



Metabolomics with Multi-Block Modelling of Mass Spectrometry and Nuclear Magnetic Resonance in order to Discriminate Haplosclerida Marine Sponges

Mehdi A Beniddir, Laurence Le Moyec, Mohamed Nawfal Triba, Arlette Longeon, Alexandre Deville, Alain Blond, Pham van Cuong, Nicole Jane de Voogd, Marie-Lise Bourguet-Kondracki

► To cite this version:

Mehdi A Beniddir, Laurence Le Moyec, Mohamed Nawfal Triba, Arlette Longeon, Alexandre Deville, et al.. Metabolomics with Multi-Block Modelling of Mass Spectrometry and Nuclear Magnetic Resonance in order to Discriminate Haplosclerida Marine Sponges. *Analytical and Bioanalytical Chemistry*, 2022, 414 (19), pp.5929-5942. 10.1007/s00216-022-04158-5 . mnhn-03867254

HAL Id: mnhn-03867254

<https://mnhn.hal.science/mnhn-03867254>

Submitted on 26 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Metabolomics with Multi-Block Modelling of Mass Spectrometry and Nuclear Magnetic Resonance in order to Discriminate Haplosclerida Marine Sponges

Mehdi A. Beniddir¹ . Laurence Le Moyec^{2,3} . Mohamed N. Triba⁴ . Arlette Longeon² . Alexandre Deville² . Alain Blond² . Pham Van Cuong⁵ . Nicole J. de Voogd^{6,7} . Marie-Lise Bourguet-Kondracki²

¹ Équipe “Chimie des substances naturelles” BioCIS, CNRS, Université Paris-Saclay, 5 rue J.-B. Clément, 92290 Châtenay-Malabry, France.

² Muséum National d'Histoire Naturelle, Unité Molécules de Communication et Adaptation des Micro-organismes (MCAM-UMR 7245 MNHN-CNRS, Sorbonne Université), 57 rue Cuvier (CP54), 75005 Paris, France.

³ Université d'Evry Val d'Essonne-Université Paris-Saclay, Boulevard François Mitterrand, 91360 Evry, France.

⁴ Laboratoire CSPBAT UMR7244 UFR Santé, Médecine et Biologie Humaine Université Paris 13, 74 rue Marcel Cachin, 93017 Bobigny, France.

⁵ Institute of Marine Biochemistry (IMBC), Vietnam Academy of Science and Technology (VAST), 18 Hoang Quoc Viet, Cau Giay, Hanoi, Vietnam.

⁶ Naturalis Biodiversity Center, PO Box 9517 2300, RA Leiden, The Netherlands.

⁷ Institute of Environmental Sciences, Leiden University, Einsteinweg 2, 2333 CC, Leiden, The Netherlands.

✉ Mehdi A. Beniddir, Laurence Le Moyec, Marie-Lise Bourguet-Kondracki
mehdi.beniddir@universite-paris-saclay.fr, laurence.le-moyec@mnhn.fr, marie-lise.bourguet@mnhn.fr

Abstract

A comprehensive metabolomic strategy, integrating ¹H NMR and MS-based multi-block modelling in conjunction with multi-informational molecular networking, has been developed to discriminate sponges of the order Haplosclerida, well known for being taxonomically contentious. A in house collection of 33 marine sponge samples belonging to three families (Callyspongiidae, Chalinidae, Petrosiidae) and four different genera (*Callyspongia*, *Haliclona*, *Petrosia*, *Xestospongia*) was investigated using LC-MS/MS, molecular networking and the annotations processes combined with NMR data and multivariate statistical modelling. The combination of MS and NMR data into supervised multivariate models led to discriminate, out of the four genera, three groups based on the presence of metabolites, not necessarily previously described in the Haplosclerida order. Although these metabolomic methods have already been applied separately, it is the first time that a multi-block untargeted approach using MS and NMR is combined with molecular networking and statistically analyzed, pointing out the pros and cons of this strategy.

Keywords Haplosclerida Sponges; Mass Spectrometry; Nuclear Magnetic Resonance Spectroscopy; Molecular Networking; Multivariate Statistical Analyzes

Introduction

Metabolomics, combining multiparametric chemical analysis technique and multivariate statistical analysis, has been applied to a variety of fields from human health to microbiology. In this perspective, marine biology is continuously benefiting from the methodological advances for metabolite identification in chemotaxonomy. Nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) often coupled with separation techniques, are the two most powerful analytical methods, used in metabolomic analyzes [1]. They both generate complementary metabolic features from the same sample; hence, their

combination greatly improves the coverage of the metabolome and enhances the accuracy of metabolites identification [2]. Recently, tremendous efforts have been directed toward the integration of MS and NMR data through the development of combined chemoinformatics strategies [3, 4]. SUMMIT MS/NMR and NMR/MS translator illustrate important developments in this field [5-8]. In many cases, the concatenation of MS and NMR data can potentially improve the reliability and predictive accuracy of statistical methods. In this regard, multi-block strategies emerged as an appealing approach for biomarker discovery [9-11]. The latter consists of combining the preprocessed individual blocks at the data level without performing any variable selection prior to modelling and interpreting the outcome. An interesting manner to unravel the structural identity of those metabolites could rely on tandem MS/MS molecular networking (MN). This approach constitutes a powerful strategy to visualize and organize MS/MS data sets and automate database searches for metabolites identification of marine organisms with the help of *in silico* generated data for annotation [12-14], especially for sponges for which several taxonomic issues remain.

Sponges are active filter-feeder animals, which belong to the phylum Porifera. These sessile organisms, which generally live on hard, or in soft, substrates, ensure their defense and their survival by their capacity to produce diverse specialized molecules, which provide protective, defense, and ecological roles [15], in addition to other defense mechanisms such as physical spicule defense. Their classification is, among others, based on the presence, size and arrangement of calcareous or siliceous skeletal structures called spicules [16]. However, these characteristic elements are lacking in some species, which reveal only the presence of spongin fibers and fibrillar collagen, leading to, at times, challenging taxonomic identification. Therefore, a chemical taxonomy classification was expected to aid with the morphological classification limitations. Nowadays, systematics associates multiple and complementary sources of data, called integrative taxonomy, combining external morphology, spicules, embryology, geography, reproduction, genetic sequences, and molecular results [17]. In this vein, chemotaxonomy investigations of sponges of the order Haplosclerida are still a work in progress (class Demospongiae). The latter encompasses more than 650 species, which are classified into the families Calcifibrospongiidae, Callyspongiidae, Chalinidae, Niphatidae, Petrosiidae and Phloeodictyidae, according to World Porifera Database [18]. The order Haplosclerida is one of the most speciose demosponge orders and, due to non-monophyletic relationships of its (morphologically defined) families and genera, one of the most challenging in terms of species identification [19-23]. The phylogeny of this order is the biggest enigma in sponge systematics due to the paucity of complex morphological features between putative sponge species and the incongruence between morphological and molecular characters. Marine sponges of the order Haplosclerida have been chemically widely investigated, leading to the discovery of diverse and varied molecules, mainly alkyl-piperidine- or alkyl pyridinium-type alkaloids and polyacetylene compounds. However, so far, no chemical markers were definitively identified, as illustrated with the study of Tribalat *et al.* [22], which led to the conclusion that the chemical diversity within the order Haplosclerida does not fit well with the taxonomic identification [22]. In this study, we applied a metabolomic strategy integrating ¹H NMR and HRMS-based multi-block modelling in conjunction with taxonomically informed molecular networking for studying 33 Haplosclerida marine sponge samples of three families (Callyspongiidae, Chalinidae, Petrosiidae) and four different genera (*Callyspongia* Duchassaing & Michelotti 1864, *Haliclona* Grant 1841, *Petrosia* Vosmaer 1885, *Xestospongia* de Laubenfels 1932). The aim of this study was not to identify all key chemical markers of this order but to select those discriminating the four genera by leveraging the complementary information provided by MS and NMR data. Although these “omics” approaches have already been explored separately, it is the first time that a multi-block approach using MS and NMR is combined with molecular networking and statistically analyzed for the classification of samples of the order Haplosclerida.

Experimental section

Sponge Material

Among our marine sponge collection, 33 samples from the order Haplosclerida within the three Callyspongiidae, Chalinidae and Petrosiidae families and four different genera (*Callyspongia* Duchassaing & Michelotti 1864, *Haliclona* Grant 1841, *Petrosia* Vosmaer 1885, *Xestospongia* de Laubenfels 1932) were selected for statistical analyses. Samples were mainly collected in Indonesia (off North and South Sulawesi Island) and also in the

Persian Gulf (off Oman), the Chinese Sea (Vietnam), and in the North Sea (off The Netherlands and France in Europe). Voucher specimen were deposited at the Naturalis Biodiversity Center.

