Chemical Structure Association Trust

NEWSLETTER

Summer 2007 Issue 15

In this Newsletter, we gratefully acknowledge the efforts of our Editor, Frances Daniel, whose dedicated service to the CSA Trust in creating our newsletters was recently recognised when she was awarded the 2006 Ernie Hyde Award, given for outstanding contribution to the Trust. We hope you enjoy this issue and look forward to many more edited by Frances, pictured right. See Page 2 for details of CSA Trust awards and grants.

CSA Trust Awards and Grants

One of the main objectives of the CSA Trust is to promote education, research and development in fields related to the handling of chemical structures, reactions and compounds. It is not surprising, therefore, that a substantial part of each Annual General Meeting is devoted to a discussion about the awards and grants administered by the Trust.

At the 2006 AGM, Trustees voted to award the Jacques-Émile Dubois Grant to Dr Rajarshi Guha from Indiana University. He receives $1300 to further his work on the development of chemoinformatics web services and the development of algorithms for ensemble feature selection and cluster enumeration. He will attend the 2007 Gordon Conference on Computer-Aided Design. Details of how to apply for a grant for 2008 are given on Page 8 of this Newsletter.

Every three years, the CSA Trust Mike Lynch Award is given in recognition of outstanding accomplishment in education, research and development activities related to systems and methods used to store, process and retrieve information about chemical structures, reactions and properties. Candidates are currently being considered for the next award which will be made at the 8th International Conference on Chemical Structures, to be held in Noordwijkerhout, the Netherlands, in June 2008.

A further award, made annually, is the Ernie Hyde Award. This is given to someone who has made an outstanding contribution to the work of the Trust. I'm sure all our readers will agree that this year's recipient is particularly well-deserving. The Ernie Hyde Award was given at the AGM to Frances Daniel, our Newsletter Editor, for all her dedicated work in producing this excellent, lively and informative Newsletter over the past 12 years. Thank you, Frances, and long may you continue!

Janet Ash

Gary Wiggins to retire

This year sees the retirement on October 1 of one of the leading figures in chemical information handling, Dr Gary Wiggins. Gary was head of the Chemistry Library and Chemical Information Center at Indiana University from 1976–2003, during which time he developed widely lauded courses in chemical information handling, and authored a seminal book for chemistry librarians, *Chemical Information Sources* (McGraw-Hill, 1991). Gary has recently and generously updated the information from the book and made it available freely on a wiki website (http://cheminfo.informatics.indiana.edu/cicc/cis).

In 2000, he was appointed Director of Chemical Informatics and Bioinformatics in the newly formed School of Informatics, where he instigated and developed a pioneering chemoinformatics educational program (see http://cheminfo.informatics.indiana.edu) and oversaw the creation of a new research group in the field. He is also noted for creating the widely read Chemical Information Sources Discussion List (CHMINF-L), and served as a trustee of the CSA Trust. He has been the recipient of numerous awards, including the ACS Herman Skolnik Award (1998), distinguished service awards from Indiana University, and very recently the ACS Patterson Crane Award (2007). His retirement was recently marked at the ACS National Meeting in Chicago (March 2007) with a symposium in his honour. Gary will be truly missed on all kinds of levels by his students, co-workers and other practitioners in the field, but we wish him a happy and long retirement!

David Wild

People and Places

Robert Parker

The new managing director of the RSC's Publishing Division is Dr Robert Parker, who has been promoted from his existing role as editorial director at the RSC.

Rudy Potenzone

Rudy Potenzone, previously with CambridgeSoft, is now WW Technology Strategist, Pharmaceuticals Enterprise and Partner Group, at Microsoft Corporation.

BioMed Central

BioMed Central has appointed two publishers to its management team to lead its expansion into new disciplines. Bryan Vickery joined the company as deputy publisher, with responsibility for the Chemistry Central portal, and for the development of a portfolio of open access journals in chemistry. Chris Leonard has joined as associate publisher and will be leading the development of PhysMath Central, a portfolio of open access titles in physics, mathematics and computer science.

15 years of Warr

Wendy Warr & Associates (http://www.warr.com) are celebrating 15 years of business. They are mounting some memorable items at http://warr.com/15years.html.

Frank Brown

In September 2006, Frank Brown (formerly of Johnson & Johnson) was appointed chief science officer at Accelrys.

Tripos Discovery Informatics

Jim Hopkins has been named chief executive officer of the newly private company Tripos Discovery Informatics. The Discovery Informatics business was recently purchased by Vector Capital from Tripos, Inc. and will function as an independent company. Dr Richard Cramer has been appointed chief scientific officer, and Dr Robert Clark, vice president of research.

233rd ACS National Meeting, Chicago, Illinois

Computer Assisted Drug Design: Reminiscing about the Future

A Symposium Honouring Yvonne C. Martin

Around 200 people attended the full day symposium honouring the retirement of Yvonne C. Martin, one of the leaders of the field of computer-assisted drug design, specifically in the area of Quantitative Structure-Activity Relationships (QSAR). She has contributed to the field with 135 publications, 7 books, and 7 patents. Her 1978 book entitled *Quantitative Drug Design* has received over 500 citations.

