a new synthetic route to D-galactosamine building blocks, a crucial component of the tunicamine scaffold. The chief driving force of these investigations is to address the limitations of current literature procedures – low overall yields, regiospecificities and stereoselectivities – through the adoption of a completely novel synthetic strategy.

5.2 Targeting an improved synthesis of D-galactosamine

5.2.1 Introduction

Aside from forming the central core of the tunicamycin skeleton, derivatives of D-galactosamine play an important role in an array of biological processes. Its parent family, the 2-amino-2-deoxysugars, represent a class of building blocks present in a great number of biologically important glycoconjugates. These include glycoproteins, peptidoglycans, glycolipids and glycosaminoglycans, as well as blood group oligosaccharides.[3-8] Functional analysis of a whole host of glycoconjugates is facilitated by their chemical synthesis, overcoming their difficult isolation in homogenous form and low natural abundance.

The most abundant amino-sugar observed in these structures is *N*-acetyl-D-glucosamine (D-GlcNAc), which is commercially available and inexpensive, allowing it to be converted to synthetic building blocks through simple protecting group manipulations.[408,409] *N*-Acetyl-D-galactosamine (D-GalNAc) is the next most abundant, and was originally obtained from hydrolysis of chondroitin sulfate from shark fin cartilage.[11-14] However, it remains in limited supply and is very expensive,[410] necessitating more elaborate routes for its incorporation into synthetic glycoconjugates.

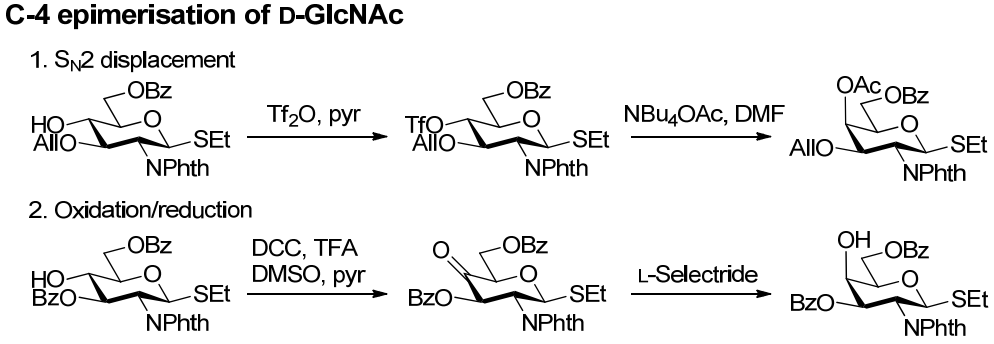


Fig. 5.2 Overview of existing syntheses of D-galactosamine derivatives – part I

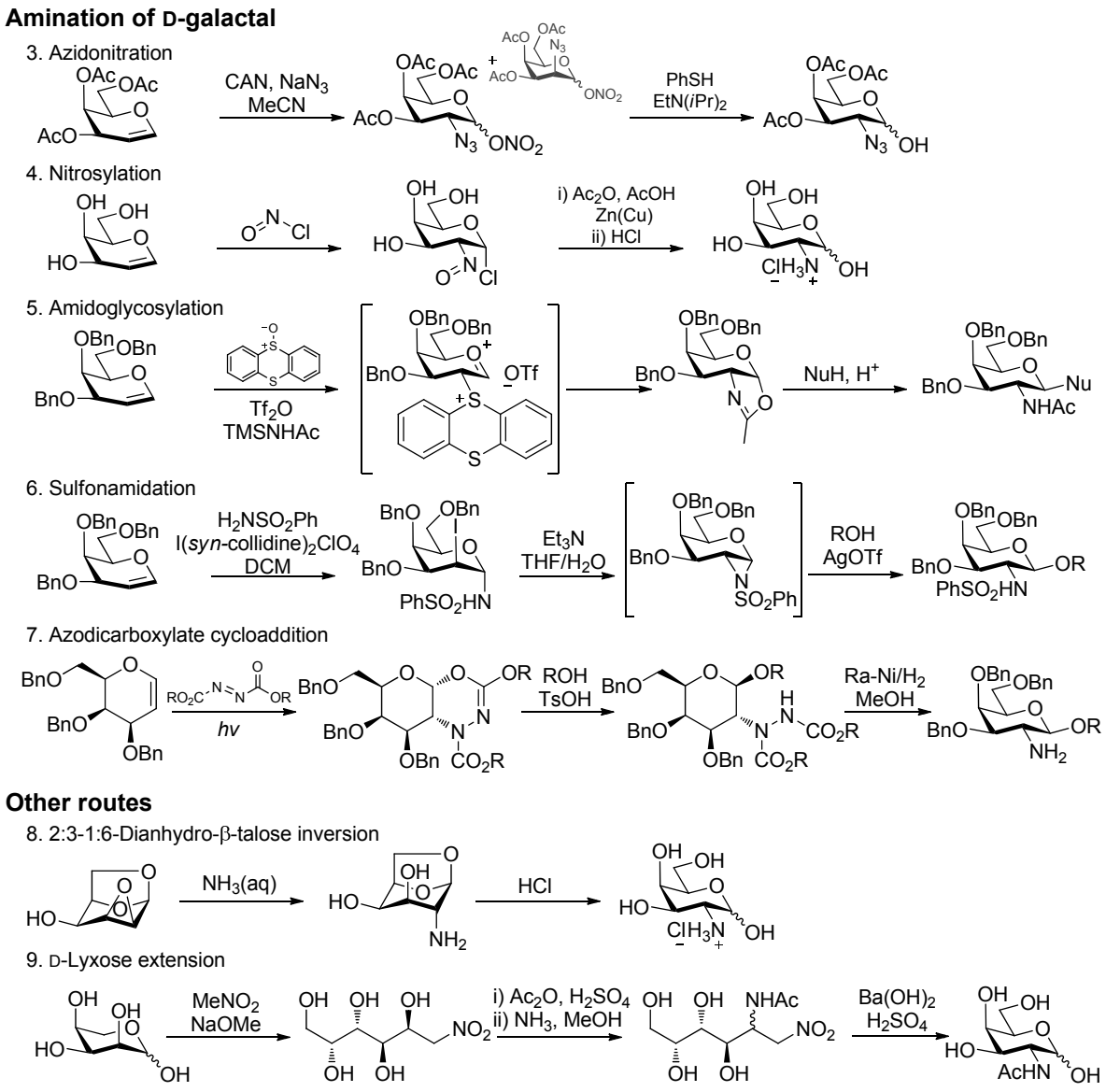


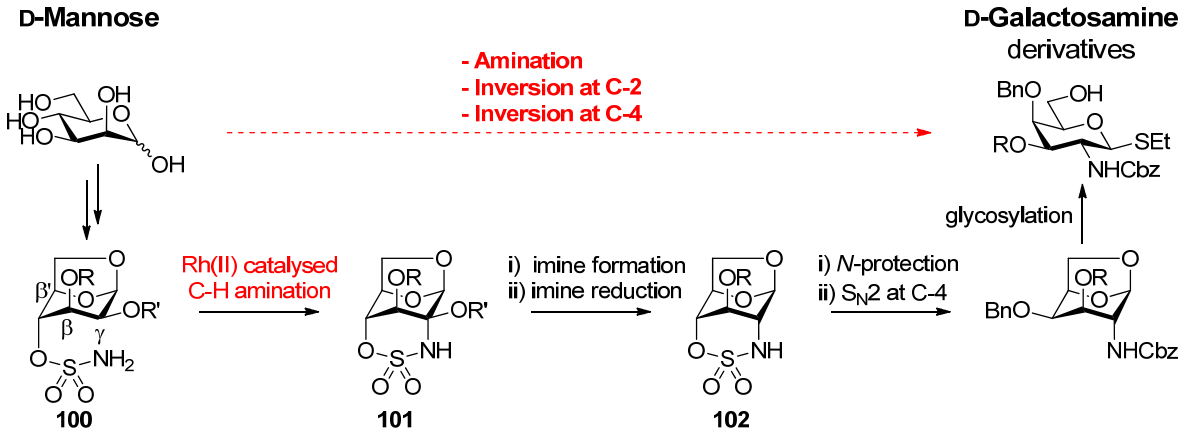
Fig. 5.3 Overview of existing syntheses of D-galactosamine derivatives – part II

Existing synthetic approaches to D-GalNAc have been summarised in Fig. 5.2 and Fig. 5.3. The most common strategy involves epimerising the C-4 position of D-GlcNAc to yield the requisite *galacto*- configuration. First performed by Gross *et al.* in 1965,[411] selective activation of 4-OH with a mesylate or triflate is followed by S_N2 inversion by an *O*-nucleophile such as acetate or benzoate.[411-418] In another report, inversion was achieved through an oxidation-stereoselective reduction sequence.[418] Unfortunately, these methods require numerous time- and cost-intensive protecting group manipulations and typically harsh reaction conditions for the inversion, providing only modest overall yields. A popular alternative is functionalisation of D-galactal, which easily obtained from the readily available D-galactose. Numerous *N*-functionalities have been introduced at the C-2 position

to yield D-galactosamine derivatives, including azide,[419] amide,[25,26] sulfonamide,[420] nitroso,[421] hydrazine,[29,30] triazene,[422] nitro[423] and carbamate[424] functional groups. However, many of these syntheses suffer from poor yields relating to unselective addition to the glycal and the non-trivial conversion of somewhat exotic adducts to final glycosylated products. In fact, one of the most widely used methods is also one of the oldest, that of glycal azidonitration developed by Lemieux *et al.* in 1979.[419] This method has remained attractive due to the direct incorporation of a synthetically useful azide handle at C-2, despite the formation of epimeric by-products in appreciable quantities. Alternative routes have also been demonstrated, including nucleophilic attack on dianhydro-β-D-talose derivatives,[425] chain extension of D-lyxose[426] and Heyns' rearrangement of D-tagatose.[427] Unfortunately, the rare monosaccharide starting materials involved have hindered the practical application of these approaches. Considering the limits of all the aforementioned methods, there is significant scope for improvement and a novel route to D-galactosamine derivatives would be of great value, particularly if it possesses key attributes of: a) expedient synthesis from an abundant precursor, b) tolerance of a wide range of protecting groups and c) high yielding glycosylation following activation.

5.2.2 Synthetic plan

The aim is a simple and expedient synthetic route to D-galactosamine derivatives starting from cheap and abundant hexoses, with regiospecific introduction of diverse protecting groups for facile functionalisation downstream into specific glycosylation donors and acceptors. Our proposed route utilises methodology involving rhodium catalysed C–H activation with accompanying nitrene insertion.[428-430]

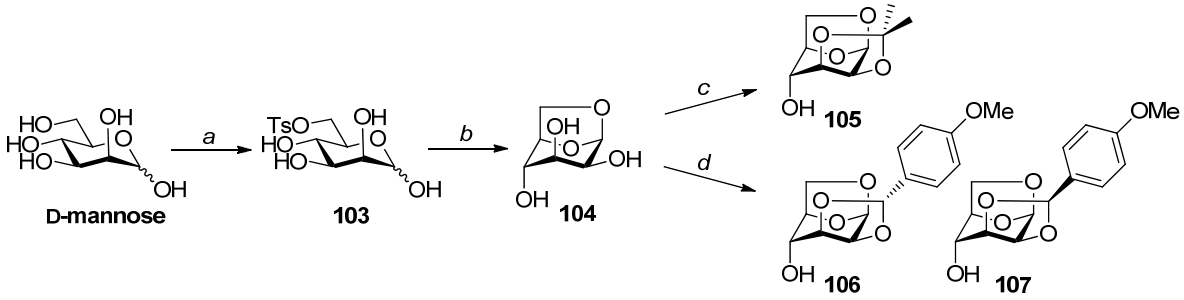


Scheme 5.1 Proposed route to 2-amino-2-deoxy-galactose derivatives

As outlined in Scheme 5.1, the starting material is readily available D-mannose. Dehydration to 1,6-anhydro-D-mannose will lock the sugar in an axial-rich conformation, prior to selective hydroxyl protection and sulfamoylation of 4-OH. Sulfamate-derived nitrenes have been shown to greatly favour insertion into γ- rather than β-C–H bonds.[428-430] It has been proposed this is due to the elongated S–O and S–N bonds (1.58 Å) and the obtuse angle of the N–S–O motif in the sulfamate (103°), matching the metric parameters of the six-membered heterocycle and thus making the formation of a 5-membered ring less kinetically favourable.[428-430] The irregular $^{1}C_{4}$ conformation of our 1,6-anhydro-D-mannose scaffold is required to bring the γ-C_2–H into close proximity of the sulfamate nitrogen in **100** for efficient nitrene insertion. Other methods of locking the hexose in an axial rich conformation were considered - including 2,6-[431] and 3,6-anhydro[432] derivatives as well as alternative bridging groups[433] – but the ease of installation and the potential of the 1,6-anhydro ring to act as a glycosyl donor proved decisive. Following Rh^{II} catalysed oxathiazinane formation (**101**), hydride attack on a transiently formed iminium ion will yield **102** in which the nitrogen can be selectively protected prior to nucleophilic inversion of the C-4 centre. Anomerically assisted opening of the 1,6-anhydro ring completes the

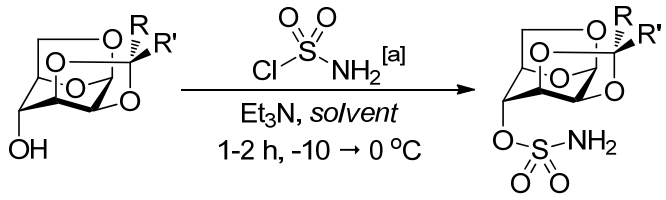
synthesis, with scope to incorporate a wide range of nucleophiles, including glycosyl acceptors. The resulting products will be selectively protected D-galactosamine derivatives, with the starting D-mannose having undergone a selective amination and inversion at both C-2 and C-4.

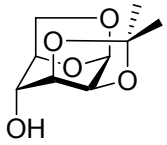
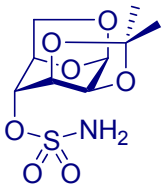
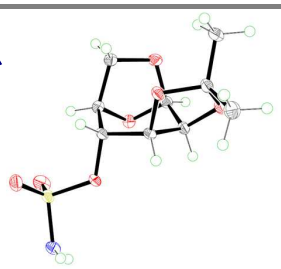
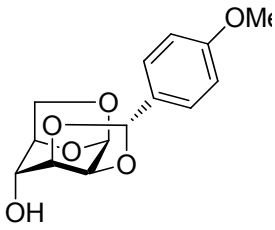
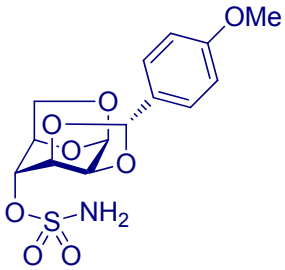
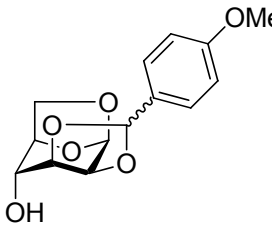
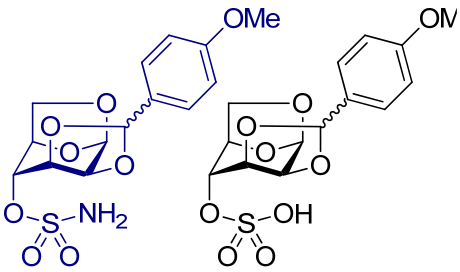
5.2.3 Synthesis of sulfamate ester building blocks



Scheme 5.2 a) TsCl, pyr, 48 h; b) 1M aq. NaOH, pH 9, 18 h; c) 2,2-dimethoxypropane, TsOH, acetone, 72 h; **105** 25% (from D-mannose); d) 4-methoxybenzaldehyde dimethyl acetal, TsOH, DMF, 50 ℃, 300 mbar, 4 h; 8% **106**, 15% **106** + **107** (from D -mannose);

To access the necessary axial-rich 1,6-anhydromannose derivatives, D-mannose was first reacted with TsCl in pyridine and then exposed to base, whereupon intramolecular displacement of the resulting 6-*O*-tosylate afforded the dehydrated sugar **104**.[434] When working on a larger scale the TsOH and NaCl by-products were successfully removed by anionic exchange with Dowex 550A resin (OH^- form) and precipitation with methanol respectively. Although minor sugar impurities remained, the product was sufficiently pure to undergo acetal formation without requiring further purification by flash chromatography. Isopropylidene and 4-methoxybenzylidene acetals were subsequently formed in 22-25 % yield from D-mannose using standard conditions (Scheme 5.2).[434,435] The moderate observed yields can be attributed to incomplete selectivity during tosylation and subsequent displacement reactions, although it should not be forgotten that a structurally rich scaffold has been obtained from a cheap and abundant starting material in the process. The anisaldehyde acetal was obtained as a mixture of *endo* and *exo* isomers, in a 2.3:1 ratio (*endo*/*exo*). Although inseparable by chromatography, *endo* isomer **106** was selectively crystallised from acetone and petroleum ether, leaving behind an *exo*-enriched mother liquor. This allowed us to test whether the acetal stereochemistry has any effect on the regioselectivity of the subsequent Rh(II) catalysed C-H amination.



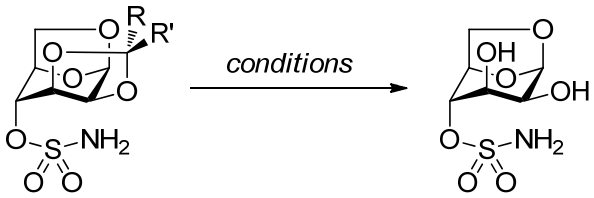
Entry	Alcohol	Conditions	Product	Yield
1	 **105**	DCM	 **108**	45%
2		MeCN		19%
3		DMA		80%
4	 **106**	DCM	 **109**	76%
5		DCM, 4 h		**109**+**110**: 39% **111**: 30%
6	**106** + **107**	DCM, 2 h	**109** + **110** **111**	**109**+**110**: 46% **111**: 0%

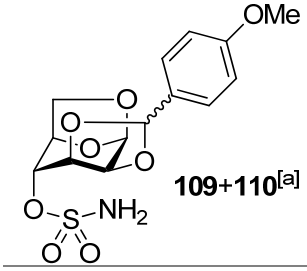
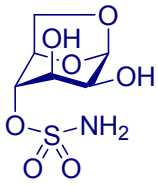
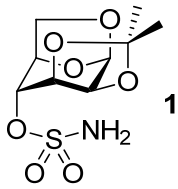
[a] Formed *in situ* from $ClSO_2NCO$ and HCOOH, stirring for 17 h in dry DCM (NB. MeCN for entry 2) prior to addition of the alcohol;

Table 5.1 Sulfamate formation from 1,6-anhydro-D-mannose derivatives

These acetal protected 1,6-anhydromannose derivatives next required conversion to sulfamate esters (Table 5.1). Using existing procedures acetonide **108** was isolated in 45 % yield, although work-up and purification steps were modified to accommodate for the high water solubility of the products.[428,436] In order to improve the modest yield observed with DCM, a number of reaction solvents were screened. DMA proved most effective, resulting in an 80 % conversion to sulfamate **108**, of which an X-ray crystal structure was obtained (Entries 1-3). The use of DMF gave a complex mixture of products (results not shown),

attributed to the previously reported propensity of this solvent to form adducts with sulfamates.[437] DMA was thus chosen as an alternative because its carbonyl group was less electrophilic and therefore less prone to forming these adducts. Anisaldehyde acetal **106** and the *exo*-rich epimeric mixture **106**+**107** were also converted to their corresponding sulfamate esters, although in the latter case a sulfate by-product was isolated (**111**). This could be explained by further sulfamoylation of the product sulfamoyl group, which could hydrolyse during purification; this has been shown to occur under extended reaction times.[438] A reduced reaction time alleviated this problem (Entries 5-6).



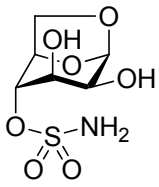
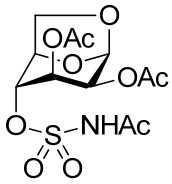
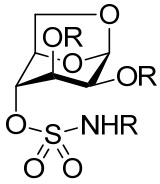
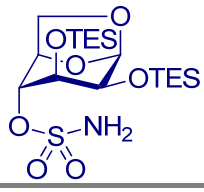
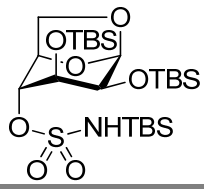
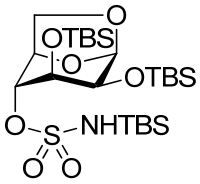
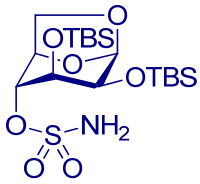
Entry	Sulfamate Ester	Conditions	Product	Yield
1	 **109**+**110**[a]	$Pd(OH)_2$ / H_2 *EtOH, 24 h*	 **112**	-
2		CAN *$MeCN/H_2O$ (9:1), 48 h*		- [b]
3		80% AcOH *24 h*		25%
4		TsOH (0.2 eq) *MeOH, 24 h*		77%
5	 **108**	TsOH (0.2 eq) *MeOH, 24 h*		-
6		TsOH (0.6 eq) *$MeOH/H_2O$ 4:1, 2 h, 70 °C*		- [c]
7		PPTS (0.6 eq) *$MeOH/H_2O$ 4:1, 3 h, 70 °C*		- [c]
8		TFA/H_2O 4:1 *18 h*		100%

[a] The benzylidene sulfamate was used as a mixture of isomers; [b] Some oxidation to the 4-methoxybenzoyl derivative was detected, but the reaction could not be driven to completion; [c] Some conversion to product was observed, although the reaction could not be driven to completion;

Table 5.2 Sulfamate acetal deprotection

In order to test the effect of ring conformation on C–H amination, portions of the acetal-protected sulfamates were converted to acyclic derivatives, in order to increase the flexibility of the tricyclic skeleton. To this end, a variety of conditions were trialled for the deprotection of existing acetals **108**, **109** and **110** (Table 5.2). In the case of the anisaldehyde acetal, the most effective method involved treatment of the epimeric mixture with TsOH in methanol, affording diol **112** in 77% yield (Entry 4). Acetonide **108** was found to be much more stable to hydrolysis, and the aforementioned conditions were not successful even with elevated temperatures and a near stoichiometric amount of acid. However, upon treatment with 80% aqueous TFA, the acetonide was successfully removed in quantitative yield (Entry 8). These results show that the sulfamate ester groups in these molecules are more robust in strongly acidic conditions than had been anticipated.

With deprotected sulfamate **112** in hand, selective protection of the 2,3-diol was attempted (Table 5.3). Acetate and silyl protecting groups were chosen as they would not form any additional rings, thereby avoiding the extra conformational rigidity present in the acetal derivatives. Standard acetylation with acetic anhydride and pyridine led purely to the formation of triacetate **113** (Entry 1). A more controlled approach was attempted, where it was hoped the greater reactivity of the two hydroxyl groups could be utilised to avoid concomitant sulfamate acetylation. Reaction time and equivalents of acetic anhydride were varied, along with the amounts of bases/nucleophilic catalysts such as triethylamine and pyridine. However, a complex distribution of acetylation states was always observed, with prolonged persistence of diol **112** and monoacetate **117** accompanying appearance of triacetate **113** (Entries 2-5). Even when diacetate **118** was isolated – running as a single spot by TLC – it was found by NMR to contain a mixture of all three possible isomers. It appears that the hydroxyl groups have a much closer level of reactivity to the sulfamate than expected, thereby greatly complicating their selective acetylation.

Entry	Sulfamate Ester	Conditions	Product	Yield
1	 **112**	Ac_2O/pyridine 1:1	 **113**	69%
2		Ac_2O (2 eq) pyridine, 21 h		**0×Ac**: 5%[a] **1×Ac**: 30% **2×Ac**: 45% **3×Ac**: 20%
3		Ac_2O (2 eq) Et_3N (1 eq) 1,4-dioxane, 10 h		**0×Ac**: 25% **1×Ac**: 60% **2×Ac**: 15% **3×Ac**: 0%
4		Ac_2O (6 eq) Et_3N (3 eq) 1,4-dioxane, 48 h	R = Ac or H **117**: 1xAc = Monoacetate **118**: 2xAc = Diacetate **113**: 3xAc = Triacetate	**1×Ac**: 30% **2×Ac**: 45% **3×Ac**: 25%
5		Ac_2O (2.1 eq) DCM, 7 d		**0×Ac**: 10% **1×Ac**: 40% **2×Ac**: 45% **3×Ac**: 5%
6		TESCl (10 eq) Imidazole (15 eq) *1,4-dioxane, -10→25 °C*	 **115**	36%
7		TBSCl (10 eq) Imidazole (15 eq) *DMA, 48 h* 0.8M	 **114**	**diTBS**: 10% **triTBS**: 53%[b]
8	 **114**	Dowex 50WX8-H^+ MeOH	 **116**	-
9		TFA/THF/H_2O 1:10:5 *0°C, 90 min*		79%

[a] Product ratios estimated by TLC and MS analysis; [b] Decomposed slowly to di-TBS compound on prolonged storage at room temperature;

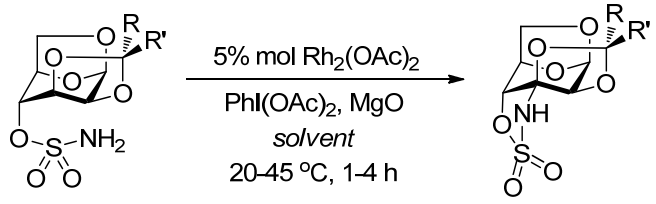
Table 5.3 Selective protection of sulfamate diol

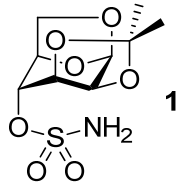
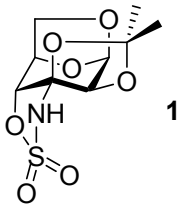
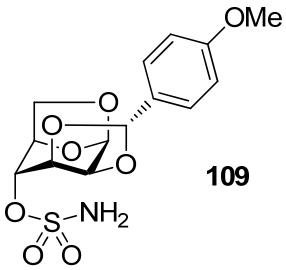
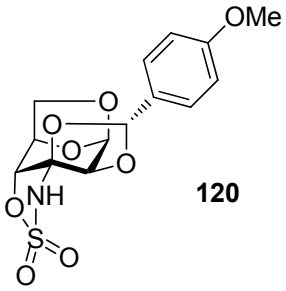
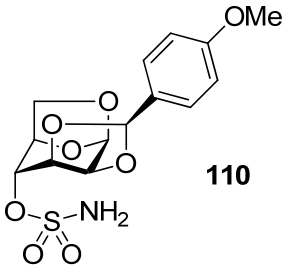
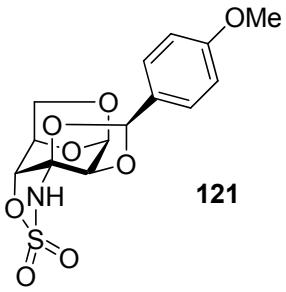
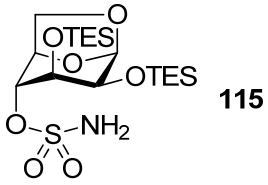
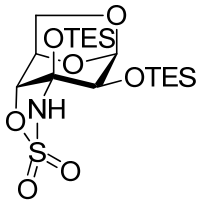
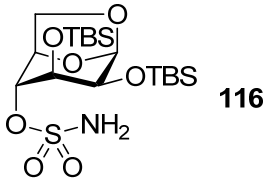
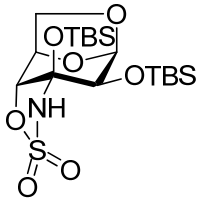
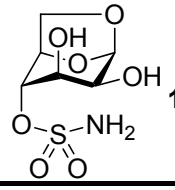
Attention was next switched to silyl group protection. Since Si–N bonds are known to be significantly weaker than Si–O bonds,[439] our strategy envisaged global silylation followed by selective *N*-desilylation – achieved either *in situ* as part of a standard work-up procedure under or under mild acidic conditions. Initially, reaction of the diol with TES-Cl

gave disilyl ether **115** in 19 % yield (Entry 6). Attempts to improve this yield with more reactive TES-OTf were unsuccessful. Silylation with TBS-Cl was extremely sluggish, and required a very concentrated reaction. Under these conditions, trisilylated derivative **114** was isolated as the major product (Entry 7). Fortunately, an efficient method was found to selectively deprotect the *N*-TBS sulfamate in the presence of both silyl ethers (Entry 9).[440]

5.2.4 Rhodium (II) catalysed nitrene insertion

With six differentially protected 1,6-anhydro-D-mannose derivatives in hand, the stage was set for Rh(II) catalysed nitrene insertion at C-2. These sulfamates were exposed to C–H amination conditions developed by Du Bois *et al.* (Table 5.4).[428] All of the derivatives tested successfully underwent sulfamate insertion with the exception of diol **112**, whose free hydroxyls presumably formed complexes with the rhodium catalyst, inactivating it and thereby precluding nitrene formation (Entry 7). It was found that a high concentration of sugar was required for efficient conversion; fortunately the reaction was tolerant to a number of aprotic solvents allowing the reaction solvent to be optimised based on the solubility profile of individual substrates. With the anisaldehyde acetal derivatives, the *endo* isomer was converted to the insertion product **120** in 92 % yield (Entry 2). The *exo* isomer was reacted as an *exo*-enriched mixture of acetal isomers, and a single spot by TLC was isolated in 49 % yield. The NMR contained two distinct products, in the same 2:1 ratio as the epimeric mixture of starting materials. The minor compound corresponded exactly to the *endo* product **120**, meaning the other was derived from the *exo* sulfamate (**110**), allowing the *exo* product to be successfully assigned (Entry 4). The poor yield observed with the di-TES derivative (**115**) may be from losses during flash chromatography, caused by the lability of *O*-TES in the mildly acidic conditions of a silica column (Entry 5).



