Chapter 3

Informatics in Compound Library Management

Mark Warne and Louise Pemberton

Abstract

The ability to accurately and efficiently manage inventory is critical to ensure cost-effectiveness and guarantee integrity of samples in drug discovery. While many large companies utilise both customised and off-the-shelf automated systems for compound library management, these systems do not come cheaply. Without doubt, for large pharma the return on investment for one of these systems can be justified; however, the upfront cost is typically prohibitive for smaller businesses looking to stretch their limited cash reserves as far as possible. At Exelgen we have shown that for any business with the combination of fit-for-purpose informatics, relatively inexpensive laboratory hardware and well-constructed SOPs (standard operating procedures), it is possible to undertake cost-effective, large-scale compound library management in a small business environment. The informatics and hardware environment deployed at Exelgen are described in detail.

Key words: Informatics, Drug discovery, Compound management.

1. Introduction

For Exelgen *(1)*, when providing small-molecule compounds to the pharmaceutical and life science industries, the ability to accurately and efficiently manage inventory is critical to ensure cost-effectiveness and guarantee integrity of samples for the customer.

Equally, companies receiving compounds, whether established pharmaceutical giants or emerging biotechnology companies, will have a need to manage their ever-growing compound collections that serve as a bedrock for biological screening efforts. To ensure the financial outlay associated with high-throughput screening (HTS) is justifiable, it is critical that compounds for testing be located simply and that one be confident of the material identity, purity and weight without duplicated effort.

W.P. Janzen, P. Bernasconi (eds.), *High Throughput Screening, Methods and Protocols, Second Edition, vol. 565*

DOI 10.1007/978-1-60327-258-2_3, Springerprotocols.com

While many larger companies utilise both customised and off-the-shelf automated systems for compound library management, these systems do not come cheaply. For example, the fully bespoke system implemented by AstraZeneca *(2)* costs many millions of dollars *(3)*. Without doubt, for large pharma the return on investment for one of these systems can be justified; however, the upfront cost is typically prohibitive for smaller businesses looking to stretch their limited cash reserves as far as possible.

At Exelgen we have shown that for any business with the combination of fit-for-purpose informatics, relatively inexpensive laboratory hardware and well-constructed SOPs (standard operating procedures), it is possible to undertake cost-effective, large-scale compound library management in a small business environment.

2. Key Components in Compound Library Management

Compound library management brings together all the principle components of daily laboratory work processes in a single cohesive entity as depicted in **Fig. 3.1**.

To be able to rapidly locate stored compounds and be confident of the material identity, purity and weight, effective compound library management system requires three key components:

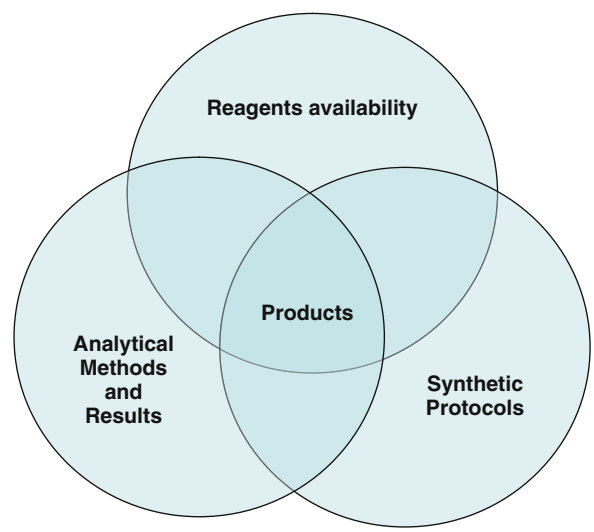

Fig. 3.1. Principle components of daily laboratory work processes are brought together in a single cohesive entity.

- An integrated informatics environment allowing rapid access to all data associated with a compound
- Analytical hardware and infrastructure
- Suitable storage and compound-handling infrastructure

The discussion of analytical methodologies and standards are well known and extensively covered in the literature *(4)*. Exelgen has not looked to reinvent the wheel in its implementation of these techniques and methodologies. Using standard analytical equipment combined with some customised assay protocols/conditions, data are interpreted using commercially available software packages *(5)*. The data generated are then simply parsed from the analytical instruments directly into the informatics system and stored in its raw format for retrieval as necessary.

Discussion of storage solutions was the primary focus of previous editions of this book *(6)*, and therefore, in this edition of the chapter we shall focus on the implementation of an informatics environment and the use of appropriate SOPs.