The symposium was jointly organised by the ACS's COMP and CINF divisions. The final talk was from Yvonne Martin. Sponsored by the *Journal of Computer-Aided Molecular Design* (JCAMD) the sessions provided breakfast and snacks, and a brief reception immediately following the completion of the technical program. The attendees were also provided with free copies of the special issue of JCAMD honouring Yvonne Martin.

The technical program contained the following papers:

- Johnny Gasteiger: Using knowledge on chemical reactions for drug design
- Philip Hajduk: Experimental and computational approaches to measuring compound reactivity
- Kent D. Stewart: Drug Guru: A new kind of computational tool for medicinal chemists
- Richard D. Cramer: Pushing the boundaries of 3-D QSAR
- Ki H. Kim: Outliers in SAR and QSAR: What are the possible sources
- Steven W. Muchmore: Do multiconformer queries enhance 3-D lead hopping?
- Dimitris K. Agrafiotis: Advances in conformational sampling
- Peter Willett: Bibliometric analysis of cheminformatics
- Robert D. Brown: Still searching for the perfect fingerprints
- Ingo A. Muegge: Virtual screening for new chemotypes using compound similarity measures
- Tudor I. Oprea: Lead-like, drug-like, or 'pub-like': How different are they?
- John Van Drie: Computer-aided drug design: The next twenty-five years
- Yvonne C. Martin: What I learned from a career in computer assisted molecular design.

Four of the papers are described below:

Using knowledge on chemical reactions for drug design

Johnny Gasteiger

Johnny gave an overview of how chemical reactions are involved in many different stages of drug design, starting with the biochemical pathways. He also described the changing priorities at different stages of drug design: for example, synthesising chemicals for their biological activity in the early stages; and focusing on the pharmacokinetic properties during the late stages.

With the notion that a better understanding of chemical reactions could help make the drug design process more efficient, Johnny presented methods which attempt to mimic the decision-making process of an organic chemist, by quantifying concepts such as partial atomic charges, or polarisability, inductive, resonance and steric effects. These methods focus on modelling of chemical reactivity, and are applicable to a broad range of chemical reactions; for example, analyses of large sets of chemical reactions from gas phase reactions to biochemical pathways. The methods include prediction of acidity values, prediction of regioselectivity in organic reactions, designing the synthesis of organic molecules and combinatorial libraries, providing information on enzyme-catalysed reactions, and drug metabolism.

Whereas much progress has been made in predicting chemical properties through statistical methods such as QSAR and pharmacophore-based models, little has been made in understanding chemical reactivity. This has delayed the development of predictive models for metabolism, for example. New 3D-screening methods help scientists identify many possible novel molecules that can bind to a given receptor site, but their synthetic feasibility has become an important consideration. Johnny's group also tackled this problem by developing a method for estimating synthetic accessibility by analysing the molecular structure for its complexity, degree of functionality, and number of stereocentres. It also provides a means for the comparison of the query structure through similarity to available compounds, and, by searching a reaction database, provides proposed reactions for synthesis of the query compound.

Pushing the boundaries of 3-D QSAR

Richard D. Cramer

Dick presented a brief historical account and survey of predictive 3D methods and provided several observations related to the use of these methods, raising some important questions. In his survey, he used 15 biological measurements taken from 11 publications and 11 distinct sets of structures. These 15 sets represent the early 11 publications that reported successful 3-D QSAR derivation, and provided Dick with a basis with common trends for making observations.

One of the observations that he presented was the use of r^2 or q^2 as the measure of merit for a 3-D QSAR. As cross-validation has become a standard method of providing the statistical validity of a predictive model, it has become widely used. Several authors questioned the reliability of q^2 as an indicator of the quality of a model for external predictions. Dick went even further to point out that the cross-validation may, in fact, discard useful SAR information. For example, an active compound with some unique features (different from other active molecules in the data set) on which its activity depends is typically eliminated from the analysis, whereas this compound should be "treated as an opportunity to be validated and exploited by further synthesis and testing, rather than an artifact to be discarded." Despite this limitation, q^2 is certainly a very important metric to help eliminate the risk of 'chance correlation' at a time we are facing increasing number of large datasets. One suggestion is to experiment with the use of more parameters than are typically justified by cross-validation.

Continued overpage

233rd ACS National Meeting, Chicago, Illinois

Advanced Mining of Life Science Information

A symposium co-sponsored by the CSA Trust

Chemical information is now being used in diverse and interesting ways to aid the data mining of life science information. The diversity of research in this area was reflected in the breadth of topics covered in this symposium, from the creation of ontologically rich journal articles to a detailed study of how models built on chemical descriptors can be used to aid in the discovery of HIV Integrase inhibitors.

