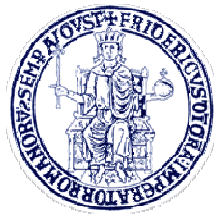


UNIVERSITY OF NAPLES FEDERICO II  
PHARMACY DEPARTMENT



*Ph.D. programme in*  
**PHARMACEUTICAL SCIENCE**  
*XXVII CYCLE (2012-2015)*

***From natural products chemistry to HTS by  
NMR: different approaches in drug discovery***

Ph.D. Chiara Del Gaudio

Tutors:

Prof. F. Zollo  
(University Federico II)

Prof. M. Pellecchia  
(Sanford-Burnham MRI)

Coordinator:

Prof. M.V. D'Auria



*Give the ones you love*

*wings to fly,*

*roots to come back,*

*reasons to stay.*



# INDEX

<b>ABSTRACT.....</b>	<b>1</b>
<b>INTRODUCTION.....</b>	<b>4</b>
<b>CHAPTER 1.....</b>	<b>8</b>
ISOLATION AND STRUCTURAL CHARACTERIZATION OF MARINE NATURALLY-OCCURRING COMPOUNDS	
1.1 Research steps.....	10
1.1.1 Isolation and purification procedures.....	10
1.1.2 Structural characterization methods.....	12
1.1.2.1 Stereochemistry determination.....	14
1.2 <i>Sinularia ineleigans</i> .....	16
1.2.1 Isolation and structural characterization of norcembranoids.....	18
1.2.2 Structural determination of 5- <i>epi</i> -norcembrenolide.....	20
1.2.3 Conclusions.....	24
<b>CHAPTER 2.....</b>	<b>26</b>
DESIGN, SYNTHESIS AND STRUCTURE-ACTIVITY RELATIONSHIP OF BILE ACIDS DERIVATIVES.	
2.1 Dual agonists of nuclear and membrane bile acid receptors.....	27
2.2 Total synthesis of INT-767.....	28
2.3 Total synthesis of INT-767 derivatives.....	31
2.3.1 Chenodeoxycholan sulfate derivatives.....	31
2.3.2 Ursodeoxycholan sulfate derivatives.....	32
2.4 Pharmacological evaluation.....	33
2.5 Docking studies.....	38
2.6 Conclusions.....	44
<b>CHAPTER 3.....</b>	<b>46</b>

## FRAGMENT-INSPIRED HTS BY NMR OF COMBINATORIAL LIBRARIES: APPLICATION TO ANTAGONISTS OF PROTEIN-PROTEIN INTERACTIONS

3.1 Design and optimization of potent inhibitors of XIAP, a member of the Inhibitor of Apoptosis Proteins family.....	50
3.1.1 Library, screening and results.....	52
3.1.2 Lead optimization.....	55
3.1.3 Computational studies .....	58
3.1.4 Conclusions.....	59
3.2 Discovery and optimization of hit compounds binding EphA3, a tyrosine kinase receptor involved in angiogenesis and metastasis formation.....	60
3.2.1 EphA3-LBD expression .....	62
3.2.2 Binders design and optimization .....	63
3.2.3 Conclusions.....	66

## EXPERIMENTAL SECTION

I. Experimental procedures for <i>Sinularia inelegrans</i> .....	68
II. Experimental procedures for FXR/GP-BAR1 dual agonists.....	79
III. Experimental procedures for XIAP-BIR3 inhibitors.....	107
IV. Experimental procedures for EphA3 inhibitors.....	112

REFERENCES.....	114
-----------------	-----

**ABSTRACT**

---

The process of drug discovery, by which new potential medicines are discovered, is a long and intricate path that can be approached with many different strategies. For a long time, drugs have been discovered by studying secondary metabolites produced by plants and other organisms mainly as defense or as a way to interact with competitors and mutualists. During recent years, high-throughput screening (HTS) of large compounds libraries has become very popular since it promised to drastically fasten the process of hit identification though requiring consistent initial investments. However, despite the rise of combinatorial chemistry as an integral part of the lead discovery process, natural products still play a major role as starting material for drug discovery. In fact, a 2012 report<sup>1</sup> found that of the 175 small molecule new chemical entities developed as anticancer agents since 1940s, 48.6% were natural derived or semi synthetic derivatives of natural products.

During my PhD, I had the chance to work in different laboratories thus gaining knowledge in different fields ranging from natural products chemistry to HTS by NMR, two approaches that converge in the common purpose of identifying and optimizing pharmacologically active molecules.

At the department of Pharmacy, in Naples, under the guidance of Prof. D'Auria and Prof. Zampella, I focused on isolation, characterization and synthetic studies of naturally-occurring compounds working on two different projects: isolation and characterization of secondary metabolites from natural sources; and design, synthesis and pharmacological evaluation of dual agonists of bile acid receptors.

The investigation of bioactive natural products from the Indian soft coral *Sinularia ineleans* led to the isolation of a novel norcembranoid, named 5-*epi*-norcembrenolide, along with twelve known compounds, which were characterized by means of high resolution mass spectrometry and 1D and 2D NMR experiments. All the isolated compounds, including sinuleptolide and 5-*epi*-sinuleptolide for which anti-inflammatory and antiviral activities had already been reported, are actually under pharmacological investigation to evaluate their antimicrobial activity and binding to nuclear receptors.

As regards dual agonists of bile acid receptors, the project involved a preliminary synthetic study in order to produce the reference compound INT-767 and several analogs that were all synthesized starting from the commercially available chenodeoxycholic and ursodeoxycholic acids. Semi synthetic derivatives were tested *in vitro* and *in vivo* to evaluate their activity on the bile acid receptors FXR and GP-BAR1. One of the compounds turned out to be the most potent dual FXR/GP-BAR1 agonist so far reported and docking studies elucidated its binding mode in the ligand binding domains of the two receptors revealing the structural requisites to achieve potent GP-BAR1/FXR dual agonism. This study is relevant for further investigations on the functional mechanism of these two receptors and for the design of novel dual GP-BAR1/FXR agonists, providing new opportunities for the treatment of enterohepatic and metabolic disorders.

The second part of my thesis was carried out at Sanford-Burnham Medical Research Institute, in San Diego, California, tutored by prof. Pellecchia and Dr. Barile. I applied a new screening technique known as HTS by NMR - which involves a target-based screening of combinatorial libraries using NMR as detection method - to the study of protein-protein interactions.

The research conducted on the BIR3 domain of XIAP, an inhibitor of apoptosis protein, resulted in the identification of an inhibitor with a potency comparable to the one of the clinical candidate GDC-0152 by Genentech ( $K_d \sim 30$  nM), which went through a phase I clinical trial as anticancer agent, but endowed with a more favorable thermodynamic profile and a higher binding specificity. The compound was subjected to a process of medicinal optimization and three prodrugs are currently being tested to evaluate their ability of delivering the active molecule inside the cells.

At the same time, I studied the ligand binding domain of EphA3, a tyrosine kinase receptor overexpressed in several types of cancer and involved in stem cells maintenance. The EphA3-LBD has been expressed for the first time in a soluble  $^{13}\text{C}$ -Met-labeled form. By using HTS by NMR as screening method we identified first small molecule EphA3 binders endowed with good selectivity between the homolog receptors EphA3 and EphA4, and with dissociation constants against EphA3 in the range of 5-10  $\mu\text{M}$ . These binders will be further optimized in order

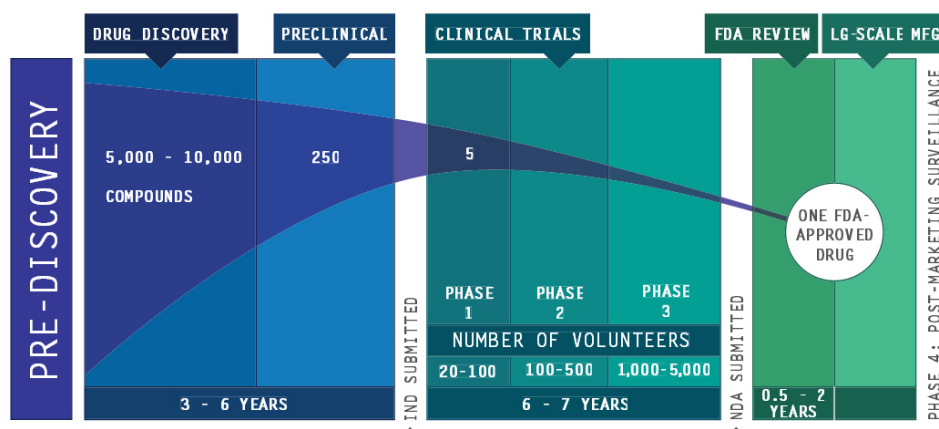
to afford a useful tool to deepen our knowledge of EphA3 pathological role in cancer development.

## **INTRODUCTION**

---

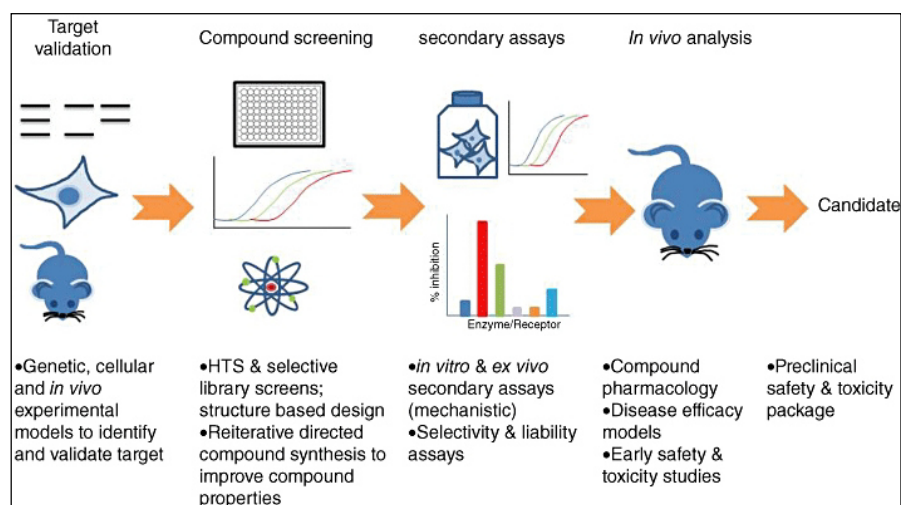
The search for small molecule therapeutics for the treatment of diseases has played a critical role in drug discovery for many years. In fact, the use of herbs and natural extracts for healing purposes goes back thousands of years; however, it has only been in the past half century or so that searching for new drugs has become a proper science. In 1900, one-third of all deaths were caused by three main diseases - pneumonia, tuberculosis, and diarrhea - that are today preventable and/or treatable thanks to scientific research. Among them, pneumonia is the only one remaining a common cause of death, which is now usually due to more complex conditions such as cardiovascular diseases and cancer. The availability of drugs to control infections, hypertension, hyperlipidemia, and to some extent even cancer, together with improved sanitation and vaccination, increased the life expectancy during the twentieth century from less than 50 years in 1900 to more than 77 years in 2000. These are certainly great achievements for the scientific society but still many life-threatening conditions are in need of a cure and plenty of new drugs are being developed to provide it.

The process of drug discovery, by which new potential medicines are discovered, is a long and complex procedure. Developing a new drug from the original idea to the launch of a finished product can take 12–15 years and cost more than \$1 billion.<sup>2</sup> Briefly, a drug discovery programme initiates because there is a disease or clinical condition without suitable medications available and this unmet clinical need represents the driving motivation for a new project.



**Figure 1.** Drug discovery process from target identification and validation to filing of a compound and the approximate timescale for these processes. FDA, Food and Drug Administration; IND, Investigational New Drug; NDA, New Drug Application. Image taken from Alzheimer's Drug Discovery Foundation.

The initial research (pre-discovery phase), often occurring in academia, aims at developing a hypothesis that the inhibition or activation of a specific protein or pathway will result in a therapeutic effect for a disease state. The outcome of this activity is the selection and validation of a target which brings to the lead or drug discovery phase. During lead discovery, a rigorous research results in finding a drug-like small molecule or a biological therapeutic, typically defined development candidate that, after intensive and iterative optimization, will progress into preclinical and, if successful, into clinical development and ultimately will become a marketed medicine (Figures 1 and 2).



**Figure 2.** Overview of drug discovery screening assays. Picture taken from Hughes *et al.*, *Br J Pharmacol.* **2011.**<sup>2</sup>

My research has been focused on the hit identification and hit-to-lead optimization processes. A ‘hit’ molecule can be defined as a compound which has the desired pharmacological activity in a preliminary screen and whose activity is confirmed upon deeper retesting. A variety of approaches exist to identify hit molecules but they can be basically divided into two main groups: phenotypic and target-based screening. The former measures the effects that compounds induce in cells, tissues or pathology models, while the latter looks at the effect of compounds on isolated target proteins via *in vitro* assays. At the beginning, drug discovery was mainly performed by phenotypic assays which don’t require previous knowledge of the molecular basis of diseases. Traditionally, many drugs have been discovered by phenotypic screening of natural extracts and identifying the single molecules responsible for the activity. On the other hand, since the beginning of the genomics era in the 1990s, the main focus of drug discovery has been on drug targets, leading to the so called “rational” drug design. Tremendous advances have been made in the development of new tools to identify compounds interacting with protein targets, for example high-throughput target-based screening assays specifically designed for protein families such as G protein-coupled receptors and kinases. Structure-based tools that can be used to aid lead identification and optimization for some targets have also been developed, including X-ray crystallography and computational modeling and screening (virtual screening). All these progresses were made thanks to huge investments in drug research and development (R&D) in recent decades. However, the number of innovative new medicines approved per year by the US Food and Drug Administration (FDA) has not increased as expected, bringing investors to question the effectiveness of the new approaches.

In the 10 years between 1999 and 2008, 100 new molecular entities (NMEs) were discovered through target-based approaches, 58 NMEs were discovered using phenotypic-based approaches, 18 NMEs were derived from modifications of natural substances and 56 agents were biologics.<sup>3</sup> These data show how new technologies, such as high-throughput target-based screening, are noteworthy for the science progression but also that traditional approaches such as natural products chemistry are still valuable and productive. Newman and Cragg

periodically publish a review about the impact of natural products as source of new drugs.<sup>1</sup> From the data they presented, it is clear that the utility of natural products as sources of novel structures, but not necessarily the final drug entity, is still alive and thriving. For example, from the 1940s to date, 175 new small molecules were found to work as anticancer agents and 48.6% of them are either natural products or directly derived from them.

During my PhD course, I was lucky enough to have an insight into the different approaches described so far so that I can claim quite a deep knowledge of the drug discovery process. In the following chapters, I will describe in details my research work in the natural products chemistry and in the target-based high-throughput screening fields. My results demonstrate how both approaches have been successfully applied in the discovery of pharmacologically active small molecules that represent lead compounds for the development of a variety of therapeutic agents.

**CHAPTER 1**

---

**ISOLATION AND STRUCTURAL CHARACTERIZATION OF MARINE NATURALLY-OCCURRING COMPOUNDS**

Nature represents an endless arsenal of new bioactive molecules and the study of secondary metabolites has historically been of immense benefit in the drug discovery process. The study of terrestrial natural-occurring compounds can easily be traced back to the beginning of the XIX century when scientists aimed at the discovery of the molecules responsible for the pharmacological activity of plants used for centuries by the popular medicine. On the other hand, the natural products chemistry associated with marine species only emerged over the past 65 years mainly as a result of the improvement of collection techniques such as scuba diving.

Oceans cover more than 70% of the earth's surface and contain more than 300,000 species of plants and animals.<sup>4, 5</sup> The diversity in species is extraordinarily rich on coral reefs and the Indo-Pacific Ocean exhibits the world's greatest tropical marine biodiversity. This huge level of biodiversity lead to a massive number of interactions between different organisms that is directly related to the abundance and the activity of secondary metabolites produced by the above-mentioned organisms. These secondary metabolites, supposed to be the result of evolutionary pressure such as predation or competition for space and resources, are shaped into structurally diverse and usually stereochemically complex compounds, many of which belonging to novel chemical groups never found before in terrestrial organisms. They often possess specific and pronounced biological activity that has been demonstrated in antitumor, anti-inflammatory, analgesia, immunomodulation, allergy, and anti-viral assays.<sup>6</sup> Secondary metabolites are mainly produced by marine invertebrates (sponges, tunicates, soft corals), often with the contribution of symbiotic microorganisms such as bacteria, cyanobacteria and fungi, even though the effective role of microorganisms to the secondary metabolism of marine invertebrates has not been fully understood yet. The production of secondary metabolites represents a strategy for survival and, since a potent biological activity should be considered as a rare molecular

property, the wider is the number of secondary metabolites produced by an organism, the more chances it has to win the evolutionary competition. For this reason, it is not surprising that a thorough chemical analysis of a single marine invertebrate, carried out with non-destructive spectroscopic techniques, can afford tens or even hundreds of secondary metabolites.

Summarizing, simple natural organisms, as marine invertebrates, are able to generate new chemical structures through a multitude of different pathways and thus, natural products constitute a potentially infinite source of chemical diversity, unmatched by any synthetic chemical collection or combinatorial library. Natural libraries are valuable as starting point to design and develop new therapeutics and to gain preliminary structure-activity relationships data. However, it should be addressed that the discovery of a new bioactive natural molecule represents just the starting point of the long path leading to a marketable drug so that natural products are often referred to as “lead compounds”. The identification of a lead compound must be followed by a deep investigation of the mode of action and the drug-likeness of the molecule. In this context, the most interesting and promising natural products are small to medium molecular weight compounds.

In spite of the difficulties associated with the limited availability of the compounds under investigation, which is strictly related to the limited supply of the biological material correctly protected for environmental concerns, some interesting results have been obtained. Through the combined efforts of marine natural product chemists and pharmacologists, an remarkable array of promising compounds have been identified so far. Some of these molecules are either at advanced stages of clinical trials or have been selected as promising candidates for pre-clinical evaluations. The majority of these compounds show antimicrobial and/or anticancer activities. Just to cite an example, trabectedin is the first approved anticancer drug derived from marine sources. The extract from the sea squirt *Ecteinascidia turbinata* was found to have anticancer activity back in 1969.<sup>7</sup> In 2007, after almost 40 years of investigations and technology improvements, the EMEA gave the authorization for the marketing of trabectedin for the treatment of advanced soft tissue sarcoma.

## 1.1 RESEARCH STEPS

Marine sampling starts with freeze-drying of the collected material until taken back to the laboratory. Samples are frozen at  $-20^{\circ}\text{C}$  till they are ready to be processed. Usually, a room temperature extraction affords the crude extract that is fractionated by liquid-liquid partitioning procedures followed by separation and isolation of the individual components using chromatographic techniques (liquid chromatography, column chromatography and high performance reversed phase liquid chromatography). Structural elucidation of isolated compounds is carried out using spectroscopic techniques, especially 2D nuclear magnetic resonance (NMR) and mass spectrometry (MS).

Summarizing, the steps for the study of natural-occurring compounds can be outlined as follows:

1. Isolation and purification of compounds from biologic material;
2. Structural determination of isolated compounds;
3. Determination of absolute and relative stereochemistry of new compounds.

### *1.1.1 ISOLATION AND PURIFICATION PROCEDURES*

Extraction is the crucial first step in the analysis of marine organisms, because it is necessary to extract the desired chemical components for further separation and characterization. Proper actions must be taken to assure that potentially active constituents are not lost, distorted or destroyed during the preparation of the extracts. Basic operations include steps such as pre-washing and grinding to obtain a homogenous sample and to improve the kinetic of extraction increasing the contact of the sample surface with the solvent system. The election solvent is methanol or a mixture methanol-chloroform (2:1). Usually, three overnight extractions with fresh solvent are sufficient for a quantitative recovery of the constituents, affording the so called crude extract.

The isolation of the constituents poses several problems, because many compounds may only be present in sub-milligram amount. The nature of the separation problem varies considerably, from the isolation of small quantities (milligrams or less) for structure determination purposes to the isolation of larger amounts (hundreds of milligrams to gram quantities) for comprehensive

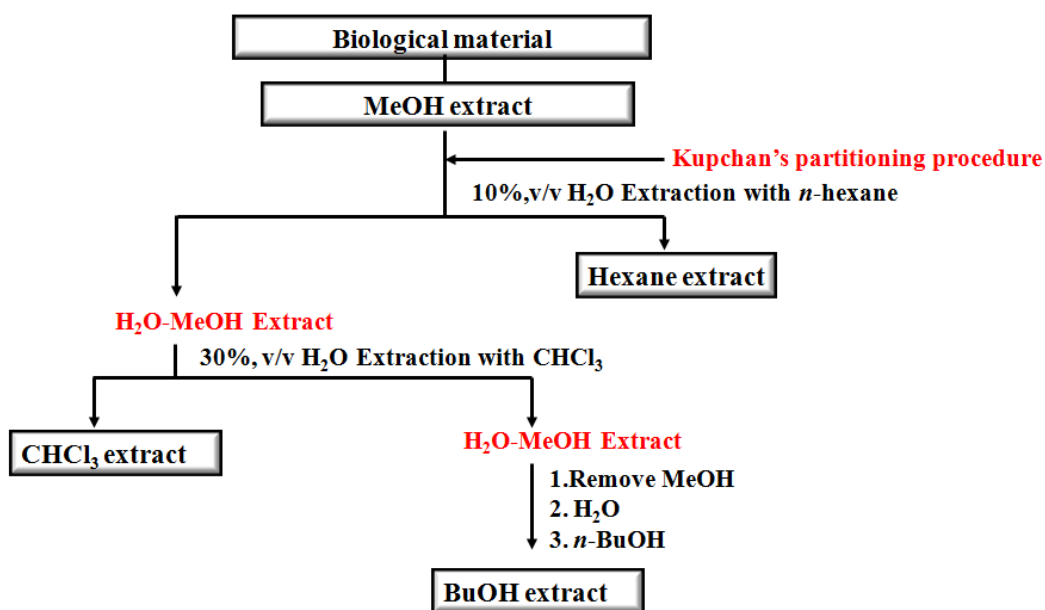
biological studies and semi-synthetic optimization. For these purposes, a good selection of different techniques and approaches is essential.

Several partitioning procedures can be applied however, the election method is typically the modified Kupchan's partitioning methodology,<sup>8</sup> which allows to fractionate the crude extract into four enriched extracts. During the process, the constitution of the aqueous phase is adjusted sequentially as follows (Figure 3):

- The methanol extract is dissolved in a mixture of 10% H<sub>2</sub>O/MeOH and partitioned against *n*-hexane; the so obtained hexane extract is rich in lipophilic compounds.
- The water content (% v/v) of the aqueous phase is adjusted to 30% and partitioned against CHCl<sub>3</sub>; compounds with medium polarity (peptides, macrolides, polyfunctionalized sterols, and so on) are usually found in this fraction.
- The aqueous phase is concentrated to remove any trace of MeOH and then extracted with *n*-BuOH; this extract contains polar compounds and salts.

The obtained fractions (*n*-hexane, CHCl<sub>3</sub>, *n*-BuOH and H<sub>2</sub>O extracts) can be tested in preliminary bioassays to identify interesting pharmacological activities or analyzed by NMR looking for chemical diversity.

The most promising extracts are further fractioned through sequential chromatographic methods. At an early stage, methods with a high loading capacity and cheap stationary phases are required. These have traditionally been column chromatography on silica gel or liquid-liquid methods. On the other hand, following steps take advantage of smaller scale and more performing techniques like HPLC.



**Figure 3.** Modified Kupchan's partitioning procedure.

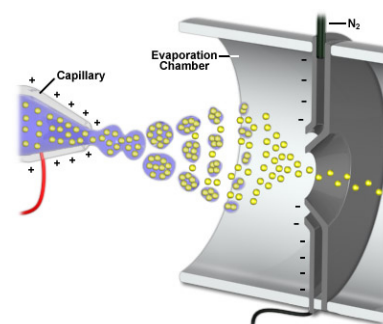
One of the preparative techniques I used in my research is the Medium Pressure Liquid Chromatography (MPLC). This is a liquid-solid chromatography in which the liquid mobile phase is forced through the solid stationary phase at medium pressure. As far as resolution is concerned, MPLC lies somewhere between flash chromatography and semi-preparative HPLC, being faster than flash chromatography and having a higher loading capacity compared to HPLC. The solid stationary phase can be a direct phase, like silica gel, or a bonded phase (RP-8, RP-18). MPLC works under a pressure of 5-40 bar and can easily be loaded with samples in the range of 100 mg-100 g.

High Performance (or High Pressure) Liquid Chromatography (HPLC) with direct or reversed-phase, is the most common chromatographic method and can be used with both analytical and preparative purposes. Usually, HPLC is applied as last step in the purification process affording compounds with a purity of 98-99%, suitable for biological tests.

### 1.1.2 STRUCTURAL CHARACTERIZATION METHODS

Structural characterization of complex organic structures is accomplished in a non-destructive way, with sub-milligram samples through the combined use mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR).

**MS** is an analytical technique 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 interest. Even being a destructive method, it just requires few



**Figure 4.** ESI source

micrograms of compound. A mass spectrometer can be schematized as follows: the *source* converts sample molecules into ions in the gas phase, the *analyzer* separates them according to their  $m/z$  ratio and the *detector* records the individual ion current intensities at each mass giving the mass spectrum. Many different types of sources, as well as analyzers, can be used. The compounds described later were analyzed by electrospray (ESI) mass spectrometry. The ESI source is particularly useful in generating ions from polar compounds or macromolecules. The sample is simply prepared by dissolving the compound in a volatile solvent like  $\text{H}_2\text{O}$ ,  $\text{MeOH}$ , or  $\text{CH}_3\text{CN}$  and volatile acids, bases or buffers are often added to the solution. The solution is forced through an electrospray needle to which a high potential difference is applied (Figure 4). This causes the spraying of droplets with a surface charge of the same polarity of the needle. Droplets are repelled from the needle towards the source sampling cone on the counter electrode. As evaporation of the solvent occurs, droplets shrink till they reach the Rayleigh limit, that is the point that the surface tension can no longer sustain the charge, then a "Coulombic explosion" occurs and droplets are ripped apart. The explosion generates smaller droplets that repeat the process until ionized molecules escape from the droplets by electrostatic repulsion. Electrospray is a very mild ionization method and its major disadvantage is that very little, usually no fragmentation at all, is produced. Therefore, the requirement of more complex mass spectrometry studies (e.g. tandem MS where the analyte molecules can be fragmented) for a complete structural elucidation.

**NMR** is the most important spectroscopic technique used for structure characterization of the isolated compounds. In addition to standard 1D  $^1\text{H}$  and  $^{13}\text{C}$

NMR spectra, a combination of 2D NMR experiments allows to fully characterize even extremely complex molecules.

COSY (COrrrelation SpectroscopY) is a homonuclear 2D experiment that allows the identification of protons that are scalarly coupled, spin systems, and is very useful when the multiplets overlap or when extensive second-order couplings complicate the 1D spectrum.

TOCSY (TOtal Correlation SpectroscopY) shows the correlations between all protons within a given spin system, not only between geminal or vicinal protons as in COSY. Correlations are outlined between distant protons as long as there are couplings between every intervening proton. This is extremely useful to identify protons on sugar rings in oligosaccharides or amino acids in peptides: all protons on a given sugar ring will have a correlation with all other protons on the same ring but not with protons on different rings.

ROESY (Rotating-frame Overhauser Effect SpectroscopY) outlines spatial proximity so, protons that are not directly bonded but are close in space (1-3 Å) usually show a ROE correlation. ROESY is really valuable to determine relative stereochemistry and conformational exchange. ROE is similar to NOE (Nuclear Overhauser Effect), 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 medium-size molecules, which would show a NOE close to zero.

The HSQC (Heteronuclear Single Quantum Correlation) is a 2D NMR correlation in which only one-bond proton-carbon couplings ( $^1J_{CH}$ ) are observed. Instead, HMBC (Heteronuclear Multiple Bond Correlation) identifies  $^{2,3}J_{CH}$  coupling constants, with  $^3J$  correlations usually stronger than  $^2J$ .

#### *1.1.2.1 STEREOCHEMISTRY DETERMINATION*

The determination of absolute and relative stereochemistry is crucial for the characterization of natural compounds. The stereochemistry involves the study of the relative spatial arrangement of atoms within molecules, their three-dimensional position in space. The stereochemistry of a molecule is a fundamental concept inextricably linked to molecular recognition. In fact, biological activity is

strictly related not to the planar structure but to the 3D arrangement of the molecule. It is now widely recognized that enantiomeric drugs may have different interactions with the human system, thus behaving very differently as regards their activity, toxicology, adsorption, metabolism, pharmacokinetics and bioavailability.<sup>9</sup> Thus, knowledge of the stereochemistry is important to understand the biological activity, the drug-receptor interaction, and also in case there is the need to synthesize greater amount of the product. The majority of natural products has one or more stereogenic centers. Usually, in a step-by-step structural elucidation, the first objective is the identification of relative stereochemistry of chiral centers. Then, after the elucidation of absolute stereochemistry even at a single stereogenic carbon, it will be possible to infer the absolute configuration of the entire structure. Structural information from NMR experiments comes primarily from the chemical shift of a nucleus that is sensitive to the environment and is influenced by geminal and vicinal nuclei. Coupling constants ( $J$ ) between pairs of protons separated by three or less covalent bonds can be measured. The value of a three-bond  $J$  contains information about the intervening torsion angle ( $\theta$ ). This is called Karplus relationship<sup>10</sup> and has the form:

$$^3J = A \cos^2 \theta + B \cos \theta + C$$

where  $A$ ,  $B$ , and  $C$  are empirically derived constants for each type of coupling constant. The value of these couplings is generally smaller when the torsion angle is close to  $90^\circ$  ( $^3J_{H-H} \sim 0-1.5$  Hz) and larger at angles of  $0^\circ$  and  $180^\circ$  (usually  $J_{180^\circ} > J_{0^\circ}$ ). Thus, it is relatively easy to discriminate between axial-axial and axial-equatorial or equatorial-equatorial correlations of two protons within a six-membered ring because the  $^3J_{ax-ax}$  is about 7-9 Hz while  $^3J_{ax-eq}$  and  $^3J_{eq-eq}$  are around 2.5 Hz. Moreover, it is possible to distinguish between *cis* and *trans* configuration of double bond protons ( $^3J_{cis} \sim 6-12$  Hz,  $^3J_{trans} \sim 14-20$  Hz).

The other major source of structural information are NOE<sup>11</sup> and ROE<sup>12</sup>, sensitive probes for short intra-molecular distances. NOE can be observed upon irradiation on a specific nucleus during the acquisition phase. In this way, the relaxation times of all the protons surrounding the irradiated one, even though not belonging to the same spin system, are influenced and this effect reflects into peaks intensity

changes. NOE intensity is proportional to the inverse of the sixth power of the distance separating two protons, so it decreases very quickly with distance and only protons that 2.5Å or less apart can be detected. NOEs are classified according to the location of the two protons involved. Intra-residual NOEs occur between protons within the same residue, whereas sequential, medium, and long range NOEs occur respectively between protons on residues sequentially adjacent, separated by 1, 2 or 3 residues, and separated by four or more residues in a polypeptide sequence. NOE effects cannot be evidenced for large molecules in which case ROESY can be used.

## 1.2 *SINULARIA INELEGANS*

Alcyonaceans, or soft corals, are among the most prolific producers of secondary metabolites in the marine environment with diterpenes and sterols as the majority of isolated compounds. As part of my research activity, I carried out a study of the secondary metabolites from a specimen of the soft coral *Sinularia inelegans*, kindly provided by Dr. Lisette D'Souza from the CSIR-National Institute of Oceanography, India.



**Figure 5.** *Sinularia sp.*

Soft corals are colonies of small animals known as polypoid cnidarians (shortened to polyps). These polyps rarely exceed 5 mm in diameter and are arranged in soft,

fleshy, irregularly shaped colonies of up to 1 meter in size.<sup>13</sup> Special organs, called ameobocytes, secrete calcium carbonate spicules which give rigidity and support to the colony. Spicules differ in shape and size within each genus and are actually the only way to identify different soft coral species.<sup>14</sup>

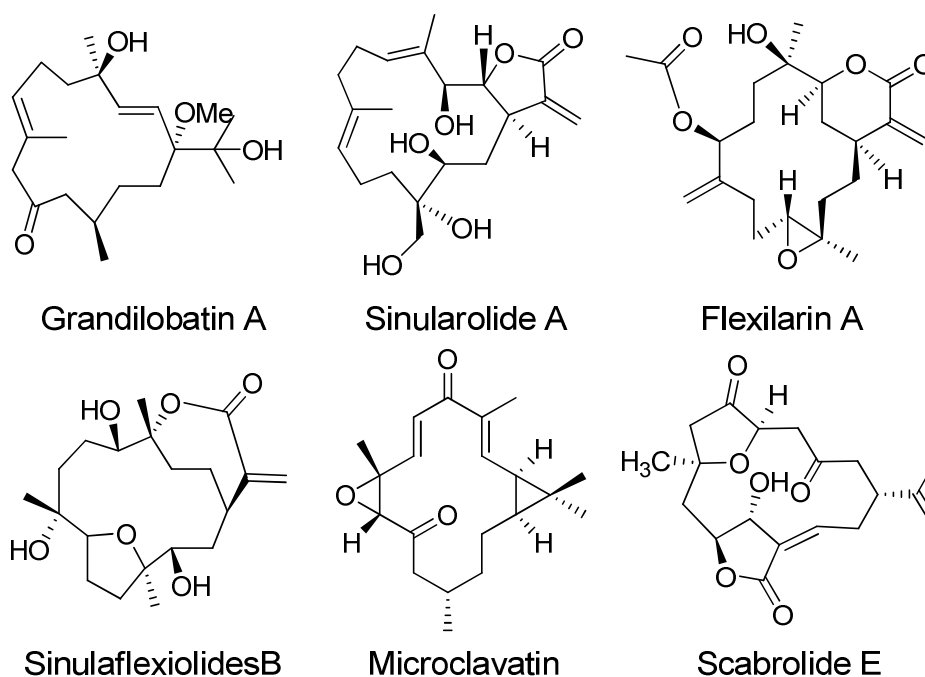
*Sinularia* corals are abundant on reefs in the Indo-Pacific Ocean including Fiji, Tonga, Solomon islands, Indonesia and the great barrier reef. The genus *Sinularia* consists of almost 90 species of which more than 50 have been chemically evaluated.<sup>15</sup> Metabolites which have been reported from the genus *Sinularia* include sesquiterpenes, diterpenes, polyhydroxylated steroids, and polyamine compounds. The peculiar structure and potential medicinal value of these metabolites have drawn increasing attention from chemists and pharmacologists. Cembranes are the most frequent secondary metabolites isolated from various *Sinularia* species. They are believed to have an important role within the chemical defense arsenal against other reef organisms.<sup>16</sup> Anti-inflammatory,<sup>17</sup> antimicrobial,<sup>18</sup> cytotoxic,<sup>19</sup> and antiviral<sup>20</sup> activities have been disclosed for members of the cembrane class. Their structures are based on a 14-membered diterpenoid, named cembrane, and are characterized by an isopropyl- and three methyl-substituted 14-membered rings, with structural changes in the position of double bonds, epoxidation, allylic and isopropyl oxidation, and carbon cyclization. Cembranoids can be divided into five categories according to the following structural features:

- 1) Isopropyl(ene)-type cembranoid diterpenes: the alteration in the structure of these diterpenes lies in the migration or increase of double bonds and a high degree of oxidation and hydroxylation (see grandilobatin A,<sup>21</sup> figure 6);
- 2)  $\gamma$ -lactone-type cembranoid diterpenes: they play a crucial role in the soft coral *Sinularia*. Many of them showed cytotoxicity and antitumor activity. In addition, the 5-membered lactone ring is the fundamental active center of the molecule (see sinularolide A,<sup>22</sup> figure 6);
- 3)  $\delta$ -lactone-type cembranoid diterpenes: to date only one compound belongs to this class (flexilarin A,<sup>23</sup> figure 6);
- 4)  $\epsilon$ -lactone-type cembranoid diterpenes:  $\epsilon$ -lactone-type cembranoid diterpenes often possess anti-cancer activities, especially compounds with a  $\beta$ -exocyclic

double bond. A large number of these diterpenes were reported from *S. flexibilis* (see sinulaflexiolide B,<sup>24</sup> figure 6);

5) casbane-type cembranoid diterpenes: are strictly related to the cembrane structure and only differ for the presence of a dimethyl-cyclopropyl moiety rather than the isopropyl residue fused to the 14-membered ring. Casbane diterpenes are extremely rare in nature and are usually found in some species of plants (see microclavatin,<sup>25</sup> figure 6).

Norcembranoids are those that lack a C18 carbon substituent in comparison with C20-cembranoids. The most common norcembranoid diterpenes are the 14-membered macrocyclic 3(2H)-furanone-based norcembranoids (see scabrolides E,<sup>26</sup> figure 6).



**Figure 6.** Some cembranoids and norcembranoids isolated from *Sinularia* sp.

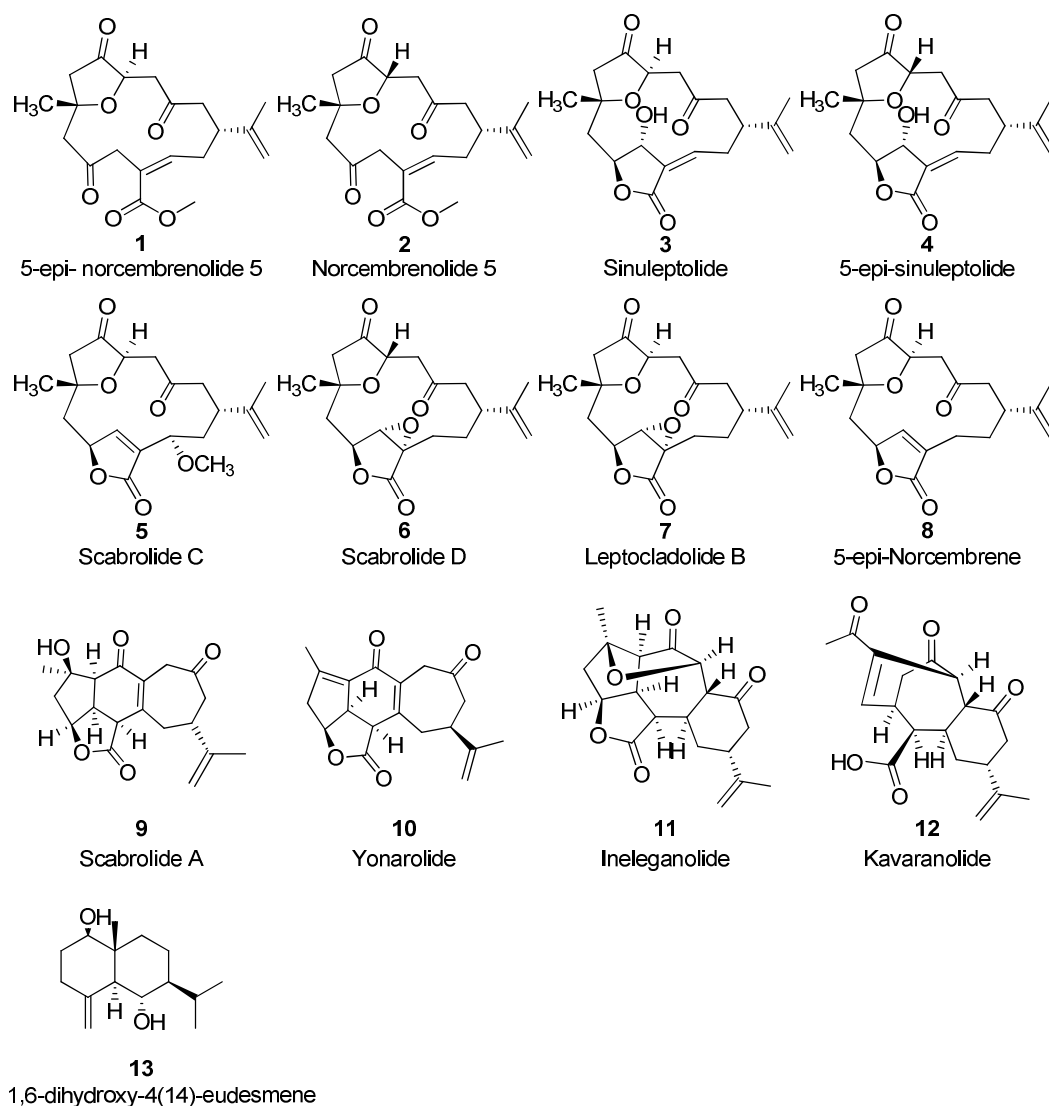
### 1.2.1 ISOLATION AND STRUCTURAL CHARACTERIZATION OF NORCEMBRANOIDS

A sample of *Sinularia ineleans* (order Alcyonacea, family Alcyoniidae) was collected in the Indian Ocean, immediately frozen and lyophilized. The crude methanolic-chloroformic (2:1) extract, provided by Dr. Lisette D'Souza, was

fractionated according to the Kupchan partitioning procedure<sup>8</sup>, affording four extracts.

The extracts were analyzed by NMR and the chloroformic extract, which seemed very rich in metabolites, was purified by MPLC (CH<sub>2</sub>Cl<sub>2</sub>:MeOH), followed by reverse-phase HPLC (MeOH aqueous 45%) to give one new compound, named 5-*epi*-norcembrenolide (**1**) and twelve known C-4 norcembranoids (**2-13**)<sup>27,28,29,30,31,32,33,34</sup> (figure 7).

The isolated known compounds were: seven 14-membered macrocyclic norcembranoids (norcembrenolide **2**<sup>31</sup>, sinuleptolide **3**<sup>35</sup>, 5-*epi*-sinuleptolide **4**<sup>33</sup>, scabrolide C **5**<sup>27</sup>, scabrolide D **6**<sup>27</sup>, leptocladolide B **7**<sup>30</sup> and 5-*epi*-norcembrene **8**<sup>31</sup>); two tetracyclic norcembranoids (scabrolide A **9**<sup>27</sup> and yonarolide **10**<sup>29</sup>); one C19-norcembranoid diterpene (ineleganolide **11**<sup>28</sup>); one tricyclic carbocycle norcembranoid (kavaranolide **12**<sup>34</sup>); one eudesmene derivative (1 $\beta$ ,6 $\alpha$ -dihydroxy-4(14)-eudesmene **13**<sup>32</sup>). The structures of the above mentioned compounds were defined by detailed analyses of their spectroscopic data (1D-2D NMR and MS) in comparison with values reported in literature.



**Figure 7.** New (1) and known norcembranoids isolated from *Sinularia inelegans*

### 1.2.2 STRUCTURAL DETERMINATION OF 5-EPI-NORCEMBRENOLIDE 5

5-*epi*-norcmbrenolide (1) was isolated as a white amorphous solid. Its molecular formula of  $C_{20}H_{26}O_6$ , implying eight degrees of unsaturation, was established by high-resolution ESI-MS based on the pseudo molecular ion  $[M + H]^+$  at  $m/z$  363.1763. The gross structure was deduced from detailed analysis of the  $^1H$  NMR spectroscopic data (Table 1), aided by 2D NMR experiments ( $^1H$ - $^1H$  COSY, HSQC, HMBC and ROESY).

**Table 1.**  $^1\text{H}$  and  $^{13}\text{C}$  NMR data (700 MHz,  $d_6$ -DMSO) of 5-*epi*-norcembrenolide (**1**).

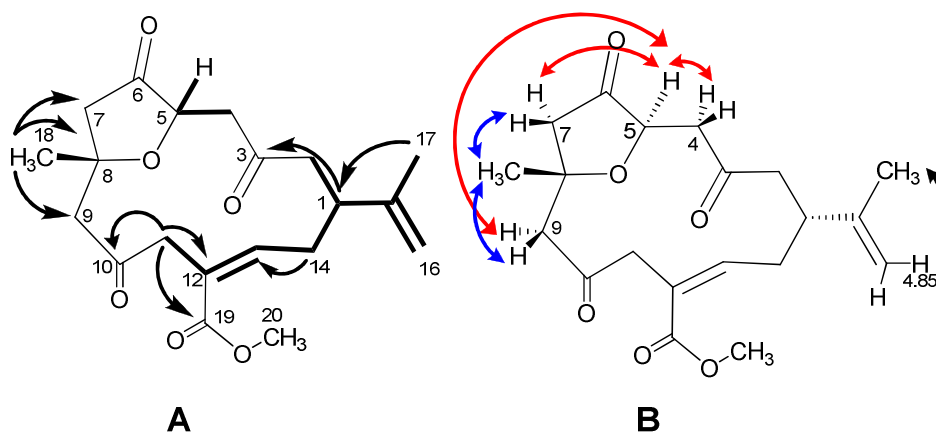
	$\delta_{\text{H}}^{\text{a}}$	$\delta_{\text{C}}$
<b>1</b>	2.71 m	40.1
<b>2</b>	2.65 s 2.37dd (13.4, 4.0)	44.6
<b>3</b>	-	208.0
<b>4<math>\alpha</math></b>	2.31 dd (20.1, 5.4)	46.4
<b>4<math>\beta</math></b>	2.75 dd (13.7, 1.5)	
<b>5</b>	4.48 d (11.3)	76.8
<b>6</b>	-	213.8
<b>7<math>\alpha</math></b>	2.68 <sup>c</sup>	48.9
<b>7<math>\beta</math></b>	2.50 <sup>b</sup>	48.9
<b>8</b>	-	79.8
<b>9<math>\alpha</math></b>	3.10 d (13.0)	50.1
<b>9<math>\beta</math></b>	2.32 <sup>c</sup>	50.1
<b>10</b>	-	206.6
<b>11</b>	3.51 d (18.0) 3.42 d (16.8)	43.5
<b>12</b>	-	129.0
<b>13</b>	6.69 t (6.9)	141.3
<b>14</b>	2.16 q (8.0) 2.25 m	29.4
<b>15</b>	-	145.6
<b>16</b>	4.68 s 4.85 s	112.9
<b>17</b>	1.69 s	20.6
<b>18</b>	1.34 s	27.1
<b>19</b>	-	167.0
<b>20</b>	3.63 s	52.2

<sup>a</sup> Coupling constants are in parenthesis and given in hertz.  $^1\text{H}$  and  $^{13}\text{C}$  assignments were aided by COSY, HSQC and HMBC experiments; <sup>b</sup> overlapped with the solvent signal; <sup>c</sup> overlapped with other signals.

The  $^{13}\text{C}$  NMR data confirmed the presence of 20 carbons, including three ketone signals at  $\delta_{\text{C}}$  206.6, 208.0 and 213.8, one methyl ester carbonyl signal at  $\delta_{\text{C}}$  167.0, one trisubstituted double bond ( $\delta_{\text{C}}$  129.0 and 141.3) and one disubstituted double bond ( $\delta_{\text{C}}$  112.9 and 145.6). The carbonyl and the olefinic carbons accounted for six degrees of unsaturation; hence the compound was bicyclic. Analysis of the  $^1\text{H}$  NMR spectrum revealed an isopropenyl group [ $\delta_{\text{H}}$  1.69 (3H, s), 4.68, (1H, s) and 4.85 (1H, s)], a deshielded olefin proton at  $\delta_{\text{H}}$  6.69 (1H, t,  $J = 6.9$  Hz) and a O-methyl singlet at  $\delta_{\text{H}}$  3.63.

Careful analysis of the COSY spectrum allowed to build up four different spin systems (figure 8), a small one made up by protons H<sub>2</sub>-4 and H-5; a large one

comprising H-11 to H<sub>2</sub>-14 and then H-1 and H<sub>2</sub>-2. The linkage between C-11 and C-13 across the quaternary C-12 was determined by the allylic coupling H<sub>2</sub>-11 to H-13, as well as the linkage of the iso-propenyl residue (C-15 to C-17) to H-1 was inferred by the allylic coupling of H<sub>2</sub>-16 and H<sub>3</sub>-17 to H-1. The two additional spin systems consisted of two diastereotopic isolated methylenes H<sub>2</sub>-7 ( $\delta_{\text{H}}$  2.50, 2.68) and H<sub>2</sub>-9 ( $\delta_{\text{H}}$  2.32, 3.10) flanked by quaternary carbons. The location of the three ketone functions at C-3, C-6 and C-10 was easily inferred from HMBC correlations H-5/C-3, C-4, C-6, C-7; H<sub>2</sub>-2/C3, H<sub>2</sub>-4/C-2, C-3, C-5, C-6 and H<sub>2</sub>-9/C-7, C-8, C-10, C-11 and C-18, and H<sub>2</sub>-11/C10, C12. The carboxyl ester group was placed at C-12 on the basis of HMBC correlations H<sub>2</sub>-11/C-12, C-19 and H-13/C-12, C-19. The chemical shift values of C-5 and C-8, together with MS data, were indicative of the ether linkage between these two centers to form a tetrahydrofuranone ring, common to several norcembranoids from *Sinularia*.<sup>37</sup> These data, together with additional HMBC correlations (figure 8) defined the planar structure of 5-*epi*-norcembranolide (**1**). The relative stereochemistry of compound **1** was deduced from the analysis of some key dipolar couplings evidenced by the ROESY spectrum (figure 8).

