
Chapter 1

Introduction

1.1 Overview

1.1.1 Natural products

Natural products have evolved to include a broad spectrum of chemical and functional diversity. Reactions occurring in nature are biased towards function, so it follows that every natural product must have a binding partner. If one accepts this premise, it explains the enormous biochemical expense associated with the biosynthesis of natural products. Evolutionary pressure has allowed natural products to target a huge variety of biomolecules, often in a highly selective fashion¹. Although not necessarily evolving alongside human proteins, “natural products emerged to interact with something and that something may not be so different from human proteins”². This also explains the great success they’ve seen as therapeutic agents. Newman and co-workers surveyed newly introduced drugs worldwide, between 1981-2010³⁻⁵. From the 1325 small-molecule, new chemical entities (NCEs) analysed, only one *de novo* NCE reported in the public domain resulted from combinatorial chemistry. The majority of approved drugs were natural products, natural product derivatives or mimics thereof. When analysing the compounds according to disease categories, almost 70% of anti-infectives and 77% used for cancer treatment were naturally, derived or inspired⁵. These numbers correlate to the larger chemical space natural products cover compared to synthetic compounds. On average, natural products have higher molecular weights, incorporate less nitrogen and sulfur (but more oxygen), and are sterically more complex, by encompassing more chiral centres, bridgehead atoms, rings (but fewer aromatic rings), than purely synthetic compounds.

Although natural products are a rich source of potential drug leads, only a small percentage have been investigated chemically or biologically. Even for those compounds demonstrating potent biological activity, the underlying mechanism is rarely elucidated. Generally, in the process of discovering new drugs, the biological target is often identified last.

There is an urgent need for techniques allowing the rapid identification of the cellular targets of biologically active natural products (or any small molecule), for application in the early stages of NCEs identification. Rather than left to last, if this was done early on, rational drug design would be accelerated and off-target interactions minimised through selective knowledge of exactly how the small molecule achieves its effects.

1.1.2 Methods for the isolation of natural product receptors

Traditional approaches to address this question involved testing the biologically active compound against a myriad of possible targets based on clues from the biological activity, followed by further research focusing on the target of highest selectivity⁶. Unfortunately, this is a biased methodology, as the most selective binding partner may simply not have been tested for. As Wittgenstein put it: "Wovon man nicht sprechen kann, darüber muss man schweigen" (what we cannot speak of we must pass over in silence)⁷. Additionally, conclusions derived from any test result, on the underlying mode of action, would try to support or refute an hypothesis based on the current state of knowledge and understanding. This creates a somewhat circular argument that lacks objectivity. Hence, the major limitation for such traditional approaches is the serendipity upon which the identification of the correct target relies.

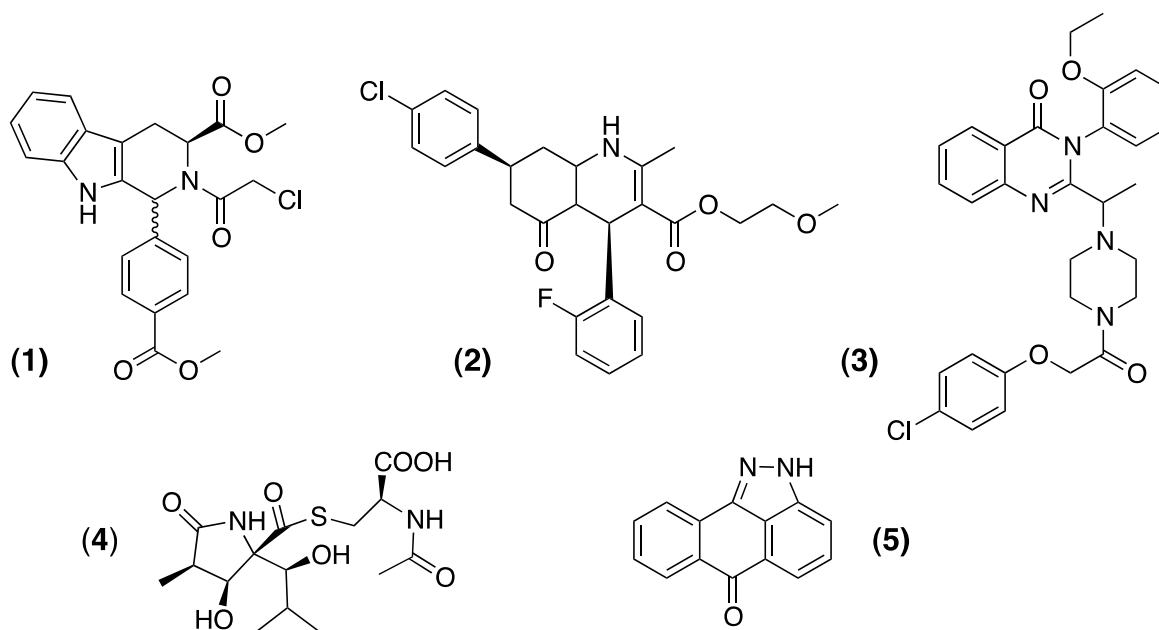
Unbiased ways of elucidating molecular mechanisms, based on genome-wide approaches, have arisen from genetics and proteomics. In classical genetics, a gene knockout aims to indirectly elucidate a gene's function by observing what effect the lack of this gene has on the phenotype produced. However, even after successful knocking out a single gene, a distinct phenotype may not be produced. The gene may be redundantly encoded or compensatory cascades may mask the expected outcome.

1.1.2.1 Chemical genetics

In 1998, Stuart Schreiber introduced the concept of "chemical genetics" and thereby the idea of using small molecules, such as natural products, as "probes" in genetics⁸. Schreiber's concept allowed for the regulation of gene products through the binding of the probes to proteins in cells. This approach is based on the characteristic ability of a biologically active small molecule to activate or inhibit specific biological processes. Furthermore, a major advantage of chemical genetics is temporal flexibility. In comparison, a gene knockout is permanent and impacts the development of the cell from the very beginning. Through chemical genetics, metabolic and genetic masking effects are bypassed as the small molecule can be applied at any time point or withdrawn.

Applications of this concept can be divided into forward or reverse methodologies⁹. In forward chemical genetics a diverse library of small molecules is screened to produce phenotypes of interest. Any compound producing a particular phenotype is

further investigated to identify the genes responsible for that phenotype. For example, the Ras protein family (originally isolated from rat sarcoma) belongs to the signal transmitting GTPase family of proteins. When the underlying gene is "switched on", the expressed protein turns on genes involved in cell growth, differentiation and survival. The *ras* gene is the most frequently mutated oncogene and observed in 20-25% of all tumours. If the *ras* gene is permanently switched on, this ultimately leads to excessive cellular growth and cancer. Stockwell and co-workers screened combinatorial libraries and found two compounds; RSL3 (**1**) and RSL5 (**2**) exhibited increased lethality for mutant cells in the presence of oncogenic *ras*. In counter screens with erastin (**3**), a previously identified oncogenic-*ras*-signalling inhibitor, the authors provided further insights into the selectivity-determining factors for each compound, as well as a better understanding of oncogenic *ras*-dependent cellular mechanisms¹⁰.



One prominent example of the application of natural products in chemical genetics is with lactacystin (**4**) that has been shown to induce neurite outgrowth¹¹. Ongoing research has revealed several important facts. Firstly, the mammalian proteasome subunit X was identified as its protein binding partner, specifically a highly conserved *N*-terminal threonine domain¹². This finding challenged the prevailing dogma that proteasomes were just "garbage disposal" units, digesting unneeded or damaged proteins by proteolysis. Since then, the proteasome has been recognised as an important regulatory complex in many essential cellular processes, such as cell-cycle progression and inflammatory responses¹³.

Reverse chemical genetics screens combinatorial libraries against specific targets (proteins) or systems. For example, olfactory sensory neurons (OSN) experience a cell death process that has long been elusive. When a combinatorial library was screened to find compounds that prevented this cell death, the study revealed SP600125 (5; a synthetic *jun* N-terminal Kinase (JNK) inhibitor), as a potential candidate. Interactions between SP600125 and JNK were later validated both *in vitro* and *in vivo*¹⁴.

More recently, classical and chemical genetics have merged to create a "combination chemical genetics" approach, whereby chemical libraries are screened against multiple phenotypes⁹. This approach most commonly incorporates a forward, but can also implement a reverse, approach similar to that used in chemical genetics. This results in inherently more complex data, as control experiments are required for identifying the effects of both synergism and antagonism on the various simultaneously used perturbans.

An array of chemicals have been tested and many genetic modulations been implemented⁹. Combination chemical genetics is yet to be established as a standard, impartial method for isolating protein-binding partners for small molecules. It has to overcome the bias toward introduced modulations and ensure that cascading effects do not mask the underlying modes of action.

Natural products and libraries thereof are also suitable for application as probes in chemical genetics and combination chemical genetics. They can certainly assist in the overall goal to identify a small-molecule modulator for each gene product⁸. Yet, in order to identify a protein-binding partner for every natural product, unbiased methods are required that provide a comprehensive set of potential protein binding partners from the organism of interest.

1.1.2.2 Genome-wide approaches

Genome-wide approaches provide this unbiased means, as they allow for the selection of a binding partner for a natural product from within the genomic repertoire of any given organism. In general, proteomics-based approaches are categorised into forward and reverse chemical proteomics¹⁵. The first category, forward chemical proteomics (FCP), uses the entire proteome of a cell as a library. The binding partners of the natural product are isolated from within this library. Reverse chemical proteomics (RCP) differs in the starting point. Rather than the

proteome of a cell, its transcriptome (mRNA) is employed to create the library. The genetic information of each library member is then translated to yield the cell's gene products (proteins) that are linked in some way.

Both FCP and RCP detect an interaction between the small molecule and protein, the major difference between FCP and RCP is that in FCP, both the isolation of a natural product's binding partner and the analysis of the latter are performed at the protein level. In contrast, in RCP, only the isolation of a protein-binding partner (target) is driven by protein-small molecule interactions and the analysis is carried out at the gene level. This has the advantage that the level of protein expression is irrelevant and even one molecule of cDNA can be amplified (PCR) and sequenced. In FCP, one is always limited by the amount of protein expressed and interference from high abundance proteins with weak binding to the target and non-specific binding proteins (e.g. albumin). In RCP, the iterative nature means that even one molecule can be isolated and reamplified to enrich the most avid binding partner in the presence of non-specific and weak binding proteins. Sequence searches (BLAST) then reveal the identity of the protein and homologies to other genes.

Both FCP and RCP aim to overcome the limitations of traditional approaches by identifying a drug's target spectrum as thoroughly as possible¹⁶.

1.1.2.2.1 Chemical proteomics

Over the past decades, the drug development process has seen fundamental pragmatic changes regarding the approach to the starting point of each development. Until the 1980s, small molecules were prioritised and researchers sought subsequently to elucidate its target later. Since then there has been a graduate shift, due largely to advances in genomics, to identification of drug targets before suitable ligands were investigated¹⁷. The discovery of new drugs progressed very much in accordance to the Nobel laureate James Black's statement, "the most fruitful basis for the discovery of a new drug is to start with an old drug"¹⁸. Using this approach, the drug's interactions with the active site were examined and chemical modifications of the known drug led to more potent or less toxic analogues. This is however, self-limiting as after derivatives of all known drugs are made, there can be no more. Thus, more recently, investigators have once again begun to incorporate a more chemical view by going back to the search for new bioactive small molecules with the aim of identifying new target classes¹⁷.

These disparate approaches are also mirrored in chemical proteomics, which can be described as two basic flavours: activity-based probe profiling (ABPP)¹⁹ and a compound centric approach (CCCP)¹⁶. In both approaches, it is necessary to tag the small molecule in some form and employ traditional affinity chromatography-based approaches to isolate a group of proteins based either on their catalytic activity (ABPP) or their binding affinity to some structural motif (CCCP) (see 1.1.2.2.1.3).

1.1.2.2.1.1 Affinity chromatography

Affinity chromatography methods firstly require an affinity matrix and secondly tethering of the probe to this matrix through some sort of linker. Most commonly agarose and Sepharose are used as matrices. Other materials, such as glycidyl methacrylate, polymethacrylate, polyacrylamide, polystyrene, and polyethyleneglycol-based resins, have also been utilised²⁰. To immobilise a natural product to the affinity support, a linker that can be conjugated with common groups such as amines, carboxylic acids, thiols, or alcohols is required. Thereby, a covalent bond between the probe and the linker or resin is established. Typically, the probe is initially derivatised with a biotinylated linker. The near-covalent affinity between biotin and avidin (see Section 3.1) allows the probe to be captured on an avidin-coated solid support. Alternatively, other haptens such as digoxigenin or the FLAG peptide are suitable to immobilise the probe to a solid support. Whatever method is chosen, the small molecule is structurally altered and this may affect its biological activity.

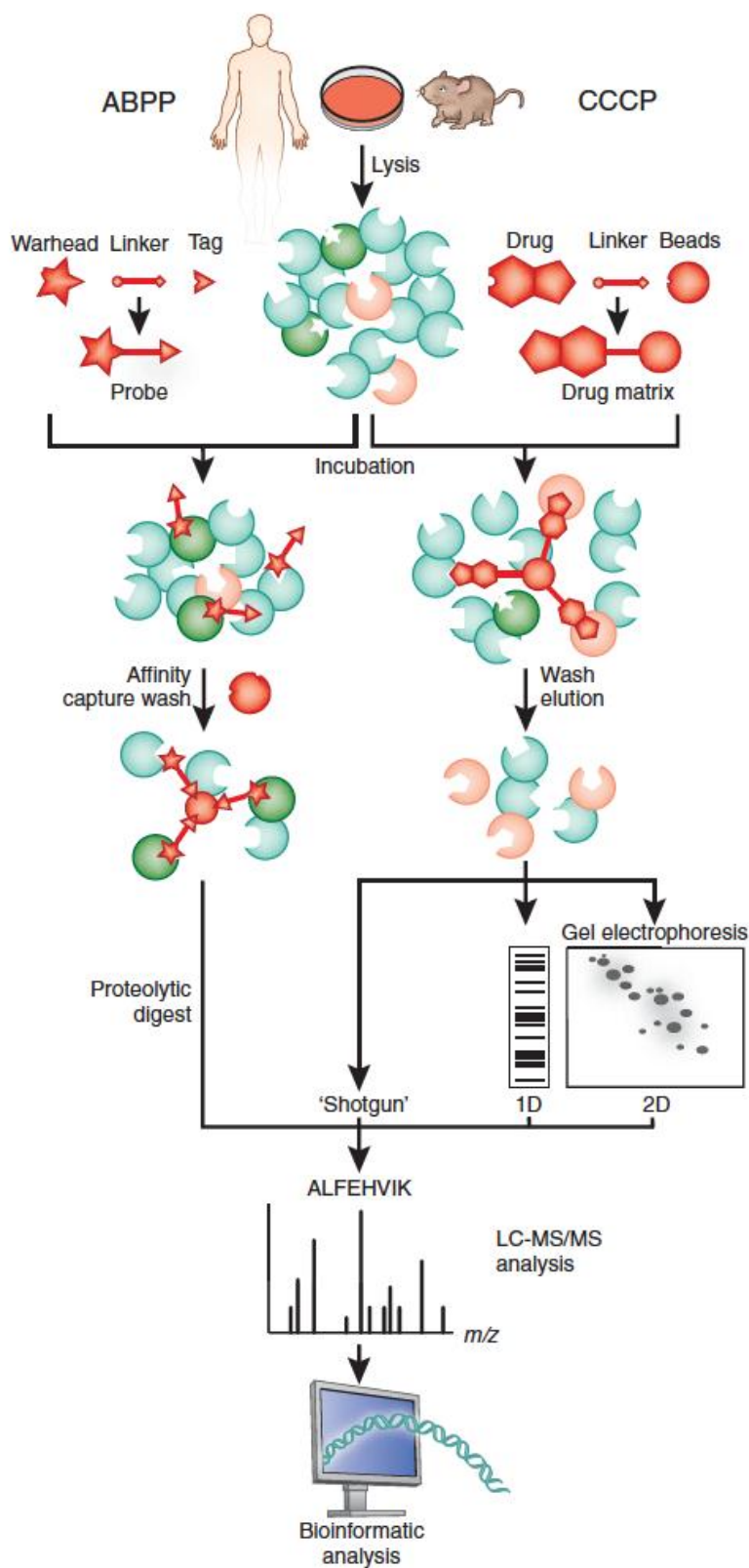


Figure 1: Two flavours of chemical proteomics: Activity-based probe profiling (ABPP) and compound-centric chemical proteomics (CCCP). Figure reproduced from Rix and Superti-Furga¹⁶, with authors' consent.

It is therefore of utmost importance to thoroughly consider the site of attachment and confirm that the affinity probe retains its biological activity.

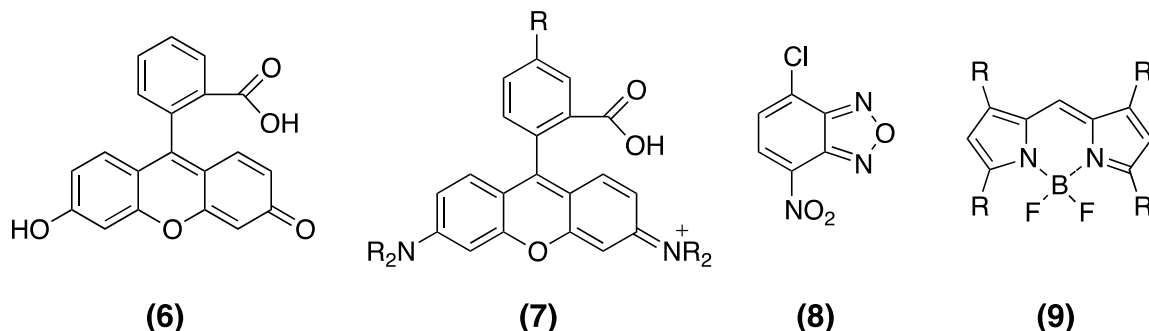
Upon immobilisation, a typical affinity experiment starts with the incubation of a cell lysate or tissue homogenate in the presence of the probe-matrix complex and protease inhibitors. Temperature and time of incubation depend on the specific sample conditions. They can vary from 4 °C to room temperature and 1 h to overnight to maximise the contact of the proteome with the displayed probe. Subsequently, the resin is washed with buffer to remove non-specific binding proteins. The addition of a small amount of detergent, such as TWEEN, usually enhances the washing efficiency and inhibits non-specific binding. However, non-specific binding is an inherent complication of all affinity-based selection strategies. To elute the specific binding proteins from the affinity support, various strategies can be used: most commonly, binding proteins are non-specifically eluted under denaturing conditions (e.g. 2% SDS); alternatively, specific elution is achieved by the use of untagged small molecule, which competitively binds to the target site and hence releases it from the solid support. The eluted proteins are then usually resolved by 1D or 2D gel electrophoresis (sodium dodecyl sulfate polyacrylamide gel electrophoresis; SDS-PAGE) and the putative binding partners excised from the gel and identified by peptide mass fingerprinting (PMF).

1.1.2.2.1.2 Activity-based probe profiling

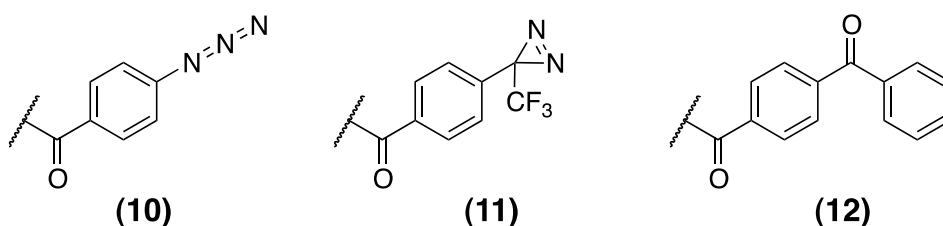
ABPP reports on the enzymatic activity of protein families, and has been successfully employed for proteases, kinases, glycosidases, oxidoreductases and others¹⁹. Using this method, an active site-directed, small molecule-based covalent probe targets a subset of the proteome, which is selected for by shared principles of binding or activity. A key to this method is the molecular design of the probe, which generally consists of three parts (see Figure 1): a reactive functional group ("warhead"; mainly electrophiles) with high affinity to the active site, a spacer (usually a hydrophilic short chain linker, such as PEG), and a tag in order to enrich, identify or visualise the target enzyme.

Warheads binding irreversibly to the target create a physical link, which for example allows for the immobilisation of the protein-probe complex to a solid support, permitting affinity isolation. This is most commonly achieved through the use of biotin as a tag. Alternatively, fluorophores, such as fluorescein (6), rhodamines (7)

nitrobenz-2-oxa-1,3-diazole (**8**; NBD-Cl) or BODIPY (**9**), can be employed as tags. These allow location of the target proteins through fluorescence microscopy or detection in gel-based read-outs (e.g. SDS-PAGE). This also requires covalent links between the warhead and active sites²¹.



However, if the warhead binds reversibly to the active site, the physical link between the probe and target protein has to be established by introducing an additional element to the probe, i.e. "tandem labelling", by using, for example, a photoreactive element as a second tag²¹. The photoreactive groups used today are generally phenylazides (**10**), phenyldiazirines (**11**), or benzophenones (**12**), which produce nitrenes, carbenes, or diradicals, respectively upon UV irradiation.



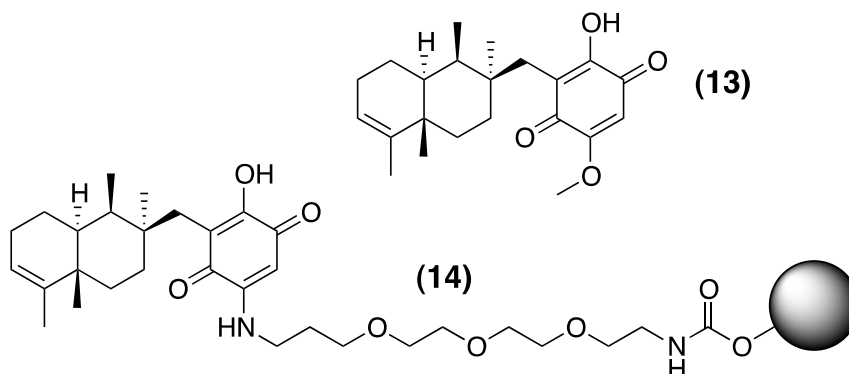
After the initial interaction between the probe and target is established, the photoreactive moiety can be activated and reacts with the nearest molecule and hence creates the physical link required for affinity purification.

ABPP is used extensively in the analysis of system-wide changes in enzyme activity such as comparing normal with different cancer states. However, if the compound's target is unknown, a shift to CCCP is required.

1.1.2.2.1.3 Compound-centric chemical proteomics

CCCP does not rely on an activity probe binding to an enzymes active site but rather a specific interaction (covalently or usually non-covalently) between small molecule and protein. It is thus much more flexible than ABPP¹⁶. As with ABPP, CCCP incorporates classical affinity chromatography methods and takes advantage of the enormous technological development in high-resolution mass spectrometry (HRMS) for the identification of target-ligand relationships.

For example, the marine metabolite bolinaquinone (**13**; BLQ) was originally isolated from *Dysidea* sp. in 1998 by Ireland and co-workers²². BLQ has been shown to be cytostatic against human colon cancer cells²³. It has also exhibited an ability to interfere with DNA and to inhibit PLA2^{24,25}.



In a compound-centric chemical proteomics approach by Margarucci *et al.*²⁶, BLQ was immobilised to an affinity support matrix with a polyethyleneglycol-based linker (**14**) and screened against the proteome of human acute monocytic leukaemia (THP1) macrophage cell lysate. The BLQ-carrying beads and underivatized control matrix were incubated with the proteome solution separately. After several washing steps, bound proteins were eluted from the beads, separated via SDS-PAGE and stained to reveal a 190 kDa protein, present only in the eluant from the BLQ-carrying beads. Further purification and enzymatic digestion identified this protein as clathrin (LC-MS/MS), an important carrier protein involved in receptor-mediated endocytosis²⁷. Target validation was performed *in vivo* by showing bolinaquinone inhibited the clathrin dependent albumin internalisation into cells. Clathrin-mediated endocytosis plays an important role in the selective uptake of proteins, viruses and other biologically relevant macromolecules at the plasma membrane of eukaryotic cells^{28,29}. The authors highlighted BLQ as a new tool to study endocytosis and clearly showed the power of chemical proteomics to isolate unknown targets in an unbiased way.

1.1.2.2.1.4 Strength and limitations of chemical proteomics

The strength of chemical proteomics lies in its ability to annotate functions to proteins and is applicable to any cell type as it relies simply on cell lysates. However, the integrity of the proteome is still a key limitation. Depending on the method, not all proteins are equally well solubilised in particular membrane proteins provide difficulties in this regard, as they do in classical proteomics. The limited protein supply and the non-iterative nature of this method make it particularly hard to

identify low abundance proteins. Other problems include low expression levels, denaturation or partial degradation during the isolation step and aggregation. However, chemical proteomics has been used to identify type I transmembrane proteins^{30,31}.

1.1.2.2.1.5 Current advances in chemical proteomics

In very recent studies, Huang and co-workers established a method they call drug affinity response target stability (DARTS)^{32,33}. This approach addresses the requirement of tagging the bioactive compound and hence eliminates one significant limitation of chemical proteomics. The concept is based on the enhanced protease stability of a protein-ligand complex compared to the unbound protein potentially preventing proteolysis^{34,35}. The shift in the thermodynamic landscape of the protein to favour the ligand-bound state was confirmed to be exploitable in DARTS³². In this study, FK506 initiated a stabilised, high-energy backbone conformation of FKBP12, which is otherwise rarely occupied. Thus, proteolytic digest resulted in reduced background (the unbound subset of the proteome was removed) and the stabilised protein-ligand complex remained for isolation and analysis. Most affinity protocols pursue a positive enrichment of the target (compared to the background proteins). In DARTS, the target protein is negatively enriched, by eliminating background and hence non-specific binders. However, since the original report of DARTS in 2009, no examples of natural product protein targets have been reported. This is clearly an intriguing, if limited, approach to isolating drug targets without the need for chemical modification of the compound.

1.1.2.2.2 Reverse chemical proteomics

The main characteristic of reverse chemical proteomics is the genotype-phenotype link³⁶, which facilitates the amplification of the gene-product through various display strategies and was first shown in phage display³⁷. An emerging field that uses this technique is the screening of random peptide libraries for sequences targeting specific proteins. The peptides displaying the desired binding properties are selected serially and can be amplified according to the physical link with the encoding gene sequence.

The expression of alien proteins on the surface of cells, organelles, and viruses can on the other hand also allow for the analysis of protein-protein interactions and studies on directed evolution of proteins (e.g. enzymes and antibodies).

For the isolation of receptors for small epitopes, such as natural products, it is arguably of advantage to employ protein libraries that represent a naturally occurring proteome, rather than combinatorial libraries. Natural products have evolved in close relationship with their naturally occurring cellular targets. Protein libraries translated from cDNA libraries most comprehensively represent those cellular targets. Hence, display technologies based on cDNA libraries are of great interest for further investigation.

Applicable technologies can be categorised as cell-based display methods (phage, bacterial, yeast, or human display) or *in vitro* technologies (mRNA, DNA, plasmid, ribosome display). Of these, phage display is the most commonly used and is probably also the best understood³⁸. Many of the above-mentioned strategies have not yet provided examples beyond "proof of concept", or are applicable for a diverse set of natural products, and are hence omitted from this review.

1.1.2.2.2.1 Cell-based display systems

Cell-based display methods include for example bacterial, yeast and mammalian cell display. Furthermore they include systems where the primary vector incorporating the cDNA library depends on a cellular setup, such as retroviral display. The latter requires mammalian host cells for replication of the retroviral particles. This display method has been reviewed recently^{39,40}. Although it is conceptually suitable for the aim to be to isolate natural product receptors it has as yet not been successful in achieving this goal beyond "proof of concept". Furthermore, this original report was based on a random peptide library and the size of displayed peptides was very limited (<8 aa).

Like retroviral display, other means (e.g. bacterial display⁴¹ or human cell display⁴²) may establish themselves in future as suitable methods for the isolation of small molecule binding proteins. The yeast two-hybrid⁴³ and three-hybrid systems⁴⁴ have also not proven suitable, as these systems have been reported to require a very low binding constant of the ligand to its receptor ($K_d < 50$ nM)⁴⁵. Interactions usually observed between natural products and binding partners will often display higher binding constants.

Thus, only those cell-based systems that provide display of cDNA libraries and have successfully yielded the unbiased isolation of receptors for natural products over the last decades are discussed hereafter.

1.1.2.2.1.1 Phage display

The first development of a phage display (PD) method dates back to the efforts of Smith, Scott and colleagues. In 1985, they reported a method for displaying polypeptides on the surface of *E. coli* specific, filamentous, M13 bacteriophage (phage)³⁷.

1.1.2.2.1.1.1 Filamentous phage

Filamentous phages replicate and assemble without killing their host. The biology of the bacteriophage is well understood^{46,47}, and they are known to assemble in the bacterial membrane periplasm and be secreted through the outer membrane. The M13 phage particles (see Figure 2) are approximately 930 nm long and consist of single stranded DNA (ssDNA) enclosed by a protein coat (capsid).

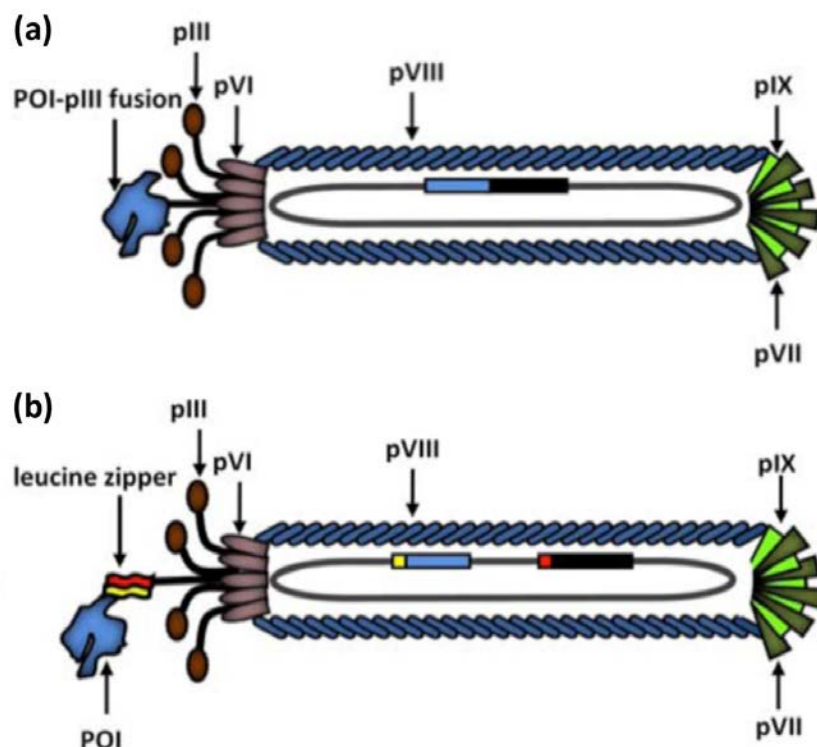


Figure 2: Schematic representation of M13 bacteriophage with different display types based on pIII coat protein. (a) The protein of interest (POI) is directly fused to a truncated pIII. (b) The POI is indirectly fused to pIII by means of a leucine zipper structure (from Georgieva and Konthur⁴⁸).

The most abundant capsid protein is pVIII, which is represented by approximately 2700 copies. Five molecules each of two minor coat proteins (pIII and pVI), cap the filaments and are involved in bacterial cell binding. They are also involved in the termination of phage particle assembly during morphogenesis. Two other minor capsid proteins (pVII and pIX) are located at the other end of the filament. There are also five copies of each and they are required for initiation and maintenance of phage assembly in the host bacteria. Today, the most commonly used format to display the

M13 bacteriophage surface involves fusion of alien DNA to the coat proteins pIII and pVIII. Fusion of alien cDNA to the M13 genome can only take place at the *N*-terminal end, as *C*-terminal derivatisation is not tolerated. As a result, the bacterial ribosome first translates the alien DNA and reaches a stop codon before translating the capsid protein. In addition, the size of the chimera is limited. For these two reasons, this system has not proven overly popular. There have been reports on *C*-terminal fusion, but these are limited to the minor coat protein pVI⁴⁹.

To bypass the *N*-terminal fusion, Cramer and Suteri⁵⁰ established a pIII-based cDNA display system which makes use of the leucine-zipper interaction of two regulatory proteins (c-Fos and c-Jun). The alien cDNA is fused to the gene encoding for one of the regulatory proteins (c-Fos) and likewise is the capsid protein coding DNA fused to the gene of the second protein (c-Jun). During replication of the M13 proteome, the leucine-zipper proteins associate in the cytoplasm and form a disulfide bond. This covalent bond links the alien protein to the pIII capsid protein before the latter is incorporated during assembly (see Figure 2). This approach however is limited in the size of the fusion protein. The assembly and infectivity of the M13 particles depends on the pIII proteins. If large proteins (> 30 aa) are displayed, a bias toward phages with no or very small DNA inserts may occur, as their replication progresses more efficiently.

Further limitations of M13-based display strategies are related to the non-lytic secretion out of the bacterial cells. As the phages have to penetrate the bacterial membrane, physicochemical characteristics (e.g. charge or lipophilicity) of the displayed proteins result in restriction of the secretion and thus bias of the selection. Overall, there are no reports on implementing cDNA libraries into M13 phage to search for natural product receptors. Related efforts, based on random peptide libraries containing M13 display, have resulted in the identification of cellular targets³⁹, but these are considered unconvincing and ambiguous and subject to the conceptual limitations outlined above¹⁵.

1.1.2.2.1.1.2 Bacteriophage T7

To address the above limitations, alternative bacteriophage display strategies were established based on lytic bacteriophage such as T4⁵¹, T7⁵² and P4⁵³. Furthermore, bacteriophage lambda has been shown to be suitable for the display of random and cDNA libraries⁵⁴. Since its invention by Rosenberg *et al.* in 1996⁵⁵, the use of T7 phage

has become the most popular method because it a lytic phage, has a rapid life cycle, is robust (perhaps too robust) and can easily display functional proteins in low copy number over 100 kD in size. Most importantly for cDNA library display, the T7 phage is a C-terminal display system⁵⁶.

T7 phage particles pack their double-stranded DNA genome into an icosahedral capsid, with an internal diameter of approximately 55 nm, with a 23 nm long and 21 nm wide tail including fibres for attachment and the initiation of infection of the host cell (see Figure 3). The detailed biology and the roles of each individual phage protein have been extensively studied and are well reviewed⁵⁷⁻⁶².

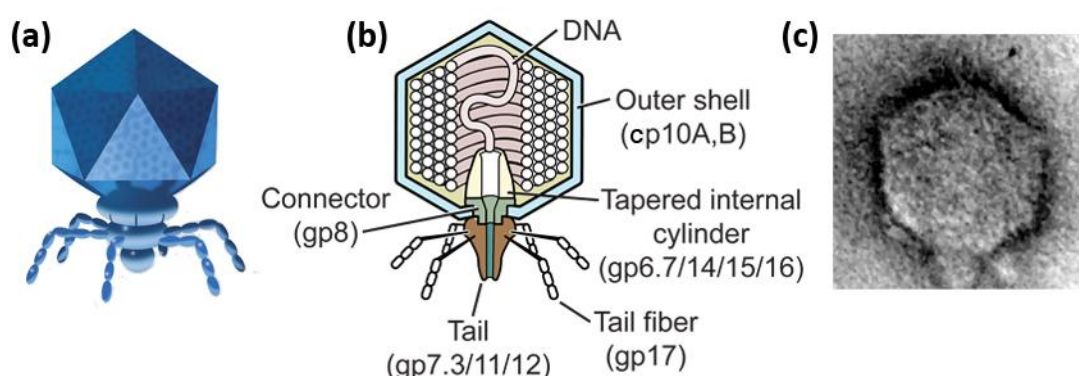
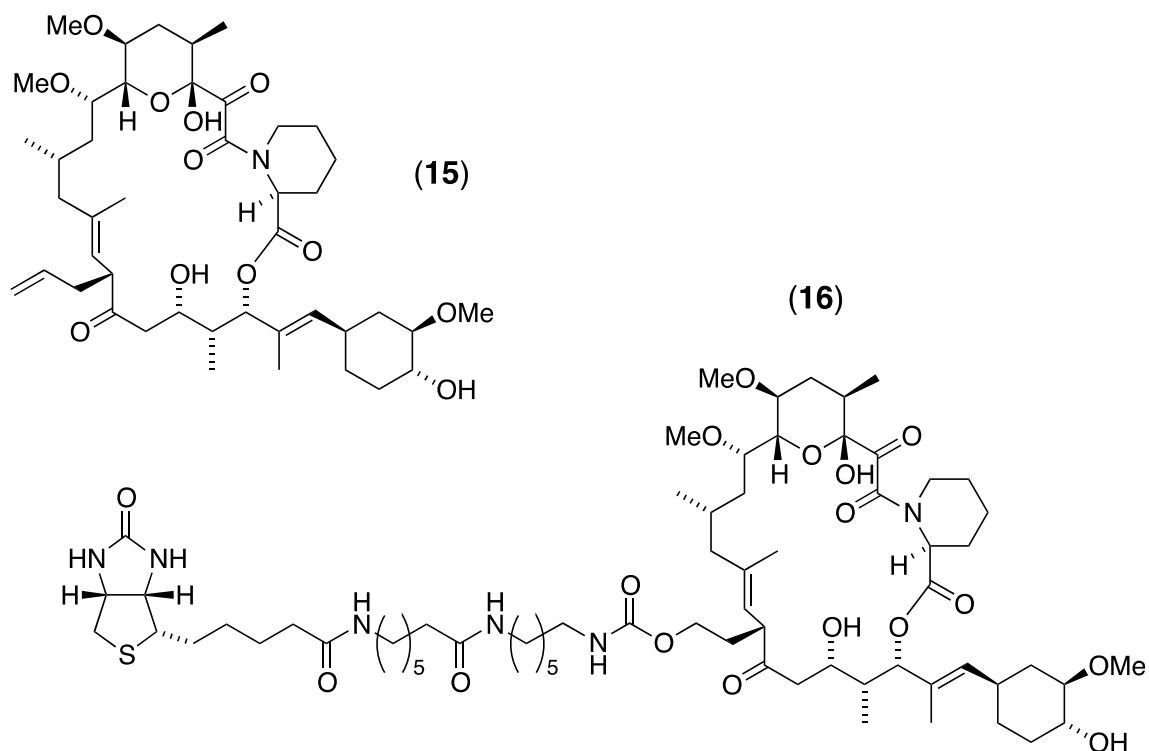


Figure 3: (a) A stylised representation of a T7 phage particle indicating the icosahedral shape of the capsid, the tail and the six fibres (Novagen Inc.). (b) A schematic representation of a T7 phage particle indicating the tight DNA packaging inside the capsid and the various proteins constituting different parts of the particle. (c) A transmission electron micrograph of a T7 phage particle (Spires and Brown, University of Texas).

In 1999, Austin and co-workers conducted the first example of using T7 phage display to isolate a natural product receptor (binding partner)⁶³. They used the immunosuppressant drug FK506 (**15**) to isolate its known target FK506-binding protein (FKBP) from a human brain cDNA library. FK506 was biotinylated to perform affinity selection. Thereby, the allyl group present in (**15**) was chosen as the site of alkylation, as it was well known that this functionality is oriented away from the active site during the binding of FK506 to FKBP⁶⁴. Several rounds of affinity selection ("biopanning") against immobilised (avidin-agarose column) biotin-FK506 (**16**) were conducted.



The T7 phage cDNA library was amplified from the host *E. coli* (BLT5403) and the lysate serially passed through two, control affinity matrices (underivatized avidin-agarose, biotin-derivatized avidin-agarose) to remove non-specific binders. After extensive washing with buffer, the lysate was loaded onto the affinity support derivatized with the biotinylated probe. Thorough washing with buffer eliminated further non-specific binders. The high binding constant between FK506 and its target FKBP ($K_d = 0.4$ nM) suggested eluting the entire complex of biotin-FK506 and the binding phage-displayed protein through application of free biotin. Thereafter, the newly gained sublibrary was transfected into fresh log-phase host cells and amplified. The phages were precipitated from the lysate, centrifuged and washed, to remove remaining free biotin. The resulting pellet was resuspended in buffer and the solution applied onto a fresh affinity column in the next round of selection. Each round of biopanning was followed by quantification of the phage titre from the eluate (i.e. plaque forming units per millilitre [pfu/mL]). Plaques appearing on a lawn of phage-transfected bacteria were randomly picked after each round. Austin and colleagues discovered a 0.450 kb phage clone to entirely dominate the round two sublibrary. This suggested the library had converged. The inserted alien DNA was isolated from the phage and sequenced, resulting in a full-length, in-frame gene encoding for human FKBP. This suggested that within only two rounds of selection with a biotin-FK506 derivatized support, phage clones displaying a human FKBP were selectively isolated

from a myriad of different clones. However, a colleague pointed out that the FKBP clone was actually a synthetic, positive-control FKBP clones and did not derive from the human cDNA library⁶⁵.

The sturdy and resilient nature of T7 phage particles causes a problem commonly observed during biopanning: cross-contamination. This necessitates laboratory precautions such as the use of disposable plastic ware, minimising aerosol formation and regular disinfection of all exposed surfaces, pipettes, incubators etc. through bleach or UV light.

Austin and co-workers repeated the experiments and took particular care to avoid cross-contamination⁶⁵. They performed seven rounds of biopanning. From the sixteen clones picked from round seven, the DNA amplification and sequencing yielded eleven containing the entire coding sequence of FKBP1a. The enrichment of the FKBP1a containing clones was already observed in round six, with five out of sixteen clones matching the FKBP1a sequence. The cross-contamination issue had successfully been resolved and thus an important "proof of concept" was established. It was clearly shown that it is possible to isolate a binding protein for a small molecule by using a cDNA library display-system based on T7 phage.

In 2006, Makowksi and colleagues investigated the diversity of M13 and T7 polypeptide-displaying libraries by implementing a bioinformatics-assisted computational approach. They found that T7 phage libraries had fewer amino acid (aa) biases, increased peptide diversity and had more normal distributions of peptide net charge, hydropathy than the M13 systems and exhibit great diversity, containing between 10^7 - 10^8 different clones⁵⁶. In comparison to filamentous phage, T7-based systems exhibit a much higher efficiency in displaying cytoplasmic proteins.

Some favourable characteristics for use in a chemistry laboratory include: a rapid life cycle that yields fast plaque formation in plate-based cell culture (~ 3 h), which enhances the overall throughput capabilities for this assay; a high number of progeny per infected cell (~ 100); and a high degree of resilience of the phage particles to severe conditions such as 1% SDS, 4 M urea, or 2 M guanidine-HCl. One remarkable feature of T7 phage is the ability to infect its host *E. coli*, even if the displayed protein on the capsid is involved with binding to the target molecule, thereby indirectly immobilising the phage to the affinity support.

As with other methods discussed, this method also has its shortcomings. As the phage-displayed proteins are expressed inside a prokaryotic cell, proteins are not exposed to post-translational modifications as in eukaryotic systems and are not, for example, phosphorylated or glycosylated. However, as shown in the filamentous phage system with a leucine-zipper based protein display, disulfide bonds can be formed during or after translation. These disulfide bonds provide a fundamental structural arrangement of the displayed protein. As with forward chemical proteomics, the prerequisite for tagging the analyte (e.g. natural product) can be limiting, especially if no SAR studies were reported for the chosen probe, to indicate which functional groups are most relevant for biological activity. In this case, it is most favourable to design and synthesise several probes to position the small molecule in various orientations to the binding partners.

Since the development of T7-phage display in 1996 by Rosenberg⁵⁵ and the "proof of concept" that this system could be used to isolate drug-binding proteins in 1999 by Austin^{63,65}, this method has led to the successful isolation of protein binding partners for several natural products (see Section 1.2).

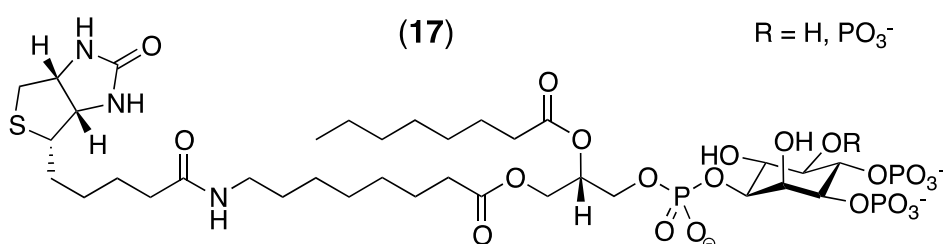
1.1.2.2.1.2 Yeast display

Yeast (*Saccharomyces cerevisiae*) is a eukaryotic single-celled organism. It was first used for cell-surface display in vaccine development by Schreuder and colleagues in 1993⁶⁶. In yeast display, the alien protein is displayed on the cell surface as an adduct to a cell wall anchored protein (e.g. α -agglutinin) as a carrier. α -Agglutinin and its counterpart a-agglutinin belong to the glucanase-extractable mannoproteins⁶⁷. They function as the mating-type specific proteins that mediate the direct cell adhesion between the opposite mating types and were reported to be located on the outermost surface⁶⁸. The agglutinin based Aga2p-Aga1p display system is the most commonly used one, although other proteins have been examined for their utility in surface display⁶⁷.

