Chemical Structure Association Trust

NEWSLETTER

Summer 2005 Issue 10

We make no apologies for the fact that a large section of this summer Newsletter is taken up the International Conference on Chemical Structures held in June in Noordwijkerhout, the Netherlands. The Trust is a co-organiser of these triennial conferences, and this year students from the Sheffield University Department of Information Studies received the 2005 Trust Grant to attend the conference (report on pages 4–7). The conference opened with the presentation of the 2005 CSA Trust Mike Lynch Award to Johnny Gasteiger, who gave the keynote address. Johnny is pictured on the right relaxing on board the Willem Barentz.

Left to right: CSA Trustees Peter Willett, Johnny Gasteiger, Guenter Grethe and Peter Nichols enjoying a refreshing trip on the IJsselmeer in the Netherlands

Applications Invited for CSA Trust Jacques-Émile Dubois Grant for 2006

The CSA Trust has created a unique Grant Program in support of its Charter, and is currently inviting the submission of grant applications for 2006. The Grant has been renamed as the 'CSA Trust Jacques-Émile Dubois Grant' in memory of Jacques-Émile who died earlier this year, and who made such a great contribution to chemoinformatics.

Purpose of the Grants:
The Grant Program has been created to provide funding for the career development of young researchers who have demonstrated excellence in their education, research or development activities that are related to the systems and methods used to store, process and retrieve information about chemical structures, reactions and compounds. A Grant will be awarded annually up to a maximum of three thousand US dollars ($3,000), subject to suitable applications being received. Grants are awarded for specific purposes, and within one year each grantee is required to submit a brief written report detailing how the grant funds were allocated.

Who is Eligible?
Applicant(s), age 35 or younger, who have demonstrated excellence in their chemical information related research and who are developing careers that have the potential to have a positive impact on the utility of chemical information relevant to chemical structures, reactions and compounds, are invited to submit applications. While the primary focus of the Grant Program is the career development of young researchers, additional bursaries may be made available at the discretion of the Trust. All requests must follow the application procedures noted below and will be weighed against the same criteria.

What Activities are Eligible?
Grants may be awarded to acquire the tools necessary to support research activities, or for travel to collaborate with research groups, to attend a conference relevant to one's area of research, to gain access to special computational facilities, or to acquire unique research techniques in support of one's research.

Applications must include the following documentation:

1. A letter that details the work upon which the Grant application is to be evaluated, as well as details on research recently completed by the applicant;
2. The amount of Grant funds being requested and the details regarding the purpose for which the Grant will be used (eg cost of equipment, travel expenses if the request is for financial support of meeting attendance, etc). The relevance of the above-stated purpose to the Trust's objectives and the clarity of this statement are essential in the evaluation of the application);
3. A brief biographical sketch, including a statement of academic qualifications;
4. Two reference letters in support of the application.

Additional materials may be supplied at the discretion of the applicant only if relevant to the application and if such materials provide information not already included in items 1–4.

Four copies of the complete application documentation must be supplied for distribution to the Grants Committee.

Deadline for Applications:
Applications must be received by October 15, 2005. Successful applicants will be notified by December 16, 2005.

Address for Submission of Applications:
Four copies of the application documentation should be forwarded to: Bonnie Lawlor, CSA Trust Grant Committee Chair, 276 Upper Gulph Road, Radnor, PA 19087, USA. E-mail submissions, if complete, may be forwarded to the Grant Committee at chescot@aol.com.

People and Places

John L. Regazzi has decided to accept a position as Dean, College of Information and Computer Science of the Long Island University and to retire from Elsevier. He will maintain his position as Chairman of The Elsevier Foundation.

Jan Velterop is leaving Biomed Central to concentrate on promoting Open Access to societies, funding institutions and publishers. Matthew Cockerill and Anne Greenwood will take joint responsibility for publishing and other activities of BioMed Central.

Joe Donahue has been appointed corporate chief business officer and president of InforSense North America. He will also lead the company's worldwide sales operations. He was previously with LION bioscience.

Trevor Heritage has left Tripos to join MDL as a senior vice president and lead the Workflow Applications Group.

Rudy Potenzone has moved from Ingenuity and recently joined CambridgeSoft as their new Vice President of Enterprise Solutions.

Tudor Oprea has been elected as the new Chair of the International QSAR and Modelling Society for 2006–2010. The other officers are Corwin Hansch, Alexander Tropsha, Hugo Kubinyi, Klaus Gundertofte, Stefan Balaz and Curt Breneman.

Roger Beckman and Brian Winterman of Indiana University Chemistry Library have taken over the running of CHMINF-L (Chemical Information Sources Discussion List) from Gary Wiggins.

Ray Carhart has gone back to Symyx Technologies.

Eugene Garfield has been elected a Fellow of the American Academy of Arts and Sciences.

Scott Kahn has left Accelrys and joined Illumina.

Joint CSAT/CINF Symposium:

'De Novo Design and Synthetic Accessibility'

To take place during the 231st ACS Meeting in Atlanta, USA, in Spring 2006 (26–30 March).