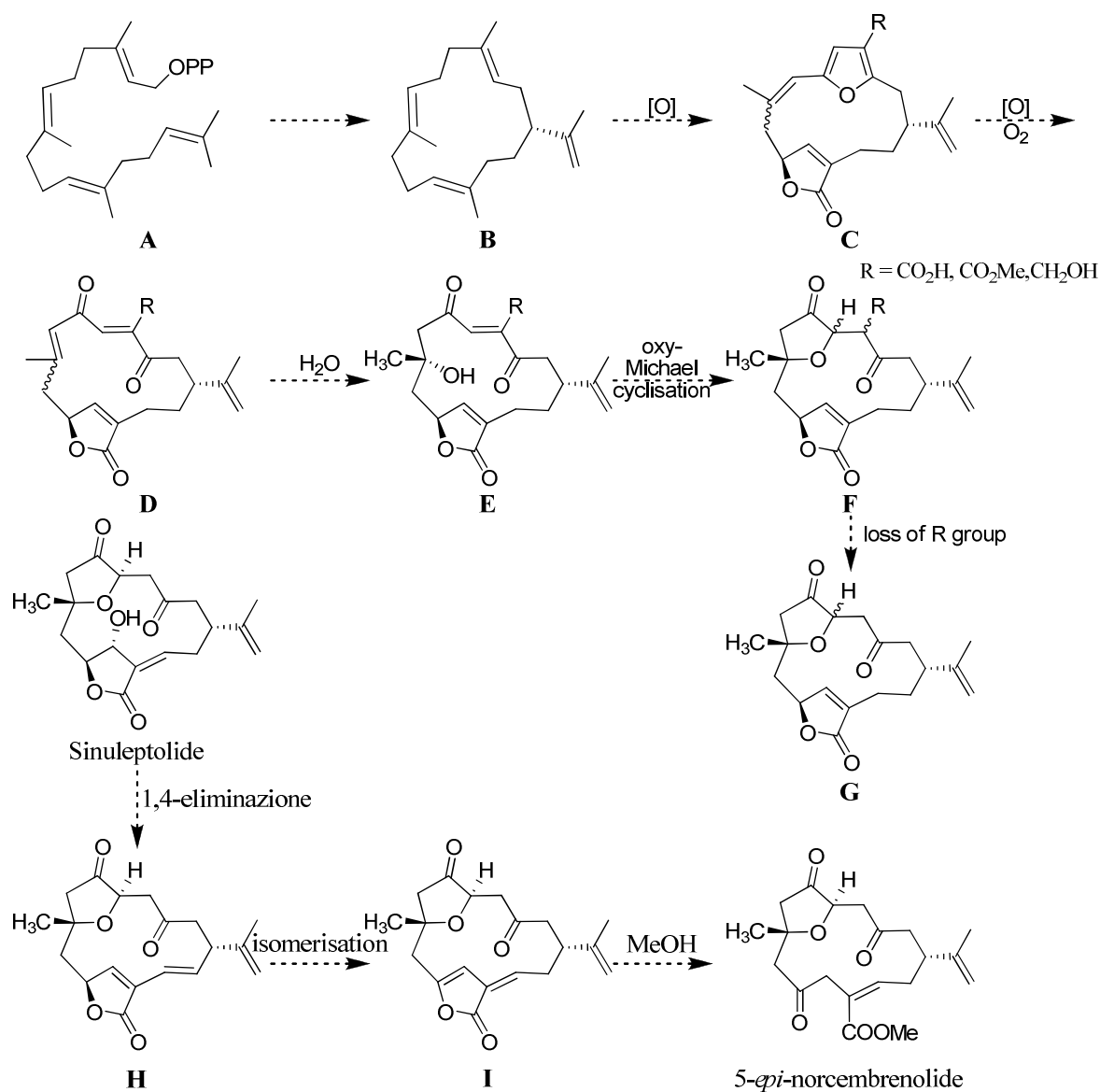


**Figure 8.** (a) COSY connectivities (bold bonds) and key HMBC correlations (black arrows); (b) key ROE correlations (colored arrows) for compound **1**.

The ROE correlations between H-5/H-4 $\alpha$ /H-7 $\alpha$  and H-9 $\alpha$  suggested that they are *syn* to each other. Furthermore, the correlations H<sub>3</sub>-18 and H-7 $\beta$ /H-9 $\beta$  indicated that the methyl substituent at C-8 is located on the  $\beta$ -face. Hence, H-5 and CH<sub>3</sub>-18 are *trans* to each other, that is in contrast with NMR data previously reported for norcembranolide<sup>31</sup>, co-occurring in the same specimen, which has H-5 and CH<sub>3</sub>-18

both on the same face. By analogy with sinuleptolide and 5-*epi*-sinuleptolide, co-occurring in the same specimen, whose *IR* absolute stereochemistry was defined by application of the Mosher's method,<sup>36</sup> we assigned the *IR* configuration. Therefore, the stereostructure of 5-*epi*-norcembranolid (1) was determined as depicted in figure 7.

Hypothesis on biosynthetic pathways for macrocyclic C19-norcembranoids, like sinuleptolide, have been reported previously by Li *et al.*<sup>37</sup> They speculated that the C19 macrocycle could derive from the furanobutenolide-based macrocyclic compounds<sup>38</sup> (**C**, figure 9), by far the largest group of C20-cembranoids found in the marine milieu, through the pathway shown in figure 9. Following this pathway, many compounds with a similar nucleus can be generated, such as sinuleptolide, norcembrene, leptocladolides, scabrolides and so on. Epimers at C5 presumably result from alternative, facially selective, hydrations and/or oxidations of the C7–C8 alkene bonds in the precursors **C** and **D** during their biosynthesis from furanocembranoids. Structure **G** is also the precursor for polycyclic norcembranoids such as ineleganolide. We can also suppose that norcembranolid and 5-*epi*-norcembranolid derive from 5-*epi*-sinuleptolide and sinuleptolide respectively. Sinuleptolide (**3**) can be converted into the corresponding 1,3-diene (**H**) by a 1,4-elimination process between C11 and C14, even though macrocyclic norcembranoids containing alkene functionality at C13–C14, e.g. compound **H**, have not yet been reported. The  $\beta$ -keto-ester functionality in 5-*epi*-norcembranolid (**1**) could be produced *in vivo* from the 1,3-diene unit between C11 and C14 in **H**. Thus, deprotonation at C10 in **H** accompanied by isomerisation to the enol lactone intermediate **I**, followed by lactone ring-opening and methanolysis, leads to **1** (figure 9). In a similar way, 5-*epi*-sinuleptolide would lead to generation of norcembranolid (**2**).



**Figura 9.** Proposal for the biosynthetic origin of **1**

### 1.2.3 CONCLUSIONS

The investigation of bioactive natural products from the Indian soft coral *Sinularia inelegrans* led to the isolation of a novel norcembranoid, named 5-*epi*-norcembrenolide (**1**) along with twelve known compounds (**2–13**) which were characterized by means of high resolution mass spectrometry and 1D and 2D NMR experiments. Many norcembranoids have been previously reported to exert cytotoxic effects against the proliferation of several cancer cell lines<sup>39</sup> and also to

have anti-inflammatory<sup>40,34</sup> and antiviral activities against human cytomegalovirus (HCMV)<sup>41</sup> and Chikungunya virus (CHIKV)<sup>34</sup>. The most active compounds of the series are sinuleptolide and its C5 epimer, probably due to the  $\alpha,\beta$ -unsaturated  $\gamma$ -lactone functionality. However, more exhaustive studies are needed to fully understand norcembranoids therapeutic potential. To this extent, all the isolated compounds are actually under pharmacological investigation to evaluate their antimicrobial activity and binding to nuclear receptors.

## CHAPTER 2

---

### DESIGN, SYNTHESIS AND STRUCTURE-ACTIVITY RELATIONSHIP OF BILE ACIDS DERIVATIVES

---

Bile acids (BAs) are the main components of bile and are produced from cholesterol catabolism. Primary bile acids, cholic acid (CA) and chenodeoxycholic acid (CDCA), are synthesized by the liver and then converted into secondary bile acids, deoxycholic and lithocholic acids (LCA), by the bacteria residing in the colon. Even though their physiologic role is mostly linked to nutrient metabolism and absorption,<sup>42</sup> BAs have recently been identified as key signaling molecules endowed with both genomic and non-genomic effects induced by the interaction with a number of nuclear receptors (FXR, PXR, CAR, VDR, LXRs) and G-protein-coupled receptors such as TGR5.

Farnesoid-X-receptor (FXR) is highly expressed in liver, intestine, kidney, adrenal glands, and adipose tissue and its main function is to maintain bile acids homeostasis. In fact, primary bile acids, like CDCA and its conjugated, are the most potent endogenous ligands.<sup>43,44</sup> FXR activation results in the formation of a heterodimer with the retinoid-x-receptor (RXR) that binds to specific response elements on the promoter of target genes, thus modulating the expression of enzymes/proteins involved in bile acid synthesis, transport, conjugation, and excretion. Activation of FXR inhibits BAs synthesis and protects against their toxic accumulation via increased conjugation in liver and increased secretion into bile canaliculi, thereby promoting bile flow. These effects can be beneficial in the prevention of gallstone formation<sup>45</sup> and in the treatment of cholestatic liver diseases such as primary biliary cirrhosis,<sup>46</sup> in which case targeting FXR can protect the liver from the development of fibrosis and cirrhosis via multiple pathways.<sup>47,48</sup> Besides the canonical function of bile acid sensor, FXR possesses several and pharmacologically interesting actions linked to lipids<sup>49</sup> and glucose homeostasis.<sup>50</sup> FXR activation, enhancing fibroblast growth factor 15 secretion in the small intestine, increases metabolic rate, reduces body weight, and reverses diabetes in mice fed high-fat diets and in leptin-deficient mice.<sup>51</sup> FXR agonists

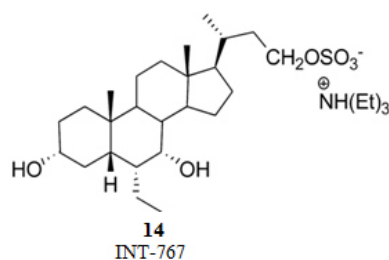
have been demonstrated to exert anti-fibrotic and anti-inflammatory<sup>52</sup> effects, therefore they could be valuable for the treatment of common human diseases such as metabolic syndrome and cancer.<sup>53</sup>

TGR5 (or M-BAR) is a member of the rhodopsin-like superfamily of G-protein-coupled receptors<sup>54</sup> also called GP-BAR1 (G-protein-coupled bile acid receptor 1). GP-BAR1 is mainly expressed in liver, small intestine, colon, adipose tissue, and skeletal muscle cells as well as in macrophages/monocytes. Upon ligation by secondary bile acids, lithocolic acid (LCA) and tauroolithocolic acid (TLCA),<sup>55,56</sup> GP-BAR1 triggers intracellular signals that allow the regulation of energy expenditure and glucose homeostasis.<sup>57</sup> In fact, GP-BAR1 activation in enteroendocrine L cells stimulates the secretion of glucagon-like peptide 1 (GLP-1), an insulintropic factor that enhances insulin release, gastrointestinal motility, and appetite.<sup>58</sup> Besides, this receptor has an important role in immunology and inflammation as it is able to reduce lipopolysaccharide (LPS) induced release of pro-inflammatory cytokines in animal models.<sup>59</sup> In rodent models of colitis, GP-BAR1 expression increases in inflamed tissues while its absence is linked to an increased intestinal permeability and enhanced susceptibility to colitis development.<sup>60</sup> Furthermore, GP-BAR1 seems to be crucial in maintaining the integrity of gastric and intestinal mucosa thanks to its protective action against gastrointestinal injury caused by non-steroidal anti-inflammatory drugs (FANS).<sup>61</sup>

## 2.1 DUAL AGONISTS OF NUCLEAR AND MEMBRANE BILE ACID RECEPTORS

Bile acids are potent signaling molecules that, through the activation of the nuclear receptor FXR and the membrane G protein-coupled receptor GP-BAR1, modulate their own homeostasis but also inflammation and lipid and glucose metabolism. These two receptors have already shown promising pharmacological activities. FXR activation inhibits BAs synthesis and has anti-inflammatory effects in atherosclerosis, inflammatory bowel disease and experimental cholestasis, whereas GP-BAR1 activation, via cAMP-mediated pathways, reduces proinflammatory cytokine production in macrophages and Kupffer cells. In addition, FXR and GP-BAR1 mutations have been identified in intrahepatic cholestasis of pregnancy and primary sclerosing cholangitis PSC, respectively.

Because FXR and GP-BAR1 have partially overlapping activities, ligands endowed with dual activity toward these receptors represent an interesting strategy to target entero-hepatic and metabolic disorders.<sup>62, 63</sup> It has been previously reported that INT-767 (6 $\alpha$ -ethyl-3 $\alpha$ ,7 $\alpha$ -dihydroxy-24-*nor*-5 $\beta$ -cholan-23-yl-23-triethylammonium sulfate, **14**, figure 10), one of the first dual FXR/GP-BAR1 agonist identified, exerts beneficial therapeutic mechanisms on liver inflammation and fibrosis in rodent models of metabolic syndrome.<sup>64,65,66,67</sup>



**Figure 10.** Previously reported dual GP-BAR1/FXR agonist.

My research activity was aimed at the development of FXR/GP-BAR1 dual agonists. Using INT-767 as template, my research group synthesized a new generation of dual bile acid receptor agonists among which compound **31** was revealed to be the most potent dual agonist of bile acid receptors so far identified. We believe this compound could pave the way towards the creation of a new therapeutic class for the treatment of diabetes.<sup>68</sup> We also elucidated its binding mode in the two receptors through a series of computations using the homology model of GP-BAR1 and the X-ray structure of FXR. Our study provide comprehensive structure-activity relationships for GP-BAR1/FXR dual modulation and is valuable for further investigations on bile acid receptors mode of action.

## 2.2 TOTAL SYNTHESIS OF INT-767

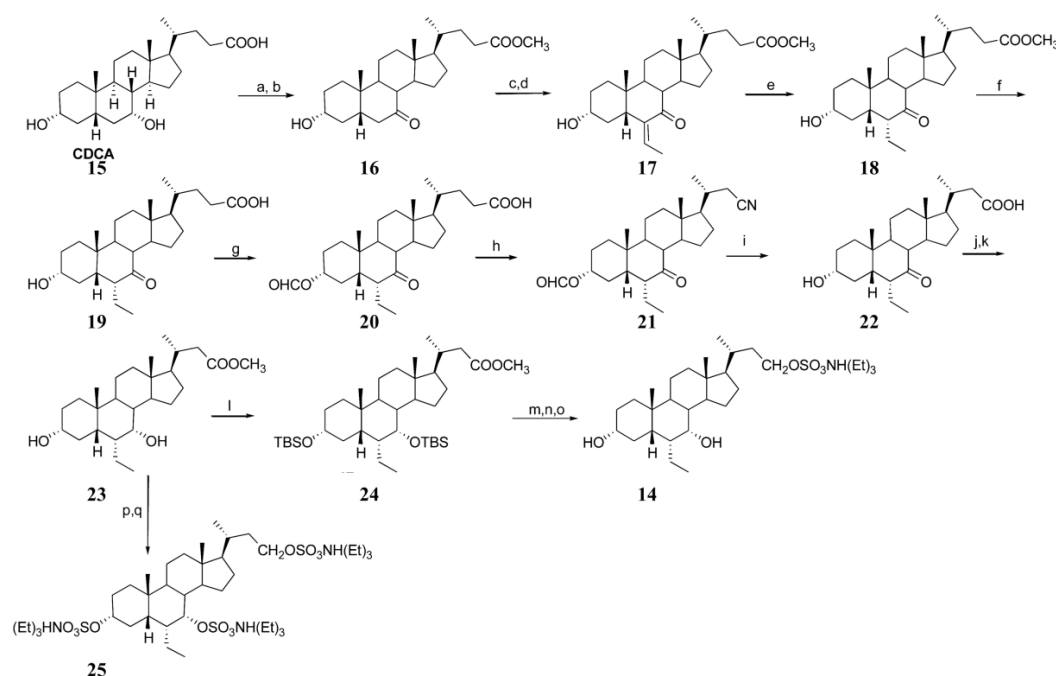
Even though INT-767 activity as dual GP-BAR1/FXR agonist<sup>65,66</sup> has been previously reported, to date no efficient synthetic procedure has been published. Therefore, my research work started with the total synthesis of INT-767 (**14**) so that we could have a positive control for biological assays. We derived compound **14** from the commercially available chenodeoxycholic acid (CDCA, **15**). Main differences between the two compounds are: the side chain that is C24 with a

carboxyl group in CDCA and C23 with a sulfate end-terminus group in INT-767; and the 6-ethyl substituent on ring B of compound **14**. The synthesis, reported in scheme 1, was realized over 15 steps and proceeded with a overall yield of 20%.

The starting material (CDCA, **15**) was regioselectively oxidized to the C7 hydroxyl group<sup>69</sup> using a mild oxidizing agent (sodium hypochlorite) and then the carboxylic acid at C24 was esterified to give the methyl ester of 7-ketolithocholic acid (**16**) (65% chemical yield over two steps). We used *in-situ* generated lithium diisopropylamide (LDA) on compound **16** to produce a silyl enol ether intermediate<sup>70</sup> that underwent a Knoevenagel aldolic addition to afford the methyl 3 $\alpha$ -hydroxy-6-ethylidene-7-keto-5 $\beta$ -cholan-24-oate **17** (77% yield). The use of low temperature (-78°C) and a strong base (LDA) accounted for the selectivity of enolate formation on C6. Hydrogenation of the exocyclic double bond, realized with H<sub>2</sub> on Pd/C, proceeded in a stereoselective way affording exclusively the 6 $\alpha$ -ethyl functionality that has an equatorial disposition demonstrated by the large coupling constant observed for the axial H-6 proton signal at  $\delta_H$  2.83 (dd,  $J = 13.0, 5.5$  Hz). Subsequently, alkaline hydrolysis of methyl ester group on the side chain gave the acid 3 $\alpha$ -hydroxy-6 $\alpha$ -ethyl-7-keto-5 $\beta$ -cholan-24-oic (**19**, quantitative yield). We protected the hydroxyl group on C3 by Fisher's esterification with formic acid and acetic anhydride<sup>71</sup> (quantitative yield) and then the performate derivative (**20**) was subjected to "Beckmann rearrangement"<sup>72</sup> with the purpose of shortening the side chain of one carbon unit from C24 to C23. Prolonged alkaline treatment of 23-nitrile derivative **21** allowed to hydrolyze the nitrile to carboxylic acid and simultaneously deprotect the 3-hydroxyl group, affording 6 $\alpha$ -ethyl-24-*nor*-7-ketolithocholic acid (**22**) (69% three-step chemical yield). C23 methyl ester derivative **23** was produced by sodium borohydride reduction of the C7-ketone followed by methanol/*p*-toluenesulfonic acid treatment (two steps, 89% yield). Again, this reduction was stereoselective as well as the previous one and 7 $\alpha$ -hydroxy configuration was confirmed by <sup>1</sup>H NMR spectrum. In fact, the shape of H-7 ( $\delta_H$  3.62, br s) was consistent with an equatorial disposition for this proton and the axial  $\alpha$ -orientation of the hydroxyl group. Final steps aimed at modifying the side chain. First, we protected the alcoholic functions at C3 and C7 as *tert*-butyldimethylsilyl derivatives then, methyl ester at C23 was reduced with lithium borohydride to a primary alcohol

that was finally sulfated with triethylammonium–sulfur trioxide complex at 95 °C (72% three-step chemical yield). Compound **14** was retrieved as triethylammonium salt after removal of hydroxyl protecting groups (MeOH/HCl) and purification by reversed-phase solid extraction on a C18 cartridge. HR ESIMS, negative ion mode  $m/z$  471.2795 (calculated 471.2780 for  $C_{25}H_{43}O_6S$ ), and NMR data (see experimental section) confirmed the chemical structure reported.

We also synthesized a derivative (**25**) of INT-767 characterized by three sulfate groups, respectively at C3 on ring A, C7 on ring B and C23 on the side chain, using compound **23** as starting material (scheme 1).



**Scheme 1.** Total synthesis of INT-767 and its trisulfate derivative **25**.

Reagents and conditions: (a) NaBr, TBABr, NaClO 10%, MeOH/AcOH/H<sub>2</sub>O/AcOEt 3:1:0.25:6.5; (b) p-TsOH, MeOH dry, 65% over two steps; (c) DIPA, n-BuLi, TMSCl, TEA dry, THF dry, -78 °C; (d) acetaldehyde, BF<sub>3</sub>(OEt)<sub>2</sub>, CH<sub>2</sub>Cl<sub>2</sub>, -60 °C, 77% over two steps; (e) H<sub>2</sub>, Pd/C, THF/MeOH 1:1, quantitative yield; (f) NaOH 5% in MeOH/H<sub>2</sub>O 1:1 v/v, quantitative yield; (g) HCOOH, HClO<sub>4</sub>, 81%; (h) TFA, trifluoroacetic anhydride, NaNO<sub>2</sub>; (i) KOH 30% in MeOH/H<sub>2</sub>O 1:1 v/v, 86% over two steps; (j) NaBH<sub>4</sub>, THF/H<sub>2</sub>O 4:1 v/v; (k) p-TsOH, MeOH dry, 89% over two steps; (l) 2,6-lutidine, tert-butyldimethylsilyl trifluoromethanesulfonate, CH<sub>2</sub>Cl<sub>2</sub>, 0 °C, 78%; (m) LiBH<sub>4</sub>, MeOH dry, THF, 0 °C, 95%; (n) Et<sub>3</sub>N·SO<sub>3</sub>, DMF, 95 °C; (o) HCl 37%, MeOH, 87% over two steps; (p) LiBH<sub>4</sub>, MeOH dry, THF; (q) Et<sub>3</sub>N·SO<sub>3</sub>, DMF, 95 °C, 72% over two steps.

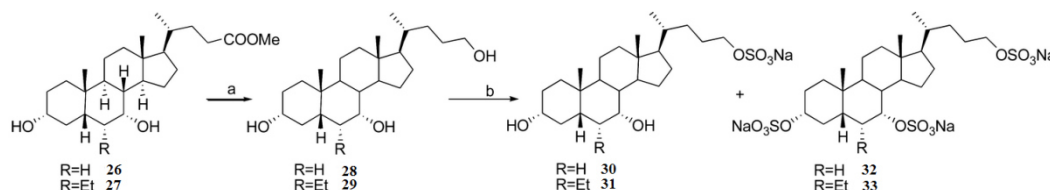
### 2.3 TOTAL SYNTHESIS OF INT-767 DERIVATIVES

We used the commercially available CDCA (**15**) and ursodeoxycholic acid (UDCA) for the development of a series of INT-767 related compounds in order to investigate the structural requisites for dual GP-BAR1/FXR agonism and to increase the agonism potency with reference to the starting compound (**14**). Therefore, we modified the bile acid scaffold at different levels:

- the side chain length;
- the presence of sulfate groups on the side chain and on rings A and B of the tetracyclic nucleus;
- the configuration at C7;
- the alkylation at C6.

#### 2.3.1 CHENODEOXYCHOLAN SULFATE DERIVATIVES

The same synthetic procedure was applied on the methyl esters of both chenodeoxycholic acid (**26**) and 6-ECDCA (**27**) to produce C24 mono and trisulfate derivatives, alkylated (compounds **31** and **33**) or not (compounds **30** and **32**) on C6. The methyl ester was reduced ( $\text{LiBH}_4$ , MeOH/THF) affording the triol (**28**, **29**) in high chemical yield (96%, scheme 2). Then, exhaustive sulfation ( $\text{Et}_3\text{N}\cdot\text{SO}_3$ ) afforded a mixture of mono and trisulfate derivatives that were purified by HPLC.

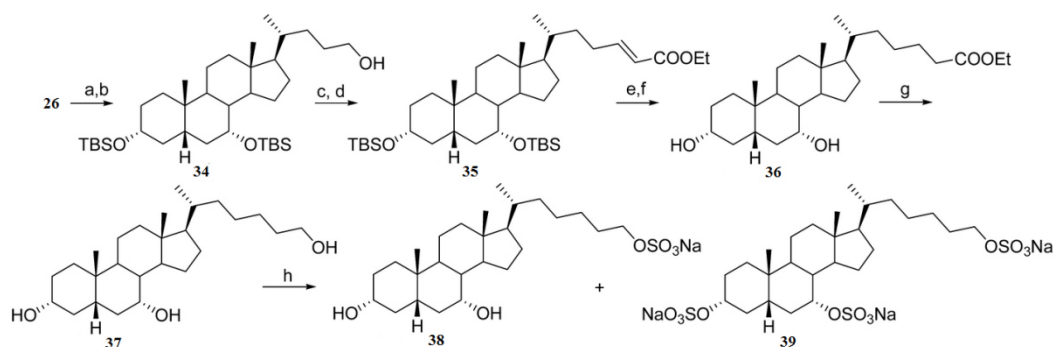


**Scheme 2.** Chenodeoxycholan Sulfate Derivatives.

Reagents and conditions: (a)  $\text{LiBH}_4$ , MeOH dry, THF, 0 °C; (b)  $\text{Et}_3\text{N}\cdot\text{SO}_3$ , DMF, 95 °C.

Starting from CDCA methyl ester **26** we also realized the elongation of the side chain from C24 to C26 thanks to a Swern oxidation<sup>73</sup>/C2 Horner<sup>74</sup> homologation. First, we protected C3 and C7 hydroxyl groups as *tert*-butylsilyl ethers and then  $\text{LiBH}_4$  reduction afforded alcohol **34** (scheme 3). At that point, the sequence of

Swern/Horner reactions gave the conjugate bis-*homo* ethyl ester **35**. This was subjected to double bond reduction and silyl deprotection to afford ethyl ester **36** (61% yield over six steps). Once again we repeated the steps of  $\text{LiBH}_4$  reduction, sulfation and HPLC separation to afford pure mono (**38**) and trisulfate (**39**) derivatives (scheme 3).



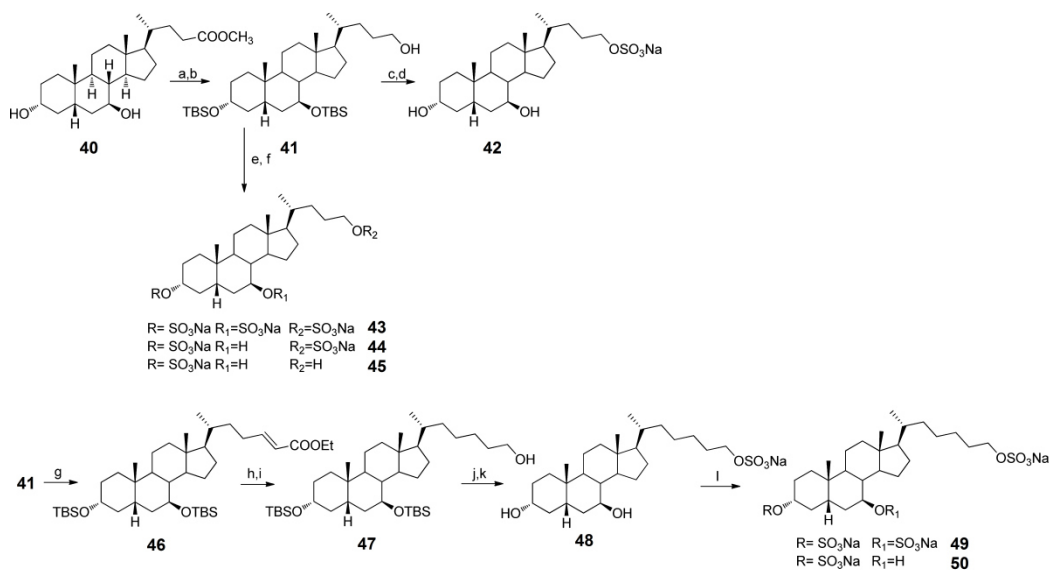
**Scheme 3.** Bis-*homochenodeoxycholan* Sulfate Derivatives.

Reagents and conditions: (a) 2,6-lutidine, tert-butyldimethylsilyl trifluoromethanesulfonate,  $\text{CH}_2\text{Cl}_2$ , 0 °C, 83%; (b)  $\text{LiBH}_4$ , MeOH dry, THF, 0 °C, quantitative yield; (c) DMSO, oxalyl chloride, TEA dry,  $\text{CH}_2\text{Cl}_2$ , -78 °C; (d)  $\text{LiOH}$ , TEPA, THF dry, reflux, 91% over two steps; (e)  $\text{H}_2$ ,  $\text{Pd}(\text{OH})_2$ , degussa type, THF/MeOH 1:1, quantitative yield; (f)  $\text{HCl}$  37%, EtOH, 81%; (g)  $\text{LiBH}_4$ , MeOH dry, THF, 77%; (h)  $\text{Et}_3\text{N}\cdot\text{SO}_3$ , DMF, 95 °C.

### 2.3.2 URSODEOXYCHOLAN SULFATE DERIVATIVES

A similar sequence of reactions was carried out on ursodeoxycholic acid (UDCA) that differs from other BAs for the  $\beta$ -orientation of the hydroxyl group at the C7 position which seems to be responsible for the lack of activity on FXR.<sup>43,44,75</sup> Hence, the modification of the 7 $\beta$ -hydroxylcholanoid scaffold would afford selective GP-BAR1 agonists.

UDCA methyl ester **40** was protected with TBS triflate and reduced at C24 to give intermediate **41** that was used to produce mono- di- and tri-sulfated compounds. The sequence of sulfation and deprotection gave ursodeoxycholan sulfate **42** (55% chemical yield in four steps, scheme 4, steps a–d), whereas removal of the silyl protecting groups followed by sulfatation gave a mixture of ursodeoxycholan derivatives (**43**, **44**, **45**) differing in the sulfation pattern and efficiently separated by HPLC (scheme 4). The C2 homologation of the side chain, already reported for chenodeoxycholan derivatives, afforded bis-*homoursodeoxycholan* derivatives (**48**, **49**, **50**).



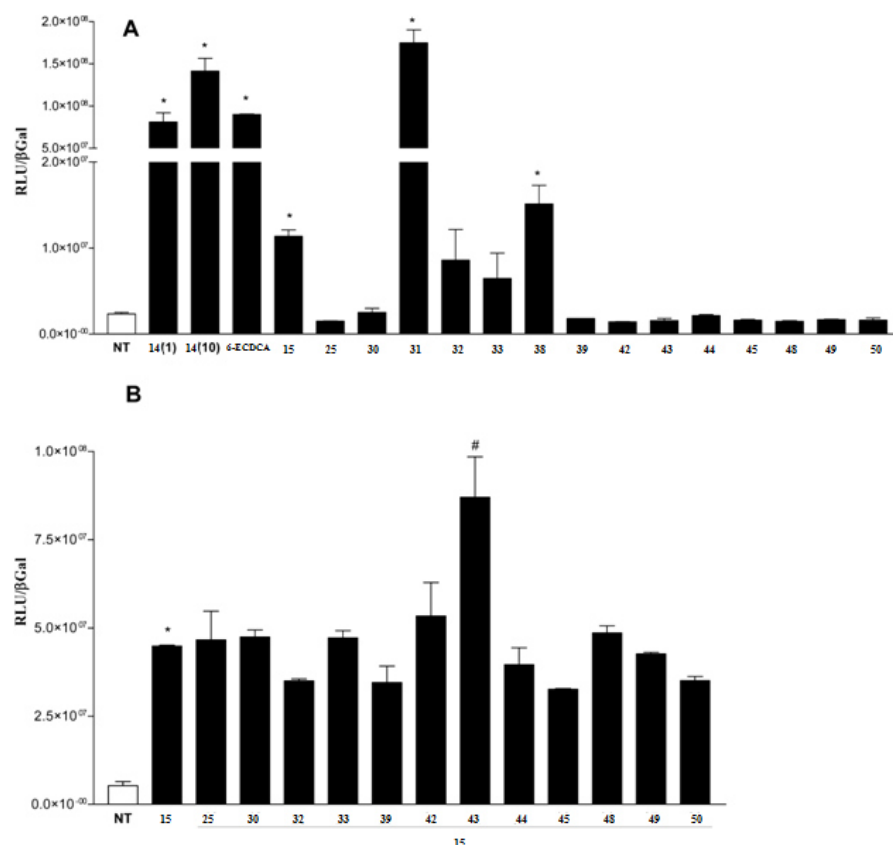
**Scheme 4.** Ursodeoxycholan Sulfate and Bis-homoursodeoxycholan Sulfate Derivatives.

Reagents and conditions: (a) 2,6-lutidine, tert-butyldimethylsilyl trifluoromethanesulfonate,  $\text{CH}_2\text{Cl}_2$ , 0 °C; (b)  $\text{LiBH}_4$ , MeOH dry, THF, 0 °C, 80% over two steps; (c)  $\text{Et}_3\text{N} \cdot \text{SO}_3$ , DMF, 95 °C; (d) HCl 37%, MeOH, 69% over two steps; (e) HCl 37%, MeOH, 96%; (f)  $\text{Et}_3\text{N} \cdot \text{SO}_3$ , DMF, 95 °C; (g) DMSO, oxalyl chloride, TEA dry,  $\text{CH}_2\text{Cl}_2$ , -78 °C then LiOH, TEPA, THF dry, reflux, 40% over two steps; (h)  $\text{H}_2$ , Pd/C, THF/MeOH 1:1, quantitative yield; (i)  $\text{LiBH}_4$ , MeOH dry, THF, 88%; (j)  $\text{Et}_3\text{N} \cdot \text{SO}_3$ , DMF, 95 °C; (k) HCl 37%, MeOH, 75% over two steps; (l)  $\text{Et}_3\text{N} \cdot \text{SO}_3$ , DMF, 95 °C.

## 2.4 PHARMACOLOGICAL EVALUATION

The synthetic derivatives described so far were tested for their activity on FXR and GP-BAR1. We performed luciferase reporter assays to evaluate transcriptional activity in HepG2 cells transfected with human FXR and in HEK-293T cells transfected with human GP-BAR1.

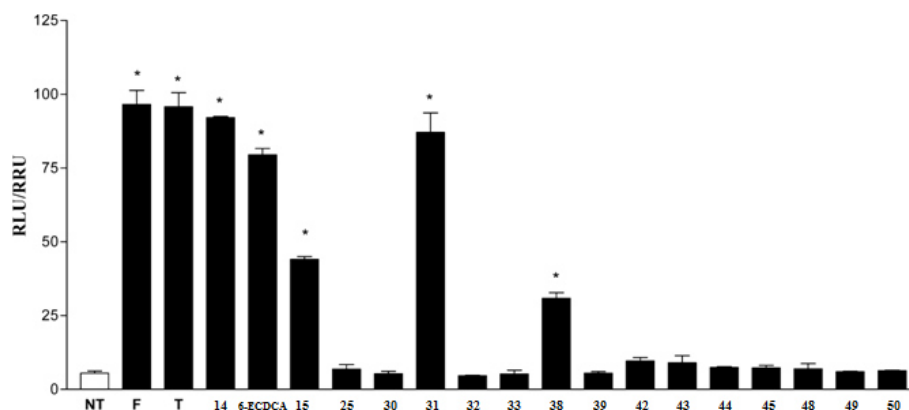
As regards the activity on FXR, we used INT-767 and 6-ECDCA, already reported as potent FXR agonists, as reference compounds. As shown in figure 11, the synthesized compounds were tested for both agonistic and antagonistic activity. We found out that compounds **31** and **38**, 6-ethylchenodeoxycholan sulfate and bis-homochenodeoxycholan sulfate respectively, are potent inducers of FXR transactivation. Specifically, compound **31** transactivates FXR with a potency comparable to INT-767 and 6-ECDCA. None of the tested compounds showed FXR antagonistic activity (figure 11B).



**Figure 11.** Transactivation assays on FXR.

(A) HepG2 cells were transfected with pSG5-FXR, pSG5-RXR, pCMV-βgal, and p(hsp27)TKLUC vectors. Cells were stimulated with 10 μM all tested compounds and with 1 μM compound **14**. CDCA (**15**) (10 μM) was used as a positive control. Results are expressed as the mean ± standard error: (\*)  $p < 0.05$  vs nontreated cells (NT). (B) HepG2 cells were transfected with pSG5-FXR, pSG5-RXR, pCMV-βgal, and p(hsp27)TKLUC vectors. Cells were stimulated with CDCA, 10 μM, in combination with 50 μM tested compounds. Results are expressed as the mean ± standard error: (\*)  $p < 0.05$  vs nontreated cells (NT); (#)  $p < 0.05$  vs CDCA.

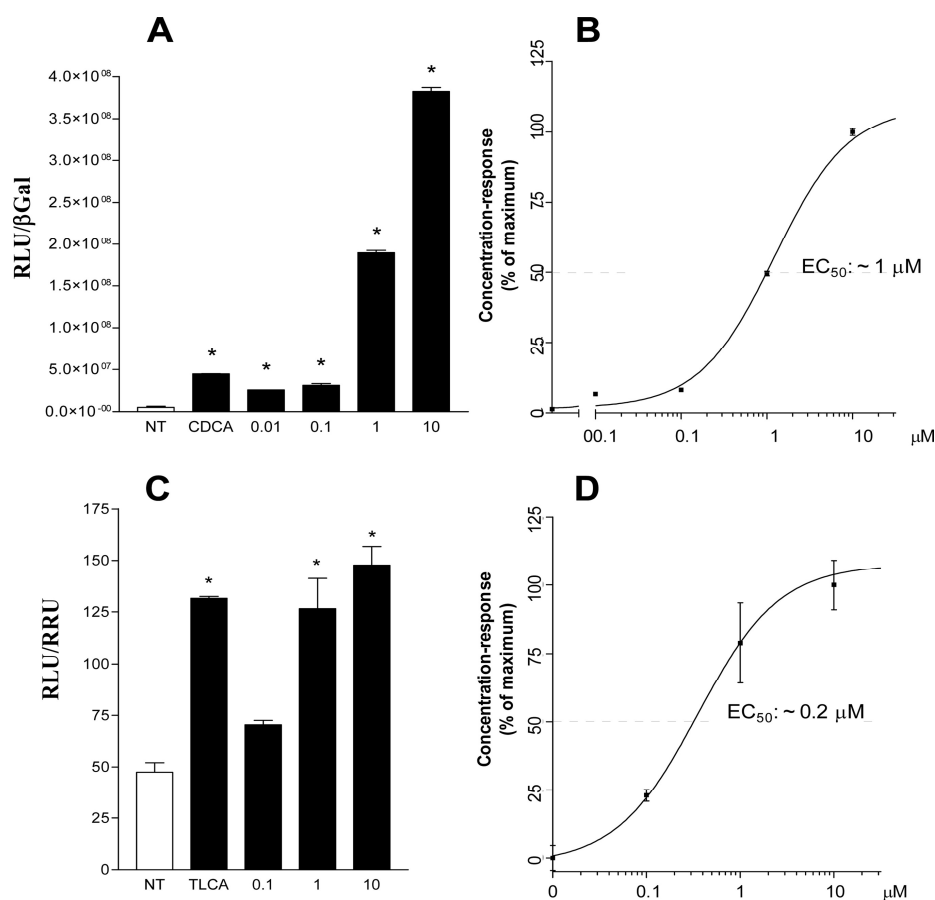
For GP-BAR1 biological assays, tauroolitholic acid (TLCA), the most potent endogenous agonist for the receptor, was used as reference while forskolin, which increases cAMP in a receptor-independent manner, was used as positive control. Results of transactivation of CREB-responsive elements in HEK-293T transiently transfected with the human GP-BAR1 revealed that compounds **31** and **38** were potent inducers of cAMP-luciferase reporter gene (figure 12) and compound **31** had a potency comparable to that of TLCA. Besides, none of the tested compounds was able to induce cAMP-luciferase reporter gene in the absence of GP-BAR1, thus indicating that their activity was GP-BAR1 mediated.



**Figure 12.** Activation of GP-BAR1.

HEK-293T cells were co-transfected with human GP-BAR1 and a reporter gene containing a cAMP responsive element in front of the luciferase gene. At 24 h after transfection, cells were stimulated with 10  $\mu$ M compounds. Luciferase activity served as a measure of the rise in intracellular cAMP following activation of GP-BAR1. TLCA (T, 10  $\mu$ M) stimulates cAMP production in a GP-BAR1 dependent manner. Forskolin (F, 10  $\mu$ M) stimulated cAMP production independently of GP-BAR1. Results are expressed as the mean  $\pm$  standard error: (\*)  $p < 0.05$  vs nontreated cells (NT).

Compounds **31** and **38** were both dual agonists however, **31** was the most potent one, thus we investigated its activity in more details. Concentration-response curves for compound **31** are reported in figure 13. Compound **31** has an  $EC_{50}$  of  $\sim 1$   $\mu$ M against FXR (panels A and B) and an  $EC_{50}$  of  $\sim 0.2$   $\mu$ M against GP-BAR1 (panels C and D). Both values are comparable to those previously reported for the reference dual agonist **14**.<sup>66</sup> Moreover, the efficacy of **31** was superior to that of CDCA (852%) and TLCA (112%) tested simultaneously, thus indicating that compound **31** is a potent GP-BAR1/FXR dual agonist.



**Figure 13.** Concentration–response curves for compound **31**.

(A, B) HepG2 cells were transfected with human FXR and stimulated with increasing concentrations of **31** (0.01, 0.1, 1, and 10 μM). (C, D) HEK-293T were transfected with human GP-BAR1 and stimulated with increasing concentration of **31** (0.1, 1, and 10 μM). CDCA (10 μM) was used as a positive control to evaluate the FXR transactivation. TLCA (10 μM) was used as a positive control to evaluate the GP-BAR1 activity. Results are expressed as the mean ± standard error: (\*)  $p < 0.05$  vs nontreated cells (NT).

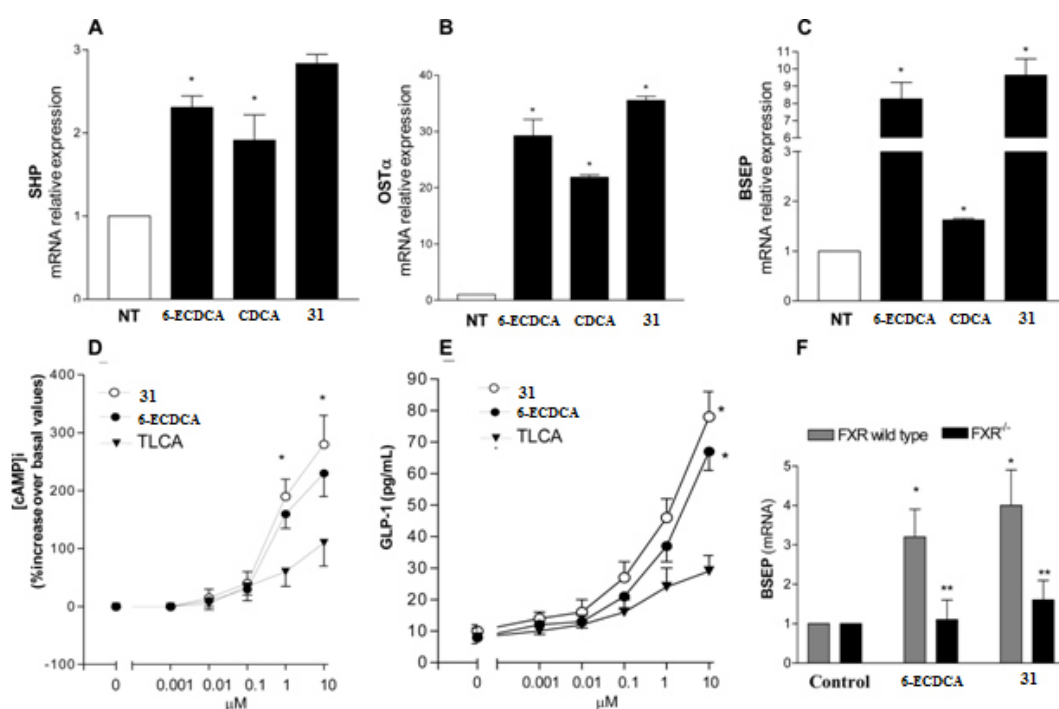
To further investigate the FXR-mediated effects of compound **31**, we realized a gene expression analysis in HepG2 cells focusing on three FXR regulated genes: small heterodimer partner (SHP), organic solute transporter (OST)  $\alpha$  and bile salt export pump (BSEP). Results shown in figure 14 (panels A–C) confirmed previous transactivation data and proved that compound **31** is a potent FXR agonist that enhances the expression of FXR target genes more efficiently than CDCA (**15**), FXR endogenous ligand, and similarly to 6-ECDCa, the most potent FXR steroid ligand so far available.<sup>76,77</sup>

We have also examined the effects of **31** on GP-BAR1 in GLUTAg cells,<sup>78</sup> that are generally considered a suitable model to evaluate the interactions between GP-

BAR1 and its ligands. They are an intestinal enteroendocrine cell line that overexpresses GB-BAR1 and reacts to receptor binding by a robust generation of cAMP followed by release of glucagon-like peptide 1 (GLP-1). As illustrated in figure 14, panel D, compound **31** is significantly more potent than TLCA and 6-ECDCA in increasing intracellular concentration of cAMP in GLUTAg cells. As expected, the three tested compounds caused a robust increase of GLP-1 concentrations in cell supernatants (panel E).

Releasing of GLP-1 from the intestine is an interesting pharmacological effect that has been proved to be beneficial for the treatment of type 1 diabetes. In fact, incretin mimetics that are able to activate GLP-1 receptor have been used in diabetes therapy for many years. However, concerns on their safety have emerged recently since they seem to be associated with an increased risk of pancreatitis and pre-cancerous findings.<sup>79,80,81,82,83</sup> The fact that **31**, a compound that is structurally different from incretins, releases GLP-1 from the intestine could result in beneficial effects without incurring in the risks associated with the long-acting GLP-1 derivatives.

Finally, the ability of **31** to modulate FXR was further investigated in FXR<sup>-/-</sup> cells. As reported in figure 14, panel F, while 6-ECDCA and compound **31** increased BSEP expression in primary hepatocytes isolated from wild type mice, this effect was lost when hepatocytes were prepared from FXR<sup>-/-</sup> mice, confirming that their activity was FXR-mediated.



**Figure 14.** Pharmacological evaluation on derivative **31**, the most potent GP-BAR1/FXR dual agonist so far identified.

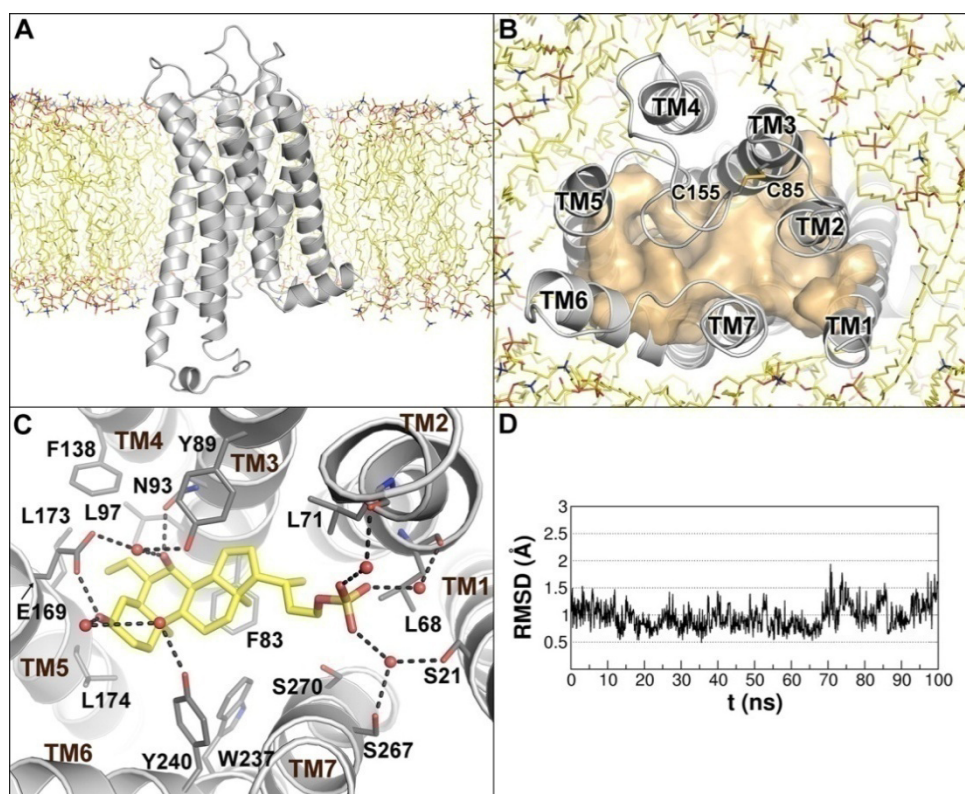
(A–C) Real-time PCR analysis of mRNA expression of FXR target genes (A) SHP, (B) OST $\alpha$ , and (C) BSEP in HepG2 cells primed with CDCA, 6-ECDCA and compound **31**. Values are normalized relative to GAPDH mRNA and are expressed relative to those of non treated cells (NT), which are arbitrarily set to 1: (\*)  $p < 0.05$  vs NT. (D, E) Effect on intracellular generation of cAMP and GLP-1 release in GLUTag cells. The data are the mean  $\pm$  SE of four to five experiments: (\*)  $p < 0.05$  vs NT. (F) Compound **31** stimulates BSEP expression in wild type hepatocytes but lose its efficacy in hepatocytes obtained from FXR<sup>-/-</sup> mice. Data are the mean  $\pm$  SE of four experiments: (\*)  $p < 0.05$  vs control and (\*\*)  $p > 0.05$  vs wild type cells.

## 2.5 DOCKING STUDIES

We investigated the binding mode of compounds **14** and **31**, the most potent dual agonists identified, in both FXR and GP-BAR1 and elucidated the structural requirements for dual GP-BAR1/FXR agonism. While the X-ray structure of FXR was already available, the tridimensional structure of GP-BAR1 has not been determined yet. Therefore, we applied the homology modeling techniques to build a tridimensional model of GP-BAR1 using the adenosine A<sub>2A</sub> receptor (PDB code 2ydo)<sup>84</sup> as template structure because of its high sequence identity (20%) and similarity (55%) values (figure 15A and figure 15B).

