7

Compound Library Management

James A. Chan and Juan A. Hueso-Rodríguez

1. Introduction

The growing number of compounds synthesized and screened in the pharmaceutical industry due to the rapid development of high-throughput chemistry and screening technologies are a response to the need of more and better quality compounds in the industry pipelines. But these inmense collections (from several hundred thousands to several millions of compounds) pose a tremendous logistical problem to be overcome in order to harvest all their tremendous potential. Some reviews have appeared to deal with this topic ***(1,2)***, but an update seemed necessary due to the rapid evolution of this field. The scope of this chapter is the management of different types of compound collections from both physical and electronic points of view, including some aspects of natural product-extracts collections. This management is a very difficult and demanding process involving the use of sophisticated equipment and databases. The first thing to bear in mind when implementing this process is that a specialized and dedicated group should be created to be responsible for maintaining the collection and processing the orders or requests from the rest of the company. Failure to do so usually ends up with a chaotic situation where no samples can be retrieved in due time and proper format, with no control on the available amounts and locations of the samples. Proper management of the compound collection is the foundation of a quality screening organization. If this function does not operate properly, the most advanced assay technologies will fail to afford reproducible lead compounds.

From: *Methods in Molecular Biology, vol. 190: High Throughput Screening: Methods and Protocols*
Edited by: W. P. Janzen

2. Types of Collections and Stores

The sources of compounds used in a screening program are the company's historical collection, new chemical entities prepared for the current projects, prospective or unbiased chemical libraries prepared by combinatorial or parallel synthesis and, finally, external purchases.

The compounds in a collection can be stored in two different ways: neats (usually as solids) and solutions. The most common situation in any high-throughput screening (HTS) group is to both store solids and solutions if sufficient compound is available. The obvious reason is that solutions in plates are more suited for HTS even though neats are more stable under long-term storage conditions. Solids in pre-tared bar-coded vials are normally stored at room temperature, usually in automated stores. They should be easily retrievable for confirmation of activity, secondary testing, and structural verification. Compounds in solution are stored in plates or tubes, typically at two different concentrations. One is a high concentration solution (10–5 m*M*) in vials or plates at low temperature (4° to –20°C) for preparing lower concentration plates for HTS, long-range storage, and used as a repository and/or for dose-response experiments. The other one is a lower concentration solution in plates (1–0.1 m*M*) at temperatures ranging from ambient to –20°C used as a source for primary and secondary screening. In some cases, two sets of the lower concentration plates are prepared, one kept at low temperature for long-range storage, and a second one as a working set at ambient or slightly below temperatures. This second collection usually is date stamped and a shelf life is assigned after which the set is discarded and substituted for a new one.

When the collection is from natural-products extracts, all that was described previously can be applied, just substituting neat compounds by dried extracts and the compound solutions by extract solutions. In this case, absolute concentration cannot be determined since we do not know the components of the extract and their quantities present, but a relative concentration can be used just by taking original concentrations in the fermentation broth as the unit. In case of plants, marine organisms, or other materials, the initial concentration usually is in the range of 10–1 mg/mL of dried extract.

The stores used for all these types of collections can vary from shelves in a cold room to independent freezers or containers that are kept at the designed temperature. In some cases, humidity conditions are also controlled and even inert atmospheres are used. All these require manual input and output of the samples which is resource extensive and very prone to mistakes. To solve this problem, large automated stores have been devised and are now available in the market. These systems are described in the following chapter. Nevertheless, for small and medium size collections as those generated *de novo* for

starting an HTS program or the historic collections of small pharma companies, intermediate solutions can be applied that combine an automated storage system with the basic liquid-handling required for all these systems.

Liquid handling is a critical component of any storage system and should be capable of producing the samples in the formats required by the HTS systems, like plate density (96-, 384-, or 1536-well plates), controls and standards positions, amounts, concentrations, single- or multiple-well replicates and dose-response formats. This require to have systems capable of dispensing liquids from few nanoliters to several microliters with high accuracy, precision, and adequate throughput so they are able to provide the samples in a timely manner. The usual requirements are at least one system for dissolving dry compounds from vials to the appropriate concentration, another system for plate replication and reformatting, and a third system with the ability to cherry pick rows, individual wells or tubes to create adhoc plates for mixture deconvolution or dose-response analysis. The number of these systems can then be scaled-up or combined into one single platform commensurate with the collection size and the throughput required by the HTS process. Examples of the equipment available in the market for all the tasks described in this and remaining sections can be found in the Appendix at the end of the chapter.

