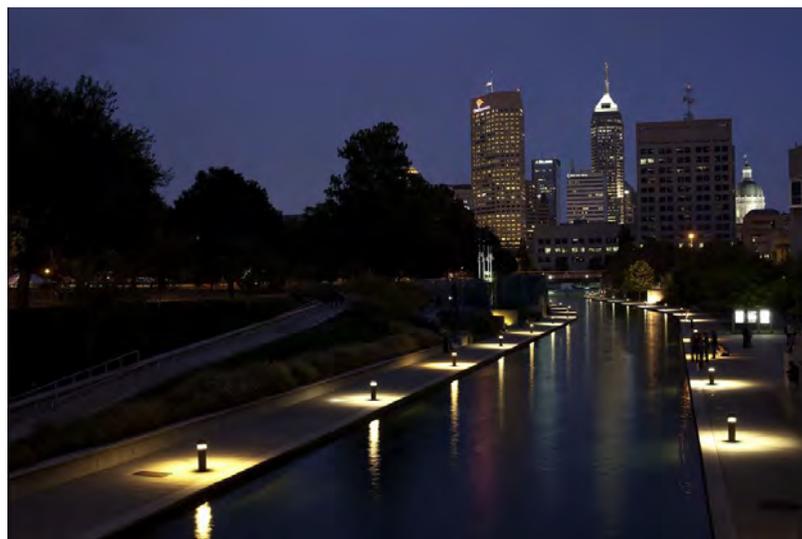


CHEMICAL Information BULLETIN



ACS
Chemistry for Life®

Winter 2013 — Vol. 65, No. 4

A Publication of the
Division of Chemical Information
of the ACS

ISSN: 0364-1910



Chemical Information Bulletin

A Publication of the Division of Chemical Information of the ACS

Winter 2013 — Vol. 65, No. 4

Svetlana Korolev, Editor,
University of Wisconsin, Milwaukee
skorolev@uwm.edu

Message from the Chair	2
Letter from the Editor	3
Awards and Scholarships	4
2013 Herman Skolnik Award Presented	4
2013 CINF Lifetime Award Presented	5
2013 CINF Scholarship for Scientific Excellence Presented	6
2013 Lucille Wert Scholarship Presented	7
2014 Lucille Wert Scholarship: Call for Applications	7
2014 Herman Skolnik Award Announced	8
Chemical Structure Association Trust Award	9
Applications Invited for CSA Trust Grants for 2014	10
Technical Program	11
CINF Technical Program Highlights	11
Chemistry on Tablet Computers	12
Integrative Chemogenomics Knowledge Mining using NIH Open Access Resources	15
Exploring the Role and Value of Social Networking in Advancing the Chemical Sciences	18
Science-Based Policy Development in the Environment, Food, Health, and Transport Sectors	20
Herman Skolnik Award Symposium honoring Richard Cramer	23
Before and After Lab: Instructing Students in 'Non-Chemical' Research Skills	43
Exchangeable Molecular and Analytical Data Formats	46
Multidisciplinary Program Planning Group	48
Interview with Kristin Briney, 2013 Lucille Wert Scholarship Winner	50
Book Reviews	54
CRC Handbook of Chemistry and Physics Celebrates its 100 th Anniversary	54
Does Science Need a Global Language?	56
Every Molecule Tells a Story	57
Committee Reports	58
ACS Council	58
Joint Board-Council Committee on Publications	62
CINF Social Networking Events	63
CINF Officers	64
Contributors to this Issue	66

ISSN: 0364-1910

Chemical Information Bulletin, © Copyright 2013 by the Division of Chemical Information of the American Chemical Society

Message from the Chair



As the memories of ACS Indianapolis meeting start to drift I reflect on what a great meeting it was. The weather was superb, the sessions that I attended were wonderful and the chance to meet with friends and colleagues in a social setting to discuss our changing world of chemical information is always good. The theme of the meeting was “Chemistry in Motion” and I would say that CINF certainly kept things moving for the attendees in terms of a great program and well-attended social events.

The highlights of the meeting for me were the excellent programming, the honor of presenting Dick Cramer with the 2013 Herman Skolnik award, hosting the CINF Luncheon with our guest speaker Katy Borner, and attending the reinvigorated Harry’s Party. I was also running in many directions between sessions I was presiding, many presentations I was giving, and a multitude of social and networking gatherings, so “Chemistry in Motion”....yes, it was! Hopefully, each of you attending walked away with similar fond memories of the gathering as well as having learned something new and exciting that is happening in our domain of chemical information.

After each ACS meeting I ask myself the question “Was it worth attending?” Certainly, attending conferences takes us away from our families, distracts us from our mainstream work for many days, and can be a burden to our bodies as we generally start early and finish late. My answer for “Indy 2013” was a resounding “Yes, it was definitely worth attending!” I came away feeling simultaneously exhausted by the pace and exhilarated by the conversations and the possibilities for the growing importance of our domain and what our Division can offer to the community. We are positively an example of a division working hard to make a difference. As usual, it was great to come together with the CINF Division leadership team to work on making “ACS Indy” a success and to plan the future meetings and activities. Hopefully, you are sensing our commitment to continued growth and success for the Division.

That said, one sad piece of information reported in Indy is the continued decline in our membership despite sincere efforts over the past few months to encourage sign up. We are hoping that the registration of the new members from the CINF reception in New Orleans will help us turn the corner on the declining membership and, again, we encourage you to renew your membership and to encourage others to join us, so that we all work together to continue to ensure that CINF is not only relevant, but is a vibrant ACS Division for the foreseeable future.

With that I will bow out and thank you for your support during the period during which I have served the Division as Chair. It has been my pleasure to serve with an incredible team of dedicated individuals to ensure superior programming at the meetings and to keep the Division moving onwards and upwards. I look forward to supporting Judith Currano, the next Division Chair, as she steps into the role in 2014. So until Dallas...have a wonderful holiday season and hope to see you all next year.

Antony Williams, Chair, ACS Division of Chemical Information

Letter from the Editor

Welcome to the post-conference issue of *Chemical Information Bulletin (CIB)*! Focusing on the CINF technical program, it includes the Program Chair's highlights, full reports of seven CINF symposia, and updates from the *Multidisciplinary Program Planning Group*. Since its establishment in 1976, the Herman Skolnik Award Symposium has been the central point of the CINF program at a fall meeting. Richard Cramer, the 2013 Herman Skolnik Award recipient, organized a symposium for the occasion. Cramer's Award Address, as well as four other presentations of this symposium, was recorded for the [Presentations on Demand](#). Wendy Warr wrote a comprehensive overview (20 pages, 84 references) of the Award Symposium honoring Richard D. (Dick) Cramer for this issue.



