Basics and Principles for Building Natural Product–based Libraries for HTS

Ronald J. Quinn

Natural product libraries require collection of samples of biota and extraction of compounds for screening against a biological target. Although this is a simple task, the establishment of a quality natural product–based library requires a good understanding of the modern drug discovery paradigm. A quality natural product library will allow both the discovery of natural products that may be able to be developed into therapeutic agents as they occur in nature and the identification of starting points for medicinal chemistry optimization. The process of establishing a natural product screening library for use in high-throughput screening (HTS) is illustrated in Figure 7.1. The collected sample may be plant material, marine organisms, or terrestrial or marine microorganisms. Once collected, the samples need to be processed to provide the compounds contained in the samples in a soluble form to allow HTS. Extraction with organic and aqueous solvents provides the soluble constituents that may be processed further into fractions or pure compounds.

1. REQUIREMENTS PRIOR TO COLLECTING NATURAL PRODUCTS

In recent years, it has been recognized that genetic material is a sovereign right of the country of origin. Collection of biota samples must be undertaken according to the United Nations Convention on Biological Diversity (CBD). The CBD opened for signature on June 5, 1992, at the United Nations Conference on Environment and Development (UNCED), also known as the Earth Summit, in Rio de Janeiro, Brazil. The text of the CBD is available on the Web site of the Convention of Biological Diversity at www.cbd.int/convention/convention.shtml.

The objectives (Figure 7.2) of the Convention are

1) the conservation of biological diversity;
2) the sustainable use of its components; and
3) the fair and equitable sharing of the benefits arising out of the utilization of genetic resources.

Figure 7.3 highlights important aspects of the CBD for those interested in biodiscovery, the use of biological diversity as a discovery resource for new potential therapeutic agents. Article 15 of the CBD recognizes the sovereign rights of States over their natural resources and that the authority to determine access to genetic resources rests with national governments and is subject to national legislation. Article 15 also specifies that access shall be on mutually agreed terms (Article 15.4), subject to the prior informed consent of the Contracting Party (Article 15.5) and the benefits arising from the commercial and other utilization of genetic resources shared on mutually agreed terms (Article 15.7).

All collections should adopt “best practices” in access and benefit sharing. It is necessary to have formal contracts between the parties and approval of the government of the country that has the sovereign rights to the genetic resources. Negotiations can be time consuming and require a good knowledge of the commercialization pathway, risks, and rewards from all sides involved. It may realistically take several years to negotiate appropriate benefit sharing and access agreements. The chance of finding a new therapeutic agent are low; there is significant attrition in the pathway to market; and large numbers of samples will need to be investigated in HTS campaigns. Although rewards for a successful drug may be large, not all drugs introduced to the market repay the investment in their discovery. The holder of genetic resources needs to have an excellent understanding of the commercialization pathway in order to understand the value of the resource and enact the mutually agreed terms of the CBD. Investors need a guarantee that they can acquire intellectual property rights to specific compounds, and the CBD provides the mechanism to ensure certainty that commercialization can take place.

2. COLLECTION

The next step in the discovery process is the collection of samples. A biota library may consist of microbial cultures or macro biota such as plants and marine and terrestrial invertebrates. There are a number of approaches taken in collecting samples. It is possible to rely on prior knowledge, such as that generated from traditional use of the biota, or to undertake random collections. Irrespective of the
CHEMICAL GENOMICS

High Throughput Screening

Nature
Plants, marine organisms, microorganisms

Biota Library
Store of samples

Extraction / Pre-fractionation / Pure Compounds

Natural Product Library

High Throughput Screening

FIGURE 7.1: Building a natural product screening library.

philosophy underpinning the strategy, it is essential to document and curate the collection. Getting this aspect right will be important for downstream activities (see section 3 on recollection). Two approaches are generally adopted with respect to collection. Collection may be undertaken by the biodiscovery team that brings the screening and isolation expertise or by botanists and biologists who bring expertise in the taxonomy and environmental areas. Regardless of the approach, there is no doubt that biota samples collected for biodiscovery purposes can play a vital role in conservation knowledge and biodiversity research, and placing samples into national collections will facilitate research into conservation and understanding of genetic resources. Taxonomic identification of the species is vital for both increasing the chances of finding novel species containing novel constituents and avoiding known compounds (see section 5 on dereplication). Current strategies for collection of biota samples require a great deal less material than was previously required by screening programs. [1] Advances in screening technology, in particular 384-well and 1536-well assays, mean that a sample of 200 mg can provide sufficient extract for screening in multiple HTS campaigns. Advances in spectroscopy for structure elucidation mean that 1 mg of compound can, in most cases, provide enough material for both structure elucidation and biological profiling in the primary HTS assay as well as several selectivity assays. Initial collection needs to be undertaken bearing in mind conservation of the species and the habitat from which it was collected. Development of a compound of interest will require progressively larger amounts of material. Collection of smaller quantities at the beginning followed by recollection of larger quantities only when results indicate the presence of a compound that is of continuing interest has the least environmental impact. In undertaking biodiscovery, it needs to be recognized that recollection of larger quantity may not be possible, and supply of compound by synthesis may be required.

