Chapter 2

Creation of a Small High-Throughput Screening Facility

Tod Flak

Abstract

The creation of a high-throughput screening facility within an organization is a difficult task, requiring a substantial investment of time, money, and organizational effort. Major issues to consider include the selection of equipment, the establishment of data analysis methodologies, and the formation of a group having the necessary competencies. If done properly, it is possible to build a screening system in incremental steps, adding new pieces of equipment and data analysis modules as the need grows. Based upon our experience with the creation of a small screening service, we present some guidelines to consider in planning a screening facility.

Key words: High-throughput screening, HTS, Cell-based assay, Instrumentation, Design, Planning, Robotic system.

1. Introduction

High-throughput screening (HTS) has been used for several decades in the search for new lead compounds. The large investment in instrumentation and development of the necessary expertise is a large commitment and has previously been largely in the realm of major pharmaceutical companies. However, in recent years, many smaller organizations have become interested in carrying out HTS campaigns. This is due to the development of many small biotechnology companies with only a few drug targets, as well as academic groups that have become interested in doing their own drug discovery. There are now multiple HTS service providers who will run HTS campaigns for a fee, often also having the option of providing chemical libraries. Nevertheless, some small research organizations would prefer to create an internal HTS facility to screen their own drug targets, or possibly to offer screening services.

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

DOI 10.1007/978-1-60327-258-2_2, Springerprotocols.com

In this chapter, we would like to present one particular view of how to build a small high-throughput screening facility. Specifically, we will speak to our experience of evolving from the use of small pipetting workstations to a complete screening system, the creation of data handling systems, and the growth of a functional HTS group. We will discuss the pros and cons of do-it-yourself versus relying upon vendors, for both instrumentation and software. And we will show the evolution of a screening system over time. Please be aware that this is essentially a case study. Vendors that are mentioned are not necessarily the best choice but represent our particular choices.

In examination of the evolution of this small screening group, we will point out some of the important choices made along the way. Hopefully this case study can give some insights to guide others in building up an HTS facility.

2. Screening System Design

2.1. Evolution from Workstations

Axxam SpA in Milan, Italy, is a small organization that evolved from a group within Bayer HealthCare, which had been responsible for assay development and configuration to feed the ultra-high-throughput screening carried on by Bayer. The group was spun off as an independent organization in 2002 and started to provide assay development services for paying customers. Eventually the customers requested that the developed assay be screened on site rather than transferred back to them, creating the need for an HTS setup.

Within the organization there was already a strong background in cell-based and biochemical assay configuration, cell culture facilities, and related technical capabilities. However no one had ever examined more than a few hundred compounds, and there was little automated equipment to support a true HTS campaign. Axxam was faced with the task of building a complete screening facility, including selection of equipment, management and analysis of the data, and development of the necessary skilled personnel. Since there were no pre-existing contracts in place to pay for screening services, the budget was limited, dictating the decisions in two ways: the need to start small without limiting the ability to expand and favoring in-house solutions as opposed to relying upon commercial software and vendors.

Starting small meant evolving from individual workstations with built-in scalability to reach a fully automated robotic screening system. It should be emphasized that, with proper organization, a great deal of work can be accomplished using only

workstations and relatively simple data analysis methodologies. For example, it is quite feasible to screen several hundred thousand compounds using only pipetting workstations to prepare plates, placing plates manually into reader instruments, and processing data using Microsoft Excel©. That approach will not perform as fast as an automated HTS system, but if the number of screening campaigns is limited, it will be adequate, with faster implementation and troubleshooting than a fully automated operation.

2.2. Considerations for System Design

If you already understand pretty clearly the types of assays that will be run, the reading instrumentation that is necessary, and the ultimate level of automation desired in a screening system, you may be able to design a system from the outset that will meet your needs for many years. In this case, future expandability may not be a high priority. You may be able to design a compact system with limited expansion options.

However, if your goal is to start with a small instrumentation investment and gradually expand, then obviously you must select a system that is flexible and open. In general, all vendors offer packages that are customizable with your desired components, but some are more amenable to future expansion than others. You should consider how much work will be necessary to add new instruments. Some systems may be difficult to expand simply because of physical limitations of space, such as the case of a central robotic arm that can access only a limited work envelope. Other systems may have limitations that are more in the realm of software. You should ask the vendor about the possibilities of integration of various types of equipment that you might imagine using in the future. If possible, it should even be possible to add a simple component into the system with little or no reliance upon the original vendor.

2.2.1. Type of Transport

There are several different robotic transport mechanisms that are prevalent. The reasons to consider the differences in transport mechanisms are that the mechanism you choose will have an impact upon the system throughput, the ease of expansion, and the overall size of the final system. For the facility that desires to start small and expand in several directions, ease of expansion will be quite important.

The most flexible systems should include plate transport mechanisms that can be readily adapted to accessing new instruments. The most typical example of this is the anthropomorphic arm with a plate gripper. With such arms the physical process of adding a new instrument into the system can be as simple as affixing the instrument to fixed position of the work surface and then teaching the plate loading/unloading movements to the robot. The issue of logically adding the new instrument into the

control software system can be much more complicated (discussed below). Such arms are also capable of reaching into confined spaces, such as shelves in a plate "hotel."

Anthropomorphic arm robots are flexible, but they do have some drawbacks. The gripping fingers need to have a great degree of structural strength, as that mechanism of gripping a plate is inherently difficult to maintain without slipping. Some key things to look for would include sharp points used to firmly grasp the plastic; pivoting pressure points to ensure even application of force; strong metal fingers; and secure mounting to the robot wrist assembly to ensure minimal play. It is also very important that the robots have some mechanism to sense gripping force so that it will sense a plate presence or absence. Another important force sensing that should be present is the ability to quickly stop if it encounters an unexpected obstacle; all axes of the robot should be able to detect collision with an obstacle. This capability is important to avoid the robot breaking or bending itself or some fixed labware when the inevitable mistake is made, for example, when teaching a new position or when something is left inside the work envelope that should not be there. It will happen, the only question is when. If the arm cannot understand that it is impacting something, it may very well have sufficient power to bend or break some components.

