# Novel peptide natural products from the NZ thermophilic microorganism *Thermogemmatispora* strain T81

by
Emma Jayne Aitken

Victoria
UNIVERSITY OF WELLINGTON
*Te Whare Wānanga*
*o te Ūpoko o te Ika a Māui*

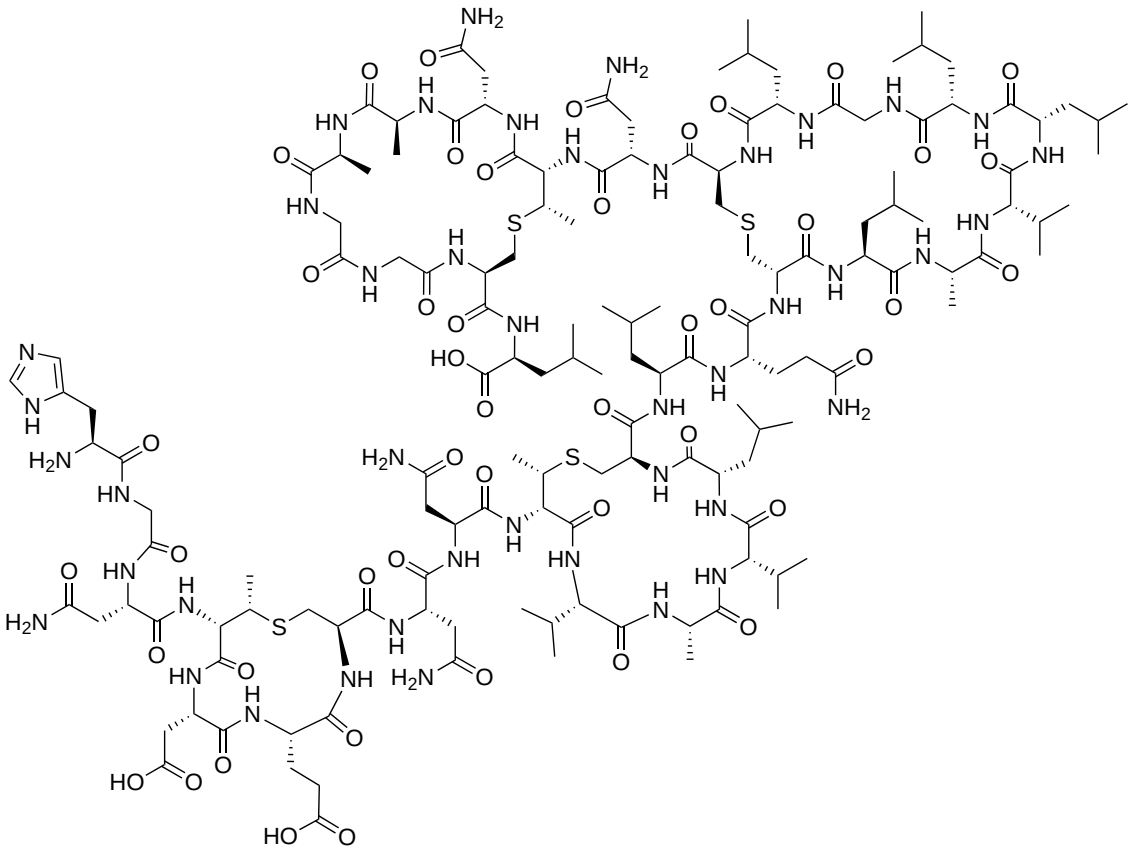
A thesis
submitted to Victoria University of Wellington
in partial fulfilment of the
requirements for the degree of
MSc
in Chemistry.

Victoria University of Wellington
2013

# Abstract

Genome mining of the newly described *Thermogemmatispora* strain T81, a thermophile from the Taupo Volcanic Zone, NZ, revealed the potential to produce novel ribosomally synthesised and post-translationally modified peptide natural products. Previous work established that strain T81 exhibits antimicrobial activity against a wide range of extremophilic bacteria. This thesis describes the mass spectrometry-guided screening of strain T81, and the subsequent isolation and structure elucidation of a novel lanthipeptide, tikitericin (**32**).

Tikitericin is a class II lanthipeptide which bears no sequence homology to known lanthipeptides. Comprised of 35 amino acids, the three-dimensional structure of tikitericin is conformationally restricted by four macrocyclic structures formed by the non-proteinogenic residues methyllanthionine and lanthionine. The amino acid sequence, predicted through bioinformatic analysis, was confirmed by chemical degradation experiments and subsequent tandem mass spectrometry. Characterisation of tikitericin's ring topology was performed by tandem mass spectrometry and stereochemical configuration of the (methyl)lanthionine residues was determined by gas chromatography mass spectrometry.



**32**

[*SS*4-7, *SS*10-15, *S*18-26, *SS*28-34-C$\beta$S]-HGNAbuDECNNAbuVAVLCLQALAVLLGLCNAbuNAAGGCL

# Acknowledgments

Firstly, thank you to my supervisor Rob. Your energy and enthusiasm has been a source of inspiration and your guidance invaluable.

This research was a collaboration between Victoria University of Wellington, GNS Science, and Callaghan Innovation. Genome sequencing of *Thermogemmatispora* strain T81 and bioinformatic analysis of the peptide natural product gene clusters was conducted at GNS Science and forms the basis of this thesis. Thanks to Matt and the extremophile research group at GNS Science for the sequence analysis and culturing of *Thermogemmatispora* strain T81, especially Jean Powers for painstakingly cultivating more than 1400 Petri dishes of bacteria. Thanks to Jason at Callaghan Innovation for supplying strain T81 in liquid media.

To everyone in the Natural Products Research Groups, thank you for your support over the past two years. To Nathaniel Dasyam for your help with bioassays, your patience and humour. Special thanks also to Ian Vorster, Jono Singh, Helen Woolner, Jacqui Barber and Vidhiya Damodaran.

Thank you to Bill Jordan for allowing me to work in your lab, and the Proteomics Research Group members: Danyl McLauchlan, Jonathan Dunne, Christine Stockum and Sarah Cordiner for making me feel welcome, helping me with the MALDI-TOF mass spectrometer, and sharing your lab equipment. To Paul Teesdale-Spittle for allowing me to hijack your biomolecular modelling laptop for three months.

Thank you to the trusts who have supported me financially throughout this project. GNS Science for the Masters Scholarship, Victoria University for the Vic Grad Award, Masters by Thesis and Sarah Anne Rhodes Research Scholarships, SCPS for the Curtis-Gordon Scholarship, and Studylink for the Studylink Bonded Merit Scholarship.

To my family, Martin, Veronica, Robin and Thomas, and Sean for encouraging me and being there.

# Table of Contents

# List of Figures

# List of Schemes

# List of Tables

# Glossary

| | |
|---|---|
| **ATP** | Adenosine triphosphate |
| **BAC** | Bacterial artificial chromosome |
| **BLASTp** | Basic local alignment search tool for proteins |
| **C18** | Octadecyl derivatised silica gel |
| **CHCA** | $\alpha$-cyano-4-hydroxycinnamic acid |
| **COSY** | Correlation spectroscopy |
| **CRI** | Crown research institute |
| **DAD** | Diode array detector |
| $\delta$ | Chemical shift (ppm) |
| **Dha** | Dehydroalanine |
| **Dhb** | Dehydrobutyrine |
| **DNA** | Deoxyribose nucleic acid |
| **eDNA** | Environmental DNA |
| **ESI** | Electrospray ionisation |
| **FDA** | United States Food and Drug Administration |
| **FTICR** | Fourier transform ion cyclotron resonance |
| $\mathbf{GI_{50}}$ | Concentration required to inhibit cell growth by 50% |
| $\mathbf{^1H}$ **NMR** | Proton nuclear magnetic resonance |
| **HP20** | PSDVB stationary support |
| **HPLC** | High pressure (performance) liquid chromatography |
| **HR MS** | High resolution mass spectrometry |
| $\mathbf{IC_{50}}$ | Dose that is inhibitory to 50% of test subjects |
| **IHD** | Index of hydrogen deficiency |
| **IMS** | Imaging mass spectroscopy |
| **Lan** | Lanthionine |
| **LAP** | Linear azol(in)e-containing peptide |
| **LH20** | Crosslinked dextran-based size exclusion resin |
| ***m/z*** | Mass to charge ratio |
| $\mathbf{M_W}$ | Molecular weight |
| **MALDI** | Matrix assisted laser desorption/ionisation |
| **MeLan** | Methyllanthionine |
| **MeOH** | Methanol |
| **NCE** | New chemical entity |
| **NCI** | National Cancer Institute |
| **NMR** | Nuclear magnetic resonance |
| **NRPS** | Non-ribosomal peptide synthetase |
| **ORF** | Open reading frame |
| **OSMAC** | One strain many compounds |
| **PC** | Principal component |
| **PKS** | Polyketide synthase |
| **ppm** | Parts per million |
| **PSDVB** | Poly(stytrene-divinylbenzene) |
| **PTM** | Post-translational modification |
| **Q** | Linear quadrupole |
| **R & D** | Research and development |
| **RiPP** | Ribosomally-synthesised and post-translationally modified peptide |
| **RNA** | Ribose nucleic acid |
| **RP** | Reversed-phase |
| **rRNA** | Ribosomal RNA |

| | |
|---|---|
| **SAM** | *S*-adenosylmethionine |
| **SAR** | Structure-activity relationship |
| **TOF** | Time of flight |
| **VUW** | Victoria University of Wellington |

# Chapter 1

# Introduction

## 1.1 The Importance of Natural Products in Drug Discovery and Development

In 1928, Fleming, Florey and Chain discovered penicillin, the first microbial-derived, pharmacologically active natural product.[1] Henceforth, cultivable microorganisms became a recognised source of active therapeutic molecules with original chemical scaffolds. The value of microbial natural products as sources of chemotherapeutic agents has been immense and undisputable, with the use of natural isolates and analogues such as doxorubicin for cancer therapy and erythromycin, tetracycline and vancomycin as anti-infectious agents.

Chemical molecules obtained from living organisms are divided into two classes. Primary metabolites (proteins, fatty acids, nucleotides and sugars) are globally shared by almost all living species. Secondary metabolites, commonly referred to as 'natural products,' are extremely diversified, species-specific low molecular weight molecules ($< 3000$). Although not essential for reproduction or survival, secondary metabolites give a competitive edge to an organism; a flexible secondary metabolism is therefore advantageous to environmental adaptation. Naturally existing compounds often provide highly specific biological activities, consistent with the proposition that essentially all natural products have some receptor binding capacity.[2,3] Many successful chemotherapeutic agents were originally synthesised to mimic the action of molecules found in nature.[4]

Ongoing analyses by Newman and Cragg provide evidence that natural products remain a major source of human medicine (Figure 1.1).[5,6] Between 1981 and 2010, 52% of the 1355 approved new chemical entities (NCEs) were formally synthetic.[6] However, a large proportion of these NCEs correspond to synthetic molecules that are natural product-inspired (S*, S*/NM, S/NM categories), only 28% can be classified as truly synthetic in origin.[6] Even discounting the influence of biologicals and vaccines, in the area of anti-

infectives 68% of approved drugs are natural product-derived or inspired.[6]

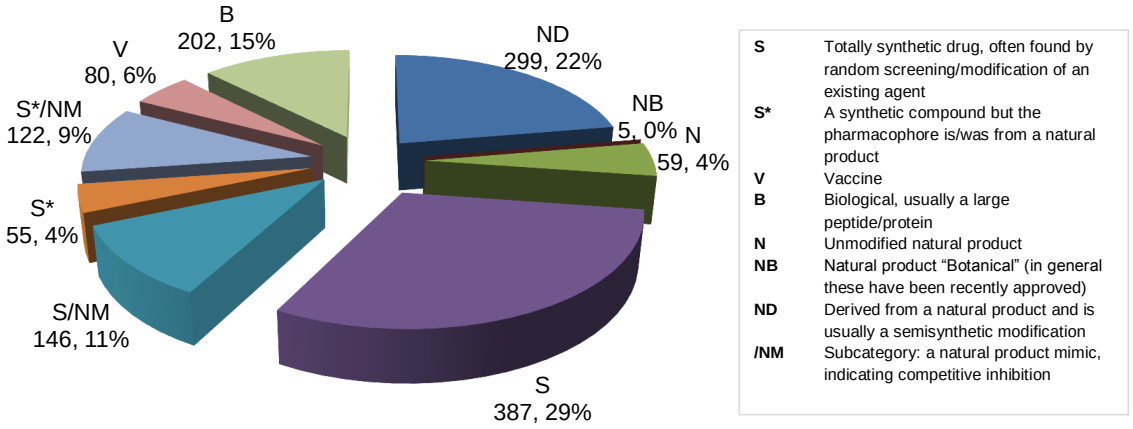



**Figure 1.1.** Sources of all approved drugs between 1981 – 2010, adapted from Newman and Cragg.[6]

Although the historical contribution of natural products is evident, the abandonment of pharmaceutical company natural product R & D programmes over the past 20 years has seen natural products research fall out of favour.[7] Some large pharmaceutical companies have downsized their research for new anti-infective compounds, instead focusing on drugs that combat chronic disease and offer a better return than antimicrobial agents.[8] With recurring incidences of multidrug-resistance and increasing demand for new antimicrobial agents, this is of great concern. The number of NCEs reached a minimum in 2004 with only 24 new NCEs approved,[5] the decline undoubtedly due to the emergence of new chemical techniques such as high throughput screening and combinatorial chemistry that hoped to circumvent the redundant isolation of analogous or minor compounds by natural product isolation methods.[9,10] It is now almost unanimously agreed that the scaffold and functional group diversity generated by combinatorial approaches falls short of natural product complexity.[11] In 2003, Feher and Schmidt published a comparison of the chemical and structural properties of naturally derived compounds versus purely synthetic compounds.[12] The study showed that combinatorial compounds cover a well-defined area of diversity space defined by properties including the number of chiral centers, rotatable bonds, ratio of aromatic atoms to ring atoms, ring fusion degree, number of C-N bonds, C-O bonds, C-halogen bonds, and C-S bonds. In contrast, natural products and drugs cover almost all of, and approximately the same chemical space. It is common knowledge in medicinal chemistry that "the removal of chiral centres, introducing additional flexibility into the molecule and decreasing its size generally leads to a less specific and weaker activity."[12] In contrast, the "generation of natural product diversity has occurred not only within the constraints of available biosynthetic reactions and precursors but also in the context of biological utility."[12] Most natural products have a function, and the biosynthetic routes generating these compounds have undergone countless cycles of mutation and selection to create the molecules we know today.

The realisation that the number of NCEs in drug development was declining may have led to rekindled interest in "rediscovering natural products,"[13] as well as the heightened appreciation of natural product-like scaffolds in "improving efficiency"[14] of diversity-oriented synthesis. Of the 20 small-molecule NCEs approved in 2010, 50% were unmodified natural products.[6] This is not to say that natural products are the only viable source of new drugs. A recent analysis has suggested that natural products preferentially target proteins that are essential to an organism, presumably because these are effective defence substances, and drew the conclusion that "natural products may not display enough versatility to be suitable for treatment of all heritable human diseases."[4,15] The conclusion is undoubtedly correct, and this introduction is not in denial of the contributions of non-natural product-derived drugs to human health. However, the urgent need for the discovery and development of new pharmaceuticals to respond to the current health care situation: increasing problems with chemotherapy (nosocomial strains,* virus evolution, mycobacterial reappearance, tumour resistance, etc.) and emergence of multidrug-resistant pathogens and infectious agents, demands that all approaches to drug discovery be exploited aggressively. Natural product-derived drugs have been especially important in the treatment of non-heritable diseases such as infectious diseases, where their ability to target proteins coded by essential genes is a powerful factor in their success.

The realignment of natural products chemistry with current biological and medical research is likely to rely on advances in genomics and synthesis.[10] Genetic sequencing of well-known natural product-producing bacteria, such as genera *Bacillus* and *Streptomyces*, has revealed gene clusters that encode metabolites vastly outnumbering the molecules discovered thus far. Research that focuses on identifying, detecting and characterising so-called 'cryptic' metabolites will bridge the disconnect between what can be inferred from the genome and what can be detected in the laboratory. Additionally, the application of enzymes with low substrate specificity will provide access to modified natural product analogues that could, in turn, provide new structure-activity relationships (SARs). These same approaches hold promise for expanding chemical diversity through design of new scaffold-rearranging reactions. Such chemical capabilities complement, and expand on, the rapidly moving field of metabolic engineering, which is ever more capable of delivering large quantities of potent and precious complex molecules. It is clear that nature has played, and will continue to play, a vital role in the drug discovery process.

*Also known as hospital-acquired. Nosocomial infection development is favoured by hospital environments and aggravated by the reduced resistance of individual patients.

## 1.2 Sourcing Diverse Biomass: Extremophile Natural Products

### 1.2.1 The unexplored potential of microbial diversity

Estimates generally agree that less than 1% of microorganisms in the environment have been cultivated in pure cultures.[8,16–18] Given that "a handful of soil contains billions of microbial organisms,"[19] and the assertion that "the workings of the biosphere depend absolutely on the activities of the microbial world,"[20] it is clear that the microbial universe presents a vast untapped resource for drug discovery. Improvements in genetic sequencing methods have led to an upsurge in sequenced bacterial genomes, the total currently standing at 4050 (Genomes Online Database).[21] Uncovering the chemical diversity available from microorganisms requires a dual approach of sourcing new microbes and applying metagenomic techniques to enhance our understanding of primary and secondary metabolism, and hence provide opportunities to manipulate these chemical pathways.

Despite the remarkable opportunities that unusual microorganisms present for chemotherapeutic and biotechnological applications, few instances can be reported for actual exploitation. This lack of progress from research findings at the laboratory-scale to the development of pilot and large-scale production is correlated with the difficulties encountered in cultivation.[17] Until recently, the study of natural microbial ecosystems has been severely limited due to an inability to cultivate most naturally occurring organisms.[8]

### 1.2.2 Phylogenetics

To fully appreciate the intellectual wealth that can be gained from studying microorganisms, knowledge of phylogenetic and microbial metabolism is beneficial. The construction of a phylogenetic tree based on ribosomal RNA (rRNA) gene sequences shows that all organisms can be divided into three domains: Archaea, Bacteria and Eukarya, which branch from a 'universal ancestor,' Archaea being the most ancient lineage (Figure 1.2).[22] Together, Archaea and Bacteria are termed prokaryotes and comprise unicellular organisms that differ in many aspects, such as their lipid chemistry (ester- or ether-linked in Bacteria or Archaea, respectively), and evolutionary histories. Members of the domain Eukarya share the protein synthesis machinery of Archaea but have ester-linked lipids, a bacterial trait.

For 3.5 billion years, prokaryotes have been key players in Earth's ecology. Their ubiquity on Earth is evidence of their ability to adapt to, and utilise, prevailing energy

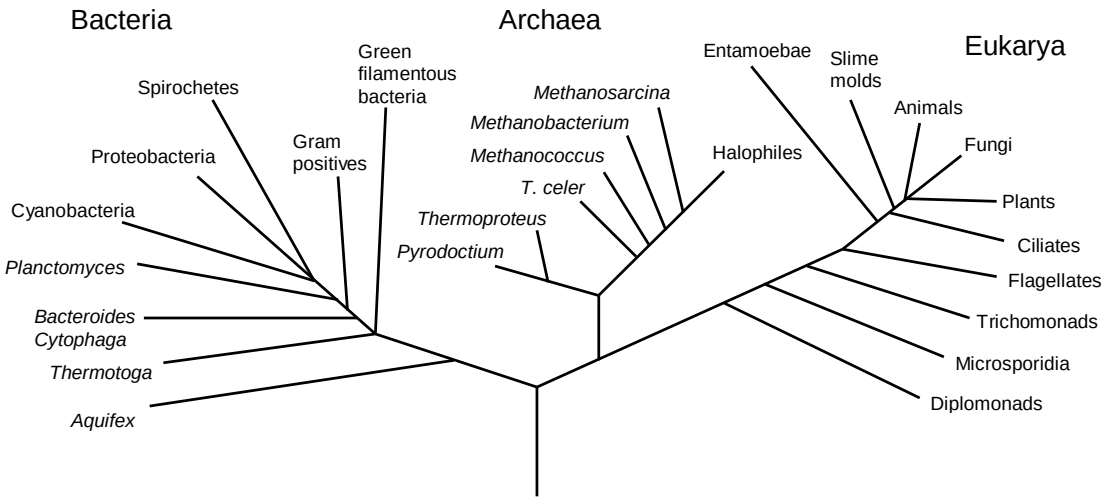



**Figure 1.2.** Universal tree of life (approximate root positions) predicted from rRNA gene phylogenies, adapted from Pakchung *et al.*[23]

and nutritional sources.[24] Compared with the complexity of a multi-cellular eukaryote, the molecular machinery of a single-celled prokaryote is relatively simple, a factor that contributes to the adaptive faculties of these organisms.[23] Within this variable energy-nutrition regime, prokaryotes can exploit distinct biochemical energy-producing mechanisms, such as respiration, using a variety of terminal electron acceptors, such as $O_2$, $NO_3^-$, $SO_4^{2-}$, $Fe^{3+}$ or fermentation reactions.[†] The energy required by organisms, derived from either light or chemical sources, is stored as ATP for use in subsequent anabolic processes. Discovery of life in seemingly prohibitive environments continues to challenge conventional concepts of growth-limiting conditions; the metabolic diversity of prokaryotes provides an insight into how Bacteria and Archaea have been so successful in colonising microenvironments on Earth that are viewed as 'extreme.' In contrast, human survival is limited to a single energy source and a single terminal electron acceptor, classifying humans as obligate heterotrophic aerobic respirers.

### 1.2.3 Extremophiles

R. D. MacElroy first coined the term 'extremophile' in a 1974 paper entitled "Some comments on the evolution of extremophiles," and in doing so, formally designated environments hostile to man as 'extreme.'[25] Extreme habitats are characterised by the presence of physio-chemical stressors (abiotic) but differ in aspects of the biotic environment, such as resource availability or the presence of competitors. Some of these environments, such as hot springs, arctic, acidic and alkaline water, saturated salt brines and pressurised abyssal waters, had originally been considered too extreme to support any form of microbial life. With the improvement of culture conditions, however, came the discovery that these hostile environments were, in fact, full of 'extremophilic'

[†]All primary energy-yielding reactions utilised by living organisms are oxidation-reduction reactions.

microorganisms that are adapted to grow optimally at, or near to, the extreme ranges of environmental variables.[‡ 26] Most extremophiles are of the domain Archaea and the terms are often used interchangeably, a misnoma since species of Bacteria and Eukarya too have extremophilic life cycles; the presently known upper optimum growth temperature is 113 °C for Archaea, 95 °C for Bacteria and 62 °C for single-celled Eukaryotes.[23,26]

An extremophile is considered to be an organism whose optimal growth conditions are found outside of 'normal' conditions: temperature = 37 °C, pH = 7.4, pressure = 1 atm.[24] Definitions of extreme are, of course, anthropocentric because from the point of view of the organism its environment is that to which it is adapted and hence is completely normal. The literature contains several terms that are used to describe extremophiles and extremotrophs and sub-definitions exist for organisms that present moderate, extreme, hyper-extreme, and/or obligate extremotrophy (Figure 1.3). The phenomena of polyextremophily refers to organisms that are adapted to two or more environmental extremes, for example, organisms living in geothermal water and soils thrive in temperature, pressure and pH extremes, as well as tolerating high concentrations of heavy metals. When describing extremophilic and extremotrophic organisms, care must be taken to discriminate between the mere presence of an organism, and those that are growing or metabolically active in an extreme environment.

| | |
|---|---|
| **Acidophile** | an organism with optimal growth at pH $\leq$ 3 – 4 |
| **Alkaliphile** | an organism with optimal growth at pH $\geq$ 10 |
| **Endolith** | an organism that lives inside rocks |
| **Hypolith** | an organism that lives inside rocks in cold deserts |
| **Halophile** | an organism requiring $\geq$ 1 M salt for growth |
| **Metalotolerant** | organisms capable of tolerating high levels of heavy metals |
| **Oligotroph** | an organism capable of living in nutritionally deplete habitats |
| **Piezophile** | an organism that lives optimally at hydrostatic pressures $\geq$ 40 MPa |
| **Psychrophile** | an organism with optimal growth at temperatures $\leq$ 10 °C |
| **Thermophile** | an organism that thrives in the temperature range 60 °C – 85 °C |
| **Toxitolerant** | organisms able to withstand high levels of damaging agents |

**Figure 1.3.** Classification of extremophilic organisms. Adapted from Pakchung *et al.* and Horikoshi *et al.*[23,26]

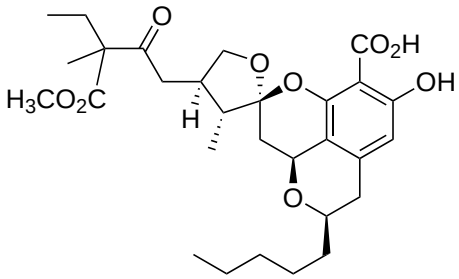
### 1.2.4 Recent advances in microbial natural product isolation

Extremophiles have evolved biomolecules and molecular mechanisms to ensure metabolic flux is maintained at conditions that would ordinarily be considered abiological, these adaptations are reflected in the sophistication of extremophile genomes.[27] *Thermotoga*, *Sulfolobus* and *Picrophilus* species have more complex biochemistries than mesophiles

[‡] A large diversity of organisms are known that can tolerate extreme conditions, and grow but not necessarily optimally, in extreme habitats. These organisms are defined as extremotrophs.[26]

in terms of unique proteins, metabolites and enzymes (with up to 40% genes coding for hypothetical proteins).[27] Research has been heavily weighted towards the isolation of thermophilic and hyperthermophilic enzymes, yet novel metabolites sourced from extremophiles will undoubtedly offer a rich resource for chemotherapy and applications in biotechnology.

The diversity of secondary metabolites isolated from extremophilic microorganisms has been reviewed by Wilson and Brimble.[28] Don and Andrea Stierle, leaders in the field, have isolated more than 20 bioactive novel compounds from a single extremophilic fungus, *Penicillium rubrum*. The fungus was isolated from the waters of Berkeley Pit Lake, an abandoned copper mine that has filled with leached groundwater, now acidic (pH 2.7) and contaminated with heavy metals (Cu, Cd, As, Al, Fe, Zn). The extreme environment naturally selects for new species or varieties of microbes that could be effective bioremediators of the mine wastewater in which they grow. Their metabolic by-products could impact the overall ecology of the Pit Lake by raising the water's pH, providing nutrients for other heterotrophs, or by sorbing metal contaminants. Bioassay-guided isolation led to the discovery of a novel spiroketal, berkelic acid (**1**), which exhibited selective activity in enzyme inhibition assays for two different signal transduction enzymes: matrix metalloproteinase-3 ($GI_{50}$ 1.87 $\mu$M) and caspase-1 ($GI_{50}$ 98 $\mu$M), as well as against the ovarian cancer cell line, OVCAR-3 ($GI_{50}$ 91 nM).[29]




**1**

## Optimisation of culturing conditions

Cultivation of extremophilic microorganisms is confounded by many issues. Firstly, there are many species, both known and unknown, for which traditional culturing techniques are not suitable, or which have entered a nonculturable state.[30,31] Specialised culturing conditions are time-consuming and expensive, often requiring tailor-designed equipment such as high temperature fermentation processes or corrosive media. Secondly, even at very low innoculum extremophiles are known to be characterised by low growth rates and, as a result, product inhibition and low biomass production is often observed.[8] The practical application of extremophilic biomasses to produce enzymes and biomolecules is therefore strictly dependent on the ability of interdisciplinary groups to improve yield to permit secondary metabolite characterisation through techniques such as high-resolution

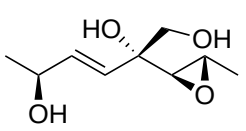
mass spectrometry (HR MS) and nuclear magnetic resonance (NMR), and subsequently give essential information for pilot-scale production and scale-up processes.[17]

A major challenge is the mimicry of 'extreme' environments from which extremophiles are collected. To overcome the limiting factors in cultivating deep-sea thermophilic piezophiles on a large scale, Clark's group at the University of California, Berkeley, are developing a unique production-scale bioreactor capable of continuously supplying growth medium and removing biomass at pressure exceeding 600 bar and temperatures of up to 200 °C.[32] The bioreactor will make a significant contribution in the assessment of culture conditions for deep-sea extremophiles, permitting research under the optimal growth conditions.[32] In an alternative approach, the research of Fenical, Jenson and the Stierles has been able to utilise standard continuous fermentation bioprocesses, demonstrating that extremophiles do not necessarily require extreme culture conditions. Continuous culture processes are especially effective in highlighting the correlation between substrates and biosynthesis of enzymes and other metabolites.

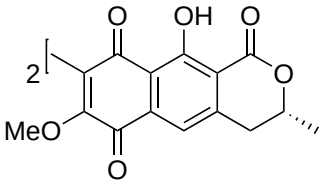
## Triggering cryptic natural product biosynthesis

Advances in genomics have revealed that the number of bacterial genes encoding for secondary metabolites greatly outnumbers the metabolites that have been isolated from them. This is true, for example, for the bacteria that make avermectin and erythromycin,[4] the conclusion being that the biosynthetic potential of microorganisms has been greatly unexplored using traditional approaches.[33] One reason for this is that only a fraction of biosynthetic pathway genes are expressed under laboratory culture conditions. Such silent genetic loci are referred to as 'cryptic' or 'orphan' pathways. To gain access to this untapped reservoir of potentially bioactive structures, the biosynthesis of these putative metabolites needs to be induced.[33] Readers are directed to recent reviews for further information.[33,34]

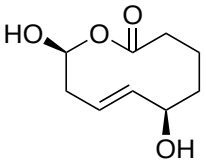
Laboratory cultivation parameters, both nutritional and environmental, are critical to the quantity and metabolic profile produced by microorganisms.[33] If a microbial strain can be cultured under non-natural laboratory conditions, it is important to be aware that its metabolic profile will vary from that of the natural isolate. Zeeck and co-workers deliberately elaborated fermentation parameters to increase the number of microbial metabolites, and coined the term 'OSMAC' (One Strain Many Compounds) for this approach. By altering easily accessible cultivation parameters, such as media composition, aeration or temperature, they were able to isolate unknown natural products from various fungi and actinomycetes.[35–37] Applying several culture conditions to an *Aspergillus ochraceus* strain known to biosynthesise only aspinonene (**2**), resulted in the biosynthesis and identification of 15 additional compounds of diverse biogenetic origin, including xanthomegnin (**3**), aspinolide A (**4**), asperloxin A (**5**), and aspergamide C (**6**).[35]



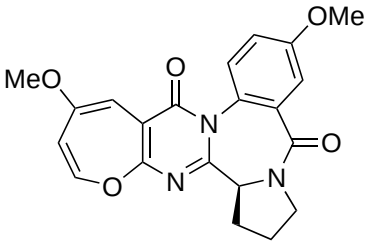
**2**



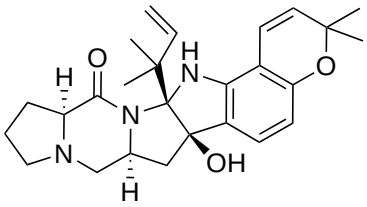
**3**



**4**

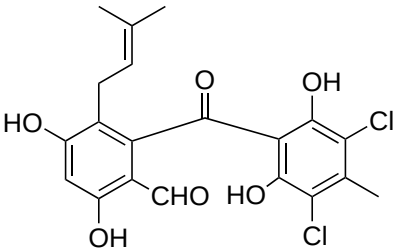


**5**

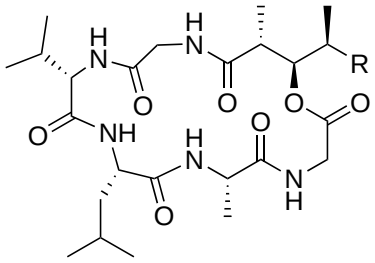


**6**

The study of secondary metabolite profiles in complex situations such as microbial communities is intriguing and the interaction between organisms of the same or different species seems to have resulted in huge natural product diversity (e.g. pheromones, predator-prey molecules, metabolites of symbiotic associations).[33] The idea of inter-species communication leading to chemical diversity has been applied to the laboratory; Fenical and colleagues investigated whether mixed fermentations can stimulate the production of secondary metabolites. When they cultured a marine *Pestalotia* species together with an unidentified, antibiotic-resistant marine bacterium, the biosynthesis of pestalone (**7**), a novel benzophenone, was elicited.[38] This compound, which displayed potent antibacterial activity against methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant *Enterococcus faecium*, could not be detected when either strain was cultured individually.[38] In a separate study, the production of two new cyclic depsipeptides, emericellamide A (**8**) and B (**9**), from a marine-derived Emericella species by challenging the fungus with the actinomycete *Salinispora arenicola*.[39]



**7**



**8** R = $CH_3(CH_2)_3CH_3$
**9** R = $(CH_3)_2(CH_2)_5CH_3$

## Genomic approaches

Combining chemical screening of extract libraries with metagenomics[§] has emerged as a powerful tool to circumvent the issue of persistent cryptic gene clusters and nonculturable

[§]Metagenomics involves the direct cloning of eDNA into large libraries to facilitate the analysis of expressed functionalities.

microbes.[8,33] With the application of efficient cloning vectors such as bacterial artificial chromosomes (BACs) and fosmids, together with enhanced DNA isolation and high throughput screening methodologies, large eDNA fragments can be expressed and the libraries generated subsequently screened for bioactivity.[8] Advances have been made towards understanding of the role of multifunctional polyketide synthetases (PKSs) in bacterial aromatic polyketide biosynthesis, and many such enzymes have been identified, together with their encoding genes. The same applies to non-ribosomal peptide synthetases (NRPSs) responsible for the biosynthesis of NRPs. Through the analysis of microbial genomes, gene clusters encoding for polyketides, NRPs, and hybrid polyketide-NRP metabolites have been identified, thereby providing the tools for engineering the biosynthesis of novel 'non-natural' natural products through gene shuffling, domain deletions and mutations (a general strategy is provided in Figure 1.4).[34]

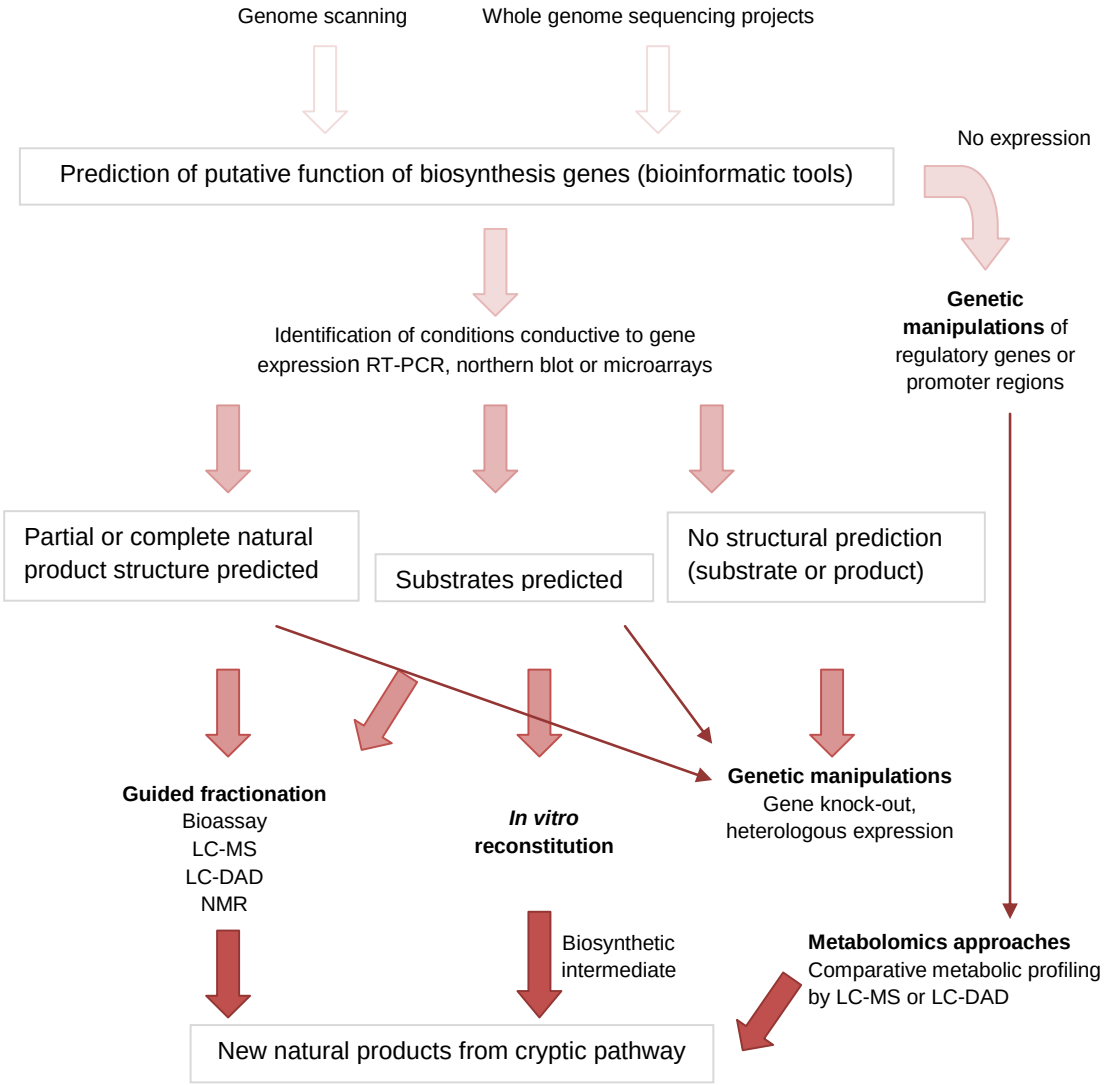
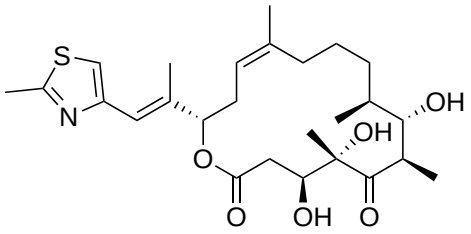


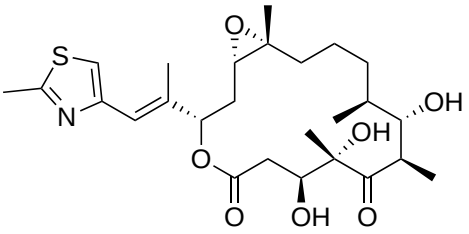

**Figure 1.4.** Genome mining-based strategies for natural products discovery, adapted Corre and Challis.[34]

The power of combinatorial biosynthesis, when applied to natural products, was demonstrated in the scale-up production of epothilone D.[40] Epothilone D entered clinical trials

as a potential anti-cancer agent but has now been discontinued in favour of a congener, 9,10-didehydroepothilone D, the most active of the epothilone series isolated from the myxobacterium, *Sorangium cellulosm*. The isolation and sequencing of the polyketide gene cluster producing epothilone B from two *S. celulosum* strains has been reported. The role of the last gene in the cluster, epoK, encoding a cytochrome P450, is the epoxidation of epothilone D (**10**) to epothilone B (**11**). Heterologous expression of the gene cluster minus the epoK in *Myxoccus xanthus* resulted in large-scale production of crystalline epothilone D.[40]

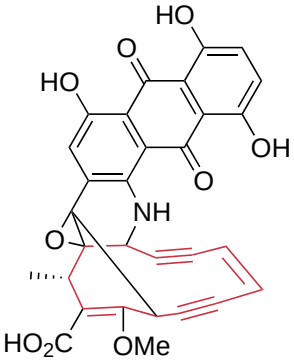



**10**

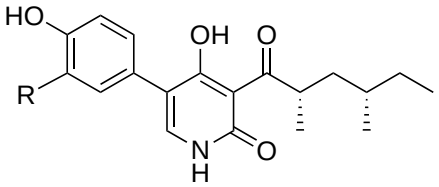



**11**

*In silico* prediction of putative chemical structures encoded by cryptic genetic loci enables the directed search for relevant metabolites.[41,42] Zazopoulos *et al.* analysed the genomes of several *Actinomycetes* for genes putatively involved in the biosynthesis of enediyne anti-tumour agents.[42] Optimised growth conditions finally induced the expression of the gene clusters, resulting in the production of enediynes, for example calicheamicin (**12**).[42] A more targeted approach is the activation of pathway-specific regulatory genes.[33] Such genes encoding putative activator proteins are present in many secondary metabolite gene clusters. Their over-expression may allow the concerted expression of all pathway genes. The advantage of this technique is that only a small gene needs to be handled and an ectopic integration is sufficient, bypassing all limitations of homologous recombination. The success of the strategy has recently been demonstrated in the genome mining of *A. nidulans* for cryptic secondary metabolite genes, revealing the presence of a putative hybrid PKS-NRPS gene. As no corresponding natural product could be identified, it was assumed that this gene locus is silent under standard fermentation conditions. Within the cluster, the presence of a putative activator gene was noticed, and homologous overexpression of this gene under the control of an inducible promoter led to the activation of the entire biosynthetic pathway. Comparative profiling of the metabolite production of the transformant strain under inducing and non-inducing conditions then revealed the production of new metabolites, and scale-up fermentation enabled the isolation of two novel pyridone alkaloids, aspyridones A (**13**) and B (**14**).[43,44] This strategy might be applicable to other orphan gene clusters and thus represents a promising new option to promote the discovery of numerous novel metabolites.




**12**




**13** R = H
**14** R = $CH_3$

**Mass spectrometry-guided isolation**

Modern mass spectrometry has been revolutionised by several developments, most notably the introduction of soft ionisation sources such as electrospray ionisation (ESI) by Fenn *et al.*[45], soft laser desorption (SLD) by Tanaka,[46] and matrix-assisted laser desorption/ionisation (MALDI) by Karas and Hillenkamp.[47] These developments enabled the ionisation, and hence analysis, of smaller biomolecules such as drugs and metabolites, as well as larger biomolecules such as lipids, peptides and even proteins. MALDI is an ionisation technique that is capable of producing intact higher molecular weight ions through the use of a pulsed laser beam combined with energy absorbing matrix molecules.[48–51] A biological sample is mixed with an acidic matrix solution and applied to a target plate. As solvent molecules evaporate from the matrix solution, analyte molecules are homogeneously incorporated into the growing matrix crystals; the incorporation of the analyte into the matrix crystals ultimately prevents fragmentation of larger analyte molecules during desorption. The resulting crystal surface is irradiated with pulsed laser light of sufficient fluence, leading to desorption and ionisation of matrix and analyte molecules. Matrix molecules absorb the laser energy, reducing the direct interaction of the analyte molecules with the laser light, and facilitate the explosive desorption and ionisation process (Figure 1.5). Fragmentation of the acidic matrix molecules typically leads to an excess of protons above the surface that in turn results in gas-phase ion-molecule reactions producing pseudomolecular ions. The ionised molecules are subsequently separated inside the mass analyser, which can be any of the following components: time of flight (TOF), Fourier transform ion cyclotron resonance (FTICR), linear quadrupole (Q), or Orbitrap.

The use of MALDI in conjunction with imaging mass spectroscopy (IMS) by Caprioli *et al.* in 1997 was pioneering in the field of biological imaging and has since been used to identify novel bioactive targets.[51,53,54] IMS collects mass spectral information about the composition and distribution of molecules present at very low concentrations on complex biological surfaces, taking full advantage of high sensitivity instrumentation to detect molecules present in concentrations as low as 500 attomol.[48] IMS instruments can be used

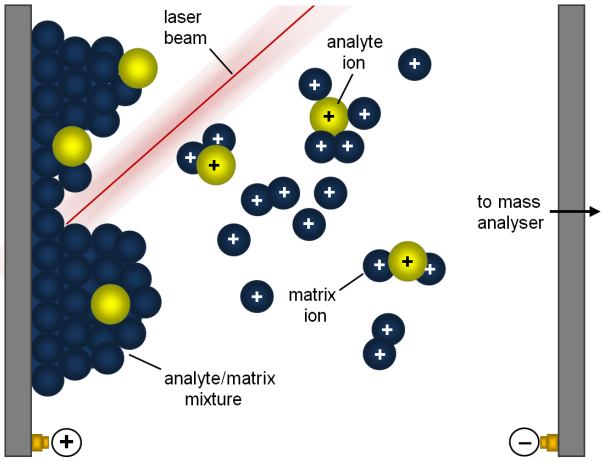



**Figure 1.5.** Schematic representation of the MALDI desorption and ionisation process, adapted from Magnet Lab.[52]

for imaging unknown compounds directly from a biological sample without any *a priori* knowledge or labels, however, the method of whole-cell MALDI-MS is also effective if a molecular target has already been identified. Rather than applying the matrix solution to the surface of a culture, whole-cell MALDI-MS involves mixing a pinhead amount of culture with the matrix solution.[55] In this process, the correlation of metabolites with spatially localised phenotypic changes is reduced. Nevertheless, this technique has been successfully applied to detect peptide antibiotics from *Bacillus* strains.[55]

The many advantages of MALDI-MS present a convincing argument for its superiority in investigating bacteriocins[¶] produced by extremophilic microorganisms. Traditional bioassay-guided isolation focuses on a single activity and, as a consequence, microbial interactions are considered in terms of an individual, predominant chemical activity.[51] Redundant isolation of metabolites is a recurring problem because the activity of minor compounds is often masked by the stronger leads of known bioactives. In contrast, MS enables a known compound to be quickly identified and dismissed as a lead by screening against libraries of natural products (e.g. Antibase[56]) without additional structural knowledge. MALDI-MS is effective for bacteriocins with molecular masses ranging from 500 to 4000 Da, where the mass window is clear and a biomarker can be identified and used to guide the purification process. A MALDI-MS based screening method requires significantly less biomass than bioassay-guided isolation and is therefore particularly appealing given the laborious task of cultivating extremophilic microorganisms. Similarly, an NMR-guided isolation strategy is less suited to microbial studies due to the relative insensitivity of NMR spectroscopy, requiring a large amount of source biomass that is feasible for the study of marine invertebrates but far less practical for extremophilic microorganisms. Finally, the extraordinary sensitivity of MALDI-MS and the very limited handling required for sample preparation opens up great potential to explore the OSMAC theory.

[¶]Bacteriocins are proteinaceous toxins produced by bacteria to inhibit the growth of similar or closely related bacterial strains.

### 1.2.5 Extremophilic bacteria isolated from the Taupo Volcanic Zone

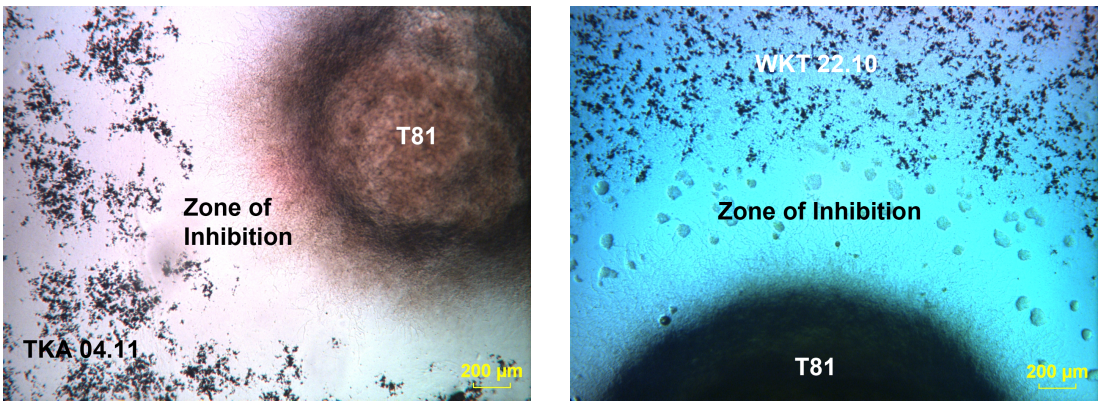
Natural geothermal areas are found in all parts of the globe that are associated with tectonic activity and are usually concentrated in small areas. New Zealand is internationally reputed for its incredible array of geothermal features located in the Taupo Volcanic Zone, associated with the subduction of the Pacific tectonic plate under the Australasian plate. Most of these geothermal areas are colonised by extremophilic microorganisms in the domains Bacteria and Archaea, which have adapted to thrive in prohibitive conditions characterised by high temperature, either high or low pH, significant heavy metal content, and high pressure. GNS Science have compiled an evolving culture library of over 1500 extremophilic microorganisms from New Zealand's unique geothermal and hydrothermal environments, including hot springs, silica sinters, mud pools and steaming ground. The isolated microbial communities are described by rRNA gene sequencing, a culture-independent technique which enables rapid identification of the species present. Through evolution, the random mutation of 16S rRNA genes ubiquitous to Bacteria and Archaea has led to each individual species having a unique 'fingerprint' sequence. Radiata Pool (Figure 1.6), Ngatamariki, has shown exceptional diversity with representative strains from 34 Bacterial phyla and all four Archaea phyla. This endemic NZ biomass provides immense potential for biodiscovery.

**Figure 1.6.** Radiata Pool, Taupo Volcanic Zone.

Stott and co-workers have investigated microbial species diversity in geothermal soils from three areas in the Taupo Volcanic Zone: Waikite, Tikitere and Ngauruhoe.[57] In addition to containing the high concentrations of reduced metals and sulfides observed in most geothermal environments, geothermal soils are rich in organic carbon energy

sources from decaying plant matter that can be utilised by organotrophic bacteria. It was hypothesised that the known diversity of all soil material in combination with the extreme conditions imparted by volcanic activity would support the adaptation of novel microbes, and a large number of chemoorganotrophic thermophilic bacteria have since been isolated and studied. 16S rRNA gene studies were conducted from DNA extracts of the soils and showed that almost all of the detected bacteria represented uncultivated species. Furthermore, the majority were less than 90% similar to any cultured species. Although the three communities differed considerably, phylogenetic analysis showed that the common phyla present were *Chloroflexi*, *Acidobacteria*, *Actinobacteria*, *Proteobacteria*, and the recently described *Armatimonadetes* (previously known as candidate division OP10).

Sequence analysis of a newly discovered bacterial thermophile species isolated from steam-affected geothermal soil (pH 4.5, 55 °C) in Tikitere, *Thermogemmatispora* strain T81, identified the biosynthetic potential for novel ribosomally-produced peptide (RiPP) antimicrobial agents.[58] *Thermogemmatispora* strain T81 is an aerobic, Gram-positive, filamentous spore-forming bacterium which was isolated from geothermally heated soil. On the basis of 16S rRNA gene sequence similarity, T81 was shown to belong to the phylum *Chloroflexi*, to the class *Ktedonobacteria*, and to the genus *Thermogemmatispora*. It is most closely related to *Thermogemmatispora onikobensis* (98.1%) and *Thermogemmatispora foliorum* (96.7%). Preliminiary work conducted by Hauch and Stott identified the ability of strain T81 to inhibit competing strains of extremophilic bacteria and fungi (Figure 1.7).[58] The discovery of biosynthetic potential and observation of inhibition, however, does not connect the source of inhibition with the production of the genetically-encoded RiPPs.




**(a)** TKA 04.11. **(b)** WKT 22.10.

**Figure 1.7.** Co-culture experiments showing co-strain inhibition by strain T81 (GNS Science).[58]

## 1.3 Ribosomally Produced and Post-Translationally-Modified Peptide Natural Products

Research in the 20th century identified four predominant classes of natural products: terpenoids, alkaloids, polyketides, and NRPs. It was incorrectly believed that the structural complexity observed in NRPs could not be reached by ribsomal biosynthesis but, through extensive genome mining, it is now understood that many known anti-infective molecules thought to be NRPs are actually ribosomal in origin. RiPP biosynthetic pathways utilise homologous enzymes to perform post-translational modifications (PTMs). This highlights the potential application of substrate-general enzymes in creating libraries of natural product analogues, in addition to the direct application of novel RiPPs themselves.

RiPPs are initially synthesised as a long precursor peptide, typically containing $\approx$ 20 – 110 amino acids (aa), encoded by a structural gene (Figure 1.8).[59] The precursor peptide contains two regions: the N-terminal leader sequence, and the C-terminal core peptide domain that is transformed into the natural product. The unmodified precursor peptide, designated 'A' (encoded by the XxxA gene), is tailored by a collection of enzymes into a modified precursor peptide abbreviated as mXxxA. In most cases the leader and modified core peptide, still attached, are exported from the cell whereupon the mature peptide is proteolytically released. The leader peptide is thought to play multiple roles in recognition by many of the post-translational modification enzymes, and for export, and immunity.[60]

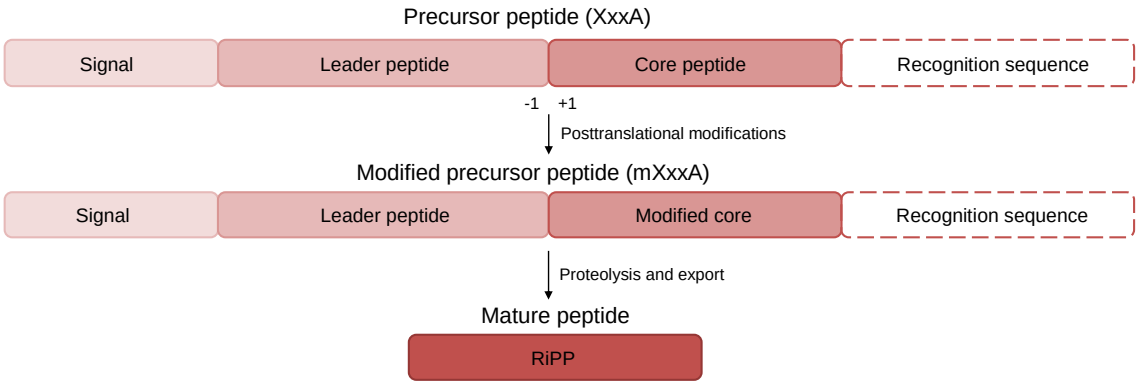



**Figure 1.8.** General RiPP biosynthetic pathway, adapted from Arnison *et al.*[59]

This rapidly expanding class of natural products is now known to include, amongst many others, the lanthipeptides, linear azol(in)e-containing peptides (LAPs), cyanobactins, thiopeptides, bottromycins, and microcins.[59]

## 1.4 Research Aims

The research aim of this thesis was, foremost, to conduct the targeted isolation and characterisation of a genetically-encoded lanthipeptide from strain T81 and, secondly, to investigate production of genetically-encoded thiopeptides by strain T81.

# Chapter 2

# Thiopeptide Screening

## 2.1 Introduction

Thiopeptides were established as a class of natural products in the 1940s after the discovery of microccin.[61] Despite over 100 reported members, the biosynthetic origin of these macrocyclic peptide antibiotics was under debate until 2009 when a series of publications revealed biosynthetic gene clusters leading to their formal recognition as RiPPs.[62–64] Also referred to as thiazolyl peptides or pyridinyl polythiazolyl peptides, thiopeptides are the most heavily post-translationally modified RiPPs known. All thiopeptides are rich in Cys derived thiazole/thiazoline rings, Thr derived oxazole/oxazoline rings, as well as Thr- and Ser-derived dehydroalanine (Dha) and dehydrobutyrine (Dhb) residues (recently reviewed by Bagley[65]). Thiopepides are defined by the oxidation state and substitution pattern of a central nitrogenous ring that forms the core macrocycle and is present as either a piperidine, dehydropiperidine, or pyridine (Figure 2.1). Thiopeptides of the series *a*-*c*, such as thiostrepton A (**15**) and siomycin,[64,66] contain a second macrocycle formed by incorporation of a tryptophan-derived quinaldic acid residue linking one side chain from the core macrocycle to the core peptide N-terminus. Similarly, series *e* thiopeptides contain a second macrocycle formed by an indolic acid residue. The most commonly observed series is *d* (e.g. thiomuracin A (**16**)) featuring the single, core macrocyle, with the exception of cyclothiazomycin (**17**) which contains an unusual tertiary thioether macrocycle.

