



**REAXYS MINI-SYMPOSIUM UK
WORKFLOW TOOLS IN CHEMICAL RESEARCH
BRITISH LIBRARY CONFERENCE CENTRE, OCTOBER 22, 2009**

A report by Dr. Wendy A. Warr



CONTENTS

INTRODUCTION: CHEMISTRY RESEARCH IN THE UK.....	1
FARHANA HOQUE, ELSEVIER, SENIOR MANAGER, KEY ACCOUNTS, UK	1
CURRENT TRENDS IN SYNTHETIC AND ANALYTICAL CHEMISTRY.....	1
PROFESSOR ELIZABETH HOUNSELL, BIRKBECK COLLEGE, UNIVERSITY OF LONDON	1
USE OF REAXYS IN SYNTHETIC CHEMISTRY RESEARCH.....	3
DR. JONATHAN GOODMAN, UNIVERSITY OF CAMBRIDGE.....	3
INNOVATIONS IN CHEMISTRY: AN ELSEVIER PERSPECTIVE	5
KAREL NEDERVEEN, PUBLISHING DIRECTOR CHEMISTRY, ELSEVIER	5
ENHANCING EFFICIENCY IN CHEMISTRY RESEARCH WORKFLOW	6
DR. DAVID EVANS, DIRECTOR SCIENTIFIC AFFAIRS, ELSEVIER PROPERTIES.....	6
CLOSING REMARKS AND DISCUSSION	8
REFERENCES	8

REAXYS MINI-SYMPOSIUM UK

WORKFLOW TOOLS IN CHEMICAL RESEARCH

A report by Dr. Wendy A. Warr

Wendy Warr & Associates wendy@warr.com; <http://www.warr.com>

INTRODUCTION: CHEMISTRY RESEARCH IN THE UK

Farhana Hoque, Elsevier, Senior Manager, Key Accounts, UK

Farhana announced the launch of Reaxys in the UK in November 2009, noting that a migration path from CrossFire will be offered and that some universities are already carrying out Reaxys trials.

She presented some data from Scopus on global chemistry research in the years 2005-2006. In terms of total article count, the UK comes third after the United States and China; and in terms of total citation count, it comes third after the United States and Japan. The Universities of Cambridge, Oxford and Manchester, and Imperial College, London produced the most articles (365, 303, 237, and 204 respectively), all highly cited.

Elsevier has a 26% share of all English language articles published each year, significantly more than any other publisher. Chemistry and chemical engineering is Elsevier's fourth biggest publishing sector, after life sciences, health sciences, and materials science and engineering. The company publishes about 30,000 articles a year in chemistry (excluding chemical engineering), which is about 30% of the world's output in chemistry and 10% of Elsevier's annual output of 300,000 articles. Elsevier also has other claims to excellence such as high 2008 impact factors, and a number of Nobel Laureates as authors or editorial advisory board members, including the 2008 recipients for the Nobel Prize in Chemistry.

CURRENT TRENDS IN SYNTHETIC AND ANALYTICAL CHEMISTRY

Professor Elizabeth Hounsell, Birkbeck College, University of London

Both synthetic and analytical chemistry have equal prominence in biomedical research with present advanced analytical tools leading to synthetic targets. Historically, Elizabeth worked in synthesis at ICRF, and the UK Medical Research Council from 1977-1994, before moving to work in analysis for University College London Medical School and Biochemistry, and now at the Birkbeck School of Biological and Chemical Sciences. Note the trend for chemistry to become part of other subject areas: in Elizabeth's case biology. Elizabeth edits the Elsevier journal *Carbohydrate Research* which covers primarily the medical field. She is also editor of a Springer biophysics encyclopaedia; the biophysics of glyco-lipid-protein systems and NMR of carbohydrates, lipids and membranes are her specialities.