2.1. Processing of Compounds

As an example of the processing of compounds for different types of orders, we present here the procedure followed at GlaxoSmithKline and graphically schematized in **Fig. 1**. Samples are processed for therapeutic teams and HTS based on whether compounds are prepared by traditional synthetic methods or parallel synthesis. Compounds prepared by traditional methods are registered, submitted in pre-tared barcoded vials, entered into inventory and 1–2 mg solid weighed out immediately and submitted to the therapeutic team for testing. Another sample enough to make 1–2 mL of 10 m*M* DMSO solution will also be weighed out into a vial, dissolved, and dispensed into a plate. Compounds prepared by parallel synthesis use an in-house database that track the design, synthesis, purification, quality control (QC), registration, and submission into inventory in a per library basis. The molecular weight and weight can then be downloaded as text file into a liquid handler with spanable tips, e.g., Tecan Genesis or CCS/Packard Multiprobe and appropriate amount of dimethyl sulfoxide (DMSO) added to the vials to make 10 m*M* solution. Aliquots can then be transferred into a stock plate for the preparation of 1 m*M* source plates. These source plates are then used to dispense compounds into different types of screening plates based on the different assay requirements. There are numerous liquid handlers in the market with 96 or 384 heads that can dispense

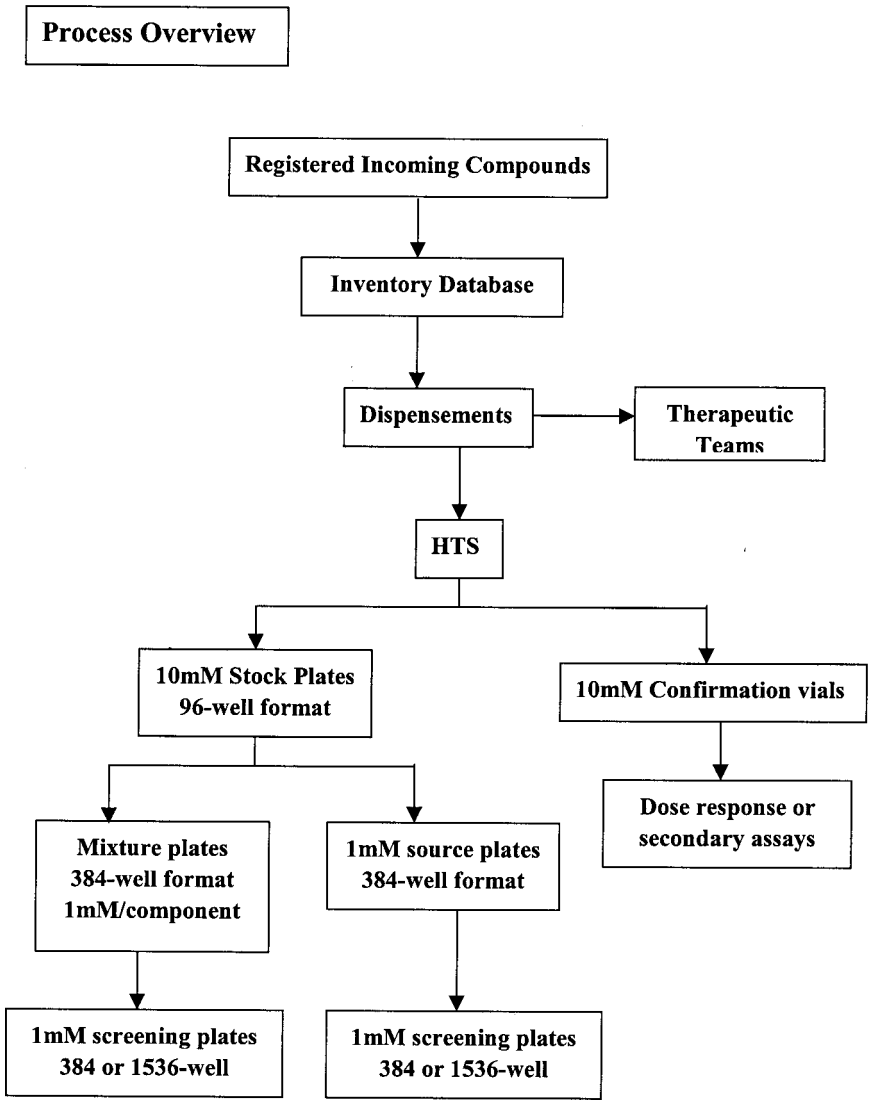

Fig. 1. A schematic representation of the Process on the Management of the Compound Collection is shown.

volumes as low as 50 nL into 384- or 1536-well plates for HTS, e.g., Robbins Hydra/Tango and CCS/Packard PlateTrack. The remaining volume in the vials is reserved for dose-response analysis and structural verification. The vials and plates can then be stored in stand-alone freezers and processed using stand-alone liquid handlers with stackers or stored in automatic refrigerated stores

like Tecan's Molbank that carries its own integrated liquid-handler dispensing unit. For very large collections there are now automated liquid stores in the market that store plates and vials integrated with all types of liquid handlers needed to process plates for HTS and cherry pick vials for dose-response analysis in very high throughputs.

