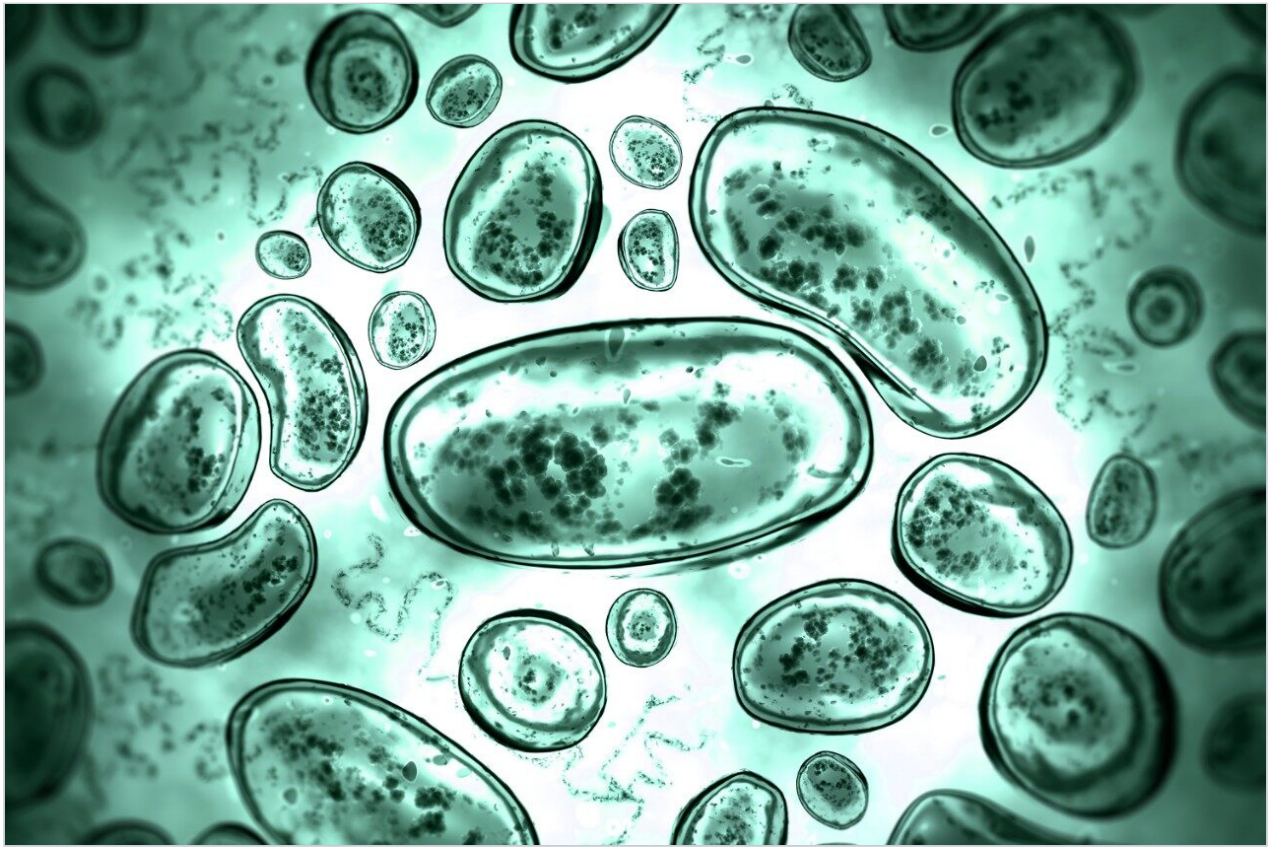
Waters™

Nota applicativa

Microbial Databases for Rapid Screening and Dereplication of Microbial Natural Products Using UPLC-QTof-MS Coupled to Novel Informatics Platform

Giorgis Isaac, Jimmy Yuk, Mark Wrona, Robert S. Plumb

Waters Corporation, Department of Chemistry, Simon Fraser University



This is an Application Brief and does not contain a detailed

Experimental section.