It is possible to design a system using only a central arm, with devices arrayed radially around the arm. A good commercially available example of such a system is the Velocity11 BioCel® product line, which features a central three-axis robot arm that moves plates in a tightly packed workspace. Having only an anthropomorphic arm at a fixed location obviously limits the instruments to a work envelope that can be reached by that arm. For some applications, this can be quite sufficient. Such systems can be perfect if you know clearly what you need and if you would be satisfied to buy a different system in the future to address new desires instead of expanding on the current system. These systems also have the benefit of being relatively compact, taking minimal floor space. One disadvantage is that the equipment integrated in the system may be very difficult to utilize in "manual" mode for some simple operations, since the operational side of the instruments is generally facing into the central robot. Some of these systems can be expanded, basically by handing a plate over from one work cell to another work cell, which may have its own robot arm.

A more traditional approach is to have some linear translation ability. This can mean simple carriages that move the plate or more complex tracks that move the anthropomorphic arm. One of the earliest robotic systems used in biological laboratories was the ORCA from Beckman Coulter (Sagian). This is a four-axis arm with a plate gripper, mounted on a linear track of 1–3 meters. They

have designed CORE systems around this transport system, with instruments lined up on both sides of the robot track. Similar systems have been developed by multiple companies, many with more robust robots (for example, Thermo CRS). A common problem with such systems is that the transport system becomes a bottleneck, the limitation of overall throughput, since it is responsible for all movements of plates. Some thought must be given to this issue, especially if you imagine adding many more instruments in the future.

There are also systems that rely upon conveyor belts to move plates from one device to the next in a sequential arrangement, with the plate generally moving under the active position of the device. Examples of this type include PerkinElmer Minitrak. This type of transport mechanism is simple and robust. However, these devices are limited in their ability to add random instruments and in the ability to use the devices in a different order. Incubations can become more complicated because they must remove a plate from the flow into a storage device. These devices would not be recommended for a start-up HTS group, unless the assays that you want to implement are very clear and amenable to such an assembly-line approach.

Another similar type of simple linear conveyor relies upon carriages to move plates from one part of the system to another, but with vertical movements to transfer plates to instruments or additional carriages. Some vendors that provide such mechanisms include CyBio AG and Hudson Control Group. At each instrument there must exist a simple mechanism to move the plate from the carriage to the instrument. In the case of CyBio this is accomplished by plate lifters and simple grippers, while Hudson uses their PlateCrane device which can lift a plate from above and rotate. Also fixed anthropomorphic robot arms can be added at strategic positions within such systems to move plates from the carriage to instruments.

2.2.2. Other Design Considerations

Typically the robot arm needs to be enclosed in a work cell with a safety interlock. Even robots that can detect collisions can cause significant harm if they hit a person. They have plenty of power, and they are heavy enough that they have a lot of momentum when moving. Another problem is that the robot may be still during long periods of time during execution of a protocol and then unexpectedly start moving. It is quite easy for a person to enter a room, unaware that the system is even active, and reach for a plate in the robot's work area at the same moment that the robot starts to move. For this reason, it is good to have a signal light on the system that indicates when it is active.

However, care must be taken to not make the safety system too difficult to work with. Whenever possible the interlock should be clear plastic; or in some cases, even a light curtain is acceptable.

When opened, it should not hinder access to the work cell. Also it should be possible to override the safety interlock – many times a technician needs to see up close something operating. Maybe the company safety officer will not be pleased by this idea, but it is important that someone has the ability to override the safety interlock.

2.2.3. Vendor

Obviously you should consider the reputation of the vendor, but equally important is the quality and relationship with the local representatives. Before making a final choice of vendor, you should make an effort to meet the technical people, not just the sales representatives. You need to meet people such as the system design engineers, software specialists, application specialists, and even service engineers/technicians, if possible. Having a good working relationship with these people is critical to getting the system that you want and in resolving problems that will occur in the future.

2.2.4. System Software

Another large consideration in choice of a screening system is the quality of the system control software. This software is the glue that unites all of the disparate methods that must run on each instrument, executing each in a carefully orchestrated sequence to move plates through the process with the maximal amount of parallelism to optimize the overall system throughput. There are many approaches to system control software, and every vendor will want to convince you that their software is wonderful. Some of the buzz words that are frequently seen are "dynamic rescheduling," "flexible timing," "LIMS connectivity," and many others. Some of these features are important and others are not. For example, the number of times that "dynamic rescheduling" is a valuable feature for you may be rather small, but this is a feature that can make the software much more complicated and expensive. You must decide which features are important for your application, and whether the lack of some feature is a strong enough reason to look to another vendor. The list of features that we would consider important to understand would include the following:

- How easy is it to integrate new instruments? Does the vendor have already written device drivers for most of the instruments you might imagine adding in the future? If not, what is the cost and time necessary to create the driver to allow integration? Is it possible for a skilled person in your staff to add a new instrument?
- How good is the error handling? Can the system automatically (or semiautomatically) recover from problems? Is there an easy way of continuing a run after a particular cycle has problems? Are there notification events, such as email, SMS, and flashing lights on the system?

- Is the scheduling dynamic or fixed? Can you add more plates into the run after it has been started? Can you use parts of the system in stand-alone mode (for example, a reader) if not being used in a schedule? Can the schedule be interrupted for a short run of another protocol, such as using the reader for a single plate?
- How "open" is the software for development by your staff? Is it well documented? Can users write external programs that can interact with the system? In our opinion, it is very important that there exists the possibility of execution of a user program (an executable program or a VBScript or a JavaScript program) and that these external programs should be able to exchange information with the system (for instance, plate barcodes, cycle number, database information, error conditions, etc.).

2.2.5. Future Expansion Possibilities

When planning a system, try to imagine where you might want to physically expand in the future. If possible, leave extra space for future instruments around an anthropomorphic arm. If you have a system that includes linear transport mechanisms, it may be possible to expand the system at the end of one of these transport sections by including a robot arm that can access the linear carriage. If possible, also leave extra space within the facility on the side system that might be a future expansion point. However, if space is at a premium in the room where the system is installed, it is possible to move the system in the future to a new location when more space is needed for expansion.

2.2.6. Flexibility and Throughput Considerations

Many managers are immediately interested in the potential throughput of a system. This is natural, since obviously the goal of high-throughput screening is to achieve high throughput! But it is important not to overemphasize the speed of the system, for several reasons. First, with some assays, it is the length of the reading that is the essential limiting factor. This is especially true for cell-based fluorescent dye assays. For example, a typical antagonist screening experiment using a membrane-potential dye may involve first injection of the test compounds, waiting several minutes for the compound effects to stabilize, then injection of the agonist and observation for another few minutes to record the possible antagonistic effects of the test compounds. Under these conditions, one experiment may require 5–10 minutes. In such a situation, the speed of the plate handling has a relatively minor effect upon the overall throughput of the system. Much more important to the overall output of the screening facility is the reliability of the screening system, the ability of the system to run unattended dependably (including overnight runs), and the capabilities of the support systems such as a cell culture.

