Waters™

De novo Discovery of Natural Products Using Progenesis QI and Natural Product Atlas Library

Suraj Dhungana, David Heywood, Jeff Goshawk, Giorgis Isaac

Waters Corporation

For research use only. Not for use in diagnostic procedures.

This is an Application Brief and does not contain a detailed Experimental section.

Abstract

Discovery of novel compounds with biological and pharmacological activities is of great interest to the scientific and pharmaceutical communities. Over the years, researchers have looked at various sources of natural products (*i.e*. botanical, microbial, marine, etc.) to discover new therapeutics or compounds of health benefits. Identification of a novel compound or a novel activity of a compound demands comprehensive workflow. Optimized extraction and activity-based assays are required for the early part of the workflow, while complex separation, detection, and characterization tools are needed to complete the workflow. Liquid chromatography coupled with mass spectrometry (LC-MS) is a powerful approach that enables the separation and characterization of compounds present in natural product extracts. Extracts that show biological activity are often complex and demand an on-line chromatographic fractionation before tandem mass spectrometry analysis on a high-resolution Q-Tof instrument. Ion mobility separation of molecular ions in gas phase, following a chromatographic separation, adds further benefit to complex sample analysis by proving added separation and collisional cross section (CCS) value useful for compound identification. Manual interrogation of a complex natural product data set is not practical and requires an informatics solution along with a well-developed compound library for streamlined data processing and identification of compounds. The discovery workflow highlighted here discuss the use of Natural Product Atlas library (>25,000 compounds) in conjunction with Progenesis QI informatic solution for the discovery and automated identification of natural product compounds.

Benefits

- UPLC Q-Tof MS and ion mobility workflow integration with NP Atlas library for natural product research

- NP Atlas library within Progenesis QI for complete discovery solution

- Batch searches for automatic identification of compounds

- Custom database creation for reliable confident compound identification

- Perform cluster analysis, node analysis, or query GNPS database

Introduction

Advancing natural product research is key to finding novel compounds of therapeutic value and understanding the health benefits of natural products that are used as traditional medicines. Characterizing the molecular species of interest within a natural product is an involved process because of the complexity of samples. Fractionation of a

given extract is a standard first step used to reduce the analytical complexity. Despite multiple fractionations, a natural product extract fraction can contain hundreds of molecular species that require advanced analytical tools for separation and detection, as well as streamlined workflows and libraries for efficient characterization.

Liquid chromatography coupled with high resolution mass spectrometry (LC-MS) workflow for natural product discovery (Figure 1) provides the analytical separation and high-resolution tandem mass spectrometry data needed for capturing the chemical and structural complexity. With this approach complex natural product extracts are separated into their components and accurate mass information on precursor ions and fragment ions are captured for structural elucidation. Use of an ion mobility enabled instrument will further separate the molecular ions in the drift time dimension, orthogonal to chromatographic separation, and provide collisional cross section (CCS) information that can be used in the identification of compounds. Once the data is captured it is imported into Progenesis QI software for analysis and appropriate library matching for known compound identification or database searching and network/cluster analysis for narrowing the candidates for *de novo* discovery.
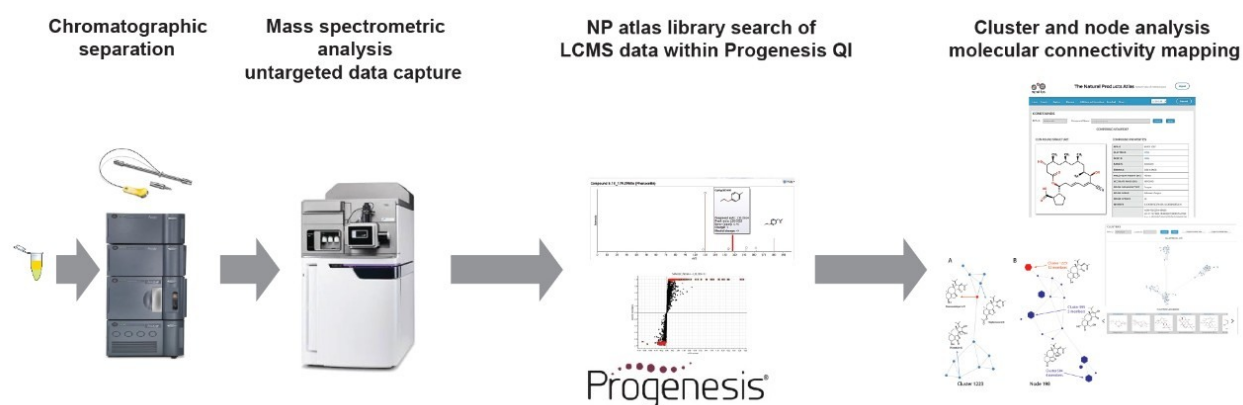

Figure 1. De novo discovery workflow for natural product research with Natural Product Library within Progenesis QI informatics solution.

