

DISS ETH NO. 27437

**IDENTIFICATION AND CHARACTERIZATION OF
NATURAL PRODUCT BIOSYNTHETIC PATHWAYS ENCODED IN
THE MICROBIOME OF THE SPONGE *MYCALE HENTSCHELI***

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

MICHAEL RUST

MSc Interdisciplinary Sciences, ETH Zurich

born on 11.11.1991

citizen of Walchwil ZG, Switzerland

accepted on the recommendation of

Prof. Dr. Jörn Piel

Prof. Dr. Shinichi Sunagawa

Prof. Dr. Peter Kast

2021

Abstract

Specialized metabolites from natural sources (animals, plants, microbes) have been pillars of traditional and modern medicine alike. The structural diversity of natural products gives rise to various bioactivities, which are exploited to combat infectious diseases, cancer, and numerous other illnesses. Traditional bioactivity-guided approaches led to the identification of numerous blockbuster molecules, but rapidly became limited by the frequent re-isolation of known metabolites. Technological advances have enabled the exploration of new habitats and the mining of genome data for the discovery of novel natural products. Within the marine environment, sponges have been a particularly rich source of specialized metabolites. Sponges, which represent the oldest extant animals, co-exist with astonishing numbers of microorganisms. For some species, it has been shown that microbial symbionts produce bioactive molecules known from the sponge host.

This work centers around the biosynthetic potential encoded in the microbiome of the chemically rich sponge *Mycale hentscheli*. We pursued diverse strategies to (i) identify biosynthetic genes, (ii) link them to complete biosynthetic gene clusters (BGCs) and predict the structure of encoded products, (iii) assign the BGCs to producing organisms, and (iv) provide functional data by expressing individual enzymes in heterologous hosts.

In **Chapter II**, we demonstrated that multiple symbionts jointly generate the rich chemistry in *M. hentscheli*, a contrasting scenario to other sponges, in which single "superproducer" symbionts within more complex microbiomes synthesize most of the bioactive compounds known from their host. In addition to bacterial pathways for the three cytotoxic polyketides (mycalamide, pateamine, and peloruside), we identified numerous orphan BGCs distributed across a broad phylogenetic range of bacteria. Characterization of enzymes from an orphan polytheonamide-type pathway suggests a novel member of this rare family as its product. The microbiome is particularly rich in pathways of a distinct family of polyketide synthases (PKSs), termed *trans*-acyltransferase (*trans*-AT) PKSs, that are often found in uncultivated symbiotic bacteria, reinforcing uncultured microbes as promising sources of chemical novelty.

Uncultivated sponge symbionts often harbor orphan BGCs, raising intriguing questions about their origin and function. Strategies to study them are usually limited to cultivation approaches or heterologous expression studies. In **Chapter III**, we pursued an alternative strategy by searching for similar clusters in sequenced genomes of cultured bacteria. We analyzed a small hybrid cluster consisting of non-ribosomal peptide synthetase and PKS parts in the metagenome of *M. hentscheli* and identified orthologous clusters in bacteria of the *Rhodobacteraceae* family. Establishing a transformation protocol for one of the cultured strains enabled a reporter assay to identify conditions under which the cluster is active. The results provide the basis for targeted product isolation and will facilitate the discovery of similar compounds in sponge extracts.

The identification of the BGC for the pharmaceutically relevant pelorusides paved the way toward biotechnological production. However, the lack of an assigned producer genome restricts targeted isolation and heterologous expression experiments. In **Chapter IV**, we applied two strategies to obtain insights into bacterial symbionts in *M. hentscheli*, a novel indirect sequence capture method to enrich DNA fragments that contain parts of a target BGC and long-read sequencing. We putatively linked the peloruside pathway to the mycalamide producer affiliated with a marine taxon previously not known as source of natural products. Extending both ends of a large orphan polyketide cluster to a member of the Verrucomicrobiota phylum further demonstrated the potential of long-read sequencing to assign BGCs to producing organisms in complex datasets.