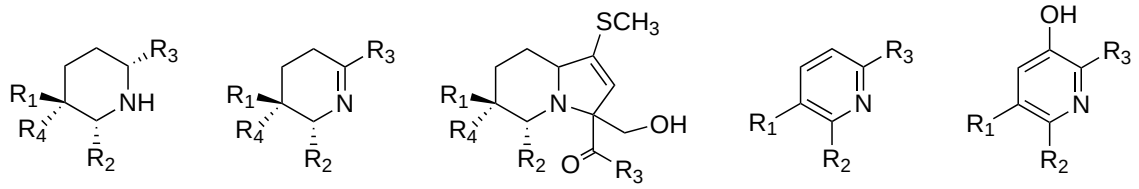
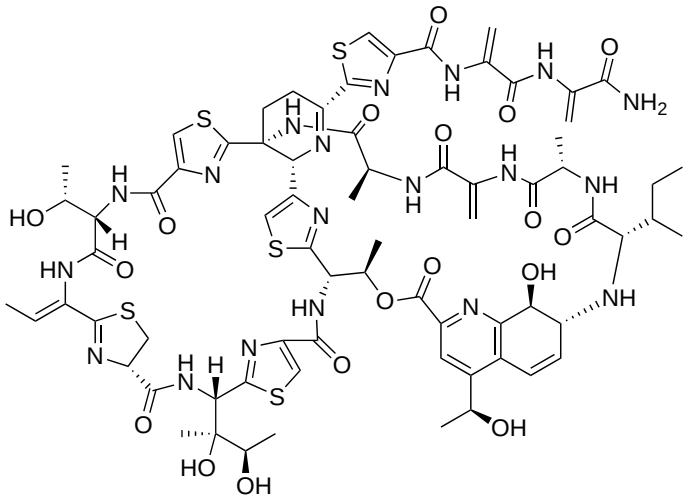
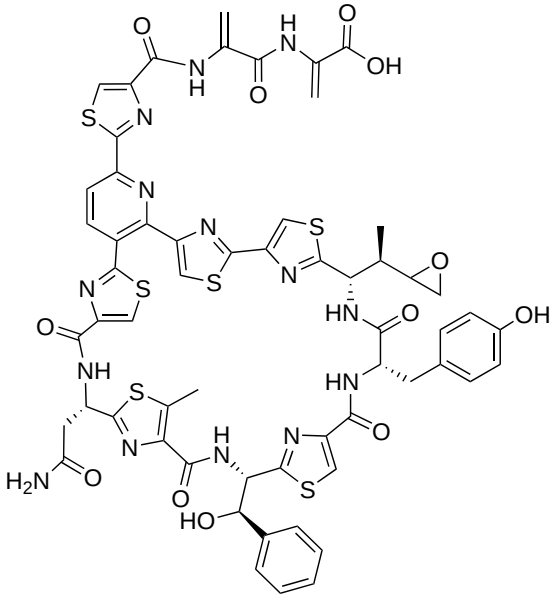


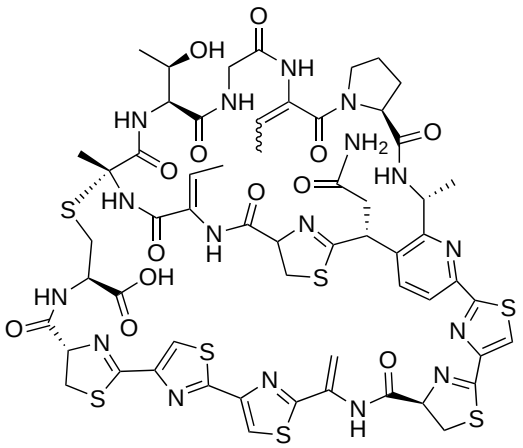

**Figure 2.1.** The structural series of thiopeptides, showing the typical substitution pattern of the central nitrogenous heterocycle. Adapted from Arnison *et al.*[59]



**15**



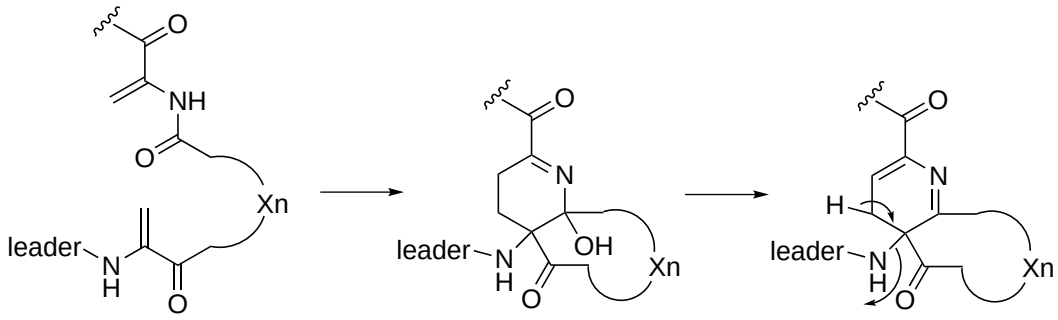
**16**



**17**

To date, ten thiopeptide gene clusters have been described.[62–64,67–72] Insertational mutagenesis experiments performed by Kelly *et al.*,[62] Wieland Brown *et al.*,[63] and Liao *et al.*[64] confirmed the involvement of an array of proteins within thiopeptide gene clusters. In addition to the precursor peptide, at least six proteins are found within each cluster to provide the minimal set of PTMs required to construct the defining thiopeptide scaffold.[67] Just as the mature peptides have features in common with lanthipeptides, linear azol(in)e-containing peptides (LAPs) and cyanobactins, thiopeptide gene clusters also contain homolgous enzymes responsible for these modifications.[73] Dehydration of Ser and Thr residues to Dha and Dhb, respectively, is performed by one or two proteins similar to the class I lanthipeptide LanB dehydratases (see Chapter 3).[59] Proteins similar to the cyclodehydratases and dehydrogenases that construct the polyazoline rings of the cyanobactins and LAPs are predicted to perform a similar role in thiopeptide maturation.[59] The central nitrogenous ring is predicted to form through a [4 + 2] cycloaddition of two Dha residues (Figure 2.1),[74] although it is not yet known whether this process proceeds through a concerted or step-wise mechanism. The function of a sixth protein is unknown and does

not show any significant homology to proteins of known function.[59] Further complexity is introduced by ancillary enzymes, including amidotransferases, cytochrome P450s, nonheme iron-dependent dioxygenases, and *S*-adenosylmethionine-dependent (SAM or AdoMet-MTase) C-, O-, and N-methyltransferases.[62,67,72,75]




**Scheme 2.1.** Bycroft proposal for generation of the pyridine ring in thiopeptides from two Dha residues. Adapted from Arnison *et al.*[59,74]

## 2.2 Identification of a Novel Thiopeptide Locus in *Thermogemmatispora* Strain T81

The genome of *Thermogemmatispora* strain T81 was sequenced by Stott and co-workers (GNS Science) after DNA extraction. Annotation of the draft genome identified a single thiopeptide gene cluster coding for three precursor peptides and an array of modification enzymes responsible for the maturation of all three precursor peptides. The gene cluster is represented in Figure 2.2, annotated in the format proposed by Arnison *et al.* in the 2012 NPR review.[59] Table 2.1 shows the proposed function of each protein, as well as the closest homolog of each gene.

Although uncommon, multiple precusor peptides formed from a single series of biosynthetic genes does have precedence. Multiple copies of thiopeptide precursor genes were reported in the gene cluster of TP-1161, where a second thiopeptide precursor gene (TpaX) was located about 12 kb downstream of TpaA.[69] The two precursor peptides encoded by these genes have the same length (48-aa) and share 70% identity on the amino acid level. Attempts to detect the mature product of TpaX in *Nocardiopsis* sp. strain TFS65-07 fermentation extracts were unsuccessful. Similarly, the proposed berninamycin cluster in the genome of *Propionibacterium acnes* KPA171202, which was identified in a bioinformatics search for thiazolyl peptide-encoding genes by Wieland Brown *et al.*, contains two genes encoding distinct putative precursor peptides.[63] One of these genes is predicted to encode a 47-aa precursor peptide with the 15-aa berninamycin core peptide at its N-terminus, whereas the second 58-residue precursor peptide predicts a 15-aa core peptide which differs in four positions from the structural peptide of berninamycin. A mature peptide product for this second precursor has not been reported so far. The

**Table 2.1.** Deduced peptides and proteins from the thiopeptide gene cluster of T81 (GNS Science).

| ORF | Length (aa) | Proposed Function | Closest homolog in Tpd series | Best Homologue [identities (%similarity, E score), GenBank accession no.] | Trans-membrane helices | Signal peptide |
|---|---|---|---|---|---|---|
| CRISPR | | CRISPR | | n/a | | |
| ORF1 | 314 | Methyltransferase | | Methyltransferase type 11 (Ktedonobacter racemifer) 46% E: 5e-76 ZP_06972504.1 | No | Yes |
| ORF2 | 108 | Unknown | | Possible choline transporter (Sarcophilus harrisii) 29%, E: 7.9 XP_002286313.1 46% E: 5e-76 ZP_06972504.1 | No | No |
| Lan1 | 69 | Thiopeptide precursor peptide | TpdA | thiomuracin prepropeptide ([Nonomuraea sp. Bp3714-39) 40%, E; 5.9, ACS83781.1 | No | No |
| Lan2 | 66 | Thiopeptide precursor peptide | TpdA | thiomuracin prepropeptide ([Nonomuraea sp. Bp3714-39) 40%, E; 0.12, ACS83781.1 | No | No |
| Lan3 | 65 | Thiopeptide precursor peptide | TpdA | thiomuracin prepropeptide ([Nonomuraea sp. Bp3714-39) 42%, E; 1.7, ACS83781.1 | No | No |
| SagD_1 | 731 | Pyridine ring formation | TpdD | protein of unknown function DUF181 (Ktedonobacter racemifer) 36% E: 8e-148 ZP_06965552.1 | No | No |
| SagD_2 | 679 | Cyclodehydratase | TpdG | protein of unknown function DUF181 (Ktedonobacter racemifer) 44% E: 7e-175 ZP_06965551.1 | No | No |
| SagB | 541 | Dehydrogenase: mcbC-like oxidoreductase | TpdE | NADH oxidase (Streptomyces hygroscopicus) 39%, E:1e-58 YP_006249963.1 | Yes | No |
| LdH | 930 | Lantibiotic dehydrogenase | TpdB | lantibiotic dehydratase domain protein( Streptomyces hygroscopicus) 38%, E:2e-08 | No | No |
| ORF3 | 342 | Lantibiotic dehydrogenase | TpdC | Lantibiotic biosynthesis protein (Streptomyces hygroscopicus) 34%, E:7e-43 YP_006249967.1 | No | No |
| ORF4 | 394 | Lantibiotic dehydrogenase | TpdC | Lantibiotic biosynthesis protein (Streptomyces hygroscopicus) 38%, E:2e-476 YP_006249968.1 | No | No |
| ORF5 | 568 | Dehydrogenase: mcbC-like oxidoreductase | TdpE | sagB-type dehydrogenase domain protein (Brevibacillus laterosporus) 33%, E: 6e-59YP_006249968.1 | No | No |
| ABC trans_1 | 327 | ABC-type multidrug transport system, ATPase component | | ABC transporter [Sphaerobacter thermophilus DSM 20745] 47%, E: 3e-92 YP_003320907.1 | No | No |
| ABC trans_2 | 252 | ABC-type multidrug transport system, permease component | | ABC-2 type transporter [Roseiflexus sp. RS-1] 28%, E: 2e-21 YP_001278549.1 | Yes | Yes |
| ORF6 | 415 | Putative dipeptidase | | Dipeptidase (Streptomyces hygroscopicus) 35%, E: 9e-42 YP_006242305.1 | No | No |
| ORF7 | 173 | Unknown | | Polypeptide electron transport complex RsxE subunit [Pseudoalteromonas rubra) 33%, E: 3.3 E: ZP_10296090.1 | Yes | Yes |
| ORF8 | 289 | Unknown | | hypothetical protein Krac_7547 (Ktedonobacter racemifer) 52%, E: 4e-30 ZP_06968715.1 | Yes | Yes |
| ORF9 | 314 | Methyltransferase | | Methyltransferase type 11 (Ktedonobacter racemifer) 45% E: 5e-78 ZP_06973633.1 | No | No |

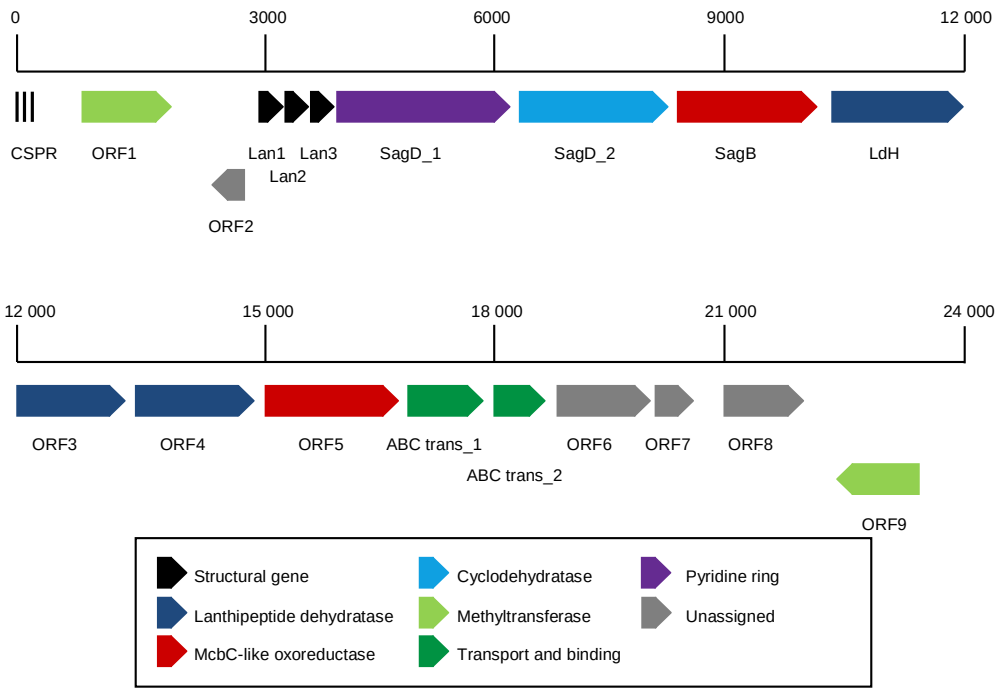



**Figure 2.2.** Thiopeptide gene cluster. Each putative gene is shown to scale, with its position marked in kb. Adapted from original figure by Hauch, GNS Science.[58]

thiocillin gene cluster from *Bacillus cereus* contains four identical and contiguous open reading frames (ORFs) from which an array of eight mature thiopeptides were observed during bioinformatic studies.[63] Seven of these eight thiopeptides had already been discovered but had been isolated in sets of two or three from three different strains of *B. cereus*. The differences arise from a set of three auxiliary modifications that occur on only a subset of the molecules: C-hydroxylation of Val-6, O-methylation of Thr-8, and reduction of the presumed Thr-14 decarboxylation product to an aminoisopropanol. Considering these examples, it was expected that the biosythetic machinery of strain T81 may not transcribe all three encoded precursor peptides under any single growth condition.

Comparative analysis to available thiopeptide gene clusters revealed a cluster of genes located downstream of Lan1 – 3 which code for the necessary maturation enzymes to create a thiopeptide framework (Table 2.1). The putative products of SagD_2, ORF5 and SagB pair bear similarity to cyclodehydratase and dehydrogenase enzymes for construction of polyazoline rings. SagD_2 contains a bacteriocin biosynthesis cyclodehydratase docking scaffold and ORF5 shows similarity to a SagB-type dehydrogenase domain protein. The proteins LdH and ORF3/4 show weak similarity to class I lanthipeptide LanB dehydratases responsible for dehydration of Ser and Thr residues. Formation of the pyridine core may be performed by SagD_1, which includes a SagD docking station, a putative peptidase containing docking domain, and is similar to the YcaO-like family. SagD_1 shows similarity to CltD of cyclothiazomycin (**17**).[70] The gene cluster contains

two proteins of unknown function, ORF6 and ORF7, which do not show significant similarity to the protein of unknown function present in other thiopeptide gene clusters.

In addition to the core modification proteins, the cluster of T81 contains a putative dipeptidase and two N-methyl transferases. The function of a dipeptidase within thiopeptide gene clusters has no precedence in reported gene clusters as yet and shares only weak similarity (35%) to the dipeptidase of *Streptomyces hygroscopius*, the closest homologue (Table 2.1). Flanking the core-modification genes, ORFs 1 and 9 encode for almost identical methyltransferases and are likely to catalyse methylation of the core peptide. These genes bear $\approx 45\%$ similarity to SAM-dependent methyltransferases [*Ktedonobacter racemifer*], enzymes that use *S*-adenosyl-L-methionine (also known as AdoMet) as a substrate for methyl transfer, creating the product *S*-adenosyl-L-homocysteine (AdoHcy). As shown in Table 2.2, genes encoding for methyltransferases have precedence among the available thiopeptide biosynthetic gene clusters. The clusters of the thiomuracins (e.g. thiomuracin A **16**) and GE2270 contain genes encoding radical SAM-dependent proteins with a significant similarity to the coproporphyrinogen III oxidases, known to catalyse unusual methylations, and are thought to be responsible for thiazole methylation.[63,67] Specific to GE2270A is TpdT, encoding an N-methyltransferase thought to catalyse the methylation of the asparagine side chain, and an additional radical SAM-dependent protein that is coupled with an O-methyl transferase to form a methoxyethyl group. An alternative function of methyltransferase enzymes is observed in the 23s rRNA N-methyl transferases of *Nocardiopsis* sp. and *Streptomyces azureus*, which are not involved in thiopeptide maturation but confer host strain resistance to the thiopeptide antibiotics produced, TP-1161 and thiostrepton.[62,69] There is no instance of SAM-dependent methyltransferases conferring host resistance, therefore the products of ORFs 1 and 9 are likely to be ancillary enzymes that facilitate methyltransfer to the core peptide.

**Table 2.2.** Methyl transferase enzymes in known thiopeptide biosynthetic gene clusters.

| Thiopeptide | Gene | Proposed function | Function of closest homologue | GenBank accession no. |
|---|---|---|---|---|
| Lan1 – 3 | ORF1 | | type 11 AdoMet_MTases | ZP_06972504.1 |
| Lan1 – 3 | ORF9 | | type 11 AdoMet_Mtases | ZP_06972504.1 |
| Thiocillin[76] | BC5080 | O-methylation (Thr) | type 11 methyltransferase | YP_001277818.1 |
| Thiomuracin[67] | TpdI | C-methylation (thiazole) | coproporphyrinogen Dhase | YP_003651165.1 |
| GE2270[67] | TpdU | C-methylation (thiazole) | coproporphyrinogen III oxidase | ACS_83777.1 |
| GE2270[67] | TpdL | C-methylation (thiazole) | coproporphyrinogen DHase | 3651165.1 |
| GE2270[67] | TpdM | O-methylation (thiazole) | O-Mtase | YP_005463210.1 |
| GE2270[67] | TpdT | N-methylation (Asp) | AdoMet_MTases | YP_004303843.1 |
| TP-1161[69] | TpaI | Host resistance | rRNA (adenosine-2'-O-)-Mtase | ZP_09960304.1 |
| Thiostrepton[62] | TsrP | Host resistance | MTase, TIGR00027 family | ZP_04607699.1 |

The genes Lan1 – 3 encode 69-, 66- and 65-aa precursor peptides with 21-, 18- and 18-aa core peptides, respectively. A BLASTp search of the three precursor peptides highlights cyclothiazomycin (**17**) as the closest known thiopeptide,[70] but a BLASTp search of the core peptide alone highlights the thiomuracins (**16**) as the closest homologs.[67] Leader and core peptide sequences were aligned with Lan1 – 3 using Clustal Omega,[77] depicted

in Figures 2.3 and 2.4, respectively. Sequence alignment is colour coded according to amino acids with similar structure and physical properties. The three leader peptides contain a high proportion of acidic residues, which are proposed to be important for modification enzyme recognition, and C-terminal GA cleavage sites characteristic of series *d* thiopeptides. Current understanding of the mechanism of pyridine core formation in series *d* and *e* thiopeptides involves elimination of the leader peptide during dehydration of a dehydropiperidine intermediate, a mechansim which, unlike other RiPPs, does not rely on proteolytic cleavage by a specific peptidase.[73]

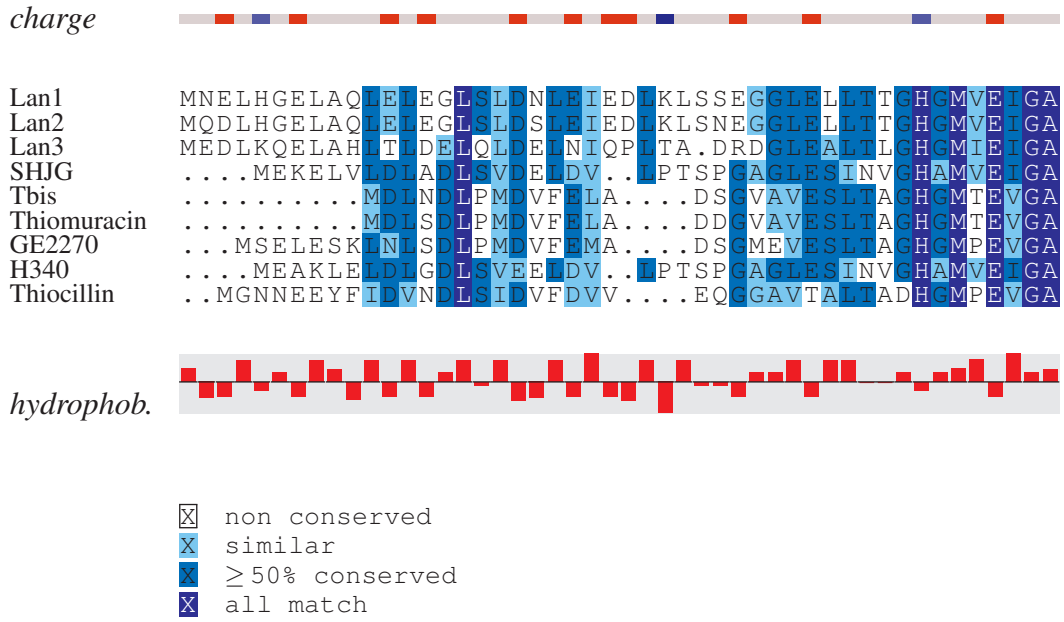



**Figure 2.3.** Sequence alignment of the leader peptides of thiopeptides from strain T81 and known thiopeptide BLASTp hits.

Schemes 2.2, 2.3 and 2.4 present the proposed maturation of the precursor peptides Lan1 – 3 into a series of putative mature peptides, which have several novel features when compared to known thiopeptides. Firstly, all isolated thiopeptides to date contain 26-, 29- or 35-membered core macrocycles.[78,79] Two pyridine core precursor residues, Ser1 and Ser9, are conserved between the precursor peptides Lan1 – 3. The pyridine core of each thiopeptide is most likely to be constructed between these residues, creating a novel 23-membered macrocyclic scaffold. Biological mode of action is strongly correlated to the core macrocycle ring size,[71,79] discovery of a novel scaffold may lead to novel modes of action. Secondly, until the discovery of cyclothiazomycin, a 2,3,6-thiazole substitution pattern was regarded as a common feature of series *d* thiopeptides and was thought to be required for construction of the pyridine core.[78] Cyclothiazomycin instead possesses an Ala residue in the 2-position and an Asp residue in the 3-position, resulting in the possibility of two isomeric configurations at the Ala-derived stereocentre.[70] Azole-substituted at the 2- and 6-positions, the putative Lan1 and Lan2 thiopeptides feature a Ser-derived substituent in the 3-position, which may be present as: (i) Ser, (ii) Dha, (iii) cyclised to an oxazole, or (iv) dehydrated to an oxazoline moiety.

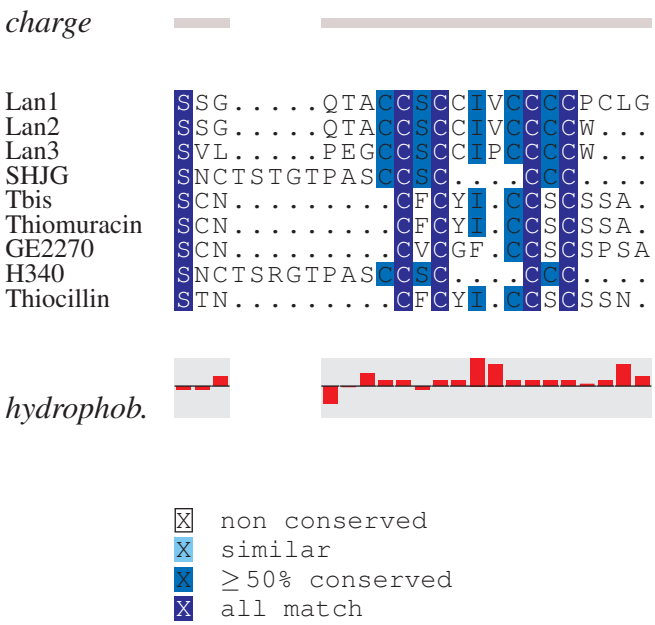



**Figure 2.4.** Sequence alignment of the core peptides of thiopeptides from strain T81 and known thiopeptide BLASTp hits.

Uncertainty in the PTM process extends to other residues in the thiopeptides since knowledge of the maturation enzymes cannot predict the extent of core peptide dehydrogenation. Schemes 2.2 – 2.4 propose a range of Ser, Thr and thiazoline dehydration. Examples of specific dehydrogenation are present amongst known thiopeptides, for example, the dehydrogenation protein CltC in cyclothiazomycin biosynthesis is able to recognise the bithiazoline functional group and selectively dehydrogenate the N-terminal residue of Cys12 – Cys13, and is able to perform two consecutive dehydrogenations in the trithiazoline cluster Cys15 – Cys16 – Cys17.[70] There are many examples where unmodified Thr and Ser residues remain in the mature peptide. Two such examples are seen in thiostrepton A (**15**) and cyclothiazomycin (**17**), both of which contain Dha/Dhb residues in addition to unmodified Thr residues. In the putative Dha/Dhb-containing thiopeptides of Schemes 2.2 and 2.3, cyclisation to oxazoline moieties is plausible but would not result in a change in molecular mass.

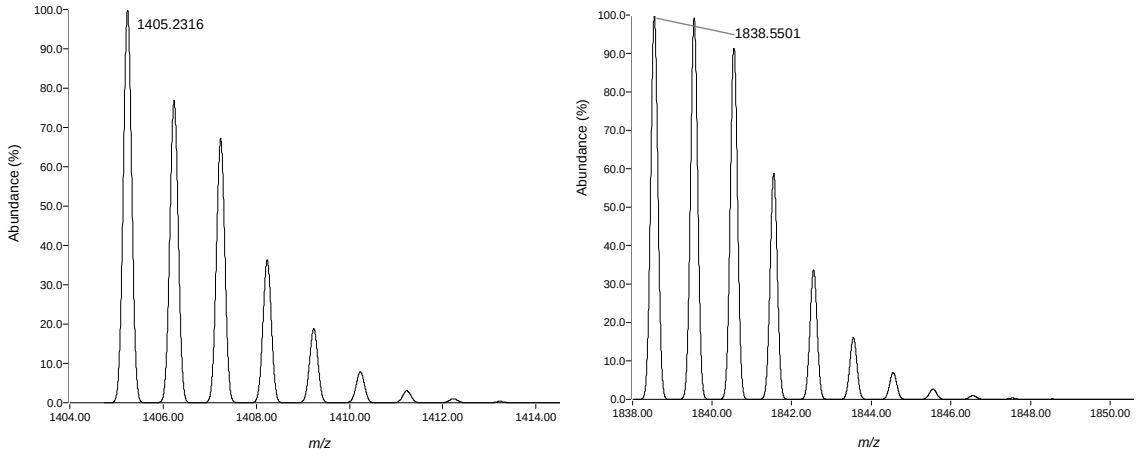
Post-translational modifications catalysed by the additional maturation enzymes ORFs 1, 8 and 9 were also considered in Schemes 2.2 – 2.4. One- and two-fold methylation by the products of ORFs 1 and 9 is shown in several of the proposed mature thiopeptides, although the schemes do not account for the full range of mature peptides. Considering the role of the putative dipeptidase product of ORF8, several analogs with C- and N-terminal residue cleavage are proposed. PTMs involving installation of a series *a* – *c* or series *e*-type tryptophan-derived moiety were not considered, nor was the presence of a tertiary thioether linked macrocycle similar to that of cyclothiazomycin (**17**), as there is no evidence of homolgous modification enzymes in the thiopeptide gene cluster of T81.

## 2.3 Thiopeptide Screening

Cultures of *Thermogemmatispora* strain T81 were screened for the presence of thiopeptides. Plated co-cultures were screened by whole-cell MALDI-TOF MS, liquid cultures were fractionated by RP HP20 chromatography and screened by MALDI-TOF MS, both procedures as described in Chapter 5. Examples of whole-cell MALDI-TOF MS spectra are provided in Appendix D.3. The overwhelming collection of hundreds of plausible mature thiopeptides required a generally applicable strategy to identify thiopeptides produced by strain T81 (not all candidates are depicted in Schemes 2.2 – 2.4). Selection criteria were derived by considering the proposed thiopeptides from strain T81, as well as analysing 88 known thiopeptides (Appendix F).

**Selection Criteria for MALDI-TOF MS Sreening**

1. Ion clusters in the range *m/z* 1200 – 2000
2. Ions for which cationised or oxidised ion clusters were identified
3. Ion clusters with an isotopic distribution pattern that fits predicted simulated spectra
4. Ions whose exact mass holds a first decimal place value $\leq 6$




**(a)** Lowest $M_W$ thiopeptide: $C_{58}H_{56}N_{18}O_9S_8$. **(b)** Highest $M_W$ thiopeptide: $C_{76}H_{103}N_{21}O_{15}S_9$.

**Figure 2.5.** Examples of simulated mass spectra for predicted thiopeptides produced by strain T81.

Criterion 1 was determined by the lowest and highest nominal masses of the strain T81 thiopeptides, with a tolerance of $\pm$ 200 Da to account for unpredicted modifications. The lowest proposed mass (1405 Da, Scheme 2.3, Figure 2.5a) is that of a Lan2 thiopeptide with all azole moieties present as thiazoles and hydrolysis of the C-terminal Trp residue. The highest proposed mass (1838 Da, Figure 2.5b) is that of a Lan1 thiopeptide with all azole moieties present as thiazolines, two methylations, and unmodified Ser2 and Thr5 residues. Criterion 2 was selected due to the tendency of peptides to sequester $Na^+$ and

**Table 2.3.** Statistical analysis of known thiopeptides.

| | | | Mean | | | |
|---|---|---|---|---|---|---|
| Series | Members | Mass range | Exact mass | 1st dp | IHD : Mass | C : H |
| a/b | 16 | 1797 – 1509 | 1625 | .5 | 2.38 | 0.85 |
| c | 1 | | 1940 | .6 | 2.16 | 0.81 |
| d | 61 | 760 – 1370 | 1195 | .2 | 2.92 | 1.04 |
| e | 10 | 1281 – 1452 | 1370 | .2 | 2.77 | 1.05 |

$K^+$ cations, as well as $H^+$ ions. Criteria 3 and 4 were determined by simulating the mass spectra of Lan1 – 3 thiopeptides (e.g. Figure 2.5). These criteria were supported by the analysis of known thiopeptides (summarised in Table 2.3), which shows the correlation between series and the ratio of index of hydrogen deficiency (IHD) and exact mass. Thiopeptides of series *d* and *e* have, on average, a higher IHD : mass ratio and a lower first decimal place value due to the central pyridine core. In contrast, series *a* – *c* thiopeptides contain a central piperidine heterocycle, a lower average IHD : mass ratio and a higher first decimal place value. Across all series, the high proportion of N, O and S relative to H has an effect on the exact mass of the thiopeptides: only one of the 88 known thiopeptides has an exact mass whose first decimal place value is $\geq 5$ (Sch 40832, 1940.**5973** Da). All mature thiopeptides presented in Schemes 2.2 – 2.4 have an exact mass whose first decimal place value is $\leq 5$ and are presumed to be of series *d*.

The criteria were used to generate 'hits' from MALDI-TOF MS spectra to continue for further analysis by HR ESI-MS. Most ions in the range *m/z* 1200 – 2000 were discounted on the basis of criterion 4. Several ions in the mass range were identified by MALDI-TOF MS as potential candidates but were later excluded by HR ESI-MS. For example, the whole-cell MALDI-TOF MS spectrum of a strain T81 co-culture with TKA 04.11 (Figure 2.6) shows ions at *m/z* 1468.6, 1506.5, 1664.6, 1778.8, and 1929.9. Investigation of these ions by HR ESI-MS revealed that the exact masses did not match any plausible thiopeptide chemical formula to $\leq \Delta$ 200 ppm. Moreover, these ion clusters did not show sodiated or potassiated versions and the total ion count was often very low.

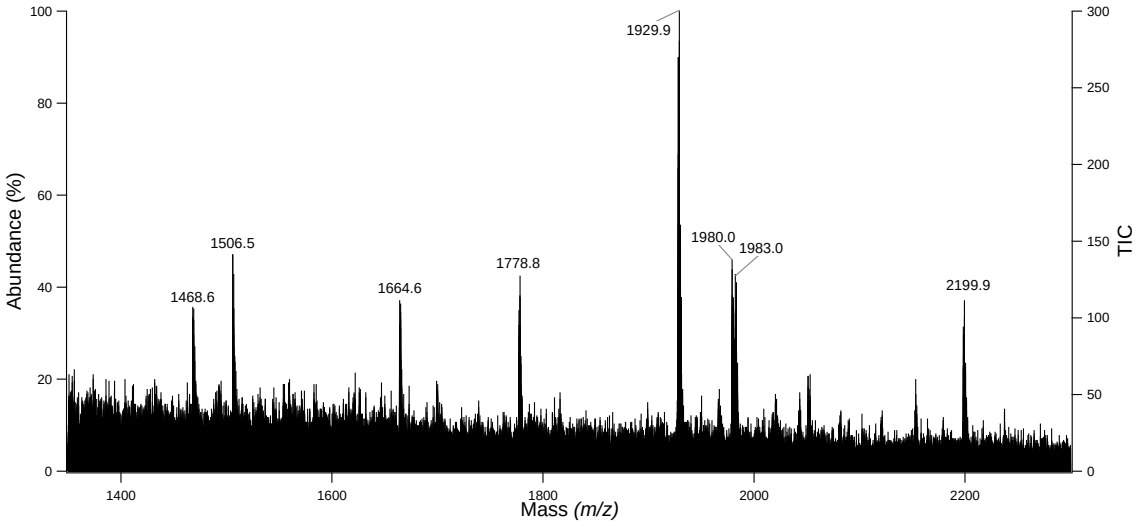
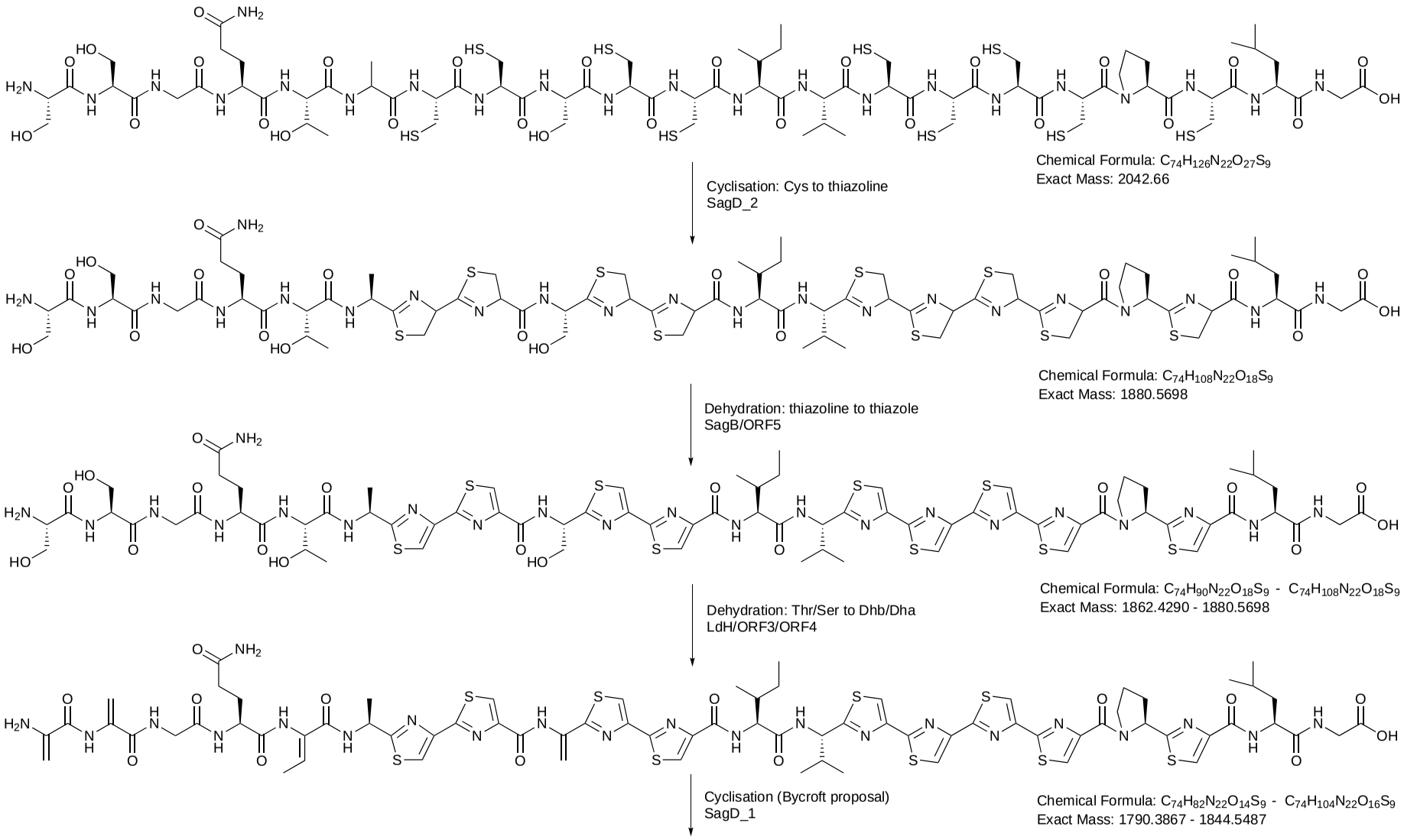
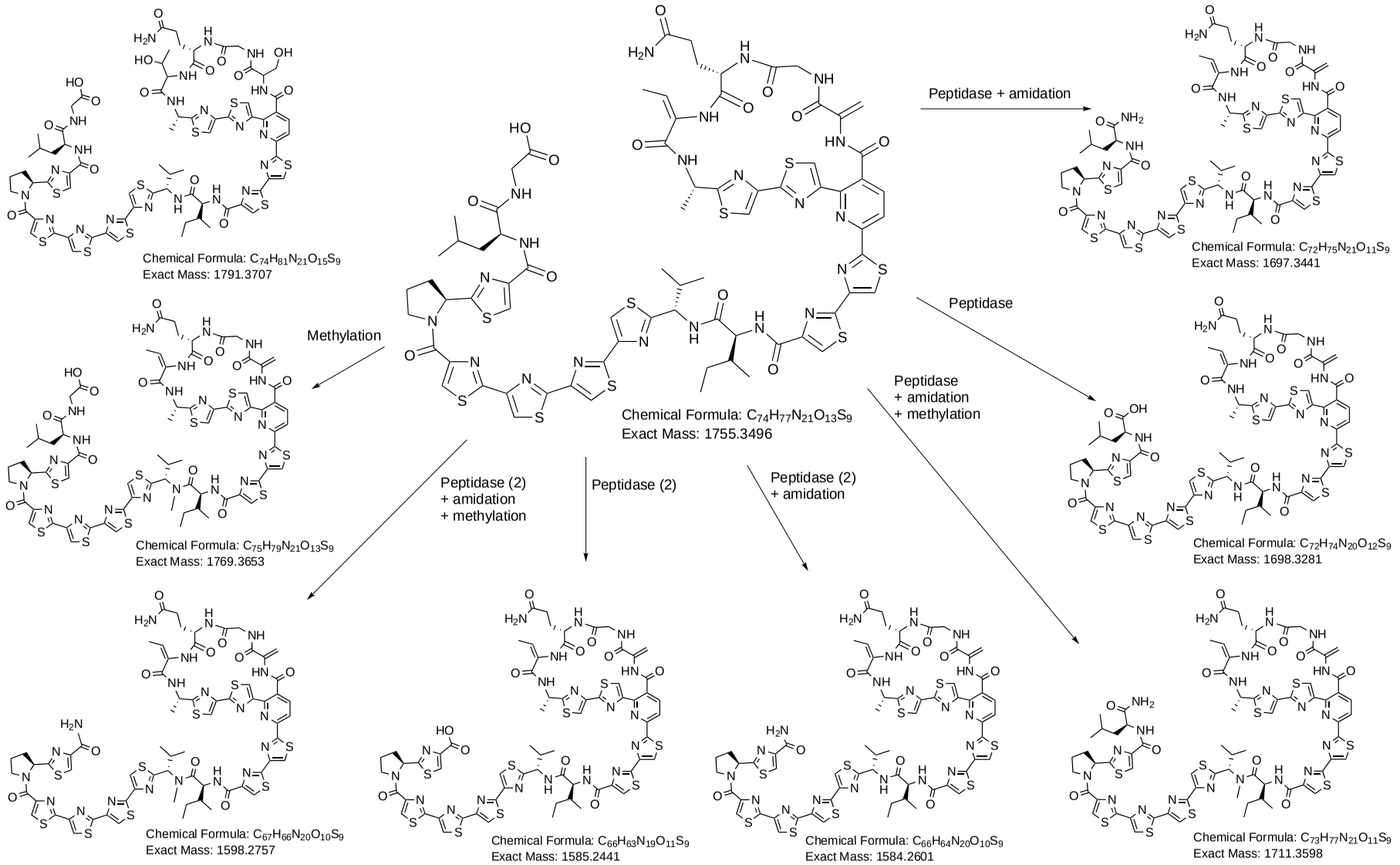



**Figure 2.6.** Whole-cell MALDI-TOF MS spectrum of a strain T81/TKA 04.11 co-culture.

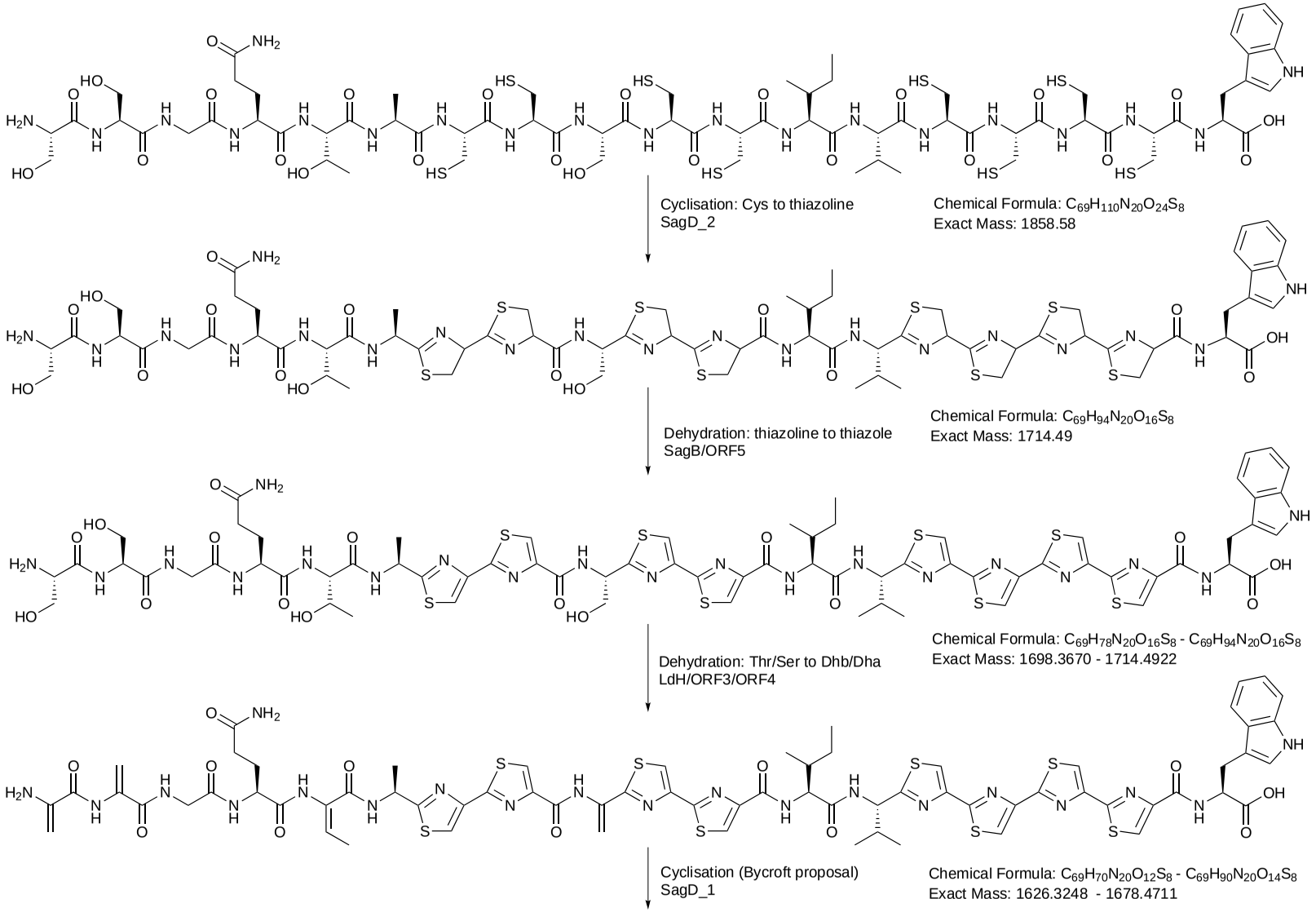
The strain T81 cultures failed to reveal a peptide corresponding in mass to the bioinformatically predicted thiopepetides or to any analogs differing in the number of dehydrations, ancillary PTMs, site of leader peptide cleavage, or that had retained the leader peptide. It appears that the thiopeptide gene cluster is not expressed in its natural host under the growth conditions explored. Considering the sheer complexity of the structures and number of mature peptides that could be formed, MS may not be the most suitable screening approach for these compounds. Microbial extracts could be screened for the presence of thiopeptides by IR and UV spectroscopic methods, searching for the characteristic thiazole/oxazole CH stretching frequency and comparing UV spectra to known thiopeptides.



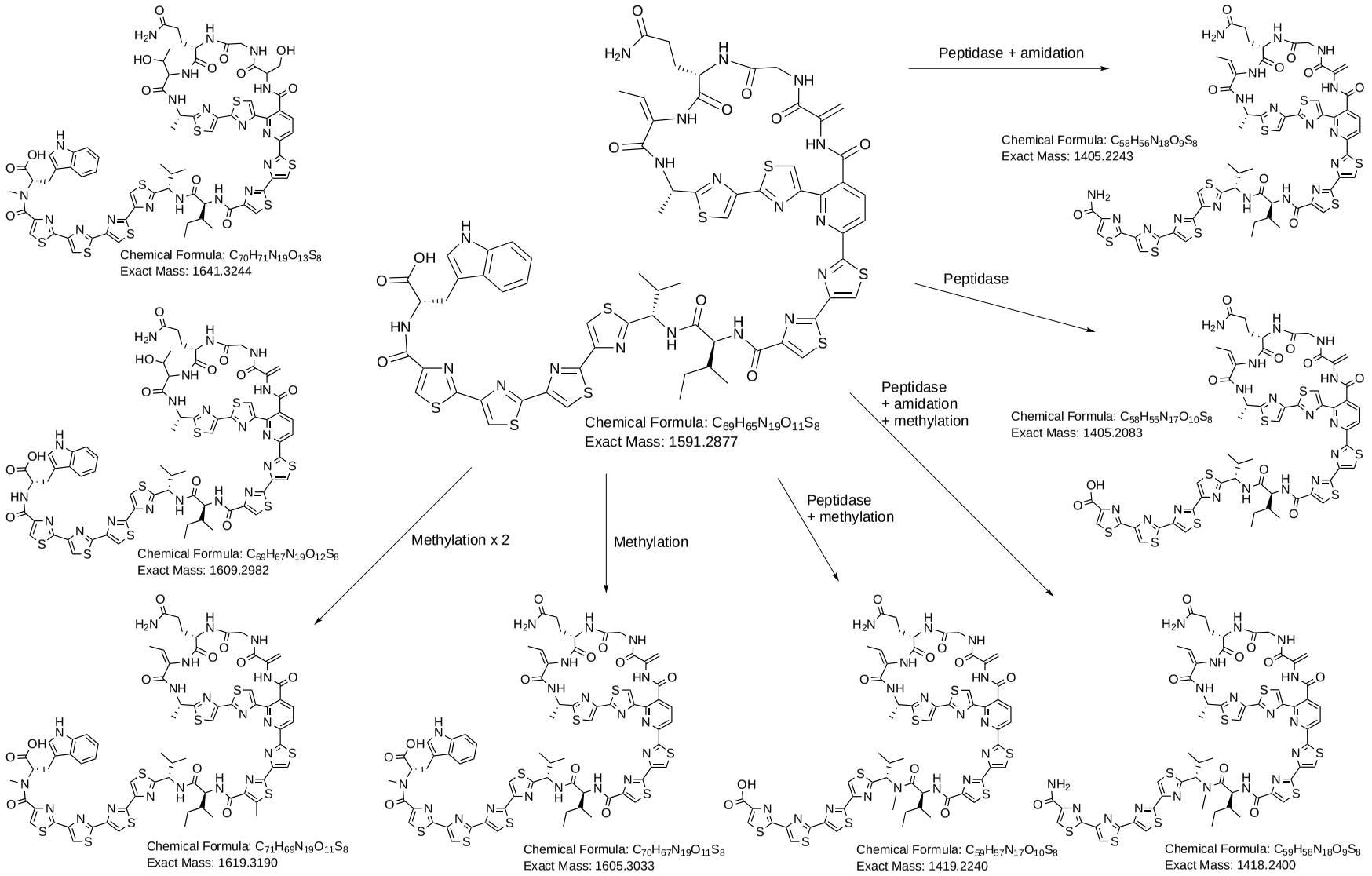

**Scheme 2.2.** Proposed biosynthesis of Lan1 thiopeptides.



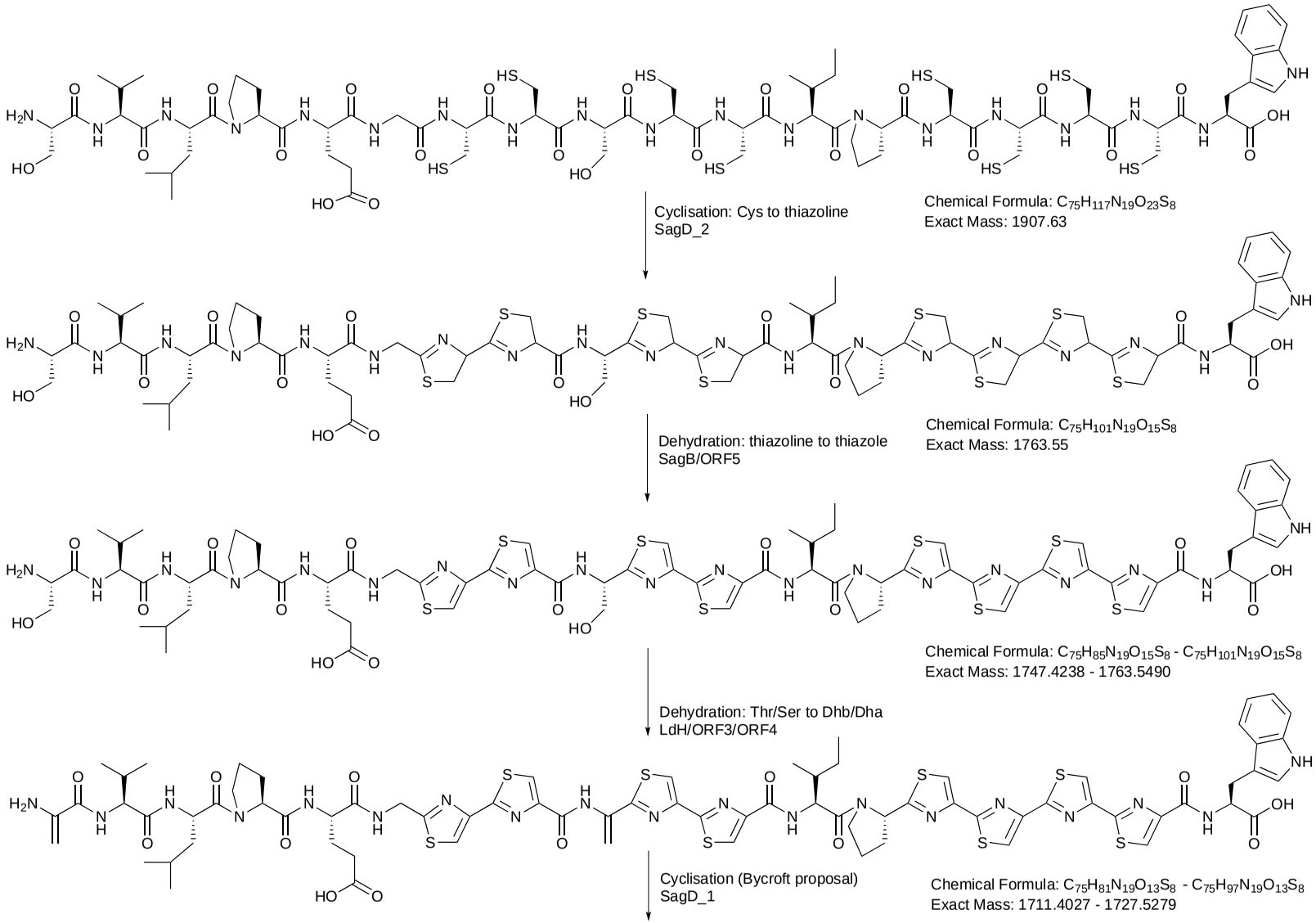

**Scheme 2.2.** Proposed biosynthesis of Lan1 thiopeptides. (cont.)



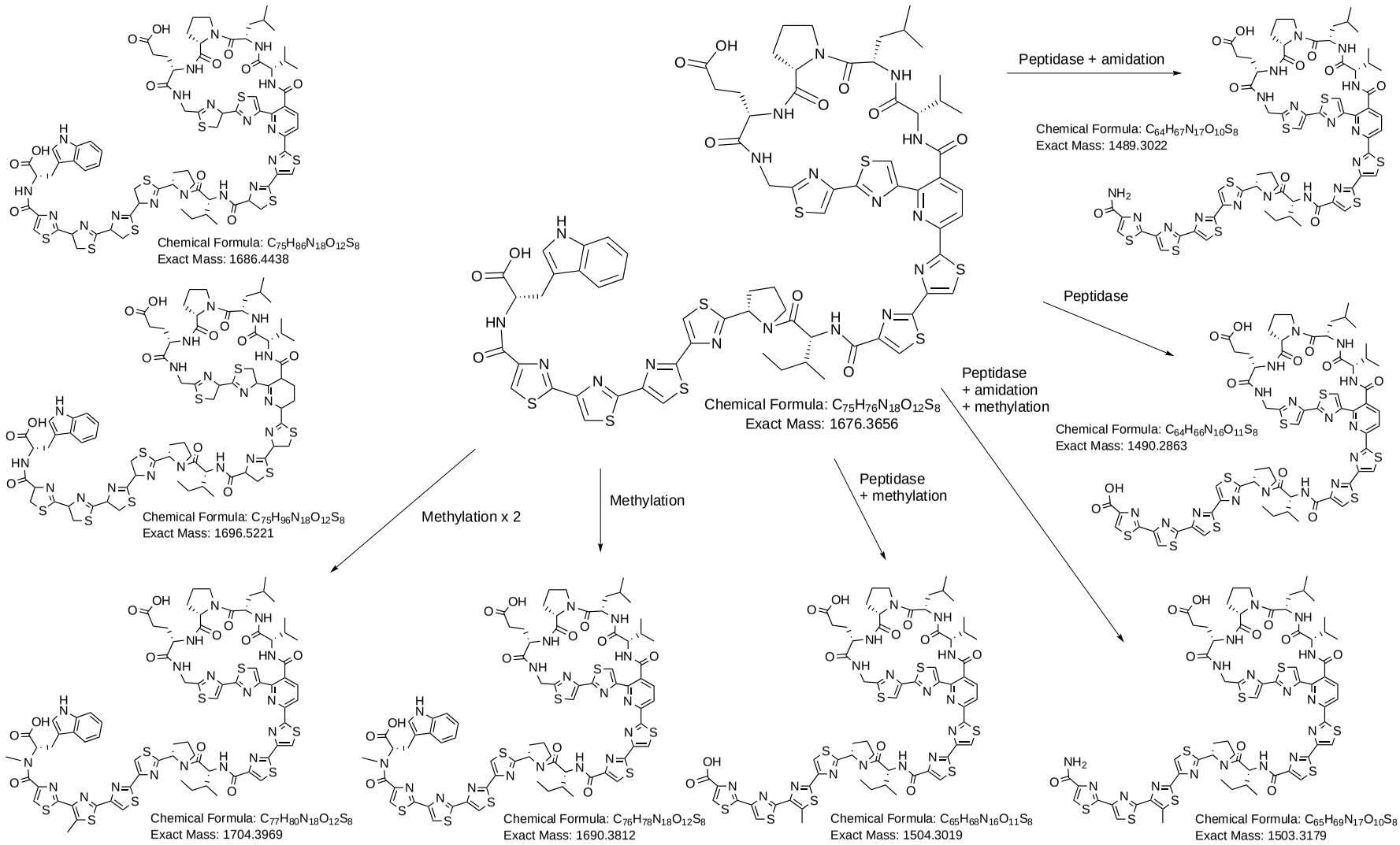

**Scheme 2.3.** Proposed biosynthesis of Lan2 thiopeptides.




**Scheme 2.3.** Proposed biosynthesis of Lan2 thiopeptides. (cont.)




**Scheme 2.4.** Proposed biosynthesis of Lan3 thiopeptides.



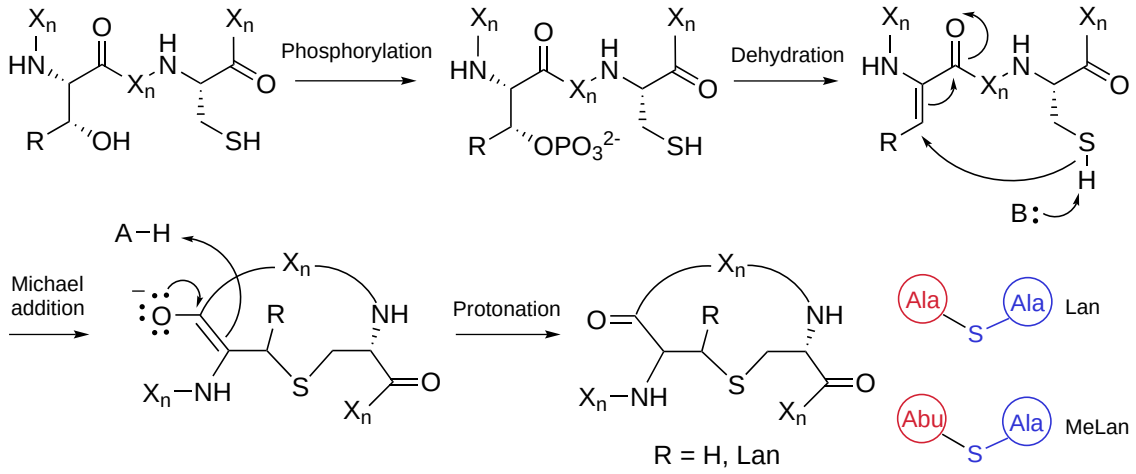

**Scheme 2.4.** Proposed biosynthesis of Lan3 thiopeptides. (cont.)

# Chapter 3

# Tikitericin, A Novel Lanthipeptide Produced by *Thermogemmatispora* Strain T81

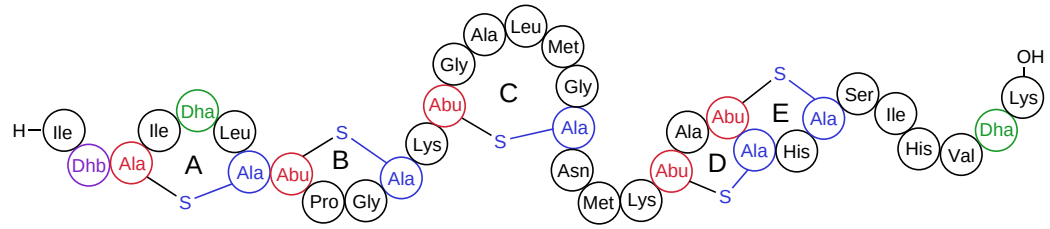
## 3.1 Introduction

Lanthipeptides are produced by Gram-positive bacteria and are clearly distinguished among the bacterial peptide toxins by intramolecular rings formed by the thioether linked amino acids lanthionine (Lan) and 3-methyllanthionine (MeLan).[60,67] Over 90 molecules sharing the (Me)Lan chemical motif have been reported, many of which have antimicrobial activity and are referred to as lantibiotics.[60,80]



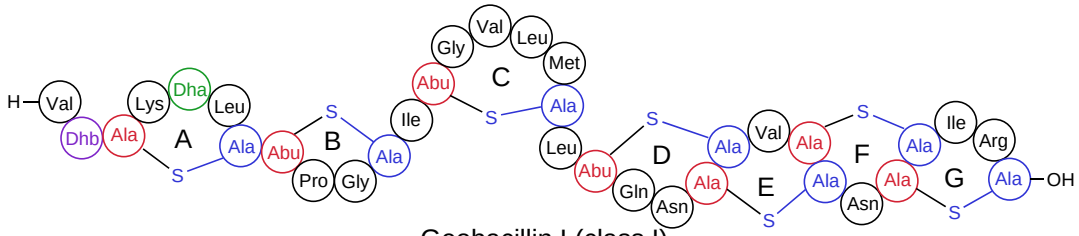

**Scheme 3.1.** Biosynthetic pathway to lanthipeptide (Me)Lan residues, adapted from Knerr and van der Donk.[81]

The characteristic (Me)Lan bridges are introduced in a two-step post-translational modification process. Firstly, dehydration of Ser and Thr residues in the precursor peptide occurs to give the rare unsaturated amino acids dehydroalanine (Dha) and



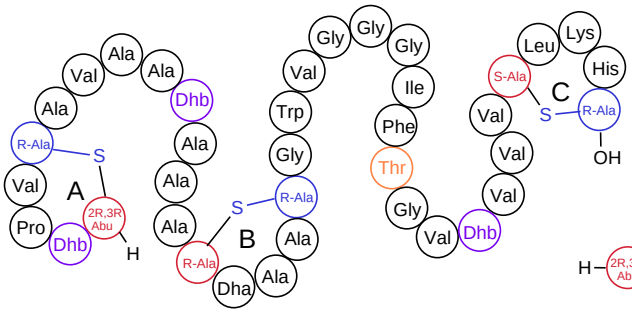

Nisin A (class I)

**18**

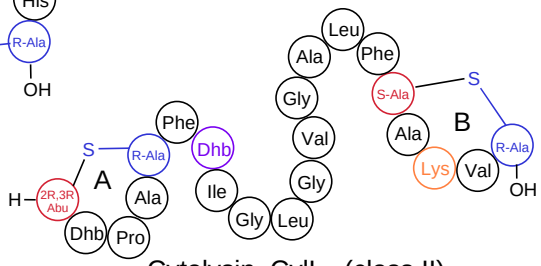



Geobacillin I (class I)

**19**

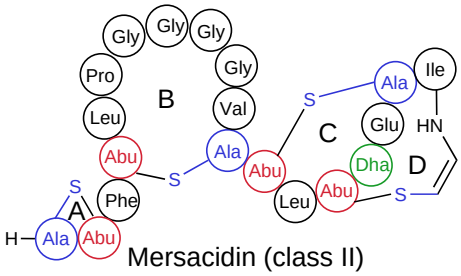



Cytolysin, CylL$_L$ (class II)

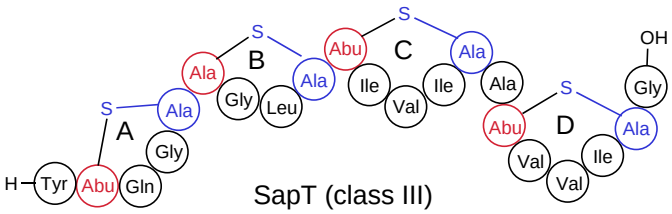



Cytolysin, CylL$_S$ (class II)

**20** **21**

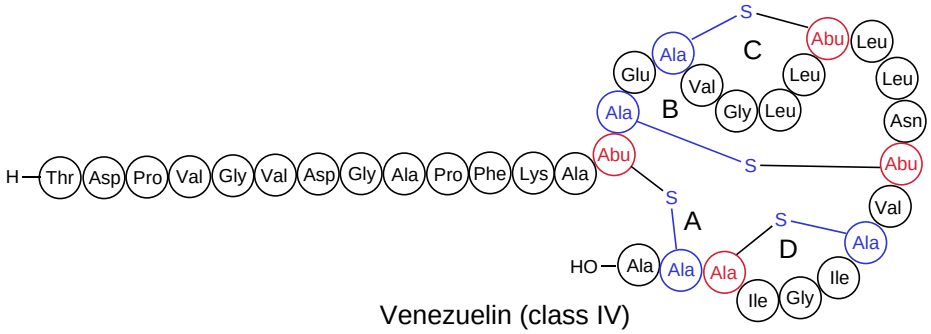



Mersacidin (class II)




SapT (class III)

**22** **23**




Venezuelin (class IV)

**24**

**Figure 3.1.** Representative examples of lanthipeptides. For (Me)Lan residues, the segments derived from Ser/Thr are in red and those derived from Cys are in blue. Additional dehydro residues are in green (Dha) and purple (Dhb). Sites of further PTM are in orange.

dehydrobutyrine (Dhb), respectively, a process that usually occurs via a transient O-phosphorylated intermediate.[82,83] The dehydro amino acids are subsequently involved in a stereospecific intra-molecular Michael-type reaction with thiol-containing cysteine residues to form conformationally restricting thioether linkages (Scheme 3.1). As shown in Figure 3.1, there is remarkable diversity in the ring topology created by lanthipeptide (Me)Lan residues.