The Aga2p-Aga1p system allows for *N*- or *C*-terminal fusion of the alien protein to the Aga2p protein ("Aga2p-fusion"). During expression of these agglutinin proteins, two disulfide bonds form between them. The Aga1p-Aga2p complex is linked to the cell surface through a glycosylphosphatidylinositol-anchor attached to Aga1p. The GPI-anchor has glycan linkages and its fatty acid tails are lodged into the cell membrane.

Hence, co-expression of Aga1p and the Aga2p-fusion results in display of the alien protein on the yeast cell surface.

Yeast display has been used in antigen development⁶⁹, there are also reports of displaying cDNA libraries on the yeast cell surface⁷⁰⁻⁷⁴. To date, there is no example of isolating the receptor for a natural product. However, a recent example for isolating a binding partner of a small molecule suggests that yeast display is suitable for the isolating natural product receptor. Liu and co-workers⁷¹ first constructed a yeast surface-displayed human cDNA library and then selected for yeast cells displaying binding proteins for a biotinylated phosphatidylinositive (**17**) probe.



Incubation of the small molecule probe with cDNA library displaying yeast cells allowed for interaction of the probe with specific binding proteins. Subsequently, the biotinylated probes were fluorescently labelled through incubation with (fluorescent) phycoerythrin-streptavidin and selected for via fluorescence-activated cell sorting (FACS). Three rounds of selection were performed and the resulting sublibrary analysed for the cDNA inserts integrated into the yeasts' plasmids. The majority of isolated clones contained a full-length pleckstrin homology (PH) domain that was known to code for phosphatidylinositol binding. This approach should be directly transferrable to biotinylated probes used in chemical proteomics (forward and reverse) and offers various advantages over for example phage display.

As yeast is a eukaryotic cell, some post-translational modifications (e.g. phosphorylation) are incorporated during protein display. The protein folding and secretory machineries (e.g. endoplasmic reticulum) are similar to those occurring in mammalian cells. Thus, this system provides a more functional display of mammalian proteins than for instance phage display. Furthermore, in yeast display 104-105 identical copies of the alien protein can be displayed on each yeast cell, which provides a benefit during the selection and analysis steps. The analysis via FACS constitutes a major advantage. As the yeast cells display many copies of the alien protein, the target-displaying cells can be heavily fluorescently labelled and easily

sorted from cells not displaying probe-binding partners. This allows for fast specific selection unlike phage, which are generally non-specifically eluted and are too small to be FACS sorted.

However, they grow much slower than for example bacteria, which clearly slows down a high throughput approach. More importantly, the transformation rates are lower as compared to for example phage display and it is not guaranteed that the entire genome is covered when expressing the transcriptome. It is also discussed that yeast displays a biased use of codons compared to humans (when constructing human libraries). This may result in an expression bias for codons rarely used in yeast and thus cause protein sequence mutations, where certain leucine residues are replaced by serines in some proteins⁷⁵.

Significant technical advances since its first report in 1993 and inherent advantages over other means (e.g. post-translational modifications), suggest that this method is suitable as a means to isolate cellular targets for natural products in near future.

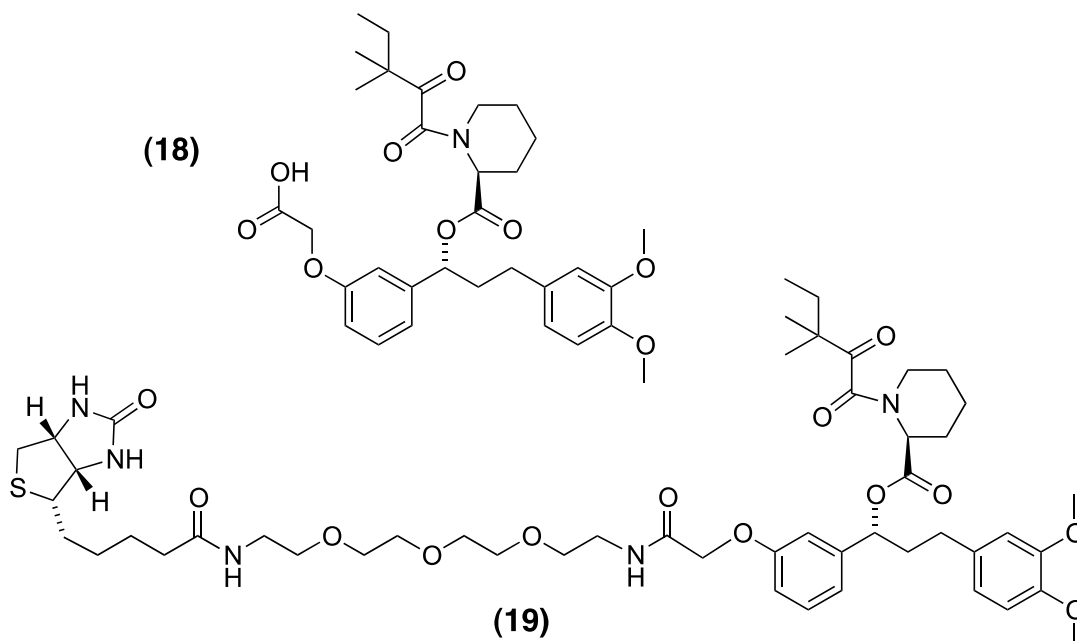
1.1.2.2.3 *In-vitro* display technologies

The genotype-phenotype link in *in vitro* display technologies refers to the connection between the mRNA and its transcribed protein. Here, a chemical process, the polymerase chain reaction (PCR or real time-PCR), transcribes the mRNA into the encoded proteins. This brings both benefits and shortcomings. A clear advantage is the size of the constructible library ($> 10^{14}$), which exceeds the dimensions for any cDNA libraries by 10^9 times. Yet, considering the overall dimension of sequence space, there are for example 10^{104} possible 80-mer peptides, which is more atoms than exist in the known universe. Hence each man-made library has to be considered small. In cell-based assays (e.g. phage or yeast display), the involvement of living cells restricts library generation during biopanning due to the maximum concentration of cells achievable in any used volume. However, cDNA libraries have the major advantage of being selected for function rather than by chance.

Some techniques used for *in vitro* display of proteins are for example ribosomal, mRNA, DNA and plasmid display, which have been reviewed recently^{15,76,77}. Each method has its distinct advantage and disadvantage, with some (e.g. mRNA display) showing potential as an alternative means to isolate cellular targets for natural products. Yet, none of these methods have so far been employed to achieve this goal, and thus are not part of this review.

1.2 Previous work on cDNA libraries in T7 phage display

Briefly after Austin's initial experiments with FK506, the research group went on to perform phage display on AP1497 (**18**), a synthetic analogue of FK506⁷⁸. In this experiment, streptavidin-coated polystyrene microtitre plates were used as solid support instead of the previously used streptavidin-coated agarose beads.



The biotinylated compound (**19**) was immobilised and screened against a human brain T7 phage-displayed cDNA library. After three rounds of selection, 96 random plaques were chosen and tested for their ability to bind to an AP1497-derivatised plate. From a selection of 30 clones, which showed the highest affinity, 22 clones contained one of three different FKBP proteins (FKBP1a, FKBP1b, or FKBP2), all of which are known receptors for FK506. Considering the diversity of the cDNA library (10^6 - 10^7 different clones), this experiment proved several points: it confirmed the selectivity of T7 phage display to isolate the most avid binding partners over the most abundant clones; despite the non-specific binders present in the initial library, this isolation can already be achieved with the level of derivatisation attainable on the microtitre plate surface; and, if multiple avid binding partners are present, they are co-selected and isolated.

1.2.1 Platforms for phage display selection

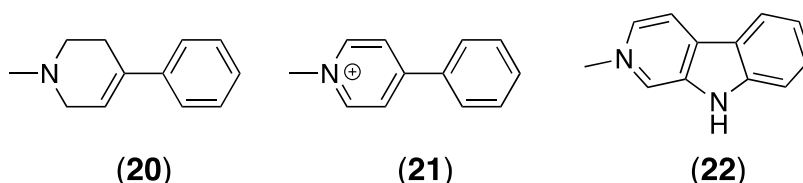
For the immobilisation of small molecule probes ("baits"), the microtitre-based method has become the most popular technique⁷⁸⁻⁹³. Its multi-well format allows screening several biotinylated probes in parallel against various cDNA libraries, and

using multi-channel pipettes for rapid throughput. The required material is comparably cheap and the disposable nature of the microtitre wells minimises cross-contamination concerns. Yet, it is generally rather labour intensive as the washing and elution steps have to be manually performed through repetitive pipetting. A few studies have employed bead-based methods that can be distinguished into two distinct forms: either affinity chromatography, or as a batch method. In affinity chromatography, the selection is performed as continuous flow of phage lysate through a set of affinity columns filled with bait-carrying beads (for detailed description see Section 1.1.2.2.1.1.2)⁹⁴⁻⁹⁷. In the batch approach, the solid support beads or resin are incubated with the phage lysate and thereafter pelleted through centrifugation to pull down phage particles displaying specific binders⁹⁸⁻¹⁰⁰. This “pull-down” approach is less frequently used as it requires several more steps¹⁰¹. An alternative technology, which was first reported for phage display in 1999¹⁰², is based on a quartz crystal microbalance (QCM) biosensor. It was further developed for T7 phage display-based, small molecule bait systems by Sugawara and co-workers in 2007¹⁰³. The cuvette type biosensor is a very sensitive and accurate mass measuring device that works by measuring changes in resonant vibrational frequency caused by a mass increase. The biotinylated probe can be immobilised on the avidin-coated biosensor and the absolute binding of the phage library to the probe can be monitored in real time. After washing the biosensor to remove non-specific binders (and still monitoring the frequency variations through the QCM device), a fresh log-phase *E. coli* was directly reinfected in the cuvette, thus eliminating the elution step. This method provides a few key advantages. Most importantly, the optimal phage absorption could be determined by monitoring the frequency decrease. This technique thereby avoids the need for determination of the washing and elution steps. However, the required instrumentation is very costly and requires expertise for operation, as opposed to a microtitre plate format. The same research group also developed a high-throughput version of affinity chromatography^{104,105}. This technique differed from generic affinity chromatography in the way that the column wash-off and the eluate were collected in a 96 well plate format. Subsequent real-time PCR (SYBR green) indicated which fractions contained high DNA (therefore phage) concentrations. These fractions were either subjected to further rounds of selection or sequenced. Like the QCM biosensor method, the washing and elution steps did not

have to be optimised. Although the authors reported this selection procedure with a synthetic small molecule, there is no reason why it should not be applicable to natural products as well.

1.2.2 Examples for the identification of targets of small-molecule natural products

Gearhard and co-workers investigated binding partners for *N*-methylated β -carboline¹⁰⁶. The premise for their work was the β -carboline-mediated cytotoxicity and possible etiologic factors of these compounds for Parkinson's disease. The structural similarity to parkinsonian-inducing toxins such as 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP; **20**) and its brain cell metabolic product 1-methyl-4-phenylpyridinium cation (MPP; **21**) led the authors to utilise 2-methylnorharman (**22**) as probe in a T7 phage biopanning approach with a human brain cDNA library.

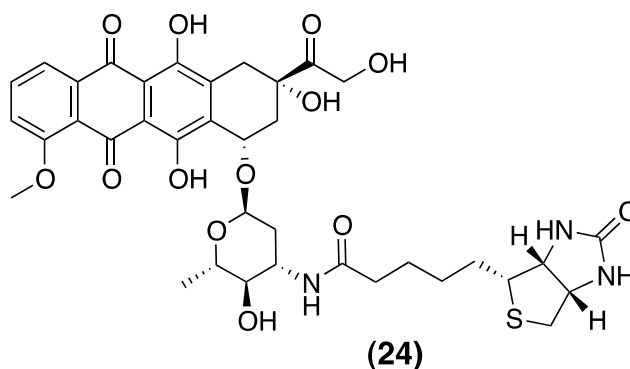
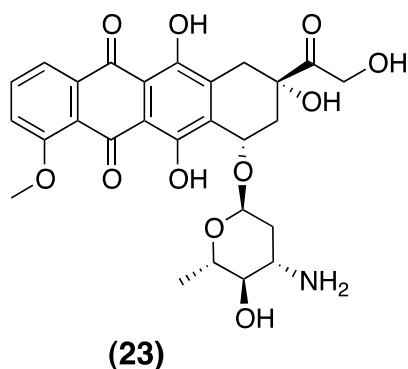


Therefore they immobilised the hydrochloride salt of **22** directly onto a nitrocellulose matrix, which was added as discs into regular microtitre plate wells as an affinity support. After six rounds of selection the sublibrary did not converge. Regardless, 43 single plaques were picked, their DNA amplified and sequenced. For the majority of clones, homology searches revealed similarities to human proteins with unknown function or identities. The remaining sequences encoded fragments of six known human proteins: human nucleolar phosphoprotein p130, dorfins, α -tubulin, paraoxonase 2, fatty acid-binding protein 5, and platelet-activating factor acetylhydrolase 1B1. The protein sequences were analysed *in silico* to predict physico-chemical properties of the toxin-binding sequences. The alignment of all six sequences did not reveal any clear consensus. However, three major characteristics for all six isolated proteins were described: abundance of highly charged amino acid residues, in particular glutamic and aspartic acids; net negative charge; and presence of hydrophobic amino acids.

Although it is possible that 2-methylnorharman binds to highly charged residues, the presence of such highly charged protein fragments may as well be due to non-specific binding to the nitrocellulose matrix. The setup did not include a control probe to remove non-specific binders before binding to 2-methylnorharman, nor did the

authors validate to what degree the phage-displayed protein fragments would adhere to the affinity support itself.

Doxorubicin (**23**) is a cytotoxic anthracycline antibiotic, first isolated in 1963 from cultures of the actinomycete *Streptomyces peucetius* var. *caesius*¹⁰⁷. Yu and co-workers used doxorubicin in a T7 phage display assay and screened a human liver cDNA library for binding partners⁷⁹.



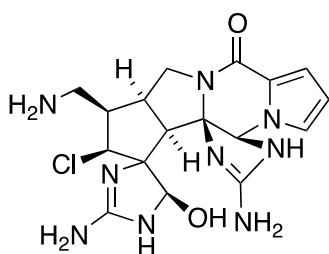
The selection progressed in a microtitre plate format and bound phage were eluted with 1% SDS, before fresh *E. coli* was reinfected for the next round of selection. After four rounds of selection two major clones were obtained. The most dominant clone (18 out of 20 plaques) was about 600-700 bp long, but expressed only a 20 aa sequence, which was also out of frame. The other 2 phage clones contained inserts of about 870 bp and encoded for an in frame, 216 aa long sequence of the C-terminal section of human nucleolar phosphoprotein, hNopp140. The authors then overexpressed the protein sequence from the 870 bp cDNA insert by cloning it into *E. coli*, to verify the binding between the protein and doxorubicin in two different assays. They employed surface plasmon resonance and fluorescence spectroscopy. In both assays, the signal intensity (either fluorescence or binding on SPR chip) increased. Comparative binding studies with the phosphorylated hNopp140 and doxorubicin revealed that neither the fluorescence intensity increased, nor did the authors observe specific binding to immobilised doxorubicin in the SPR studies. The authors determined a dissociation constant for doxorubicin to hNopp140 in the low micromolar range.

The compound was biotinylated without a linker between the NP and the biotin-tag (**24**). The immobilisation of a probe directly onto the solid support raises concerns, as highlighted earlier (i.e. with *N*-methylnorharman¹⁰⁶).

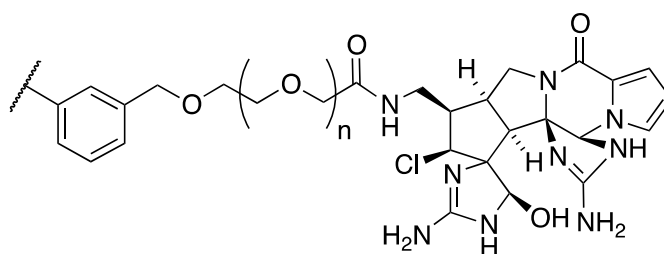
The lack of a spacer may cause steric hindrance of the protein binding part to approach or fold around the bait (see Section 3.2.2). If the small molecule is not

orientated away from the solid support, non-specific binding between the phage and the affinity matrix become an inherent problem.

Interestingly, Piggott and Karuso also isolated hNopp140 while screening a human liver cDNA library, although their studies aimed to isolate a binding partner for the marine natural product palau'amine (**25**)⁸⁴. For this experiment, the natural product was covalently immobilised onto a polystyrene microtitre plate, but a polyethyleneglycol-based linker was incorporated to distance the probe from the solid support. After five rounds of selection (which were performed in similar fashion to Yu and co-workers experiment⁷⁹), only a minor number of the clones isolated had the cDNA insert orientated correctly and displayed proteins in frame. Of these, some displayed identical homology with hNopp140 compared to the clones isolated by Yu *et al.*



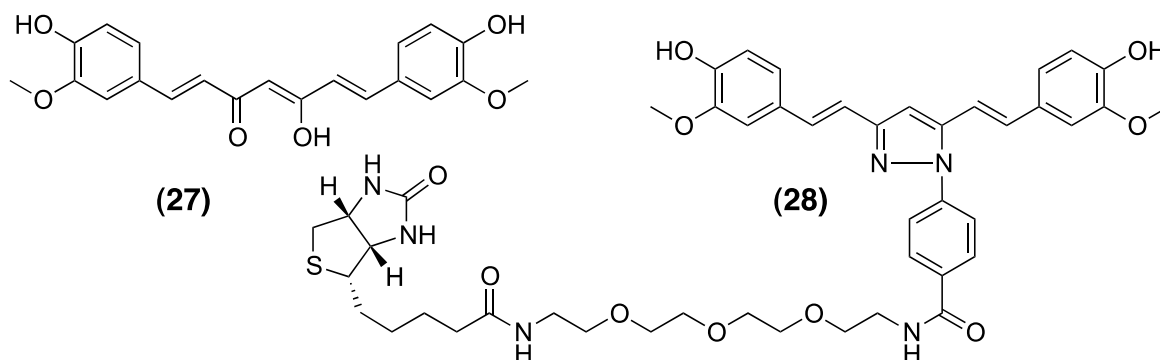
(25)



(26)

This is highly suspicious and the authors concluded this finding was likely based on non-specific interaction of hNopp140 as promiscuous binder with the solid support⁸⁴. Considering the high number of charged amino acids in hNopp140, with long stretches of serine residues, and the low abundance of the corresponding clones in both reported experiments make this explanation very plausible. The isolation of highly charged human nucleolar phosphoprotein from norharman, doxorubicin and palau'amine must be considered suspicious. It is unlikely that this is the target for any of these natural products and that it is instead simply a promiscuous interaction that highlights the major pitfall of phage display-based receptor isolation. Clearly steps need to be taken to clear these types of interactions before meaningful results can be expected from this technique.

HBC, a synthetic derivative of curcumin (**27**) that exhibits potent inhibitory activity against the proliferation of several tumour cell lines, was used in T7 phage display by Kwon *et al.*⁸⁰ in 2004. The authors immobilised a biotinylated analogue of HBC (**28**) in a microtitre plate-based format and screened five different human cDNA libraries.

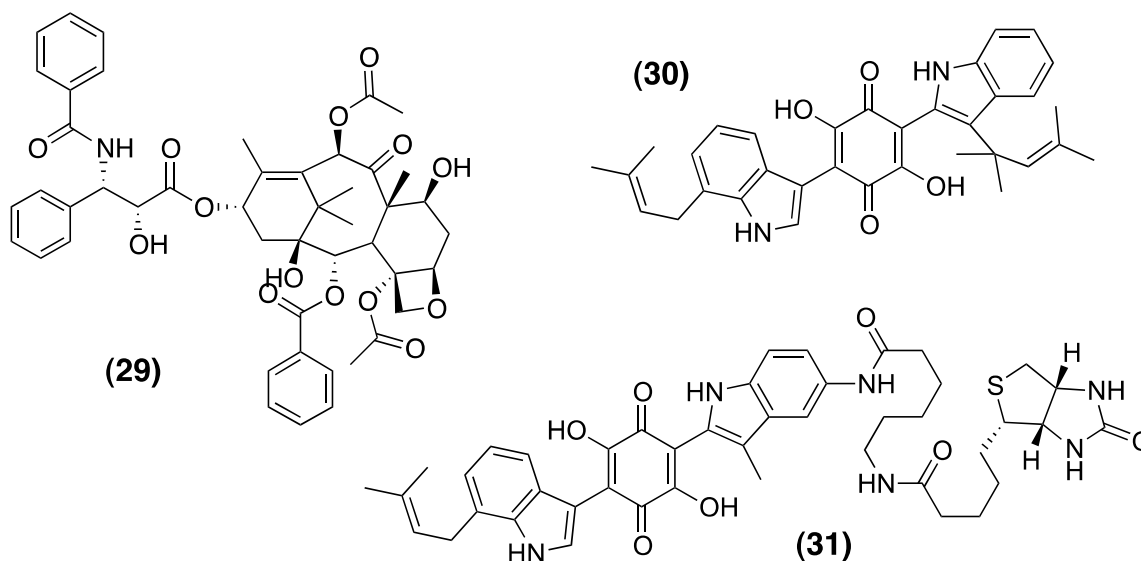


After four rounds of selection, the HBC-binding phage clones were analysed and revealed a human calmodulin fragment as potential binder. The authors⁸⁰ performed binding studies with surface plasmon resonance (SPR) and showed a clear dose-response. Additional docking studies also suggested calmodulin could be a real target for HBC. However, the isolated fragment only covered a minor fraction of calmodulin and displayed a high number of charged amino acids.

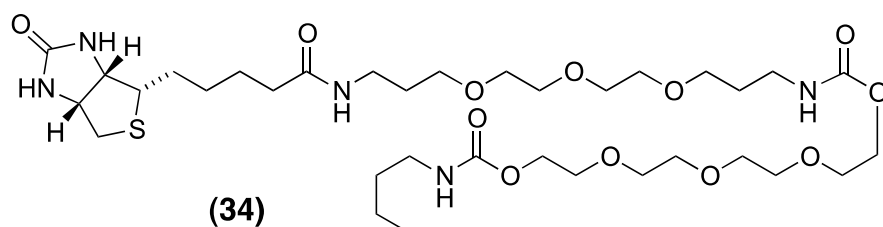
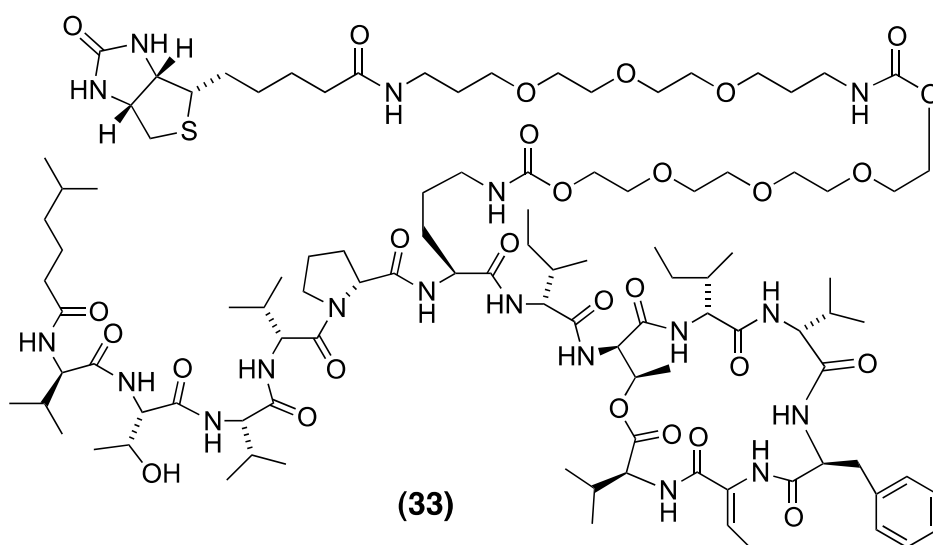
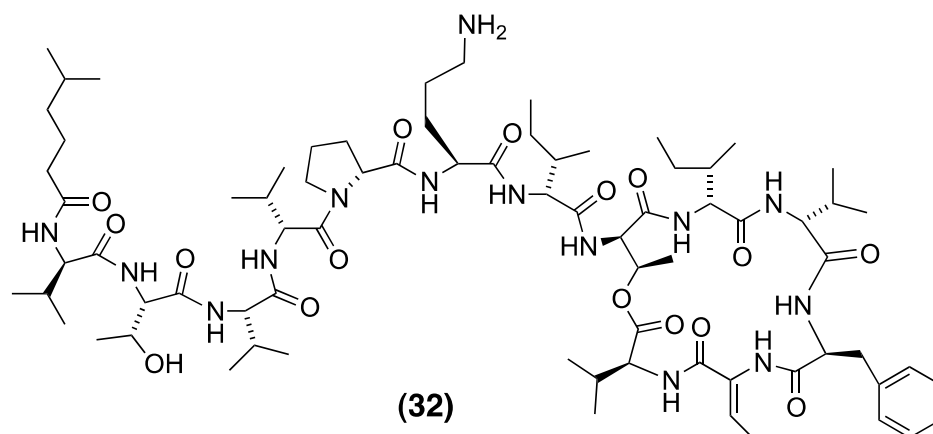
Paclitaxel (**29**), a mitotic inhibitor isolated from the bark of *Taxus brevifolia*, exhibits a well-documented tubulin-stabilising effect. Yet, the causes of many observed side effects are still elusive. In 2006, Austin *et al.* found that nanomolar concentrations of paclitaxel, mimicking that found in cancer patients, induced cytosolic Ca^{2+} oscillations in human neuronal cells⁹⁵. These oscillations were independent of extracellular and mitochondrial calcium but dependent on intact signalling through the phosphoinositide signalling pathway. They also screened a C7-biotinylated analogue of paclitaxel against a T7 phage human brain cDNA library and isolated neuronal Ca^{2+} sensor 1 (NCS-1) as binding protein. In a knockout chemical genetics approach, the selectivity of paclitaxel to NCS-1 was verified in comparison to other biotinylated natural products. This example shows that phage display is also suitable to detect off-targets.

Demethylasterriquinone (DAQ) B1 (**30**) is derived from *Aspergillus terreus* and acts as orally active insulin mimic^{108,109}. Pirrung and co-workers identified glyceraldehyde 3-phosphate dehydrogenase (GADPH) as binding partner by screening a biotinylated DAQ analogue (**31**) against a T7 phage human liver cDNA library⁸⁷. After five rounds of selection, multiple phage clones encoding for a GAPDH fragment were identified. As GAPDH reportedly binds to the glucose transporter and disrupts phosphatase(s) acting on PI lipids, it thereby potentiates insulin signalling via the PI3Kinase pathway. Hence, the identification of GAPDH as putative target for DAQ is interesting and directs further research on DAQ into therapeutic relevance of

this interaction. However, the authors did not perform any binding studies (SPR *etc.*) and therefore quantification of the interaction between DAQ and GAPDH are required in the first instance.

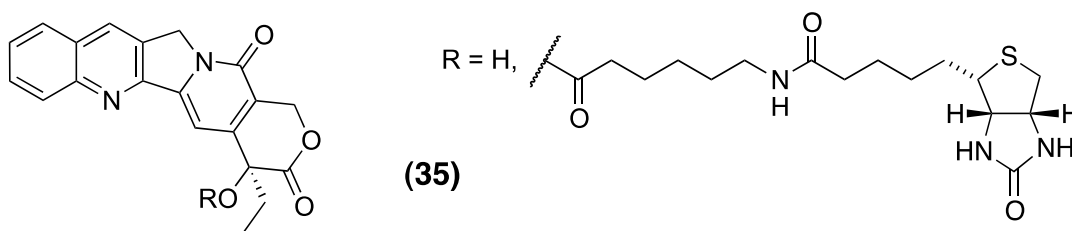


The marine natural product Kahalalide F (**32**; KF) was first isolated from the marine mollusk *Elysia rufescens* in 1993¹¹⁰. Its synthetic analogue elisidepsin is currently in phase II clinical trials as anticancer drug against a range of solid tumours. Piggott and Karuso⁸⁹ employed a biotinylated analogue (**33**) and screened three human cDNA libraries. Additionally, the authors constructed a fluorescently labelled kahalalide probe (KF-NBD), which was used for fluorescence microscopy to determine the localisation of the natural product in specific features of the cytoplasm. Both probes were shown to retain their cytotoxic activity despite derivatisation at the ornithine side chain. In addition, the authors constructed a control probe (**34**), mimicking the ornithin chain of KF. This species was utilised to remove non-specific binders before biopanning on biotin-KF was performed. The phage display studies revealed ribosomal protein S25 (RPS25) as the dominant clone from all libraries tested after nine rounds of biopanning. SPR analysis confirmed specific binding of kahalalide F to RPS25 and further suggested RPS25 as a putative target for KF. From these results, the authors postulated that KF was involved in a hitherto unidentified signalling pathway that triggers a series of events resulting in cell lysis and oncosis.



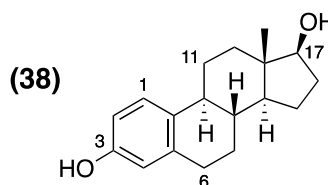
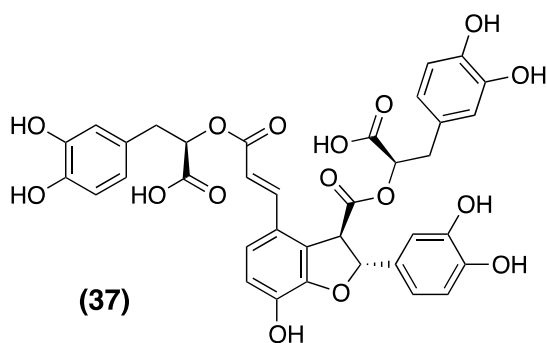
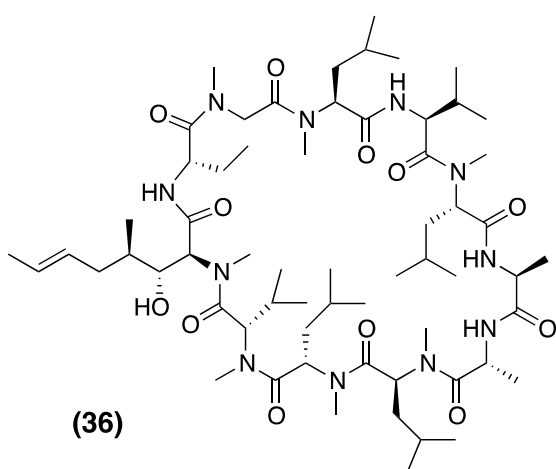
The authors also constructed the first prokaryotic T7-phage libraries and therein explored potential binding partners for a biotinylated FK506 analogue. Their efforts resulted in gDNA libraries from *Pseudomonas stutzeri* and *Vibrio fischeri*, and a cDNA library from *Arabidopsis thaliana*. They furthermore isolated several new FKBP from these environmental DNA libraries⁹¹.

Sugawara and co-workers reported several examples of phage display selections for small-molecule binding partners. They performed several “proof of concept” studies for new PD platforms^{85,103,111} and used camptothecin, a cytotoxic quinoline alkaloid (35) from *Camptotheca acuminata*⁸⁵, as the model compound.



In 2005, the authors screened a biotinylated camptothecin analogue against a random peptide library in a microtitre plate-based setup to isolate an octamer peptide sequence (NSSQSARR) as binding motif. The same probe was later immobilised on a biosensor and screened in a new experimental setup based on a quartz crystal microbalance (QCM) approach¹⁰³. After three rounds of selection with the QCM biosensor, the rescued clones revealed the same displayed binding motif as previously identified. Since then, Sugawara *et al.* have repeatedly employed the QCM technology to isolate protein-binding motifs for several bioactive compounds, such as methotrexate¹¹² and trimannoside¹¹³.

The simultaneous isolation of two cellular targets was previously mentioned (see Section 1.2). Another such example was reported from T7 phage experiments with cyclosporine A, an immunosuppressant drug first isolated from the fungus *Tolypocladium inflatum*. In 2008, He *et al.* co-isolated cyclophilin A and B from their biopanning experiments with a T7 phage human brain cDNA library⁹⁶. Both proteins are known targets for cyclosporin (36), with binding constants of approximately 0.5 μM . The compound was covalently linked to Ni-nitrotri-acetic acid polymer beads. The authors precleared the phage lysate with underivatised matrix. The elution was omitted and the host *E. coli* directly infected with the affinity matrix.



The polyphenolic oxidant salvianolic acid B (**37**) originates from *Salvia miltiorrhiza* and is held accountable for the anti-atherosclerotic effects of the plant extract¹¹⁴. In 2010, Xu and co-workers reported the implementation of magnetic beads as solid support in T7 phage display¹⁰⁰. During biopanning, the beads can be retained magnetically and the supernatant removed. In this study, several clones were identified as binding partners, but no clear target was described for salvianolic acid B. Very recently, Van Dorst *et al.* searched a T7 phage human breast cDNA library for protein binding partners of the human sex hormone 17 β -estradiol (**38**)⁹². After three rounds of selection, biopanning with biotinylated 17 β -estradiol yielded two major phage clones, autophagy/beclin-1 regulator 1 (beclin 1) and ATP synthase F0 subunit 6 (ATP6). Specific binding between the corresponding proteins and estradiol was validated in SPR studies. However, known cellular targets of estradiol, such as estrogen receptor α (ER α) and estrogen receptor β (ER β), were not isolated from the cDNA library. The authors argue that the isolation of ER α and ER β was not expected from this T7 phage display experiment. In this *E. coli*-based expression system full length ER α and ER β will be absent, as the corresponding DNA binding domain is highly toxic to the host organism. Such restrictions in functional display of library members is certainly a shortcoming of this otherwise relatively unbiased technology. However, the two isolated clones are proposed as potential off-targets. Beclin 1 is an essential mediator in autophagy, a catabolic process occurring in unfavourable environmental conditions, that involves the lysosomal degradation of cellular compartments. F0F1-ATP6 synthase is a key mitochondrial enzyme that modulates the cellular energy metabolism. Another important point to be made here is that contrary to popular belief, phage display is capable of isolating membrane proteins if the relevant intra- or extra-cellular domains are represented in the library. Other examples include membrane proteins such as UQCRB (mitochondrial inner membrane) and the α -domain of F1-ATP synthase, were identified as binding partners for terpestacin and a synthetic ATP mimic, respectively^{94,115}.

1.3 Relevance of studies

Today, there are a variety of methods available to help elucidate the molecular mechanisms underlying the mode of action of natural products. Each of these methods has its benefits and shortcomings. Phage display is still in its developmental stage and is clearly no different, but holds the potential to reveal new protein binding

partners for small molecules in a way that cannot be achieved by other methods. It will not work for every system but neither will any other genome-wide technique. Specific shortcomings include the lack of post-translational modifications, codon usage and the expression of protein complexes or membrane proteins.

Reverse chemical proteomics is an exciting technique to identify cellular protein targets of small, biologically active molecules. In particular, the combination of affinity chromatography and phage display allow for the isolation of the most avid protein-binding partner for the investigated probe and thereby avoid the problem of high abundance, low binding affinity interference common in forward chemical proteomics. Given the urgent need to rapidly identify cellular targets for compounds with known bioactivity, this technique can potentially speed the drug discovery process and help re-fill drug pipelines.

This project therefore aims to elucidate protein-binding partners for three structurally diverse natural products with hitherto unclear cellular targets. A secondary aim is the bioassay guided isolation of natural products from marine organisms on a milligram scale. Hence, isolation and structure elucidation of marine derived natural products is a first step in the global drug discovery process.

Chapter 2

Marine Natural Products

2.1 Introduction

Secondary metabolites are not waste products. They play an important role in the survival strategies of the organisms that produce them or their symbionts¹¹⁶. Although not directly required for the survival of the individual, these metabolites reportedly help the organism communicate with its environment, assist for reproduction, or protect the organism by deterring predators¹¹⁷. Secondary metabolites clearly exhibit biological potential. This has long been employed by humans, in their consumption of digestible plants and spices, their use of remedies made out of plants or animals and in their extraction of purified compounds for medicinal purposes¹¹⁸. The ease of access to plants and other land organisms has resulted in a constantly expanding knowledge of terrestrial natural products. Nonetheless, only an estimated 10% of the biodiversity has been investigated and tested for bioactivity in at least one context, leaving many useful natural compounds yet to be investigated for potential drug leads¹¹⁹. The gulf between available and known natural products (NP) is even bigger for those derived from marine sources, as attempts to investigate marine natural products emerged only in the mid-20th century. This leaves marine natural products (MNP) as a vastly untapped source of both new compounds and compound classes¹²⁰. Hence, drug discovery can directly benefit from investigations into marine natural products and their biological activity. The first step in reverse chemical proteomics is the isolation of biologically active natural products (see 1.1.1). Laboratory experience in isolation and structure elucidation of small secondary metabolites is therefore essential for this discipline. Thus, although not the major focus of this thesis, an in depth engagement with MNP chemistry is provided within this chapter.

2.1.1 Marine natural products

Bergmann and Feeney (1951) were the first to describe the potential of marine natural products to medicine by the isolation of unusual *arabino*- and *ribo*-pentosyl nucleosides from a sea sponge¹²¹. This led subsequently to the clinically used derivatives ara-A (**39**) and ara-C (**40**) as anticancer and antiviral agents, respectively, that are still in use today. Great efforts to isolate and characterise new marine natural products have been made ever since. Aside from a basic scientific interest resulting from the grand challenge to catalogue every naturally occurring small molecule,

commercial interests have driven the search for new antimicrobial, anticancer, antiprotozoal and herbicidal compounds¹²².

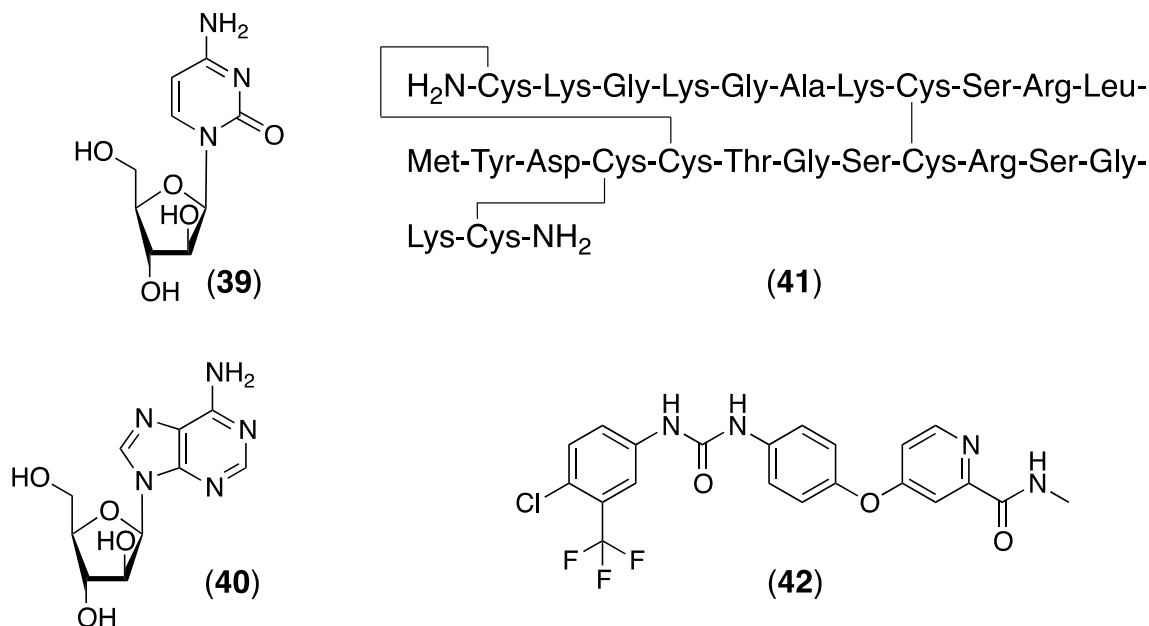
The first unaltered, marine natural product drug approved by the FDA was ω -conotoxin MVIIA (**41**) from the marine snail *Conus magus*. ω -Conotoxin MVIIA is one of many peptides found in the venom of this snail, and was marketed in 2004, as ziconotide, for chronic pain from spinal cord injuries¹²³. Since then, several other marine compounds have entered the pharmaceutical market or are currently in clinical trials^{124,125} (see Table 1).

Table 1: Marine natural products marketed as drugs or drug candidates in clinical trials (reproduced from Mayer *et al.*^{125,126}, and updated from the NIH database "clinicaltrials.gov"¹²⁷).

Clinical Status	Compound Name [year of approval]	Trade-mark	Marine Organism	Chemical Class	Disease Area
FDA-Approved	Cytarabine (Ara-C) [1969]	Cytosar-U®	Sponge	Nucleoside	Cancer
	Vidarabine (Ara-A) [1976]	Vira-A®	Sponge	Nucleoside	Antiviral
	Ziconotide [2004]	Prialt®	Cone Snail	Peptide	Analgesic
	Eribulin Mesylate (E7389) [2010]	Halaven®	Sponge	Macrolide	Cancer
	Omega-3-Fatty Acid Ethyl Esters [2004]	Omacor® Lovaza®	Fish	Omega-3 Fatty Acids	Hypertriglyceridemia
	Trabectedin (ET-743) EU Approved only [2007]	Yondelis®	Tunicate	Alkaloid	Cancer
	Brentuximab Vedotin (SGN-35) [2011]	Adcetris®	Mollusk	Antibody-Drug Conjugate (MM Auristatin E)	Cancer
Phase III	Plitidepsin	Aplidin®	Tunicate	Depsipeptide	Cancer
Phase II	DMXBA (GTS-21)	N/A	Worm	Alkaloid	Cognition, Schizophrenia
	Plinabulin (NPI 2358)	N/A	Fungus	Diketopiperazine	Cancer
	Elisidepsin	Irvalect®	Mollusk	Depsipeptide	Cancer
	PM00104	Zalypsis®	Nudibranch	Alkaloid	Cancer
	Glembatumumab Vedotin (CDX-011)	N/A	Mollusk	Antibody Drug Conjugate (MM Auristatin E)	Cancer
	PM01183	N/A	Tunicate	Alkaloid	Cancer
Phase I	Marizomib (Salinosporamide A)	N/A	Bacterium	Beta-Lactone-Gamma Lactam	Cancer
	PM060184	N/A	Sponge	Unavailable	Cancer
	SGN-75	N/A	Mollusk	Antibody Drug Conjugate (MM Auristatin F)	Cancer
	ASG-5ME	N/A	Mollusk	Antibody Drug Conjugate (MM auristatin E)	Cancer

Abbreviations: FDA = Food & Drug Administration, N/A = not available.

Research by major pharmaceutical companies in this field declined in the mid 1990s, coinciding with and, most likely, driven by the establishment of combinatorial chemistry. Combinatorial chemistry promised an endless opportunity to design new molecules and thereby create new marketable compounds¹²⁴. For various reasons, this promise has not been realised, leaving the pharmaceutical industry unable to meet the high demand for new structural entities displaying high levels of bioactivity.



In numerous papers over the last 10 years, Newman *et al.* elaborately highlighted the importance of natural products for the development of drugs against microbial infections and particularly against cancer^{4,5,128-130}. Newman argues that approximately 80% of all anticancer drugs are natural products or derived therefrom⁵. Since the release of ziconotide in 2004, a combination of factors has produced a minor renaissance in the niche field of MNP research. Combinatorial methods failed to produce new lead compounds (only one new drug based on combinatorial chemistry has ever been released in its 25-year history (sorafenib; **42**)⁴. However, over the years, the analytical and instrumental improvements have allowed analysis of ever-smaller samples. Complete structural characterisation can now be achieved using quantities of less than one milligram. High throughput screening (HTS) approaches, using microgram quantities of a sample, allow rapid evaluation of biological activity. Compounds are now efficiently tested against multiple cell types or in cell free assays. The successful marketing of marine compounds as drugs has also boosted the credibility of this research and resulted in higher fiscal support from funding bodies and industry. Given the apparently high

strike rate as drug leads from MNPs, two questions remain; Why are MNPs such a biologically active class of compounds? And why are MNPs so prevalent as antimicrobial and anticancer drugs or drug leads?