"It has been quite a while since ACS has been in Indianapolis – last time was in 1931 – and this marks the 3rd time ACS is holding a national meeting in this city," noted Marinda Li Wu, 2013 ACS President, in her *Welcome to Indianapolis* (ACS Show Daily, September 9, 2013). Saluting the city hosting "Indy 500," the theme of the meeting was "Chemistry in Motion." Our Division supported the theme by organizing a symposium "Exchangeable molecular and analytical data formats and their importance in facilitating data exchange" by Robert Lancashire and Antony Williams. They summarized their session for this Bulletin. Antony Williams, Division Chair, remarked on the "Chemistry Motion" theme: "I would say that CINF certainly kept things moving for the attendees in terms of a great program and well-attended social events." Read more on the broadening perspectives of the ACS thematic program by Guether Grethe in this issue.

Speaking of chemical information, we celebrated two milestone anniversaries at the Indianapolis meeting. The first edition of the *CRC Handbook of Chemistry and Physics* was published a hundred years ago. Mickey Haynes, the Editor-in-Chief of the Handbook, wrote a "book review" for this Bulletin. For the view from history, I appended a brief book review published in *JACS* in 1917. The following excerpt from one of the earliest reviews summarizes the high quality standard of this publication sustained over the century: "In continuing their policy of frequent revision and giving special consideration to the suggestions of those who have used former editions, the authors are succeeding remarkably well in developing a book along lines most acceptable to those interested in a volume of this type." (*Ind. Eng. Chem.* **1928**, 20 (7), 776–777)

The 90th Anniversary of *C&EN* (*Chemical & Engineering News*) was widely publicized at the Indiana convention center. *C&EN* began its historic run as the *News Edition of the Industrial and Engineering Chemistry* on January 10, 1923, and it continues to be the leading weekly newsmagazine "Serving the chemical, life sciences, and laboratory worlds" (*C&EN* logo) to the ACS members and other subscribers. The 90th Anniversary Special Issue was published on September 9, 2013. Debra Davis from ACS provided highlights and presentation slides from the Joint Board Council Committee on Publications for this issue of *CIB*.

As usual, the Bulletin includes award announcements and calls for nominations, Editor's choice interview with the 2013 Lucille Wert Scholarship recipient Kristin Briney, book reviews, ACS Council report, and a summary of the CINF social networking events. Hopefully, the readers of *Chemical Information Bulletin* can emulate the experience of the meeting attendees, who contributed timely reports to this publication (66 pages). I express my sincere thanks to all writers.

Svetlana Korolev, Editor, Chemical Information Bulletin

Awards and Scholarships

2013 Herman Skolnik Award Presented

Dr. Richard (“Dick”) Cramer was awarded the 2013 Herman Skolnik Award by the Division of Chemical Information (CINF) during the 246th American Chemical Society National Meeting held in Indianapolis, IN, September 8-12, 2013.

This award recognizes Cramer’s outstanding contributions to and achievements in the theory and practice of chemical information science. Cramer’s scientific breakthroughs include the invention of Comparative Molecular Field Analysis (CoMFA), the first and most widely-used 3D-QSAR technique for molecular discovery. Researchers use CoMFA to build statistical and graphical models that relate the chemical and biological properties of molecules to their 3D structures and the 3D steric and electrostatic properties. These models are then used to predict the activity of novel compounds. This information helps researchers to decide which molecules will likely make the best new drug candidates. This development earned Cramer one of the earliest cheminformatics patents. He continues to refine this technique through his work on topomeric descriptors and other QSAR innovations.

Dr. Cramer organized a one-day symposium for the occasion, which included the following presentations (links to abstracts below, a symposium report by Wendy Warr is in this Issue):

[Adventures in CoMFAland](#) / Robert D Clark

[Adventures in drug discovery: For now we see through a glass, darkly](#) / Robert C Glen

[Three paradigm shifts in computer-assisted drug design: The inventors and by-standers](#) / Yvonne C Martin

[Look back at 3D-QSAR and Dick Cramer](#) / Anton J. Hopfinger

[Evolution of QSAR from regression analysis to physical modeling](#) / Ajay N Jain

[Scientific analysis of baseball performance](#) / David W. Smith

[Synthesis planning: Something about reactions, representation, relationships, and reasoning](#) / W. Todd Wipke

[Think local, act global: Some challenges in cheminformatics and drug research](#) / Tudor I Oprea

[From library design to off-target prediction: A wide array of topomer applications](#) / Bernd Wendt

2013 Herman Skolnik Award address:

[Whole template CoMFA: The QSAR grail?](#) / Richard D Cramer

Andrea Twiss-Brooks, Chair, CINF Awards Committee

2013 CINF Lifetime Award Presented



Guenter Grethe receiving congratulations from Antony Williams, CINF Chair

Guenter Grethe was presented the CINF Lifetime Award during the Division Luncheon in Indianapolis on Tuesday, September 10, 2013.

The award was established by the Executive Committee of the Division of Chemical Information in 2006. It recognizes long-term membership, and outstanding service and active contributions to the Division over the years. The recipient must have been a member of the Division for at least 20 years.

Guenter has served CINF in various offices, and has been especially active in promoting the development of international partnerships and activities for CINF, including the XCITR project and the International Conference on Chemical Structures.

Guenter was the recipient of the 2001 Herman Skolnik Award and the 2004 Val Metanomski Meritorious Service Award from the Division.

Congratulations, Guenter!

Andrea Twiss-Brooks, Chair, CINF Awards Committee

2013 CINF Scholarship for Scientific Excellence Presented

The scholarship program of the Division of Chemical Information (CINF) of the American Chemical Society (ACS) is designed to reward graduate and post-graduate students in chemical information and related sciences for scientific excellence and to foster their involvement in CINF. At each of the ACS National Meetings since 2005, the program has awarded a total of 49 scholarships. The awards at the Fall 2013 National Meeting in Indianapolis were sponsored by the [Royal Society of Chemistry](#). Applicants presented their posters at the CINF Welcoming Reception and the Sci-Mix session. Two scholarships valued at \$1,000 each were presented to the winners at the CINF Luncheon during the same meeting.

The names of the recipients and the titles of their posters are:

Johannes Hachmann, Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, "The Harvard Clean Energy Project: From big data and cheminformatics to the rational design of molecular OVP materials."

Co-authors: Roberto Olivares-Amaya, Alan Aspuru-Guzik

Abhik Seal, School of Informatics and Computing, Indiana University, Bloomington, IN, "The enhanced ranking of PknB inhibitors using data fusion methods."