Until recently, lanthipeptides have been classified on the basis of structural differences in the modified peptide, assigned to either type A or type B. Mersacidin (**22**), actagardine, and lantibiotics of the innamycin type constitute the latter group, which are rigid globular peptides with no net positive or negative charge, also distinguished by an outlying long leader peptide. Type A lantibiotics such as nisin (**18**) and subtilin are flexible, elongated peptides. With the increasing discovery of novel, structurally diverse lanthipeptides, including those with no antimicrobial activity (e.g. venezuelin (**24**)),[80,83] classification has shifted to a 'class' nomenclature defined according to the producing strain's biosynthetic machinery (Figure 3.2).[59,60] Lanthipeptide precursors are encoded by short genes, generically termed LanAs. For class I lantibiotics, dehydration and cyclisation of Ser and Thr residues in the (Me)Lan precursor is carried out by a dedicated dehydratase (LanB) and a LanC cyclase.[84,85] For class II,[86] III,[87] and IV[83] lanthipeptides, dehydration and cyclisation are carried out by a bifunctional lanthionine synthetase. The class II LanM lanthionine synthetases contain a N-terminal dehydration domain, which bears no homology to other enzymes in the databases, and a C-terminal LanC-like cyclase domain. Class III lanthipeptide precursors are modified by a trifunctional synthetase bearing N-terminal lyase and central kinase domains, and a putative C-terminal cyclase domain which lacks many of the conserved active-site residues found in LanC/LanM. Finally, the recently introduced class IV lanthipeptides are modified by the synthetase LanL, which contains N-terminal lyase and kinase domains homologous to class III synthetases, but its C-terminal cyclase domain is analogous to LanC. These distinct pathways demonstrate the convergent evolution of efficient ribosome-based biosynthetic strategies towards a common functional group that imparts significant chemical stability to the mature peptide.[60]

The archetypical lantibiotic nisin (**18**) has been used as a food preservative in the dairy industry since 1969 with no substantial microbial resistance,[88] promoting the application of lantibiotics as chemotherapeutics for human pathogens. A combination of multiple modes of action is believed to be a factor in nisin escaping resistance mechanisms.[80,89] Although the mechanism of antimicrobial activity has not been extensively studied across all classes, most lantibiotics are believed to act on pathogenic bacteria by pore formation and/or inhibition of peptidoglycan biosynthesis.[89] Specific binding interactions for the cell wall precursor, lipid II, have been identified for nisin, which bonds to the pyrophosphate moiety of lipid II through the A and B rings conserved in several other class I lantibiotics, including microbisporicin and mutacin 1140.[60] Once bound to

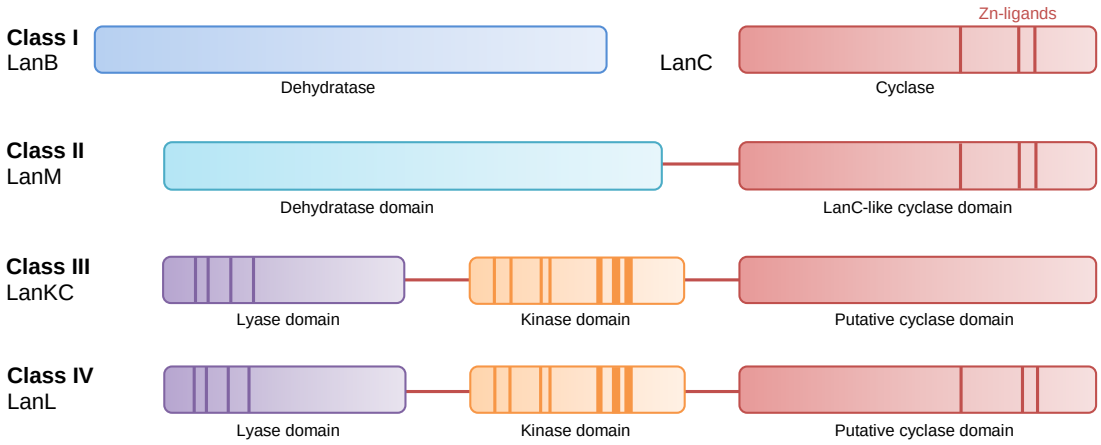



**Figure 3.2.** Classification of lanthipeptides, adapted from Arnison *et al.*[59]

lipid II, nisin is able to insert into the membrane and form stable pores.[90] Lantibiotics generally possess potent activity against many clinically relevant Gram-positive bacteria, including drug-resistant strains of *Staphylococcus*, *Streptococcus*, *Enterococcus*, and *Clostridium*, as well as against select Gram-negative pathogens, such as *Neisseria* and *Helicobacter*.[91] Several lantibiotics are currently in clinical development, including duramycin for the symptomatic treatment of cystic fibrosis and a derivative of actagardine for the treatment of *Clostridium difficile* infections.[92–94] Mutacin 1140 is currently in preclinical development for the treatment of Gram-positive bacterial infections.[95] Other applications of lantibiotics in agriculture, veterinary medicine and molecular imaging are also emerging.[96,97]

## 3.2 Identification of a Novel Lanthipeptide Locus in *Thermogemmatispora* strain T81

The genome of *Thermogemmatispora* strain T81 was sequenced by Stott and co-workers (GNS Science) after DNA extraction. Annotation of the draft genome identified a gene cluster encoding for putative lanthionine synthetase and associated lanthipeptide biosynthesis proteins (Figure 3.3, GNS Science). Deduced peptides and proteins derived from T81 are shown in Table 3.1.

A BLASTp search of the lanthionine synthetase gene (TikM) showed 99% homology to a lanthionine synthetase C family protein from *Knedonobacter racemifer* (phylum *Chloroflexi*), representative biosynthetic machinery of class II lanthipeptides (Figure 3.4, Table 3.1).

**Table 3.1.** Deduced peptides and proteins from T81 (GNS Science).

| ORF | Length (aa) | Proposed function | Best homologue [identities (% peptide sequence similarity, no. of amino acids), GenBank accession no.] | Trans-membrane helices | Signal peptide |
|---|---|---|---|---|---|
| ORF1 | 329 | Dehydratase | Serine/threonine dehydratase family protein *Rhodococcuserythropolis* (97%, 329 aa), ZP 04383500 | No | No |
| ORF2 | 172 | Acetyltransferase | Acetyltransferase, gnat family *Clostridium botulinum* (100%, 170 aa) YP 001886336 | No | No |
| ORF3 | 364 | | Peptide release factor 2 (RF-2) *Lactobacillus pentosus* (96%, 357 aa), CCB82692 | No | No |
| ORF4 | 361 | | RNA polymerase, sigma 70 subunit, RpoD subunit *Lactobacillus acidophilus* (88%, 365 aa) YP 194063 | No | No |
| ORF5 | 199 | | Anti-RNA polymerase sigma 70 factor *Shewanellaamazonesis* (32%, 154 aa), YP 926267 | No | Yes |
| ORF6 | 71 | | Conserved hypothetical protein *Ktedonobacterracemifer* (98%, 71 aa), ZP 06968110 | No | No |
| ORF7 | 340 | | Signal recognition particle-docking protein FtsY *Ktedonobacterracemifer* (99%, 337 aa), ZP 06968111 | No | No |
| ORF8 | 327 | | Acetyl-CoAcarboxylase, carboxyltransferase, $\beta$ subunit (86%, 295 aa) ZP 06968112 | No | No |
| ORF9 | 319 | | Acetyl-CoA carboxylase, carboxyl transferase, $\alpha$ subunit *Ktedonobacterracemifer* (98%, 328 aa) ZP 06968113 | No | No |
| ORF10 | 84 bp | Anti codon | tRNALeu CAA | - | - |
| ORF11 | 201 | Unknown | Hypothetical protein RB9488 *Rhodopirellulabaltica* (34%, 543 aa) NP 868909 | Yes | No |
| TikM | 1114 | Lanthionine synthetase | Lanthioninesynthetase C family protein, *Ktedonobacterracemifer* ( 99%, 1099 aa), ZP 06966366 | Yes | No |
| TikA | 82 | Lanthipeptide | Lantibiotic mersacidin *Bacillus psuedomycoides* (57%, 75 aa), ZP 04153726 | No | No |
| ORF12 | 853 | Exporter protein | Cyclic nucleotide-regulated ABC bacteriocin/lantibiotic exporter *Nodulariaspumigena* (%, 1033 aa), ZP 01630014 | Yes | No |
| ORF13 | 201 | Transferase | Alpha 1,6 mannosyltransferase *Ajellomycescapsulatus* (93%, 404 aa), EGC47259 | Yes | Yes |
| ORF14 | 80 | Unknown | Conserved hypothetical protein *Ktedonobacterracemifer* (100%, 80 aa) ZP 06969678 | No | No |
| ORF15 | 412 | | Coenzyme PQQ biosynthesis protein E *Geodermatophilus obscures* (%, 370 aa) YP 003409498 | No | No |
| ORF16 | 101 | | Coenzyme PQQ biosynthesis protein D *Marinobacteralgicola* (96%, 91 aa) ZP 01893577 | No | No |
| ORF17 | 258 | | Coenzyme PQQ biosynthesis protein C *Streptomyces bingchenggensis* (95%, 244 aa) ADI05215 | No | No |
| ORF18 | 310 | | Coenzyme PQQ biosynthesis protein B *Pseudonocardia* sp. P1 (100%, 297 aa) ZP 08123741 | No | No |
| ORF19 | 37 | Unknown | No significant similarity found | No | No |
| ORF20 | 64 | Unknown | No significant similarity found | No | No |
| ORF21 | 667 | | Acetoacetyl-CoA synthase *Ktedonbacterracemifer* (99%, 667 aa) ZP 06967063 | No | No |
| ORF22 | 364 | | DNA primase *Thermodesulfatatorindicus* (93%, 565 aa), YP 004624884 | No | No |
| ORF23 | 159 | | Hypothetical protein Krac 8420 *Ktedonobacterracemifer* (74%, 150 aa), ZP 06969556 | Yes | No |
| ORF24 | 302 | | ABC-type bacteriocin transporter *Streptococcus suis* (43%, 716 aa), ZP 03624214 | Yes | Yes |
| ORF25 | 40 | Unknown | No significant similarity found | No | No |
| ORF26 | 82 | | DNA binding domain protein, excisionase family *Ktedonobacterracemifer* (%, 67 aa), ZP 06966399 | No | No |
| ORF27 | 302 | | Reiske (2Fe-2S) iron-sulfur domain *Ktedonobacterracemifer* (92%, 292 aa) ZP 06970773 | Yes | No |
| ORF28 | 225 | | Putative transcriptional regulator *Rubrobacterxylanophilus* (93%, 216 aa) YP 645543 | No | No |
| ORF29 | 401 | | FMN-dependent $\alpha$-hydroxy acid dehydrogenase *Haloterrigenaturkmenica* (99%, 431 aa) YP 003404447 | No | No |
| ORF30 | 156 | | CBS domain-containing membrane protein *Chloroflexusaggregans* (98%, 155 aa) YP 002462951 | No | No |
| ORF31 | 350 | | Pyruvate dehydrogenase E1 component $\alpha$ subunit *Ktedonobacterracemifer* (99%, 351 aa) ZP 06969295 | Yes | No |
| ORF32 | 333 | | Dehydrogenase complex E1 component $\beta$ subunit *Saccharopolysporaerythraea* (98%, 331 aa) YP 001107786 | No | No |
| ORF33 | 449 | | Catalytic domain of various dehydrogenase complexes *Ktedonobacterracemifer* (100%, 437 aa), ZP 06969297 | No | No |
| ORF34 | 472 | | Macrolide-efflux protein *Syntrophusaciditrophicus* (92%, 537 aa), YP 462499 | Yes | No |

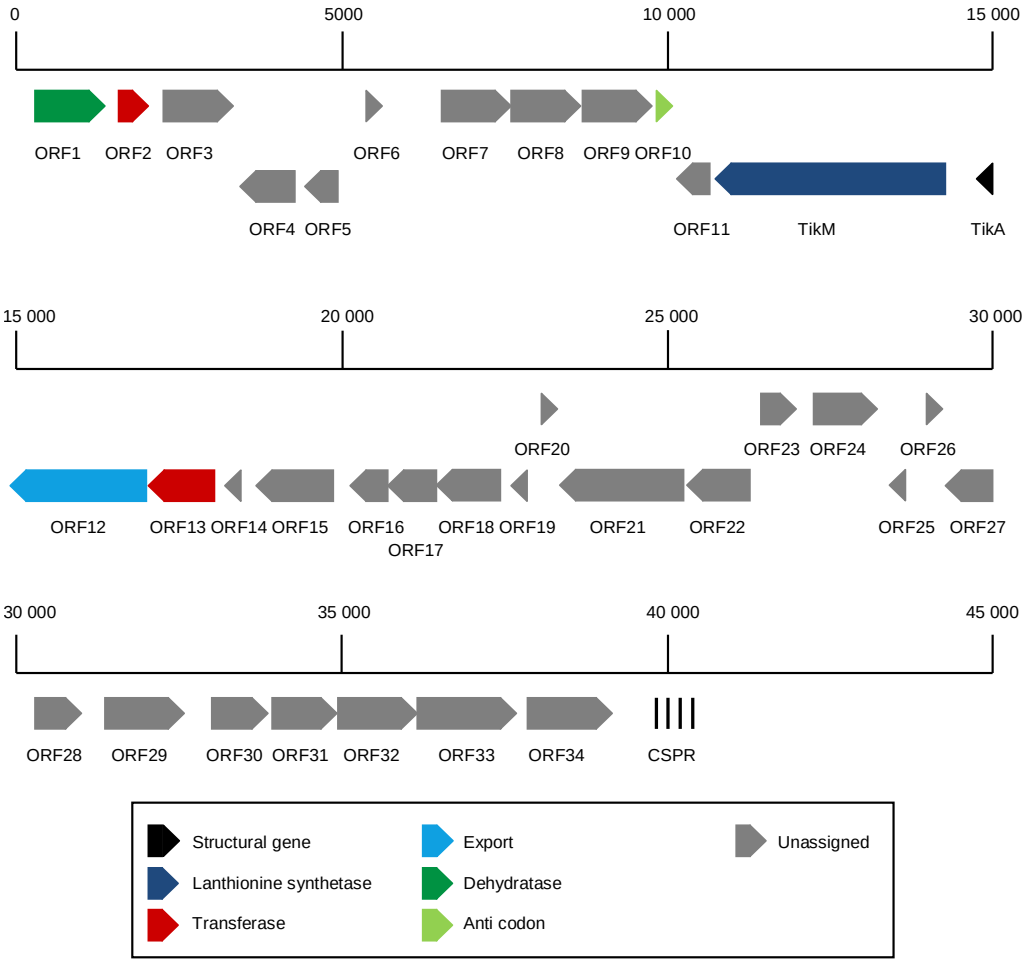



**Figure 3.3.** Tikitericin gene cluster. Each putative gene is shown to scale, with its position in contig 61 marked in kb. Adapted from original figure by Hauch, GNS Science.[58]

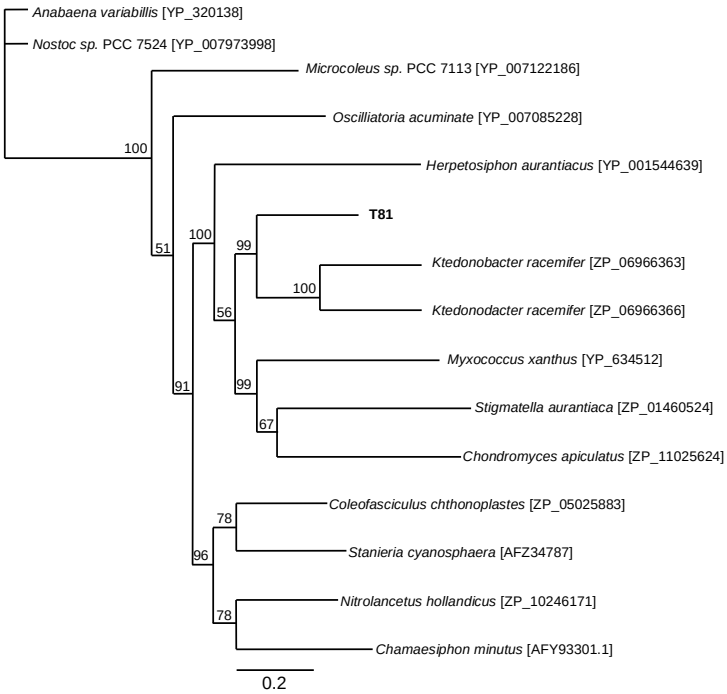



**Figure 3.4.** An un-rooted Neighbour-Joining dendrogram showing the relationship between the T81 lanthionine synthetase (TikM) and other lanthionine synthetase family C proteins. Adapted from original figure by Stott, GNS Science.

The lanthipeptide precursor peptide (TikA) consists of a 47 amino acid-long leader peptide and a 35 amino acid-long core peptide, presented in Figure 3.5. The precursor peptide contains a C-terminal leader peptide cleavage site, observed in other class II lanthipeptide gene clusters,[98] as well as the potential for cyclisation of the core peptide by the bifunctional TikM encoded immediately downstream of the TikA gene sequence (Figure 3.3). Thr, Ser, and Cys residues are shaded to show the potential for post-translational modifications to form (Me)Lan bridges in the mature peptide. Four cysteine and hydroxyl-containing residue pairs suggest the potential for up to four (Me)Lan-linked macrocycles in the mature peptide.


**Figure 3.5.** Tikitericin precursor peptide. The C-terminal leader peptide cleavage site is shaded in red, and Thr, Ser and Cys residues are shaded in blue.

A BLASTp search returned a close relationship of TikA with other 'unknown' gene sequences contained in the genomes of various Cyanobacteria including *Stanieria*, *Anabaena*, *Nostoc*, and *Microcoleus*; *Actinomycetes* including *Streptomyces*; and *Chloroflexi* including *Nitrolancetus* and *Ktedonobacter*.[99] The search highlighted the following known type II lantibiotics: lacticin 481, nukacin, salivaricin, mersacidin (**22**); two known class I lantibiotics: streptococcin and mutacin; and the non-(Me)Lan containing bacteriocins: curvacin A, sakacin A, and carnobacteriocin A, as closely related metabolites.

Leader and core peptide sequences of the known BLASTp hits were aligned with TikA using Clustal Omega,[77] depicted in Figures 3.6 and 3.7, respectively. Sequence alignment is colour coded according to amino acids with similar structure and physical properties. Very little sequence similarity is evident in the core peptide of TikA. Alignment of the leader peptide shows that with 43.5% homology, mersacidin (class II) is the most closely related of the characterised lantibiotics (Figure 3.6). Bar graphs showing residue charge and hydrophobicity are presented for both alignments. The tikitericin leader peptide contains a significant proportion of acidic residues, which are proposed to be important for modification enzyme recognition.[100]

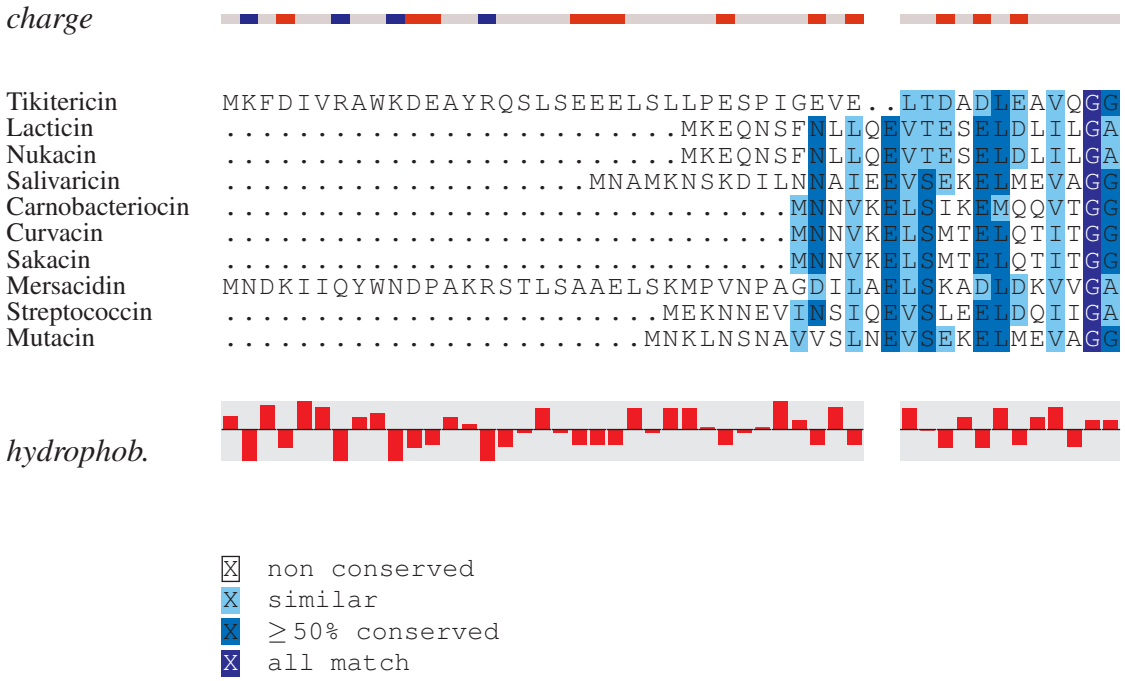



**Figure 3.6.** Sequence alignment of the leader peptides of tikitericin and known lanthipeptide BLASTp hits.

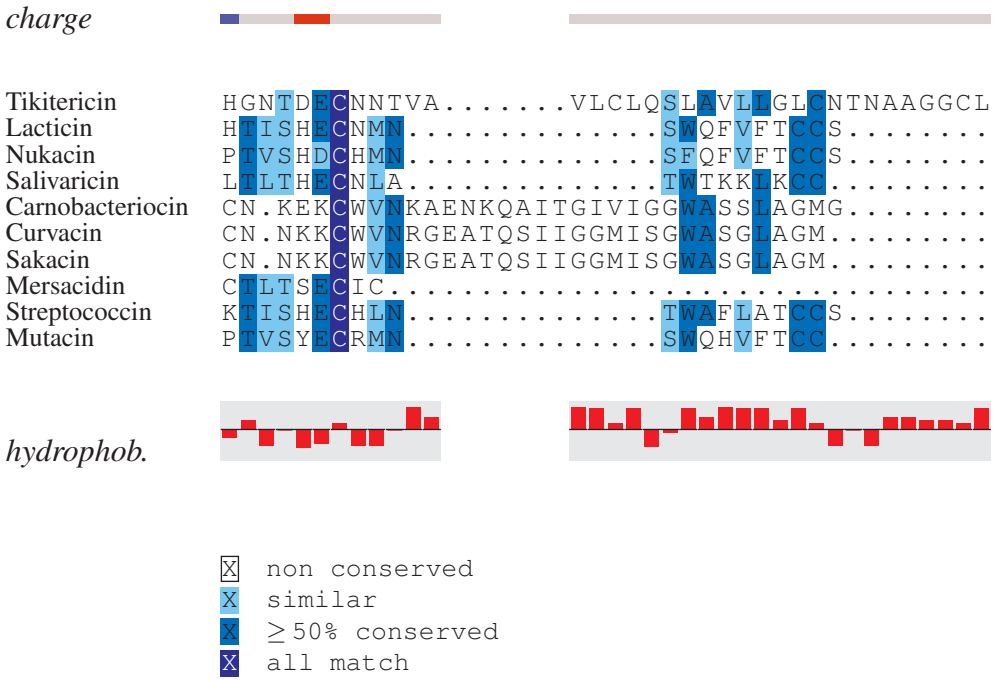



**Figure 3.7.** Sequence alignment of the core peptides of tikitericin and known lanthipeptide BLASTp hits.

## 3.3 Detection and Isolation of the Lanthipeptide, Tikitericin

### 3.3.1 Investigation of lanthipeptide production

#### Liquid media

Initial studies involved the screening of strain T81 cultured as a single strain in liquid media. Bacterial cultures were fractionated by the cyclic loading process developed by Northcote and co-workers, described in Appendix C and tailored for microbial biomass. The cultures were filtered over 3 $\mu$m cellulose filter paper to separate supernatant from bacterial cell mass. The filtrate was immediately passed through a column of reversed-phase HP20 resin, whereas the cell mass was extracted twice in MeOH and the extracts cyclic loaded onto HP20 resin. The 30%, 75% and 100% $Me_2CO$ fractions for both the supernatant and cell mass were screened by $^1$H and COSY NMR. Liquid cultures of three volumes were screened: 2 L shaken culture, 16 L bioreactor culture (Callaghan Innovation), 20 L bioreactor culture (Callaghan Innovation). To distinguish between bacterial metabolites and artefacts of the culture media, volumes of non-inoculated blank AOM1 media were fractionated and screened as a control. Differences were observed between $^1$H NMR spectra of nutrient media and culture fractions, although it was evident that the majority of the isolated biomass was merely recovered culture media. A comparison of the 2 and 16 L cultures highlighted the change in metabolic profile according to the scale and method of cultivation (Figure 3.8), bioreactor cultures showing a greater proportion of aliphatic proton resonances and a reduced number of signals in the aromatic region. The 75% $Me_2CO$ fractions were tested in a bacteriostatic assay against the clinical isolates, *S. aureus* and *E. coli*. No biological activity was observed at the highest concentration tested, 400 $\mu$g/mL. The 30%, 75% and 100% $Me_2CO$ fractions were analysed by MALDI-TOF MS at concentrations of 800, 400, and 200 $\mu$g/mL using 'dried-droplet' and 'bottom-up' methods of sample preparation. Two matrices, $\alpha$-cyano-4-hydroxycinnamic acid (CHCA) and 2,5-dihydroxybenzoic acid (DHB) were employed but it was quickly discovered that the best results in terms of signal intensities were observed using DHB matrix. No ions were detected in tikitericin's expected mass region *m/z* 3000 – 4000.

Strain T81 was co-cultured with strain TKA04.11 in a liquid culture (2 L). Previous growth inhibition studies by Hauch (GNS Science) had shown that the growth of strain TKA04.11 is inhibited by strain T81 (Figure 1.7). This culture, together with single-strain cultures (2 L) of strains T81 and TKA04.11 were fractionated by RP HP20 chromatography and the 30%, 75% and 100% $Me_2CO$ fractions analysed by $^1$H NMR. The 75% $Me_2CO$ fractions were tested in a bacteriostatic assay. No bioactivity was

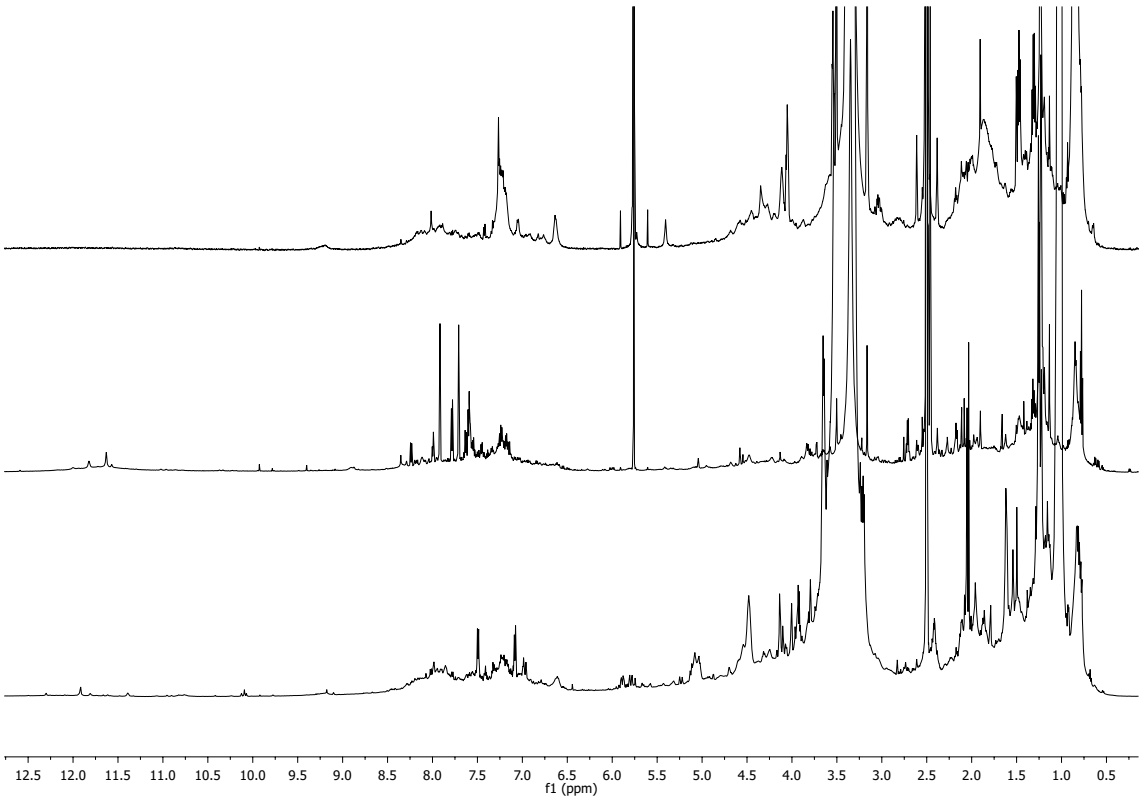



**Figure 3.8.** $^{1}$H NMR spectra ($(CD_3)_2SO$) of the 75% $Me_2CO$ in $H_2O$ HP20 fractions of liquid AOM1 media (top) and strain T81 cultures, 2 (middle) and 16 L (bottom).

observed at 400 $\mu$g/mL against *S. aureus* or *E. coli*. The 30%, 75% and 100% $Me_2CO$ fractions were analysed by MALDI-TOF MS at concentrations of 800, 400, and 200 $\mu$g/mL in DHB matrix using a dried-droplet method. No ions were detected in the mass region *m/z* 3000 – 4000.

**Solid media**

Competition experiments with five co-strains of bacteria (XS 01.96, TKA 04.11, WRG 1.1, WKT 21.8, and TKA 04.12) were cultured for analysis by whole-cell MALDI-TOF MS. Using a pipette tip, a pinhead amount of cells was carefully scraped from the agar surface in three different regions for each plate, a colony of either strain and from within the zone of growth inhibition, and MALDI target spots were prepared in duplicate for each scraping using a dried-droplet method of sample preparation. MS spectra collected from the sample spots revealed clusters of ions in the mass region *m/z* 3000 – 4000. Using the predicted isotopic distribution pattern for the tikitericin core peptide ($C_{144}H_{241}N_{43}O_{50}S_4$), the monoisotopic peak was identified for each ion cluster and recorded (Table 3.2). No ions of $m/z \geq 3000$ were detected in samples of a blank media control.

**Table 3.2.** Whole-cell MALDI-TOF MS screening of co-culture experiments.

| T81 vs | Region | Ions detected, monoisotopic (*m/z* 3000 – 4000) |
|---|---|---|
| XS 01.96 | T81 | 3123.4, 3291.5, 3307.5, 3323.5, 3430.6 |
| XS 01.96 | XS01.96 | 3123.4, 3291.5, 3307.5, 3323.5, 3430.6 |
| XS 01.96 | Zone | 3010.3, 3066.3, 3123.3, 3141.4, 3222.6, 3236.5, 3291.5, 3297.5, 3307.6, 3323.5 |
| TKA 04.11 | T81 | 3291.3, 3307.3, 3322.3 |
| TKA 04.11 | TKA04.11 | 3123.3, 3291.3, 3307.3, 3314.3, 3329.3 |
| TKA 04.11 | Zone | 3235.4, 3259.4, 3274.4, 3291.5, 3307.3, 3323.3 |
| WRG 1.1 | T81 | 3291.5, 3307.5, 3323.5, 3430.6, 3430.5, 3452.5, 3468.5 |
| WRG 1.1 | WRG | 3291.5, 3307.5 |
| WRG 1.1 | Zone | 3291.5, 3307.5, 3330.5, 3452.5, 3469.5 |
| WKT 21.8 | T81 | 3291.5, 3309.5, 3430.5, 3452.5, 3468.6, 3470.6 |
| WKT 21.8 | WKT | 3431.5, 3469.5, 3470.6 |
| WKT 21.8 | Zone | 3430.4, 3451.5, 3468.5 |
| TKA 04.12 | T81 | 3291.5, 3307.5, 3324.5, 3429.6, 3452.5, 3469.5 |
| TKA 04.12 | TKA04.12 | 3291.5, 3307.5, 3313.5, 3329.5 |
| TKA 04.12 | Zone | 3123.5, 3291.5, 3307.5, 3324.5 |
| Media | | |

A quasimolecular ion of *m/z* 3429.6 was detected by whole-cell MALDI-TOF MS across multiple co-strains and regions (Figure 3.9). This mass is consistent with four-fold dehydration of the genetically-encoded core peptide ($C_{144}H_{241}N_{43}O_{50}S_4$, monoisotopic mass = 3500.6520 Da) and subsequent cyclisation to form the putative lanthipeptide, tikitericin ($C_{144}H_{233}N_{43}O_{46}S_4$, monoisotopic mass = 3428.6098 Da), detected as the $[M + H]^+$ ion. The following cationised ions were detected: *m/z* 3451.5 $[M + Na]^+$; *m/z* 3467.6 $[M + K]^+$; *m/z* 3473.6 $[M – H + 2Na]^+$.

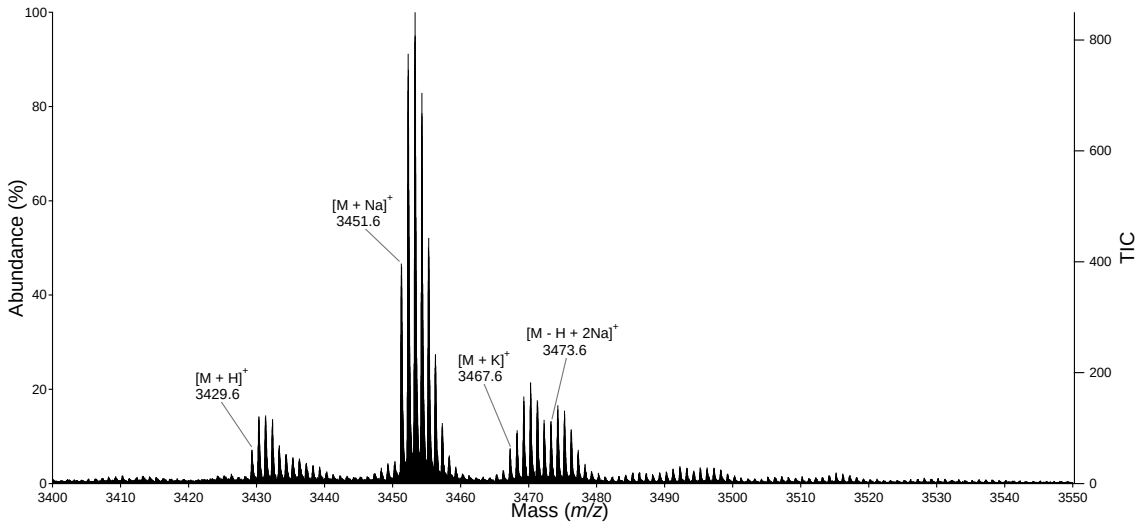
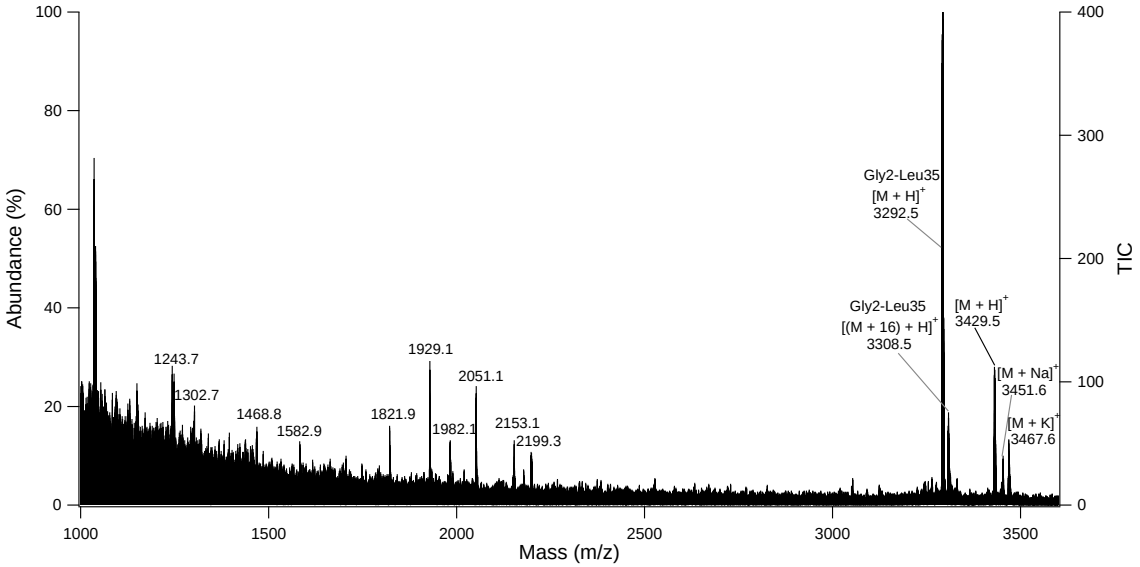


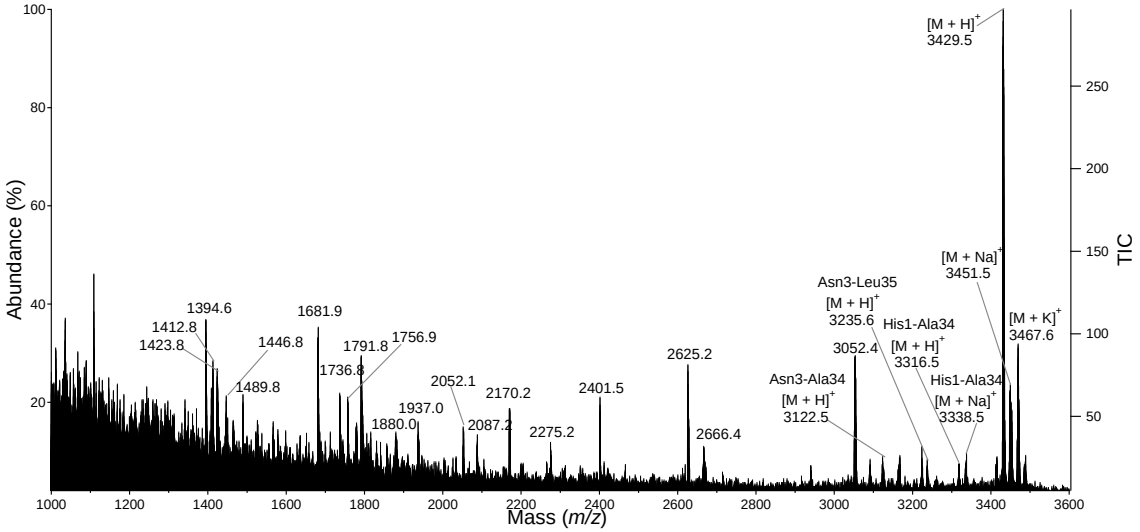

**Figure 3.9.** Whole-cell MALDI-TOF MS spectrum of tikitericin, scraping taken from a single-strain T81 culture.

Exemplified by Figure 3.10, a number of other ions were detected by whole-cell MALDI-TOF MS, several of which are consistent with N- and/or C-terminal hydrolysis products of tikitericin. The truncated versions of tikitericin were tentatively characterised as arising from: N-terminal hydrolysis of the peptide bonds His1 – Gly2 (to form Gly2-Leu35, M = 3291.5509 Da) and Gly2 – Asn3 (Asn3-Leu35, M = 3234.5494 Da); C-terminal hydrolysis of Ala34 – Leu35 (His1-Ala34, M = 3315.5257 Da); and hydrolysis of Gly2 – Asn3 and Ala34 – Leu35 (Asn3-Ala34, M = 3121.4453 Da). Truncated peptides were detected in greater abundance than tikitericin in many of the whole-cell MALDI MS

spectra, e.g. Figure 3.10a. Figure 3.10 additionally demonstrates the significant variation between spectra collected from within a single sample spot in the ions detected and their relative abundance compared to tiktitericin. The relevance of these peptides is discussed further in Section 3.3.4.




**(a)** Sample 1.




**(b)** Sample 2.

**Figure 3.10.** Whole-cell MALDI-TOF MS spectra of a strain T81/XS 01.96 co-culture, scrapings taken from the zone of inhibition.

MALDI-TOF MSMS of tikitericin (precursor ions: *m/z* 3429 $[M + H]^+$, *m/z* 3451 $[M + Na]^+$) weakly detected three fragment ions of *m/z* 1535.6, *m/z* 1648.7 and *m/z* 1776.8, corresponding to fragmentation across amide bonds and the detection of b15, b16 and b17 ions, respectively (Figure 3.11). Detection of these ions suggests two 'halves' of the molecule linked by a short chain, giving the four possible structures shown (Figure 3.12). The suppressed fragmentation pattern supports a cyclised peptide with overlapping rings; cyclic structures are less susceptible to fragmentation than linear peptide regions.[86] Alternatively, the suppression could be attributed to low precursor ion abundance or a reflection of low proton mobility. The ions b15 – b17 are products of the bx –

yz fragmentation pathway (see Section 3.4.2), a model that was initially proposed by Biemann and has since been refined by many research groups. According to this 'mobile proton' model, fragmentation across a peptide bond requires the involvement of a proton at the cleavage site.[101] Strongly basic amino acids such as lysine, arginine and histidine have a deleterious effect on proton migration and a higher energy input is required to move the proton to sites along the peptide backbone. As a consequence, compared to peptides that do not contain basic residues, less sequence information is acquired under the same collision conditions. Willard and Kinter showed that the phenomenon is highly dependent on the position of the basic residue, with a reduction in sequence information for histidine residues close to the N-terminus.[102] The poor fragmentation pattern observed in the tandem MSMS of tikitericin is consistent with the idea that an N-terminal histidine impedes mobility of the N-terminal amide proton and, subsequently, fragmentation of the singly-charged precursor ion. No additional fragmentation could be induced by increasing the laser energy, and the ring topology of tikitericin could therefore not be assigned on the basis of MALDI-TOF MSMS data alone. Four plausible structural isomers are depicted in Figure 3.12, having either linear or bridged connectivity of ring A/B and C/D pairs.

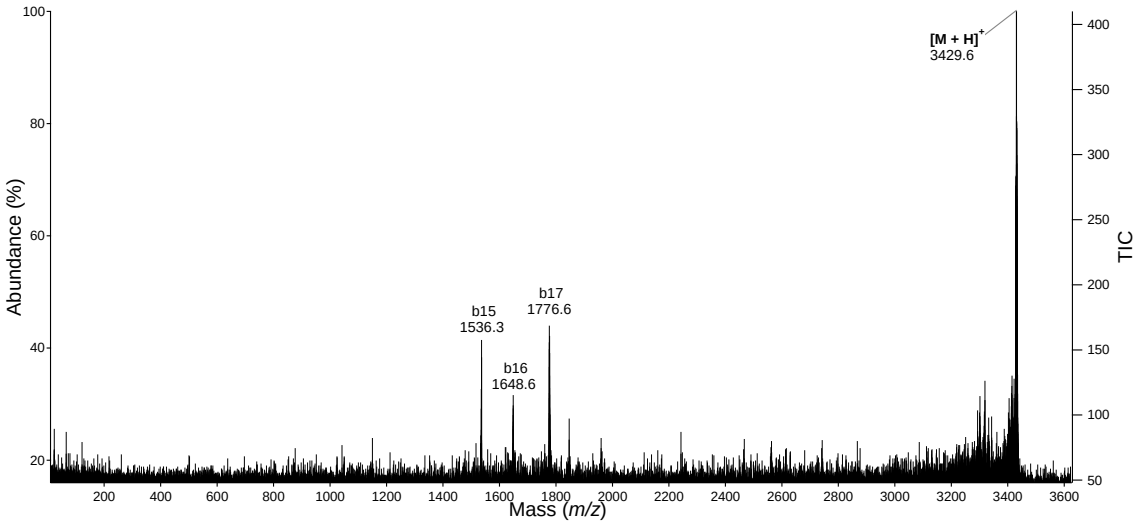
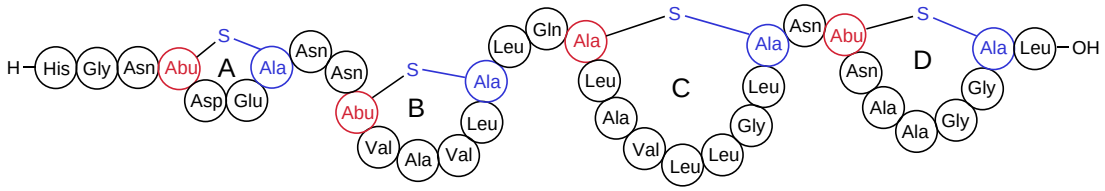
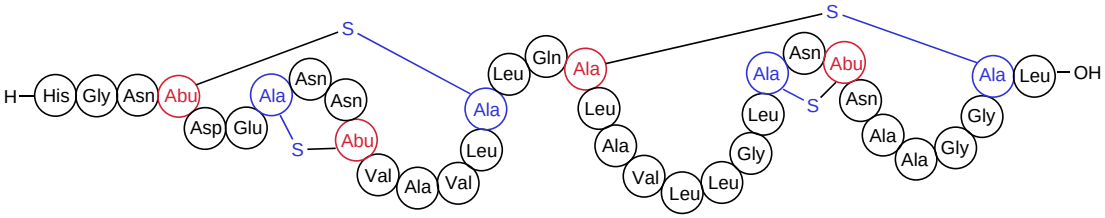


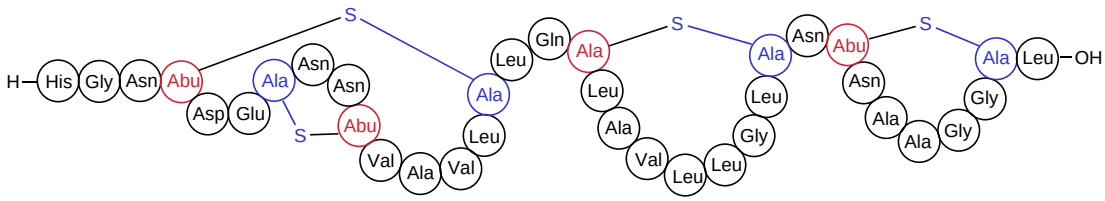

**Figure 3.11.** Whole-cell MALDI-TOF MSMS spectrum of tikitericin (M = 3428.6098 Da), precursor ion: *m/z* 3429 $[M + H]^+$.



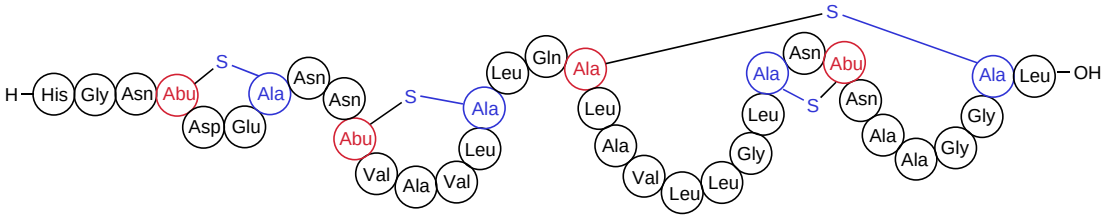

**(a)** Linear.




**(b)** Bridged.




**(c)** Bridged.




**(d)** Bridged.

**Figure 3.12.** Proposed tikitericin ring topology.

There are several complications surrounding the use of whole-cell MALDI-TOF MS for characterising novel RiPPs. The method of sample preparation does not include a concentrating (de-salting) step; if metabolites are not abundant on the agar surface, low S/N ratios will be observed. In turn, low ion count reduces the ability to perform informative fragmentation analyses, as observed in the MALDI-TOF MSMS analysis of the tikitericin $[M + H]^+$ ion. Additionally, low S/N ratios prevent the confident assignment of monoisotopic peaks since isotopic clusters often overlap or are not resolved from the background noise. Secondly, despite calibration to $\pm$ *m/z* 0.05, detected ions showed variation to $\pm$ *m/z* 0.2, revealing the instrument's low accuracy in the case of low ion abundance. The crude method of sample preparation resulted in areas of localised concentration within target spots and therefore variance in MS spectra across the sample spot area. Attempts to improve the matrix crystallisation were made and some improvement was observed by assisting the drying process with hot air, increasing both the ratio of MeCN and the matrix concentration, ensuring that spectra were collected from the crystal edges, and using a 'bottom-up' rather than 'dried-droplet' method of sample preparation.[55] Although useful as a means of viewing the total metabolite level in a sample, low resolution and weak ion signals prevented the use of whole-cell MALDI-

TOF MS for the complete characterisation of tikitericin or other metabolites.

In order to fully characterise tikitericin it was necessary to isolate the lanthipeptide from cultures of strain T81. Early attempts by M. Stott (GNS Science) to clone the lanthipeptide gene cluster into a mesophile host were unsuccessful due to the size and instability of the resulting plasmid vector. Screening of liquid cultures had revealed that the lanthipeptide gene cluster was likely to be inactive under the batch-reactor fermentation growth conditions employed by Callaghan Innovation or standard shaken methods employed by GNS Science. Two late-growth cultures of strain T81 were fractionated by RP HP20 and analysed by MALDI-TOF MS; the $[M + H]^+$ ion was detected with a low ion count in a single fraction obtained from the extracted cell mass of an overgrown strain T81 liquid culture (2 L). A comparison of tikitericin abundance under nine solid media growth conditions was therefore performed (Table 3.3). For each culture, pinhead amounts were carefully picked from the agar surface with a pipette, mixed with aliquots of DHB matrix (5 and 10 $\mu$L), spotted onto a MALDI MS target plate and dried with hot air. Ten MS spectra were collected for each culture condition and the tikitericin $[M + H]^+$ ion count range recorded. Spectra were collected using 5000 shots/spectrum, a laser intensity of 7000 and random continuous stage motion. Areas of localised metabolite concentration resulted in broad ion count ranges, limiting the accuracy in determining optimum growth conditions. Nevertheless, two overarching trends were observed: (i) degradation of tikitericin was only observed in co-cultures, and (ii) the ion count of the tikitericin $[M + H]^+$ peak increased in proportion to the incubation time of strain T81. For the purpose of isolating tikitericin, therefore, single-strain cultures and longer incubation times were preferable.

**Table 3.3.** Optimisation of strain T81 cultivation conditions for tikitericin biosynthesis.

| T81 incubation (d) | Co-strain | Co-strain incubation (d) | $[M + H]^+$ ion count | Degradation (Y/N) |
|---|---|---|---|---|
| 3 | – | – | 60 – 451 | N |
| 5 | – | – | 50 – 309 | N |
| 26 | – | – | 334 – 462 | N |
| 5 | TKA04.11 | 1 | 58 – 85 | Y |
| 5 | XS01.96 | 1 | 50 – 76 | Y |
| 26 | TKA04.11 | 1 | 50 – 649 | Y |
| 26 | XS01.96 | 1 | 58 – 719 | Y |
| 26 | TKA04.11 | 4 | 51 – 219 | Y |
| 26 | XS01.96 | 4 | 23 – 52 | Y |

### 3.3.2 Isolation of tikitericin

Tikitericin was isolated from batches of single-strain T81 bacilli grown on plated AOM1 media at 50 °C for 14 d. MeOH extracts were chromatographically separated over HP20 resin (see Appendix C). Tikitericin was detected by MALDI-TOF MS in 40%, 50% and 60% $Me_2CO$ fractions, and the purification strategy was subsequently condensed to generate only three fractions of 30%, 60% and 100% $Me_2CO$. The $^1$H NMR spectra of all

HP20 fractions (exemplified by the 50% $Me_2CO$ HP20 fractions, Figure 3.13) indicated that media components constituted the majority of mass present in strain T81 fractions.

