ELSEVIER

Defining and maintaining a high quality screening collection: the GSK experience

Stephen J. Lane[1], Drake S. Eggleston[2], Keith A. Brinded[3], John C. Hollerton[3], Nicholas L. Taylor[3] and Simon A. Readshaw[4]

[1]CASS, DR, GlaxoSmithKline, Gunnels Wood Road, Stevenage, SG1 2NY, UK
[2]CASS, DR, GlaxoSmithKline, Five Moore Drive, PO Box 13398, RTP, NC 27709, USA
[3]CASS, DR, GlaxoSmithKline, Gunnels Wood Rd, Stevenage, SG1 2NY, UK
[4]CASS, DR, GlaxoSmithKline, NFSP, Harlow, CM19 5AW, UK

Understanding the quality of a screening collection is the first step to improving it and, as a result, the quality of the screening process. This article outlines how this issue was approached at GlaxoSmithKline and some of the hurdles that needed to be overcome to achieve success. The article focuses specifically on the necessary software and hardware infrastructure needed, and at some of the extra benefits of such a project in terms of data mining and data modelling.

Introduction and rationale

The quality of the compound collection of a pharmaceutical company has a major influence on the success of biological screening in drug discovery programmes. It is acknowledged today that the success of HTS is dependent on three main aspects: the quality of the biology of a screen; the size and diversity of a screening collection; and the chemical and physical quality of compounds in a screening collection. The last aspect has not always been considered as important and was sacrificed to a degree to enable the new 'compound hungry' paradigm of HTS to be fed effectively. The combinatorial synthesis of large libraries had an accepted focus on quantity of compounds at the expense of quality, arguing that a hit would shine through and would be readily deconvoluted. Drug discovery has evolved in a way that has conserved many of the throughput advantages of HTS and automated synthesis. It has pulled back from the vision that size is everything to a more sensible position that size is important but so is quality.

The quality of collections, particularly those that have been assembled over time, can be impacted by compounds that have degraded because of their storage conditions. This is often a consequence of lack of knowledge or consideration of compound stability at the time of inclusion. Also, the information technology infrastructure needed to maintain the integrity of the collections has often failed to keep pace with the changing requirements of the collection. The combination of these factors has given rise to screening collections containing 'bad' compounds.

Subsequently, it has been calculated that screening 'bad' samples (where 'bad' samples can be considered as impure, duplicated or structurally incorrect) results in considerable waste in screening costs alone. In addition, it is not known how much cost is associated with unsuccessfully pursuing false hits. At the time of the merger of SmithKline Beecham (SB) and GlaxoWellcome (GW) to form GlaxoSmithKline (GSK), it was recognized that both heritage companies had 'bad' compounds in their collections but it was not known how many, and whether the existing knowledge merely represented the tip of an iceberg. Undoubtedly all pharmaceutical companies that maintain collections for screening will encounter similar circumstances. GSK quickly realized that a key part of its top level strategy (to increase the number of quality 'hits' and leads and reduce optimization cycle times) was to have the highest quality screening collection.

Following the merger, the amalgamated GW and SB screening collection exceeded 1.4 million compounds, all of which could potentially be included in the new GSK ultra-HTS (uHTS) collection. The combined collection was also required to be formatted for loading into the new automated liquid stores in Tres Cantos, Spain, and Upper Providence, PA, USA. Because of these two events, every sample in the collection had to be physically handled, which provided a unique opportunity to assess fully their integrity and quality before incorporation into the screening collection.

Corresponding author: Lane, S.J. (stephen.j.lane@gsk.com).

