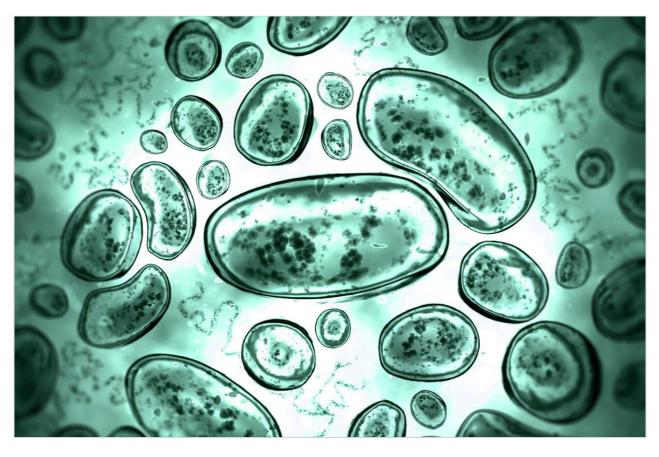
Waters[™]

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

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.

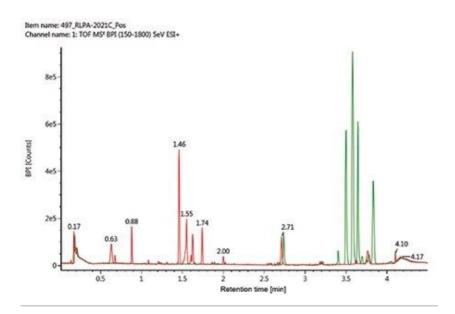


Figure 1. Representative overlaid UPLC-MS chromatograms of the early (red) and late (green) eluting different marine microbial fractions in positive ion mode.

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.

🗄 🕶 🔠 🔹 o 🔬 My W	ork	A Feb2018_RL_Marine NP	🏙 Scie *	Search folders	P 🗇
cientific Library 1 🖒 🍷 Home + Import	Items				
36	Import files				
Links Home Manage library items Manage libraries	File to import Library to create:	Spreadsheet file (Foi Screen C-Wsers\usgmez\Document Microbial Ubrary_V01		<pre>/P_Collaborations/;</pre>	
Import library items Configure Scientific Ubr			9 8 - 14 - 1 di fermat 14 di	does november Au lend homewaking * (2*3 two factor * (2*0 des * (2))?	
Emport library items	Kare Bart	Ange Layout Assessed Data	Previous View Bind Contract () General - () Contract () 8 - 16 - () Strender () 50 - () Strender () Contract () Contr	does november Au lend homewaking * (2*3 two factor * (2*0 des * (2))?	······································
Emport library items	Kare Bart	Nage Layout Auswahl Other (M)	Previous View Bind Contract () General - () Contract () 8 - 16 - () Strender () 50 - () Strender () Contract () Contr	Sons nonartal na Internation - Pro Internation - Pro Internation - Pro International I	1000 8 - 7 1001 8 - 27 A 1001 8 - 27 A 1001 8 - 1001 8 1001 8 - 1001 8

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.