2.1.2 Ecological aspects of secondary metabolites in marine organisms

In extremely densely populated environments such as coral reefs, an organism's resources are limited and constantly threatened by competitors. An organism thriving in challenging conditions cannot afford to waste energy on producing entities, such as secondary metabolites, which do not enhance fitness. It must therefore be assumed that these compounds contribute in some way to the organism's survival. The biosynthesis of small molecules requires valuable building blocks such as isoprenes, amino acids, acetate, etc. Biosynthesis also consumes high amounts of energy (ATP) and occupies separate biochemical pathways and storage vacuoles. Thus, the current opinion on the metabolites' purpose postulates diverse functions and selection for essential roles in organismal survival. For example metabolites may represent chemically-based self protection mechanisms, such as anti-fouling agents, semiochemicals for inter- and intra-species communication (pheromones, allomones, kairomones) or UV-protection¹³¹.

Many secondary metabolites are regarded as vital frontline defence mechanisms in the given organism. In particular, sessile, soft-bodied organisms, such as many plants, fungi or various invertebrates, are assumed to rely mainly on chemical defence.

Although subject to frequent instances of pathogenic attack, these organisms often lack obvious morphological defence mechanisms or the physical ability to flee¹³²⁻¹³⁴.

In marine environments, a number of factors increase selective pressure on defence chemicals. If defence-related compounds are exuded into the surrounding water body, they are immediately diluted. Only high potency and selectivity in the mode of action (MOA) of these small molecules can compensate for loss of concentration.

Compounds may also need to pass through the cellular membranes of another (competitor or predator) organisms' cells, whether by active transport or passive diffusion, to have an effect. These are all drug-like characteristics.

Evolution of compounds in marine invertebrates has not occurred in isolation, but in close contact with sympatric organisms; predators, prey, parasites, opportunists.

Chemical experimentation arguably dates back to the first unicellular organisms. The chemical repertoire of early unicellular organisms probably produced selection for

multicellular organisms. Primitive multicellular species, such as sea sponges (Porifera), biosynthesise their secondary metabolites in relationship to, or as a response to, the metabolism of unicellular species. In an on-going response to a grand-scale natural experiment, sea sponges have built up complex chemical repertoires since their development more than 600 million years ago¹³⁵. Yet, Porifera share approximately a third of their genome and the same basic biochemical pathways with higher vertebrates and even humans^{136,137}. These genetic and biochemical similarities allow the chemical variety and flexibility from millions of years of natural combinatorial chemical selection to be directly related to our own cellular processes. This makes natural products a logical place to look for new chemical entities. Especially the marine environment as this has been the least explored.

2.1.3 Biological activity as a concept in drug discovery

Traditionally, the search for biologically active compounds resulted from the urgency to treat a particular disease. After choosing or establishing a relevant assay, an array of potential sources was screened for effectors and the resulting hits were investigated in detail. However, through this approach, the majority of screened candidates are often deemed "inactive". Potential activity, which candidates may exhibit in different assays or in the originating biological system, were dismissed¹³⁴. These candidates (anything from crude extracts to pure compounds) may not display activity in the tested environment whilst still possessing high potency under different circumstances. Thus, any comment on the compound's bioactivity must be discussed exclusively in relation to the test environment, rather than in general terms⁷.

"Negative" results or "inactivity" simply indicates that the compounds' activity has yet to be elucidated, whether or not it is exploitable for human interest.

More recently, genome-wide approaches in genetics and in proteomics (see Section 1.1.2.2), or even whole animal screens (e.g. zebrafish), provide compound centric approaches to screen for biological activity in unbiased ways.

2.1.4 Considering the phylogenetic origin of MNP drug leads

If we categorise bioactive MNPs according to their activity as well as the species from which they originate, clear phylogenetic trends can be observed. For example, Singh *et al.*¹³⁸ reviewed the literature published between 2004 and 2010, on all natural products displaying anti-HIV activity. From the 159 compounds highlighted, 11 were

of marine origin, and 6 of these were isolated from sponges. Several reviews have also analysed anti-mycobacterial drug leads from natural sources¹³⁹⁻¹⁴². Most recently, Garcia *et al.*¹⁴² reviewed a total of 259 anti-tubercular natural products of plants, bacteria, fungi, and marine organisms appearing in scientific literature from 2006 to 2011. Of the 9 mycobactericidal marine natural products, 6 were derived from sponges. A similar picture has emerged for MNPs with anti-tumour activity, where predominantly sponge-derived compounds have been identified as potential drug leads (see Table 1).

Generally, the low abundance of MNPs reflected by such reviews is likely due to a historical focus on terrestrial organisms, rather than an inferior efficacy of MNPs. This makes the high percentages that are active even more striking.

2.1.5 Sponge-derived marine natural products

Sponges (Porifera) are the oldest and lowest members of the metazoan family, which have steadily evolved over the last 635 million years¹³⁵. They flourish in habitats from tropical to polar regions, from shallow waters to the deep-sea. The over 9000 extant species described provide an incredibly rich source of chemical diversity. As sessile filter feeders sponges are particularly exposed to and dependent upon their environment. Their bodies are perforated with pores leading into a channel system. This system is responsible for oxygenation, capturing food, removing excretory matter and for reproduction¹⁴³.

Chemical defences have been identified as important in competition for space between sessile benthic species. Chemical defences may also repel and deter predators and may protect against fouling organisms or parasitic bacteria or fungi¹⁴⁴⁻¹⁴⁶. The chemical basis of defence mechanisms can be traced to both inorganic and organic matter. For instance, *Tedania charcoti*, a demosponge from deep-sea Antarctica, has been shown to contain extraordinarily high levels of cadmium and zinc, which are extremely effective antibiotics¹⁴⁷.

Organic compounds isolated from sea sponges include metabolites from all six major biochemical pathways (isoprenoid, acetogenin, amino acid, shikimate, nucleic acid and carbohydrate) and have been extensively reviewed by Faulkner (in the years 1977 and 1984 - 2002)¹⁴⁸ and Blunt (since 2003)¹⁴⁹. These compounds may be present in their biologically active form or as "protoxins". Upon mechanical tissue damage, defense protoxins are enzymatically metabolised into more potent

compounds, with a higher protective capacity. For example, *Aplysina aerophoba* has been reported to bioconvert the stored isoxazoline alkaloids, such as isofistularin-3 (**43**) or aerophobine-2 (**44**), into aeroplysinin-1 (**45**) and the dienone (**46**)¹⁵⁰. Both these compounds show highly deterrent activity towards predatory fish such as the wrasse, *Thalassoma bifasciatum* (see Figure 4). The presence of protoxins is also well documented in terrestrial natural products, such as the cyanoglycosides from milkweed and even culinary plants such as onions, garlic and wasabi.

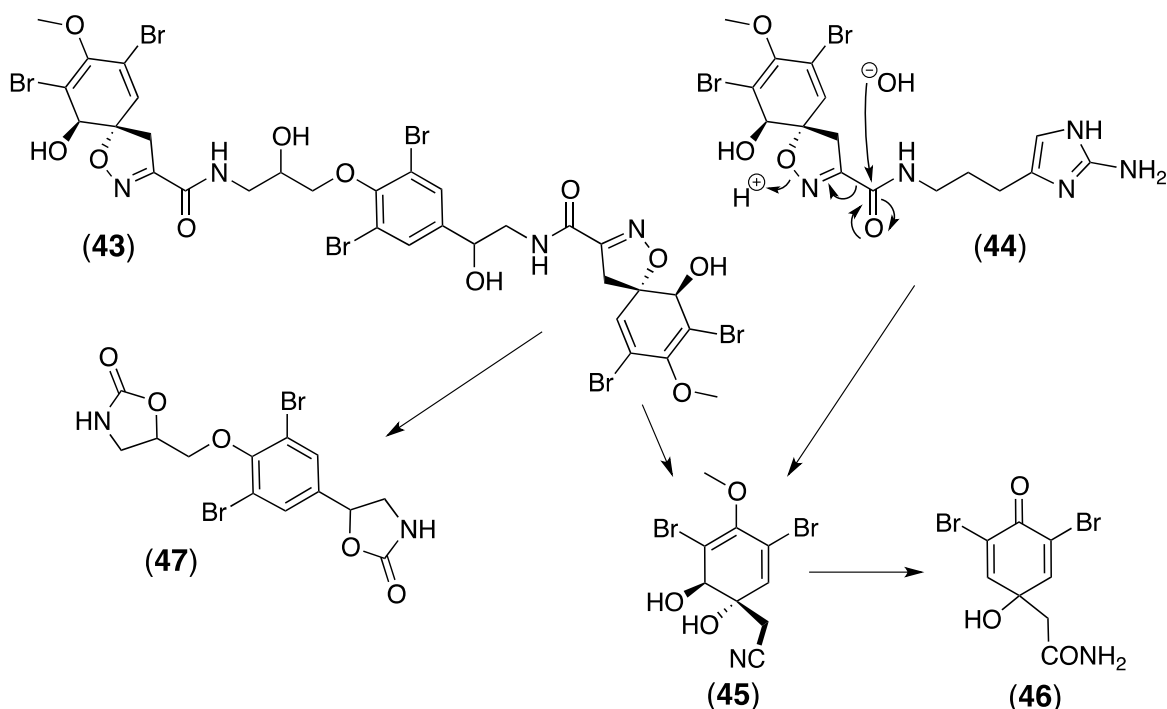


Figure 4: Protoxin isoxazoline alkaloids isofistularine-3 (**43**) and aerophobin-2 (**44**) from *Aplysina aerophoba* and bioconversion into toxins aeroplysinin-2 (**45**) and dienone (**46**). When isofistularin-3 (**43**) is used as a substrate, the bisoxazolidinone derivative (**47**) is recovered as a second product of the reaction (reproduced from Ebel *et al.*¹⁵⁰).

Marine sponges are very primitive animals. They are sessile, soft-bodied and apparently defenceless, which exposed them to a high selective pressure. Hence, Porifera had to develop survival strategies to maintain their ecological niche, which they've succeeded in over millions of years. They are a living example of the biological potential of their defense strategies, which appear to rely on their secondary metabolites. In addition, poriferan secondary metabolites have repeatedly been shown to be of high pharmaceutical relevance. Thus, sponges are an excellent source for the discovery of new drug leads.

2.1.5.1 Antimicrobial activity from sponge secondary metabolites

Porifera have been reported to be an extremely rich source for antimicrobial compounds as displayed in Table 2¹⁵¹.

Table 2: Examples of antimicrobial substances from marine sponges (from Laport *et al.*¹⁵¹)

Substance	Class	Species	Action spectrum
(S)-(+)-curcuphenol	Sesquiterpene	<i>Myrmekioderma styx</i>	<i>M. tuberculosis</i>
6-hydroxymanzamine E	Alkaloid	<i>Acanthostrongylophora</i> sp.	<i>M. tuberculosis</i>
Agelasine D	purine derivative	<i>Agelas</i> sp.	Gram-positive and Gram-negative bacteria, <i>M. tuberculosis</i>
Arenosclerins A-C	Alkaloids	<i>Arenosclera brasiliensis</i>	<i>S. aureus</i> (MRSA strain), <i>P. aeruginosa</i> (antibiotic-resistant strain)
Axinellamines B-D	Alkaloids	<i>Axinella</i> sp./ <i>Halichondrida</i>	<i>H. pylori</i>
C14 acetylenic acid	Fatty acid	<i>Oceanapia</i> sp.	<i>E. coli</i> , <i>P. aeruginosa</i> , <i>B. subtilis</i> and <i>S. aureus</i>
Caminosides A-D	Glycolipids	<i>Caminus sphaeroconia</i>	<i>E. coli</i>
Corallidictyals A-D	Hydroquinones	<i>Aka coralliphaga</i>	<i>S. aureus</i>
Cribrostatin 3	Alkaloid	<i>Cribrochalina</i> sp.	<i>N. gonorrhoeae</i> (antibiotic-resistant strain)
Cribrostatin 6	Alkaloid	<i>Cribrochalina</i> sp.	<i>S. pneumoniae</i> (antibiotic-resistant strain)
CvL	Lectine	<i>Cliona varians</i>	<i>B. subtilis</i> , <i>S. aureus</i>
Cyclostelletamines A-I, K-L	Nitrogenous	<i>Pachychalina</i> sp.	<i>S. aureus</i> (MRSA strain), <i>P. aeruginosa</i> (antibiotic-resistant strain), <i>M. tuberculosis</i>
Deoxytopsentin	Alkaloids	<i>Spongosorites</i> sp.	<i>S. aureus</i> (MRSA strain), <i>S. aureus</i> (MRSA strain), <i>P. aeruginosa</i> (antibiotic-resistant strain)
Haliclonaclamamine E	Alkylpiperidine alkaloids	<i>Arenosclera brasiliensis</i>	<i>S. aureus</i> (MRSA strain)
Hamacanthin A	Alkaloids	<i>Spongosorites</i> sp.	<i>S. aureus</i> (MRSA strain)
Ingenamine G	Nitrogenous	<i>Pachychalina</i> sp.	<i>S. aureus</i> (MRSA strain), <i>E. coli</i> , <i>M. tuberculosis</i>
Isojaspic acid, cacospongin D and jaspinquinol	Meroditerpenes	<i>Cacospongia</i> sp.	<i>S. epidermidis</i>
Latrunculins	Macrolides	<i>Negombata magnifica</i>	<i>S. aureus</i> and <i>B. cereus</i> .
Melophlin C	Nitrogen heterocycles	<i>Meloplus sarassinorum</i>	<i>B. subtilis</i> and <i>S. aureus</i>
Petrosamine B	Alkaloids	<i>Oceanapia</i> sp.	<i>H. pylori</i>
Polydiscamide A	Peptide	<i>Discodermia</i> sp.	<i>B. subtilis</i>
Psammaphin A	Bromotyrosine-derived	<i>Psammaphysilla</i>	<i>S. aureus</i> (MRSA strain)

There is a diverse antibacterial spectrum of sponge-derived MNPs, including important human pathogens such as mycobacteria, Gram-positive, Gram-negative, or

antibiotic-resistant strains of bacteria (including both Gram-positive and -negative strains). The latter are particularly note-worthy, as resistance has developed against the majority of available antibiotics (including last resort antibiotics such as vancomycin), and there is an urgent need to discover new antibiotic substances¹⁵².

2.1.5.2 Antifouling

Sponges are very important in the ecology of rocky and coral reefs, where they usually flourish, exhibiting high diversity and abundance. They are exposed to constant biotic competition for limited space in littoral zones, especially against autotrophs such as phytoplankton, algae, and other fouling organisms. Despite harbouring symbiotic organisms on their porous surface as well as within sponges are surprisingly rarely fouled. Sponges appear to contain chemicals that are repellent to plants that could potentially shade them, overgrow them or clog their pores. Repellents that work against marine plants are also likely to have an effect on terrestrial plants. Sponges are therefore an excellent potential source of new chemical entities displaying antifouling or herbicidal properties.

The literature on sponge metabolites which act as herbicides is limited. One highly interesting cellular target for herbicides has been described by Llewellyn *et al.*¹⁵³. They conducted by far the largest screening for herbicidal compounds from marine organisms, employing a collection from the Australian Institute of Marine Science of 5023 extracts from diverse taxonomic groups. Extracts were screened using a novel plant biomolecular assay, enabling identification of inhibition of individual enzymes (pyruvate orthophosphate dikinase, PPDK) involved in the C4 acid cycle. The assay has advantages over other herbicidal assays in that it generates comparative data between C3 plants and C4 plants. The C4 plant group contains the world's worst weeds. Of the 1472 screened sponge extracts, 14% showed activity, with 9% showing selectivity for the C4 specific enzymes. The disadvantage of this assay is its limitation to a single enzyme system - it does not provide any additional information, such as route of uptake or application methods. More recent studies from the same research group reveal seven poriferan sesquiterpene quinones with promising activities against the PPDK¹⁵⁴. The PPDK enzyme is an interesting target for herbicides. Both invertebrates and vertebrates lack PPDK, thereby reducing the likelihood that these herbicides will have unwanted toxicological or environmental impacts.

Besides environmental concerns and toxicology of herbicide use, the rise of genetically modified crops and the limited number of target biochemical pathways in plants reduces the number of herbicides possible¹⁵³. Additionally, after intensive use of a limited number of market-dominating agrochemicals over the last few decades, herbicidal resistance has emerged and currently occurs in response to all known modes of action¹⁵⁵.

The search for alternatives to the existing herbicides is therefore of utmost importance. Given that most compounds currently used derive from natural products¹⁵⁶, MNPs seem to be a most promising and extensive source for new chemical entities and important lead compounds.

2.1.6 Transfer of chemical defence compounds between species

Chemicals involved in organismal defence may also be used by other species for the same or similar purposes¹⁵⁷⁻¹⁵⁹. Other animals that feed on sponges can sequester, concentrate and modify sponge metabolites for their own purposes, either making the sponge molecule less or more toxic¹⁴³. For instance, the nudibranch gastropod, *Hypselodoris orsini* feeds on sponges of the family Thorectidae. Initial metabolic conversion of acquired toxic sesterterpenes, such as scalaradial, yields the selectively reduced and less toxic deoxoscalarin. During a second step, the selective oxidation to 6-ketodeoxoscalarin, the metabolites are deposited in the nudibranch's mantle dermal formations. There they are re-purposed as a predator deterrent¹⁶⁰.

There are several examples of prominent MNPs being carried from their source throughout the food chain and accumulating in the predator, as in the case of the lipodepsipeptide kahalalide F. Its analogue, Elisidepsin, is currently undergoing phase II clinical cancer trials. Kahalalide F was first isolated from the sacoglossan *Elysia rufescens* with its food source, the green algae *Bryopsis* spp.¹¹⁰. This lipopeptide was later shown to provide a defence against predation from both the algae and the sea slug¹⁶¹, which accumulated kahalalide F at up to 5000 fold concentrations, as compared to the algae. In 2007, Hill *et al.*¹⁶² established that the kahalalides were derived from a bacterium (*Vibrio* sp.), living in symbiosis with *Bryopsis* sp.

2.1.7 Symbionts as potential source for bioactive small molecules

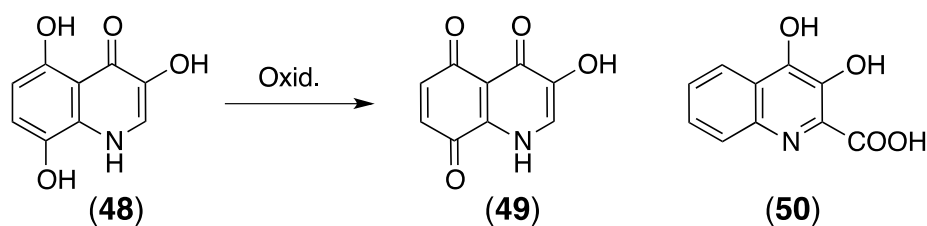
Many sponges selectively harbour symbionts, such as macroalgae or cyanobacteria, on their porous surfaces as well as within their channel system¹⁶³. In such cases, the Porifera provide morphological protection to the microorganism in exchange for

nutrients from the symbiont and their chemical repertoires are therefore highly intertwined¹⁶⁴. This has to be taken into account when isolating secondary metabolites from sea sponges, as potential candidates for any screened bioactivity may originate from the sponge itself, or may instead originate from associated microorganisms. The true origin of most sponge compounds have not yet been proved unambiguously¹⁶⁵.

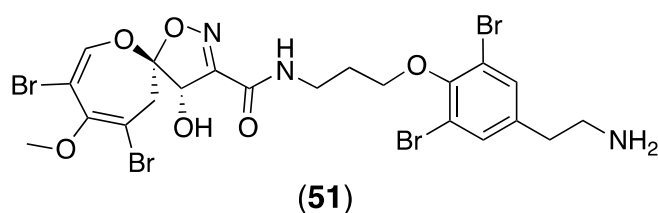
The origin of sponge compounds is important in the general drug discovery process. The limited quantity of rare compounds from marine sources normally constitutes a major bottleneck. For the procurement and manufacture of potential new drug candidates, identifying the original source of the compound is of utmost importance. The prospect that sponge-associated microorganisms may be the original source of at least some of the isolated bioactive MNPs, has raised hopes of obtaining sustainable and limitless supplies of these metabolites. The question of the origin of many MNPs remains a key issue among MNP chemists and will take many more decades to resolve¹⁶⁵.

2.1.8 Order Verongida

The order of Verongida sponges (Berquist, P. R. 1978) belongs to the class Demospongiae and harbours four families: Aplysinellidae, Aplysinidae, Ianthellidae, and Pseudoceratinidae¹⁶⁶. The order members lack spicules and are typically fleshy and soft¹⁴³. Their skeletons are incorporate large, widely spaced spongin fibres, which form dendritic or reticulate structures. The fibres may be aggregated into bundles. Verongida sponges don't exhibit differentiation between primary and secondary elements. Also, they do usually not incorporate detritus into their fibres. The fibres display a laminated cortical region and a distinct central pith of fine spongin fibrils. In some species, the cortex may be reduced or disappear entirely. The gelatinous matrix within the sponge (mesohyl) contains abundant collagenous fibrils. They typically possess a pigment, which oxidises to a deep purple colour when exposed to air. For example, in 1984 Cimino *et al.* reported "uranidine(s)" (**48**) as the predominant zoochrome in *Verongida aerophoba*¹⁶⁷. The authors showed the uranidine undergoes air oxidation to yield the corresponding quinone (**49**). They suggested a 3,4-dihydroxyquinoline-2-carboxylic acid (**50**) as the biosynthetic precursor. This compound was first isolated in 1971 by Fatturoso *et al.*¹⁶⁸ and reported to act as a strong feeding deterrent¹⁶⁹.



Verongida sponges are typically unfouled and they possess distinct morphological and biochemical defense mechanisms¹⁷⁰. The most notable biochemical characteristic is the consistent biosynthesis of dibromotyrosine derivatives¹⁷¹. This suggests that these compounds are of not of microbial origin as the same types of compounds are isolated from verongid sponges from around the world. Many of these bromotyrosines exhibit biological activities. For example, molokai'iamine (**51**) exhibits selective antitumour activity against various cell lines in the micromolar range^{172,173}. Hamann and coworkers recently reviewed the isolation, occurrence, biosynthesis and structural characteristics of all brominated tyrosines reported until 2005 (i.e. 284)¹⁷⁴.



Verongida sponges have continuously proven to be a rich source of biologically active secondary metabolites^{159,175}, and are ideal candidates for bioassay-guided isolation of marine natural products.

2.1.9 Bioassay guided fractionation

The search for new compounds, new compound classes and ultimately new drug leads is not reliant on the initial identification of the producing organism (e.g. sponge or symbiont bacterium), but can be elucidated by focussing on their bioactivity profile.

Rather than simply describing secondary metabolites isolated from organisms (as part of a *basic science* approach), *applied natural product research* aims to drive isolation of NP through applicability, i.e. their potential for commercial use¹²². Firstly, this requires screening crude extracts against a chosen biological system and assessing the general activity profile of the extracts within the test environment. Any extract exhibiting biological activity in these assays is then fractionated via liquid-liquid partitioning according to the polarity of the constituents. Thorough bioassay-

guided monitoring of the fractions yielded from the partitioning process aims to narrow down the bioactive principle. Ultimately this provides a single chemical structure as the compound responsible for the observed bioactivity¹⁷⁶.

Bioassay-guided isolation of natural products faces major limitations in that choices often need to be made as to which activity to test for. Unless a laboratory is extremely well equipped and has facilities and funding to screen each extract or compound against tens or hundreds of targets, it is necessary to focus the activity profiling to circumstances which are expected to result in a high hit rate. The ecological and phylogenetic circumstances of the chosen organisms producing the NP are therefore extremely informative (although they can also be potentially misleading, see sections 2.1.6 and 2.1.7). The ethical restrictions and biosafety approvals required for many bioactivity tests also limit the choice of assays for many laboratories.

The aim of this chapter therefore is to identify a selection of preliminary biological assays to guide the isolation of sponge metabolites as the first step in reverse chemical proteomics.

2.1.9.1 Antimicrobial activity

Even within a traditional organic chemistry laboratory, simple antibacterial bioassays can be easily performed to guide the isolation of MNPs according to their bioactivity profile against common human pathogens.

2.1.9.1.1 Disc diffusion

The disc diffusion method is based on the Kirby-Bauer assay¹⁷⁷, which is commonly used for the screening of antibacterial agents¹⁷⁸. This assay is useful as it is a cheap, easy and quick method allowing simultaneous screening of several test samples (e.g. extracts or pure compounds) on a single agar plate against any one organism (bacteria or fungi). A bacterial lawn is cultivated on an agar plate and a paper disc containing a known concentration of the test sample is placed on the surface. The assay relies on the ability of the test material to diffuse into the agar, and to prevent the growth of the organism.

The zone of inhibition (ZOI) of growth of the organism is measured by the diameter of no growth (including the diameter of the disc, \varnothing 6 mm), indicating the efficacy of the test substance. Generally, the test sample can be regarded as having no (< 7 mm), low (< 10 mm), moderate (10 - 15 mm) or high (> 15 mm) levels of activity depending on the diameter of the inhibition zone.

This assay is limited by the requirement that the active components diffuse fully into the agar¹⁷⁹. It is therefore an inappropriate means for testing non-polar compounds that diffuse poorly into agar¹⁸⁰. Additionally, this assay is not commonly employed to determine the minimum inhibitory concentration (MIC) of a substance, as a range of concentrations have to be tested separately, requiring the use of a great deal of sample.

2.1.9.1.2 MTT microdilution assay

Since its development¹⁸¹, the MTT cell proliferation assay (using the tetrazolium salt 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (**52**; MTT) as an indicator of cell viability), has been widely used for assessing antiproliferative activity of test samples such as crude extracts or pure compounds. Originally developed as an assay for the assessment of anticancer agents, it has also been adapted for use as an antibacterial screening assay¹⁸².

This colourimetric assay is based on the reductive cleavage of MTT by live cells to yield blue formazan (**53**; see Figure 5)¹⁸³. A colour change is an indicator of the presence of viable cells and lack of antiproliferative activity of the test sample.

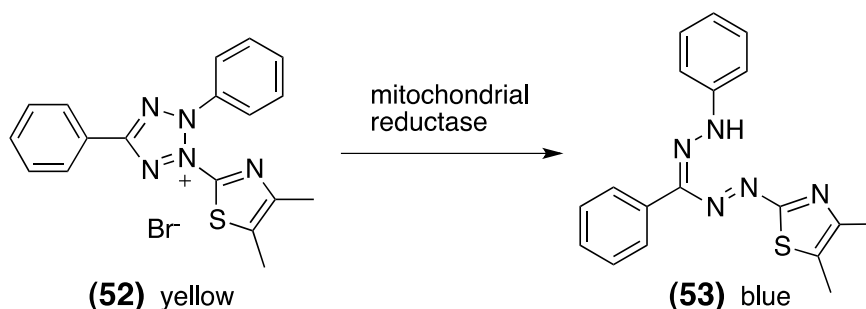


Figure 5: MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide) (yellow) is converted to formazan (dark blue) through mitochondrial reductase in live cells.

Although the reductive mechanism of MTT in bacteria is not well understood, this assay is generally accepted as one of the standard antimicrobial tests and is widely used^{184,185}.

The MTT assay is applicable for high throughput screening, as it uses significantly less material than the disc diffusion test for determining the MIC of the test sample. Serial dilutions are prepared directly in a 96-well microtitre plate format and easily produce a dose-response curve. As this assay relies on a colourimetric indicator and chemical reduction, the initial colour of the test sample and its redox activity will impact on the interpretation of the results. Also, solubility issues with non-polar samples may occur similar to the disc diffusion test.

2.1.9.1.3 Turbidity assay

Prepared in a similar manner to the MTT assay, the turbidity microdilution method can also be used in a 96-well plate format to allow the assay of multiple concentrations of several samples at once. This assay relies on a change in turbidity of the sample due to growth of the organism. This change is measured by absorption (at a wavelength of 600 nm) as measurement of the cell density and does not require the addition of an indicator after the incubation step, as compared to the MTT assay. This method provides an objective measure of MIC. With appropriate equipment (96-well plate reader), this assay allows for the rapid screening of many concentrations of several extracts and is suitable for automation^{186,187}. It can furthermore be used to assess the growth kinetics of the organism, both for determining a dose response curve and for use in time lapse. It provides quantitative data in form of the proportional inhibition determined via comparison to the control. Hence, the data on sample concentration required to inhibit 50% growth allow for calculating IC₅₀ values¹⁸⁸.

However, it has to be taken into account that this assay only provides information on the turbidity of the cultures. The viability of the organisms or abnormal morphological characteristics of the cells cannot be determined. The presence of insoluble material (e.g. from an extract) will also interfere with the measurement. However, this assay can be applied in combination with a colourimetric assay, such as the MTT assay, by simply adding the indicator after incubation and turbidity measurements. A change in colour will thus provide evidence of the viability of the cells.

2.1.9.2 Herbicidal activity

A bioassay based on the duckweed *Wolffia arrhiza* has been used for testing herbicidal activity of sponge-derived metabolites. This assay is especially suitable for testing lake and river pollution, sediment toxicity, and herbicidal pollution. The term “duckweed” describes species belonging to the aquatic, vascular angiosperm plant family Lemnaceae, which includes four different genera, *Spirodela*, *Lemna*, *Wolffiella*, and *Wolffia*. Lemnaceae are free floating monocotyledonous plants that display exponential growth by clonal propagation. The use of Lemnaceae in ecotoxicological research is widespread and standardised by various agencies to assess phytotoxicity¹⁸⁹. *Wolffia arrhiza* is the smallest higher order plant in the world with

an average size of approximately 2 mm³. Leaf, stem and root system are unified, making it an ideal organism for screening large libraries of phytotoxic compounds in a 96-well plate format.

2.2 Results and discussion

2.2.1 General extraction

There are several methods that can be used to prepare sponge extracts¹⁷⁶. In this study, frozen sponge samples were extracted with 70% (v/v) aqueous ethanol and the solvents were evaporated under reduced pressure and low temperature (40 °C) to afford the crude extracts. The crudes were then partitioned against solvents with increasing polarity of the solvent to afford the petroleum ether (PE), dichloromethane (DCM), ethyl acetate (EtOAc) and *n*-butanol (BuOH) and water (Water) fractions (see Table 3). JB07-S1 and the molluscs (see Section 2.2.4) were partitioned against PE, EtOAc and BuOH.

Table 3: List of investigated marine sponge samples including original weight and weights of different extracts.

Sponge Sample	Initial Weight (g)	Partition				
		PE	DCM	EtOAc	BuOH	Water
JB07-S1	654	0.9076	n.a.	6.5420	0.2241	5.7681
Mollusc	13	0.0401	n.a.	0.0120	0.0956	0.1375
JB08-S6	505	1.1978	1.2471	0.2497	1.3086	3.4372
JB08-S7	220	1.1576	0.1373	0.0067	0.0823	1.6463
JB08-S8	395	2.2706	2.0970	0.1566	1.0535	1.4294
JB08-S9	320	0.1147	0.0777	0.0436	0.0475	0.7075
JB08-S10	185	0.3444	0.7788	0.0311	0.1627	1.3847
JB08-S14	125	0.7722	0.1555	0.0036	0.0901	0.9837
JB08-S15	183	0.6416	0.2114	0.0423	0.3717	0.6353
JB09-S1	45	0.3147	0.0798	0.0057	0.1283	1.0852
JB09-S2	56	0.2942	0.2051	0.0395	0.3668	0.6441
JB09-S3	18	0.1209	0.1173	0.0779	0.2233	0.5404
JB09-S4	225	1.7568	0.0214	0.0151	0.2933	0.2534

Abbreviations: PE = petroleum ether, DCM = dichloromethane, EtOAc = ethyl acetate, BuOH = *n*-butanol, n.a. not available.

2.2.2 Screening of crude sponge partitions

2.2.2.1 Antimicrobial assay

All preliminary antibacterial screening of marine sponge extracts was performed against three human pathogenic microorganisms, *Staphylococcus aureus*, *Escherichia coli* and *Pseudomonas aeruginosa*. These are some of the most frequently isolated pathogens from a variety of clinical conditions and are therefore ideal candidates to be used for screening antimicrobial agents¹⁹⁰.

2.2.2.1.1 Disc diffusion

All 59 sponge partitions were screened in a disc diffusion assay against the three pathogens and the results are summarised in Table 4. Generally, all samples were dissolved in acetone/water (1:1 v/v) mixtures to minimise solubility issues for non-polar compounds. Thirteen partitions (22%) yielded antibacterial activity (low - high) against at least one bacterial strain. For all remaining partitions, there was either no antibiotic activity detected or the zone of inhibition (ZOI) was only minimal ($\varnothing < 7$ mm), also being characterised as "inactive".

Table 4: Results from disc diffusion assay of sponge fractions are displayed in diameter (in mm) of zone of inhibition.

Sponge Sample or Antibiotic	Bacterial strains and diameter of inhibition zone (in mm)		
	<i>E. coli</i>	<i>S. aureus</i>	<i>P. aeruginosa</i>
JB07-S1	11 (EtOAc)	35 (EtOAc)	< 7
JB08-S6	< 7	< 7	< 7
JB08-S7	< 7	7 (DCM)	< 7
JB08-S8	< 7	8 (PE, DCM)	< 7
JB08-S9	9 (PE, DCM)	8 (DCM)	< 7
JB08-S10	< 7	< 7	< 7
JB08-S14	< 7	< 7	< 7
JB08-S15	< 7	< 7	< 7
JB09-S1	< 7	< 7	< 7
JB09-S2	< 7	< 7	< 7
JB09-S3	13 (DCM, EtOAc)	15 (DCM, EtOAc), 12 (PE)	< 7
JB09-S4	< 7	< 7	< 7
Ampicillin	25	17	-
Chloramphenicol	-	-	20

Each sponge sample represents the PE, DCM, EtOAc, But and water partition of an ethanolic extract of the sponge. The concentration of all samples was 30 µg/disc, except for ampicillin (10 µg/disc). The diameter of a disc was 6 mm.

The ethyl acetate fraction of sponge JB07-S1 clearly demonstrated the strongest level of antibacterial activity, being characterised as "high" against *S. aureus* (\varnothing 35 mm) and "moderate" against *E. coli* (\varnothing 11 mm). JB09-S3 also demonstrated antibacterial activity from a range of partitions, with both DCM and EtOAc partitions showing "moderate" activity against *S. aureus* (\varnothing 15 mm) and *E. coli* (\varnothing 13 m), as well as the PE fraction displaying "moderate" activity against *S. aureus* (\varnothing 12 mm).

Unsurprisingly, not all partitions contained compounds showing bactericidal activity. However, that none of the samples subjected to *P. aeruginosa* showed any impact is somewhat surprising. There are two feasible explanations for this. Firstly, perhaps none of the extracts tested possessed antibacterial activity against this antibiotic-sensitive strain of *P. aeruginosa*. Secondly, the concentrations of those compounds within each partition carrying the biological potential may have been too low so that the bacterial growth outperformed any inhibitory effects. The inherent risk in testing any unknown mixture of compounds in a biological assay as a whole is to miss substances that are present in very small quantities, but would display activity in higher concentrations. For the disc diffusion assay, recommendations for the extract concentration vary between 0.5-5 mg per disc^{191,192}. It is likely that a higher concentration would demonstrate a higher level of antibacterial activity. A concentration of 0.5 mg of extract per disc was used for these assays. Our aim was simply to narrow down the search for bioactive compounds and to guide the isolation and structure elucidation towards compounds present in quantities sufficient for derivatisation (milligrams) in reverse chemical proteomics. We were not aiming to isolate every single minor constituent of each extract. Other research laboratories, with a special focus on NP chemistry, have access to specialised instrumentation to perform the analysis of compounds present in microgram amounts much more efficiently.

2.2.2.2 Herbicidal

55 sponge partitions were also tested for their herbicidal potential. Unfortunately, the data on the herbicidal activity for all partitions from JB07-S1 is missing.

The herbicide *N*-(phosphonomethyl)glycine (glyphosate) was used as positive control. Glyphosate is known to show only weak activity if introduced to media and to have a much higher herbicidal efficacy if for instance sprayed onto the leaves of plants^{193,194}. However, taking this into account, we applied glyphosate in the identical way as the sponge extracts and determined an MIC of 0.45 mg/mL.

To standardise the conditions throughout the assay, the starting concentration of extract solutions applied in the assay was limited to 1 mg/mL, as higher concentrations occurred in solubility issues of the apolar fractions. The results from all partitions tested in this assay are summarised in .

20 of the 55 partitions showed complete growth inhibition of the duckweed at this level. Another 16 samples inhibited at least 50 % growth of *W. arrhiza* compared to growth control (inhibitory concentration (IC_{50}) ≤ 1 mg/mL), and for 19 partitions the starting concentration did not result in significant inhibition (i.e. > 10 % growth inhibition compared to control). None of the samples provided a lower MIC as the positive control glyphosate (MIC = 0.45 mg/mL).

Table 5: Herbicidal activity of sponge partitions tested against *W. arrhiza*.

Sponge Sample	Inhibition of <i>Wolffia arrhiza</i>				
	PE	DCM	EtOAc	But	Water
JB07-S1	n.a.	n.a.	n.a.	n.a.	n.a.
JB08-S6	++	++	+	++	-
JB08-S7	+	++	+	+	-
JB08-S8	++	++	++	+	-
JB08-S9	++	+	-	-	-
JB08-S10	++	+	+	-	-
JB08-S14	++	+	++	++	+
JB08-S15	++	++	-	-	-
JB09-S1	++	+	-	+	-
JB09-S2	++	++	+	-	-
JB09-S3	++	++	++	-	+
JB09-S4	+	++	+	-	-
Glyphosate	MIC = 0.45 mg/mL				

Activity is represented as high (++, MIC ≤ 1 mg/mL), moderate (+; $IC_{50} \leq 1$ mg/mL) or missing (-).

Interestingly, all five partitions of sample JB08-S14 displayed herbicidal activity, which has not been observed with any other sample. The polarity of the extracts and thus compounds reflect an overall trend in activity, whereby the less polar partitions (PE, DCM, EtOAc) generally showed a higher degree of herbicidal activity compared to polar one (water and butanol). However, both butanol and water partitions of JB08-S14 (at [1 mg/mL]) inhibited plant growth. The butanol one, similar to the PE and EtOAc partitions of the same sponge, resulted in the complete inhibition of *W. arrhiza* at 1 mg/mL. The DCM and water partition only limited the plant growth to less than half maximum growth.

2.2.2.3 Summary

From all investigated fractions showing bioactivity (antimicrobial or herbicidal), we decided to analyse the ethyl acetate partition of JB07-S1 in detail. It clearly displayed the highest antibacterial effect observed from within our collection. Additionally, the co-collection of its immediate predator (mollusc) instigated our interest as to find both biologically active small molecules and learn more about the transport of biologically active compounds throughout the food chain.

2.2.3 *Pseudoceratina purpurea*

Sample JB07-S1 was identified as *Pseudoceratina purpurea* (Carter 1880; family Pseudoceratinidae Carter; order Verongida; class Demospongiae) according to Bergquist *et al.*¹⁶⁶ and Hooper¹⁴³. There are several synonymised taxa (*Aplysina purpurea*, *Dendrilla verongiformis*, *Druinella ramosa*, *Druinella tyroeis*, *Hexadella pleochromatum*, *Korotnewia desiderata*, *Psammaplysilla purpurea*, *Thorectopsamma xana*)¹⁹⁵.

The poriferan genus *Pseudoceratina* is characterised by a sparse fibrous skeleton organised on the dendritic plan, in which bark elements are absent. Its texture is firm, often hard and incompressible. Widely spaced tubercles or low raised ridges interrupt the otherwise smooth sponge surface. Its matrix is extremely dense and contains heavy collagen. As typical for all species of the Verongida order, *Pseudoceratina* are oviparous rather than viviparous.

In general, Verongida differ in their chemistry from many other sponge orders. The majority of all sponge secondary metabolites described previously, derive from the isoprenoid pathway. However, compounds isolated from this order are predominantly amino acid derivatives (87%). Moreover, one of the first chemotaxonomic observations for all Porifera denoted the high abundance of brominated tyrosine compounds in all Verongida genera^{131,170}.

2.2.3.1 Bioassay guided fractionation

The ethyl acetate partition was subjected to gel permeation filtration (in Sephadex, see Section 8.1.1.1). Correlating subfractions were combined and tested using the disc diffusion assay against three bacterial strains (see Section 2.2.2.1). Sephadex fractions S-9-12, S-13-14 and S-15-22 all displayed bactericidal activity against both *S. aureus* and *E. coli* (see Table 6).

Table 6: The Sephadex fractions of JB07-S1 (*P. purpurea*) were tested for antibacterial activity in the disc diffusion assay and fractions showing activity are summarised below.

<i>P. purpurea</i> EtOAc partition Sephadex Fractions	Antibacterial activity in disc diffusion test		
	<i>E. coli</i>	<i>S. aureus</i>	<i>P. aeruginosa</i>
JB07-S1-EA-S-9-12	7	14	< 7
JB07-S1-EA-S-13-14	8	16	< 7
JB07-S1-EA-S-15-22	8	10	< 7
Ampicillin	25	17	-
Chloramphenicol	-	-	20

Results are displayed in diameter (in mm) of zone of inhibition. The concentration of all samples was 30 µg/disc, except for ampicillin (10 µg/disc). The diameter of a disc is 6 mm.

From these bioactive fractions, S-13-14 produced the largest zone of inhibition against *S. aureus* (Ø 16 mm) and can hereby be ranked as highly active. In comparison to the antibacterial results from the initial EtOAc partition the activity was significantly reduced. Fractions S-9-12 and S-15-22 both ranked as "moderately" active (Ø 14 mm and Ø 10 mm, respectively). This suggested that the compounds primarily responsible for antibiotic effect were enriched in fractions S-14-15, and partially present in the previous and following Sephadex fraction. The activity against *E. coli* was very low with a maximum of 8 mm ZOI for any of the three bioactive Sephadex fractions.

While stability issues may have contributed partially to this reduced activity compared to the parent fraction¹⁹⁶, it is well known that synergism effects of complex natural product mixtures are often responsible for the overall biological activity, hence isolated pure constituents will show a reduced effect¹⁹⁷⁻¹⁹⁹.

It is likely that during the fractionation process one or more synergistically acting compounds were separated from the major set of bioactive components (during size exclusion chromatography). This is a common phenomenon seen, for example, in a review on the antiplasmodial efficacy of plant extracts compared to the effect of their fractionated active compounds¹⁹⁸. In this study, the authors describe the activity of the extract from the plant *Garcinia cowa* in comparison to the isolated constituents. When tested individually, the isolated compounds exhibited only up to 3% of the overall activity and thus a synergistic mechanism of the mixture is postulated as a feasible explanation.

2.2.3.2 HPLC purification of brominated tyrosines

Subsequent LC-MS analysis proved the three subfractions to be very similar in their chemical composition, although with varying concentrations of each compound. The fractions were therefore combined and collectively purified by preparative HPLC (see Figure 6).

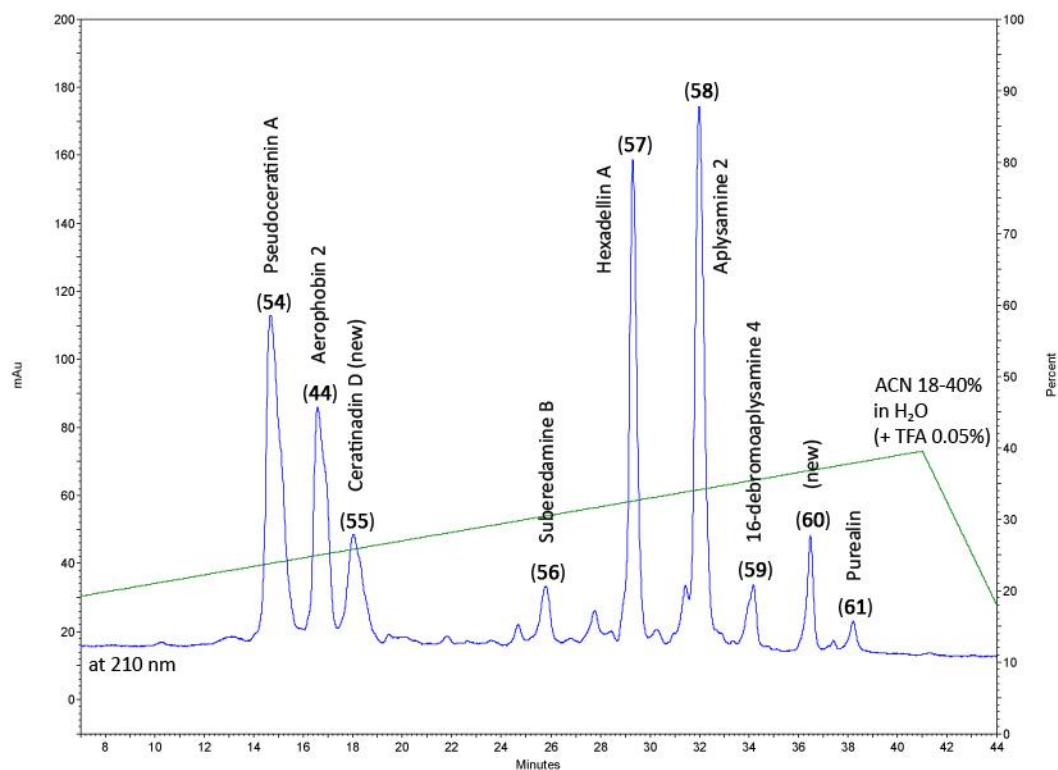


Figure 6: HPLC trace of bioactive Sephadex fractions 9-22 isolated from *Pseudoceratina purpurea*. Numbered peaks indicate isolated and structurally elucidated compounds. A gradient of 18-40 % acetonitrile in aqueous TFA (0.05 %) was run over 41 min on a preparative Gemini C18 column and peaks were detected at 210 nm. The structures of compounds 44 and 54-61 were elucidated by NMR and MS.

