UNIVERSITY OF NAPLES FEDERICO II
SCHOOL OF MEDICINE
DEPARTMENT OF PHARMACY

PhD programme in Pharmaceutical Sciences

An innovative platform for the discovery of new marine bioactive natural substances

PhD student
Viviana Di Matteo

Tutor
Prof. Roberta Teta

Co-Tutor
Prof. Alfonso Mangoni

Coordinator
Prof. ROSARIA MELI

CYCLE (a.y. 2024/2025)

CONTENTS

Abstract in English I
Abstract in Italian V
INTRODUCTION 1
References 8

Chapter 1 11
Structural elucidation methods 11
1.1 Mass spectrometry 12
1.2 Nuclear Magnetic Resonance 15
1.3 Methods for the determination of relative/absolute configurations 17
1.3.1 Electronic circular dichroism (ECD) 20
1.3.2 Computational methods 22
1.3.3 Murata method 24
1.3.4 Marfey's method 27
References 29

Chapter 2 31
Molecular networking 31
References 35

Chapter 3 36
Genome Mining 36
References 41

PART 1 Exploration of bacterial natural products 42
Chapter 4 43
4.1 Extremophiles 43
4.2 Thermophiles 47
4.3 Thermoactinomyces vulgaris 48
References 51

Chapter 5 53
Unraveling Thermoactinoamides: Molecular Networking, Biosynthesis, and Unexpected Epimerization 53
5.1 Synopsis of this chapter 53

5.2 Discovery, Isolation, and Stereostructural Elucidation of Thermoactinoamide A from the Thermophilic Bacterium Thermoactinomyces vulgaris 55
5.3 Biological Activity of Thermoactinoamide A 60
5.4 Non-ribosomal peptides 62
5.5 Non-ribosomal peptide synthetase 64
5.6 Identification of the biosynthetic gene cluster of thermoactinamide 73
5.7 Molecular networking and structural determination of thermoactinamide A congeners 79
5.8 Synthesis of the cyclic peptide thermoactinoamide A: unexpected epimerization 82
5.9 Structure Elucidation of Synthetic Thermoactinoamide A 84
5.10 Cyclization of linear peptide YI-6 86
5.11 Cyclization of linear peptide LV-6 89
..... 91
5.12 Experimental section 91
5.12.1 General experimental procedures 91
5.11.2 Collection, extraction, isolation 92
5.11.3. Bioinformatic analysis 94
5.11.4. LC-HRMS and LC-HRMS/MS 95
5.11.5. LC-HRMS/MS Data Processing and Molecular Networking 95
5.9.6. Bioactivity Assays 97
5.11.7. Advanced Marfey's analysis of thermoactinoamide A (**1a**) 98
5.11.8. Partial hydrolysis of thermoactinoamide A (**1a**)..... 98
5.11.9. Advanced Marfey's analysis of **1b** and **1c**..... 98
5.11.10. Head-to-tail cyclization of linear peptides YI-6 (**2**) and LV-6 (**3**)..... 99
5.12. Spectroscopic data 101
References 126

PART 2 Exploration of cyanobacterial natural products..... 129
Chapter 6 131
Cyanobacteria 131
References 146

Chapter 7 149

Cyanochelin B: a siderophore produced by cyanobacteria with photolytic properties that prevent the monopoly of iron in UV light. 149
7.1 Synopsis of this chapter 149
7.2 Siderophores 150
7.3 Cyanobacterial siderophores 154
7.4 Polyketides and polyketide synthase 157
7.5 PKS-NRPS Hybrid 161
7.6 Cyanochelin B: a cyanobacterial siderophore with double mode of action 162
7.7 Planar Structure Elucidation 163
7.7 Stereostructural Elucidation 168
7.7 Kinetics of photolysis and structure of photolytic fragments 172
7.7 Effect of cyanochelin B on Synechocystis monoculture 176
7.7 Effect of UV-light on iron acquisition during co-cultivation 178
7.7 Phylogenetic distribution and biosynthetic gene clusters of cchB producers 184
7.8 Experimental section 188
7.8.1 General methods 188
7.8.2 Strains and cultivation 188
7.8.3 Siderophore inhibition test 189
7.8.4 Co-cultivation experiments 189
7.8.4 HPLC-HRMS analysis 191
7.8.4 Compound isolation 192
7.8.4 Advanced Marfey's analysis 193
7.8.4 Field sampling and strain isolation 194
7.8.4 Cyanochelin B PCR protocol 195
7.8.4 Whole-genome sequencing and assembly 196
7.8.4 Phylogenetic analysis 196
7.8.4 Analysis of the cyanochelin BGCs 197
7.9 Cyanochelin B biosynthesis and Spectroscopic data 198
References 216

Chapter 8 219
Bio-tracking, bio-monitoring and bio-magnification interdisciplinary studies to assess cyanobacterial harmful algal blooms (cyanoHABs)' impact in complex coastal systems 219
8.1 Synopsis of this chapter 219

8.2 Environmental triggers, bioaccumulation, and ecotoxicology of cyanobacterial toxins 220
8.3 Regulatory efforts to monitor cyanotoxins in water and food 222
8.4 Lake Avernus 2022 Red Bloom 223
8.5 Remote and proximal sensing analysis 225
8.6 Study Area and Sample Collection 227
8.7 Extraction of Mussel Tissues 230
8.8 LC-MS/MS Analysis and Molecular Networking 232
8.9 Quantitative analysis of cyanotoxins 233
8.10 Experimental section 236
8.10.1 General methods 236
8.10.2 Remote and proximal sensing 236
8.10.3 Sampling and handling of mussels 238
8.10.4 Extraction and fractionation 238
8.10.5 LC-MS/MS analysis 239
8.10.6 Computational workflow in MZmine 2.53 239
8.10.7 Feature-based molecular networking 240
References 251

Conclusion 253

Abstract in English

Research on natural products continues to link molecular discoveries with ecological and biomedical knowledge, but progress often stalls at the crossroads between complex mixtures and structural certainties. This thesis proposes and demonstrates an integrated pathway that flows smoothly from genes to networks, to spectra, and, when necessary, to landscapes. High-resolution LC-MS/MS, multidimensional NMR, and targeted derivatization strategies provide the analytical framework; feature-based molecular networking focuses attention on coherent families rather than isolated peaks; and genome mining provides a biosynthetic rationale that governs structural assignments and selects what should be isolated. Throughout the work, this combination allows for early dereplication, more precise targeted isolation, and more robust justification of stereochemical assignments, while opening up pathways to link molecules to function in living systems. The scientific results of this thesis are described in the chapters 5, 7 and 8.

The first section focuses on a family of lipophilic cyclohexapeptides produced by a thermophilic bacterium. Starting from a classic antibacterial assay, the study reexamines thermoactinamide A and reveals a broader constellation of congeners that were previously obscured by low abundance and spectral overlap. Genome sequencing reveals a greater than expected capacity for non-ribosomal peptide synthesis, while domain predictions help reconcile MS/MS fragmentation patterns with plausible sequences. Molecular networking highlights weak but chemically consistent signals; taken together, these tools expand the thermoactinamide family and highlight flexible and potentially iterative behaviour within specific NRPS modules. The story then moves on

to synthesis, not simply as a source of material, but as proof of truth. Under widely used PyBOP-mediated macrocyclization conditions, the synthesized thermoactinamide does not match the natural compound. NMR comparison, enhanced Marfey analysis, and LC-HRMS demonstrate that the macrocyclization reaction reverses the α stereogenic centre of the activated residue during ring closure, thus questioning the paradigm of chemical synthesis as the final proof of structure. This discovery accounts for the discrepancies present in the literature and has broader implications: if macrocyclization may alter the stereochemistry of the target, then an entire structure-activity campaign risks being built on epimeric surrogates.

The second section presents the discovery and determination of the stereochemical structure of cyanochelin B, a photolabile siderophore produced by a filamentous cyanobacterium. High-resolution mass spectrometry and comprehensive 1D/2D NMR establish the planar structure. Genome mining identify a hybrid NRPS-PKS cluster whose predicted adenylation specificities matched the observed structure and the unusual presence of β-hydroxyaspartate. The absolute configuration of the molecule is elucidated using Marfey derivatization and NMR-based Murata method, complemented with bioinformatic analysis.

Functionally, the Fe(III) complex of cyanochelin B undergoes rapid photolysis under UV-A light, fragmenting into a small set of defined products while simultaneously photoreducing iron from Fe(III) to Fe(II). In a membrane co-culture, the siderophore privatizes iron in the dark but loses its exclusivity under illumination, allowing a non-producing competitor to capture iron and grow more rapidly. This light-dependent mechanism links

daytime irradiation to the microbial iron economy: the same molecule that imposes a monopoly at sunset becomes a channel for sharing at midday. The distribution of the biosynthetic gene cluster within related cyanobacteria, sometimes on mobile elements, suggests that this strategy is both transferable and responsive to environments where iron limitation intersects with strong light fluctuations. By uncovering a mechanistic link between photochemistry and nutrient flux, the study calls for a reinterpretation of laboratory competition experiments that ignore light regimes and proposes ecological investigations along depth and clarity gradients to map where siderophores act.

The third section extends the workflow from the lab benches to a real-world setting during an episode in which a small volcanic lake took on an unexpected red hue. A rapid detection strategy combines retrospective satellite screening with targeted UAV surveys to locate and quantify surface heterogeneity. Field teams then collect samples for microscopy, amplicon sequencing, and untargeted metabolomics. Feature-based molecular networking quickly organizes the resulting spectra, revealing that the bloom's metabolome is dominated by anabaenopeptins, including several variants not previously reported at the site. This chemical fingerprint is not confined to the lake: hydrological connectivity carries the signal downstream to a coastal area with shellfish aquaculture, transforming an optical anomaly into a food safety issue. The case demonstrates that multiscale detection combined with rapid chemotyping can reduce the time between detection and assessment.

Overall, the thesis proposes a practice rather than a single technique. It demonstrates that the search for novelty should start with families and

biosynthetic expectations; that synthesis should be used not only to create molecules, but also to test our structural claims; and that environmental questions are best answered when optics and omics engage in dialogue with each other. From a methodological point of view, this approach reduces friction in research and protects against reassuring but erroneous structural narratives. The broader message is that the shortest line from nature to insight is triangular, connecting genomes, networks, and spectra.

Abstract in Italian

La ricerca sui prodotti naturali si è da sempre collegata a conoscenze ecologiche e biomediche, ma i progressi spesso si arrestano al confine tra miscele complesse e certezze strutturali. Questa tesi propone e dimostra un percorso integrato che procede dai geni all'analisi metabolomica, agli spettri e, quando necessario, all'aspetto ecologico. LC-MS/MS ad alta risoluzione, NMR multidimensionale e strategie di derivatizzazione mirate costituiscono la parte sperimentale; il molecular networking concentra l'attenzione su famiglie di analoghi anziché su picchi isolati; il genome mining fornisce una razionalizzazione biosintetica che guida le assegnazioni strutturali e seleziona ciò che va isolato. Nel complesso, questa combinazione consente dereplicazione precoce, isolamento mirato più preciso e giustificazioni stereochimiche più robuste, aprendo vie per collegare le molecole alla funzione nei sistemi viventi. I risultati scientifici sono descritti nei capitoli 5, 7 e 8.

Nella prima sezione è studiata una famiglia di cicloesapeptidi lipofili prodotti da un batterio termofilo. A partire da un classico saggio antibatterico, lo studio riesamina il thermoactinamide A e rivela una più ampia costellazione di congeneri in precedenza oscurati da bassa abbondanza e sovrapposizione spettrale. Il sequenziamento genomico indica una capacità di sintesi NRPS superiore al previsto, mentre le predizioni di dominio aiutano a riconciliare i pattern di frammentazione MS/MS con sequenze plausibili. Il molecular networking evidenzia segnali deboli ma chimicamente consistenti; nel complesso, gli strumenti espandono la famiglia dei thermoactinamidi e

indicano un comportamento flessibile e potenzialmente iterativo all'interno di specifici moduli NRPS. La trattazione passa quindi alla sintesi, non solo come fonte di materiale ma come verifica strutturale: in condizioni di macrociclizzazione mediate da PyBOP, ampiamente utilizzate, il prodotto sintetico non coincide con il naturale. Confronti NMR, analisi di Marfey potenziata e LC-HRMS dimostrano che, durante la chiusura dell'anello, la reazione comune inverte il centro stereogenico α del residuo attivato, generando riproducibilmente una miscela dominata da un epimero. Questa osservazione spiega le discrepanze in letteratura e ha implicazioni più ampie: se la macrociclizzazione altera la stereochimica del bersaglio, intere campagne struttura-attività rischiano di poggiare su surrogati epimerici.

Nella seconda sezione è presentata la scoperta e la determinazione stereochimica della cianochelina B, un sideroforo fotolabile prodotto da un cianobatterio filamentoso. La spettrometria di massa ad alta risoluzione e un set completo di NMR 1D/2D stabiliscono l'architettura planare. Il genome mining identifica un cluster ibrido NRPS-PKS le cui specificità di adenilazione previste corrispondono alla struttura osservata e alla presenza inusuale di β-idrossiaspartato. La configurazione assoluta è chiarita mediante derivatizzazione di Marfey e metodo di Murata basato sull'NMR, con supporto di analisi bioinformatiche. Dal punto di vista funzionale, il complesso Fe(III) della cianochelina B subisce fotolisi rapida sotto luce UV-A, frammentandosi in un insieme limitato di prodotti definiti e fotoriducendo simultaneamente il ferro da Fe(III) a Fe(II). In co-coltura separata da membrana, il sideroforo privatizza il ferro al buio ma perde l'esclusività in luce, consentendo a un competitore non produttore di acquisire ferro e

crescere più rapidamente. Questo meccanismo dipendente dalla luce collega l'irraggiamento diurno all'economia microbica del ferro: la stessa molecola che impone un monopolio al buio diventa un canale di condivisione alla luce. La distribuzione del cluster biosintetico in cianobatteri affini, talora su elementi mobili, suggerisce una strategia trasferibile e sensibile a contesti in cui la limitazione di ferro si interseca con forti fluttuazioni luminose. Si propone una reinterpretazione degli esperimenti di competizione in laboratorio che ignorano i regimi di luce e si raccomandano indagini ecologiche lungo gradienti di profondità e trasparenza per mappare dove agiscono i siderofori.

Nella terza sezione il workflow è esteso dal laboratorio a un contesto reale durante un episodio in cui un piccolo lago vulcanico ha assunto una inaspettata colorazione rossa. Una strategia di rilevamento rapido combina screening satellitare retrospettivo e rilievi mirati con UAV per localizzare e quantificare l'eterogeneità superficiale, seguito da campionamento in loco per analisi al microscopio, sequenziamento del 16S e metabolomica non mirata. Il molecular networking organizza rapidamente gli spettri di MS/MS, rivelando che il metaboloma della fioritura è dominato da anabaenopeptine, incluse varianti non precedentemente riportate nel sito. L'impronta chimica non è confinata al lago: la connettività idrologica trasporta il segnale a valle fino a un'area costiera con mitilicoltura, trasformando un'anomalia ottica in un problema di sicurezza alimentare. Il caso mostra che la rilevazione multiscala, combinata con chemotipizzazione rapida, può ridurre il tempo tra rilevamento, valutazione e risposta.

In sintesi, la tesi propone una prassi più che una singola tecnica. Dimostra che la ricerca di nuovi metaboliti secondari dovrebbe partire da famiglie e aspettative biosintetiche; che la sintesi va impiegata anche per testare le ipotesi strutturali; e che le questioni ambientali sono affrontate al meglio quando ottiche e “omiche” sono integrate.

INTRODUCTION

Natural products have historically been a cornerstone of drug discovery, providing a rich reservoir of chemically diverse and biologically potent compounds.[1] From the isolation of morphine in 1803 to Fleming's discovery of penicillin in 1929, natural substances have produced innovative drugs and continue to be an important part of our pharmacopoeia. In fact, natural compounds possess high chemical and stereochemical diversity, specific and efficient target binding, and favourable bioactivity profiles, characteristics that make them excellent starting points in drug development.

To date, hundreds of thousands unique natural molecules have been identified, covering a wider swath of the biologically relevant chemical space than typical synthetic libraries.[2,3] The historical importance of natural products in medicine underlies the ongoing search for new bioactive molecules from nature.

Advances in biochemistry and physiology in the 19th-20th centuries made it clear that organisms produce two broad classes of metabolites: primary and secondary. Primary metabolites (e.g., sugars, fatty acids, amino acids, nucleic acids) are ubiquitous compounds essential for organism growth, development, and reproduction. They participate in basic life processes and are present in all living cells. Secondary metabolites, on the other hand, are not directly necessary for survival or growth, but often mediate specialized ecological interactions that provide a selective advantage to the producer. These so-called natural products (secondary metabolites) are typically

exclusive to specific lineages or species and play roles in interspecies relationships, such as defence from herbivores, predator deterrence, competition, or symbiosis.

While terrestrial plants and soil microbes provided the first drugs, in recent decades the marine environment has emerged as a treasure trove of structurally new and potent bioactive molecules.[4] Oceans cover more than two-thirds of the planet and are home to immense biodiversity ranging from simple microbes to complex invertebrates and algae. Marine organisms are exposed to unique and often extreme conditions—high salinity, variable pressure and temperature, low light, nutrient scarcity, intense competition in crowded coral reefs, etc.—which have led to the evolution of an exceptional chemical diversity not found in terrestrial organisms.

Over time, harsh marine conditions have favoured the production of a wide variety of secondary metabolites with unusual molecular scaffolds, high stereochemical complexity, and powerful biological activities.

As a result, marine natural products (MNPs) often have completely different structures than terrestrial natural products, expanding the chemical space for drug discovery.

The extent of marine chemodiversity is astonishing. By early 2010, researchers had catalogued about ten thousand unique marine-derived metabolites; current estimates approach forty thousand, with hundreds of new reports per year. However, the path from discovery to therapeutic approval remains arduous. To date, only fifteen to twenty marine-derived drugs have come to market, including cytarabine for leukaemia, ziconotide for intractable pain, and trabectedin for soft tissue sarcoma—a success ratio of about one

approval for every two thousand isolated compounds.[5] Though modest, these triumphs underscore both the promise and the formidable challenge of translating marine natural products into drugs.

Many characteristic marine drugs do not originate from the macroscopic host, but from embedded or free-living microorganisms.[6] The realization that sponges, corals, and tunicates often harbour microbial symbionts responsible for their metabolites has shifted the focus of marine biodiscovery to bacteria and fungi. For example, actinomycetes isolated from marine sediments, mangroves and seawater have provided dozens of new scaffolds. A six-year survey recorded nearly one hundred new compounds from thirty-eight strains of “rare” marine actinomycetes, including potent anticancer agents and antibiotics with unique modes of action.[7]

However, turning marine chemical potential into drugs comes with significant practical hurdles. Field collection often requires submersibles, remotely operated vehicles or deep-sea expeditions, all logistically complex and expensive. Collecting benthic invertebrates in sufficient quantities risks harming the environment, while yields of target metabolites from wild biomass are usually measured in parts per million. Ethical constraints, biodiversity treaties, and national sovereignty over marine genetic resources further complicate access.[8]

Microbial cultivation poses its own obstacles: many marine bacteria are recalcitrant to standard laboratory media, and even when a strain can be cultured, the environmental cues-nutrient limitation, interspecies signalling, osmotic stress-that trigger biosynthesis of secondary metabolites is easily lost in vitro.

Consequently, traditional “grind and find” discovery has been slow and burdened by frequent rediscovery of known compounds.

Advances across multiple disciplines are, however, transforming this landscape. High-density fermentation and metabolic engineering platforms provide scalable production routes for microbial metabolites.

Analytical innovations, ultra-high resolution mass spectrometry, cryoprobe multidimensional NMR, microcrystal electron diffraction, permit structural elucidation from microgram samples, making low yield metabolites accessible to rigorous characterization.[9,10]

Meanwhile, genome sequencing has unveiled an unexpected treasure trove: marine microbial genomes may encode twenty, thirty, or more biosynthetic gene clusters, the vast majority silent under laboratory conditions. Bioinformatic “genome mining” pipelines detect those clusters, predict core scaffolds, and guide targeted elicitation strategies.[11]

Coupling such predictions with the “One Strain, Many Compounds” (OSMAC) approach, systematically varying salinity, nutrients, and co cultivation partners, has awakened cryptic pathways, revealing metabolites that classical screens missed.[12,13]

Mass spectrometry-based molecular networking further accelerates dereplication by organizing data from complex extracts into families of related molecules so that never-before-seen analogues can be instantly recognized.[14]

Synthetic and systems biology extend these advantages by moving entire gene sets into heterologous hosts or inserting strong promoters to activate silent operons in the native producer. These converging technologies are dissolving

long-standing barriers of supply and accessibility, enabling the transition from oceanic discovery to clinical development with unprecedented speed.

In summary, it can be said that over the past two decades, research on natural products has quietly but surely transformed. The classic "top-down" pathway-collect a sample, extract all that it contains, screen the crude mixture for biological activity, and then isolate the responsible molecule-remains an essential art of natural products chemistry; however, on its own, this empirical pathway struggles with rediscovery, the intractability of many cultured marine microbes, and the ecological and logistical costs of collecting large amounts of biomass from the sea.

Marine microbiology, meanwhile, has been revolutionized by affordable whole-genome sequencing, which has revealed a vast underground world of biosynthetic gene clusters that are silent under normal laboratory conditions. As a result, the field has moved toward an integrative paradigm in which genetic prediction, metabolic elicitation, high-resolution metabolomics, and conventional spectroscopy continuously dialogue rather than operate in isolation.[15]

In this context, the research conducted during this PhD project operated at this methodological intersection. The three studies presented demonstrate how this integrated approach enables the discovery of chemical spaces that would otherwise remain hidden.

The first case, which concerns the lipophilic cyclic peptide thermoactinoamide from *Thermoactinomyces vulgaris*, started in the most traditional way: a zone of antibacterial inhibition on an agar plate led to the isolation of a novel cyclohexapeptide. Genome sequencing revealed that *T.*

vulgaris harbours a single, multimodular NRPS operon whose capacity exceeds that implied by the initially detected thermoactinoamide. In fact, bioinformatic inspection of that cluster, followed by an MS/MS-based molecular network, revealed several novel congeners.

Separately, upon externalization of the chemical synthesis to obtain material for drug testing, NMR and Marfey analyses indicated that the PyBOP-mediated cyclization step unexpectedly epimerized a residue, affording the α-epimer of the natural product.

The discovery of cyanochelin B followed an opposite trajectory. Analysis of the genome of a filamentous cyanobacterium, *Leptolyngbya* sp., uncovered a hybrid NRPS-PKS cluster whose domain architecture included a β-hydroxy-aspartate-containing siderophore. Thinking that the pathway should activate under iron-starved conditions, the organism was grown in Fe-depleted medium and, as hoped, the appearance of a high-mass ion was observed, which the MS/MS network placed in a completely new chemical family. NMR spectroscopy and advanced Marfey analysis established the complete structure and absolute configuration; photochemical studies then revealed an ecologically surprising behaviour: upon mild UVA irradiation, the iron-cyanochelin complex fragments, releasing soluble Fe^{2+} and thus short-circuiting the iron monopoly that siderophores usually impose.

A third study broadens the lens to include the ecosystem. In 2022, a vibrant red bloom of *Planktothrix rubescens* coloured Lake Avernus, a volcanic crater lake on the Phlegrean coast. Combining Sentinel-2 satellite images, drone-led multispectral surveys, and in situ surveys, we followed the progression of the bloom from the lake basin through a short river conduit to

the adjacent Tyrrhenian Sea and finally into a coastal mussel farm. Targeted LC-HRMS/MS analyses confirmed that toxic anabaenopeptins produced in the lake accumulated weeks later in mussel tissues.

The results obtained during the Ph.D. program have been reported in three published articles.

1. Di Matteo V, Esposito G, Costantino V, Della Sala G, Teta R, Mangoni A. When Synthesis Gets It Wrong: Unexpected Epimerization Using PyBOP in the Synthesis of the Cyclic Peptide Thermoactinoamide A. *J Nat Prod.* **2024**; 87(4):948-953.
2. Esposito G, De Rosa T, Di Matteo V, Ciccarelli C, Ajaoud M, Teta R, Lega M, Costantino V. Bio-tracking, bio-monitoring and bio-magnification interdisciplinary studies to assess cyanobacterial harmful algal blooms (cyanoHABs)' impact in complex coastal systems. *Sci Total Environ.* **2025**; 978:179480.
3. Falcao, B. P.,* Di Matteo, V.,* Hrouzek, P., Štenclová, L., Urajová, P., Mareš, J., et al. Cyanochelin B: a cyanobacterium-produced siderophore with photolytic properties that negate iron monopolization in UV light. *Appl Environ Microbiol.* **2025**; 0:e02566-24.

References

1. Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; Supuran, C. T. Natural Products in Drug Discovery: Advances and Opportunities. *Nat. Rev. Drug Discov.* **2021**, 20, 200–216, doi:10.1038/s41573-020-00114-z.

2. Sorokina, M.; Steinbeck, C. COCONUT Online: Collection of Open Natural Products. *J. Cheminf.* **2021**, 13, 2, doi:10.1186/s13321-020-00478-9.

3. Rutz, A.; Sorokina, M.; Galgonek, J.; Mietchen, D.; Willighagen, E.; Gaudêncio, S. P.; Steinbeck, C. The LOTUS Initiative for Open Knowledge Management in Natural Products. *Sci. Data* **2022**, 9, 137, doi:10.1038/s41597-022-01243-0.

4. Carroll, A. R.; Copp, B. R.; Davis, R. A.; Keyzers, R. A.; Prinsep, M. R. Marine Natural Products (2022). *Nat. Prod. Rep.* **2024**, 41, 1057–1166, doi:10.1039/d4np00005c.

5. Haque, N.; Profeta, R.; Fabbrocini, A.; Lauro, G.; Limatola, A.; Evidente, A.; Menna, M.; Campiani, G.; Novellino, E.; Di Micco, S.; et al. Marine Natural Products in Clinical Use. *Mar. Drugs* **2022**, 20, 528, doi:10.3390/md20080528.

6. Piel, J. A Polyketide Synthase–Peptide Synthetase Gene Cluster from an Uncultured Bacterial Symbiont of Paederus Beetles. *Proc. Natl. Acad. Sci. U.S.A*. **2002**, 99, 14002–14007, doi:10.1073/pnas.212163199.

7. Subramani, R.; Sipkema, D. Marine Rare Actinomycetes: A Promising Source of Structurally Diverse and Unique Novel Natural Products. *Mar. Drugs* **2019**, 17, 249, doi:10.3390/md17050249.

8. Romano, G.; Costantini, M.; Sansone, C.; Lauritano, C.; Ruocco, N.; Ianora, A. Marine Organisms: Potential Sources of Bioactive Compounds for Biomaterials and Therapy—Overcoming the Supply Problem. *Mar. Drugs* **2022**, 20, 219, doi:10.3390/md20040219.

9. Ma, J.; Li, L.; Guan, X.; Zhao, M. UHPLC–HRMS for Structure Elucidation and Complex Sample Profiling. *Pharmaceutics* **2022**, 14, 1223, doi:10.3390/pharmaceutics14061223.

10. Hilton, B. D.; Martin, G. E.; Pinto, P. A.; et al. Microcryoprobe NMR: Pushing the Limits of Small-Sample Natural Product Structure Elucidation. *J. Nat. Prod.* **2010**, 73, 1171–1176, doi:10.1021/np100121b.

11. Medema, M. H.; Fischbach, M. A. Computational Approaches to Natural Product Discovery. *Nat. Chem. Biol.* **2015**, 11, 639–648, doi:10.1038/nchembio.1884.

12. Romano, S.; Jackson, S. A.; Patry, S.; Dobson, A. D. W. Extending the "One Strain Many Compounds" (OSMAC) Approach: A Deep Dive into Marine Microbes. *Mar. Drugs* **2018**, 16, 244, doi:10.3390/md16070244.

13. Peng, X.; Zhang, Y.; Bai, J.; et al. Co-Culture as a Powerful Elicitation Strategy to Activate Silent Biosynthetic Gene Clusters in Microorganisms. *Microorganisms* **2021**, 9, 720, doi:10.3390/microorganisms9040720.

14. Nothias, L. F.; Petras, D.; Schmid, R.; et al. Feature-Based Molecular Networking in the GNPS Analysis *Environment. Nat. Methods* **2020**, 17, 905–908, doi:10.1038/s41592-020-0933-6.

15. Luo, Y.; Cobb, R. E.; Zhao, H. Recent Advances in Natural Product Discovery. *Curr. Opin. Biotechnol.* **2014**, 30, 230–237, doi:10.1016/j.copbio.2014.09.002.

Chapter 1

Structural elucidation methods

Before the era of modern spectroscopy, researchers had to rely on purely chemical manipulations, stepwise degradation or derivatization of functional groups, to deduce the complete architecture of newly discovered organic molecules. These approaches were destructive and required relatively large amounts of material, often tens to hundreds of milligrams. Although these classical methods enabled chemists to work out the structures of several remarkably complex but abundant natural products, they were not practical for the large number of metabolites produced only in trace amounts (at the microgram level).

Today, continued advances in spectroscopic instrumentation have transformed structure elucidation. Mass spectrometry (MS) and nuclear magnetic resonance (NMR) now allow chemists to unravel connectivity and stereochemistry from just a few micrograms of sample. In the investigations described in this thesis, MS and NMR are the main analytical pillars. Optical rotation measured at the sodium D-line (589 nm) provides additional information on chirality, while circular dichroism (CD) provides complementary data for assigning absolute stereochemistry. For amino acid residues, absolute configuration was established by selective hydrolysis followed by Marfey derivatization, an elegant microscale chemical method that remains indispensable even in the age of spectroscopy.

The following sections present a detailed description of these analytical techniques, highlighting recent improvements and updated literature for each method.

1.1 Mass spectrometry

Structural elucidation of a new natural product usually begins with the assignment of its elemental composition, obtained by high-resolution mass spectrometry (HR-MS). In a mass spectrometer, the analyte is converted into ions in the gas phase, and the instrument records the mass-to-charge (*m/z*) ratios of these ions. Three modules are essential: an ion source that generates ions, a mass analyser that separates them by *m/z*, and a detector that converts the ion flux into an electrical signal.

Because detection depends entirely on the presence of stable gaseous ions, the ionization step is critical; depending on the instrument, molecules may arrive pre-charged or acquire charge within the source.

In this study, compounds were examined using electrospray ionization (ESI) (Figure 1.1) coupled with an Orbitrap analyser (Figure 1.2).

Electrospray ionization is often interfaced with high-performance liquid chromatography (HPLC), allowing in-line separation of complex mixtures before they enter the mass spectrometer.

Operating at atmospheric pressure and considered a “soft” technique, ESI gently transfers analytes from solution to the gas phase with minimal fragmentation. A dilute solution of the compound in a polar, volatile solvent, typically water, methanol or acetonitrile, is pushed through a metal capillary set at 3-5 kV. The high potential disperses the liquid into a plume of charged microdroplets, which are desolvated by a stream of heated nitrogen. As each

droplet shrinks, the surface charge density increases until the electrostatic repulsion exceeds the surface tension at the Rayleigh limit; at this point the droplet undergoes Coulomb fission and releases multiply charged ions such as $[M + H]^+$ or $[M - H]^-$, directly into the gas phase. These intact molecular ions retain the true molecular weight of the analyte and are directed to the mass analyser.

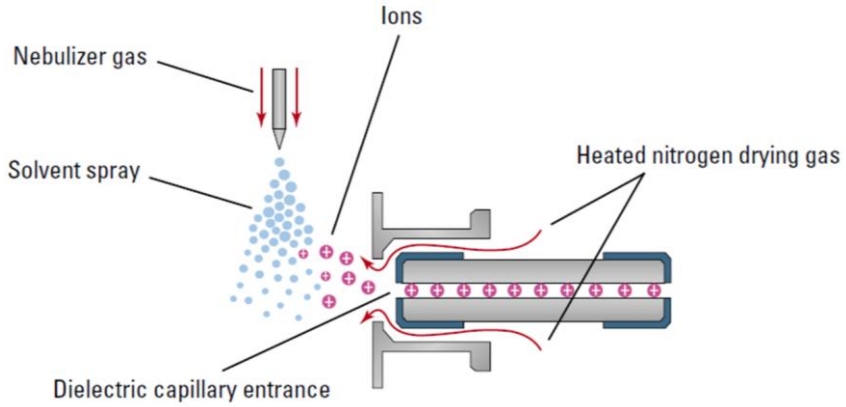


Figure 1.1. Representation of an ESI source

For complex extracts, the ESI source is routinely coupled online to high-performance liquid chromatography (LC-ESI-MS). The chromatographic phase separates co-eluting constituents, allowing for cleaner spectra and more confident assignments. Once ions enter the analyser region, their precise *m/z* values are measured with an Orbitrap detector, an electrostatic trap invented by Alexander Makarov in the early 2000s. The ions injected into the Orbitrap spiral around a central electrode while oscillating axially; the frequency of the oscillation is inversely proportional to the square root of *m/z*. A fast Fourier transform (FFT) of the image current produces a mass spectrum with resolution powers exceeding 100,000 and a typical mass error of less than 1 ppm.

The instrument used in this case is an LTQ-Orbitrap, which combines a linear ion trap with the Orbitrap detector. The linear trap acts as a first mass filter and collision cell for tandem MS (MS/MS) experiments. In an MS/MS workflow, a precursor ion selected in the trap collides with an inert gas such as argon or nitrogen, fragmenting by collision-induced dissociation (CID) or high-energy collision dissociation (HCD).

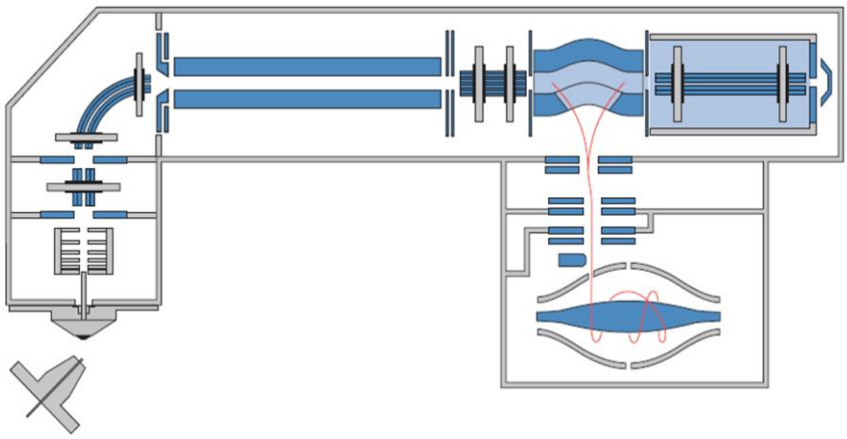


Figure 1.2. Representation of an LTQ-Orbitrap

The resulting ions are transferred to the Orbitrap, where their *m/z* values are measured with the same high precision as the precursor. Interpretation of these fragmented ions (neutral losses, diagnostic cleavages, and characteristic substructures) reveals functional groups and connectivity patterns, greatly accelerating dereplication and de novo structural determination.

Together, LC-ESI-Orbitrap MS and its tandem-MS capability provide three pillars of information from sub-microgram samples: (1) an exact molecular mass that locks in the elemental formula, (2) isotopic patterns and adduct distributions that corroborate that formula, and (3) fragmented ion data that illuminate the skeleton of the molecule.

Such sensitivity and precision are indispensable in natural product chemistry, where target compounds are often low in abundance and new molecular scaffolds are the norm rather than the exception.[1]

1.2 Nuclear Magnetic Resonance

Nuclear magnetic resonance spectroscopy (NMR) is indispensable for clarifying the detailed structure of organic molecules. When a sample containing active spin nuclei, most commonly ^{1}H and ^{13}C, is placed in a strong static magnetic field and irradiated with radiofrequency energy, these nuclei absorb at resonance frequencies that depend on both the field strength and their immediate electronic environment. Careful analysis of these chemical shifts, together with the fine structure resulting from scalar (*J*) couplings, reveals how atoms are connected and what functional groups surround them.

A one-dimensional (1D) ^{1}H spectrum typically shows discrete peaks for protons in different environments, and a ^{13}C spectrum does the same for carbons; however, spectral crowding in complex molecules often makes complete assignment impossible with 1D means alone. Modern structure determination therefore relies heavily on two-dimensional (2D) experiments, which distribute the signals over two frequency axes and correlate them through specific nuclear interactions, greatly alleviating peak overlap.

In routine workflows for structure determination, a sequence of complementary 2D NMR experiments is applied. The starting point is usually the homonuclear COSY (Correlation Spectroscopy) spectrum, whose off diagonal crossed peaks reveal pairs of scalarly coupled protons through two

or three bonds; these correlations establish the basic proton spin systems and trace short-range connectivity.

The magnetization can then be propagated further with a TOCSY (Total Correlation Spectroscopy) experiment: within a given spin network, each proton becomes correlated with all others, a particularly powerful feature for repetitive motifs such as peptide chains or oligosaccharide chains, where TOCSY conveniently collects all resonances belonging to each individual subunit.

Then, attention then shifts to heteronuclear experiments. An HSQC (Heteronuclear Single-Quantum Correlation) spectrum correlates each proton with the carbon directly bonded via the $^{1}J_{CH}$ coupling to a bond, thus coupling the ^{1}H and ^{13}C chemical shifts and distinguishing CH, CH_2, and CH_3 groups immediately.[2]

Finally, a HMBC (Heteronuclear Multiple-Bond Correlation) experiment detects longer-range $^{2}J_{CH}$ and $^{3}J_{CH}$ couplings, linking the fragments defined by COSY and HSQC into an unbroken carbon structure and, crucially, assigning the quaternary carbons without attached protons.[3]

It is often necessary to optimize experimental delays and band-selective HSQC or HMBC sequences, which restrict radio frequency excitation to narrow spectral windows, can refine resolution and recover crowded regions without introducing aliasing.

Together, this series of complementary 1D and 2D experiments allows chemists to assign nearly all proton and carbon resonances, establish atom-atom connectivity, and even extract stereochemical information through analysis of coupling constants and NOE studies (see Section 1.3). Advances

such as ultra-high-field magnets (≥500 MHz for ^{1}H) and cryogenically cooled probes have brought NMR sensitivity to a level that makes complete elucidation possible even when only a few micrograms of sample are available.

1.3 Methods for the determination of relative/absolute configurations

Determining the relative and absolute stereochemistry of a natural product is a decisive step in its complete structural characterization. Stereochemical information is essential not only for understanding how a molecule interacts with biological receptors, but also for guiding synthetic chemists seeking to replicate or modify complex architectures. Since most natural products have one or more stereogenic centres, NMR spectroscopy, together with chiroptical and derivatization methods, forms the backbone of modern stereochemical analysis.

In NMR, three experimentally accessible quantities are exploited to probe relative configuration: chemical shifts (δ), scalar coupling constants (*J*), and nuclear Overhauser enhancements through space (NOE) or Overhauser in rotating systems (ROE). Chemical shifts can occasionally differentiate diastereoisomers because protons in distinct stereochemical environments often resonate at slightly different frequencies. Unfortunately, these shifts depend on a multitude of subtle electronic and conformational variables, so they rarely provide a direct route to configuration assignment in an unknown structure.

Much more diagnostic are vicinal coupling constants. The amplitude of a $^{3}J_{HH}$ coupling varies with the dihedral angle θ between the coupled protons in

accordance with Karplus's relation: values close to 0-1 Hz are observed when θ is close to 90°, medium couplings (≈2-6 Hz) occur around 60° (gauche), and large couplings (≈8-14 Hz) indicate antiperiplanar arrangements close to 180°.[4]

In flexible aliphatic chains, this information can indicate whether adjacent stereogenic centres assume trans or cis relationships, while in six-membered rings they differentiate between axial-axial (large *J*), axial-equatorial (medium *J*), and equatorial-equatorial (small *J*) orientations.

The same principle distinguishes the geometries of alkenes, where $^{3}J_{trans}$ couplings (≈14–18 Hz) far exceed $^{3}J_{cis}$ values (≈6–12 Hz). Heteronuclear couplings ($^{3}J_{CH}$) follow similar trends and can reinforce these conclusions.

Complementing the bond data, NOE correlations reveal which protons are located at a spatial distance of approximately 5 Å from each other. Selective irradiation of a proton alters the intensity of nearby signals via cross-dipole relaxation; for small molecules, this manifests as a positive increase, while for large macromolecules, it becomes negative.

The two-dimensional NOESY experiment simultaneously records all these interactions, allowing chemists to map which substituents share the same molecular face or approach each other in bent conformations. For compounds in the intermediate molecular weight range, where conventional NOEs weaken, the ROESY experiment offers an alternative: ROE enhancements are always positive and therefore remain observable when NOEs pass through zero. ROESY is therefore ideal for medium-sized natural products, large enough to dampen standard NOEs but too small to show the characteristic negative NOE of proteins.[5]

It should be noted that NMR, being inherently a time-averaging technique, provides information that often represents an average of all conformations populated by the molecule in solution. For rigid structures (e.g., many cyclic systems), a single conformation dominates, and NMR parameters can be directly correlated to a fixed geometry. In contrast, flexible molecules (such as many acyclic chains or macrocycles) can adopt multiple conformations in solution that rapidly interconvert on the NMR time scale. In these cases, coupling constants and NOE distances may be averaged values, which can complicate stereochemical interpretation. Traditional approaches (simple Karplus analysis and qualitative NOE observation) can give ambiguous or misleading results if multiple conformers contribute.

This challenge has stimulated the development of specialized methods for stereochemical determination in flexible systems, one of which is the Murata method, discussed in detail in Chapter 1.3.3.

Finally, it is important to distinguish relative configuration from absolute configuration. NMR methods (couplings and NOE) generally provide information on relative stereochemistry, i.e., how chiral centres are arranged relative to each other (e.g., all cis relative to mixed trans/cis in a polyol chain). NMR alone cannot determine the absolute configuration (the overall "handedness" of the molecule) because the NMR spectrum of an enantiomer is superimposable on that of its mirror image. As a non-chiral technique, NMR will treat enantiomers identically in an achiral environment.

To assign the absolute configuration, it is necessary to introduce a chiral reference (e.g., by derivatization with a chiral reagent so that the enantiomers give NMR-distinguishable diastereomeric derivatives. The Mosher method is

a classic example) or to use other chiral and chemical methods described below.

1.3.1 Electronic circular dichroism (ECD)

Electronic circular dichroism spectroscopy (ECD), previously known simply as "circular dichroism (CD)" until the advent of vibrational circular dichroism (VCD), which necessitated clearer terminology, is now one of the most widely used chiral tools for determining the absolute configuration of natural products. In an ECD experiment, a chiral molecule is irradiated with circularly polarized ultraviolet or visible light, and the instrument records the difference in absorption ($\Delta A = A_L - A_R$, often reported as molar circular dichroism $\Delta\varepsilon$) across the wavelength. Since the electric field vectors of the two circular polarizations trace helices in opposite directions, a chiral chromophore will interact more strongly with one than the other, producing positive or negative Cotton effects depending on the wavelength in the ECD spectrum. Enantiomers therefore give mirror-image traces, a feature that makes ECD exceptionally useful for distinguishing them.

A prerequisite for this technique is the presence of a UV/Vis chromophore. Many natural products contain intrinsic π systems (e.g., aromatic rings, conjugated carbonyls) that absorb in the required region, but even molecules lacking such groups can be made active for ECD by adding an auxiliary chromophore via derivatization. Classic examples are Mosher-type esters and exciton-coupled bis-chromophore adducts, in which an added benzoate, naphthoate, or similar aromatic group introduces absorption and locks in a stereochemically sensitive conformational relationship.[6] Typically, only a

few micrograms of material are required (modern detectors can even approach the sub-microgram range), so ECD is perfectly suited to scarce metabolites.

Translating an experimental ECD curve into an absolute stereochemical assignment is less straightforward than simply recording it, because the sign and amplitude of each Cotton effect depend in a complex way on the electronic structure and three-dimensional arrangement of the chromophores. Empirical guidelines exist for some functional groups: the octet rule rationalizes the ECD of n→π* transitions in carbonyl compounds, while the chiral exciton rule applies when two chromophores interact through space. However, most natural products do not fall within the scope of these heuristics. The current solution is to compare the experimental spectrum with the spectra predicted for all plausible stereoisomers using quantum chemical methods, most commonly time-dependent density functional theory (TD-DFT). The workflow involves generating a conformational set for each candidate, calculating the rotational forces for the lowest-energy conformers, Boltzmann weighting these contributions, and summing them to obtain a simulated ECD trace. The calculated curve that best matches the experimental data identifies the correct absolute configuration with a high degree of reliability; discrepancies in the sign, shape, or position of the peak quickly eliminate incorrect stereochemical models.

Because ECD combines nanomolar sensitivity with broad chromophore compatibility and a solid theoretical framework for interpretation, it has become the method of choice for absolute configuration determination in natural product chemistry. When used in conjunction with other

spectroscopic and computational tools, ECD provides a rapid and unambiguous path from trace sample to fully resolved three-dimensional structure.[7]

1.3.2 Computational methods

Although careful analysis of vicinal coupling constants and NOESY/ROESY contacts often provides a reliable picture of the relative stereochemistry of an organic molecule, interpretation of NMR data alone can be ambiguous, especially for flexible skeletons where multiple conformers contribute to the observed averages. In this case, computational chemistry offers a powerful complement. By applying the laws of quantum and statistical mechanics with modern software, professionals can model molecular geometry, energetics, NMR parameters, and chiral responses, thereby verifying each stereochemical assumption against explicit theoretical predictions.

Computational approaches fall into two broad categories. Molecular mechanics (MM) methods use classical force fields in which atoms behave as masses connected by springs; they are fast and capable of sampling large systems, but their accuracy depends on well-parameterized potentials. Quantum mechanics (QM) methods deal directly with the electronic structure. QM calculations are more demanding, but have become accessible on PCs for small-to-medium-sized molecules. They can reproduce subtle electronic and steric effects with high fidelity, making them valuable for predicting NMR chemical shifts, coupling constants, electronic CD spectra, optical rotation, and other observables that depend sensitively on electron distribution.

A typical workflow begins with an exhaustive conformational search. Systematic torsional scans, stochastic searches, or molecular dynamics simulations (often accelerated by simulated annealing) generate a diverse set of candidate geometries. Each geometry is then minimized energetically to its nearest local minimum; since minimization alone may fall into a local rather than global well, the preliminary conformer search is critical to ensure coverage of all low-energy regions. The resulting set of conformers, weighted according to their Boltzmann populations at room temperature, represents the behaviour of the molecular solution.

With this set available, researchers calculate the properties most relevant for stereochemical discrimination. NMR parameters are routinely predicted: ^{1}H and ^{13}C chemical shifts can be calculated for each conformer and averaged to provide a theoretical spectrum for each candidate stereoisomer. Statistical tools such as DP4 or DP4+ probability analysis then quantify the match between each calculated data set and the experiment, often identifying the correct configuration with high reliability. Similarly, the vicinal coupling constants ($^{3}J_{HH}$, $^{3}J_{CH}$) derived from DFT geometries can be compared directly with those measured in the laboratory, providing an orthogonal verification of the relative stereochemistry.

When NMR data prove inconclusive, chiroptic properties come into play. Time-dependent DFT (TD-DFT) calculations of circular electronic dichroism, optical rotation, or vibrational CD allow direct comparison between simulated and experimental spectra for each enantiomeric or diastereomeric model. The stereostructure who's calculated ECD curve best

reproduces the sign, position, and amplitude of the experimental Cotton effects typically emerges as the correct absolute configuration.

Since the calculated results always depend on the theoretical level and the chosen basis set, they are interpreted together with the experimental evidence rather than in isolation. However, even with these caveats, the synergy between experiment and calculation has transformed stereochemical analysis. For complex or conformational mobile natural products, quantum-assisted NMR and ECD calculations now routinely resolve ambiguities that once required complex degradation studies or total synthesis. With the continued growth in computing power and the increasing accessibility of software, computational chemistry is set to remain an indispensable partner for clear elucidation.

1.3.3 Murata method

As discussed in Chapter 1.3., NMR coupling constants and NOE correlations are powerful indicators of stereochemistry in rigid molecules. However, flexible acyclic molecules pose a particular challenge for stereochemical determination. Many bioactive natural products, such as polyketides and other linear or macrocyclic compounds, have multiple contiguous stereocentres but also considerable conformational freedom. In such cases, standard NMR approaches (analysis of $^{3}J_{HH}$, $^{3}J_{CH}$, and NOE data) often provide ambiguous or mediated results because the molecule does not assume a fixed conformation.

To solve this problem, Murata and colleagues introduced an ingenious NMR-based method to determine the relative configuration of adjacent chiral

centres in flexible acyclic systems using homo- and heteronuclear *J* coupling information. They recognized that even in such flexible structures, conformational freedom is not unlimited, but conformations can be represented by three staggered rotamers. When each stereogenic centre carries an oxygenated substituent (a hydroxyl, an alkoxy, or a methoxy adjacent to the carbonyl, as found in all polyketides), gauche and anomeric effects often cause one of these staggered rotamers to predominate.[8] Considering that there are two possible relative configurations between the two chiral centres (usually defined as *erythro* or *threo* in this context) and three possible conformers for each of them, six possibilities exist. They can be distinguished by measuring the vicinal coupling ($^3J_{HH}$) between the two protons on the stereogenic centres and the two-bond ($^2J_{CH}$) and the three-bond ($^3J_{CH}$) couplings of these protons.

The heteronuclear couplings can be measured using standard phase-sensitive HMBC spectra or, when greater quantitative accuracy is required, more sophisticated sequences such as IPAP-HSQMBC or J-resolved. Rather than requiring accuracy to the nearest hertz, however, Murata's protocol qualitatively classifies each coupling. A vicinal $^3J_{HH}$ coupling of approximately 8-10 Hz, or a vicinal heteronuclear $^3J_{CH}$ coupling greater than approximately 4 Hz, is considered “large” and is characteristic of an anti (≈180°) dihedral angle between the coupled nuclei, while values between 2 and 4 Hz are “small” and are indicative of a gauche (≈60°) dihedral angle. In addition, the value of a two-bond heteronuclear $^2J_{CH}$ coupling is sensitive to the geometrical relationship between the proton involved and the oxygenated substituent, an anti relationship leading to a small coupling and a gauche

relationship leading to a large coupling. Once all the coupling information is available, the analyst only has to decide which of the six possibilities is supported by the measured data; the answer follows immediately.

The power of this double-coupling fingerprint became apparent in Murata's original study on 1,3-diols and amino alcohols and has since been applied to a wide variety of oxygen-rich polyketides, including zooxanthellatoxin fragments, crocacines, and the polyether archazolid B. Success is greatest when a single rotamer dominates the equilibrium, precisely the situation expected for vicinal oxygen stereocenters, because medium-sized couplings can confound the large/small distinction.

Where the method reaches its limits, researchers typically reinforce it with Kishi's empirical ^{13}C shift correlations, which are less sensitive to conformational averaging, or with density functional calculations that predict Boltzmann-weighted *J* values for each candidate stereoisomer.[9] Such computational integrations have proven decisive for highly flexible macrocycles and for cases where experimental couplings oscillate in the intermediate regime.

Over the past two decades, advances in pulse sequence design have lowered the practical detection limit for long-range couplings, while quantitative NOE accumulation experiments and, more recently, residual dipole coupling measurements have provided complementary constraints. Despite these refinements, the conceptual core remains intact: by exploiting the intrinsic conformational bias imposed by heteroatom substituents, Murata's method extracts stereochemical information that would otherwise be obscured by motor mediation. Its economy of experimental effort, two couplings from a

single HMBC acquisition, has made it a standard first-line tactic in modern natural product chemistry, simplifying the stereochemical assignment of flexible acyclic fragments that once required much more elaborate manoeuvres.

1.3.4 Marfey's method

Among the few classic chemical protocols that continue to rival modern spectroscopic and computational tools, Marfey's derivatization remains the method of choice for uniquely assigning the L or D series of amino acid residues in peptides and other natural products.

Introduced in 1984, the method converts enantiomeric amino acids, which would otherwise be inseparable on a standard achiral HPLC column, into diastereomeric derivatives that exhibit different chromatographic behaviours. The reagent that causes this transformation is 1-fluoro-2,4-dinitrophenyl-5-L-alaninamide, universally known as L-FDAA.[10]

The workflow begins with a total acid hydrolysis of the peptide, typically by heating in 6 N HCl for twenty-four hours, to release the free amino acids.

The hydrolysate is then neutralized and treated with an excess of L-FDAA in a slightly basic medium. Each amino group replaces the fluorine atom of the dinitrophenyl ring by aromatic nucleophilic substitution, producing an L-FDAA-amino acid adduct. An L amino acid gives rise to L-FDAA-L-AA, while a D residue gives rise to L-FDAA-D-AA; the two products are diastereoisomers and therefore have distinct retention times on a reversed-phase column (C18).

To interpret these retention times, reference derivatives are prepared. The simplest approach is to derivatize authentic L and D standards of each amino acid with L-FDAA, generating the two diastereoisomers.

If a D standard is not available, an equally valid alternative is to derivatize the L standard twice, once with L-FDAA and once with its mirror image reagent D-FDAA. Since D-FDAA-L-AA is the enantiomeric counterpart of L-FDAA-D-AA, in both cases the pair of diastereomers required for comparison is obtained.

Analytical separation is usually performed by RP-HPLC with UV detection (the dinitrophenyl chromophore absorbs strongly) or, for maximum sensitivity and selectivity, by LC-MS. This reduces the sample requirement to a few micrograms. Each derivative in the hydrolysate co-elutes with one of the reference diastereomers, immediately revealing whether the original residue was L or D.

Marfey's strategy also extends to non-proteogenic amino acids for which no commercially available standards are available. A large body of empirical data shows that, under standard C18 conditions, the L-FDAA derivatives of L amino acids usually elute earlier than their L-FDAA-D analogues.[11] This systematic shift is attributed to the trans versus cis arrangement of the bulky substituents on the α carbons of the derivatized pair, which modulates the overall hydrophobicity. Exploiting this trend, the unknown amino acid is derivatized separately with L- and D-FDAA, and the elution order of the products is observed. If the L-FDAA derivative elutes first, the unknown amino acid is L; if it elutes last, the amino acid is D.

References

1. Hu, Q.; Noll, R.J.; Li, H.; Makarov, A.; Hardman, M.; Cooks, R.G. The Orbitrap: A new mass spectrometer. *J. Mass Spectrom.* **2005**, *40*, 430–443, doi:10.1002/jms.856.

2. Palmer, A.G.; Cavanagh, J.; Byrd, R.A.; Rance, M. Sensitivity improvement in three-dimensional heteronuclear correlation NMR spectroscopy. *J. Magn. Reson.* **1992**, *96*, 416–424, doi:10.1016/0022-2364(92)90097-Q.

3. Bax, A.; Summers, M.F. 1H and 13C Assignments from Sensitivity-Enhanced Detection of Heteronuclear Multiple-Bond Connectivity by 2D Multiple Quantum NMR. *J. Am. Chem. Soc.* **1986**, *108*, 2093–2094, doi:10.1021/ja00268a061.

4. Karplus, M. Contact electron-spin coupling of nuclear magnetic moments. *J. Chem. Phys.* **1959**, *30*, 11–15, doi:10.1063/1.1729860.

5. Sanders, J.K.M.; Mersh, J.D. Nuclear magnetic double resonance; the use of difference spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **1982**, *15*, 353–400, doi:10.1016/0079-6565(82)80011-3.

6. Dale, J. A.; Dull, D. L.; Mosher, H. S. α-Methoxy-α-trifluoromethylphenylacetic Acid, a Versatile Reagent for the Determination of Enantiomeric Composition of Alcohols and Amines. *J. Org. Chem.* **1969**, *34*, 2543–2549, doi:10.1021/jo01261a013.

7. Grauso, L.; Teta, R.; Esposito, G.; Menna, M.; Mangoni, A. Computational prediction of chiroptical properties in structure elucidation of natural products. *Nat. Prod. Rep.* **2019**, *36*, 1005–1030, doi:10.1039/c9np00018f.

8. Matsumori, N.; Kaneno, D.; Murata, M.; Nakamura, H.; Tachibana, K. Stereochemical Determination of Acyclic Structures Based on Carbon–Proton Spin–Coupling Constants. A Method of Configuration Analysis for Natural Products. *J. Org. Chem.* **1999**, *64*, 866–876, doi:10.1021/jo981810k.

9. Higashibayashi, S.; Czechtizky, W.; Kobayashi, Y.; Kishi, Y. Universal NMR Databases for Contiguous Polyols. *J. Am. Chem. Soc.* **2003**, *125*, 14379–14393, doi:10.1021/ja0375481.

10. Marfey, P. Determination of o-amino acids . II Use of a bifunctional. *Carlsb. Res. Commun.* **1984**, *49*, 591–596.

11. Fujii, K.; Ikai, Y.; Mayumi, T.; Oka, H.; Suzuki, M.; Harada, K.I. A nonempirical method using LC/MS for determination of the absolute configuration of constituent amino acids in a peptide: Elucidation of limitations of Marfey's method and of its separation mechanism. *Anal. Chem.* **1997**, *69*, 3346–3352, doi:10.1021/ac9701795.

Chapter 2

Molecular networking

Marine organisms are widely recognized as a rich source of structurally diverse small-molecule natural products, many with notable pharmacological activities, particularly anticancer and antibiotic effects. Early studies in marine natural products chemistry focused primarily on abundant, easily accessible coastal species. In contrast, current research targets organisms from highly specialized environments, including polar regions, tropical coral reefs, and deep-sea habitats. This expansion has significantly broadened the known chemical diversity but also intensified a persistent challenge: marine extracts are highly complex mixtures that may contain thousands of components, ranging from primary and secondary metabolites to trace contaminants. Consequently, isolating and identifying genuinely new bioactive compounds remains a demanding task.

To simplify this search, chemists rely on dereplication, which is the rapid recognition of previously reported molecules so that resources can be focused on genuine novelties. In practice, dereplication is performed most efficiently with high-resolution liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS).[2]

However, traditional approaches, which rely on laborious manual checking of LC-MS/MS traces, quickly become unmanageable when confronted with the enormous data sets generated by modern instruments.

Fortunately, rapid advances in genomics, metabolomics, and data science tools have ushered in a new era of high-throughput discovery. Among the most influential innovations is molecular networking, a computational workflow that organizes LC-MS/MS data into an intuitive visual map and can be performed by any researcher using the freely accessible Global Natural Products Social Molecular Networking (GNPS) platform.[3]

The technique exploits a simple but powerful observation: since the architecture, functional groups, and intrinsic stability of a molecule determine how it fragments in the collision cell of a mass spectrometer, two structurally similar compounds will tend to generate comparable MS/MS spectra. GNPS quantifies this similarity by aligning each pair of spectra, counting the fragment peaks that share the same *m/z* or derive from identical neutral losses, and then calculating a cosine similarity score ranging from 0 (no overlap) to 1 (perfect match). Each consensus spectrum becomes a node in a graph; nodes with high similarity scores are connected by edges, and clusters of nodes emerge that correspond to families of related metabolites.

Visualization software such as *Cytoscape* can further enrich the graph by resizing nodes based on precursor ion intensity or modulating edge thickness to reflect cosine similarity, giving the researcher an immediate overview of the chemical relationships within the extract (Figure 2.1).[4]

The main advantage of molecular networking lies in its ability to juxtapose experimental spectra with the vast libraries of reference spectra now available online, thus flagging known constituents at a very early stage and saving chemists the effort of isolating them again. Compounds that do not have exact matches but are found in the same cluster as a known metabolite can still be

recognized as close analogues, potentially new but clearly related in structure, to guide targeted isolation.

The GNPS library contains over 200,000 reference spectra, and every dataset uploaded to the platform is stored in the MassIVE repository, encouraging open data sharing within the community.[5]

GNPS also allows users to add rich metadata (organism of origin, hydrophobicity, abundance, and even bioactivity scores) to each node, greatly improving the interpretability of a network.

A suite of in silico auxiliary tools further extends its reach. For example, Dereplicator and Dereplicator+ match experimental spectra to predicted fragmentation trees and then annotate natural peptide and non-peptide products, while MolNetEnhancer overlays automated chemical class information on each cluster.[6,7]

Classic molecular networking workflows, however, neglect the chromatographic dimension; as a result, isomers that elute at different retention times may be merged into a single node, and isotopic or adduct peaks may be misinterpreted as separate entities.

The Feature-Based Molecular Networking (FBMN) approach addresses these shortcomings by first processing raw LC-MS/MS files in software such as *MZmine*.[8,9]

FBMN preserves retention time information, distinguishes isomers, correctly groups isotopes and adducts, and quantifies each feature, producing networks that more accurately reflect the chemical reality of a complex extract.[10]

Thanks to its flexibility and compatibility with a wide range of ion source technologies, molecular networking has quickly become an indispensable

tool. It accelerates the discovery of new lead compounds by simultaneously identifying known molecules and highlighting unexplored analogues.[11]

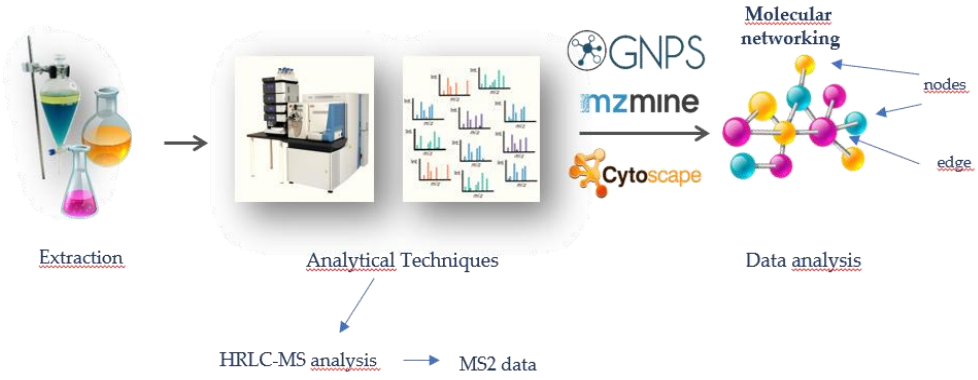


Figure 2.1. Schematic workflow of molecular networking

References

1. Gaudêncio, S. P.; Pereira, F. Dereplication: Racing to Speed up the Natural Products Discovery Process. *Nat. Prod. Rep.* **2015**, *32*, 779–810, doi:10.1039/C4NP00134F.

2. Wolfender, J.L.; Litaudon, M.; Touboul, D.; Queiroz, E.F. Innovative omics-based approaches for prioritisation and targeted isolation of natural products-new strategies for drug discovery. *Nat. Prod. Rep.* **2019**, *36*, 855–868, doi:10.1039/c9np00004f.

3. Yang, J.Y.; Sanchez, L.M.; Rath, C.M.; Liu, X.; Boudreau, P.D.; Bruns, N.; Glukhov, E.; Wodtke, A.; De Felicio, R.; Fenner, A.; et al. Molecular networking as a dereplication strategy. *J. Nat. Prod.* **2013**, *76*, 1686–1699, doi:10.1021/np400413s.

4. Shannon, P.; Markiel, A.; Ozier, O.; **et al.** Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504, doi:10.1101/gr.1239303.

5. Watrous, J.; Roach, P.; Heath, B.; **et al.** Mass Spectral Molecular Networking of Living Microbial Communities. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, E1743–E1752, doi:10.1073/pnas.1203689109.

6. Mohimani, H.; Gurevich, A.; Shlemov, A.; Mikheenko, A.; Korobeynikov, A.; Cao, L.; Shcherbin, E.; Nothias, L.F.; Dorrestein, P.C.; Pevzner, P.A. Dereplication of microbial metabolites through database search of mass spectra. *Nat. Commun.* **2018**, *9*, 1–12, doi:10.1038/s41467-018-06082-8.

7. Ernst, M.; Kang, K. Bin; Caraballo-Rodríguez, A.M.; Nothias, L.F.; Wandy, J.; Chen, C.; Wang, M.; Rogers, S.; Medema, M.H.; Dorrestein, P.C.; et al. Molnetenhancer: Enhanced molecular networks by integrating metabolome mining and annotation tools. *Metabolites* **2019**, *9*, doi:10.3390/metabo9070144.

8. Nothias, L. F.; Petras, D.; Schmid, R.; **et al.** Feature-Based Molecular Networking in the GNPS Analysis Environment. *Nat. Methods* **2020**, *17*, 905–908, doi:10.1038/s41592-020-0933-6.

9. Schmid, R.; Heuckeroth, S.; Korf, A.; **et al.** Integrative Analysis of Multimodal Mass Spectrometry Data in MZmine 3. *Nat. Biotechnol.* **2023**, *41*, (article), doi:10.1038/s41587-023-01690-2.

10. Nothias, L.F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; et al. Feature-based molecular networking in the GNPS analysis environment. *Nat. Methods* **2020**, *17*, 905–908, doi:10.1038/s41592-020-0933-6.

11. Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; **et al.** ClassyFire: Automated Chemical Classification with a Comprehensive, Computable Taxonomy. *J. Cheminf.* **2016**, *8*, 61, doi:10.1186/s13321-016-0174-y.

Chapter 3

Genome Mining

Genome mining has transformed the discovery of natural products by allowing researchers to read an organism's chemical potential directly from its DNA, rather than starting with biological tests and working backwards.[1]

In a modern top-down workflow, entire genomes and even metagenomes are scanned for clusters of biosynthetic genes (BGCs).[2] The combination of genes and domains in these loci not only signals the broad class of metabolites encoded, but often provides clues about key structural features of the final molecule. Thanks to the ready availability of high-quality genomic sequences, shared curation standards, and a mature software ecosystem, in silico predictions now regularly link gene cluster detection to plausible scaffolds and monomer compositions.[3] In practice, this means that we can make defensible hypotheses, prior to any isolation work, that a given microbe is a source of new chemistry, and then use those hypotheses to guide cultivation, heterologous expression, and analytical follow-up.