Within the Callyspongiidae family, 10 samples from the genus *Callyspongia* were studied, including five samples from the South Sulawesi Island (Indonesia), which are *Callyspongia (Cladochalina) aerizusa* SS12 and SS19, *Callyspongia joubini* SS28, *Callyspongia (Cladochalina) subarmigera* SS18 and *Callyspongia robusta* SS53, as well as five unidentified *Callyspongia* species, whose two samples named NS101, NS105 from the North Sulawesi Island (Indonesia) and three samples from Oman (Persian Gulf) with the code name IO03 and IO28 and IO41.

Within the Chalinidae family, 13 samples from the genus *Haliclona* have been investigated, including three North Sulawesi sample *Haliclona (Reniera)* aff. *fascigera* NS107, *Haliclona (Gellius) amboinesis* NS69, *Haliclona (Halichoclona) vanderlandi* NS86, and two South Sulawesi samples *Haliclona (Reniera) fascigera* SS06A and SS06B, , as well as two samples from the North Sea off the Netherlands, namely *Haliclona (Soestella) xena* NL01A and NL01B, three samples from North Sea France collected off Roscoff *Haliclona (Reniera) cinerea* FR 17, *Haliclona (Haliclona) simulans* FR12 and FR60, and three unidentified samples one from North Sulawesi named NS110 and two from South Sulawesi named SS33 and SS51.

Within the Petrosiidae family, ten samples were studied: four samples were from the genus *Petrosia* and six from the genus *Xestospongia*. The *Petrosia* samples included one sample *Petrosia (Petrosia) hoeksemai* NS94 from North Sulawesi and two samples *P. hoeksemai* SS03 and *Petrosia (Petrosia) nigricans* SS44 from South Sulawesi as well as one unidentified sample (VPH) from Vietnam in the Chinese Sea. In addition, six *Xestospongia* samples have completed the panel of the family Petrosiidae samples: three samples from South Sulawesi *Xestospongia testudinaria* SS09 and SS10, *X. viridenigra* SS57, as well as three unidentified and most likely undescribed samples encompassing one from South Sulawesi SS40 and two from North Sulawesi NS66 and NS109. For further details, see Electronic Supplemental Material Table S1.

Sponge Extracts Preparation

The same protocol was used for all samples. Sponge samples (500 g) were cut into small pieces and immediately immersed in MeOH (1 L) after collection. After filtration, an aliquot of 20 mL was evaporated, and 150 mg of the obtained crude extract were solubilized in methanol and mixed to 2 g of C18 powder. After evaporation, the obtained homogeneous dry extract was deposited as a powder on a C18 Sep-Pack cartridge (Phenomenex 200 mg/10 mL) for being eluted firstly with 20 mL H₂O in order to eliminate salt and secondly with 20 mL MeOH to obtain the desalted extracts.

After solvent evaporation, an aliquot of 2 mg of each dried extract was dissolved in 500 μ L deuterated DMSO for NMR analyses and an aliquot of 200 μ g dissolved in 200 μ L MeOH was used for mass analyses.

Mass Spectrometry: Data Dependent LC-ESI-HRMS² Analysis and Processing

LC-ESI-HRMS² analyses were achieved using an Agilent LC-MS system comprising an Agilent 1260 infinity HPLC coupled to an Agilent 6530 Q-TOF-MS (Agilent Technologies, Massy, France) equipped with an ESI source, operating in positive-ion mode. A Sunfire analytical C₁₈ column (150 $\times$ 2.1 mm; i.d. 3.5 μ m, Waters) was used, with a flow rate of 250 μ L/min and a linear gradient from 5% B (A: H₂O + 0.1% formic acid, B: ACN) to 100% B in 20 min and then 100% B over 10 min for a total runtime of 30 min. Injection volume was set at 10 μ L. Source parameters were set as follows: capillary temperature at 320°C, source voltage at 3500 V, sheath gas flow rate at 10 L.min⁻¹. The divert valve was set to waste for the first 3 min. MS scans were operated in full-scan mode from m/z 100 to 1,700 (0.1s scan time) with a mass resolution of 11,000 at m/z 922. MS¹ scan was followed by MS² scans of the five most intense ions above an absolute threshold of 5,000 counts. Selected parent ions were fragmented at a collision energy fixed at 45 eV and an isolation window of 1.3 amu. Calibration solution, containing two internal reference masses (purine, C₅H₄N₄, m/z 121.050873, and HP-921 [hexakis-(1H,1H,3H-tetrafluoropentoxo)phosphazene], C₁₈H₁₈O₆N₃P₃F₂₄, m/z 922.0098). A permanent MS/MS exclusion list criteria was set to prevent oversampling of the internal calibrant. LC-UV and MS data acquisition and processing were performed using MassHunter Workstation software (Agilent Technologies, Massy, France).

Mass Spectrometry: LC-MS/MS Data Processing

The MS² data files, related to the 33 extracts were converted from the .d (Agilent) standard data-format to .mzXML format using the MSConvert software, part of the ProteoWizard package [24]. All .mzXML were then processed using MZmine 2 v52 [25]. The mass detection was realized keeping the noise level at 1.2E3 at MS¹ and at 2E1 at MS². The ADAP chromatogram builder was used using a minimum group size of scans of 2, a group intensity threshold of 2E3, a minimum highest intensity of 2E3 and *m/z* tolerance of 10 ppm [26]. The ADAP wavelets deconvolution algorithm was used with the following standard settings: S/N threshold = 100, minimum feature height = 3000, coefficient/area threshold = 5, Peak duration range 0.02 – 1.5 min, RT wavelet range 0.02 - 0.2. MS² scans were paired using a *m/z* tolerance range of 0.02 Da and RT tolerance range of 1.5 min. Isotopes were grouped using the isotopic peaks grouper algorithm with a *m/z* tolerance of 10 ppm and a RT tolerance of 1.3 min with the most intense peak. The “identification” module was employed to annotate the peak list, using the “adduct search” which compiled common charge carriers (H⁺, Na⁺, K⁺, Li⁺, CH₃OH⁺, ACN⁺), and “complex search”. The resulted peak list was filtered to keep only rows with MS² features corresponding to pseudo-molecular ions. The .mgf and .csv files were generated using the dedicated “Export/Submit to GNPS/FBMN” option. A X1 matrix was generated, in which each line is a sponge sample extract and each column a detected *m/z* (obtained from the above mentioned .csv file).

Mass Spectrometry: Feature-Based Molecular Networking Parameters

A molecular network was created using the online FBMN workflow [27] (version release_20) at GNPS (<http://gnps.ucsd.edu>) with a parent mass tolerance of 0.02 Da and an MS/MS fragment ion tolerance of 0.02 Da. A network was then created where edges were filtered to have a cosine score above 0.65 and more than 6 matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other’s respective top 10 most similar nodes. The spectra in the network were then searched against GNPS spectral libraries. All matches kept between network spectra and library spectra were required to have a score above 0.65 and at least 6 matched peaks. The molecular networking data were analyzed and visualized using Cytoscape (ver. 3.6.0) [28].

Mass Spectrometry: *In Silico* spectral Data Base of Dictionary Natural Products (ISDB-DNP) and Taxonomic Weighting Annotation

The clustered spectra of the molecular network were further dereplicated against the ISDB, following the guidelines suggested by Allard and co-workers [29, 30]. Spectral matching parameters were set as follows: tolerance = 0.02, score threshold = 0.2, Top K results = 5. A taxonomically informed score (Kingdom = Porifera) was further applied to the *in silico* annotated matches. The generated molecular network was visualized using Cytoscape (ver. 3.6.0) [28] and ChemViz 2 for structure display.

MarinLit-based Dereplication

Discriminant variables were searched against MarinLit [31] using their deprotonated exact masses (tolerance = ± 0.01) and “Haplosclerida” as a taxonomic filter.

NMR Data Acquisition and Processing

Proton spectra were acquired at 600 MHz and 298 K on a Bruker Avance III HD spectrometer with a 5 mm reversed TCI cryoprobe. One-dimensional free induction decays (FID) were acquired with a single 90° pulse sequence on 32K data points for a 12 ppm spectral width (acquisition time 2.3 sec), a 2 s relaxation delay and 256 scan accumulations. Signal processing was automatically performed in Mnova software including the Fourier transform with a 0.3 Hz line broadening, baseline correction, chemical shift calibration (DMSO at δ_H 2.50 ppm), spectral alignment. Spectral phasing was performed manually. The processing method also included an integral calculation of 0.001 ppm regions from 9.5 to 0 ppm (referred to as bins). The bin size was optimized considering the modelling results and the loading-plot interpretation (bin resolution v.s. spectral resolution). This generated an X2 matrix, used for multivariate statistical analysis, in which each line is a spectrum (i.e. a sample sponge extract) and each column is a bin (labeled with its mean chemical shift). For several specific samples, 2D experiments were also performed with TOCSY (DIPSI) and HSQC sequences in order to help the identification of discriminant metabolites.