The HPLC separation yielded 9 bromotyrosine compounds (**44**, **54** - **61**; see Figure 7). Inspection of the ^1H and ^{13}C NMR and MS indicated that seven were known (pseudoceratinine A (**54**)²⁰⁰; aerophobin-2 (**44**)²⁰¹; suberedamine B (**55**)²⁰²; hexadellin A (**57**)²⁰³; aplysamine-2 (**58**)²⁰⁴; 16-debromoaplysamine-4 (**59**)²⁰⁵; purealin (**61**)²⁰⁶. There is no evidence in the available literature that compounds **55** and **61** have been previously described.

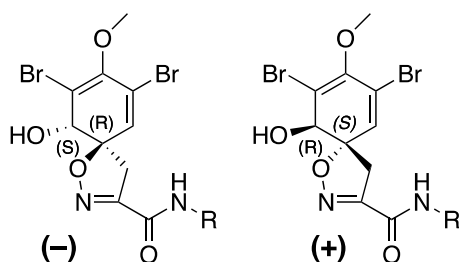
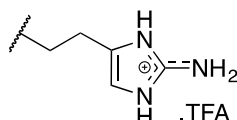
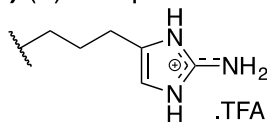
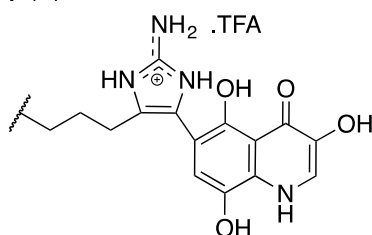
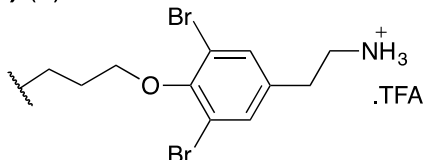
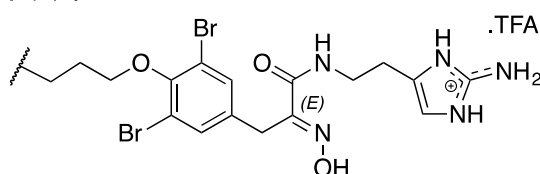
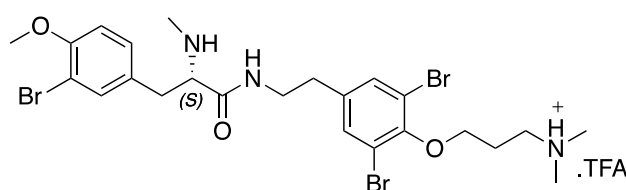
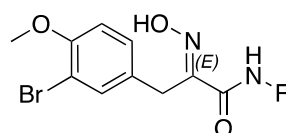
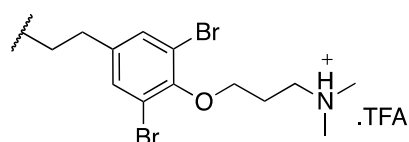
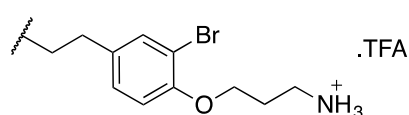
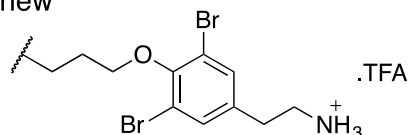
spiro-cyclohexadienylisoxazolines:**(54) (-)-pseudoceratinine A****(44) (+)-aerophobin-2****(55) (+)-ceratinadin-D****(57) (-)-hexadellin A****(61) (-)-purealin****(56) (+)-suberedamine B****bromotyrosine oximes****(58) aplysamine-2****(59) 16-debromoaplysamine-4****(60) new**

Figure 7: Structures of nine brominated tyrosine derivatives isolated from the ethyl acetate partition of the marine sponge *Pseudoceratina purpurea*. Compounds 55 and 60 have not previously been described in literature.

Brominated tyrosine compounds are structurally divided into six categories¹⁷⁴: simple bromotyrosine derivatives, spirocyclohexadienylisoxazolines, spirooxepinisoxazolines, oximes, bastadins, and other structural classes. In a detailed review on bromotyrosines, Peng *et al.*¹⁷⁴ discussed the structural features, biogenetic origin and spectral data on all 284 different bromotyrosines that had been reported up to the year 2005. Since then several papers were published on new brominated

tyrosines from various verongid genera, with 20 new compounds exclusively derived from *Pseudoceratina* sp.²⁰⁷⁻²¹⁹.

Bromotyrosines display a wide range of biological activities, ranging from antimicrobial, anticancer, antifouling, antiviral, ATPase regulator, calcium channel modulator, and others.

Five of the bromotyrosines isolated in this study (**44**, **54**, **55**, **57**, **61**) share the spiro-cyclohexadienyl-isoxazoline moiety and compounds **44**, **54**, **55** and **61** additionally incorporate histamine, a common feature for many verongid derived bromotyrosines. Compounds **58** – **60** can be classified as oxime-bromotyramines, which are widespread within the *Pseudoceratina* genus as well. Suberedamine B, originally described from the genus *Suberea* sp. (family Aplysinellidae, order Verongida)²⁰², is a representative of bromotyrosine alkaloids having a primary amino or an *N*-methyl amino group at the α -carbon.

2.2.3.2.1 (+)-Ceratinadin D (**55**)

(+)-Ceratinadin D was obtained as an optically active solid ($[\alpha]_D^{20} +52$) with UV maxima at 234, 276, and 337 nm, suggesting the presence of a spiro-cyclohexadienyl-isoxazoline and a uranidine moiety. The IR absorptions indicated the existence of OH and/or NH (3420, 3380, and 3290 cm^{-1}) and an amide (1677 cm^{-1}). All values matched reasonably well with those reported for (+)-ceratinadins A and B²¹⁷. The low resolution mass spectrum displayed an isotopic cluster (694.95, 696.95, 699.00 amu in ratio 1:2:1) suggesting two bromines were present. High resolution mass spectrometry suggested a molecular formula $\text{C}_{25}\text{H}_{24}^{79}\text{Br}_2\text{N}_6\text{O}_8$ (3.6 ppm error). The ^{13}C NMR spectrum indicated the presence of 25 carbons (see Table 7), one O-methyl (δ 59.7), four methylenes (δ 39.0, 38.2, 28.1, 21.6), three sp^2 methines (δ 131.2, 124.2, 113.7), one sp^3 methine (δ 73.5), 13 sp^2 quaternaries (δ 154.5, 149.6, 147.6, 146.6, 139.9, 137.4, 128.9, 122.4, 120.9, 117.8, 113.1, 112.6, 103.2), one sp^3 quaternary (δ 90.1) and two carbonyls (δ 173.2, 158.9). All protonated carbons were correlated to their protons by HSQC experiments²²⁰.

Use of one and two-dimensional NMR data (see Table 7) enabled the construction of four substructures (see Figure 8). Inspection of ^1H , ^{13}C , and ^1H - ^1H COSY NMR spectra suggested that the following proton signals belonged to the same spin system: δ_{H} 6.54 (sp^2 methine), δ_{H} 3.90 (sp^3 methine), δ_{H} 3.63 (OMe), and an AB system (δ_{H} 3.59 and δ_{H} 3.15 each 1H) characteristic of a spirocycloisoxazoline ring previously published from

other verongid compounds¹⁷⁴. HMBC correlations from H-5 to C-1, C-2, C-3, H-8 to C-3, and H-10a/b to C-1, C-6, C-11 confirmed the 1-hydroxy-2,4-dibromo-3-methoxy-11-carbonyl spirocyclohexadienyl isoxazole (substructure **A**) and the connection between substructures **A** and **B** was obtained by HMBC correlations from H-10a/b to C-14.

The signal at δ_{H} 8.51 (NH) showed a coupling to δ_{H} 3.13 (H-17) and a HMBC correlation to the amide carbonyl (C-14; δ 158.9). Further COSY correlations from H-17 to H-18 (δ_{H} 1.73), and in turn H-19 (δ_{H} 2.51) supported the conclusion that substructure **B** was an amide unit connected to a propyl chain.

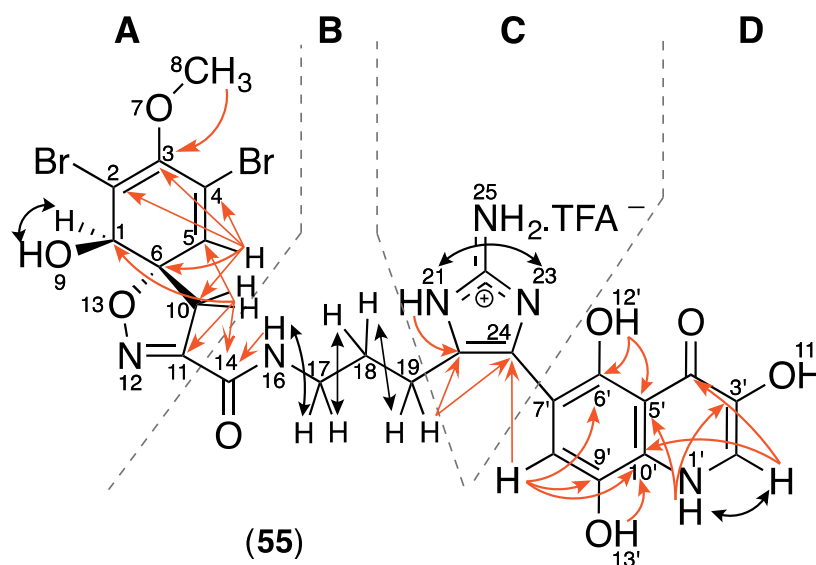


Figure 8: Structure and selected 2D-NMR correlations of new bromotyrosine compound (55). Colours: orange (HMBC), black (COSY).

HMBC correlations between H-19 and C-20 and C-24 (see Figure 9) indicated that the CH_2 at δ_{C} 21.2 ppm was connected to the imidazole ring. The w-COSY correlation between two broad singlets at δ_{H} 11.93 (N-21H) and δ_{H} 12.15 (N-23H), and the 2H singlet at δ_{H} 7.77 suggested the imidazole is present in the form of an imidazolium salt. The presence of a homohistamine unit is commonly observed in verongid metabolites¹⁷⁴ and was also found in other metabolites from this sponge. For example, aerophobin, pseudoceratinin A and purealin show similar NMR data for this substructure and provided good evidence for unit **C**. Although there were no correlations to C-22, a quaternary carbon at δ 146.6 was observed, which matches literature values for 2-aminoimidazole. Because of resonance and long T1 relaxation times, it is common not to observe correlations with guanidine carbons.

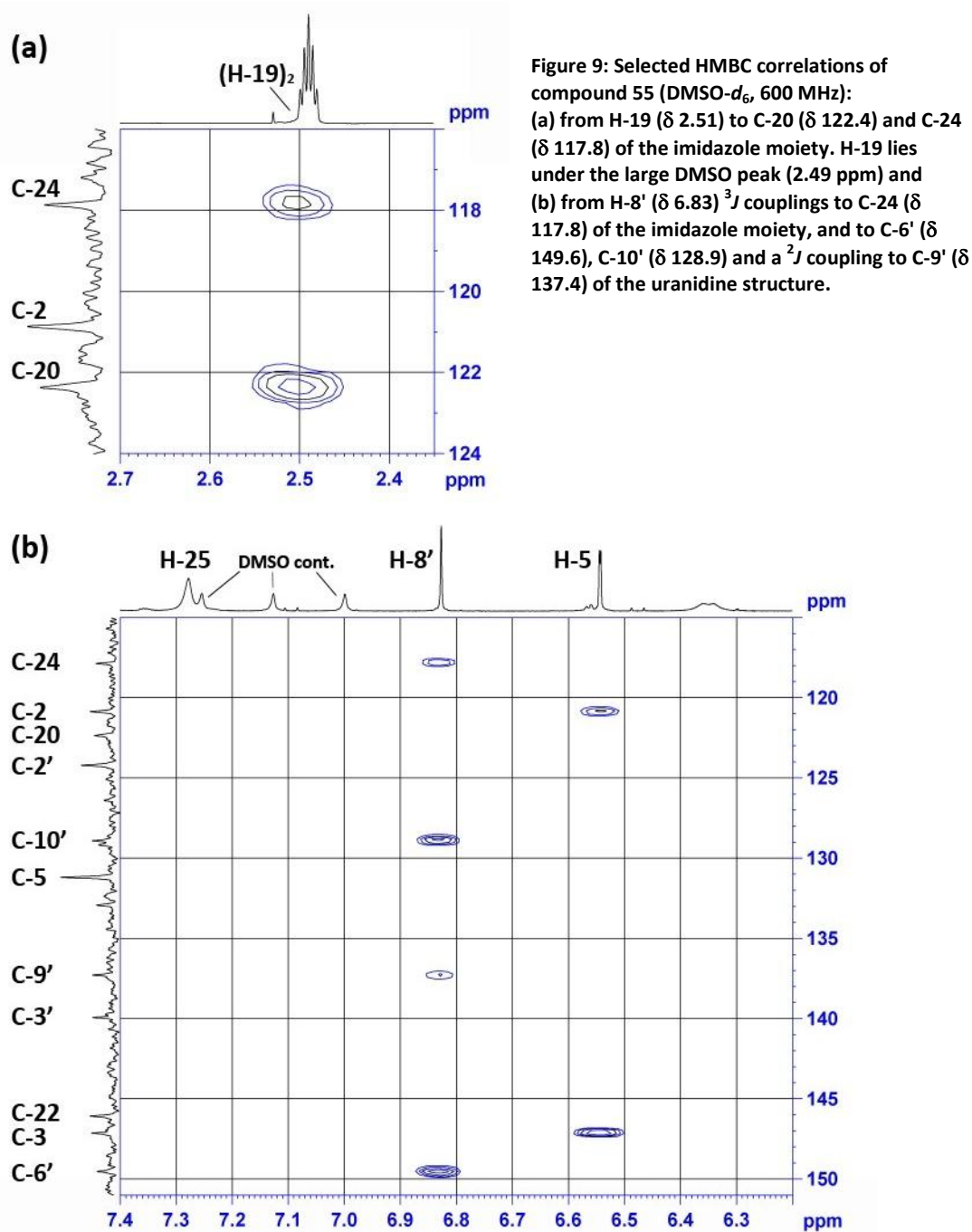


Figure 9: Selected HMBC correlations of compound 55 (DMSO-*d*₆, 600 MHz): (a) from H-19 (δ 2.51) to C-20 (δ 122.4) and C-24 (δ 117.8) of the imidazole moiety. H-19 lies under the large DMSO peak (2.49 ppm) and (b) from H-8' (δ 6.83) 3J couplings to C-24 (δ 117.8) of the imidazole moiety, and to C-6' (δ 149.6), C-10' (δ 128.9) and a 2J coupling to C-9' (δ 137.4) of the uranidine structure.

The HMBC correlations in the trihydroxy quinolinone ring system (H-1' to C-2' and C-10') matched those previously reported for 7-substituted 3,6,9-trihydroxy quinolinone²¹⁵. In addition, H-8' also showed a correlation to C-10' (see Figure 9b), which could only be observed with 7-substitution (*c.f.* from H-7' a $^4J_{CH}$ coupling would not be observable). If the uranidine moiety was attached at C-8'; $^3J_{CH}$ couplings should be observed from H-7' to C-9' and C-5'. The site of attachment at the imidazole was clearly indicated by strong correlations between C-24 and H-19 and H-8' (see Figure 9). Further evidence for the presence of unit **D** arose from 1H - 1H coupling between H-2' (δ_H 7.64, d, J = 5.8 Hz) and H-1' (δ_H 11.76, d, J = 5.8 Hz).

The chemical shifts for subunits **A**, **C** and **D** all aligned very well with those described for ceratinadin B²¹⁷, a compound similar to the one described here. However, ceratinadin B contains a histamine unit instead of homohistamine (subunit **B**) and has a molecular weight 14 lower than this compound. Due to the similarity of these two compound, and with ceratinadins A-C already reported, we suggest to name the new bromotyrosine described within this study as ceratinadin D. Finally, the optical rotation suggested the same absolute stereochemistry as ceratinadin B.

Table 7: NMR (DMSO-*d*₆, 600 MHz) data for new bromotyrosine compound (55).

Pos.	δ_c	δ_H , m (J in Hz)	COSY (H no.)	¹ H- ¹³ C HMBC (C no.)
1	73.5	3.90 d (7.7)	9	3, 4, 5, 6
2	120.9	-		
3	147.6	-		
4	113.1	-		
5	131.2	6.54 s	1	2, 3, 4, 6(w), 10
6	90.1	-		
8	59.7	3.63 s		3
9	-	6.36 d (7.7)	1	1(w), 4(w), 6(w)
10	39.0	3.59 d (3.6), 3.15 d (3.6)	10	1, 5, 6, 11
11	154.5	-		
14	158.9	-		
16	-	8.51 t (5.8)	17	14
17	38.2	3.13 m	18	18, 19
18	28.1	1.73 m	17, 19	17, 19
19	21.6	2.51 m	18	18(w), 20, 24
20	122.4	-		
21	-	11.93 bs	23	20(w)
22	146.6	-		
23	-	12.15 bs	21	—
24	117.8	-		
25	-	7.28 bs		
1'	-	11.76 d (6.4)	2'	3', 5'
2'	124.2	7.64 d (6.4)	1'	4', 10'
3'	139.9	-		
4'	173.2	-		
5'	112.6	-		
6'	149.6	-		
7'	103.2	-		
8'	113.7	6.83 s		24, 6', 9', 10'
9'	137.4	-		
10'	128.9	-		
11'	-	8.93 bs		
12'	-	14.3 bs		5', 6'(w)
13'	-	10.25 bs		10'

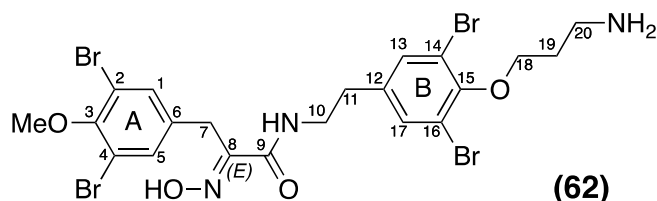
a (w) denotes a weak correlation.

2.2.3.2.2 Bromotyrosine (**60**)

Compound **60** was a simple bromotyrosine derivative showing the typical UV absorption at 278 nm and M+H cluster (m/z 698, 700, 702, 704, 706) suggesting the presence of four bromine atoms (ratio of M+H 1:4:6:4:1).

HRMS was consistent with a molecular formula $C_{21}H_{23}Br_4N_3O_4$ (3.8 ppm error). Three characteristic bands in the infrared spectrum, at wavenumbers 3572 (O-H), 1672 (C=N) and 1001 (N-O), confirmed the presence of an oxime.

These data suggested the compound was aplysamine-4²²¹ (**62**; atom numbering as in original report by Jurek *et al.*).



However, comparison of the 1H -NMR spectrum (see Table 8) reported for aplysamine-4 differed to our compound in the following ways: In aplysamine-4, the aromatic protons from ring A and B are all at $\delta_H \sim 7.44$. We found two different environments for the aromatic ring protons, where H-1/H-5 was at δ_H 7.44 but H-13/H-17 were at 7.55. This could be attributed to the use of different solvents but what was also different was the chemical shift of the ethyl and propyl-chain carbons. In aplysamine-4 the propyl group terminated with an ammonium ion (C-20 = 39.0 ppm) whereas in our compound, this carbon was at 36.2 ppm. Similarly the carbon shifts for the ethyl group did not match those of aplysamine-4 (see Table 8).

Table 8: 1H - and ^{13}C -NMR data of bromotyrosine aplysamine-4 in d_4 -methanol (reproduced from Jurek *et al.* 1993²²¹).

Pos.	δ_c	δ_H , m (J in Hz)	Pos.	δ_c	δ_H , m (J in Hz)
1	134.5	7.43 s	12	140.3	-
2	118.6	-	13	134.4	7.44 s
3	152.1	-	14	118.7	-
4	118.6	-	15	152.1	-
5	134.5	7.43 s	16	118.7	-
6	137.4	-	17	134.4	7.44 s
7	28.8	3.81 s	18	71.6	4.06 t (5.8)
8	152.2	-	19	29.0	2.18 tt (5.8, 7.8)
9	165.5	-	20	39.0	3.29 t (7.8)
10	41.3	3.43 t (7.2)	MeO	61.0	3.81 s
11	35.2	2.75 t (7.2)	NH ₂	-	Not observed.

The ^{13}C NMR spectrum of our compound indicated the presence of 21 carbons (see Table 9), one O-methyl (δ 60.4), six methylenes (δ 71.3, 39.4, 36.2, 31.5, 29.6, 27.9), four sp^2 methines (δ 133.2×2 , 132.9×2), nine sp^2 quaternary (δ 151.8, 151.3, 151.0, 136.8, 136.3, 117.6×2 , 117.1×2) and one amide carbonyl (δ 163.0). All protonated carbons were assigned by HSQC experiment. Use of one and two-dimensional NMR data (see Table 9) enabled the construction of three substructures (see Figure 10). Inspection of ^1H , ^{13}C , and ^1H - ^1H COSY NMR spectra suggested that the following proton signals belonged to substructure **A**: δ_{H} 12.02 (oxime), 7.44 (sp^2 methine), 3.76 (sp^3 methine) and 3.75 (OMe). HMBC correlations from H-2 to C-2, C-3, H-6 to C-4, and H-7a/b to C-8 and C-11 provided good evidence for the oxime-tyrosine unit **A**. The connection between substructures **A** and **B** was obtained by HMBC correlations from H-7a/b to carbonyl C-11 and from the amide proton H-13 to oxime carbon C-8. C-8 and C-11 could be easily distinguished by their typical chemical shifts (see Table 9) as reported in the literature¹⁷⁴. The geometry of the oxime was determined as *E* from the up-field ^{13}C chemical shift of C-7 (δ_{C} 27.9), *c.f.* δ_{C} 35.7 for *Z*- observed in (*E,Z*)-*N,N'*-bis-[3-(3'-bromo-4'-hydroxyphenyl)-2-oximidopropionyl] cystamine²²², which has been used as a reference in determining the geometry of other bromotyrosine alkaloids^{174,204,213,221,223,224}.

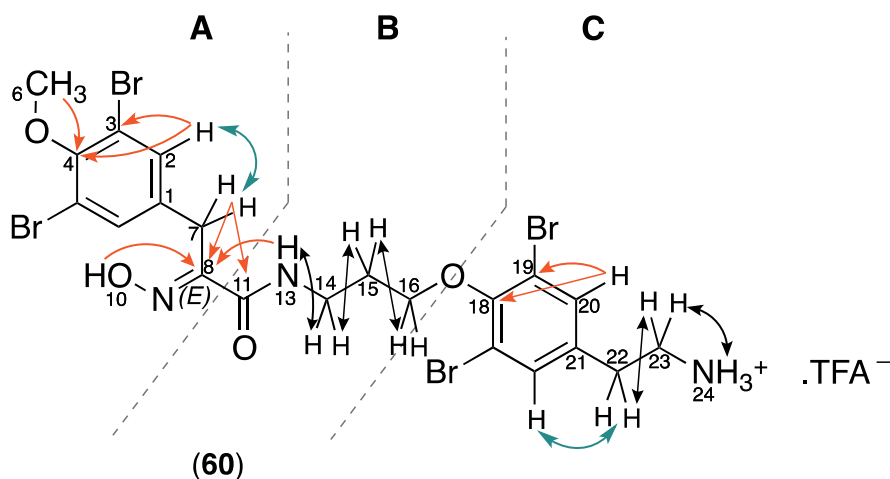


Figure 10: Structure and selected 2D-NMR correlations of new bromotyrosine compound (60).
Colours: orange (HMBC), blue (ROESY), black (COSY).

The amide at δ_{H} 8.12 showed coupling to a methylene (δ_{H} 3.38) assigned to H-14. Further COSY correlations to δ_{H} 1.96, and then to δ_{H} 3.88 supported the assumption that substructure **B** was an amide unit connected to a propyl chain. The protons at 3.88 ppm were attached to a carbon at 71.3 ppm (HSQC), which would only be explained by attachment of an oxygen, suggesting attachment to unit C via an ether.

Indeed, a weak correlation between C-18 (151.3 ppm) and H-16 was observed in the HMBC spectrum (see Figure 8) provided strong evidence for a link between the propyl chain and unit C.

The signal at δ_{H} 7.55 (H-20) showed HMBC correlations to sp^2 quaternary carbons (C-18, C-19) and a ROESY correlation to a methylene at δ_{H} 2.81/ δ_{C} 31.5, consistent with attachment of an alkyl group directly to the aromatic ring, in support of substructure C. Direct coupling of δ_{H} 2.81 to δ_{H} 3.05, and of δ_{H} 3.05 to the primary amine protons at δ_{H} 7.80 were in good accordance to literature values¹⁷⁴ for a tyramine unit.

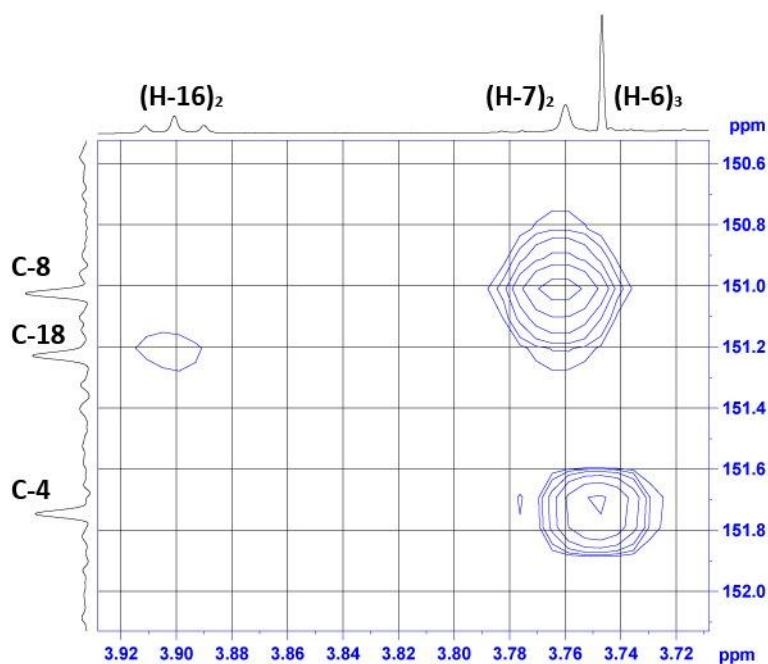


Figure 11: Selected HMBC correlations found in the unknown bromotyrosine (60; DMSO- d_6 , 600 MHz). A weak correlation between H-16 (δ 3.88) and C-18 (δ 151.3) indicates the link of substructures B and C.

Putting this together, we conclude that this is a new brominated tyrosine derivative (60).

Table 9: NMR (DMSO-*d*₆, 600 MHz) data for new bromotyrosine compound (60).

Pos.	δ_c	δ_H , m (J in Hz)	COSY (H no.)	1H - ^{13}C HMBC (C no.)	ROESY (H no.)
1	136.3	-			
2	132.9	7.44 s	7	3, 4	7
3	117.1	-			
4	151.8	-			
6	60.4	3.75 s		4	
7	27.9	3.76 s	2	1, 2, 8, 11	2
8	151.0	-			
10	-	12.02 s		8(w)	
11	163.0	-			
13	-	8.12 t (6.0)	14	14	7, 14, 16, 15(w)
14	36.2	3.38 m	13, 15	11, 15, 16	15, 16
15	29.6	1.96 m	14, 16	14, 16	13(w), 14, 16
16	71.3	3.88 t (6.4)	15	14, 15	14, 15
18	151.3	-			
19	117.6	-			
20	133.2	7.55 s		18, 19, 22	22, 23
21	136.8	-			
22	31.5	2.81 t (7.4)		20, 21, 23	20, 23
23	39.4	3.05 m	22, 24	21(w)	20, 22, 24
24	-	7.81 b	23	-	23

a (w) denotes a weak correlation.

2.2.3.3 MTT microdilution assay

The purified bromotyrosine compounds from the bioactive ethyl acetate fraction of *P. purpurea* were subjected to antimicrobial MTT-microdilution assays against *E. coli* and *S. aureus*. The crude Sephadex fractions did not display any inhibitory effect on *P. aeruginosa*. Hence, no microdilution tests were performed against this bacterium. MTT tests of compounds **44** and **54-61** were performed as triplicates. The determined MICs of bioactive compounds are summarised in Table 10. Four of the tested brominated tyrosines (**57-60**) inhibited the growth of both *E. coli* and *S. aureus*. The other five compounds did not show any activity.

Table 10: MICs determined from MTT assay of HPLC purified bromotyrosines from *P. purpurea*.

Compound	MIC ($\mu\text{g/mL}$) against bacterial strains	
	<i>E. coli</i>	<i>S. aureus</i>
Hexadellin (57)	250 (n.r.)	250 (n.r.)
Aplysamine-2 (58)	250 (n.a. ²⁰⁴)	125 (n.a. ²⁰⁴)
16-debromoaplysamine-4 (59)	250 (250 ²²⁵)	125 (200 ²²⁵)
New bromotyrosine (60)	125	31.3
Ampicillin	1.3 (0.25-1.0)	1.1 (< 0.25)

MIC expressed as median of replicates (n = 3). Reference values in brackets. n.r. = not reported. n.a. = not active in literature. Ampicillin control reference ranges²²⁶

The previously unknown bromotyrosine (**60**) showed an MIC of 31.3 $\mu\text{g/mL}$ against *S. aureus*. As compound (**60**) exhibited significantly higher activity than any other isolated compound, it could be that it accounts for most of the crude ethyl acetate extract's activity discussed in Section 2.2.2.1.1. However, the results from the MTT microdilution assay used here and the disc diffusion assay used in Section 2.2.2.1.1 to assess the crude extracts are not directly comparable. The MTT microdilution test suggest the isolated bromotyrosines had only moderate activity.

The high antibacterial activity observed for the crude extract may be due to a cumulative effect of several weak antibacterially active compounds. Additionally, synergistic effects of the compounds isolated, or minor constituents that have not been isolated could be responsible for an overall higher antibacterial activity of the crude extract. In such cases, the bioactivity (or toxicity) has been reported to be enhanced several-fold by addition of as little as a few precents of minor constituents, which hardly show activity by themselves¹⁴⁴.

There is no doubt that high throughput screening (HTS) approaches can be used to isolate the bioactive principles of extracts in shorter time spans than the setup used for this study allowed to. Recent technical developments for example allow for an automated use of solid phase extraction cartridges to desalt the extracts; directly link LCMS analysis to HTS bioassays in multiwell formats; or utilise FTICR-MS or LC-UV/MS-SPE-NMR workflow procedure^{227,228}. Additionally, the analysis of complex mixtures in the field of metabolomics has driven extensive efforts to develop software and spectral databases capable of classifying spectra and detecting the presence of individual specific compounds within a mixture of hundreds of small molecules without requiring compound isolation²²⁹.

The major goal in this study was to perfect experience in isolating rare marine natural products, and the identification of at least one novel compound. The combination of applying simply bioassays to lead the way to separation revealed only limited activity remained for the individual compounds, but yielded two new structures.

Although possible, the application of the isolated compounds in reverse chemical proteomics (see Chapter 4) is of limited interest due to stability issues observed for most bioactive brominated tyrosines²³⁰. Hence, their prospects for development as therapeutic agents are remote so these compounds were not prioritised for the work described in Chapters 3 and 4.

2.2.4 *Tylodina corticalis*

During the sample collection of *Pseudoceratina purpurea* (JB07-S1) a bright yellow opisthobranch was co-collected while feeding on this sponge. This was an opportunity to elucidate similarities in the chemical composition of prey and predator; identified as *Tylodina corticalis* (Tate 1889; family Tylodinidae; superfamily Umbraculoidea; class Gastropoda; phylum Mollusca).

Although opisthobranchs of the genus *Tylodina* are found in distant regions, they are exclusively associated with sponges of the order Verongida (in particular the family Aplysinellidae)²³¹. *Tylodina corticalis* typically feed on the genus *Pseudoceratina*²³². It has also been shown, that they accumulate the secondary metabolites from their prey, including many bromotyrosine compounds, sequestered and utilised as feeding deterrents for themselves and their eggs^{231,233-235}.

2.2.4.1 LC-MS analysis of ethyl acetate partition

The extraction of *T. corticalis* differed from the general procedure (see Section 2.2.1) to minimise degradation of any secondary metabolites. An ethanolic extract of the entire organism was obtained at low temperature and subsequent partitioning with DCM, EtOAc and BuOH, respectively. All partitions were checked by LC-MS and the ethyl acetate fraction revealed a similar chromatogram to the one obtained from the ethyl acetate partition of *P. purpurea* (EA-S-9-22; see Section 2.2.3.2).

The UV and MS chromatograms of the ethyl acetate fractions of mollusc and sponge JB07-S1-EA-S-9-22 were compared to identify compounds present in both species (see Figure 12). The UV spectra and correlating molecular ions of each peak were matched with the compounds isolated from JB07-S1-EA-S-9-22 (see Figure 14).

In the UV trace of JB07-S1-EA-S-9-22 the first two major peaks show tailing, indicating a strong interaction with the silica or overload of the column. The same peaks in the *T. corticalis* extract were tiny and with a slightly longer retention time. Injecting volumes above saturation level of the HPLC column commonly results in shorter retention times but as the UV spectra and MS matched they were shown to be identical. Very polar or amphiphilic compounds such as the bromotyrosines found in the verongid sponges can also have slight variations in retention times depending on the precise composition of the injection solvent, or small variations in the solvent gradient and modifier concentrations. Hence, the information on UV absorption and

corresponding MS data (isotopic [M+H]⁺ cluster) of compounds with close retention times was prioritised to identify each component.

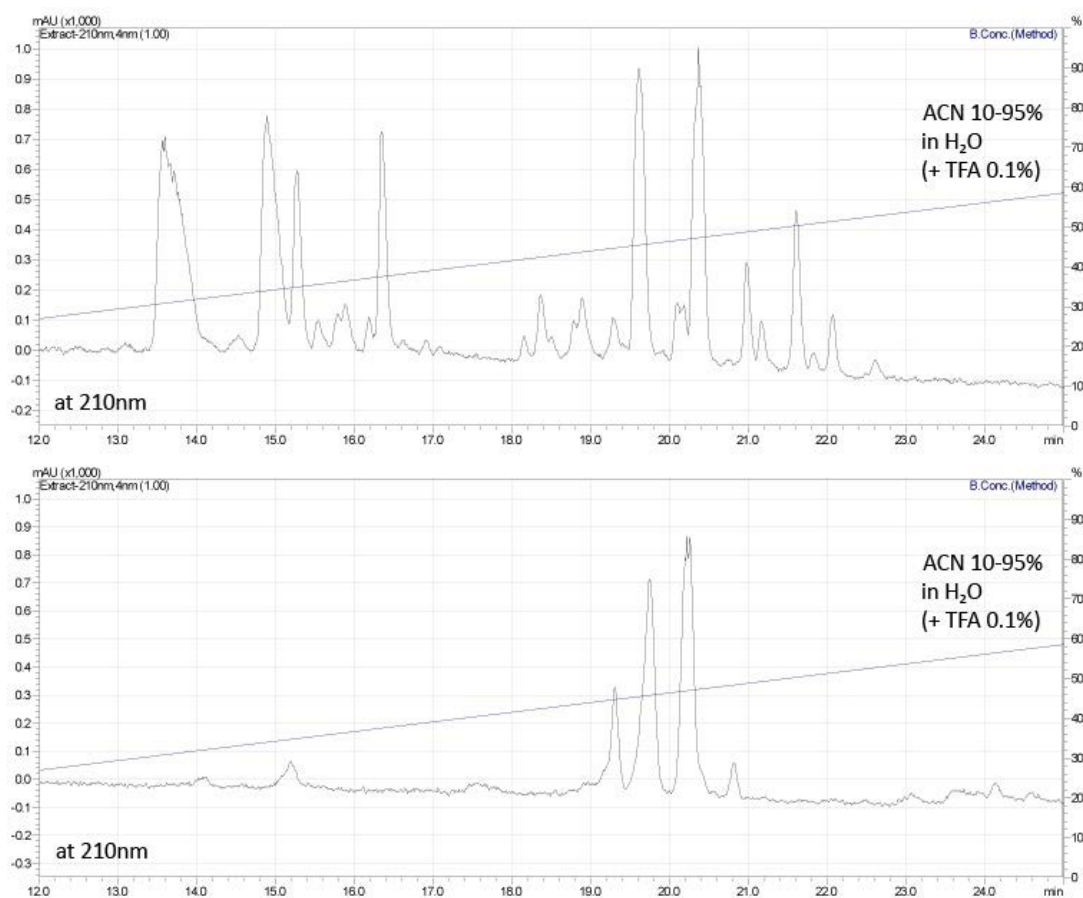


Figure 12: RP-C18 HPLC UV traces (at 210 nm) of bioactive Sephadex fractions of ethyl acetate partitions of *P. purpurea* and the crude EtOAc-partition of *T. corticalis*.

The first compound appearing in the UV trace from *T. corticalis*, (13.7-14.2 min) showed similar spectroscopic properties as the first major peak from the sponge extract eluting at 13.4–14.0 min, which was previously identified as pseudoceratinine A (**54**; see Section 2.2.3.2). They both displayed the typical UV absorption at 286 nm and M+H⁺ cluster (m/z 490, 492, 494) suggesting the presence of two bromine atoms (ratio of M+H⁺ 1:2:1). The difference in elution time was attributed to a slight overload of the HPLC column with pseudoceratinine A. These data suggested that the two compounds were identical and hence isolated from both prey and predator. It was isolated as a major metabolite from the sponge, yet, only present as a trace constituent in the ethylacetate extract from the mollusc. Pseudoceratinine A has not previously been reported from *Tyrodina* sp.

Similarly, the second peak visible in the UV trace of the mollusc extract (at 15.1 min) matched the isotopic cluster (m/z 502, 504, 506; ratio of M+H⁺ 1:2:1) and the UV

maxima (226, 285 nm) of the compound eluting at 14.7-15.2 min from the sponge extract. This compound was identified as aerophobin-2 (**44**) in Section 2.2.3.2. Aerophobin-2 has previously been isolated from *Tylodina* sp. The mollusc commonly sequesters this compound and bioconverts it into aeroplysinin-2 (**45**) and the dienone (**46**), both highly active feedings deterrents (see Figure 4)¹⁵⁰.

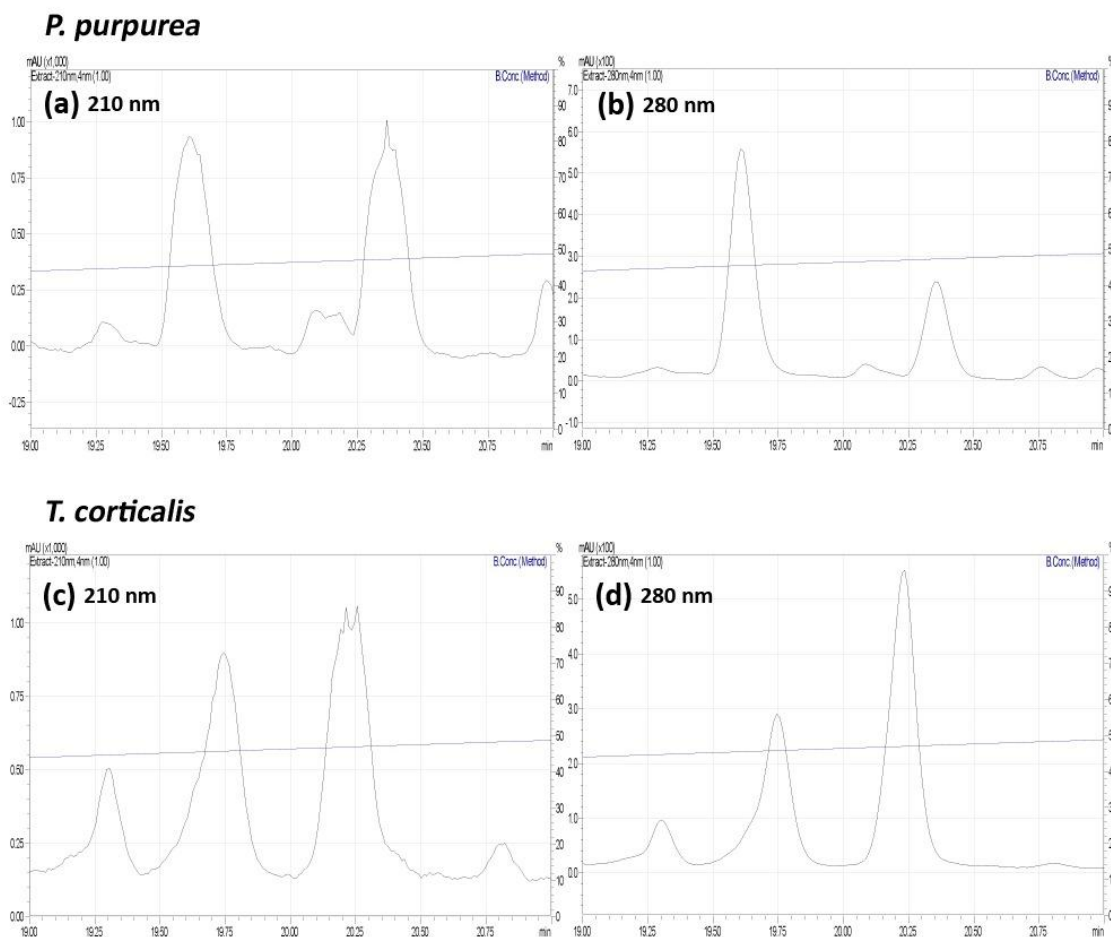


Figure 13: UV traces of the bioactive Sephadex fractions from the ethyl acetate extract of *P. purpurea* at 210 nm (a) and 280 nm (b) and of the crude ethyl acetate fraction of *T. corticalis* at 210 nm (c) and 280 nm (d).

The majority of compounds extracted from *T. corticalis* eluted after 19-21 min. Comparison of the UV traces at 210 and 280 nm of both sponge and mollusc ethyl acetate extracts (see Figure 13) suggested only limited matches. The UV traces at 340 nm did not show any peaks indicating that the compounds containing the uranidine pigment are retained by the mollusc. However, in combination with the correlating MS data several similarities were found (see Figure 14).