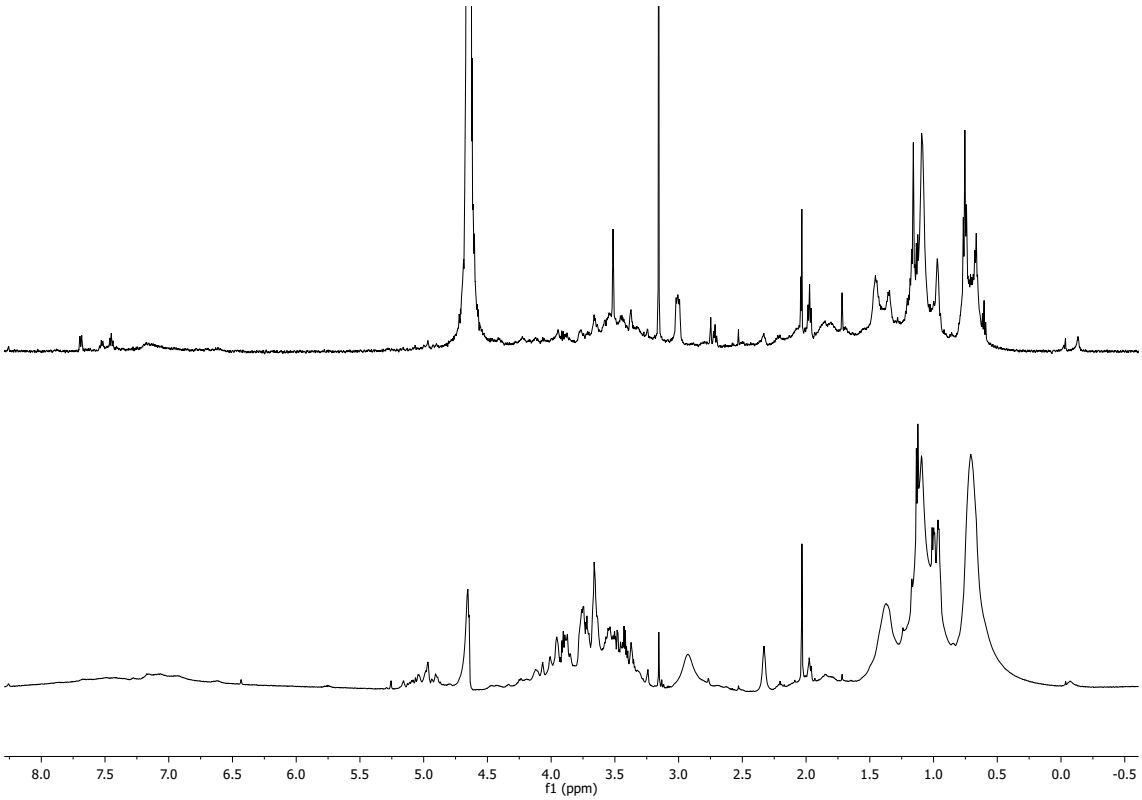



**Figure 3.13.** $^1$H NMR spectra ($D_2O$) of the 50% $Me_2CO$ in $H_2O$ HP20 fractions of solid AOM1 media (top, 6 plates) and strain T81 (bottom, 60 plates).

The HP20 fractions from three batches of strain T81 ($\approx$ 300 plates) were combined and partitioned further over Sephadex® LH-20 size-exclusion resin with an isocratic solvent composition of 50% MeOH/$H_2O$ and a flow rate of 0.4 mL/min. Fraction combination was guided by MALDI-TOF MS analysis, which showed that tikitericin eluted earlier than mid-range molecular weight compounds of *m/z* 1000 – 2000 but co-eluted with compounds of *m/z* 2500 – 3500. LH-20-purified tikitericin was adequate for the purpose of amino acid sequence characterisation and sub-samples were used in proteolytic digest and linearisation experiments (Sections 3.4 and 3.5). However, an abundance of low molecular weight ions in MALDI-TOF and HR ESI-MS spectra suggested that tikitericin represented only a minor proportion of the fraction. Further separation of tikitericin from presumed culture media contamination was required for stereochemical analysis (Section 3.6) and studying biological activity (Section 3.7).

A gradient method for reversed-phase C18 high-performance liquid chromatography (HPLC) purification was developed using HR ESI-MS. Tikitericin eluted with approximately 39% MeCN. This method was subsequently transferred to a semipreparative HPLC, guided by simultaneous UV and evaporative light scattering detection (ELSD). An adjustable flow splitter with a split ratio of 20:1 was used to split the flow from the DAD to the ELSD detector. Collection tubing lengths and diameters were adjusted to generate the smallest loss in efficiency and acceptable backpressure. It was found

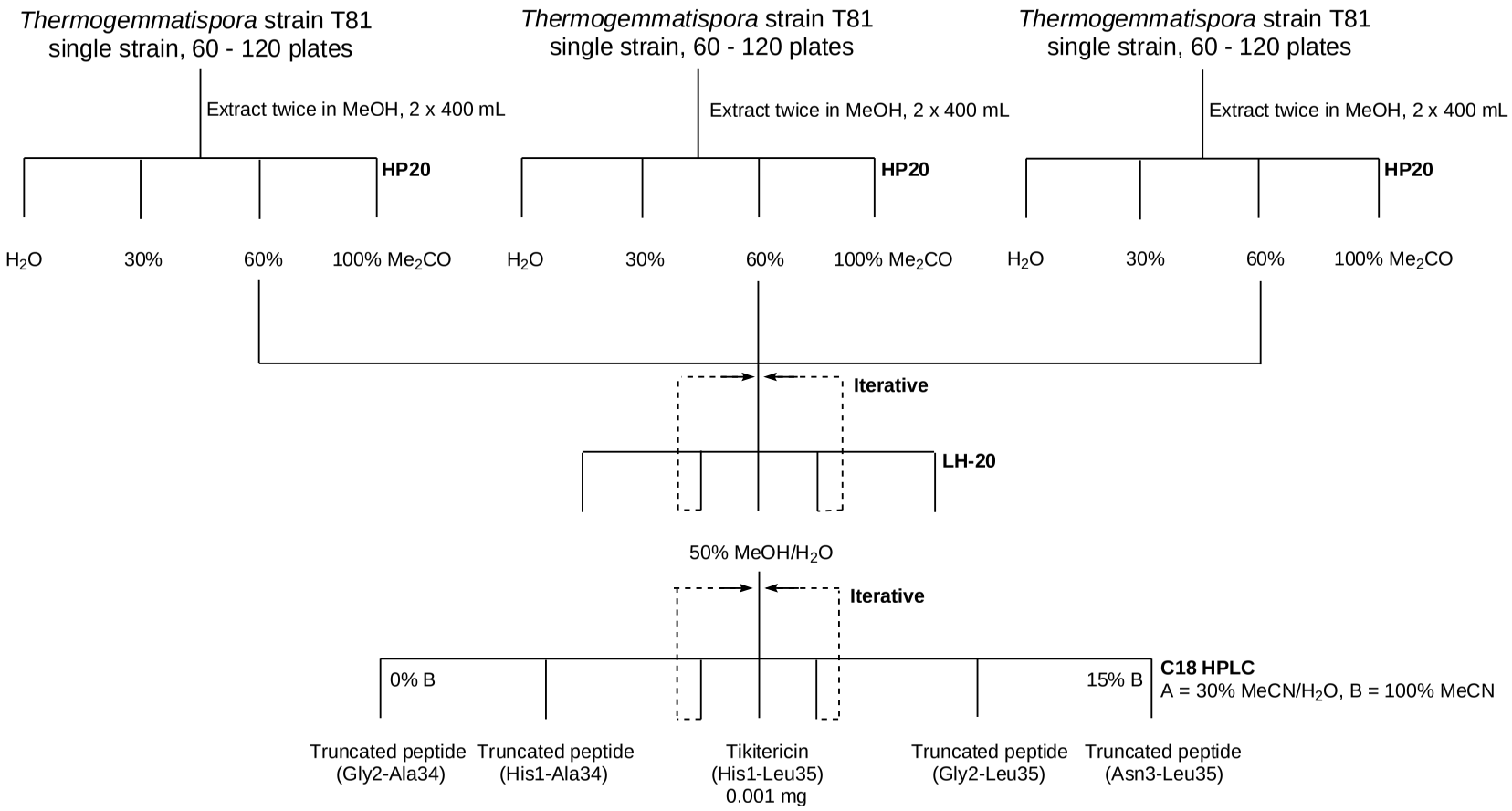



**Figure 3.14.** Purification strategy employed in the isolation of tikitericin, repeated four times.

that a pre-filtration step of the extract was required due to solubility problems related to injection. An injection solvent of elution strength equivalent to the initial mobile phase could not be used because of the wide polarity range of the compounds in the extract and the large difference in concentration between major compounds and tikitericin led to column overloading and thus coelutions. The extract was therefore suspended in MeOH and filtered over a 0.45 $\mu$m filter prior to injection.

An overview of the isolation of tikitericin is depicted in Figure 3.14. This process was repeated four times to generate a single fraction of highest purity from 1200 plates, required for biological activity studies. Fractions of lesser purity were re-cycled into previous steps.

### 3.3.3 Identification of tikitericin by HR ESI-MS

The molecular formula of tikitericin was determined from HR ESI-MS data obtained from LH-20 purified material. The *m/z* of singly charged ions lies outside the instrument mass range (*m/z* 100 – 3200) but tikitericin was detected as a series of multiply-charged quasimolecular ions.

**Tikitericin:** M = 3428.6098 Da, *m/z* 858.1594 $[M + 4H]^{4+}$ (calculated, 858.1597 Da; $\Delta$ – 0.40 ppm), *m/z* 880.1420 $[M + 4Na]^{4+}$ (calculated, 880.1417 Da; $\Delta$ 0.33 ppm), *m/z* 1143.8779 $[M + 3H]^{3+}$ (calculated, 1143.8772 Da; $\Delta$ 0.60 ppm), *m/z* 1166.1927 $[M + 3Na]^{3+}$ (calculated, 1166.1934 Da; $\Delta$ – 0.59 ppm), *m/z* 1715.3153 $[M + 2H]^{2+}$ (calculated, 1715.3122 Da; $\Delta$ 1.82 ppm), *m/z* 1737.3008 $[M + 2Na]^{2+}$ (calculated, 1737.2941 Da; $\Delta$ 3.86 ppm).

### 3.3.4 *In situ* Tikitericin degradation products

A second proteolytic cleavage of N-terminal residues from the propeptide has been reported for several class II lantibiotics,[103–108] and in some cases the removal of these residues is necessary for bioactivity.[103,104] There is no evidence in the tikitericin gene cluster for genetically-controlled N-terminal hydrolysis of the core peptide (Table 3.1). The hydrolysis may be: (i) catalysed by a protease secreted by the co-strain as a defense mechanism, (ii) catalysed by a protease secreted by strain T81, triggered by the presence of a co-strain, and required to activate the mature peptide, (iii) catalysed randomly by the action of non-selective proteases, or (iv) a result from thermal degradation of the bioactive metabolite.

A series of truncated peptides were initially detected by whole-cell MALDI-TOF MS in plated cocultures but were later detected in single-strain plated cultures of T81.

Purification by size-exclusion chromatography and C18 HPLC enabled the partial characterisation of these minor peptides by HR ESI-MS but the isolated biomass was insufficient to perform full structural analyses by tandem MSMS. Seven truncated analogues of tikitericin were identified. The identification of peptides derived from hydrolysis across every residue not contained within a macrocycle, rather than the existence of a single terminally-cleaved peptide, suggests that these peptides are the result of degradation rather than activation.

**Gly2-Leu35:** M = 3291.5509 Da, *m/z* 1098.8615 $[M + 3H]^{3+}$ (calculated, 1098.8591 Da; $\Delta$ 2.24 ppm), *m/z* 1646.788 $[M + 2H]^{2+}$ (calculated, 1646.7827 Da; $\Delta$ 3.22 ppm).

**Asn3-Leu35:** M = 3234.5494 Da, *m/z* 1079.1821 $[M + 3H]^{3+}$ (calculated, 1079.1837 Da; $\Delta$ – 1.56 ppm), *m/z* 1618.2797 $[M + 2H]^{2+}$ (calculated, 1618.2720 Da; $\Delta$ 3.80 ppm).

**His1-Ala34:** M = 3315.5257 Da, *m/z* 1106.1855 $[M + 3H]^{3+}$ (calculated, 1106.1825 Da; $\Delta$ – 2.71 ppm), *m/z* 1658.7720 $[M + 2H]^{2+}$ (calculated, 1658.7701 Da; $\Delta$ – 1.13 ppm).

**Gly2-Ala34:** M = 3178.4668 Da, *m/z* 1060.4867 $[M + 3H]^{3+}$ (calculated, 1060.4962 Da; $\Delta$ 8.92 ppm), *m/z* 1607.2241 $[M + 2H]^{2+}$ (calculated, 1607.2672 Da; $\Delta$ 4.06 ppm), *m/z* 1612.2273 $[M + 2Na]^{2+}$ (calculated, 1612.2226 Da; $\Delta$ – 2.88 ppm).

**Asn3-Ala34:** M = 3121.4453 Da, *m/z* 1041.4844 $[M + 3H]^{3+}$ (calculated, 1041.4891 Da; $\Delta$ 4.45 ppm), *m/z* 1561.7300 $[M + 2H]^{2+}$ (calculated, 1561.7299 Da; $\Delta$ – 0.05 ppm).

**Asn9-Leu35:** M = 2576.3277 Da, *m/z* 1289.1720 $[M + 2H]^{2+}$ (calculated, 1289.1711 Da; $\Delta$ 0.67 ppm), *m/z* 2577.3343 $[M + H]^{+}$ (calculated, 2577.3350 Da; $\Delta$ – 0.25 ppm).

**Abu10-Leu35:** M = 2462.2848 Da, *m/z* 1232.1443 $[M + 2H]^{2+}$ (calculated, 1232.1497 Da; $\Delta$ – 4.33 ppm), *m/z* 2463.2901 $[M + H]^{+}$ (calculated, 2463.2921 Da; $\Delta$ – 0.25 ppm).

## 3.4 Determination of Tikitericin Ring Topology

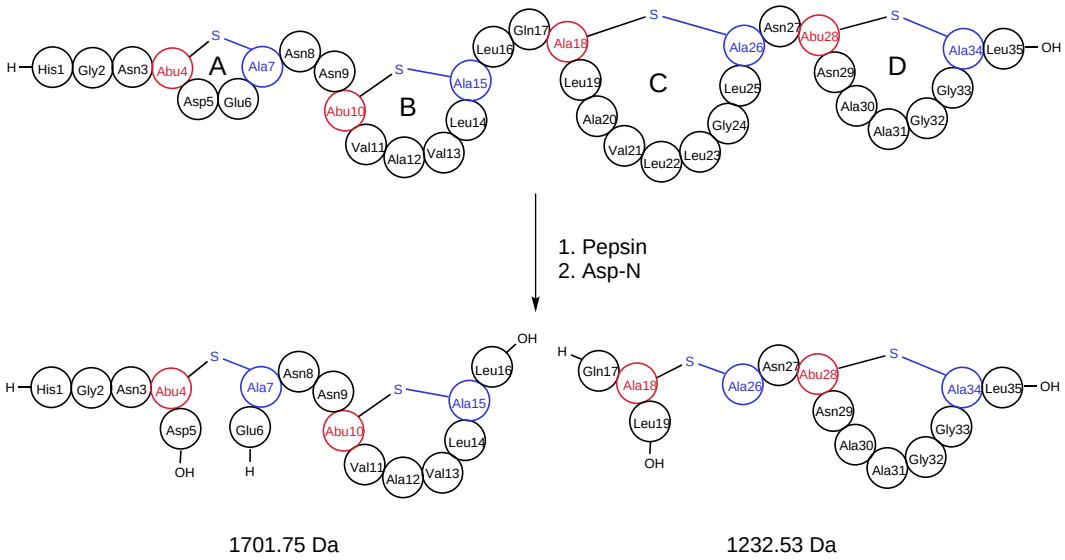
### 3.4.1 In-solution proteolytic digestion

Proteolytic enzymes were selected to determine between the four possible cyclised variants of tikitericin according to the specific hydrolytic activity of commercially available enzymes and the resulting characteristic peptide fragments. The reaction conditions for each digestion experiment were evaluated using samples of commercially available nisin A as a model prior to tikitericin digestion.

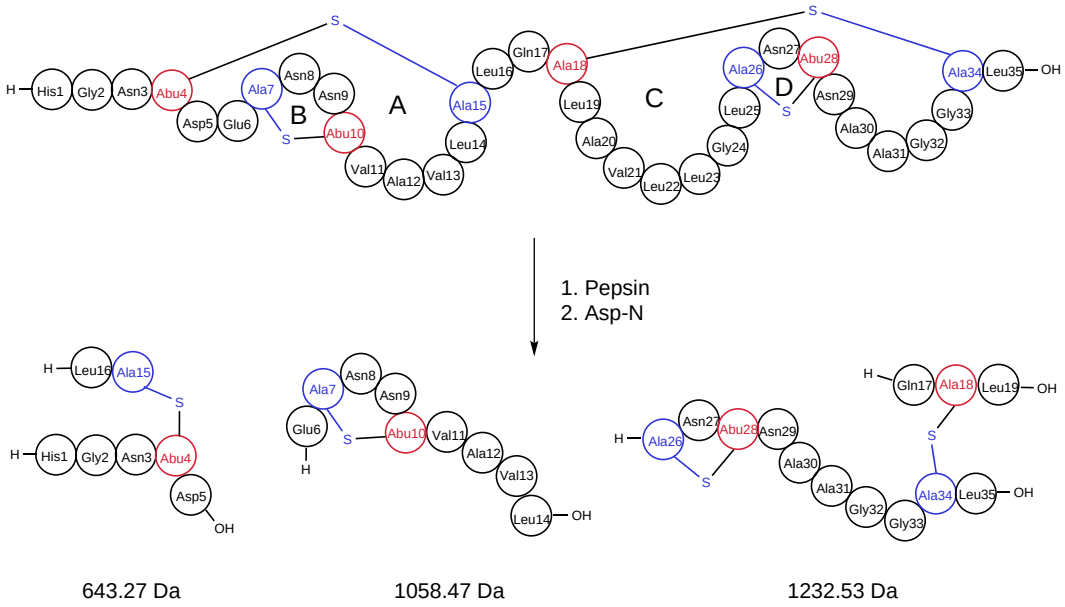
#### Asp-N/thermolysin and Asp-N/pepsin digests

To determine the topology of the ring A/B pair, a dual digest approach was applied. Pepsin or thermolysin used in conjunction with endopeptidase Asp-N would produce characteristic ions for the linear and bridged isomers of rings A and B (Scheme 3.2). The specificity of pepsin includes cleavage at the C-terminus of Phe and Leu, and to a lesser extent at the C-terminus of Glu residues. Pepsin will not cleave at Val, Ala, or Gly linkages. Thermolysin is an endopeptidase that preferentially cleaves at the N-terminus of the hydrophobic residues Leu, Phe, Val, Ile, Ala and Met. Asp-N is a metalloendoprotease that hydrolyses peptide bonds on the N-terminal side of Asp and cysteic acid residues. For the Asp-N/thermolysin (or Asp-N/pepsin) digest, a characteristic fragment of *m/z* 1702.8 $[M + H]^+$ would support linear connectivity of rings A and B (Scheme 3.2a), while characteristic fragments of *m/z* 1058.5 $[M + H]^+$ and *m/z* 644.3 $[M + H]^+$ would support bridged connectivity of rings A and B (Scheme 3.2b).

The activity of pepsin was successfully trialled with nisin but was not successful in producing hydrolysed products when used with tikitericin. Similarly, digestion of tikitericin with Asp-N and thermolysin did not produce products that could be assigned to logical hydrolysis fragments. MALDI-TOF MS after digestion with Asp-N showed an insignificant amount of hydrolysis, with enzyme self-digestion occurring at longer reaction times and elevated temperatures. After incubation with thermolysin for 6 h, none of the predicted fragments were observed. These results are consistent with similar experiments reported in the literature.[109] Efficient enzyme activity generally requires a protein to be linearised by reduction and alkylation of the peptide disulfide bonds prior to digestion. In the peptide's native state, proteolytic enzymes cannot gain access to the amino acid residues.




**(a)** Linear topology of rings A and B.

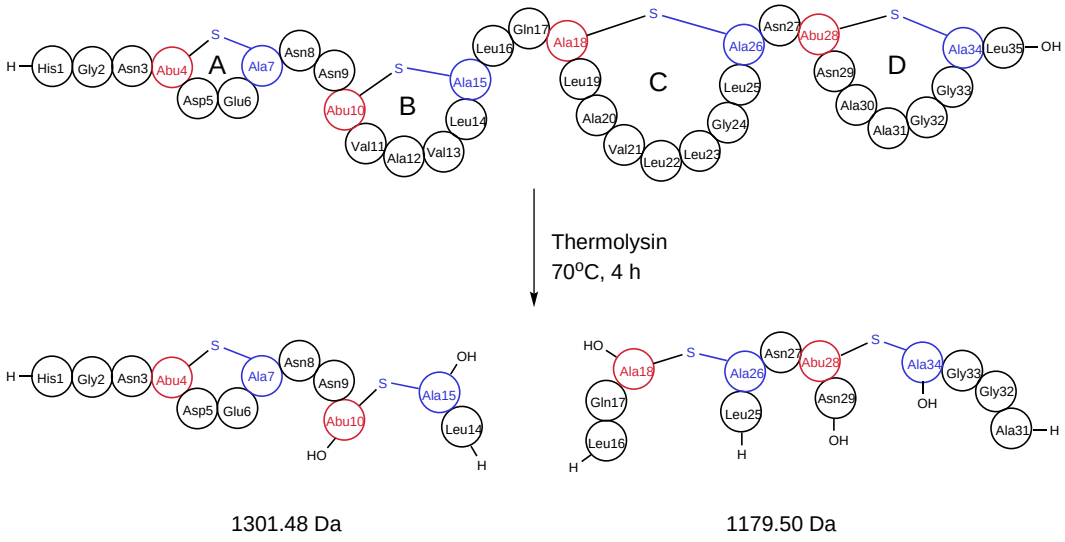



**(b)** Bridged topology of rings A and B.

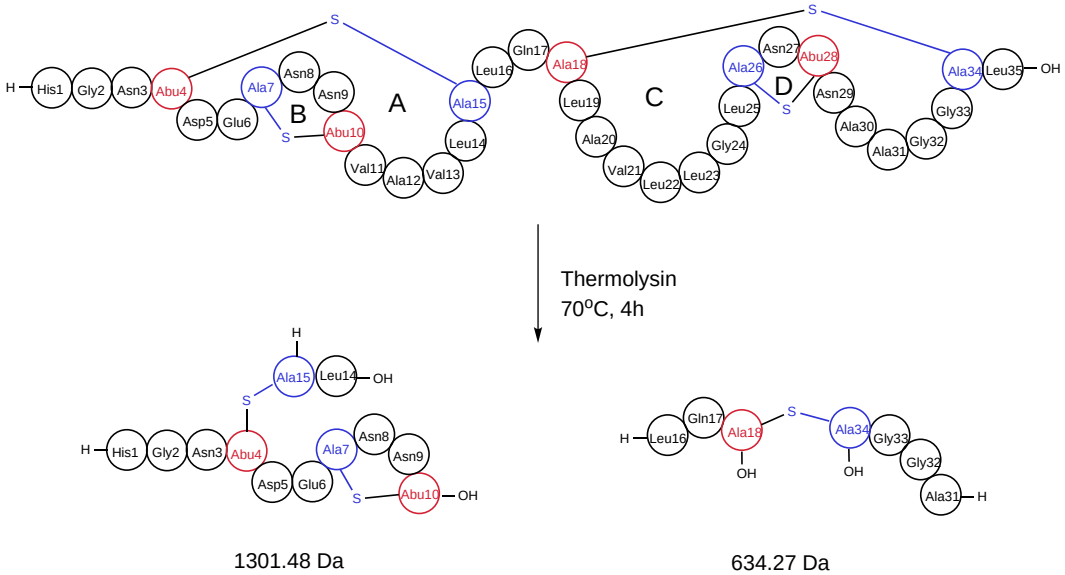
**Scheme 3.2.** Asp-N/pepsin digest strategy to determine connectivity of tikitericin A and B rings.

**Thermolysin digest**

It was proposed that an in-solution digest of tikitericin with thermolysin, and specific cleavage at Leu, Val and Ala residues, would produce characteristic ions for the linear and bridged isomers of the ring C/D pair (Scheme 3.3). A unique ion of *m/z* 1180.5 $[M + H]^+$ would support linear connectivity of rings C and D (Scheme 3.3a), and a unique ion of *m/z* 635.3 $[M + H]^+$ would support bridged connectivity of rings C and D (Scheme 3.3b).



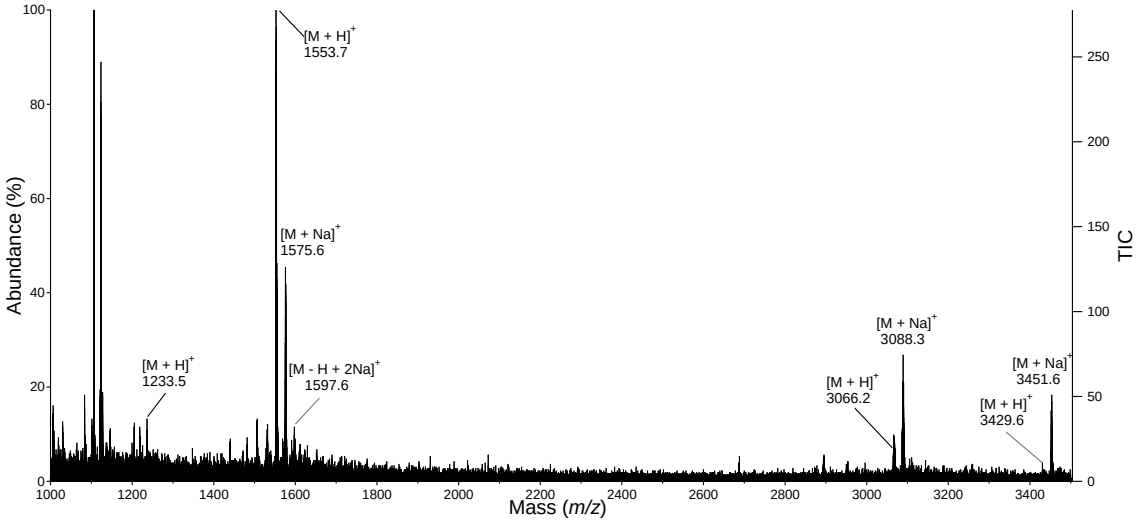

**(a)** Linear topology of rings C and D.



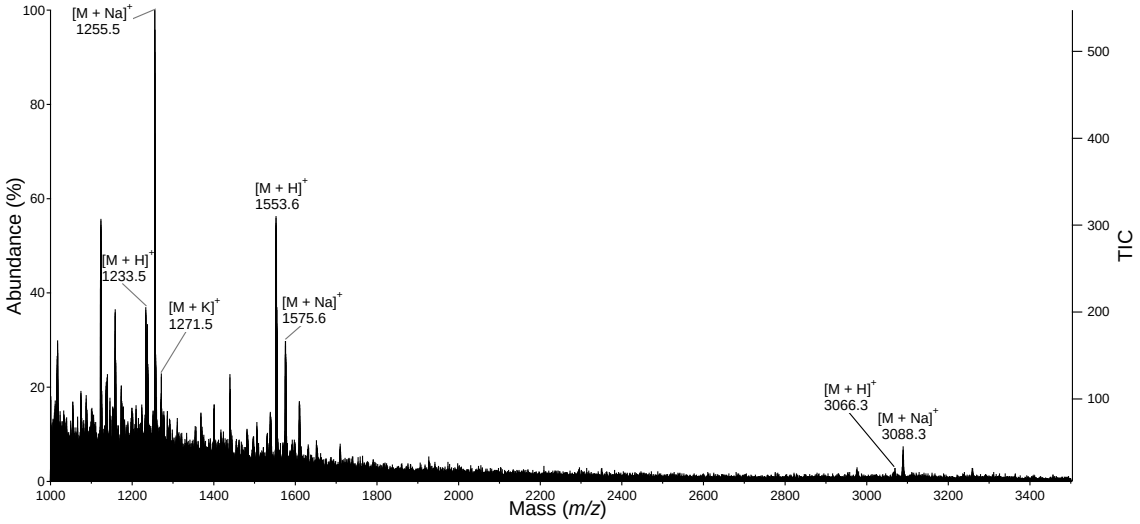

**(b)** Bridged topology of rings C and D.

**Scheme 3.3.** Thermolysin digest strategy to determine connectivity of tikitericin C and D rings.

Over a range of reaction parameters, cleavage of lanthipeptide peptide bonds was shown to occur at the N-terminus of residues Leu16, Leu23, Leu25, Leu35 but not Leu14 or Ala12, as required to determine the Me(Lan) bridge connectivity of rings C and D. Ions of two peptide fragments were observed by MALDI-TOF MS (Figures 3.16 and 3.15).



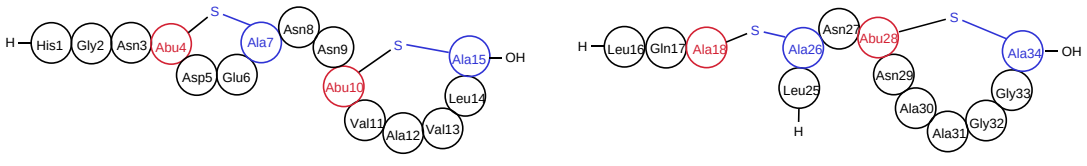

**(a)** Reaction time, 4 h.




**(b)** Reaction time, 6 h.

**Figure 3.15.** MALDI-TOF MS spectra of tikitericin after thermolysin digestion.

Ions of fragment i (**25**) were detected as follows: *m/z* 1553.6 $[M + H]^+$, *m/z* 1575.6 $[M + Na]^+$, and *m/z* 1597.6 $[M – H + 2Na]^+$. Ions of fragment ii (**26**) were detected as: *m/z* 1233.5 $[M + H]^+$, *m/z* 1255.5 $[M + Na]^+$, and *m/z* 1271.5 $[M + K]^+$. HR ESI-MS data was subsequently obtained.




**25** **26**

**Figure 3.16.** Tikitericin-derived hydrolysis fragments observed after in-solution thermolysin digestion.

**Fragment i (25):** M = 1552.6399 Da, *m/z* 777.3282 $[M + 2H]^{2+}$ (calculated, 777.3272 Da; $\Delta$ 1.28 ppm), *m/z* 1553.6422 $[M + H]^{+}$ (calculated, 1553.6471 Da; $\Delta$ – 3.21 ppm).

**Fragment ii (26):** M = 1232.5278 Da, *m/z* 617.2704 $[M + 2H]^{2+}$ (calculated, 617.2712 Da; $\Delta$ – 1.23 ppm), *m/z* 1233.5301 $[M + H]^{+}$ (calculated, 1233.5351 Da; $\Delta$ – 3.99 ppm).

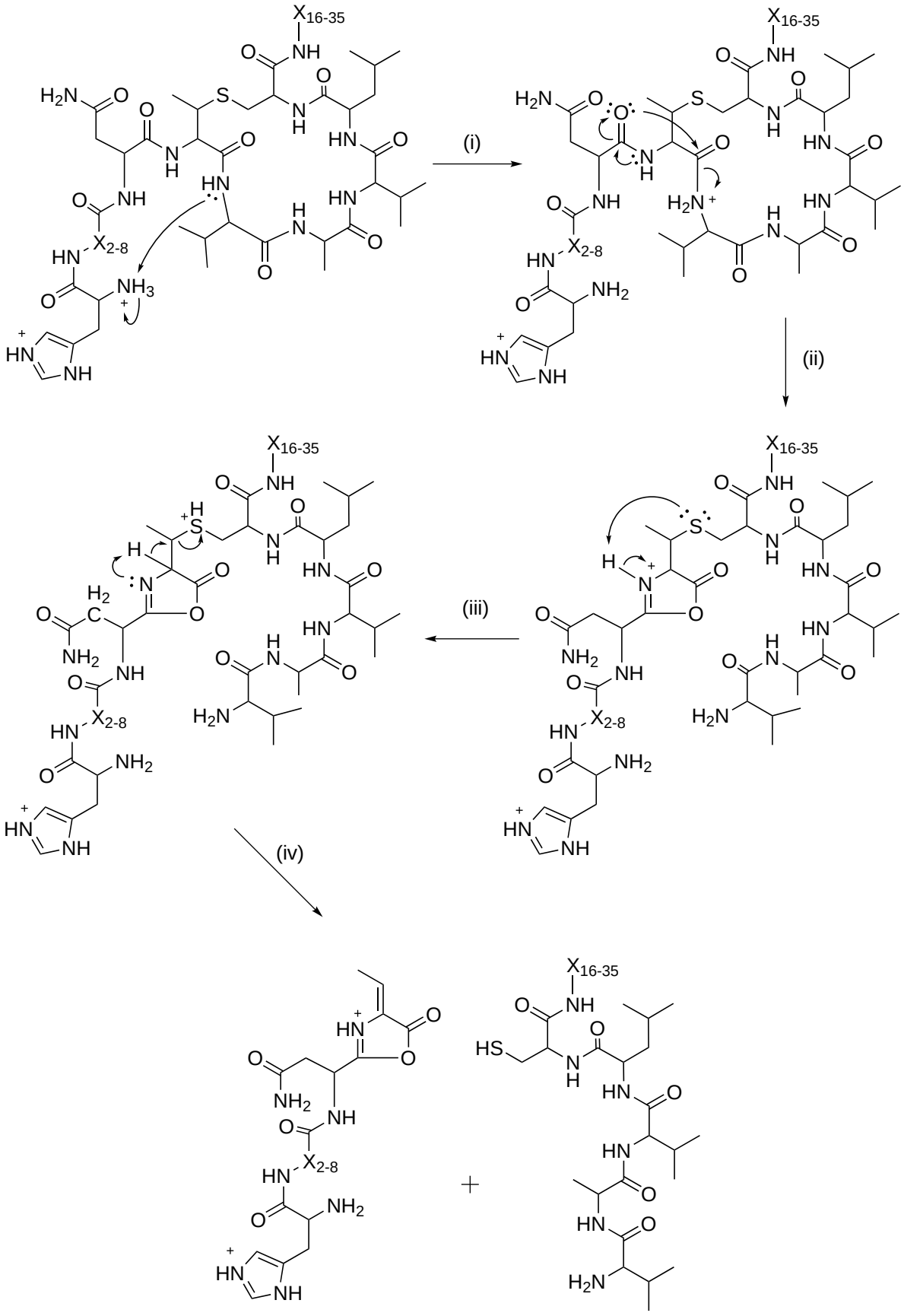
The limited hydrolysis observed in these digest experiments shows the reduced susceptibility of tikitericin to enzymatic degradation compared to linear peptides, a feature that is likely to be mirrored in the naturally-occurring environment into which the putative lantibiotic is secreted. Furthermore, non-hydrolysed parent ions of tikitericin were detected after 4 h at 70 °C (Figure 3.15a), supporting significant thermal stability of tikitericin, another consequence of the conformational rigidity imparted by (Me)Lan residues. Although the two hydrolysis fragments could not fully characterise the (Me)Lan bridge connectivity, the experiments demonstrated the accessibility of thermolysin to several hydrophobic Leu residues.

### 3.4.2 MSMS analysis

To provide insight into the (Me)Lan ring topology, LH-20-purifed tikitericin was subjected to HR ESI-MSMS. It was necessary to apply high CID energies of 50 – 70 arbitrary units to induce even the most facile dissociation pathway, charge-directed fragmentation of peptide bonds, a consequence of the previously mentioned deleterious effect of the N-terminal histidine. Product ion spectra contained both the corresponding b- and y-type so-called 'sequencing ions' (Table 3.4), and were more informative than data from the MALDI-TOF MSMS spectra. The improved dissociation is a direct result of the precursor ion charge states. Doubly and triply charged ions of basic residue-containing peptides require lower energy input for dissociation;[101] a proton can remain sequestered by the histidine residue and additional protons can serve as mobile protons for the dissociation pathway.

It can be seen that, with the exception of ions b10, b11, b18 and b19, fragmentation was suppressed in sequences comprising residues 4 – 7, 10 – 15, 18 – 26 and 28 – 34 (Figures 3.17 and 3.18). This pattern clearly supports linear connectivity of all four (Me)Lan residues (Figure 3.12a, Table 3.4). A mechanism for internal ring fragmentation is proposed in Scheme 3.4, demonstrated by formation of the b10 ion. Following migration of a mobile proton (i), the nitrogen-protonated peptide bond is cleaved by nucleophilic attack of the adjacent N-terminal amide oxygen (ii).*[101] For linear peptides this results in a protonated oxazolone b-ion or, via an ion-molecule complex, a neutral N-terminal fragment and a corresponding y-ion.[101] After dissociation of the tikitericin precursor

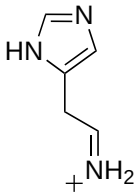
*A second mechanism is accepted whereby the peptide backbone fragmentations occur mainly from protonated oxonium carbonyl forms.



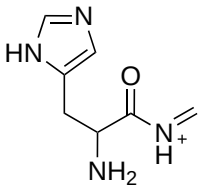

**Scheme 3.4.** Proposed mechanism for the internal ring fragmentation of tikitiericin. Steps (i) and (ii) show the bx – yz dissociation pathway, which forms either a bn or the corresponding yn ion in the dissociation of linear peptides, adapted from Wysocki *et al.*[101] Steps (iii) – (iv) show proton migration and subsequent fragmentation of the thioether bond.

peptide backbone, however, the peptide is still intact, linked by the MeLan thioether bond. Step (iii) shows further proton migration to the adjacent thioether, followed by charge-directed fragmentation to form a conjugated, charge stabilised oxazolone (iv). The tendency of peptides to sequester multiple cations led to the b10 ion being detected as both $[M + H]^+$ and $[M + 2H]^{2+}$ ions, with the additional proton most likely to reside on the N-terminal histidine. This model is a simplification because in a large peptide, protons are probably nested by the carbonyl oxygens of the peptide backbone.[110] Similar internal fragmentation has been observed in the A ring of nisin (**18**) and geobacillin II (**19**).[103,111]

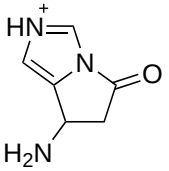
In addition to the key sequencing ions, two a-type immonium ions were observed from the loss of CO from the corresponding bn ion: a1 (**27**) *m/z* 110.0717 (calculated, 110.0718 Da; Δ – 0.91 ppm), and a2 (**28**) *m/z* 167.0917 (calculated, 167.0933 Da; Δ – 9.58 ppm). The a1 ion was observed in greater abundance than b1 (**29**); b1-type oxazolium (acylium resonance form) ions are usually not observed at all because of facile decomposition,[110] however, in this case, the b1 ion is stabilised by the N-terminal His residue. This has been previously observed for N-terminal Lys, His and Met residues.[110]



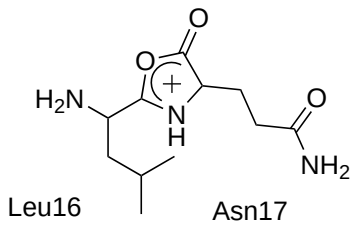
**27**



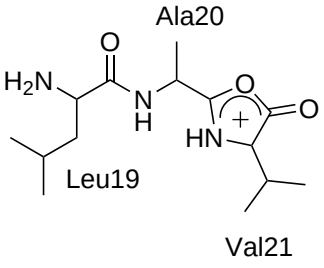
**28**



**29**




**30**




**31**

The high collision energy required to induce proton mobilisation was seen to induce other secondary fragmentation ions, such as ions b15/b17 (**30**) *m/z* 242.1489 (calculated, 242.1499 Da; Δ – 4.13 ppm) and b18/b21 (**31**) *m/z* 284.1968 (calculated, 284.1969 Da; Δ – 0.35 ppm), which result from further b-type fragmentation of the existing fragment ions b17 and b21, respectively. Several prominent non-sequencing ions in the product ion spectra could not be identified and may be the result of additional side-chain or small molecule fragmentations.

Fundamental full characterisation of tikitericin therefore required further experimental study.

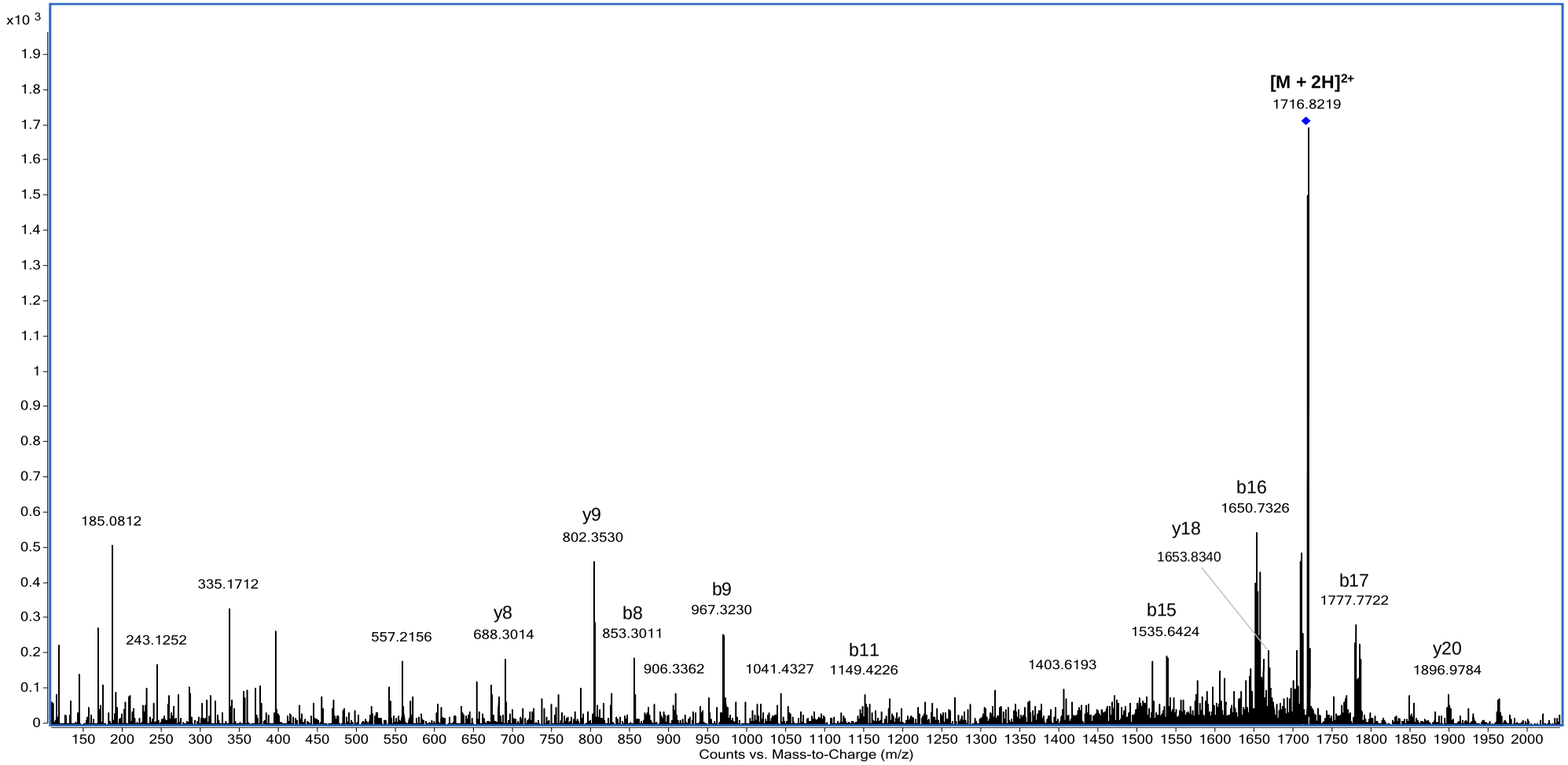



**Figure 3.17.** HR ESI-MSMS spectrum of tikitericin (M = 3428.6098 Da), precursor ion: *m/z* 1716.32 $[M + 2H]^{2+}$, CID energy = 50.

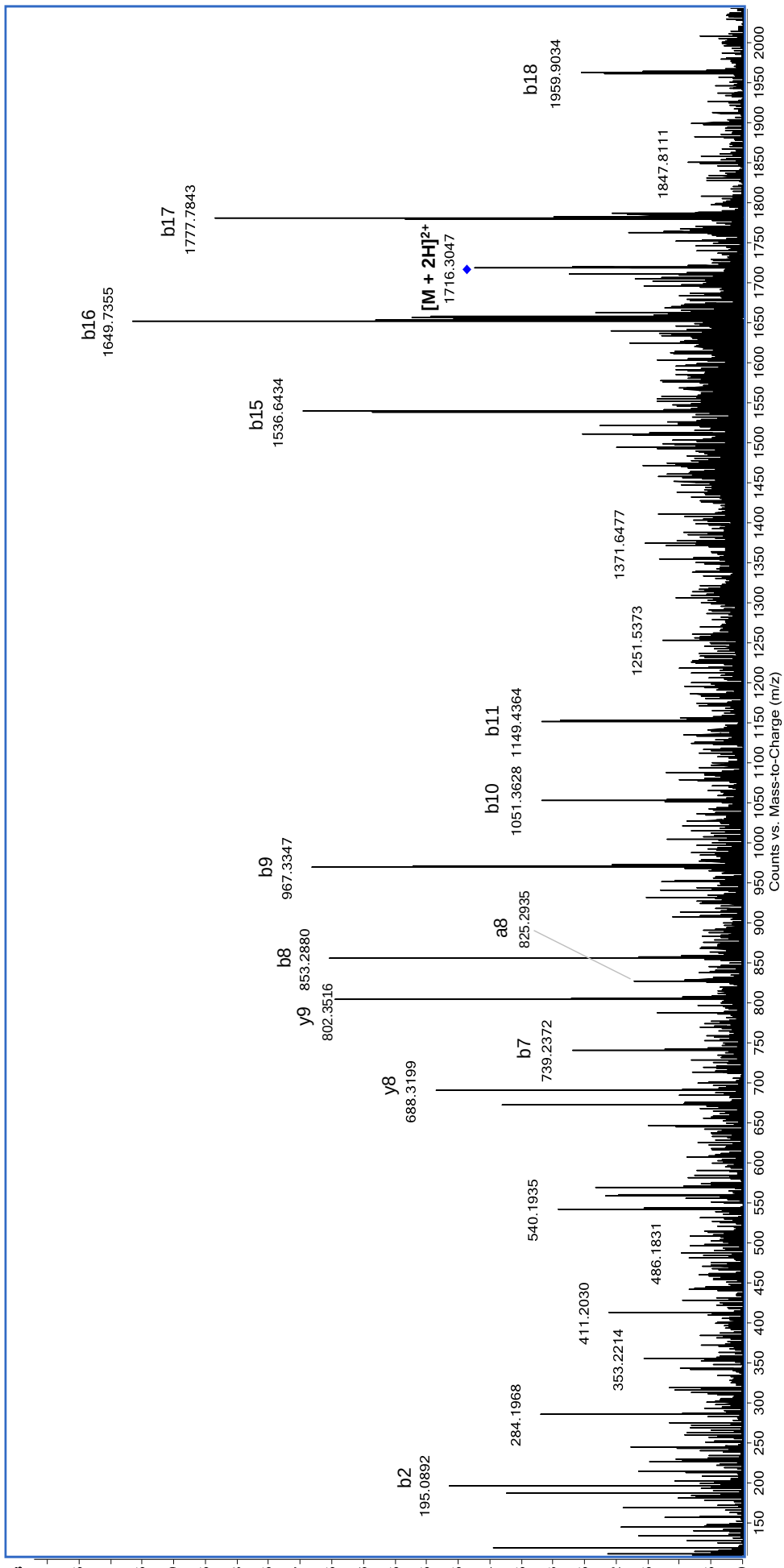
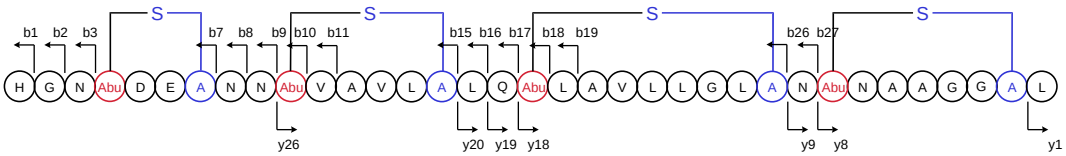



**Figure 3.18.** HR ESI-MSMS spectrum of tikitericin (M = 3428.6098 Da), precursor ion: *m/z* 1716.32 $[M + 2H]^{2+}$, CID energy = 60.

**Table 3.4.** HR ESI-MSMS fragmentation pattern of tikitericin (**32**) (M = 3428.6098 Da), precursor ions: *m/z* 1144.55 $[M + 3H]^{3+}$, *m/z* 1716.32 $[M + 2H]^{2+}$.

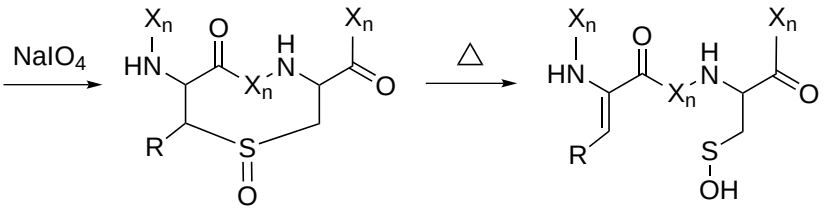



**32**

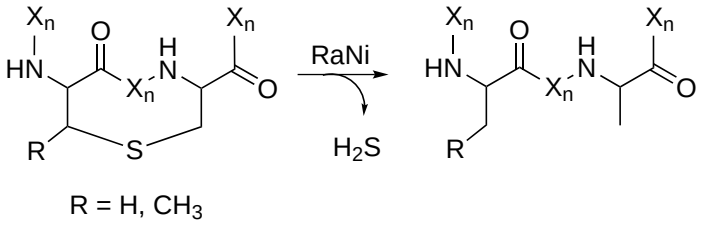
| b ions | | | | y ions | | | |
|---|---|---|---|---|---|---|---|
| Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) | Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) |
| 1 | 138.0667 | 138.0628 | – 28.25 | 34 | 3292.5587 | | |
| 2 | 195.0882 | 195.0884 | 1.03 | 33 | 3225.5373 | | |
| 3 | 309.1311 | 309.1315 | 1.29 | 32 | 3121.4943 | | |
| 4 | 392.1682 | | | 31 | 3038.4572 | | |
| 5 | 507.1952 | | | 30 | 2794.3877 | | |
| 6 | 636.2378 | | | 29 | 2691.3785 | | |
| 7 | 739.2470 | 739.2469 | – 0.14 | 28 | 2691.3785 | | |
| 8 | 853.2899 | 853.2898 | – 0.12 | 27 | 2577.3356 | | |
| 9 | 967.3328 | 967.3331 | 0.31 | 26 | 2463.2927 | 2463.2780 | – 5.83 |
| 10 | 1050.3700 | 1050.3689 | – 1.05 | 25 | 2380.2555 | | |
| 11 | 1149.4383 | 1149.4377 | – 0.52 | 24 | 2281.1871 | | |
| 12 | 1220.4745 | | | 23 | 2210.1500 | | |
| 13 | 1319.5439 | | | 22 | 2111.0816 | | |
| 14 | 1432.6279 | | | 21 | 1997.9975 | | |
| 15 | 1535.6371 | 1535.6386 | 0.98 | 20 | 1894.9883 | 1894.9920 | 1.72 |
| 16 | 1648.7212 | 1648.7233 | 1.27 | 19 | 1781.9043 | 1781.9060 | 0.87 |
| 17 | 1776.7798 | 1776.7806 | 0.45 | 18 | 1653.8457 | 1653.8340 | – 7.35 |
| 18 | 1845.8012 | 1845.7977 | – 1.90 | 17 | 1584.8242 | | |
| 19 | 1958.8853 | 1958.8878 | 1.28 | 16 | 1471.7402 | | |
| 20 | 2029.9224 | | | 15 | 1400.7031 | | |
| 21 | 2128.9908 | | | 14 | 1301.6347 | | |
| 22 | 2242.0749 | | | 13 | 1188.5506 | | |
| 23 | 2355.1589 | | | 12 | 1075.4665 | | |
| 24 | 2412.1804 | | | 11 | 1018.4451 | | |
| 25 | 2525.2645 | | | 10 | 905.3610 | | |
| 26 | 2628.2737 | 2628.2564 | – 6.58 | 9 | 802.3518 | 802.3516 | – 0.19 |
| 27 | 2742.3166 | 2742.3631 | 16.96 | 8 | 688.3089 | 688.3077 | – 1.67 |
| 28 | 2825.3537 | | | 7 | 605.2718 | | |
| 29 | 2939.3966 | | | 6 | 491.2288 | | |
| 30 | 3010.4337 | | | 5 | 420.1917 | | |
| 31 | 3081.4709 | | | 4 | 349.1546 | | |
| 32 | 3138.4923 | | | 3 | 292.1332 | | |
| 33 | 3195.5138 | | | 2 | 235.1117 | | |
| 34 | 3298.5230 | | | 1 | 132.1025 | 132.1029 | 3.41 |

## 3.5 Linearisation of Tikitericin

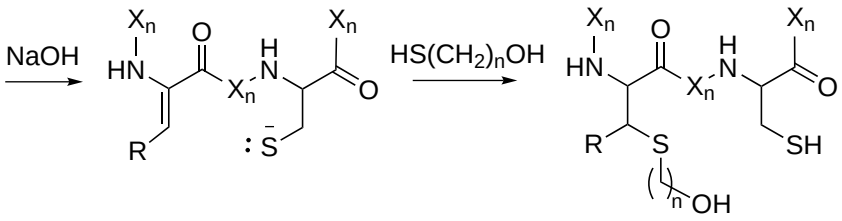
Required for primary sequence determination, most 'normal' disulfide bridged peptides and proteins can be linearised by a simple process of sulfide reduction and alkylation. The high stability of lantibiotic thioether bridges, however, requires stronger methods of linearisation. Several methods have been reported in the literature but none has yet reached the general applicability of disulfide bond reduction.[109,112–117] The methods attempted in this research are shown in Scheme 3.5.




**(a)** (i) Oxidation (ii) Thermal elimination




**(b)** (i) Raney Nickel reduction




**(c)** (i) Base-induced elimination (ii) Addition

**Scheme 3.5.** Lanthipeptide linearisation methods.

### 3.5.1 Oxidative thermal elimination

Tikitericin linearisation was attempted by the oxidation of (Me)Lan thioether bonds with sodium periodate, followed by heating under reflux to induce thermal elimination (Scheme 3.5a). This method has been used successfully in the elimination of volatile sulfenic acids,[118] but has not yet been applied to lanthipeptide linearisation. The reaction was trialled with nisin but was not pursued due to the difficulty in controlling stoichiometric oxidation of four (Me)Lan thioether groups to sulfoxide functional groups on the microgram scale. A range of nisin : sodium periodate ratios and reaction times were explored and analysed by MALDI-TOF MS. Under all reaction conditions, nisin was oxidised to a multitude of oxidised products, many of which contained sulfone groups from over oxidation (Appendix D.9). Refluxing oxidised nisin did not result in linearisation.

### 3.5.2 Raney Nickel-catalysed reduction

In 2011 Fuchs *et al.* reported a new method of linearisation via Raney Nickel (RaNi)-catalysed thioether bond reduction (Scheme 3.5b).[109] This method was successfully trialled in the current study, MALDI-TOF MS analysis showing reductive linearisation of nisin. The reaction was repeated with LH-20 purified tikitericin. MALDI-TOF MS analysis showed reductive linearisation with partial hydrolysis to several truncated peptides (Figure 3.19).

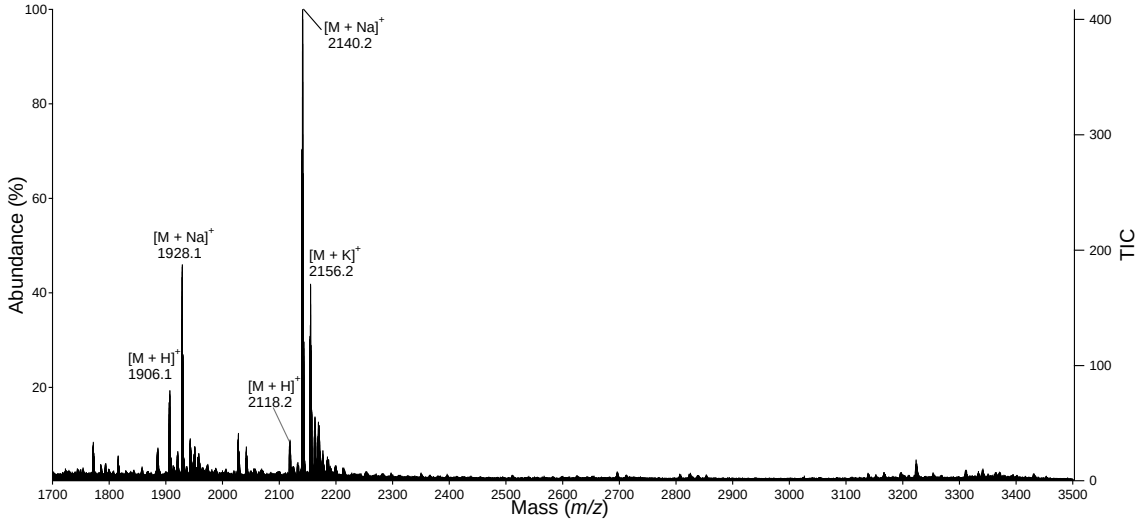



**Figure 3.19.** MALDI-TOF MS spectrum of Raney Nickel linearisation products.

HR ESI-MS data was obtained for the following linear peptides:

**RaNi linearised tikitericin Asn8-Leu35:** M = 2600.5014 Da, *m/z* 867.8403 $[M + 3H]^{3+}$ (calculated, 867.8411 Da; $\Delta$ – 0.84 ppm), *m/z* 1301.2618 $[M + 2H]^{2+}$ (calculated, 1301.2580 Da; $\Delta$ 2.94 ppm).

**RaNi linearised tikitericin Asn9-Leu35:** M = 2486.4584 Da, *m/z* 829.8262 [M + 3H]$^{3+}$ (calculated, 829.8268 Da; Δ – 0.65 ppm), *m/z* 1244.2360 [M + 2H]$^{2+}$ (calculated, 1244.2365 Da; Δ – 0.41 ppm), *m/z* 1266.2098 [M + 2Na]$^{2+}$ (calculated, 1266.2184 Da; Δ – 6.86 ppm).

**RaNi linearised tikitericin Abu10-Leu35:** M = 2372.4155 Da, *m/z* 1187.7077 [M + 2H]$^{2+}$ (calculated, 1187.7165 Da; Δ – 7.36 ppm), *m/z* 1209.6992 [M + 2Na]$^{2+}$ (calculated, 1209.6984 Da; Δ 0.64 ppm).

**RaNi linearised tikitericin Ala12-Leu35:** M = 2188.2807 Da, *m/z* 1095.6449 [M + 2H]$^{2+}$ (calculated, 1095.6559 Da; Δ – 10.06 ppm).