Abstract

This application brief describes the use of UNIFI Scientific Information Software with the Waters scientific microbial database and online ChemSpider libraries such as Marinlit, for confident compound identification from complex natural product microbial samples.

Benefits

Custom and online microbial natural product databases coupled with UNIFI Scientific Information System Software that allows confident compound identification from complex samples.

Introduction

In natural products (NPs) discovery research, crude extracts of various origins such as plants, marine organisms, and microorganisms, containing thousands of metabolites have to be characterized, either as part of bioactive guided isolation studies for drug discovery or in the frame of a metabolomics investigation for biomarker identification.[1] Despite the advances in hardware and the development of numerous derivatization, labeling, and analytical methods, unambiguous compound identification remains one of the major bottlenecks in natural products discovery.[1,2] Known NPs are frequently re-identified during NP discovery research and having a known microbial compounds database can aid in the elimination of known active compounds for further consideration.

Bioinformatics tools are becoming essential in natural products research, as advances in experimental throughput and the complexity of data obtained from biological profiling can make manual interpretation and compound identification a difficult task.[2] Here we describe the use of UNIFI Scientific Information Software with the Waters scientific microbial database and online ChemSpider libraries such as Marinlit, for confident compound identification from complex natural product microbial samples. The custom microbial database was developed in collaboration with Prof. Roger Linington from Simon Fraser University.

Results and Discussion

Microbes were isolated from marine sediment, grown under standard fermentation conditions with XAD-16 resin, extracted with 1:1 methanol/dichloromethane, and fractionated on a reverse phase ACQUITY UPLC C_{18} Column. The chromatographic separation was carried out on an ACQUITY UPLC BEH C_{18} Column (2.1 mm x 50 mm, 1.8 µm) in 4.5 min at a temperature of 50 °C. Mobile phase consisted of eluent A (0.1% formic acid in water) and eluent B (0.1% formic acid in acetonitrile) with a flow rate of 0.8 mL/min. Both ESI^+ and ESI^- experiments were performed with data acquired over a range of *m/z* 50 to 1800. A representative positive ion total ion chromatograms of the early and late eluting different marine microbial fractions are shown in Figure 1.

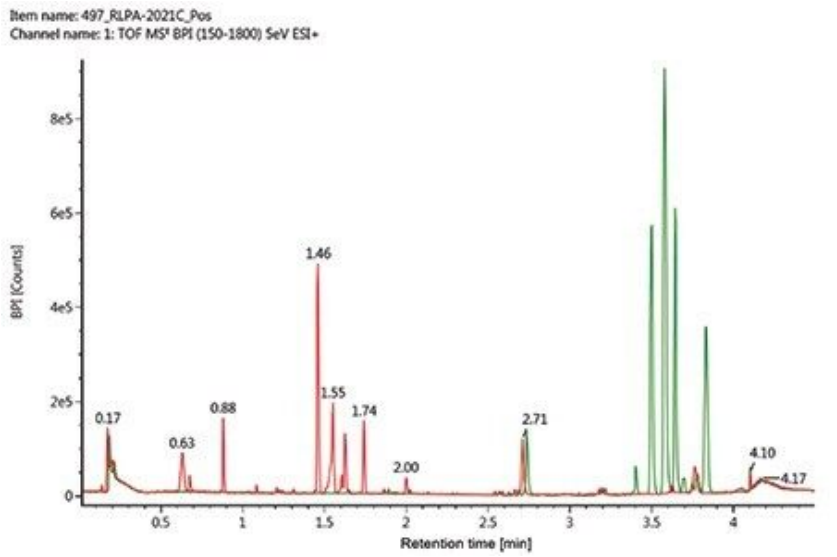


Figure 1. Representative overlaid UPLC-MS chromatograms of the early (red) and late (green) eluting different marine microbial fractions in positive ion mode.