This symposium will be organised on behalf of the CSA Trust by Professor Dr. Johann Gasteiger, johann.gasteiger@chemie.uni-erlangen.de

Details will be available after the 2005 ACS Fall meeting at

http://www.chemistry.org/portal/a/c/s/1/acsdisplay.html?DOC=meetings%5Cfuture.html

Welcome to new Trustee

We welcome Steve Heller, who has recently been appointed as a new Trustee. Steve is currently a consultant and a guest researcher at NIST working on the IUPAC INChI chemical identification project, which he initiated. The objective of the IUPAC Chemical Identifier Project is to establish a unique label, the IUPAC International Chemical Identifier (INChI), a non-proprietary identifier for chemical substances that could be used in printed and electronic data sources thus enabling easier linking of diverse data compilations. Steve is also a member of the NIH PubChem Roadmap project, and is a consultant to and the director of specialised databases for the Yahoo of chemistry, Chem-industry.com. He received a BS in Chemistry from SUNY-Stony Brook and a PhD in Chemistry from Georgetown University. From 2000–2002 he was a strategic planner for MDL Information Systems, and has been a member of the editorial board of the *Internet Journal of Chemistry*, and editor of the Elsevier *Trends in Analytical Chemistry* (TrAC) internet computer column. Steve was also the creator of the NIH/EPA/NIST Mass Spec database and the NIH/EPA CIS (Chemical Information System).

The 2005 International Chemical Information Conference and Exhibition, Nîmes, France 16–19 October 2005

Full details at http://www.infonortics.com/chemical/ch05/05chempro.html

Sunday 16 October

Registration 15.00–19.00 followed by Reception and Dinner

Monday 17 October

Keynote Talk:

Controversies, Conflicts and Challenges in Scientific Publishing

Robert Bovenschulte, ACS Publications, DC, USA

Session 1: The Shifting Balances in Publication & Media

Martin G. Hicks, Beilstein Institut, Germany
Sabine Brünger-Weilandt, FIZ Karlsruhe, Germany
Willem-Geert Lagemaat, Univentio, The Netherlands

Session 2: Integrated Tools for Discovery in R&D

Bruce E. Wilson, Dow Chemical Company, Michigan, USA
Yike Guo, InforSense, UK
Brian Bartell, Notiora, USA
Alexander Lawson, Elsevier MDL, Germany
Lewis Jardine, Intellidos, UK

Tuesday 18 October

Keynote Talk:

Where is Google going? What information professionals should be watching, and why

Stephen E Arnold, AIT, Kentucky, USA

Session 3: Innovation and the Next Great Things

John Lewis Needham, Google, California, USA
Leslie Lees, Knovel, New York, USA
Corilee Christou, Reed Business Information, USA
Miriam Heller, National Science Foundation, DC, USA

Session 4: Data Analysis

Christine McCue and Rainer Stuike-Prill
STN/CAS, Ohio, USA
Patrick Perrin, ExerGen Biosciences, USA
Louis Graziano, The Rohm and Haas Company, USA
Christoph Schaub and Achim Zielesny
Bayer Business Services & University of Applied Sciences, Gelsenkirchen, Recklinghausen, Germany

Evening: Outing and Dinner

Wednesday 19 October

Session 5: Tools & Strategies

Martin White, Intranet Focus, UK
Peter Kallas, BASF, Germany
Nigel Clarke, European Patent Office, Austria

Session 6: Information Management & Organisation

Robert Scoffin, CambridgeSoft, UK
Morten Christoffersen, Novo Nordisk, Denmark
Richard Neale, Thomson Scientific, UK

Close of conference 12.45pm

The 7th International Conference on Chemical Structures

This triennial conference was held, as on all previous occasions, at the Conference Centre in Noordwijkerhout. It has become one of the major chemoinformatics conferences, as can be seen by the fact that over 200 people attended the meeting from 21 different countries. A number of student bursaries were offered through the Trust and we give particular thanks to AstraZeneca, GlaxoSmithKline and the Cambridge Crystallographic Data Centre, who sponsored the following students:

- Kaido Tämm, University of Tartu, Estonia (CCDC)
- Alex Perryman, University of California at San Diego (GSK)
- Alexandre Carvalho, Universidade do Porto (AstraZeneca)

The Trust sponsored Naparat Kammasud, Institut Curie–Laboratoire Raymond Latarjet in France and in addition, the 2005 Trust Grant was awarded to Sheffield University Department of Information Studies to enable eight students to attend the conference. This report was written by Kirstin Moffat (Monday am), Linda Hirons (Monday pm and posters), Yogendra Patel (Tuesday am), Simon Cottrell (Tuesday pm and posters), Jerome Hert (Wednesday am), David Wood (the excursion on Wednesday pm) and Kris Birchall (Thursday am).

The keynote address on the Sunday evening was given by Johnny Gasteiger (pictured below), who was presented with the 2005 CSA Trust Mike Lynch award by Guenter Grethe. The title of Johnny's talk was 'My Love Affair with Molecules – and Reactions'. Although the international language of chemistry is the two-dimensional structure diagram, Johnny compares molecules to human beings, and he

treasures the fact that they are three-dimensional species which have skin, left and right hands and can change shape. The keynote address covered computational approaches to the generation of 3D molecular models, the calculation of molecular surface properties, the generation of multiple conformations and the identification of molecular chirality. The address concluded with a description of some of the computer methods for modelling chemical reactions. The full talk given by Johnny Gasteiger is at http://www2.chemie.uni-erlangen.de/presentations/. Following the keynote address, delegates were treated to a reception, courtesy of Elsevier MDL and a splendid rijsttafel sponsored by CAS.

The main part of the conference started on the Monday with the cheminformatics session. A number of interesting and informative presentations were given, covering topics such as similarity searching, reaction classification, activity prediction, and pharmacophore generation.

