Chemical Information

BULLETIN

Spring 2012 Volume 64 No. 1

Division of Chemical Information

Chemical Information Bulletin

A Publication of the Division of Chemical Information of the ACS

Volume 64 No. 1 (Spring) 2012

Judith N. Currano, Editor
University of Pennsylvania
currano@pobox.upenn.edu

In This Issue

Cover design by Mark Luchetti
ISSN: 0364-1910

Message from the Chair

It is with a sense of excitement and trepidation that I start my tenure as Chair of the Division of Chemical Information. While the Division continues to be a vital component of the chemical information community, there are many issues that affect our field and our members. Terms such as "Big Data" and "Data Scientist" are all the rage now, and I believe that CINF is already representative of much of "data science." When I look at what the CINF membership does with chemical data - collection, curation, representation, management, and analysis - we're much more than just chemical information. Instead, I see the CINF community addressing the complexity and subtlety of chemical data by developing techniques, tools, processes, and policies that allow us to go from raw signals, numbers, and structures to actionable information. The broad scope of the Division allows us to easily span areas that may not be chemistry, but that need to be chemically aware. I see great opportunities for the Division and its membership by positioning ourselves to take advantage of the resurgent interest in data science.

We must also be aware of policies and advocacy efforts that affect our field. Specifically, the Open Data, Open Access, and Open Science movements have gathered strength over the last few years, and I believe that many of the issues and concerns that are the focus of these movements impact the activities of many CINF members. Given the interlinked nature of today's world, siloed data is a hindrance to modern research, and it is worrying that there are efforts afoot to hinder and block certain aspects of the Open movements. I believe that CINF, with links to chemistry databases, analytics and publishing, is in a unique position to support various aspects of the Open movements that directly touch our community. I hope that we will be able to make some concrete progress, in at least some of these areas, during the coming year.

How does this help you as a CINF member? The CINF leadership is actively discussing and developing strategies to make CINF more relevant in today's world: enhancing topicality of programming, working on local networking opportunities, and so on. But we need to hear from you, our members. What's missing? What could we do better? Who could we reach out to? I welcome anybody who is interested in getting involved with the Division to get in touch with me (or any other committee member). Obviously, the Division benefits from your involvement. Conversely, there's a lot for you to gain: have your say in how the Division is run and where it's going, and, most importantly, network with some very smart, very nice people. While there is ample scope to meet with CINF members at the San Diego meeting, don't restrict yourself to the National Meetings. The excellent, newly revamped CINF Website has contact details for the various committees, and we are all on the ACS Network, if you prefer sending messages that way. While we do not have an official presence on other social media networks, a number of our members are regulars on sites like Twitter, FriendFeed, and others. So, speak up, and you will be heard.

I am looking forward to an exciting year ahead and I hope to meet many of you in San Diego in a few weeks.

Rajarshi Guha, *Chair*
ACS Division of Chemical Information (CINF)

Letter from the Editor

Greetings from Philadelphia, where March is coming in like a lamb!

I usually find the concept of "spring semester" in Philadelphia to be somewhat oxymoronic. Annually, the semester begins with icy temperatures and snowfall in January and February and moves to "unseasonably" hot days in April and May. However, this year's mild winter has brought thoughts of spring early, and so it was not difficult to begin to envision the spring issue of the *Chemical Information Bulletin* prior to the vernal equinox. Like the fall issue, the main thrusts of the spring *CIB* are the events of the Spring ACS National Meeting. Once again, the combined efforts of David Martinsen, Graham Douglas, and Danielle Dennie have brought us a presentation of technical and social events at the upcoming meetings that should help everyone arrive in the right place at the right time!

If you will be in San Diego on Saturday afternoon, please consider attending a CINF committee meeting. Several of our committees, including the Communications & Publications Committee, the Education Committee, and the Program Committee will be meeting in open session on Saturday, and the chairs have provided agendas so that you'll know what to expect. The names and contact information for all of the CINF committee chairs appear at the end of the *CIB*, as well as on the CINF Web site (http://www.acscinf.org), so please feel free to contact any of them if you are interested in serving on a committee.

I realize that not all of our members will be attending the ACS National Meeting, so, we have tried to intersperse the meeting information with other features that should be interesting to many. In this issue, Svetla Baykoucheva presents a delightful interview with Andrea Twiss-Brooks from the University of Chicago, who has held a wide variety of positions within and outside of ACS. Bob Buntrock supplies us with two insightful book reviews. Finally, in a reincarnation of an old column, Song Yu draws our attention to articles of note that have appeared in the literature in 2011. We hope to continue the literature digest section in future issues, so, if you have authored or read an article that you think would be of interest to your fellow CINF members, please be sure to let me know!

I am trying to expand the feature section of the pre-meeting *CIB* to include even more content that will appeal to the broader membership. If you have an idea for a feature or would like to write a brief piece, please feel free to contact me at any point during the year.

Once again, I thank the army of volunteers who helped me to produce another *Chemical Information Bulletin* at an extremely busy time of year. I hope that you will enjoy their efforts.

Judith N. Currano
Editor, Chemical Information Bulletin
currano@pobox.upenn.edu

Awards and Scholarships

Call for Nominations: Herman Skolnik Award 2013

About the Award
The ACS Division of Chemical Information established the Herman Skolnik Award to recognize outstanding contributions to and achievements in the theory and practice of chemical information science. The Award is named in honor of the first recipient, Herman Skolnik. For a list of previous honorees, please visit http://www.acscinf.org/awards/skolnik.php

By this Award, the Division of Chemical Information is committed to encouraging the continuing preparation, dissemination and advancement of chemical information science and related disciplines through individual and team efforts. Examples of such advancement include, but are not limited to, the following:

- Design of new and unique computerized information systems;
- Preparation and dissemination of chemical information;
- Editorial innovations;
- Design of new indexing, classification, and notation systems;
- Chemical nomenclature;
- Structure-activity relationships;
- Numerical data correlation and evaluation;
- Advancement of knowledge in the field.

The Award consists of a $3000 honorarium and a plaque. The recipient is expected to give an address at the time of the Award presentation. In recent years, the Award Symposium has been organized by the recipient.

Nomination Procedure
Nominations for the Herman Skolnik Award should:

- describe the nominee's contributions to the field of chemical information;
- include supportive materials such as a biographical sketch and a list of publications and presentations;
- be accompanied by three seconding letters;
- be received prior to June 1, 2012;
- be sent by email to Andrea Twiss-Brooks, atbrooks@uchicago.edu

Please note: paper submissions are not accepted.

Division of Chemical Information

CINF Scholarship for Scientific Excellence

We are very pleased to announce that FIZ CHEMIE will sponsor the Fall 2012 CINF Scholarship for Scientific Excellence. Up to three scholarships, valued at $1,000 each, will be awarded at the 244th ACS National Meeting in Philadelphia, PA, August 19 – 23, 2012.

Call for Applications

Graduate and postdoctoral students in chemical information and related sciences are invited to apply for the 2012 CINF Scholarship for Scientific Excellence, sponsored by FIZ CHEMIE. The scholarship program of the Division of Chemical Information (CINF) of the American Chemical Society (ACS) is designed to reward young scientists who deliver outstanding contributions to the field of chemical information that aim at better use of knowledge to support research in chemistry and computer-aided preparation. Applicants must be enrolled at a certified college or university. The deadline for applications is March, 31, 2012, and a maximum of three scholarships of $1,000 each will be awarded at the 244th ACS National Meeting in Philadelphia, PA, August 19 – 23, 2012.

The scholarship is designed to honor scientific excellence and to foster students' involvement in CINF. Posters accepted for the competition will be presented during the Welcoming Reception of the division on Sunday evening at the National Meeting. Winners will be chosen based on content, presentation and relevance of the poster and they will be recognized during the reception. Winning posters will be marked "Winner of FIZ Chemie-CINF Scholarship for Scientific Excellence" at the poster session. Additionally, the applicants will have the option to also show their poster at the Sci-Mix session on Monday night.

To apply, please inform the Chair of the selection committee, Guenter Grethe at ggrethe@att.net that you are applying for a scholarship. Abstracts for the poster must be submitted electronically through PACS, the abstract submission system of ACS. The contents shall reflect upon the student's work and describe research in the field of chemical information and related sciences. Submit your abstract to http://abstracts.acs.org using your ACS ID. If you do not have an ACS ID, follow the registration instructions and submit your abstract in the CINF program for the session "CINF Scholarship for Scientific Excellence". PACS will be open for abstract submissions from January 23 to March 31, 2012. Additionally, please send a 2,000-word abstract describing the work to be presented in electronic form to the Chair of the selection committee ggrethe@att.net by June 30, 2012. Any questions related to applying for one of the scholarships should also be directed to ggrethe@att.net .

Talking about eScience, Libraries and Other Things:

An Interview with Andrea Twiss-Brooks, Co-director of the Science Libraries Division of the University of Chicago's John Crerar Library

By Svetla Baykoucheva and Andrea Twiss-Brooks

For many years, Andrea Twiss-Brooks has been one of the most noticeable members of the Chemical Information Division of the American Chemical Society (CINF). She has worked in the University of Chicago Library's Science Libraries Division since 1993, starting as Chemistry Librarian. As time progressed, she added collection development and reference services in other subject areas, including physics, geophysical sciences and the history of science, medicine and technology. Prior to pursuing a career as a chemistry librarian, Andrea worked as laboratory technician intern in industry, a teaching assistant in chemistry, project manager for an update of the Wiley-NBS Mass Spectral Database, and an editorial associate for the journal *Accounts of Chemical Research*. Andrea currently holds the position of Co-Director, Science Libraries Division, where she has administrative responsibilities for all collections in science, medicine, and technology at the University of Chicago.

Andrea received her B.S. in chemistry from Texas Christian University in Fort Worth and her M.S. in chemistry from Cornell University. These degrees were followed by a M.S. in library science from the University of North Texas. She is a member of the American Chemical Society and its Division of Chemical Information, the Geoscience Information Society, the Association of Academic Health Sciences Libraries, and the Medical Libraries Association, and she has served on a variety of advisory bodies for scientific and medical publishers.

Svetla Baykoucheva: *Andrea, we recently met with you at a conference on eScience in Phoenix, Arizona. This was the last event of a six-month long initiative from the Association of Research Libraries (ARL) known as the "eScience Institute." More than 70 universities, with three representatives from each institution, participated in this program. What are the main conclusions that you have drawn from the discussions at this conference? At what stage of involvement are most of the universities in eScience/eResearch?*

Andrea Twiss-Brooks: Our eScience Institute team gained a lot from the interaction with other universities' teams at the Phoenix event. The universities represented at the event in Phoenix are primarily large research institutions. We discovered that there is a broad range of involvement by libraries in supporting eScience (or, even more broadly, eResearch). Coming up to the event, we believed that collaboration with other stakeholders on campus is critical to the success of any eResearch support services that our Library wants to develop, and we were even more firmly convinced of this after discussions with various colleagues at the event. At the University of Chicago, the Director of Research Computing (organizationally located in the office of the Vice President for Research and National Laboratories) provided a senior staff member from his organization for our Library team, and we gained valuable insight into issues from this member. The development of a community of practice among institutions and their libraries will also be crucial. The continued sharing of information, best practices,

and innovative approaches among the members of the Institute should be one of the outcomes of this program.

SB: *How do you see the role that science librarians could play in eScience/eResearch? Will having a science background be helpful to deal with scientific data? What kind of skills do librarians need to develop to meet the challenges of this new area?*

ATB: Science librarians could play a number of roles in supporting eResearch. Roles that librarians might play in managing scientific data include consulting faculty on the creation of data management plans, selection and application of appropriate existing metadata schema, development of ontologies and metadata schema for scientific data in areas lacking existing schema, providing DOI assignments for improving citability of datasets, managing institutional data repositories for discovery and sharing of datasets, and designing long-term preservation archives for scientific data. There are several models that are already emerging at various institutions. After the Phoenix event, I had an opportunity to visit with colleagues at the University of New Mexico, where they have hired several data librarians, including a life sciences and an engineering data librarian. The science backgrounds of these data librarians are an important asset, allowing the librarians to discuss research data and the research process knowledgeably with their faculty. While librarians have some skills that could be adapted to the needs of eResearch support, we will also need to develop new skills in our staff, including a better understanding of disciplinary research techniques and the structure of the resulting data, knowledge of XML and other important standards, programming and scripting language ability, and ability to communicate effectively with both researchers and technical staff involved in providing the eResearch support services.

SB: *What are the technical challenges in accessing, assessing, managing, and using data? How could such problems as variable data quality, coupling heterogeneous data together, and inadequate metadata be solved?*

ATB: One of the biggest challenges is figuring out where to start. Our researchers are pumping out huge amounts of data, in every conceivable format. I think we have to think strategically, but act tactically. If we wait until we have a "perfect" solution, we won't do anything. Whatever the technical solutions may be,

Positions held by Andrea Twiss-Brooks in the ACS Chemical Information Division (CINF)

Divisional Web Site Creator/Coordinator (1995-2006)

Division Chair (2001)

GDCh CIC – ACS CINF Collaboration Steering Committee (2005-present)

Bylaws and Procedures Manual Committee, Chair (2003-2006)

Awards Committee (2007-present), Chair (2012-2015)

Program Committee (2008-present)

Division Councilor (2006-present)

Positions related to or as a result of Andrea Twiss-Brooks' various roles in CINF

ACS Joint Board-Council Committee on Chemical Abstracts Service

ACS Committee on Community Activities

ACS Committee on Nominations and Elections

Organized or co-organized and/or presented at a number of CINF technical symposia

Editor, Special Topics in Intellectual Property *(*ACS Symposium Series *1055), Oxford University Press, 2011 based on a CINF technical symposium*

Feature editor, "Chemical Information Instructor," Journal of Chemical Education *(2002-2009)*

I also think it's going to be critical to work with the researchers to provide solutions that fit into their existing workflows as much as possible. I don't have too many answers to this question, but there are a lot of smart people out there working on various aspects—collaboration and partnerships on our campuses and between institutions is going to be the key to finding sustainable solutions.

SB: *What kind of new services may researchers need in order to preserve, manage, or find scientific data? At what stage of the research process do you think librarians could have an input in this process?*

ATB: I think that librarians should have input from the beginning. We have already been consulted by a few of our faculty during the grant writing phase of the process. This has been a terrific opportunity not only to be involved in thinking through the data management planning, but also to educate researchers in what librarians can do for them. We should have a role in educating the future researcher by offering instruction and outreach to our graduate student populations. Librarians are well placed to provide the framework for data management principles, understanding of rights and responsibilities, introduction to concepts such as data curation, and more. When we interviewed faculty as part of the eScience Institute work, we heard, over and over, that the library is seen as trusted, neutral, and stable. This places libraries in a unique position, with our communities looking to us to provide long-term archiving and preservation solutions that serve the entire community.

SB: *What kind of developments could force a shift from a "provenance" scientific culture to one of shared data, open access, community models?*

ATB: That's a hard one. While our current system of determining tenure and promotion remains entrenched, it's hard to envision a culture of more sharing, open access and community models. In some disciplines we do see more sharing and more openness, but it seems to me that we need a "carrot" to encourage more sharing, and that the reward for more sharing and more openness needs to be tied to academic success.

SB: *Now you have higher responsibilities, but for many years you were a chemistry librarian. What does your current job entail?*

ATB: I am actually still a chemistry librarian, but these days I also have more administrative responsibilities. Except for the substantial increase of time spent in meetings, I have to say I find it very rewarding. I am responsible for overall management of all collections-related activities in the science libraries at the University of Chicago, including the John Crerar Library. In this capacity, I have been fortunate to have been involved in the planning and implementation of our Google book scanning effort in the sciences, a two-year effort that we will wrap up this summer. Because many gifts to the Library focus on collections, I have had increasing opportunities to interact with donors and potential donors in a variety of contexts. I'm also a member of the senior management team for the Library which works together to provide a long term vision for the role of the Library in the future. This planning includes not only thinking about services, budgets, and staff development, but also space planning.

SB: *How will the many organizational and technological changes currently happening affect the field of chemical publishing and chemical information?*

ATB: We've already seen a dramatic change in publishing and chemical information in the last decade. I'm not sure I can predict what the next decade holds for publishing, but I'm sure it's going to be just as exciting as the last one. A lot will depend on how quickly shared data, open access, and community models we talked about earlier take hold in the chemistry community.

SB: *You have played many roles in the ACS Chemical Information Division (CINF). How did you get actively involved with the Division, and which of the many positions you have held has been the most satisfying to you? Could you tell us an interesting story or fact related to CINF that might not be known to our readers?*

ATB: I originally became involved with CINF as the creator of our very first Website, or as I preferred to style myself in those days, the Web diva. I don't want to say how long ago that was, but the first CINF Web page was hand coded in plain HTML and hosted on the library's web server. Through my work as the Web diva, I interacted with the officers, committee chairs and membership of the Division. It's hard to say which position has been the most rewarding, since each position had its unique aspects. I have enjoyed being Division Councilor for the past several years, primarily because of the interaction with the larger ACS organization as well as the opportunity to raise awareness of CINF as a Division among other Councilors and members of ACS. I don't know if I have an interesting story about CINF, but I do think it's a fact that our Division is the friendliest Division in ACS. Networking with colleagues is one of the most important benefits that I get out of my CINF membership; CINF is a collection of some of the smartest people I know, and I get lots of great ideas and good advice from them.

SB: *If you are to consult those involved in the curriculum of library schools, what advice would you give them? If you need to hire a librarian in the near future, what skills are you going to look for in the candidates?*

ATB: I think that library schools need to be flexible in their curricula, while not abandoning the basic principles of librarianship. New librarians need more technical knowledge today, but they also need to understand broad principles and problems. They need to be able to apply principles of metadata, information organization, enhancing discoverability of information, and curation, preservation, and archiving principles and practice. One of the most important skills we look for in candidates is actually a non-technical skill, and a hard one to define: potential. In my estimation, a candidate with potential should have at least some of the technical knowledge needed, but demonstrate a strong willingness and an ability to continue to learn and develop, an intense curiosity about the institution's research and learning activities, excellent communication skills, and flexibility/adaptability. We don't know exactly what challenges we'll be facing next, and having competent staff who are not afraid to try something new is one of the keys to our future success.

SB: *Tell us something about yourself. What hobbies (if any) do you have? What books do you read? What particular interests do you have? In most of the interviews that I have done for the* Bulletin*, I have often asked this question, and I have always been surprised by the answers I got.*

ATB: I am an avid reader, mainly of mysteries with historical settings (I'm particularly fond of Ruth Downie, Charles Todd, and Anne Perry), although I do belong to a book group where the selections serve to "stretch" me a bit. One recent book and discussion I particularly enjoyed was "Angle of Repose," by Wallace Stegner. I'm also a fan of debut novels; I feel like these books are the ones that the authors "had" to get written, and I don't have to live up to any preconceptions of their previous works. I enjoy hiking and nature and was thrilled to have an opportunity to try one of the trails in Phoenix's South Mountain Park following the eScience Institute event. Recently, I have been spending quite a lot of time quilting. It provides me with a creative outlet that results in an actual physical product, which is a nice counterpoint to my professional activities, which usually result in something a bit less concrete. If anyone is interested in seeing my projects, I've just launched a new blog: http://andreasquilting.wordpress.com

SB: *Thanks, Andrea, for discussing eScience and other questions that I am sure will be of great interest to the readers of the* Bulletin.

Svetla Baykoucheva
University of Maryland
College Park, MD
sbaykouc@umd.edu

Andrea Twiss-Brooks
University of Chicago
Chicago, IL
atbrooks@uchicago.edu

Contributors, Volume 64 No.1

The following individuals contributed time or material to this issue of the Chemical Information Bulletin.

Articles and Features
Svetla Baykoucheva
Robert E. Buntrock
Rajarshi Guha
Song Yu

Committee Reports
Rachelle Bienstock
Chuck Huber
Bonnie Lawlor
Bill Town
Andrea Twiss-Brooks

Sponsor Information
Graham Douglas

Technical Program
David Martinsen

Production
Judith N. Currano, *Editor*
Danielle Dennie, *Webmaster*
Mark Luchetti, *Cover Design*
Svetla Baykoucheva, *Copyeditor*
Svetlana Korolev, *Copyeditor*

Literature Digest

The "Literature Digest" section of the Chemical Information Bulletin *is designed to highlight recent articles by or of interest to the CINF membership. If you have recently authored an article, please tell us about it! If you have read an article that interested you and that you think would interest others, we would be delighted to hear about it, as well. Just send us the reference, and we'll be happy to review it for inclusion in this column!*

The following articles were published since 2011, and were assembled by Song Yu in the month of February.

Backman, T. W. H., Cao, Y. Q., & Girke, T. (2011). ChemMine tools: an online service for analyzing and clustering small molecules. *Nucleic Acids Research, 39*, W486-W491.

Baykoucheva, S. (2011). Comparison of the Contributions of CAPLUS and MEDLINE to the Performance of SciFinder in Retrieving the Drug Literature. *Issues in Science and Technology Librarianship*(Summer 2011).

Bjork, B.-C. (2011). A study of innovative features in scholarly open access journals. *J Med Internet Res, 13*(4), e115.

Castro, E. A., & Haghi, A. K. (2012). *Advanced methods and applications in chemoinformatics : research progress and new applications* Hershey, PA: Engineering Science Reference.

Choi, J. Y., Bae, S. H., Qiu, J., Chen, B., & Wild, D. (2011). Browsing large-scale cheminformatics data with dimension reduction. *Concurrency and Computation-Practice & Experience, 23*(17), 2315-2325.

Downs, G. M., & Barnard, J. M. (2011). Chemical patent information systems. *Wiley Interdisciplinary Reviews-Computational Molecular Science, 1*(5), 727-741.

Guha, R., & Bender, A. (2011). *Computational Approaches in Cheminformatics and Bioinformatics* (1st ed.). Hoboken: Wiley-Blackwell.

Little, J. L., Cleven, C. D., & Brown, S. D. (2011). Identification of "known unknowns" utilizing accurate mass data and Chemical Abstracts Service databases. *J. Am. Soc. Mass Spectrom., 22*(2), 348-359.

Seiler, K. P., Bodycombe, N. E., Hawkins, T., Shell, R., Lemon, A., de Souza, A., et al. (2011). Master Data Management: Getting your House in Order. *Combinatorial Chemistry & High Throughput Screening, 14*(9), 749-756.

Umashankar, V., & Gurunathan, S. (2011). Chemoinformatics and its applications. *Handb. Syst. Toxicol., 1*, 347-359.

Urbaczek, S., Kolodzik, A., Fischer, J. R., Lippert, T., Heuser, S., Groth, I., et al. (2011). NAOMI: On the Almost Trivial Task of Reading Molecules from Different File formats. *J. Chem. Inf. Model., 51*(12), 3199-3207.

Wagner, A. B. (2011). Searching Coordination and Organometallic Compounds in SciFinder. *Issues in Science and Technology Librarianship*(Fall 2011).

Wagner, A. B. (2011). Searching Inorganic Substances in SciFinder. *Issues in Science and Technology Librarianship*(Winter 2011).

Warr, W. A. (2011). Fragment-based drug discovery: what really works. An interview with Sandy Farmer of Boehringer Ingelheim. *J. Comput.-Aided Mol. Des., 25*(7), 599-605.

Warr, W. A. (2011). Protein Data Bank Japan (PDBj): an interview with Haruki Nakamura of Osaka University. *J. Comput.-Aided Mol. Des., 25*(9), 791-793.

Warr, W. A. (2011). Representation of chemical structures. *Wiley Interdiscip. Rev. Comput. Mol. Sci., 1*(4), 557-579.

Warr, W. A. (2011). Some trends in chem(o)informatics. *Methods Mol. Biol. (N. Y., NY, U. S.), 672*(Chemoinformatics and Computational Chemical Biology), 1-37.

Zhang, Y., & Beckman, R. (2011). E-book Usage among Chemists, Biochemists and Biologists: Findings of a Survey and Interviews. *Issues in Science and Technology Librarianship*(Spring 2011).

Song Yu
Columbia University
New York, NY
sy2133@columbia.edu

Book Reviews

Blandon, Peter, Gorton, John, and Hammond, Robert B. Molecular Modeling: Computational Chemistry Demystified; RSC Publishing: Cambridge, UK. £69.99 (softcover) 310 pp. ISBN: 978-1-84973-352-6.

I have been intrigued by molecular models ever since I first saw them demonstrated in undergraduate chemistry classes. This fascination continued in graduate school, and, although carefully-guarded model sets were available in some research groups, my concentration on heterocyclic and aromatic chemistry made the expensive Dreiding models less attractive to a "starving student." I eventually acquired various modeling sets, through purchase or discards, but I left the lab before I could justify any work with the embryonic modeling programs (although several models still adorn my bookshelves).

When JCIM (and JCICS) still published book and program reviews, most of the reviews of modeling programs and books understandably went to practitioners in the field. Although usually left out of this group of experts, I am a visual species of chemist, and I am still fascinated with molecular modeling and have felt left out for decades. When RSC approached me about reviewing this book for CIB, I jumped at the chance.

This review will be "static," focusing on the content and presentation of the book, since at this time I choose not to load all of the necessary programs onto my aging tower PC for a full scale demo. Therefore, a "dynamic," hands-on work-through will, for now, be left to the reader.

This book provides an excellent, well documented introduction to molecular models and modeling, including a terse and sometimes humorous history. On the assumption that students at a "well endowed university" taking computational chemistry would find all of the necessary resources, the book is aimed at molecular modelers who, for whatever reason, work alone. CDs of Interprobe (INTERCHEM, PRESTO, 3D databases, etc.) open source software are included, and links are provided for additional free resources. Problems and questions appear at the end of three of the chapters.

The introduction outlines chemical structure and structure representation, including SMILES strings. Chapter 2 describes the necessary computer resources to use the programs, plus a synopsis of operating systems and other computational processes. Chapter 3 begins a primer on bonding and other molecular phenomena and on data as a basis for modeling. Chapter 4 enables the user to "get down and dirty" by using INTERCHEM for molecular modeling. Modeling of proteins, nucleic acids, and the solid state, as well as the basics and applications of stereochemistry and conformation analysis, are discussed and demonstrated in Chapters 5-7. 3D structures and modeling in medicinal chemistry and for drug discovery are described in Chapters 8-10.

Appendices cover additional data and program information. The text is well illustrated with structure line drawings, screen shots, and ball and stick computer drawn models. Two errors were noted. In ref. 1 on p. vii, the lead author is P. Gund. Ref. 19, noted in the list on p. 11, is not cited in the text (the publisher has been informed).

The book is recommended for use in collegiate and graduate student molecular modeling courses, as well as by post graduate professionals.

Waller, Francis J. *Writing Chemistry Patents and Intellectual Property: A Practical Guide*; John Wiley & Sons, Hoboken, NJ, $79.95 (hardcover) 256 pp. ISBN: 978-0-470-49740-1.

Based on an ACS Short Course (Practical Approaches to Patents and Other Forms of Intellectual Property), this book is aimed at a fairly wide audience. The author has extensive experience as a scientist and inventor (46 patents) with du Pont and Air Products and Chemicals. The book's audience includes those totally uniformed about intellectual property (IP), new employees in chemical or chem./tech organizations where patents are important (both academic and industrial), the interested general public, and those writing their own patent applications (with the collaboration of a patent attorney, of course). The book could possibly be used as a text for scientists or engineers.