The level of data makes all this possible. Public archives such as GenBank/INSDC and institutional resources such as IMG/M serve as raw

material for reanalysis, while community curation closes the loop from genes to molecules. The MIBiG standard captures experimentally validated BGC-product pairs along with metadata such as enzyme specificities, providing a gold standard against which new predictions can be benchmarked and trained. In addition, large-scale resources such as antiSMASH-DB with its precalculated BGC calls, IMG-ABC as a large knowledge base of predicted and known clusters, and atlas-style initiatives such as BGC Atlas organize millions of BGCs into gene cluster families with environmental context. Together, these repositories support both rigorous dereplication against known chemistry and prioritization of truly novel loci.

From a detection standpoint, antiSMASH has become the workhorse.[2] Recent versions recognize an expanded range of cluster classes and make assembly pipelines modular in a way that is immediately interpretable for NRPS and PKS systems. Under the hood, hidden Markov models detect characteristic constellations of biosynthetic and adaptive enzymes, while ClusterBlast, KnownClusterBlast, and SubClusterBlast compare predicted regions with precalculated entries and MIBiG references.

These comparisons quickly highlight similarities with known clusters and flag substructures, such as desoxysaccharides, that may appear in the product. When assemblies are fragmented, as is often the case in metagenomic datasets, graphical methods such as biosyntheticSPAdes help reconstruct long, contiguously represented clusters. Classic sequence analysis tools continue to prove their worth: BLAST and multiple sequence alignment with Clustal Omega remain useful for validating gene identities, refining cluster boundaries, and inspecting conserved motifs that govern substrate selection.

From there, the prediction moves from “which cluster?” to “which molecule?”. Two families of enzymes dominate microbial biosynthesis in assembly lines: non-ribosomal peptide synthetases (NRPS) and modular polyketide synthetases (PKS). Their nearly colinear modular architecture provides a natural scaffold for inference. In NRPS, adenylation domains (A) choose and activate amino acid building blocks, while in PKS, acyltransferase domains (AT) load the extender units that make up the carbon structure.

antiSMASH annotates these modules and deduces their order of operation; specialized predictors then estimate the specificities of the domains.

Motif-based and machine learning tools such as NRPSpredictor2 and SANDPUMA help assign substrates to the A domain, while AdenylPred extends the scope to the broader family of enzymes that form adenylate, including β-lactone synthetases, capturing the biosynthetic logic that goes beyond the canonical NRPS.[4] For whole-molecule reasoning, PRISM 4 links gene content to a curated library of virtual reactions and hidden Markov models to reconstruct candidate structures in many classes of secondary metabolites (NRPS/PKS, β-lactams, nucleosides, lincosamides, RiPPs, and others). The result is typically a small set of plausible scaffolds with enumerated adaptation variants, which can then be ranked based on chemical similarity to known metabolites or orthogonal genetic evidence (e.g., the presence of resistance markers or pathway-specific regulators).

Machine learning complements these rule-based strategies. DeepBGC, for example, uses bidirectional LSTM models and word embedding representations of Pfam domains to reduce false positives and generalize to BGC classes other than those the models have seen before.[5] Currently, such

approaches are most effective at deciding whether a region is indeed a BGC and assigning general product classes, but when paired with curated resources such as MIBiG and atlas-scale cluster collections, they also provide increasingly useful information for downstream chemical inference.

Since discovery is not only about individual clusters but also about landscapes of biosynthetic potential, it is important to organize diversity. BiG-SCAPE groups BGCs into gene cluster families by comparing domain content and synteny, while CORASON provides multilocus phylogenies that reveal evolutionary relationships and highlight subclade specificities.[6]

On a scale of millions of clusters, BiG-SLiCE supports rapid assignments across large datasets, and BiG-FAM exposes a searchable universe of families spanning reference genomes. Combined with antiSMASH-DB and BGC Atlas, these tools make dereplication efficient and novelty assessment explicit, so that efforts are focused where they are most likely to yield results.[7]

Finally, sequence alone rarely allows structures to be determined with accuracy, so genomics is increasingly paired with metabolomics to triangulate gene-molecule links, a practice often referred to as metabologenomics.

Feature-based molecular networking (FBMN) in the GNPS environment, discussed earlier, organizes large LC-MS/MS datasets into molecular families and resolves isomeric features, allowing annotations to propagate across a network of related spectra.[8] Community platforms such as the Paired Omics Data Platform record validated connections between BGCs and MS/MS features, while frameworks such as NPLinker combine strain correlation scores with feature-based predictors to rank plausible matches between genes

and spectra. BiG-MAP adds another layer by profiling the abundance and expression of BGCs in metagenomes and metatranscriptomes, providing an ecological and normative context that helps decide which clusters to activate or express next. Together, these complementary lines of evidence reduce false positives, guide targeted isolation, and transform genome-based predictions into experimental hypotheses.

References

1. Medema, M. H.; Fischbach, M. A. Computational Approaches to Natural Product Discovery. Nat. Chem. Biol. **2015**, *11*, 639–648. doi:10.1038/nchembio.1884.

2. Blin, K.; Shaw, S.; Kloosterman, A. M.; **et al.** antiSMASH 6.0: Improving Cluster Detection and Comparison Capabilities. *Nucleic Acids Res.* **2021**, *49*, W29–W35. doi:10.1093/nar/gkab335.

3. Kautsar, S. A.; Blin, K.; Shaw, S.; Navarro-Muñoz, J. C.; **et al.** MIBiG 2.0: A Repository for Biosynthetic Gene Clusters of Known Function. *Nucleic Acids Res.* **2020**, *48*, D454–D458. doi:10.1093/nar/gkz882.

4. Röttig, M.; Medema, M. H.; Blin, K.; **et al.** NRPSpredictor2—A Web Server for Predicting NRPS Adenylation Domain Specificity. *Nucleic Acids Res.* **2011**, *39*, W362–W367. doi:10.1093/nar/gkr323.

5. Hannigan, G. D.; Prihoda, D.; Palicka, A.; **et al.** A Deep Learning Genome-Mining Strategy for Biosynthetic Gene Cluster Prediction. *Nucleic Acids Res.* **2019**, *47*, e110. doi:10.1093/nar/gkz654.

6. Navarro-Muñoz, J. C.; Selem-Mojica, N.; Mullowney, M. W.; **et al.** A Computational Framework to Explore Large-Scale Biosynthetic Diversity. *Nat. Chem. Biol.* **2020**, *16*, 60–68. doi:10.1038/s41589-019-0400-9.

7. Bağcı, C.; Blin, K.; Wolf, T.; Collemare, J.; Medema, M. H.; Ziemert, N.; Kautsar, S. A. BGC Atlas: A Comprehensive Resource for Exploring Biosynthetic Gene Cluster Diversity. *Nucleic Acids Res.* **2024**, *52*, D586–D594. doi:10.1093/nar/gkad984.

8. Nothias, L. F.; Petras, D.; Schmid, R.; **et al.** Feature-Based Molecular Networking in the GNPS Analysis Environment. *Nat. Methods* **2020**, *17*, 905–908, doi:10.1038/s41592-020-0933-6.

PART 1

Exploration of bacterial natural products

Chapter 4

Thermoactinomyces vulgaris

4.1 Extremophiles

The name extremophiles refers to organisms that can survive and thrive in physical and chemical conditions that are hostile to most of the biosphere, which is dominated by mesophilic species. They are mostly prokaryotic microorganisms, bacteria and archaea, with a few rare eukaryotic examples (lichens, fungi, protozoa) that colonize niches characterized by extreme temperatures, very acidic or alkaline pH, high salinity, high hydrostatic pressure, ionizing radiation, and/or high concentrations of metals.

Their extraordinary diversity has suggested a classification based on the predominant environmental property:

- thermophiles and hyperthermophiles in hot environments up to temperatures close to or above the boiling point of water;
- psychrophiles in permanently cold habitats, sometimes well below 0 °C;
- acidophiles and basophiles (alkaliphiles), respectively in conditions of high acidity or alkalinity;
- halophiles and osmophiles, adapted to high salt content and high osmolarity;
- piezophiles (barophiles), capable of tolerating pressures up to thousands of atmospheres.

This functional taxonomy, although simplified, provides a useful framework for understanding how extreme selective pressures have shaped physiology and metabolism along multiple evolutionary lines.[1]

The differences between the two domains, prokaryotic and eukaryotic, are reflected in their adaptation strategies.

Prokaryotes, the oldest and most numerous inhabitants of the planet, owe much of their success to metabolic plasticity: they rapidly modulate biochemical pathways, membrane composition, and protection systems to maintain homeostasis in otherwise lethal scenarios. In a thermophile, for example, proteins tend to have a higher content of salt bridges and hydrophobic interaction networks, a reduced length of surface loops, and a strengthening of the hydrophobic "core," all of which increase their denaturation temperature. Plasma membranes incorporate more saturated and branched lipids, with polar heads and fatty acid/steroid ratios that limit the increase in fluidity induced by heat and preserve transporter functionality. Acidophiles and alkalophiles maintain a cytoplasmic pH close to neutrality through highly efficient proton pumps and antiporters, which expel or sequester H^+/OH^- ions, counteracting the external gradient. Halophiles and osmophiles maintain osmotic balance by accumulating compatible internal solutes (e.g., K^+) and regulating ionic strength through active transport and the release of cationic organic amines such as putrescine.

Finally, in psychrophiles, the membrane contains high fractions of unsaturated or polyunsaturated fatty acids, rarely seen in mesophilic prokaryotic cells, which remain liquid at low temperatures and prevent

deleterious stiffening: this ensures lipid homeoviscosity and the correct dynamics of membrane proteins even in sub-zero environments.

These ‘structural’ adaptations are accompanied by ‘chemical’ solutions: the production of extremozymes (or, in the case of thermophiles/hyperthermophiles, thermozymes) that maintain optimal catalytic activity under extreme temperature, pH, or salinity conditions; and the biosynthesis of secondary metabolites capable of shaping a protective microenvironment, shielding oxidative damage, modulating local pH, or interfering with competitors.[1]

The result is a biochemical arsenal capable not only of ensuring survival, but also of offering biotechnological advantages over mesophilic counterparts: greater operational stability, wider temperature/pH windows, resistance to inhibitors and solvents, and sometimes substrate specificity peculiarities useful for industrial processes.[2]

A case in point is Taq polymerase, isolated from the thermophilic bacterium *Thermus aquaticus* in the hot springs of Yellowstone Park. Its thermal stability has made cyclic PCR possible without enzyme replenishment at each denaturation cycle, revolutionizing DNA amplification from micro-samples and paving the way for modern molecular, genomic, and forensic diagnostics.[3] This example clearly shows how biochemical innovation selected in nature under extreme environmental constraints translates into enabling technologies for research and industry.

Today, interest in extremophiles is fuelled not only by the prospect of new biocatalysts, but also by their potential in drug discovery. Numerous metabolites of extremophilic origin exhibit antibacterial, antifungal, antiviral,

and anticancer activities, as well as effects on physiological pathways such as coagulation. In this context, advances in genetics and omics techniques have greatly accelerated discovery.

Two approaches are particularly transformative. Genome mining uses prediction algorithms to identify “silent” or uncharacterized biosynthetic clusters in genomes, for example, non-ribosomal peptide synthetases (NRPS) or polyketide synthases (PKS) pathways, and correlate genes with potential natural products.[4] Metagenomic screening, on the other hand, explores entire environmental microbiomes without the need for cultivation, cloning and testing segments of environmental DNA in model hosts in search of new enzymatic functions. Applied to thermophilic and hyperthermophilic communities, these methods are leading to the identification of new thermozymes with exceptional activity and stability, as well as new bioactive chemical entities that are difficult to access through cultivation alone.[5]

The application value of extremozymes is already evident in many supply chains. In the medical field, enzymes that are stable at extreme temperatures or pH levels improve diagnostic and sample preparation protocols, as well as serving as targets or tools for therapeutic protein engineering. In the agri-food sector, robust enzymatic activities enable more efficient processes (e.g., controlled hydrolysis and fermentation) under conditions that reduce microbial contamination. In the industrial sector (bioenergy, green chemistry), catalysts resistant to solvents and harsh operating conditions reduce costs and environmental impacts, expanding compatibility with unpurified feedstocks.[6] Finally, from a pharmacological point of view, the secondary metabolites of extremophiles constitute chemical scaffolds with

often favourable stability and permeability properties, making them candidates for SAR optimization series.

4.2 Thermophiles

Thermophiles are extremophiles that thrive at high temperatures, typically between 41-122 °C (106-252 °F). Many belong to the Archaea, but they are also common among Bacteria; their distinctive feature is their ability to maintain metabolism and growth under thermal conditions that are prohibitive for most mesophilic organisms. This adaptation is largely due to the production of thermostable enzymes (thermozymes), which not only remain active at high temperatures but often have faster kinetics than their mesophilic counterparts, in accordance with the increase in reaction rates at high temperatures.[7] From a genetic point of view, a correlation between G+C content and thermal proliferation window has been proposed in some coding regions.[8]

Classification can follow two complementary criteria:

1. Thermal requirements: a distinction is made between facultative thermophiles, which grow well both above and below ~50 °C, and obligate thermophiles, which require >50 °C.
2. Optimum growth: simple thermophiles (50-64 °C), extreme thermophiles (65-79 °C) and hyperthermophiles (≥80 °C).

Ecologically, thermophiles inhabit hot, sulphur-rich geothermal niches-hot springs, geysers, sulphur springs, fumaroles, often linked to active volcanism. These sites, found from Italy to Iceland, from Yellowstone (USA) to Kamchatka, function as biogeographical “islands,” favouring local speciation. A recurring feature is the temperature gradient within the springs:

the deeper zones are warmer, the edges colder, and this determines a zonation of microbial communities based on the thermal optimum. In systems with pigmented microorganisms, this zoning manifests itself as multi-coloured microbial carpets covering the rocks, clearly visible at the Grand Prismatic Spring in Yellowstone (Figure 4.1), where each colour band reflects specific biochemical signatures and populations adapted to distinct temperature ranges.[9]

Figure 4.1. Grand Prismatic Spring in Yellowstone Park, USA. The colours are caused by microbial mats formed by thermophiles.

4.3 Thermoactinomyces vulgaris

One of the most interesting regions for the isolation of thermophilic bacteria is undoubtedly Iceland, which is particularly rich in geothermal areas. With the most active volcanic islands on Earth and secondary volcanic manifestations such as hot springs, fumaroles, and geysers, it is the ideal habitat for the growth of these species of microorganisms.

Considering how valuable these territories can be for research as a reservoir of microorganisms, the Icelandic Strain Collection and Records (ISCaR) was created, a national repository of microbial cultures and samples from Iceland's unique natural environment. It mainly comprises cultures of prokaryotes and, to a lesser extent, phytoplankton, fungi, viruses, and complex biomass samples from microbial communities, with a total of over 150 bacterial strains initially isolated. The samples deposited in ISCaR have been collected from a multitude of sites in and around Iceland and from a variety of environments, including marine, freshwater, hot springs, and glaciers.

Once isolated, these microorganisms are systematically screened to assess their biological activity.

Some of these strains have shown interesting pharmacological activity, in particular strain ISCAR 2354, originating from a hot spring on the coast of the Snæfellsnes peninsula in western Iceland (640 57' 52''N, 230 5' 21.53" W).

Thanks to the sequencing of the 16S rRNA gene (GenBank accession number M372920), this strain was identified as *Thermoactinomyces vulgaris* (Thermoactinomycetaceae, Bacillales, Firmicutes), a thermophilic, Gram-positive and filamentous, which grows at temperatures between 55 and 60 °C (Figure 4.2).

From the ISCAR 2354 strain, in a previous study by the research group with which I collaborated during my thesis work, the compound thermoactinoamide A was isolated, a cyclic hexapeptide with antibiotic and antiproliferative activity composed of mixed D and L amino acids, together with minor analogues.[10]

Figure 4.*2*. *Thermoactinoamyces vulgaris* colture

References

1. Irwin, J. A. Overview of Extremophiles and Their Food and Medical Applications. In *Physiological and Biotechnological Aspects of Extremophiles*; Salwan, R.; Sharma, V., Eds.; *Academic Press: London*, **2020**; Chapter 6, pp 65–87. doi:10.1016/B978-0-12-818322-9.00006-X.

2. Somayaji, A.; Dhanjal, C. R.; **et al.** An Insight into the Mechanisms of Homeostasis in Extremophiles. *Microbiol. Res.* **2022**, *263*, 127115. doi:10.1016/j.micres.2022.127115.

3. Chien, A.; Edgar, D. B.; Trela, J. M. Deoxyribonucleic Acid Polymerase from the Extreme Thermophile *Thermus aquaticus*. *J. Bacteriol.* **1976**, *127*, 1550–1557. doi:10.1128/jb.127.3.1550-1557.1976.

4. Blin, K.; Shaw, S.; Kloosterman, A. M.; **et al.** antiSMASH 6.0: Improving Cluster Detection and Comparison Capabilities. *Nucleic Acids Res.* **2021**, *49*, W29–W35. doi:10.1093/nar/gkab335.

5. Li, J.; Jia, H.; Zhang, L.; Shen, Q.; Han, T. Metagenomic Exploration of Thermophilic Microbiomes Uncovers Diverse Industrially Relevant Enzymes. *Bioresour. Technol.* **2015**, *182*, 363–370. doi:10.1016/j.biortech.2015.01.130.

6. Littlechild, J. A. Enzymes from Extreme Environments and Their Industrial Applications. *Front. Bioeng. Biotechnol.* **2015**, *3*, 161. doi:10.3389/fbioe.2015.00161.

7. Vieille, C.; Zeikus, G. J. Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 1–43. doi:10.1128/MMBR.65.1.1-43.2001.

8. Zheng, H.; Wu, H. Gene-Centric Association Analysis for the Correlation between the Guanine–Cytosine Content Levels and Temperature Range Conditions of Prokaryotic Species. *BMC Bioinf.* **2010**, *11* (Suppl. 11), S7. doi:10.1186/1471-2105-11-S11-S7.

9. Ward, D. M.; Castenholz, R. W.; Miller, S. R. Cyanobacteria in Geothermal Habitats. In *Ecology of Cyanobacteria II*; Whitton, B. A., Ed.; Springer: Dordrecht, **2012**; pp 39–63. doi:10.1007/978-94-007-3855-3_3.

10. Teta, R.; Marteinsson, V. T.; Longeon, A.; Klonowski, A. M.; Groben, R.; Bourguet-Kondracki, M.-L.; Costantino, V.; Mangoni, A. Thermoactinoamide A, an Antibiotic Lipophilic Cyclopeptide from the Icelandic Thermophilic Bacterium *Thermoactinomyces vulgaris*. *J. Nat. Prod.* **2017**, *80*, 2530–2535. doi:10.1021/acs.jnatprod.7b00560.

Chapter 5

Unraveling Thermoactinoamides: Molecular Networking, Biosynthesis, and Unexpected Epimerization

5.1 Synopsis of this chapter [1]

Previous work had already identified thermoactinoamide A as a lipophilic cyclic hexapeptide derived from the thermophile *Thermoactinomyces vulgaris*.[1] From Icelandic isolates (ISCAR2354/2850), LC-MS/MS with feature-based molecular networking outlined a small family of congeners (A-F, later G-K). Genome mining of *T. vulgaris* DSM 43016 identified the multimodular NRPS Thd cluster (initially in two contigs, reassembled by reference mapping), and domain/substrate predictions (NRPSpredictor2, NaPDoS) explained the mixed D/L pattern and tolerated substitutions.[2] The structure of thermoactinoamide A (Figure 5.1) was assigned by HRMS, multidimensional NMR, and Marfey-enhanced LC-MS.

These results constitute the preliminary framework within which the present work focuses on head-to-tail macrocyclization and unexpected α-epimerization.[3]

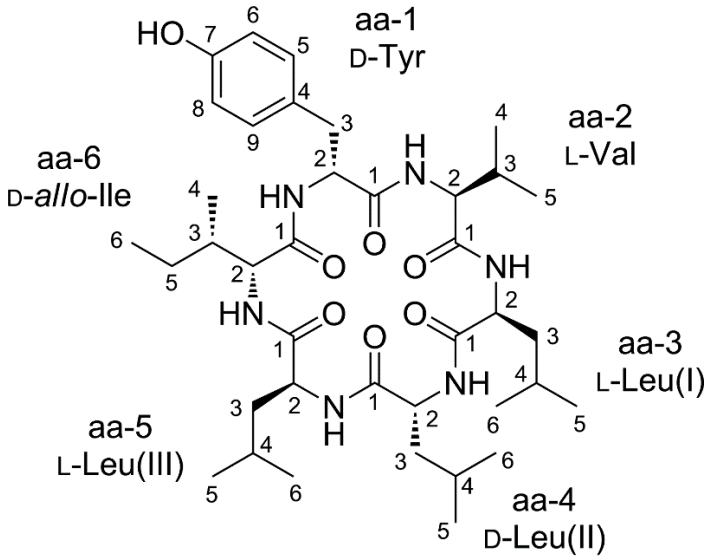


Figure 5.1. Structure of thermoactinoamide A (**1a**)

5.2 Discovery, Isolation, and Stereostructural Elucidation of Thermoactinoamide A from the Thermophilic Bacterium Thermoactinomyces vulgaris

Thermoactinoamide A is a lipophilic cyclic hexapeptide produced by the thermophilic bacterium *Thermoactinomyces vulgaris*. It was discovered during an antimicrobial study of Icelandic isolates recovered from geothermal environments, niches where ecological pressure often goes hand in hand with chemical peculiarity. During this campaign, strain ISCAR 2354, sampled from a coastal hot spring on the Snæfellsnes peninsula (western Iceland), repeatedly showed antibacterial activity and was identified as *T. vulgaris* by 16S rRNA sequencing. An independent isolate, ISCAR 2850, from a deep-water hydrothermal sediment core, exhibited the same metabolic signature, strengthening the association between this species and the observed peptide family. These observations motivated a comprehensive isolation and structural elucidation program focused on the major component, subsequently named thermoactinoamide A.[1]

To access the molecule, the producing strain was grown at 60°C for 24 hours in 166 medium supplemented with 1% NaCl, under conditions consistent with a thermophilic growth profile. The cells and broth were treated together: brief sonication disrupted the biomass, and two liquid-liquid separations with H_2O-saturated *n*-butanol containing 10% methanol enriched the organic phase in peptide-like constituents. After filtration on PTFE (0.45 µm) and gentle concentration at 40 °C, the crude extract was analyzed by LC-ESI-HRMS on a 5 µm Kinetex C18 (50 × 2.1 mm) at 25 °C with a gradient of water (0.1% formic acid)/acetonitrile and Orbitrap detection.

The chromatogram revealed a compact family of six related peaks with peptide-like MS/MS fragmentation; the dominant member was identified as thermoactinoamide A. Milligram quantity was obtained by preparative reverse-phase HPLC on a 10 µm Kinetex C18 (250 × 10 mm) using a MeOH/H_2O (0.1% formic acid) with DAD monitoring at 280 nm, yielding material of sufficient purity for detailed spectroscopic analysis ($t_R \approx 18.7$ min; $[\alpha]^{25}_D$ +80 in MeOH; weak ECD at 280 nm).

High-resolution MS and MS/MS established the molecular formula and the peptide nature. In addition to the base peak at *m/z* 715.4752 corresponding to the $[M+H]^+$ ion, the spectrum showed $[M+Na]^+$ at *m/z* 737.4570, a double charge $[M+Ca]^{2+}$ at *m/z* 377.2147, and $[M+Ca+HCOO]^+$ at *m/z* 799.4277, all within 1 ppm of the values calculated for $C_{38}H_{63}N_6O_7$ or the corresponding adducts.

The MS/MS fragmentation pattern was diagnostic of a small lipophilic peptide containing Tyr, Val, and multiple Leu/Ile residues; together with the exact mass, these data indicated a head-to-tail cyclized hexapeptide structure (Figure 5.2).

A comprehensive suite of 1D/2D NMR (^1H, ^{13}C, COSY, TOCSY, HSQC, HMBC, NOESY) acquired at 700 MHz in CD_3SOCD_3 provided a sequence-level picture. Six distinct amide NH signals and six α-CH signals confirmed a six-residue peptide. TOCSY and COSY defined the side chain spin systems for Tyr, Val, one Ile residue, and three Leu residues. Fundamentally, sequential NOESY correlations between each α proton and the amide NH of the next aminoacid residue of the chain, supported by HMBC connectivity, reconstructed the order as a cycle (Tyr-Val-Leu(I)-Leu(II)-Leu(III)-*allo*-Ile).

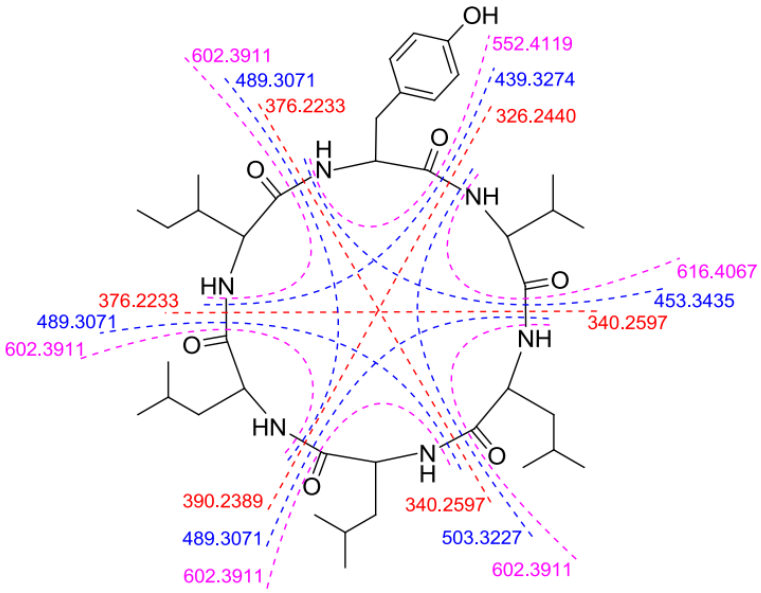


Figure 5.2. Thermoactinoamide A (**1a**) fragmentation pattern[1]

The absolute configuration of each residue was established using an advanced Marfey workflow enabled for LC-MS, adapted to the well-known Leu/Ile/*allo*-Ile challenge. After acid hydrolysis (6 N HCl/AcOH, 120 °C) and derivatization with D-FDAA, a pentafluorophenyl (PFP) stationary phase was used (2.6 µm Kinetex PFP, 100 × 4.6 mm; 25 °C; 200 µL min^{-1}; 60 → 100% MeOH) to achieve base separation of the D/L isobaric pairs, which cannot be reliably achieved with conventional C18. Comparison with D-/L-FDAA standards for Tyr, Val, Leu, Ile, and D-*allo*-Ile resolved the set as D-Tyr, L-Val, D-Leu, L-Leu, L-Leu, D-*allo*-Ile. Since Marfey identifies configurations but not positions, partial hydrolysis was performed to generate informative fragments: a linear pentapeptide (*m/z* 620.4018) corresponding to the loss of a Leu/Ile and a tetrapeptide (*m/z* 521.3333) corresponding to the loss of Val plus a Leu/Ile. Marfey's on these fragments, coupled with their MS/MS sequences (H-Leu(III)-allo-Ile-Tyr-Val-Leu(I)-OH; H-Leu(II)-

Leu(III)-allo-Ile-Tyr-OH), placed the single D-Leu in Leu(II), completing the stereochemical map of the macrocycle (Figure 5.3).

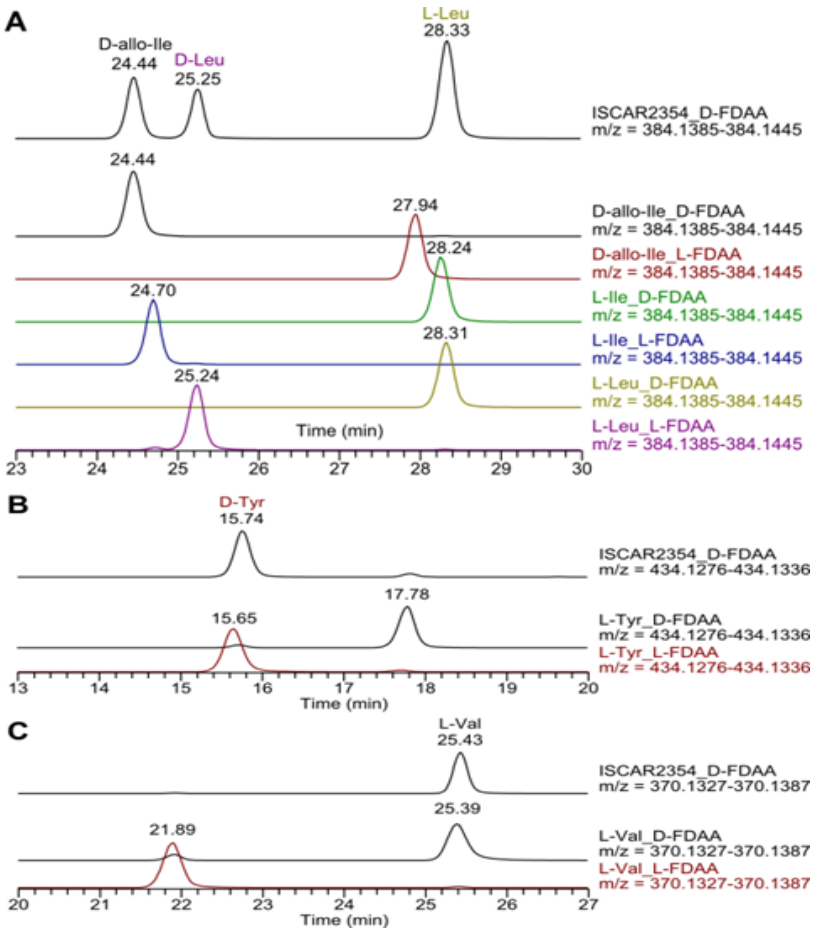


Figure 5.3. Advanced Marfey's analysis of compound **1** using a pentafluorophenyl (PFP) bonded-phase column. (A)Extracted-ion chromatograms at m/z 384.1415 of the D-FDAA derivatives from the hydrolysis of 1 and of D- and L-FDAA derivatives L-Leu, L-Ile, and D-allo-Ile. (B) Extracted-ion chromatograms at m/z 434.1306 of the D-FDAA derivatives from the hydrolysis of 1 and of D-and L-FDAA derivatives L-Tyr. (C)Extracted-ion chromatograms at m/z 370.1357 of the D-FDAA derivatives from the hydrolysis of 1 and of D- and L-FDAA derivatives L-Val.[1]

In addition to the main metabolite, LC-MS investigation consistently revealed ten minor co-produced analogues, insufficient for isolation, but amenable to putative structural annotation by MS/MS.

By analogy with thermoactonamide A, these congeners differ at one position (e.g., Val↔Ile/*allo*-Ile, Tyr↔Phe/Tyr, or a Leu/Ile substitution), while

retaining the overall architecture of the cyclic hexapeptide and the mixed D/L pattern. Although preliminary, this microfamily indicates an enzyme system with relaxed substrate specificity at selected positions and reinforces the idea that mixed-configuration lipophilic cyclic hexapeptides from thermophilic organisms constitute a treatable scaffold for chemical biology.

5.3 Biological Activity of Thermoactinoamide A

Some structural features of thermoactinoamide A, such as being a cyclic hexapeptide with mixed D/L configurations, an aromatic amino acid residue, and a prevalence of lipophilic residues, can be considered an interesting starting point for defining new scaffolds in the search for antibiotic drugs. The activity of thermoactinoamide A was tested against a group of three Gram-positive bacterial strains (*Staphylococcus aureus* ATCC 6538, *Enterococcus faecalis* ATCC 29212*, Bacillus cereus* ATCC 14579) and three Gram-negative bacterial strains (*Escherichia coli* ATCC 8739, *Pseudomonas aeruginosa* ATCC 13388, *Klebsiella pneumoniae* ATCC 11296). The compound was shown to exert significant and selective bacterial growth inhibition against *S. aureus* with a minimum inhibitory concentration (MIC) value of 35 μM, while it was inactive against the other strains up to 140 μM.[1] In addition, potential inhibitory effects on cell growth were tested by exposing a single dose of 5 μM for 72 hours against three different tumor cell lines, namely BxPC-3 (pancreatic carcinoma), PANC1 (pancreatic carcinoma), and 3AB-OS (osteosarcoma tumor stem cell line). Cell proliferation was monitored by the xCELLigence System Real-Time Cell Analyzer (RTCA), based on dynamic measurements of electronic impedance changes, indicative of cell viability and morphology. Unlike PANC1 and 3AB-OS carcinoma cells that were not sensitive to drug treatment (Figures 5.4 and 5.5), the compound was shown to exert selective and moderate antiproliferative activity against BxPC-3 tumor cells at a concentration of 5 μM (Figure 5.6), inducing exclusively in these cells a significant reduction in the slope of the

real-time proliferation curve and a simultaneous increase in the doubling time of tumor cells.

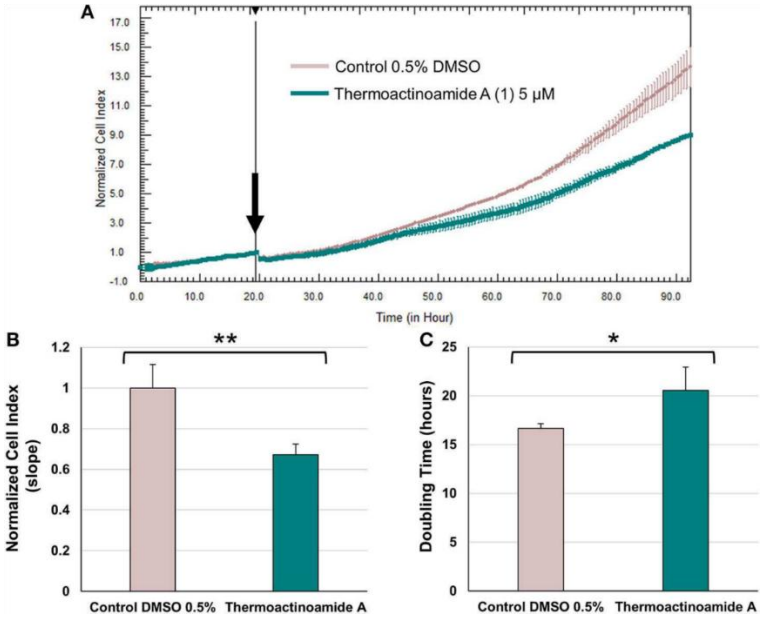


Figure 5.4. Effects of thermoactinoamide A on the proliferation of PANC-1 cancer cells.[2]

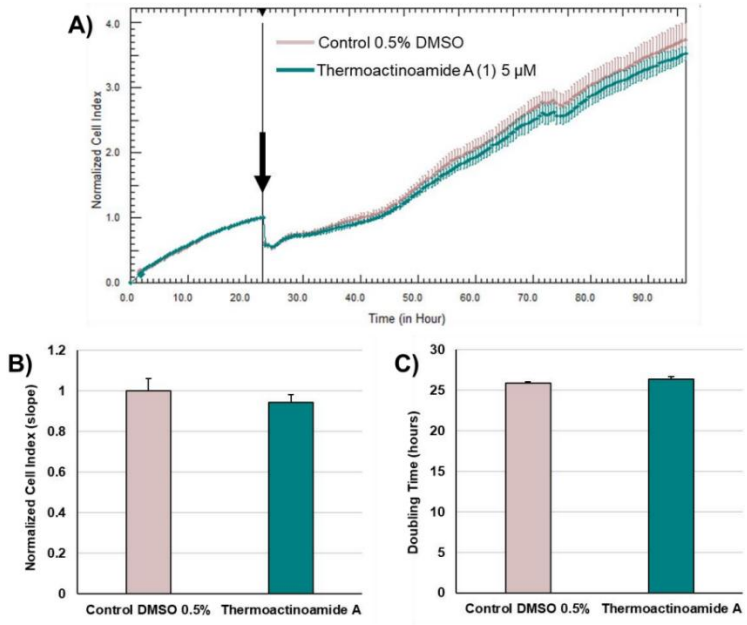


Figure 5.5. Effects of thermoactinoamide A on the proliferation of 3AB-OS cancer cells.[2]

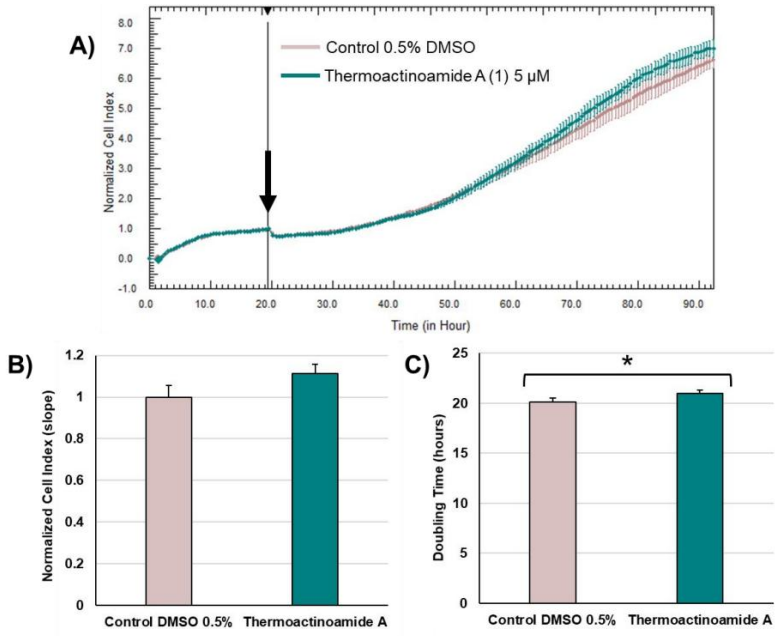


Figure 5.6. Effects of thermoactinoamide A on the proliferation of BxPC-3 cancer cells.[2]

5.4 Non-ribosomal peptides

Non-ribosomal peptides (NRPs) are a class of secondary metabolites, usually produced by microorganisms such as bacteria and fungi. They are classified as such because they are synthesized by enormous enzyme complexes, Non-Ribosomal Peptide Synthetases (NRPS), through a sequential process like an assembly line and independent of messenger RNA. These compounds exhibit great structural diversity and, consequently, a wide range of biological activities and pharmacological properties.[4] This makes them ideal candidates in the search for new lead compounds from natural sources, exploitable in various areas, such as plant, animal, and human health. The function of these secondary metabolites and their use for the producing organism is still the subject of scientific debate. For those from non-pathogenic bacterial strains, such as various actinomycetes, a communication or signalling function seems likely.

They often consist of complex cyclic and/or branched structures of concatenated repeated sequences, which may also contain non-proteinogenic amino acids (NPAAs), including D-amino acids and amino acids modified with N-methyl and N-formyl groups, glycosylations, acylations, halogenations, or hydroxylations. Non-proteinogenic residues are not naturally encoded by the human genetic code, nor can they be found in the polypeptide chains of our organism.[5] However, in bacteria, fungi, plants, and marine animals, they are essential building blocks of polypeptide chains, as they confer stability against proteolytic degradation and provide stereochemical constraints that promote downstream biosynthetic steps and the resulting tertiary structure of the natural peptide product.

Therefore, the biosynthetic processing and bioactivity of peptides often depend on their presence. Examples of non-proteinogenic amino acids are: the enantiomer D-alanine, which contributes to the structure of peptidoglycan (murein), a constituent of the cell wall of most bacteria; dehydroalanine (2-amino-2-propenoic acid), a non-proteinogenic monomer of microcystin toxins produced by some cyanobacteria of the genus *Mycrocystis*; butyrin (2-aminobutyric acid), a constituent of the antibiotic cyclosporine produced by the fungus *Beauveria nivea*; the D-valine enantiomer, a precursor of several antibiotics, including penicillins and cephalosporins.

The currently known structures of non-ribosomal peptides reflect the complexity and abundance of the different structural classes that characterize them. Head-to-tail macrocycles with rings of various sizes (e.g., gramicidin S, cyclosporin) represent, together with lipocycle peptides (e.g., polymyxin B, colistin), the largest group of NRPs. Cyclization, in addition to increasing

proteolytic stability, constrains the conformation of the peptide, thereby promoting the establishment of effective interaction with the biological target and the subsequent activity. Linear peptide structures are also abundant and include tripeptide compounds (e.g., bialafos) up to peptides with twenty amino acid units (e.g., alameticine). They often have oxazolidine or thiazolidine nuclei, which can be further oxidized or reduced, while residues such as serine can undergo dehydration reactions to form dehydroalanine.

Several non-ribosomal peptides have been shown to have useful therapeutic activities as antibiotics, cytostatics, or immunosuppressants, to the extent that they are included in medicinal formulations currently on the market. Examples include the antibiotics penicillin and vancomycin, anticancer drugs such as bleomycin, and immunosuppressants such as cyclosporine A.

5.5 Non-ribosomal peptide synthetase

Non-ribosomal peptides are synthesized by an equally complex category of biocatalysts, non-ribosomal peptide synthetases (NRPS). Unlike ribosomes, these multi-enzymatic complexes do not involve messenger RNA.[6] The structural heterogeneity typical of non-ribosomal peptides is primarily due to the ability of NRPS to incorporate a wider range of monomers than the biosynthetic pathway of ribosomal peptides. Suffice it to say that of the 500 amino acids present in nature and identified, the 20 proteogenic amino acids represent a minority of only 4%. Furthermore, the peptide scaffold thus produced can undergo extensive modifications both during and after the assembly of the amino acid chain. This is thanks to the particular architecture of NRPS, in which each module governs, according to the rule of collinearity,

the incorporation of a monomer into the growing amino acid chain. Each module is separated from the next by short spacer regions of about 15 amino acids and is itself made up of at least four distinct domains, identified as central or essential domains, each with defined functions. The catalytic domains of NRPS select, activate, or modify covalently bound reaction intermediates, guiding the entire process of amino acid chain elongation and the subsequent release of the peptide product (Figure 5.7). The basic domains of an NRPS are:[7]

- Adenylation domain (A): The A domain is a member of the adenylating enzyme superfamily (ANL), which also includes acyl-CoA synthetases and luciferases. Composed of approximately 550 amino acids, during the first phase, the adenylation domain selects the appropriate amino acid with high specificity, activating it by adenylation with hydrolysis of adenosine 5′-triphosphate (ATP). This forms an aminoacyl-AMP, which is then transferred to the peptidyl carrier protein (PCP) domain. An A domain is characterized by the presence of 10 conserved motifs (designated a1 to a10) distributed throughout the entire sequence. Scattered between motifs a3 and a6 are 10 amino acids of fundamental importance, as they represent the binding pocket and confer substrate specificity. The relationship between the nature of these sequences and selectivity during monomer recruitment has been the subject of many studies to develop bioinformatic tools capable of predicting in silico the monomer most likely to be recruited by each A domain.

However, unlike the ribosomal genetic code, the NRPS code still has some pitfalls and has not been completely deciphered.

Finally, a certain promiscuity of A domains has been observed, thanks to which co-production of variants of the same compound can be achieved. This probably represents an adaptive tool, as it allows the synthesis of multiple analogues based on the substrate present at a given time.

- Peptidyl carrier protein (PCP) domain: Once selected and adenylated, the activated monomer is transferred to the peptidyl carrier protein (PCP) domain, which consists of approximately 100 amino acids and is also referred to as the thiolation domain (T). PCPs play a central role in the transfer of amino acids and peptides between different catalytic domains.

Typical NRPS assembly lines contain one PCP per module, which means that there are the same number of PCP domains in the assembly chain as monomers in the final peptide product. Before they can perform their role, PCPs need to be activated by a post-translational modification: phosphopantetheine transferase (PPTase) catalyzes the transfer of a portion of phosphopantetheine (PPant) to a serine residue located at the apex of the second helix that makes up the PCP.

The PPant fraction derives from coenzyme A (CoASH) and ends with a reactive thiol group: in this phase, the aminoacyl-AMP establishes a covalent bond with the free thiol group of the cofactor 4′-phosphopantetheine (4'-PP) bound to the PCP, releasing AMP. The swinging arm with the terminal SH thiol group, whose flexibility is essential for the functioning of the NRPS, allows the bound substrates to reach the catalytic sites within the domains, moving the growing peptide along the assembly line.

At this stage, various modifications can be made to the substrate by optional domains, such as epimerization (E domain) to obtain the D enantiomers of

amino acid residues, N-methylation (NMT domain), cyclization (Cy domain) to produce heterocycles, or oxidation (Ox domain). T

The PCP domain can also serve as a platform to present the amino acid or peptide chain to enzymes that introduce in-trans modifications, such as halogenases, transferases, or monooxygenase enzymes.

- Condensation domain (C): this is the third central domain, approximately 450 amino acids long. It is located upstream of the adenylation domain and mediates the formation of the peptide bond between the amino acids linked to two neighbouring thiolation domains. Specifically, the bond involves the thioester group of the growing peptide chain and the amino group of the newly activated extending amino acid.

Several subtypes of C domains have been identified based on the nature of the condensed monomers: LCL domains catalyse the formation of a peptide bond between two L monomers, while DCL domains condense a monomer with a D configuration to an L monomer loaded on the next module.

Although C domains do not possess substrate-specific binding pockets as in A domains, they have been shown to participate in a correction process that prevents them from performing the condensation reaction if, for example, certain modifications of the aminoacyl thioesters bound to the T domain are missing, as found during the biosynthesis of glycopeptide antibiotics.

- Thioesterase (TE): Finally, a C-terminal thioesterase domain (TE) catalyses the release of the newly formed non-ribosomal peptide from the enzyme complex. The peptide can be released by hydrolysis, producing linear compounds such as aminoadipic acid-cysteine-valine (ACV), a precursor of the antibiotic penicillin, or through an intramolecular reaction to produce a

cyclic peptide, using hydroxyls or amines as nucleophiles. Once released, the peptide is often further modified by glycosylation, acylation, halogenation, or hydroxylation reactions to produce the mature natural product.

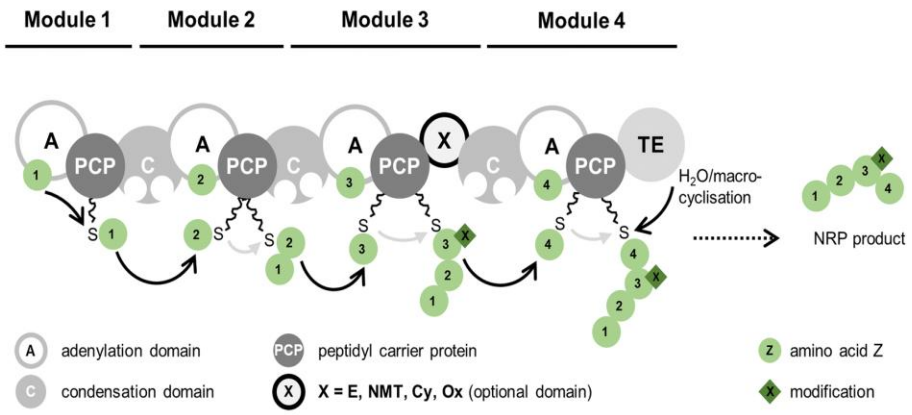


Figure 5.7. Schematic representation of the domains that make up NRPS and the general biosynthetic principles of NRPs.

Several cyclization strategies based on the three-dimensional architectures of TE domains have been described: *cis*-acting type I TEs catalyse the release of the product as macrocycles of various head-to-tail (e.g., tirocidine) or side chain-to-tail (e.g., bacitracin) sizes, while autonomous type II TEs act in *trans* mainly as correction/modification enzymes.

At the same time, in addition to the central domains, secondary domains have also been described, which can modify the incorporated amino acids during non-ribosomal synthesis, catalysing epimerization, methylation, cyclization, oxidation, and, although rarely found, formylation reactions.

These modifications give the secondary metabolites structural characteristics that lead to greater resistance to proteolytic degradation.

By far the most frequently found secondary domain among NRPSs is the epimerization domain (E), which converts an L-form monomer to its D-form

through α-carbon isomerization.[8] When present in a given module, the E domain follows the T domain and catalyses the epimerization of the monomer within the growing peptide bound to it. Consequently, a C domain following an E domain is classified and traced back to subtype DCL.

Another secondary domain is the methylation domain (MT), which catalyses the transfer of a methyl group from S-adenosylmethionine (SAM) to the carbon or nitrogen atom of the peptide bond for methylation of the peptide scaffold. Generally, the MT domain is located upstream of the T domain to which the target residue is covalently bound. N-methylation is a structural requirement for maintaining an active conformation of the peptide.

Formylation (F) domains are often listed among the secondary NRPS domains, although they are rarely observed in nature. To date, only three cases of compounds bearing N-formylation due to an F domain in the initial module of the assembly lines have been described: anabaenopeptilides, linear gramicidin, and kolossin A. F domains in NRPSs show a high similarity to methionine-tRNA-formyltransferase, and their activity occurs once the amino acid is loaded onto the subsequent T domain.

All these modifications can be made by secondary domains acting in cis during peptide assembly, but they can also result from external enzymes acting in trans while the intermediates are still covalently bound to the NRPS or after the release of the assembled peptide. For example, while F domains acting in cis are active only on the N-terminal amino acid, formyltransferases can catalyse the modification of a monomer located also within the peptide portion, which would explain the higher prevalence of formylated secondary

metabolites compared to what would be expected based on the F domains present.

The recruitment and modification of amino acids used during non-ribosomal peptide synthesis can occur via different pathways.

For example, cytoplasmic enzymes encoded by the same gene clusters as NRPSs use proteinogenic amino acids or other primary metabolism substrates to generate non-proteinogenic amino acids, which are subsequently involved in NRPS domains (Figure 5.8, type A).

An alternative mechanism involves the loading of a proteinogenic amino acid onto an autonomous NRPS module, modification by trans-acting enzymes, and subsequent release (Figure 5.8, type B).

Alternatively, the activated building block can be transferred to a shuttle protein or an autonomous T domain for translocation to the main assembly line (Figure 5.8, type C).

However, in some cases, the autonomous NRPSs themselves represent the main assembly line in which synthesis takes place, and the substrates bound to them can be released directly as mature products (Figure 5.8, type D).

An additional mechanism involves the loading of the amino acid onto the multimodular NRPS assembly line and modification by one or more cis-acting NRPS domains (Figure 5.8, type E).

An NRPS can also recruit other sartorial enzymes that act in trans (Figure 5.8, type F).

Finally, post-NRPS modification can occur after the release of the peptide from the NRPS (Figure 5.8, type G).

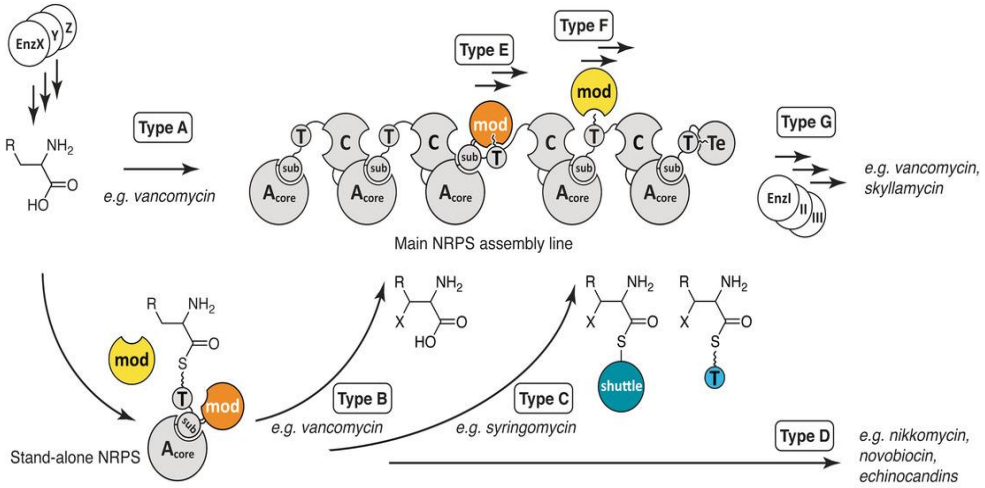


Figure 5.8. Recruitment pathways and modifications of amino acids used by NRPS.

Depending on the characteristics of NRPS complexes, the biosynthesis modes of non-ribosomal peptides are classified as linear, iterative, and non-linear. The latter includes modes that are not strictly linear and not strictly iterative. Linear biosynthesis refers to the co-linearity between the order of the modules in the non-ribosomal assembly chain and the sequence of the monomers incorporated into the final peptide.

This means that the number and order of the modules in the NRPS coincide with the number and order of the monomers in the peptide.

An example of iterative mode, on the other hand, concerns the biosynthesis of gramicidin S. This antibiotic produced by the bacterium *B. brevis* is a cyclic decapeptide, but its biosynthetic pathway involves a total of five modules. In this case, the TE domain plays a key role in allowing the enzyme to repeat the synthesis process twice. An iterative mechanism is usually found in the synthesis of non-ribosomal siderophores such as enterobactin or bacillibactin, where the presence of trimers in the structure is a necessary condition for effective iron chelation.

The non-linear biosynthesis mode includes all modes that do not fall into the previous categories and mainly leads to branched structures. These NRPS are characterized by an unusual arrangement of the central domains, which leads to unusual internal cyclization or branching points. While the first two assembly lines were linear, these NRPS use an unexpected logic, given their architecture. This makes it more difficult to predict the products of such enzyme complexes.

Due to their modular organization, NRPSs vary in size, but some of them can be exceptionally large (up to over one megadalton) and are therefore encoded by genes or groups of genes which, together with those encoding cell surface proteins, are considered among the largest in the microbial world.

In bacteria, genes for the biosynthesis of secondary metabolites are commonly grouped into so-called biosynthetic gene clusters (BGCs). Therefore, based on knowledge of a given gene cluster, the structure of the corresponding non-ribosomal peptides can be predicted thanks to the correlation by collinearity between the amino acid sequence of the metabolite and the sequence of modules in the enzyme structure, which in turn is dictated by the gene sequence encoding the individual modules. This has led to the development of bioinformatics tools capable of detecting NRPS biosynthetic pathways through the analysis of their gene clusters and predicting the structure of non-ribosomal peptides directly from genome sequences. Although biological analyses will be always necessary to confirm the function and activity of newly discovered compounds, this genome- and metabolome-based isolation strategy results in considerable time savings compared to previous bioanalysis approaches.

In fact, it avoids the rediscovery of already known NRPs, fully exploiting unexplored microbial resources and the application potential of the diverse secondary metabolites synthesized by NRPS.

5.6 Identification of the biosynthetic gene cluster of thermoactinamide

The study of the biosynthetic cluster responsible for thermoactinoamide began with the analysis of the genome of *Thermoactinomyces vulgaris* DSM 43016 (GenBank; Figure 5.9), which is 2.56 Mb in size, with GC 47.9% and 15 contigs. The contigs, sequences obtained by assembling the reads based on overlapping regions, redistribute the sequencing fragments into longer, continuous units (in the order of 50,000–200,000 bp), minimizing the interruptions typical of the whole-genome shotgun approach.

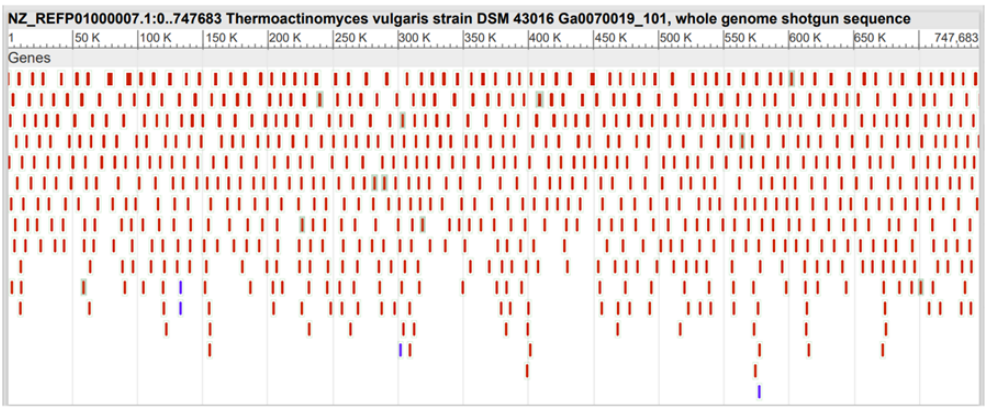


Figure 5.9. Genome of *Thermoactinomyces vulgaris* DSM 43016 from GenBank.[2]

To define the taxonomic scope and enable comparative analysis, the Average Nucleotide Identity (ANI) was applied, which fragments the query genome in silico (blocks of 1020 bp) and evaluates its average identity against the

intact subject.[9] ANI showed values >95% (taxonomic cut-off) between DSM 43016 and the deposited genomes of *Thermoactinomyces sp.* AS95, Gus2-1, and CDF, supporting their assignment to *T. vulgaris* DSM 43016 and enabling integrated genome mining (Table 5.1).

Genome 1 - Accession	Genome 1	Genome 2	ANI (%)
NZ_REFP00000000	*T. vulgaris* DSM 43016	*Thermoactinomyces* sp. AS95	99.65
NZ_REFP00000000	*T. vulgaris* DSM 43016	*Thermoactinomyces* sp. CDF	99.56
NZ_REFP00000000	*T. vulgaris* DSM 43016	*Thermoactinomyces* sp. Gus2-1	99.67
LSVF00000000	*Thermoactinomyces* sp. AS95	*Thermoactinomyces* sp. Gus2-1	99.71
LFJU00000000	*Thermoactinomyces* sp. CDF	*Thermoactinomyces* sp. AS95	99.61
JPZM00000000	*Thermoactinomyces* sp. Gus2-1	*Thermoactinomyces* sp. CDF	99.58

Table 5.1. ANI calculations show that *Thermoactinomyces* sp. AS95, *Thermoactinomyces* sp. CDF, and *Thermoactinomyces* sp. Gus2-1 are associated with *Thermoactinomyces vulgaris* DSM 43016, as the corresponding ANI values exceed 95%.[2]

A preliminary screening with antiSMASH and PRISM identified five putative secondary metabolite pathways in the DSM 43016 genome:

- a type III polyketide synthase;
- a small NRPS system with separately encoded A, PCP, and E proteins;
- a multimodular NRPS cluster distributed over two contigs;
- a lassopeptide;
- a siderophore cluster.

The same biosynthetic architecture was found in AS95, Gus2-1, and CDF. Considering that cyclic peptides with D-amino acids (such as thermoactinoamide A) are typically produced by NRPS and that module→residue co-linearity allows mapping of structure and genetics, it

was deduced that the thermoactinoamide family is directed by the multimodular NRPS pathway, designated “Thd”.

The two contigs containing the NRPS BGC of DSM 43016 were assembled manually using the contig with the intact Thd cluster of *Thermoactinomyces* AS95 as a reference; the putative genes obtained from the de novo assembly are summarized in Table 5.2.

ORF	Protein ID	Position [nt]	No. of aa	Putative function	Closest homolog (accession#) organism	expect value	identity/positives [% aa]
ORF1	RMB00104	24613-26286	557	ribonuclease J	WP_037996450, *Thermoactinomyces*	0.0	99/100
ORF2	RMB00105	26442-26999	185	hypothetical protein	WP_037996506, *Thermoactinomyces*	2e-134	100/100
ORF3	RMB00106	27109-28356	415	peptidoglycan amidohydrolase	WP_052186787, *Thermoactinomyces*	0.0	100/100
ORF4	RMB00107	28668-28895	75	hypothetical protein	WP_037996451, *Thermoactinomyces*	3e-45	100/100
ORF5	RMB00108	28979-29302	107	multidrug efflux SMR transporter	WP_037996453, *Thermoactinomyces*	9e-68	100/100
ThdA	RMB00109	29719-41271	3850	**NRPS (C*-A-PCP-C-A-PCP-E-C-A-PCP-E)**	WP_049720208, *Thermoactinomyces* sp. CDF	0.0	99/99
ThdB	KYQ85937	41347-51942	3531	**NRPS (C-A-PCP-C-A-PCP-C-A-PCP-E)**	WP_138614946, *Thermoactinomyces vulgaris*	0.0	99/99
ORF8	RMB00111	51988-52482	164	hypothetical protein	WP_049720207, *Thermoactinomyces*	2e-118	100/100
ORF9	RMB00112	52923-53867	314	dihydroxyacetone kinase subunit DhaK	WP_037996514, *Thermoactinomyces*	0.0	100/100
ORF10	RMB00113	53869-54501	210	dihydroxyacetone kinase subunit L	WP_037996459, *Thermoactinomyces*	2e-153	100/100
ORF11	RMB00114	54498-54884	128	dihydroxyacetone kinase subunit DhaM	WP_037996461, *Thermoactinomyces*	9e-86	100/100
ORF12	RMB00115	54971-55693	240	ATP-dependent protease ClpP protease subunit	WP_037996464, *Thermoactinomyces*	2e-173	99/100
ORF13	RMB00116	55690-55911	73	YlzJ-like protein	WP_037996465, *Thermoactinomyces*	3e-46	100/100

Table 5.2. Putative genes identified on the de novo assembled contig containing the thermoactinamide cluster (highlighted in green and bold) from *T. vulgaris* DSM 43016.[2]

The Thd cluster measures 22,223 bp and encodes two NRPS, named ThdA and ThdB, each consisting of three modules, in line with the cyclohexapeptide structure of thermoactinoamides (Figure 5.10).

Using NRPSpredictor2 (integrated into antiSMASH), the selectivity of the adenylation domains was predicted, reconstructing the possible sequence of incorporated amino acids. The integration of new genetic-biosynthetic information with chemical-spectroscopic data has made it possible to clarify

structural ambiguities for low-abundance congeners or those with symmetries/isobars (which complicate the interpretation of MS/MS spectra), to outline the stereostructural characteristics of already known variants, and to guide the targeted detection of new congeners produced by *T. vulgaris* DSM 43016.[2]

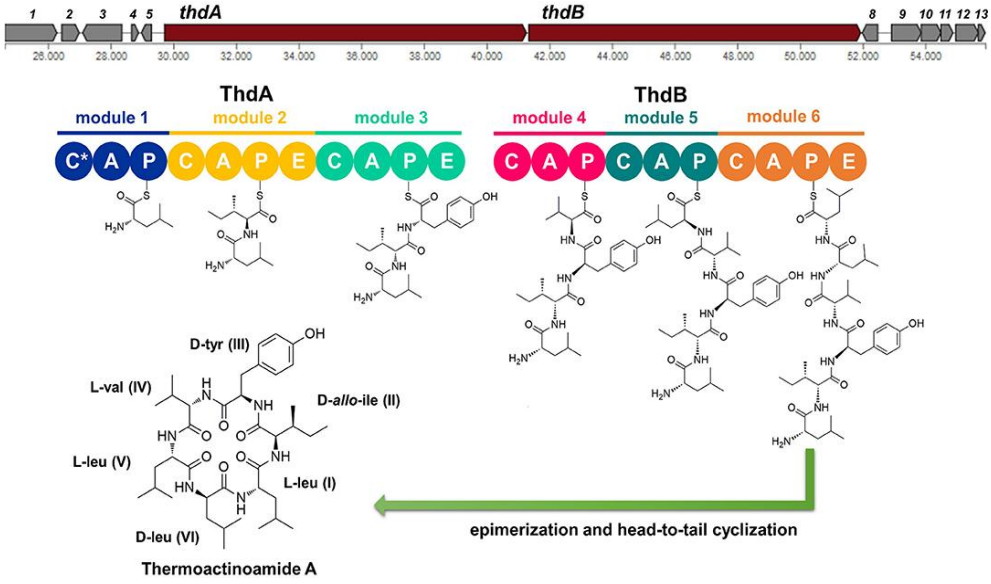


Figure 5.10. Presumed biosynthesis of thermoactinamide A (C, condensation domain; A, adenylation; P, peptidyl carrier protein; E, epimerase; C*, truncated condensation domain, probably inactive).[2]

The peptide chain begins with the activation of L-Leu by domain A1 of the first module of thdA, which adenylates the monomer by consuming ATP and forms aminoacyl-AMP; the residue is then transferred to PCP1 via a thioester bond with 4′-phosphopantetheine (4′-PP), with the release of AMP. The specificity for leucine of the $ThdA_{A1}$ domain is supported by both the Stachelhaus code and the predictions of NRPSpredictor2 (Table 5.3).

	Antismash prediction	Binding pocket signatures (residue position[a])								
		235	236	239	278	299	301	322	330	331
ThdA_A1	Leu	D	A	W	F	L	G	N	V	V
ThdA_A2	Ile/Val	D	G	F	F	L	G	V	V	F
ThdA_A3	Tyr	D	A	A	T	I	A	A	V	C
ThdB_A4	Ile/Val	D	G	F	F	V	G	G	V	F
ThdB_A5	Leu	D	A	W	F	L	G	N	V	V
ThdB_A6	Leu	D	A	W	F	L	G	N	V	V

[a]as reported by Teta et al. (2015).

Table 5.3. Distinctive features of adenylation domains from Thd synthetase.[2]

The second module of thdA selects L-Ile (A2 domain with amino acid-specific motif), charges it on PCP2, and the condensation domain (C2) catalyses the formation of the L-Leu-L-Ile peptide bond between the two PCPs.

The NaPDoS analysis[10] (Table 5.4) classifies C2 as LCL (L/L condensation) and confirms the presence of an epimerization domain (E) in the module: the latter converts L-Ile to D-*allo*-Ile, in accordance with the configuration experimentally assigned to the natural product.

The third thdA module incorporates Tyr (A3 selectivity predicted by NRPSpredictor2; Stachelhaus code match ≈ 90%) and the DCL-type C3 domain condenses D-*allo*-Ile with L-Tyr. A second E domain in the same module epimerizes tyrosine to D-Tyr, generating the tripeptide L-Leu–D-*allo*-Ile–D-Tyr. The chain then passes to NRPS ThdB, consisting of three modules in series. The A4 domain shows partially relaxed selectivity for Val/Ile (hydrophobic binding pocket), in line with the presence of variants of the family with Val or Ile in that position; the C4 domain is classified as DCL and couples D-Tyr with L-Val/Ile. The subsequent modules A5 and A6 (100% sequence identity) are both specific for L-Leu and the corresponding C5/C6

are LCL, completing the elongation. The D-Leu of the final product is attributed to the thdB$_{E3}$ domain, which epimerizes the last L-Leu installed.