3. Compound Library Management Informatics

When Exelgen was founded back in 1997, throughout the industry the idea of fully integrated compound library management was still in its infancy. Today, off-the-shelf and customised solutions are commonplace; however, at the time Exelgen was looking to implement its own system, it was reliant on its then parent company Tripos *(7)* to develop a solution. While not an option available to all businesses, this enabled the system to be built entirely fit-for-purpose.

Named ChemCore™, Exelgen's informatics system combines chemical registration, inventory and sample management. The system goes beyond the traditional definition of compound library management, specifically logging reagent purchase, product synthesis, analysis, use, movement and transfer of all compounds, essentially following a compound from cradle to grave. Used throughout Exelgen by computational, research, analytical and high-throughput chemists, chemical inventory managers and sample preparation scientists, this informatics system exists in a single environment, ensuring its effectiveness in an integrated workflow. In addition to being fully integrated, the ChemCore system is able to easily adapt to scale, an essential requirement since Exelgen synthesizes and analyses in excess of 300,000 compounds annually. This means the ChemCore system is required to track 20 million containers, 10 million structures, 6 million analyses, 1 million purifications and 5 million reactions.

While Exelgen was fortunate that it could have an entirely tailor-made informatics solutions, it is now possible to purchase a number of software tools off-the-shelf, or semi-customized for this task *(8)*. ChemCore itself was also commercialized to some extent, and in addition to being deployed at Exelgen, it has been implemented in a customized form at Schering Ag, now part of Bayer Healthcare *(9)*.

The following narrative describes what we believe are the essential features to be considered when implementing an informatics solution for a compound library management system.

4. Informatics System Requirements

4.1. Ensuring an Informatics System Is Fit for Purpose

One must ensure that when implementing an informatics system for compound library management, it is fit for purpose. The Exelgen process for library production is depicted in **Fig. 3.2** and the ChemCore system was designed to integrate with each element of the process.

Off-the-shelf systems and semi-customizable systems are usually built using multiple components and selecting the correct components is key. At Exelgen the system consists of three main pieces:

- Registration
- Reagent selection and ordering
- Inventory management

Additionally, a graphical user interface (GUI) front end provides an effective electronic lab notebook where synthetic protocols and alike can be associated with product containers.

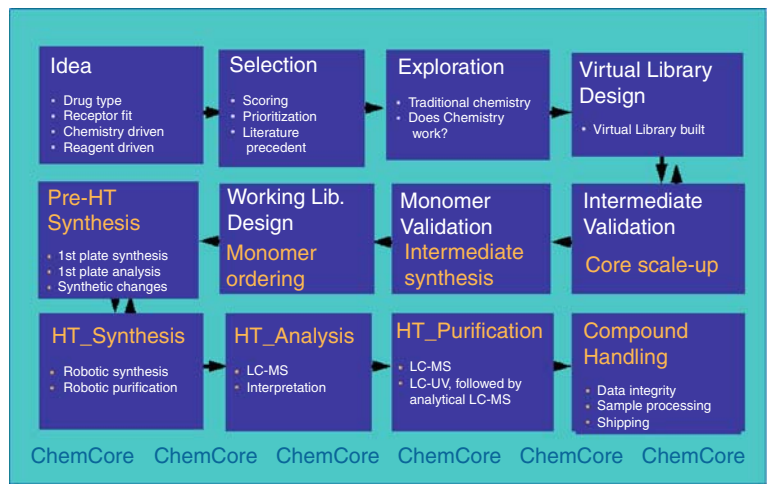

Fig. 3.2. Exelgen library production process.

4.2. Registration of Compounds/Structures

Compound registration is central to any compound library management system. At the entry level, either as a single structure or as a series of structures from an SD file, compounds/structures should be registered with a unique compound identifier.

In a departure from the standard methodology of compound registration, Exelgen scientists register samples they intend to make rather than, as is the norm for most scientists, final characterized samples. The rationale for this is twofold. First, the fact that the scientist did not get the expected product should not be considered a failure, but rather an opportunity to gather further information about the proposed reaction and potentially seek learning into the root causes for the reaction not working. Second, at Exelgen, synthesis of between 1 to 200 samples at a time would preclude the scientist from registering each product individually. Proposed structures are therefore registered using high-throughput computational methodologies.

