

Simultaneous visualisation of attrition and timelines

funnel diagrams

In drug discovery both attrition and timelines are important factors to consider when applying process improvement to lead optimisation. In our attempts to improve the visualisation of both of these factors simultaneously, we developed funnel diagrams, which allow turnaround time and percentage progression to be viewed in a glance. This tool has been rolled out to the whole of research, encouraging an open environment that involves everyone. In this article, we would like to present examples of the funnel diagrams, implementation details and discussions of issues, the people aspects of successful deployment and possible future developments.

**By Dr Linda Hiron,
Dr Craig Johnstone
and Colin
Sambrook-Smith**

The problems faced by the pharmaceutical industry have been widely reported, and have been approached from a variety of angles. The scientific and technical challenges, the fulfilment of 'low hanging' opportunities, the increasing demands for breakthrough medicines from payors, and the increasing demands for improved safety from regulators have probably all played a contributory role in the current state of the industry. The heart of our problems can be articulated in Paul's¹ recent estimate that it costs as much as \$1.8 billion to create a successful drug. This is a cumulative cost which is highly sensitive to the contributions from both attrition and the consequences of the many man-years of time invested. In recent years, an appreciation has emerged that the science of drug discovery and development runs on underlying processes and attempts to improve these supportive processes have been described²⁻¹⁶.

At Prosidion, we wished to examine and improve the systems and processes in research, with a particular focus on the testing process in the

first instance. As is common in test cascades, tests are typically arranged sequentially based on capacity and complexity, and compounds which pass certain criteria progress to the next test. From a science perspective, one is primarily interested in the test results, but from a process perspective, there are two principal pieces of information which are useful to gather, track and manage: attrition rate and time.

Herein we describe our attempt to create a simple simultaneous visualisation of these two key elements. We discuss benefits and issues, and propose extensions to the basic format which may have broader, useful applications to the visualisation of time and attrition in other domains of the drug R&D process.

The funnel diagrams and their current application

A single funnel (Figure 1) represents the associated assay timelines and attrition rates through the drug discovery cascade for new compounds synthesised and registered during a given month for one

research project. Each rectangle within a funnel represents an 'event' being tracked through the cascade. In this example, an 'event' is when a compound is first synthesised and registered or when an assay is completed. The width of each rectangle represents the number of compounds processed through that event. This volume is also annotated to the right of each rectangle. Hence a single funnel shows the attrition level by its funnel-like shape. Time is represented by the elongation of the funnel, since the distance between each rectangle and the top of the funnel represents the average number of days from compound registration to the completion of that event. These average (mean) times are also written to the left of each rectangle. It is important to note that the timescale is linear above a three-month border and that time is not scaled below the border. This change of scale is required to accommodate visualisation of short times at the top of the funnel and long timescales at the end. From **Figure 1** it is evident that in October 2009, in this research project, 50 new compounds were registered. 46 of these went through Assay 1, 41 through Assay 2, 10 through Assay 3 and so forth, ending with just two of the original 50 being tested in Assay 7. It took an average of four days from compound registration to get results from Assay 1 and an average of 59 days to get results from Assay 4. Note that Assays 5, 6 and 7 took more than three months to be completed and, despite being placed together at the bottom of the funnel, they differ greatly in completion time. This visualisation of new compounds made in a single month prompts us, at a glance, to ask useful and insightful process improvement questions, such as: what happened to the four compounds that were not put through the front line assay?; what is driving high attrition between Assay 2 and 3?; why is it taking so long to get results from Assays 5, 6 and 7?