There is also the consideration of what plate format you want to be using. For most standard screening applications, there are three choices for plate density in common use today: 96-well, 384-well, and 1536-well plates. While there may be the temptation to work with 1536-well plates in order to maximize throughput and conserve reagents, working at this density is typically rather difficult. Furthermore, there are limited options available for simultaneous pipetting of 1536 wells. The typical approach is to use a 384-channel pipettor head four times, indexed to the four quadrants of the plate. In this case, some of the throughput gain of the higher density will be sacrificed. Unless there is some overwhelming reason to use such a high-density plate, it is probably not a good choice for a young screening group.

The 384-well plate is a good compromise between miniaturization and ease of use. Most assays that work in 96-well plates can be translated to 384-well format with little difficulty. Virtually all vendors offer equipment that works with 384-well plates (pipettors, washers, readers, etc.). Furthermore, some equipment that works with 384-well plates can also be made to work with 1536-well plates so that you can utilize that format if necessary. A more significant consideration for many groups is whether they want to select equipment that is capable of working with both 384-well and 96-well plates. For example, it may or may not be possible to use only 96 tips on a pipettor that has 384 channels. Some equipments are designed with this as an explicit feature, so you can select these devices if you believe that you will need this capability.

2.3. Case Study: A Scalable System

The example which follows is based on our experience in gradually building an HTS system from individual workstations to a fully automated system. It should be noted that, although the building was done in four phases, each phase yielded a functional system which was used to perform HTS campaigns. Therefore, this example will show that productive work can be done while the system is being built.

2.3.1. Individual Workstations

We decided to initially buy a standardized system from CyBio AG (Jena, Germany) called "CyBi®-Cellight," which featured a luminescent plate reader (Lumax flash HT), along with a pipetting device and two sets of stackers. A diagram of this system is shown in **Fig.2.1**. We selected this system because it was a preconfigured design, it was simple to use, and it incorporated the type of reader that we were interested in. The heart of this system was a CyBi-Well 384-well pipettor, which had a four-position carriage to move plates to the pipettor. Stackers provided storage of compound, dilution, and assay plates. The mechanism of transporting plates between the various devices consisted of a rotary arm, plus plate lifters at each transfer position.

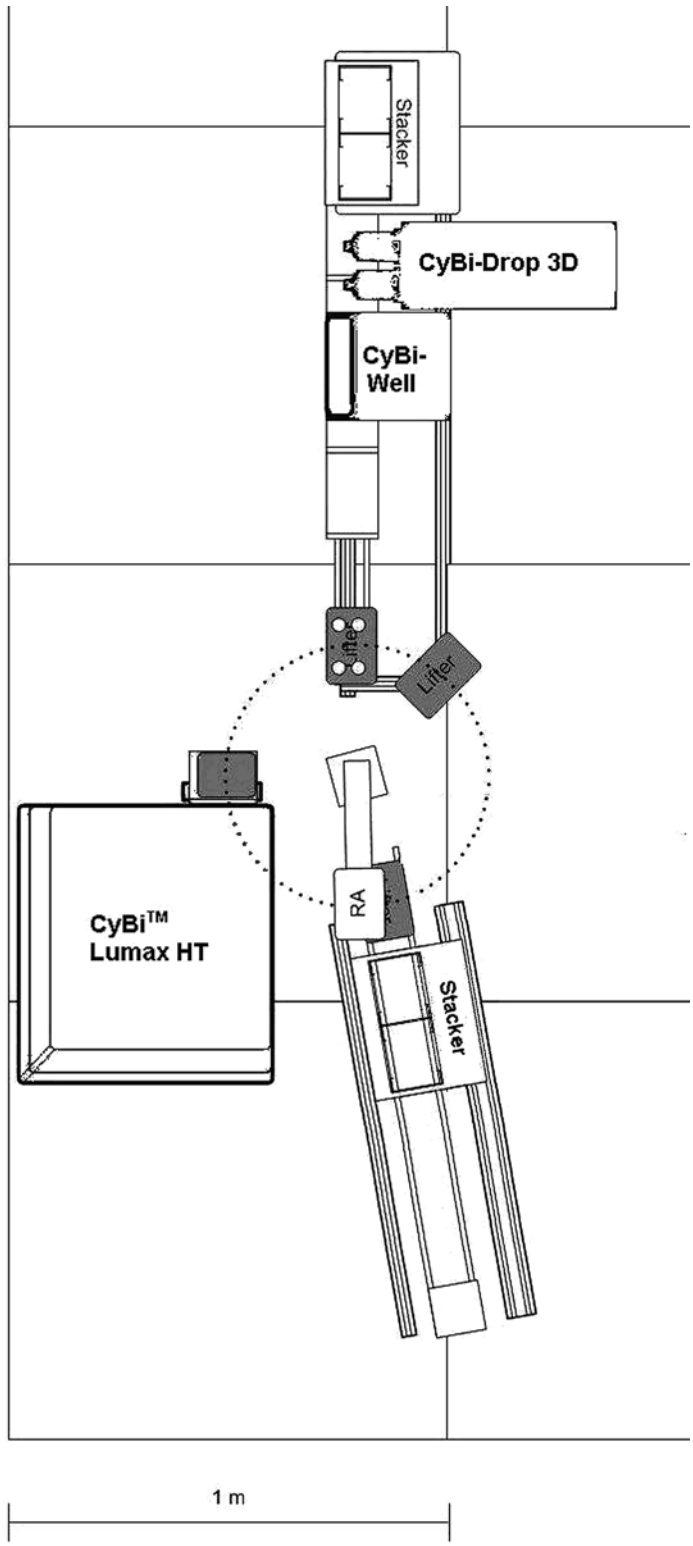

Fig. 2.1. The initial standardized "CyBi-Cellight" system from CyBio AG (Jena, Germany), which features a luminescent plate reader (Lumax flash HT) along with a pipetting device and two sets of stackers.