	📷 🕄 🔹 🔕 My Wo	rk	_≜_ Feb200	8_BL_Marine NP	Scientific L	itrary »	Secroh folders	P
Só	entific Library							
	Thome + Manage	Library Ibens • Li	bory Hem Edit	Df.				G 1 H - 4 C
les	ults	-	liorobia (Bara)					Catook - +
4	6			10000				Shops
	Search results (419 no a	Property		Value		-		
d,	Name	tare type		Compound				
÷.	CYCLOSPORIN A	Dem description					3 1	1
į.	98.128.140-7-ACETON-5	LIPAC name					2	1
í.	ACLACINOM/ICIN A	Formula		C46H77NO17			3 1-1	1
2	PROCEDANANYCH	Hill formula		C46H77NO17			744	L
ř	TUBLAYSIN-A	Average molar r	nets	916.5001			NO	- ALL
	MURAYAMIYEN 87	Menoisotapic #	ante	915.5192			AL '	
5	FURMICIN	Den tap		Streptomyces Glycosylated_p	hygroscopicus. solyitetide		.75	1
5	LINCOMISCIN	-			c7/e1-13-33-30	2		
1	NEOPOLYCKIN C	6			57-12041/56-11(57/5212 x2114-15-35(49x24(3)	10		
8.	ACTINOSPECTACIN			19-29(16-17-4	8)39(25)4132(50)20-34			
,	5-DEOXYENTEROCIN	(The second sec	attend in		Calla	and the second sec		EX
5	PREADOMYON	Detection re	anse				the second s	B
È	BARLOMYCIN AS	Properties Spectra	_	111			121	10 82 3
2	TVLOSEN	References	Intensity	e · Formula Mes	attel Mass. Adduct	t- Charge	Fragmentation type	Expected m/u Observe
5	ISATIN	Documents Retention time da		ent model: Unknown d from spreadsheet		x. Mensally c	reated, Greated by adir	ninistrator on Jan 09, 200
1	TALOSIN A	Compare spectra	ra mpone	o non spierospeet				-
5	BARE OMYCEN I	Detection results			915.5392 -H	-1	None	934,5119
	ANISOMYCIN	Optimization resu Audit trails	its					
	ACTAVINTION	The second second						
1	×	141						18

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.

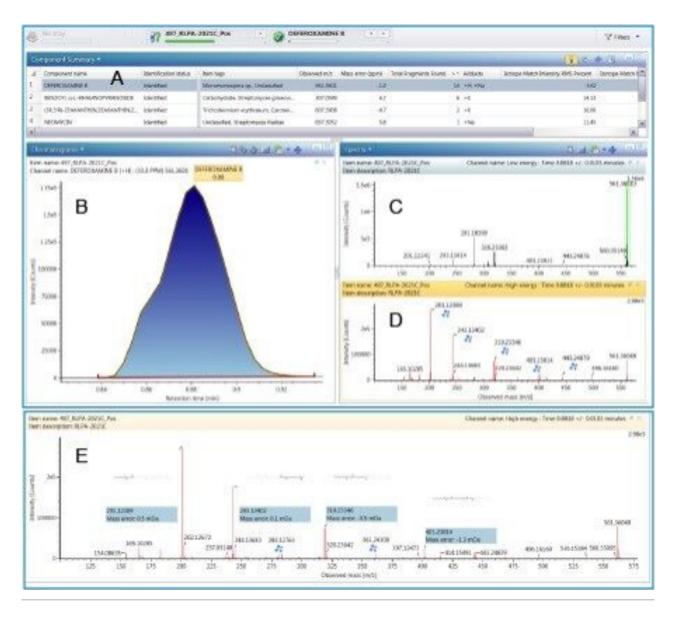


Figure 4. Compound identification result from the UNIFI microbial database. (A) The component summary interface; (B) Selected ion chromatogram of single component corresponding to panel A (C) The respective low energy precursor exact mass spectrum and (D) The corresponding high energy fragment ion spectra 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.