Multivariate Statistical Analysis

The X1 and X2 matrices were normalized using the probabilistic method according to spectra and the normalized centered method according to variables. In a first step, PCA and O2PLS analysis were performed for the MS and the NMR data separately before using the MB methods. In, a second step, multiblock (MB) statistical analysis was performed using the two X1 (MS data) and X2 (NMR data) according to the method described by Qannari *et al.* [32].

The first MB model calculated is a non-supervised principal component analysis (CPCA) [33]. Prior to the analysis blocks scaling was performed by dividing each block by its variance. Therefore, the result does not depend on the number of variables in each block. The score plot result aim to demonstrate the homogeneity of the samples based on the variability of data. PCA reveals possible outliers that may occur in case of technical bias. PCA may also detect separation between groups. The MBPLS method is a supervised model derived from CPCA and PLS [33]. Prior to the analysis blocks scaling was performed. A score plot and a loading plot were computed to illustrate the results of the MBPLS model. The loading plot colors represent the variable weight in scores building. The intensity also represents the variable weight but rescaled to consider variable standard deviation. Spectral regions were considered to be discriminating when they corresponded to bins with a weight higher than the medium value. Each region considered in the Table 2 was checked within the original spectra for the chemical shift and multiplicity. In such a case, it was considered that the level or relative amount of the corresponding molecule was modified according to the supervising factor.

The statistical performances of the models (OPLS or MBPLS) were assessed by calculating the R^2Y fit parameter and the Q^2Y cross-validated coefficient of determination parameter. R^2Y corresponds to the variance explained by the Y matrix. Q^2Y (computed using the “leave-one-out” cross-validation method) estimates the model’s predictability. $R^2Y = 1$ indicates that the model describes the data perfectly, whereas $Q^2Y = 1$ indicates perfect predictability.

SMART-based Annotation

1H and ^{13}C chemical shifts mined from HSQC spectra corresponding to NMR discriminant variables were submitted to the online tool SMART 2.1 (<https://smart.ucsd.edu/classic>) and the substructure proposition with a cosine higher than 0.9 were taken into consideration.

Results

Workflow

The workflow strategy showing the multiblock-molecular networking-multi-informed annotation is depicted in Fig.1. For this case study, 33 samples comprising 27 different species were selected in our MNHN sponge collection. The genera were considered as the main supervising factor for statistical classification. The

geographical origin was also tested as a classification factor. This sample collection was used for an accurate investigation into the chemical composition by analyzing the metabolome of 33 marine sponges with a multiblock approach using ^1H NMR and HRMS-based data as well as taxonomically informed molecular networks (Fig. 1) as a potential complementary tool for the structural annotations of the variables that discriminate the four sponge genera of this study within the order Haplosclerida.

Multivariate Statistical Analysis of HRMS and ^1H NMR Data

To perform multivariate statistical analyzes, data were collected into three matrices. The response matrix called Y contained the designation of the four genera. The X1 matrix contained the MS data and the X2 matrix the NMR data (see Experimental section for data acquisition and preprocessing). The results of the computed model were appreciated with two statistical parameters R^2Y and Q^2Y (see material and methods section).

In the first step HRMS and ^1H NMR data were analyzed separately. When MS or NMR data were included in a principal component analysis (PCA) model, no outlier could be detected. In the same time, PCA models were neither able to separate the samples according to the genera nor the geographical origin. Using the supervised O2PLS model, both types of data, MS and NMR achieved a computed model according to genera. However, the statistical parameters, in particular the predictability Q^2 , were too low (0.45 for MS and 0.39 for NMR) to consider these models.

In a second step, the HRMS and ^1H NMR data were combined to be analyzed in the multiblock-projection on latent structure model (MBPLS). The joint data analysis by the mean of MBPLS model improved the statistical performances as the predictability of the model reached the 0.58 value and the R^2Y value of 0.95 for a calculation of seven components. The score plot obtained is shown in Fig. 2 demonstrating that the first component discriminates *Petrosia* and *Xestospongia* from *Haliclona* and *Callyspongia*. The second component discriminates *Haliclona* from *Callyspongia*. It is noteworthy that the two genera *Xestospongia* and *Petrosia* could not be separated in this model, probably this phenomenon decreases the statistical performance of the method based on classification of four groups. Interestingly, the score plot demonstrates a clear classification of the sponges according to their family rather than their genera.

The MBPLS classification obtained was based on the calculation of principal components in which the variables interfere with weights values in relation to their importance in the component equation. The loading plots, presented in Fig. 3 for the first component, shows the weights of each variable for the calculation of the component. The variables with the higher weight are considered as discriminating variables (thus m/z or chemical shifts δ). The variables with a positive value along the y axis of the loading plot are increasing when the predictive axis increases and those with a negative value along y axis are decreasing when the predictive increases. With such results, it is possible to highlight the variable responsible for discrimination of *Petrosia* and *Xestospongia* from *Haliclona* and *Callyspongia* with weight of variables on predictive axis 1 (pred1) and variables responsible for discrimination of *Haliclona* from *Callyspongia* with weight of variables on predictive axis 2 (pred2).

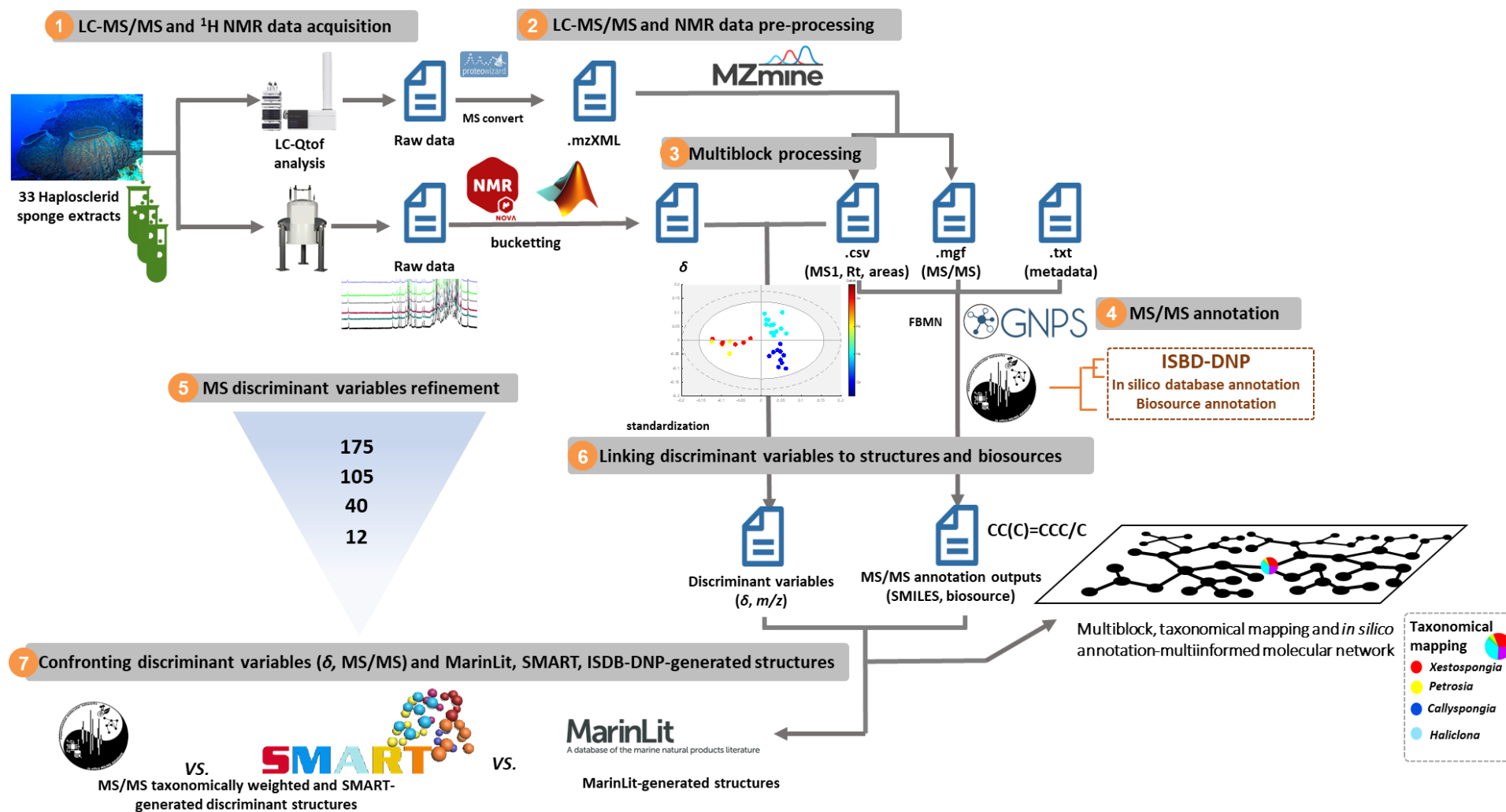


Fig. 1 Overview of the Multiblock-Molecular Networking-multi-informed annotation workflow.