Entry	Sulfamate Ester	Conditions	Product	Yield
1	 **108**	THF 0.05M	 **119**	45% (rsm: 54%)[b]
2	 **109**	1,4-dioxane 0.2M	 **120**	92%
3	 **110**	THF 0.02M	 **121**	- [a]
4		1,4-dioxane 0.2M		49%[a]
5	 **115**	DCM 0.2M	 **122**	19% (rsm: 14%)
6	 **116**	DCM 0.2M	 **123**	61%
7	 **112**	1,4-dioxane 0.2M	-	0% (rsm: 93%)

[a] Reaction performed on *exo*-enriched mixture of isomers **109**+**110**; [b] recovered starting material

Table 5.4 Rh(II) catalysed C–H amination of 1,6-anhydro-D-mannose sulfamates

Careful analysis of the NMR spectra obtained from all the sulfamidate products confirmed *N*-insertion had indeed taken place - one of the seven hexose-derived ^{1}H shifts in each starting material had disappeared and a peak in the ^{13}C NMR of each compound was missing in HSQC and DEPT spectra, indicating a quaternary carbon centre. However, it became apparent that *N*-insertion had occurred at C-3 rather than C-2. This was unexpected because with sulfamates, it is known that insertion into the γ C–H is preferred over the β C–H, and as such the 6-membered product is favoured over a 5-membered oxathiazolidine ring. It has been proposed this is due to the elongated S–O and S–N bonds (1.58 Å) and the obtuse angle of the N–S–O motif in the sulfamate (103°), matching the metric parameters of the six-membered heterocycle and thus making the formation of a 5-membered ring less kinetically favourable.[37-39] Our unexpected result was confirmed in each case by a clear $^{3}J_{1\text{-}2}$ coupling of ~ 2 Hz confirmed by COSY, and lack of any $^{3}J_{2\text{-}3}$ or $^{3}J_{3\text{-}4}$ coupling. In addition, a HMBC spectrum of acetonide **119** showed distinct coupling of C-3 in this compound to H-1, H-2, H-4 and H-5.

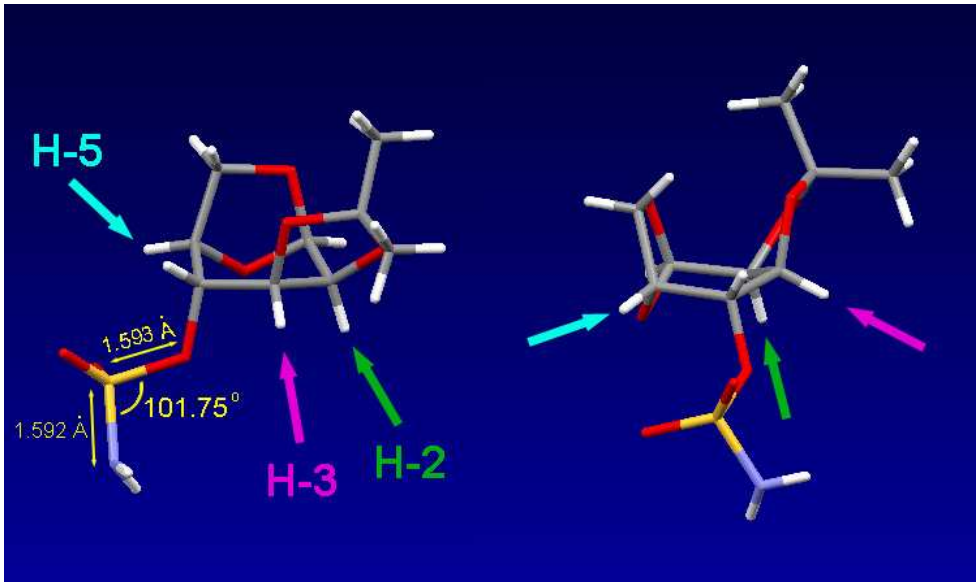


Fig. 5.4 X-Ray crystal structure of 1,6-anhydro-2,3-*O*-isopropylidene-4-*O*-sulfamoyl-β-d-mannopyranose **108** visualised using Mercury 1.4.2 software[441]

We were able to obtain an X-ray crystal structure of acetonide sulfamate **108**, which shed some light on the unusual regioselectivity observed. As can be seen in Fig. 5.4, the C–H bonds at C-2, C-3 and C-5 are all accessible to cyclisation with the sulfamate. They are all tertiary C–H bonds with an α-ethereal centre, a motif shown to strongly favour nitrene insertion.[436] On this basis, the crystal structure of sulfamate **108** would support insertion at C-2 – the S–O^{4} and S–N bond lengths are 1.593 and 1.592 Å respectively and

the O4–S–N angle is 101.75°, closely matching literature data.[428] Therefore it was especially surprising that in each of the five examples only a single product was observed: cyclisation to C-3 to form a five-membered oxathiazolidine. This is despite the presence of two alternative cyclisation pathways at C-2 and C-5, one of which offering a much more favoured γ insertion.

There are several potential origins of this clear selectivity. The rigid bi- or tricyclic scaffold might present the β-C_3–H bond in a much more favourable conformation for nitrene insertion than the γ-C_2–H bond. The crystal structure does indeed show the pyranose ring to be distorted away from a chair conformation, leaving the axial bonds at C–2 and C–4 no longer parallel but pointing slightly away from each other: the O–C_2–C_4–O torsion angle is -157° and the angle between the C_2–O and C_4–O vectors is 23°. A potential determining effect consistent with the observed specificity of attack at C–3 over C–5 can be considered through stereoelectronic effects. Both C–H bonds are β to the sulfamate and both show similar spatial orientation relative to C_4–O. However, the C_3–H bond is selectively weakened by hyperconjugation from the adjacent oxygen lone pair, which is held in a suitable conformation to overlap well with the C_3–H σ* orbital (Fig. 5.5). In contrast, the lone pair orbitals of the pyranose oxygen are more poorly aligned with σ*(C_5–H), making this bond less activated for nitrene insertion. Such stereoelectronic effects may explain why the identity and/or stereochemistry of the hydroxyl protecting groups at O–2 and O–3 had no bearing on the regioselectivity of the insertion reaction.

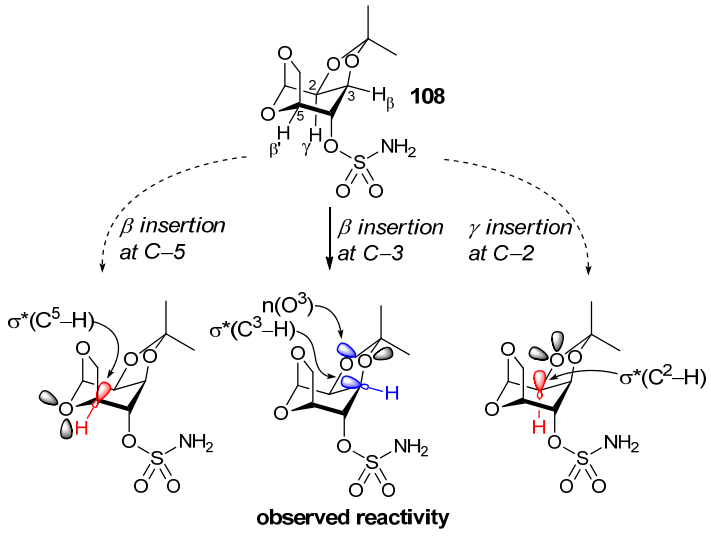


Fig. 5.5 Stereoelectronic analysis of observed regioselectivity in C–H insertion amination

5.3 Conclusions

With respect to the original goals, the results obtained here would unfortunately lead to 3-amino-3-deoxy-D-idose rather than the desired 2-amino-2-deoxy-D-galactose derivatives. However, these studies have important broader implications. The study represents an exceptional example of selective C–H activation in the presence of multiple potential insertion sites. The inverted regioselectivity exemplified by selective β-insertion took place in the presence of an equally activated γ-C–H, which would normally be greatly favoured.[37-39] Additionally, remarkable selectivity was observed between the two β-C–H centres at C-3 and C-5; this selectivity has been rationalised through consideration of stereoelectronic effects.

To the best of our knowledge, this is also the first example of a sulfamate cyclisation forming a five membered ring in preference to a six membered ring based on conformation alone. Although β-insertion has been reported with sulfamates, this has only been observed when no γ pathway is available,[37,51] or if an electron-donating group is present at the β but not the γ position, providing a much more active C–H for insertion.[442] Formation of 7 and even 8-membered sulfamidates has been reported, but the presence of a heteroatom adjacent to the site of insertion was the driving force for specific C–H activation, rather than pure conformational constraints.[443]

In conclusion, this is an interesting and unique observation which will further our understanding of the influencing factors and mechanistic aspects involved in C–H activation and nitrene insertion methodologies.[444]

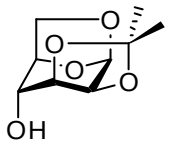
5.4 Experimental

5.4.1 General Considerations

For details of equipment, analytical parameters, reagents, solvents and other general considerations please refer to section 3.5.3.

5.4.2 Synthetic procedures

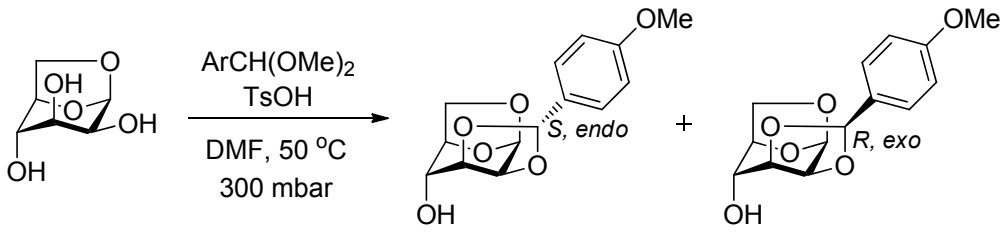
1,6-Anhydro-2,3-*O*-isopropylidene-β-D-mannopyranose 105[434]



D-Mannose (10.0 g, 55.5 mmol) and *p*-toluenesulfonyl chloride (14.8 g, 77.7 mmol) were dissolved in dry pyridine (70 mL) and the solution cooled to 0 °C. After stirring for 5 days, TLC (MeOH/EtOAc 1:4) showed a spot at R_f 0.5 (6-*O*-*p*-toluenesulfonyl-D-mannose). 1 M aq. NaOH (50 mL) was added to the golden solution, followed by slow addition of sat. aq. NaOH until a pH of 9 was reached. An associated colour change was observed, with the golden solution turning dark brown when an alkaline pH was achieved. After stirring for 18 h TLC (MeOH/EtOAc 1:4) showed spots at R_f 0.5 (6-*O*-*p*-toluenesulfonyl-D-mannose) and 0.3 (1,6-anhydro-D-mannose). The reaction was neutralised with 37% aq. HCl, and the solvent removed *in vacuo*. The orange residue was dissolved in MeOH and dry loaded onto silica. Crude purification by flash chromatography (MeOH/EtOAc 3:17) through a short silica plug was performed, yielding 1,6-anhydro-D-mannose (7.40 g) as a pale brown solid. This crude product was suspended in dry acetone (200 mL) and cooled to 0 °C, before *p*-toluenesulfonic acid (1.01 g, 5.28 mmol) and 2,2-dimethoxypropane (6.73 mL, 54.7 mmol) were added. After stirring for 72 h at room temperature, TLC (EtOAc/Petrol 4:1) showed a single product at R_f 0.3. Triethylamine (1.5 mL) was added and the solvent removed *in vacuo*. The residue was dissolved in DCM (200 mL), washed with brine (150 mL) and dried over $MgSO_4$, before being filtered and concentrated *in vacuo*. The resulting solid was recrystallised from ethyl acetate and petroleum ether to yield 1,6-anhydro-2,3-*O*-isopropylidene-β-D-mannopyranose **105** (1.91 g, 9.45 mmol, 16% from D-mannose) as white crystals. Flash chromatography (EtOAc/Petrol 4:1) of the mother liquor afforded further product (931 mg, 4.60 mmol, 9% from D-mannose); **M.p.** = 152.5-155 °C, EtOAc, (Lit.[445] = 154-155 °C); $\mathbf{[\alpha]_D^{21}}$ = -51.3, c = 1.0, $CHCl_3$, (Lit.[446]: $[\alpha]_D^{21}$ = -58, c = 2.3, H_2O); **^{1}H NMR** (400 MHz, $CDCl_3$) δ 1.32, 1.53 (2 × s, 2 × 3 H, 2 × -CH_3), 2.74 (br. s, 1 H, 4-O*H*), 3.76 (dd, $J_{6a,6b}$ = 7.3 Hz, $^4J_{6,4}$ = 6.4 Hz, 1 H, H-6a), 3.96 (br. s., 1 H, H-5), 4.02 (dd, $J_{6b,6a}$ = 7.3 Hz, J = 0.8 Hz, 1 H, H-6b), 4.08 (dd,

$J_{2,3}$ = 6.1 Hz, $J_{2,1}$ = 2.9 Hz, 1 H, H-2), 4.21 (dd, $J_{3,2}$ = 6.1 Hz, $J_{3,4}$ = 1.3 Hz, 1 H, H-3), 4.52 (dd, $J_{4,6}$ = 6.4 Hz, $J_{4,3}$ = 1.3 Hz, 1 H, H-4), 5.34 (d, $J_{1,2}$ = 2.9 Hz, 1 H, H-1); **^{13}C NMR** (100 MHz, $CDCl_3$) δ 25.8, 25.9 (2 × -C(CH_3)$_2$), 64.3 (C-6), 69.4 (C-5), 72.0 (C-2), 75.7 (C-4), 76.1 (C-3), 99.5 (C-1), 110.2 (-C(CH_3)$_2$); **IR** (thin film) ν : 3500 (br, O-H), 3000 (w), 2950 (m), 2910 (m), 1390 (m) cm^{-1}; **MS** *m/z* (ESI$^-$) : 201 [(M - H)$^-$, 20%], 403 [(2M - H)$^-$, 100%];

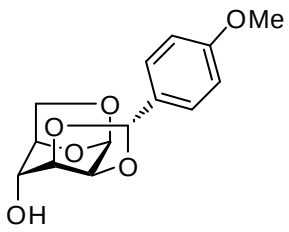
4-Methoxybenzylidene protection of 1,6-anhydro-β-D-mannopyranose 104 [447,448]



D-Mannose (2.61 g, 16.1 mmol) and *p*-toluenesulfonyl chloride (4.30 g, 22.5 mmol) were dissolved in dry pyridine (20 mL) and the solution cooled to 0 °C. After stirring for 5 days, TLC (MeOH/EtOAc 1:4) showed a spot at R_f 0.5 (6-*O*-*p*-toluenesulfonyl-D-mannose). 1 M aq. NaOH (12 mL) was added to the golden solution, followed by slow addition of sat. aq. NaOH until a pH of 9 was reached. An associated colour change was observed, with the golden solution turning dark brown when an alkaline pH was achieved. After stirring for 18 h TLC (MeOH/EtOAc 1:4) showed spots at R_f 0.5 (6-*O*-*p*-toluenesulfonyl-D-mannose) and 0.3 (1,6-anhydro-D-mannose). The reaction was neutralised with 37% aq. HCl, and the solvent removed *in vacuo*. The orange residue was dissolved in MeOH and dry loaded onto silica. Crude purification by flash chromatography (MeOH/EtOAc 3:17) through a short silica plug was performed, yielding 1,6-anhydro-D-mannose (7.40 g) as a pale brown solid. This crude solid was concentrated to dryness first from toluene/H_2O (9:1, 25 mL), then from toluene (25 mL), before being dissolved in dry DMF (5 mL). *p*-Toluenesulfonic acid (43 mg, 0.33 mmol) and 4-methoxybenzaldehyde dimethyl acetal (2.68 mL, 17.8 mmol) were added. After stirring for 4 h at 50 °C under reduced pressure (300 mbar), TLC (EtOAc/Petrol 4:1) showed spots at R_f 0.7 (4-methoxybenzaldehyde derivatives) and R_f 0.4 (product). Triethylamine (1 mL) was added and the solvent removed *in vacuo*. Flash chromatography (EtOAc/Petrol 1:1), afforded 1,6-anhydro-2,3-*O*-(4-methoxybenzylidene)-β-D-mannopyranose (1.01 g, 3.92 mmol, 23% from D-mannose) as a mixture of 4-methoxybenzylidene isomers (*endo*/*exo* 2.3:1). This white solid was recrystallised from acetone and petroleum ether to yield 1,6-anhydro-2,3-*O*-*endo*-(4-methoxybenzylidene)-β-D-mannopyranose **106** (373 mg, 1.29 mmol, 8% from D-mannose) as white crystals. The

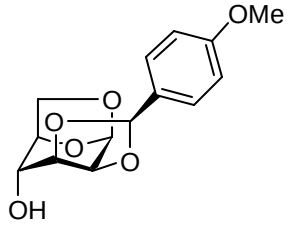
orange mother liquor contained 1,6-anhydro-2,3-*O*-(4-methoxybenzylidene)-β-D-mannopyranose (630 mg, 2.63 mmol, 15% from D-mannose) as an *exo*-enriched mixture of 4-methoxybenzylidene isomers (*endo*/*exo* 1.3:1);

1,6-Anhydro-*endo*-2,3-*O*-(4-methoxybenzylidene)-β-D-mannopyranose 106



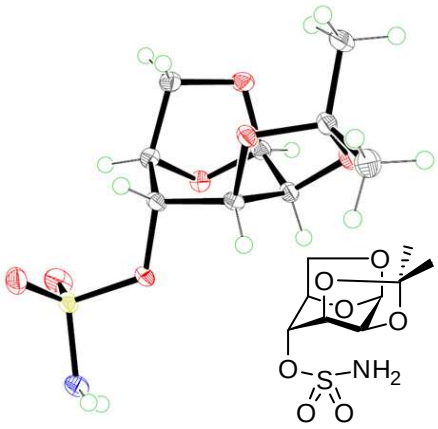
M.p. = 139-140 °C, EtOAc, (Lit.[448] = 141-142 °C); $[\alpha]_D^{21}$ = -80.1, c = 1.0, MeOH, (Lit.[448]: $[\alpha]_D^{20}$ = -81.5, c = 1.0, EtOH); **^{1}H NMR** (400 MHz, DMSO-d_6) δ ppm 3.66 (dd, $J_{6b,6a}$ = 7.3 Hz, $J_{6b,5}$ = 6.4 Hz, 1 H, H-6b), 3.76 (s, 3 H, -OCH_3), 3.83 (dd, $J_{6a,6b}$ = 7.3 Hz, $J_{6a,5}$ = 1.3 Hz, 1 H, H-6a), 3.88 (dt, $J_{4,4\text{-OH}}$ = 4.8 Hz, $J_{4,3}$ = 1.0 Hz, $J_{4,5}$ = 1.0 Hz, 1 H, H-4), 4.02 (dt, $J_{3,2}$ = 6.8 Hz, $J_{3,4}$ = 1.0 Hz, $^4J_{3,5}$ = 1.0 Hz, 1 H, H-3), 4.06 (dd, $J_{2,3}$ = 6.8 Hz, $J_{2,1}$ = 2.8 Hz, 1 H, H-2), 4.48 (app. dq, $J_{5,6b}$ = 6.4 Hz, $J_{5,6a}$ = 1.3 Hz, $J_{5,4}$ = 1.0 Hz, $^4J_{5,3}$ = 1.0 Hz, 1 H, H-5), 5.38 (d, $J_{1,2}$ = 2.8 Hz, 1 H, H-1), 5.57 (d, $J_{4\text{-OH},4}$ = 4.8 Hz, 1 H, 4-O*H*), 5.63 (s, 1 H, C$H_{\text{benzyllic}}$), 6.94 (d, *J* = 8.6 Hz, 2 × 1H, 2 × CH_{Ar}), 7.56 (d, *J* = 8.8 Hz, 2 × 1H, 2× CH_{Ar}); **^{13}C NMR** (100 MHz, DMSO-d_6) δ ppm 55.8 (-O*C*H$_3$), 64.7 (C-6), 69.2 (C-4), 71.4 (C-2), 76.1 (C-5), 78.7 (C-3), 98.8 (C-1), 103.6 (*C*H$_{\text{benzyllic}}$), 114.1 (2 C, 2 × *C*H$_{\text{Ar}}$), 129.6 (4°$_{\text{Ar}}$), 129.8 (2 C, 2 × *C*H$_{\text{Ar}}$), 160.8 (4°$_{\text{Ar}}$); **MS** *m/z* (ESI$^-$) : 279 [(M-H)$^-$, 70%], 559 [(2M-H)$^-$, 100%];

1,6-Anhydro-*exo*-2,3-*O*-(4-methoxybenzylidene)-β-D-mannopyranose 107



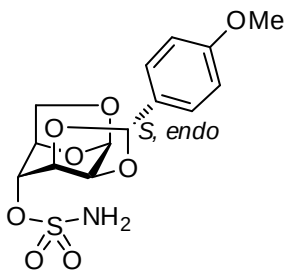
^{1}H NMR (400 MHz, DMSO-d_6) δ ppm 3.63 (dd, $J_{6b,6a}$ = 7.3 Hz, $J_{6b,5}$ = 6.1 Hz, 1 H, H-6b), 3.75 (s, 3 H, -OCH_3), 3.88 (d, $J_{4,4\text{-OH}}$ = 4.7 Hz, 1 H, H-4), 3.92 (dd, $J_{6a,6b}$ = 7.3 Hz, $J_{6a,5}$ = 1.0 Hz, 1 H, H-6a), 4.04 (m, 1 H, H-3), 4.33 (dd, $J_{2,3}$ = 5.9 Hz, $J_{2,1}$ = 2.8 Hz, 1 H, H-2), 4.48 (m, 1 H, H-5), 5.44 (d, $J_{1,2}$ = 2.8 Hz, 1 H, H-1), 5.49 (d, $J_{4\text{-O}H,4}$ = 4.8 Hz, 1 H, 4-O*H*), 6.10 (s, 1 H, C$H_{\text{benzyllic}}$), 6.91 (d, *J* = 8.6 Hz, 2 × 1H, 2 × CH_{Ar}), 7.31 (d, *J* = 8.6 Hz, 2 × 1H, 2 × CH_{Ar}), shifts were assigned from a spectrum containing an inseperable mixture of isomers, enriched in the *exo* isomer; **^{13}C NMR** (100Hz, DMSO-d_6) δ ppm 55.1 (-O*C*H$_3$), 64.7 (C-6), 68.7 (C-4), 72.8 (C-2), 75.3 (C-5 76.0 (C-3), 98.9 (C-1), 103.9 (*C*H$_{\text{benzyllic}}$), 113.6 (2 C, 2 × *C*H$_{\text{Ar}}$), 127.6 (2 C, 2 × *C*H$_{\text{Ar}}$), 132.1, 159.7 (2 × 4°$_{\text{Ar}}$), shifts were assigned from a spectrum containing an inseperable mixture of isomers, enriched in the *exo* isomer; **MS** *m/z* (ESI$^-$) : 279 [(M-H)$^-$, 70%], 559 [(2M-H)$^-$, 100%];

1,6-Anhydro-2,3-*O*-isopropylidene-4-*O*-sulfamoyl-β-D-mannopyranose 108



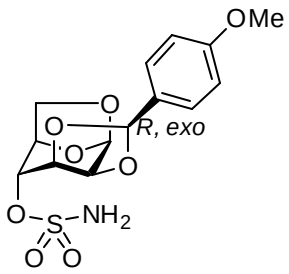
A flame dried flask was charged with DCM (2 mL) and cooled to 0 °C. Chlorosulfonyl isocyanate (700 µL, 8.15 mmol) was added, followed by dropwise addition of formic acid (300 µL, 8.15 mmol), which resulted in the formation of a white precipitate. After stirring for 17 h, the flask was cooled to -10 °C in a MeOH/ice bath. 1,6-Anhydro-2,3-*O*-isopropylidene-β-D-mannopyranose **104** (659 mg, 3.25 mmol) was dissolved in dry DMA (2 mL) and transferred *via* cannula to the reaction flask. After 1 h, triethylamine (1.36 mL, 9.75 mmol) was added, upon which a heavy cream precipitate formed in the clear yellow solution. After stirring at room temperature for 1 h TLC (EtOAc/Petrol 1:1) showed a single product at R_f 0.5. The solvent was removed *in vacuo* and the residue dry loaded onto silica from MeOH. Flash chromatography (EtOAc/Petrol gradient elution: 9:11→1:0), afforded 1,6-anhydro-2,3-*O*-isopropylidene-4-*O*-sulfamoyl-β-D-mannopyranose **108** (730 mg, 2.60 mmol, 80%) as colourless crystals; **Crystal Structure:** The X-ray crystal structure of **108** was solved by Dr. Amber Thompson at the Oxford University Chemical Crystallography Laboratory. Crystal data and structural parameters can be found in Appendix A6; $\mathbf{[\alpha]_D^{18}}$ = -49, *c* = 2, MeOH; **^{1}H NMR** (500 MHz, $CDCl_3$) δ ppm 1.33, 1.55 (2 × s, 2 × 3 H, 2 × -CH_3), 3.78 (dd, $J_{6a,6b}$ = 7.4 Hz, $J_{6a,5}$ = 6.3 Hz, 1 H, H-6a), 4.05 (dd, $J_{6b,6a}$ = 7.4 Hz, $J_{6b,5}$ = 1.3 Hz, 1 H, H-6b), 4.07 (dd, $J_{2,3}$ = 6.3 Hz, $J_{2,1}$ = 2.8 Hz, 1 H, H-2), 4.19 (d, $J_{3,2}$ = 6.3 Hz, 1 H, H-3), 4.62 (dd, $J_{5,6a}$ = 6.3 Hz, $J_{5,6b}$ = 1.3 Hz, 1 H, H-5), 4.89 (br. s., 2 H, -NH_2), 4.95 (s, 1 H, H-4), 5.38 (d, $J_{1,2}$ = 2.8 Hz, 1 H, H-1); **^{13}C NMR** (125 MHz, $CDCl_3$) δ ppm 25.8, 25.9 (2 × -C(*C*H_3)$_2$), 64.5 (C-6), 71.6 (C-4), 72.1 (C-2), 73.5 (C-5), 73.8 (C-3), 99.2 (C-1), 110.2 (-*C*(CH$_3$)$_2$); **IR** (thin film) *v* : 3264 (br, N-H), 3268 (br, N-H), 2980 (m), 2360 (s), 2341 (m), 1373 (s, -SO_2NH_2), 1185 (s, -SO_2NH_2) cm^{-1}; **MS** *m/z* (ESI$^-$) : 280 [(M-H)$^-$, 100%], 316 [(M+^{35}Br)$^-$, 70%], 318 [(M+^{37}Br)$^-$, 40%]; **HRMS** *m/z* (ESI$^+$) : calc. for $C_9H_{13}NNaO_7S$ (M + Na)$^+$ = 302.0305. Found 302.0299;

1,6-Anhydro-*endo*-2,3-*O*-(4-methoxybenzylidene)-4-*O*-sulfamoyl-β-D-mannopyranose 109