Methods discussed for similarity searching included Peter Willett's presentation on data fusion methods that have been carried out at Sheffield University, such as turbo similarity searching, where data fusion is carried out using the nearest neighbours of a target compound. Andreas Bender also gave a snappy presentation on how COSMO screening charges can be used to describe surface properties, which can then be used in similarity searching, and illustrated this with various examples.

In the session on activity prediction, Joerg Wegner looked at whole molecules and presented a comparison of the use of vector based coding and structure based coding methods for the prediction of ADMETox data. Peter Ertl focused on whether the structural and property features of active heterocyclic rings could be used to discriminate between activity and non-activity. This information can then be used to search a database of reasonable ring systems for those that may be active for a certain target.

Pharmacophore generation was also considered. Robert E. Clark illustrated that a MOGA using multiplet pharmacophores, generated from a number of different molecules, and steric information, could be used to identify molecular alignments, while Gerhard Wolber defined a pharmacophore by looking at where a ligand interacts with a protein in complexes, using LigandScout. Initial improvement of the ligand information, by considering things such as hybridisation states and bond orders, is followed by 3D pharmacophore generation. It is then possible to overlay the pharmacophores from comparable complexes to find chemically and geometrically comparable features.

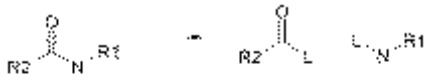
This session ended with Guenter Grethe describing how hierarchical tree structures can be used to classify reactions by type or name. The tree structure allows addition of reactions not already covered, and can include named reactions that may be considered as a combination of two or more reactions.

The afternoon began with a humorous yet informative talk from Stephen Heller, full of great quotations and amusing analogies. The effect of the internet upon scientific publishing was discussed with our need to confront open access issues. In the second half of the presentation the IUPAC International Chemical identifier (InChi) was presented and compared to SMILES strings. Unlike SMILES, InChi is a unique string representation for a chemical structure. The ASCII code is generated quickly and contains all connectivity information required to analyse a molecule further. Nikolous Stiefl then introduced us to a novel descriptor ErG, an Extension to reduced Graphs for application to scaffold hopping. Promising results were presented with retrieval rates similar to DAYLIGHT fingerprints. The third presentation by Matthew Stahl covered the importance of conformer generation with suggestions for improvements to both their construction and application to structure-based drug design. The afternoon session ended with two product reviews, the first from Eva Seip on MDL's Patent Chemistry Database and Discovery Gate, and the second from Miklos Vargyas on ChemAxon's Platform for Cheminformatics.

After coffee, the first poster session began with 35 presentations. A wealth of research interests were covered and summarised by many colourful and creative illustrations and catchphrases. For exam-

ple, a Novartis intranet web tool called 'Treasure Island' trains neural networks with three-dimensional molecular structure to make property predictions. Other topics included the use of the docking to identify likely drug binding conformations to use as an alignment in 3D QSAR studies of androgen receptor ligands (presented by Annu Söderholm). The day finished with a reception courtesy of Accelrys and a superb buffet dinner.

The sessions on Tuesday covered structure-based design and virtual screening. In his presentation on FlexNovo, Jorg Dregen (Hamburg University) described a docking program in which the fragments used are based on molecules from the WDI. 35,000 molecules were cut at strategic points, creating 16,780 unique fragments and 21,386 linking points (cf. retrosynthesis), an example of which is below:



where L is the linking point.

The procedure of FlexNovo is as follows:

1. Fragments are placed in the binding site using pharmacophoric principles (so that there are favourable interactions such as H-bonding or hydrophobic contacts between the fragment and the binding site) and scored based on their interaction with the receptor (using the same scoring function as FlexX).
2. The highest scoring fragments are then taken through to the next stage in which other fragments are added to the original fragment via linking rules (which account for: compatibility of fragments being joined; modifications of topology and geometry; availability of terminal linking points for further extension of the generated molecule).
3. The generated poses are then filtered according to their geometry so the unsaturated buried H-bonds, close contacts of equally charged groups and high energy conformations are filtered out. Other filtering methods (such as Lipinski's rule of 5) can also be applied.

In examples given molecules that are similar to known ligands, and bind in a similar fashion, were generated at a speed of 1–2 seconds of processing time.

Kristztina Boda (University of Leeds) described recent advances in de-novo ligand design and optimisation, in particular additions made to the SPROUT program. Structure generation in SPROUT is based on the following method:

1. Find binding pockets in a receptor
2. Find potential H-bonds
3. Dock structures
4. Link fragments
5. Score and cluster solutions.

Large numbers of solutions are generated during the sequential structure generation so heuristics are used to reduce the number of potential solutions. A complexity score is implemented so that atom patterns in potential solutions are compared to those found in typical drugs (taken from the MDDR, WDI, etc) and the high rank solutions are likely to be good candidates for screening/drug development programs. A problem with this is that as the patterns are based on existing drugs, simple but novel solutions may be identified as being poor candidates.

The last part of the presentation focused on SYNSPROUT. Here all possible molecules using a core fragment and a monomer library (eg for R groups) are generated. Pharmacophoric features for the core and library fragments are identified and placement of the core is defined by either size or the location of H-bonds. Each monomer has a set of low energy conformations (to account for flexibility of the ligand – SPROUT adds fragments but does not account for flexibility) and those with a high energy or with boundary violations are rejected. The remaining poses are scored with the best ones kept for analysis.

