

High-Throughput Experimentation in Pharmaceutical Process R&D: Developing a New Software Workflow to Overcome Downstream Data-Analysis Bottlenecks and Improve Productivity

Paul D. Higginson and Neal W. Sach*

Pfizer Global Research & Development, Pfizer Ltd., Ramsgate Road, Sandwich, Kent, CT13 9NJ, UK

Abstract:

A new software workflow is described to overcome our data analysis, data-modification, and visualisation bottlenecks. Through a combination of in-house and external supplier expertise, a software workflow has been developed to tackle the mass of data generated from high-throughput experimentation. Also described is our solution to high-throughput HPLC reaction analysis. In overcoming these bottlenecks, the productivity of our screening technology has dramatically increased.

Introduction

Combinatorial methods and high-throughput experimentation have radically improved the productivity of Discovery Research departments within the pharmaceutical industry. New technology has enabled vast numbers of compounds to be quickly synthesised and screened leading to a dramatic increase in the potential number of new drug candidates entering the development pipeline.^{1,2} In addition, tighter controls on drug candidate nomination have improved attrition rates leading to an increased workload within Process R&D departments across the pharmaceutical industry.

The sector has responded by developing innovative new technologies and combining these with statistical tools such as Principle Component Analysis (PCA) and Design of Experiments (DOE) to improve productivity and efficiency.^{3,4} This combination of tools has enabled a step-change in the number of experiments that can be run in the Process R&D laboratories allowing Discovery synthetic routes to drug candidates to be screened, optimised, and scaled faster and with less resource.⁵ Presently however the workload demands from Discovery Research departments continue to increase, and in particular there is now a new pressure within Development to minimise the time and resources spent delivering an early toxicological batch from which key-decision points can be reached. The effect of this strategy on Process R&D is 3-fold:

(a) The process chemist must initially focus on rapid short-term bulk delivery rather than longer-term process development.

(b) With the reduced time lines, extra attention must be paid to the safety and robustness of the synthesis as little experience will be achievable prior to scaling the chemistry.

(c) New technology and separations science must be fully utilised to meet the vast resource shortfall created by this approach.

As part of a strategic approach to meet these increasing demands, the Pfizer Development laboratories at Sandwich began examining high-throughput experimentation tools that could meet this experimental shortfall. A robotic microreaction screening system was developed by modifying commercially available solids and liquid handling technology^{6,7} that allowed the completion of hundreds of reactions per day whilst using only small quantities of valuable starting materials (typically 1–5 g per screen of 100 reactions). The technology operates at the Discovery–Development interface and enables the rapid mapping of chemical space for a given transformation; two outcomes are generally obtained:

(a) A range of alternative conditions are identified from the chemical space enabling the chemist to focus on only those reactions that are likely to yield the optimum process.

(b) The current conditions are confirmed as being amongst the best available, and resources may be better spent investigating alternative routes than optimising a poor reaction.

In both cases a wealth of information can be gained from only a small investment in compound; however, the time spent analysing, visualising, and mining our data was reducing the impact of the technology. We present a software workflow developed in-house and with external partners to overcome these bottlenecks that greatly improves the productivity when utilising this approach.

Developing the Workflow

An examination of our existing high-throughput experimentation workflow immediately highlighted a number of bottlenecks (Figure 1).

The workflow analysis illustrated that whilst the majority of our protocols could be prepared and run within a day using

* E-mail: neal.sach@pfizer.com.

- (1) Bauser, Marcus; Stakemeier, Hubertus. Instrumentation for combinatorial chemistry. *Handbook of Combinatorial Chemistry* 2002, 1, 190–224.
- (2) Balkenhohl, Friedhelm; von dem Bussche-Huennefeld, Christoph; Lansky, Annegret; Zechel, Christian. Combinatorial synthesis of small organic molecules. *Angew. Chem., Int. Ed. Engl.* 1996, 35 (20), 2288–2337.
- (3) Emiabata-Smith, David F.; Crookes, Derek L.; Owen, M. R. A Practical Approach to Accelerated Process Screening and Optimization. *Org. Process Res. Dev.* 1999, 3 (4), 281–288.
- (4) McKay, B.; Hoogenraad, M.; Damen E. W.; Smith, A. A. Advances in multivariate analysis in pharmaceutical process development. *Curr. Opin. Drug Discov. Dev.* (2003 Nov), 6 (6), 966–77.
- (5) Laird, T. Special feature section: Laboratory automation in process R&D-Editorial. *Org. Process Res. Dev.* 2001, 5 (3), 272.