**Binding mode of INT-767 (14) in GP-BAR1 and FXR.** To take into account the receptor flexibility, 10 different GP-BAR1 conformations, obtained from the MD simulation on the apo form of the receptor, were considered for the docking

calculations of compound **14**. Docking solutions, obtained with the AutoDock4.2 docking program (AD4.2),<sup>85</sup> showed the ligand occupying the region underneath the extracellular loop II. However, most of the best scored binding poses didn't match with the previously reported requirements for the binding of bile acid derivatives to GP-BAR1.<sup>86</sup> The inaccuracy of docking calculations was probably due to the high flexibility of the protein that has different conformations in the apo form and in the agonist-bound state. To overcome this limitation, we had to step up the computational strategy and consider the full receptor flexibility.<sup>87,88,89</sup> Hence, we ran molecular dynamics (MD) simulations on the **14**/GP-BAR1 complex embedded in a POPC bilayer surrounded by explicit water. In this calculation, the docking pose that fulfilled most of the aforementioned structural requisites for the binding was selected as starting conformation. Then, the ligand binding pose was refined by steered MD simulations to optimize distance and orientation of the ligand-protein interactions. The obtained complex was equilibrated and underwent an over-100 ns MD simulation during which the whole system was fully flexible. The ligand binding conformation, depicted in figure 15C, was stable throughout the whole simulation, as shown by the low rmsd value reported in figure 15D, with all the ligand-protein interactions conserved and was in full agreement with the structural features required for GP-BAR1 binding.<sup>86</sup>



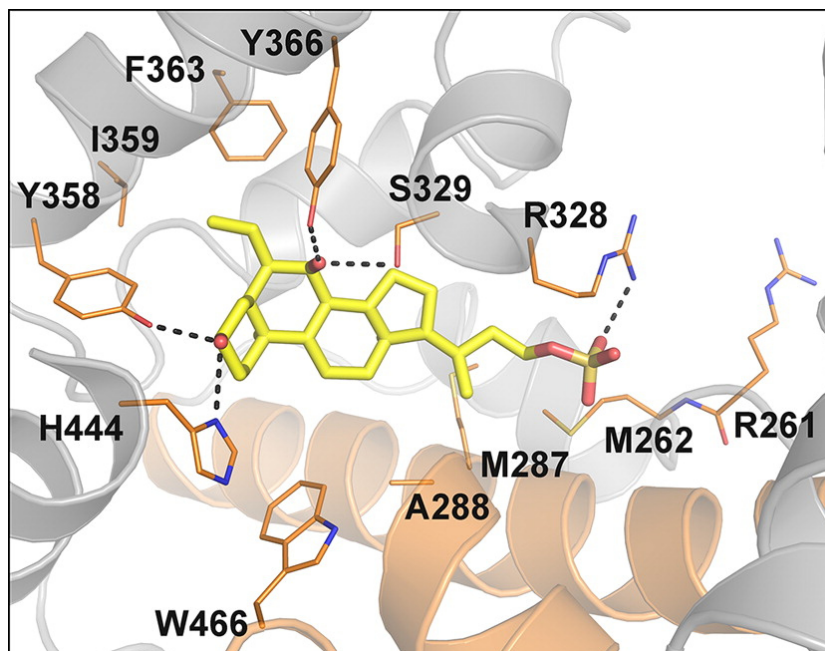
**Figure 15.** Tridimensional model of GP-BAR1 and binding mode of **14**.

Side (A) and top (B) views of the tridimensional model of the GP-BAR1 receptor embedded in the POPC bilayer. The receptor is shown as gray cartoons. Lipids are depicted as yellow sticks. The ligand binding cavity (B) is shown as transparent orange surface. Residue numbers were assigned according to the amino acid sequence of hGP-BAR1. (C) Binding mode of **14** obtained through over 100 ns long MD simulation. The ligand and interacting residues are represented as yellow and gray licorice, respectively, while the GP-BAR1 receptor is shown as gray cartoons. Water molecules are displayed as spheres, while all the hydrogens are omitted for clarity reasons. (D) The rmsd plot of the ligand heavy atoms during the production MD run. Prior to the rmsd calculations, trajectory frames were aligned using the coordinates of the C $\alpha$  carbons of the transmembrane helices.

In details, the ligand 3 $\alpha$ -hydroxyl and 7 $\alpha$ -hydroxyl groups are involved in key H-bonds with Glu169 and Asn93, and interact via a water molecule with Tyr240 and Tyr89, respectively. These interactions bring the sulfate side chain pointing toward helices I, II, and VII where the ligand establishes water-mediated interactions with the side chains of Ser21 and Ser267 and with the backbone CO groups of Leu68 and Leu71. The steroidal scaffold further stabilizes the binding through several hydrophobic interactions with Leu71, Phe83, Leu174, and Trp237. The 6-ethyl group deepens in a hydrophobic pocket, made up by Leu97, Phe138, Leu173, and Leu174, whose shape and lipophilicity suggest the importance of having short and hydrophobic chains at the C-6 position to achieve

optimal GP-BAR1 agonist activity. The proposed binding model is further substantiated by recent mutagenesis data,<sup>90</sup> reporting a decrease of the binding affinity for ligands similar to **14** in Asn93Ala, Glu169Ala, and Tyr240Ala mutant forms of GP-BAR1.

To elucidate the binding mode of **14** to FXR ligand binding domain (FXR-LBD), computations were performed on the crystal structure of the FXR-LBD from *Rattus norvegicus* (rFXR) in complex with 6-ECDCA and GRIP-1 coactivator peptide NID-3 (PDB code 1osv).<sup>91</sup> The homology between rFXR-LBD and human FXR-LBD (hFXR-LBD) is very high (95%) and all the residues of the binding pocket are conserved among the two species. Docking studies, realized again with the AutoDock4.2 software, showed compound **14** bound to the cavity formed by helices H3, H5, H7, H11, and H12 (figure 16). The sulfate group establishes key salt-bridges with the Arg328 guanidinium group, while the steroidal scaffold is responsible for several hydrophobic interactions with the side chains of Met262, Met287, Ala288, and Trp466. In addition, the 3 $\alpha$ -OH and 7 $\alpha$ -OH H-bond with Tyr358 and His444, and with the Ser329 and Tyr366, respectively. The 6-ethyl substituent is fundamental to further stabilize the binding pose thanks to its hydrophobic contacts with Tyr358, Ile359, Phe363 and Tyr366. Overall, the binding mode is very similar to that experimentally found for 6-ECDCA.<sup>91</sup> In fact, like 6-ECDCA, **14** induces a conformational change in the protein, making it capable of accommodating coactivator peptides, thus inducing the transcription of target genes. Moreover, the carboxylate group of 6-ECDCA as well as the sulfate moiety of **14** are essential for the activation of FXR through their salt-bridges with the Arg328 side chain.

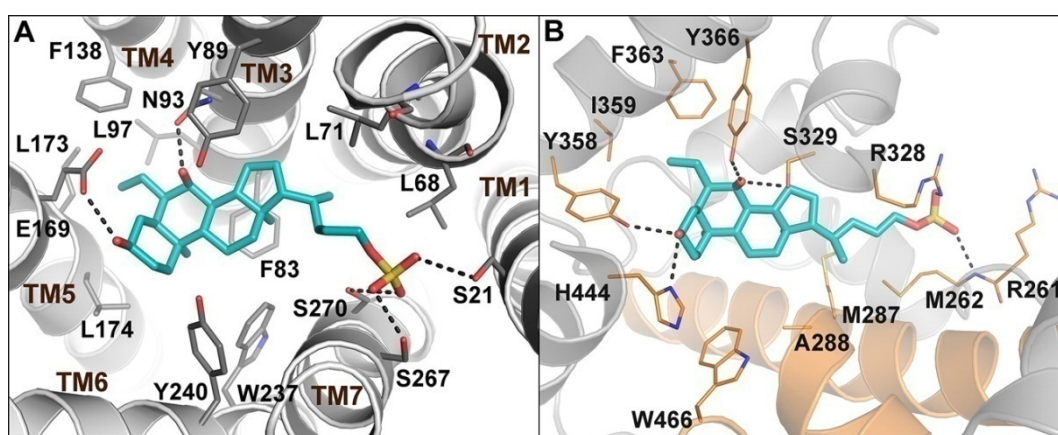


**Figure 16.** Binding mode of **14** in FXR.

The ligand is depicted as yellow sticks, respectively. FXR is shown as orange (helices H3, H4, and H12) and gray cartoons. Amino acids involved in ligand binding are shown as orange sticks. Hydrogens are omitted for clarity.

**Structural requisites for dual GP-BAR1/FXR agonism.** We performed docking calculations in the GP-BAR1 and FXR binding sites also on compound **31**, the most potent dual agonist identified in our study. The GP-BAR1 conformation obtained from the MD simulation on the **14**/GP-BAR1 complex was used. The Ramachandran analysis assured the quality of the structure showing 100% of residues in allowed conformational regions. The binding poses of compound **31** were found to be very similar to those of **14** in the two receptors (figure 17). Comparing these results, we were able to define the structural requirements for dual GP-BAR1/FXR agonism. In fact, although the two receptors are structurally very different, the active ligands show common structural features. In particular, derivative **31**, which only differs from INT-767 because of a one carbon unit longer side chain, can engage direct (i.e., not water-mediated) H-bonds with Ser21, Ser267, and Ser270 on helices TM1 and TM7 of GP-BAR1 receptor (figure 17A), thus explaining the improved agonist profile of **31** in comparison with **14**. In the FXR binding site, the sulfate group of both **31** and **14** engages a salt-bridge interaction with the side chain of Arg328 but, in compound **31** it can also interact with the Arg261 side chain further stabilizing the binding (figure

17B). Moreover, derivative **31** is able to H-bond with the backbone NH of Met262 in the loop connecting helices H1 and H2. Shorter side chains, as in the case of 6-ECDCA, are still able to activate the receptors because they can interact with at least one serine residue in GP-BAR1 and with Arg328 in FXR. Comparing docking results for different ligands, we realized that the sulfate group on the side chain is better than the carboxylate, being able to establish a larger number of interactions with both water molecules and surrounding residues. Finally, a negatively charged group is strictly required to activate FXR through salt-bridge interactions with Arg328 or/and Arg261, while it seems not necessary to activate GP-BAR1, for which a polar group able to form H-bond interactions is sufficient.



**Figure 17.** Binding mode of **31** in GP-BAR1 and in FXR.

(A) Conformation obtained from MD simulations on the **14**/GP-BAR1 complex. (B) Conformation obtained from MD simulations within rFXR-LBD (PDB code 1osv). Compound **31** is represented as cyan sticks. The receptors are shown as gray and orange (helices H3, H4, and H12 in FXR) cartoons. Amino acids involved in ligand binding are depicted as gray (in GP-BAR1) and orange (in FXR) sticks. Hydrogens are omitted for clarity.

On the other side, the configuration and the position of the ethyl and hydroxyl groups at C6 and C7, respectively, are important for the dual activity. In fact, the most potent compounds (**14** and **31**), have both these groups in the  $\alpha$  configuration. In GP-BAR1, the 6 $\alpha$ -ethyl group engages in non polar interactions that are important to stabilize the protein-ligand complex. Indeed, compounds lacking the 6 $\alpha$ -ethyl group exhibit a lower activity (e.g., compare 6-ECDCA vs CDCA and **30** vs **31**). As for GP-BAR1, the absence of the 6 $\alpha$ -ethyl group leads to less active agonists also for FXR. Actually, compound **38**, that lacks the 6-ethyl group, is less active than **14** and **31** in transactivation assays on both GP-BAR1

and FXR (figure 11A and 12). However, the 6-ethyl deficiency is partially overcome by the presence of a longer side chain that tightly interacts with the serine residues on helices TM1 and TM7 in GP-BAR1 and with the Arg261 and Arg328 side chains in FXR.

Furthermore, the introduction of additional sulfate groups on rings A and B of the steroidal scaffold leads to a loss of activity towards both receptors (compare **14** vs **25**, **38** vs **39**, and **31** vs **33** in figures 11 and 12). This could be ascribed to the excessive bulkiness and negative charge of these groups that negatively influence the binding.

Finally, as shown in figures 11 and 12, all UDCA derivatives (**42–45** and **48–50**) were inactive toward both receptors demonstrating that the configuration at C7 plays a fundamental role to properly accommodate the molecule in the binding sites of the receptors. In fact, the 7- $\beta$ OH group cannot interact with Tyr89 and Asn93 on TM3 in GP-BAR1 and with Ser329 and Tyr366 in FXR. The loss of these interactions, together with the lack of the 6 $\alpha$ -ethyl substituent on the UDCA scaffold, can explain the inability of these derivatives to activate either GP-BAR1 or FXR in transactivation assays.

## 2.6 CONCLUSIONS

In the present study, bile acid derivatives have been developed leading to the identification of compound **31**, the most potent GP-BAR1/FXR dual agonist reported so far. In particular, **31** transactivates FXR and increases the expression of FXR-regulated genes in the liver. Moreover, **31** is a potent GP-BAR1 agonist whose potency was evaluated HEK-293T cells transfected with the membrane receptor. *In vivo* and *in vitro* studies have shown that **31** increases the intracellular concentration of cAMP in a GP-BAR1-dependent manner and stimulates the release of the potent insulinotropic hormone GLP-1 in GLUTAg cells, a cell line known to respond to GP-BAR1 ligands. This effect makes of **31** a promising lead compound for the treatment of insulin-dependent conditions including diabetes. Furthermore, the binding mode of **31** in the LBD of the two receptors was elucidated through a series of computational studies. These simulations also revealed the structural requisites to achieve potent GP-BAR1/FXR dual agonism. Despite its potent GP-BAR1 activity, **31** maintains the ability to stimulate FXR

and FXR-target genes including BSEP which remains a potential drawback since FXR activation worsens liver injury in rodent model of cholestasis. In conclusion, this study is relevant for further investigations on the functional mechanism of these two receptors and for the design of novel dual GP-BAR1/FXR agonists, providing new opportunities for the treatment of enterohepatic and metabolic disorders.

## CHAPTER 3

---

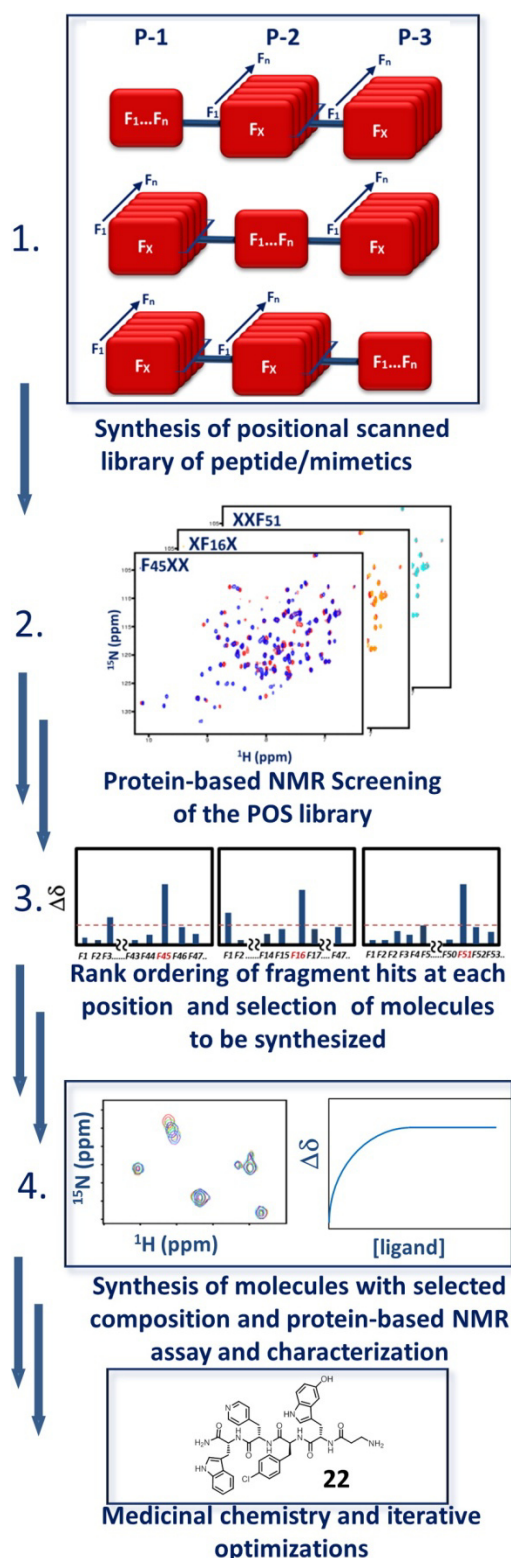
### FRAGMENT-INSPIRED HTS BY NMR OF COMBINATORIAL LIBRARIES: APPLICATION TO ANTAGONISTS OF PROTEIN-PROTEIN INTERACTIONS

High-throughput screening by NMR is an innovative target-based screening strategy, developed by prof. Pellecchia,<sup>92</sup> that improves the traditional fragment-based ligand discovery (FBLD) approach by combining the robustness of protein NMR spectroscopy as the detection method and the basic principles of combinatorial chemistry to enable the screening and deconvolution of large libraries of fragments ( $>10^5$  compounds).

FBLD, also known as fragment-based drug discovery (FBDD), has become popular as an alternative to high-throughput screening (HTS) drug discovery in both academia and big pharmaceutical companies.<sup>93,94,95,96</sup> The two approaches FBLD and HTS have always been in contrast one each other. In fact, FBLD is based on the concept that chemical space can be more efficiently probed by screening collections of small fragments rather than libraries of larger molecules that instead are usually preferred by HTS campaigns. Moreover, the FBLD approach, focusing on fragment molecules (molecular weight is usually less than 250 Da) is able to identify binders with low affinity to their target but endowed with high binding efficiency per atom. Therefore, the screening techniques employed in FBLD are much more sensitive and reliable than HTS bioassays, usually flawed by false positive and artifacts. On the other hand, HTS approaches have been widely used in industries because of the advantage of testing large libraries rapidly and, despite the young age of this technique that was first cited in 1991, many recently approved drugs, such as the anti-HIV Efavirenz<sup>97</sup> and anticancers like Gefitinib and Erlotinib,<sup>98</sup> derive from the optimization process of HTS hits.<sup>99</sup>

HTS by NMR has been able to overcome the limitations of both approaches and to combine their advantages. This new technique is based on the screening of large libraries of fragments, preassembled on a common backbone, using NMR

spectroscopy as the screening method. In order to make the process amenable to NMR time-scale, it is crucial to minimize the number of samples to be screened. Therefore, the library is assembled in mixtures that can be easily deconvoluted thanks to the positional scanning approach, meaning that in each mixture one position of the common backbone is fixed while the others are populated by all possible residues.<sup>100,101,102,103</sup> For example, a library made up by three fragments linked together on a common backbone has three positions of diversity. If we select 100 different residues that can occupy each of three positions, the library will include  $10^6$  molecules ( $100 \times 100 \times 100$ ). Conversely, arranging the library in a positional scanning fashion, the same residues can be evaluated by testing only  $100 + 100 + 100$  mixtures. Thus, we would be able to evaluate  $10^6$  compounds by testing only 300 samples, a number that can be screened by NMR in a very reasonable time. Binding information is usually obtained by measuring chemical shift variations of  $^{15}\text{N}$  and/or  $^{13}\text{C}$ -labeled proteins. The chemical shift of a given nucleus represents its local environment, which can be perturbed upon the binding of a compound in its proximity, causing the corresponding signal in the protein spectrum to change position. Since each mixture contains a dominating concentration of the fixed fragment, we assume that this has to be the main responsible for the observed chemical shift perturbations. Therefore, mixtures containing most effective fragments at a given position would produce the largest changes in the NMR spectra of the target. As a result, most potent binders are deduced from the combination of the mixtures producing the highest chemical shift perturbations at each position. Consequent synthesis and testing of individual compounds would result in the identification of the most active compounds among the possible  $10^6$  molecules. These hits are immediately amenable to subsequent hit-to-lead optimization backed up by robust NMR-based binding data, without any concern about artifacts that can be readily recognized by NMR and without the need of establishing the structure of the protein complex to guide fragment linking or growing, as usually happens in FBDD.



**Figure 18.** Schematic representation of HTS by NMR approach. The library is arranged in a combinatorial positional scanning fashion. (1) In the example reported, a three-position combinatorial library is prepared with  $n$  fragments ( $F_1, F_2, \dots, F_n$ ). This requires the synthesis of  $3n$  mixtures, each containing  $n^2$  molecules. Such mixtures are tested using protein-based NMR assays (2) so that preferential binding fragments at each position are deduced (3). Consequently, individual compounds with proper combinations of the most active fragments are synthesized and tested (4) for follow-up hit to lead optimizations. Picture taken from Barile and Pellecchia, *Chem. Rev.*, **2014**.<sup>104</sup>

The use of NMR as detection method is associated with many other advantages. First of all, because NMR is an unbiased screening technique, ligands for different areas of the protein surface could be identified simultaneously, possibly defining new protein's hot spots and/or allosteric sites. Furthermore, most of the previous positional scanning assay studies were based on either fluorescence or absorbance readouts,<sup>105,106</sup> thus requiring a specific knowledge of the protein function for assay establishment that is not necessary anymore since NMR can detect any protein regardless of its function. Moreover, the HTS by NMR approach not only identifies preferred scaffolds and initial hit compounds, but

also determines their binding affinity and, in most cases, provides initial structural information about the site of binding allowing for more direct follow-up hit-to-lead optimizations. In fact, given that resonance assignments for the protein are known, perturbation of protein NMR signals can be used to identify the residues interacting with the binder. Finally, as already mentioned, aggregators and other

reactive small molecules are responsible for false positives that affect screening results from most spectrophotometric assay platforms<sup>107, 108</sup> and even other biophysical screening techniques such as surface plasmon resonance.<sup>109</sup> However, since NMR data allow to assess integrity and folding of proteins, compounds (or mixtures in the primary screens) causing aggregation or protein denaturation are readily identified and rejected.

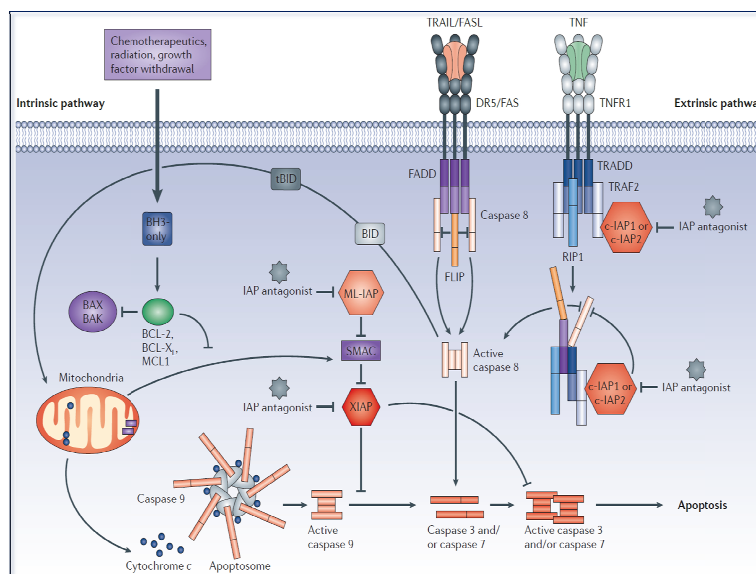
HTS by NMR has been previously applied in a *de novo* ligand discovery project against the ligand-binding domain (LBD) of the EphA4 tyrosine kinase receptor leading to the discovery of a compound endowed with good affinity ( $IC_{50}=3.4$   $\mu$ M) and high selectivity for EphA4 that represents a lead compound for the development of EphA4 antagonists.

During the months I spent at Sanford-Burnham Medical Research Institute in San Diego, tutored by Prof. M. Pellecchia and doctor E. Barile, I successfully applied an extension of the HTS by NMR approach, which has been called Fragment-Inspired HTS by NMR, to the investigation of protein-protein interactions. I focused on two different targets: the BIR3 domain of the antiapoptotic protein XIAP and EphA3, a tyrosine kinase receptor involved in cells migration and differentiation.

The fragment-inspired HTS by NMR is characterized by a positional scanned library assembled after a previous identification of a binding fragment which will act as an anchoring point. Indeed the library is built by keeping the anchoring fragment fixed at one position, while all the other positions are randomized as previously described for the more general HTS by NMR approach. In this way, a library containing hundred thousand derivatives of that single fragment can be promptly screened by NMR which is able to unambiguously identify the most effective fragments in the neighboring positions. This implies the method can provide a quick avenue for the non trivial purpose of fragment growing or linking that is actually the most time-consuming step in the traditional FBLD campaigns. For this reason, we are confident that the approach may find general and widespread utility in hit optimization and fragment maturation projects.

### 3.1 DESIGN AND OPTIMIZATION OF POTENT INHIBITORS OF XIAP, A MEMBER OF THE INHIBITOR OF APOPTOSIS PROTEINS FAMILY

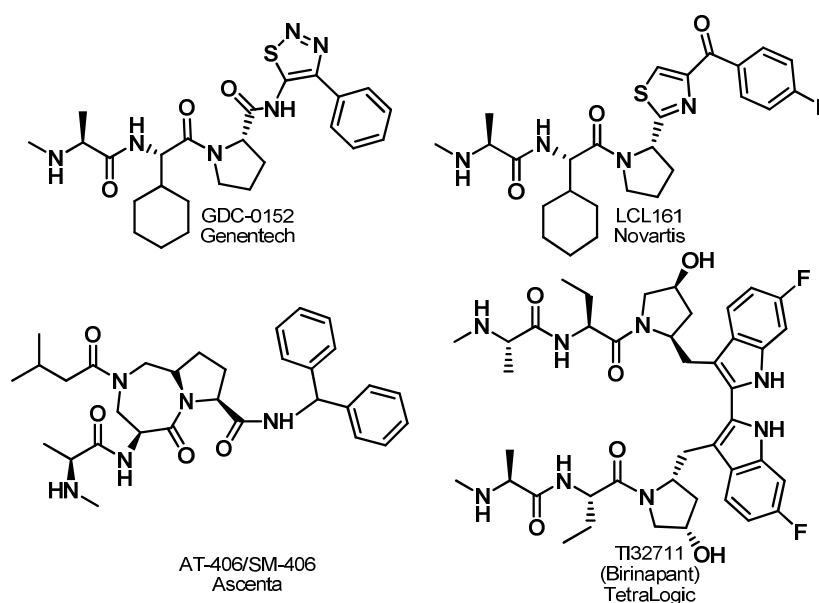
Apoptosis, a controlled programmed cell death, is a crucial event in cell cycle that is responsible for the removal of undesired or damaged cells so that defects in this process play important roles in cancer pathogenesis and progression.<sup>110</sup> Caspases, a family of intracellular proteases, are the principal mediators of apoptosis. They are activated by several cell death stimuli through both intrinsic and extrinsic pathways and trigger cascade biochemical reactions responsible for cellular death.<sup>111,112</sup> Certain caspases are inhibited by IAPs, a family of anti-apoptotic proteins<sup>113</sup> whose over-expression has been documented in several types of human cancers.<sup>114,115</sup> Moreover, previous studies suggested an important role for IAPs in maintaining tumor survival and promoting resistance to apoptosis induction by anticancer drugs.<sup>116,117,118,119</sup> The X-linked inhibitor of apoptosis protein (XIAP) is a member of the IAP family known for its ability to directly bind and suppress caspase-9, an initiator caspase that cleaves a range of substrates in a cascade, leading to cellular death.<sup>120</sup> The interaction XIAP/caspase-9 is mediated by the third baculoviral IAP repeat (BIR3) domain of XIAP that sequesters caspase-9 in a monomeric thus inactive state.<sup>121</sup> In turn, XIAP can be inhibited by the protein SMAC (also known as DIABLO) through an interaction mediated by the same BIR3 domain.



**Figure 39.** Schematic representation of cell death pathways. Picture taken from Fulda *et al.*, *Nature Reviews*, 2012.<sup>122</sup>

Three-dimensional structures of the BIR3 domain in complex with peptides and full-length binding partners (caspase-9 and SMAC) have been derived by both X-ray and NMR<sup>123,124</sup> showing that the binding to BIR3 is dependent solely on the tetrapeptidic sequence AVPI or AVPF and its N-terminus.<sup>120,125</sup> Many studies have also reported the absolute requirement for Ala at the N-terminus and for a Pro as third residue, which together are sufficient for binding to BIR3.

The identification of a *consensus motif* resulted in the generation of few small molecule inhibitors of XIAP that entered the clinical trials as anticancer agents (figure 20).<sup>126,127,128</sup> The first inhibitor came from the Wang laboratory at the University of Michigan who identified SM-406, taken into phase I clinical trials in partnership with Ascenta. Soon after, Genentech identified GDC-0152, the first SMAC mimetic that entered clinical trials in June 2007, while Novartis identified LCL161, a potent orally available SMAC mimetic. Maybe inspired by the fact that SMAC itself is a dimer in solution and it can bind both the BIR2 and BIR3 domains of XIAP simultaneously, a bivalent antagonist called Birinapant, able to engage the two BIR domains at the same time, came out by Tetralogic. Looking at the chemical structures of the above mentioned inhibitors, is readily noticed how similar they are to each other, differing very slightly from the original *consensus motif*.



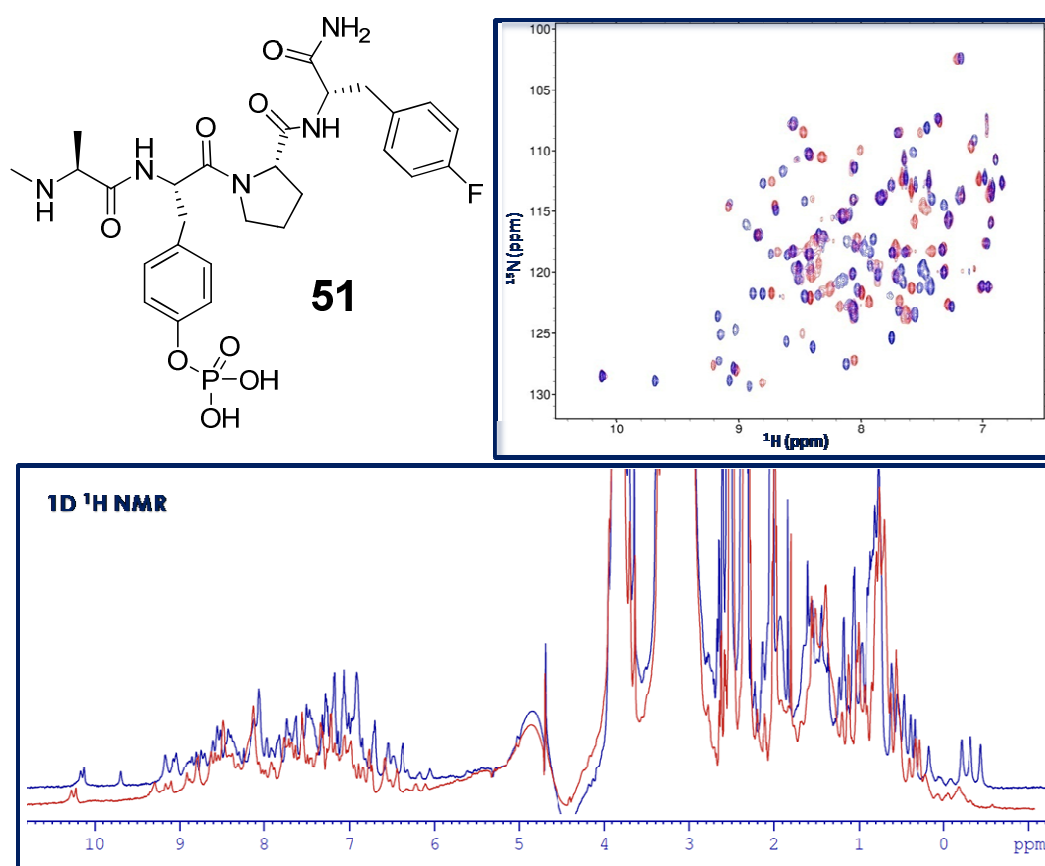
**Figure 40.** Selection of XIAP antagonists under clinical investigation.<sup>122</sup>

Because the interaction of XIAP-BIR3 with its binding partners has been so well characterized, we chose this protein as target for a proof of concept application of the Fragment-Inspired HTS by NMR. Our results not only demonstrated the feasibility of the approach but also led us to the discovery of a new class of potent XIAP inhibitors, which significantly differ from all the previously reported antagonists.

### 3.1.1 LIBRARY, SCREENING AND RESULTS

Knowing the AVPF *consensus motif* to bind XIAP-BIR3 and giving the absolute requirement for an alanine at the N-terminus, we used a tetra-peptoid library with the N-terminal alanine as anchoring point. The following positions were occupied by 46 residues, selected among natural and non natural amino acids, assembled on a common backbone in a positional scanning fashion and yielding compounds with an average molecular weight of about 450 Da. The choice of using non natural amino acids was crucial to sample a wider chemical space thus affording innovative results. The library was made up by 138 mixtures (46+46+46), each containing 2,116 compounds (1x46x46). Herein, the construction of this peculiar library allowed the NMR screening of almost 100,000 compounds (46x46x46=97,336) in just a few days. Usually, the screening of the mixtures can be realized with one-dimensional (1D)  $^1\text{H}$ -NMR spectra monitoring a region below than 1 ppm, which contains the  $^1\text{H}$  aliphatic resonances from methyl groups of protein residues and that is rarely populated by aspecific signals from organic small molecules.<sup>129,104</sup> However, because of the high expression yield in *E. coli* cells, we decided to perform also 2D [ $^1\text{H}$ ,  $^{15}\text{N}$ ]-HSQC experiments that required a greater amount of protein but allowed a better evaluation of the mixtures effectiveness concurrently giving insights on the changes occurring on the protein binding surface. Samples were prepared with 25 $\mu\text{M}$  protein in presence of each of the 138 mixtures at an overall concentration of 500 $\mu\text{M}$  which, assuming a hit rate as low as 1-2%, afforded a minimum protein/ligand ratio of 5:1. Binding mixtures were promptly identified since they caused a dramatic change of the protein spectrum, with most of the peaks experimenting large shifts and new peaks appearing. To rank order all the fragments in each relative position, we took into account the chemical shift variation ( $\Delta\delta$ ) of selected sensitive peaks and the

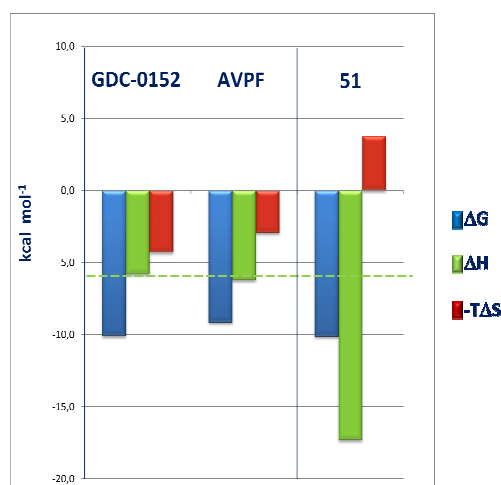
numbers of new peaks occurring in each spectrum. Among the most active residues we were able to identify both the *consensus motif* AVPF and the sequence of GDC-0152 (Ala-cyclohexylGly-Pro-Phe). Moreover, by rank-ordering the hits, we derived and confirmed some previously reported SARs.<sup>130,131</sup> For example Ile is known to be more effective than Val in P2, Pro in P3 cannot be replaced and a number of aromatic residues are well tolerated in P4. In addition, we found that D-amino acids are not allowed in P2 while they are tolerated in P3. However, the most interesting outcome of this screening was that for the second position, the most effective residue in perturbing the BIR3 spectrum was not Val or cyclohexylglycine, but a new unexpected residue of phosphotyrosine, never investigated before in this context. All the considerations above, led us to the synthesis of a single peptide representing the most effective combination of the screened residues, which was a tetra-peptide having sequence Ala-phosphoTyr-Pro-4-fluoro-Phe (compound **51**, figure 21).



**Figure 21.** 1D  $^1\text{H}$  and 2D [ $^1\text{H}$ ,  $^{15}\text{N}$ ]-HSQC spectra of XIAP-BIR3 domain in absence (red) and presence (blue) of compound **51** (1:1 ratio).

The dissociation constant ( $K_d$ ) value of compound **51** binding to the XIAP-BIR3 was calculated via isothermal titration calorimetry (ITC) along with the natural peptide AVPF and the reference compound GDC-0152. AVPF showed a  $K_d$  of 174 nM while compound **51** demonstrated exactly the same dissociation constant of GDC-0152 that is 33 nM. Besides confirming the potency of **51**, ITC experiments also provided insights into the nature of the forces responsible for the complex formation. We found out that, while having the same Gibbs energy of binding ( $\Delta G$ ), the two peptides show completely different enthalpy and entropy contributions. In fact, while GDC-0152 exhibits an entropy-driven interaction resulting mostly from hydrophobic interactions ( $\Delta H=0.17$  Kcal/mol), compound **51** demonstrates a much more favorable enthalpy contribution ( $\Delta H=0.32$  Kcal/mol), probably due to specific interactions of the phosphate group with the receptor, that is completely compensated by an entropic loss resulting in the two inhibitors to have similar binding affinities (figure 22). Compensating entropy changes usually arise from conformational entropy losses associated with the molecular structuring induced by hydrogen bonds formation and/or by a reduced desolvation and are typical of binders with nanomolar affinity.

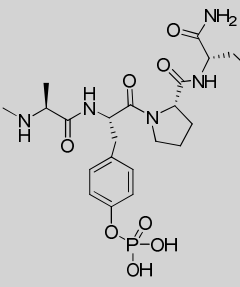
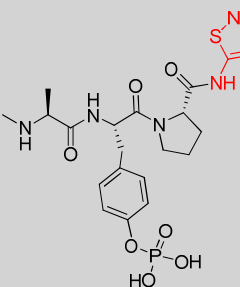
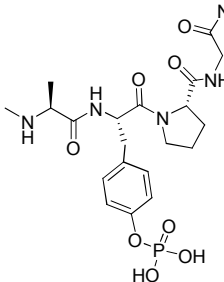
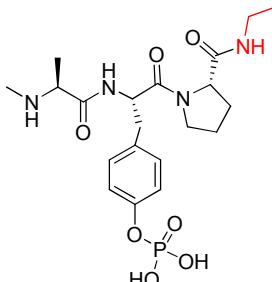
Summarizing, the synthesis of just one peptide, resulting from the combination of the most potent fragments identified by HTS by NMR, afforded a compound with affinity to the target protein in the nanomolar range and comparable to that of the clinical candidate GDC-0152.



**Figure 22.** Thermodynamic profiles of selected BIR3 binders: GDC-0152, AVPF and compound **51**.

## 3.1.2 LEAD OPTIMIZATION

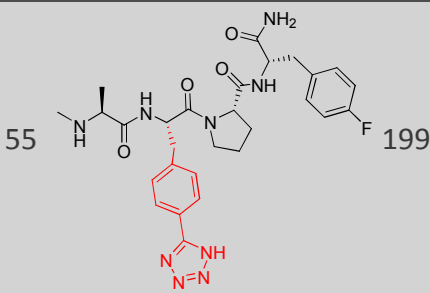
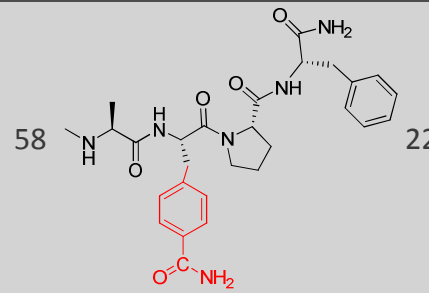
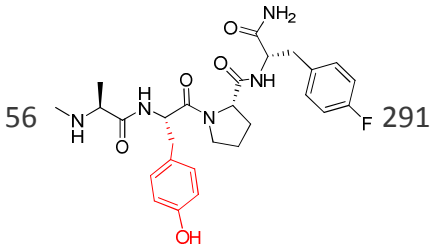
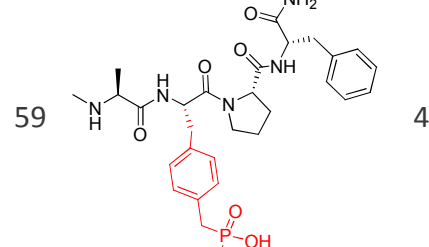
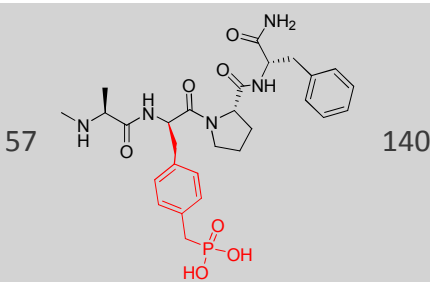
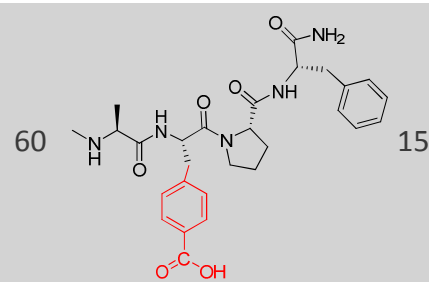
Since compound **51** showed a very promising activity, it was subsequently used as a starting point for iterative optimizations. At this lead optimization stage we used ITC to validate the compounds and measure their affinity to the target. First, we investigated the importance of the C-terminal residue by synthesizing three derivatives (**52-54**) reported in table 2. ITC measurements showed that the Phe residue has a slightly better activity ( $K_d=24$  nM) than the 4-fluoro-Phe present in our library so we proceeded to further optimize compound **52**.

Cmpd n°	Structure	$K_d$ (nM)	Cmpd n°	Structure	$K_d$ (nM)
51		33	53		76
52		24	54		80

**Table 2.** Analogs derived from optimization of P4 position and their relative dissociation constants.

Our main issue was that the phosphate group, which is responsible for the great enthalpic contribution to the binding, is usually associated to undesirable pharmacokinetic properties, in particular it is subjected to a fast hydrolysis in blood and, due to the negative charge, cannot permeate cell membranes. In order to improve the drug-likeness of our compounds, we tried to replace the phosphate group with a bioisoster that could retain compound activity while extending its half-life. Six analogs of compound **52**, namely compounds **55-60**, were synthesized (table 3) but most of them showed a drastic reduction in the activity except for compound **59** bearing the isosteric and isoelectronic phosphonate

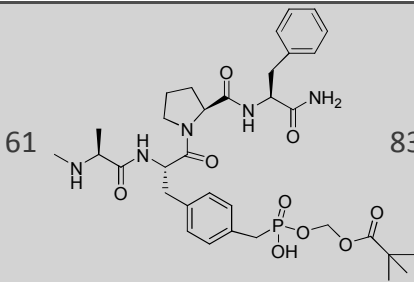
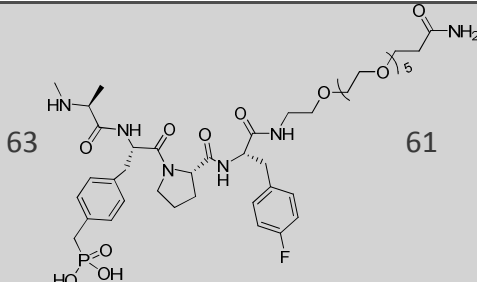
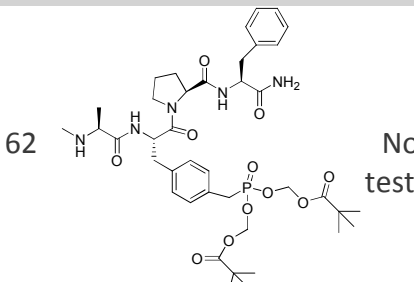
group. Compound **59** was equally active ( $K_d=49$  nM) compared to the parent compound and also retained the strong enthalpic contribution to the binding ( $\Delta H=0.29$  Kcal/mol). Moreover, due to its chemical nature, compound **59** is chemically and enzymatically more stable than the corresponding phosphate. Indeed, the  $\text{CH}_2\text{-P}$  bond, different from the  $\text{O-P}$  linkage, is not susceptible to phosphodiesterase and phosphatase mediated hydrolysis which often causes a short duration of action.

Cmpd n°	Structure	$K_d$ (nM)	Cmpd n°	Structure	$K_d$ (nM)
55		199	58		227
56		291	59		48
57		140	60		1510

**Table 3.** Compound **52** analogs varying by different residues in P2 position. The corresponding  $K_d$  derived by ITC measurements are also reported.

Next, we focused on the improvement of oral absorption and membrane permeation. In fact, it is well known that drugs containing a phosphate, phosphonate or phosphinate functional group carry one or two negative charges at physiological pH values making them very polar thus unable to undergo passive

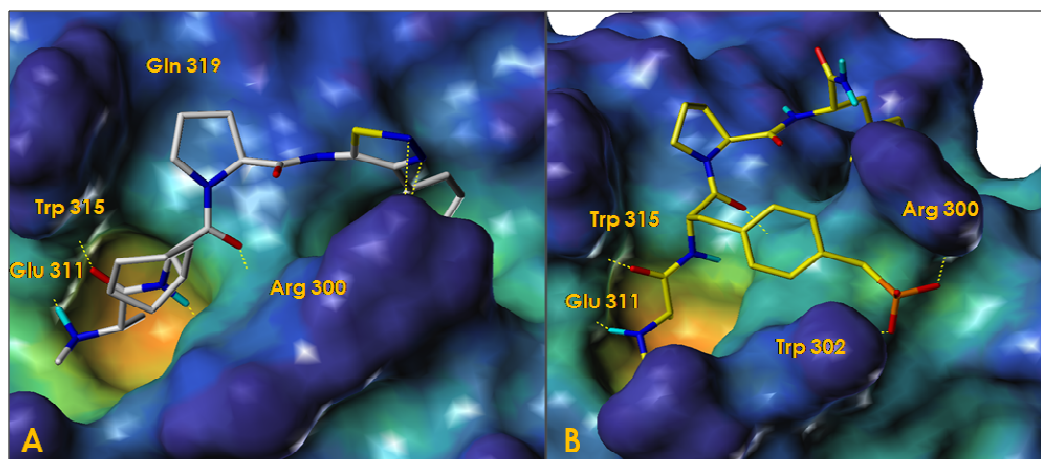
diffusion across cellular membranes. Moreover, because of their high polarity, these agents often exhibit a low volume of distribution and, therefore, are subjected to efficient renal clearance as well as possibly biliary excretion. Many approaches have been investigated to enhance cellular bioavailability of organic phosphates and the majority of them involves masking the phosphoryl moiety with prodrug protecting groups that can be removed enzymatically within the cell.<sup>132</sup> A simple ethyl derivatization is usually not advisable since the cleavage of the protecting groups would be very slow due to the absence of specific enzymes. However, the hydrolysis becomes faster when the alkyl size increases and when a spacer group, like a methylene, is introduced because enzymes such as carboxyl-esterases get involved. The pivaloyloxymethyl (POM) moiety is an esterase-cleavable group that has found wide utility in protecting phosphate and phosphonic acid functionality in peptide mimetics.<sup>133</sup> The first acyloxyalkyl group is cleaved by an esterase-like enzyme to release one phosphonic acid group. The phosphonate monoester formed, carrying a negative charge, is a poor substrate for a common esterase however the second prodrug moiety can fall off rapidly through a simple chemical hydrolysis. In this way only one enzyme is needed to trigger release of both moieties and the efficiency of delivering the phosphonate diacid is high.<sup>134</sup> After these considerations, we synthesized two prodrugs: the phosphonate monoester (**61**) and the diester (**62**) with POMs protecting groups. The monoester, which is still negatively charged, cannot permeate the lipidic bilayer but retains activity on XIAP ( $K_d=83$  nM) while the diester is inactive *in vitro*. In another approach we linked a polyethylene glycol residue to the C-terminus of the peptide (compound **63**, table 4) in order to increase the global lipophilicity of the molecule and make the partition in lipids easier. Compounds **61-63** will be shortly tested in cells to evaluate their ability to permeate membrane cells and to investigate the ability of the POM prodrugs to deliver the active molecule inside the cells.

Cmpd n°	Structure	K <sub>d</sub> (nM)	Cmpd n°	Structure	K <sub>d</sub> (nM)
61		83	63		61
62		Not tested			

**Table 4.** Prodrugs of compound **59** and their relative dissociation constants.

### 3.1.3 COMPUTATIONAL STUDIES

To disclose the molecular basis for the highly specific binding of compound **59** to the BIR3 domain of XIAP we performed docking studies on both **59** and GDC-0152 in order to outline the differences between the binding poses of the two peptides. For these calculations the X-ray resolved structure of the cIAP1/XIAP chimeric BIR3 domain (PDB code 3UW4)<sup>134</sup> was used. Docking computations revealed that the terminal ammino group of the N-methyl-Ala, that is common to the two peptides, salt-bridges with Glu311, while the CO of the same residue H-bonds with Trp315. These are the only interactions that are shared by both peptides while the rest of the molecule arranges in very different poses. As regard GDC-0152, the binding to the protein is mediated by few H-bonds with the backbone of the molecule and some hydrophobic interactions with the Pro. Instead, the binding of compound **59** is supported by two H-bonds between the phosphate group of the peptide and Arg300 and Trp302 of the protein, interactions that are totally lacking in the binding pose of GDC-0152. It is likely that the binding to the phosphate group forces the protein into a conformational rearrangement that accounts for the previously observed entropy loss.



