COMPUTER SOFTWARE REVIEWS

Reaxys. . Elsevier Properties SA, 360 Park Avenue South, New York, NY 10010-1710. www.info.reaxys.com

Chemical knowledge is hard to extract from the vast quantity of molecular data that is stored in journals, patents, and databases. Finding and comparing relevant information from different sources is a time-consuming and challenging part of the process of developing chemical syntheses. Errors and omissions are hard or impossible to eliminate altogether. Reaxys is a new resource which provides a user-friendly method to search and to analyze data taken from three key databases: Beilstein, Gmelin, and the Patent Chemistry Database. Reaxys has been constructed with synthesis design particularly in mind.

Reaxys is a web-based system, so no specific software installation is required. I opened the webpage: //https:www.reaxys.com/ and was immediately able to access the resource. Access is controlled by IP address, so I did not have to go through a log-in procedure because my institution has trial access to Reaxys. There is also an option to log in with a user name and password, if you are using a computer which does not have a network address from a subscribing organization. The welcoming webpage is an attractive form with only a few buttons, including "help". However, the straightforward design tempted me to start experimenting without consulting the documentation. Fortunately, the interface is intuitive enough for this to be quite an effective strategy.

There is a window for drawing molecules, but the first button that attracted my attention was the one labeled "generate structure from name". Clicking on this allowed me to type a chemical name, a SMILES string, an InChIKey, or a CAS registry number, and possibly other chemical identifiers. The name is compared with a database, and the molecular structure is generated automatically, provided there is a match. My first attempts, dolabriferol and brasilinolide, did not produce any structures, even though they are both reported in a number of papers in chemical literature. I searched Beilstein using DiscoveryGate and found that brasilinolide is listed in this database.

My third attempt with the "generate structure from name" option was discodermolide, and this generated the correct structure for discodermolide, although all stereochemistry, including double-bond geometries, had been omitted. This process is a very convenient and quick way of generating complex molecules that can then be edited by the built-in structure editor. In this case, I used the stereochemistry-free structure, which had been automatically generated as the search query, but I could have easily added stereochemistry, restricted the molecule to a particular role in a reaction, searched using the molecule as a substructure, or used any of many other options.

Discodermolide was put into clinical trials as an anticancer agent. The only way to get a sufficient quantity of the compound for the trials was by total synthesis, despite the great complexity of the molecule. I knew, before running the search, that there are many papers about the synthesis of this molecule, and also a number of patents. Would Reaxys be able to cope with this? Could it find the large amount of information, and could it present the complex data in a way that is both orderly and helpful?

Clicking on the "search" produced a list of citations to papers describing the synthesis, ordered by "Reaxys ranking". The algorithm used for this is not clear to me, but the ones at the top of the list were all interesting papers. There are options to change the order of the papers, if required. Small pictures of the molecules appeared with each citation, with buttons beneath them. Clicking on a button gave a menu with a series of options to find out more about the molecule illustrated, and there was also an option to "plan a synthesis". Selecting this produced a scheme with the final synthetic step of the sequence leading to discodermolide from the selected paper and with the option to "synthesize" the final intermediate. Clicking this button produced a series of papers which had made the final intermediate, any of which could be added to the scheme. After tracing back the synthesis for a few more steps, I 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. However, every step of this synthesis has been carried out, according to the literature abstracted into the Reaxys databases (see Figure 1).

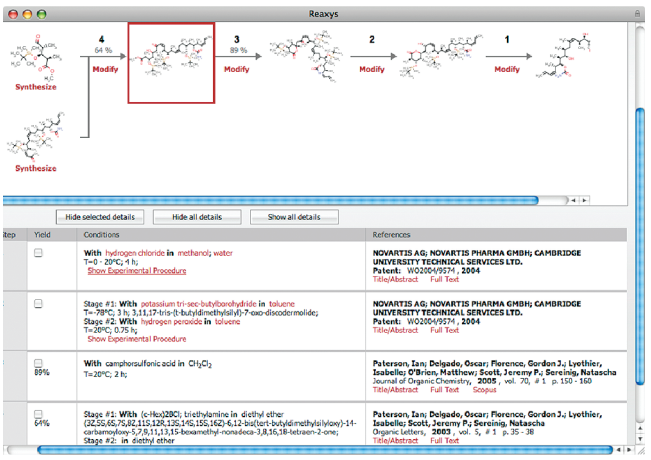

Figure 1. An example from the Reaxys database.

The facility to move easily between patents and journals and to generate schemes using reactions from different papers was extremely impressive and will be a great help for synthetic chemists. The synthetic chemistry literature contains very few citations of patents, despite the large amount of chemistry contained in the patent literature. This may be due to chemists' reluctance to read patents because they are not always written with the primary aim of clear communication of the synthetic information or because of a tradition of focusing on journal articles, which have to get through peer-review by experts in synthesis. Reaxys makes searching the patent literature as easy as searching journals and produces clearly presented graphical summaries of the key synthetic transformations. The option to switch easily between patents and journals should mean that reactions from patents can be easily compared with reactions from journals, leading to an increased use of reactions from patents in those cases where patented procedures work better than reactions in journals.