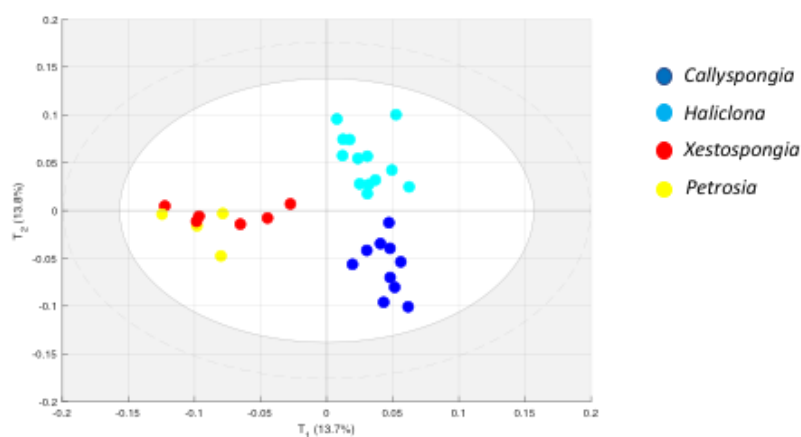


Fig. 2 Score plot of the Multiblock-PLS model computed with HRMS and ^1H NMR data. The axes correspond to the two first component scores of the model (T_1 and T_2). The percentages are the fraction of all data variance explained by each component. Each dot corresponds to data of a sample colored according to genera.

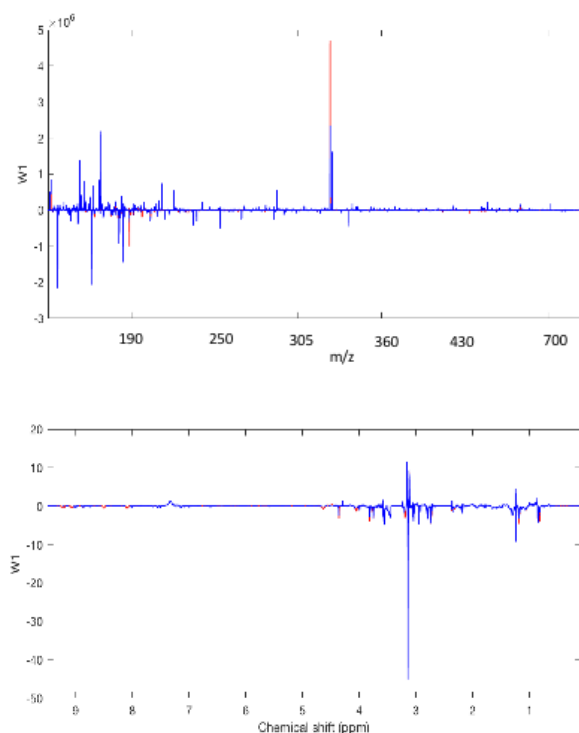


Fig. 3 Loading plots associated with the first component score (T_1). The top chart corresponds to MS data and the bottom chart to NMR data. The vertical axis represents the rescaled weight of each variable for the calculation of the T_1 . Variables presenting a y negative value increase in *Petrosia* and *Xestospongia* samples and those with a positive value increase in *Haliclona* and *Callyspongia*. A colored scale is used to highlight the discriminant variables. For weight over 0.02 (median absolute value of W1), the variable was colored in red.

At first, all variables with relevant weight on pred1 and pred2 were collected as discriminant variables. This first selection included 175 variables to be assigned to metabolites or chemical structures in the refinement process as described below.

Beside this statistical classification of the four sponge genera, the four groups were pairwise compared using MBPLS calculation. The score plots obtained are presented in Fig. 4. These score plots in relation with the Q^2Y values achieved confirmed the result of the first MBPLS calculation. Best models were those comparing *Callyspongia* to *Xestospongia* or *Petrosia* while those comparing *Haliclona* to *Xestospongia* or *Petrosia* achieved lower Q^2Y values. These models correspond to the classification along the first axis of MBPLS model. At the same time, classification of *Haliclona* vs. *Callyspongia* was not statistically reliable. It is noteworthy that *Petrosia* and *Xestospongia* could not be discriminated due to the low number of samples for these genera.

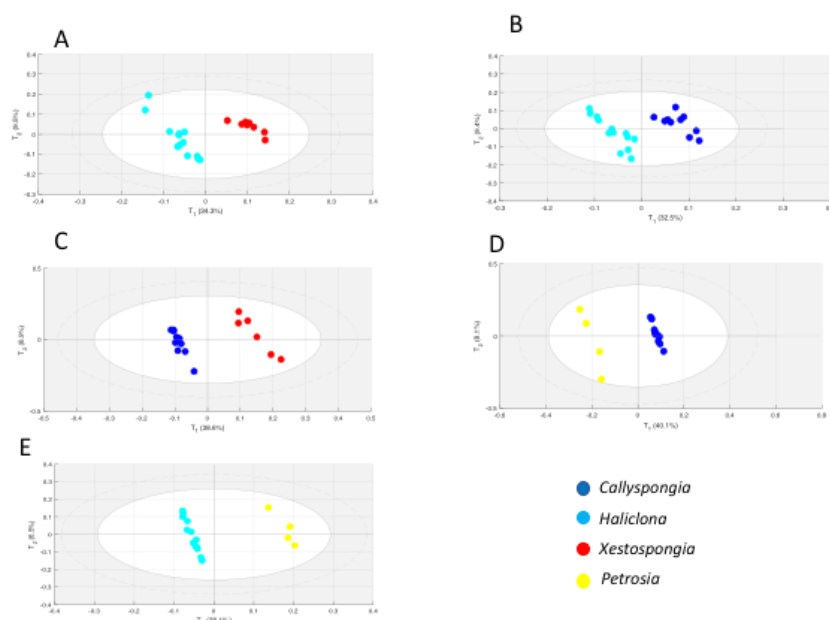


Fig. 4 Score plots of the Multiblock-PLS models computed with the MS and NMR data in order to classify genera from each other. The axes correspond to the two first component scores of the model (T_1 and T_2). The percentages are the fraction of all data variance explained by each component. Each dot corresponds to data of a sample colored according to genera. A: *Haliclona* vs. *Xestospongia*; $Q^2Y = 0.66$. B: *Haliclona* vs. *Callyspongia*; $Q^2Y = 0.395$. C: *Callyspongia* vs. *Xestospongia*; $Q^2Y = 0.764$. D: *Petrosia* vs. *Callyspongia*; $Q^2Y = 0.625$. E: *Haliclona* vs. *Petrosia*; $Q^2Y = 0.314$. For *Petrosia* and *Xestospongia* no model could be computed.

Metabolite Identification by Combining Taxonomically Informed Feature-Based Molecular Networking and NMR

Data Refinement by Blending Discriminant Variables and Biosources

The discriminant variables from the MS block revealed 175 m/z values. As a way to reduce those variables and to ascribe them reasonable molecular structure candidates, we applied a filtering strategy that leveraged the biosource information obtained from the taxonomically informed scoring feature-based molecular networking workflow [29, 34]. In this regard, the 33 sponge extracts were profiled by LC-HRMS/MS and the resulting MS/MS data were preprocessed and organized following the Feature-Based Molecular Networking (FBMN) workflow. Unfortunately, an examination of the Global Natural Product Social Molecular Networking (GNPS) [35] library hits yielded nonspecific annotations such as nucleotides, diketopiperazines and phosphocholines (https://gnps.ucsd.edu/ProteoSAFe/result.jsp?task=3012a113c5544344a8e31a156259234f&view=view_all_annotations_DB). To further extend the annotation coverage in a reliable manner, we turned toward the potential of the taxonomical scoring of *in silico* MS/MS annotations (ISDB-DNP). In brief, this strategy enables the dereplication of experimental MS/MS data against CFM-ID [36] predicted mass spectrometric data of the *Dictionary of Natural Products*. Next, the confidence of the chemical structure annotations can be further enhanced by integrating taxonomic information (biosource) of the investigated species. Thus, this strategy allowed to link discriminant variables to molecular structures and biosources. Out of these 175 variables, 105 of

them could be annotated using *in silico*-DNP (ISDB-DNP). Then, the selection of variables identified as marine biosource produced a list of 40 annotated m/z with ISDB-DNP, including sponges, bacteria and fungi. Next, this selection was restrained to sponges as biosource which ended with eight annotated features. At last, four additional features (belonging to the 175 variables) that could not be annotated by ISDB-DNP were added to the selection (Table 1).