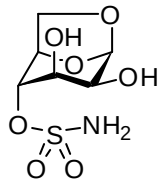
A flame dried flask was charged with DCM (5 mL) and cooled to 0 ºC. Chlorosulfonyl isocyanate (326 μL, 2.68 mmol) was added, followed by dropwise addition of formic acid (100 μL, 2.68 mmol), which resulted in the formation of a white precipitate. After stirring for 16 h, the flask was cooled to -10 ºC in a MeOH/ice bath. 1,6-Anhydro-*endo*-2,3-*O*-(4-methoxybenzylidene)-β-D-mannopyranose **106** (300 mg, 1.07 mmol) was dissolved in dry DCM (10 mL) and triethylamine (750 μL, 5.35 mmol), and transferred *via* cannula to the reaction flask. After 1 h stirring at room temperature under N_2, TLC (EtOAc/Petrol 3:1) showed a single product at R_f 0.5. Triethylamine was added until the solution was no longer acidic. The solvent was removed *in vacuo* and the residue dry loaded onto silica from MeOH. Flash chromatography (EtOAc/Petrol gradient elution: 1:1→6:4), afforded 1,6-anhydro-*endo*-2,3-*O*-(4-methoxybenzylidene)-4-*O*-sulfamoyl-β-D-mannopyranose **109** (292 mg, 0.81 mmol, 76%) as a white amorphous solid; $\mathbf{[\alpha]_D^{18}}$ = -26, *c* = 1, MeOH; **^{1}H NMR** (400 MHz, MeOD-*d*$_4$) δ ppm 3.79 (s, 3 H, -OC*H*$_3$), 3.80 (dd, $J_{6b,6a}$ = 7.6 Hz, $J_{6b,5}$ = 6.6 Hz, 1 H, H-6b), 4.03 (dd, $J_{6a,6b}$ = 7.6 Hz, $J_{6a,5}$ = 1.3 Hz, 1 H, H-6a), 4.15 (dd, $J_{2,3}$ = 6.8 Hz, $J_{2,1}$ = 3.0 Hz, 1 H, H-2), 4.38 (dq, $J_{3,2}$ = 6.8 Hz, $J_{3,4}$ = 1.0 Hz, $^4J_{3,5}$ = 1.0 Hz, $^4J_{3,1}$ = 0.8 Hz, 1 H, H-3), 4.77 (t, $J_{4,5}$ = 1.0 Hz, $J_{4,3}$ = 1.0 Hz, 1 H, H-4), 4.79 (dq, $J_{5,6b}$ = 6.6 Hz, $J_{5,6a}$ = 1.3 Hz, $J_{5,4}$ = 1.0 Hz, $^4J_{5,3}$ = 1.0 Hz, 1 H, H-5), 5.44 (dd, $J_{1,2}$ = 3.0 Hz, $^4J_{1,3}$ = 0.8 Hz, 1 H, H-1), 5.71 (s, 1 H, C*H*$_{benzyllic}$), 6.91 (d, *J* = 8.6 Hz, 2 × 1H, 2 × C*H*$_{Ar}$), 7.59 (d, *J* = 8.6 Hz, 2 × 1H, 2 × C*H*$_{Ar}$); **^{13}C NMR** (100 MHz, MeOD-*d*$_4$) δ ppm 55.7 (-O*C*H$_3$), 65.3 (C-6), 72.4 (C-2), 75.2 (C-5), 77.2 (C-3), 77.3 (C-4), 100.1 (C-1), 105.6 (*C*H$_{benzyllic}$), 114.4 (2 C, 2 × *C*H$_{Ar}$), 129.7 (4°$_{Ar}$), 130.3 (2 C, 2 × *C*H$_{Ar}$), 162.2 (4°$_{Ar}$); **IR** (thin film) *v* : 3288 (br, N-H), 3272 (br, N-H), 2955 (s), 2930 (s), 2897 (s), 2858 (s), 1718 (s, C=C^{Ar}) 1363 (s, -SO_2NH_2), 1256 (s), 1185 (s, -SO_2NH_2) cm^{-1}; **MS** *m/z* (ESI$^-$) : 358 [(M-H)$^-$, 100%], 717 [(2M-H)$^-$, 80%]; **HRMS** *m/z* (ESI$^-$) : calc. for $C_{14}H_{16}NO_8S$ (M - H)$^-$ = 358.0602. Found 358.0603;

Exo-enriched 1,6-anhydro-2,3-_O_-(4-methoxybenzylidene)-4-_O_-sulfamoyl-β-D-manno pyranose 110



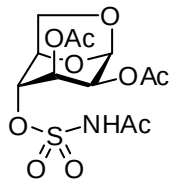
A flame dried flask was charged with DCM (10 mL) and cooled to 0 °C. Chlorosulfonyl isocyanate (684 μL, 5.62 mmol) was added, followed by dropwise addition of formic acid (212 μL, 5.62 mmol) which resulted in the formation of a white precipitate. After stirring for 16 h, the flask was cooled to -10 ºC in a MeOH/ice bath. An *exo*-enriched isomeric mixture of 1,6-anhydro-2,3-*O*-(4-methoxybenzylidene)-β-D-mannopyranose (630 mg, 2.25 mmol) was dissolved in dry DCM (10 mL) and triethylamine (1.57 mL, 11.25 mmol), and transferred *via* cannula to the reaction flask. After 2 h stirring at room temperature under N_2, TLC (EtOAc/Petrol 3:1) showed a spot at R_f 0.5 (product). Triethylamine was added until the solution was no longer acidic. The solvent was removed *in vacuo* and the residue dry loaded onto silica from MeOH. Flash chromatography (EtOAc/Petrol gradient elution: 1:1→7:3), afforded *exo*-enriched 1,6-anhydro-2,3-*O*-(4-methoxybenzylidene)-4-*O*-sulfamoyl-β-D-mannopyranose **110** (373 mg, 1.04 mmol, 46%) as an orange oil (*endo*/*exo* 1.3:1); **^{1}H NMR** (400 MHz, $CDCl_3$) δ ppm 3.81 (s, 3 H, -OCH_3), 3.84 (dd, $J_{6b,6a}$ = 7.8 Hz, $J_{6b,5}$ = 5.8 Hz, 1 H, H-6b), 4.13 (dd, $J_{6a,6b}$ = 7.8 Hz, $J_{6a,5}$ = 1.0 Hz, 1 H, H-6a), 4.39 (d, $J_{3,2}$ = 6.0 Hz, 2 H, H-3), 4.43 (dd, $J_{2,3}$ = 6.0 Hz, $J_{2,1}$ = 2.7 Hz, 1 H, H-2), 4.82 (m, 2 × 1H, H-4, H-5), 5.57 (d, $J_{1,2}$ = 2.7 Hz, 1 H, H-1), 6.28 (s, 1 H, C$H_{benzyllic}$), 6.91 (d, J = 8.8 Hz, 2 × 1H, 2 × CH_{Ar}), 7.33 (d, J = 8.8 Hz, 2 × 1H, 2 × CH_{Ar}), shifts were assigned from a spectrum of an inseparable mixture of isomers, enriched in the *exo* isomer; **^{13}C NMR** (100 MHz, $CDCl_3$) δ ppm 55.3 (-OCH_3), 65.0 (C-6), 73.2 (C-2), 73.6 (C-5), 75.4 (C-3), 77.4 (C-4), 99.7 (C-1), 105.5 (C$H_{benzyllic}$), 113.9 (2 C, 2 × CH_{Ar}), 127.6 (2 C, 2 × CH_{Ar}), 130.7, 160.4 (2 × 4°$_{Ar}$), shifts were assigned from a spectrum of an inseperable mixture of isomers, enriched in the *exo* isomer; **MS** *m/z* (ESI^-) : 358 [$(M-H)^-$, 100%], 717 [$(2M-H)^-$, 80%]; **HRMS** *m/z* (ESI^-) : calc. for $C_{14}H_{16}NO_8S$ $(M - H)^-$ = 358.0602. Found 358.0603;

1,6-Anhydro-4-*O*-sulfamoyl-β-D-mannopyranose 112



A mixture of 1,6-anhydro-2,3-*O*-isopropylidene-4-*O*-sulfamoyl-β-D-mannopyranose isomers (754 mg, 2.68 mmol) was dissolved in TFA/H_2O 4:1 (8 mL). After stirring at room temperature for 18 h, TLC (MeOH/EtOAc 1:49) showed a single product at R_f 0.25. Removal of solvent *in vacuo* afforded 1,6-anhydro-4-*O*-sulfamoyl-β-D-mannopyranose **112** (646 mg, 2.68 mmol, 100%) as an off white amorphous solid; **$[\alpha]_D^{18}$** = -77, *c* = 0.8, MeOH; **^{1}H NMR** (400 MHz, MeOD-d_4) δ ppm 3.58 (dd, $J_{2,3}$ = 5.5 Hz, $J_{2,1}$ = 2.0 Hz, 1 H, H-2), 3.70 (dd, $J_{6b,6a}$ = 7.3 Hz, $J_{6b,5}$ = 6.1 Hz, 1 H, H-6b), 4.06 (dq, $J_{3,2}$ = 5.5 Hz, $^4J_{3,5}$ = 1.8 Hz, $J_{3,4}$ = 1.8 Hz, $^4J_{3,1}$ = 1.7 Hz, 1 H, H-3), 4.25 (dd, $J_{6a,6b}$ = 7.3 Hz, $J_{6a,5}$ = 0.8 Hz, 1 H, H-6a), 4.55 (t, $J_{4,5}$ = 1.8 Hz, $J_{4,3}$ = 1.8 Hz, 1 H, H-4), 4.68 (m, 1 H, H-5), 5.28 (dd, $J_{1,2}$ = 2.0 Hz, $^4J_{1,3}$ = 1.7 Hz, 1 H, H-1); **^{13}C NMR** (100 MHz, MeOD-d_4) δ ppm 65.7 (C-6), 67.7 (C-2), 70.3 (C-3), 75.5 (C-5), 80.5 (C-4), 103.2 (C-1); **IR** (thin film) *v* : 3282 (br, N-H, O-H), 1362 (s, -SO_2NH_2), 1181 (s, -SO_2NH_2) cm^{-1}; **MS** *m/z* (ESI^-) : 240 [$(M-H)^-$, 50%], 354 [$(M+TFA-H)^-$, 100%], 481 [$(2M-H)^-$, 80%]; (ESI^+) : 305 [$(M+MeCN+Na)^+$, 60%], 505 [$(2M+ Na)^+$, 100%]; **HRMS** *m/z* (ESI^-) : calc. for $C_6H_{10}NO_7S$ $(M - H)^-$ = 240.0183. Found 240.0191;

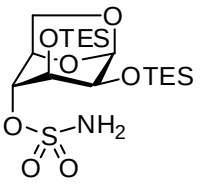
2,3-O-Acetyl-1,6-anhydro-4-*O*-(*N*-acetyl)sulfamoyl–β-D-mannopyranose 113



1,6-Anhydro-4-*O*-sulfamoyl-β-D-mannopyranose **112** (21 mg, 0.087 mmol) was dissolved in dry pyridine (500 µL) and acetic anhydride (500 µL). After stirring for 16 h at room temperature, TLC (MeOH/EtOAc 1:9) showed a single product at R_f 0.2. After the solvent removed *in vacuo*, flash chromatography (MeOH/EtOAc 3:17) afforded 2,3-*O*-acetyl-1,6-anhydro-4-*O*-(*N*-acetyl)sulfamoyl–β-D-mannopyranose **113** (22 mg, 0.060 mmol, 69%) as a clear oil; **$[\alpha]_D^{18}$** = -59, *c* = 1.0, $CHCl_3$; **^{1}H NMR** (400 MHz, DMSO-d_6) δ ppm 1.75 (s, 3 H, -CH_3^{NHAc}), 1.95 (s, 3 H, -CH_3^{Ac}, coupled to C-2 by HMBC), 2.06 (s, 3 H, -CH_3^{Ac}, coupled to C-3 by HMBC), 3.69 (dd, $J_{6b,6a}$ = 7.7 Hz, $J_{6b,5}$ = 6.1 Hz, 1 H, H-6b), 4.13 (dd, $J_{6a,6b}$ = 7.7 Hz, $J_{6a,5}$ = 0.8 Hz, 1 H, H-6a), 4.50 (t, $J_{4,5}$ = 1.5 Hz, $J_{4,3}$ = 1.5 Hz, 1 H, H-4), 4.64 (m, 1 H, H-5), 4.72 (dd, $J_{2,3}$ = 5.4 Hz, $J_{2,1}$ = 1.9 Hz, 1 H, H-2), 5.22 (dq, $J_{3,2}$ = 5.4 Hz, $J_{3,4}$ = 1.5 Hz, $^4J_{3,5}$ = 1.5 Hz, $^4J_{3,1}$ = 1.5 Hz, 1 H, H-3), 5.36 (dd, $J_{1,2}$ = 1.9 Hz, $^4J_{1,3}$ = 1.5 Hz, 1 H, H-1); **^{13}C NMR** (100 MHz, DMSO-d_6) δ ppm 21.2, 21.4, 26.8 (3 × -CH_3^{Ac}), 65.4 (C-6), 67.4 (C-2), 68.5 (C-3), 75.3 (C-5), 76.3 (C-4), 99.2 (C-1), 170.0, 170.0 (2 × C=O^{Ac}), 175.2 (C=O^{NHAc}); **IR** (thin film) *v* : 3500 (br, N-H), 2979 (w), 1742 (s, C=O^{Ac}), 1592 (s, C=O^{NAc}), 1374 (s, -SO_2NH_2) cm^{-1}; **MS** *m/z* (ESI^-) : 366 [$(M-H)^-$, 100%], 733 [$(2M-H)^-$, 40%]; (ESI^+) : 390

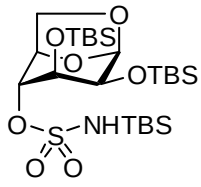
[(M+ Na)$^+$, 100%]; **HRMS** *m/z* (ESI$^-$) : calc. for $C_{12}H_{16}NO_{10}S$ (M - H)$^-$ = 366.0500. Found 366.0500;

1,6-Anhydro-2,3-bis-O-(triethylsilyl)-4-O-sulfamoyl-β-D-mannopyranose 115



1,6-Anhydro-4-*O*-sulfamoyl-β-D-mannopyranose **112** (51 mg, 0.212 mmol) and imidazole (182 mg, 2.67 mmol) were dissolved in dry DMA (420 µL). Chlorotriethylsilane (300 µL, 1.78 mmol) was then added. After stirring for 18 h at room temperature, TLC (EtOAc/Petrol 1:4) showed a single spot at R_f 0.2 (product). The solution was diluted with H_2O (5 mL). The aqueous layer was extracted with DCM (3 × 5 mL) and the combined organic layers washed with sat. aq. $NaHCO_3$ (5 mL) and brine (5 mL), before being dried over $MgSO_4$. Flash chromatography (EtOAc/Petrol gradient: 3:14→3:7), afforded 1,6-anhydro-2,3-bis-*O*-(triethylsilyl)-4-*O*-sulfamoyl–β-D-mannopyranose **115** (36 mg, 0.077 mmol, 36%) as a white amorphous solid; $\mathbf{[\alpha]_D^{18}}$ = -55, *c* = 1.0, MeOH; **^{1}H NMR** (400 MHz, MeOD-d_4) δ ppm 0.69 (m, 12 H, TES), 1.01 (m, 18 H, TES), 3.68 (m, 2 H, H-2, H-6b), 4.17 (m, 1 H, H-3), 4.31 (dd, $J_{6a,6b}$ = 7.1 Hz, $J_{6a,5}$ = 0.8 Hz, 1 H, H-6a), 4.47 (app t, $J_{4,3}$ = 2.0 Hz, $J_{4,5}$ = 2.0 Hz, 1 H, H-4), 4.65 (app d, $J_{5,6b}$ = 5.4 Hz, $J_{5,6a}$ = 0.8 Hz, $J_{5,4}$ = 2.0 Hz, 1 H, H-5), 5.22 (dd, $J_{1,2}$ = 1.7, $^4J_{1,3}$ = 1.2 Hz, 1 H, H-1); **^{13}C NMR** (100 MHz, MeOD-d_4) δ ppm 4.8 (6 C, TES), 6.2 (3 C, TES), 6.3 (3 C, TES), 64.9 (C-6), 69.4 (C-2), 72.1 (C-3), 74.3 (C-5), 80.2 (C-4), 102.7 (C-1); **IR** (thin film) *v* : 3350 (br, N-H), 3271 (br, N-H), 2955 (s), 2911 (m), 2877 (s), 1373 (s, -SO_2NH_2), 1185 (s, -SO_2NH_2) cm^{-1}; **MS** *m/z* (ESI$^-$) : 468 [(M-H)$^-$, 100%]; (ESI$^+$) : 487 [(M+NH_4)$^+$, 30%], 492 [(M+Na)$^+$, 100%], 533 [(M+MeCN+Na)$^+$, 70%]; **HRMS** *m/z* (ESI$^-$) : calc. for $C_{18}H_{39}NNaO_7SSi_2$ (M + Na)$^+$ = 492.1878. Found 492.1878;

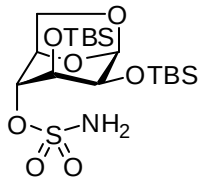
1,6-Anhydro-2,3-bis-*O*-(*tert*-butyldimethylsilyl)-4-*O*-(*N*-*tert*-butyldimethylsilyl)sulfamoyl–β-D-mannopyranose 114



1,6-Anhydro-4-*O*-sulfamoyl-β-D-mannopyranose **112** (293 mg, 1.22 mmol) and imidazole (1.25 g, 18.3 mmol) were dissolved in dry DMA (1.5 mL). *tert*-Butyldimethylchlorosilane (1.84 g, 12.2 mmol) was then added. After stirring for 48 h at room temperature, TLC (EtOAc/Petrol 1:4) showed spots at R_f 0.6 (product) and 0.2 (disilylated product). The yellow suspension was diluted with water (20 mL), extracted with DCM (3 × 20 mL) and the combined organic layers washed with sat. aq. $NaHCO_3$ (20 mL) and brine (15 mL), before drying over $MgSO_4$. Flash chromatography (EtOAc/Petrol gradient: 1:4→1:1) afforded 1,6-anhydro-2,3-bis-*O*-(*tert*-butyldimethylsilyl)-4-*O*-(*N*-*tert*-butyldimethylsilyl)sulfamoyl–β-D-

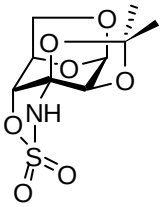
mannopyranose **114** (375 mg, 0.643 mmol, 53%) as a white amorphous solid and 1,6-anhydro-2,3-bis-*O*-(*tert*-butyldimethylsilyl)-4-*O*-sulfamoyl–β-D-mannopyranose **116** (58 mg, 0.124 mmol, 10%) as a white amorphous solid; $[\alpha]_D^{18}$ = -43, *c* = 0.8, $CHCl_3$; **^{1}H NMR** (400 MHz, $CDCl_3$) δ ppm 0.10, 0.11, 0.11, 0.13 (4 × s, 4 × 3 H, TBS), 0.30, 0.31 (2 × s, 2 × 3 H, TBS), 0.91, 0.93, 0.97 (3 × s, 3 × 9 H, TBS), 3.68 (dd, $J_{2,3}$ = 4.7 Hz, $J_{2,1}$ = 1.4 Hz, 1 H, H-2), 3.74 (dd, $J_{6b,6a}$ = 7.1 Hz, $J_{6b,5}$ = 6.1 Hz, 1 H, H-6b), 4.11 (m, 1 H, H-3), 4.31 (dd, $J_{6a,6b}$ = 7.1 Hz, $J_{6a,5}$ = 1.3 Hz, 1 H, H-6a), 4.46 (t, $J_{4,3}$ = 1.8 Hz, $J_{4,5}$ = 1.8 Hz, 1 H, H-4), 4.65 (m, 1 H, H-5), 5.29 (dd, $^4J_{1,3}$ = 1.5 Hz, $J_{1,2}$ = 1.4 Hz, 1 H, H-1); **^{13}C NMR** (100 MHz, $CDCl_3$) δ ppm -5.0, -4.9, -4.9, -4.8, -4.8, -4.0 (6 × 1 C, TBS), 17.4, 18.0, 18.4 (3 × 1 C, 4°$_{TBS}$), 25.7, 25.8, 26.1 (3 × 3 C, TBS), 64.8 (C-6), 69.1 (C-2), 71.4 (C-3), 73.8 (C-5), 79.6 (C-4), 102.6 (C-1); **IR** (thin film) *v* : 3263 (br, N-H), 2954 (s), 2931 (s), 2896 (s), 2859 (s), 1472 (m), 1375 (s, -SO_2NH_2), 1256 (s), 1176 (s, -SO_2NH_2) cm^{-1}; **MS** *m/z* (ESI^-) : 583 [(M-H)$^-$, 100%], 696 [(M+TFA-H)$^-$, 40%]; (ESI^+) : 601 [$(M+NH_4)^+$, 80%], 606 [$(M+Na)^+$, 100%], 647 [$(M+MeCN+Na)^+$, 90%];

1,6-Anhydro-2,3-bis-*O*-(*tert*-butyldimethylsilyl)-4-*O*-sulfamoyl-D-mannopyranose 116



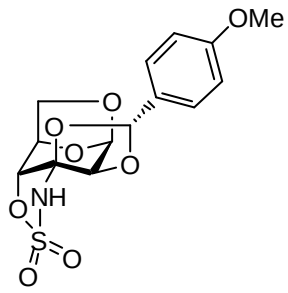
1,6-Anhydro-2,3-bis-*O*-(*tert*-butyldimethylsilyl)-4-*O*-(*N*-*tert*-butyldimethylsilylsulfamoyl)–β-D-mannopyranose **114** (62 mg, 0.11 mmol) was dissolved in TFA/THF/H_2O 1:10:5 (5.4 mL) at 0 ºC. After stirring at room temperature for 90 min, TLC (EtOAc/Petrol 3:7) showed a single product at R_f 0.25. Removal of solvent *in vacuo* followed by flash chromatography (EtOAc/Petrol 3:7) afforded 1,6-anhydro-2-*O*-(*tert*-butyldimethylsilyl)-4-*O*-sulfamoyl–β-D-mannopyranose **116** (39 mg, 0.083 mmol, 79%) as a white amorphous solid; $[\alpha]_D^{18}$ = -53, *c* = 1.5, $CHCl_3$; **^{1}H NMR** (400 MHz, $CDCl_3$) δ ppm 0.11, 0.12, 0.13, 0.13 (4 × s, 4 × 3 H, TBS), 0.92, 0.94 (2 × s, 2 × 9 H, TBS), 3.71 (dd, $J_{2,3}$ = 4.8 Hz, $J_{2,1}$ = 1.5 Hz, 1 H, H-2), 3.77 (dd, $J_{6b,6a}$ = 7.3 Hz, $J_{6b,5}$ = 5.9 Hz, 1 H, H-6b), 4.13 (dq, $J_{3,2}$ = 4.8 Hz, $J_{3,4}$ = 1.9 Hz, $^4J_{3,5}$ = 1.8 Hz, $^4J_{3,1}$ = 1.4 Hz, 1 H, H-3), 4.34 (dd, $J_{6a,6b}$ = 7.3 Hz, $J_{6a,5}$ = 1.0 Hz, 1 H, H-6a), 4.53 (td, $J_{4,3}$ = 1.9 Hz, $J_{4,5}$ = 1.0 Hz, 1 H, H-4), 4.69 (m, 1 H, H-5), 5.04 (s, 2 H, -NH_2), 5.31 (dd, $J_{1,2}$ = 1.5 Hz, $^4J_{1,3}$ = 1.4 Hz, 1 H, H-1); **^{13}C NMR** (100 MHz, CDCl) δ ppm -5.0, -4.8, -4.8, -4.0 (4 × 1 C, TBS), 18.0, 18.4 (2 × 1 C, 4°$_{TBS}$), 25.8, 26.1 (2 × 3 C, TBS), 64.8 (C-6), 68.9 (C-2), 71.2 (C-3), 73.7 (C-5), 81.2 (C-4), 102.5 (C-1); **IR** (thin film) *v* : 3290 (br, N-H), 3271 (br, N-H), 2955 (s), 2930 (s), 2897 (s), 2858 (s), 1563 (m), 1472 (m), 1363 (s, -SO_2NH_2) cm^{-1}; **MS** *m/z* (ESI^-) : 468 [(M-H)$^-$, 80%], 582 [(M+TFA-H)$^-$, 100%]; (ESI^+) : 487 [$(M+NH_4)^+$, 80%], 492 [$(M+Na)^+$, 100%], 533 [$(M+MeCN+Na)^+$, 90%]; **HRMS** *m/z* (ESI^-) : calc. for $C_{18}H_{38}NO_7SSi_2$ $(M - H)^-$ = 468.1913. Found 468.1910;

3-Amino-1,6-anhydro-2,3-*O*-isopropylidene-β-D-mannopyranose 3,4-sulfamidate 119



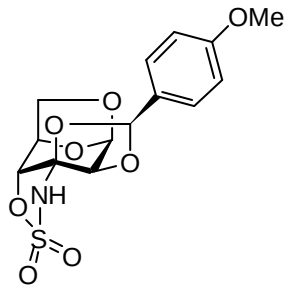
Powdered magnesium oxide (9.0 mg, 0.213 mmol) was flame dried under reduced pressure (1 mbar). 1,6-Anhydro-2,3-*O*-isopropylidene-4-*O*-sulfamoyl-β-D-mannopyranose **108** (26.0 mg, 0.093 mmol) was dissolved in dry THF (4 mL) and added *via* cannula. Iodobenzene diacetate (33 mg, 0.100 mmol) and Rh(II) acetate dimer (1.6 mg, 4 μmol, 0.04 eq) were added. After stirring the pale blue solution at room temperature for 5 h under N_2, TLC (acetone/toluene 1:4) showed spots at R_f 0.45 (product) and 0.3 (starting material). The reaction mixture was filtered through celite and washed with THF. Flash chromatography (acetone/toluene gradient: 1:9→1:4), afforded 3-amino-1,6-anhydro-2,3-*O*-isopropylidene-β-D-mannopyranose 3,4-sulfamidate **119** (14 mg, 0.050 mmol, 54%) as a white amorphous solid which degraded over time. Starting material was also recovered (12 mg, 0.042 mmol, 46%); $\mathbf{[\alpha]_D^{21}}$ = -57, *c* = 0.7, MeOH; **^{1}H NMR** (400 MHz, MeOD-*d*$_4$) δ ppm 1.47, 1.51 (2 × s, 2 × 3 H, 2 × -C*H*$_3$), 3.83 (dd, $J_{6a,6b}$ = 8.1 Hz, $J_{6a,5}$ = 6.3 Hz, 1 H, H-6a), 3.92 (dd, $J_{6b,6a}$ = 8.1 Hz, $J_{6b,5}$ = 1.3 Hz, 1 H, H-6b), 4.21 (d, $J_{2,1}$ = 3.5 Hz, 1 H, H-2), 4.74 (d, $J_{4,5}$ = 0.8 Hz, 1 H, H-4), 4.86 (ddd, $J_{5,6a}$ = 6.3 Hz, $J_{5,6b}$ = 1.3 Hz, $J_{5,4}$ = 0.8 Hz, 1 H, H-5), 5.46 (d, $J_{1,2}$ = 3.5 Hz, 1 H, H-1); **^{13}C NMR** (100 MHz, MeOD-*d*$_4$) δ ppm 26.1, 26.6 (2 × -C(*C*H$_3$)$_2$), 65.0 (C-6), 72.3 (C-5), 80.3 (C-2), 83.4 (C-4), 92.5 (C-3, quaternary C by lack of signal in HSQC and DEPT, coupled to H-1, H-2, H-4, H-5 by HMBC), 98.4 (C-1), 113.6 (-*C*(CH$_3$)$_2$); **IR** (thin film) *v* : 3262 (br, N-H), 2970 (m), 2300 (s), 2321 (m) cm^{-1}; **MS** *m/z* (ESI$^-$) : 278 [(M-H)$^-$]; **HRMS** *m/z* (ESI$^+$) : calc. for $C_9H_{15}NNaO_7S$ $(M + Na)^+$ = 304.0461. Found 304.0462;

3-Amino-1,6-anhydro-2,3-*endo*-*O*-(4-methoxybenzylidene)-β-D-mannopyranose 3,4-sulfamidate 120