(6) ReactArray Workstation: Available from Anachem, Anachem House, Charles Street, Luton, Bedfordshire, LU2 0EB, UK., <http://www.reactarray.com>.

(7) Autodose Weighing Station: Available from Autodose, 18 Chemin des Aulx, CH 1228, Geneva, Switzerland., <http://www.autodose.ch>.

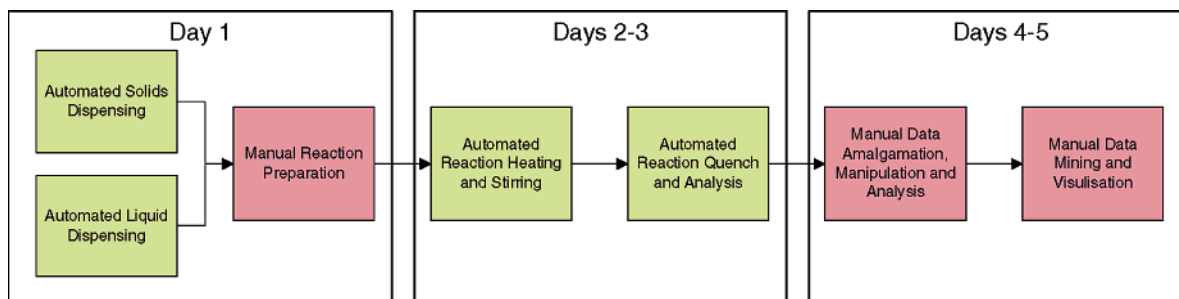


Figure 1. Illustration of high-throughput experimentation bottlenecks. Green denotation indicates automated processing step, whilst red indicates significant manual intervention required.

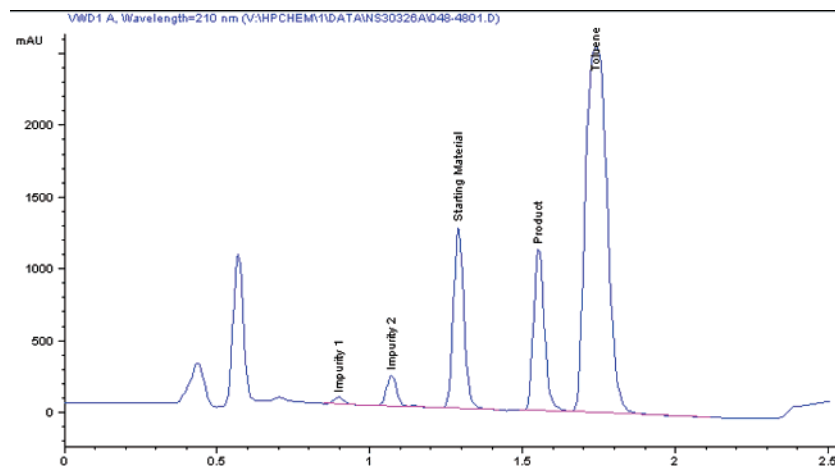


Figure 2. Example chromatogram using Chromolith column technology for high-throughput reaction analysis.

automated instrumentation, our downstream processing required extensive manual intervention and considerably lengthened the time from start to finish. To maximize the productivity of our high-throughput experimentation three areas were identified for improvement and investment: reaction analysis, data analysis, and data visualisation.

Reaction Analysis

The majority of our high-throughput protocols end with an automated quench to obtain homogeneity before sampling for HPLC or occasionally GC analysis to determine conversion and purity. The limitation of our analysis methods was 2-fold: first, the analysis of a 100-reaction protocol could take over a day and added to the length of our screen; second, with an analysis cycle time of over a day, no intermediary sample points were possible without time-consuming offline analysis.