In any non-targeted discovery experiment, the goal is to identify as many compounds as possible. However, compound identification has always been a challenging step due to sample complexity and the absence of relevant microbial databases. In order to address this issue, a natural product analytical workflow was used to analyze the mass spectra of marine extracts using a custom microbial natural product database with precursor exact mass, fragment ion information, and theoretical isotopic distribution that allows confident identification of compounds from a complex sample. As shown in Figure 2, it is simple and straight forward to

create a compound database in the UNIFI Scientific Library. In order to create a UNIFI Scientific Library, .mol files are required for each of the compounds of interest and an Excel spreadsheet containing the name of the compound of interest, molecular formula, the name of the corresponding .mol file and item tag (such as compound class and Latin name) were created in the same folder. Other measured parameters such as fragment ion, retention time, and collision cross section can also be added to the library. Then, the spreadsheet was imported into UNIFI as a scientific library container (.ULC) to create the microbial database with a total of 439 compounds. Detailed step-by-step explanation on how to create a scientific library using UNIFI is provided elsewhere.[3] The basic infrastructure of the UNIFI microbial database is shown in Figure 3 with compound name, chemical formula, structure, average molecular mass, monoisotopic mass, and item tag of each selected compounds. Figure 4 shows the compound identification result from the UNIFI microbial database. A data-independent acquisition (DIA) was used in order to acquire both precursor exact ion in low energy and corresponding fragment ion spectra in high energy from one injection. Figure 4, Panel A, shows the component table that lists all components that are identified from the UNIFI microbial library. For each identified compound, the extracted ion chromatogram (Figure 4, Panel B), low energy parent ion spectra (Figure 4, Panel C), and corresponding high energy spectra (Figure 4, Panel D) producing likely structural fragment ion matches are displayed. UNIFI compares and matches predicted in silico fragments with high energy experimentally derived fragments.