Co-authors: David J. Wild



Johannes Hachmann, Abhik Seal, Guenter Grethe

The next scholarships jointly sponsored by InfoChem and Springer will be awarded at the 2014 Spring ACS National Meeting in Dallas, TX.

Guenter Grethe, Coordinator, CINF Scholarship for Scientific Excellence

2013 Lucille M. Wert Scholarship Presented



Kristin Briney receiving congratulations from Andrea Twiss-Brooks, CINF Awards Chair

Kristin Briney was presented [the 2013 Lucille M. Wert Scholarship](#) during the ACS-Milwaukee Section Dinner Meeting, in Klemmer's Banquet Center, on October 17, 2013.

An interview with Kristin Briney is featured in this issue of *Chemical Information Bulletin*.

Andrea Twiss-Brooks, Chair, CINF Awards Committee

2014 Lucille M. Wert Scholarship: Call for Applications

Designed to help persons with an interest in the fields of Chemistry and Information to pursue graduate study in Library, Information, or Computer Science, the Scholarship consists of a \$1,500 honorarium. This scholarship is given yearly by the Division of Chemical Information of the American Chemical Society.

The applicant must have a bachelor's degree with a major in Chemistry or related disciplines (related disciplines are, for example, Biochemistry or Chemical Informatics). The applicant must have been accepted (or currently enrolled) into a graduate Library, Information, or Computer Science program in an accredited institution. Work experience in Library, Information or Computer Science is preferred.

The deadline to apply for the 2014 Lucille M. Wert Scholarship is **February 1, 2014**.

Details on the application procedures can be found at:
<http://www.acscinf.org/content/lucille-m-wert-student-scholarship>.

Applications (email preferred) can be sent to: margaret.matthews@thomsonreuters.com

Contact information: Marge Matthews, CINF Awards Committee
633 Dayton Rd., Bryn Mawr, PA 19010-3801
Phone: 215-823-3922

Marge Matthews, Coordinator, Lucille M. Wert Scholarship

2014 Herman Skolnik Award Announced

Dr. Engelbert Zass has been selected as the recipient of the American Chemical Society Division of Chemical Information's 2014 Herman Skolnik Award.

The award recognizes outstanding contributions to and achievements in the theory and practice of chemical information science and related disciplines. The prize consists of a \$3,000 honorarium and a plaque. Dr. Zass will also be invited to present an award symposium at the Fall 2014 ACS National Meeting to be held in San Francisco.



Dr. Zass, Head of the Chemistry Biology Information Center at ETH Zürich (retired), is being recognized for outstanding contributions and achievements in the practice of chemical information science, notably for his lifelong work in education, research and development activities. Throughout his career he has been a true bridge-builder and mediator between database producers, vendors, publishers, librarians and end-users in chemistry, contributing to advancing chemical information as a whole.

Dr. Zass specialized in chemical information after receiving his Ph.D. in organic chemistry, and has more than 30 years of experience in searching, operating, and designing chemistry databases, as well as in the support, training and education of users of chemical information. He has given numerous lectures and courses in Europe and the U.S., is author of more than 60 papers on chemical information, and served on several publisher advisory boards. From 1999 till 2004, he was a partner in the BMBF (German Federal Ministry of Education and Research) Project "Vernetztes Studium – Chemie," where he was engaged in the design of multimedia educational material for chemical information. Through his leadership, vision, and collaborative efforts with his staff, he developed a model 21st century library that serves chemists and biologists at ETH.

To his many friends and admirers, "Bert" has been a true leader in the profession, generous in sharing his expertise with colleagues. His dedicated, sustained efforts and the transformative impact he has had on chemical information systems, database producers, chemists, and librarians make him a worthy recipient for the Herman Skolnik Award.

Dr. Zass did his undergraduate studies in chemistry at Universität zu Köln, followed by a Master's degree (Diplom) in Chemistry with Prof. E. Vogel. He went on to complete graduate studies with Prof. A. Eschenmoser at the Laboratorium für organische Chemie der ETH Zürich, culminating with a Ph.D. in Organic Chemistry (Dr. sc. nat.). Upon completing his education, Dr. Zass was a lecturer and senior scientist at the ETH, Laboratorium für organische Chemie ETH / Chemistry Information Center, later serving as Head of the expanded ETH Chemistry Biology Pharmacy Information Center until his retirement in 2012.

Andrea Twiss-Brooks, Chair, CINF Awards Committee



Chemical Structure Association Trust Award

The Chemical Structure Association Trust is an internationally recognized organization established to promote the critical importance of chemical information to advances in chemical research. In particular, the Trust strives to create a heightened and sustained awareness of the essential role that is played in scientific research by the systems and methodologies used for the storage, processing and retrieval of information related to chemical structures, reactions and compounds. In support of its charter, the Trust has created the Chemical Structure Association Trust Award as well as a unique Grant Program. These programs are financed by investments managed by the Trust and through funds donated by industrial, academic, and government organizations that recognize the value of and benefit from research, development, and education in the fields supported by the Trust.

The Trust is now seeking submission of nominations for its Award for its next round of deliberations.

Award Program

Purpose: To recognize and encourage *outstanding* accomplishments in education, research and development activities that are related to the systems and methods used to store, process and retrieve information about chemical structures, reactions and compounds.

Nature: The Award, given on a tri-annual basis beginning in 2002, consists of five hundred U.S. dollars (\$500.00) and an appropriate memento. The Award will be presented at a prestigious, relevant conference to be identified prior to each presentation. The awardee will be asked to give a presentation at the conference.

Eligibility: The Award shall be granted to an individual without regard to age or nationality for outstanding achievement in education, research or development in the area of systems and methods used for the storage, processing and retrieval of information about chemical structures, reactions and compounds. Nominations of persons known to be deceased will not be considered. Posthumous awards will be made only when knowledge of the recipient's death is received after the Award Committee has announced their decision.

Nominations: Any individual may submit one nomination or one seconding letter for the Award in any given year. In nomination by petition, the person whose signature is first will be considered to be the nominator. The required nominating documentation is as follows: 1) a letter that evaluates the nominee's accomplishments and the specific relevant work that is to be recognized; 2) a biographical sketch, including a statement of academic qualifications, as well as contact information; 3) at least two seconding letters that support the nomination and provide additional factual information with regard to the scientific achievements of the nominee. If appropriate, a list of the nominee's publications and/or patents may also be submitted. The Award will not be given in any year in which the nominees do not meet the award criteria.