Since this CyBi-Cellight system had no capability for plate incubation, the system was essentially a workstation. For screening applications, plates were run in small batches to minimize the time that they were sitting at room temperature. While the functionality was somewhat limited, it presented good opportunities for future expansion. The rotary arm was an obvious point through which additional instrumentation could be linked. In addition, each of the stacker carriages was on rails that were open at the end, thus allowing the possibility of expansion in both of those directions. This was further supported by the fact that this vendor had multiple options for ways to expand the system at these open integration points.

2.3.2. Integrating a Third-Party Reader

After some time using this system we were ready to move to the next step. We desired to add an automated incubator and a second plate reader, the Molecular Devices FLIPR^{TETRA} for fluorescence plate imaging. This presented the challenge to move away from a single-vendor solution to a composite system, integrating equipment from multiple vendors. After consultation with CyBio, they added a rail/carriage transport system that took plates from the existing rotary arm device. In this system, shown in **Fig.**2.2, the location of the original components remained essentially the same, and the transport system could move plates between that part of the system and to the incubator and the FLIPR^{TETRA}.

The transfer of the plate from the linear carriage to the FLIPR^{TETRA} presented a small challenge. The FLIPR^{TETRA} has a robotic interaction location, which is a top-loaded plate holder. Thus CyBio added a Z-motion arm with a simple plate gripper, which can grasp a plate that is presented under the gripper on the linear motion carriage. After grasping the plate, the carriage is moved out of the way, and the plate is lowered directly down onto the FLIPR^{TETRA} plate holder.

Along with the new devices, we added the CyBio scheduling software. This software was necessary to manage the relatively complex plate movements of our assays. A typical cell-based fluorescence assay involves bringing a cell plate out of the incubator; adding fluorescent dye using the 384-well pipetting device; placing the plate back into the incubator for 1 hour; doing a small-volume transfer of compound from the compound library plate to a compound dilution plate, again using the same 384-well pipetting device; taking into the FLIPR^{TETRA} the cell plate, the compound dilution plate, and an activator plate filled with the agonist; running the experiment in the FLIPR^{TETRA}; and finally putting all of the plates into output stacks. The scheduling software allows us to manage all of these plate movements, optimize the timing for maximal parallelism, and track the information associated with each cycle.

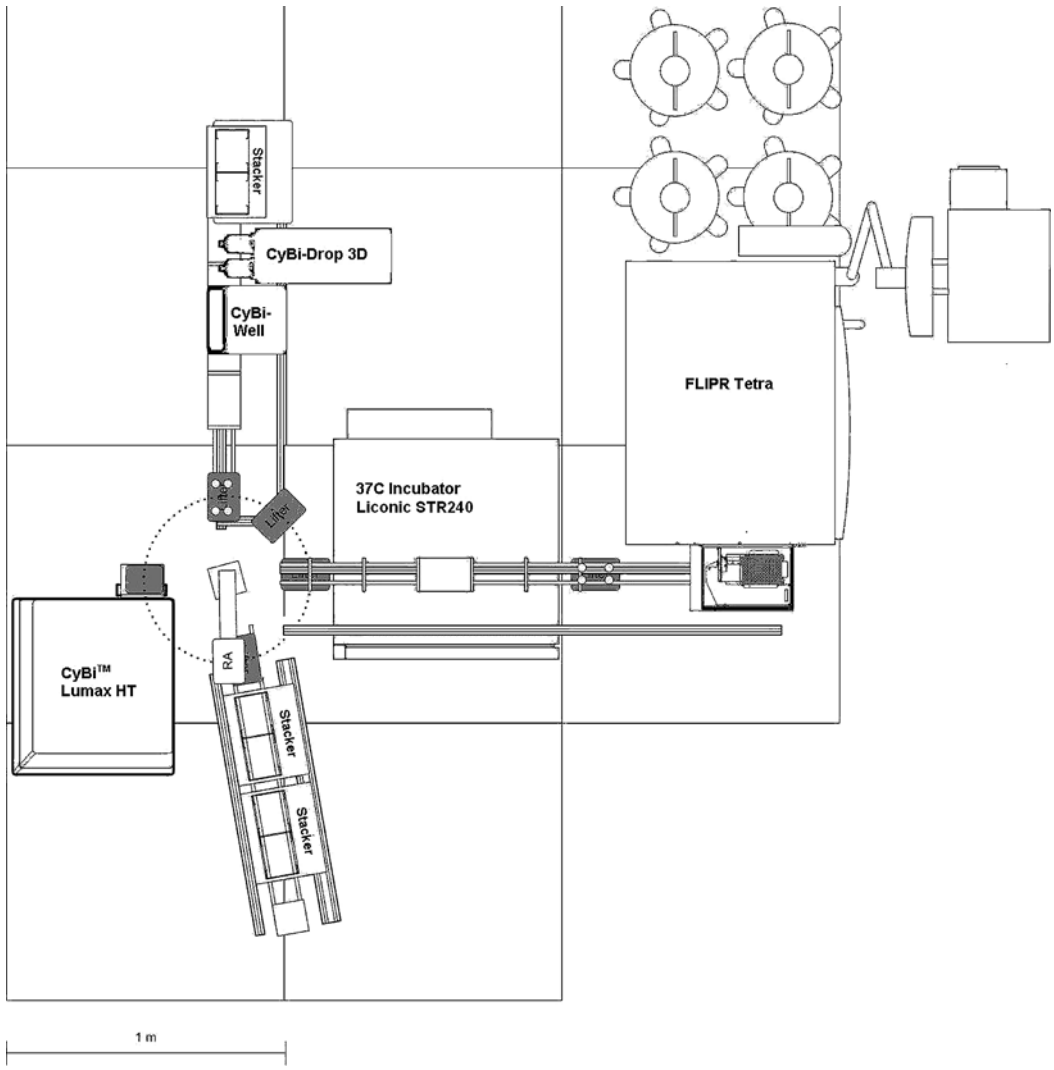

Fig. 2.2. The initial system was extended with a transport system that moved plates to a 37°C incubator and FLIPR^{TETRA} plate reader.