Subject to the nature and size of your business, it may also be important to consider how you enforce compound security within your management system. At a minimum, one should consider a structure-based system that one can control for personnel access, ensuring any requirements one may have with regard to exclusivity and confidentiality.

The registration of structures at Exelgen is depicted in **Fig. 3.3**, demonstrating the registration and security dialogues.

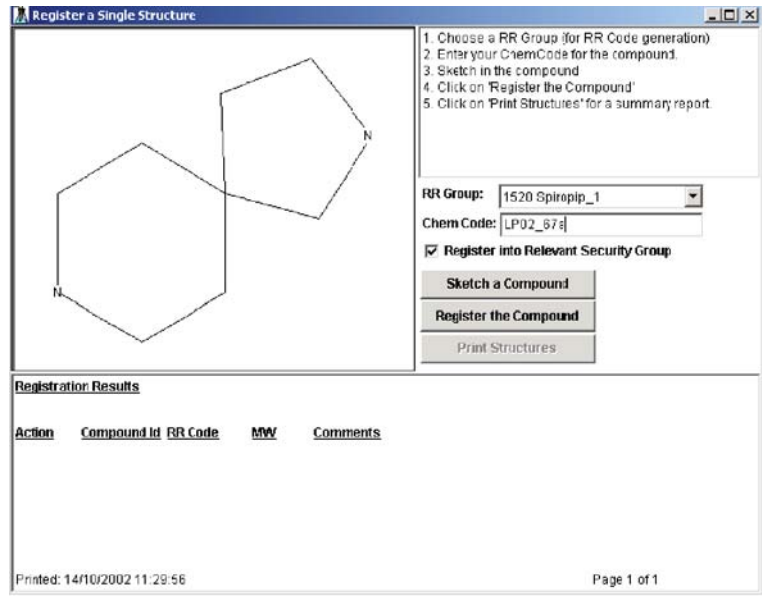

Fig. 3.3. Registration and security dialogues.

Beyond the scope and requirements of most compound library management systems is the capability to implement an ideas management. However, the contract research services nature of the Exelgen business necessitates the provision of security around not just physical compounds but also ideas. This means we have also implemented the registration of ideas and reactions as well. These dialogues are depicted in **Figs. 3.4** and **3.5**.

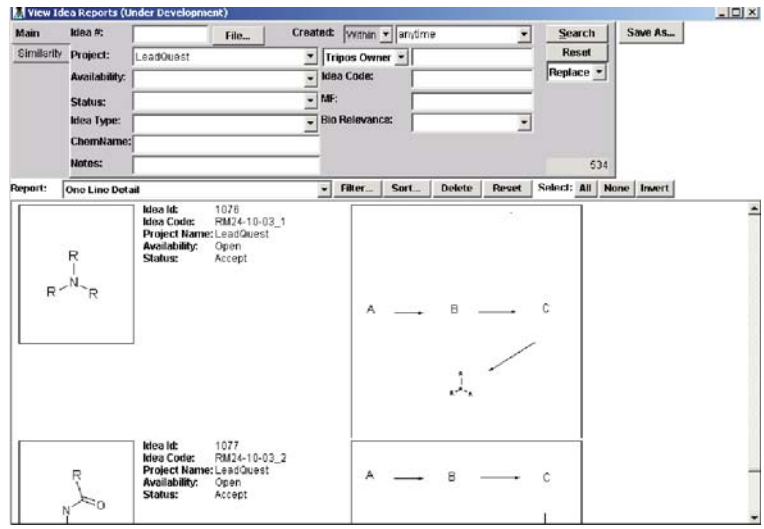

Fig. 3.4. Idea registration and reaction registration dialogues.

4.3. Reagent Management

While not essential for all companies, should one be engaged in compound synthesis, a reagent management component is a particularly useful utility within a compound library management system. In the same way that it is useful to be able to track purity, age and stock quantity of a screening compound, when reagents are entered into the system, it becomes possible to determine whether to retrieve reagents from in-house stores or purchase new reagents. By adopting a reagent management process, in addition to a more cost-effective management of internal reagent inventory, one can enable improvements in timing and quality of reagent delivery from suppliers.

In considering the wealth of reagent suppliers available, it is not inconceivable that precious scientific resource is deployed in selecting samples from one of these reagent suppliers. However, once implemented, a simple algorithm that scores the supplier based on a variety of metrics including quality of service, reagent pack size and cost per unit of material can equally propose a supplier. When considering single reagent purchase, the time saved is inconsequential; however, for the generation of sparse matrix sample libraries, where selection of several hundred reagents from a variety of suppliers is a requirement, the task becomes much more cumbersome.