The information should be submitted via email to René Deplanque, Chair of the Awards Committee, (deplanque@gmx.de) and to Bonnie Lawlor, CSA Trust Secretary, (blawlor@nfais.org) by **November 15, 2013**.

Bonnie Lawlor, Secretary, CSA Trust

Applications Invited for CSA Trust Grants for 2014

The Chemical Structure Association (CSA) Trust is an internationally recognized organization established to promote the critical importance of chemical information to advances in chemical research. In support of its charter, the Trust has created a unique Grant Program and is currently inviting the submission of grant applications for 2014.

Purpose of the Grants: The Grant Program has been created to provide funding for the career development of young researchers who have demonstrated excellence in their education, research or development activities that are related to the systems and methods used to store, process and retrieve information about chemical structures, reactions and compounds. One or more Grants will be awarded annually up to a total combined maximum of ten thousand U.S. dollars (\$10,000). Grants are awarded for specific purposes, and within one year each grantee is required to submit a brief written report detailing how the grant funds were allocated. Grantees are also requested to recognize the support of the Trust in any paper or presentation that is given as a result of that support.

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

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

Application Requirements: Applications must include the following documentation:

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

Additional materials may be supplied at the discretion of the applicant only if relevant to the application and if such materials provide information not already included in items 1-4. Three copies of the complete application must be supplied for distribution to the Grants Committee.

Deadline for Applications: Applications must be received no later than **March 13, 2014**. Successful applicants will be notified no later than May 2, 2014.

Address for Submission of Applications: Three copies of the application documentation should be forwarded to: Bonnie Lawlor, CSA Trust Grant Committee Chair, 276 Upper Gulph Road, Radnor, PA 19087, USA. If you wish to enter your application by e-mail, please contact Bonnie Lawlor at chescot@aol.com prior to submission so that she can contact you if the e-mail does not arrive.

Bonnie Lawlor, Chair, CSA Trust Grant Committee

Technical Program

CINF Technical Program Highlights

I would like to thank all of the organizers, speakers, and poster presenters who contributed to the Indianapolis Meeting. Indianapolis is considered a hometown to me as I am a native Hoosier, so I was very excited that the city would be hosting. While attendance may have been lower than some previous meetings, we had an extremely interesting and diverse program. Some of the Indianapolis organizers have written symposia reports that can be found in this issue, so please check out what you missed or refresh your memories as to what you saw presented. We ended up with eleven symposia plus the Herman Skolnik Award Symposium and the General Papers session. Final count after withdrawn presentations and posters was ~110 presentations and ~10 posters for Sci-Mix.



Three of our sessions were recorded as part of the ACS Presentations on Demand program and will be made available to registered attendees (<http://presentations.acs.org/common/default.aspx>). At the Indianapolis Meeting, ACS announced it will also be expanding the Presentations on Demand service as a member benefit, regardless of whether you attend the Meeting or not. The sessions were:

- Current Challenges in Cheminformatics: Exploiting Information and Knowledge in Structured and Unstructured Environments
- Herman Skolnik Award Symposium
- Joint CINF-CSA Trust Symposium: Semantic Technologies in Translational Medicine and Drug Discovery

General Papers

With only two presentations in the General Papers session, and it being the Thursday afternoon of the meeting, we were still pleasantly surprised that our audience fluctuated between six and eight attendees (that is not counting the organizers/speakers!). And as luck would have it, both speakers were also able to give their talks in other symposia during the week where we had withdrawn papers. While they had to present twice, they were able to reach a wider audience. Stuart Chalk of University of North Florida discussed the Eureka Research Workbench, an open source ELN he has developed. Rachelle Bienstock, CINF's very own Chair-Elect 2014, discussed working with creating chemotypes within the EPA's ToxCast project.

Looking forward

Erin Bolstad (erin.bolstad@mso.umt.edu) will be primary Program Chair for the Spring Meeting in Dallas (March 16-20, 2014) with me supporting her, especially on the PACS side. The theme of the Dallas meeting will be "Chemistry and Materials for Energy." The [program](#) is shaping up and will be published in the January 20, 2014 issue of *C&EN* and at <http://www.acs.org/dallas2014>. Registration and housing will open in mid-December. The Fall 2014 Meeting will be held in San Francisco (August 10-14, 2014).

Jeremy Garritano, Chair, CINF Program Committee

Chemistry on Tablet Computers

Sunday morning at 8:10 am in Indianapolis saw the opening of the technical symposia for the Chemical Information Division, and one of the two sessions at that time was on “Chemistry on Tablet Computers.” Chemistry on mobile devices has been a constant theme for the last several meetings, appearing both in the COMP and CHED Divisions, as well as in CINF. This meeting highlighted a number of interesting developments in this area.

Tony Williams kicked off the session with a talk on “Apps and approaches to mobilizing chemistry from the Royal Society of Chemistry.” He discussed a strategy of using mobile apps to tap into the resources on the cloud for retrieval of data as well as computations. While the RSC Mobile app provides access to all of the journal content, RSC has supported development of a number of third party apps that integrate data from ChemSpider as well as other web resources. In addition to basic name searching and structure drawing and searching, these supported apps include things like Green Solvents, Lab Solvents, and Open Drug Discovery Teams. RSC is also working on a robust API for ChemSpider so developers can create their own apps for chemical information. In this way, RSC is focused on making all of their content to be mobile accessible – journal content, database content, and structure and substructure searching across all of their content. ([slideshare](#))

Layne Morsch, University of Illinois Springfield (UIS), and Hans Keil, PerkinElmer, described collaboration between PerkinElmer, McGraw Hill, Saint Louis University, and UIS. The project involved using tablet computers (iPads provided by PerkinElmer) equipped with ChemDraw for iPad and a “flick-to-share” collaboration service to study the use of these tools in the university chemistry classroom.

For PerkinElmer, this represented a new market opportunity and a chance to see how the ChemDraw app performed with real users in somewhat stressful circumstances. For the faculty members, it represented a chance to examine the benefits and challenges of using new technology in teaching and learning.

Morsch used the iPad to lecture in his classroom, including drawing structures, with the iPad projected to a screen. Using the flick-to-share functionality, problems involving structure drawing could be sent to student iPads, students could enter a structure and send the result back to Morsch. These techniques were used during lectures, but were also used for tests. The first 2 or 3 questions on an exam made use of the flick-to-share functionality, followed by distribution of the rest of the exam on paper. He did not use the flick-to-share questions on the final exam because students could flick answers to each other, in addition to the professor.

Morsch found a 100% level of engagement in classroom exercises, since the structure drawings were sent back to him. In addition, the students learned to use ChemDraw, a benefit for those continuing in chemistry, and because of the capabilities of ChemDraw, were able to draw more accurate reaction mechanisms.