	Query id	Database match id	percent identity	align length	e-value	pathway product	domain class
☐	ThdA_C*	tyroc3_C1_DCL	24	275	1e-18	tyrocidin	DCL
☐	ThdA_C1	tyroc3_C3_LCL	47	419	2e-112	tyrocidin	LCL
☐	ThdA_E1	surfa5_C4E	44	446	2e-113	surfactin	epim
☐	ThdA_C2	bacit3_C1_DCL	40	426	4e-98	bacitracin	DCL
☐	ThdA_E2	tyroc2_C4E	48	451	9e-106	tyrocidin	epim
☐	ThdB_C3	grami2_C1_DCL	42	430	2e-104	gramicidin	DCL
☐	ThdB_C4	tyroc3_C2_LCL	48	419	1e-113	tyrocidin	LCL
☐	ThdB_C5	tyroc3_C4_LCL	50	419	6e-117	tyrocidin	LCL
☐	ThdB_E3	grami1_C1E	42	442	3e-94	gramicidin	epim

Table 5.4. NaPDoS analysis in the detection and prediction of condensation and epimerization domains within the Thd gene cluster. A specific class was assigned to each domain. An $^{L}C_{L}$ domain catalyses a peptide bond between two L-amino acids; a $^{D}C_{L}$ domain connects an L-amino acid to a growing peptide with a terminal D-amino acid; an E domain catalyses the inversion of stereochemistry in amino acids.[2]

During release/cyclization, the Thd cluster is distinguished by the absence of a canonical thioesterase/cyclase and a terminal CT domain, nor are any trans-acting cyclases (such as SurE in surugamide) detected. To explain the head-to-tail macrolactamization, an epimerase-assisted mechanism analogous to that hypothesized for AnaPS (acetylaszonalenine) is proposed: thdB$_{E3}$ could orient the NH of the N-terminal Leu towards the thioester carbonyl of the C-terminal Leu, favouring ring closure.

Furthermore, given that domains E and C share sequence/structure elements, a catalytic role for epimerase in the final condensation/release step is also plausible. Overall, the modular ThdA/ThdB logic explains the mixed stereo-configuration of thermoactinoamide A and the formation of congeners through relaxed selectivity in A4.

5.7 Molecular networking and structural determination of thermoactinamide A congeners

The actual production of thermoactinoamide A and its congeners by *T. vulgaris* DSM 43016 was confirmed by MeOH/$CHCl_3$ extraction of the entire strain and LC-HRMS/MS analysis on LTQ Orbitrap, with data processing using Feature-Based Molecular Networking (FBMN).[11]

The raw LC-MS files were pre-processed in MZmine (feature detection, alignment, deconvolution), generating the .mgf of the MS^2 spectra for upload to GNPS and the construction of the molecular network; Cytoscape was used to map the chromatographic information and visualize the network. The thermoactinoamide cluster shows 11 nodes: that of thermoactinoamide A (1a, *m/z* 715.48) is the most abundant, while the remaining 10 congeners are significantly less abundant (Figure 5.11).

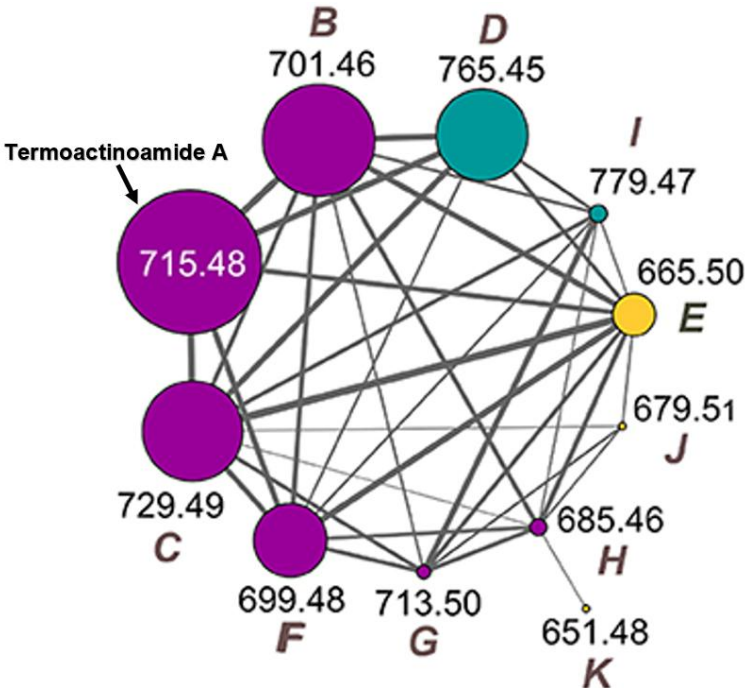


Figure 5.11. Cluster of thermoactinoamides. The nodes of thermoactinoamides containing one aromatic residue are marked in purple, while those containing two

and no aromatic residues are marked in blue and orange, respectively. The size of the node is relative to the areas of the peaks observed on the LC-MS chromatograms, and the thickness of the border is relative to the similarity of the cosine score.[2]

In the network, the area of the nodes reflects the relative quantity, and the thickness of the edges reflects the degree of spectral similarity.

Since there are no other related NRPS clusters in the genome of *T. vulgaris* DSM 43016, it can be inferred that all thermoactinoamide variants are assembled by the same multimodular BGC that synthesizes compound 1.

The high-resolution fragmentation pattern of thermoactinoamide A was taken as a model for interpreting the congeners. As in many cyclic peptides, fragmentation typically proceeds by cleavage of two amide bonds, with the loss of one, two, or three residues (α, β, γ ions), providing information on the sequence but not discriminating between Leu/Ile isomers or stereochemical aspects.[1]

These elements previously required extensive spectroscopic and chemical analysis; however, knowledge of the biosynthetic cluster could have provided decisive constraints from the outset.

In this context, the integration of molecular networking and genomic-biosynthetic information allowed the stereostructural determination of the new thermoactinamide variants, overcoming the limitations of mass spectrometry alone and guiding a more solid assignment consistent with the NRPS logic of the BGC of *T. vulgaris* DSM 43016.[2] In particular:

- B, C, F: differ from A by a single substitution consistent with the relaxed selectivity of domains A and with the aromatic recognition of $ThdA_{A3}$: B (aa2 Ile→Val), C (aa4 Val→Ile), F (aa3 Tyr→Phe).

- D and E did not match the BGC predictions and were therefore isolated. D shows Tyr in aa1 and aa3 (instead of L-Leu in aa1), a sequence confirmed by LC-HRMS/MS and 2D NMR (HMBC); the configurations of Tyr were assigned with Marfey (1-L-Tyr, 3-D-Tyr). This implies imperfect specificity for $ThdA_{A1}$, likely influenced by the truncated C domain* which may act as a gatekeeper on the selectivity of the downstream A domain.
- E, obtained in small quantities, is entirely aliphatic (L-Ile, 2×L-Leu, D-*allo*-Ile, D-Leu, L-Val) with two D residues; Marfey advanced/MS data suggest cyclodimerization of the L-Ile/Val-L-Leu-D-Leu/*allo*-Ile triad and an iterative mode of ThdB alone (single E domain → two D centres).

Molecular networking also revealed five new nodes (thermoactinoamide G–K), too scarce for isolation; putative structures (including configurations) were assigned by combining HR-MS/MS (fragmentation patterns) and NRPS logic of the BGC:

- G: Tyr→Phe (aa3) and Val→Ile/Leu (aa4);
- H: Ile/Leu→Val (aa2) and Tyr→Phe (aa3);
- I: analogue of D with $+CH_2$ for Val→Ile/Leu (aa4) and Leu→Tyr (aa1);
- J and K: all aliphatic series like E, consistent with iteration of ThdB only; putative sequences proposed (J: L-Ile-L-Leu-D-allo-Ile-L-Ile–L-Leu-D-Leu; K: L-Val-L-Leu-D-allo-Ile-L-Val-L-Leu-D-Leu).

In summary, the integration of MS/MS and biosynthetic genomics consolidates the NRPS ThdA/ThdB programming with flexible selectivity

and, in specific cases, iterative behaviour, clarifying known structures and supporting the assignment of new congeners.

5.8 Synthesis of the cyclic peptide thermoactinoamide A: unexpected epimerization

Although we were able to cultivate the productive thermophilic strain and modulate its secondary metabolism through an OSMAC (One Strain-Many Compounds) approach, our fermentations consistently produced low and variable yields of thermoactinoamide A. Heterologous expression of the relevant biosynthetic gene cluster was ongoing but still immature, with uncertainties regarding construct design, expression levels, and post-translational activation of the NRPS modules. In this context, obtaining the higher and reliable quantities required for an expanded biological program, would have required a significant increase in scale, time, and resources, with no guarantee of supply stability.

In contrast, chemical synthesis offered a faster and more cost-effective route to obtaining material on demand, uncoupled from cultivation bottlenecks, seasonal variability in metabolite production, or subtle effects of culture media that can alter the metabolic profile.

A synthetic supply also promised practical advantages beyond mere quantity: access to high-purity reference material for LC-MS/NMR benchmarking, a platform for isotope labelling or site-selective modifications to support mechanistic studies, and an orthogonal route to verify the proposed stereostructure. Given that thermoactinoamide A is a relatively compact cyclic hexapeptide, standard Fmoc-SPPS assembly followed by head-to-tail

macrocyclization seemed technically straightforward and well within the capabilities of specialized suppliers.

For reasons of speed and cost, we therefore outsourced the synthesis to a company specializing in custom peptides (Primm s.r.l.). Only later did we discover that the NMR spectra of the synthetic product differed from those of the natural compound, a discrepancy that was ultimately traced back to an unexpected epimerization during cyclization, but this was a result that we could not have foreseen at the time of the decision and does not invalidate the original rationale behind the choice of synthetic route.

This study led to the publication of the following research papers:

- Di Matteo V, Esposito G, Costantino V, Della Sala G, Teta R, Mangoni A. When Synthesis Gets It Wrong: Unexpected Epimerization Using PyBOP in the Synthesis of the Cyclic Peptide Thermoactinoamide A. *J Nat Prod.* **2024**; 87(4):948-953.

5.9 Structure Elucidation of Synthetic Thermoactinoamide A

A comparative analysis of the ^{1}H NMR spectra of the purchased synthetic peptide (**1b**) and the natural thermoactinoamide A (**1a**) (Figure 5.14) immediately revealed non-overlapping fingerprints, indicating that the two materials are not the same compound. Assuming that the natural structure had been correctly assigned and given that the amino acid sequence in the synthetic route was not in doubt, the most plausible explanation was a configuration error introduced during synthesis, namely epimerization/racemization at one of the six α carbons during macrocyclization. To verify this, we reapplied the enhanced Marfey protocol with LC-MS previously used for **1a** to **1b**.

An initial portion of 25 μg of **1b** was hydrolyzed (6 N HCl/AcOH, 1:1, 120 °C, 18 h) and the resulting amino acids were derivatized with D-FDAA (100 μL, 1%). The HPLC-ESI-HRMS retention times of the D-FDAA adducts were then compared with those of the authentic standards prepared from L-Tyr, L-Val, L-Leu, L-Ile, D-allo-Ile and with the D-FDAA derivatives obtained in parallel from natural **1a** under identical conditions.

The chromatograms (Figure 5.12) showed the presence of L-Val, D-Tyr, and D-Leu in **1b**, as in **1a**, but no trace of D-allo-Ile, indicating the conversion of D-allo-Ile to L-Ile during synthesis. Since L-Leu and L-Ile are not separated at the baseline as D-FDAA derivatives, we treated a second 25 μg aliquot of **1b** with L-FDAA. In this configuration, the chromatogram clearly showed both L-Leu and L-Ile, identifying the site of epimerization at the α carbon atom of isoleucine.

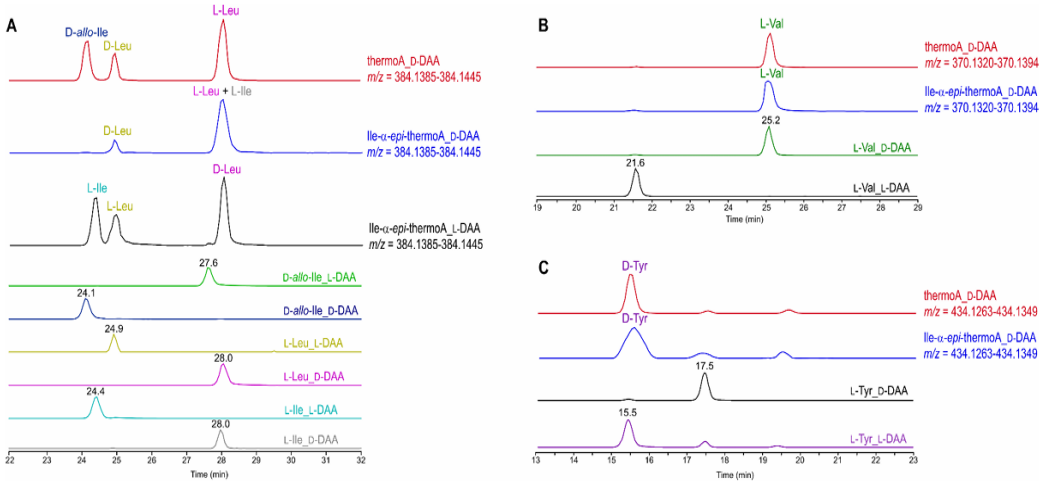


Figure 5.12. LC-MS-enhanced Marfey's analysis of compound **1b**. (A) Extracted-ion chromatograms at *m/z* 384.1415 of the D-DAA derivatives from the hydrolysis of **1a** (red trace) and **1b** (blue trace) and of the L-DAA derivatives from the hydrolysis of **1b** (black trace), along with reference chromatograms of D- and L-DAA derivatives of L-Leu, L-Ile, and D-allo-Ile. (B) Extracted-ion chromatograms at *m/z* 370.1357 of the D-DAA derivatives from the hydrolysis of **1a** (red trace) and **1b** (blue trace) and of D- and L-DAA derivatives of L-Val. (C) Extracted-ion chromatograms at *m/z* 434.1306 of the D-DAA derivatives from the hydrolysis of **1a** (red trace) and**1b** (blue trace) and of D- and L-DAA derivatives of L-Tyr. LC conditions: 2.6 µm Kinetex PFP column (100×4.6mm) maintained at 25°C eluted at 200 µL min^{-1} with 0.1% HCOOH in H_2O and MeOH; gradient program: 60% MeOH 5 min, 60%→100% MeOH over 30 min, 100% MeOH 15min.[3]

To consolidate the assignment, we recorded a complete set of 1D/2D NMR data for **1b** (^{1}H, ^{13}C, COSY, TOCSY, HSQC, HMBC) in CD_3OD, the same solvent used in the independent total synthesis report, and, for direct comparison, we compiled the complete NMR assignments of natural **1a** also in CD_3OD in Table 5.5. The TOCSY/COSY traces allowed complete mapping of the side chain from each α proton, confirming the amino acid composition of the peptide. Two key HMBC correlations, between H-2 (α-H) of L-Ile at δ 4.33 ppm and C-1 (CO) of L-Leu(III) at δ 174.3 ppm, and between H-2 of D-Tyr at δ 4.42 ppm and C-1 of L-Ile at δ 172.5 ppm, located L-Ile exactly where D-allo-Ile is found in **1a**, providing decisive proof that the commercial material is the Ile-α epimer of thermoactinoamide A. We have

therefore designated this unnatural variant as Ile-α-epi-thermoactinoamide A (**1b**).

It should be noted that the "thermoactinoamide A" described in the independent report on total synthesis (Sun et al.)[12] is identical to **1b**: its published ^{1}H and ^{13}C NMR spectra correspond to those of our synthetic sample, indicating that both synthetic efforts led to the same epimerized product rather than the natural configuration found in **1a**.

5.10 Cyclization of linear peptide YI-6

After realizing that two independent synthetic campaigns converged on the same unexpected compound, we sought to understand how a standard peptide route could lead to an incorrect stereochemical outcome.

Short peptides are routinely assembled by solid-phase peptide synthesis (SPPS): the protected C-terminal amino acid is anchored to the resin, and the chain is extended towards the N-terminal through iterative cycles of N-deprotection and coupling.[13]

Once the linear precursor is obtained, it is cleaved from the resin and, for cyclic targets, subjected to a macrocyclization step. Although SPPS is generally robust, allowing different sequences to be obtained with high purity and yield, macrocyclization remains the critical operation, in that side processes such as racemization/epimerization, dimeric formation, or oligomerization can occur.

There are numerous head-to-tail (and side chain-to-tail) ligation strategies to mitigate these problems, but each has trade-offs and none is universally risk-free.[14] This picture led us to suspect that the D-allo-Ile epimerization observed

for thermoactinoamide A likely arose during the ring closure event rather than during linear assembly.

This suspicion was reinforced by the conditions reported in Sun et al, where the linear precursor YI-6 (**2**) has D-allo-Ile at the C-terminal, the residue that must be activated for cyclization and, therefore, the most vulnerable to configuration drift under coupling conditions.

We therefore reproduced the cyclization exactly as described: out of resin, using PyBOP (benzotriazol-1-yloxy-tripyrrolidinophosphonium hexafluorophosphate) as the coupling reagent and DIEA to maintain the pH > 8. Before closing the ring, the synthetic linear peptide D-Tyr-Val-Leu-D-Leu-Leu-D-allo-Ile (YI-6, **2**) was profiled by LC-HRMS and compared with the six linear fragments obtained by partial hydrolysis of natural thermoactinoamide A. The retention time and HRMS/MS matched one of these six standards, providing further confirmation of the sequence and configuration of the linear intermediate.

The linear peptide YI-6 (*m/z* 733.4858, $[M+H]^+$, $C_{38}H_{65}N_6O_8^+$) was then cyclized under the conditions reported (Figure 5.13).

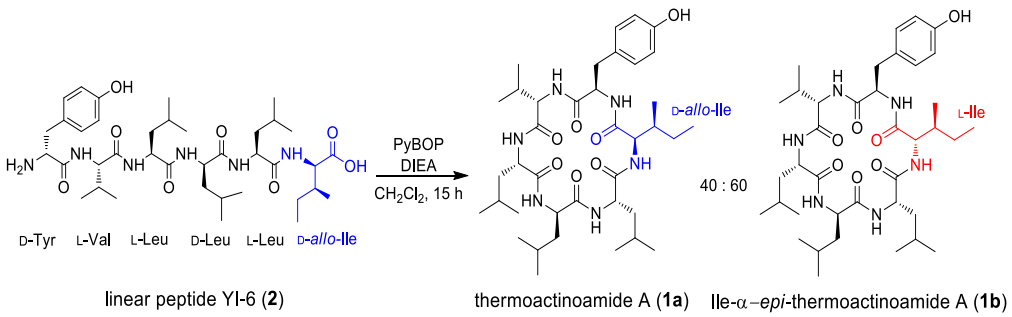


Figure 5.13. Head-to-tail cyclization of peptide YI-6 (**2**), yielding thermoactinoamide A (**1a**) and its epimer with L-Ile replacing D-allo-Ile (**1b**), i.e. Ile-α-epi-thermoactinoamide A. The acronym YI-6 indicates the linear hexapeptide **2** having D-Tyr at the N-terminus and D-allo-Ile at the C-terminus.[3]

LC-HRMS analysis of the crude showed the expected cyclic ion (*m/z* 715.4753, $[M+H]^+$, $C_{38}H_{63}N_6O_7^+$). However, the chromatogram of the ions extracted at *m/z* 715.4753 (Figure 5.29) showed two peaks: one at t_R = 15.5 min, coinciding with the natural thermoactinoamide A (**1a**), and a second at t_R = 15.3 min. Purification of the crude mixture by reverse-phase HPLC yielded two pure fractions (t_R = 19.6 and 19.1 min, Figure 5.30) in an approximate ratio of 4:6. NMR comparison established that these were thermoactinoamide A (**1a**) and Ile-α-epi-thermoactinoamide A (**1b**), respectively, the latter corresponding to the commercial synthetic sample described above.

Thus, the cyclization of linear peptide **2** reproducibly generates both epimers, with the epimerized product 1b as the major component (~60%).

A reasonable explanation is that the PyBOP-activated form of YI-6 cyclizes more slowly than its α-epimer, allowing epimerization to compete and ultimately favour the formation of **1b**. This kinetic preference could stem from greater steric strain in the natural macrocycle **1a** compared to **1b**; computational analyses dedicated to ring strain and transition state energetics are underway to verify this hypothesis.

It is also plausible that Sun et al., as well as the supplier who provided our synthetic material, obtained a similar result with two products. In their case, it appears that the most abundant cyclization product was assigned to the natural target without recognizing the minor component as **1a**. A direct NMR cross-check would likely have revealed the discrepancy, but the spectra were acquired in a different solvent, and the observed differences appear to have

been attributed to solvent effects rather than a change in configuration at the Ile α centre.

5.11 Cyclization of linear peptide LV-6

To verify whether the epimerization observed during macrocyclization was related to the intrinsic tendency to racemization of the C-terminal residue in the linear precursor (in peptide **2**, this is D-allo-Ile), we repeated the ring closure starting from a different linear permutation among the six sequences from which thermoactinoamide A can be assembled (Figure 5.28).

We selected LV-6 (**3**), i.e., Leu-D-Leu-Leu-D-allo-Ile-D-Tyr-Val, and subjected it to the same off-resin conditions used for YI-6 (**2**): PyBOP as the coupling reagent and DIEA to maintain a basic pH during cyclization. Analysis of the crude product by LC-HRMS at *m/z* 715.4753 ($[M+H]^+$, $C_{38}H_{63}N_6O_7^+$) revealed two chromatographic peaks in an approximate ratio of 1:1, one coeluting with the natural thermoactinoamide A (**1a**) at t_R = 15.5 min and a second at t_R = 14.7 min, distinct from the previously observed Ile-α-epimer (**1b**).

This result indicates that, under these conditions, epimerization occurred at the C-terminal valine, generating a Val-α-epimer that we denote **1c**.

Purification of the reaction mixture by reverse-phase HPLC (Figure 5.32) yielded both species. The fraction with the longest retention time was confirmed as **1a** by direct comparison of $^1H/^{13}C$ NMR data with those of the natural product. The fastest eluting fraction was identified as Val-epi-thermoactinoamide A (**1c**) through a combination of extensive NMR experiments and Marfey analysis. Using the improved LC-MS-assisted

Marfey protocol described above, we determined the configurations at the residue level in 1c: 2 × L-Leu, 1 × D-Leu, 1 × D-allo-Ile, 1 × D-Tyr, as in 1a, but with D-Val instead of L-Val (Figure 5.39).

Taken together, these experiments demonstrate that both linear precursors, YI-6 (**2**) and LV-6 (**3**), can provide natural thermoactinoamide A (**1a**) after cyclization, also diverging into epimeric products (**1b** from α-Ile epimerization and **1c** from α-Val epimerization). This dual result provides chemical confirmation of the original structural assignment for 1a and emphasizes that macrocyclization, rather than linear SPPS assembly, is the critical step in which α-centre epimerization can occur.

		thermoactinoamide A, **1a**		Ile-α-*epi*-thermoactinoamide A, **1b**		Val-*epi*-thermoactinoamide A, **1c**	
		δ_C (type)	δ_H (*J* in Hz)	δ_C (type)	δ_H (*J* in Hz)	δ_C (type)	δ_H (*J* in Hz)
1 (CO)		174.5, C	-	174.8, C	-	174.7, C	-
2 (α)		56.4, CH	4.84, m	57.5, CH	4.42, dd (9.9, 6.6)	54.5, CH	4.74, dd (12.2, 4.0)
3 a		39.4, CH_2	2.95, d (3.0)	37.1, CH_2	2.91, m	37.9, CH_2	3.12, dd (14.1, 4.0)
b	D-Tyr		2.94, d (3.9)	D-Tyr	2.87, m	D-Tyr	2.93, dd (14.0, 12.3)
4		128.6, C	-	128.1, C	-	129.8, C	-
5/9		131.5, CH	7.06, d (8.5)	131.4, CH	7.05, d (8.5)	131.3, CH	7.16, d (8.5)
6/8		116.3, CH	6.70, d (8.5)	116.4, CH	6.69, d (8.5)	116.1, CH	6.68, d (8.5)
7		157.5, C	-	157.6, C	-	157.1, C	-
1 (CO)		173.9, C	-	173.5, C	-	175.1, C	-
2 (α)		63.0, CH	3.65, d (5.6)	60.1, CH	4.04, d (4.0)	64.0, CH	3.65, d (9.9)
3	L-Val	30.9, CH	1.93, m	L-Val 30.1, CH	2.25, m	D-Val 29.6, CH	2.02, m
4		18.6, CH_3	0.85, d (6.9)	16.8, CH_3	0.73, d (7.0)	20.2, CH_3	1.09, d (6.6)
5		19.3, CH_3	0.79, d (7.0)	19.4, CH_3	0.50, d (7.0)	19.5, CH_3	0.95, d (6.6)
1 (CO)		174.0, C	-	174.0, C	-	174.2, C	-
2 (α)		52.1, CH	4.55, t (7.4)	52.7, CH	4.52, dd (9.2, 5.8)	53.1, CH	4.34, dd (11.9, 3.6)
3 a		41.4, CH_2	1.70, m	42.2, CH_2	1.73, m	40.6, CH_2	1.82, m
b	L-Leu(I)		1.64, m	L-Leu(I)	1.68, m	L-Leu(I)	1.62, m
4		25.9, CH	1.61, m	25.9, CH	1.62, m	25.8, CH	1.69, m
5		23.2, CH_3	0.97, d (6.5)	23.3, CH3	0.94, d (6.6)	23.9, CH_3	0.98, d (6.6)
6		22.7, CH_3	0.94, d (6.5)	22.1, CH3	0.90, d (6.6)	20.8, CH_3	0.89, d (6.6)
1 (CO)		173.9, C	-	174.5, C	-	173.1, C	-
2 (α)		52.9, CH	4.41, dd (10.8, 4.6)	54.0, CH	4.25, dd (11.2, 3.5)	54.1, CH	4.40, dd (9.3, 3.9)
3 a		40.2, CH_2	1.72, m	41.5, CH_2	1.69, m	41.9, CH_2	1.76, m
b	D-Leu(II)		1.72, m	D-Leu(II)	1.59, m	D-Leu(II)	1.56, m
4		25.9, CH	1.57, m	26.2, CH	1.69, m	26.5, CH	1.58, m
5		23.7, CH_3	0.93, d (6.7)	21.3, CH_3	0.90, d (5.8)	22.4, CH_3	0.98, d (6.0)
6		21.1, CH_3	0.87, d (6.5)	23.6, CH_3	0.97, d (5.8)	23.8, CH_3	0.91, d (6.0)
1 (CO)		176.6, C	-	174.3, C	-	175.8, C	-
2 (α)		54.1, CH	4.39, t, (7.7)	53.2, CH	4.34, t, (7.4)	54.5, CH	4.39, dd (8.3, 7.2)
3 a		40.6, CH_2	1.56, m	40.3, CH_2	1.59, m	40.6, CH_2	1.62, m
b	L-Leu(III)		1.56, m	L-Leu(III)	1.59, m	L-Leu(III)	1.52, m
4		26.0, CH	1.61, m	25.8, CH	1.59, m	26.0, CH	1.58, m
5		22.8, CH_3	1.00, d (6.5)	23.0, CH_3	0.91, d (4.7)	22.6, CH_3	0.99, d (6.5)
6		22.8, CH_3	0.94, d (6.5)	22.6, CH_3	0.97, d (4.7)	23.1, CH_3	0.93, d (6.5)
1 (CO)		173.4, C	-	172.5, C	-	174.5, C	-
2 (α)		58.3, CH	4.31, d (4.1)	57.9, CH	4.33, d (8.1)	58.6, CH	4.23, d (4.1)
3	D-*allo*-Ile	37.3, CH	2.13, m	L-Ile 39.0, CH	1.81, m	D-*allo*-Ile 36.3, CH	1.92, m
4		14.7, CH_3	0.81, d (7.0)	15.4, CH_3	0.87, d (6.9)	14.6, CH_3	0.55, d (7.0)
5 a		27.4, CH_2	1.27, m	25.9, CH_2	1.50, m	27.4, CH_2	1.20, m
b			1.27, m		1.07, m		1.68, m
6		12.0, CH_3	0.89, t, (7.4)	11.5, CH_3	0.89, t, (7.4)	12.0, CH_3	0.82, t (7.5)

Table 5.5. NMR Spectroscopic Data (700 MHz, CD_3OD) for Natural Thermoactinoamide A (**1a**) and its Synthetic Epimers **1b** and **1c**.[3]

5.12 Experimental section

5.12.1 General experimental procedures

Optical rotation was measured with a Jasco P-2000 polarimeter (Jasco Europe s.r.l., Cremella, Italy) at the sodium D-line. 1D NMR and 2D NMR

experiments were carried out at 600 and 700 MHz on a Bruker Avance Neo spectrometer using CD_3OD; chemical shifts were referenced to the residual solvent signal (CD_3OD: δ_H 3.31, δ_C 49.00). The HSQC spectra were optimized for $^1J_{CH}$ = 142 Hz, and the HMBC experiments for $^{2,3}J_{CH}$ = 8.3Hz. A Thermo LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific Inc.,Waltham, MA, USA) combined with a Thermo U3000 HPLC system, was used to record high-resolution ESI-MS and HR-ESI-HPLC experiments. High-performance liquid chromatography (HPLC) separations were achieved on an Agilent 1260 Infinity quaternary LC apparatus (Agilent Technology, Cernusco sul Naviglio, Italy), equipped with a diode-array detector (DAD).

5.11.2 Collection, extraction, isolation

The thermophilic producer *Thermoactinomyces vulgaris* ISCAR 2354, used for the isolation and characterization of thermoactinoamide A, was collected from a coastal hot spring in Icelandic seawater (64°57′52″ N, 23°5′21.53″ W). A second isolate of the same species, ISCAR 2850, obtained from a sediment core sample from a hydrothermal vent at a depth of 400 m (66°36′51″ N, 17°39′09″ W), also contained thermoactinoamide A (**1a**). Both strains were assigned to *T. vulgaris* by 16S rRNA gene sequencing (GenBank accessions M372920 and M372921), showing 100% identity with the *T. vulgaris* entries in GenBank.

For cultivation and extraction, strains were grown for 24 hours at 60 °C in 250 mL flasks containing 166 medium with 1% NaCl. The entire culture was then sonicated for 3 minutes with a Branson Sonifier 250 (continuous duty cycle; controls set to 4/20) to disrupt the cells. The broth was transferred to a

1 L centrifuge bottle, mixed with 100 mL of n-butanol saturated with water containing 10% MeOH, and centrifuged (10 min, 10,000 g, 4 °C).

The upper organic phase was collected and extraction with n-butanol was repeated. The combined organic layers were filtered through a 0.45 μm PTFE membrane and concentrated on a rotary evaporator at 40 °C, yielding the extract used for subsequent analyses.

The organic fraction ISCAR 2354 (739 mg) was examined by LC-ESI-HRMS on a Thermo LTQ Orbitrap XL, using a Kinetex C18 column (5 μm, 50 × 2.10 mm) maintained at 25 °C. The mobile phases were A = H_2O with 0.1% HCOOH and B = CH_3CN; the program started at 30% B for 5 min, followed by a linear ramp from 30% to 99% B in 17 min, with a flow of 200 μL min^{-1}. The TIC was dominated by six closely related components; MS/MS spectra were acquired for each ion, and their fragmentation signatures indicated cyclic peptides. The main constituent was purified by reverse-phase HPLC using a Kinetex C18 (10 μm, 250 × 10 mm) and A = 0.1% HCOOH in H_2O / B = MeOH according to the following program: 60% B for 5 min, then 60% → 100% B for 17 min, with a 13 min hold at 100% B; flow 5 mL min^{-1}, UV detection at 280 nm. This yielded thermoactinoamide A (**1a**) as a pure product (1.2 mg, t_R 18.7 min).

Thermoactinoamide A (**1a**): colorless, amorphous solid; $[\alpha]^{25}_D$ +80 (c 0.5, MeOH). ECD in MeOH (412 μM) showed λ_{max} ($\Delta\varepsilon$) 280 (+0.5). $^1H/^{13}C$ NMR data are reported in Table xxx. HRESIMS (positive mode; 0.1% HCOOH in MeOH) gave *m/z* 715.4752 $[M+H]^+$ (calcd for $C_{38}H_{62}N_6O_7$, 737.4753), *m/z* 737.4570 $[M+Na]^+$ (calcd for $C_{38}H_{62}N_6O_7Na^+$, 737.4572), *m/z* 377.2147

$[M+Ca]^{2+}$ (calcd for $C_{38}H_{62}N_6O_7Ca^{2+}$, 377.2147), and *m/z* 799.4277 $[M+Ca+HCOO]^+$ (calcd for $C_{39}H_{63}N_6O_9Ca^+$, 799.4277).

5.11.3. Bioinformatic analysis

For the comparison of bacterial genomes, the Average Nucleotide Identity (ANI) was calculated using the Pairwise ANI tool (Chen et al., 2019)[15] available on the DOE Joint Genome Institute portal. The biosynthetic pathways of secondary metabolites were identified by genome mining with antiSMASH 4.0 (bacterial version, stringency set to "relaxed") activating the options KnownClusterBlast, ClusterBlast, SubClusterBlast, ActiveSiteFinder, Cluster Pfam and Pfam-based GO, and with PRISM 3 (parameters: structure limit = 50, window = 10,000; search enabled for all biosynthetic domain families; ORF prediction performed with all available methods). The selectivity of NRPS adenylation domains was estimated with NRPSpredictor2 (Röttig et al., 2011)[16], while identification and classification of condensation (C) and epimerization (E) domains were obtained with NaPDoS (Ziemert et al., 2012)[10].

In the genome of *Thermoactinomyces vulgaris* DSM 43016, the thd cluster was divided into two contigs (Ga0070019_105, accession REFP01000011; Ga0070019_114, accession REFP01000005). The two contigs were reassembled using a guided method, using the contig with the intact thd cluster from *Thermoactinomyces* AS95 (NODE_4; accession LSVF01000006) as a template. Alignment and assembly were performed with blastn (Altschul et al., 1990)[17] and SeqMan (DNASTAR, v5.00.

5.11.4. LC-HRMS and LC-HRMS/MS

A 500 mL culture of *Thermoactinomyces vulgaris* DSM 43016 (DSMZ) was incubated for 24 h at 50 °C in CYC medium (Czapek Dox agar 48.0 g/L, yeast extract 2.0 g/L, casamino acids 6.1 g/L, tryptophan 0.02 g/L, sterile Milli-Q H_2O). The culture was frozen and lyophilized, then rehydrated with 4 mL of distilled water and sonicated for 5 min. The resulting slurry was extracted with MeOH/$CHCl_3$ (2:1, 6 mL), paper-filtered, and evaporated to give 20.6 mg of crude extract, following a previously reported protocol for low-molecular-weight metabolites (Costantino et al., 2012)[18]. The residue was dissolved in MeOH (10 mg/mL) for LC-HRMS/MS measurements.

Analyses were carried out on a Thermo LTQ Orbitrap XL high-resolution ESI mass spectrometer coupled to a Thermo U3000 HPLC (solvent reservoir, in-line degasser, binary pump, refrigerated autosampler). Separation used a Kinetex C18 column (5 μm, 50 × 2.10 mm) at ambient temperature, with a 200 μL min^{-1} flow. Mobile phases were A = H_2O + 0.1% HCOOH and B = CH_3CN; the gradient was 30% B for 5 min, then 30 → 99% B over 30 min, followed by 100% B for 3 min. Spectra were acquired in positive-ion mode with spray voltage 4.8 kV, capillary temperature 285 °C, sheath N_2 32 units (~150 mL min^{-1}), and auxiliary N_2 15 units (~50 mL min^{-1}).

Data collection employed data-dependent acquisition (DDA): within *m/z* 610-800, the top three ions from each full scan were subjected to HRMS/MS (CID) using an isolation width 2.0 Th, normalized collision energy 35, Activation Q 0.250, and activation time 30 ms.

5.11.5. LC-HRMS/MS Data Processing and Molecular Networking

Raw LC-HRMS/MS files were handled following a feature-based molecular networking workflow. Data were imported into MZmine 2.53 (Pluskal et al., 2010) and subjected to mass detection at MS^1 and MS^2 levels with a noise threshold of 1,000. Chromatogram building used the ADAP module (Myers et al., 2017) with minimum height 1,000 and *m/z* tolerance 0.05 Da or 20 ppm. Deconvolution employed the Local Minimum Search with chromatographic threshold 5%, minimum RT range 0.50 min, minimum relative height 30%, minimum absolute height 10,000, peak top/edge ratio ≥1.3, and peak duration 0.0-6.0 min. Alignment was performed using the Join aligner (*m/z* tolerance 0.05 Da/20 ppm, absolute RT tolerance 0.6 min). The following adducts/isotopes were removed: [M+Na–H], [M+K–H], [M+Mg–2H], [$M+NH_3$], [$M–Na+NH_4$], and [M+1, ^{13}C]; features without MS/MS were also excluded. Feature tables were exported as .mgf (for GNPS) and .csv (RT, peak area, peak height). The dataset was deposited to MassIVE (MSV000085201).

A molecular network (Wang et al., 2016) was generated on GNPS (Metabolomics workflow) with parent tolerance 0.02 Da, fragment tolerance 0.2 Da, cosine ≥ 0.71, and ≥ 6 matched peaks, retaining only reciprocal top-10 neighbors. Library searches used cosine ≥ 0.70 with ≥ 7 matched peaks. The completed task (https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=a365d81e928f4d0dbc5c76b0dff4515b) was mapped and visualized in Cytoscape 3.7.2 (Shannon et al., 2003) by linking the .csv chromatographic metrics to network nodes.

5.9.6. Bioactivity Assays

In order to evaluate the growth-inhibitory effects of thermoactinoamide A (**1a**), the xCELLigence System Real Time Cell Analyzer (ACEA Biosciences, San Diego, CA, USA) was used for real-time monitoring of cancer cell proliferation after drug exposure, as previously described (Caso et al., 2019)[19]. Data from biological assay represent the mean (± standard deviation, SD) of three independent experiments. Two-group comparisons were performed using Student's t-test. $p < 0.05$ were considered to be statistically significant. Statistical analysis was performed using the GraphPad Prism Software Version 5 (GraphPad Software Inc., San Diego, CA, USA).

Broth microdilution assays in 96-well plates were run against Gram-positive strains (*S. aureus* ATCC 6538, *E. faecalis* ATCC 29212, *B. cereus* ATCC 14579) and Gram-negative strains (*E. coli* ATCC 8739, *P. aeruginosa* ATCC 13388, *K. pneumoniae* ATCC 11296). For each organism, a 5 mL LB preculture was started from a single colony and incubated overnight at 37 °C with shaking. Cultures were then standardized to OD_{620} = 0.03 and 200 µL aliquots dispensed into wells containing twofold serial dilutions of the peptide. Plates were incubated overnight at 34 °C, 450 rpm, and read at 620 nm. The MIC (triplicate determinations) was the lowest concentration producing complete growth inhibition. Positive-control MICs were: ampicillin - *S. aureus* 0.11 µM, E. faecalis 4.59 µM; gentamicin - *B. cereus* 17.2 µM, *P. aeruginosa* 10.5 µM; chloramphenicol - *E. coli* 19.3 µM, *K. pneumoniae* 3.17 µM.

5.11.7. Advanced Marfey's analysis of thermoactinoamide A (**1a**)

An aliquot of compound **1a** (12 µg) was hydrolysed in 6 N HCl/AcOH (1:1) at 120 °C for 12 h; residual acid was removed under N_2. The hydrolysate was dissolved in TEA/acetone (2:3, 100 µL) and derivatized with D-FDAA (1% in CH_3CN/acetone 1:2; 100 µL) at 50 °C for 1.5 h. After drying, the D-FDAA adducts of Tyr, Val, Leu, Ile were re-dissolved in MeOH (100 µL). Authentic L-Tyr, L-Val, L-Leu, L-Ile, D-allo-Ile were derivatized with L-FDAA and D-FDAA to provide reference standards. Derivatives of 1a were analysed by HPLC-ESI-HRMS on a Kinetex PFP, 2.6 µm, 100 × 4.6 mm at 25 °C, 200 µL min^{-1}; gradient: 60% MeOH, 5 min → 60→100% MeOH over 30 min → 100% MeOH for 15 min. Spectra were acquired in positive-ion mode and processed with Xcalibur. Retention-time matching versus standards provided the configuration assignments.

5.11.8. Partial hydrolysis of thermoactinoamide A (**1a**)

A 34 µg portion of **1a** was treated with 6 N HCl/AcOH (1:1) at 100 °C for 2 h, concentrated under N_2, reconstituted in MeOH (50 µL), and purified by HPLC under the same conditions, with the HRMS used as detector via a post-column split (10% to MS). Two linear peptides (t_R 23.8 and 24.9 min) were isolated, then fully hydrolysed and derivatized with D-FDAA as above for subsequent analysis.

5.11.9. Advanced Marfey's analysis of **1b** and **1c**

Compounds **1b** (25 µg) and **1c** (100 µg) were hydrolysed in 6 N HCl/AcOH (1:1) at 120 °C for 18 h; residual acid was removed under N_2. Each

hydrolysate was dissolved in TEA/acetone (2:3, 100 µL) and derivatized with D-FDAA (1% in CH_3CN/acetone 1:2, 100 µL) at 50 °C for 2 h; mixtures were dried and reconstituted in MeOH (200 µL). A second 25 µg aliquot of **1b** was processed identically but reacted with L-FDAA. Authentic standards (L-Tyr, L-Val, L-Leu, L-Ile, L-allo-Ile) were derivatized with both L- and D-FDAA to provide references.

Derivatives (D-/L-DAA for 1b, D-DAA for 1c) were analyzed by HPLC-ESI-HRMS on a Kinetex PFP, 2.6 µm, 100 × 4.6 mm column at 25 °C, 200 µL min^{-1}. Elution: 60% MeOH (5 min), 60 → 100% MeOH (30 min), then 100% MeOH (15 min) in 0.1% HCOOH/H_2O–MeOH.

MS operated in positive-ion mode with spray 4.8 kV, capillary 285 °C, sheath N_2 32 units (~230 mL min^{-1}) and aux N_2 15 units (~150 mL min^{-1}).

Data-dependent acquisition (DDA) triggered four HRMS/MS scans after each full MS for the four most intense ions over *m/z* 150–2000 at 60,000 resolution. CID settings: isolation width 4.0, NCE 35, Activation Q 0.250, activation time 30 ms. Raw data were processed with the Xcalibur software suite.

5.11.10. Head-to-tail cyclization of linear peptides YI-6 (**2**) and LV-6 (**3**)

The macrocyclization of linear peptide **2** (1.8 mg) was performed following Sun et al., 2019. Peptide **2**, dissolved in dry CH_2Cl_2 (4.8 mL), was treated with PyBOP in dry CH_2Cl_2 (400 µL, 20 mg mL^{-1}) and, after addition of DIEA (0.7 µL), the mixture was stirred overnight. The reaction was evaporated and checked by LC-HRMS on a 5 µm Kinetex C18 (50 × 2.10 mm) at 25 °C with A = 0.1% HCOOH/H_2O, B = CH_3CN; gradient: 30% B, 5 min → 30→99%

B, 17 min → 99% B, 1 min → 99→100% B, 1 min → 100% B, 10 min; 200 µL min^{-1}. The residue was then purified by RP-HPLC on a Luna C18 (250 × 4.60 mm, 5 µm) under 30% B, 5 min → 30→95% B, 17 min → 95% B, 13 min at 1 mL min^{-1}, 280 nm, affording **1a** (t_R 19.5 min, 0.5 mg) and **1b** (t_R 19.0 min, 0.7 mg) in pure form.

Ile-α-epi-thermoactinoamide A (**1b**). Colorless amorphous solid; $[\alpha]^{25}_D$ 60 (c 0.2, MeOH); ECD (1.3 mM, MeOH) λ_{max} ($\Delta\varepsilon$) 226 (−8); $^1H/^{13}C$ NMR: see table 5.5. HRESIMS (positive, 0.1% HCOOH/MeOH): *m/z* 715.4763 $[M+H]^+$ (calc. 715.4753), *m/z* 737.4582 $[M+Na]^+$ (calc. 737.4572).

Cyclization of linear peptide **3** (1.2 mg) was carried out analogously: 3.9 mL dry CH_2Cl_2, PyBOP (200 µL, 20 mg mL^{-1}) and DIEA (0.5 µL), then the same work-up and chromatographic conditions, giving **1a** (t_R 19.5 min, 0.4 mg) and **1c** (t_R 18.4 min, 0.4 mg).

Val-epi-thermoactinoamide A (**1c**). Colorless amorphous solid; $[\alpha]^{25}_D$ +20 (c 0.1, MeOH); ECD (0.28 mM, MeOH) λ_{max} ($\Delta\varepsilon$) 231 (−4); $^1H/^{13}C$ NMR: see Table 5.5. HRESIMS (positive, 0.1% HCOOH/MeOH): *m/z* 715.4765 $[M+H]^+$ (calc. 715.4753), *m/z* 737.4583 $[M+Na]^+$ (calc. 737.4572).

5.12. Spectroscopic data

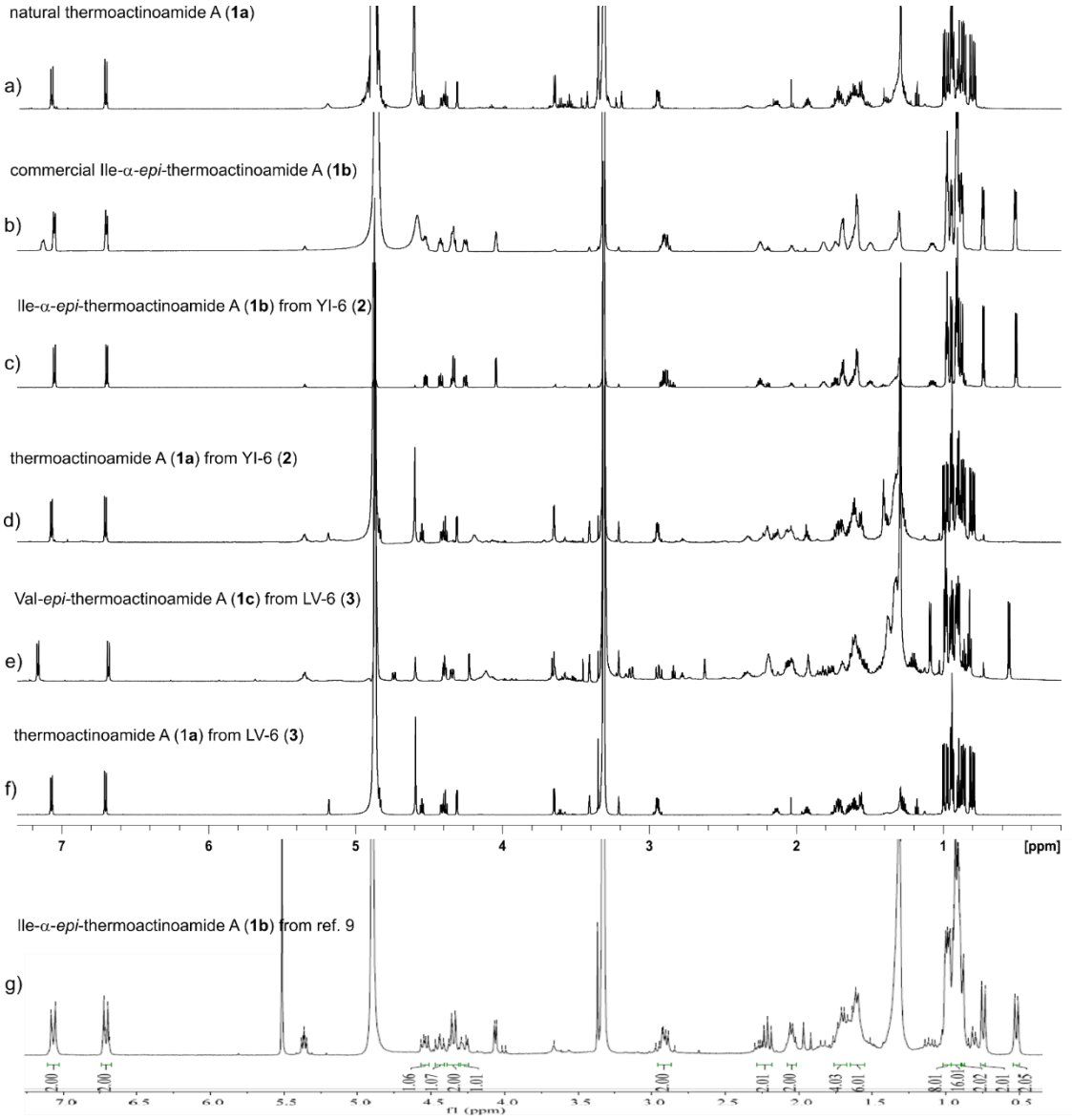


Figure 5.14. Comparison of ^{1}H NMR spectra (CD_3OD) of natural (a) and synthetic (d, f) thermoactinoamide A (**1a**) and of its two epimers **1b** (b, c, g) and **1c** (e).[3]

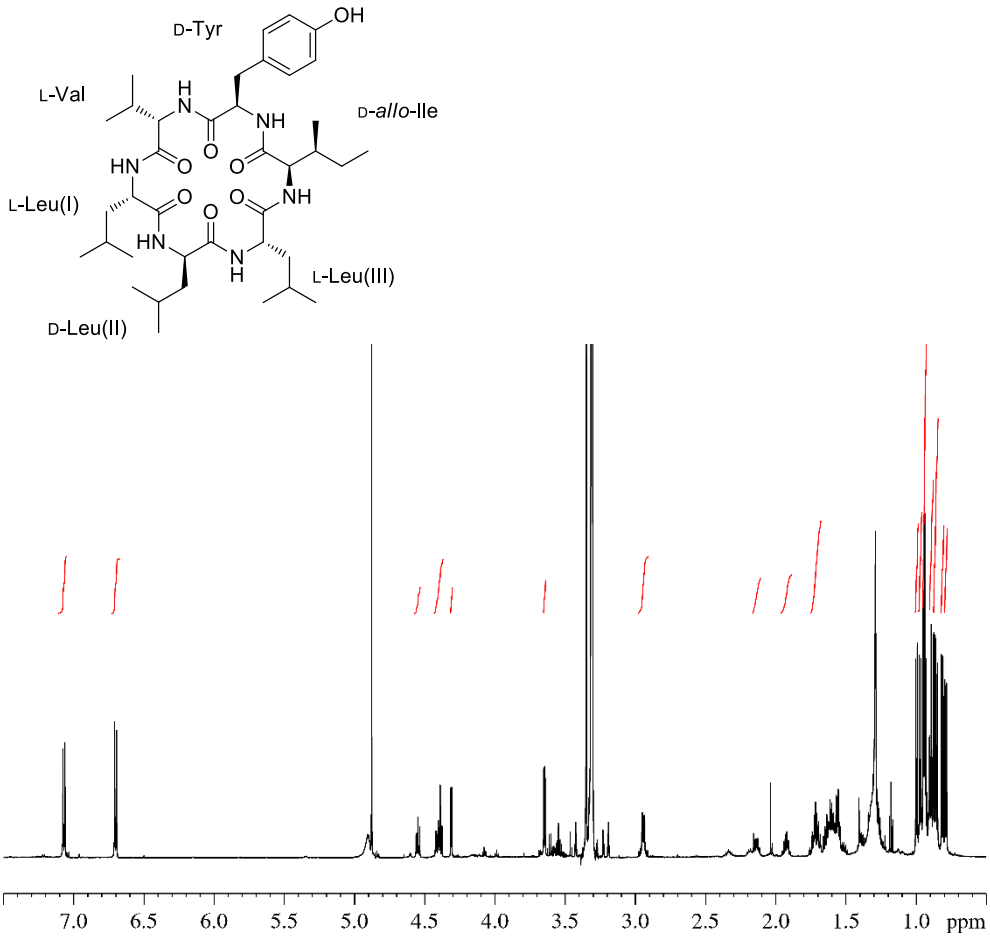


Figure 5.15. ^{1}H-NMR spectrum of thermoactinoamide A (**1a**) (600 MHz, CD_3OD).[3]

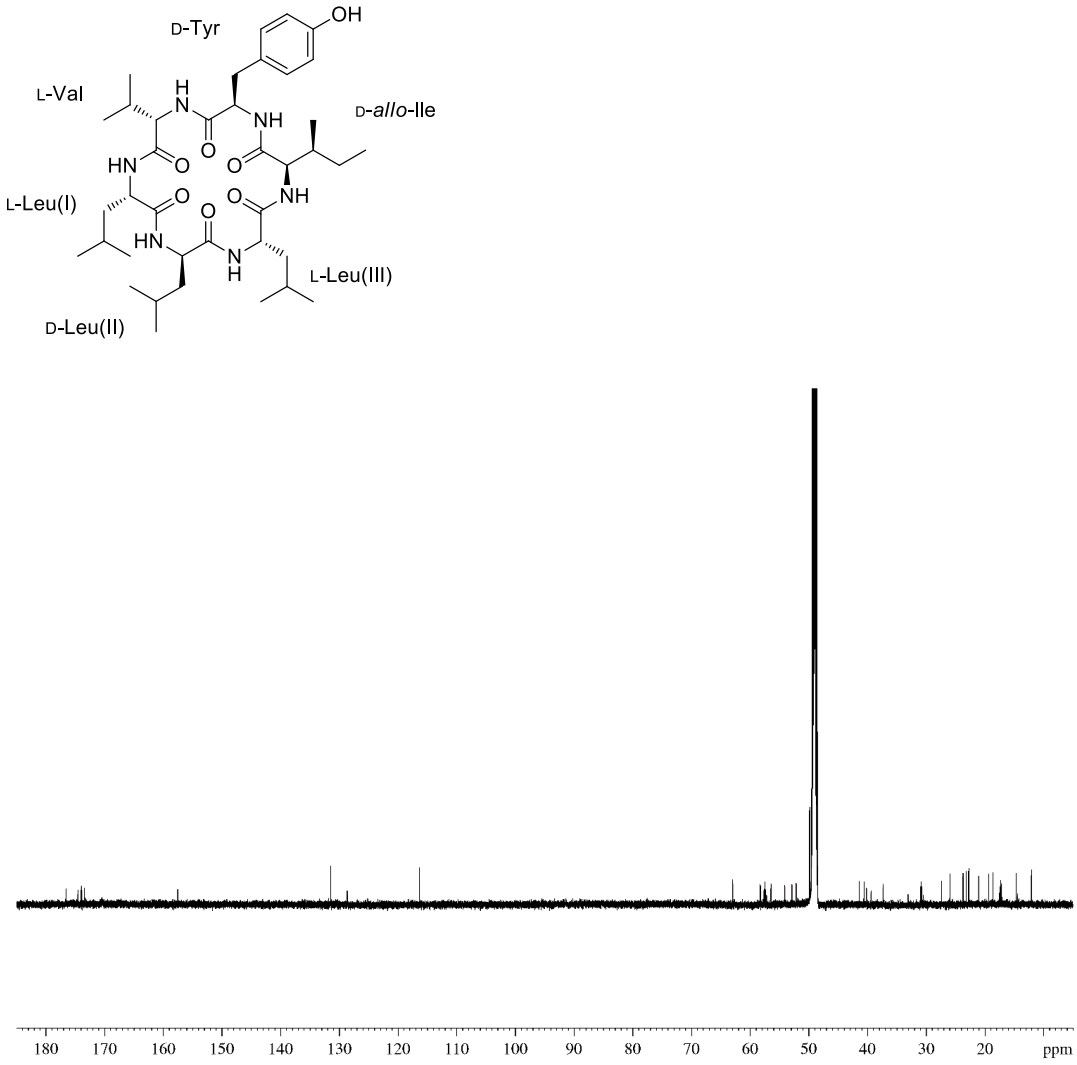


Figure 5.16. ^{13}C-NMR spectrum of thermoactinoamide A (**1a**) (600 MHz, CD_3OD).[3]

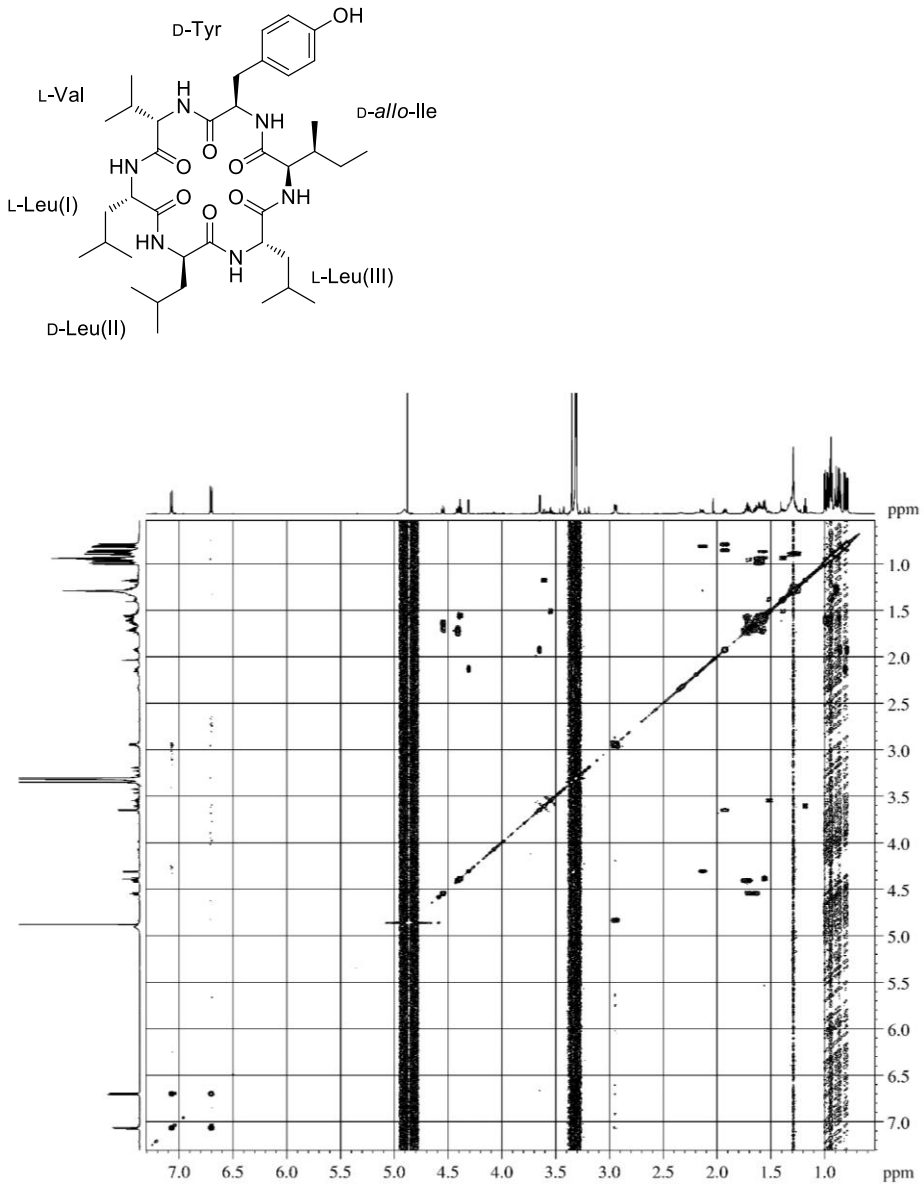


Figure 5.17. COSY spectrum of thermoactinoamide A (**1a**) (600 MHz, CD_3OD).[3]

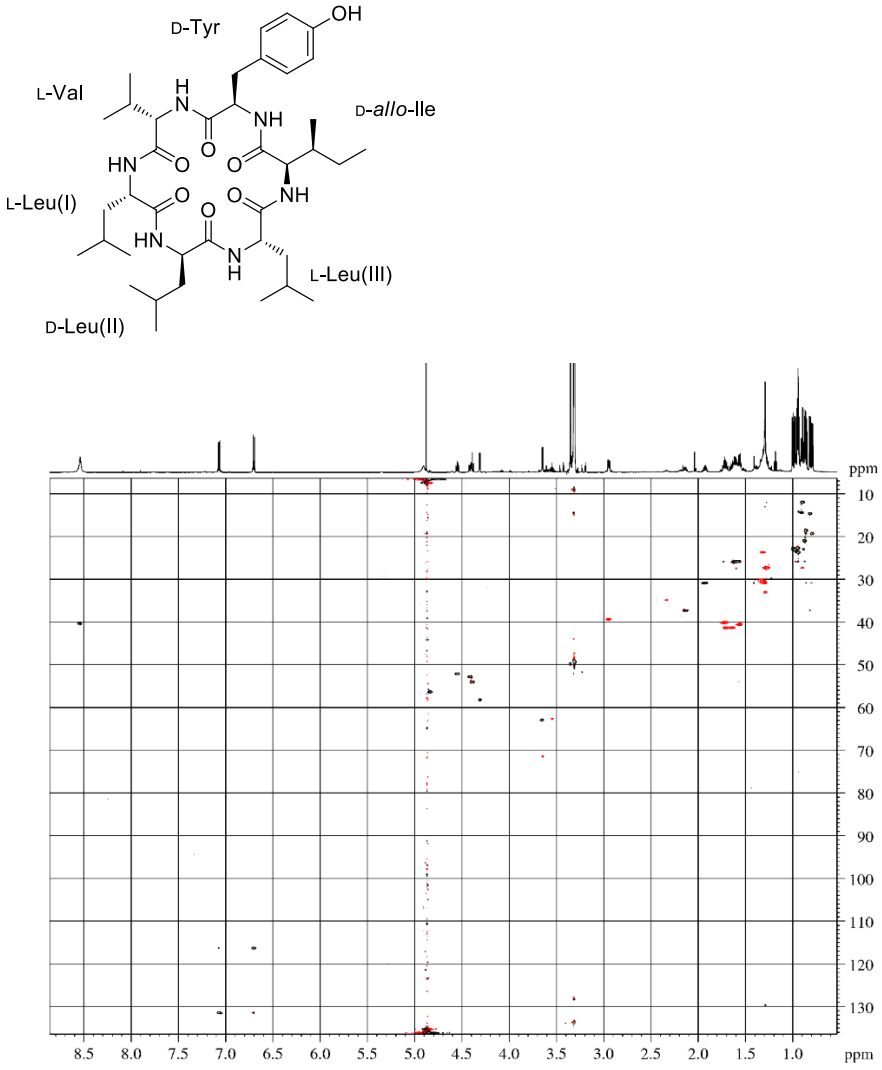


Figure 5.18. HSQC spectrum of thermoactinoamide A (**1a**) (600 MHz, CD_3OD).[3]

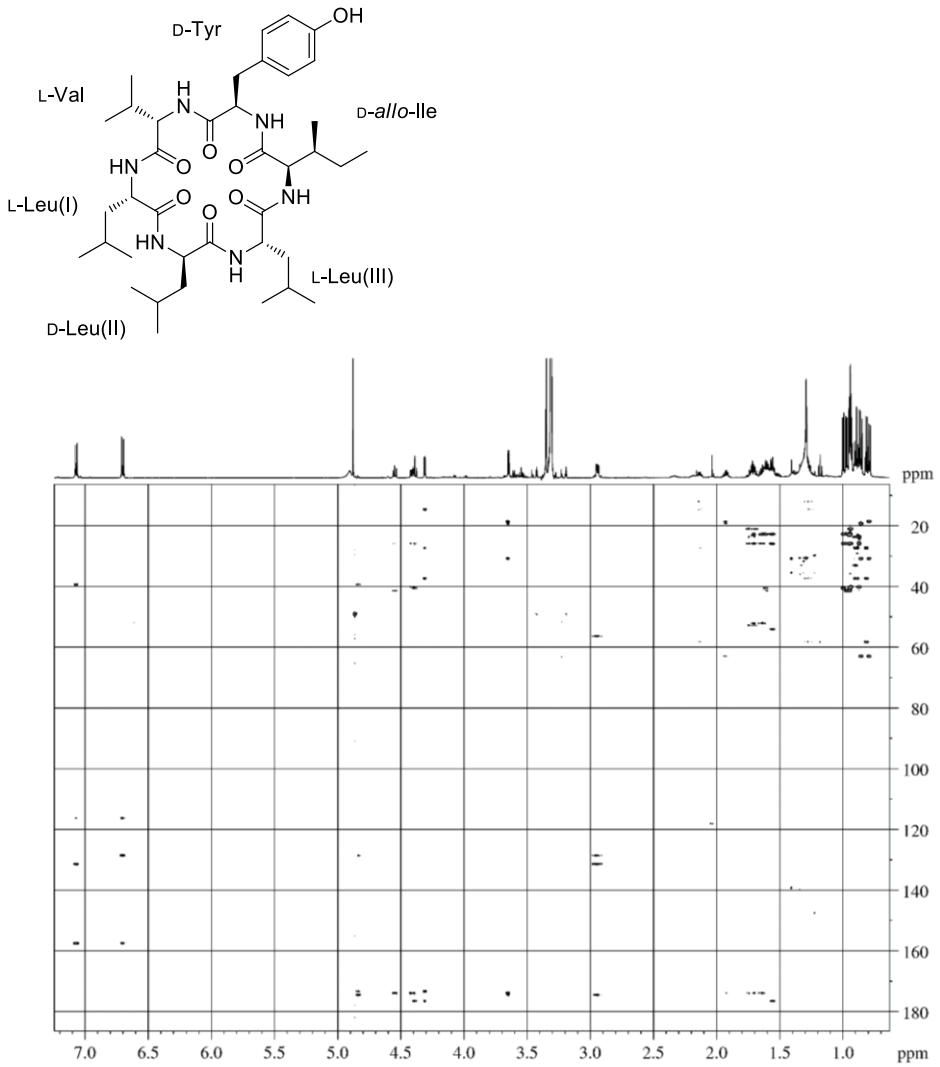


Figure 5.19. HMBC spectrum of thermoactinoamide A (**1a**) (600 MHz, CD_3OD).[3]