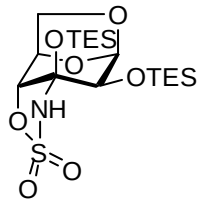
Powdered magnesium oxide (9.0 mg, 0.218 mmol) was flame dried under reduced pressure (1 mbar). 1,6-Anhydro-*endo*-2,3-*O*-(4-methoxybenzylidene)-4-*O*-sulfamoyl-β-D-mannopyranose **109** (34 mg, 0.094 mmol) was dissolved in dry 1,4-dioxane (470 μL) and added *via* syringe. Iodobenzene diacetate (33 mg, 0.103 mmol) and Rh(II) acetate dimer (2.0 mg, 4.7 μmol, 0.05 eq) were added. After stirring the pale blue solution under N_2 at room temperature for 1 h and 45 °C for a further 90 min, TLC (EtOAc/Petrol 1:1) showed a single product at R_f 0.5. The reaction mixture was filtered through celite and washed with 1,4-dioxane, and the filtrate concentrated *in vacuo*. Flash chromatography (EtOAc/Petrol 2:3), afforded 3-amino-1,6-anhydro-2,3-*endo*-*O*-(4-methoxybenzylidene)-β-D-mannopyranose 3,4-sulfamidate **120** (31 mg, 0.087 mmol, 92%) as a white foam which degraded over time; $\mathbf{[\alpha]_D^{21}}$ = -41, *c* = 0.5, MeOH; **^{1}H NMR** (400 MHz, DMSO-d^6) δ ppm 3.77 (s, 3 H, -OC*H*$_3$), 3.85 (dd, $J_{6b,6a}$ = 8.3 Hz, $J_{6b,5}$ = 6.3 Hz, 1 H, H-6b), 3.92 (dd, $J_{6a,6b}$ = 8.3 Hz, $J_{6a,5}$ = 1.5 Hz, 1 H, H-6a), 4.21 (d, $J_{2,1}$ = 3.8 Hz, 1 H, H-2), 4.95 (d, $J_{4,5}$ = 0.7 Hz, 1 H, H-4), 4.98 (ddd, $J_{5,6b}$ = 6.3 Hz, $J_{5,6a}$ = 1.5 Hz, $J_{5,4}$ = 0.7 Hz, 1 H, H-5), 5.75 (d, $J_{1,2}$ = 3.8 Hz, 1 H, H-1), 5.80 (s, 1 H, C*H*$_{benzyllic}$), 6.98 (d, *J* = 8.7 Hz, 2 H, 2 × 1H, 2 × C*H*$_{Ar}$), 7.54 (d, *J* = 8.7 Hz, 2 H, 2 × 1H, 2 × C*H*$_{Ar}$), 9.80 (br. s., 1 H, -N*H*); **^{13}C NMR** (100 MHz, DMSO-d^6) δ ppm 55.2 (-OC*H*$_3$), 64.8 (C-6), 71.3 (C-5), 77.4 (C-2), 82.4 (C-4), 91.5 (C-3, quaternary C by absence from DEPT, HSQC), 97.1 (C-1), 102.9 (*C*H$_{benzyllic}$), 113.7 (2 C, 2 × *C*H$_{Ar}$), 127.1 (4°$_{Ar}$), 129.3 (2 C, 2 × *C*H$_{Ar}$), 160.7 (4°$_{Ar}$); **IR** (thin film) *v* : 3280 (br, N-H), 2965 (s), 2933 (s), 2890 (s), 1713 (s, C=C^{Ar}); **MS** *m/z* (ESI$^-$) : 356 [(M-H)$^-$, 100%]; **HRMS** *m/z* (ESI$^-$) : calc. for $C_{14}H_{14}NO_8S$ (M - H)$^-$ = 356.0440. Found 356.0442;

Exo-enriched 3-amino-1,6-anhydro-2,3-_O_-(4-methoxybenzylidene)-β-D-mannopyranose 3,4-sulfamidate 121



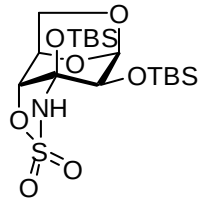
Powdered magnesium oxide (14.0 mg, 0.346 mmol) was flame dried under reduced pressure (1 mbar). An *exo*-enriched isomeric mixture (*exo*/*endo* 2:1) of 1,6-anhydro-2,3-*O*-(4-methoxybenzylidene)-4-*O*-sulfamoyl-β-D-mannopyranose (54 mg, 0.150 mmol) was dissolved in dry 1,4-dioxane (750 µL) and added *via* syringe. Iodobenzene diacetate (53 mg, 0.165 mmol) and Rh(II) acetate dimer (3.3 mg, 7.5 µmol, 0.05 eq) were added. After stirring the pale blue solution under N_2 at room temperature for 3 h and 45 ºC for a further 2 h, TLC (EtOAc/Petrol 3:2) showed a single product at R_f 0.6 (product). The reaction mixture was filtered through celite and washed with 1,4-dioxane, and the filtrate concentrated *in vacuo*. Flash chromatography (EtOAc/Petrol 1:1), afforded *exo*-enriched 3-amino-1,6-anhydro-2,3-*O*-(4-methoxybenzylidene)-β-D-mannopyranose 3,4-sulfamidate (24 mg, 0.067 mmol, 49%) as a white foam (*exo*/*endo* 2:1); **^{1}H NMR** (400 MHz, MeOD-d_4) δ ppm 3.81 (s, 3 H, -OCH_3), 3.89 (dd, $J_{6b,6a}$ = 8.2 Hz, $J_{6b,5}$ = 4.3 Hz, 1 H, H-6b), 4.10 (dd, $J_{6b,6a}$ = 8.2 Hz, $J_{6b,5}$ = 0.9 Hz, 1 H, H-6a), 4.46 (d, $J_{2,1}$ = 3.0 Hz, 2 H, H-2), 4.96 (m, 3 H, H-4, H-5), 5.65 (d, $J_{1,2}$ = 33.8 Hz, 1 H, H-1), 6.29 (s, 1 H, C$H_{\text{benzyllic}}$), 6.94 (d, J=8.8 Hz, 2 H, 2 × 1H, 2 × CH_{Ar}), 7.42 (d, J=8.8 Hz, 3 H, 2 × 1H, 2 × CH_{Ar}), 7.97 (s, 1 H, -N*H*), shifts were assigned from a spectrum of an inseperable mixture of isomers, enriched in the *exo* isomer; **^{13}C NMR** (100 MHz, MeOD-d_4) δ ppm 54.8 (-O*C*H$_3$), 65.6 (C-6), 72.4 (C-5), 81.3 (C-2), 83.4 (C-4), 92.1 (C-3, quaternary C by absence from DEPT, HSQC), 99.7 (C-1), 108.1 (*C*H$_{\text{benzyllic}}$), 113.7 (2 C, 2 × *C*H$_{\text{Ar}}$), 128.6 (2 C, 2 × *C*H$_{\text{Ar}}$), 129.6, 163.9 (2 × 1 C, 2 × 4°$_{\text{Ar}}$), shifts were assigned from a spectrum of an inseperable mixture of isomers, enriched in the *exo* isomer; **MS** *m/z* (ESI$^-$) : 356 [(M-H)$^-$, 100%];

3-Amino-1,6-anhydro-2,3-bis-*O*-(triethylsilyl)-β-D-mannopyranose 3,4-sulfamidate 122



1,6-Anhydro-2,3-bis-*O*-(triethylsilyl)-4-*O*-sulfamoyl–β-D-mannopyranose **115** (36 mg, 0.077 mmol) was dissolved in dry 1,4-dioxane (390 μL). Powdered magnesium oxide was flame dried under reduced pressure (1 mbar) and a sample (7.1 mg, 0.176 mmol) added to the solution. Iodobenzene diacetate (27 mg, 0.085 mmol) and Rh(II) acetate dimer (1.7 mg, 3.9 μmol, 0.05 eq) were added. After stirring the solution under Ar at room temperature for 2 h, the reaction was heated to 45 °C for 3 h, upon which TLC (EtOAc/Petrol 3:7) showed a product at R_f 0.5, alongside remaining starting material at R_f 0.3. The reaction mixture was concentrated *in vacuo*. Flash chromatography (EtOAc/Petrol gradient: 1:9→3:7) afforded 3-amino-1,6-anhydro-2,3-bis-*O*-(triethylsilyl)-β-D-mannopyranose 3,4-sulfamidate **122** (7 mg, 0.015 mmol, 19%) as a white amorphous solid which degraded rapidly over time. Starting material **115** was also recovered (5 mg, 0.011 mmol, 14%); $\mathbf{[\alpha]_D^{22}}$ = -42, *c* = 0.5, $CHCl_3$; **^{1}H NMR** (400 MHz, $CDCl_3$) δ ppm 0.71 (m, 12 H, TES), 1.00 (m, 18 H, TES), 3.88 (dd, $J_{6b,6a}$ = 7.6 Hz, $J_{6b,5}$ = 5.9 Hz, 1 H, H-6b), 3.91 (d, $J_{2,1}$ = 1.5 Hz, 1 H, H-2), 4.14 (dd, $J_{6a,6b}$ = 7.6 Hz, $J_{6a,5}$ = 1.0 Hz, 1 H, H-6a), 4.50 (d, $J_{4,5}$ = 1.8 Hz, 1 H, H-4), 4.80 (m, 2 H, H-5, N-H), 5.34 (d, $J_{1,2}$ = 1.5 Hz, 1 H, H-1); **^{13}C NMR** (100 MHz, $CDCl_3$) δ ppm 4.9 (3 C, TES), 5.9 (3 C, TES), 6.7 (3 C, TES), 6.9 (3 C, TES), 65.3 (C-6), 71.6 (C-5), 75.5 (C-2), 85.5 (C-4), 89.7 (C-3, quaternary C by absence from DEPT, HSQC), 101.5 (C-1); **IR** (thin film) *v* : 3258 (br, N-H), 2948 (s), 2929 (s), 2891 (s), 2867 (s), 1470 (m); **MS** *m/z* (ESI^-) : 466 [$(M-H)^-$, 100%]; **HRMS** *m/z* (ESI^-) : calc. for $C_{18}H3_6NO_7SSi$ $(M - H)^-$ = 466.1751. Found 466.1749;

3-Amino-1,6-anhydro-2,3-bis-*O*-(*tert*-butyldimethylsilyl)-β-D-mannopyranose 3,4-sulfamidate 123



Powdered magnesium oxide (4.5 mg, 0.113 mmol) was flame dried under reduced pressure (1 mbar). 1,6-Anhydro-2,3-bis-*O*-(*tert*-butyldimethylsilyl)-4-*O*-sulfamoyl–β-D-mannopyranose **116** (23 mg, 0.049 mmol) was dissolved in dry DCM (245 μL) and added *via* syringe. Iodobenzene diacetate (17 mg, 0.054 mmol) and Rh(II) acetate dimer (1.1 mg, 2.5 μmol, 0.05 eq) were added. After stirring the solution under Ar at room temperature for 4 h, TLC (EtOAc/Petrol 3:7) showed a single product at R_f 0.5. The reaction mixture was concentrated *in vacuo*. Flash chromatography (EtOAc/Petrol gradient: 3:14→3:7) afforded

3-amino-1,6-anhydro-2,3-bis-*O*-(*tert*-butyldimethylsilyl)-β-D-mannopyranose 3,4-sulfamidate **123** (14 mg, 0.03 mmol, 61%) as a clear amorphous solid which degraded over time; $[\alpha]_D^{22}$ = -63, *c* = 0.5, $CHCl_3$; **^{1}H NMR** (400 MHz, $CDCl_3$) δ ppm 0.18, 0.19, 0.22, 0.26 (4 × s, 4 × 3 H, TBS), 0.92, 0.96 (2 × s, 2 × 9 H, TBS), 3.88 (dd, $J_{6b,6a}$ = 7.7 Hz, $J_{6b,5}$ = 5.7 Hz, 1 H, H-6b), 3.93 (d, $J_{2,1}$ = 1.4 Hz, 1 H, H-2), 4.18 (dd, $J_{6a,6b}$ = 7.7, $J_{6a,5}$ = 0.8 Hz, 1 H, H-6a), 4.50 (d, $J_{4,5}$ = 1.5 Hz, 1 H, H-4), 4.81 (ddd, $J_{5,6b}$ = 5.7 Hz, $J_{5,6a}$ = 0.8 Hz, $J_{5,4}$ = 1.5 Hz, 1 H, H-5), 4.85 (s, 1 H, N-H), 5.35 (d, $J_{1,2}$ = 1.4 Hz, 1 H, H-1); **^{13}C NMR** (100 MHz, $CDCl_3$) δ ppm -4.5, -3.9, -3.3, -2.8 (4 × 1 C, TBS), 18.1, 18.3 (2 × 1 C, 4°$_{TBS}$), 25.7, 26.0 (2 × 3 C, TBS), 65.3 (C-6), 71.7 (C-5), 76.1 (C-2), 84.8 (C-4), 89.8 (C-3, quaternary C by absence from DEPT, HSQC), 101.6 (C-1); **IR** (thin film) *v* : 3273 (br, N-H), 2957 (s), 2933 (s), 2899 (s), 2856 (s), 1560 (m), 1465 (m) cm^{-1}; **MS** *m/z* (ESI^-) : 466 [$(M-H)^-$, 100%]; **HRMS** *m/z* (ESI^-) : calc. for $C_{18}H_{36}NO_7SSi_2$ $(M - H)^-$ = 466.1756. Found 466.1745;

5.5 References

406. Xu L, Appell M, Kennedy S, Momany Frank A, Price Neil PJ (2004): **Conformational analysis of chirally deuterated tunicamycin as an active site probe of UDP-N-acetylhexosamine:polyprenol-P N-acetylhexosamine-1-P translocases**. *Biochemistry* **43**: 13248-13255.

407. Brandish PE, Kimura K, Inukai M, Southgate R, Lonsdale JT, Bugg TDH (1996): **Modes of action of tunicamycin, liposidomycin B, and mureidomycin A: Inhibition of phospho-N-acetylmuramyl-pentapeptide translocase from *Escherichia coli***. *Antimicrob. Agents Chemother.* **40**: 1640-1644.

408. Banoub J, Boullanger P, Lafont D (1992): **Synthesis of oligosaccharides of 2-amino-2-deoxy sugars**. *Chem. Rev.* **92**: 1167-1195.

409. Debenham J, Rodebaugh R, B. F-R (1997): **Recent Advances in N-Protection for Amino Sugar Synthesis**. *Liebigs Ann.* **1997**: 791-802.

410. Prices were obtained from Sigma-Aldrich and were accurate as of July 2010: *N*-acetyl-D-glucosamine (£10/g), *N*-acetyl-D-galactosamine (£118/g).

411. Brendel K, Gross PH, Zimmerma.Hk (1965): **Aminozuckersynthesen .13. Zum Mechanismus Des Nucleophilen Ersatzes Von Methansulfonylgruppen - Die Gleichzeitige Bildung Von D-Gulosamin- Und D-Galaktosamin-Derivatenausglucosamin-Derivaten**. *Justus Liebigs Ann. Chem.* **683**: 182-186.

412. Agoston K, Kerekgyarto J, Hajko J, Batta G, Lefeber DJ, Kamerling JP, Vliegenthart JFG (2002): **Synthesis of fragments of the glycocalyx glycan of the parasite *Schistosoma mansoni***. *Chem. Eur. J.* **8**: 151-161.

413. Gross PH, Jeanloz RW (1967): **Optically Pure *N*-Substituted Derivatives of Benzyl 2-Amino-2-Deoxy-α- and -β-D-Glucopyranoside**. *J. Org. Chem.* **32**: 2759-2763.

414. Hill J, Hough L (1968): **Nucleophilic Replacement Reactions of Sulphonates. 4. Transformation of 2-Amino-2-Deoxy-D-Glucose into Derivatives of 2-Amino-2-Deoxy-D-Galactose and 2-Amino2,4,6-Trideoxy-D-*Xylo*-Hexose**. *Carbohydr. Res.* **8**: 398-404.

415. Horner MW, Hough L, Richards.Ac (1970): **Selective Benzoylation of Methyl 2-Benzamido-2-Deoxy-Alpha-D-Glucopyranoside - a Convenient Preparation of Derivatives of 2-Amino-2-Deoxy-D-Galactose**. *J. Chem. Soc. C*: 1336-1340.

416. Ito N, Tokuda Y, Ohba S, Sugai T (2004): **An expeditious route to N-acetyl-D-galactosamine from N-acetyl-D-glucosamine based on the selective protection of hydroxy groups**. *Bull. Chem. Soc. Jpn.* **77**: 1181-1186.

417. McGeary RP, Wright K, Toth I (2001): **Conversion of glucosamine to galactosamine and allosamine derivatives: Control of inversions of stereochemistry at C-3 and C-4**. *J. Org. Chem.* **66**: 5102-5105.

418. Hederos M, Konradsson P (2005): **Efficient routes to ethyl-2-deoxy-2-phthalimido-1-beta-D-thio-galactosamine derivatives via epimerization of the corresponding glucosamine compounds**. *J. Carbohydr. Chem.* **24**: 297-320.

419. Lemieux RU, Ratcliffe RM (1979): **Azidonitration of Tri-O-Acetyl-D-Galactal**. *Can. J. Chem.* **57**: 1244-1251.

420. Griffith DA, Danishefsky SJ (1990): **Sulfonamidoglycosylation of Glycals - a Route to Oligosaccharides with 2-Aminohexose Subunits**. *J. Am. Chem. Soc.* **112**: 5811-5819.

421. Lemieux RU, Nagabhus.Tl (1968): **Synthesis of 2-Amino-2-Deoxyhexoses - D-Glucosamine D-Mannosamine D-Galactosamine and D-Talosamine**. *Can. J. Chem.* **46**: 401-403.

422. Weng M, Geyer A, Friemel A, Jochims JC, Lutz M (2000): **Cycloaddition reactions of 1,3-diaza-2-azoniaallene salts and glycals**. *J. Prakt. Chem. (Weinheim, Ger.)* **342**: 486-493.

423. Winterfeld GA, Schmidt RR (2001): **Nitroglycal concatenation: A broadly applicable and efficient approach to the synthesis of complex O-glycans**. *Angew. Chem., Int. Ed.* **40**: 2654-2657.

424. Kozlowska-Gramsz E, Descotes G (1981): **Addition of Nitrenes to Cyclic Vinylethers - Application to the Synthesis of Amino-Sugars**. *Tetrahedron Lett.* **22**: 563-566.

425. James SP, Smith F, Stacey M, Wiggins LF (1946): **The Action of Alkaline Reagents on 2:3-1:6- and 3:4-1:6-Dianhydro-β-Talose - a Constitutional Synthesis of Chondrosamine and Other Amino-Sugar Derivatives**. *J. Chem. Soc.*: 625-628.

426. Richard K, Werner K (1956): **2-Amino-2-desoxy-zucker durch katalytische Halbhydrierung von Amino-, Arylamino- und Benzylamino-nitrilen; D- und L-Glucosamin. Aminozucker-synthesen II**. *Justus Liebigs Ann. Chem.* **600**: 115-125.

427. Wrodnigg TM, Lundt I, Stutz AE (2006): **Synthesis of N-protected galactosamine building blocks from D-tagatose via the Heyns rearrangement**. *J. Carbohydr. Chem.* **25**: 33-41.

428. Espino CG, Wehn PM, Chow J, Du Bois J (2001): **Synthesis of 1,3-Difunctionalized Amine Derivatives through Selective C-H Bond Oxidation**. *J. Am. Chem. Soc.* **123**: 6935-6936.

429. Wehn PM, Lee J, DuBois J (2003): **Stereochemical Models for Rh-Catalyzed Amination Reactions of Chiral Sulfamates**. *Org. Lett.* **5**: 4823-4826.

430. Fleming JJ, Fiori KW, Du Bois J (2003): **Novel Iminium Ion Equivalents Prepared through C-H Oxidation for the Stereocontrolled Synthesis of Functionalized Propargylic Amine Derivatives**. *J. Am. Chem. Soc.* **125**: 2028-2029.

431. Eades EDM, Ball DH, Long L (1965): **Methyl 2,6-Anhydro-3,4-di-*O*-α-D-mannoside. An Intramolecular Nucleophilic Displacement of Mesylate with Inversion**. *J. Org. Chem.* **30**: 3949-3951.

432. Wolfrom ML, Chakravarty P, Horton D (1966): **Amino Derivatives of Starches. 2-Amino-3,6-anhydro-2-deoxy-D-mannose**. *J. Org. Chem.* **31**: 2502-2504.

433. Hiranuma S, Kanie O, Wong C-H (1999): **Formation of the 1,2,6-orthoester of mannose and its utilization in the glycosylation reaction**. *Tetrahedron Lett.* **40**: 6423-6426.

434. Georges M, Fraser-Reid B (1984): **A simple, one-flask, two-step synthesis of 1,6-anhydro-β-D-mannopyranose (D-mannosan) from D-mannose**. *Carbohydr. Res.* **127**: 162.

435. Florent J-C, Monneret C (1980): **Ouverture par le N-bromosuccinimide des acétals benzylidéniques de méthyl-hexosides de configuration manno et synthèse de la N-acétyl--mycosamine**. *Carbohydr. Res.* **85**: 243-257.

436. Fiori KW, Fleming JJ, Du Bois J (2004): **Rh-Catalyzed Amination of Ethereal C^{α}-H Bonds: A Versatile Strategy for the Synthesis of Complex Amines**. *Angew. Chem., Int. Ed.* **43**: 4349-4352.

437. Schwarz S, Thieme I, Richter M, Undeutsch B, Henkel H, Elger W (1996): **Synthesis of estrogen sulfamates: Compounds with a novel endocrinological profile**. *Steroids* **61**: 710-717.

438. Okada M, Iwashita S, Koizumi N (2000): **Efficient general method for sulfamoylation of a hydroxyl group**. *Tetrahedron Lett.* **41**: 7047-7051.

439. Greene TW, Wuts PGM (1999): *Protective Groups in Organic Synthesis* 3rd edn. John Wiley & Sons, Inc.: Chichester.

440. Kocienski PJ (2005): *Protecting Groups* 3rd (revised) edn. Thieme Publishing Group: Stuttgart.

441. Mercury 1.4.2 (Build 2) - http://www.ccdc.cam.ac.uk/mercury/.

442. Toumieux S, Compain P, Martin OR (2005): **Intramolecular metal-catalyzed amination of pseudo-anomeric C-H bonds**. *Tetrahedron Lett.* **46**: 4731-4735.

443. Toumieux S, Compain P, Martin OR, Selkti M (2006): **New Aspects of Catalytic Intramolecular C-H Amination: Unexpected Formation of a Seven-Membered Ring in Nitrogen-Containing Systems**. *Org. Lett.* **8**: 4493-4496.

444. Wyszynski FJ, Thompson AR, Davis BG (2010): **Inverted regioselectivity of C–H amination: Unexpected oxidation at β- rather than γ-C–H**. *Org. Biomol. Chem.* **8**: 581-589.

445. Manna S, McAnalley BH, Ammon HL (1993): **2,3,4-Tri-*O*-acetyl-1,6-anhydro-β-D-mannopyranose, an artifact produced during carbohydrate analysis. A total synthesis of 2,3,5-tri-*O*-acetyl-1,6-anhydro-β-D-mannofuranose**. *Carbohydr. Res.* **243**: 11-27.

446. Furneaux RH, Shafizadeh F (1979): **Pyrolytic production of 1,6-anhydro-β-D-mannopyranose**. *Carbohydr. Res.* **74**: 354-360.

447. de la Paz ML, Ellis G, Pérez M, Perkins J, Jiménez-Barbero J, Vicent C (2002): **Carbohydrate Hydrogen-Bonding Cooperativity - Intramolecular Hydrogen Bonds and Their Cooperative Effect on Intermolecular Processes - Binding to a Hydrogen-Bond Acceptor Molecule**. *Eur. J. Org. Chem.* **2002**: 840-855.

448. Kloosterman M, de Nijs MP, van Boom JH (1984): **A new approach to the synthesis of 1,6-anhydro-2-azido-2-deoxy-β-D-glucopyranose derivatives**. *Recl.: J. R. Neth. Chem. Soc.* **103**: 243-246.

Chapter 6: Summary

6.1 Original hypotheses and objectives

In the introduction to this thesis we set out the hypothesis that rationally designed analogues of tunicamycin could act as nucleotide-sugar substrate or even bi-substrate transition state mimics of particular NDP-sugar-dependent glycosyltransferases, thus acting as potent and specific inhibitors (section 1.5). Bacterial peptidoglycan biosynthesis enzyme MraY and mammalian GPT from the *N*-linked protein glycosylation pathway are both specifically targeted by native tunicamycins, but potent inhibition of the latter has precluded the clinical use of this nucleoside antibiotic as an antibacterial drug. In the first instance it was proposed that modifying the terminal portions of the tunicamycin structure could improve its selectivity for MraY over GPT, thus potentially yielding a novel antibiotic compound exhibiting minimal cytotoxicity and with a mode of action orthogonal to that of existing drugs. As an extension of this theory, further modifications to the tunicamycin scaffold were proposed which would yield compounds potentially inhibiting a much wider range of medicinally important glycosyltransferases.

In pursuit of this long term objective and in order to begin to test this hypothesis, a variety of routes to tunicamycin analogues were devised, incorporating a combination of relay, synthetic and chemoenzymatic strategies (section 1.5.2). Dissection of the tunicamycin biosynthetic pathway was highlighted as a particularly valuable facet of investigation, offering a number of advantages over adopting a purely synthetic route to functional analogues. First of all, tunicamycin is not scarce and can be isolated in preparative quantities by isolation from fermentation broths of *S. chartreusis*. A relay synthesis together with the use of biosynthetic enzymes for the chemoenzymatic diversification of advanced intermediates offered the most efficient route to tunicamycin analogues. Additionally, since the biosynthesis of tunicamycin – and that of nucleoside antibiotics in general – was poorly understood, targeting its elucidation would provide the greatest scope for added value and academic insight beyond the long term objective of producing rationally designed tunicamycin analogues.

6.2 Chapter two

In chapter 2 we initiated the elucidation of the tunicamycin biosynthetic pathway by whole genome sequencing of tunicamycin producer *Streptomyces chartreusis* NRRL 3882. Subsequent *in silico* genome mining identified a fourteen-gene putative biosynthetic cluster (*tun*). Molecular cloning and heterologous expression of this *tun* gene cluster in *S. coelicolor* hosts unambiguously defined the minimal set of genes required for tunicamycin production. The amino acid sequences encoded by individual *tun* genes were also subjected to in-depth bioinformatic analysis, allowing the proposal of a detailed putative metabolic pathway to tunicamycin production from common cellular precursors (Fig. 6.1).

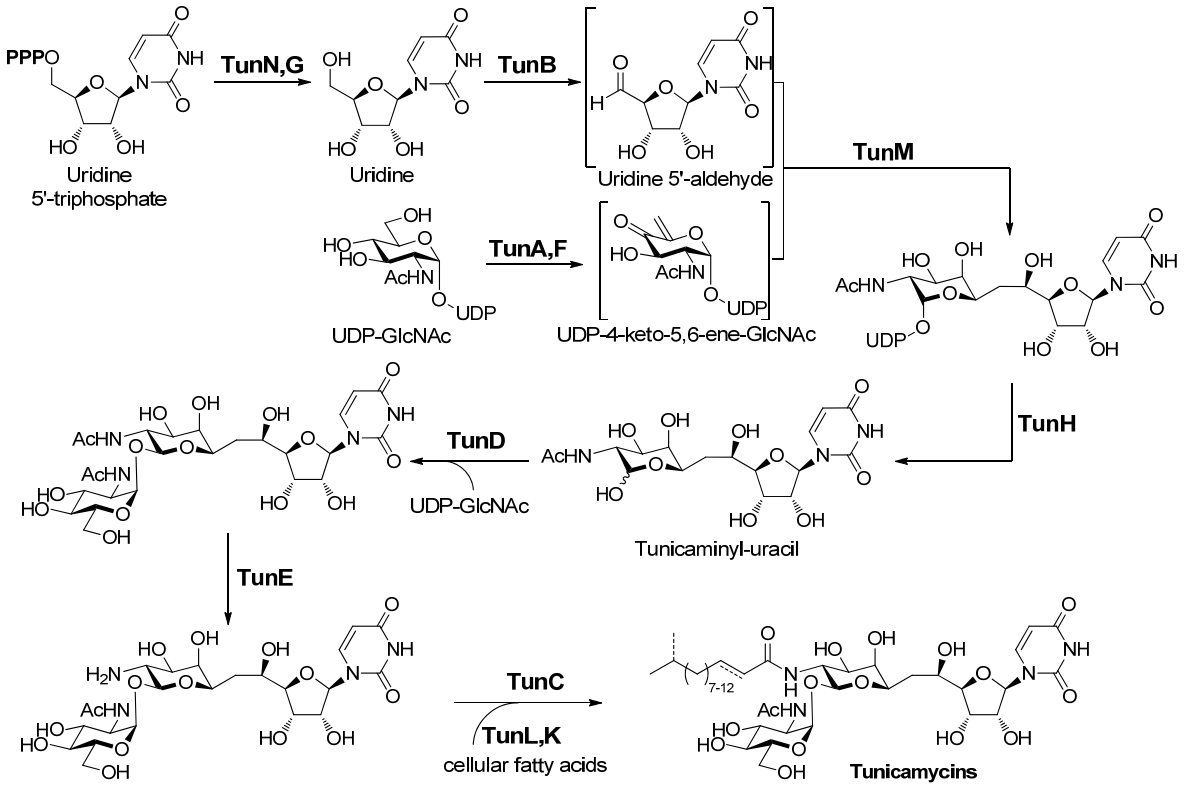


Fig. 6.1 Proposed biosynthetic pathway for the tunicamycins

Homologous gene clusters were identified within the publicly available genomes of *Actinosynnema mirum* DSM 43287 and *Streptomyces clavuligerus* ATCC 27064; the latter is known to produce an antibiotic closely related to tunicamycin – MM19290.[449] We thus claimed to have identified the biosynthetic gene cluster responsible for MM19290 biosynthesis in *S. clavuligerus* ATCC 27064 and furthermore proposed the core structure of this compound is identical to that of the tunicamycins, differing only in the nature of its acyl side chains.