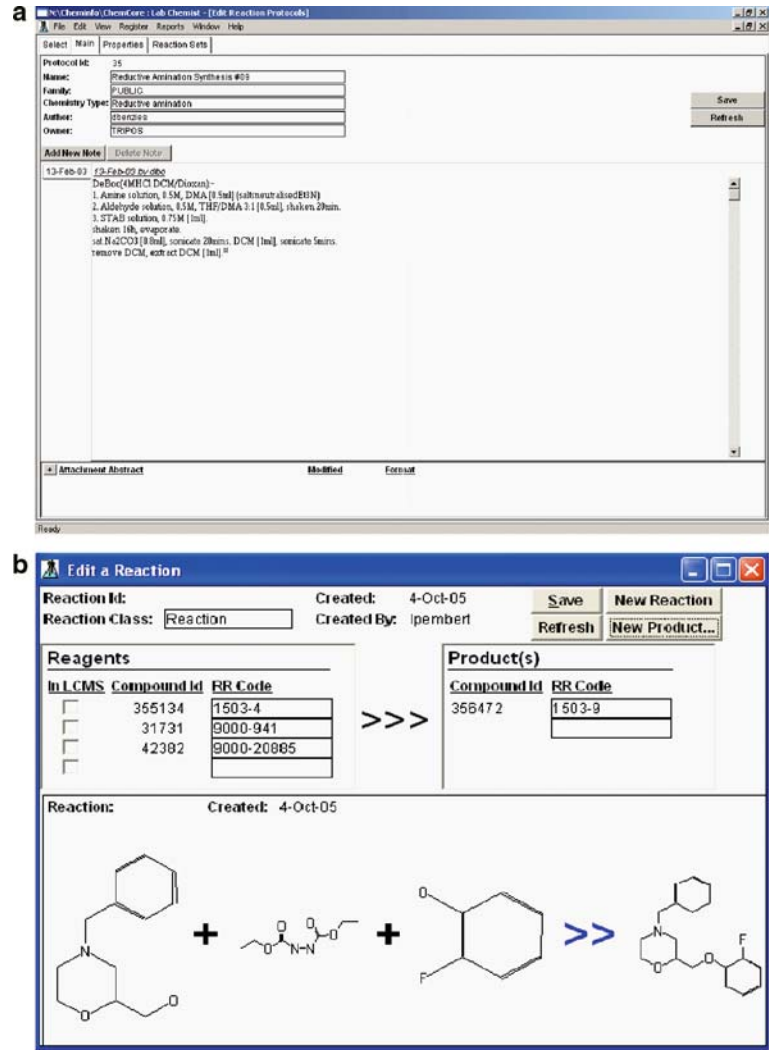

Fig. 3.5. (**A**) Addition of synthetic protocols tied to reaction types. (**B**) Addition of synthetic protocols ties a unique identifier and text-based synthetic protocol to a series of products.

Having the system select, based on the criteria already outlined, a suitable supplier can be a saving of several hours of manpower. This system, called the price-supplier score, is depicted in **Fig. 3.6**. In addition to using an algorithm scoring system for suppliers, at Exelgen the reagent management group supports the scientific team with the provision of reagent samples, ultimately allowing the scientists to focus their area of expertise, scientific research and development. The price-supplier score is further supplemented with financial actuals reported as part of the acquisition process. With the advent of online ordering and invoicing, the importation of this data back into the system is simple and worthwhile.

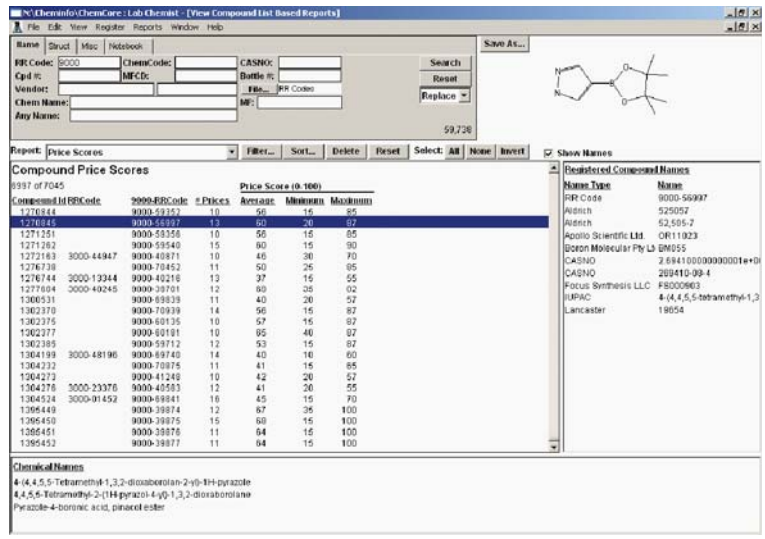

Fig. 3.6 Exelgen's price-supplier scoring system.

