

**UNIVERSITA' DEGLI STUDI DI NAPOLI
"FEDERICO II"**



DOTTORATO DI RICERCA IN "SCIENZA DEL FARMACO"
XXI CICLO 2005-2008

***Marine Natural Products:
from Structure Elucidation toward Sustainable
Production***

Dott.ssa Roberta Teta

Tutor
Prof. E. Fattorusso

Coordinatore
Prof.ssa M. V. D'Auria

INDICE

| | |
|---------------------------------------------------------------------------------------|-----------|
| INTRODUCTION | 5 |
| CHAPTER 1 | 11 |
| PORIFERA | 11 |
| CHAPTER 2 | 15 |
| STRUCTURAL DETERMINATION METHODS | 15 |
| 2.1 MASS SPECTROMETRY | 15 |
| 2.2 NUCLEAR MAGNETIC RESONANCE | 17 |
| 2.3 CIRCULAR DICHROISM | 19 |
| CHAPTER 3 | 23 |
| GLYCOLIPIDS | 23 |
| 3.1 GLYCOSPHINGOLIPIDS | 24 |
| 3.2 ISOLATION PROCEDURES | 25 |
| 3.3 STRUCTURE DETERMINATION OF GLYCOLIPIDS | 27 |
| CHAPTER 4 | 35 |
| TERPIOSIDE A FROM THE MARINE SPONGE <i>TERPIOS</i> SP. | 35 |
| 4.1 ISOLATION OF TERPIOSIDE | 35 |
| 4.2 STRUCTURE DETERMINATION OF TERPIOSIDE | 36 |
| 4.3 EXPERIMENTAL SECTION | 48 |
| CHAPTER 5 | 55 |
| PLANAR STRUCTURE ELUCIDATION OF TERPIOSIDE B FROM <i>TERPIOS</i> SP. | 55 |
| 5.1 ISOLATION OF TERPIOSIDE B | 55 |
| 5.2 STRUCTURE DETERMINATION | 56 |
| 5.3 EXPERIMENTAL SECTION | 63 |
| CHAPTER 6 | 67 |
| AMPHICERAMIDES A AND B FROM THE SPONGE <i>AMPHIMEDON COMPRESSA</i> | 67 |

| | |
|-------------------------------------------------------------------------------------|------------|
| 6.1 ISOLATION OF AMPHICERAMIDES | 69 |
| 6.2 STRUCTURE DETERMINATION OF AMPHICERAMIDE A | 70 |
| 6.3 STRUCTURE DETERMINATION OF AMPHICERAMIDE B..... | 79 |
| 6.4 STRUCTURE DETERMINATION OF 3A, 4A, AND 5A | 87 |
| 6.5 EXPERIMENTAL SECTION | 88 |
| CHAPTER 7..... | 101 |
| SPONGE-MICROBE ASSOCIATIONS..... | 101 |
| 7.1 CELLULAR LOCALIZATION OF METABOLITES FROM <i>P.SIMPLEX</i> | 102 |
| CHAPTER 8..... | 107 |
| PLAKORTIN | 107 |
| 8.1 THE BIOGENESIS OF PLAKORTIN | 111 |
| 8.2 POLYKETIDE SYNTHASES | 112 |
| 8.3 AN HYPOTHESIS ON THE BIOGENETIC PATHWAY TO PLAKORTIN..... | 115 |
| 8.4 STRATEGY FOR THE IDENTIFICATION OF PLAKORTIN BIOSYNTHETIC GENE CLUSTER | 119 |
| CONSTRUCTION OF A METAGENOMIC DNA LIBRARY | 123 |
| 9.1 DNA EXTRACTION | 123 |
| 9.2 CONSTRUCTION OF THE GENE LIBRARY | 124 |
| 9.3 PCR SCREENING | 128 |
| 9.4 EXPERIMENTAL SECTION | 133 |

INTRODUCTION

Seas and oceans represent an endless source of compounds with interesting pharmacological activity. The remarkable biodiversity within the marine environment is proven by the coexistence of an enormous number of species, which interact with each other and with the environment in different ways. In this process, chemistry plays an essential role; all the marine organisms synthesize or obtain from symbiotic microorganisms “secondary metabolites”, molecules with even very complex structures, which are characteristic of specific species or specific taxa. Many secondary metabolites are involved in the interactions within or between species. Considering the number of different species and their immeasurable possible interactions, it’s not a surprise that a wide variety of secondary metabolites are produced to preserve and improve the life of the producing organisms.

It’s very well known that plants are particularly rich in secondary metabolites, and their properties have been long exploited for the treatment of diseases. In the last decades, the progress in some technologies and the wider availability of diving equipment made it possible to extend this research to marine organisms.

A larger number of organisms and less genetic homogeneity between separate populations of the same species exist in the marine compared to the terrestrial environment. This results in a higher structural diversity at the molecular level. Among marine organisms, the chance of finding bioactive

compounds is remarkably higher in some invertebrates, like corals, tunicates, and sponges. Many of these compounds are involved in their chemical defence, which is essential for the survival of sessile organism, often lacking any physical defence from their predators.

Even when showing interesting and specific pharmacological activities, natural compounds isolated from marine organisms rarely raise the interest of pharmaceuticals companies, because they are difficult to obtain in sufficient amounts for clinical use. Total synthesis is a possible way to overcome this problem, but this is usually prevented by the complex structure of most natural products, often including many chiral centers, which makes this option economically unfeasible.

In this case, tapping the natural source is the best option. The problem is to conciliate development dynamics necessary for man and protection of the marine environment and biodiversity. Oceans are showing increasing signs of overexploitation and degradation, resulting in loss of both productivity and biodiversity. In this context, a massive collection of marine organisms producing compounds of industrial interest appears unrealistic. Alternative approaches have therefore been proposed, such as the cultivation of the organism of interest under controlled conditions (aquaculture), and the laboratory production of metabolites in bioreactors from cell cultures.

A more recent approach, which in the light of the explosive development of molecular biotechnologies is rapidly gaining a prominent position worldwide, is the study of the metabolic processes leading to the

synthesis of secondary metabolites. The understanding of these processes at the genetic level is paving the way to the possibility to control them, and eventually to produce the desired metabolites in large scale using a "green technology" such as gene cloning.

The research work described in this PhD Thesis was performed at the Dipartimento di Chimica delle Sostanze Naturali of Università di Napoli "Federico II", and was directed to different aspects of the research on marine natural products. On one hand, part of the research was directed to the "core activity" of natural product chemistry, i.e. isolation and structure elucidation of new compounds from marine organisms and their symbiotic microorganisms. In this respect, particular attention was paid to glycolipids from sponges, and several structurally interesting new glycosphingolipids were isolated from the marine sponges *Terpios* sp. and *Amphimedon compressa*.

On the other hand, a new research line was started up that focused on the study of the biosynthetic pathway of plakortin, polyketide peroxide with interesting antimalarial properties which is present in the marine sponge *Plakortis simplex* and is probably produced by an unculturable symbiotic microorganism of the sponge. The aim of this work was to identify and isolate the biosynthetic gene cluster for this compound, with the future objective of expressing the pathway heterologously to produce plakortin by fermentation.

Therefore, the results of research work sketched above have been organized into two sections:

- ✓ Isolation and structure elucidation of glycolipids from the marine sponges *Amphimedon compressa* and *Terpios* sp.
- ✓ Study of the biosynthetic pathway of plakortin from the marine sponge *Plakortis simplex*.

PART I

ISOLATION AND STRUCTURAL DETERMINATION OF GLYCOLIPIDS FROM MARINE SPONGES

Chapter 1

Porifera

Porifera commonly known as sponges are the simplest and ancient phylum among metazoans. They are ubiquitous animals and live permanently attached to a location in the water. There are from 5,000 to 10,000 known species of sponges. Most sponges live in salt water; only about 150 species live in fresh water.

Their body consists of specialized tube-like cells called porocytes which control channels leading to the interior of mesohyl (ostia).

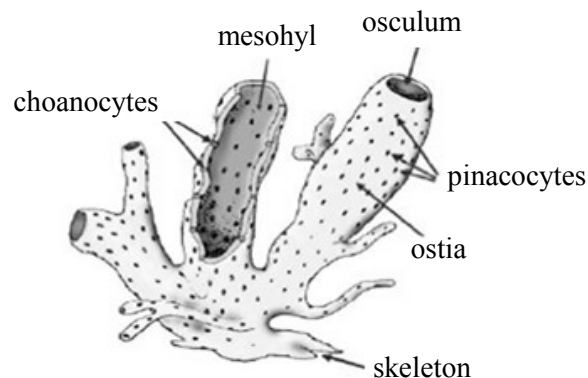


Figure 1: A sponge body structure

Mesohyl is the gelatinous matrix within the sponge made of collagen and covered by choanocytes, cylindrical flagellated cells. The outer layer is formed by pinacocytes, plate-like cells which digest food particles too large to enter the ostia. The body is reinforced by the skeleton, collagen fibers and spicules: Calcareous sponges produce spicules made of silica calcium carbonate while the larger class (90%) of Demospongiae produce a special

form of collagen called spongin besides silica spicules; glass sponges, common in polar water and in the depths of temperate and tropical seas contain syntria in their structure which enable them to extract food from these resource-poor waters with the minimum of effort. Sponges are filters feeders; they obtain nourishment, tiny and floating organic particles, plankton and oxygen from flowing water that they filter through their body.

Water flows into a sponge through porocytes and flows out of it through large opening called oscula. The flowing water also carries out waste products.

The simplest body structure in sponges is a tube or vase shape known as asconoid; in siconoids the body wall is pleated and the pumping capacity is increased; leuconoids contain a network of chambers lined with choanocytes and connected to each other and to the water intakes and outlet by tubes.

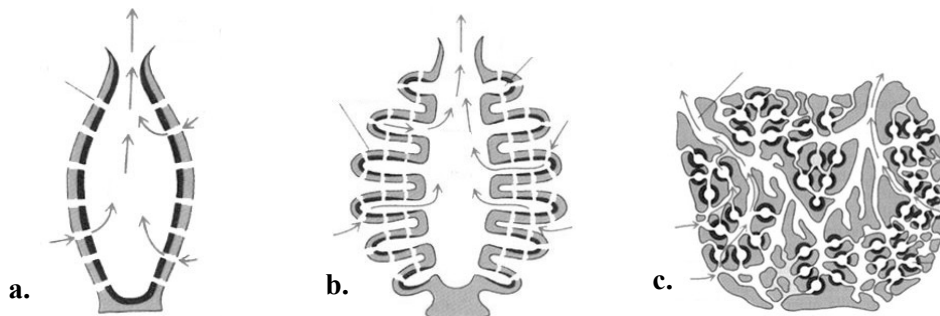


Figure 2: Sponge classification: **a.** asconoid; **b.** siconoids; **c.** leuconoids

Most sponges are hermaphrodites (each adult can act as either the female or the male in reproduction). Fertilization is internal in most species; some released sperm randomly float to another sponge with the water current. If a sperm is caught by another sponge's collar cells (choanocytes), fertilization of an egg by the travelling sperm takes place inside the sponge.

The resulting tiny larva is released and is free-swimming; it uses tiny cilia (hairs) to propel itself through the water. The larva eventually settles on the sea floor, becomes sessile and grows into an adult.

Some sponges also reproduce asexually; fragments of their body (buds) are broken off by water currents and carried to another location, where the sponge will grow into a clone of the parent sponge (its DNA is identical to the parent's DNA).

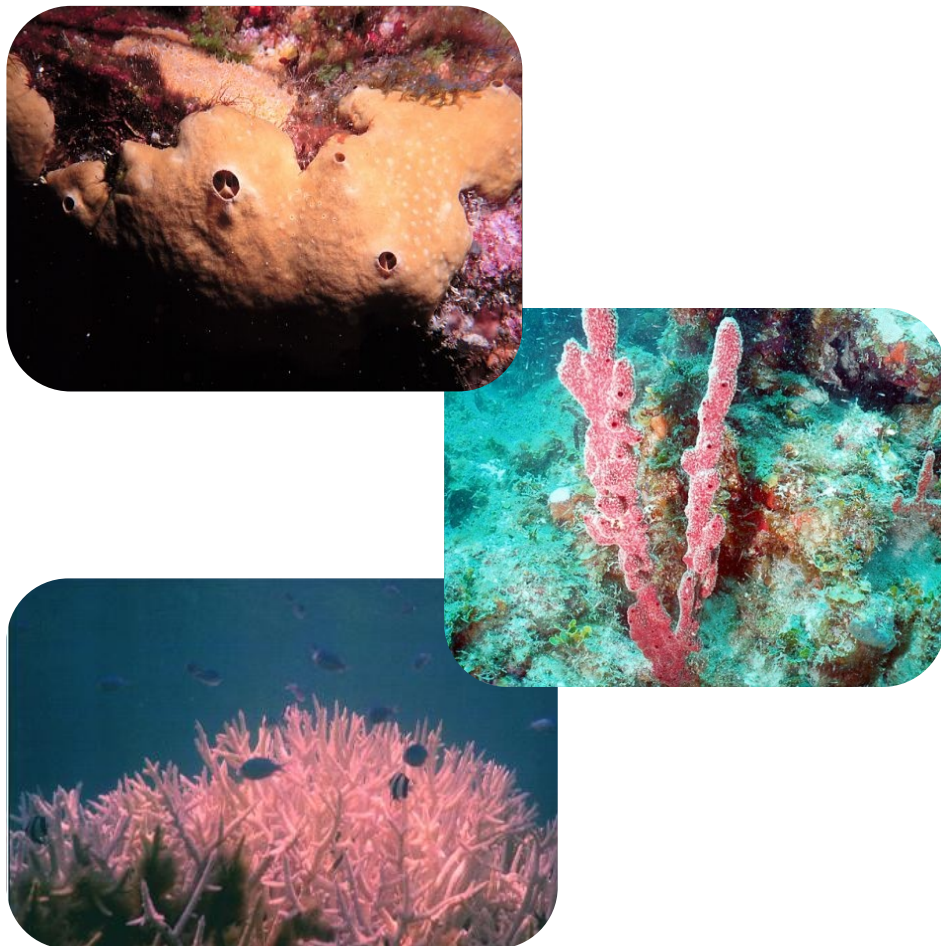


Figure 3: Sponges discussed in this PhD Thesis: *Plakortis simplex*, *Amphimedon compressa* and *Terpios* sp.

Chapter 2

Structural Determination Methods

Until a few decades ago structural determination of new organic compounds was only pursued through the use of chemical techniques (degradation and interconversion of functional groups). The development of spectroscopic techniques dramatically changed this approach. Today, it is possible to determine complex organic structures completely in a non-destructive way, with submilligram samples. Structural determination described in this thesis is largely based on spectroscopic techniques, mostly mass spectrometry (MS) and nuclear magnetic resonance (NMR), even if in some cases degradation methods were used.

2.1 Mass Spectrometry

The first step in the study of a new bioactive compound is the determination of the molecular formula through the high resolution mass spectrometry.

Mass spectrometry is an analytical technique that is used to determine the molecular mass of a compound on the basis of the mass-to-charge ratio (m/z ratio) of ions produced from the molecules. A very accurate measurement of the molecular mass (high resolution mass spectrometry) can also provide the molecular formula of the molecule under study.

The *source* is the component of the mass spectrometer which produces ions from the molecule, while the *analyzer* measures the mass-to-charge

ratio of the ions. There are many different types of sources, as well as of analyzers.

During or after ionization the molecule may fragment, and the mass of the fragments provides information on the structure of the molecule under examination. If the ions do not fragment by themselves, they may be induced to fragment by letting them collide with gas molecules. In this case, a second analyzer is used to measure the mass of the fragments. This is known as tandem mass spectrometry or MS/MS.

Most of compounds described in the following sections were analyzed by ESI mass spectrometry. The ESI source is a widely used technique for polar and/or charged macromolecules. The sample is dissolved in a volatile solvent like H₂O, MeOH, and CH₃CN; volatile acids, bases or buffers are often added to the solution. This solution is pumped through a charged metal capillary and, in coming out of the capillary, forms a spray. Because of the electric potential of the capillary, each droplet of the spray carries an excess positive or negative charge, and this causes extensive protonation or deprotonation of the molecules of the sample, which become ions. An uncharged carrier gas such as nitrogen is used to help the liquid to nebulize and the neutral solvent in the droplets to evaporate.

As the solvent evaporates, the ionized analyte molecules become closer and closer, until they can escape from the droplet by electrostatic repulsion. For molecules with a high molecular weight, the ions may take more than one proton (up to some tens), and therefore may have multiple charge. Formation of multiply charged ions allows the analysis of high molecular

weight molecules such as proteins, because it reduces the m/z ratio of the ions, which is therefore easier to measure.

2.2 Nuclear Magnetic Resonance

The most important spectroscopic technique used for structure elucidation of the isolated secondary metabolites was nuclear magnetic resonance (NMR). In addition to standard ^1H and ^{13}C NMR spectra, a large use of 2D NMR experiments was made. They are superior to their 1D NMR counterparts both for the shorter acquisition times, and for the easier assignment of nuclei resonating in crowded regions of the spectra (signal overlapping is much less likely in two dimensions than in one).

The COSY (Correlation SpectroscopY) experiment is one of the simplest and yet most useful 2D NMR experiment. It allows determination of the connectivity of a molecule by identifying which protons are scalarly coupled. In spite of the many modifications which have been proposed along the years, the very basic sequence composed of two $\pi/2$ pulses separated by the evolution period t_1 is still the best choice if one is simply dealing with the presence or the absence of a given coupling, but not with the value of the relevant coupling constant.

The TOCSY (Total Correlation SpectroscopY) experiment is a 2D NMR experiment very useful in the analysis of molecules composed of many separate spin systems, such as oligosaccharides or peptides. The TOCSY spectrum shows correlation peaks between nuclei that may be not directly coupled, but are still within the same spin system. The appearance

of a TOCSY spectrum resembles in all aspects a COSY; the difference is that the cross peaks in a COSY result from coupled spins, whereas in the TOCSY spectra they arise from coherence transfer through a chain of spin-spin couplings, and therefore any pair of protons within a spin system may give rise to a peak. The range of the coherence transfer (i.e. through how many couplings the coherence may be transferred) increases with increasing mixing times (Δ), but a mixing time too long may reduce sensitivity.

The HSQC (Heteronuclear Single Quantum Correlation) and HMQC (Heteronuclear Multiple Quantum Correlation) experiment are 2D NMR heteronuclear correlation experiments, in which only one-bond proton-carbon couplings ($^1J_{\text{CH}}$) are observed. In principle, the HSQC experiment is superior to HMQC in terms of selectivity and additionally allows DEPT-style spectral editing. However, the sequence is longer and contains a larger number of π pulses, and is therefore much more sensitive to instrumental imperfections than HMQC.

The HMBC (Heteronuclear Multiple Bond Correlation) experiment is a heteronuclear two- and three-bond ^1H - ^{13}C correlation experiment; its sequence is less efficient than HSQC because the involved $^{2,3}J_{\text{CH}}$ couplings are smaller (3-10Hz). Moreover, while $^1J_{\text{CH}}$ are all quite close to each other, $^{2,3}J_{\text{CH}}$ can be very different, making impossible to optimize the experiment for all couplings. As a result, in most HMBC spectra not all of the correlation peaks which could be expected from the structure of the molecule are present.

The ROESY (Rotating-frame Overhauser Spectroscopy) experiment is a chemical shift homonuclear correlation which can detect ROEs (Rotating-frame Overhauser Effect). ROE is similar to NOE, being related to dipolar coupling between nuclei, and depending on the geometric distance between the nuclei. While NOE is positive for small molecules and negative for macromolecules, ROE is always positive. Therefore, the ROESY experiment is particularly useful for mid-size molecules, which would show a NOE close to zero. The ROESY sequence is similar to the TOCSY sequence, and unwanted TOCSY correlation peaks may be present in the ROESY spectra. Fortunately, these unwanted peaks can be easily recognized, because their phase is opposite compared to ROESY correlation peaks. It is important to acquire ROESY spectrum in phase-sensitive mode for a correct interpretation of the spectrum.

2.3 Circular Dichroism

Circular dichroic (CD) spectroscopy of optically active compounds is a powerful method for studying the three-dimensional structure of organic molecules, and can provide information on absolute configurations, conformations, reaction mechanisms, etc.

The CD spectroscopy takes advantage of the different absorption shown by chiral compounds of left and right circularly polarized UV/Vis light. In circularly polarized light, the electric field vector rotates about its propagation direction forming a helix in the space while propagating. This

helix can be left-handed or right-handed, hence the names left and right circularly polarized light.

At a given wavelength, circular dichroism of a substance is the difference between absorbance of left circularly polarized and right circularly polarized light:

$$\Delta A = A_L - A_R$$

Since circular dichroism uses asymmetric electromagnetic radiations, it can distinguish between enantiomers. Two enantiomers have the same CD spectra, but with reversed sign.

Of course, in order to show a *differential* absorbance, the molecules needs to absorb the UV/Vis light, and therefore must possess at least one chromophore. If the molecule does not have a chromophore, this can be introduced using a derivatization reaction. This is why methyl glycosides are benzoylated (see *Chapter 3*) to determine their absolute configuration.

One of the most important methods to establish the absolute configuration of a molecule is the exciton chirality method. This method is based on the interaction between two chromophores. When two or more strongly absorbing chromophores are located nearby in space and constitute a chiral system, their electric transition moments interact spatially (exciton coupling) and generate a circular dichroism. Because the theoretical basis of exciton coupling are well understood, it is possible to correlate the CD spectrum of an exciton-coupled chromophore system with the spatial orientation of the chromophores, which in turn can be related to the absolute configuration of the molecule. It is important to point out that, unlike for

example optical rotation, the exciton chirality method does not require any reference compound to provide the absolute configuration of the molecule under study (if its conformation is known).

Chapter 3

Glycolipids

Glycolipids represent a class of molecules of widespread diffusion in the animal and vegetable kingdoms; because of their chemical and biochemical diversity and of their biological role, they appear a very interesting class of compounds. Glycolipids are ubiquitous components of the cells plasmatic membranes, where they play different and significant functions, such as cellular differentiation and signal transduction modulation, as well as adhesion and cell-cell recognition.

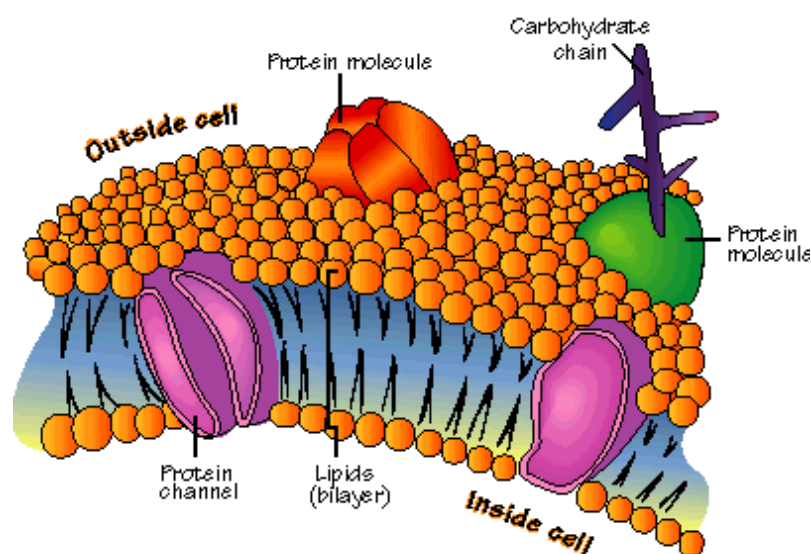


Figure 4: Membrane phospholipid bilayer organization

Many studies on glycolipids demonstrated their activity on the immune system and their influence on angiogenesis and tumor cells proliferation, and therefore they are used in therapy against several diseases. Glycolipids are amphiphilic compounds: the polar portion of the molecule is composed of an oligosaccharide chain, which is the main responsible for the biological

role of glycolipids. It is commonly acknowledged that the bioactivity depends essentially on the nature of the sugar head and the adjacent functionalized part of the molecule. In contrast, the alkyl chains serve to position the glycolipid in the membrane and affect its fluidity.

Depending on their lipophilic portion, they are currently divided into two groups: glycoacylglycerolipids (GALs) and glycosphingolipids (GSLs). There is a third important group, comprising glycolipids whose lipid portion is derived from mevalonate, such as steroidal and terpenic glycosides. The occurrence of polyisoprenoidic glycolipids is generally confined to species belonging to a few taxa, where they frequently play peculiar biological functions; as for marine organisms, they are mostly present in invertebrates belonging to phylum Echinodermata.

3.1 Glycosphingolipids

The structure of a representative GSL is depicted in Figure 5. A carbohydrate chain and a fatty acyl group are linked to a long-chain aminoalcohol, called sphingoid base or long-chain base (LCB). The fatty acyl chain is amide-linked to the LCB, and together they make up what is called a ceramide; the monosaccharide or oligosaccharide group is bound to the primary alcoholic function of the ceramide.

Sphingosine is the LCB most commonly found in higher animals, so that LCBs are often referred to as sphingosines. In plant glycolipids, the trihydroxylated LCB phytosphingosine is frequently found; the name phytosphingosine is also often used more generally, meaning any

trihydroxylated LCB. Since an increasing number of different sphingoid bases are being described in the literature, a semisystematic nomenclature for LCBs has been proposed, which is based on sphinganine (2-amino-1,3-octadecandiol).¹

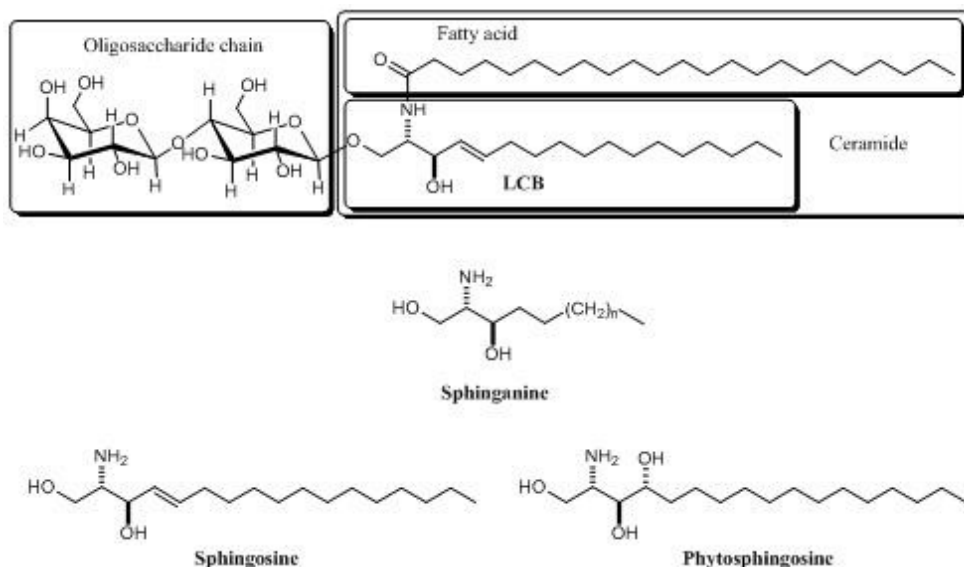


Figure 5: A generic glycosphingolipid

3.2 Isolation Procedures

A matter which appears to be worthy of a preliminary discussion is the level of purity of a mixture of glycolipids. Virtually all natural glycolipids occur as mixtures of homologues, differing in the length and in the branching of the alkyl chains of the lipid portion of the molecule. Often, structural studies are performed on the mixtures of homologues, while the nature of the fatty acids and/or the sphingoid bases is subsequently established by chemical degradation followed by GC/MS analysis. The difficulty (or more often the impossibility) to obtain a chemically homogeneous glycolipid is evident, if one considers that, for example, a

glycosphingolipid which appear pure by normal phase HPLC often contains 7-8 different sphingoid bases and fatty acids, and this corresponds to over 50 different compounds. Attempts of chromatographic separation of such complex mixtures are reported in the literature,^{2,3} but they afforded only a small number of homogenous fractions, the other ones being still mixtures which were characterized for the major component only, or not at all. Generally, a mixture of homologue glycolipids can be regarded to be a "single" compound, the purity being referred to the biological role of the material that for glycolipids depends on the saccharidic chain.

The most recent procedures for the isolation of neutral marine glycolipids take advantage of their amphiphilic character; basically, the homogenate of the tissue is extracted with a mixture of methanol and chloroform. The combined extracts are usually partitioned between water and butanol, and the glycolipid material can be recovered in the butanol phase. The crude extract obtained after the evaporation of the phase is successively subjected to a reversed-phase column chromatography as the first step of the purification of the organic phase: glycolipids, in spite of their remarkable polarity, are conspicuously retained by the stationary phase and are eluted together with quite non-polar substances, which can be easily removed through a successive adsorption chromatography on SiO₂. A relatively pure glycolipids fraction is therefore obtained.

The next steps of the separation procedure must be optimized for each single compound. In many cases, the whole glycolipids mixture is acetylated, because acetylated glycolipids are easily separated on normal

phase HPLC and analyzed with NMR. Moreover, the acetylation reaction with deuterated acetic anhydride $(\text{CD}_3\text{CO})_2\text{O}$ allows one to distinguish between the natural acetyl groups and those introduced by the acetylation reaction.

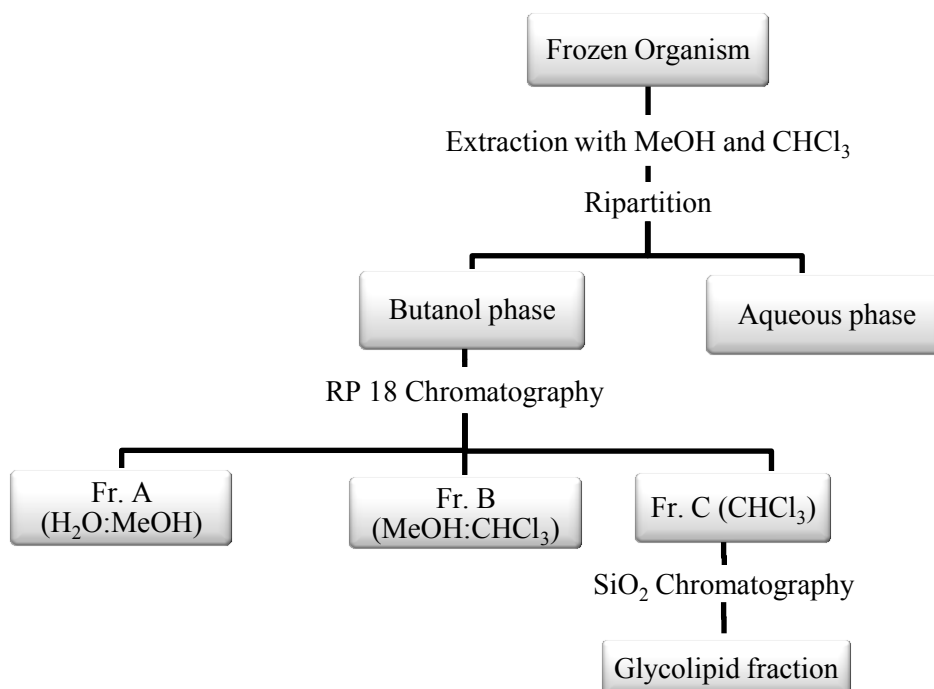


Figure 6: Glycolipids isolation scheme

3.3 Structure Determination of Glycolipids

The first reports on complete structural assignments of glycolipids from marine sources are relatively recent, dating from the seventies. From the beginning, spectroscopic data remarkably assisted chemical analysis to clarify the structures of these metabolites. Of course, the importance of the physical methods has dramatically increased during the last ten years, when the modern two-dimensional techniques improved the NMR spectroscopy as a precious tool of chemical analysis, able to provide decisive information

for the elucidation of complex structures, including the stereochemical details, with just a few milligrams of a reasonably pure product.

According to our procedures, the starting point of the structural analysis of a natural product is the determination of molecular formula. This is currently achieved by interpretation of mass spectral data; the successive step involves the assignment of the structure to the lipid part of the molecule and to the sugar moiety, which are generally investigated in separate experiments and through quite different analytical techniques.

Structure of the Sugar Portion

For the analysis of the structure of the polar portion of glycolipids, two-dimensional homonuclear (^1H - ^1H) and heteronuclear (^1H - ^{13}C) correlation experiments are mainly performed. The NMR studies are carried out using the peracetylated glycolipid because of the better proton dispersion in the ^1H -NMR spectrum of this derivative, and the possibility of discriminating between ether and ester oxymethine proton resonances on the basis of their different chemical shift ranges (δ 3.5-4.5 and 4.7-5.7, respectively).