Chapters describe the history of patenting; relevant vocabulary, terminology, and definitions; the differences between trade secrets and patents; provisional and non-provisional US patents; reasons patent applications are rejected; reasons for invalidation of patents; techniques of writing patent applications based on analysis of eight US patents; format of claims; the need for confidentiality agreements; practical information on copyrights and trademarks; global patent filing strategy; information on patenting for academic scientists, including whether to publish or patent; IP resources (primarily books); and future developments. Several chapters pose questions, and responses are provided.

Overall, this is an excellent book, recommended for several audiences including undergraduate and graduate students, as well as "rookie" industrial chemists. However, the importance of prior art searching is only briefly mentioned, and the recommended information resources listed are mostly books. There is scattered and brief mention of Google searching but no mention of Google Patents. For more information on searching the reader is referred to two other books.

Robert E. Buntrock
Buntrock Associates
Orono, ME
buntrock16@myfairpoint.net

ACS Chemical Information Division (CINF)
Fall 2011 ACS National Meeting

Denver, CO (August 28 – September 1)

Technical Program Schedule

Rachelle Bienstock, *Program Chair*

SUNDAY MORNING

Section A
San Diego Convention Center, Room 27A

Drug Polypharmacology Prediction and Design
Cosponsored by LIFE
S. Zhang, Organizer, Presiding
S. Ekins, Presiding

9:00 Introductory Remarks.
9:05-1. Polypharmacology, drug repurposing and collaborative drug discovery: Shining light or flash in the pan? **Christopher A Lipinski**
9:25-2. High accuracy polypharmacology models for large datasets. **S. Joshua Swamidass**
9:45-3. QSARome of GPCRs. **Eugene Muratov**
10:05 Intermission.
10:15-4. Gaussian ensemble screening (GUESS): A new approach to polypharmacology and virtual screening. **Violeta Isabel Perez Nueno**
10:35-5. Physical binding site modeling for quantitative prediction of biological activities. **Rocco Varela**, Ajay N Jain
10:55-6. Virtual screening of multi-target agents from large chemical libraries by machine learning approach. **Yu Zong Chen**
11:15-7. Structure-based identification of a dual FAAH/COXs inhibitor: Tackling inflammation with a single molecule acting synergistically on multiple proteins. **Angelo D Favia**, Andrea Cavalli, Marco De Vivo

Section B
San Diego Convention Center, Room 25C

Instructional Tools for Chemical Information
C. Huber, Organizer, Presiding

8:15 Introductory Remarks.
8:20-8. Embedding chemistry information literacy skills into the curriculum at James Madison University. **Meris A Mandernach**, **Barbara A Reisner**
8:40-9. Resources for introducing crystal structure information into undergraduate teaching. **Gary Battle**
9:00-10. If they build it, will they use it: Using input from students in a chemical literature class in the redisgn of the library's chemistry webpages. **Allan K Hovland**, Rob C Sloan
9:20-11. Explore chemical information teaching resources (XCITR). **Guenter Grethe**, Grace Baysinger, Rene Deplanque, Gregor Fels, Ira Fresen, Andrea Twiss-Brooks, Gregor Zimmermann
9:40-12. Engaging the wired generation. **Jessica A Parr**, Norah Xiao
10:00 Intermission.
10:10-13. Chemical information instruction at ETH Zurich: Review and trends. **Martin P. Braendle**, Engelbert Zass, Lukas Korosec, Peter A Limacher, Hans P. Luethi
10:30-14. One-shot wonder: Integrating chemical information literacy throughout the curriculum. **Linda M Galloway**.
10:50-15. Feedback and training examples from user communities using Elsevier's Reaxys. **Christine Flemming**
11:10-16. Teaching new graduate students: Chemical information as a research tool. **Bonnie L. Fong, Darren B Hansen**
11:30-17. SpringerMaterials: The world's largest resource for chemical and physical properties in materials science. **Mikail Shaikh**
11:50 Concluding Remarks.

Drug Discovery Receptors Not Big-box Stores
Sponsored by COMP,
Cosponsored by BIOL, CINF, and MEDI

Molecular Mechanics Electrostatics and Polarization. The New Black.
Sponsored by COMP,
Cosponsored by BIOL, CINF, MEDI, and PHYS

SUNDAY AFTERNOON

Section A
San Diego Convention Center, Room 27A

Drug Polypharmacology Prediction and Design
Cosponsored by LIFE
S. Zhang, S. Ekins, Organizers, Presiding

1:30 Introductory Remarks.
1:35-18. Develop novel predictive polypharmacology models with high-quality data. **Shuxing Zhang**
1:55-19. Predicting drug polypharmacology using secondary structure element information. **Oliver Koch**
2:15-20. Assessing drug target association using semantic linked data. **David Wild**
2:35 Intermission.
2:45-21. 3D Pharmacophore-based activity profiling for multitarget screening. **Gerhard Wolber**, Fabian Bendix, Goekhan Ibis, Thomas Seidel.
3:05-22. Where have all the good drugs gone? **Gisbert Schneider**
3:25-23. Finding promiscuous old drugs for new uses. **Sean Ekins**, Antony J Williams
3:45 Concluding Remarks.

Section A
San Diego Convention Center, Room 27A

CINFlash
R. Guha, Organizer, Presiding

4:00 Discussion.

Section B
San Diego Convention Center, Room 25C

Instructional Tools for Chemical Information
C. Huber, Organizer, Presiding

1:30-24. Use of course reserves as a gentle introduction to the chemical literature. **Donna T. Wrublewski**
1:50-25. Faculty-librarian collaboration yields innovative chemistry seminar program. **Valerie K. Tucci, Benny Chan**
2:10-26. Blind assessment: The unexpected benefits of peer review in a classroom setting. **Judith N. Currano**.
2:30-27. Faculty-librarian partnership for a student research presentation in a physical chemistry laboratory course. **Donna T Wrublewski**, Mine G Ucak-Astarlioglu
2:50 Intermission
3:00-28. CAS learning solutions: Training at the point of need. **Jayne A. Knoop**
3:20-29. Chemistry Referece Resolver: A tool to simplify reference retrieval. **Oleksandr Zhurakovskyi**
3:40-30. Using LibGuides to enhance large-enrollment chemistry lab courses. **Jeremy R Garritano**
4:00-31. Learning about cheminformatics through an education wiki. **Martin A Walker**, Aileen E Day, Antony J Williams, Lorna M Thomson.
4:20 32. Bringing faculty, students and librarians together: Lessons and opportunities for ACS on campus after two years
S. Sara Rouhi

Chemical Networks in Biology
Sponsored by LIFE, Cosponsored by BIOL, BIOT, CINF, and MEDI

Collaborative Drug Discovery for Neglected Diseases
Sponsored by COMP, Cosponsored by BIOL, BIOT, CINF, MEDI, TOXI, and YCC

Drug Discovery Target-based is Sooooo Cool
Sponsored by COMP, Cosponsored by BIOL, CINF, and MEDI

SUNDAY EVENING

Section A
San Diego Convention Center, Hall D

CINF Scholarship for Scientific Excellence
G. Grethe, Organizer, Presiding

6:30 - 8:30
33. Use of screening results to validate a diversity subset of an HTS library. **Rohan S Patil**, Maureen Beresini, Nicholas Skelton
34. Development of a screening informatics system at the UNM Center for Molecular Discovery, an NIH MLP specialty center. **Jeremy J Yang**, Oleg Ursu, Stephen L Mathias, Cristian G Bologa, Anna Waller, Annette M Evangelisti, Gergely Zahoransky-Kohalmi, Tudor I Oprea.
35. Thermodynamical properties of small Pd clusters on the stoichiometric and defective TiO2 (110) surfaces studied with first-principle methods. **Jin Zhang**, Anastassia Alexandrova
36. Development of a hybrid method combining quantum mechanical calculations and discrete molecular dynamics for metallo-protein modeling. **Manuel Sparta**, Anastassia N Alexandrova
37. Learning to predict more chemical reactions: Model extensions and an expanded training set. **Matthew A Kayala**, Pierre Baldi
38. COBRA: Computational brewing approach to predicting the molecular composition of organic aerosols. **David R Fooshee**, Tran B Nguyen, Sergey A Nizkorodov, Julia Laskin, Alex Laskin, Pierre Baldi

39. High-throughput 3D structure prediction of small molecules. **Peter Sadowski**, Arlo Randall, Pierre Baldi
40. Predicting inactive and active conformations of the dopamine D2 receptor. **Fan Liu**, Ravinder Abrol, Dennis A Dougherty, William A Goddard III

41. Impact of retractions on the chemical literature. **Elsa Alvaro**
42. Cheminfomatic modeling of human CC chemokine receptorome. **Terry-Elinor Reid**, Huzefa Rangwala, Samantha McCullough, Muhammad Habib, Simon Wang
43. Evolutionary computational modeling of β-diketo acids for virtual screening of HIV-1 integrase inhibitors. **Gene M Ko**, A. Srinivas Reddy, Rajni Garg, Sunil Kumar, Ahmad R Hadaegh
44. Molecular dynamics of the Hsp70 chaperone in response to nucleotide and substrate: A coarse-grained perspective. **Ewa I. Golas**, Gia G. Maisuradze, Patrick Senet, Stanislaw Oldziej, Cezary Czaplewski, Harold A. Scheraga, Adam Liwo
45. Searching putative targets in silico for anti–prion compounds. **Jorge Valencia**, Beining Chen, Val Gillet
46. On the accuracy of chemical structures found on the internet. **Andrew D. Fant**, Eugene Muratov, Denis Fourches, Antony J. Williams, Alexander Tropsha
47. Spectral clustering of chemical data: A Lanczos-based approach. **Sonny Gan**, Valerie J Gillet, Eleanor J Gardiner, David A Cosgrove
48. Structure based pharmacophore screening for new P-gp inhibitors. **Freya Klepsch**, Katharina Prokes, Zahida Parveen, Peter Chiba, Gerhard F Ecker

MONDAY MORNING

Section A
San Diego Convention Center, Room 27A

Computer-Aided Drug Design: Hopes, Reality and Prospects How Has Computational Chemistry Transformed Drug Discovery, and What Can Increase its Impact
Cosponsored by COMP
C. Corbeil, J. Cross, Organizers
O. Ravitz, Organizer, Presiding

8:15 Introductory Remarks
8:20 49. Perspective in computational approaches applied to drug discovery problems. **Christine Humblet**
9:05-50. Rational, data-driven approach to lead optimization. **Dan J Warner**
9:50 Intermission
10:05-51. WOMBAT and WOMBAT-PK: Ten years. **Tudor I Oprea**
10:50-52. Developing "Best Practices" in predictive cheminformatics for drug-discovery applications. **Curt M Breneman,** Michael Krein, Margaret McLellan, Tao-wei Huang, Lisa Morkowchuk, Dimitris K. Agrafiotis

Section B
San Diego Convention Center, Room 25C

Joint CINF-CSA Trust Symposium Beyond Small Molecules: Pushing the Envelope for Chemical Structure Representation
Financially supported by Chemical Structure Association Trust
K. Taylor, Organizer, Presiding

8:00 Introductory Remarks
8:05-53. Cheminformatics for material discovery: Representation, searching and screening of porous materials. **Richard L Martin**, Maciej Haranczyk
8:30-54. New strategies to normalize chemical structure representations and weed-out impractical small molecules. **Evan Bolton**
8:55-55. Efficient perception of proteins and nucleic acids from atomic connectivity. **Roger A Sayle**
9:20-56. Organization and analysis of information for biotherapeutics research. **Hugo O Villar**, Mark R. Hansen, Eric Feyfant
9:45 Intermission
9:55-57. Markush structure usability in patent and combinatorial chemistry: New approaches and software tools. **Wei Deng**
10:20-58. Rendering the stages of structure elucidation: ACD/Labs Markush representation. **Andrey Yerin**, Ian Peirson
10:45-59. New developments in Markush structure searching. **Donald Walter**
11:10-60. Representing and retrieving non specific structures. **Keith T Taylor**
11:35 Concluding Remarks

Drug Discovery Data Alchemy
Sponsored by COMP, Cosponsored by BIOL, CINF, and MEDI

Drug Discovery Structural Bioinformatics: Modeling Protein-Protein Interactions and Novel Drug Targets
Sponsored by COMP, Cosponsored by BIOL, CINF, and MEDI

Perspectives in Applied Computational Methods
Sponsored by COMP, Cosponsored by CINF and MEDI

MONDAY AFTERNOON

Section A
San Diego Convention Center, Room 27A

Computer-Aided Drug Design: Hopes, Reality and Prospects
Cosponsored by COMP
C. Corbeil, J. Cross, Organizers
O. Ravitz, Organizer, Presiding

1:00 Introductory Remarks
1:05-61. Toward a computational pipeline from antibody homology modeling to docking to design. **Jeffrey J Gray**
1:50-62. MD simulations in pharmaceutical research - examples and lessons learnt. Hannes G Wallnoefer, Klaus R Liedl, Clara Christ, Daniel Seeliger, **Thomas Fox**,
2:35 Intermission
2:50-63. Fragment-to-lead using fragment molecular orbital QM calculations. **Richard J Law**, Osamu Ichihara, Michael P Mazanetz, Michelle Southey, Mark Whittaker, David Hallett
3:35-64. Docking: This might be heaven or this might be…. **Martha S Head**
4:20 Concluding Remarks
4:25 Intermission
4:30 CINF Open Meeting

Section B
San Diego Convention Center, Room 25C

Mobile Space and E-Books
R. Apodaca, Organizer, Presiding

12:45 Introductory Remarks
12:50-65. Having a mobile app presence - necessary or nice to have? **Steven M Muskal**
1:25-66. Molecular visualization apps in education and research. Jason Vertrees, Blaine Bell, Woody Sherman
2:00-67. Building a mobile app ecosystem for chemistry collaboration. Alex M. Clark
2:35 Intermission.
2:45-68. Chemistry made mobile – the expanding world of chemistry in the hand. **Antony Williams**
3:20-69. ChemDoodle Mobile: Leveraging mobile apps in chemistry. **Kevin J Theisen**
3:55-70. Mobile apps for drug discovery. Antony J Williams, **Sean Ekins**, Alex Clark

Drug Discovery Structural Bioinformatics: Exploring Structure-Function Relationships
Sponsored by COMP, Cosponsored by BIOL, CINF, and MEDI

Perspectives in Applied Computational Methods
Sponsored by COMP, Cosponsored by CINF and MEDI

MONDAY EVENING

Section A
San Diego Convention Center, Hall D

Sci-Mix
R. Bienstock, Organizer
8:00 - 10:00

1, 4, 5, 7, 12, 14, 16, 17, 18, 19, 24, 27, 29, 33, 34, 37, 38, 40, 42, 43, 44, 45, 47, 48, 53, 66. *See previous listings.*
84, 88, 91, 122, 132, 133. *See later listings.*

71. Efficient one-pot preparations of PI3Kd inhibitors using algorithmic network detection. **Chris M. Gothard**, Nosheen A. Gothard, Siowling Soh, Bartosz A. Grzybowski
72. Statistical analysis of microarray gene expression data from a mouse model of Toxoplasmosis. **Shrikant d Pawar**, Claire Rinehart, Cheryl Davis

TUESDAY MORNING

Section A
San Diego Convention Center, Room 27A

Recent Advances in Reaction Searching
R. Schenck, D. Evans, Organizers, Presiding

8:30 Introductory Remarks.
8:35-73. Synthetic information challenges for the medicinal chemist. **Haiying He**
9:00-74. SOS 4.0: Advances in text, structure, and reaction searching. **M. Fiona Shortt de Hernandez**, Rolf Hoppe, Guido F. Herrmann, Peter Loew
9:25-75. Automated extraction of reactions from the patent literature. **Daniel M Lowe**, Peter Murray-Rust, Robert C Glen
9:50 Intermission
10:00-76. Efficient searching and similarity of unmapped reactions: Applications to pharmaceutical ELN analysis. **Roger A Sayle**, Thierry Kogej, David Drake

10:25-77. Novel tools and techniques in reaction searching: "Name Reaction" and "All-In-One" reaction searches. **Valentina Eigner-Pitto**, Hans Kraut, Heinz Saller, Heinz Matuszczyk, Peter Loew
10:50-78. Catalyzing information retrieval for organometallic and metal-mediated reactions. **Judith N. Currano**
11:15-9. Finding synthetic chemistry in global literature and patents. **Kurt Zielenbach**, Jeffrey M Wilson, Jeffrey D Schloss, Bryan J Harkleroad

Section B
San Diego Convention Center, Room 25C

Systems Chemical Biology and Other "Systems" Approaches in Chemistry and Biology
T. Oprea, J. Kuras, Organizers, Presiding

8:00 Introductory Remarks
8:05-80. WITHDRAWN
8:35-81. Development of a human diet interactome map. **Irene Kouskoumvekaki**
9:05-82. Studying the chemical interactome space between the human host and the genetically defined metabotypes of our gut. **Gianni Panagiotou**
9:35-83. Comparative study of small molecule inhibition of Mycobacterium tuberculosis and Francisella tularensis. Sandra V Bennun, **Elebeoba E May**
10:05 Discussion
10:20 Intermission
10:30-84. Identifying druggable targets by mining open chemical biology data. **Yanli Wang**
11:00-85. Exploiting semantic networks of public data for systems chemical biology. **David J Wild**
11:30-86. Enhancing chemoinformatics with pathway analysis tools: An integrated approach to drug discovery. **Tatiana Khasanova**, Eugene Myshkin, Sirimon O'Charoen, Yuri Nikolsky, **Svetlana Bureeva**

Drug Discovery Looking for a Few Good Methods? We Got 'Em Here
Sponsored by COMP, Cosponsored by BIOL, CINF, and MEDI

Perspectives in Applied Computational Methods
Sponsored by COMP, Cosponsored by CINF and MEDI

TUESDAY AFTERNOON

Section A
San Diego Convention Center, Room 27A

Recent Advances in Reaction Searching
R. Schenck, D. Evans, Organizers, Presiding

1:30 Introductory Remarks
1:35-87. Updated tools, techniques and data sources for effective reaction retrieval in support of synthetic methodology and drug discovery. **Matthew A Kellett**
1:55-88. Extremely rapid searching of in-house reaction databases: Turning ELN data into a searchable library. **Philip J Skinner**, Scott Flicker, Joshua Wakefield, Sean Greenhow, Megean Schoenberg, Kate Blanchard, Phil McHale, Sandra Sessoms, Robin Smith
2:15-89. Understanding search results: From a single reaction to scope and limitations of the reaction route in the ChemInform Reaction Library (CIRX). **Yana Steudel**, Ulrike Schramke
2:35-90. Dealing with chemical reality: Handling reactions plus associated data and branching reaction schemes. Jonathan S Brecher, Harold Helson, **Phil J McHale**
2:55 Intermission
3:05-91. Reaction searching for compounds which do not even exist yet. **Carsten Detering**, Christian Lemmen, Marcus Gastreich
3:25-92. Helping you make the right choices for your next synthetic route! **Juergen Swienty-Busch**, David A. Evans
3:45-93. Advanced reaction searching: A comprehensive treatment of stereoselectivity in reactions. **Peter Johnson**, Anthony P Cook, James Law, Aniko Simon, Orr Ravitz
4:05-94. Algorithmic network detection of reaction sequences: From novel "one-pot" reactions to unanticipated synthetic routes to chemical weapons. **Chris M. Gothard**, Nosheen A. Gothard, Siowling Soh, Bartosz A. Grzybowski
4:25 Concluding Remarks

Section B
San Diego Convention Center, Room 25C

Systems Chemical Biology and Other "Systems" Approaches in Chemistry and Biology
T. Oprea, J. Kuras, Organizers, Presiding

1:30-95. Framework for systematic prediction of pharmacologically relevant targets of small molecules. **Emmanuel R Yera**, Ann E Cleves, Ajay N Jain
2:00-96. Designing ligands against multi-target profiles. **Andrew L Hopkins**

2:30-97. CARLSBAD (Confederated Annotated Research Libraries for Small molecule BioActivity Data): A database and its platform

Gergely Zahoranszky-Kohalmi, Jeremy J Yang, Cristian G Bologa, Stephen L Mathias, Oleg Ursu, Jarrett Hines-Kay, Tudor I Oprea
3:00 Intermission
3:10-98. Drug combinations to reduce adverse drug reactions and improve intrapatient differences in response. **John P Overington**
3:40-99. Integrating targets, drugs and clinical outcomes into systems medicine. **Tudor I Oprea**
4:10 Panel Discussion
4:55 Concluding Remarks

Drug Discovery Methods Make Us Smile
Sponsored by COMP, Cosponsored by BIOL, CINF, and MEDI

Molecular Mechanics Methodologies: That is All
Sponsored by COMP, Cosponsored by BIOL, CINF, MEDI, and PHYS

Perspectives in Applied Computational Methods
Sponsored by COMP, Cosponsored by CINF and MEDI

WEDNESDAY MORNING

Section A
San Diego Convention Center, Room 27A

InChI Symposium
A. Tropsha, Organizer
A. J. Williams, Organizer, Presiding

8:30 Introductory Remarks
8:35-100. IUPAC InChI project: A status report. **Stephen Heller**
9:05-101. Great promise of navigating the internet using InChIs. **Antony J Williams**
9:35-102. InChI names and keys: Do they add value to commercial software and databases. **Keith T Taylor**, Carmen Nitsche
10:05 Intermission
10:25-103. Use of InChI in wikis. **Martin A Walke**, Aileen Day
10:55-104. InChIKey collision safety: Experimental estimation for algorithmically generated structure libraries. Andrey Yerin, Kirill Blinov
11:25-105. InChI here, InChI there, InChIs everywhere. **Juergen Swienty-Busch**, David A. Evans

Section B
San Diego Convention Center, Room 25C

Beyond the Database: New Models of Scholarship in an eScience World
P. Bourne, Organizer, Presiding

9:00 Introductory Remarks.
9:05-106. New searching paradigms in drug discovery enabled by semantic integration of public data. **David J Wild**, Erik A Stolterman, Michael S Lajiness
9:25-107. Collaborative computational technologies for biomedical research: An enabler of more open drug discovery. **Sean Ekins**, Antony J Williams
9:45-108. Enabling biomolecular simulation data sharing across institutions using a Grid architecture. **Julien C Thibault**, Thomas E Cheatham, Julio C Facelli
10:05 Intermission
10:15-109. Representing chemical information by URLs: The chemical identifier resolver as a general chemoinformatics tool. **Marc C Nicklaus**, Markus Sitzmann
10:35-110. Publication@Source: The Lab as a database. **Jeremy G Frey**, Mark I Borkum, Simon J Coles, Tim Parkinson
10:55 Concluding Remarks

Computational Approaches to Spectroscopy Analysis Spectroscopy of Small Things
Sponsored by COMP, Cosponsored by ANYL, CINF, and PHYS

Drug Discovery No Madness, Just Methods
Sponsored by COMP, Cosponsored by BIOL, CINF, and MEDI

Molecular Mechanics Methodologies are Still Cool
Sponsored by COMP, Cosponsored by BIOL, CINF, MEDI, and PHYS

WEDNESDAY AFTERNOON

Section A
San Diego Convention Center, Room 27A

InChI Symposium
A. J. Williams, Organizer
A. Tropsha, Organizer, Presiding

1:00-111. InChIs as building blocks for complex substance identifiers. **Yulia Borodina**, Lawrence Callahan, Frank Switzer
1:30-112. Accessing NCI/CADD web resources by InChI. **Markus Sitzmann**, Marc C. Nicklaus

2:00-113. InChI vs IUPAC nomenclature: Aspects to be aware of when using Standard InChI. **Daniel M Lowe**
2:10-114. InChI adoption at the Royal Society of Chemistry. **Richard Kidd**
2:20-115. Registration system of mcule: InChI is the key. **Ferenc Szalai**, Robert Kiss, Mark Sandor
2:30 Intermission
2:45-116. "UniChem": A prototype unified chemical structure cross-referencing and identifier tracking system. **Jon Chambers**, Anna Gaulton, Anne Hersey, Mark Davies, John P Overington
3:15-117. Update on project to introduce InChI to researchers in the Department of Chemistry at Louisiana State University. **William W Armstrong**, Karen L Salazar
3:45-118. Past, present and future of the InChI Trust. **Jason N Wilde**
4:15-119. InChiKey insertion technique for compound-specific and any-compound proximity search. **Stephen K. Boyer**, Thomas Griffin, Alfredo Alba, Su Yan, Ying Chen, Scott Spangler, Eric Louie, Jeff Kreulen
4:35-120. Exploring almost every InChI of nature.com. **Laura J Croft**
4:55 Concluding Remarks

Section B
San Diego Convention Center, Room 25C

Libraries and Institutional Research Evaluation
L. Solla, Organizer
A. Twiss-Brooks, Organizer, Presiding

1:30 Introductory Remarks
1:35-121. Finding the future: Using research analytical tools with journal article databases and social media data to identify high-impact research leaders and programs. **Elizabeth A. Brown**
2:00-122. Providing comparative data on published research impact (internally and externally). **Donna T. Wrublewski**, Denise B. Bennett, Valrie I. Davis, Michelle Leonard
2:25-123. Social networking tools as public representations of a scientist. **Antony J Williams**
2:50 Intermission
3:00-124. Next era of research productivity evaluation: A multidimensional research assessment framework. **Daniel Calto**
3:25-125. Measuring research: Beyond H. **Daniel Hook**
3:50-126. Methods and solutions for measuring and benchmarking the impact of research. **Daphne Grecchi**

Computational Approaches to Spectroscopy Analysis Spectroscopy of Slightly Bigger Things
Sponsored by COMP, Cosponsored by ANYL, CINF, and PHYS

Drug Discovery Stomping Bugs, Drug Style (Anti-infectives) *Sponsored by COMP, Cosponsored by BIOL, CINF, and MEDI*

Material Science
Sponsored by COMP, Cosponsored by CINF, PHYS, PMSE, and POLY

Molecular Mechanics Applications: No Abbreviations
Sponsored by COMP, Cosponsored by BIOL, CINF, MEDI, and PHYS

THURSDAY MORNING

Section A
San Diego Convention Center, Room 27A

General Papers Chemical Databases, Drug Discovery, and Chemical Structure Representation
R. Bienstock, Organizer, Presiding

8:30-127. Can we really do computer-aided drug design? **Matthew D Segall**
8:50-128. Where screening starts: Effective preprocessing of chemical libraries. Matthias Hilbig, Adrian Kolodzik, Sascha Urbaczek, **Matthias Rarey**
9:10-129. Toward a gold standard: Improving the quality of public domain chemistry databases. Antony J Williams, **Sean Ekins**
9:30 Intermission
9:40-130. ChemSpider as a knowledge base. **Valery Tkachenko**, Antony Williams, Aileen Day, Jon Steel
10:00-131. ChemSpider as a chemical term resolver. Valery Tkachenko, **Antony Williams**
10:20-132. How to design chemical patterns easily with an interactive editor. **Karen T. Schomburg**, Lars Wetzer, Matthias Rarey
10:40-133. Lexichem TK 2.1.0. **Edward O Cannon**
Drug Discovery Inside of a Ligand
Sponsored by COMP, Cosponsored by BIOL, CINF, and MEDI

Material Science
Sponsored by COMP, Cosponsored by CINF, PHYS, PMSE, and POLY

Molecular Mechanics Proteins are Just Plain Interesting

Sponsored by COMP, Cosponsored by BIOL, CINF, MEDI, and PHYS

THURSDAY AFTERNOON

Drug Discovery Talking About Ligands
Sponsored by COMP, Cosponsored by BIOL, CINF, and MEDI

Material Science
Sponsored by COMP, Cosponsored by CINF, PHYS, PMSE, and POLY

Molecular Mechanics Application of Our Hip Methodologies
Sponsored by COMP, Cosponsored by BIOL, CINF, MEDI, and PHYS

Molecular Mechanics Proteins: There is Nothing Plain and Simple About 'Em.
Sponsored by COMP, Cosponsored by BIOL, CINF, MEDI, and PHYS

Social Networking Events at the Spring ACS National Meeting

The ACS Division of Chemical Information is pleased to host the following social networking events at the Spring 2012 ACS National Meeting in San Diego, CA. Please join us

CINF Sunday Welcoming Reception & CINF Scholarships for Scientific Excellence Posters
6:30-8:30pm, Sunday, March 25th – Grand Ballroom Salon C, Omni San Diego Hotel

Reception co-sponsored by **ACS Symposium Series**, **InfoChem**, **Thieme Publishers**, and ***Journal of Chemical Information and Modeling***.