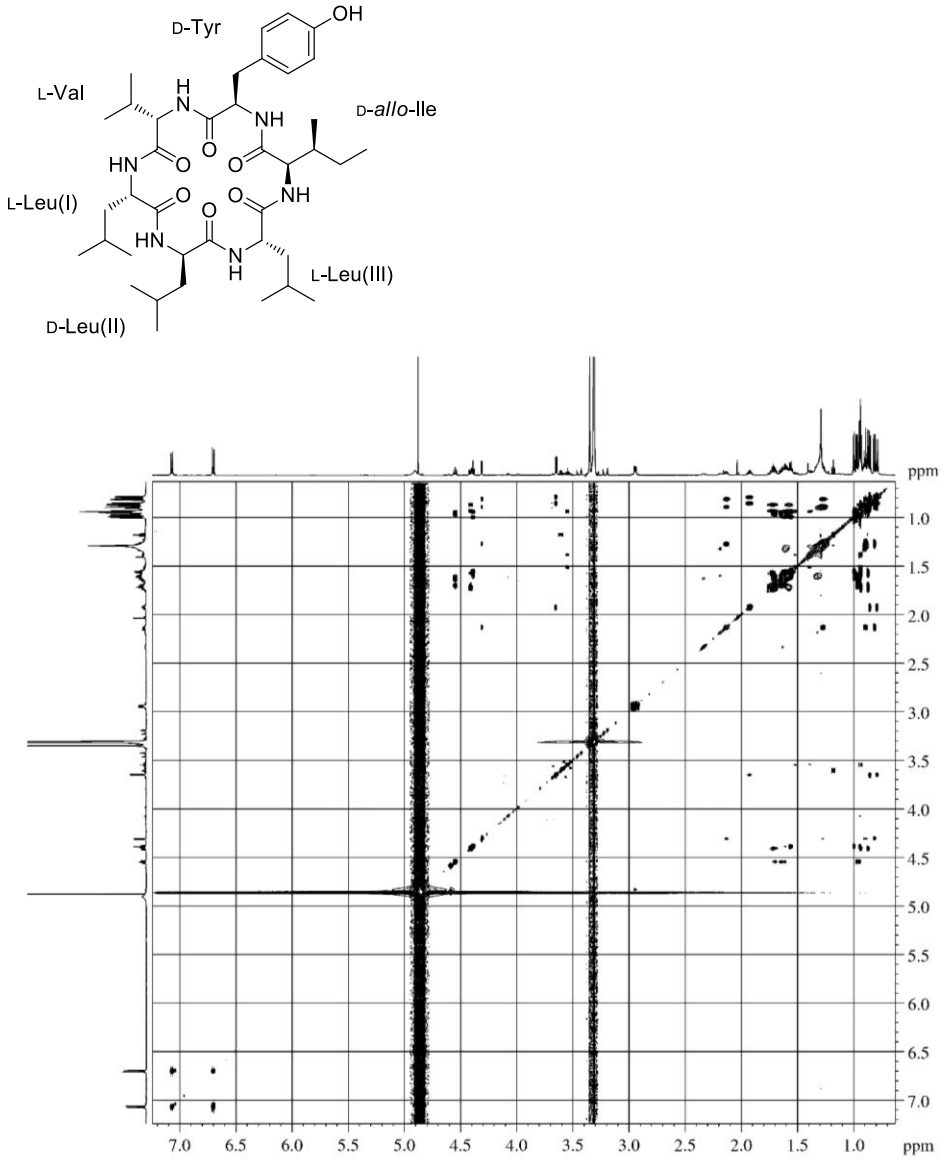


Figure 5.20. TOCSY spectrum of thermoactinoamide A (**1a**) (600 MHz, CD_3OD).[3]

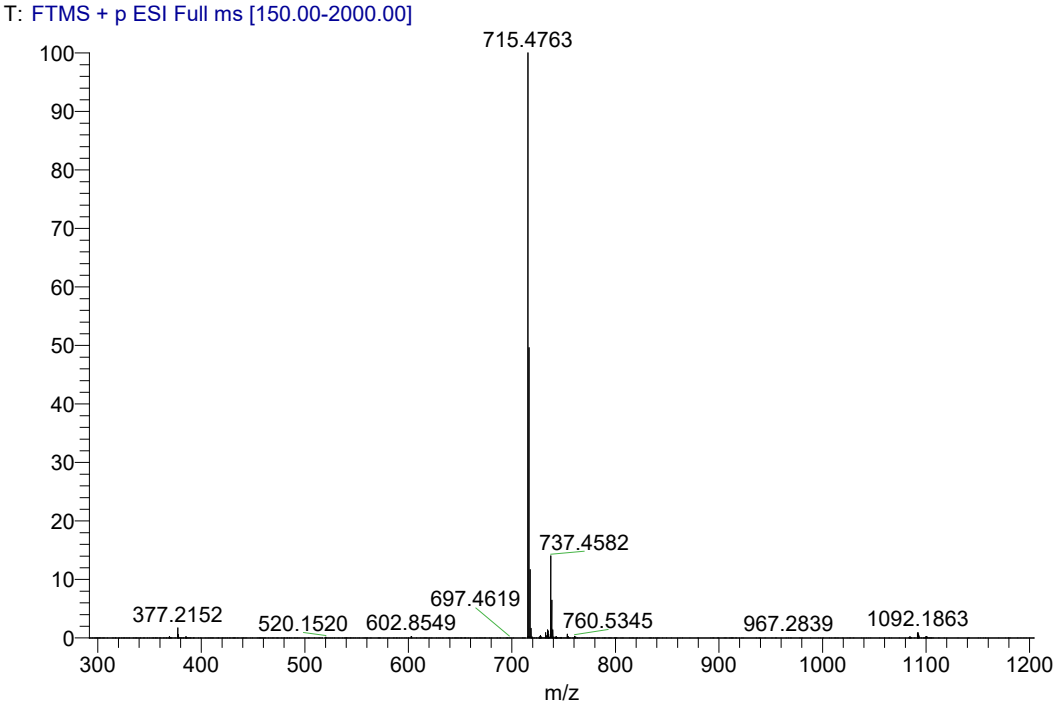


Figure 5.21. High resolution ESI mass spectrum of Ile-α-*epi*- thermoactinoamide A (**2b**).[3]

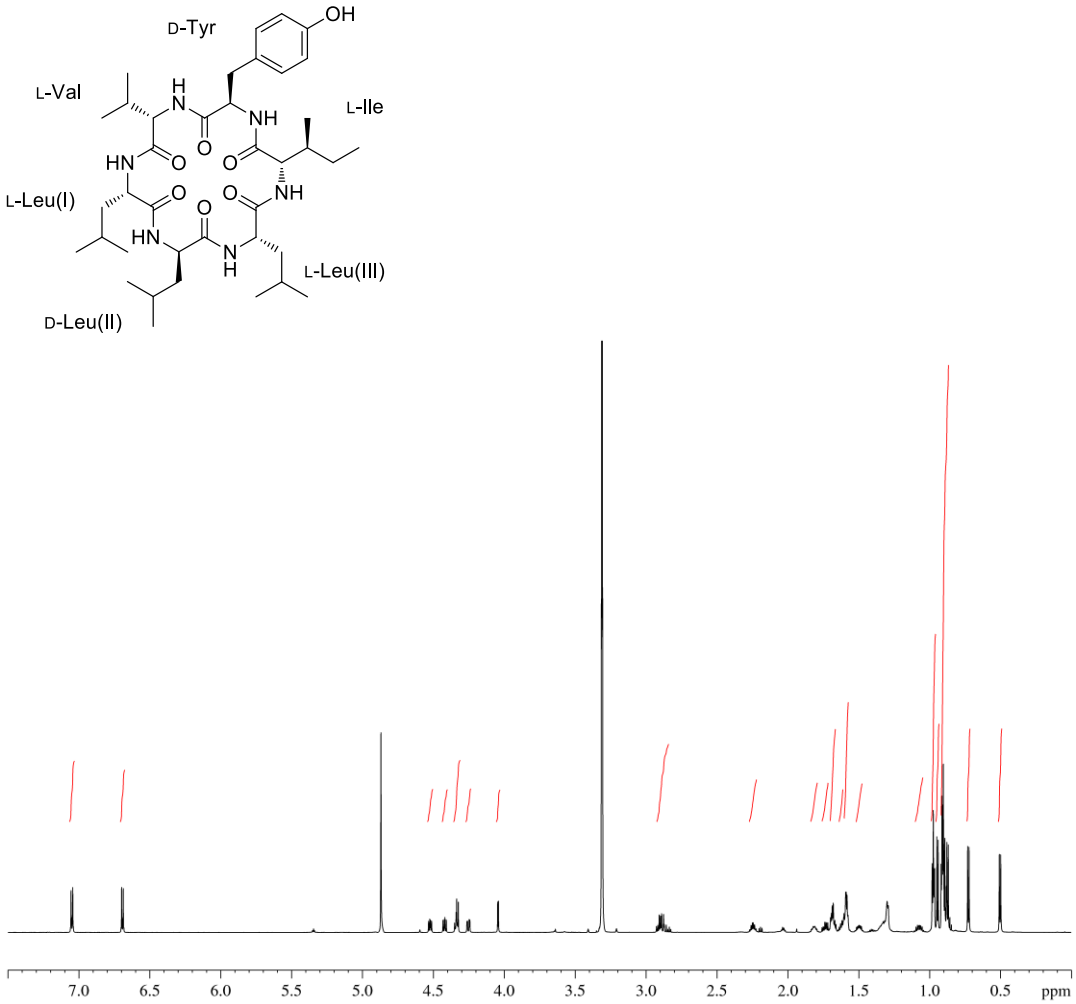


Figure 5.22. ^{1}H-NMR spectrum of Ile-α-*epi*-thermoactinoamide A (**2b**) (700 MHz, CD_3OD).[3]

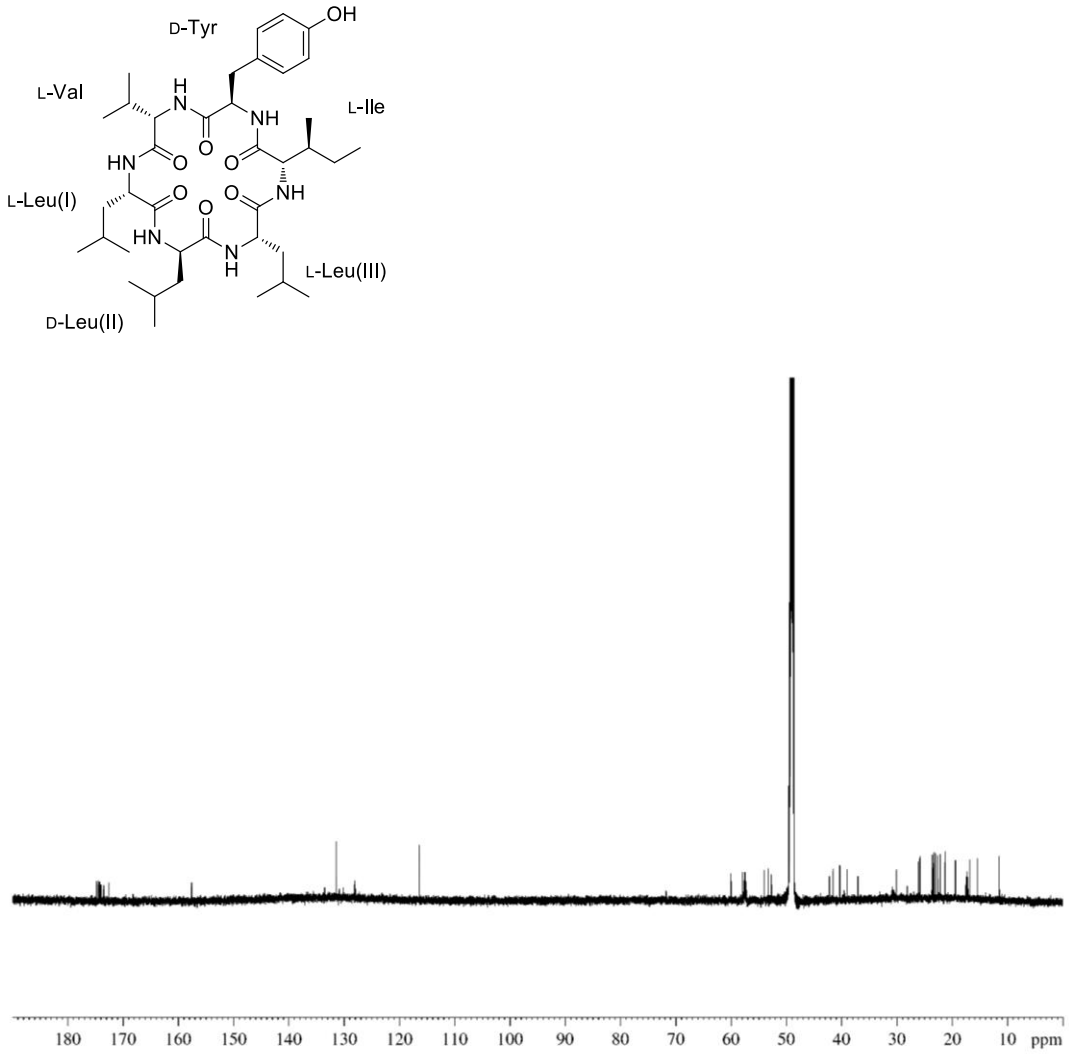


Figure 5.23. ^{13}C-NMR spectrum of Ile-α-*epi*-thermoactinoamide A (**2b**) (700 MHz, CD_3OD).[3]

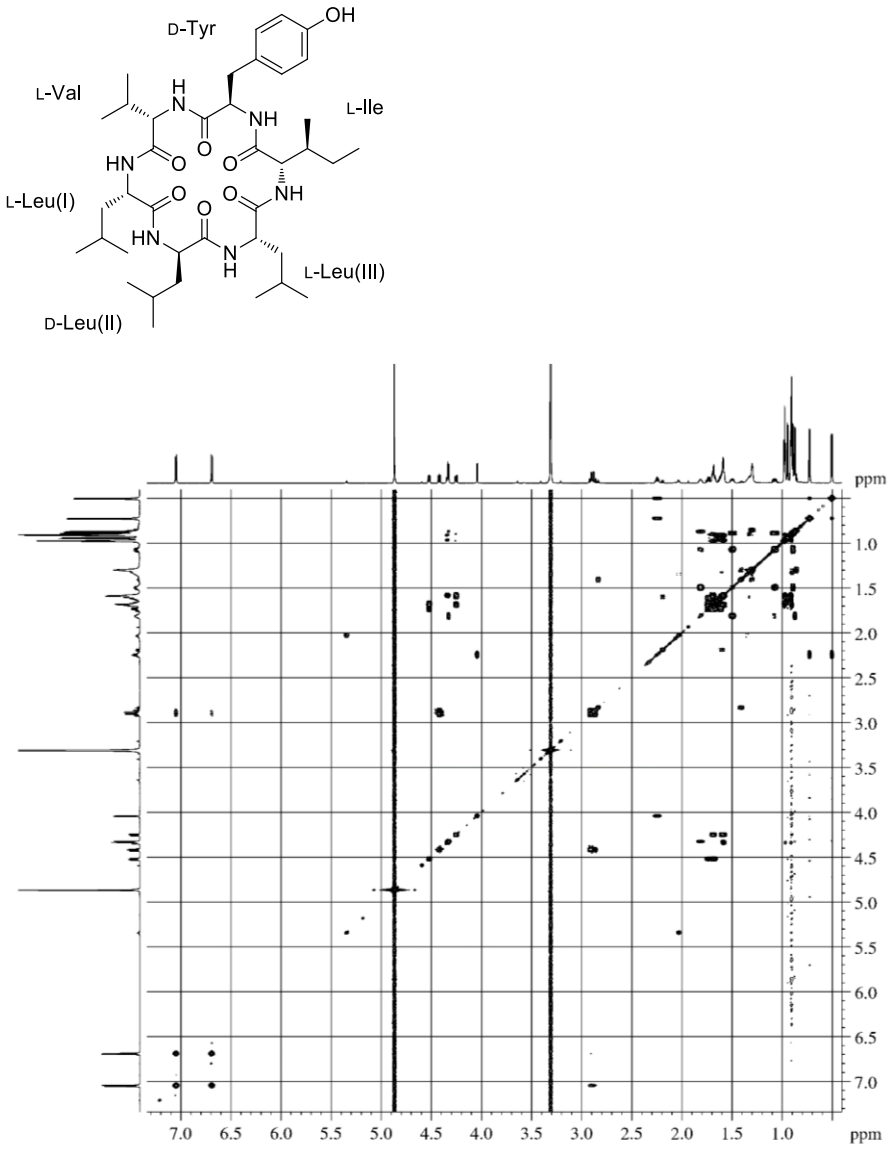


Figure 5.24. COSY spectrum of Ile-α-*epi*-thermoactinoamide A (**2b**) (700 MHz, CD_3OD).[3]

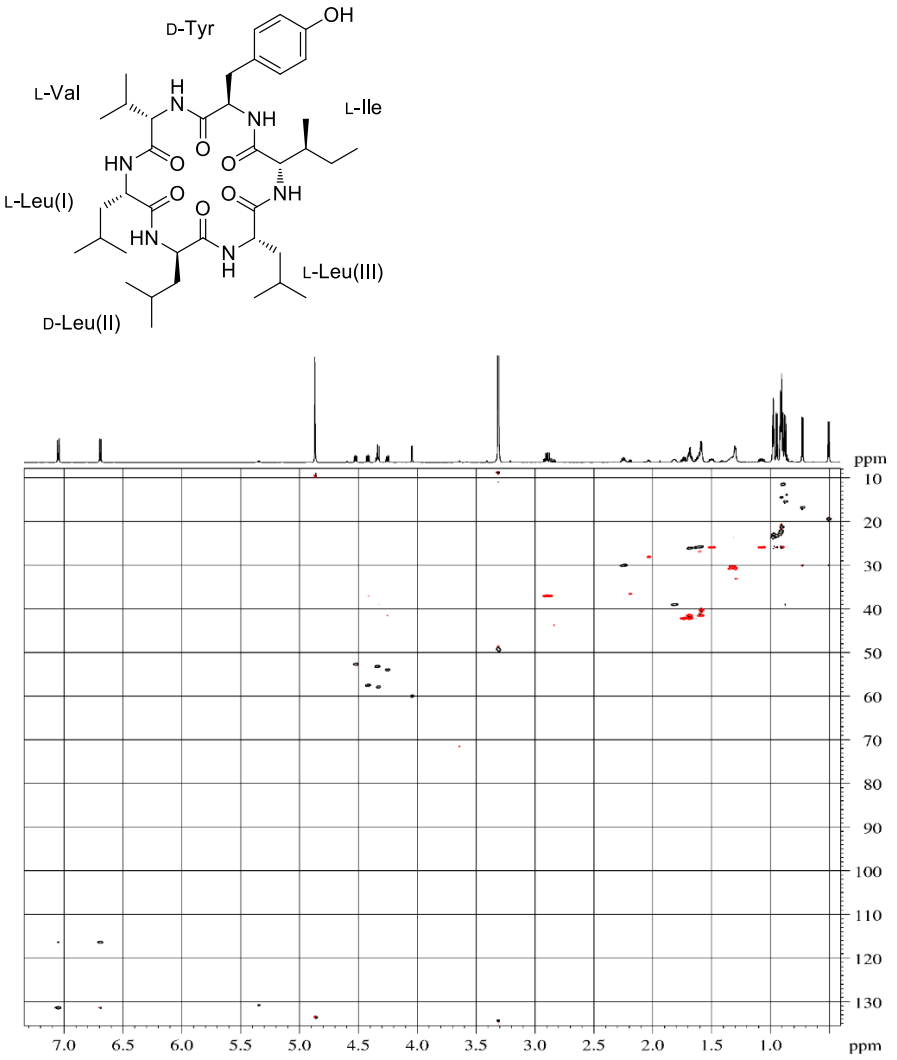


Figure 5.25. HSQC spectrum of Ile-α-*epi*-thermoactinoamide A (**2b**) (700 MHz, CD_3OD).[3]

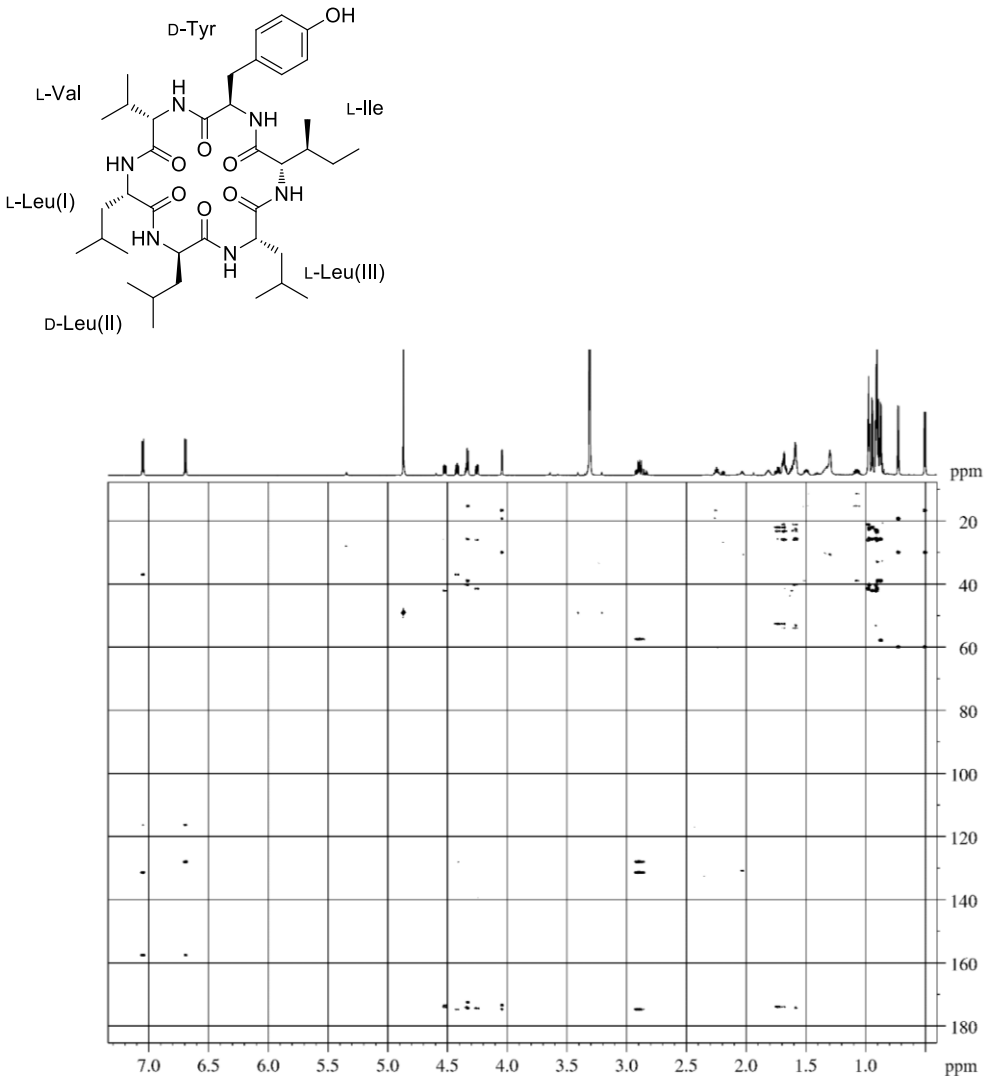


Figure 5.26. HMBC spectrum of Ile-α-*epi*-thermoactinoamide A (**2b**) (700 MHz, CD_3OD).[3]

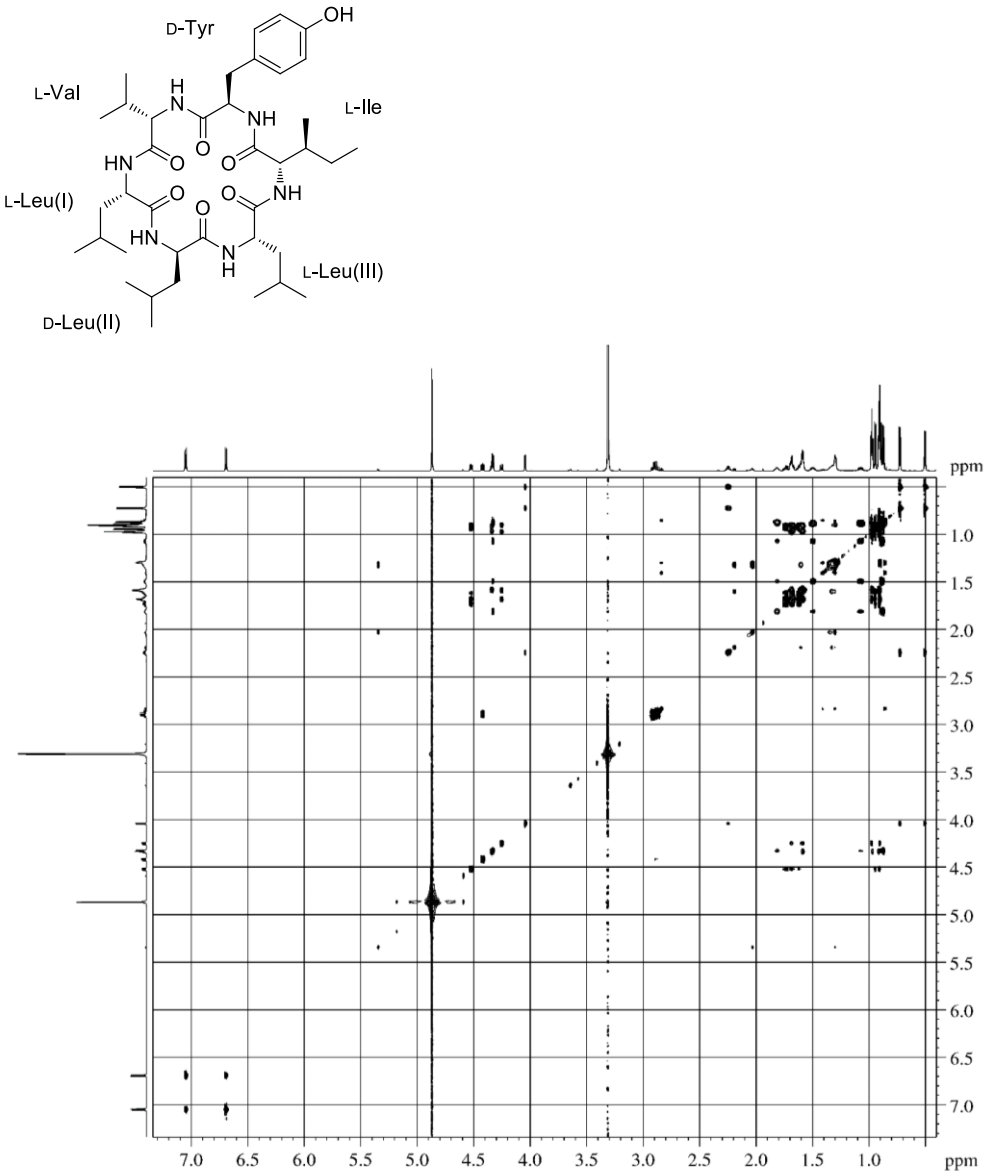


Figure 5.27. TOCSY spectrum of Ile-α-*epi*-thermoactinoamide A (**2b**) (700 MHz, CD_3OD).[3]

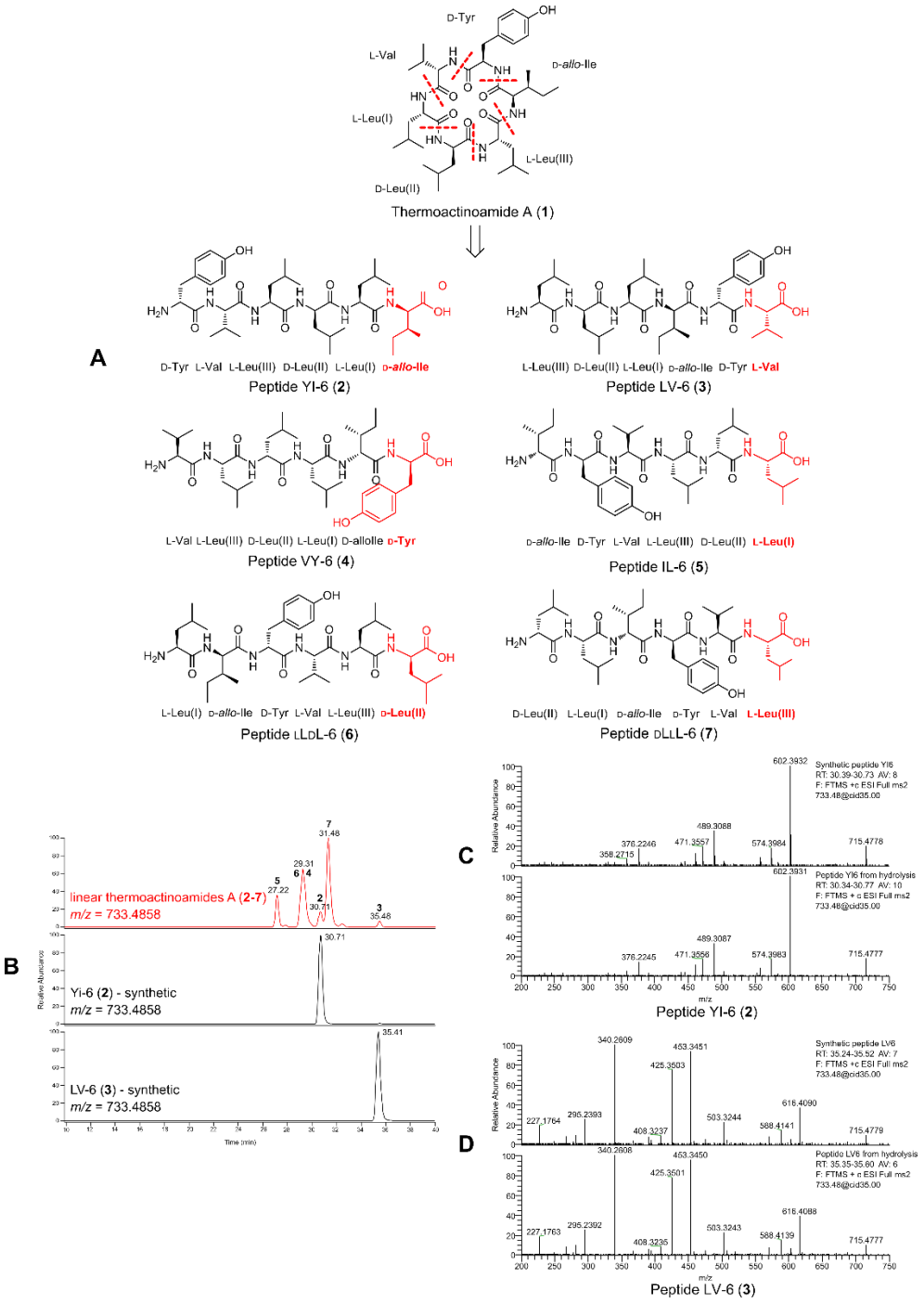


Figure 5.28. (A) The six linear linear peptides yielding thermoactinoamide A (**1a**) by cyclization. (B) Extracted ion chromatogram at *m/z* 733.4858 of partially hydrolized thermoactinoamide A (**1a**) and linear peptides YI-6 (**2**) and LV-6 (**3**). LC conditions: Kinetex PFP column (100 × 4.6 mm) maintained at 25 °C was eluted at 200 µL min^{-1} with 0.1% HCOOH in H_2O and MeOH; gradient: 60% MeOH 5 min, 60% → 100% MeOH over 30 min, 100% MeOH 15 min. (C) Comparison of the MS/MS spectra of peptide YI-6 (**2**) and the corresponding linear hexapeptide from **1a**. (D) Comparison of the MS/MS spectra of peptide LV-6 (**3**) and the corresponding linear hexapeptide from **1a**. 2.6 µm.[3]

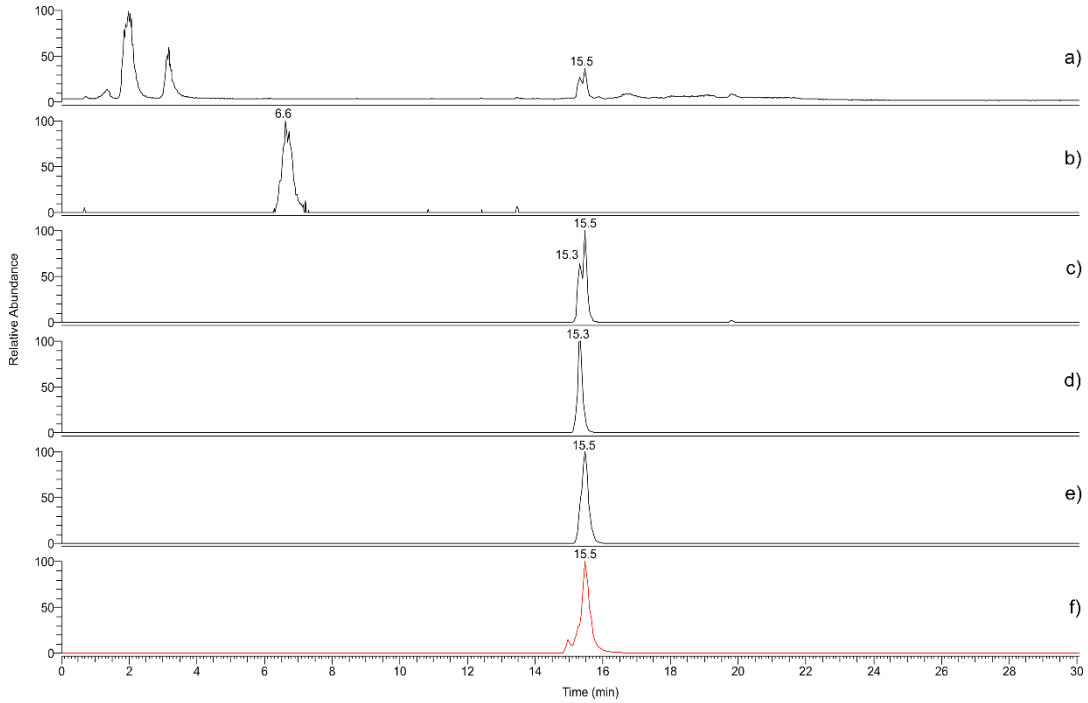


Figure 5.29. LC-MS chromatogram of the crude mixture from cyclization of linear peptide YI-6 (**2**). (a) Crude reaction mixture, TIC. (b) Crude reaction mixture, extracted ion chromatogram at *m*/*z* 733.4858 ($[M+H]^+$ for linear peptide YI-6, **2**). (c) Crude reaction mixture, extracted ion chromatogram at *m*/*z* 715.4753 ($[M+H]^+$ for **1a** and **1b**). (d) Isolated synthetic **1b**. (e) Isolated synthetic **1a**. (f) Natural thermoactinoamide A (**1a**). LC conditions: 5µm Kinetex C18 column (50×2.10 mm) maintained at 25 °C; gradient system [eluent A: 0.1% HCOOH in H_2O; eluent B: CH_3CN; gradient program: 30% B 5 min, 30% → 99% B over 17 min, 99% B 1 min, 99% → 100% B over 1 min; 100% B 10 min; flow rate 200 µL min^{-1}].[3]

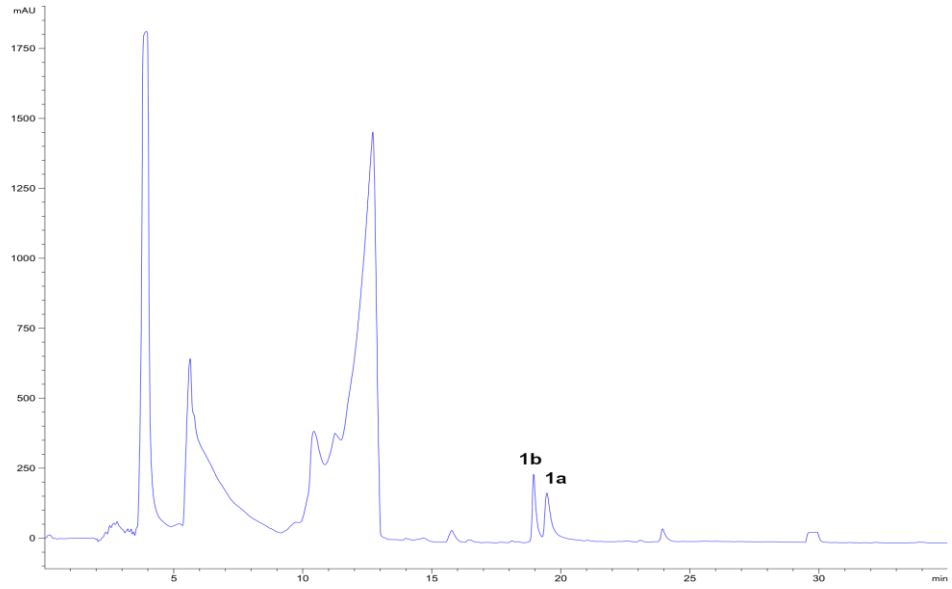


Figure 5.30. HPLC chromatogram of the reaction product from cyclization of linear peptide YI-6 (**2**) HPLC conditions: Luna C18 column (250 × 4.60 mm, 5 µm) [eluent A: 0.1% HCOOH in H_2O; eluent B: CH_3CN; gradient program: 30% B 5 min, 30% → 95% B over 17 min, 95% B 13 min; flow rate 1 mL min^{-1}, UV detection at 280 nm].[3]

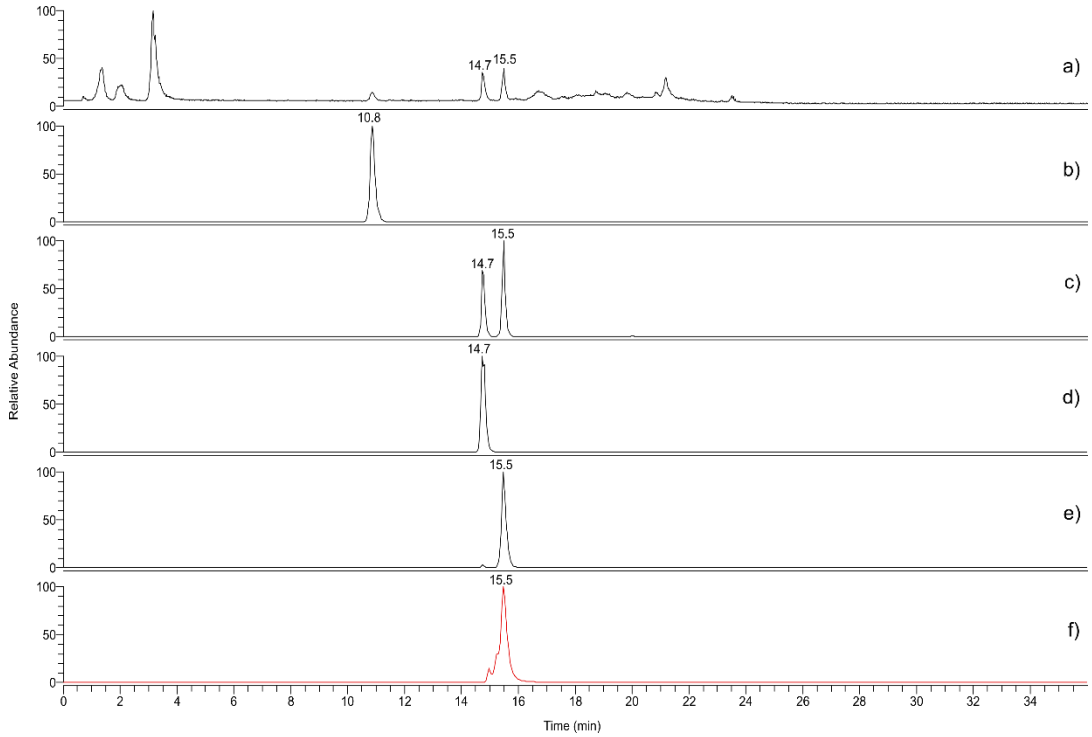


Figure 5.31. LC-MS chromatogram of the crude mixture from cyclization of linear peptide LV-6 (**3**). (a) Crude reaction mixture, TIC. (b) Crude reaction mixture, extracted ion chromatogram at *m*/*z* 733.4858 ($[M+H]^+$ for linear peptide LV-6, **3**). (c) Crude reaction mixture, extracted ion chromatogram at *m*/*z* 715.4753 ($[M+H]^+$ for **1a** and **1c**). (d) Isolated synthetic **1c**. (e) Isolated synthetic **1a**. (f) Natural thermoactinoamide A (**1a**). LC conditions as in Figure S16.[3]

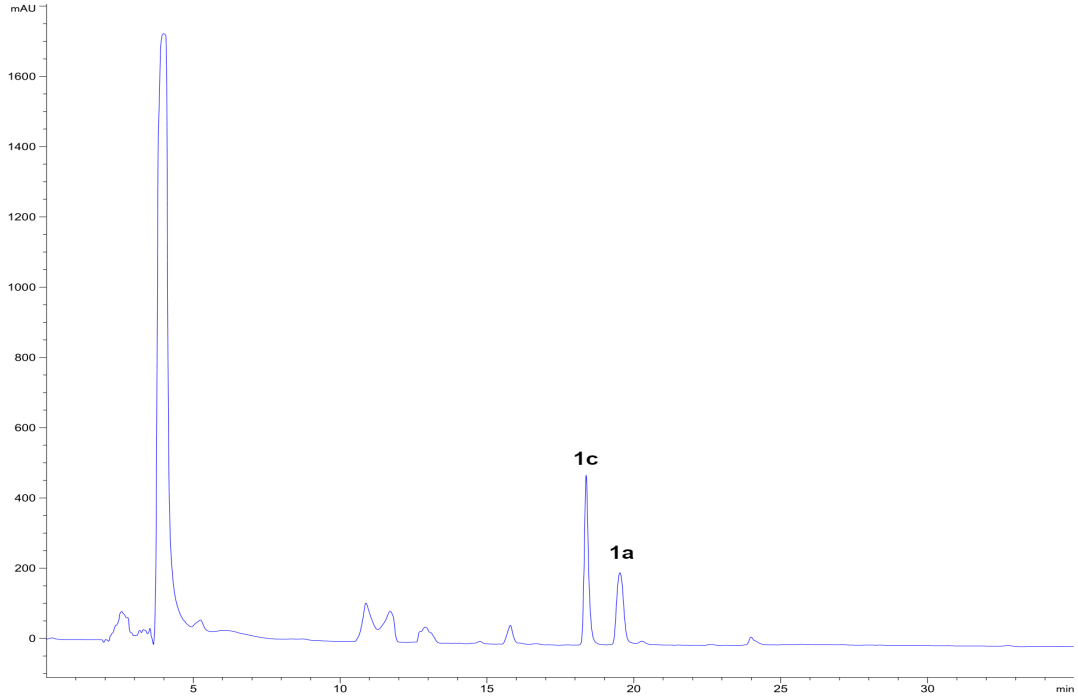


Figure 5.32. HPLC chromatogram (UV, 280 nm) of the reaction product from cyclization of linear peptide LV-6 (**3**). HPLC conditions as in Figure S17.[3]

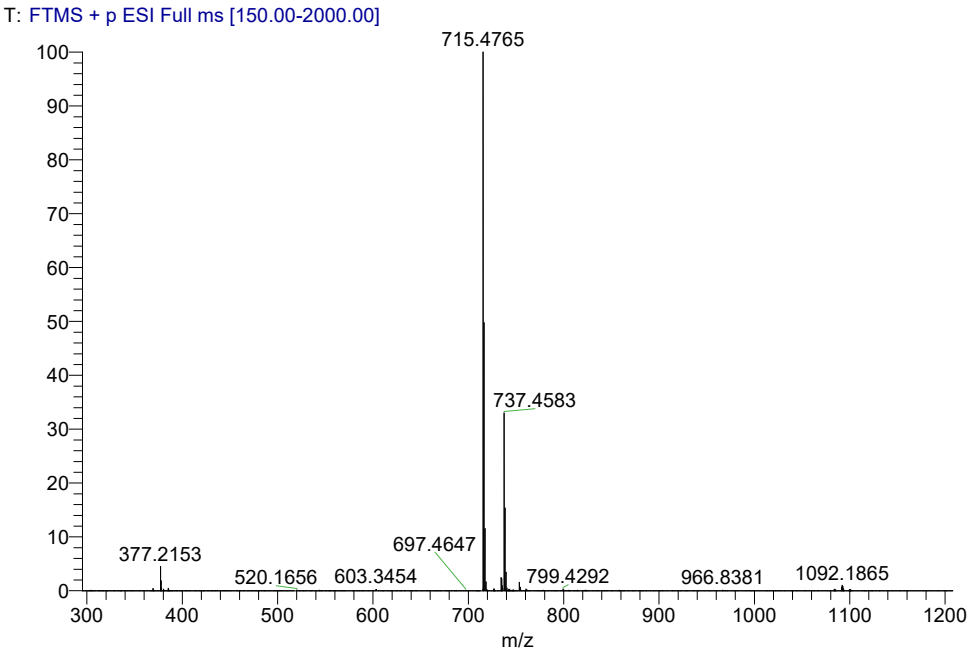


Figure 5.33: High resolution ESI mass spectrum of Val-*epi*-thermoactinoamide A (**1c**).[3]

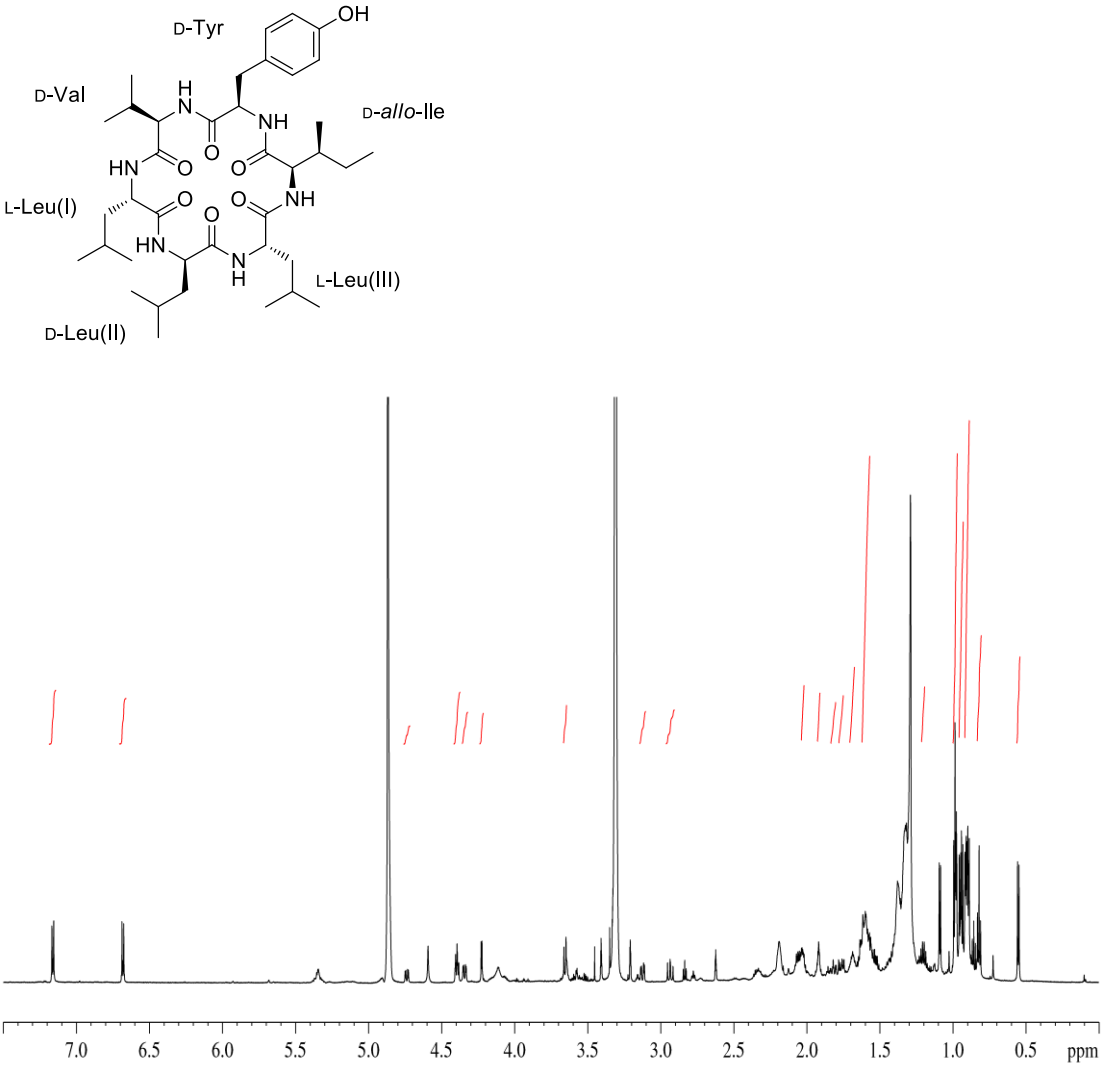


Figure 5.34. ^{1}H-NMR spectrum of Val-α-*epi*-thermoactinoamide A (**1c**) (700 MHz, CD_3OD).[3]

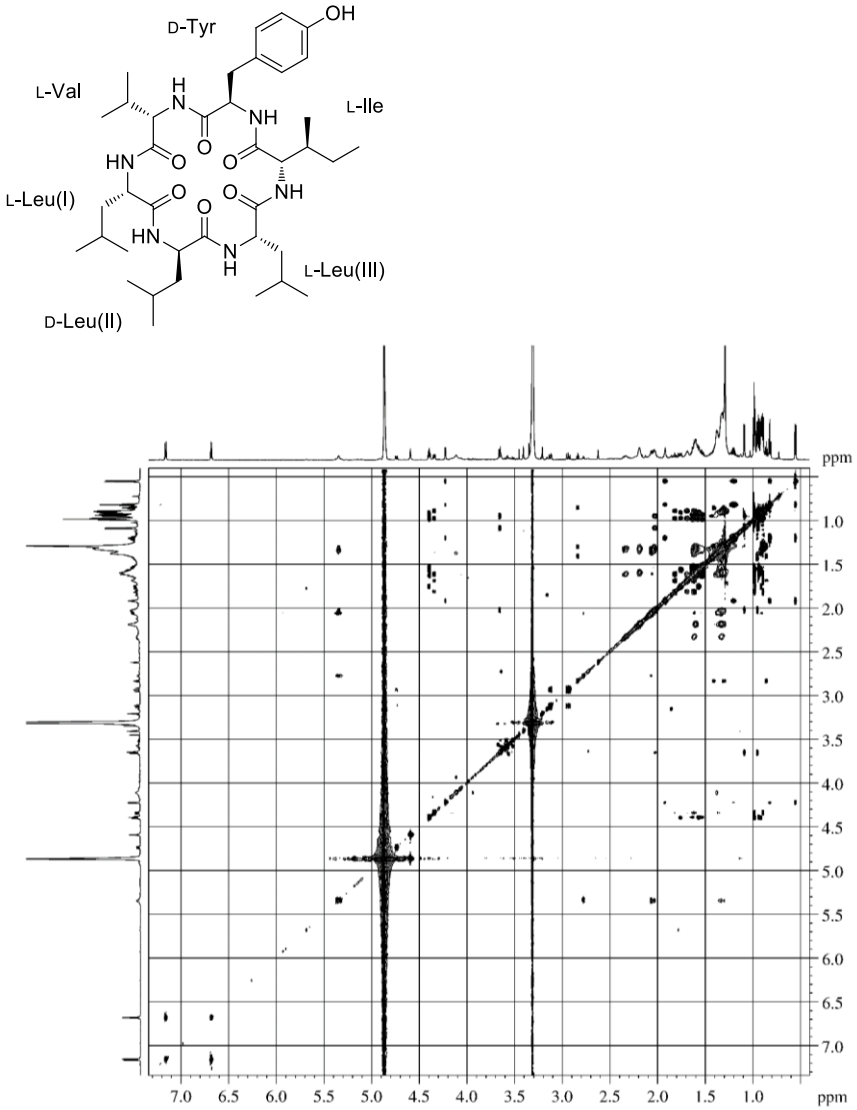


Figure 5.35. COSY spectrum of Val-α-*epi*-thermoactinoamide A (**1c**) (700 MHz, CD_3OD).[3]

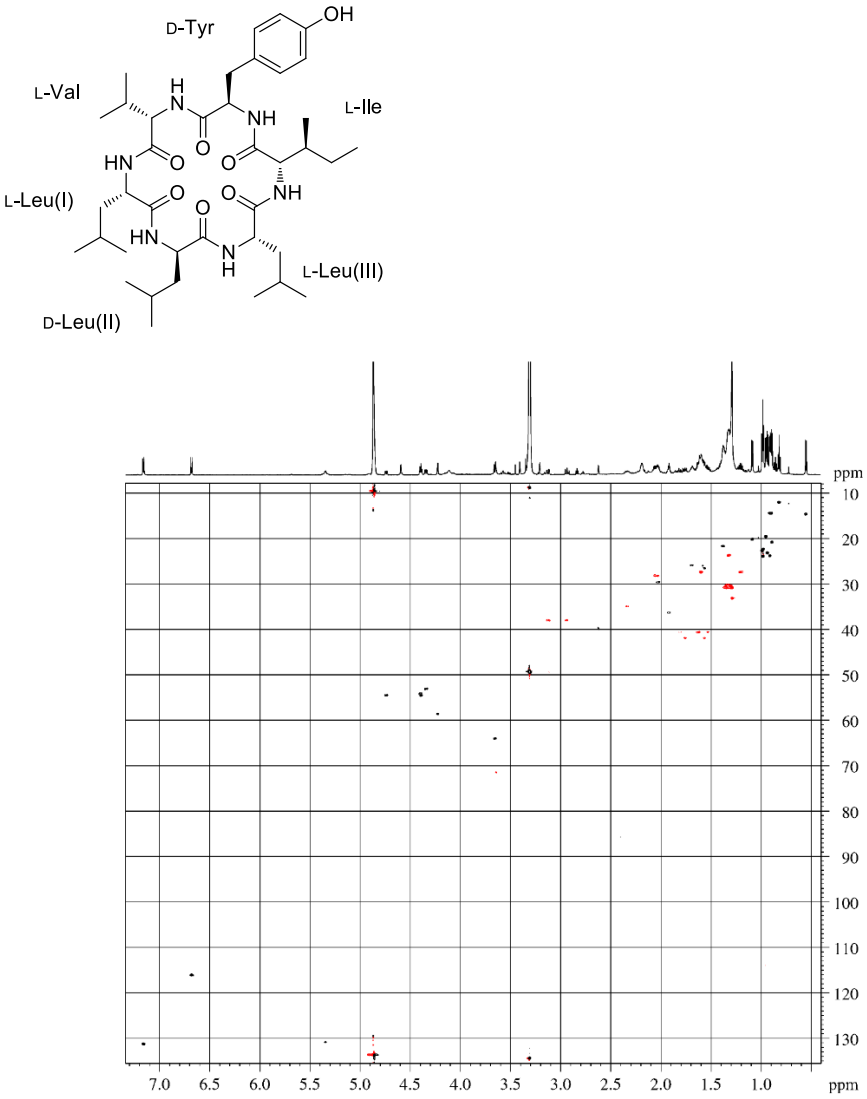


Figure 5.36. HSQC spectrum of Val-α-epi-thermoactinoamide A (**1c**) (700 MHz, CD_3OD).[3]

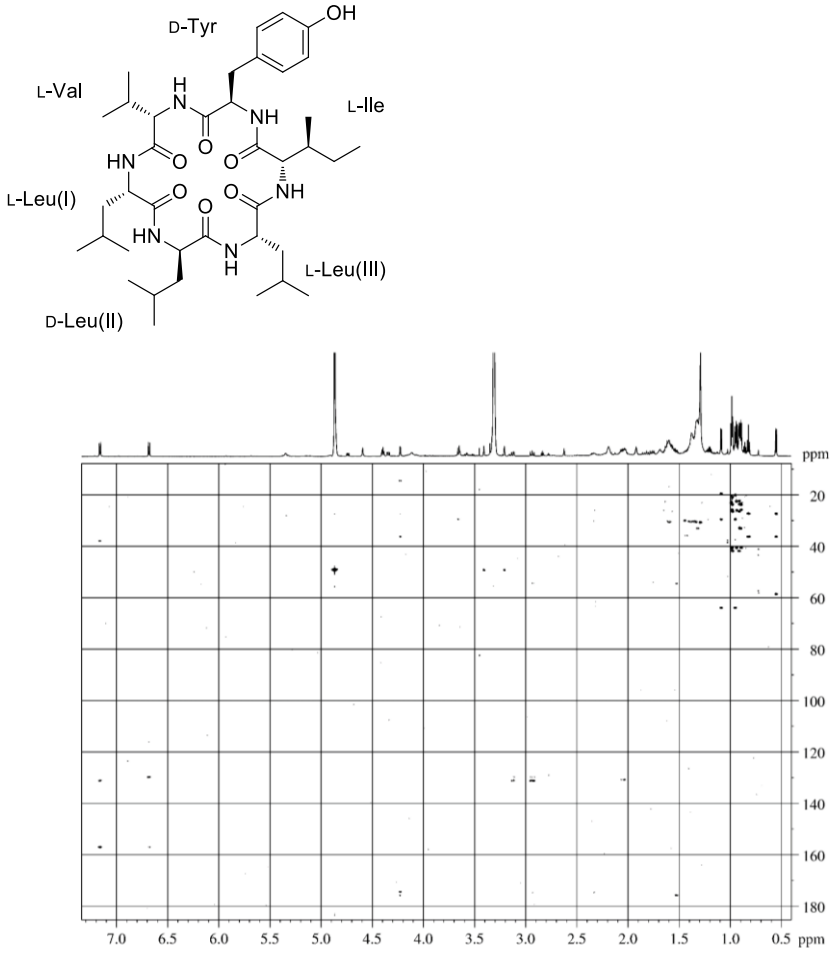


Figure 5.37. HMBC spectrum of Val-α-*epi*-thermoactinoamide A (**1c**) (700 MHz, CD_3OD).[3]

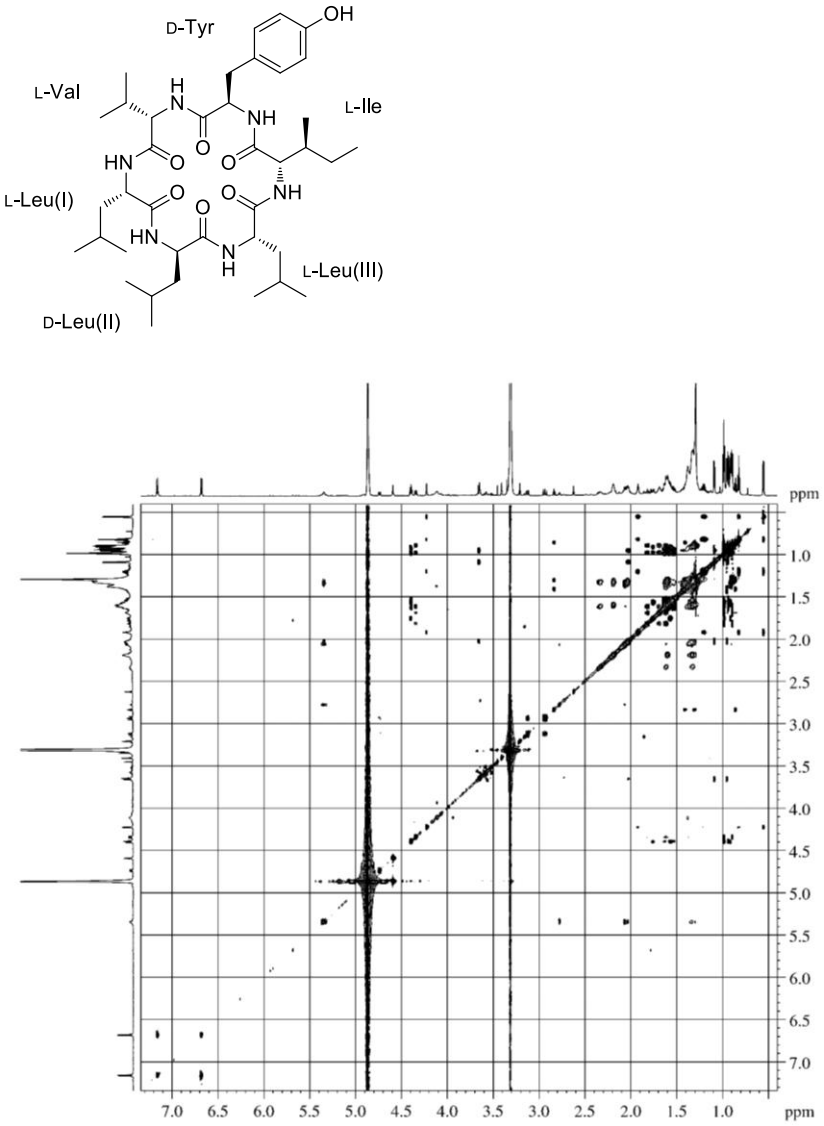


Figure 5.38. TOCSY spectrum of Val-α-*epi*-thermoactinoamide A (**1c**) (700 MHz, CD_3OD).[3]

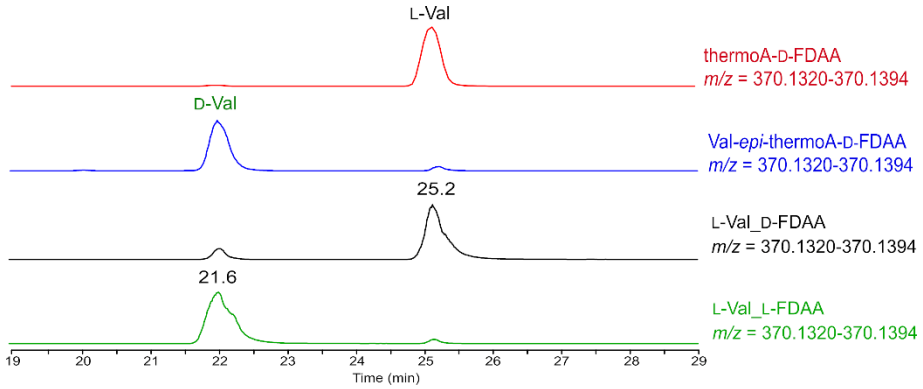


Figure 5.39. LC-MS-enhanced Marfey's analysis of Val-*epi*-thermoactinoamide A (**1c**). Extracted-ion chromatograms at *m/z* 370.1357 of the D-DAA derivative from the hydrolysis of thermoactinoamide A (**1a**) (red trace), Val-*epi*-thermoactinoamide A (**1c**) (blue trace), and of D-and L-DAA derivatives of L-Val (black and green trace, respectively). LC conditions: 2.6 μm Kinetex PFP column (100 × 4.6 mm) maintained at 25 °C was eluted at 200 μL min^{-1} with 0.1% HCOOH in H_2O and MeOH; gradient program: 60% MeOH 5 min, 60% → 100% MeOH over 30 min, 100% MeOH 15 min.[3]

References

1. Teta, R.; Marteinsson, V. T.; Longeon, A.; Klonowski, A. M.; Groben, R.; Bourguet-Kondracki, M.-L.; Costantino, V.; Mangoni, A. Thermoactinoamide A, an Antibiotic Lipophilic Cyclopeptide from the Icelandic Thermophilic Bacterium Thermoactinomyces vulgaris. *J. Nat. Prod.* **2017**, 80, 2530–2535. doi:10.1021/acs.jnatprod.7b00560.

2. Della Sala, G.; Mangoni, A.; Costantino, V.; Teta, R. Identification of the Biosynthetic Gene Cluster of Thermoactinoamides and Discovery of New Congeners by Integrated Genome Mining and MS-Based Molecular Networking. *Front. Chem.* **2020**, 8, 397. doi:10.3389/fchem.2020.00397.

3. Di Matteo, V.; Esposito, G.; Costantino, V.; Della Sala, G.; Teta, R.; Mangoni, A. When Synthesis Gets It Wrong: Unexpected Epimerization Using PyBOP in the Synthesis of the Cyclic Peptide Thermoactinoamide A. *J. Nat. Prod.* **2024**, 87, 948–953. doi:10.1021/acs.jnatprod.3c01229.

4. Duban, A.; Cociancich, R.; Leclère, V. Nonribosomal Peptide Synthesis Definitely Working Out of the Rules. *Microorganisms* **2022**, 10, 577. doi:10.3390/microorganisms10030577.

5. Mordhorst, S.; Ruijne, F.; Vagstad, A. L.; Kuipers, O. P.; Piel, J. Emulating Nonribosomal Peptides with Ribosomal Biosynthetic Strategies. RSC *Chem. Biol.* **2023**, 4, 7–36. doi:10.1039/D2CB00169A.

6. Walsh, C. T.; O'Brien, R. V.; Khosla, C. Nonproteinogenic Amino Acid Building Blocks for Nonribosomal Peptide and Hybrid Polyketide Scaffolds. Angew. *Chem. Int. Ed.* **2013,** 52, 7098–7124. doi:10.1002/anie.201306346.

7. Mordhorst, S.; Ruijne, F.; Vagstad, A. L.; Kuipers, O. P.; Piel, J. Emulating Nonribosomal Peptides with Ribosomal Biosynthetic Strategies. RSC *Chem. Biol.* **2023**, 4, 7–36. doi:10.1039/D2CB00169A.

8. Iacovelli, R.; Bovenberg, R. A. L.; Driessen, A. J. M. Nonribosomal Peptide Synthetases and Their Biotechnological Potential in Penicillium rubens. *J. Ind. Microbiol. Biotechnol.* **2021**, 48, kuab045. doi:10.1093/jimb/kuab045.