Some drawbacks were that using ChemDraw was slower than drawing freehand, no text tool was available and note taking required switching apps, and the variability between students was much greater as well. However, Morsch and Keil felt that the experiment was very much a success.

Jeff Lang, American Chemical Society, gave a presentation on “Can I get that to go? Reading research articles on a tablet.” He presented a number of statistics showing that while the sales of tablets and other mobile devices are increasing, many people, and many in the student demographic,

use tablets largely for games and entertainment. Even though there is a lot of usage from both the website and the mobile app, it still accounts for only 5% of total usage. Reasons for this include the availability of scientific content via mobile devices when researchers are off campus, and the fact that 90% of downloads are still for the PDF format of the article. While the PDF format is readable on mobile devices, especially tablets, there is a difference between what Lang termed a grazing mode, looking to make connections between works as well as scanning at a cursory level for articles of interest, and digestion mode, reading for deeper understanding of relevant articles. Lang described the development of a more interactive version of the PDF, which provides the composed version of the text, but also allows for incorporation of linking capabilities normally associated with HTML. He also described a responsive approach, where the server would identify the device characteristics and connection speed, and automatically deliver the best version of an article.

The Java-based JMol has become one of the standards for molecular visualization on the Internet, but in addition to lack of support on tablets, within the last year, Java has been the target of many security problems. The U.S. Department of Homeland Security recommended disabling Java in web browsers in January, 2013. Bob Hanson, St Olaf College, presented a talk on “JSmol: Full-service molecular visualization on the Web without Java,” describing the migration of the Java-applet based software to a purely JavaScript environment. Hanson used Java2Script, an open source program developed by Zhou Renjian, to begin the conversion. Over the course of several months, Hanson got the JavaScript version, called JSmol, working in basic form, and then was able to get all of the Jmol functionality working, and finally optimized it so it was nearly as fast as the Java applet version. JSmol now works on tablets and smartphones as well as desktop and laptop operating systems. In order to compensate for the speed of the mobile devices, JSmol detects the speed of the device and uses that to decide how to render the molecule. Hanson is now working on a conversion of JSpecView to JavaScript.

Tamsin Mansley, from Dotmatics, presented on “Enabling Chemistry on the Go.” The philosophy behind Dotmatics is that scientists should be able to access their own data, with whatever tools, wherever they are. To implement that philosophy, Dotmatics employs cloud hosting and web-based and app-based access to that cloud. The data workflow comprises an electronic laboratory notebook, and includes importing, querying, analyzing, visualizing, and sharing data. The web-based tools are designed to be tablet friendly, allowing all of the dotmatics capabilities across devices in a familiar interface. Dotmatics also has an app version of their structure drawing utility, called Elemental, which also includes property calculation, and database searching.

Simon Coles, University of Southampton, discussed “Tablets in the lab: enabling the flow of chemical synthesis data into a chemistry repository.” Coles has been involved with chemical data management and cheminformatics for a number of years. These have resulted in a rich environment with ontologies to describe reaction plans, enactments, observations, and outcomes, expressed as RDF triples. Web-based LIMS and ELN environments, for example, labtrove.org, have been created. A recent paper published in *Chemistry Central Journal* included electronic supporting information submitted directly from an ELN. One of the challenges, though, is that in order to start the digital lifecycle of the experimental data, the experimental procedures from the lab must be created in an ELN environment, usually by transcribing from a paper lab notebook. Coles and coworkers are currently looking at how to introduce tablets into the lab as a way to aid in that process. This can be a challenge, especially in organic synthesis, where introducing a tablet into the lab might be a source of problems, not the least of which could be damage to the tablet. The study currently underway has focused on surveys and observations about how researchers use their paper lab notebooks for capturing data about the experiments, and what those same researchers do to prepare the data for publication. The main findings were that researchers cannot lose the functionality of the paper notebook. They need to be

able to “scribble” and to have the flexibility to deviate from the plan. A tablet app does not have to be a full ELN, but a lightweight version that interfaces to the main ELN and only performs operations that are absolutely necessary for recording in the lab. A number of apps have been created to support these requirements, including Notelus and Plan Buddy. These allow for scribbling, note taking, and incorporating of photos.

Steve Muskal of Eidogen-Sertanty finished the session with a talk on “New strategy to engage mobile computing users and developers.” Muskal has described a number of standalone mobile apps in recent meetings, with increasing capabilities for those apps to talk to each other, and to store and access data in the cloud. However, as with all apps, those were limited because they were basically stand-alone, vertical applications and it was cumbersome to move between them. In this talk, Muskal described an app, PP Mobile, which provides an interface to the Accelrys Pipeline Pilot Science Cloud. The app allows any Pipeline Pilot report, in either HTML or PDF, to be deployed to a mobile device. In addition, a dashboard allows interaction between the app and the Pipeline Pilot cloud. For example, the camera and GPS on a smartphone could be used to take a picture of a barcode on a sample bottle and record time, date, and location in addition to the sample identity. Protocols can be launched from the mobile device to run on the server. The user can come back later to review the results. Through this environment, users now have broad access to a variety of tools and services within Pipeline Pilot. Muskal concludes that this could be a game-changing app for tablet users to interact with their chemistry in the cloud. ([slides](#))

Martin Brändle and David Martinsen, Symposium Organizers



Sourcing information about SciApps

- SciMobileApps Wiki: scimobileapps.com



- Add your own apps or one you discovered
- Open content and community-driven
- Can someone add ChemDraw for iPad?



Slide courtesy of Antony Williams

Integrative Chemogenomics Knowledge Mining Using NIH Open Access Resources

The symposium took place on Monday, September 9, 2013 from 8:50 AM until approximately noon in the Indiana Convention Center in downtown Indianapolis. Five speakers, primarily from member Centers of the NIH Molecular Libraries Program (MLP), made presentations to 40-60 attendees. The topic of the session was "Open Access resources in chemogenomics," with a particular emphasis on the BioAssay Research Database (BARD, <http://bard.nih.gov/>), a new software project aimed to integrate and contextualize 10 years of MLP data resident in PubChem.

Brief opening remarks by Tudor Oprea (University of New Mexico) introduced the BARD project and the outreach goal of the symposium: raising awareness of BARD in the cheminformatics and chemogenomics communities.