Annotating MS Block Discriminant Variables with MarinLit and ISDB-DNP

As a way to complement the *in silico* annotations obtained with ISDB-DNP, the 12 MS block discriminant variables were searched against MarinLit using their deprotonated exact masses (± 0.01) and “Haplosclerida” as a taxonomic filter. The results of this dereplicative process are detailed in the Table 1 and Figure 5. In case of non-annotated discriminant variables, the ability of molecular networking in clustering structurally similar ions was exploited in a way that an annotated analogue belonging to the same “molecular family” was selected instead. Among, these 12 discriminant variables (Fig. 5), only six were previously identified within sponge genera investigated in this study, namely, (m/z 463.3609, sarcotride A), (m/z 455.3861, petroformynic acid), (m/z 469.324, petrosianyne A), (m/z 471.3417, 1, 12, 18, 29-triacontatetrayne-3,14,17,28-tetraol) from the genus *Petrosia*, and (m/z 449.3439, ingamines A and E), (m/z 455.3861, petrosteryl acetate), previously identified within the genus *Xestospongia*.

Confronting Discriminant NMR Chemical Shifts to MS/MS-in silico Annotated, and MarinLit-generated Structures

In the NMR block, 10 spectral regions could be observed at the higher rate in *Xestospongia* and *Petrosia* than *Callyspongia* and *Haliclona* samples (Table 2). Considering the properties of this spectroscopic method (*i.e.*, NMR), these signals can be considered as arising from relatively small molecules and with a relative high concentration in the final sample (over 50 μM). Among these signals, four chemical shifts at δ_{H} 0.82 (green stars), 3.19 and 3.82 (red stars), and 4.36 ppm (golden stars) matched some steroids, cyclitols, and polyacetylenic metabolites proposed from the MS-block dereplication (Fig. 5). Accordingly, the latter seemed to discriminate efficiently *Xestospongia* and *Petrosia* from *Callyspongia* and *Haliclona* samples. Two additional chemical shifts increased in *Haliclona* and *Callyspongia* samples, respectively at δ_{H} 3.06 (purple stars) and 3.68 ppm (blue stars), matched sesquiterpene and polyacetylenic metabolites. Contrasting with the observations obtained from the MS block annotations, some of these compounds were also increased within *Xestospongia* and *Petrosia* samples. Nevertheless, this confrontation of the MS and NMR discriminant data should be considered as complementary information rather than confirming data.

Table 1 MS block discriminant variables filtered through biosources criteria with their putative annotations dereplicated against MarinLit and ISDB-DNP.

<i>m/z</i> ^a	Analogue ^b	MarinLit annotation (±0.01) (Taxonomy) ^c	ISDB-DNP annotation (±0.02) (Taxonomy)	MS/MS spectral score	Taxonomic pondered spectral score	Axis 1: increased in	Axis 2: increased in
224.999	225.1317	Calyxolane A or B [37] (<i>Calyx podatypa</i>)	Calyxolane A or B [37] (<i>Calyx podatypa</i>)	0.34	0.51		<i>Callyspongia</i>
265.2241	251.2005	Fulvanin 1 [38] (<i>Reniera fulva</i>)	Methyl <i>trans</i> -monocyclofarnesate [39] (<i>Halichondria panicea</i>)	0.29	0.44	<i>Xestospongia</i> and <i>Petrosia</i>	
305.2191	305.2157	No hit	<u>Cinachylenic acid A</u> [40] (<i>Cinachyrella australiensis</i>)	0.30	0.45	<i>Xestospongia</i> and <i>Petrosia</i>	
325.2043		No hit	Methyl 6-methoxy-3,6-peroxyhexadeca- 4,10,12- trienoate [41] (<i>Plakortis simplex</i>)	0.28	0.42		<i>Callyspongia</i>
351.2594	351.2571	No hit	Plakortide E [42] (<i>Plakortis halichondroides</i>)	0.27	0.40	<i>Xestospongia</i> and <i>Petrosia</i>	
431.3518		No hit	Stylisterol A [43] (<i>Stylissa</i> sp.)	0.27	0.40	<i>Xestospongia</i> and <i>Petrosia</i>	
449.3439		Ingamine A [44] (<i>Xestospongia ingens</i>)	<u>-2-Methyl-2-[(3E,7E,11E)-4,8,12,16- tetramethylheptadeca-3,7,11,15- tetraenyl]chromen-6-ol</u> (<i>Ircinia</i> sp.)	0.32	0.478	<i>Xestospongia</i> and <i>Petrosia</i>	
		Ingamine E [45] (<i>Xestospongia ingens</i>)	2-Pentaprenylbenzoquinone [46] (<i>Dysidea pallescens</i>)	0.31	0.476		
			Difurospinosulin [47] (<i>Ircinia spinosula</i>)	0.2736	0.41		
463.3609	463.361	Sarcotride A [48] (<i>Petrosia</i> sp.)	Sarcotride A [48] (<i>Petrosia</i> sp.),	0.24	0.48	<i>Xestospongia</i> and <i>Petrosia</i>	
			Sarcotride D [49] (<i>Sarcotragus</i> sp.)	0.216	0.32		

Table 1 (following): MS block discriminant variables filtered through biosources criteria with their putative annotations dereplicated against MarinLit and ISDB-DNP

<i>m/z</i> ^a	Analogue ^b	MarinLit annotation (±0.01) (Taxonomy) ^c	ISDB-DNP annotation (±0.02) (Taxonomy) ^d	MS/MS spectral score	Taxonomic pondered spectral score	Axis 1: increased in	Axis 2: Increased in
455.3861	Petrosteryl acetate [50] (<i>Xestospongia</i> sp.) Petroformynic acid [51] (<i>Petrosia ficiformis</i>)	No hit				<i>Xestospongia</i> and <i>Petrosia</i>	
469.324	Petrosianyne A [52] (<i>Petrosia</i> sp.) Amphimedoside D [53] (<i>Amphimedon</i> sp.)	No hit					<i>Callyspongia</i>
471.3417	1, 12, 18, 29- Triacontatetrayne- 3,14,17,28-tetraol [54] (<i>Petrosia</i> sp.) 2-Ethoxycarbonyl-2- hydroxy-A-nor- ergosta-5,24(28)-dien- 4-one [55] (<i>Haliclona oculata</i>)	No hit				<i>Xestospongia</i> and <i>Petrosia</i>	
525.2085	No hit	No hit				<i>Xestospongia</i> and <i>Petrosia</i>	

^a These values were considered as $[M+H]^+$

^b These *m/z* values are parent masses belonging to the same cluster as the proposed variables referred in the *m/z* column

^c Annotation score was assigned Level 4: “Tentative candidate”[56]

^d Annotation score was assigned Level 5: “Unequivocal molecular formula”[56]