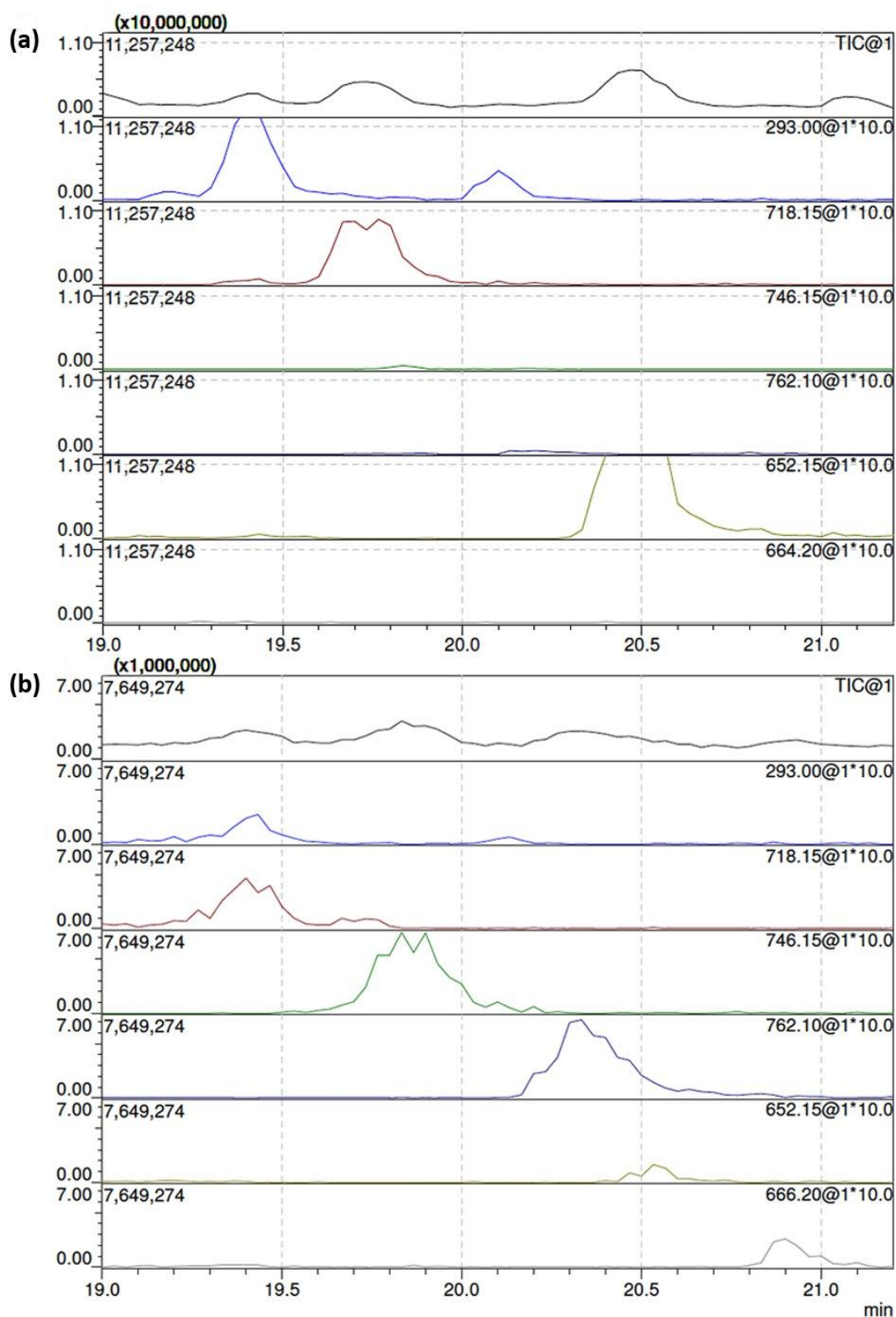


Figure 14: MS chromatograms (TIC) and selective ion traces of ethyl acetate extracts of (a) *P. purpurea* and (b) *T. corticalis*. For brominated compounds, the selective ion fragments represent the lowest centre mass peaks of isotopic clusters (for improved MS signal).

The presence of a non-brominated compound with a molecular ion of 293 (M+H) was observed in both extracts at identical retention times (19.4 and 20.1 min). It is possible that this is a sterol or fatty acid, but UV and MS data were insufficient to clearly identify the compound(s). Interestingly, the major compounds found in the mollusc were not detected in the sponge (porealidin T and P). As the feeding history of the mollusc is unknown, it is possible that these compounds arise from the mollusc feeding on other related sponges with a slightly different metabolite mix. It is also possible that the mollusc accumulates very minor components from the sponge or modifies dietary compounds for its own purposes.

Table 11: Comparison of HPLC retention times, UV and MS data from brominated compounds of *P. purpurea* (JB07-S1-EA-S9-22) and *T. corticalis* (EtOAc).

R_t (MS) (min)	λ max (nm)	m/z (^{79}Br only)	n_{Br}	<i>P. purpurea</i>	<i>T. corticalis</i>	compound name
13.7-14.2	226, 286	490.15	2	+	+	pseudoceratinin A
14.67	232, 286	494.15	2	+		porealidin L*
15.1-15.4	228, 285	504.15	2	+	+	aerophobin-2
15.41	234, 276, 340	681.20	2	+		ceratinadin B
15.61	-	392.55	2	+		?
16.00	-	414.45	1	+		?
16.35	234, 276, 337	695.25	2	+		new (ceratinadinD)
18.48	209, 286	490.10	2	+		pseudoceratine* or porealidine J*
19.00	209, 286	504.15	2	+		?
19.41	210	293.00	0	+	+	sterol*
19.3-19.5	223, 280	714.05	4	~	+	hexadellin B*
19.6-19.8	223, 280	714.05	4	+	~	hexadellin A
19.85	203, 280	742.10	4	~	+	porealidin P or Q*
20.08	210	293.00	0	~	~	sterol*
20.13	203, 280	762.10	4		+	porealidin T*
20.30	impure	634.20	3	+		suberedamine A*
20.47-20.53	222, 280	648.20	3	+	+	aplysamine-2
20.90	234, 279	664.25	3		+	purpuramine J*
20.98	210, 278	620.10	3	+		16-debromo-aplysamine-4
21.71	209, 278	698.15	4	+		new
22.25	209, 278	880.20	4	+		purealin
24.27	240, 308, 363	405.10	2		+	?

* tentative assignment based on MS and UV. ~ indicates trace amounts detected.

Grey indicates similarities between *P. purpurea* and *T. corticalis*.

Italic indicates low accuracy for UV spectrum due to low sample concentration

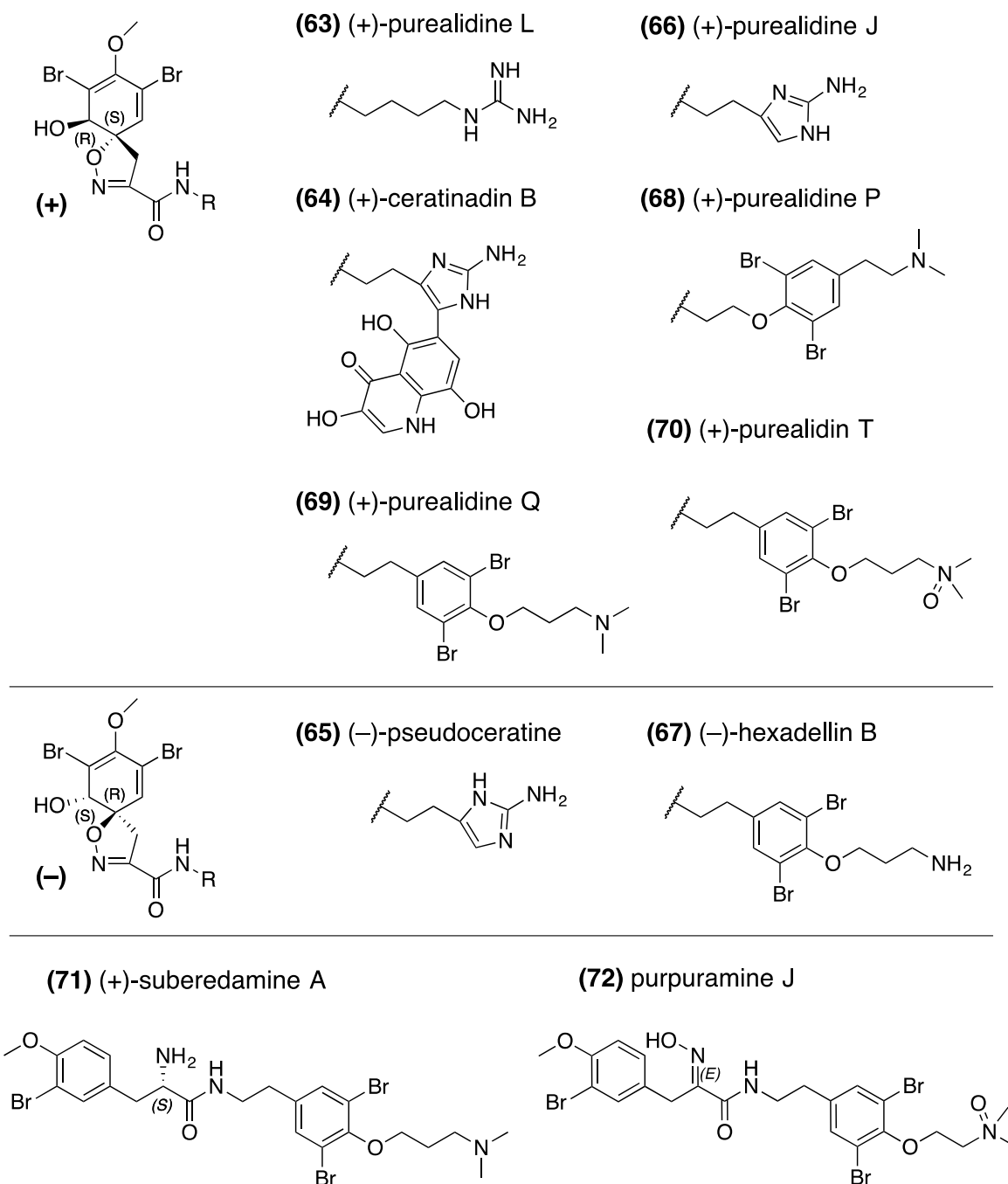


Figure 15: Bromotyrosine derivatives identified from the ethyl acetate extracts of *P. purpurea* and *T. corticalis* based on UV and MS analysis. Stereochemistry shown as reported in literature.

2.2.4.2 High Resolution LC-MS of ethyl acetate partitions

The ethyl acetate partitions of *P. purpurea* and *T. corticalis* were consequently analysed by HR-LC-MS to support the tentative identification of the compounds described in Section 2.2.4.1. The chromatography conditions from Section 2.2.4.1 were adapted and optimised for the new instrumentation setup, aiming for best

separation of the ethyl acetate fraction of *T. corticalis*. These optimised conditions were applied identically to the ethyl acetate fraction of *P. purpurea*.

The retention times of all eluting compounds varied slightly compared to the LR LC-MS analysis, but due to the mass, the elution time and the isotopic cluster pattern (see Figure 16 and Section 8.1.2.2) the data of both LR- and HR-LC-MS could easily be matched. For simplicity of presentation, all HR-MS data were correlated to the retention times arising from the LR-LC-MS experiments (see Table 12).

Similar to the low resolution analysis, the ethyl acetate fraction of *P. purpurea* resulted in partial overload of the MS detector, causing peak broadening and splitting of the four largest peaks (see Figure 16). However, this did not impact the quality of analysis or the accuracy of the masses of the analysed compounds.

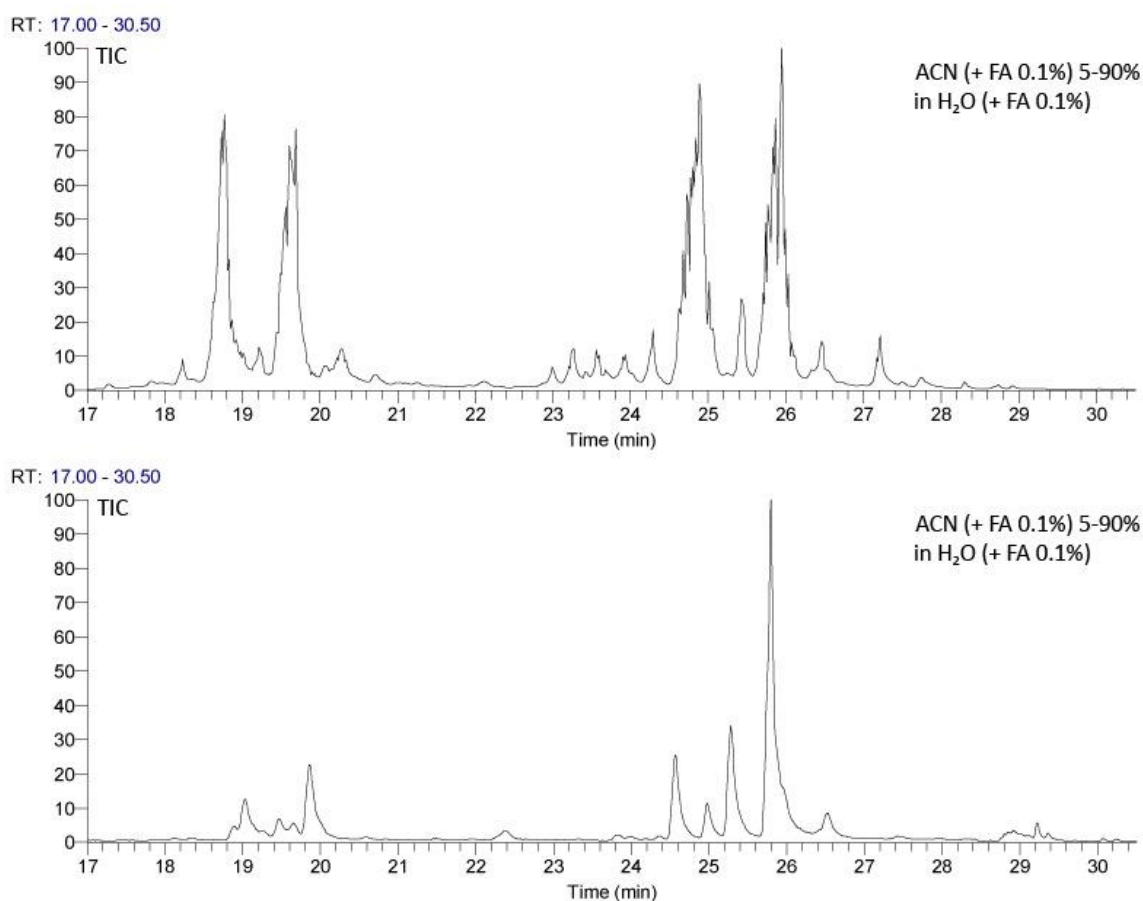


Figure 16: High resolution LC-MS chromatograms (TIC) of bioactive Sephadex fractions of ethyl acetate partitions of *P. purpurea* (JB07-S1-EA-S-9-22) and the crude EtOAc-partition of *T. corticalis*.

For all compounds from Table 11 the molecular formulas were calculated based on the accurate mass and the number of bromines present (using XCalibur software). The calculations were restricted by an error of less than 1 ppm accuracy, which mostly resulted in only one suitable formula. However, in those cases where two or

more formulas fitted the exact mass within 1 ppm accuracy, each of the peaks from the isotopic clusters were analysed with the appropriate number of ^{79}Br and ^{81}Br atoms present in the formula. This allowed identification the most suitable molecular formula for all compounds. The data are summarised in Table 12.

The majority of identified molecular formulas were in agreement with the structures tentatively assigned and presented in Section 2.2.4.1. However, two constituents, eluting at 19.00 and 24.27 min respectively, are not present in the current scientific literature. Unfortunately, there was insufficient material available to isolate the quantities required for NMR analysis (< 200 μg).

The compound eluting at 18.48 min showed an accurate mass of 489.9719 amu ($[\text{M}+\text{H}]^+$), which matched the molecular formula ($\text{C}_{15}\text{H}_{17}\text{N}_5\text{O}_4\text{Br}_2$; 0.805 ppm) of pseudoceratine (**65**), purealidine J (**66**) or purealidine M (**66a**). The UV spectrum (λ_{max} 209 and 286 nm) however, suggest purealidine M as the best fit, given that the spiro-cyclohexadienylisoxazolines of **65** and **66** should absorb around 230 and 286 nm¹⁷⁴.

The compounds eluting from 19.3 to 19.8 min matched previously described hexadellin A (**57**) or B (**67**; see Section 2.2.4.1) and was found in both the sponge and mollusc extracts. The compound was identified as hexadellin A by NMR analysis of a purified fraction from the sponge.

Finally, suberedamine A could not be verified as the compound eluting after 20.30 min. Its molecular formula of $\text{C}_{23}\text{H}_{30}\text{Br}_3\text{N}_3\text{O}_3$ correlates to an accurate mass of 633.9915 ($[\text{M}+\text{H}]^+$). Yet, the accurate mass found for this peak was 633.9557 amu ($[\text{M}+\text{H}]^+$), fitting best to $\text{C}_{22}\text{H}_{27}\text{N}_3\text{O}_4\text{Br}_3$ (0.513 ppm). This formula suits purpurealidine H (**72a**), and purpuramine I (**72b**) and L (**72c**). As there is insufficient UV data available for this peak, it is unfortunately impossible to confirm the correct identification.

Overall, it is clear that the genus *Tylodina* accumulates and sequesters sponge bromotyrosine metabolites for their own chemical defense²³⁶. The uptake of the yellow sponge pigment (uranidine) further gives the distinct colour to the mollusc, providing a passive defence (camouflage). Proksch and coworkers have reported several studies on the feeding habits of the mediterranean *T. perversa* on its food source *Aplysina aerophoba* and *A. cavernicola*^{150,169,231,233,237,238}. For example, during this study aerophobin-2 was identified from *T. corticalis* and its food source *P.*

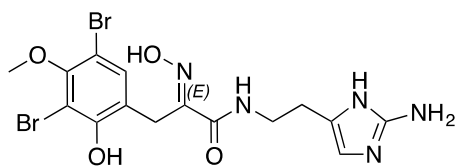
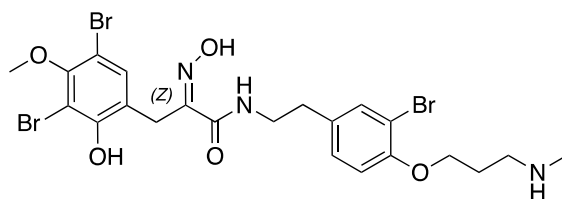
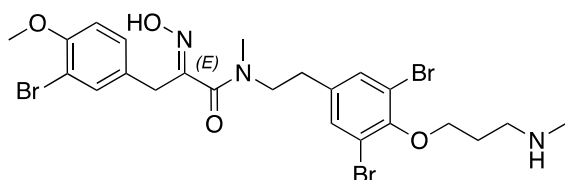
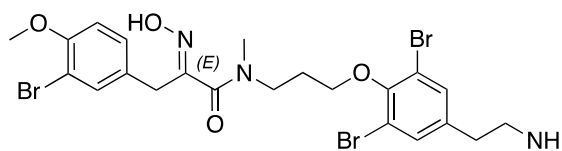
Table 12: High resolution MS data matched to the compounds described in Table 11. The molecular formulas were calculated based on *n* atoms of bromine.

R_t (LR-MS) (min)	λ max (nm)	LR m/z (⁷⁹ Br only)	HR m/z (⁷⁹ Br only)	$n_{(Br)}$	molecular formula [MF + H] ⁺	error (ppm)	<i>P. purpurea</i>	<i>T. corticalis</i>	compound name
13.7- 14.2	226, 286	490.15	489.9718	2	C ₁₅ H ₁₈ N ₅ O ₄ Br ₂	0.601	+	+	pseudoceratinin A
14.67	232, 286	494.15	494.0031	2	C ₁₅ H ₂₂ O ₄ N ₅ Br ₂	− 0.619	+	+	purealidine L
15.1- 15.4	228, 285	504.15	503.9878	2	C ₁₆ H ₂₀ N ₅ O ₄ Br ₂	0.485	+	+	aerophobin-2
15.41	234, 276, 340	681.20	680.9969	2	C ₂₄ H ₂₃ N ₆ O ₈ Br ₂	− 0.317	+		ceratinadin B
15.61	-	392.55		2			+		?
16.00	-	414.45		1			+		?
16.35	234, 276, 337	695.25	695.0096	2	C₂₅H₂₅N₆O₈Br₂	− 0.307	+		new (ceratinadinD)
18.48	209, 286	490.10	489.9719	2	C ₁₅ H ₁₈ N ₅ O ₄ Br ₂	0.805	+		purealidine M*
19.00	209, 286	504.15	503.9878	2	C ₁₆ H ₂₀ N ₅ O ₄ Br ₂	0.088	+		unknown
19.41	210	293.00	-	0			+	+	sterol*
19.3- 19.8	223, 280	714.05	713.8458	4	C ₂₁ H ₂₄ N ₃ O ₅ Br ₄	− 0.490	+	+	hexadellin A
19.85	203, 280	742.10	741.8754	4	C ₂₃ H ₂₈ N ₃ O ₅ Br ₄	− 0.587	~	+	purealidin P or Q*
20.08	210	293.00	-	0			~	~	sterol*
20.13	203, 280	758.10	757.8897	4	C ₂₃ H ₂₈ N ₃ O ₆ Br ₄	− 0.682		+	purealidin T*
20.30	impure	634.20	633.9557	3	C ₂₂ H ₂₇ N ₃ O ₄ Br ₃	0.513	+		purpurealidine H*, or purpuramine I or L*
20.47- 20.53	222, 280	648.20	647.9705	3	C ₂₃ H ₂₉ N ₃ O ₄ Br ₃	0.189	+	+	aplysamine-2
20.90	234, 279	664.25	663.9655	3	C ₂₃ H ₂₉ N ₃ O ₅ Br ₃	− 0.430		+	purpuramine J*
20.98	210, 278	620.10	619.9397	3	C ₂₁ H ₂₅ N ₃ O ₄ Br ₃	− 0.039	+		16-debromo-aplysamine-4
21.71	209, 278	698.15	697.8505	4	C₂₁H₂₄N₃O₄Br₄	0.091	+		new
22.25	209, 278	880.20	879.8942	4	C ₂₇ H ₃₀ N ₇ O ₇ Br ₄	0.872	+		purealin
24.27	240, 308, 363	405.10	404.9809	2	C ₁₄ H ₁₉ N ₂ O ₂ Br ₂	− 0.572		+	unknown

* tentative assignment based on HR-MS and UV. ~ indicates trace amounts detected.

Grey indicates similarities between *P. purpurea* and *T. corticalis*.

Italic indicates low accuracy for UV spectrum due to low sample concentration

(66a) purealidine M**(72a)** purpurealidine H**(72b)** purpuramine I**(72c)** purpuramine L

purpurea. It is well-known that members of this mollusc family can accumulate and use highly active feeding deterrents aeroplysinin-2 (**45**) and the dienone (**46**) (see Section 2.1.5), but even more so accumulate the protoxins, such as isofistularin-4 (**43**) or aerophobin-2 (**44**).

The finding of identical bromotyrosines in verongid sponges and *Tylodina* spp, here and in published literature, highlights how closely the biochemistry of these two genera is intertwined.

However, studies have suggested that the majority of the biomass of *Tylodina*'s prey is comprised of sponge-associated cyanobacteria rather than poriferan tissue²³⁵ and challenge both the carnivorous classification of the opisthobranch and the biogenetic origin of bromotyrosines. Our observations do not support this as we observed *T. corticalis* feeding deeply on sponge tissue, where cyanobacteria are unlikely to reside. In addition, bromotyrosines are typical metabolites of verongid sponges and related structures have not been isolated from cyanobacteria.

This feeds into the ongoing debate in marine natural product chemistry on the origin of natural products, micro- or macro-organism. In either case, marine natural products remain a vast source for new compounds, compound classes and ultimately an immense pool for discovering new drug leads.

Chapter 3

Synthesis of Linkers, Reagents and Biotinylated Probes

3.1 Overview

Throughout the last 50 years, since Merrifield first used a cross-linked polystyrene resin for solid phase organic synthesis²³⁹, a myriad of linkers for diverse applications has been described. Some of the features required for solid support organic synthesis are to be cleavable, be it under nucleophilic, electrophilic, reductive or oxidative conditions, or be it metal assisted or photo induced²⁴⁰; to provide immobilised protecting groups; or to assist in monitoring solid-phase organic reactions, such as applied in single-bead direct analysis based on MALDI-TOF mass spectrometry²⁴¹.

When aiming to immobilize small molecules onto a solid support such as required in reverse chemical proteomics (RCP), it is most important to choose a linker suitable for the conditions of the biological system used. The prerequisites for RCP are water-solubility, length, shape and bio-orthogonal chemistry so that the linker survives the myriad of enzymes native to biological systems. To perform affinity selections on proteins, a linker also has to allow the protein access to the immobilised small molecule. This requires the analyte to be extending far enough away from the solid support so as not to interfere with the binding and allow the small molecule to penetrate into even the deepest binding site.

Every linker also needs a reactive end to ease conjugation with common functional groups present in drugs and natural products, such as carboxylic acids, amines, alcohols or phenols.

This chapter describes the synthesis of various hydrophilic linkers suitable for bioconjugation with natural products. The biotinylated linkers consist of long (33-38 Å) or short (16-17 Å) building blocks, which are readily combined utilising Huisgen 1,3-dipolar cycloadditions. Additionally, this chapter explores the milligram-scale derivatisation of bioactive natural products with these linkers without affecting biological activity.

3.2 Linkers and reagents

3.2.1 Poly(ethyleneglycol) linkers

3.2.1.1 Introduction

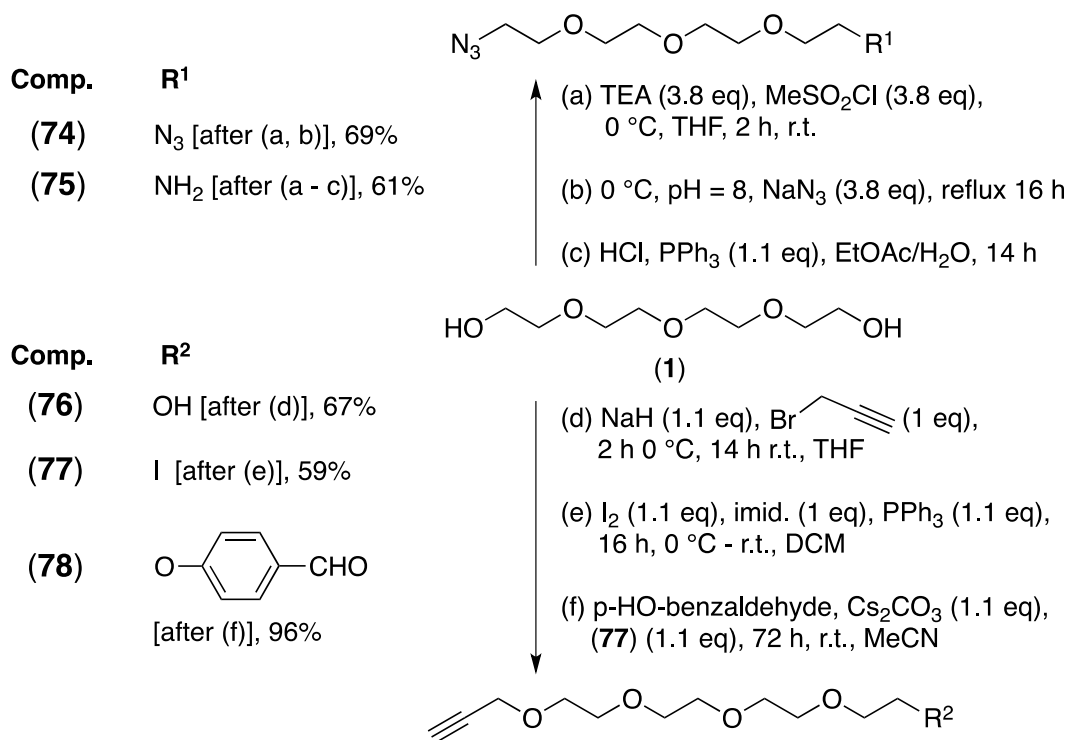
Polyethyleneglycol (PEG) constitutes an ideal backbone for synthesising probes for chemical proteomics due to its hydrophilicity, chemical stability and bioorthogonality. Monodisperse PEGs, such as tetraethyleneglycol (TEG), allow the synthesis and characterisation of discrete, fixed-length linkers, while incorporating

polydisperse PEGs increases display options due to their statistical distribution of lengths.

3.2.1.2 Results and discussion

Initially, different heterobifunctional (azide-amine; acetylene-alcohol; acetylene-halide) spacers were synthesised from commercially available TEG (**73**, see Scheme 1). Although longer, fixed-length PEG species are commercially available, these are usually very expensive and these costs can be easily bypassed by our approach. Azide and acetylene groups were incorporated to perform Cu(I)-catalysed azide-alkyne Huisgen 1,3-dipolar cycloaddition ("click chemistry") between two spacers and thereby create linkers of extended, fixed length, increasing the size from around 17 Å (for single TEG unit) up to 38 Å (for single TEG-triazole-TEG unit). The selectivity of Cu for 1,4-substituted triazoles, ensures the linkers are more or less linear. Utilising long spacers is particularly important in methods such as phage display (see 1.1.2.2.1.1). Such methods require immobilising an analyte (small molecule or peptide) on a solid support to allow for binding partners, such as single members from a protein library, to approach the analyte and initiate binding. To allow for this binding to take place, a minimum of steric hindrance must be presented and the analyte has to be orientated away from the solid support. The actual binding site may be located deep inside the protein so the linker should be long enough to allow the analyte unrestricted access. Non-specific interactions between the protein and the solid support must also not interfere, or sterically hinder, the binding to the analyte. For example, in human indoleamine-2,3-dioxygenase-1 (IDO-1) the active site spreads over 16 Å from the residue Ser₁₆₇ at the rear of the pocket to Arg₂₃₁ and the 7-propanoate arm of heme to either side of the entrance²⁴²⁻²⁴⁴. Small analytes with dimensions of less than 10 Å, such as various tryptophan-analogue inhibitors will end up entirely submerged in the binding pocket, therefore necessitating the attached spacer to firstly extend the probe far enough away from the solid support and secondly reach into the binding site itself. The binding site of a protein may be masked through a flexible loop or be concealed close to the protein's centre. Similar deep penetration is typical for kinases (ATP binding site) and tetrahydrofolate transferases (pterin binding site) for example). With these considerations in mind, we aimed to provide long, hydrophilic, narrow and extended linkers for our natural products.

To achieve long spacers, click chemistry was applied as it generally provides the benefit of very mild, highly effective and fast reaction conditions. The orthogonal nature of the chemistry means that it is safe for most natural products that might be sensitive to acid/base or oxidative/reductive conditions. Our only concern was the possible ability of the TEG chains to complex copper, hindering purification.



Scheme 1: Synthesis of heterobifunctional TEG-based linkers.

TEG-diazide (74) was achieved by formation of the dimesylate intermediate via addition of triethylamine and dropwise addition of methanesulfonyl chloride under nitrogen. The molar excess of base and acid chloride allowed for quantitative conversion, confirmed by electron spray ionisation mass spectrometry (ESI MS). This intermediate was not isolated but directly subjected to substitution of the mesylates with sodium azide under reflux, yielding the diazido product quantitatively (ESI MS). The subsequent aqueous workup reduced the overall good yield to a moderate level because of the hydrophilicity of the compound and losses into the aqueous phase during workup (checked by MS). Infrared spectroscopy revealed two signals indicative for azides, at wavenumbers 2251 and 2108 cm⁻¹.

TEG-diazide (74) was then mono-reduced with triphenylphosphine in a biphasic system (EtOAc/1 M HCl). This procedure favours mono-reduction and provides N3-TEG-NH₂ (75) as the stable hydrochloride salt in reasonably good yields.

To achieve the acetylene counterpart for the 1,3-dipolar cycloaddition, TEG was stirred with sodium hydride with dropwise addition of propargyl bromide, which resulted predominantly in the mono-alkylated product (**76**). Unreacted starting material and minor amounts of TEG-diacetylene were easily removed by column chromatography.

The primary alcohol (**76**) was then converted to the iodide (**77**) with triphenylphosphine/imidazole/I₂. The triphenylphosphine oxide and imidazole were removed by silica chromatography.

The iodide (**77**) was converted to the ether (**78**) under standard Williamson ether conditions with 4-hydroxybenzaldehyde in excellent yields. The introduced aromatic aldehyde allows for reductive amination with amines, which are commonly present in natural products, without affecting protonation (c.f. amide formation) or basicity of the biomolecule. Initially, attempts were also made to introduce aliphatic aldehydes into the TEG linkers. These efforts were aborted, as the oxidation of primary alcohols via the Onodera procedure²⁴⁵ (activated DMSO/P₂O₅) or chromium based Jones reagent²⁴⁶ (CrO₃) resulted in very low yields, sluggish reaction times and hence difficulties during chromatographic workup. In contrast, the introduction of the aromatic aldehyde progressed smoothly and without side products and was therefore employed for bioconjugation reactions incorporating primary amines.

Overall, the linkers described above provide a set of terminally functionalised chains that allow derivatisation of amines, carboxylic acids and alcohols under mild conditions. The next step was to link the spacers to a solid support.

3.2.2 Biotinylated linkers

3.2.2.1 Introduction

Also known as Vitamin H or Coenzyme R, biotin is essential in the synthesis of fatty acids, branched-chain amino acid catabolism, and is involved in gluconeogenesis. As a cofactor it is responsible for carbon dioxide transfer in carboxylase enzymes.

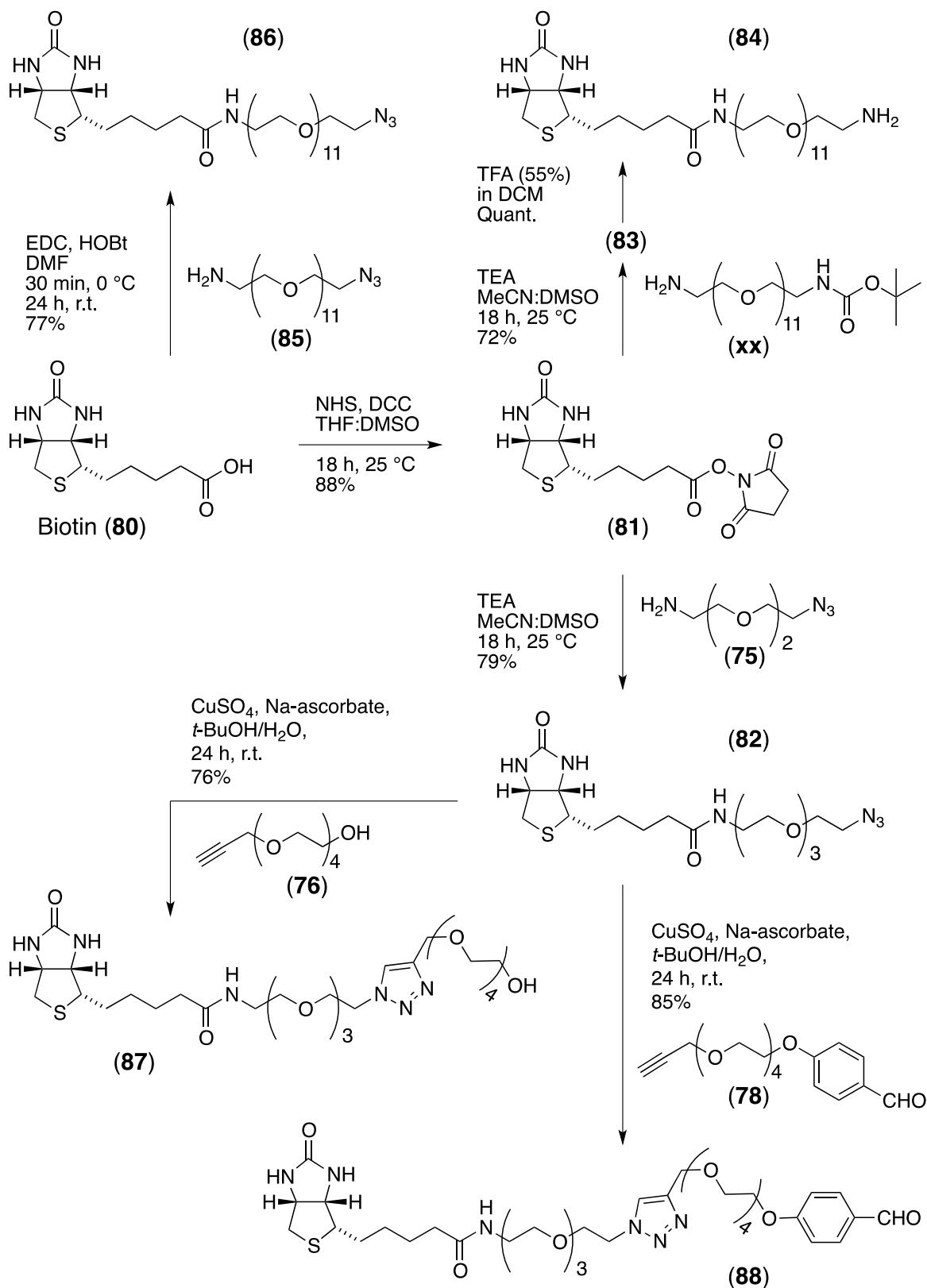
Interestingly, it also shows an extremely high affinity for avidin ($K_D = 0.4$ nM), a homotetrameric glycoprotein found in egg white. Inside the egg, the strong biotin-avidin interaction aims to capture all free biotin, which bacterial growth depends on, and thus shows antibacterial properties²⁴⁷.

Biotin consists of a bicyclic ring system and a carboxylic acid side chain. The ring system derives from a tetrahydroimidizalone ring fused to tetrahydrothiophene,

which carries a valeric acid side chain. In the biotin-binding pocket of avidin, only the bicyclic core of biotin is recognised, leaving the carboxyl acid side chain free for chemical derivatisation. It is for this fortuitous interaction that biotin has been exploited for a huge range of different bioanalytical applications such as ELISA-type assays and affinity isolations of many kinds²⁴⁸. For example, biotinylated small molecules can be immobilised on avidin-derivatised resins or microtitre plates to generate affinity probes for chemical proteomics.

3.2.2.2 Results and discussion

To provide various PEG-based biotinylated linkers ready for coupling to natural products, the compounds described in 3.2.1.2 and commercially available linkers (**75**, **79**) were coupled to biotin (**80**) under standard peptide coupling conditions (NHS ester (**81**), DCC; see Scheme 2). An alternative route, based on the direct coupling of biotin to the bifunctional amine-azide dodecaethyleneglycol (**85**) via 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC), resulted in the amide (**86**) in yields comparable to those achieved through DCC coupling conditions. The azide containing biotinylated linker (**82**) was subjected to Cu(I)-catalysed Huisgen 1,3-dipolar cycloaddition with acetylenes (**76** or **78**) forming the 1,4-substituted triazoles.



Scheme 2: Synthesis of fixed length, biotinylated, poly(ethyleneglycol) based linkers.

There is overwhelming evidence for the high regioselectivity of the described click chemistry conditions to favour the formation of 1,4-substituted triazoles over their 1,5-substituted isomers²⁴⁹. However, the synthesised triazoles were analysed by 2D

NMR (ROESY, HSQC, HMBC) experiments (see Figure 17 and Table 1) to confirm the linkage had been formed. As the ROESY and HMBC only showed correlations of H-30 to H-25 and H-24 (w), 1D selective ROE experiment had to be performed (see Figure 18) and confirmed ROE enhancement of H-25 and H-31 upon irradiation of H-30. This confirmed integrity of the linker and 1,4-substitution of the triazole.

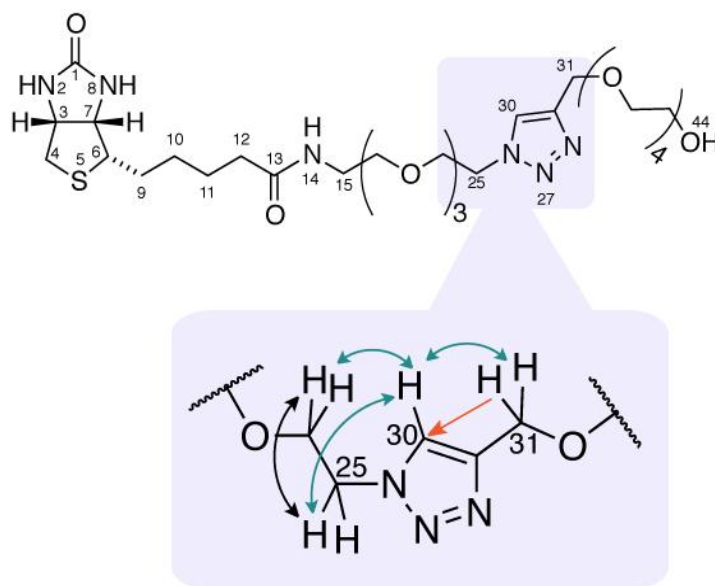


Figure 17: Selected 1D and 2D NMR correlations of the 1,4 disubstituted triazole **87**. Colours of arrows: black (COSY), orange (HMBC), blue (ROESY).

1D ROE 200 msec spin lock @ 8.056 ppm

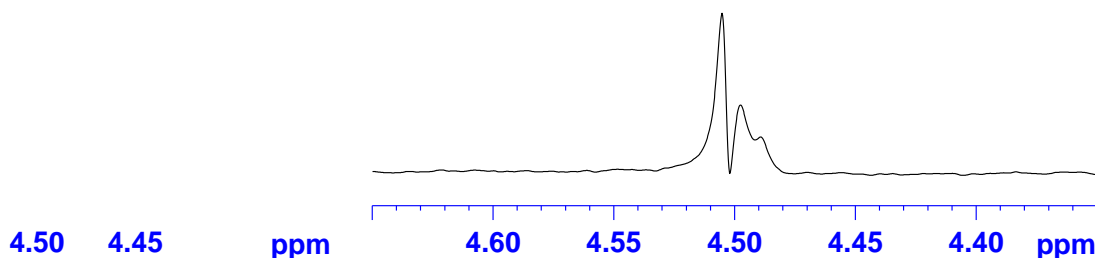


Figure 18: Selective 1D selective ROE of triazole (**87**) with spin lock on H-30 (8.06 ppm), showing ROE to both H-31 (s, 4.51) and H25 (t, 4.50).

These results are fundamentally important as the orientation and overall appearance of each linker is crucial for display of the natural product.

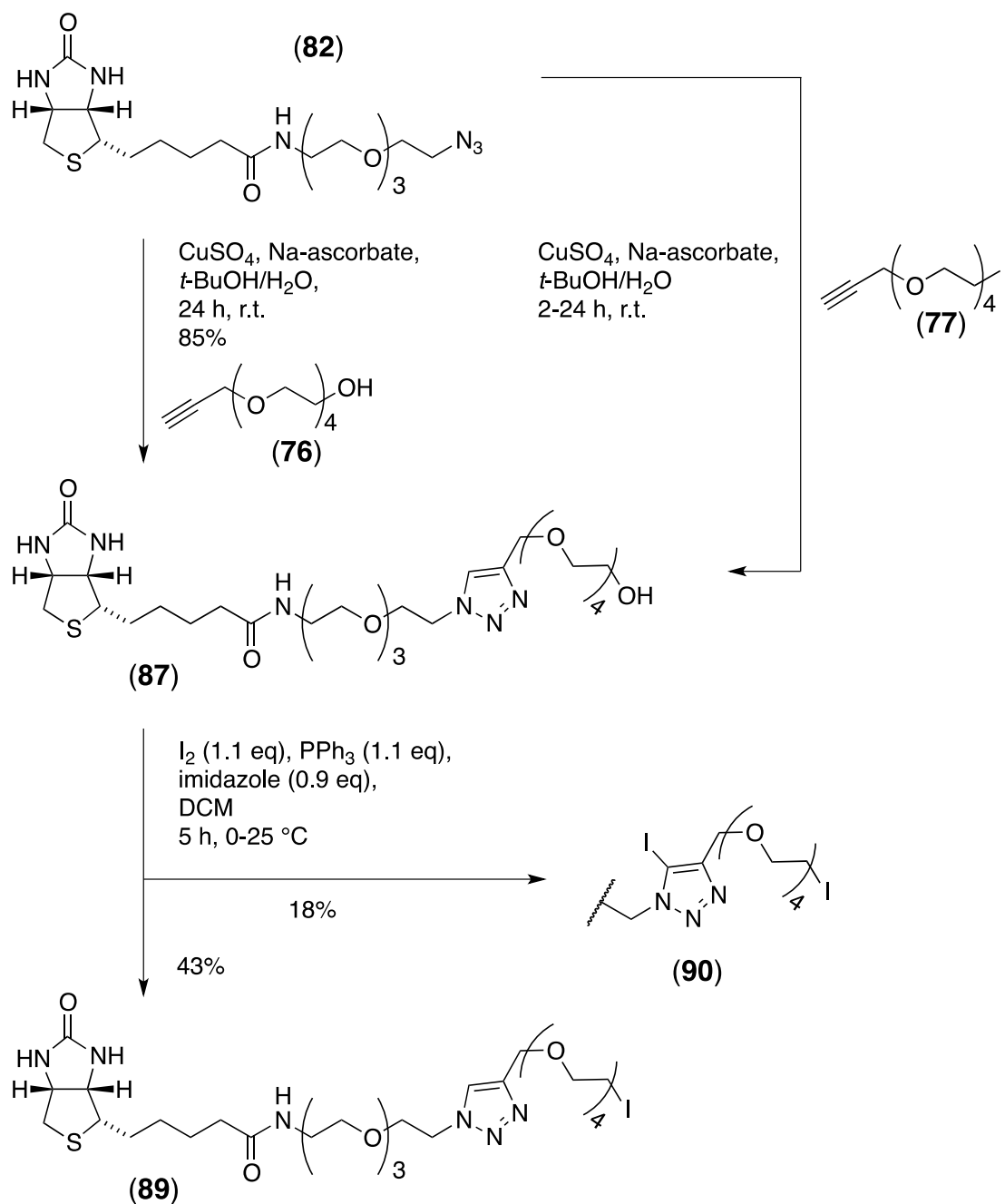
The resulting linkers provide fixed lengths of 33 Å (**87**) or 38 Å (**88**) respectively (see Scheme 2).

Table 13: NMR (DMSO-*d*₆, 600 MHz) data for compound (87).