Farhana Hoque mentioned journals and the stress on metrics. Journals such as *Carbohydrate Research* have a very conscientious and expert team of editors and referees that ensures that only a high standard of original research is published, but it is difficult to see how

standards are going to be maintained in the long term, as the number of journals worldwide, and the number of papers in those journals increases, because of the stress on metric assessments; and as number of submissions increases from countries that have less of a tradition of publication; and as the trend to open access publishing continues.

There is also a trend for researchers not to read the literature as it is published within a journal, getting an overview of their scientific disciplines, but to go to the literature electronically, to the Web and to databases on a need to know basis, that is, starting from scratch when beginning to write a grant proposal, a review, or an essay. Indeed, they may review the background before carrying out a new piece of synthetic research: hence the increased use of interfaces such as Reaxys.

Elizabeth went on to discuss synthetic targets and analytical methods. Natural products are highly complex. Woodward's team concluded the synthesis of Vitamin B12 in 1972; 100 co-workers spent 11 years devising a synthesis involving nearly 100 steps, yet Vitamin B12 is not a highly complex molecule, compared with many that Elizabeth has studied. Syntheses of other natural products can be admired monthly in a special column in *Chemistry World*. Antibiotics, anticancers, and vaccines are much more complex than Vitamin B12. Vaccines are conjugates of a protein and a carbohydrate, so more than one molecule would have to be synthesised.

Other targets are intracellular, mammalian cell products; cell surface glycoproteins and glycolipids (studied by rational drug design, NMR, X-ray, and MDS approaches); and cell-cell interactions, where solid phase synthesis, microarray analysis, nanotechnology, and synthetic biology play a part. In synthetic biology, for example, Ben Davis's group at Oxford, do interesting work making artificial cells.

Elizabeth has a particular interest in integrated systems from glycosylation of proteins and lipids [1]. The chemistry of mammalian cell membranes involves protein N-glycosylation (Asn), protein O-glycosylation (Ser/Thr), glycolipids, glycosylphosphatidylinositol (GPI)-anchored (glyco)proteins (prion protein is usually bound to membranes by a GPI anchor) and glycosaminoglycans (PG GAGs). These are usually studied by NMR (or mass spectrometry) rather than by X-ray crystallography. Tools to data mine the NMR structures became available from 1984, when Elizabeth produced the first program to do high resolution proton NMR of oligosaccharides [2].

A protein with an oligosaccharide attached sits on a cell surface. Elizabeth showed a 3D diagram of the C-terminal part of a prion structure. The protein structure was simplified as ribbons, and two carbohydrate structures formed "rabbit ears" on it. The carbohydrate is as important as the protein. The design of vaccines also involves looking at both carbohydrate and protein. Elizabeth's next structure was even more complex: a GPI with a protein and three glycans anchored to a cell membrane. The intricate details of these stunning structures are worth examining at <http://sites.google.com/site/reaxysminisymposiumuk/>. Other highlights of Elizabeth's research over the years are microbial glycosylation [3-5] and O-glucosaminylation of mammalian cytoplasmic proteins involved in diabetes [6-7].

USE OF REAXYS IN SYNTHETIC CHEMISTRY RESEARCH

Dr. Jonathan Goodman, University of Cambridge

Jonathan first explained why Google fails when used in searching chemistry. A search on "galactose" produces lots of hits but glucose is a high-ranking false drop. Entering an IUPAC Chemical Identifier (InChI) into Bing or Google is not always successful since search engines may truncate the string. Even the InChIKey is not ideal: the odds of two structures having the same key are very slim indeed but it could happen.

So Jonathan tried Reaxys. At the time of his talk he had only limited experience of the software but he has since written a detailed review for the *Journal of Chemical Information and Modeling*. The welcoming Reaxys Web page is an attractive form, with only a few buttons, including "Help", although the intuitive design tempted Jonathan to start experimenting without consulting the documentation. Fortunately, the interface allows this to be a fairly effective strategy.