Trans-AT PKSs are giant multimodular enzymes that are responsible for the biosynthesis of polyketides exhibiting remarkably diverse structures and bioactivities. These systems often originate in uncultivated and symbiotic bacteria, rendering their analysis particularly challenging. In **Chapter V**, we established a cloning strategy to assemble complete *trans*-AT PKS clusters into plasmids compatible with expression in different hosts. We initially captured clusters with known products from cultivated sources that can be used as model systems to study PKS function. Cloning orphan clusters from a sponge metagenome expanded the set of expression plasmids and opened avenues for the characterization of unknown pathways. Finally, we implemented an inducible expression system in two non-canonical host strains that provides a basis for the development of a heterologous production system.

This thesis sheds light on the biosynthetic potential encoded in the microbiome of the sponge *M. hentscheli* and presents strategies to characterize natural product pathways from uncultivated sources. A better understanding of the biosynthetic machineries will spur the development of synthetic biology tools to expand the natural product chemical space.

Zusammenfassung

Spezialisierte Metaboliten aus natürlichen Quellen (Tiere, Pflanzen, Mikroben) stellen Grundpfeiler der traditionellen und modernen Medizin gleichermaßen dar. Die strukturelle Vielfalt der Naturstoffe widerspiegelt sich in verschiedenen Bioaktivitäten, die zur Bekämpfung von Infektionen, Krebs und zahlreichen anderen Krankheiten genutzt werden. Klassische bioaktivitätsgeleitete Ansätze zur Isolierung von Naturstoffen führten zur Identifizierung zahlreicher Blockbuster-Moleküle, wurden jedoch rasch durch die häufige Re-Isolierung bereits bekannter Metaboliten eingeschränkt. Inzwischen haben technologische Fortschritte die Erkundung neuer Lebensräume und die Auswertung von Genomdaten die Entdeckung neuer Naturstoffe ermöglicht. Innerhalb der marinen Umwelt sind Schwämme eine besonders reiche Quelle an spezialisierten Metaboliten. Schwämme repräsentieren die ältesten lebenden Tiere und koexistieren mit einer erstaunlichen Vielzahl von Mikroorganismen. Für einige Schwammarten wurde gezeigt, dass mikrobielle Symbionten die bioaktiven Naturstoffe produzieren, die von diesem Wirt bekannt sind.

Diese Arbeit konzentriert sich auf das biosynthetische Potential, das im Mikrobiom des chemisch reichen Schwammes *Mycale hentscheli* kodiert ist. Wir verfolgten verschiedene Strategien, um (i) biosynthetische Gene zu identifizieren, (ii) sie mit kompletten biosynthetischen Genclustern (BGC) zu verknüpfen und die Struktur der kodierten Produkte vorherzusagen, (iii) die BGC den produzierenden Organismen zuzuordnen und (iv) funktionelle Daten durch Expression einzelner Enzyme in heterologen Wirten zu bekommen.

In **Kapitel II** zeigten wir, dass mehrere Symbionten gemeinsam die Fülle an Naturstoffen in *M. hentscheli* erzeugen. Dies steht im Gegensatz zu anderen Schwämmen, in denen einzelne "Superproduzenten" als Teil komplexerer Mikrobiome den Grossteil der bioaktiven Substanzen ihres Wirtes synthetisieren. Zusätzlich zu den bakteriellen Biosynthesewegen für die drei zytotoxischen Polyketide (Mycalamid, Pateamin und Pelorusid) identifizierten wir zahlreiche unbekannte BGC, die über einen grossen phylogenetischen Bereich von Bakterien verteilt sind. Die Charakterisierung von Enzymen eines unbekanntes Polytheonamid-ähnlichen Biosyntheseweges deutet auf ein neues Produkt dieser seltenen Naturstofffamilie hin. Das Mikrobiom von *M. hentscheli* ist besonders reich an Stoffwechselwegen einer bestimmten Familie von Polyketid-Synthasen (PKS). Diese werden als *trans*-Acyltransferase (*trans*-AT) PKS bezeichnet und treten häufig in unkultivierten symbiotischen Bakterien auf, was unkultivierte Mikroben zu einer vielversprechenden Quelle für neuartige Moleküle macht.