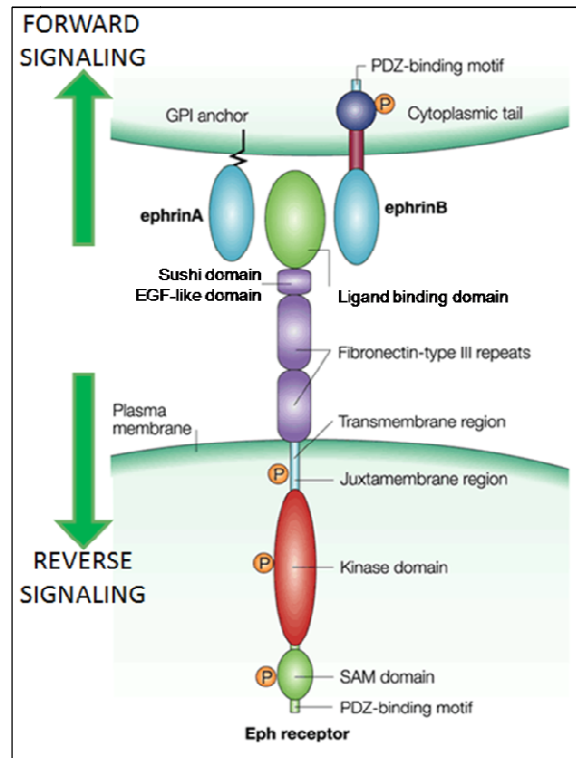
**Figure 23.** Solvent-accessible surface representation of the peptide-binding site of the 1.79-Å resolution crystal structure of the cIAP1/XIAP chimeric BIR3 domain (cXBIR3CS) (PDB ID 3UW4)<sup>135</sup> in complex with GDC-0152 (panel A) and compound **59** (panel B). The protein surfaces generated with MOLCAD<sup>136</sup> are color-coded according to the cavity depth (blue, shallow; yellow, deep). Protein residues involved in hydrogen bonds are highlighted.

#### 3.1.4 CONCLUSIONS

To our knowledge, we reported on a new strategy for fragment growing, Fragment-Inspired HTS by NMR. We demonstrated the feasibility of this method by applying it to a test case that also led to the discovery of a new class of XIAP antagonists whose most active member is compound **59**, which exhibits significant binding affinity and a favorable thermodynamic profile for the targeted BIR3 domain of XIAP. Three prodrugs are now being tested to evaluate their ability of delivering the active molecule inside the cells. Our data clearly demonstrate the effectiveness and usefulness of our strategy for the rapid growing of a previously identified fragment with low affinity for a target protein. Based on the data reported, we believe the approach may find its utility especially in hit-to-lead optimization and fragment maturation projects.

### 3.2 DISCOVERY AND OPTIMIZATION OF HIT COMPOUNDS BINDING EPHA3, A TYROSINE KINASE RECEPTOR INVOLVED IN ANGIOGENESIS AND METASTASIS FORMATION.

The Eph receptor family is the largest among tyrosine kinases and is unique in that both the ligands and receptors are membrane bound providing the possibility for bidirectional cell-cell signaling. Fourteen Eph receptors have been identified in vertebrates. They are divided into two subclasses which differ primarily in the structure of their ligand binding domains. EphA receptors usually bind GPI-linked ephrin-A peptides while EphB receptors show greater affinity for transmembrane ephrin-B ligands. However, interactions are promiscuous within each class and some Eph receptors can also bind to ephrins of the other class.<sup>137</sup> Upon binding with the ephrins, receptor-ligand complexes assemble into clusters which are responsible for signals that affect both of the contacting cells. Depending on the cellular context, including the identity of the Eph-ephrin receptor-ligand pairs and their relative density on interacting cells,<sup>138</sup> the binding results in either repulsion or adhesion between the two cells. Cellular repulsion, that is the most common response, requires disruption of the receptor-ligand complexes that is brought about either by enzymatic cleavage of the tethered ephrin ligands or by endocytosis of Eph-ephrin complexes. These mechanisms allow the regulation of many different phenomena such as organ development, tissue remodeling, neuronal signaling, insulin secretion and bone metabolism.<sup>139,140</sup> Moreover, the involvement of Eph/ephrin-signaling in tumorigenesis has been extensively investigated due to frequent upregulation of Eph receptors in numerous cancer types.<sup>141,142,143</sup>



**Figure 24.** Schematic representation of Eph receptors and ephrins. The diagram shows two adjacent cells, one bringing an ephrin (top) and one expressing an Eph receptor (bottom). Ephrins can bind to two different sites within the ligand-binding domain as well as a site on the adjacent Sushi domain. Together the Sushi and EGF-like domains comprise a cysteine-rich linker region that is important in mediating cluster formation with other Ephs. Two fibronectin type III repeats complete the extracellular domain. The intracellular domain consists of a juxtamembrane region and a kinase domain which are required for receptor phosphorylation and activation, a sterile  $\alpha$ -motif (SAM) domain responsible for the interaction with other receptors and a PDZ motif which may serve as docking site for signaling molecules. Picture adapted from Kullander and Klein, *Nat Rev Mol Cell Biol*, **2002**.<sup>144</sup>

Among the Eph family, EphA3 is presently one of the most promising targets for cancer therapy due to the fact that it is expressed in a variety of solid and hematopoietic tumor cells,<sup>145,146,147,148</sup> while there is very little evidence for physiologic adult expression or function. Moreover, this receptor has shown multiple tumor-promoting roles including stem cell maintenance,<sup>145, 149</sup> enhancement of angiogenesis and metastasis formation.<sup>150</sup> The high affinity ligands for EphA3 are ephrin-A2 and ephrin-A5, but it also binds ephrin-A3 and ephrin-A4 with lower affinity. While important details of EphA3 signaling have been determined, more complete understanding of the activity of this receptor will require complete knowledge of its interacting proteins. To date, the majority of investigations into functions of Eph receptors has been accomplished using

genetic and biochemical methods while pharmacological approaches, crucial to understand acute Eph inhibition effects, have been hampered by the lack of potent and selective inhibitors.<sup>151</sup> To date, KB004,<sup>152</sup> a Humaneered® antibody, is the only agent known to bind EphA3 and is actually in phase 2 of clinical trials for the treatment of hematologic malignancies.

My research project had the ambitious purpose of finding selective small molecule inhibitors of EphA3 receptor. We expressed for the first time the ligand binding domain (LBD) of EphA3 in a <sup>13</sup>C-Met-labeled form allowing NMR studying of the protein in presence of best ligands. Next, starting from compounds with high affinity for EphA4,<sup>92</sup> a receptor sharing high sequence identity with EphA3, we developed new EphA3 inhibitors endowed with good selectivity and affinity for the target. These compounds pave the way to the creation of useful pharmacological tools to deepen our knowledge of EphA3 pathological role in cancer development.

### 3.2.1 EPHA3-LBD EXPRESSION

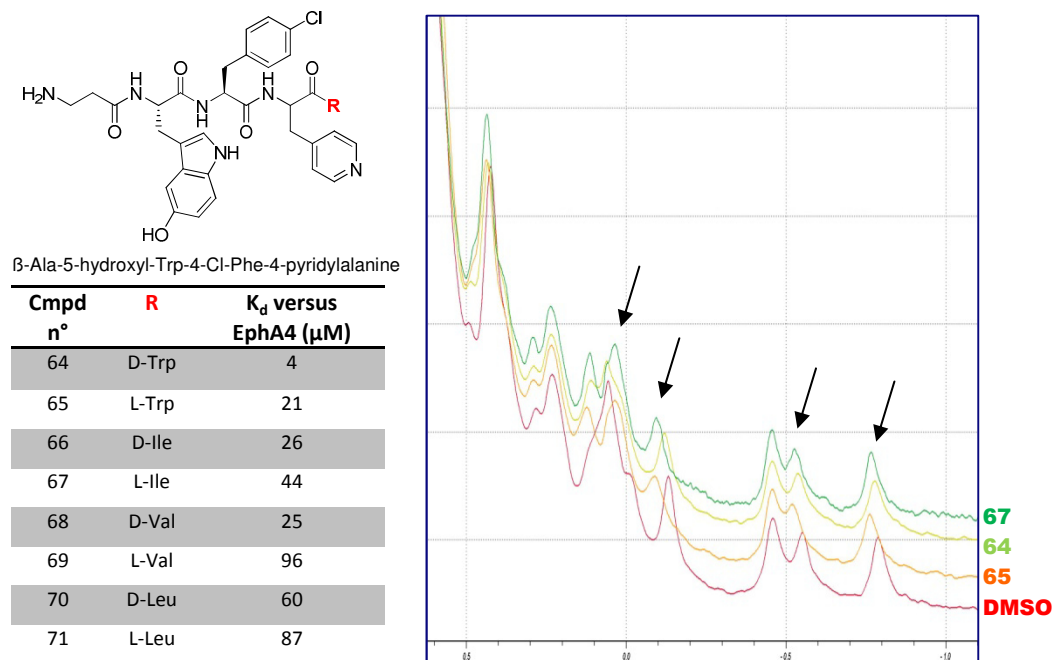
My work started with the expression of the LBD of EphA3 which has never been reported before. First attempts using a gene encoding for the LBD sequence (28-207) cloned downstream to a His-tag and following the standard procedures for expression in *E. coli*, were unsuccessful since the protein was insoluble and accumulated in inclusion bodies. Formation of inclusion bodies could result from the failure to control misfolded or unfolded proteins depending on their propensity to trigger a cellular stress response, when recombinant protein is expressed at high rates, or when the protein is expressed under unfavorable protein-folding conditions.<sup>153,154,155</sup> In order to increase the solubility and to aid the proper folding of the protein, we explored different strategies. First, we lowered the bacterial growth temperature after induction from 37°C to 18°C. This decreases the rate of protein synthesis and usually more soluble protein is obtained. Then, since the protein contains five Cys residues, we employed Rosetta-gami B (DE3) and SHuffle T7 Express Competent *E. coli* cells, two strains created to enhance the formation of target protein disulfide bonds in the bacterial cytoplasm. However, since the protein was still insoluble we thought to mutate the gene. One strategy consisted in replacing the Cys202 with an Ala residue and then in mutating some

hydrophobic residues with highly soluble aminoacids (Arg, Glu and Lys) in key positions. Finally, we tried the addition of fusion partners.<sup>156,157</sup> Commercially available fusion tags include thioredoxin (Trx)<sup>158</sup> and glutathione-S-transferase (GST)<sup>159</sup> that could enhance solubility since the fusion polypeptide is highly soluble. Moreover, Trx has an intrinsic oxido-reductase activity that catalyzes disulfide bond formation (Trx), while GST can protect its target protein from the proteolytic degradation, stabilizing it into the soluble fraction.<sup>160</sup> The combination of conditions that gave the greatest amount of soluble protein was the gene sequence GST-His<sub>10</sub>-TEV-LBD<sub>(29-201)</sub> trasformed in SHuffle T7 Express Competent cells. The fusion protein was purified by Ni<sup>+</sup>-affinity chromatography thanks to the His-tag included in the sequence. Then, an enzymatic cleavage with a double mutant TEV protease<sup>161</sup> followed by further purification afforded the pure EphA3-LBD. More details on the expression and purification protocols are reported in the experimental section. More difficulties arose when we tried to express the <sup>15</sup>N-labeled protein required for NMR studies. Since SHuffle T7 Express *E.coli* cells grew very slowly in the minimum broth (M9) containing <sup>15</sup>NH<sub>4</sub>Cl and produced insufficient amounts of recombinant protein. To overcome this limitation, we chose to express the protein selectively labeled with <sup>13</sup>C on the methyl of the methionines. This was easily attained by supplementing the enriched bacterial growth medium with an excess of labeled Met just prior to the induction of protein expression.<sup>162,163</sup>

### 3.2.2 BINDERS DESIGN AND OPTIMIZATION

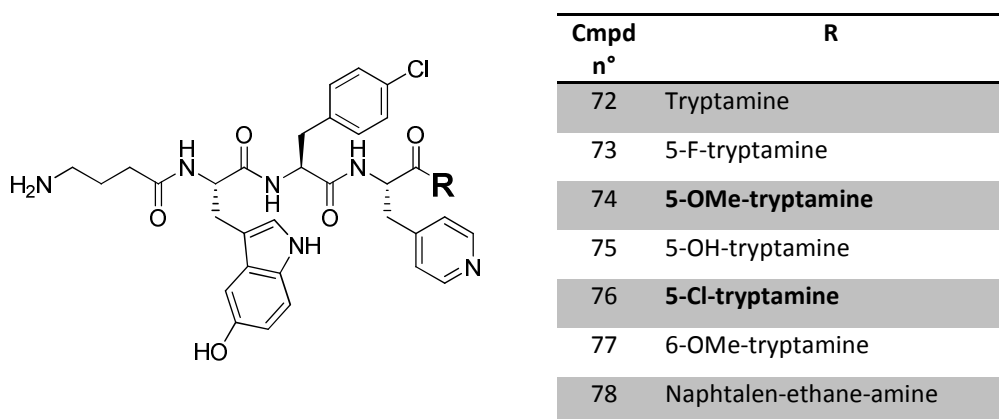
From a previous HTS by NMR of a tetrapeptidic library, EphA4 inhibitors<sup>92</sup> with affinity in the low micromolar range for the target had been identified. Considering the high homology of sequence between the ligand binding domains of EphA4 and EphA3 (73% identity), we examined these compounds (**64-71**, figure 25) in order to select the one with highest affinity and selectivity towards EphA3 to use as starting compound for EphA3 binding optimization process. Initial tests were realized through 1D <sup>1</sup>H NMR experiments of the EphA3-LBD in absence and in presence of the peptides in a 1:10 molar ratio by monitoring the aliphatic region of the spectra below 0.7 ppm since no spectral overlapping between ligand and protein resonances occurred in this region.<sup>129</sup> Largest

variations in chemical shifts were observed for compound **65** that has a L-Trp as C-terminal residue, while the most active compound against EphA4 was **64** with a C-terminal D-Trp.



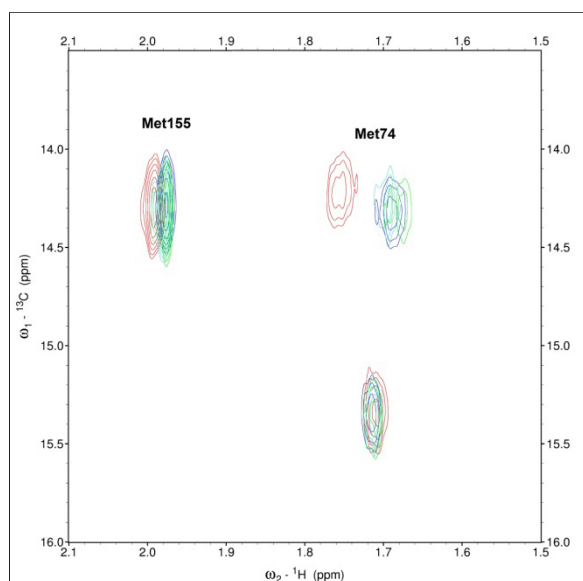
**Figure 25.** Selected peptides with corresponding binding constants versus EphA4<sup>92</sup> and aliphatic region of 1D <sup>1</sup>H-NMR spectra of EphA3-LBD in absence (red) and presence (other colors) of selected peptides (1:10 molar ratio). Arrows indicate protein peaks sensitive to ligand binding.

Using **65** as hit compound, we went through an iterative optimization process analyzing the SARs for each of the five positions in the sequence so that we could infer the structural requisites to gain selectivity on EphA3 receptor. Being Trp the most active residue in P5 position, we synthesized seven nor-Trp-derivatives (**72-78**), tested again by 1D <sup>1</sup>H-NMR experiments analyzing the aliphatic region.



**Figure 26.** Analogs derived from optimization of position P5.

5-OMe-tryptamine and 5-Cl-tryptamine were the most active residues in this position so we compared the activity of derivatives **74** and **76** to compound **65** by means of 2D [ $^1\text{H}$ - $^{13}\text{C}$ ]-HSQC experiments that allowed a better evaluation of the binding. The LBD of EphA3 has three methionines, that we had labeled with  $^{13}\text{C}$ , of which only two (Met74 e Met155) are close enough to the binding site so that they experiment a variation in chemical shift upon ligand binding. Therefore, the HSQC spectra reported in figure 27, showing just three signals, were very simple to analyze and gave unambiguous results.



**Figure 27.** 2D- $[^1\text{H}, ^{13}\text{C}]$ -HSQC experiments of the EphA3-LBD in absence (red) and in presence of compounds **65** (blue), **74** (green), **76** (light blue).

Since compound **74** seemed the best candidate among the three, it was also tested by ITC showing a dissociation constant of about  $5\mu\text{M}$ . However, while ITC is perfect to detect and characterize binding of very potent ligands ( $K_d < 1\mu\text{M}$ ), it is far less reliable for detecting weaker binders and, in our case, ITC experiments were not fully reproducible probably due to the low enthalpic contribution to the binding. Hence, we estimated the dissociation constant of the complex by measuring the chemical shift variation upon ligand titration.<sup>164,165</sup> This method is very sensitive to weak binding events and is the election technique for ligands that bind with fast off rates on the NMR time scale (as a rule of thumb,  $K_d$  ranging from  $1\mu\text{M}$  to  $1\text{mM}$  or above). The binding curve for compound **74** was determined by titrating the peptide into a sample of  $^{15}\text{N}$ -EphA3-LBD, acquiring

[ $^{15}\text{N}$ - $^1\text{H}$ ]-HSQC spectra at different ligand:protein ratios till the protein was saturated, and monitoring the chemical shift changes of the backbone amide. The  $K_d$  was then calculated and resulted in a value of  $9.94\ \mu\text{M}$ .

Then, we investigated the role of the N-terminal residue. In fact, from previous studies on EphA4, emerged that this receptor tolerates only Ala or  $\beta$ -Ala as first residue while the presence of bulkier amino acids prevents the binding event. By testing compounds **79-84** (table 5) we demonstrated that the presence of methyl functionalizations on the first amino acid, makes the peptide able to discriminate between the two receptors, since compounds **82** and **83** with a 3-amino-3-methylbutanoic acid and 3-amino-2,2-dimethylpropionic acid respectively as first residue, retain good affinity for EphA3 while unable to bind EphA4. Moreover, we found out that significant modifications introduced in the positions two, three and four of the peptoid negatively influenced the binding to both receptors.

Cmpd n°	Structure	Cmpd n°	Structure
79		82	
80		83	
81		84	

**Table 5.** Analogs derived from optimization of position P1.

### 3.2.3 CONCLUSIONS

In the present study, the ligand binding domain of EphA3 has been expressed for the first time in a  $^{13}\text{C}$ -Met-labeled form. By using HTS by NMR as screening

method we identified first small molecule EphA3 binders, which are endowed with good selectivity between the two homolog receptors EphA3 and EphA4, with dissociation constants against EphA3 in the range of 5-10  $\mu$ M. These binders will be further optimized in order to afford a useful tool to deepen our knowledge of EphA3 pathological role in cancer development.

## EXPERIMENTAL SECTION

---

### I. EXPERIMENTAL PROCEDURES FOR *SINULARIA INELEGANS*

High-resolution ESI-MS spectra were performed with a Micromass Q-TOF mass spectrometer. ESI-MS experiments were performed on an Applied Biosystem API 2000 triple-quadrupole mass spectrometer. Specific rotations were measured on a Jasco P-2000 polarimeter.  $^1\text{H}$  (500 and 700 MHz) and  $^{13}\text{C}$  (125 and 175 MHz) NMR spectra were measured on Varian INOVA spectrometers.  $J$  values are in hertz, chemical shifts ( $\delta$ ) are reported in ppm, and were referenced to the residual solvent signal ( $\text{CHCl}_3$ :  $\delta_{\text{H}}$  7.26,  $\delta_{\text{C}}$  77.0;  $\text{CHD}_2\text{OD}$ :  $\delta_{\text{H}}$  3.30,  $\delta_{\text{C}}$  49.0;  $\text{CHD}_2\text{SOCD}_3$ :  $\delta_{\text{H}}$  2.50,  $\delta_{\text{C}}$  39.5). Homonuclear  $^1\text{H}$  connectivities were determined by the COSY experiment; one-bond heteronuclear  $^1\text{H}$ - $^{13}\text{C}$  connectivities by the HSQC experiment; two- and three-bond  $^1\text{H}$ - $^{13}\text{C}$  connectivities by gradient-HMBC experiments optimized for a  $^{2,3}J$  of 8 Hz. Thin-layer chromatography (TLC) was performed on Alugram silica gel G/UV254 plates. Medium pressure liquid chromatography was performed on a Büchi apparatus using a silica gel (230-400 mesh) column. HPLC was performed using a Waters Model 510 pump equipped with Rheodine injector and a differential refractometer, Waters model 401.

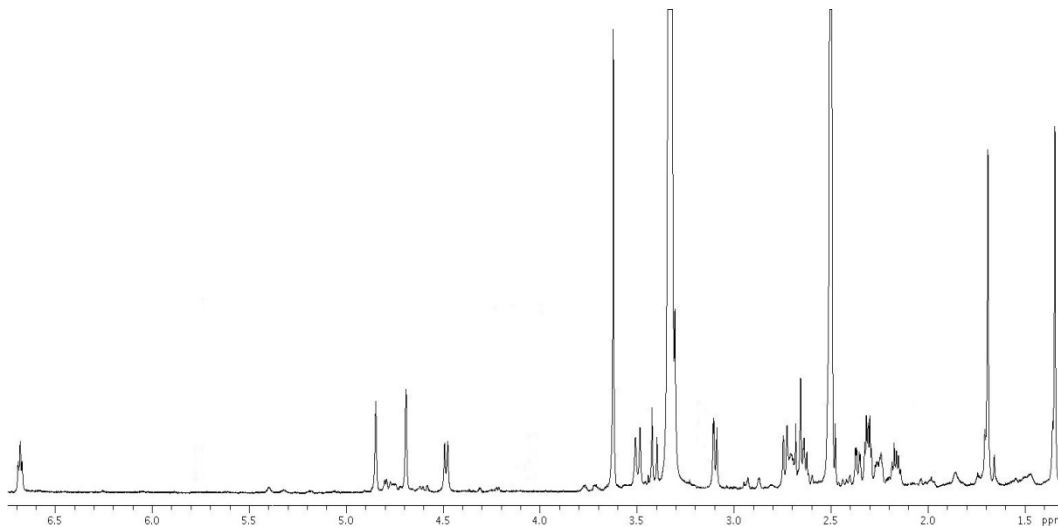
**Organic material and separation of individual norcembranoids.** A sample of *Sinularia ineleqans* (order Alcyonacea, family Alcyoniidae) was collected in the Indian Ocean, off shore from Mandapam, Tamil Nadu, India (Lat - 9° 17' 16"N Long - 79° 07' 55"E) in December 2010. The sample was frozen immediately after collection and lyophilized. A voucher specimen is deposited under the accessing number 14S009 at CSIR-National Institute of Oceanography, India. The lyophilized material was extracted with methanol:chloroform (2:1) at room temperature and the crude methanolic:chloroformic extract (34.75 g) was subjected to a modified Kupchan's partitioning procedure as follows. The methanol extract was dissolved in a mixture of MeOH/H<sub>2</sub>O containing 10% H<sub>2</sub>O and partitioned against *n*-hexane. The water content (% v/v) of the MeOH extract was adjusted to 30% and partitioned against  $\text{CHCl}_3$ . The aqueous phase was concentrated to remove MeOH and then extracted with *n*-BuOH. The  $\text{CHCl}_3$  extract (5.3 g) was chromatographed by MPLC ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$ ) and 23 fractions were collected and combined on the basis of their similar TLC retention factors. Fraction 9, eluted with  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  96:4 (645.3 mg) was purified by HPLC on a Nucleodur C18 column (3 $\mu\text{m}$ , 150 $\times$ 4.60 mm, flow rate 1.0 mL/min) with 45% MeOH/H<sub>2</sub>O as eluent to give 0.9 mg of kavaranolide (**12**) ( $t_{\text{R}}$ =8.4 min), 3.2 mg of scabrolide A (**9**) ( $t_{\text{R}}$ =9.3 min), 2.6 mg of ineleqanolide (**11**) ( $t_{\text{R}}$ =14.7 min), 1.4 mg of scabrolide C (**5**) ( $t_{\text{R}}$ =23.1 min), 0.4 mg of yonanolide (**10**) ( $t_{\text{R}}$ =25.3 min), 0.6 mg of 5-*epi*-norcembrene (**8**) ( $t_{\text{R}}$ =29.5 min), 0.4 mg of scabrolide D (**6**) ( $t_{\text{R}}$ =34.2 min), 0.4 mg

of leptocladolide B (**7**) ( $t_R$ =43.2 min), 0.9 mg of 5-*epi*-norcembrenolide (**1**) ( $t_R$ =45 min), 1.8 mg of norcembrenolide (**2**) ( $t_R$ =51.6 min) and 0.4 mg of 1 $\beta$ ,6 $\alpha$ -dihydroxy-4(14)-eudesmene (**13**) ( $t_R$ =55.8 min). Fraction 11, eluted with CH<sub>2</sub>Cl<sub>2</sub>/MeOH 93:7 (194 mg) was purified by HPLC on a Nucleodur C18 column (3 $\mu$ m, 150 $\times$ 4.60 mm, flow rate 1.0 mL/min) with 45% MeOH/H<sub>2</sub>O as eluent to give 1.3 mg of 5-*epi*-sinuleptolide (**4**) ( $t_R$ =17.5 min) and 3.6 mg of sinuleptolide (**3**) ( $t_R$ =20.5 min).

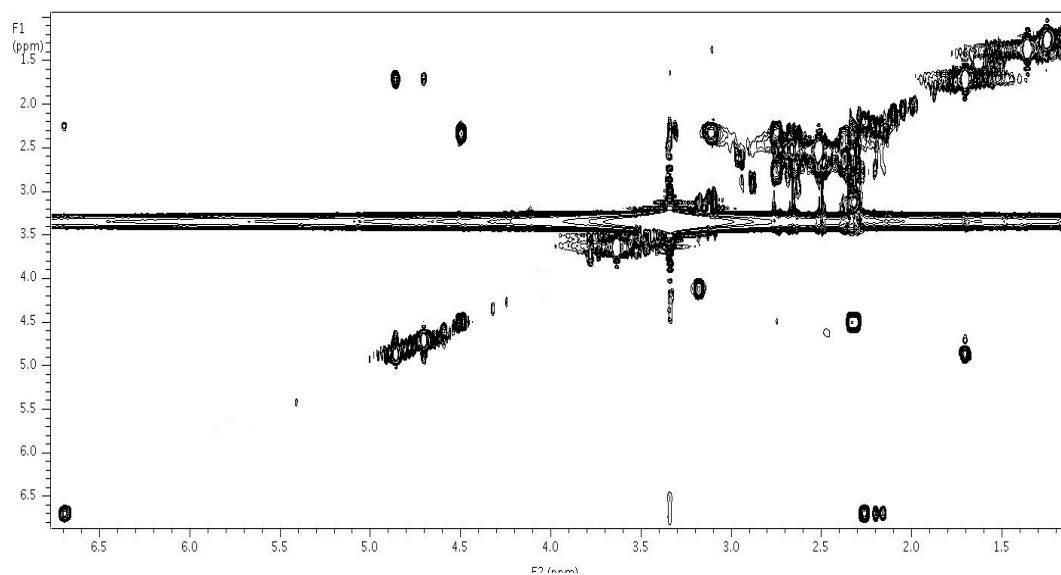
**5-*epi*-norcembrenolide** : white amorphous solid;  $[\alpha]_D^{25}$  +70.0 ( $c$  0.04, chloroform); <sup>1</sup>H and <sup>13</sup>C NMR data in *d*<sub>6</sub>-DMSO given in table 1, chapter 1; MW for C<sub>20</sub>H<sub>26</sub>O<sub>6</sub>: 362.4, HRMS (ESI): 363.1763 [M+H]<sup>+</sup>.

**SPECTROSCOPIC DATA**

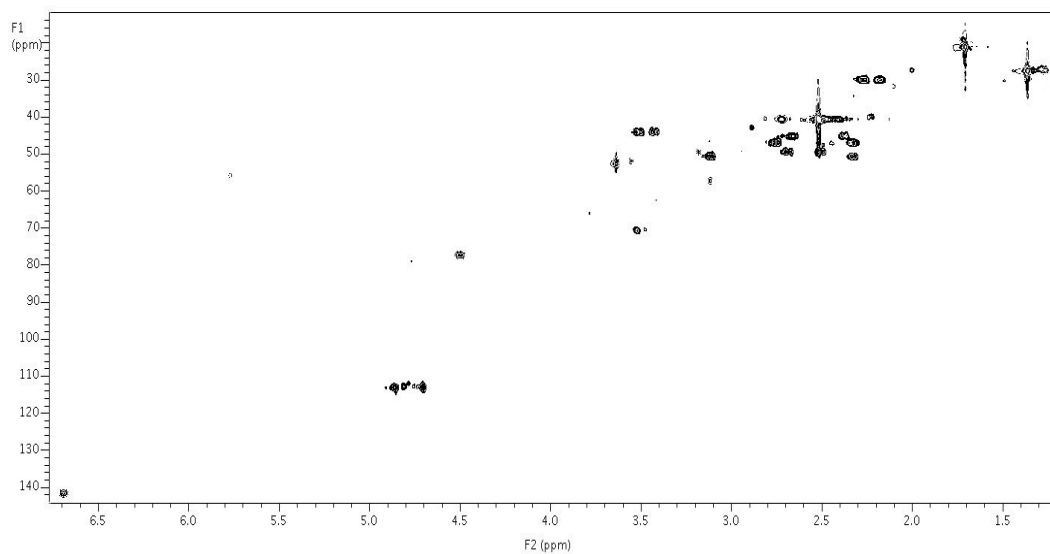
$^1\text{H}$ -NMR spectrum (700 MHz,  $d_6$ -DMSO) of compound **1**



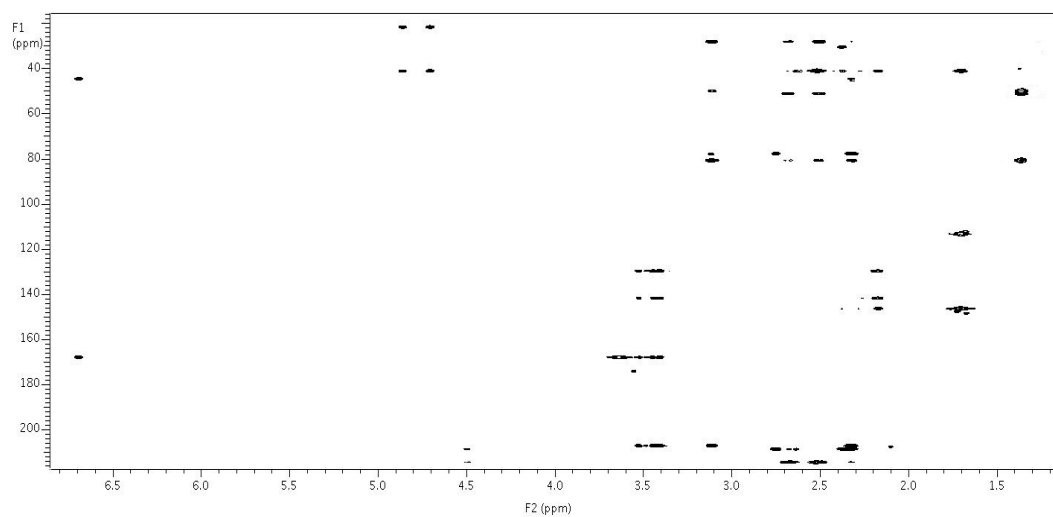
COSY spectrum (700 MHz,  $d_6$ -DMSO) of compound **1**



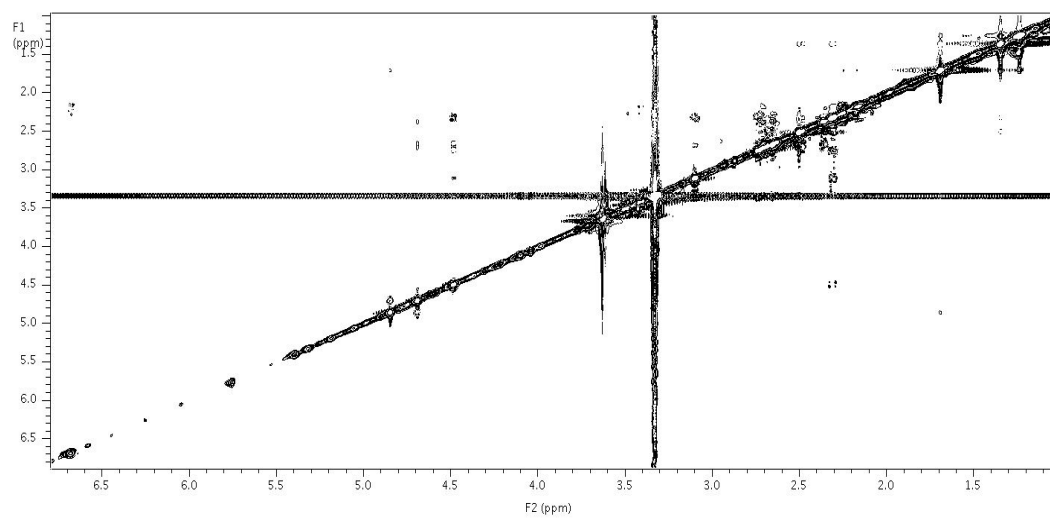
HSQC spectrum (700 MHz,  $d_6$ -DMSO) of compound **1**



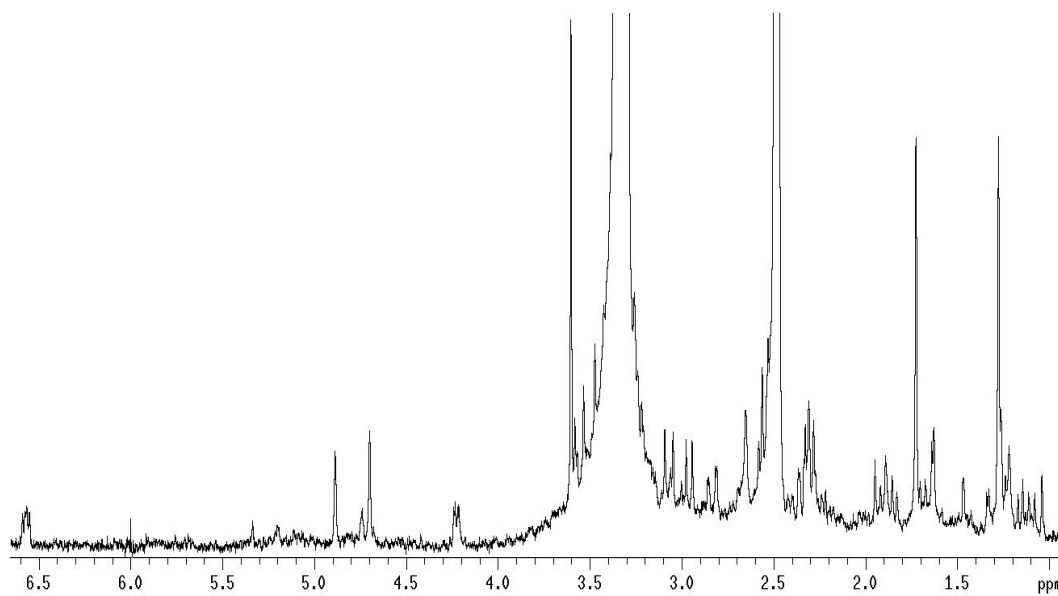
HMBC spectrum (700 MHz,  $d_6$ -DMSO) of compound **1**



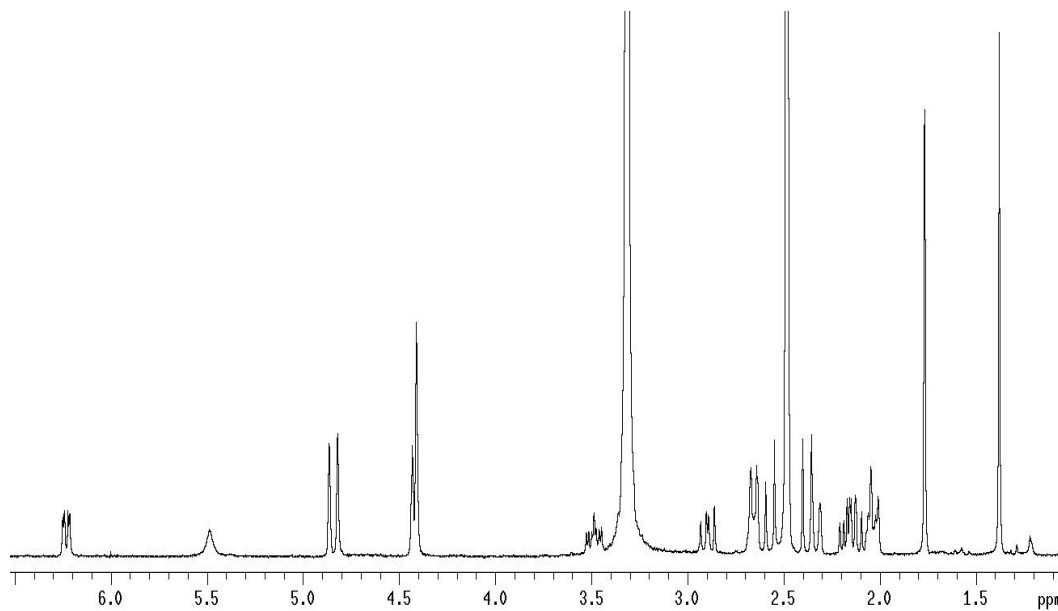
ROESY spectrum (700 MHz,  $d_6$ -DMSO) of compound **1**



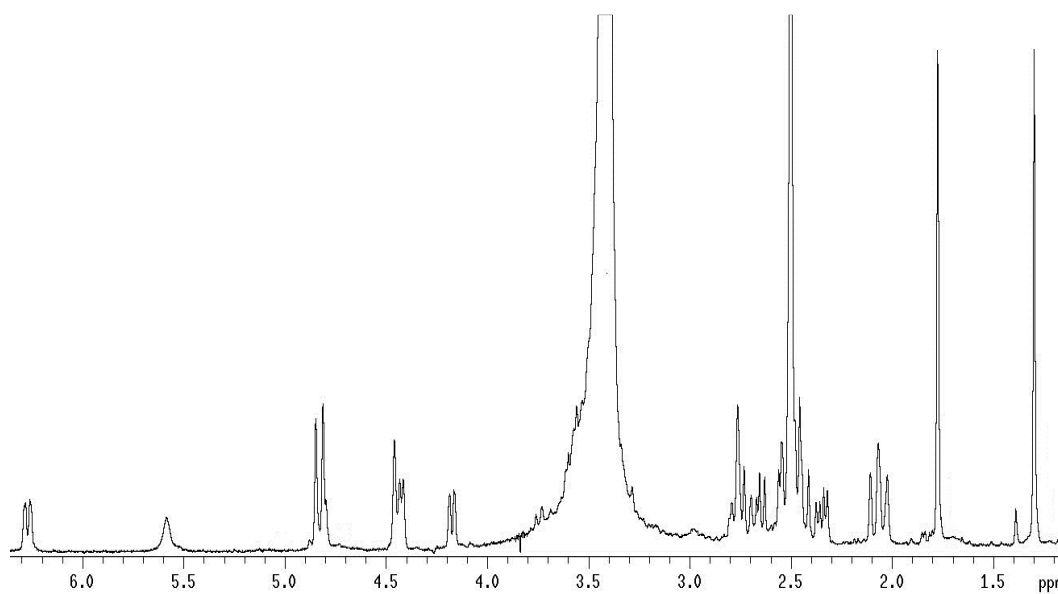
$^1\text{H}$ -NMR spectrum (400 MHz,  $d_6$ -DMSO) of compound **2**



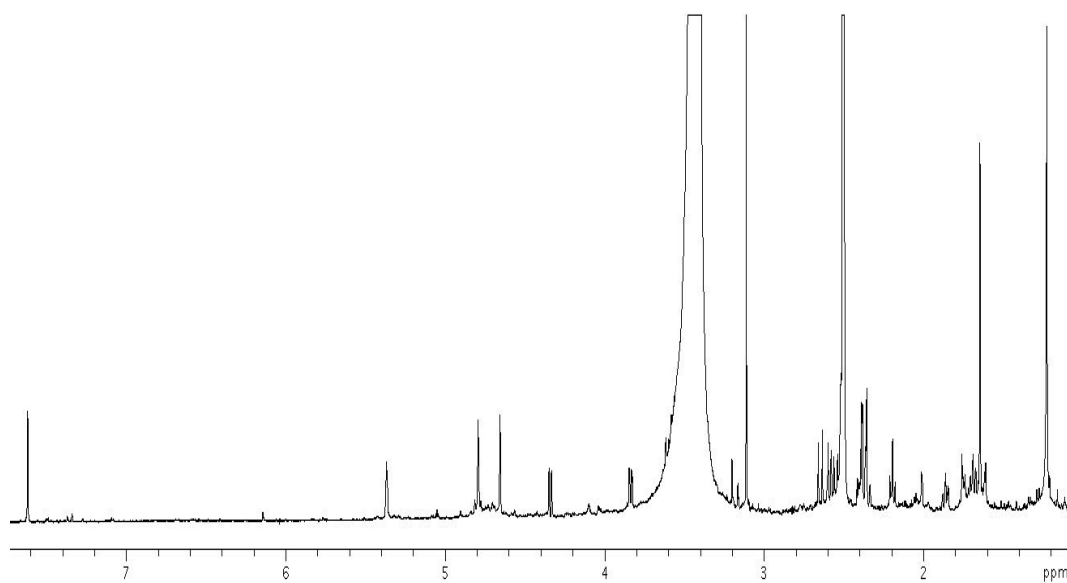
$^1\text{H}$ -NMR spectrum (400 MHz,  $d_6$ -DMSO) of compound **3**



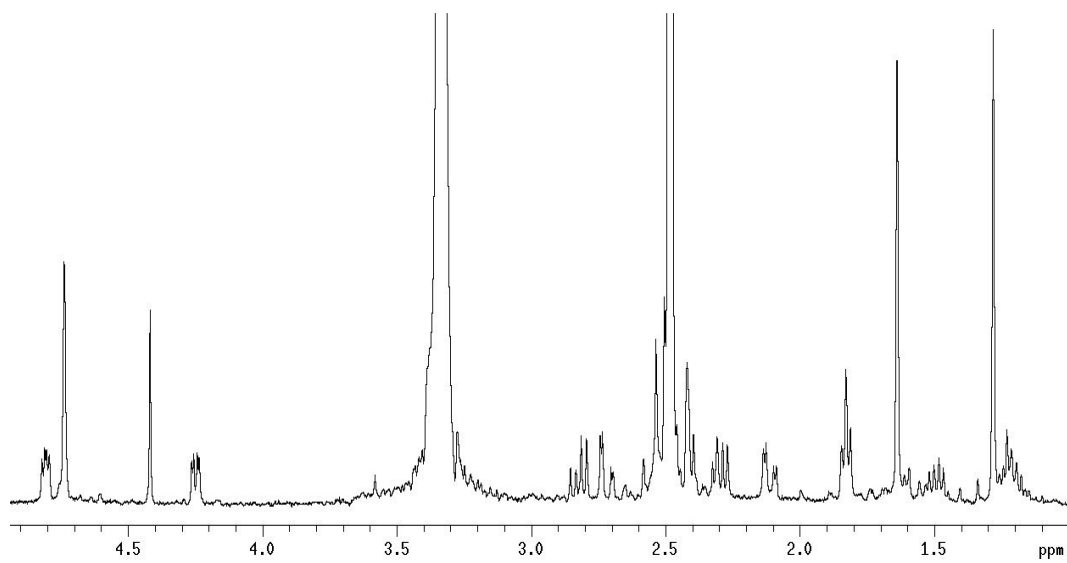
$^1\text{H}$ -NMR spectrum (400 MHz,  $d_6$ -DMSO) of compound **4**



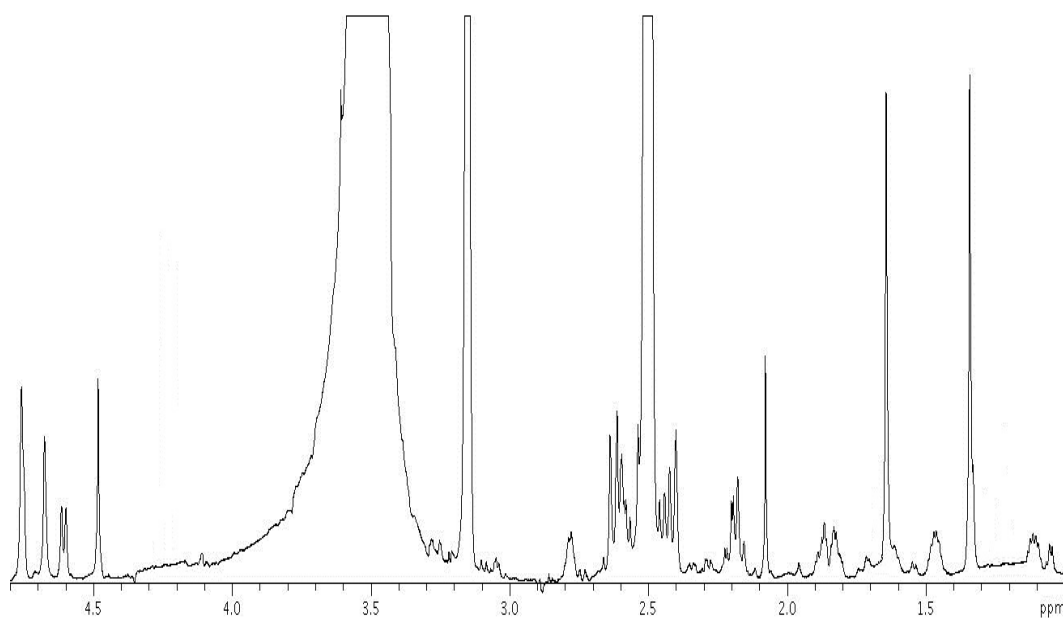
$^1\text{H}$ -NMR spectrum (700 MHz,  $d_6$ -DMSO) of compound **5**



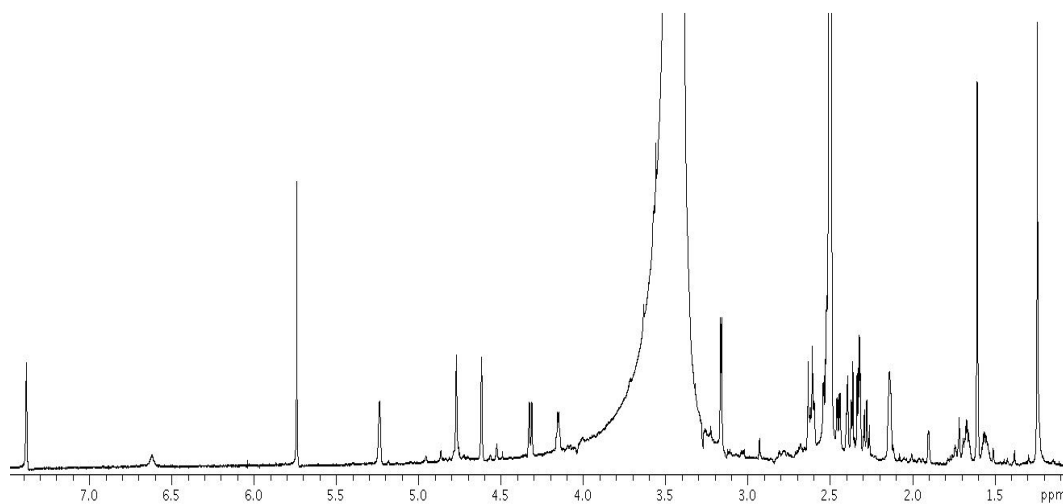
$^1\text{H}$ -NMR spectrum (400 MHz,  $d_6$ -DMSO) of compound **6**



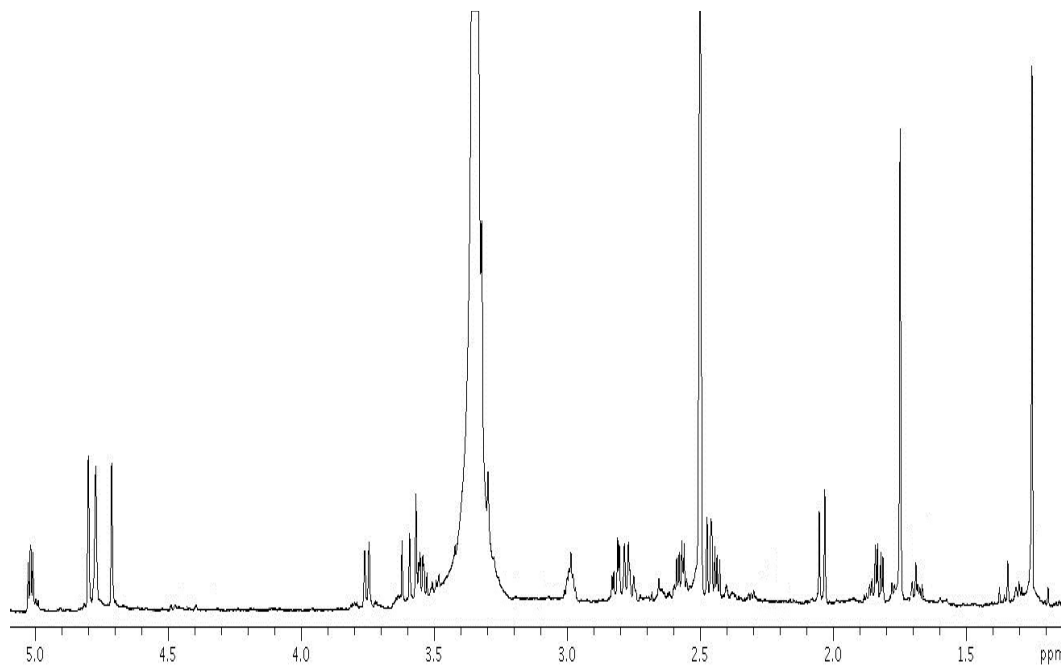
$^1\text{H}$ -NMR spectrum (700 MHz,  $d_6$ -DMSO) of compound **7**