To overcome these hurdles we consulted our Discovery and Analytical colleagues regarding the latest technology in high-throughput HPLC and were recommended to Merck Chromolith chromatography columns.⁸ Unlike conventional columns that are filled with micron-sized matrix particles, Chromolith columns have a highly porous rodlike structure that produces an extremely low backpressure and enables flow rates up to 5 times higher than those of traditional columns. As a consequence of these extended flow rates, extremely fast separations are possible with the added benefit

that system and column reequilibrations are considerably reduced. In applying this technology to our existing instrumentation, a cycle time of just 2.5 min becomes typical (Figure 2).

By reducing our HPLC cycle time so radically, a typical protocol of 100 reactions can be analysed in hours permitting the capture of multiple time points for each reaction. These extra data points are significant as they reduce our reliance on the final sample point and generate important reaction rate data; however, our data-analysis issues only worsen. To take full advantage of our new reaction analysis technology, our data analysis issues needed resolving, as a typical protocol was now capable of producing over 400 data points in 24 h.

Data Analysis

The majority of our high-throughput reaction analysis is carried out using Agilent or Gilson instrumentation. In both cases acquisition methods can be created to automatically analyse chromatograms and output data to an electronic file. In the case of our Gilson instrumentation the output from multiple chromatograms was in the form of a single large text (TXT) file that then required significant data manipulation before some level of knowledge could be ascertained. In the case of our Agilent instrumentation the output was much more fragmented and in the form of comma separated value (CSV) files, one per acquisition; i.e., 400 chromatograms would lead to 400 CSV files. In both cases manually manipulating this amount of data was unfeasible and a software solution was urgently required. Before this issue

(8) Chromolith Performance RP-18e: Available from Merck, Harrier House, High Street, Yiewsley, West Drayton, Middlesex, UB7 7QG, UK., <http://www.merck.de>.

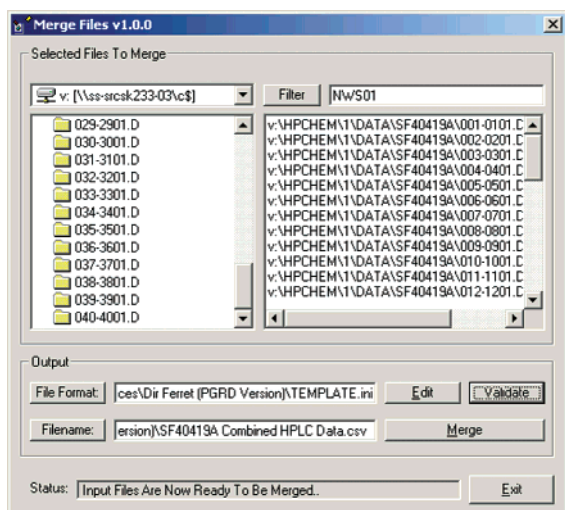


Figure 3. A screen shot from our VB based program designed to merge multiple CSV files from Agilent instrumentation.

could be tackled a solution was first required to amalgamate the fragmented output from our Agilent instrumentation and obtain a single file as with our Gilson instruments.

Working with in-house experts, we created a program using Visual Basic (VB) to simply merge the multiple CSV files from the Agilent instrumentation such that a very similar output to our Gilson instrumentation was achieved (Figure 3).

Having harmonised the output from our instruments, we next searched for software that was capable of formatting this raw instrument data into a comprehensible form. Our requirements for the software were 3-fold: (a) Accept our instrumentations TXT and CSV file formats and output the reformatted data into CSV or Excel (XLS) format. (b) Have a range of data-formatting functions allowing the user to simply and quickly build up a sequence of stepwise data-manipulation steps. (c) Be generic such that once built the method could be used without further modification on all future input instrument files.