Pos.	δ_c	δ_H , m (J in Hz)	COSY (H no.)	1H - ^{13}C HMBC (C no.)	ROESY (H no.)	sel 1D ROE (H no.)
1	162.7	-				
2	-	6.34, br s				
3	59.2	4.29, dd (7.7, 5.0)	4 β	1, 4, 6	4 β , 6(w), 7	
4	39.8	2.81, dd (β , 12.7, 4.8); 2.56, d (α , 12.6)	3	3, 6, 7	3	
6	55.4	3.08, m	7, 9	10	3(w), 4 β , 7, 9	
7	61.0	4.12, m	6	1, 4, 6	3, 4 β (w), 6	
8	-	6.40, br s		3		
9	28.1	1.60, m; 1.45, m	6, 10	7, 11	6, 11	
10	28.2	1.31, m	9, 11	6, 7, 11	9, 11	
11	25.3	1.49, m	10, 12	6, 11	12	
12	35.9	2.05, t (7.5)	11	6w, 10, 11	11	
13	174.6	-		10		
14	-	7.80, t (5.4)	15		15, 24(w)	
15	38.4	3.17, m	14, 16		14	
16-23	72.3 - 69.5	3.63 - 3.26		15		
24	68.8	3.80, t (5.3)	25		25, 30(w)	
25	49.3	4.50, t (5.7)	24, 30		24, 30	24
29	144.8	-				
30	123.9	8.03, s	25		24(w), 25	24, 25, 30
31	63.5	4.51, s		33		30
33-43	70.9 - 69.1	3.83 - 3.47				

(w) denotes a weak correlation.

Initially, attempts were also made to synthesise biotin-TEG- triazole-TEG-I (**89**) and to provide a biotinylated linker for immediate alkylation of natural compounds containing phenols. As expected, click chemistry of azide (**82**) and acetylene (**77**) went smoothly but the reaction conditions also resulted in Cu(I) catalysed hydrolysis of the alkyl-iodine²⁵⁰, so was not a feasible route. Conversion of the alcohol (**87**) to the iodide (PPh₃/I₂/imidazole) was attempted. Unfortunately, this resulted in iodination of the triazole as side product (**90**), a reaction that has some literature precedent²⁵¹.



Scheme 3: Synthesis of biotinylated linker incorporating a terminal halide.

This problem was easily by-passed, by simply reacting the propargyl iodide with the natural product of interest and doing the click chemistry last (see 3.2.3.1.2. for example).

3.2.3 Biotinylated natural products

3.2.3.1 Manzamine A

3.2.3.1.1 Introduction and review of SAR studies

The manzamine alkaloids are characterised by a unique polycyclic ring system consisting of 5, 6, 8 and 13 membered rings with 2 bridgeheads coupled to a β -carboline moiety. It is assumed that the complex ring structure is biosynthetically derived from three simple building blocks, i.e. ammonia, C₁₀ and C₃ polyketide units. The incorporated β -carboline structure arises from condensation of tryptamine (Pictet-Spengler reaction) with the polycyclic unit²⁵².

Manzamine A (**91**) was isolated as the first representative of this class of alkaloid from the Okinawan sponge *Haliclona* sp. in 1986²⁵³. Since then about 80 manzamine-related alkaloids have been reported from more than 16 species within 5 families of marine sponges, located anywhere from Indonesia to the Red Sea. The different alkaloids vary widely in structure and are classified either as β -carboline-containing or ircinal-related alkaloids.

Many representatives of both classes showed a variety of bioactivities ranging from antimalarial²⁵⁴, cytotoxic^{253,255}, antimicrobial²⁵⁶, pesticidal²⁵⁷⁻²⁵⁹ to anti-inflammatory²⁶⁰, but also effective against HIV and AIDS-related opportunistic pathogens^{261,262}.

Out of all the described activities, the highest potential for the manzamine alkaloids appears to be against malaria and neuroinflammation. The antimalarial potency of manzamine A (**91**) and 8-hydroxymanzamine A (**92**) are promising with IC₅₀ values in low nanomolar range *in vitro* against chloroquine-sensitive and resistant strains of *Plasmodium falciparum*. These two compounds outperformed the clinically used antimalarial drugs chloroquine and artemisinin both *in vitro* and *in vivo*^{254,263,264}. Interestingly, both compounds also exhibited the ability to inhibit human glycogen synthase kinase 3 β (GSK-3 β)-dependent *tau* phosphorylation while simultaneously displaying an excellent cytotoxicity profile when tested against mammalian kidney fibroblasts²⁶⁵. GSK-3 is an omnipresent serine-threonine kinase and involved in the regulation of many cell functions, including signalling by insulin, growth factors and nutrients, control of cell division, apoptosis and microtubule function²⁶⁶. GSK-3 dependent hyperphosphorylation of microtubule associated protein tau is an early event in neurodegenerative diseases, such as Alzheimer's disease. Given the

inhibitory capacity of several manzamines, a potential application as drug leads in the treatment of neurodegenerative pathologies was suggested^{265,267}. Although the K_d values of around 5 μ M are significant, they are clearly outperformed by those of ATP-competitive GSK-3 inhibitors, such as the arylindolemaleimide SB-216763 and anilino maleimide SB-415286, displaying competitive inhibitory constants in a low nanomolar range²⁶⁸. Given the high potential of the manzamine alkaloids, such reasonably weak binding suggests other cellular targets.

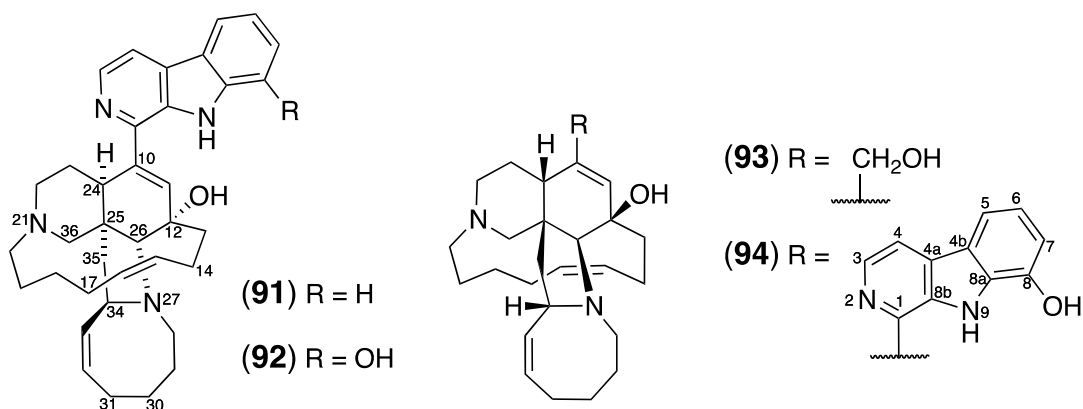


Figure 19: Structure of the selected manzamine alkaloids: manzamine A (91), 8-Hydroxymanzamine A (92), ircinol A (93), and *ent*-8-hydroxymanzamine A (94).

In addition to very low cytotoxicity against non-pathological cells, potent anticancer activity was reported for many different manzamine alkaloids, with manzamine A exhibiting the most pronounced bioactivity²⁶⁹. It affected human colon tumour cells, lung carcinoma cells and breast cancer cells at concentrations of around 0.5 μ g/mL²⁷⁰. Additionally, significant *in vitro* activity was found against KB, LoVo and HSV cell lines with IC_{50} values varying between 50-150 ng/mL²⁷¹. Clearly, manzamine A displays a wide range of biological activities and is potentially a highly interesting lead for drug development based on various cellular targets.

The modes of action underlying the different bioactivities are not yet known and further research must aim to evaluate suspected targets as well as to describe the cellular receptors for each of the reported bioactivities. The previously described phage display technology (see Section 1.1.2.2.1.1) offers a valuable tool to search for the most avid protein binding partners. To apply the phage display technique, biotinylated manzamine is required. However, if one aims to derivatise a bioactive molecule, one has to be careful not to affect the bioactivity.

In the case of the diversely bioactive manzamine alkaloids this decision can partly be based on published structure activity relationship (SAR) studies discussed hereafter.

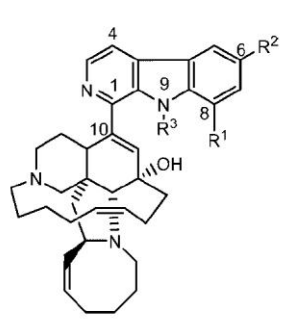
Due to the low abundance and stepwise discovery of many of these alkaloids, not all manzamines have been included in each SAR study.

To maintain the antimalarial activity of the manzamine alkaloids, β -carboline-based structures are required, for instance shown by the lack of antimalarial activity of ircinol A (**93**) compared to *ent*-8-hydroxymanzamine A (**94**)²⁷². Different functional groups in positions C-6, C-8 or N-9 lead to either increased or decreased activity, displaying the relevance of the carboline moiety for the activity. Studies on N-9 derivatised β -carbolines proved the aliphatic, polycyclic structure of manzamines were essential as well, as no activity was observed in structures missing the polycyclic ring system²⁷³. Particularly the 8-membered ring is required, shown by decreased activity due to changes of either the C-32-C-33 olefin (oxidation, reduction, Grubbs metathesis ring opening), oxidation of C-31, or reduction of the N-27-C-34 bond to give the ring-opened product^{272,274}.

When investigating likely pharmacophores relevant for the antimicrobial effects, manzamine alkaloids were tested against AIDS-opportunistic pathogens like methicillin-sensitive and resistant *Staphylococcus aureus*, *Mycobacterium tuberculosis* and *M. intracellulare*, and *Cryptococcus neoformans*. The published SAR studies provide similar results as for the antimalarial activity, outlining the importance of the β -carboline-based structure including the 8-membered ring as present in the natural product manzamine A. Surprisingly, the antileishmanial activity does not depend on a β -carboline structure - displayed by the high activity of ircinol-related compounds against *Leishmania donovani*, but again was impacted by structural changes of the 8-membered ring²⁷⁵. This suggests an unrelated target site and mode of action compared to those for *Plasmodium*.

Furthermore, the GSK-3 inhibitory activity of manzamine alkaloids depends on the presence of both the aliphatic ring system and β -carboline substructure, as confirmed by lack of activity for both individual building blocks, ircinal A and carboline.

Interestingly, the replacement of hydrogen at the carboline 6 or 8 position with OH, OMe, OEt or OTs groups is well tolerated by the GSK-3 enzyme. These analogues provided equipotent inhibition of GSK-3 (compare Table 14) suggesting that this ring is solvent exposed in the bound conformation. However, substitution of N-9 of the carboline moiety, particularly with larger groups, resulted in decreased GSK-3 inhibition or complete inactivation²⁶⁵.

Table 14: GSK-3 Inhibition by Manzamine A and Analogues^a taken from Hamann et al.²⁶⁵


compounds	R ¹	R ²	R ³	GSK-3 β percent inhibition at 25 μ M	GSK-3 β IC ₅₀ (μ M)
manzamine A	H	H	H	73.2	10.2
1	OH	H	H	86.7	4.8
2	H	OH	H	74.3	16.6
3	OTs	H	H	80.4	
4	OMe	H	Me	72.4	nd
5	OEt	H	Et	78.0	10.4
6	O- <i>i</i> -But	H	<i>i</i> -But	24.2	nd
7	H	H	(CH ₂) ₁₁ CH ₃	0	nd
8	H	H	<i>t</i> -BuOCOMe	3	nd

^a nd = not determined.

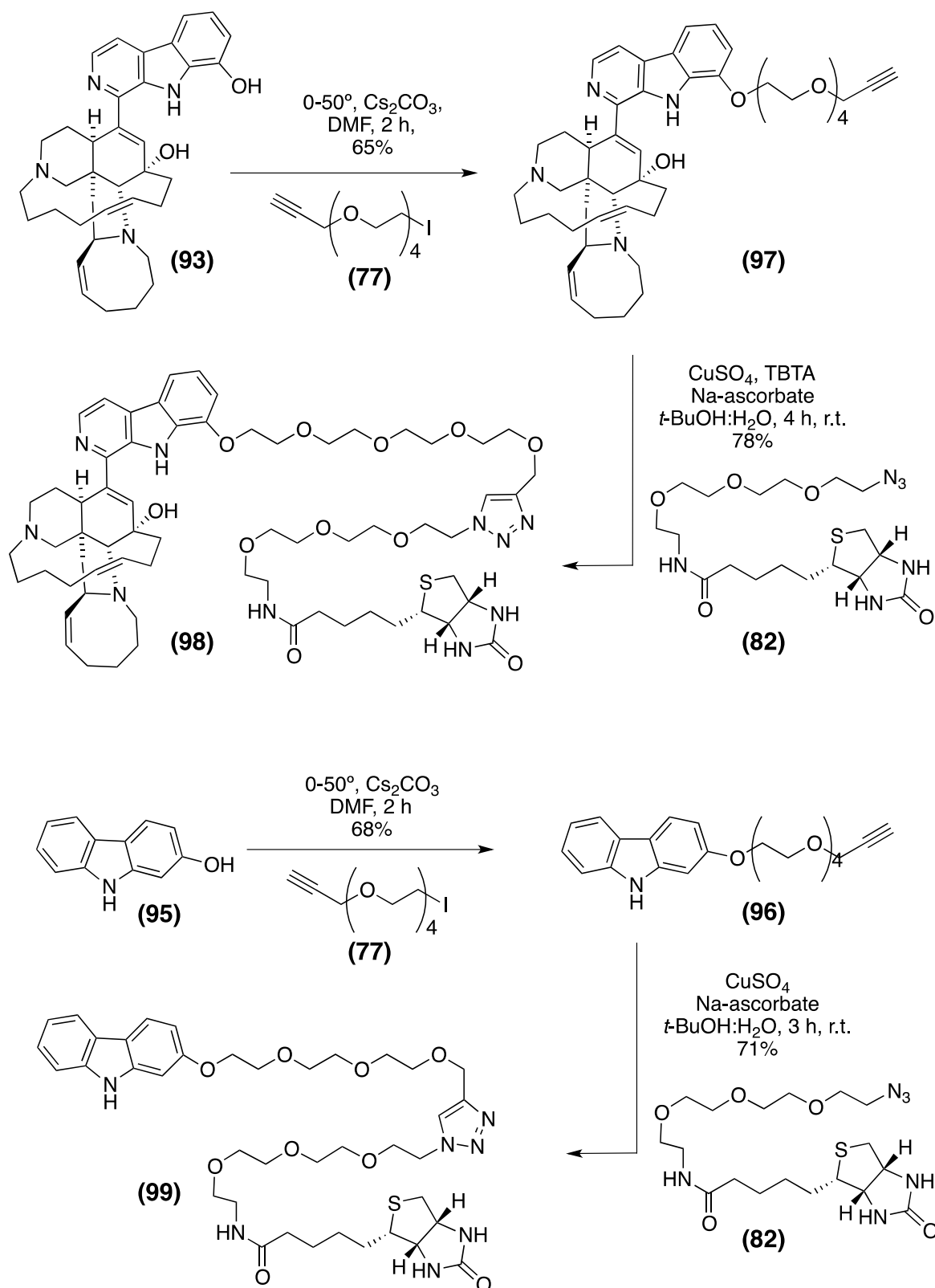
So far, there is only a limited number of SAR studies published focusing on the anticancer activities of the manzamines^{269,275,276}. Carboline- and ircinal-related compounds, both oxidised and reduced derivatives, and enantiomers displayed cytotoxic activity against various human cancer cells. Hence, it is at this point not feasible to determine the structural features most relevant for the anticancer effects. According to the previously mentioned structure-activity relationship studies, the kinase inhibition, the antiprotozoal and antimicrobial activities appear to depend predominantly on the structure of manzamine A's 8-membered ring. Although the β -carboline moiety is required, alkylation of the 8-hydroxy group was well tolerated in biological assays.

In regards of the biotinylation, which is essential for chemical proteomics, it appeared that attachment of the PEG-based biotinylated linker at the phenol of 8-hydroxylmanzamine (**92**) would be the most feasible route (see Scheme 4).

3.2.3.1.2 Results and discussion

3.2.3.1.2.1 Derivatisation

The phenol in 8-hydroxymanzamine (**92**) is the most acidic functionality in the molecule. However, initial attempts from within the research group to alkylate 8-OH-manzamine A with Na_2CO_3 or K_2CO_3 disappointingly lead to alkylation of the phenol and pyrrole, and even dialkylation. It is well known that the use of carbonates with larger alkali metal cations such as rubidium or cesium often provides advantages in terms of obtainable yields or rates²⁷⁷. As early as in preparing Merrifield's resin²³⁹, the authors for example observed that esterification proceeded more readily for carboxylates in conjunction with larger alkali cations. With DMF as solvent, the use of cesium as cation yielded up to tenfold reaction rates compared to lithium. These results provided good support for the presumption that carboxylates with bigger counter cations are far more dissociated in aprotic solvents and hence more reactive. To validate different alkali metals as suitable cations for the alkylation of the β -carboline moiety, lithium, potassium, sodium and cesium carbonates were used in 0.5, 1 or 1.5 equivalents with 2-OH-carbazole (**95**) and monitored by TLC and ESI-LC-MS. This compound is structurally similar to the β -carboline and was used as a model system to optimise the alkylating conditions. Additionally, 2-OH-carbazole could be directly utilised as the control probe for the phage display biopanning experiments. Over 24 hours, Li_2CO_3 did not result in any alkylated product, whereas sodium and potassium carbonates provided mixtures of mono- and dialkylated products. According to the retention time in reverse phase LC-MS, only one type of monoalkylated product was formed and assumed to be the O-alkylated one. Even with 1.5 equivalents, none of the latter two carbonates pushed the reaction to completion. However, cesium carbonate gave rise to only one product, which was isolated and verified by 2D-NMR experiments as O-alkylated product (**96**; see Scheme 4).



Scheme 4: Synthesis of biotin-manzamine (98) affinity probe and biotin-carbazole (99) control.

The reaction reached its maximum yield after 30 min at room temperature (monitored by TLC) and could be pushed to complete consumption of the iodide (monitored by TLC) by increasing the temperature to 50 °C. Pyrrole usually requires

a strong base to be deprotonated, however, transalkylation could result in the formation of N-alkylated product.

For 8-OH-manzamine A, employing half an equivalent of Cs_2CO_3 at 0 °C and slowly allowing the reaction temperature to rise to 50 °C afforded predominantly the O-alkylated product (**98**) and a minor amount of the dialkylated product (confirmed by ESI-MS, not further characterised), which were separated by RP-HPLC. The reaction did not reach completion (monitored by HPLC trace), but prolonged reaction time favoured the formation of dialkylated product and hence the reaction time was kept to a total of 2 h and the starting material recycled (HPLC).

The HPLC-purified acetylene (**97**) was then subjected to click chemistry with the biotinylated azide (**82**). *tert*-Butyl alcohol/water as solvents were favoured over for example THF/water or ethanol/water mixtures due to faster reaction times. Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA) was added as copper(I) stabilising ligand²⁷⁸ to enhance the copper(I) reactivity, increase solubility, prevent disproportionation to Cu(0) and Cu(II) and to prevent the copper binding to the TEG. The reaction was monitored by ESI-LC-MS and purified by HPLC to isolate biotin-manzamine (**98**) in good yields.

To validate the site of alkylation for biotin-manzamine, 2D-NMR experiments were performed. The most important correlations (see Table 15) found during HSQC, HMBC and ROESY experiments are displayed in Figure 20. Most importantly, the site of alkylation was confirmed by an ROE cross peak between H-7 and H-38 (see Figure 21a) and HMBC correlations from both H-50 and H-56 to C-55 (see Figure 21b) as evidence for the 1,4-triazole and that the linkage between manzamine and biotin was as expected.

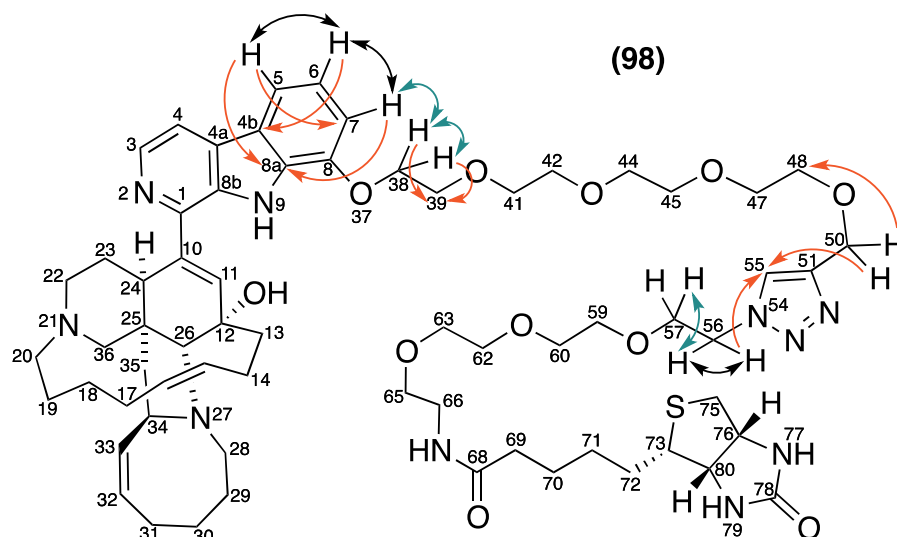


Figure 20: Structure of biotin-manzamine (**98**) and selected 2D-NMR correlations. Key correlations confirming the site of alkylation and stereochemistry of the triazole are shown. HMBC in red, ROESY in blue and COSY in black.

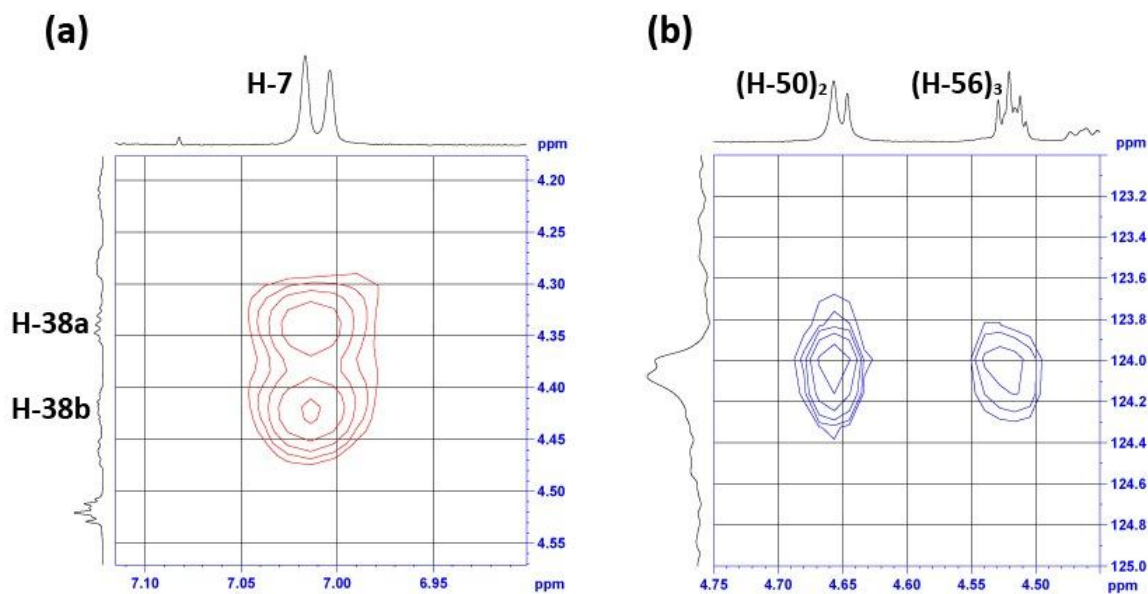


Figure 21: Selected 2D-NMR correlations of biotin-manzamine (**98**; CDCl_3 , 600 MHz). (a) ROE cross peak between H-7 and (H-38)₂. (b) HMBC correlations of C-55 to both (H-50)₂ and (H-56)₂.

All proton and carbon shifts and relevant correlations for biotin-manzamine (**98**) were assigned and are summarised in Table 15.

Table 15: Selected NMR (CDCl₃, 600 MHz) data for biotin-manzamine (98).

Pos.	δ_C	δ_H , m (J in Hz)	COSY (H no.)	1H - ^{13}C HMBC (C no.)	ROESY (H no.)
1	143.1	-			
3	137.6	8.34 d (5.1)	4	5	
4	114.1	7.83 d (5.1)	3	4b, 8b	
4a	130.1	-			
4b	122.8	-			
5	113.4	7.66 d (7.9)	6	7, 8a	
6	120.3	7.15 t (7.8)	5, 7	4b	
7	109.6	7.01 d (7.7)	6	8a	38
8	146.2	-			
8a	132.5	-			
9	-	11.07 s			
9a	133.3	-			
10	141.7	-			
11	135.5	6.70 m		1, 23, 26	
12	71.1	-			
13	39.1	2.10, 1.73	14	12	13
14	20.7	2.23 m	13, 15	16	
15	127.0	5.58 m	14, 16		
16	133.0	5.55 m	15, 17		
17	25.0	2.49, 1.62 m	16, 18		
18	26.3	1.51, 1.21 m	17, 19		
19	24.5	1.85, 1.50 m	18, 20		19
20	53.5	2.63, 2.43 m	19		
22	49.2	2.94, 1.93 m	23		
23	33.8	2.96, 1.74 m	22, 24		
24	40.5	2.69 m	23		
25	47.0	-			
26	77.7	3.70 m		12, 13, 25, 35, 36	
28	53.2	4.06, 3.24 m			28, 29
29	26.5	2.55, 2.03 m	30		28
30	24.5	2.01, 1.47 m		31	
31	28.3	2.38, 2.29 m			32
32	142.3	6.31 m	33		
33	124.1	5.45 t (9.8)	32	31	
34	57.0	4.92 m		25, 26, 32	
35	44.6	2.48, 1.92 m			
36	70.5	2.96, 2.43 m		12(w), 26	
38	68.6	4.42, 4.35 m	39		7
39	69.6	4.21, 4.07 m	38	41 - 48	
40 - 48	70.9 - 69.3	3.93 - 3.51			
50	64.6	4.66 s		48, 51, 55	
51	144.8	-			
55	124.0	7.78 s		51(w), 56	

(Table 14 continued)

Pos.	δ_c	δ_H , m (J in Hz)	COSY (H no.)	1H - ^{13}C HMBC (C no.)	ROESY (H no.)
56	50.1	4.52 t (5.1)	57	55	
57	69.2	3.86 t (5.1)	56	59 - 63	
59 - 63	70.9 - 69.3	3.93 - 3.51			
65	69.6	3.50 m	66		
66	39.2	3.42 m	65		
67	-	7.79 b	66		
68	172.9	-			
69	35.7	2.17 m	70	68(w)	
70	25.4	1.64 m	69, 71		
71	27.9	1.70 m, 1.60 m	72	70	
72	27.8	1.41 m	70		
73	55.1	3.11 m	72, 80		
75	40.5	2.87, 2.71 m	76		
76	60.1	4.47 m	75	78(w)	
78	163.2	-			
80	61.8	4.29 m	73	78(w)	

a (w) denotes a weak correlation.

3.2.3.1.2.2 Purity and stability of biotin-manzamine (98)

For phage biopanning as with biological testing, it is important to have the analyte absolutely pure to exclude the possibility of impurities contributing to the observed effect. Preparative separations were checked by analytical HPLC (see Figure 22).

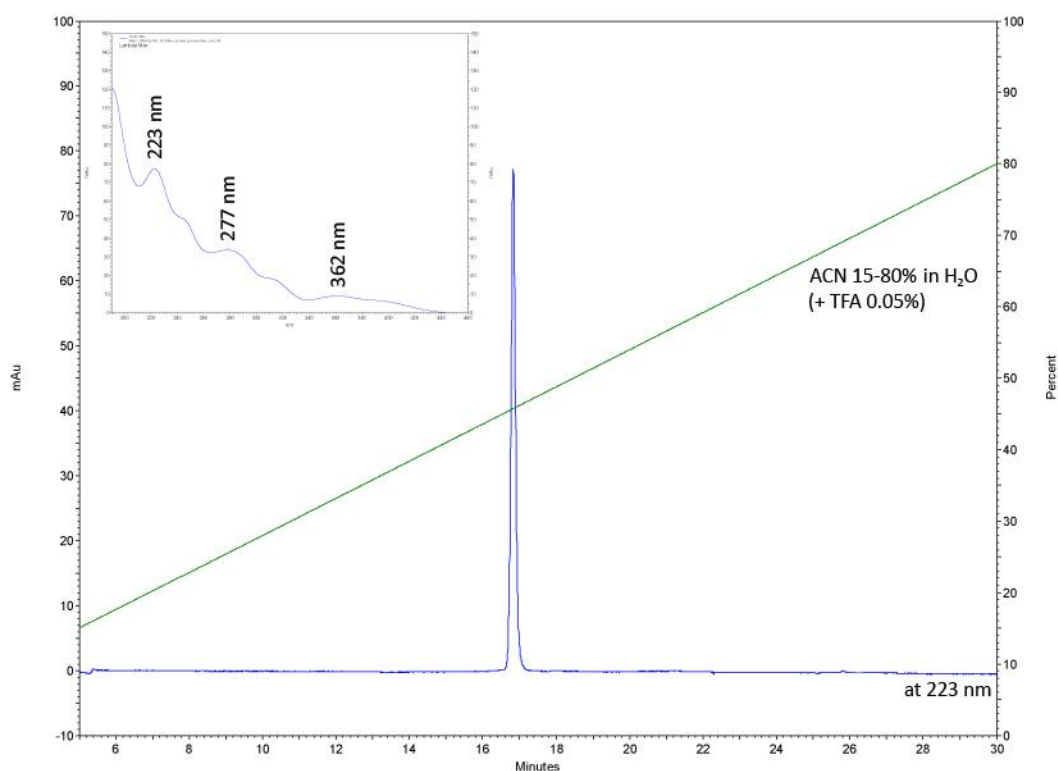


Figure 22: RP-HPLC chromatogram of biotin-manzamine (98; 0.5 mg/mL) at 223 nm and the compounds UV spectrum indicating the absorbance maxima at 223 nm, 277 nm and 362 nm.

Throughout the biopanning process, the biotinylated probe is also required to be stable for a minimum of 6 h (i.e. 2 h for coating the streptavidin derivatised microtitre plates and 4 h during the phage selection procedure). Consequently, a solution of biotin-manzamine stored in PBS (0.5 mg/mL) was analysed by RP-HPLC (see Figure 23), showing that no degradation occurs over the tested time period.

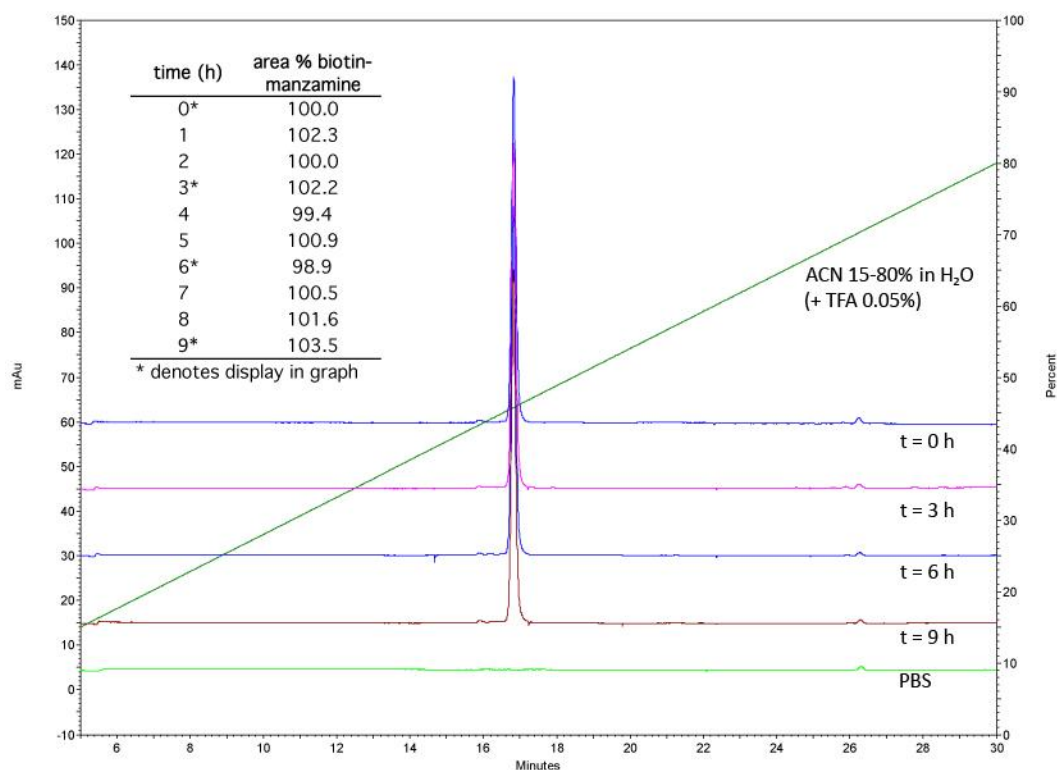


Figure 23: RP-HPLC traces of biotin-manzamine (98) in PBS (0.5 mg/mL) detected at 223nm after 0 h, 3 h, 6 h and 9 h relative to time zero (100%).

Without the use of an internal standard, slight variation in integrals is to be expected due to instrumental inaccuracy.

3.2.3.2 Daptomycin

3.2.3.2.1 Introduction

Daptomycin (**100**) is a branched cyclic, 13-residue anionic depsipeptide antibiotic originally isolated from the soil actinobacterium *Streptomyces roseosporus* as a member of the antibiotic complex, named A21978C²⁷⁹. It constitutes the only member of a new class of antibiotics, namely cyclic lipopeptides to have reached the market. It was first approved by the Food and Drug Administration (FDA) for non-topical use in 2003 for the treatment of skin and skin structure infections caused by Gram-positive pathogens, and in 2006 for non-topical treatment of bacteremia and right-side endocarditis caused by *Staphylococcus aureus* strains and MRSA²⁸⁰.

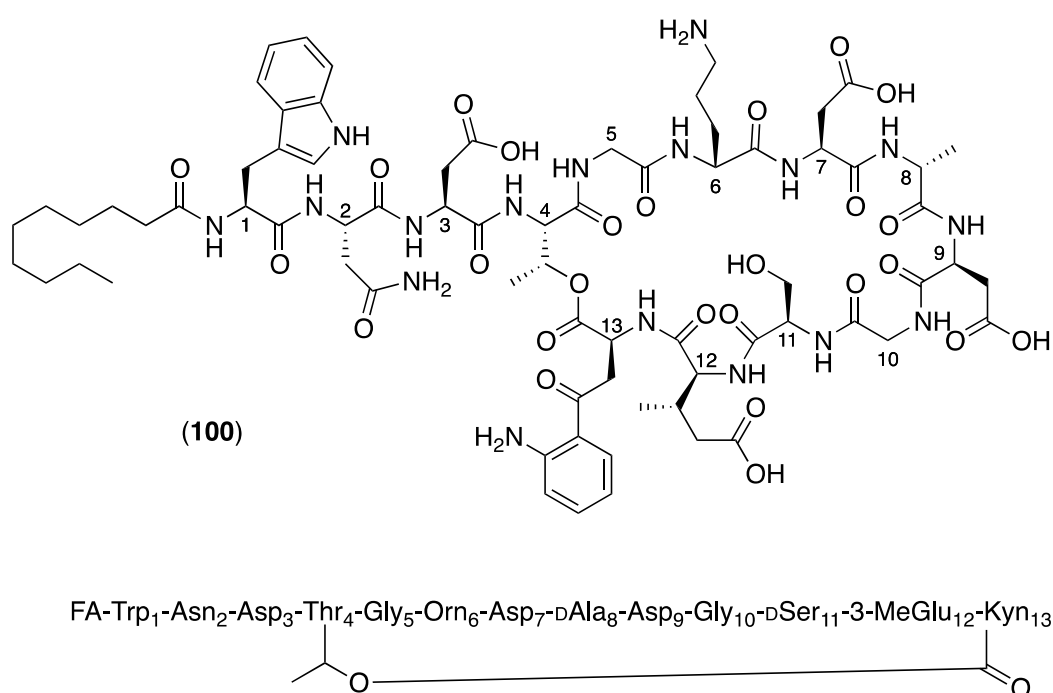


Figure 24: Structure and amino acid sequence of daptomycin (**100**)

Its mode of action (MOA) as an antibiotic is yet to be completely elucidated.

3.2.3.2.1.1 Mode of action as antibiotic

Current opinion proposes a multi-step mechanism of insertion into the bacterial membrane and membrane disrupting effects targeting a two-component system relevant for cell membrane biosynthesis^{280,281}. Yet, this explanation is still disputed and hence, along with other hypotheses on the MOA, reviewed below. Calcium ions are essential for the antimicrobial activity. However, the following models for the MOA for the antibiotic activity have been proposed:

(a) daptomycin inhibits lipoteichoic acid (LTA) biosynthesis in the presence of calcium ions²⁸²⁻²⁸⁴. LTA is a bacterial cell surface macromolecule anchored to cytoplasmic membrane, and extends into the peptidoglycan cell wall layer.

(b) Daptomycin dissipates the membrane potential causing cell death²⁸⁵. Therefore the decanoyl side chain of calcium-free daptomycin ("apo-daptomycin") weakly binds to the cytoplasmic membrane, followed by a calcium chelating step that causes deeper penetration into the membrane and aggregation of daptomycin which creates channels allowing potassium ions to permeate. This leads to membrane depolarisation and ultimately cell death²⁸⁶.

(c) Based on observations from NMR experiments, circular dichroism (CD) measurements and fluorescence spectroscopy, it was suggested that Ca^{2+} initially binds to daptomycin in solution and induces a conformational change, reducing its charge and increasing its amphipathicity²⁸⁷. This process facilitates oligomerisation and leads to micelle formation, which allows for interaction of daptomycin with acidic or neutral membranes. In a following step, Ca^{2+} bridges the gap between daptomycin and the acidic phospholipids. CD measurements indicated a second conformational change, allowing deeper insertion of daptomycin into the membrane layer.

(d) NMR studies in phosphatidylcholine (PC) micelles showed that daptomycin did not undergo major conformational changes prior to membrane insertion, but was inserted as its apo-structure²⁸⁸. Once it bound calcium, daptomycin experienced only a minor conformational rearrangement upon binding to PC. The MOA was therefore revised as follows: apo-daptomycin binds to calcium in a 1:1 ratio, upon which micelle formation of 14-16 daptomycin units occurs²⁸⁹. These micelles approach the bacterial membrane and dissociate in close proximity of the lipid bilayer. Calcium-bound daptomycin subsequently inserts into the membrane and may oligomerise. Pores are formed within the bilayer and the resulting potassium efflux causes depolarisation of the membrane. Membrane-associated processes are disturbed, which ultimately leads to cell death.

(e) Most recent studies suggest, however, that the MOA is more complex. Jones *et al.*²⁹⁰ reported no membrane depolarisation in clinical isolates of *S. aureus* that developed reduced susceptibility to daptomycin over the course of treatment of recalcitrant endocarditis. These strains showed elevated lysyl-phosphatidylglycerol (LPG) in the membrane outer leaflet. LPG imparts increased positive charge and, if

present in the membrane, repels daptomycin penetration by less favourable electrostatic interaction. Hobbs *et al.*²⁹¹ demonstrated that membrane depolarisation and potassium efflux did not cause cell death, but only occurred after cell death of *S. aureus*. Transcriptome studies suggested an additional mechanism, targeting inhibition of cell wall biosynthesis through the two-component regulatory system YycFG²⁹². This contains YycG, a membrane-spanning sensor/histidine kinase, and YycF, its response regulator. YycFG is required for viability and functions as a master regulatory system for cell wall metabolism. Muraih *et al.*²⁹³ confirmed daptomycin oligomerisation on liposome membranes via fluorescence resonance energy transfer at concentrations similar to those required for antimicrobial activity. They detected oligomers of fluorescently labeled daptomycin on membranes containing the negatively charged phosphatidylglycerol (PG), combined with neutral PC or phosphatidylethanolamine plus cardiolipin, as well as on membrane vesicles prepared from *Bacillus subtilis* cells. No oligomerisation occurred on neutral, PC-exclusive membranes or in solution.

These data limit the accuracy to which the MOA of daptomycin may be described and suggest the current opinion only on a general level: The lipid tail and the Ca²⁺ ions bound to the acidic peptide help direct daptomycin into the cell division septum rich in PG. It disrupts membrane functions, potentially by oligomerisation and formation of pores allowing for membrane depolarisation. In addition, daptomycin binds to YycG, disrupting the function of this key sensor kinase and leading to cell death without lysis.

Although daptomycin has been approved for more than 8 years, fundamental aspects of its antimicrobial MOA are still ambiguous. It is unclear if there is one dominant cause of cell death and if so, which one it is, or, in what order these steps occur. In addition, no human binding proteins have been proposed to explain its side effects.

3.2.3.2.1.2 Impact on human cells and body

Despite the research on daptomycin's effect on bacterial cells, very little is reported about its interaction with human cells.

As early as 1990, Canepari *et al.*²⁸⁴ exposed human epithelial cell membranes (HEp-2 cultures) to radioactively labeled daptomycin and observed binding to occur.

Similarly to the antimicrobial MOA, binding towards cell membrane predominantly occurred in the presence of calcium. According to this study, daptomycin did not

enter the cytoplasm. In addition, bound antibiotic was almost completely removed by washing with EDTA (Ca^{2+} binding). Based on those findings, it was concluded that no specific interaction between human cell membranes and daptomycin occurs.

Surprisingly, there is hardly any more literature on daptomycin's interactions with human cell surface proteins or transporter systems. From the molecule's amphipathic character, it seems reasonable to believe that the lipophilic tail anchors the molecule similarly into the human phosphatidylcholine lipid bilayer, as it does to bacterial membranes. It was furthermore observed in phospholipid bilayer studies that daptomycin induces, in a Ca^{2+} dependent manner, substantial lipid flip-flop²⁸⁷. This phenomenon is potentially relevant for daptomycin's entry into human cells but what it interacts with in the lipid bilayer or the cytoplasm is unknown.

The most common side effects of a treatment with this antibiotic include non-specific symptoms like nausea, headache, diarrhea, vomiting²⁹⁴. Severe adverse effects like myalgia, showing skeletal muscle toxicity, have also been reported in up to 40% of the patients²⁹⁵. As early as in Phase I clinical trials muscle toxicity was recognised²⁹⁶, while administering 4 mg/kg every twelve hours.

Cubist Pharmaceuticals then initiated an animal study on dogs to guide clinical dosing regimens. The data suggested that skeletal muscle effects were more closely related to the dosing interval than to either the maximum concentration of the drug in plasma or the area under the concentration-time curve. Administering daptomycin once daily seemed to minimise effects on skeletal muscle, possibly by allowing more time between doses for repair of subclinical effects. Noticeably, once daily administration also resulted in less increase of serum creatinine phosphokinase (CPK) compared to equal quantities over 24 h if administered in portions either every 8 h or 12 h.

In addition, a recent toxicity study on rat muscle cells demonstrated that daptomycin has a concentration-dependent and time-dependent effect on the plasma membrane of primary cultures of differentiated, spontaneously contracting rat myotubes²⁹⁷. The membrane damaging effects only occur at daptomycin concentrations above 1 mg/mL, affecting up to 20-30% of the cells at a concentration of 2 mg/mL. On the other hand, no membrane damaging effects were evident in non-differentiated myoblasts or other mono-nucleated cells present in the cultures, even at the highest test concentration (6 mg/mL). In addition, myotubes lost their ability to contract

either transiently (at daptomycin concentration above 750 $\mu\text{g/mL}$) or permanently (above 2 mg/mL). The authors concluded that putative targets for daptomycin effects on skeletal muscle are structures present only on the surface of the plasma membrane of highly differentiated myotubes, but this hypothesis is still under dispute.

Our work focuses on determining cellular targets from within human cells for two main reasons. First, we aim to help elucidate the mode of action underlying the observed side effects. In addition, many bioactive small molecules possess extensive polypharmacology, often across target boundaries²⁹⁸. Therefore, the search for the most avid human protein binding partner may reveal unknown potential drug targets for this unique compound.

3.2.3.2.1.3 *Physicochemical properties*

Daptomycin carries a predominantly acidic structure and its negative charge at neutral pH explains its high aqueous solubility. However, the lipid tail and the presence of some hydrophobic amino acid residues, lead to overall amphipathic behaviour so we would expect that this molecule would be attracted to positively charged species with hydrophobic binding areas such as phospholipid bilayers. This characteristic is crucial for its antibiotic activity²⁹⁹.

Like other acidic lipopeptide antibiotics, its antibacterial activity is Ca^{2+} -dependant. Its potency reaches its maximum in the presence of a calcium concentration of around 50 mg L^{-1} (1.25 mM)³⁰⁰. This equates to the concentration of ionized Ca^{2+} normally found in human serum. In our experiments, performed in M9TB, Mg^{2+} is added (1 mM) and calcium is present in the tryptone (0.019%), which equates to a final concentration of about 45 μM . Daptomycin also self-associates in aqueous solutions to form micelles. These aggregation effects are concentration dependent and complicate structural studies. For example NMR experiments prove challenging due to line broadening^{287,288,301-303}. Conformational changes upon Ca^{2+} binding were proposed, but none of the authors indicated at which position this might occur^{303,304}. Conversely, the conformation of calcium free daptomycin ("apo-daptomycin") in solution has been described derived from simulated annealing molecular dynamics using NMR derived Nuclear Overhauser Effect (NOE) constraints, yielding a distorted hairpin formed by Gly₅-DAla₆ type II' β -turn³⁰².