Scholarships for Scientific Excellence sponsored exclusively by **Accelrys.**

Harry's Party
5:30-8:00pm Monday, March 26th– Suite 577, Seaport Tower, Manchester Grand Hyatt

Sponsored exclusively by FIZ CHEMIE Berlin**. * 6 minutes north of the Convention Center**

CINF Luncheon *(Ticketed Event – Purchase tickets at registration or see Rajarshi Guha on site.)*
12:00-1:30pm Tuesday March 27th – Sails Pavilion, San Diego Convention Center

Luncheon sponsored exclusively by: RSC Publishing.

Speaker: Dr. Kirsten Skorgerson (Monsanto). "The Chemistry of Wine"

CINF Tuesday Reception in support of the CINF InChI Symposium
Reception: 6:30-8:30pm Tuesday, March 27th – Ballroom D. Mancehster Grand Hyatt
Symposium: 8:30am-5:00pm Wednesday March 28th – Room 27A, San Diego Convention Center

Reception co-sponsored by **ACS Division of Chemical Information**, **Accelrys**, **ChemAxon** and **PerkinElmer**.

InChI Symposium Speaker Support sponsored by the **InChI Trust**.

Committee Information

The following CINF committee and governance meetings will occur on Saturday, March 24, 2012, at the San Diego Convention Center. The CINF Executive Committee Meeting is a closed meeting; if you wish to attend it, contact the division chair.

11:30 AM - 1:00 PM	Long Range Planning Luncheon	Room 25C
1:00 PM - 2:00 PM	Awards Committee Meeting	Room 26B
2:00 PM - 3:00 PM	Finance & Fundraising Committee Meetings	Room 25A
1:00 PM - 3:00 PM	Education Committee Meeting	Room 27A
1:00 PM - 3:00 PM	Communications & Publications Committee Meeting	Room 26A
1:00 PM - 4:00 PM	Program Committee Meeting	Room 25B
4:00 PM - 6:00 PM	Executive Committee Meeting	Room 25C

The chairs of several CINF committees that are meeting in San Diego have included agendas and descriptions of their work. Many are actively recruiting new members; if you are interested in joining a committee, please contact its chair. A list of committees and their chairs appears on page 73 of this publication.

Awards Committee

The Awards Committee will meet in closed session at the San Diego ACS National Meeting.

Due to the sensitive nature of awards selection, the CINF Awards Committee meeting is a closed session. However, members of the Division who are interested in serving on the Awards Committee should contact the chair of the committee, Andrea Twiss-Brooks (atbrooks@uchicago.edu).

During the San Diego meeting, the Awards Committee will be reviewing applications to select the winner of the Lucille Wert Scholarship, discussing the future of the CINF Best Presentation Award, and planning awards activities for the fall meeting in Philadelphia, including the presentation of the Herman Skolnik Award. On Sunday evening, they will select and announce the winners CINF Scholarships for Scientific Excellence (sponsored in San Diego by Accelrys) during the Welcome Reception and poster session.

Communications & Publications Committee

The following is a preliminary meeting agenda for the Communications & Publications Committee.

1. CINF website and the new Drupal implementation (Danielle Dennie)
2. New e-CIB content: an update (Judith Currano)
3. Editors for the four e-CIB issues in 2013 (Judith Currano/Svetlana Korolev)
4. Committee membership: search for new members
5. Search for new committee chair (2013-2015)
6. ACS Network and its relationship to the CINF website: an update
7. Procedures manual

If you have questions or comments, please contact the Chair of the Communications & Publications Committee, Bill Town, at bill_town@mac.com.

Education Committee

The following is the meeting agenda for the Education Committee.

1. Review of San Diego meeting
    a. Program – Symposium: Instructional Tools for Chemical Information, Sunday, March 25
    b. Other
2. Philadelphia meeting – August 19 – 23, 2012
    a. Program
    b. Other
3. BCCE 2012 – Penn State (University Park, PA) July 29 – August 2, 2012
4. Looking ahead
    a. New Orleans – April 7-11, 2013
    b. Indianapolis – September 8-12, 2013
5. Information Competencies for Chemistry Undergraduates – what next?
6. Information Competencies for Chemistry Graduates – is there still any interest?
7. New business
    a. Looking ahead – new chair for 2013-15
    b. Other?

For more information, please contact the chair, Chuck Huber, at huber@library.ucsb.edu.

Program Committee

The Program Committee will hold an open meeting on Saturday, March 24, 2012, in Room 25B of the San Diego Convention Center, from 1:00 - 4:00 PM. Anyone can attend; the more the merrier!

NOTE: People who will not be attending the Program Planning meeting in San Diego are free to email programming suggestions for future meetings,(if they are interested in chairing a symposium on a specific topic) to the program chair.

The agenda for the Program Committee is:

1. Review the current CINF San Diego Spring 2012 program.
2. Review the Philadelphia Fall 2012 program and deadline dates.
3. Introduce Jeremy Garritano, taking over as Program Chair after the Philadelphia Meeting.
4. Discussion and review (along with with Jeremy) of proposed programs for New Orleans and Indianapolis, Spring and Fall 2013 meetings.
    - New Orleans Theme: Chemistry of Energy and Food (Confirmed)
    - Indianapolis Theme: Chemistry in Motion (Confirmed)
5. Begin outlining programs, obtaining suggestions and proposing symposia and chairs to contact for Spring and Fall 2014 meetings.
    - Spring 2014, Dallas: Chemistry of Energy/Advanced Materials for New Opportunities (Approved by MPPG; to be confirmed by ACS Divisions/Committees)
    - Fall 2014, San Francisco: Chemistry and Stewardship of the World (Approved by MPPG; to be confirmed by ACS Divisions/Committees)
6. Conclude and wrap up.

Rachelle Bienstock, CINF Program Committee Chair
rachelleb1@gmail.com

Report on the Council Agenda for March 28, 2012

The Council of the American Chemical Society will meet in San Diego, CA, on Wednesday, March 28, 2012, from 8:00am until approximately 12:00pm in the Marriott Hall of the San Diego Marriott Marquis & Marina Hotel All ACS members are welcome to attend, although only Councilors are permitted to vote. A continental breakfast is usually available at 7:00am for all attendees.

There are a number of items for Council Action, but the majority is routine and not major. The action items are summarized below.

Nominations and Elections

President-Elect

The Committee on Nominations & Elections (N&E) has identified four nominees for the office of 2013 ACS President-Elect. They are as follows: Thomas J. Barton, Luis A. Echegoyen, William A. Lester, Jr., and Barry M. Trost. The four nominees will answer questions at the Town Hall meeting that will be held on Sunday, March 25th, at 4:45pm in the San Diego Ballroom Salon A at the Marriott Marquis & Marina Hotel, and each will give a three minute presentation at

the Council meeting. Council will select the final two candidates whose names will appear on the fall ballot.

District I and District V Directors

The Committee on Nominations and Elections has announced the list of nominees to represent District I and District V on the Board of Directors for the term 2013-2015. Nominees for District I include Barbara J. Garrison, Thomas R. Gilbert, Neil D. Jespersen, and Julianne M. D. Smist. Nominees for District V are John E. Adams, Frank D. Blum, Peter K. Dorhout, and Frankie K. Wood-Black. Ballots have been mailed to Councilors in the two districts and the results will be announced in San Diego. The final ballots with two candidates for each district will be mailed on or before October 10 to all members in District I and District V for election of a Director from each District.

Director-at-Large

The Committee on Nominations and Elections also announced the selection of the following candidates for Director-at-Large for a 2013-2015 term: Carol A. Duane, Valerie J. Kuck, Bonnie Lawlor, and Ingrid Montes. The election of two Directors-at-Large will be conducted in the fall. Ballots will be mailed to the Council on or before October 10, 2012.

Town Hall Meeting

A Town Hall meeting organized by the Committee on Nominations and Elections is scheduled for Sunday, March 25, 2012 in the Marriott Marquis & Marina Hotel, Marriott Hall, from 4:45pm - 5:45pm. It will highlight a Q&A session with the candidates for President-Elect. All ACS members are encouraged to attend. It is a great way to gather first-hand information and decide for whom you might want to vote in the fall election

ACS Dues for 2013

Council will vote on the recommendation from the Committee on Budget and Finance with regard to the 2013 membership dues (an increase of $3.00 - from $148 to $151). The increases to ACS dues are based upon an escalator defined in the ACS Bylaws (Bylaw XII, Section 3,a). The dues are calculated by multiplying the base (current) rate "by a factor which is the ratio of the revised Consumer Price Index for Urban Wage Earners and Clerical Workers (Service Category) for the second year previous to the dues year to the value of the index for the third year previous to the dues year, as published by the United States Department of Labor, with the fractional dollar amounts rounded to the nearest whole dollar".

Base rate 2012: $148.00

Change in the Consumer Price Index, Urban Wage Earners, Services Category:

December 2011 CPI-W Services:	262.954
December 2010 CPI-W Services:	257.382
Change in CPI-W Index:	2.16%

2013 Dues, fully escalated: $148.00 x 1.0216 = $151.20
2013 Dues, Rounded: $151.00

Change in Committees

The Committee on Committees (ConC) will put forth a recommendation for the continuation of the committees that have been reviewed as required by the Committee Sunset Laws. The committees were not named in the Agenda Book.

Approval of Chemical Professional's Code of Conduct

Council will vote to accept the newest edition of the Chemical Professional's Code of Conduct (last revision was in 2007). The changes are minor. A statement that says "comply with safety policies and procedures" has been added to the section on the responsibilities of chemical professionals to their employers. In the section on responsibilities of chemical professionals to their employees, the wording now reads "… respect the professionalism of their subordinates and have concern for their well-being…" Other minor changes address the environment and respect for colleagues despite their level of education and whether they come from government, academia or industry or from other scientific and engineering disciplines.

Petition on Market Data Collection

On occasion, an ACS President will request that a special discussion item be placed on the Council agenda. In Denver, the topic "How can ACS best cultivate a culture of safety in U.S. universities and colleges?" will be discussed in response to devastating incidents in academic laboratories and observations by many that graduates do not have strong safety skills.

In June 2011, the Safety Culture Task Force (SCTF) held a retreat to discuss and begin efforts to prepare publications that might assist academia in strengthening its safety cultures. Specifically, SCTF seeks to identify the best elements and best practices of a good safety culture, identify academic institutions that are currently using these best practices, identify specific recommendations that could be used by universities and colleges to strengthen their safety culture, and identify tools and resources that could be beneficial in these efforts.

Elements that SCTF believe to be critical to strengthening safety cultures are: leadership; teaching basic laboratory and chemical safety; safety ethic/attitude/safety awareness; learning lessons from laboratory incidents; collaborative interactions; promoting and communicating safety; and encouraging institutional support of safety by budgeting for safety programs and supplies.

SCTF seeks: 1) input, suggestions, and recommendations for strengthening the safety culture in academia; 2) "bright spots" – those places within academia that provide exemplary examples of one or more of the above mentioned safety culture elements; and 3) good resources and instructional materials that can be used to build strong safety cultures. Send your comments and suggestions to Marta Gmurczyk, ACS Staff Liaison to the Committee on Chemical Safety at M_Gmurczyk@acs.org

Bylaw Changes for Council Vote

Petition on Market Data Collection
Council is being asked to vote on a petition related to market tests for data collection. This petition seeks to add a provision to allow for recommended controlled market testing to collect data before the Society changes benefits, dues or membership categories. The Committee on Budget and Finance cannot assess with reasonable accuracy the range of potential costs that would result from the implementation of this petition as they have no idea how many market tests may ultimately be conducted. The petitioners have thus added additional wording that will impose limits on the size and length of tests. The wording is as follows:

"The Committee on Membership Affairs may conduct membership recruitment and retention market testing of SOCIETY memberships, services, and benefits, including special dues categories, to provide data as the basis for recommendations to the Council. The number of individuals included in any test related to dues discount shall not exceed a number equal to 10% of the total membership. Each test shall be limited to three years' duration unless granted prior approval by Council. All test results will be reported to Council at least once a year."

Petition to Amend Recorded Vote Request Procedure
Council will vote on a change in wording to BYLAW III, Section 4, regarding recorded votes taken at Council meetings. The wording is as follows:

"Any member of Council may call for a recorded vote on the current action before the Council, other than an election, at any time before voting using a method from which it can be determined how each Councilor voted has commenced."

This wording has been introduced as Council now uses "clickers" for voting. The computer that totals the votes can list votes by clicker number from which it can be determined how each Councilor voted.

The petition has no impact on the finances of the Society.

Bylaw Changes for Consideration Only

Petition on Candidate Comment in C&EN
The wording to BYLAW V, Section13 is recommended to have the following added:

"A candidate for election to the Board of Directors may not publish an ACS Comment in C&EN six monts before the ballots are mailed for the corresponding election."

This is to avoid an unfair campaign advantage to incumbents who may choose to make official statements on their task force or committee activities prior to elections.

Petition on International Chemical Sciences Chapters Funds
The proposed revised wording of BYLAW IX, Section 4, is as follows:

"An International Chemical Sciences Chapter shall receive no allotment of funds from the SOCIETY and shall not be entitled to elected representatives on the Council. However, the Board of Directors may allocate funds to a Chapter for a specific SOCIETY activity in which participation of the Chapter is deemed necessary for carrying out that SOCIETY activity effectively."

The financial implications of this petition are still being assessed.

Respectfully submitted March 2, 2012

CINF Councilors

Bonnie Lawlor
National Federation of Advanced Information Services (NFAIS)
blawlor@nfais.org

Andrea Twiss-Brooks
University of Chicago
atbrooks@uchicago.edu

News from Our Sponsors

The American Chemical Society Division of Chemical Information (CINF) is very fortunate to receive generous financial support from our sponsors to maintain the high quality of the Division's programming and to promote communication between members at social functions at the ACS Spring 2012 National Meeting in San Diego, CA, and to support other divisional activities during the year, including scholarships to graduate students in Chemical Information.

The Division gratefully acknowledges contributions from the following sponsors:

Gold	FIZ CHEMIE Berlin
Silver	Accelrys RSC Publishing
Bronze	ACS Symposium Series ChemAxon Chemical Computing Group InChI Trust InfoChem *Journal of Chemical Information and Modeling* PerkinElmer Thieme Publishers
Contributor	OpenEye SimBioSys

Opportunities are available to sponsor Division of Chemical Information events, speakers, and material. Our sponsors are acknowledged on the http://www.acscinf.org Web site, in the *Chemical Information Bulletin*, on printed meeting materials, on banners at the symposia, and at any events for which we use your contribution.

Please feel free to contact me if you would like more information about supporting CINF.

Graham Douglas
Chair, Fundraising Committee
Email: Fundraising@acscinf.org
Tel: 510-407-0769

The ACS CINF Division is a non-profit tax-exempt organization with taxpayer ID no. 52-6054220.

News from ChemAxon

TIBCO Spotfire and ChemAxon to Deliver Integrated Solution for Chemical Science Community

(http://www.chemaxon.com/news/tibco-spotfire-and-chemaxon-to-deliver-integrated-solution-for-chemical-science-community/): Bench discovery scientists can now use their existing Spotfire licenses with Instant JChem (IJC). IJC (Instant Jchem) is a desktop device for managing chemical searching, enumerating, property prediction, and storing related project data, and one of its unique features is that it is suitable for working with large data sets. Spotfire is fairly well used in the industry, and we have had major success with IJC as companies migrate from legacy applications like ISISbase. The fact that IJC comes with its own database and is quite 'cost effective' also makes this interesting for smaller organizations and individuals, as IJC as they can get started quickly.

ChemAxon Adds Document Viewer for PDF Files to chemicalize.org

(http://www.chemaxon.com/news/introducing-the-document-viewer-on-chemicalize-org/): chemicalize.org is a free and very useful website we created to improve chemistry on the web. It lets you paste a URL and see a "chemicalized" version of it (text names for chemistry extracted and annotated with an image), get the structures off the page, perform structure searches across pages already chemicalized, and see predicted properties for molecules. Users tell us it is intuitive and useful, and it recently received a positive review in JCIM (http://pubs.acs.org/doi/abs/10.1021/ci300046g).

If you want to get a sense of its function, try viewing the Wikipedia page on penicillin V using chemicalize: http://www.chemicalize.org/?url=http%3A%2F%2Fen.wikipedia.org%2Fwiki%2FPhenoxymethylpenicillin. You can easily toggle to the properties view for a substance (http://www.chemicalize.org/structure/#!mol=penicillin%20V&source=parser) by mousing over any instance of its name in the text and clicking on the structure. You can do similar things with a pdf, as seen in the following example: http://www.chemicalize.org/document/#u=http%3A%2F%2Fwww.chemaxon.com%2Fwp-content%2Fuploads%2F2011%2F09%2F1758-2946-3-S1-O3.pdf. Finally, chemicalize.org allows you to perform a chemical similarity search for a molecule of interest (see http://www.chemicalize.org/search/#m=penicillin+V/t=t/h=0), as well as performing a text search for the substance, using all common names to populate a "pre-chemicalized" Google search (http://www.chemicalize.org/websearch/#m=Penicillin%20V/p=0). For more information or to try it out for yourself, please visit http://www.chemicalize.org.

ChemAxon Moves to New European Headquarters

(https://www.chemaxon.com/news/chemaxon_moves_to_new_european_headquarters/): We recently moved from our lovely villa in the Buda hills to dedicated new offices at GraphiSoft Technology Park with space to expand.

ChemAxon Activities at ACS National Meetings

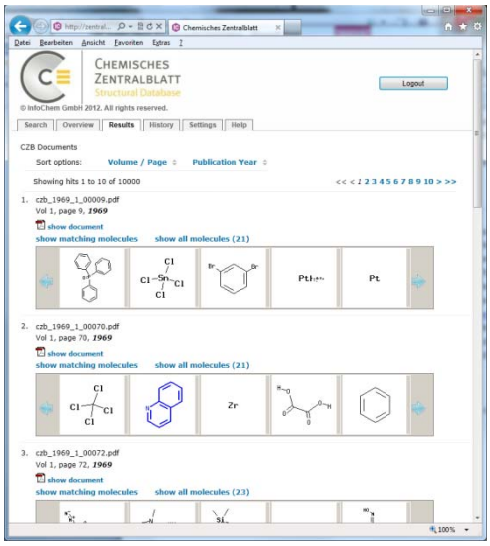
ChemAxon is exhibiting at the Spring ACS National Meeting. We are also organising a session at the Fall ACS National Meeting, entitled "Hunting for hidden treasures: chemical information in patents and other documents." For more information, please see www.ccl.net/chemistry/resources/messages/2012/02/04.001-dir/index.html.

ChemAxon User Meetings

We have just launched our European User meeting agenda, which can be viewed at http://www.chemaxon.com/events/ugm-budapest-2012/#program. Our US meeting (http://www.chemaxon.com/events/ugm-boston-2012/) will be held in September, in Boston.

You can view these and other news stories at: http://www.chemaxon.com/news/.

Chemisches Zentralblatt Structural Database version 1.1 successfully released

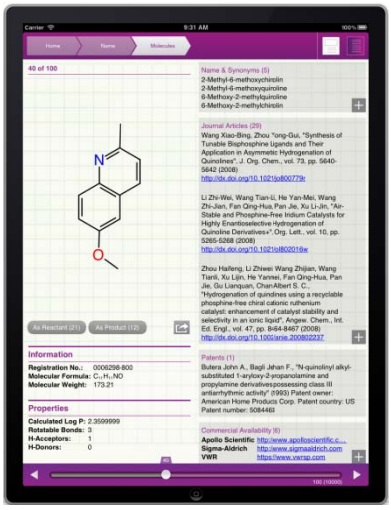


Chemisches Zentralblatt Structural Database is a Web-based application that provides language independent access to the information contained in Chemisches Zentralblatt, one of the most important abstracts journals for the time period 1830-1969. In January, version 1.1 was successfully launched with a considerably improved results display and user interface. The possibility of browsing between pages in the pdf documents has also been enhanced. For more information about the product, please visit http://infochem.de/products/databases/czb.shtml.

Contact us (*http://www.infochem.de/company/contact.shtml*) if you are interested in evaluating Chemisches Zentralblatt Structural Database.

SPRESImobile App launch in March

InfoChem is pleased to announce the launch of the SPRESImobile App in March. This new App for iPhone and iPad will enable users to perform structure and reaction searches also on their mobile devices. The App, developed in cooperation with Eidogen Sertanty, will be downloadable free of charge on iTunes. The free App will give access to a subset of the SPRESI reaction data, ChemReact (*http://infochem.de/products/databases/chemreact41.shtml*). With SPRESImobile, InfoChem is one of the first companies supporting reaction searches on mobile devices.



For more information about SPRESI[web] please visit http://www.infochem.de/products/databases/spresiweb.shtml or www.spresi.com.

Please feel free to contact us *(http://www.infochem.de/company/contact.shtml)*for more information about InfoChem, our current research projects and our products.

ARChem

SimBioSys, Inc. is pleased to present its implementation of predictive stereochemistry within the synthesis planning system ARChem. Various aspects of stereochemical perception, rule generation, and predictions will be discussed in a talk by Professor A. Peter Johnson during the session "Recent Advances in Reaction Searching" in San Diego. Please see the abstract for his talk (CINF-93) on page 61 of this publication.

Title: ***Advanced reaction searching: A comprehensive treatment of stereoselctivity in reactions***

Date/Time: Tuesday, March 27, 2012 - 03:45 PM

Location: San Diego Convention Center**,** Room 27A

CLiDE 5

CLiDE 5 is the latest release of the leading chemical structure extraction software application. CLiDE extracts editable structures from molecular images embedded in journal papers, patents, reports, Web pages, and other types of media, and interacts seamlessly with chemical editors. The program offers, via an easy-to-use GUI, a high-accuracy, convenient, and efficient alternative to copying molecular structures manually.
CLiDE 5 features the following:

- Three different flavors for different usages:
  - CLiDE Standard (for all chemists' everyday use) – automatic extraction of one structure at a time.
  - CLiDE Professional – extraction of multiple structures, and up to a full document in one click, facilitated by a split view showing source image vs extracted structure
  - CLiDE Batch (for database creation and mass data production) – batch handling of many documents
- User-friendly Acrobat-Reader-like GUI
- Seamless integration with ChemDraw, Marvin Sketch, and AccelrysDraw
- Windows, Mac OS and Linux support.
- Supported input documents formats: PDF, doc, docx, HTML
- Supported input image file formats: BMP, GIF, JPEG, PNG, TIFF, and others
- Supported output formats: MDL Molfile, Sdfile, Rgfile, XML, CML, CDX, MRV
- Automatic detection of potential interpretation errors
- Contextual interpretation of generic structures

Both ARChem and CLiDE 5 can be evaluated for free by submitting a demo request at:
http://www.simbiosys.com/products/demo_request.html

Contact Information: info@simbiosys.com ; SimBioSys Inc. 135 Queen's Plate Drive, Unit 520 Toronto, ON M9W 6V1, CANADA. Ph: +1-416-741-4263 Fax: +1-416-741-5084 www.simbiosys.com

Thieme Chemistry

Science of Synthesis.

Best methods. Best results.

Science of Synthesis provides a **critical review of synthetic methodology** developed to-date in the fields of organic and organometallic chemistry. Features include:

- Selection of molecular transformations by world-renowned experts with elaboration on scope and limitations
- Full-text descriptions of synthetic methods with practical experimental procedures immediately applicable in the lab
- Community of over 1,000 experts involved in the review and updating of methods
- Logical organization of the synthetic methods for each functional group
- Intuitive search functions to allow rapid lead generation and route optimization

What's new in SOS 4.0?

Functionality:• Completely revised and optimized interface • Intuitive search functionality, allowing you to quickly enter a (sub-)structure or term • Filtering and sorting of search results • Saving of personal queries and (manually revised) search results in a MySoS account • Citation export as well as enhanced print and export functionalities.

Content: Four Science of Synthesis Knowledge Update volumes (over 2000 printed pages) and three Stereoselective Synthesis volumes from the Science of Synthesis Reference Library (over 3000 printed pages)

Want to know more?

Then come and visit us at the Thieme booth #1029 or attend the following presentation:

Title: ***SOS 4.0: Advances in text, structure, and reaction searching***

Date/Time: March 27, 2012, from 9:00 am to 9:25 am

Location: San Diego Convention Center, Room: Room 27A

The abstract for this talk (CINF-74) appears on p. 55 of this publication.

ACS Chemical Information Division (CINF)
Spring 2012 ACS National Meeting
San Diego, CA (March 25 – March 29)

Abstracts

CINF 1 Polypharmacology, drug repurposing and collaborative drug discovery: Shining light or flash in the pan?

Christopher A Lipinski,
clipinski@meliordiscovery.com, Scientific Advisory Board, Melior Discovery, Waterford, CT 06385-4122, United States

Technical and scientific arguments strongly support polypharmacology, drug repurposing and collaborative drug discovery approaches. In opposition are people / cultural issues that tend to a pessimistic viewpoint. Societal value of efforts in academia versus industry in these areas is unclear. Is there an aspect of academic culture that directs to a higher error rate and lack of confidence compared to work performed in industry? Bias and error in academic biology is well documented as are errors in public chemistry databases. Both academic target identification errors and public database chemistry structural errors are common. Peer reviewed publication pressure induces bias and error. "Hypothesis driven research" can select for high error rates. Unknown is how industry compares to academia in terms of bias and error. How do the internal pressures of metrics, stage gates and timelines in industry compare to the pressures of publish or perish in academia?