The programming logic for reagent management mimics that of screening compound management, but fields associated with the structures are modified to be fit for purpose.

4.4. Retrieving Compound Structures and Associated Information

With structures recorded in an informatics system, key to being able to retrieve the information quickly and effectively is a flexible search engine. One does not have to be worrying about whether the correct naming convention is being utilized or on which plate a particular compound may be located. A multi-parameter search algorithm that can search compound identifiers, exact structures, substructures and all associated data fields is essential in any system.

At Exelgen our customized system uses the approach to enable the searching of reaction schemes encoded in the system, including fields such as exact product, product substructure, reagent substructure and product and reagent substructure and the specific synthetic protocol. Dialogues depicting these search options are depicted in **Fig. 3.7**.

4.5. Benefits of Associating Newly Generated Physical Data with Structures in the Informatics System

With structures already registered into the informatics system, at the point of synthesis, a unique bar code on the reaction vessel is linked to the structure by association with the protocol ID. This bar code serves as the linchpin for all the associated physical transformations during the sample lifetime. Post-synthesis, the relationship between the parent reaction vessel and the derived samples, be it synthetic or analytical ('daughter samples'), is maintained.

For example, analysis of samples is performed using Waters and Agilent *(10, 11)* samples, with the results being interpreted through the use of the standard MassLynx *(5)* software. The raw

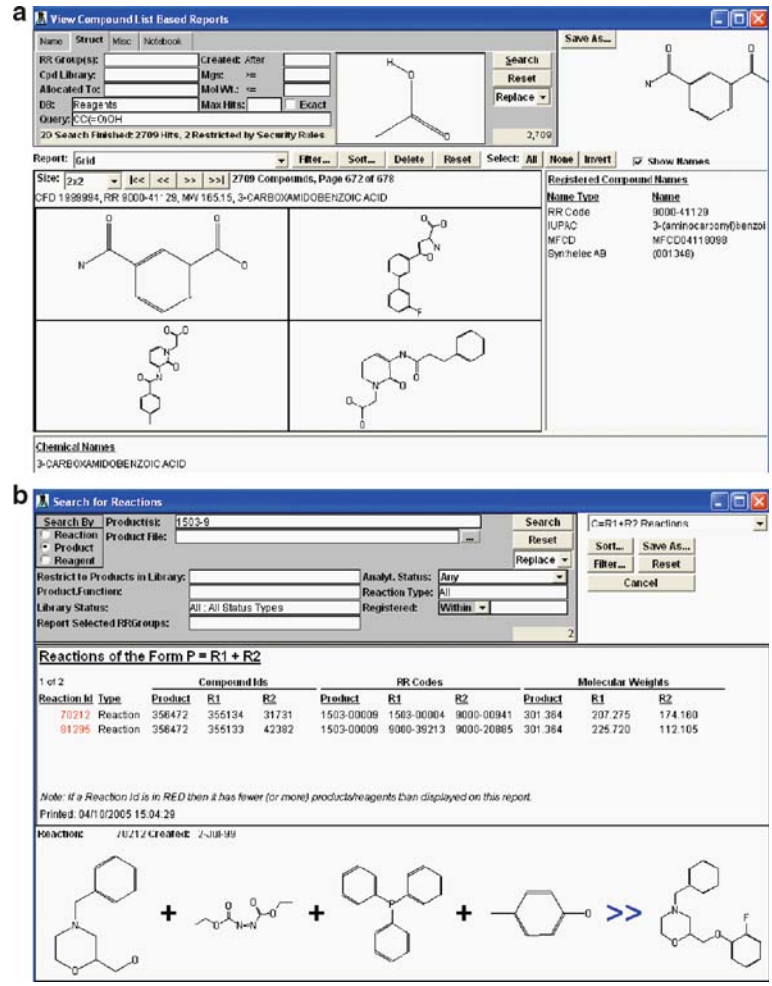

Fig. 3.7. Compound and reaction search dialogues.

data are stored, with the interpreted purity value being parsed into the compound management system. There can be multiple analytical results associated with each structure, but only one unique analysis per bottle. Consequently, for a final product or reagent, when daughter samples are created directly from the parent bottle, the analytical ID of the parent is automatically transferred. These recorded purity values can then be consolidated and utilized to provide decision support in synthetic operations.