Further insight into higher-level project considerations can be derived by examining the changes in funnels across a series of months presented in a trellis. To illustrate this, we present a six-month snapshot view of one of our lead optimisation research projects (**Figure 2**). Looking across the top of the funnels at their widths, we can quickly see the large increase from May to June in the number of new compounds being synthesised. In fact, we had moved additional chemistry resources on to this project in May, and so the increase in volume of new compounds is the detected consequence of the resource change. Notice in July, June and April the funnels fold back on themselves. This shows that the assays have been carried out

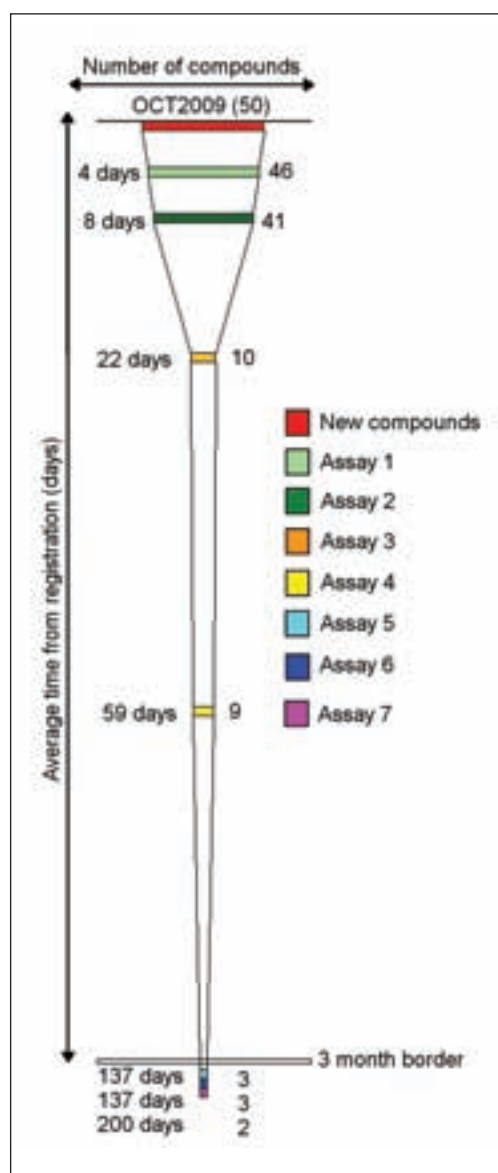


Figure 1

A single funnel. Simple simultaneous visualisation of assay timelines and attrition rates for the test cascade of a particular project in October 2009. Each rectangle represents an event being tracked through the cascade. The width of a rectangle represents the number of compounds processed by that event (also annotated to the right of each rectangle). The distance between a rectangle and the top of the funnel represents the average number of days from compound registration to completion of that event (also annotated to the left of each rectangle)

in a different order to the stated cascade, with Assay 4 results coming out before Assay 3. Assays 1 and 2 are consistently completed within a week of compound registration, regardless of the volume of requests. This suggested that this is a minimal impact area for process improvements. However, the funnels highlight more variable timelines for Assay 3, again seemingly independent of the assay's throughput. As a result, the processes around assay three were examined during June and July.

It was discovered that a significant and variable amount of time was taken in submission of Assay 3 requests, due to the nature of decision making based on primary assay SAR information from a set of compounds (not just the compound under

Informatics

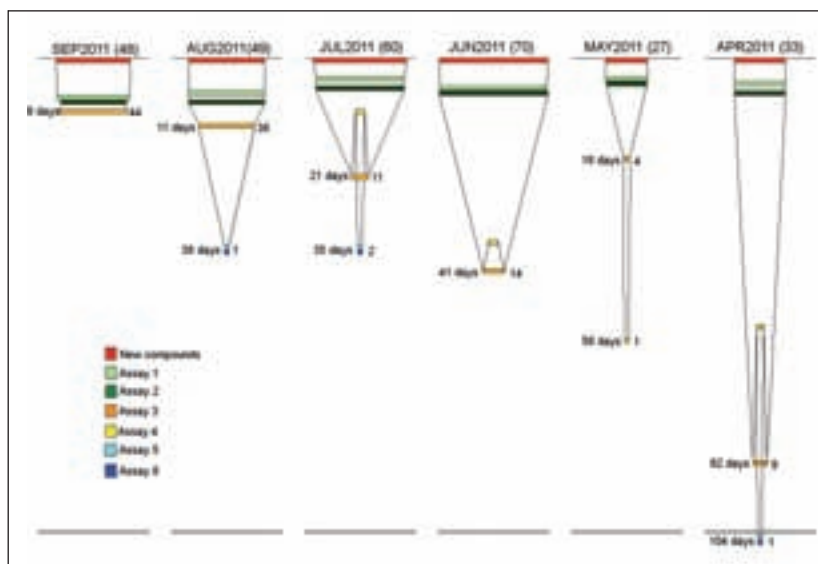


Figure 2
Six-month snapshot of a lead optimisation project using a trellis of monthly funnels. See Figure 1 for a detailed explanation of a single funnel and its annotations