The eventual objective of the natural product library is to obtain a source of diverse compounds for HTS evaluation. Compound diversity correlates with biota diversity. Therefore, it is worth bearing in mind that the seventeen megadiverse countries of the world (Australia, Brazil, China, Colombia, Democratic Republic of Congo, Ecuador, India, Indonesia, Madagascar, Malaysia, Mexico, Papua New Guinea, Peru, Philippines, South Africa, the United States, and Venezuela) account for more than 70% of all biodiversity [2]. Figure 7.4 shows the seventeen megadiverse countries of the world. Figure 7.5 shows the number of endemic plant species for the seventeen megadiverse countries as one indicator of biodiversity.

The establishment of a database, at the stage of collection, is vital to track the obligations under the CBD and also to track samples through the HTS campaigns. Data such as taxonomy, collection date and location, collecting institution and individual collector, and species abundance should be captured in the database. This assists with both tracking and monitoring samples throughout the research process for access and benefit-sharing purposes, recollection (including any concerns associated with sustainability), and the identification of factors that contribute to bioactivity, such as season, location, and stage in reproductive cycle.

A choice can be made to extract the whole biota sample, to extract a subsample of the biota sample, or to extract a fresh sample prior to each screening campaign. Extraction of the whole biota sample ensures that one extract is available for the screening, isolation, and structure elucidation stage. Sequoia, for example, extracted 150 g of dried plant

Objectives
Conservation of biological diversity
Sustainable use of biological diversity
Fair and equitable sharing of the benefits arising out of the utilization of genetic resources

FIGURE 7.2: Objectives of the Convention on Biological Diversity.

Article 15
Affirms the sovereign right of countries to control the use of their own resources

15.4 Access shall be on mutually agreed terms
15.5 Access shall be subject to the prior informed consent of the Contracting Party
15.7 Benefits arising from commercial and other utilization of genetic resources shared on mutually agreed terms

FIGURE 7.3: Article 15 of the CBD recognizes the sovereign rights of states over their natural resources and that the authority to determine access to genetic resources rests with national governments and is subject to national legislation.
material in generating its fractionated library [3]. Compound degradation may occur in extracts that are held for a long time. The cost in time and solvent will be greater than those for a procedure that extracts a subsample of the biota sample. In our experience, plant samples that are air-dried and ground to a powder and marine samples that are freeze-dried and ground retain compound integrity. Extraction prior to each assay is not efficient. Figure 7.6 illustrates aspects of a biota store. The humidity needs to be controlled to prevent moisture from entering the samples; high humidity may result in fungal or other microorganism spoilage. Samples have been stored in individual containers that are bar coded. Individual samples are stored in larger containers that are bar coded. The database system tracks the location of individual containers in each of the larger containers.
3. RECOLLECTION

Larger quantities of compound will be required to progress any interesting compounds. The probability of recollection will be enhanced if the initial collections are well documented. The ability to have the original collector return to the same site armed with the GPS location, photographs (underwater and above water for marine samples), and taxonomic knowledge cannot be underestimated. Again, conservation requirements must be observed when larger quantities are sought.

4. EXTRACTION

The methodology for natural product screening should conform to well-established screening practice. HTS campaigns are usually conducted on compounds dissolved in dimethyl sulfoxide (DMSO). Compound libraries can be screened for many years, and preparation of a natural product screening library should ensure that it can be used for the same time span. To avoid continual extraction of the entire biota library for every HTS campaign, it is much more efficient to store sufficient amount of the biota extract to last five years (>100 screens). Figure 7.7 shows a multiparallel extraction apparatus designed to extract ninety-six solid biota samples simultaneously. An extraction cycle comprising a hexane wash and two extraction solvents with overnight drying using nitrogen and vacuum dryers allowed a twenty-four-hour cycle.

