

Chemical informatics: where are we headed?

Robert Glen

Unilever Centre for Molecular Sciences Informatics,
University Chemical Laboratory,
Lensfield Road, Cambridge, CB2 1EW, UK

Chemistry has a language that is in some ways quite formal and systematic. However, like most languages, there is a degree of 'artistic licence' and converting this into a computer representation is challenging, but allows a new way of doing science, knowledge discovery. In earlier years, we have been to some extent limited in that we have applied the standard scientific method of hypothesis and experiment followed by acceptance or revision of the hypothesis (QSAR is an excellent example). There are of course many examples of successes in drug discover or property prediction using this approach. However, with large amounts of data, the internet and cheminformatics tools, science can be performed in a new way in which hypotheses are generated from relationships between the data, sometimes revealing unexpected results. Indeed, serendipity could potentially be systematised. Although we are not there yet, I believe that this is where we are headed. I will present some examples of how we are describing chemical structures, how these methods can be used in lead discovery and also some examples using recent applications designed to both improve the quality of data, and also to discover new relationships between cheminformatics data.