1359-6446/06/$ – see front matter ©2006 Elsevier Ltd. All rights reserved. PII: S1359-6446(05)03730-X

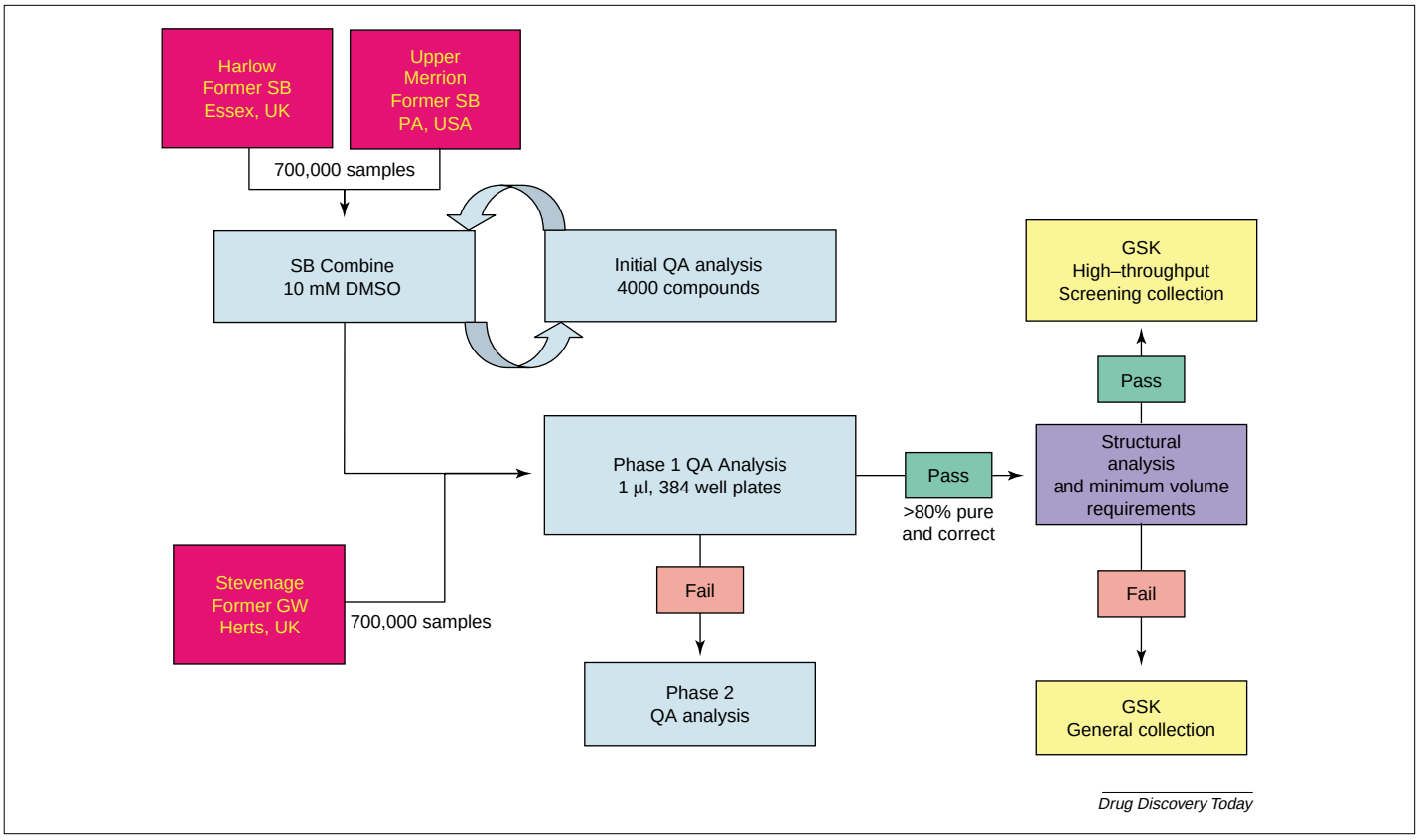

FIGURE 1

Phase 1 of the quality assurance process.

To ensure the highest quality screening collection for the future of GSK, we took the opportunity to analyze all samples for purity and identity, thereby setting the baseline standard for the creation of the new GSK screening collection. Secondary objectives were:

- To set a quality control (QC) standard for all future compounds entering the screening collection.
- To develop a chromatographic system-suitability test to ensure compliance to a minimum standard for all instrumentation used to obtain analytical data.
- To produce a massive and unique high-quality dataset of UV and mass spectrometry (MS) spectra, chromatograms, extracted retention times and NMR spectra, acquired on a single validated method (for data mining projects).
- To provide the core dataset for a global analytical data repository (GADR).

This initial triage of the heritage collections had to be achieved in 18 months – a mammoth task!