The ^1H - ^{13}C heteronuclear correlation (HSQC) spectrum allows to easily identify anomeric carbons and the corresponding anomeric protons and so the number of saccharidic units, because in peracetylated glycolipids anomeric carbons have characteristic resonances compared to the other oxygenated sugars carbons, being they linked to two carbon atoms instead of one. Methylene diastereotopic protons can also be identified because they are coupled with the same carbon. Starting from the anomeric proton all the

protons in the corresponding spin system are assigned with a TOCSY spectrum whereas a subsequent COSY spectrum is used to find out the proton sequence in each spin system. The chemical shift values of the protons contribute to discriminate between acetylated hydroxyl groups and those involved in glycosidic bonds, providing useful data to determine glycosylation sites and furanose or pyranose structure of each monosaccharide of the carbohydrate chain.

The nature of the sugar units and the configuration of the glycosidic bonds are ascertained by ^1H - ^1H coupling constants analysis. When coupling constant analysis is not sufficient for an unambiguous structural assignment, additional information can be provided by a ROESY spectrum, evidencing inter-proton spatial proximities.⁴ Linkages between couples of sugars and between the carbohydrate chain to the aglycon are evidenced by a long-range ^1H - ^{13}C shift correlation (HMBC) experiment,⁵ which shows connectivity between proton and carbon atoms separated by two or three σ bonds, or by the ROESY spectrum.

Structure of the Ceramide Portion

Also in this case, NMR experiments are useful to clarify the structure and the relative stereochemistry of the polar portion of the sphingosine as well as to identify alkyl chain branching. It is to be noted, however, that as said above the material to be investigated is often a mixture of homologues differing in the lipid part of the molecule, and this requires the use of appropriate analytical procedures.

Sphingoid bases are normally removed from the molecule by treatment with acidic methanol, then purified through a SiO₂ column chromatography, and identified and quantized as free bases or as appropriate derivatives by GC analysis.⁶ If reference compounds are not available, the identification can be accomplished by periodate/permanganate oxidation⁷ followed by methylation⁸ of the sphingosine homologues, and gas-chromatographic identification of the resulting fatty acid methyl esters.

Absolute configuration

Absolute configuration can be assigned to sphingosines, α -hydroxyacids, and methyl glycosides through Circular Dichroism (CD) measurements. Methanolysis of GSLs provides a mixture of fatty acid methyl esters and/or α -hydroxyacid methyl esters, sphingosines and methyl glycosides. The mixture is benzoylated and separated on normal phase HPLC, so that individual compounds can be analyzed separately. Their ¹H NMR and CD spectra are recorded and compared with those of authentic sample of known stereochemistry.

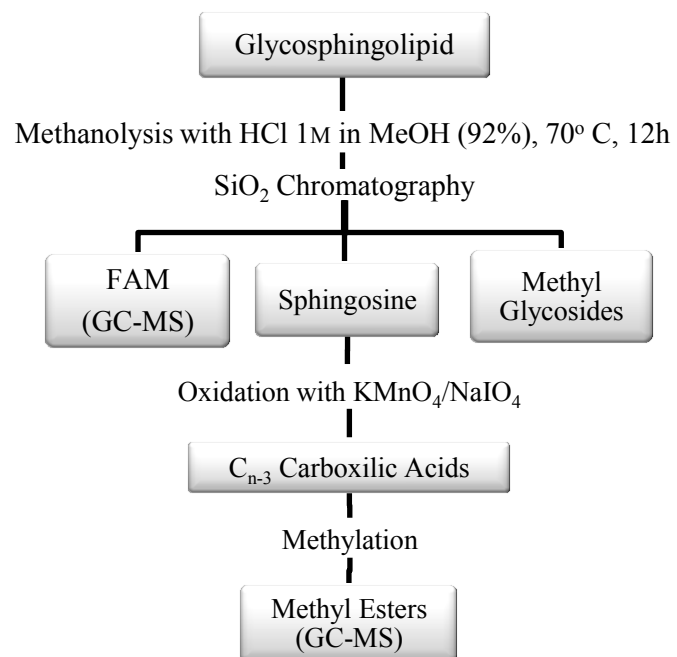


Figure 7: Microscale degradation scheme

References

1. Sweeley, C.C: Sphingolipids. In: New Comprehensive Biochemistry (Vance, D.E., and J. Vance eds.) Vol. 20, Biochemistry of Lipids, Lipoproteins and Membranes, p. 327. Amsterdam: Elsevier. 1991.
2. Higuchi, R., T. Natori, and T. Komori: Glycosphingolipids from the Starfish *Asterina pectinifera*. Isolation and Characterization of Acanthacerebroside B and Structure Elucidation of Related, Nearly Homogeneous Cerebrosides. *Liebigs Ann. Chem.*, 51 (1990).
3. Higuchi, R., M. Inagaki, K. Togawa, T. Miyamoto, and T. Komori: Constituents of Holothuriodeae. IV. CE-2b, CE-2c and CE-2d, Three New Sphingosine-type Glucocerebrosides from the Sea Cucumber *Cucumaria echinata*. *Liebigs Ann. Chem.*, 79 (1994).
4. Costantino, V., E. Fattorusso, A. Mangoni, M. Akinin, and E.M. Gaydou: Axyceramide A and B, Two Novel Tri-alpha-glycosylceramides from the Marine Sponge *Axinella* sp. *Liebigs Ann. Chem.*, 181 (1994).
5. Cafieri, F., E. Fattorusso, Y. Mahajnah, and A. Mangoni: Longiside, a Novel Digalactosylceramide from the Caribbean Sponge *Agelas longissima*. *Liebigs Ann. Chem.*, 1187 (1994).
6. Higuchi, R., S. Matsumoto, M. Fujita, T. Komori, and T. Sasaki: Glycosphingolipids from the Starfish *Astropecten latespinosus*, 2. Structure of Two New Ganglioside Molecular Species and Biological Activity of the Ganglioside. *Liebigs Ann.*, 545 (1995).

7. Kubo, H., G.J. Jiang, A. Irie, M. Morita, T. Matsubara, and M. Hoshi: A Novel Ceramide Trihexoside from the Eggs of the Sea Urchin, *Hemicentrotus pulcherrimus*. *J. Biochem.*, **111**, 726 (1992).
8. Costantino, V., E. Fattorusso, and A. Mangoni: Glycolipids from Sponges. III. Glycosyl Ceramide Composition of the Marine Sponge *Agelas conifera*. *Liebigs Ann.*, 2133 (1995)

Chapter 4

Terpioside A from the Marine Sponge *Terpios* sp.

The new and interesting GSL terpioside A (**1a**)¹ was isolated from the Caribbean sponge *Terpios* sp. It is the first example of a natural diglycosylceramide having a L-fucofuranose unit in the sugar chain of the molecule.

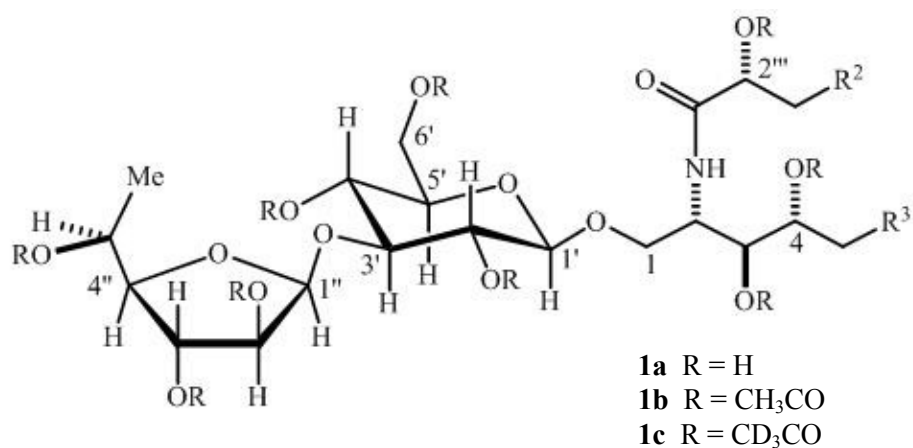


Figure 8: Structure of terpioside from the sponge *Terpios* sp.

4.1 Isolation of Terpioside

Samples of *Terpios* sp. were collected along the coast of Key Largo (Florida) and kept frozen until extraction. The specimens were extracted three times with methanol and then with chloroform. The methanol extract was partitioned between water and BuOH, in order to separate lipophilic metabolites as glycolipids from hydrophilic molecules as proteins and sugars. The organic phase was dried, and according to our standard procedure a glycolipid fraction was obtained by subsequent reversed phase and normal phase column chromatography. The glycolipid fraction was

acetylated, and the peracetylated glycolipids were subjected to repeated HPLC on SiO₂ columns to give 62.6 mg of compound **1b**. This was deacetylated with MeOH/MeONa to yield the natural glycolipid **1a**.

4.2 Structure Determination of Terpioside

Extensive NMR techniques, microscale degradation and mass spectrometric analysis were used to establish the structure of terpioside A (**1a**).

Natural terpioside **1a** was subjected to mass spectrometric experiments. The ESI mass spectrum of compound **1a** showed a series of sodiated pseudomolecular ion peaks at $m/z = 986, 1000, 1014, 1028, 1042,$ and $1056,$ which suggested a series of homologues differing in the size of the alkyl chains. A high-resolution measurement performed on the most abundant ion at $m/z = 1028.7567$ indicated the molecular formula C₅₅H₁₀₇NO₁₄ for the dominant homologue.

A preliminary analysis of the ¹H NMR spectrum of the peracetyl derivative (**1b**) showed it to be a glycosphingolipid, as suggested by:

- an strong aliphatic chain signal at $\delta 1.25$
- an amide NH doublet at $\delta 6.76$
- several signals due to oxymethine and oxymethylene groups between $\delta 5.4$ e 3.4 .

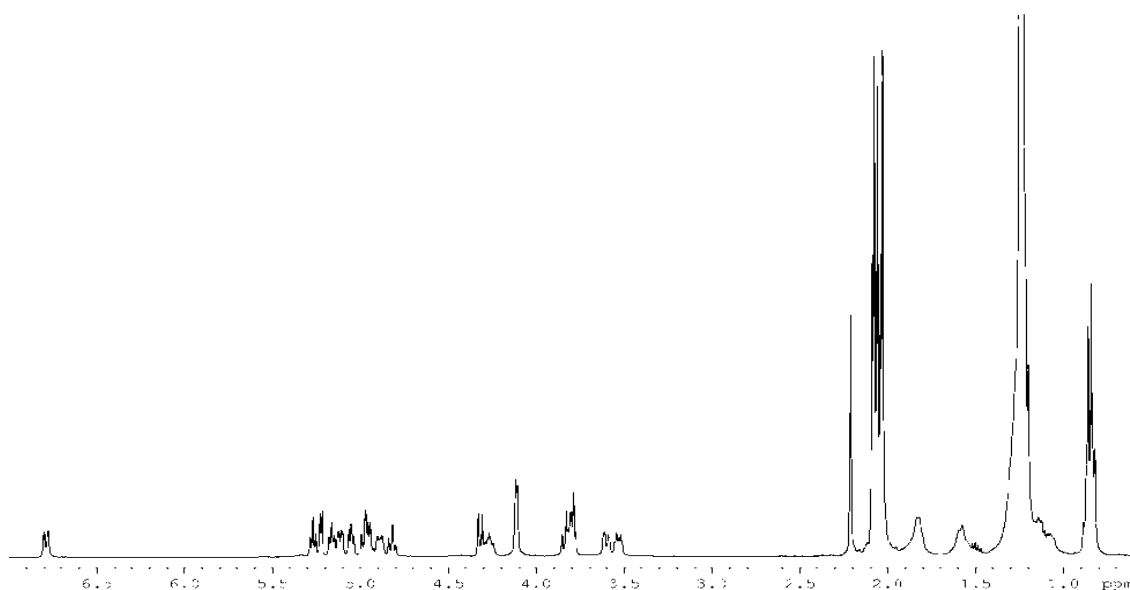


Figure 9: ^1H NMR spectrum of peracetyl terpioside (**1b**) in CDCl_3 at 400MHz

The ^1H NMR spectrum also showed in the methyl region a triplet at δ 0.87 (ethyl terminus) and a doublet at δ 0.85 ppm (isopropyl terminus), the intensities of which were not in an integral ratio relative to those of other signals in the spectrum. This showed that the alkyl chains differ not only in the length, but also in the branching of the alkyl chains.

Planar Structure

More detailed analysis of one- and two-dimensional NMR spectra showed that the ceramide portion of the molecule is composed of a 3,4-dihydroxysphinganine and an α -hydroxy fatty acid residue. Starting from the amide NH doublet at δ 6.78 (2-NH) all the protons of the polar part of the sphinganine from H₂-1 up to H₂-6 were assigned by analysis of the COSY spectrum (**Table 1**). The α -hydroxy substitution of the fatty acid residue was revealed by the absence in the ^1H NMR spectrum of **1b** of the

characteristic triplet at δ 2.3 for the fatty acid α -protons, whereas the spectrum displayed an acetoxymethine proton resonance at δ 5.17 (2^{III} -H) coupled with a methylene resonance at δ 1.83 ppm (3^{III} -H), in turn coupled with proton resonances in the broad signal, relative to alkyl chain protons, at δ 1.24.

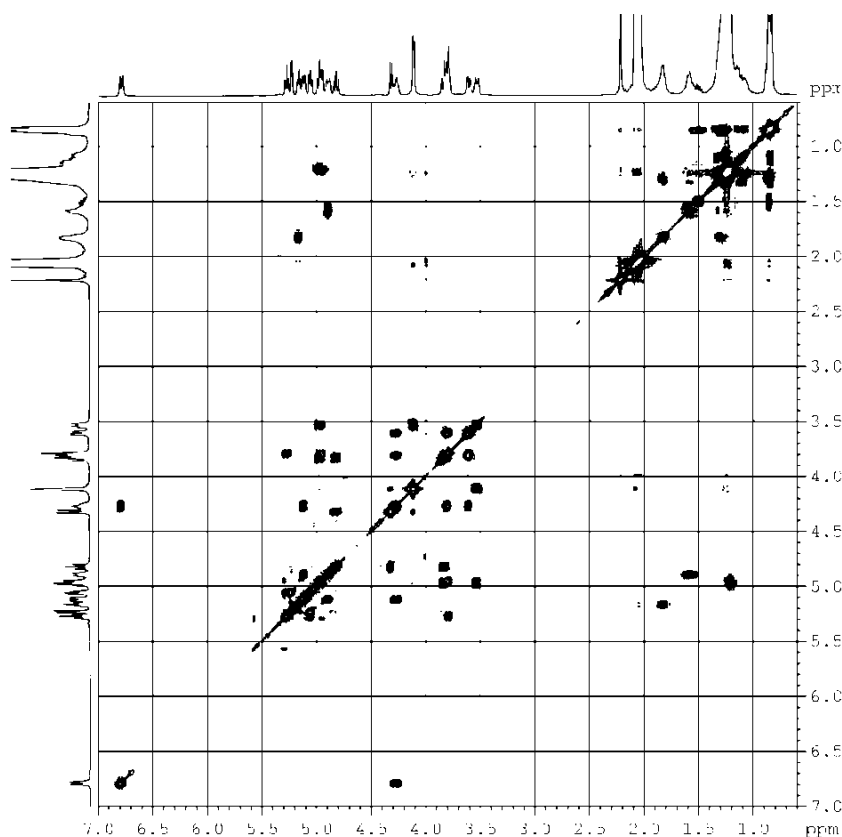


Figure 10: 2D NMR COSY spectrum of the compound **1b** in CDCl_3 at 400MHz

Evidence that the sphingosine was linked to the fatty acid was provided by the ROESY spectrum. In the ROESY spectrum of compound **1b** the oxymethine proton at δ 5.17 (2^{III} -H) showed a correlation peak with the NH doublet at δ 6.78. This demonstrated the amide bond between the α -hydroxy fatty acid carbonyl and the sphingosine and completed the planar structure determination of the ceramide portion of the molecule.

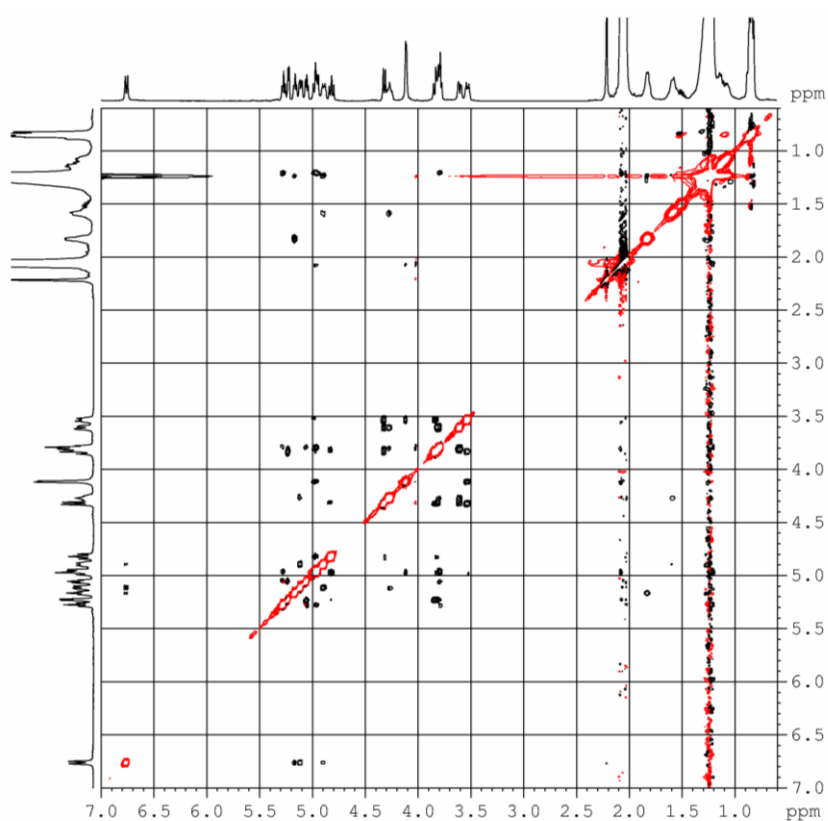


Figure 11: 2D NMR ROESY spectrum of terpioside (**1b**) in CDCl_3 at 500MHz

The two sugar units were revealed by the presence of two anomeric carbons in the ^{13}C NMR spectrum at δ 100.1 (C-1^{I}) and 100.8 (C-1^{II}). Anomeric carbons have characteristic resonances around δ 100, instead of δ 60-75 as the other oxygenated sugars carbons because they are linked to two oxygen atoms. They were correlated through the HSQC spectrum with the corresponding anomeric protons, resonating as doublets at δ 4.32 (1^{I}-H) and 5.23 (1^{II}-H).

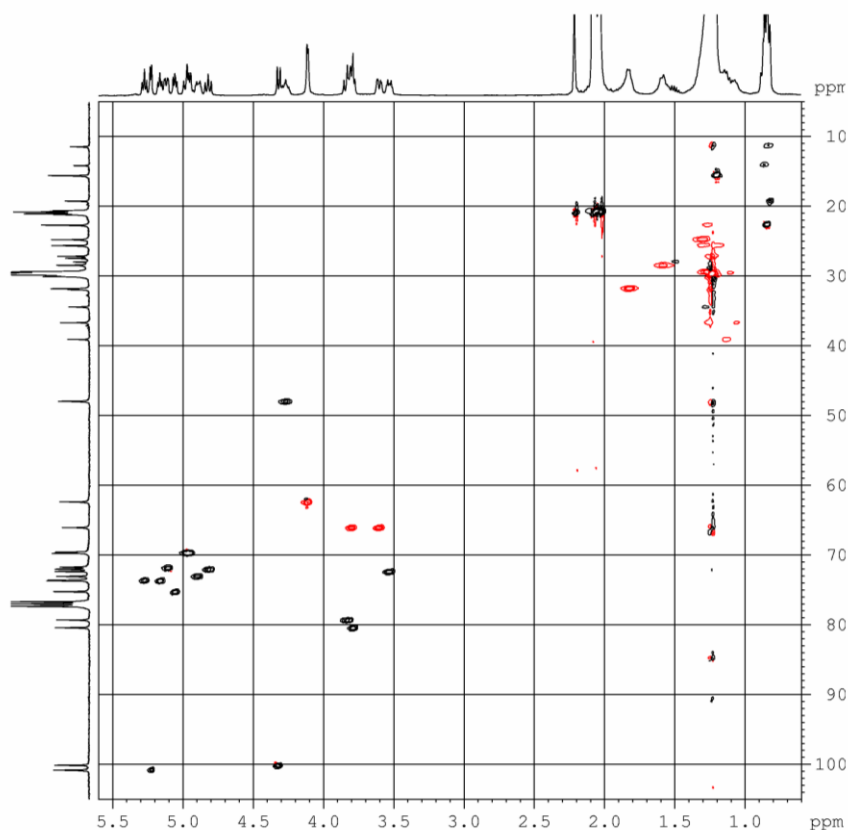


Figure 12: 2D NMR HSQC spectrum of peracetylated terpioside (**1b**) in CDCl_3 at 500MHz

Starting from the anomeric proton doublet at δ 4.30, it was possible to assign through the COSY and TOCSY spectra the signals of all the protons in the first monosaccharide unit: four oxymethine protons [δ 4.81, 3.83, 4.97 and 3.53 (2^1-H , 3^1-H , 4^1-H e 5^1-H)] and a couple of oxymethylene protons (6^1-H), suggesting it is a hexose.

The ROESY e HMBC spectra indicated that this sugar unit is linked to the primary hydroxyl group of the ceramide, as shown by the ROESY correlation peak of 1^1-H with 1-H_a e 1-H_b and by the HMBC correlation peak between 1^1-H and C-1.

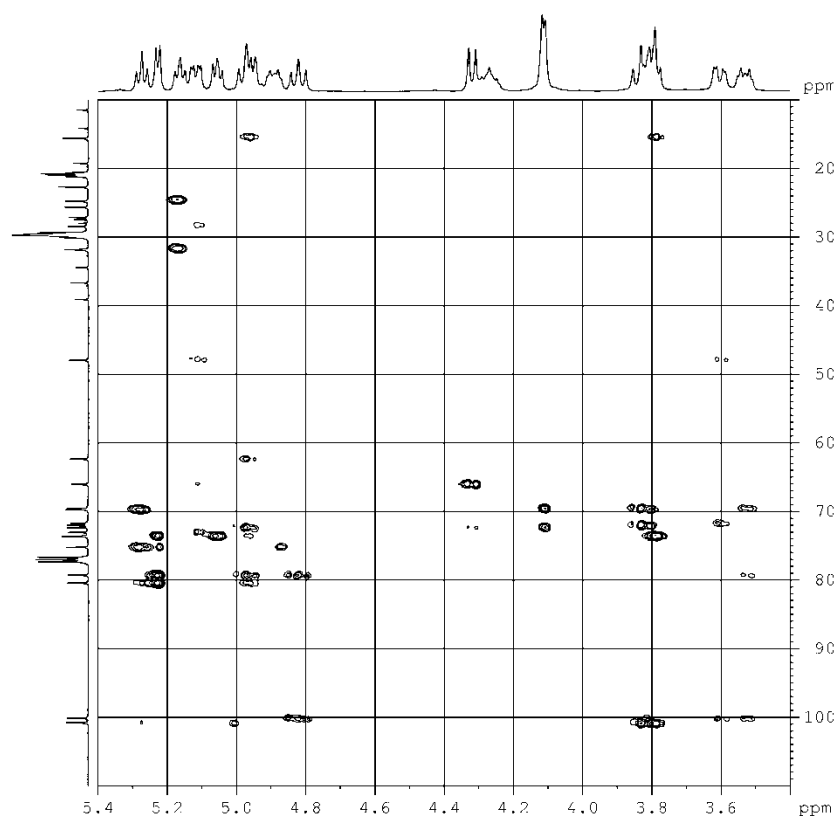


Figure 13: Mid-field region of the 2D NMR HMBC spectrum of teriposide (**1b**) in CDCl_3 at 500MHz

In addition, the deshielded chemical shift of the oxymethine protons at the 2^I-, 4^I-, and 6^I- positions indicated that the relevant hydroxy groups are acetylated. In contrast, the high-field chemical shift of 3^I-H (δ 3.83) suggested glycosylation at this position, whereas the high-field chemical shift of the 5^I-H (δ 3.51) confirmed that the relevant carbon atom C-5^I is involved in the pyranose acetal function.

Similarly, the second sugar spin system was recognized from the TOCSY and COSY experiments as that of a 6-deoxyhexose. This was indicated by the methyl doublet at δ 1.21 (6^{II}-H₃), which pointed to the presence of a methyl group at the 6-position instead of the usual $-\text{CH}_2\text{OH}$ group.

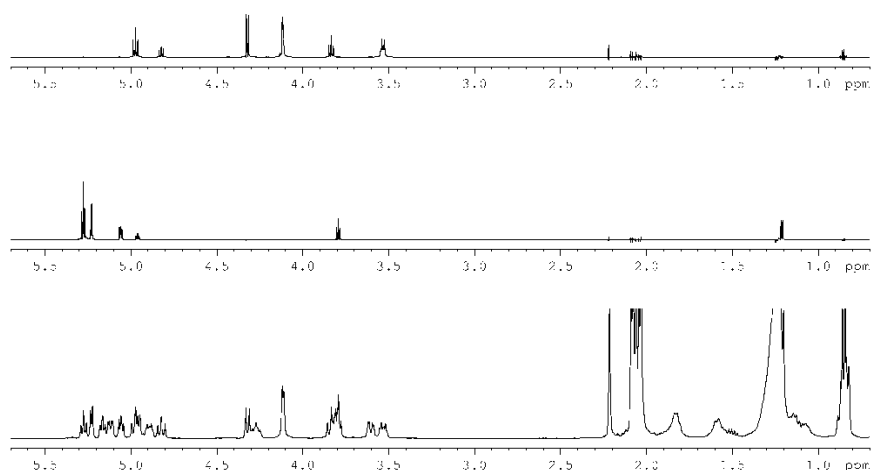


Figure 14: The spin systems of the two sugars spectrum of compound (**1b**) evidenced by the *z*-TOCSY spectrum (CDCl₃, 500MHz)

The high-field chemical shift of the 4^{II}-H signal (δ 3.79), compared with the low-field chemical shift of the 5^{II}-H signal (δ 4.96), clearly indicated that C-4^{II} links an oxygen atom involved in an acetal function instead of an ester function; thus, the sugar must be in the furanose form and not in the more common pyranose form. This second sugar unit is linked to the first sugar at position 3, as shown by the HMBC correlation between C-1^{II} and 3^I-H and the ROESY correlation between 1^{II}-H and 3^I-H.

Table 1: ^1H and ^{13}C NMR spectroscopic data of terpioside peracetate **1b** (CDCl_3)

| Position | | δ_{H} (mult, J (Hz)) ¹ | δ_{C} (mult) ² |
|----------|-----------------|---------------------------------------------------|-----------------------------------------|
| 1 | a | 3.80 (dd, 10.4, 2.9) | 66.0 (CH ₂) |
| | b | 3.61 (dd, 10.4, 3.4) | |
| 2 | | 4.27 (m) | 47.9 (CH) |
| 2-NH | | 6.78 (d, 9.0) | - |
| 3 | | 5.12 (dd, 8.4, 3.2) | 71.7 (CH) |
| 4 | | 4.89 (ddd, 10.0, 3.2, 3.2) | 73.0 (CH) |
| 5 | | 1.59 (m) | 28.5 (CH ₂) |
| 6 | a | 1.30 (m) | 25.6 (CH ₂) |
| | b | 1.19 (m) | |
| 1' | | 4.32 (d, 8.0) | 100.1 (CH) |
| 2' | | 4.81 (dd, 9.5, 8.0) | 72.0 (CH) |
| 3' | | 3.83 (t, 9.5) | 79.3 (CH) |
| 4' | | 4.97 (t, 9.5) | 69.6 (CH) |
| 5' | | 3.53 (ddd, 9.5, 3.9, 3.9) | 72.3 (CH) |
| 6' | a | 4.12 (m) | 62.3 (CH ₂) |
| 1'' | | 5.23 (d, 4.6) | 100.8 (CH) |
| 2'' | | 5.06 (dd, 6.3, 4.6) | 75.2 (CH) |
| 3'' | | 5.28 (t, 6.3) | 73.5 (CH) |
| 4'' | | 3.79 (t, 6.3) | 80.4 (CH) |
| 5'' | | 4.96 (dd, 6.6, 6.3) | 69.6 (CH ₂) |
| 6'' | | 1.21 (d, 6.6) | 15.5 (CH ₃) |
| 1''' | | - | 169.7 (C) |
| 2''' | | 5.17 (dd, 6.8, 4.9) | 73.6 (CH) |
| 3''' | | 1.83 (m) | 31.8 (CH ₂) |
| 4''' | | 1.32 (m) | 24.7 (CH ₂) |
| Ac's | CH ₃ | 2.22, 2.09, 2.08, 2.08, 2.06, 2.10, 2.05, 2.03 | 21.0-20.6 (CH ₃) |
| | CO | - | 170.9-169.1 |

- a. Additional ^1H signals: δ 1.24 (br., alkyl chain protons), 0.87 (t, $J = 7.0$ Hz, *n*-chain Me groups), 0.85 (d, $J = 6.5$ Hz, *iso*-chain Me groups) ppm.
- b. Additional ^{13}C signals: δ 31.9 (CH₂, ω -2), 22.7 (CH₂, ω -1), 22.7 (CH₃, *iso*-chain Me groups), 14.1 (CH₃, ω) ppm.

Stereostructure and Alkyl Chain

The analysis of coupling constants is a useful method to elucidate the relative configurations of the chiral centers of sugars in the pyranose form and, consequently, to establish the nature of the sugar. In our case, the first hexose unit was readily recognized as a β -glucopyranoside because of the large coupling constants between all the pairs of vicinal ring protons which pointed to their axial orientation.