CINF 2 High accuracy polypharmacology models for large datasets

S. Joshua Swamidass, *swamidass@gmail.com, Bradley Calhoun, Department of Pathology and Immunology, Washington University in St Louis, St Louis, MO 63108, United States*

To predict the targets and off-targets of molecules we developed predictive models of the small molecule inhibitors of several hundred proteins. Surprisingly, support vector machine (SVM) predictive models could not reliably separate known inhibitors from a set of a half million commonly screened molecules. In contrast, carefully predictors that specifically encode critical pieces of chemistry knowledge (like the similarity principle) yield more powerful models that can reliably extrapolate to large, diverse sets of molecules. In most cases, these models correctly identify---with accuracy greater than 95%---inhibitors in the same test the SVM fails. This study highlights the pitfalls of relying on models outside their domain of applicability, but also suggests predictive models specifically designed to incorporate chemistry knowledge can dramatically outperform generic predictive algorithms.

CINF 3. QSARome of GPCRs

Eugene Muratov[1,2], *murik@email.unc.edu, Guiyu Zhao*[1], *Denis Fourches*[1], *Chris Grulke*[1], *Alexander Tropsha1. (1) University of North Carolina, Chapel Hill, NC 27599, United States (2) A.V. Bogatsky Physical-Chemical Institute NAS of Ukraine, Odessa, Ukraine*

Many marketed drugs for treating CNS disorders have complex G Protein-Coupled Receptors (GPCR) polypharmacology leading either to favorable pharmaceutical outcomes or undesired adverse events. We have curated and integrated binding data for thousands of GPCR ligands extracted from both ChEMBL and PDSP databases. We have developed and extensively validated a panel of Quantitative Structure Activity Relationships (QSAR) classification models for 34 different receptors (i.e., the QSARome). Most models achieved high prediction performances according to a rigorous 5-fold external validation. The QSARome was applied to assess the GPCR binding profiles of 13 external drugs not present in the modeling set, reaching an external prediction accuracy of 70.5%. The QSARome was also used to identify novel compounds with unique target-selective GPCR binding profiles. The QSARome panel is integrated within the Chembench webportal (chembench.mml.unc.edu) providing an effective in silico means to search for novel molecules with the desired GPCR polypharmacology.

CINF 4. Gaussian ensemble screening (GUESS): A new approach to polypharmacology and virtual screening

Violeta Isabel Perez Nueno,
violeta.pereznueno@inria.fr, Vishwesh Venkatraman, Lazaros Mavridis, David W. Ritchie, Orpailleur Team, INRIA Nancy – Grand Est, Vandoeuvre-lès-Nancy, France

We previously introduced a spherical harmonic (SH) approach to compare the 3D shapes of ligands and target binding pockets [1][2][3]. Here, we present a novel extension of this approach to predict relationships between drug classes, which we call Gaussian Ensemble Screening (GUESS). This allows promiscuous ligands and targets to be predicted rapidly without requiring thousands of bootstrap comparisons as in current promiscuity prediction approaches [4].

When using GUESS to find relationships between drug classes in a subset of the MDDR, our approach detects interesting relationships between targets such as GABA A and tyrosine-specific protein kinase, and ACE and neutral endopeptidase, for example, whose dual inhibitors have both been confirmed experimentally [5][6]. Hence, GUESS is a useful way to study polypharmacology relationships, and could provide a novel approach for drug repositioning.

1. Using spherical harmonic surface property representations for ligand-based virtual screening. Pérez-Nueno, V. I. Venkatraman, V. Mavridis, L. Clark, T. Ritchie, D.W. (2011) Molecular Informatics 30, 151-159.

2. Using Consensus-Shape Clustering to Identify Promiscuous Ligands and Protein targets and to Choose the Right Query for Shape-Based Virtual Screening. Pérez-Nueno, V. I. Ritchie, D.W. (2011) J. Chem. Inf. Model. 51, 1233-1248.

3. Predicting drug polypharmacology using a novel surface property similarity-based approach. Pérez-Nueno, V. I. Venkatraman, V. Mavridis, L. Ritchie, D.W. (2011) Journal of Cheminformatics 3 (Suppl 1), O19.

4. Predicting new molecular targets for known drugs. Keiser, M. J. et al. (2009) Nature 462, 175-181.

5. Regulation of GABAA receptor by protein tyrosine kinases in frog pituitary melanotrophs. Castel H, Louiset E, Anouar Y, Le Foll F, Cazin L, Vaudry H. (2000) J Neuroendocrinol. 12, 41-52.

6. Dual ACE and Neutral Endopeptidase Inhibitors: Novel Therapy for Patients with Cardiovascular Disorders. Tabrizchi, Reza (2003) Drugs 63, 2185-2202.

CINF 5. Physical binding site modeling for quantitative prediction of biological activities

Rocco Varela, *rocco.varela@ucsf.edu, Ajay N Jain, Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, California 94158, United States*

Prediction of polypharmacology can be useful at the level of binary putative target identification. Additional utility derives from estimation of the potency of an off-target effect. The latter goal is challenging: in the surprising cases of off-target effects, the off-target ligands will be dissimilar to on-target ligands being designed. Many QSAR methods have utility in making predictions within highly related chemical series, but cannot generally be fruitfully applied for off-target activity quantification due to limited domains of applicability. The Surflex-QMOD approach has been demonstrated to produce accurate and scaffold-independent predictions of binding affinity by constructing an interpretable physical model of a binding site based only on the structures and activities of ligands. Results will be presented establishing that QMOD-derived models produce accurate predictions in cases ranging from limited data of heterogeneous scaffolds to ample data containing related scaffolds. The potential for quantifying off-target effect potencies will also be examined.

CINF 6. Virtual screening of multi-target agents from large chemical libraries by machine learning approach

Yu Zong Chen, *phacyz@nus.edu.sg, Pharmacy, National University of Singapore, Singapore, Singapore*

Selective multi-target agents have been increasingly explored for enhanced therapeutic efficacy. Because they are more sparsely distributed in the chemical space, more efficient methods are needed for searching them.

We have explored machine leaning methods for searching active compounds from large chemical

libraries, Here we present our recent work in exploring machine learning methods for searching dual-target kinase inhibitors and serotonin reuptake inhibitors from large chemical libraries. The dual inhibitor yields, target selectivity, and false hit rates of our methods, trained on individual target inhibitors, are 25%-57% (majority >36%), 95%-99% (against inhibitors of other family members), and 0.007%-0.1% (against 13.56M-17M PubChem and168K MDDR compounds) and 0.0%-3% (against MDDR compounds similar to the dual-inhibitors. They outperformed Surflex-Dock, DOCK Blaster, kNN and PNN in searching 1.02M Zinc clean-leads or MDDR dataset.

Machine learning methods are potentially useful to complement conventional methods for facilitating multi-target drug lead discovery from large chemical libraries.

CINF 7. Structure-based identification of a dual FAAH/COXs inhibitor: Tackling inflammation with a single molecule acting synergistically on multiple proteins

Angelo D Favia, *angelo.favia@iit.it, Andrea Cavalli, Marco De Vivo, Department of Drug Discovery and Development, Istituto Italiano di Tecnologia, Genoa, GE 16163, Italy*

The modulation with a single compound of diverse proteins involved in a complex disease represents one of the frontiers of drug discovery programs.1, 2 Here, the rational structure-based identification of a dual-target hit that simultaneously inhibits Fatty Acid Amide Hydrolase (FAAH) and Cyclooxygenases (COXs) is reported. The dual hit, identified through an interdisciplinary drug discovery effort, active in the low υM range versus both FAAH and COXs represents a suitable starting point for the rational design of a novel drug with superior therapeutic profile. Progresses on the computer-assisted step-wise growth of the hit are here reported.

References

1. Morphy, R. Rankovic, Z., Fragments, network biology and designing multiple ligands. Drug Discov. Today 2007.

2. Bottegoni, G. Favia, A. D. Recanatini, M. Cavalli, A., The role of fragment-based and computational methods in polypharmacology. Drug Discov Today 2011.

CINF 8. Embedding chemistry information literacy skills into the curriculum at James Madison University

Meris A Mandernach[1], *manderma@jmu.edu,* ***Barbara A Reisner***[2], *reisneba@jmu.edu. (1) Libraries & Educational Technologies, James Madison University, Harrisonburg, VA 22807, United States (2) Department of Chemistry and Biochemistry, James Madison University, Harrisonburg, VA 22807, United States*

Historically, chemistry information literacy skills have been taught in an independent course at James Madison University. As part of program assessment, we discovered that students were not making significant gains during the course because much of the content was covered in earlier courses. In 2011, a chemistry librarian and chemistry faculty member revised the course content in order to map critical information literacy skills, as identified by both ACS and Special Libraries Association, into the core chemistry curriculum. Through the use of online tutorials and web guides, content was integrated into courses where students use the information. In this presentation, we will describe how these skills and content have been mapped into individual courses. We will detail the creation of online content and its delivery to several courses with learning management software. We will also discuss successes and roadblocks we have encountered and preliminary assessment data.

CINF 9. Resources for introducing crystal structure information into undergraduate teaching

Gary Battle, *battle@ccdc.cam.ac.uk, Cambridge Crystallographic Data Centre, Cambridge, United Kingdom*

Visualizations and analyses of crystallographically-determined 3D molecular structures can greatly enhance student learning, and are ideally suited to teaching fundamental chemistry concepts including stereochemistry, conformation, chirality and reaction mechanisms. In spite of this, crystallography has historically been poorly represented in University teaching curricula and at best is simply noted at high school level. As a consequence, there is sometimes a

limited understanding of how to retrieve and interpret crystallographic information.

The Cambridge Structural Database (CSD) serves as the worldwide repository of experimentally-measured 3D crystal structures. Built over 45 years and containing more than half-a-million structures the CSD is a vast and ever growing compendium of accurate 3D molecules and is finding increasing application in chemical education.

This talk will focus on continuing efforts to broaden the use of crystallographic data, and in particular to make CSD structures accessible and comprehensible to students and teachers. We will showcase a teaching subset of more than 500 CSD structures created specifically to illustrate key chemical concepts, and a growing collection of teaching materials that make use of this subset in classroom and laboratory environments.

These educational tools are freely available on the web and their utility has recently been recognised by the Chemistry Division of the Special Libraries Association and ACS Division of Chemical Information in their publication Information Competencies for Chemistry Undergraduates: the elements of information literacy. 2nd ed. Sept 2011.

CINF 10. If they build it, will they use it: Using input from students in a chemical literature class in the redisgn of the library's chemistry webpages

Allan K Hovland[1], *akhovland@smcm.edu, Rob C Sloan*[2]. *(1) Department of Chemistry and Biochemistry, St. Mary's College of Maryland, St. Mary's City, MD 20686, United States (2) Library, St. Mary's College of Maryland, St. Mary's City, MD 20686, United States*

The chemical information resources located on the college library's webpages were arranged in a less-than user-friendly fashion. In searching for models for an improved interface, the chemical information instructor (AKH), found several interesting websites and noticed the common denominator was the use of the Springshare software Libguides. Coincidently, the college had just gotten a license to use this software. In discussionswith the library's science liasion (RCS), it was decided to have the introduction to chemical literature class participate in building the new chemistry webpages. A first class assignment had the students go to websites of schools using the Libguide software and required them to identify features they liked and didn't like. An objective for this project is to have the students design pages that will be most useful to their fellow chemistry classmates. Our thought is that if they build it, they will use it.

CINF 11. Explore chemical information teaching resources (XCITR)

Guenter Grethe[1], *ggrethe@att.net, Grace Baysinger*[2], *Rene Deplanque*[3], *Gregor Fels*[4], *Ira Fresen*[3], *Andrea Twiss-Brooks*[5], *Gregor Zimmermann*[3]. *(1) Consultant, Alameda, CA 94502, United States (2) Swain Chemistry and Chemical Engineering Library, Stanford University, Stanford, CA 94305, United States (3) FIZ Chemie Berlin, Berlin, Germany (4) Department of Chemistry, University of Paderborn, Paderborn, Germany (5) John Crerar Library, University of Chicago, Chicago, ILL 60637, United States*

Several years ago, the Division of Chemical Information of the American Chemical Society and the Division of Computer-Information-Chemie of the German Chemical Society established a Collaborative Working Group to foster a transnational dialogue in order to develop a shared approach for the access, exchange and management of chemical information. Within the larger context of the overall approach, the working group developed XCITR, an international repository of chemical information educational material to be used by librarians and instructors in chemical information. XCITR makes full use of features in Web 2.0 technology and is meant to be a hub in which instructors at all levels can deposit and access important teaching materials. We will discuss the history and organization of XCITR, describe technical details and provide examples from this freely available source.

CINF 12. Engaging the wired generation

Jessica A Parr[1], *parr@usc.edu,* ***Norah Xiao***[2]. *(1) Department of Chemistry, University of Southern California, Los Angeles, Ca 90089, United States (2) Science and Engineering Library, University of Southern California, Los Angeles, CA 90089, United States*

Undergraduates are so used to being plugged in and having all desired information at their fingertips. How often has a first years lab report had Wikipedia as a reference? This is not a bad place to start, but not the best source for scientific information. An Information Literacy program has been developed to introduce new students to the scientific databases and the resources available to them in the Science and Engineering Library. With a combination of lecture from the Chemistry librarian and interactive activities, the students had a good time while learning the basics of gathering scientific information. An online tutorial has also been developed to help students, faculty and staff become acquainted with the information resources available to them. This talk will report the details and results of this program, as well as plans to expand it to a larger audience.

CINF 13. Chemical information instruction at ETH Zurich: Review and trends

Martin P. Braendle[1]*, braendle@chem.ethz.ch, Engelbert Zass*[1]*, Lukas Korosec*[2]*, Peter A Limacher*[2]*, Hans P. Luethi*[2]*. (1) Chemistry Biology Pharmacy Information Center, ETH Zuerich, Zuerich, Switzerland (2) Laboratory of Physical Chemistry, ETH Zuerich, Zuerich, Switzerland*

Given the tightly subject matter packed curricula, it is often difficult for librarians to obtain time for scientific information instruction. At ETH, we have developed a two-pillar strategy that tries to meet the students where they have information needs, and to provide tools offering improved access to sources.

We will review our approach that includes problem-oriented units integrated in lab courses (Bachelor course). They are complemented by supporting material for major databases on the web site and individual end-user support. We also investigated ways to improve how students assess information and report of a large study with second-semester students who rated the German Wikipedia and Roempp Online chemistry encyclopedia content with regards to chemical thermodynamics.

Because our instruction is focused on the most important sources, it is complemented with information services that support the user in locating appropriate sources, e.g. a recently introduced textbook portal connecting to our library navigator.

CINF 14. One-shot wonder: Integrating chemical information literacy throughout the curriculum

Linda M Galloway*, galloway@syr.edu, Library, Syracuse University, Syracuse, NY 13244, United States*

Integrating chemical information literacy into an undergraduate chemistry program via a series of guest librarian presentations is proposed as a viable alternative to a formal course. A recent large-scale analysis conducted by Hong Kong Baptist University demonstrated a statistically significant correlation between student performance and library instruction, but only if a certain minimum amount of instruction is provided. This finding, coupled with desired proficiencies articulated in Information Competencies for Chemistry Undergraduates by the Special Libraries Association, Chemistry Division, spurred the development of a set of "one-shot" instruction sessions tied to explicit information competencies. This skill-specific sequence of instruction, attached to defined classes, will enable chemistry students to develop a thorough understanding of chemical literature and how it fits into scholarly communication in the sciences. This paper will detail a plan to systematically integrate, at point-of-need, information literacy skills into the chemistry undergraduate curriculum.

CINF 15. Feedback and training examples from user communities using Elsevier's Reaxys

Christine Flemming*, c.flemming@elsevier.com, Elsevier, New York, NY, United States*

Securing the maximum institutional value for any chemistry resource is reliant on the successful introduction to, and training of, the entire user community on the appropriate use of the e-resource.

This presentation will show some methods and examples developed based on user feedback and will highlight how professionals in academia use Reaxys as a teaching tool for chemistry. Included in this presentation will be examples of the use of social networks, online forums, and other current practices in developing and supporting user communities.

CINF 16. Teaching new graduate students: Chemical information as a research tool

Bonnie L. Fong[1]*, bonnie.fong@rutgers.edu,* ***Darren B Hansen***[2]*, dbhansen@rutgers.edu. (1) John Cotton Dana Library, Rutgers University, Newark, NJ 07102, United States (2) Department of Chemistry, Rutgers University, Newark, NJ 07102, United States*

Chemistry majors do not always have the opportunity to learn information-seeking skills while they are undergraduates. However, as graduate students, they are expected to know how to find chemical information. This session will discuss a collaborative effort between a chemistry professor and a physical sciences librarian at Rutgers University to design a mini-course that helps the students achieve this goal. It will focus on course development, such as: identifying which resources to include (e.g., reference materials, databases), selecting hot topics to discuss (e.g., data management), crafting appropriate assignments, deciding on any supporting readings, etc. Attempts at encouraging collaborative learning, with the assistance of an online site, will also be addressed.

CINF 17. SpringerMaterials: The world's largest resource for chemical and physical properties in materials science

Mikail Shaikh*, email.mikail@gmail.com, eProduct Management, Springer Science and Business Media, new york, New York 10013, United States*

As part of “Springer Databases”, SpringerMaterials is currently one of the world's largest resources for materials properties in chemistry and physics, based off the famous Landolt-Bornstein book series. It contains over 100,000 documents about 250,000 compounds and 3000 compounds in a database format, ranging from nuclear and molecular data to multi-phase systems and advanced materials! With a collection of such magnitude, the ability to efficiently search and navigate becomes almost as important as the content, if not more. Springer Science is actively committed to interacting with researchers to aid the learning curve through sample searches, social media discussion, conference participation, newsletters, flash tutorials, webinars, contests, etc and with librarians through newsletters, surveys, statistics, etc. We also constantly strive to follow the cadence of science, and develop our content in sync.

Find out more about how having a scientifically trained e-Product team encourages a three way conversation between the scientific, information and publishing communities!

CINF 18. Develop novel predictive polypharmacology models with high-quality data

Shuxing Zhang*, shuzhang@mdanderson.org, Department of Experimental Therapeutics, MD Anderson Cancer Center, Houston, TX 77054, United States*

Prediction of drug polypharmacology is of great importance. To this end, we embarked on the construction of high quality chemical and biological databases. Starting from 11,863 co-crystallized ligands, we curated through a rigorous workflow 643 multi-targeting ligands available in 2,131 high-resolution crystal structures with known binding affinities. These data were then employed to build polypharmacology models using our novel diffusion-based target prediction (DTaP) algorithm which is account for the whole multi-targeting profiles for each ligand. With these models, we demonstrated that at least one real target is ranked on the top 5 for 46 known ligands while at least two real targets are ranked on the top 5 for 31 ligands. Interestingly we also found that dasatinib is ranked on the top as an ACK1 inhibitor, and the following-up binding assay confirmed this discovery. These results demonstrated the predictive power of our polypharmacology modeling method trained with high quality data.

CINF 19. Predicting drug polypharmacology using secondary structure element information

Oliver Koch*, oliver.koch@msd.de, BioChemInformatics, Intervet Innovation GmbH, Schwabenheim, GermanyMOLISA GmbH, Magdeburg, Germany*

The protein interactions in protein-ligand binding and protein-protein interfaces can be regarded based on structural similarity of the secondary structure elements. The most prominent example is the protein fold of a protein domain that is more conserved than the amino acid sequence. Proteins with similar fold but

dissimilar sequence and function can bind similar ligands and interact with similar proteins. The next level corresponds to the spatial arrangement of the secondary structure elements around the ligand binding site ("ligand-sensing cores") or in the protein interface ("interface-sensing surfaces"). These similarities in otherwise unrelated proteins can be useful in predicting drug polypharmacology. The successful applications in drug design using predicted polypharmacology in protein-ligand binding will be shown and the analogy in the design of protein-protein interface inhibitors and the potential of polypharmacology prediction will be discussed.

Reference: Koch, O. Future Med. Chem. 2011; 3(6): 699-708.

CINF 20. Assessing drug target association using semantic linked data

***David Wild**, djwild@indiana.edu, Bin Chen, School of Informatics and Computing, Indiana University, Bloomington, IN 47401, United States*

The rapidly increasing amount of public data in chemistry and biology relating to drug discovery provides new opportunities for large-scale data mining for drug discovery. Systematic integration of these heterogeneous sets and provision of algorithms to data mine the integrated sets permits investigation of complex mechanisms of action of drugs. In this work we integrated and annotated data from public datasets relating to drugs, chemical compounds, protein targets, diseases, side effects and pathways, building a semantic linked network consisting of over 200,000 nodes and 1.5 million edges. We developed a statistical model to assess the association of drug target pairs based on their relation with other linked objects. Validation experiments demonstrate the model can identify direct drug target pairs with high precision (AUROC=0.92). Indirect drug target pairs (for example drugs which change gene expression level) are also identified but not as strongly as direct pairs. We further calculated the association scores for 174 drugs from 10 disease areas against 1683 human targets, and measured their similarity using a 174*1683 score matrix. The similarity network indicates that drugs from the same disease area tend to cluster together in ways that are not captured by structural similarity, with several potential new drug pairings being identified. This work thus provides a novel, validated alternative to existing drug target prediction algorithms.

CINF 21. 3D Pharmacophore-based activity profiling for multitarget screening

***Gerhard Wolber**[1], gerhard.wolber@fu-berlin.de, Fabian Bendix[2], Goekhan Ibis[2], Thomas Seidel[2]. (1) Department of Pharmacy, Freie Universitaet Berlin, Berlin, Germany (2) Inte:Ligand GmbH, Vienna, Austria*

With the improved performance of computer hardware, virtual screening methods tend at aiming to increase throughput and not at improving prediction quality. While high restrictivity is suitable for single-target screening campaigns, sensitivity becomes more important in case of activity profiling where the virtual screening process is reversed, i.e. a small number of molecules is screened versus a big number of targets. Current enrichment metrics are not suitable for measuring prediction performance any more and need to be investigated in a more differentiated way. We discuss the challenge of multi-target screening using 3D pharmacophore based activity profiling and suggest methods, protocols and visualization techniques to use activity prediction model collections for virtual screening against multiple targets.

CINF 22. Where have all the good drugs gone?

***Gisbert Schneider**, gisbert.schneider@pharma.ethz.ch, Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland, Zurich, Switzerland HCI H411, Switzerland*

It has been realized that 'druglike' compounds often bind to multiple macromolecular targets. Consequently, predicting target profiles of hit and lead structure candidates has become both a challenge and an opportunity for computer-assisted drug design. No longer is ligand interaction with an individual target the sole objective in molecular design, but multi-dimensional functions guide in silico compound assembly and selection. We will present and discuss the concept of 'adaptive' fitness landscapes and their potential for drug discovery. The computational framework of this approach is based on data projection and Gaussian density estimation, which results in a probabilistic multi-dimensional 'SAR landscape' as a

visual aid for compound prioritization. A related method employs self-organizing map representations of chemical space, which were successfully used for a selection of compounds exhibiting desired activity. Target profile prediction represents a consequent next step towards fully automated de novo drug design that satisfies multiple objectives in parallel.

[1] Reutlinger, M. Guba, W. Martin, R. E. Alanine, A. I. Hoffmann, T. Klenner, A. Hiss, J. A. Schneider, P. Schneider, G. Neighborhood-preserving visualization of adaptive structure-activity landscapes and application to drug discovery. Angew. Chem. Int. Ed. 2011 , doi: 10.1002/anie.201105156.

[2] Schneider, G. Geppert, T. Hartenfeller, M. Reisen, F. Klenner, A. Reutlinger, M. Hähnke, V. Hiss, J. A. Zettl, H. Keppner, S. Spänkuch, S. Schneider, P. Reaction-driven de novo design, synthesis and testing of potential type II kinase inhibitors. Future Med. Chem. 2011 , 3, 415-424.

[3] Schneider, P. Stutz, K. Kasper, L. Haller, S. Reutlinger, M. Reisen, F. Geppert, T. Schneider, G. Target profile prediction for a Biginelli-type dihydropyrimidine compound library and practical evaluation. Pharmaceuticals 2011 , 4, 1236-1247.

[4] Schneider, G. Tanrikulu, Y. Schneider, P. Self-organizing molecular fingerprints: a ligand-based view on drug-like chemical space and off-target prediction. Future Med. Chem. 2009 , 1, 213-28.

CINF 23. Finding promiscuous old drugs for new uses

Sean Ekins[1,2], *ekinssean@yahoo.com, Antony J Williams*[3]*. (1) Collaborations in Chemistry, Fuquay Varina, NC 27526, United States (2) Collaborative Drug Discovery, Burlingame, CA 94010, United States (3) Royal Society of Chemistry, Wake Forest, NC 27587, United States*

In the last 6 years high-throughput screening has been used to identify FDA approved drugs that are active against multiple targets (also termed promiscuity). We have identified 34 studies that have screened libraries of FDA approved drugs against various whole cell or target assays. Each study has identified one or more compound with a new bioactivity that had not been previously described. Thirteen of these drugs were active against more than one additional disease, thereby suggesting a degree of promiscuity. The 109 molecules identified by screening in vitro were statistically more hydrophobic than orphan designated products with at least one marketing approval for a common disease indication or one marketing approval for a rare disease (FDA rare disease research database). We have created a database of in vitro data on old drugs for new uses that could be applied for repositioning these or other molecules for neglected and rare diseases.

CINF 24. Use of course reserves as a gentle introduction to the chemical literature

Donna T. Wrublewski, *dtwrublewski@ufl.edu, George A. Smathers Libraries, University of Florida, Gainesville, Florida 32611, United States*

For most upper level chemistry undergraduates, the Physical/Biophysical chemistry laboratory course is their first required exposure to reading and citing chemical literature, specifically in regards to preparing experimental reports. A basic library literacy lecture is scheduled early on in the semester, designed to assist them with these tasks. Working with the rotating team of instructors who teach these courses, all references in the laboratory manuals were checked for accuracy, updated where needed, and verified to be accessible through online subscriptions or the course reserves system. Specific course reserves system instruction was included in the literacy lecture as a way to guide students to the recommended references and familiarize them with accessing library resources. This ease of access, in conjunction with targeted in-class instruction, should lead to (1) increased library usage and (2) improved quality of students' reports.

CINF 25. Faculty-librarian collaboration yields innovative chemistry seminar program

Valerie K. Tucci[1], *vtucci@tcnj.edu,* ***Benny Chan***[2], *chan@tcnj.edu, Lynn Bradley*[2]*, Stephanie Sen*[2]*, Abby R. O'Connor*[2]*. (1) Library, The College of New Jersey, Ewing, New Jersey 08628-0718, United States (2) Chemistry, The College of New Jersey, Ewing, New Jersey 08628-0718, United States*

Faculty-librarian collaboration at The College of New Jersey has created an innovative Chemistry Seminar Program which enhances the real-world skills of

undergraduate chemistry majors and raises the students' awareness of the value of a commitment to lifelong professional and personal enrichment. The seminar program consists of three interrelated segments: Advising, Chemical Information Literacy and Good Laboratory Practice (GLP). Advising begins as prescriptive advising and grows into developmental advising with emphasis on resume writing, interviewing skills and career options. Chemical information literacy consists of formalized instructional sessions followed by three assessments. The first assessment evaluates student knowledge of basic library skills, the second assessment concentrates on evaluating SciFinder searching capabilities and the third assessment features patent searching using the USPTO database and SciFinder. The CAS Learning Solutions tutorials are integrated into the instructional sessions along with online practice. GLP begins with teaching fundamental principles and often culminates in supporting faculty-student research.