question). This factor was removed by moving Assay 3 up to the top of the cascade and therefore eliminating the need for assay submission decisions. Secondly, the scientists performing the assay sometimes found it hard to locate the compounds. This was solved by sending a separate vial of the compound straight to Assay 3 without going via Assay 1 and 2. The third change was to run the assay once a week, regardless of volume of submission. These changes were implemented during August, and the August and September funnels clearly signal the impact these changes have had on Assay 3 turnaround. Furthermore, since Assay 3 is an *in vitro* surrogate of the more complex *in vivo* Assay 5, there is a large knock-on effect on Assay 5 and 6, the data from which are available two months faster. Finally, by removing the gated selection step between assay 2 and 3, most of the compounds are tested in all three frontline assays (1, 2 and 3), providing more scientific knowledge and understanding of the structure, activity and property relationships in the class of compounds, and therefore providing designers with better insight into how to design better compounds that have a greater chance of passing further down the cascade in the next cycle^{16,17}.

Roll out and implementation

The funnel diagrams have been rolled out as part of Prosidion's centralised reporting system, meaning that everyone within research has access to them. They are updated automatically once a day, giving an up-to-date view of the cascade within a particular drug discovery project. The funnels are embedded in, and drawn from, a wider informat-

ics infrastructure which involves: multiple scientific data repositories, a data warehouse, a reporting system and the code to draw the funnels (Figure 3).

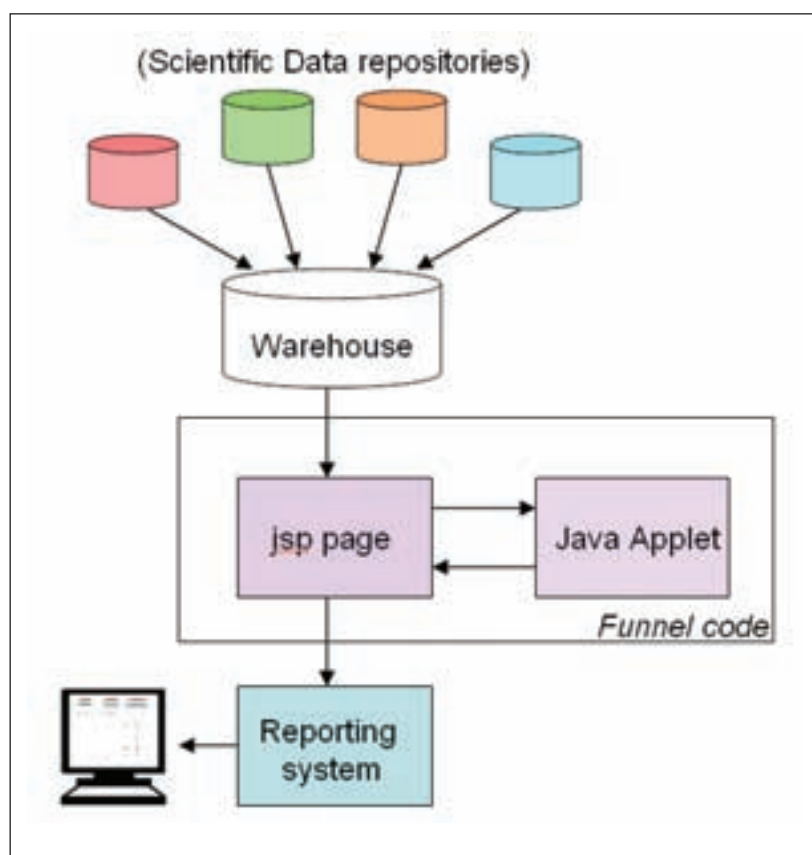
At Prosidion we use Browser, from Dotmatics (www.dotmatics.com), as our reporting system. This is a highly versatile web-based reporting tool, which can be easily extended to either link out to or contain in-house functionality. The roll out of the funnel diagrams are a good example of this flexibility. In our distribution of Browser there is a tab placed at the top of each research project's query page that allows users to go directly to the funnel diagrams for that project. This is in line with the aim of users only having to go to one centralised location from which they can then link out to other tools. The funnels are embedded into Browser via a jsp page. This jsp page requests the data required to draw the funnels from our data warehouse. Our data warehouse is a centralised database of all our assay and compound data gathered together from several data sources and was implemented in collaboration with Raptor Informatics (www.raptorinformatics.com). It provides us with a centralised repository that our reporting system can query efficiently. A set of materialised views written especially for the funnels, that refresh once a day, collate and pivot the relevant compound and assay attrition and timeline data out of our repositories into a format that can be easily translated into the funnel diagrams. Our main repository of chemical and biological data is ActivityBase™, from IDBS (www.idbs.com). Finally the jsp page performs some additional formatting and sends the data to a java applet, which draws the funnels.