3. Informatic Tools

As it can be clearly seen from the previous section, the management of such huge collections with a tremendous variety of demands can not be done without the help of an Inventory Database, not only to being able to process all the types of incoming orders and outgoing samples and plates but also to keep track of all the samples locations, current amounts, and dispensement history for solids or solutions. The Inventory Database should be linked to other informations of much interest such as lot identification, source, reference to the obtaintion procedure, molecular weight, molecular formula, structure, program, or project the compound was originally prepared for, submission date, inventory entry date and internal identification number and QC data. In the specific case of natural-products extracts other information like producing organism (taxonomy), part or tissue used for the extraction, method of extraction, geographical origin of the organism, method of cultivation, etc., should be added. It is important to note that the more information kept for any given sample, the better, but trying not to overload the database so its use and maintenance is kept at a reasonable level. Another important issue is the interface of this Database, whose use for data entry should be restricted to a limited number of people due to its complexity and the high value of the information stored. This is usually solved through an intermediate interface where scientists can enter their orders to be then processed by Compound Bank staff.

3.1. Requesting of Compounds

A Lotus Notes Database is used by the Therapeutic Teams to request compounds in solid or solution form for testing. This database is critical in tracking the status of the compound request especially for a large organization with different research sites in different countries. This database has automatic notification feature via e-mail for both the requester and recipient when the request is completed. The requester can check for availability of compounds in solution or solid form using another in-house database accessible through desktop PCs that also allows the scientist to query published biological and physical data. This request database also tracks the compounds from the time they leave the Compound Bank to the time they reach the recipient's laboratory. For a list of compounds, a dispensement barcode will be provided and the recipient can scan the barcode to generate a spreadsheet enumerating pertinent informa-

tion (e.g., compound number, molecular weight, amount dispensed, barcode, etc.) in order to prepare the testing solutions. This database has the following required fields for the requester:

Requested by: Name of the requestor.

Requested for: Name of the recipient if different to the requestor.

Department: Name of department.

Ship to site: Specific location where the samples should be shipped.

Program/Project name: Program name or code.

List of compounds: Internal numbers of the compounds requested (Chemical names are not admitted, since the Inventory Database only recognize the internal codes.)

Quantity needed: Amount required. (Normally 1–2 mg for solids or 20 μL of a 10 m*M* DMSO solution. Requester has to justify request when amount exceeds the norm.)

DMSO solution acceptable? Only for those cases where not solid is available.

Compounds checked against Inventory: Availabilities checked prior to sending the order (This process is highly recommended in order to avoid the iterative cycles of questioning if solutions and amounts are acceptable, delaying the processing of the order.)

Purpose: General and short description on what the samples are intended for (for tracking purposes only.)

Attach a file: For Excel or text attachments.

Comments: General comments on urgency or specific requirements. The Compound Bank personnel is required to complete the following fields:

Responsible person: Name of the person that processes the order.

Dispatch date: For requests outside the site. The requester and recipient are automatically notified via e-mail when a dispatch date is typed in.

Shipment number: Needed to track the shipment.

Ready for pickup: For requests within the same site. The requester and recipient are automatically notified via e-mail when the samples are ready for pickup.

The complexity of this Database varies directly with the complexity of the research organization due to the number and types of requests.

3.2. Tracking of Compounds in Plates and Vials

The Inventory Database is central to the purpose of tracking compounds on a vial/plate/well basis by assigning a barcode to each container, e.g., the stock plate is a container and has a barcode and every well/compound in this particular plate is also assigned a barcode. It is important to track the compound into the well basis, as the integrity of each sample might be different due to differ-

