

Computer and Database Tools for the Identification of Novel Lead Compounds in Natural Products Extracts

Paul Klausmeyer*, T. Luke Simmons, Thomas G. McCloud
SAIC-Frederick, Inc., National Cancer Institute at Frederick, Frederick, MD

Abstract

A unique software configuration was assembled to facilitate the discovery of novel bioactive compounds from natural product extracts.

MassLynx/OpenLynx/FractionLynx v3.5

- Simultaneous acquisition of LC-MS, UV, ELSD and fluorescence data.
- Extracts and exports peak spectra into user-defined searchable libraries.
- Purification of bioactive compounds by selective isolation of targeted ion peaks.
- Enables sharing of processed chromatographic data with other laboratories through network connections.

ISISBase/ISISDraw v.2.3

- Structures can be manipulated in order to rationalize molecule fragmentation phenomenon.
- Substructures created in Draw can be exported to Base.
- Allows merging of Beilstein STR, CN, INP data, LC/MS data, and biological data into a unified structure searchable database.

Supported by NCI Contract NO1-CO-12400

Introduction

The NCI is collaborating with SAIC-Frederick, Inc. to discover antibiotics from natural sources for the purpose of treating secondary infections in immunocompromised individuals. Our approach to drug discovery is heavily dependent on robotic sample preparation, high throughput biological screening and fast chromatographic methods. Chromatographic data generated in the Natural Products Laboratory is used to fingerprint an extract, which can be depicted through the correlation of antibiotic activity to a peak on a chromatogram. An effort is made to rapidly identify the active compound by mining the HPLC-MS data for clues that might lead to a structure. To date, we have identified, to the level of chemistry or exact structure, about 40 bioactive compounds, as well as several compounds that may be novel in structure.

The fractionation and biological screening of the NCI's natural products library has resulted in the creation of a vast store of raw data, which may be viewed locally on workstations running MassLynx, Millennium 32, and Excel software. Interpretation of the data in the absence of a unifying interface has been time-consuming, and has been restricted primarily to the chromatographers. We are attempting to use ISIS/Base software to merge in-house results and commercial databases in order to improve accessibility and searchability of the data, which we believe will lead to a greater rate of novel compound discovery. The utility of the ISIS database is here presented in the context of potential drug discovery from extracts of the fungus *Aspergillus fumigatus*.

Materials and Methods

a. Chromatographic Analysis

The multi-detector HPLC system consists of a Waters 600 pump controlled by Waters Millennium 32 software and a Waters 2700 autosampler which allows for samples to be loaded in either vials or microtiter plate format. Eluate from the column is split initially 115:1 with a high precision LCP splitter which directs 0.8% volume to a Kratos fluorescence detector, a Waters 996 Photodiode Array Detector, and a Waters Micromass API Electrospray Mass Spectrometer. The bulk of the flow goes to a 25:1 LCP splitter, with the major flow directed to an ISCO Foxy fraction collector which collects in microtiter plate format. The remainder of the flow is directed into a SEDEK75 Evaporative Light Scattering Detector. This arrangement gives 95% sample recovery. For routine analysis, 50 mg of crude extract is dissolved in acetonitrile/water (1:1) at 10 mg/ml and injected onto a 250 mm x 21 mm Dynamax 60A-C18 phase bonded column eluted with acetonitrile/20 mM NH₄OAc pH 4.0 at 15 ml/min under gradient conditions 0-5 min (30:70), 25 min (50:50), 45-60 min (100:0). A typical total volume of eluate is 645 ml which is collected into 4 x 24 well microtiter plates. These are reformatted by a TECAN High Speed Robotic Liquid Handling System into two 96 well master microtiter plates with 2 ml in each well, plus 15 replicate daughter plates at 250 ul in each well. Three of these daughter plates are 96 well polypropylene flat bottom containing one 5 mm filter paper disc in each well. All plates are dried for several hours in a Savant Speed Vac to remove the majority of the organic solvent, then placed into dry ice, and when solidly frozen, into a freeze drier. Once thoroughly dried, these fractions are suitable for biological evaluation, while the bulk of the sample mass is retained in 96 well plate format for further chromatography or spectroscopy. All the data is captured electronically, and following biological evaluation, a combined physical/biological data chart is assembled. When the biological test data is received, the chemical complexity of the active wells is determined by a universal analytical HPLC method. A 50x21 cm, 3.5 µm, 100Å Waters C18 Xterra column is heated to 60°C and eluted under an acetonitrile-20 mM NH₄OAc pH4 (aq) gradient at 0.8 ml/min. Retention times and mass spectra of the unknowns are compared to those of standard compounds. Known biologically active compounds are eliminated from further study.