A number of packages met these criteria; however, one in particular performed exactly the function we required, Data Analysis Package (DAP) by Avantium.⁹

Avantium are experts in high-throughput experimentation and developed the DAP software to quickly turn raw electronic instrumentation data into information. In developing the software, Avantium recognised DAP must integrate into a larger workflow and have provided exceptionally generic data-import and -export functions. These functions are the software's strength, and DAP is capable of working with almost any instrument that provides an electronic data file. Once inside DAP, the data can be manipulated as required and exported in any number of formats allowing the software to sit almost seamlessly between processes in a workflow. Pfizer collaborated with Avantium along with a number of other companies to develop DAP from an in-house software to a commercially viable product and in doing so ensured the product is completely compatible with a number of popular analytical instruments, including our Gilson and Agilent instruments.

Developing a data-manipulation method in DAP is a process of wiring together a number of functional components to achieve the desired data transformation. A small graphical icon represents each component, and dragging a series of these icons from the tree toolbar onto the workspace and wiring them together creates a method. A typical method begins with a data-import function (Figure 4a) that is then often linked to a data-stripping or reformatting function. Once the data is in the correct format, a number of transformations are possible within DAP from simple calculations to advanced statistical manipulation. Once transformed the data can either be visualised using one of the many types of plot (Figure 4b) or exported for visualisation and mining in secondary software (Figure 4a).

Once a method is created within DAP, it can be saved and reused on all future data sets of that format without any further modification. This is an important feature of DAP,

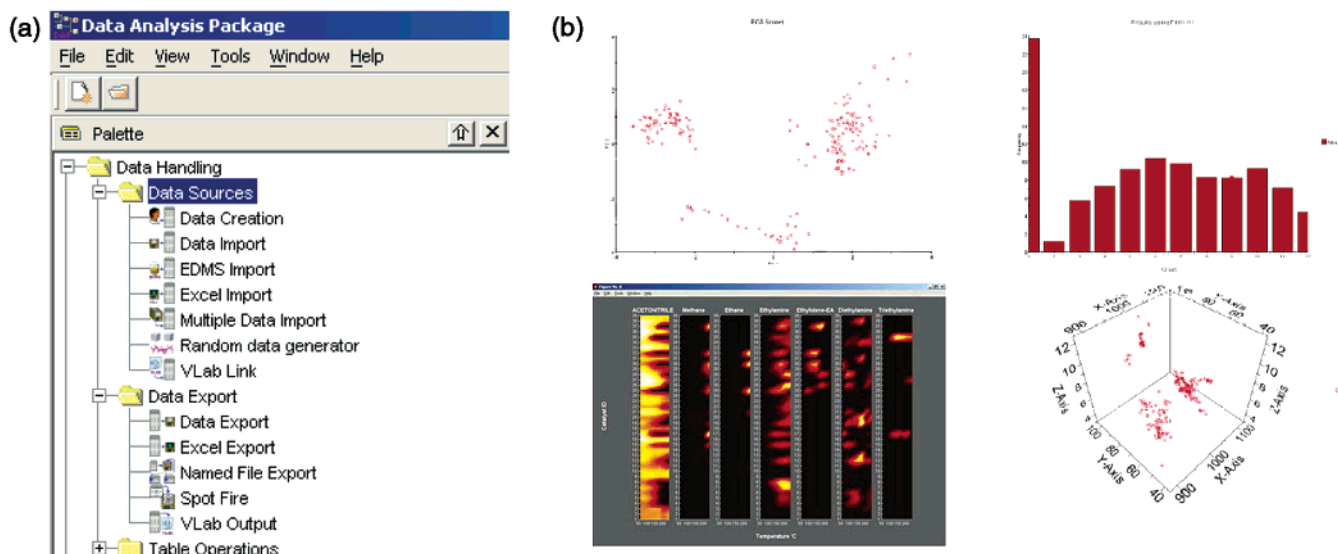


Figure 4. (a) An illustration of the extensive number of data-import and -export functions available in DAP. (b) Examples of the data-visualisation options available that include sunset plots, 2-D and 3-D scatter plots, and simple bar chart plots.