Accessible and LC-MS data compatible libraries and databases are critical to timely and confident completion of natural product discovery projects. The Natural Products Atlas (npatlas.org <http://www.npatlas.org/>), created by a consortium of data curators from around the world and maintained by researchers in professor Roger Linington's Research Group at Simon Fraser University, is a great resource for the natural product research community and provides tremendous knowledgebase for advancing natural product research.[1] The Natural Products Atlas (NP Atlas) is designed to cover all microbially-derived natural products published in the peer-reviewed primary scientific literature and in its current version as of June 2021 contains >29,000 compounds (~11,000 bacterial compounds and 18,000 compounds of fungal origin) and the database contains taxonomical information to inquiry the distribution of compounds originating from different species. The NP Atlas also has search, explore, and

discover features to help researchers further their knowledge. Compounds can be searched based on their structure, compound name, molecular weight, chemical formula, InChiKey, SMILES, etc. or by using a structure drawing tool. The explore feature on NP Atlas allows clusters and node analysis on compounds with structural similarities, while the discover feature allows interrogation from an alternate point of view such as authors, journals, etc.

Despite being a tremendous repository and tool for natural product discovery, the open access NP Atlas library lacks the ability to be directly interrogated using the LC-MS data that is generated from a natural product extract. In partnership with NP Atlas, Waters has generated desktop or local version of NP Atlas libraries that are compatible with Progenesis QI and UNIFI software. Both the Progenesis QI and UNIFI version of the NP Atlas libraries can be downloaded at marketplace.waters.com <http://marketplace.waters.com> . This development gives much needed flexibility to advance natural product research by automating the search and compound identification process against the NP Atlas. In addition to NP Atlas, there are other natural product libraries freely available for download at the Waters marketplace and use with Progenesis QI and UNIFI.