CINF 26. Blind assessment: The unexpected benefits of peer review in a classroom setting

Judith N. Currano, *currano@pobox.upenn.edu, Chemistry Library, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6323, United States*

Peer review is an important part of the scholarly communication process but one in which students receive relatively little formal training. Students in the University of Pennsylvania's graduate-level chemical information course are required to prepare a guide to the literature on a subject of their choice, with the goal of teaching someone to search Penn's resources. The addition of double-blind peer review to the project, in which students reviewed one another's term projects as though they were scientific reviewers for a journal, gave the students practice assessing one another's work. The assignment had the unexpected result of improving the overall quality of the term projects.

CINF 27. Faculty-librarian partnership for a student research presentation in a physical chemistry laboratory course

Donna T Wrublewski[1], *dtwrublewski@ufl.edu, Mine G Ucak-Astarlioglu*[2]*. (1) George A. Smathers Libraries, University of Florida, Gainesville, Florida 32611, United States (2) Department of Chemistry, University of Florida, Gainesville, Florida 32611, United States*

For most upper level chemistry undergraduates, the Physical/Biophysical chemistry laboratory course is their first experience searching and citing chemical literature. A basic library literacy lecture is given early in the semester. At the invitation of one of the class instructors, the chemistry librarian participated in the planning and evaluation of a new library component for the Fall 2011 semester. Currently students are asked to prepare a talk on a paper from the chemical literature; the new component requires students to give a second talk to explain paper selection, describe search strategies, and evaluate additional subject competency gained. Based on the second talk, expected student outcomes are (1) increased chemical literature literacy and general subject knowledge; (2) improved abilities using search engines and correct keywords; (3) improved library course content from developing and adapting more efficient search techniques; (4) improved education quality and student empowerment from fostering closer chemistry department/library collaborations.

CINF 28. CAS learning solutions: Training at the point of need

Jayne A. Knoop, *jknoop@cas.org, Department of Learning and Support Solutions, CAS, Columbus, OH 43204, United States*

As the world's authority for chemical information, CAS analyzes and organizes complex content for an international customer base. SciFinder is the chemistry research tool of choice for students at more than 1,800 academic institutions and research scientists at government and commercial organizations worldwide. Patent experts rely on the advanced search and analysis capabilities of STN. To most effectively serve the needs of its diverse customers, CAS offers a variety of training options, including in-product Help, e-learning tutorials, webinars, virtual classes, workshops, patent

forums, and custom private training. Session attendees will discover the substantial value that CAS training adds at the point of need for SciFinder and STN customers.

CINF 29. Chemistry Referece Resolver: A tool to simplify reference retrieval

Oleksandr Zhurakovskyi, *oleksandr.zhurakovskyi@chem.ox.ac.uk, Department of Chemistry, University of Oxford, Oxford, United Kingdom*

High quality research requires thorough literature screening and numerous databases are available to ease this process. However, chemists regularly need to find a paper manually from its reference details. We have developed Chemistry Reference Resolver (http://chemsearch.kovsky.net) as a tool to facilitate this process. This tool accepts, as input, a reference in a number of styles (ACS, NPG style, DOI, etc) and redirects the user directly to the corresponding online abstract/PDF link. As currently configured, the Resolver has a number of plugins available for major browsers.

CINF 30. Using LibGuides to enhance large-enrollment chemistry lab courses

Jeremy R Garritano, *jgarrita@purdue.edu, Purdue University Libraries, Purdue University, West Lafayette, IN 47906, United States*

At a large research university, teaching chemical information literacy skills to the thousands of students in chemistry lab courses can be difficult. In the fall of 2011, using LibGuides, the M.G. Mellon Library of Chemistry at Purdue University created a site to support students in first- and second-year chemistry labs and to help with knowledge retention between the two years and beyond. Gathering feedback from faculty, teaching assistants, advisors, and others within the Chemistry Department, the site is a collaborative effort designed to bring together information on writing, communication, and information seeking skills related to the chemistry lab courses. This site helps to supplement sections of the lab manual and presents a single, stable resource for the different versions of first- and second-year chemistry lab courses taught at the University. Implementation, usage statistics and user feedback will be shared.

CINF 31. Learning about cheminformatics through an education wiki

Martin A Walker[1], *walkerma@potsdam.edu, Aileen E Day*[2], *Antony J Williams*[2], *Lorna M Thomson*[2]. *(1) Department of Chemistry, State University of New York at Potsdam, Potsdam, New York 13676, United States (2) Royal Society of Chemistry, Cambridge, United Kingdom*

The Royal Society of Chemistry (RSC) recently unveiled a new chemical education wiki, called RSC LearnChemistry:Share. This site is designed to bring the features of RSC ChemSpider into an education site, as well as to create a workspace where educators can share their resources (quizzes, lab experiments, tutorials, etc.).

The wiki is built around a cheminformatics platform, and educators will naturally be exposed to basic cheminformatics concepts as they use the site and contribute content. For example, InChIs can be generated in the site and entered by teachers, to define answers to their quiz questions. Substance pages are found by InChIKey-directed structure searches. These and other pages can display "live" ChemSpider data, including spectra. As a community of educators develops on the site, these active contributors will learn to use cheminformatics tools in the best way – by using them to educate their students.

CINF 32. Bringing faculty, students and librarians together: Lessons and opportunities for ACS on campus after two years

S. Sara Rouhi, *s_rouhi@acs.org, Library Relations, ACS Publications, Washington,, DC 20009, United States*

ACS on Campus, the Publications Divisions' campus outreach program is now almost two years old. The curriculum was developed by ACS Pubilcations Library Relations to reintroduce faculty and students to their libraries' resources. As the program enters its third year questions arise: Is ACS on Campus meeting the needs of the librarian community? To what extent will librarians continue to embrace a program that requires outreach work on their part? Does the lack of chatter

about the program in the librarian community indicate weaknesses within the program that need to be addressed?

This presentation provides an overview of the program including feedback from librarians, students, and faculty. It will examine potential weaknesses within the program and opportunities for improvement in light of expansion opportunities. The presenter will be looking for direct feedback from session participants about the relevance of the program in light of their current situations at their institutions.

CINF 33. Use of screening results to validate a diversity subset of an HTS library

Rohan S Patil[1,2], *rspatil@umail.iu.edu, Maureen Beresini*[1], *Nicholas Skelton*[3]. *(1) Department of Biochemical and Cellular Pharmacology, Genentech, South San Francisco, California 94080, United States (2) Department of Chemical Informatics, Indiana University, Bloomington, Indiana 47408, United States (3) Department of Computational Chemistry and Cheminformatics, Genentech, South San Francisco, California 94080, United States*

High-throughput screening (HTS) is an important element in the discovery of small molecule leads. We have routinely screened the full Genentech library with over 1 million compounds. It is not always possible or desirable to screen the full library. Screening a portion of the library may speed the screening and streamline the hit triage process. Additionally, there may be cases in which reagent limitations or lower assay throughput warrant a smaller compound set. Consequently, an effort was undertaken to select five diverse subsets of the entire library, then compare the performance of each of them with that of others as well as to that of the entire library. Finally, we identify the diversity compound set that could be used in a subset screen in place of a whole library screen.

CINF 34. Development of a screening informatics system at the UNM Center for Molecular Discovery, an NIH MLP specialty center

Jeremy J Yang[1,2,3], *jjyang@salud.unm.edu, Oleg Ursu*[1,2], *Stephen L Mathias*[1,2], *Cristian G Bologa*[1,2], *Anna Waller*[2], *Annette M Evangelisti*[2], *Gergely Zahoransky-Kohalmi*[1,2], *Tudor I Oprea*[1,2]. *(1) Department of Biochemistry and Molecular Biology, University of New Mexico, Albuquerque, NM 87131, United States (2) Center for Molecular Discovery, University of New Mexico, Albuquerque, NM 87131, United States (3) School of Informatics & Computing, Indiana University, Bloomington, IN 47405, United States*

Robotic, automated, molecular screening against biological targets is not new, nor is the need for supporting informatics systems. However, continued advances in contributing technologies, plus increased expectations, pose continuing challenges for informatics systems developers. These advances include: (1) New methodology, such as high-content and multiplex bioassays, (2) More relevant public data, and (3) New privacy and collaboration models, and (4) Advances in cheminformatics and bioinformatics methodology. In this poster we present the screening informatics system developed at the University of New Mexico Center for Molecular Discovery, which combines industry-standard commercial with open-source software components, and custom code developed at UNM. The choices made as to components and overall design were rationally and pragmatically driven, with time and resource constraints combining with technical objectives, resulting in a novel, hybrid solution. A functioning system was required continuously, for ongoing projects, necessitating an evolutionary approach. These challenges are both difficult and typical for advancing informatics systems at productive organizations, where upgrading infrastructure takes place in a contex of operational imperatives.

CINF 35. Thermodynamical properties of small Pd clusters on the stoichiometric and defective TiO2 (110) surfaces studied with first-principle methods

Jin Zhang, *jinzhang@chem.ucla.edu, Anastassia Alexandrova, Department of Chemistry and Biochemistry, UCLA, Los Angeles, California 90095, United States*

Using first-principle methods, we studied the adsorption properties of the sub-nano Pd clusters on the stoichiometric and the defective titania surfaces. In particular, we mapped the potential energy surfaces (PES) of a Pd atom on three types of titania surfaces. With the data obtained from these calculations, we constructed a square lattice model describing the movement of Pd monomers with the Monte Carlo method to simulate the cluster growth and sintering processes at various temperatures. We found that on the stoichiometric surface or surface with Ti-interstitial atom, the Pd monomers tend to sinter into larger clusters, whereas the Pd dimer, trimer and tetramer appear to be relatively stable below 600 K. This result agrees with the standard sintering model of transition metal clusters and experimental observations.

CINF 36. Development of a hybrid method combining quantum mechanical calculations and discrete molecular dynamics for metallo-protein modeling

Manuel Sparta, *sparta@chem.ucla.edu, Anastassia N Alexandrova, Department of Chemistry and Biochemistry, UCLA, LOS ANGELES, CALIFORNIA 90095-1569, United States*

Natural metallo-enzymes are known for their outstanding catalytic dexterity. The quality modeling of metallo-enzymes is highly desirable, both to understand the mechanism of their proficiency, and to eventually design artificial metallo-enzymes. However, the challenge of this modeling is great, because it is needed to account for both the cooperative dynamic moves of the protein, and electronic structure of the metal. We report a new method that combines an extensive statistical mechanical sampling of the protein, achieved with Discrete Molecular Dynamic (DMD), and the quantum mechanical (QM) description of the active site, QM/DMD. The method is fast and robust. Testing of QM/DMD at several levels was done on Rubredoxin. QM/DMD successfully predicts the structures of the Fe(II) and Fe(III) forms of Rubredoxin, and their mutants, properly describes the unusual weak H-bonds between SCys and aliphatic C-H groups near the active site, and captures the response of the RedOx potential to mutations.

CINF 37. Learning to predict more chemical reactions: Model extensions and an expanded training set

Matthew A Kayala, *mkayala@ics.uci.edu, Pierre Baldi, Department of Computer Science, University of California, Irvine, Irvine, California 92627, United States*

In previous work, we introduced a machine learning approach to predict productive mechanistic reactions. The representations and methods presented allowed for practical mechanistic reaction prediction over the set of polar organic reaction. Here we describe several improvements to our previous approach. First, we describe extensions to the orbital interaction representation to cover pericyclic, radical, and stereospecific reactions. Next we describe a new larger and more diverse training set of chemical reactions, derived from extending the use of the Reaction Explorer expert system along with manual curation from literature and graduate level texts. Then, we show how our general machine learning approach, with the new representations and expanded dataset, exhibits excellent performance results. Finally, a multi-step pathway prediction application is made available.

CINF 38. COBRA: Computational brewing approach to predicting the molecular composition of organic aerosols

David R Fooshee[1], *dfooshee@uci.edu, Tran B Nguyen*[2], *Sergey A Nizkorodov*[2], *Julia Laskin*[3], *Alex Laskin*[4], *Pierre Baldi*[1]. *(1) Department of Computer Science, University of California, Irvine, Irvine, CA 92697, United States (2) Department of Chemistry, University of California, Irvine, Irvine, CA 92697, United States (3) Chemical and Materials Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States (4) Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352, United States*

We introduce a novel Computational Brewing Approach (COBRA) to model oligomerization chemistry stemming from repetitive condensation and addition reactions of monomers in isoprene photooxidation organic aerosols. COBRA takes two sets of data as input: a list of the initial chemical structures making up the starting molecular pool, and a list of reaction rules defining potential chemical transformations within the system. The reactions are propagated through several iterations, with products of previous iterations serving as reactants for the next one. A set of four reactions including esterification, aldol condensation, and hemiacetal formation, along with 27 seed molecules, were used to predict products of oligomerization in isoprene photooxidation secondary organic aerosol (SOA). The simulation generated thousands of unique structures in the mass range of 120-500 Da, and correctly predicted greater than 70% of the peaks observed by high-resolution mass spectrometry (HR-MS) of isoprene SOA. Selected structures predicted by the simulation were confirmed with tandem mass spectrometry (MSn). The model aids in structure elucidation from tandem mass spectrometry by offering up to a 100-fold reduction in the number of possible isomers for a given molecular formula. COBRA is not limited to atmospheric aerosol chemistry; it can also be applied to the prediction of reaction products in other environmental complex mixtures for which reasonable reaction mechanisms and seed molecules can be supplied by experimental or theoretical methods.

CINF 39. High-throughput 3D structure prediction of small molecules

Peter Sadowski[1,2], *peter.j.sadowski@uci.edu, Arlo Randall*[1,2], *Pierre Baldi*[1,2]. *(1) Department of Computer Science, University of California Irvine, Irvine, CA 92697, United States (2) Institute for Genomics and Bioinformatics, Irvine, CA 92697, United States*

Although databases such as PubChem contain 3D structures for millions of molecules, there is a much larger space of small drug-like molecules that researchers wish to explore. Next-generation drug discovery projects require a high-throughput way of predicting accurate 3D structures of virtual molecules. State-of-the-art density functional theory (DFT) methods are accurate but slow, while lower levels of theory such as molecular mechanics models are insufficient for describing the complex bonding of organometallic molecules. A recent system named COSMOS has demonstrated its ability to quickly predict 3D structures for millions or billions of virtual molecules given a large library of precomputed rigid fragments. Here we present an extensible library of 100,000 unique, highly accurate fragment structures that we have produced from isomeric SMILES codes, using a combination of pattern matching, molecular mechanics, and DFT.

CINF 40. Predicting inactive and active conformations of the dopamine D2 receptor

Fan Liu, *liufan.brooks@gmail.com, Ravinder Abrol, Dennis A Dougherty, William A Goddard III, Chemistry, California Institute of Technology, Pasadena, California 91125, United States*

G-protein coupled receptors (GPCRs) achieve their functional versatility by adopting various structural conformations defined by different orientations of their characteristic seven transmembrane helices. We predicted inactive and active ensembles of conformations for the dopamine D2 receptor using GEnsemble, which efficiently samples trillions of possible conformations based on different helix orientations. Using the predicted conformational ensembles for inactive and active receptor states, we predicted the binding sites for dopamine, which provides insights into potential receptor activation mechanisms. The binding sites suggest mutagenesis experiments involving the D2 receptor that will provide validation to the binding sites and the activation mechanism proposed based on those sites.

CINF 41. Impact of retractions on the chemical literature

Elsa Alvaro, *ealvaro@indiana.edu, School of Library and Information Science, Indiana University, Bloomington, Indiana 47405, United StatesChemistry Library, Indiana University, Bloomington, Indiana 47405, United States*

Article retractions have recently attracted a lot of scholarly and popular interest. While most of the studies are focused on the biomedical literature, an analysis of the extent, impact, and causes of retractions in the chemical literature is still missing.

In this work, we report a longitudinal study of retractions across chemistry journals starting from 1990. We have carried out statistical analysis on data collected from retraction notices, including reasons for retraction, agents, and rate, and studied potential correlations with parameters such as the impact factor of the journal. We have also performed bibliometric studies and applied network theory principles to understand the impact and propagation of invalid research in the chemical literature. The results of this work show that while some of the findings are consistent with those reported in other fields, others appear to be distinct of the chemical literature.

CINF 42. Cheminfomatic modeling of human CC chemokine receptorome

Terry-Elinor Reid[1], *yasmanii@hotmail.com, Huzefa Rangwala*[2], *Samantha McCullough*[1], *Muhammad Habib*[1], *Simon Wang*[1]. *(1) Pharmaceutical Sciences, Howard University, United States (2) George Mason University, United States*

CC chemokine receptors (CCRs) represent one subfamily of chemokine receptors. Among them, CCR2 is implicated in the inflammatory responses while CCR5 acts as the primary co-receptor by which HIV infects human T cells. Thus CCRs represent important targets for modern drug discovery. To study the complex binding profiles of CCRs we have collected data sets of structurally diverse molecules with known affinities for the whole human CC chemokine receptorome. They had been rigorously curated prior to the molecular descriptors calculation. Externally predictive cheminformatic models were developed using multiple algorithms and further validated by five-fold external validation. We also employ advanced principles such as semi-supervised learning and multi-task learning in order to capture the inter-targets (CCRs subtypes) information. The multitude of predictive models at the receptorome scale provide valuable tools for virtual screening of chemical libraries to identify structurally novel ligands as well as to address the complex selectivity issues.

CINF 43. Evolutionary computational modeling of β-diketo acids for virtual screening of HIV-1 integrase inhibitors

Gene M Ko[1], *gko@sciences.sdsu.edu, A. Srinivas Reddy*[2], *Rajni Garg*[1], *Sunil Kumar*[3], *Ahmad R Hadaegh*[4]. *(1) Computational Science Research Center, San Diego State University, San Diego, CA 92182-1245, United States (2) La Jolla Bioengineering Institute, San Diego, CA 92121, United States (3) Department of Electrical and Computer Engineering, San Diego State University, San Diego, CA 92182-1309, United States (4) Department of Computer Science and Information Systems, California State University San Marcos, San Marcos, CA 92096-0001, United States*

We have used a differential evolution-binary particle swarm optimization (DE-BPSO) feature selection method to develop a QSAR model for 91 structurally diverse β-diketo acids which are potent HIV-1 integrase inhibitors. DE-BPSO is a novel feature selection method which requires fewer generations than BPSO to select a good small subset of descriptors. These descriptors are then used for developing QSAR models. 387 constitutional, geometrical, topological, electrostatic, and quantum-chemical descriptors were computed for each of the 91 structures and QSAR models were developed. The top ranked model satisfying predictive statistical constraints ($r2 > 0.6$, $r2validation > 0.5$, $r2test > 0.5$) was considered for analysis of the physiochemical features of β-diketo acids conducive for inhibition of HIV-1 integrase. The model suggests that molecular volume of the chemical compounds plays a dominant role in the inhibition of HIV-1 integrase. We also used this model successfully as a virtual screening tool to predict the biological activities closer to their experimental values of 18 2-pyrrolinone derivatives and 32 rhodanine containing compounds. This model can be used to identify novel compounds which may have similar structural properties as β-diketo acids with inhibitory effects towards HIV-1 integrase. We believe that DE-BPSO is a novel feature selection method for QSAR model development of other chemical compounds.

CINF 44. Molecular dynamics of the Hsp70 chaperone in response to nucleotide and substrate: A coarse-grained perspective

Ewa I. Golas[1,2], *ewa.golas.chem@gmail.com, Gia G. Maisuradze*[2], *Patrick Senet*[2,3], *Stanislaw Oldziej*[2,4], *Cezary Czaplewski*[1,2], *Harold A. Scheraga*[2], *Adam Liwo*[1,2]. *(1) Department of Chemistry, University of Gdansk, Gdansk, Poland (2) Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, United States (3) Laboratoire Interdisciplinaire Carnot de Bourgogne, Universite de Bourgogne, Dijon Cedex, France (4) Laboratory of Biopolymer Structure, Intercollegiate Faculty of Biotechnology, University of Gdansk, Gdansk, Poland*

The 70kDa heat-shock (Hsp70) proteins form a class of chaperones recognized for their diverse and essential roles in the domain of protein repair, folding assistance, and agglomerate prevention. The focus of the present work was to determine and characterize the motion of the bacterial Hsp70 from Escherichia Coli via canonical molecular dynamics simulation. The UNRES forcefield was used to model the whole chaperone, with the implicit placement of nucleotide in the nucleotide binding domain (NBD), and the explicit introduction of a guest peptide in the substrate binding domain (SBD). The definition of an 'implicit' nucleotide was achieved though the application of harmonic restraints on the NBD. The characterization of the observed motions included an analysis of internal angles, distances, and inter-domain interactions. A comparison of the behavior of the chaperone with previous simulations preformed without substrate (earlier work) was also realized. Two systems with variant guest peptide were independently studied.

CINF 45. Searching putative targets in silico for anti–prion compounds

Jorge Valencia[1], *lip09jmv@sheffield.ac.uk, Beining Chen*[2], *Val Gillet*[1]. *(1) Information School, The University of Sheffield, Sheffield, South Yorkshire S1 4DP, United Kingdom (2) Chemistry, The University of Sheffield, Sheffield, South Yorkshire S3 7HF, United Kingdom*

Transmissible Spongiform Encephalopathies are fatal neurological disorders caused by a proteinaceous infectious particle (PrPSC). The PrPSC transforms the normal prion isoform (PrPC) to the infective conformation by a mechanism which remains unknown. In previous work we identified a set of anti-prion compounds with EC50 in the range 1–10 nM in a mouse cellular model, however, the target(s) of these compounds is unknown. In this project, we describe an in silico protocol for target prediction using inverse docking. A set of 333 differentially expressed genes involved in the transformation and identified using microarray analysis were collected from the literature; 168 corresponding structures were downloaded from the PDB; and a diverse set of five anti-prion compounds were docked to the proteins using GOLD. From the results, we have identified a set of putative targets shared by the compounds. Next, we aim to corroborate the results through a proteomics analysis.

CINF 46. On the accuracy of chemical structures found on the internet

Andrew D. Fant[1], *andrew.fant@unc.edu, Eugene Muratov*[1], *Denis Fourches*[1], *Antony J. Williams*[2], *Alexander Tropsha*[1]. *(1) Division of Chemical Biology and Medicinal Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7363, United States (2) Royal Society of Chemistry, Wake Forest, NC 27587, United States*

The Internet has been widely lauded as a great equalizer of information access. However, the absence of any central authority on content places the burden on the end-user to verify the quality of the information accessed. We have examined the accuracy of the chemical structures of ca. 150 major pharmaceutical products that can be found on the internet. We have demonstrated that while erroneous structures are commonplace, it is possible to determine the correct structures by utilizing a carefully defined structure validation workflow. In addition, we and others have shown that the use of un-curated structures affects the accuracy of cheminformatics investigations such as QSAR modeling. Furthermore, models built for carefully curated datasets can be used to correct erroneously reported biological data. We posit that chemical datasets must be carefully curated prior to any cheminformatics investigations. We summarize best practices developed in our groups for data curation.

CINF 47. Spectral clustering of chemical data: A Lanczos-based approach

Sonny Gan[1], *liq09sg@shef.ac.uk, Valerie J Gillet*[1], *Eleanor J Gardiner*[1], *David A Cosgrove*[2]. *(1) Information School, University Of Sheffield, Sheffield, United Kingdom (2) AstraZeneca, Alderly Park, United Kingdom*

The application of traditional clustering algorithms to partition chemical datasets is well established. Recently, clustering methods which partition a dataset based upon the eigenvectors of an input matrix have gained considerable attention in computer vision, providing excellent results for a variety of tasks. Despite this, their application to chemical data has been limited. A non-overlapping spectral clustering approach (L-NOSC), which utilizes a modified Lanczos algorithm to identify the eigenpairs of a matrix, has been developed. This L-NOSC affords considerable computational advantages when compared with other spectral clustering methods that rely on a full matrix diagonalization procedure. The ability of the L-NOSC algorithm to cluster several activity datasets, described by five different descriptors, has been evaluated using the Quality Clustering Index. Finally, the performance of the algorithm has been compared to both the leading traditional clustering methods and a spectral clustering algorithm which uses a full matrix diagonalization, with promising results.

CINF 48. Structure based pharmacophore screening for new P-gp inhibitors

Freya Klepsch[1], *freya.klepsch@univie.ac.at, Katharina Prokes*[1], *Zahida Parveen*[2], *Peter Chiba*[2], *Gerhard F Ecker*[1]. *(1) Department of Medicinal Chemistry, University of Vienna, Vienna, Vienna 1090, Austria (2) Institute of Medical Chemistry, Medical University of Vienna, Vienna, Vienna 1090, Austria*

Overexpression of the xenotoxin transporter P-glycoprotein (P-gp) is one major reason for the development of multidrug resistance (MDR) leading to the failure of antibiotic and cancer therapies. Inhibitors of P-gp have thus been advocated as promising candidates for overcoming the problem of MDR.

By applying an exhaustive docking protocol that implied SAR information into the pose selection process, a validated binding hypothesis for propafenone analogs in P-gp could be determined. The docking complex was further used for the generation of a structure based pharmacophore model that comprised important interaction points and exclusion volumes. The model was validated by our in house data set as well as by spiked DUD sets. Furthermore, screening the Life Chemicals database retrieved a number of hits that were tested experimentally for their P-gp inhibiting activity. Among those, four compounds showing new chemical scaffolds were found to be active in the µmol range.

CINF 49. Perspective in computational approaches applied to drug discovery problems

Christine Humblet, *christine.humblet@comcast.net, Discovery Chemistry Research & Technologies, Lilly Research Laboratories, Indianapolis, IN 46285, United States*

Computational chemistry and cheminformatics have seen considerable progress over the past decades. This presentation will highlight progress seen over time, illustrate current best practices, and build a forward-looking perspective based on a practitioner's point of view.

CINF 50. Rational, data-driven approach to lead optimization

Dan J Warner, *Dan.Warner@astrazeneca.com, R&D Montreal, AstraZeneca, St. Laurent, QC H4S 1Z9, Canada*

At the very heart of the role of a medicinal chemist or drug designer is the ability to link chemical structure to molecular properties. The traditional approach to deriving these structure-property relationships (SPR) has been to encode known compounds in the form of molecular 'descriptors' and link them to experimentally determined properties in a quantitative fashion. At AstraZeneca, we have been at the forefront of the emerging field of matched molecular pairs analysis (MMPA) or inverse-QSAR, with a number of recent publications in the area1,2. As the name implies, this turns the traditional approach of investigating these relationships on its head in the identification of changes in structure that correspond to a desirable changes in properties3. The presentation will attempt to summarize

the current state-of-the-art with respect to the literature, AstraZeneca's in-house system for MMPA, and how we can use it to expedite the multi-objective optimization of chemical leads.

1. Griffen, E. Leach, A. G. Robb, G. R. Warner, D. J. Matched Molecular Pairs as a Medicinal Chemistry Tool. J. Med. Chem. Perspectives. 2011, A.S.A.P.