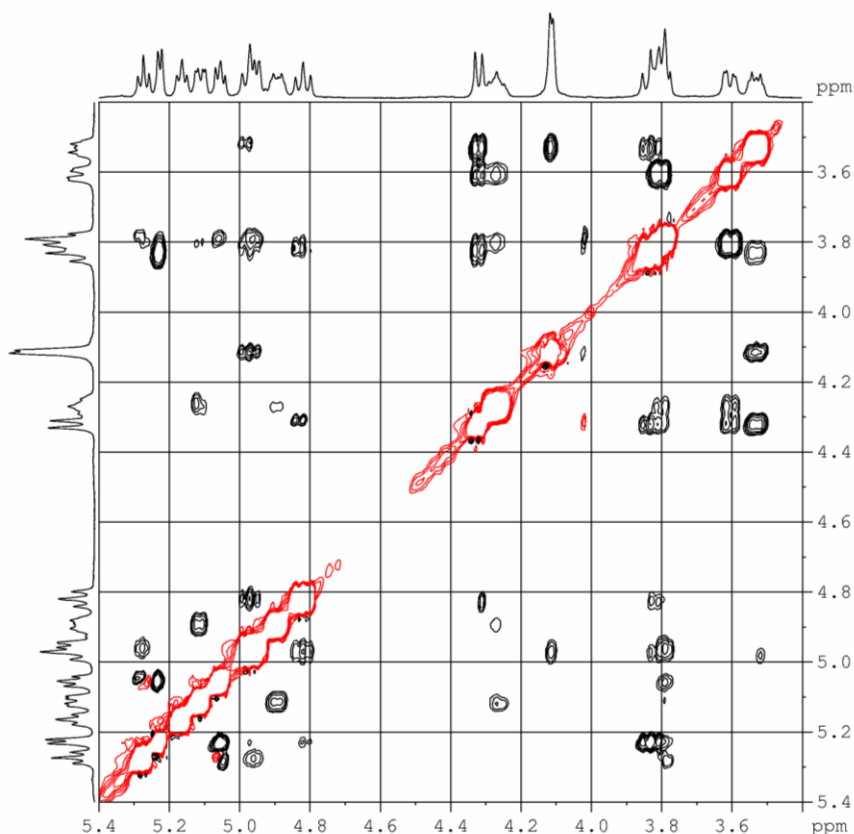
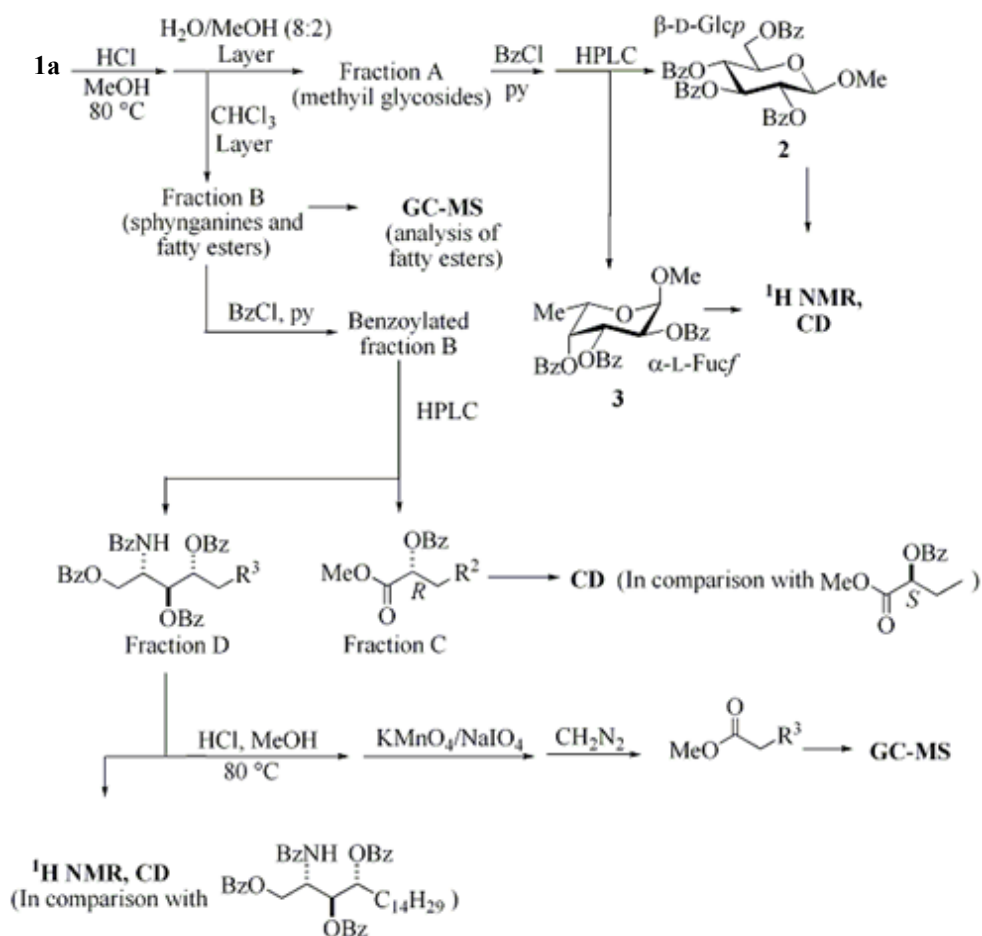


Figure 15: Mid-field region of the 2D NMR ROESY spectrum of terpioside (**1b**) in CDCl_3 at 500MHz

As for the furanose sugar unit, the ROESY spectrum showed correlation peaks between $1^{\text{II}}\text{-H}/4^{\text{II}}\text{-H}$, $2^{\text{II}}\text{-H}/4^{\text{II}}\text{-H}$ e $3^{\text{II}}\text{-H}/5^{\text{II}}\text{-H}$, which indicated that 1^{II}-H , 2^{II}-H e 4^{II}-H lie on the same side of the five-membered ring, whereas 3^{II}-H

H and the CHOHCH_3 group at C-4^{II} are on the other side. The configuration at C-5 could not be established on the basis of the ROESY data, so the sugar could be either an α -fucosylfuranoside (6-deoxy- α -galactofuranoside) or a 6-deoxy- β -altrosylfuranoside. Degradation analysis showed the former to be the case.

The remaining structural features of terpioside were established by chemical degradation. The natural glycolipid **1a** was subjected to acidic methanolysis, and the reaction products were separated, perbenzoylated, and analyzed by CD and GC-MS (**Scheme 1**).



Scheme 1: Microscale degradation analysis of terpioside A (**1a**)

This procedure, set up in our laboratory,² allowed us to establish the absolute configuration of each sugar and that of the hydroxy acid, and the relative and absolute configurations of the phytosphingosine, as well as to identify the alkyl chains of sphingosines (**Table 2**) and fatty acids (**Table 3**).

| Sphinganine | Comp. [%] |
|---------------------------------------------------------------------------------------------------------------------|-----------|
| (2 <i>S</i> ,3 <i>S</i> ,4 <i>R</i>)-2-Amino-16-methyl-1,3,4-heptadecanetriol (<i>iso</i> -C ₁₈) | 4.5 |
| (2 <i>S</i> ,3 <i>S</i> ,4 <i>R</i>)-2-Amino-15-methyl-1,3,4-heptadecanetriol (<i>ante-iso</i> -C ₁₈) | 15.5 |
| (2 <i>S</i> ,3 <i>S</i> ,4 <i>R</i>)-2-Amino-1,3,4-octadecanetriol (<i>n</i> -C ₁₈) | 4.1 |
| (2 <i>S</i> ,3 <i>S</i> ,4 <i>R</i>)-2-Amino-16-methyl-1,3,4-octadecanetriol (<i>iso</i> -C ₁₉) | 70.1 |
| (2 <i>S</i> ,3 <i>S</i> ,4 <i>R</i>)-2-Amino-1,3,4-nonadecanetriol (<i>n</i> -C ₁₉) | 5.8 |

Table 2: Sphinganine composition of terpioside **1a**

| Fatty acid methyl esters | Composition [%] |
|-------------------------------------------------------------------------------|-----------------|
| Methyl 2-hydroxy-20-methylenicosanoate (<i>iso</i> -C ₂₂) | 3.6 |
| Methyl 2-idrossidocosanoate (<i>n</i> -C ₂₂) | 1.7 |
| Methyl 2-hydroxy-21-methyldocosanoate (<i>iso</i> -C ₂₃) | 17.0 |
| Methyl 2-hydroxy-20-methyldocosanoate (<i>ante-iso</i> -C ₂₃) | 4.2 |
| Methyl 2-hydroxytricosanoate (<i>n</i> -C ₂₃) | 4.4 |
| Methyl 2-hydroxy-22-methyltricosanoate (<i>iso</i> -C ₂₄) | 2.2 |
| Methyl 2-hydroxy-21-methyltricosanoate (<i>ante-iso</i> -C ₂₄) | 5.3 |
| Methyl 2-hydroxytetracosanoate (<i>n</i> -C ₂₄) | 22.3 |
| Methyl 2-hydroxy-23-methyltetracosanoate (<i>iso</i> -C ₂₅) | 15.5 |
| Methyl 2-hydroxy-22-methyltetracosanoate (<i>ante-iso</i> -C ₂₅) | 5.0 |
| Methyl 2-hydroxypentacosanoate (<i>n</i> -C ₂₅) | 11.7 |
| Methyl 2-hydroxy-24-methylpentacosanoate (<i>iso</i> -C ₂₆) | 3.2 |
| Methyl 2-hydroxy-23-methylpentacosanoate (<i>ante-iso</i> -C ₂₆) | 1.8 |
| Methyl 2-hydroxyhexacosanoate (<i>n</i> -C ₂₆) | 2.1 |

Table 3: Fatty acid composition of terpioside **A 1a**

In particular, analysis of the perbenzoylated methyl glycoside fraction led to the isolation of the 6-deoxyglycoside **3**, which was tentatively identified as tri-*O*-benzoyl- α -fucopyranoside on the basis of its ¹H NMR spectrum. An authentic sample of tri-*O*-benzoyl- α -fucopyranoside was prepared from L-fucose by acidic methanolysis and subsequent

perbenzylation, and its ^1H NMR and CD spectra were recorded. The spectra were identical to those of compound **3**, thus confirming that the outer sugar of terpioside is a fucose and showing that its absolute configuration is L.

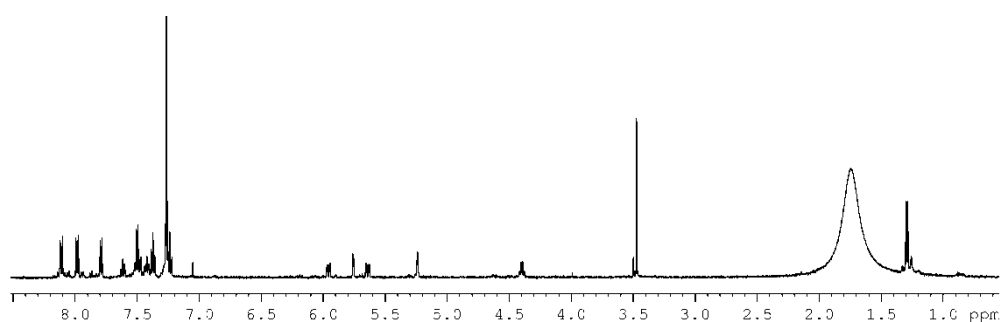


Figure 16: ^1H NMR spectrum of synthetic benzoylated L-fucose

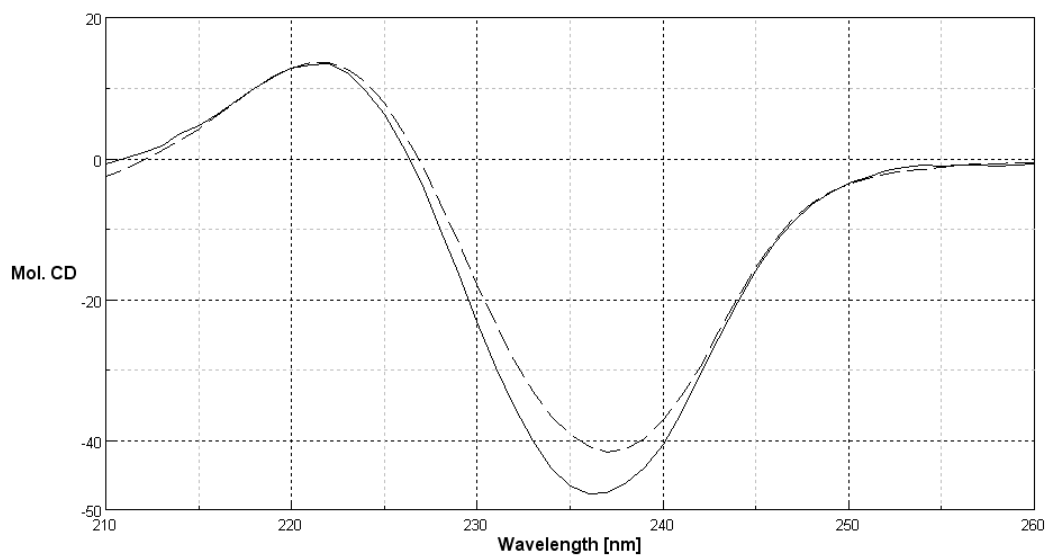


Figure 17: CD spectrum (MeCN) of Methyl tri-*O*-benzoyl- α -fucopiranoside (**3**) from L-fucose (dashed line) and terpioside **1a** (solid line)

Once the structure of the peracetyl derivative **1b** had been elucidated, the one- and two-dimensional NMR spectra of the natural GSL **1a** were recorded and analyzed. The information provided by the COSY and

HMQC NMR spectra of **1a** confirmed all the structural features determined for its peracetylated derivative and allowed the assignment of all the resonances in its ^1H and ^{13}C NMR spectra (**Exp. Sect.**).

To make it sure that no acetyl groups were present in the natural terpioside **1a** before the peracetylation reaction used for its isolation, we performed the acetylation reaction on the glycolipid fraction using trideuterioacetic anhydride, and then purified the mixture to give the deuterio derivative **1c**. The ^1H NMR spectrum of **1c** was identical to that of **1b**, except that no acetyl methyl singlet was present.

4.3 Experimental Section

Collection, Extraction, and Isolation: Specimens of *Terpios* sp. were collected in December 2005 from along Key Largo coast (Florida) and identified by Prof. S. Zea (University of Colombia). They were frozen immediately after collection and kept frozen until extraction. The sponge (220 g of dry weight after extraction) was homogenized and extracted with methanol (3x1L), and then with chloroform (3x1 L); the combined extracts were partitioned between H_2O and *n*-BuOH. The organic layer was concentrated in vacuo and afforded 20 g of dark green oil, which was purified by chromatography on a column packed with RP-18 silica gel. A fraction eluted with CHCl_3 (5.9 g) was further purified by chromatography on an SiO_2 column, giving a fraction [820 mg; eluent: EtOAc/MeOH (7:3)] mainly composed of glycolipids. This fraction was peracetylated with Ac_2O in pyridine for 12 h. The acetylated glycolipids were subjected to HPLC

separation on an SiO₂ column [eluent: *n*-hexane/EtOAc (6:4)], thus affording 62 mg of terpioside peracetate **1b**.

Terpioside Peracetate 1b: Colorless oil, $[\alpha]_{\text{D}}^{25} = -8$ ($c = 0.15$, CHCl₃). ESI-MS (positive ion mode, MeOH): $m/z = 1434, 1420, 1406, 1392, 1378, 1364$ ([M + Na]⁺ series). ¹H and ¹³C NMR: **Table 1**. Composition of the sphinganine: **Table 2**. Composition of the fatty acids: **Table 3**.

Deacetylation of 1b: Compound **1b** (30 mg) was dissolved in MeOH (950 μL) and a 0.4 M solution of MeONa in MeOH (50 μL) was added. The reaction was allowed to proceed at 25 °C for 18 h and then the reaction mixture was dried under nitrogen and the residue partitioned between water and chloroform. After removal of the solvent, the organic layer gave 22 mg of the native glycosphingolipid **1a**.

Terpioside (1a): Colorless amorphous solid, $[\alpha]_{\text{D}}^{25} = -25$ ($c = 0.15$, MeOH). HR-MS (ESI, positive ion mode, MeOH): calcd. for [C₅₅H₁₀₇NNaO₁₄]⁺ 1028.7589; found 1028.7567 [M + Na]⁺. ¹H NMR ([D₅]pyridine): δ 0.86 (*n*- and *iso*-chain Me groups), 1.28 (large band, alkyl chains), 1.54 (d, $J = 6.5$ Hz, 3 H, 6^{II}-H), 1.65 (m, 1 H, 6-Hb), 1.69 (m, 1 H, 4^{III}-Hb), 1.75 (m, 1 H, 4^{III}-Ha), 1.88 (m, 1 H, 5-Hb), 1.90 (m, 1 H, 5-Ha), 1.99 (m, 1 H, 3^{III}-Hb), 2.22 (m, 1 H, 5-Ha), 3.75 (m, 1 H, 5^I-H), 3.95 (ddd, $J = 7.8, 7.8, 3.7$ Hz, 1 H, 2^I-H), 4.08 (ddd, $J = 9.2, 9.2, 3.5$ Hz, 1 H, 4^I-H), 4.15 (m, 1H, 4-H), 4.16 (overlapped, 4^{II}-H), 4.23 (overlapped, 3^I-H), 4.24 (overlapped, 6^I-Hb), 4.26 (m, 1 H, 3-H), 4.36 (overlapped, 5^{II}-H), 4.56 (m, 1 H, 2^{III}-H), 4.47 (dd, $J =$

10.8 and 7.0 Hz, 1 H, 1-Ha), 4.65 (dd, $J = 10.8, 3.9$ Hz, 1 H, 1-Hb), 4.85 (d, $J = 7.9$ Hz, 1 H, 1^I-H), 5.24 (m, 2-H), 6.08 (br. s, 1 H, 4-OH), 6.14 (d, $J = 4.5$ Hz, 1 H, 1^{II}-H), 6.25 (br. s, 1 H, 2^{II}-OH), 6.43 (br. s, 1 H, 6^I-OH), 6.53 (br. s, 1 H, 5^{II}-OH), 6.86 (br. s, 1 H, 3-OH), 7.10 (br. s, 1 H, 4^I-OH), 7.36 (br. s, 1 H, 2^I-OH), 7.45 (br. s, 1 H, 3^{II}-OH), 7.73 (br.s, 1 H, 2^{III}-OH), 8.53 (d, $J = 9.5$ Hz, 1 H, 2-NH) ppm. ¹³C NMR ([D₅]pyridine): δ 14.3 (CH₃, *n*-chain Me groups), 20.6 (CH₃, C-6^{II}), 22.8 (CH₃, *iso*-chain Me groups), 22.9 (CH₂, *n*-chain ω -1 CH₂ groups), 25.6 (CH₂, C-4^{III}), 26.3 (CH₂, C-6), 30.5–29.5 (several CH₂, alkyl chains), 32.1 (CH₂, *n*-chain ω -2 CH₂ groups), 33.8 (CH₂, C-5), 35.3 (CH₂, C-3^{III}), 51.4 (CH, C-2), 61.9 (CH₂, C-6^I), 65.9 (CH, C-5^{II}), 61.9 (CH₂, C-6^I), 69.2 (CH, C-4^I), 72.1 (CH, C-4; CH, C-2^{III}), 74.9 (CH, C-3^{II}), 77.9 (CH, C-5^I), 79.2 (CH, C-2^{II}), 84.1 (CH, C-3^I), 86.4 (CH, C-4^{II}), 105.1 (CH, C-1^I), 103.1 (CH, C-1^{II}), 105.3 (CH, C-1^I), 175.3 (C, C-1^{III}) ppm. Composition of the sphinganine: **Table 2**. Composition of the fatty acids: **Table 3**.

Methanolysis of 1a: Compound **1a** (100 μ g) was dissolved in 1 N HCl in 91% MeOH (500 μ L), and the solution obtained was kept in a sealed tube at 80 °C for about 12 h. The reaction mixture was dried under nitrogen and partitioned between CHCl₃ and H₂O/ MeOH (8:2). The aqueous layer was concentrated to give a mixture of methyl glycosides (Fraction A), whereas the organic layer contained a mixture of α -hydroxy acid methyl esters and sphinganine (Fraction B).

Methyl Tri-O-benzoyl- α -L-fucopyranoside (3): L-Fucose (2.0 mg) was subjected to acidic methanolysis as described above. The resulting methyl glycosides were benzoylated with benzoyl chloride (50 μ L) in pyridine (500 μ L) at 25 °C for 16 h. The reaction was then quenched with MeOH; after 30 min, the mixture was dried under nitrogen. Methyl benzoate was removed by keeping the residue under vacuum with an oil pump for 24 h. The residue was purified by HPLC [column: Luna SiO₂, 5 μ ; eluent: *n*-hexane/ *i*PrOH (99:1); flow: 1 mLmin⁻¹; UV detector: 280 nm] affording the glycoside **3** (t_R = 6.6 min). ¹H NMR (CDCl₃): δ 1.29 (d, J = 6.5 Hz, 3 H, 6-H), 3.47 (s, 3 H, OMe), 4.39 (br. q, J = 6.5 Hz, 1 H, 5-H), 5.24 (d, J = 3.6 Hz, 1 H, 1-H), 5.64 (dd, J = 10.7, 3.6 Hz, 1 H, 2-H), 5.76 (br. d, J = 3.4 Hz, 1 H, 4-H), 5.95 (dd, J = 10.7, 3.4 Hz, 1 H, 3-H), 7.24 (t, J = 7.7 Hz, 2 H, benzoyl *meta* protons), 7.37 (t, J = 7.4 Hz, 2 H, benzoyl *meta* protons), 7.42 (t, J = 7.6 Hz, 1 H, benzoyl *para* proton), 7.53–7.45 (overlapping signals, 3 H, benzoyl protons), 7.61 (t, J = 7.5 Hz, 1 H, benzoyl *para* proton), 7.79 (d, J = 7.9 Hz, 2 H, benzoyl *ortho* protons), 7.98 (d, J = 7.9 Hz, 2 H, benzoyl *ortho* protons), 8.11 (d, J = 8.0 Hz, 2 H, benzoyl *ortho* protons) ppm. CD (MeCN): λ_{max} = 237 ($\Delta\epsilon$ = -42), 222 ($\Delta\epsilon$ = +14) nm.

Absolute Stereochemistry of Methyl Glycosides from Compound 6a:

Fraction A from the methanolysis of compound **1a** was benzoylated with benzoyl chloride (20 μ L) in pyridine (200 μ L) at 25 °C for 16 h. The reaction was then quenched with MeOH; after 30 min, the mixture was dried under nitrogen. Methyl benzoate was removed by keeping the residue under vacuum with an oil pump for 24 h. The residue was purified by HPLC

[column: Luna SiO₂, 5 μ; eluent: *n*-hexane/*i*PrOH (99:1); flow: 1 mL min⁻¹]. The chromatogram showed two peaks: methyl tetra-*O*-benzoyl- α -D-glucopyranoside (**2**), identified by comparison of its ¹H NMR and CD spectra with those reported in the literature,² and methyl tri-*O*-benzoyl- α -L-fucopyranoside (**3**), identified by comparison of its ¹H NMR and CD spectra with those of the authentic samples prepared from L-fucose.

Analysis of Fatty Acid Methyl Esters: Fraction B from the methanolysis of compound **1a** was analyzed by GLC-MS and its components identified by comparison of their retention times and mass spectra with those of authentic samples. The results are compiled in **Table 3**.

Analysis of Fraction B: Fraction B from the methanolysis of compounds **1a** was benzoylated as described above and the crude product of the reaction was purified by HPLC [column: Luna SiO₂, 5 μ; eluent: *n*-hexane/*i*PrOH (99:1); flow: 1 mLmin⁻¹]. The chromatogram shows two peaks, which were identified as a mixture of homologues of benzoylated fatty acid methyl esters (fraction C, *t*_R = 4.0 min) and a mixture of perbenzoylated sphinganine (Fraction D, *t*_R = 6.9 min) on the basis of their respective ¹H NMR spectra.

Methyl (R)-2-Benzoyloxyalkanoate: Fraction C from the mixture of homologues. CD (MeCN): $\lambda_{\text{max}} = 229$ ($\Delta\epsilon = -3.4$) nm. The ¹H NMR spectrum was identical to that reported in the literature.²

(2S,3S,4R)-1,3,4-O-Benzoyl-2-benzoylamino-1,3,4-alkanetriol: Fraction D from the mixture of homologues. CD (MeCN): $\lambda_{\max} = 233$ ($\Delta\epsilon = -1$), 222 ($\Delta\epsilon = +2$) nm. The ^1H NMR spectrum was identical (apart from the methyl region) to that of an authentic sample of D-ribo-phytosphingosine perbenzoate.³

Oxidative Cleavage and GC-MS Analysis of Sphinganines: Fraction D was debenzoylated by acidic methanolysis as described above and subjected to oxidative cleavage with $\text{KMnO}_4/\text{NaIO}_4$ as described in the literature.⁴ The resulting carboxylic acids were methylated with CH_2N_2 . The esters obtained were analyzed by GC-MS, and the results are compiled in **Table 2**, expressed in terms of original sphinganines.

References

1. V. Costantino, E. Fattorusso, C. Imperatore, A. Mangoni, R. Teta, *Eur. J. Org. Chem.* **2008**, 2130-2134
2. V. Costantino, E. Fattorusso, C. Imperatore, A. Mangoni, S. Freigang, L. Teyton, *Bioorg. Med. Chem.*; DOI:10.1016/j.bmc.2007.10.098
3. V. Costantino, E. Fattorusso, C. Imperatore, A. Mangoni, *J. Org. Chem.* **2004**, *69*, 1174-1179
4. V. Costantino, E. Fattorusso, A. Mangoni, M. Di Rosa, A. Ianaro, P. Maffia, *Tetraedron* **1996**, *52*, 1573-1578

Chapter 5

Planar Structure Elucidation of Terpioside B from *Terpios* sp.

The investigation of the GSL composition of the marine sponge *Terpios* sp. revealed the presence of terpioside B, a new furanose-rich pentaglycosylated glycosphingolipid. Planar structure of terpioside B (**1**) was elucidated using extensive 2D NMR and mass spectrometry studies.

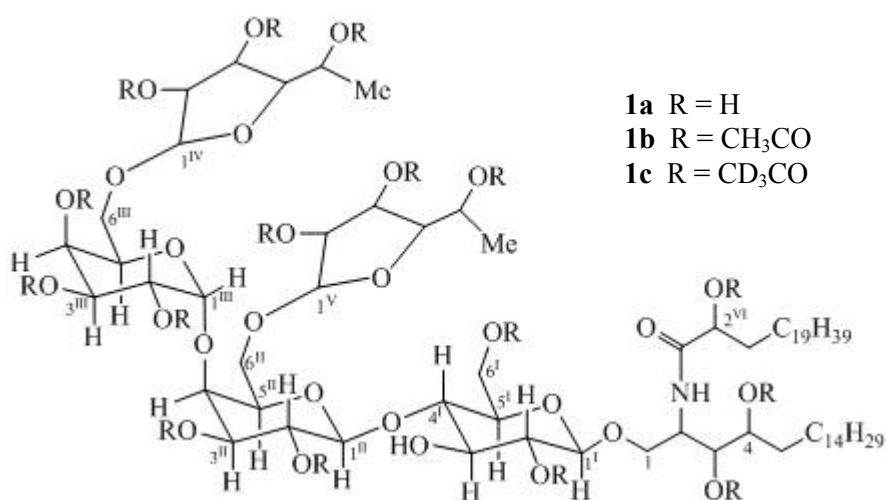


Figure 18: Terpioside B from *Terpios* sp.

5.1 Isolation of Terpioside B

A further analysis of the peracetylated glycolipid fraction obtained from *Terpios* sp. (see Chapter 4) was performed. The extract showed the presence of more polar glycosphingolipids and, after repeated HPLC separation on SiO₂ columns, pure terpioside B peracetate (**1b**) was obtained.

5.2 Structure Determination

The ESI mass spectrum of peracetylated terpioside B (**1b**) gave a series of $[M+2Na]^{2+}$ pseudomolecular ion peaks at m/z 1104.8, 1111.7, 1118.2, and 1125.5, in accordance to molecular ions of m/z 2163, 2177, 2191, 2205. The singly charged $[M+Na]^+$ ions could not be detected, because their m/z ratio was beyond the upper limit of the analyzer in the mass spectrometer we used. This molecular weights are in accordance with the molecular formulas $C_{105}H_{167}NO_{45} + n CH_2$ ($n = 0-3$), and showed that terpioside B, like most glycolipids of marine origin, is present in *Terpios* as a mixture of homologues.

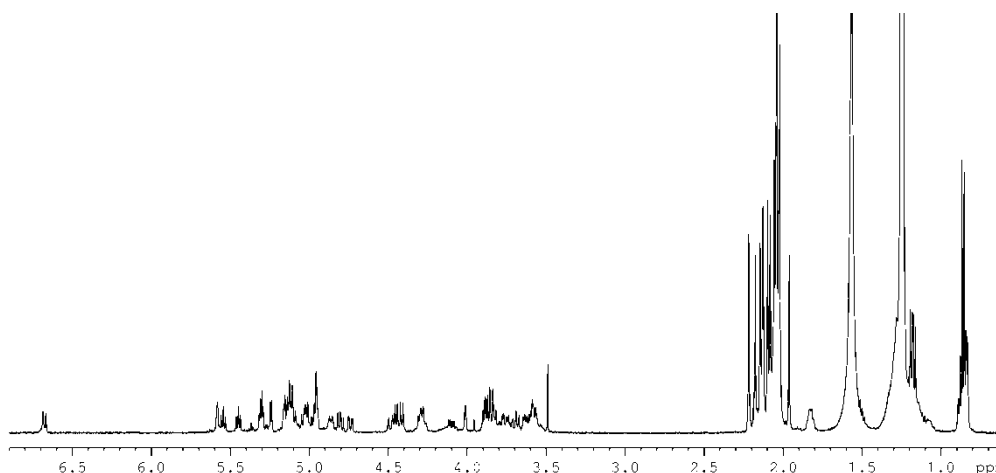


Figure 19: 1H NMR spectrum of terpioside B peracetate in $CDCl_3$ at 700 MHz

The general features of the proton spectrum of peracetylated terpioside B clearly resembled those of glycolipids, as illustrated by (a) the large band of alkyl chain methylene protons at δ 1.24, (b) many overlapping oxymethine and oxymethylene protons between δ 3.4 and 5.6 and (c) $-NH$ doublet at δ 6.65 ppm.

All the NMR experiments directed to the structure elucidation were performed on the peracetate derivative, to take advantage of the better signal dispersion in the proton spectrum of a peracetylated sugar chain and the possibility of identifying the glycosylation site of each sugar.

The ceramide portion of terpioside B is the one most commonly found in glycolipids from sponges, and the same as in terpioside A, composed of a trihydroxylated, saturated sphinganine and a 2-hydroxy fatty acid.

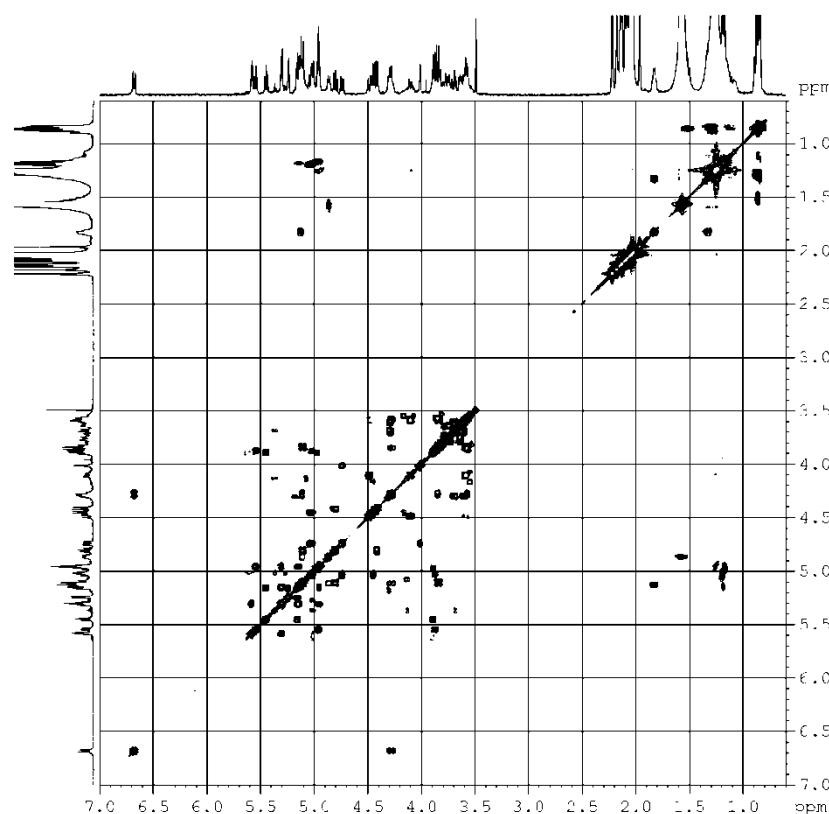


Figure 20: 2D NMR COSY spectrum of compound **1b** in CDCl_3 at 700 MHz

Using -NH proton as a starting point, analysis of the 2D COSY allowed identification of all the proton of the polar part of the sphinganine up to H-5. As for the ceramide fatty acid residue, a resonance at δ 5.13 (H-2^{VI}) in the ^1H NMR spectrum, instead of the characteristic triplet at δ 2.3 of the

fattyacid α -protons, indicated the α -hydroxy substitution. This resonance showed in the ROESY spectrum an intense correlation peak with the amide NH doublet, confirming the $-\text{NHCOCHOH}-$ functionality. Both H-2^{VI} and 2-NH were shown to be coupled with the CO carbon atom at δ 170.0 (C-1^{VI}) by the HMBC spectrum.