The session was kicked off by Darryl León from Active Motif in Carlsbad, who gave an overview of literature text mining in drug discovery (including a discussion of semantic, syntactic and lexical ambiguity), and then proceeded to give fascinating examples of how existing software packages can be used to carry out a variety of analyses. Darryl's talk was followed by a presentation by Richard Kho of Inforsense, who described how pharmaceutically relevant predictive models can be built within a workflow environment. Next, Colin Batchelor described the Project Prospect at the RSC (www.projectprospect.org) which is making available onotologically rich versions of journal articles and abstracts in RSS feeds, including machine readable chemical structure information and chemical and biological ontological markup (see Page 11). Chihae Yang of Leadscope described data mining approaches which can be used to link chemistry and biology; in particular, methods for selection of HTS hits that combine bioassay and toxicity profiles. Tony Trippe demonstrated innovations at the Chemical Abstracts Service that allow for the visualisation of chemical space by organising and clustering large numbers of chemical substances. By integrating this with existing CAS software capabilities, novel kinds of mining can be carried out. Next, Hugo Villar of Altoris explored ways in which massive amounts of biological and chemical information can be sensibly organised in a manner which simplifies the process of analysis whilst retaining accuracy, through the use of a variety of tools and statistical approaches. The session was completed by two talks from Indiana University: Xiao Dong described research into service ontologies for chemoinformatics which permit web services to be classified and organised so that it may be possible in the future to allow automatic discovery, composition, invocation and execution of workflows which can answer complex life science queries; Rajarshi Guha described a tiered screening protocol for HIV Integrase inhibitors that can be used to virtual screen vendor databases for potential hits.

Overall, this was an inspiring session that showed that innovation in chemical information and chemoinformatics is far from dead. New developments in text mining, the semantic web, journal publishing, predictive models and techniques for handling large volumes of chemical and biological information are opening up all kinds of opportunities and challenges for future research in chemical information applied to the life sciences.

David Wild

Continued from Page 3

Outliers in SAR and QSAR: What are the possible sources?

Ki H. Kim

Ki presented an interesting concept derived from a basic assumption made in early QSAR and SAR studies that structurally similar active compounds bind to a receptor in a similar manner. Those active compounds with a different binding mode are usually eliminated from the training set as outliers, but some of these outliers may provide an important source of leads.

Ki explained sources of outliers as due to possible experimental (or even typographical) errors, a lack of certain descriptors or parameters, inappropriate calculation of parameter values used, inappropriate mathematical models, a different mode of mechanism, or a different binding mode.

Ki examined binding modes of individual ligands for 14 structural data sets, and observed that in 10 of the 14 data sets multiple binding of structurally close ligands occurred (despite the same binding site being exploited in each case). This is the first such study focusing on multiple binding modes, and the findings indicate a surprisingly large number of compounds were treated as outliers. Ki also indicated that outliers could present us with opportunities in drug discovery research.

Lead-like, drug-like, or 'pub-like': How different are they?

Tudor I. Oprea

As the focus in drug discovery has increasingly moved into new lead discovery through techniques such as HTS, the question of how similar lead compounds are to drugs has also become more important. These lead compounds are then subjected to certain processes before, perhaps, promotion to active drugs.

Using various databases such as MDDR, Molecular Libraries Initiative Small Molecule Repository (MLSMR), Nature Chemical Biology (NCB), and WOMBAT, Tudor extracted the lead compounds, the drugs, and active and inactive compounds from these databases. He then examined various properties including MW, molecular complexity, flexibility, the number of H-bond donors and acceptors, LogP, LogSw, and the number of rule of five criteria violation.

Within certain statistical confidence levels, based on the above listed properties, there were no significant differences between leads of known drugs and active compounds (and their inactive counterparts as well) from MLSMR and NCB. Despite the similarity in some of the physical properties, he concluded that active compounds, in general, were less complex, less flexible, and more soluble than drugs.

Osman Güner

3rd International Conference of Molecular Simulations and Applied Informatics Technologies

This conference was held in Hangzhou, the capital city of Zhejiang Province, China, in April 2007. The conference attracted about 600 registrations, almost double the number for the second conference, which was held in Beijing in September 2005.

The Hangzhou conference was a combination of an academic meeting and a user meeting of NeoTrident (http://www.neotrident.com/), the leading representative of Accelrys, MDL, Thomson and Thermo in China. The meeting was well organised with plenary and invited lectures in three sessions, with dedicated computational applications in life sciences, materials design and experiences in pharmaceutical industry.

Some eminent scientists from China were among the plenary speakers. Professor Kaixian Chen, the founder of Drug Discovery and Design Center (DDDC, http://www.dddc.ac.cn/), Shanghai Institute of Materia Medica, Chinese Academy of Sciences, reviewed new trends in drug discovery and progress in China. Professor Nanxian Chen from Tsinghua University, now serving as a member of the National Advanced Materials Committee of China, introduced his philosophy of 'problems and inverse problems'. Professor Chen proposed the inverse problems on the boson system, the fermion system and the lattice system, and he obtained a concise and unified solution for the inverse capacity problem which includes both Einstein's model and Debye's model as special cases. Chen also obtained a concise and general solution for the inverse lattice problem, which is important for establishing the interatomic potentials base. Professor Chen's work provides an important foundation for the performance prediction of complex materials.