Logistical challenge

This project was without precedent at GSK, or even in the pharmaceutical industry, and involved the combined and closely aligned efforts of the Compound Management, Analytical Sciences and Cheminformatics departments at a scale and complexity never undertaken before. Within the size constraints of this article, the analytical challenges of the project will be the focus of the discussion.

Following development, testing and assembly of the source plates, the compound quality assurance (QA) process for the HTS

BOX 1

LC–MS method

The following chromatographic method was used for Phase 1 and Phase 2 of the project:

Column: 2.1 × 30 mm 3.5 u Xterra.MS C18
Flow rate: 1 ml/min
Stop time: 4.5 min
Cycle time: 5.4 min
Oven temperature: 25°C
Solvent A: 0.1% formic acid in water
Solvent B: 0.1% formic acid in acetonitrile
Gradient timetable

Time	A%	B%	Flow (ml/min)
0.00	100.0	0.0	1.000
0.10	100.0	0.0	1.000
3.10	0.0	100.0	1.000
3.80	0.0	100.0	1.000
3.90	100.0	0.0	1.000
4.00	100.0	0.0	2.000
4.40	100.0	0.0	2.000
4.50	100.0	0.0	1.000

Injection volume: 1 µl
MS scan duration: 0.5 s
MS inter-scan delay: 0.2 s
Ionization mode: ES+ / ES– alternate scans
Mass range: 100–1000 amu
DAD wavelength range: 210–350 nm

collection involved three main elements (Phase 1, 2 and 3). This followed a pilot study on the compound combination process for former SB compounds, the reformatting of samples from a 96-well into 384-well plate format (for former GW compounds only) and the combination of samples from tubes held in USA and UK collections (former SB only).

- Phase 1 (Figure 1) involves using liquid chromatography (LC), diode array (DAD), MS and evaporative light scattering (ELSD) as a cost-effective way to rapidly screen 1.4 million compounds for purity, molecular weight and molecular ion-isotope pattern confirmation.
- Phase 2 involves using NMR and LC–MS to analyze fresh solution of compounds that do not pass the Phase 1 criteria (Box 1).
- Phase 3 involves the purification of high-value impure compounds and reformatting of the selected former GW and SB compounds for the new screening collection.

Analyzing 1.4 million compounds in 18 months to give a highly reliable estimation of compound quality required a vigorous strategy. The analytical process needed to be industrialized while ensuring that the data generated was suitable for high-throughput automated estimation of purity and surety – NMR is impractical at this rate. It is just possible to acquire the data, but reliable automated processing is not available and it is impossible to interpret this volume of NMR data without a small army of trained spectroscopists. However, LC–MS, although daunting, was considered possible with modern robust systems. Commercially available automated processing was unsophisticated and by no means infallible but provided a core to build on. Inherent ambiguity of MS response was allowed for by including a second phase (Phase 2), where certain categories of Phase 1 'failure' would be reassessed using LC–MS and NMR approaches. It was very tempting to use new technology, such as MUX systems [1], for this type of project, after all, high-throughput analysis is what such systems were designed for. However, a pilot study convinced us that the compromise in data quality and uncertain robustness (at that time) of such systems weighed too heavily in the risk assessment of the project.

Strategy

The key requirements for Phase 1 was that it enabled 50,000 compounds to be analyzed, processed and validated each and every month.

The method of choice was fast gradient high-performance liquid chromatography (HPLC) with DAD and ELSD as well as switched positive- and negative-ion electrospray mass spectrometry (ES+/ES–) detection on a modern quadrupole mass spectrometer. The chosen HPLC autosampler could accommodate two 384-well plates, which satisfied the extreme requirements of running 24 h a day, 7 days a week (i.e. with a cycle time of 5 min per sample, a 384-well plate took 32 h to analyze). The chosen generic method, on a modern C18 stationary phase, gave good chromatographic resolution in a reasonable analysis time. All samples presented for assay were dissolved in dimethyl sulfoxide (DMSO). The addition of ELSD enabled purity estimation for poorly retained components that were masked by the UV response given by the DMSO solvent peak in DAD detection. The ELSD also allowed us to detect compounds that exhibited little or no UV chromophore. Summed diode array UV was used for the primary purity estimation with >80% pure