3.2.3.2.1.4 Pharmacodynamics

Daptomycin is administered as once-daily single dose intravenous infusion of 4 mg/kg body weight over 7-14 days. It is reversibly bound to human plasma proteins, primarily to serum albumin, in a concentration-independent manner. The overall mean binding ranges from 90 to 93%. The site of metabolism has not been identified. Minor amounts of three oxidised metabolites and one unidentified compound have been detected in urine. Primarily the kidneys excrete daptomycin. In a mass balance study³⁰⁵ of 5 healthy subjects using radiolabeled daptomycin, approximately 78% of the administered dose was recovered from urine based on total radioactivity (approximately 52% of the dose based on microbiologically active concentrations) and 5.7% of the dose was recovered from faeces (collected for up to 9 days) based on total radioactivity. Because renal excretion is the primary route of elimination, dosage adjustment is necessary in patients with severe renal insufficiency (CLCR <30 mL/min).

3.2.3.2.1.5 Chemical modifications and SAR studies

For daptomycin, several studies give clear indication of which sites are suitable for derivatisation without significant loss of activity and which must not to be modified. Combinatorial biosynthesis³⁰⁶ and chemoenzymatic investigations²⁸⁹ have proven particular residues to be crucial for antibiotic activity.

For instance, the β -methylglutamate (MeGlu₁₂) is a common feature of daptomycin and other acidic lipopeptide antibiotics such as Calcium Dependant Antibiotic (CDA) and A54145 and is in all three cases situated at the same relative position in the peptide ring. If this amino acid is substituted with non-methylated glutamate, the minimum inhibitory concentration (MIC) increases 7-fold. In addition, the exchange of Kyn₁₃ with Trp also increases the MIC, suggesting the need for a less rigid aromatic residue carrying more polar functionalities. Substitution of the D-serine in position 11 seems to have little effect on the antibacterial activity. Asp₇ and Asp₉ are proposed to be part of a specific EF-hand Ca²⁺-binding motif. Considering that Ca²⁺-binding is essential for the activity, it is not surprising that substitution of either of these acidic residues with Asn results in complete loss of activity.

Orn₆ has been derivatised to yield less, equally or more active compounds. Amidation of the γ -amino group results a decrease of activity. When performing synthetic SAR studies on Orn₆, it became clear that the γ -amino group is not essential for whole cell

activity, but derivatisation with various functional groups impacted the antibiotic activity³⁰⁷. The authors determined that a non-amide amine appears to be required for activity, the introduction of an electron-deficient aryl ring via reductive amination and the use of heterocyclic spacers or more polar functionalities helped maintain drug-like properties. Furthermore, the overall charge of the molecule (derived from carboxylic acids and the primary amine) was not affected (as opposed to by introducing an amide).

Consequently, arylation of the ornithine via reductive amination is a reasonable choice for derivatising daptomycin without destroying activity.

3.2.3.2.1.6 Maintaining antibiotic activity

After derivatising daptomycin it is important to verify the remaining antibiotic potential. According to the recommendations set by the National Committee for Clinical Laboratory Standards (NCCLS), the MICs are determined through a MTT broth microdilution assay under temperature-adapted conditions³⁰⁸.

3.2.3.2.1.7 Stability and purity

Daptomycin is susceptible to both alkaline and acidic degradation, giving rise to three major degradation products³⁰⁹. In alkaline conditions, ester hydrolysis between Thr₄ and Kyn₁₃ results in a ring opened product. In mildly acidic conditions (pH = 3-6) a two-step pathway results in the succinimido intermediate ("anhydro-daptomycin") at Asp₉ and subsequent reversible formation of two aspartic acid isomers. At lower pH, unknown degradation pathways occur.

During the biopanning process the biotinylated probe is dissolved in PBS at very mildly basic conditions (pH = 7.4) for up to 6 h. The question remains is daptomycin stable under these conditions?

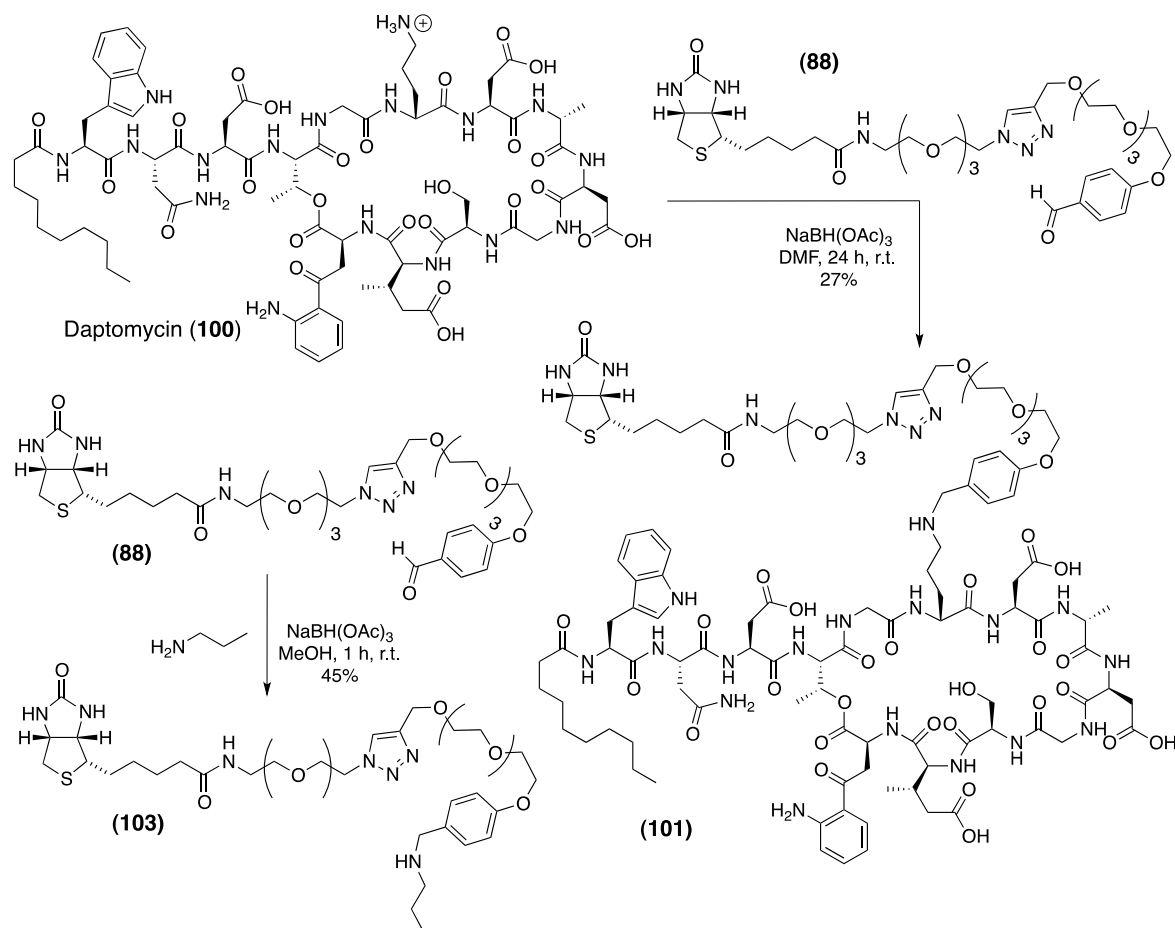
3.2.3.2.2 Results and discussion

3.2.3.2.2.1 Derivatisation

Initially, we envisaged that a probe, randomly derivatised at one of the four carboxy groups would give a probe where the daptomycin is displayed in several orientations and would maximise the ability for the probe to find its binding partner. In this way, if any of the carboxy groups are critical, identification of the protein binding partner(s) would still be possible. Initial attempts with DCC and 0.2 equivalents of NHS resulted in the LC-ESI-MS detection of a single NHS-ester species ($[M+2H]^{2+}$), but coupling to the amine could not be observed. Various conditions³¹⁰⁻³¹² were adapted to improve the yield of the mono-NHS ester (e.g. solvents; water; phosphate buffers at pH = 3 or 5.3, which correlate to the pKa values of the carboxylic acids³⁰⁹, DMF, DCC, EDC, HOBt), but only small amounts of NHS esters could be observed. Due to the poor results and the likelihood of intramolecular or intermolecular reaction of the NHS ester with other nucleophilic groups, this route was quickly abandoned in favour of the established reductive amination of daptomycin⁶⁹.

This successful strategy (see Scheme 5) involved dissolving daptomycin and aldehyde in DMF and the resulting imine reduced *in situ* with excess sodium triacetoxyborohydride (STAB-H). The reaction was quenched with water and passed through a C₁₈ cartridge. After freeze-drying, the residue was purified by reverse phase HPLC in an ammonium phosphate buffer (pH = 5.0) to prevent irreversible degradation of daptomycin. ESI-MS yielded the diprotonated ion and HR-ESI MS confirmed the correct molecular formula for the derivatised product (**101**). A control probe (**103**) was synthesized incorporating propylamine as an analogue for the ornithine side chain (see Scheme 5).

Compared to other hydride reducing agents, STAB-H exhibits remarkable selectivity to reducing imines derived from aldehyde rather than ketones³¹³. On the other hand, the bulkiness of the reducing agent combined with the size and flexibility of both daptomycin and the polyethyleneglycol linker simply limit the accessibility of the imine resulting in only a moderate yield for this step.



Scheme 5: Synthesis of biotin-daptomycin (**101**) and biotin-propyl (**103**) as control probe.

Unreduced imine was always recovered and consequent hydrolysis back to the aldehyde and daptomycin was observed during workup and purification. The low yield was tolerated in favour of degradation resulting from long exposure to excess reducing agent.

To prove the site of alkylation, 2D-NMR data (HSQC, HMBC, COSY, ROESY and NOESY) were acquired for biotin-daptomycin (**101**). The data failed to provide sufficient evidence due to spectral crowding. Difficulties arose from the high flexibility of the species and therefore not every proton and carbon of daptomycin (**100**) could be assigned (see Appendix 8.2.2).

However, these data revealed that the linker moiety was present as a 1:2 mixture of biotinylated daptomycin (**101**) and an oxidised analogue incorporating biotin sulfoxide (**102**; see Figure 27). This observation was initiated by the presence of two subsets of urea N-H signals (δ_{H} 6.78 / 6.69 and 6.42 / 6.37, see Figure 25). COSY, HMBC and ROESY correlations showed the presence of two biotin A+B-ring systems. The determined chemical shifts for biotinsulfoxide were in good accordance to literature values³¹⁴.

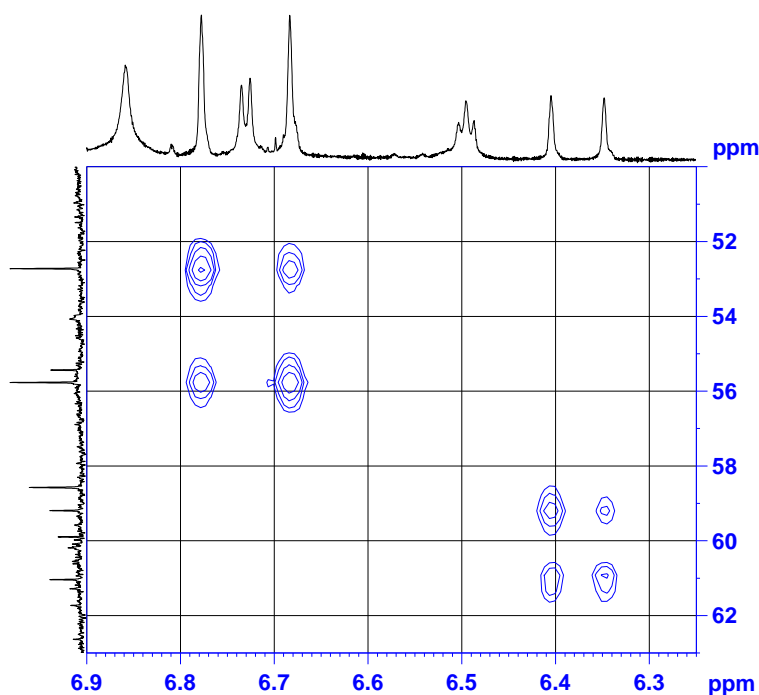


Figure 25: Selected HMBC couplings (DMSO- d_6 , 600 MHz) of the biotin urea N-H protons of biotin-daptomycin (**101**; δ_H 6.42, 6.37) and biotinsulfoxide-daptomycin (**102**; δ_H 6.78, 6.69).

Both species co-eluted as one peak during RP-HPLC separation but could be detected by NMR spectroscopy and MS. The presence of the sulfoxide was unnoticed during the preparation of the linker (**88**), and therefore carried forward into the synthesis of biotin-daptomycin (**101**). Backtracking, it was possible to trace the source of oxidation to the click chemistry reaction of biotin-azide (**82**) and the acetylene (**78**). We speculate that copper(II)-mediated oxidation of biotin resulted in some sulfoxide being produced.

Generally, the oxidation of biotin has been reported to occur as a side product during conjugation of a Ni(Salen)-biotin complex to oligopeptides^{315,316}. In this case, the authors utilised a reactive oxygen species (magnesium salt of monoperoxyphthalic acid) for their conjugation conditions, which clearly provides a good source of oxygen. In another publication⁷⁶, the oxidation of biotin was observed during solid support oligonucleotide synthesis. The authors therein postulated that the oxidative conditions (I_2 /pyridine/water) triggered the sulfoxide formation, which was reported to be quantitative for the utilised borosilicate based solid support. In the most extensive review on the copper(I)-catalysed azide-alkyne cycloaddition, Meldal *et al.* discussed several examples of oxidative side reactions of standard click chemistry conditions²⁴⁹. Yet, none of these examples was related to oxidation of biotin like compounds.

An attempt to separate those two species proved unnecessary for two reasons.

Firstly, the oxidation of the biotin sulfur will not affect which and how any protein binding partner will interact with daptomycin. And secondly, the affinity of biotin-sulfoxide to avidin is greatly reduced compared to biotin³¹⁷, causing biotin-daptomycin to outperform its oxidised counterpart during the coating of the microtitre plates. Hence, probes **101** and **102** were applied as mixture.

As the NMR data failed to provide sufficient proof for the site of alkylation, MS-MS data were acquired and the fragmentation pattern of the probe was analysed. The results provided strong evidence for successful reductive amination at precisely the ornithine amine.

In the first instance, a full scan between 400-2000 amu was performed to identify the doubly charged proton and potassium ion of biotin-daptomycin (**101**; $[M_1+H+K]^{2+}$) at 1212.3 amu (see Figure 26, A and B) as the most abundant signal. The biotin-sulfoxide probe (**102**) was identified as di-protonated species ($[M_2+2H]^+$) at 1201.3 amu, and furthermore as the potassium and sodium adducts ($[M_2+K+H]^{2+}$ or $[M_2+Na+H]^{2+}$). The signal for ($[M_1+H+K]^{2+}$) was isolated for MS-MS (see Figure 26).

In alternating collision induced detection (CID) and electron transfer detection (ETD) fragmentation of this ion (see Figure 26, C), the major two fragments were identified as the biotinylated linker including the amine group of ornithine and the potassium

salt of daptomycin lacking the ornithine amino group, i.e. incorporating an L-norvaline in place of the L-ornithine ("nor-daptomycin"; Figure 27).

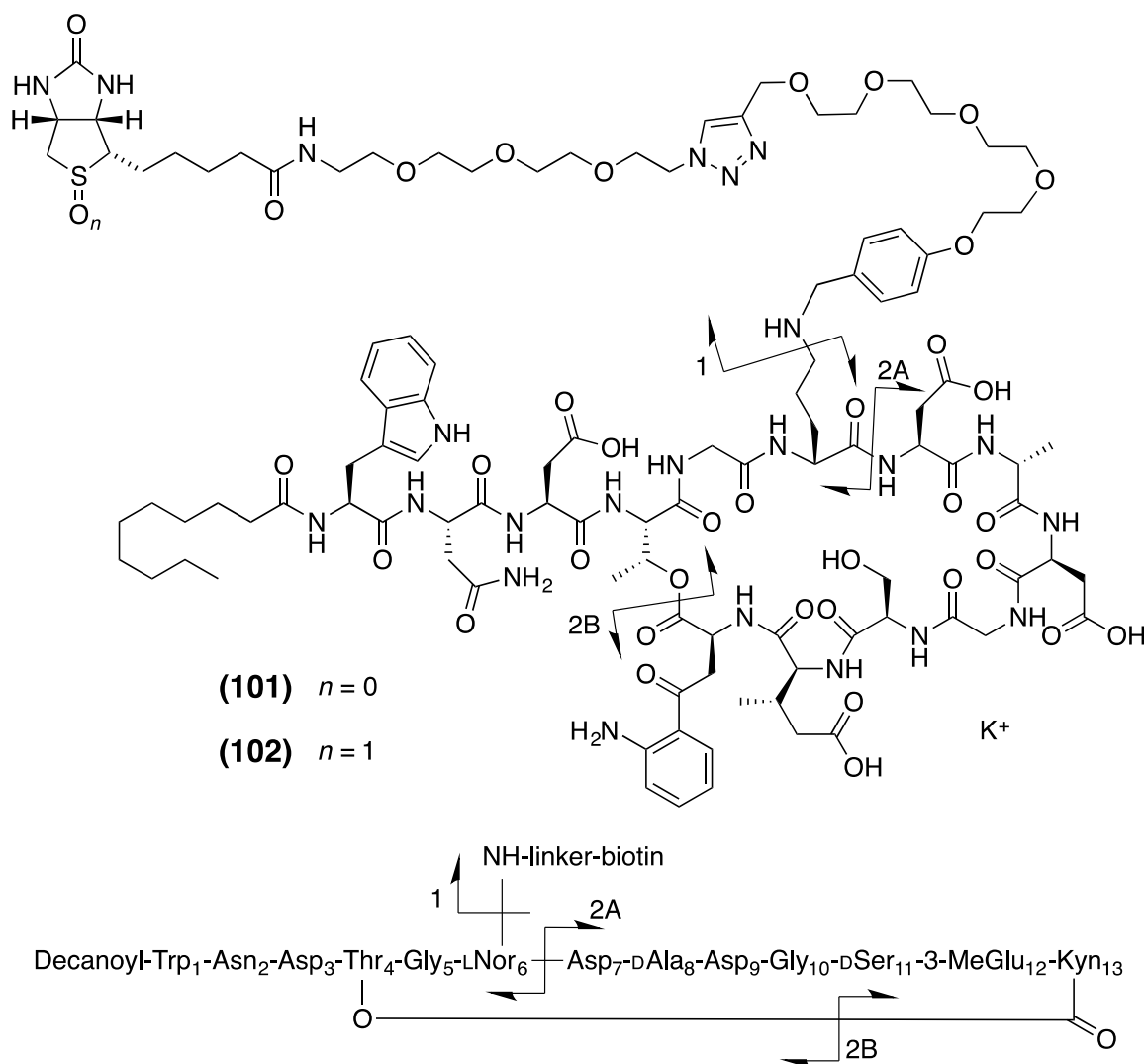


Figure 27: Structure and amino acid sequence of biotin-daptomycin (101) and fragmentation pattern during direct injection ESI-MS/MS analysis (1, 2A, 2B).

MS³ studies of the daptomycin fragment revealed two primary fragmentation pathways of nor-daptomycin (see Figure 27 and Figure 28; pathways 2A and 2B). The starting point for both pathways was the potassium adduct of nor-daptomycin. Pathway 2A started with a ring opening between Nor₆ and Asp₇ and included an immediate loss of water. In parallel, pathway 2B was observed, which showed a ring-opening of nor-daptomycin at the lactone between Thr₄ and Kyn₁₃. Further fragmentation patterns perfectly matched the amino acid sequences for the ring opened initial fragment. Due to the initial loss of the TEG-biotin with an amino group and the consequent fragmentation pattern showing that the rest of the molecule was

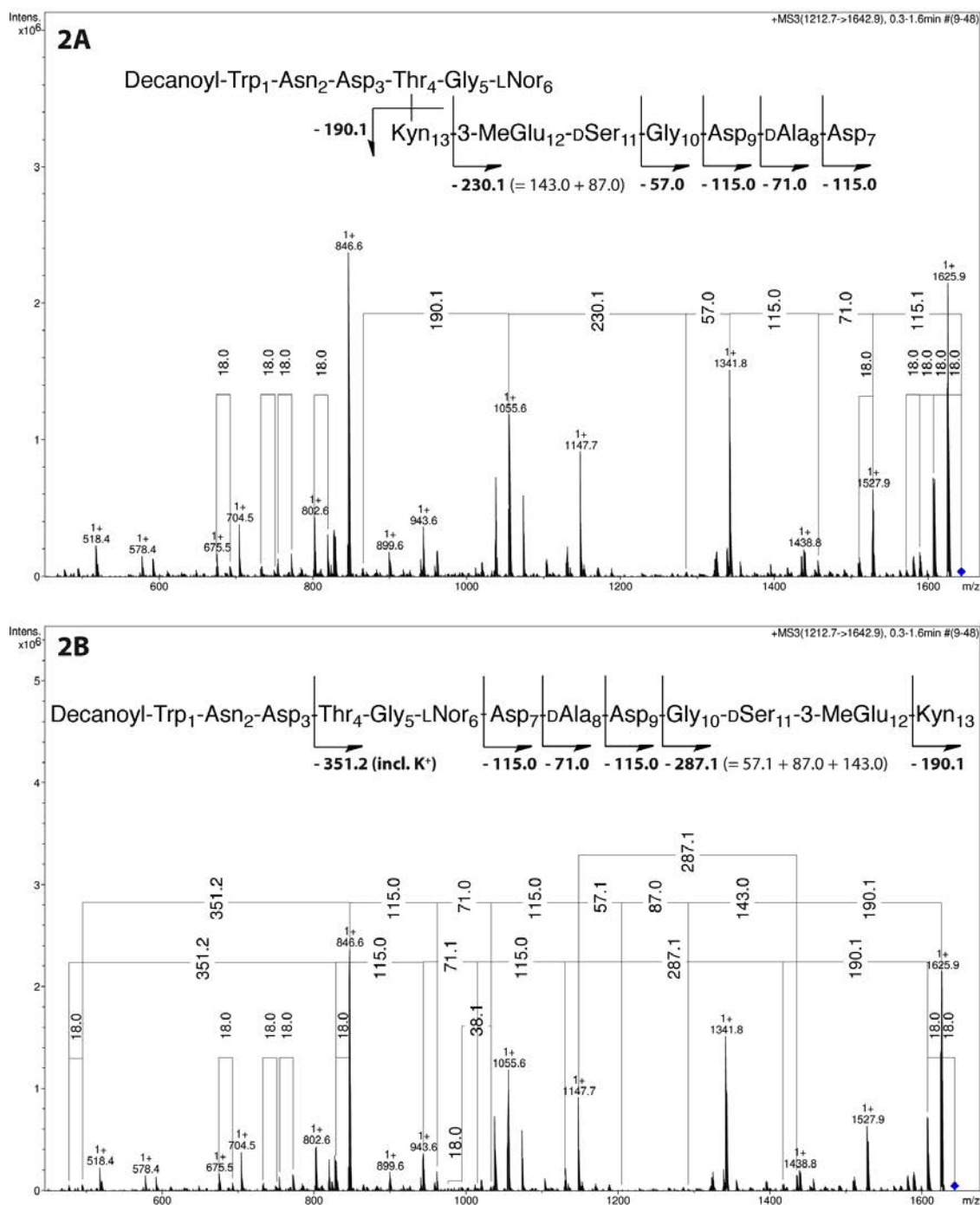


Figure 28: MS3 data of Nor-daptomycin, showing fragmentation pattern following two pathways (2A, 2B).

intact, it was concluded that reductive amination had occurred on the ornithine side chain as outlined in Scheme 5.

Finally, the aldehyde-containing linker (**88**) was reacted with propylamine to yield the control probe biotin-propyl (**103**), mimicking the alkyl side chain of ornithine. The yield for this reaction was significantly higher than for biotin-daptomycin, suggesting that sterical hindrance and limited reaction site accessibility were indeed contributing to the low yields of biotin-daptomycin.

3.2.3.2.2 Stability of daptomycin and purity of biotin-daptomycin

To guarantee stability throughout the biopanning process, daptomycin was dissolved in PBS and kept at room temperature for 12 hours. The formation of degradation products was monitored by reverse phase HPLC and the data presented in Figure 29.

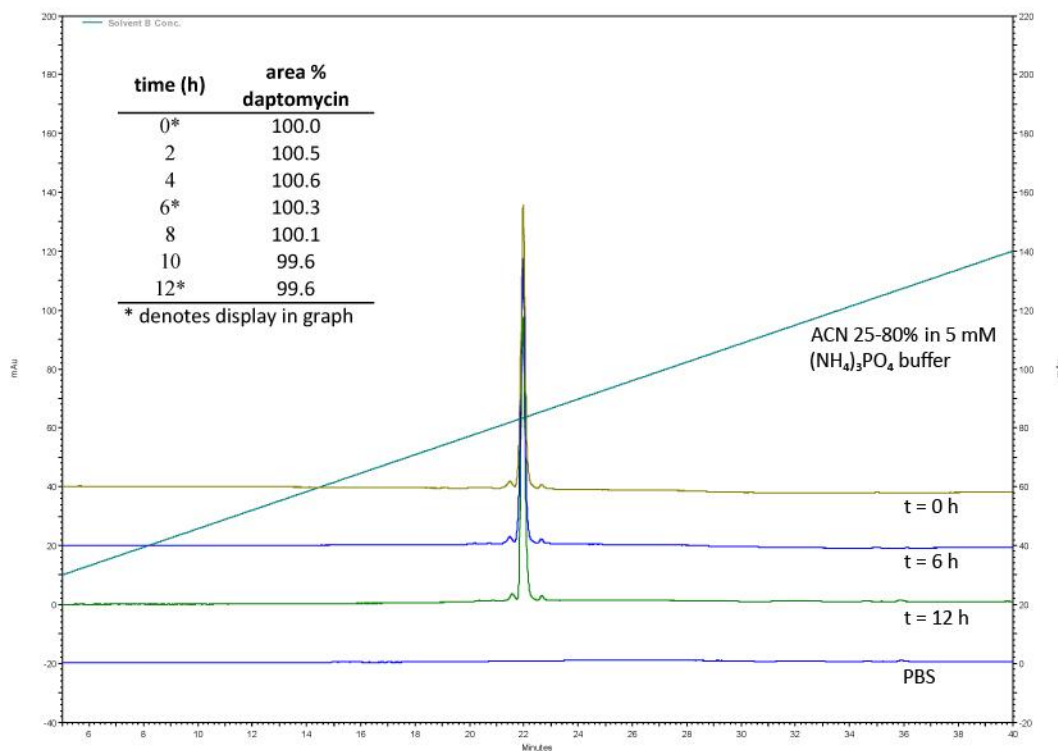


Figure 29: Reverse phase HPLC traces of daptomycin (100) in PBS (1 mg/mL) detected at 230 nm after 0 h, 6 h and 12 h

Clearly, daptomycin is degrading but slowly (1% over 12 h), and is easily stable enough for the biopanning experiments (~ 6 h).

In addition, the biotinylated daptomycin was checked for purity via RP-HPLC analysis and found to elute off the column as one single peak (see Figure 30) despite containing a mixture of biotin and biotin-sulfoxide linker.

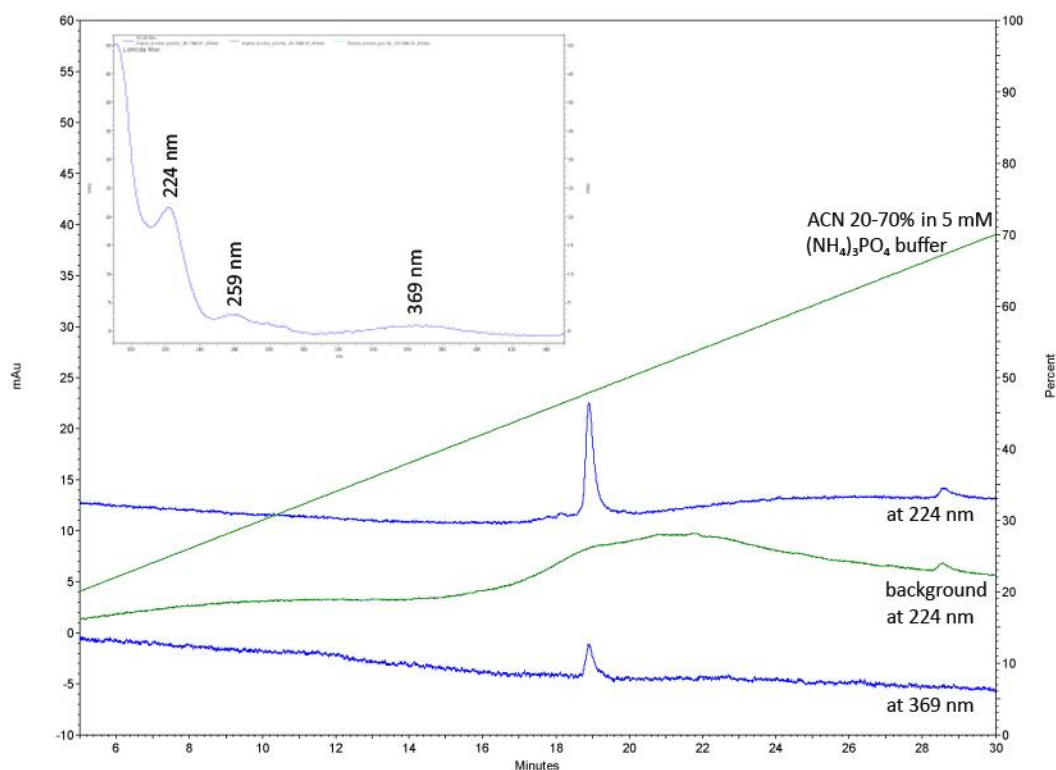


Figure 30: Reverse phase HPLC analysis of purified biotin-daptomycin including UV traces at 224 nm and 365 nm and the compound's UV spectrum

3.2.3.2.2.3 Antibiotic activity

To verify that the site of alkylation did not abolish the antimicrobial activity of daptomycin, we performed a turbidity microdilution bioassay on *Staphylococcus aureus* and the data fitted to a sigmoidal dose-response curve (variable slope). In our assay, the determined $LD_{50} = 1.26 \times 10^{-9}$ M (95% CI [1.18×10^{-9} , 1.35×10^{-9}], $R^2 = 0.9957$) for daptomycin compares favourably with the literature (MICs against methicillin-susceptible *S. aureus* is reported as $3 - 6 \times 10^{-7}$ M^{299,318}). The biotinylated daptomycin showed slightly reduced activity with an $LD_{50} = 6.27 \times 10^{-8}$ M (95% CI [4.53×10^{-8} , 8.68×10^{-8}], $R^2 = 0.9626$) but still better than literature values.

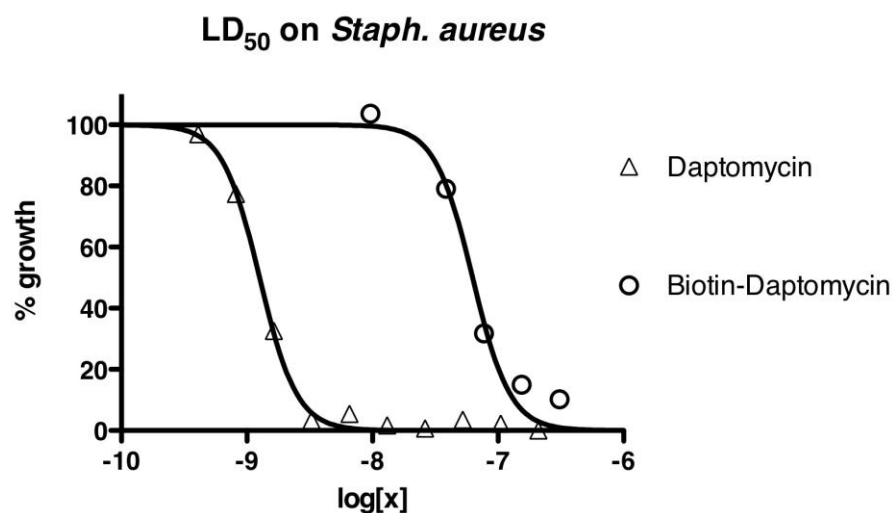


Figure 31: Antibacterial assay of daptomycin (100) and biotin-daptomycin (101) on *S. aureus*, displayed as percent growth versus concentration in mol/L. Data was fitted to a sigmoidal dose-response curve with variable slope (Hill plot) with the top and bottom values set to 100 and 0 respectively.

It is not surprising given that the linker would affect uptake of the depsipeptide by the bacteria. Nevertheless, biotin-daptomycin still performs as a potent antibiotic, which gives strong evidence in favour of the chosen site of alkylation.

3.2.3.3 Artemisinin

3.2.3.3.1 Introduction and review of SAR studies

Artemisinin (**104**; see Figure 32) is a highly oxygenated, polycyclic sesquiterpene lactone that naturally occurs in the leafy parts of the sweet wormwood, *Artemisia annua*. Artemisinin incorporates an endoperoxide bridge to give a 1,2,4-trioxane substructure, and its bioactive form dihydroartemisinin (**105**; DHA) includes a cyclic hemiacetal. DHA proved to be the substance responsible for the plant's reputed medicinal action, which had been employed in Chinese traditional medicine against fever and malaria for many centuries. Artemisinin (or "Qinghaosu") was first isolated by Chinese chemists in 1971^{319,320} and since then has attracted an immense amount of attention and research as an affordable and renewable malaria cure. Artesunate (**106**) is part of the artemisinin group of drugs and is semisynthetically produced from DHA and succinic anhydride to provide a more bioavailable version of DHA. The artemisinin group is recommended by the WHO for antimalarial treatment against both chloroquine-sensitive and chloroquine-resistant strains of *Plasmodium falciparum*.

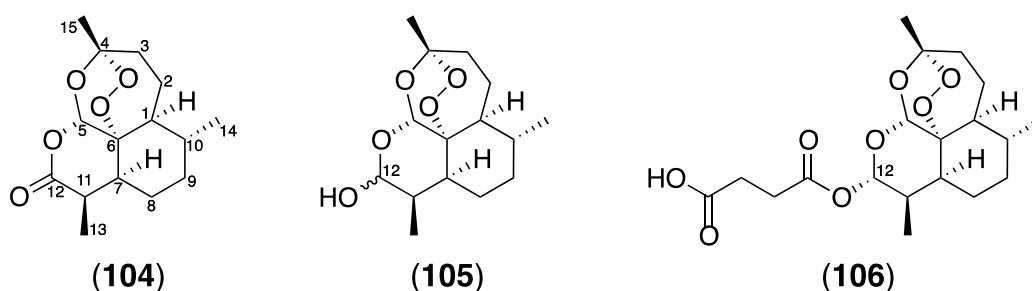


Figure 32: Structure of artemisinin (**104**), and synthesis of the reduction product dihydroartemisinin (**105**) and the succinic acid hemiester; α -artesunate (**106**).

The unusual structure and therapeutic significance of this compound class resulted in wide-ranging research into its chemistry and pharmacology, not only as an antimalarial but also as a new drug for anticancer treatment³²¹. Various members of the artemisinin group showed cytotoxicity in nano- and low micromolar range against breast, colon, lung and ovarian tumour cells as well as central nervous system tumours and leukemia³²².

The mode of action underlying the artemisinin group's antimalarial activity is extensively researched, but still highly disputed³²³⁻³²⁷. Similarly, there is yet no conclusive understanding of the mode of action for its potent and highly diverse

anticancer activity profile. Even more so, a myriad of different human cellular targets have been reported to be affected^{328,329}.

3.2.3.3.1.1 *Malaria and artemisinin's mode of action as antimalarial*

The gold standard for treatment of severe malaria infections by the most pathogenic parasite *Plasmodium falciparum* is an Artemisinin-based Combination Therapies (ACT), which combines DHA or artesunate with standard non-endoperoxide antimalarials such as quinolines and antifolates (e.g. piperaquine, mefloquine, amodiaquine or lumefantrine)³³⁰. This combinatorial approach aims to prevent the development of resistance to the currently most effective and seemingly last resort antimalarial compound class, the artemisinins. Unfortunately, resistance is already being observed throughout clinical treatments in several highly infected areas³²⁴. Hence, comprehensive understanding of the compound's activity is critical to effectively provide alternative compounds in the near future and allow for new forms of treatment.

The current opinion on the mode of action (MOA) is that artemisinin inhibits *Plasmodium falciparum* sarcoplasmic, endoplasmic calcium ATPase PfATP6, or, in a wider context of action against non-plasmodial cells, the sarco/endoplasmic reticulum calcium ATPase (SERCA)³²⁴⁻³²⁶. Other hypotheses include parasite heme, electron transport chain in mitochondria or other proteins as targets^{323,331,332}.

Whereas the MOA based on SERCA inhibition claims the calcium transporter protein as a single target, the alternative mechanisms suggest an initial electron transfer to activate artemisinin's endoperoxide bridge, resulting in formation of different alkoxy and thence carbon centred radicals, suggesting a less specific alkylation of potentially multiple targets. It was formerly assumed³³³, that the free artemisinin radicals alkylate heme, preventing the hemozoin formation during hemoglobin digest within the parasite in a chloroquine like manner and therefore resulting in accumulation of heme to toxic levels, ultimately causing death of the parasite. This hypothesis also accommodated for heme-Fe²⁺ to activate the endoperoxide bridge initially, as heme-Fe²⁺ and exogenous Fe²⁺ enrich to relatively high levels during the parasite's required digestion of hemoglobin³³².

However, artemisinins also exhibit activity against stages of the parasite that do not actively manufacture heme. Also, other experiments demonstrated that the site of action is not exclusively within the food vacuole^{334,335}. Coghi *et al.*³³³ unambiguously

elaborated that the antimalarial activity does not rely on endoperoxide activation through hemoglobin-Fe²⁺ or heme-Fe²⁺. The authors chose the well-established chemistry behind carbon monoxide based functional inhibition of metalloproteins, especially reduced heme proteins, to prove an increased antimalarial activity during reduced metalloprotein activity.

On the other hand, if PfATP6, or in other organisms SERCA, is to be assumed to be the main target for artemisinins, most of the reported effects are coherently explained. Several studies proved inhibitory effects of artemisinin against SERCA from different organisms like rabbit muscle preparations³³⁶, human colon cancer cells³³⁷, *Toxoplasma gondii* SERCA after heterologous expression in Ca-ATPase defective yeast³³⁸ or *Trypanosoma cruzi* membranes³³⁹. In a very recent study by Moore *et al.*³⁴⁰ on *Saccharomyces cerevisiae* the authors not only provided further evidence for the inhibition of ATP-dependent calcium transporters (Pmc1p and Pmr1p) but also indicate mitochondria as secondary targets for artemisinin derived radicals. It is yet to be determined which of the postulated two major pathways is most dominant *in vivo* both for parasites and in human cells.

As we aim to employ artemisinin in reverse chemical proteomics to isolate the most avid human protein-binding partner, it is of utmost importance to consider the molecule's reactive sites before designing a chemical route for biotinylation.

3.2.3.3.1.2 Mode of action as anticancer agent

Artemisinin appears as a very versatile drug, with effects against various unrelated cancer cells lines such as colon, breast, lung and pancreatic cancer, or leukaemias, all of which are recently reviewed in excellent publications^{322,329,341}. General mechanisms identified as being impacted by artemisinin derivatives include inhibition of angiogenesis in tumours³⁴²⁻³⁴⁶, inhibition of proliferation, migration and tube formation of human umbilical vein endothelial cells (HUVEC)³⁴⁷, or reduced expression of vascular endothelial growth factor (VEGF) on HUVECs³⁴⁸. Additionally, downregulation of B-cell leukemia/lymphoma 2 (BCL-2) and upregulation of BCL-2 associated X protein (BAX) were linked to the observed HUVEC apoptosis³⁴⁹. Very recently, artemisinin was reported to trigger apoptosis, or programmed cell death (PCD) in breast cancer cells with hierarchical signalling from lysosomes to mitochondria³²⁸.

The overall picture is highly diverse, with multiple general biochemical pathways seemingly targeted by artemisinin or its derivatives, thus so far, no coherent, single target mode of action has been proposed. Given the various effects observed, a fundamental impact on cellular homeostasis, as for instance caused by inhibiting Ca-ATPase transporters, seemed reasonable as an underlying mechanism, but this clearly requires extensive further research. It is furthermore speculated, that artemisinin exhibits a multilayered mode of action, involving multiple specific targets and/or electron transfer initiated activation required to exert any of its effects. The anticancer effects were shown to be partly iron(II) dependent³⁵⁰ - such as observed in the MOA against *Plasmodium sp.*

If the presence of ferrous ions is required to maintain activity, it can be concluded that the endoperoxide moiety is activated by electron transfer to give rise to alkylating C-centred radicals and radical oxygen species (ROS). When tested against cell lines that overexpress oxidative stress enzyme the antitumour activity was lost, which provides additional evidence for the above hypothesis, as these enzymes apparently eliminate ROS.

In summary, artemisinin radicals, ROS and remaining, unactivated artemisinin endoperoxides will all cause effects on different cellular targets and functions and no unifying target is likely to be described at all.

3.2.3.3.1.3 Derivatisation

As the information on artemisinin's targets is highly ambivalent and the occurring structural rearrangements are altering the molecule's surface area, it is difficult to define those functionalities of the molecule, which are most important for the optimal binding to the protein target. As outlined above, it is a well established view that the endoperoxide bridge is crucial for the bioactivity and SAR studies clearly portray the relation of activity and presence of this functionality³⁴¹.

Artemisinins are most commonly derivatised on the C-12 position, primarily to yield esters and ethers, or alternatively, by reducing the hemiacetal to an ether and substitute directly via a carbon-carbon bond to C-12. Haynes *et al.*³⁵¹ provided an extensive review on common strategies to form esters and ethers of DHA. Most importantly, derivatisation at C-12 has been shown to increase activity³⁵¹⁻³⁵³.

We aim to contribute to the understanding of the biology behind this highly intriguing small molecule by applying it to reverse chemical proteomics (see Section 4.5). Given

the existing evidence, alkylation at C-12 of the artemisinin is the most likely to not effect biological activity.

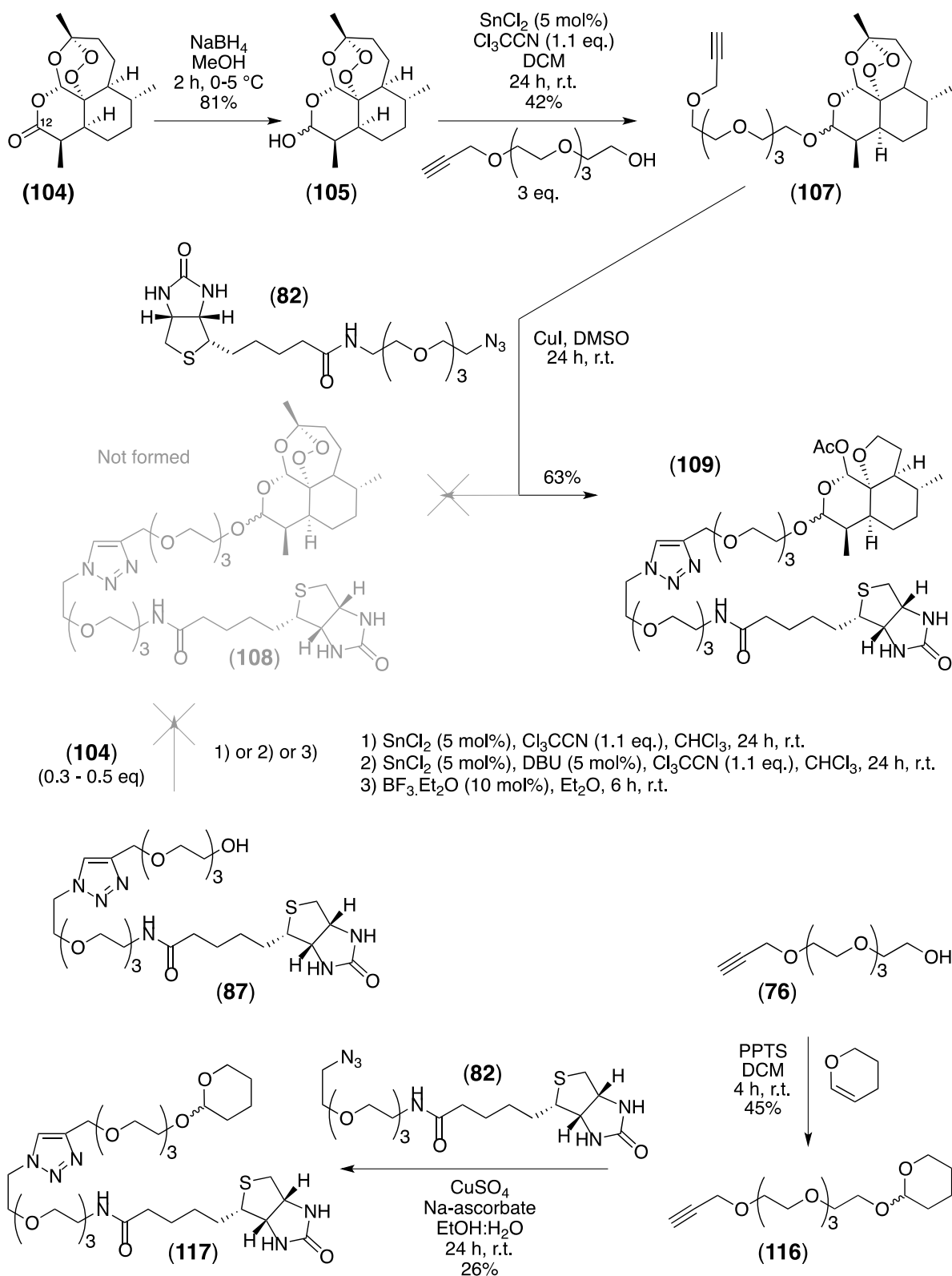
3.2.3.3.1.4 Purity and stability

Generally, artemisinins degrade in acidic conditions yielding various species lacking the endoperoxide bridge^{354,355}. In non-biological environments, certain transition metals besides iron may cause reductive activation of the endoperoxide bridge and result in ring opened species with no or drastically reduced bioactivity. Additionally, artesunate, the more bioavailable prodrug of active compound dihydroartemisinin, quickly hydrolyses in acidic, aqueous conditions to free DHA and succinic acid³⁵⁶. At physiological pH, hydrolysis occurs as well, but at greatly reduced rate. Nevertheless, this necessitates solution-free drug preparations, i.e. as freeze dried powders, for storage and transport of the medication (days to months). However, phosphate buffered saline solutions at physiological pH provide a good medium to maintain stability on a short term scale (hours)³⁵⁶.