**RaNi linearised tikitericin Val13-Leu35 (33):** M = 2117.2575 Da, *m/z* 706.7598 [M + 3H]$^{3+}$ (calculated, 706.7597 Da; Δ 0.14 ppm), *m/z* 1059.6369 [M + 2H]$^{2+}$ (calculated, 1059.6359 Da; Δ 1.00 ppm), *m/z* 1081.6209 [M + 2Na$^{2+}$ (calculated, 1081.6178 Da; Δ 2.84 ppm), *m/z* 2118.2645 [M + H]$^{+}$ (calculated, 2118.2645 Da; Δ 0.01 ppm), *m/z* 2140.2574 [M + Na]$^{+}$ (calculated, 2140.2464 Da; Δ 5.11 ppm).

**RaNi linearised tikitericin Leu14-Leu35:** M = 2018.1888 Da, *m/z* 673.7369 [M + 3H]$^{3+}$ (calculated, 673.7369 Da; Δ 0.04 ppm), *m/z* 1010.1017 [M + 2H]$^{2+}$ (calculated, 1010.1017 Da; Δ – 0.01 ppm).

**RaNi linearised tikitericin Ala15-Leu35 (34):** M = 1905.1047 Da, *m/z* 953.5612 [M + 2H]$^{2+}$ (calculated, 953.5596 Da; Δ 1.64 ppm), *m/z* 975.5445 [M + 2Na]$^{2+}$ (calculated, 975.5416 Da; Δ 3.01 ppm), 1906.1103 [M + H]$^{+}$ (calculated, 1906.1120 Da; Δ – 0.9 ppm), *m/z* 1928.0924 [M + H]$^{+}$ (calculated, 1928.0940 Da; Δ – 0.8 ppm).

MALDI-TOF MSMS data was obtained for the protonated and sodiated precursor ions of **33** (Table 3.5, Figure 3.20) and **34** (Table D.1 and Figure D.10, see Appendix D.4), and HR ESI-MSMS data was obtained for the protonated [M + 2H]$^{2+}$ precursor ion of **34** (Table 3.6, Figure 3.21). Note that for the tandem MSMS of **33** and **34**, the b1 ion corresponds to cleavage of the Val13 – Leu14 and Ala15 – Leu16 peptide bonds, respectively. Data sets from both instruments produce a contiguous series of backbone cleavage sequence ions and a remarkable improvement in spectrum quality (S/N, ion count) following hydrolysis of the N-terminal histidine. In the case of the ESI-MS, sequence information was obtained with 35% of the collision energy required to induce fragmentation of naturally-occurring tikitericin. Despite exploration of reaction parameters such as RaNi and buffer concentrations, reaction time and temperature, the fully linearised peptide was not detected. This degradation prevented sequence data from being obtained for three crucial peptide bonds between residues Abu4 and Ala7.

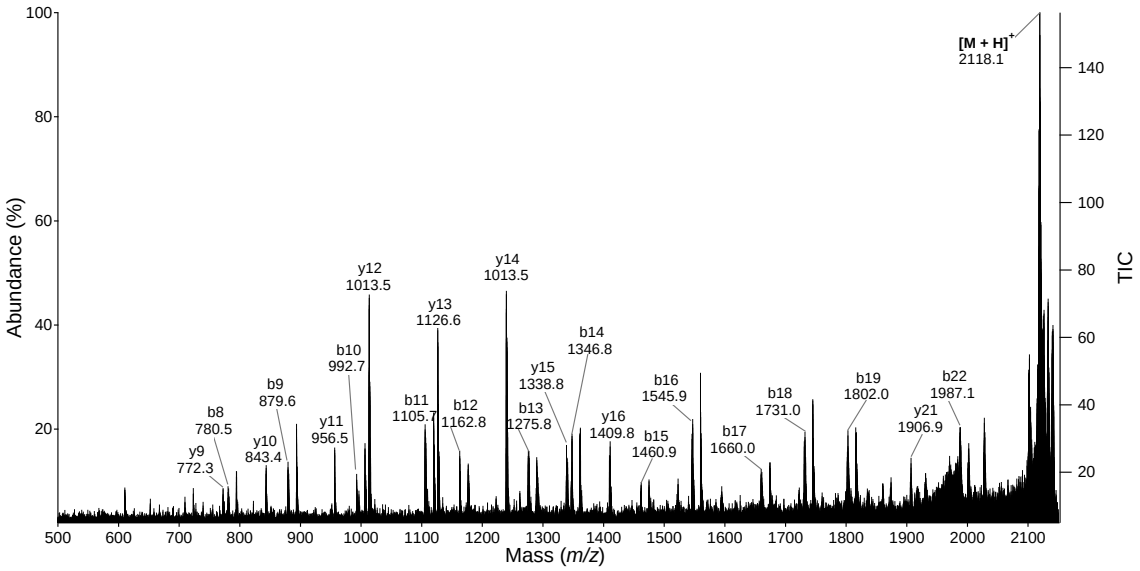
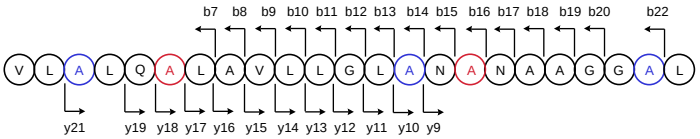



**Figure 3.20.** MALDI-TOF MSMS spectrum of Raney Nickel linearisation product Val13-Leu35 (**33**) (M = 2117.2575 Da), precursor ion: *m/z* 2140 $[M + Na]^+$.

**Table 3.5.** MALDI-TOF MSMS fragmentation pattern of Raney Nickel linearisation product Val13-Leu35 (**33**) (M = 2117.2575 Da), precursor ions: *m/z* 2118 $[M + H]^+$, *m/z* 2140 $[M + Na]^+$.




**33**

| b ions | | | y ions | | |
|---|---|---|---|---|---|
| Pos | Exact (*m/z*) | Observed (*m/z*) | Pos | Exact (*m/z*) | Observed (*m/z*) |
| 1 | 100.08 | | 22 | 2019.20 | |
| 2 | 213.16 | | 21 | 1906.11 | 1905.99 |
| 3 | 284.20 | | 20 | 1835.08 | |
| 4 | 397.28 | | 19 | 1721.99 | 1721.92 |
| 5 | 525.34 | | 18 | 1593.93 | 1593.84 |
| 6 | 596.38 | | 17 | 1522.90 | 1522.88 |
| 7 | 709.46 | 709.49 | 16 | 1409.81 | 1409.82 |
| 8 | 780.50 | 780.49 | 15 | 1338.77 | 1338.77 |
| 9 | 879.57 | 879.56 | 14 | 1239.71 | 1239.69 |
| 10 | 992.65 | 992.65 | 13 | 1126.62 | 1126.62 |
| 11 | 1105.73 | 1105.73 | 12 | 1013.54 | 1013.55 |
| 12 | 1162.76 | 1162.77 | 11 | 956.52 | 956.52 |
| 13 | 1275.84 | 1275.85 | 10 | 843.43 | 843.43 |
| 14 | 1346.88 | 1346.89 | 9 | 772.40 | 772.40 |
| 15 | 1460.92 | 1460.91 | 8 | 658.35 | |
| 16 | 1545.97 | 1545.97 | 7 | 573.30 | |
| 17 | 1660.02 | 1659.92 | 6 | 459.26 | |
| 18 | 1731.05 | 1731.06 | 5 | 388.22 | |
| 19 | 1802.09 | 1802.10 | 4 | 317.18 | |
| 20 | 1859.11 | 1859.01 | 3 | 260.16 | |
| 21 | 1916.13 | | 2 | 203.14 | |
| 22 | 1987.17 | 1987.18 | 1 | 132.10 | |

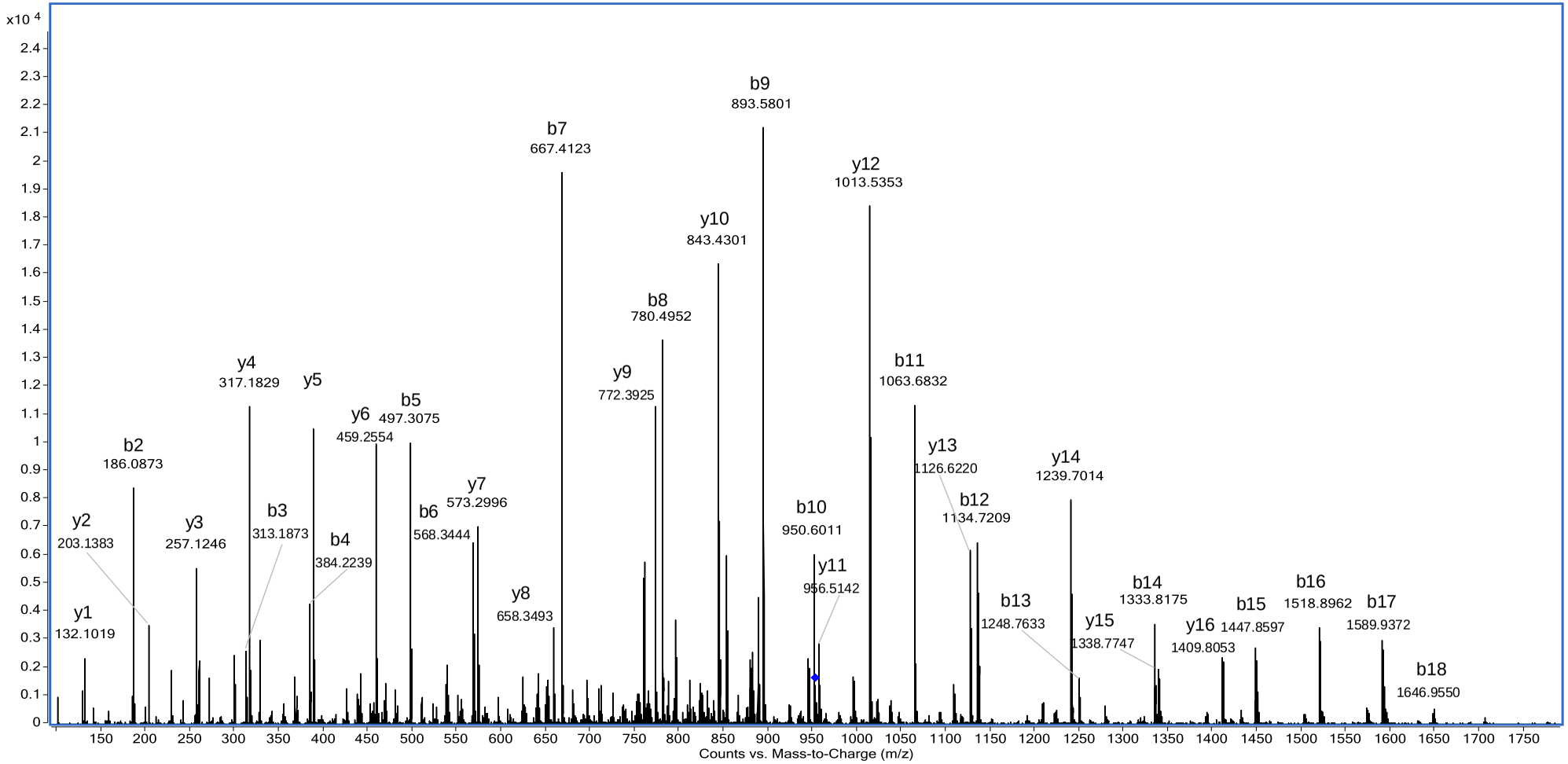
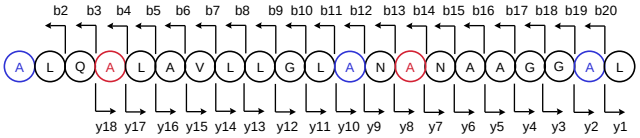



**Figure 3.21.** HR ESI-MSMS spectrum of Raney Nickel linearisation product Ala15-Leu35 (**34**) (M = 1905.1077 Da), precursor ion: *m/z* 954.06 $[M + 2H]^{2+}$, CID energy = 25.

**Table 3.6.** HR ESI-MSMS fragmentation pattern of Raney Nickel linearisation product Ala15-Leu35 (**34**) (M = 1905.1077 Da), precursor ion: *m/z* 954.06 $[M + 2H]^{2+}$.




**34**

| b ions | | | | y ions | | | |
|---|---|---|---|---|---|---|---|
| Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) | Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) |
| 1 | 72.0449 | | | 20 | 1835.0755 | | |
| 2 | 185.1290 | 185.1284 | – 3.24 | 19 | 1721.9914 | | |
| 3 | 313.1876 | 313.1873 | – 0.96 | 18 | 1593.9329 | 1593.9300 | – 1.79 |
| 4 | 384.2247 | 384.2239 | – 2.08 | 17 | 1522.8958 | 1522.8800 | – 10.34 |
| 5 | 497.3088 | 497.3069 | – 3.82 | 16 | 1409.8117 | 1409.8053 | – 4.50 |
| 6 | 568.3459 | 568.3444 | – 2.64 | 15 | 1338.7746 | 1338.7747 | 0.11 |
| 7 | 667.4143 | 667.4138 | – 0.75 | 14 | 1239.7062 | 1239.7030 | – 2.54 |
| 8 | 780.4943 | 780.4959 | 2.05 | 13 | 1126.6221 | 1126.6220 | – 0.04 |
| 9 | 893.5824 | 893.5798 | – 2.91 | 12 | 1013.5380 | 1013.5365 | – 1.43 |
| 10 | 950.6039 | 950.6014 | – 2.63 | 11 | 956.5166 | 956.5142 | – 2.46 |
| 11 | 1063.6879 | 1063.6863 | – 1.50 | 10 | 843.4325 | 843.4320 | – 0.53 |
| 12 | 1134.7251 | 1134.7228 | – 2.03 | 9 | 772.3954 | 772.3925 | – 3.69 |
| 13 | 1248.7680 | 1248.7633 | – 3.76 | 8 | 658.3525 | 658.3493 | – 4.78 |
| 14 | 1333.8207 | 1333.8181 | – 1.95 | 7 | 573.2997 | 573.2993 | – 0.61 |
| 15 | 1447.8637 | 1447.8597 | – 2.76 | 6 | 459.2608 | 459.2554 | – 11.65 |
| 16 | 1518.9008 | 1518.8972 | – 2.37 | 5 | 388.2197 | 388.2183 | – 3.48 |
| 17 | 1589.9379 | 1589.9372 | – 0.44 | 4 | 317.1825 | 317.1815 | – 3.00 |
| 18 | 1646.9594 | 1646.9550 | – 2.67 | 3 | 260.1611 | 260.1609 | – 0.58 |
| 19 | 1703.9808 | 1703.9800 | – 0.47 | 2 | 203.1396 | 203.1383 | – 6.15 |
| b0 | 1775.0179 | 1774.9800 | – 21.35 | 1 | 132.1025 | 132.1012 | – 9.46 |

### 3.5.3 Base-induced elimination/ thiol addition

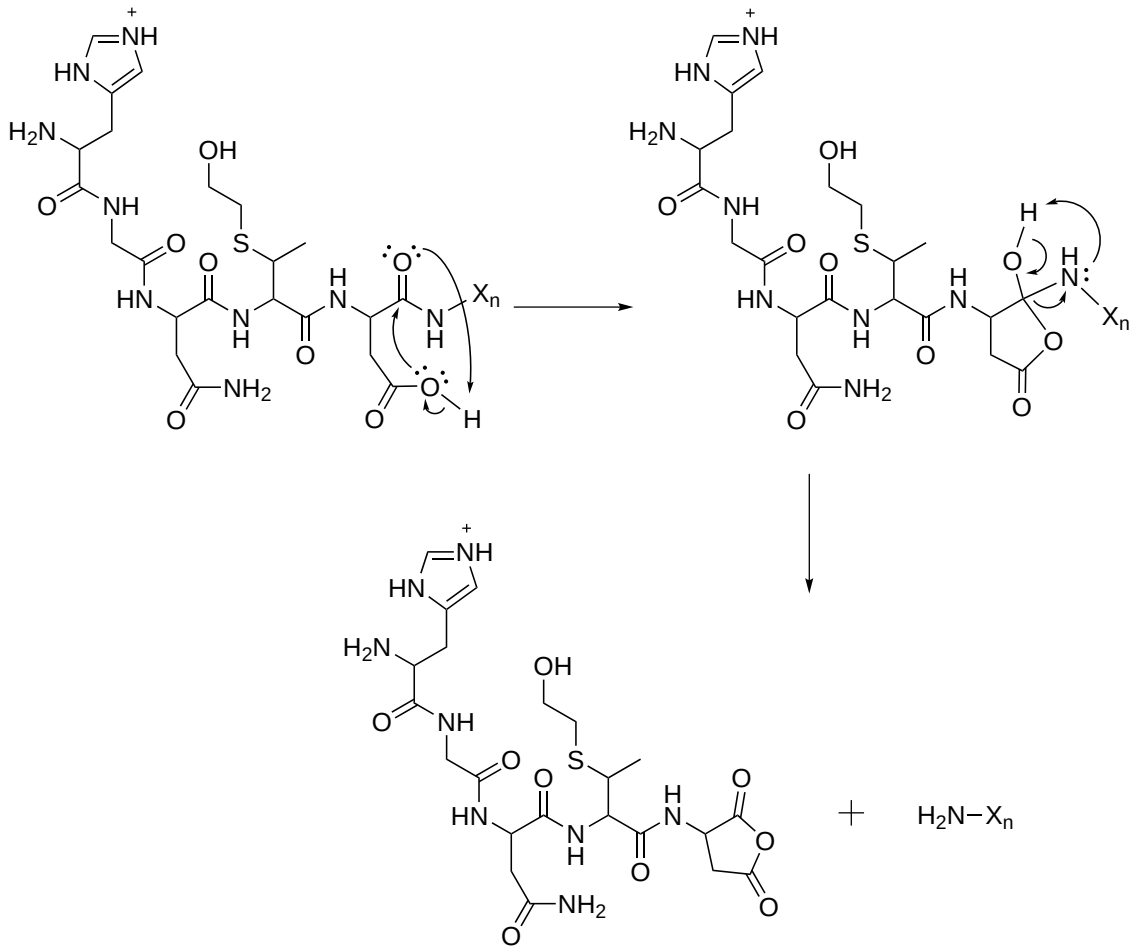
A base-induced elimination strategy (Scheme 3.5c),[113] followed by addition of 2-mercaptoethanol was successful in generating the fully linearised peptide. Analysis of the reaction products by MALDI-TOF MS revealed four major products at *m/z* 3529.6, 3607.6, 3685.7, and 3763.7, corresponding to the addition of one, two, three and four molecules of 2-mercaptoethanol, all detected as $[M + Na]^+$ quasimolecular ions. The peptides were detected as multiply charged ions by HR ESI-MS (Figure 3.22).

**Fully linearised tikitericin (35):** M = 3740.6651 Da, *m/z* 1247.8920 $[M + 3H]^{3+}$ (calculated, 1247.8958 Da; Δ 3.06 ppm), *m/z* 1871.3369 $[M + 2H]^{2+}$ (calculated, 1871.3400 Da; Δ – 1.68 ppm), *m/z* 1909.3432 $[M + 2Na]^{2+}$ (calculated, 1909.2959 Da; Δ 8.04 ppm).

**Triple addition tikitericin (36):** M = 3662.6516 Da, *m/z* 1221.8911 $[M + 3H]^{3+}$ (calculated, 1221.8911 Da; Δ – 0.07 ppm), *m/z* 1832.3317 $[M + 2H]^{2+}$ (calculated, 1832.3331 Da; Δ – 0.74 ppm).

**Double addition tikitericin (37):** M = 3584.6376 Da, *m/z* 1195.8866 $[M + 3H]^{3+}$ (calculated, 1195.8865 Da; $\Delta$ – 0.13 ppm), *m/z* 1793.3268 $[M + 2H]^{2+}$ (calculated, 1793.3261 Da; $\Delta$ 0.38 ppm), *m/z* 1815.3174 $[M + 2Na]^{2+}$ (calculated, 1815.3080 Da; $\Delta$ 5.14 ppm).

**Single addition tikitericin:** M = 3506.6237 Da, *m/z* 1169.8831 $[M + 3H]^{3+}$ (calculated, 1169.8818 Da; $\Delta$ 1.08 ppm), *m/z* 1192.2031 $[M + 3Na]^{3+}$ (calculated, 1192.1980 Da; $\Delta$ 4.23 ppm), *m/z* 1754.3199 $[M + 2H]^{2+}$ (calculated, 1754.3134 Da; $\Delta$ 0.46 ppm), *m/z* 1776.3016 $[M + 2Na]^{2+}$ (calculated, 1776.3011 Da; $\Delta$ 0.31 ppm).




**Scheme 3.6.** Mechanism of charge remote fragmentation: the Aspartic and Glutamic acid effect.[101]

High-resolution LC-ESI MSMS data was obtained for the fully linearised peptide (Table 3.7), as well as for the doubly and triply linearised products (Tables 3.8 and 3.9). As expected, the MSMS spectra were dominated by the low energy sequencing ions. It was immediately apparent that crucial sequencing ions b4, b5 and b6, not previously observed in the MSMS spectrum of tikitericin, were abundant in the product ion spectra of all three compounds (Figures 3.23, 3.25 and 3.24). Suppression caused by the N-terminal histidine required higher energy for proton mobilisation than was, in this case, required to initiate 'charge remote' fragmentation pathways.† The b5 (*m/z* 585.21) and b6 (*m/z* 714.25) ions

†Pathways with no proton involvement.

are products of charge remote fragmentation due to the Aspartic and Glutamic acid effect, the mechanism of which is depicted in Scheme 3.6 for the b5 ion.[101] In this mechanism, the N-terminal histidine actually enhances the fragmentation pathway by tightly binding the proton and allowing the acidic hydrogen of the aspartic or glutamic acid side chain to initiate cleavage.[101] These ions were not observed in the tandem MSMS of tikitericin prior to linearisation of ring A and, fortuitously, the fragmentation enhanced sequence information from two of the crucial peptide bonds.

Characterisation of the partially-linearised products (**36**) and (**37**) was complicated by sequence ions common to multiple isomers. Furthermore, fragment ions of $m/z \geq 2000$ were not detected with high accuracy or abundance and the sequencing data is therefore presented as follows. For residues His1 – Gln17, bn ions are tabulated (b1 – b17) for three isomers: linearisation of rings A and B, A only, and B only. For residues Ala18/Abu18 – Leu35 (y1 – y18), yn ions are tabulated for three isomers, linearisation of rings C and D, D only and C only. Ions arising from internal fragmentation are tabulated but are not depicted diagrammatically unless the *m/z* was diagnostic (e.g. b4, b5 and b6). In the case of triple linearisation (**36**), four possible isomers may form, with linearisation of rings: (i) A, B and C, (ii) A, B and D, (iii) A, C and D, or (iv) B, C and D. Based on the most abundant sequencing ions, all four isomers may be present, although the definite presence or absence of any isomer cannot be ascertained. For example, the abundance of ions b10 – b14 showed the addition of two 2-mercaptoethanol moieties to rings A and B but fragment ions $\geq 2000$ were not detected and the precursor could therefore not be assigned as (i), (ii), or a combination of both isomers.

In the case of double linearisation (**37**), six structural isomers may form, with linearisation of rings: (i) A and B, (ii) A and C, (iii) A and D, (iv) B and C, (v) B and D, or (vi) C and D. Non-sequencing ions were observed in the MSMS spectra of **37**, including cyclic z-type ions, z8 and z9, which form from the decomposition of y-type ions through the loss of $NH_3$ and are likely to be stabilised by conjugative effects.[110]

Although interesting, it was not necessary to fully characterise these peptides because each of the sequencing ions, regardless of which isomer was the precursor ion, supports the primary amino acid sequence of tikitericin. The sum total of all the preceding data confirmed the full linear sequence of amino acids as predicted by the initial genomic data.

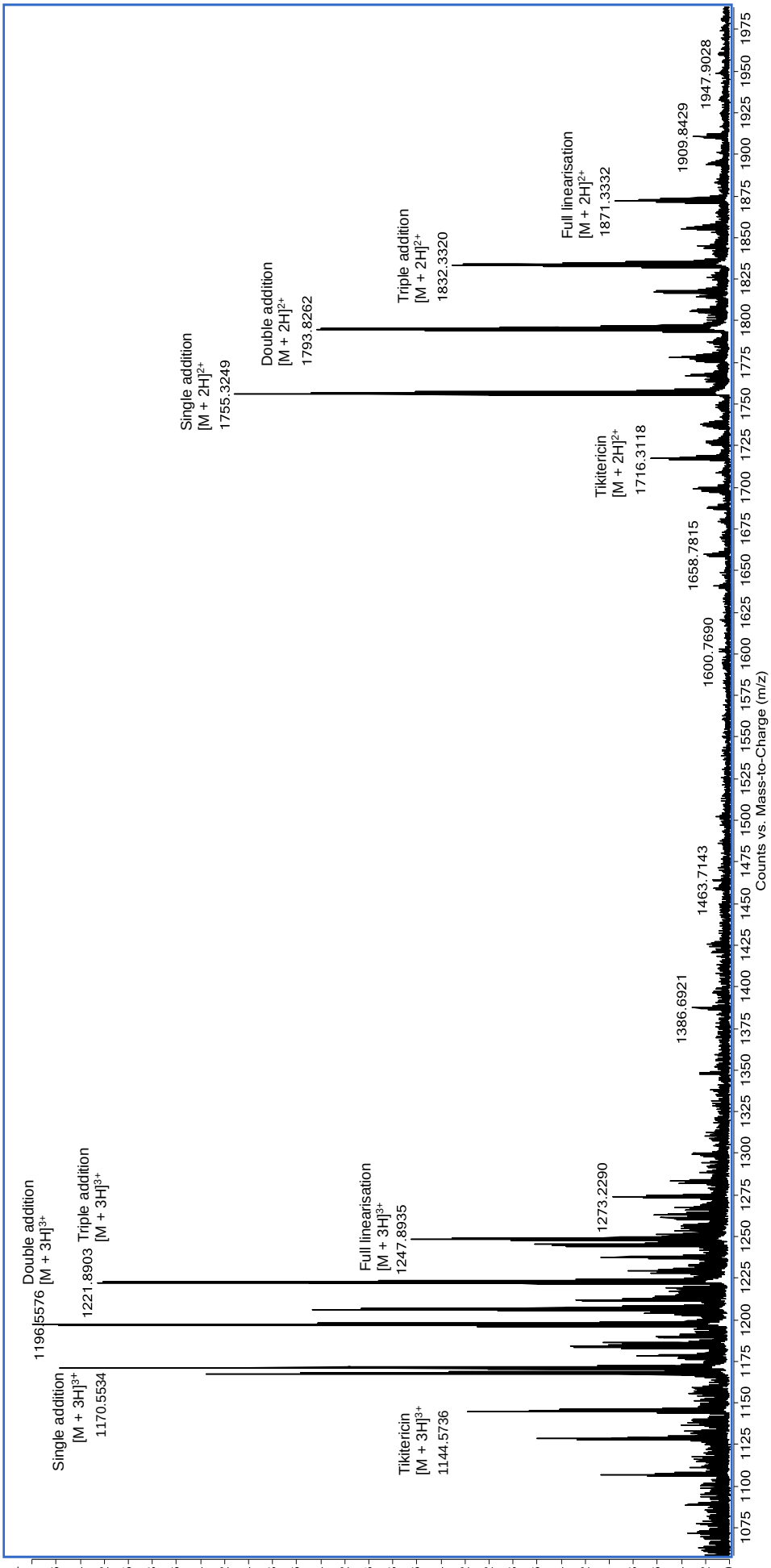



**Figure 3.22.** HR ESI-MS spectrum of base-induced elimination/thiol addition linearisation products.

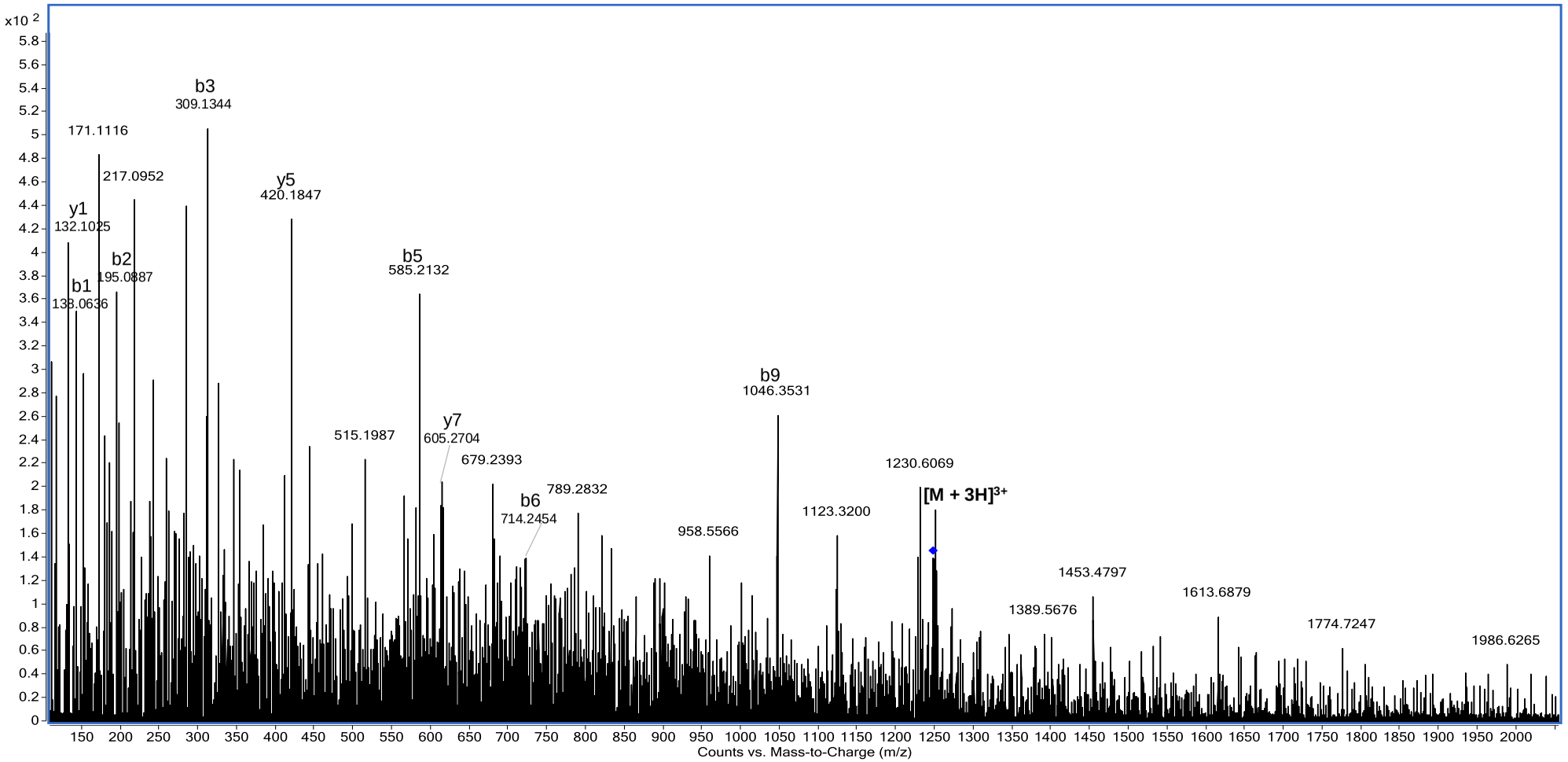
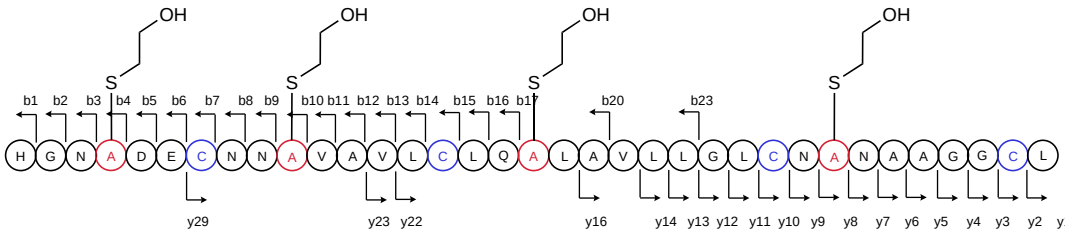



**Figure 3.23.** HR ESI-MSMS spectrum of fully linearised tikitericin (**35**) (M = 3740.6651 Da): precursor ion *m/z* 1247.89 $[M + 3H]^{3+}$, CID energy = 60.

**Table 3.7.** HR ESI-MSMS fragmentation pattern of fully-linearised tikitericin (**35**) (M = 3740.6641 Da): precursor ions *m/z* 1247.89 $[M + 3H]^{3+}$, *m/z* 1871.34 $[M + 2H]^{2+}$.




**35**

| b ions | | | | y ions | | | |
|---|---|---|---|---|---|---|---|
| Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) | Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) |
| 1 | 138.0667 | 138.0636 | – 22.45 | 34 | 3604.6145 | | |
| 2 | 195.0882 | 195.0887 | 2.56 | 33 | 3547.5930 | | |
| 3 | 309.1311 | 309.1344 | 10.68 | 32 | 3433.5501 | | |
| 4 | 470.1822 | 470.1757 | – 13.82 | 31 | 3272.4990 | | |
| 5 | 585.2091 | 585.2085 | – 1.03 | 30 | 3157.4721 | | |
| 6 | 714.2517 | 714.2454 | – 8.82 | 29 | 3028.4295 | 3028.3215 | – 35.65 |
| 7 | 817.2609 | 817.2633 | 2.94 | 28 | 2925.4203 | | |
| 8 | 931.3038 | 931.2961 | – 8.27 | 27 | 2811.3774 | | |
| 9 | 1045.3468 | 1045.3462 | – 0.57 | 26 | 2697.3345 | | |
| 10 | 1206.3978 | 1206.3864 | – 9.45 | 25 | 2536.2834 | | |
| 11 | 1305.4662 | 1305.4589 | – 5.59 | 24 | 2437.2150 | | |
| 12 | 1376.5033 | 1376.5100 | 4.87 | 23 | 2366.1779 | 2366.1098 | – 28.76 |
| 13 | 1475.5717 | 1475.5808 | 6.17 | 22 | 2267.1095 | 2267.1243 | 6.55 |
| 14 | 1588.6558 | 1588.6620 | 3.90 | 21 | 2154.0254 | | |
| 15 | 1691.6650 | 1691.6530 | – 7.09 | 20 | 2051.0162 | | |
| 16 | 1804.7491 | 1804.7396 | – 5.26 | 19 | 1937.9322 | | |
| 17 | 1932.8076 | 1932.8320 | 12.62 | 18 | 1809.8736 | | |
| 18 | 2079.8430 | | | 17 | 1662.8382 | | |
| 19 | 2192.9271 | | | 16 | 1549.7541 | 1549.7666 | 8.10 |
| 20 | 2263.9642 | 2263.9020 | – 27.47 | 15 | 1478.7170 | | |
| 21 | 2363.0326 | | | 14 | 1379.6486 | 1379.6282 | – 14.75 |
| 22 | 2476.1167 | | | 13 | 1266.5645 | 1266.6136 | 38.81 |
| 23 | 2589.2008 | 2589.0815 | – 46.08 | 12 | 1153.4805 | 1153.4870 | 5.68 |
| 24 | 2646.2222 | | | 11 | 1096.4590 | 1096.4632 | 3.88 |
| 25 | 2759.3063 | | | 10 | 983.3749 | 983.3423 | – 33.10 |
| 26 | 2862.3155 | | | 9 | 880.3657 | 880.3662 | 0.62 |
| 27 | 2976.3584 | | | 8 | 766.3228 | 766.2818 | – 53.44 |
| 28 | 3137.4094 | | | 7 | 605.2718 | 605.2704 | – 2.23 |
| 29 | 3251.4524 | | | 6 | 491.2288 | 491.2300 | 2.54 |
| 30 | 3322.4895 | | | 5 | 420.1917 | 420.1847 | – 16.54 |
| 31 | 3393.5266 | | | 4 | 349.1546 | 349.1515 | – 8.74 |
| 32 | 3430.5481 | | | 3 | 292.1332 | 292.1311 | – 7.02 |
| 33 | 3507.5695 | | | 2 | 235.1117 | 235.1147 | 12.97 |
| 34 | 3610.5787 | | | 1 | 132.1025 | 132.1025 | 0.38 |

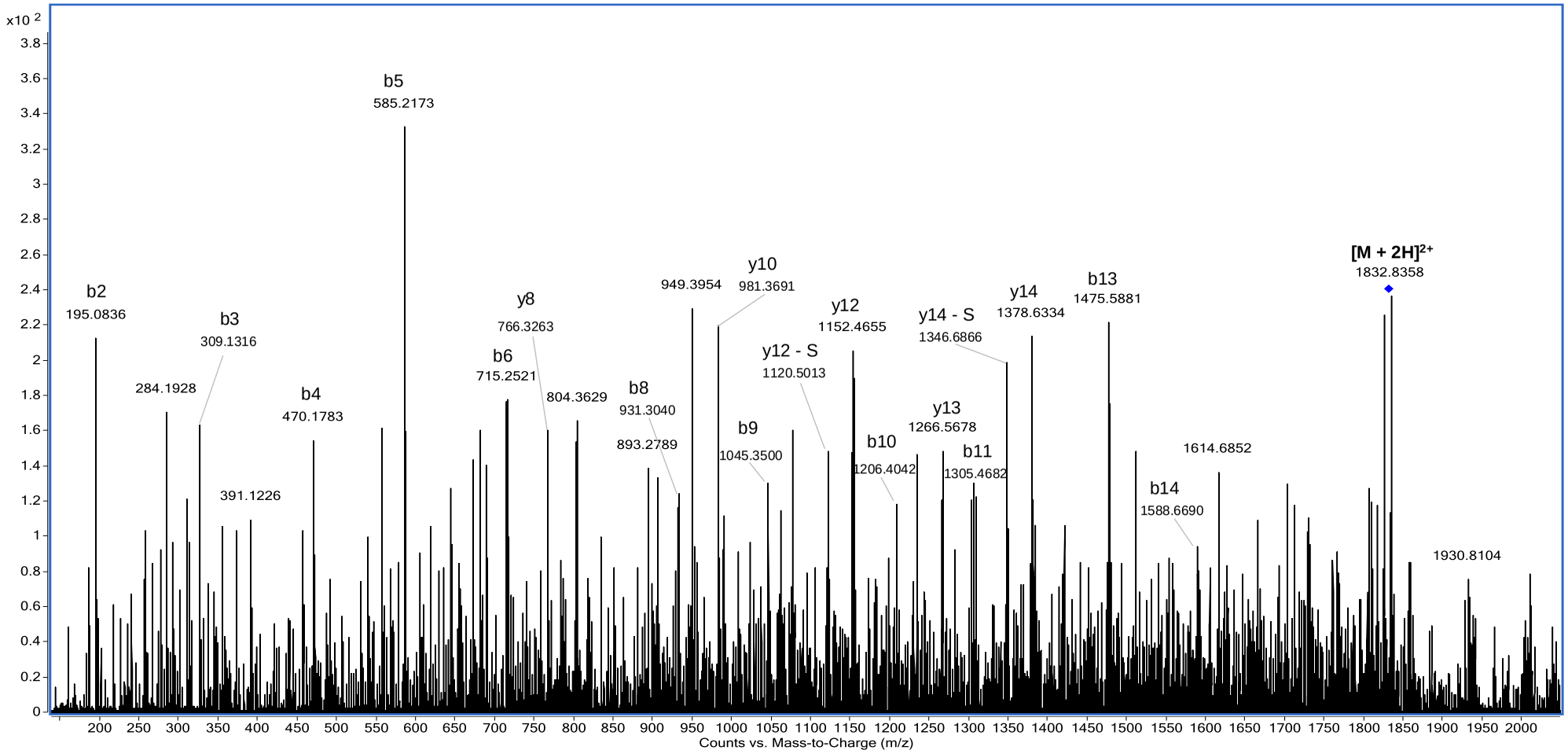
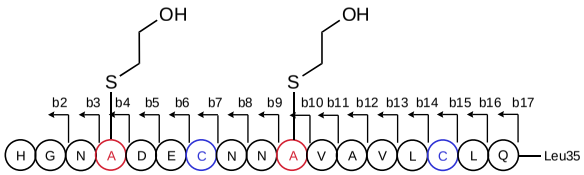


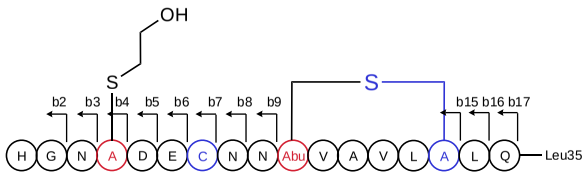

**Figure 3.24.** HR ESI-MSMS spectrum of triple addition tikitericin (**36**) (M = 3662.6516 Da): precursor ion *m/z* 1832.33 $[M + 2H]^{2+}$, CID energy = 70.

**Table 3.8.** HR ESI-MSMS fragmentation pattern of triple addition tikitericin (**36**) (M = 3662.6516 Da): precursor ions *m/z* 1221.89 $[M + 3H]^{3+}$, *m/z* 1832.33 $[M + 2H]^{2+}$.



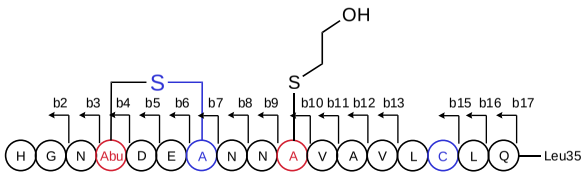

b ions

| Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) |
|---|---|---|---|
| 1 | 138.0667 | | |
| 2 | 195.0882 | 195.0890 | 4.10 |
| 3 | 309.1311 | 309.1316 | 1.62 |
| 4 | 470.1822 | 470.1740 | – 17.44 |
| 5 | 585.2091 | 585.2097 | 1.03 |
| 6 | 714.2517 | 714.2531 | 1.96 |
| 7 | 817.2609 | 817.2601 | – 0.98 |
| 8 | 931.3038 | 931.3040 | 0.21 |
| 9 | 1045.3468 | 1045.3500 | 3.06 |
| 10 | 1206.3978 | 1206.4042 | 5.31 |
| 11 | 1305.4662 | 1305.4682 | 1.53 |
| 12 | 1376.5033 | 1376.5052 | 1.38 |
| 13 | 1475.5717 | 1475.5806 | 6.03 |
| 14 | 1588.6558 | 1588.6690 | 8.31 |
| 15 | 1691.6650 | 1691.6655 | 0.30 |
| 16 | 1804.7491 | 1804.7504 | 0.72 |
| 17 | 1932.8076 | 1932.8205 | 6.67 |




b ions

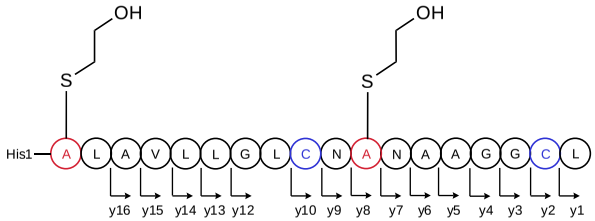
| Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) |
|---|---|---|---|
| 1 | 138.0667 | | |
| 2 | 195.0882 | 195.0890 | 4.10 |
| 3 | 309.1311 | 309.1316 | 1.62 |
| 4 | 470.1822 | 470.1740 | – 17.44 |
| 5 | 585.2091 | 585.2097 | 1.03 |
| 6 | 714.2517 | 714.2531 | 1.96 |
| 7 | 817.2609 | 817.2601 | – 0.98 |
| 8 | 931.3038 | 931.3040 | 0.21 |
| 9 | 1045.3468 | 1045.3500 | 3.06 |
| 10 | 1128.3839 | 1128.3724 | – 10.19 |
| 11 | 1227.4523 | 1227.4557 | 2.77 |
| 12 | 1298.4894 | 1298.4791 | – 7.93 |
| 13 | 1397.5578 | 1397.5548 | – 2.15 |
| 14 | 1510.6419 | | |
| 15 | 1613.6511 | 1613.6474 | – 2.29 |
| 16 | 1726.7351 | 1726.7408 | 3.30 |
| 17 | 1854.7937 | 1854.8182 | 13.21 |



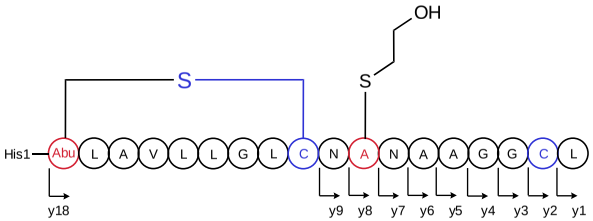

b ions

| Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) |
|---|---|---|---|
| 1 | 138.0667 | | |
| 2 | 195.0882 | 195.0890 | 4.10 |
| 3 | 309.1311 | 309.1316 | 1.62 |
| 4 | 392.1682 | 392.1681 | – 0.25 |
| 5 | 507.1952 | 507.1996 | 8.68 |
| 6 | 636.2378 | 636.2355 | – 3.62 |
| 7 | 739.2470 | 739.2562 | 12.45 |
| 8 | 853.2899 | 853.2885 | – 1.64 |
| 9 | 967.3328 | 967.3488 | 16.54 |
| 10 | 1128.3839 | 1128.3724 | – 10.19 |
| 11 | 1227.4523 | 1227.4557 | 2.77 |
| 12 | 1298.4894 | 1298.4791 | – 7.93 |
| 13 | 1397.5578 | 1397.5548 | – 2.15 |
| 14 | 1510.6419 | | |
| 15 | 1613.6511 | 1613.6474 | – 2.29 |
| 16 | 1726.7351 | 1726.7408 | 3.30 |
| 17 | 1854.7937 | 1854.8182 | 13.21 |

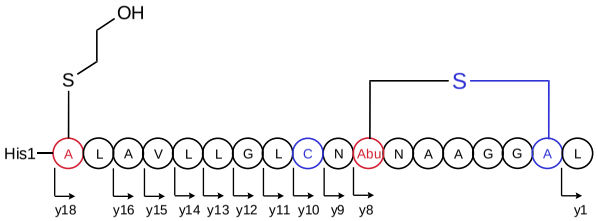
**Table 3.8.** HR ESI-MSMS fragmentation pattern of triple addition tikitericin (**36**) (M = 3662.6516 Da): precursor ions *m/z* 1221.89 $[M + 3H]^{3+}$, *m/z* 1832.33 $[M + 2H]^{2+}$. (cont.)




| y ions | | | |
|---|---|---|---|
| Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) |
| 18 | 1809.8736 | 1809.8380 | – 19.67 |
| 17 | 1662.8382 | | |
| 16 | 1549.7541 | 1549.7669 | 8.26 |
| 15 | 1478.7170 | 1478.7499 | 22.25 |
| 14 | 1379.3786 | 1379.6342 | 185.30 |
| 13 | 1266.5645 | 1266.5660 | 1.18 |
| 12 | 1153.4805 | 1153.4805 | 0.00 |
| 11 | 1096.4590 | | |
| 10 | 983.3749 | 983.3727 | – 2.24 |
| 9 | 880.3657 | 880.3736 | 8.97 |
| 8 | 766.3229 | 766.3263 | 4.44 |
| 7 | 605.2719 | 605.2760 | 6.77 |
| 6 | 491.2289 | 491.2294 | 1.02 |
| 5 | 420.1918 | 420.1956 | 9.04 |
| 4 | 349.1547 | 349.1541 | – 1.72 |
| 3 | 292.1333 | 292.1320 | – 4.45 |
| 2 | 235.1118 | 235.1105 | – 5.53 |
| 1 | 132.1025 | 132.1007 | – 13.25 |




| y ions | | | |
|---|---|---|---|
| Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) |
| 18 | 1731.8596 | 1731.8591 | – 0.26 |
| 17 | 1662.8382 | | |
| 16 | 1549.7541 | 1549.7669 | 8.26 |
| 15 | 1478.7170 | 1478.7499 | 22.25 |
| 14 | 1379.3786 | 1379.6342 | 185.30 |
| 13 | 1266.5645 | 1266.5660 | 1.18 |
| 12 | 1153.4805 | 1153.4805 | 0.00 |
| 11 | 1096.4590 | | |
| 10 | 983.3749 | 983.3727 | – 2.24 |
| 9 | 880.3657 | 880.3736 | 8.97 |
| 8 | 766.3229 | 766.3263 | 4.44 |
| 7 | 605.2719 | 605.2760 | 6.77 |
| 6 | 491.2289 | 491.2294 | 1.02 |
| 5 | 420.1918 | 420.1956 | 9.04 |
| 4 | 349.1547 | 349.1541 | – 1.72 |
| 3 | 292.1333 | 292.1320 | – 4.45 |
| 2 | 235.1118 | 235.1105 | – 5.53 |
| 1 | 132.1025 | 132.1007 | – 13.25 |




| y ions | | | |
|---|---|---|---|
| Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) |
| 18 | 1731.8596 | 1731.8591 | – 0.26 |
| 17 | 1584.8242 | | |
| 16 | 1471.7402 | 1471.7737 | 22.80 |
| 15 | 1400.7031 | 1400.6729 | – 21.52 |
| 14 | 1301.6347 | 1301.6393 | 3.57 |
| 13 | 1188.5506 | 1188.5584 | 6.60 |
| 12 | 1075.4665 | 1075.4699 | 3.21 |
| 11 | 1018.4451 | 1018.4366 | – 8.30 |
| 10 | 905.3610 | 905.3566 | – 4.80 |
| 9 | 802.3518 | 802.3531 | 1.68 |
| 8 | 688.3090 | 688.3169 | 11.48 |
| 7 | 605.2719 | 605.2760 | 6.77 |
| 6 | 491.2289 | 491.2294 | 1.02 |
| 5 | 420.1918 | 420.1956 | 9.04 |
| 4 | 349.1547 | 349.1541 | – 1.72 |
| 3 | 292.1333 | 292.1320 | – 4.45 |
| 2 | 235.1118 | 235.1105 | – 5.53 |
| 1 | 132.1025 | 132.1007 | – 13.25 |

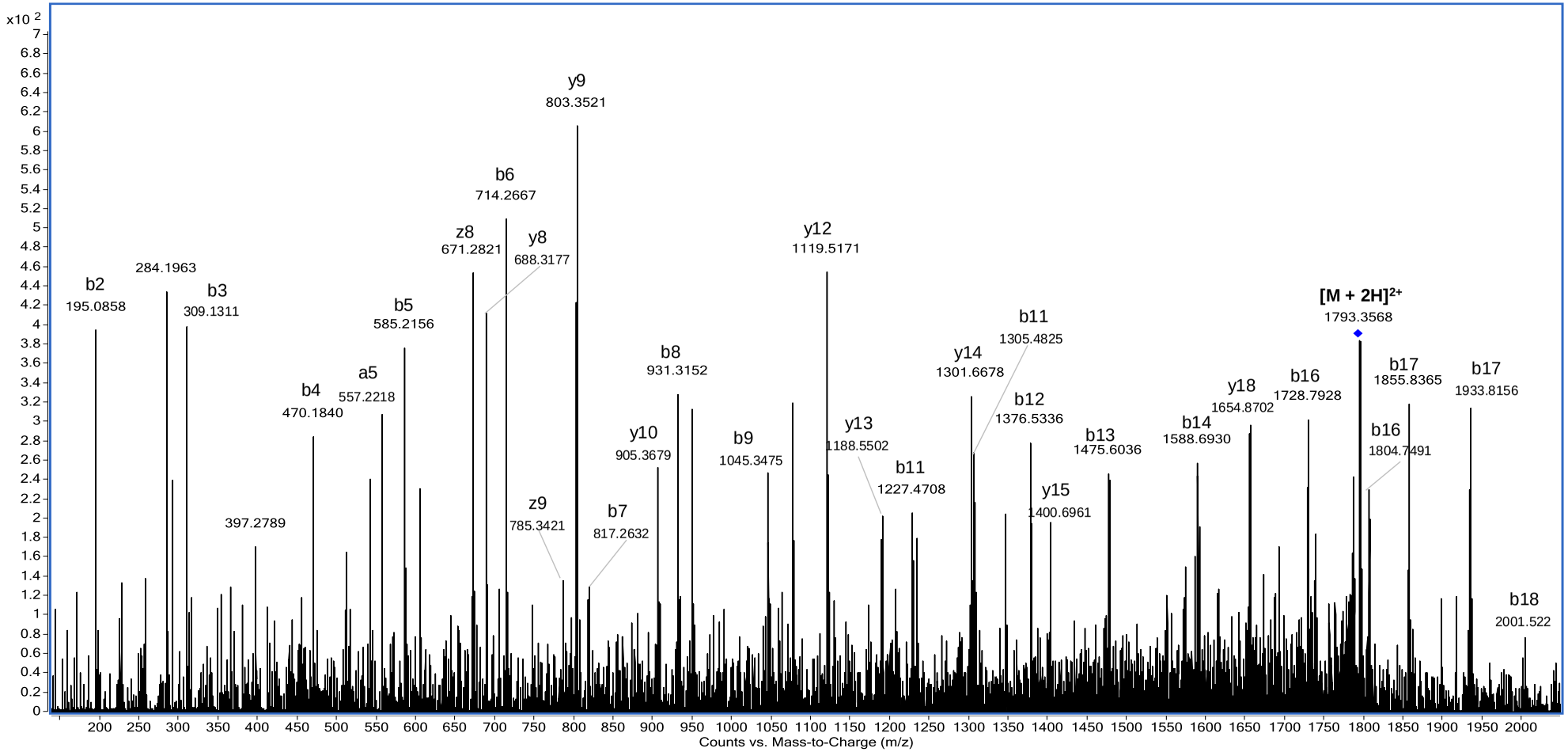
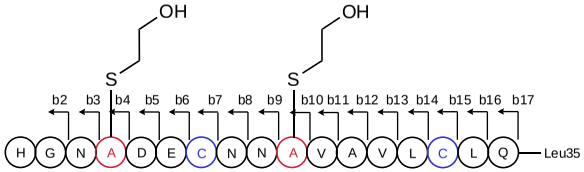


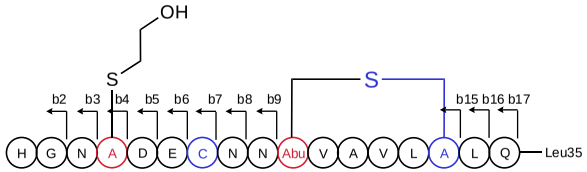

**Figure 3.25.** HR ESI-MSMS spectrum of double addition tikitericin (**37**) (M = 3584.6376 Da): precursor ion *m/z* 1793.33 $[M + 2H]^{2+}$, CID energy = 70.

**Table 3.9.** HR ESI-MSMS fragmentation pattern of double addition tikitericin (**37**) (M = 3584.6376 Da): precursor ions *m/z* 1195.89 $[M + 3H]^{3+}$, *m/z* 1793.33 $[M + 2H]^{2+}$.