Drawing a very complex molecule has little appeal but Reaxys has a facility for generating a structure from a chemical name. This failed when Jonathan tried "dolabriferol". A correct structure *was* generated in the case of "discodermolide", although all stereochemistry was omitted. Jonathan could have easily added stereochemistry, restricted the molecule to a particular role in a reaction, or searched using the molecule as a substructure etc., but he chose to use the stereochemistry-free structure.

Discodermolide was put into clinical trials as an anti-cancer agent. The only way to get a sufficient quantity of the compound for the trials was by total synthesis, despite the complexity of the molecule. Jonathan already knew that there are published articles and patents covering the synthesis of discodermolide, so this search was a good test for Reaxys. Clicking on "Search" produced a list of "Reaxys-ordered" citations describing the synthesis. The papers at the top of the list were all interesting, but the order of the papers can be changed if necessary.

Structures of the reactant and product molecules appeared with each citation, with buttons underneath them. Clicking on a button gave a menu with options for finding out more about the molecule illustrated, and also an option to "Plan a synthesis". Selecting "Plan a synthesis" produced a scheme with the final synthetic step of the sequence leading to discodermolide from the selected paper, and the option to "synthesise" the final intermediate. Choosing that option produced papers reporting synthesis of the final intermediate, any of which could be added to the scheme. After tracing back the synthesis for a few more steps, Jonathan had generated a scheme that showed the final part of a synthesis of discodermolide, with some steps taken from a patent, and others from different papers. This synthesis might be different from all the synthetic routes which have been reported, as it selected the reactions for different steps from different sources, but every step of this synthesis has been carried out, according to the literature abstracted into the Reaxys databases.

The facility to move easily between patents and journals and to generate schemes using reactions from different papers was impressive. The synthetic chemistry literature contains

very few citations of patents, despite the large amount of chemistry contained in the patent literature. Reaxys makes searching patents as easy as searching the scientific literature and produces clearly presented graphical summaries of the key synthetic transformations. The option to switch easily between patents and journals facilitates comparison, and may lead to an increased use of reactions from patents in some cases.

Following the discodermolide synthetic scheme further backwards eventually led Jonathan to molecules which can be purchased: icons under structures give links to the Symyx Available Chemicals Directory and to the free resource eMolecules. Articles cited in the synthesis are linked to the Scopus database and to full text if you have the required subscription. Maybe a better total synthesis of discodermolide could be devised using Reaxys.

Reactions with the same reaction scheme, but different reaction conditions extracted from different publications (patents and journals) are merged into just one reaction profile. The reaction procedure text from patent publications is provided immediately: there is no need to follow a link. Reaxys can gather large amounts of data quickly, and structure them in an orderly and accessible way, making data from journals and patents equally easy to assimilate.

One of Jonathan's research interests is molecules that cannot possibly be made. He used the Java-based structure-drawing option in Reaxys and was able to draw structures without referring to the help pages. He quickly sketched tetra-tertiary-butylmethane, a molecule that it is not possible to synthesise [8]. He was surprised to find two literature references and to see that the database included physical data on this improbable compound, but he soon found that the database contained no synthesis of the molecule and that the physical data were calculated not measured. Jonathan was not sure why the molecule has a CAS Registry Number.

He first tested text search by looking for his own name but was disappointed to find that recent papers of his are missing. The good news was that the chemistry described in a paper that *was* already in Reaxys was accurately and clearly transcribed into a scheme. There were buttons under all the molecules for finding more information, and for suggesting syntheses and suppliers. Most of this information was not in the paper itself. Thus Reaxys adds value through linking.

Piperidine was mentioned in the paper but not characterised as it is a standard reagent. Reaxys gave information on suppliers and also alternative names and physical properties of piperidine, including a proton NMR spectrum. One patent gave erroneous NMR data but Jonathan would have been able to link to other literature and patents to check this.