Professor Chunting Zhang (http://tubic.tju.edu.cn/) from Tianjin University overviewed his Z curve method for recognizing protein coding genes in bacterial and archaeal genomes with highly accurate gene start prediction. Professor Xiaojie Xu from Peking University, one of the key promoters of the conference, reported 'components aggregation', a newly observed experimental phenomena in TCM (Traditional Chinese Medicine) that may offer some insight into TCM's multi-formula effects. Professor Xu's group is actively involved in ADME/T prediction and the mechanism of TCM effects.

Professor Baixin Liu of Tsinghua University discussed the theoretical modelling of formation of amorphous alloys. Dr Li Chen, Head of Research and CSO at Roche R&D centre in China, Professor Hualiang Jiang, also from DDDC, and Professor Han Zhou from the Research Institute of Petroleum Processing, SINOPEC, also presented plenary lectures. Some of the invited presentations deeply impressed the session's audiences. Professor Minbo Chen from the Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, questioned the theoretical bases when MD (Molecular Dynamics) are applied to biomolecules. Professor Luhua Lai of the Center for Theoretical Biology, Peking University, introduced her group's progress in network based drug design. More than 200 full papers were accepted for the conference proceedings, which suggests increasing interest in computational applications in China.

Xiaoxia Li
Institute of Process Engineering,
Chinese Academy of Sciences

UK PubMed Central (UKPMC)

UKPMC is a new online resource designed to promote free access to research and further biomedical discovery. It has now gone live and is freely available at http://www.ukpmc.mimas.ac.uk/.

Based on a US model, UKPMC provides free access to an online digital archive of peer-reviewed research papers in the medical and life sciences.

UKPMC is part of a network of PMC International (PMCI) repositories. PMCI is a collaborative effort between the US National Library of Medicine (NLM), the publishers whose journal content makes up the PMC archive, and organisations in other countries that share NLM's interest in archiving life sciences literature.

UKPMC is fully searchable and will provide context-sensitive links to other online resources, such as gene and chemical compound databases. Currently all documents in UKPMC are linked to databases hosted by the National Center for Biotechnology Information (NCBI) at the US National Institutes of Health (NIH). Over time, however, additional links to resources hosted in the UK and Europe are planned.

The first phase of the project involves mirroring the US PubMed Central database. UKPMC also provides a manuscript submission system (UKMSS) to enable researchers to deposit articles that have been accepted for publication in a peer-reviewed journal.

UKPMC is run by a partnership between the British Library, the University of Manchester and the European Bioinformatics Institute (EMBL-EBI). The British Library runs the service, promotes it to researchers, and offers support for those who want to include their research papers in UKPMC. The University of Manchester hosts the service, on servers based at MIMAS (Manchester Information and Associated Services), and will support the process of engaging with higher education users. EBI contributes its biomedical domain knowledge and text-mining tools to integrate the research literature with the underlying bioinformatics databases.

The set-up, maintenance and ongoing development of UKPMC is funded by a mix of UK governmental and charitable research funders.

Meet the New Trustees

The CSA Trust is pleased to announce that five new members have been accepted onto the Board of Trustees. These distinguished new members represent a wide variety of aspects of chemical structure handling, chemoinformatics and computer aided drug design, and we are delighted to have them on board.

Dimitris Agrafiotis

Dimitris K. Agrafiotis is Vice President of Informatics at Johnson & Johnson Pharmaceutical Research & Development. He also serves as Adjunct Professor of Informatics at Indiana University School of Informatics.

He earned a BS in chemistry from the University of Patras, Greece, in 1985, and a PhD in theoretical organic chemistry from Imperial College, University of London, in 1988, with Professor H. Rzepa. After post-doctoral training with Professor A. Streitwieser at the University of California, Berkeley, and Professor E. J. Corey at Harvard University, he joined Parke-Davis Pharmaceutical Research (now Pfizer) as a Senior Scientist in the Computer-Aided Drug Design group. In 1994, he moved to 3-Dimensional Pharmaceuticals where he focused on the development of intelligent computational tools for combinatorial chemistry and structure-based drug design, serving as Executive Director of Informatics.

Following the acquisition of 3DP by Johnson & Johnson PRD in 2003, he was appointed Senior Research Fellow and Team Leader of the Molecular Design and Informatics group at the US Research & Early Development site, and directed the development of ABCD, a new global informatics platform for J&J PRD. In his current role, he oversees all information management systems for research and early development worldwide, and is responsible for setting global information technology strategy and architecture for J&J PRD.

His research interests include computer-assisted drug design, combinatorial chemistry and organic synthesis, molecular diversity, QSAR, artificial intelligence, and software engineering. He is the author of more than 60 scientific publications and book chapters, and is a co-inventor of 3DP's proprietary DirectedDiversity technology. He has been a fellow of the Alexander Onassis Foundation, and serves on the Board of the International QSAR & Modelling Society, the Editorial Boards of the *Journal of Molecular Graphics and Modelling*, *Chemical Biology & Drug Design*, and *Cheminformatics*, and the Scientific Advisory Boards of the NIH Exploratory Centers for Chemoinformatics Research at Indiana University, University of North Carolina, and Rensselaer Polytechnic Institute.