Issues to be aware of

The benefits of simply and transparently visualising turnaround time and attrition within a drug discovery project are, we believe, fairly self evident. However, our experiences have highlighted that there are some issues which need to be considered and navigated, ideally in advance of implementation.

The first of these is the choice of date from which the turnaround time is calculated. This is a simple problem to state, but it can be deceptively intricate to get right. Many of the issues stem directly from the perceptions of people working within the processes about how these data might be interpreted, used (and potentially abused). To illustrate the point, a database which tracks a testing workflow may have a number of date entries, such as those in Figure 4.

From an overarching process or value stream perspective, the time taken from when the opportunity to test a compound arises (its original preparation date) to when the data is available for interpretation is important. This time lapse includes a number of sub-processes, as well as waiting time between steps, and the time taken to reach decisions. However, a project team may be interested in the turnaround time from when they decide they want a test run, until they get the result. This excludes the time it has taken them to reach the view that they want a test run, perhaps with reasonable justification in some cases, since perhaps they were awaiting other results in order to decide. Underneath this project cycle, there may be a number of subprocesses, which often lie within a departmental boundary, such as a compound management group which receives requests and despatches samples to testers and the testing groups such as pharmacology and DMPK. In each of these departmental cases, individuals are often directly associated with the execution of the test, and are therefore sensitive to the transparency and exposure these data can bring, unless cycle times are short. Inevitably, since there are sensitivities about prompt decision-making, and individual work cycles, debate and discussion can ensue around which is the 'correct' cycle time to monitor. The nature of the test in question can also play a major role: for example, rapid set-up and cycle-time *in vitro* assays are often uncontroversial, but downstream *in vivo* assays, which can involve extensive run-in times, and may also include extensive dosing periods, can appear to have long cycle times. They are usually placed downstream, and the compounds are selected based on a battery of upstream results and decisions, and therefore the overall registration to result cycle times can look very substantial when expressed as a proportion of the entire project lifetime. Many practitioners may feel that these long cycle times are beyond influence and manipulation, and as such, they can feel that making these long cycles transparent is unhelpful and frustrating. On the other hand, for a project manager, it can aid planning to know at the outset that the project may only get one or two shots at a pivotal sub-chronic study, and the compound needs to be synthesised as much as nine months before the deadline for the data. In this way, it becomes clear that there is no 'correct' timestamp or use of the data, since it depends on what questions one is asking, what one is trying to achieve, and where one starts from as a baseline. Fortunately, the funnel drawing tool is sufficiently flexible to enable visualisation of any timestamp,



as long as the raw data is available in the source databases. Whatever the sensitivities, it is clear that from a drug discovery improvement point of view, the total time taken from opportunity to outcome is the one which adds most delay, contains most waiting and decision-making, and offers the greatest potential improvement opportunity to the research manager.