9. Figueras, M. J.; Beaz-Hidalgo, R.; Hossain, M. J.; Liles, M. R. Taxonomic Affiliation of New Genomes Should Be Verified Using Average Nucleotide Identity and Multilocus Phylogenetic Analysis. *Genome Announc.* **2014**, 2, e00927-14. doi:10.1128/genomeA.00927-14.

10. Ziemert, N.; Podell, S.; Penn, K.; Badger, J. H.; Allen, E.; Jensen, P. R. The Natural Product Domain Seeker NaPDoS: A Phylogeny-Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. *PLoS One* **2012**, 7, e34064. doi:10.1371/journal.pone.0034064.

11. Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; et al. Feature-Based Molecular Networking in the GNPS Analysis Environment. *Nat. Methods* **2020**, 17, 905–908. doi:10.1038/s41592-020-0933-6.

12. Sun, Y.; Li, H.; Zhou, R.; Wang, H.; Feng, P.; Xu, S. Total Synthesis of Thermoactinoamide A and Its Analogue. *Heterocycles* **2019**, 98, 789–797. doi:10.3987/COM-19-14076.

13. Chandrudu, S.; Simerska, P.; Toth, I. Chemical Methods for Peptide and Protein Production. *Molecules* **2013**, 18, 4373–4388. doi:10.3390/molecules18044373.

14. Hayes, H. C.; Luk, L. Y. P.; Tsai, Y.-H. Approaches for Peptide and Protein Cyclisation. *Org. Biomol. Chem.* **2021**, 19, 3983–4001. doi:10.1039/D1OB00411E.

15. Chen, I.-M. A.; Chu, K.; Palaniappan, K.; Pillay, M.; Ratner, A.; Huang, J.; et al. IMG/M v.5.0: An Integrated Data Management and Comparative Analysis System for Microbial Genomes and Microbiomes. *Nucleic Acids Res.* **2019**, 47, D666–D677. doi:10.1093/nar/gky901.

16. Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. NRPSpredictor2-A Web Server for Predicting NRPS Adenylation Domain Specificity. *Nucleic Acids Res.* **2011**, 39, W362–W367. doi:10.1093/nar/gkr323.

17. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, 215, 403–410. doi:10.1016/S0022-2836(05)80360-2.

18. Costantino, V.; Della Sala, G.; Mangoni, A.; Perinu, C.; Teta, R. Blurring the Boundary between Bio- and Geohopanoids: Plakohopanoid, a C32 Biohopanoid Ester from Plakortis cf. lita. *Eur. J. Org. Chem.* **2012**, 5171–5176. doi:10.1002/ejoc.201200676.

19. Caso, A.; Esposito, G.; Della Sala, G.; Pawlik, J. R.; Teta, R.; Mangoni, A.; Costantino, V. Fast Detection of Two Smenamide Family Members Using Molecular Networking. *Mar. Drugs* **2019**, 17, 618. doi:10.3390/md17110618.

PART 2

Exploration of cyanobacterial natural products

Chapter 6

Cyanobacteria

Cyanobacteria, often mistakenly called “blue-green algae,” are photosynthetic single-celled prokaryotes whose evolutionary history dates back at least 3.5 billion years, as evidenced by microfossils and laminated stromatolites. They appeared on the primordial Earth, when there was still no free molecular oxygen. By evolving oxygenic photosynthesis, which uses water as an electron donor, releasing O_2 and producing ATP and NADPH, they gradually transformed the planet's atmosphere and oceans. This planetary-scale change, culminating in the Great Oxidation Event, allowed the emergence of aerobic metabolisms, the formation of the protective ozone layer, and ultimately the diversification of complex life. In parallel, many cyanobacteria also developed the ability to fix atmospheric nitrogen (N_2) into ammonia (NH_3/NH_4^+), a bioavailable form that fertilizes ecosystems and fuels primary productivity in waters and soils where combined nitrogen is scarce. Together, these two biogeochemical feats, oxygen production and nitrogen fixation, make cyanobacteria the fundamental engineers of the Earth's biosphere.

Cyanobacteria occupy an unparalleled variety of environments.[1] They thrive wherever there is light and liquid water: from lakes, reservoirs, and rivers to shallow coastal waters, wetlands, and brackish estuaries. They cover hot springs and geysers, pigmenting silica terraces with bright bands; they inhabit hypersaline lagoons and alkaline lakes; they withstand polar deserts and

seasonal snowfields; they form endolithic communities within translucent rocks and living crusts in arid landscapes. Their global distribution reflects remarkable physiological plasticity: the ability to harvest dim or spectrally filtered light, regulate buoyancy and position within the water column, move or aggregate, store nutrients, and change metabolism in response to changing environmental conditions. Not surprisingly, cyanobacteria are also frequent symbiotic partners, contributing fixed carbon and nitrogen to various hosts. They serve as photobionts in lichens; populate specialized roots of cycads; live within the fronds of the aquatic fern *Azolla*; and associate with marine sponges and other invertebrates. These relationships extend the influence of cyanobacteria into nutrient-poor niches and stabilize ecological networks in both terrestrial and aquatic systems.

Although cyanobacteria are natural constituents of planktonic and benthic communities, human activities have increasingly altered the balance in their favour. Excessive inputs of nitrogen and phosphorus from wastewater, agricultural runoff, and urban fertilizers, combined with rising temperatures, increased water column stratification, and reduced mixing, favour floating and fast-growing taxa. The result is more frequent and persistent blooms of cyanobacteria in freshwater and brackish waters around the world.

6.1 Cellular characteristics and properties

Cyanobacteria are Gram-negative eubacteria. Their cell envelope consists of an inner cytoplasmic membrane, a peptidoglycan layer, and an outer membrane typical of Gram-negative bacteria.[2] Inside the cytoplasm, an intricate system of photosynthetic thylakoid membranes houses the light-

harvesting and photochemical mechanisms: photosystems I and II, the cytochrome b_6f complex, and ATP synthase. The light-harvesting antennae are dominated by phycobilisomes: supramolecular arrays of phycobiliproteins, particularly phycocyanin (blue), allophycocyanin, and phycoerythrin (red-orange). These pigments expand the usable spectrum to the green, yellow, and orange wavelengths that chlorophyl absorbs poorly, allowing efficient energy capture in low light conditions or with spectral shifts.

Unlike plant chloroplasts, cyanobacterial thylakoids are not stacked in grana, but form parallel lamellae distributed throughout the cytoplasm. Morphologically, cyanobacteria are extraordinarily diverse. Picocyanobacteria such as *Prochlorococcus* and *Synechococcus* measure less than a micrometre and dominate vast areas of the open ocean.

At the opposite extreme, filamentous taxa produce long trichomes, some enclosed in mucilaginous sheaths, which can reach hundreds of micrometres in length. Many species form macroscopic colonies; for example, *Microcystis* aggregates into floating clusters bound by mucilage that can coalesce into surface foam. Some groups, such as members of the Stigonematales, exhibit true branching. This spectrum of forms reflects not only differences in cell size, but also distinct strategies for nutrient acquisition, light harvesting, and resistance to grazing or environmental stress.

Despite their prokaryotic organization, lacking a nucleus and membrane-bound organelles, cyanobacteria possess specialized intracellular compartments and inclusions. The nucleoid region contains the genomic DNA and is often located in a central position. Carboxysomes, protein-bound

bacterial microcompartments, concentrate CO_2 around the enzyme RuBisCO and improve carbon fixation efficiency, especially under conditions of low CO_2 concentration or high O_2 concentration.[3] Cells also store resources in granules of glycogen, polyphosphate, and cyanophycin (a nitrogen-rich polymer composed of aspartate and arginine), which act as buffers against fluctuations in nutrient supply. Many planktonic species regulate their vertical position using gas vesicles, rigid structures with protein walls filled with gas that give them buoyancy, allowing them to exploit light near the surface during periods of stratification and retreat downward when radiation is excessive.[4]

These gas vesicles are sometimes mistakenly described as “vacuoles”; unlike eukaryotic vacuoles, they contain gas rather than aqueous solution. Cyanobacteria exhibit both behavioural and evolutionary versatility. They can glide on surfaces and exhibit phototaxis, optimizing their exposure to favourable light intensities and spectral qualities. Several filamentous lineages undergo cell differentiation, producing heterocysts and akinetes.

Heterocysts are specialized cells with thick walls that create a micro-oxygenated environment for nitrogenase, the oxygen-sensitive enzyme that reduces N_2 to ammonia. By spatially separating nitrogen fixation in heterocysts while nearby vegetative cells perform oxygenic photosynthesis, genera such as *Dolichospermum*/*Anabaena* and *Nostoc* maintain growth in nitrogen-poor waters.

Akinetes function as resting spores, improving survival to cold, desiccation, and nutrient deprivation and contributing to the seasonal recurrence of blooms from benthic seed banks.

Pigment composition and regulatory circuits underlie the success of cyanobacteria in climates with different light conditions. Through chromatic acclimation, some species remodel their phycobiliprotein content to better adapt to environmental spectra, for example by increasing phycoerythrin under green light or enhancing phycocyanin under red light, maintaining efficient energy transfer to chlorophyll in different habitats, from turbid lakes to shaded crevices and cave walls.

Complementing these strategies, cyanobacteria use high-affinity absorption systems for scarce nutrients and secrete siderophores to chelate and acquire iron, an essential element for photosynthetic electron transport but often limiting in both fresh and marine waters. Together, these characteristics (versatility in light harvesting, control of motility and buoyancy, aggregation into colonies or filaments, nitrogen fixation, toxin biosynthesis, and iron removal) explain the ecological breadth and competitive advantage of cyanobacteria across the planet.

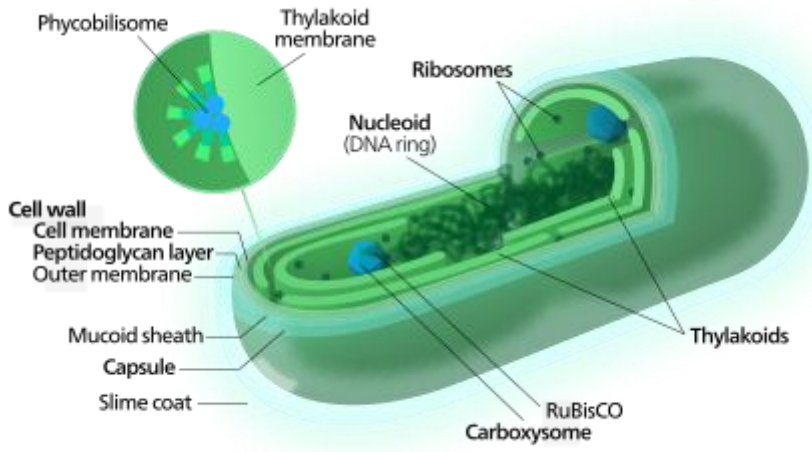


Figure 6.1. Cyanobacterial cell

6.2 Ecology, habitat, symbiosis, and anthropogenic impacts

The ecological footprint of cyanobacteria covers the entire gradient from oligotrophic to hyper-eutrophic waters and from extreme cold to hot temperatures. In nutrient-poor environments, picocyanobacteria dominate due to their small size and high surface area to volume ratio; in nutrient-rich, stratified, and warm conditions, colonies and filaments that form blooms often prevail. Benthic mats stabilize sediments and channel biogeochemical fluxes; microbialites formed by cyanobacterial growth and mineral precipitation echo ancient stromatolitic architectures. Symbiotic associations extend the reach of cyanobacteria into otherwise prohibitive niches: lichens combine host fungi with cyanobacterial photobionts; freshwater ferns such as Azolla host endosymbiotic cyanobacteria that provide fixed nitrogen; cycads host them in specialized coralloid roots; and marine sponges integrate them into complex holobionts that influence reef productivity.[4] Through carbon fixation, oxygen evolution, nitrogen supply, and iron cycling, cyanobacteria are fundamental to the metabolism of many ecosystems, past and present.

Over the last century, accelerating nutrient loading, altered hydrology, and climate warming have reshaped aquatic ecosystems. Cyanobacterial blooms now occur earlier in the season, persist longer, and extend into previously unaffected regions. In addition to producing toxins, bloom biomass can deplete oxygen during decomposition, causing fish kills; it can clog water intakes and filters; and it can impair recreational activities and aesthetics. Effective mitigation depends on reducing external nutrient inputs (both

phosphorus and nitrogen), remediating nutrient reserves in sediments where possible, restoring mixing or water exchange where physically possible, and using monitoring tools, from in situ fluorometers to satellite remote sensing, to track the onset, spread, and toxicity of blooms. Since not all cyanobacteria are toxic and toxin production varies depending on strains and environmental contexts, risk assessment requires both taxonomic surveillance and direct measurement of toxins. With proper context, cyanobacteria are recognised as ancient and vital to a habitable planet, while their opportunistic bloom dynamics necessitate careful management of today's water resources.

6.3 The multifaceted nature of cyanobacteria

6.3.1 Cyanobacteria as sources of bioactive natural products

Cyanobacteria are increasingly recognised as prolific sources of diverse natural products with valuable bioactivities. These photosynthetic microorganisms produce a wide spectrum of secondary metabolites, biosynthesised through pathways such as non-ribosomal peptide synthetases and polyketide synthetases. Consequently, cyanobacterial compounds exhibit broad biological activities: many demonstrate anticancer, antibacterial, antifungal, antiviral, and even protease inhibitory properties.[6] Notable examples include unique polysulphated polysaccharides with antiviral effects, various polyketides and peptides with cytotoxic or antimicrobial activity, and other specialised metabolites.[7] This high chemical diversity makes cyanobacteria a promising reservoir of new pharmaceutical compounds, with the potential to yield novel drug candidates.

Another significant feature of cyanobacteria is their ability to synthesise antioxidant molecules. They produce compounds such as polyunsaturated fatty acids (PUFAs), β-carotene and other carotenoids, as well as various phenolic pigments with antioxidant properties.[8] In addition, cyanobacteria generate photoprotective substances such as mycosporin-like amino acids (MAAs) and scytonemin (Figure 6.2), as well as sterols with antimicrobial effects.

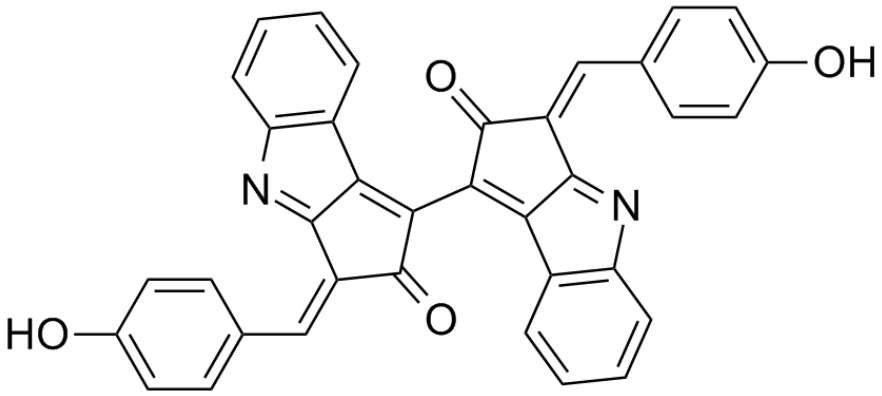


Figure 6.2. Structure of scytonemin

The growing interest in these natural antioxidants stems from concerns about synthetic preservatives and artificial antioxidants, some of which have toxicity or undesirable side effects. Therefore, there is a concerted effort to explore safer and more sustainable alternatives: for example, cyanobacterial carotenoids and polyphenols are being studied as health-promoting nutraceuticals. By exploiting the intrinsic diversity of cyanobacterial metabolites in this way, antioxidant compounds could be obtained that improve health and reduce dependence on synthetic additives.

It is important to emphasise that cyanobacteria offer practical advantages as a bioresource. In fact, these microorganisms can be cultivated relatively easily in laboratories or photobioreactors, providing a renewable and controlled source of biomass. This solves the “supply problem” that often arises in the discovery of natural products: harvesting large quantities of rare marine organisms is not only unsustainable, but often unfeasible. Random harvesting of wild marine organisms can have serious ecological repercussions and usually yields insufficient quantities of the desired compounds.

In contrast, cyanobacterial strains can be cultivated on a large scale under defined conditions, in line with the principle of “do no significant harm” to ecosystems, while ensuring a consistent production of bioactive metabolites. In essence, cyanobacteria serve as environmentally friendly production platforms for valuable natural products. This sustainable supply is highly consistent with modern environmental guidelines (e.g., EU sustainability criteria) and circumvents the bottleneck of limited natural supply. Given their metabolic versatility and cultivability, cyanobacteria are excellent candidates for bioprospecting efforts aimed at discovering new drugs and nutraceuticals. Indeed, the remarkable chemical repertoire of these organisms is increasingly considered a prolific source of new compounds for drug development.

6.3.2 Cyanobacteria as toxin producers and environmental bioindicators

On the other hand, cyanobacteria are known for their role in producing toxins and degrading water quality under certain conditions. When environmental factors become favourable, such as in the presence of excess nutrients (high

phosphorus and nitrogen loads), abundant sunlight, warm temperatures, and stratified or stagnant waters, cyanobacteria can proliferate explosively.

A major concern is that some bloom-forming cyanobacteria produce potent cyanotoxins. Cyanotoxins are toxic secondary metabolites encompassing a wide range of chemical classes and biological targets. Broadly, these toxins are often categorized by their primary toxic effects into hepatotoxins, neurotoxins, and endotoxins (lipopolysaccharides).[9] The most well-known hepatotoxins are the microcystins, cyclic heptapeptides produced by genera such as *Microcystis*, *Planktothrix*, and others. Microcystins (along with the related nodularins from *Nodularia*) specifically damage the liver, and over 100 microcystin variants have been identified. Notably, microcystin-LR (Figure 6.3) is one of the most toxic and extensively studied congeners.

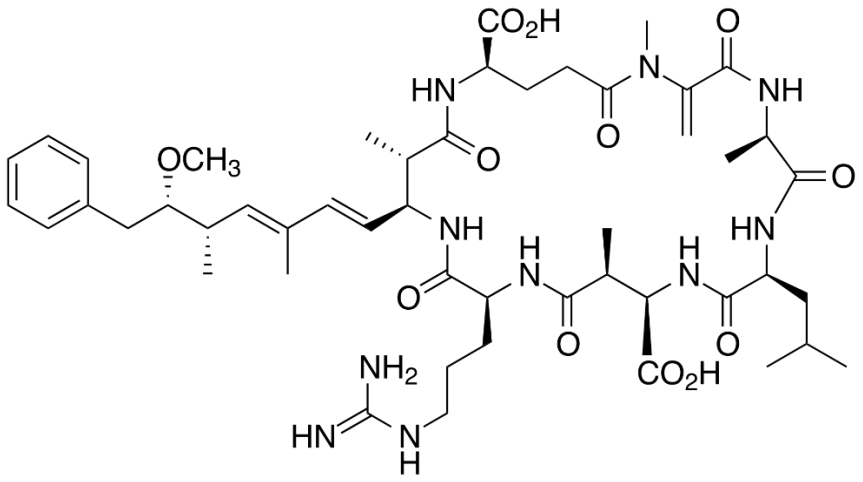


Figure 6.3. Structure of microcystin-LR

Another hepatotoxic compound is cylindrospermopsin, a tricyclic alkaloid initially isolated from *Cylindrospermopsis*; it too primarily targets the liver.

In contrast, neurotoxic cyanotoxins include anatoxins and saxitoxins. Anatoxin-a (Figure 6.4), for example, is a deadly neurotoxin produced by certain strains of *Dolichospermum* (formerly *Anabaena*) and other genera. It acts as a potent agonist at nicotinic acetylcholine receptors, effectively mimicking acetylcholine at neuromuscular junctions and causing a depolarizing neuromuscular blockade. This leads to rapid paralysis of respiratory muscles- hence its nickname "Very Fast Death Factor."

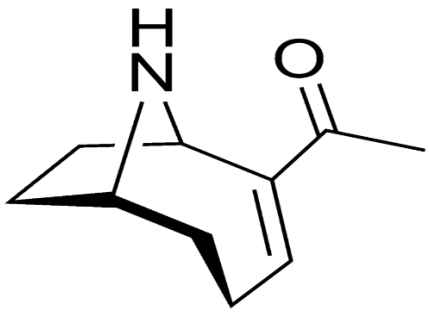


Figure 6.4. Structure of anatoxin-a

Saxitoxins (STXs), on the other hand, are a group of alkaloid neurotoxins best known for causing paralytic shellfish poisoning in marine settings (produced by dinoflagellates), but some freshwater cyanobacteria like *Aphanizomenon* can produce them as well. Saxitoxin and its analogues block voltage-gated sodium channels in nerve cells, preventing normal nerve impulse transmission; this blockade results in paralysis and can be lethal due to respiratory failure.

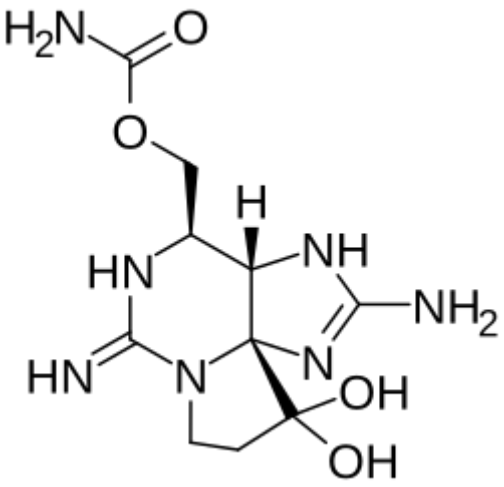


Figure 6.5. Structure of saxitoxin

More recently, additional cyanotoxin classes have been recognized, underscoring the chemical variety of cyanobacterial poisons. For instance, anabaenopeptins are a family of cyclic hexapeptides (originally discovered in *Anabaena*) that contain an unusual ureido-linked amino acid in their ring structure (Chapter 8). These compounds were found to inhibit certain proteases and phosphatases, suggesting a mechanism for their toxicity and a role in the producing organisms' physiology.

Likewise, other peptide toxins like cyanopeptolins, oscillatoxins, and lyngbyatoxins add to the spectrum of bioactive molecules cyanobacteria can generate. Each class of cyanotoxin comprises numerous structural variants (congeners). For example, dozens of microcystin analogues and many anabaenopeptin variants are known, each differing slightly in amino acid composition or functional groups. This diversity of toxin structures presents a challenge for monitoring and risk assessment, as analytical methods must detect a range of toxins with potentially different potencies.

Crucially, a single cyanobacterial bloom can produce multiple toxins simultaneously, especially if it contains mixed species or strains. In some cases, one cyanobacterial genus can synthesize several toxin types during the same bloom. For instance, a *Dolichospermum* bloom might release both microcystin (a hepatotoxin) and anatoxin-a (a neurotoxin), or a mixed bloom of *Microcystis* and *Anabaena* might collectively contribute a cocktail of microcystins, saxitoxins, and anabaenopeptins. The co-occurrence of multiple toxins amplifies the hazard of HAB events, different toxins can affect different organ systems or species, and their combined presence complicates treatment of contaminated water. This makes environmental monitoring of cyanobacterial blooms and their toxins an essential public health task.[10]

Given the above, the detection of cyanobacteria in water is a serious environmental warning sign. Their sudden proliferation typically signals eutrophication (excessive nutrient pollution) and ecological imbalance. In fact, cyanobacteria are often used as bioindicators of water quality, since their abundance is generally triggered by high nutrient loads from anthropogenic sources. Put simply, a bloom outbreak is telling us that a water body is under stress from nutrient enrichment and favourable conditions for algae. Moreover, cyanobacteria are notable indicators because of the very toxins and nuisance compounds they produce, these substances pose significant threats to water supplies and aquatic food webs.[11] For example, the presence of toxin-producing cyanobacteria in a reservoir may foreshadow the contamination of drinking water with microcystin or anatoxin, prompting water managers to take pre-emptive action. Even before toxins are measured, a visible bloom

alerts officials to potential health risks for humans and animals (through ingestion, skin contact, or inhalation of aerosols). It also indirectly flags issues like agricultural runoff or wastewater discharge that may be polluting the system.

In summary, while cyanobacteria can be beneficial bioresources in biotechnology, they also represent biohazards in polluted waters. Their dual nature, on one side, source of novel bioactive compounds; on the other, producers of deleterious toxins, makes them a subject of intense scientific interest. In environmental monitoring, cyanobacteria serve as sentinels: their presence (or bloom proliferation) is a biological indicator of water health status, signalling when an ecosystem's balance is upset. By tracking cyanobacterial populations and the toxins they generate, researchers and public health officials can gauge environmental quality and mitigate risks.[12] Thus, the many "faces" of cyanobacteria range from valuable bioresources for sustainable bioproducts to critical bioindicators (and agents) of environmental change and pollution.

During my PhD, my research on cyanobacteria linked toxicology, bloom monitoring, and cyanotoxin profiling with a focus on their bioaccumulation and biomagnification in complex coastal systems, and natural products, particularly the discovery and structural/functional study of cyanobacterial siderophores such as the photolabile cyanochelin B containing β-hydroxyaspartate, in collaboration with Dr. Pavel Hrouzek research group at the Algatech Center, Institute of Microbiology, Czech Academy of Sciences (CAS), in Třeboň, Czech Republic.

These studies led to the publication of two research papers:

1. Esposito G, De Rosa T, Di Matteo V, Ciccarelli C, Ajaoud M, Teta R, Lega M, Costantino V. Bio-tracking, bio-monitoring and bio-magnification interdisciplinary studies to assess cyanobacterial harmful algal blooms (cyanoHABs)' impact in complex coastal systems. *Sci Total Environ*. **2025**; 978:179480.

2. Falcao BP, Di Matteo V, Hrouzek P, Štenclová L, Urajová P, Mareš J, et al. Cyanochelin B: a cyanobacterium-produced siderophore with photolytic properties that negate iron monopolization in UV light. *Appl Environ Microbiol.* **2025**; 0:e02566-24.

References

1. Álvarez, C.; Jiménez-Ríos, L.; Iniesta-Pallarés, M.; Jurado-Flores, A.; Molina-Heredia, F. P.; Ng, C. K. Y.; Mariscal, V. Symbiosis between Cyanobacteria and Plants: From Molecular Studies to Agronomic Applications. *J. Exp. Bot.* **2023**, 74, 6145–6157. doi:10.1093/jxb/erad261.

2. Cassier-Chauvat, C.; Blanc-Garin, V.; Chauvat, F. Genetic, Genomics, and Responses to Stresses in Cyanobacteria: Biotechnological Implications. *Genes* **2021**, 12, 500. doi:10.3390/genes12040500.

3. Rae, B. D.; Long, B. M.; Badger, M. R.; Price, G. D. Functions, Compositions, and Evolution of the Two Types of Carboxysomes: Polyhedral Microcompartments That Facilitate CO_2 Fixation in Cyanobacteria and Some Proteobacteria. *Microbiol. Mol. Biol. Rev*. **2013**, 77, 357–379. doi:10.1128/MMBR.00061-12.

4. Tashiro, Y.; Monson, R. E.; Ramsay, J. P.; Salmond, G. P. C. Molecular Genetic and Physical Analysis of Gas Vesicles in Buoyant Enterobacteria. *Environ. Microbiol.* **2016**, 18, 1264–1276. doi:10.1111/1462-2920.13203.

5. Chang, A. C. G.; Chen, T.; Li, N.; Duan, J. Perspectives on Endosymbiosis in Coralloid Roots: Association of Cycads and Cyanobacteria. *Front. Microbiol.* **2019**, 10, 1888. doi:10.3389/fmicb.2019.01888.

6. Singh, R. K.; Tiwari, S. P.; Rai, A. K.; Mohapatra, T. M. Cyanobacteria: An Emerging Source for Drug Discovery. *J. Antibiot.* **2011**, *64*, 401–412. doi:10.1038/ja.2011.21.

7. Cock, I. E.; Cheesman, M. J. A Review of the Antimicrobial Properties of Cyanobacterial Natural Products. *Molecules* **2023**, *28*, 7127. doi:10.3390/molecules28207127.

8. Guerreiro, A.; Andrade, M. A.; Menezes, C.; Vilarinho, F.; Dias, E. Antioxidant and Cytoprotective Properties of Cyanobacteria: Potential for Biotechnological Applications. *Toxins* **2020**, *12*, 548. doi:10.3390/toxins12090548.

9. Falfushynska, H.; Kasianchuk, N.; Siemens, E.; Henao, E.; Rzymski, P. A Review of Common Cyanotoxins and Their Effects on Fish. *Toxics* **2023**, *11*, 118. doi:10.3390/toxics11020118.

10. Song, L.; Jia, Y.; Qin, B.; Li, R.; Carmichael, W. W.; Gan, N.; Xu, H.; Shan, K.; Sukenik, A. Harmful Cyanobacterial Blooms: Biological Traits, Mechanisms, Risks, and Control Strategies. *Annu. Rev. Environ. Resour.* **2023**, *48*, 123–147. doi:10.1146/annurev-environ-112320-081653.

11. Chorus, I.; Welker, M. (Eds.). *Toxic Cyanobacteria in Water: A Guide to Their Public Health Consequences, Monitoring and Management*, 2nd ed.; CRC Press/WHO: London, **2021**. doi:10.1201/9781003081449.

12. U.S. Environmental Protection Agency. *EPA Drinking Water Health Advisories for Cyanotoxins*. **2025**.

Chapter 7

Cyanochelin B: a siderophore produced by cyanobacteria with photolytic properties that prevent the monopoly of iron in UV light.

7.1 Synopsis of this chapter [1]

A photolabile siderophore containing β-hydroxyaspartate, cyanochelin B, produced by filamentous Leptolyngbyaceae, has been structurally resolved and linked to light-dependent iron redistribution in microbial communities. The NRPS/PKS-derived lipoheptapeptide was elucidated by high-resolution MS/MS and comprehensive 1D/2D NMR with stereochemical assignment using the Marfey and Murata methods. The ferric complexes of cyanochelin B undergo rapid UV-A-induced photolysis according to zero-order kinetics ($t½ \approx 2.3$ min at 19.6 μmol m^{-2} s^{-1}), while they remain largely intact in the absence of UV or Fe^{3+}; photolysis produces defined fragments (PF1-PF4), at least one of which retains the ability to bind iron, in line with the simultaneous reduction of Fe^{3+} to the more soluble Fe^{2+}. In separate membrane co-culture with immobilized ferric iron, cyanochelin B allows monopolization of the producer in the dark, but UV-A rays abolish this exclusivity and redirect iron to a non-producing competitor (approximately 3.1-fold increase in growth), demonstrating that photolysis loosens siderophore-based privatization.

Biosynthetic gene clusters for cyanochelin B are present in several genera of Leptolyngbyaceae, including *Leptolyngbya*, *Phormidesmis*, and *Myxacorys*, with evidence of plasmid localization that could facilitate horizontal transfer; these taxa often populate subaerophytic mats that experience intermittent iron deficiency and strong light fluctuations. More generally, β-OH-Asp siderophores constitute a photochemically active class capable of reducing ferric iron to ferrous iron under UV-A light, potentially coupling daylight regimes to the iron cycle and 'losing' exchanges between species.[1]

7.2 Siderophores

Cyanobacteria depend on iron as an essential cofactor in the fundamental reactions of oxygenic photosynthesis, respiration, and, where applicable, nitrogen fixation. Since the photosynthetic electron transport chain is densely populated by iron-sulfur proteins, cytochromes, and other metalloenzymes, the cellular iron requirement in cyanobacteria is significantly higher than that of many non-photosynthetic microbes.[2] In fact, intracellular iron concentrations can exceed those of the surrounding water by four to six orders of magnitude, implying that acquisition via steep chemical gradients requires considerable energy investment and tightly regulated transport systems. However, despite iron being one of the most abundant elements in the Earth's crust, its bioavailability in aquatic environments is severely limited. In oxidized waters, soluble ferrous iron (Fe^{2+}) is rapidly oxidized to ferric iron (Fe^{3+}), which hydrolyses and precipitates or forms poorly soluble complexes.

For planktonic cells, iron scarcity is therefore a chronic stress that shapes physiology, gene regulation, and community composition.

To cope with iron limitation, cyanobacteria employ several complementary strategies. At the energy metabolism level, cells can remodel photosynthetic complexes to reduce iron shares; they store iron in ferritin-like proteins to balance supply and demand; and, where possible, they replace iron-containing enzymes with functional analogues that do not require iron. Fundamentally, they also activate high-affinity iron import systems, foremost among which is the biosynthesis and export of siderophores.[3] Siderophores are low molecular weight (typically 400-1000 Da) organic ligands produced by microorganisms and plants that exhibit exceptionally high affinity and selectivity for Fe^{3+}. Their production is regulated upwards when intracellular iron reserves fall below the thresholds necessary to sustain biochemical processes, and secretion into the medium expands the effective range within which cells can recover ferric iron.

The canonical siderophore cycle begins when the secreted ligand encounters available ferric iron and forms a stable iron-siderophore complex. Most classic siderophores coordinate Fe^{3+} through oxygen donor groups (Lewis bases) that bind to the acidic Lewis metal centre, often generating a hexadentate octahedral chelate with remarkable thermodynamic stability. The iron-siderophore complex is then recognized by specific receptors on the outer membrane of the producing cell (and sometimes by competitors), translocated across the envelope via dedicated uptake mechanisms, and released into the cytoplasm. The reduction of Fe^{3+} to Fe^{2+} often facilitates the

release of iron, as many siderophores bind ferrous iron only weakly compared to ferric iron.[4]

The ligand can then be recycled or degraded, depending on the organism and the chemistry of the siderophore. Structurally, siderophores are commonly classified into four archetypal families based on their main donor groups: hydroxamate, catecholate (phenolate), carboxylate, and mixed-type ligands that combine two or more motifs.

Hydroxamate siderophores include paradigmatic chelators such as deferoxamine;[3]

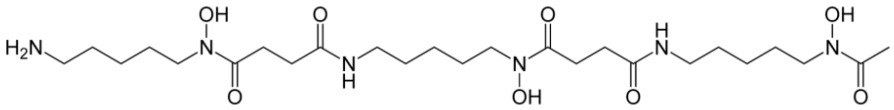


Figure 7.1. Structure of deferoxamine.

catecholate siderophores include enterobactin with its tris-catecholate macrocycle;

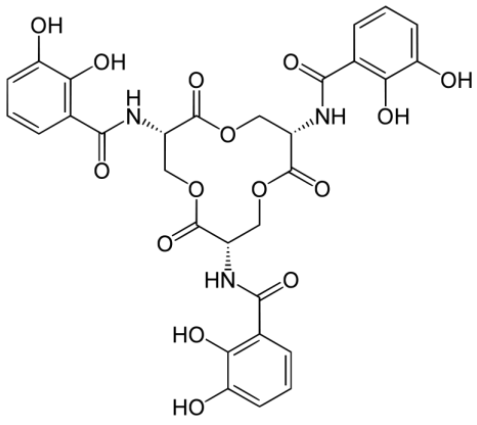


Figure 7.2. Structure of enterobactin.

carboxylate-dominant examples include vibrioferrin;

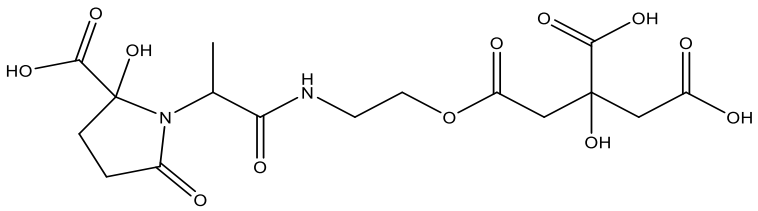


Figure 7.3. Structure of vibrioferrin.

and mixed-type systems include yersiniabactin, which integrates characteristics of thiazoline and phenolate.

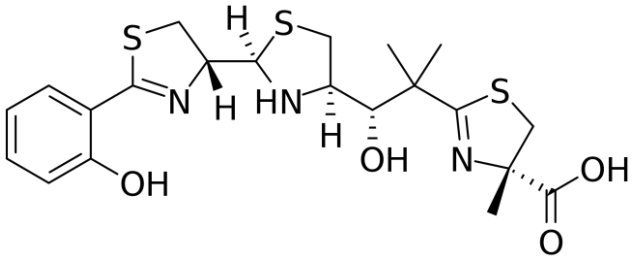


Figure 7.4. Structure of yersiniabactin.

Although this taxonomy is convenient, the chemistry of natural siderophores is highly diverse, and many compounds feature hybrid architectures and auxiliary functional groups that precisely regulate solubility, membrane permeability, metal selectivity, and susceptibility to enzymatic treatment.

In addition to their ecological role, siderophores are of considerable practical interest. As natural iron chelators, they inspire or directly provide therapies for iron overload disorders; the clinical agent deferoxamine (marketed as deferoxamine mesylate, Desferal) is used to remove excess iron in patients, for example during the management of transfusion-dependent thalassemia,

although its intravenous administration limits its practicality and drives the search for orally available alternatives.[5]

In the development of antimicrobials, “Trojan horse” strategies combine antibiotics with siderophore structures to hijack active uptake pathways and enhance the activity of drugs against Gram-negative bacteria. In environmental technology, siderophores contribute to bioremediation by mobilizing iron and associated metals, influencing nutrient cycles and the fate of contaminants in soil and water.

7.3 Cyanobacterial siderophores

Within cyanobacteria, siderophore production is an important response to iron deficiency, complementing physiological adaptations to the photosynthetic mechanism. However, unlike the exhaustive catalogues of heterotrophic bacteria, only a limited number of cyanobacterial siderophores have been documented, and even fewer have been fully characterized structurally.

To date, three examples are often cited as structurally elucidated representatives: schizokinen and synechobactin, both predominantly hydroxamate-type chelators, and anachelin, who’s binding to iron is mainly based on catecholate residues.[6]

These molecules demonstrate that cyanobacteria can access the same chemical strategies used by other bacteria, employing hydroxamate or catecholate donors to achieve a strong bond with iron, while adapting peripheral characteristics to their ecological context, such as life in the photic zone and association with mucilaginous colonies or benthic mats.

The relative scarcity of described cyanobacterial siderophores is likely due to a discovery bias rather than a true scarcity. The recent availability of sequenced cyanobacterial genomes and the identification of multiple “silent” or cryptic biosynthetic gene clusters suggest the existence of a reservoir of undiscovered chelators. Many of these clusters remain unexpressed under standard laboratory conditions. However, deliberate cultivation in iron-free media may trigger their activation, revealing new siderophore patterns and associated transport systems. Knowledge gained from siderophore chemistry in other bacteria provides valuable insights into what to expect: a wide range of structural motifs capable of chelating iron, including residues such as β-hydroxyaspartate (β-OH-Asp), now recognized in several bacterial siderophores (e.g., serobactin, cupriachelin, and pacifibactin). The recurrence of β-OH-Asp in phylogenetically distant producers suggests convergent solutions to the dual challenge of strong binding to Fe^{3+} and efficient uptake. From a mechanistic point of view, the siderophore-mediated iron acquisition pathway in cyanobacteria follows the same general logic observed in other Gram-negative bacteria (export, capture, receptor-mediated import, and intracellular release), but is integrated into the unique constraints of photosynthetic metabolism. Iron scarcity not only reduces the capacity of photosystems I and II, but also alters the stoichiometry of the cytochrome b6f complex and decreases the pool of ferredoxins and other iron-sulfur proteins, with cascading effects on carbon fixation and redox balance. By externalizing iron acquisition through high-affinity ligands, cyanobacteria can maintain growth and pigment synthesis under conditions that would otherwise arrest their proliferation. Furthermore, siderophore production and uptake are often

coordinated with other fitness-enhancing traits, such as buoyancy regulation via gas vesicles (which position cells in iron-rich microenvironments), the secretion of exopolysaccharides (which can bind metals and modulate diffusion), and community-level interactions in which shared siderophores (or “xenosiderophores”) are opportunistically recovered by non-producers.

From an applied perspective, cyanobacterial siderophores occupy a promising, but still largely unexplored niche. Their ligand structures, evolved to function in sunlight-exposed, often alkaline, iron-poor waters, may offer distinctive stability profiles, photochemical behaviours, or receptor specificities that are advantageous for biotechnology. As in the case of deferoxamine and related clinical chelators, medicinal chemistry can exploit cyanobacterial skeletons for iron overload therapy; conversely, synthetic biologists can graft cyanobacterial siderophore transport cassettes into chassis organisms to enhance biolixation or targeted metal recovery. The observation that adding iron-like siderophore groups to antibiotic cores can amplify antimicrobial potency also invites the design of conjugates tuned to cyanobacterial-derived receptors, potentially broadening the spectrum of susceptible pathogens.

In summary, siderophores are fundamental to how cyanobacteria reconcile an iron-hungry metabolism with chronically iron-poor environments. They exemplify a broader adaptive program that includes metabolic reconfiguration, storage, and substitution, but offer a unique advantage by extending the organism's reach into the extracellular space to collect ferric iron with high selectivity and affinity. Although only a handful of cyanobacterial siderophores (schizokinen, synechobactin, and anachelin) are

currently well defined, genomic investigations and inducible expression strategies point to a rich and largely unexplored chemical space. Systematic exploration of this space, guided by environmental triggers such as iron deficiency and informed by recurring motifs such as β-hydroxyaspartate, promises to reveal new molecules and mechanisms at the interface between microbial ecology, coordination chemistry, and practical applications in medicine and environmental remediation.[7]

7.4 Polyketides and polyketide synthase

Polyketides are one of the most fascinating families of natural products and can best be understood as the result of modular, programmable chemistry encoded in giant enzymatic machines. Their field of application is broad: from familiar drugs such as the antibiotic erythromycin and the anticancer drug doxorubicin to cholesterol-lowering agents such as lovastatin and a series of marine metabolites discovered in symbiotic consortia. A useful way to approach these compounds is to start from the idea that a very simple operation, namely the repeated extension of two carbon atoms of an acyl chain, can be orchestrated in many different ways to produce macrocycles, polycycles, densely oxygenated chains, aromatic compounds, and structures rich in stereocentres. This orchestration is carried out by polyketide synthases (PKSs), and the way PKSs are constructed largely explains why polyketides look the way they do.

The central step in the biosynthesis of polyketides is a decarboxylative Claisen condensation that adds a two- or three-carbon extension unit to a growing chain. The extender typically comes in the form of malonyl-,

methylmalonyl-, or, occasionally, ethylmalonyl-CoA, while the initiator can be acetyl-, propionyl-, or butyryl-CoA. The intermediates are not free in solution but are held by a swinging arm of 4'-phosphopantetheine attached to an acyl carrier protein (ACP), which transports them into the catalytic pockets in the correct order. What makes PKS special is what happens after each chain extension cycle. In fatty acid synthases, each elongation is followed by a fixed sequence of ketoreduction, dehydration, and enoyl reduction, so that the product tends toward complete saturation. PKSs, in contrast, can perform all, some, or none of these modifications in a given module. The result is a programmed pattern of β-keto groups, β-hydroxy groups, and double bonds that encode functional handles and stereochemical information along the main chain. When the chain has reached the desired length and oxidation state, it is typically released by a thioesterase domain, often with a macrocyclization that locks the ring size and geometry; other chemical release reactions are also known, including reductive discharge. Late-stage adaptation enzymes, oxidases, halogenases, glycosyltransferases, and others, often decorate specific positions to regulate activity and solubility.

PKS are available in three general architectural solutions.[8]

Type I systems are enormous, multifunctional proteins that connect catalytic domains like stations on an assembly line. In the modular variant, each module is used only once to perform a single cycle of extension and optional processing; the famous erythromycin pathway (the DEBS assembly line) is a system of this type and illustrates the so-called collinearity rule: the number and order of modules tend to mirror the number and order of the product's construction steps.

A second mode of operation, common in fungi, is iterative: the same set of domains cycles several times, with built-in timing that decides when to reduce, when to dehydrate, and when to stop; lovastatin biosynthesis is a classic example of this more economical choreography.

Type II PKSs divide the work between several small monofunctional proteins that work together to form polycyclic aromatics.

A minimal engine consisting of two components, ketosynthase and an ACP, builds a chain of defined length through iterative additions of malonyl; Type III PKS cyclases take minimalism even further: they do not use an ACP at all, but directly accept CoA-activated starters, performing multiple condensations of malonyl and driving intramolecular cyclization in the same pocket to give chalcones, stilbenes, alkylresorcinols, and related polyphenols.

It is useful to keep the domain in mind because their interaction determines the product pattern.

A ketosynthase (KS) receives the growing chain and catalyses the formation of the C-C bond; an acyltransferase (AT) selects and loads the extension unit onto the ACP, which binds and transports intermediates between stations. Optional reductive domains then modify the β position: a ketoreductase (KR) transforms a keto group into a β alcohol with defined stereochemistry, a dehydratase (DH) removes water to install a double bond, and an enoyl reductase (ER) can restore saturation.

Methyltransferase (MT) and epimerase (E) introduce subtle but significant changes in the carbon skeleton, while thioesterase (TE) perform release,

often via macrocycle. In a minimal modular unit, KS-AT-ACP is expected; the presence or absence of KR, DH, and ER determines the oxidation state written in that segment of the chain, and grouping within KR domains can even predetermine the absolute configuration of the emerging β-hydroxy center and influence the geometry of the downstream double bond.

Within type I PKS, two evolutionary strategies are currently recognized.

In cis-AT systems, which are historically better known, each module has its own AT domain. This arrangement tends to respect collinearity and allows modules to specialize in different extensor units, such as malonyl and methylmalonyl, so reading the module map often allows the basic structure and oxidation pattern of the product to be outlined with some precision.

The second strategy, increasingly common in bacteria, is the trans-AT architecture. In this case, the modules do not have built-in AT domains; instead, one or a few autonomous AT enzymes patrol multiple modules, repeatedly loading ACPs, most commonly with malonyl-CoA. On paper, these machines appear modular, but in practice they behave with much greater freedom. The order of domains can be unconventional; modules can be split into distinct polypeptides; apparently redundant segments appear; and specialized cassettes for β-branching or double bond isomerization are common. The net effect is a system that frequently breaks collinearity, complicating product prediction based solely on gene order, but greatly expanding the accessible chemical space. Many marine and soil bacteria use this strategy to produce macrolactins, bacillaenes, and other boldly designed metabolites.

Modern annotation pipelines then predict domain boundaries, identify cis-trans-AT logic, and apply any clues of collinearity to propose a candidate structure for the encoded metabolite. Chemical data obtained from mass spectrometry can be overlaid, features that covary with the presence of the cluster, isotopic labelling reporting the use of extensors, to strengthen the link between genes and molecules.

7.5 PKS-NRPS Hybrid

An additional level of complexity, and opportunity, emerges when PKS modules are interspersed with non-ribosomal peptide synthetase (NRPS) modules. NRPS assembly lines have a modular appearance like that of PKS, but their building blocks are amino acids rather than acyl units.

As said above, adenylation domains (A) select and activate specific residues, peptidyl carrier proteins (PCPs) retain them via the same 4'-phosphopantetheine arm used by ACPs, while condensation domains (C) form amide bonds to elongate the peptide. Accessory domains can epimerize L-amino acids to D-, install N-methyl groups, and cyclize side chains into thiazolines and oxazolines, which often further oxidize to thiazoles and oxazoles. In hybrid pathways, the growing intermediate passes seamlessly from one platform to another: from ACP to a C domain when a polyketide chain is about to receive an amino acid extension (PKS→NRPS), or from PCP to a KS domain when a short peptide fragment is ready for polyketide elongation (NRPS→PKS).[9] This choreography depends on compatible binding surfaces and linkers between domains, subtle recognition motifs on the ACP/PCP helices, and timing that keeps both assembly lines

synchronized. Hybrids earn their place by unlocking chemistry that neither system can handle alone. Macrocycles can fuse ester and amide bonds in the same ring; main chains can be punctuated by heterocycles; side chains can carry precisely positioned hydroxyls, halogens, or methyl groups, supplied by tailor-made enzymes recruited in the cluster. Epotilones are a clear example: a macrolactone core built by PKS that incorporates a thiazole via an NRPS-like module that channels the cysteine-derived chemistry into the structure. Rapamycin offers another model, in which a PKS assembly line inserts a pipecolate unit supplied by an incorporated NRPS module, decisively shaping the conformation and binding properties of the macrocycle.[10]

Cyanobacterial toxins, such as microcystins and nodularins, exemplify the opposite polarity (NRPS→PKS), in which short peptide blocks are extended by polyketide units to regulate ring size and lipophilicity. In several marine trans-AT systems, the interconnection is even deeper: entire NRPS cassettes are inserted midway through assembly to give rise to families whose potent bioactivity depends on this hybridization of logics.

7.6 Cyanochelin B: a cyanobacterial siderophore with double mode of action

Siderophores, as mentioned above, are low molecular weight chelating agents secreted to capture iron in conditions of scarcity. In addition to simple removal, they structure microbial interactions, sometimes allowing a producer to monopolize iron, other times acting as shared resources.

A noteworthy subgroup that coordinates Fe(III) via β-hydroxy-aspartate (β-OH-Asp) is also photolabile: under UV light, these siderophores cleave and photoreduce Fe(III) to more soluble Fe(II), which neighbouring cells can

absorb non-exclusively. This photochemistry has been implicated in mutualism between algae and bacteria and suggests that siderophores may mediate complex community dynamics in systems with limited iron availability.

In collaboration with Dr. Pavel Hrouzek research group, from the Algatech Center, Institute of Microbiology, Czech Academy of Sciences (CAS), in Třeboň, Czech Republic, we report the presence of the siderophore, containing β-OH-Asp, cyanochelin B in the photoautotrophic filamentous cyanobacterium *Leptolyngbya* sp. NIES-3755.[11] We demonstrate that UV light cleaves iron-cyanochelin B complexes within minutes and that some fragments may undergo secondary photolysis. In co-culture, we found that UV exposure abolishes the iron monopoly of *Leptolyngbya* and releases photoreduced iron that directly supports a competing non-producing phototroph (*Synechocystis* sp. PCC 6803).

We also isolated additional producers of cyanochelin B from environmental samples and discuss how photolytic siderophores reshape iron availability and competition in natural communities.

7.7 Planar Structure Elucidation

Cyanochelin B (approximately 2 mg, approximately 99% purity) was obtained by HPLC fractionation of the extract from iron-starved *Leptolyngbya* sp. NIES-3755 (see Methods for cultivation and purification details).

The purified compound was subjected to a full suite of 1D and 2D homonuclear and heteronuclear NMR experiments (^{1}H, ^{13}C, COSY, HSQC,

HMBC, NOESY), complemented by high-resolution mass spectrometry (HRMS). Evaluation of the cyanochelin B structure was further supported by bioinformatic analysis of its putative biosynthetic gene cluster (BGC).

HRMS analysis of cyanochelin B revealed a protonated molecular ion $[M+H]^+$ at *m/z* 1026.5111, consistent with the molecular formula $C_{47}H_{75}N_7O_{16}S$.

MS/MS fragmentation produced a characteristic fragment ion at *m/z* 786.2613, corresponding to the loss of the N-terminal hydroxylated acyl chain (fragment indicated as M-FA, where FA indicates the fatty acyl fraction). The subsequent MS/MS spectrum of the peptide core showed sequential losses of residues from the C-terminus, producing fragment ions at *m/z* 725.2580 (b6-FA, loss of ethanolamine), 594.1889 (b5-FA, loss of β-OH-Asp), 507.1543 (b4-FA, loss of Ser), 376.1325 (b3-FA, loss of β-OH-Asp), 319.1107 (b2-FA, loss of Gly), and 144.0474 (a1-FA, loss of Phe), (Figure 7.12).

This fragmentation pattern was consistent with the predicted amino acid sequence and with the specificity/order of the adenylation domains in the hybrid NRPS-PKS gene cluster encoded in *Leptolyngbya* sp. NIES-3755.

In particular, the detection of the *m/z* 144.0474 fragment (containing sulfur) indicated a thiazoline moiety derived from cysteine at the N-terminal end. NMR data confirmed the peptide core and identified all substructures. The ^1H NMR spectrum in DMSO-d_6 clearly indicated a lipopeptide, with signals for a saturated hydrocarbon chain and multiple NH peptide bonds and α protons. A prominent methyl triplet at δ_H 0.86 (H-47) indicated a terminal CH_3 group; in the COSY spectrum (Figure 7.15), this proton was correlated with a methylene at δ_H 1.26 (H-46), which in turn is part of a long aliphatic chain

extending over 14 methylene units (protons δ_H 1.23-1.51, carbons δ_C 22.1-31.3). This aliphatic chain C15 is attached to a methine carbon containing oxygen at δ_H 3.46 (H-32, δ_C 76.0), which is further connected to a quaternary carbon at δ_C 76.8 (C-30) containing an oxygen substituent and a methyl group (C-31 at δ_H 1.34, δ_C 24.0).

An HMBC correlation (Figure 7.17) from methyl H-31 to quaternary carbon C-30 confirmed this branching point. The quaternary carbon C-30 is connected to a methylthiazoline heterocycle, evidenced by HMBC correlations from H-31 to C-29 and protons on C-27 to C-29. The presence of a 2-methylthiazoline ring is fully consistent with a cysteine residue that was cyclized and N-methylated during biosynthesis (as predicted by BGC, see Figure 7.11).

The thiazoline substructure was further supported by key HMBC and NOESY correlations (Figure 7.17, Figure 7.19) and by the sulfur isotope pattern of the *m/z* 144 fragment in HRMS. The peptide nature of cyanochelin B was evident from NMR signals of multiple amides and α protons. Five distinct amide NH signals were observed in DMSO-d_6 (δ_H 7.65, 8.27, 8.07, 7.83, 7.93 for NH of residues 17, 15, 11, 8, 4, respectively). Correspondingly, four α-proton signals (δ_H 4.69, 4.45, 4.76, 4.60 for H-4, H-8, H-11, H-17, respectively) appeared. Two methylene protons at δ_H 3.89 and 3.80 (doublet, H-15a and H-15b) were attributed to a glycine residue. In addition, two oxymethylene proton signals at δ_H 4.60 (H-5) and 4.51 (H-12) indicated protons on oxygen-containing carbons, consistent with two β-hydroxylated aspartate residues.

Finally, the presence of a fraction of ethanolamine at the C-terminal (a glycine reduced to primary alcohol) was confirmed by an amide proton at δ_H 7.7 (NH-

2) and by COSY/HMBC correlations linking this fragment: for example, COSY connectivity and HMBC cross peaks were observed between the methylene protons of ethanolamine (H-1/H-2) and the carbonyl β-OH-Asp (C-3). These data establish that the terminal residue is ethanolamine, linked via an amide bond to the C-terminal of a β-OH-aspartate.

Combining the above evidence, we deduced that cyanochelin B is a lipo-heptapeptide consisting of a 2-methylthiazoline (methylCys), Phe, Gly, β-OH-Asp, Ser, β-OH-Asp, and ethanolamine (in this sequence from the N-terminal to the C-terminal). This planar structure is summarized in Figure 7.5.

Residue	Position		δH [mult., J (Hz)]	δC [mult.]	HMBC ($^1H \rightarrow ^{13}C$)	NOESY
	OH-1					
	1		3.37, t (6.5)	59.6, CH_2	2	2, NH-2
	2		3.13, m	41.8, CH_2	1, 3	1, NH-2
	NH-2		7.68, t (5.6)		1, 2, 3	1,2, NH-4
	3			168.8, C		
	4		4.69, dd (9.1, 2.3)	56.0, CH	3, 6, 7	NH-2, NH-4
D-β-OH-Asp	5		4.60, d (2.3)	69.7, CH	3, 6, 7	NH-2, NH-4
	OH-5					
	6			173.3, C		
	NH-4		7.93, d (9.1)		4, 7	NH-2, NH-8, NH-11
	7			170.2, C		
	8		4.45, td (7.4, 5.8)	54.5, CH	7, 10	NH-4, NH-8
D-Ser	9	a	3.67, dd (5.8, 10.2)	62.0, CH	7	NH-4, NH-8
		b	3.47, dd (7.3 ,10.2)		7	NH-4, NH-8
	OH-9					
	NH-8		7.83, d (7.5)		8, 10	NH-4, NH-11,11
	10			169.0, C		
	11		4.76, dd (9.1, 2.7)	55.4, CH	10,13,14	NH-8, NH-11
L-β-OH-Asp	12		4.51, d (2.7)	70.2, CH	10, 13,14	NH-8, NH-11
	OH-12					
	13			172.9, C		
	NH-11		8.07, d (9.1)		10,11, 14	NH-8, NH-15
	14			169.0, C		
Gly	15	a	3.89, dd (17.0, 5.6)	41.8, CH_2	14, 16	NH-15, NH-11
		b	3.80, dd (17.0, 5.6)		14, 16	NH-15, NH-11
	NH-15		8.27, t (5.6)		15, 16	NH-11, NH-17, 28
	16			171.0, C		
	17		4.60, td (9.5, 4,4)	53.6, CH	16, 19,25	NH-15, NH-17
	18	a	3.05, dd (13.6, 4.4)	38.0, CH_2	16,17, 19	NH-15, NH-17
		b	2.79, dd (13.6, 10.1)		16,17,19	NH-15, NH-17
L-Phe	19			137.6, C		
	20/24		7.20, m	129.3, CH	18, 19	17
	21/23		7.20, m	127.8, CH	18, 19	
	22		7.16, m	126.1, CH		
	NH-17		7.65, d (9.0)		17, 25	NH-15, 28
	25			173.5, C		
	26			84.0, C		
	27	a	2.97, d (11.3)	40.7, CH_2	25, 26, 28, 29	28
		b	2.82, d (11.3)		25, 26, 28, 29	28
	28		1.25, s	23.8, CH3	25, 26, 27	27a ,27b
	29			178.7, C		
	30			76.8, C		
	OH-30					
	31		1.34, s	24.0, CH_3	29, 30, 32	33a, 33b, 32
	32		3.47, br.d (9.9)	76.0, CH_2	29,34	31
	OH-32					
	33	a	1.51, m	30.5, CH_2	34,35/43	31,32
		b	1.23, m		34,35/43	
	34	a	1.50, m	26.2, CH_2	32,33, 35/45, 44	32,35/43
		b	1.25, m		32,33, 35/45, 44	
	35/43		1.23, m	29.0, CH_2	34,44	34a,34b
	44		1.24, m	28.6, CH_2	35/43,45,46	
	45		1.24, m	31.3, CH_2	35/43,44,46	
	46		1.26, m	22.1, CH_2	44, 45, 47	47
	47		0.86, t (7.1)	13.9, CH_3	45, 46	46

Table 7.1. Summary of NMR analyses of cyanochelin B. (^{1}H 700 MHz, ^{13}C 175 MHz, DMSO-d6, 308K).[1]

7.7 Stereostructural Elucidation

Cyanochelin B contains a total of nine stereocentres: the α carbons of Phe and Ser, two chiral centres in each of the two β-OH-Asp residues, the α carbon of the terminal methylcysteine (in the thiazoline ring), and two adjacent chiral centres (C-30 and C-32) in the N-terminal acyl side chain (a 2,3-dihydroxy-2-methyl fatty acid).

The relative configuration in both β-hydroxyaspartate residues was established by applying Murata's coupling constant analysis for flexible acyclic systems, in which the vicinal values $^{3}J_{H,H}$ and the long-range diagnostic couplings $^{3}J_{C,H}$ are interpreted together, to define the dominant rotamers and the mutual orientation of the substituents.[12] For β-OH-Asp in position 6, the α proton H-4 (δ_H 4.69, dd) shows a small vicinal coupling with H-5 ($^{3}J_{H4,H5}$ = 2.3 Hz); according to Karplus's relationship, this value is characteristic of a *gauche* arrangement. Consistent with this, HMBC sections optimized for long-range couplings (Figure 7.20) reveal a weak $^{3}J_{C,H}$ for H-4→C-6 (carboxyl) and H-5→C-3 (carbonyl), indicating that each carbonyl is also *gauche* relative to the corresponding proton in the preferred conformer. A Newman projection on C4–C5 that satisfies (i) H-4/H-5 *gauche*, (ii) H-4/C-6 *gauche*, and (iii) H-5/C-3 *gauche* uniquely leads to a *threo* relationship between the α and β substituents in Asp^{6} (Fig. 7.20), a conclusion that remains unchanged after conversion to the zig-zag representation. The β-OH-Asp at position 4 behaves similarly: H-11 (δ_H 4.76, dd) couples weakly with H-12 ($^{3}J_{H11,H12}$ = 2.7 Hz), placing these protons in the gauche position; HMBC shows weak H-11→C-13 (carboxyl) and H-12→C-14 (carbonyl) couplings, again consistent with *gauche* arrangements. Corresponding Newman analysis

on C11–C12 similarly yields a *threo* arrangement for Asp[4] (Figure 7.21). The NOESY cross-peaks within each residue are compatible with these preferred rotamers and do not contradict the Murata-based assignments. In summary, both β-OH-Asp residues in cyanochelin B assume *threo* relative configurations to the Cα–Cβ diads, as required by the observed pattern of small vicinal couplings $^{3}J_{H,H}$ and small diagnostic couplings $^{3}J_{C,H}$.

To determine the absolute configurations of the aminoacidic residues in cyanochelin B, advanced Marfey analysis (derivatization with FDAA followed by LC-MS) was performed on the acid hydrolysate of the molecule. Hydrolyzed cyanochelin B was derivatized with 1-fluoro-2,4-dinitrophenyl-5-L-alaninamide (L-FDAA) and, in a separate aliquot, with the D enantiomer of the reagent (D-FDAA), to obtain diastereomeric derivatives of its constituent amino acids.[13] The LC-MS retention times of the derivatized hydrolysate were compared with those of the derivatized standard aminoacids. From these analyses, we identified L-phenylalanine and D-serine as constituents of cyanochelin B (i.e., S configuration at the α carbon of Phe and R at the α carbon of Ser). In the chromatograms of the L-DAA derivatives, the Phe and Ser derived from cyanochelin coeluted with the L-Phe-L-DAA and D-Ser-L-DAA standards at 26.65 min (Phe) and 16.80 min (Ser), respectively. Derivatization with D-FDAA provided a confirmatory result: the D-DAA-derivatized hydrolysate showed peaks corresponding to L-Phe (28.41 min) and D-Ser (16.56 min), in line with the expected retention shifts for the enantiomeric reagent (Figure 7.22).

The assignment of β-OH-Asp enantiomers was more complex, as no commercial D-β-OH-Asp standard was available and β-OH-Asp derivatives

exhibited complex behaviour. However, Marfey's analysis suggested the presence of both enantiomers of *threo*-β-OH-Asp among the hydrolysis products. In other words, cyanochelin B contains one L-*threo*-β-OH-Asp residue and one D-*threo*-β-OH-Asp residue (Figure 7.23). Spectroscopic data did not allow us to determine with certainty at which position (4 or 6) the D form is located relative to the L form.

However, bioinformatic evidence provided by the BGC provides an important clue. A study conducted by Reitz and colleagues[14] showed that, with one exception, β-OH-Asp residues of siderophores occur exclusively as D-*threo* or L-*threo*; furthermore, β-hydroxylases fused with NRPS typically install the S configuration in Cβ, while autonomous β-hydroxylases produce R. In cyanochelin B, the module that should incorporate residue 4 (*CcsI*) lacks epimerase and is preceded by a β-hydroxylase fused with NRPS (*CcsFGH*), thus favouring L-*threo*-β-OH-Asp (11S,12S). In contrast, the module for residue 6 (*CcsL*; BAU16061) is specific for Asp, contains an epimerase domain, and is preceded by an independent β-hydroxylase (*CcsK*), which encodes D-*threo*-β-OH-Asp (4R,5R)

Hydrolysis of cyanochelin B for Marfey analysis also produced a methylcysteine residue (from the opening of the 2-methylthiazoline ring). Due to the lack of D/L methylcysteine standards, we relied on literature data to determine its configuration. Carmeli et al.[15] reported that the L-methylcysteine derivative (prepared as L-methylCys with L-FDAA, forming a diastereomer L-DAA) elutes later than the corresponding D-methylcysteine (or L-methylCys with D-FDAA).

In our analysis, the Marfey products derived from methylcysteine showed retention behaviour corresponding to that of an L-methylcysteine-L-DAA diastereomer. We therefore conclude that methylcysteine in cyanochelin B has the L configuration (i.e., R at the α carbon).

Finally, in the N-terminal dihydroxyacyl residue (C-30, C-32), the stereochemical outcome at C-32 is constrained by the ketoreductase (KR) domain of *CcsC* (BAU16015). Sequence analysis indicates the absence of the canonical LDD and HXXY motifs, which predicts a stereospecific reduction of the β-keto group at C-32 to give an S-configured hydroxyl (C-32S).[16,17,18] Regardless, NMR data support a *threo* relationship between the neighboring diols C-30/C-32: H-32 resonates as a broad doublet (J = 9.9 Hz), a multiplicity reported for *threo* isomers in 2,3-dihydroxy-2-methyl fatty acid analogues (e.g., methyl 2,3-dihydroxy-2-methyl octanoate).[19,20] Considering together the C-32S configuration imposed by KR and the *threo* assignment based on NMR, a definitive configuration (30R,32S) is obtained for the vicinal diol of cyanochelin B.