Rajarshi Guha (National Institutes of Health), one of the primary leaders of BARD development at the NIH, presented a technical perspective on the application programming interface (API) underlying BARD. After talking about the high-level architecture of BARD and the main components, he dove into the details of the RESTful API that BARD provides to scientists and developers. The API provides programmatic access to all the entities stored in the BARD warehouse such as assays, projects, experiments, and so on. Currently the API serves JSON and provides a variety of system-level resources that provide structure information about available resources, schema, etc. He then went on to highlight how the REST resource hierarchy could be extended by user-contributed plug-ins. After describing the workflow for plug-in development he highlighted a few of the plug-ins that are currently available including the BADAPPLE promiscuity method (University of New Mexico) and SMARTCyp prediction tool (Technical University of Denmark). He went on to highlight the flexibility of the plug-in architecture that allows a plug-in to accept any data type (strings, files) and output arbitrary data types and formats (plain text, HTML, SVG, and so on). The result of this architecture is that plug-in functionality can range from a simple descriptor plug-in (taking a SMILES and returning a number) to a fully fledged, HTML5-rich, interactive interface to the API and database. Guha ended the presentation by stressing the fact that BARD is more than just a data store. Instead, it represents a platform that co-locates data and the methods to analyze, annotate, and interpret the data. Combined with the extensibility features built into the platform, BARD represents a hub for collaborations between experimentalists and computational scientists.

Alexander Tropsha (University of North Carolina) spoke about how his group developed a BARD plug-in that connects BARD to the Chembench (<https://chembench.mml.unc.edu/>) online QSAR modeling system. He described the QSAR modeling workflow that they have settled upon and highlighted key aspects of the workflow (including data cleaning and model validation) that are well defined in the Chembench suite of tools and will be made accessible via the BARD plug-in.

Tudor Oprea then presented a case study in which external data were highly curated and annotated with annotations of targets and other descriptors from the BioAssay Ontology (BAO) (<http://bioassayontology.org/>), developed at the University of Miami. DrugMatrix, which is an open-access dataset available through the National Toxicology Program at NIEHS (<https://ntp.niehs.nih.gov/drugmatrix/index.html>) was originally downloaded from ChEMBL (<https://www.ebi.ac.uk/chembl/>) in December 2012. This dataset required significant manual curation: assay details from the Eurofins Panlabs (<https://www.eurofinspanlabs.com/Panlabs>) Assay Catalog

needed to be matched with DrugMatrix data on record; targets needed further data mining (e.g., species, exact target annotation); and substrate/reference compound information needed completion for each biochemical and pharmacological screen. For example, two receptors "Imidazole I2" and "Sigma 2" and one enzyme "phorbol ester" needed re-mapping, while the exact chemical structure for 11 compounds remains undetermined; and a total of 37 targets required additional curation. Comparison attempts between DrugMatrix and another matrix-style dataset (CEREP Bioprint, <http://www.cerep.fr>) illustrate why BARD needs assay ontologies: although a number of target - chemical pairs (e.g., for target UniProt ID) can be identified, numerical bioactivity value comparisons remain meaningless in the absence of assay similarity information (e.g., agonist vs. antagonist, radio-ligand binding vs. functional assay, etc.). Establishment of a standardized research data format (RDF) as implemented in BARD to provide contextual information across assays using language familiar to research scientists and linking back to established ontologies (e.g. BAO) offers a potential platform for providing a formal assay similarity definition.

Eric Dawson (Vanderbilt University), a key outreach coordinator for the BARD project, described the way in which active engagement of end users (medicinal chemists and biologists) in BARD development has enhanced requirements-gathering and user-interface elements. Emphasis was placed on the collaborative nature of MLP Centers working together to bring the research data management (RDM, Broad Institute), application programming interface (API), and database warehouse architecture (National Chemical Genomics Center, NCGC) all together while simultaneously coordinating with an engaged user base of experienced scientists from participating Centers that leverages industrial backgrounds from a current perspective working in academia. The development of a potential local, private installation of BARD behind an organization's firewall was also described with targets for deployment at Vanderbilt's High Throughput Screening (HTS) Center core and St. Jude's Children's Hospital (Kip Guy laboratory). Dawson articulated that such a version of the database and tools would promote novel development of intellectual property and seed new collaborations between academic medical centers and the pharmaceutical industry.

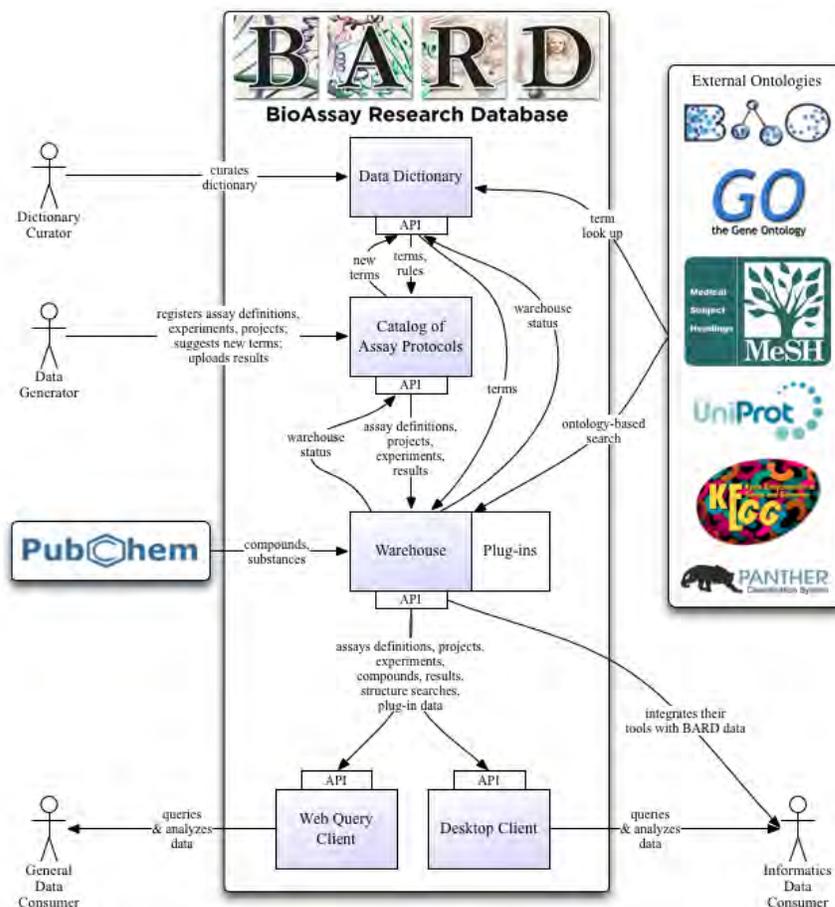
Jeremy Yang (University of New Mexico) presented the BADAPPLE (BioActivity Data Associative Promiscuity Pattern Learning Engine) plug-in for BARD, which is an evidence-based estimator of scaffold promiscuity that relies on historical screening data to assign promiscuity to compound scaffolds on the basis of the performance of compounds containing those scaffolds. Importantly, as the first plug-in written for BARD, BADAPPLE provides a pathway to be emulated by future potential plug-in developers. The BADAPPLE algorithm generates a score based on scaffold-family membership, which is derived solely from empirical BARD activity data. This score reflects both a pan-assay "batting average," as well as weighted evidence, with high scores indicative of highly promiscuous patterns. The score is "evidence-based," meaning that the algorithm evaluates data "as is," and score values are subject to change as new information becomes available. The BARD annotations and bioassay ontology enable improvements, extensions, and customizations for BADAPPLE. Somewhat surprisingly, 1.4% of the scaffolds (i.e., 1,979 scaffolds out of over 146,000, extracted from nearly 374,000 compounds) capture 50% of the bioactivity observed in 528 assays (over 30 million bioactivity observations or wells). BADAPPLE is available both as a BARD plugin, and as a web-based tool (<http://pasilla.health.unm.edu/tomcat/badapple/badapple>), and can be used to identify suspicious screening results.