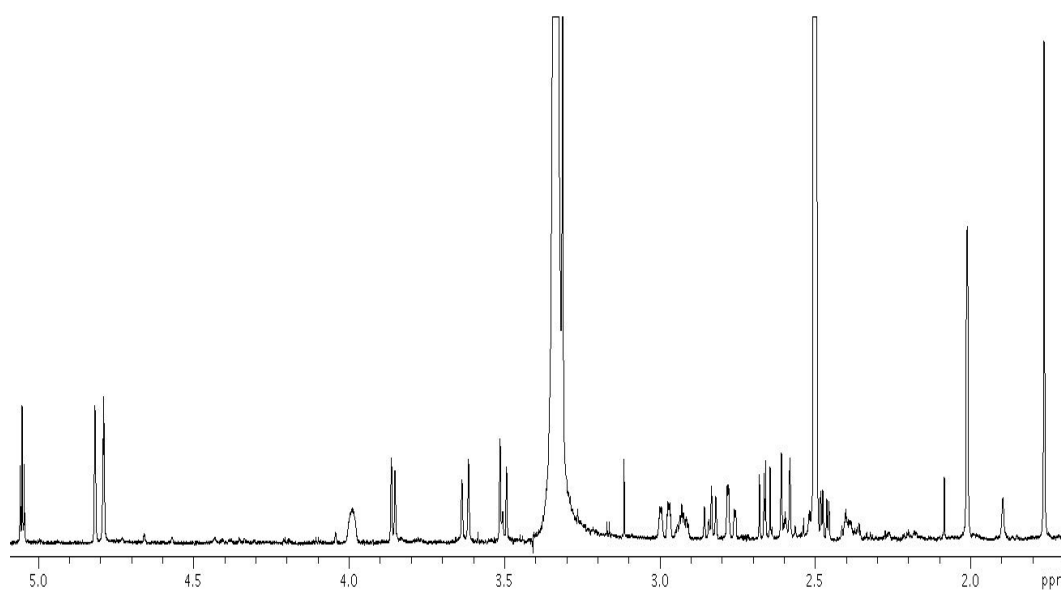
$^1\text{H}$ -NMR spectrum (700 MHz,  $d_6$ -DMSO) of compound **8**



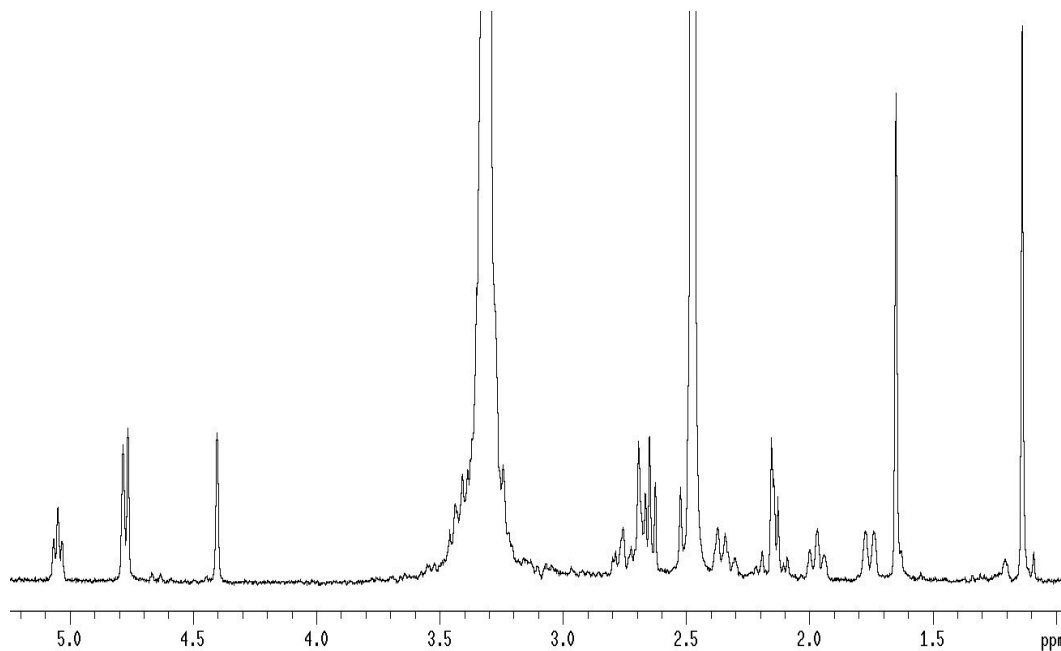
$^1\text{H}$ -NMR spectrum (700 MHz,  $d_6$ -DMSO) of compound **9**



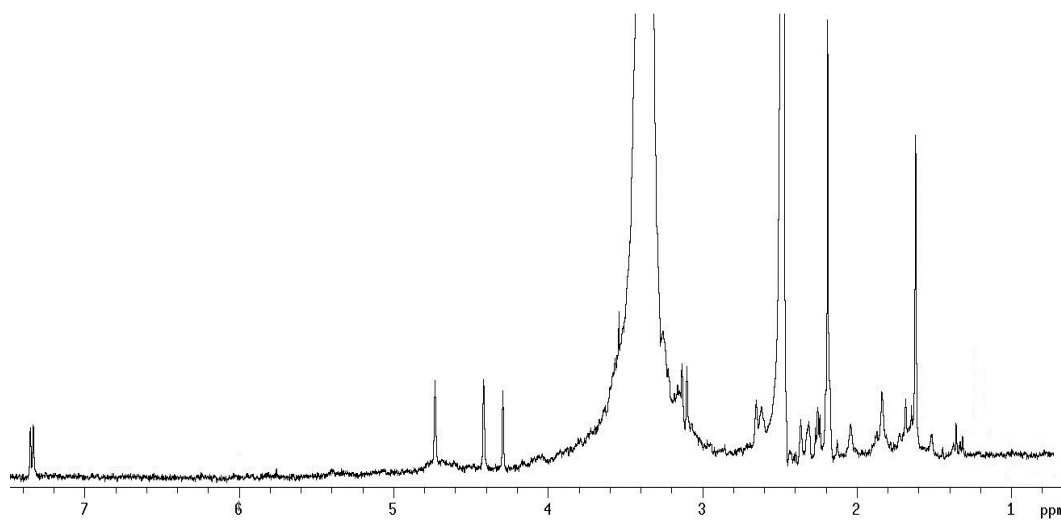
$^1\text{H}$ -NMR spectrum (700 MHz,  $d_6$ -DMSO) of compound **10**



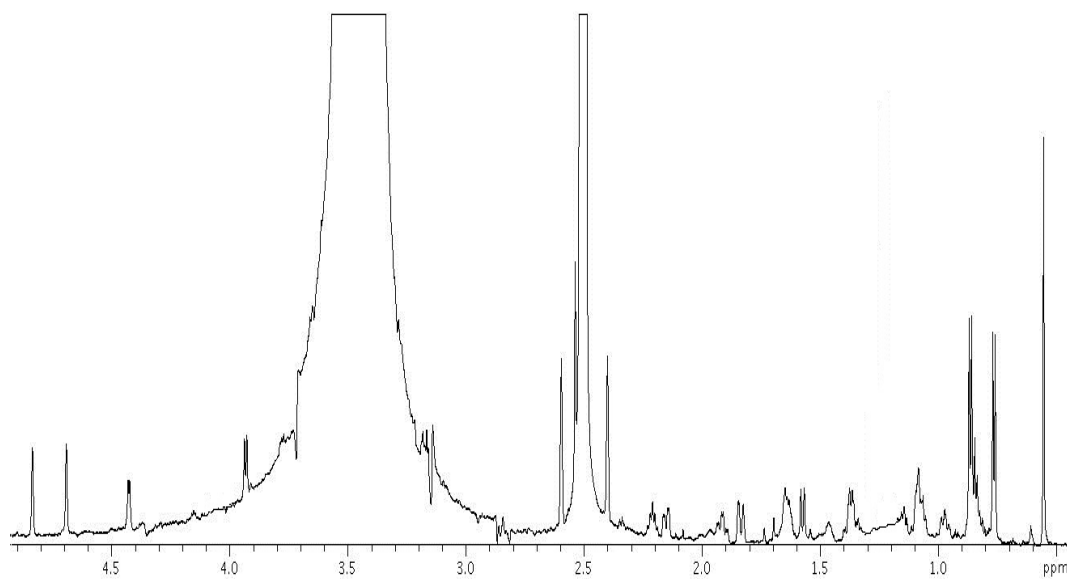
$^1\text{H}$ -NMR spectrum (400 MHz,  $d_6$ -DMSO) of compound **11**



$^1\text{H}$ -NMR spectrum (400 MHz,  $d_6$ -DMSO) of compound **12**



$^1\text{H}$ -NMR spectrum (700 MHz,  $d_6$ -DMSO) of compound **13**



## II. EXPERIMENTAL PROCEDURES FOR FXR/GP-BAR1 DUAL AGONISTS

**Chemistry.** For specific rotations, ESI-MS, NMR and HPLC data see experimental section of *Sinularia inelegrans*, pag.68.

Reaction progress was monitored via thin-layer chromatography (TLC) on Alugram silica gel G/UV254 plates. Silica gel MN Kieselgel 60 (70–230 mesh) from Macherey-Nagel Company was used for column chromatography. All chemicals were obtained from Sigma-Aldrich, Inc. Solvents and reagents were used as supplied from commercial sources with the following exceptions. Tetrahydrofuran, dichloromethane, diisopropylamine, and triethylamine were distilled from calcium hydride immediately prior to use. Methanol was dried from magnesium methoxide as follows. Magnesium turnings (5 g) and iodine (0.5 g) are refluxed in a small (50–100 mL) quantity of methanol until all of the magnesium has reacted. The mixture is diluted (up to 1 L) with reagent grade methanol, refluxed for 2–3 h, and then distilled under nitrogen. All reactions were carried out under argon atmosphere using flame-dried glassware. The purity of all of the intermediates, checked by  $^1\text{H}$  NMR, was greater than 95%. The purity of tested compounds was determined to be always greater than 95% by analytical HPLC analysis as reported for each compound.

Compounds **16** and **17** were prepared as previously reported.<sup>69,70</sup>

**Methyl 6 $\alpha$ -Ethyl-3 $\alpha$ -hydroxy-7-keto-5 $\beta$ -cholan-24-oate (18).** A solution of methyl 3 $\alpha$ -hydroxy-6-ethyliden-7-keto-5 $\beta$ -cholan-24-oate **5** (1.3 g, 3.06 mmol) in dry THF/dry MeOH (50 mL, 1:1 v/v) was hydrogenated in the presence of palladium, 5 wt %, on activated carbon (50 mg). The flask was evacuated and flushed first with argon and then with hydrogen. The mixture was stirred at room temperature under  $\text{H}_2$  for 8 h. The catalyst was filtered through Celite, and the recovered filtrate was concentrated under vacuum to give **18** (1.13 g, quantitative yield). An analytic sample was obtained by silica gel chromatography, eluting with hexane/EtOAc 8:2 and 0.5% of triethylamine.  $[\alpha]_{25}^{\text{D}} -3.7$  (c 0.5,  $\text{CHCl}_3$ ). Selected  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  3.67 (3H, s), 3.57 (1H, m), 2.57 (1H, t,  $J = 11.5$  Hz), 2.37 (1H, m), 2.24 (1H, dd,  $J = 6.6, 9.6$  Hz), 2.20 (1H, m), 1.22 (3H, s), 0.93 (3H, d,  $J = 6.2$  Hz), 0.85 (3H, t,  $J = 7.3$  Hz), 0.67 (3H, s).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  215.1, 178.4, 71.6, 56.3, 51.3, 50.6, 50.4, 47.5, 46.4, 44.4, 43.8, 40.4, 38.2, 36.6, 36.3, 35.2, 32.4, 32.0, 30.6, 29.3, 25.8, 23.5 (2C), 22.8, 18.8, 12.5, 12.0. HRMS-ESI  $m/z$  433.3323  $[\text{M} + \text{H}]^+$ ,  $\text{C}_{27}\text{H}_{45}\text{O}_4$  requires 433.3318.

**6 $\alpha$ -Ethyl-3 $\alpha$ -hydroxy-7-keto-5 $\beta$ -cholan-24-oic Acid (19).** Compound **18** (1.1 g, 2.5 mmol) was hydrolyzed with a methanol solution of sodium hydroxide (5%, 30 mL) in  $\text{H}_2\text{O}$  (6 mL) overnight under reflux. The resulting solution was then concentrated under vacuum, diluted with water, acidified with HCl, 6 N, and extracted with ethyl acetate ( $3 \times 50$  mL). The collected organic phases were washed with brine, dried over anhydrous  $\text{Na}_2\text{SO}_4$ , and evaporated under reduced

pressure to give **19** in quantitative yield (1.1 g). An analytic sample was obtained by silica gel chromatography, eluting with CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5.  $[\alpha]_{25}^D$  -21.5 (c 0.35, CH<sub>3</sub>OH). Selected <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  3.46 (1H, m), 2.83 (1H, dd, J = 13.0, 5.5 Hz), 2.50 (1H, t, J = 11.2 Hz), 2.34 (1H, m), 2.20 (1H, m), 1.22 (3H, s), 0.96 (3H, d, J = 6.6 Hz), 0.81 (3H, t, J = 7.3 Hz), 0.71 (3H, s). <sup>13</sup>C NMR (100 MHz, CD<sub>3</sub>OD):  $\delta$  215.5, 187.0, 71.7, 56.4, 53.3, 52.2, 51.2, 50.5, 45.4, 43.8, 40.4, 36.8, 36.6, 35.3, 32.6, 32.3, 32.0, 30.6, 29.3, 25.6, 23.9, 23.0, 20.8, 18.8, 12.5, 12.3. HRMS-ESI  $m/z$  419.3164 [M + H]<sup>+</sup>, C<sub>26</sub>H<sub>43</sub>O<sub>4</sub> requires 419.3161.

**6 $\alpha$ -Ethyl-3 $\alpha$ -formyloxy-7-keto-5 $\beta$ -cholan-24-oic Acid (20).** A solution of **19** (1.0 g, 2.6 mmol) in 30 mL of 90% formic acid containing 90  $\mu$ L of 70% perchloric acid was stirred at 47–50 °C for 6 h. The temperature of the heating bath was lowered to 40 °C. Then 24 mL of acetic anhydride was added over 10 min, and the mixture was stirred for additional 10 min. The solution was cooled to room temperature, poured into 50 mL of water, and extracted with diethyl ether. The organic layers were washed with water to neutrality, dried over Na<sub>2</sub>SO<sub>4</sub>, and evaporated to give 940 mg of **20** (81%). An analytic sample was obtained by silica gel chromatography, eluting with CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5.  $[\alpha]_{25}^D$  -25.9 (c 0.56, CH<sub>3</sub>OH). Selected <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  7.99 (1H, s), 4.79 (1H, m), 2.71 (1H, dd, J = 5.9, 12.8 Hz), 1.29 (3H, s), 0.93 (3H, d, J = 6.3 Hz), 0.80 (3H, t, J = 7.2 Hz), 0.66 (3H, s). <sup>13</sup>C NMR (100 MHz CDCl<sub>3</sub>):  $\delta$  212.5, 180.0, 160.6, 73.2, 54.8, 51.9, 50.6, 49.9, 48.9, 43.6, 42.6, 38.9, 35.7, 35.1, 33.8, 30.9, 30.7, 30.3, 28.2, 27.6, 25.9, 24.5, 23.4, 21.8, 18.8, 12.0, 11.9. HRMS-ESI  $m/z$  447.3116 [M + H]<sup>+</sup>, C<sub>27</sub>H<sub>43</sub>O<sub>5</sub> requires 447.3110.

**6 $\alpha$ -Ethyl-3 $\alpha$ -formyloxy-7-keto-24-nor-5 $\beta$ -cholan-23-nitrile (21).** Crude **20** (930 mg, 2.08 mmol), 6.7 mL of cold trifluoroacetic acid, and 1.8 mL (15.6 mmol) of trifluoroacetic anhydride were stirred at 0–5 °C until dissolution. Sodium nitrite (435 mg, 6.3 mmol) was added in small portions. After the addition was complete, the reaction mixture was stirred first at 0–5 °C for 1 h, then at 38–40 °C for 2 h. On completion, the mixture was neutralized with NaOH, 2N, and then the product was extracted with 50 mL of diethyl ether (3  $\times$  50 mL), followed by washing with brine and dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>. The ether was removed under reduced pressure to afford 860 mg of **21** in quantitative yield, which was subjected to the next step without any purification.

**6 $\alpha$ -Ethyl-3 $\alpha$ -hydroxy-7-keto-24-nor-5 $\beta$ -cholan-23-oic Acid (22).** Crude compound **21** (860 mg, 2.08 mmol) was refluxed with 30% KOH in about 50 mL of methanol/water, 1:1. After the mixture was stirred for 48 h, the basic aqueous solution was neutralized with HCl, 6N. Then methanol was evaporated and the residue was extracted with AcOEt (3  $\times$  50 mL). The combined organic layers were washed with brine, dried, and evaporated to dryness to give white solid residue **22** (723 mg, 86%). An analytic sample was obtained by silica gel chromatography, eluting with CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5.  $[\alpha]_{25}^D$  -25.2 (c 0.22, CH<sub>3</sub>OH). Selected <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  3.46 (1H, m), 2.82 (1H, dd, J = 6.4, 12.6 Hz), 1.25 (3H, s), 1.25 (3H, d ovl), 0.81 (3H, t, J = 7.3 Hz), 0.74 (3H, s). <sup>13</sup>C

NMR (100 MHz, CD<sub>3</sub>OD):  $\delta$  215.7, 177.6, 71.7, 56.3, 53.3, 51.2, 50.6, 50.5, 45.3, 43.8, 42.5, 40.4, 36.8, 35.3, 35.0, 32.6, 30.5, 29.4, 25.6, 23.9, 22.9, 20.9, 19.6, 12.5, 12.3. HRMS-ESI  $m/z$  405.3008 [M + H]<sup>+</sup>, C<sub>25</sub>H<sub>41</sub>O<sub>4</sub> requires 405.3005.

**Methyl 6 $\alpha$ -Ethyl-3 $\alpha$ , 7 $\alpha$ -dihydroxy-24-nor-5 $\beta$ -cholan-23-oate (23).** Compound **22** (715 mg, 1.7 mmol) was dissolved in a solution of tetrahydrofuran/water (50 mL, 4/1 v/v) and treated at 0 °C with NaBH<sub>4</sub> (320 mg, 8.5 mmol). After 1 h, water and MeOH were added dropwise during a period of 15 min at 0 °C with effervescence being observed. Then after evaporation of the solvents, the residue was diluted with water, acidified with HCl, 1N, and extracted with AcOEt (3  $\times$  50 mL). The combined organic phases were washed with brine, dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>, and evaporated under reduced pressure. The crude residue was purified by flash chromatography on silica gel, using dichloromethane/methanol, 9:1, as eluent, to afford 640 mg of 3 $\alpha$ ,7 $\alpha$ -dihydroxy-6 $\alpha$ -ethyl-24-nor-5 $\beta$ -cholan-23-oic acid (93% yield).  $[\alpha]_{25}^D$  +3.7 (c 1.48, CH<sub>3</sub>OH). Selected <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  3.62 (1H, br s), 3.30 (1H, m ovl), 0.97 (3H, d, J = 7.9 Hz), 0.88 (3H, s), 0.87 (3H, t ovl), 0.70 (3H, s). <sup>13</sup>C NMR (100 MHz, CD<sub>3</sub>OD):  $\delta$  177.8, 73.2, 71.2, 57.5, 51.7, 46.9, 43.9, 43.8, 42.5, 41.6, 41.0, 36.7, 35.1, 34.5, 34.4, 31.3, 29.4, 24.6, 23.8, 23.5, 21.9, 19.6, 12.3, 12.0. HRMS-ESI  $m/z$  407.3166 [M + H]<sup>+</sup>, C<sub>25</sub>H<sub>43</sub>O<sub>4</sub> requires 407.3161. The above compound (640 mg, 1.6 mmol) was dissolved in 50 mL of dry methanol and treated with p-toluenesulfonic acid (1.5 g, 8.0 mmol). The solution was left to stand at room temperature overnight. The mixture was quenched by addition of NaHCO<sub>3</sub> solution until neutrality. Most of the solvent was evaporated, and the residue was extracted with EtOAc. The combined extract was washed with brine, dried with Na<sub>2</sub>SO<sub>4</sub>, and evaporated to give **23** as an amorphous solid (645 mg, 96%). An analytic sample was obtained by silica gel chromatography, eluting with hexane/EtOAc 6:4.  $[\alpha]_{25}^D$  + 0.95 (c 0.96, CHCl<sub>3</sub>). Selected <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  3.61 (3H, s), 3.62 (1H, s ovl), 3.28 (1H, m), 0.95 (3H, d, J = 6.9 Hz), 0.88 (3H, s), 0.87 (3H, t, J = 7.2 Hz), 0.69 (3H, s). <sup>13</sup>C NMR (100 MHz, CD<sub>3</sub>OD):  $\delta$  175.7, 73.2, 71.2, 57.8, 51.8, 51.6, 46.9, 43.9, 43.8, 43.1, 42.3, 41.5, 41.0, 36.7, 35.2, 35.1, 34.4, 31.2, 29.3, 24.5, 23.8, 23.5, 21.9, 19.6, 12.3, 12.0. HRMS-ESI  $m/z$  421.3322 [M + H]<sup>+</sup>, C<sub>26</sub>H<sub>45</sub>O<sub>4</sub> requires 421.3318.

**Methyl 3 $\alpha$ ,7 $\alpha$ -Di(*tert*-butyldimethylsilyloxy)-6 $\alpha$ -ethyl-24-nor-5 $\beta$ -cholan-23-oate (24).** To a solution of **23** (600 mg, 1.4 mmol) in 20 mL of CH<sub>2</sub>Cl<sub>2</sub> at 0 °C were added 2,6-lutidine (14.3 mmol, 1.6 mL) and *tert*-butyl dimethylsilyltrifluoromethanesulfonate (4.2 mmol, 960  $\mu$ L). After the mixture was stirred for 24 h at 0 °C, the reaction was quenched by addition of aqueous NaHSO<sub>4</sub> (1 M, 50 mL). The layers were separated, and the aqueous phase was extracted with CH<sub>2</sub>Cl<sub>2</sub> (3  $\times$  50 mL). The combined organic layers were washed with NaHSO<sub>4</sub>, water, saturated aqueous NaHCO<sub>3</sub>, and brine. Purification by flash chromatography on silica gel using hexane/ethyl acetate, 9:1, and 0.5% of triethylamine as eluent gave **24** (722 mg, 78%) as a clear, colorless oil.  $[\alpha]_{25}^D$  +2.44 (c 0.62, CHCl<sub>3</sub>). Selected <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  3.69 (1H, br s),

3.48 (3H, s), 3.40 (1H, m), 1.02 (3H, d,  $J = 6.3$  Hz), 0.90 (3H, t,  $J = 7.3$  Hz), 0.89 (3H, s), 0.86 (18H, s), 0.70 (3H, s), 0.042 (12H, s).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  176.3, 74.1, 72.9, 56.0, 51.6, 51.6, 49.4, 45.1, 43.2, 42.4, 41.9, 41.5, 40.8, 35.7, 34.9, 34.1, 33.9, 31.8, 31.3, 26.7, 26.6, 25.7 (3C), 25.6 (3C), 24.4, 24.0, 23.0, 19.3, 18.8, 12.0, 11.9,  $-3.0$  (2C),  $-2.8$  (2C). HRMS-ESI  $m/z$  641.5051  $[\text{M} + \text{H}]^+$ ,  $\text{C}_{38}\text{H}_{73}\text{O}_4\text{Si}_2$  requires 641.5047.

**6 $\alpha$ -Ethyl-3 $\alpha$ ,7 $\alpha$ -dihydroxy-24-nor-5 $\beta$ -cholan-23yl-23-triethylammonium**

**Sulfate (14).** Dry methanol (133  $\mu\text{L}$ , 3.3 mmol) and  $\text{LiBH}_4$  (1.7 mL, 2 M in THF, 3.3 mmol) were added to a solution of **24** (722 mg, 1.1 mmol) in dry THF (15 mL) at  $0^\circ\text{C}$  under argon, and the resulting mixture was stirred for 8 h at  $0^\circ\text{C}$ . The mixture was quenched by addition of NaOH (1 M, 2.2 mL) and then allowed to warm to room temperature. Ethyl acetate was added, and the separated aqueous phase was extracted with ethyl acetate ( $3 \times 30$  mL). The combined organic phases were washed with water, dried ( $\text{Na}_2\text{SO}_4$ ), and concentrated. Purification by silica gel (n-hexane/ethyl acetate 8:2) gave C23 alcohol derivative as a colorless oil (650 mg, 95%).  $[\alpha]_{25}^{\text{D}} -0.25$  (c 0.5,  $\text{CH}_3\text{OH}$ ). Selected  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  3.70 (1H, m), 3.67 (1H, br s), 3.60 (1H, m), 3.37 (1H, m), 0.94 (3H, d,  $J = 6.4$  Hz), 0.87 (3H, s), 0.87 (3H, t, ovl), 0.86 (18H, s), 0.66 (3H, s), 0.042 (12H, s).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  73.6, 71.2, 61.0, 56.6, 50.8, 45.6, 43.0, 41.5, 40.3, 39.9, 36.0, 35.8, 34.4, 33.5, 33.2 (2C), 31.3, 28.6, 26.2 (7C), 23.9, 23.4, 22.5, 21.0, 19.0, 12.0, 11.8,  $-4.3$  (2C),  $-4.4$  (2C). HRMS-ESI  $m/z$  621.5093  $[\text{M} + \text{H}]^+$ ,  $\text{C}_{37}\text{H}_{73}\text{O}_3\text{Si}_2$  requires 621.5098. The triethylamine–sulfur trioxide complex (906 mg, 5 mmol) was added to a solution of C23 alcohol (620 mg, 1 mmol) in DMF dry (7 mL) under an argon atmosphere, and the mixture was stirred at  $95^\circ\text{C}$  for 24 h. The solution was then concentrated under vacuum. To the solid dissolved in methanol (30 mL) was added three drops of HCl, 37% v/v, and the mixture was stirred for 5 h at room temperature. At the end of reaction, silver carbonate was added to precipitate chloride. Then the reaction mixture was centrifuged and the supernatant was concentrated in vacuo. The residue was poured over a RP18 column. The fraction eluted with  $\text{H}_2\text{O}/\text{MeOH}$  (1:1) gave compound **14** as a white solid (500 mg, 87%). A sample (20 mg) was subjected to purification by HPLC on a Nucleodur 100-5 C18 (5  $\mu\text{m}$ , 10 mm i.d.  $\times$  250 mm) with  $\text{MeOH}/\text{H}_2\text{O}$  (65:35) as eluent (flow rate 3 mL/min ( $t_{\text{R}} = 24.8$  min)).  $[\alpha]_{25}^{\text{D}} +2.9$  (c 0.17,  $\text{CH}_3\text{OH}$ ).  $^1\text{H}$  and  $^{13}\text{C}$  NMR data in  $\text{CD}_3\text{OD}$  are given in table 6. HR ESIMS  $m/z$  471.2775  $[\text{M} - \text{Na}]$ ,  $\text{C}_{25}\text{H}_{43}\text{O}_6\text{S}$  requires 471.2780.

**Table 6.**  $^1\text{H}$  and  $^{13}\text{C}$  NMR data (700 MHz,  $\text{CD}_3\text{OD}$ ) for **14**

position	$\delta_{\text{H}}^{\text{a}}$	$\delta_{\text{C}}$	HMBC
1	0.99 ovl <sup>b</sup> 1.86 ovl	36.4	
2	1.32 m 1.59 m	31.0	
3	3.28 m	72.9	
4	1.72 1.88	34.1	C3, C5
5	1.28 m	46.7	C4
6	1.54 m	42.9	C5
7	3.64	70.8	C5, C8
8	1.45	41.4	
9	1.80 m	34.2	
10	-	47.6	
11	1.30 1.49	21.6	
12	1.20 ovl 2.00	40.8	
13	-	43.7	
14	1.51	51.3	
15	1.10 1.70	24.3	
16	1.33 1.90	29.1	C14
17	1.18 m	57.5	
18	0.73 s	12.0	C12, C13, C14, C17
19	0.92 s	23.2	C1, C5, C9, C10
20	1.60 m	33.9	
21	0.99 d (6.4)	18.8	C17, C20, C22
22	1.30 ovl	36.3	
23	4.02 m 4.06 m	67.0	C24, C25
24	1.38 1.51	23.3	C27
25	0.90 t (7.2)	11.9	
$\text{NH}(\text{CH}_2\text{CH}_3)_3^+$	3.19 q (7.4)	46.8	
$\text{NH}(\text{CH}_2\text{CH}_3)_3^+$	1.30 t (7.4)	8.3	

<sup>a</sup>Coupling constants are in parentheses and given in hertz.<sup>1</sup>H and <sup>13</sup>C assignments aided by COSY, HSQC and HMBC experiments.<sup>b</sup>Ovl: overlapped with other signals.

**Compound 25.** To a solution of **23** (30 mg, 0.07 mmol) in dry THF (5 mL) at 0 °C, dry methanol (20  $\mu\text{L}$ , 0.5 mmol) and  $\text{LiBH}_4$  (250  $\mu\text{L}$ , 2 M in THF, 0.5 mmol) were added. The resulting mixture was stirred for 8 h at 0 °C. The mixture was quenched by addition of NaOH (1 M, 240  $\mu\text{L}$ ), and then ethyl acetate was added. The separated aqueous phase was extracted with ethyl acetate (3  $\times$  30 mL). The combined organic phases were washed with water, dried ( $\text{Na}_2\text{SO}_4$ ), and concentrated. Purification by silica gel ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$  9:1) gave triol derivative as a white solid (27 mg, quantitative yield).  $[\alpha]_{25}^{\text{D}} +1.97$  (c 0.5,  $\text{CH}_3\text{OH}$ ). Selected  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  3.67 (1H, br s), 3.60 (2H, m), 3.43 (1H, m), 0.93 (3H, d,  $J = 6.6$  Hz), 0.88 (3H, s), 0.87 (3H, t,  $J = 7.1$  Hz), 0.67 (3H, s).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  73.2, 71.2, 60.8, 57.9, 51.7, 51.6, 46.9, 43.9, 43.1, 41.5, 41.0, 39.8, 36.7, 36.6, 34.5, 34.4, 34.2, 31.2, 29.3, 24.5, 23.7, 23.5, 21.9, 19.5, 19.3, 12.2, 12.0. HRMS-ESI  $m/z$  393.3367  $[\text{M} + \text{H}]^+$ ,  $\text{C}_{25}\text{H}_{45}\text{O}_3$  requires 393.3369. Triethylamine–sulfur trioxide complex (127 mg, 0.7 mmol) was added to triol (27 mg, 0.07 mmol) in DMF dry (10 mL) under an argon atmosphere, and

the mixture was stirred at 95 °C for 48 h. The reaction mixture was quenched with water (1.6 mL), and the solution was poured over a C18 silica gel column to remove excess Et<sub>3</sub>N·SO<sub>3</sub>. The fraction eluted with H<sub>2</sub>O/MeOH (9:1) gave a mixture, which was further purified by HPLC on a Nucleodur 100-5 C18 (5 µm, 10 mm i.d. × 250 mm) with MeOH/H<sub>2</sub>O (35:65) as eluent (flow rate 3 mL/min) to give compound **25** as a white solid (47 mg, 72%). [ $\alpha$ ]<sub>25</sub><sup>D</sup> -2.3 (c 0.24, CH<sub>3</sub>OH). Selected <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  4.62 (1H, br s), 4.11 (1H, m), 4.03 (2H, m), 3.19 (18H, q, J = 7.2 Hz), 1.30 (27H, t, J = 7.2 Hz), 0.99 (3H, d, J = 6.5 Hz), 0.96 (3H, s), 0.93 (3H, t, J = 7.4 Hz), 0.71 (3H, s). HR ESIMS *m/z* 833.4322 [M – Et<sub>3</sub>NH], C<sub>37</sub>H<sub>73</sub>N<sub>2</sub>O<sub>12</sub>S<sub>3</sub> requires 833.4326.

**Chenodeoxycholan Sulfate Derivative Synthesis. Methyl 3 $\alpha$ ,7 $\alpha$ -Dihydroxy-5 $\beta$ -cholan-24-oate (26).** A mixture of CDCA (**15**, 100 mg, 0.25 mmol) and p-toluenesulfonic acid (237 mg, 1.25 mmol) in dry methanol (10 mL) was left to stand at room temperature for 1 h. The mixture was quenched by addition of NaHCO<sub>3</sub> solution until neutrality. Most of the solvent was evaporated, and the residue was extracted with EtOAc. The combined extract was washed with brine, dried with Na<sub>2</sub>SO<sub>4</sub>, and evaporated to give the desired methyl ester as colorless amorphous solids (102 mg, quantitative yield). An analytic sample was obtained by silica gel chromatography, eluting with CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5. [ $\alpha$ ]<sub>25</sub><sup>D</sup> +7.81 (c 2.33, CHCl<sub>3</sub>). Selected <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  3.73 (1H, br s), 3.57 (3H, s), 3.30 (1H, m ov), 2.26 (1H, m), 2.13 (1H, m), 0.84 (3H, d, J = 6.0 Hz), 0.82 (3H, s), 0.60 (3H, s). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>):  $\delta$  174.4, 71.2, 67.6, 55.3, 51.0, 49.8, 42.0, 41.1, 39.2, 38.8 (2C), 34.9, 34.8, 34.5, 34.2, 32.3, 30.5, 30.4, 29.9, 27.7, 23.0, 22.3, 20.1, 17.7, 11.2. HRMS-ESI *m/z* 407.3165 [M + H]<sup>+</sup>, C<sub>25</sub>H<sub>43</sub>O<sub>4</sub> requires 407.3161.

**3 $\alpha$ ,7 $\alpha$ ,24-5 $\beta$ -Cholantriol (28).** Methyl ester **26** (100 mg, 0.25 mmol) was dissolved in dry THF (5 mL) at 0 °C in the presence of dry methanol (70 µL, 0.84 mmol) and LiBH<sub>4</sub> (875 µL, 2 M in THF, 1.75 mmol). After 8 h, a solution of 1 M NaOH (500 µL) was added and then allowed to warm to room temperature. Ethyl acetate was added, and the separated aqueous phase was extracted with ethyl acetate (3 × 30 mL). The combined organic phases were washed with water, dried (Na<sub>2</sub>SO<sub>4</sub>), and concentrated. Purification by silica gel (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5) gave triol derivative **28** as a colorless oil (90 mg, 96%). [ $\alpha$ ]<sub>25</sub><sup>D</sup> +0.82 (c 2.07, CH<sub>3</sub>OH). Selected <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  3.80 (1H, br s), 3.56 (2H, m), 3.40 (1H, m), 0.91 (3H, d, J = 6.0 Hz), 0.87 (3H, s), 0.63 (3H, s). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>):  $\delta$  71.8, 68.4, 63.2, 56.0, 50.3, 42.5, 41.5, 39.6, 39.3 (2C), 35.5, 35.3, 34.9, 34.5, 32.8, 30.5 (2C), 29.3, 28.2, 23.6, 22.7, 20.5, 18.6, 11.7. HRMS-ESI *m/z* 379.3216 [M + H]<sup>+</sup>, C<sub>24</sub>H<sub>43</sub>O<sub>3</sub> requires 379.3212.

**3 $\alpha$ ,7 $\alpha$ -Dihydroxy-5 $\beta$ -cholan-24-yl-24-sodium Sulfate (30).** At a solution of triol **28** (40 mg, 0.1 mmol) in DMF dry (3 mL) was added triethylamine–sulfur trioxide complex (36 mg, 0.2 mmol) under an argon atmosphere, and the mixture was stirred at 95 °C for 1 h. Most of the solvent was evaporated, and the residue was poured over a RP18 column to remove excess SO<sub>3</sub>·NEt<sub>3</sub>. The fraction eluted

with H<sub>2</sub>O/MeOH 1:1 gave a mixture that was further purified by HPLC on a Nucleodur 100-5 C18 (5  $\mu$ m, 4.6 mm i.d.  $\times$  250 mm) with MeOH/H<sub>2</sub>O (65:35) as eluent (flow rate 1 mL/min) to give 35 mg (73%) of compound **18** ( $t_R$  = 11.2 min).  $[\alpha]_{25}^D$  +4.93 (c 0.05, CH<sub>3</sub>OH). Selected <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  3.95 (2H, t,  $J$  = 6.4 Hz), 3.78 (1H, br s), 3.30 (1H, m ovl), 0.96 (3H, d ovl), 0.95 (3H, s), 0.70 (3H, s). HR ESIMS  $m/z$  457.2627 [M – Na], C<sub>24</sub>H<sub>41</sub>O<sub>6</sub>S requires 457.2624.

**5 $\beta$ -Cholan-3 $\alpha$ ,7 $\alpha$ ,24-tryl-3,7,24-sodium Trisulfate (32).** The triethylamine–sulfur trioxide complex (91 mg, 0.5 mmol) was added to a solution of triol **28** (40 mg, 0.1 mmol) in dry DMF (3 mL) under an argon atmosphere, and the mixture was stirred at 95 °C for 3 h. Most of the solvent was evaporated, and the residue was poured over a RP18 column to remove excess SO<sub>3</sub>·NEt<sub>3</sub>. The fraction eluted with H<sub>2</sub>O/MeOH, 9:1, gave a mixture that was further purified by HPLC on a Nucleodur 100-5 C18 (5  $\mu$ m, 4.6 mm i.d.  $\times$  250 mm) with MeOH/H<sub>2</sub>O (35:65) as eluent (flow rate 1 mL/min) to give 20 (44 mg,  $t_R$  = 8 min, 65%).  $[\alpha]_{25}^D$  +4.78 (c 0.12, CH<sub>3</sub>OH). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  4.45 (1H, br s), 4.15 (1H, m), 3.95 (2H, t,  $J$  = 6.5 Hz), 2.35 (1H, d,  $J$  = 12.5 Hz), 2.3 (1H, d,  $J$  = 14.0 Hz), 0.97 (3H, d ovl), 0.97 (3H, s), 0.70 (3H, s). HR ESIMS  $m/z$  661.1395 [M – Na], C<sub>24</sub>H<sub>39</sub>Na<sub>2</sub>O<sub>12</sub>S<sub>3</sub> requires 661.1399. The same sequence of reactions and purification procedure was carried out on 6 $\alpha$ -ethylchenodeoxycholic acid (**27**) to give a mixture of compounds **31** and **33**. Purification by HPLC on a Nucleodur 100-5 C18 (5  $\mu$ m, 4.6 mm i.d.  $\times$  250 mm) with MeOH/H<sub>2</sub>O (65:35) as eluent (flow rate 1 mL/min) gave compound **31** ( $t_R$  = 22.0 min) as a white solid (15 mg). Purification by HPLC on a Nucleodur 100-5 C18 (5  $\mu$ m, 4.6 mm i.d.  $\times$  250 mm) with MeOH/H<sub>2</sub>O (35:65) as eluent (flow rate 1 mL/min) gave compound **33** ( $t_R$  = 24.0 min) as a white solid (27 mg).

**6 $\alpha$ -Ethyl-3 $\alpha$ ,7 $\alpha$ -dihydroxy-5 $\beta$ -cholan-24-yl-24-sodium Sulfate (31).**  $[\alpha]_{25}^D$  +3.18 (c 0.51, CH<sub>3</sub>OH). Selected <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  3.96 (2H, t,  $J$  = 6.3 Hz), 3.65 (1H, br s), 3.30 (1H, m ovl), 0.97 (3H, d,  $J$  = 6.5 Hz), 0.92 (3H, s), 0.90 (3H, t,  $J$  = 7.0 Hz), 0.70 (3H, s). <sup>13</sup>C NMR (100 MHz, CD<sub>3</sub>OD):  $\delta$  72.2, 70.9, 69.6, 56.4, 50.5, 45.9, 42.5, 41.1, 39.9, 39.6, 35.7, 35.6 (2C), 35.4, 33.4, 32.0, 31.9, 29.3, 28.2, 23.6 (2C), 23.1, 20.5, 18.9, 12.2, 11.9. HR ESIMS  $m/z$  485.2935 [M – Na], C<sub>26</sub>H<sub>45</sub>O<sub>6</sub>S requires 485.2937.

**6 $\alpha$ -Ethyl-5 $\beta$ -cholan-3 $\alpha$ ,7 $\alpha$ ,24-tryl-3,7,24-sodium Trisulfate (33).**  $[\alpha]_{25}^D$  –1.9 (c 0.23, CH<sub>3</sub>OH). Selected <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  4.55 (1H, br s), 4.09 (1H, m), 3.96 (2H, t,  $J$  = 6.0 Hz), 0.97 (3H, d,  $J$  = 6.4 Hz), 0.92 (3H, s), 0.90 (3H, t,  $J$  = 7.0 Hz), 0.70 (3H, s). HR ESIMS  $m/z$  689.1715 [M – Na], C<sub>26</sub>H<sub>43</sub>Na<sub>2</sub>O<sub>12</sub>S<sub>3</sub> requires 689.1712.

**Bis-homochenodeoxycholan Sulfate Derivatives. 3 $\alpha$ ,7 $\alpha$ -Di(*tert*-butyldimethylsilyloxy)-5 $\beta$ -cholan-24-ol (34).** 2,6-Lutidine (290  $\mu$ L, 2.5 mmol) and *tert*-butyl dimethylsilyltrifluoromethanesulfonate (171  $\mu$ L, 0.75 mmol) were added at 0 °C to a solution of methyl ester **26** (100 mg, 0.25 mmol) in 10 mL of CH<sub>2</sub>Cl<sub>2</sub>. After the mixture was stirred for 2 h at 0 °C, the reaction was quenched

by addition of aqueous NaHSO<sub>4</sub> (1 M, 50 mL). The layers were separated, and the aqueous phase was extracted with CH<sub>2</sub>Cl<sub>2</sub> (3 × 50 mL). The combined organic layers were washed with NaHSO<sub>4</sub>, water, saturated aqueous NaHCO<sub>3</sub>, and brine. Purification by flash chromatography on silica gel using hexane/ethyl acetate, 99:1, and 0.5% of triethylamine as eluent gave the corresponding methyl ester (131 mg, 83%) as a clear, colorless oil. To a solution of methyl ester (100 mg, 0.16 mmol) in dry THF (5 mL) at 0 °C were added dry methanol (45 µL, 1.12 mmol) and LiBH<sub>4</sub> (560 µL, 2 M in THF, 1.12 mmol). The resulting mixture was stirred for 3 h at 0 °C. The mixture was quenched by addition of 1 M NaOH (320 µL) and then ethyl acetate. The organic phase was washed with water, dried (Na<sub>2</sub>SO<sub>4</sub>), and concentrated. Purification by silica gel (hexane/ethyl acetate 8:2) gave compound **34** as a white solid (97 mg, quantitative yield).  $[\alpha]_{25}^D +7.47$  (c 0.12, CHCl<sub>3</sub>). Selected <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 3.82 (1H, br s), 3.47 (2H, t, J = 6.5 Hz), 3.41 (1H, m), 0.93 (3H, d, J = 6.0 Hz), 0.88 (9H, s), 0.88 (3H, s ovl), 0.86 (9H, s), 0.65 (3H, s), 0.08 (6H, s), 0.04 (6H, s). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>): δ 72.8, 69.7, 63.4, 56.3, 50.1, 42.5, 42.0, 40.7, 40.6, 39.8, 36.0, 35.8, 35.2, 34.8, 32.4, 32.2, 31.2, 29.7 (3C), 28.3, 26.3 (3C), 26.0 (3C), 24.1, 23.0, 20.7, 18.8, 12.0, -2.2, -4.4, -4.5, -5.4. HRMS-ESI *m/z* 607.4946 [M + H]<sup>+</sup>, C<sub>36</sub>H<sub>71</sub>O<sub>3</sub>Si<sub>2</sub> requires 607.4942.

**Ethyl 3a, 7a-Di(tert-butyldimethylsilyloxy)-25,26-bis-homo-5β-chole-24-en-26-oate (35).** DMSO (4.18 mL, 0.75 mmol) was added dropwise for 5 min to a solution of oxalyl chloride (14.7 mL, 0.37 mmol) in dry dichloromethane (5 mL) at -78 °C under argon atmosphere. After 30 min, a solution of **34** (90 mg, 0.15 mmol) in dry CH<sub>2</sub>Cl<sub>2</sub> (2 mL) was added dropwise and the mixture was stirred at -78 °C. After 30 min, Et<sub>3</sub>N dry (6.83 mL, 0.75 mmol) was added dropwise to the solution and the mixture was allowed to warm to room temperature. After 1 h, the reaction was quenched by addition of aqueous NaHSO<sub>4</sub> (1 M, 50 mL). The layers were separated, and the aqueous phase was extracted with CH<sub>2</sub>Cl<sub>2</sub> (3 × 50 mL). The combined organic layers were washed with saturated aqueous NaHSO<sub>4</sub>, saturated aqueous NaHCO<sub>3</sub>, and brine. The organic phase was then dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated to give the corresponding aldehyde (90 mg, quantitative yield) as a colorless oil, which was used without any further purification. To a solution of aldehyde (0.15 mmol) in THF dry (10 mL) was added LiOH (250 mg, 10.5 mmol) and TEPA (triethylphosphonoacetate, 2.07 mL, 10.5 mmol). The reaction mixture was stirred for 45 min at room temperature and then quenched with water (10 mL). The mixture was then extracted with EtOAc (3 × 30 mL), and the organic phase was concentrated in vacuo. Flash chromatography (hexane/EtOAc, 99:1) afforded pure **35** (92 mg, 91% over two steps).  $[\alpha]_{25}^D +3.31$  (c 0.5, CHCl<sub>3</sub>). Selected <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 6.93 (1H, dt, J = 5.1, 15.6 Hz), 5.78 (1H, d, J = 15.6 Hz), 4.1 (2H, m), 3.77 (1H, br s), 3.38 (1H, m), 0.91 (3H, d ovl), 0.90 (9H, s), 0.90 (3H, s ovl), 0.87 (9H, s), 0.61 (3H, s), 0.06 (6H, s), 0.03 (6H, s). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>): δ 178.5, 149.8, 120.8, 72.5, 69.5, 59.9, 55.8, 49.9, 42.3, 41.8, 40.5, 40.4, 39.5, 35.6 (2C), 35.0, 34.6, 34.3, 32.2, 30.9,

29.6, 29.1 (2C), 28.0, 26.0 (3C), 25.8 (3C), 23.8, 22.7, 20.5, 18.4, 14.2, 12.0, -2.4, -4.5, -4.6, -5.6. HRMS-ESI  $m/z$  675.5207  $[M + H]^+$ ,  $C_{40}H_{75}O_4Si_2$  requires 675.5204.

**Ethyl 25,26-Bis-homo-3 $\alpha$ ,7 $\alpha$ -dihydroxy-5 $\beta$ -cholan-26-oate (36).** Compound **35** (90 mg, 0.13 mmol) and THF dry (15 mL) were mixed and deoxygenated with flowing nitrogen for 5 min. The catalyst Pd(OH)<sub>2</sub> (5 mg, mmol), 20 wt % on carbon (Degussa type), was added. The mixture was transferred to a standard PARR apparatus and flushed with nitrogen and then with hydrogen several times. The apparatus was shaken under 50 psi of H<sub>2</sub>. After 2 h, the reaction was complete. The catalyst was filtered through Celite, and the recovered filtrate was concentrated under vacuum to afford 92 mg of ethyl ester in quantitative yield, which was subjected to the next step without any purification. To the ethyl ester dissolved in ethanol (30 mL) was added 1 mL of HCl, 37% v/v, and the mixture was stirred for 5 h at room temperature. At the end of reaction, silver carbonate was added to precipitate chloride. Then the reaction mixture was centrifuged and the supernatant was concentrated in vacuo to give the desired ethyl ester **36** (50 mg, 81% over two steps) as a colorless amorphous solid. An analytic sample was obtained by silica gel chromatography, eluting with CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5.  $[\alpha]_{25}^D$  +6.72 (c 0.25, CHCl<sub>3</sub>). Selected <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  4.1 (2H, q, J = 7.0 Hz), 3.8 (1H, br s), 3.38 (1H, m), 2.29 (2H, t, J = 6.7 Hz), 1.24 (3H, t, J = 7.0 Hz), 0.94 (3H, d, J = 6.6 Hz), 0.92 (3H, s), 0.68 (3H, s). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>):  $\delta$  175.4, 72.8, 69.0, 61.4, 57.4, 51.5, 43.6, 43.1, 41.0, 40.7, 40.3, 37.0, 36.7, 36.5, 35.8, 35.2, 34.8, 34.0, 31.3, 29.3, 26.6, 26.5, 24.6, 23.5, 21.8, 19.3, 14.6, 12.3. HRMS-ESI  $m/z$  449.3635  $[M + H]^+$ ,  $C_{28}H_{49}O_4$  requires 449.3631.

**25,26-Bis-homo-3 $\alpha$ ,7 $\alpha$ ,26-5 $\beta$ -cholantriol (37).** Ethyl ester **36** (40 mg, 0.09 mmol) was reduced with LiBH<sub>4</sub> (2 M in THF dry, 135  $\mu$ L, 0.27 mmol) and dry MeOH (11  $\mu$ L, 0.27 mmol) in dry THF at 0 °C for 5 h. The mixture was quenched by addition of 1 M NaOH solution (180  $\mu$ L), and then ethyl acetate was added. The organic phase was washed with water, dried (Na<sub>2</sub>SO<sub>4</sub>), and concentrated. Purification by silica gel (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 9:1) gave compound **37** as a white solid (28 mg, 77%).  $[\alpha]_{25}^D$  +15.0 (c 0.32, CH<sub>3</sub>OH). Selected <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  3.82 (1H, br s), 3.61 (2H, t, J = 6.6 Hz), 3.43 (1H, m), 0.89 (3H, d, J = 6.0 Hz), 0.88 (3H, s), 0.64 (3H, s). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>):  $\delta$  71.8, 68.4, 62.8, 56.0, 50.4, 42.5, 41.4, 39.6 (2C), 39.4, 35.8, 35.6, 35.2, 35.0, 34.5, 32.8 (2C), 30.5, 28.2, 26.1, 25.8, 23.6, 22.7, 20.5, 18.6, 11.6. HRMS-ESI  $m/z$  407.3523  $[M + H]^+$ ,  $C_{26}H_{47}O_4$  requires 407.3525.