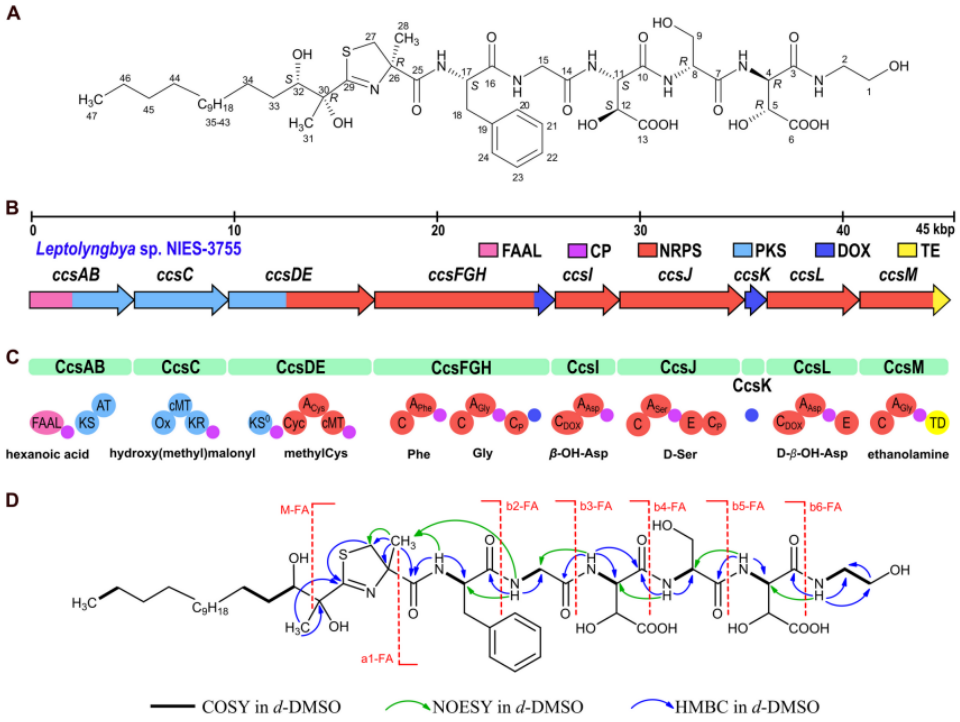


Figure 7.5. Overview of structural features and biosynthesis of cyanochelin B. (A) Structure and stereochemistry of cyanochelin B. (B) Architecture of the gene cluster encoding hybrid PKS/NRPS biosynthesis of cyanochelin B (BGC). The BGC is found on plasmid 2 of *Leptolyngbya* sp. NIES-3755, NCBI accession number: AP017310.1, on coordinates 85398-32159 on the complement strand, please note that the BGC spans over beginning/end of the circular sequence. Genes are color-coded according to the type of biosynthetic step: FAAL, fatty acyl-AMP ligase, pink; CP, carrier protein, purple; NRPS, non-ribosomal peptide synthetase, red; PKS, polyketide synthase, light blue; DOX, aspartate β-hydroxylase, blue; TE, thioesterase, yellow. (C) Modules of the PKS/NRPS encoded by the cluster: KS, β-ketoacyl synthase; KS0, non-elongating β-ketoacyl synthase; AT, acyltransferase; Ox, flavin-containing monooxygenase; cMT, C-methyltransferase domain; KR, β-ketoacyl reductase; Cyc, cyclization (thiazole-forming) condensation domain; C, condensation domain; A, adenylation domain; E, epimerase domain; TD, thioester reductase domain. Subscript specifies the type of C domains or the specificity of A domains. (D) Most relevant 2D NMR correlations and MS fragments of cyanochelin B.[1]

7.7 Kinetics of photolysis and structure of photolytic fragments

Siderophores containing β-OH-Asp can reduce bound ferric iron when exposed to UV light, with simultaneous photolytic cleavage of the siderophore molecule.[17] To understand how cyanochelin B behaves under natural conditions and in culture, we quantified its photolysis rate under UV light. In the absence of UV light (dark controls) or in the absence of iron,

cyanochelin B was relatively stable and showed no significant decomposition. However, under UV-A irradiation, iron-bound cyanochelin B underwent rapid photolysis following apparent zero-order kinetics.

The half-life was approximately 8.9 minutes at our maximum UV intensity and approximately 24.1 minutes at the lower UV conditions used subsequently in co-culture experiments (approximately half the intensity). These half-lives indicate that once cyanochelin B chelates Fe(III), the complex photolyzes within a few minutes under UV exposure equivalent to natural sunlight.

Photolysis produced several distinct products, which we named photolytic fragments (PF1-PF7). LC-MS analysis detected new peaks after UV treatment of Fe-cyanochelin B, and MS/MS spectra of these fragments showed characteristic ions derived from cyanochelin, confirming their origin as siderophore fragments. In several cases (PF1-PF4), we were able to identify the cleavage site of the photoreaction.

Three primary cleavage sites were identified (Fig. 7.6):

1. between C30 and C32 in the aliphatic chain (i.e., between the two hydroxyl carbons of the fatty acyl moiety),
2. between C12 and C13 in β-OH-Asp at position 4 (cleavage of the β–γ bond of that Asp),
3. between the Gly3 amide bond and β-OH-Asp4 (N–C cleavage). These specific breaks suggest that photolysis affects certain vulnerable bonds, probably those adjacent to iron-chelating groups.

Among the photolysis products, the most interesting was fragment PF1 (*m/z* 784.2454).

In particular, when photolysis was performed in the presence of excess ferric iron and analyzed under HPLC conditions that preserve Fe(III) complexes, a species corresponding to PF1 plus iron was observed at *m/z* 837.1553 (corresponding to $[PF1 – 2H + Fe]^+$), with an isotopic pattern consistent with one Fe atom. This indicates that PF1 retains the ability to coordinate iron. Based on MS/MS, PF1 derives from the cleavage of the aliphatic chain between C30 and C32, with concomitant oxidation (the hydroxyl in C32 probably oxidized to carbonyl). PF1 still contains both β-OH-Asp residues, so it remains a hexadentate chelator capable of binding iron. It is not yet clear whether the iron in PF1 is Fe(III) or photoreduced Fe(II); however, we detected an additional fragment PF3 (*m/z* 738.2399) that appears to derive from further photolysis of PF1 (loss of the C12-C13 fragment from β-OH-Asp4), suggesting that PF1 (and similarly PF2) may undergo secondary photolytic cleavage. The structures of the main photolytic fragments inferred from MS/MS are shown in Figure. 7.6

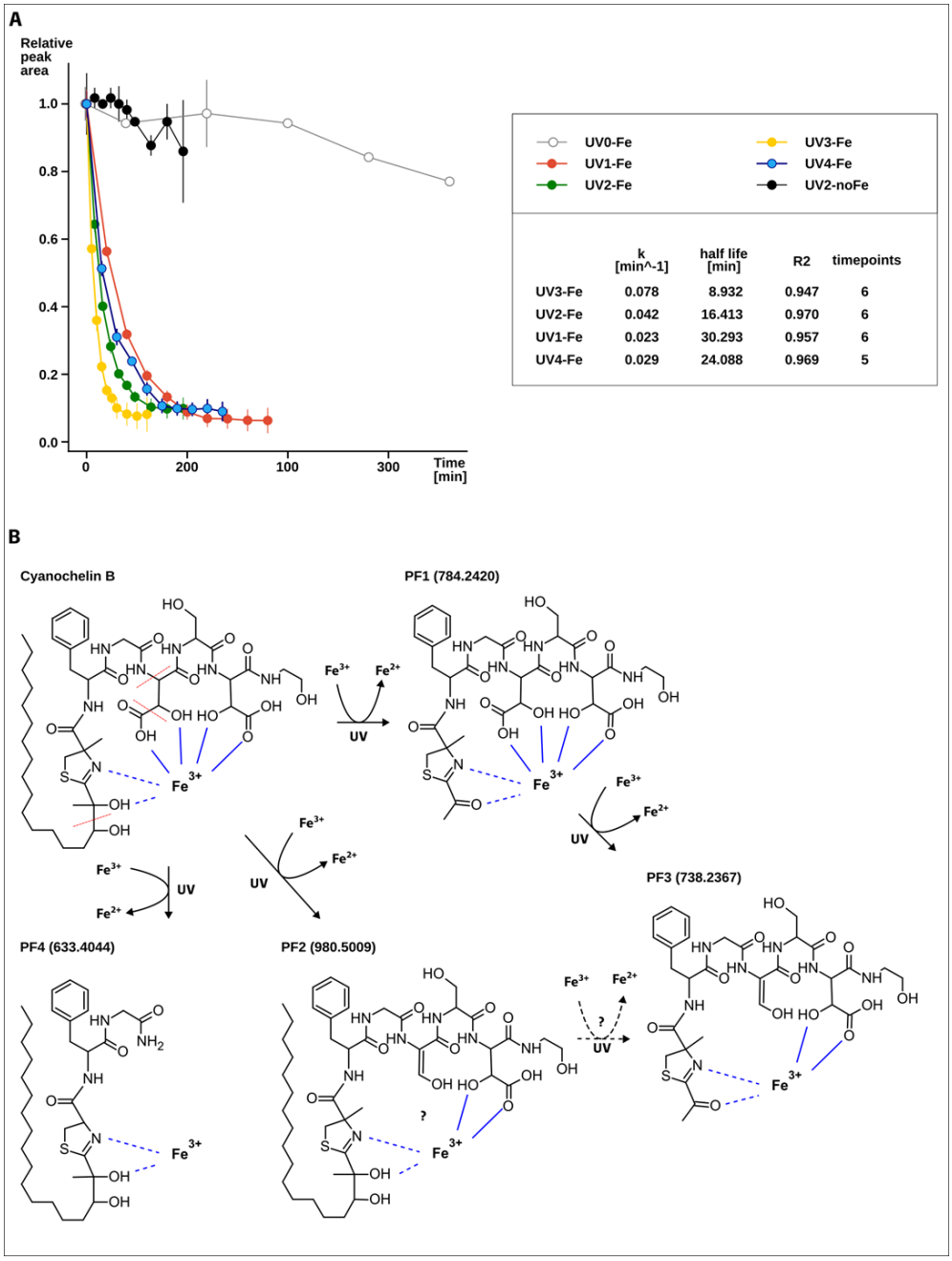


Figure 7.6. (A) The photolysis of Fe-cyanochelin B under UV light produces identifiable fragments (PF1-PF7) through specific cleavage sites (red lightning bolts): between the two diol carbons of the acyl chain (producing PF1/PF2), between the β and γ carbons of β-OH-Asp4 (producing PF3/PF4), and between Gly^3 and β-$OH\text{-}Asp^4$ (producing PF5/PF6). **(B)** Proposed structures for key photolytic fragments PF1–PF4. PF1 retains both β-OH-Asp residues and can bind to iron (observed as Fe complex at *m/z* 837), with C32 oxidized to ketone. PF2 is analogous to PF1 but lacks the terminal ethanolamine (from cleavage b6). PF3 corresponds to PF1 minus the C12-C13 piece of β-$OH\text{-}Asp^4$ (after the second photolysis). PF4 is analogous to PF3 minus the ethanolamine. (The structures of PF5-PF7 have not been fully determined).[1]

7.7 Effect of cyanochelin B on Synechocystis monoculture

To assess the extent to which *Leptolyngbya* could monopolize iron through the production of cyanochelin B, we first verified whether cyanochelin B itself had a direct inhibitory effect on the growth of *Synechocystis* sp. PCC 6803 (a model phototroph that does not produce siderophores). *Synechocystis* and *Leptolyngbya* have similar growth rates and nutritional requirements, so *Synechocystis* serves as a suitable reporter/competitor strain. *Synechocystis* cultures were grown in iron-rich medium (standard BG-11 medium) or iron-deficient medium and exposed to various concentrations of purified cyanochelin B. In iron-deficient medium, *Synechocystis* showed virtually no difference in growth over a wide range of cyanochelin B concentrations, indicating that cyanochelin B does not impose any additional growth inhibition beyond iron deficiency itself. For example, with the addition of 0 μM versus 180 μM cyanochelin B, the OD_{750} of iron-depleted *Synechocystis* cultures increased from 0.2 to 0.65 and 0.60, respectively, during the analysis period, with substantially identical growth in the absence or presence of excess cyanochelin. Therefore, cyanochelin B shows no apparent toxicity or growth suppression towards *Synechocystis*, apart from its role in iron chelation.

In iron-rich BG-11 medium, *Synechocystis* grew to reach higher overall densities, but increasing cyanochelin B concentrations eventually limited this growth by reducing iron availability. *Synechocystis* cultures in standard BG-11 with cyanochelin B up to 2.2 μM reached approximately 1.5 times the cell density of iron-free cultures (as expected, since iron allows for better growth). However, at cyanochelin B concentrations ≥ 6.7 μM, the final density of

Synechocystis in iron-rich medium dropped to approximately the same level as in iron-free conditions. In other words, above a certain cyanochelin: iron ratio, added cyanochelin B sequesters enough iron to deprive *Synechocystis* of iron despite its initial presence. It is important to note that at these higher concentrations, we did not observe any further decrease in growth beyond the effect of iron, supporting the conclusion that the impact of cyanochelin B is solely due to competition with iron and not to any other growth inhibition mechanism. It should be noted that the addition of cyanochelin B to the medium slightly reduces the effective iron concentration; in our setup, cyanochelin was dissolved in iron-free BG-11 before being mixed with the culture, resulting in a final medium with about half the usual iron content (~10 μM Fe). Probably some of this iron also precipitated, meaning that in our tests 6-7 μM of cyanochelin B was sufficient to chelate virtually all bioavailable iron.

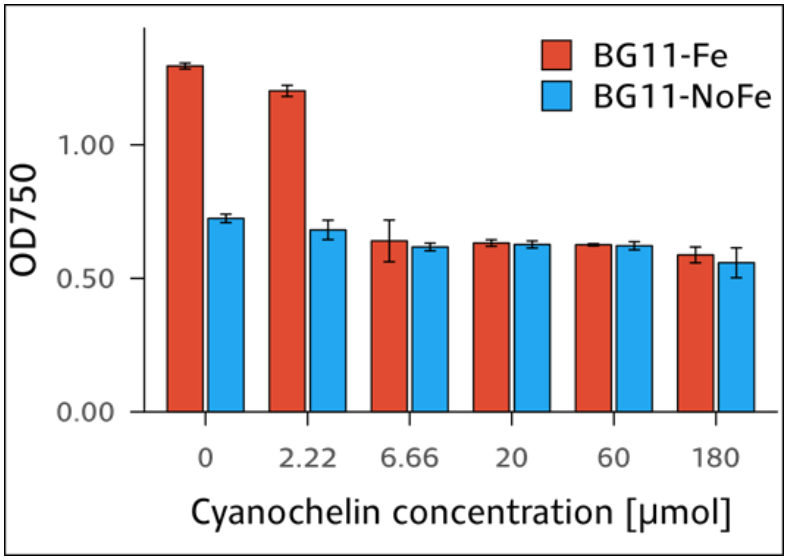


Figure 7.7. Final optical densities (OD_{750}) of *Synechocystis* cultures grown in 96-well plates under different conditions: iron-rich BG-11 (blue bars) vs. iron-deficient BG-11 (orange bars), each with varying concentrations of cyanochelin B (from 0 to 180 µM). The addition of cyanochelin B significantly reduces the growth of *Synechocystis* only in iron-containing media, reaching an inhibition plateau at ≥ 6.7 µM, where growth is as low as in iron-free conditions.[1]

7.7 Effect of UV-light on iron acquisition during co-cultivation

We then examined how the presence of UV light (which allows photolysis of cyanochelin) influences the competition for iron between *Leptolyngbya* (producer) and *Synechocystis* (non-producer) in mixed culture. To this end, we designed a co-culture plate test in which the two strains were grown in the same well but separated by a permeable membrane (which allows the exchange of siderophores and nutrients but prevents cell mixing).

We tested three iron availability scenarios: (1) iron-free medium, (2) soluble ferric iron supplied as ferric ammonium citrate (20 µM Fe; "iron-rich") and (3) an insoluble iron source accessible to siderophores: ferric hydroxide precipitate immobilized in alginate beads ("immobilized iron in beads"). In scenario 3, iron is present in a particulate form accessible only to siderophores

(which diffuse into the alginate spheres and chelate Fe), mimicking iron in its natural bioavailable mineral form. Each of these conditions was tested with and without UV-A illumination. Cultures were prepared in 6-well plates with membrane inserts (ThinCerts®) dividing each well into two compartments (upper insert and lower well). *Leptolyngbya* was inoculated into the insert and *Synechocystis* into the lower compartment, each with an initial OD_{750} of 0.02, and cultured for 14-15 days under continuous light (with or without UV). This setup allowed us to monitor the growth (OD) of each strain separately, while sharing the same culture medium and diffusible components.

As a reference, we first confirmed the growth of each monoculture under these conditions. In monoculture (single strain in an insert, with identical medium in the lower part to ensure equilibrium), *Leptolyngbya* grew well in standard BG-11 medium: after 14 days, it reached an OD of 1.48 under UV-free conditions and 1.64 with UV (slight increase). With iron immobilized in the spheres as an iron source, *Leptolyngbya* reached a lower OD (0.89 without UV, 1.05 with UV), about 60-70% of the growth with completely soluble iron. This indicates that *Leptolyngbya* is able to acquire iron from the spheres, albeit less efficiently than free iron, and that UV light does not hinder iron uptake when present alone (if anything, UV slightly improved the growth of *Leptolyngbya* under all conditions, albeit not significantly). Under conditions of complete iron deficiency, *Leptolyngbya* reached an OD of 0.54 (without UV) and 0.71 (with UV), approximately 36-43% of its growth in an iron-rich medium. These results in monoculture demonstrate that *Leptolyngbya* is able to use iron immobilized by its siderophore and that UV photolysis of its

siderophore does not drastically reduce its ability to acquire iron in the absence of competitors.

Synechocystis monocultures also showed the expected trends. In standard BG-11, *Synechocystis* grew to a higher density than *Leptolyngbya* under the same conditions, reaching an OD of 1.67 (without UV) and 2.74 (with UV). The stimulation by UV rays in the *Synechocystis* monoculture is interesting and could be due to *Synechocystis*' tolerance to UV rays or the use of photoreactive processes (even though it does not produce cyanochelin). With iron immobilized in the spheres, *Synechocystis* reached a final OD of 1.24 (without UV) and 1.71 (with UV), equal to approximately 62-74% of that achieved with soluble iron. This indicates that *Synechocystis* can also access some of the iron present in the spheres, probably by scavenging ferric iron that dissociates or perhaps through weak siderophores or reductive dissolution (although *Synechocystis* is not known to produce siderophores, it could still benefit from iron leaching out of the spheres). In iron-free monoculture, *Synechocystis* grew to OD 0.99 (without UV) and 0.82 (with UV), about 30-60% of its growth in the presence of iron. Interestingly, even without an iron source, *Synechocystis* continued to divide approximately 5-fold, likely drawing on internal reserves or trace contaminants. Overall, both strains were able to utilize the immobilized iron source to some extent, and neither was fatally hindered by the presence of UV in monoculture.

To verify the amount of iron accessible from the beads, we performed ICP-MS measurements on the alginate beads and culture media. Each 15 μL bead was found to contain an average of 2.08 μg of Fe. In 13 days in an iron-free medium, a set of three beads released only 35 ng of Fe into 3 mL of medium

(measured as 12.47 µg/L of Fe in the medium, compared to 1.43 µg/L background in fresh iron-free BG-11). This confirms that the spheres retain most of the iron (several micrograms) and that only minimal traces are released without the action of siderophores, demonstrating that the immobilized iron condition effectively forces the strains to rely on siderophores for iron access.

In co-culture, the dynamics changed significantly. In iron-rich medium (excess soluble iron, no competitive pressure), *Synechocystis* outcompeted *Leptolyngbya* regardless of UV. *Synechocystis* reached an OD of 1.79 (without UV) and 1.57 (with UV), which corresponds to 95-107% of its OD in monoculture under the same conditions. In contrast, *Leptolyngbya* in these co-cultures only reached an OD of 0.14 (without UV) and 0.13 (with UV), which is very low, only 12% (without UV) or 8% (with UV) of what it achieved alone in an iron-rich medium. In essence, when iron is abundant and both strains grow together, *Synechocystis* grows luxuriantly and *Leptolyngbya* barely grows. This probably reflects the fact that *Synechocystis* is the fastest-growing strain under nutrient-rich conditions (perhaps outcompeting its competitors for other resources such as CO_2 or phosphorus) and/or some interference effects. In any case, in the absence of iron deficiency, cyanochelin B is not necessary and *Leptolyngbya* does not thrive in the presence of *Synechocystis*.

The most revealing scenario was co-culture with iron immobilized on microspheres (i.e., iron present but accessible only via siderophores). In the absence of UV, *Leptolyngbya* was able to largely monopolize this iron source. *Leptolyngbya* reached an OD 1.17 in co-culture (without UV), which is 125%

of its OD in monoculture with microspheres. Meanwhile, *Synechocystis* reached only an OD 0.49 (without UV), which is about 40% of its monoculture with beads. Therefore, under dark conditions, *Leptolyngbya* clearly dominated access to iron from the beads via cyanochelin, growing at approximately twice the density of *Synechocystis*. However, under UV exposure, this advantage was negated. With UV, the OD of *Leptolyngbya* in co-culture dropped to 0.65 (only 56% of its OD in co-culture without UV), while the OD of *Synechocystis* rose to 1.84 (about 3.75 times higher than without UV and even slightly higher than that obtained with *Leptolyngbya* in an iron-rich medium). The final cell density ratio changed to approximately 3:1 in favor of *Synechocystis* (with UV), while it was 1:2 in favor of *Leptolyngbya* without UV. In other words, UV-induced photolysis of cyanochelin B completely abolished the iron monopoly of *Leptolyngbya* and allowed *Synechocystis* to access the previously sequestered iron and outperform the producer. This result demonstrates that photolysis of cyanochelin B can "level the playing field" in the competition for iron: the Fe(II) released from the complex becomes available to the non-producer and stimulates its growth at the expense of the producer.

In iron-free co-cultures (total absence of iron sources), neither strain grew well and UV rays had little effect (since no iron-siderophore complexes that could be photolyzed were formed). Both *Leptolyngbya* and *Synechocystis* achieved low OD values (*Leptolyngbya* 0.40, *Synechocystis* 0.20-0.30), roughly corresponding to those obtained in their respective monocultures under iron-deficient conditions. This condition confirms that, in the absence of iron, cyanochelin does not confer any benefit to either strain (and UV rays

are irrelevant because there is no Fe-cyanochelin to photolyze). Any slight difference between UV and non-UV in this configuration is likely due to noise. Overall, the co-culture experiments indicate that cyanochelin B allows *Leptolyngbya* to secure an otherwise inaccessible source of iron (the spheres) and to outcompete *Synechocystis* in the dark. But under UV light, the benefit of cyanochelin to *Leptolyngbya* is compromised: photolysis releases iron in a form that *Synechocystis* can capture, allowing the non-producer to thrive despite not producing a siderophore. In essence, *Leptolyngbya*, by producing a photolabile siderophore, ends up altruistically feeding its competitor in the presence of UV. This finding is consistent with scenarios proposed in natural communities, where photolytic siderophores could support organisms other than the producer.

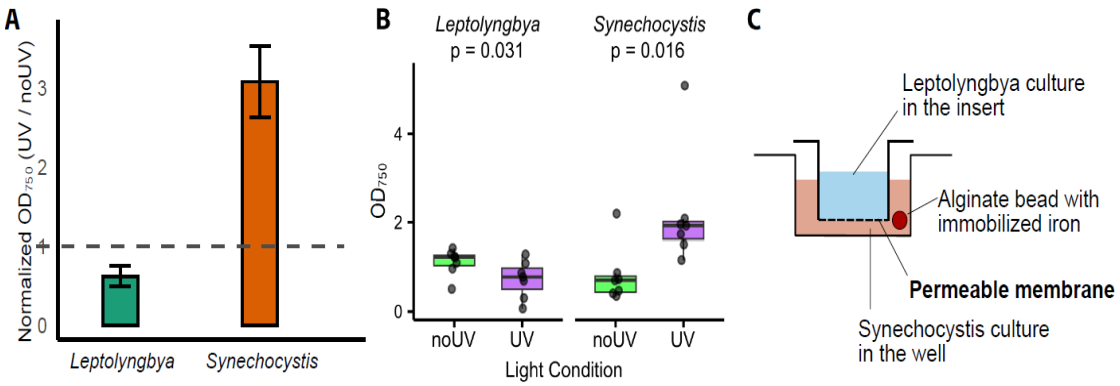


Figure 7.8. Growth results of *Leptolyngbya* (cyanochelin producer) and *Synechocystis* (non-producer) in co-culture under different iron conditions, with and without UV. **(A)** Final OD_{750} of each strain in co-culture with iron immobilized on beads: without UV, *Leptolyngbya* >> *Synechocystis*; with UV, *Synechocystis* >> *Leptolyngbya*. **(B)** Co-culture in iron-rich medium: *Synechocystis* dominates under both light conditions. **(C)** Co-culture in iron-free medium: both strains suppressed (low OD), UV has a negligible effect. Photolysis of cyanochelin under UV specifically affects the condition of iron immobilized in the beads, eliminating the monopoly of iron by *Leptolyngbya*. Error bars indicate ±SD (n=3).[1]

7.7 Phylogenetic distribution and biosynthetic gene clusters of cchB producers

During a field investigation of cyanobacterial mats potentially deficient in iron, we obtained an environmental sample (designated S146) from a wooden bridge wet with spray near a waterfall in the Croatian hinterland. The sample consisted of a subaerophytic microbial mat from a limestone region. Microscopic examination revealed a variety of filamentous cyanobacteria, including *Microcoleus*, *Coleofasciculus*, *Calothrix*, and several morphotypes of Leptolyngbyaceae (small filamentous forms). We cultivated this mixed community in iron-free medium for about 3 months to induce siderophore production. HPLC-MS analysis of the culture supernatant confirmed the production of cyanochelin B (*m/z* 1026.5) by the enriched community.

During enrichment, four distinct filamentous morphotypes similar to *Leptolyngbya* became predominant. To isolate clonal strains, single filaments of each morphotype were selected and transferred to fresh iron-free medium. We obtained several isolates and screened them for cyanochelin B production by LC-MS. Four strains (labelled S146-12, S146-33, S146-35, S146-36) were confirmed as producers of cyanochelin B. Interestingly, all four producing isolates had a very similar morphology corresponding to the genus *Phormidesmis* (a member of the family Leptolyngbyaceae). This suggested that the *Phormidesmis*-like filaments in the original sample were responsible for the production of cyanochelin B. Indeed, three of these isolates (S146-12, -33, -35) were morphologically identical, and a fourth similar strain (S146-20) was also obtained that did not show cyanochelin production.

To genetically identify the producers of cyanochelin B, we designed PCR primers targeting the known cyanochelin B BGC (*cchB* cluster). We based

the primer design on the gene sequence of the *Leptolyngbya* sp. NIES-3755 cluster (reported by Galica et al.)[7] and on a homologous cluster found in the ULC467 strain of *Myxacorys almedinensis* (via BLAST search). Five loci scattered across the approximately 45 kbp BGC *cchB* were targeted. Using genomic DNA from the mixed sample and *Leptolyngbya* NIES-3755 as a positive control, we amplified fragments of the cluster. The environmental sample S146 produced PCR products for three of the five loci (suggesting the presence of *cchB* genes in the community), and sequencing of these fragments indicated slight sequence polymorphisms, suggesting the presence of multiple clusters in the community. We then applied the same PCR screening to our isolated strains. Three of the *Phormidesmis* isolates (S146-12, -33, -35) were PCR positive for multiple *cchB* loci, while the non-productive strain S146-20 was also positive for the cluster (surprisingly, given that it did not produce detectable cyanochelin). The remaining isolate (S146-36) had problems (its genome was contaminated, and we were unable to obtain valid sequences). These results confirmed that at least three strains of *Phormidesmis* carry the cyanochelin B biosynthesis genes and are likely to be the producers observed. We constructed a phylogenetic tree of the 16S rRNA gene sequences of these isolates and related cyanobacteria to classify them taxonomically. The cyanochelin B-producing strains belong to the Leptolyngbyaceae family, grouping into a subgenus that includes *Leptolyngbya*, *Phormidesmis*, *Plectolyngbya*, *Myxacorys*, *Monilinema*, *Alkalinema*, *Planktolyngbya*, etc. Our isolates are grouped with *Phormidesmis* and are closely related to *Leptolyngbya* NIES-3755 (the original strain) and *Myxacorys almedinensis* A, all containing the *cchB* cluster. Figure 7.9 shows a phylogenetic tree

highlighting the known producers of cyanochelin B (marked with an asterisk) in this clade. The presence of the *cchB* cluster in these taxa suggests that the trait may be ancestral in this sublineage, although frequent loss events (or horizontal transfer) may have distributed the cluster sporadically.

In particular, *Leptolyngbya* NIES-3755 and *Myxacorys* A carry the cluster on plasmids rather than on the chromosome, facilitating horizontal gene transfer. This raises interesting evolutionary questions about the conservation of this siderophore trait within the "Black Queen" hypothesis: perhaps non-producing strains persist in communities by relying on photolytic release of iron provided by producers.

We sequenced the entire genome of the three confirmed producing isolates (S146-12, -33, -35) and performed a bioinformatic analysis of their *cchB* biosynthetic gene clusters. All three strains possessed a cluster essentially identical in organization to that of *Leptolyngbya* NIES-3755, with nine core genes encoding a hybrid PKS/NRPS assembly line, plus additional adaptive enzyme genes.[21,22,23] Figure 7.10 compares these clusters, showing that the architecture of the PKS-NRPS module is conserved in all strains. Minor differences included the absence of three terminal genes (presumably an ABC transport system, DevA/B/C homologs) in the *Phormidesmis* clusters compared to *Leptolyngbya*. The core genes of the assembly chain (including adenylation domain specificities, epimerase domains, etc.) were the same, which explains why all these strains produce cyanochelin B.

Furthermore, the genome of *Myxacorys almedinensis* was found to contain a very similar cluster (via BLAST), strongly suggesting that it may also

produce cyanochelin B or a very similar analogue. Note that we did not experimentally confirm the production of *Myxacorys* in our study.

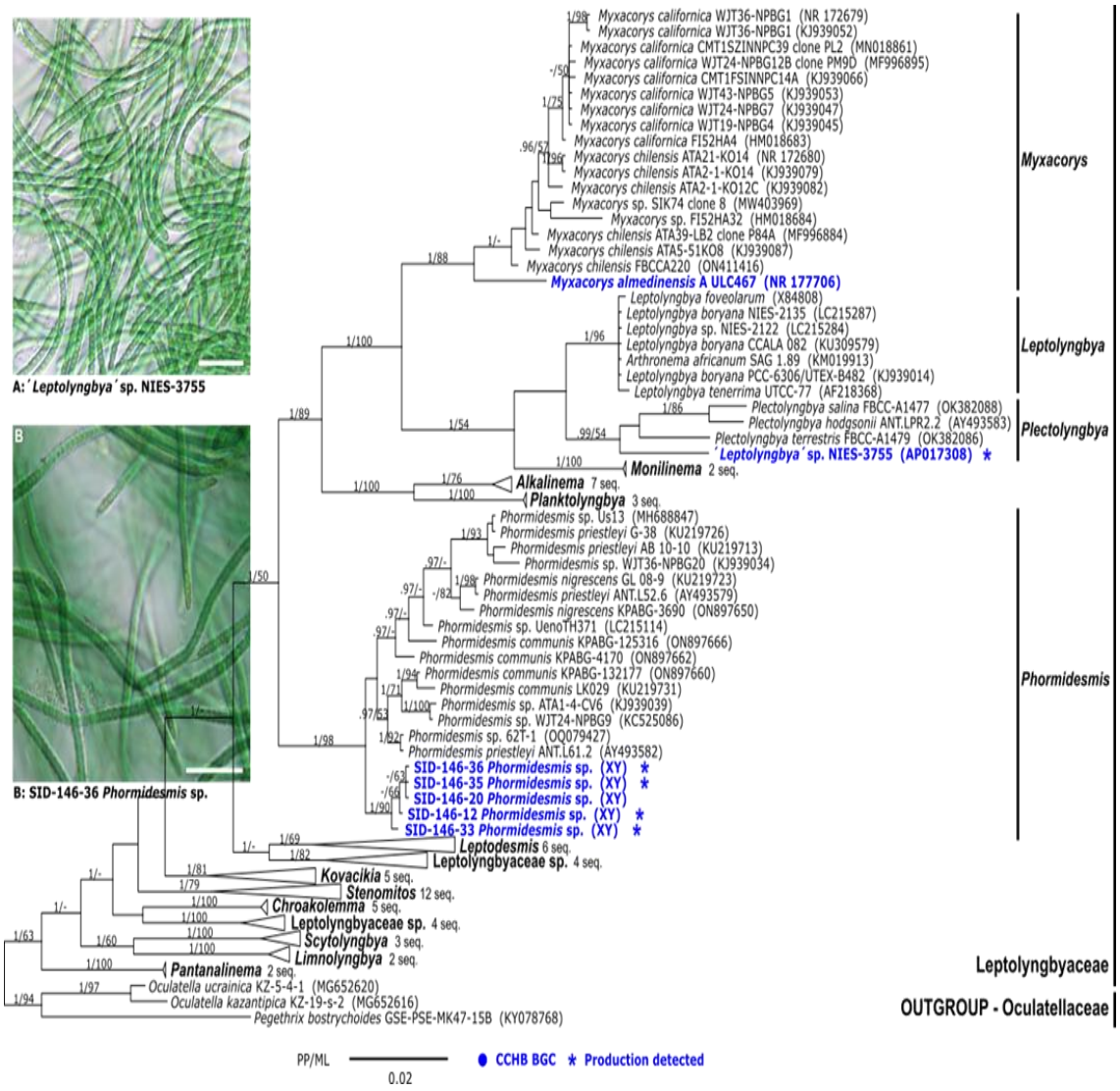


Figure 7.9. Phylogeny of the 16S rRNA gene of selected strains of Leptolyngbyaceae, highlighting those with cyanochelin B biosynthetic gene clusters. Strains confirmed to produce cyanochelin B (by HPLC-MS) are marked with an asterisk. *Leptolyngbya* sp. NIES-3755 (previously known as a producer) and the new isolates of *Phormidesmis* from sample 146 belong to the same clade. *Myxacorys almedinensis* A also contains a homologous cluster (by genome analysis). The tree was inferred using Bayesian methods; branch support values (posterior probability and bootstrap) are indicated on the nodes. Scale bar = 0.02 substitutions per site.[1]

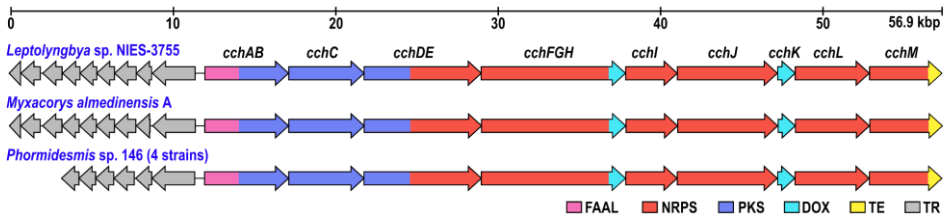


Figure 7.10. Comparison of cyanochelin B biosynthetic gene clusters from different cyanobacteria. The central PKS/NRPS genes (pink arrows) covering approximately 45 kb are highly conserved in gene order and domain content among strains (*Leptolyngbya* NIES-3755, *Phormidesmis* S146-12/33/35, *Myxacorys* A). The colored domains correspond to those in Fig. 7.5. Epimerase domains (E) are present in modules for Ser5 and Asp6 (blue stars). The *Phormidesmis* clusters lack three genes (devABC homologs) present at the end of the *Leptolyngbya* cluster (probably an export transporter). The position of the plasmid or chromosome of each cluster is indicated (the NIES-3755 and *Myxacorys* A clusters are plasmids). The high similarity suggests a recent common origin or horizontal transfer of the cluster between these filamentous cyanobacteria.[1]

7.8 Experimental section

7.8.1 General methods

1D and 2D NMR experiments were carried out at 35°C on a Bruker AvanceNeo 700 MHz spectrometer (Billerica, MA, US) equipped with a triple resonance CHN cryoprobe, using d-DMSO (Sigma Aldrich, Milan, Italy); chemical shifts were referenced to the residual solvent signal (d-DMSO: δ_H 2.50 ppm, δ_c 39.51 ppm.). The HSQC spectra were optimized for $^{1}J_{CH}$ = 145 Hz and the HMBC experiments for $^{2,3}J_{CH}$ = 8 Hz. Abbreviations for signal couplings are as follows: s = singlet, d = doublet, br.d = broad doublet, dd = doublet of doublets, td = triplet of doublets, t = triplet, q = quartet, m = multiplet.

7.8.2 Strains and cultivation

Leptolyngbya sp. NIES-3755 was obtained from the NIES culture collection (Japan) and cultivated in BG-11 medium. For the production of cyanochelin

B, the cultures were subjected to successive transfers to iron-depleted BG-11 medium until siderophore production was detectable by HPLC-MS, then harvested in bulk. *Synechocystis* sp. PCC 6803 (Nixon strain) was kindly provided by Prof. Peter Sobotka. All strains were maintained at 21 ± 2 °C under continuous low-intensity light (5–10 μE m^{-2} s^{-1}, diffused white fluorescent light).

7.8.3 Siderophore inhibition test

To evaluate the possible direct inhibitory effect of cyanochelin B, *Synechocystis* was cultured for approximately 7 days in iron-free BG-11 until OD_{750} ≈ 0.82 (measured in a 1 cm cuvette). The culture was then harvested by centrifugation (3000 rpm, 5 min) and washed with iron-free medium to remove residual iron, then resuspended in standard BG-11 or iron-free BG-11 until OD_{750} = 0.4. These cell suspensions (100 μL per well) were mixed 1:1 with cyanochelin B solutions (in iron-free medium) at various concentrations so that the final OD_{750} of *Synechocystis* was 0.2 and the cyanochelin B concentrations were 0, 2.2, 6.7, 20.0, 60.0, and 180.0 μM. The cultures were incubated in 96-well plates (final volume of 200 μL per well) at 25 °C under white light of ~300 μE m^{-2} s^{-1} for 10 days. Growth was monitored by measuring OD_{750} in a plate reader (BMG Labtech Fluostar Omega). Each treatment was performed in triplicate wells. The final OD_{750} readings were compared between the different concentrations of cyanochelin.

7.8.4 Co-cultivation experiments

Leptolyngbya was inoculated into the insert compartment and *Synechocystis* in the lower well compartment; each had an initial OD_{750} = 0.02 in 3 mL of medium. Three iron conditions were tested: (1) iron-deprived BG-11 (no added iron source), (2) standard BG-11 with 20 μM ferric ammonium citrate (soluble iron), and (3) BG-11 with no soluble iron but containing ferric oxyhydroxide precipitate immobilized in alginate beads (bead-immobilized iron). Ferric-iron alginate beads were prepared by precipitating $FeCl_3$ (0.5 M) with NaOH (1 M) in a 1:9 ratio, collecting the precipitate, mixing it into 2% sodium alginate, and dropping the mixture into 0.1 M $CaCl_2$ to form 3-4 mm diameter beads (each bead 15 μL volume, containing roughly 2-3 μg Fe as determined by ICP-MS). Three beads were added per well for the immobilized iron condition. The pH of all media was adjusted to 7.5 before inoculation.

Plates were incubated at 25 °C either under constant white light (~50 μE) or the same white light supplemented with UV-A LEDs (3.5 μE, peak 360 nm). "UV" and "non-UV" treatments were done in parallel with identical inocula. Cultures were gently shaken (50–75 rpm on an orbital shaker) to aid diffusion. The inserts allowed diffusion of molecules (siderophores, nutrients) but kept cells physically separate. After 15 days, cultures were harvested; each compartment was thoroughly mixed and OD_{750} measured for each strain separately (for insert, contents were measured; for well, contents measured). Growth yield (final OD) of *Leptolyngbya* and *Synechocystis* in each condition was compared to monoculture controls. All experiments were done in triplicate.

For the iron bead condition, inductively coupled plasma mass spectrometry (ICP-MS) was used to measure iron content in beads and media. Samples of dried beads were digested in acid and analysed against Fe standards. Media from bead-containing wells (after incubation) were also analysed for dissolved Fe. These measurements confirmed bead Fe content (2.08 μg Fe per bead) and minimal Fe leakage into iron-free media (0.012 μg/mL over 13 days).

7.8.4 HPLC-HRMS analysis

The concentrations of cyanochelin B and its photoproducts were analysed using a Dionex Ultimate 3000 UHPLC system coupled with a Bruker Impact HD II Q-TOF mass spectrometer (electrospray ionization, positive mode). Chromatographic separations were performed on a Waters MaxPeak XBridge Premier BEH C18 column (50 × 2.1 mm, 2.5 μm, 130 Å). The mobile phases were 0.1% formic acid in water (solvent B) and 0.002% formic acid in acetonitrile (solvent A). The flow rate was 0.4 mL/min. A gradient elution was used as follows: 0-2 min 5% A (95% B is aqueous); 2-8 min ramp from 5% to 100% A (increase in acetonitrile, using a convex gradient curve for a gradual increase); 8-10 min hold at 100% A; 10-11 min return to 5% A; 11-12 min re-equilibration to 5% A. The column was maintained at 40 °C.

Mass spectrometer settings: capillary voltage 4500 V, capillary temperature 220 °C, N_2 coating gas 12 L/min, nebulizer 1.0 bar. High-resolution MS and MS/MS data were acquired with automatic fragmentation dependent on the data of the first 3 ions per scan. Exact mass calibration was performed with

sodium formate clusters. Data were processed with Bruker Compass DataAnalysis 4.4.

7.8.4 Compound isolation

To isolate cyanochelin B, we used adsorption and preparative HPLC. The exhausted culture media (approximately 4-6 L of iron-depleted *Leptolyngbya* culture) were gently mixed with Amberlite XAD-16 resin (20 g/L) overnight to adsorb the organic compounds. The resin was then washed with water and 10% methanol to remove salts and polar impurities and eluted with 100% methanol. The methanol extract was evaporated under reduced pressure (rotary evaporator, 40 °C) to obtain a crude extract, which was stored at –20 °C.

Before HPLC, the extract was redissolved in minimal methanol and filtered. Preparative HPLC was performed on a Waters μBONDAPAK Phenyl reverse phase column (7.8 mm × 300 mm, 10 μm particles, 125 Å pores).

Mobile phase A was water with 0.08% TFA (trifluoroacetic acid) and mobile phase B was acetonitrile with 0.002% TFA. The flow rate was 3-4 mL/min with a linear gradient: 0–2 min 100% B (equilibrate the column); 2-8 min ramp up to 50% B; 8-24 min ramp up to 65% B; 25-28 min hold at 100% B; then re-equilibrate. (Note: the initial gradient was 100% aqueous; in practice, we started with 15% A/85% B for better retention, then increased B over 25 min to 100%.) Fractions were collected every 0.75 min. Each fraction was analysed by analytical HPLC-HRMS as above to identify those containing *m/z* 1026 (cyanochelin). The target fractions were collected and dried. If necessary, a second purification step was performed on the same column with

a shallower MeOH/water gradient to improve purity. The final yield of cyanochelin B was approximately 2 mg with >99% purity (by UV and MS).

7.8.4 Advanced Marfey's analysis

Cyanochelin (10 μg) was hydrolysed with 600 μL of 6 N HCl/AcOH (1:1) at 120 °C for 18 h. At the end, residual HCl fumes were removed under a flow of N_2. The hydrolysate was dissolved in TEA/acetone (2:3, 200 μL) and derivatized by adding 200 μL of a 1% solution of 1-fluoro-2,4-dinitrophenyl-5-L-alaninamide (L-FDAA) in CH_3CN/acetone (1:2), according to Marfey (1984).[13] An additional aliquot of cyanochelin (10 μg) was treated in the same way and derivatized with 200 μL of 1% 1-fluoro-2,4-dinitrophenyl-5-D-alaninamide (D-FDAA). The vials were incubated at 50 °C for 2 h, then the samples were dried, and the L-DAA and D-DAA derivatives of the free amino acids were redissolved in MeOH (200 μL) for subsequent analysis. Authentic standards of L-Tyr, D-Ser, and L-β-OH-Asp were derivatized with L-FDAA and D-FDAA as above to obtain the corresponding L-DAA and D-DAA standards.

Marfey derivatives of cyanochelin B were analysed with a Thermo U3000 HPLC system coupled to a Thermo LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). Retention times were compared with those of the derivatized standards. Separation was performed on a Kinetex C18 column, 5 μm (50 × 2.10 mm) at 25 °C, with a flow rate of 200 μL min^{-1} and mobile phase 0.1% HCOOH in H_2O and ACN.

Gradient: 5% ACN for 3 min, 5–50% ACN in 30 min, 50–90% ACN in 1 min, then 90% ACN for 6 min. Mass spectra were acquired in positive ESI

mode with the following parameters: spray voltage 4.8 kV, capillary temperature 285 °C, nebulizing gas (N_2) 32 units (230 mL min^{-1}) and auxiliary gas (N_2) 15 units (150 mL min^{-1}). The DDA mode involved four HRMS/MS scans on the most intense ions after each full MS scan; *m/z* range 150–2000, resolution 60,000. The HRMS/MS scans were obtained with CID, isolation width 4.0, normalized collision energy 35, activation Q 0.250, and activation time 30 ms. Data processing was performed using the Xcalibur suite.

For the analysis of β-OH-Asp (standard and cyanochelin-L-DAA/D-DAA derivatives), a Phenomenex Luna Omega Polar C18 column, 3 μm (100 × 2.10 mm) at 45 °C, with a flow rate of 400 μL min^{-1} and the same mobile phase (0.1% HCOOH in H_2O/ACN). Gradient: 10% ACN for 1.5 min, then 10–95% ACN in 8.5 min, and 95% ACN for 2 min.

Since the amino acid formed several adducts with the derivatization reagent, a targeted evaluation of the chromatograms by LC-MS in SRM mode was also performed. In this approach, the analysers were set to specific *m/z* ratios to follow precursor/product pairs: the MS method included two HRMS/MS events to monitor the *m/z* 356.1-358.1 and 280.1-282.1 fragments starting from the *m/z* 402.2 precursor ion,

7.8.4 Field sampling and strain isolation

Environmental sample S146 was collected from a microbial mat on a wooden bridge in the spray zone of a waterfall (Croatia, GPS N 44.04152, E 16.23488, altitude ~87 m). The mat was scraped off and collected in sterile 50 ml tubes. Part of the fresh sample was examined under a microscope, confirming the

predominance of filamentous cyanobacteria. The sample was divided for (a) cultivation, (b) DNA preservation (in liquid nitrogen at -80 °C), and (c) preservation in glycerol (20% glycerol, -80 °C). For cultivation, pieces of the mat were placed in 250 mL bottles with iron-free Z medium (a freshwater medium) and incubated at room temperature with low light. After approximately 3 months, LC-MS analysis indicated the production of a siderophore (m/z 1026). The enrichment culture was then seeded on agar and diluted to extinction in liquid to isolate clonal strains. Single filaments were picked up under a microscope and transferred to fresh iron-free BG-11. Four isolates similar to *Leptolyngbya* growing in monoculture were obtained; these were further cultivated under iron-free conditions to test cyanochelin production. S146-12, -33, -35, -36 showed a peak at *m/z* 1026 by LC-MS, while S146-20 did not (although morphologically similar). The producing isolates were cryopreserved and used for genetic analysis.

7.8.4 Cyanochelin B PCR protocol

PCR primers were designed to amplify five loci across the cyanochelin B BGC using sequences from *Leptolyngbya* sp. NIES-3755 (12) and a homologous cluster in *Myxacorys almedinensis* A (ULC467; GCA_010091945.1). Genomic DNA from sample 146 and NIES-3755 (positive control) was extracted with the NucleoSpin Soil Mini kit (Macherey-Nagel, Düren, Germany). Locus-specific PCRs were performed and Sanger sequenced (SeqME, Dobříš, Czech Republic) as in Ref. 24. Sample 146 was PCR-positive in loci 1, 4, and 5, and observed SNPs indicated cluster variants. The same protocol was used to screen clonal

isolates (three strains per *Leptolyngbya-like* morphotype); PCR-positive amplicons were sequenced as above. Taxonomic identity of three strains per morphotype was verified by 16S rRNA PCR/sequencing using VRF2 (5′-GGG GAA TTT TCC GCA ATG GG-3′) and VRF1 (5′-CTC TGT GTG CCT AGG TAT CC-3′) with conditions from Johansen et al.[23]

7.8.4 Whole-genome sequencing and assembly

For each of the five *Phormidesmis* clonal strains from sample 146, single filaments were isolated by glass capillary technique. Genomic DNA was amplified by multiple-displacement amplification (MDA) using REPLI-g Mini (Qiagen).[25] MDA products from 5–8 filaments per strain passing 16S QC were pooled and submitted for de novo Illumina HiSeq paired-end sequencing (150 bp reads; 2.5 Gbp per sample) at the Institute of Experimental Botany (Olomouc, Czech Republic). Reads were trimmed, assembled, binned (to remove minor bacterial contamination), and taxonomically classified with the nf-core/mag metagenomics pipelineon the Institute of Hydrobiology server (Biology Centre, CAS).[25]

7.8.4 Phylogenetic analysis

Complete rRNA operons of *Phormidesmis* sp. SID-146-12, -20, -33, and -35 were excised from genomes (Geneious Prime 2024.0.7). For SID-146-36, a partial 16S was obtained by Sanger sequencing of a PCR product (primers VRF2 and 1494, 5′-CTA CGG CTA CCT TGT TAC GA-3′).[23,26] Sequences are deposited in GenBank (PQ675769–PQ675773). Alignment included 108 taxa (Leptolyngbyaceae + 3 Oculatellaceae outgroup), 1112 positions, built

in MEGA11[27] with MUSCLE[28] and manually curated. Bayesian inference was run in MrBayes v2.6[29] with default settings for 5 million generations; posterior probabilities were recorded. Maximum likelihood analysis used PhyML[30] under TN93+R (model chosen by SMS) (55) with 1000 nonparametric bootstraps. Trees were merged; the figure was prepared in Inkscape (https://inkscape.org/). Supports shown: ML ≥50, PP ≥0.95. p-distance checks (rRNA) and BGC similarity among novel strains were calculated in MEGA11.

7.8.4 Analysis of the cyanochelin BGCs

Genome bins of *Phormidesmis* strains were screened with antiSMASH v7.0 (bacteria mode) . Contigs with KnownClusterBlast similarity to cyanochelin A were examined for NRPS/PKS domain composition, A-domain substrate specificity (Stachelhaus code), epimerase (E) domains, and tailoring enzymes, to connect cluster content with cyanochelin B structure. The stereochemical outcome of the PKS ketoreductase (KR) was predicted by aligning the *ccsAB*-encoded KR with reference bacterial KRs and analyzing active-site groove residues.

7.9 Cyanochelin B biosynthesis and Spectroscopic data

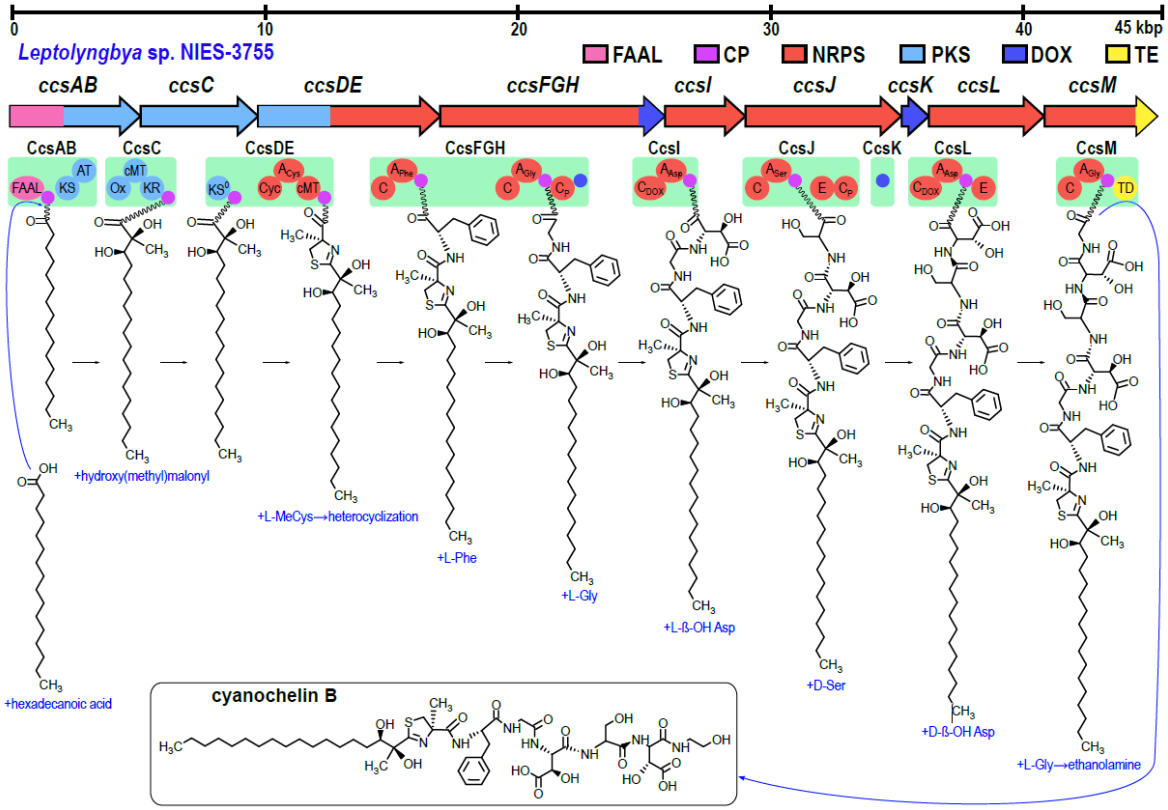
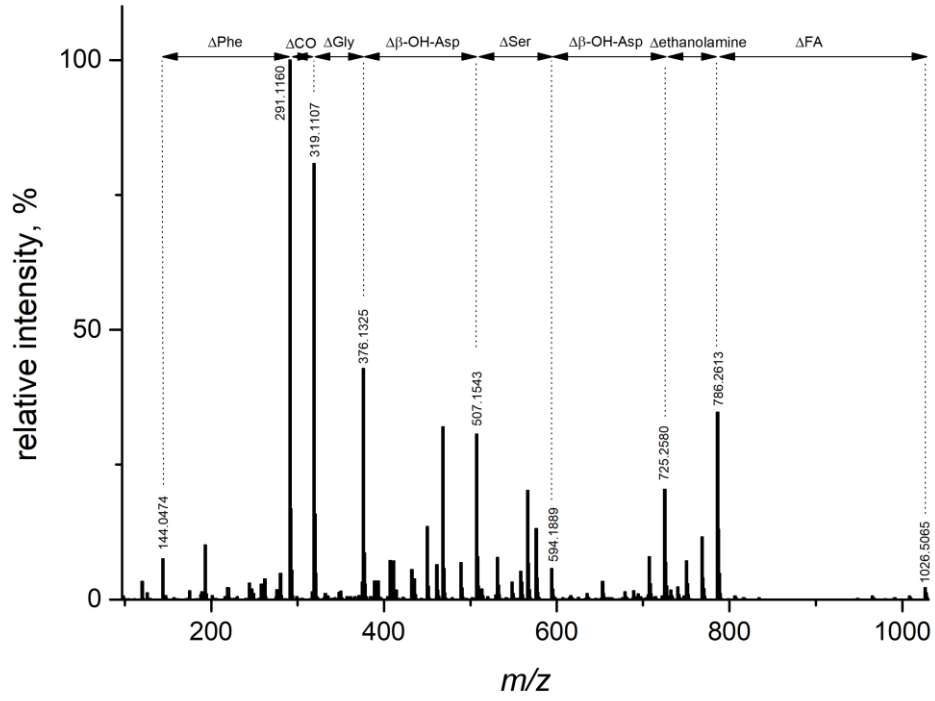


Figure 7.11. Reconstruction of cyanochelin B biosynthesis in *Leptolyngbya* sp. NIES-3755. Cyanochelin B is synthesized by a PKS/NRPS pathway enconded by 9 nine genes grouped in ~45 kbp long BGC depicted by arrows. Genes are depicted in scale and portion of the genes are colorized by according to type of biosynthetic step they perform. Multidomain proteins (green frames) are depicted with individual domains and attached intermediate. Abbreviations: FAAL - fatty acyl-AMP ligase, CP - carrier protein, NRPS - non-ribosomal peptide synthetase, PKS - polyketide synthetase, DOX- aspartate β-hydroxylase, TE - thiesterase.[1]

Strain	A-domain	Protein	Position	Stachelhaus Prediction	Stachelhaus Code	Reconstructed Substrate	Observed AA
M. almedinensis A	A1	CcsE	1407-1803	Cys	DLYNLSLIWK	Cys	Cys
	A2	CcsFGH	505-896	Phe	DAWTIAAVCK	Phe	Phe
	A3		1551-1942	Gly	DILQLGLIWK	Gly	*Gly*
	A4	CcsI	450-850	Asp	DLTKIGHVGK	Asp	*β*-OH-Asp
	A5	CcsJ	510-911	Ser	DVWHFSLIDK	Ser	Ser
	A6	CcsL	450-853	Asp	DLTKIGHVGK	Asp	*β*-OH-Asp
	A7	CcsM	470-863	Gly	DILQLGLIWK	Gly	Ethanolamine
Phormidesmis sp. 146-12	A1	CcsE	1405-1802	Cys	DLYNLSLIWK	Cys	Cys
	A2	CcsFGH	504-895	Phe	DAWTIAAVCK	Phe	Phe
	A3		1550-1941	Gly	DILQLGLIWK	Gly	*Gly*
	A4	CcsI	448-848	Asp	DLTKVGHVGK	Asp	*β*-OH-Asp
	A5	CcsJ	510-911	Ser	DVWHFSLIDK	Ser	Ser
	A6	CcsL	451-854	Asp	DLTKIGHVGK	Asp	*β*-OH-Asp
	A7	CcsM	470-862	Gly	DILQLGLIWK	Gly	Ethanolamine
Phormidesmis sp. 146-20	A1	CcsE	1406-1802	Cys	DLYNLSLIWK	Cys	Cys
	A2	CcsFGH	504-895	Phe	DAWTIAAVCK	Phe	Phe
	A3		1550-1941	Gly	DILQLGLIWK	Gly	*Gly*
	A4	CcsI	448-848	Asp	DLTKVGHVGK	Asp	*β*-OH-Asp
	A5	CcsJ	510-911	Ser	DVWHFSLIDK	Ser	Ser
	A6	CcsL	452-855	Asp	DLTKIGHVGK	Asp	*β*-OH-Asp
	A7	CcsM	470-862	Gly	DILQLGLIWK	Gly	Ethanolamine
Phormidesmis sp. 146-33	A1	CcsE	1406-1802	Cys	DLYNLSLIWK	Cys	Cys
	A2	CcsFGH	504-895	Phe	DAWTIAAVCK	Phe	Phe
	A3		1550-1941	Gly	DILQLGLIWK	Gly	*Gly*
	A4	CcsI	449-849	Asp	DLTKVGHVGK	Asp	*β*-OH-Asp
	A5	CcsJ	510-911	Ser	DVWHFSLIDK	Ser	Ser
	A6	CcsL	452-855	Asp	DLTKIGHVGK	Asp	*β*-OH-Asp
	A7	CcsM	470-862	Gly	DILQLGLIWK	Gly	Ethanolamine
Phormidesmis sp. 146-35	A1	CcsE	1406-1802	Cys	DLYNLSLIWK	Cys	Cys
	A2	CcsFGH	504-895	Phe	DAWTIAAVCK	Phe	Phe
	A3		1550-1941	Gly	DILQLGLIWK	Gly	*Gly*
	A4	CcsI	448-848	Asp	DLTKVGHVGK	Asp	*β*-OH-Asp
	A5	CcsJ	510-911	Ser	DVWHFSLIDK	Ser	Ser
	A6	CcsL	452-855	Asp	DLTKIGHVGK	Asp	*β*-OH-Asp
	A7	CcsM	470-862	Gly	DILQLGLIWK	Gly	Ethanolamine

Table 7.2. Specificities of the A-domains in the identified biosynthetic gene clusters. A-domains in the order of synthesis are coded as E, F, G, I, J, L and M.[1]



Ion	m/z	Molecular Formula	Error (ppm)
$[M+H]^+$	1026.5065	$C_{47}H_{76}N_7O_{16}S^+$	0.1
$[M-FA+H]^+$	786.2613	$C_{31}H_{44}N_7O_{15}S^+$	0.3
$[M-FA-ethanolamine+H]^+$	725.2580	$C_{29}H_{37}N_6O_{14}S^+$	0.5
$[M-FA-ethanolamine-\beta-OH-Asp+H]^+$	594.1889	$C_{25}H_{32}N_5O_{10}S^+$	0.1
$[M-FA-ethanolamine-\beta-OH-Asp-Ser+H]^+$	507.1543	$C_{22}H_{27}N_4O_8S^+$	1.4
$[M-FA-ethanolamine-\beta-OH-Asp-Ser-\beta-OH-Asp+H]^+$	376.1325	$C_{18}H_{22}N_3O_4S^+$	0.2
$[M-FA-ethanolamine-\beta-OH-Asp-Ser-\beta-OH-Asp-Gly+H]^+$	319.1107	$C_{16}H_{19}N_2O_3S^+$	1.1
$[M-FA-ethanolamine-\beta-OH-Asp-Ser-\beta-OH-Asp-Gly-CO+H]^+$	291.1160	$C_{15}H_{19}N_2O_2S^+$	0.6
$[M-FA-ethanolamine-\beta-OH-Asp-Ser-\beta-OH-Asp-Gly-CO-Phe+H]^+$	144.0474	$C_6H_{10}NOS^+$	2.2

Figure 7.12. MS/MS fragmentation spectrum of cyanochelin B and annotation of ions found.[1]

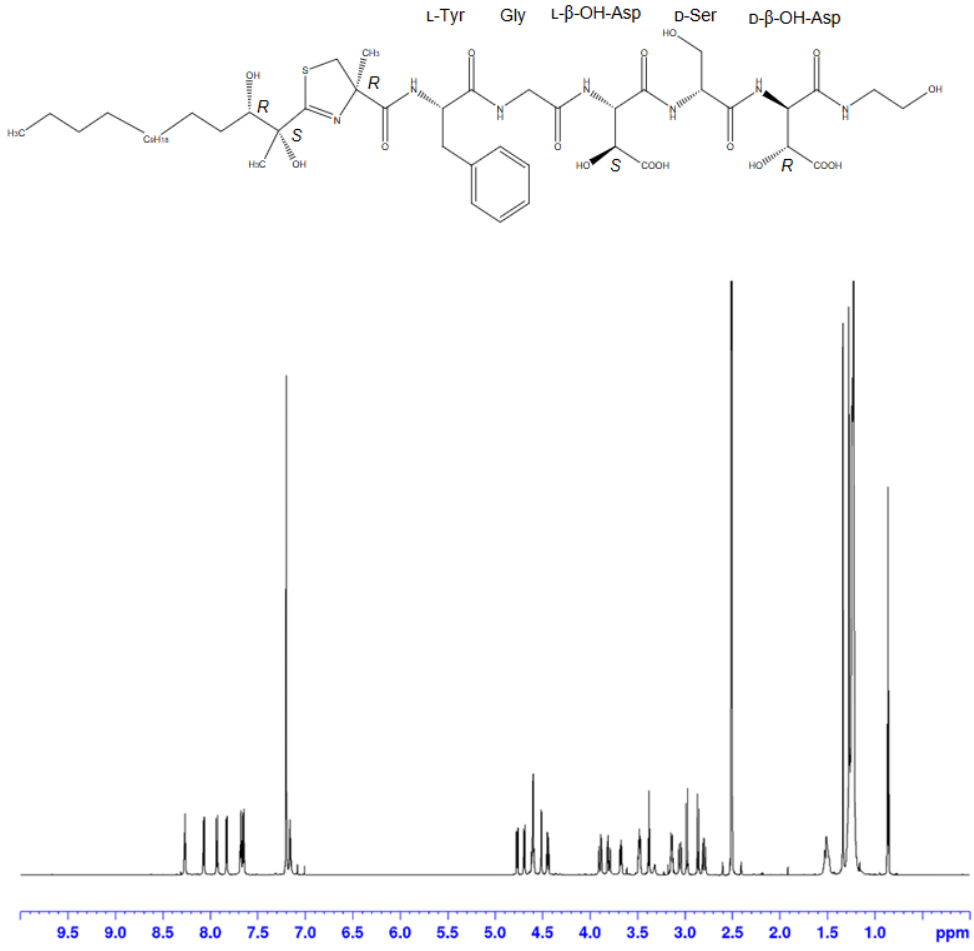


Figure 7.13. ^{1}H-NMR spectrum of cyanochelin B (700 MHz, DMSO-d6, 308K).[1]

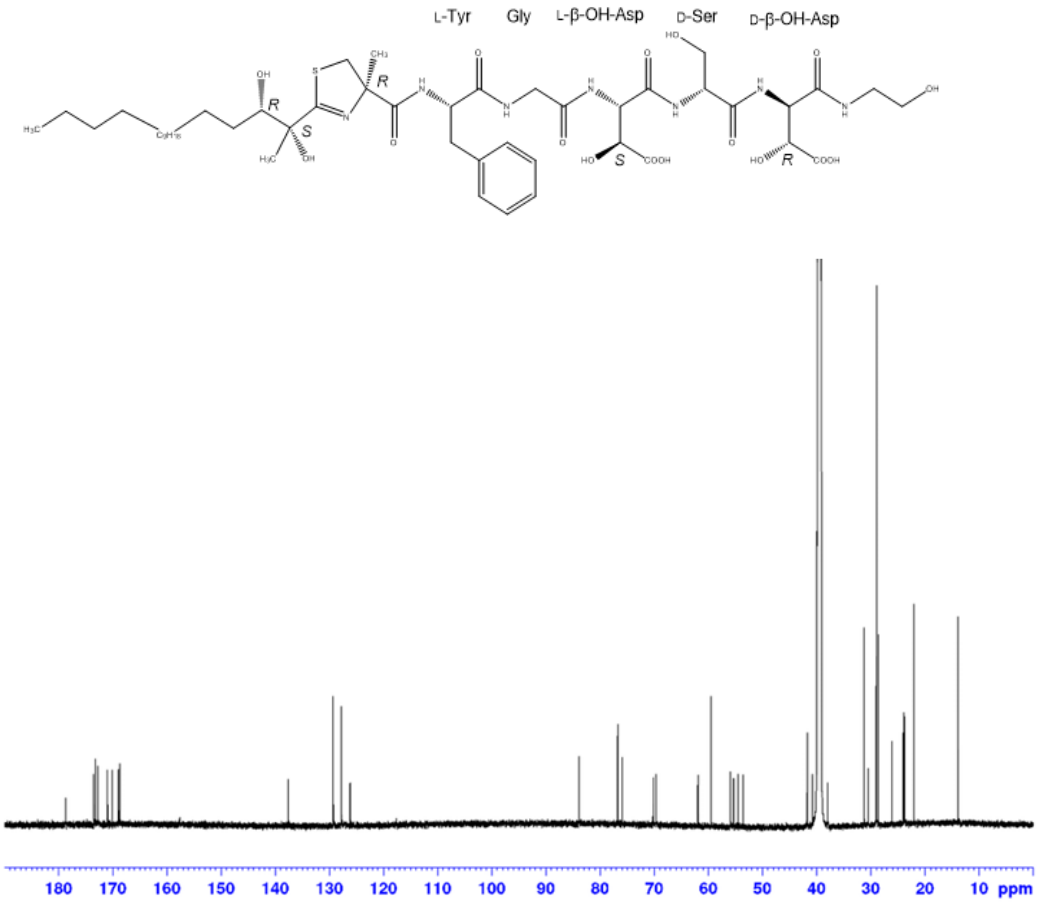


Figure 7.14. ^{13}C-NMR spectrum of cyanochelin B (700 MHz, DMSO-d6, 308K).[1]