Following the formal presentations, Paul Clemons (Broad Institute) walked through recent demonstration screenshots of BARD web-client development, highlighting features that will be available on BARD's public release later this fall. BARD web query provides a simple search with auto-complete that guides users toward controlled vocabulary terms and yields tabbed search results for Projects, Assays, and Compounds. Facet-based browsing allows rapid filtration of results based on additional controlled vocabulary terms. Projects and Assays can be navigated to the level of individual compound results, and search results can be saved to a Query Cart for further analyses, including Molecular Spreadsheet views and linked hierarchy visualizations that permit rapid assessment of compound performance across target classes, phenotypes or assay types.

Following the presentations, the organizers and speakers formed a panel to engage the audience in Q&A and discussion. Much discussion was directed at how to sustain BARD as a community resource into the future, both from the standpoint of continued funding of the project beyond its initial two-year timeframe, and in terms of community adoption of BARD as a useful tool that will promote deposit of non-MLP data to BARD in the future.

Paul Clemons, Eric Dawson, Rajarshi Guha, Tudor Oprea, Symposium Organizers and Participants

Slide presentations are at https://bard.nih.gov/bardtext_acsSlides.html



BARD architecture https://bard.nih.gov/bardtext_architecture.html

Exploring the Role and Value of Social Networking in Advancing the Chemical Sciences

“Role and Value of Social Networking in Advancing the Chemical Sciences” was a well-attended full day symposium on the CINF track on Monday, September 9, 2013 at the ACS Indianapolis Meeting. The symposium was subdivided into four areas with invited speakers specifically addressing the following topics: Social Media for the Individual Scientist, Social Media to Support Education, Social Media to Share Science with the Community, and Social Media to Share Chemical Information. This allowed for participants, both in the room and online, to achieve the goal of the symposium which was to “review how these [social media] tools are presently being used and what the opportunities are for the future for improved engagement with the existing systems or the development of new and improved tools.”

I had the distinct honor of co-organizing this symposium alongside CINF Chair, Antony Williams. It is amusing that Tony and I had the opportunity to work together on this “social media in science” symposium since we met via social media: Twitter to be specific. He attended the [ACS Denver Tweetup](#) in 2011 that I organized. We connected as co-organizers for this CINF session because my astute ACS Division of Small Chemical Businesses Program Chair, Joe Sabol, spotted Tony’s [blog post](#) on LinkedIn announcing this session. Joe immediately alerted me (using the old fashioned phone call which is still sometimes the quickest) about this since I was in the process of organizing a symposium titled “Small Businesses Grow by Using Social Media” on the SCHB track for the Indianapolis meeting. One phone call (again using the spoken word communication) between Tony and I determined that we should merge two endeavors and co-organize a symposium under the CINF banner with SCHB as a co-sponsor. Tony and I worked together to publicize the session via our respective social media vehicles. Two weeks prior to the meeting the symposium was mentioned in a tweet by Egon Willighagen commenting positively on the utilization of Twitter handles for our morning and afternoon session announcements. Morning session speaker David Wild, whose talk centered around cheminformatics, wikis and Google forums, suggested a session hashtag [#smchem](#) in addition to the conference hashtag [#ACSindy](#) for use by both those live tweeting at the session and those following along on Twitter. The additional hashtag filter turned out to be critical due to the alacrity of the [#ACSindy](#) feed.

Tony kicked off the morning session with his “Personal experiences in participating in the expanding social networks for science” with an emphasis on how he got started with social media, including the re-branding efforts with his social media self: from [@ChemSpiderman](#) to [@ChemConnector](#). Personal branding continued to be a theme throughout the morning session as I spoke at length on myself and my father/business partner as the PID brand: [@pidgirl](#) and [@pidguy](#) respectively. Tony also illustrated his use of altmetrics from Plum Analytics which served nicely as an introduction to Andrea Michalek’s talk on how altmetrics is gaining momentum and an overview of the functionality of their PlumX tool. ACS Network gurus Chris McCarthy and Christine Brennan-Schmidt spoke about using tools collaboratively to communicate and advance science. Bob Belford shared with us the twenty year story from ChemConf to ConfChem. The morning session was closed by “Grace Baysinger, chem librarian extraordinaire at Stanford” as Donna W [@CaltechChemLib](#) tweeted during the session. [@ChemConnector](#) chimed in with his tweet “Grace Baysinger talks about [#XCITR](#). Originally hosted by Fiz Chemie now hosted by Royal Society of Chemistry <http://www.xcitr.org/> [#smchem](#).”

For the afternoon session we had invited Carmen Drahl, Senior Editor of *C&EN*, to moderate our panel discussion and, additionally, Carmen agreed to live tweet from our session as part of her [#WHERE SCARMEN](#) crowdsourcing experiment, which, of course, garnered more Twitter attention for our symposium. This can correlate to increased attendance in the session at the conferences.

Bibiana Campos-Seijo, Editor of *Chemistry World*, opened the afternoon session emphasizing that “You cannot ignore the power of social media” as tweeted by Carmen Drahl. George Ruger from the ACS Mid-Hudson Local Section addressed the ways in which social media can be used to communicate science to the community. Evan Bolton concluded the talks with his experienced tale of social networking and PubChem. Next, the session speakers Andrea, George and Evan were joined by Joe Sabol, SCHB Program Chair, and Mark Jones, Communications Fellow for Dow Chemical, for the panel discussion moderated by Carmen Drahl. Carmen asked the panelists the following questions:

- How do you advise a colleague who tells you they have an interest in social media, but has no idea how to use it or where to start?
- What social tools do you think chemists have not explored enough yet?
- Are there any social media tools that you feel are over-utilized?