Knowing the key dates one wishes to capture and how to use them is important but in practice it can be challenging to capture these in a consistent manner. Such consistency is important for correct and insightful interpretation of the funnels, but collecting data from a wide variety of experimental paradigms can make this problematic. For example, the definition of experiment start date for a short duration experiment is straightforward, and the opportunity for error due to variation is minimal. However, the start of longer term experiments can be ambiguous (eg the first day of dosing?; the day the animals commence pre-conditioning?). Furthermore, not all organisations are in a position to record elaborate studies in their in-house databases, since the data is often complex to describe for electronic capture and the number of such studies is often relatively small.

Figure 3
Informatics Infrastructure. The funnels are connected up to a reporting system and to several scientific data repositories via a highly structured data warehouse

Informatics

References

- 1 Paul, SM et al (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* 9, 203-214.
- 2 Petrillo, AL. Lean thinking for drug discovery – better productivity for pharma, *Drug Discovery World*, Spring 2007, 9-14.
- 3 Houston, JG et al. Technologies for Improving Lead Optimisation. *American Drug Discovery*, 1 (3), Oct/Nov 2006, page 6-15.
- 4 Hammond, C and O'Donnell, CJ (2008). Lean six sigma – its application to drug discovery. *Drug Disc. World* Spring 11-18.
- 5 Andersson, S et al (2009). Making medicinal chemistry more effective – application of lean sigma to improve processes speed and quality. *Drug Discov. Today* 14, 598-604.
- 6 Russell, K (2008). Improving pharmaceutical R&D using lean sigma. *PharmaFocus Asia* 7, 48-51.
- 7 Weller, HN et al (2006). Application of Lean manufacturing concepts to drug discovery: rapid analogue library synthesis. *J. Comb. Chem.* 8, 664-669.
- 8 Sewing, A et al (2008). Helping science to succeed: improving processes in R&D. *Drug Discov. Today* 13, 227-233.
- 9 Sewing, A (2008). Evolution in thinking and processes? *Drug Discov. Today. Technol.* 5, e9-e14.
- 10 Barnhart, T (2008). Lean in R&D: the surprising fit. *Future State* Spring 1-3.
- 11 Allen, M and Wigglesworth, MJ (2009). Innovation leading the way: application of lean manufacturing to sample management. *J. Biomol. Screen.* 14, 515-522.
- 12 Ullman, F and Boutellier, R (2008). A case study of lean drug discovery: from project driven research to innovation studios and process factories. *Drug Discov. Today* 13, 543-550.

These layers of complexity can be exacerbated when experiments are being conducted by multiple external organisations such as CROs. All of these issues can be overcome with good operational definitions, but it is helpful to consider and control these sources of variation upfront before implementing visualisations to avoid confusion and misinterpretation later.

The example funnels (Figure 2) are representations of attrition and average cycle times (we have used mean times, but note that medians are often chosen to reduce the effect of outliers). The need to average the cycle times for simplicity of visualisation comes at the cost of loss of important insight on variation. From a process perspective, it is usually more desirable to have predictable cycle times with low variation, rather than an apparently short average time with high underlying variation, since reliable times enable forward planning. In our experience, at the outset of this kind of process work, cycle times are long and variation is high, and so after the first cycle of intervention, it is often common and sufficient to measure reduction in cycle time as the primary goal. However, as an improvement system evolves and matures, reduction in variation around the average becomes more prominent. Consequently, one of the possible future developments of the funnel representations could be representation of the underlying variation (see next section).