**25,26-Bis-homo-3 $\alpha$ ,7 $\alpha$ -dihydroxy-5 $\beta$ -cholan-26-yl-26-sodium Sulfate (38) and 25,26-Bis-homo-5 $\beta$ -cholan-3 $\alpha$ , 7 $\alpha$ , 26-tryl-3,7,26-sodium Trisulfate (39).** The triethylamine-sulfur trioxide complex (132 mg, 0.7 mmol) was added to a solution of triol **37** (28 mg, 0.07 mmol) in DMF dry (3 mL) under an argon atmosphere, and the mixture was stirred at 95 °C for 3 h. Most of the solvent was evaporated, and the residue was poured over a RP18 column to remove excess SO<sub>3</sub>·NEt<sub>3</sub>. The fraction eluted with H<sub>2</sub>O/MeOH 1:1 gave a mixture that was

further purified by HPLC on a Nucleodur 100-5 C18 (5  $\mu$ m, 4.6 mm i.d.  $\times$  250 mm) with MeOH/H<sub>2</sub>O (65:35) as eluent (flow rate 1 mL/min) to give 4 mg of compound **38**. The fraction eluted with H<sub>2</sub>O/MeOH (9:1) gave a mixture that was further purified by HPLC (Nucleodur 100-5 C18, 5  $\mu$ m, 4.6 mm i.d.  $\times$  250 mm), with MeOH/H<sub>2</sub>O (35:65) as eluent (flow rate 1 mL/min), to give 33 mg of compound **39**.

**25,26-Bis-homo-3 $\alpha$ ,7 $\alpha$ -dihydroxy-5 $\beta$ -cholan-26-yl-26-sodium Sulfate (38).**  $[\alpha]_{25}^D +9.13$  (c 0.05, CH<sub>3</sub>OH). Selected <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  3.98 (2H, t, J = 6.5 Hz), 3.79 (1H, m), 3.31 (1H, m, overlapped), 0.95 (3H, d, J = 6.8 Hz), 0.93 (3H, s), 0.63 (3H, s). HR ESIMS  $m/z$  485.2939 [M – Na], C<sub>26</sub>H<sub>45</sub>O<sub>6</sub>S requires 485.2937.

**25,26-Bis-homo-5 $\beta$ -cholan-3 $\alpha$ ,7 $\alpha$ ,26-tryl-3,7,26-sodium Trisulfate (39).**  $[\alpha]_{25}^D +1.35$  (c 1.13, CH<sub>3</sub>OH). Selected <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  4.45 (1H, br s), 4.13 (1H, m), 3.98 (2H, t, J = 6.5 Hz), 0.95 (3H, d, J = 6.6 Hz), 0.93 (3H, s), 0.69 (3H, s). HR ESIMS  $m/z$  689.1716 [M – Na], C<sub>26</sub>H<sub>43</sub>Na<sub>2</sub>O<sub>12</sub>S<sub>3</sub> requires 689.1712.

**Ursodeoxycholan Sulfate and Bis-homoursodeoxycholan Sulfate Derivatives.**

**3 $\alpha$ ,7 $\beta$ -Di(*tert*-butyldimethylsilyloxy)-5 $\beta$ -cholan-24-ol (41).** To a solution of methyl 3 $\alpha$ ,7 $\beta$ -dihydroxy-5 $\beta$ -cholan-24-oate (2 g, 5 mmol) in CH<sub>2</sub>Cl<sub>2</sub> dry (50 mL) was added at 0 °C 2,6-lutidine (49 mmol, 5.72 mL) and *tert*-butyl dimethylsilyltrifluoromethanesulfonate (15 mmol, 3.43 mL). After 2 h, the reaction was quenched by addition of aqueous NaHSO<sub>4</sub>, 1 M. The layers were separated, and the aqueous phase was extracted with CH<sub>2</sub>Cl<sub>2</sub> (3  $\times$  50 mL). The combined organic layers were washed with NaHSO<sub>4</sub>, saturated aqueous NaHCO<sub>3</sub>, and finally water. The organic phase was then dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated to give the corresponding methyl ester **40** (3.10 g, quantitative yield) as a colorless oil, which was used without any further purification. Dry methanol (600  $\mu$ L, 15 mmol) and LiBH<sub>4</sub> (7.50 mL, 2 M in THF, 15 mmol) were added to a solution of methyl ester (3.10 g, 5 mmol) in dry THF (15 mL) at 0 °C under argon, and the resulting mixture was stirred for 8 h at 0 °C. The mixture was quenched by addition of NaOH (1 M, 10 mL) and then allowed to warm to room temperature. Ethyl acetate was added, and the separated aqueous phase was extracted with ethyl acetate (3  $\times$  30 mL). The combined organic phases were washed with water, dried (Na<sub>2</sub>SO<sub>4</sub>), and concentrated. Purification by silica gel (n-hexane/ethyl acetate 99:1) gave compound **41** as a colorless oil (2.40 g, 80%).  $[\alpha]_{25}^D +21.7$  (c 0.49, CHCl<sub>3</sub>). Selected <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  3.61 (2H, m), 3.50 (2H, m), 0.88 (3H, d, overlapped), 0.88 (3H, s), 0.88 (9H, s), 0.83 (9H, s), 0.60 (3H, s), 0.003 (6H, s), –0.01 (6H, s). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>):  $\delta$  72.7, 72.5, 61.0, 55.5, 55.4, 44.0, 43.8, 42.8, 40.0, 39.0, 38.8, 38.0, 37.8, 35.4, 35.1, 34.1, 32.8, 31.8 (2C), 30.8, 28.7, 27.3, 26.4 (3C), 25.9 (3C), 23.5, 21.2, 19.0, 12.1, –2.8 (2C), –4.6 (2C). HRMS-ESI  $m/z$  607.4947 [M + H]<sup>+</sup>, C<sub>36</sub>H<sub>70</sub>O<sub>3</sub>Si<sub>2</sub> requires 607.4942.

**3 $\alpha$ ,7 $\beta$ -Dihydroxy-5 $\beta$ -cholan-24-yl-24-sodium Sulfate (42).** To a solution of compound **41** (100 mg, 0.16 mmol) in DMF dry (7 mL) was added 58 mg of

triethylamine–sulfur trioxide complex (0.32 mmol). The mixture was stirred at 95 °C for 4 h. Then the solution was concentrated under vacuum. To the solid dissolved in methanol (30 mL) was added three drops of HCl, 37% v/v, and the mixture was stirred for 3 h at room temperature. At the end of reaction, silver carbonate was added to precipitate chloride. Then the reaction mixture was centrifuged and the supernatant was concentrated in vacuo. The residue was poured over a RP18 column. The fraction eluted with H<sub>2</sub>O/MeOH (7:3) gave 53 mg (69% over two steps) of compound **42**.  $[\alpha]_{25}^D +15.3$  (c 0.97, CH<sub>3</sub>OH). Selected <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD): δ 3.96 (2H, t, J = 6.6 Hz), 3.48 (2H, m), 0.97 (3H, d, J = 6.4 Hz), 0.96 (3H, s), 0.71 (3H, s). <sup>13</sup>C NMR (100 MHz, CD<sub>3</sub>OD): δ 72.3 (2C), 69.6, 57.0, 56.4, 44.8, 43.0 (2C), 41.2, 40.7, 37.3 (2C), 36.7, 36.0, 35.5, 32.7, 30.6, 29.8, 27.8, 26.5, 23.3, 22.6, 20.0, 12.0. HR ESIMS *m/z* 457.2621 [M – Na], C<sub>24</sub>H<sub>41</sub>O<sub>6</sub>S requires 457.2624.

**5β-Cholan-3α,7β,24-tryl-3,7,24-sodium Trisulfate (43), 7β-Hydroxy-5β-cholan-3α,24-diyl-3,24-sodium Disulfate (44), and 7β,24-Dihydroxy-5β-cholan-3α-yl-3-sodium Sulfate (45).** To compound **41** (100 mg, 0.16 mmol) dissolved in methanol (30 mL) was added 500 μL of HCl, 37% v/v, and the mixture was stirred for 2 h at room temperature. At the end of reaction, silver carbonate was added to precipitate chloride. Then the reaction mixture was centrifuged and the supernatant was concentrated in vacuo to give 60 mg (96%) of the desired triol as colorless amorphous solids. An analytic sample was obtained by silica gel chromatography, eluting with CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5. The triethylamine–sulfur trioxide complex (118 mg, 0.65 mmol) was added to a solution of triol (50 mg, 0.13 mmol) in DMF dry (3 mL). The mixture was stirred at 95 °C for 2 h. Then the solution was concentrated under vacuum. The residue was poured over a RP18 column. The fraction eluted with H<sub>2</sub>O/MeOH (99:1) gave a mixture that was further purified by HPLC on a Nucleodur 100-5 C18 (5 μm, 10 mm i.d. × 250 mm) with MeOH/H<sub>2</sub>O (35:65) as eluent (flow rate 3 mL/min) to give 13 mg of **43** and 8 mg of **44**. The fraction eluted with H<sub>2</sub>O/MeOH (1:1) gave a mixture that, purified by HPLC on a Nucleodur 100-5 C18 (5 μm, 10 mm i.d. × 250 mm) with MeOH/H<sub>2</sub>O (65:35) as eluent (flow rate 3 mL/min), afforded 14 mg of **45**.

**5β-Cholan-3α,7β,24-tryl-3,7,24-sodium Trisulfate (43).**  $[\alpha]_{25}^D +20.6$  (c 0.08, CH<sub>3</sub>OH). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD): δ 4.28 (1H, m), 4.21 (1H, br s), 3.96 (2H, t, J = 6.7 Hz), 0.98 (3H, s), 0.97 (3H, d, J = 6.8 Hz), 0.71 (3H, s). HR ESIMS *m/z* 661.1395 [M – Na], C<sub>24</sub>H<sub>39</sub>Na<sub>2</sub>O<sub>12</sub>S<sub>3</sub> requires 661.1399.

**7β-Hydroxy-5β-cholan-3α,24-diyl-3,24-sodium Disulfate (44).**  $[\alpha]_{25}^D +35.7$  (c 0.11, CH<sub>3</sub>OH). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD): δ 4.23 (1H, m), 3.96 (2H, t, J = 6.7 Hz), 3.48 (1H, m), 0.98 (3H, s), 0.97 (3H, d, J = 6.8 Hz), 0.71 (3H, s). HR ESIMS *m/z* 559.2015 [M – Na], C<sub>24</sub>H<sub>40</sub>NaO<sub>9</sub>S<sub>2</sub> requires 559.2011.

**7β,24-Dihydroxy-5β-cholan-3α-yl-3-sodium Sulfate (45).**  $[\alpha]_{25}^D +37.4$  (c 0.08, CH<sub>3</sub>OH). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD): δ 4.23 (1H, m), 3.50 (2H, m), 3.48 (1H,

m), 0.97 (3H, s), 0.96 (3H, d, ovl), 0.71 (3H, s). HR ESIMS  $m/z$  457.2627 [ $M - Na$ ],  $C_{24}H_{41}O_6S$  requires 457.2624.

**Ethyl 3 $\alpha$ ,7 $\beta$ -Di(*tert*-butyldimethylsilyloxy)-25,26-bis-homo-5 $\beta$ -chol-24-en-26-oate (46).** Compound **46** (450 mg, 40% over two steps) was synthesized, starting from compound **41** (1.0 g, 1.6 mmol), by analogous procedures to those detailed above for compound **35**.  $[\alpha]_{25}^D +11.7$  (c 0.58,  $CHCl_3$ ).  $^1H$  NMR (400 MHz,  $CDCl_3$ ):  $\delta$  6.9 (1H, dt,  $J = 7.3, 15.5$  Hz), 5.78 (1H, d,  $J = 15.5$  Hz), 4.1 (2H, q,  $J = 7.3$  Hz), 3.64 (1H, m), 3.50 (1H, br s), 1.26 (3H, t,  $J = 7.3$  Hz), 0.91 (3H, d, ovl), 0.90 (3H, s), 0.86 (18H, s), 0.62 (3H, s), 0.04 (6H, s), 0.03 (6H, s).  $^{13}C$  NMR (100 MHz,  $CDCl_3$ ):  $\delta$  167.0, 150.0, 120.9, 72.7, 72.5, 60.0, 55.5, 54.9, 43.9, 43.7, 42.8, 39.9, 38.8, 37.9, 37.8, 35.3 (2C), 35.1, 34.3, 34.1, 30.9 (2C), 30.8, 28.9, 27.3, 26.4 (3C), 25.9 (3C), 23.5, 21.2, 18.6, 14.3, 12.1, -2.8, -3.4, -4.6 (2C). HRMS-ESI  $m/z$  675.5207 [ $M + H$ ] $^+$ ,  $C_{40}H_{75}O_4Si_2$  requires 675.5204.

**3 $\alpha$ ,7 $\beta$ -Di(*tert*-butyldimethylsilyloxy)-25,26-bis-homo-5 $\beta$ -cholan-26-ol (47).** A solution of compound **46** (400 mg, 0.6 mmol) in THF dry/MeOH dry (10 mL/10 mL, v/v) was hydrogenated in the presence of palladium, 5 wt %, on activated carbon (10 mg). The flask was evacuated and flushed first with argon and then with hydrogen. After 2 h, the reaction was complete. The catalyst was filtered through Celite, and the recovered filtrate was concentrated under vacuum to give the ethyl ester, which was used without any further purification. Dry methanol (73  $\mu$ L, 1.8 mmol) and  $LiBH_4$  (900  $\mu$ L, 2 M in THF, 1.8 mmol) were added to a solution of ethyl ester (400 mg, 0.6 mmol) in dry THF (15 mL) at 0  $^{\circ}C$  under argon, and the resulting mixture was stirred for 8 h at 0  $^{\circ}C$ . The mixture was quenched by addition of 1 M NaOH (1.2 mL) and then allowed to warm to room temperature. Ethyl acetate was added, and the separated aqueous phase was extracted with ethyl acetate (3  $\times$  30 mL). The combined organic phases were washed with water, dried ( $Na_2SO_4$ ), and concentrated. Purification by silica gel (n-hexane/ethyl acetate 99:1) gave compound **47** as a colorless oil (334 mg, 88% over two steps).  $[\alpha]_{25}^D +9.3$  (c 0.4,  $CHCl_3$ ).  $^1H$  NMR (400 MHz,  $CDCl_3$ ):  $\delta$  3.75 (2H, m), 3.65 (1H, m), 3.52 (1H, m), 0.92 (3H, s), 0.89 (3H, d, ovl), 0.89 (18H, s), 0.64 (3H, s), 0.05 (6H, s), 0.04 (6H, s).  $^{13}C$  NMR (100 MHz,  $CDCl_3$ ):  $\delta$  72.7, 72.5, 63.2, 55.5, 55.1, 43.9, 43.7, 42.8, 40.0, 38.8, 38.0, 37.8, 35.9 (2C), 35.6, 35.2, 34.1, 30.9 (2C), 30.8, 30.3, 29.7, 28.6, 27.3, 26.4 (3C), 25.9 (3C), 23.5, 21.2, 18.8, 12.1, -2.8, -3.3, -4.6 (2C). HRMS-ESI  $m/z$  635.5257 [ $M + H$ ] $^+$ ,  $C_{38}H_{75}O_3Si_2$  requires 635.5255.

**3 $\alpha$ ,7 $\beta$ -Dihydroxy-25,26-bis-homo-5 $\beta$ -cholan-26-yl-26-sodium Sulfate (48).** Compound **48** (180 mg, 75% over two steps) was synthesized, starting from compound **47** (300 mg, 0.47 mmol), by procedures analogous to those detailed above for compound **42**.  $[\alpha]_{25}^D +14.7$  (c 0.17,  $CH_3OH$ ). Selected  $^1H$  NMR (400 MHz,  $CD_3OD$ ):  $\delta$  3.98 (2H, t,  $J = 6.5$  Hz), 3.47 (2H, m), 0.96 (3H, s), 0.95 (3H, d,  $J = 6.0$  Hz), 0.71 (3H, s). HR ESIMS  $m/z$  485.2935 [ $M - Na$ ],  $C_{26}H_{45}O_6S$  requires 485.2937.

**25,26-Bis-homo-5 $\beta$ -cholan-3 $\alpha$ ,7 $\beta$ ,26-tryl-3,7,26-sodium Trisulfate (49)** and **25,26-Bis-homo-7 $\beta$ -hydroxy-5 $\beta$ -cholan-3 $\alpha$ ,26-diyl-3,26-sodium disulfate (50)**. The triethylamine–sulfur trioxide complex (214 mg, 1.2 mmol) was added to a solution of compound **48** (60 mg, 0.12 mmol) in DMF dry (3 mL). The mixture was stirred at 95 °C for 5 h. Then the solution was concentrated under vacuum. The residue was poured over a RP18 column. The fraction eluted with H<sub>2</sub>O/MeOH (7:3) gave a mixture that was further purified by HPLC, on a Nucleodur 100-5 C18 (5  $\mu$ m, 10 mm i.d.  $\times$  250 mm) with MeOH/H<sub>2</sub>O (35:65) as eluent (flow rate 3 mL/min), to give **49** and **50**, 8 and 5 mg, respectively.

**25,26-Bis-homo-5 $\beta$ -cholan-3 $\alpha$ ,7 $\beta$ ,26-tryl-3,7,26-sodium Trisulfate (49)**.  $[\alpha]_{25}^D$  +3.34 (c 0.34, CH<sub>3</sub>OH). Selected <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  4.30 (1H, m), 4.23 (1H, m), 3.98 (2H, t, J = 6.6 Hz), 0.98 (3H, s), 0.95 (3H, d, J = 6.1 Hz), 0.70 (3H, s). HR ESIMS  $m/z$  689.1715 [M – Na], C<sub>26</sub>H<sub>43</sub>Na<sub>2</sub>O<sub>12</sub>S<sub>3</sub> requires 689.1712.

**25,26-Bis-homo-7 $\beta$ -hydroxy-5 $\beta$ -cholan-3 $\alpha$ ,26-diyl-3,26-sodium Disulfate (50)**.  $[\alpha]_{25}^D$  +24.2 (c 0.11, CH<sub>3</sub>OH). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  4.23 (1H, m), 3.96 (2H, t, J = 6.3 Hz), 3.48 (1H, m), 0.97 (3H, s), 0.95 (3H, d, J = 6.6 Hz), 0.70 (3H, s). HR ESIMS  $m/z$  587.2329 [M – Na], C<sub>26</sub>H<sub>44</sub>NaO<sub>9</sub>S<sub>2</sub> requires 587.2324.

**Transactivation.** HepG2 cells were cultured at 37 °C in minimum essential medium with Earl's salts containing 10% fetal bovine serum (FBS), 1% l-glutamine, and 1% penicillin/streptomycin. The transfection experiments were performed using Fugene HD (Promega) according to manufactured specifications. Cells were plated in a 24-well plate at 5  $\times$  10<sup>4</sup> cells/well. For FXR mediated transactivation, HepG2 cells were transfected with 75 ng of pSG5-FXR, 75 ng of pSG5-RXR, 100 ng of pCMV- $\beta$ -galactosidase, and 250 ng of the reporter vector p(hsp27)-TK-LUC containing the FXR response element IR1 cloned from the promoter of heat shock protein 27 (hsp27). For GP-BAR1 mediated transactivation, HEK-293T cells were transfected with 200 ng of pGL4.29 (Promega), a reporter vector containing a cAMP response element (CRE) that drives the transcription of the luciferase reporter gene luc2P, with 100 ng of pCMVSPORT6-human GP-BAR1, and with 100 ng of pGL4.70 (Promega), a vector encoding the human Renilla gene. In control experiments HEK-293T cells were transfected only with vectors pGL4.29 and pGL4.70 to exclude any possibility that compounds could activate the CRE in a GP-BAR1 independent manner. At 24 h post-transfection, cells were stimulated 18 h with CDCA (**15**) at 10  $\mu$ M, **14** (1 and 10  $\mu$ M), 6-ECDCA and compounds **25**, **30–33**, **38**, **39**, **42–45**, and **48–50**, 10  $\mu$ M; in another experimental setting, cells were primed with the combination of CDCA (**15**) at 10  $\mu$ M and compounds **25**, **30**, **32**, **33**, **34**, **42–45**, and **48–50** (50  $\mu$ M). After treatments, cells were lysed in 100  $\mu$ L of lysis buffer (25 mM Tris-phosphate, pH 7.8; 2 mM DTT; 10% glycerol; 1% Triton X-100), and 20  $\mu$ L of cellular lysate was assayed for luciferase activity using the luciferase assay system (Promega). Luminescence was measured using Glomax 20/20 luminometer (Promega). For FXR mediated transactivation, luciferase activities

were normalized for transfection efficiencies by dividing the luciferase relative light units by  $\beta$ -galactosidase activity expressed from cells co-transfected with pCMV $\beta$ gal. For GP-BAR1 mediated transactivation, luciferase activities were normalized with Renilla activities. To realize dose–response curves for compound **31**, cells were transfected as described previously and then treated with increasing concentrations of **31** (0.01, 0.1, 1, and 10  $\mu$ M). Luciferase activities were assayed and normalized as described above in order to calculate EC<sub>50</sub>. GLUTAg cells were kindly donated by Dr. D. J. Drucker, Banting and Best Diabetes Centre, University of Toronto, Toronto General Hospital, 200 Elizabeth Street MBRW-4R402, Toronto, Canada M5G 2C4. FXR<sup>-/-</sup> mice were originally donated by Dr. F. Gonzalez, Laboratory of Metabolism, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, U.S., as previously reported.<sup>166</sup> Cells were plated in a 24-well plate at  $5 \times 10^4$  cells/well, and the transfection experiments were performed using Fugene HD (Promega) according to manufactured specifications.

**Real-Time PCR.** PCR was performed using Primers, and experimental conditions were as described previously.<sup>167</sup>

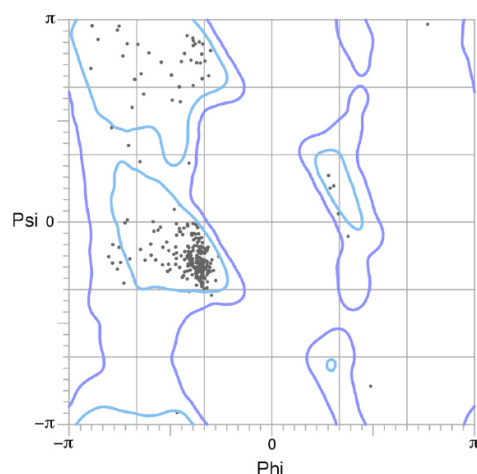
**cAMP and GLP-1 Measurement.** cAMP and GLP-1 release in GLUTAg cells were measured as described previously.<sup>61</sup>

**Homology Modeling.** The crystal structure of human adenosine A<sub>2A</sub> receptor (PDB code 2ydo)<sup>84</sup> was used as template to build the homology model of human GP-BAR1. First a multiple alignment of the sequence of different GPCRs (adenosine receptors, histamine H<sub>2</sub> receptor, sphingosine 1-phosphate receptors, cannabinoid receptors, muscarinic acetylcholine receptors, dopamine receptor, rhodopsin, etc.) was performed using ClustalW<sup>168, 169</sup> to identify the most conserved regions along the GP-BAR1 sequence (Swissprot ID code q8tdu6). Then we chose as template the agonist conformation of the adenosine A<sub>2A</sub> receptor that shows one of the highest sequence identity and similarity values, 20% and 55%, respectively (figure 28). The alignment between the model (GP-BAR1) and the template (A<sub>(2A)</sub>R-GL31) primary sequence was refined considering highly conserved amino acid residues among GPCRs. In particular, in the extracellular region, where the ligand binding site is located, the presence of the disulfide bond between Cys85 and Cys155 of the extracellular loop II (EL2) and the transmembrane helix III (TM3), respectively, was considered. Finally, using Modeller, version 9.11,<sup>170</sup> we built the tridimensional structure of the GP-BAR1 receptor. The homology model of GP-BAR1 was embedded in a POPC phospholipids bilayer to mimic the physiological environment and submitted first to minimization and then to a 100 ns long MD simulation to check its stability (see section on molecular dynamics methods). A geometric validation of the tridimensional structure of the receptor used in the docking calculations was

carried out using the MolProbity server (<http://molprobity.biochem.duke.edu>). The model shows about 94% of all residues in favored conformational regions and 100% in allowed ones according to the Ramachandran analysis and 0% of backbone bonds and angles with bad values (figure 29).

		<-----TM1----->	IL1	<-----TM2----->	
hAA2AR(2ydo)	1	SSVYITVELAIAVLAILGNVLCWAVWLNNSL	-QNV	TNYFVVSAAAADIL	49
hGP-BAR1	1	PKGALGLSLALASLIITANLLALGIAWDRRLRSP	PAGCF	FLSLLLAGLL	50
		.. : :.***: * * .*:~. :. : :. * . :. *~* *~*~			
		----->	EL1	<-----TM3----->	
hAA2AR(2ydo)	50	VGVLAIFFAIAISTGFCACHGCLFIACFVLVL	TASSIF	SLLAIAIDRYI	99
hGP-BAR1	51	TG-LALPTLPGLWNQSRRGYWSCLLVYLPNFS	FLSLLAN	LLLVHGERYM	99
		.~* ~*~* .: . . .***~: . * : ~*~* : ~*~*~			
		->	IL2	<-----TM4----->	
hAA2AR(2ydo)	100	AIRIPLRYNGLVTGTRAKGIIAICWVLSFAIGL	TPMLGWN	NCGQPKGKA	149
hGP-BAR1	100	AVLRPLQPPG--SI-R--LALLTWAGPLLFASL	PALGWN	-----H	135
		*: ~*~* : * : * : : * .~. :. : * ~*~*~			
		EL2	<-----TM5----->		
hAA2AR(2ydo)	150	HSQCGEGQVACLFEDVVPNMVMVYFNFFACVL	VPLLML	GVYLRIFLAA	199
hGP-BAR1	136	WTPGAN-----CSSQAIFPAPYL-YLEVYGL	LLPAVGAA	FLSVRLATA	179
		: ~.~* * : :.~* ~* : ~*~*~. : * : : ~*~*~*~			
		----->	IL3	<-----TM6----->	
hAA2AR(2ydo)	200	RRQLKQ-M--ES-----TLQKEVHAAKSLAI	IVGLFAL	CWLPLHI	236
hGP-BAR1	180	HRQLQDICRLERAVCRDEPSALARALTWRQARA	QAGAMLL	FGLCWGPYVA	229
		: ~*~*~*~* : * : : ~*~*~*~* : ~*~*~*~*~			
		----->	EL3	<-----TM7----->	
hAA2AR(2ydo)	237	INCFTFFCPDCSHAP-LWLMYLAIVLSHTNSV	VNPFII	YAIRIREFRQTFR	285
hGP-BAR1	230	TLLSVLAYEQRPPLGPGTLLSLLSLGSASAAA	VPVAMGL	GDQRYTAPWR	279
		: :~.~* : . : : ~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~*~			
hAA2AR(2ydo)	286	KIIRSHVLRQQEPFKAAAAENLYF			309
hGP-BAR1	280	AAAQ-----			283

**Figure 28.** Alignment of the primary sequence of the template structure, human adenosine A2A receptor (PDB code: 2YDO), and the target, human GP-BAR1, used in the homology modeling process. Only residues of the template structure resolved by X-ray are reported. The transmembrane helices, intracellular and extracellular loops are also shown. For sake of clarity the alignment numbering starts from 1, which corresponds to Ser6 and Pro14 in adenosine A<sub>2A</sub> receptor and GP-BAR1, respectively.

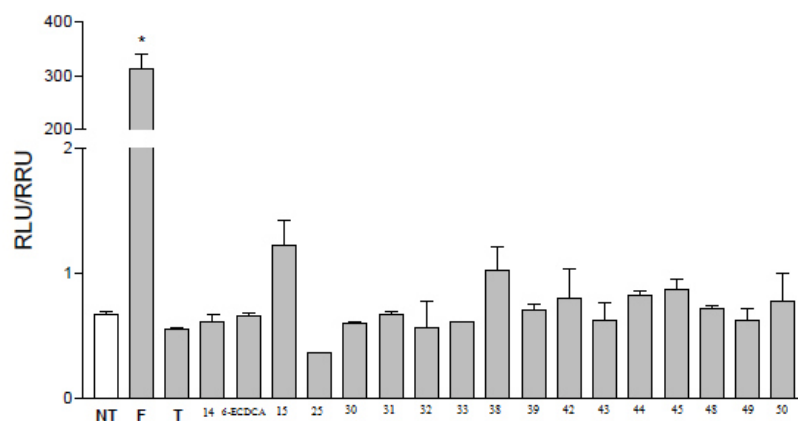


**Figure 29.** Ramachandran plot of the GP-BAR1 structure obtained from the MD simulation on the 14/GP-BAR1 complex, and used in the docking calculations. The cyan and violet regions represent the favored and allowed conformational regions, respectively, as defined by ProCheck.

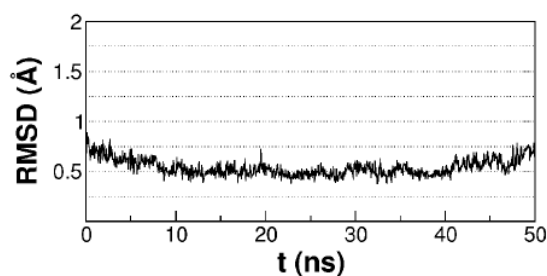
**Molecular Docking.** Molecular docking calculations in the rFXR-LBD X-ray structure (PDB code 1osv)<sup>91</sup> and in the three-dimensional model of hGP-BAR1 were carried out using the AutoDock4.2 software package.<sup>85</sup> Ligand tridimensional structures were generated with the Maestro build panel.<sup>171</sup> For each ligand, an extensive ring conformational sampling was performed with the OPLS-AA force field<sup>172</sup> and a 2.0 Å rmsd cutoff using MacroModel (version 9.9)<sup>173</sup> as implemented in Maestro 9.3.<sup>171</sup> All conformers were then refined using LigPrep<sup>174</sup> as implemented in Maestro 9.3. Protonation states at pH 7.0 were assigned using Epik.<sup>175</sup> Protein structure was prepared through the Protein Preparation Wizard through the graphical user interface of Maestro 9.3. Water molecules were removed, and hydrogen atoms were added and minimized using the OPLS-2005 force field.<sup>172</sup> Ligands and receptor structures were converted to AD4 format files using ADT, and the Gasteiger–Marsili partial charges were then assigned. Grid points of 70 × 70 × 70 for FXR and of 65 × 80 × 55 for GP-BAR1 with a 0.375 Å spacing were calculated around the binding cavity using AutoGrid4.2. Thus, 100 separate docking calculations were performed for each run. Each docking run consisted of 10 million energy evaluations using the Lamarckian genetic algorithm local search (GALS) method. Otherwise default docking parameters were applied. Docking conformations were clustered on the basis of their rmsd (tolerance = 2.0 Å) and were ranked based on the AutoDock scoring function.<sup>85</sup>

**Molecular Dynamics.** The same computational protocol was used to set up both the apo GP-BAR1-membrane and the GP-BAR1/1-membrane complexes. A 94 Å × 94 Å (in x and y axes) pre-equilibrated POPC phospholipid bilayer was initially generated using the membrane-builder tool of CHARMM-GUI.org (<http://www.charmm-gui.org>). In order to place the receptor into the bilayer, a hole was generated, and all lipids in close contact (<1 Å distance from any protein atoms) were deleted. Each complex was solvated and neutralized using the solvation and autoionize modules of VMD 1.9.1.<sup>176</sup> The TIP3 water model<sup>177</sup> was used to treat the solvent. The *ff99SBildn*,<sup>178,179</sup> *gaff*,<sup>180</sup> and *lipid11*<sup>181</sup> Amber force fields were used to parametrize the protein, the ligand, and the lipids, respectively. All the simulations were performed with the NAMD 2.8 MD code.<sup>182</sup> A 10 Å cutoff (switched at 8.0 Å) was used for atom–pair interactions. The long-range electrostatic interactions were computed by means of the particle mesh Ewald (PME) method<sup>183</sup> using a 1.0 Å grid spacing in periodic boundary conditions. The SHAKE algorithm was applied to constrain bonds involving hydrogen atoms, and thus an integration 2 fs time step interval could be used. Amber charges were applied to the proteins and water molecules, whereas the ligand charges were computed using the restrained electrostatic potential (RESP) fitting procedure.<sup>184</sup> The ESP was first calculated by means of the Gaussian 03 package<sup>185</sup> using a 6-31G\* basis set at Hartree–Fock level of theory, and then the RESP charges were obtained by a two-stage fitting procedure using Antechamber.<sup>186</sup> The system was equilibrated in the NPT ensemble using a target temperature and pressure of 300

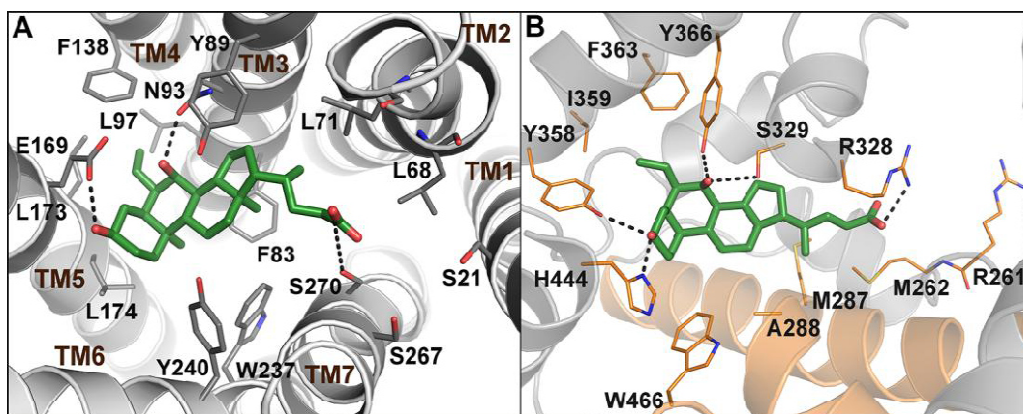
K and 1 atm, respectively. Harmonic constraints were applied to the protein and the ligand, which were gradually released along the equilibration process. Production runs were performed in NPT conditions at 1 atm and 300 K. All the residue labels were taken from crystal structure of rFXR-LBD with PDB code 1osv and the wild-type amino acidic sequence of human GP-BAR1. All figures were rendered using PyMOL (<http://www.pymol.org>). The tridimensional model of GP-BAR1 is available upon request.



**Figure 30.** Activation of cAMP responsive gene in absence of GP-BAR1. HEK293 cells were transfected with a reporter gene containing a cAMP responsive element in front of the luciferase gene. Twenty-four hour post transfection cells were stimulated with 10  $\mu$ M compounds **14-15**, **25**, **30-33**, **38**, **39**, **42-45**, **48-50**. Luciferase activity served as a measure of the activation of cAMP responsive element regardless of GP-BAR1. TLCA (T, 10  $\mu$ M) stimulates cAMP production in a GP-BAR1 dependent manner. Forskolin (F, 10  $\mu$ M) stimulated cAMP production regardless of GP-BAR1. Results are expressed as mean  $\pm$  standard error. \* $p < 0.05$  versus not treated cells.

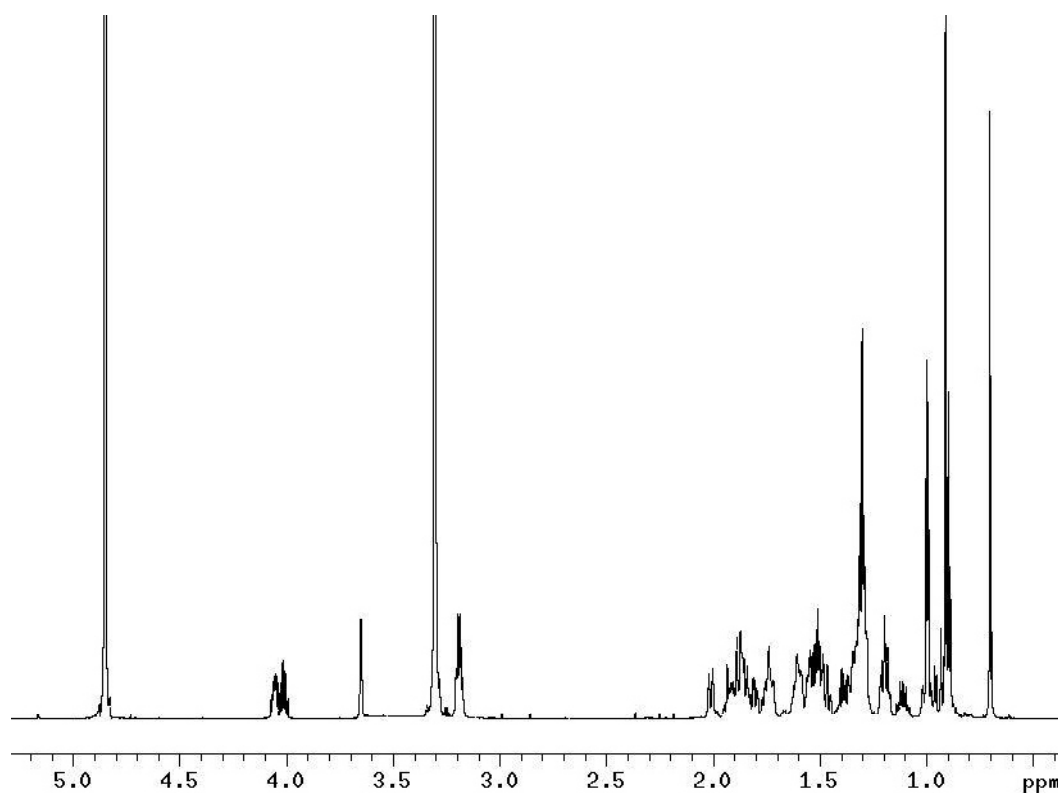


**Figure 31.** Rmsd plot of the GP-BAR1 backbone  $C\alpha$  atoms during the MD run with the receptor embedded in the membrane. The receptor was very stable during the whole simulation with a low averaged rmsd value ( $0.54 \pm 0.08$  Å).

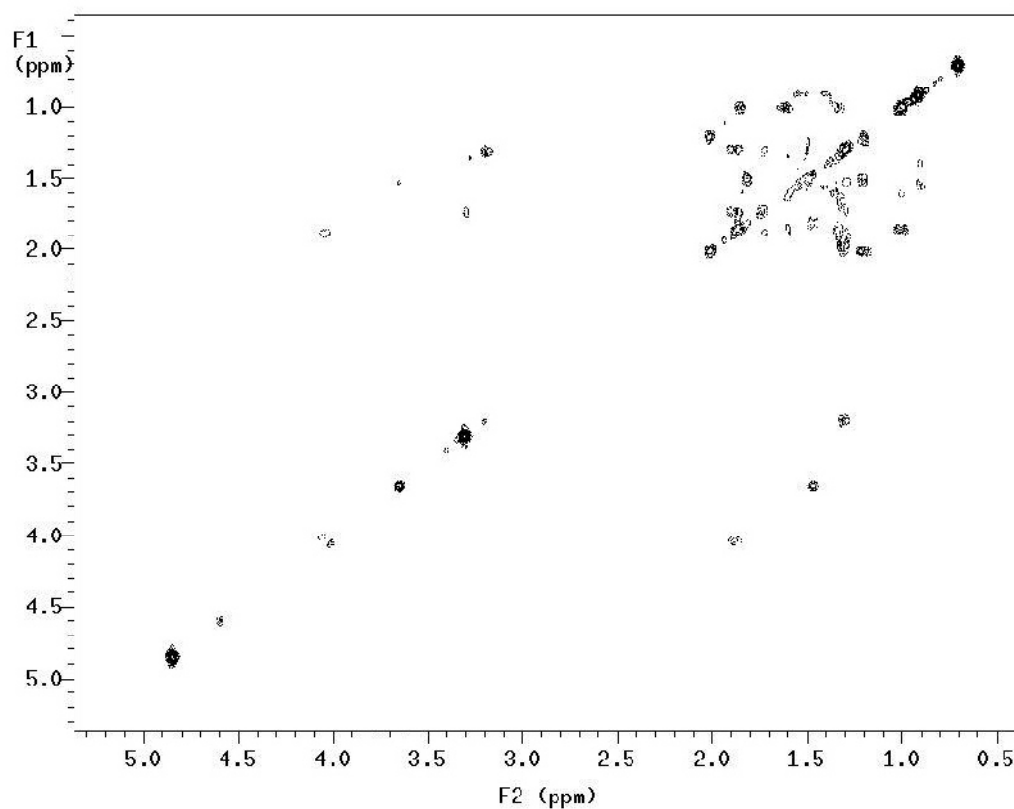


**Figure 32.** Binding modes of 6-ECDCA in the GP-BAR1 conformation obtained from MD simulations on the **14**/GP-BAR1 complex (A) and within the rFXR-LBD (PDB ID: 1osv) (B). The ligand is represented as green sticks. The receptors are shown as grey and orange (helices H3, H4 and H12 in FXR) cartoons. Amino acids involved in ligand binding are depicted as grey and orange sticks, in GP-BAR1 and in FXR, respectively. Hydrogens are omitted for clarity.

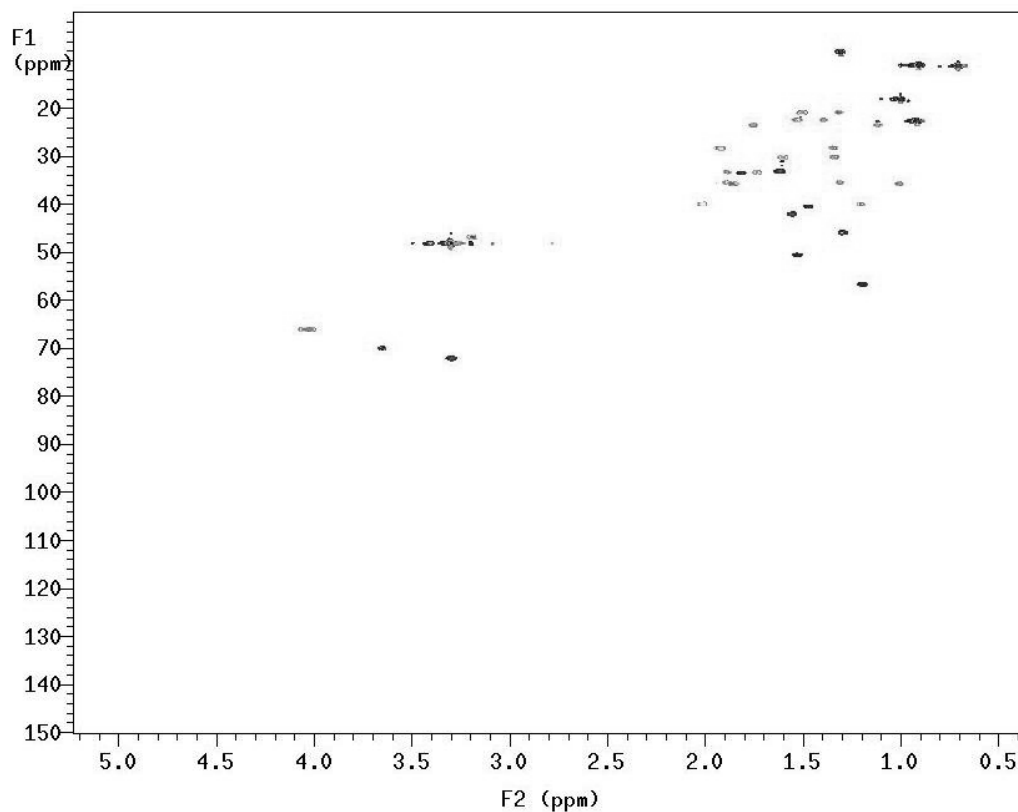
$^1\text{H}$ -NMR spectrum (700 MHz,  $\text{CD}_3\text{OD}$ ) of compound **14**



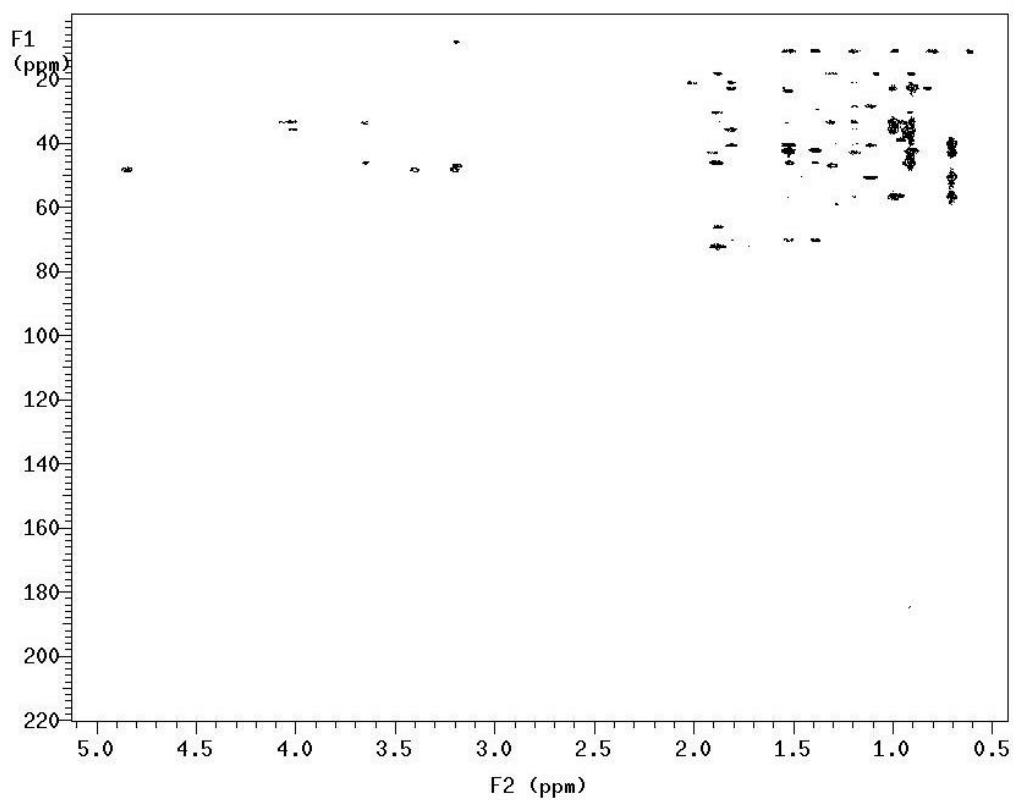
COSY spectrum (700 MHz,  $\text{CD}_3\text{OD}$ ) of compound **14**



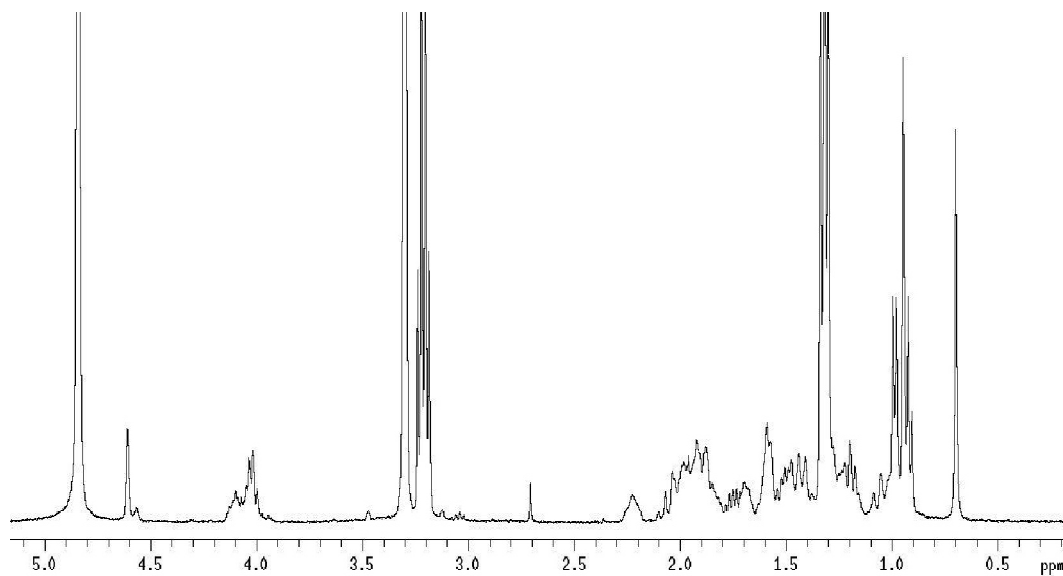
HSQC spectrum (700 MHz, CD<sub>3</sub>OD) of compound **14**