2.3.3. Automation of a Manual Step, Custom Automation

We embarked upon a third phase of expansion after about 1 year. We desired to add into the system a Biotek plate washer in order to speed up the processing of cell plates. We also discovered that some of our assays would benefit from a period of room-temperature incubation. In order to accomplish a room-temperature incubation without blocking the system, there must exist a location to temporarily store plates during the incubation time. Therefore, we also added into the system a plate "hotel," a rack with multiple shelves that could store individual plates. As no "off-the-shelf" solution existed, we needed to work with a vendor to create the desired solution. We went back to CyBio to design these system additions. They proposed the use of a four-axis anthropomorphic robotic arm,

the KiNEDx SCARA robot from Peak Robotics, with a plate gripper. The use of this robot simultaneously resolved the issue of loading the plate washer from above and moving plates in and out of the plate hotel. The new devices were installed above the existing automated incubator, as shown in **Fig.2.3**.

2.3.4. Final, Fully Automated System

For further expansion, we have designed the inclusion of instruments to better manage compound source plates: a cold incubator for compound storage, a plate piercer and plate sealer, and some warming stations to thaw the DMSO compound plates. Since the system is now rather densely packed, we developed a plan to

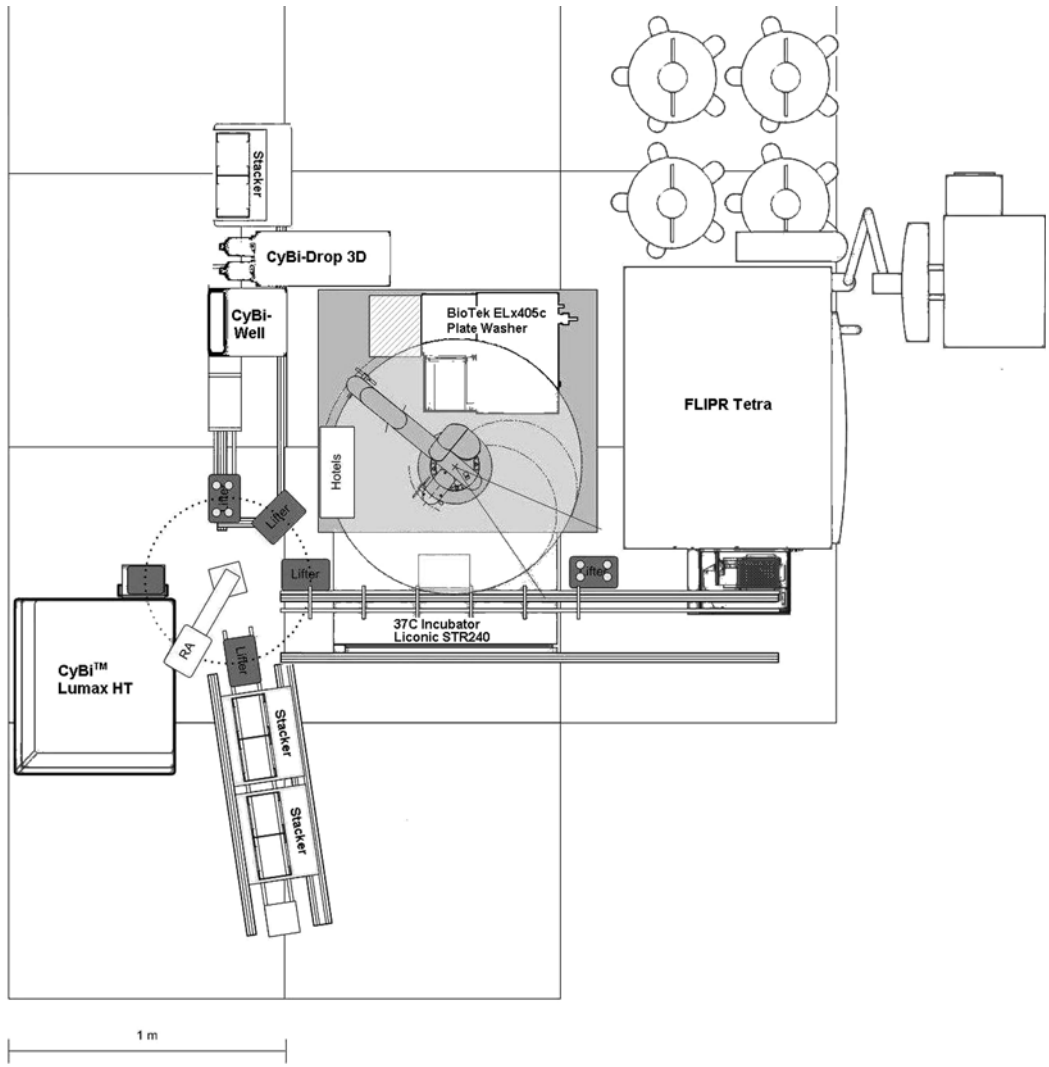

Fig. 2.3. The system was expanded by the inclusion of a SCARA robot that could transport plates to a plate washer and a room-temperature incubation hotel.

include these new instruments at the open end of the pipettor carriage, using another anthropomorphic robotic arm to move plates among the devices. This design is shown in **Fig.2.4**. However, in the end we decided that we would make the investment in a second screening system that incorporated these features.