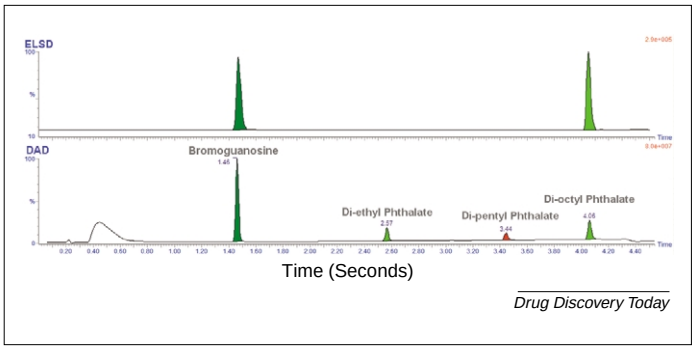

FIGURE 2

System suitability test mix with typical responses on summed diode array and evaporative light scattering detection.

being the criteria for passing Phase 1. A system-suitability test mix (Figure 2) was developed to ensure that all data were acquired from appropriate instruments under suitable conditions for good quality chromatography and mass spectrometry.

The system suitability test mix was run at the beginning, after every 80 samples and at the end of each 384-well plate to confirm that all acquired data conformed to the guidelines set out for data integrity and system suitability. The test mix was used to check the following points:

- Correct plate orientation;
- Correct operation of the autosampler;
- Gradient equilibration;
- Chromatographic resolution and integrity;
- Column integrity;
- Correct solvent delivery;
- DAD and ELSD response, MS response (+ve and –ve), MS resolution and MS calibration.

Infrastructure

To accomplish the required throughput of 50,000 compounds per month every month, ten systems were installed in a purpose-designed laboratory with a spare system for process redundancy. In addition, the following infrastructure components were required:

- Uninterrupted power supply;
- Stringent method development (i.e. good chromatography and 'catch all' detection);
- System QC and robustness testing;
- Proactive column replacement;
- A planned service regime – in partnership with vendors – with readily available spares and in-house engineers;
- Remote system access and web system monitoring;
- Weekend work;
- Training and expert backup;
- A team of trained technicians.

The use of LC, DAD, ELSD and MS as the first filter ensured that the acquisition capability was sufficiently high to enable industry-standard purity estimation. However, the compound dependency of MS, ELSD and, to a degree, UV (chromophore) response necessitated a higher level of categorization of pass–fail than commercial browsers offer, as well as a decision tree to direct 'fails' from Phase 1 through a route that took into account the variability of response. This was achieved using in-house software development.

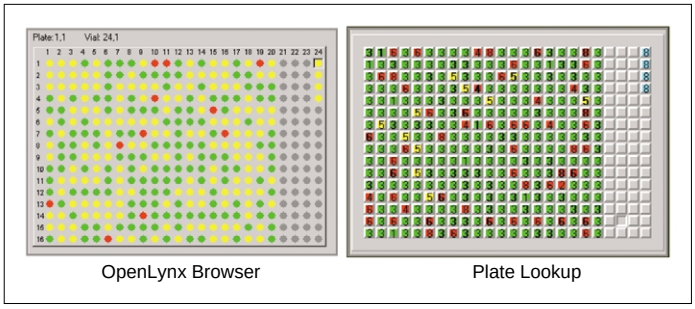

FIGURE 3

Comparison of commercial browser view against in-house PlateLookup. The view has added numerical categorization of pass/fail/tentative categories.

Software development

QA Software (PlateLookup) was written in-house to reprocess the typical green, red and yellow results for 'pass', 'fail' and 'tentative' from the Micromass OpenLynx Browser and convert them to a numerical categorisation of 1–10 using PlateLookup (Figure 3). The categories were:

1. Pure, right, polar;
2. Pure, wrong, polar;
3. Pure, right;
4. Pure, wrong;
5. Present by MS;
6. Impure, right;
7. Pure, right, isomers;
8. Impure, wrong;
9. Did not ionise;
10. No peaks.

