

## Cheminformatics Education

### Introduction

Despite some skepticism at the turn of the century<sup>1,2</sup> the terms “cheminformatics” and “chemoinformatics” are now in common parlance. The term “chemical informatics” is used less. The premier journal in the field, the *Journal of Chemical Information and Modeling*, does not use any of these terms and its cheminformatics papers are scattered across multiple sections including “chemical information”.

Unfortunately it is not just the name of the discipline that is undefined: opinions also vary on the scope. Paris supplied the following definition<sup>1</sup> (shortened here): “Chem(o)informatics is a generic term that encompasses the design, creation, organization, storage, management, retrieval, analysis, dissemination, visualization and use of chemical information, not only in its own right, but as a surrogate or index for other data, information and knowledge”. Brown<sup>3</sup> says the discipline is “mixing of information resources to transform data into information, and information into knowledge, for the intended purpose of making better decisions faster in the arena of drug lead identification and optimization”.

Gasteiger<sup>4</sup> says that cheminformatics is the application of informatics methods to solve chemical problems and Bajorath<sup>5</sup> agrees that a broad definition such as that is needed to cover all the different scientific activities that have evolved, or have been assimilated, under the cheminformatics umbrella. Varnek’s definition<sup>6,7</sup> is rather different. He considers that cheminformatics is a part of theoretical chemistry based on its own molecular model; unlike quantum chemistry in which molecules are represented as ensemble of electrons and nuclei, or force fields molecular modeling dealing with classical “atoms” and “bonds”, cheminformatics considers molecules as objects (graphs and vectors) in multidimensional chemical space.

The Web site of Michigan Technological University (“Michigan Tech”)<sup>8</sup> states: “Cheminformatics integrates a comprehensive knowledge of chemistry with an extensive understanding of information technology. The intersection of chemistry and information technology embraces an expanding territory; it includes computational modeling of individual molecules, thermodynamic methods of estimating chemical properties, methods of predicting biological activity of hypothetical compounds, and organization and classification of chemical information”. In addition, Bajorath<sup>5</sup> notes that “Despite their conceptual diversity, many developments in this discipline continue to be driven in pharmaceutical environments. In fact, without its strong foundation in drug discovery, chemoinformatics as we understand it today would probably not be the same”.

The field of cheminformatics is well established as a discipline in its own right, and its importance is clear to all its practitioners whatever their definition of the discipline, but it is not still recognized as a scientific field by the ministries of education and other governing academic organizations in most European countries. Academic positions in cheminformatics are rarely advertised which can lead to loss of continuity and expertise for cheminformatics research groups. PhD fellowships and post doctoral funding are lacking and very few programs are funded at the national and European levels.<sup>6</sup>

In a recent editorial, Bajorath<sup>5</sup> wrote: “In contrast to bioinformatics, chemoinformatics is still fairly under-represented in academic research and teaching, which might cause problems for the further development of this field. Fortunately, it is increasingly being recognized that the methodological arsenals of bioinformatics and chemoinformatics are much more similar to each other than the diversity of their applications might suggest. Consequently, boundaries between these fields are becoming fluid and we can expect that interest in the computational study of small molecules and ligand-target interactions will further increase, also in the bioinformatics community.”

The National Science Foundation in the United States funds cyberinfrastructure, computational science, interdisciplinary research and bioinformatics; some workers have realized that claiming to be in those disciplines is useful when making grant applications for cheminformatics research. It is, however, generally agreed that there is a need for more teaching programs that integrate chemistry and computer science. A declaration<sup>9</sup> about the need for cheminformatics education was approved at Obernai, France in 2006.

### **Undergraduate courses**

One of the ways to raise its public awareness is to integrate cheminformatics into chemistry curricula. At the Louis Pasteur University of Strasbourg, cheminformatics is taught as an individual discipline at the undergraduate, masters and PhD levels.<sup>6,7</sup> An “introduction to cheminformatics” course is delivered to third year undergraduate students.

Research is the main focus of the Unilever Centre for Molecular Science Informatics at the University of Cambridge’s Department of Chemistry. However, the center also provides both undergraduate and postgraduate teaching in molecular informatics (“cheminformatics plus bioinformatics”).<sup>10</sup> The undergraduate course in molecular informatics is the most advanced course in the chemistry degree program that is taken by all the students in the class. At this stage, the students have already had the opportunity to specialize in particular areas of chemistry, and will soon move on to specific research projects. The center provides students with information about the molecular informatics tools available to them, as they move towards the research environment, and demonstrates how data they gather can be combined to provide knowledge that would be inaccessible without co-operative effort and suitable analysis.