2. Warner, D. J. Griffen, E. J. St-Gallay, S. A. WizePairZ: A Novel Algorithm to Identify, Encode, and Exploit Matched Molecular Pairs with Unspecified Cores in Medicinal Chemistry. J. Chem. Inf. Model. 2010, 50, 1350-1357.

3. Leach, A. G. Jones, H. D. Cosgrove, D. A. Kenny, P. W. Ruston, L. MacFaul, P. Wood, J. M. Colclough, N. Law, B. Matched Molecular Pairs as a Guide in the Optimization of Pharmaceutical Properties; a Study of Aqueous Solubility, Plasma Protein Binding and Oral Exposure. J. Med. Chem. 2006, 49, 6672-6682.

CINF 51. WOMBAT and WOMBAT-PK: Ten years

Tudor I Oprea, *tudor@sunsetmolecular.com, Sunset Molecular Discovery LLC, Santa Fe, NM 87505, United States*

In 2001, the first version of WOMBAT was assembled (over 20,000 entries). This database, now with over 300,000 entries, indexes medicinal chemistry literature. WOMBAT is used for predictive chemical biology efforts. Some use cases for WOMBAT (from literature) will be highlighted. WOMBAT-PK, initially centered on pharmacokinetics data, was developed in 2003 to take advantage of the wealth of information for approved drugs. The use of WOMBAT-PK, most recently in relationship to BDDCS (Biopharmaceuticsl Drug Disposition and Classification System) and its application in drug discovery and development, will be detailed.

CINF 52. Developing "Best Practices" in predictive cheminformatics for drug-discovery applications

Curt M Breneman[1], *brenec@rpi.edu, Michael Krein*[1], *Margaret McLellan*[1], *Tao-wei Huang*[1], *Lisa Morkowchuk*[1], *Dimitris K. Agrafiotis*[2]. *(1) Department of Chemistry & Chemical Biology, Rensselaer Polytechnic Institute, Troy, NY 12180, United States (2) Johnson & Johnson Pharmaceutical Research & Development, LLC, Spring House, PA 19477, United States*

After decades of development and nearly continuous use (and abuse) in support of drug discovery efforts, QSAR and related descriptor-based statistical learning methods have earned mixed reviews throughout their checkered past. Why is this? In practice, it is necessary for users to choose from a long list of available descriptors and then select a machine learning method with the hope that the resulting model will have a chance of representing the physical effects that actually control the endpoint of interest. All too often, the resulting models might have been over-trained using small datasets, or could have limited applicability domains. Inappropriate descriptor choices can also doom such efforts. Why and when does this happen? In this talk, the evolution of "Best Practices" in predictive cheminformatics will be illustrated by way of a series of example scenarios, concluding with some guidelines and recommendations for both model builders and end users.

CINF 53. Cheminformatics for material discovery: Representation, searching and screening *of porous materials*

Richard L Martin, *richardluismartin@lbl.gov, Maciej Haranczyk, Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States*

We summarize the recent advancements in material discovery facilitated by the application of cheminformatics concepts to crystalline porous materials. We discuss challenges involved in developing descriptors and comparison techniques for these structures, and comment on the role of the similar property principle in material analysis.

Our approach is to focus on void space, rather than the overall structure – this allows us to describe a material from the point of view of a guest molecule. We perform the Voronoi decomposition to obtain a periodic graph representation of a material's void space, and construct novel descriptors – Voronoi holograms – by automatically inspecting the graph. Through application of a modified Tanimoto similarity coefficient and MaxMin diversity selection, we illustrate the

calculation of (dis)similarity using this descriptor, and the retrieval of a diverse and representative subset of promising candidate materials for CO2 capture not obtainable through the use of existing structural descriptors.

CINF 54. New strategies to normalize chemical structure representations and weed-out impractical small molecules

Evan Bolton, *bolton@ncbi.nlm.nih.gov, National Center for Biotechnology Information, Bethesda, Maryland 20894, United States*

What is a reasonable chemical structure representation? A typical chemist reply is: one that I can accurately and unambiguously interpret. A practical answer, however, a chemist interpretation and a computer interpretation may be at distinct odds when defining what "reasonable" means. There are many implicit vs. explicit aspects to a chemist interpretation lost when interpreting a chemical structure. There are missing formal charges (e.g., nitro groups represented by "*N(=O)O"), implied resonance by using double bonds (e.g., carboxylic acid represented by "*C(=O)=O" rather than "*C(=O)O"), and so on. A chemist might overlook and implicitly understand and correct these aspects in their understanding, but a computer is often confused and happy to reject a chemical structure as being unreasonable without substantial chemical intuition programmed into the interpretation. Alternatively, a chemical structure may seem completely reasonable by all known valence rules, etc., to a compute but is completely reject by a chemist as being impossible to exist (e.g., "OOOOOOOOO", poly-peroxides).

This presentation will outline new strategies being explored at PubChem to determine whether a chemical structure is unreasonable or impractical. The basic approach is to take advantage of a comprehensive survey of first, second, and third order nearest-neighbor environments and develop a self-learning and automated statistical-based approach to reliably predict the likelihood a chemical is reasonable (or its better representation!) without the need for developing and maintaining a set of transforms (e.g., SMIRKS).

CINF 55. Efficient perception of proteins and nucleic acids from atomic connectivity

Roger A Sayle, *roger@nextmovesoftware.com, NextMove Software, Cambridge, Cambridgeshire CB4 0EY, United Kingdom*

A common problem in the conversion of molecular data file formats is the annotation of amino acid and nucleic acid residues not explicitly represented in "small molecule" file formats describing only element type and 3D co-ordinates or atomic connectivity. This problem has limited the interoperability between chemical information processing programs and has led to the situation where molecular graphics programs currently treat the same molecule differently depending upon the file format that it is stored in. An algorithm has been developed to rapidly identify polypeptides and nucleic acids from simple connectivity that can assign standard atom names, residue names, residue numbers and chain identifiers to each atom, and bond orders to each bond. One of the features of this algorithm is a very efficient method for identifying a sidechain from a set of rooted graphs, which has running time linear in the number of atoms in the sidechain.

CINF 56. Organization and analysis of information for biotherapeutics research

Hugo O Villar[1], *hugo@altoris.com, Mark R. Hansen*[1], *Eric Feyfant*[2]*. (1) Altoris, Inc., La Jolla, CA 92037, United States (2) Global Biotherapeutic Technologies Dept, Pfizer, Inc., Cambridge, MA 02140, United States*

For decades, different chemoinformatics tools have guided the identification of structure activity relationships in chemical series towards the development of new therapeutic agents. Now, active research aimed towards the development of biopolymers as therapeutic agents has generated the need for new tools to organize and identify structure activity relationships (SAR) in large volumes of data being generated as has been done historically in medicinal chemistry. Many of the concepts and visualization tools used in chemoinformatics can be adapted to deal with the challenges presented with peptides, proteins and nucleic acids. We will present a new desktop application, SARvision|Biologics to mine and visualize trends in data generated in biologics research. We will show how some concepts used in

cheminformatics research need to be tailored and can be adapted for biologics research.

CINF 57. Markush structure usability in patent and combinatorial chemistry: New approaches and software tools

Wei Deng, *ddeng@chemaxon.com, Szabolcs Csepregi, ChemAxon, United States*

Markush structures are widely used in combinatorial libraries and patents. However, the flexibility and complexity of Markush structures make them difficult to create, index, visualize, search, and enumerate. Recent improvement of various ChemAxon applications to toggle this problem will be introduced in this presentation.

It will be shown how Markush structures can be created automatically from a library of specific chemical structures, their static and dynamic structures visualized (Markush Viewer, Enumeration and Reduction) and searched. Interactive navigation and searching of Thomson Reuters patent content will also be described, including Markush and specific structures and other patent data. Recent developments make the handling of these databases easier, faster and more accurate. The query features and various visualization options of search results all help the casual or more experienced users to understand the vast amount and complex data that are contained in the patent literature.

CINF 58. Rendering the stages of structure elucidation: ACD/Labs Markush representation

Andrey Yerin, *erin@acdlabs.ru, Ian Peirson, Advanced Chemistry Development, Inc. (ACD/Labs), Toronto, Ontario M5C 1T4, Canada*

The Markush structure is a favourite tool of patents, allowing a large number of discreet structures by definition of a single object. While recent cheminformatics tools provide a possibility to operate with Markush structure, it remains almost totally the object for patents, being beyond everyday chemical applications. The workflow of metabolite identification or impurity and degradant profiling now demands the rendering of stages corresponding to the specific degree of knowledge about the chemical structure.

ACD/Labs have developed several tools that allow encoding and visualization by Markush tools multiple variable substitution points, mass and formula modifications. The ability to create and search a database of such structures extends possibilities to retain, extract and leverage knowledge in the organization. The implemented structure representation can be encoded by traditional structure formats and may become a standard tool for the exchange of partially defined structures between various chemical applications.

CINF 59. New developments in Markush structure searching

Donald Walter, *don.walter@thomsonreuters.com, Thomson Reuters, New York, NY 10036, United States*

Patents protect compounds that are specifically disclosed (e.g. butylated hydroxytoluene) and disclosed as families of compounds (e.g. hindered phenols), called Markush structures. Since a single Markush structure can represent an enormous number of specific structures, representing and searching Markush structures represent special challenges. Furthermore, determining which embodiments of the Markush match your query is often a time consuming process. This talk will focus on powerful new ways to search Markush structures and analyze them, all as part of an integrated cheminformatics system to speed discovery and legal evaluation.

CINF 60. Representing and retrieving non specific structures

Keith T Taylor, *keith.taylor@accelrys.com, Accelrys Inc, San ramon, California 94583, United States*

The valence model has allowed the encoding of most of the structures used in lifescience research. In some cases, it is necessary to standardize on a form for a substructure but these do not present significant challenges to understanding. Substances that are used in industrial, and consumer products industries often present significant challenges to simple representations; examples include polymers, and substances derived from nature such as vegetable oils. The growing importance of biotherapeutics brings significant challenges due to their size, and natural and unnatural post translational modifications.

The management of substances and their representation in electronic systems will be reviewed and remaining challenges identified.

CINF 61. Toward a computational pipeline from antibody homology modeling to docking to design

Jeffrey J Gray, *jgray@jhu.edu, Chemical & Biomolecular Engineering, Johns Hopkins University, Baltimore, MD 21218, United States*

This talk will discuss the successes and limitations of homology modeling and protein structure prediction methods in general, what impact they have had on drug discovery, and how they will transform drug design in the future.

CINF 62. MD simulations in pharmaceutical research - examples and lessons learnt

Hannes G Wallnoefer[1,2], Klaus R Liedl[2], Clara Christ[1], Daniel Seeliger[1], ***Thomas Fox***[1]*, thomas.fox@boehringer-ingelheim.com. (1) Computational Chemistry, Lead Identification and Optimization Support, Boehringer Ingelheim Pharma GmbH & Co KG, Biberach, Germany (2) Institute of General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innsbruck, Austria*

Due to methodological advances and increased computer power, molecular dynamics (MD) based techniques have become feasible for application in drug design. MD simulations and Free Energy calculations can be used to investigate conformational variability of target proteins and to estimate ligand-receptor binding free energies.

We will show examples how the results of such simulations can be used in lead identification and optimization of NCEs. In the NBE field, we use MD simulations to study the conformational dynamics of antibodies and nanobodies, in particular to assess the flexibility of CDRs. The trajectories, for instance, reveal the average solvent accessibility of residues that are prone to post-translational modification, e.g. oxidation of Methionine or deamidation of Asparagine. Insights obtained from these analyses influence sequence optimization considerations.

Whether these techniques will become standard tools in drug discovery will depend not only on the accuracy of their results, but also on the availability of standardized and reliable setup and analysis procedures, as often human time rather than computing time turns out to be the bottleneck.

CINF 63. Fragment-to-lead using fragment molecular orbital QM calculations

Richard J Law, *richard.law@evotec.com, Osamu Ichihara, Michael P Mazanetz, Michelle Southey, Mark Whittaker, David Hallett, Evotec, United Kingdom*

Screening of low molecular weight weak binders, “fragments”, and obtaining hits is a well understood process that can be achieved by many different assay techniques. Less well defined is how to proceed once a hit is obtained. Computational chemistry, and the application of multiple techniques, plays a vital role in understanding and ranking the many potential routes for fragment expansion design. Protein-ligand interactions are routinely investigated by docking and the results are often ranked using molecular mechanics (MM) based scoring functions. MM scoring functions have many limitations and as a consequence scoring functions do not adequately predict ligand binding affinity nor do they describe the interactions in sufficient detail as to accurately and illustratively guide medicinal chemistry. To rationalize binding at a quantum level, we demonstrate the application of the fragment molecular orbital (FMO) method as a novel computational methodology and its use in structure-based drug design to guide medicinal chemistry. As well as using FMO to prioritize fragment hits for expansion and rank docking results, it can also be used to perform virtual fragment expansion to help guide subsequent rounds of fragment-to-lead chemistry. The method can also be applied to the scoring of molecular probes, such as water, to assess the nature of unoccupied pockets within proteins in order to further guide compound design.

CINF 64. Docking: This might be heaven or this might be…

Martha S Head, *Martha.S.Head@gsk.com, Computational and Structural Chemistry, GlaxoSmithKline Pharmaceuticals, United States*

GSK has somewhat famously conducted and published an evaluation of docking programs. In that publication, we argued that in general docking programs can find

well-docked poses but cannot reliably score those poses, can for at least some protein targets identify actives in a virtual screen but can not a priori be expected to do so for any new protein targets, and cannot rank order molecules by potency. And yet, I continue to assert that docking is a technology that can have impact in structure-based lead optimization. This talk will discuss that apparent paradox and will discuss the role of expertise in maximizing the utility of less-than-perfect computational tools.

CINF 65. Having a mobile app presence - necessary or nice to have?

Steven M Muskal, *smuskal@eidogen-sertanty.com, Eidogen-Sertanty, Inc., Oceanside, CA 92056, United States*

The growth of mobile computing devices from smart-phones to tablets has been explosive, routinely enabling 24x7x365 connectivity and access into workflows previously constrained to the office. Coupled with cloud computing environments (e.g. Amazon's EC2 and RDS environments), mobile devices and their respective apps have become necessary tools in day-to-day communication and scientific workflow. We will discuss the advantages and disadvantages of native- vs. web-based apps running on mobile devices as well as lessons learned over the last two years after having developed and deployed several mobile apps including iKinase(Pro), iProtein, MobileReagents, Reaction101, Yield101, and others.

CINF 66. Molecular visualization apps in education and research

Jason Vertrees, Blaine Bell, ***Woody Sherman****, woody.sherman@schrodinger.com, Schrodinger, Inc., New York, NY 10036, United States*

The power and ubiquity of mobile devices has increased dramatically in recent years, allowing for high-performance molecular visualization deployable to the masses. Here, we present the current status of molecular visualization on mobile devices, with a focus on two apps. The first, Ball & Stick, is geared toward middle school and high school educational. Common tasks are made very simple and a wizard-like workflow manager guides students through lessons. The second, Mobile PyMOL, is a recent development to port the most powerful and highly used desktop molecular visualization software to mobile devices. Mobile PyMOL is geared toward college-level education and research at all levels. We discuss the advantages of these apps for various tasks and provide a roadmap for each app moving forward. We also discuss the challenges of developing in mobile environments and potential ways to overcome those challenges.

CINF 67. Building a mobile app ecosystem for chemistry collaboration

Alex M. Clark*, aclark.xyz@gmail.com, R&D, Molecular Materials Informatics, Montreal, Quebec H3J2S1, Canada*

The number of mobile apps for chemistry has grown rapidly in the last year. Some of these apps feature cheminformatics capabilities that are powerful and mature enough for integration into a real world workflow. Apps have many mechanisms for sharing data, including interprocess communication, remote procedure calls, mailing attachments, and cloud data storage. This presentation will focus on recent new features and new apps, with a particular emphasis on data sharing, collaboration, and finding ways to modularly substitute mobile apps for traditional cheminformatics tools.

CINF 68. Chemistry made mobile – the expanding world of chemistry in the hand

Antony Williams*, williamsa@rsc.org, Department of Informatics, Royal Society of Chemistry, Wake Forest, NC 27587, United States*

Mobile devices are now mainstream handheld computers providing access to computational power and storage that a decade ago was available only on desktop computers. In terms of chemistry informatics the majority of capabilities that were previously found only on desktop computers is fast migrating to mobile devices making use of the combination of powerful visualization capabilities, fast cloud-based calculations, websites optimized for the mobile platforms, and delivering “apps”. This presentation will provide an overview of how access to chemistry continues to be made increasingly mobile and specifically on how the Royal Society of Chemistry is contributing to this computing environment.

CINF 69. ChemDoodle Mobile: Leveraging mobile apps in chemistry

Kevin J Theisen, *kevin@ichemlabs.com, iChemLabs, LLC., Piscataway, NJ 08854, United States*

The mobile devices market introduces new dynamics for start-up and established companies alike. Benefiting from this market requires significant investment. ChemDoodle Mobile is an HTML5 based mobile app that is popular on both the Apple iTunes Store and the Android Market. The tools involved in the development and deployment of ChemDoodle Mobile are discussed, with a focus on four libraries, ChemDoodle Web Components, jQuery Mobile, Sencha Touch and PhoneGap. After an app is created, it is still a difficult task to make it a successful product. Given the limited scientific market, alternative means of revenue generation are discussed. We will review some data for ChemDoodle Mobile. If done in an affordable manner, mobile apps can provide a company with a significant product to attract and satisfy customers.

CINF 70. Mobile apps for drug discovery

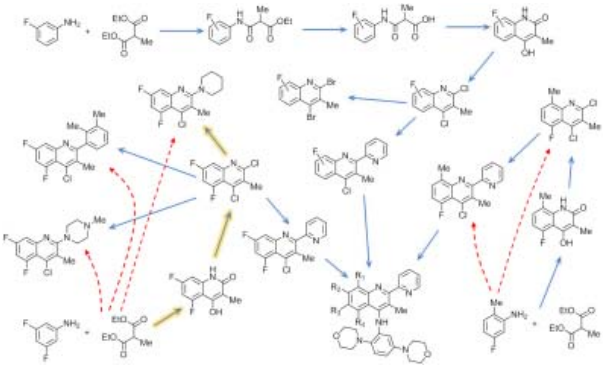
Antony J Williams[1], ***Sean Ekins[2]***, *ekinssean@yahoo.com, Alex Clark[3]. (1) Royal Society of Chemistry, Wake Forest, NC 27587, United States (2) Collaborations in Chemistry, Fuquay Varina, NC 27526, United States (3) Molecular Materials Informatics, Montreal, Quebec H3J 2S1, United States*

Mobile hardware and software technology continues to evolve very rapidly and presents drug discovery scientists with new platforms for accessing data and performing data analysis. Smartphones and tablet computers can now be used to perform many of the operations previously addressed by laptops or desktop computers. Although the smaller screen sizes and requirements for touch screen manipulation can present user interface design challenges, especially with chemistry related applications, these limitations are driving innovative solutions. We will present an introduction to some of the mobile apps we have been involved with most closely. One example is the Green Solvents app which utilizes data created by the ACS Green Chemistry Institute Pharmaceutical roundtable. We will also describe a wiki to capture information about scientific mobile apps (www.scimobileapps.com) and provide our perspective on what mobile platforms may provide the drug discovery scientist in the future as this disruptive technology takes off.

CINF 71. Efficient one-pot preparations of PI3Kd inhibitors using algorithmic network detection

Chris M. Gothard, *cgothard@northwestern.edu, Nosheen A. Gothard, Siowling Soh, Bartosz A. Grzybowski, Department of Chemistry, Northwestern University, Evanston, IL 60208, United States*

One-pot reactions are central to the development of efficient chemical syntheses of complex and biologically important substrates in modern industrial-scale chemistry. Algorithmic detection of 'one-pot' reaction sequences can assist synthetic chemists in developing novel chemical processes and serve as a good starting point in the search for tandem reactions. We have identified novel multistep one-pot routes to medicinally important targets, such as PI3Kd inhibitors. Optimization of the synthetic routes has led to the efficient and high yielding preparation of these compounds.



CINF 72. Statistical analysis of microarray gene expression data from a mouse model of Toxoplasmosis

Shrikant d Pawar, *pawar1550@gmail.com, Claire Rinehart, Cheryl Davis, Department of Biology, Western Kentucky University, Bowling Green, KY 42101, United States*

Toxoplasmosis, caused by the protozoan parasite, Toxoplasma gondii is a major cause of morbidity and mortality in patients with AIDS and an important cause of miscarriage, stillbirth and congenital disease in newborns. Previous studies have provided evidence that dietary supplementation with vitamin E and selenium is

harmful during experimental toxoplasmosis in mice, whereas a diet deficient in vitamin E and selenium results in decreased numbers of tissue cysts in the brain and dramatically reduced brain pathology. The overall goal of the present study was to determine the impact of dietary supplementation with antioxidants on gene expression in the brains of non-infected mice and in mice infected with T. gondii using microarray analysis. RNA was isolated from the brains of C57BL/6 mice, and an Agilent Oligo Whole Mouse Genome Microarray (Agilent Technologies, Inc.) was performed. A total of 48 chips were normalized by calculating Z scores. Differentially expressed genes were identified by performing ANOVA and forming patterns. These differentially expressed genes and their respective fold change ratios were used in Ingenuity Pathway Analysis (IPA) software to analyze the pathways involved with these genes.

CINF 73. Synthetic information challenges for the medicinal chemist

Haiying He, *he_haiying@wuxiapptec.com, Department of Medicinal Chemistry, WuXi AppTec Co., Shanghai, Waigaoqiao Free Trade Zone 200131, China*

As a contract research organization, our chemists may be responsible for many aspects of the drug development process: from the synthesis of lead candidates to the start of clinical trials. Since it can take less time to design a drug-like molecule than to synthesize it in the lab, much of our chemists' time is spent with electronic information products developing a synthetic action plan. With a broad range of problems to solve, our chemists need information sources that cover a broad range of chemistry. This talk will cover the information sources that we have acquired and discuss how they are used to solve the problems that our synthetic chemists face on a daily basis.

CINF 74. SOS 4.0: Advances in text, structure, and reaction searching

M. Fiona Shortt de Hernandez[1], *fiona.shortt@thieme.de, Rolf Hoppe*[1], *Guido F. Herrmann*[1], *Peter Loew*[2]. *(1) Thieme Publishers Stuttgart, Stuttgart, Germany (2) InfoChem GmbH, Munich, Germany*

We have taken a major chemistry reference work in print and designed an interactive Web version from first principles. The online product combines full-text browsing functionality together with InfoChem's modern structure and reaction search capabilities. Science of Synthesis is a unique, structure/reaction searchable, full-text resource that provides the user with expert-evaluated methods and reactions.

Science of Synthesis covers synthetic methodology developed from the early 1800s to-date for the entire field of organic and organometallic chemistry. World-renowned chemists have chosen important molecular transformations for a class of organic compounds and elaborated on their scope and limitations. The logical, structured order of content within Science of Synthesis means that it is simple to gain an overview of the wider context in a particular subject field.

The user can search for specifically defined full-text fields and advanced text searching options are available. It is possible to search the manually prepared named reaction index, which associates transformations with specific named reactions even if they are not mentioned as such in the full text i.e. deep indexing.

A number of filters, called fields are available allowing the hitlist to be refined. The hitlist generated by a text search can be filtered depending upon the location in the full text that the result occurs e.g. in the title, full text, or references. It may also be sorted by relevance - using an algorithm that weights the text results – or by publication date.

The hitlist generated by a structure search may be filtered by the role of the structure e.g. product, catalyst, or solvent. It may also be filtered by the best match criteria (exact or substructure result). The hitlist itself is ordered by relevance, using an internal algorithm that weights exact structure and substructure results. Sophisticated search operators (rating!) allow for efficient and convenient reaction searching.

CINF 75. Automated extraction of reactions from the patent literature

Daniel M Lowe, *daniel_lowe_uk@yahoo.co.uk, Peter Murray-Rust, Robert C Glen, Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Cambridge, Cambridgeshire CB2 1EW, United Kingdom*

We have created a pipeline of recently enhanced open source components for extracting chemical reactions from full text chemical literature. OSCAR4 is used to recognise chemical entities and resolve to structures where appropriate. OPSIN is used to resolve systematic chemical names to structures. Chemical Tagger performs part of speech tagging allowing the interpretation of phrases in chemical syntheses. The final output is a semantic representation (chemical components and their roles, reaction conditions, actions including workup, yield and properties of the product). We then attempt to map all atoms in the product(s) to reactants. If successful we also attempt to calculate the stoichiometry of the reaction. The system has been deployed on over 56,000 USPTO patents published since 2008. The level of recall is useful and most extracted reactions make chemical sense. The pipeline is generally applicable to reactions in chemical literature including journals and theses.

CINF 76. Efficient searching and similarity of unmapped reactions: Applications to pharmaceutical ELN analysis

Roger A Sayle[1], *roger@nextmovesoftware.com, Thierry Kogej*[2], *David Drake*[3]. *(1) NextMove Software, Cambridge, United Kingdom (2) DECS, AstraZeneca, Molndal, Sweden (3) RDI, AstraZeneca, Alderley Park, Cheshire, United Kingdom*

Complex queries of reaction databases typically require every database entry to be atom-to-atom mapped. Such atom mapping provides insights in to the reaction mechanism and conveniently allows determination of the location of the reaction center and which bonds are made or broken during the course of a reaction. Unfortunately, such annotations are not routinely captured in some real world applications, such as Electronic Lab Notebooks (ELNs), precluding ready analysis of the reaction transformations described and reducing the knowledge exploitation from these data sources. Although automated atom mapping algorithms exist, their performance on difficult (e.g. unbalanced) reactions and ambiguous alternatives make methods for processing unmapped reactions desirable.

We describe several algorithms for efficiently searching such "noisy" reaction databases, including a measure of reaction similarity that does not require prior explicit atom mapping. This measure is used to cluster synthesis experiments in an ELN, identifying areas of related or novel chemistry.

CINF 77. Novel tools and techniques in reaction searching: "Name Reaction" and "All-In-One" reaction searches

Valentina Eigner-Pitto, *ve@infochem.de, Hans Kraut, Heinz Saller, Heinz Matuszczyk, Peter Loew, InfoChem GmbH, Munich, Germany*

Beginning in the late 1980s InfoChem started to develop a profound understanding of the storage and handling of chemical structure and reaction information. A major challenge in reaction searching emerged in 1989, when InfoChem acquired an exclusive license to a reaction database (SPRESI) of (initially) 1.8 million reaction records. Since the reaction database management systems (REACCS and ORAC) commercially available at that time could not handle more than 500,000 records, InfoChem was forced to conceive concepts for the selection of meaningful subsets of reaction databases and for performing reaction search algorithms. Based on a high quality reaction center detection module (ICMAP), InfoChem developed a sophisticated reaction type classification application (CLASSIFY) that is still unique to this day. Besides clustering of reaction databases and linking of different reaction databases, this software allows a reaction 'similarity' search (RTS). The world's major vendors of chemical information have adopted this technology to enhance the reaction retrieval capabilities of their products. More recent developments at InfoChem have resulted in a processing tool for the algorithmic detection of name reactions in any reaction database, and the development of smart reaction search operators such as the "All-in-one" reaction search, both of which are based on the company's fundamental software and technology assets. This talk will briefly outline the background and technology of these algorithms, and present some end-

user orientated applications derived from these technologies in detail: "ICNameRXN" and "all-in-one" reaction search (RSA).