The two most popular ways of achieving long-term storage that maintains the integrity of the samples in DMSO are microtubes and minitubes held anywhere between −20°C and 20°C under an atmosphere of nitrogen or low relative humidity. Figure 7.8 shows typical microtubes. Figure 7.9 shows a store that has a capacity of 300,000 microtubes and 1,500,000 wells in 384-well microtiter plate format. Figure 7.10 shows a formatting robot for conversion from tubes to microtiter plates.

There are a number of ways in which biota samples can be processed into a form suitable for screening:

- Crude extracts – an extract using organic or organic/aqueous solvent mixtures;
- Prefractionated libraries of crude extracts – crude extracts that are fractionated using solid-phase extraction (SPE), conventional liquid chromatography techniques, or a combination of both; and
- Pure natural products.
5. HANDLING DEREPLICATION

Dereplication, the use of biological profiling and chromatographic/spectroscopic analysis to recognize known substances present in an extract, has been used widely [4–6]. Dereplication aims to identify known compounds prior to investment in isolation. Another important outcome of dereplication is the identification of multiple extracts or fractions that contain the same active component or biological profile. Dereplication has been particularly necessary for those groups working on microbial extracts [7]. Differential analysis of two-dimensional (2D) nuclear magnetic resonance (NMR) spectra has been used to identify new compounds from fungal extracts [8].

**FIGURE 7.7:** Parallel extraction – Six banks of sixteen syringes for solid phase extraction. Acknowledgement to Stephen Toms (Eskitis Institute) and J-Kem Scientific for design and construction of the 96-way parallel extraction unit. Photo: Stuart Newman, Griffith University.

**FIGURE 7.8:** Bar-coded microtubes that hold solubilized compound. Compounds are stored in individual tubes. Photo: Chris Stacey, Griffith University.
6. PRO AND CONS OF SCREENING CRUDE EXTRACTS/FRACTIONS/PURE COMPOUNDS

A quality natural product–based library requires that the compounds within the screening set be able to be developed into therapeutic agents. A good understanding of the modern drug discovery paradigm is therefore necessary. Analysis of the 126,000 unique entries in the Dictionary of Natural Products [9] against Lipinski’s “Rule of Five” indicated that 60% of natural products comply with Lipinski’s rules. [10] Only about 10% of NPs exhibited two or more violations, the same as for trade drugs [11].

Compound diversity is greatest in crude extracts, but many components are commonly occurring or redundant. Much of the chemical space in crude extracts consists of
compounds that are frequent hitters, that is, have promis-
cuous pharmacology, do not conform to the physicochem-
ical parameters required of drugs, or indeed may not be
druggable in the sense that they are not easily synthesized
and it is difficult to develop chemical libraries or structure-
activity relationships (SARs) around the compounds
[12].

In addition to the screening of crude extracts, two
approaches have been utilized. One has been to purify as
many components of an extract as possible and determine
the structures of the compounds [10, 13, 14]. This approach
has several obvious advantages: 1) The compounds isolated
can be treated the same as any other compound in a library;
and 2) immediate assessment of the compound’s potential
can be made during hit evaluation, thus eliminating the
time delay between hit extract identification and isolation
of the hit natural product. With a structure in hand, it is pos-
sible to exclude compounds from the screening set that do
not conform to the physicochemical parameters of choice
or have chemical alerts associated with the structures. Two
major disadvantages of this approach are 1) minor compo-
nents may be overlooked and thus never reach a screen-
ing set and 2) limited throughput that results in a limited
diversity.

A second approach has been to generate libraries
of semipurified fractions for screening. Prefractionation
involves the separation of crude extracts by SPE, column
chromatography, high-performance liquid chromatogra-
phy (HPLC), liquid-liquid partitioning, or some combi-
nation of the above to obtain fractions containing simpler
mixtures of compounds. Prefractionation can eliminate
highly hydrophilic and hydrophobic fractions, and com-
ounds found within each fraction can be tested at a higher
dose than would be possible in a crude extract. Even though
active fractions require purification to isolate the active
constituent, the use of MS and NMR data as well as bioassay
data can guide further purification.