The Applied Chemistry degree from Faculdade de Ciências e Tecnologia (FCT) at the Universidade Nova de Lisboa, Portugal is now an EU first-cycle “Bologna degree”, recently certified by the European Chemistry Thematic Network with the Eurobachelor label. An undergraduate introductory cheminformatics module is part of the third year program of the Applied Chemistry degree. Learning cheminformatics reinforces basic chemical concepts (e.g. stereochemistry), develops the ability for multidisciplinary approaches (e.g., in QSAR studies), and promotes transferable competences related to data management.<sup>11</sup>

At the University of Erlangen-Nuremberg, Germany a study course for a bachelors degree in Molecular Science was established a few years ago. In this course, the students are required to take classes in cheminformatics during their studies to prepare them for their masters degree studies.

The Department of Information Studies at the University of Sheffield is better known for its research work and its masters programs but recently, its activities have been extended to include undergraduate teaching in a move to integrate cheminformatics teaching into mainstream chemistry teaching.<sup>12</sup> Undergraduates now have the option of studying for an MChem/BSc course in Chemistry with Computational Drug Design. This has been designed by the cheminformatics research group in collaboration with the Department of Chemistry. It is a basic chemistry degree with increasing cheminformatics content (replacing some inorganic chemistry); 50% of year four of the MChem involves a cheminformatics research project. The first students are expected in 2008. The level one module (introduction to cheminformatics) is already delivered as an elective on all chemistry degree programs and had 50 students in 2007/08.

The Department of Chemistry at Michigan Tech<sup>8</sup> offers [undergraduate](#) degrees in pharmaceutical chemistry, biochemistry and molecular biology, and cheminformatics, in addition to the basic chemistry degree.

### **Masters courses**

Undergraduate courses may familiarize students with the practical aspects of cheminformatics but the students are unlikely to be aware of many of the important applications in, for example, drug discovery, and they will not have many of the requisite IT skills. Multi-disciplinary programs involving informatics and chemistry are thus better offered at the masters level. Such courses can be offered to graduates in computer science with a good knowledge of chemistry as well as to chemists,<sup>13</sup> but at Sheffield and Strasbourg, chemistry, not computer science, is a prerequisite for the MSc.

The cheminformatics research group at the University of Sheffield has several decades of experience in training scientists to PhD level, many of whom now hold senior positions in the field throughout the world. In 2000, its teaching activities were extended to masters level provision<sup>12,13</sup> with the introduction of the world's first MSc in cheminformatics supported financially by the UK government for five years, and designed in collaboration with the U.K. pharmaceutical industry. The one-year course is divided into three semesters: two semesters of taught courses and a third semester spent on an industrially based dissertation project. The designers of the course were able to draw on existing modules already delivered at masters level (in information systems and information management) in the department.

Unfortunately the five year period is now over and the end of financial support from the UK government has led to a reduction in the number of applicants; studentships are no longer available. Changes in the funding of UK higher education mean that students are now faced with debt following undergraduate studies. Reduced student numbers mean that the previous structure for an MSc in cheminformatics is no longer effective. An MSc by research has a smaller taught component and a larger in-house research project: six months. Students join the cheminformatics research group at Sheffield.

Student supply and demand were well balanced during the period of government funding, but the end of funding has led to a decrease in recruitment and the future of the masters level training at Sheffield is currently uncertain, something that has met with considerable dismay by colleagues in industry. Two publications review developments in cheminformatics education.<sup>14,15</sup> Unfortunately the MSc course in Manchester, England<sup>16</sup> described in those papers has suffered from some of the same problems as the Sheffield course and has been discontinued.

A new one year MSc course at the University of York, England<sup>17</sup> focuses on the use of mathematics and statistics for real-world data analysis and interpretation. The program, which starts in October 2008, is said to be suitable for mathematics and computer science graduates with an interest in the application of mathematical and computational techniques to chemical problems, as well as mathematically inclined students with a background in chemistry, physics or biological science. The course is available with a flexible two-year part-time structure and Diploma and Certificate options are also available. The course is jointly run between the Mathematics and Chemistry departments.

Masters degree students in Molecular Science at the University of Erlangen-Nuremberg can specialize either in Drug Design or Materials Science. In both branches they have to take classes in theoretical chemistry that also encompass lectures and seminars in cheminformatics.

