



The
University
Of
Sheffield.

Postgraduate Education Programmes in Chemoinformatics

Peter Willett, University of Sheffield

Presented at Obernai, 30th May 2006



Overview

- The need for formal education programmes
- Overview of three such programmes
 - Organisation
 - Taught components
 - Critical success factors
- Conclusions



From chemical information to chemoinformatics

- Chemical information systems available for many years
 - Covered in mainstream chemistry programmes
- Recent prominence of chemoinformatics
 - Huge volumes of data about molecular structures and bioactivities resulting from developments in combi-chem and HTS
 - Otherinformatics subjects as a result of comparable data explosions
 - Substantially increased employment opportunities

Sources of trained person-power

- Established sources of skilled staff insufficient to meet market needs
 - Limited number of academic institutions with active chemoinformatics research groups
 - Larger number carrying out related research, e.g., chemometrics, crystallography, computational chemistry
 - In-house training of chemists who developed computational interests
- Need for dedicated educational programmes to address this world-wide shortage

Educational programmes: I

- Much current interest at BSc and MSc levels
- D.J. Wild and G.D. Wiggins (2006) “Challenges for chemoinformatics education in drug discovery” *Drug Discovery Today*, 11, 436-439, lists the following programmes *inter alia*:
 - Louis Pasteur University of Strasbourg
 - Johann Wolfgang Goethe University of Frankfurt
 - University of Massachusetts at Lowell
 - University of Noiada in India
 - Michigan Technical University



Educational programmes: II

- Focus in this presentation on three programmes
 - MSc Chemoinformatics at Sheffield
 - MSc Cheminformatics at Manchester (formerly UMIST)
 - MS Chemical informatics at Indiana
- Rationale for this focus is that I have at least some personal knowledge of all of these
- Commonality of background for the two UK programmes



Funding support: I

- The Engineering and Physical Sciences Research Council (EPSRC) is one of the main research funding agencies in the UK.
- Will consider support for MSc programmes if, and only if, demonstrable and significant industrial need
- Discussions with them and industry 1998-99 led to a call for new Masters Training Packages (MTPs) that included chemoinformatics as one of several priority areas



Funding support: II

- Both Sheffield and Manchester bid, with MTP awards being announced mid-July 2000
 - Funding over 5 years (ca. £500K) to cover (principally) student tuition fees and maintenance costs
 - Expectation that courses would become self-funding at end of period
- Sheffield start in September 2000 and Manchester in September 2001



Sheffield and Manchester courses

- Add informatics skills to existing chemical knowledge (cf some bioinformatics courses)
- The differences in content arise mainly from the different natures of the two departments
 - Stronger chemistry focus in Manchester
 - Stronger informatics focus and industrial involvement in Sheffield
- Common structure
 - Two semesters of taught modules
 - Dissertation project in the summer



Taught modules: Manchester

- Core
 - Applications of cheminformatics, bioinformatics (two modules), chemical information sources, database management and computer programming (Fortran and Perl), molecular simulation (two modules), spectroscopy and crystallography
- Electives
 - Algorithm design, combinatorial chemistry, technology enterprise
- NB some variations in detailed composition at both institutions



Taught modules: Sheffield

- Core
 - Chemoinformatics (two modules), computer programming (Java and Perl, from Computer Science), database design (Oracle), information retrieval, information systems modelling
- Elective
 - Advanced information retrieval, electronic publishing, healthcare information, human-computer interaction, molecular modelling (from Chemistry)



Dissertation

- This runs from ca. mid-June till September
- At Manchester, the dissertation can be carried out internally (the majority) or externally on placement with a company
- At Sheffield, nearly all are carried out externally on placement with a company
 - Initially just in UK but since 2003-04 involvement also from France, Holland, Spain and Switzerland
 - Not necessarily the aim but several have resulted in articles (four thus far, and two to be done)



Industrial placements

- Industry is keen to support placements
 - A key part of the Sheffield proposal
- Four main types of project
 - Development and/or testing of an existing or novel piece of software for some specific application
 - Development of a Web front-end to an existing system or service
 - Comparison of different programs for some specific application
 - Analysis of chemical and/or biological dataset(s)

Continuing professional development: I

- Little industrial interest in distance-learning modules (cf Nottingham molecular modelling)
- Limited take-up of industry secondment and Manchester day release schemes
- But strong interest expressed in a short course, “A Practical Introduction to Chemoinformatics”
 - Sponsored by MGMS and CSA Trust
 - Four days in June
 - Delivered to ~16 delegates from around the world; mostly from industry plus two student bursaries



Continuing professional development: II

- Material from Sheffield MSc programme focussing on drug discovery applications
- Both academic and industrial tutors
- Mixture of lectures, hands-on sessions, and a group problem-solving exercise
- Over-subscribed every year that it has been run (annually since 2001)



