Laboratorio di Preparazioni Estrattive

Estrattiva
Fasi Preliminari alla Estrazione da Piante

SAPIENZA
UNIVERSITÀ DI ROMA

Dereplication

Since the "Golden Age of Antibiotics" in the 1950s, natural products chemists have faced the steadily increasing problem of how to maximize the discovery of new compounds and minimize the re-evaluation of natural products already described in the literature. If a compound has been isolated, identified, and reported, it should be possible to use the published information to identify the compound when it appears again, without having to repeat the entire isolation and structure-determination process. In addition, in many instances, the questions being asked in a study can be answered simply by partial identification of the unknown structure. These complementary processes of rapid identification of known compounds from a partially purified mixture and identification of enough of an unknown structure to prioritize or conclude an isolation, have come to be termed "dereplication" by the natural products community.

Dereplication

The origin of the term dereplication as currently used in natural products chemistry is obscure, but it seems to have predated 1978, since that year's edition of the CRC Handbook of Antibiotic Compounds *(1)* contains the following historical analysis:

> . . .early recognition of duplication in a new active agent was essential. . . In the earliest procedures extensive use of biological activity and resistance patterns served to detect similarities. . . Now however, the great convenience of chemical and physical instrumental methods lays the groundwork for more specific identification and *dereplication*.

The aims of dereplication and partial identification procedures are to identify a chemical entity to a level that will successfully conclude the extraction as efficiently as possible. This can mean either a full identification of a natural product after only partial purification, or a partial identification to the level of a family of known compounds. Full identification in such cases normally relies on comparison of the compound with a characterized standard.

Dereplication

Reasons for carrying out a partial identification include:

1. To identify the compound as a member of a set that is not "desired." These classes might include common "interfering" groups of compounds such as tannins and polyphenols, fatty acids, or proteins, that may show nonspecific biological activity. The natural product might be one or a mixture of such compounds and may even include novel compounds.
2. To group a number of samples into related chemical classes and therefore prioritize samples for the extraction procedure. Often when natural extracts are assayed in high-throughput screens against biological targets, a relatively large primary "hit" rate is observed, which tends to suggest that one or more commonly recurring compounds are active in that assay. This filtering mechanism should identify rapidly the most diverse extracts on which to concentrate further efforts.
3. To obtain information that will progress the extraction; e.g., determination of the compound as an alkaloid means that it is basic and might be isolated by ion-exchange techniques.

Dereplication

This process of dereplication is becoming increasingly important and occupies a significant proportion of the time of many natural products chemists. Certain groups of organisms, such as marine invertebrates, were once very productive sources of novel compounds, but as they have been progressively investigated, the isolation of structurally novel components is becoming less routine, and more time is spent on dereplication. Faulkner's observations about marine natural products can be extended to many areas of natural products discovery:

> The good old days of 'grind and find'—when nearly every organism contained new and interesting molecules—are fast disappearing. Most bioactive extracts now contain only known compounds and developing rapid methods to avoid duplication of earlier research has become a priority *(2)*.

This is not to say that the large proportion of natural products are not waiting to be discovered, but that we need methods for bypassing those already discovered that effectively stand in the way.

Dereplication

Primary screen hit rates in natural products screening programs typically range from 0.1% to 4%. Thus, from a screen of 100,000-250,000 samples, the number of hits identified may be quite large (several hundred to several thousand). Most of these hits will ultimately be of no interest for a variety of reasons, eg. commonly-occurring classes of natural products may be responsible for the activity in a number of different hits, many hits may be due to compounds of low intrinsic potency but high abundance, others may be due to compounds of higher potency but with undesirable physicochemical and/or biological properties, etc.

The process of dereplication seeks to select a small sub-population of hits identified in a primary screen that are most likely to contain active compounds with the desired characteristics. With ultimate success being governed by the quality of those selections which, in turn, is dependent upon the accumulation of sound experimental data on all screen hits, dereplication has been a critical success-determining and rate-limiting step in natural products drug discovery.

Dereplication

Cerylid has developed an effective and efficient proprietary dereplication procedure that yields high quality data on highly-resolved components from all screen hits at high capacity — these being the necessary requirements for timely and effective prioritization of hits. All hits from screens are first subjected to a high-capacity fractionation procedure designed to generate information about the relative polarity of all active compounds present. Based on this information, all extracts displaying bioactivity in one or more of these initial fractions are progressed for HPLC separations using short gradients tailored to provide high resolution over the appropriate polarity ranges. With coupled UV/visible detection of eluates, testing of fractions for bioactivity both in the primary screen and relevant secondary assays, and analysis of active fractions by LC-MS, a package of physicochemical and bioactivity data on pure or nearly pure active HPLC fractions from all screen hits is generated.

Analysis of data arising from the dereplication process drives informed selection of those few extracts that are most likely to yield useful lead compounds.

Dereplication

Tecniche di separazione
- Ripartizione tra solventi
- TLC/HPLC (bioautografia)
- SPE
- Cromatografia controcorrente
- Elettrocromatografia capillare

Identificazione chimica

Tecniche spettroscopiche
- U.V.
- M.S.
- N.M.R.
- I.R.
- Simulazioni computazionali

Profilo biologico
- attivita' antibiotica
- Attivita' antitumorale
- Affinity fingerprints
- Tecniche immunochimiche

Database di letteratura

Dereplication

Databases Containing Natural Product Information

Database	Number of compounds	Number of natural products	Structure search	Mol wt	UV	NMR, ^{1}H, ^{13}C	Mass spec.	Bio-activity	Source organism	Access	Contents
CAS *(65)*	>12,000,000		Yes	Yes				Yes	Yes	STN	All sources
Beilstein *(66)*	>5,000,000	91,000	Yes	Yes	Yes	Ref.	Ref.	Yes	Yes	STN	All sources
NAPRALERT *(67)*	100,000	100,000	Yes (search by registry file)	Yes (search by registry file)				Yes	Yes	STN	Mainly plants
Dictionary Natural Products *(68)*	115,000	70,000	Yes	Yes					Yes	CD-ROM	All sources
Berdy *(69)*	26,000	26,000		Yes	Yes			Yes	Yes	MS-DOS	All sources
KMC-plus *(70)*	16,000	16,000		Yes	Yes			Yes	Yes	MS-DOS	All sources
Antibase *(71)*	14,000	14,000	Yes	Yes	Yes	Yes	Yes	Yes	Yes	ISIS/Base	Microorganisms
DEREP *(72)*	7000	7000		Yes	Yes			Yes	Yes	MS-DOS	All sources
Marinlit *(73)*	6000	6000	Yes	Yes	Yes			Yes	Yes	Mac	Marine organisms
MNP *(74)*	4000	4000	Yes	Yes	Yes			Yes	Yes	Mac	Marine organisms

Dereplication

7. Scope for Natural Products Dereplication

7.1. History of Natural Product Identification

The pace of natural product discovery has been increasing throughout this century. The number of structurally assigned natural products that have been reported in the literature now stands at >100,000 ***(66)***, of which >26,000 ***(69)*** have reported biological activity (*see* **Fig. 13**).

The majority of the reported natural products have been isolated from plants, fungi, or bacteria. Since a large percentage of natural products have shown some sort of biological activity, it can be assumed that much natural product isolation has been bioassay-directed. Except in a few cases, little effort has been expended in determining the complete range of secondary metabolites that are biosynthesized by any one organism.