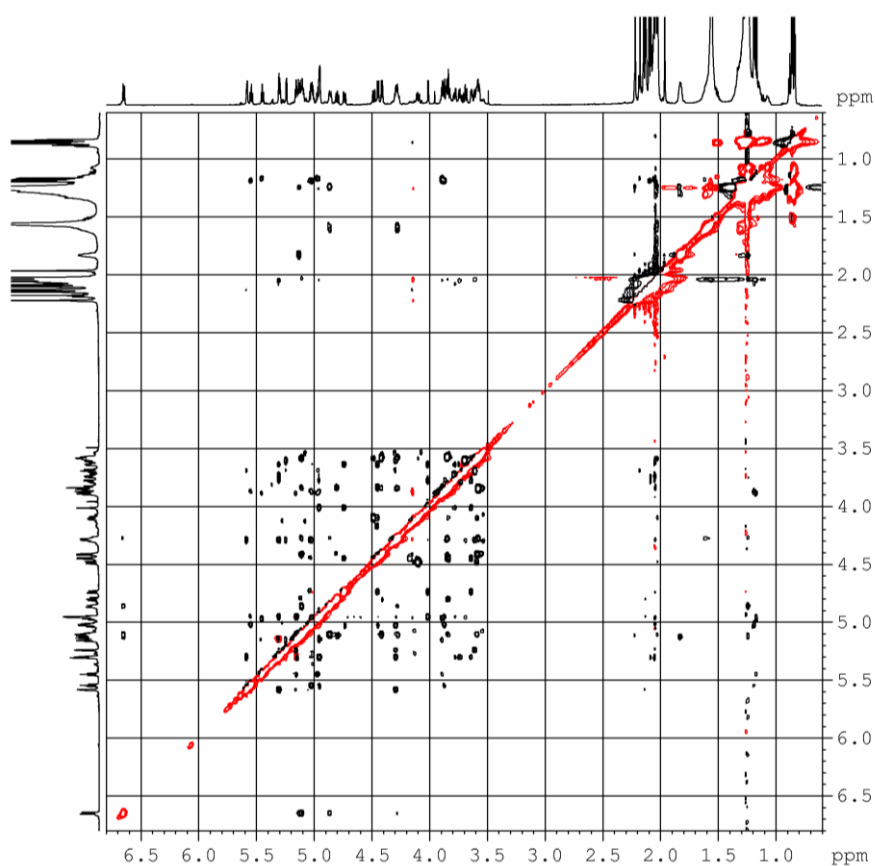


Figure 21: 2D NMR ROESY spectrum of terpioside B peracetate (**1b**) in CDCl_3 at 700 MHz

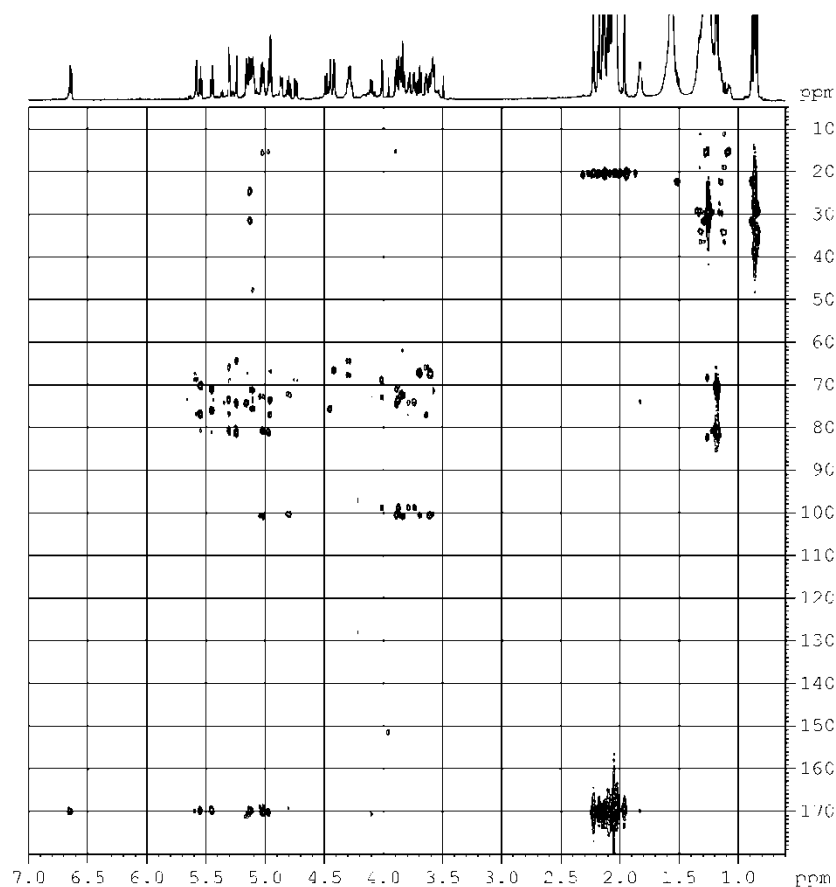


Figure 22: 2D NMR HMBC spectrum of compound **1b** in CDCl_3 at 700 MHz

Analysis of HSQC spectrum of terpioside B peracetate showed the sugar portion of the molecule to be composed by a pentasaccharide chain. In the spectrum, resonances for the five anomeric protons (δ 4.41, 4.44, 4.95, 5.24, and 5.30) and the relevant carbons (δ 100.2, 100.7, 98.8, 100.4, and 98.7 respectively) were identified.

Elucidation of the structure of the pentasaccharide chain was more challenging for the severe signal overlapping and the presence of two sugars in the furanose form. Decisive information came from the z -filtered version of the TOCSY 2D NMR experiment. **Figure 24** shows sections of the z -TOCSY experiment displaying subspectra of the six sugar spin-systems,

which allowed us to identify the signals of each sugar, starting from the relevant anomeric protons.

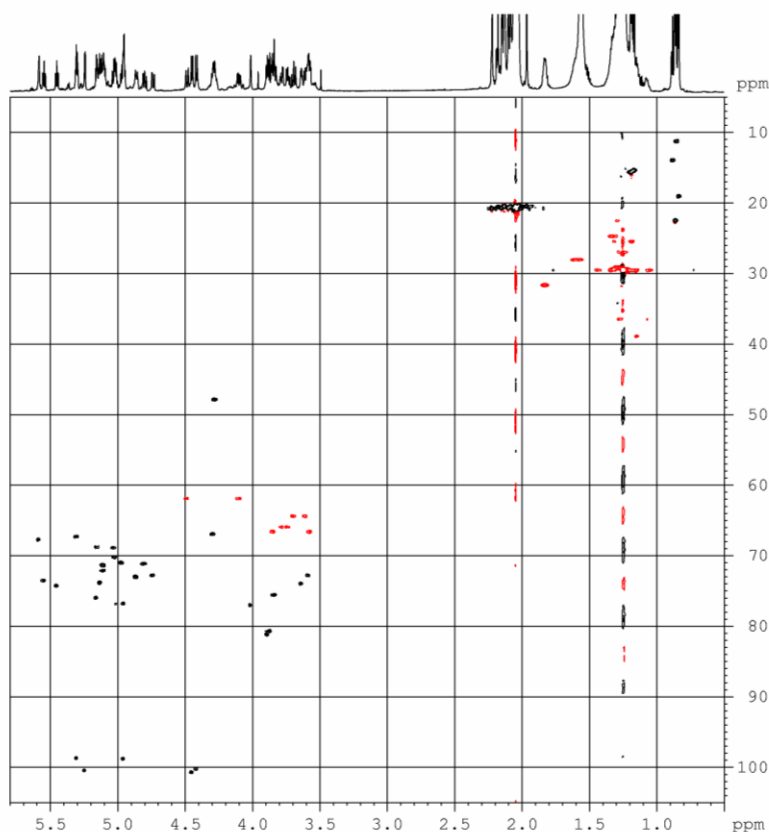


Figure 23: 2D NMR HSQC spectrum of terpioside B peracetate (**1b**) in CDCl_3 at 700 MHz

The first sugar residue of the saccharide chain was identified as a β -glucopyranoside. The shielded chemical shift of H-5^{I} indicated a pyranose ring, while the large coupling constant (7.8 Hz) of the anomeric proton was indicative of its axial orientation. The large coupling constants between each vicinal pair of ring protons clearly indicated that all the ring protons are axial. The HMBC spectrum showed that this sugar is linked to the sphinganine (correlation peak between H-1^{I} and C-1), and glycosylated at 4^{I} (correlation peak between H-1^{II} and C-4^{I}).

The second sugar residue is a hexopyranose as indicated by the high-field chemical shift of H-5^{II}. Coupling constants (**Table 4**) showed that H-1^{II}, H-2^{II} and H-3^{II} are axial, while H-4^{II} is equatorial. This sugar residue is therefore a β -galactopyranoside. The shielded chemical shift of H-4^{II} and H-6^{II} suggested glycosylation at these positions, which was confirmed by the HMBC correlation peak between H-4^{II} and C-1^{III} and between H-6^{II} and C-1^V.

The third and last pyranose sugar residue is a α -galactopyranoside. The pattern of coupling constants (**Table 4**) showed the axial orientation of H-2^{III} and H-3^{III}, and the equatorial orientation of H-1^{III} and H-4^{III}. As for H-5^{III}, it shows a ROESY correlation peak with H-3^{III}, and is therefore axial. Finally, this sugar is glycosylated at position 6, as shown by the chemical shift of H_{a,b}-6^{III} (δ 3.60, 3.69) and by the HMBC correlations between H-1^{IV} and C-6^{III} and between H_{a,b}-6^{III} and C-1^{IV}.

The sugar glycosylating the third sugar at position 6 is a 6-deoxyhexofuranoside, as indicated by the presence of five methine protons and the characteristic methyl doublet at δ 1.17 ppm, which pointed the presence of a methyl group instead of the usual $-\text{CH}_2\text{OH}$ group at the 6-position in the relevant spin system. The high-field chemical shift of H-4^{IV} (δ 3.89), compared with the low-field chemical shift of the H-5^{IV} signal (δ 4.97), clearly indicated that C-4^{IV} links an oxygen atom involved in an acetal function instead of an ester function; thus, the sugar must be in the furanose form and not in the more common pyranose form. The low-field chemical shifts of H-2^{IV}, H-3^{IV}, H-5^{IV} and H₂-6^{IV} show that hydroxyl groups

at positions 2, 3, 5 and 6 are all acetylated, and therefore that the sugar is not further glycosylated.

The fifth sugar residue, glycosylating the β -galactopyranoside at position 6 and not further glycosylated, was also a hexose in the furanose form (HMBC correlation between H-1^V and C-4^V). Also this sugar is a terminal sugar, as argued by the chemical shifts of H-2^V, H-3^V, H-5^V and H₂-6^V.

Determination of the stereochemistry of furanose sugars, as well as of the absolute configuration of the whole molecule, is still in progress and will be performed using methods similar to those described for terpioside A.

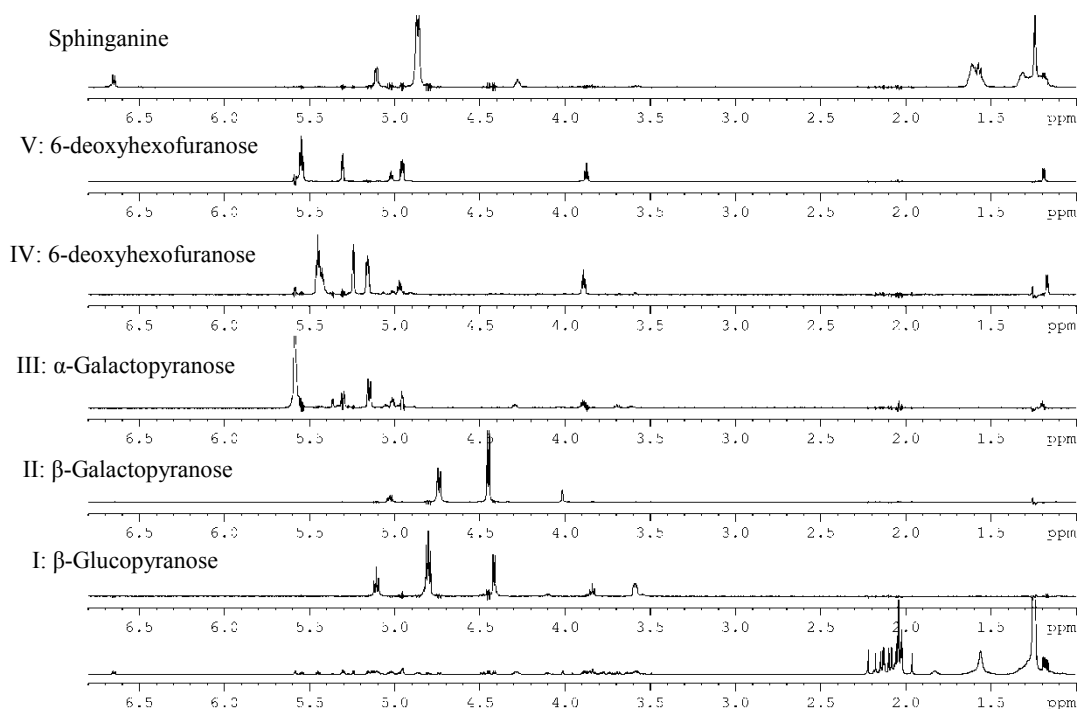


Figure 24: The spin systems of the five sugars spectrum of compound **1b** evidenced by the *z*-TOCSY spectrum (CDCl₃, 700MHz)

5.3 Experimental Section

Collection, extraction and isolation: Specimens of *Terpios* were treated as previously described for terpioside A (*Chapter 4*). The sponge homogenized was extracted with methanol and chloroform; the combined extracts were partitioned between H₂O and *n*-BuOH. The organic layer was purified by subsequent chromatography on RP-18 silica gel column and SiO₂. This procedure allowed to obtain a fraction mainly composed of glycolipids. This fraction was peracetylated with Ac₂O in pyridine for 12 h. The acetylated glycolipids were subjected to HPLC separation on an SiO₂ column [eluent: *n*-hexane/EtOAc (6:4)], thus affording a fraction which was further chromatographed with HPLC on a SiO₂ column [eluent: *n*-hexane/*iso*-PrOH (8:2)] affording to 7 mg of pure terpioside peracetate (**1b**).

Terpioside B peracetate (1b): The peracetylated derivative **1b** was obtained as colorless oil. ¹H- and ¹³C-NMR are reported in **Table 4**.

Table 4: ^1H and ^{13}C NMR data of terpioside B peracetate (**1b**)

| Position | | δ_{H} (mult, J (Hz)) | δ_{C} (mult) |
|-----------------|------|--------------------------------------|----------------------------|
| 1 | a | 3.85 ^a | 66.6 (CH ₂) |
| | b | 3.57 (dd, 10.8, 3.5) | |
| 2 | | 4.28 (m) | 47.8 (CH) |
| 2-NH | | 6.65 (d, 8.9) | |
| 3 | | 5.1 (dd, 8.3, 2.5) | 71.3 (CH) |
| 4 | | 4.86 (dt, 10.0, 2.9, 2.9) | 73.0 (CH) |
| 5 | a, b | 1.58 ^a | 28.0 (CH ₂) |
| 1' | | 4.41 (d, 7.8) | 100.2 (CH) |
| 2' | | 4.80 (dd, 9.5, 7.8) | 71.1 (CH) |
| 3' | | 5.1 (t, 9.5) | 72.1 (CH) |
| 4' | | 3.84 (t, 9.5) | 75.5 (CH) |
| 5' | | 3.58 ^a | 72.7 (CH) |
| 6' | a | 4.48 (dd, 12.1, 1.7) | 61.9 (CH ₂) |
| | b | 4.10 (dd, 12.1, 5.5) | |
| 1'' | | 4.44 (d, 8.1) | 100.7 (CH) |
| 2'' | | 5.03 (dd, 10.7, 8.1) | 68.8 (CH) |
| 3'' | | 4.74 (dd, 10.7, 2.6) | 72.8 (CH) |
| 4'' | | 4.01 (br. d, 2.6) | 77.0 (CH) |
| 5'' | | 3.64 (m) | 73.9 (CH) |
| 6'' | a | 3.78 (dd, 11.6, 4.3) | 65.9 (CH ₂) |
| | b | 3.73 (dd, 11.6, 6.5) | |
| 1''' | | 4.95 (d, 3.3) | 98.8 (CH) |
| 2''' | | 5.15 (dd, 11.1, 3.3) | 68.7 (CH) |
| 3''' | | 5.30 (dd, 11.1, 1.8) | 67.2 (CH) |
| 4''' | | 5.58 (br. d, 3.2) | 67.7 (CH) |
| 5''' | | 4.29 (m) | 66.9 (CH) |
| 6''' | a | 3.60 (dd, 10.4, 5.5) | 64.4 (CH ₂) |
| | b | 3.69 (br. t, 10.4) | |
| 1 ^{IV} | | 5.24 (d, 4.5) | 100.4 (CH) |
| 2 ^{IV} | | 5.15 (dd, 6.7, 4.5) | 75.9 (CH) |
| 3 ^{IV} | | 5.45 (t, 6.7) | 74.2 (CH) |
| 4 ^{IV} | | 3.89 (t, 6.7) | 81.2 (CH) |
| 5 ^{IV} | | 4.97 (t, 6.4) | 71.0 (CH) |
| 6 ^{IV} | | 1.17 (d, 6.4) | 15.3 (CH ₃) |
| 1 ^V | | 5.30 (d, 4.6) | 98.7 (CH) |
| 2 ^V | | 4.95 (dd, 6.9, 4.6) | 76.8 (CH) |
| 3 ^V | | 5.54 (t, 6.9) | 73.5 (CH) |
| 4 ^V | | 3.87 (t, 6.9) | 80.7 (CH) |
| 5 ^V | | 5.02 (t, 6.9) | 70.2 (CH) |
| 6 ^V | | 1.19 (d, 6.5) | 15.6 (CH ₃) |
| 1 ^{VI} | CO | | 169.7 |
| 2 ^{VI} | | 5.13 ^a | 73.8 (CH) |
| 3 ^{VI} | a, b | 1.83 (m) | 31.6 (CH ₂) |

| | | | |
|--------------|-----------------|-------------|---------------|
| <i>Ac</i> 's | CH ₃ | 2.22 – 1.96 | 20.1 – 20.3 |
| | CO | – | 169.8 – 170.2 |

- a. Overlapped signals
- b. Additional ¹H signals: $\delta = 1.24$ (br., alkyl chain protons), 0.88 (t, $J = 7.0$ Hz, *n*-chain Me groups), 0.86 (d, $J = 6.7$ Hz, *iso*-chain Me groups) ppm.
- c. Additional ¹³C signals: $\delta = 29.4$ (CH₂, $\omega-2$), 22.7 (CH₂, $\omega-1$), 22.7 (CH₃, *iso*-chain Me groups), 14.1 (CH₃, ω) ppm

Chapter 6

Amphiceramides A and B from the Sponge *Amphimedon compressa*

Chemical analysis of the Caribbean sponge *Amphimedon compressa* showed it to contain two novel glycosphingolipids, amphiceramide A (**1**) and B (**2**), possessing an unusual Δ^6 -phytosphingosine. The saccharide chain of amphiceramide A is composed of a β -glucose residue glycosylated at position 6 by an *N*-acetyl- β -glucosamine. This diglycoside has never been found before in a natural product. The saccharide chain of amphiceramide B consists of an allolactose [Gal(1 β →6)Glc] residue β -linked to the ceramide. In addition, the sponge contains a new molecular species of acetamidoglucosyl ceramide (**3**), and the known glucosyl ceramide **4** (halicerebroside A)¹ and melibiosyl ceramide **5** (amphimelibioside C).² The elucidation of the stereostructure of amphiceramides has been accomplished by a combination of extensive one- and two-dimensional NMR spectroscopy, MS and MS/MS spectrometry, and chemical degradation.

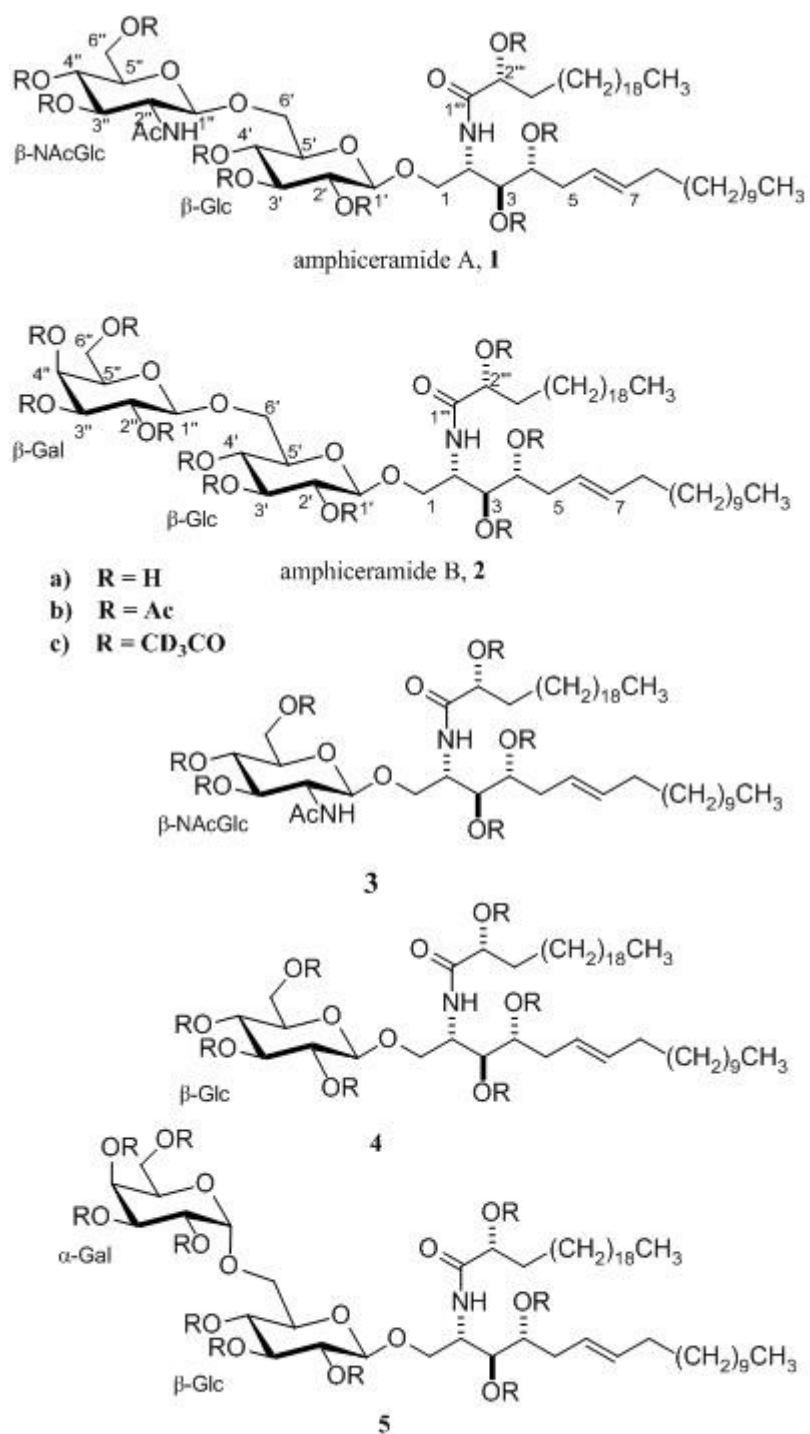


Figure 25: Glycosphingolipids from *Amphimedon compressa*

6.1 Isolation of Amphiceramides

Following our usual procedure, the BuOH soluble fraction obtained from the combined chloroform and methanol extracts of *A. compressa* was subjected first to reversed-phase and then to normal-phase silica gel column chromatography, to get a fraction mainly composed of glycolipids. The fraction was acetylated with Ac₂O in pyridine to give a peracetylated fraction, which is easier to purify by HPLC on a SiO₂ column. The HPLC separation yielded five peracetylated glycolipids, each isolated as a mixture of homologues. Unlike what we have observed in other sponges, these mixtures were relatively simple, with one single major homologue in each fraction, as judged by the respective MS and ¹H NMR spectra. Therefore, the five glycolipid mixtures were subjected to reversed-phase HPLC separation, which yielded the five pure peracetylated glycolipids **1b-5b**.

As for terpioside A, to distinguish any acetyl group possibly present in the natural glycolipid from those introduced during the acetylation reaction, the isolation procedure was repeated using deuterated acetic anhydride in the acetylation step, giving the pertrideuteroacetylated derivatives **1c-5c**. The ¹H NMR spectra of compounds **2c**, **4c**, and **5c** were identical to those of the respective peracetylated derivatives, except that no acetyl methyl singlet was present in the spectrum. This showed that all the acetyl groups came from the acetylation reaction. The ¹H NMR spectrum of compound **1c** showed one acetyl methyl singlet, indicating that one acetyl group is present in the natural compound **1a**. The subsequent structure elucidation work showed that this acetyl group is part of an amide function. Likewise, the ¹H

NMR spectrum of compound **3c** also contained one acetyl signal, showing that the natural compound **3a** contains an acetamido group. Finally, the peracetylated glycolipids were deacetylated using MeONa/MeOH (the acetamido groups present in compounds **1a** and **3a** survive these reaction conditions) to give the respective natural glycolipids **1a-5a**.

6.2 Structure Determination of Amphiceramide A

The positive-ion ESI mass spectrum of amphiceramide A (**1a**) showed a sodiated pseudomolecular ion peak at m/z 1041, in accordance with the molecular formula $C_{72}H_{120}N_2NaO_{24}^+$. This was confirmed by a high-resolution measurement at m/z 1041.7172. The ESI mass spectrum of the peracetyl derivative **1b** showed an $[M + Na]^+$ pseudomolecular ion peak at m/z 1419, indicating the presence of 9 additional acetyl groups. Compound **1b** was used for the subsequent NMR analysis.

The general 1H NMR spectral features of compound **1b** showed it to be a glycosphingolipid: the NH doublet at $\delta = 6.94$ of the ceramide amide function, a series of overlapped signals in the region between δ 5.4 and 3.4, nine signals of acetyl groups around δ 2.0, an intense signal at δ 1.24 indicative of a long aliphatic chain, and 6H methyl triplet at δ 0.88 as the sole signal in the high-field region of the spectrum, indicating that the alkyl chains of the ceramide were both unbranched. A second amide NH doublet at δ 6.48 suggested the presence of an aminosugar in the molecule.

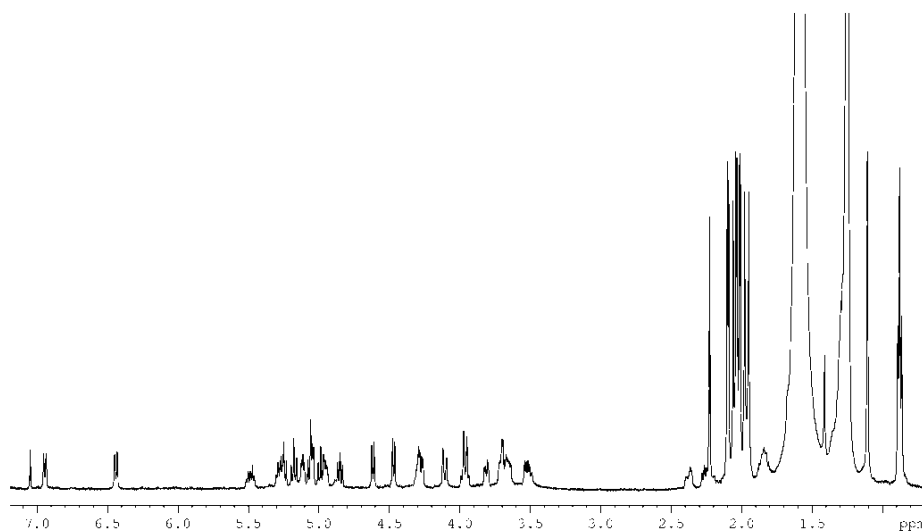


Figure 26: ^1H NMR spectrum of amphiceramide A (**1b**) in CDCl_3 at 700MHz

The absence of the characteristic triplet of the fatty acid α -protons at δ 2.3 suggested the presence in the ceramide moiety of an α -hydroxy acid, as usually found in GSLs from marine sponges. In addition, the presence of two methine groups resonating at δ 134.9 and 123.7 evinced from the ^{13}C NMR and HSQC spectra indicated a disubstituted double bond. Starting from the NH doublet at δ 6.94 (2-H), the COSY spectrum allowed sequential assignment of all the protons of the sphingosine. In particular, the scalar coupling of the two diastereotopic methylene protons at δ 2.37 and 2.24 (5- H_2) with the oxymethine proton at δ 4.94 (4-H) and the olefinic proton at δ 5.24 (6-H), located the double bond on the sphingosine at position 6. In addition, the coupling constant between 6-H and 7-H (15.2 Hz) indicated the *E* stereochemistry of the double bond. The HMBC correlation of coupling of 2-H with C-1^{III} (δ 170.2) and C-2 (δ 48.9) confirmed the amide linkage between the sphingosine and the fatty acid.

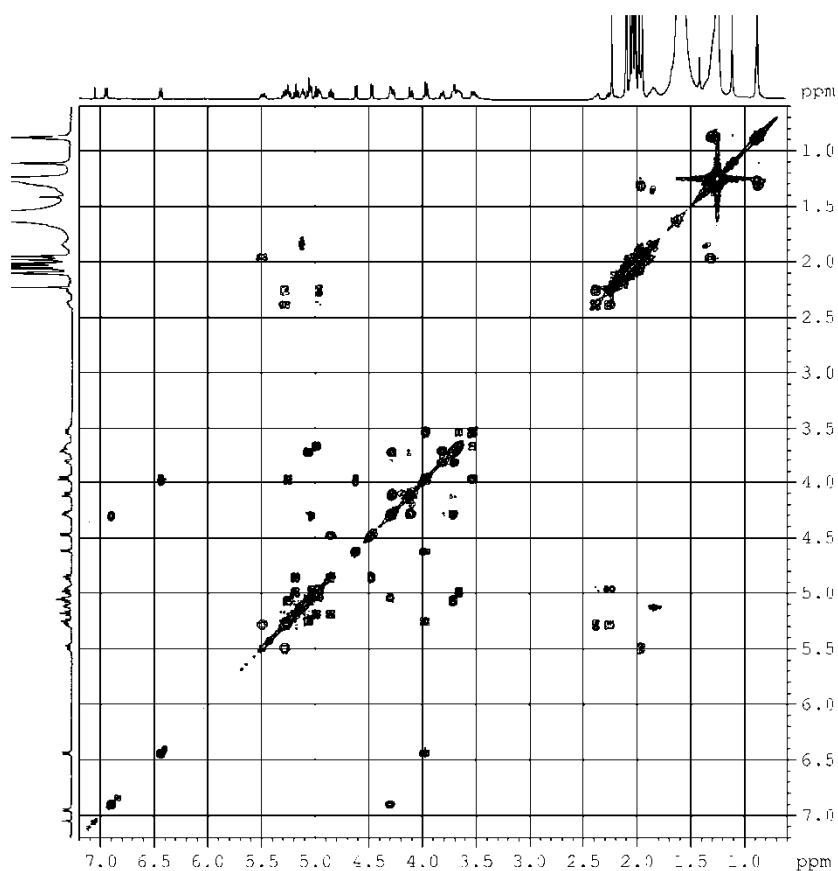


Figure 27: 2D NMR COSY spectrum of amphiceramide A (**1b**) in CDCl_3 at 500MHz

The key step in structural elucidation of a GSL is the identification of the nature of the sugars composing the saccharide chain. This can be achieved collecting data from COSY, TOCSY and HSQC to identify the spin system of each sugar unit and to assign the protons within each spin system. If the sugars are in the pyranose form, the nature of each sugar can then be easily identified by coupling constants analysis. Finally, identification of the positions involved in the sugar-ceramide and in the sugar-sugar linkages is based on HMBC and/or ROESY data.

Following this protocol, the two anomeric protons (δ 4.61 and 4.46) were identified from their HMQC correlation peaks with the respective

anomeric carbons signals, resonating at δ 101.4 and 100.1 in the ^{13}C NMR spectrum.

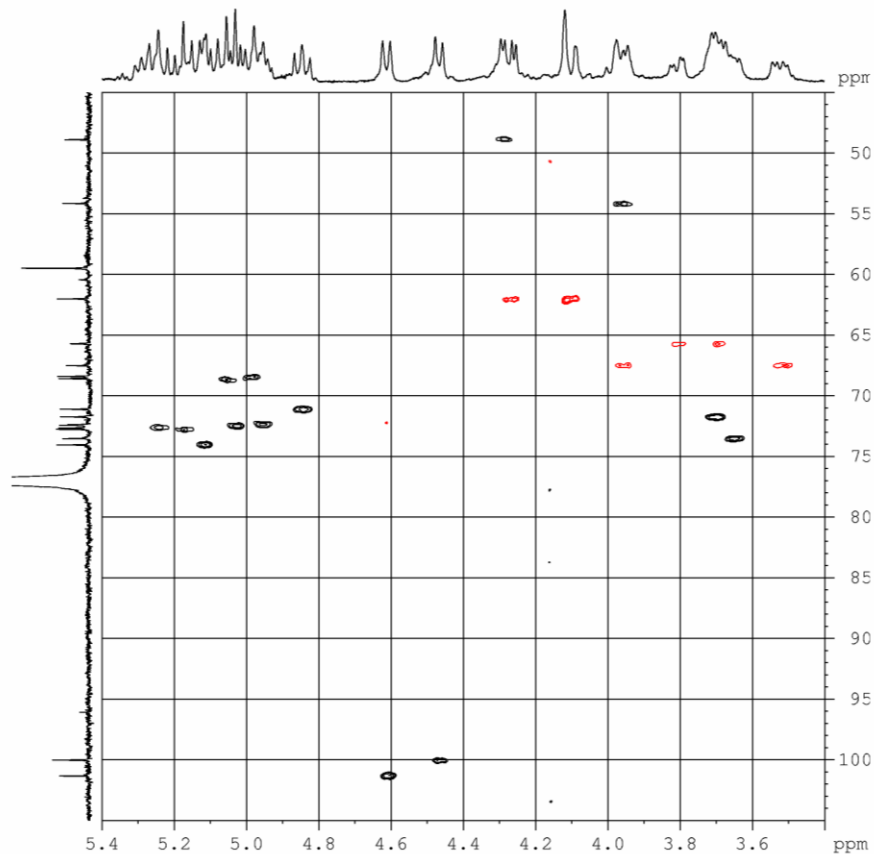


Figure 28: 2D NMR HSQC spectrum of ampiceramide A (**1b**) in CDCl_3 at 700MHz

Then, starting from each anomeric proton resonance, each sugar spin system was assigned using the TOCSY and COSY spectra. Both sugar residues contain, in addition to the anomeric CH, four methine and one methylene groups, and are therefore hexoses. The upfield chemical shift of 5^{I}-H and 5^{II}-H showed that the respective oxymethine groups are not involved in an ester function, indicating that the two sugar residues are both in the pyranose form. Coupling constant analysis allowed to establish that

all the ring protons of the sugar with the anomeric proton at δ 4.46 are axial, and therefore that this sugar residue is a β -glucopyranoside.