Grace Baysinger

Grace Baysinger studied at the University of Michigan, Ann Arbor, where she earned a BS in Botany in 1978, and an Accredited Masters in Library Science in 1981.

She has worked as Reference Librarian, Natural Sciences Library, University of Michigan (1981–1984); Head Librarian, Chemistry Library, University of Michigan (1985–1989); Head Librarian and Bibliographer, Swain Library of Chemistry and Chemical Engineering, Stanford University (1989–present); and Head, Science and Engineering Resource Group Libraries, Stanford University (1999–2005).

Her current professional activities include Chair of the American Chemical Society's Joint Board Council Committee on Publications (JBCCP), Chair of *Chemical & Engineering News* Editorial Board, member of JBCCP's Copyright Subcommittee, consultant for the ACS Committee on Community Activities, Advisory Board member for *Science of Synthesis* (Thieme), Advisory Committee for the *CRC Handbook of Chemistry and Physics*, member of ACS CINF Education Committee, and member of Collaborative Working Group for ACS CINF and GDCh Computers in Chemistry.

Professional honours include Stanford's Marsh D. O'Neill Award, 1996, the Stanford University Libraries Distinguished Service Award, 1997 and the CINF Meritorious Service Award, 2004.

Her research interests are in end-user searching and the development of search interfaces, collection development and management, and issues associated with scholarly publications.

Jonathan Goodman

Dr Jonathan Goodman is a Senior Lecturer in the Department of Chemistry at the University of Cambridge, UK, Deputy-Director of the Unilever Centre for Molecular Science Informatics, and a Fellow of Clare College.

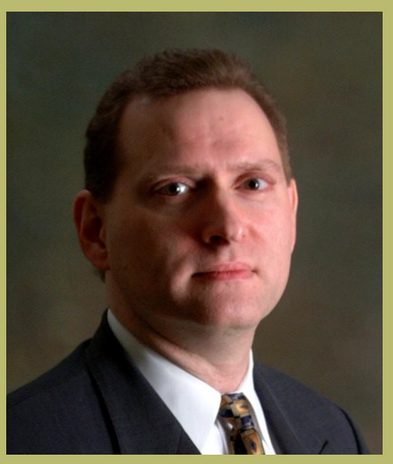
He did his PhD at Cambridge with Professor Ian Paterson FRS, using experimental and computational methods to study the aldol reaction. He then worked with Professor Clark Still at Columbia University developing the Macromodel molecular mechanics program and studying molecular recognition.

His current research interests cover experimental organic chemistry, molecular modelling and chemical informatics. His research has led to about a hundred publications and a book in these areas. Experimentally, recent work has included the development of asymmetric epoxidation reagents, and the investigation of solubility properties of drug-like compounds. Within chemical informatics, he set up one of the world's first chemistry WWW servers, and develops databases and resources that are used around the world. His computational studies of molecular structure and reactivity have led to widely used methods of conformational analysis and to the solution of puzzling aspects of organic reactivity.

Gregory Banik

Gregory M. Banik joined the Informatics Division of Bio-Rad Laboratories as the Division General Manager in May 1999. Dr Banik conceived and launched the Bio-Rad's KnowItAll Informatics System, which has won numerous awards, including Frost & Sullivan's Drug Discovery Technology Market Leadership of the Year Award in 2005, *Scientific Computing & Instrumentation* magazine's Reader's Choice Award in 2001, 2002, 2003, 2004 and 2006 and back-to-back R&D 100 Awards from *R&D* magazine in 2004 and 2005.

Dr Banik joined Bio-Rad from Molecular Simulations Inc. (MSI, now Accelrys) in San Diego, where he was Director of Business Development. He joined MSI in January 1998 from UMI (now ProQuest) based in Ann Arbor, Michigan. At UMI, Dr Banik was Director of Marketing and Product Management for UMI's Corporate and International Division. Prior to UMI, Dr Banik was New Product Manager, Chemical Information Products, at Thomson Corporation's Institute for Scientific Information (ISI) in Philadelphia, Pennsylvania. He began his professional career with Abbott Laboratories in North Chicago, Illinois, where he was Scientific Information



Manager, heading the cheminformatics group in the Pharmaceutical Products Division from 1990 to 1993.

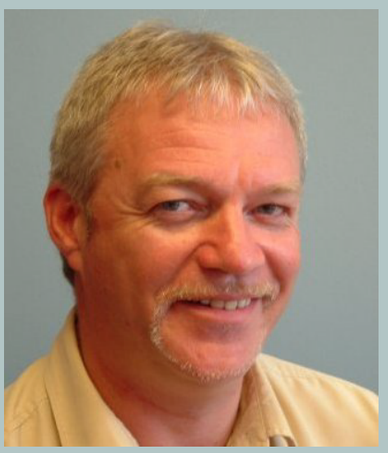
Dr Banik earned his PhD in Chemistry at Northwestern University in Evanston, Illinois in 1990 and was a Lecturer in the University's chemistry department that same year. He also received a Masters of Science in Chemistry from Northwestern and a Bachelors of Science in Chemistry and Computer Studies from Grinnell College in Grinnell, Iowa.