The School of Informatics at Indiana University was established in 2000. Indiana offers MSc and PhD courses in cheminformatics (but not undergraduate education). It is the only center in the United States offering formal qualifications in cheminformatics. Distance learning is available too and is surprisingly successful. Degrees available are an MS in chemical information, a PhD in informatics, with a specialism in cheminformatics, and a graduate certificate.<sup>18</sup> There is also a strong bioinformatics group at the university.

At the Louis Pasteur University of Strasbourg, a one-year masters program in chemical informatics was created in 2001. Since 2005 the course "master en cheminformatique" has been a two year program:<sup>6,7</sup> three semesters of formal instruction and hands-on tutorials followed by a

semester of industrial or academic training. There are 8-10 students a year. For admission a strong background in chemistry is required. With the MSc courses in cheminformatics at Sheffield and Manchester closing, the masters courses in Indiana and in Strasbourg will be the only remaining masters courses in the world in “cheminformatics” as such. Strasbourg organizes these courses for chemists, noting that Indiana organizes them for computer scientists.

The University of Massachusetts Lowell offers an MS in computer science with a bio/cheminformatics option.<sup>19</sup> Applicants for admission to the program typically have an undergraduate degree in computer science or a related discipline such as mathematics, physics, biochemistry or engineering.

The Beilstein Institute has funded an endowed chair for cheminformatics at Johann Wolfgang Goethe University, Frankfurt-am-Main, Germany; Prof. Dr. Gisbert Schneider is the professor for cheminformatics and bioinformatics. The university has a new MSc program in computational science<sup>20</sup> targeting outstanding students with a bachelors degree in biology, chemistry, geoscience or physics, but open also to exceptional students with a degree in mathematics, computer science or engineering who also have a solid background in one of the natural sciences.

This MSc program leads students towards the current state-of-the-art of research, both with respect to their specific field within the natural sciences and with respect to computational techniques. The program covers both the basic tools for numerical simulation, provided by applied mathematics and computer science, and the formulation of efficient, yet realistic models in the individual disciplines, which requires an understanding of the scales, interactions and correlations involved. The MSc in computational science is administered jointly by the departments of biochemistry, chemistry and pharmacy, geoscience/geography, physics and computer science and mathematics, together with the Frankfurt Institute for Advanced Studies and the Center for Scientific Computing. The working language of the entire MSc program is English, consistent with the international background of the students.

The center for bioinformatics at the University of Hamburg, Germany offers a four-semester course<sup>21</sup> leading to a masters degree in bioinformatics. It is designed for undergraduates in biology, chemistry, medicine, molecular life sciences, computer science and other related fields who are interested in an interdisciplinary specialization in bioinformatics or cheminformatics. This accredited study program offers excellent study conditions with three research groups in different areas. Students first learn the foundations of the subjects not studied previously. Thus graduates in the life sciences first learn the foundations of computer science and programming, while graduates in computer science first learn the foundations of chemistry and molecular biology. The students can then specialize in genome informatics, structural bioinformatics or cheminformatics. The program is completed with comprehensive elective courses in computer and life sciences.

### **Distance learning and other resources**

The textbooks most commonly used in cheminformatics courses are those by Leach and Gillet<sup>22</sup> and Gasteiger and Engel.<sup>23</sup> Books by Bajorath<sup>24</sup> and by Oprea<sup>25</sup> are also recommended. A few universities in China and India are also teaching cheminformatics but details of these are not well known in Europe and the United States.

The cheminformatics research group at Sheffield runs an intensive one week short course in cheminformatics (“A Practical Introduction to Cheminformatics”) which provides hands-on training for scientists already in employment. It is sponsored by student bursaries from the Molecular Graphics and Modelling Society and CSA Trust. It has been run annually over four days since 2001, delivered to about 18 delegates from the United Kingdom, mainland Europe and the United States, mostly from industry. Numbers are restricted due to the intense nature and practical component of the course. The Institute of Cheminformatics Studies in India offers 12-

month distance training programs<sup>26</sup> for industry. The Louis Pasteur University of Strasbourg will also be offering short courses for industry,<sup>27</sup> starting in May 2009.