b. Mass-Directed Isolation and Purification

Mass directed autopurification is achieved through implementation of Waters FractionLynx v3.5 Application Manager controlled ZQ 2000 Mass Detector, 2767 Sample Manager, Delta 600 high pressure pump and 2896 Photodiode Array Detector. User-defined acquisition parameters are entered into a MassLynx v3.5 sample list. Threshold limits for signal intensity of the target mass ion(s) relative to base line intensity (ex. $10^3/10^3$ -fold higher) are defined within the FractionLynx Fraction Editor. During data acquisition, the data channel monitoring function screens the mass chromatogram in real time and triggers the 2767 Sample Manager to deposit the desired analyte into a collection vessel. Selected ion fractions are indicated according to appropriate mass target, retention time and collection vessel.

c. Construction of the Isis Databases

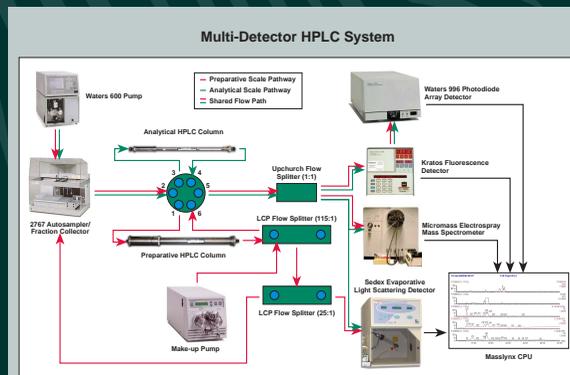
Beilstein natural products data and structures were downloaded in SD file format from the NCI library and provided to the Natural Products Support Group. In order to more readily mine the NCI's wealth of biological data, a link from the Beilstein data to the NCI's pure compound repository was created. This was accomplished by exporting the registry number field from the Beilstein database to the NCI's enhanced data browser, followed by the download of the NSC number. The NSC number was then uploaded to the Beilstein database via the registry number link.

A second ISIS database was created to provide access to critical chromatographic and biological data from a single-chromatogram system. The data channel monitoring function screens the mass chromatogram in real time and triggers the 2767 Sample Manager to deposit the desired analyte into a collection vessel. Selected ion fractions are indicated according to appropriate mass target, retention time and collection vessel.

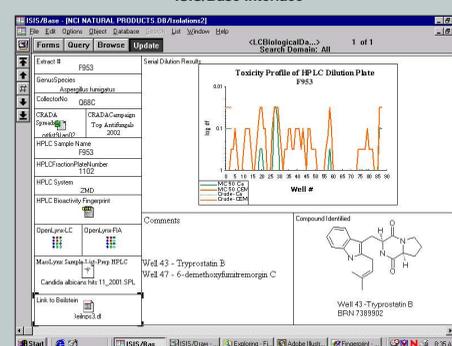
During data acquisition, the data channel monitoring function screens the mass chromatogram in real time and triggers the 2767 Sample Manager to deposit the desired analyte into a collection vessel. Selected ion fractions are indicated according to appropriate mass target, retention time and collection vessel.

During data acquisition, the data channel monitoring function screens the mass chromatogram in real time and triggers the 2767 Sample Manager to deposit the desired analyte into a collection vessel. Selected ion fractions are indicated according to appropriate mass target, retention time and collection vessel.

During data acquisition, the data channel monitoring function screens the mass chromatogram in real time and triggers the 2767 Sample Manager to deposit the desired analyte into a collection vessel. Selected ion fractions are indicated according to appropriate mass target, retention time and collection vessel.

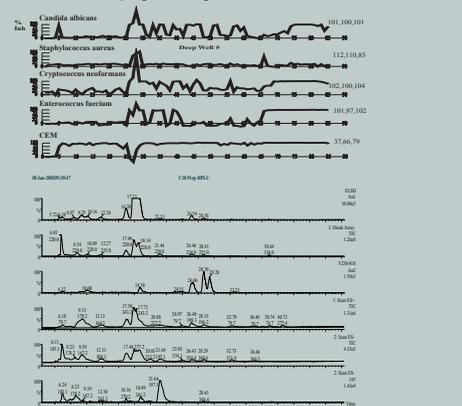


ISIS/Base Interface



The ISIS/Base serves as the departure point for accessing the numerous data files generated in the drug discovery process.

Natural Products Support Group HPLC Trace/Biological Activity Fingerprint for a Natural Products Crude Extract F953 *Aspergillus fumigatus*, C18 HPLC



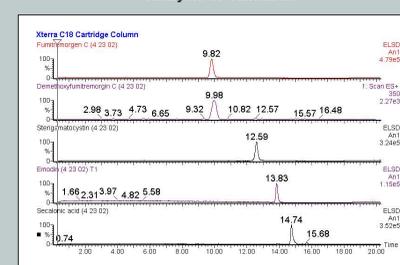
Search Results from the Beilstein ISIS/Base

Known Compounds Isolated from *Aspergillus*

	CN	BRN	NSC
	Gliotoxin	50675	77672
	Fumitremorgin C	4586854	719655
	Citreoviridin	5311865	159630
	Secalonic Acid D	1279816	159631
	Sterigmatocystin	1299457	201423