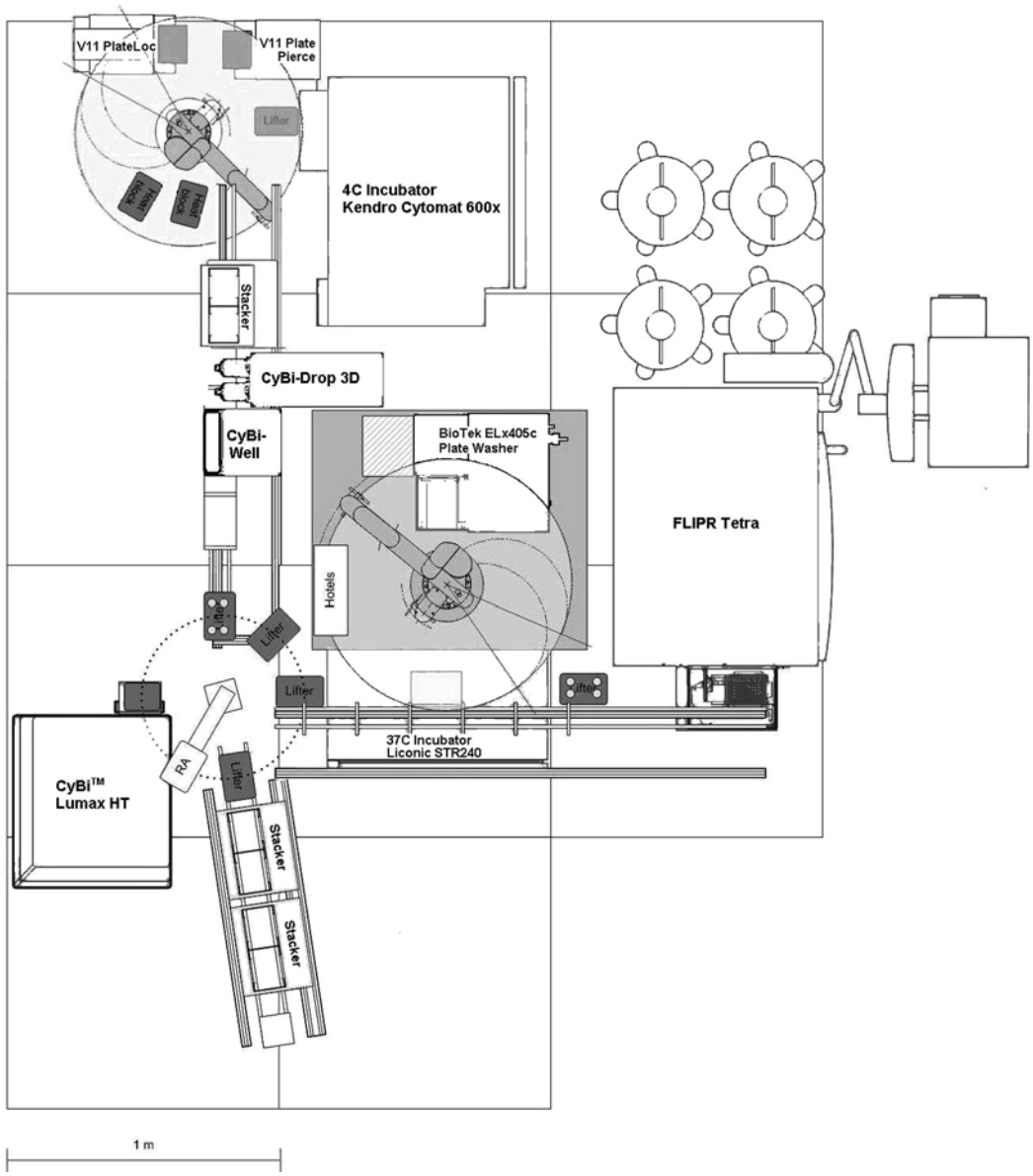

Fig. 2.4. This design included compound plate handling by expanding the system on the left side of the pipetting device. A SCARA robot could move compound plates from a 4°C incubator to heat blocks, a plate piercer to open sealed plates, and a plate sealer to reseal plates after processing.

2.3.5. Comments

We have described the gradual buildup of a fully automated HTS system. While the example is based on our experience with CyBio, similar systems can be built with other vendors. It is important to note that

- Starting with individual workstation eliminates the need for internal customization.
- A good collaboration with the primary vendor is essential. By now, they all understand that they cannot provide the optimum solution for every detector or liquid handler. A vendor who emphasizes compatibility with other equipment should be selected over one who insists on using only their equipment.
- Special attention should be paid to scheduling software since its quality drives the quality of the setup.
- A combination of off-the-shelf and custom solution will almost always be necessary. It is important to minimize the latter as it is more expensive.
- Maintenance should not be underestimated. Again, standard equipment will be cheaper and simpler to maintain than custom solutions.

2.4. Additional Equipment

For an HTS group, there are numerous additional instruments necessary to make a functional pipeline. For example, for follow-up after the primary screening, a typical practice is to retest the putative hit compounds. For this confirmation test, it is essential to have some pipetting device that is capable of cherry picking the desired wells. The minimum requirement of such a device is that it should take plates from a stack, read barcodes, and work from a worklist that describes which wells of each plate need to be picked. Additional desirable features include the ability to open and reseal the source plates and possibly to perform dilutions if that is desired for the confirmation experiments. In our case, we chose to use a Hamilton Starlet, with an eight-channel pipetting system and some features that allow it to use plates in stacks. This system is essentially a stand-alone workstation.

It is possible to automate some aspects of compound management. At the minimum, there should be investment in good-quality freezers for storage of compound plates and plate sealing and piercing devices. For more extensive compound management functionality, there are several good systems available for fully automated sample processing and storage. For a small screening group, such systems are typically not necessary, but they may become attractive after the growth of the internal screening library. Vendors of such systems include REMP, The Automation Partnership, Hamilton Storage Technologies (TekCel), MatriCal, and TTP LabTech.

3. Data Handling

3.1. Requirements

Well-managed data handling is a critical, and maybe sometimes underappreciated, issue for a successful HTS group. There are basically two things that must be considered: first, how to manage the data, store them, and get the data into whatever program will be used for analysis and second, how to analyze the data.

In most cases, there must be a central database to act as a repository for the data coming from the screening system and to hold information regarding plate contents (barcodes, compound identifiers, etc.). With only a small number of compounds, it is possible to work without a database. One can import data directly into the analysis software and work with the entire data set in one analysis session. What the limit is for that sort of approach depends upon your analysis software. We have analyzed the primary screening results for 100,000 compounds within a single Spotfire session. Using pipeline processing tools such as PipelinePilot, even larger data sets can be processed. However, for most practical purposes, a centralized database system is appropriate when working with more than a few hundred plates worth of data.

Another requirement is a flexible data analysis engine. While the software included with most reader instruments includes some mechanism to compute the results, this is usually not sufficient for sophisticated screening assays. Some framework must exist that allows the specification of formulas and rules for the computation of the results, the detection of outliers, the determination of average responses, and the determination of hits. This framework should allow easy configuration with various processing options.

Finally, there is great benefit to be obtained from tools that allow visualization of your data, in many different formats, ranging from the most detailed view of the signal from individual wells up to the most gross view of the performance of one screening campaign versus another. Such tools are useful to understand trends in the data that may be difficult to understand by simply looking at numbers. For example, you will want to be able to see things such as the response across the plate to see if there is a gradient due to temperature or cell growth; the Z' factor throughout the day, or from one day to another, to look for temporal effects; the average response from all wells across many plates to look for instrument problems related to individual wells, rows, or columns. The more visualization options that are present in the analysis software, the better.

3.2. Build or Buy?

There are many commercial products available to help manage and analyze the data. Some of these include Genedata Screener®, IDBS ActivityBase XE, SciTegic Pipeline Pilot, MDL® Assay Explorer, and Spotfire DecisionSite® for Lead Discovery.