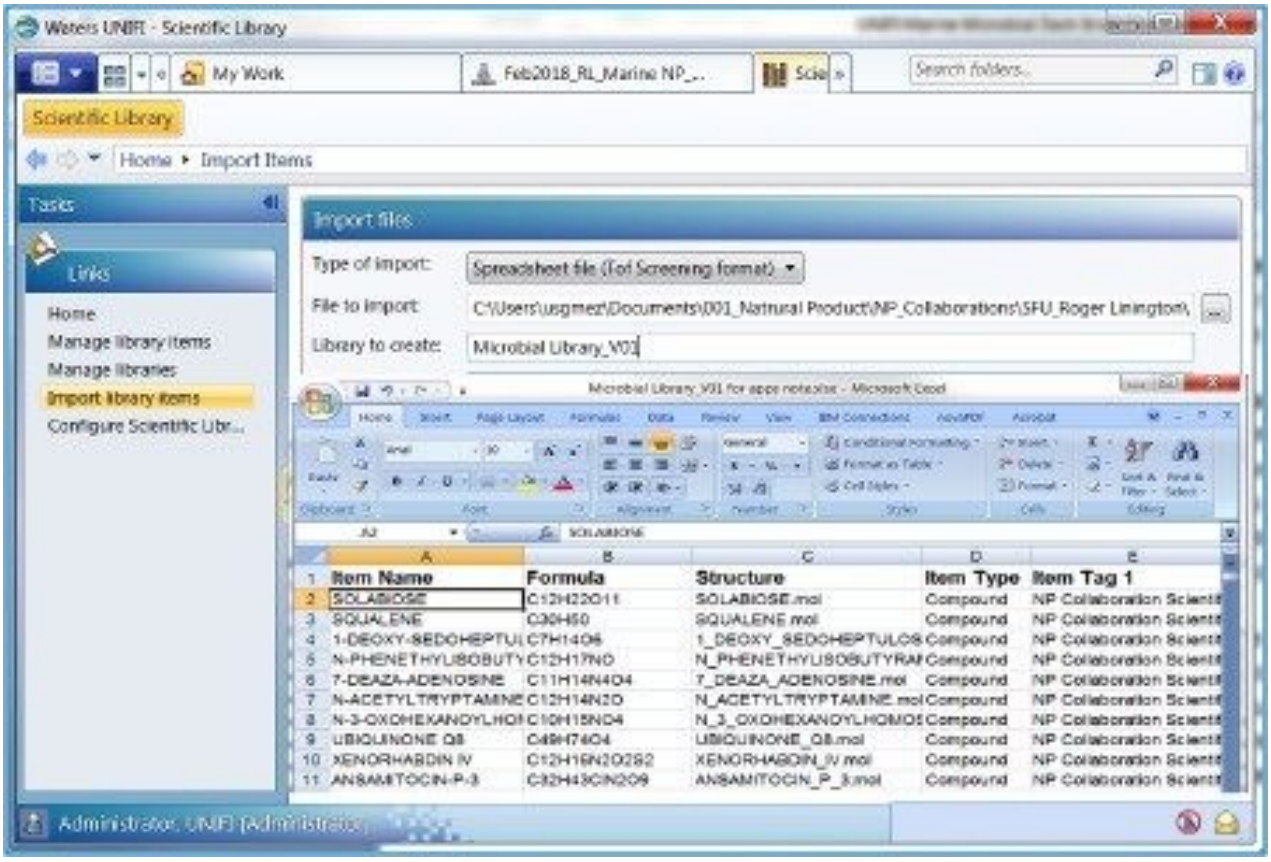


Figure 2. Creating a compound database in the UNIFI Scientific Library. In addition to item name, formula, structure, and item tag; other measured parameters such as fragment ion, retention time, and collision cross section can also be added to the library.