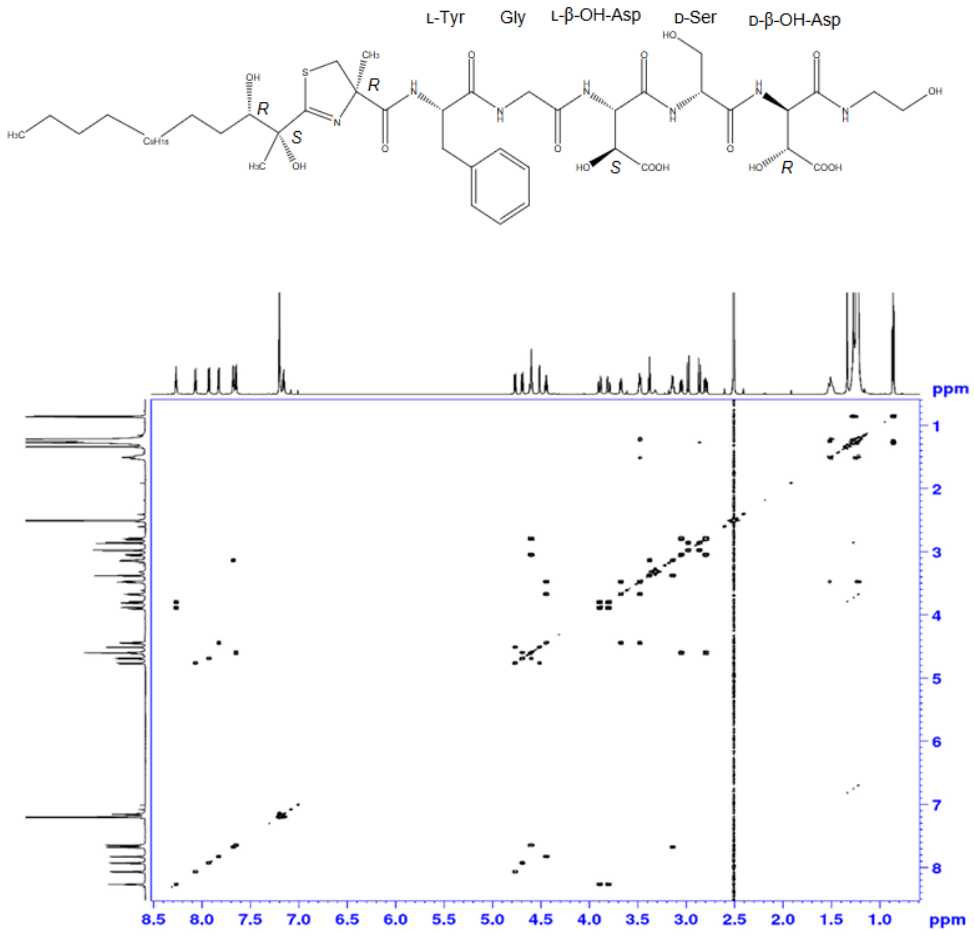


Figure 7.15. COSY spectrum of cyanochelin B (700 MHz, DMSO-d6, 308K).[1]

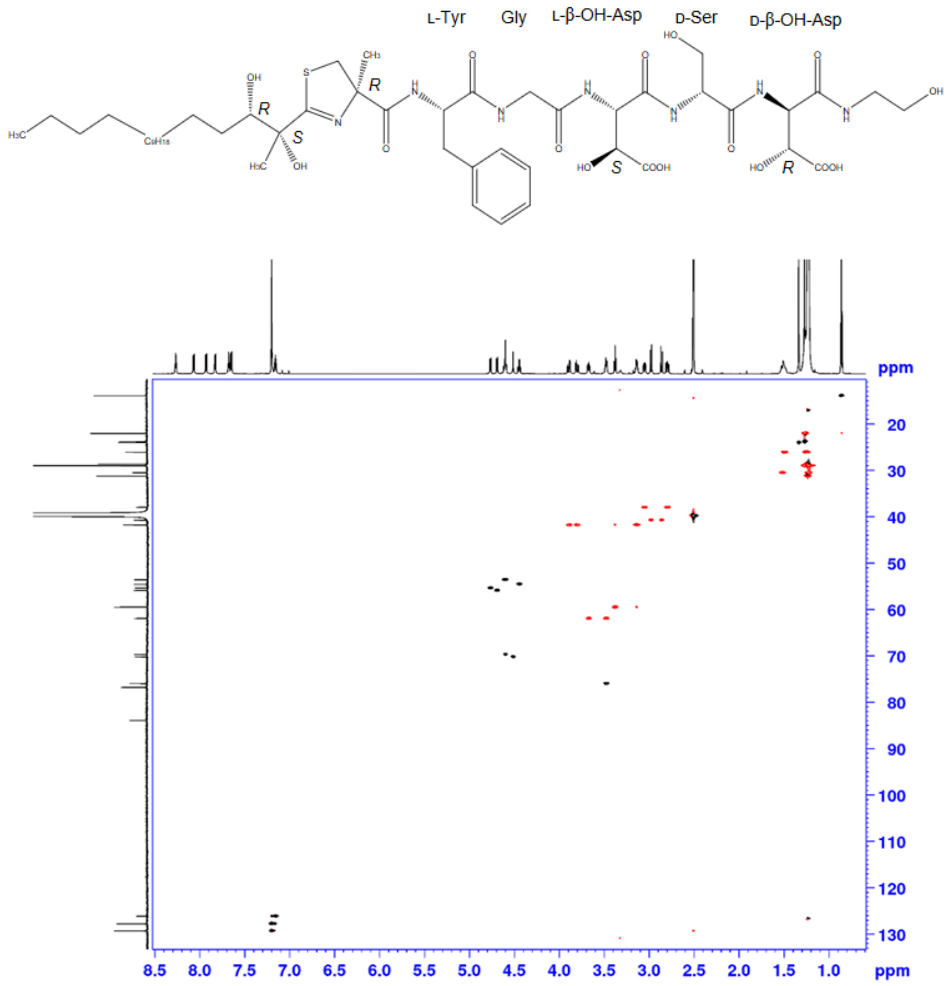


Figure 7.16. HSQC spectrum of cyanochelin B (700 MHz, DMSO-d6, 308K).[1]

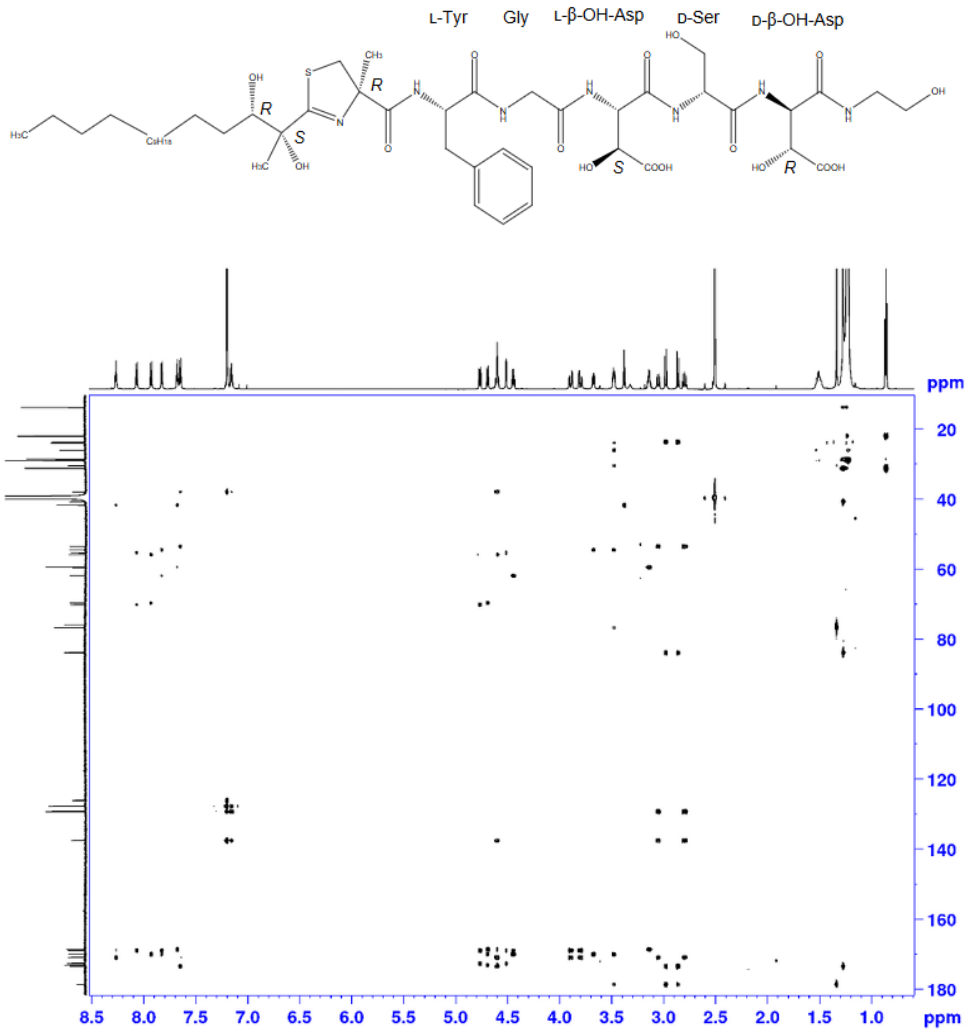


Figure 7.17. HMBC spectrum of cyanochelin B (700 MHz, DMSO-d6, 308K).[1]

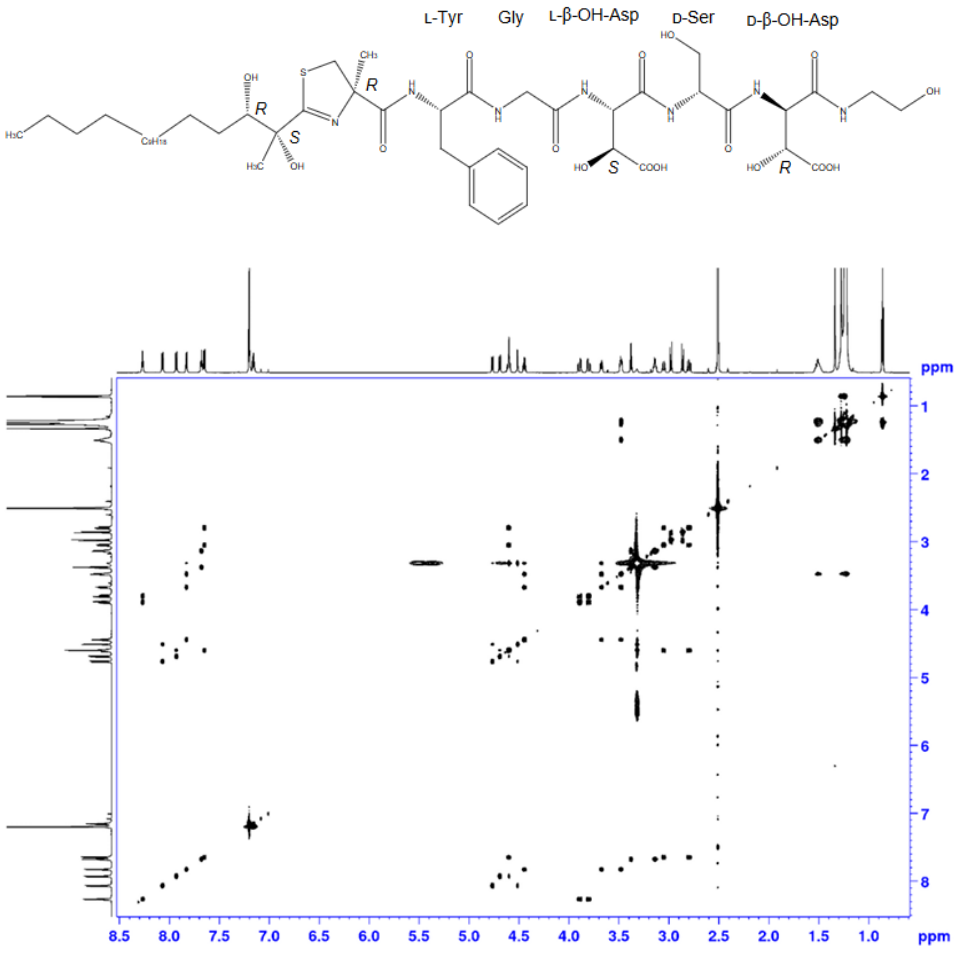


Figure 7.18. TOCSY spectrum of cyanochelin B (700 MHz, DMSO-d6, 308K).[1]

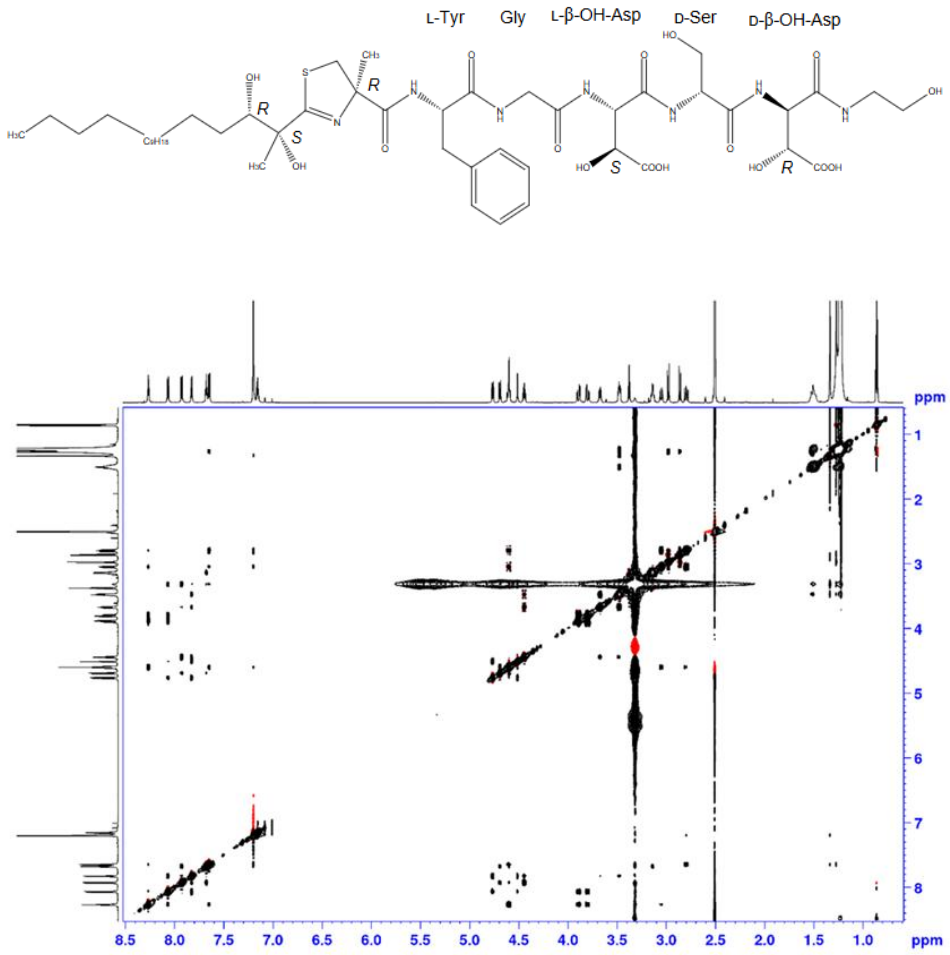


Figure 7.19. NOESY spectrum of cyanochelin B (700 MHz, DMSO-d6, 308K).[1]

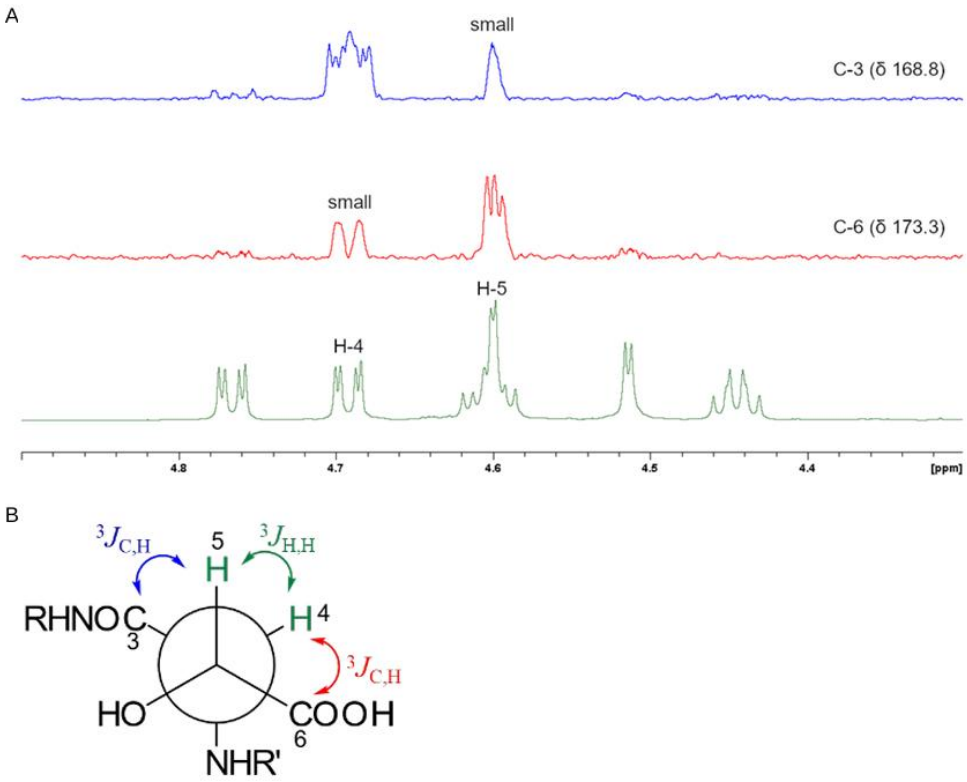


Figure 7.20. (A) One-dimensional sections of the HMBC spectrum of cyanochelin B at δ 168.8 (C-3, blue trace) and δ 173.3 (C-6, red trace), compared with the standard ^{1}H NMR spectrum (green trace). The HMBC spectrum was not 13C decoupled, so any $J_{C,H}$ would cause additional splitting of the relevant signal. The absence of any remarkable additional splitting showed that the $^3J_{C,H}$ between H-5 and C-3 and that between H-4 and C-6 (red trace) are both small. (B) Together with the small $^3J_{H,H}$ between H-4 and H-5 (Table 7.1), these data showed gauche relationship between C-3 and H-5, H-4 and H-5, and H-4 and C-6, unequivocally determining the relative configuration of β-OH-Asp^6 as shown in the picture, i.e. *threo*.[1]

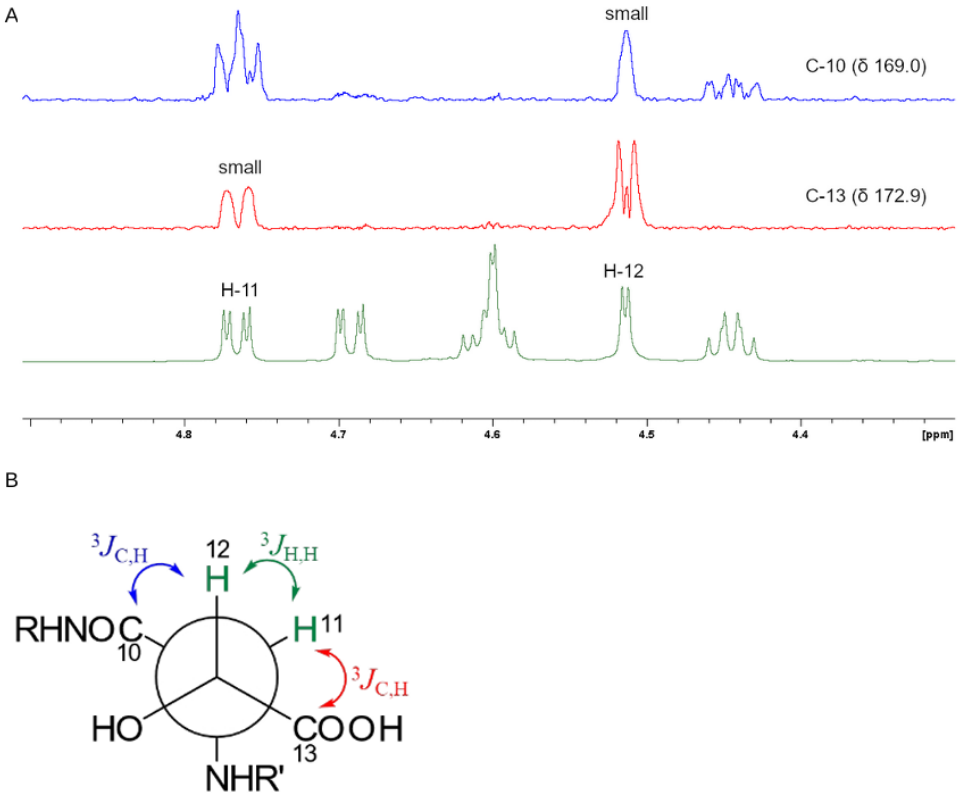


Figure 7.21 (A) One-dimensional sections of the HMBC spectrum of cyanochelin B at δ 169.0 (C-10, blue trace) and δ 172.9 (C-13, red trace), compared with the standard ^{1}H NMR spectrum (green trace). The HMBC spectrum was not ^{13}C decoupled, so any $J_{C,H}$ would cause additional splitting of the relevant signal. The absence of any remarkable additional splitting shows that the $^3J_{C,H}$ between H-12 and C-10 and that between H-11 and C-13 (red trace) are both small. (B) Together with the small $^3J_{H,H}$ between H-11 and H-12 (Table 7.1), these data showed gauche relationship between C-10 and H-12, H-10 and H-11, and H-10 and C-13, unequivocally determining the relative configuration of β-OH-Asp[4] as shown in the picture, i.e. *threo*.[1]

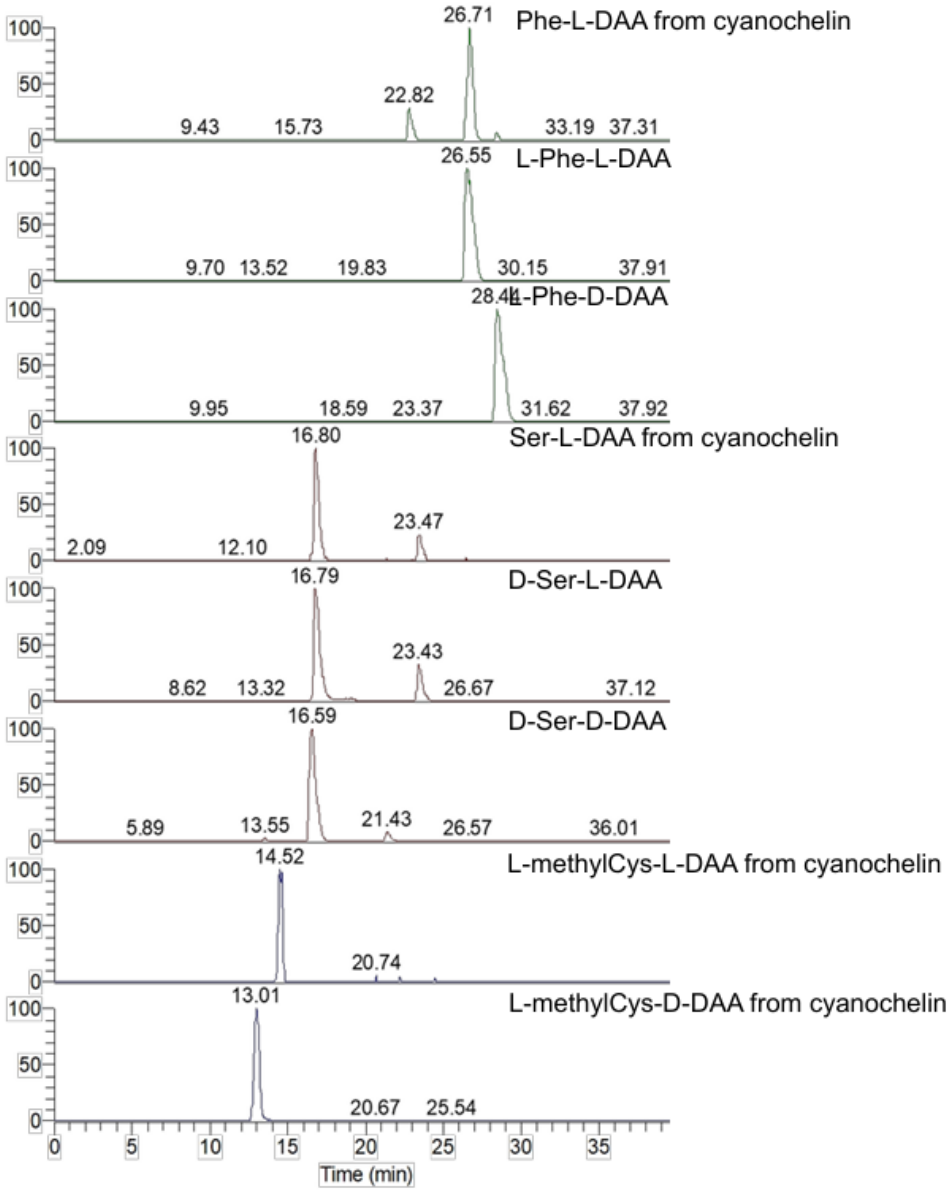


Figure 7.22 LC-MS-enhanced Marfey's analysis of cyanochelin B. Extracted-ion chromatograms at *m/z* 418.1357 of the L-FDAA derivative from the hydrolysis of cyanochelin B and of D-and L-FDAA derivatives of L-Tyr (green trace). Extracted-ion chromatograms at *m/z* 358.0993 of the L-FDAA derivative from the hydrolysis of cyanochelin B and of D-and L-FDAA derivatives of D-Ser (red trace). Extracted-ion chromatograms at *m/z* 388.0921 of the D- and L-FDAA derivatives from the hydrolysis (blu trace).[1]

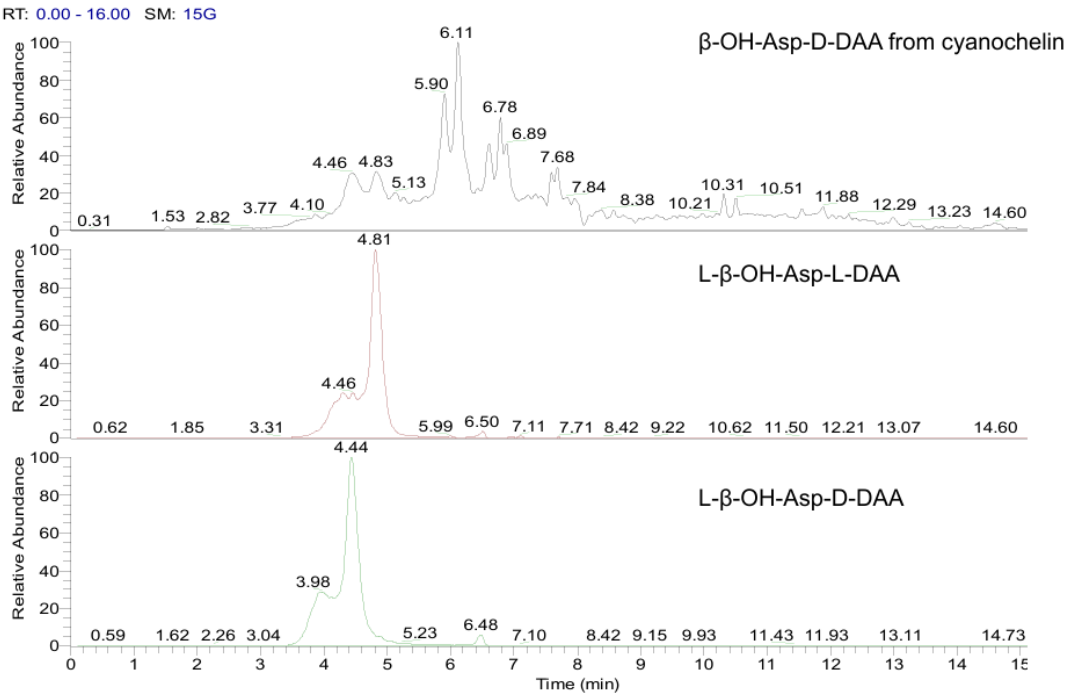


Figure 7.23. LC-MS SRM-enhanced Marfey's analysis of cyanochelin B. In SRM, mass analyzers are set to a selected mass-to-charge ratio, focusing on specific precursor and product ions, the MS method involved two HRMS/MS scan events, following two product ions: *m/z* 356.1-358.1 and *m/z* 280.1-282.1 from the *m/z* precursor 402.2. Extracted-ion chromatograms at *m/z* 356.1-358.1 of the D-FDAA derivative from the hydrolysis of cyanochelin B (black trace) and of D- (green trace) and L-FDAA (red trace) derivatives L-β-OH-Asp.[1]

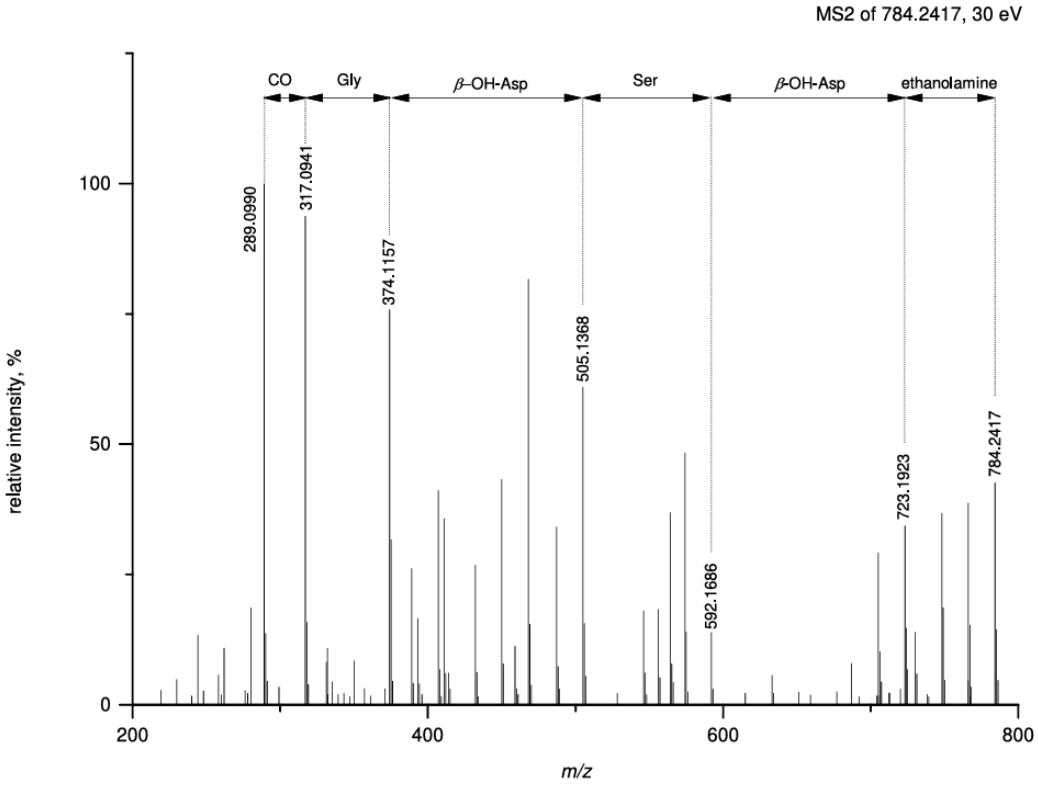


Figure 7.24. Annotated MS/MS fragmentation spectrum of photolytic fragment *m/z* 784.2417 (PF1).[1]

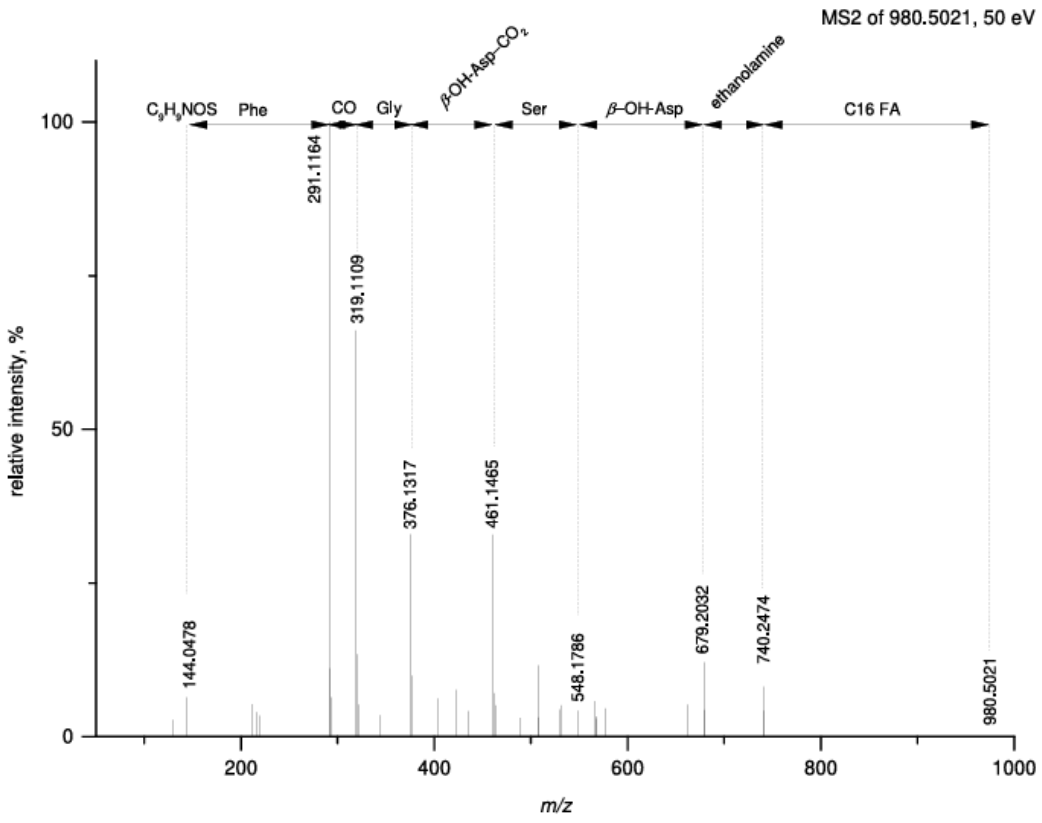


Figure 7.25. Annotated MS/MS fragmentation spectrum of photolytic fragment *m/z* 980.5021 (PF2).[1]

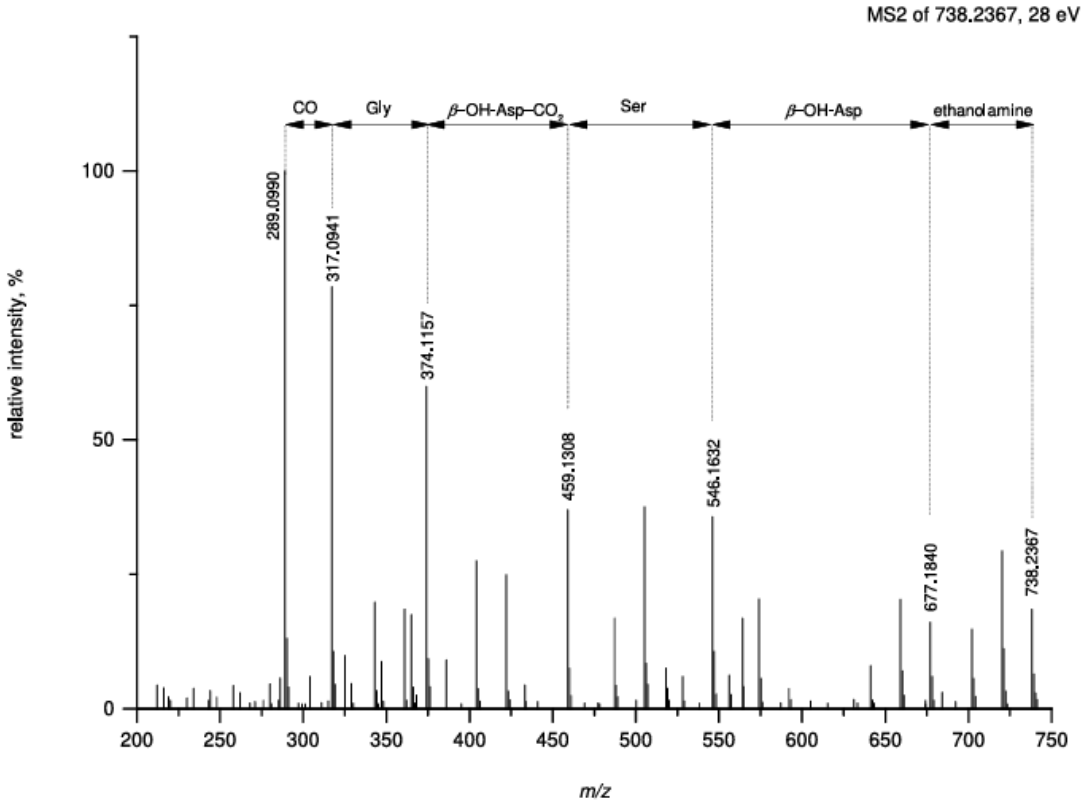


Figure 7.26. Annotated MS/MS fragmentation spectrum of photolytic fragment *m/z* 738.2367 (PF3).[1]

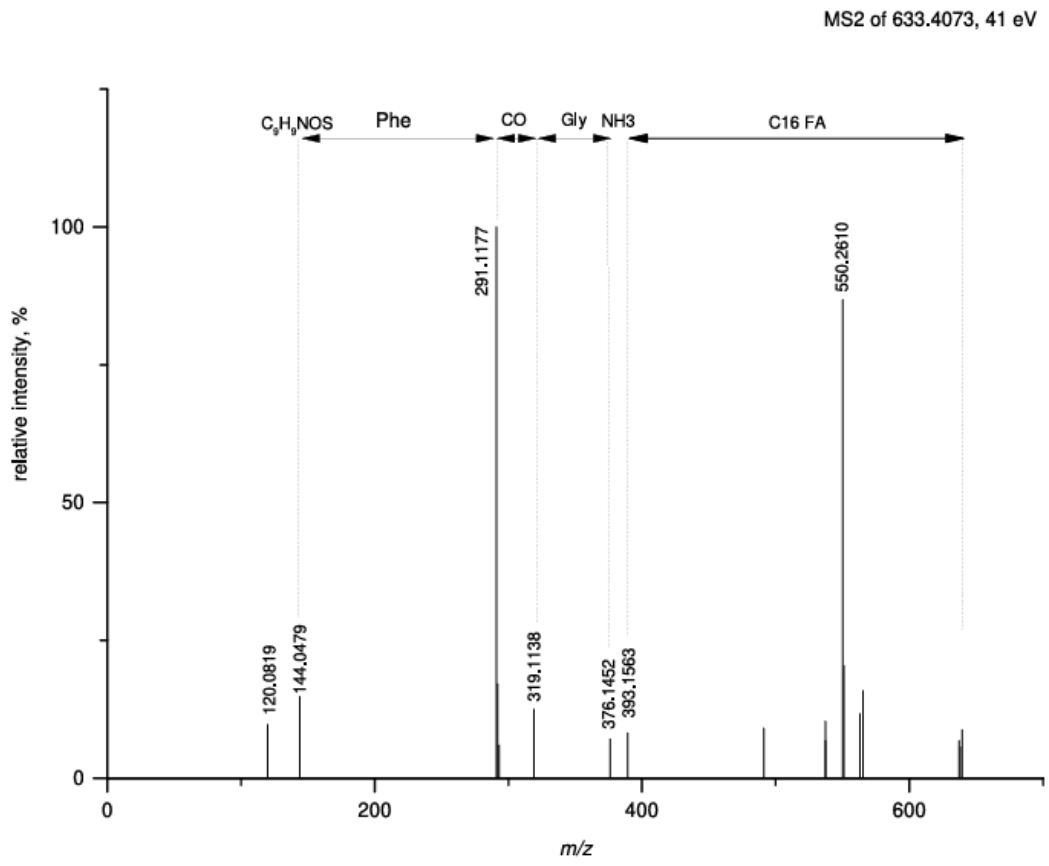


Figure 7.27. Annotated MS/MS fragmentation spectrum of photolytic fragment *m/z* 633.4073 (PF4).[1]

References

1. Falcao, B. P., Di Matteo, V., Hrouzek, P., Štenclová, L., Urajová, P., Mareš, J., et al. Cyanochelin B: Cyanochelin B: a cyanobacterium-produced siderophore with photolytic properties that negate iron monopolization in UV light. *Appl Environ Microbiol*. **2025**; 0:e02566-24.

2. Rao, P. V.; Gupta, N.; Bhaskar, A. S.; Jayaraj, R. Toxins and Bioactive Compounds from Cyanobacteria and Their Implication on Human Health. *J. Environ. Biol*. **2002**, 23, 215–224.

3. Årstøl, E.; Hohmann-Marriott, M. F. Cyanobacterial Siderophores: Physiology, Structure, Biosynthesis, and Applications. *Mar. Drugs* **2019**, 17, 281. doi:10.3390/md17050281.

4. Miethke, M.; Marahiel, M. A. Siderophore-Based Iron Acquisition and Pathogen Control. *Microbiol. Mol. Biol. Rev.* **2007**, 71, 413–451. doi:10.1128/MMBR.00012-07.

5. Zhou, T.; Ma, Y.; Kong, X.; Hider, R. C. Design of Iron Chelators with Therapeutic Application. *Dalton Trans.* **2012**, 41, 6371–6389. doi:10.1039/c2dt12159j.

6. Morrissey, J.; Bowler, C. Iron Utilization in Marine Cyanobacteria and Eukaryotic Algae. *Front. Microbiol.* **2012**, 3, 43. doi:10.3389/fmicb.2012.00043.

7. Galica, T.; Borbone, N.; Mareš, J.; Kust, A.; Caso, A.; Esposito, G.; Saurav, K.; Hájek, J.; Reháková, K.; Urajová, P.; Costantino, V.; Hrouzek, P. Cyanochelins, an Overlooked Class of Widely Distributed Cyanobacterial Siderophores, Discovered by Silent Gene Cluster Awakening. *Appl. Environ. Microbiol.* **2024**, e02566-24. doi:10.1128/aem.02566-24.

8. Nivina, A.; Yuet, K. P.; Hsu, J.; Khosla, C. Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem. Rev.* **2019**, *119*, 12524–12547. doi:10.1021/acs.chemrev.9b00525

9. Galica, T.; Hrouzek, P.; Mareš, J. Genome Mining Reveals High Incidence of Putative Lipopeptide Biosynthesis NRPS/PKS Clusters Containing Fatty Acyl-AMP Ligase Genes in Biofilm-Forming Cyanobacteria. *J. Phycol.* **2017**, 53, 985–998. doi:10.1111/jpy.12555.

10. Molnár, I.; Schupp, T.; Ono, M.; Zirkle, R.; Milnamow, M.; Nowak-Thompson, B.; Engel, N.; Toupet, C.; Stratmann, A.; Cyr, D. D.; Görlach, J.; Mayo, J. M.; Hu, A.; Goff, S.; Schmid, J.; Ligon, J. M. The Biosynthetic Gene Cluster for the Microtubule-Stabilizing Agents Epothilones A and B from *Sorangium cellulosum* So ce90. *Chem. Biol.* **2000**, *7*, 97–109. doi:10.1016/S1074-5521(00)00075-2.

11. Hardy, C. D.; Butler, A. Ambiguity of NRPS Structure Predictions: Four Bidentate Chelating Groups in the Siderophore Pacifibactin. *J. Nat. Prod.* **2019**, 82, 990–997.

12. Matsumori, N.; Kaneno, D.; Murata, M.; Nakamura, H.; Tachibana, K. Stereochemical Determination of Acyclic Structures Based on Carbon-Proton Spin Coupling Constants. A Method of Configuration Analysis for Natural Products. *J. Org. Chem.* **1999**, 64, 866–876.

13. Marfey, P. Determination of D-Amino Acids. II. Use of a Bifunctional Reagent, 1,5-Difluoro-2,4-Dinitrobenzene. *Carlsberg Res. Commun.* **1984**, 49, 591–596.

14. Reitz, Z. L.; Hardy, C. D.; Suk, J.; Bouvet, J.; Butler, A. Genomic Analysis of Siderophore β-Hydroxylases Reveals Divergent Stereocontrol and Expands the Condensation Domain Family. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, 116, 19805–19814

15. Carmeli, S.; Paik, S.; Moore, R. E.; Patterson, G. M. L.; Yoshida, W. Y. Revised Structures and Biosynthetic Studies of Tantazoles A and B. *Tetrahedron Lett.* **1993**, 34, 6681–6684. doi:10.1016/S0040-4039(00)61674-X.

16. Keatinge-Clay, A. T. A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. *Chem. Biol.* **2007**, 14, 898–908. doi:10.1016/j.chembiol.2007.07.009.

17. Butler, A.; Harder, T.; Ostrowski, A. D.; Carrano, C. J. Photoactive Siderophores: Structure, Function and Biology. *J. Inorg. Biochem.* **2021**, 221, 111457.

18. Kreutzer, M. F.; Kage, H.; Nett, M. Structure and Biosynthetic Assembly of Cupriachelin, a Photoreactive Siderophore from the Bioplastic Producer Cupriavidus necator H16. *J. Am. Chem. Soc.* **2012**, 134, 5415–5426.

19. Chatzimpaloglou, A.; Yavropoulou, M. P.; Rooij, K. E.; Biedermann, R.; Mueller, U.; Kaskel, S.; Sarli, V. Total Synthesis and Biological Activity of the Proposed Structure of Phaeosphaeride A. *J. Org. Chem.* **2012**, 77, 9659–9667. doi:10.1021/jo301662e.

20. Chatzimpaloglou, A.; Kolosov, M.; Eckols, T. K.; Tweardy, D. J.; Sarli, V. Synthetic and Biological Studies of Phaeosphaerides. *J. Org. Chem.* **2014**, 79, 4043–4054. doi:10.1021/jo500545d.

21. Hirose, Y.; Fujisawa, T.; Ohtsubo, Y.; Katayama, M.; Misawa, N.; Wakazuki, S.; Shimura, Y.; Nakamura, Y; Kawachi, M.; Yoshikawa, H.; Eki, T.; Kanesaki, Y. Complete Genome Sequence of Cyanobacterium *Leptolyngbya* sp. NIES-3755. *Genome Announc.* **2016**, 4, e00090-16. doi:10.1128/genomeA.00090-16.

22. Strunecký, O.; Ivanova, A. P.; Mareš, J. An Updated Classification of Cyanobacterial Orders and Families Based on Phylogenomic and Polyphasic Analysis. *J. Phycol.* **2023**, 59, 12–51.

23. Johansen, J. R.; Kováčik, L.; Casamatta, D. A.; Iková, K. F.; Kaštovský, J. Utility of 16S-23S ITS Sequence and Secondary Structure for Recognition of Intrageneric and Intergeneric Limits within Cyanobacterial Taxa:

Leptolyngbya corticola sp. nov. (Pseudanabaenaceae, Cyanobacteria). *Nova Hedwigia* **2011**, 92, 283–302. doi:10.1127/0029-5035/2011/0092-0283.

24. Štenclová, L.; Wilde, S. B.; Schwark, M.; Cullen, J. L.; McWhorter, S. A.; Niedermeyer, T. H. J.; Henderson, W. M.; Mareš, J. Occurrence of Aetokthonotoxin Producer in Natural Samples. A PCR Protocol for Easy Detection. *Harmful Algae* **2023**, *125*, 102425. doi:10.1016/j.hal.2023.102425.

25. Krakau, S.; Straub, D.; Gourlé, H.; Gabernet, G.; Nahnsen, S. nf-core/mag: A Best-Practice Pipeline for Metagenome Hybrid Assembly and Binning. *NAR Genom. Bioinform*. **2022**, 4, lqac007. doi:10.1093/nargab/lqac007.

26. Wilmotte, A.; Demonceau, C.; Goffart, A.; Hecq, J. H.; Demoulin, V.; Crossley, A. C. Molecular and Pigment Studies of the Picophytoplankton in a Region of the Southern Ocean (42–54° S, 141–144° E) in March 1998. Deep-Sea Res. II **2002**, 49, 3351–3363. doi:10.1016/S0967-0645(02)00087-5.

27. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, 38, 3022–3027. doi:10.1093/molbev/msab120.

28. Edgar, R. C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* **2004**, 32, 1792–1797. doi:10.1093/nar/gkh340.

29. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D. L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M. A.; Huelsenbeck, J. P. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space. *Syst. Biol.* **2012**, 61, 539–542. doi:10.1093/sysbio/sys029.

30. Guindon, S.; Dufayard, J.-F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **2010**, 59, 307–321. doi:10.1093/sysbio/syq010.

Chapter 8

Bio-tracking, bio-monitoring and bio-magnification interdisciplinary studies to assess cyanobacterial harmful algal blooms (cyanoHABs)' impact in complex coastal systems

8.1 Synopsis of this chapter [1]

Cyanobacteria thrive from polar salt marshes to tropical lagoons, adapting quickly to light, salinity and nutrient loadings. When nutrient-rich conditions tip the balance, these ancient microbes can erupt into expansive surface blooms whose pigments stain the water and whose secondary metabolites, cyanotoxins, threaten ecosystems, aquaculture and public health. Because the organisms respond so visibly and so fast, they are powerful natural indicators of changing coastal conditions.

To turn that ecological warning system into a management tool, in collaboration with the research group of Prof. Massimiliano Lega from University of Naples Pathenope, we developed the Fast Detection Strategy (FDS), a multidisciplinary workflow that marries satellite and drone imagery with high-resolution mass-spectrometry metabolomics. In practice the method works like this: space-borne and proximal sensors flag unusual optical signatures on the water; targeted field sampling follows within hours; and molecular-networking algorithms rapidly screen extracts for cyanobacterial chemotypes, all without lengthy chromatographic

purification. The result is a same-week picture of "who is blooming, where, and with which toxins," delivered in language that local agencies can act on.

The approach proved its worth during the crimson bloom of *Planktothrix rubescens* in Lake Avernus (Naples). Remote sensing traced the bloom's downstream journey through a short emissary channel to nearby mussel farms, while mass-spectrometric networking revealed the presence of anabaenopeptins, a class of hepatotoxic peptides now attracting international attention. Early warning allowed authorities to issue precautionary harvest closures before the toxins entered the market.[1,2]

Our experience shows that pairing wide-area vision with fast laboratory tools turns cyanobacteria into practical sentinels of water quality. Timely alerts give local authorities room to act, closing shellfish beds or advising swimmers, before toxins pose a significant risk. Continued, integrated monitoring will therefore be essential to safeguard both marine ecosystems and human health as nutrient pressures and climate change intensify.

8.2 Environmental triggers, bioaccumulation, and ecotoxicology of cyanobacterial toxins

Harmful cyanobacterial blooms (cyanoHAB) occur when environmental conditions allow cyanobacteria to outcompete other plankton, particularly under eutrophic conditions, where high levels of phosphorus and nitrogen promote intense growth, and are further favoured by warm temperatures, high solar radiation, water column stability, and slow circulation, which together favour the presence of floating taxa near the surface; climate change is increasing the frequency of such warm and stratified periods. In addition to promoting biomass, stress factors such as nitrogen limitation, iron stress, or

intense light can modulate the expression of toxin genes, although regulation is highly species- and context-specific. Anthropogenic factors (agricultural runoff, wastewater, urban effluents) amplify these risks by providing nutrients that trigger and sustain blooms. Once formed, cyanotoxins released by spill or cell lysis can persist and interact with food webs: relatively stable toxins such as microcystins accumulate in zooplankton, in fish and especially in filter-feeding bivalve molluscs (e.g. mussels),[3] creating exposure vectors for humans and raising concerns about bioaccumulation and, in some cases, trophic transfer; while true biomagnification remains controversial and probably toxin-dependent, passage to higher trophic levels has been documented. Ecological consequences include acute fish mortality (liver lesions from microcystins, neurotoxicity from anatoxins/saxitoxins), suppression of grazers that contribute to prolonging blooms, and sublethal effects that impair growth, reproduction, and behaviour in several species; terrestrial wildlife, livestock, and domestic animals are also at risk when they access contaminated water.[4] Human exposure occurs through drinking water, recreational activities, and consumption of contaminated fish, shellfish, or algal products, with outcomes ranging from acute liver failure or respiratory paralysis to potential chronic effects (e.g., liver disease and carcinogenicity) with low-dose exposure over long periods.[5] Overall, the dynamics of cyanotoxins reflect the interaction between nutrient enrichment, climatic conditions, and physiological stress, highlighting the need for vigilant monitoring, early warning systems, and nutrient management in watersheds to protect the ecosystem and public health.[6]

8.3 Regulatory efforts to monitor cyanotoxins in water and food

Growing evidence of the risks associated with cyanotoxins has led to the development of various guidelines and regulations aimed at protecting human health. At the international level, the World Health Organization (WHO) has been at the forefront of establishing health-based guidelines. The WHO interim guideline for microcystin-LR in drinking water is 1 µg/L, a level intended to protect even lifelong consumption. This is supplemented by a provisional tolerable daily intake (TDI) of 0.04 µg of microcystin-LR per kg of body weight per day. Many countries have adopted or referred to these values to establish their own standards. For example, the US Environmental Protection Agency (EPA) has issued health advisories for microcystins in drinking water and recommended thresholds for recreational water in the order of a few micrograms per liter (with specific limits varying by state and age group).[7] In the European Union (EU), microcystin-LR is included in the list of priority substances to be monitored. EU Directive 2013/39/EU recommends that microcystin concentrations do not exceed 0.5 µg/L in surface water used for drinking water abstraction. This value essentially serves as an early warning limit for water services. The EU also maintains strict standards for marine biotoxins in shellfish (e.g., a saxitoxin limit of 800 µg/kg in shellfish meat) due to the risk of algal toxins entering the food chain.[8] However, currently most cyanotoxins, with the exception of saxitoxin, are not explicitly regulated in food at the EU level. National regulations can vary considerably. Italy, for example, has issued specific guidelines for recreational waters: a 2010 decree by the Ministry of Health set alert thresholds for blooms at 100,000 cyanobacterial cells/mL or 25 µg/L of

microcystin in bathing waters. If these levels are exceeded or if cyanobacteria producing other toxins are present, authorities are required to carry out risk assessments on a case-by-case basis, in consultation with national health institutes. It should be noted that Italian regulations (like many others) do not yet set numerical limits for cyanotoxin contamination in fish or shellfish, even though this is recognized as a potential risk factor. Australia and some US states have gone further, setting provisional limits for certain toxins in food (e.g., the Australian guideline of 50 µg/kg for total microcystins in shellfish or the Oregon limit of 1 µg/kg for microcystin in food supplements). Overall, the regulatory landscape is still evolving. The lack of universally adopted food safety standards for freshwater cyanotoxins highlights a gap that researchers and policymakers are working to fill. What is clear is that routine monitoring programs for cyanotoxins in water have become more common in many countries, particularly for drinking water sources and recreational lakes. Agencies often take a multi-tiered approach, with rapid field tests or satellite observations triggering detailed chemical analyses in cases of suspected blooms. Continued refinement of guidelines by the WHO, EU, and national authorities, based on toxicological research, is critical to managing the risks of cyanotoxins in both water and food.

8.4 Lake Avernus 2022 Red Bloom

Lake Avernus, a small volcanic lake located in the Phlegraean Fields near Naples, has long been immortalized as the “Gateway to Hades,” a reputation born of unusual natural phenomena, from harmful volcanic gases that once killed birds in flight to sudden mass deaths. In recent decades, the lake has

also shown episodic and dramatic changes in water colour, now interpreted as the visible imprint of recurring cyanobacterial blooms. These events result from the interaction between natural nutrient inputs linked to the lake's volcanic environment and anthropogenic enrichment from the surrounding watershed. Historical observations indicate an alternation in the dominance of different cyanobacterial taxa and toxic profiles, with a notable episode of red coloration documented in 2007, highlighting both the rarity and periodic recurrence of such events. In March 2022, the lake turned a dull red color again, prompting an in-depth investigation to capture the onset of the bloom, its biological identity, and toxicological signature.[2]

We applied a rapid detection strategy (FDS), in collaboration with Prof. Massimiliano Lega from University of Naples Pathenope, a coordinated process that integrates satellite and drone remote sensing, targeted in situ sampling, and advanced laboratory analysis (molecular biology and metabolomics). Multi-year observations from Sentinel-2 flagged the anomaly via spectral indices and were quickly confirmed by high-resolution UAV imagery, which mapped the surface extent and heterogeneity of the event.[9] Microscopy and 16S rRNA sequencing identified the organism responsible for the bloom as *Planktothrix rubescens*, a filamentous, red-pigmented cyanobacterium that thrives in stratified lakes (Figure 8.5). High-resolution LC-HRMS/MS, coupled with feature-based molecular networking, revealed a complex assembly of cyanotoxins dominated by anabaenopeptins: fourteen variants were detected, including seven known congeners and seven recently reported analogues (Figure 8.6). Surprisingly, anabaenopeptin B accounted

for approximately 4% of the extractable metabolome, indicating an unusually high toxin load.[2]

Beyond characterizing a single episode, the study positions Lake Avernus as an informative model for understanding the dynamics of harmful algal blooms (cyanoHAB) in small stratified volcanic lakes subject to mixed natural and anthropogenic pressures. The lake's geochemistry and meromictic structure can trap and periodically release nutrients, while runoff from agriculture and low-density urban areas likely increases baseline eutrophication, together creating windows of opportunity for toxigenic blooms. Hydrological connectivity with nearby coastal waters and shellfish aquaculture areas further increases the importance of rapid detection and risk communication. Methodologically, the Avernus case demonstrates how combining synoptic satellite surveillance with proximal detection and untargeted metabolomics can provide early warning, organism attribution, and toxin profiling on actionable time scales, an approach that is widely transferable to watersheds, lakes, and coastal systems facing similar threats.[2] In this thesis, the 2022 Avernus bloom serves both as a detailed case study, linking environmental triggers to toxic outcomes, and as a model for scalable, interdisciplinary monitoring and management of cyanobacterial risks.[11,12,13,14]

8.5 Remote and proximal sensing analysis

An integrated remote-proximal sensing workflow captured both the onset and the entire footprint from lake to coast of the cyanobacterial bloom in Lake Avernus in March 2022. Multi-year Sentinel-2 time series (2018-2022; multispectral 10 m) first signalled spring 2022 as anomalous: NDVI deviated sharply from baseline between late March and April, and NDTI, sensitive to

turbidity, increased in step with the observed reddening. True-colour images from March 31 showed a reddish-brown hue throughout the basin, but lacked local-scale detail due to the size of the lake and the optics, which led to a UAV campaign on April 5 (DJI Mavic 2 Pro; high-resolution RGB plus multispectral-orange, cyan, near-infrared) in a programmed grid under clear skies and synchronized with the Sentinel-2 overflight for cross-calibration. Hundreds of georeferenced frames were processed using motion structure (Agisoft Metashape) into a centimeter-scale orthomosaic that mapped red and turbid waters throughout the basin, revealing a bloom throughout the lake with localized intensifications consistent with metalimnetic populations of *Planktothrix rubescens* dispersing toward the surface during mixing.

Drone-derived reflectance was then used to calibrate satellite indices, yielding a close match between NDTI peak and areas of deeper surface reddening and validating NDTI as a robust proxy for blooming. Extending the analysis beyond the lake, we combined NDVI/NDTI with customized/enhanced indices in the red band tuned to the phycoerythrin-rich signal to delineate a continuous source-path-destination corridor from Avernus, along the emissary channel, to the bay of Pozzuoli: the coastal impact zone extended approximately 300 m offshore (the first 200 m visible in RGB, the outer 100 m resolved only through multispectral classification) toward sensitive receptors (swimming beaches and a mussel farm approximately 1 km offshore). The DJI Mavic 3 missions equipped with RTK (RGB, thermal, multispectral; MAPIR OCN/RGN payloads) over the channel mouth and coastal targets fixed the spectral signature in situ for subsequent satellite calibration and produced orthomosaics/DSMs (PIX4D→QGIS) that

identified the discharge, plume axis, and likely marine exposure areas. Together, satellite data and data collected by UAVs provided early warning, temporal context, and fine-scale spatial mapping for an unusually intense event, absent in the 2019-2021 period, providing direct information for targeted in situ sampling, downstream biomonitoring, and risk communication for aquaculture and recreational waters.

8.6 Study Area and Sample Collection

The main study area extended from Lake Averno (in Italian: Lago d'Averno), a volcanic lake (approximately 0.5 km² in area) in the Campi Flegrei district, to the adjacent coastal waters of the Bay of Pozzuoli in the Tyrrhenian Sea. Lake Averno is separated from the sea by a narrow strip of land, through which flows a small outlet channel connecting the lake's outflow to the sea (historically modified as a drainage channel). The coastal area in question is part of the Campi Flegrei marine zone, characterized by semi-enclosed bay conditions and a mix of natural and aquaculture sites. In particular, in these coastal waters, about 1 km off the mouth of the lake channel and close to several local bathing beaches, there is a mussel farm. This scenario is unique in that it allows freshwater cyanobacteria to have a direct impact on the marine ecosystem. In a previous study on the same red bloom event, my research group also analyzed the lake waters at the bloom site and identified the species responsible for the bloom as *Planktothrix rubescens* through microscopy and 16S rRNA sequencing, further confirmed by phylogenetic analysis (neighbor-joining) (Figure 8.5).

Following remote sensing indications that the 2022 cyanobacterial bloom had moved from Lake Averno to the sea, we conducted field sampling of Mediterranean mussels (*Mytilus galloprovincialis*) to assess toxin uptake.

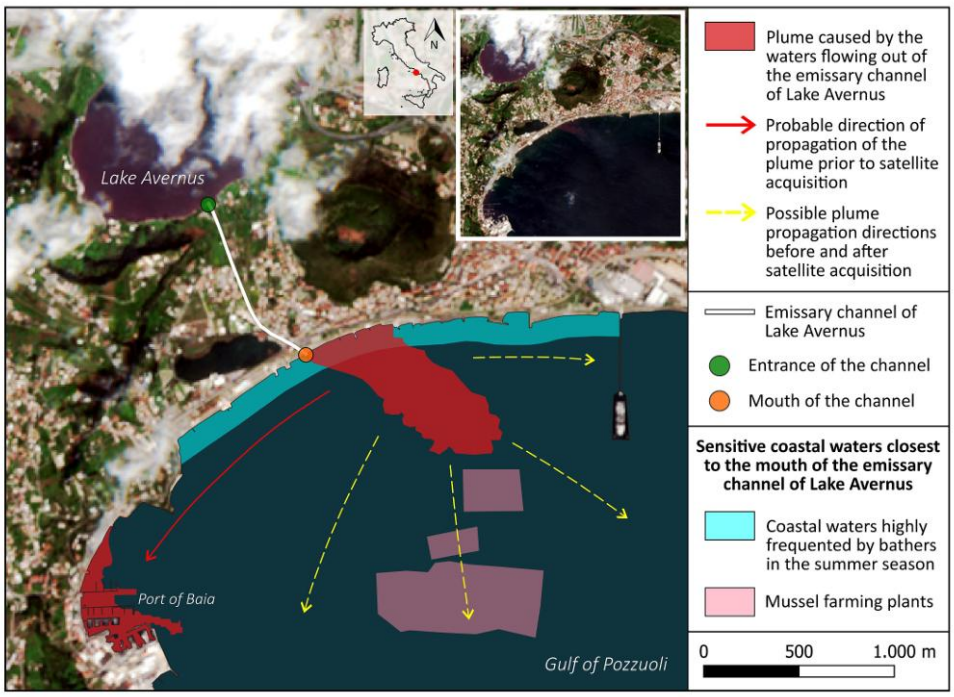


Figure 8.1 Simplified illustration of the phenomenon, showing the areas affected by the bloom, the regions of origin and destination, and the directions of advance of the flow front.[1]

The mussels were collected on April 21, 2022, from the fish farm off the coast of Pozzuoli (Campi Flegrei), very close to the Averno outlet terminal. This date was about three weeks after the peak of the bloom in the lake (which occurred at the end of March 2022), allowing the bloom water to mix with the coastal environment and the mussels to accumulate any toxins.