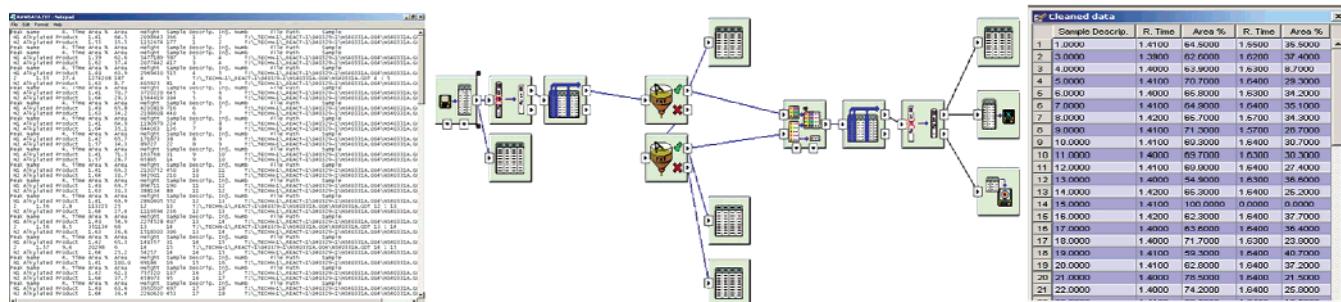


Figure 5. Illustration of the raw data entering the DAP workflow, a DAP workflow screen-shot, and an example output.