mooned Summary *								🕦 t. 🕷 (8 C
Component name	Identification strata	- inter	Observed m/	c Obwried RT (m	n) Denector co	ита Акаротан и	- Ablacts Mass	error (gom) Expected PE (min) Mas	a ensr (m
Candidate Mass 768,4989	None	1.000	768.40	0069 1	150	83062 3053	89		
Candidate Marc 795 5104	None		796.52	204 (165 3	82664 1739	е (Bucidate.	- 1
Canadidate Visio 102.0174	None		301.85	574 0	149	10400 909	91	Label camponews.	
Candidate Mass 707.4681	None		767,46	iai i	158	25530 795	EA	Aution as in source fragment	
Candidate Visio 1100-6317	Note	1.	1:00.62	105		538	93	View absentite antigometrix	- 1
and all the second s	Management of the second second second	-			1004				
оторатк т	19-20 H 10-1	÷	C121	seedin 🕋	2100			Yew companient details	
	and a first state			ten sane Gil, Rif.	A-2032E Pos		Channel name:	View component details	
m name: 641, FLPA-30306, Par			• •	tem name (41, RIA tem description, RU	A-2032E Pos		Channel name: 796/52238		
m same: 641,8194-30306,9ts	3 MMU +797 51331333		• •	tem name (41, RIA tem description, RU	A-2032E Pos		2.220.020	Creav	
m came: 641, FLPA-30301, Por aned same: 1 +7965204 (3).			• •	tem name 641, Rif. Tem description, RJ 1	A-3031 <u>8</u> Aos PA-20008		2.220.020	Crepy Inux Report selected components	
n came: 641,8124-30301,Por aned same 1: +7965204 (3).	3 MMU +797 51331333		•	tem name (41, RIA tem description, RU	A-2032E Pos	16.3027 501.448	796.52036	Creav	
n came: 641,8124-30301,Por sned same 1: +7965204 (3).	3 MMU +797 51331333		•	tem name 641, RJA tem description, RJJ	A-3031 <u>8</u> Aos PA-20008	96.30227 652.448 606	796.52036	Copy Nove Neport selected components Report all	
n came: 641, 8124-30301, Por sned same 1: +7965304 (3).	3 MMU +797 51331333			Sen sare (40, 40) ten deoription (10)	A-3038 Aos PA-2008 281.20785	and the second	796.52038 =797.522 41 -801.475 800	Crepy Inux Report selected components	•
n came: 641, 8124-30301, Por sned same 1: +7965304 (3).	3 MMU +797 51331333			ben name 64, 816 ten devription RD	A-3038, Pos PA-2008 R81.70785 480 A-2008, Pos	and the second	79652038 -790.522 41 -001.475 810 Channel name Hi	Copy Nov. Neport selected components Report all Properties.	rune -
n came: 641, F124-30306, Par	3 MMU +797 51331333			tern name 641, 816 tern description, RU	A X000 Aos FA 2000E 181.70785 480 A 2000E Aos FA 2000E	and the second	796.52038 =797.522 41 -801.475 800	Copy Nov. Neport selected components Report all Properties.	•
n came: 641,8124-30301,Por sned same 1: +7965204 (3).	3 MMU +797 51331333			2000 1000 641, 416 2000 description 100	A X000 Aos FA 2000E 181.70785 480 A 2000E Aos FA 2000E	and the second	79652038 -790.522 41 -001.475 810 Channel name Hi	Crepy Nex Neport selected companients Report all Properties gh energy: Time 2.6464 +/- 0.0202 and	rune -

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.

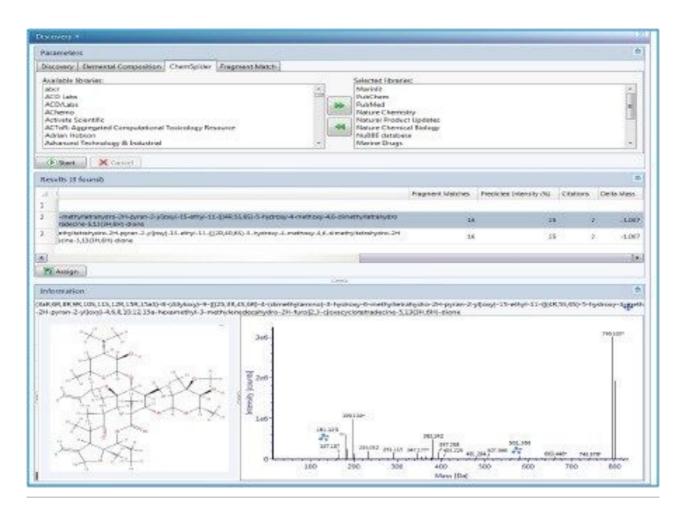


Figure 5B. Structural elucidation for the identification and confirmation of the unmatched component at 796.5204 m/z. The proposed structure was found to have a total of 16 matched high energy fragment ions with 1.007 ppm precursor mass error which indicates a possible correct compound assignment.

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

- 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 Chem.* 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

©2019 Waters Corporation. All Rights Reserved.