When synthesizing hit sets around a particular product of interest, it may be a requirement for the same reagent to be used multiple times. In a set of 20 reactions using the same reagent, the poor recovery of product from one reaction would not necessarily be a cause for alarm. However, should this be the case for 15 reactions, further investigation would be warranted. In reviewing the results for a given synthesis effort, the analytical results can be used to depict the following:

- Impact of reagent
- Impact of reaction conditions
- Investigation of poor success rates

Once armed with this information, ChemCore can be interrogated to determine which reagent batch was used and further investigation can be initiated.

Additionally, when considering the purification of samples, whether it be the selection of samples for purification or the selection of purification method, data can be extracted from out informatics system allowing us to proactively implement changes to our process to improve sample recovery in terms of purity and sample amount.

4.6. Inventory Management

As with all warehouses, being able to locate the item of interest is vital to the operation of the business. With sample inventory management, the compound products and reagents are likely to be in multiple locations, be that stored at –20°C in a solution or stored in various quantities 'neatly' on the shelf. At Exelgen the ChemCore system will accommodate a strategy to return the material to exactly the same location, as required for reagents, which by their very nature have specific storage requirements or segregation conditions. In addition, the system will have to put the samples back in an empty location and tell the system where the samples are stored as in the case of our product storage. Exelgen, as a small business, does not have the luxury of a multi-million dollar storage facility, as is frequent within Major Pharma; however, by working to our strengths, we have built a robust inventory management process that, with a small team of dedicated sample technicians, enables us to compete on the same playing field for sample delivery.

Each sample is tracked using a unique system-assigned bar code encoded with a code-128 symbology or 2D bar codes; different samples of the same batch or of the same compound are tracked using their own unique bar code. Plates, storage bins, rooms and sites are also hierarchically bar coded in a similar manner to provide a full, unique inventory location for each sample. All movement operations and material-handling operations, such as liquid handling and hand weighing, are tracked for each sample.

Once the parent containers are released into circulation, and daughter samples generated from the parent, the history for each sample diverges. Equally, should the same structure be remade, and the material stored, the bar code will record the batch history as being different from the original sample. In a simple search for the structure, both batches of material will be returned; on further investigation, the differences in sample history will be obvious.

An additional twist with sample management is one of sample integrity. Samples that have been stored for many years may have the tendency, however good the storage system be, to degrade. At Exelgen we have a rolling QC process, where, with multiple

analysis associated with each sample bottle, we can evaluate the potential degradation of samples and cost-effectively determine what the fate of such a sample should be.

4.7. Sample Formatting and Delivery

With the advent of HTS, samples formats have been developed internally to match each company's internal process. As a service provider to the pharma/life science industry, Exelgen is required to be the ultimate flexible partner, since each company has a subtly different delivery requirement based on these 'in-house' processes. Over the past decade, as methodology in HTS has become more sophisticated, we have seen the requirements move from the provision of a single sample, such as 1 mg in a 1.4-ml tube, to samples being requested as micromolar concentrations and deliver higher density plate formats with more complex delivery strategies. Since the process within each company is different, the data requirements are also different. Delivery of samples to compound acquisition teams must be supported with corresponding data in their specified format. Data are typically requested as a single SD file containing all standard information; however, requirements have been as diverse as sketch files, with corresponding data supplied in .csv format or in a Microsoft® Access™ database *(12)* not unheard of.

While this may not be perceived as arduous, typically pharma provide their specific containers to meet their HTS sample format requirements. In an effort to keep the integrity of the internal bar code numbering system within ChemCore intact, we have had to customize our system to ensure that each sample can be assigned a proxy bar code, usually the customer-supplied vial identifier, which can be in a variety of formats including 1D symbologies and 2D bar codes.