But there is also the possibility to build in-house a great deal of the data handling system if you have sufficient time and talent available. You have to make a decision early on whether you want to buy or build the software. In general, we have gravitated toward the do-it-yourself approach. The reasons for this preference include the following:

- If you gradually grow from a small operation to become a large HTS group, then there is time to organically grow the data handling mechanisms. It can be almost a natural growth, in which you build new parts as you have new needs.
- The cost of some commercial systems can be quite high.
- In general, someone (you or the vendor or both) *must* do customization of commercial systems, and often the customizations take as much time and effort to manage as it would have been to just build it yourself. To make the commercial software truly valuable, it must provide a very solid framework, which would be prohibitively expensive and complicated for you to create from scratch (for example, complex visualization programs).
- To run an effective screening group, you need to have people on staff who can write computer code, even if only to customize a commercial product. With some good guidance and sufficient time, in general those people also have enough skills to build an in-house designed system.
- If you build it yourself, you have complete control over how it works. Of course, you also have the complete responsibility for doing it correctly and validating the results.

There are certainly disadvantages to building from scratch. There is a greater up-front time investment, and there may be certain sophisticated analysis tools that are too difficult to build. The algorithms for most basic HTS data analysis are not very complicated, really just simple math. But tools that allow visualization and pattern detection may be beyond the scope of what you want to build. If you decide to build some part of your data system on your own, inevitably you will make some mistakes in your implementation. You should accept the fact that you will probably have to re-engineer some parts of your system after working with it for a while and possibly even completely rebuild some parts.

3.3. Database

The database system can range in complexity. At the most simple, it should be able to hold plate identifiers, compound information for each well, and the biological result for each well. At the most complex, the database will also contain the molecular structure of each compound, and additional information such as performance in other screenings and results of secondary tests.

Early on, there must be a decision on several points:

- Do you want to store chemical information in the database or simply an identifier of the compound? Because chemically aware database are typically rather specialized, the most common approach is to have the screening database hold only screening data and simply contain a reference to the chemical compound; the chemistry can be managed in a separate cheminformatics database system.
- Should there be cross-screening knowledge or is each project distinct and unrelated to the others? For a service provider, it may be quite acceptable to have no relation between the compounds in one screening and another. However, if you will be running the same chemical library on several different targets, there can be some advantages in performing meta-analysis across several screening campaigns to allow identification of things such as frequent hitters and toxic compounds.
- Should the database contain only screening data or should it also incorporate other related data such as dose–response data or the results of secondary testing? Incorporating this type of data will typically be done in a fashion that is quite distinct from the plate-based screening data.

Regardless of whether you make the decision to build or buy a database system, you will have to decide how to integrate the database with other existing data within the organization. If you buy a commercial system, it is important that the system be open, properly documented, and employ standard technologies to allow easy federation to other informatics systems.

3.4. Analysis

There are several levels of analysis that must be considered. First, there is the conversion of the raw data collected by the reader into some meaningful number. For example, a biochemical endpoint assay may simply require a single absorbance reading; a biochemical kinetic assay may need the computation of the slope of the absorbance curve over time; and for a cell-based kinetic response, the appropriate value may be the Max–Min or the area under the curve. In most cases, the instrument software is capable of producing some "reduced" value that is appropriate for use. In some cases, however, there can be some advantage to capturing all of the raw data in the data system and performing the data reduction during the analysis process.

The second level of analysis is the conversion of the well response into a biologically meaningful value. For example, if you are performing an antagonist screening, you may choose to compute the percent inhibition, based upon the average response of control wells. From this biological response, one can determine putative hit compounds based upon some criteria, such as a simple threshold.

Important parameters that should be computed include the plate quality metrics such as Z' and Z factors, the number of outlier control wells, and the number of putative hits per plate. Other analysis capabilities that are good to have include the ability to easily detect systematic effects, such as bad wells, rows, or columns, which can indicate instrumentation problems; repetitive patterns, such as every third plate having a lower response, which can indicate a problem in the overall processing of plates; and gradients across the plate, which can relate to temperature effects, cell seeding problems, etc.

3.5. Data Handling Case Study

In the case of Axxam, we decided to build almost all of the data management and analysis system in-house. The only commercial product that we rely upon is Spotfire, with some customizations as noted below.

The data analysis system that we designed is constructed around a central database, which is implemented in Microsoft SQL Server. The components include the database schema itself, importer code to read the instrument data files, stored procedures to perform the analysis, and some reporting tools.

The database schema consists of about 30 tables. We decided early on to store in the database all of the raw data from the instrument. Since most of our experiments are kinetic readings (occurring over the course of 30 seconds to 10 minutes), this means that we store the luminescence or the fluorescence value of every time point for every well. Therefore, this table alone will have several millions of rows of data related to a single screening. Obviously proper management of indexes and table partitions is important for acceptable performance.

There is the possibility of storing only reduced data, such as a single number that represents the kinetic response of each well (such as Max–Min or some similar reduction of the kinetic response). However, storing all of the data is beneficial in case some change needs to be made in the computation of the response and also in the examination of the raw kinetic responses. When trying to understand a problem, it is sometimes advantageous to be able to see all of the data from a certain range of plates; this is made much easier by being able to retrieve the data from the database as opposed to retrieving it from many original instrument files.

The first step in analyzing screening data involves importing the data from the file produced by the instrument into the database. For this we have created specific importer code, written in a scripting language. Since the output of each instrument is unique, there must be some specific code written for each file type. These importers are then further automated by an additional application that scans folders on the instrument computers to effect the import of new data files. In this way, data that are generated can be imported into the database immediately after the plate is run and the analysis triggered.

For the analysis, we implemented a type of configurable execution. The analysis is broken into many steps, each of which is accomplished by running a stored procedure. The particular procedures to run and the order of execution are dictated by entries on a database table. Thus, different steps can be configured for different projects. When defining a new project, we typically copy the analysis steps from a similar project processed in the past and make any changes necessary. Furthermore, we have another database table containing specific values defined for each project, such as the Hit Threshold, acceptable Z' value, and acceptable number of bad wells. The database also holds the plate layout for each project, which defines the location of control and test wells.