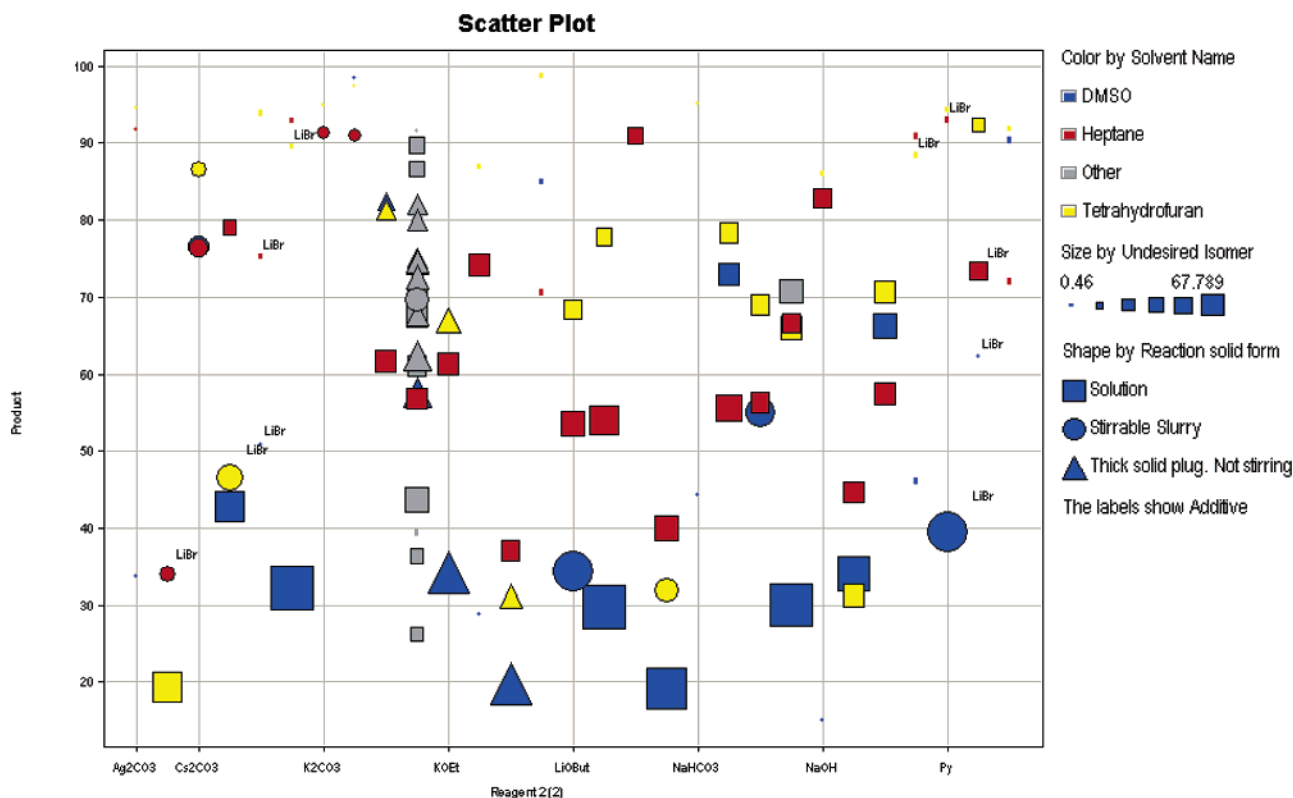


Figure 6. An example 2D Spotfire scatter plot simultaneously illustrating six variables, yield (y axis), base (x axis), solvent (colour), reaction solid form (shape), impurity levels (size), and reaction additives (label).

as it allows methods to be created specifically for a type of instrumentation data output that can then be shared and reused throughout an organisation without further modification. As a result only a handful of advanced users are required to generate instrument data-handling methods that can be used by all.

In our workflow the DAP software sits between the reaction analysis instrumentation and visualisation software. The user imports the raw electronic instrumentation data files into the appropriate DAP method which then transforms the data into a meaningful and understandable form before exporting in a file format compatible with our visualisation software (Figure 5). By incorporating DAP in this way, our raw data can be turned into knowledge in minutes compared with our original process that could take days.

Data Visualisation and Mining

The requirement to visualise and mine large experimental data sets is a common challenge, particular in the Discovery arena, where software solutions have existed for some time. In that context drug activity is plotted against various structure property relationships; however, such software is equally capable of serving wider applications. One such example is Spotfire DecisionSite (SDS)¹⁰ that has evolved from the Discovery arena as a leading data-visualisation and mining tool and has found applications across several fields. The software stands apart from a number of other commercial products on the market due its tolerance of raw data sets containing missing or slightly misformatted fields and its ability to turn these into clear visual scatter plots that illustrate key information.

(9) Avantium Data Analysis Package: Available from Avantium Technologies, Zekeringstraat 29, 1014BV Amsterdam, The Netherlands. <http://www.avantium.com>.

(10) Spotfire Decision Suite: Available from Spotfire AB, Första Långgatan 26, SE-413 28, Göteborg, Sweden. <http://support.spotfire.com>.

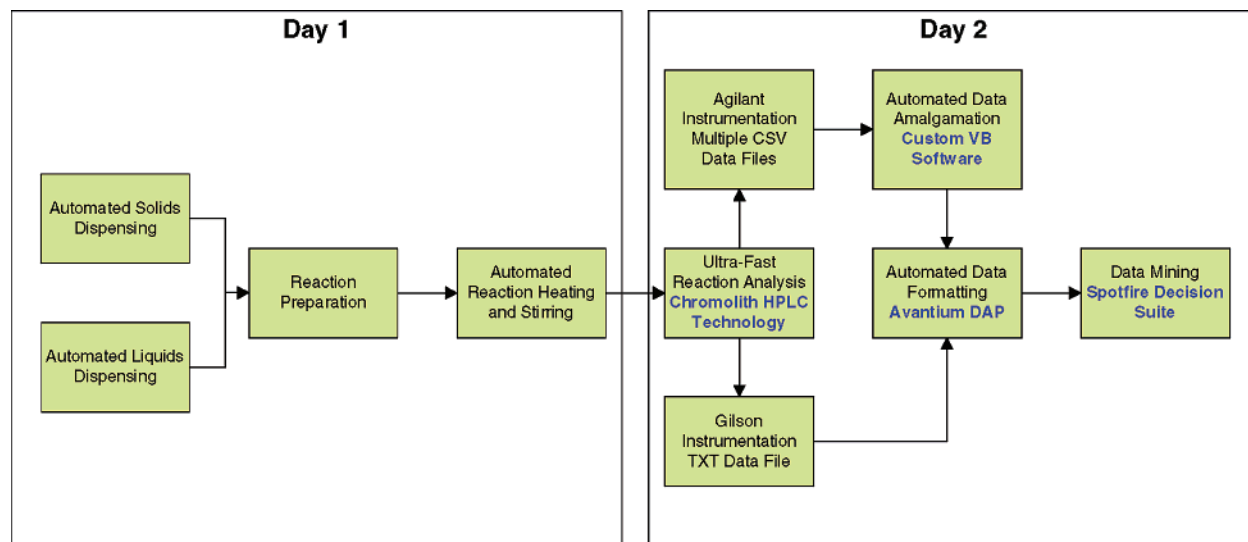
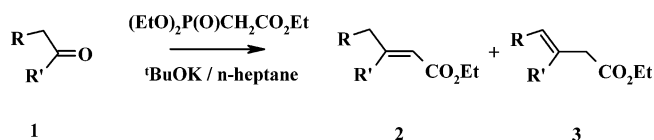