Holger Claussen (BioSolveIT) discussed combining the power of combinatorial chemistry with the efficiency of pharmacophore-based docking, and described methods for using docking in combination with core fragments for a molecule. The principle of their work is that core fragments are docked, the substituents are added and the new molecule are then scored. They found that the scores by using the method were vastly different to when the whole molecule is docked in one go. By increasing the size of the core fragment to include one of the substituents, the correlation between the scores improved. They found this method to be much faster than docking compounds individually. The second part of his presentation focused on using pharmacophoric constraints during the docking procedure. Using this method 15 out of 16 actives were found. The speed up of their method compared to normal sequential docking (where compounds are docked sequentially) was found to be vast. Using a 60 node cluster a sequential docking program would take around 60 days; their method took only two days.

Andy Vinter (Cresset BioMolecular Design Ltd) showed the usage of molecular fields for identifying bound conformations of molecules. The fields created have coverage for hydrophobic, hydrophilic, H-bond acceptors, H-bond donors and steric bulk of a molecule. The fields of two molecules are then overlaid, with the aim of identifying the conformation of the second molecule which is bound (the first field is derived from the bound conformation of a molecule). For applications in virtual screening the 3D structure of molecules has to be known, along with various conformations of each molecule as the align-

Sheffield students at the conference: back row, left to right, Simon Cottrell, David Wood, Jerome Hert, Kris Birchall; frontrow from left to right, Linda Hirons, Yogendra Patel, Kirstin Moffatt, Chido Mpamhanga

ment is dependent on conformations. A database of 1.5 million compounds contained 75 million entries, as there were up to 60 different conformations per molecule. From this method, new active compounds were found for GPCR ligands, using two known actives to form the initial fields for a query.

Henriette Willems (De Novo Pharmaceuticals) used a de novo structure-based design tool, SkelGen, which is a generation program used in combination with pharmacophore constraints and fragments generated from the WDI to create novel micromolar ligands against the oestrogen receptor. In addition to the de novo scoring functions, molecules were scored on the ease of their synthesis. This process was based on 107 reactions. All possible synthetic routes for a molecule were generated, and the score obtained was based on these routes.

Donovan Chin (Biogen Idec) described interaction fingerprints for use in designing combinatorial libraries. A bit string is created which describes the interactions of a binding site with a ligand. Combining the fingerprint for a set of related targets (eg a family of kinases) results in 'fingerprint profiles'. Using these fingerprints it is possible to remove compounds that are often scored highly by a program. In addition, using these fingerprints as descriptors for the ligands, a SAR tree can be developed, which can visually explain why some compounds are poor drugs even though they are good inhibitors. In one example, a tree showed that the region which contained compounds that were good inhibitors was also a region with poor oral delivery.

In the last session of the morning, Simon Folkertsma (University of Nijmegen) described a similar technique. Using the aligned binding domains of 101 Nuclear Receptors (NRs) a graph was obtained for the average number of contacts per position (of the NR) with a ligand. This was then used to see which positions were important for the binding of agonists and which were important for the binding of antagonists.

The Tuesday afternoon session continued the themes of the morning, with two talks on virtual screening followed by one on de novo design. Steffen Renner of the University of Frankfurt opened the session with description of his fuzzy pharmacophore approach, known as SQUID. In SQUID, pharmacophore models are represented by Gaussian feature density functions. Steffen found there was little overlap between the virtual screening hits found using SQUID pharmacophore models and conventional pharmacophore models. Therefore, the two techniques appear to be complementary.

Tímea Polgár of Richter Gedeon Ltd, Budapest, then described the development of a virtual screening protocol for β-secretase. There is presently no effective treatment for Alzheimer's disease, but a potential therapy would involve inhibition of the β-secretase pathway. Virtual screening was performed using BioSolveIT's FlexX software. The basic FlexX module was compared with FlexX-Pharm, which allows pharmacophore-type constraints to be defined, and FlexE, which allows protein flexibility to be taken into account. FlexX-Pharm provided vastly improved enrichment rates compared to FlexX, using a previously derived pharmacophore model for β-secretase inhibition. However, FlexE performed badly compared to FlexX.

The last speaker was Suresh Singh of Vitae Pharmaceuticals, Pennsylvania, on de novo design of PPAR-α agonists using Vitae Pharmaceuticals' ConTour technology. ConTour uses a course-grained knowedge-based potential which was derived from high-resolution protein–ligand structural data, and has been validated by synthesising and measuring the activity of the compounds generated.

The talks were followed by the second poster session, at which over 40 posters were presented. Patrick Fricker of the University of Hamburg presented a novel technique for scaffold replacement, in which replacements are suggested for a central element linking two or more terminal groups. Sandra Handschuh of Boehringer Ingelheim showed how the GRID program could be used to characterise potential binding sites, in particular, to identify regions important to ligand selectivity. Chrysi Kirtay of the University of Cambridge presented a comparison of various docking scoring functions, including her own BLEEP scoring function, which performed favourably in relation to other scoring functions. Esther Kellenberger of the University of Strasbourg described the generation of the scPDB – a collection of structures of potential drug binding sites, derived by extracting the structures of all the binding sites of drug-like ligands from the Protein Data Bank. The scPDB can be used for docking and for analysing similarity between different binding cavities.

On Wednesday morning, more techniques to provide support for drug discovery, generating compounds, predicting their properties or toxicity, and selecting them for biological testing were described.

A. IJzerman from Leiden University presented a de novo design tool based on an evolutionary algorithm. The 'molecular evaluator' is not based on fragments mutations like other methods but rather on one-atom/one-bond mutations and is therefore supposed to cover a larger area of the chemical space. The fitness function is also interactive and allows the user to decide the ease of synthesis and the interest provided by the generated molecules.