CINF 78. Catalyzing information retrieval for organometallic and metal-mediated reactions

Judith N. Currano, *currano@pobox.upenn.edu, Chemistry Library, University of Pennsylvania, Philadelphia, PA 19104-6323, United States*

Organometallic substances can be extremely difficult to find in many databases, given the inconsistencies in data entry. This makes retrieving reactions involving organometallics even more challenging, particularly organic reactions that are catalyzed by a metal-containing species. This paper presents some interesting ways of getting around the limitations of today's tools and finding reactions catalyzed by classes of organometallic substances, as well as introducing methods of going beyond substructure searching to locate reactions in which both starting material and product contain metals.

CINF 79. Finding synthetic chemistry in global literature and patents

Kurt Zielenbach[1], *kzielenbach@cas.org, Jeffrey M Wilson*[2], *Jeffrey D Schloss*[2], *Bryan J Harkleroad*[2]. *(1) Marketing, CAS, Columbus, Ohio 43202, United States (2) Product Development, CAS, Columbus, Ohio 43202, United States*

CAS, the world's authority in chemical information, has seen steady growth in chemists' need for synthetic pathways. With more than 50 million chemical reactions now available in the CAS databases, chemists are asking for useful ways to navigate through large answer sets to find the best reaction. This talk will focus on the features that SciFinder provides to solve the synthetic research problems typically faced by chemists. Among the features that chemists need, and this presentation will illustrate, are relevancy ranking, direct access to experimental procedures, and tools for organizing search results, creating synthetic schemes from individual reactions, and communicating proposed reaction schemes with their peers.

CINF 81. Development of a human diet interactome map

Irene Kouskoumvekaki, *irene@cbs.dtu.dk, Department of Systems Biology, Technical University of Denmark, Kgs Lyngby, Denmark*

Similar to pharmaceuticals, food contains compounds that act as modifiers of biological functions. However, the level of complexity is increased by the simultaneous presence of a variety of components, with diverse chemical structures and numerous biological targets. Nowadays, it is widely recognized that systems chemical biology has the potential to increase our understanding of how small molecules interact with biological systems. A fruitful strategy to approach and explore the field of nutritional research is, therefore, to borrow methods that are well established in pharmaceutical research.

We have recently initiated a project at CBS/DTU, where we used text mining to construct a unique database with state-of-the-art information concerning food and its molecular components. During the talk, I will present the steps we followed for developing a database that consists of 1,500 food types and 35,000 small molecules and I will highlight applications through linking the nutritional chemical space with the human proteome and disease.

CINF 82. Studying the chemical interactome space between the human host and the genetically defined metabotypes of our gut

Gianni Panagiotou, *gpa@bio.dtu.dk, Department of Systems Biology, Technical University of Denmark, Lyngby, Europe 2800, Denmark*

The bacteria that colonize the gastrointestinal tracts of mammals represent a highly selected metagenome that has a profound influence on human physiology by shaping the host's metabolic and immune network activity. Despite the recent advances on the biological principles that underlie microbial symbiosis in the gut of mammals, mechanistic understanding of the contributions of the gut microbiome and the links of variations in the metabotypes to the host health are obscure. Here we mapped the entire biosynthetic potential of the gut microbiome based on metagenomics sequencing data that derived from fecal samples of 267

European individuals. These metabolic signatures were used to study the signaling cascade triggered in humans through chemical (bacterial metabolites)-protein interaction networks and provide evidence of how specific changes in the gut microbial community/metabolism might affect or counteract the development of IBDs, obesity and related diseases.

CINF 83. Comparative study of small molecule inhibition of Mycobacterium tuberculosis and Francisella tularensis

*Sandra V Bennun, **Elebeoba E May**, eemay@sandia.gov, Nanobiology Department, Sandia National Laboratories, Albuquerque, NM 87185, United States*

The ability of intracellular pathogens to persist depends on their capacity to biochemically adapt to changes in the host's intracellular environment modulated by immune response mechanisms. Targeted introduction of small molecules can interfere and reduce the pathogen's metabolic capacity. We use computational systems chemical biology (SCB) methods to comparatively investigate the effects of inhibitory molecules on the tricarboxylic acid cycle (TCA) of M. tuberculosis and F. tularensis (Ft), two pathogens that infect host macrophages. The reconstructed metabolic pathway for Ft is missing reactions present in Mtb, a slower growing pathogen. Differences in metabolic capacity may impact several factors including intracellular localization, response to oxidative stress, and potentially response to small molecule inhibitors. We evaluate inhibition of isocitrate lyase (ICL) during aerobic and oxidative stress conditions, and simulate the metabolic consequence of ICL disruption in both systems. We will discuss observations regarding Mtb and Ft's response given the variations in their representative metabolic models.

CINF 84. Identifying druggable targets by mining open chemical biology data

***Yanli Wang**, ywang@ncbi.nlm.nih.gov, National Institutes of Health, National Center for Biotechnology Information, Room 5S506, Building 38A, 8600 Rockville Pike, Bethesda, MD 20894, United States*

Molecular target identification plays a central role for drug discovery and small molecule probe development. A in-silicon approach is presented to suggest molecular target and chemical-target association network based on comparing and combining multiple bioactivity endpoints from chemical biology experiments and target annotations available in open-access databases. A clustering analysis based on bioactivity profile similarity reveals strong correlations between chemical structures, across-panel biological responses, and chemical-target associations, suggesting novel compound candidates with desired pharmacological properties may be identified by bioactivity profile comparison. A computational approach was further developed based on the BioActivity Profile Similarity Search (BASS) to mutually identify compound-target associations among neighbor compounds with similar bioactivity spectrum. An overall success rate of 45% was obtained for the predicted compound-target associations. Analysis shows that BASS not only could identify structurally similar compounds but also could suggest novel chemical scaffolds for the aimed targets.

CINF 85. Exploiting semantic networks of public data for systems chemical biology

***David J Wild**, djwild@indiana.edu, School of Informatics and Computing, Indiana University, Bloomington, IN 47405, United States*

We have developed a systems chemical biology data resource called Chem2Bio2RDF (www.chem2bio2rdf.org) that integrates publicly available datasets pertaining to chemical compounds, drugs, drug side effects, targets, genes, pathways, diseases and scholarly publications. The dataset is semantically annotated using ontologies including a new chemogenomic ontology called Chem2Bio2OWL. We have developed a variety of graph-based and other network algorithms to look for chemogenomic and other associations in this data, including association search tools, integration of the literature with a novel BioLDA topic modeling method, a method called SLAP for missing link prediction, and rule-based inference of new relationships. In this talk, I will describe the Chem2Bio2RDF resource and give an overview of the algorithms and how they are being applied in drug discovery problems.

CINF 86. Enhancing chemoinformatics with pathway analysis tools: An integrated approach to drug discovery

Tatiana Khasanova, *tatiana.khasanova@thomsonreuters.com, Eugene Myshkin, Sirimon O'Charoen, Yuri Nikolsky,* ***Svetlana Bureeva****, svetlana.bureeva@thomsonreuters.com, IP and Science, Thomson Reuters, Carlsbad, California 92008, United States*

Modern approach to drug discovery encompasses multifauceted, integrated consideration of both chemical and biological processes. System biology methods provide tools necessary to analyze complicated relationships between molecular entities in both normal and disease states. This approach enables understanding fine details of a drug's mode of action, which keeps shifting from consideration of a target to an analysis of a whole affected pathway. Integrating system biology approach with chemoinformatics methods is a key to a successful application of this new approach to drug discovery process. Over the last four years, GeneGo, a Thomson Reuters company, has created a unique systems pharmacology suite (MetaDrug™) that leverages the power of systems biology (target CVs, OMICs data analysis), flexibility of classical chemical tools (QSAR, metabolic rules) and reliability of a comprehensive manually curated "knowledge base" for analysis of biological effects of new and known small molecules. The multi-step analysis workflow proceeds from compound targets to affected pathways, from pathways to associated diseases and toxicities. The same approach enables researchers to go backwards and identify targets for compounds with known phenotype or effect on pathway level. In our presentation, we will discuss several use cases that illustrate how this approach can be used for drug repositioning, solving mechanism of action and discovery of synergistic drug combinations.

CINF 87. Updated tools, techniques and data sources for effective reaction retrieval in support of synthetic methodology and drug discovery

Matthew A Kellett, *matthew.kellett@thomsonreuters.com, Life Sciences Editorial Team, Thomson Reuters, Philadelphia, PA 19130, United States*

Chemical reaction databases are a key element in support of synthetic chemistry research and multiple stages of the drug discovery and development process. While substructure searching provides useful results, newer tools in conjunction with the integration of enhanced indexing allow for more effective analysis of results. One example is the inclusion of reaction mapping and synthesis planning capabilities which enable the incorporation of these results directly into the workflow process. The use of separate databases that include unique features such as: detailed condition information, keywords, biological activity indexing, reference citations, multiple examples of new methodologies, and/or specific drug and natural product syntheses improve the overall retrieval of relevant chemical literature and synthetic methods.

CINF 88. Extremely rapid searching of in-house reaction databases: Turning ELN data into a searchable library

Philip J Skinner, *philip.skinner@perkinelmer.com, Scott Flicker, Joshua Wakefield, Sean Greenhow, Megean Schoenberg, Kate Blanchard, Phil McHale, Sandra Sessoms, Robin Smith, PerkinElmer Informatics, United States*

Electronic Laboratory Notebooks (ELNs) are increasing used to capture experiments, a majority of early deployed systems to capture chemical reactions. However, ELNs optimized for data capture are less so for data searching, that organizations with the largest databases and the most to benefit from their in-house ELN reaction libraries are those least able to access them. To facilitate reaction searching we developed a novel fast algorithm applied to data extracted from an ELN. We use fragment-based fingerprinting to determine transformation fingerprints. Substructure search algorithms, used to determine fragments within a molecule, generate the transformation fingerprints used in searches as strings. Searches compare target reactions with pre-calculated transformations to find similar reactions. Lookups for transformation strings are performed in a cascading similarity order. Results are bucket-sorted, displaying closest matches first. This results in extremely rapid return of relevant search results, allowing researchers to quickly mine the ELN for information hitherto unavailable.

CINF 89. Understanding search results: From a single reaction to scope and limitations of the reaction route in the ChemInform Reaction Library (CIRX)

Yana Steudel, *steudel@fiz-chemie.de, Ulrike Schramke, ChemInform Databank, FIZ CHEMIE Berlin, Berlin, Germany*

The main goal of a synthetic chemist is to design new compounds with specific properties. Most databases deliver synthetic procedures for the preparation of known compounds. ChemInform with its unique full reaction schemes can be used to plan the preparation of yet unknown derivatives. The reaction scheme provides the full information on the scope and limitations of a given reaction with one glimpse and imparts a better understanding of the general synthetic value of a reaction to the chemist. The presentation will demonstrate the possibilities offered by ChemInform to implement reaction search results in the current research projects.

CINF 90. Dealing with chemical reality: Handling reactions plus associated data and branching reaction schemes

Jonathan S Brecher, Harold Helson, ***Phil J McHale****, phil.mchale@perkinelmer.com, PerkinElmer Informatics, Cambridge, MA 02140, United States*

Experiments in chemical ELNs include the synthetic route, text, and other compounds or reactions involved (e.g. a side reaction or by-product). Intelligent retrievability requires storing this information as a unified entity, so the structures, information and context can be found. Many systems store these pieces of information separately, thus losing their inter-relationships. Another complication is branched synthetic routes. Most systems can handle single step reactions and linear multi-step reactions: but branched reactions are problematic.

We describe methods for unified indexing of reactions and associated data and structures and for handling branched reaction schemes in the following categories:

• Divergent: reaction product is a reactant in two or more different reactions;

• Convergent: compound is the product of two or more different reactions;

• Cyclic: reaction scheme starts and ends with the same compound.

We have implemented these methods in a chemical cartridge and exposed them through a chemistry ELN.

CINF 91. Reaction searching for compounds which do not even exist yet

Carsten Detering, *detering@biosolveit.de, Christian Lemmen, Marcus Gastreich, BioSolveIT GmbH, St. Augustin, NRW 53757, Germany*

We present an industry-proven method which – extremely rapidly – searches through textbook or corporate reaction recipes and assembles novel molecules similar to any real or virtual query/starting molecule.

This is remarkable because due to the combinatorial nature of reactions the search space is gigantic and the needed time for computation is extremely little: Only one reaction (for example, 10 acids A1-A10 and 10 different amines N1-N10) can formally form 100 products (here: amides) already; the search space amounts up to 1013virtual molecules which are searched in a few minutes only.

The technology thus generates new intellectual property (IP), alongside with the recorded reaction information – proposing how to have the associated synthesis.

The paper will explain the basics of the similarity concept (FTrees-FS [1]) and the search technology and highlight dozens of successful examples including many from the pharmaceutical industry.

[1] www.biosolveit.de/FTrees

CINF 92. Helping you make the right choices for your next synthetic route!

Juergen Swienty-Busch[1]*, J.Swienty-Busch@elsevier.com, David A. Evans*[2]*. (1) Elsevier Information Systems, Frankfurt, Germany (2) Elsevier Properties SA, Neuchatel, NE 2000, Switzerland*

With a steadily increasing amount of data chemists today face the challenge of finding the best possible

path to a desired molecule and to optimize the synthesis. While the size of a reaction database is important to find enough examples for making a good choice it is equally important that this information can be clustered, ranked, categorized and visualized intuitively to get to the desired result quickly. This talk will discuss a few examples showing how Reaxys addresses these challenges.

CINF 93. Advanced reaction searching: A comprehensive treatment of stereoselectivity in reactions

Peter Johnson[1], *p.johnson@leeds.ac.uk, Anthony P Cook*[1], *James Law*[2], *Aniko Simon*[2], *Orr Ravitz*[2]. *(1) Department of Chemistry, University of Leeds, Leeds, W Yorks LS17 8JQ, United Kingdom (2) Simbiosys Inc, Toronto, Ontario M9W 6V1, Canada*

The ARChem system for automated retrosynthetic analysis makes use of rules describing retrosynthetic transformations which are generated by automated mining of reaction databases. The application of these rules in a controlled fashion minimises the potential combinatorial explosion of possible routes to a target structure. In the past the system has ignored any stereochemical designators in the target molecule and the retrosynthetic rules were devoid of any stereochemical information, a major omission, given the huge advances in enantioselective synthesis in the past few decades. Recent work addresses this problem through:

a) comprehensive perception of stereochemical information in targets, available starting materials, and reactants and products of literature examples

b) inclusion of the stereochemical course of each reaction into an abstracted rule so that retrosynthetic searches provide stereochemically sound suggestions

Details of the work will be given as will examples of the retrosynthetic analysis of stereochemically complex targets.

CINF 94. Algorithmic network detection of reaction sequences: From novel "one-pot" reactions to unanticipated synthetic routes to chemical weapons

Chris M. Gothard, *cgothard@northwestern.edu, Nosheen A. Gothard, Siowling Soh, Bartosz A. Grzybowski, Department of Chemistry, Northwestern University, Evanston, IL 60208, United States*

Using the known network of chemical reactions, computational searches across the network have enabled us to discover (1) novel tandem reactions (e.g. 'one-pot' syntheses) and (2) unanticipated synthetic routes to dangerous substances (e.g. Chemical Weapons). One-pot reactions are central to the development of efficient chemical syntheses of complex and biologically important substrates in modern industrial-scale chemistry. By screening reactions for both compatibility among functional groups and reaction conditions, we have identified multistep one-pot synthetic routes to medicinally important PI3Kd inhibitors. Optimization of these reaction sequences has led to efficient and high yielding preparations of PI3Kd inhibitors. In addition to medicinal targets, we have also employed network-based detection to uncover unanticipated routes to dangerous substances. Although common precursors to chemicals weapons are well regulated, there exists' alternative routes that utilize only unregulated substances. Detection of these unanticipated routes is an important strategy in limiting access to chemical weapons.

CINF 95. Framework for systematic prediction of pharmacologically relevant targets of small molecules

Emmanuel R Yera, *eyera@ucsf.edu, Ann E Cleves, Ajay N Jain, Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA 94158, United States*

Drug discovery requires the design of molecules that modulate the activity of specific biological targets with minimal effects on other targets. To systematically predict pharmacologically relevant relationships between small molecules and protein targets, including undesirable off-targets, we have developed a probabilistic framework based on molecular similarity. Small molecules may be quantitatively compared based

on 2D or 3D characteristics, the latter being directly related to binding. Given a new molecule along with a set of molecules sharing some biological effect, a single score based on the comparison to the known set is produced, reflecting either 2D similarity, 3D similarity or their combination. The results of a systematic application to a large set of drugs will be presented along with a critical analysis examining what can be learned about drug pharmacology based on different molecular similarity methods. The potential for association of phenotypic effects with specific biological targets will also be discussed.

CINF 96. Designing ligands against multi-target profiles

***Andrew L Hopkins**, a.hopkins@dundee.ac.uk, College of Life Sciences, University of Dundee, Dundee, Scotland DD1 5EH, United Kingdom*

The prospect of multi-target drug design has been recently advanced by the development of computational polypharmacology prediction method. We describe a new approach for the automated design of ligands against profiles of multiple drug targets. The method is demonstrated by the evolution of an approved enzyme inhibitor drug into, brain penetrable ligands with specific polypharmacology or exquisite selectivity profiles for G-protein couple receptors. Overall, 800 ligand-target predictions, of prospectively designed ligands, were tested experimentally, of which 75% were confirmed correct. The method demonstrates automated design can be a useful method to solve the complexity of optimising multiple structure-activity relationships. The validated method shows promise to be a potential source of drug leads where multi-target profiles are required to achieve either selectivity over other drug targets or a desired polypharmacology.

CINF 97. CARLSBAD (Confederated Annotated Research Libraries for Small molecule BioActivity Data): A database and its platform

***Gergely Zahoranszky-Kohalmi**, GZahoransky-Kohalmi@salud.unm.edu, Jeremy J Yang, Cristian G Bologa, Stephen L Mathias, Oleg Ursu, Jarrett Hines-Kay, Tudor I Oprea, Department of Biochemistry and Molecular Biology, University of New Mexico, Albuquerque, New Mexico 87131, United States*

Identifying key interactions and the determinant structural patterns between small molecules and their biological targets remains a major challenge in the field of drug discovery and drug repurposing. Here we introduce CARLSBAD, a platform and database designed to guide the network-based discovery of such complex patterns with the help of maximal overlapping substructures and hierarchical scaffolds. Bioactivity data are delivered on a basis of consensus of leading bioactivity databases (IUPHAR-DB, PDSP, WOMBAT, PubChem, ChEMBL) enhanced by systematic confidence annotations. A web application and a Cytoscape plugin provide a convenient interface for exploring and analyzing interactions between millions of molecules and thousands of biological targets. Applications using the Cytoscape plugin based on CARLSBAD data will be presented.

CINF 98. Drug combinations to reduce adverse drug reactions and improve intrapatient differences in response

***John P Overington**, jpo@ebi.ac.uk, Computational Chemical Biology, EMBL-EBI, Hinxton, Cambs CB10 1SD, United Kingdom*

ADRs are a major cause of hospitalization and a large cost burden to healthcare systems. A further complicating factor is that many ADRs are rare, but serious, events, dependent on factors such as metabolic state, drug-drug interactions (DDIs) and pharmacogenetic variation within the patient population. We have applied a simple systems-based theoretical model of poly-/network-pharmacology and PK/PD models, combined with a data-mining approach, to the generation of a new strategy to identify drug combinations that have improved safety profiles. This drug combination approach is also theoretically predicted to reduce intra-patient pharmacogenetic-based differences in drug response/efficacy, as an additional emergent property. The theoretical basis for this approach will be outlined, along with representative examples of specific drug combinations with synergistic safety features. Finally, we outline our plans to test this hypothesis across a variety of cardiovascular diseases.

CINF 99. Integrating targets, drugs and clinical outcomes into systems medicine

Tudor I Oprea, *toprea@salud.unm.edu, Department of Biochemistry and Molecular Biology, UNM School of Medicine, Albuquerque, NM 87131, United States*

For the therapeutic management of chronic diseases, the systems medicine approach shows great promise. Starting from such therapeutic indications that are relevant for complex, multifactorial chronic diseases (e.g., diabetes, asthma, cancer), we examine the relationship between drugs, targets, current indications and counter-indications, as well as serious adverse reactions. The analysis is aided by the use of controlled vocabularies such as those available in MedDRA and ICD-10, and benefits from manual curation. Practical aspects such as single vs. multiple (chronic) dosing, as well as temporal patterns (evolution of disease states) will be considered. Such aspects are likely to pave the way towards a systems medicine approach and improve our understanding of the therapeutic control of chronic diseases.

CINF 100. IUPAC InChI project: A status report

Stephen Heller, *steve@hellers.com, IUPAC, Silver Spring, MD 20902, United States*

The current status and use the InChI algorithm will be presented. Future use and extension of the algorithm will be described.

CINF 101. Great promise of navigating the internet using InChIs

Antony J Williams, *williamsa@rsc.org, Department of Informatics, Royal Society of Chemistry, Wake Forest, NC 27587, United States*

The InChI, the International Chemical Identifier, has been the basis of both indexing and deduplication of the ChemSpider database since the inception of the platform. When the InChI was adopted we envisaged a future whereby the identifier would proliferate across journals, databases and the internet in general providing us a basis for "structure searching the internet". This presentation will provide an overview of how the InChI has facilitated the integration of ChemSpider to chemistry on the internet, some of the surprising findings that have resulted from this work and extrapolate the influence of InChIs into the future for a chemically enabled web.

CINF 102. InChI names and keys: Do they add value to commercial software and databases

Keith T Taylor, *keith.taylor@accelrys.com, Carmen Nitsche, Accelrys Inc, San Ramon, California 94583, United States*

InChI names and keys provide a general identifier for a chemical structure that can be used to correlate chemical information in public and private data repositories. They are exposed, generated, and processed in many of Accelrys' software products and they are recorded in Accelrys' commercial databases.

Although often thought of as a unique identifier, in many cases this is not true. How their uniqueness (or lack of uniqueness) creates opportunities and problems will be discussed.

CINF 103. Use of InChI in wikis

Martin A Walker[1], *walkerma@potsdam.edu, Aileen Day*[2]*. (1) Department of Chemistry, State University of New York at Potsdam, Potsdam, New York 13676, United States (2) Royal Society of Chemistry, Cambridge, United Kingdom*

The International Chemical Identifier (InChI) provides a useful shortcut representation of chemical structures. On Wikipedia the InChI (and then InChIKey) was originally provided as a service to chemists, but now it can draw traffic from web searches (for example, from ChemSketch). It also provides a reference point, intersecting article with structure when performing structure validations.

In ShareChemistry, the new RSC education wiki, new extensions have been written to make the wiki "structure-friendly". In conjunction with the Ketcher drawing tool, the InChI is used to provide full structure searching within the wiki. In addition, the InChI is used in "predict the product" quiz questions or similar, where the student has to draw the product structure; this also utilizes the InChI structure to identify "nearly correct" answers and provide appropriate feedback. By allowing quizzes to go beyond simple multiple choice questions, the use of InChI can greatly enhance student learning.

CINF 104. InChIKey collision safety: Experimental estimation for algorithmically generated structure libraries

Andrey Yerin, *erin@acdlabs.ru, Kirill Blinov, Advanced Chemistry Development, Inc. (ACD/Labs), Toronto, Ontario M5C 1T4, Canada*

The InChIKey is a hash-based fixed length representation of the IUPAC International Chemical Identifier (InChI) and has growing importance in chemical informatics as a basis for searching and indexing chemical structures. Since it is composed of 22 variable letters the InChIKey theoretically has an extremely low collision rate but certainly cannot uniquely encode the whole of chemical space. While InChIKey collisions have already been reported experimental tests of collision rates for extremely large databases have not yet been performed.

A protocol allowing for the generation of InChIKeys for algorithmically created virtual structure databases has been launched at ACD/Labs. We will report on our work analyzing large generated data sets and provide reliable statistical estimations of InChIKey collisions.

CINF 105. InChI here, InChI there, InChIs everywhere

Juergen Swienty-Busch[1], *J.Swienty-Busch@elsevier.com, David A. Evans*[2]*. (1) Elsevier Information Systems, Frankfurt, Germany (2) Elsevier Properties SA, Neuchatel, NE 2000, Switzerland*

We will describe how Elsevier is using InChIs everyday in SciVerse ScienceDirect and Reaxys.

CINF 106. New searching paradigms in drug discovery enabled by semantic integration of public data

David J Wild[1], *djwild@indiana.edu, Erik A Stolterman*[1]*, Michael S Lajiness*[2]*. (1) School of Informatics and Computing, Indiana University, Bloomington, IN 47405, United States (2) Eli Lilly, Indianapolis, IN 46285, United States*

The recent explosion of publicly available sources of data relating to drug discovery, along with electronic access to journal articles offers many new possibilities for knowledge discovery, but navigating the millions of data points relating to compounds, drugs, targets, genes, diseases and publications spread over hundreds of online datasets can be overwhelming. We will present work done at Indiana to semantically integrate many of these public data sources to provide a framework for identifying networks of information spread across datasets and publications of interest to a particular researcher, research question or hypothesis (for example, what can we find in public data that pertains to the relationship of Ibuprofen to Parkinson Disease?). Examples will be given of how prototype tools developed at Indiana and Eli Lilly can be used to effect these kinds of search, and what this might reveal about the future of search tools and paradigms for drug discovery. We will also discuss some of the barriers to performing these kinds of advanced search effectively, including data quality issues, properly accessing data in the text of journal articles, and extracting important relationships from "background noise".

CINF 107. Collaborative computational technologies for biomedical research: An enabler of more open drug discovery

Sean Ekins[1], *ekinssean@yahoo.com, Antony J Williams*[2]*. (1) Collaborations in Chemistry, Fuquay Varina, NC 27526, United States (2) Royal Society of Chemistry, Wake Forest, NC 27587, United States*

The current paradigm in the pharmaceutical industry is that products can only be created and developed by massive collaborative teams. Each company has to build their own costly R&D platforms and IT infrastructure. Other research industries realized decades ago that they had to share data and methods because of cost. The pharmaceutical industry has been slow to realize this. Expanding beyond our recent book (Collaborative Computational Technologies for Biomedical Research) in which a growing number of technologies, consortia, precompetitive initiatives and complex collaboration networks are described, we suggest a more open drug discovery is being enabled by collaborative computational technologies. Academia however, is not training the next generation of scientists to practice open science or even collaborate, this represents challenges and opportunities. We will describe our observations and make recommendations that impact everyone from technology developers to granting agencies. This may enable future discoveries to be made outside traditional institutions.