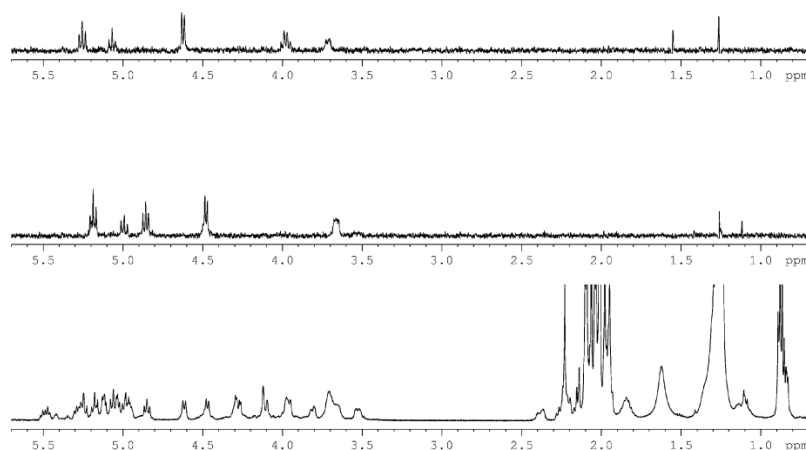


Figure 29: The spin systems of the two sugars spectrum of compound **1b** evidenced by the *z*-TOCSY spectrum (CDCl₃, 500MHz)

As for the second unit, the chemical shift values of C-2^{II} (δ 54.2) and 2^{II}-H (δ 3.95) were indicative of an *N*-linked methine carbon, and this was confirmed by the coupling of 2^{II}-H with the NH doublet at δ 6.48. Coupling constant analysis showed that also for this sugar the ring protons are all axial, and therefore the sugar residue was identified as a 2-amino-2-deoxyglucopyranoside. Finally, the HMBC correlation peak of 1^I-H with C-1 showed that the glucose was directly linked to ceramide, while that of 1^{II}-H with C-6^I showed that the glucosamine is glycosylating the glucose at position 6^I. Once the gross structure of the peracetyl derivative **1b** was elucidated, NMR spectra of **1a** were collected and fully assigned, providing further support to the proposed structure (Experimental Part).

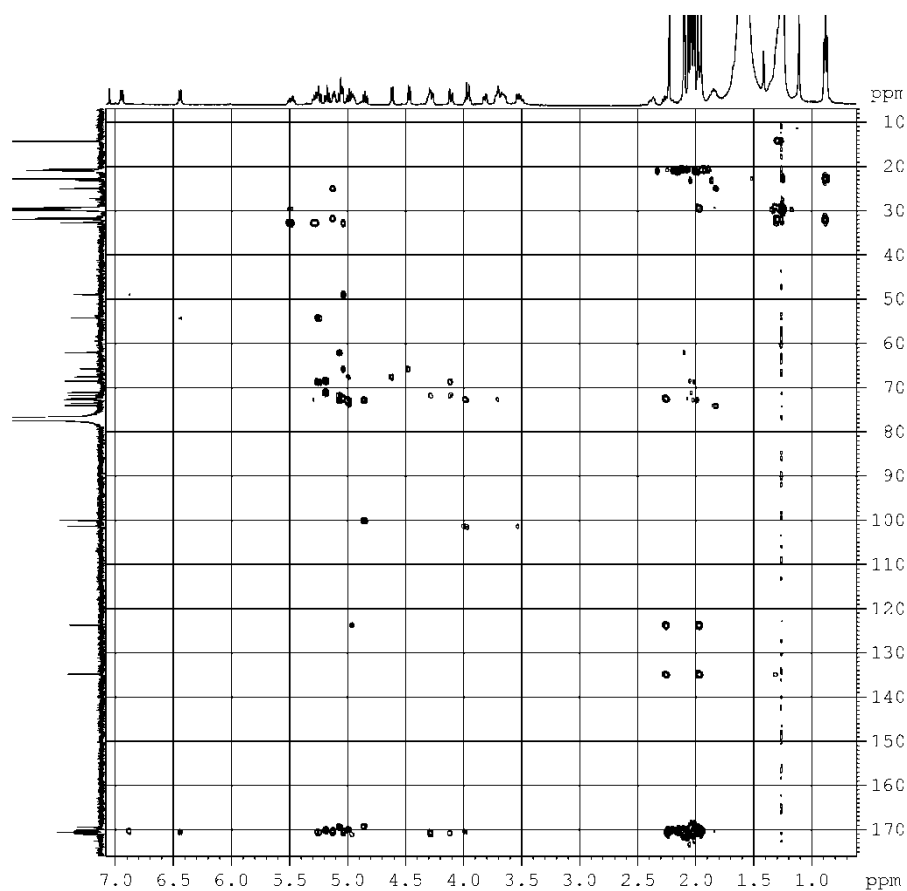
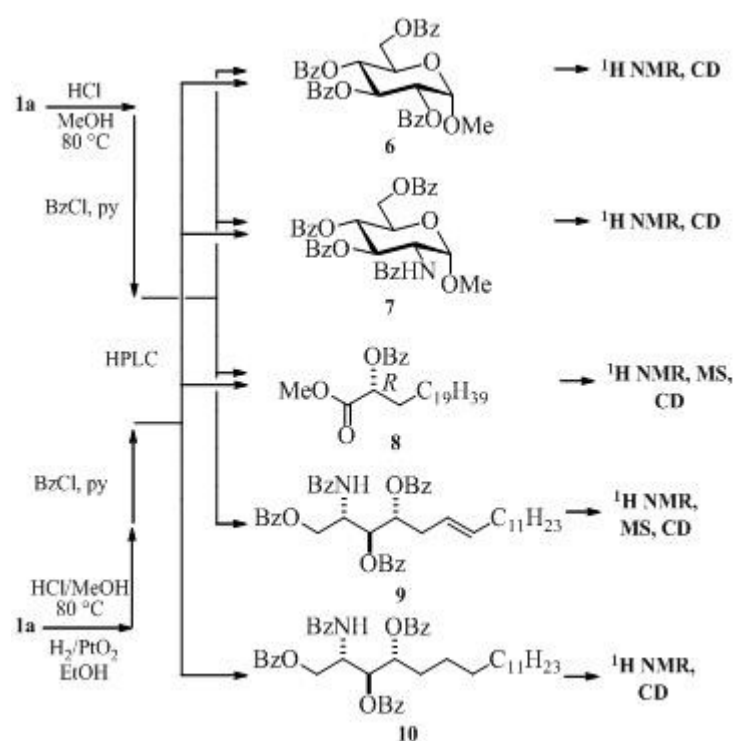


Figure 30: 2D HMBC spectrum of ampiceramide A (**1b**) in CDCl_3 at 700MHz

The absolute configuration of the two sugars, the configuration at C-2^{III} of the α -hydroxy fatty acid, and the configuration at C-2, C-3 and C-4 of the sphingosine were determined by chemical degradation. We used a simplified version of the micro-scale procedure we have developed for the analysis of glycolipids isolated as mixtures of homologues.³ Compound **1a** was subjected to acidic methanolysis, and the reaction products were perbenzoylated and subjected to HPLC, giving compounds **6–9**. The methyl glycoside **6** was identified as methyl tetra-*O*-benzoyl- α -D-glucopyranoside by comparison of its ^1H NMR and CD spectra with those reported in the literature.⁴ The methyl glycoside **7** was identified as methyl tri-*O*-benzoyl-2-benzamido-2-deoxy- α -D-glucopyranoside by comparison of its ^1H NMR

and CD spectra with those of an authentic sample prepared from *N*-acetyl-D-glucosamine with the same procedure used for compound **1a**. The fatty ester **8** was identified as methyl (*R*)-2-benzoyloxycosanoate from the EI mass spectrum (M^+ at m/z 474) and from the ^1H NMR and CD spectra, which matched those reported. The stereochemistry of the perbenzoylated Δ^6 -phytosphingosine **9** could not be determined directly from its spectra because reference data were not available in the literature.



Scheme 2: Microscale degradation analysis of Amphiceramide A (**1a**)

In the ^1H -NMR spectrum of **9**, chemical shifts and coupling constant of protons at C-1 through C-4 were close to those of *d-ribo*-phytosphingosine perbenzoate (**10**), but the presence of the additional double bond significantly affected the CD spectrum of **9**, which was remarkably different from that of the reference compound, and did not allow a reliable assignment of its configuration. Therefore, compound **1a** was subjected to

catalytic reduction with H_2/PtO_2 , and the reaction product was subjected to the degradation procedure described above. HPLC separation gave a compound whose $^1\text{H-NMR}$ and CD spectra were identical to those of *ribo*-phytosphingosine perbenzoate **10**, thus defining also the configuration of the the perbenzoylated Δ^6 -phytosphingosine **9**.

Structure **1a** was further supported by an MS/MS spectrum, carried out by dissolving the sample in MeOH with 1 mM LiCl as previously reported and using the $[\text{M} + \text{Li}]^+$ pseudomolecular ion at m/z 1025 as the precursor ion. The spectrum contained three peaks at m/z 660 and 687 and 822 arising from, respectively, the loss of the dehydrated whole saccharide chain, the loss of a C_{22} 2-hydroxyacyl group and the loss of the dehydrated *N*-acetylglucosamine.

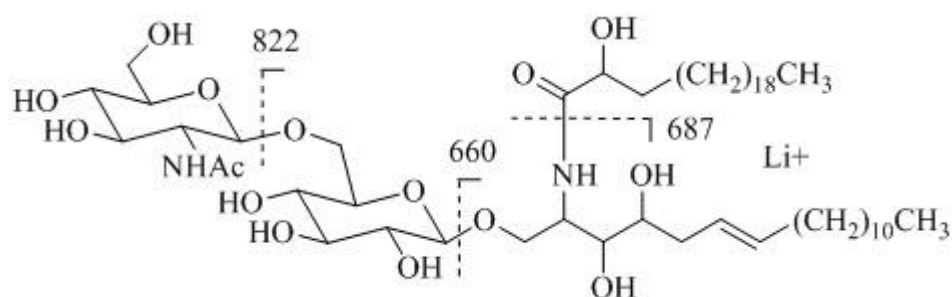


Figure 31: MS/MS fragmentation pathways of the lithiated adduct of (**1a**)

Table 5: ^1H and ^{13}C NMR spectroscopic data of amphiceramide A **1b** (CDCl_3)

| Position | δ_{H} (mult., J [Hz]) | δ_{C} (mult.) |
|-----------------|---------------------------------------------------------|-----------------------------|
| 1 | a 3.80 (dd, 10.9, 4.0) | 65.8 (CH_2) |
| a O | b 3.69 (dd, 10.9, 4.2) | |
| 2 v | 4.28 ^a | 48.9 (CH) |
| 2-NH | 6.94 (d, 8.4) | - |
| 3 r | 5.04 (dd, 6.2, 4.9) | 72.5 (CH) |
| 4 l | 4.94 (ddd, 8.9, 4.9, 3.5) | 72.4 (CH) |
| 5 a | a 2.37 (m) | 32.7 (CH_2) |
| | b 2.24 ^a | |
| 6 p | 5.24 (ddd, 15.2, 6.9, 6.9) | 123.7 (CH) |
| 7 p | 5.48 (ddd, 15.2, 6.8, 6.8) | 134.9 (CH) |
| 8 e | 1.96 ^a | 32.6 (CH_2) |
| 9-15, 17, 19''' | 1.24 ^a | 29.7 (CH_2) |
| 16, 20''' | 1.25 ^a | 31.9 (CH_2) |
| 17, 31''' | 1.28 ^a | 22.7 (CH_2) |
| 18, 22''' | 0.88 (t, 6.8) | 14.1 (CH_3) |
| 1' g | 4.46 (d, 8.0) | 100.1 (CH) |
| 2' n | 4.85 (dd, 9.6, 8.0) | 71.2 (CH) |
| 3' a | 5.18 (t, 9.5) | 72.8 (CH) |
| 4' l | 4.98 (dd, 10.0, 9.5) | 68.5 (CH) |
| 5' s | 3.65 (ddd, 10.0, 5.5, 1.7) | 73.6 (CH) |
| 6' s | a 3.95 ^a | 67.6 (CH_2) |
| | b 3.52 (dd, 11.9, 5.5) | |
| 1'' | 4.61 (d, 8.4) | 101.4 (CH) |
| 2'' | 3.95 ^a | 54.2 (CH) |
| 2''-NH | 6.48 (d, 8.9) | - |
| 2''-Ac | 1.95 | 23.1 (CH_3) |
| | - | 170.7 (C) |
| 3'' | 5.24 (dd, 10.5, 9.4) | 72.7 (CH) |
| 4'' | 5.06 (dd, 10.0, 9.4) | 68.7 (CH) |
| 5'' | 3.70 ^a | 71.8 (CH) |
| 6'' | a 4.28 (dd, 12.3, 4.7) | 62.1 (CH_2) |
| | b 4.10 (dd, 12.3, 2.3) | |
| 1''' | - | 170.2 (C) |
| 2''' | 5.11 (dd, 7.5, 4.6) | 74.1 (CH) |
| 3''' | 1.84 (m) | 31.7 (CH_2) |
| Ac's | 2.22, 2.10, 2.09, 2.06, 2.04, 2.03, 2.02, 2.01, 1.98 | 21.1-20.6 (CH_3) |
| | - | 171.1-169.3 (C) |

a. Overlapped signals

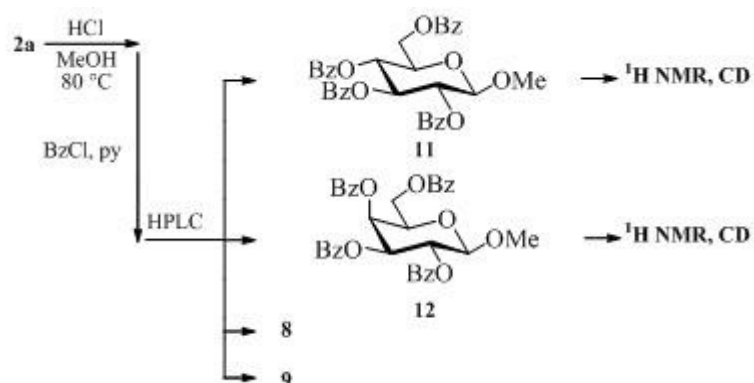
6.3 Structure Determination of Amphiceramide B

Compound **2a** showed a $[M + Na]^+$ pseudomolecular ion peak in the high resolution ESI mass spectrum at m/z 1000.6907, in accordance with the molecular formula $C_{53}H_{101}NO_{14}$. The 1H NMR spectrum of the peracetate **2b** was similar to that of **1b**, but only one NH doublet was present. This data suggested a diglycosylceramide with no aminosugar present in the structure. The ceramide part of the molecule appeared the same as that of **1b**, each proton showing chemical shift close to those observed in **1b**, and nearly identical coupling constants.

The structure of the saccharide chain of **2b** was studied with the same methods used for **1b**. The sugar linked to the ceramide is, again, a β -glucopyranoside, glycosylated at position 6¹ by the second sugar, which was shown to be a β -galactopyranoside on the basis of the following evidence. The 2^{II}-H proton showed a large coupling with both 1^{II}-H and 3^{II}-H, and these three protons are therefore axial. In contrast, the coupling constant between 3^{II}-H and 4^{II}-H (3.4 Hz) showed the latter proton to be equatorial. As for 5^{II}-H, its axial orientation was demonstrated by its intense correlation peak with the axial 3^{II}-H in the ROESY spectrum.

The absolute configuration of the sugar was determined with the same degradation scheme described for **1a** (Scheme 2). In this case, however, the benzoylated α -glycosides of glucose and galactose (which are the major products of the reaction sequence) coeluted from the HPLC column, and could not be obtained in the pure form. The corresponding β -glycosides **11** and **12**, although less abundant, gave rise to distinct peaks in the chromatogram. Structure and absolute configuration of compounds **11** and

12 were identified by comparison of their ^1H NMR and CD spectra with those reported in the literature.⁵



Scheme 3: Microscale degradation analysis of Amphiceramide B (**2a**)

Finally, the structure of the ceramide of **2a**, and in particular the length of the two alkyl chains, was confirmed by MS/MS spectrometry. The $[\text{M} + \text{Li}]^+$ pseudomolecular ion of **2a** at m/z 984 showed the same fragmentation pattern (**Figure 32**), with the loss of a C_{22} α -hydroxy acyl group leading to the the fragment ion at m/z 646.

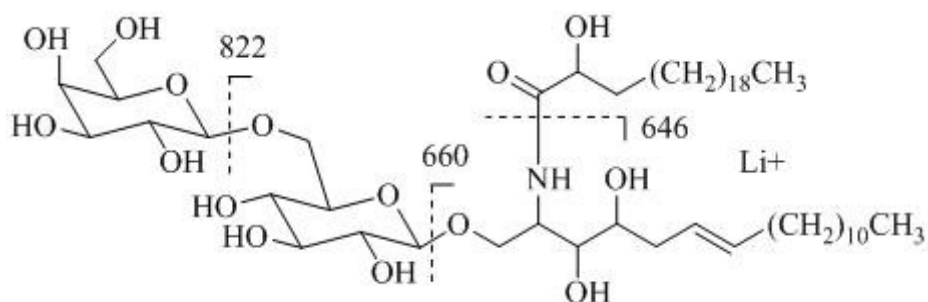


Figure 32: MS/MS fragmentation pathways of the lithiated adduct of (**2a**)

Table 6: ^1H and ^{13}C NMR spectroscopic data of ampiceramide B **2b** (CDCl_3)

| Position | | δ_{H} (mult., J [Hz]) | δ_{C} (mult.) |
|------------------|---|---------------------------------------|-----------------------------|
| 1 | a | 3.90 ^a | 66.9 (CH_2) |
| | b | 3.61 ^a | |
| 2 | | 4.32 (m) | 48.4 (CH) |
| 2-NH | | 6.88 (d, 8.9) | - |
| 3 | | 5.09 (dd, 7.3, 4.1) | 72.1 (CH) |
| 4 | | 4.91 (ddd, 9.3, 4.1, 3.6) | 72.6 (CH) |
| 5 | a | 2.36 (m) | 32.3 (CH_2) |
| | b | 2.24 ^a | |
| 6 | | 5.26 (ddd, 15.2, 6.7, 6.7) | 124.1 (CH) |
| 7 | | 5.47 (ddd, 15.2, 6.8, 6.8) | 134.8 (CH) |
| 8 | | 1.95 (m) | 32.7 (CH_2) |
| 9-15, 4'''-19''' | | 1.24 ^a | 29.8 (CH_2) |
| 16, 20''' | | 1.24 ^a | 32.0 (CH_2) |
| 17, 21''' | | 1.28 ^a | 22.8 (CH_2) |
| 18, 22''' | | 0.88 (t, 6.8) | 14.2 (CH_2) |
| 1' | | 4.44 (d, 8.0) | 100.6 (CH) |
| 2' | | 4.84 (dd, 9.7, 8.0) | 71.5 (CH) |
| 3' | | 5.15 (t, 9.6) | 72.9 (CH) |
| 4' | | 4.93 (t, 9.9, 9.5) | 69.0 (CH) |
| 5' | | 3.64 (m) | 73.4 (CH) |
| 6' | a | 3.88 (dd, 11.2, 1.7) | 68.0 (CH_2) |
| | b | 3.60 ^a | |
| 1'' | | 4.52 (d, 8.0) | 101.3 (CH) |
| 2'' | | 5.17 (t, 10.5, 8.0) | 68.8 (CH) |
| 2''-NH | | - | - |
| 2''-Ac | | - | - |
| | | - | - |
| 3'' | | 5.04 (dd, 10.5, 3.4) | 71.0 (CH) |
| 4'' | | 5.38 (dd, 3.4, 1.2) | 67.2 (CH) |
| 5'' | | 3.92 (ddd, 6.7, 6.7, 1.2) | 70.8 (CH) |
| 6'' | a | 4.20 (dd, 11.3, 6.5) | 61.3 (CH_2) |
| | b | 4.10 (dd, 11.3, 6.9) | |
| 1''' | | - | 169.8 (C) |
| 2''' | | 5.13 (dd, 7.3, 4.8) | 74.1 (CH) |
| 3''' | | 1.84 ^a | 31.9 (CH_2) |
| Ac's | | 2.22, 2.15, 2.09, 2.07, 2.06, | 21.1-20.7 (CH_3) |
| | | 2.04, 2.04, 2.02, 1.98, 1.98 | |
| | | - | 170.7-169.3 |

a. Overlapped signal

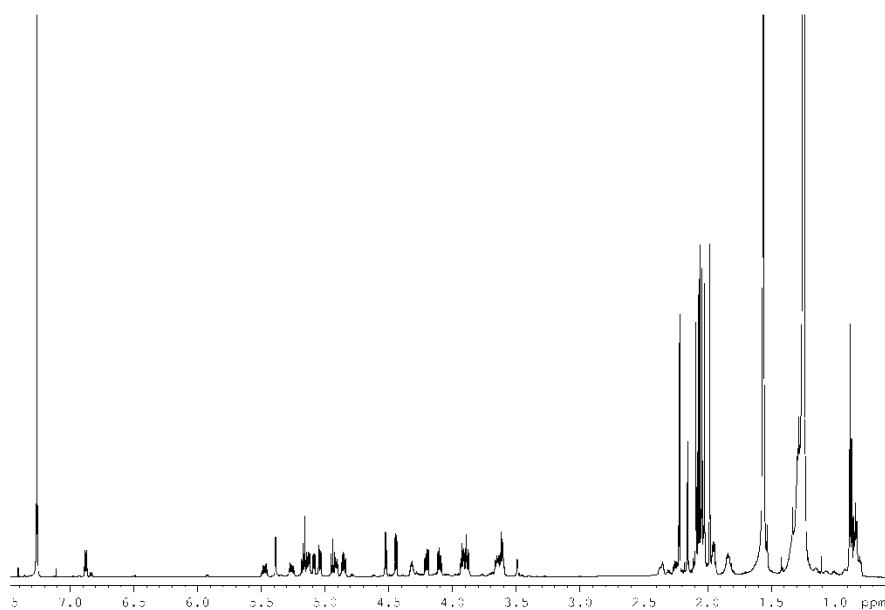


Figure 33: ^1H NMR spectrum of amphiceramide B peracetate (**2b**) in CDCl_3 at 700MHz

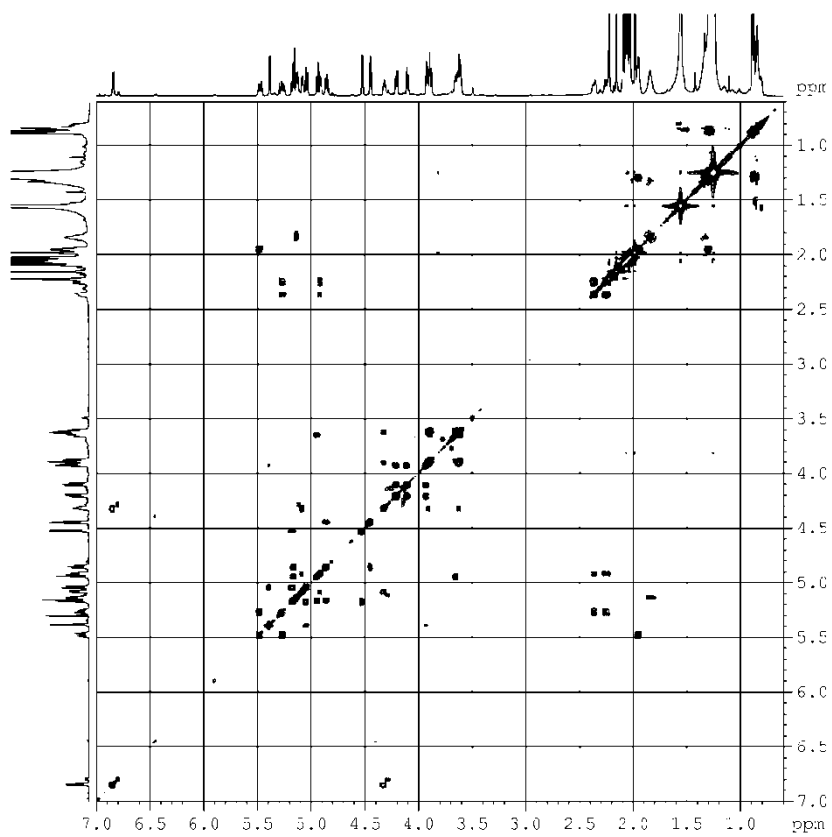


Figure 34: 2D NMR COSY spectrum of **2b** in CDCl_3 at 700MHz

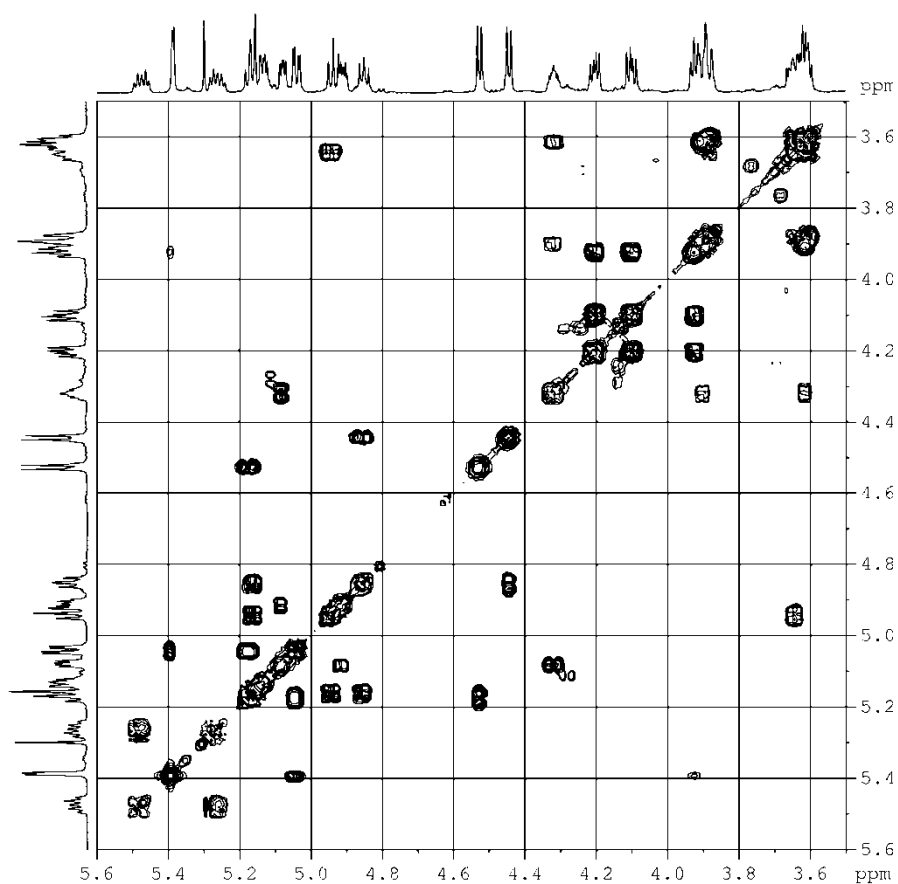


Figure 35: Middle-field 2D NMR COSY spectrum of peracetylated amphiceramide B in CDCl₃ at 700MHz

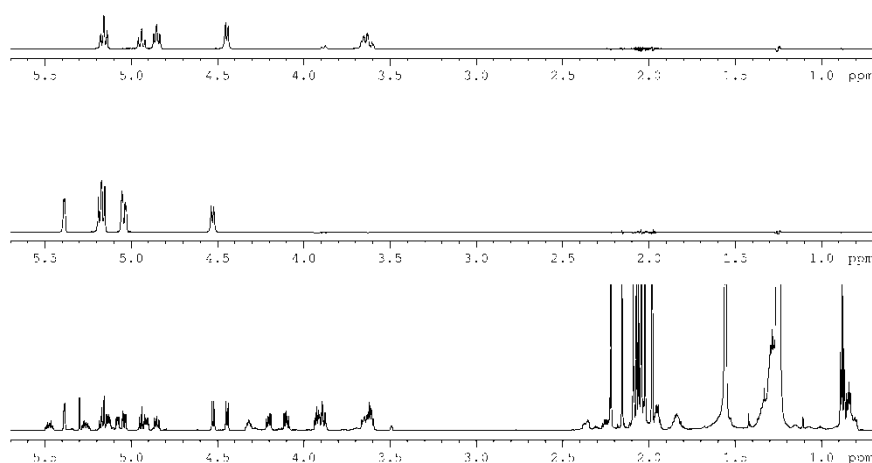


Figure 36: Spin systems in the Z – TOCSY spectrum of compound **2b** in CDCl₃ at 700MHz