Discarding compounds with undesirable properties as early as possible is a major goal of chemoinformatics. The toxicity of molecules can be predicted using the fragment-based method MC4PC, conceptually similar to substructural analysis and developed at the Case Western University by G. Klopman. The oxidative metabolism can also be predicted using QSAR techniques with the Bioprint database developed at Cerep Inc; the database includes over 2000 marketed drugs which have been screened across a panel including the inhibition of 8 CYPs. At Bayer Healthcare, the solubility of molecules is predicted using 10 ANNs that have each been trained with 10% of a dataset containing more than 6000 in-house, consistently measured, experimental solubilities. This method is effective, relatively fast and robust. Compounds can also be discarded on the basis of their metabolic clearance. C. H. Arnby presented the method used to mine the MDL Metabolic database at AstraZeneca. Both substrates and reaction centres are represented by atom-based fingerprints, which are then used to predict metabolic sites of a candidate molecule.

At Novartis, A. Schuffenhauer compared several methods to measure the chemical complexity of libraries, of which the number of features, ie the count of Similog keys or circular substructures, is the most relevant. Measures derived from those count (eg the ligand efficiency) are used to ensure the suitability for screening since diversity selection is affected by the size

bias of the similarity coefficient and can hence affect the overall complexity of the screening libraries.

Clustering and ligand-based virtual screening methods are used at Unilever Research to identify the structural features characterising bitterness. S. Rodgers described the use of phylogenetic-like trees and Bayes classifiers to predict the bitterness taste of molecules. Using a database of 833 bitter compounds plus a few thousand background structures, Bayes classifiers were accurate in 80% of the cases, while the phylogenetic-like tree were successful only in 50%.

The conference outing on Wednesday afternoon was a boat trip across the IJsselmeer, a large lake created by damming the Zuider Zee, to the typical old Dutch town of Hoorn. Delegates were driven by coach to the port of Enkhuizen, where we boarded three boats and set sail on the IJsselmeer. Our boat, named the Willem Barentsz, was a large wooden sail boat built in 1913. After leaving the port some of us were asked to assist the deck hands with the raising of the sails, and we began a leisurely voyage across the lake. We spent a pleasurable half hour enjoying the small and attractive town of Hoorn. Back on the boat, we made the most of the drinks provided for us, with the Old Dutch Genever, and the Shipper Bitter proving to be a popular choice. Later, we dined down in the hull of the boat, courtesy of the Chemical Computing Group.

Given the festivities of the boat journey the night before, the turnout for the final morning was surprisingly good, with everyone on good form. The topic for the first half was 'The Analysis of Large Datasets', with Kimito Funatsu presiding.

Iain McFadyen from Wyeth Research explored the role of modern high throughput screening (HTS). He highlighted the need for changes from traditional HTS approaches and detailed several methods that can be used to improve HTS efficiency and increase hit quality and diversity. Keen to validate the Wyeth HTS approach, he presented plenty of positive results to support their methodology.

Robert Brown of SciTegic presented a Bayesian learning approach that uses fragment descriptors to construct activity class specific statistical models, allowing predictions of the likelihood of a molecules

Sunset over the IJsselmeer, as the delegates enjoyed a cruise and dinner on board

biological activity. He went on to show how the methods fragment-based nature allows interpretability and how such information could potentially be used to customise activity class selectivity.

Eleanor Gardiner of the University of Sheffield spoke on a very hot subject – scaffold hopping. Using graph-matching similarity measures applied to Reduced Graphs in similarity searches of the MDDR, she demonstrated that not only are such methods effective for active retrieval, but that scaffold hopping is achievable when using this approach, with the actives found being topologically distinct from the query molecules.

Closing the session was Dmitrii Rassokhin of Johnson & Johnson with an overview of their in-house cheminformatics system. This is a large platform encompassing an integrated structure and biological/chemical database, along with a whole array of modelling and mining applications.

The final session presided over by Matthias Rarey was on 'Bridging the Cheminformatics–Bioinformatics Gap', an issue of increasing importance for the future of drug discovery. Wolf Ihlenfeldt of the National Institutes of Health began with 'The Integration of Chemical and Biological Data: the NCBI PubChem Project'. The need to have access to large amounts of biological and chemical information, particularly screening data, is of prime importance to those attempting to model biological activity. However, the competitive and proprietary nature of the pharmaceutical industry is antagonistic to this need. In light of this, it seems that the PubChem project holds great promise. However, an interesting point raised by the audience was that while the NCBI's own screening effort will no doubt be standardised, the data submitted by other groups may not be, and so there is a need to encourage other groups to standardise or at least provide as much assay information as possible with the data they submit to the PubChem database.

Edith Chan of Inpharmatica followed with an outline of their StARLITe knowledgebase, which attempts to relate protein target choice with the chemical space to be explored and hence allows the user to select the most appropriate protein target and focus on the relevant chemical space.

Richard Jackson of the University of Leeds gave the final presentation of the conference, examining how the study of similarity between protein–ligand binding sites could be used to identify potential binding sites on novel proteins. The implications of this are interesting both in evolutionary terms and for the discovery of new targets for existing drugs.

Then all that remained was to hear the conference closing remarks of reflection and praise from ICCS vice chair Markus Wagener, and then off to an eagerly awaited last free lunch!

ChIN – A Unique Chemistry Web Guide

ChIN (http://chin.csdl.ac.cn/) is a comprehensive web guide or directory of internet chemical resources, which was started in 1997. It aims to be an authoritative tool for finding current, good quality chemical information on the internet for the chemical community in China and globally. The construction of the ChIN site has been fully enhanced as the Chemistry Portal, Chinese National Science Digital Library (CSDL) since 2002. More than 10,000 web sites or pages carefully selected by the ChIN team have been indexed, accessible both in Chinese and English.