Distance education at Indiana uses Web conferencing and teleconferencing. There are links with the University of Michigan, to run the class at both Universities (but centered in Indiana). Lectures are recorded. Out of 75 students, 39 have been distance education students. It has been reported<sup>18</sup> that the benefits of distance education win out over the challenges. A wiki is used for the introductory cheminformatics class. The University of Erlangen-Nuremberg offers teaching modules for cheminformatics on the Web,<sup>28</sup> in collaboration with FIZ CHEMIE Berlin, developed within the scope of "Vernetztes Studium-Chemie", a lead project of the German Federal Ministry of Education and Research. Currently, only some chapters are available.

DouglasConnect runs both public and private workshops and training programs in the areas of knowledge management and innovation. It has also an eCheminfo "community" and runs hands-on eChemInfo workshops.<sup>29</sup>

One problem in developing cheminformatics teaching courses is the lack of freely available software. In bioinformatics much software has been developed in academia and is available at zero or low cost but only recently has free cheminformatics software begun to appear. For example, at Indiana, ChemBioGrid<sup>30</sup> is exploiting publicly available chemical information. Several databases are exposed through Web service wrappers, and the research team has also created Web service wrappers around several free and commercial cheminformatics tools. It also has a close working relationship with the Murray-Rust group in the Unilever Centre for Molecular Informatics at Cambridge University and has implemented several of their Web services locally. The researchers have also implemented a large amount of the functionality of the Chemistry Development Kit (CDK)<sup>31</sup> as Web services, plus many services relating to the R statistical package.

In the Cheminformatics Community Wiki<sup>32</sup> a number of researchers and teachers in cheminformatics have joined forces to define the field in terms of a curriculum and to create open teaching materials in cheminformatics and chemical information based on open technologies and standards.

The practical sessions of FCT Lisbon's cheminformatics course are based on software freely available to students (free cross-platform Java software available for different operating systems), as well as public data sets and Web services. All required resources are integrated into a Web portal for the course,<sup>33</sup> enabling students to study and practice on their own, at the practical sessions and after classes. Use of the material on the Web is authorized to other academic institutions provided the source is acknowledged.

Synergix Ltd. has developed commercial educational resources for drug design, molecular modeling, cheminformatics and medicinal chemistry. In 2002 the Company launched a computer-based course<sup>34</sup> entitled "Molecular Conceptor" which takes advantage of interactive multimedia systems.

## **Conclusion**

The catalog of courses and resources compiled in this paper might suggest that cheminformatics education is flourishing. It is not. Many examples of isolated efforts are cited here but there is no European or international coordination. Cheminformatics practitioners have still not defined their discipline and its impact, let alone successfully made a case to governments and funding agencies. Perhaps the pharmaceutical industry, which has much to lose, should be further collaborating with academia to address the challenge.

## **References**

1. Warr, W. A. Balancing the needs of the recruiters and the aims of the educators. Presented at 218<sup>th</sup> ACS National Meeting. New Orleans, Louisiana, August 22-26, 1999. <http://www.warr.com/warrzone2000.html> (accessed October 12, 2008).
2. Hann, M.; Green, R. Chemoinformatics - a new name for an old problem? *Curr. Opin. Chem. Biol.* **1999**, *3*, 379-383.
3. Brown, F. K. Cheminformatics: what it is and how does it impact drug discovery? *Annu. Rep. Med. Chem.* **1998**, *33*, 375-384.
4. Gasteiger, J. The central role of chemoinformatics. *Chemom. Intell. Lab. Syst.* **2006**, *82*, 200-209.
5. Bajorath, J. Cheminformatics and its relationship to drug discovery, bioinformatics, basic research, and teaching. *Drug Discovery Today* **2007**, *2*(9) <http://www.drugdiscoverytoday.com/echoice/nov2007/theview.html> (accessed October 12, 2008).
6. Varnek, A. Cheminformatics: recognition through teaching. Presented at 235<sup>th</sup> ACS National Meeting. New Orleans, Louisiana, April 6-10, 2008.
7. *Chemoinformatics Approaches to Virtual Screening*; Varnek, A., Tropsha, A, Eds.; RSC Publishing, Cambridge, England, 2008.
8. Michigan Technological University Undergraduate Chemistry Programs. <http://www.chemistry.mtu.edu/pages/undergrad/index.php> (accessed October 13, 2008).
9. Obernai Declaration. [http://infochim.u-strasbg.fr/chemoinformatics/Obernai\\_declaration.php](http://infochim.u-strasbg.fr/chemoinformatics/Obernai_declaration.php) (accessed October 12, 2008).
10. Goodman, J. M. Molecular informatics: research and learning. Presented at 235<sup>th</sup> ACS National Meeting. New Orleans, Louisiana, April 6-10, 2008.
11. Aires-de-Sousa, J. Mounting an undergraduate chemoinformatics course with free software. Presented at 235<sup>th</sup> ACS National Meeting. New Orleans, Louisiana, April 6-10, 2008.
12. Gillet, V. J.; Holliday, J.; Willett, P. Graduate training in chemoinformatics at the University of Sheffield. Presented at 235<sup>th</sup> ACS National Meeting. New Orleans, Louisiana, April 6-10, 2008.
13. Willett, P. Education for cheminformatics. Presented at Cheminformatics in Europe: Research and Teaching. Obernai, France, May 29-31, 2006. [http://infochim.u-strasbg.fr/chemoinformatics/Education\\_Chemoinformatics.pdf](http://infochim.u-strasbg.fr/chemoinformatics/Education_Chemoinformatics.pdf) (accessed October 12, 2008).
14. Wild, D. J.; Wiggins, G. D. Challenges for cheminformatics education in drug discovery. *Drug Discovery Today* **2006**, *11*(9/10), 436-439.
15. Schofield, H.; Wiggins, G.; Willett, P.; Recent developments in cheminformatics education. *Drug Discovery Today* **2001**, *6*(18) 931-934.
16. Cooke, H.; Willett, P. Masters level training in chem(o)informatics in the UK. Presented at 235<sup>th</sup> ACS National Meeting. Philadelphia PA, August 22-26, 2004. <http://acscinf.org/docs/meetings/228nm/presentations/228nm38.pdf> (accessed October 13, 2008).