Numerous prefractionation strategies have been re-
ported [3, 14–19]. Approaches are wide-ranging, from the
preparation of four to two hundred fractions per sample.
MerLion has taken the approach of using reverse-phase
C18 HPLC and collecting four fractions with collection
commencing after the solvent front peak (Figure 7.11) [15].
Sequioa has performed an organic extraction followed by silica
flash chromatography and an aqueous extract followed by
C18 flash chromatography. Subsequent HPLC collected
forty fractions per each of four flash fractions from the
organic extract. The aqueous extract was passed through
C18, polyamide, and a 3,000 molecular cutoff filter, and
the single aqueous flash fraction was subjected to HPLC,
again collecting forty fractions. The HPLC fractions that
contained quantifiable compounds, approximately 60% of
the total, constituted the library. This process led to a
library of 36,000 fractions containing one to five com-
ponds per well (Figure 7.12) [3]. Wyeth has used pre-
fractionation to generate ten fractions plus crude from the
combined extract from at least two different culture condi-
tions (Figure 7.13) [18]. The Ireland group has used SPE
followed by HPLC to generate eighty fractions per sample
(Figure 7.14) [19].

Prefractionation together with taxonomic identifica-
tion, GPS location, photographs, and voucher specimens

![FIGURE 7.11: Prefractionation reported by MerLion using
reverse-phase C18 HPLC and collection of four fractions per
extract [15].](image1)

![FIGURE 7.12: Prefractionation reported by Sequioa using flash
chromatography followed by HPLC. Forty fractions were col-
lected per each of five flash fractions from the biota sample. This
process led to a library of 36,000 fractions containing one to five
compounds per well (Figure 7.12) [3].](image2)

![FIGURE 7.13: Prefractionation reported by Wyeth generating
ten fractions plus crude from the combined extract form at least
two different culture conditions [18].](image3)
provide a systematic documentation of a biota collection (Figure 7.15). A number of questions can be answered with this data set. On the one hand, Figure 7.16 shows five chromatograms of the same genus that clearly demonstrated chemical diversity at the species level. On the other hand, Figures 7.17 and 7.18 shows a cluster of four and two biota, respectively, displaying similar chromatographic profiles. If biological screening results were to indicate activity in the same fraction, a useful dereplication would be achieved.

HTS data have been used to evaluate prefractionation strategies. Of the 1,882 active cultures following nine HTS campaigns on a microbial natural product library, 79.9% of the activity was found only in the fractions, 12.5% of the activity was detected only in the crude extract, and 7.6% was found in both fractions and crude extract (Figure 7.19) [18]. A very similar result was achieved using the simpler four fractions per extract approach. For 80% of the 1,700 active fractions from eleven HTS screens, the associated crude extract did not show activity (Figure 7.19) [15].

In considering the approach to developing the natural product library, the throughput of the HTS technology as well as the cost need to be taken into account. The Wyeth publication quotes assay costs at $0.25 to $0.50 per well and the screening of one million wells at between $250,000 and $500,000 per HTS campaign. If ten thousand biota samples were extracted to give ten thousand extracts, they would produce twenty-nine 384-well microtiter plates in which the plate has two columns empty for controls. If four fractions were produced per biota sample, the resultant forty thousand fractions would produce one hundred...
FIGURE 7.16: Chromatograms of the same genus that demonstrate chemical diversity at the species level.

and fourteen 384-well microtiter plates. Increasing the prefractionation to give twenty fractions per sample would require 569 microtiter plates to house the two hundred thousand fractions. Isolation of pure compounds from ten thousand biota samples, which may contain approximately two hundred thousand unique compounds, will result in a smaller number of compounds because of the time and effort involved in isolation and structure elucidation. A set of twenty thousand pure compounds would be accommodated in six microtiter plates. Figure 7.20 shows the cost for the resultant HTS campaigns at a cost of $0.10 per well. A cost-benefit analysis (Figure 7.21) indicates that prefractionation provides a significant benefit in screening relative to the investment in enrichment of the screening set. The cost associated with preparation of a pure natural product library is significantly greater than the effort to prefractionate. The arrow indicates a balance along this continuum where the cost-benefit ratio is optimal.

CONCLUSION

Natural products interact with a variety of proteins during their biosynthesis and can target specific therapeutic target proteins. [20] If modern drug discovery principles are applied in the preparation of the natural product screening library using the general approaches discussed, the resultant high-quality libraries will be more efficient in harnessing the natural product world in the modern drug discovery paradigm.
FIGURE 7.17: Prefractionation aids dereplication.

FIGURE 7.18: Prefractionation aids dereplication.

FIGURE 7.20: The greater the prefractionation, the greater the resultant cost of the HTS campaign.

FIGURE 7.21: Cost-benefit analysis of screening sets.
REFERENCES