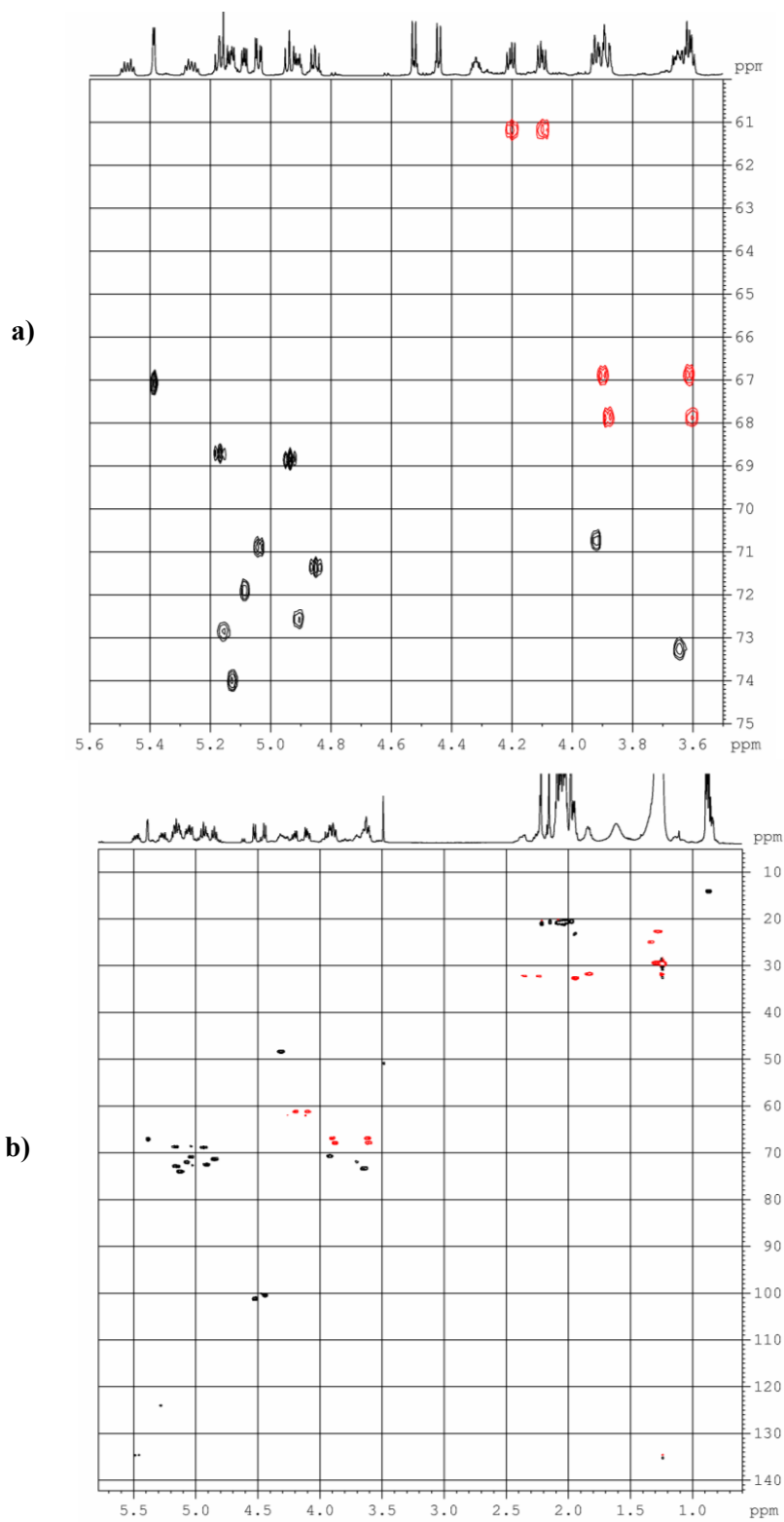
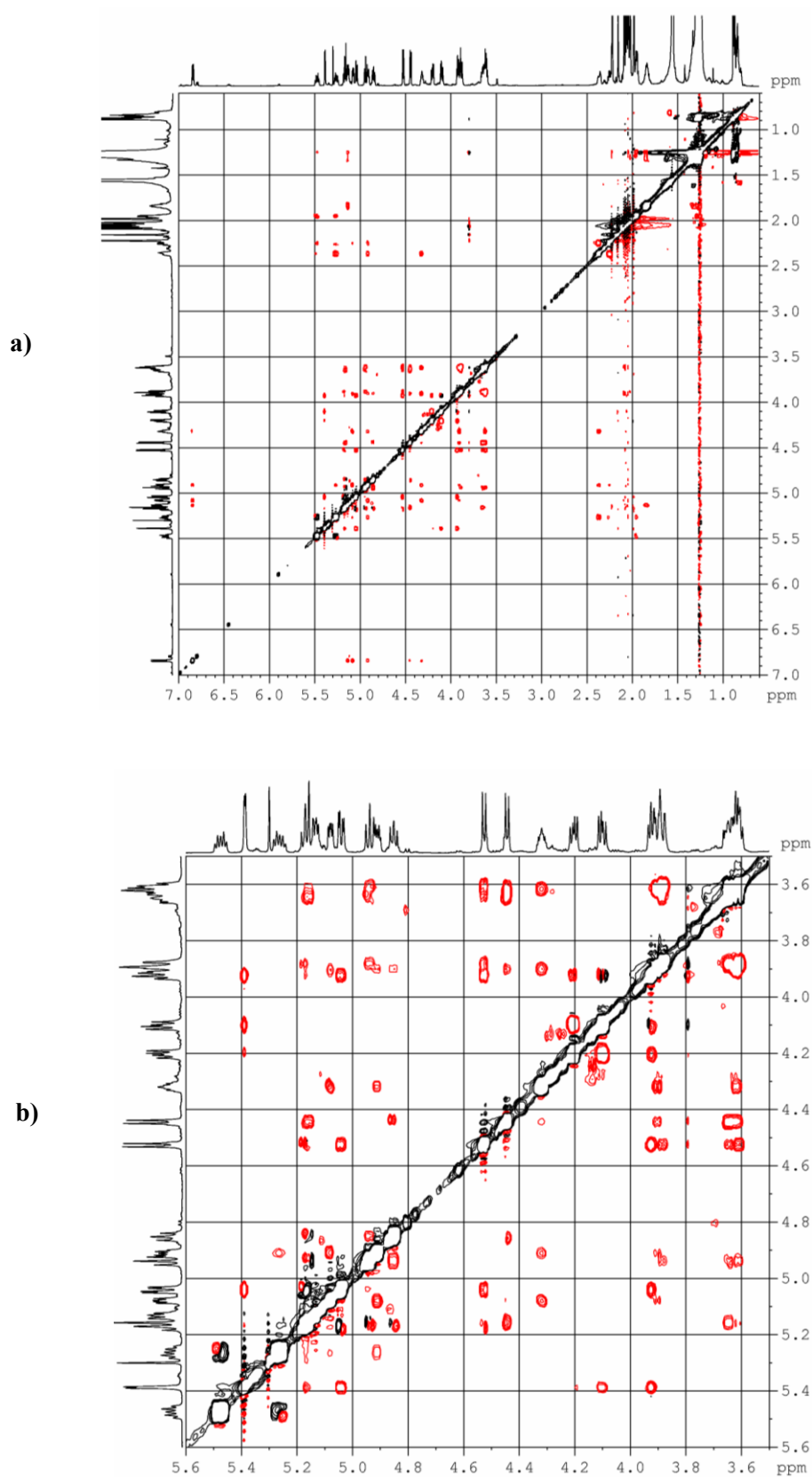


Figure 37a, b: 2D NMR HSQC spectra of **2b** in CDCl₃ at 700MHz and 500MHz



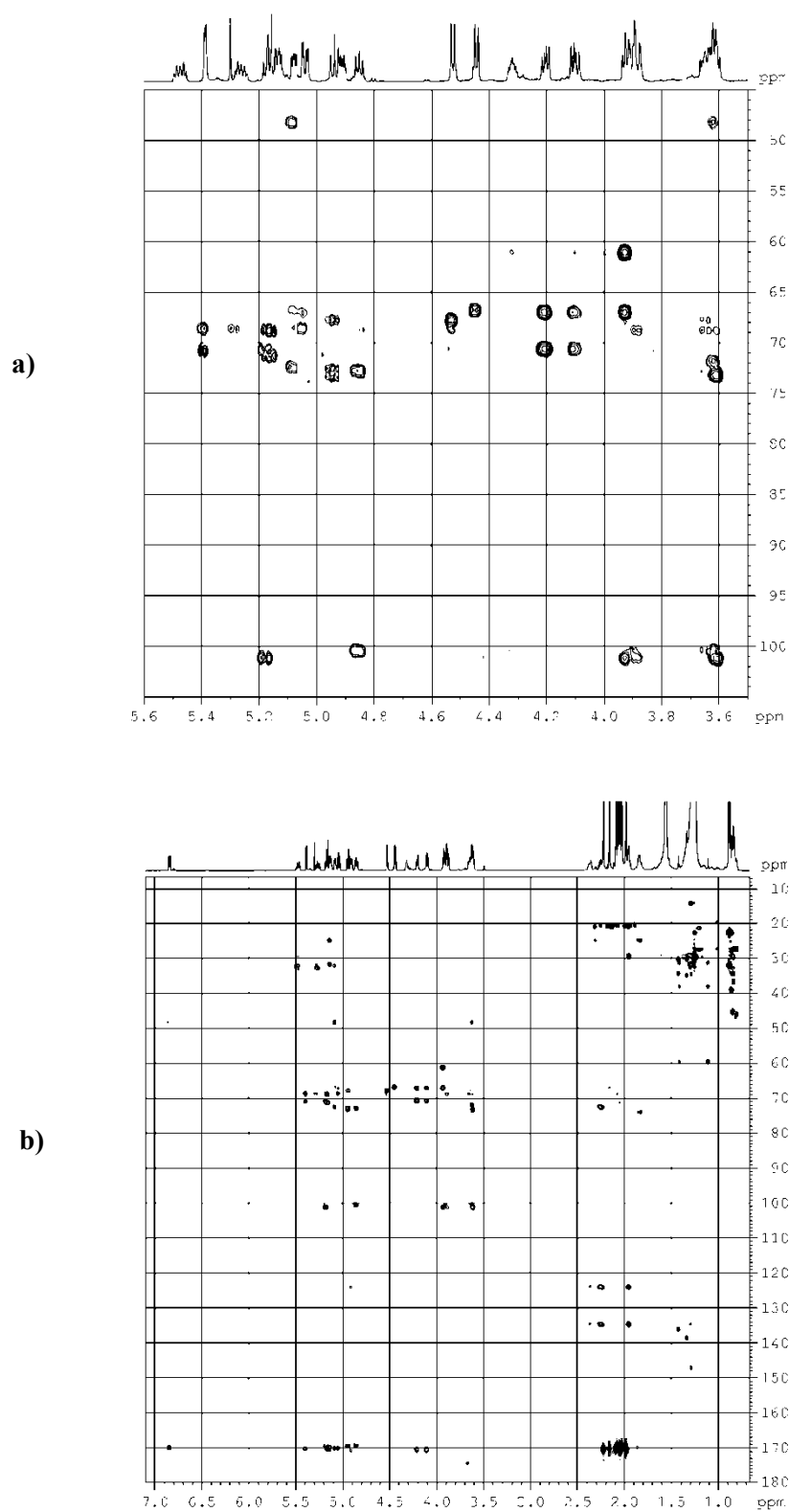


Figure 39 a,b: 2D NMR HMBC spectra of **2b** in $CDCl_3$ at 700MHz

6.4 Structure Determination of **3a**, **4a**, and **5a**

The ^1H and ^{13}C NMR data of the peracetate **3b** were very similar to those reported for ampicerebroside B heptaacetate, a monoglycosyl ceramide composed of a Δ^6 -unsaturated phytosphingosine and an *N*-acetyl- β -glucosamine. The only difference between the two compounds, the unbranched C_{18} sphingosine of **3b** instead of the *iso* C_{19} sphingosine of ampicerebroside B, was reflected in the 6H methyl triplet at δ 0.88 in the ^1H NMR spectrum of **3b**. This was confirmed by the mass spectrum of the natural compound **3a**, showing a lithiated $[\text{M} + \text{Li}]^+$ pseudomolecular ion at m/z 863 indicative of a C_{40} ceramide. The fragment ion at m/z 525 in the MS/MS spectrum ($[\text{M} + \text{Li} - \text{C}_{22}\text{H}_{42}\text{O}]^+$) indicated a C_{22} α -hydroxy fatty acid and, consequently, a C_{18} sphingosine.

The ESI mass spectrum of compound **4a** ($[\text{M} + \text{Na}]^+$ at m/z 838) was consistent with the molecular formula $\text{C}_{46}\text{H}_{89}\text{NO}_{10}$ and indicative of a monohexosylceramide with the same ceramide of compounds **1a** and **2a**. Examination of the ^1H NMR spectrum of the peracetate **4b** evidenced the presence of the signals of a β -glucopyranoside (chemical shift and coupling constants matched well those reported for a β -glucosylceramide peracetate), while the signals of the ceramide proton were very similar to those of **1b** and **2b**. Degradation analysis as described above lead to the formation of the methyl glycoside **6**, the fatty ester **8**, and the sphingosine **9**, thus defining completely structure **4a**. Compound **4a** had been previously reported as halicerebroside A from the Red Sea sponge *Halilona* sp, but due to the limited spectroscopic data present in the original paper identification by spectral comparison was not possible.

Finally, compound **5a** was identified by comparison of its MS, ^1H NMR, and ^{13}C NMR data with those reported for amphimelibioside C from the Japanese sponge *Amphimedon* sp.

6.5 Experimental Section

Collection, Extraction, and Isolation Procedures: Specimens of *Amphimedon compressa* were collected in the summer of 2005 along the coast of Key Largo (Florida). They were frozen immediately after collection and kept frozen until extraction. The sponge (290 g of dry weight after extraction) was homogenized and extracted with methanol (3·5 L) and then with chloroform (3·5 L); the combined extracts were partitioned between H_2O and *n*-BuOH. The organic layer was concentrated *in vacuo* and afforded 53.3 g of a dark oil, which was chromatographed on a column packed with RP-18 silica gel. A fraction eluted with CHCl_3 (4.7 g) was further chromatographed on a SiO_2 column, giving a fraction [1.01 g, eluent: EtOAc/MeOH (7:3)] mainly composed of glycolipids. This fraction was peracetylated with Ac_2O in pyridine for 12 h. The acetylated glycolipids were subjected to HPLC separation on an SiO_2 column [eluent: *n*-hexane/EtOAc (6:4)]. Five major fractions were obtained, each mainly composed of one of compounds **1b-5b**. Amphiceramide A (5.6 mg) and 4 Amphiceramide B (1.6 mg), were obtained in the pure form from the respective fractions after a further normal phase HPLC purification [eluent: *n*-hexane/*i*-PrOH (85:15)] followed by reversed-phase HPLC using MeOH as eluent.

Amphiceramide A peracetate (1b): Amorphous solid $[\alpha]_D^{25} = +2$ ($c = 1.5$, CHCl_3); ESIMS (positive ion mode, MeOH) m/z 1419 ($[\text{M} + \text{Na}]^+$). ^1H and ^{13}C NMR: see **Table 5**.

Amphiceramide B peracetate (2b): Amorphous solid. $[\alpha]_D^{25} = +4$ ($c = 0.5$, CHCl_3); ESIMS (positive ion mode, MeOH) m/z 1420 ($[\text{M} + \text{Na}]^+$). ^1H and ^{13}C NMR: see **Table 6**.

Compound (3b): Amorphous solid. ESIMS (positive ion mode, MeOH) m/z 1131 ($[\text{M} + \text{Na}]^+$). ^1H NMR (CDCl_3) $\delta = 0.88$ (n -chain Me groups), 1.28 (large band, alkyl chains), 1.94 (s, 3 H, Ac), 1.95 (overlapped, 2 H, 8- H_a , 8- H_b), 2.00-2.07 (several protons acetyl Me groups), 1.94 (s, 3 H, NH-Ac), 2.23 (overlapped, 1 H, 5- H_b), 2.34 (m, 1 H, 5- H_a), 3.48 (m, 1 H, 2'-H), 3.74 (overlapped, 2H, 1- H_a , 1- H_b), 3.74 (overlapped, 1 H, 5'-H), 4.10 (br. d, $J = 12.4$ Hz, 1 H, 6'- H_b), 4.24 (dd, $J = 12.4, 4.1$ Hz, 1 H, 6'- H_a), 4.30 (m, 1 H, 2-H), 4.85 (d, $J = 8.2$ Hz, 1 H, 1'-H), 4.95 (overlapped, 1 H, 3-H), 4.95 (overlapped, 1 H, 4-H), 5.00 (t, $J = 9.6$ Hz, 1 H, 4'-H), 5.07 (dd, $J = 4.5, 7.3$ Hz, 1 H, 2''-H), 5.23 (ddd, $J = 6.9, 15.2, 6.9$ Hz, 1 H, 6-H), 5.37 (t, $J = 9.6$ Hz, 1 H, 3'-H), 5.44 (ddd, $J = 6.6, 15.2, 6.6$ Hz, 1 H, 7-H), 6.07 (br. s, 1H, 2'-NH), 7.1 (d, $J = 8.6$ Hz, 1H, 2-NH) ^{13}C NMR (CDCl_3) $\delta = 14.2$ (CH_3 , 18-C, 22''-C), 20.5-23.2 (several CH_3 , acetyl Me groups), 22.7-31.8 (several CH_2 , alkyl chain), 32.7 (CH_2 , 5-C), 32.7 (CH_2 , 8-C), 47.8 (CH , 2-C), 55.5 (CH , 2'-C), 61.9 (CH_2 , 6'-C), 66.2 (CH_2 , 1-C), 68.8 (CH , 4'-C), 71.9 (CH , 5'-C), 72.2 (CH , 3'-C), 72.6 (CH , 4-C), 72.6 (CH , 3-C), 98.9 (CH , 1'-C), 123.7

(CH, 6-C), 134.7 (CH, 7-C), 170.1 (CO, 1"-C), 171.2 (CO, NH-Ac) 177.1-169.6 (several C, acetyl CO groups).

Deacetylation of compounds 1b-4b: Compound **1b**, **2b**, **3b** and **4b** were dissolved in MeOH (1 mL) and a solution of solution of MeONa in MeOH (0.4 M, 50 μ L) was added. The reactions were allowed to proceed at 25 °C for 18 h and then the reaction mixtures were dried under nitrogen and the residues partitioned between water and chloroform. After removal of the solvent, the organic layers gave the native glycosphingolipids (**1a**, 1.5 mg; **2a**, 0.9 mg; **3a**, 11.3 mg; **4a**, 2.8 mg).

Amphiceramide A (1a): Colorless solid $[\alpha]_D^{25} = -4$ (*c* 0.9, CHCl₃); HR-MS (ESI positive ion mode, MeOH) calcd. for [C₅₄H₁₀₂N₂NaO₁₅]⁺ 1041.71724. ¹H and ¹³C NMR: see **Table 7**.

Amphiceramide B (2a): Colorless solid $[\alpha]_D^{25} = -7$ (*c* 0.5, CHCl₃); HR-MS (ESI positive ion mode, MeOH) calcd. for [C₅₂H₉₉NNaO₁₅]⁺ 1000.69069. ¹H and ¹³C NMR: see **Table 8**.

Compound 3a: White solid; HR-MS (ESI positive ion mode, MeOH) calcd. for [C₄₈H₉₂N₂NaO₁₀]⁺ 879.66442. ¹H NMR (C₅D₅N) δ = 0.86 (t, *J* = 6.84 Hz, 3H, 18-H, 22"-H), 1.27 (overlapped, 2H, 9-H, 15-H), 1.27 (overlapped, 2H, 16-H, 20"-H), 1.27 (overlapped, 2-H, 17-H, 21"-H), 1.97 (overlapped, 1H, 3"-H_a), 2.02 (overlapped, 2H, 8-H), 2.18 (overlapped, 1H, 3"-H_b), 2.18 (overlapped, 3H, NH-Ac), 2.67 (multiplet, 1H, 5-H_b), 3.00 (multiplet, 1H, 5-

H_a), 3.82 (multiplet, 1H, 5'-H), 4.18 (t, $J = 9.2$ Hz, 1H, 4'-H), 4.23 (overlapped, 1H, 4-H), 4.24 (t, $J = 9.2$ Hz, 1H, 3'-H), 4.29 (overlapped, 1H, 6'-H_b), 4.29 (overlapped, 1H, 3-H), 4.44 (overlapped, 1H, 6'-H_a), 4.48 (multiplet, 1H, 2'-H), 4.54 (dd, $J = 10.9, 3.9$, 1H, 1-H_b), 4.59 (overlapped, 1H, 1-H_a), 4.59 (overlapped, 1H, 2''-H), 5.12 (d, $J = 8.3$, 1H, 1'-H), 5.24 (multiplet, 1H, 2-H), 5.70 (ddd, $J = 15.2, 6.6$ Hz, 1H, 7-H), 5.94 (ddd, $J = 15.2, 6.8$ Hz, 1H, 6-H), 8.5 (d, $J = 9.3$, 1H, 2'-NH), 8.85 (d, $J = 7.7$, 1H, 2-NH): ¹³C NMR (C₅D₅N) $\delta = 14.0$ (CH₃, 18-C, 22''-C), 22.8 (CH₂, 17-C, 21''-C), 23.2 (CH₃, NH-Ac), 29.7 (CH₂, 9-C, 15-C), 32.9 (CH₂, 8-C), 35.2 (CH₂, 3''-C_{a, b}), 36.7 (CH₂, 16-C, 20''-C), 37.4 (CH₂, 5-C_{a, b}), 50.4 (CH, 2-C), 57.5 (CH, 2'-C), 62.2 (CH₂, 6'-C_{a, b}), 68.6 (CH₂, 1-C_{a, b}), 72 (CH, 4'-C), 72.2 (CH, 2''-C), 72.5 (CH, 3'-C), 75.1 (CH, 3-C), 76.7 (CH, 4-C), 78.2 (CH, 5'-C), 101.7 (CH, 1'-C), 127.8 (CH, 6-C), 132.6 (CH, 7-C), 172.2 (CO, NH-Ac), 175.7 (C, 1''-C).

Methanolysis of Amphiceramide A (1b): Compound **1b** (1.2 mg) was dissolved in 1N HCl in 91% MeOH (500 μ L), and the solution obtained was kept at 80 °C for about 12 h. The reaction mixture was dried under nitrogen and benzoylated with benzoyl chloride (50 μ L) in pyridine (500 μ L) at 25° C for 16h. The reaction was then quenched with MeOH; after 30 min, the mixture was dried under nitrogen. Methyl benzoate was removed by keeping the residue under vacuum with an oil pump for 24 h. The residue was purified by HPLC [column: Luna SiO₂, 5 μ ; eluent: *n*-hexane/*i*PrOH (99:1); flow: 1 mL min⁻¹]. The chromatogram showed four peaks: methyl glycoside **6**, identified as methyl tetra-*O*-benzoyl- α -D-glucopyranoside ($t_R =$

10.8 min) by comparison of its ^1H NMR and CD spectra with those reported in the literature, methyl tri-*O*-benzoyl-2-benzamido-2-deoxy- α -glucopyranoside **7** ($t_{\text{R}} = 43.3$ min) identified by comparison of its ^1H NMR and CD spectra with those of an authentic sample, the benzoylated fatty acid methyl ester (**8**, $t_{\text{R}} = 4.0$ min), recognized as methyl (*R*)-2-benzoyloxydocosanoate because of its EI mass, ^1H NMR and CD spectra, corresponding to those reported, the perbenzoylated Δ^6 -phytosphingosine (**9**, $t_{\text{R}} = 10.5$ min).

(2*S*,3*S*,4*R*)-1,3,4-*O*-Benzoyl-2-benzoylaminoctadec-6-ene-1,3,4-triol (9):

^1H -NMR (CDCl_3): $\delta = 0.87$ (d, $J = 6.9$ Hz, 3H, 18- H_3), 1.34-1.16 (large band, alkyl chain), 1.88 (m, 2H, 8- H_2), 2.62 (m, 1H, 5- H_b), 2.74 (m, 1H, 5- H_a), 4.59 (dd, $J = 12.0$ and 6.1 Hz, 1H, 1- H_b), 4.67 (dd, $J = 12.0$ and 6.0 Hz, 1H, 1- H_a), 5.15 (m, 1H, 2-H), 5.41 (ddd, $J = 15.2, 7.2,$ and 7.2 Hz, 1H, 6-H or 7-H), 5.52 (ddd, $J = 15.2, 7.0,$ and 7.0 Hz, 1H, 7-H or 6-H), 5.56 (m, 1-H, H-4), 5.72 (t, $J = 5.2$ Hz, 1H, 3-H), 7.14 (1H, d, $J = 9.1$ Hz, NH-2), 7.33-7.63 (12 H, overlapping signals, benzoyl protons), 7.83 (2H, d, $J = 7.7$ Hz, benzoyl ortho protons), 7.93 (2H, d, $J = 7.7$ Hz, benzoyl ortho protons), 8.01 (2H, d, $J = 7.7$ Hz, benzoyl ortho protons), 8.04 (2H, d, $J = 7.7$ Hz, benzoyl ortho protons); CD (MeCN): $\lambda_{\text{max}} (\Delta\epsilon) = 250$ nm (-2), 222 nm (+8).

Methyl tri-*O*-benzoyl-2-benzamido-2-deoxy- α -glucopyranoside (7): *N*-

acetyl-*D*-glucosamine (11.2 mg) was subjected to acidic methanolysis as described above. The resulting methyl glycosides were benzoylated with benzoyl chloride (200 μL) in pyridine (2 mL) at 25 $^\circ\text{C}$ for 16 h. The reaction

was quenched with MeOH; after 30 min, the mixture was dried under nitrogen. Methyl benzoate was removed by keeping the residue under vacuum with an oil pump for 24 h. The residue was purified by HPLC [column: Luna SiO₂, 5 μ ; eluent: *n*-hexane/*i*PrOH (99:1); flow: 1 mLmin⁻¹; UV detector: 260 nm] and afforded the glycoside **7** (t_R = 6.6 min). ¹H NMR (CDCl₃): δ = 3.50 (s, 3H, OMe), 4.37 (ddd, J = 9.9, 4.9, and 2.8 Hz, 1H, 5-H), 4.48 (dd, J = 12.1 and 4.9 Hz, 1H, 6-H_b), 4.62 (dd, J = 12.1 and 2.8 Hz, 1H, 6-H_a), 4.76 (ddd, J = 10.6, 9.1, and 3.5 Hz, 1H, 2-H), 5.02 (d, J = 3.5 Hz, 1H, 1-H), 5.76 (t, J = 9.7 Hz, 1H, 4-H), 5.83 (dd, J = 10.6 and 9.5 Hz, 1H, 3-H), 6.62 (1H, d, J = 9.1 Hz, NH-2), 7.28-7.59 (12 H, overlapping signals, benzoyl protons), 7.68 (2H, d, J = 7.8 Hz, benzoyl ortho protons), 7.90 (2H, d, J = 7.9 Hz, benzoyl ortho protons), 7.93 (2H, d, J = 7.9 Hz, benzoyl ortho protons), 8.06 (2H, d, J = 8.0 Hz, benzoyl ortho protons); CD (MeCN): λ_{\max} ($\Delta\epsilon$) = 231 nm (+4).

Catalytic Reduction and Methanolysis of Compound 1a: Compound **1a** (0.5 mg) was dissolved in 1 mL of EtOH, and a small amount of PtO₂ was added. The suspension was kept under H₂ at 1 atm for 12 h, then filtered on TITAN PTFE (0.2 μ m), and taken to dryness. The residue was dissolved in 1 mL of 1 n HCl in 92% MeOH, and the obtained solution was kept for about 12 h at 80°C. The resulting mixture was dried under nitrogen and benzoylated with benzoyl chloride (100 μ L) and pyridine (1mL) and purified by HPLC [column: Luna SiO₂, 5 μ ; eluent: *n*-hexane/*i*-PrOH (99:1); flow: 1mL min⁻¹].

(2S,3S,4R)-1,3,4-O-Benzoyl-2-benzoylamino-1,3,4-octadecanetriol (10):
CD (MeCN): $\lambda_{\text{max}} = 233$ ($\Delta\epsilon = -1$), 222 ($\Delta\epsilon = +2$) nm. The ^1H NMR spectrum was identical to that of an authentic sample of D *-ribo-*phytosphingosine perbenzoate.⁶

Methanolysis of Amphiceramide B and analysis of methyl glycosides:

Amphiceramide B (**2b**, 0.7 mg) was subjected to acidic methanolysis as described above for compound **1b**. The reaction mixture was dried under nitrogen and benzoylated with benzoyl chloride (100 μL) and pyridine (1mL) at 25 °C for 16h. The reaction was quenched with MeOH and dried under nitrogen. The residue was then purified by HPLC [column: Luna SiO₂, 5 μ ; eluent: *n*-hexane/*i*PrOH (99:1); flow: 1 mLmin⁻¹; UV detector: 260 nm] and two peaks were identified as methyl tetra-O-benzoyl- α -D-glucopyranoside (**10**) ($t_{\text{R}} = 28.8$ min) and methyl tetra-O-benzoyl- α -D-galactopyranoside (**11**) ($t_{\text{R}} = 30.7$ min) by a comparison of their retention times, ^1H NMR spectra and CD spectra with those reported.

Table 7: ^1H and ^{13}C NMR spectroscopic data of ampicramid A **1a** (Py)

| Position | | δ_{H} (mult, J (Hz)) | δ_{C} (mult) |
|-----------|-----------------|--------------------------------------|----------------------------|
| 1 | a | 4.75 ^a | 69.8 (CH ₂) |
| | b | 4.50 (dd, 10.3, 3.5) | |
| 2 | | 5.27 ^a | 51.2 (CH) |
| 2-NH | | 8.55 (d, 8.9) | - |
| 3 | | 4.33 ^a | 75.1 (CH) |
| 4 | | 4.28 (m) | 72.4 (CH) |
| 5 | a | 3.02 (m) | 37.2 (CH ₂) |
| | b | 2.68 (m) | |
| 6 | | 5.97 (ddd, 6.9, 15.2, 6.9) | 127.8 |
| 7 | | 5.70 (ddd, 6.9, 15.2, 6.9) | 132.5 |
| 8 | | 2.03 (m) | 32.9 (CH ₂) |
| 9-15 | | 1.27 ^a | 29.8 (CH ₂) |
| 16, 20''' | | 1.31 ^a | 32.0 (CH ₂) |
| 17, 21''' | | 1.25 ^a | 22.7 (CH ₂) |
| 18, 22''' | | 0.86 ^a | 14.1 (CH ₃) |
| 1' | | 4.85 (d, 7.7) | 104.7 (CH) |
| 2' | | 3.93 ^a | 71.3 (CH) |
| 3' | | 4.11 (m) | 77.8 (CH) |
| 4' | | 3.94 ^a | 74.6 (CH) |
| 5' | | 3.95 ^a | 76.8 (CH) |
| 6' | a | 4.75 ^a | 69.7 (CH ₂) |
| | b | 4.14 (dd, 11.2, 6.3) | |
| 1'' | | 5.25 (d, 8.3) | 102.7 (CH) |
| 2'' | | 4.53 (br. t, 8.1) | 57.6 (CH) |
| 2''-NH | | 8.92 (d, 7.5) | |
| 3'' | | 4.40 (m) | 76.2 (CH) |
| 4'' | | 4.21 (m) | 72.0 (CH) |
| 5'' | | 3.91 ^a | 78.1 (CH) |
| 6'' | a | 4.47 (m) | 62.2 (CH ₂) |
| | b | 4.32 ^a | |
| 1''' | | - | 175.7 (C) |
| 2''' | | 4.57 (m) | 72.1 (CH) |
| 3''' | a | 2.18 ^a | 35.2 (CH ₂) |
| | b | 1.94 (m) | |
| 4''' | a | 1.67 (m) | 25.7 (CH ₂) |
| | b | 1.75 (m) | |
| NH-Ac | CO | - | 171.2 |
| | CH ₃ | 2.17 | 23.1 |

a. Overlapped signals

Table 8: ^1H and ^{13}C NMR spectroscopic data of amphiceramide B **2a** (Py)

| Position | | δ_{H} (mult, J (Hz)) | δ_{C} (mult) |
|-----------|------|--------------------------------------|----------------------------|
| 1 | a | 4.75 (dd, 10.3, 6.3) | 70.4 (CH ₂) |
| | b | 4.43 ^a | |
| 2 | | 5.26 ^a | 51.1 (CH) |
| 2-NH | | 8.57 (d, 10.4) | - |
| 3 | | 4.35 (m) | 75.2 (CH) |
| 4 | | 4.28 ^a | 72.5 (CH) |
| 5 | a | 3.02 (m) | 37.1 (CH ₂) |
| | b | 2.69 (m) | |
| 6 | | 5.98 (ddd, 6.9, 15.1, 6.9) | 128.0 (CH) |
| 7 | | 5.71 (ddd, 6.4, 15.1, 6.4) | 132.4 (CH) |
| 8 | | 2.03 (m) | 33.0 (CH ₂) |
| 9-15 | | 1.33 ^a | 29.7 (CH ₂) |
| 16, 20''' | | 1.20 ^a | 32.0 (CH ₂) |
| 17, 21''' | | 1.23 ^a | 22.7 (CH ₂) |
| 18, 22''' | | 0.84 (t, 5.1) | 14.1 (CH ₃) |
| 1' | | 4.85 (d, 7.7) | 105.2 (CH) |
| 2' | | 3.91 (d, 7.7) | 74.7 (CH) |
| 3' | | 4.10 ^a | 78.0 (CH) |
| 4' | | 4.06 (t, 5.9) | 71.2 (CH) |
| 5' | | 3.95 (br. t, 8.2) | 77.1 (CH) |
| 6' | a | 4.81 (br. d, 11.3) | 69.6 (CH ₂) |
| | b | 4.25 (dd, 11.3, 6.3) | |
| 1'' | | 4.99 (d, 7.8) | 105.7 (CH) |
| 2'' | | 4.49 (t, 7.8) | 72.4 (CH) |
| 3'' | | 4.17 (br. s) | 74.9 (CH) |
| 4'' | | 4.53 (m) | 70.1 (CH) |
| 5'' | | 4.06 (br. dd, 14.1, 6.9) | 76.7 (CH) |
| 6'' | a, b | 4.41 ^a | 62.2 (CH ₂) |
| 1''' | | - | 170.5 (C) |
| 2''' | | 4.58 (m) | 72.1 (CH) |
| 3''' | a, b | 1.97 ^a | 35.5 (CH ₂) |

a. Overlapped signals

References

1. S. Hirsch, Y. Kashman, *Tetrahedron* **1989**, *45*, 3897–3906.
2. C. Emura, R. Higuchi, T. Miyamoto, R.W.M. VanSoest, *J. Org. Chem.*, **2005**, *70*, 3031–3038.
3. V. Costantino, E. Fattorusso, C. Imperatore, A. Mangoni, *Eur. J. Org. Chem.* **2005**, 368–373.
4. V. Costantino, E. Fattorusso, C. Imperatore, A. Mangoni, S. Freigang, L. Teyton, *Bioorg. Med. Chem.* **2008**, *106*, 2077–2085.
5. V. Costantino, E. Fattorusso, C. Imperatore, A. Mangoni, *J. Org. Chem.*, **2008**, *73*, 6158–6165, and previous paper of the series "Glycolipids from sponges".
6. V. Costantino, E. Fattorusso, C. Imperatore, A. Mangoni, *J. Org. Chem.* **2004**, *69*, 1174–1179.