As a chemistry web directory, ChIN has some unique features:

- **Resource classifications**: The resources indexed in ChIN can be browsed by two hierachical categories. One category is based on resource types such as databases, software etc. The other is based on chemistry disciplines including chemical engineering.
- **Navigation based on resource description**: ChIN is a collection of chemistry links and also a knowledge base of chemical information sources that is implemented by creating resource descriptions. A summary page is created for each resource indexed in ChIN, using different data models for different resource types in order to describe the resource properly. In addition, subjects/keywords and cross links to closely related summary pages are added to the summary page. Each summary page serves as the resource description, and as a navigator to related resources by clicking on its classification category paths, related links and hyperlinked subjects/keywords for automated searching of the ChIN database.
- **Resource rating**: Each resource added to the ChIN site can be rated according to its quality on five levels by ChIN editors. Resources with first and second level ratings are recommended in the Top Picks column. The resources under one category then can be listed according to rank. This may help users to find better resources more quickly when browsing the ChIN site.
- **Combined Search of Classifications and Title/Keywords**: Combined searching of the two classification category headings and title/keywords is possible in ChIN's Advanced Search. This can be very useful when one is interested in a specific type of resources on a topic. For such a combined search, one can browse the classification trees provided on the Advanced Search page and click a heading so that the full heading will appear automatically as input in the current searching field box.

ChIN has become a widely recognised chemical web guide in China. Ten books have been published since 2000 in Chinese to introduce or recommend ChIN. ChIN was introduced in chemical information related courses for undergraduate and graduate students in some universities in China. ChIN has been recommended and linked by top research institutes, universities, scientific libraries and organisations in China, such as the website of the Chinese Chemical Society, the Chemistry Division of NSFC (National Natural Science Foundation of China) and the National Library of China. Since 1998, the total successful requests for pages is over 19,000,000.

Xiaoxia Li,
Chinese Academy of Sciences

Practical Chemoinformatics Training Course

As a recipient of a CSA Trust bursary to attend the Practical Chemoinformatics course at the University of Sheffield, I would like to express my thanks for this award, which has been of significant help in extending my knowledge of molecular modelling and chemoinformatics.

As a first year PhD student in the Molecular Design Group at Trinity College Dublin, my research is focused on delivering a novel utility to facilitate rapid flexible superpositioning of molecular pharmacophore features, volume and shape descriptors to enable high-throughput large database searching in ligand-based drug discovery. Currently I am exploring tools for 2D database searching and data analysis of drug molecules, such as the Daylight toolkit. The first day of this course provided a very good introduction to 2D similarity searching, and practical exercises in substructure searching with Daylight software really helped me to explore and familiarise myself with some new utilities of the Daylight software.

During the lecture and practical exercises in the Pharmacophore and Combinatorial Libraries sessions, I realised how I might use pharmacologically relevant features of an active ligand for in silico virtual screening, to identify alternate chemotypes with equivalent potency and mechanism of action, as a new approach in my research project.

Coming from a bioinformatics background, the content of this chemoinformatics training course covered exactly the topics which I am currently learning, and the software used in this course was closely related to my research project. I also learned a lot and received much valuable advice from talks with other delegates who came from industry. The knowledge I learned from this course also has been shared with my colleagues in my research group. I consider this training program has been highly successful and very much needed. I hope this course will be continued to be successful in the future and that more students may benefit from it.

Yidong Yang
Molecular Design Group
Department of Biochemistry
Trinity College Dublin
Ireland

SPONSOR'S SPOTLIGHT

FINGER ON THE PULSE: Wendy Warr & Associates

Since January 1992, Wendy Warr & Associates (http://www.warr.com) has been supplying business and competitive intelligence services to a broad spectrum of clients in the United States, Europe, Australia, the Middle East, and Asia. Our success stems from our extensive network and our specialised knowledge of cheminformatics and high throughput chemistry. Pharmaceutical companies, venture capitalists, publishers, software companies, and scientific database producers have benefited from our expert counsel and services in recent years.

Pharmaceutical and chemical companies:

- Multi-year contract with a pharma for current awareness services on drug discovery systems, including ADME prediction and virtual high throughput screening.
- Competitive intelligence study on research information systems for a major pharmaceutical company (just one of many CI studies).
- Design of training course on combinatorial chemistry for sales staff in an analytical chemistry company.
- Presentations on integration of bioinformatics and cheminformatics.
- *Strategies for Improving Pharmaceutical R&D Productivity* written for Decision Resources.
- *Directory of Molecular Diversity Suppliers* written for Select Biosciences.

Software companies and publishers:

- Major study on trends in several types of research information systems for a software vendor.
- Market research study for a major scientific publisher on the future of a major reference work.
- Advice on trends in chemical reaction databases.
- Competitive intelligence study for a company in the data mining field.
- Advice to a cheminformatics start-up on user requirements and pricing.
- Marketing and PR strategy for a company strong in databases and computation but weak in market placement.
- Intelligence and current awareness services to a consultancy specialising in trends in electronic publishing.

For venture capitalists and management consultants:

- Advice to venture capitalists on strategic information sources necessary to the pharmaceutical and biotech industries.
- Work with venture capitalists on licensing of a software solution.
- Advice to financial analysts on future prospects for two specific bio/pharma companies.