ent storage condition and usage. The inventory database is linked to the registration database that contains all the relevant compound properties and the screening database, which contains all the biological results. The main customer of the Compound Bank is usually the Screening Departments. For primary screening, scientists generate a request indicating which plates are required to carry out a specific screening run. These plates are barcoded and then provided to the Compound Bank for assay-plate dispensement. The source plates used for the dispensation are in 384-well format containing 50 µL of a 1 m*M* solution in neat DMSO. These plates exist in the inventory database and volumes are debited on dispensement. The volume normally requested ranges from 300 nL to 1 µL, and thus these source plates have enough volume for 30–40 screens. Dose-response analysis and structural-verification samples are retrieved from vials containing a 10 m*M* solution in neat DMSO and the volume normally requested ranges from 5–10 µL and dispensed directly into the first column (for 96-well plates) or the first two columns (for 384-well plates) of plates suitable for serial dilution and IC_{50}s determination or for liquid chromatography/mass spectometry (LC/MS) structural integrity analysis. A barcode is generated for each sample using the inventory database and these barcodes are imported to the screening or analytical database as sample IDs. The barcode system allows the screening or analytical team to quickly access data necessary to assay the plate and calculate raw data to generate results. The available volume for the 10 m*M* source (vial or plate) is debited upon dispensements.

4. QC of the Collection

A very important issue, usually overlooked when starting a compound collection, is the quality of the compounds introduced into it. It is very important that the compounds that enter the collection are checked to verify that they comply with a certain requirement of purity and integrity. Only misleading results can be expected for a collection with compounds with very different or poor levels of purity. The second issue is that many compounds slowly degrade and that the stability is highly dependent on the storage conditions, but at the same time there is no existing storage method that can guarantee no decomposition of the compounds in the collection. It is clear then that any hit from the HTS should be considered as preliminary until a confirmation of the integrity of the screening sample has been obtained.

4.1. QC of Screening Hits

One of the key steps after the finding of active compounds with desirable chemical and biological properties is the structural verification of these hits. The source of these samples for QC could either be a sample of the solution

from the confirmation vial or a sample of the solid if available, or both. It is important to track the sample to container level with dispensement barcode. Ideally, when screening hits are reported to the team, the information submitted should include not only activity but also recently obtained physical QC data. With the advent of parallel synthesis, solids normally are not available for confirmation of activity and structure and in these cases, solution samples are used. If even this type of samples is not available for a given compound, then a resynthesis of the desired compound is undertaken.

One important issue to consider is that many times hits from screening solutions which yielded >95% purity by LC/MS and nuclear magnetic resonance (NMR) failed to confirm activity upon resynthesis. This can be caused by small nuisance impurities present in the sample, which inhibit the target nonspecifically. This information should be noted in the database for future reference.

This QC topic is a very controversial one and there is not a clear response yet to questions such as what the best conditions for storing compounds are, how often should samples be QCed (only when showing up as hits or in a regular schedule?), should all file compounds be QCed before inclusion into the HTS collection, what should be the minimum purity for a compound to be maintained in the screening solution collection, and how practical is it to continually QC and remove bad compounds that are part of the HTS collection? These and other questions do not yet have clear answers, but our opinion is that all the compounds should be QCed prior to inclusion into the HTS collection, that all hits should also be analyzed, and that a regularly scheduled QC of all the compounds would be ideal, even though this creates an extra load of work. These results would be very useful in the medium term to define which are the best storage conditions and to define patterns of instability based on the structure of the compounds, so that shelf life could be assigned based on those data. Once a compound does not pass the QC criteria, ideally it should be removed from the collection or flagged in the database.

Another important consideration in the management of a compound collection is that all liquid handlers and the informatic management systems should be QCed periodically in order to ensure that the materials provided to HTS are always in the requested amount and concentration. The system should provide correct locations and remaining amounts for samples and the liquid handlers should dispense the right amounts in the right wells or tubes. The QC of the informatic systems is inherent to the process of retrieving samples to be assayed.

4.2. QC of Instruments

A process has been developed by us for the QC of the instruments ***(3)***. In summary, a standard operating procedure was used for liquid handling quality

control (LHQC) that has enabled us to evaluate liquid handler performance on two levels. The first level provides for routine daily testing on existing instrumentation while the second level allows for evaluation of new products available in the market as candidates for our HTS process. In practice, prior to the dispensement of compounds onto assay plates, three dye plates are run according to established guidelines. An Excel template has been created to help in the analysis of the results. In order to pass this test, the results from the three plates will have to comply with all the following rules:

- Systematic errors: not acceptable if all three plates have the same well empty.
- Nonsystematic errors (total number of unacceptable errors in the three plates).
  - ♦ Overdispensement
    - – Assay sensitive to DMSO: <1% wells with >50% extra volume dispensed with respect to the average.
    - – Assay nonsensitive to DMSO: <1% wells with >100% extra volume dispensed with respect to the average.
  - ♦ Underdispensement
    - – <1% wells with <50% volume dispensed with respect to the average.
    - – <0.5% wells with <10% volume dispensed with respect to the average.
- Accuracy: every plate should have as average the expected volume of solution ± 20%.
- Precision: every plate should show an overall CV<10% for volumes >200 nL and <20% for volumes <200 nL.