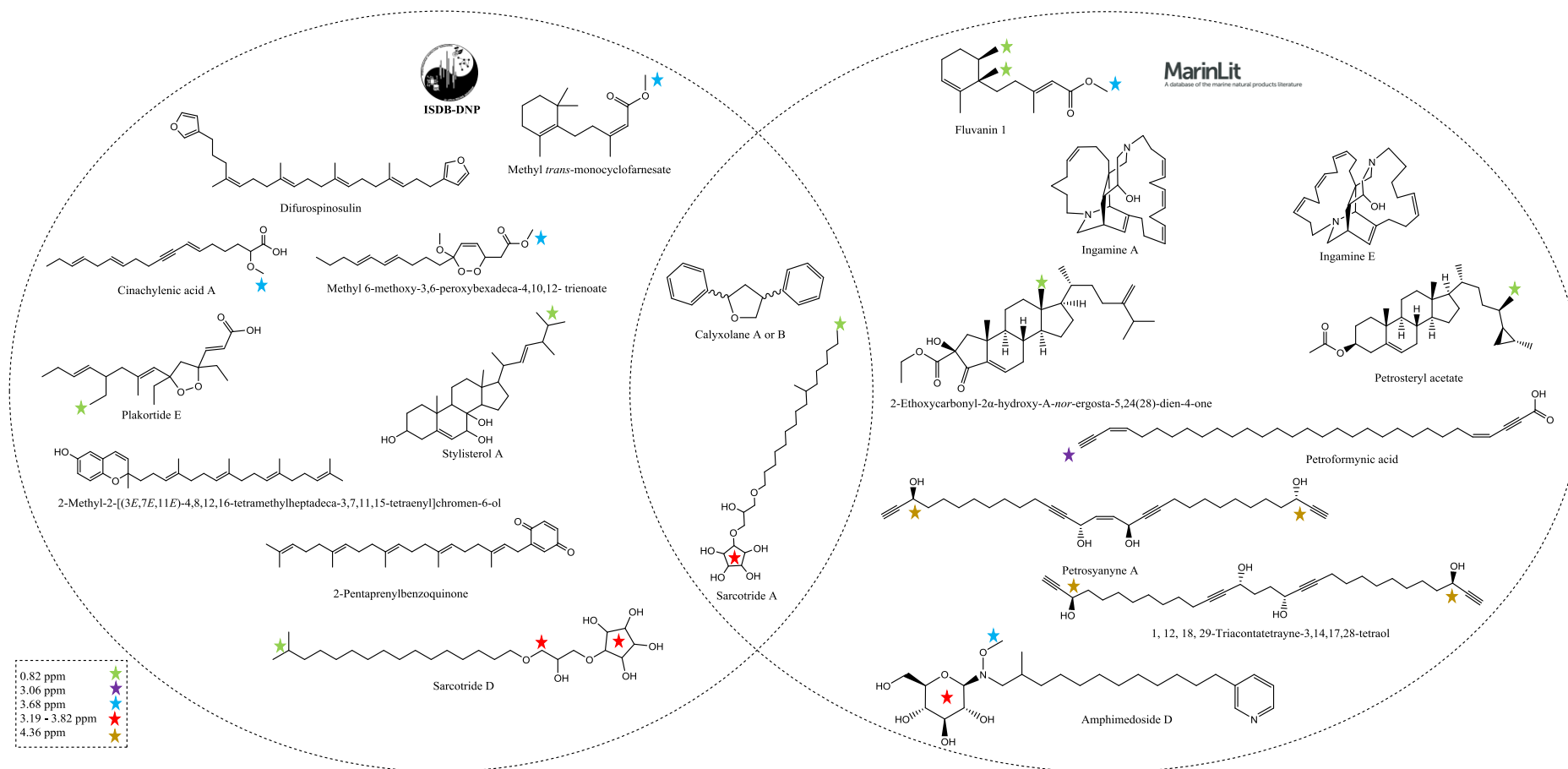


Fig. 5 Putative chemical structures obtained from ISDB-DNP (left side) and MarinLit (right side) annotations for MS block discriminant variables obtained with the statistical MB-PLS model of 33 sponge samples of the order Haplosclerida (see Table 1). The colored stars correspond to NMR block discriminant variable chemical shifts (see Table 2) (The chemical shifts were often not exactly comparable with the literature because the solvents used were not the same).

Table 2 Discriminant variables from the NMR block and 2D TOCSY and HSQC correlations relevant for bis-alkyl pyridinium identification.

Discriminant chemical shift on axis 1 (ppm)	Increased in:	Discriminant chemical shift on axis 2 (ppm)	Increased in:	TOCSY (δ_H/δ_H)	HSQC (δ_H/δ_C)	Increased in:
0.82	<i>Xestospongia</i> and <i>Petrosia</i>	2.94	<i>Haliclona</i>			
3.19	<i>Xestospongia</i> and <i>Petrosia</i>	2.97	<i>Haliclona</i>			
3.82	<i>Xestospongia</i> and <i>Petrosia</i>	3.06	<i>Haliclona</i>			
4.06	<i>Xestospongia</i> and <i>Petrosia</i>	3.11	<i>Haliclona</i>			
4.36	<i>Xestospongia</i> and <i>Petrosia</i>	3.68	<i>Callyspongia</i>			
4.6	<i>Xestospongia</i> and <i>Petrosia</i>					
8.10	<i>Xestospongia</i> and <i>Petrosia</i>			8.09/8.48	8.07/127.34	<i>Xestospongia</i> and <i>Petrosia</i>
8.50	<i>Xestospongia</i> and <i>Petrosia</i>			9.01/8.48	8.48/144.86	<i>Xestospongia</i> and <i>Petrosia</i>
9.10	<i>Xestospongia</i> and <i>Petrosia</i>			9.01/8.09	9.01/142.19	<i>Xestospongia</i> and <i>Petrosia</i>
9.28	<i>Xestospongia</i> and <i>Petrosia</i>					
				8.02/8.74	8.02/126.50	<i>Petrosia</i>
				8.74/8.86	8.75/143.64	<i>Petrosia</i>
				8.86/8.02	8.86/144.38	<i>Petrosia</i>

Among those 10 discriminant NMR signals, the deshielded chemical shifts between δ_H 8 and 9 ppm (Table 2), that allowed to discriminate *Petrosia* and *Xestospongia* from *Haliclona* and *Callyspongia*, attracted our attention. Owing to their peculiar chemical shifts, we first thought that structures, obtained from the MS block discriminant variables (Fig. 5), that may match those values will be easily identified. Unfortunately, none of them (Fig. 5) were compatible with such values. As a way to address this issue, we used the recently introduced tool SMART [57] to annotate structurally the above-mentioned NMR discriminant variables by acquiring additional HSQC data. This dereplicative process annotated alkylpyridinium substructure using the values: $\delta_{H/C}$ [8.07/127.34, 8.48/144.8, 9.01/142.1]. Satisfyingly, this annotation was further confirmed using TOCSY experiment that highlighted the spin system that encompasses the NMR discriminant variables between δ_H 8 and 9 ppm (Table 2).

Discussion

This study reported a unique strategy for combining 1H NMR and MS datasets with molecular networking for discriminating *Petrosia* and *Xestospongia* (family Petrosiidae) from *Haliclona* (family Chalinidae) and *Callyspongia* (family Callyspongiidae) sponges. Sponges of the order Haplosclerida of genera *Xestospongia*, *Petrosia*, *Callyspongia* and *Haliclona* have been widely chemically investigated; the genus *Xestospongia* being the most prolific as illustrated in Electronic Supplementary Material Fig. S2.

Our first goal was to propose a methodology to discriminate sponges from the same order according to their genera. Therefore, we have investigated a set of 33 samples obtained from four genera of the order Haplosclerida. To these samples analyzed with MS and NMR, the multiblock statistical analysis was applied and these particular multivariate statistics were effective in our set of samples to classify three out of four groups. Using multiblock bilinear factorization algorithms that capitalize on the availability of blocking information, we achieved greater levels of model interpretability with the NMR and MS data than available from single-block PCA and PLS methods.

As a way to annotate MS block discriminant variables, the acquired mass spectra were processed and annotated following the taxonomically informed feature-based molecular networking workflow. Furthermore, the biosource data obtained from this workflow enabled to filter the discriminant variables efficiently. We also demonstrated the combined use of SMART NMR and *in silico* MS/MS data for the rapid and accurate identification of discriminant variables. Despite a relatively low number of samples, we were able to statistically discriminate three groups out of the four genera, revealing molecules not necessarily reported as the most abundant in the order Haplosclerida. Table 1 merges two levels of interpretation of our data, an experimental one with the results of the multi-block statistical model and a predicted one with MS/MS-based *in silico* database (ISDB-DNP) and MarinLit annotations. As a second goal, the help of molecular networking of MS data and 2D NMR experiments, allowed a deeper interpretation of data. With the use of the discriminant spectral regions (i.e. spectral regions having the highest importance in the calculation of the model), we were able to propose putative structural annotations that may differentiate the three groups. Those molecules, summarized in Table 1 and Fig. 5., highlighted 12 discriminant variables. As shown in Table 1, using MS/MS *in silico* annotation, all the cosine scores are lower than 0.34, which is commonly observed in workflows using *in silico* MS/MS annotation [30, 58]. In addition, the predictability over 0.5 is an acceptable threshold in multivariate statistical analysis. This value is reasonable taking into account that *Petrosia* and *Xestospongia* are definitively not discriminated. On another hand, the fact that the four samples of the genus *Petrosia* did not produce a separated cloud ensures that the model was not over-fitted, as this type of algorithm may be susceptible of over-fitting problems. For clarity, we decided to exclusively dedicate Table 1 to MS and MS/MS data and Table 2 to NMR data.