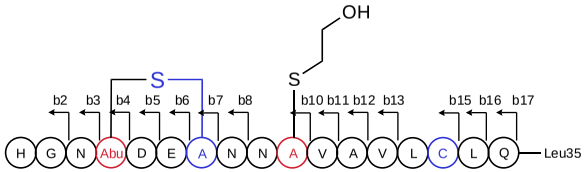

b ions

| Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) |
|---|---|---|---|
| 1 | 138.0667 | | |
| 2 | 195.0882 | 195.0877 | – 2.56 |
| 3 | 309.1311 | 309.1309 | – 0.65 |
| 4 | 470.1822 | 470.1828 | 1.28 |
| 5 | 585.2091 | 585.2050 | – 7.01 |
| 6 | 714.2517 | 714.2558 | 5.74 |
| 7 | 817.2609 | 817.2632 | 2.81 |
| 8 | 931.3038 | 931.3177 | 14.93 |
| 9 | 1045.3468 | 1045.3475 | 0.67 |
| 10 | 1206.3978 | 1206.4238 | 21.55 |
| 11 | 1305.4662 | 1305.4825 | 12.49 |
| 12 | 1376.5033 | 1376.5185 | 11.04 |
| 13 | 1475.5717 | 1475.5610 | – 7.25 |
| 14 | 1588.6558 | 1588.6399 | – 10.01 |
| 15 | 1691.6650 | 1691.6817 | 9.87 |
| 16 | 1804.7491 | 1804.7605 | 6.32 |
| 17 | 1932.8076 | 1932.8244 | 8.69 |




b ions

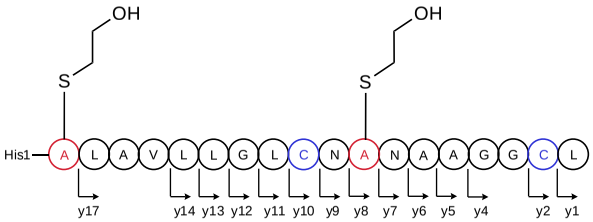
| Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) |
|---|---|---|---|
| 1 | 138.0667 | | |
| 2 | 195.0882 | 195.0877 | – 2.56 |
| 3 | 309.1311 | 309.1309 | – 0.65 |
| 4 | 470.1822 | 470.1828 | 1.28 |
| 5 | 585.2091 | 585.2050 | – 7.01 |
| 6 | 714.2517 | 714.2558 | 5.74 |
| 7 | 817.2609 | 817.2632 | 2.81 |
| 8 | 931.3038 | 931.3177 | 14.93 |
| 9 | 1045.3468 | 1045.3475 | 0.67 |
| 10 | 1128.3839 | 1128.3786 | – 4.70 |
| 11 | 1227.4523 | 1227.4505 | – 1.47 |
| 12 | 1298.4894 | 1298.5027 | 10.24 |
| 13 | 1397.5578 | 1397.5484 | – 6.73 |
| 14 | 1510.6419 | | |
| 15 | 1613.6511 | 1613.6772 | 16.17 |
| 16 | 1726.7351 | 1726.7705 | 20.50 |
| 17 | 1854.7937 | 1854.7890 | – 2.53 |



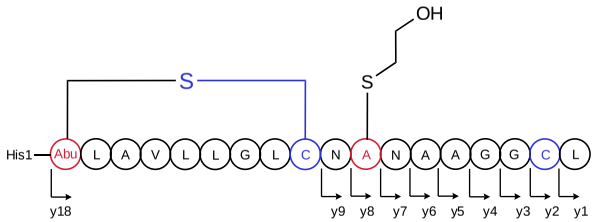

b ions

| Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) |
|---|---|---|---|
| 1 | 138.0667 | | |
| 2 | 195.0882 | 195.0877 | – 2.56 |
| 3 | 309.1311 | 309.1309 | – 0.65 |
| 4 | 392.1682 | 392.1674 | – 2.04 |
| 5 | 507.1952 | 507.1944 | – 1.58 |
| 6 | 636.2378 | 636.2336 | – 6.60 |
| 7 | 739.2470 | 739.2391 | – 10.69 |
| 8 | 853.2899 | 853.2967 | 7.97 |
| 9 | 967.3328 | | |
| 10 | 1128.3839 | 1128.3786 | – 4.70 |
| 11 | 1227.4523 | 1227.4505 | – 1.47 |
| 12 | 1298.4894 | 1298.5027 | 10.24 |
| 13 | 1397.5578 | 1397.5484 | – 6.73 |
| 14 | 1510.6419 | | |
| 15 | 1613.6511 | 1613.6772 | 16.17 |
| 16 | 1726.7351 | 1726.7705 | 20.50 |
| 17 | 1854.7937 | 1854.7890 | – 2.53 |

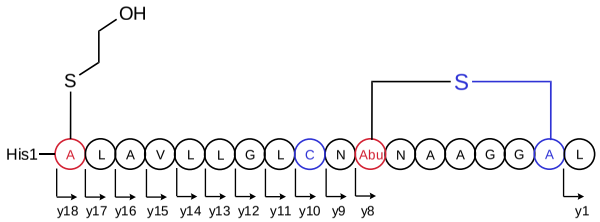
**Table 3.9.** HR ESI-MSMS fragmentation pattern of double addition tikitericin (**37**) (M = 3584.6376 Da): precursor ions *m/z* 1195.89 $[M + 3H]^{3+}$, *m/z* 1793.33 $[M + 2H]^{2+}$. (cont.)




| y ions | | | |
|---|---|---|---|
| Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) |
| 18 | 1809.8736 | | |
| 17 | 1662.8382 | 1662.8601 | 13.17 |
| 16 | 1549.7541 | | |
| 15 | 1478.7170 | | |
| 14 | 1379.3786 | 1379.3856 | 5.07 |
| 13 | 1266.5645 | 1266.5994 | 27.55 |
| 12 | 1153.4805 | 1153.4762 | – 3.73 |
| 11 | 1096.4590 | 1096.4553 | – 3.37 |
| 10 | 983.3749 | 983.3993 | 24.81 |
| 9 | 880.3657 | 880.3617 | – 4.54 |
| 8 | 766.3229 | 766.3319 | 11.74 |
| 7 | 605.2719 | 605.2641 | – 12.89 |
| 6 | 491.2289 | 491.2246 | – 8.75 |
| 5 | 420.1918 | 420.1927 | 2.14 |
| 4 | 349.1547 | 349.1497 | – 14.32 |
| 3 | 292.1333 | | |
| 2 | 235.1118 | 235.1108 | – 4.25 |
| 1 | 132.1025 | 132.1025 | 0.38 |




| y ions | | | |
|---|---|---|---|
| Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) |
| 18 | 1731.8596 | 1731.8420 | – 10.13 |
| 17 | 1662.8382 | 1662.8601 | 13.17 |
| 16 | 1549.7541 | | |
| 15 | 1478.7170 | | |
| 14 | 1379.3786 | 1379.3856 | 5.07 |
| 13 | 1266.5645 | 1266.5994 | 27.55 |
| 12 | 1153.4805 | 1153.4762 | – 3.73 |
| 11 | 1096.4590 | 1096.4553 | – 3.37 |
| 10 | 983.3749 | 983.3993 | 24.81 |
| 9 | 880.3657 | 880.3617 | – 4.54 |
| 8 | 766.3229 | 766.3319 | 11.74 |
| 7 | 605.2719 | 605.2641 | – 12.89 |
| 6 | 491.2289 | 491.2246 | – 8.75 |
| 5 | 420.1918 | 420.1927 | 2.14 |
| 4 | 349.1547 | 349.1497 | – 14.32 |
| 3 | 292.1333 | | |
| 2 | 235.1118 | 235.1108 | – 4.25 |
| 1 | 132.1025 | 132.1025 | 0.38 |




| y ions | | | |
|---|---|---|---|
| Pos | Exact (*m/z*) | Observed (*m/z*) | Δ (ppm) |
| 18 | 1731.8596 | 1731.8420 | – 10.13 |
| 17 | 1584.8242 | 1584.7214 | – 64.83 |
| 16 | 1471.7402 | | |
| 15 | 1400.7031 | 1400.6961 | – 4.96 |
| 14 | 1301.6347 | 1301.6322 | – 1.88 |
| 13 | 1188.5506 | 1188.5502 | – 0.29 |
| 12 | 1075.4665 | 1075.4653 | – 1.07 |
| 11 | 1018.4451 | 1018.4660 | 20.57 |
| 10 | 905.3610 | 905.3679 | 7.68 |
| 9 | 802.3518 | 802.3520 | 0.31 |
| 8 | 688.3090 | 688.3077 | – 1.89 |
| 7 | 605.2719 | 605.2641 | – 12.89 |
| 6 | 491.2289 | 491.2246 | – 8.75 |
| 5 | 420.1918 | 420.1927 | 2.14 |
| 4 | 349.1547 | 349.1497 | – 14.32 |
| 3 | 292.1333 | | |
| 2 | 235.1118 | 235.1108 | – 4.25 |
| 1 | 132.1025 | 132.1025 | 0.38 |

## 3.6 Determination of (Me)Lan Residue Stereochemistry

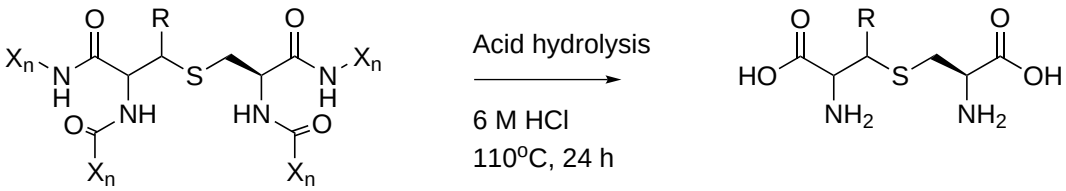
Surprisingly, full characterisation of only a small number lanthipeptides has been completed by assigning stereochemistry at the C$\alpha$ and C$\beta$ stereocenters in the (Me)Lan residues. Until 2013, all investigated lanthipeptides possessed *S*-configuration at both C$\alpha$- and C$\beta$-stereocenters of (Me)Lan cross-links and it was therefore assumed, based on the homology of tailoring enzymes, that (2*S*,6*R*)-lanthionine (*meso*-lanthionine) and (2*S*,3*S*,6*R*)-3-methyl-lanthionine were the 'natural' configurations installed by all enzymes.[80,103,119–122] Tang and van der Donk have recently reported that both MeLan residues in the two-component lantibiotic, cytolysin (**20** and **21**), are (2*R*,3*R*,6*R*)-3-methyl-lanthionine.[123] Furthermore, the two Lan residues in modified $CylL_L$ (**20**) are present in a 1:1 ratio of (2*S*,6*R*) and (2*R*,6*R*) configurations, while the Lan residue in modified $CylL_S$ (**21**) has the traditional (2*S*,6*R*) configuration. The stereochemistry of (Me)Lan cross-linking residues is now recognised as an important factor in the biological activity of lanthipeptides; the chemical synthesis of lacticin 481 analogues by Knerr showed that bioactivity was completely abolished in all isomers with non-natural configurations.[124]

Gas chromatography-mass spectrometry (GCMS) was employed to analyse the configuration at chiral carbon centres formed during the Michael-type addition of cysteine thiols to the dehydro residues, Dha and Dhb. Stereochemically-pure (Me)Lan standards were compared with hydrolysed and chemically-derivatised naturally-occurring (Me)Lan amino acid residues as described by Vederas,[125] and as has been employed by the van der Donk research group in the characterisation of cytolysin,[123] as well as lantibiotics from *Geobacillus thermodenitrificans*.[103]

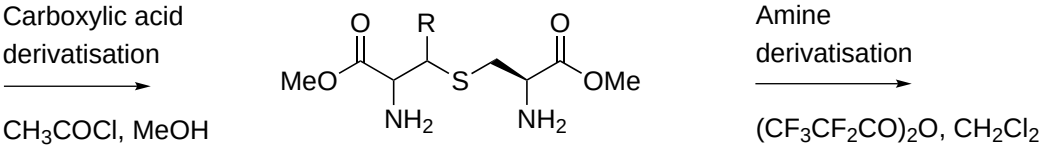
Nisin and tikitericin were hydrolysed in 6 M HCl and the C- and N-termini functionalised as methyl esters and pentafluoropropionic amides, respectively (Scheme 3.7). Formation of the derivatised products was confirmed by HR ESI-MS analysis.

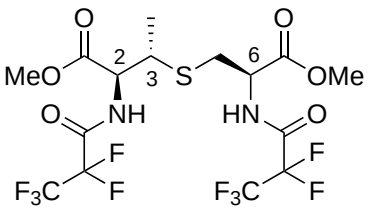
**Derivatised Lan, tikitericin:** M = 528.0413 Da, *m/z* 529.0554 $[M + H]^+$ (calculated, 529.0486 Da; $\Delta$ 12.92 ppm), *m/z* 546.0754 $[M + NH_4]^+$ (calculated, 546.0751 Da; $\Delta$ 0.53 ppm), *m/z* 551.0304 $[M + Na]^+$ (calculated, 551.0305 Da; $\Delta$ – 0.15 ppm), *m/z* 567.0043 $[M + K]^+$ (calculated, 567.0044 Da; $\Delta$ – 0.22 ppm).

**Derivatised MeLan, tikitericin:** M = 542.0569 Da, *m/z* 560.0905 $[M + NH_4]^+$ (calculated, 560.0908 Da; $\Delta$ – 0.54 ppm), *m/z* 565.0456 $[M + Na]^+$ (calculated, 565.0462 Da; $\Delta$ – 0.97 ppm), *m/z* 581.0161 $[M + K]^+$ (calculated, 581.0201 Da; $\Delta$ – 1.76 ppm).

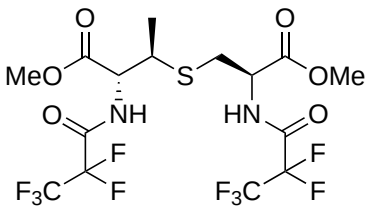



Lan R = H
MeLan R = $CH_3$

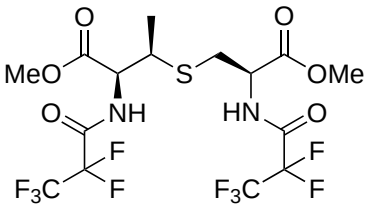


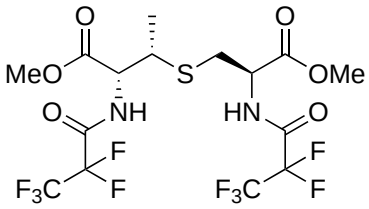
**(2*S*,3*S*,6*R*)-MeLan**



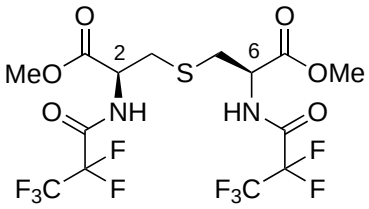
(2*R*,3*R*,6*R*)-MeLan



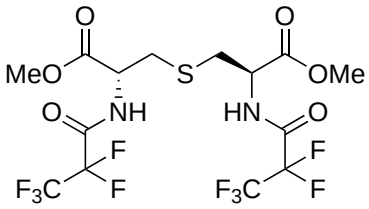
(2*S*,3*R*,6*R*)-MeLan



(2*R*,3*S*,6*R*)-MeLan



**(2*S*,6*R*)-Lan**

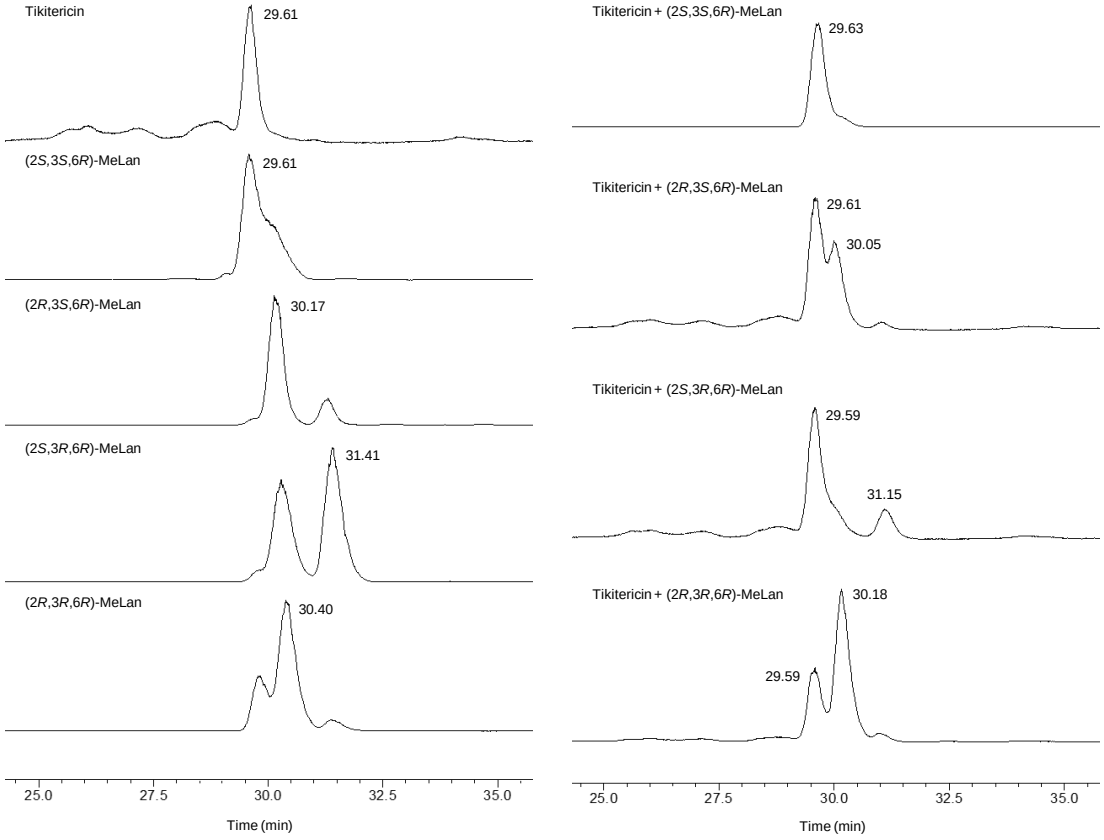


(2*R*,6*R*)-Lan

**Scheme 3.7.** Acid hydrolysis and derivatisation of (Me)Lan residues. The stereoisomers of derivatised Lan and MeLan are shown, with the most commonly observed isolates in bold.

### 3.6.1 Determination of MeLan stereochemistry

Chromatographic separation of the MeLan standards was achieved by GCMS with an achiral HP-VOC column. The hydrolysed tikitericin derivative eluted with identical retention time (RT) to the (2*S*,3*S*,6*R*) standard (Figure 3.26a), the same RT as the nisin derivative (data not shown). The (2*R*,3*S*,6*R*) and (2*R*,3*R*,6*R*) standards in particular were seen to have undergone isomerisation but a characteristic peak was identified for each standard. Characterisation was confirmed by performing a co-injection of the tikitericin derivative with each standard; a single peak was observed for the co-injection with the (2*S*,3*S*,6*R*) standard (Figure 3.26b). Fortuitously, it can therefore be concluded that all three MeLan residues of tikitericin have (2*S*,3*S*,6*R*) configuration.



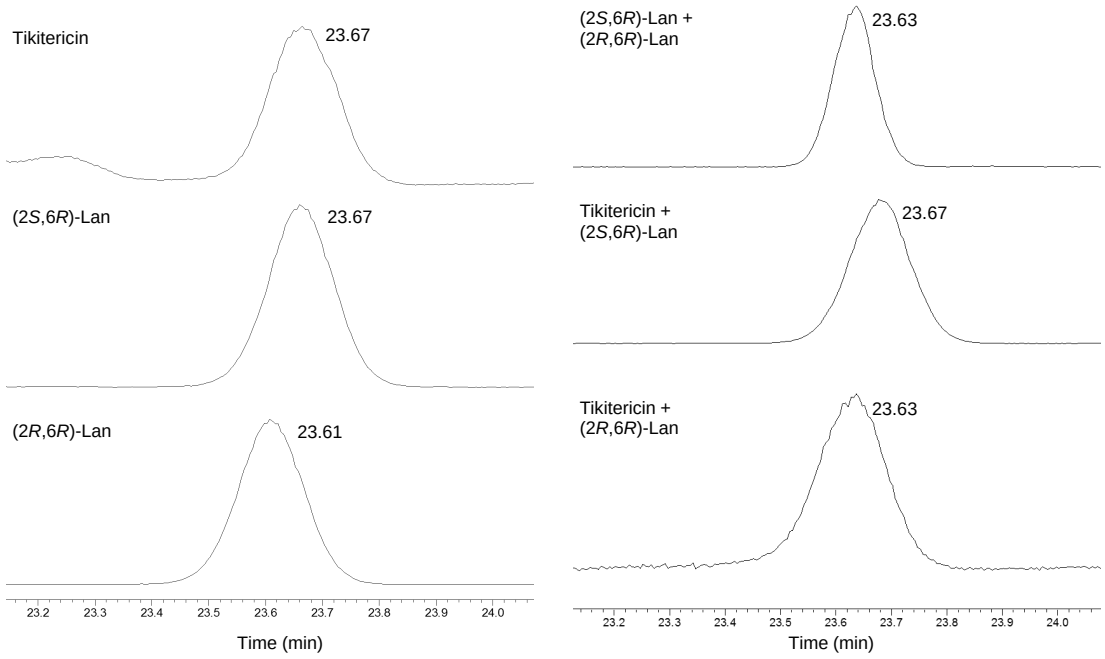

**(a)** Individual standards. **(b)** Co-injection.

**Figure 3.26.** GCMS traces for synthetic, derivatised MeLan standards and hydrolysed/ derivatised MeLan residues obtained from tikitericin. SIM mode, ions 248 and 379 Da detected.

### 3.6.2 Determination of Lan stereochemistry

Stereochemical analysis of the tikitericin Lan derivative proved to be more challenging. The two Lan synthetic standards could not be separated using chiral stationary phases. Lan derivatives were not observed using CHIRALDEX B-DM or a Chirasil-Val columns, presumably due to: (i) the column upper temperature limit (200 °C for both columns)

being lower than the boiling point of the Lan compounds, or (ii) an irreversible interaction of the compounds with the stationary phase that prevented their elution. Previous analyses of derivatised Lan compounds reported elution from an identical Chirasil-Val column at 180 °C,[103,123] although both the van der Donk and Vederas research groups have observed irreversible binding of the Lan compounds to chiral stationary phases. Temperature programmed and isothermal methods were explored using the achiral HP-VOC column but a mixture of the standards could not be resolved into two peaks. When injected individually, the (2*R*,6*R*) standard consistently eluted earlier than the (2*S*,6*R*) standard across all methods. A maximum difference in RT of 4 sec was observed for the two standards and derivatives obtained from both tikitericin and nisin repeatedly aligned with the (2*S*,6*R*) diastereomer (Figure 3.27). When the two standards were co-injected, a RT between those of the individual standards was observed (23.63 min), this shift in RT was similarly observed for a co-injection of tikitericin with the (2*R*,6*R*)-Lan standard. A coinjection of tikitericin with the (2*S*,6*R*)-Lan standard did not shift the previously observed RT of tikitericin, supporting (2*S*,6*R*) configuration of the tikitericin Lan residue.

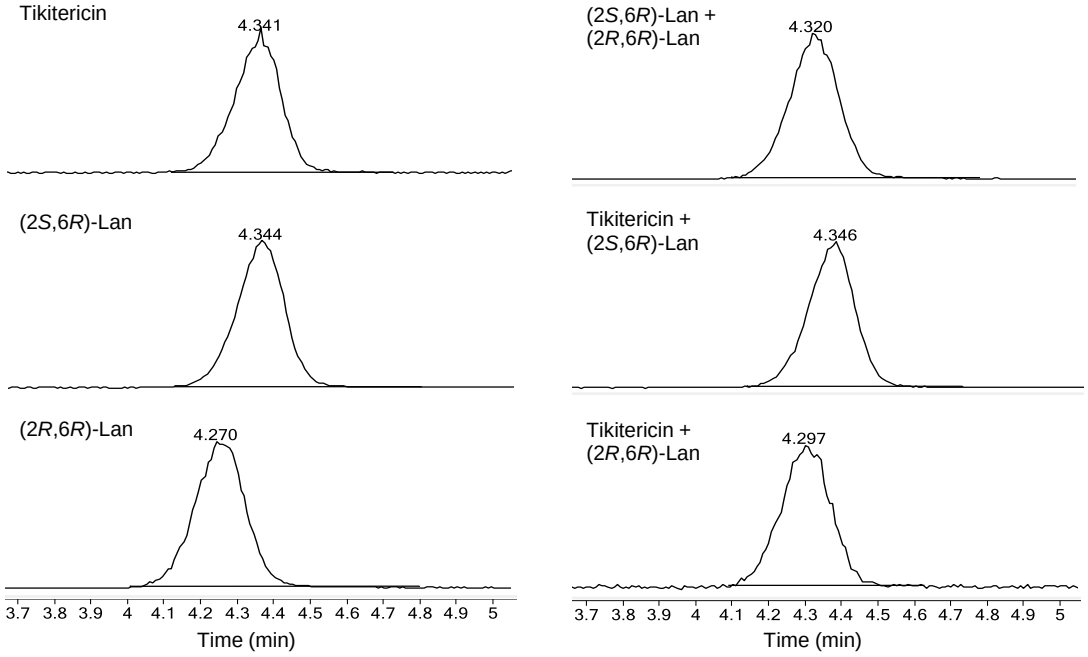



**(a)** Individual standards. **(b)** Co-injection.

**Figure 3.27.** GCMS traces for synthetic, derivatised Lan standards and hydrolysed/ derivatised Lan residues obtained from tikitericin. SIM mode, ions 248 and 365 Da detected.

To corroborate the GCMS-based assignment, lanthionine standards were analysed by HR ESI-MS (Figure 3.28). A method was developed which observed a 4 sec difference in RT between the two Lan diastereomers: 4.27 and 4.34 min for (2*R*,6*R*) and (2*S*,6*R*) configurations, respectively. A mixture of the standards could not be fully resolved into two peaks, the RT of this mixture was between those of the individual standards (4.32 min). Longer elution times observed under isocratic conditions with a lower proportion of eluent B resulted in peak broadening and no improvement in separation. A sample

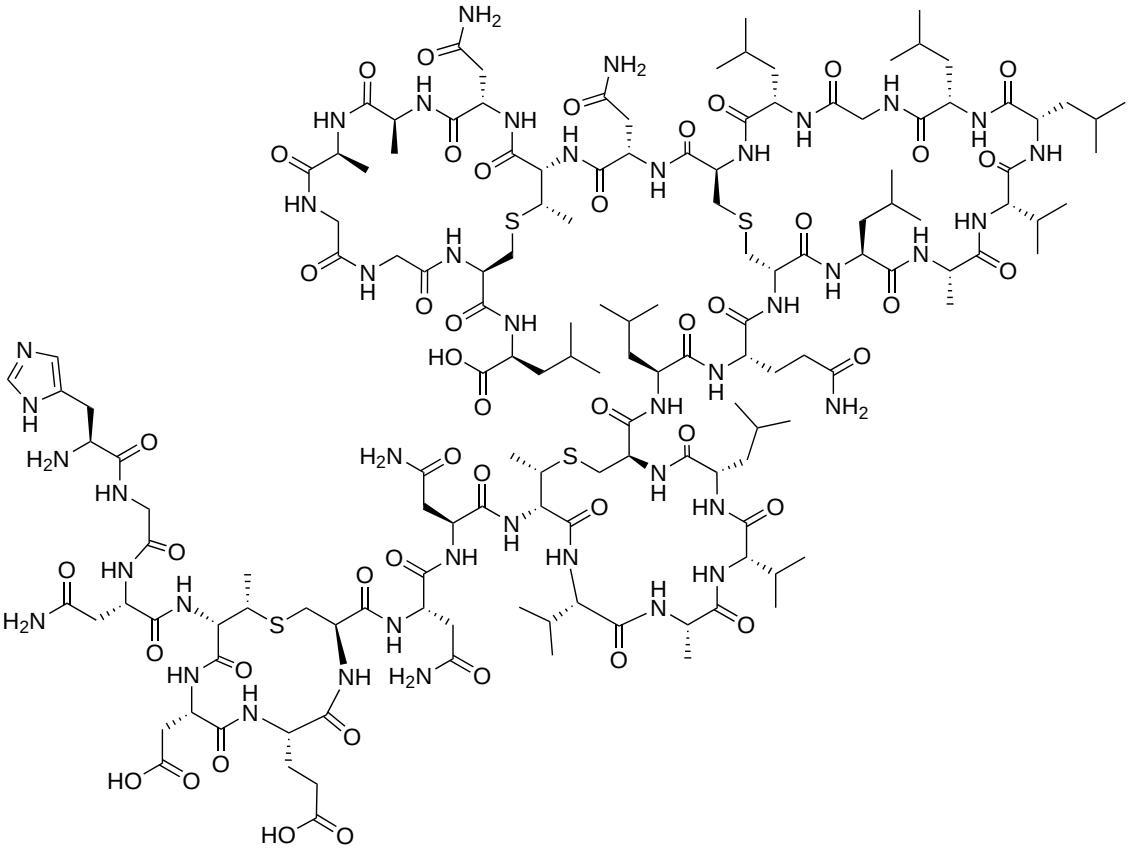
of derivatised Lan obtained from tikitericin was shown to elute with the same RT as the (2*S*,6*R*) standard. A shift in RT was not observed for a coinjection of tikitericin with the (2*S*,6*R*) standard, while a coinjection of tikitericin with the (2*R*,6*R*) standard saw a shift in RT (4.30 min). The Lan residue of tikitericin (**32**) is therefore tentatively assigned as having (2*R*,6*R*) configuration based upon the combination of GCMS and HR ESI-MS data, and similar to the majority of known lanthipeptides.




**(a)** Individual standards. **(b)** Co-injection.

**Figure 3.28.** HR ESI-MS traces for synthetic, derivatised Lan standards and hydrolysed/derivatised Lan residues obtained from tikitericin.

The novel, ribosomally synthesised lanthipeptide, tikitericin (**32**), contains four macrocyclic residues formed by the non-proteinogenic residues, (2*S*,6*R*)-lanthionine and (2*S*,3*S*,6*R*)-3-methyllanthionine. These macrocycles are constructed with linear ring topology from serine, threonine and cysteine residues in the precursor peptide, TikA, by the class II lanthionine synthetase modification enzyme, TikM. The structure of tikitericin can be represented using the shorthand notation recently proposed by Arnison *et al.*[59]



**32**

[*SS*4-7, *SS*10-15, *S*18-26, *SS*28-34-C$\beta$S]-HGNAbuDECNNAbuVAVLCLQALAVLLGLCNAbuNAAGGCL

## 3.7 Biological Activity of Tikitericin

Purified tikitericin (obtained from 1200 plates of bacilli) was quantified by $^1$H NMR using the acquisition parameters described by West.[126] The relative integration of tikitericin resonances in the amide region, compared to the methyl proton resonance of $CH_3NO_2$, quantified the sample as $\leq 1$ $\mu$g. This method has previously been utilised at VUW by the Northcote research group and is particularly useful for the quantification of sub-milligram quantities.

The biological activity of tikitericin was tested in a bacteriostatic assay against *S. aureus* (see Appendix E) at a maximum concentration of $\approx 2.5$ $\mu$g/mL (0.73 nM). No inhibitory activity was observed. This result does not prove that tikitericin is inactive against Gram

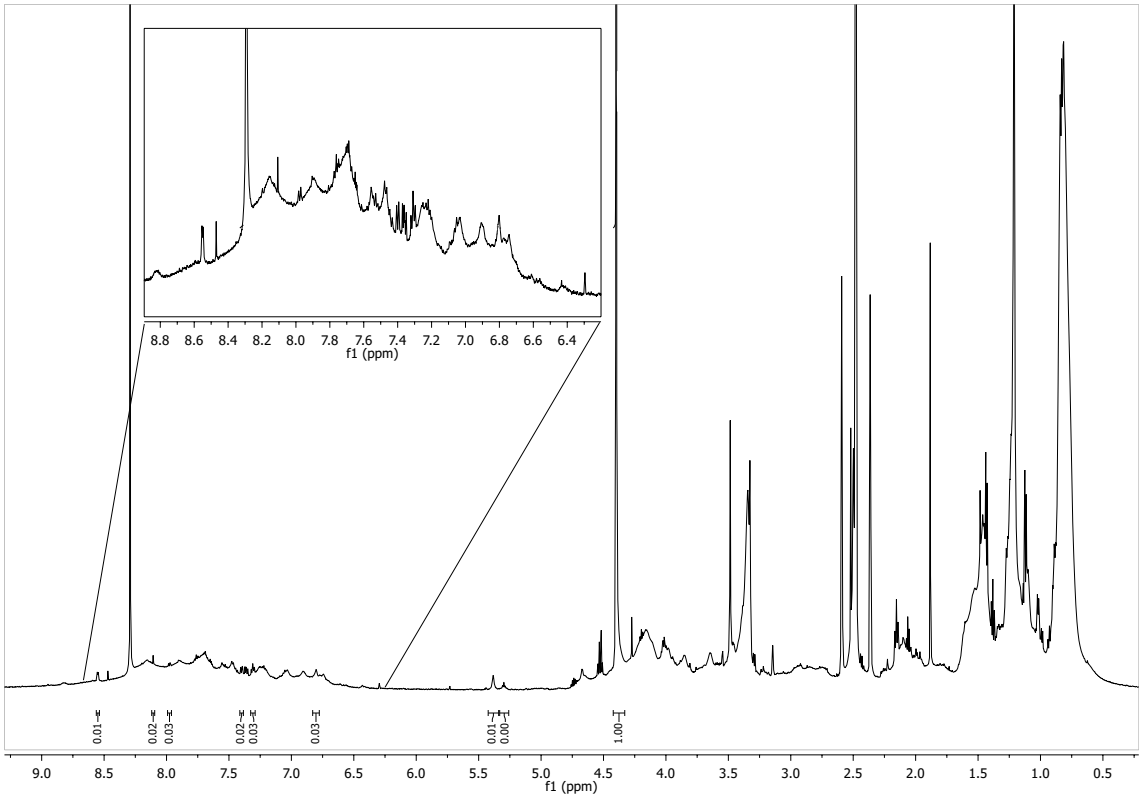



**Figure 3.29.** $^{1}$H NMR spectrum ($(CD_3)_2SO$) of tikitericin.

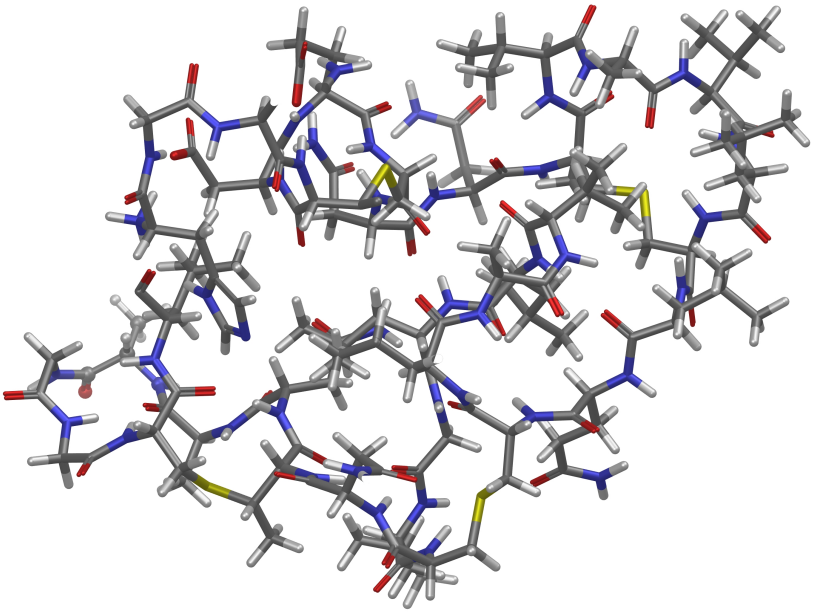
positive strains of bacteria, instead it simply highlights the need for greater quantities of material. A comprehensive study by Mota-Meira *et al.* (Table 3.10) shows that the minimum inhibitory concentration of lantibiotics mutacin B-Ny266, nisin A, vancomycin and oxacillin against Gram positive bacteria varies significantly between species and strains.[127]

**Table 3.10.** Activities of mutacin B-Ny266, nisin A, vancomycin, and oxacillin against Gram-positive bacteria, adapted from Mota-Meira *et al.*[127]

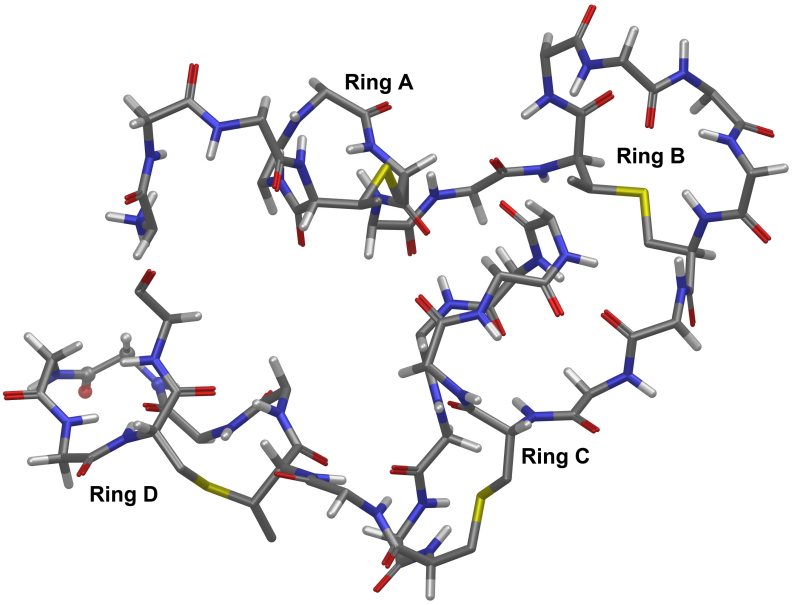
| | MIC($\mu$g/mL) of: | | | |
|---|---|---|---|---|
| Organism (no. of strains) | Mutacin B-Ny266 | Nisin A | Vancomycin | Oxacillin |
| *Micrococcus* spp. (4) | 0.05 (0.03 - 0.08) | 1.1 (0.3 - 16.7) | 1.0 (0.5 - 2.0) | 2.1 (0.01 - 4.2) |
| *Enterococcus* spp. (12) | 12.8 (1.6 - 25.6) | 16.7 (8.4 - 33.4) | 3.9 (1.9 - 120) | 11.9 (4.2 - 63.4) |
| *Listeria* spp. (23) | 0.8 (0.4 - 2.0) | 4.2 (1.1 - 16.7) | 1.0 (0.3 - 2.1) | 4.2 (0.5 - 15.9) |
| *Staphylococcus* spp. (19) | 1.6 (0.1 - 18.1) | 4.2 (1.5 - 83.6) | 3.8 (0.3 - 15.1) | 7.9 (0.1 – 127) |
| *Streptococcus* spp. (17) | 0.4 (0.03 - 6.4) | 8.4 (0.3 - 83.6) | 0.5 (0.3 - 1.0) | 0.1 (0.02 – 1.1) |
| *Bacillus* spp. (5) | 0.8 (0.01 - 6.4) | 4.2 (0.01 - 10.5) | 0.5 (0.1 – 0.5) | 0.3 (0.002 - 127) |
| *Clostridium* spp. (4) | 0.2 (0.01 - 3.2) | 1.1 (0.1 - 4.2) | 0.5 (0.3 – 4.0) | 1.1 (0.1 - 4.2) |
| *Actinomyces viscosus* (2) | 0.4 (0.4 - 0.8) | 83.6 (41.8 - 83.6) | 0.5 (0.5 - 1.0) | 0.5 (0.5 – 0.8) |
| *Peptostreptococcus* spp. (2) | 0.4 (0.2 - 0.8) | 2.1 (1.1 - 2.1) | 0.5 (0.5 - 1.0) | 0.3 (0.1 – 0.5) |
| *Corynebacterium diphtheriae*(1) | 0.4 (0.4 - 0.8) | 4.2 (4.2 - 8.4) | 1.0 (0.5 - 1.0) | 2.1 (1.1 – 2.1) |
| *Gardnerella vaginalis* (1) | 0.03 (0.01 - 0.05) | 1.1 (1.1 - 2.1) | 0.4 (0.3 - 0.5) | 0.1 (0.07 – 0.1) |
| *Mycobacterium smegmatis* (2) | 32 (25 - 70) | 8.4 (7.4 - 15) | 8.0 (8.0 - 15) | 95.1 (63.4 – 126.8) |
| *Propionibacterium acnes* (2) | 1.2 (0.8 - 1.6) | 2.1 (1.1 - 4.2) | 1.0 (0.5 - 1.0) | 0.5 (0.5 – 1.1) |

## 3.8 Computer Modelling of Tikitericin

Ribosomal peptide antibiotics that have undergone post-translational modification of the side chains or peptide backbone are conformationally restricted for both increased binding affinity to a specific molecular target and resistance against proteases. To explore the rigidity of tikitericin, the three-dimensional structure was modelled by Monte Carlo methods using Maestro software (Schrödinger). The tikitericin model was constructed



**(a)** Full depiction




**(b)** Amino acid side chains are omitted for clarity

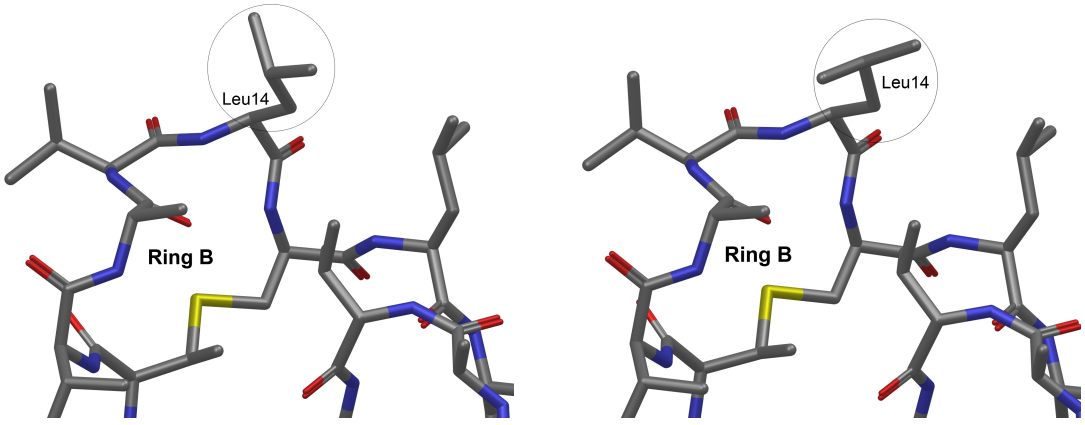
**Figure 3.30.** Three-dimensional rendering of tikitericin in its lowest energy conformation. Carbon = grey, oxygen = red, nitrogen = blue, hydrogen = white, sulfur = yellow.

with assigned configurations for (Me)Lan residues and with charged residues in their zwitterionic form. An energy minimisation was performed using an OPLS forcefield suitable for proteins and small molecules and a water solvent, 5000 iterations and a convergence of 0.005. A mixed torsional/low mode conformational search was performed with 10,000 steps and an energy window of 5 kJ/mol.

**Table 3.11.** Conformations of tikitericin.

| Conf | Times found | Potential energy (kJ/mol) | Rel. potential energy |
|---|---|---|---|
| 1 | 73 | – 6767.96 | 0.00 |
| 2 | 82 | – 6767.78 | 0.17 |
| 3 | 330 | – 6766.76 | 1.19 |
| 4 | 418 | – 6766.33 | 1.62 |
| 5 | 13 | – 6766.22 | 1.73 |
| 6 | 32 | – 6766.13 | 1.83 |
| 7 | 17 | – 6765.83 | 2.12 |
| 8 | 74 | – 6764.70 | 3.26 |
| 9 | 2 | – 6764.49 | 3.46 |
| 10 | 5 | – 6764.47 | 3.48 |
| 11 | 44 | – 6764.37 | 3.58 |
| 12 | 10 | – 6764.29 | 3.66 |
| 13 | 48 | – 6764.16 | 3.80 |
| 14 | 17 | – 6764.03 | 3.93 |
| 15 | 56 | – 6763.99 | 3.96 |
| 16 | 1 | – 6763.87 | 4.09 |
| 17 | 85 | – 6763.77 | 4.18 |
| 18 | 1 | – 6763.68 | 4.27 |
| 19 | 1 | – 6763.50 | 4.45 |
| 20 | 5 | – 6763.42 | 4.54 |
| 21 | 1 | – 6763.14 | 4.81 |

After three months, 8900 steps had been performed. The calculation returned 20 conformers within 5 kJ/mol of the global minimum (Table 3.11), which was found 73 times and is depicted in Figure 3.30. The majority of high energy conformers arise from subtle vibrational differences. The most significant conformational change is observed in three of the high energy conformers (entries 9, 15 and 19), which show rotation around the C$\beta$ – C$\gamma$ bond of Leu14 (Figure 3.31).




**(a)** Low energy rotamer **(b)** High energy rotamer

**Figure 3.31.** Tikitericin rotamers. Hydrogen atoms are omitted for clarity.

The similarity of the 21 conformers generated in this calculation supports the hypothesis that the (Me)Lan residues serve the function of constraining the conformational flexibility of tikitericin, thereby conferring improved affinity for a biological target.

# Chapter 4

# Concluding Remarks and Future Directions

The research described in this thesis is the first example of both genomic- and mass spectrometry-guided natural product isolation, as well as the first example of microbial natural product isolation to be conducted at VUW. Consequently, a major focus of the project was to adapt the protocol currently employed by the Northcote and Keyzers research groups, and develop methodology suited to microbial natural product isolation.

The remarkable sensitivity of mass spectrometry enabled complete characterisation of a novel lanthipeptide, tikitericin, isolated from cultures of *Thermogemmatispora* strain T81 grown on solid media. This putative lantibiotic exceeds the generally accepted molecular weight boundary of $< 3000$ Da for natural products and was detected in low abundance. Low biomass is a trend which has been previously observed for extremophilic microorganisms; the presence of tikitericin would undoubtedly have been missed using historically favoured techniques of bioassay or NMR-guided isolation. Structurally-related lanthipeptides show broad spectrum inhibitory activity against Gram positive strains of bacteria and tikitericin may have future antibacterial applications. However, targeted isolation of tikitericin from more than 1000 plates of strain T81 was unsuccessful in yielding sufficient biomass for biological studies or to obtain 2D NMR data. To enable the study of tikitericin's biological function, research must be continued in one of three directions.

Firstly, the culture conditions under which strain T81 maximises tikitericin production could be determined by exploring batch reactor processes employed by Callaghan Innovation. At present, the cultivation of strain T81 by Callaghan Innovation is focused towards the production of lipids as opposed to secondary metabolites.

Studies of RiPP maturation enzymes have improved *in vivo* and *in vitro* engineering systems and the large-scale production of lantibiotics for chemotherapeutic application is now accessible. *In vitro* expression of the lanthipeptide gene cluster in a mesophile host

such as *E. coli* is therefore a logical progression. Synthesising only the precursor peptide and lanthionine sythetase genes for expression in *E. coli* would circumvent the plasmid vector instability observed by M. Stott. Site-directed mutagenesis to substitute the C-terminal Gly residue of the leader peptide for a Lys and incorporate a trypsin cleavage site would enable the modified core peptide to be cleaved extra-cellularly from the leader residues, an approach utilised by van der Donk and co-workers in the identification of lantibiotics from *G. thermodenitrificans*.[103] Furthermore, fusion of N-terminal $His_6$ tags to the leader peptide prior to trypsin digestion would improve the purification process through use of immobilised metal affinity chromatography, also utilised by the van der Donk research group in the expression of enterococcal cytolysin in *E. coli*.[123]

An alternative avenue is the chemical synthesis of tikitericin, a method which would not rely on biosynthetic machinery. The advantages of solid-supported synthesis can be seen in recent publications by the Vederas and van der Donk research groups.[60,124,125] This method is not restricted to reproducing the naturally occurring (Me)Lan stereochemistry introduced by modification enzymes. Solid-supported synthesis is also remarkably efficient, with Knerr reporting the total synthesis of lacticin 481 in only 10 – 12 days,[124] and would enable milligram-scale quantities to be isolated for comprehensive NMR and biological studies.

Mass spectrometric screening was unable to identify thiopeptide products of a second bioinformatically-predicted biosynthetic cluster, and it appears that the thiopeptide gene cluster is not expressed in its natural host under the growth conditions explored. It is highly unlikely that the thiopeptides could be characterised without pursuing the chemoenzymatic platform. While MS characterisation of tikitericin was achieved with nanogram quantities, structure elucidation of the heavily post-translationally modified thiopeptides will require greater quantities of material. The extent of peptide backbone modification renders standard tandem MSMS sequencing methods ineffective because a high proportion of rigid thiazol(in)e moieties reduces the number of facile fragmentation pathways. Instead, milligram quantities would be required for structural characterisation by NMR. In addition to mass spectral and NMR analyses, microbial extracts could be screened for the presence of thiopeptides by IR and UV spectroscopic methods, searching for the characteristic thiazole CH stretching frequency and comparing UV spectra to known thiopeptides. This study included thorough analysis of the thiopeptide gene cluster of strain T81, the biosynthetic predictions described will provide the basis for future research towards isolation of these novel natural products.

# Chapter 5

# General Experimental Procedures

Supelco Diaion HP20® and HP20SS® polystyrene (divinylbenzene) resins were used for reversed-phase benchtop column chromatography unless otherwise stated. Size-exclusion chromatography was performed using Sephadex® LH-20. HPLC-grade MeOH (Fisher) and MeCN (Fisher) were used for general purposes and for HPLC purification. $H_2O$ was distilled immediately prior to use. MS-grade MeOH (Fisher) and MeCN (Fisher) were used for MS analyses. Chemical reagents were purchased from Sigma-Aldrich and used without further purification, unless otherwise stated. Filtration was performed over Whatman Grade 6 (3 $\mu$m) cellulose filter paper. Millipore C18 ZipTips (10, 100 $\mu$L) were used for desalting proteolytic digestion and linearisation reactions.

For MS analyses, an AB Sciex TOF/TOF 5800 MALDI MS (Applied Biosystems) equipped with a diode pulse laser (355 nm, 1 kHz) and a 6530 Q-TOF LC/MS (Agilent Technologies) were used.

MALDI MS parameters were as follows: operation mode, MS reflector positive; CID, off; mass range, *m/z* 700 – 4000; total shots/spectrum, 5000; stage motion, random continuous 600 $\mu$m/s; average laser intensity, 7000; pulse rate, 400 Hz. Mass spectra were taken from the edge of the crystals and the laser fluence was optimised to achieve the highest analyte S/N ratios. Mass spectra were internally calibrated using CalMix2 (Applied Biosystems), with peak matching performed against the monoisotopic peaks of bradykinnin fragment 1-7 (*m/z* 757.4000), angiotensin II (*m/z* 1046.5420), P14R (*m/z* 1533.8580), ACTH fragment 18-39 (*m/z* 2465.1990) and insulin oxidised B chain (*m/z* 3494.6510). Peak matching criteria for calibration were as follows: S/N ratio, $\geq$ 200; mass tolerance, $\pm$ *m/z* 0.05; matched peaks, 4; outlier error, $\leq$ 3 ppm. Data Explorer (Applied Biosystems) was used for the analysis of the MS spectra using the following parameters: baseline correction; Gaussian smooth, 5; S/N ratio, $\geq$ 20; mass peak filter, monoisotopic peaks; outlier error, $\leq$ *m/z* 0.8.

MALDI MSMS parameters were as follows: operating mode, MSMS 1 KV positive; CID, on; precursor width range, $\pm$ 5 Da; metastable suppressor, on; total shots/spectrum, 5000;

stage motion, random continuous 600 $\mu$m/s; laser intensity, 6000; pulse rate, 1000 Hz. The monoisotopic angiotensin 1 peak (*m/z* 1296.800) was selected as the precursor ion for calibration; MSMS spectra were then calibrated using known fragment ion masses from this precursor ion. Peak matching criteria for calibration were as follows: S/N ratio, $\geq$ 10; mass tolerance, $\pm$ *m/z* 0.3; matched peaks, 3; outlier error, $\leq$ 20 ppm. Data Explorer was used for peak analysis.

Q-TOF ESI-MS parameters were as follows: positive ion mode electrospray ionisation; gas temperature, 300 °C; gas flow, 9 L/min; nebulizer, 30 psi; capillary voltage, 3500 V; nozzle voltage, 500 V; fragmentor voltage, 140 V; acquisition rate, 2 spectra/s; mass range, *m/z* 100 – 3200. An Eclipse plus reversed-phase C18 column 30 mm x 2.1 mm, 3.5 $\mu$m (Agilent Technologies) at a flow rate of 0.4 mL/min was used for chromatographic separation, unless otherwise stated. Column temperature held at 35 °C. Eluent A was 99.9% $H_2O$/0.1% $NH_4HCO_2$. Eluent B was 99.9% MeCN/0.1% $HCO_2H$. Unless otherwise stated, LC analyses were performed with the following gradient: 5% eluent B for 0.5 min, eluent B increased to 100% at 4 min, eluent B held at 100% until 4.5 min, eluent B reduced to 5% at 4.51 min and held until 7 min. Q-TOF ESI-MSMS parameters were as follows: max time between MS spectra, 5 s; mass range, *m/z* 100 – 3200; MSMS scan rate, 2 spectra/s; CID collision gas, $N_2$; CID gas pressure, 16 psi; isolation width, 4 amu; collision energies, 50 – 70 arbitrary units unless otherwise stated.

NMR spectra were obtained using a 600 MHz Varian DirectDrive spectrometer equipped with a triple resonance HCN cryogenic probe, operating at 600 MHz for $^1$H nuclei. Spectra were recorded in $CD_3OD$ (Cambridge Isotopes), $(CD_3)_2SO$ (Cambridge Isotopes), $CD_3CN$ (Cambridge Isotopes) and $D_2O$ (Cambridge Isotopes) and were referenced to the residual solvent peak [($CD_3OD$: $\delta_H$ 3.31), (($CD_3)_2SO$: $\delta_H$ 2.50), ($CD_3CN$: $\delta_H$ 1.94) and ($D_2O$: $\delta_H$ 4.79)].[128]

## 5.1 MALDI-TOF MS Screening of Strain T81

Liquid and plated single and co-cultured bacterial strains were cultured by GNS Science unless otherwise stated. Details of the strains of bacteria and fungi are provided in Appendix A.

### 5.1.1 Whole-cell screening of plated cultures

Using pipette tips, pinhead amounts of single bacterial colonies were carefully picked without destroying the surface of the agar and mixed with aliquots of matrix solution (5, 10, and 20 $\mu$L, 10 mg/mL 2,5-dihydroxybenzoic acid (DHB) solubilised in 50%

MeCN/0.1% trifluoroacetic acid (TFA), v/v in $H_2O$). Aliquots (1 $\mu$L) were spotted onto a stainless steel MALDI MS target plate and dried with hot air. From each competition experiment, scrapings were taken from the centre of the T81 colony, zone of inhibition (if observed), and centre of the co-strain colony.

### 5.1.2 Screening of liquid cultures

Liquid cultures were filtered over celite and filter paper. The filtrate was passed through a column of HP20 resin (15 mL) and the column washed with $H_2O$ then eluted generating three 45 mL fractions of 30%, 75% and 100% $Me_2CO$, v/v in $H_2O$. Cell mass fractions were extracted twice with MeOH overnight and cyclic loaded onto a column of HP20 resin (15 mL), the column washed with $H_2O$ then eluted generating three 45 mL fractions of 30%, 75% and 100% $Me_2CO$, v/v in $H_2O$. Solvent was removed *in vacuo* and samples stored at 4 °C until further use. Samples were prepared for MALDI-TOF MS analysis by mixing aliquots of each fraction (1 $\mu$L), suspended to a concentration of 4 mg/mL in $H_2O$, with aliquots of matrix solution (5 and 10 $\mu$L, 10 mg/mL DHB solubilised in 50% MeCN/0.1% TFA, v/v in $H_2O$). Aliquots (1 $\mu$L) were spotted onto a stainless steel MALDI MS target plate and dried with hot air.

## 5.2 Purification of Nisin A

Nisin A (1 g, 2.5% in salts and milk protein) was dissolved in $H_2O$ (20 mL) and passed through a column of HP20-SS resin (2 mL), the column washed with $H_2O$ then eluted generating five 6 mL fractions of 30%, 40%, 50%, 60% and 100% $Me_2CO$, v/v in $H_2O$. Solvent was removed *in vacuo* and samples stored at 4 °C until further use. Samples were prepared for MALDI-TOF MS analysis by mixing aliquots of each fraction (1 $\mu$L), suspended to a concentration of 4 mg/mL in $H_2O$, with aliquots of matrix solution (5 and 10 $\mu$L, 10 mg/mL DHB solubilised in 50% MeCN/0.1% TFA; v/v in $H_2O$). Aliquots (1 $\mu$L) were spotted onto a stainless steel MALDI MS target plate and dried assisted by hot air.

## 5.3 Isolation of Tikitericin

The agar was scraped from 100 plates of single-strain T81 cultures and extracted twice for 24 h in MeOH. The MeOH extracts were filtered over celite and filter paper, and cyclic loaded onto a column of HP20 resin (15 mL), the column washed with $H_2O$ then eluted generating three 45 mL fractions of 30%, 60% and 100% $Me_2CO$, v/v in $H_2O$. Solvent

was removed *in vacuo* and samples stored at 4 °C until further use. Samples were prepared for MALDI-TOF MS analysis by mixing aliquots of each fraction (1 $\mu$L), suspended to a concentration of 4 mg/mL in $H_2O$, with aliquots of matrix solution (5 and 10 $\mu$L, 10 mg/mL DHB solubilised in 50% MeCN/0.1% TFA, v/v in $H_2O$). Aliquots (1 $\mu$L) were spotted onto a stainless steel MALDI MS target plate and dried with hot air.

Tikitericin-containing fractions (60% $Me_2CO$, v/v in $H_2O$, 3 batches of 100 plates) were combined and loaded onto a 1.6 m column of LH-20 resin (100 g), and the column eluted with 50% MeOH, v/v in $H_2O$ with a flow rate of 0.4 mL/min. LH20 fractions were collected using an automated fraction collector (15 min/tube, 6 mL) and analysed by MALDI-TOF MS. Aliquots of LH20-purified fractions (1 $\mu$L) and matrix solution (5 $\mu$L, 10 mg/mL DHB solubilised in 50% MeCN/0.1% TFA, v/v in $H_2O$) were mixed and spotted (1 $\mu$L) onto a stainless steel MALDI MS target plate, dried with hot air.

The final purification of tikitericin was optimised by HR ESI-MS using a Zorbax C18 column 100 mm x 4.6 mm i.d., 3.5 $\mu$m (Agilent Technologies). The binary pump was operated with a flow rate of 1 mL/min and a solvent system of eluent A: 0.2 M $HCO_2H$ in 30% MeCN/70% water, and eluent B: 0.2 M $HCO_2H$ in 100% MeCN. The solvent was held at 0% B for 10 min, increased to 15% B over 30 min, increased to 100% B over 1 min and held for 4 min, decreased to 0% B over 2 min, returned to 0% B over 1 min and held for 2 min. Tikitericin eluted at approximately 39% MeCN with a retention time of 26 min. This method was transferred to analytical-scale HPLC, guided by simultaneous ELSD and UV detection. UV absorption was monitored at 220 nm and the ELSD operated with the following parameters: evaporator temp, 80 °C; nebuliser temp, 70 °C; gas flow, 1.6 SLM; LED intensity, 100%. An adjustable flow splitter Quicksplit® (Analytical Scientific Instruments) was used to split the mobile phase between collection and ELSD in a 20:1 ratio. Stainless steel capillaries (0.17 mm i.d) of 55 cm and 35 cm lengths were used to direct the mobile phase from the DAD to splitter, and the splitter to ELSD, respectively. Supelco PEEK capillary tubing (0.0508 mm i.d.) of 2.25 m was connected from the splitter for simultaneous ELSD detection and collection.

**Tikitericin (32):** *m/z* 858.1594 $[M + 4H]^{4+}$ (calculated, 858.1597 Da; $\Delta$ – 0.40 ppm), *m/z* 880.1420 $[M + 4Na]^{4+}$ (calculated, 880.1417 Da; $\Delta$ 0.33 ppm), *m/z* 1143.8779 $[M + 3H]^{3+}$ (calculated, 1143.8772 Da; $\Delta$ 0.60 ppm), *m/z* 1166.1927 $[M + 3Na]^{3+}$ (calculated, 1166.1934 Da; $\Delta$ – 0.59 ppm), *m/z* 1715.3153 $[M + 2H]^{2+}$ (calculated, 1715.3122 Da; $\Delta$ 1.82 ppm), *m/z* 1737.3008 $[M + 2Na]^{2+}$ (calculated, 1737.2941 Da; $\Delta$ 3.86 ppm).

## 5.4 In-solution Proteolytic Digestion of Tikitericin

The reaction conditions for each digestion experiment were evaluated using samples of nisin A prior to tikitericin digestion.

### 5.4.1 Asp-N/thermolysin digest

Asp-N (Sequencing grade) was dissolved in digestion buffer (100 mM Tris-HCl solution, pH 8.0) to a final concentration of 2 $\mu$g/mL. LH20-purified tikitericin was suspended in digestion buffer to a concentration of 1 mg/mL. Tikitericin (50 $\mu$L) was added to an equivalent volume of Asp-N solution and incubated at 37 °C. After 50 h, an aliquot of thermolysin (50 $\mu$L, 0.2 mg/mL in 100 mM Tris-HCl, 0.5 mM $CaCl_2$ solution, pH 8.5) was added and the incubation temperature was increased to 70 °C. The reaction was stopped by adding 10% $HCO_2H$ to a final concentration of 0.5% and the peptides were desalted with C18 ZipTips according to the manufacturer's protocol.