Other refinements within the software included calculated molecular weight, number of isomers possible and the ability to view the four source 96-well plates (used to make up the 384-well plate) individually (Figure 4). This capability enabled patterns (e.g. those arising from transcription errors) that look randomly distributed in the 384-well plate to be more easily recognized.

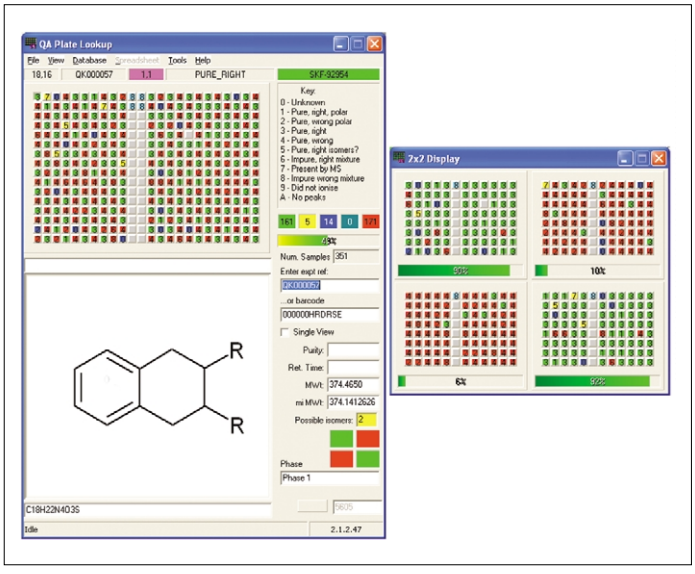

FIGURE 4

PlateLookup view showing concentration of failures in two of the 96-well plates on deconvolution of 384-well plate into 4 96-well source plates.

Phase 1 processing identified 'pass' categories (1, 3, 5 and 7) for which there was high confidence that the compound was >80% pure, with the structure supported by molecular weight confirmation. These compounds were directed to dispense into the new GSK screening collection.

Failure categories from Phase 1 correlate to the injection of 1 μl of a solution of compound at an approximate concentration of 10 mM DMSO not passing the criteria of purity and surety. This could be due to the compound being impure or wrong (e.g. incorrect molecular weight) but could also result from it being insoluble, low concentration or failing to ionise. Failure categories from Phase I (2, 4, 6, 8, 9 and 10) were therefore passed into Phase 2 where further analysis (LC, DAD, ELSD, MS and NMR) was carried out on fresh solutions of solid samples, where available. Phase 1 fails that were found to pass in Phase 2 were then subjected to a DMSO stability study at room temperature under controlled humidity. Compounds that remained stable were then directed to dispense. This was based on the rationale that different results for Phase 1 and Phase 2 might indicate a DMSO stability issue. The overall fate of samples was determined by the Phase 1 and Phase 2 results.

Results and lessons

Overall, 61% of heritage compounds were found to exceed the Phase 1 criteria of acceptance for the new GSK collection and have formed the high-quality base line. This pass rate was found to be consistent for both heritage collections, although their compositions and storage conditions (among other factors) were independent and quite different. This strongly supports the conclusion that these findings are almost certainly typical for any large collection assembled over the same period of time. Subsequently, more-detailed Phase 2 analysis rescued a further 50,000 compounds – some of which were due to incorrect categorisation of the Phase 1 data, others were due to historical sample handling errors. These data have provided an 'analysis time zero' for heritage solutions.

Data mining

The large volume of quality data allows us to be more confident in our conclusions about stability and storage. For example, when we look at the instances of category 6 failures (impure, right, mixture) for a large dataset, we see a larger population of failures in well positions around the plate edge. This, coupled with other data on water uptake at the plate edge [2], clearly demonstrates that under our heritage storage conditions in plates, decomposition related to water uptake is more likely at the edges.

These observations provide clear direction for storage strategies and, following the QA project, we have a clear idea of the initial quality of the new collection when assembled.

The acquisition of such a large number of high quality LC, DAD, ELSD, MS and NMR spectra against known structures during Phase 1 and Phase 2 has provided the opportunity to rapidly build a GADR (Figure 5) that is available to chemists and biologists at the desktop.