According to the compound identification workgroup of the Metabolomics Society, the structures that have been proposed using ISDB (*in silico* database), were assigned as level 4 (tentative candidate)[56]. For MarinLit dereplication results, obtained using HRMS data only, annotation level was assigned as level 5 (unequivocal molecular formula)[56]. Although Fig. 5 reveals several natural product structures, we must remind that the aim of this study was not to identify all key chemical markers of the order Haplosclerida but to select those which we expected to discriminate the four genera by leveraging the complementary information provided by MS and NMR data. Four out of the 12 discriminate variables, namely sarcotride A, petroformynic acid, petrosianyne A, 1, 12, 18, 29-triacontatetrayne-3,14,17,28-tetraol were previously identified from the genus *Petrosia*, and three additional ones, ingamines A and E as well as petrosteryl acetate, were previously identified within the genus *Xestospongia*. Surprisingly, calyoxanes A and B, previously reported from *Calyx podatypa*, a sponge classified within the Oceanapiidae family in the Haplosclerida order. As this putative discriminant molecule was obtained from annotation of both databases (i.e., ISDB-DNP and MarinLit) with the best MS/MS spectral score in ISDB-DNP, we can be confident about this annotation. On the other hand, the fact that these compounds might have been found in high quantities in *Calyx podatypa*, does not mean that they are absent from other sponges. It has to be kept in mind that with the metabolomic statistical model, the discriminant molecules are not necessarily the most abundant in the group. It should rather be expressed as “is more frequently found” or “in a higher quantity in *Callyspongia*”. Regarding the presence of peroxide derivatives, namely methyl 6-methoxy-3,6-peroxyhexadeca-4,10,12- trienoate and plakortide E, previously identified from sponges of the order Homosclerorpha, *Plakortis simplex* and *P. halichondroides*, respectively, we must admit that this annotation was obtained only using the ISDB-DNP (cosine score = 0.28 and 0.27, respectively). Yet, peroxides have already been reported from *Callyspongia* sp. [59]. Further analysis of Fig 5 and Table 1 revealed that, on the one hand, sesquiterpenes, steroids, cyclitols, polyprenyl furans and quinones metabolites seem to discriminate efficiently *Xestospongia* and *Petrosia* from *Callyspongia* and *Haliclona* samples. On the other hand, diphenyl butanoids seem to be characteristic of *Callyspongia* sponges. Remarkably, according to these MS-based annotations, polyacetylenic and alkylpyridine compounds failed to discriminate these genera since they were increased in

Callyspongia, *Petrosia* and *Xestospongia* samples. Accordingly, 3-alkylpyridine and 3-alkylpiperidine alkaloids are compounds typically found across Haplosclerida sponges and are considered as taxon-specific metabolite class [60]. Concerning polyacetylenic compounds, recent reports suggest that a thorough investigation of these metabolites found within and outside of Haplosclerida is necessary to evaluate their taxonomic specificity [60].

Conclusions

In this study, the confrontation of MS and NMR block variables allowed us to identify steroids, cyclitols and bis-alkylpyridinium metabolites that discriminate *Xestospongia* and *Petrosia* from *Haliclona* and *Callyspongia*. As Haplosclerida marine sponges encompass a broad diversity of species that produce pharmaceutically valuable bioactive compounds, metabolomics can allow a better understanding of the phylogeny and diversity within this group, which is of major importance to understand morphological character evolution in demosponges. The classification obtained here with two different analytic platforms and a reduced number of Haplosclerida samples shows that the metabolic and chemical variability making metabolomics a possible tool to facilitate Haplosclerida classification. However, the variability that may be assigned to associated micro-organisms, seasonality or life stage aspects could not be investigated with this set of samples. These factors deserve further experiments to support a definitive taxonomic classification using the workflow proposed in this study. Such an implemented MBPLS model could help for classification of an unknown Haplosclerida sample. At last, the strategy described herein may be applied to support others marine organism classification.

Supplementary Materials: Table S1: List of the 33 samples of Haplosclerida marine sponges analyzed with their reference in our MNHN sponge collection and their geographical localization. Figure S1: (A): ¹H spectrum obtained with the *Petrosia hoecksemai* NS94 sample (DMSO and H₂O signals have been withdrawn) and (B): TOCSY enhancement of the region between 7 and 10 ppm. Figure S2: According to MarinLit database, number of articles and isolated compounds from *Xestospongia*, *Petrosia*, *Callyspongia* and *Haliclona* genera (April, 2022).

Data Availability Statement: The raw data files, and preprocessed peak lists related to the LC-MS/MS analysis were deposited on the public MassIVE repository under the accession number: MSV000088423. The molecular network could be accessed on this link: <https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=3012a113c5544344a8e31a156259234f>

Acknowledgments and Funding Information: We thank Professor Jean-Claude Braekman for using his sponge sample's collection given to the Museum National d'Histoire Naturelle. We also thank the Muséum National d'Histoire Naturelle de Paris (ATM) and CNRS- LIA FVLC-TEL for financial support.

Declarations

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Letertre MPM, Dervilly G, Giraudeau P. Combined nuclear magnetic resonance spectroscopy and mass spectrometry approaches for metabolomics. *Anal Chem.* 2021;93(1):500-18.
2. Deng L, Gu H, Zhu J, Nagana Gowda GA, Djukovic D, Chiorean EG, et al. Combining NMR and LC/MS using backward variable elimination: metabolomics analysis of colorectal cancer, polyps, and healthy controls. *Anal Chem.* 2016;88(16):7975-83.
3. Wolfender J-L, Nuzillard J-M, van der Hooft JJJ, Renault J-H, Bertrand S. Accelerating metabolite identification in natural product research: toward an ideal combination of liquid chromatography–high-

- resolution tandem mass spectrometry and NMR profiling, in silico databases, and chemometrics. *Anal Chem.* 2019;91(1):704-42.
4. Keifer PAK, M. E. S.; Schwartz, R. E.; Hughes, R. G., Jr.; Rittschof DR, K. L. . Presented at the 26th Interscience Conference on Antimicrobial Agents and Chemotherapy, 1986.
 5. Bingol K, Bruschweiler-Li L, Yu C, Somogyi A, Zhang F, Bruschweiler R. Metabolomics beyond spectroscopic databases: a combined MS/NMR strategy for the rapid identification of new metabolites in complex mixtures. *Anal Chem.* 2015;87(7):3864-70.
 6. Bingol K, Bruschweiler R. NMR/MS translator for the enhanced simultaneous analysis of metabolomics mixtures by NMR spectroscopy and mass spectrometry: application to human urine. *J Prot Res.* 2015;14(6):2642-8.
 7. Wang C, Zhang B, Timári I, Somogyi Á, Li D-W, Adcox HE, et al. Accurate and efficient determination of unknown metabolites in metabolomics by NMR-based molecular motif identification. *Anal Chem.* 2019;91(24):15686-93.
 8. Wang C, He L, Li D-W, Bruschweiler-Li L, Marshall AG, Bruschweiler R. Accurate identification of unknown and known metabolic mixture components by combining 3D NMR with fourier transform ion cyclotron resonance tandem mass spectrometry. *J Prot Res.* 2017;16(10):3774-86.
 9. Edison AS, Colonna M, Gouveia GJ, Holderman NR, Judge MT, Shen X, et al. NMR: unique strengths that enhance modern metabolomics research. *Anal Chem.* 2021;93(1):478-99.
 10. Mishra P, Roger J-M, Jouan-Rimbaud-Bouveresse D, Biancolillo A, Marini F, Nordon A, et al. Recent trends in multi-block data analysis in chemometrics for multi-source data integration. *TrAC Trends Anal Chem.* 2021;137:116206.
 11. Li X, Luo H, Huang T, Xu L, Shi X, Hu K. Statistically correlating NMR spectra and LC-MS data to facilitate the identification of individual metabolites in metabolomics mixtures. *Anal Bioanal Chem.* 2019;411(7):1301-9.
 12. Fox Ramos AE, Evanno L, Poupon E, Champy P, Benididir MA. Natural products targeting strategies involving molecular networking: different manners, one goal. *Nat Prod Rep.* 2019;36(7):960-80.
 13. Benididir MA, Kang KB, Genta-Jouve G, Huber F, Rogers S, van der Hooft JJJ. Advances in decomposing complex metabolite mixtures using substructure- and network-based computational metabolomics approaches. *Nat Prod Rep.* 2021;38(11):1967-93.
 14. Duperron S, Benididir MA, Durand S, Longeon A, Duval C, Gros O, et al. New benthic cyanobacteria from guadeloupe mangroves as producers of antimicrobials. *Mar Drugs.* 2019;18(1):16.
 15. Carroll AR, Copp BR, Davis RA, Keyzers RA, Prinsep MR. Marine natural products. *Nat Prod Rep.* 2020;37(2):175-223.
 16. Hooper JNA, Van Soest RWM in *Systema Porifera 2002*, Kluwer Academic /Plenum Publishers, New York.
 17. Cárdenas P, Pérez T, Boury-Esnault N. Chapter two - Sponge systematics facing new challenges. In: Becerro MA, Uriz MJ, Maldonado M, Turon X, editors. *Adv Mar Biol.* 61: Academic Press; 2012. p. 79-209.
 18. de Voogd NJ, Alvarez B, Boury-Esnault N, Carballo JL, Cárdenas P, Díaz M-C, Dohrmann M, Downey R, Hajdu E, Hooper JNA, Kelly M, Klautau M, Manconi R, Morrow CC, Pisera AB, Ríos P, Rützler K, Schönberg C, Vacelet J, van Soest RWM, (2021). World Porifera Database. Accessed at <http://www.marinespecies.org/porifera> on 2021-11-24. doi:10.14284/359.
 19. McCormack GP, Erpenbeck D, Van Soest RWM. Major discrepancy between phylogenetic hypotheses based on molecular and morphological criteria within the Order Haplosclerida (Phylum Porifera: Class Demospongiae). *J Zool Sys Evol Res.* 2002;40(4):237-40.
 20. Redmond NE, McCormack GP. Large expansion segments in 18S rDNA support a new sponge clade (Class Demospongiae, Order Haplosclerida). *Mol Phylogenetics Evol.* 2008;47(3):1090-9.
 21. Redmond NE, Raleigh J, van Soest RWM, Kelly M, Travers SAA, Bradshaw B, et al. Phylogenetic relationships of the marine haplosclerida (Phylum Porifera) employing ribosomal (28S rRNA) and mitochondrial (cox1, nad1) gene sequence data. *PLOS ONE.* 2011;6(9):e24344.
 22. Tribalat M-A, Marra MV, McCormack GP, Thomas OP. Does the chemical diversity of the order haplosclerida (Phylum Porifera: Class Demospongia) fit with current taxonomic classification? *Planta Med.* 2016;82(09/10):843-56.
 23. Morrow C, Cárdenas P, Boury-Esnault N, Picton B, McCormack G, Van Soest R, et al. Integrating morphological and molecular taxonomy with the revised concept of Stelligeridae (Porifera: Demospongiae). *Zool J Linn Soc.* 2019;187(1):31-81.
 24. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol.* 2012;30:918-20.
 25. Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform.* 2010;11(1):395.