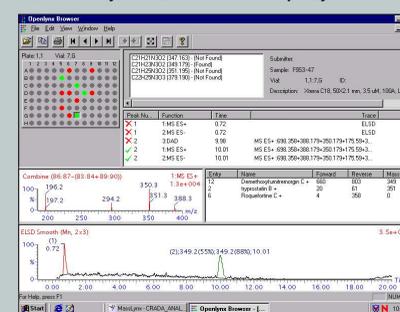
A preliminary search of the Beilstein ISIS/Base generates a table of structures previously isolated from genus *Aspergillus*. Known active compounds may be obtained from the NCI repository for LC-MS analysis and inclusion into the spectral library.

Analysis of Standards



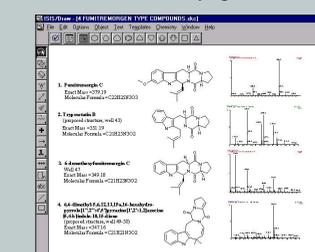
The ability to quickly eliminate known compounds from further study depends in part on the availability of standards derived from the producing organism. The HPLC analysis of 5 bioactive compounds from genus *Aspergillus* is displayed in this MassLynx screen capture. Mass spectra are extracted from the peak of interest and inserted into searchable spectral libraries.

LC-MS Analysis of Bioactive Wells - OpenLynx Software



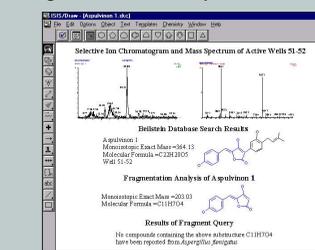
Fractions that exhibit a favorable bioactivity spectrum are assumed for purity and identity. The analysis of fraction 47 indicates the presence of 6-demethylfumitremorgin C, as indicated both by the presence of the mass ion (green shaded well) and the high value of the spectral match parameter. The OpenLynx software provides an intuitive interface for data viewing and is easily shared since all data is preprocessed and is stored within a single report file.

Active Constituents of *Aspergillus* Extract F953



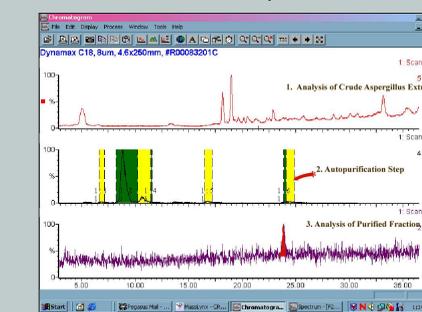
The mass spectra for compounds 1 and 3 were obtained from certified standards, while spectra 2 and 4 were derived from the preparative HPLC fractionation of the crude *Aspergillus* extract. The activity in well 43 is attributed to tyrostatin, based on similarities in structure, HPLC retention, and MS fragmentation to the standards. The mass spectrum of 4 (well 49-50) does not display the same complexity of fragmentation, possibly due to stabilization by the macrocyclic ring. In the absence of a standard, further structural studies are required to identify the active component.

Data Mining for Novel Active Compounds - ISIS/Draw Interface



The negative ion mass spectrum of the peak at 29.89 minutes (well 51-52) suggests that the compound is unrelated to the fumitremorgins, despite similar HPLC retention. A Beilstein database search for *Aspergillus*-produced compounds of molecular weight 364 AMU yields only 1 credible structure, aspulvinon 1. The blue highlighted region in 1 is imported into the ISIS/Base query engine for a substructure search. No compounds containing this fragment have been reported from the intensely studied *Aspergillus fumigatus*.

Mass-Directed Autopurification



Compounds can be isolated from a complex mixture during unattended operation. A minor component of the *Aspergillus* extract is selectively collected based on a positive MS peak at 348 AMU.

Conclusions

The ISIS/Base software enables easy and effective access to the NCI/SAIC antibiotic drug discovery data. Processing of chromatographic data by the OpenLynx software reveals known biological compounds in natural products extracts, thereby reducing the time in the following of false leads. The ready availability of a large structure searchable database enhances our ability to isolate novel compounds. With these software and database tools we have quickly found several known compounds and a possibly novel active component in an extract of *Aspergillus fumigatus*.

Acknowledgments
We would like to thank Dr. Ken Tucker, Ms. Patricia Slaty, and Ms. Cathy Hartland for the antimicrobial screening data. Thanks to Mr. Lucas Barr for creating HPLC fraction plates. Thanks also to Ms. Sally Graf for providing *Aspergillus* fungal growth and to Ms. Shyamala Nagarathan and Ms. Anne Kamata for providing the fungal extracts. Thanks to Mr. John Sun and Ms. Amy Burnham for computer and technical support. Finally, thanks to Dr. John Beutler for providing the Beilstein chemical structures.