Since receiving his PhD, Dr Banik was also enrolled in MBA programs at the University of Chicago and Temple University in Philadelphia.

Bob Stembridge

Bob Stembridge graduated from the University of Sussex with an Honours degree in chemistry. He joined Thomson Scientific in 1980 and has held various roles in editorial, marketing, sales and product development. Leaving in 1988 for interludes working as Information Analyst specialising in patent analytics at British Petroleum and European Sales Liaison with Dialog, he returned to Thomson Scientific in 1996 and most recently became Customer Relations Manager with responsibility for liaison with customer user groups for the organisation.

He is a member of ACS, Patent Information Users Group (PIUG) and Patent and Trademark Group (PATMG) and has served on the PATMG management committee for a number of years. He is the PATMG immediate past-Chair. He was elected to the CSA Trust Board of Trustees in February 2007.



Applications Invited for CSA Trust Jacques-Émile Dubois Grants for 2008

The Chemical Structure Association Trust (CSAT) is an internationally recognised organisation established to promote the critical importance of chemical information to advances in chemical research. In support of its charter, the Trust has created a unique grant programme, renamed in honour of Professor Jacques-Émile Dubois who made significant contributions to the field of cheminformatics. The Trust is currently inviting the submission of grant applications for 2008.

Purpose of the Grants

The grant programme 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 $3,000. 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 programme 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.

Application Requirements

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. Three copies of the complete application document must be supplied for distribution to the Grants Committee.

Deadline for Applications

Applications must be received no later than October 19, 2007. Successful applicants will be notified by December 21, 2007.

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 blawlor@nfais.org or chescot@aol.com.

Introducing QSAR World

QSAR World is a free and comprehensive web-based portal for the Quantitative Structure Activity Relationship modelling community, facilitated by Strand Life Sciences.

The website can be found at www.qsarworld.com. This free forum for sharing ideas and knowledge via articles, posters, freeware and other media currently offers:

- Scientific literature on in-silico chemistry, machine learning and statistics
- Manually curated datasets extracted from peer-reviewed publications
- List of select keywords-based literature references and recommended classic articles
- Reports and articles about latest issues and developments in QSAR modelling
- Visitor's questions posted on the website and answered by experts
- A rolling calendar of all QSAR related events round the globe throughout the year
- Latest and relevant news
- Editor's choice articles, columns by QSAR gurus and interviews with thought leaders.

Suggestions for enhancements, articles and posters, questions on QSAR and modelling and any other relevant information for publication are warmly welcomed.

QSAR World is fortunate in having a distinguished Editorial Advisory Board comprising Dr Wendy Warr (founder, Wendy Warr & Associates), Dr Kalyanasundaram Subramanian (CSO, Strand Life Sciences), Dr Eric Jamois (VP, Business Development, Strand Life Sciences) and Dr Andreas Bender (Presidential Fellow, NIBR, Cambridge, MA).

SPONSOR'S SPOTLIGHT

InfoChem GmbH

Founded in 1989 and based in Munich, Germany, InfoChem has now more than 15 years of experience in the development and integration of sophisticated software tools for the storage and handling of structure and reaction information.

The InfoChem team consists of 15 full time employees and up to 60 freelance abstractors. The small dimensions and the broad network allow InfoChem to be flexible and react very quickly to the needs of customers.

InfoChem itself is a subsidiary of Springer Science+ Business Media. Since 1991, the publishing house Springer has held a majority interest (80%) in InfoChem GmbH. When Springer was taken over by Bertelsmann in 1998, InfoChem became part of the Bertelsmann group. Since September 2003, the specialist publishing group Springer Science+Business Media has belonged to the investment companies Cinven and Candover.

InfoChem's customers are large, global chemical and pharmaceutical companies, small, high potential start-ups worldwide, and major publishers.

Content

InfoChem distributes one of the largest chemical structure and reaction data collections worldwide (SPRESI). It currently contains 6 million organic and organometallic compounds, and 3.8 million reactions and associated factual data abstracted from 627,000 references, including 164,000 patents. It covers the chemical literature published since 1974. SPRESI was built jointly by the All-Union Institute of Scientific and Technical Information of the Academy of Science of the USSR (VINITI) and, until 1989, the German Zentrale Informationsverarbeitung Chemie, Berlin (ZIC).

The conversion of these files into an electronic searchable database and the development of a user-friendly web application accessible via the Internet (SPRESI^{web}) has been one of InfoChem's major achievements.

Services

In close collaboration with subsidiary InfoChimia GmbH (founded in Berlin in 1990), and with a network of highly trained abstractors, InfoChem offers data processing and database building of the highest quality. The most important example in this field is the conception and development of internet/intranet versions of printed major reference works such as Science of Synthesis (Thieme Verlag), e-EROS (John Wiley & Sons), and Comprehensive Asymmetric Catalysis (Springer).