Results and Discussion

During *de novo* discovery, one of the key steps along the workflow is the identification of compounds. Compound identification is a bottleneck in natural products workflows and a comprehensive database is essential for the efficient identification of compounds. The Progenesis QI version of NP Atlas is specifically developed for this workflow to automate the identification process. Setting the Progenesis QI version of the NP Atlas as the compound database and specifying the search parameters for precursor mass tolerance and theoretical fragmentation tolerance can be achieved in a single step (Figure 2). The search feature within Progenesis also calculates theoretical isotopic distribution based on the chemical formula and takes that into consideration during the compound identification and scoring. An example showing the identification of Erythromycin G using accurate mass, theoretical fragments, and isotope similarity score is shown in Figure 3A. Upon identification the experimental data for Erythromycin G is exported to generate in-house custom library, which now contains the experimental MS and MS/MS spectra, LC retention time (RT), and CCS information generated during ion mobility enabled HRMS experiment (Figure 3B).
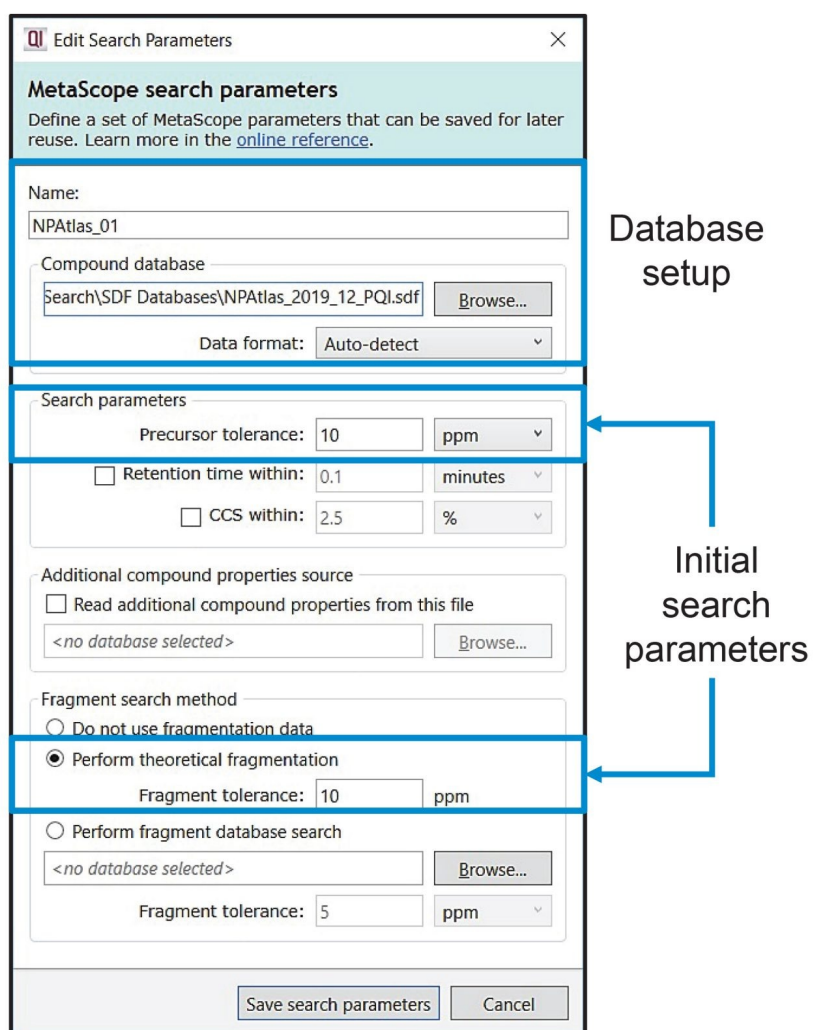
Figure 2. Simple way to add NP Atlas as a database within Progenesis QI and setting up initial search parameters.
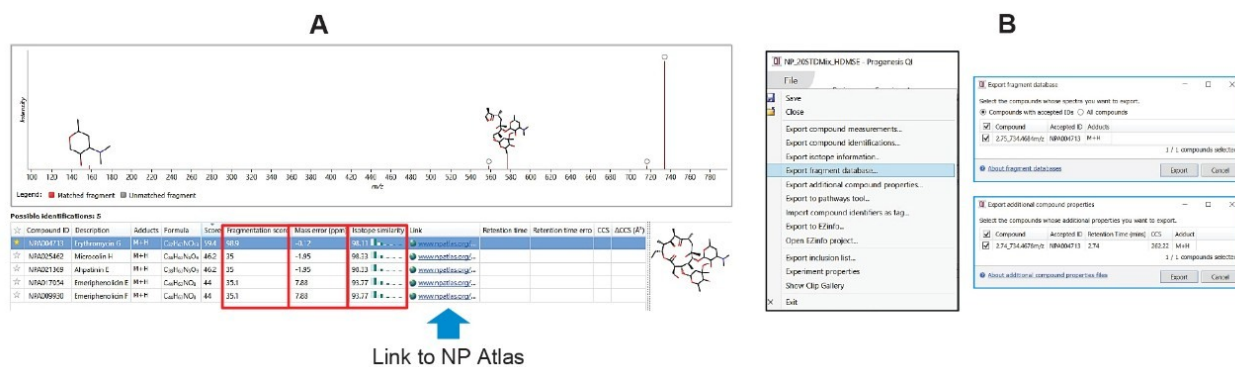
Figure 3. A) Unknown compound identification using NP Atlas as a database within Progenesis QI and B) Exporting experimentally derived MS/MS spectra, RT, and CCS parameter for in-house custom library creation.