4.8. Avoiding Errors

Having automated the tracking of samples from registration of reagents and products, through analysis, purification and preparation for shipping, one can imagine that should an error in handling occur, the ability to reproduce the same error many hundreds of times is not inconceivable. Therefore, having an intervening and final QA prior to delivery ensures that the sample integrity is absolute. Exelgen follows a series of SOPs designed to provide this assurance, which include a random correlation of actual data to the system anticipated data by intrusive sampling throughout the process.

5. Case Study: Provision of 50,000 Samples in Customer-Specified Format

Exelgen was required to provide 50,000 of its off-the-shelf LeadQuest compounds to a client in the following complex format:

1 × 625 plates, 96-well format containing 20 µl leftover LQ library

1 × 157 plates, 384-well format containing 15 µl Cherry Picking mother

1 × 157 plates, 384-well format containing 45 µl Mother leftover

1 × 157 plates, 384-well format containing 5 µl daughter plate

Source plates for the products to be delivered were 50,000 compounds, at fixed concentration in DMSO, plated as 80 compounds per rack with columns 1 and 12 empty.

Based on the customer-specified format, the following series of processes were enabled using the ChemCore informatics system, with manpower of less than four people in over 2 weeks.

(a) Source samples were thawed, but at a rate to ensure all thawed samples could be plated within a working day. This was to minimise freeze-thaw cycles.

(b) For each 96-well source plate, create 2 × 384 master plates (**Fig.** 3.8).

(c) Freeze remaining samples after plating of 2 × 384 master plates.

(d) Freeze Cherry Picking mother.

(e) Dilute master plate.

(f) Create six daughter plates.

(g) Freeze and store.

This whole process from library plate to assay plate is depicted in **Fig.** 3.9.

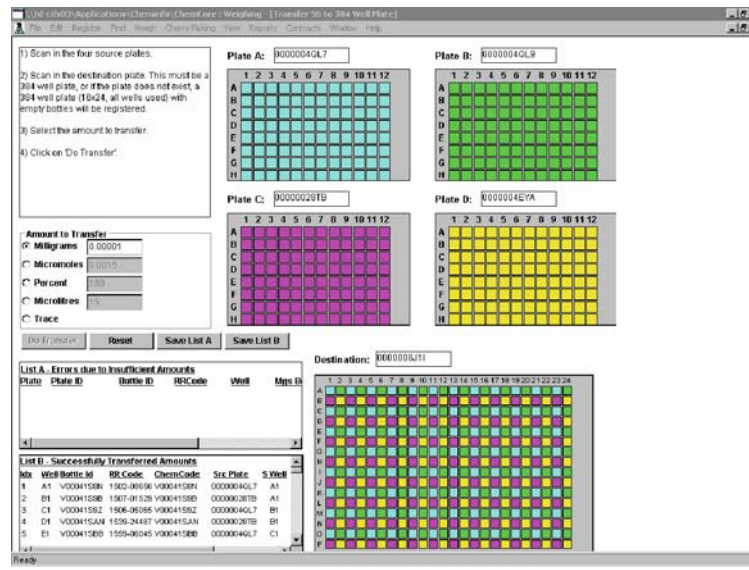

Fig. 3.8. Transfer of 96- to 384-well plate format.