Additionally, InfoChem has long-standing, successful project development and consulting collaborations with big pharmaceutical companies such as BASF AG, Bayer AG, E. Merck, and Boehringer Ingelheim, and with high potential start-up companies and non-profit institutions such as CAS, the European Patent Office and the German Umweltbundesamt (Federal Environment Agency).

Software Products

InfoChem offers a broad selection of software tools and applications to solve any problem in the field of chemical data management, focusing in particular on storage, handling and retrieval of structures and reactions.

Building on basic core technologies such as reaction mapping (IC*MAP*) and classification (CLASSIFY), R-group analysis and standardisation rules (IC*CHECK*), a high-perfomance search engine (IC*FSE*) and advanced chemistry handling (RetroSynthesis, IC*NameRXN*), and a series of indispensable tools (IC*EDIT*, IC*IMAGE*, IC*CONVERT*), InfoChem always has the ideal solution to meet the customer's needs.

Of crucial importance to our customers is the fact that all InfoChem products have been designed specifically to handle huge amounts of chemical data, without degrading the speed and performance of the system. The InfoChem Cartridge for Oracle (IC*CARTRIDGE*), designed to include chemistry data management in relational database systems, has proved to be the best performing reaction cartridge on the market.

More information can be found on the website at http://infochem.de/

EVENTS 2007–08

June		
18–20	CSA Trust/Molecular Graphics and Modelling Society 4th Joint Sheffield Conference on Chemoinformatics University of Sheffield, Sheffield, UK	E-mail: cheminf2007@sheffield.ac.uk http://cisrg.shef.ac.uk/shef2007/
25–29	Latest Advances in Drug Discovery Design & Planning Methods: a Hands-on 5 Day eCheminfo Advanced Training Workshop Week, Chemistry Research Laboratory, Oxford University, Oxford, UK	http://echeminfo.com/COMTY_training
July		
8 –11	AIMECS07 Congress, Istanbul, Turkey	http://www.aimecs07.org
August		
5–11	41st IUPAC World Chemistry Congress, Lingotto Conference Centre, Turin, Italy	E-mail: iupac.2007@unito.it; abotto@ mafservizi.it http://www.iupac2007.org
6-9	Drug Discovery and Development of Innovative Therapeutics, World Trade Center Boston and The Seaport Hotel, Boston, USA	http://www.drugdisc.com/
19–23	234th American Chemical Society Meeting and Exposition, Boston, USA	E-mail: natlmtgs@acs.org http://www.acs.org/meetings/
23–25	12ACC, Chemistry for Development, Environment and Sustainability in Asia, Kuala Lumpur, Malaysia	http://www.ikm.org.my/12acc.htm
October		
2–3	Drug Discovery: The Application of High Throughput Techniques to Target and Lead Identification, Nottingham, UK First Joint ELRIG/SBS meeting	http://www.elrig.org/
21–24	ICIC: International Conference for Science & Business Information, Barcelona, Spain (Formerly the International Chemical Information Conference)	E-mail: contact@infonortics.com http://www.infonortics.com/chemical/index.html
December		
4–6	Online Information 2007, Olympia, London, UK	http://www.online-information.co.uk/index.html

2008

June		
1–5	8th International Conference on Chemical Structures, Noordwijkerhout, The Netherlands	Contact details not yet available

Product News

DiscoveryGate adds Beilstein

Elsevier MDL has improved its online DiscoveryGate platform so that a single search lets researchers find integrated and linked information from the Beilstein Database and 22 other scientific databases covering bioactivity, synthetic methodology, chemical sourcing, drug safety, pharmacology and more. All 9.9 million chemical structures from the Beilstein Database have been added to the DiscoveryGate Compound Index, including 6.5 million unique structures. DiscoveryGate now covers over 27 million chemical structures, more than 17 million reactions and the world's largest collection of observed chemical properties, with well over 500 million experimentally measured values.

More information can be found at http://www.mdl.com.

STN Viewer

New from CAS and FIZ Karlsruhe is STN Viewer, a web-based workflow productivity tool helps researchers to manage and evaluate full-text patent documents in STN's collection of patent databases. STN Viewer provides:

- a powerful suite of tools designed for easy management of patent projects
- integration with STN, seamless access from patent searching to evaluation
- Direct access to information contained in STN's highly valued full-text patent databases reporting national (British, French, German, US), regional (European), and international (World Patent Organization) patent literature.

With STN Viewer, information and intellectual property professionals can create custom patent projects using full-text patent documents. They can easily evaluate patents, view major sections, publication stages and family members, define and highlight terms and phrases, and navigate the entire patent document via a visual summary of highlighted terms.

For further details, go to http://www.stn-international.de/ and http://www.fiz-karlsruhe.de/.

