TONY DAVIES COLUMN

eMolecules and spectroscopy

A.N. Davies

External Professor, University of Glamorgan, UK, Director, ALIS Ltd, Analytical Laboratory Informatics Solutions

Background

As any regular reader of this column will know I have been campaigning regularly over the years for better public access to reference quality data. So when my attention was drawn to the following statement on an American website it was obvious that their work would need further investigation!

eMolecules® is the leading open-access chemistry search engine. eMolecules' mission is to discover, curate and index all of the public chemical information in the world, and make it available to the public for free.[1]

Reading further it became clear that the core of the system was chemical structure oriented, but that they also claim to have access to spectra... so the investigation began!

The system is essentially a clearing-house for metadata and chemical structure metadata indexing of numerous third-party chemical information providers. The list of linked systems accessible from the eMolecules search result pages is obviously dominated by the chemistry manufacturers, with over 150 of the biggest providing their data to the search eMolecules engine.

History

eMolecules is headed by four entrepreneurs who launched a new web-based search engine called Chmoogle back in November 2005. They all have strong backgrounds in chemical informatics. Klaus Gubernator, is eMolecules founder, president and CEO. Craig A. James is a computer scientist and is Chief Technology Officer and cofounder. Rashmi Mistry is VP of Software Development and Jon M. Van Winkle is Vice President of Sales and Marketing at eMolecules.

eMolecules in action

When you login you are immediately presented with your previous search result so if you have to break off your work in the middle to attend to some minor crisis (a situation which is becoming ever more prevalent in the modern scientists life!) you can return later without losing your work. But if you are starting a new search the system does something unexpected and presents you with 5 million plus answers before you start. This doesn't effect the way the system works but did worry me at the start as I couldn't work out what I had done to get so many hits!

In search of my badly needed paracetamol, I decided to reduce the hit list by use of the CAS registry number 103–90–2 and see what suppliers would be available.

This quickly reduced the number of compounds on offer (of course actually to one!), but with a large list of potential vendors and different sources for information on the suppliers. The hit list as shown on the screen is also active and all the cells in the hit list are starting

points for further work. Where concrete links (blue underlined) exist they will navigate you away from the eMolecules website, other cells will start up the structure/substructure search window, text searching etc. (Figure 1).

However, if I want to see any structural analogues currently on the market—by clicking in the box where my paracetamol structure is displayed I start up the structure search window with the structure already loaded. This then allows me to find other compounds, which, for example, may contain this as a substructure.

eMolecules and CSEARCH

I may have found the solution to my current headache—how can this help my spectroscopy problem? Although only in the prototype stages there are a couple of spectroscopy services being trialled through the eMolecules system.

By switching to nmr.emolecules.com you are presented with the same front-end, but with an NMR spectroscopy slant!

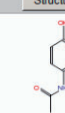Unfortunately, at the moment my user id doesn't retain its active searches

Figure 1. Once you have a hit list you can drill down or across, as all the cells in the display are also active. So if I want to look at published information on the compound and like the PubChem initiative I can click on the active PubChem link to be directed straight to the appropriate page on the NCBI website.

between the subsystems, so if I need an NMR spectrum of my paracetamol (well I always carry out an analysis before I swallow, don't you?) I will have to start again.

Again the Start Search gives me the entire database to browse through. And now we can see that this particular prototype has 140k structures. Repeating my quest for paracetamol I now get three hits. The active links take me to a "lite" version of the well-known CSEARCH NMR database of Wolfgang Robien at the University of Vienna, Austria. Interestingly I get two ^{13}C NMR spectra to review and one [17]O!

eMolecules and ChemGate

The second system developed by Dirk Hermanns for Wiley uses the SpecInfo database which we have featured in this column in the past.[2]

To access this prototype I again have to enter through a different portal called chemgate.emolecules.com—by now I hope you are starting to see the pattern emerging!

Again my profile, although storing my previous work on this sub-domain, doesn't bring results across from other sub domains.

Starting a new search here provides me with the usual listing of the entire database by structure—in this prototype just over 92,000 structures. Just to ruin the red thread running through this article paracetamol isn't present in the database so I reverted to an older remedy.

The advantage of having all these different systems available through the one interface is that as soon as you have picked up the use of the tools it doesn't matter which you switch to—you are immediately able to get your job done!