The genome scanning approach used to identify the *tun* gene cluster in these studies represented a rapid and efficient way of identifying natural product gene clusters, without the need for laborious biochemical probes and avoiding frustrating false-positive results. It utilised rationalised predictions of the chemical reactivity that is likely to be involved in the construction of tunicamycin from metabolic precursors, acting as a 'logic filter' that was applied during the genome mining process. The exponential increase in publically available gene sequences in recent years has dramatically expanded the possibilities afforded by bioinformatic analysis and the work presented in chapter 2 highlights *in silico* mining of partially assembled genome sequences as an increasingly effective tool during the early stages of dissecting a bacterial biosynthetic pathway.

In all, through a rapid and efficient whole genome-scanning approach we successfully identified and verified the biosynthetic genes of a tunicamycin-family antibiotic for the first time, offering valuable insights into the poorly understood biosynthetic pathway of this fascinating family of nucleoside antibiotics.[450]

6.3 Chapter three

In order to facilitate future studies towards the formation of tunicamycin analogue libraries, access to sufficient quantities of pure tunicamycin was required. Considering tunicamycin is only sold in milligram quantities by commercial suppliers and at an extremely high cost, in chapter 3 we detailed investigations into the detection, isolation and purification of preparative quantities of tunicamycin from fermentation broths of *S. chartreusis*. Following optimisation, the yields of pure tunicamycin obtained approached 35 mg/L culture, establishing an efficient supply route sufficient for any subsequent investigations.

These isolated tunicamycins were next subjected to degradation studies, targeting advanced biosynthetic intermediates containing the tunicamine scaffold for use both as precursors in the synthesis of tunicamycin analogues and as substrates for the *in vitro* characterisation of individual Tun enzymes. Efficient routes to compounds *N*-acetyl-tunicaminyl uracil **90** and *N*-acetyl-*N*-deacyl-tunicamycin **99** were developed, utilising a combination of acidic hydrolysis, mild methanolysis and effective protecting group strategies (Fig. 6.2). Both of these compounds were extremely valuable since they represented key intermediates for subsequent functionalisation into diverse libraries of tunicamycin analogues. Not only this, but they were also proposed to act as natural substrates for glycosyltransferase TunD and *N*-deacetylase TunE from the tunicamycin

biosynthetic pathway and it was hoped they would be useful as part of the *in vitro* functional characterisation of these enzymes – studies presented in chapter 4.

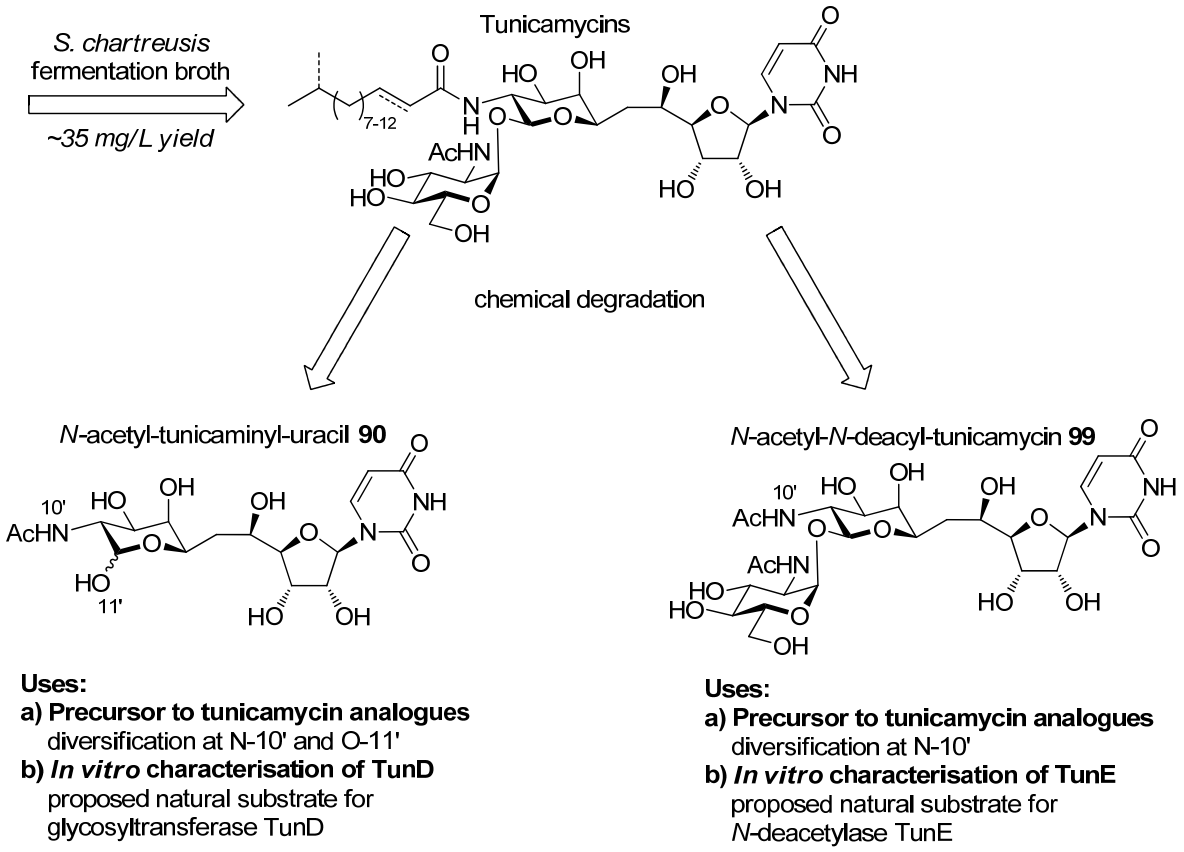


Fig. 6.2 Summary and significance of tunicamycin degradation studies from chapter three

Overall, results presented in chapter 3 complement both the characterisation of the tunicamycin biosynthetic pathway described in chapters 2 and 4, and studies towards the original objectives of forming tunicamycin analogues based around the core tunicamine scaffold.

6.4 Chapter four

The studies presented in chapter 4 sought to complement the elucidation of the tunicamycin biosynthetic gene cluster by lending weight to the proposed metabolic pathway of tunicamycin biosynthesis through the functional characterisation of individual *tun* gene products. Work was focussed on three particular gene products: epimerase TunF, glycosyltransferase TunD and *N*-deacetylase TunE. Protocols for their over-expression and purification from recombinant *E. coli* hosts were optimised, allowing subsequent *in vitro* studies to take place which have offered initial insights into the functions of these enzymes.

Epimerase TunF was successfully cloned and overexpressed as an N-terminal His_{10} fusion protein and was experimentally shown to function as a UDP-GlcNAc-4-epimerase. A novel enzymatic assay based on ^{1}H NMR was developed in the process, offering both a simple and rapid method for detecting enzymatic activity and allowing accurate elucidation of product structures directly from reaction mixtures, eliminating the need for laborious chromatographic separation and purification of reaction products. The observed epimerase activity was in line with previously reported predictions and those made in chapter 2, where UDP-GlcNAc rather than UDP-GalNAc was proposed to act as a metabolic precursor in tunicamine biosynthesis.[451]

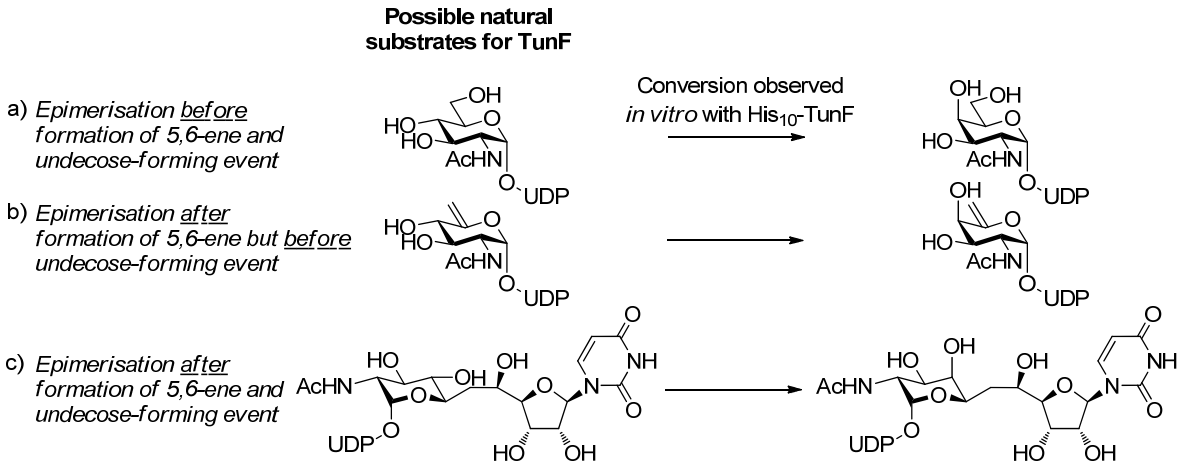
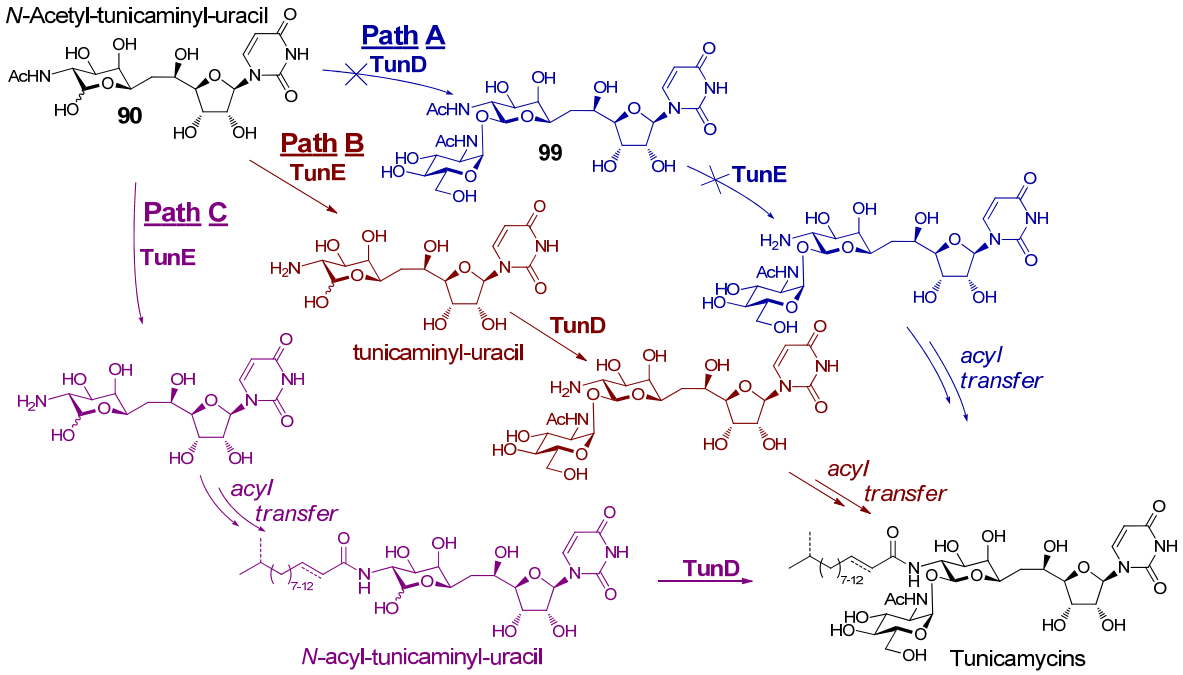


Fig. 6.3 Possible substrates for epimerase TunF

The functional characterisation of TunF established the importance of its role in the installation of the pseudo-GalNAc moiety within tunicamine, acting to epimerise a *gluco*-configured precursor at some stage during the biosynthetic pathway. Unfortunately, despite demonstrating activity against UDP-GlcNAc the precise identity of TunF's true *in vivo* substrate remained unclear. We suggested that TunF could act at a number of stages in the

biosynthetic pathway. Depending on whether it acts before, during or after the undecose C–C bond-forming event a different UDP-*N*-acetylglucosamine derivative was rationalised as a potential substrate in each case (Fig. 6.3).

Successful over-expression and purification of glycosyltransferase TunD experienced significant hurdles resulting from poor protein solubility and stability. Following the screening of a wide range of conditions for heterologous protein expression and the production of a number of fusion protein constructs, an optimised procedure for TunD expression was developed involving the production of a C-terminal His_6 fusion and subsequent purification by Ni^{2+}-affinity chromatography. Potential functional activity of this TunD-His_6 construct was analysed with a coupled spectrophotometric assay, using previously synthesised *N*-acetyl-tunicaminyl-uracil **90** as the glycosyl acceptor substrate. Biological activity was observed, although the results showed that the TunD-catalysed UDP release was not coupled with GlcNAc transfer to the glycosyl acceptor. This activity was thus attributed to enzyme-catalysed hydrolysis of UDP-GlcNAc, a reaction which has been previously observed with glycosyltransferases as a background process competing with the much more rapid glycosyl transfer.[452-455]



Scheme 6.1 Implications of observed TunD-His_6 activity on the tunicamycin biosynthetic pathway

The lack of GlcNAc transfer to the glycosyl acceptor was surprising and suggested that *N*-acetyl-tunicaminyl-uracil **90** was not the natural substrate for glycosyltransferase TunD (Scheme 6.1, Path A). A pair of alternative biosynthetic pathways were proposed,

both suggesting the reordering of *N*-deacetylation (TunE), glycosyl transfer (TunD) and *N*-acylation (TunC, L, K) events towards the end of the metabolic pathway (Scheme 4.1, Paths B and C). Two alternative substrates for glycosyltransferase TunD were suggested, both of which were derivatives of tunicaminyl-uracil; one had a free amine at N-10´ and the other was acylated with fatty acids at this position.

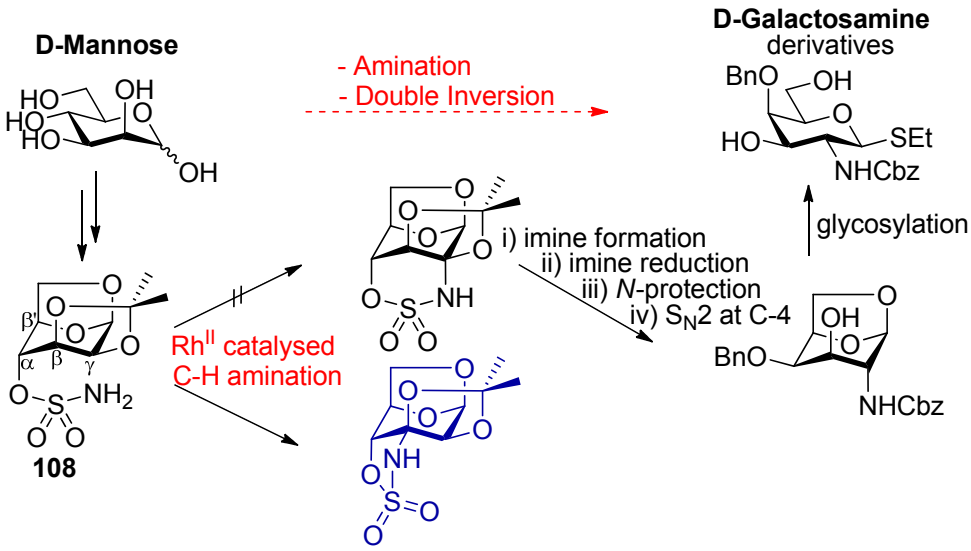
Initial investigations into the over-expression and purification of putative *N*-deacetylase TunE were also presented. Soluble TunE was obtained in low levels as an N-terminal His_6 fusion but unfortunately failed to show any biological activity in a HPLC assay utilising previously synthesised *N*-acetyl-*N*-deacyl-tunicamycin **99** as the substrate. The results suggested that this fusion was unstable in solution and based on the prediction that TunE was a metalloenzyme binding Zn^{2+} at its N-terminus, it was proposed that a C-terminal His-fusion of TunE purified by Zn^{2+}-affinity chromatography would provide a more appropriate construct for the detection of biological activity. In line with the alternative substrates proposed for glycosyltransferase TunD, the corresponding TunE substrates from these modified pathways were also presented (Scheme 6.1). Although only preliminary investigations into the expression, purification and functional verification of TunE have been presented herein, they act as a solid foundation for the future construction of alternative TunE constructs and functional enzyme assays.

Together, chapter 4 documented the functional verification of epimerase TunF, partial functional verification of glycosyltransferase TunE and significant optimisation of expression and purification protocols for TunF, E and D enzymes from the tunicamycin biosynthetic pathway. Previously unforeseen insights into the tunicamycin metabolic pathway were gained and these findings now act as a gateway, allowing rational design of future experiments and targeted investigations of specific areas of the biosynthetic pathway.

6.5 Chapter five

In chapter 5 we set out to synthesise derivatives of D-galactosamine, a key motif within the tunicamine scaffold. Synthesis of differentially protected derivatives of this sugar would not only allow future construction of unnatural tunicamine analogues – for example linked to non-uracil nucleobases – but it was also hoped to offer an improvement over existing methodologies which suffer from moderate yields and incomplete regio- and stereoselectivities. A Rh^{II}-catalysed C–H insertion strategy was envisaged, utilising a 1,6-anhydro-D-mannose precursor (Scheme 6.2). Unfortunately, the expected γ-C–H insertion at C–2 was not observed, with insertion occurring at the β-C–H of C–3 instead. An X-ray

crystal structure of sulfamate precursor **108** offered valuable insights as to the source of this unique selectivity. With respect to the original goals, the results obtained would have unfortunately led to 3-amino-3-deoxy-D-idose rather than the desired 2-amino-2-deoxy-D-galactose derivatives. However, these studies offered important broader implications. To the best of our knowledge, this was the first example of a sulfamate cyclisation forming a five membered ring in preference to a six membered ring based on conformation alone. The inverted regioselectivity exemplified by selective β-insertion has taken place in the presence of an equally activated γ-C–H, which would normally be greatly favoured. Although β-insertion has been reported with sulfamates, this has only been observed when no γ pathway is available, or if an electron-donating group is present at the β but not the γ position, providing a much more active C–H for insertion.[456]



Scheme 6.2 Unusual C-H insertion selectivity observed en route to D-galactosamine derivatives

In conclusion, the results obtained in chapter 5 represent somewhat of a digression from the original goals and the overall objectives of the thesis, but they offer an interesting and unique observation which will further our understanding of the influencing factors and mechanistic aspects involved in C-H activation and nitrene insertion methodologies.[457]

6.6 Future directions

Work is currently ongoing in our laboratory to further probe the tunicamycin biosynthetic pathway. The primary goal is to continue with *in vitro* functional characterisation of individual Tun enzymes, supplementing the insights gained both from the detailed bioinformatic analysis of the *tun* gene cluster and the experimental characterisation of recombinant TunF, TunD and TunE proteins. Aside from deciphering the true order of biosynthetic events and assigning true natural substrates for key enzymes, further biochemical investigations will also unravel the enzymology and substrate specificity of these enzymes, many of which are from poorly studied or potentially novel enzyme families. Further attempts to crystallise Tun proteins would support all of these goals by potentially providing a wealth of structural information. In all, these *in vitro* investigations will not only expand our understanding of the tunicamycin biosynthetic pathway, but they will also extend the toolbox of biosynthetic machinery available for the construction of tunicamycin analogues.

One particular target remains TunF, where further work is required in order to determine the true biosynthetic substrate. Any studies will be complemented by recent functional characterisation of TunA in our research group.[458] The dehydratase was found to accept UDP-GlcNAc – but not UDP-GalNAc – as a substrate, catalysing conversion to UDP-5,6-ene-GlcNAc. This product was further found to act as a substrate for TunF, where formation of a UDP-5,6-ene-GalNAc product was detected. In order to probe the precise ordering of these early biosynthetic transformations, access to proposed alternative substrates UDP-GlcNAc-5,6-ene and UDP-8´-*epi*-tunicamine may be required, although synthesis of these compounds is likely to be challenging. Studies into the precise role of TunF should also be complemented by acquiring a deeper understanding of the carbohydrate tail-to-tail coupling event in the formation of undecose tunicamine. The likely involvement of TunA, TunB and TunM should be investigated experimentally which will help explain the formation of such a unique C–C linkage.

Further functional characterisation of glycosyltransferase TunD and *N*-deacetylase TunE is also a key target for future work. Further optimisation of over-expression and purification procedures may be required, but the subsequent testing of TunD and TunE activity using the previously proposed alternative substrates will offer meaningful insights into the ordering of biosynthetic events. Once more is known about the enzymology and substrate specificity of these key biosynthetic enzymes, their application to the

chemoenzymatic synthesis of tunicamycin analogues can be tested experimentally, incorporating an informed choice of substrates and reaction conditions.

6.7 References

449. Kenig M, Reading C (1979): **Holomycin and an Antibiotic (MM 19290) Related to Tunicamycin, Metabolites of *Streptomyces clavuligerus***. *J. Antibiot.* **32**: 549-554.

450. Wyszynski FJ, Hesketh AR, Bibb MJ, Davis BG (2010): **Dissecting tunicamycin biosynthesis by genome mining: cloning and heterologous expression of a minimal gene cluster**. *Chem. Sci.* **1**: 581-589.

451. Tsvetanova BC, Kiemle DJ, Price NPJ (2002): **Biosynthesis of tunicamycin and metabolic origin of the 11-carbon dialdose sugar, tunicamine**. *J. Biol. Chem.* **277**: 35289-35296.

452. Zhang YN, Wang PG, Brew K (2001): **Specificity and mechanism of metal ion activation in UDP-galactose:β-Galactoside-α-1,3-galactosyltransferase**. *J. Biol. Chem.* **276**: 11567-11574.

453. Ly HD, Lougheed B, Wakarchuk WW, Withers SG (2002): **Mechanistic studies of a retaining α-galactosyltransferase from *Neisseria meningitidis***. *Biochemistry* **41**: 5075-5085.

454. Ciesla WP, Bobak DA (1998): ***Clostridium difficile* toxins A and B are cation-dependent UDP-glucose hydrolases with differing catalytic activities**. *J. Biol. Chem.* **273**: 16021-16026.

455. Morera S, Imberty A, Aschke-Sonnenborn U, Ruger W, Freemont PS (1999): **T4 phage β-glucosyltransferase: Substrate binding and proposed catalytic mechanism**. *J. Mol. Biol.* **292**: 717-730.

456. Toumieux S, Compain P, Martin OR (2005): **Intramolecular metal-catalyzed amination of pseudo-anomeric C-H bonds**. *Tetrahedron Lett.* **46**: 4731-4735.

457. Wyszynski FJ, Thompson AR, Davis BG (2010): **Inverted regioselectivity of C–H amination: Unexpected oxidation at β- rather than γ-C–H**. *Org. Biomol. Chem.* **8**: 581-589.

458. Yabe T (2010): *Functional characterizations of the tun genes in the biosynthetic cluster of tunicamycin.* Part II Thesis: University of Oxford.

Appendix

A1 Abbreviations

$[\alpha]_D^t$	Optical rotation at temperature t
A	Adenine
aa	Amino acid
Ac	Acetyl
ACP	Acyl carrier protein
All	Allyl
AMP	Adenosine monophosphate
app.	Apparent
ATP	Adenosine triphosphate
aq.	Aqueous
Bn	Benzyl
BLAST	Basic Local Alignment Search Tool
Boc	*tert*-Butyloxycarbonyl
BOM	Benzyloxymethyl
bp	Base pairs
br	Broad
Bz	Benzoyl
C	Celsius, cytosine
c	Centi
calc.	Calculated
CAN	Ceric ammonium nitrate
Cbz	Carboxybenzyl
CDD	Conserved domain database
CDI	Collision induced dissociation
CMP	Cytidine monophosphate
CoA	Co-enzyme A
COSY	Correlation spectroscopy
CTAB	Hexadecyltrimethylammonium bromide
CV	Column volumes
d	Deca, doublet
Da	Daltons = atomic mass units
dd	Doublet of doublets
dt	Doublet of triplets
DAP	*meso*-Diaminopimelic acid = (2*R*,6*S*)-2,6-diaminoheptanedioic acid
DCC	*N*,*N´*-Dicyclohexylcarbodiimide
DCM	Dichloromethane
DEPT	Distortionless Enhancement by Polarisation Transfer

DMA	*N*,*N*´-Dimethylacetamide
DMAP	4-(Dimethylamino)pyridine
DMAPP	Dimethylallyl diphosphate
DMF	*N*,*N*´-Dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
Dol	Dolichol
DPM	Dolichol-phosphate-mannose
EDTA	Ethylenediaminetetraacetic acid
Et	Ethyl
ER	Endoplasmic reticulum
ESI	Electrospray ionisation
EXSY	Exchange spectroscopy
$FADH_2$	Flavin adenine dinucleotide – reduced form
Fuc	L-Fucose = 6-deoxy-L-galactose
g	Acceleration due to gravity
G	Guanine
Gal	D-Galactose
GalNAc	*N*-Acetyl-galactosamine = *N*-acetamido-2-deoxy-D-galactopyranoside
GDP	Guanosine diphosphate
GlcNAc	*N*-Acetyl-glucosamine = *N*-acetamido-2-deoxy-D-glucopyranoside
Glc	Glucose
GPI	Glycosyl-phosphatidyl-inositol
GPT	GlcNAc-1-phosphate transferase = GlcNAc-1-phosphate translocase
GST	Glutathione *S*-transferase
GT	Glycosyltransferase
GTP	Guanosine triphosphate
h	Hours
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HexNAc	*N*-Acetyl-hexosamine
HIV	Human immunodeficiency virus
HMBC	Heteronuclear multiple bond correlation
HMDS	Hexamethyldisilazane
HMPA	Hexamethylphosphoramide
HMQC	Heteronuclear multiple quantum coherence
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
HSQC	Heteronuclear single quantum coherence
Hz	Hertz
IC_{50}	Half maximal inhibitory concentration
IMAC	Immobilised metal ion affinity chromatography
IPP	Isopentenyl disphosphate
^{i}Pr	Isopropyl

IPTG	Isopropyl-β-D-thiogalactopyranoside
IR	Infrared spectroscopy
J	Coupling constant
JIC	John Innes Centre, Norwich, UK
k	Kilo
kb	Kilobase pairs
K_i	Absolute inhibition constant
K_i*	Final steady-state inhibition constant for slow-binding inhibitors
L	Litre, litres
LB	Luria-Bertani media
LC/MS	Liquid chromatography mass spectrometry
LDA	Lithium diisopropylamide
Lipid I	Undecaprenyl-pyrophosphoryl-MurNAc pentapeptide
Lipid II	GlcNAc-β-(1→4)-Lipid I
µ	Micro
M	Molar, mega
m	Milli, metre, multiplet, medium
Man	D-Mannose
Mb	Megabase pairs
MBP	Maltose-binding protein
Me	Methyl
MEP	2-*C*-Methyl-D-erythritol 4-phosphate
MES	2-(*N*-morpholino)ethanesulfonic acid
min	Minutes
MOM	Methoxymethyl ether
MOPS	3-(*N*-Morpholino)propanesulfonic acid
M.p.	Melting point
mRNA	Messenger RNA
MRSA	Methicillin resistant *Staphylococcus aureus*
MS	Mass spectrometry
MurNAc	*N*-Acetyl-muramic acid = 2-*N*-acetyl-3-*O*-carboxyethyl-D-glucosamine
Mw	Molecular weight
MWCO	Molecular weight cut-off
m/z	Mass/charge ratio
n	Nano
NADH	Nicotinamide adenine dinucleotide – reduced form
NADPH	Nicotinamide adenine dinucleotide phosphate – reduced form
NDP	Nucleotide diphosphate
NMR	Nuclear magnetic resonance
NRP	Non-ribosomal peptide
NRPS	Non-ribosomal peptide synthetase
NTP	Nucleotide triphosphate
Nu	Nucleophile

OD_n	Optical density at a wavelength of n nm
OGT	*O*-GlcNAc transferase
ORF	Open reading frame
OST	Oligosaccharyltransferase
PAGE	Polyacrylamide gel electrophoresis
PEP	Phosphoenolpyruvate
p	Pico
P	Phosphate
PCR	Polymerase chain reaction
Ph	Phenyl
Phth	Phthaloyl
pI	Isoelectric point – the pH at which a protein carries no net electrical charge
PMB	*para*-Methoxybenzyl
PKS	Polyketide synthase
PP	Pyrophosphate, pyrophosphoryl
ppm	Parts per million
PPTS	Pyridinium *para*-toluenesulfonate
pyr	Pyridine
q	Quartet
R_f	Retention factor
RNA	Ribonucleic acid
ROESY	Rotating angle nuclear Overhauser effect
rpm	Revolutions per minute
rRNA	Ribosomal RNA
rt	Room temperature
s	Seconds, singlet, strong
SAM	*S*-Adenosyl methionine
sat.	Saturated
SDS	Sodium dodecyl sulfate
sep.	Septet
SLM	Standard litres per minute
S_N2	Bimolecular nucleophilic substitution
TBS	*tert*-Butyldimethylsilyl
^tBu	*tert*-Butyl
t	Triplet
T	Thymine
TCEP	Tris(2-carboxyethyl) phosphine
TDP	Thymidine diphosphate
TEA	Triethylamine
TES	Triethylsilyl
TEV	Tobacco etch virus
Tf	Triflyl = trifluoromethanesulfonyl
TFA	Trifluoroacetic acid