After importing the data, the primary analysis is automatically triggered. Primary analysis includes computation of the kinetic response (for example, Max–Min and area under the curve); determination of outliers for the controls; computation of the average and standard deviation of the kinetic response of the controls; computation of the desired biological response value for each test well (such as percent activation or inhibition) with respect to the controls; and computation of plate quality parameters, including Z' and Z factors, number of hits, and number of outlier control wells. A flag is set on the test wells that meet the Hit criterion for the project, and a flag is set at the level of the plate to indicate if it met the established quality criteria.

For visualization of the information, we use Spotfire DecisionSite. This software is a general-purpose data visualization and analysis tool. The software is very open to developers, and we have performed a number of customizations. We have defined numerous information links, which allow the retrieval of specific data from the screening database, such as plate quality statistics or the biological response from all wells. Examination of such statistics allows us to detect problems with the screening and to fine-tune our hit thresholds.

We also use Spotfire for some data analysis that is not amenable to the automated processing of the database. This includes processing of assay optimization experiments, small screenings and compound profiling that do not conform to the standard organization of samples, and dose–response experiments. To assist in these analyses, we have created numerous custom tools that perform specialized functions and guides that take the user step by step through the analysis process. A great advantage of using Spotfire to accomplish the data analysis is that each step of the process can be checked visually, and unexpected problems can be detected. Sometimes the best way to process the data is not clear, so some experimentation in Spotfire with different approaches can help us to configure the eventual processing rules that are defined in the database for a particular screening project. Spotfire is also very valuable to confirm the results produced by database processing;

obtaining the same results from both systems for a few plates is a good assurance that the automated database processing is configured correctly.

4. Operational Issues

4.1. Personnel

The selection of the proper personnel is truly critical for a successful screening group. Obviously there must be qualified biological scientists who can run the screening campaigns, interpret the biological responses being observed, and understand when there are problems. It may be possible to have less sophisticated technicians running the system. But having talented, knowledgeable people in charge of the execution of the screening, who are in daily contact with the process, is critical to achieving good quality results. Depending upon your organization and the type of screening assays being performed, the screening group may also include people responsible for cell culture, enzyme preparation, compound management, and management of consumables.

There are several other technical positions that are very important for a successful screening group. These include people involved in automation, instrument programming, data system development, and data analysis.

Regarding the automation personnel, there are at least two distinct roles that are important. First, someone must be an expert in the system software. This means understanding how to create automated protocols, optimize schedules for proper parallel execution and maximal throughput, interact with the database, and operate the system properly. In most circumstances, the actual users of the screening system will not be experts in developing new screening protocols, but rather will simply run the protocols that are perfected by the system expert.

Another important automation role is more of an automation development position. Depending upon the complexity of your system and the frequency with which you need new functionality, this may or may not be a full-time position. This person should be capable of debugging system errors, lower-level programming to accomplish things such as reconfiguring system communications, adding new error handling, adapting the system to new plate handling concepts, and possibly integrating new instruments. The competencies of such a person would include database programming, script programming, some knowledge of mechanics and electronics, and a great deal of generalized computer knowledge. In some cases, it is not possible or convenient to have a person on your staff with such specialized skills. For many of these functions, it is possible to rely upon the system vendor to perform

such customizations or automation consultants. However, when using a vendor in such a way, it is still critical that there be someone within the screening group with a relatively high level of technical knowledge who can interact with the vendor. The technical specialists working for the vendor absolutely need a counterpart within your organization with whom they can interact and speak the same technical "language."

On the data analysis side, there are also several important roles. Regardless of whether you are using a commercial analysis software or an in-house-created data system, there must be someone within the screening group who understands the software in detail. Obviously if the system is created in-house, there should be an expert within the organization (maybe not within the screening group) who understands all details of the data system and who can make corrections and modifications. If the system is from a commercial source, then the experts will be within that external provider. In either case, there should be a person within the screening group who understands both the software and the biology in sufficient detail to be able to communicate problems and desired improvements.

Finally, there must be one or more persons responsible for data analysis. It would be optimal if the people in charge of the execution of the screening campaigns were also responsible for the data analysis, because they are most intimately aware of the special issues for each screening. However, due to the complexities of analysis software, it is often more practical to have some expert users dedicated to overseeing the data analysis.

4.2. Making It Work Reliably

The design and establishment of a screening instrumentation platform is a substantial investment of time and effort, but it is really only the first part of creating a functional screening system. The process of improving the screening system in terms of throughput, functionality, and reliability is a never-ending process. The company's management needs to understand that the screening team will devote substantial time to small but important improvements, especially during the first year of operation. Furthermore it must be anticipated that there will be bugs and equipment failures that can result in weeks of downtime.

Regarding instrumentation problems, it is typically advisable to maintain service contracts with the equipment vendors to ensure a rapid response. Outright failure of instruments is simple enough to deal with: if the robot stops moving, or the reader does not read, the vendor will normally be able to understand and fix the problem quickly. However, the problems that cause the greatest loss of system efficiency are those that are not so simple to understand, especially problems that are not reproducible. Typical problems of this type include occasional problems in loading a plate from a stacker or an incubator, random tip loading problems, dropping plates or misplacement of instrument locations, and

random strange software problems. These types of tiny problems can interrupt a screening run and possibly cause the loss of an entire days' work. In such cases, simply calling the vendor may not result in a quick solution. It is imperative that someone watch the system carefully, make some hypothesis regarding the cause of the problem, and test various possibilities to try to reliably reproduce the problem. Only then can the vendor solve the problem easily. Of course this type of debugging takes a great amount of time. While it is possible to convince a vendor that they must send a service technician for several days until they can find the problem, this is typically not the best solution. It is much better to have someone on staff who can invest the necessary time to debug the problem.

5. Conclusion

Building a high-throughput screening facility within an organization that has no HTS experience can be a daunting proposition. Obviously it is not as simple as just purchasing a screening system and some analysis software. Regardless of whether you decide to rely entirely upon external vendors and commercial software or you decide to build much of the functionality yourselves, there is a steep learning curve and many pieces that must be put in place in order to make a functional screening facility.

The most important asset that must be acquired is knowledge – knowledge of how to design a screening system, how to accomplish the required data handling and analysis, and how to run the facility. This knowledge can be acquired by bringing in experienced people or by slowly building up the expertise from within the organization. Instrumentation and software vendors can offer a lot of possibilities, but someone within the screening team must have a clear idea of what is required.

We have presented one view of how to build a screening capacity in a step-by-step approach, growing in instrumentation and data handling abilities over the course of several years. While it is not easy, the creation of a well-functioning screening facility can add great value to a research organization.