The ability to compare physical data from different sources is particularly useful for properties such as solubility, where different values for the same molecule are often reported. Jonathan searched Reaxys for diclofenac, a widely used pharmaceutical compound, and retrieved a list with a wide range of solubility values. Reaxys displays these results in a table of the different experimental values, each with a reference and a link to the original paper. "If only this time-saving tool had been available in the past", remarked

Jonathan. Reaxys makes it clear at a glance that a careful investigation is needed because the literature records many different values.

The Reaxys user interface is more attractive than the CrossFire Commander one and it has proved popular with Cambridge researchers during a trial. It also has a useful workflow tool that shows the various steps carried out in a search procedure. The key message that Jonathan imparts in his course on chemical information is that it is not enough to use just one database. In practice, it is hard to persuade people to use more than one search tool. Crossfire was not easy to “sell”; Reaxys has much more appeal. Part of the attraction to researchers in Cambridge has been the unlimited number of simultaneous users who can use the system; there is no frustrating “all licences in use” message. Reaxys is a helpful and user-friendly tool which will make multiple searches much easier and much more attractive, and so will lead to the more effective use of the literature. A combination of Reaxys and another product would be ideal but Reaxys is good value for money right now. Jonathan quipped that he could not speak for the next 200 years: Cambridge takes a long term view of things.

INNOVATIONS IN CHEMISTRY: AN ELSEVIER PERSPECTIVE

Karel Nederveen, Publishing Director Chemistry, Elsevier

As the world’s leading publisher of science and health information, Elsevier serves more than 30 million scientists, students, and health and information professionals worldwide. Scirus is the most comprehensive science-specific search engine on the Internet. Driven by the latest search engine technology, Scirus searches over 350 million science-specific Web pages.

ScienceDirect is a leading full-text scientific database offering articles or chapters from more than 2,500 peer-reviewed journals and more than 11,000 books. It currently contains more than 9.5 million articles or chapters, and more than 500 million full text articles are downloaded in a year. Elsevier serves more than 500 scientific societies and has a global scholarly community of 7,000 journal editors, 70,000 editorial board members, 300,000 reviewers, and 600,000 authors, publishing 2,000 journals and 19,000 books. The company’s main offices are in Amsterdam, Oxford and New York.

A 2009 STM report [9] discusses formal and informal types of scholarly communication. Karel reproduced a table from this report comparing oral and written modes of communications and characteristics of old and new instances of these modes. Will Web 2.0 have an influence on the functions of the learned journal? The four functions in question are registration, certification, archiving and dissemination. Web 2.0 will not change these core functions; peer reviewed content will still deliver them. There have been many recent changes but authors’ core motivations still depend on the four functions. Publishers will continue to facilitate scholarly communication and enhance the value of articles by offering peer review, publication, quality control, and visibility.

Nevertheless, Elsevier has been studying new features of the article of the future. At a fairly simple level, these include graphical abstracts and article highlights. Users of graphical

abstracts are becoming more creative. Karel showed a typical screen that could be built with what is really just a repackaging of an article from *Cell*. The building blocks of the article are on tabs in the display; figures are one tab. A gallery of figures can be “moused over”. Elsevier has sought feedback on these features and the 500 reactions to date suggest that this presentation is much better than the current version. The new features were thus rolled out.

Another possibility is semantic enrichment: what Karel described as Natural Language Processing at Internet scale. Elsevier already has one tool, Illumin8, designed for corporate R&D professionals. Following the recent launch of the initial “Article of the Future” prototype with *Cell*, where the traditional linear journal article is displayed in a much more useful format for life scientists, the *Cell*-Reflect pilot is the next step in Elsevier’s ongoing content innovation effort with the scientific community to determine how a scientific article is best presented online. Reflect will be piloted on the research articles in the November 12 issue of *Cell*. It identifies the proteins, genes, and small molecules mentioned in *Cell* articles. It also generates pop-up windows containing relevant contextual information about those entities. Reflect was initially developed at the European Molecular Biology Laboratory, Heidelberg, Germany and earlier this year won Elsevier’s Grand Challenge 2009. Using this sort of technology, ScienceDirect could be linked to molecules in Reaxys, but this is only at the conceptual stage right now.