Figure 7. An illustration of the new software workflow.

Scheme 1



The raw data sets can be imported into SDS from a number of sources including file (XLS, TXT, and CSV formats), database, and clipboard. Once inside the software, a scatter plot can be quickly built by plotting the experimental factors against the responses. The scatter plots are capable of handling both continuous (temperature, yield) and categorical (solvent, yes/no) factors, and responses can simultaneously illustrate several of these at once through axis, shape, colour, and size. Once a scatter plot is assembled, the user can quickly mine individual data points or use the advanced query tools to identify more subtle trends and patterns.

The software is ideal for our application as DAP directly exports to SDS (Figure 4a, data export functions), and any misformatted fields are easily handled by the flexible software. In addition the SDS scatter plots are unparalleled as a method to quickly visualise the multiple categorical and continuous variables associated with high-throughput experimentation. The software completes our reaction analysis, data analysis to data visualisation workflow and also helps present complex data to our customers in an easily understandable format.

Figure 6 demonstrates the software's capacity to illustrate the results from a high-throughput olefination protocol. In this instance 120 conditions were examined at a 50 mg scale varying base, solvent, additive, and olefinating reagent to improve on the existing conditions (Scheme 1).

Whilst the conditions were high-yielding the physical processing issues caused variable levels of the undesired isomer (3) and was difficult to process on scale. To overcome these issues, a protocol was designed to encourage a desirable processing form (shape), low levels of undesired isomer (size), and good yield (axis).

Using our new downstream processing solutions, the protocol was completed within a number of days and reported to the project team using the Spotfire visualisation. Working with the project team, we modified the visualisation in real time and mined to extract the key trends (e.g., THF > heptane > DMSO) and exemplify the most important reactions (e.g., small, round spots above the 90% yield threshold). In applying the technology in this manner, the experimental space is rapidly mapped using very little material enabling the project team to quickly focus on the key reactions.

Conclusions

A new workflow has been developed to overcome our downstream bottlenecks and significantly improve the productivity of our high-throughput experimentation (Figure 7).

Whilst the downstream data-processing steps may look fragmented, the author has found, as with most technology, small fragmented operations with a low level of manual intervention require less long-term maintenance than one completely automated system. In fragmenting our data operations, trouble-shooting the workflow becomes simpler and less time-consuming.

Using the new workflow two goals are achieved: (a) Utilising the Chromolith HPLC technology and downstream data handling software multiple time points can be obtained for each reaction adding an additional level of quality and confidence to the reaction screen. (b) By reducing the reaction screen duration by half and minimising the level of manual intervention, the level of resource required is dramatically decreased.

Our screening technology is now having an impact across the early-development portfolio at the Discovery–Development interface. Through this bulk-sparing high-throughput experimentation, the experimental space can be rapidly mapped enabling the process chemist to focus on the areas of activity that are most likely to produce scalable reactions.

Acknowledgment

We gratefully acknowledge the following Pfizer staff: M. Taylor and B. Mathews for their advice and suggestions regarding Chromolith method development and the work of N. Betts and G. Williams for developing the VB based data-amalgamation software. We also thank the software develop-

ment team at Avantium for incorporating our data-modification suggestions into the commercial version of DAP.

Received for review June 24, 2004.

OP049873U