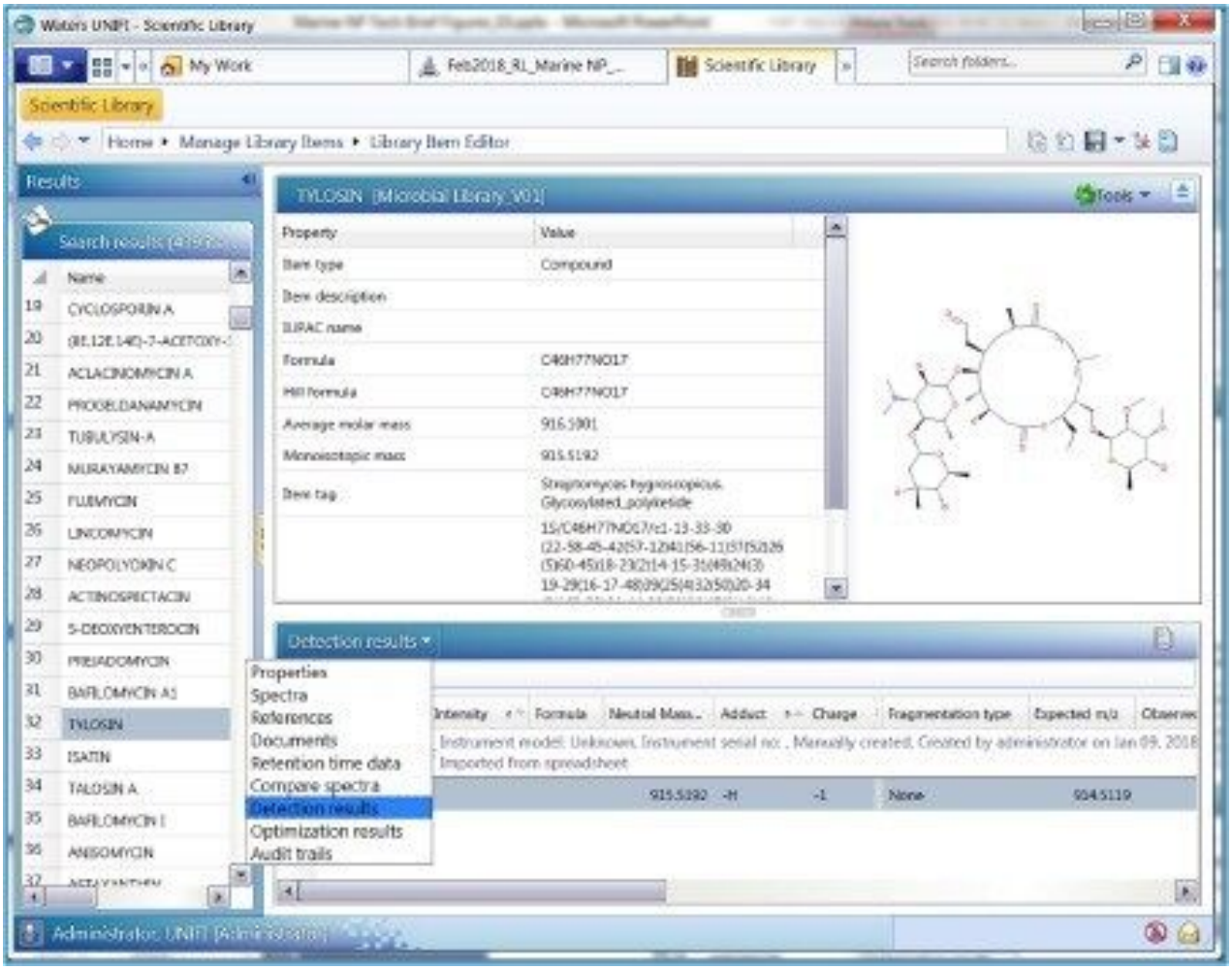


Figure 3. The basic infrastructure of the microbial natural products library in UNIFI showing the chemical formula, average molecular mass, mono-isotopic mass, chemical structure, and item tag of the selected compound.