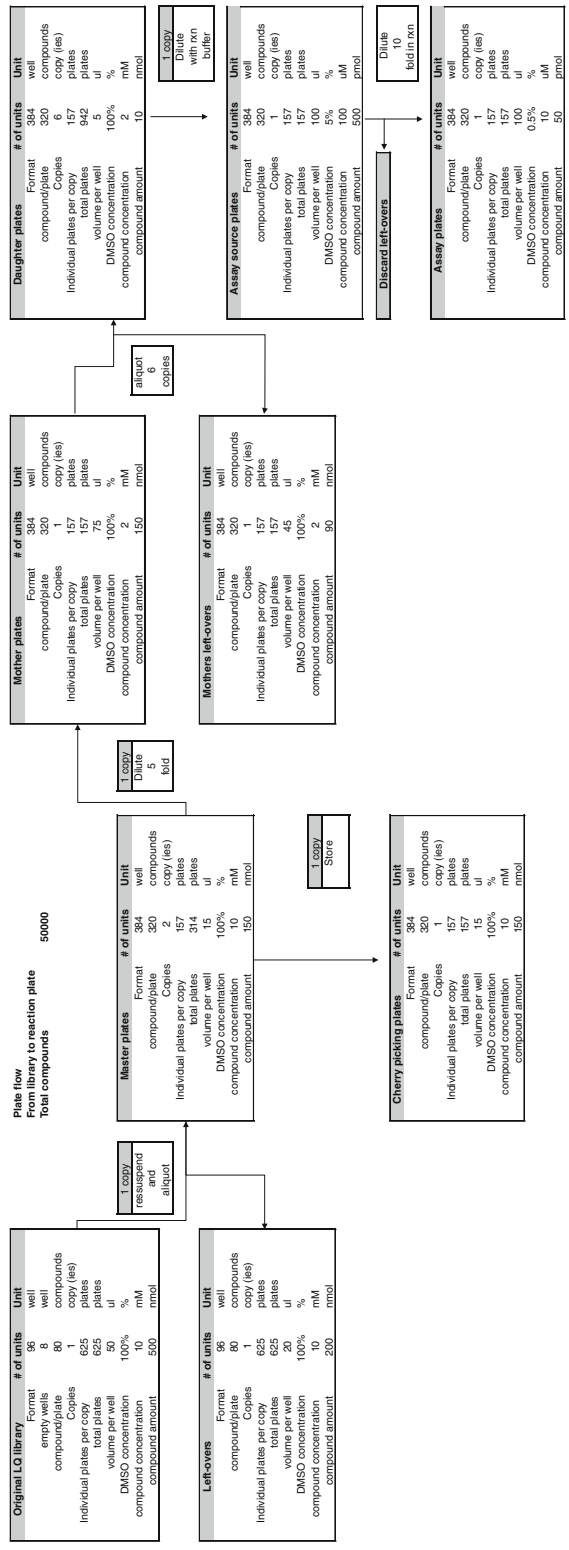

Fig. 3.9. Case study flow chart.

6. Summary

As discussed, the combination of fit-for-purpose informatics, relatively inexpensive laboratory hardware and well-constructed SOPs (standard operating procedures) enables a small business, such as Exelgen, to have a flexible compound library management system that competes with the fully automated systems deployed within large pharma. The benefits extolled on the businesses are then twofold.

First, the benefit of improved inventory since ChemCore enables us to work to a truly just-in-time philosophy:

- For example, we can manage our reagent inventory, so we know when and how much to order needed reagents.
- Similar to managing our product inventory, we know when and how much to synthesize needed compounds.
- We reduce our excess inventory and therefore minimize hazardous material on site. Because the turnover of reagents and products is relatively quick, an improvement in reagent quality is also observed.

Second, work process improvements are seen because the centralized database of information enables scientists to

- plan their work effectively, including anticipation of reagents and provision of analytical instrumentation time; and
- view data within the database and learn from past successes and failures, therefore avoiding redundant experiments.

Ultimately, the benefit to Exelgen is one of a more streamlined work flow that supports our scientists to make better decisions faster. To evidence this, we have outlined a compound plating exercise, which was done with the minimum of resource in the shortest possible timeframe in the above case study.

References

1. Exelgen Ltd. (formerly Tripos Discovery Research Ltd.), Bude-Stratton Business Park, Bude, Cornwall, EX23 8LY UK
2. AstraZeneca UK Ltd., AstraZeneca R&D Alderley Park, Alderley Park, Macclesfield, Cheshire, SK10 4TF UK
3. IQPC Best Practice for Compound Management & Integrity 2005, May 23rd–25th 2005, The Selfridge Hotel, London
4. Changing Requirements of Purification as Drug Discovery Programs Evolve from Hit Discovery, John Isbell, J. Comb. Chem., 2008; 10(2), pp. 150–157
5. MassLynx, Waters Corporation, 34 Maple Street, Milford, MA 01757 USA
6. Compound Library Management: An Overview of an Automated System Wilma W. Keighley, Terry P. Wood, High Throughput Screening: Methods and Protocols Series: Methods in Molecular Biology, Apr-18-2002; Vol. 190, pp. 129–152
7. Tripos, 1699 South Hanley Road St. Louis, MO 63144-2319 USA
8. For example, Symyx Technologies, Inc. 3100 Central Expressway Santa Clara, CA 95051 and ID Business Solutions Ltd., 2 Occam Court, Surrey Research Park, Guildford, Surrey, GU2 7QB UK
9. Bayer Healthcare, Leverkusen, Germany
10. Waters Corporation, 34 Maple Street, Milford, MA 01757 USA
11. Agilent Technologies, Inc., 5301 Stevens Creek Blv, Santa Clara CA 95051 USA
12. Microsoft Corporation, 1 Microsoft Way, Redmond WA, USA