Das Auftreten neuartiger BGC in unkultivierten Schwammsymbionten wirft Fragen über deren Ursprung und Funktion auf. Strategien zur Untersuchung dieser BGC beschränken sich normalerweise auf Kultivierungsansätze oder heterologe Expressionsstudien. In **Kapitel III** verfolgten wir eine alternative Strategie, in der wir sequenzierte Genome von kultivierten Bakterien nach ähnlichen Clustern durchsuchten. Wir analysierten einen kleinen Hybridcluster bestehend aus nicht-ribosomalen Peptid-Synthetase und PKS-Elementen aus dem Metagenom von *M. hentscheli* und identifizierten orthologe Cluster in Bakterien der Familie *Rhodobacteraceae*. Die Etablierung eines Transformationsprotokolls für einen der kultivierten Stämme ermöglichte einen Reporter-Assay zur Identifizierung von Bedingungen, unter denen der Cluster aktiv ist. Die Ergebnisse bilden die Grundlage für eine gezielte Produktisolierung und erleichtern zukünftig die Entdeckung ähnlicher Verbindungen in Schwammextrakten.

Die Identifizierung des BGC für die pharmazeutisch relevanten Peloruside ebnete den Weg zu deren biotechnologischen Produktion. Da jedoch bisher dem BGC kein Produzentengenom zugeordnet ist, sind die gezielte Isolierung und heterologe Expressionsexperimente nur eingeschränkt machbar. In **Kapitel IV** verfolgten wir zwei Strategien, um Einblicke in bakterielle

Symbionten in *M. hentscheli* zu erhalten: eine neuartige indirekte Methode zur Anreicherung von DNA-Fragmenten die Teile eines Ziel-BGC enthalten und Long-Read-Sequenzierung. Dadurch haben wir den Pelorusid-Biosyntheseweg mit dem Mycalamid-Produzenten in Verbindung gebracht. Dieser gehört zu einem marinen Taxon, das bisher nicht als Quelle von Naturstoffen bekannt war. Ausserdem konnten wir beide Enden eines großen unbekanntes Polyketid-Clusters einem Mitglied des Verrucomicrobiota-Phylums zuordnen. Unsere Daten demonstrieren das Potenzial der Long-Read-Sequenzierung, um BGC in komplexen Datensätzen produzierenden Organismen zuzuordnen.

Trans-AT PKS sind riesige multimodulare Enzyme, die für die Biosynthese von Polyketiden mit erstaunlicher struktureller Vielfalt und verschiedener Bioaktivitäten verantwortlich sind. Diese Systeme stammen oft aus unkultivierten und symbiotischen Bakterien, was ihre Analyse besonders herausfordernd macht. In **Kapitel V** etablierten wir eine Klonierungsstrategie, um komplette *trans*-AT PKS-Gencluster in Plasmiden zu assemblieren, die mit der Expression in unterschiedlichen Wirten kompatibel sind. Wir starteten mit Clustern aus kultivierten Bakterien deren Produkte bekannt sind, da diese als Modellsysteme zur Untersuchung von PKS-Funktionen verwendet werden können. Das Klonieren von unbekanntes Genclustern aus einem Schwamm-Metagenom erweiterte den Satz an Expressionsplasmiden und eröffnete Wege zur Charakterisierung neuer Biosynthesewege. Schliesslich implementierten wir ein induzierbares Expressionssystem in zwei nicht-typischen Wirtsstämmen, welches eine Grundlage für die Entwicklung eines heterologen Produktionssystems darstellt.

Diese Arbeit beleuchtet das biosynthetische Potenzial, das im Mikrobiom des Schwamms *M. hentscheli* kodiert ist und stellt Strategien zur Charakterisierung von Naturstoffbiosynthesewegen aus unkultivierten Quellen vor. Ein besseres Verständnis der biosynthetischen Mechanismen wird die Entwicklung von Werkzeugen der synthetischen Biologie vorantreiben, um die chemische Vielfalt der Naturstoffe zu erweitern.