CINF 108. Enabling biomolecular simulation data sharing across institutions using a Grid architecture

Julien C Thibault[1], *julien.thibault@utah.edu, Thomas E Cheatham*[2], *Julio C Facelli*[1,3]. *(1) Department of Biomedical Informatics, University of Utah, Salt Lake City, Utah 84112, United States (2) Department of Medicinal Chemistry, University of Utah, Salt Lake City, Utah 84112, United States (3) Center for High-Performance Computing, University of Utah, Salt Lake City, Utah 84112, United States*

Our work presents a Grid-based infrastructure that can enable biomolecular simulation data sharing across research labs. Raw and derived data include molecular dynamics atom trajectories and energies as a function of time. This would facilitate the development of new models (e.g. coarse-grain representations, force fields) and their assessment. Each data node is managed by the iBIOMES system, which creates a virtual data warehouse at the researcher's site by managing distributed file servers. The current implementation of iBIOMES offers a command-line interface that can be used to register simulation files into the system. An interactive web client is used to present simulation data to the researchers. External systems can query iBIOMES nodes through the Java API or the RESTful web service interface. Data queries are metadata-driven and supported by the iRODS framework. Using the caGrid toolkit, we will connect iBIOMES nodes together and enable federated queries across institutions.

CINF 109. Representing chemical information by URLs: The chemical identifier resolver as a general chemoinformatics tool

Marc C Nicklaus, *mn1@helix.nih.gov, Markus Sitzmann.NCI-Frederick, National Cancer Institute, NIH, Frederick, MD 21702, United States*

Traditional chemistry databases are siloed off as far as the data are concerned, and require, as an additional component, a specific user interface to access these data. Furthermore, this access is often an either-or proposition in that it either allows usage by a human on a compound-by-compound basis, or download of datasets in bulk from computer to computer. The Chemical Identifier Resolver (CIR) of the NCI/CADD Group, in contrast, allows direct access into its entire data store of currently 120 million structure records as well as to its chemical transformation capabilities. Access is by straightforward URLs that can be put together as easily by a human as by other web services, programming packages or scripting languages. We will present how CIR can be used to represent chemical structure data from, or through, InChI[Keys], chemical names, IUPAC names, SMILES, tautomeric forms, many different chemical file formats as well as calculated (physicochemical) properties.

CINF 110. Publication@Source: The Lab as a database

Jeremy G Frey[1], *j.g.frey@soton.ac.uk, Mark I Borkum*[1], *Simon J Coles*[1], *Tim Parkinson*[2]. *(1) Department of Chemistry, University of Southampton, Southampton, Hants SO17 1BJ, United Kingdom (2) Department of Electronics and Computer Science, University of Southampton, Southampton, Hants SO17 1BJ, United Kingdom*

All data exists within a context: the human or machine that generated the data; the processes that were enacted; and the environments in which it occurred. The capture and dissemination of the context of data is of vital importance affording the data new potential for creating fresh value. In contrast to the top-down software development approaches, where the structure and semantics of the data are informed by the implementation of the software applications that manage the data, the Smart Research Frameworks (SRF) project software suite of data-agnostic software applications and frameworks when used in conjunction with a novel scientific methodology, which focuses on the explicit description of scientific intent and action, facilitates the automated and semi-automated capture and dissemination of richly structure data in context. By providing an appropriate query mechanism (SPARQL endpoint) the laboratory notebooks collectively effectively form a super database of raw and analysed data.

CINF 111. InChIs as building blocks for complex substance identifiers

Yulia Borodina, *yulia.borodina@fda.hhs.gov, Lawrence Callahan, Frank Switzer, Office of the Commissioner, FDA, Silver Spring, MD 20993, United States*

The FDA Substance Registration System (SRS) registers and assigns UNique Ingredient Identifiers (UNII) to substances which may be simple chemicals or complex substances such as chemically modified biopolymers or synthetic polymers. These complex substances cannot be identified by a single InChI string however InChI strings may be used for identification of the structural elements these complex substances are comprised of. These structural elements include monomers, modifying agents and fragments. We will discuss the possibility of using InChI strings as building blocks for creating the more complex identifiers needed for registration of these complex substances.

CINF 112. Accessing NCI/CADD web resources by InChI

Markus Sitzmann, *sitzmann@helix.nih.gov, Marc C. Nicklaus, Chemical Biology Laboratory, Center for Cancer Research, National Cancer Institute/Frederick, NIH, DHHS, Frederick, MD 21702, United States*

IUPAC's International Chemical Identifiers (InChIs/InChIKeys) are a vital tool to enable web-based linking between different sources of chemical content. We present an overview about how InChI/InChIKeys can be used to access our NCI/CADD web services which are part of our Chemical Identifier Resolver (CIR). The service is publically available at http://cactus.nci.nih.gov/chemical/structure and provides a simple and programmatic URL API to access a broad range of chemical structure information and chemical structure representation formats linked to a specific InChI/InChIKey.

At the time of writing, the database utilized by CIR indexes approx. 120 million structure records which have been aggregated from various small-molecule databases and which, after careful structure normalization including calculation of our NCI/CADD Chemical Structure Identifiers, comprise a set of approx. 80 million unique chemical structures. For the entire set of normalized structures, Standard InChI/InChIKeys and those using various non-standard sets of configuration flags of the InChI algorithm have been calculated. On basis of these different identifier sets, we will discuss the differences in structure identification between the InChI/InChIKey and NCI/CADD identifier sets that we have observed in the database used by CIR, and what these discrepancies can tell us about definition and design, scope, limitations and problems of chemical structure identifiers.

CINF 113. InChI vs IUPAC nomenclature: Aspects to be aware of when using Standard InChI

Daniel M Lowe, *daniel_lowe_uk@yahoo.co.uk, Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Cambridge, Cambridgeshire CB2 1EW, United Kingdom*

Features of IUPAC nomenclature that cannot be represented in Standard InChI will be examined to draw caution to cases where the use of standard InChI (and even in some cases non-standard InChI) may result in a loss of information. These areas include the representation of tautomers and mixtures of stereoisomers.

CINF 114. InChI adoption at the Royal Society of Chemistry

Richard Kidd, *kiddr@rsc.org, Royal Society of Chemistry, Cambridge, Cambs. CB4 0WF, United Kingdom*

This flash talk will cover how the InChI standard has been used within the RSC and our contribution to the development of the standard.

CINF 115. Registration system of mcule: InChI is the key

Ferenc Szalai, *rkiss@mcule.com, Robert Kiss, Mark Sandor.mcule.com, Budapest, Hungary*

Mcule provides virtual screening services on the web to help identifying novel drug candidates by screening different databases. For these databases, it is essential to have a robust molecule registration system not depending on different drawing conventions, tautomeric states, etc. It is critical to assure that the same

compounds get the same IDs and, most importantly, different compounds never get the same ID. To the best of our knowledge, InChI provides the best solution for this problem. In this presentation we would like to summarize how InChI is implemented into the mcule registration system and how it is used effectively with our vendor database and open registration services.

CINF 116. "UniChem": A prototype unified chemical structure cross-referencing and identifier tracking system

Jon Chambers, *jon.chambers@ebi.ac.uk, Anna Gaulton, Anne Hersey, Mark Davies, John P Overington, Computational Chemical Biology Group, European Bioinformatics Institute, Cambridge, CAMBS CB10 1SD, United Kingdom*

ChEMBL is an online database of bioactivity data for a large number of organic, drug-like compounds. These data are abstracted from the primary published literature, and are utilized to address a wide range of drug-discovery and chemical biology problems. Chemical structures within ChEMBL are standardized using a series of business rules, and replicated compound structures from different publications and sources are normalized on the basis of identical standard InChIs. Cross-referencing of these structures with identical structures in other chemistry databases is useful for the purposes of comparison and integration (for example, the creation of web links between database interfaces). Unfortunately, however, the process of creating and maintaining these cross-references often involves script-based, semi-manual steps. To assist in the automation of this process within our own institution, we have developed a prototype system (called 'UniChem') for archiving and cross-referencing of chemical structures and their identifiers from multifarious sources. The design of the system is modeled on that of the UniParc database, which serves a similar cross-referencing and archival function for protein sequences. UniChem uses the standard InChI as a means of normalizing between different sources, and in addition to providing up to date cross-linking information, is also able to track changes in identifier assignments over time.

CINF 117. Update on project to introduce InChI to researchers in the Department of Chemistry at Louisiana State University

William W Armstrong, *notwwa@lsu.edu, Karen L Salazar.LSU Libraries, Louisiana State University, Baton Rouge, Louisiana 70803, United States*

In order to be effective, InChI must be understood and used by researchers, educators, and publishers on a large scale. At the 2011 ACS meeting in Denver, the authors provided details of their plans to create a teaching methodology successful at enabling researchers at Louisiana State University to understand InChI and the need it fills as a universal, non-proprietary method of identifying chemical compounds that can take full advantage of new web-based communication and search technologies. Equally important, this teaching methodology will be designed to help researchers bridge the gap from understanding to application and actually begin to integrate InChI into their regular workflows. In this talk, the authors will review their progress and provide an analysis and evaluation of the results to date, along with the next steps. The final product can be employed at similar institutions worldwide.

CINF 118. Past, present and future of the InChI Trust

Jason N Wilde, *j.wilde@nature.com, Nature Publishing Group, London, United Kingdom*

The InChI Trust was established in 2009 with the aim to develop and support the non-proprietary IUPAC InChI standard and promote its use to the scientific community. Over the last 3 years the Trust has approved new versions of the InChI algorithm and established working parties to investigate the technical solutions required to tackle some of the more complex problems related to chemical structure representation i.e. Markush, Polymers and Mixtures etc… This talk summarises the impact/success of this work and sets out a road map for future development of the InChI standard.

CINF 119. InChiKey insertion technique for compound-specific and any-compound proximity search

Stephen K. Boyer, *skboyer@gmail.com, Thomas Griffin, Alfredo Alba, Su Yan, Ying Chen, Scott Spangler, Eric Louie, Jeff Kreulen, Almaden Research Center, IBM Research, San Jose, California 95120, United States*

The combined technologies of text analytics and name-to-structure conversions for reading and processing molecular structures provide researchers the ability to build large databases of structures and derive important relationships previously inaccessible, a capability important to discovery and innovation. Our previous work took this approach to produce SMILES strings that represented chemical structures used as input for subsequent applications, rendering the scientific and patent literature searchable by structure/substructure programs. We now report the additional ability to detect, normalize, and replace chemical names in documents with InChiKeys and then index the combined text and embedded InChi's using SOLR, a Lucene-based full text-indexing engine.

The resulting index supports Boolean combinations of chemical compounds and regular text words and phrases. It also supports proximity searching. The net result is that we can now perform searches for exact chemical structures or even unspecified chemical structures within a specified context.

CINF 120. Exploring almost every InChI of nature.com

Laura J Croft, *l.croft@nature.com, Nature Publishing Group, London, United Kingdom*

With the launch of Nature Chemistry in April 2009 the number of chemical structures published on nature.com has increased rapidly. Since then we have made efforts to increase the discoverability for our readers of information relating to chemical structures both within our own article pages and elsewhere on the web. This talk will showcase some of the trials and tribulations of using InChIs to represent chemical structures on nature.com.

CINF 121. Finding the future: Using research analytical tools with journal article databases and social media data to identify high-impact research leaders and programs

Elizabeth A. Brown, *ebrown@binghamton.edu, Libraries, Binghamton University, Binghamton, NY 13902-6012, United States*

Research administrators need to identify promising researchers and future growth areas and recognize established research innovators. Journal article indexes measure past research accomplishments through citation and reference counts. This data is inadequate for identifying and ranking future high-impact research activities and programs.

Research analytical tools built from literature databases can readily identify future high-impact activities, institutional peers and potential collaborators and recruits. Libraries can use collections data to assess and identify future collections priorities. Examples of how the libraries can implement and use these research analytical tools will be shown.

Informal scholarship activities provide data on research impact. This includes the blogosphere, shared data sites, social networks, subject digital repositories, and popular media presence. Site metrics provide data on high-impact activities among peers and supplement research and library database tools. Examples of these activities and their impact will be shown.

CINF 122. Providing comparative data on published research impact (internally and externally)

Donna T. Wrublewski, *dtwrublewski@ufl.edu, Denise B. Bennett, Valrie I. Davis, Michelle Leonard, George A. Smathers Libraries, University of Florida, Gainesville, Florida 32611, United States*

Science librarians at the University of Florida have recently been involved in two data collection projects requested by university departments. The first required the collection and evaluation of comparative indicators of research productivity, with a five-day turnaround, comparing UF with selected Chemical Engineering departments at five other institutions. The second involved the development of publication lists for the University of Florida Clinical and Translational Science

Institute to use in their Executive Council meetings to identify and analyze output of affiliated Institute researchers. Both projects required negotiated understanding of the needs and expectations of each department request and involved manual manipulation and analysis of data. These studies established rough guidelines for procedures that can be used in the future for institutional requests, and the methodological issues encountered are broadly applicable to many "fact-finding" scenarios.

CINF 123. Social networking tools as public representations of a scientist

Antony J Williams, *williamsa@rsc.org, Department of Informatics, ChemSpider, Royal Society of Chemistry, Wake Forest, NC 27587, United States*

The web has revolutionized the manner by which we can represent ourselves online by providing us the ability to exposure our data, experiences and skills online via blogs, wikis and other crowdsourcing venues. As a result it is possible to contribute to the community while developing a social profile as a scientist. At present many scientists are still measured by their contributions using the classical method of citation statistics and a number of freely available online tools are now available for scientists to manage their profile. This presentation will provide an overview of tools including Google Scholar Citations and Microsoft Academic Search and will discuss how these are and other tools, when integrated with the ORCID identifier, may more fully recognize the collective contributions to science. I will also discuss how an increasingly public view of us as scientists online will likely contribute to our reputation above and beyond citations.

CINF 124. Next era of research productivity evaluation: A multidimensional research assessment framework

Daniel Calto, *d.calto@elsevier.com, Atyab Tahir, Elsevier Inc., United States*

There is a clear need for performance measures related to research productivity. But which types of content and metrics should be included in such measures? What do reasonable performance measures look like, and how should they be applied? Many advocates of new metrics state that measurements need to be more quantitative and objective. Others argue that qualitative metrics capture critical aspects about a researcher's overall effectiveness.

We propose that no single metric is adequate to capture the true dynamics of a researcher's productivity, but that an approach combining qualitative and quantitative metrics can lead to a reasonable and standards-based approach to evaluating research productivity. We would like to speak on a matrix created by Elsevier's bibliometric expert that provides the evaluator with a flexible framework to identify which elements are measured and which metrics to use, and shows how the purpose of the evaluation helps to determine the structural elements of the assessment.

CINF 125. Measuring research: Beyond H

Daniel Hook, *daniel@symplectic.co.uk, Symplectic Limited, London, United Kingdom*

Research information management (RIM) systems give us more possibilities than ever before to capture and analyze data about the research taking places in academic institutions. The key challenge in making RIM systems successful is ensuring that data is collected efficiently. To ensure quality, faculty are often involved in this process: There needs to be a tangible benefit to them and assurances that data, once collected, won't be used negatively.

A concerning issue for many faculty is use H-Index. This is symptomatic of a more general fear of inappropriate or uninformed use of bibliometric measures. Classic bibliometric measures, such as H-Index or Impact Factor, might typically be thought of as the equivalent of "mean averages". This causes unease in faculty as these averages lack context.

We will consider the role of the library in data collection and in educating faculty about bibliometric measures, which implicitly contextualise data and go beyond H.

CINF 126. Methods and solutions for measuring and benchmarking the impact of research

Daphne Grecchi,
Daphne.Grecchi@thomsonreuters.com, IP & Science, Thomson Reuters, Philadelphia, PA,, PA 19130, United States

Academic institutions are increasingly tasked with demonstrating research productivity via objective measurements. The library can play an important role in quantitative research evaluation, in terms of implementing and educating. Bibliometrics has been the primary method for decades. Traditional indicators focused on journal literature are most applicable to the sciences. The analysis of books and patents as further research outputs can augment traditional journal metrics and new approaches such as network-based metrics offer additional perspectives on the research landscape. Bibliometrics should be used in context and in conjunction with other research performance measures.

Thomson Reuters has worked in the field of citation indexing and analysis for over 50 years, beginning with our roots as the Institute for Scientific Information. We offer a suite of options from web-based evaluation tools like InCites and Research In View, to customized reports and engineered systems. Examples of metrics and methods from these tools will be discussed.

CINF 127. Can we really do computer-aided drug design?

Matthew D Segall, *matt.segall@optibrium.com, Optibrium Ltd., Cambridge, United Kingdom*

We will explore the accuracy of current computational methods in drug discovery, including 2D and 3D QSAR, docking, pharmacophore, molecular dynamics and quantum mechanical approaches. Based on this, we will address the question of whether we are truly operating in a drug design paradigm. We will compare this with the application of computational methods to the discovery of new drugs. From this alternative perspective, computational methods can add significant value to guide decisions about which chemistry to pursue and which can be rejected with confidence; focussing resources on the chemistry that is most likely to succeed, while avoiding missed opportunities. This is particularly important in the multi-parameter optimisation of high quality drug candidates that require a balance of many properties to succeed downstream.

CINF 128. Where screening starts: Effective preprocessing of chemical libraries

Matthias Hilbig, Adrian Kolodzik, Sascha Urbaczek, ***Matthias Rarey****, rarey@zbh.uni-hamburg.de, Center for Bioinformatics, University of Hamburg, Hamburg, Germany*

Today, every scientist has electronic access to more than 13 million commercially available compounds via ZINC [1] and even more when structure collections like PubChem [2] are used. Although computing speed still increases substantially, modeling tasks like docking, pharmacophore and 3D-similarity searching remain demanding tasks. Furthermore, they all rely on high quality structures including correct protonation and tautomeric states. Preprocessing chemical libraries, especially filtering them by restricting simple scalar properties and applying substructure-based exclusion rules, is therefore a frequently applied task. This step is often performed with scripts or pipelining tools, making the adaption of the library to the individual target and modeling task difficult. We propose to use a more interactive process in order to tailor-make compound collections on a case-by-case basis. We developed a tool named Mona supporting this process by handling a multitude of descriptors for large data sets in an efficient database so that compound collections can be customized on the fly.

[1] ZINC library, http://zinc.docking.org

[2] PubMed, http://pubchem.ncbi.nlm.nih.gov

CINF 129. Toward a gold standard: Improving the quality of public domain chemistry databases

Antony J Williams[1], ***Sean Ekins***[2], *ekinssean@yahoo.com. (1) Royal Society of Chemistry, Wake Forest, NC 27587, United States (2) Collaborations in Chemistry, Fuquay Varina, NC 27526, United States*

In recent years there has been a dramatic increase in the number of freely accessible online databases serving the chemistry community such that the internet now has a rich array of chemistry data. This is useful for data-

mining, computer modeling, and for integrating into other systems to expand data accessibility and aid drug discovery. With this improved data accessibility comes a responsibility to ensure that it is as high quality as possible. This will prevent scientists from wasting time performing erroneous searches, creating flawed computational models etc. Improved discoverability of online resources should not be marred by the delivery of incorrect data. We will describe our experiences with multiple chemical compound databases and other online resources. We will suggest approaches to collaborate to deliver definitive reference data sources for researchers and additionally describe the creation of a new wiki for the community to contribute and rank databases (www.scidbs.com).

CINF 130. ChemSpider as a knowledge base

Valery Tkachenko[1], *tkachenkov@rsc.org, Antony Williams*[1], *Aileen Day*[2], *Jon Steel*[2]. *(1) Department of Informatics, ChemSpider, Royal Society of Chemistry, Wake Forest, NC 27587, United States (2) Department of Informatics, ChemSpider, Royal Society of Chemistry, Cambridge, United Kingdom*

The amount of information on the internet is proliferating at such a speed that it is difficult to comprehend how much data will be available online in the coming years. The domain of Chemistry is surprisingly complex using its own multiple languages of chemical names, chemical structures, terminologies and ontologies. Data encoded in these forms is already on the web. Numerous efforts have been made to capture and host these data and enable them to be discoverable. ChemSpider has previously focused on being a "structure-centric database" but efforts are now afoot to extend the system into a "chemistry knowledgebase". Traditionally machine-to-machine communications were facilitated by the use of web services but the diversity of chemistry-related information makes it hard to provide comprehensive web services layer for knowledge bases. An alternative approach using semantic web technologies such as SPARQL has been implemented and this presentation will report our work.

CINF 131. ChemSpider as a chemical term resolver

Valery Tkachenko, ***Antony Williams****, WilliamsA@rsc.org, Department of Informatics, ChemSpider, Royal Society of Chemistry, Wake Forest, NC 27587, United States*

In recent years, in parallel with the general broad trend of information proliferation, many tens of public chemical databases have been created and made available using internet technologies. In many cases fluent data exchange has occurred between these various databases as they source information from one another. While this has the advantages of linking together multiple data sources the results also include the proliferation of errors across the various databases. The lack of a public authority to resolve such errors significantly affects the quality of freely accessible chemical information. While ChemSpider has previously allowed a crowdsourcing approach to curation efforts have now migrated to addressing this problem using a "federated resolver" approach. This presentation will report on our work in this area.

CINF 132. How to design chemical patterns easily with an interactive editor

Karen T. Schomburg*, schomburg@zbh.uni-hamburg.de, Lars Wetzer, Matthias Rarey, Center for Bioinformatics, University of Hamburg, Hamburg, Germany*

Chemical patterns are descriptions of generic chemical structures. They are essential for methods like searches in molecule databases or filtering of datasets, which practicing chemists employ frequently. However, due to their background, the representations of patterns (as linear languages like SMARTS1) are optimized for efficient computational interpretation. Their regular expression-like constitution makes them hard to use and creates an impediment to work with many fundamental chemoinformatic methods. A graphical interface to chemical patterns similar to structure diagrams is more adequate to the standards in chemical society. It supports the understanding of single patterns as well as becoming familiar with the concept of chemical patterns. The SMARTSviewer provides a visualization concept following the graphic standards of structure diagrams, along with an interactive editor, allowing intuitive design of chemical patterns from scratch. No

specific knowledge about the SMARTS language is needed to understand or design a chemical pattern.

[1] James, C. A., Weininger D., Daylight Theory Manual. Daylight Chemical Information Systems, Inc. of Aliso Viejo, CA, available at www.daylight.com

CINF 133. Lexichem TK 2.1.0

***Edward O Cannon**, ed.cannon@eyesopen.com, OpenEye Scientific Software, Santa Fe, NM 87508, United States*

Lexichem is a chemical nomenclature toolkit created by OpenEye Scientific Software that provides a fast and reliable way of converting chemical names to chemical structures and back in over 10 different languages. This presentation aims to give an overview of the new developments introduced in Lexichem toolkit version 2.1.0.

A new performance metric based on percentage round tripping of canonical isomeric smiles is introduced. The advantages of this metric, and why it will be used as a benchmark in future software releases will be discussed.

Using this new metric, we report significant increases (up to 39%) in performance from our previous release of Lexichem (v2.0.2) over four databases.

CINF 2012 Officers and Functionaries

Chair:
Dr. Rajarshi Guha
NIH Chemical Genomics Center
9800 Medical Center Drive
Rockville, MD 20852
814- 404- 5449 (voice)
812-856-3825 (fax)
rajarshi.guha@gmail.com

Chair Elect:
Dr. Antony Williams
VP Strategic Development
ChemSpider
Royal Society of Chemistry
(919) 201-1516 (voice)
williamsa@rsc.org

Past Chair/Nominating Chair:
Dr. Gregory Banik
Bio-Rad Laboratories, Inc.
Informatics Division
2 Penn Center Plaza, Suite 800
1500 John F Kennedy Blvd
Philadelphia, PA 19102-1721
267-322-6952(voice)
267-322-6953 (fax)
gregory_banik@bio-rad.com

Secretary:
Ms. Leah Solla
Cornell University
Physical Sciences Library
283 Clark Hall
Ithaca, NY 14853-2501
607-255-1361 (voice)
607-255-5288 (fax)
lrm1@cornell.edu

Treasurer:
Ms. Meghan Lafferty
University of Minnesota
Science & Engineering Library
108 Walter Library
117 Pleasant St SE
Minneapolis MN 55455
612-624-9399 (voice)
612-625-5583 (fax)
mlaffert@umn.edu

Councilor:
Ms. Bonnie Lawlor
National Federation of Advanced Information Services (NFAIS)
276 Upper Gulph Road
Radnor, PA 19087-2400
215-893-1561 (voice)
215-893-1564 (fax)
blawlor@nfais.org

Councilor:
Ms. Andrea B. Twiss-Brooks
4824 S Dorchester Avenue, Apt 2
Chicago, IL 60615-2034
773-702-8777 (voice)
773-702-3317 (fax)
atbrooks@uchicago.edu

Alternate Councilor:
Mr. Charles F. Huber
University of California, Santa Barbara
Davidson Library
Santa Barbara, CA 93106
805-893-2762 (voice)
805-893-8620 (fax)
huber@library.ucsb.edu

Alternate Councilor:
Dr. Guenter Grethe
352 Channing Way
Alameda, CA 94502-7409
510-865-5152 (voice and fax)
ggrethe@att.net

Program Committee Chair:
Dr. Rachelle Bienstock
National Institute of Environmental Health Sciences
PO Box 12233 MD F0-011
Research Triangle Park, NC 27709
919-541-3397 (voice)
rachelleb1@gmail.com

Membership Committee Chair:
Dr. Gregory Banik
See Past Chair/Nominating Chair

Archivist/Historian:
Ms. Bonnie Lawlor
See Councilor

Audit Committee Chair:
Ms. Jody Kempf
Science & Engineering Library
University of Minnesota
108 Walter Library
117 Pleasant St SE
Minneapolis MN 55455
612-624-9399 (voice)
612-625-5583 (fax)
j-kemp@umn.edu

Awards Committee Chair:
Ms. Andrea Twiss-Brook
See Councilor

***Chemical Information Bulletin* Editors:**
Ms. Judith N. Currano (Spring, Fall)
University of Pennsylvania
Chemistry Library
231 S. 34th St. 5th Floor
Philadelphia, PA 19104-6323
215-746-5886 (voice)
215-898-0741 (fax)
currano@pobox.upenn.edu

Ms. Svetlana Korolev (Winter, Summer)
University of Wisconsin, Milwaukee
2311 E. Hartford Avenue
Milwaukee, WI 53211
414-229-5045 (voice)
414-229-6791 (fax)
skorolev@uwm.edu

Communications & Publications Committee Chair:
Dr. William Town
Kilmorie Consulting
24A Elsinore Road
London SE23 2SL
United Kingdom
+44 20 8699 9764 (voice)
bill.town@kilmorie.com

Constitution, Bylaws & Procedures:
Ms. Susanne Redalje
University of Washington
Chemistry Library
Box 351700
Seattle, WA 98195
206-543-2070 (voice)
curie@u.washington.edu

Education Committee Chair:
Mr. Charles F. Huber
See Alternate Councilor

Finance Committee Chair:
Ms. Meghan Lafferty
See Treasurer

Fundraising Committee Chair:
Mr. Graham Douglas
Scientific Information Consulting
1804 Chula Vista Drive
Belmont, CA 94002
510-407-0769 (voice)
Graham_C_Douglas@hotmail.com

Webmaster:
Ms. Danielle Dennie
Concordia University
Vanier Library Building
7141 Sherbrooke St. W.,
Montréal (QC), H4B 1R6, Canada
514-848-2424 x 5237 (voice)
danielle.dennie@concordia.ca