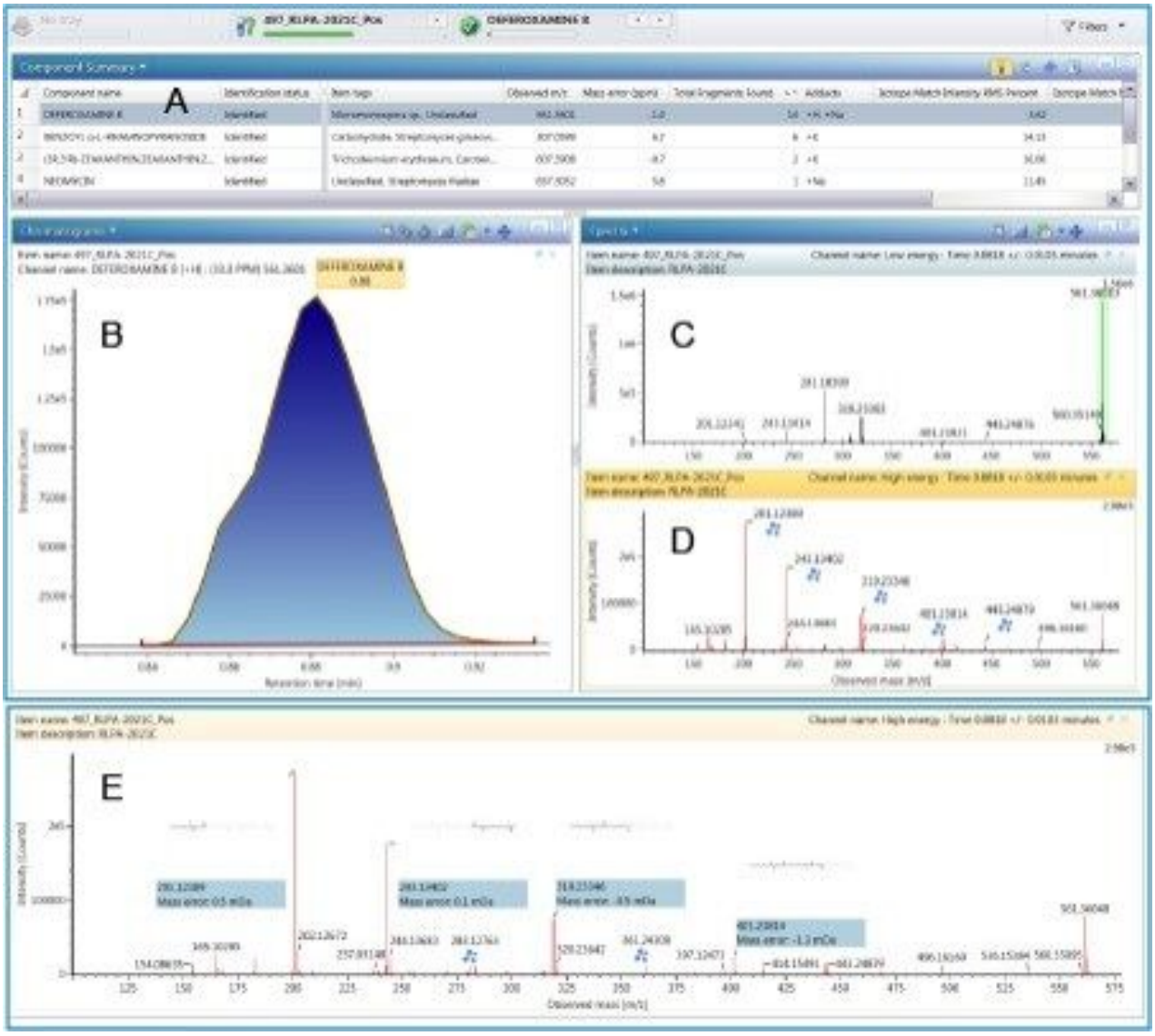


Figure 4. Compound identification result from the UNIFI microbial database. (A) The component summary interface; (B) Selected ion chromatogram of single component corresponding to panel A (C) The respective low energy precursor exact mass spectrum and (D) The corresponding high energy fragment ion spectra of Deferoxamine B with 14 matched fragment ion spectra. The blue mark indicates the experimental fragment ion that matches to the expected in-silico fragment ions generated from the mol structure using MassFragment.(E) Zoom-in high energy display of likely structural fragmentation matches for Deferoxamine B.

All components not identified from the microbial natural product database were then listed as unmatched peaks in the result browser. UNIFI provides various identification tools for the unmatched peaks. One of the options is a discovery toolset which utilizes elemental composition determination, online library searches

(such as ChemSpider), and fragment ion matching. ChemSpider is a web-based chemical structure database with access to over 58 million structures from hundreds of data sources.[4] UNIFI provides a direct connection to the ChemSpider libraries. Some of the widely used marine natural product related libraries within ChemSpider include MarinLit, Marine Drugs, and NIST.

The discovery toolset search can be launched by selecting the parent ion of interest from the list of candidate masses (796.5204 *m/z*) using the "Elucidate" module (Figure 5A). This method makes use of the ChemSpider web services, automatically exporting data from UNIFI to ChemSpider for searching according to the parameters selected, importing the results, and assigning them against the correct compounds within the software. Figure 5B shows the ranked top identification result for the candidate mass 796.5204 *m/z*. The proposed structure was found to have a total of 16 matched high energy fragment ions which indicates a possible correct identification. Structurally related compounds can be also identified by utilizing tools provided from the UNIFI platform such as common neutral mass search and common fragment ion search to connect compounds from a given structural family.

