

Introduction

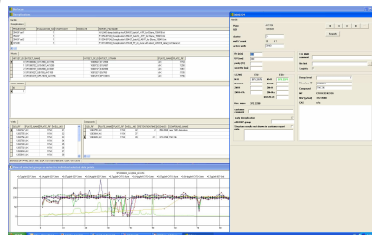
Natural products (NPs) once played a major role in drug discovery. As either the original isolated NP or as a subsequent derivatised NP, NPs accounted for more than 50% of the drugs introduced to the market. Nevertheless, the motivation to find new NPs from hit discovery campaigns dramatically decreased after 1990 due to the perceived speed of the new paradigm of high-throughput screening using enlarged small-molecule compound decks (enlarged by the new combinatorial chemistry synthesis technologies).

Today, there is a distinctive need for compounds which can affect the functions of some proteins that synthetic small molecules seem incapable of effecting (breaking protein-protein interactions, for example) and also for compounds with more novelty and/or complexity. Simultaneously, technological advances in the isolation and identification of natural products have nourished a recent renaissance of NPs in drug discovery.

BioFocus has implemented the high-throughput paradigm into natural product research by generating a library of pre-purified natural products derived from fermentation broths. Fractions are screened and active wells analyzed by LCMS in high-throughput mode. Utilizing HRESIMS data in combination with data from ACD/Labs databases and our customized LIMS, it is now possible to rapidly identify known secondary metabolites in pre-purified fractions and focus the effort on potentially new compounds for scale-up work.

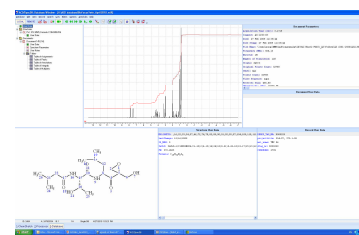
The strategy and the tools used at BioFocus for speeding up the data processing is described herein and includes the development of an ACD/Labs spectral database as well as a customized LIMS.

Customized LIMS



BioFocus has developed a customized LIMS for tracking all samples from strain information, crude extract, library fractions as well as all analytical information on each structure responsible for activity. This data storage system can help speeding up processes of dereplication in identifying active compounds that has been recorded in-house.

ACD/Labs



BioFocus is developing an ACD/Labs database with spectral information on every compounds identified during our process. In ACD/Labs the ¹H NMR database can be used to help speeding up the dereplication and structure elucidation process.

BioFocus pure natural products library: Analysis of diversity and physico-chemical properties

645 natural compounds have been identified during the follow up of subfraction screening programs. Their chemical properties have been analyzed as shown below.

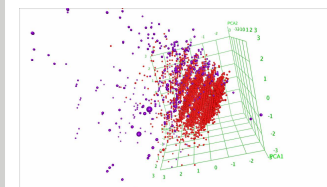


Figure 5: Diversity comparison of the collection of 645 natural compounds (purple) with BioFocus' collection of synthetic compounds (red). A principal component analysis was performed using the number of H-bond donors, H-bond acceptors, the number of rotatable bonds, and the atom-count. The coordinate axis show the first three principal components.

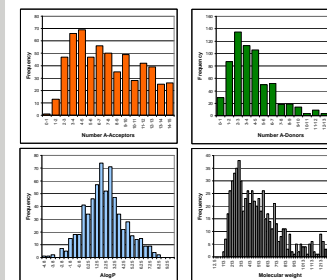


Figure 6: Distribution plots of the 4 "Lipinski properties" determined for the 645 identified natural compounds.