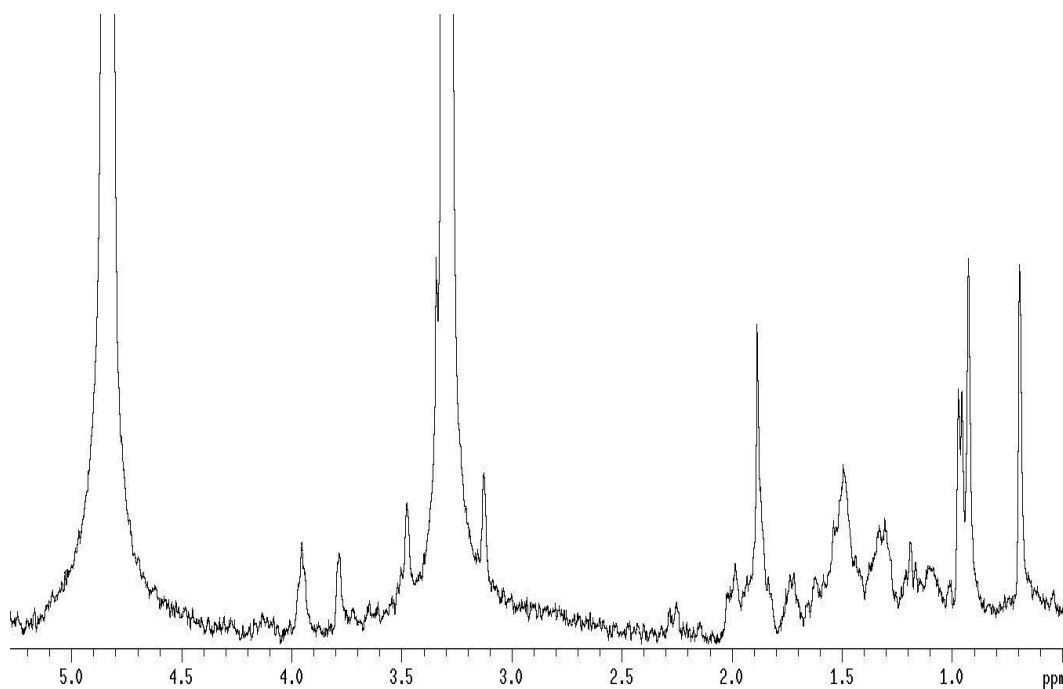
HMBC spectrum (700 MHz, CD<sub>3</sub>OD) of compound **14**



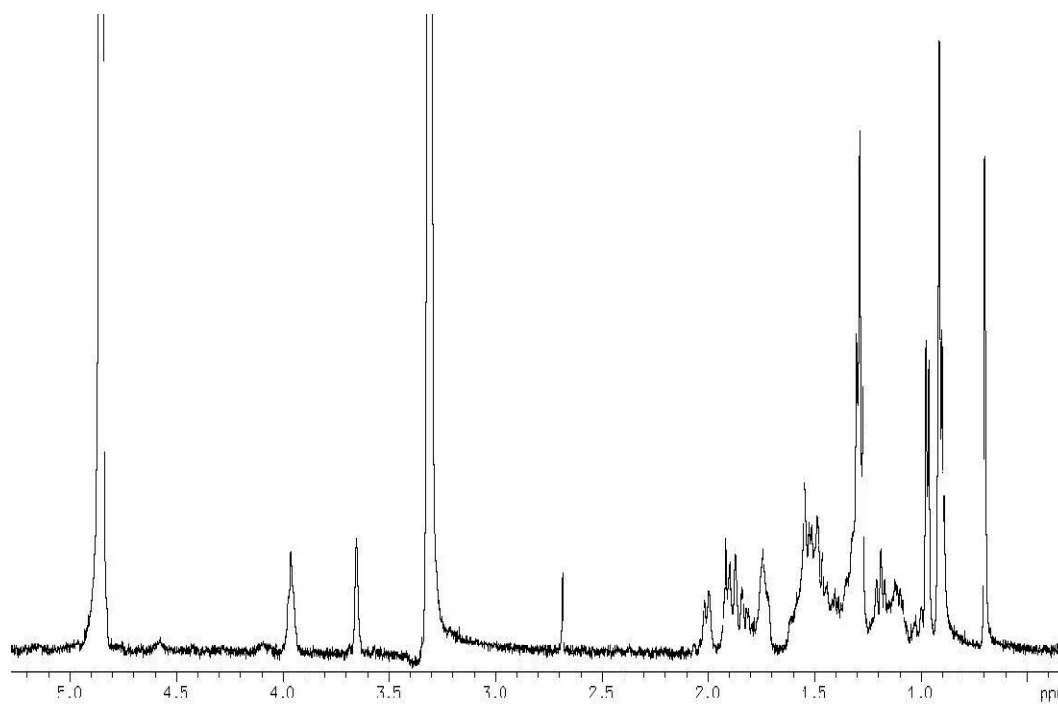
$^1\text{H}$ -NMR spectrum (700 MHz,  $\text{CD}_3\text{OD}$ ) of compound **25**



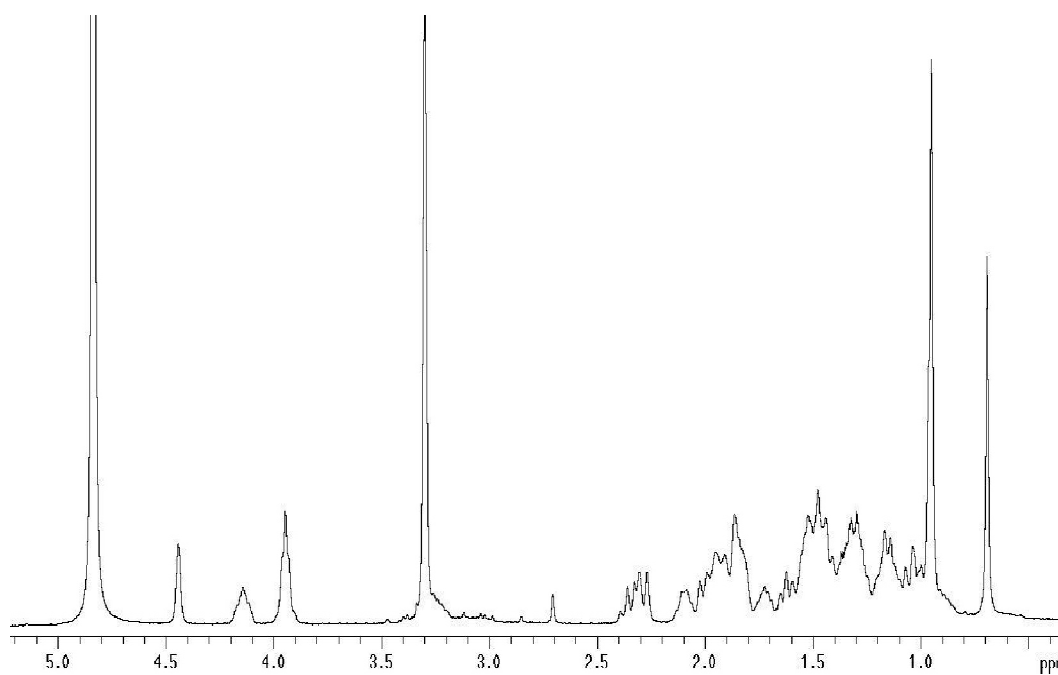
$^1\text{H}$ -NMR spectrum (400 MHz,  $\text{CD}_3\text{OD}$ ) of compound **30**



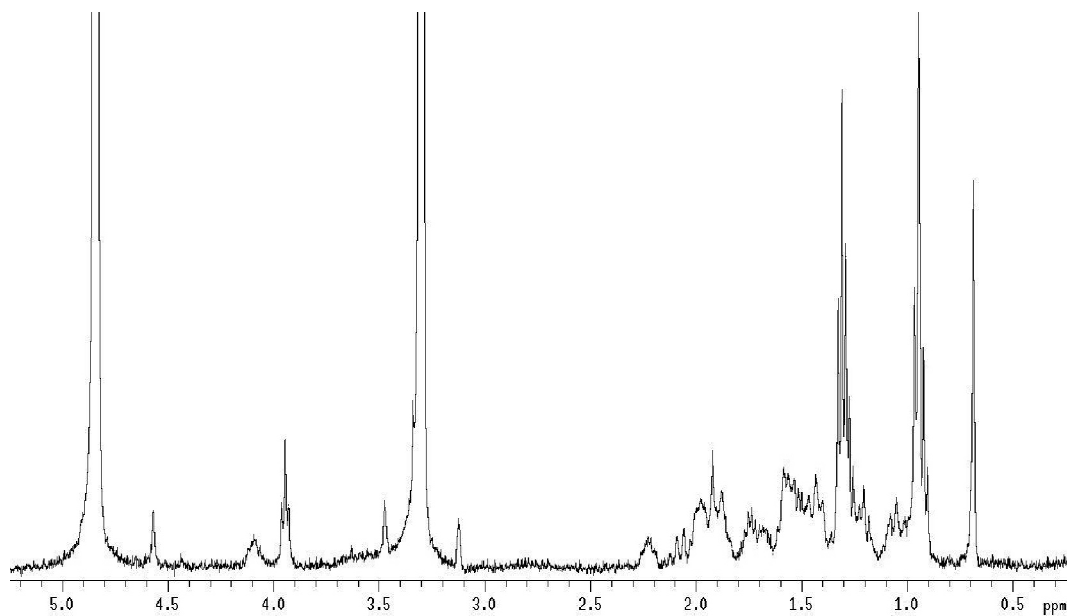
$^1\text{H}$ -NMR spectrum (400 MHz,  $\text{CD}_3\text{OD}$ ) of compound **31**



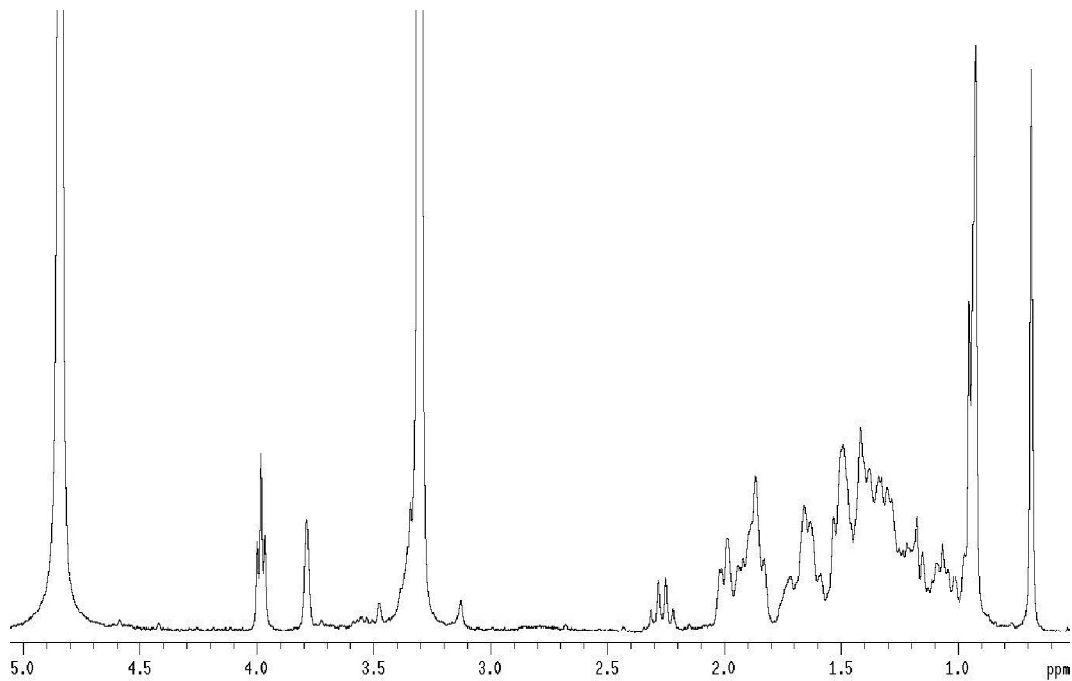
$^1\text{H}$ -NMR spectrum (400 MHz,  $\text{CD}_3\text{OD}$ ) of compound **32**



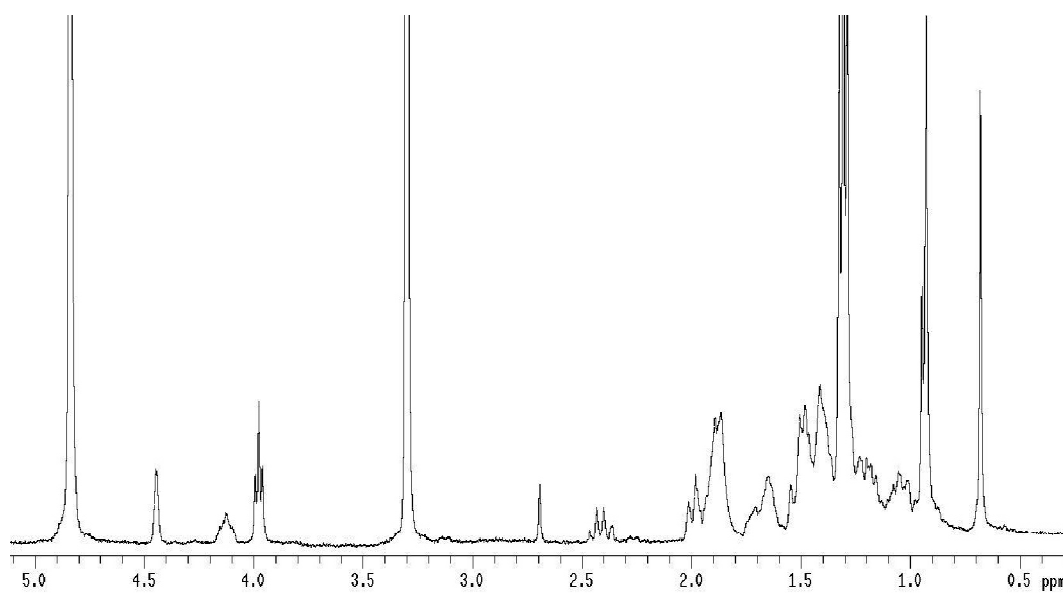
$^1\text{H}$ -NMR spectrum (400 MHz,  $\text{CD}_3\text{OD}$ ) of compound **33**



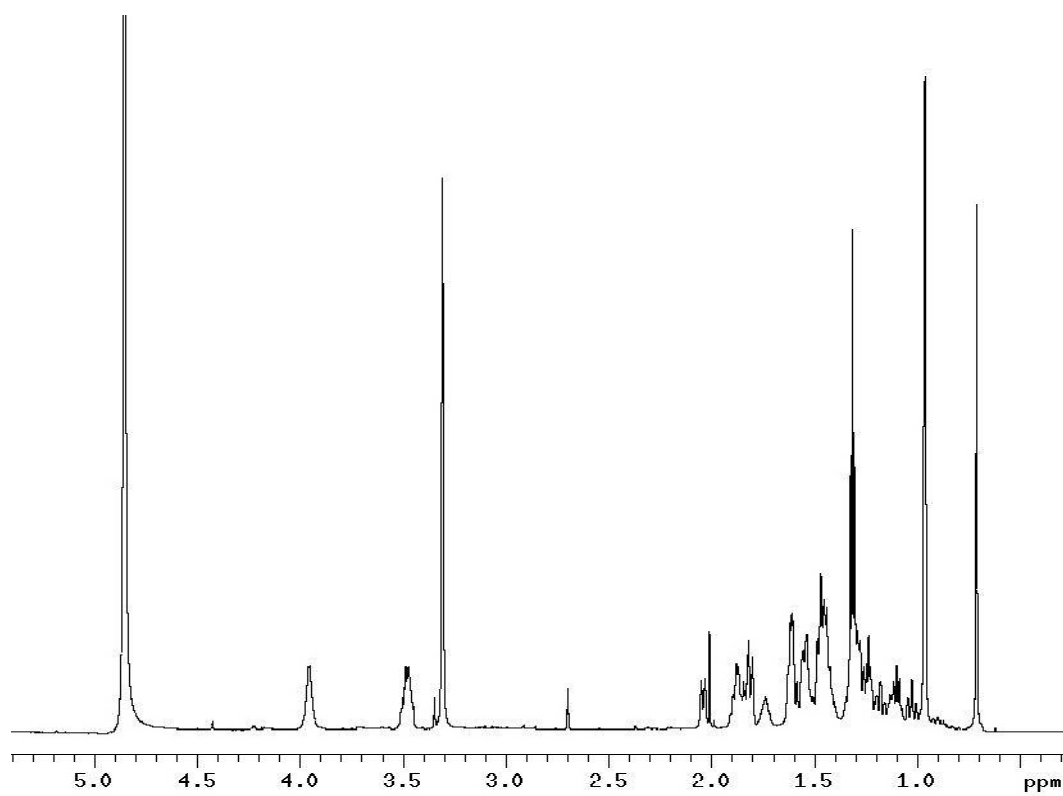
$^1\text{H}$ -NMR spectrum (400 MHz,  $\text{CD}_3\text{OD}$ ) of compound **38**



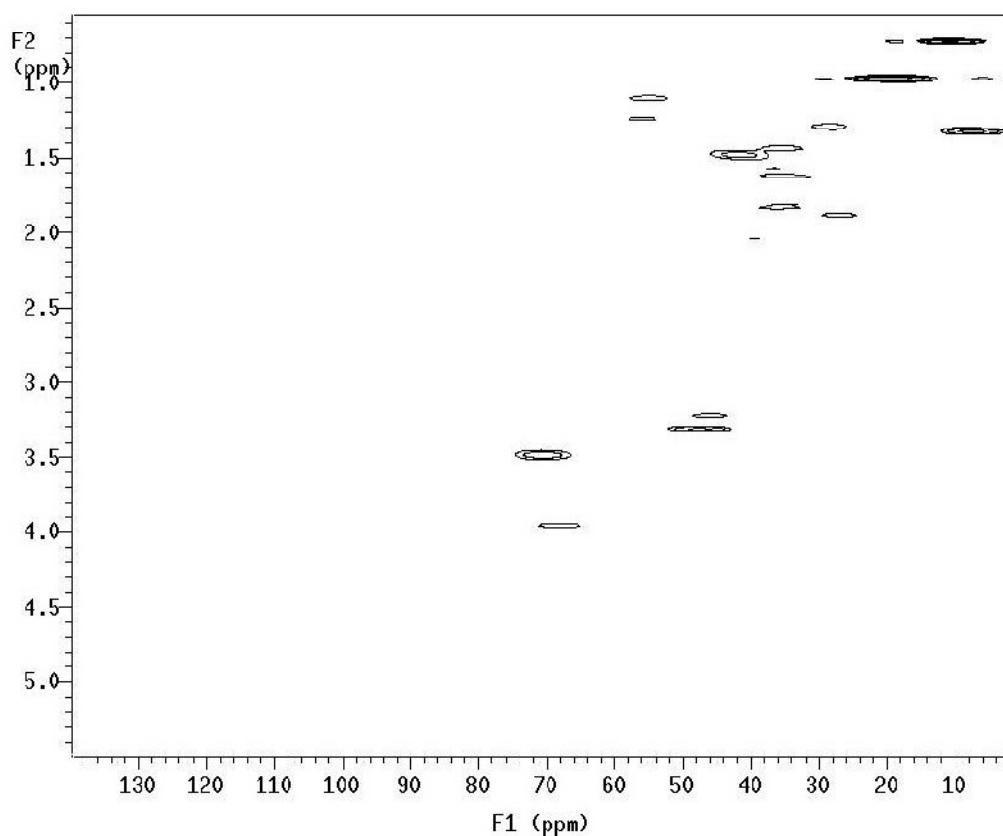
$^1\text{H}$ -NMR spectrum (400 MHz,  $\text{CD}_3\text{OD}$ ) of compound **39**



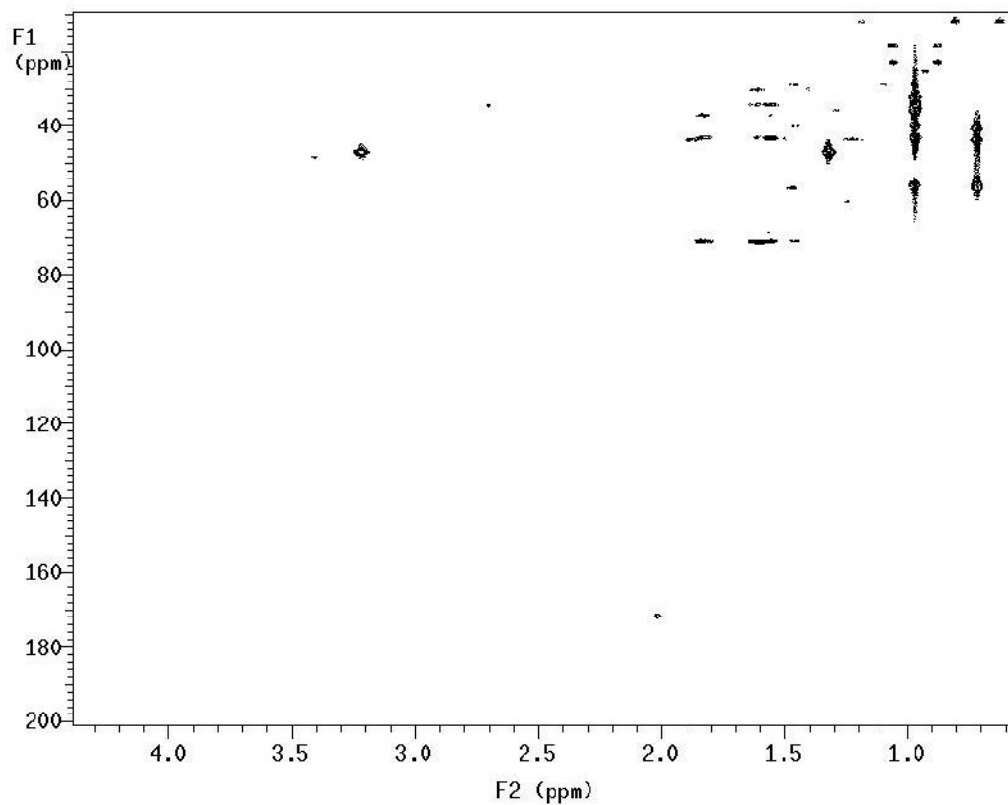
$^1\text{H}$ -NMR spectrum (700 MHz,  $\text{CD}_3\text{OD}$ ) of compound **42**



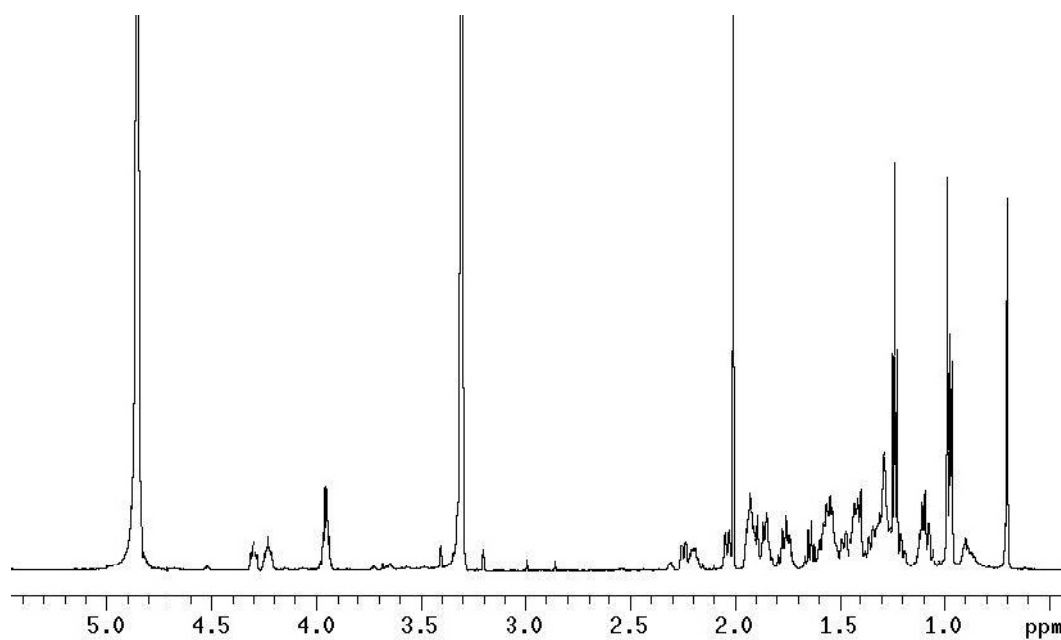
HSQC spectrum (700 MHz, CD<sub>3</sub>OD) of compound **42**



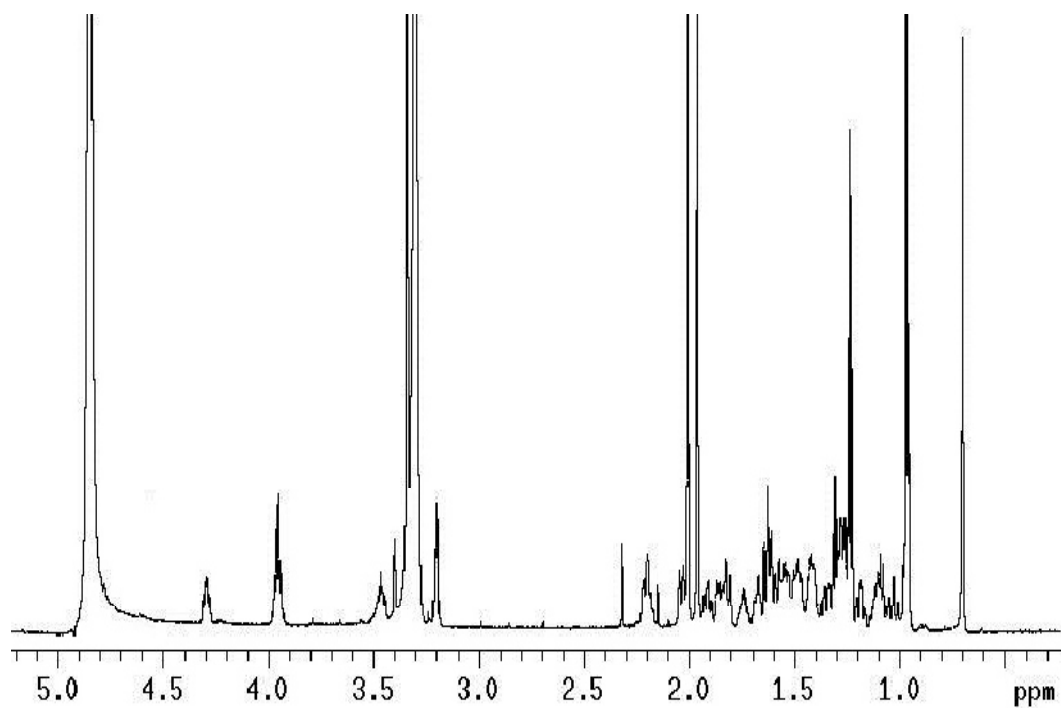
HMBC spectrum (700 MHz, CD<sub>3</sub>OD) of compound **42**



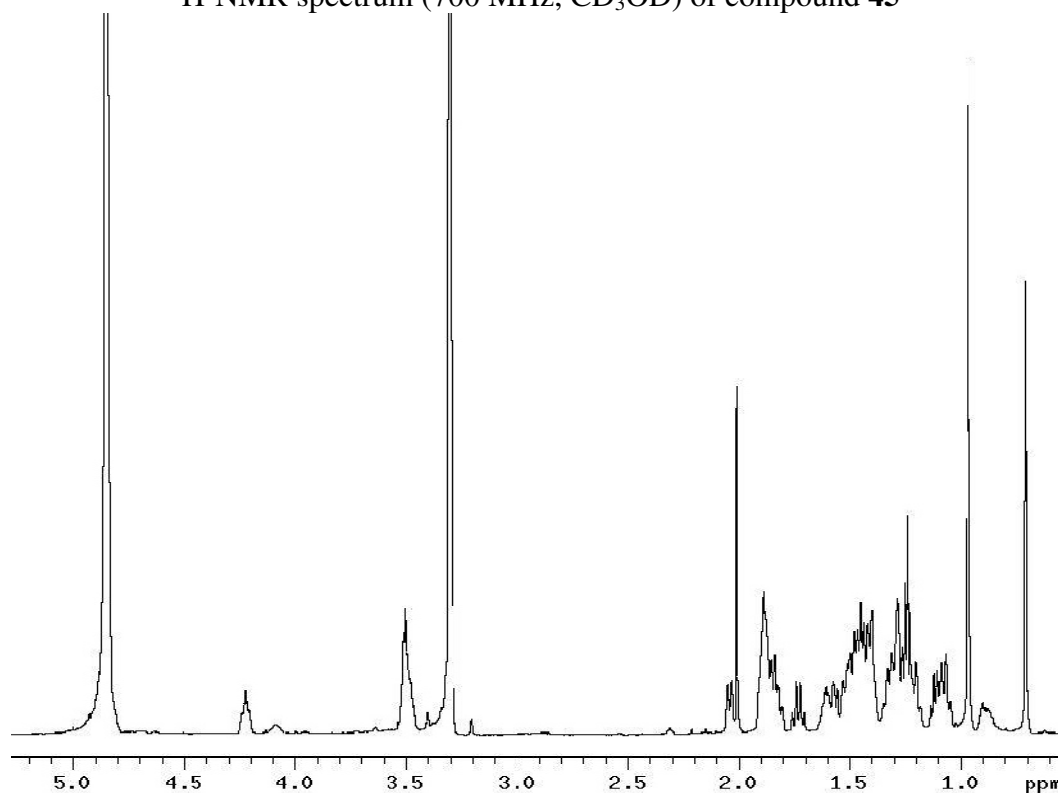
$^1\text{H}$ -NMR spectrum (700 MHz,  $\text{CD}_3\text{OD}$ ) of compound **43**



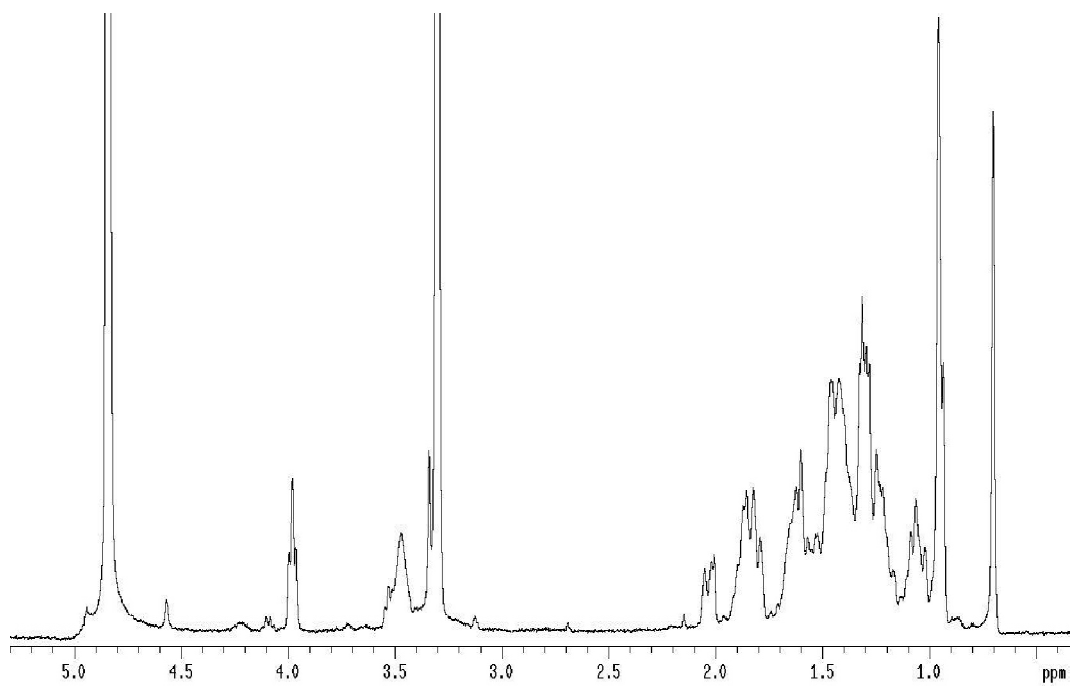
$^1\text{H}$ -NMR spectrum (700 MHz,  $\text{CD}_3\text{OD}$ ) of compound **44**



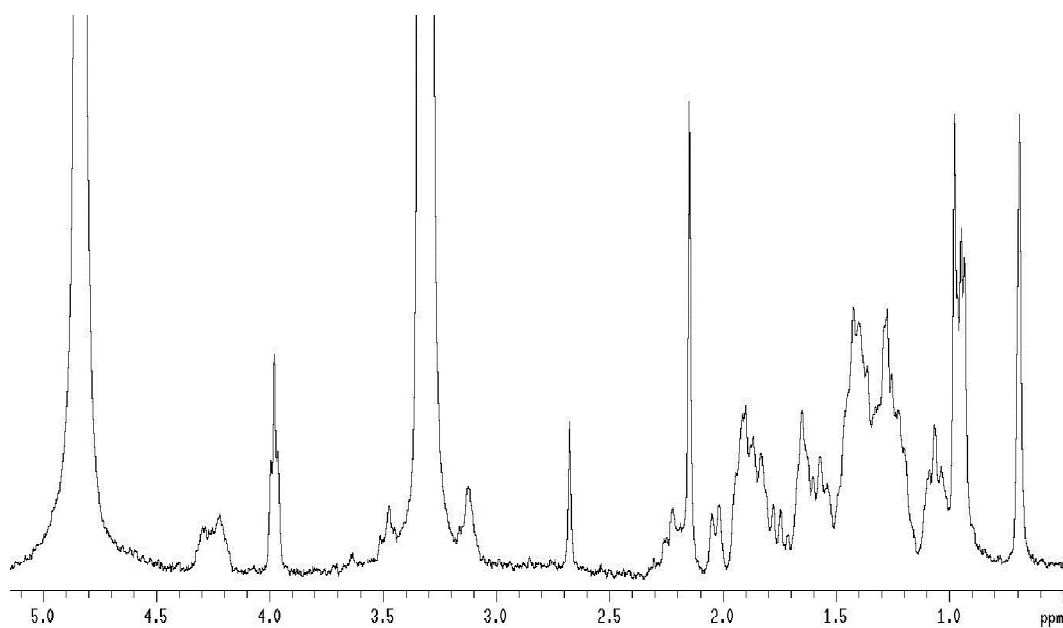
$^1\text{H}$ -NMR spectrum (700 MHz,  $\text{CD}_3\text{OD}$ ) of compound **45**



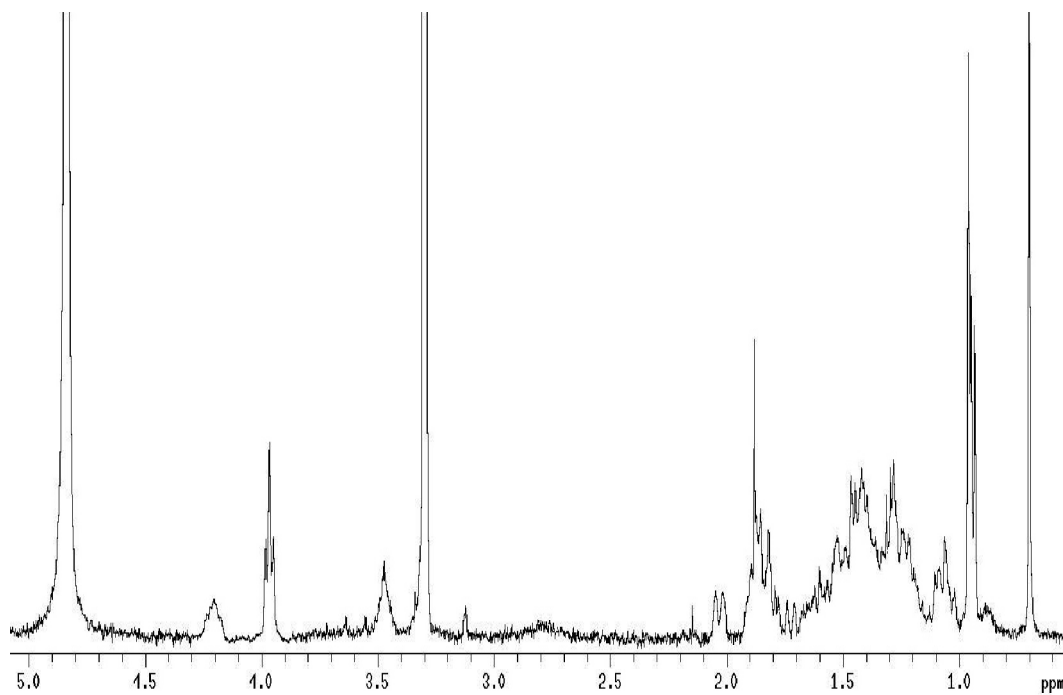
$^1\text{H}$ -NMR spectrum (400 MHz,  $\text{CD}_3\text{OD}$ ) of compound **48**



$^1\text{H}$ -NMR spectrum (400 MHz,  $\text{CD}_3\text{OD}$ ) of compound **49**



$^1\text{H}$ -NMR spectrum (400 MHz,  $\text{CD}_3\text{OD}$ ) of compound **50**



### III. EXPERIMENTAL PROCEDURES FOR XIAP-BIR3 INHIBITORS

**Protein Expression and purification.** The BIR3 domain of the human XIAP was prepared on the basis of previously reported data.<sup>187</sup> The BIR3 domain cDNA fragment (residues 253–347, confirmed by sequencing) was cloned into a pET15b vector (Novagen) and then transformed into *E. coli* BL21(DE3) Gold cells (Novagen). The transformed cells were transferred into LB medium and grown at 37°C. When optical density reached 0.6, 1 mM isopropyl 1-thio-D-galactopyranoside (IPTG) was added into the culture that continued to grow at 18°C overnight. The overexpressed protein was purified using Ni<sup>2+</sup>-affinity chromatography (HisTrap HP by GE Healthcare Life Sciences). Uniformly <sup>15</sup>N-labeled protein for NMR studies was prepared with a similar procedure by growing bacterial cells in a minimal medium containing <sup>15</sup>NH<sub>4</sub>Cl for <sup>15</sup>N labeling and a trace amount of Zn(OAc)<sub>2</sub> required to get a proper folded protein. The final protein buffer was MES 50mM, NaCl 100mM, TCEP 1mM, Zn(OAc)<sub>2</sub> 50uM, pH=6.

**Library.** The positional scanning library was produced by PEPSCAN PRESTO BV (The Netherlands). The library is tetrapeptidic with a fixed alanine at N-terminus and then arranged in a positional scanning fashion having one fragment fixed and all the other positions systematically populated by all the possible combinations. The A-XXX library is made up of 46 residues at each of the three diversity positions and includes L (11), D (12), and modified (23) amino acids (table 7) arranged in a total of 138 mixtures, each containing 2,116 compounds. All mixtures are N-terminal free and C-terminal amidated.

**NMR Spectroscopy.** NMR spectra were acquired on 600 and 700 MHz Bruker Avance spectrometer equipped with either TCI probe and z-shielded gradient coils or a TCI cryoprobe. All NMR data were processed and analyzed using TOPSPIN2.1 (Bruker Biospin, Billerica, MA, USA) and SPARKY3.1 (University of California, San Francisco, CA, USA). Compound binding was detected at 27°C by comparing the 2D-[<sup>15</sup>N, <sup>1</sup>H]-HSQC spectra of 25 μM XIAP-BIR3 in the absence and presence of mixtures at 500 μM as overall concentration.

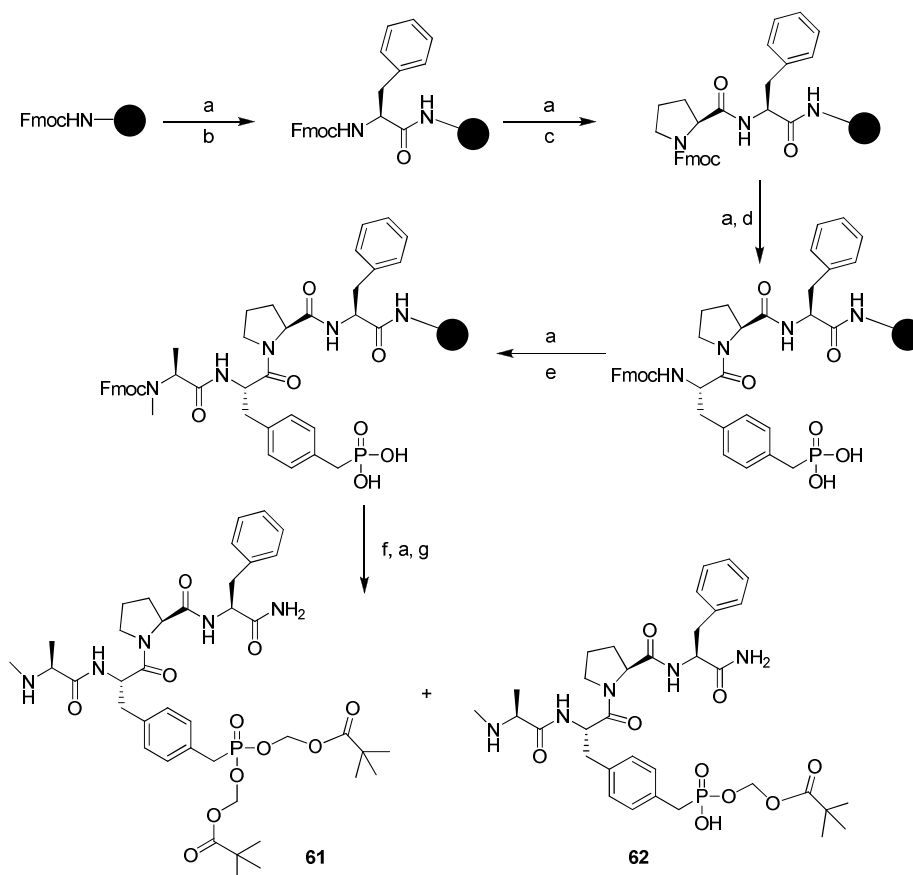
**Table 7.** Residues used to build the A-XXX-NH<sub>2</sub> library.

<b>1</b>	$\beta$ -Alanine	<b>U</b>	1-aminocyclopropane-1-carboxylic acid
<b>2</b>	$\beta$ -cyclohexyl-L-Ala	<b>a</b>	D-Ala
<b>3</b>	$\beta$ -cyclohexyl-D-Ala	<b>d</b>	D-Asp(OtBu)
<b>4</b>	$\beta$ -(3-pyridyl)-L-Ala	<b>e</b>	D-Glu(OtBu)
<b>5</b>	4-fluoro-L-phenylalanine	<b>f</b>	D-Phe
<b>6</b>	$\beta$ -cyclopropyl-L-Ala	<b>i</b>	D-Ile
<b>7</b>	4-fluoro-D-phenylalanine	<b>k</b>	D-Lys(Boc)
<b>8</b>	p-phenyl-L-Phe	<b>l</b>	D-Leu
<b>9</b>	D-homophenylalanine	<b>m</b>	D-Met
<b>B</b>	L-cyclopropylglycine	<b>p</b>	D-Pro
<b>J</b>	L-cyclopentyl-Gly	<b>v</b>	D-Val
<b>O</b>	L-cyclohexyl-Gly	<b>w</b>	D-Trp(Boc)
<b>P</b>	L-Pro	<b>y</b>	D-Tyr(tBu)
<b>Z</b>	L-homolys(Boc)	<b>Y</b>	L-Tyr(tBu)H
<b>b</b>	D-Nle	<b>W</b>	L-Trp(Boc)
<b>j</b>	L-HomoArg(Pbf)	<b>D</b>	L-Asp(OtBu)
<b>o</b>	O-Benzyl-L-Serine	<b>E</b>	L-Glu(OtBu)
<b>u</b>	$\beta$ -L-HomoTrp(Boc)	<b>K</b>	L-Lys(Boc)
<b>z</b>	L-Tyr(HPO <sub>3</sub> Bzl)	<b>R</b>	L-Arg(Pbf)
<b>c</b>	$\gamma$ -aminobutyric acid	<b>V</b>	L-Val
<b>g</b>	$\delta$ -aminovaleric acid	<b>I</b>	L-Ile
<b>h</b>	D-tetrahydroisoquinoline-1-carboxylic acid(d-Tiq)	<b>L</b>	L-Leu
<b>C</b>	L-Dap(Dnp)	<b>S</b>	L-Ser(tBu)

**Peptide Synthesis.** Unless otherwise indicated, all anhydrous solvents were commercially obtained and stored in Sure-seal bottles under nitrogen. All other reagents and solvents were purchased as the highest grade available and used without further purification. Mass spectral data were acquired on Shimadzu LCMS-2010EV for low resolution and on Bruker Daltonics Autoflex II MALDI TOF/TOF. Purity of all compounds was obtained in a HPLC Breeze from Waters Co. using an Atlantis T3 3 $\mu$ m 4.6x150 mm reverse phase column. The eluant was a linear gradient with a flow rate of 1 ml/min from 95% A and 5% B to 5% A and 95% B in 15 min followed by 5 min at 100% B (Solvent A: H<sub>2</sub>O with 0.1% TFA; Solvent B: acetonitrile with 0.1% TFA). The compounds were detected at  $\lambda$ =254 nm or 220 nm. Compounds **51**, **52** and **55-63** were synthesized manually using standard Fmoc peptide synthesis protocol with the Rink amide resin (0.61 meq/g). An example of the synthetic procedure applied to prodrugs **61** and **62** is reported in scheme 5. For each coupling reaction (for 0.1 mmol scale), 4 eq of Fmoc-amino acid, 6 eq of coupling agents (Oxyma Pure and DIC), 6 eq of DIEA in DMF dry (5 mL) were used. The coupling reaction was allowed to proceed for 2 h. Fmoc deprotection was performed by treating the resin-bound peptide with 20% piperidine in DMF (4 mL DMF, 1 mL piperidine) for 30 min in two times. After

each coupling reaction or Fmoc deprotection, the resin-bound peptide was washed with DMF (5 mL, 6 times) and DCM (5 mL, 6 times) respectively. For phosphotyrosine, mono benzyl protected phosphotyrosine (from Chem-impex) was used which was removed during cleavage reaction. For POM protection (compounds **61-62**): iodomethyl pivalate (6 eq), DMAP (10 mol%), DIEA (8 eq) in DMF dry were stirred at room temperature 24 h. The reaction was repeated twice using fresh same reagents and solvent. The phosphonic function in compounds **59** and **60** was introduced as 4-Phosphonomethyl-L-phenylalanine (L-Pmp) and D-Pmp respectively, without protection on the side chain. To obtain prodrug **63**, Fmoc-NH-dPEG<sub>6</sub>-COOH was coupled to the resin before the C-terminal amino acid using the same coupling mixture described previously. Then Fmoc group was removed using 20% piperidine in DMF, all in solid phase. The peptide was cleaved from the resin by treatment with a cleavage cocktail containing 30% TFA in DCM, 2 % water, 2% triisopropylsilane, 2% phenol for 3 h. All protecting groups also were removed during this cleavage reaction if any. TFA and DCM were removed under reduced pressure and the peptide was precipitated in diethyl ether, centrifuged, and washed with diethyl ether prior to drying in high vacuum. The crude peptide was purified by preparative reverse phase HPLC. The final compound was characterized by NMR and MALDI-Mass. All compounds were >95% purity.

Compounds **53** and **54** were synthesized on a 2-chlorotrityl resin (0.36 meq/g) charged with L-Pro using the same coupling and deprotecting conditions reported above. Peptides were cleaved from the resin maintaining the protecting groups by treatment with a solution of acetic acid, trifluoroethanol and dichloromethane (1:2:7) affording the free C-terminus. Then the three-peptides were coupled with the required ammine and the final compounds were deprotected and purified similarly to the other products.



**Scheme 5.** Synthetic scheme for compounds **61** and **62**.

Reagents and conditions: (a) 20% piperidine in DMF, rt, 30 min, repeated twice; (b) Fmoc-L-Phe-OH, Oxyma Pure, DIC, DIEA, DMF, rt, 2 h; (c) Fmoc-L-Pro-OH, Oxyma pure, DIC, DIEA, DMF, 2 h; (d) Fmoc-L-Phe(4-CH<sub>2</sub>PO<sub>3</sub>H<sub>2</sub>)-OH, Oxyma pure, DIC, DIEA, DMF, rt, 16 h; (e) Fmoc-L-Ala (N-Me)-OH (1.2 eq.) Oxyma pure, DIC, DIEA, DMF, rt, 2 h; (f) Iodomethyl pivalate, DIEA, DMAP (10 mol%), DMF, rt, 24 h, repeated twice; (g) 30% TFA in dichloromethane, phenol, TIPS, H<sub>2</sub>O, rt, 3 h. The black ball represents the Rink amide resin.

**Binding Constant Determination by ITC.** Isothermal titration calorimetry was performed on a VP-ITC calorimeter from Microcal (Northampton, MA, USA). Direct titrations were performed at 25°C in MES buffer supplemented with 10% DMSO. Experimental data were analyzed using Microcal Origin software provided by the ITC manufacturer (Microcal).

**Docking studies.** Molecular modeling studies were conducted on a Linux workstation and a 64 3.2-GHz CPUs Linux cluster. Docking studies were performed using the X-ray coordinates of the cIPA1/XIAP chimeric BIR3 domain (cXBIR3CS) in complex with GDC-0152 (PDB ID: 3UW4).<sup>135</sup> The crystal structure was extracted from the Protein Data Bank, and the complexed ligand was used to define the binding site for docking of the synthesized peptides. The genetic algorithm (GA) procedure in the GOLD docking software performed flexible docking of small molecules whereas the protein structure was

static.<sup>188,189,190</sup> For each compound, 20 solutions were generated and subsequently ranked according to GoldScore. Molecular surfaces were generated with MOLCAD<sup>190</sup> and docked structures analyzed with Sybyl (Tripos Inc., St. Louis, MO).