Indiana: background

- School of Informatics started in 2000
 - Based on departments at both the Indianapolis and Bloomington campuses
- Wide range of MS/PhD programmes
 - Bioinformatics, health informatics, human-computer interaction, laboratory informatics, music informatics, new media, social informatics etc
 - Chemoinformatics complements these, started in 2001 with 9 MS students to date (cf 72 for bioinformatics and 36 for health informatics); also PhD programme since 2005, with 2 students



Indiana: course content

- Strict pre-requisites
 - If chemistry (or computer science) UG then must cover much of a UG minor in computer science (or chemistry)
- Modules cover
 - Chemical information technology, molecular modelling, programming for chemical informatics
 - Extensive range of elective modules (bioinformatics recommended)
 - Small-scale dissertation project (20% of entire course)



Indiana: distance learning

- Considerable effort to provide distance learning modes, with clear CPD potential
- Initial use of dedicated videoconferencing to cover Bloomington and Indianapolis campuses
 - Too complex when more than two specific locations involved
 - Move to location-independent teaching and learning using Macromedia Breeze
 - Student in California, faculty in Sheffield and Philadelphia etc



Funding support: I

- NIH call in 2005 for applications for Exploratory Centers for Cheminformatics Research: \$4M to fund six centers
- Indiana is one of these
 - Chemical Informatics and Cyberinfrastructure Collaboratory (cf e-science in UK); MS programme part of the Collaboratory
 - MIT, North Carolina State, Rensselaer, UMI, UNC
- Forthcoming NIH follow-up call as part of the NIH Roadmap Molecular Libraries Initiative (see <http://www.nihroadmap.nih.gov/molecularlibraries>)
 - Seeking applications for two Centres for Chemoinformatics Research (each ca. \$10M over 5 years)



Funding support: II

- The NIH initiative is in marked contrast to the situation in the UK, where has been difficult to get research-council, as against industrial, funding.
- EPSRC : chemistry panel doesn't fund computer science and vice versa
- BBSRC: go via bioinformatics route, which has been strongly funded on back of established strengths in structural biology
- Will Roadmap encourage funders in EU?



Lessons that we can draw

- Principally (but not exclusively) based on Sheffield experience
 - Industrial involvement
 - Crucial if a meaningful programme is to be delivered
 - Sustainability
 - Or the lack thereof
 - Employability
 - Excellent prospects for students taking the programme



Industrial involvement: I

- Crucial in EPSRC agreeing to fund this area
 - Appropriately so: chemoinformatics is not a purely academic discipline so need for industrial involvement to ensure relevance
- Visiting lecturers
 - Manchester: database producers and software suppliers
 - Sheffield and Indiana: database producers, software suppliers and ag/pharm companies
- Dissertation projects and supervision (this implying significant funding of the programmes)



Industrial involvement: II

- Dissertation the most important part of a UK MSc (cf certificate and diploma courses)
 - Dissertation intended from the start to be the principal component of the programme, towards which the taught components build
- Correctness of this view, with all three stakeholders benefiting
 - Companies come back year after year
 - Students maximise their employability
 - Academic departments get research links and publications (unexpected benefit)



Sustainability: I

- The original EPSRC funding was for 5 years
 - Sheffield 2000-2004, Manchester 2001-2005
- Expectation by EPSRC that the programmes would become self-funding
 - Students (both EU and overseas) supporting themselves
 - Marketing of distance-learning modules derived from programme materials
- These expectations were not realistic



Sustainability: II

- In UK, as elsewhere, declining numbers of students studying physical sciences at UG level
 - First physics and now chemistry closures
- Even with the offer of fees, maintenance and placement, only limited numbers of applications for the programme
- Manchester programme now in its final year, with Sheffield expecting 2006-07 as final year
- Move to MRes degrees



Employability of graduates

Employment	Manchester 2002-04	Sheffield 2000-04
Industry	26	20
PhD	11	13
Other	3 patents, 2 teachers	10 IT/informatics
Unknown	12	5



Conclusions

- There is a market here, but it is not one that students realise (at least in the UK thus far)
- Extensive industrial involvement is mandatory for a successful programme
- Possible effect of Roadmap on funding chemoinformatics research



- Acknowledgements

- Val Gillet and John Holliday (Sheffield)
- Helen Cooke and Patrick O'Malley (Manchester)
- Garry Wiggins and David Wild (Indiana)

- References

- M. Hann and R. Green (1999) "Chemoinformatics - a new name for an old problem?" *Current Opinion in Chemical Biology*, 3, 379-383
- H. Schofield, G. Wiggins and P. Willett (2001) "Recent developments in chemoinformatics education", *Drug Discovery Today*, 6, 931-934
- D.J. Wild and G.D. Wiggins (2006) "Challenges for chemoinformatics education in drug discovery" *Drug Discovery Today*, 11, 436-439