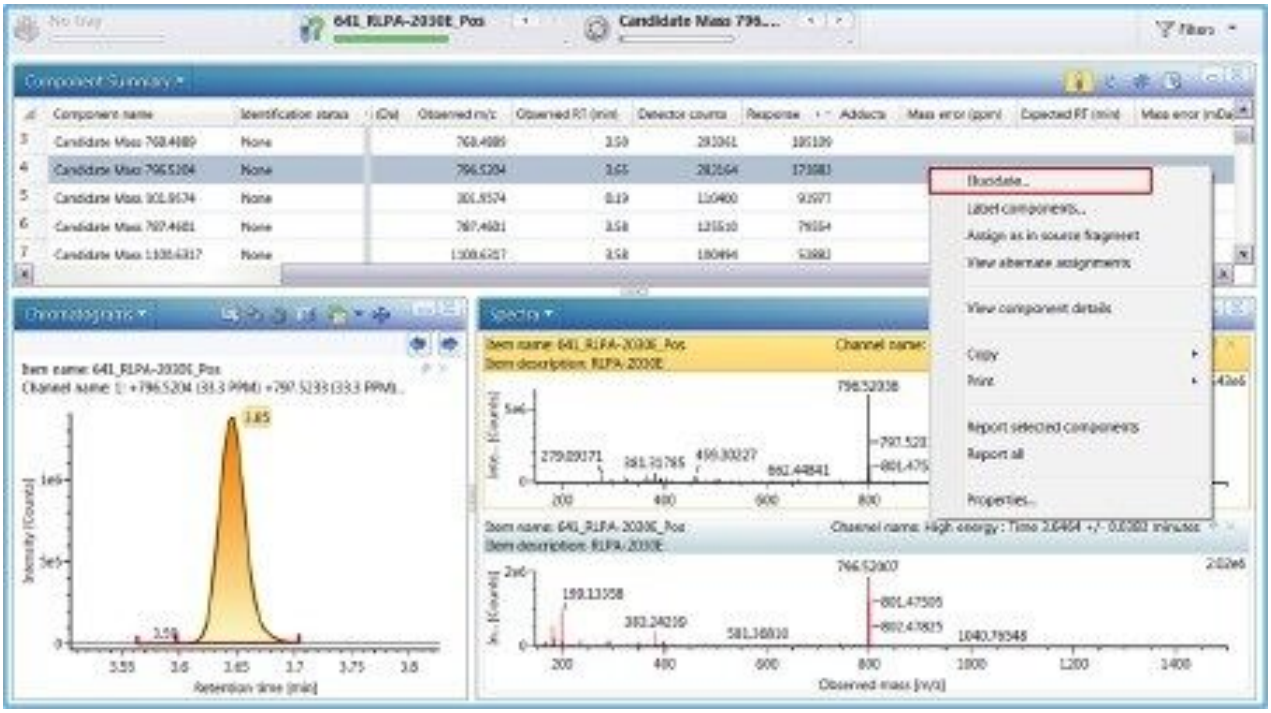


Figure 5A. Candidate mass at 796.5204 m/z and 3.65 min for structural elucidation. The discovery toolset search can be launched by selecting the parent ion from the list of candidate masses using the "Elucidate" module.