GADR facilitates the searching of structure, substructure or unique identifier and allows display and comparison of the associated high-quality analytical and spectroscopic data. The project provided an initial massive dataset with which to design and build the collection QA database. This is now the repository for all data

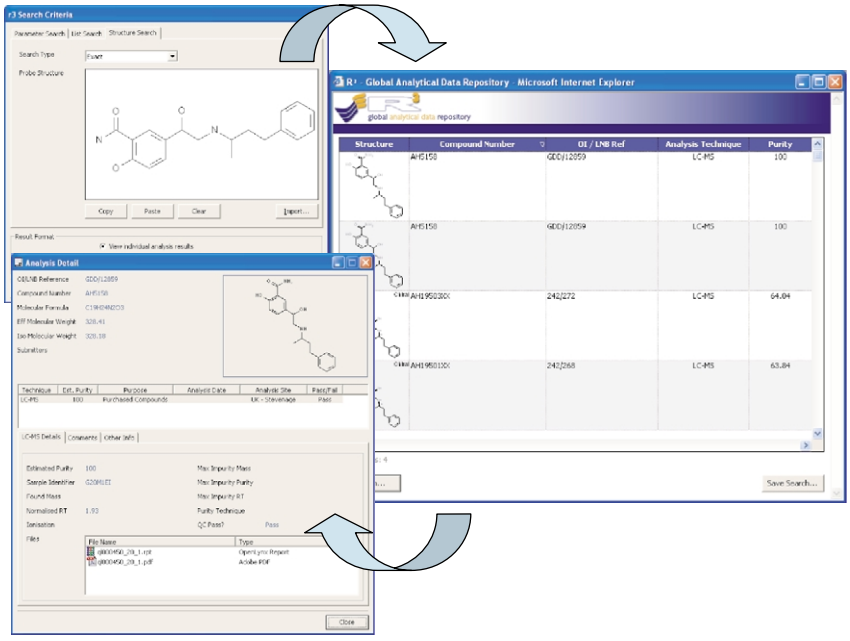

FIGURE 5

Global analytical data repository.

acquired on systems that have passed system suitability requirements during discovery.

Retention-time data collected on all compounds analyzed in the project shows the method to be well chosen, with a Gaussian distribution across the gradient. When overlaid, the system suitability mix is shown to again be well chosen to cover the retention time window (Figure 6). A small percentage of compounds were very poorly retained on the system, and for correct structures their purity was determined by a purpose-designed polar method.

This massive number of accurate retention-time measurements under a single generic gradient has enabled us to build a predictive *in silico* model of the retention time on this system that gives a mean error of 14 seconds for the predicted retention time of a given structure. For example, the predicted retention time of propranolol is 1.85 min and the observed retention time is 1.87 min (an error of 1.2 s).

Conclusions

The project was completed ahead of schedule and within budget, thus allowing the new high quality GSK screening collection to be used at our screening sites ahead of schedule. Already we see much improved retest statistics, predictable SAR and greater confidence in the quality of our samples. Subsequently the new GSK screening collection has been added to by synthesized and purchased compounds that have all been quality assured to the same level, thus defining the quality of new compounds entering the collection. This provides us with a highly valuable time-zero analytical measurement when our compounds are first dissolved in DMSO and made available for screening. This, in turn, allows an opportunity for stability-monitoring using subsequently acquired analytical data on the solutions.

We have been able to rapidly build the physical collection, as well as GADR, which now has in excess of 2 million high-quality analytical datasets (including LC, DAD, ELSD, MS and NMR spectra) of known structures that are available for viewing and comparison at the desktop. Searching on structure, substructure and unique identifier is proving invaluable in understanding screening result differences, as well as helping in structural elucidation of unknowns.