THF	Tetrahydrofuran
TIC	Total ion chromatogram
TLC	Thin layer chromatography
TMS	Trimethylsilyl
TOCSY	Total correlation spectroscopy
TOF	Time-of-flight
Tosyl, Ts	*para*-Toluenesulfonyl
tRNA	Transfer RNA
Tris	Tris(hydroxymethyl)aminomethane
UMP	Uridine monophosphate
UDP	Uridine diphosphate
UGGT	UDP-glucose:glycoprotein glucosyltransferase
unsat.	Unsaturated
UV	Ultraviolet
V	Volts
v	Volume
VRE	Vancomycin-resistant enterococci
VRSA	Vancomycin-resistant *Staphylococcus aureus*
w	Weak
W.IP.E.	Water:isopropanol:ethyl acetate

Amino acid	3-letter code	1-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Table A1.1 Abbreviations of amino acids

A2 Sequences of individual tun genes from S. chartreusis

A2.1 tunA: Putative UDP-GlcNAc dehydratase

DNA sequence of tunA gene:

```
  1 GTGCAAAGAA ACACTTTGAA ACACCGGATT CTCATCACCG GAGGCGCCGG ATTCATCGGT
 61 GGCCACCTTG CCCGCGCACT GGTGGCGTCC GGGGAAGAAG TAACAGTCCT GGACGACCTG
121 CGTGTGCCGC CGATGATCCC GCCGGAAGGC ACCGGAAAGT TCTTAGAAAA ACCCGTCTTA
181 GAGTTAGAGG AAAGGGATCT CTCTGACGTA CGCCTCGTCT ACCACCTGGC GTCGCACAAG
241 AGTGTTCCTC GCTCCTTCAA ACAGCCGCTC GACTATCTCG ACAACGTGGA CTCGGGCCGC
301 CACCTGCTGG CTCTCTGCAC GAGCGTTGGC GTGCCGAAGG TGGTCGTCGG GAGTACGTGC
361 GAGGTCTACG GCCAGGCCGA CACACTACCG ACACCCGAAG ATAGCCCTCT ATCCCCGCGC
421 TCGCCCTACG CCGCGAGCAA GGTCGGGCTG GAGATGGTGG CCGGGGCCCA CCAGCGCGCC
481 TCAGTCGCTC CCGAGGTGGG CATCGTTCGC TTCTTCAACG TCTACGGCCC CGGCGAGCGG
541 CCCGACGCCC TCGTACCCCG GCTCTGCGCC AACCTCCTGA CCAGGAACGA ACTGCCTGTC
601 GAGGGCGACG GTGAGCAGCG TCGCGACTTC ACCTACATCA CGGATGTCGT GGACAAACTC
661 GTCGCGCTGG CGAACCGACC ACTGCCCTCA GTTGTCAACT TCGGTTCCGG CCAGTCCCTC
721 TCCGTGAACG ACGTCATCCG GATCCTGCAG GCGACCTCTC CGGCGGCCGA GGTCGCCCGG
781 AAGCAGCCTC GGCCGAACGA AATCACGGAG TTCCGGGCGG ACACCGCCTT GCAGACGCGG
841 CAGATCGGAG AGAGGTCCGG CGGGATCGGA ATCGAGGAAG GCATTCGCCT CACTCTCGAA
901 TGGTGGCAGT CGCGGGACCT TGACGACATA CGCCAGCGTA TCTTCCAAGA GGAGGGGGCC
961 GAC
```

Amino acid sequence of TunA protein:

```
  1 MQRNTLKHRI LITGGAGFIG GHLARALVAS GEEVTVLDDL RVPPMIPPEG TGKFLEKPVL
 61 ELEERDLSDV RLVYHLASHK SVPRSFKQPL DYLDNVDSGR HLLALCTSVG VPKVVVGSTC
121 EVYGQADTLP TPEDSPLSPR SPYAASKVGL EMVAGAHQRA SVAPEVGIVR FFNVYGPGER
181 PDALVPRLCA NLLTRNELPV EGDGEQRRDF TYITDVVDKL VALANRPLPS VVNFGSGQSL
241 SVNDVIRILQ ATSPAAEVAR KQPRPNEITE FRADTALQTR QIGERSGGIG IEEGIRLTLE
301 WWQSRDLDDI RQRIFQEEGA D
```

A2.2 tunB: Putative Fe-S oxidoreductase

DNA sequence of tunB gene:

```
  1 ATGACCGGCT ACACCGCACC TGTACGCGCC AGTCTGGATG TCACTTACGC ATGCGACCTG
 61 CGCTGCATCC ACTGCCGCAC GAACACCGGG GAGATCCCCG TCCACATCCG CCGCAAGATG
121 CTGTCCATCG AGCAGTTGCA GGACATCATG CTCCAACTGG ACCGGATGGG GACATTCGAG
181 ATCACCCTCA CCGGTGGGGA GCCGACGATC CGCAAGGGCT TCTGGCAGCT CCTGGACGTC
241 GTTCCGAAGC TGCGCCATTC GACCGTCACT CTCATCACGA ACGCGGCAGC GCACACGCGG
301 GAACAGCTCG ACCGCATCGT GGACAGCGGC ATTTCGTCCA TCCGGGTCAG CATGGACGGC
361 ACGCGGGAGA CGTTCGCCGC CGTACGGCTC ATGGACGTGT TCGACACCGT CGTGGAGAAC
421 TCGATCTACC TCCGCGACCG TGTGCGCAGT TTCAAGGTGC TGACCACGGT CATGAAGACC
481 AACCAGCACA ACGTGTTCGA CCTCGCGCAC TACCTGCGAG CCCACGGCTT CCGGCGCCAG
541 GACCTCATCC TGGTCCGGGC GCACGGCCGC GGTGGCCGCA ACCGTCTGCT CCTGACAGAG
601 GAGGAGACCA TGGCGGTCCA CCGCAAGGTG ACCGAGTTCA AGCGGGACGT GCCGGCCACG
661 GACTTCGACC TGAACCTCAA CGCGCCCTAC CTGGTCCCGG GCGAGAAGCT GCGCCCCTTC
721 CAGGACGTCG TCATGTTCCC CTACCTGGTG CGCGACAGCA GCATCGCGAT CTCCGCCACA
781 GGCGACGTGA CCATGAGCCG CCTCTACAGT GCCGCGCCCT TGGGGAACAC CAAGGATGCT
841 CCCGTGCAAG AGATCTGGAC GCAGGGGCAG CAGCAGCTCG TCCAAGAGCA GGAGGAGTTC
901 ACCGAGGAGC GTCTGCGCGA GATCTTCTGG GACTTCGCGA CCAGTGACGG ACCGGACGGT
961 CCCCCGCTCA CCTCGCTGCT GGACCGGCAG ATCTTCGAAG GCGCGGAGGT GAGG
```

Amino acid sequence of TunB protein:

```
  1 MTGYTAPVRA SLDVTYACDL RCIHCRTNTG EIPVHIRRKM LSIEQLQDIM LQLDRMGTFE
 61 ITLTGGEPTI RKGFWQLLDV VPKLRHSTVT LITNAAAHTR EQLDRIVDSG ISSIRVSMDG
121 TRETFAAVRL MDVFDTVVEN SIYLRDRVRS FKVLTTVMKT NQHNVFDLAH YLRAHGFRRQ
181 DLILVRAHGR GGRNRLLLTE EETMAVHRKV TEFKRDVPAT DFDLNLNAPY LVPGEKLRPF
241 QDVVMFPYLV RDSSIAISAT GDVTMSRLYS AAPLGNTKDA PVQEIWTQGQ QQLVQEQEEF
301 TEERLREIFW DFATSDGPDG PPLTSLLDRQ IFEGAEVR
```

A2.3 tunC: Putative N-acyltransferase

DNA sequence of tunC gene:

```
  1 TTGTCCAGGG AGGCCCTCAT ACGCCTGTAC GACCTCACGC CGTCCCAGCC TCTGCTCGAC
 61 GCTCTGAGCC CTGCCACCGC ATCGCGGGAT ATCGCTCCTG TCGTTCCCCG GTTCAAAGGG
121 GCCGCTGGGC CACGAGCCCA GTCCTTCGTC GAACTGCACC GGGAAGGGAC GCTTCTGGGC
181 CGGTGTGGCA TCAACGTCAA GGGCCCGGGC ACAGTCGGCG CCTGCGAGGT GGCGGCCGTG
241 GTCGCCCCGG CGGAGCGGGC GGGCATGCAC TGGCTTCTCG TCCATGTCGC ACTCGAGCGG
301 CTTCAGTGGC TGGGGTATGC CTACGCGATG ACGGAAGTCA GCGAATACGC GGACCACTTC
361 CCGTCGGTGC TGCGGCAGGC GGCGTGGTGG ATTCCCGATT CCAGCGAACG CAAGAGCGCG
421 GCGGCGCGGG ACGACAAGAG CCTGGAATGG GCGGATCTCT TCATCGACTT CCGGACCTGG
481 ACGCCAAGTT CTACTCCCAC GTCTCTTACG GTCAACGGCC GAGACCTGTG GGTGCGTCGT
541 CCCGAAGCGA GTGAGGAACT GCTGATTGTC GACTGGCTCA GAGAGACCTT CGGCGGAGGC
601 TGGGCGAGCG AGATCCACCG GTCGTTCTCA CGGGATCCGA TCTCGTCTGT GATCGTGGTC
661 GACCGGAACA AGGAACTGCC GCCCAAGGAC CGCTTGCTCG GCTTCCTTGC CTACGACACT
721 GCGCGGCTCG GGATGCTGTC TGCGATTGCG CTCGTGCCCG AGACACGCGG TCGCGATCTG
781 TCCTTGGCGA CAGCGCTCAT CGAGGAGTGC CTGCGCGAAG CGCGAGCCAG CGGCATGACC
841 TATGCCGTGC TGGGAGGCGT AGGCAACGCC CGCTTGGCGG CTCTTCGCAC ATTCAGCGCT
901 CTATGGACCA TCCCTGGCTC GTGCCCGGGA ATCTTCGGGA GGGGTGTTCG GAAC
```

Amino acid sequence of TunC protein:

```
  1 MSREALIRLY DLTPSQPLLD ALSPATASRD IAPVVPRFKG AAGPRAQSFV ELHREGTLLG
 61 RCGINVKGPG TVGACEVAAV VAPAERAGMH WLLVHVALER LQWLGYAYAM TEVSEYADHF
121 PSVLRQAAWW IPDSSERKSA AARDDKSLEW ADLFIDFRTW TPSSTPTSLT VNGRDLWVRR
181 PEASEELLIV DWLRETFGGG WASEIHRSFS RDPISSVIVV DRNKELPPKD RLLGFLAYDT
241 ARLGMLSAIA LVPETRGRDL SLATALIEEC LREARASGMT YAVLGGVGNA RLAALRTFSA
301 LWTIPGSCPG IFGRGVRN
```

A2.4 tunD: Putative N-acetylglucosamine transferase

DNA sequence of tunD gene:

```
   1 ATGGAGATCA TCTTCACTGT CTCCGATCAG GTATGGGGCG GCAAACACCG CTACATGCAC
  61 GACATGGCGC TCGGGCTGGC GCAGGCAGGC CATACGGTCA CCGTCCTCGC GGAGGAGGGT
 121 GGAGCCATGC TTCAGCAGTG CCGTGCCGCC GGGGTAACGA CCGTGAGTGT GCCGGCGTTC
 181 GCCAGCGACG ACGCGGCCGA GGCGGTTCGC AAGGCGCTGC GGCATCGTCG TCCTCATATC
 241 GTGTGTGTCT CCGGCCGCGC GGACGCCGCC GCAGTCCACC ACGCCCAGTC GCAGGGCGTG
 301 ACTGATGCCG CCGTGTGCCT GTTCCGGCAC TCGGCGTTCC CTCTGGGAAC CACCGACGAG
 361 GTGCGCGACC TATTCACCGG CGTGAACCTG GTCTTCACCA CGTCACTGGA ACAGCGGCAA
 421 CGCCAGTTCG AGCCGCTGAT CAACGCCGGT GTCTTGAAGG ACGAACAGGT CGAGATACTC
 481 ACCAGTGGCG TCGGGGAGCC TCTGCTCGCG GCTCTGGACG CCGCCGACCG CGGCGCCGCT
 541 CGAAAGGAAC TCCGAGCCGA ATCGGACCAG TTCGTCTTCC TCGTGCTGGC CCGCCTGGCC
 601 TGGGAGAAGG GCATCGACCA GGTGATCGAC GCCTTCGCTG ACCTGGAGCT GCCGCCGGAC
 661 GCCGCGCCCC CGCTGCTGGT GGTCGCGGGT GAGGGGCCGC TGGAGGCAGA ACTACGAGGC
 721 CAGACCATCG AACGGGGTGT AGCCGAACGG GTCCAATTCC TCGGCCACCA GGACCACGTC
 781 GCCCCGGTGA TCAAGGCCAG TGACGCCGTG GTCCTCACCA GCACGGTTCC CGAGACCGGC
```

```
 841 CCGCTCGCGC TCAAGGAGGC CATGGCCGCC GGCCGCCCCG TCATCGCCTC TGTCCAGGGC
 901 GGCATCCCGG AGTTCGTCGA GGACGAGCGG CACGGTCTGC TGGTCATCGA CGACGAGGAC
 961 CTGCGGCAGG CGATGCAGCG GCTCCTGAGC GACCGCGAAG CCGCCGAGAC CATGGGAGCC
1021 GCCGGATCCG AGTCGGTTCG CGGCGGACAT CGTGCAGTAC GACGCGTGGA GTACCTCGCC
1081 CACCGTCTCG ATCTCCTGGC CCTGGAACAG CTCGCCCCGG ACACGGTCCT TCACGAAGTG
1141 GTGTGGGACG ACGTACGGCT GCGTGAAGAG ACACAAGGCG GCTTCGTCTT CGTGCCGCGT
1201 ACATCGCACA TCATGGAACT GGACAGCGCT ACGTACGCTG TCGTGCGCAC CGCGGTCGAA
1261 GCCGGTGACC CACTCCAGTT GATCAAGCTC CCCGAAGAGA CGCTCGGTGT GATTGCACAC
1321 CGGCTCTACG CCATGGGCGC GCTGGTCCGG CAGGATGGTC AGGCGACGCC TGCGGCCCGC
1381 ACCGCGGACG GCGAACGACC GGTTACGGAG CCCGCG
```

Amino acid sequence of TunD protein:

```
  1 MEIIFTVSDQ VWGGKHRYMH DMALGLAQAG HTVTVLAEEG GAMLQQCRAA GVTTVSVPAF
 61 ASDDAAEAVR KALRHRRPHI VCVSGRADAA AVHHAQSQGV TDAAVCLFRH SAFPLGTTDE
121 VRDLFTGVNL VFTTSLEQRQ RQFEPLINXX AGVLKDEQVE ILTSGVGEPL LAALDAADRG
181 AARKELRAES DQFVFLVLAR LAWEKGIDQV IDAFADLELP PDAAPPLLVV AGEGPLEAEL
241 RGQTIERGVA ERVQFLGHQD HVAPVIKASD AVVLTSTVPE TGPLALKEAM AAGRPVIASV
301 QGGIPEFVED ERHGLLVIDD EDLRQAMQRL LSDREAAETM GAAGSESVRG GHRAVRRVEY
361 LAHRLDLLAL EQLAPDTVLH EVVWDDVRLR EETQGGFVFV PRTSHIMELD SATYAVVRTA
421 VEAGDPLQLI KLPEETLGVI AHRLYAMGAL VRQDGQATPA ARTADGERPV TEPA
```

A2.5 tunE: Putative N-deacetylase

DNA sequence of tunE gene:

```
  1 GTGAAGGTCC TGGTCATCGC CGCACACCCG GATGACGAGG TTCTCGGTGC CGGGGCCACG
 61 GTGTCGGCAT TGAGCCAGCA GGGCGCAGTG GTCCAGATCC ACATCCTCGC CGAGGGTATC
121 TCGCTGCGGC ATTCGGGCGT GAAGATCGAG GAGGCGCGGG AACGCTGCCA GGTGGCGGCC
181 AAGGATCTCG GTGCGGCCGA GGTGTCGTTC GGCGGTCTGG CGGCCGACGG ACGCCTCCTG
241 GCCGACCTCC CTCAGCGACA GGTCGTCGAT GCGGTTAGCC ACGCCCTGCG TGAGGCAGAG
301 CCCGAAGTGG TCTTCACTCA CCACCCTGGC GACATCCACG CGGACCACCG CAGCGTCGCC
361 CATGCAGTGG GGTACGGCAC CCGCATCCTC GGCTCCGGCT CCGTTCGGCA GGTGCTGCAC
421 TTCGAGGTCC TGTCTTCCAC GGAACAGCAA ACGGGCCTGG TGGCGCCGTT CACCCCCAAC
481 CTGTTCTACG ACGTCACCGG GCACGTCGAA GCCAAGTGCC GGGCACTTGC CGCCTACCCG
541 TACGAGCTGT ACGACCCGCC CCACCCTCGC TCCCTGGCGG CCGTTCGGAC GCTGGCCTCA
601 TACCGGGGGA CGCAGGTGGG GGTGGAAGCG GCTGAGGCAT TCATGATCGG ACGAGAACTA
661 CGCGGCCCCA TGGCAGCGGT TCCCAGTGGA GTTGATGTGC GA
```

Amino acid sequence of TunE protein:

```
  1 MKVLVIAAHP DDEVLGAGAT VSALSQQGAV VQIHILAEGI SLRHSGVKIE EARERCQVAA
 61 KDLGAAEVSF GGLAADGRLL ADLPQRQVVD AVSHALREAE PEVVFTHHPG DIHADHRSVA
121 HAVGYGTRIL GSGSVRQVLH FEVLSSTEQQ TGLVAPFTPN LFYDVTGHVE AKCRALAAYP
181 YELYDPPHPR SLAAVRTLAS YRGTQVGVEA AEAFMIGREL RGPMAAVPSG VDVR
```

A2.6 tunF: UDP-GlcNAc-4-epimerase

DNA sequence of tunF gene:

```
  1 ATGAGAGTGC TCGTGACTGG GGGTGCCGGG TATGTGGGCT CCTTCACCGT ACGGCGCCTC
 61 CTTGCGGCCG GCCACGAGGT CGTGGTTGTC GACAACCTCA GCACGGGCCG CCGCGAAGCA
121 GTCCGTGGCT GCCGCCTACA CGTGGTGGAC ATCCTCGATA TCGCCAGCAT GGGCACGGTC
181 TTCGAGGAGT TCCACCCTGA GGCCGTGATC CACTTCGCGG CCCTGAAATC CTCTGAAGAG
241 TCGCTGCGTG ACATCAACAC GTACTTCTCC GTCAACCTGA CGGGGACGCA GAACGTGCTC
301 GCCCTGTGCG CACGCACCGG GGTGGAGCGC TTTGTCTTCT CCTCCTCGTG CGCCGTGTAT
361 GGCACCCCGC AAATCTGCCC GGTGGATGAG ACCGCGCCCG TACGGCCCGA AAGTCCGTAC
421 GGAGAGACGA AGTACCTGTG CGAGCGCGTG ATCGCCTCCT ACGCGCAGGC CACAGGGATG
481 CGCTACGCCA ACCTGCGGTA CTTCAACGCG GCAGGTGCGG CCGACGACGC TTCCCTGGGA
541 GAGTACGCGG GCCCGGCATC AACCCAACTC CTACCGGTGG CCATACGCTC AACCCTGGGC
601 CTGGCGCCCA CCCTGCGGAT CTTCGGCAAC GACTACCCGA CGACCGACGG GACGGCTCTG
661 CGCGACTACA TCCACGTCGA GGATCTCGCC CAGGCCCACG TGCGGGTCCT GGAGGGTCTG
721 CAAAAAGGCA GTCAGTGCGG GCCGGTGAAC CTGGGCCGCG GCCAGCCCGT CAGCGTGCGC
781 GAGCTGATCG ATGTCTTGCA CGACGTCTCC GGGAAGGAGA CCCCGGTTGC CATGGCGCCC
841 CGCCGTCCAG GGGACCCAGG TCTGTCCTGG GCCGATCCGA GTCTGGCGCT GAGCCGATTC
901 GGCTGGCGAG CAGAACGCGA CCTTGACCAC ATCGTTCGTA CGGCGTGGAA CTGGCACACC
961 AGTACCCCCG CGTCGCTGGA G
```

Amino acid sequence of TunF protein:

```
  1 MRVLVTGGAG YVGSFTVRRL LAAGHEVVVV DNLSTGRREA VRGCRLHVVD
 51 ILDIASMGTV FEEFHPEAVI HFAALKSSEE SLRDINTYFS VNLTGTQNVL
101 ALCARTGVER FVFSSSCAVY GTPQICPVDE TAPVRPESPY GETKYLCERV
151 IASYAQATGM RYANLRYFNA AGAADDASLG EYAGPASTQL LPVAIRSTLG
201 LAPTLRIFGN DYPTTDGTAL RDYIHVEDLA QAHVRVLEGL QKGSQCGPVN
251 LGRGQPVSVR ELIDVLHDVS GKETPVAMAP RRPGDPGLSW ADPSLALSRF
301 GWRAERDLDH IVRTAWNWHT STPASLE
```

A2.7 tunG: Putative UMP phosphatase

DNA sequence of tunG gene:

```
  1 ATGCACCTCT ACCTACTGCG CCACGCCGAG GACCGTTCCG GAAAGTTCGA CCCGCACCAG
 61 GATCGCCTGA CCGAGCGAGG ACGCCGACAG GCGCAGGAAG CGGCTCATTG GCTGGCTGAG
121 CGGCGTCCCA CAGAGCTTCG CACCAGTGTG CTTCCCCGCG CCCAGCAGAC AGCCGACATC
181 GTCGCGGCGC ACACCGGCCT TGCCGCACGG GAGGATCCGC GCCTCAATGA GATCGTCTGG
241 TCTCCGGTCG CGGGTGAGTT CCTGCCGAAC CCGCGGGAGG ACCGGCTCCA TCCGCCGCCA
301 CCGGGGGCGG AATCATGGAA TGTGTTCATG ACGCGGGTCG CCACATGCAT GAGCGAGCTG
361 TGTCATCAGG CGTCGGAGGA CCAGCGGATT GTGTTGGTCA CCCACAGTGG CTTCTTCGAC
421 GCTGTCAATG AACTGCTGTG TGGCGGGACC GGCAGGGTGG AGCTGCTGGT TGCGCACGCT
481 GGGATCACAC ACTGGCAGTA CCGTCCCGGA GACTCGGGTG GCGCCTGGCT TCTGCACCAA
541 CACAATGTCA CCTTCGCCCC ACAGACCCTG ATGGATCACG GATCCCTGAG CACAGCGGAG
601 TTGAGCGCA
```

Amino acid sequence of TunG protein:

```
  1 MHLYLLRHAE DRSGKFDPHQ DRLTERGRRQ AQEAAHWLAE RRPTELRTSV LPRAQQTADI
 61 VAAHTGLAAR EDPRLNEIVW SPVAGEFLPN PREDRLHPPP PGAESWNVFM TRVATCMSEL
121 CHQASEDQRI VLVTHSGFFD AVNELLCGGT GRVELLVAHA GITHWQYRPG DSGGAWLLHQ
181 HNVTFAPQTL MDHGSLSTAE LSA
```

A2.8 tunH: Putative nucleotide pyrophosphatase

DNA sequence of tunH gene:

```
   1 ATGAGCGGGC GGCGCAGGCG TATGGTCGTC ATCTGTGTCG ACGGCGGGCT TCCCGGGATC
  61 ATCCGCGAGC ACGACTTCTT CAATCTGAGC CGGGCCCTGC CGACACTTCC CGGCACTGCC
 121 GTGCACGAGC TGCGCAGCAT CTACCCCTCG TCAACAGCGC CCTCCCACGC GTCCTTTCTT
 181 ACCGGGACGT ACCCGAGCGG GCACGGGATC GTGGGGAACC GCTTCTGGGA GCGGGAGAGT
 241 GTTGAGGAGA TCCGCCGACG CTCGGACGAT CCGCTTTCAT CTTTTCACCC CTACGAGGAA
 301 AGCTCGCTCA CCGCACCGAG CCTGCTCGAC TGGTTTGCCC GGCAAGGGGC GTCGGCCGCC
 361 GCCGTCCATT TCCCGCAGAC CTTCTCCCGC AACGCCCAGC TCGCGATCCC CAGCTGCTAC
 421 TGCTTGTACG CCCCAGCCCG GAACCTCGTC GTCCCGCTCG GACCGACCGT TGACGGGGCG
 481 GCCGAGGGCG TGGTGCAGCT CTCCTACCTG GGACACGAGG TGGCCTTGTA TCTCCGGGTC
 541 GATCAGCAGA CGAACGTGAT CACGGTGGGA AGCCACCGAG AGACCGCTGT CGTGGCCGAC
 601 AGCCTCCGGC CCACGCGGCT GGACATCCCC GTTTCTTCGG GTTCCGTGTC CGTGGCGGTC
 661 TCCTGCCGCA GGCTCGACGA AGGCCAGATC GAGGTGCGTC TGGGAACCGC CGTTATCACC
 721 CTGGGATTCG GCGGGCTCGA TATGCCGGAC CGCGCCGGTG ACGGGCCCGC GTCCCTGTAC
 781 GTCGAATACA CCGCCAATCC AGGCCACACG TTCCACGAAT CGCCTCGGGC CGAGTGGGTG
 841 GAACAGACCG CCCTCGACGT CCTGAAGCAG CATGACCCAG ATGTCCTGTT CGTGCGCTTC
 901 AACCAGGCCG ACCATGCTCA GGAGTTCCTC TACTGGCATG CGGTCCGAGG CAGCTCCGGT
 961 GAACGAAGTC GCGCATGGCA GCAGATCCTT GACGTCTACG CCCGGATCGA CGCCTGCGTG
1021 ATGCGGCTCG CCCAGGCCGC CGGTGAGCAC GCGGAGTTCC TGTTCTTCTC CGACCACGGC
1081 ATCGACTACG TCGAGACCCA CCTGCGCCCG AACCACGTGC TGGCGGACCT CGGGCTCGCC
1141 GACCGCATGG TTTTCCAGGG TGACAGCAAC TGCGCCTACC TGTACGCCGA CGCACCTGTG
1201 CCACCAGCGG AGAAAAAGCG CCTCGACCAG GCCCTGCACG CACTCGATCC TTCCATCGGC
1261 GTGCTCACTC CAGCCGAGCA GGACCGGTTG AAGTTGCCGG TAGGGAGCCC GAGGCTGGGC
1321 CGTCTCACGG TGACCTGCGG CCCGCACACC GAATTCCAGT ACAGCGACGC GACACCGAGG
1381 GAGAGCGTGG CCTCCGCATC ACACGGTTAC CTGCCCAGCG ACTCGGCCAT GAGCGGCTTC
1441 TTCCGCTGCT TCGGCCAGGC AGCAGCGCGG CCGGCTCCCC CCAAGGACAT CACGGGGGCC
1501 GCAACCGCGG TCCGCAGCAT CTGGCGCCAA GCGGCGCGGG AGATG
```

Amino acid sequence of TunH protein:

```
  1 MSGRRRRMVV ICVDGGLPGI IREHDFFNLS RALPTLPGTA VHELRSIYPS STAPSHASFL
 61 TGTYPSGHGI VGNRFWERES VEEIRRRSDD PLSSFHPYEE SSLTAPSLLD WFARQGASAA
121 AVHFPQTFSR NAQLAIPSCY CLYAPARNLV VPLGPTVDGA AEGVVQLSYL GHEVALYLRV
181 DQQTNVITVG SHRETAVVAD SLRPTRLDIP VSSGSVSVAV SCRRLDEGQI EVRLGTAVIT
241 LGFGGLDMPD RAGDGPASLY VEYTANPGHT FHESPRAEWV EQTALDVLKQ HDPDVLFVRF
301 NQADHAQEFL YWHAVRGSSG ERSRAWQQIL DVYARIDACV MRLAQAAGEH AEFLFFSDHG
361 IDYVETHLRP NHVLADLGLA DRMVFQGDSN CAYLYADAPV PPAEKKRLDQ ALHALDPSIG
421 VLTPAEQDRL KLPVGSPRLG RLTVTCGPHT EFQYSDATPR ESVASASHGY LPSDSAMSGF
481 FRCFGQAAAR PAPPKDITGA ATAVRSIWRQ AAREM
```