Perhaps the most important issue arising from open and transparent visualisations of processes is the human aspect. Organisational management cultures tend to emphasise performance manage-

ment, and these principles are exerted at the individual level, with employees having salaries and/or bonuses awarded on the grounds of individual performance, albeit sometimes within a team context. Understandably, this predominant culture can give rise to friction due to an individual's sensitivities about individually attributable workflows being made openly available for scrutiny. Furthermore, department leaders can feel exposed if it appears that 'their' department has longer cycle times than others. It is unfortunate that these sensitivities arise, since the principles which underpin the desire to make the process performance visible and transparent have, at their heart, the idea that it is the system and process which is being examined and improved, not the individual or team^{18,19}. Therefore substantial efforts to communicate the distinction are required before, during and after implementation, and proactive steps to avoid blame, finger-pointing and punitive behaviours are essential to support the rhetoric if culture change is to be successfully brought about. Training and coaching at all levels may even be required. On the other hand, transparency of improvement can be very motivating to those involved. Indeed, the most positive feedback we received about the funnels came from staff who were directly involved in running Assay 3. They could see the problem, the result of their interventions, and the magnitude of their improvements directly. They were proud of their achievements, and were proud to know that everyone else could see those improvements too.

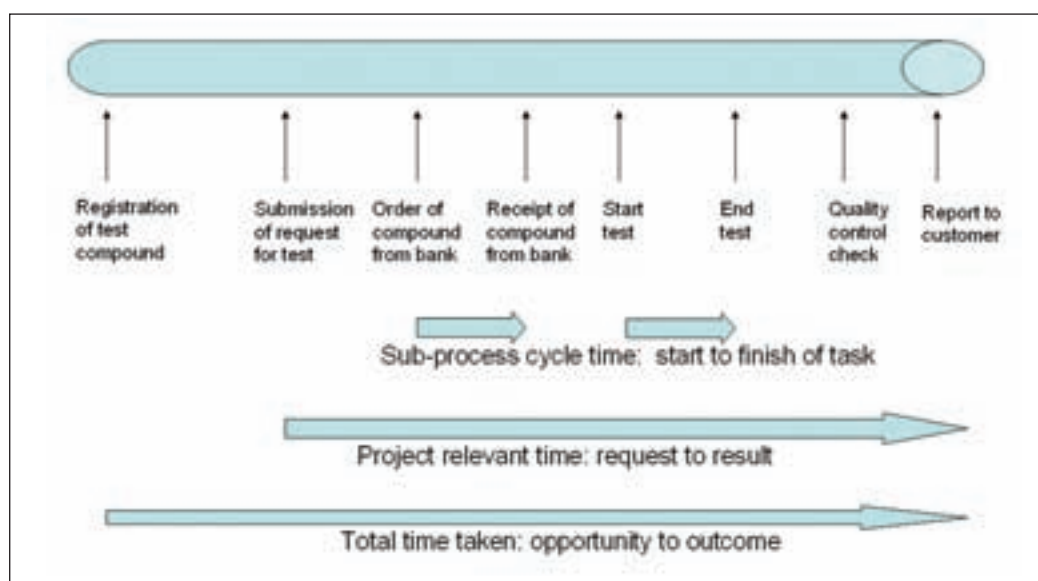


Figure 4: Testing workflow time-stamps. Choosing the most appropriate time stamps is an important precursor to maximising the use of the funnel diagrams

Continued on page 49

Future developments

As previously mentioned, showing just the average assay timelines in the funnels means that there is no indication of variation. In order to overcome this, each event rectangle could be populated with an error bar, or surrounded by spots to represent the individual compound times. For even further granularity, every compound could be plotted out in a separate column of events, rather similar to the excellent 'event-based analyses' first reported by Petrillo². However, we feel that each of these expansions of the funnels provides more information but compromises simplicity, and tends to erode the simultaneous view of attrition and time. We therefore prefer to keep the funnels in their simple form with a click to expand function, where the error bars or chosen expansion can be easily switched on and off.

A further possible extension of the funnel would be to track upstream through the conception and synthesis stages to registration. This would involve connecting the funnels up to a compound ideas database^{16,20,21}. The funnels could then be used to gain insight into: how long it takes to decide to make a compound idea, how many ideas are actually carried forward and whether the idea volume correlates with synthesis volume. Another event extension would be to integrate the funnels with an assay requesting system, in order to be able to directly visualise time for decision-making.