Following the synthetic scheme further backward eventually leads to molecules which can be purchased. These are indicated by including links to the Symyx Available Chemicals Directory and to the free resource eMolecules. My institution currently has a subscription to the first of these, taking me to a different company's Web site with its different access requirements. The articles that are cited in the synthesis are linked into the Scopus database, and clicking on the *Scopus* link takes me directly to the article's entry in this database, to which we also have a current subscription. The power of Reaxys is enhanced, therefore, if parallel subscriptions to these other resources are available.

I now tried out the structure-drawing option, which is Java-based and works in an intuitive way. I was able to draw a structure without referring to the help pages, and I quickly sketched tetra-tertiary-butylmethane, a molecule that it is not possible to synthesize.[1] I was slightly concerned that the search produced the molecule and that the database included physical data on it, as the database should contain only validated experimental data. However, it became clear both that no synthesis of the molecule was in the database and that the physical data came from calculations and not from measurements. The molecule has a CAS Registry Number, which refers to an abstract of a dissertation that was not indexed in Beilstein.

10.1021/ci900437n
Published on Web 12/04/2009

As well as searching for structures, it is also possible to search for text using the "text, authors, and more" tab. I searched for my own name (purely because I know what I have published, of course). As I typed the name, suggestions for completing the text came up very quickly, which would have been very useful had I been unsure of the spelling or initials. I was disappointed to discover I have only published two papers so far this year according to Reaxys, and this probably reflects the rate at which the database is updated. I explored the data available on one of these papers and found the chemistry described in the paper had been accurately and clearly transcribed into a scheme. All of the molecules in the scheme had buttons for finding more information and for suggesting syntheses and suppliers. Most of this information was not in the paper, and this demonstrates how Reaxys can link data from disparate sources into a convenient and accessible form.

I asked for more data on piperidine, a compound that was mentioned in the paper but not characterized, as it is a standard reagent. Reaxys gave information on suppliers via its link to the available chemicals directory and also to alternative names and physical properties of the molecule. A lot of property information was available by following the links within Reaxys, including the ^{1}H NMR spectra. The ^{1}H NMR information that I saw first had been gathered from a patent that appeared to have misreported the information. Links would have taken me to the original patent or to other papers to check this.

The ability to compare physical data from different sources is particularly useful for properties that are hard to measure unambiguously. For example, solubility is a property for which different papers often report different values for the same molecule. Diclofenac is a widely used pharmaceutical compound which has been very well characterized. A search for diclofenac gave access to its physical properties and gave a list of solubility values, each with a reference showing a wide variation. Without a search tool like Reaxys it would be tempting to use the first solubility value which was encountered. Reaxys makes it clear at a glance that the literature records many different values, and so a careful investigation is needed. Reaxys also produces a neat table of the different experimental values, each with a reference and a link to the original paper, if available. This means that the fiddly and time-consuming job of literature searching becomes almost trivial.

The database used by Reaxys is large but not comprehensive. Web of Knowledge and SciFinder seem to be more up to date than this database, but Reaxys does contain quite a lot of information from 2009. The name-to-structure facility is useful, but does not index all of the compounds available to Reaxys. Name-to-structure could not find a structure for brasilinolide, but experimentation with the "advanced search" box on the Reaxys home page showed that brasilinolide is present in the Reaxys database, once I had worked out that I needed to type: " BI.BISUB = 'brasilinolide' " and not just "brasilinolide. I got this code from the "show fields and operators" button, but use of the more advanced features would probably benefit from consulting the help files. These are clearly set out with training materials and webinars as well as frequently asked questions.

The user interface is attractive and has been popular with researchers during the trial. Part of the attraction to researchers in our department has been the unlimited number of simultaneous users that can use the system. Unlike some other products, there was no frustrating "all licenses in use" message. Even though there was no log-in procedure for me, I eventually got a message: "Your Reaxys session has reached a timeout and has been shut down automatically. Please note: all unsaved results of the closed session have been lost." This would have been irritating had I been part way through a complex analysis, but it would have taught me to use the "save query" facility.

In an ideal world, every project should begin with a literature survey that would be run through several different databases and search engines in order to be as certain as possible that all relevant information has been collected. In practice, I have found it hard to persuade people to use more than one search method. The availability of this helpful and user-friendly tool will make multiple searches much easier and more attractive and will lead to the more effective use of the literature. Reaxys is a powerful tool in the armory of structure-based literature searching. It 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. It will be very useful both for teaching and research and deserves to be widely used.

REFERENCES AND NOTES

(1) (a) 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.

Jonathan Goodman*

Unilever Centre for Molecular Science Informatics, Department of Chemistry, Lensfield Road, Cambridge CB2 1EW

CI900437N

10.1021/ci900437n