3.2.3.3.2 Results and discussion

3.2.3.3.2.1 Derivatisation

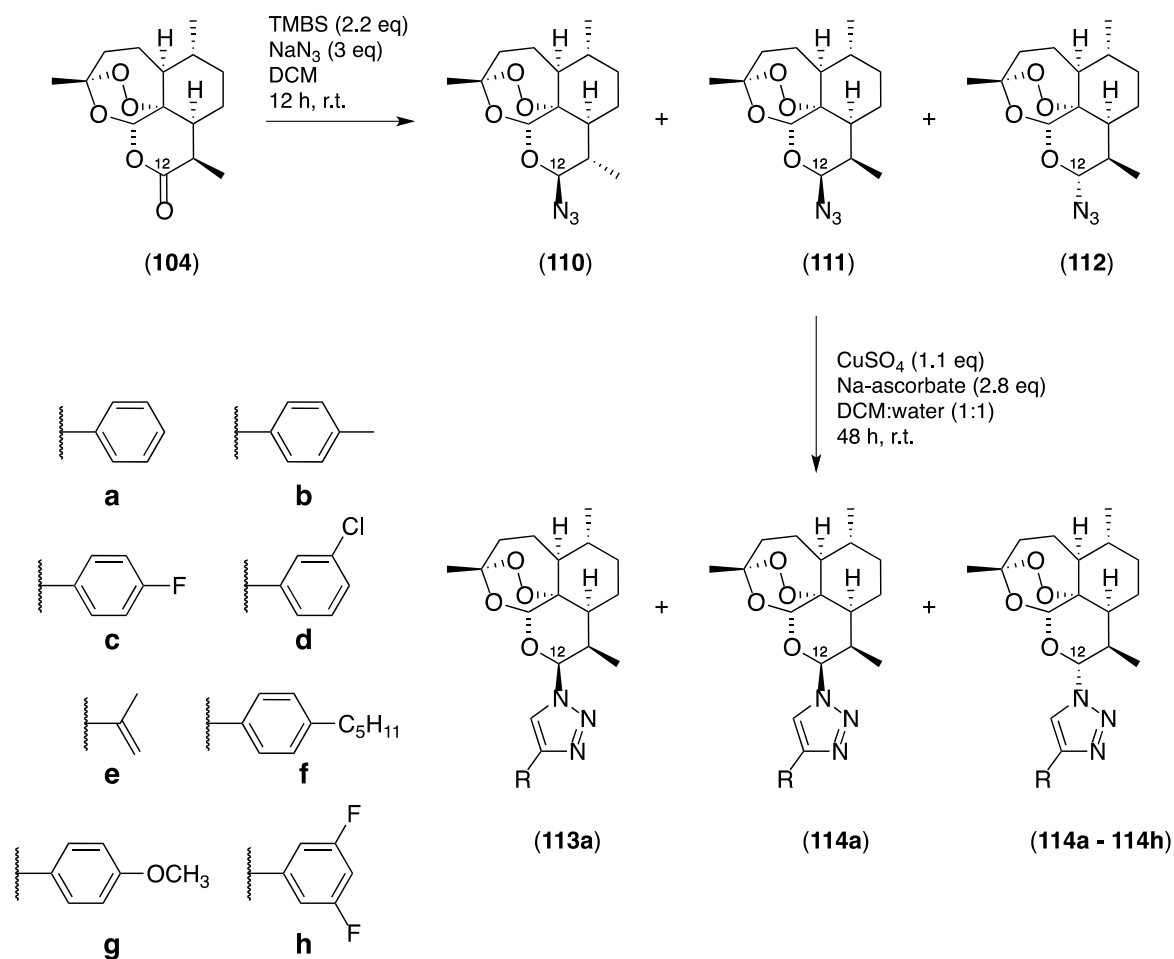
A synthetic route (Scheme 6) was established starting from artemisinin and PEG-based linkers described in 3.2.1.2 and 3.2.2.2. Initially, artemisinin (**104**) was reduced with sodium borohydride to yield an equimolar mixture of α - and β -DHA (**105**) in accordance to literature results³⁵⁷. Due to the presence of the reducing agent, it is most important to maintain low temperatures and prevent reduction of and consecutive radical formation by the endoperoxide bridge. The following step was based on DHA's hemiacetal forming trichloroacetimidate intermediate (Schmid reaction) and being activated through SnCl_2 to give both α - and β -epimers of the acetal DHA-TEG-acetylene (**107**) in a ratio of 1:2 (by NMR). We planned to react this acetylene with biotin-TEG-azide to form a triazole linker because artemisinin triazoles have been previously described to have enhanced anticancer activity³⁵⁸. When acetylene (**107**) and the biotinylated azide (**82**) were reacted ($\text{Cu(I)}/\text{DMSO}$) a product was formed with a molecular weight (ESI-MS $m/z = 966$ $[\text{M}+\text{H}]^+$) fitting the expected value for the product. However, during 2D-NMR characterisation, the elucidated structure differed from the expected product.



Scheme 6: Attempted synthesis of biotin-dihydroartemisinin (108). Sideproduct (109) formed instead. Successful synthesis of biotin-pyran control probe (117).

While Cho *et al.*³⁵⁸ observed triazole formation with a variety of artemisinin derivatives in moderate yields (see Scheme 7), we were not able to reproduce these results, under a variety of conditions ($\text{CuSO}_4/\text{Na-ascorbate}/\text{DCM}:\text{water}$; $\text{CuSO}_4/\text{Na-}$

ascorbate/THF:water/TBTA; CuI/THF:water/(ultrasound), including use of freshly prepared CuI, which allowed forgoing the reducing agent (sodium ascorbate).



Scheme 7: Scheme from Cho *et al.* (2009)³⁵⁸, the C-12-azide was formed to yield three isomers (110-112; isolated yields 110: 21%, 111: 17%, 112: 49%), the azides are then subjected to Cu(I)-catalysed azide-alkyne Huisgen cycloaddition with acetylenes to yield products (113a, 114a, 115a - 115h; isolated yields 113a: 43%, 114a: 20%, 115a: 75%, 115b: 60%, 115c: 57%, 115d: 63%, 115e: 71%, 115f: 65%, 115g: 69%, 115h: 66%).

In our case, the major product always formed was a biotinylated degradation product of DHA (**109**). The NMR characterisation revealed an unexpected signal (δ_c 169.0) suggesting an ester carbonyl. HRMS suggested the molecular formula C₄₄H₇₄N₆O₁₄S (–0.562 ppm error). The IR absorptions supported the existence of an ester (1751, 1090, 1024 cm⁻¹). The ¹³C NMR spectrum confirmed the presence of 44 carbons (see Table 16). All protonated carbons were correlated to their protons by HSQC spectroscopy. The chemical shifts of carbons and protons from positions 16 to 50 were in agreement with the structural assignment of the biotinylated alcohol (**87**). HMBC correlations of the methyl at 2.07 ppm (H-13) to the ester carbonyl at δ_c 169.0 suggested the presence of an acetate and a weak HMBC correlation of the same methyl to a carbon at 87.5 ppm (δ_H 6.07; Figure 34) suggested the acetate was

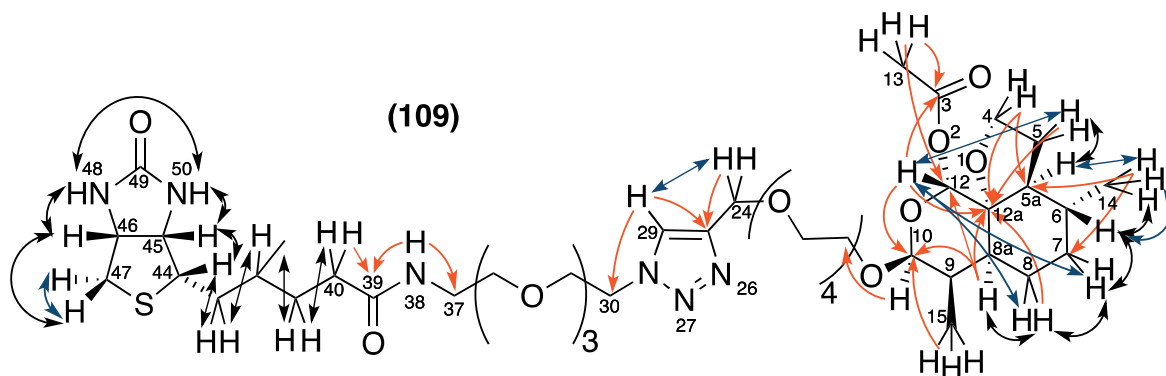


Figure 33: Structure and selected 2D-NMR correlations of biotinylated DHA-derivative (109). The numbering of the artemisinin-moiety was adapted from literature³⁵⁹. Colours: orange (HMBC), blue (ROESY), black (COSY).

attached to a hemiacetal (C-12; see Figure 34). The presence of an acetate clearly indicated a rearranged artemisin and suggested a structure similar to **121** (scheme 8). COSY showed three spin systems; O-C(O)H-CH-Me, -CH-CH₂-CH₂-CH(Me)-CH and CH-CH₂-CH₂-O. These were connected together using HMBC correlations (see Figure 33 and Table 16). In brief, H-10 (δ_{H} 4.67, δ_{C} 101.1) was easily identified as an acetal with HMBC correlations to C-12, C-16 (the first carbon of the PEG linker) and C8a, in the next spin system. The chemical shift of C-10 was in agreement with ether derivatives of dihydroartemisinins typically described in literature³⁵¹. H-8a was also

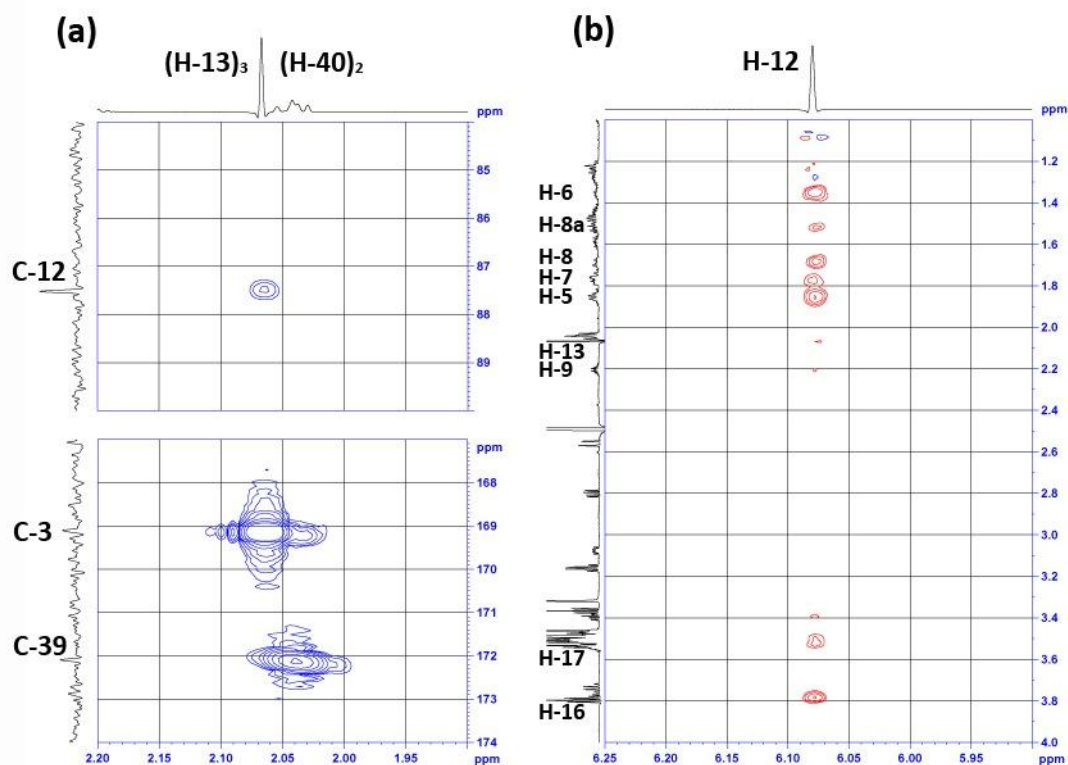


Figure 34: Selected 2D-NMR correlations of compound 109 (600 MHz, DMSO-*d*₆). (a) HMBC couplings of H-13 and C-3, and long range couplings of H-13 and C-12. (b) ROESY correlations between H-12 and H-5, H-6, H-7, H-8, H-8a, H-9, H-13, H-16 and H-17.

Table 16: Selected NMR(DMSO-*d*₆, 600 MHz) data for biotinylated artemisinin derivative (109).

Pos.	δ_c	δ_H , m (J in Hz)	COSY (H no.)	1H - ^{13}C HMBC (C no.)	ROESY (H no.)
3	169.0	-			
4	67.5	4.03 , 3.74 q (7.9)		5, 5a, 12a	
5	27.1	1.85 m, 1.75 m	5a	4, 5a(w), 6(w), 9	12
5a	55.0	1.23 m	5	5, 6, 7, 8a, 12, 12a, 14	
6	30.2	1.36 m	14	14(w)	12, 14
7	35.2	1.77 m, 0.86 m	6, 8	4, 6(w), 8a	12, 14
8	24.4	1.86 m, 1.69 m	7, 8a	7, 8a, 12a	12, 14
8a	46.2	1.52 m	8	10, 12, 12a	12, 14
9	32.8	2.21 m	10, 15	8, 8a, 10, 15	10, 12
10	101.1	4.67 d (4.2)	9	8a, 12, 16	9, 15, 16
12	87.5	6.07 s	10(w)	3, 10(w), 12a	5, 6, 7, 8, 8a, 9(w), 13(w)
12a	79.6	-			
13	21.17	2.07 s		3, 12	12(w)
14	20.3	0.86 d (6.4)	6	5a, 6, 7, 8(w), 12a(w)	6, 7, 8, 8a
15	12.2	0.81 d (7.4)	9	8a, 9, 10	10
16	66.7	3.78 m, 3.39 m	17	10	10
17	69.3	3.52 m	16	16	
18-22, 32-35	69.4 - 69.9	3.45 - 3.56 m		18-22, 32-35	
23	69.0	3.54 m		24	
24	63.5	4.50 s		25	
25	143.7	-			
29	124.2	8.04 s	24	25, 30	24, 31
30	49.3	4.49 t (5.1)	31	29	
31	68.9	3.79 m	30	30	
36	69.1	3.36 t (6.0)	37	37	
37	38.4	3.16 m	36, 38	36, 39	38
38	-	7.82 t (5.6)	37	39	37
39	172.0	-			
40	35.0	2.05 t (7.5)	41	41, 42	
41	25.2	1.48 m	40	39, 40, 43	
42	28.2	1.28 m		40, 41, 43	
43	28.0	1.59 m, 1.44 m		41, 44, 45	
44	55.3	3.08 m	45	43	
45	61.0	4.11 m	44, 46	44, 47, 49(w)	48
46	59.2	4.29 m	46, 47	44, 47, 49	47, 50
47	39.8	2.81 dd (β , 12.5, 5.1), 2.56 d (α , 13.4)	46	44, 45, 46	46
48	-	6.41 bs	50	45, 46	45
49	162.7	-			
50	-	6.34 bs	48	45, 46	46, 47

a (w) denotes a weak correlation.

correlated to C-12 (HMBC), connecting the first two spin systems. The last spin system (CH-CH₂-CH₂-O) was connected to the other two via correlations to C-5a and C-12a. H-5a showed correlations to C-5, C-6, C-7, C-8a, C-12, C-12a and C-14 that can only be accommodated with structure **109**, similar to compounds **121** produced by Fe(II).

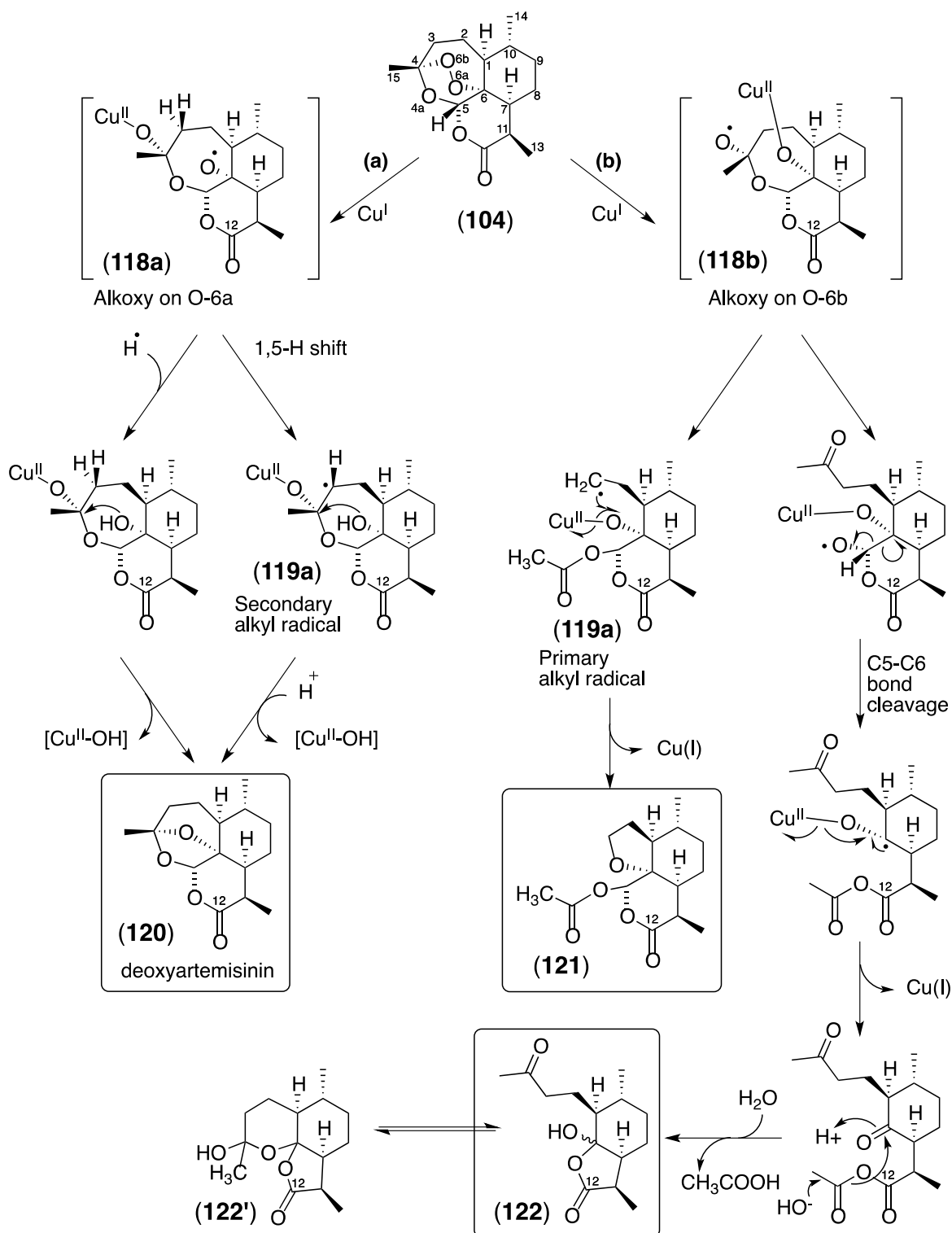
It was suggested that copper can initiate a similar electron transfer as Fe(II)³⁵⁵ (see 3.2.3.3.1.1), which could explain the rearrangement of the endoperoxide moiety of artesunate to **109**.

At about the same time this work was completed, Bousejra-El Garah *et al.*³⁵⁹ published a mechanistic study of artemisinin degradation initiated by Cu(I). Similar to the known mechanism of Fe(II), the authors postulated a reductive activation of one of the endoperoxide oxygens giving rise to an alkoxy ion, resulting in two degradation pathways (Scheme 8).

Amongst other rearrangements described (see Scheme 8), cleavage of the C₃-C₄ bond to a primary alkyl radical and subsequent decomplexation from the copper yielded a tetrahydrofuran derivative (**121**) such as isolated from our experiments. In this paper, other degradation products were also isolated but compound **120** was the major product. In our attempted click chemistry, there were several products produced but compound **109** was the only one isolated. It is clear from the NMR spectra that the others probably correspond to analogues of the other compounds in Figure 15. The authors' excellent work clearly provided a mechanistic foundation to our observations.

Cho *et al.* used a biphasic system of DCM and water (see Scheme 7). The artemisinin azides (**110-112**) and acetylenes (**a-h**) they employed are non-polar and the product would be expected to remain in the DCM layer. Mechanistic studies into the Cu(I)-catalysed azide-alkyne Huisgen 1,3-dipolar cycloaddition²⁴⁹ postulated that copper initially coordinates to the acetylene to give rise to a copper acetylide, after which the azide displaces another copper-bound ligand and binds to the copper. Thereafter, an unusual six-membered copper(III) metallacycle is formed. Ring contraction to a triazolyl-copper derivative is followed by protonolysis that delivers the triazole product and closes the catalytic cycle.

In a biphasic system and with apolar adducts, the Cu(I)-catalysed cycloaddition will only occur at the phase boundary, possibly inhibiting the copper mediated degradation observed in the current work.



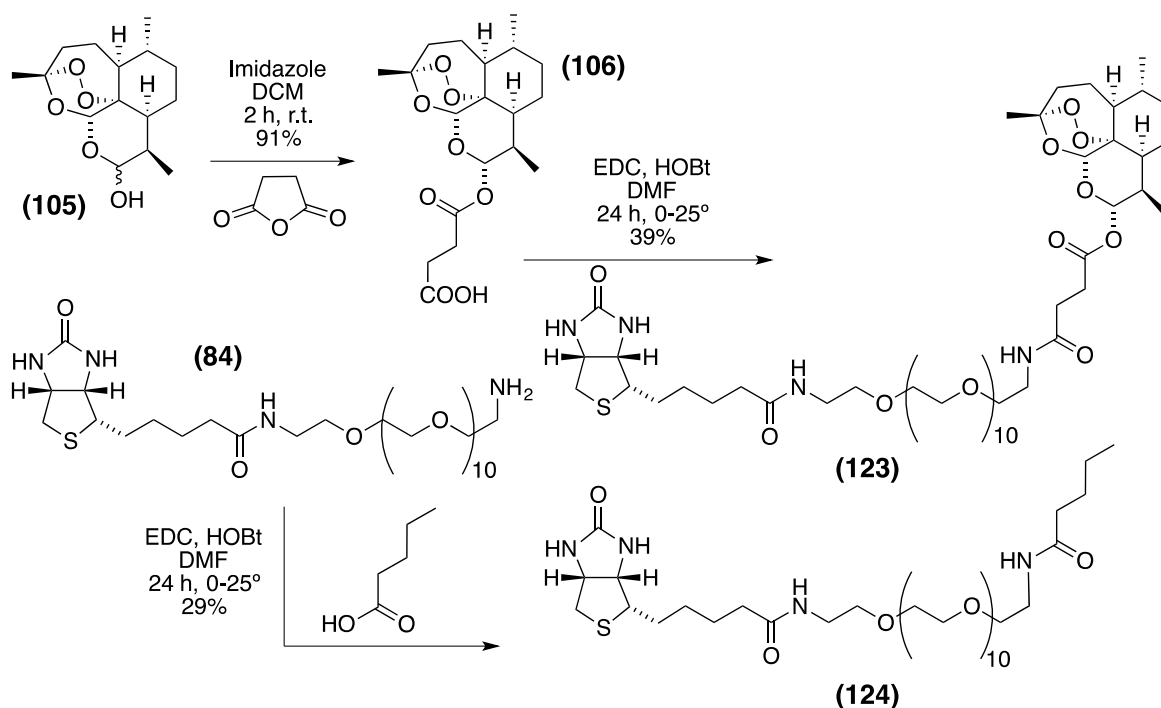
Scheme 8: Reductive activation of artemisinin (104) by CuCl₂/Na₂S₂O₄, giving rise to rearrangement products 120, 121, and 122 (Figure adapted from Bousejra-El Garah *et al.*³⁵⁹)

In our case, a biphasic system was not applicable as the biotinylated linker (**87**) is water soluble and the anticipated product (**108**) was expected also to be water soluble. Hence, we abandoned this approach. In contrast, the control probe (**117**) was easily made from tetrahydropyran albeit in moderate yield (see Scheme 6).

An alternative strategy was initiated, which involved coupling of the biotinylated TEG to artesunate via an amide coupling (see Scheme 9). Artesunate (**106**) was synthesised from DHA and succinic anhydride following literature procedure³⁶⁰.

Interestingly, this reaction exclusively yields the α -epimer only ($^3J_{H12-H13} = 9.8$ Hz), although the starting material consists of a mixture of α - and β -DHA. Because the α -epimer reacts very quickly and is in equilibrium with its aldehyde, all the β -isomer is converted to α - during the reaction^{351,357}.

The biotinylated amine (**84**; see 3.2.2.2) was added to artesunate under standard peptide coupling conditions (HOBt and EDC in DMF), yielding biotin-artesunate (**123**) in moderate yields. A control probe (**124**, biotin-valeramide) containing valeric acid was synthesised to mimic the artesunate side chain.



Scheme 9: Synthesis of artesunate (**106**) from DHA (**105**) and derivatisation with biotinylated linker (**84**) to yield biotin-artesunate (**123**). Biotin-PEG-NH₂ was also coupled with valeric acid to yield biotin-valeramide (**124**).

ESI-MS showed the signal at 591 amu, which fits the disodium salt of the expected molecular formula C₅₃H₉₂N₄O₂₀SNa₂. HRMS provided a mass peak of 1159.5920 amu, which corresponds very well with the monosodium adduct of the expected product.

NMR shifts of all carbons and protons of the artesunate and biotin units of **123** were unambiguously assigned (see Table 17). The PEG linker incorporates ten ethyleneglycol units. As their shifts all overlap within a very narrow ppm range, the ^1H -NMR signal was presaturated at 3.50 ppm before acquisition of 2D experiments. Therefore only the chemical shifts at positions 20-21 and 42-43 could be clearly assigned via long-range couplings (see Figure 35).

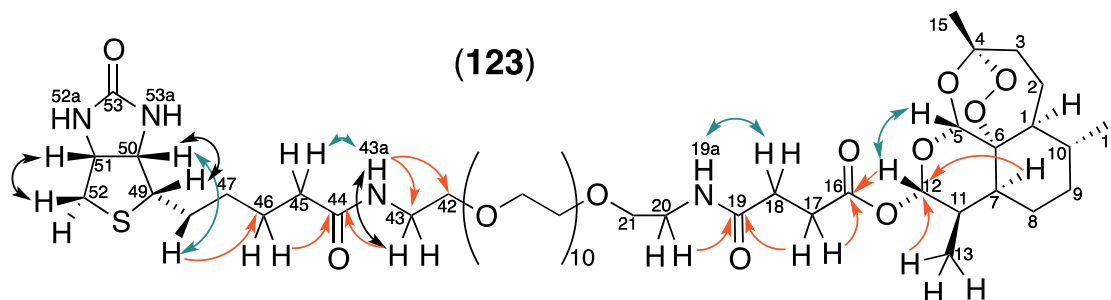


Figure 35: Structure and selected NMR correlations of biotin-artesunate (**123**; 600 MHz, $\text{DMSO}-d_6$). HMBC in red, ROESY in blue and COSY in black.

Combined 1D and 2D NMR experiments (^1H , ^{13}C , HSQC, HMBC, ROESY, COSY; see Figure 36 and Section 8.2.4) suggested that α -artesunate had been successfully coupled to the biotinylated polyethyleneglycol linker (**84**), without impacting the endoperoxide bridge. This was supported by the presence of three methyl groups (H-13, H-14, H-15) that displayed identical shifts to those determined during the synthesis of α -artesunate and were in good accordance with values reported in literature³⁶⁰. The coupling constant of H-12 ($J = 9.8$ Hz) was good evidence for the α -stereochemistry of artesunate. ROESY correlations between H-5 and H-12 (see Figure 36b) further supported the α -stereochemistry of the artesunate moiety. HMBC long range couplings of H-12 (δ_{H} 5.63) to C-16 (δ_{C} 171.4) provided good evidence for the intact ester link between dihydroartemisinin and the succinic acid moiety (see Figure 36c).

During the attempted synthesis of biotin-DHA (**108**), using Cu(I)-mediated click chemistry, the artesunate skeleton rearranged to the acetate **109**. During synthesis of biotin-artesunate (**123**), the lack of the carbonyl signal (δ_{C} 169) suggested that this rearrangement did not occur under standard peptide conditions.

ROESY correlations between H-18/H-19a (δ_{H} 2.38 and 7.96) provided strong evidence for the formation of the amide bond between the amine linker and the succinic acid moiety of artesunate (see Figure 36a). HMBC correlations from C-19 (δ_{C} 170.7) to H-20 (δ_{C} 3.17) provided further evidence for the amide formation (see

Figure 36d). The second amide bond between H-43a/H-45 (δ_{H} 7.83 and 2.05) was indicated by ROESY correlations (see Figure 36a), which clearly shows the link between biotin and the PEG linker.

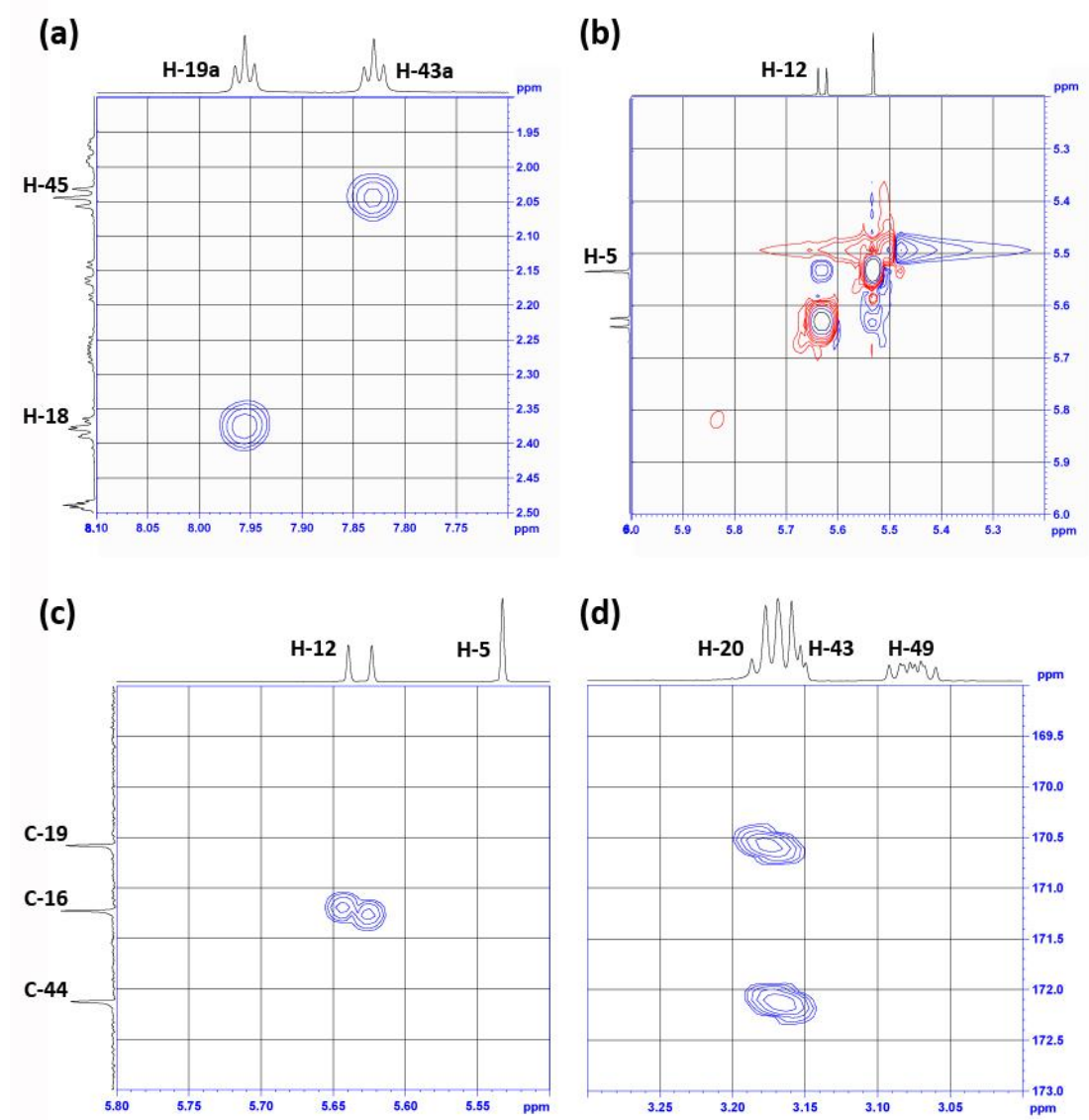


Figure 36: Selected 2D-NMR correlations of biotin-artesunate (123; 600 MHz, DMSO- d_6). (a) ROESY correlations between H-18/H-19a and H-45/H-43a. (b) ROESY correlations between H-5 and H-12. (c) HMBC coupling of H-12 and C-16. (d) HMBC long range coupling of H-20 and C-19.

Hence, we confirmed the successful synthesis of biotin-artesunate as the α -epimer, incorporating the endoperoxide bridge and all rings intact.

Table 17: Selected NMR (DMSO-*d*₆, 600 MHz) data for biotin-artesunate (123).

Pos.	δ_c	δ_H , m (J in Hz)	COSY (H no.)	1H - ^{13}C HMBC (C no.)	sel. ROESY (H no.)
1	51.1	1.16 td (11.4, 6.6)	2	2, 5, 6, 7, 9, 10, 14	
2	24.2	1.80 m, 1.31 m	1, 3	5, 6, 8, 11, 12	
3	35.9	2.17 td (14.0, 3.9), 1.98 m	2	2(w), 15	
4	103.6		-	2, 3, 5, 15	
5	90.6	5.54 s	-	1, 7, 15(w)	1, 2, 7, 8, 9, 10
6	79.9		-	1, 5, 7	
7	44.6	1.53 m	11	8(w), 9(w), 13	
8	21.0	1.62 m, 1.44 m	8, 9		
9	33.6	1.59 m, 0.93 m	8	14	
10	36.0	1.41 d (9.7)	14	14	
11	31.7	2.27 m	7, 12	8, 12, 13	
12	91.6	5.63 d (9.8)	11	7, 11, 13	7, 8, 9(w), 11, 13
13	11.7	0.75 d (7.1)	11	7, 11, 12	
14	20.1	0.88 d (6.4)	10	1, 9, 10(w)	
15	25.5	1.27 s	-	3, 4, 5	
16	171.4		-	12, 17, 18	
17	28.9	2.57 t (6.9)	18	18	
18	29.5	2.38 t (7.4)	17	17	19a
19	170.7		-	17, 18, 19a, 20	
19a	-	7.96 t (5.6)	20		
20	38.6	3.17 m	19a, 21	21	
21	69.1	3.38 t (6.1)	20	20	
22-41	69.5 - 70.3	3.48 - 3.51		22- 41	
42	69.2	3.37 t (6.0)	43		
43	38.4	3.16 m	42, 43a	42	
43a	-	7.83 t (5.6)	43		45
44	172.1		-	43, 43a, 45, 46	
45	35.1	2.05 t (7.5)		47(w), 46	43a
46	25.2	1.48 m	47	45	
47	28.2	1.28 m	46	45	
48	28.0	1.59 m, 1.45 m		49	
49	55.4	3.08 m	50	48(w), 50, 51, α -52	
50	61.1	4.11 m	49	46, α -52, 52a, 53a	
51	59.2	4.29 m	α -52, β -52	α -52, 52a, 53a	
52	39.8	2.80 dd (β , 12.5, 5.1), 2.56 d (α , 13.4)	51		
52a	-	6.35 m			
53	162.8			50, 51, 52a, 53a	
53a	-	6.41 br s			

a (w) denotes a weak correlation.

3.2.3.3.2.2 Stability and purity

For the selection experiments described in Section 4.5 it is important to ensure the stability throughout both the coating of the neutravidin-coated microtitre plates and the following biopanning procedure (6 h total). An aliquot of a freshly prepared

solution of biotin-artesunate in PBS was injected onto an analytical RP-HPLC column and the eluting peaks integrated. This analysis was repeated for eight consecutive hours and the integrals are presented as percent of the area from the initial injection at $t = 0$ h (see Figure 37). As hydrolysis is expected as the primary degradation pathway, special emphasis was given to the detection of a biotinylated succinic acid species lacking dihydroartemisinin. Therefore, a second sample of biotin-artesunate was exposed to acidic conditions ($\text{pH} = 2$), which resulted in significant formation of the hydrolysis product after only 2 h. The results are displayed as a separate trace in Figure 37. The formation of this hydrolysed species was not observed throughout the stability monitoring in PBS (9 injections from $t = 0$ to 8 h) indicating that the probe is stable under the biopanning conditions.

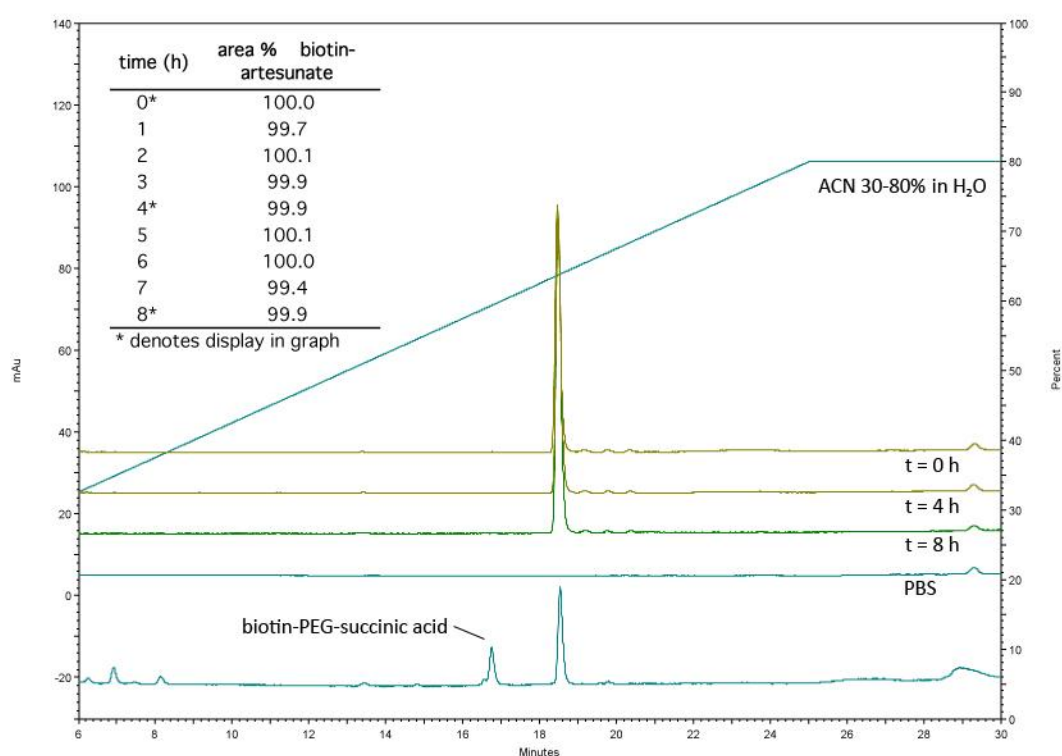


Figure 37: Stability test of biotin-artesunate in PBS over 8 h monitored via RP-HPLC (at 210 nm) in comparison to background and sample containing biotin-PEG-succinic acid as major hydrolysis product.

The preparative RP-HPLC purification of the major batch of the biotinylated probe was reinjected onto an analytical RP-HPLC column to ensure the product was absolutely pure. Figure 38 displays biotin-artesunate as a single peak, with no noticeable impurities present ($\lambda = 210$ nm).

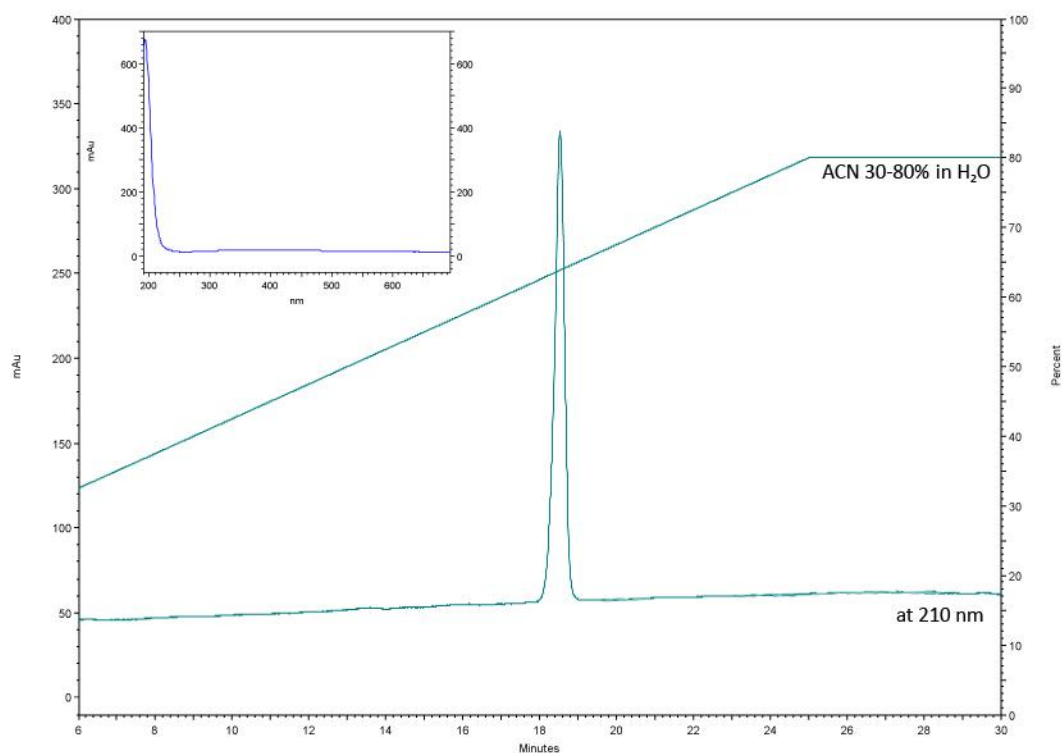


Figure 38: Single peak RP-HPLC trace and UV spectrum of biotin-artesunate (123).

Biotin-artesunate was therefore considered as pure and stable and suitable for application in the T7 phage display assay described in Section 4.5.

3.3 Summary

This chapter described the synthesis of various reagents that are required for the construction of affinity supports suitable for reverse chemical proteomics studies. Known and novel polyethyleneglycol based linkers for immobilising small molecules on polymeric supports were synthesised and characterised. Biotinylated analogues of three biologically active natural products (manzamine, daptomycin and artemisinin) were synthesised on a milligram scale for immobilisation on neutravidin-coated microtitre plates. The resulting probes are suitable for application in affinity chromatography based methods, such as in reverse chemical proteomics, to isolate protein binding partners for the bioactive compounds.