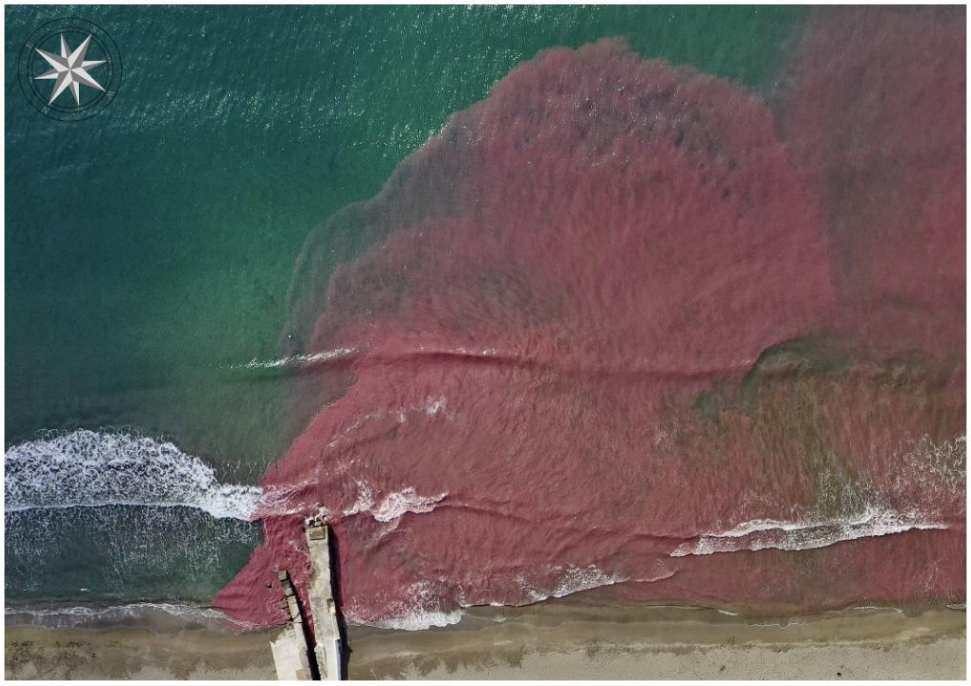


Figure 8.2 A detailed, high-resolution view captured by the drone shows the plume formed along the coast by the inflow of red-colored water with a high concentration of cyanobacteria.[1]

Divers or farm staff collected a sample of approximately 2 kg of mussels, focusing on those molluscs that were in the direct line of the outflow or in areas where discoloured (reddish) water had been observed. The sample size of 2 kg was chosen to provide sufficient biomass for repeated extractions and analyses.

Immediately after collection, the live mussels were stored in a cool, dark container with wet towels/seaweed (to prevent dehydration) and then transferred to a refrigerator (4°C) in the laboratory within a few hours. Maintaining the cold chain is important to slow down post-mortem degradation of toxins or metabolites. Before chemical processing, a portion of the mussels was set aside for microscopic examination. An optical microscope (Motic Panthera C2, 40-1000× magnification) was used to inspect smears of mussel tissue and fluids for the presence of cyanobacteria.

In practice, this involved opening some mussels and taking a drop of their interstitial fluid or a little intestinal content, then checking for characteristic cyanobacterial cells (such as filaments of *P. rubescens*) or unusual microscopic anatomy (e.g., tissue inflammation that could indicate toxin stress). Any such observations would have qualitatively supported the chemical results.

After these preliminary checks, the mussels were prepared for extraction. Each mussel was rinsed with distilled water to remove salt and particles from the surface, then shelled by cutting the adductor muscles and opening the shells. The soft tissue (edible portion) of all specimens was combined in a stainless-steel blender. The collected tissue was thoroughly homogenized for several minutes until a homogeneous suspension was obtained. Homogenization ensures that any toxins are evenly distributed throughout the sample matrix, allowing the subsamples to be representative.

8.7 Extraction of Mussel Tissues

The homogenized mussel tissue was extracted using a multi-step workflow driven by solvent polarity to isolate cyanotoxins. First, the suspension was immersed in 100% methanol and shaken vigorously, repeating the methanol extraction a second time with fresh solvent to maximize the recovery of polar metabolites. Methanol effectively solubilizes peptide toxins such as anabaenopeptins and other hydrophilic compounds, while precipitating proteins and other macromolecules that could interfere with downstream analysis. The residual pellet was then treated with a 1:1 mixture of methanol/chloroform to treat the intermediates: this co-solvent system

dissolves a broad spectrum of medium-polarity constituents because chloroform, a non-polar solvent, extracts the lipophilic fractions while methanol maintains the solubility of the polar ends in amphiphilic molecules. Finally, to identify highly hydrophobic components, the remaining solid was extracted twice with 100% chloroform, capturing very non-polar substances such as lipids, carotenoids, or any extremely hydrophobic toxin analogues.

After each extraction, the liquid phase was clarified in two steps: a coarse sieve removed large particles (e.g., connective tissue and shell fragments), and the filtrate was then passed under vacuum through a celite bed to remove fine debris and obtain a clear solution. The filtrates obtained from repeated cycles with the same solvent were collected and concentrated at reduced pressure at moderate temperature on a rotary evaporator to avoid thermal degradation. The dried residues reflected the expected polarity distribution of mussel metabolites: the methanol fractions yielded 164.2 mg, the methanol/chloroform fraction 15.3 mg, and the chloroform fractions 18.7 mg, indicating that most of the extractable organic material, which likely includes the target peptide toxins, resides in the polar pool. These crude extracts necessarily contained both the cyanotoxins of interest and a matrix of co-extracted substances (e.g., amino acids and salts in the polar fraction, lipids and pigments in the nonpolar fraction).

For LC-MS analysis, each dried extract was reconstituted in 100% methanol at a concentration of 5 mg/mL, maintaining solvent consistency with the extraction phase and ensuring compatibility with the methanol/water chromatographic system. The solutions were briefly sonicated to ensure

complete dissolution and, where necessary, passed through a 0.22 μm PTFE syringe filter to remove any residual particles prior to injection.

8.8 LC-MS/MS Analysis and Molecular Networking

We profiled mussel extracts using reverse-phase liquid chromatography coupled with high-resolution tandem mass spectrometry (LC-MS/MS). Using positive electrospray ionization, a peptide gradient separated the constituents, while the mass spectrometer acquired accurate mass full scans interspersed with MS/MS dependent on the most abundant ions. This configuration generated rich fragmentation patterns suitable for the structural elucidation of cyanobacterial peptides.

The raw data sets were processed with MZmine to detect and deconvolute chromatographic features, align them across samples, and remove common artifacts (e.g., adducts and isotopic duplicates). Only features supported by MS/MS spectra were retained for downstream analysis to ensure robust annotation.

To contextualize the chemistry, we applied feature-based molecular networking on the GNPS platform with library searching and visualized the resulting network in Cytoscape. Essentially, we searched for the cluster corresponding to anabaenopeptin toxins. Based on prior knowledge related to lake blooms, anabaenopeptin A, B, F, and oscillamide Y should produce similar peptide fragmentation patterns (often showing a series of ions fragmented from the cleavage of the cyclic peptide ring) (Figure 8.7, 8.8, 8.9, 8.10). Indeed, molecular networking revealed a distinct cluster in which our mussel extract features matched those of the lake extract and the library

spectra for anabaenopeptins. This cluster confirmed the identity of these features as anabaenopeptin analogues and demonstrated that the mussel extract contained the same family of toxins present in the lake water. If unknown analogues had been present, they would likely have appeared as nodes in this cluster even without matches in the library, providing us with clues for further investigation. For this study, the focus remained on known analogues that could be identified with certainty.

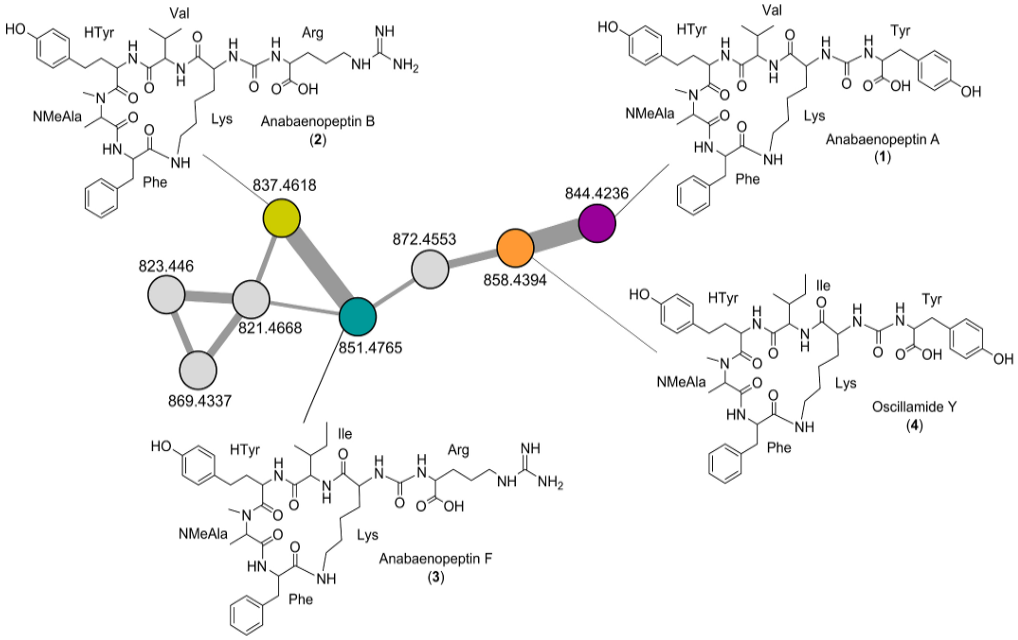


Figure 8.3 Anabaenopeptin cluster: nodes are labeled with the parent mass ratio. Gray nodes correspond to anabaenopeptins found only in water samples collected from the lake. Edge thickness is related to the cosine similarity score.[1]

8.9 Quantitative analysis of cyanotoxins

To quantify the amount of cyanotoxins in mussel samples, we performed targeted LC-MS analysis using an external calibration curve.[10] We selected anabaenopeptin B (AP B) as the calibration standard because it was one of the main toxins detected and we had isolated reference material available. A stock solution of AP B at 1 mg/mL in methanol was prepared. Serial dilutions

were made from this to obtain standard solutions at concentrations of 0.1, 0.5, 1, 5, 7.5, 10, and 50 µg/mL.

These seven points cover a wide dynamic range and were expected to produce a reliable linear calibration curve. Each standard was injected into the LC-MS (Orbitrap) under the same conditions described for the sample analysis (same column, gradient, instrument settings) to record its peak signal.

A calibration curve was constructed by plotting the peak area of AP B against known concentrations. Specifically, we extracted the ion chromatogram (EIC) for the ion *m/z* 837.3 (corresponding to protonated anabaenopeptin B) for each standard injection and integrated the peak at the retention time (~19.8 min) of AP B. The areas from 0.1 to 50 µg/mL were fitted (probably with a linear regression), and the linearity was considered good (often $r^2 > 0.99$ for these ranges). The calibration function (slope and intercept) was then obtained to convert the measured peak area to concentration.

The mussel extract was diluted or adjusted to the appropriate concentration (it was analysed at 5 mg/mL as indicated) and injected in triplicate to ensure accuracy. We again extracted the ion chromatogram at *m/z* 837.3 for the mussel extract and measured the peak area of anabaenopeptin B. Using the calibration curve, we calculated the concentration of AP B in the injected extract solution. We then performed a backward calculation to express this value in µg of AP B per kg of original mussel tissue. This step considered the dilution factor (5 mg/mL of extract corresponds to a certain wet weight of tissue; for example, if 2 kg of mussels yielded ~200 mg of total extract between fractions and AP B was only present in the polar fraction, we divided the value accordingly). The result was that the concentration of

anabaenopeptin B in mussels was determined to be approximately 21 µg per kg of mussels (wet weight).

For the other toxins (anabaenopeptin A, anabaenopeptin F, and oscillamide Y), we did not have individual standards. To estimate their levels, we used the calibration curve of AP B as a proxy. We identified their peaks in the mussel extract based on their unique *m/z* values and retention time (~19.8 min, which was curiously similar for all these peptides under our conditions). The *m/z* values for A, F, and oscillamide Y differ slightly (each anabaenopeptin has a unique mass due to differences in amino acid composition), so we extracted those ions and integrated their peaks. Assuming a similar response factor, we plugged these peak areas into the AP B calibration equation. The calculations yielded approximate concentrations, which were then converted to µg per kg of mussel.

The results were: anabaenopeptin A 19 µg/kg, anabaenopeptin F 2.8 µg/kg, and oscillamide Y 0.5 µg/kg. These values are reported with the caveat that they use AP B as a surrogate standard, but given the small differences in structure, the quantification is still considered semi-quantitative (within an acceptable factor for risk assessment).

All LC-MS quantitative analyses included adequate quality control measures. We performed blank tests with solvent to check for any contamination (none detected, given the low levels in µg). Calibration standards were interspersed throughout the analysis to ensure that the sensitivity of the instrument did not change; if it did, we corrected using a bracket calibration approach. In addition, the lake water extract (from which AP B was originally isolated) could be analysed to provide a reference point or to verify retention times.

The quantified data were then used to assess exposure. Although there are no official regulatory limits for anabaenopeptins in shellfish (as they are emerging toxins), the data serve as a baseline for understanding the environmental transfer of cyanotoxins.[15]

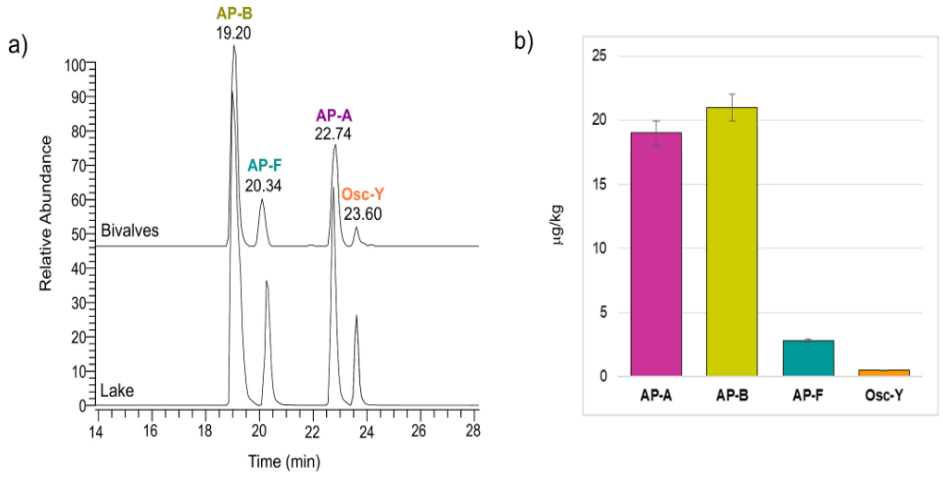


Figure 8.4 Quantitative analysis of anabaenopeptins present in bivalve mollusk samples. a) Ion chromatograms extracted at *m/z* 844.4236, 837.4618, 851.4765, 858.4394 corresponding respectively to anabaenopeptin A (1), B (2), F (3) and oscillamide Y (4) from the MeOH extract of Lake Averno (lower trace) and from bivalve mollusks (upper trace). b) Results of quantitative analysis of anabaenopeptins A (1), B (2), F (3) and oscillamide Y (4) in bivalve mollusk samples. Results are expressed in µg of cyanotoxin per kg of bivalve mollusks.[1]

8.10 Experimental section

8.10.1 General methods

High-resolution ESI mass spectra were measured on a Thermo LTQ Orbitrap XL mass spectrometer coupled to a Thermo Ultimate 3000 UPLC system.

8.10.2 Remote and proximal sensing

Satellite acquisitions and pre-processing

We used Sentinel-2 Level 2A multispectral images (Sentinel Hub / EO Browser workflow) to map the evolution of the bloom. The 10 m bands (B2,

B3, B4, B8) supported water/vegetation monitoring; the 20 m bands (B5, B6, B7, B8A, B11, B12) helped detect bio-optical anomalies; the 60 m bands (B1, B9, B10) provided information for atmospheric correction. Over Italy (~40° N), the effective revisit time is typically ~3-4 days thanks to twin satellites, allowing temporal monitoring of the event.

Drone platforms, calibration, and photogrammetry

Proximal detection used DJI Mavic 3 Enterprise platforms (thermal and multispectral) with RTK for centimeter-level geolocation. The Mavic 3 Multispectral is equipped with four 5 MP monochrome sensors (green 560 ± 16 nm; red 650 ± 16 nm; red-edge 730 ± 16 nm; NIR 860 ± 26 nm) plus a 20 MP RGB camera. Two MAPIR Survey3 cameras provided complementary filter sets: OCN (orange 600–620 nm; Cyan 480–500 nm; NIR 850 nm) and RGN (Red 640–660 nm; Green 520–540 nm; NIR 850 nm). Radiometric integrity was ensured through pre-flight calibration targets and onboard radiometers with continuous irradiance recording. The missions were planned in DJI software for high overlap and accurate georeferencing; flights at 100 m AGL produced ~2.4 cm GSD. The images were processed with PIX4D (Mapper/Fields) to generate orthomosaics and DSMs, then integrated with Sentinel-2 layers in QGIS for multiscale analysis.

Multispectral indices and classification

To biologically track the waters affected by the bloom and associated surface water bodies, we calculated standard indices (NDVI, NDTI) and, given the red chromaticity of the phenomenon determined by phycoerythrin, we designed custom red-weighted indices that emphasize the prevalence of the red band and turbidity. Supervised classification combined spectral, spatial,

and contextual features. The processing repeated top-down (satellite-driven) and bottom-up (drone-driven) cycles, with selective reprocessing of raw frames to avoid oversampling/ undersampling and to refine the delineation of critical areas (channel mouths, bathing sites, aquaculture plots). The results were summarized in thematic illustrations to guide the subsequent phases (sampling and analysis).

8.10.3 Sampling and handling of mussels

Guided by remote/proximal maps, Mediterranean mussels (*Mytilus galloprovincialis*) were collected on April 21, 2022, from an offshore farm near Pozzuoli, in the receiving area of the emissary.

Approximately 2 kg of live bivalve molluscs were collected, stored in a refrigerator, and a portion was examined under an optical microscope (Motic Panthera C2) prior to extraction. The soft tissues were quickly washed, manually shelled (by cutting the adductor muscle), collected, and homogenised with a blender to obtain a representative suspension for chemical work.

8.10.4 Extraction and fractionation

The homogenate was subjected to polarity-driven liquid-liquid extraction: 100% MeOH (2 × 2 L) for polar constituents; MeOH/$CHCl_3$ 1:1 (2 L) for medium polarity constituents; and 100% $CHCl_3$ (2 × 2 L) for non-polar constituents. The filtrates were clarified (coarse sieve → vacuum filtration with celite) and concentrated under vacuum, yielding 164.2 mg (MeOH), 15.3 mg (MeOH/$CHCl_3$), and 18.7 mg ($CHCl_3$). Each residue was redissolved in MeOH 100% (5 mg mL^{-1}) for LC-HRMS/MS analysis.

8.10.5 LC-MS/MS analysis

Chromatographic separations were performed on an Agilent 1100 HPLC coupled to a high-resolution Thermo LTQ Orbitrap XL mass spectrometer operating in positive ESI mode. The analytes were resolved on a Phenomenex Kinetex C18 column (100 × 2.1 mm, 5 µm) maintained at room temperature and eluted at 0.2 mL min^{-1} with solvent A: H_2O + 0.1% formic acid and solvent B: MeOH + 0.1% formic acid. The gradient program was: 0–3 min, 10% B; 3–33 min, 10→100% B; 33–40 min, 100% B (followed by rebalancing of the column to initial conditions). This low-flow acid regime improves peptide ionization efficiency and MS sensitivity, while providing sufficient resolving power between polar and lipophilic constituents.

The Orbitrap was adjusted for stable spraying and efficient desolvation (spray voltage 4.8 kV; capillary 285 °C; N_2 sheath 32 a.u. ≈ 150 mL min^{-1}; N_2 auxiliary 15 a.u. ≈ 50 mL min^{-1}). Data were acquired in data-dependent acquisition (DDA): Each full scan (*m/z* 150-2000) in the Orbitrap was followed by MS/MS on the top five precursors in the linear ion trap using CID with isolation width 2.0, NCE 35, Q activation 0.250, activation time 30 ms. Instrument control and inspection were performed in Thermo Xcalibur (2.2 SP1 build 48). This workflow balances coverage depth with high-precision precursor mass measurement, producing rich peptide fragmentation patterns suitable for dereplication and networking.

8.10.6 Computational workflow in MZmine 2.53

The supplier's .raw files were processed in MZmine 2.53. Mass detection was performed on data centred at MS^1 and MS^2 levels with a noise level of 1000 to exclude baseline signals. The ADAP module was used to construct the chromatogram, with a minimum peak height of 1000 and an *m/z* tolerance of 0.01 (20 ppm), generating retention time (RT) and *m/z* characteristics for each sample. For the alignment between samples, the Join aligner was used with an *m/z* tolerance of 0.05 (10 ppm) and an RT tolerance of 0.3 min to consolidate the characteristics between lake and mussel extracts and between polarity fractions.

To reduce redundancy and source-induced artifacts, common adduct/isotope species were removed from the aligned list by applying adduct filters (maximum relative height 100%) for $[M+Na–H]^+$, $[M+K–H]^+$, $[M+Mg–2H]^+$, $[M+NH_3]^+$, $[M+1, {}^{13}C]^+$, $[M–{}^{35}Cl+{}^{37}Cl]^+$. Essentially, we kept only features with associated MS/MS spectra to ensure that every node entering the network had structural information. The curated spectral bundles were exported as .mgf (MS/MS) and the table of quantitative features (RT, peak area, peak height) as .csv for submission to GNPS.

8.10.7 Feature-based molecular networking

We generated a molecular network based on features on GNPS from MZmine exports using precursor and fragment tolerances of 0.02 Da, with cosine ≥ 0.6 and ≥ 5 matching fragments to define edges. The library search in the workflow used cosine ≥ 0.6 with ≥ 6 matching peaks to annotate nodes relative to community spectral collections. Networks were visualized in

Cytoscape 3.9.1 for chemical mapping, class-level clustering, and rapid localization of cyanotoxin families.

8.11 Phylogenetic and spectroscopic data

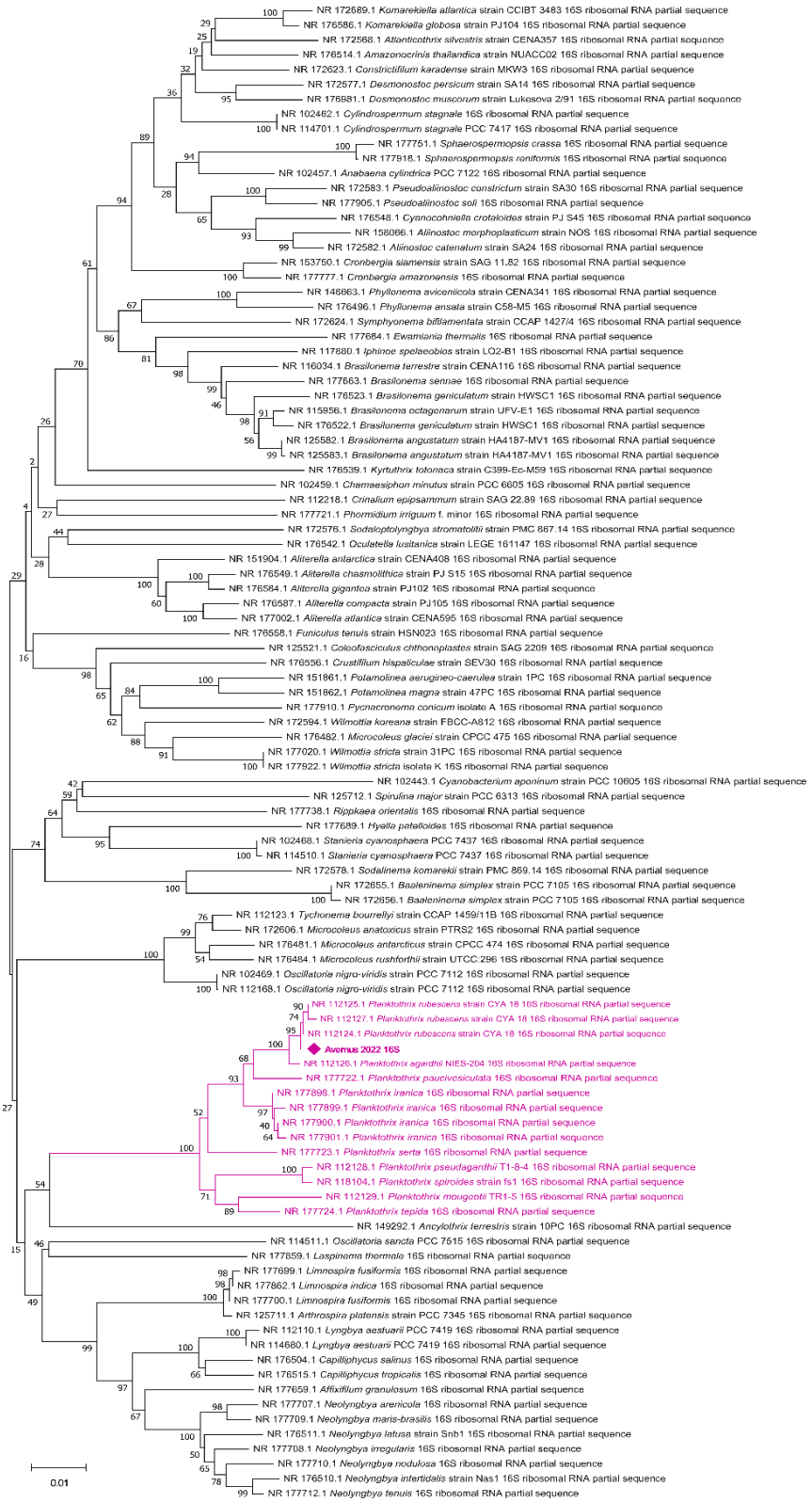


Figure 8.5. 16S rRNA gene neighbour-joining phylogenetic tree displaying the taxonomy of the cyanobacterial community inhabiting 2022 Lake Avernus bloom labelled with diamond (♦). Bootstrap values are given at nodes. Scale-bar represents the phylogenetic distance related to the number of nucleotide substitutions per site.[2]

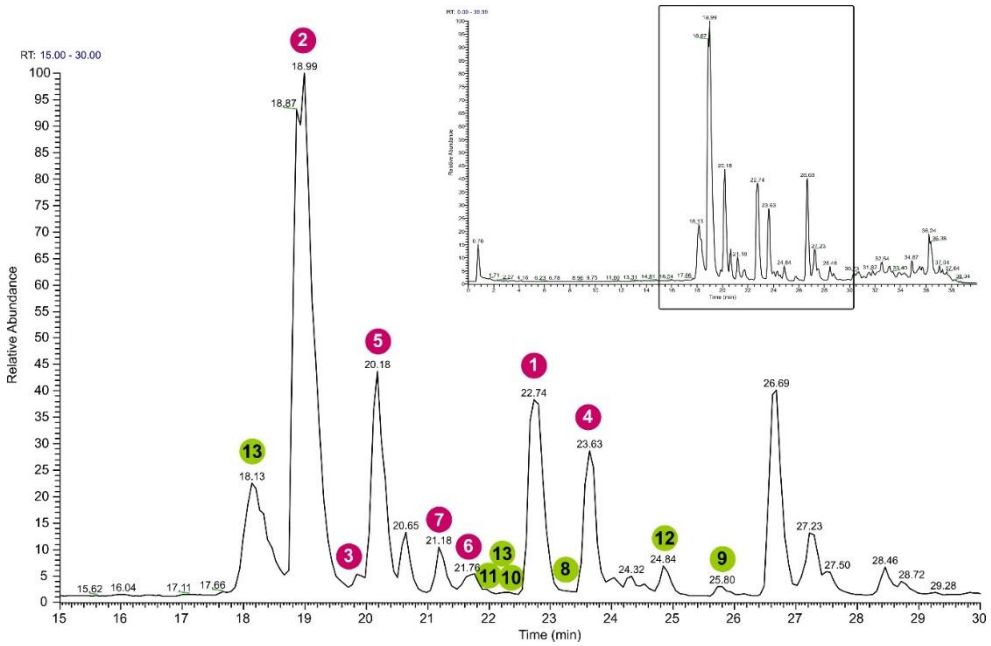


Figure 8.6. LC-HRMS total ion chromatogram (TIC) of the 2022 Lake Avernus bloom MeOH extract. Peak numbers refer to known (purple) and new (green) anabaenopeptin variants.[2]

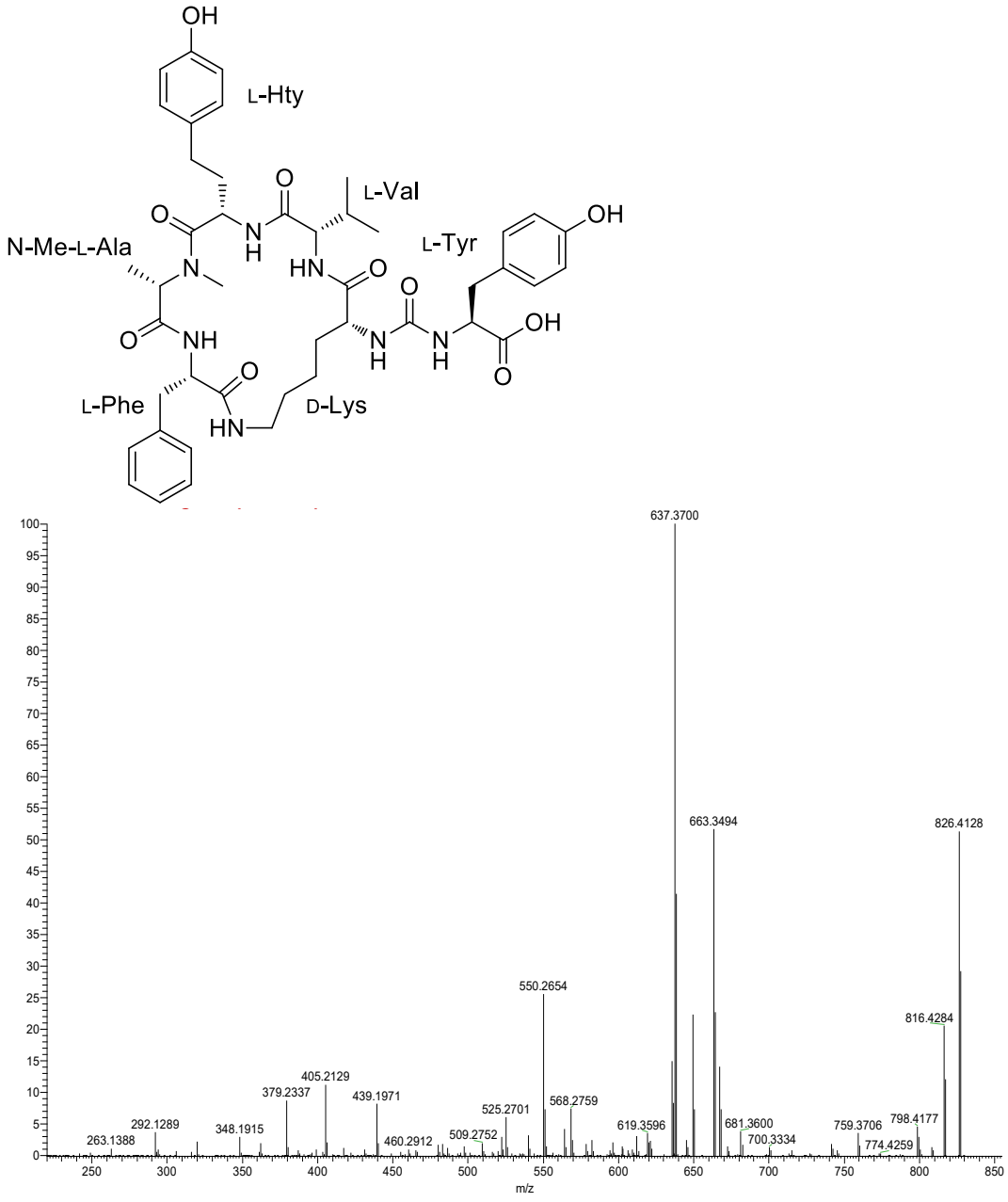


Figure 8.7. HRMS/MS spectrum of AP-A (*m/z* 844.4234).[2]

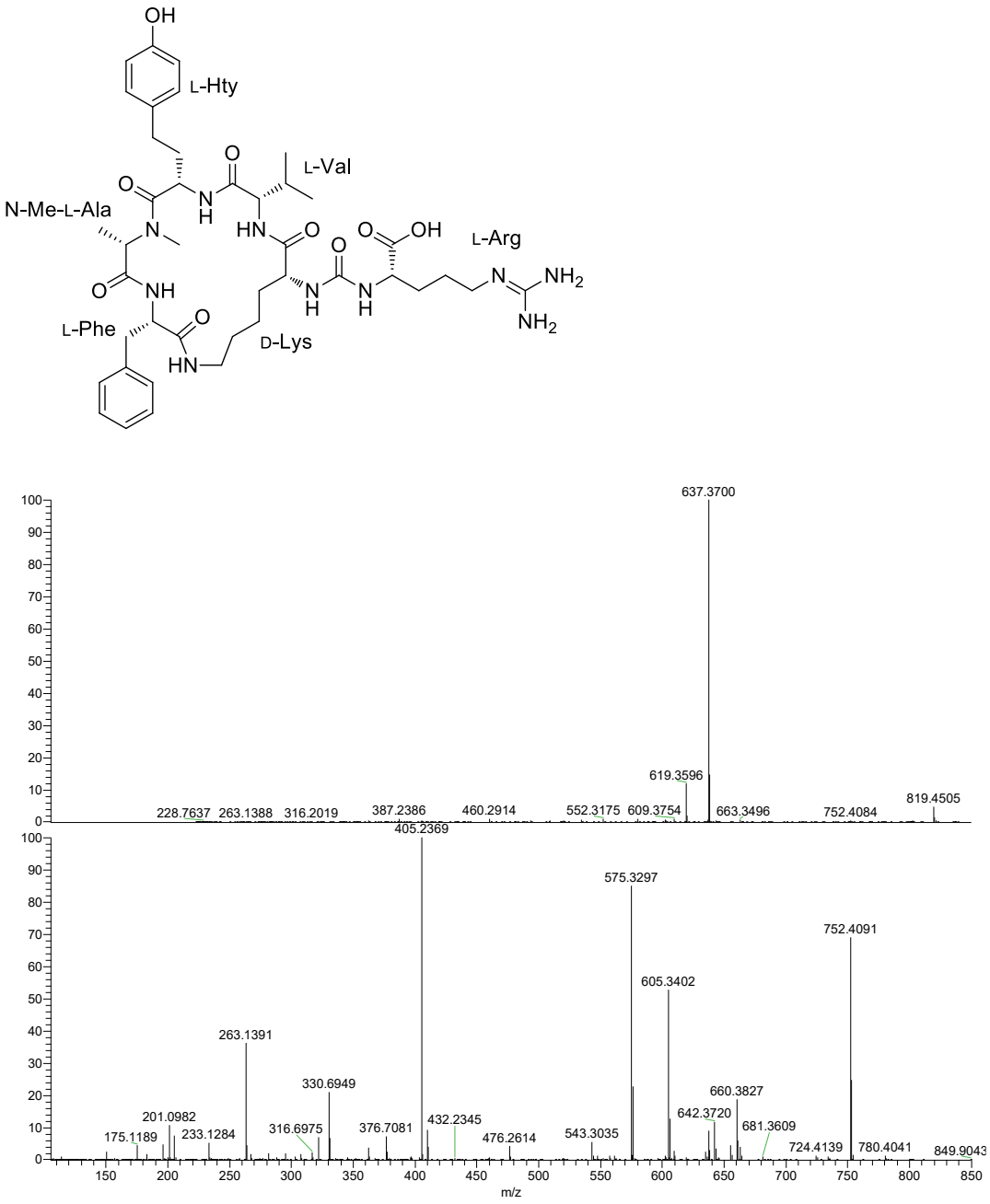


Figure 8.8. HRMS/MS spectrum of AP-B (*m/z* 837.4618).[2]

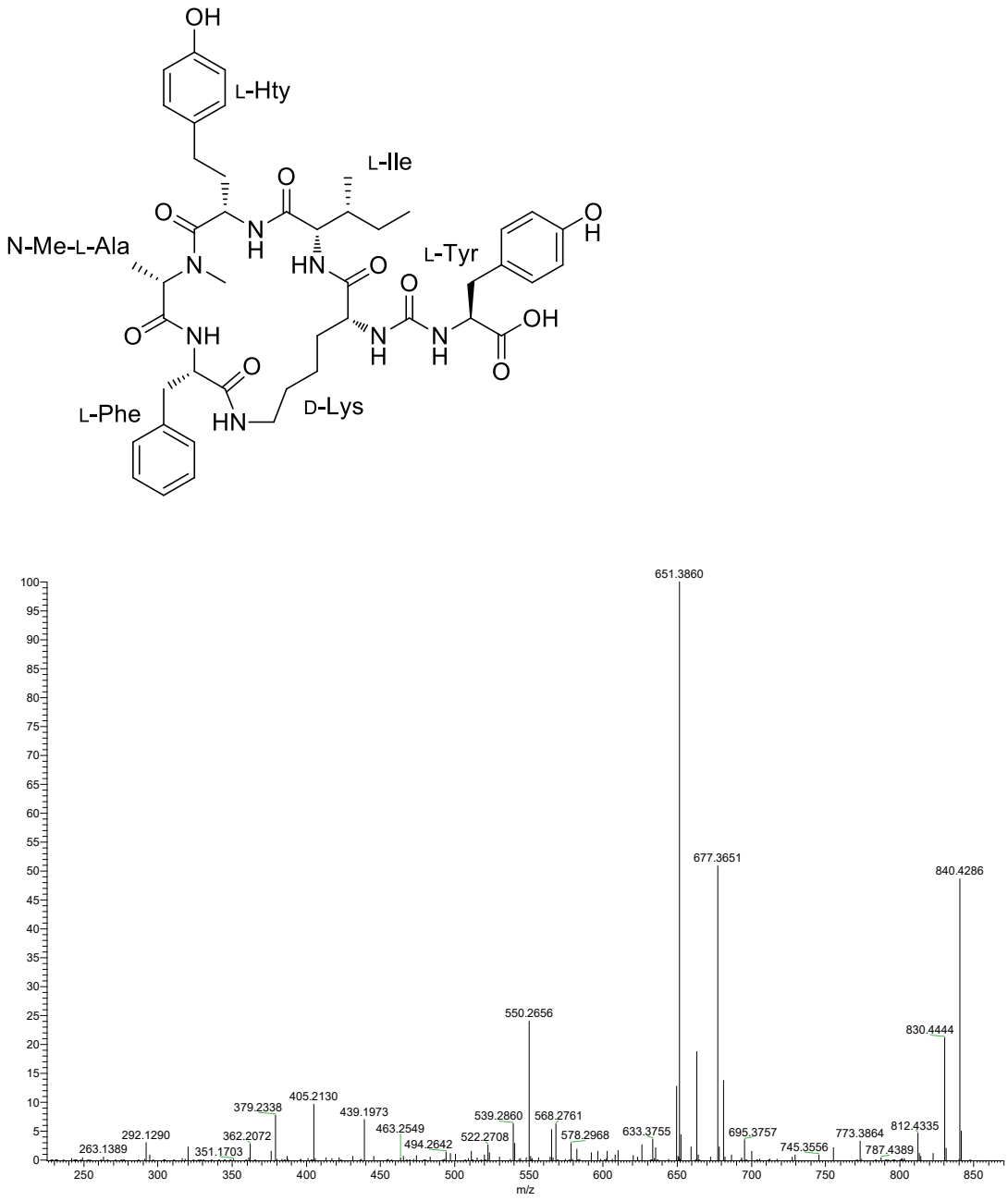


Figure 8.9. HRMS/MS spectrum of Osc-Y (*m/z* 858.4393).[2]

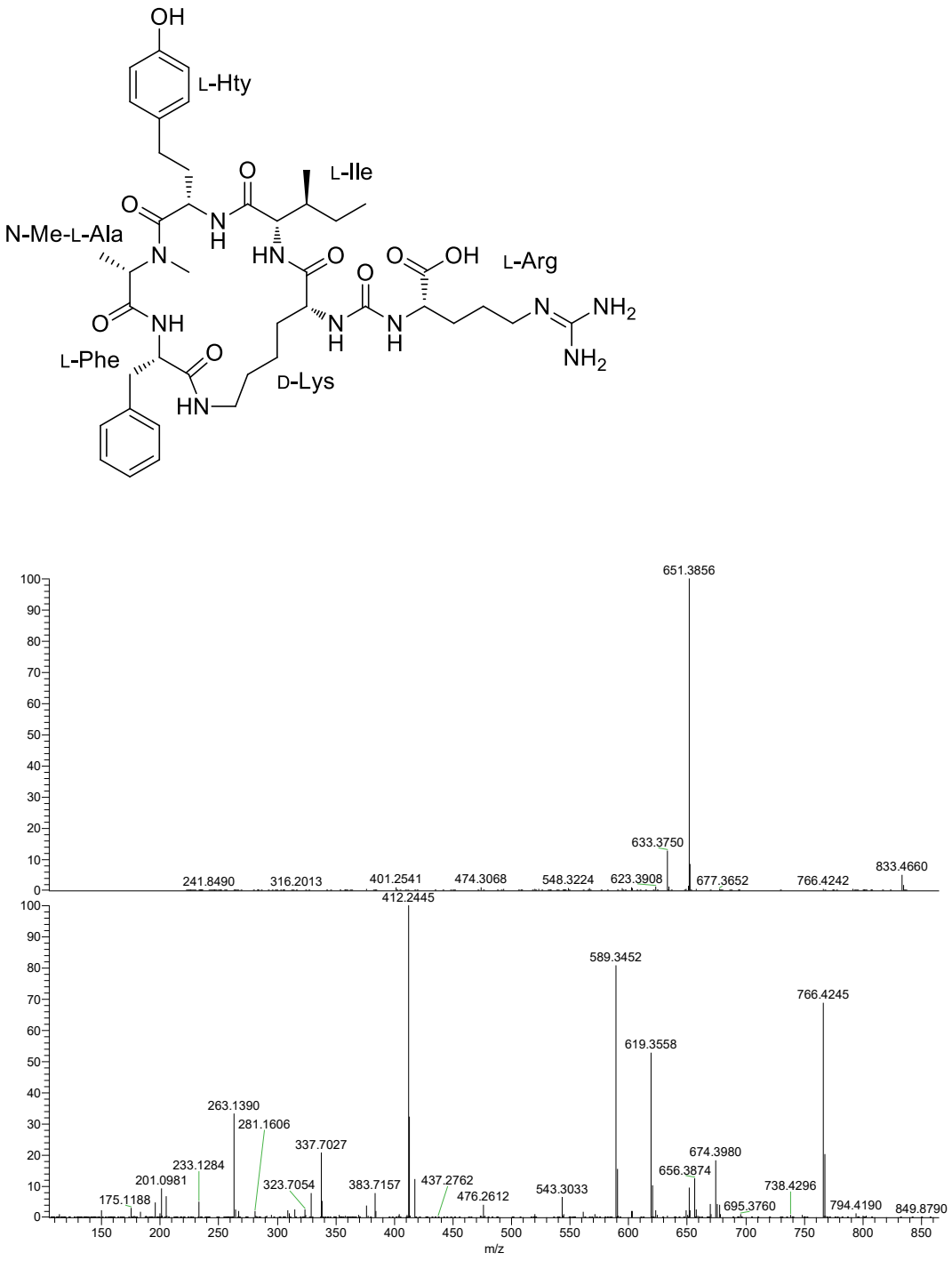


Figure 8.10. HRMS/MS spectrum of AP-F (*m/z* 851.4768).[2]

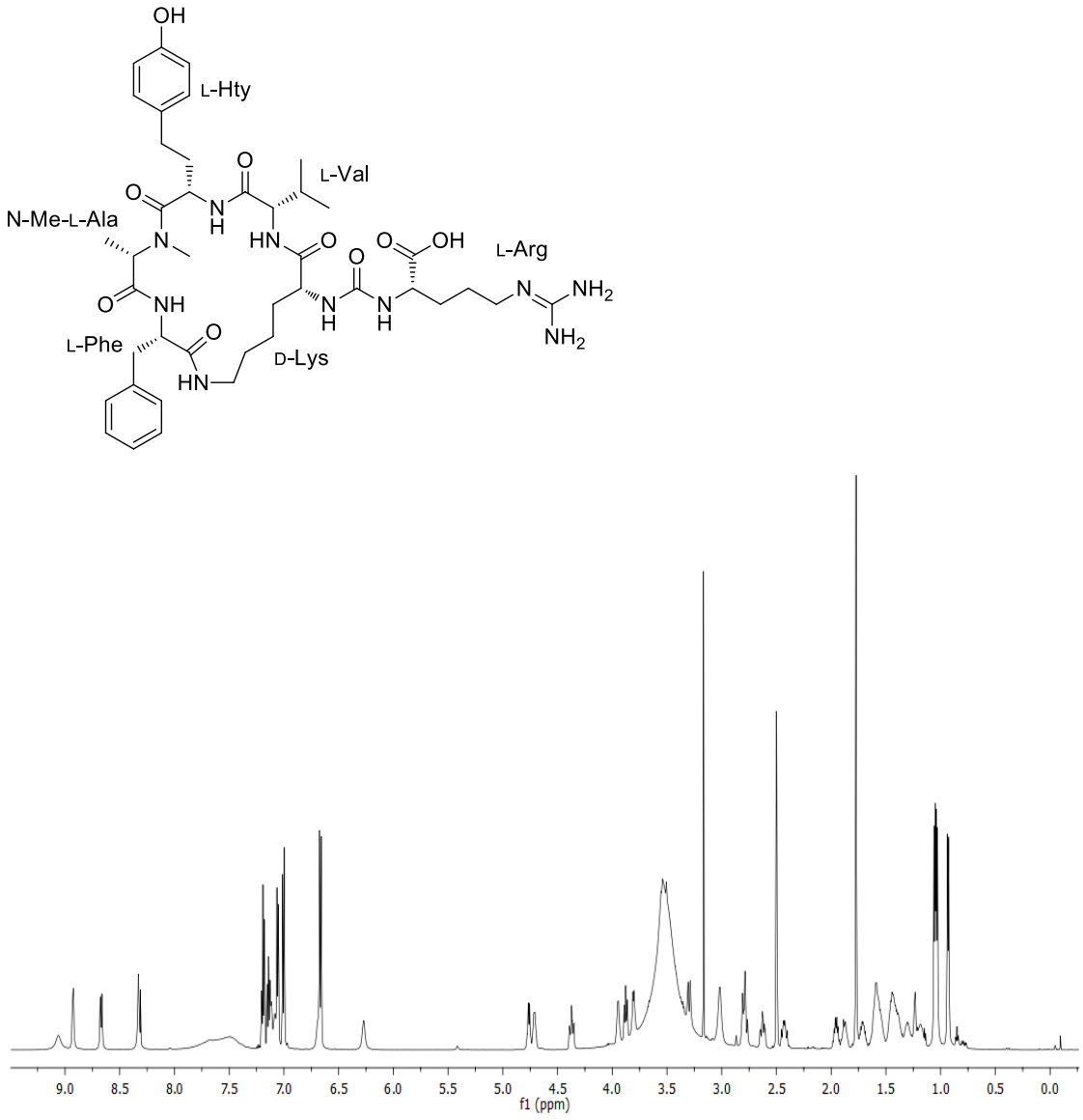


Figure 8.11. ^{1}H-NMR spectrum of AP-B (700 MHz, DMSO-d_6).[2]

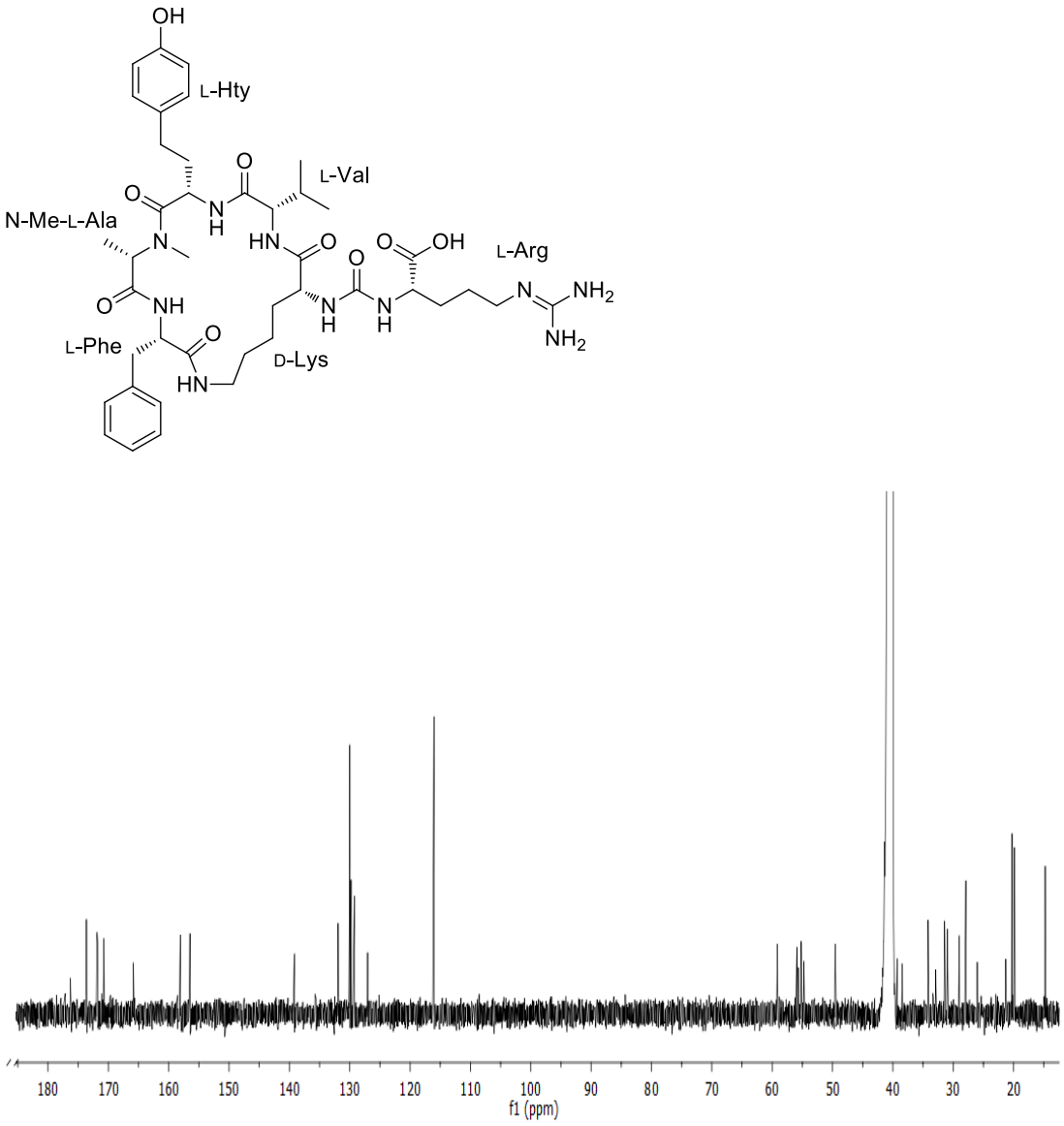


Figure 8.12. ^{13}C-NMR spectrum of AP-B (700 MHz, DMSO-d_6).[2]

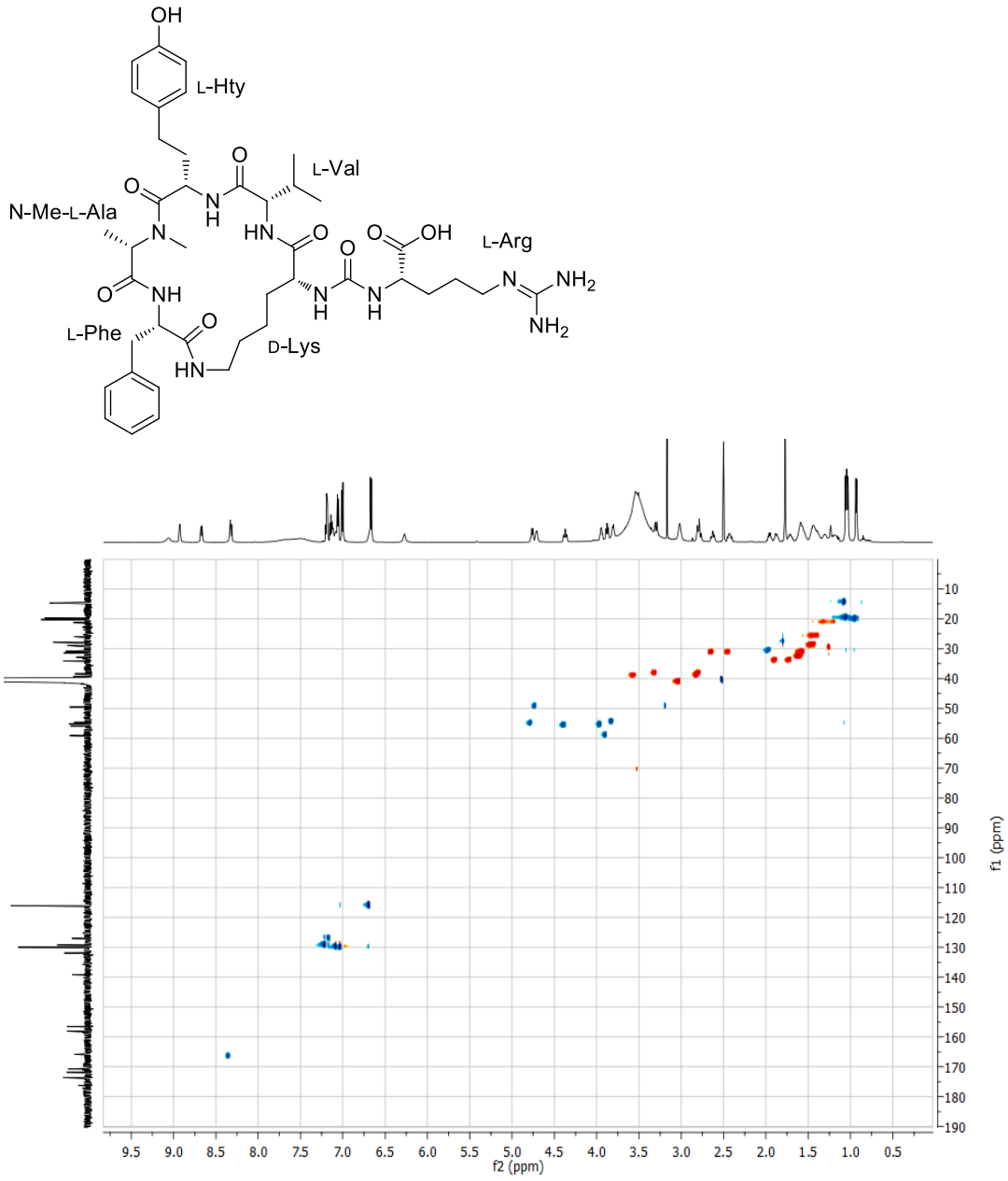


Figure 8.13. HSQC-NMR spectrum of AP-B (700 MHz, DMSO-d_6).[2]

AA	pos.	δ_C, type		δ_H, mult (J in Hz)
Phe	1	170.7, C		
	2	55.9, CH		4.41, ddd (12.5,8.7,3.2)
	3	38.4, CH_2	a	3.59, dd (13.2, 3.2)
			b	2.82, dd (13.2,12.5)
	4	139.2, C		
	5, 9	129.8, CH		7.10, d (7.0)
	6, 8	129.3, CH		7.23, dd (7.4, 7.0)
	7	127.0, CH		7.18, m
	NH*			8.71, d (8.7)
MeAla	1	165.9, C		
	2	55.2, CH		4.80, q (6.9)
	3	14.8, CH_3		1.09, d (6.9)
	N-Me	28.0, CH_3		1.82, s
Hty	1	171.8, C		
	2	49.5, CH		4.75, m
	3	34.1, CH_2	a	1.92, m
			b	1.75, m
	4	31.4, CH_2	a	2.66, ddd (13.3,10.7,3.9)
			b	2.48, ddd (13.3,10.7,6.8)
	5	131.9, C		
	6, 10	130.1, CH		7.05, d (8.4)
	7, 9	116.2, CH		6.71, d (8.4)
	8	156.5, C		
	NH*			8.97, d (4.0)
	OH*			9.11, br s
Val	1	173.6, C		
	2	59.1, CH		3.92, t (7.6)
	3	30.9, CH		1.99, m
	4	20.3, CH_3		0.97, d (6.4)
	5	19.9, CH_3		1.08, d (6.6)
	NH*			7.16, d (7.6)
Lys	1	171.9, C		
	2	55.7, CH		3.98, dd (11.6,6.3)
	3	32.9, CH_2		1.6, m
	4	21.3, CH_2	a	1.34, m
			b	1.21, m
	5	29.0, CH_2		1.48, m
	6	41.0, CH_2		3.07, m
	α-NH*			6.67, d (6.3)
	ε-NH*			7.43, m
Arg	1	176.3, C		
	2	54.7, CH		3.8, m
	3	30.2, CH_2		1.58, m
	4	26.0, CH_2	a	1.48, m
			b	1.42, m
	5	41.0, CH_2		3.07, m
	6	158.0, C		
	α-NH*			6.27, br s
	δ-NH*			7.43, m
	CO(ureido)	158.1, C		

***** *These signals were attributed by comparison with literature data*

Table 8.1 NMR data of AP-B (2) (700 MHz, DMSO-d_6).[2]

References

1. Esposito, G., De Rosa, T., Di Matteo, V., Ciccarelli, C., Ajaoud, M., Teta, R., Lega, M., Costantino, V. Bio-tracking, bio-monitoring and bio-magnification interdisciplinary studies to assess cyanobacterial harmful algal blooms (cyanoHABs)'impact in complex coastal systems. *Stoten* **2025**, 978, 179480.

2. Esposito, G.; Glukhov, E.; Gerwick, W. H.; Medio, G.; Teta, R.; Lega, M.; Costantino, V. Lake Avernus Has Turned Red: Bioindicator Monitoring Unveils the Secrets of "Gates of Hades." *Toxins* **2023**, 15. doi:10.3390/toxins15120698.

3. Alexander, D. E. Bioaccumulation, Bioconcentration, Biomagnification. In Environmental Geology, Encyclopedia of Earth Science; Kluwer Academic Publishers: Dordrecht, **1999**; pp 43–44. doi:10.1007/1-4020-4494-1_31.

4. Buratti, F. M.; Manganelli, M.; Vichi, S.; Stefanelli, M.; Scardala, S.; Testai, E.; Funari, E. Cyanotoxins: Producing Organisms, Occurrence, Toxicity, Mechanism of Action and Human Health Toxicological Risk Evaluation. *Arch. Toxicol.* **2017**, 91, 1049–1130. doi:10.1007/s00204-016-1913-6.

5. Buratti, S.; Franzellitti, S.; Poletti, R.; Ceredi, A.; Montanari, G.; Capuzzo, A.; Fabbri, E. Bioaccumulation of Algal Toxins and Changes in Physiological Parameters in Mediterranean Mussels from the North Adriatic Sea (Italy). *Environ. Toxicol.* **2013**, 28, 451–470. doi:10.1002/tox.20739.

6. Chorus, I.; Fastner, J.; Welker, M. Cyanobacteria and Cyanotoxins in a Changing Environment: Concepts, Controversies, Challenges. *Water* **2021**, 13, 2463. doi:10.3390/w13182463.

7. U.S. Environmental Protection Agency (EPA). Drinking Water Health Advisory for the Cyanobacterial Toxin Microcystins; *EPA 820R15100*, **2015**. (0.3 µg/L preschoolers; 1.6 µg/L altri)

8. European Union. *Directive (EU) 2020/2184 on the Quality of Water Intended for Human Consumption*; OJ L 435/1, 16 Dec **2020**.

9. Cook, K. V.; Beyer, J. E.; Xiao, X.; Hambright, K. D. Ground-Based Remote Sensing Provides Alternative to Satellites for Monitoring Cyanobacteria in Small Lakes. *Water Res.* **2023**, 242, 120076. doi:10.1016/j.watres.2023.120076.

10. España Amórtegui, J. C.; Pekar, H.; Retrato, M. D. C.; Persson, M.; Karlson, B.; Bergquist, J.; Zuberovic-Muratovic, A. LC-MS/MS Analysis of Cyanotoxins in Bivalve Mollusks—Method Development, Validation and First Evidence of Occurrence of Nodularin in Mussels (Mytilus edulis) and Oysters (Magallana gigas) from the West Coast of Sweden. **Toxins** *2023*, 15, 329. doi:10.3390/toxins15050329.

11. Teta, R.; Della Sala, G.; Mangoni, A.; Lega, M.; Costantino, V. Tracing Cyanobacterial Blooms to Assess the Impact of Wastewaters Discharges on Coastal Areas and Lakes. *Int. J. Sustain. Dev. Plan.* **2016**, 11, 804–811. doi:10.2495/SDP-V11-N5-804-811.

12. Teta, R.; Romano, V.; Della Sala, G.; Picchio, S.; De Sterlich, C.; Mangoni, A.; Di Tullio, G.; Costantino, V.; Lega, M. Cyanobacteria as Indicators of Water Quality in Campania Coasts, Italy: A Monitoring Strategy Combining Remote/Proximal Sensing and In Situ Data. *Environ. Res. Lett.* **2017**, 12. doi:10.1088/1748-9326/aa5649.

13. Teta, R.; Esposito, G.; Casazza, M.; Zappa, C. J.; Endreny, T. A.; Mangoni, A.; Costantino, V.; Lega, M. Bioindicator RS as a Tool in Environmental Impact Assessment: Cyanobacteria as a Sentinel of Pollution. *Int. J. Sustain. Dev. Plan.* **2019**, 14, 1–8. doi:10.2495/SDP-V14-N1-1-8.

14. Teta, R.; Della Sala, G.; Esposito, G.; Stornaiuolo, M.; Scarpato, S.; Casazza, M.; Anastasio, A.; Lega, M.; Costantino, V. Monitoring Cyanobacterial Blooms during the COVID-19 Pandemic in Campania, Italy: The Case of Lake Avernus. *Toxins* **2021**, 13. doi:10.3390/toxins13070471.

15. Zorrilla, J. G.; Siciliano, A.; Petraretti, M.; Saviano, L.; Spampinato, M.; Cimmino, A.; Guida, M.; Pollio, A.; Bravi, S.; Masi, M. Ecotoxicological Assessment of Cyclic Peptides Produced by a Planktothrix rubescens Bloom: Impact on Aquatic Model Organisms. *Environ. Res.* **2024**, 257, 119394. doi:10.1016/j.envres.2024.119394.

Conclusion

Natural products remain essential starting points for drug discovery, but progress is often limited by rediscovery, misattribution of structure, sourcing difficulties, and poor translatability to real-world settings. This thesis addresses these limitations with an integrated, multidisciplinary workflow. Genome mining and molecular networking prioritise molecular families, targeted synthesis validates and corrects structural hypotheses, and remote sensing combined with metabolomics guides sampling and measurements. In doing so, it shifts prioritisation from single peaks to families, builds in early stereochemical checkpoints, and links detection directly to management needs, addressing the introduction's key challenges.

Applied to thermophilic peptides, genome-based expectations and MS/MS networking convert a handful of suggestive spectra into a coherent family of congeners, clarifying module flexibility and gatekeeping with implications for pathway engineering and anticipation of natural diversity. The accompanying synthetic study demonstrates that macrocyclisation can be a reproducible epimerisation trap. Routine diagnosis (comparative NMR, advanced Marfey/double derivatisation) and the use of alternative closures that minimise racemisation emerge as best practices, directly reducing false leads, a recurring cause of delays and friction in natural product-based drug discovery. More broadly, the findings support the codification of macrocyclisation as a formal risk point in peptide pipelines, with decision rules on when to change reagents or ring-closing strategies.

The chapter on cyanochelin B links photochemistry to nutrient competition, demonstrating that light rapidly modulates the fate of iron bound to an organic ligand. In the dark, high-affinity chelation resembles monopolisation; under illumination, photolysis/photoreduction redistributes access. This mechanism has clear temporal and spatial dimensions (diurnal cycle, depth, turbidity, seasonality) and is consistent with mobile biosynthetic gene clusters selected under conditions of fluctuating iron scarcity. By linking structure, function, and environment, the work demonstrates how ecological context can refine target selection and mode-of-action hypotheses, addressing another obstacle introduced in the introduction: the gap between in vitro biochemical activity and in situ relevance.

The case of the lake on the coast puts into operation a pipeline that goes from detection to decision. Anomalies detected by satellites trigger drone surveys and field confirmations; untargeted metabolomics organises thousands of features into families that can be interpreted with toxin-level specificity; and monitoring of discharge into a shellfish area illustrates how watershed connectivity projects inland events onto coastal food systems. Although abundance cannot be inferred from LC-MS intensity alone, the process enables near real-time management actions, a result rarely achieved with the classic “grind-and-find” discovery method. In practical terms, this provides a practical path from chemical signal to regulatory response, combining optics (where to look) with chemistry (what matters) on management timelines.

Several general conclusions emerge. (i) Combining molecular networking with genome mining embeds biosynthetic plausibility into spectral interpretation, improving novelty assessment and limiting rediscovery.

(ii) Macrocyclization is a key stereochemical risk point that demands early, systematic verification to avoid costly downstream rework. (iii) Integrating remote sensing with metabolomics is essential for environmental forensics and public health surveillance, linking detection to action. Together, these elements form an end-to-end strategy that helps translate chemodiversity into actionable evidence.

Overall, aligning genomes, networks, spectra and environmental observations accelerates natural products research, improving reliability and decision-readiness. It targets the main bottlenecks outlined in the introduction, prioritisation, structural certainty, sourcing validation and ecological translation, and, through three case studies, shows that many can already be addressed in practice. This integration strengthens the value of discoveries for drug development and ecosystem health, offering a transferable model for future work.