Elsevier is also working on videos and mobile technologies. Authors can submit videos; about 30,000 videos on ScienceDirect are being extracted and streamed. Apple itablets may be the way of the future. Karel showed a biochemistry structural digital abstract. IUPAC International Chemical Identifiers (InChIs) are used in structural digital abstracts; it would be good to do something similar for crystallography data. Elsevier works with NIST to get thermodynamic data checked before the peer review process. Extended use of graphical abstracts could be seen as a poor man’s version of the article of the future, but the building blocks are in place for the real article of the future, with article enhancement and contextual linking to external content.

ENHANCING EFFICIENCY IN CHEMISTRY RESEARCH WORKFLOW

Dr. David Evans, Director Scientific Affairs, Elsevier Properties

Reaxys is a new workflow tool for researchers in chemistry and related sciences. It is an extensive repository of chemical properties and reaction data, and a resource for accurate and validated experimental data, designed to support the optimisation of synthetic processes and give high quality answers scientists can use with confidence. It has a simple intuitive interface, with analysis and planning tools that help chemists to plan the synthesis of target compounds. It is a Web-based product (<http://www.reaxys.com>); no installation is required.

If you wanted to know what is known about a compound, and how to make it, you could carry out a search in a bibliographic chemical database. Then you would have to follow up by reading each individual article in full text. Instead, Reaxys immediately provides factual and actionable answers. It was developed in collaboration with more than 20 development

partners and was built from deeply indexed data using an “agile” approach. Elsevier’s User-Centered Design group studied why users wanted to do each process and why they wanted to do it in a particular way.

The Reaxys database is a combination of the CrossFire databases (Beilstein for organic chemistry, Gmelin for inorganic chemistry, and patent chemistry, that is mainly organic), all of which concentrate on recording structures and their properties, and structures and their reactions. The database supports the scientific method in which verifiable experimental observations are generalised into hypotheses for ongoing modification in an iterative process.

Reaxys is navigated in structure/property/reaction space. For any given chemical structure, the scientist can discover products that could be made from that structure, precursors from which the structure could be made, intrinsic and extrinsic properties of that molecule, and related structures and properties. David illustrated this using Salbutamol as the target structure.

He listed some typical questions a chemist might ask. For example, how can Salbutamol be synthesised? What are the process details, yields and patents? How do I to buy or synthesise the starting materials? Is there supplier data? What about reactivity? How does Salbutamol react with other substances to produce new compounds? Are there spectral or physical data to identify the compound? Which bioactivity and application data does it have? Which analogous compounds exist, and how do their properties compare to those of Salbutamol? Which key publications should I read?

David ran through a worked example, to show how Reaxys answers those questions. He chose the “Reactions” tab, entered the name Salbutamol, produced a structure from it, and searched for the structure “as drawn”, removing salts, mixtures, etc. from the hit set by checking boxes. He clicked on “View results”, to see 51 reactions drawn from 38 publications. On the left hand side of the screen there is a set of filtering options. David chose to limit the number of reaction steps to five.

He saw a route starting with salicylaldehyde and by clicking on a “context” icon under a structure was able to use the synthesis planner. Now a pathway to Salbutamol was shown, including the reaction conditions for each step. Clicking on “Modify” under a step leads to alternative methods for making the intermediate. Information on where to buy it is also available.

Next David did his original Salbutamol search again (he had saved the query) in the “Substance and Properties” tab. Zooming in on any specific property in the resultant list produces a tabular display of values for that property together with literature references and details of how the property was measured. David then “zoomed back” to a substance overview, with a spreadsheet of structures and related data.