PART II

STUDY ON THE BIOSYNTHETIC PATHWAY OF PLAKORTIN FROM THE SPONGE PLAKORTIS SIMPLEX

Chapter 7

Sponge-microbe associations

Different types of Demospongiae, such as the Caribbean sponge *Plakortis simplex*, are associated with endosymbiotic micro-organisms and these symbionts contribute to 38-57% of the total sponge biomass.¹ Bacteria and fungi are mostly present in the mesohyl, which contains heterotrophic (eubacteria, archaea) and some autotrophic bacteria. As sponge-microbe interactions are widespread and some bacteria seem to be specific and permanently associated with the sponges, the existence of sponge-bacteria symbiosis is well established. The sponge-symbiont relationship can be categorized into obligatory mutualism (i.e. the symbiont play an essential role in the metabolism of their host), facultative mutualism (they have a beneficial effect on their host, but the host will survive without the symbiont) or commensalism (they are present without obvious beneficial effects to their host).

According to the literature, possible aims of sponge-microbe symbiosis are:

- Providing nutrition through intracellular digestion or translocation of metabolites²
- Access of new products via nitrogen fixation³
- Stabilization of the sponge skeleton⁴
- Prevention against predation or fouling through secondary metabolite production⁵

7.1 Cellular localization of metabolites from *P.simplex*

The bacterial content in *Plakortis simplex* is particularly high, around 90 bacteria per sponge cell, which are permanently associated with the host. The isolation of bacteriohopanoids⁶ that are typical bacterial metabolites, from the sponge led to the hypothesis that not only these compounds but also the other characteristic metabolites isolated from *P. simplex*, such as the antimalarial polyketide plakortin⁷ and several unusual glycolipids,^{8,9} may be produced by its symbionts.

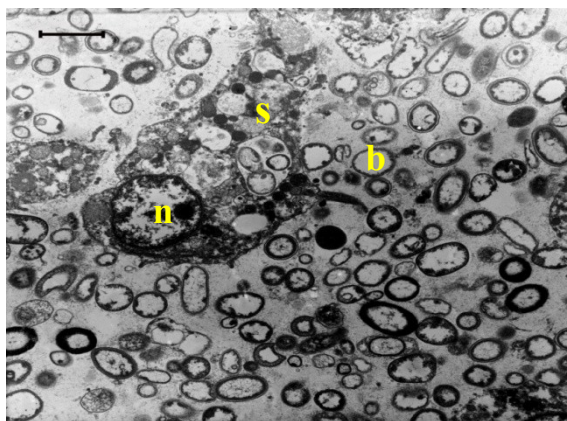


Figure 40: Transmission electron micrograph (TEM) of the *P. simplex* mesohyl. **s.** sponge cell, **n.** sponge cell nucleus, **b.** bacteria. Scale bar 2 μm

This hypothesis was first tested by localizing specific metabolites in cells using physical separation of sponge cells, bacteria and extracellular matrix by differential centrifugation.¹⁰ The obtained fractions were separately analyzed for the typical *Plakortis simplex* metabolites by NMR and mass spectrometry. The results suggested that bacterial symbionts are the producers of most of the secondary metabolites isolated from *Plakortis*.

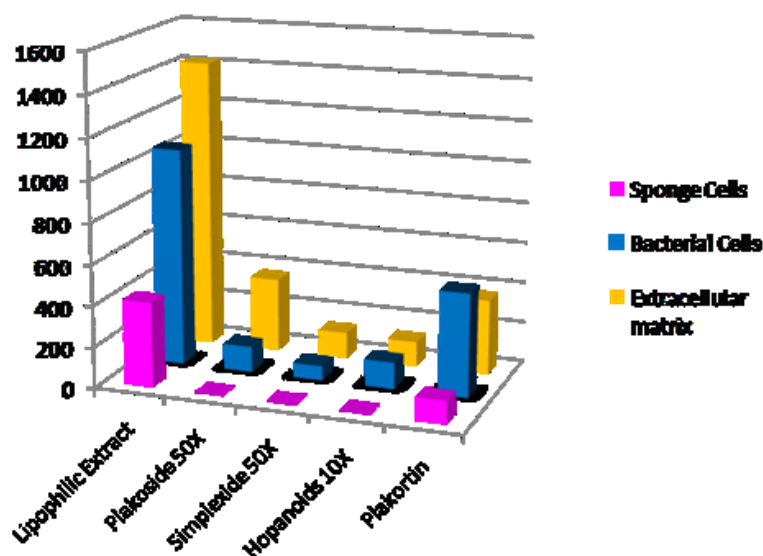


Figure 41: Results of differential centrifugation analysis

In fact, hopanoids and glycolipids were found only in bacterial cells, while the polyketide plakortin and its derivatives were present in both bacteria and sponge cells. However, the amounts were considerably different, as the bacteria fraction contained five times as much plakortin as the sponge cell fraction. All the metabolites were also present in the extracellular matrix, either because the metabolites are excreted, or because part of the cells producing them are lysed during the cell separation process.

The bacterial origin of these metabolites may be the key to overcome the problem of the limited availability of sponge material, which rises every time the commercial production of a marine compound of pharmaceutical interest from a sponge is hypothesized. If the metabolite is produced by a bacterium, then large-scale laboratory culture is possible, and the need to harvest sponges from their natural environment is eliminated.

In the attempt to identify the microorganisms producing the metabolites found in *P. simplex*, over 150 strains were isolate from the pool of its

symbionts; 30 representative species were cultivated and subjected to chemical analysis in search of glycolipids and plakortin. Unfortunately, we could not find the metabolites in any strains. These results were not surprising, because it has been estimated that less than 1% of marine microbes are cultivable in the laboratory today, and this is even more true for symbiotic bacteria. In addition, it is not uncommon that biosynthetic gene clusters remain silent unless some particular culture conditions are met. Therefore, a culture independent approach was explored, the search for genes responsible for the production of the antimalarial polyketide plakortin. The final aim of this research is to transfer the plakortin genes into a bacterial host, which would become capable of biosynthesizing plakortin, and would allow its production on large scale through a fermentative process.

References

1. Hentschel U, Hopke J, Horn M, Friedrich AB, Wagner M, Moore BS (2002) Molecular evidence for a uniform microbial community in sponges from different oceans. *Appl. Environ. Microbiol.* 68:4431–4440
2. Wilkinson CR, Garrone R (1980) “Nutrition in marine sponges. Involvement of symbiotic bacteria in the uptake of dissolved carbon. In: Smith D, et al. (eds) *Nutrition in lower Metazoa*. Pergamon, Oxford, pp 157-161.
3. Wilkinson CR, Fay P (1979) “Nitrogen fixation in coral reef sponges with symbiotic bacteria” *Nature* 279:527-529.
4. Wilkinson CR, Nowak M, Austin B, Cilwell RR, (1981) “Specificity of bacterial symbionts in Mediterranean and Great Barrier Reef Sponges”. *Microbiol. Ecol.* 7:13-21
5. Unson MD, Holland ND, Faulkner DJ (1994) “A brominated secondary metabolite synthesized by the cyanobacterial symbiont of a marine sponge and accumulation of the crystalline metabolite in the sponge tissue”. *Mar. Biol.* 119:1-11
6. Costantino V, Fattorusso E, Imperatore C, Mangoni A (2001) A biosynthetically significant new bacteriohopanoid present in large amounts in the Caribbean sponge *Plakortis simplex*. *Tetrahedron* 57:4045–4048
7. Higgs MD, Faulkner DJ (1978) Plakortin, an antibiotic from *Plakortis halichondroides*. *J Org Chem* 43:3454–3457

8. Costantino V, Fattorusso E, Mangoni A, Di Rosa M, Ianaro A (1999) Simplexides, novel immunosuppressive glycolipids from the Caribbean sponge *Plakortis simplex*. *Bioorg Med Chem Lett* 9:271–276
9. Costantino V, Fattorusso E, Mangoni A, Di Rosa M, Ianaro A (1997) A unique prenylated glycosphingolipids with immunosuppressive activity from the marine sponge *Plakortis simplex*. *J Am Chem Soc* 119:12465–12470
10. Laroche M, Imperatore C, Grozdanov L, Costantino V, Mangoni A, Fattorusso E, “Cellular Localization of Secondary Metabolites Isolated from the Caribbean Sponge *Plakortis simplex*”. *Mar Biol* (2007) 151:1365–1373 DOI 10.1007/s00227-006-0572-1

Chapter 8

Plakortin

Malaria is one of the major causes of mortality in the tropical regions. Each year, there are approximately 515 million cases of malaria, killing between one and three million people, the majority of whom are young children in Sub-Saharan Africa. Although some are under development, no vaccine is currently available for malaria; preventive drugs must be taken continuously to reduce the risk of infection. These prophylactic drug treatments are often too expensive for most people living in endemic areas. Chloroquine is very cheap and, until recently, was very effective, which made it the antimalarial drug of choice for many years in most parts of the world. However, resistance of *Plasmodium falciparum* to chloroquine has spread recently from Asia to Africa, making the drug ineffective against the most dangerous *Plasmodium* strain in many affected regions of the world. In those areas where chloroquine is still effective it remains the first choice. Unfortunately, chloroquine-resistance is associated with reduced sensitivity to other drugs, such as quinine and amodiaquine.

The more recently discovered artemisinin, a sesquiterpene endoperoxide from *Artemisia annua*, is active against *Plasmodium falciparum* and chloroquine resistant strains. Artemisinin and its semisynthetic liposoluble (artemether and artether) and water soluble (artesunate) derivatives are characterized by a 1,2,4-trioxane (6-membered ring containing 3 oxygen atoms) that is the essential pharmacophoric group for their activity. Even

though its mechanism of action hasn't been exactly determined yet, there is experimental evidence that the artemisinin endoperoxide function reacts with the heme iron atom. The reaction yields an oxygen radical, and after a rearrangement a carbon radical. This alkylating species could be responsible for the antimalarial activity.

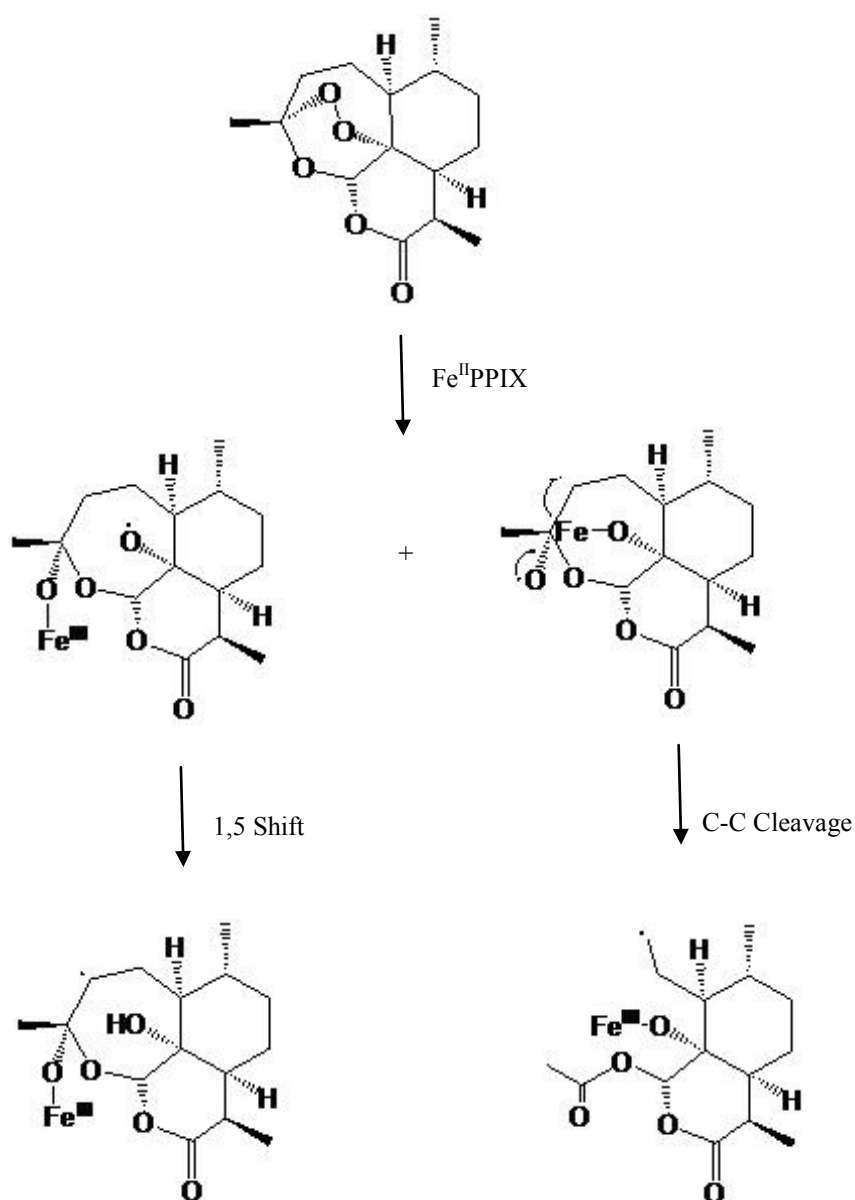


Figure 42: Mechanism of action of artemisinin

The problem with these compounds is their neurotoxic activity which is related to the endoperoxide ring; for this reason new antimalarial compounds are required that could be easier synthesized, preserved and safer. In this respect, several secondary metabolites from sponges show an endoperoxide ring similar to that of artemisinin. Therefore, these compounds are being tested in order to evaluate their possible antimalarial activity. The first active compound, isolated in 1978 by Faulkner et al. in the lipophilic extract of the sponge *Plakortis halicondroides*, is plakortin, 6-membered cyclic endoperoxide.

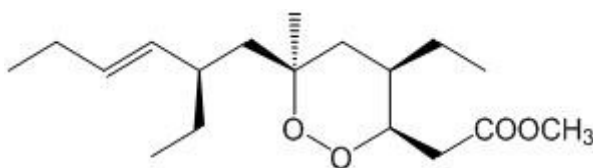


Figure 43: Plakortin

Most polyketides that are appearing in the literature have some common key features: they all contain a 1,2-dioxane ring with an acetate residue at position 3 and alkyl chains at positions 4, 6, and 6. Related molecules with a 5-membered peroxide ring (1,2-dioxolane), such as Plakortide E, were also discovered. The analysis of the lipophilic extracts of the Caribbean sponge *Plakortis simplex* proved that plakortin is the major constituent of this fraction (20%) together with 9,10-dihydroplakortin and the 5-membered Plakortide E.

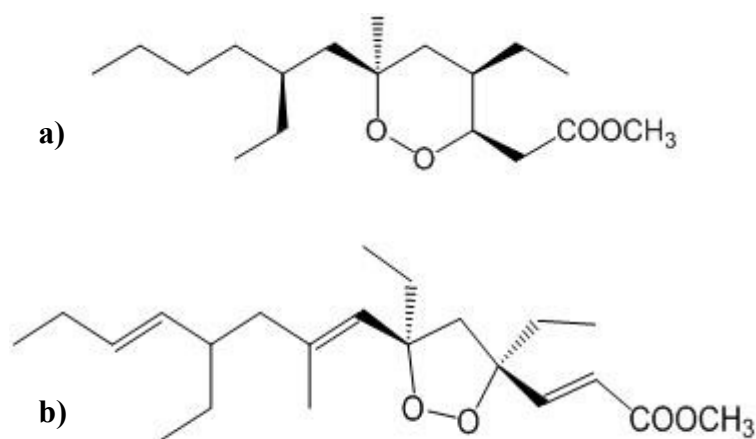


Figure 44: Dihydroplakortin (a) and Plakortide E (b)

These three metabolites were tested for their antimalarial activity on D10 strains and on W2 (chloroquine-resistant). Bacterial growth was evaluated spectrophotometrically (OD_{650}), measuring the pLDH activity according to the method by Makler and Hinrichs. Both plakortin and dihydroplakortin have a good antimalarial activity, while plakortide E is inactive, proving that a six-membered peroxide ring is essential for activity. Plakortin is very active on the chloroquine-resistant W2 strain because endoperoxide compounds are active against *Plasmodium falciparum* with a different mechanism compared to chloroquine. The biological target of plakortin is similar to that of artemisinin, but the former shows a lower activity, probably related to its 1,2-dioxane ring instead of the 1,2,4-trioxane of artemisinin. In fact, it seems that the oxygen not involved in the peroxide moiety can make the radical production easier, increasing the antimalarial activity. In spite of its lower activity, its simple structure and insignificant cytotoxicity make plakortin a very interesting compound.

8.1 The biogenesis of plakortin

The studies about cellular localisation showed plakortin to be present mostly in the bacterial cells. Because any attempt to isolate the strain of bacteria responsible for the biosynthesis of plakortin, a culture independent approach is being currently explored, searching in the metagenomic (sponge + symbiont) DNA from *P. simplex* the gene responsible for the biosynthesis of the plakortin.

The chance of success of this approach comes from the bacterial origin of plakortin in that:

- In bacteria, the genes for the biosynthesis of metabolites are often grouped in clusters
- There are no introns, and this allows to express the genes directly from chromosome
- The compactness of the genome makes it faster to identify the gene clusters and to transfer them to hosts for expression
- Large-scale production of metabolites is possible through fermentation

Moreover, plakortin is a polyketide, and polyketides are biosynthesized by polyketide synthases (PKS), enzyme complexes which are widespread in bacteria and are very well studied.

8.2 Polyketide Synthases

PKSs are large (100-10000 KDa) multienzymatic systems responsible for the production of extremely complex natural products from simple blocks with 2, 3 or 4 carbon atoms such as acetyl-CoA, propionyl-CoA, butyryl-CoA and their activated derivatives malonyl-, methylmalonyl- and ethylmalonyl-CoA. They show strong genetic, structural, and mechanistic similarity with fatty acid synthases. The key step in the biosynthesis of polyketides is chain elongation, obtained through Claisen condensation and decarboxylation. Unlike fatty acid biosynthesis, in which every chain elongation step is followed by a fixed sequence of ketoreduction, dehydration, and enoylreduction, the reaction intermediate obtained during the polyketide biosynthesis are subjected to none, some or all these modifications, yielding very complex products. An added degree of complexity comes from the possible use of different starter and elongation groups as well as from the formation of new stereogenic centres.

In the microbial kingdom, three types of PKSs were identified that differ in the architecture:

- Type I PKSs: large multifunctional proteins, which can be either modular (erythromycin, rapamycin and rifamycin) or iterative (lovastatin). Iterative PKSs are involved in the biosynthesis of polyketides such as 6-methyl salicylic acid and aflatoxin in fungi; these multidomain proteins include all the active sites for the biosynthesis of the polyketide and they reuse domains in a cyclic fashion like the fatty acid synthases.

- Type II PKSs: they show active sites similar to those of type I PKSs, but they are distributed on different small monofunctional peptides. They are involved in the biosynthesis of bacterial aromatic compounds such as doxorubicin
- Type III PKSs: iterative PKSs responsible for the production of chalcones and stilbenes in plants and polyhydroxy phenols in bacteria. Chalcone synthases consisting of a single polypeptide chain do not show ACP domain, and their substrates are directly CoA derivatives.

Modular PKSs are very interesting enzymes; they contain a sequence of separate modules; a module consists of not repeated catalytic domains with defined function separated by short spacer regions. Each module is responsible of one step of polyketide elongation and of modification of functional groups.

The most common catalytic domains found in PKSs are:

- Acyltransferase (AT): this enzyme transfers the elongation group (malonyl or a malonyl derivative) from CoA to the ACP module. This module is responsible for the selection of the proper elongation group.
- Acyl Carrier Protein (ACP): this domain binds the elongation group and, after the Claisen condensation, the growing chain through the distal –SH group of the 4'-phosphopantetheine arm. This protein is biosynthesized as an inactive form and after translation binds to 4'-phosphopantetheine through a serine residue

- Ketosynthase (KS): catalyzes the displacement of the growing polyketide chain from the ACP domain of the previous module to an -SH in the KS domain itself, and then the decarboxylative Claisen condensation with the elongation group linked to the ACP domain of the current module, leading to a β -ketoacyl.
- Ketoreductase (KR): this NADPH-dependent domain catalyzes the reduction reaction of the β -ketoacyl to a β -hydroxy acyl.
- Dehydratase (DH): this enzymatic subunit catalyzes the removal of a water molecule from the β -hydroxy acyl with the formation of a double bond.
- Enoylreductase (ER): catalyzes the reduction of the double bond produced by the DH domain
- Thioesterase (TH): this domain is present only in the last module, and catalyzes the release of the polyketide from the last ACP domain.

All the modules are at least composed of a KS, an AT, and an ACP domain. Moreover, specific combinations of KR, DH, ER (and also epimerase and methyltransferase) domains can be present in each module. Due to the non-iterative nature of these enzymes, it is possible to predict the structure of the polyketide produced by a PKS on the basis of the number of modules of the PKS and the domains present in each module. Conversely (and more interestingly from our point of view) it is possible to predict the PKS organization on the basis of the structure of the polyketide.

8.3 An Hypothesis on the Biogenetic Pathway to Plakortin

As discussed in previous section, plakortin is presumably biosynthesized by a Type I PKS.

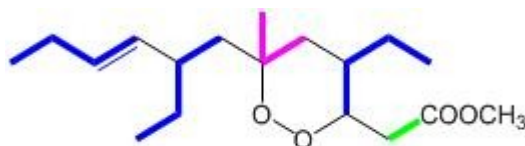


Figure 45: Biogenetic dissection of plakortin

Looking at the structure of plakortin, it is quite apparent that carbon skeleton should be produced by condensation of three butyrate, one propionate, and one acetate units. Alternatively, the C₄ unit at the end of the molecule could arise from two acetate units instead of one butyrate. The reactions that lead to the formation of plakortin can be grouped into four cycles (corresponding to four modules of a PKS), each composed of several steps catalyzed by different catalytic domains. As in all modular and non-iterative PKS, the synthesis proceeds through a cascade mechanism and the reaction product of each reaction is the substrate for the next reaction.

Based on the above discussion, the reaction sequence for the biosynthesis of plakortin can be hypothesized as follow:

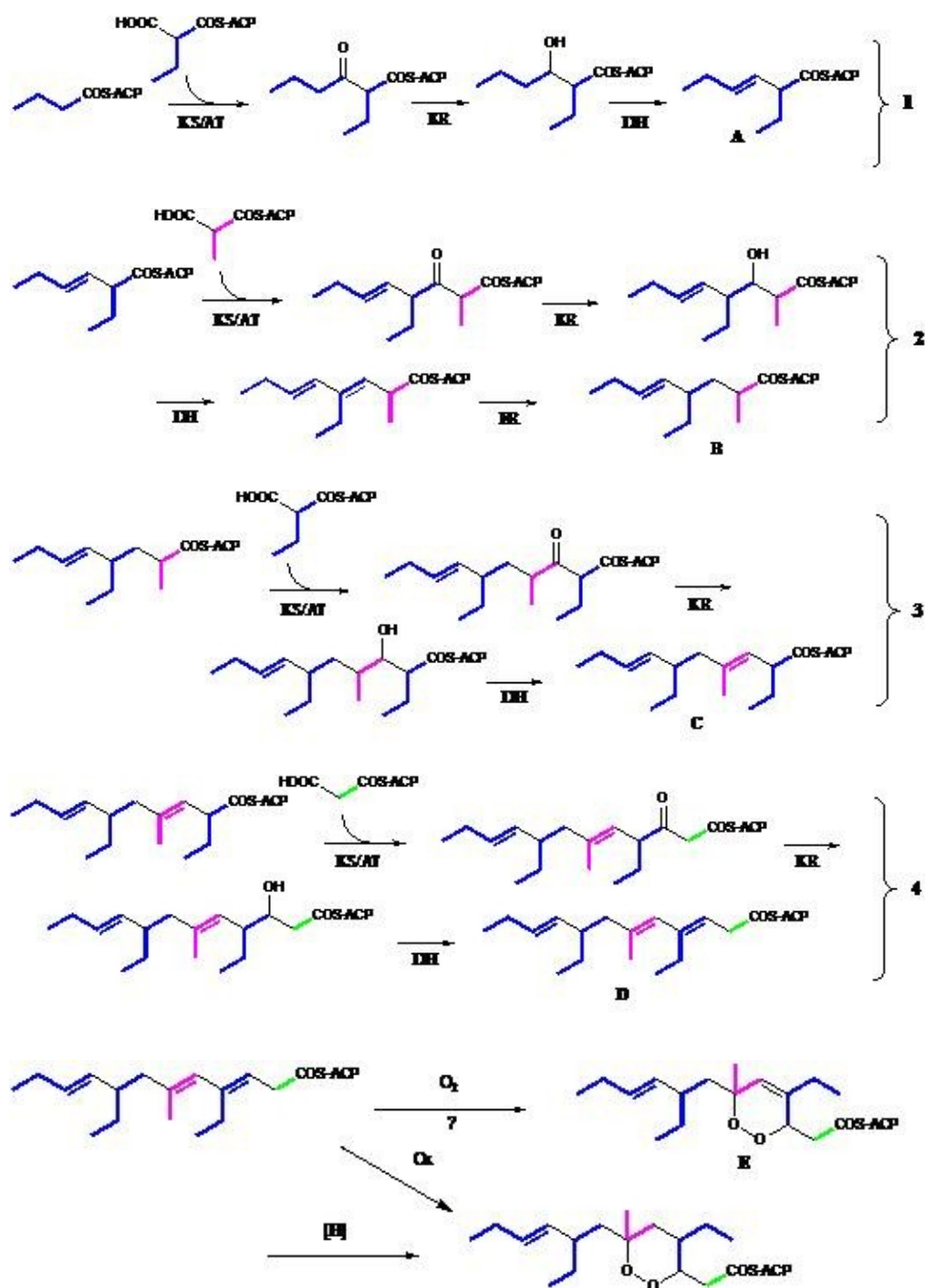


Figure 46: The cascade mechanism hypothesized for the biosynthesis of plakortin

Loading

The loading stage consists of three steps. First, the loading AT domain transfers the starter group from ethylmalonyl-CoA to the –SH group of the ACP domain, followed by decarboxylation to give butyryl-ACP.

Cycle 1

The elongation group is loaded by the current AT domain onto the ACP domain, yielding ethylmalonyl-ACP. The ACP-bound elongation group reacts with the butyryl residue bound the ACP of the starting module in a Claisen condensation and decarboxylation, catalyzed by the KS domain. The obtained β -ketoacyl group is reduced to β -hydroxy by the KR domain. Then, the DH domain splits off a water molecule, resulting in the α,β -unsaturated thioester (intermediate **A**) which is the substrate for the subsequent enzymatic reactions.

Cycle 2

The second cycle is similar to the first one, but in this case the AT domain loads a methylmalonyl-CoA molecule instead of ethylmalonyl-CoA onto the ACP module. The decarboxylative Claisen condensation (KS) is followed by reduction (KR), dehydration (DH) reactions and by the α - β double bond reduction, catalyzed by ER domain (intermediate **B**).

Cycle 3

The KS domain catalyzes the condensation between compound 2 and ethylmalonyl residue on the ACP domain. The β -ketoacyl group of the

growing chain is reduced and the subsequent β -hydroxyacyl dehydration yields the intermediate **C**.

Cycle 4

The AT domain loads a malonyl-CoA onto the ACP domain. The Claisen condensation and decarboxylation (KS) between malonyl-ACP and the intermediate **C** followed by reduction (KR) and dehydration (DH) produces the intermediate **D**.

Oxidative Cyclisation

As in literature there are not comparable examples, it is difficult to hypothesize an accurate mechanism of oxidative cyclisation. The reaction could take place in the presence of singlet oxygen through a Diels-Alder type mechanism. This reaction yields the intermediate **E** and, after the reduction of the double-bond and hydrolysis of the thioester, plakortin.

Therefore, the organization of PKS for plakortin is probably as sketched below. The PKS could be composed in one or more polypeptide chains, and therefore could be coded by one or more clustered genes.

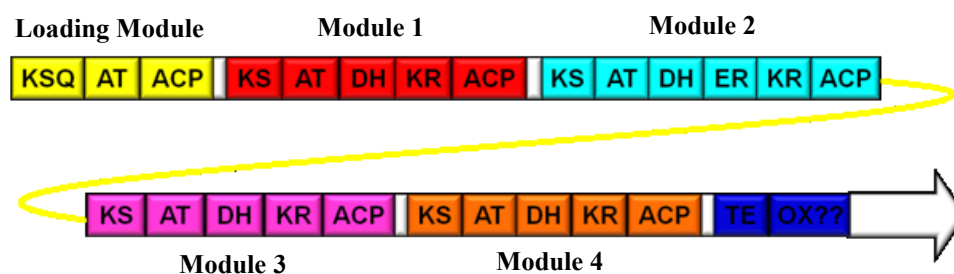


Figure 47: Hypothesized organization of the PKS for plakortin

8.4 Strategy for the Identification of Plakortin Biosynthetic Gene Cluster

In order to isolate plakortin gene cluster, a library of metagenomic DNA (DNA for sponge and DNA from symbiotic microorganisms) from *Plakortis* was constructed. DNA was isolated from the sponge, sheared and inserted in fosmid vectors, which were used to transform competent *E. coli* cells (i.e. cells which are capable to and replicate the fosmids). Dealing with a metagenome, estimation of the number of the genomes of their size is impossible, and therefore the library should be as large as possible.

Generally speaking, the selection of recombinant clones may be based on:

- DNA sequence: screening with hybridization or PCR
- Protein sequence: immunological screening
- Biochemical function: screening based on the polypeptide function
- Peptide ability to interact with other polypeptide: screening for interaction

In our case there was no guarantee about plakortin gene expression in the host, and therefore PCR-based screening of nucleotide sequence was used. PCR (Polymerase Chain Reaction) allows amplification of a specific piece of DNA exponentially, and is effective even if only a trace amount of the target DNA fragment is present in the metagenome. The couple of primers can be directly tested on the whole metagenome before the screening of the library. However, the design of appropriate primers requires prior knowledge about the sequence to amplify.

Plakortin fits this strategy very well. PKS are widespread in bacteria, and the sequences of many PKS genes are available in literature, showing conserved regions common to all modules and presumably also to plakortin PKS. Another reason why plakortin was chosen as a target is that it is mainly composed of butyrate and propionate, rather than acetate, units, and therefore is presumably biosynthesized by a quite peculiar PKS. Because the selection of methylmalonyl-CoA or ethylmalonyl-CoA depends on the AT domain, the sequences of many AT domains with different selectivity were compared. These domains show:

- Conserved regions common to all domains
- *Signature regions*: regions which are different in AT domains specific for different substrates (malonyl-CoA, methylmalonyl-CoA, ethylmalonyl-CoA)

Conserved regions were used to design a couple of "generic" primers suitable for all AT domains, regardless of their selectivity. As butyrate-specific AT domains are uncommon, specific primers were also designed, on the basis of the *signature regions* of propionate and butyrate-specific AT domains, so as to increase the chance to isolate plakortin PKS gene cluster among all the PKS genes in the metagenome.

After PCR screening, the selected fosmid need to be sequenced and their sequences analyzed to confirm their compatibility with the structure of the polyketide. PKS for plakortin should contain 4 or 5 modules and so its size should be comparable to the DNA insert contained in each fosmid (around 40Kb). This could make it possible, although not likely, to find the whole gene cluster inside a single fosmid.

The final purpose of this study is to express the plakortin gene cluster in a heterologous host. Heterologous expression of polyketides is currently a pioneering research area. Although an increasing number of examples are being reported in the literature, a number of problems remain to be resolved, such as the choice of the most suitable host, and the frequent need to provide the host with the additional biosynthetic machinery which is needed for post-translational modification of the PKS and for supply of biosynthetic precursors. However, a large amount of research is being done in this fast-developing field, and heterologous expression is expected to become a commercial reality in a few years.

Chapter 9

Construction of a Metagenomic DNA Library

9.1 DNA Extraction

A small piece of frozen *Plakortis* in RNA Later (200 mg) was minced and suspended in a buffer containing CTAB and SDS as detergents, which degrade membranes and release the cellular content in solution, and PVP, NaCl and β -mercaptoethanol, with the function of stabilizing the DNA molecule and minimizing DNA breakage. The resulted suspension was incubated for 3 h at 55 °C to allow cell lysis, and then centrifuged to precipitate spicules and other insoluble material. A volume of chloroform was added to the supernatant, containing DNA, to extract most proteins. After centrifugation and removal of the chloroform phase, 100% isopropanol (2/3 volumes) was added to the aqueous phase to precipitate DNA. After centrifugation, the supernatant was discarded and DNA was washed twice with 70% EtOH, air-dried for a few minutes, and re-dissolved in a diluted TE buffer.

The amount of DNA was evaluated spectrophotometrically by measuring the absorbance of the sample at 260 nm, the absorption maximum of nucleic acids.