A2.9 tunI: Putative ABC transporter – ATP-binding subunit

DNA sequence of tunI gene:

```
  1 ATGCCAGAAA ACCCTCCTTA CGCATATGAG GTCGAGGGCC TGCACAAGCA GTTCGACGGA
 61 GTCCCCAGGC CGGCCAATGG CGAGATCTCC TTTCACATAA AGCAGGGCGA ATTCTTCGGA
121 CTTCTGGGCA GCAATGGTGC GGGAAAGACA ACCCTGATCA AGCAGATGAT CGGTCTGGTG
181 CGACCGGATG GGGGGACGGT GAGGCTACTG GGGCGCGATG TCTTTGCCCA CCCTCGCTTC
241 ACGGCGGCCA CGGTTGGTTA CATGCCGCAG ACCGCCTTCG CCCTCAACAG CCTGACCGTC
301 AAAGAGGCCG TCTACTTCAG TGCGCACCTA CGCGGTGTCG CCGCCAAGGC GGCCCGCCGG
361 GAGCGTGACG CCCTGTTGGA GCTGCTTGGT CTGACCCACC TGGCCGGCAA GGTCGCCCGG
421 CAGCTCTCCG GCGGAGAACG CCGACTGTTG CAGATCGCGG TGACGCTGAT CGGATCACCG
481 CCTGTGCTGA TCCTGGACGA GCCCACCAAC GAGCTGGACC CGTCGCGTCG CCGCCAGATA
541 TGGCAGCTTC TCAGGGAGCT GAACCAACGC GACGGCCACA CGATCATCCT CATCACCCAC
601 AACGCCATCG AAGCGGAACA GGTGATCGAG CGCGTCGGAA TCCTCAAGGA GGGGCGGCTC
661 GTCGCGCTGG GCGCTCCCGC CGAGCTGAAG TCTTCCCTCC GCTCCCAGGT ACGCCTGTCG
721 ATCACGGTAG GCGATACACA AATGCATGGT TTGCCTGACT ATGTCCGCGT GCAGTCCGAC
781 CACCGCGGCA AGGTCGTGGC CCTCGTCGAC TCTGCCGACC TGCCCCATCT GACCAAGGAC
841 GTAGATCTCA CCCAGTTCCC CGAGTTCACG ATGCACACCG CGACCCTGGA AGACGTGTAT
901 GTCAGCTACG CG
```

Amino acid sequence of TunI protein:

```
  1 MPENPPYAYE VEGLHKQFDG VPRPANGEIS FHIKQGEFFG LLGSNGAGKT TLIKQMIGLV
 61 RPDGGTVRLL GRDVFAHPRF TAATVGYMPQ TAFALNSLTV KEAVYFSAHL RGVAAKAARR
121 ERDALLELLG LTHLAGKVAR QLSGGERRLL QIAVTLIGSP PVLILDEPTN ELDPSRRRQI
181 WQLLRELNQR DGHTIILITH NAIEAEQVIE RVGILKEGRL VALGAPAELK SSLRSQVRLS
241 ITVGDTQMHG LPDYVRVQSD HRGKVVALVD SADLPHLTKD VDLTQFPEFT MHTATLEDVY
301 VSYA
```

A2.10 tunJ: Putative ABC transporter – permease subunit

DNA sequence of tunJ gene:

```
  1 ATGTCAGCTA CGCGTGACAA GAAGCTCGGA CTGAGCTTCC TTGGGCAGGT CTGGGACCTG
 61 TTCAGGATTC AGCTCACCAA CTGGCGGTGG TCCTGGCCTC AGATGGTGCT CACTGGTGTG
121 CTTGCCCCTT TGGTCAGCAT CGCCGCCTTG GGGCTCTTCG CCCGGAACTC TGGCCACGCC
181 GCGACCGAAT ATGTCCTGAC CGGCAGTGTC ACCATGGCGA TCCTGTTCGA GACTCAAAAC
241 AAAATCGCCA GCAACTTCGC GTTCATGCGC GGAAACGGCG CCTTCGAGTA CTACGCCGCA
301 CTGCCTGTCC GTCGCGAGGC TCTCATCTTT GCCACACTCG CAGCCTTCTC CATCCTGTCG
361 CTGCCTGCCA TCGCGGTGAC CCTGCTGCTC GGCTCGGCCC TACTCGACGT CCCGCTCACG
421 TTCTCGCCAC TCGCCCCGCT GTGCCTCCTG CTTGCCCTGC TTCCTTCAGC CGGACTGGGG
481 GCCTTCATCG GCAGCAGATC CGGAACCATC GAACAGGCGT CCTCTCTCAG CCTGGCCACC
541 ACGCTCCTCA TGATGGCCGC AGGCCCGGTG GCAGTCCCAC CAGAACTGCT GCCCGACGCA
601 CTGATCTGGA TCGGCAACCT CAACCCGGCG GTCTACGCGG CCGACAGCCT GCGTCACTCC
661 CTAACGGCAC CCGACGCCTC CCGGGCCCTT TCTGACTTGG CGATCCTTGC CATCTTTGCC
721 GGCGCGGTGT GGGGCCTGAC CAGCACACGT ATGCACTGGC GTGCCCGCTA CGGGCGCCGC
781 GCCTCG
```

Amino acid sequence of TunJ protein:

```
  1 MSATRDKKLG LSFLGQVWDL FRIQLTNWRW SWPQMVLTGV LAPLVSIAAL GLFARNSGHA
 61 ATEYVLTGSV TMAILFETQN KIASNFAFMR GNGAFEYYAA LPVRREALIF ATLAAFSILS
121 LPAIAVTLLL GSALLDVPLT FSPLAPLCLL LALLPSAGLG AFIGSRSGTI EQASSLSLAT
181 TLLMMAAGPV AVPPELLPDA LIWIGNLNPA VYAADSLRHS LTAPDASRAL SDLAILAIFA
241 GAVWGLTSTR MHWRARYGRR AS
```

A2.11 tunK: Putative acyl carrier protein

DNA sequence of tunK gene:

```
  1 ATGACCGAGC AACAGTTCCT CGCAGTCCTC GCACCCCTGA TCGAGGAAAT CACCGGGAAC
 61 GCCTTCTCCC AGGAACAACT GGACGTCACC TTCGGCGATC TGGACATCGA CTCCCTCAAT
121 CAGATCGAGA TCATCAGCCG TATCGAGGAC GAGTTCGACT GCCACATAGA GGACAGCACC
181 CTGCGGACGA TCCGTACGCC CCGCGGACTG CTGGACCACA TTTCGAAGTC CACCAGCCCC
241 GCA
```

Amino acid sequence of TunK protein:

```
 1 MTEQQFLAVL APLIEEITGN AFSQEQLDVT FGDLDIDSLN QIEIISRIED EFDCHIEDST
61 LRTIRTPRGL LDHISKSTSP A
```

A2.12 tunL: Putative phospholipid phosphatase

DNA sequence of tunL gene:

```
  1 ATGATCCGAA ACCACTCCGG CGACCGGCCC CGGTCACTGG CATGGTTCGC GTGGTGGGCA
 61 GCGGTCCACG CCGGTCTGCT ACTCGTCCTC GTCACCTTCC GGTGGGACGA ACGGCTGAGC
121 GAGACGCTCT GCGCCTGGAC GAGCAGCAGC GCGTCCAGTG CCCGGCTTCT CGAGACACTG
181 ACCTCGTGGG GCAACCCGTG GCCTGTGACC ATCATTGCCG CCGGATCCGC GTGCTTCGCG
241 CTGATCCGGC GGAACGTCAC AGCTGCCGCA GTCCTCCTGG TCATTCCGTA CGCCGCCACA
301 GTGAGCTGCA ACCTGCTCAA GGGATGGGCT GAACGACAGC GTCCGGCCTT GGATTGTGCC
361 TCGATTCAAG CCGACGGCTT CAGTTTTCCC AGCGGACACG CGGTCGGTGT GACGACCGGA
421 TTCATGCTGG CCGCGCTCTA CTTCAGTTCC CAGATGCCAC GGGCCGTCTC CGCACTGATG
481 CTGTTCCTTG CTGGGCTGAG CGCCGAACTC GTCGCGTGGA GCAGGGTCCT TCTCGGTGCG
541 CACTTCAGCG TCGATGTGCT CGCTGGGCAA CTCTTGGGCG CCGGCTGGCT GGCCGTCGCC
601 GCCCTCGTGC TGCACCGCCA GTCAGTCCGC CGCTTGCAAC GCGAGGCCCG TCAAGCGGGG
661 TCGAACTCCC TGGACGAGCC CGTGTCA
```

Amino acid sequence of TunL protein:

```
  1 MIRNHSGDRP RSLAWFAWWA AVHAGLLLVL VTFRWDERLS ETLCAWTSSS ASSARLLETL
 61 TSWGNPWPVT IIAAGSACFA LIRRNVTAAA VLLVIPYAAT VSCNLLKGWA ERQRPALDCA
121 SIQADGFSFP SGHAVGVTTG FMLAALYFSS QMPRAVSALM LFLAGLSAEL VAWSRVLLGA
181 HFSVDVLAGQ LLGAGWLAVA ALVLHRQSVR RLQREARQAG SNSLDEPVS
```

A2.13 tunM: Putative radical SAM protein

DNA sequence of tunM gene:

```
  1 GTGCCGTTCA ACCACAACGA TCATTACCAT GACCTGTTAC TTCGTGAACT TCCCGAATCA
 61 TGCCGAAGGG CCCTTGAGGT GGGCTGCGGC ACAGGGAGGT TTGCACAGAA ACTAGCGCTT
121 CGAGGAATCG AAGTAGATGC GATTGACGTA GATAATGACG TCGTCACCGC CGCACGAGAG
181 ATGAATTCTA CTCTCGATCT CCCCAGGCAG ATCCGCTTCG AGCGGGCAGA TATCACACAC
241 CTAACCTTGA CACCTGACAC ATACGACTAC ATCGCGTGCT TGGCCAGCAT CCACCACGTC
301 CCCTTTGAGA CAGTGCAGAA GCTGCGCGCA TCGCTCACCA CGAACGGCGT CTTGGTGATC
361 CTCGGATGCT ACCGAGAAGC AACGATATGG GACCACGCCA TAAGCCTTCT GGCCGTCCCT
421 GTGAACGCTA TTTGGCGTGC TGCGGTGTTC TTCCGAGAGA GCTGGTCAGG AAGCCGTCCC
481 GATCGGGGCA CGCAGTCTCC TTCAGCACCT GTGGCGCCCC CTTCGATGTC GCTCGCGGAG
541 ATCCGAGAAG ACAGTGCTGA ATATCTTCCA GGCAGGCGGA TTCGGCGCCT CCTCTTCTGG
601 AGATATCTTC TGGTCTTTCA CAACATGAAG GCAGGGCTGG GAAGTGAT
```

Amino acid sequence of TunM protein:

```
  1 MPFNHNDHYH DLLLRELPES CRRALEVGCG TGRFAQKLAL RGIEVDAIDV DNDVVTAARE
 61 MNSTLDLPRQ IRFERADITH LTLTPDTYDY IACLASIHHV PFETVQKLRA SLTTNGVLVI
121 LGCYREATIW DHAISLLAVP VNAIWRAAVF FRESWSGSRP DRGTQSPSAP VAPPSMSLAE
181 IREDSAEYLP GRRIRRLLFW RYLLVFHNMK AGLGSD
```

A2.14 tunN: Putative UTP pyrophosphatase

DNA sequence of tunN gene:

```
  1 ATGATCTCTC AGCCACCGAG AAGATATGCC CTGGGGGCCA CCTGCGTTAC GTTCAACGAC
 61 GTGGGACGTG TCCTTATTGC GCGCCGCCGC TCCCCGGAAC GCTGGGAGCT GCCCGGTGGC
121 TTGGTGGACC CGGGCGAGGC ATTCCATGAT GCCGCCACTC GCGAGACCTA CGAGGAGACC
181 GGGGTCCACG TCAAAGTCCA CGGCTTGGTG GGTGTCTACC AGCATCCGAG CCGGGGCATC
241 CTCGCAGGCA TCTTCATCGC GACAGCCTTG TCGGGGGAAC CTCGCCCAAC AGCGGAATCC
301 ATCGCCGCCG AGTGGGCGGA CGTGGAGGAT GCTCTCACGC GACTGCACCC CCTCTATCGA
361 CCGCGCCTTG AAGACGCCCT CACAGCCAGG GCTTCCGTGA CGCTCCGCGT CCATGAAGGA
421 ACGGAGGTCC TCTCACTCTT CGAGGCCACA CACGTC
```

Amino acid sequence of TunN protein:

```
  1 MISQPPRRYA LGATCVTFND VGRVLIARRR SPERWELPGG LVDPGEAFHD AATRETYEET
 61 GVHVKVHGLV GVYQHPSRGI LAGIFIATAL SGEPRPTAES IAAEWADVED ALTRLHPLYR
121 PRLEDALTAR ASVTLRVHEG TEVLSLFEAT HV
```

A3 LC/MS confirmation of heterologous tunicamycin production

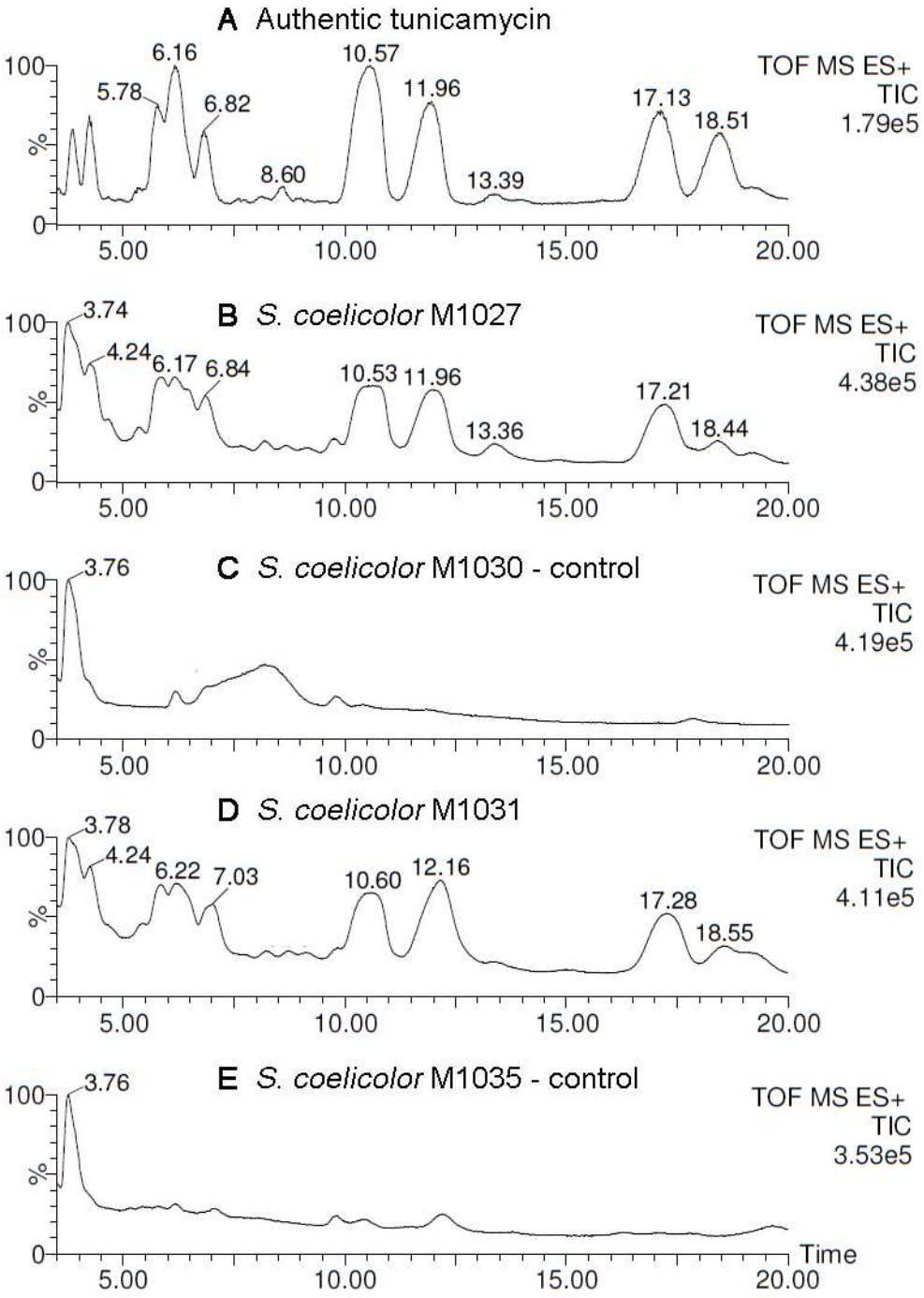


Fig. A3.1 Total ion chromatograms of extracts from recombinant *S. coelicolor* strains (for strain descriptions see Table 2.8 and 2.10)

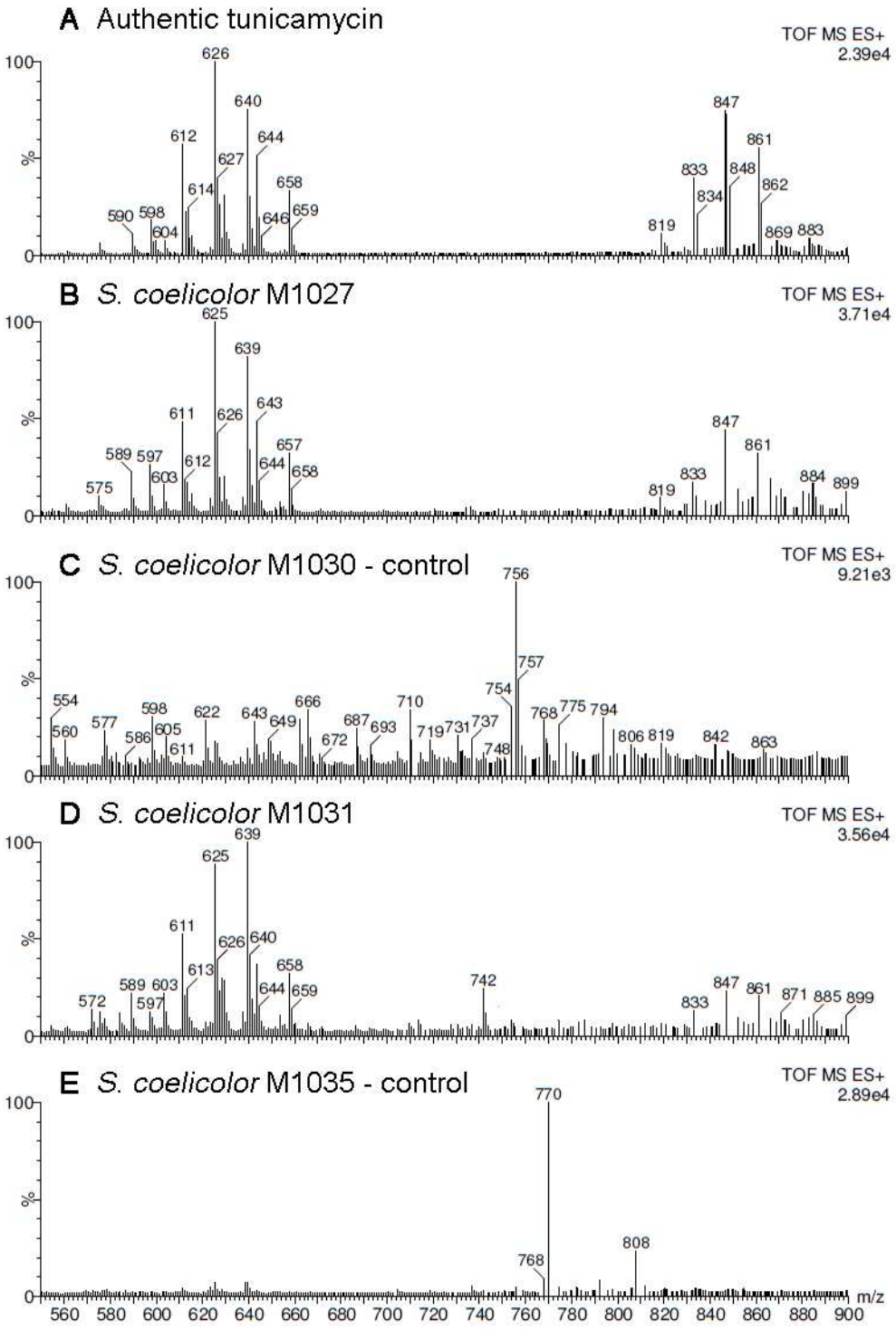


Fig. A3.2 Mass spectra from LC/MS analysis of extracts from recombinant *S. coelicolor* strains (for strain descriptions see Table 2.8 and 2.10)

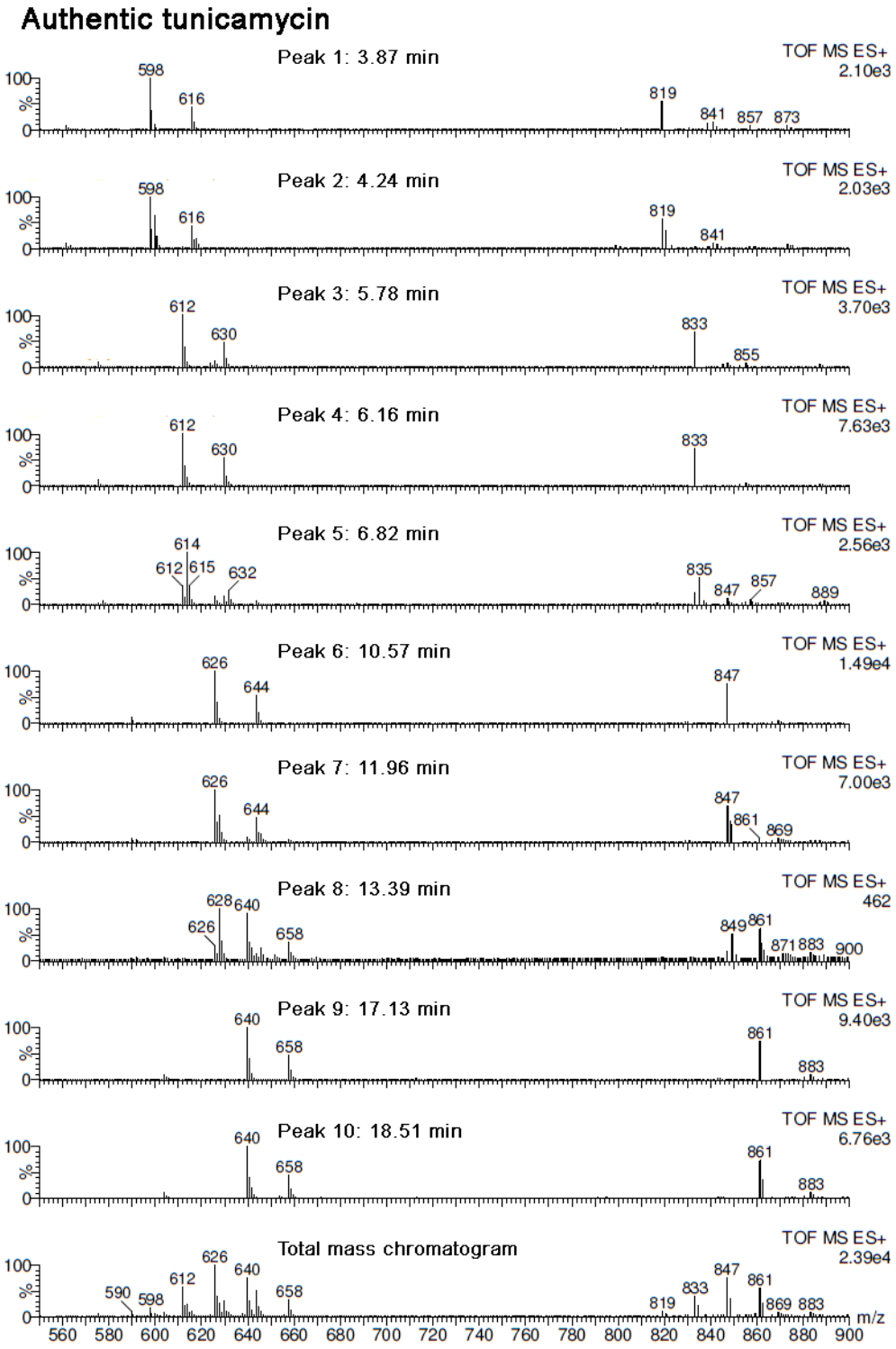


Fig. A3.3 Extracted mass spectra of individual peaks from LC/MS analysis of authentic tunicamycin

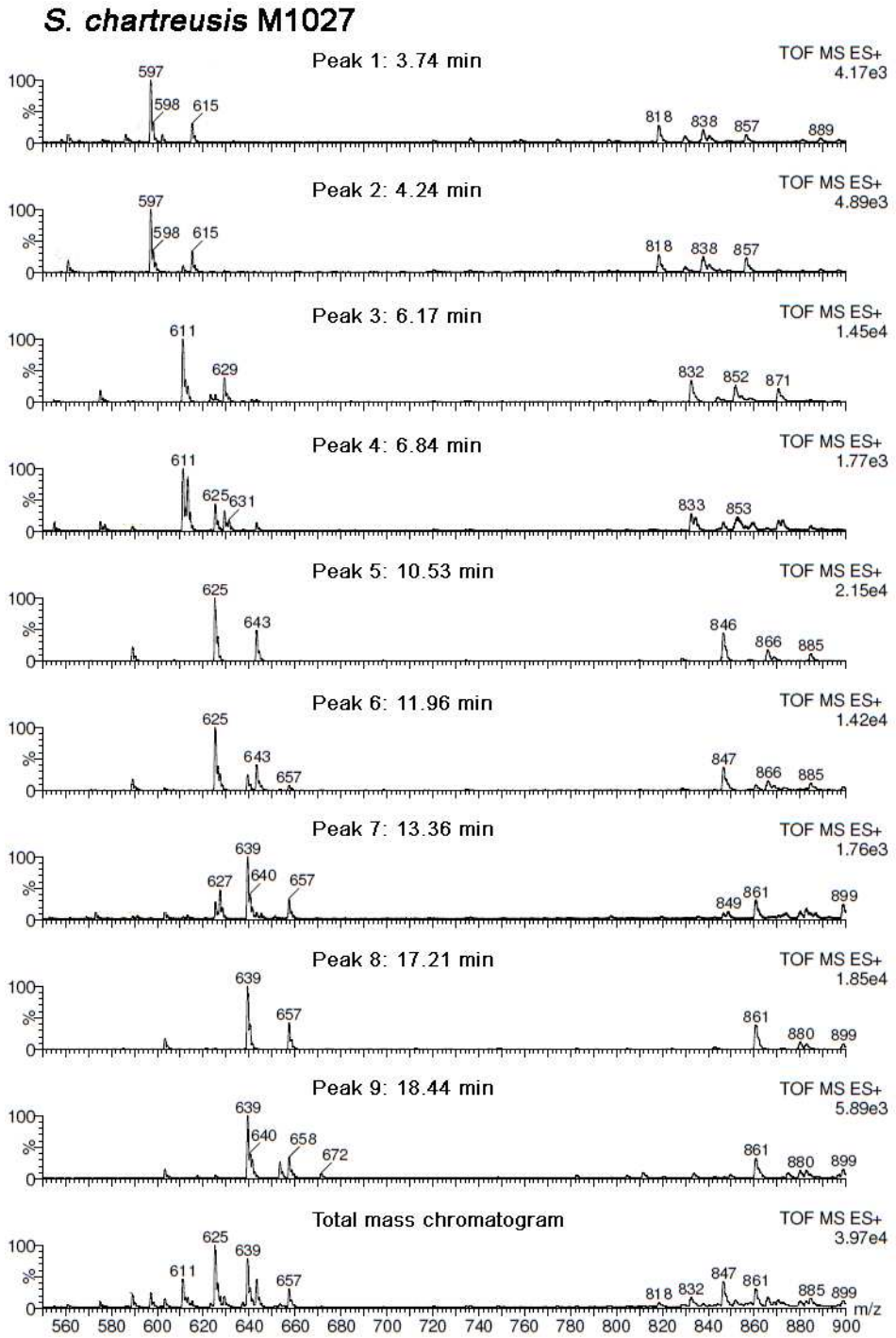


Fig. A3.4 Extracted mass spectra of individual peaks from LC/MS analysis of extracts from *S. coelicolor* M1027 (for strain descriptions see Table 2.8 and 2.10)