### 5.4.2 Thermolysin digest

Thermolysin was dissolved in digestion buffer (100 mM Tris-HCl, 0.5 mM $CaCl_2$ solution, pH 8.5) to a final concentration of 0.2 mg/mL. LH20-purified tikitericin was suspended in digestion buffer to a concentration of 1 mg/mL. An aliquot of tikitericin solution (50 $\mu$L) was added to an equivalent volume of thermolysin solution and incubated at 70 °C for 4 h. The reaction was stopped by adding 10% $HCO_2H$ to a final concentration of 0.5% and the peptides were desalted with C18 ZipTips according to the manufacturer's protocol.

**Fragment i:** M = 1552.6399 Da, *m/z* 777.3282 $[M + 2H]^{2+}$ (calculated, 777.3272 Da; $\Delta$ 1.28 ppm), *m/z* 1553.6422 $[M + H]^{+}$ (calculated, 1553.6471 Da; $\Delta$ – 3.21 ppm).

**Fragment ii:** M = 1232.5278 Da, *m/z* 617.2704 $[M + 2H]^{2+}$ (calculated, 617.2712 Da; $\Delta$ – 1.23 ppm), *m/z* 1233.5301 $[M + H]^{+}$ (calculated, 1233.5351 Da; $\Delta$ – 3.99 ppm).

## 5.5 Tikitericin Linearisation

The reaction conditions for each linearisation experiment were evaluated using samples of nisin A prior to tikitericin linearisation.

### 5.5.1 Oxidative thermal elimination

Nisin (3.8 mg) was suspended in $H_2O$ (2 mL). An aliquot (260 $\mu$L, 0.075 $\mu$mol) was added to a pear-shaped flask and the flask cooled to 0 °C in an ice bath. $NaIO_4$ (2 mg) was dissolved in $H_2O$ (4 mL). An aliquot (50 $\mu$L, 0.46 $\mu$mol) was added to the flask and the reaction brought to ambient temperature and stirred overnight. An aliquot (50 $\mu$L) was removed for analysis and the reaction was then heated to 60 °C for 6 h.

### 5.5.2 Raney Nickel-catalysed reduction

LH-20 purified tikitericin (50 $\mu$g) was suspended in a 8 M guanidine hydrochloride, 20 mM EDTA and 200 mM Tris-HCl solution (300 $\mu$L). After the addition of RaNi (25 mg, active catalyst slurry in $H_2O$), followed by shaking at 55 °C for 15 h and subsequent centrifugation at 5,000 rpm for 2 min, the supernatant was removed and desalted using C18 ZipTips according to the manufacturer's protocol. HR ESI-MS data was obtained with collision energies of 25, 50 and 75 arbitrary units.

**RaNi linearised tikitericin Asn8-Leu35:** M = 2600.5014 Da, *m/z* 867.8403 $[M + 3H]^{3+}$ (calculated, 867.8411 Da; $\Delta$ – 0.84 ppm), *m/z* 1301.2618 $[M + 2H]^{2+}$ (calculated, 1301.2580 Da; $\Delta$ 2.94 ppm).

**RaNi linearised tikitericin Asn9-Leu35:** M = 2486.4584 Da, *m/z* 829.8262 $[M + 3H]^{3+}$ (calculated, 829.8268 Da; $\Delta$ – 0.65 ppm), *m/z* 1244.2360 $[M + 2H]^{2+}$ (calculated, 1244.2365 Da; $\Delta$ – 0.41 ppm), *m/z* 1266.2098 $[M + 2Na]^{2+}$ (calculated, 1266.2184 Da; $\Delta$ – 6.86 ppm).

**RaNi linearised tikitericin Abu10-Leu35:** M = 2372.4155 Da, *m/z* 1187.7077 $[M + 2H]^{2+}$ (calculated, 1187.7165 Da; $\Delta$ – 7.36 ppm), *m/z* 1209.6992 $[M + 2Na]^{2+}$ (calculated, 1209.6984 Da; $\Delta$ 0.64 ppm).

**RaNi linearised tikitericin Ala12-Leu35:** M = 2188.2807 Da, *m/z* 1095.6449 $[M + 2H]^{2+}$ (calculated, 1095.6559 Da; $\Delta$ – 10.06 ppm).

**RaNi linearised tikitericin Val13-Leu35 (33):** M = 2117.2575 Da, *m/z* 706.7598 $[M + 3H]^{3+}$ (calculated, 706.7597 Da; $\Delta$ 0.14 ppm), *m/z* 1059.6369 $[M + 2H]^{2+}$ (calculated, 1059.6359 Da; $\Delta$ 1.00 ppm), *m/z* 1081.6209 $[M + 2Na^{2+}$ (calculated, 1081.6178 Da; $\Delta$ 2.84 ppm), *m/z* 2118.2645 $[M + H]^{+}$ (calculated, 2118.2645 Da; $\Delta$ 0.01 ppm), *m/z* 2140.2574 $[M + Na]^{+}$ (calculated, 2140.2464 Da; $\Delta$ 5.11 ppm).

**RaNi linearised tikitericin Leu14-Leu35:** M = 2018.1888 Da, *m/z* 673.7369 $[M + 3H]^{3+}$ (calculated, 673.7369 Da; $\Delta$ 0.04 ppm), *m/z* 1010.1017 $[M + 2H]^{2+}$ (calculated, 1010.1017 Da; $\Delta$ – 0.01 ppm).

**RaNi linearised tikitericin Ala15-Leu35 (34):** M = 1905.1047 Da, *m/z* 953.5612 $[M + 2H]^{2+}$ (calculated, 953.5596 Da; $\Delta$ 1.64 ppm), *m/z* 975.5445 $[M + 2Na]^{2+}$ (calculated, 975.5416 Da; $\Delta$ 3.01 ppm), 1906.1103 $[M + H]^{+}$ (calculated, 1906.1120 Da; $\Delta$ – 0.9 ppm), *m/z* 1928.0924 $[M + H]^{+}$ (calculated, 1928.0940 Da; $\Delta$ – 0.8 ppm).

### 5.5.3 Base-induced elimination/thiol addition

LH-20 purified tikitericin (50 $\mu$g) was suspended in a 0.24 M NaOH, 0.63 mM 2-mercaptoethanol, 25% EtOH solution (40 $\mu$L) and incubated with shaking at 50 °C for 1 h. The temperature was raised to 85 °C and the reaction incubated for a further 1 h. After the addition of acetic acid (10 $\mu$L) the reaction was diluted with $H_2O$ (150 $\mu$L) and desalted using C18 tips according to the manufacturer's protocol.

**Fully linearised tikitericin (35):** M = 3740.6651 Da, *m/z* 1247.8920 $[M + 3H]^{3+}$ (calculated, 1247.8958 Da; $\Delta$ 3.06 ppm), *m/z* 1871.3369 $[M + 2H]^{2+}$ (calculated, 1871.3400 Da; $\Delta$ – 1.68 ppm), *m/z* 1909.3432 $[M + 2Na]^{2+}$ (calculated, 1909.2959 Da; $\Delta$ 8.04 ppm).

**Triply linearised tikitericin (36):** M = 3662.6516 Da, *m/z* 1221.8911 $[M + 3H]^{3+}$ (calculated, 1221.8911 Da; $\Delta$ – 0.07 ppm), *m/z* 1832.3317 $[M + 2H]^{2+}$ (calculated, 1832.3331 Da; $\Delta$ – 0.74 ppm).

**Doubly linearised tikitericin (37):** M = 3584.6376 Da, *m/z* 1195.8866 $[M + 3H]^{3+}$ (calculated, 1195.8865 Da; $\Delta$ –0.13 ppm), *m/z* 1793.3268 $[M + 2H]^{2+}$ (calculated, 1793.3261 Da; $\Delta$ 0.38 ppm), *m/z* 1815.3174 $[M + 2Na]^{2+}$ (calculated, 1815.3080 Da; $\Delta$ 5.14 ppm).

**Singly linearised tikitericin:** M = 3506.6237 Da, *m/z* 1169.8831 $[M + 3H]^{3+}$ (calculated, 1169.8818 Da; $\Delta$ 1.08 ppm), *m/z* 1192.2031 $[M + 3Na]^{3+}$ (calculated, 1192.1980 Da; $\Delta$ 4.23 ppm), *m/z* 1754.3199 $[M + 2H]^{2+}$ (calculated, 1754.3134 Da; $\Delta$ 0.46 ppm), *m/z* 1776.3016 $[M + 2Na]^{2+}$ (calculated, 1776.3011 Da; $\Delta$ 0.31 ppm).

## 5.6 Determination of (Me)Lan Residue Stereochemistry

### 5.6.1 Preparation of derivatised (Me)Lan residues

Nisin A (300 $\mu$g) was dissolved in 6 M HCl (3 mL) and heated with stirring at 110 °C in a sealed tube for 24 h. The reaction was cooled and solvent removed *in vacuo*. Methanol (5 mL) was cooled in an ice-water bath and acetyl chloride (1.5 mL) was added dropwise.

This solution was added to the hydrolysed nisin A and heated at 110 °C for 45 min. The sample was cooled and dried *in vacuo*. $CH_2Cl_2$ (3 mL) and pentafluoropropionic anhydride (1 mL) were added and the material was heated at 110 °C for 15 min, then allowed to cool and dried under a stream of argon. The hydrolysed and derivatised product was taken up in MeOH, transferred to a sample vial and dried under a stream of argon. The presence of derivatised (Me)Lan compounds was confirmed by HR ESI-MS. HPLC purified tikitericin (300 $\mu$g) was derivatised using the procedure described above, and the presence of derivatised (Me)Lan compounds confirmed by HR ESI-MS.

**Derivatised Lan, nisin:** M = 528.0413 Da, *m/z* 529.0484 $[M + H]^+$ (calculated, 529.0486 Da; $\Delta$ – 0.41 ppm), *m/z* 546.0755 $[M + NH_4]^+$ (calculated, 546.0751 Da; $\Delta$ 0.75 ppm), *m/z* 551.0305 $[M + Na]^+$ (calculated, 551.0305 Da; $\Delta$ – 0.03 ppm), *m/z* 567.0016 $[M + K]^+$ (calculated, 567.0044 Da; $\Delta$ – 5.07 ppm), *m/z* 1079.0738 $[2M + Na]^+$ (calculated, 1079.0718 Da; $\Delta$ – 1.87 ppm).

**Derivatised MeLan, nisin:** M = 542.0569 Da, *m/z* 543.0642 $[M + H]^+$ (calculated, 543.0642 Da; $\Delta$ 0.03 ppm), *m/z* 560.0912 $[M + NH_4]^+$ (calculated, 560.0908 Da; $\Delta$ 0.69 ppm), *m/z* 565.0463 $[M + Na]^+$ (calculated, 565.0462 Da; $\Delta$ 0.17 ppm), *m/z* 581.0195 $[M + K]^+$ (calculated, 581.0201 Da; $\Delta$ – 1.06 ppm), *m/z* 1107.1056 $[2M + Na]^+$ (calculated, 1107.1031 Da; $\Delta$ 2.26 ppm).

**Derivatised Lan, tikitericin:** M = 528.0413 Da, *m/z* 529.0554 $[M + H]^+$ (calculated, 529.0486 Da; $\Delta$ 12.92 ppm), *m/z* 546.0754 $[M + NH_4]^+$ (calculated, 546.0751 Da; $\Delta$ 0.53 ppm), *m/z* 551.0304 $[M + Na]^+$ (calculated, 551.0305 Da; $\Delta$ – 0.15 ppm), *m/z* 567.0043 $[M + K]^+$ (calculated, 567.0044 Da; $\Delta$ – 0.22 ppm).

**Derivatised MeLan, tikitericin:** M = 542.0569 Da, *m/z* 560.0905 $[M + NH_4]^+$ (calculated, 560.0908 Da; $\Delta$ – 0.54 ppm), *m/z* 565.0456 $[M + Na]^+$ (calculated, 565.0462 Da; $\Delta$ – 0.97 ppm), *m/z* 581.0161 $[M + K]^+$ (calculated, 581.0201 Da; $\Delta$ – 1.76 ppm).

### 5.6.2 Determination of MeLan stereochemistry by GCMS

The derivatised amino acids from tikitericin, nisin A, as well as synthetic MeLan standards (provided by J. C. Vederas) were analysed using a Shimadzu QP2010 Plus GCMS equipped with an HP-VOC column 60 m x 0.32 mm i.d., 0.18 $\mu$m (Agilent Technologies). The samples were dissolved in EtOAc and introduced to the instrument via a splitless injection at an inlet temperature of 270 °C and a linear velocity of 40 cm/sec using He as the carrier gas. The temperature programme was held at 200 °C for 1 h, increased to 280 °C at 25 °C/min, and held at 280 °C for 10 min. The MS was operated in SIM mode, monitoring fragment masses of 248 and 379 Da.

**Tikitericin RT:** 29.61 min.

**Nisin RT:** 29.60 min.

**(2*S*,3*S*,6*R*)-MeLan RT:** 29.61 min (M).

**(2*R*,3*S*,6*R*)-MeLan RT:** 30.17 min (M), 31.32 min (m).

**(2*S*,3*R*,6*R*)-MeLan RT:** 31.41 min (M), 30.28 min (m).

**(2*R*,3*R*,6*R*)-MeLan RT:** 30.40 min (M), 29.79 min (m), 31.41 min (m).

**Tikitericin + (2*S*,3*S*,6*R*)-MeLan RT:** 29.63 (M).

**Tikitericin + (2*R*,3*S*,6*R*)-MeLan RT:** 29.61 min (M), 30.05 min (m).

**Tikitericin + (2*S*,3*R*,6*R*)-MeLan RT:** 29.59 min (M), 31.15 min (m).

**Tikitericin + (2*R*,3*R*,6*R*)-MeLan RT:** 30.18 min (M), 29.59 min (m).

### 5.6.3 Determination of Lan stereochemistry by GCMS

The derivatised amino acids from tikitericin, nisin A, as well as synthetic Lan standards (provided by J. C. Vederas) were analysed using a Shimadzu QP2010 Plus GCMS. Two chiral stationary phases were employed: a CHIRALDEX B-DM column 30 m x 0.25 mm i.d., 0.12 $\mu$m (Astec®, Sigma-Aldrich), and a CP-Chirasil-Val column 25 m x 0.25 mm i.d., 0.12 $\mu$m (Agilent Technologies). An HP-VOC column 60 m x 0.32 mm i.d., 0.18 $\mu$m (Agilent Technologies) was also employed. The samples were dissolved in EtOAc and introduced to the instrument via a splitless injection at an inlet temperature of 270 °C and a linear velocity of 40 cm/sec using He as the carrier gas. The temperature programme was increased from 150 °C to 205 °C at a rate of 10 °C/min, held at 205 °C for 15 min, increased to 250 °C at 3 °C/min, increased to 280 °C at 30 °C/min and held at 280 °C for 5 min. The MS was operated in SIM mode, monitoring fragment masses of 248 and 365 Da.

**Tikitericin RT:** 23.67 min.

**Nisin RT:** 23.72 min.

**(2*S*,6*R*)-Lan RT:** 23.67 min.

**(2*R*,6*R*)-Lan RT:** 23.61 min.

**(2*S*,6*R*)-Lan + (2*R*,6*R*)-Lan RT:** 23.64 min.

**Tikitericin + (2*S*,6*R*)-Lan RT:** 23.66 min.

**Tikitericin + (2*R*,6*R*)-Lan RT:** 23.64.

### 5.6.4 Determination of Lan stereochemistry by HR ESI-MS

The derivatised amino acids from tikitericin, nisin A, as well as synthetic Lan standards were analysed by HR ESI-MS using a Zorbax C18 column 100 mm x 4.6 mm i.d., 3.5 $\mu$m (Agilent Technologies). The binary pump was operated with a flow rate of 0.4 mL/min and a solvent system of eluent A: 99.9% $H_2O$/0.1% $NH_4HCO_2$, and eluent B: 99.9% MeCN/0.1% $HCO_2H$. The solvent was held at 40% B for 5 min, increased to 100% B over 30 sec, held at 100% B for 5 min, returned to 40% B over 1.5 min and held for 2 min.

**Tikitericin RT:** 4.34 min.

**Nisin RT:** 4.34 min.

**(2*S*,6*R*)-Lan RT:** 4.34 min.

**(2*R*,6*R*)-Lan RT:** 4.27 min.

**(2*S*,6*R*)-Lan + (2*R*,6*R*)-Lan RT:** 4.32 min.

**Tikitericin + (2*S*,6*R*)-Lan RT:** 4.35 min.

**Tikitericin + (2*R*,6*R*)-Lan RT:** 4.30 min.

## 5.7 NMR Quantification of Tikitericin

An internal standard of $CH_3NO_2$ was prepared to a concentration of 19 mM (10 $\mu$L $CH_3NO_2$ in 10 mL $CDCl_3$), 5 $\mu$L of this was added to tikitericin in $d_6$-DMSO (200 $\mu$L) and the solvent transferred to a 3 mm NMR tube. A $^1$H NMR experiment was performed with the following parameters: spectral width, 12 – -2 ppm; scans, 1024; acquisition time, 8 sec; d1, 15; $H_2O$ suppression, $\delta$ 3.3 ppm.

## 5.8 Biological Activity of Strain T81 Fractions

The following fractions were tested in a bacteriostatic assay against *S. aureus* and *E. coli* at a starting concentration of 400 $\mu$g/mL: 75% $Me_2CO$ fraction off HP20, 2 L strain T81

liquid culture (filtrate); 75% $Me_2CO$ fraction off HP20, 2 L strain T81 liquid culture (cell mass); 75% $Me_2CO$ fraction off HP20, 16 L strain T81 liquid culture (filtrate); 75% $Me_2CO$ fraction off HP20, 2 L strain T81/TKA04.11 liquid co-culture (filtrate).

HPLC-purified Tikitericin was tested in a bacterostatic assay against *S. aureus* at a starting concentration of $\approx 2.5$ $\mu$g/mL.

## 5.9 Computer Modelling of Tikitericin

The three-dimensional structure of tikitericin was modelled by Monte Carlo methods using Maestro software (Schrödinger). The tikitericin model was constructed with the natural stereochemical configurations for (Me)Lan residues and with charged residues in their zwitterionic form. An energy minimisation was performed using the following parameters: solvent, water; forcefield, OPLS_2005; PRCG; number of iterations, 5000; convergence, 0.005. A mixed torsional/low mode conformational search was performed with the following parameters: steps, 10,000; energy window, 5 kJ/mol.

# Appendix A

# Strains of Bacteria and Fungi (GNS Science)

**Table A.1.** Strains of bacteria and fungi.

| **Strain** | **Code** | **Description** | **Source** | Inhibited by strain T81 (Y/N) |
|---|---|---|---|---|
| *Thermogemmatispora* strain T81 | T81 | Thermophilic bacteria | Tikitere | |
| *Geobacillus thermoleovorans* (100%) | MOK 14.5 | Thermophilic bacteria | Taupo | N |
| *Bacillus smithii* | MYA 1.7 | Thermophilic bacteria | Taupo | N |
| *Thermoactinomyces* | MYA 1.9 | Thermophilic bacteria | Taupo | N |
| *Thermoactinomyces* | MYA 1.10 | Thermophilic bacteria | Taupo | N |
| *Thermomyceslanuginosus* | MYA 1.12 | Thermophilic fungi | Taupo | N |
| *Bacillus* sp. 2n | MYA 1.21 | Thermophilic bacteria | Taupo | N |
| *Thermoactinomyces* sp. JAM-FM1 | MYA 1.23 | Thermophilic bacteria | Taupo | N |
| *Alicyclobacillus* sp. (94.5%) | TKA 2.1 | Thermophilic bacteria | TeKopia | Y |
| *Alicyclobacillus* sp. (99%) | TKA 3.1 | Thermophilic bacteria | TeKopia | Y |
| *Alicyclobacillus* sp. (98.7%) | TKA 4.11 | Thermophilic bacteria | TeKopia | Y |
| *Alicyclobacillus acidocaldarius* (99.5%) | TKA 4.12 | Thermophilic bacteria | TeKopia | Y |
| *Geobacillus vulcani* (98.8%) | TKA 4.13 | Thermophilic bacteria | TeKopia | N |
| *Alicyclobacillus acidocaldarius* (98.9%) | TKA 4.14 | Thermophilic bacteria | TeKopia | Y |
| *Alicyclobacillus acidocaldarius* (99.3%) | TKA 4.15 | Thermophilic bacteria | TeKopia | Y |
| *Microbisporarosea* (98%) | TKT 20.44 | Thermophilic bacteria | Tikitere | N |
| Candidate division OP10 bacterium T49 (99.4%) | TKT 38.2 | Thermophilic bacteria | Tikitere | N |
| *Chloroflexi* bacterium T104 (99.2%) | TKT 38.14 | Thermophilic bacteria | Tikitere | N |
| *Geobacillus* sp. CICC 10315 (97.7%) | TKT 38.20 | Thermophilic bacteria | Tikitere | N |
| *Bacillus fumarioli* (99%) | WKT 16.21 | Thermophilic bacteria | Waikite | Y |
| *Alicyclobacillus pomorum* (94%) | WKT 17.14 | Thermophilic bacteria | Waikite | Y |
| *Bacillus fumarioli* (99%) | WKT 17.15 | Thermophilic bacteria | Waikite | N |
| *Bacillus fumarioli* (99%) | WKT 17.25 | Thermophilic bacteria | Waikite | Y |
| *Alicyclobacillus tengcongensis* (95%) | WKT 21.8 | Thermophilic bacteria | Waikite | Y |
| *Bacillus fumarioli* (99%) | WKT 22.5 | Thermophilic bacteria | Waikite | Y |
| *Alicyclobacillus acidocaldarius* (98%) | WKT 22.10 | Thermophilic bacteria | Waikite | Y |
| *Thermusbrockianus* (98%) | WKT 32.35 | Thermophilic bacteria | Waikite | N |
| *Geobacillus vulcani* (100%) | WRG 1.1 | Thermophilic bacteria | Wairakei | N |
| *Geobacillus stearothermophilus* (99%) | XKA 55.2 | Thermophilic bacteria | Kermadec hydrothermal vents | N |
| *Geobacillus stearothermophilus* (99%) | XKA 55.3 | Thermophilic bacteria | Kermadec hydrothermal vents | N |
| *Geobacillus stearothermophilus* (99%) | XKA 55.5 | Thermophilic bacteria | Kermadec hydrothermal vents | N |
| *Geobacillus stearothermophilus* (99%) | XKA 55.6 | Thermophilic bacteria | Kermadec hydrothermal vents | N |
| *Geobacillus stearothermophilus* (99%) | XKA 55.7 | Thermophilic bacteria | Kermadec hydrothermal vents | N |
| *Geobacillus stearothermophilus* (99%) | XKA 100.2 | Thermophilic bacteria | Kermadec hydrothermal vents | N |
| Unknown strain | XS01.96 | Thermophilic bacteria | ? | Y |

# Appendix B

# *Thermogemmatispora* Strain T81 (GNS Science)

## B.1 *Thermogemmatispora* Strain T81

*Thermogemmatispora* strain T81 was isolated from steam-affected geothermal soil (pH 4.5, 55 °C) in Tikitere, New Zealand,[57] and was recovered on plates of AOM1 solid medium incubated aerobically at 60 °C. Colonies were isolated through successive rounds of colony picking and plating, and isolate purity was confirmed by 16S rRNA gene sequencing using the universal primers 9F and 1492R34.[129] Strain T81 was able to grow on this media at 45 – 70 °C, but not at 37 °C or 75 °C. The optimum pH was 4.8 – 6.5; growth occurred from pH 3.1 – 7.1, but no growth was observed $\leq$ 2.6, or $\geq$ 7.4. Growth at 60 °C was observed on the commercial media R2A (Difco) and NB nutrient broth (Sigma) at pH 4.5 and 6.5, but not on Luria-Bertani broth or tryptic soy broth (Difco) at pH 4.5 or pH 6.5. Strain T81 has been deposited at Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, MascheroderWeg 1b, D-38124, Braunschweig, Germany, Accession No. DSM 21103 (deposited 30 January 2008, patent deposit).

*Thermogemmatispora* strain T81 has the following identifying characteristics: Gram-positive, filamentous bacteria with branching hyphae and cells of 0.2 – 0.4 $\mu$m in width and $\geq$ 30 $\mu$m in length, has white vegetative mycelia with yellow under-layer aerobe, pH growth range 4.0 – 7.5, at temperatures between 40 °C – 75 °C, growth on gellan/phytagel, agar, carboxymethylcellulose, xylan, pectin, xanthan, Avicell® (crystalline cellulose), sodium alginate, R2A, and nutrient broth. The major fatty acids were *i*18:0, *i*19:0, and 12,17-di-Me 18:0.

## B.2 Growth Conditions

Plated cultured were prepared by inoculating petri dishes containing 10 mL of AOM1 media supplemented with 2% Luria broth (LB, Sigma-Aldrich) with T81 colonies and incubating at 60 °C for 14 d. Liquid cultures were prepared by inoculating 2 L AOM1 media supplemented with 2% LB broth and incubating with shaking at 60 °C for 7 d. For co-culturing competitive assays, species of bacterial and fungal organisms were inoculated into liquid AOM1 + 2% LB and allowed to grow to an optical density (OD) of 0.4 - 0.6. All species were incubated at 60 °C and took approximately 14-18 h to reach an OD of 0.4 - 0.6. T81 was streaked for isolated colonies on AOM1 + 2% LB and incubated for 96 h at 60 °C. In duplicate, each strain of bacteria or fungi was co-cultured with T81 by pipetting the bacteria (or fungi) culture onto the incubated T81 plates to cover, 1 mL with excess liquid removed. The plates were incubated at 60 °C for a further 16 h.

**AOM1 media composition**

$(NH_4)_2SO_4$, 4.0 g/L; $K_2HPO_4$, 0.45 g/L; $MgSO_4 \cdot 7\,H_2O$, 0.60 g/L; $MgCl_2 \cdot 6\,H_2O$, 1.0 g/L; $CaCl_2 \cdot 2\,H_2O$, 0.005 g/L; FeEDTA solution, 3 mL/L; Trace element solution (Methanotrophs), 3 mL/L; Methanogen trace metal solution (Wolin), 1 mL/L; Phytagel 15.0 g/L; pH 6.5, pH adjusted with 1 M $H_2SO_4$. Wolin trace metal solution: nitrilotriacetic acid, 1.5 g/L; $Fe(NH_4)_2(SO_4)_2 \cdot 6\,H_2O$, 0.2 g/L; $Na_2SeO_4 \cdot 10\,H_2O$, 0.44 g/L; $CoCl_2 \cdot 6\,H_2O$, 0.1 g/L; $MnSO_4 \cdot 4\,H_2O$, 0.12 g/L; $Na_2MoO_4 \cdot 2\,H_2O$, 0.1 g/L; $NaWO_4 \cdot 2\,H_2O$, 0.1 g/L; $ZnSO_4 \cdot 7\,H_2O$, 0.1 g/L; $AlCl_3 \cdot 6\,H_2O$, 0.04 g/L; $NiCl_2 \cdot 6\,H_2O$, 0.025 g/L; $H_3BO_3$, 0.01 g/L; $CuSO_4 \cdot 5\,H_2O$, 0.01 g/L; pH 6.5, pH adjusted with KOH. Trace metal solution for methanogens: $ZnSO_4 \cdot 7\,H_2O$, 0.44 g/L; $Na_2MoO_4 \cdot 2\,H_2O$, 0.60 g/L; $MnCl_4H_2O$, 0.19 g/L; $CuSO_4 \cdot 5\,H_2O$, 0.06 g/L; $H_3BO_3$, 0.10 g/L; $CoCl_2 \cdot 6\,H_2O$, 0.08 g/L. FeEDTA: $FeSO_4 \cdot 7\,H_2O$, 1.54 g/L; NaEDTA, 2.06 g/L.

## B.3 Genome Sequencing

The genome of *Thermogemmatispora* strain T81 was sequenced after DNA extraction via paired-end 454 pyrosequencing and assembled using Newbler. The genome is approximately 5.9 MB and has a %G+C content of $\approx$ 60.84%. Annotation of the draft genome using IMG/ER identified a gene cluster encoding for genes putatively encoding for lanthionine synthetase and associated lanthipeptide biosynthesis genes.

## B.4 Preliminary Investigation of Lanthipeptide Production

In competitive growth assays, strain T81 was able to inhibit the following bacterial strains (Figures 1.7a, 1.7b): TKA 2.1, TKA 3.1, TKA 4.11, TKA 4.12, TKA 4.14, TKA 4.15, WKT 16.21, WKT 17.14, WKT 17.25, WKT 21.8, WKT 22.5, WKT 22.10. T81 was unable to inhibit the following bacterial strains: MOK 14.5, MYA 1.7, MYA 1.9, MYA 1.10, MYA 1.12, MYA 1.21, MYA 1.23, TKA 4.13, TKT 20.44, TKT 38.2, TKT 38.14, TKT 38.20, WKT 17.15, WKT 32.35, WRG 1.1, XKA 55.2, XKA 55.3, XKA 55.5, XKA 55.6, XKA 55.7, XKA 100.2.

# Appendix C

# Cyclic Loading Protocol

The technique of cyclic loading was developed for extracts of marine sponges by Northcote and West and was adapted during this research to fractionate crude bacterial extracts. The technique allows undesirable highly polar growth media salts and carbohydrates, and non-polar fats to be separated from the intermediate polarity secondary metabolites of interest. PSDVB (poly(styrene-divinylbenzene)) cross-linked polymeric resin beads are employed as the stationary phase. PSDVB is a macro-porous, rigid resin which is chemically inert and stable across a large pH range, and can be reused a number of times without decreased performance. The beads are devoid of polar functional groups so irreversible binding or degradation of polar substrates has not been observed.

The crude (MeOH) extract is passed through a PSDVB column, allowing adhesion of non-polar metabolites within the extract to the column. The eluent is diluted 1:1 with $H_2O$ and re-cycled through the column iteratively until all metabolites of interest have adsorbed, typically with 12.5% or 25% MeOH concentration. Increasing eluent polarity facilitates adsorption of increasingly polar metabolites to the stationary phase, in effect the opposite of conventional chromatography. After adsorption of the desired material to the column, it is eluted with mixtures of $H_2O$ and an organic modifier of decreasing polarity, commonly 30%, 75% and 100% $Me_2CO$ in $H_2O$.

The 30% and 75% $Me_2CO$ in $H_2O$ fractions are usually backloaded onto the screening column using the same cyclic technique, the column is air dried and then eluted with organic solvent. This generates fractions devoid of $H_2O$ which can be easily evaporated under reduced pressure. In this research the mixed $Me_2CO/H_2O$ fractions were evaporated directly under reduced pressure.

This laboratory uses water-miscible solvents such as MeOH and $Me_2CO$, which are cheaper and considerably more environmentally friendly than the halogenated solvents frequently used in other approaches such as liquid-liquid partitioning. Large quantities of eluent are generated during the procedure, for example, a 100-plate workup of bacteria produces 8 L of eluent. This is a major disadvantage of the technique.

# Appendix D

# MALDI-TOF MS Spectra

## D.1 Culture Media

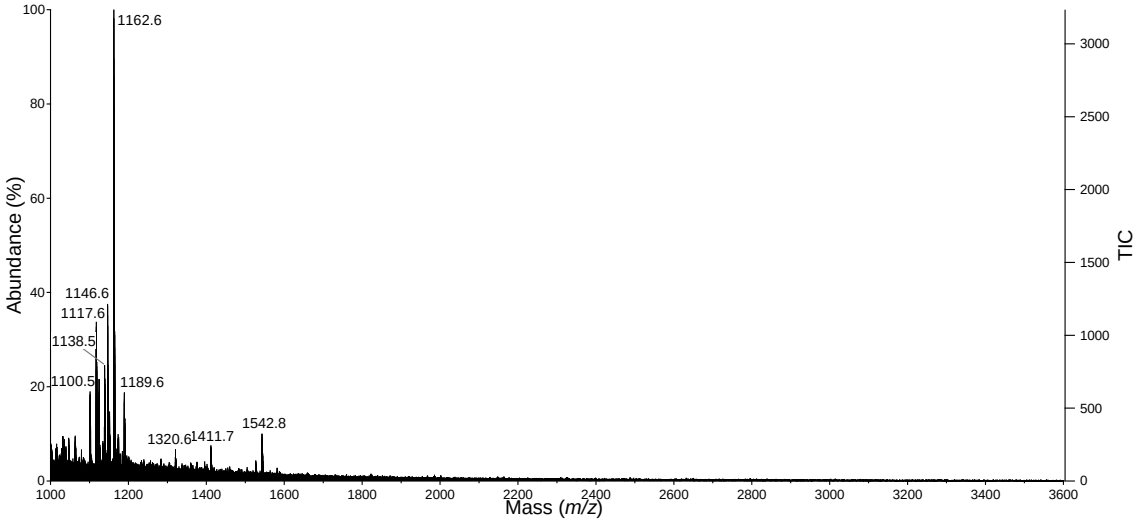
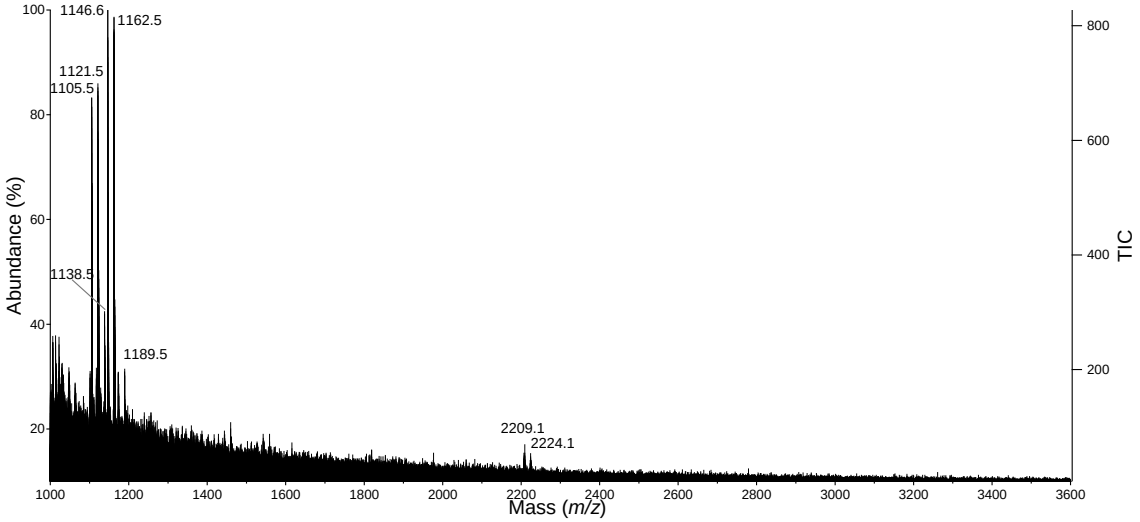
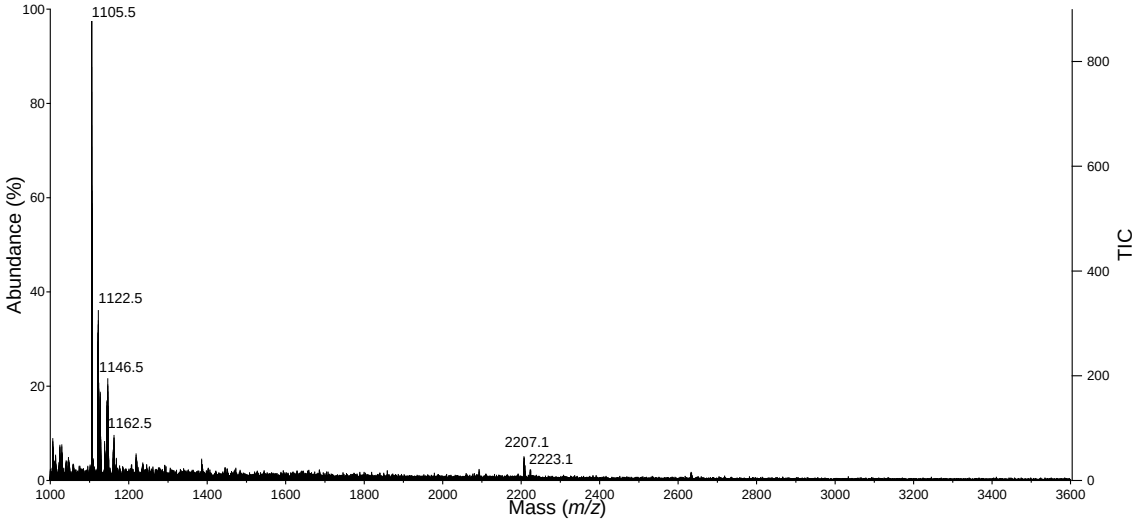



**Figure D.1.** MALDI-TOF MS spectrum of solid AOM1 media.



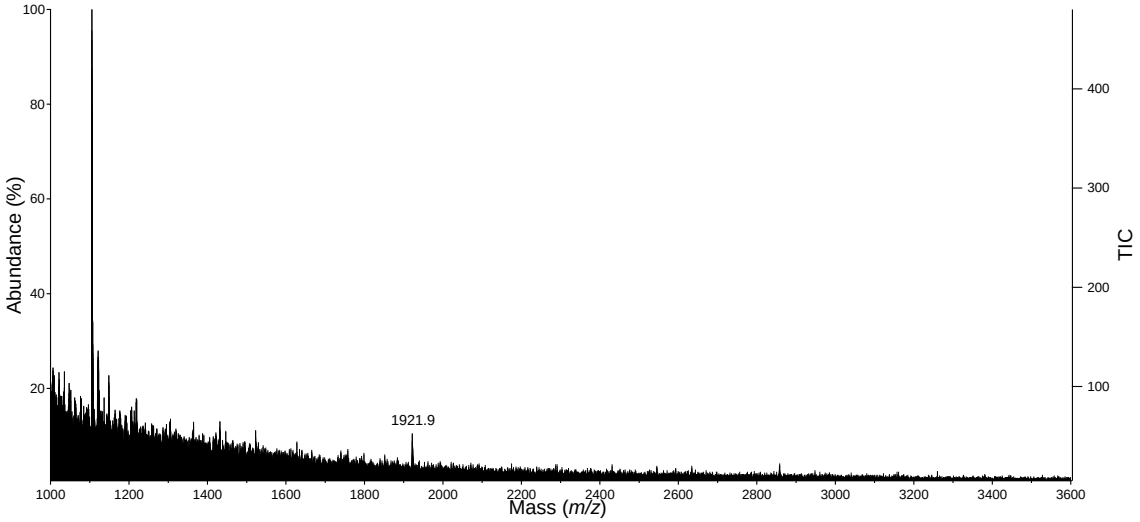

**(a)** 30% $Me_2CO$ fraction.



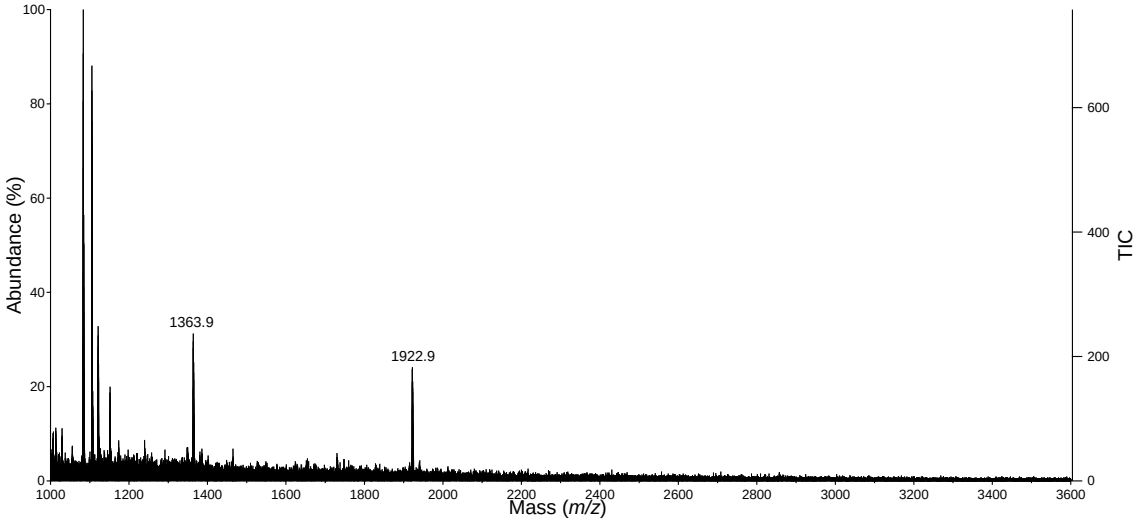

**(b)** 75% $Me_2CO$ fraction.

**Figure D.2.** MALDI-TOF MS spectra of $Me_2CO$ fractions off HP20 from liquid AOM1 media.

## D.2 Strain T81 Liquid Cultures



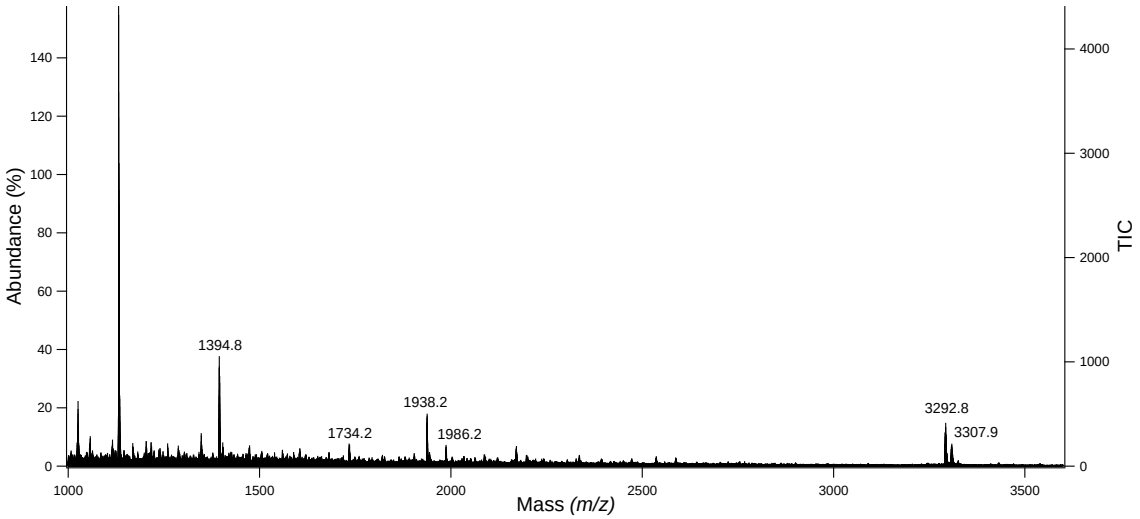

**(a)** 30% $Me_2CO$ fraction.



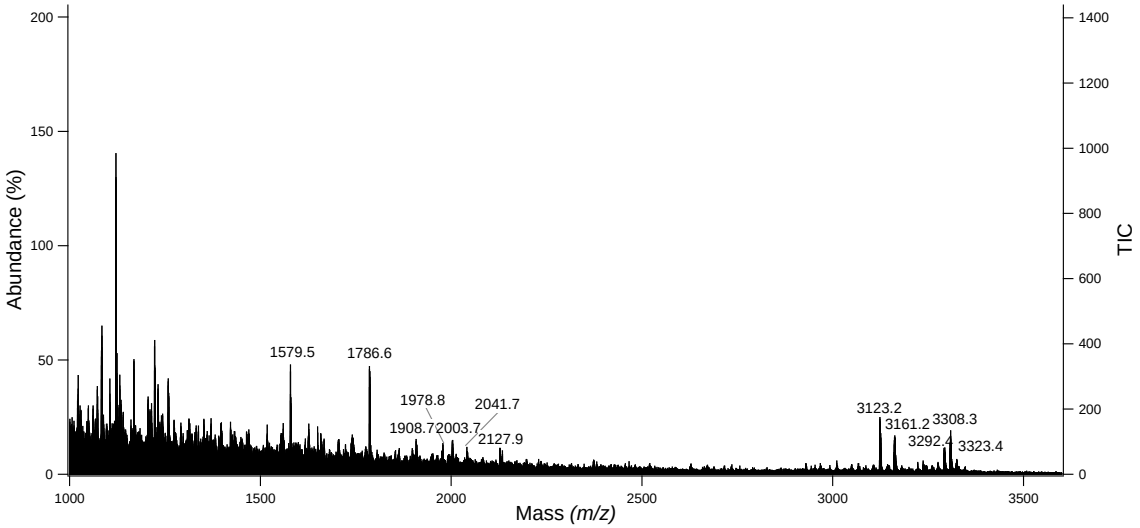

**(b)** 75% $Me_2CO$ fraction.

**Figure D.3.** MALDI-TOF MS spectra of $Me_2CO$ fractions off HP20 from a liquid coculture of strain T81 with strain TKA 04.11.

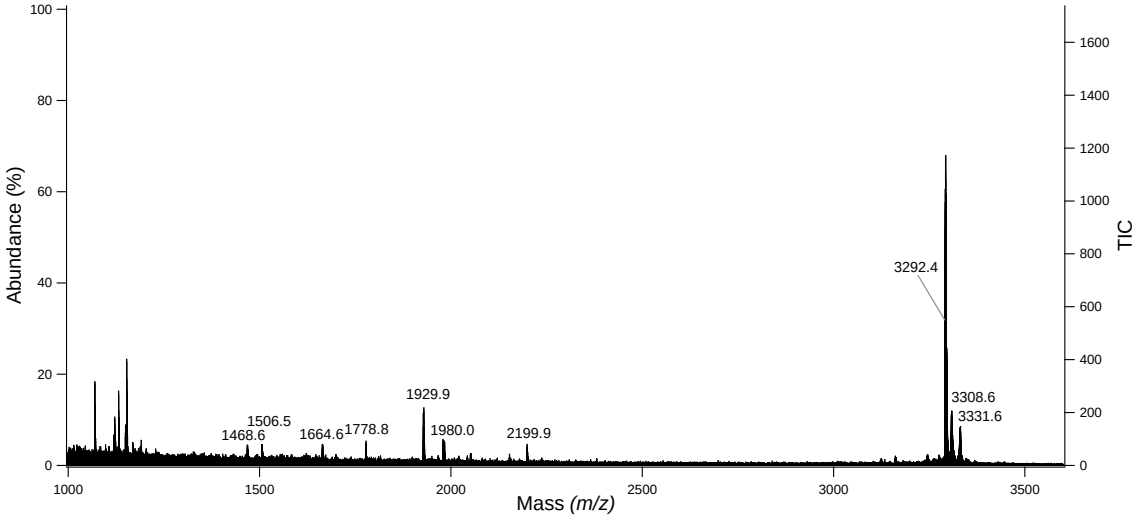
## D.3 Strain T81 Competition Experiments




**(a)** Strain T81.

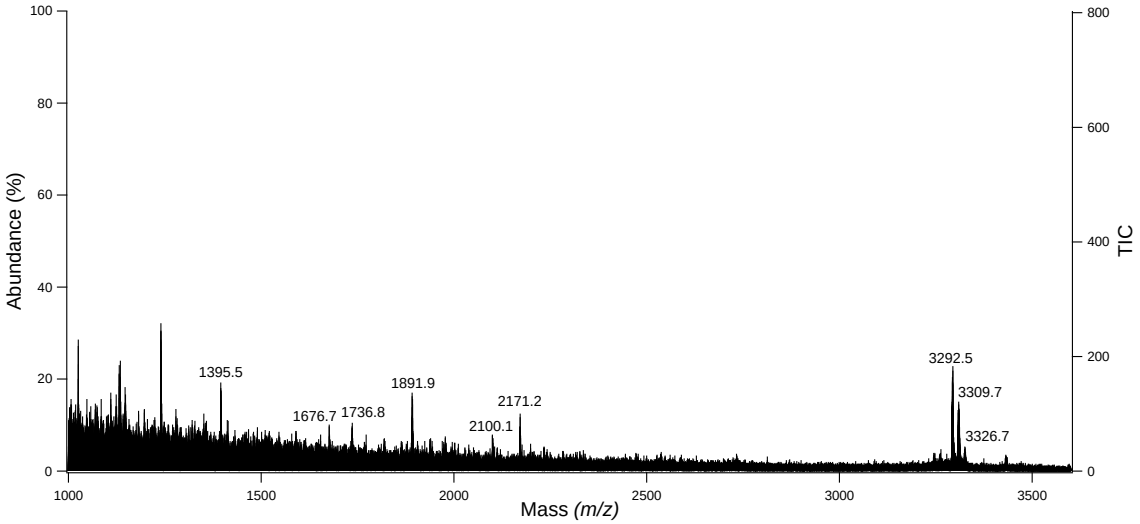



**(b)** Zone of Inhibition.

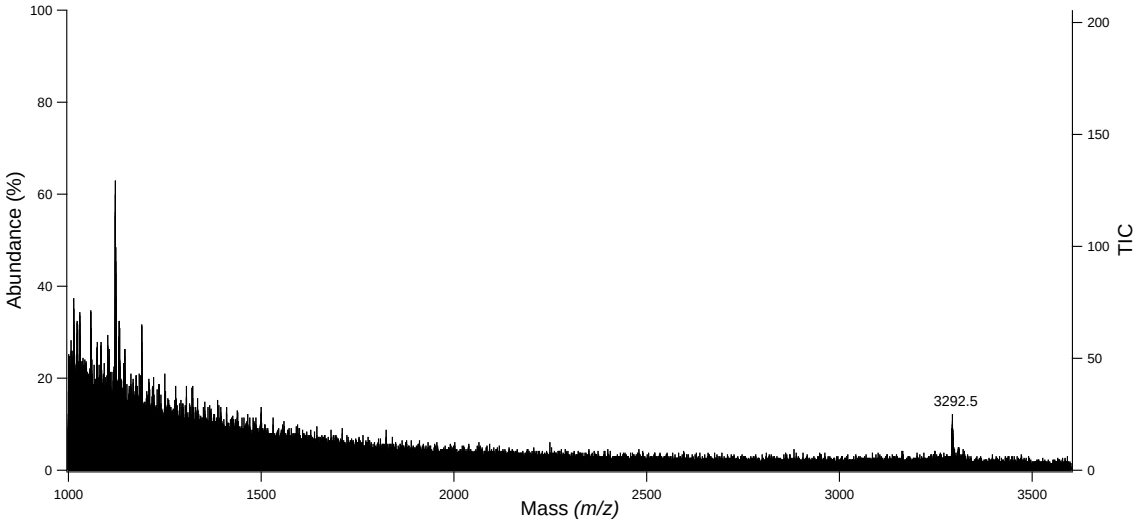



**(c)** Strain TKA 04.11.

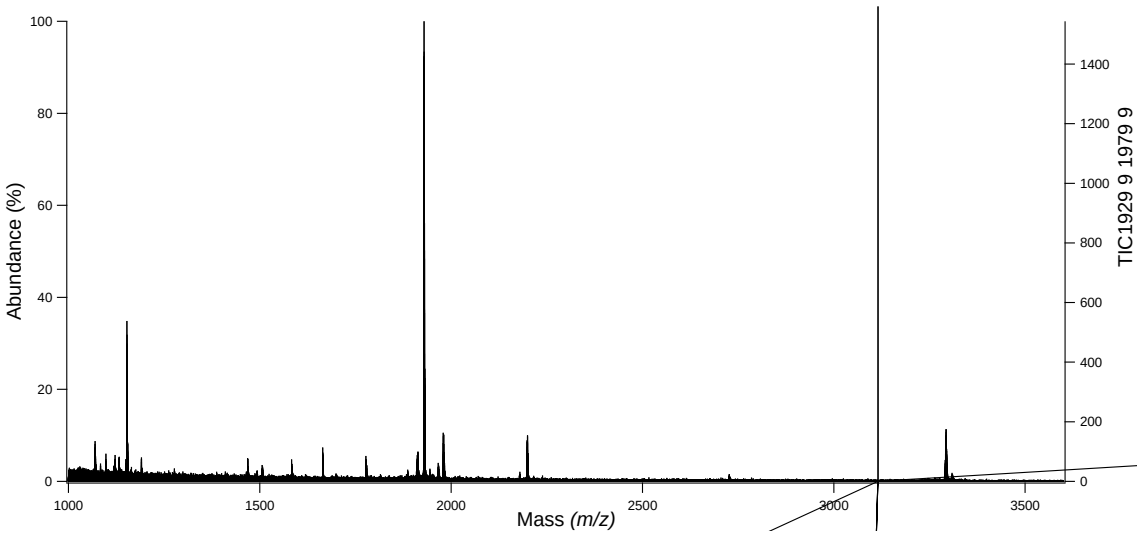
**Figure D.4.** Whole-cell MALDI-TOF MS spectra of a strain T81/TKA 04.11 coculture.




**(a)** Strain T81.

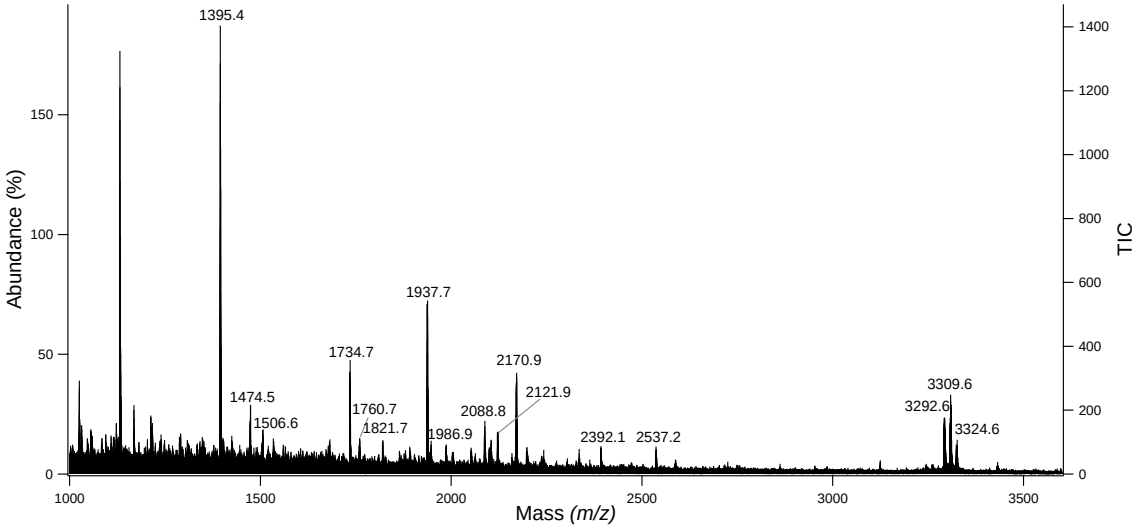



**(b)** Zone of Inhibition.

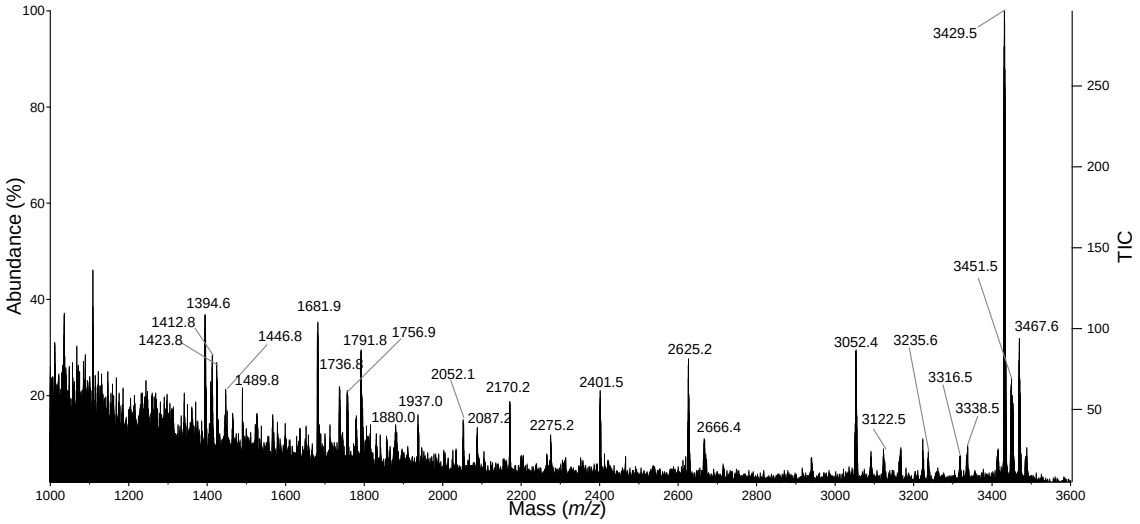



**(c)** Strain TKA 04.12.

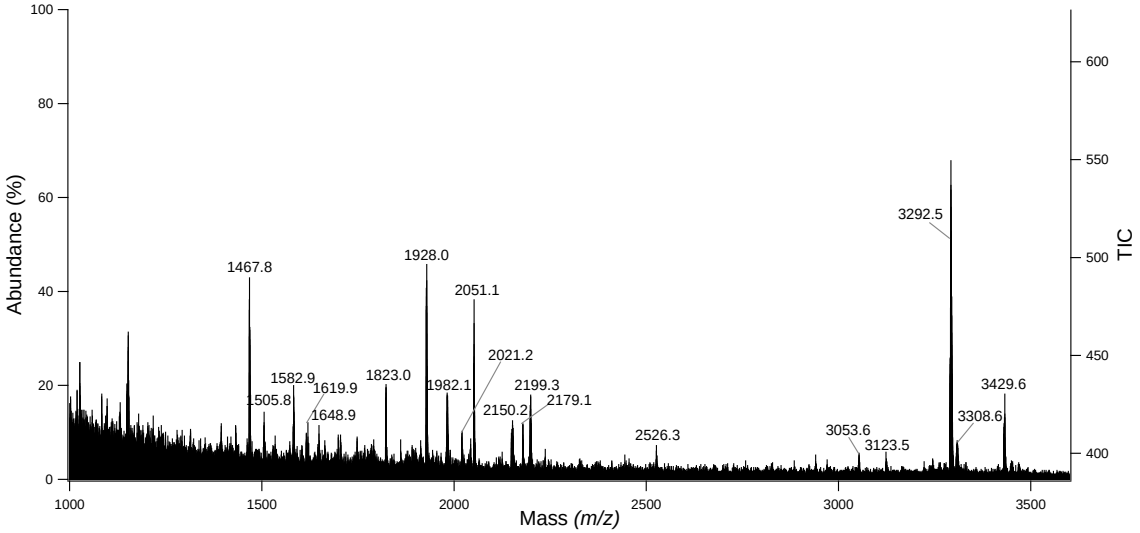
**Figure D.5.** Whole-cell MALDI-TOF MS spectra of a strain T81/TKA 04.12 coculture.