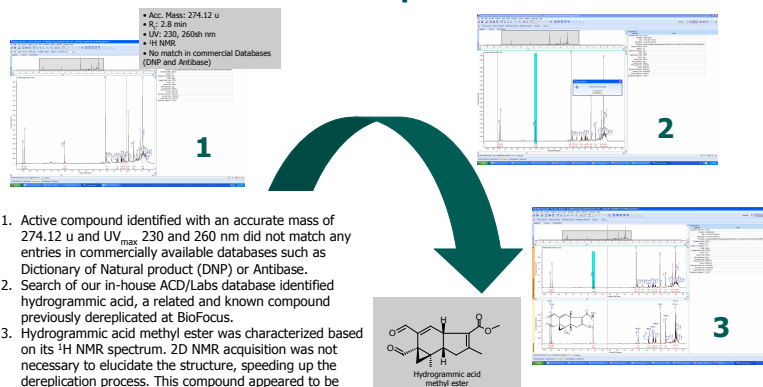
Novelty

Searches in natural product databases revealed that 135 of the 645 identified natural compounds have not been described in literature before, i.e. are novel compounds (21%).

Discussion

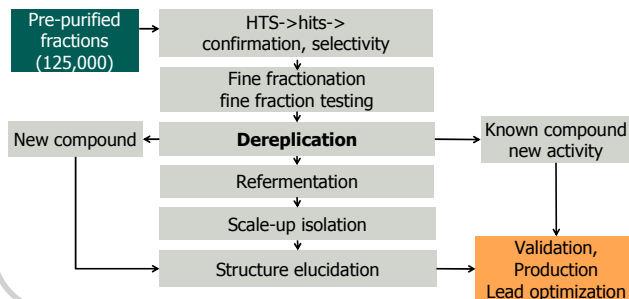
- The diversity analysis showed that BioFocus' natural compound collections complement the BioFocus diverse library of synthetic compounds.
- The natural compounds cover a broad range of physicochemical properties, even beyond the "Lipinski space". Lipinski's rule of five is not directly applicable to natural compounds and many natural compounds, that do not comply with Lipinski's rule of five, are orally bioavailable (Curr Opin Chem Biol 2008, 12, 306).
- More than 20% of the natural compounds discovered in the screens are novel, i.e. deliver valuable IP.

Example

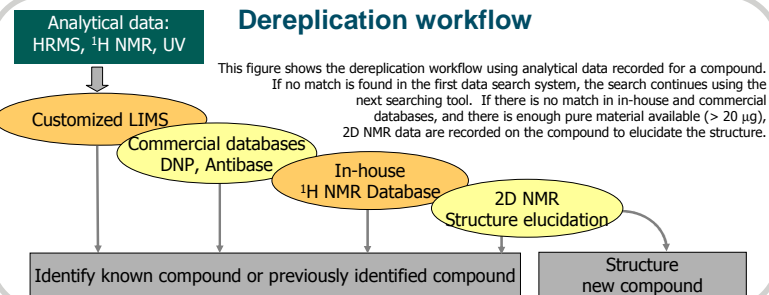


- Active compound identified with an accurate mass of 274.12 u and UV_{max} 230 and 260 nm did not match any entries in commercially available databases such as Dictionary of Natural product (DNP) or Antibase.
- Search of our in-house ACD/Labs database identified hydrogrammic acid, a related and known compound previously dereplicated at BioFocus.
- Hydrogrammic acid methyl ester was characterized based on its ¹H NMR spectrum. 2D NMR acquisition was not necessary to elucidate the structure, speeding up the dereplication process. This compound appeared to be also known (in Chemical Abstract).

From hit to lead: natural product follow-up strategy at BioFocus



Dereplication workflow



Conclusion

The property distribution of our pure natural products isolated at BioFocus (purple) compared with BioFocus' collection of synthetic compound (red) shows that natural products cover a broader space, highlighting the importance of natural product chemical diversity needed for drug discovery. In order to continue feeding drug discovery with chemical diverse compounds, natural product research is adapting to the high-throughput paradigm, and libraries of pre-purified natural compounds suitable for HTS have been developed in industry. Screening of such libraries generate a massive number of data that need to be processed. In order to optimize the resources allocated and to follow a successful strategy, it is important to identify known active compounds as early as possible in the natural product follow-up. In this regard, BioFocus has developed a spectral database and customized LIMS to help identify structures at an early stage of the processing. BioFocus is currently expanding this database to include HRMS spectra for all compounds.