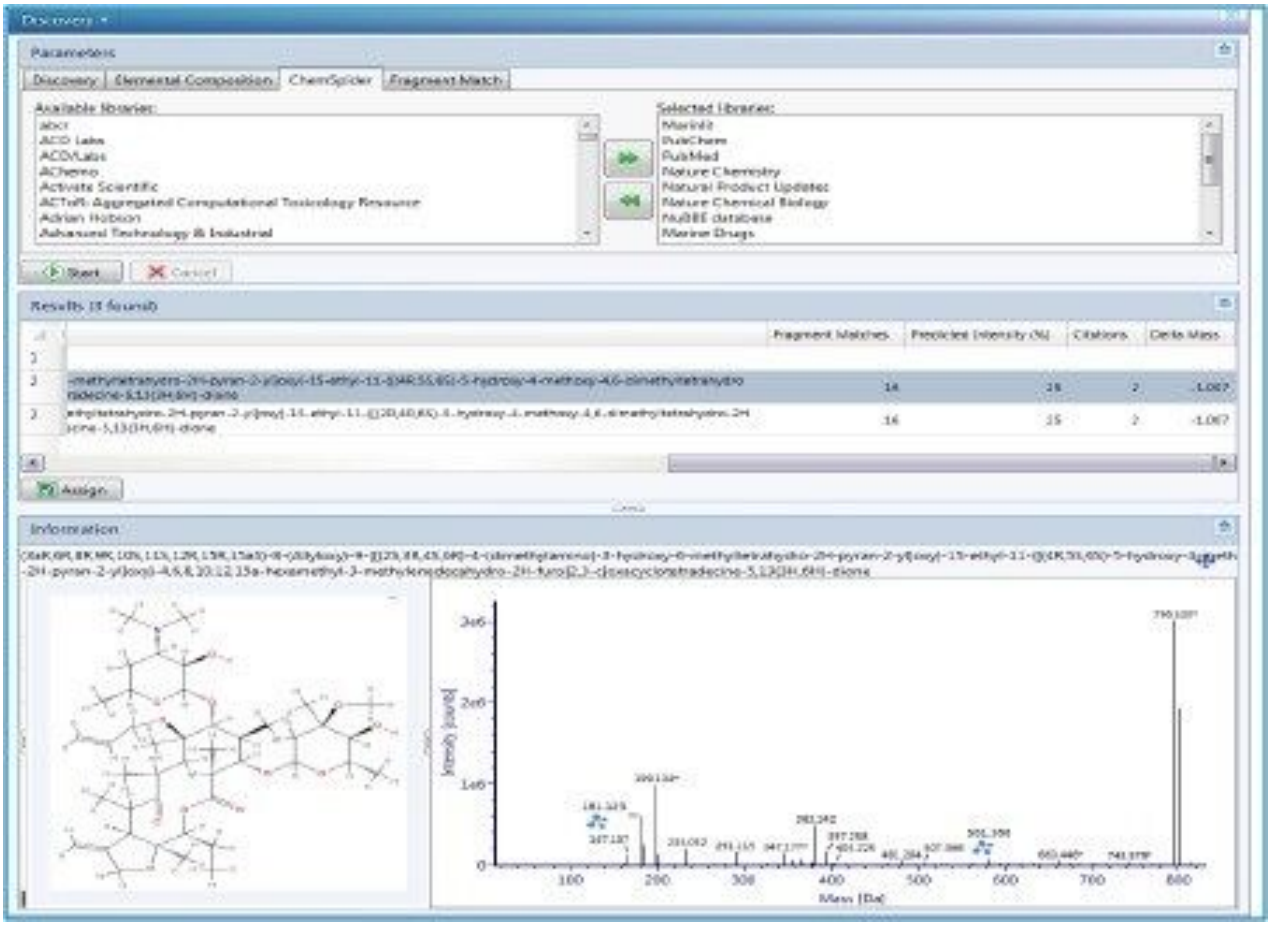


Figure 5B. Structural elucidation for the identification and confirmation of the unmatched component at 796.5204 m/z. The proposed structure was found to have a total of 16 matched high energy fragment ions with 1.007 ppm precursor mass error which indicates a possible correct compound assignment.

Conclusion

Custom and online marine microbial databases were used for high-throughput and rapid automated identification of marine microbial compounds using high-resolution UPLC-QTof-MS technology coupled to UNIFI informatics platform. The use of data independent full spectra acquisition with alternating collision energy provides confident compound identification based on low energy precursor exact mass and corresponding high energy fragment ion information from the custom microbial database. All additional masses of interest not identified from the custom microbial natural product database were then identified using the discovery toolset search which utilizes elemental composition determination, ChemSpider online

library search with Marinlit, and fragment ion matching.

References

1. Allard PM, Péresse T, Bisson J, Gindro K, Marcourt L, Pham VC, Roussi F, Litaudon M, and Wolfender JL; "Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication", *Anal Che*m. 2016;88(6):3317–3323.
2. Kurita LK and Linington RG; "Connecting Phenotype and Chemotype: High-Content Disovery Strategies for Natural Product Research"; *J. Nat. Prod*. 2015, 78, 587–596.
3. Mullin L, Cleland G, Graham K, Shah D, and Burgess J; "Building a UNIFI Scientific Library for HRMS Screening Experiments". Waters Application Note. 720004927EN. 2014.
4. http://www.chemspider.com/

Featured Products

UNIFI Scientific Information System <https://www.waters.com/134801359>

ACQUITY UPLC System <https://www.waters.com/514207>

720006412, October 2018