9.2 Construction of the Gene Library

After extraction, metagenomic DNA was used to construct a fosmid library using the CopyControl™ Fosmid Library Production Kit (Epicentre). A gene library is a population of bacterial colonies, each carrying a different DNA fragment that is inserted in a cloning vector. There are many types of cloning vectors which can be used to build genomic library. Plasmids are circular and double-stranded extra-chromosomal DNA molecules that are capable of replicating independently of the chromosomal DNA. Cosmids are hybrid plasmids. They contain the *cos* sequences from the Lambda phage, which allow the recognition by the phage capsid proteins. In this way cosmids, packaged into capsids, can be easily transferred into bacterial cells where they can replicate as plasmids, so that many copies of cosmids can be present in each bacterial cell. Cosmids are able to contain 37 to 52 kbp of DNA while normal plasmids carry only 1-20 kbp. Fosmids are more sophisticated than cosmids; they are derived from the bacterial plasmid F, and contain a partitioning and replication system which allows the bacterial host (*E. coli*) to contain a single copy of the fosmid, offering a higher stability to the library.

The CopyControl pCC1FOS vector, which was used to build metagenomic DNA fosmid library from *Plakortis simplex*, has distinctive features, such as chloramphenicol-resistance for selection of the transformed bacterial cells, and the inducible high-copy *oriV* origin of replication in addition to the single-copy origin. Initiation of replication from *oriV*

requires an inducer, and can be activated during the fosmid isolation to increase its amounts.

The following steps were involved in *P. simplex* DNA library construction:

- DNA extraction and purification
- Random shearing of the DNA to approximately 40 Kb fragments
- End-repair reaction of the sheared DNA to blunt, 5'-phosphorylated ends
- Size selection of end-repaired DNA by LMP (low melting point) agarose gel electrophoresis
- Ligation of blunt end DNA to the CopyControl pCC1FOS vector
- Packaging of ligated DNA and transformation of EPI 300-T1^R competent cells
- Plating of transformed cells and antibiotic selection of positive clones
- Picking of selected fosmid clones and screening

End-Repair

During this step DNA fragments ends, which could have been damaged during DNA extraction and shearing, were repaired and 5'-phosphorylated. A reaction mixture containing *End-repair 10X Buffer*, dNTPs, ATP, and *End-repair Enzyme Mix* was added to the appropriate volume of DNA solution in TE Buffer. *End-repair Enzyme Mix* contains T4 polynucleotide kinase which catalyses γ -phosphate transfer from ATP to free –OH groups

at DNA 5'-position and T4 polymerase with both 5' → 3' polymerase and 3'→5' exonuclease activities.

Reaction was incubated at room temperature and then at 70° C for 10 minutes to inactivate *End-repair Enzyme mix*.

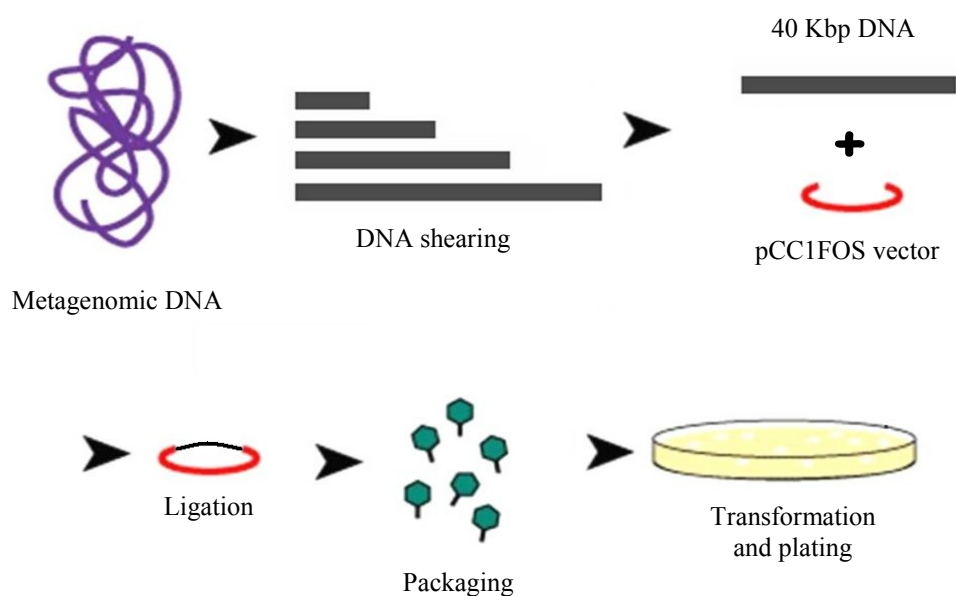


Figure 48: Metagenomic DNA Library Construction

Size Selection

Size selection of DNA fragments is required before ligation. The optimal fragment size is around 40 Kb according to the capacity of fosmid cloning vector. It is important that the DNA recovered is ≥ 25 Kb in order to avoid unwanted chimeric clones.

End-repaired DNA was size selected by low melting point (LMP) agarose gel electrophoresis. GELase, an enzyme responsible for hydrolyzing β -agarose leaving nucleic acids intact, was used to extract the DNA fragment (1.2 μ g) of correct size back from the gel matrix.

Ligation

During ligation size-fractionated DNA (0.25 µg) was inserted into the pCC1FOS vector. Ligation involves creating a phosphodiester bond between double-stranded DNA molecules having 3'-hydroxylated and 5'-phosphorylated blunt ends. *Fast-Link DNA Ligase* was used for this purpose, which originates from the T4 bacteriophage.

Packaging

After ligation fosmids were mixed with the capsid proteins (*MaxPlax Lambda Packaging Extract*) of bacteriophages, giving phage particle capable to transfer the fosmid into bacterial cells. Titering of this mixture showed packaged fosmid clones concentration of 50.000 CFU/mL.

Transformation

Transformation is the process by which competent bacterial cells take up foreign DNA molecules and allow its replication and expression. When DNA is transferred to one bacterium by a virus, transformation is called transduction. During transduction bacteriophages attach to a specific receptor on the bacterial cell's surface through tail fibers. Lysozyme dissolves a hole in the cell wall, the phage contracts and the nucleic acid is injected into the cell.

Serial dilution of phage particles (packaged fosmid clones) into *Phage Dilution Buffer* (PDB) were mixed with EPI 300-T1^R host cells and incubated at 37° C for 20min. Then the infected bacteria were spread on LB

plates + 12.5 µg/mL chloramphenicol and incubated overnight to select the fosmid clones.

The bacterial colonies grown on the plates were transferred in 96-well plates containing 300 µL of LB medium + chloramphenicol and subjected to PCR technique to evaluate the presence of the plakortin gene cluster.

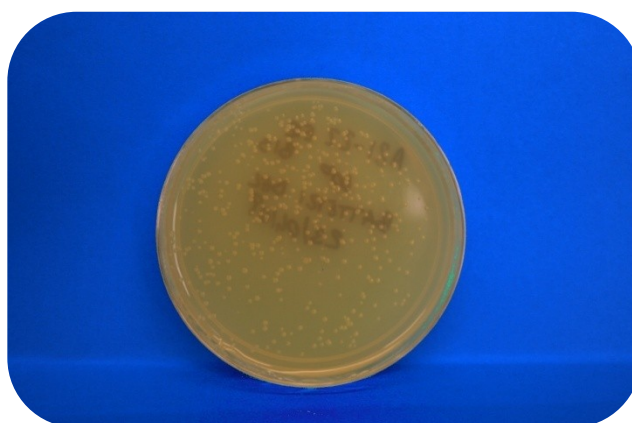


Figure 49: Metagenomic DNA Library from *Plakortis simplex*

9.3 PCR Screening

The Polymerase Chain Reaction, developed by Kary Mullins in 1984, is a widely used technique for a variety of application like DNA cloning and sequencing, the diagnosis of hereditary disease, and the identification of genetic fingerprint in forensic science. It allows to exponentially amplify a single or few copies of a specific sequence of DNA, called target sequence, from a mixture without any purification

DNA molecules are melted by raising temperature to 90-95° C (denaturation); after the helixes are separated, the temperature is lowered to 45-65° C, so that a pair of short oligonucleotide (primers), specific for the 3'-

ends of the target sequence can bind to the appropriate strands of DNA, one primer on each strand (annealing). The primers direct the DNA polymerase, called Taq, to synthesized complementary DNA strands from free nucleotide triphosphates (extension/elongation). In the entire mixture, only the DNA containing the target sequence is duplicated, because the Taq polymerase can only duplicates molecules having the primers attached. These three phases are repeated 20-40 times, and the number of target sequences doubles at each cycle.

One of the most important factors affecting the quality and the selectivity of PCR is the choice of primers: they need to match as closely as possible the beginning and the end of the DNA fragment to be amplified, so as to anneal to these starting and ending points exclusively. DNA polymerase binds to them and begins the synthesis of the new DNA strands.

A pair of generic primers, designed on the conserved amino acid sequences of Type I PKSs AT domains were used in the screening of a preliminary metagenomic DNA library from *Plakortis*, constructed as a test.

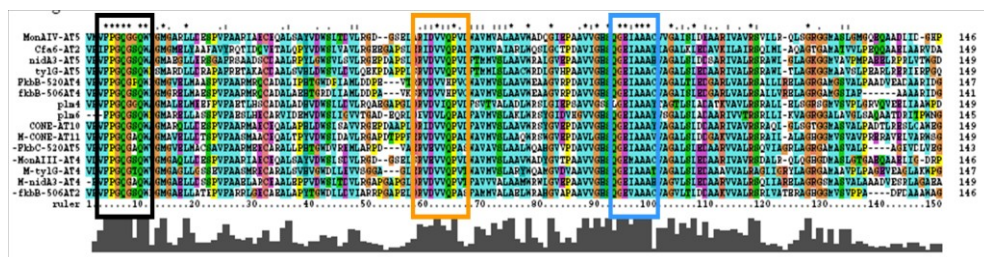


Figure 50: Signature region identified in PKs gene clusters from GenBank.

| Primer | Nucleotide Sequence | Conserved Amino Acid Sequence |
|------------------------|--------------------------------|-------------------------------|
| PKSAT1F (24bp) | 5'-TTYCCIGGICARGGINSSCAGTGG-3' | FPGQGSQW |
| PKSAT3R2 (18bp) | 5'- GCIGCIGCNATCTCNCCC-3' | GEIAAA |

Table 9: Primers for conserved regions of AT domains of type-I PKSs. Y = C+T, R = A+G, S = G+C, N = A+G+C+T, I = inosine.

A number of *degenerated nucleotides* (mixtures of different nucleotides) and inosines (inosine can indiscriminately pair with all of A, T, C, and G) were used in the synthesis of these primers, because the conserved amino acid sequences cannot be univocally translated into nucleotide sequences due to codon degeneracy (each amino acid is coded by more than one codon).

Only a single positive clone was obtained called 11G3, that was partially sequenced (~4 Kb) by primer walking. It appeared to contain a Type I PKS, and the AT domain resembled more closely to propionyl/butyryl-specific ATs than the acetate-specific ones.

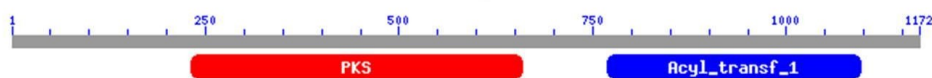


Figure 51: Primer walking result on 11G3 positive fosmid

This partial sequence was used to design two new couples of primers specific for this particular PKS (one couple was designed on the AT domain and one couple on the KS domain), which were used for the screening of a new, larger library in order to obtain more positive fosmids.

| Primer | Nucleotide Sequence | Aminoacid Sequence |
|---------------------|----------------------------------------|--------------------|
| ATPEF (26bp) | 5'-ATGGTGTTCGGGGCAGGGCACGCA-3' | MVFSGQGTQ |
| ATPER (30bp) | 5'-GGCGGCGGCCACCTCGCCCGAGCTGTGTCC-3' | GHSSGEVAAA |
| KSPEF (32bp) | 5'-AACGAGATGGCGAACATGGATCCGCAGATCCG-3' | NEMANMDPQIR |
| KSPER (29bp) | 5'-CGGTCGCCCACGACGGTGCCGGTCGCATG-3' | HATGTVVGDR |

Table 10: Primers pairs designed on positive fosmid sequence

So far, over 8.000 of the 50.000 colonies of the library were screened using these primers to identify a PKS gene.

A positive clone has been obtained, called α 11D7. It was full sequenced by shotgun sequencing, but unfortunately the PKS gene cluster was located at one end of the insert, so that only a small part of the gene cluster (about 9 Kbp) was present in the insert.

| bp | Conserved domains | AA |
|-------------|----------------------------------------|------|
| 44-1114 | Glycosyl transferase | 356 |
| 1126-2310 | Heparinase (?) | 394 |
| 5482-6723 | Integrase | 413 |
| 6723-7628 | Integrase | 301 |
| 7625-8617 | Integrase | 330 |
| 8941-9987 | WcaG | 348 |
| 10471-11847 | Sugar Transferase | 458 |
| 11858-20813 | PKS module (incomplete ORF, see below) | 2985 |
| 20814-21658 | Vector PCC1FOS | |

Table 11: Open reading frames identified in contig0003 and their putative function as identified by BLAST Search CD.

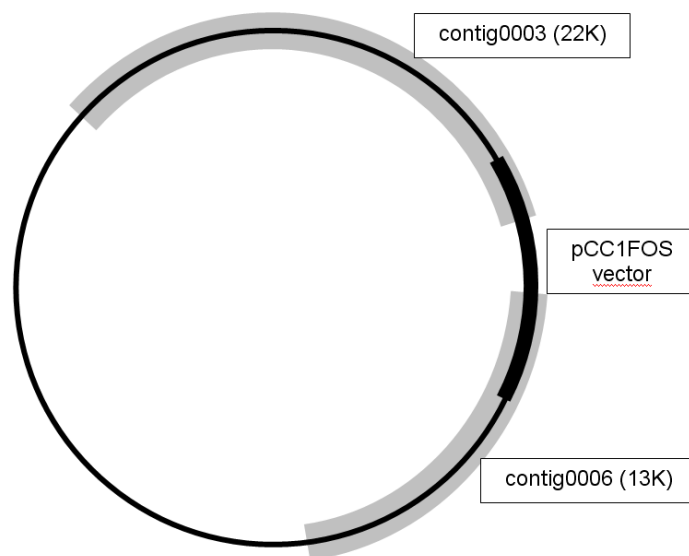


Figure 52: The two biggest contigs from shotgun sequencing of fosmid α 11D7. Contig0003 contains part of a PKS gene cluster.

The nucleotide sequence was translated in amino acid sequence and analyzed with BLAST in search of homologous proteins. BLAST *CD Search* module, used to search conserved domains, revealed the presence of the first 9 Kb of a PKS gene, and its AT domain is similar to AmbH gene from *Polyangium cellulosum*, responsible for the biosynthesis of antibiotic polyketide ambrucitin. We still do not know whether or not this module is related to plakortin. New primers were designed using Primer-BLAST for the end region of the known sequence and these are used for screening of more fosmid to search the remaining part of the gene cluster.

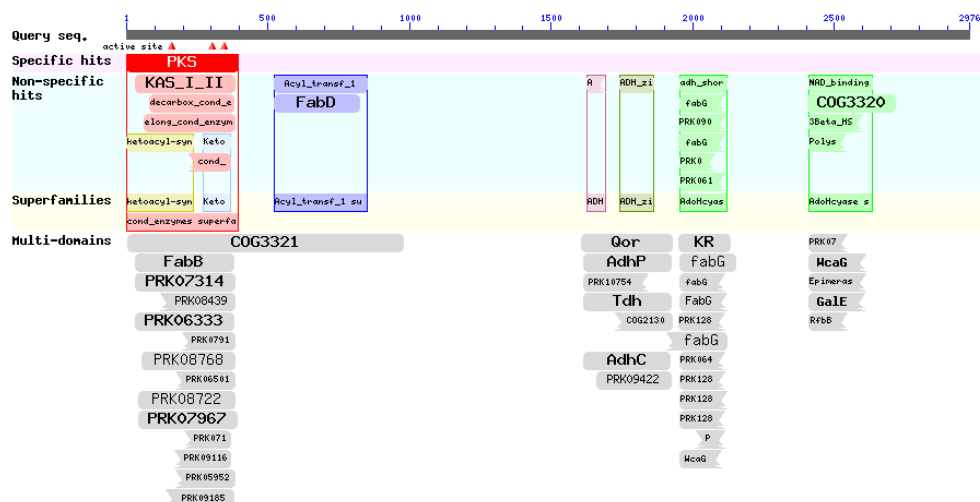


Figure 53: A PKS module in α 11D7 positive clone

9.4 Experimental Section

SB Buffer: 20X SB Buffer was prepared by dissolving 40 g of boric acid in 90 mL of distilled water, and leaving the solution under magnetic stirring for one day. Next day, solid NaOH (around 10g) was added to the solution, while monitoring the pH with a pH meter, up to pH 8.3. The SB Buffer was autoclaved at 120° C for 20 min, and then a 1:20 dilution was used for gel electrophoresis.

TE Buffer: TE buffer was prepared dissolving in 500 ml of distilled water the following compounds:

- 10 mM Tris-HCl (pH = 7.5)
- 1 mM EDTA

Cell Lysis Buffer: The following compounds were dissolved in 500 mL of distilled water:

- 2% CTAB
- 200mM tris HCl pH 8
- 50mM EDTA
- 1.4M NaCl
- 0.5% PVP

LB Medium + Chloramphenicol: In 500 mL of distilled water, 12.5 g of LB medium (Sigma) were dissolved. The solution was sterilized in autoclave at 120° C for 20min. After cooling to 55° C, 250 µL of a 25 µg/mL chloramphenicol solution in 96% ethanol were added.

LB Medium + Chloramphenicol - Agar Plates: 12.5 g of LB and 7.5 g of agar powder (Fluka) were dissolved in 500 mL of distilled water, and after autoclaving and cooling to 50 °C, the chloramphenicol solution described above was added. After mixing, the medium was poured into 100mm, sterile Petri dishes. The plates were allowed to cool down to room temperature and then stored at 4 °C.

Metagenomic DNA Extraction: 190 mg of *Plakortis simplex* tissue, stored in RNA Later at -80 °C, was minced in small piece and subjected to cellular lysis with the following reagents:

- cell lysis buffer (18 mL)
- 10% SDS (180 µL)

➤ β -mercaptoethanol (90 μ L)

The suspension was kept at 55 °C for 3 h, mixing by inversion every 30min. the tube was spun for 4min at 4500rpm to remove spicules and other insoluble material (white cotton-like precipitate). The supernatant (17.1 ml), containing the DNA, was gently transferred to a clean tube. One volume of chloroform was added and the solution was mixed by shaking, in order to separate DNA from proteins, soluble in chloroform. The tube was spun for 20min at 4500rpm at room temperature. The upper aqueous phase (15.3 mL) was again washed with chloroform. After centrifugation 100% isopropanol (9.5 mL, 2/3 vol.) was added to the aqueous phase (14.2 mL) and the solution was mixed by inversion several times to precipitate DNA. The tube was spun for 10 min at 4500 rpm. A thick pellet formed. The supernatant was removed and the pellet was washed twice with 70% ethanol. The tube was spun at 4500 rpm for 4 min. The ethanol was discarded and the pellet was air-dried for 30 min, than dissolved in 250 μ L of TE Buffer. The amount of obtained DNA was evaluated spectrophotometrically.

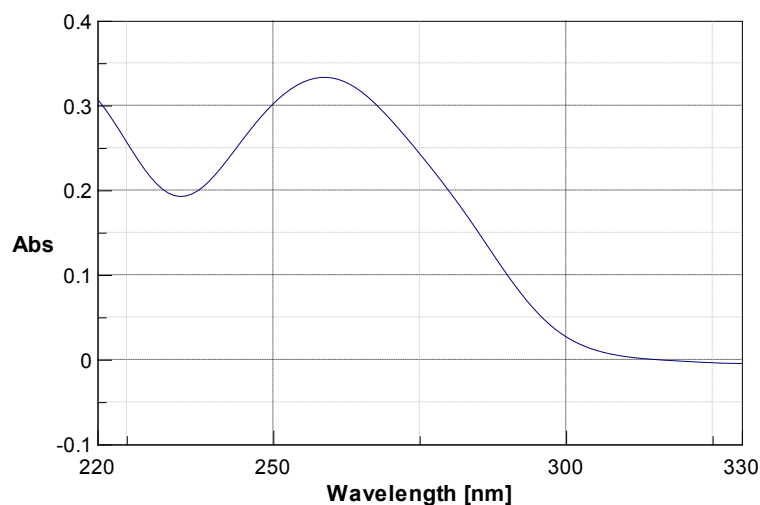


Figure 54: UV spectrum of DNA (1:50 dil., 400 μ L) extracted from *Plakortis simplex*. $A_{260} = 0.35$, $A_{280} = 0.22$, ratio = 1.6, 875 ng/ μ L, c = 1 cm

Shearing and End-Repair: Purified DNA was randomly sheared through the tips of a pipette (100 μ L) for around 100 times. Sheared DNA was subjected to *End-Repair* reaction according to the *CopyControl Fosmid Library Production Kit* (Epicentre) procedure. The following reagents were thaw on ice and mixed:

- 0 μ L sterile water
- 8 μ L 10X *End-Repair* Buffer
- 8 μ L 2.5 mM dNTP Mix
- 8 μ L 10 mM ATP
- 52 μ L sheared DNA (~0.5 mg/mL)
- 4 μ L *End-Repair* Enzyme Mix

(80 μ L total reaction volume). Mixture was incubated at room temperature for 45 min and then at 70 °C for 10 min to inactivate the *End-Repair Enzyme Mix*.

Size Selection of the End-Repaired DNA: 16 μ L (1/5 vol.) of 10X Blue Juice[™] Gel Loading Buffer (Invitrogen) were added to End-Repaired DNA solution (80 μ L). The mixture was loaded into a low melting point (LMP) agarose gel (20 cm long), as follows. This was prepared dissolving 0.36 g of LMP agarose (BioRad) in 90 mL of SB Buffer 1X.

- | | |
|-----------------------------|-----------------------------------|
| 1. HMW Marker | 10 μ L + 2 μ L Blue Juice |
| 2. DNA not sheared | 2 μ L + 1 μ L Blue Juice |
| 3. Fosmid Control DNA (1:5) | 2 μ L + 1 μ L Blue Juice |
| 4. End-Repaired DNA sol. | 90 μ L |
| 5. End-Repaired DNA sol. | 6 μ L |

The sample was resolved by electrophoresis at room temperature overnight at a constant voltage of 45 V.

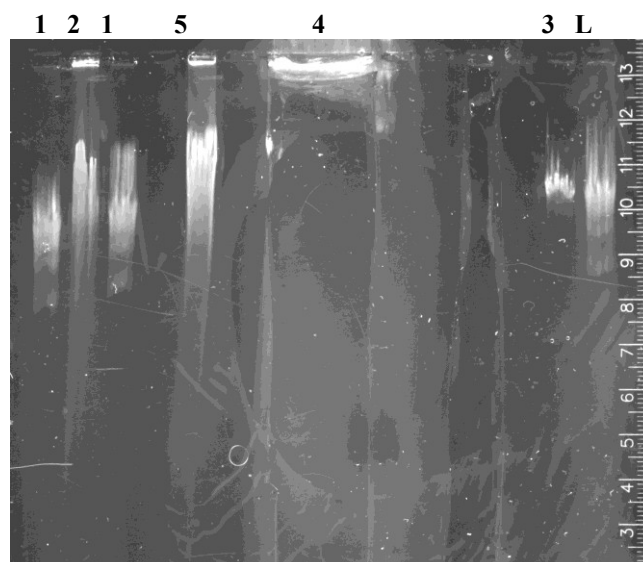


Figure 55: Agarose gel electrophoresis of end-repaired DNA.

The next day the lane 5 containing DNA was cut off and separated from the gel. The gel was stained with SYBR Safe DNA Gel Stain (Invitrogen) solution in 1% SB Buffer 1X and the position of the Fosmid Control DNA markers were visualized with UV light. After reassembling the gel a 0.5cm wide gel slice containing migrated DNA was excised from the lane 5 just slightly above the position of the 36 Kb Fosmid Control DNA markers. DNA was then extracted from the gel slice, as follows.

Recovery of the Size-Fractionated DNA from Agarose Gel: The gel slice (340.6 mg), containing the DNA of correct size was incubated at 70 °C for 15 min to melt the LMP agarose and then at 45 °C. 3.4 µL of GELase Enzyme Preparation (1 µL for every 100 µL melted agarose) were added and the solution was kept at 45 °C for one hour. After incubation the solution was transferred to 70 °C for 10 min to inactivate the GELase

enzyme. The solution was chilled in an ice bath for 5 min and then centrifuged at 13000 rpm for 20 min to pellet any insoluble oligosaccharides. A gelatinous pellet formed. The upper 326.8 μL of the supernatant, containing DNA, were removed into a clean microcentrifuge tube. The following reagents were added to precipitate the DNA:

- 32.7 μL 3 M Sodium Acetate (pH 7)
- 898.7 μL of 99.6% Ethanol

The mixture was inverted several times and DNA was allowed to precipitate for 30 min at $-20\text{ }^{\circ}\text{C}$. It was then centrifuged for 20 min at 13000 rpm. DNA pellet was washed twice with 75% EtOH, then dried under laminar flow for 10 min and redissolved in 25 μL of TE Buffer.

Ligation: The following reagents were thawed on ice, combined and mixed:

- 0 μL ultrapure sterile water
- 1 μL 10X Fast-Link *Ligation* Buffer
- 1 μL 10 mM ATP
- 1 μL CopyControl pCC1FOS Vector (0.5 $\mu\text{g}/\mu\text{L}$)
- 6 μL insert DNA (0.25 μg of $\sim 40\text{ Kb}$)
- 1 μL Fast-Link DNA Ligase

(10 μL total reaction volume). The mixture was incubated for 2 h at room temperature and then transfer to $70\text{ }^{\circ}\text{C}$ for 10 min to inactivate Fast-Link DNA Ligase.

Packaging: 50 mL of LB + 10 mM MgSO_4 were inoculated with a single EPI 300-T1^R Plating Strain cells and kept under shaking overnight at $37\text{ }^{\circ}\text{C}$.

The next day 5 mL of EPI 300-T1^R overnight culture were used to inoculate 45 mL of LB + 10 mM MgSO₄. Bacterial growth at 37 °C was monitored by UV spectroscopy to an OD₆₀₀ = 0.8. One tube of *MaxPlax Lambda Packaging Extract* was thawed on ice; 25 µL were transferred in a clean eppendorf tube and all the 10 µL of the *Ligation* reaction were added. The mixture was mixed by pipetting several times and incubated at 30 °C for 90 min. After the 90 min packaging reaction was completed the remaining 25 µL of *MaxPlax Lambda Packaging Extract*, previously thawed on ice, were added and the reaction was incubated for an additional 90 min at 30 °C. At the end of the second 90 min Phage Dilution Buffer was added to 1 mL final volume and after mixing 25 µL of chloroform. A viscous precipitate formed, which didn't interfere with library production.

Titering and Transformation: Phage particles in PDB were diluted as follows:

1. 1:10¹: 1 µL phage particles + 10 µL PDB
2. 1:10²: 10 µL phage particles + 990 µL PDB
3. 1:10³: 10 µL of 1:10² dilution + 90 µL PDB

10 µL of each dilution were added to 100 µL of the prepared EPI300-T1^R host cells and the suspensions were incubated for 20 min at 37 °C. The infected EPI300-T1^R were spread on LB plates + 12.5 µg/ml chloramphenicol. After overnight incubation at 37 °C the colonies on 1:10² dilution plate were 50 therefore the titer of phage particles was about 50.000 CFU. Colonies were picked from the plates using sterile toothpicks, and used to inoculate 96-wells plates, each well containing 250 µL of LB + 12.5

$\mu\text{g/mL}$ chloramphenicol in duplicate. The colonies in the same row of every plate were pooled in the same well in one new 96-well plate, as shown below:

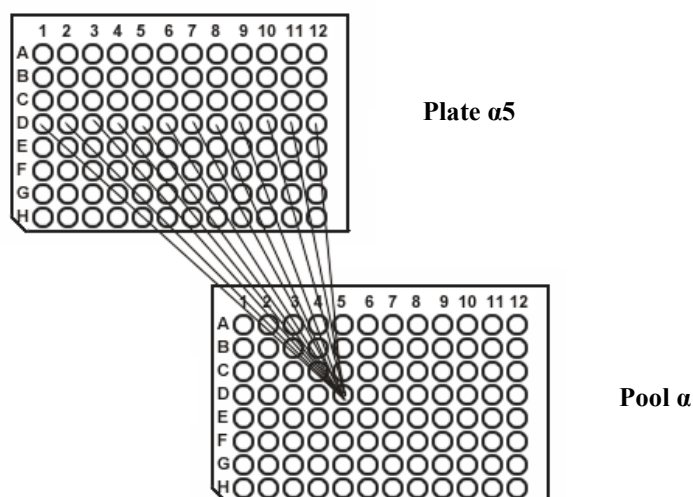


Figure 56: Pooling procedure

One copy of the pools was stored at $-80\text{ }^{\circ}\text{C}$, after adding $50\text{ }\mu\text{L}$ of glycerol to each well ($200\text{ }\mu\text{L}$).

High Copy Number Amplification of Fosmid Clones: Before screening the fosmids contained in the pool plates were induced to high copy number. 20 mL of LB medium + $12.5\text{ }\mu\text{g/mL}$ chloramphenicol and $20\text{ }\mu\text{L}$ of *1000X CopyControl Induction Solution* were mixed and used to load a 96-well plate ($200\text{ }\mu\text{L}$ in each well). $50\text{ }\mu\text{L}$ from the cultures of one pools were added in each well, the plates were shaken for 5 h at $37\text{ }^{\circ}\text{C}$.

PCR Screening: All the colonies from the pools were screened with PCR using specific primers for AT and KS modules shown above. The following reagents were mixed in a eppendorf tube for PCR ($0.2\text{ }\mu\text{L}$):

-
1. 8 μL *Master Mix* (Eppendorf)
 2. 0.8 μL Mg^{2+}
 3. 2.5 μL forward primer
 4. 2.5 μL reverse primer
 5. 6.2 μL sterile H_2O

The thermocycler program was:

- | | | | |
|---------------------------|----------|---|-----|
| 1. 95 °C (cellular lysis) | 30 min | | |
| 2. 95 °C (denaturation) | 30 s | } | 32X |
| 3. 52 °C (annealing) | 30 s | | |
| 4. 72 °C (elongation) | 1 min | | |
| 5. 72 °C (polymerization) | 5 min | | |
| 6. 4 °C | ∞ | | |

5 μL of every PCR product were mixed with 1 μL of DNA intercalating agent Blue JuiceTM Gel Loading Buffer and loaded in the agarose gel (1%).

PCR products were allowed to migrate for 30 min at 150 V. The only positive fosmid has been obtained from 7D well of $\alpha 11$ plate; it was extracted using *QIAPrep Spin MiniPrep Kit* (Qiagen).

Plasmid DNA Purification: The positive colony was picked from the selective plate, inoculated in a culture of 5 mL LB medium + 12.5 $\mu\text{g}/\text{mL}$ chloramphenicol and incubated overnight at 37 °C with vigorous shaking.

The next day in two clean falcons the following reagents were mixed:

1. 4.5 mL LB + 12.5 $\mu\text{g}/\text{mL}$ chloramphenicol
2. 500 μL fresh overnight culture of $\alpha 11\text{D}7$
3. 5 μL 1000X CopyControl Induction Solution

Positive fosmids were amplified to high copy number at 37 °C for 5 h and the transformed bacterial cultures were then centrifuged at 10.000 rpm for 10min. The supernatant was discarded and the pelleted bacterial cells were resuspended in 250 µL of P1 Buffer + RNase A. Buffer P2 was added for cellular lysis and the suspension was mixed by inverting the tube until the color of the solution turned to blue (for the presence of Lyse Blue reagent in the Buffer P2). 350 µL of Buffer N3 were added quickly to avoid DNA degradation and the suspension was mixed by inverting the tube until all trace of blue was gone, that indicated the SDS was precipitated. The suspension was then centrifuged for 10 min at 13.000 rpm. The supernatant was applied to the QIAprep spin column and centrifuged for 30-60 s. The column was washed by adding 0.5 mL Buffer PB and centrifuging for 30-60 s, and then 0.75 mL Buffer PE and again centrifuging. To elute DNA 50 µL Buffer EB were added to the center of each QIAprep spin column and centrifuged for 1 min. The obtained fosmid amount was spectrophotometrically measured.