Aspirin provided a single hit and four different ^{13}C NMR spectra to view. So taking any example of a $C_{13}H_{18}O_2$ compound, (I had been trying for Ibuprofen this time) and clicking the active link I was redirected to the Wiley ChemGate interface.

The first time around the extra Chemgate login page defeats the transfer of my hit list information. Once logged in to Chemgate, clicking the link from
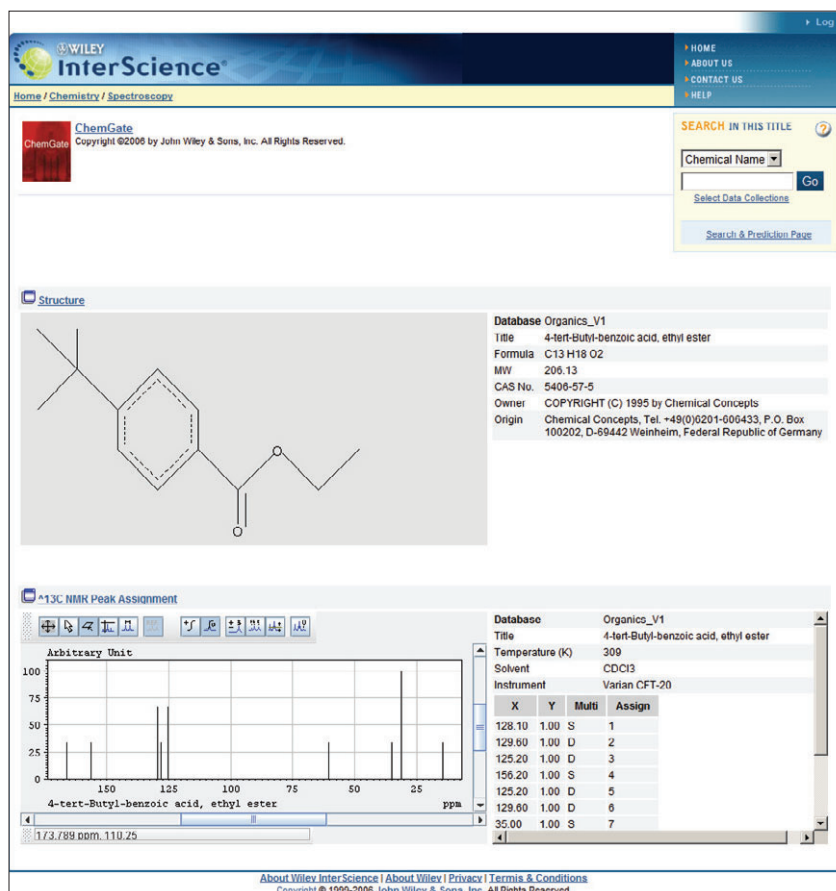

Figure 2.

the eMolecules website provided me with a number of options for reviewing the spectroscopic data available in the ChemGate prototype. Figure 2 is the display you receive when you highlight the peak assignments.

ACD/Labs ChemSketch

Finally, if you are one of the over 700,000 users of ACD/labs ChemSketch software you may not be aware of a feature allowing you to search directly on the eMolecules site from ChemSketch.

Hidden up on the top right hand side of the ChemSketch interface are
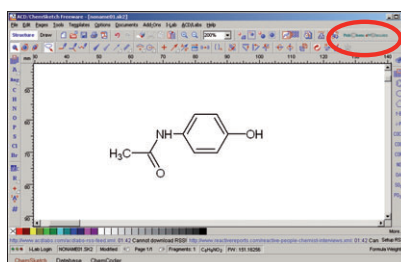
PubChem and eMolecules logos. Clicking here brings up an option box. I found I had to be rather selective, as at the time of trying only eMolecules link to Browser view actually worked but then you get a somewhat different display—probably because you are not logged in and therefore have no preferences active.

Conclusion

In May 2006 the name Chmoogle morphed into eMolecules, the company name emphasising its unique identity. I can only applaud the ease of use of this resource, and the fact that the same tools migrate from domain to domain makes life very simple. Yes, it is a commercial website which will survive through the amount of traffic it can generate, but the functionality available already—only some 14 months from the original launch—bodes well for the future.

References
1. http://www.emolecules.com