## IV. EXPERIMENTAL PROCEDURES FOR EPHA3 INHIBITORS

**Protein Expression and purification.** The Ligand Binding Domain of EphA3 (residues 29–201) was expressed as fusion protein with glutathione-S-transferase (GST). The cDNA encoding for the sequence GST-His<sub>10</sub>-TEV-LBD<sub>(29-201)</sub> was cloned into a modified pET32b vector (Novagen) and then transformed into SHuffle® T7 Express Competent *E. coli* cells (New England Biolabs). Transformed cells were transferred into Luria-Broth medium and grown at 30°C. When optical density reached 0.8, 1 mM isopropyl 1-thio-D-galactopyranoside (IPTG) was added to the culture that continued to grow at 18°C for 12 hours. The fusion protein was purified using Ni<sup>2+</sup>-affinity chromatography (HisTrap HP by GE Healthcare Life Sciences). The protein was dialyzed to remove imidazole and to buffer exchange to 25 mM CHES, 150 mM NaCl, pH= 9.0, buffer required for the following cleavage realized with a double mutant (L56V/S135G) TEV protease<sup>161</sup> (molar ratio 10:1) shaking overnight at 5°C. The LBD was released and purified again by reverse Ni<sup>2+</sup>-affinity chromatography to remove the GST-His<sub>10</sub> fragment. Uniformly <sup>15</sup>N-labeled protein for NMR studies was prepared with a similar procedure by growing bacterial cells in a minimal medium containing <sup>15</sup>NH<sub>4</sub>Cl for <sup>15</sup>N labeling. <sup>13</sup>C-Met-labeled protein was obtained by supplementing the LB medium with an excess of the labeled methionine (200 mg/L of culture) 5 minutes before the induction of protein expression.<sup>191,192</sup> The final protein buffer was 20 mM TRIS, 150mM NaCl, pH=8.0.

**NMR Spectroscopy.** NMR spectra were acquired as described in Experimental section III, pag. 107. Compound binding was detected at 27°C by comparing the 1D <sup>1</sup>H spectra of 10 μM EphA3-LBD in the absence and presence of compounds at mole ratio 1:10, respectively. For 2D-[<sup>13</sup>C, <sup>1</sup>H]-HSQC experiments we used 20 μM of <sup>13</sup>C-Met-labeled protein and compounds were added at a molar ratio 1:5. Dissociation equilibrium constants (K<sub>d</sub>) were derived by titrating the peptide into a 30 μM sample of <sup>15</sup>N-labeled EphA3-LBD and acquiring [<sup>15</sup>N-<sup>1</sup>H]-HSQC spectra at different ligand:protein ratios, starting from 1:0.5 till the protein was saturated. Titration analysis was done by fitting chemical shift data into the following quadratic equation:<sup>164,193</sup>

$$\Delta\delta_{obs} = \Delta\delta_{max} \frac{(K_d + [L]_0 + [P]_0) - \sqrt{(K_d + [L]_0 + [P]_0)^2 - 4[P]_0[L]_0}}{2[P]_0},$$

where  $\Delta\delta_{obs}$  is the observed chemical shift perturbation value at each titration point,  $\Delta\delta_{max}$  is the maximum chemical shift perturbation value of the fully complexed protein and  $[L]_0$  and  $[P]_0$  are the total concentrations of ligand and protein, respectively.

**Peptide Synthesis.** Peptide synthesis has been performed manually using standard Fmoc peptide synthesis protocol with the Rink amide or the 2-chlorotrityl resin. For general reaction conditions see scheme 5, pag. 110. All final compounds were

purified to >95% purity, as determined by a HPLC (Breeze HPLC system from Waters Co. using an Atlantis T3 reverse-phase column: 3 $\mu$ m; 4.6 $\times$ 150mm). Mass spectral data were acquired on Shimadzu LCMS-2010EV for low resolution and on Bruker Daltonics Autoflex II MALDI TOF/TOF.

## REFERENCES

- <sup>1</sup> Newman DJ, Cragg GM Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod.* **2012**;75(3):311-35.
- <sup>2</sup> Hughes J, Rees S, Kalindjian S, Philpott K. Principles of early drug discovery. *British Journal of Pharmacology.* **2011**;162(6):1239-1249.
- <sup>3</sup> David C. Swinney and Jason Anthony How were new medicines discovered? *Nature reviews* **2011** vol.10.
- <sup>4</sup> Jimeno J. M. A clinical armamentarium of marine derived anti-cancer compounds. *Anticancer Drugs*, **2002**,13, 15–19.
- <sup>5</sup> Pomponi S. A. The bioprocess-technological potential of the sea. *J. Biotechnol.*, **1999**, 70, 5–13.
- <sup>6</sup> Newman, D. J.; Cragg, G. M. Marine Natural Products and Related Compounds in Clinical and Advanced Preclinical Trials. *J. Nat. Prod.* **2004**, 67, 1216–1238.
- <sup>7</sup> Lichter et al. Worthen LW, ed. "Food-drugs from the sea. Proc: Aug 20–23, 1972.". *Marine Tech Soc.* **173**, 117–127.
- <sup>8</sup> Kupchan S. M., Britton R. W., Ziegler M. F., Sigel C. W. *J. Org. Chem.*, **1973**, 38, 178.
- <sup>9</sup> Jamali, F.; Mehahvar, R.; Pautto, F.M. *J. Pharm. Sci.*, **1989**, 78, 685
- <sup>10</sup> Karplus, M. *J. Chem. Phys.*, **1959**, 11.
- <sup>11</sup> Derome, A.E. Modern NMR techniques for chemistry research; Pergamon Press: Oxford, **1987**, 98.
- <sup>12</sup> Neuhaus, D.; Williamson, M. The Nuclear Overhauser Effect in structural and conformational analysis; VCH publisher, Inc.: New York Weinheim Cambridge, **1989**.
- <sup>13</sup> Ellis S.; Sharron L. The Culture of Soft Corals for the Marine Aquarium Trade. *Center for Tropical and Subtropical Aquaculture*, **1999**, 13.
- <sup>14</sup> Ellis, S. Farming Soft Corals for the Marine Aquarium Trade. *Center for Tropical and Subtropical Aquaculture*, **1999**, 140.
- <sup>15</sup> Wen-ting Chen, Yan Li, Yue-wei Guo. Terpenoids of *Sinularia* soft corals: chemistry and bioactivity *Acta Pharmaceutica Sinica B* **2012**;2(3):227–237
- <sup>16</sup> Coll, J. C.; Bowden, B. F.; Tapiolas, D. M.; Willis, R. H.; Djura, P.; Streamer, M.; Trott, L. Studies of australian soft corals—XXXV : The terpenoid chemistry of soft corals and its implications. *Tetrahedron*, **1985**, 41, 1085-1092.
- <sup>17</sup> Cheng, S.-Y.; Wen, Z.-H.; Chiou, S.-F.; Hsu, C.-H.; Wang, S.-K.; Dai, C.-F.; Chiang, M. Y.; Duh, C.-Y. Durumolides A–E, anti-inflammatory and antibacterial cembranolides from the soft coral *Lobophytum durum*. *Tetrahedron*, **2008**, 64, 9698-9704.
- <sup>18</sup> Yan, P.; Deng, Z.; van Ofwegen, L.; Proksch, P.; Lin, W. Lobophytones O–T, New Biscembranoids and Cembranoid from Soft Coral *Lobophytum pauciflorum*. *Mar. Drugs*, **2010**, 8, 2837-2848.
- <sup>19</sup> Coval, S. J.; Patton, R.W.; Petrin, J. M.; James, L.; Rothofsky, M. L.; Lin, S. L.; Patel, M.; Reed, J. K.; McPhil, A. T.; Bishop, W. R. A cembranolide diterpene farnesyl protein transferase inhibitor from the marine soft coral *Lobophytum cristagalli*. *Bioorg. Med. Chem. Lett.* **1996**, 6, 909–912.
- <sup>20</sup> Cheng, S.; Chuang, C.; Wang, S.; Wen, Z.; Chiou, S.; Hsu, C.; Dai, C.; Duh, C. Antiviral and anti-inflammatory diterpenoids from the soft coral *Sinularia gyrosa*. *J. Nat. Prod.* **2010**, 73, 1184–1187.

- 
- <sup>21</sup> Ahmed, Tai, Wen, Su, Wu, Hu, et al. A C-3 methylated isocembranoid and 10-oxocembranoids from a Formosan soft coral, *Sinularia grandilobata*. *J Nat Prod* **2008**;71:946–51.
- <sup>22</sup> Li, Zhang, Deng, Ofwegen, Proksch, Lin. Cytotoxic cembranoid diterpenes from soft coral *Sinularia gibberosa*. *J Nat Prod* **2005**;68:649–52.
- <sup>23</sup> Lin, Chen, Liaw, Chen, Kuo, Shen. Cembrane diterpenoids from the Taiwanese soft coral *Sinularia flexibilis*. *Tetrahedron* **2009**;65:9157–64.
- <sup>24</sup> Wen, Ding, Deng, Ofwegen, Proksch, Lin. Sinulaflexiolides A–K, cembrane-type diterpenoids from the Chinese soft coral *Sinularia flexibilis*. *J Nat Prod* **2008**;71:1133–40.
- <sup>25</sup> Zhang, Yan, Zhang, Lu, Su, Zeng, et al. Cytotoxic diterpenoids from the soft coral *Sinularia microclavata*. *J Nat Prod* **2005**;68:1087–9.
- <sup>26</sup> Ahmed, Su, Kuo, Sheu. Scabrolides E–G., three new norditerpenoids from the soft coral *Sinularia scabra*. *J Nat Prod* **2004**;67:2079–82.
- <sup>27</sup> Sheu JH, Ahmed AF, Shiue RT, Dai CF, Kuo YH Scabrolides A–D, four new norditerpenoids isolated from the soft coral *Sinularia scabra*. *J Nat Prod.* **2002**; 65(12): 1904–8.
- <sup>28</sup> Chang-Yih Duh, Shang-Kwei Wang, Min-Chi Chia, Michael Y. Chiang. A novel cytotoxic norditerpenoid from the Formosan soft coral *Sinularia inelegans*. *Tetrahedron Letters* **1999**; 40(33): 6033–6035.
- <sup>29</sup> Kazuo Iguchi, Kinzo Kajiyama, Yasuji Yamada. Yonarolide: a new marine norditerpenoid possessing a novel tricyclic skeleton, from the Okinawan soft coral of the genus, *Sinularia*. *Tetrahedron Letters* **1995**; 36(48): 8807–8808.
- <sup>30</sup> Atallah F Ahmed, Ru-Ting Shiue, Guey-Horng Wang, Chang-Feng Dai, Yao-Haur Kuo, Jyh-Horng Sheu Five novel norcembranoids from *Sinularia leptoclados* and *S. parva* *Tetrahedron* **2003**; 59(37): 7337–7344.
- <sup>31</sup> Sato, Fenical, Qi-tai, Clardy Norcembrene diterpenoids from pacific soft-corals of the genus *Sinularia*. *Tetrahedron* **1985**; 41(19): 4303–4308.
- <sup>32</sup> Gonzalez, Barrera, Yanes, Diaz, Rodriguez Perez. Chromenes and benzofurans from *Ageratina glechonophylla*. *Phytochemistry* **1989**; 28(9): 2520–2522.
- <sup>33</sup> Kamel HN, Fronczek FR, Khalifa SI, Slaterry M. Microbial transformation of 5-episinuleptolide. *Chem Pharm Bull* **2007**;55(4):537–40.
- <sup>34</sup> Festa C, Adel H, De Marino S, Lombardi V, Tilvi S, Nawrot DA, Zampella A, D'Souza L, D'Auria MV, Tammela P. Bioactive cembrane derivatives from the Indian Ocean soft coral, *Sinularia kavarattiensis*. *Mar Drugs*. **2014**;12(7):4045–68.
- <sup>35</sup> Shoji N., Umeyama A., Arihara S.A. Novel norditerpenoid from the Okinawan soft coral *Sinularia* sp. *J. Nat. Prod.* **1993**;56:1651–1653.
- <sup>36</sup> Parker. NMR determination of enantiomeric purity *Chem. Rev.*, **1991**, 91 (7), 1441–1457
- <sup>37</sup> Yi Li and Gerald Pattenden Novel macrocyclic and polycyclic norcembranoid diterpenes from *Sinularia* species of soft coral: Structural relationships and biosynthetic speculations *Nat. Prod. Rep.*, **2011**, 28, 429
- <sup>38</sup> P. A. Roethle and D. Trauner, *Nat. Prod. Rep.*, **2008**, 25, 298–317.
- <sup>39</sup> Tsai, Tsung-Chang; Wu, Yu-Jen; Su, Jui-Hsin; Lin, Wei-Tung; Lin, Yun-Sheng A new spatane diterpenoid from the cultured soft coral *Sinularia leptoclados* *Marine Drugs* **2013**;11:114–123

- <sup>40</sup> Takaki H, Koganemaru R, Iwakawa Y, Higuchi R, Miyamoto T. Inhibitory effect of norditerpenes on LPS-induced TNF- $\alpha$  production from the Okinawan soft coral, *Sinularia* sp. *Biol Pharm Bull.* **2003**;26(3):380-2.
- <sup>41</sup> Yen-Ju Tseng, Shang-Kwei Wang and Chang-Yih Duh Secosteroids and Norcembranoids from the Soft Coral *Sinularia nanolobata* Mar. *Drugs* **2013**, *11*(9), 3288-3296.
- <sup>42</sup> de Aguiar Vallim, T. Q.; Tarling, E. J.; Edwards, P. A. Pleiotropic roles of bile acids in metabolism. *Cell Metab.* **2013**, *17*, 657–669.
- <sup>43</sup> Makishima, M.; Okamoto, A. Y.; Repa, J. J. ; Tu, H.; Learned, R. M.; Luk, A.; Hull, M. V.; Lustig, K. D.; Mangelsdorf, D. J.; Shan, B. Identification of a nuclear receptor for bile acids *Science* **1999**, *284*, 1362– 13
- <sup>44</sup> Parks, D. J.; Blanchard, S. G.; Bledsoe, R. K.; Chandra, G.; Consler, T. G.; Kliewer, S. A.; Stimmel, J. B.; Willson, T. M.; Zavacki, A. M.; Moore, D. D.; Lehmann, J. M. Bile acids: natural ligands for an orphan nuclear receptor *Science* **1999**, *284*, 1365– 1368
- <sup>45</sup> Moschetta A, Bookout AL, Mangelsdorf Prevention of cholesterol gallstone disease by FXR agonists in a mouse model. *J. Nat Med.* **2004**;10(12):1352-8
- <sup>46</sup> Zollner G, Marschall HU, Wagner M, Trauner. Role of nuclear receptors in the adaptive response to bile acids and cholestasis: pathogenetic and therapeutic considerations. *Mol Pharm.* **2006**;3(3):231-51.
- <sup>47</sup> Fiorucci S, Antonelli E, Rizzo G, Renga B, Mencarelli A, Riccardi L, Orlandi S, Pellicciari R, Morelli A. *Gastroenterology.* **2004** ;127(5):1497-512.
- <sup>48</sup> Fickert P, Fuchsbichler A, Moustafa T, Wagner M, Zollner G, Halilbasic E, Stöger U, Arrese M, Pizarro M, Solís N, Carrasco G, Caligiuri A, Sombetzki M, Reisinger E, Tsybrovskyy O, Zatloukal K, Denk H, Jaeschke H, Pinzani M, Trauner M. Farnesoid X receptor critically determines the fibrotic response in mice but is expressed to a low extent in human hepatic stellate cells and periductal myofibroblasts. *Am J Pathol.* **2009**;175(6):2392-405.
- <sup>49</sup> Zhang, Y.; Lee, F. Y.; Barrera, G.; Lee, H.; Vales, C.; Gonzalez, F.J.; Willson, T. M.; Edwards, P. A. Activation of the nuclear receptor FXR improves hyperglycemia and hyperlipidemia in diabetic mice. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1006–1011.
- <sup>50</sup> Renga, B.; Mencarelli, A.; D’Amore, C.; Cipriani, S.; Baldelli, F.; Zampella, A.; Distrutti, E.; Fiorucci, S. Glucocorticoid receptor mediates the gluconeogenic activity of the farnesoid X receptor in the fasting condition. *FASEB J.* **2012**, *26*, 3021–3031.
- <sup>51</sup> Fu L, John LM, Adams SH, Yu XX, Tomlinson E, Renz M, Williams PM, Soriano R, Corpuz R, Moffat B, Vandlen R, Simmons L, Foster J, Stephan JP, Tsai SP, Stewart. Fibroblast growth factor 19 increases metabolic rate and reverses dietary and leptin-deficient diabetes *Endocrinology.* **2004**;145(6):2594-603
- <sup>52</sup> Wang, Y. D.; Chen, W. D.; Wang, M.; Yu, D.; Forman, B. M.; Huang, W. Farnesoid X receptor antagonizes nuclear factor  $\kappa$ B in hepatic inflammatory response. *Hepatology* **2008**, *48*, 1632–1643.
- <sup>53</sup> Baptissart, M.; Vega, A.; Maqdasy, S.; Caira, F.; Baron, S.; Lobaccaro, J. M.; Volle, D. H. Bile acids: from digestion to cancers. *Biochimie* **2013**, *95*, 504–517.

- 
- <sup>54</sup> Takeda, S.; Kadowaki, S.; Haga, T.; Takaesu, H.; Mitaku, S. Identification of G protein-coupled receptor genes from the human genome sequence *FEBS Lett.* **2002**, 520, 97– 101
- <sup>55</sup> Maruyama, T.; Miyamoto, Y.; Nakamura, T.; Tamai, Y.; Okada, H.; Sugiyama, E.; Nakamura, T.; Itadani, H.; Tanaka, K. Identification of membrane type receptor for bile acids (M-BAR) *Biochem. Biophys. Res. Commun.* **2002**, 298, 714– 719
- <sup>56</sup> Kawamata, Y.; Fujii, R.; Hosoya, M.; Harada, M.; Yoshida, H.; Miwa, M.; Fukusumi, S.; Habata, Y.; Itoh, T.; Shintani, Y.; Hinuma, S.; Fujisawa, Y.; Fujino, M. A G protein-coupled receptor responsive to bile acids *J. Biol. Chem.* **2003**, 278, 9435– 9440
- <sup>57</sup> Watanabe, M.; Houten, S. M.; Matak, C.; Christoffolete, M. A.; Kim, B. W.; Sato, H.; Messaddeq, N.; Harney, J. W.; Ezaki, O.; Kodama, T.; Schoonjans, K.; Bianco, A. C.; Auwerx, J. Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation *Nature* **2006**, 439, 484– 489
- <sup>58</sup> Thomas, C.; Gioiello, A.; Noriega, L.; Strehle, A.; Oury, J.; Rizzo, G.; Macchiarulo, A.; Yamamoto, H.; Matak, C.; Pruzanski, M.; Pellicciari, R.; Auwerx, J.; Schoonjans, K. TGR5-mediated bile acid sensing controls glucose homeostasis *Cell Metab.* **2009**, 10, 167– 177
- <sup>59</sup> Kawamata, Y.; Fujii, R.; Hosoya, M.; Harada, M.; Yoshida, H.; Miwa, M.; Fukusumi, S.; Habata, Y.; Itoh, T.; Shintani, Y.; Hinuma, S.; Fujisawa, Y.; Fujino, M. A G Protein-Coupled Receptor Responsive to Bile Acids *J. Biol. Chem.* **2003**, 278, 9435– 9440
- <sup>60</sup> Cipriani, S.; Mencarelli, A.; Chini, M. G.; Distrutti, E.; Renga, B.; Bifulco, G.; Baldelli, F.; Donini, A.; Fiorucci, S. The bile acid receptor GPBAR-1 (TGR5) modulates integrity of intestinal barrier and immune response to experimental colitis *PLoS One* **2011**, 6, e25637.
- <sup>61</sup> Cipriani, S.; Mencarelli, A.; Bruno, A.; Renga, B.; Distrutti, E.; Santucci, L.; Baldelli, F.; Fiorucci, S. Activation of the bile acid receptor GPBAR1 protects against gastrointestinal injury caused by non-steroidal anti-inflammatory drugs and aspirin in mice *Br. J. Pharmacol.* **2013**, 168, 225– 237
- <sup>62</sup> Fiorucci, S.; Mencarelli, A.; Palladino, G.; Cipriani, S. Bile-acid-activated receptors: targeting TGR5 and farnesoid-X-receptor in lipid and glucose disorders *Trends Pharmacol. Sci.* **2009**, 30, 570– 580
- <sup>63</sup> Fiorucci, S.; Cipriani, S.; Baldelli, F.; Mencarelli, A. Bile acid-activated receptors in the treatment of dyslipidemia and related disorders *Prog. Lipid Res.* **2010**, 49, 171– 185
- <sup>64</sup> Cipriani, S.; Mencarelli, A.; Palladino, G.; Fiorucci, S. FXR activation reverses insulin resistance and lipid abnormalities and protects against liver steatosis in Zucker (fa/fa) obese rats *J. Lipid Res.* **2010**, 51, 771– 784
- <sup>65</sup> Baghdasaryan, A.; Claudel, T.; Gumhold, J.; Silbert, D.; Adorini, L.; Roda, A.; Vecchiotti, S.; Gonzalez, F. J.; Schoonjans, K.; Strazzabosco, M.; Fickert, P.; Trauner, M. Dual farnesoid X receptor/TGR5 agonist INT-767 reduces liver injury in the Mdr2<sup>-/-</sup> (Abcb4<sup>-/-</sup>) mouse cholangiopathy model by promoting biliary HCO<sub>3</sub><sup>-</sup> output *Hepatology* **2011**, 54, 1303– 1312

- <sup>66</sup> Rizzo, G.; Passeri, D.; De Franco, F.; Ciaccioli, G.; Donadio, L.; Rizzo, G.; Orlandi, S.; Sadeghpour, B.; Wang, X. X.; Jiang, T.; Levi, M.; Pruzanski, M.; Adorini, L. Functional characterization of the semisynthetic bile acid derivative INT-767, a dual farnesoid X receptor and TGR5 agonist *Mol. Pharmacol.* **2010**, *78*, 617–630
- <sup>67</sup> Pellicciari, R.; Fiorucci, S.; Pruzanski, M. Preparation of Bile Acid Derivatives as FXR Ligands for the Prevention or Treatment of FXR-Mediated Diseases or Conditions. PCT Int. WO 2008002573 A2 20080103, **2008**.
- <sup>68</sup> D'Amore, Saverio Di Leva, Sepe, Renga, Del Gaudio, D'Auria, Zampella, Fiorucci, Limongelli. Design, Synthesis, and Biological Evaluation of Potent Dual Agonists of Nuclear and Membrane Bile Acid Receptors *J. Med. Chem.*, **2014**, *57*(3), 937–954
- <sup>69</sup> Sepe, V.; Ummarino, R.; D'Auria, M. V.; Chini, M. G.; Bifulco, G.; Renga, B.; D'Amore, C.; Debitus, C.; Fiorucci, S.; Zampella, A. Conicasterol E, a small heterodimer partner sparing farnesoid X receptor modulator endowed with a pregnane X receptor agonistic activity, from the marine sponge *Theonella swinhoei*. *J. Med. Chem.* **2012**, *55*, 84–93.
- <sup>70</sup> Gioiello, A.; Macchiarulo, A.; Carotti, A.; Filippini, P.; Costantino, G.; Rizzo, G.; Adorini, L.; Pellicciari, R. Extending SAR of bile acids as FXR ligands: discovery of 23-N-(carbocinnamyloxy)-3 $\alpha$ ,7 $\alpha$ -dihydroxy-6 $\alpha$ -ethyl-24-nor-5 $\beta$ -cholan-23-amine. *Bioorg. Med. Chem.* **2011**, *19*, 2650–2658.
- <sup>71</sup> Tserng, K. Y.; Klein, P. D. Formylated bile acids: improved synthesis, properties, and partial deformylation. *Steroids* **1977**, *29*, 635–648.
- <sup>72</sup> Schteingart, C. D.; Hofmann, A. F. Synthesis of 24-nor-5 betacholan-23-oic acid derivatives: a convenient and efficient one-carbon degradation of the side chain of natural bile acids. *J. Lipid Res.* **1988**, *29*, 1387–1395.
- <sup>73</sup> Mancuso, A. J.; Swern, D. Activated dimethyl sulfoxide: useful reagents for synthesis. *Synthesis* **1981**, *3*, 165–185.
- <sup>74</sup> Nagoka, H.; Kishi, Y. Further synthetic studies on rifamycin S. *Tetrahedron* **1981**, *37*, 3873–3888.
- <sup>75</sup> Fujino, T.; Une, M.; Imanaka, T.; Inoue, K.; Nishimaki-Mogami, T. Structure–activity relationship of bile acids and bile acid analogs in regards to FXR activation. *J. Lipid Res.* **2004**, *45*, 132–138.
- <sup>76</sup> Pellicciari, R.; Fiorucci, S.; Camaioni, E.; Clerici, C.; Costantino, G.; Maloney, P. R.; Morelli, A.; Parks, D. J.; Willson, T. M. 6- $\alpha$ -Ethyl-chenodeoxycholic acid (6-ECDCA), a potent and selective FXR agonist endowed with anticholestatic activity *J. Med. Chem.* **2002**, *45*, 3569–3572
- <sup>77</sup> Fiorucci, S.; Cipriani, S.; Mencarelli, A.; Baldelli, F.; Bifulco, G.; Zampella, A. Farnesoid X receptor agonist for the treatment of liver and metabolic disorders: focus on 6-ethyl-CDCA *Mini-Rev. Med. Chem.* **2011**, *11*, 753–762
- <sup>78</sup> Parker, H. E.; Wallis, K.; le Roux, C. W.; Wong, K. Y.; Reimann, F.; Gribble, F. M. Molecular mechanisms underlying bile acid-stimulated glucagon-like peptide-1 secretion *Br. J. Pharmacol.* **2012**, *165*, 414–423
- <sup>79</sup> Petersen, A. B.; Christensen, M. Clinical potential of lixisenatide once daily treatment for type 2 diabetes mellitus *Diabetes, Metab. Syndr. Obes.* **2013**, *6*, 217–231

- <sup>80</sup> Nauck, M. A.; Baranov, O.; Ritzel, R. A.; Meier, J. J. Do current incretin mimetics exploit the full therapeutic potential inherent in GLP-1 receptor stimulation? *Diabetologia* **2013**, 56, 1878– 1883
- <sup>81</sup> Scheen, A. J. Cardiovascular effects of dipeptidyl peptidase-4 inhibitors: from risk factors to clinical outcomes *Postgrad. Med.* **2013**, 125, 7– 20
- <sup>82</sup> Cohen, D. Reports of pancreatitis are 20–30 times more likely with GLP-1 drugs, analysis finds *Br. Med. J.* **2013**, 346, f2607
- <sup>83</sup> Butler, A. E.; Campbell-Thompson, M.; Gurlo, T.; Dawson, D. W.; Atkinson, M.; Butler, P. C. Marked expansion of exocrine and endocrine pancreas with incretin therapy in humans with increased exocrine pancreas dysplasia and the potential for glucagon-producing neuroendocrine tumors *Diabetes* **2013**, 62, 2595-2604
- <sup>84</sup> Lebon, G.; Warne, T.; Edwards, P. C.; Bennett, K.; Langmead, C. J.; Leslie, A. G.; Tate, C. G. Agonist-bound adenosine A<sub>2A</sub> receptor structures reveal common features of GPCR activation *Nature* **2011**, 474, 521– 525
- <sup>85</sup> Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A semiempirical free energy force field with charge-based desolvation *J. Comput. Chem.* **2007**, 28, 1145– 1152
- <sup>86</sup> Tiwari, A.; Maiti, P. TGR5: an emerging bile acid G-protein-coupled receptor target for the potential treatment of metabolic disorders *Drug Discovery Today* **2009**, 14, 523– 530
- <sup>87</sup> Grazioso, G.; Limongelli, V.; Branduardi, D.; Novellino, E.; De Micheli, C.; Cavalli, A.; Parrinello, M. Investigating the mechanism of substrate uptake and release in the glutamate transporter homologue GltPh through metadynamics simulations *J. Am. Chem. Soc.* **2012**, 134, 453– 463
- <sup>88</sup> Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel metadynamics as accurate binding free-energy method *Proc. Natl. Acad. Sci. U.S.A.* **2013**, 110, 6358– 6363
- <sup>89</sup> Limongelli, V.; Marinelli, L.; Cosconati, S.; La Motta, C.; Sartini, S.; Mugnaini, L.; Da Settimo, F.; Novellino, E.; Parrinello, M. Sampling protein motion and solvent effect during ligand binding *Proc. Natl. Acad. Sci. U.S.A.* **2012**, 109, 1467– 1472
- <sup>90</sup> Macchiarulo, A.; Gioiello, A.; Thomas, C.; Pols, T. W. H.; Nuti, R.; Ferrari, C.; Giacche, N.; De Franco, F.; Pruzanski, M.; Auwerx, J.; Schoonjans, K.; Pellicciari, R. Probing the binding site of bile acids in TGR5 *ACS Med. Chem. Lett.* **2013**, 4, 1158– 1162
- <sup>91</sup> Mi, L. Z.; Devarakonda, S.; Harp, J. M.; Han, Q.; Pellicciari, R.; Willson, T. M.; Khorasanizadeh, S.; Rastinejad, F. Structural basis for bile acid binding and activation of the nuclear receptor FXR *Mol. Cell* **2003**, 11, 1093– 1100
- <sup>92</sup> Wu B, Zhang Z, Noberini R, Barile E, Giulianotti M, Pinilla C, Houghten RA, Pasquale EB, Pellicchia M. HTS by NMR of combinatorial libraries: a fragment-based approach to ligand discovery *Chem Biol.* **2013**;20(1):19-33.
- <sup>93</sup> Congreve M, Chessari G, Tisi D, Woodhead AJ. Recent developments in fragment-based drug discovery. *J Med Chem.* **2008**; 51(13):3661-80.
- <sup>94</sup> Fischer M, Hubbard RE. Fragment-based ligand discovery. *Mol Interv.* **2009**; 9(1):22-30.

- <sup>95</sup> Hajduk PJ, Greer J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov.* **2007**; 6(3):211-9.
- <sup>96</sup> Murray CW, Rees DC. The rise of fragment-based drug discovery. *Nat Chem.* **2009**; 1(3):187-92
- <sup>97</sup> De Corte, B. L. From 4,5,6,7-tetrahydro-5-methylimidazo[4,5,1-jk](1,4)benzodiazepin-2(1H)-one (TIBO) to etravirine (TMC125): fifteen years of research on non-nucleoside inhibitors of HIV-1 reverse transcriptase. *J. Med. Chem.* **48**, 1689–1696 (2005).
- <sup>98</sup> Fry, D. W. *et al.* A specific inhibitor of the epidermal growth factor receptor tyrosine kinase. *Science* **265**, 1093–1095 (1994).
- <sup>99</sup> Macarron, Banks, Bojanic, Burns, Cirovic, Garyantes, Green, Hertzberg, Janzen, Paslay, Schopfer & Sittampalam Impact of high-throughput screening in biomedical research *Nature Reviews Drug Discovery* **2011**, 10,188-195
- <sup>100</sup> Dooley CT, Houghten RA. The use of positional scanning synthetic peptide combinatorial libraries for the rapid determination of opioid receptor ligands. *Life Sci.* **1993**; 52(18):1509-17.
- <sup>101</sup> Dooley CT, Ny P, Bidlack JM, Houghten RA Selective ligands for the mu, delta, and kappa opioid receptors identified from a single mixture based tetrapeptide positional scanning combinatorial library. *J Biol Chem.* **1998**; 273(30):18848-56.
- <sup>102</sup> Houghten RA, Pinilla C, Blondelle SE, Appel JR, Dooley CT, Cuervo JH Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature.* **1991**; 354(6348):84-6.
- <sup>103</sup> Pinilla C, Appel JR, Blanc P, Houghten RA Rapid identification of high affinity peptide ligands using positional scanning synthetic peptide combinatorial libraries. *Biotechniques.* **1992**; 13(6):901-5.
- <sup>104</sup> Elisa Barile and Maurizio Pellecchia NMR-Based Approaches for the Identification and Optimization of Inhibitors of Protein–Protein Interactions *Chem. Rev.*, **2014**;114 (9):4749–4763
- <sup>105</sup> Diamond SL Methods for mapping protease specificity. *Curr Opin Chem Biol.* **2007**; 11(1):46-51.
- <sup>106</sup> Lim MD, Craik CS Using specificity to strategically target proteases. *Bioorg Med Chem.* **2009**; 17(3):1094-100.
- <sup>107</sup> Baell JB, Holloway GA New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem.* **2010**; 53(7):2719-40.
- <sup>108</sup> Shoichet BK Interpreting steep dose-response curves in early inhibitor discovery. *J Med Chem.* **2006**; 49(25):7274-7.
- <sup>109</sup> Giannetti AM, Koch BD, Browner MF Surface plasmon resonance based assay for the detection and characterization of promiscuous inhibitors. *J Med Chem.* **2008**; 51(3):574-80.
- <sup>110</sup> Tapan Kumar Palai and Smruti Ranjan Mishra. Caspases: An apoptosis mediator *J. Adv. Vet. Anim. Res.*, **2015**; 2(1): 18-22
- <sup>111</sup> Thornberry NA, Lazebnik Y. Caspases: enemies within. *Science.* **1998**; 281(5381):1312-6.
- <sup>112</sup> Zhiliang Wang, Michael Cuddy, Temesgen Samuel, Kate Welsh, Aaron Schimmer, Farid Hanaii, Richard Houghten, Clemencia Pinilla and John C. Reed

Cellular, Biochemical, and Genetic Analysis of Mechanism of Small Molecule IAP Inhibitors *J Biol Chem*, **2004**, 279, 48168-48176.

<sup>113</sup> Salvesen, G. S., and Duckett, C. S. Apoptosis: IAP proteins: blocking the road to death's door *Nat. Rev. Mol. Cell Biol.* **2002**;3, 401-410

<sup>114</sup> Kasof, G. M., and Gomes, B. C. Livin, a Novel Inhibitor of Apoptosis Protein Family Member *J. Biol. Chem.* **2001** 276, 3238-3246

<sup>115</sup> Vucic, D., Stennicke, H. R., Pisabarro, M. T., Salvesen, G. S., and Dixit, V. M. ML-IAP, a novel inhibitor of apoptosis that is preferentially expressed in human melanomas *Curr. Biol* **2000**; 10, 1359-1366

<sup>116</sup> Chen J, Wu W, Tahir SK, Kroeger PE, Rosenberg SH, Cowser LM, Bennett F, Krajewski S, Krajewska M, Welsh K, Reed JC, Ng SC Down-regulation of survivin by antisense oligonucleotides increases apoptosis, inhibits cytokinesis and anchorage-independent growth. *Neoplasia*. **2000**;2(3):235-41.

<sup>117</sup> Gavin J. Gordon, Krishnarao Appasani, Jeremy P. Parcells, Nishit K. Mukhopadhyay, Michael T. Jaklitsch, William G. Richards, David J. Sugarbaker and Raphael Bueno Inhibitor of apoptosis protein-1 promotes tumor cell survival in mesothelioma *Carcinogenesis* **2002**; 23 (6):1017-1024.

<sup>118</sup> Hiromasa Sasaki, YingLun Sheng, Fumikazu Kotsuji, and Benjamin K. Tsang Down-Regulation of X-linked Inhibitor of Apoptosis Protein Induces Apoptosis in Chemoresistant Human Ovarian Cancer Cells *Cancer Res* **2000**; 60;5659

<sup>119</sup> Martin Holcik, Chiaoli Yeh, Robert G Korneluk and Terry Chow Translational upregulation of X-linked inhibitor of apoptosis (XIAP) increases resistance to radiation induced cell death. *Oncogene* **2000**, 19,(36):4174-4177

<sup>120</sup> Fischer U, Janicke RU, Schulze Osthoff K Many cuts to ruin: a comprehensive update of caspase substrates. *Cell Death and Differentiation*, **2003**; 10: 76- 100.

<sup>121</sup> Shiozaki EN, Chai J, Rigotti DJ, Riedl SJ, Li P, Srinivasula SM, Alnemri ES, Fairman R, Shi Y. Mechanism of XIAP-mediated inhibition of caspase-9. *Mol. Cell.* **2003**; 11:519-527.

<sup>122</sup> Fulda S et al., Targeting IAP proteins for therapeutic intervention in cancer., *Nature Reviews* **2012**;11, 109-123

<sup>123</sup> Liu Z, Sun C, Olejniczak ET, Meadows RP, Betz SF, Oost T, Herrmann J, Wu JC, Fesik SW. Structural basis for binding of Smac/DIABLO to the XIAP BIR3 domain *Nature*. **2000**; 21-28;408(6815):1004-8.

<sup>124</sup> Sun C, Cai M, Meadows RP, Xu N, Gunasekera AH, Herrmann J, Wu JC, Fesik SW. NMR structure and mutagenesis of the third Bir domain of the inhibitor of apoptosis protein XIAP. *J Biol Chem*. **2000**;275(43):33777-81.

<sup>125</sup> Srinivasula SM, Hegde R, Saleh A, Datta P, Shiozaki E, Chai J, Lee RA, Robbins PD, Fernandes-Alnemri T, Shi Y, Alnemri ES. A conserved XIAP-interaction motif in caspase-9 and Smac/DIABLO regulates caspase activity and apoptosis. *Nature*. **2001**; 410:112-116.

<sup>126</sup> Bank A, Wang P, Du C, Yu J, Zhang L. SMAC mimetics sensitize nonsteroidal anti-inflammatory drug-induced apoptosis by promoting caspase-3-mediated cytochrome c release. *Cancer Res*. **2008**;68:276-284.

<sup>127</sup> Flygare JA, Fairbrother WJ. Small-molecule pan-IAP antagonists: a patent review. *Expert Opin. Ther.Pat.* **2010**; 20:251-267

<sup>128</sup> Li J, Li Q, Xie C, Zhou H, Wang Y, Zhang N, Shao H, Chan SC, Peng X, Lin SC, Han J. Beta-actin is required for mitochondria clustering and ROS generation

- in TNF-induced, caspase-independent cell death. *J. Cell Sci.* **2004**; 117:4673–4680.
- <sup>129</sup> Stebbins JL, Zhang Z, Chen J, Wu B, Emdadi A, Williams ME, Cashman J, Pellecchia M. Nuclear magnetic resonance fragment-based identification of novel FKBP12 inhibitors. *J. Med. Chem.* **2007**; 50:6607–6617.
- <sup>130</sup> Sharma SK, Straub C, Zawel L. Development of Peptidomimetics Targeting IAPs. *Int J Pept Res Ther* **2006**;12(1):21-32.
- <sup>131</sup> Sun W, Nikolovska-Coleska Z, Qin D, Sun H, Yang CY, Bai L, Qiu S, Wang Y, Ma D, Wang S. Design, synthesis, and evaluation of potent, nonpeptidic mimetics of second mitochondria-derived activator of caspases. *J Med Chem.* **2009**;52(3):593-6.
- <sup>132</sup> C. Schultz, Prodrugs of biologically active phosphate esters. *Bioorg. Med. Chem.* **2003**, 11, 885.
- <sup>133</sup> S. Zhao, F. A. Etzkorn, *Bioorg. Med. Chem. Lett.* **2007**, 17, 6615.
- <sup>134</sup> Valentino J. Stella Prodrugs: challenges and rewards. *P. Springer Science & Business Media*, **2007**
- <sup>135</sup> Flygare, Beresini, Budha, Chan, Chan, Cheeti, Cohen, Deshayes, Doerner, Eckhardt, Elliott, Feng, Franklin, Frankovitz  
Reisner, Gazzard, Halladay, Hymowitz, La, Lo  
Russo, Maurer, Murray, Plise, Quan, Stephan, Young, Tom, Tsui, Um, Varfolomeev, Vucic, Wagner, Wallweber, Wang, Ware, Wen, H Wong, J M.  
Wong, M Wong, S Wong, Yu, Zobel and Fairbrother Discovery of a Potent Small-Molecule Antagonist of Inhibitor of Apoptosis (IAP) Proteins and Clinical Candidate for the Treatment of Cancer (GDC-0152) *J. Med. Chem.*, **2012**, 55 (9), 4101–4113
- <sup>136</sup> Teschner, M.; Henn, C.; Vollhardt, H.; Reiling, S.; Brickmann, J. Texture mapping: a new tool for molecular graphics. *J. Mol. Graphics* **1994**, 12, 98–105.
- <sup>137</sup> Stringer, B ; Day, B ; McCarron, J ; Lackmann, M ; Boyd, A EphA3 (Eph receptor A3) *Atlas Genet Cytogenet Oncol Haematol.* **2010**;14(3)
- <sup>138</sup> Pasquale EB Eph receptor signalling casts a wide net on cell behaviour. *Nat Rev Mol Cell Biol.* **2005**;6(6):462-75.
- <sup>139</sup> D. Arvanitis, A. Davy Eph/ephrin signaling: networks *Gene Dev.* **2008**; 22:416
- <sup>140</sup> E.B. Pasquale Eph-Ephrin Bidirectional Signaling in Physiology and Disease *Cell*, **2008**;133:38
- <sup>141</sup> W. Mao, E. Luis, S. Ross, J. Silva, C. Tan, C. Crowley, C. Chui, G. Franz, P. Senter, H. Koeppen, P. Polakis EphB2 as a therapeutic antibody drug target for the treatment of colorectal cancer. *Cancer Res.*, **2004**;64:781
- <sup>142</sup> A. Lugli, H. Spichtin, R. Maurer, M. Mirlacher, J. Kiefer, P. Huusko, D. Azorsa, L. Terracciano, G. Sauter, O.P. Kallioniemi, S. Mousses, L. Tornillo EphB2 expression across 138 human tumor types in a tissue microarray: high levels of expression in gastrointestinal cancers. *Clin. Cancer Res.* **2005**;11:6450
- <sup>143</sup> T.N. Campbell, S.M. Robbins The Eph receptor/ephrin system: an emerging player in the invasion game. *Curr. Issues Mol. Biol.*, **2008**;10:61
- <sup>144</sup> Klas Kullander and Rüdiger Klein Mechanisms and functions of eph and ephrin signalling *Nat Rev Mol Cell Biol* **2002**;3, 475-486

- <sup>145</sup> Day BW, Stringer BW, Al-Ejeh F, Ting MJ, Wilson J, Ensbey KS, et al. EphA3 maintains tumorigenicity and is a therapeutic target in glioblastoma multiforme. *Cancer Cell* **2013**;23:238–48
- <sup>146</sup> Chiari R, Hames G, Stroobant V, Texier C, Maillere B, Boon T, et al. Identification of a tumor-specific shared antigen derived from an Eph receptor and presented to CD4 T cells on HLA class II molecules. *Cancer Res* **2000**;60:4855–63.
- <sup>147</sup> Xi HQ, Wu XS, Wei B, Chen L. Aberrant expression of EphA3 in gastric carcinoma: correlation with tumor angiogenesis and survival. *J Gastroenterol* **2012**;47:785–94.
- <sup>148</sup> Keane N, Freeman C, Swords R, Giles FJ. EPHA3 as a novel therapeutic target in the hematological malignancies. *Expert Rev Hematol* **2012**;5:325–40.
- <sup>149</sup> Ashton JM, Balys M, Neering SJ, Hassane DC, Cowley G, Root DE, et al. Gene sets identified with oncogene cooperativity analysis regulate *in vivo* growth and survival of leukemia stem cells. *Cell Stem Cell* **2012**;11:359–72.
- <sup>150</sup> Janes PW, Slape CI, Farnsworth RH, Atapattu L, Scott AM, Vail ME. EphA3 biology and cancer *Growth Factors*. **2014**;32(6):176–89.
- <sup>151</sup> Y Choi, F Syeda, J R. Walker, P J. Finerty Jr., D Cuerrier, A Wojciechowski, Q Liu, S Dhe-Paganon, N S. Gray Discovery and structural analysis of Eph receptor tyrosine kinase inhibitors *Bioorg Medicinal Chem Lett* **2009**; 19(15):4467–4470
- <sup>152</sup> KB004, a Novel Non-Fucosylated Humanized Antibody, Targeting EphA3, is Active and Well Tolerated in a Phase 1/2 Study of Advanced Hematologic Malignancies *ASH Annual Meeting*, December **2014**
- <sup>153</sup> S. Hunke, J.M. Betton Temperature effect on inclusion body formation and stress response in the periplasm of *Escherichia coli* *Mol. Microbiol.*, **2003**;50:1579–1589
- <sup>154</sup> G. Georgiou, P. Valax Expression of correctly folded proteins in *Escherichia coli* *Curr. Opin. Biotechnol.*, **1996**;7:190–197
- <sup>155</sup> E.D.B. Clark Refolding of recombinant proteins *Curr. Opin. Biotechnol.*, **1998**;9:157–163
- <sup>156</sup> R.C. Stevens Design of high-throughput methods of protein production for structural biology *Structure*, **2000**;8:R177–R185
- <sup>157</sup> P. Turner, O. Holst, E.N. Karlsson Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in *Escherichia coli* by using a soluble fusion-tag and by tuning of inducer concentration *Protein Expr. Purif.*, **2005**;39:54–60
- <sup>158</sup> Stewart EJ, Aslund F, Beckwith J. Disulfide bond formation in the *Escherichia coli* cytoplasm: an *in vivo* role reversal for thioredoxins. *EMBO J.* **1998**;17(19):5543–50.
- <sup>159</sup> P. Braun, J. LaBaer High throughput protein production for functional proteomics *Trends Biotechnol.*, **2003**;21:383–388
- <sup>160</sup> Kaplan W, Hüsler P, Klump H, Erhardt J, Sluis-Cremer N, Dirr H. Conformational stability of pGEX-expressed *Schistosoma japonicum* glutathione S-transferase: a detoxification enzyme and fusion-protein affinity tag. *Protein Sci.* **1997**;6(2):399–406.

- <sup>161</sup> George Kostallas, Lisa Sandersjö, Mohammed Al-Askeri and Patrik Samuelson Construction, expression and characterization of TEV protease mutants engineered for improved solubility DiVA: [diva2:416059](#)
- <sup>162</sup> Religa, Tomasz L.; Kay, Lewis E Optimal methyl labeling for studies of supra-molecular systems *J Biomol NMR* **2010**;47 (3),163-169
- <sup>163</sup> Tugarinov, Vitali; Kanelis, Voula; Kay, Lewis E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy *Nat Prot* **2006**; 1 (2):749-754
- <sup>164</sup> Pellecchia M Solution nuclear magnetic resonance spectroscopy techniques for probing intermolecular interactions *Chem Biol.* **2005**;12(9):961-71.
- <sup>165</sup> Pellecchia M, Becattini B, Crowell KJ, Fattorusso R, Forino M, Fragai M, Jung D, Mustelin T, Tautz L. NMR-based techniques in the hit identification and optimisation processes. *Expert Opin Ther Targets.* **2004**;8(6):597-611.
- <sup>166</sup> Mencarelli, A.; Renga, B.; Migliorati, M.; Cipriani, S.; Distrutti, E.; Santucci, L.; Fiorucci, S. The bile acid sensor farnesoid X receptor is a modulator of liver immunity in a rodent model of acute hepatitis. *J. Immunol.* **2009**, 183, 6657–6666.
- <sup>167</sup> Di Leva, F. S.; Festa, C.; D'Amore, C.; De Marino, S.; Renga, B.; D'Auria, M. V.; Novellino, E.; Limongelli, V.; Zampella, A.; Fiorucci, S. Binding mechanism of the farnesoid X receptor marine antagonist suvanine reveals a strategy to forestall drug modulation on nuclear receptors. Design, synthesis, and biological evaluation of novel ligands. *J. Med. Chem.* **2013**, 56, 4701–4717.
- <sup>168</sup> Chenna, R.; Sugawara, H.; Koike, T.; Lopez, R.; Gibson, T. J.; Higgins, D. G.; Thompson, J. D. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **2003**, 31, 3497–3500.
- <sup>169</sup> Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. ClustalW and ClustalX version 2. *Bioinformatics* **2007**, 23, 2947–2948.
- <sup>170</sup> Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, 234, 779–815.
- <sup>171</sup> *Maestro*, version 9.3; Schroëdinger, LLC: New York, NY, **2012**.
- <sup>172</sup> Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, 118, 11225–11236.
- <sup>173</sup> *MacroModel*, version 9.9; Schroëdinger, LLC: New York, NY, **2012**.
- <sup>174</sup> *LigPrep*, version 2.5; Schroëdinger, LLC: New York, NY, **2012**.
- <sup>175</sup> *Epik*, version 2.3; Schroëdinger, LLC: New York, NY, **2012**.
- <sup>176</sup> Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, 14, 33–38.
- <sup>177</sup> Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, 79, 926–935.
- <sup>178</sup> Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, 117, 5179–5197.

- <sup>179</sup> Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, 78, 1950–1958.
- <sup>180</sup> Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, 25, 1157–1174.
- <sup>181</sup> Skjevik, Å. A.; Madej, B. D.; Walker, R. C.; Teigen, K. LIPID11: a modular framework for lipid simulations using Amber. *J. Phys. Chem. B* **2012**, 116, 11124–11136.
- <sup>182</sup> Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, 26, 1781–1802.
- <sup>183</sup> Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: an  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, 98, 10089–10092.
- <sup>184</sup> Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well behaved electrostatic potential based method using charge restraints for determining atom-centered charges: the RESP model. *J. Phys. Chem.* **1993**, 97, 10269–10280.
- <sup>185</sup> Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, **2004**.
- <sup>186</sup> Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanics. *J. Mol. Graphics Modell.* **2006**, 25, 247–260.
- <sup>187</sup> Sun, Cai, Meadows, Xu, Gunasekera, Herrmann, Wu, Fesik NMR Structure and Mutagenesis of the Third Bir Domain of the Inhibitor of Apoptosis Protein XIAP *J Biol Chem*, **2000**, 275, 33777-33781.
- <sup>188</sup> Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, 267, 727–748.
- <sup>189</sup> Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, 11, 425–445.
- <sup>190</sup> Teschner, M.; Henn, C.; Vollhardt, H.; Reiling, S.; Brickmann, J. Texture mapping: a new tool for molecular graphics. *J. Mol. Graphics* **1994**, 12, 98–105.

- 
- <sup>191</sup> Religa, Tomasz L.; Kay, Lewis E Optimal methyl labeling for studies of supra-molecular systems *J Biomol NMR* **2010**;47 (3),163-169
- <sup>192</sup> Tugarinov, Vitali; Kanelis, Voula; Kay, Lewis E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy *Nat Prot* **2006**; 1 (2):749-754
- <sup>193</sup> Lee Fielding NMR methods for the determination of protein–ligand dissociation constants *Curr Top Med Chem.* **2003**;3(1):39-53