INPADOCDB

INPADOCDB (INternational PAtent DOCumentation Data Base) has over 63 million patents and utility models from 80 patent organisations, which can be accessed via STN International. Time coverage for the most important industrialised countries goes back to the 19th century. Database documents contain bibliographical information on the patent documents and are linked to the full texts. Patent family information, patent classification and legal status data are also available. Most documents contain abstracts in English providing speedy information on the described invention for which a patent has been filed.

INPADOCDB was created by merging and revising the EPO's INPADOC and DOCDB. The database obtained from this is offered by the EPO in a new format, giving STN the opportunity to review and reload the INPADOC file. The 49 million documents from INPADOC have now become part of the 63 million documents contained in INPADOCDB. Automated patent monitoring functions, the so-called Selective Dissemination of Information [SDI]/Alert Services, are also available in INPADOCDB.

Further information can be found at http://www.fiz-karlsruhe.de.

IO Informatics

IO Informatics, Inc. has extended Sentient Suite to deliver the next generation of informatics software products. The new Web-Query, Process Manager and Knowledge Explorer perform excellently as stand-alone products, but they also take great advantage of Data Manager to deliver a complete integration informatics research portal suite.

- Web Query helps users with no database expertise to query and browse multiple data stores from any location using a web browser and to save the results in a number of different formats.
- Process Manager is a complete workflow solution providing drag and drop modelling of the business process, with a web-based user interface to guide users through their tasks and graphical dashboards allowing managers instantaneous status reports on project progress.
- Knowledge Explorer leverages the latest semantic web technologies to model, visualise and explore data, revealing networks and novel relationships in complex data sets from one or multiple sources.

For more information, please email info@io-informatics.com

New SPRESI^{web} 2.5

This new version of SPRESI^{web} provides several new features: data from 1974–2005 are now available, the number of chemical supplier catalogues included has been increased, a quick search textbox enables searching of the basic index of structures, reactions and references with one keyword and a new EXISTS operator has been introduced. In addition, an alphabetical list of existent Name Reactions has been integrated in version 1.0 of IC$_{NameRXN}$. SPRESI^{web} 2.5 also offers a structure similarity search, a search for calculated chemical descriptors, and IC$_{EDIT}$, a structure and reaction drawing tool developed in-house by InfoChem.
The webpage is at http://infochem.de/en/products/spresiweb.shtml

Project Prospect

RSC Publishing, the publishing arm of the Royal Society of Chemistry, has launched 'Project Prospect' for its journals. Electronic RSC journal papers have been enhanced so that their data can be read, indexed and intelligently searched by machine. Readers will be able to click on named compounds and scientific concepts in an electronic journal article to download structures, understand topics, or link to electronic databases.

More details can be found at http://www.rsc.org/Publishing/Journals/ProjectProspect/index.asp

CAS celebrates 100th Anniversary

Throughout 2007, Chemical Abstracts Service (CAS) is celebrating its 100th year of serving chemists and other scientists by providing access to chemical literature and patents (http://www.cas.org/aboutcas/cas100/index.html).

Over the decades, CAS databases have reflected the progress of science and the explosion of published research around the world:

- 1907: William A. Noyes enlarged the *Review of American Chemical Research*, an abstracting publication begun by Arthur Noyes in 1895 that was the forerunner of *Chemical Abstracts*.
- 1965: Introduction of the CAS Chemical Registry System, which gave substances a CAS Registry Number for easy identification.
- 1980: CAS ONLINE was launched, making it possible for users to search the CAS Registry database.
- 1983: ACS and the German organisation, FIZ Karlsruhe, signed an agreement in 1983 to cooperate in forming an international online network. STN, the scientific and technical information network, was launched in 1984.
- 1995: CAS introduced the SciFinder research tool to allow access to CAS databases without needing to learn a command language.
- 2007: CAS databases today contain over 27 million bibliographic records from journal and patent literature, accompanied by more than 170 million citations. The Registry contains more than 31 million identified molecular substances, 1.5 billion predicted and experimental properties, and nearly 12 million reactions.

Thanks to our Corporate Sponsors

AstraZeneca
ChemIndustry.com, Inc.
Digital Chemistry
FIZ-Chemie
Hampden Data Services
InfoChem
Kilmorie Clarke Ltd
Elsevier-MDL
Questel.Orbit
Rusch Consulting Group
Sheffield Chemoinformatics Research Group
Symyx Technologies Inc.
Thomson Scientific
Wendy Warr & Associates

Contact us

Chair of Trustees
Dr Andreas Barth
Hermann-von-Helmholtz-Platz 1
D-76344 Eggenstein-Leopoldshafen
Germany
Tel: +49-7247-808-450
Fax: +49-7247-808-133
E-mail: andreas.barth@FIZ-Karlsruhe.de

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

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

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

Chair – Public Relations
David Wild
USA
Tel: +1-(812)-856-1848
E-mail: djwild@indiana.edu

Chair – Grants
Bonnie Lawlor
276 Upper Gulph Road
Radnor
PA 19087, USA
E-mail: blawlor@nfais.org

Chair – Meetings and Training
Vacant

Chair – Fund Raising
Vacant

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
August 31 2007