Publications:

- Our six-monthly, cutting edge reports on the state of the art in chemical information and computation (http://www.warr.com/more pubs.html) are renowned.
- Our CEO and namesake, Wendy Warr, is in constant demand the world over for presentations on cheminformatics.

For more information, please call:
Dr. Wendy A. Warr
Wendy Warr & Associates
6 Berwick Court, Holmes Chapel
Cheshire, CW4 7HZ, England
Tel./fax +44 (0)1477 533837
wendy@warr.com http://www.warr.com

EVENTS 2005–2006

August		
14–19	40th IUPAC Congress, Beijing, China.	http://www.ccs.ac.cn/IUPAC2005.htm
24–26	11th Asian Chemical Congress, Seoul	http://www.11acc.org/
28 August – September 1	230th American Society Meeting, Washington DC	http://acswebcontent.acs.org/nationalmeeting/dc05/home.html
September		
11–14	Biomolecular Simulations: from Prediction to Practice MGMS International Meeting 2005 Trinity College Dublin, Ireland	E-mail: mgms2005@tchpc.tcd.ie http://www.tchpc.tcd.ie/mgms/
October		
5–7	Euro-MUG 2005: Daylight User Group Meeting Cambridge Garden Moat House, Cambridge, UK	E-mail: peter@daylight.com http://www.daylight.com/meetings/f_emug.html
16–19	The 2005 International Chemical Information Conference and Exhibition, Nîmes, France	E-mail: contact@infonortics.com http://www.infonortics.com/chemical/index.htm
21–26	International Computational Methods in Sciences and Engineering 2005 (ICCMSE 2005), Hotel Poseidon Resort, Loutraki, Korinthos, Greece	E-mail: secretary@ieccs.net
29 October – November 1	3rd International Symposium on Computational Methods in Toxicology and Pharmacology Integrating Internet Resources (CMTPI-2005), Shanghai, China	http://www.sioc.ac.cn/CMTPI2005/
November		
13–15	1st German Conference on Chemoinformatics, 19th CIC-Workshop, Goslar, Germany	http://www.cic-workshop.de/
14–16	AccelrysWorld 2005, Thistle Tower Hotel, London, UK	http://www.accelrys.com/accelrysworld/
29 November–December 1	Online Information, Olympia, London, UK	http://www.online-information.co.uk
December		
15– 20	Pacifichem 2005 (International Chemical Congress of Pacific Basin Societies) Honolulu, Hawaii, US	E-mail: Pacifichem@comcast.net http://www.pacifichem.org
2006		
March		
5– 8	IPI-ConfEx 2006, International Patent Information Conference and Exposition, Athens, Greece	Trudi@IPI-Institute.com www.IPI-ConfEx.com
26–30	ACS Spring Meeting, Atlanta, Georgia, US Includes CSAT/CINF symposium, 'De Novo Design and Synthetic Accessibility'	http://www.chemistry.org/portal/a/c/s/1/acsdisplay.html?DOC=meetings%5cfuture.html

Product News

Text Analytics Collection on Pipeline Pilot

Notiora's Text Analytics collection for Pipeline Pilot from SciTegic includes over 50 new Pipeline Pilot components for the search, characterisation, and mining of text resources. With these components, local and third party data sources can be searched with a single universal query, returning the search results to Pipeline Pilot in a common structure. The suite presently supports online engines including: Google, PubMed, NIH Crisp Awarded Grants, and PatFT US Patent search. Trends and correlations in the scientific literature or with experimental results can be discovered, including for example, new correlations between diseases and genes or compounds of interest. There is further information at http://www.notiora.com/scitegic.shtml

Ogham and LexiChem from Open Eye

OpenEye Scientific Software has released Ogham and LexiChem, state-of-the-art programs that handle the 2D rendering, naming of compounds and the generation of structures from names. Ogham, named after the ancient Irish pictographic alphabet of lines, hashes and wedges, is designed to render 2D Kekule drawings from 3D structures or connectivity information. Ogham offers a variety of display controls such as colour, size, optional use of superatoms, or style of representation for aromatic and dative bonds. Ogham provides a consistent and specifiable orientation of core substructures, making it easy to spot common and alternate moieties in a series of compounds.

LexiChem provides efficient conversion from chemical names into chemical structures and vice versa. Compounds can be named in a number of styles including fully systematic, preferred IUPAC 2005, acceptable common-usage IUPAC, CAS naming conventions, and non-IUPAC traditional common names. Both LexiChem and Ogham are available either as standalone applications or as C++ toolkits for in-house or third-party software development purposes. For further information please contact: business@eyesopen.com

STN AnaVist

Chemical Abstracts Service and FIZ Karlsruhe have developed STN AnaVist as an advanced tool to help analyse search results from scientific literature and patent databases, and to visualise patterns and trends. Searchers can assimilate and present information more effectively to support competitive intelligence, research and development strategy and management decision-making. Key features of STN AnaVist include:

- Analysis of information from three leading scientific resources: CAS' CAplus database of scientific literature and patent information and the USPATFULL and PCTFULL patent databases containing the full text of U.S. and Patent Cooperation Treaty (PCT) patents;
- An interactive workspace displaying a range of data visualisations, dynamically integrated, including cluster and contour maps, histograms, and co-occurrence matrices;
- Harmonisation and standardisation of data prior to visualisation using algorithms based on intellectual data analysis;
- A data grouping feature that helps to minimise scattering of results by permitting editing and customising data elements across the databases;
- Flexible tools for deeper analysis of data subsets;
- Two options for creating results for analysis: integrated concept search capability within STN AnaVist, and import of search results into STN AnaVist from STN Express.