Here, we have described a simple method of visualising attrition and speed in a single, easy-to-interpret view, as applied to the post-synthesis test cascade in drug discovery projects. However, given that speed and attrition are at the heart of the industry's problems, it is trivial to envisage wider, more strategic applications of visual representation of these two issues simultaneously. For example, within a large company with numerous projects at various stages through the discovery and development process, one could readily apply the funnel diagram to represent real-time residence time and attrition across the portfolio. Similarly, a portfolio containing a number of in-licensing opportunities which are being evaluated in parallel at various stages of maturity could also be readily visualised in this way. Furthermore, the funnel diagram could find utility in cross-company benchmarking reports. Participating companies routinely compare their success/attrition rates and residence time per phase. These dimensions are usually compared separately, leaving questions about whether the 'fast' companies suffer higher attrition later, or operate a different attrition model. The funnel diagram would enable company speed and attrition profiles to be compared directly.

In drug discovery we need to improve the speed and quality simultaneously. Going fast in the wrong direction is not fruitful and producing the highest quality too slowly will be commercially unattractive. Thus it is valuable to practising scientists, managers and leaders to be able to view speed and attrition simultaneously. Therefore we offer this simple funnel representation as a small yet scalable and, we believe, more widely applicable tool to visualise these two most important parameters in addressing the challenges of improving drug discovery.

DDW

Dr Linda Hirons is currently CEO at Amethyst Informatics Ltd (linda.hirons@amethystinformatics.co.uk), having spent the past 3½ years working at Prosidion Ltd, a research-driven biotech company in Oxford, within the Research Technologies team as an informatics scientist. Previously she was a Post Doc at Lilly, using the KNIME pipelining technology to develop an inverse structure-based design tool. She obtained her PhD at Sheffield University under the supervision of Professor Peter Willett and Professor Chris Hunter, looking at activity fingerprints in DNA.

Dr Craig Johnstone joined (Astra)Zeneca Pharmaceuticals in 1994. He has worked in oncology, inflammation and cardiovascular research programmes. As the Director of Chemistry, Cardiovascular & Gastrointestinal Research Area in the UK, he became interested in the improvement of the discovery process and in 2008, in addition to his line management role, he was appointed Value Chain Leader, CV&GI at AstraZeneca. He joined Prosidion in 2011, as Head of Medicinal Chemistry.

Colin Sambrook-Smith graduated in chemistry from the University of Bristol. He joined Courtaulds plc (later Akzo Nobel) and spent more than 10 years working on materials modelling and simulation. In 1999 he joined OSI Pharmaceuticals and focused on the structure-based design of kinase inhibitors for oncology. He transferred to Prosidion in 2006 and is Head of Research Technologies, responsible for informatics, computational chemistry and array chemistry.

Continued from page 48

13 Carleysmith, SW et al (2009). Implementing lean sigma in pharmaceutical research and development: a review by practitioners. *R&D Manag.* 39, 95-105.

14 Uitdehaag, JCM (2011). The seven types of drug discovery waste: toward a new lean for the drug industry. *Drug Discov. Today*, 16, 369-371.

15 Walker, SM and Davies, BJ (2011). Deploying continuous improvement across the drug discovery value chain. *Drug Discov. Today*, 16, 467-471.

16 Plowright, AT et al (2011). Hypothesis driven drug design: improving quality and effectiveness of the design-make-test-analyse cycle. *Drug Discov. Today*, (in press), doi:10.1016/j.drudis.2011.09.012.

17 Gleeson, MP et al (2011). Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nat. Rev. Drug Disc.*, 10, 197-208.

18 Shook, J (2010). How to change a culture: lessons from NUMMI. *MIT Sloan Manag. Rev.* 51, 63-68.

19 Deming, WE (1993). *The New Economics for Industry, Government, Education*, second edition.

20 Lee, M et al (2011). DEGAS: Sharing and tracking target compound ideas with external collaborators. *J. Chem. Inf. Model* (in press), DOI: 10.1021/ci2003297.

21 Brodney, MD et al (2009). Project-focused activity and knowledge tracker: a unified data analysis, collaboration and workflow tool for medicinal chemistry project teams. *J. Chem. Inf. Model.* 49, 2639-2649.