5. Summary

The increasing size of the collections used for drug-discovery purposes has demanded both hardware and software automation of compound management in order to cope with the increasing demands of HTS. Splitting the collection into a number of copies in different formats is a desirable approach to keep a balance between rapid response to the demands and best storage conditions. Flexibility to different assay configurations can be provided with the appropriate selection of liquid handlers, and the informatic management systems should be accessible to keep track of the samples and link them to a variety of information that can help interpret HTS data. In this respect, QC data on the compounds and quality checks of the equipment used are highly desirable. It is also prudent for a large organization with different research sites to have a unified database and compatible plate format and concentration in order to be able to exchange samples and share screening results. The accomplishment of all the previous requirements is the only way to ensure an efficient and effective compound library management.

References

1 Gosnell, P. A., Hilton, A. D., Anderson, L. M., Wilkins, L., and Janzen, W. P. (1997) Compound library management in high throughput screening. *J. of Biomol. Screening* **2,** 99–102.

2. Kim, S. S., Gong, Y. D., and Yoo, S. (2000) Chemical Library Management. *Korean J. Med. Chem.* **10,** 22–38.
3. Taylor, P. B., Ashman, S., Bond, B. C., Hertzberg, R. P., Kubala, S. M., Macarron, R., et al. (2000) A Standard Operating Procedure for Assessing Liquid Handler Performance in HTS. Society for Biomolecular Screening, 6th Annual Convention and Exhibition, Vancouver, British Columbia, Canada, September 6–9.

Appendix

In this appendix some information about existing equipment is described with references to the vendors web pages when available.

1. Haywain. Compounds are being weighed manually in most organizations due the impracticability of handling different physical form of the solids. The Automation Partnership has the Haywain™ module that takes racks of dry powders in vials and weighs 0.1–3 mg accurately into vials to weigh free-flowing dry powder. Web address: http://www.autoprt.co.uk
2. Liquid-dispensing units.
   a. Multidrop. This is a rapid dispensing system very well suited for the quick dispensation of buffer onto plates. It provides a great throughput with good levels of accuracy. Web address: http://www.labsystems.fi/products/dispensers/index.htm
   b. Tomtec's Quadra: There are a number of different models in the market able to dispense from 96–1536 wells. When equipped with a low volume head, it is able to dispense down to 0.5 μL with great accuracy. It is also possible to integrate stackers for the automatic feeding of the machine. Web address: http://www.tomtec.com/Pages/products.html
   c. Robbins' Hydra and Tango: These machines are very well-suited for the dispensation of very low volumes (50 nL) and have 384–1536 reformatting capacity. The difference between the Tango and the Hydra is that the former has more deck positions to accommodate a larger number of plates. The system has been fitted with sensors to avoid mispositioning of the plates and with flexible needles to avoid replacement when any crash occurs. Web address: http://www.tomtec.com/Pages/products.html
   d. CCS/Packard's PlateTrack: The PlateTrak is an automated microplate processing system with stackers that in a single platform can accommodate plate replication (96 to 96, 384 to 384) and reformatting (4×96 to 384 and 4×384 to 1536) with acceptable CVs dispensing 0.5 μL to dry plate. The bi-directional conveyor provides secure microplate handling, accurate dispensing and efficient automation in configuration from 4–16 modules. Web Address: http://www.ccspackard.com
   e. Tecan's Genesis. This is a very well-known piece of equipment that is able to provide great flexibility in the reformatting process due to its cherry picking capability and its ability to recognize a great number of containers from flasks to plates and vials. It is fully programmable so any mapping can be achieved.

When equipped with piezo tips it can dispense very low volumes. Web address: http://www.tecan.com/index_tecan.htm

f. Beckman's BIOMEK2000: Also a pretty standard piece of equipment that shares many of the functions of the Genesis. Web address: http://www.beckman.com/Beckman/biorsrch/prodinfo/automated_solutions/biomek2k.asp

3. Integrator systems.
   a. Hudson's Plate crane. Web address: http://www.hudsoncontrol.com/products/platecrane.htm
   b. Tecan's Twister. Web address: http://www.tecan.com/tec_main_twister.htm

   These two systems are able to move plates to given positions and greatly help to make process run unattended.

4. Middle-size store and liquid-handlers systems.
   a. Tecan's Molbank: This system integrates an automated storage system able to automatically retrieve plates and boxes stored at the user's defined conditions. It is linked to a Genesis system so it can create plates directly without human intervention. It is ideally suited for small collections. Web address: http://www.tecan.com/index_tecan.htm