This was a project on an unprecedented scale and complexity that owes its success to careful planning and the initiation of an enthusiastic, multidisciplinary project team. Guiding the project at the outset was a focused steering committee to set the strategy, anticipate issues, perform risk assessments and take the critical decisions in a timely and informed manner. The steering committee consisted of representatives from all parts of the process,

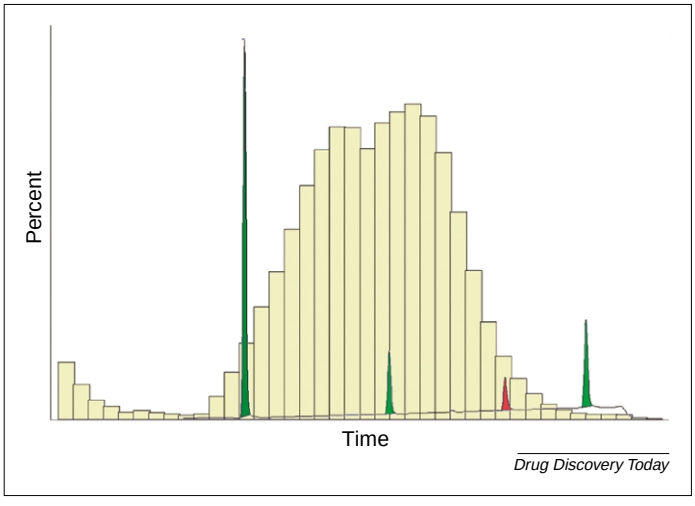

FIGURE 6

Retention time distribution of >700,000 compounds chromatographed by a single generic method with the system suitability test mix overlaid.

including compound management, facilities management, analytical sciences and cheminformatics. This ensured that no part of the process was considered in isolation and proposed changes to the process were visible to all participants.

Lessons from the project about the immense value in standardization and inherent process suitability standards have enabled us to harmonize analytical practice across discovery. Procedures and protocols that have been developed and established as a direct result of the project are now routinely applied to synthesized and purchased new additions to the collection. We have a greater understanding of compound-dependent analytical responses that is supported by validated reference data. We have also become aware of how important it is to be able to readily QC components of a process and to know fundamentally what we are screening, how pure it is and how much of it is present.

We are now starting to see the benefits of this project in higher retest success rates and fewer samples undergoing structural elucidation because they were wrong in the screening collection.

Future challenges

The major challenge, with all of the pressures of productivity and return of investment in the modern pharmaceutical industry, is not to allow the new standard to deteriorate. At GSK, we recognize the immense value of our new 'pure and sure' screening collection. As a result, we have maintained the project infrastructure, albeit at a reduced level, so all new chemical entities entering the screening collection are subject to the same high level of scrutiny.

During the course of the project we have become more aware of the potential pitfalls of DMSO storage (along with the many benefits) and feel that we have an ongoing strategy that allows us to ask questions about the purity, stability and concentration of our screening solutions [3] at critical process points. Having established a compound-integrity standard for our collection, we are actively quantifying our screening solutions at these crucial points to understand the potential source of compound losses during the drug discovery process.

The mass of data generated by this project offers many other rewards from data mining. Other areas that we will be looking into include automated quantification, prediction of UV spectra and generation of a structural model of chemical stability in DMSO. There may be many other patterns that arise from the data and we will continue to look for other ways of using this invaluable resource.

Acknowledgements

The authors acknowledge the contribution of the following people from Analytical Sciences, Compound Management, Cheminformatics and Facilities Management in making the project a success: Zoe Blaxill, Steve Besley, Ken Murray, Sue Holland, Steve Trowbridge, Rob Hughes, Rob Lifely, Paul Wallace, Paul Burke, Jason Ejimadu, Helen Dawson, Mark Scott, Irene Areri, Elina Snell, Pardeep Sandhu and Jane Henshaw.

References

1 De Biasil, V. *et al.* (1999) High Throughput Liquid Chromatography/Mass Spectrometric Analyses Using a Novel Multiplexed Electrospray Interface. *Rapid Commun. Mass Spectrom.* 13, 1165

2 Blaxill, Z. *et al.* (2001) The Effect of Sealing Versus Lidding Plates on Degradation and Water Uptake of Compounds Held in DMSO at 4°C. *Poster at the 7th annual conference of the Socioety for Biomolecular Screening*, Baltimore, September 2001

3 Lane, S.J. *et al.* (2005) Toward single-calibrant quantification in HPLC A comparison of three detection strategies: evaporative light scattering, chemiluminescent nitrogen, and proton NMR. *Anal. Chem.* 77, 4354–4365