17. University of York courses. <http://maths.york.ac.uk/www/Chemoinformatics> (accessed October 24, 2008).
18. Wild, D. J. Developing a cheminformatics education and teaching center for the Web 2.0 world. Presented at 235<sup>th</sup> ACS National Meeting. New Orleans, Louisiana, April 6-10, 2008.
19. University of Massachusetts Lowell M.S. Computer Science Bio/cheminformatics Option. [http://genome.uml.edu/programs\\_grad.htm](http://genome.uml.edu/programs_grad.htm) (accessed October 13, 2008).
20. MSc program in computational science. <http://www.physik.uni-frankfurt.de/mpcs/> (accessed October 13, 2008).
21. Study program for bioinformatics/cheminformatics. <http://www.zbh.uni-hamburg.de/study.php> (accessed October 17, 2008).
22. Leach, A. R.; Gillet, V. J. *An introduction to cheminformatics*. Kluwer Academic Publishers: Dordrecht, The Netherlands, 2003.
23. Gasteiger, J.; Engel, T. *Cheminformatics: a textbook*. Wiley-VCH, Weinheim, Germany, 2003.
24. *Cheminformatics: concepts, methods, and tools for drug discovery (methods in molecular biology)*; Bajorath, J., Ed.; Humana Press: Totowa, NJ, 2004.
25. Oprea, T. I. *Cheminformatics in drug discovery*. John Wiley & Sons: New York, NY, 2005.
26. Institute of Cheminformatics Studies. [http://www.cheminformaticscentre.org/files/academic\\_program.asp](http://www.cheminformaticscentre.org/files/academic_program.asp) (accessed October 13, 2008).
27. Initiation à la chemoinformatique. <http://www.seforco.fr/stage-684-0809.html> (accessed October 17, 2008).
28. Vernetztes Studium-Chemie teaching modules for cheminformatics. <http://www2.chemie.uni-erlangen.de/projects/vsc/chemoinformatik-engl.html> (accessed October 12, 2008).
29. eCheminfo. [http://echeminfo.com/COMTY\\_training](http://echeminfo.com/COMTY_training) (accessed October 13, 2008).
30. ChemBioGrid. <http://www.chembiogrid.org/index.html> (accessed October 12, 2008).
31. Steinbeck, C.; Han, Y.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E. The Chemistry Development Kit (CDK). An Open-source Java Library for Chemo- and Bioinformatics. *J. Chem. Inf. Comput. Sci.* **2003**, 43(2), 493-500.
32. Cheminformatics community wiki. <http://wiki.cubic.uni-koeln.de/cheminfwiki> (accessed October 12, 2008).
33. FCT course materials. <http://www.dq.fct.unl.pt/staff/jas/qc> (accessed October 13, 2008).
34. Molecular Conceptor. <http://www.molecular-conceptor.com/home.html> (accessed October 13, 2008).