Once a custom database is created, the next set of discovery samples can be searched against the custom database for confident compound identification. A custom database uses 5 parameters to identify a compound in an ion mobility enabled workflow (accurate mass of the precursor ion, isotope similarity score, RT, accurate mass fragment ion information, and CCS). If the custom library is generated in the absence of ion mobility, the identification will be based on 4 parameters. Addition of RT and CCS (in an ion mobility experiment) provide orthogonal information to increase the confidence of identification. Figure 4 illustrates an example of a search against a custom created database. Once again, we are looking at the identification of Erythromycin G, but in this example the search is performed against the custom database. In this search, the library match includes the accurate mass of the precursor ion, isotope similarity score, RT, accurate mass fragment ion information, and CCS value. Once the identification is confirmed, hyperlink within Progenesis can be followed to visit the compound specific page on the NP Atlas website to further interrogate the compound by utilizing all the features that are available within NP Atlas for compound exploration and discovery. NP Atlas provides additional external links, such as GNPS and MIBiG. Figure 5 shows an example linking NP Atlas with GNPS for further interrogation of compound within GNPS for molecular network analysis (Figure 5).
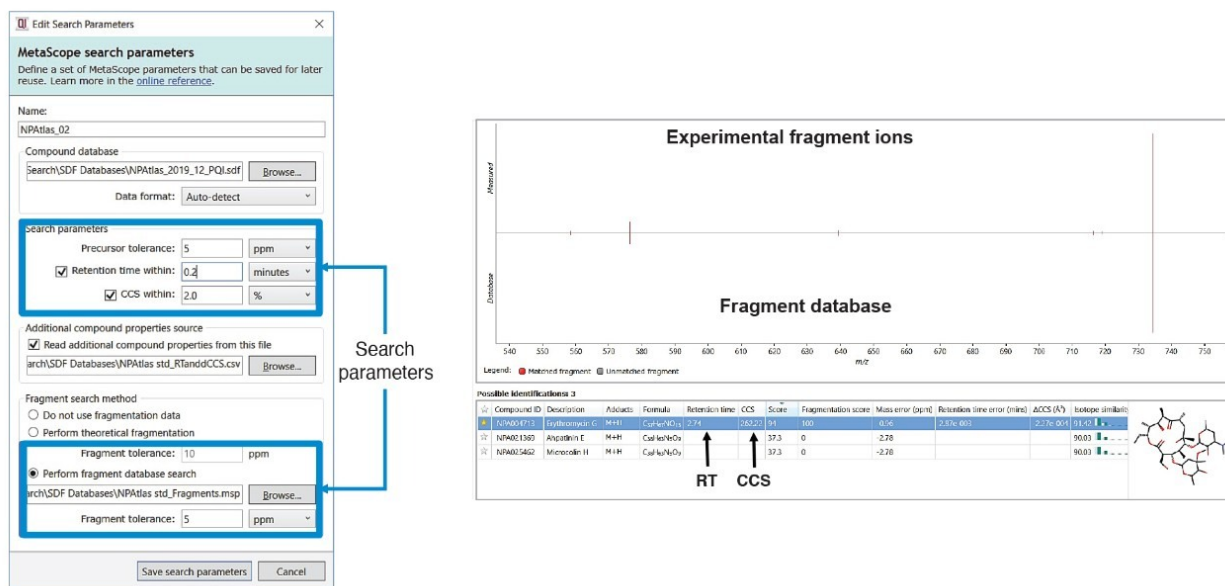
Figure 4. Setting up search parameters and performing search against a custom created library on an ion mobility enabled instrument.

Figure 5. NP Atlas provides direct access to GNPS for molecular network analysis.

To maximize the utility of NP Atlas libraries the recommended configuration of instrumentation involves the use of ACQUITY Premier System for the separation of complex natural product extracts with an ion mobility enabled Q-Tof SYNAPT XS High Resolution Mass Spectrometer for sample analysis. Ion mobility enabled data independent acquisition mode (HDMS[E]) is recommended to capture the sample complexity and both precursor exact mass and fragment ion information from a single injection. Following an untargeted data capture for a natural product discovery, data is processed using Progenesis QI workflow with NP Atlas library for database search.

An alternate instrument configuration to achieve similar results involves the use of Xevo G2-XS Q-Tof HRMS System. This platform does not support ion mobility separation so the ability to separate co-eluting compounds using

gas phase ion mobility separation will not be available. However, following a data independent acquisition MS^{E} mode both precursor exact mass and fragment ion information can be captured from a single injection. Without the CCS values available through an ion mobility experiment, the compound identification on a Xevo G2-XS platform will be based on 4 total parameters instead of 5, as discussed earlier.

Conclusion

Natural Product Atlas is outstanding resource for the natural product research community. To further the benefit of NP Atlas a desktop version of the NP Atlas library is developed. The primary benefit of the desktop version is the compatibility with the LC-MS data and ability to directly search LC-MS data against NP Atlas for compound identification, which is currently not available on the OpenAccess version. This compatibility further enables batch searches and automatic identification of compounds based on precursor exact mass, theoretical fragments and theoretical isotopic distribution, retention time and ion mobility derived CCS value. The desktop version of the NP Atlas library is freely available and can be download by visiting marketplace.waters.com < http://marketplace.waters.com> .

References

1. Santen, Jeffrey A. van, Grégoire Jacob, Amrit Leen Singh, Victor Aniebok, Marcy J. Balunas, Derek Bunsko, Fausto Carnevale Neto, *et al.* "The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery." ACS Central Science 5, no. 11 (November 27, 2019): 1824–33. https://doi.org/10.1021/acscentsci.9b00806 <https://doi.org/10.1021/acscentsci.9b00806> .

Acknowledgements

Featured Products

ACQUITY UPLC I-Class PLUS System <https://www.waters.com/134613317>

SYNAPT XS High Resolution Mass Spectrometer <https://www.waters.com/waters/nav.htm?cid=135020928>

Xevo G2-XS QTof Quadrupole Time-of-Flight Mass Spectrometry <https://www.waters.com/134798222>

MassLynx MS Software <https://www.waters.com/513662>

Progenesis QI <https://www.waters.com/134790652>

720007297, June 2021