There is further information at http://www.cas.org/stnanavist/

STN Express with Discover! Analysis Edition (Version 8.0)

New features in this release allow information professionals to present their search results in improved reports or tables while also boosting their ability to search, analyse, visualise and review scientific information. The Variable Group (R-Group) Analysis Tool can analyse multi-component chemical substances and improve researchers' ability to preview substances to ensure the inclusion of relevant structures from the CAS Registry. Several other enhancements have been added for chemical structure analysis and reporting. New capabilities include:

- CAS Registry Number and Role Report Wizard provides an analysis of chemical substances and their corresponding CAS roles to find new areas of research
- Analyze Plus Wizard provides more analysis options and editing capabilities
- More secure internet connection with Telnet encryption option
- Automatic upload of individual gene sequences to DGENE and PCTGEN sequence databases, which makes searching easier
- Post-processing tools provide flexible cell formatting in tables for better client presentations
- An ability to export search results for use with the new STN AnaVist software.

There is further information at http://www.cas.org/ONLINE/STN/discover.html

InfoChem Cartridge and DialogLink 5

InfoChem GmbH and Dialog have joined forces to provide users of the Dialog online service with chemical structure searching of more than 10 million chemical compounds. The chemical structure searching feature is powered by the InfoChem Cartridge, a tool developed by InfoChem to integrate chemical structure and reaction retrieval in relational database management systems. The application uses the InfoChem Fast Search Engine to handle and search within millions of structures and reactions. Through its alliance with InfoChem, Dialog also is offering substructure highlighting making it easier for searchers to locate their structure within search results. Once a chemical compound is identified, Dialog users may quickly link to related research sources, such as patent documents, articles published in scholarly journals, news reports about manufacturers and an extensive range of other content options. There is further information at http://www.infochem.de and http://www.dialog.com.

Chemical Information for the Non-Chemist

Tuesday, October 11 2005

Royal Society of Chemistry, Thomas Graham House, Science Park, Milton Road, Cambridge, CB4 0WF, UK

Following the success of the one-day training course last October in London, another training day has been organised jointly by the CSA Trust, the RSC-CIG and USTLG (the University Science and Technology Librarians Group)

Programme

10.00 – 10.30am Registration and Coffee

- What makes Chemical Information different – David Walsh (Pfizer)
- Searching for Chemicals on the Internet – Teresa Loughbrough (Unilever)
- Chemical Structure Drawing Packages – Don Parkin (Daresbury)

12.30 – 13.30pm Lunch

- Overview of Structure Databases on STN – Jan Davies (STN)
- How to search for chemical structures – Substructure searching – Barry Dunne (CAS)
- RSC Information Services - including Knovel – Nazma Masud (RSC)
- Introduction to Reaction Searching – Jeanette Eldridge (AstraZeneca)

16.30pm Conclusion

For further information, or to book, please contact either of the following:
Doug Veal doug@dovertonltd.freeserve.co.uk
Peter Nicholls hdspeter@aol.com

Thanks to our Corporate Sponsors

AstraZeneca
Barnard Chemical Information
Cambridge Crystallographic Data Centre
Chemical Abstracts Service
GlaxoSmithKline
Hampden Data Services
Kilmorie Consulting
MDL Information Systems Inc. (San Leandro, US)
Omniviz Inc.
Pfizer
Questel.Orbit
Sheffield Chemoinformatics Research Group
Syngenta, Jealott's Hill
Thomson Scientific
Wendy Warr & Associates

Contact us

Chair of Trustees
Dr Guenter Grethe
352 Channing Way
Alameda
CA 94502-7409, USA
Tel/Fax: +1-(510)-865-5152
E-mail: ggrethe@comcast.net

Secretary
David J Walsh
Pfizer Global Research and Development
Sandwich
Kent CT13 9NJ
Tel: +44-(0)1304-645619
E-mail: david.j.walsh@pfizer.com

Donations Secretary and Website Administrator
Dr John Holliday
Department of Information Studies
University of Sheffield
Sheffield, S10 2TN, UK
Tel: +44-(0)114-2222685
E-mail: j.d.holliday@sheffield.ac.uk

Treasurer
Dr Geoff Downs
30 Kiveton Lane
Todwick
Sheffield, S26 1HL, UK
Tel: +44-(0)1909-515444
E-mail: geoff@bci.gb.com

US Donations
Dr Peter Rusch
162 Holland Court
Mountain View
CA 94040, USA
Tel: +1-(650)-961-8120
E-mail: pfrusch@aol.com

Chair – Public Relations
Janet Ash
UK
Tel/Fax: +44-(0)1580-852270
E-mail: janet_ash@btopenworld.com

Chair – Grants
Bonnie Lawlor
276 Upper Gulph Road
Radnor
PA 19087, USA
E-mail: chescot@aol.com

Joint Chairs – Meetings and Training
Peter Nichols
HDS
E-mail: hdspeter@aol.com

Suzanne Pears
AstraZeneca
E-mail: suzannepears@astrazeneca.com

Chair – Fund Raising
Dr Bill Town
Kilmorie Consulting
Tel: +44-20-8699-9764
E-mail: bill.town@kilmorie.com

Published by the CSA Trust
Views expressed in the Newsletter are those of individual authors and not of the Editor

Newsletter Editor
Frances Daniel
UK
Tel: +44-(0)1949-844837
E-mail: francesdaniel@bottesford65.fsnet.co.uk

NEXT COPY DATE
September 30 2005