He clicked on “6 prep out of 7 reactions” in the “Number of preparations” column to zoom in to a source document (a journal). An article title showed that this research concerned asymmetric reduction with a certain catalyst. Clicking on “Scopus” to zoom in would lead to

“cited by” information, or full text from ScienceDirect, and also patent details. Zooming in to Scopus allows other relevant papers and reviews to be found. David also displayed details of a relevant patent and showed a click through to the full text of the patent on EspaceNet. Below the spreadsheet part of the earlier display are sections from which the user can zoom in to identification data such as the IR spectrum, and to bioactivity and application information. The original list of typical questions had by now been fully answered.

CLOSING REMARKS AND DISCUSSION

There was a useful discussion session, the interest of which was heightened by the fact that there was one attendee from a major pharmaceutical company among the largely academic audience. It was clear that requirements for patent information differ between the two communities. Much of the discussion centred on Reaxys content and currency. Elsevier confirmed that it continuously reviews the journals and patents that should be indexed.

There was discussion on the complementarity of factual and bibliographic databases: one must not try to compare chalk with cheese. Nevertheless the pharmaceutical end users tend to use only one tool. Very few will use two tools and users vary as to which one they prefer. Reaxys is seen to be very intuitive compared with CrossFire with the Commander interface, and the graphics displays are much better. An IT manager at a major university also said how much the chemists at his site like Reaxys: it is reported to be much more usable than CrossFire, and the Web-based approach is appreciated.

REFERENCES

1. Hounsell, E.F. NMR of Carbohydrates Lipids and Membranes. *RSC Specialist Periodical Reports in NMR*. **2009**, 38; **2008**, 37, 274-304; **2007**, 36 (and volumes 29-35).
2. Hounsell, E.F.; Wright, D.J.; Donald, A.S.R.; Feeney, J. A computerised approach to the analysis of oligosaccharide structure by high resolution proton NMR. *Biochem. J.* **1984**, 223, 129-143.
3. Hounsell, E.F. Glycoprotein Analysis in Biomedicine. Humana Press: Totowa NJ, 1993.
4. Rangarajan, M.; Hashim, A.; Aduse-Opoku, J.; Paramonov, N.; Hounsell, E. F.; Curtis M. A. Expression of Arg-Gingipain RgpB is required for correct glycosylation and stability of monomeric Arg-gingipain RgpA from *Porphyromonas gingivalis* W50. *Infection and Immunity* **2005**, 73(8), 4864-4878.
5. Young, M., Davies, M.J., Bailey, D., Gradwell M.J. and Hounsell, E.F. Characterisation of oligosaccharides from an antigenic mannan of *Saccharomyces cerevisiae*. *Glycoconjugate J.* **1998**, 15, 815-822.
6. King, I.A.; Hounsell, E.F. Cytokeratin 13 contains O-glycosidically linked N-acetylglucosamine residues. *J. Biol. Chem.* **1989**, 264, 14022-14028
7. Hashmi, F.; Malone-Lee, J; Hounsell, E.F. Plantar skin in type II diabetes: an investigation of protein glycation and biomechanical properties of plantar epidermis. *European Journal of Dermatology* **2006**, 16(1), 23-32.
8. de Silva; K. M. N.; Goodman J. M. What Is the Smallest Saturated Acyclic Alkane that Cannot Be Made? *J. Chem. Inf. Model.* **2005**, 45, 81-87; Paton R. S.; Goodman, J. M.

Exploration of the Accessible Chemical Space of Acyclic Alkanes. *J. Chem. Inf. Model.* **2007**, *47*, 2124-2132.

9. Ware, M. Mabe, M. The STM report: an overview of scientific and scholarly journals publishing. STM, 2009. [http://www.stm-
assoc.org/news.php?id=255&PHPSESSID=3c5575d0663c0e04a4600d7f04afe91f](http://www.stm-
assoc.org/news.php?id=255&PHPSESSID=3c5575d0663c0e04a4600d7f04afe91f)
(accessed November 17, 2009)