**(a)** Strain T81.

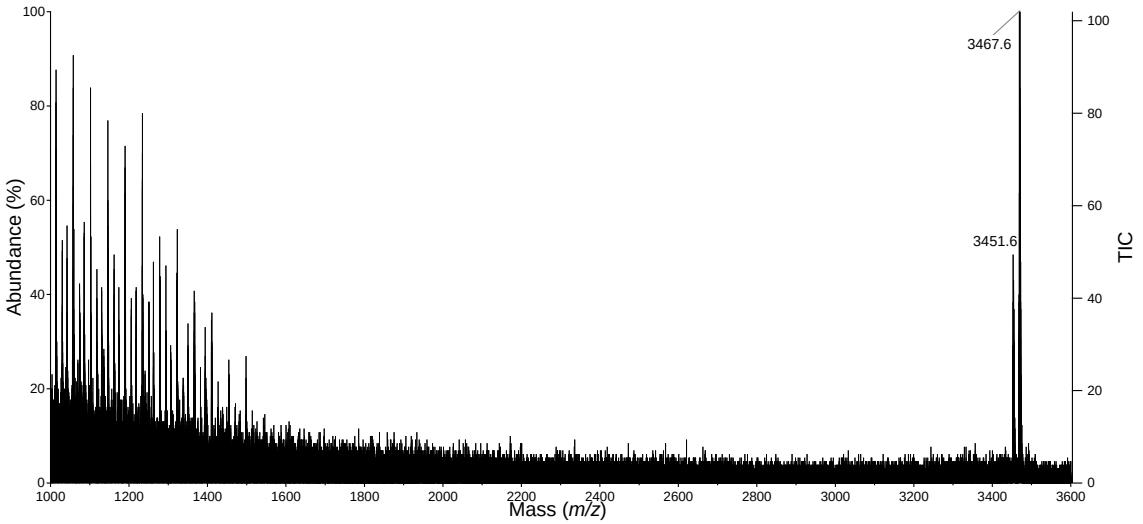



**(b)** Zone of Inhibition.

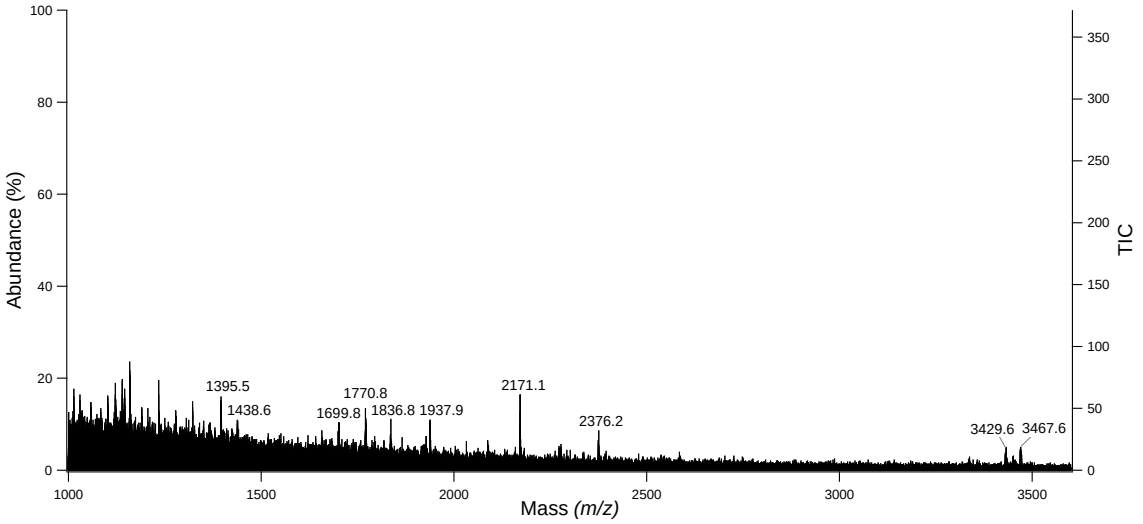



**(c)** Strain XS 01.96.

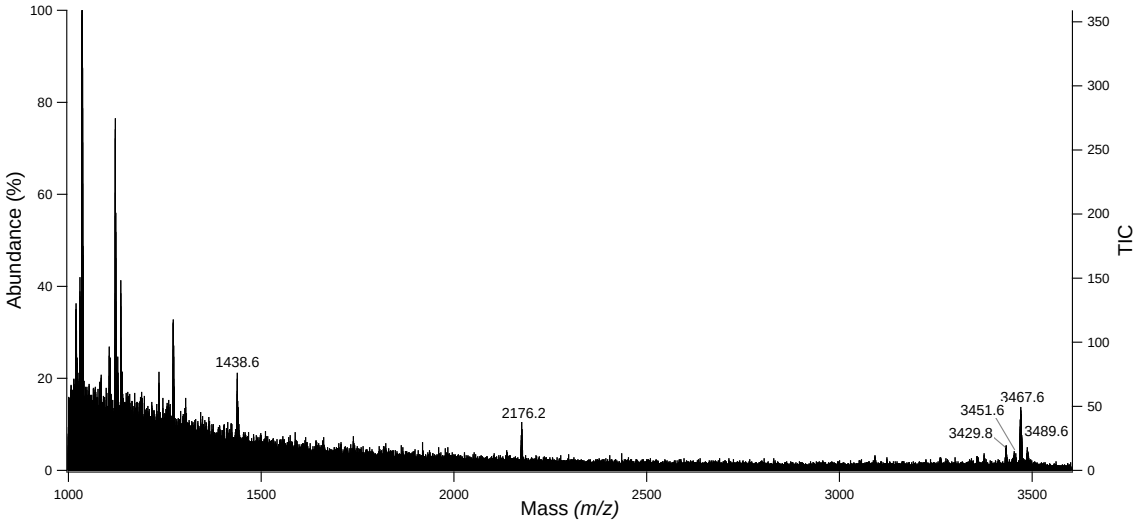
**Figure D.6.** Whole-cell MALDI-TOF MS spectra of a strain T81/XS 01.96 coculture.




**(a)** Strain T81.

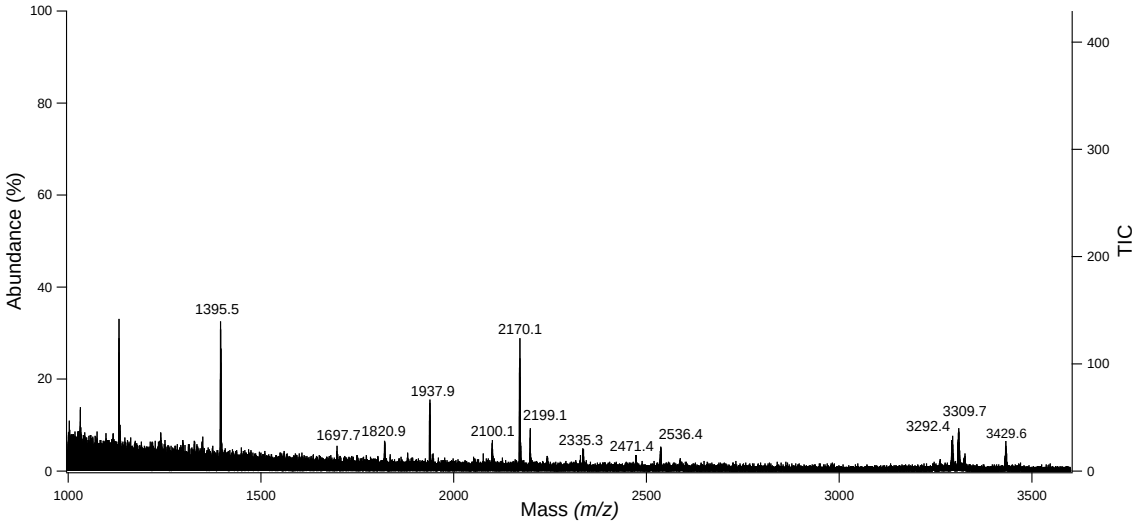



**(b)** Zone of Inhibition.

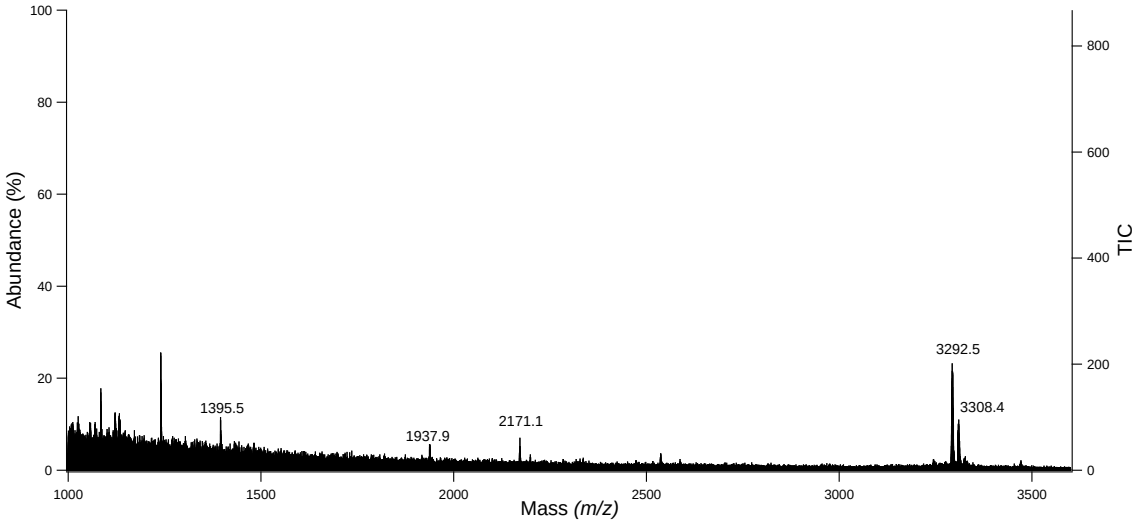



**(c)** Strain WKT 21.8.

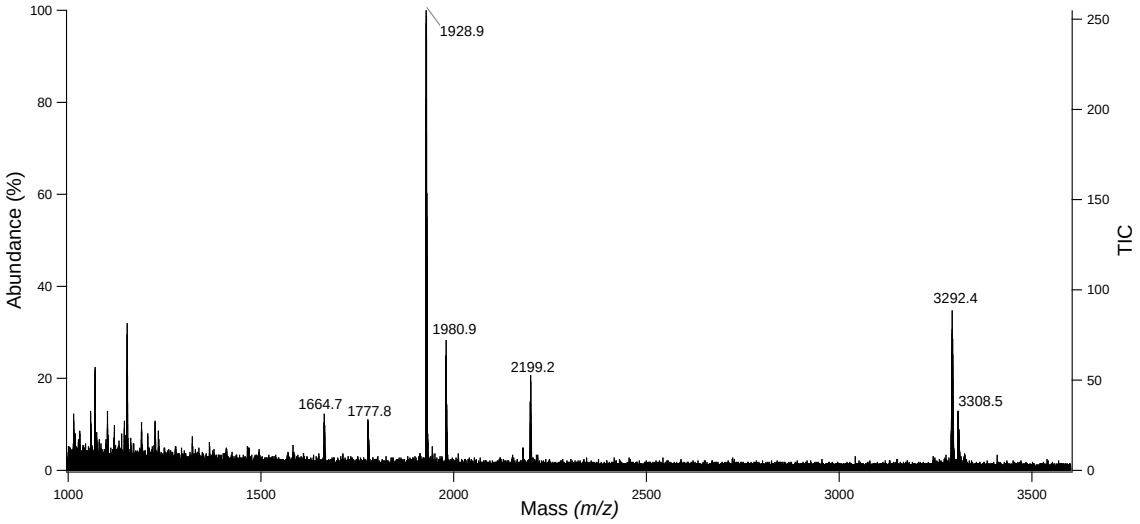
**Figure D.7.** Whole-cell MALDI-TOF MS spectra of a strain T81/WKT 21.8 coculture.




**(a)** Strain T81.




**(b)** Zone of Inhibition.




**(c)** Strain WRG 1.1.

**Figure D.8.** Whole-cell MALDI-TOF MS spectra of a strain T81/WRG 1.1 coculture.

# D.4 Linearisation Experiments

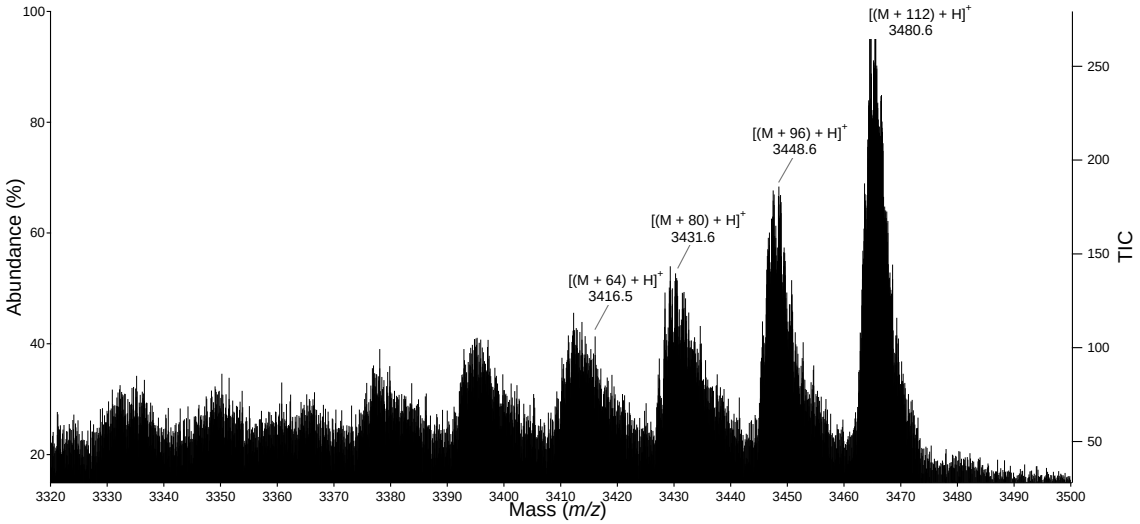



**Figure D.9.** MALDI-TOF MS spectrum of nisin oxidation reaction.

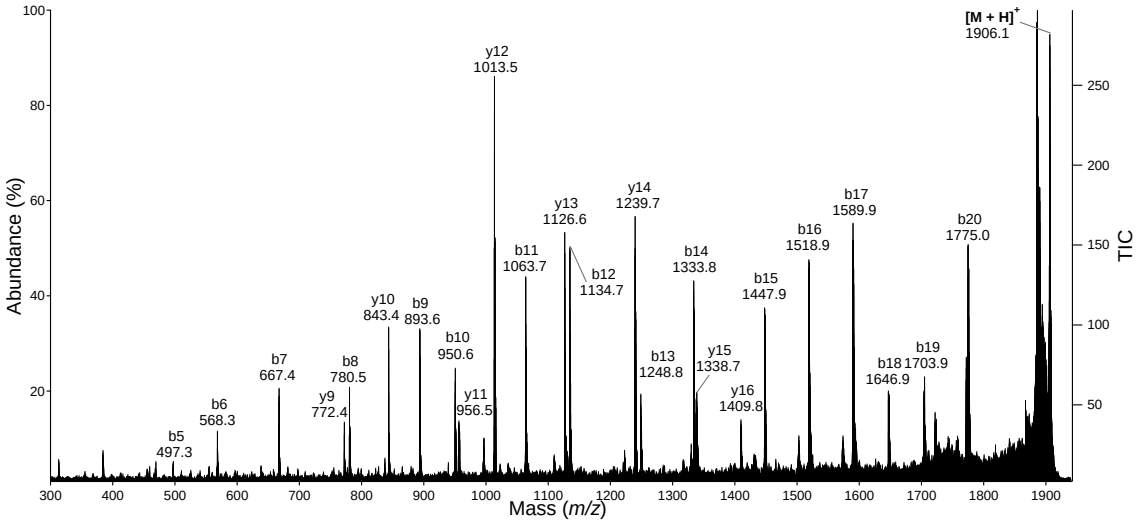
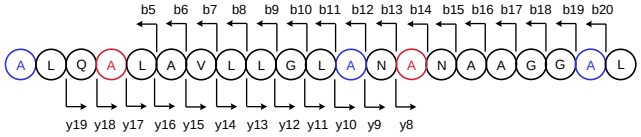



**Figure D.10.** MALDI-TOF MSMS spectrum of Raney Nickel linearisation product Ala15-Leu35 (**34**) (M = 1905.1077 Da), precursor ions: *m/z* 1906 $[M + H]^+$, *m/z* 1928 $[M + Na]^+$.

**Table D.1.** MALDI-TOF MSMS fragmentation pattern of Raney Nickel linearisation product Ala15-Leu35 (**34**) (M = 1905.1077 Da), precursor ions: *m/z* 1906 $[M + H]^+$, *m/z* 1928 $[M + Na]^+$.




**34**

| b ions | | | y ions | | |
|---|---|---|---|---|---|
| Pos | Exact (*m/z*) | Observed (*m/z*) | Pos | Exact (*m/z*) | Observed (*m/z*) |
| 1 | 72.04 | | 20 | 1835.08 | |
| 2 | 185.13 | | 19 | 1721.99 | 1721.92 |
| 3 | 313.19 | | 18 | 1593.93 | 1593.92 |
| 4 | 384.22 | | 17 | 1522.90 | 1522.93 |
| 5 | 497.31 | 497.33 | 16 | 1409.81 | 1409.80 |
| 6 | 568.35 | 568.39 | 15 | 1338.77 | 1338.71 |
| 7 | 667.41 | 667.41 | 14 | 1239.71 | 1239.69 |
| 8 | 780.49 | 780.50 | 13 | 1126.62 | 1126.61 |
| 9 | 893.58 | 893.56 | 12 | 1013.54 | 1013.53 |
| 10 | 950.60 | 950.63 | 11 | 956.52 | 956.50 |
| 11 | 1063.69 | 1063.70 | 10 | 843.43 | 843.44 |
| 12 | 1134.73 | 1134.73 | 9 | 772.40 | 772.39 |
| 13 | 1248.77 | 1248.75 | 8 | 658.35 | 658.41 |
| 14 | 1333.82 | 1333.81 | 7 | 573.30 | |
| 15 | 1447.86 | 1447.86 | 6 | 459.26 | |
| 16 | 1518.90 | 1518.90 | 5 | 388.22 | |
| 17 | 1589.94 | 1589.93 | 4 | 317.18 | |
| 18 | 1646.96 | 1646.96 | 3 | 260.16 | |
| 19 | 1703.98 | 1703.99 | 2 | 203.14 | |
| 20 | 1775.02 | 1775.03 | 1 | 132.10 | |

# Appendix E

# Bioassay Protocol

## E.1 *Staphylococcus aureus*

Plated cultures were prepared by inoculating petri dishes containing 10 mL of LB media with freezer stocks of a clinical isolate *S. aureus* S917 and incubating aerobically at 37 °C for 24 h. Liquid cultures were prepared by inoculating 10 mL LB media and incubating aerobically at 37 °C with shaking (200 rpm) for 24 h. To assess culture purity, microscope slides were smeared with culture (10 $\mu$L), air-dried and Gram-stained. Cells were diluted to an OD of 0.02 prior to the bioassay.

**LB media composition**

Solid media: 30 g/L LB (Sigma-Aldrich); liquid media: 20 g/L LB.

## E.2 Bacteriostatic Assay

In order to determine the minimum inhibitory concentration (MIC) of crude and purified extracts of strain T81, bioactivity was tested in duplicate against the clinical isolate *S. aureus* across a range of concentrations. In this bioassay, extracts are two-fold serial-diluted in 96 well plates. The extract is usually prepared to a concentration of 16 $\mu$g/$\mu$L in DMSO. Wells of column 2 are prepared to a final volume of 100 $\mu$L by adding media (95 $\mu$L) and extract (5 $\mu$L) such that the concentration in column 2 is 400 $\mu$g/mL. Following this, 50 $\mu$L of the liquid is aspirated from column 2 and pipetted into column 3, this step repeated through columns 3 – 11, and the final 50 $\mu$L discarded. Cells (50 $\mu$L) are added to each well to achieve a final OD of 0.01. The plates are sealed and incubated at 37 °C with shaking (220 rpm) for 24 h. Control wells are prepared: a positive control provided by the broad-spectrum polyketide antibiotic, tetracycline; negative controls provided by

wells of media and media/cells. An incubation time of 24 h allows for drug exposure as *S. aureus* growth proceeds through lag, log and stationary phases.

Growth inhibition is measured by two methods: OD and fluorescence with resazurin. After 24 h, the fluorescence (excitation, 385 nm; emission, 509 nm) and OD (600 nm) of each well is measured. Resazurin (0.02% w/v solution, 30 $\mu$L) is added to each well, the plates incubated at 37 °C for 20 min, and fluorescence measured (excitation, 530 nm; emission, 590 nm). In the presence of actively metabolising bacteria, resazurin dye changes colour from a non-fluorescent blue to a fluorescent pink.

# Appendix F

# Known Thiopeptides

**Table F.1.** Known Thiopeptides

| Thiopeptide | Series | Chemical formula | Exact mass | Thiopeptide | Series | Chemical formula | Exact mass |
|---|---|---|---|---|---|---|---|
| 68-1147 [130] | a/b | $C_{72}H_{87}N_{19}O_{17}S_6$ | 1681.4852 | QN3323 A [131] | d | $C_{48}H_{47}N_{13}O_{10}S_6$ | 1157.1893 |
| A10255B [132] | d | $C_{53}H_{48}N_{16}O_{15}S_3$ | 1245.2414 | QN3323 B [131] | d | $C_{49}H_{49}N_{13}O_{10}S_6$ | 1171.205 |
| A10255E [132] | d | $C_{54}H_{50}N_{16}O_{15}S_3$ | 1258.2804 | QN3323 Y1 [131] | d | $C_{47}H_{47}N_{13}O_{10}S_6$ | 1145.1893 |
| A10255G [132] | d | $C_{52}H_{46}N_{16}O_{15}S_3$ | 1230.2491 | Radamycin [133] | d | $C_{48}H_{47}N_{15}O_{11}S_3$ | 1105.2742 |
| A10255J [132] | d | $C_{43}H_{38}N_{14}O_{11}S_3$ | 1022.2007 | S 54832 A-I [134] | e | $C_{59}H_{55}N_{13}O_{19}S_5$ | 1409.2341 |
| Amythiamicin A [135] | d | $C_{50}H_{51}N_{15}O_8S_6$ | 1181.2369 | Sch 18640 [136] | a/b | $C_{72}H_{87}N_{19}O_{17}S_6$ | 1681.4852 |
| Amythiamicin B [135] | d | $C_{50}H_{53}N_{15}O_9S_6$ | 1199.2475 | Sch 40832 [137] | c | $C_{84}H_{104}N_{18}O_{26}S_5$ | 1940.5973 |
| Amythiamicin C [135] | d | $C_{50}H_{50}N_{14}O_9S_6$ | 1182.2209 | Siomycin A [138] | a/b | $C_{77}H_{88}N_{19}NaO_{21}S_5$ | 1797.4903 |
| Amythiamicin D [135] | d | $C_{43}H_{42}N_{12}O_7S_6$ | 1030.1624 | Siomycin B [139] | a/b | $C_{65}H_{75}N_{17}O_{16}S_5$ | 1509.4181 |
| Berninamycin A [140] | d | $C_{51}H_{51}N_{15}O_{15}S$ | 1145.3410 | Siomycin C [139] | a/b | $C_{72}H_{82}N_{18}O_{19}S_5$ | 1662.4607 |
| Berninamycin B [140] | d | $C_{51}H_{51}N_{15}O_{14}S$ | 1129.3461 | Siomycin D [139] | a/b | $C_{70}H_{79}N_{19}O_{18}S_5$ | 1633.4454 |
| Berninamycin C [140] | d | $C_{48}H_{48}N_{14}O_{14}S$ | 1076.3195 | Sulfomycin I [141] | d | $C_{53}H_{50}N_{16}O_{17}S_2$ | 1247.1911 |
| Berninamycin D [140] | d | $C_{45}H_{45}N_{13}O_{13}S$ | 1007.298 | Sulfomycin II [142] | d | $C_{52}H_{48}N_{16}O_{16}S_2$ | 1216.2876 |
| Cyclothiazomycin B1 [143] | d | $C_{59}H_{64}N_{18}O_{14}S_7$ | 1472.2894 | Sulfomycin III [142] | d | $C_{51}H_{46}N_{16}O_{17}S_2$ | 1218.2668 |
| GE2270A [144] | d | $C_{56}H_{55}N_{15}O_{10}S_6$ | 1289.2581 | Thioactin [145] | d | $C_{43}H_{40}N_{14}O_{11}S_4$ | 1056.1884 |
| GE2270B1 [146] | d | $C_{55}H_{53}N_{15}O_{10}S_6$ | 1275.2424 | Thiocillin I [79] | d | $C_{48}H_{49}N_{13}O_{10}S_6$ | 1159.2050 |
| GE2270B2 [146] | d | $C_{55}H_{53}N_{15}O_9S_6$ | 1259.2475 | Thiocillin II [79] | d | $C_{49}H_{51}N_{13}O_{10}S_6$ | 1173.2206 |
| GE2270C1 [146] | d | $C_{54}H_{51}N_{15}O_9S_6$ | 1245.2318 | Thiocillin III [79] | d | $C_{49}H_{51}N_{13}O_9S_6$ | 1157.2257 |
| GE2270C2A [147] | d | $C_{56}H_{55}N_{15}O_{11}S_6$ | 1305.253 | Thiocillin IV [79] | d | $C_{49}H_{49}N_{13}O_9S_6$ | 1155.21 |
| GE2270C2B [146] | d | $C_{55}H_{53}N_{15}O_{10}S_6$ | 1275.2424 | Thiomuracin A [67] | d | $C_{60}H_{52}N_{14}O_{12}S_6$ | 1352.2213 |
| GE2270D1 [146] | d | $C_{53}H_{49}N_{15}O_9S_6$ | 1231.2162 | Thiomuracin B [67] | d | $C_{60}H_{54}N_{14}O_{12}S_6$ | 1354.237 |
| GE2270D2 [147] | d | $C_{55}H_{53}N_{15}O_{10}S_6$ | 1275.2424 | Thiomuracin C [67] | d | $C_{60}H_{54}N_{14}O_{11}S_6$ | 1338.2421 |
| GE2270E [147] | d | $C_{54}H_{51}N_{15}O_{10}S_6$ | 1261.2268 | Thiomuracin D [67] | d | $C_{60}H_{54}N_{14}O_{10}S_6$ | 1322.2472 |
| GE2270T [146] | d | $C_{55}H_{51}N_{15}O_{10}S_6$ | 1273.2268 | Thiomuracin E [67] | d | $C_{60}H_{54}N_{14}O_{13}S_6$ | 1370.2319 |
| GE37468A [148] | d | $C_{59}H_{52}N_{14}O_{12}S_5$ | 1308.2493 | Thiomuracin F [67] | d | $C_{60}H_{52}N_{14}O_{12}S_6$ | 1352.2213 |
| Geninthiocin [149] | d | $C_{50}H_{49}N_{15}O_{15}S$ | 1131.3253 | Thiomuracin G [67] | d | $C_{60}H_{52}N_{14}O_{13}S_6$ | 1368.2163 |
| Glycothiohexide A [150] | e | $C_{58}H_{57}N_{13}O_{15}S_6$ | 1367.2421 | Thiomuracin H [67] | d | $C_{60}H_{54}N_{14}O_{12}S_6$ | 1354.237 |
| JBIR-83 [151] | d | $C_{34}H_{36}N_{10}O_7S_2$ | 760.221 | Thiomuracin I [67] | d | $C_{59}H_{50}N_{14}O_{12}S_6$ | 1338.2057 |
| JBIR-84 [151] | d | $C_{37}H_{39}N_{11}O_8S_2$ | 829.2424 | Thiopeptin A1a [152] | a/b | $C_{72}H_{86}N_{18}O_{19}S_5$ | 1666.492 |
| Methylsulfomycin [153] | d | $C_{55}H_{54}N_{16}O_{16}S_2$ | 1258.3345 | Thiopeptin A1b [152] | a/b | $C_{72}H_{84}N_{18}O_{19}S_5$ | 1664.4764 |
| Micrococcin P1 [154] | d | $C_{48}H_{49}N_{13}O_9S_6$ | 1143.2100 | Thiopeptin A3a [152] | a/b | $C_{65}H_{79}N_{17}O_{16}S_5$ | 1513.4494 |
| Micrococcin P2 [154] | d | $C_{48}H_{47}N_{13}O_9S_6$ | 1141.1944 | Thiopeptin A3b [152] | a/b | $C_{65}H_{77}N_{17}O_{16}S_5$ | 1511.4338 |
| MJ347-81F4A [155] | e | $C_{61}H_{60}N_{14}O_{18}S_5$ | 1436.2814 | Thiopeptin A4a [152] | a/b | $C_{68}H_{82}N_{18}O_{17}S_5$ | 1582.4709 |
| MJ347-81F4B [155] | e | $C_{60}H_{58}N_{14}O_{18}S_5$ | 1422.2657 | Thiopeptin A4b [152] | a/b | $C_{68}H_{80}N_{18}O_{17}S_5$ | 1580.4552 |
| Nocardithiocin [156] | d | $C_{48}H_{48}N_{12}O_{11}S_6$ | 1160.189 | Thiopeptin Ba [157] | a/b | $C_{71}H_{84}N_{18}O_{18}S_6$ | 1668.4535 |
| Nocathiacin I [158] | e | $C_{61}H_{60}N_{14}O_{17}S_6$ | 1452.2585 | Thiopeptin Bb [157] | a/b | $C_{71}H_{82}N_{18}O_{19}S_5$ | 1650.4607 |
| Nocathiacin II [159] | e | $C_{61}H_{60}N_{14}O_{16}S_6$ | 1436.2636 | Thiostrepton A [160] | a/b | $C_{72}H_{85}N_{19}O_{18}S_5$ | 1663.4924 |
| Nocathiacin III [159] | e | $C_{52}H_{43}N_{13}O_{15}S_6$ | 1281.1326 | Thiostrepton B [160] | a/b | $C_{66}H_{79}N_{17}O_{16}S_5$ | 1525.4494 |
| Nocathiacin IV [161] | e | $C_{58}H_{57}N_{13}O_{15}S_6$ | 1367.2421 | Thiotipin [162] | d | $C_{52}H_{47}N_{15}O_{16}S_2$ | 1201.2767 |
| Nosiheptide [163] | e | $C_{51}H_{43}N_{13}O_{12}S_6$ | 1221.1478 | Thioxamycin [164] | d | $C_{52}H_{48}N_{16}O_{15}S_4$ | 1264.2368 |
| O-$CH_3$-Glycothiohexide A [150] | e | $C_{59}H_{59}N_{13}O_{15}S_6$ | 1381.2578 | TP-1161 [69] | d | $C_{50}H_{47}N_{15}O_{13}S_3$ | 1161.264 |
| Promoinducin [165] | d | $C_{57}H_{54}N_{16}O_{18}S_2$ | 1314.3243 | Val-Geninthiocin [166] | d | $C_{50}H_{49}N_{15}O_{14}S$ | 1115.3304 |
| Promothiocin A [167] | d | $C_{36}H_{37}N_{11}O_8S_2$ | 815.2268 | YM-266183 [168] | d | $C_{48}H_{47}N_{13}O_{10}S_6$ | 1157.1893 |
| Promothiocin B [167] | d | $C_{42}H_{43}N_{13}O_{10}S_2$ | 953.2697 | YM-266184 [168] | d | $C_{49}H_{49}N_{13}O_{10}S_6$ | 1171.2050 |

# References


1. Fleming, A. *Brit. J. Exp. Pathol.* **1929**, *10*, 226–236.
2. Verdine, G. L. *Nature (London)* **1996**, *384*, 11–13.
3. Walsh, C. T. *Acc. Chem. Res.* **2008**, *41*, 4–10.
4. Kingston, D. G. I. *The Practice of Medicinal Chemistry*; Academic Press: London, **1996**.
5. Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2007**, *70*, 461–477.
6. Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2012**, *75*, 311–335.
7. Kingston, D. G. I. *J. Nat. Prod.* **2011**, *74*, 496–511.
8. Li, X.; Guo, J.; Dai, S.; Ouyang, Y.; Wu, H.; Sun, W.; Wang, G. *Curr. Top. Med. Chem. (Sharjah, United Arab Emirates)* **2009**, *9*, 1525–1535.
9. Lefevre, F.; Robe, P.; Jarrin, C.; Ginolhac, A.; Zago, C.; Auriol, D.; Vogel, T. M.; Simonet, P.; Nalin, R. *Res. Microbiol.* **2008**, *159*, 153–161.
10. Miller, S. J.; Clardy, J. *Nat. Chem.* **2009**, *1*, 261–263.
11. Walsh, C. T.; Fischbach, M. A. *J. Am. Chem. Soc.* **2010**, *132*, 2469–2493.
12. Feher, M.; Schmidt, J. M. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 218–227.
13. Rouhi, M. *Chem. Eng. News* **2003**, *81*, 261–263.
14. Borman, S. *Chem. Eng. News* **2006**, *84*, 56–78.
15. Dancik, V.; Seiler, K. P.; Young, D. W.; Schreiber, S. L.; Clemons, P. A. *J. Am. Chem. Soc.* **2010**, *132*, 9259–9261.
16. Connon, S. A.; Giovannoni, S. J. *Appl. Environ. Microbiol.* **2002**, *68*, 3878–3885.
17. Schiraldi, C.; De Rosa, M. *Trends Biotechnol.* **2002**, *20*, 515–521.
18. Newman, D. J.; Cragg, G. M.; Snader, K. M. *J. Nat. Prod.* **2003**, *66*, 1022–1037.
19. Pace, N. R. *Science* **1997**, *276*, 734–740.
20. Madigan, M. T.; Martinko, J. M.; Parker, J. *Brock Biology of Microorganisms*, 8th ed.; Prentice Hall: Upper Saddle River, NJ, **1996**.
21. *Genomes Online (GOLD)*, **2013**. `http://www.genomesonline.org/cgi-bin/GOLD/index.cgi`.
22. Javaux, E. J. *Res. Microbiol.* **2006**, *157*, 37–48.
23. Pakchung, A. A. H.; Simpson, P. J. L.; Codd, R. *Environ. Chem.* **2006**, *3*, 77–93.
24. Kristjánsson, J. K.; Hreggvidsson, G. O. *World J. Microbiol. Biotechnol.* **1995**, *11*, 17–24.
25. MacElroy, R. D. *Biosystems* **1974**, *6*, 74–75.

26. Horikoshi, K.; Antranikian, G.; Bull, A. T.; Robb, F. T.; Stetter, K. O. *Extremophiles Handbook*; Springer-Verlag: Tokyo, **2011**; pp 3–15.

27. Ferrer, M.; Golyshina, O.; Beloqui, A.; Golyshin, P. N. *Curr. Opin. Microbiol.* **2007**, *10*, 207–214.

28. Wilson, Z. E.; Brimble, M. A. *Nat. Prod. Rep.* **2009**, *26*, 44–71.

29. Stierle, A. A.; Stierle, D. B.; Kelly, K. *J. Org. Chem.* **2006**, *71*, 5357–5360.

30. Xu, H.-S.; Roberts, N.; Singleton, F. L.; Attwell, R. W.; Grimes, D. J.; Colwell, R. R. *Microbial Ecology* **1982**, *8*, 313–323.

31. Colwell, R. R.; Grimes, D. J. In *Nonculturable Microorganisms in the Environment*; ASM Press: Washington DC, **2000**; pp 1–6.

32. Park, C. B.; Clark, D. S. *Appl. Environ. Microbiol.* **2002**, *68*, 1458–1463.

33. Scherlach, K.; Hertweck, C. *Org. Biomol. Chem.* **2009**, *7*, 1753–1760.

34. Corre, C.; Challis, G. L. *Nat. Prod. Rep.* **2009**, *26*, 977–986.

35. Bode, H. B.; Bethe, B.; Hofs, R.; Zeeck, A. *ChemBioChem* **2002**, *3*, 619–627.

36. Grond, S.; Papastavrou, I.; Zeeck, A. *Eur. J. Org. Chem.* **2002**, 3237–3242.

37. Bode, H. B.; Walker, M.; Zeeck, A. *Eur. J. Org. Chem.* **2000**, 1451–1456.

38. Cueto, M.; Jensen, P. R.; Kauffman, C.; Fenical, W.; Lobkovsky, E.; Clardy, J. *J. Nat. Prod.* **2001**, *64*, 1444–1446.

39. Oh, D.-C.; Kauffman, C. A.; Jensen, P. R.; Fenical, W. *J. Nat. Prod.* **2007**, *70*, 515–520.

40. Lau, J.; Frykman, S.; Regentin, R.; Ou, S.; Tsuruta, H.; Licari, P. *Biotechnol. Bioeng.* **2002**, *78*, 280–288.

41. Udwary, D. W.; Zeigler, L.; Asolkar, R. N.; Singan, V.; Lapidus, A.; Fenical, W.; Jensen, P. R.; Moore, B. S. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 10376–10381.

42. Zazopoulos, E.; Huang, K.; Staffa, A.; Liu, W.; Bachmann, B. O.; Nonaka, K.; Ahlert, J.; Thorson, J. S.; Shen, B.; Farnet, C. M. *Nat. Biotechnol.* **2003**, *21*, 187–190.

43. Bergmann, S.; Schuemann, J.; Scherlach, K.; Lange, C.; Brakhage, A. A.; Hertweck, C. *Nat. Chem. Biol.* **2007**, *3*, 213–217.

44. Brakhage, A. A.; Schuemann, J.; Bergmann, S.; Scherlach, K.; Schroeckh, V.; Hertweck, C. *Prog. Drug. Res.* **2008**, *66*, 1, 3–12.

45. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science (Washington, DC, 1883-)* **1989**, *246*, 64–71.

46. Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 151–153.

47. Karas, M.; Bachmann, D.; Hillenkamp, F. *Anal. Chem.* **1985**, *57*, 2935–2939.

48. Chughtai, K.; Heeren, R. M. A. *Chem. Rev. (Washington, DC, U. S.)* **2010**, *110*, 3237–3277.

49. Cornett, D. S.; Reyzer, M. L.; Chaurand, P.; Caprioli, R. M. *Nat. Methods* **2007**, *4*, 828–833.

50. Svatos, A. *Trends Biotechnol.* **2010**, *28*, 425–434.

51. Yang, Y.-L.; Xu, Y.; Kersten, R. D.; Liu, W.-T.; Meehan, M. J.; Moore, B. S.; Bandeira, N.; Dorrestein, P. C. *Angew. Chem., Int. Ed.* **2011**, *50*, 5839–5842, S5839/1–S5839/30.

52. *National High Magnetic Field Laboratory*, **2013**. `http://www.magnet.fsu.edu/education/tutorials/tools/ionization_maldi.html`.

53. Caprioli, R. M.; Farmer, T. B.; Gile, J. *Anal. Chem.* **1997**, *69*, 4751–4760.

54. Yang, Y.-L.; Xu, Y.; Straight, P.; Dorrestein, P. C. *Nat. Chem. Biol.* **2009**, *5*, 885–887.

55. Stein, T. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 1146–1152.

56. *Antibase*, **2012**. `http://www.wiley-vch.de/stmdata/antibase.php`.

57. Stott, M. B.; Crowe, M. A.; Mountain, B. W.; Smirnova, A. V.; Hou, S.; Alam, M.; Dunfield, P. F. *Environ. Microbiol.* **2008**, *10*, 2030–2041.

58. Hauch, E. *Putative lantiobiotic gene cluster derived from Thermogemmatispora lignivorax*; GNS Science, **2011**.

59. Arnison, P. G. et al. *Nat. Prod. Rep.* **2013**, *30*, 108–160.

60. Knerr, P. J.; van der Donk, W. A. *J. Am. Chem. Soc.* **2012**, *134*, 7648–7651.

61. Su, T. L. *Br. J. Exp. Pathol.* **1948**, *29*, 473–481.

62. Kelly, W. L.; Pan, L.; Li, C. *J. Am. Chem. Soc.* **2009**, *131*, 4327–4334.

63. Wieland Brown, L. C.; Acker, M. G.; Clardy, J.; Walsh, C. T.; Fischbach, M. A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 2549–2553.

64. Liao, R.; Duan, L.; Lei, C.; Pan, H.; Ding, Y.; Zhang, Q.; Chen, D.; Shen, B.; Yu, Y.; Liu, W. *Chem. Biol. (Cambridge, MA, U. S.)* **2009**, *16*, 141–147.

65. Bagley, M. C.; Dale, J. W.; Merritt, E. A.; Xiong, X. *Chem. Rev. (Washington, DC, U. S.)* **2005**, *105*, 685–714.

66. Benazet, F.; Cartier, M.; Florent, J.; Godard, C.; Jung, G.; Lunel, J.; Mancy, D.; Pascal, C.; Renaut, J. *Experientia* **1980**, *36*, 414–416.

67. Morris, R. P.; Leeds, J. A.; Naegeli, H. U.; Oberer, L.; Memmert, K.; Weber, E.; LaMarche, M. J.; Parker, C. N.; Burrer, N.; Esterow, S.; Hein, A. E.; Schmitt, E. K.; Krastel, P. *J. Am. Chem. Soc.* **2009**, *131*, 5946–5955.

68. Ding, Y.; Yu, Y.; Pan, H.; Guo, H.; Li, Y.; Liu, W. *Mol. BioSyst.* **2010**, *6*, 1180–1185.

69. Engelhardt, K.; Degnes, K. F.; Zotchev, S. B. *Appl. Environ. Microbiol.* **2010**, *76*, 7093–7101.

70. Wang, J.; Yu, Y.; Tang, K.; Liu, W.; He, X.; Huang, X.; Deng, Z. *Appl. Environ. Microbiol.* **2010**, *76*, 2335–2344.

71. Young, T. S.; Walsh, C. T. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 13053–13058, S13053/1–S13053/25.

72. Yu, Y.; Duan, L.; Zhang, Q.; Liao, R.; Ding, Y.; Pan, H.; Wendt-Pienkowski, E.; Tang, G.; Shen, B.; Liu, W. *ACS Chem. Biol.* **2009**, *4*, 855–864.

73. Walsh, C. T.; Malcolmson, S. J.; Young, T. S. *ACS Chem. Biol.* **2012**, *7*, 429–442.

74. Bycroft, B. W.; Gowland, M. S. *J. Chem. Soc., Chem. Commun.* **1978**, 256–258.

75. Zhang, Q.; Li, Y.-X.; Chen, D.-D.; Yu, Y.; Duan, L.; Shen, B.; Liu, W. *Nat. Chem. Biol.* **2011**, *7*, 154–160.

76. Ivanova, N. et al. *Nature (London)* **2003**, *423*, 87–91.

77. *Clustal Omega*, **2013**. `http://www.ebi.ac.uk/Tools/msa/clustalo/`.

78. Walsh, C. T.; Acker, M. G.; Bowers, A. A. *J. Biol. Chem.* **2010**, *285*, 27525–27531.

79. Bowers, A. A.; Walsh, C. T.; Acker, M. G. *J. Am. Chem. Soc.* **2010**, *132*, 12182–12184.

80. Willey, J. M.; van der Donk, W. A. *Annu. Rev. Microbiol.* **2007**, *61*, 477–501.

81. Knerr, P. J.; van der Donk, W. A. *Annu. Rev. Biochem.* **2012**, *81*, 479–505.

82. Chatterjee, C.; Miller, L. M.; Leung, Y. L.; Xie, L.; Yi, M.; Kelleher, N. L.; van der Donk, W. A. *J. Am. Chem. Soc.* **2005**, *127*, 15332–15333.

83. Goto, Y.; Li, B.; Claesen, J.; Shi, Y.; Bibb, M. J.; van der Donk, W. A. *PLoS Biol.* **2010**, *8*, 1–10.

84. Kluskens, L. D.; Kuipers, A.; Rink, R.; De, B. E.; Fekken, S.; Driessen, A. J. M.; Kuipers, O. P.; Moll, G. N. *Biochemistry* **2005**, *44*, 12827–12834.

85. Li, B.; Yu, J. P. J.; Brunzelle, J. S.; Moll, G. N.; van der Donk, W. A.; Nair, S. K. *Science (Washington, DC, U. S.)* **2006**, *311*, 1464–1467.

86. Xie, L.; Miller, L. M.; Chatterjee, C.; Averin, O.; Kelleher, N. L.; van der Donk, W. A. *Science (Washington, DC, U. S.)* **2004**, *303*, 679–682.

87. Müller, W.; Schmiederer, T.; Ensle, P.; Süssmuth, R. *Angew. Chem., Int. Ed.* **2010**, *49*, 2436–2440.

88. Lubelski, J.; Rink, R.; Khusainov, R.; Moll, G.; Kuipers, O. *Cell. Mol. Life Sci.* **2008**, *65*, 455–476.

89. Li, J.; Girard, G.; Florea, B. I.; Geurink, P. P.; Li, N.; van der Marel, G. A.; Overhand, M.; Overkleeft, H. S.; van Wezel, G. P. *Org. Biomol. Chem.* **2012**, *10*, 8677–8683.

90. Hasper, H. E.; de Kruijff, B.; Breukink, E. *Biochemistry* **2004**, *43*, 11567–11575.

91. Piper, C.; Cotter, P. D.; Ross, P. R.; Hill, C. *Curr. Drug Discov. Technol.* **2009**, *6*, 1–18.

92. Grasemann, H.; Stehling, F.; Brunar, H.; Widmann, R.; Laliberte, T. W. *Chest* **2007**, *131*, 1461–1466.

93. Sedgwick, T.; Dawson, M. J. *MedNous May* **2009**, *6*, 8–9.

94. Donadio, S.; Maffioli, S.; Monciardini, P.; Sosio, M.; Jabes, D. *J. Antibiot.* **2010**, *63*, 423–430.

95. Ghobrial, O.; Derendorf, H.; Hillman, J. D. *J. Pharm. Sci.* **2010**, *99*, 2521–2528.

96. Cotter, P. D.; Hill, C.; Ross, R. P. *Nat. Rev. Microbiol.* **2005**, *3*, 777–788.

97. Zhao, M.; Li, Z.; Bugenhagen, S. *J. Nucl. Med.* **2008**, *49*, 1345–1352.

98. Yang, W. A., X.; van der Donk *Chem. Eur. J.* **2013**, *19*, 7662–7677.

99. *BLAST®*, **2013**. `http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins`.

100. Oman, T. J.; van der Donk, W. A. *Nat. Chem. Biol.* **2010**, *6*, 9–18.

101. Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. *J. Mass Spectrom.* **2000**, *35*, 1399–1406.

102. Willard, B. B.; Kinter, M. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 1262–1271.

103. Garg, N.; Tang, W.; Goto, Y.; Nair, S. K.; van der Donk, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 5241–5246, S5241/1–S5241/10.

104. Cox, C. R.; Coburn, P. S.; Gilmore, M. S. *Curr. Protein Pept. Sci.* **2005**, *6*, 77–84.

105. Holo, H.; Jeknic, Z.; Daeschel, M.; Stevanovic, S.; Nes, I. F. *Microbiology* **2001**, *147*, 643–651.

106. McClerren, A. L. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 17243–17248.

107. Caetano, T.; Krawczyk, J. M.; Mosker, E.; Sussmuth, R. D.; Mendo, S. *Chem. Biol.* **2011**, *18*, 90–100.

108. Cooper, L. E.; McClerren, A. L.; Chary, A.; van der Donk, W. A. *Chem. Biol.* **2008**, *15*, 1035–1045.

109. Fuchs, S. W.; Jaskolla, T. W.; Bochmann, S.; Koetter, P.; Wichelhaus, T.; Karas, M.; Stein, T.; Entian, K.-D. *Appl. Environ. Microbiol.* **2011**, *77*, 1698–1707.

110. Watson, J. T.; Sparkman, O. D. Analysis of Proteins and Other Biopolymers. In *Introduction to Mass Spectrometry: Instrumentation, Applications, and Strategies for Data Interpretation*, 4th ed.; Wiley: West Sussex, **2007**; pp 689–803.

111. Shi, Y.; Bueno, A.; van der Donk, W. A. *Chem. Commun. (Cambridge, U. K.)* **2012**, *48*, 10966–10968.

112. Li, B.; Sher, D.; Kelly, L.; Shi, Y.; Huang, K.; Knerr, P. J.; Joewono, I.; Rusch, D.; Chisholm, S. W.; van der Donk, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 10430–10435.

113. Zimmermann, N.; Metzger, J. W.; Jung, G. *Eur. J. Biochem.* **1995**, *228*, 786–797.

114. Kawulka, K. E.; Sprules, T.; Diaper, C. M.; Whittal, R. M.; McKay, R. T.; Mercier, P.; Zuber, P.; Vederas, J. C. *Biochemistry* **2004**, *43*, 3385–3395.

115. Meyer, H. E.; Heber, M.; Eisermann, B.; Korte, H.; Metzger, J. W.; Jung, G. *Anal. Biochem.* **1994**, *223*, 185–90.

116. Smith, L.; Novak, J.; Rocca, J.; McClung, S.; Hillman, J. D.; Edison, A. S. *Eur. J. Biochem.* **2000**, *267*, 6810–6816.

117. Lohans, C. T.; Huang, Z.; van Belkum, M. J.; Giroud, M.; Sit, C. S.; Steels, E. M.; Zheng, J.; Whittal, R. M.; McMullen, L. M.; Vederas, J. C. *J. Am. Chem. Soc.* **2012**, *134*, 19540–19543.

118. Bartlett, P. A. *J. Am. Chem. Soc.* **1976**, *98*, 3305–3312.

119. Kusters, E.; Allgaier, H.; Jung, G.; Bayer, E. *Chromatographia* **1984**, *18*, 287–293.

120. Allgaier, H.; Jung, G.; Werner, R. G.; Schneider, U.; Zahner, H. *Eur. J. Biochem.* **1986**, *160*, 9–22.

121. Kellner, R.; Jung, G.; Horner, T.; Zahner, H.; Schnell, N.; Entian, K. D.; Gotz, F. *Eur. J. Biochem.* **1988**, *177*, 53–59.

122. Kellner, R.; Jung, G.; Josten, M.; Kaletta, C.; Entian, K. D.; Sahl, H. G. *Angew. Chem., Int. Ed.* **1989**, *28*, 616–619.

123. Tang, W.; van der Donk, W. A. *Nat. Chem. Biol.* **2013**, *9*, 157–159.

124. Knerr, P. J.; van der Donk, W. A. *J. Am. Chem. Soc.* **2013**, *135*, 7094–7097.

125. Ross, A. C.; Liu, H.; Pattabiraman, V. R.; Vederas, J. C. *J. Am. Chem. Soc.* **2010**, *132*, 462–463.

126. West, L. M. *The Isolation of Secondary Metabolites from NZ Marine Sponges*; Ph. D Thesis, Victoria University of Wellington, **2001**.

127. Mota-Meira, M.; La Pointe, G.; Lacroix, C.; Lavoie, M. C. *Antimicrob. Agents Chemother.* **2000**, *44*, 24–29.

128. Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. *J. Org. Chem.* **1997**, *62*, 7512–7515.

129. Weisburg, W. G.; Barns, S. M.; Pelletier, D. A.; Lane, D. J. *J. Bacteriol.* **1991**, *173*, 697–703.

130. Marquez, J. A.; Testa, R. T.; Wagman, G. H.; Weinstein, M. J. *Antibiotic complex from Micromonospora arborensis*; **1978**, US4078056 A, 6 pp.

131. Kamigiri, K.; Watanabe, M.; Nagai, K.; Arao, N.; Suzumura, K.; Suzuki, K.; Kurane, R.; Yamaoka, M.; Kawano, Y. *Thiopeptide compounds suitable for treatment of multidrug resistant bacteria infection*; **2002**, WO2002072617A1, 27 pp.

132. Boeck, L. D.; Berry, D. M.; Mertz, F. P.; W., R., Wetzel *J. Antibiot.* **1992**, *45*, 1222–1230.

133. Castro Rodríguez, J.; González Holgado, G.; Santamaría Sánchez, R. I.; Cañedo, L. M. *J. Antibiot.* **2002**, *55*, 391–395.

134. Keller-Juslen, C.; Kuhn, M.; King, H. D. *Antibiotics, pharmaceutical compositions and their use*; **1984**, US4478831A, 17 pp.

135. Shimanaka, K.; Takahashi, Y.; Iinuma, H.; Naganawa, H.; Takeuchi, T. *J. Antibiot.* **1994**, *47*, 1145–1152.

136. Puar, M. S.; Ganguly, A. K.; Afonso, A.; Brambilla, R.; Mangiaracina, P.; Sarre, O.; MacFarlane, R. D. *J. Am. Chem. Soc.* **1981**, *103*, 5231–5233.

137. Puar, M. S.; Chan, T. M.; Hegde, V.; Patel, M.; Bartner, P.; Ng, K. J.; Pramanik, B. N.; MacFarlane, R. D. *J. Antibiot.* **1998**, *51*, 221–224.

138. Tori, K.; Tokura, K.; Yoshimura, Y.; Okabe, K.; Otsuka, H.; Inagski, F.; Miyazawa, T. *J. Antibiot.* **1979**, *32*, 1072–1077.

139. Okabe, K.; Tokura, K.; Hayashi, K.; Tori, K.; Terui, Y.; Yoshimura, Y.; Otsuka, H.; Matsushita, K.; Inagaki, F.; Miyazawa, T. *Pept. Chem.* **1980**, *17*, 19–24.

140. Liesch, J. M.; Rinehart, K. L. *J. Am. Chem. Soc.* **1977**, *99*, 1645–1646.

141. Abe, H.; Kushidab, K.; Shiobarac, Y.; Kodamac, M. *Tetrahedron Lett.* **1988**, *29*, 1401–1404.

142. Kohno, J.; Kameda, N.; Nishio, M.; Kinumaki, A.; Komatsubara, S. *J. Antibiot.* **1996**, *49*, 1063–1065.

143. Hashimoto, M.; Murakami, T.; Funahashi, K.; Tokunaga, T.; Nihei, K.-i.; Okuno, T.; Kimura, T.; Naoki, H.; Himeno, H. *Bioorg. Med. Chem.* **2006**, *14*, 8259–8270.

144. Selva, E.; Beretta, G.; Montanini, N.; Saddler, G. S.; Gastaldo, L.; Ferrari, P.; Lorenzetti, R.; Landini, P.; Ripamonti, F.; Goldstein, B. P.; Berti, M.; Montanaro, L.; Denaro, M. *J. Antibiot.* **1991**, *44*, 693–701.

145. Yun, B.-S.; Hidaka, T.; Furihata, K.; Seto, H. *J. Antibiot.* **1994**, *47*, 1541–1545.

146. Colombo, L.; Stella, S.; Selva, E. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 717–22.

147. Lociuro, S.; Tavecchia, P.; Ciabatti, R.; Restelli, E. *Derivatives of antibiotic GE2270 factors C2a, D2, and E*; **1997**, EP0880541 B1, 78 pp.

148. Stella, S.; Montanini, N.; Le Monnier, F.; Ferrari, P.; Colombo, L.; Marinelli, F.; Landini, P.; Ciciliato, I.; Goldstein, B. P.; Selva, E.; Denaro, M. *J. Antibiot.* **1995**, *48*, 780–786.

149. Yun, B.-S.; Hidaka, T.; Furihata, K.; Seto, H. *J. Antibiot.* **1994**, *47*, 969–975.

150. Northcote, P. T.; Williams, D.; Manning, J. K.; Borders, D. B.; Maiese, W. M.; Lee, M. D. *J. Antibiot.* **1994**, *47*, 894–900.

151. Takagi, M.; Motohashi, K.; Nagai, A.; Hashimoto, J.; Shin-Ya, K. *J. Antibiot.* **2010**, *63*, 405–408.

152. Hensens, O. D.; Albers-Schonberg, G. *J. Antibiot.* **1983**, *36*, 832–845.

153. Vijaya Kumar, E. K.; Kenia, J.; Mukhopadhyay, T.; Nadkarni, S. *J. Nat. Prod.* **1999**, *62*, 1562–1564.

154. Hall, G. E.; Sheppard, N.; Walker, J. *J. Chem. Soc. Perkin 1.* **1966**, *16*, 1371–1373.

155. Sasaki, T.; Otani, T.; Matsumoto, H.; Unemi, N.; Hamada, M.; Takeuchi, T.; Hori, M. *J. Antibiot.* **1998**, *51*, 715–721.

156. Mukai, A.; Fukai, T.; Hoshino, Y.; Yazawa, K.; Harada, K.; Mikami, Y. *J. Antibiot.* **2009**, *62*, 613–619.

157. Tori, K.; Tokura, K.; Yoshimura, Y.; Terui, Y.; Okabe, K.; Otsuka, H.; Matsushita, K.; Inagaki, F.; Miyazawa, T. *J. Antibiot.* **1981**, *34*, 124–129.

158. Constantine, K. L.; Mueller, L.; Huang, S.; Abid, S.; Lam, K. S.; Li, W.; Leet, J. E. *J. Am. Chem. Soc.* **2002**, *124*, 7284–7285.

159. Leet, J. E.; Li, W.; Ax, H. A.; Matson, J. A.; Huang, S.; Huang, R.; Cantone, J. L.; Drexler, D.; Dalterio, R. A.; Lam, K. S. *J. Antibiot.* **2003**, *56*, 232–242.

160. Hang, P. C.; Honek, J. F. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1471–1474.

161. Regueiro-Ren, A.; Ueda, Y. *J. Org. Chem.* **2002**, *67*, 8699–8702.

162. Yun, B.-S.; Hidaka, T.; Furihata, K.; Seto, H. *Tetrahedron* **1994**, *50*, 11659–11664.

163. Mocek, U.; Chen, L.-C.; Keller, P. J.; Houck, D. R.; Beale, J. M.; Floss, H. G. *J. Antibiot.* **1989**, *42*, 1643–1648.

164. Matsumoto, M.; Kawamura, Y.; Yasuda, Y.; Tanimoto, T.; Matsumoto, K.; Yoshida, T.; Shoji, J. *J. Antibiot.* **1989**, *42*, 1465–1469.

165. Yun, B.-S.; Seto, H. *Biosci. Biotech. Bioch.* **1995**, *59*, 876–880.

166. Sajid, I.; Shaaban, K. A.; Frauendorf, H.; Hasnain, S.; Laatsch, H. *Z. Naturforsch.* **2008**, *63b*, 1223–1230.

167. Yun, B.-S.; Fujita, K.; Furihata, K.; Seto, H. *Tetrahedron* **2001**, *57*, 9683–9687.

168. Nagai, K.; Kamigiri, K.; Arao, N.; Suzumura, K.; Kawano, Y.; Yamaoka, M.; Zhang, H.; Watanabe, M.; Suzuki, K. *J. Antibiot.* **2003**, *56*, 123–128.