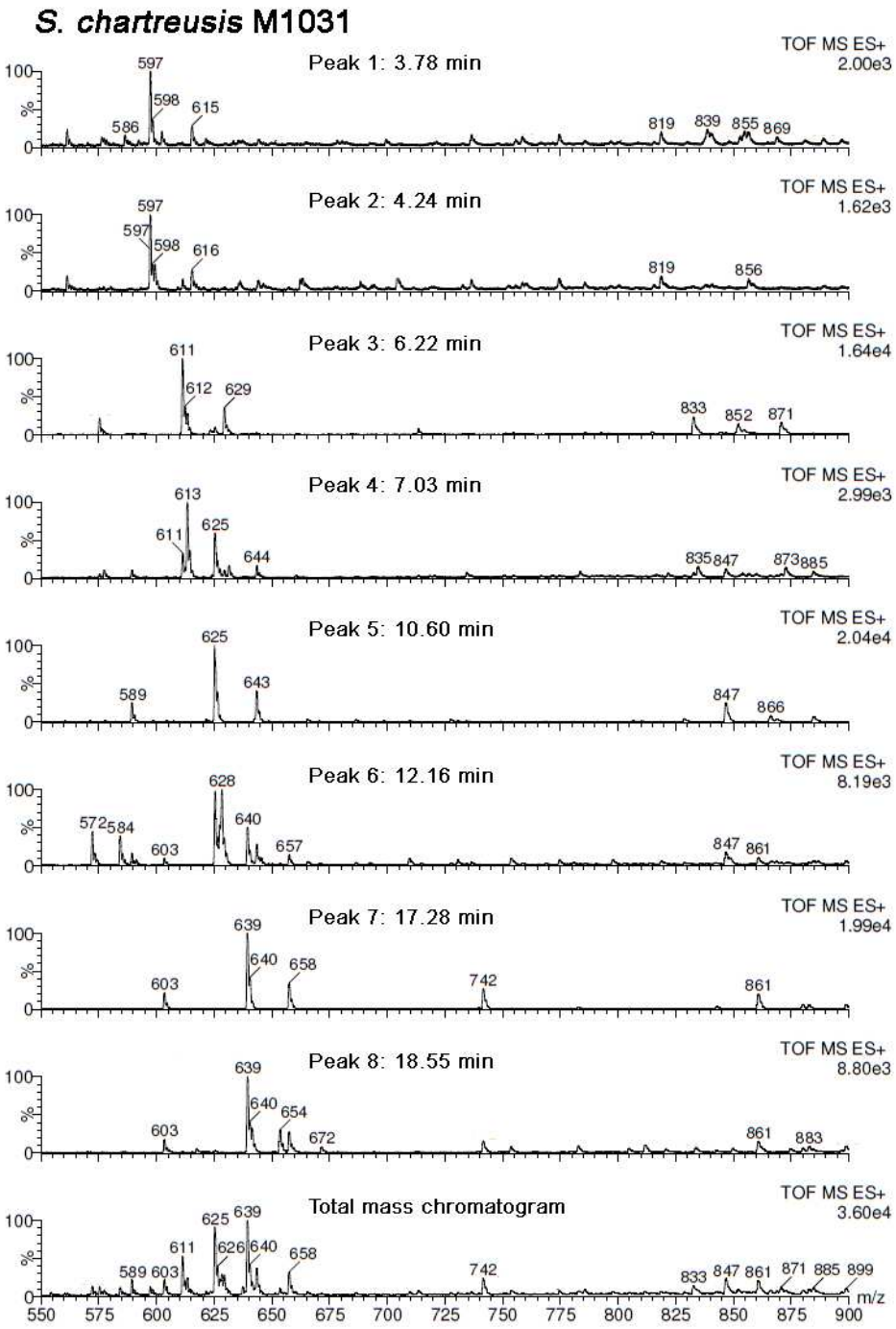


Fig. A3.5 Extracted mass spectra of individual peaks from LC/MS analysis of extracts from *S. coelicolor* M1031 (for strain descriptions see Table 2.8 and 2.10)

A4 ^{1}H NMR confirmation of heterologous tunicamycin production

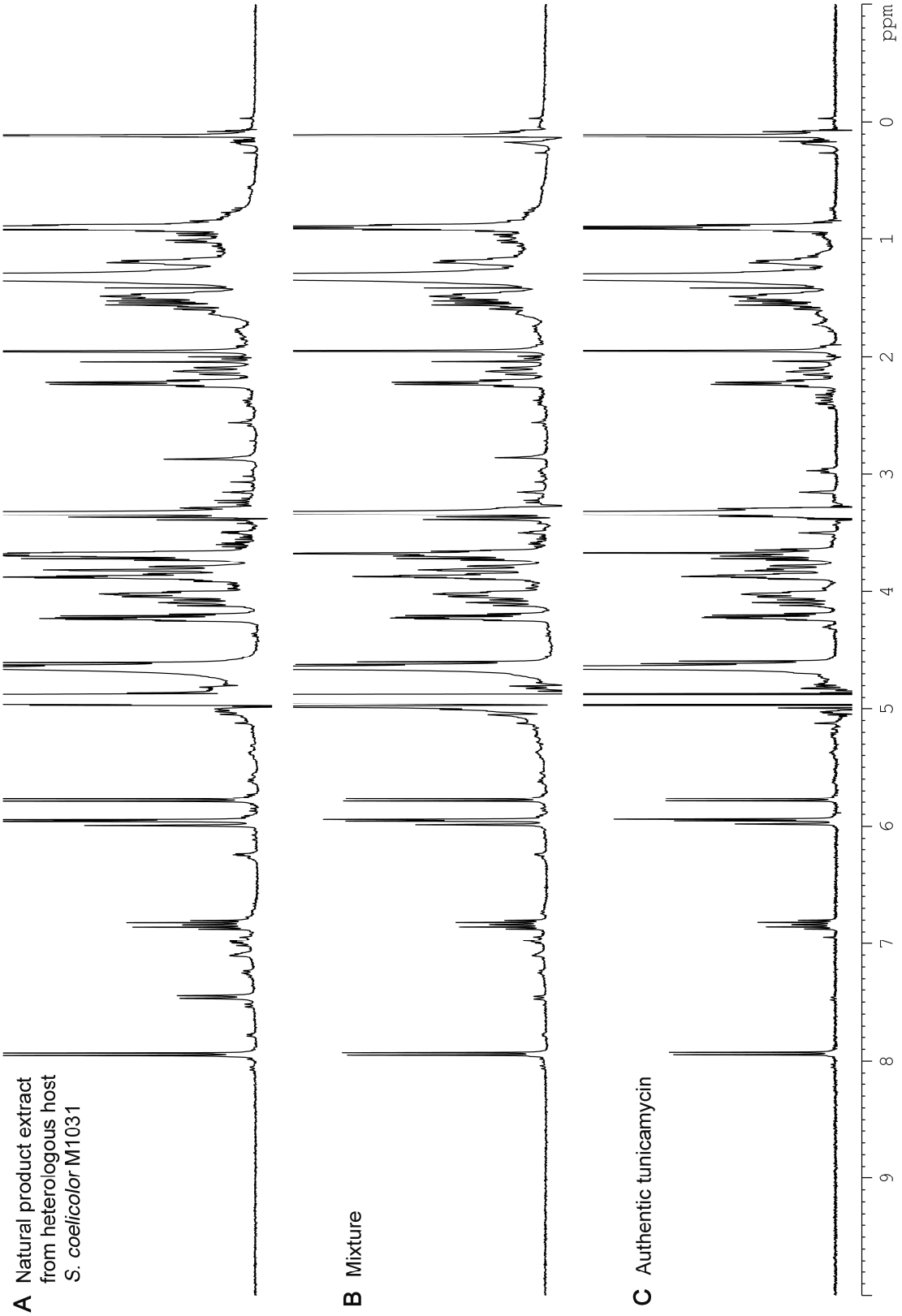


Fig. A4.1 ^{1}H NMR of extract from heterologous host with an authentic tunicamycin sample (see section 2.3.5)

A5 Selected spectra from degradation of isolated tunicamycins

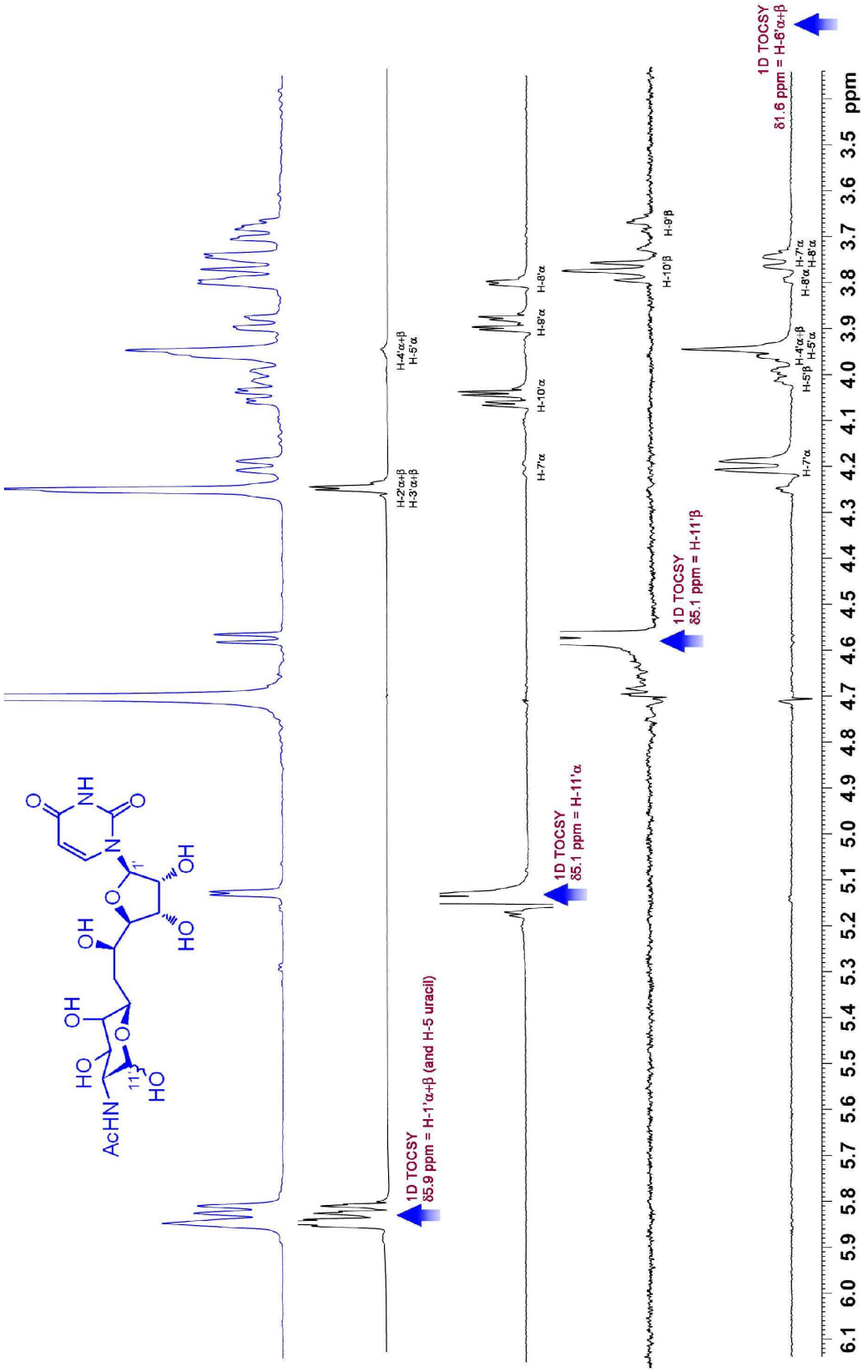


Fig. A5.1 1D TOCSY spectra of *N*-acetyl-tunicaminyl-uracil **90** (see section 3.3.1)

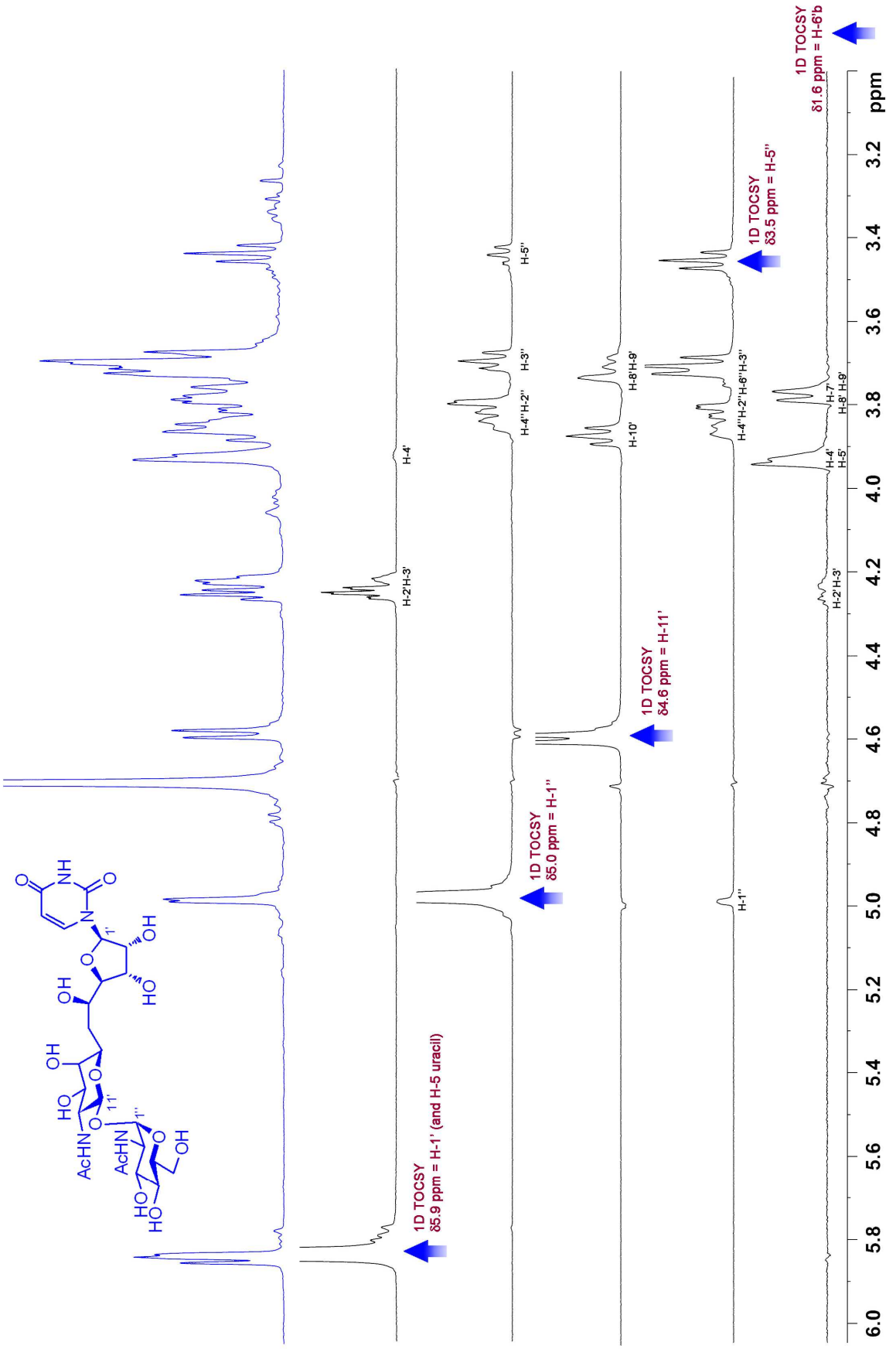


Fig. A5.2 1D TOCSY spectra of *N*10´-Acetyl-*N*10´-deacyl-tunicamycin **99** (see section 3.3.2)

A6 Crystal supplementary information for 1,6-anhydro-2,3-O-isopropylidene-4-O-sulfamoyl-β-D-mannopyranose 108

Table A6.1: Crystal data and structure refinement

Oxford University Chemical Crystallography Laboratory Identification code	5940
Empirical formula	C9 H15 N1 O7 S1
Formula weight	281.29
Temperature	150 K
Wavelength	0.71073 Å
Crystal system	Triclinic
Space group	P 1
Unit cell dimensions	a = 6.0756(2) Å α = 80.5318(12)°. b = 6.6423(2) Å β = 84.2434(12)°. c = 7.3505(2) Å γ = 89.0269(12)°.
Volume	291.120(15) $Å^3$
Z	1
Density (calculated)	1.604 Mg/m^3
Absorption coefficient	0.306 mm^{-1}
F(000)	148
Crystal size	0.40 × 0.25 × 0.24 mm^3
Theta range for data collection	5.483 to 27.494°
Index ranges	-7<=h<=7, -8<=k<=8, -9<=l<=9
Reflections collected	6964
Independent reflections	2545 [R(int) = 0.034]
Completeness to theta = 26.944°	99.0 %
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	0.93 and 0.86
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	2545 / 3 / 170
Goodness-of-fit on F^2	1.0079
Final R indices [I>2sigma(I)]	R1 = 0.0232, wR2 = 0.0567
R indices (all data)	R1 = 0.0238, wR2 = 0.0570
Absolute structure parameter	-0.02(4)
Largest diff. peak and hole	0.20 and -0.24 $e.Å^{-3}$

Table A6.2: Atomic coordinates (× 10^4) and equivalent isotropic displacement parameters (Å^2 ×10^3)

	x	y	z	U(eq)[a]
S(1)	5470(1)	5488(1)	3347(1)	17
O(2)	6667(2)	6972(2)	2000(2)	25
O(3)	3453(2)	6076(2)	4285(2)	27
N(4)	7041(2)	4484(2)	4857(2)	19
O(5)	4993(2)	3601(2)	2357(2)	16
C(6)	3972(2)	3950(2)	614(2)	15
C(7)	1559(2)	4577(2)	924(2)	16
O(8)	399(2)	2948(2)	2178(2)	17
C(9)	91(2)	1510(2)	985(2)	17
O(10)	-563(2)	2670(2)	-665(2)	19
C(11)	356(3)	4696(2)	-823(2)	19
C(12)	2299(2)	400(2)	696(2)	15
C(13)	4215(2)	1870(2)	-23(2)	14
O(14)	4130(2)	2120(2)	-1991(2)	16
C(15)	3523(2)	161(2)	-2387(2)	17
O(16)	2290(2)	-845(2)	-713(2)	18
C(17)	5585(3)	-1070(2)	-2788(2)	24
C(18)	2047(3)	484(2)	-3941(2)	23

[a] U(eq) is defined as one third of the trace of the orthogonalised Uij tensor;

Table A6.3 Bond lengths (Å) for **108**

S(1)-O(2)	1.4284(11)	C(11)-H(111)	0.965
S(1)-O(3)	1.4259(11)	C(11)-H(112)	1.003
S(1)-N(4)	1.5921(12)	C(12)-C(13)	1.5276(18)
S(1)-O(5)	1.5930(10)	C(12)-O(16)	1.4277(16)
N(4)-H(41)	0.90(2)	C(12)-H(121)	0.979
N(4)-H(42)	0.87(2)	C(13)-O(14)	1.4350(15)
O(5)-C(6)	1.4622(16)	C(13)-H(131)	0.97
C(6)-C(7)	1.5227(18)	O(14)-C(15)	1.4404(16)
C(6)-C(13)	1.5300(18)	C(15)-O(16)	1.4432(17)
C(6)-H(61)	0.99	C(15)-C(17)	1.515(2)
C(7)-O(8)	1.4438(16)	C(15)-C(18)	1.5081(19)
C(7)-C(11)	1.530(2)	C(17)-H(171)	0.992
C(7)-H(71)	0.979	C(17)-H(172)	0.988
O(8)-C(9)	1.4251(17)	C(17)-H(173)	0.976
C(9)-O(10)	1.4147(16)	C(18)-H(181)	1.007
C(9)-C(12)	1.5327(18)	C(18)-H(182)	0.999
C(9)-H(91)	1.003	C(18)-H(183)	0.985
O(10)-C(11)	1.4483(17)		

Table A6.4 Bond angles (°) for **108**

O(2)-S(1)-O(3)	118.91(7)	O(10)-C(11)-H(112)	109.9
O(2)-S(1)-N(4)	110.04(7)	H(111)-C(11)-H(112)	112.4
O(3)-S(1)-N(4)	108.49(7)	C(9)-C(12)-C(13)	112.49(11)
O(2)-S(1)-O(5)	107.49(6)	C(9)-C(12)-O(16)	111.96(11)
O(3)-S(1)-O(5)	108.79(6)	C(13)-C(12)-O(16)	103.22(10)
N(4)-S(1)-O(5)	101.75(6)	C(9)-C(12)-H(121)	109
S(1)-N(4)-H(41)	118.4(13)	C(13)-C(12)-H(121)	110.5
S(1)-N(4)-H(42)	109.7(13)	O(16)-C(12)-H(121)	109.5
H(41)-N(4)-H(42)	114.3(18)	C(6)-C(13)-C(12)	114.59(11)
S(1)-O(5)-C(6)	119.67(8)	C(6)-C(13)-O(14)	109.68(10)
O(5)-C(6)-C(7)	111.03(11)	C(12)-C(13)-O(14)	101.99(10)
O(5)-C(6)-C(13)	102.57(10)	C(6)-C(13)-H(131)	110.1
C(7)-C(6)-C(13)	112.05(11)	C(12)-C(13)-H(131)	111.4
O(5)-C(6)-H(61)	108.9	O(14)-C(13)-H(131)	108.7
C(7)-C(6)-H(61)	111.7	C(13)-O(14)-C(15)	106.32(10)
C(13)-C(6)-H(61)	110.1	O(14)-C(15)-O(16)	105.45(10)
C(6)-C(7)-O(8)	108.16(10)	O(14)-C(15)-C(17)	109.85(12)
C(6)-C(7)-C(11)	112.49(12)	O(16)-C(15)-C(17)	109.08(11)
O(8)-C(7)-C(11)	102.51(11)	O(14)-C(15)-C(18)	108.95(11)
C(6)-C(7)-H(71)	110.9	O(16)-C(15)-C(18)	109.48(12)
O(8)-C(7)-H(71)	107.4	C(17)-C(15)-C(18)	113.69(12)
C(11)-C(7)-H(71)	114.7	C(15)-O(16)-C(12)	109.36(10)
C(7)-O(8)-C(9)	102.00(10)	C(15)-C(17)-H(171)	110.1
O(8)-C(9)-O(10)	105.63(10)	C(15)-C(17)-H(172)	110.9
O(8)-C(9)-C(12)	107.19(11)	H(171)-C(17)-H(172)	109
O(10)-C(9)-C(12)	113.24(11)	C(15)-C(17)-H(173)	111.3
O(8)-C(9)-H(91)	107.3	H(171)-C(17)-H(173)	108.3
O(10)-C(9)-H(91)	109.9	H(172)-C(17)-H(173)	107.2
C(12)-C(9)-H(91)	113.1	C(15)-C(18)-H(181)	105.8
C(9)-O(10)-C(11)	107.24(10)	C(15)-C(18)-H(182)	107.2
C(7)-C(11)-O(10)	103.52(11)	H(181)-C(18)-H(182)	110.6
C(7)-C(11)-H(111)	109.3	C(15)-C(18)-H(183)	105.5
O(10)-C(11)-H(111)	109.4	H(181)-C(18)-H(183)	114.2
C(7)-C(11)-H(112)	112	H(182)-C(18)-H(183)	112.8

Table A6.5 Anisotropic displacement parameters ($Å^2 \times 10^3$) for **108**. The anisotropic displacement factor exponent takes the form: $-2\pi^2[h^2 a^{*2} U^{11} + ... + 2 h k a^* b^* U^{12}]$

	U^{11}	U^{22}	U^{33}	U^{23}	U^{13}	U^{12}
S(1)	18(1)	17(1)	17(1)	-7(1)	-3(1)	0(1)
O(2)	34(1)	19(1)	23(1)	-1(1)	-6(1)	-8(1)
O(3)	23(1)	34(1)	28(1)	-17(1)	-4(1)	8(1)
N(4)	18(1)	24(1)	16(1)	-6(1)	-3(1)	0(1)
O(5)	18(1)	16(1)	15(1)	-6(1)	-5(1)	-1(1)
C(6)	16(1)	17(1)	14(1)	-4(1)	-4(1)	-2(1)
C(7)	17(1)	15(1)	16(1)	-2(1)	-1(1)	0(1)
O(8)	15(1)	18(1)	17(1)	-3(1)	1(1)	-1(1)
C(9)	16(1)	18(1)	16(1)	-2(1)	-3(1)	-3(1)
O(10)	15(1)	21(1)	22(1)	-1(1)	-7(1)	-2(1)
C(11)	17(1)	18(1)	22(1)	-2(1)	-6(1)	1(1)
C(12)	17(1)	15(1)	14(1)	-3(1)	-2(1)	-3(1)
C(13)	13(1)	17(1)	11(1)	-3(1)	-1(1)	0(1)
O(14)	19(1)	16(1)	13(1)	-4(1)	-2(1)	-3(1)
C(15)	20(1)	16(1)	15(1)	-4(1)	-2(1)	-3(1)
O(16)	24(1)	17(1)	15(1)	-4(1)	-2(1)	-6(1)
C(17)	27(1)	24(1)	22(1)	-9(1)	-1(1)	4(1)
C(18)	27(1)	24(1)	18(1)	-4(1)	-9(1)	-3(1)

Table A6.6 Hydrogen coordinates ($\times 10^4$) and isotropic displacement parameters ($Å^2 \times 10^3$) for **108**

	x	y	z	U(eq)
H(41)	6430(30)	3660(30)	5870(30)	26
H(42)	8220(30)	3980(30)	4310(30)	26
H(61)	4840	4992	-278	18
H(71)	1423	5799	1511	19
H(91)	-1140	575	1602	20
H(111)	-829	5677	-786	23
H(112)	1397	5002	-1982	21
H(121)	2598	-449	1866	17
H(131)	5622	1258	272	15
H(171)	6537	-1153	-1757	36
H(172)	6438	-444	-3953	36
H(173)	5215	-2457	-2928	36
H(181)	776	1344	-3519	35
H(182)	2931	1240	-5057	36
H(183)	1618	-894	-4106	34

Table A6.7 Hydrogen bonds for **108** [Å and °].

D-H...A	d(D-H)	d(H...A)	d(D...A)	<(DHA)
N(4)-H(41)...O(14)#1	0.90(2)	2.13(2)	3.006(2)	163
N(4)-H(42)...O(8)#2	0.87(2)	2.14(2)	2.977(2)	160
C(7)-H(71)...O(3)	0.98	2.52	3.137(2)	121
C(13)-H(131)...O(10)#2	0.97	2.5	3.201(2)	129
C(18)-H(181)...O(10)	1.01	2.47	3.276(2)	137
C(18)-H(182)...O(5)#3	1	2.5	3.492(2)	171

Symmetry transformations used to generate equivalent atoms:
#1: x,y,z+1 #2: x+1,y,z #3: x,y,z-1

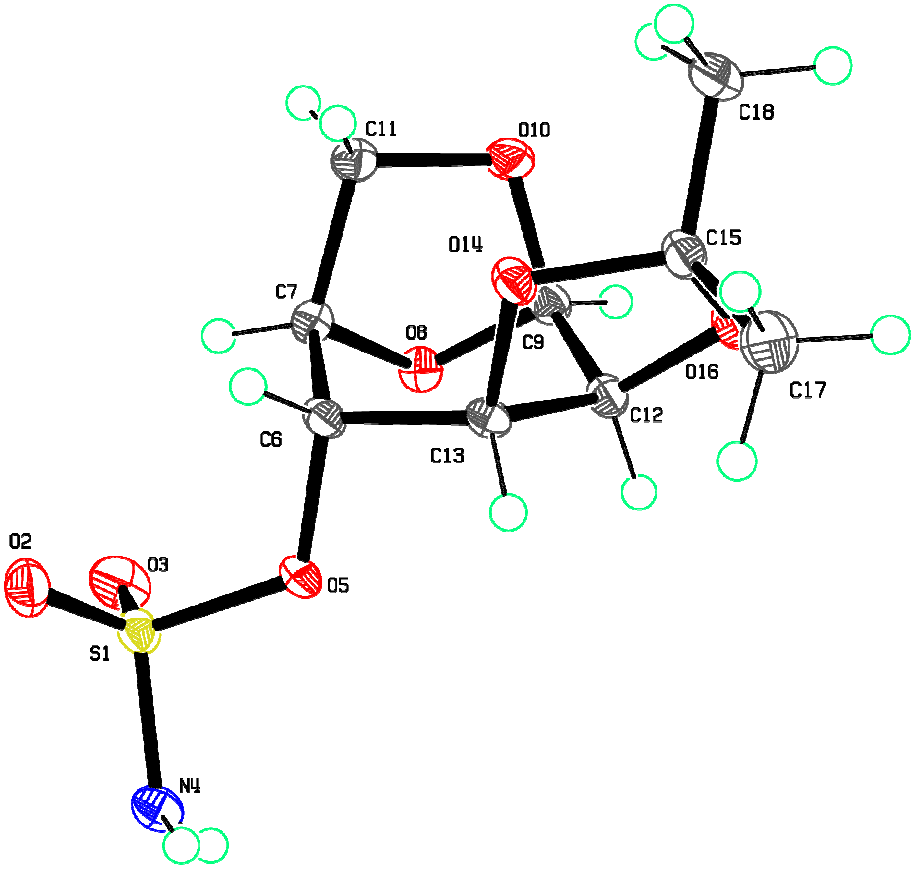


Fig. A6.1 ORTEP representation of compound **108**, created in ORTEP-3 v2.02[459]

6.8 References

459. Farrugia L (1997): **ORTEP-3 for Windows - a version of ORTEP-III with a Graphical User Interface (GUI)**. *J. Appl. Crystallogr.* **30**: 565.