26. Myers OD, Sumner SJ, Li S, Barnes S, Du X. One step forward for reducing false positive and false negative compound identifications from mass spectrometry metabolomics data: new algorithms for constructing extracted ion chromatograms and detecting chromatographic peaks. *Anal Chem.* 2017;89(17):8696-703.
27. Lagarde A, Mambu L, Mai P-Y, Champavier Y, Stigliani J-L, Beniddir MA, et al. Chlorinated bianthrone from the cyanolichen *Nephroma laevigatum*. *Fitoterapia.* 2021;149:104811.
28. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-504.
29. Rutz A, Dounoue-Kubo M, Ollivier S, Bisson J, Bagheri M, Saesong T, et al. Taxonomically informed scoring enhances confidence in natural products annotation. *Front Plant Sci.* 2019;10(1329).
30. Allard P-M, Péresse T, Bisson J, Gindro K, Marcourt L, Pham VC, et al. Integration of molecular networking and in-silico MS/MS fragmentation for natural products dereplication. *Anal Chem.* 2016;88(6):3317-23.
31. MarinLit. <http://pubs.rsc.org/marinlit/> Accessed Nov 2021. [
32. Qannari EM, Courcoux P, Vigneau E. Common components and specific weights analysis performed on preference data. *Food Qual Pref.* 2001;12(5):365-8.
33. Wehrens R, van der Linden WE. Bootstrapping principal component regression models. *J Chemometr.* 1997;11(2):157-71.
34. Nothias L-F, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, et al. Feature-based molecular networking in the GNPS analysis environment. *Nat Methods.* 2020;17(9):905-8.
35. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol.* 2016;34(8):828-37.
36. Allen F, Pon A, Wilson M, Greiner R, Wishart D. CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic Acids Res.* 2014;42:W94-9.
37. Rodríguez AD, Cobar OM, Padilla OL. The calyxolanes: new 1,3-diphenylbutanoid metabolites isolated from the caribbean marine sponge *Calyx podatypa*. *J Nat Prod.* 1997;60(9):915-7.
38. Casapullo A, Minale L, Zollo F. Paniceins and related sesquiterpenoids from the mediterranean sponge *Reniera fulva*. *J Nat Prod.* 1993;56(4):527-33.
39. Cimino G, de Stefano S, Minale L. Methyltrans-monocyclofarnesate from the sponge *Halichondria panicea*. *Experientia.* 1973;29(9):1063-4.
40. Mokhlesi A, Hartmann R, Kurtán T, Weber H, Lin W, Chaidir C, et al. New 2-methoxy acetylenic acids and pyrazole alkaloids from the marine sponge *Cinachyrella* sp. *Mar Drugs.* 2017;15(11):356.
41. Rudi A, Talpir R, Kashman Y, Benayahu Y, Schleyer M. Four new C16 1,2-dioxene-polyketides from the sponge *Plakortis aff. simplex*. *J Nat Prod.* 1993;56(12):2178-82.
42. Patil AD, Freyer AJ, Bean MF, Carte BK, Westley JW, Johnson RK, et al. The plakortones, novel bicyclic lactones from the sponge *Plakortis halichondrioides*: activators of cardiac SR-Ca²⁺-pumping ATPase. *Tetrahedron.* 1996;52(2):377-94.
43. Mitome H, Shirato N, Hoshino A, Miyaoka H, Yamada Y, Van Soest RWM. New polyhydroxylated sterols stylisterols A–C and a novel 5,19-cyclosterol hatomasterol from the Okinawan marine sponge *Stylissa* sp. *Steroids.* 2005;70(1):63-70.
44. Kong F, Andersen RJ, Allen TM. Ingamines A and B, new cytotoxic alkaloids from the marine sponge *Xestospongia ingens*. *Tetrahedron.* 1994;50(21):6137-44.
45. Kong F, Andersen RJ. Ingenamine alkaloids isolated from the sponge *Xestospongia ingens*: Structures and absolute configurations. *Tetrahedron.* 1995;51(10):2895-906.
46. Cimino G, De Luca P, De Stefano S, Minale L. Disidein, a pentacyclic sesterterpene condensed with an hydroxyhydroquinone moiety, from the sponge *Disidea pallescens*. *Tetrahedron.* 1975;31(3):271-5.
47. Cimino G, De Stefano S, Minale L. Polyprenyl derivatives from the sponge *Ircinia spinosula*: 2-Polyprenylbenzoquinones, 2-polyprenylbenzoquinols, prenylated furans and a C-31 difuranoterpene. *Tetrahedron.* 1972;28(5):1315-24.
48. Kim D-K, Lim YJ, Kim JS, Park JH, Kim ND, Im KS, et al. A cyclitol derivative as a replication inhibitor from the marine sponge *Petrosia* sp. *J Nat Prod.* 1999;62(5):773-6.
49. Liu Y, Jung JH, Xu T, Long L, Lin X, Yin H, et al. New cyclitol derivative from a sponge *Sarcotragus* species. *Nat Prod Res.* 2011;25(6):648-52.
50. Cheng Z, Liu D, de Voogd NJ, Proksch P, Lin W. New sterol derivatives from the marine sponge *Xestospongia* sp. *Helv Chim Acta.* 2016;99(8):588-96.
51. Guo Y, Gavagnin M, Salierno C, Cimino G. Further petroformynes from both atlantic and mediterranean populations of the sponge *Petrosia ficiformis*. *J Nat Prod.* 1998;61(3):333-7.
52. Juan Y-S, Lee C-C, Tsao C-W, Lu M-C, El-Shazly M, Shih H-C, et al. Structure elucidation and cytotoxic evaluation of new polyacetylenes from a marine sponge *Petrosia* sp. *Int J Mol Sci.* 2014;15(9):16511-21.
53. Takekawa Y, Matsunaga S, van Soest RWM, Fusetani N. Amphimedosides, 3-alkylpyridine glycosides from a marine sponge *Amphimedon* sp. *J Nat Prod.* 2006;69(10):1503-5.

54. Masamitsu O, Saori A, Akira T, Hiyoshizo K, Yoshiyasu F, Kozo S. Bioactive polyacetylenes from the marine sponge *Petrosia* sp. Chem Lett. 1994;23(1):89-92.
55. Yu S, Deng Z, Proksch P, Lin W. Oculatol, oculatolide, and a-nor sterols from the sponge *Haliclona oculata*. J Nat Prod. 2006;69(9):1330-4.
56. Blaženović I, Kind T, Ji J, Fiehn O. Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. Metabolites. 2018;8(2):31.
57. Reher R, Kim HW, Zhang C, Mao HH, Wang M, Nothias L-F, et al. A convolutional neural network-based approach for the rapid annotation of molecularly diverse natural products. J Am Chem Soc. 2020;142(9):4114-20.
58. Klein-Júnior LC, Cretton S, Allard P-M, Genta-Jouve G, Passos CS, Salton J, et al. Targeted isolation of monoterpene indole alkaloids from *Palicourea sessilis*. J Nat Prod. 2017;80(11):3032-7.
59. Toth SI, Schmitz FJ. Two new cytotoxic peroxide-containing acids from a new guinea sponge, *Calyspongia* sp. J Nat Prod. 1994;57(1):123-7.
60. Galitz A, Nakao Y, Schupp PJ, Wörheide G, Erpenbeck D. A soft spot for chemistry—current taxonomic and evolutionary implications of sponge secondary metabolite distribution. Mar Drugs. 2021;19(8):448.