The discussions proceeded in the fast pace and brought the apparent interest from the audience: all eyes on the panelists’ responses (less live tweeting for those not in attendance). It would have been better if the panel discussion was videotaped.

The day concluded with a short interactive workshop organized by Antony Williams, Teri Vogel and Andrea Michalek (the facilitators were assembled by using crowdsourcing via Tony’s call-to-action power of blogging) with the intended agenda to discuss online forums, public profile tools, altmetrics, reference managers and collaborative platforms.

There was definitely in-room live-tweeting going on throughout the day. Tony was stationed in the back of the room and I was on the stage presiding, timing the speakers, and tweeting simultaneously. Based on feedback from colleagues, next time I will attempt presiding and tweeting from the front row rather than the stage, in order to avoid a distraction to the audience members in being engaged with the speakers. Tony and I are collaborating again in organizing a symposium (in four half day sessions) to explore “Evolving Nature of Scholarly Publishing: Connecting Scholars to Each Other and to Society” at Pacificchem. If a trip to Hawaii in December 2015 sounds good to you, I hope you will join us on our next “social media in science” quest.



Jennifer Maclachlan, Carmen Drahl, Antony Williams

Jennifer Maclachlan, Symposium Co-Organizer

Science-Based Policy Development in the Environment, Food, Health, and Transport Sectors

This one-day symposium explored the interaction between science and policy development in the regulation of the environment, food, health and transport. It consisted of a series of case studies illustrating the impact of science on policy development. The controversy surrounding the science behind the study of global warming and the resulting focus on the reduction of carbon dioxide emissions by international agreement and by national and international regulation is one example of such an area where science and policy development are inextricably intertwined. The symposium is one of a series which is seeking to identify other areas where science-based policy development is of increasing importance and was cosponsored by CINF, AGFD, ANYL, ENVR and MEDI.

The first speaker in the half-day session was Thomas A. Duster, who spoke about “Adaptive management tools for engineered nanomaterials in municipal wastewater effluents.” Engineered nanomaterials in consumer products are everywhere and result in delivery to municipal wastewater treatment systems where they may be subsequently discharged to the environment. At sufficient concentrations, many common nanomaterials, including titanium dioxide nanoparticles and carbon nanotubes, are toxic or disruptive to aquatic organisms. Application of contemporary environmental policies poses significant challenges when trying to mitigate these potential impacts. For example, the traditional standards-to-permits approach of the Clean Water Act (CWA), which applies to most wastewater treatment plant effluents in the United States, typically involves the development of contaminant-specific water quality criteria. However, existing research regarding the detection, fate, and toxicology of nanomaterials is still in its infancy and rapidly changing, thereby limiting the ability of policymakers to justify and establish static effluent discharge standards for these emerging contaminants.

Thomas described an adaptive nanomaterials management approach that strives to bridge the gap between significant scientific uncertainties and an ostensive need for some type of policy structure. At the core of this adaptive management procedure is a robust mechanism for information and data organization, which is programmed to alert policymakers of convergence in the literature among: (a) observed and/or anticipated concentrations of target nanomaterials in wastewater effluents; (b) demonstrated impacts of these concentrations on aquatic organisms or ecological function; and (c) our technological capacity to reliably detect these target nanomaterial concentrations. The confluence of these factors is expected to be a significant trigger in evaluating the need for specific management actions and/or expansion of policies related to the release of engineered nanomaterials to environmental systems. Finally, Thomas described how specific elements of this approach may be applied to policy challenges for other emerging contaminants.

Our second speaker was Frederick W. Stoss, who described the “Role of STEM data and information in an environmental decision-making scenario: the case of climate change.” The 1997 Kyoto Protocol to the United Nations Framework Convention on Climate Change (FCCC) established agreements for reducing greenhouse gas (GHG) emissions. Every national academy of science states that anthropogenic sources of GHGs, caused by human activities, impact the Earth’s climate. However, “climate deniers” claim there is no scientific basis for climate change and that it is a well orchestrated hoax. So contentious were these allegations that computers of the Climatic Research Unit at the University of East Anglia were “hacked” and email messages and reports became “evidence” of this “scientific hoax.” Results included disruptions of FCCC policy negotiations and erosion of public confidence in the science of climate change. In his presentation, Fred investigated the growth of climate information, defined different levels of understanding of and access to information, provided a context by which information is generated, and presented a model demonstrating the role of scientific data and information in an environmental decision-making model.

The third speaker was Helena Hogberg, who presented work on the “Identification of pathways of toxicity to predict human effects” which she coauthored with Thomas Hartung. The 2007 National Research Council report “Toxicity Testing in the 21st Century: a vision and a strategy” has created an atmosphere for change in the U.S. It suggested moving away from traditional (animal) testing to modern technologies based on pathways of toxicity. These toxicity pathways could be modeled in relatively simple cell tests. The NIH is funding, by a transformative research grant, The Human Toxome project led by Center for Alternatives to Animal Testing. The project also involves U.S. EPA ToxCast, Hamner Institute, Agilent and members of the Tox-21c panel. The goal is to develop a public database of pathways, the Human Toxome, to enable scientific collaboration and exchange.

An area of toxicology where Tox-21c could have significant impact is developmental neurotoxicity (DNT). Current animal tests for DNT have several limitations, including high costs (\$1.4 million per substance), and require substantial time. In addition, there are scientific concerns regarding the relevance of these studies for human health effects. Consequently, only few substances have been identified as developmental neurotoxicants. This is a concern as evidence shows that exposures to environmental chemicals contribute to the increasing incidence of neuro-developmental disorders in children. Moving towards a mechanistic science could help identify the perturbed pathways that are likely to lead to these adverse effects. DNTox-21c is a CAAT project funded by FDA that is aiming to identify pathways of developmental neurotoxicity using a metabolomics approach.

Beside the technical development of new approaches, a case was made that we need both conceptual steering and an objective assessment of current practices by evidence-based toxicology. Applying an approach modeled on Evidence-based Medicine (EBM) was suggested, which over the last two decades has demonstrated that rigorous systematic reviews of current practices of studies provides powerful tools to provide health care professionals and patients with the current best scientific evidence for diagnostic and treatment options.

The first speaker after the intermission was Rodger Curren, who addressed the topic of the “Role of education and training in supporting science-based policy development” that was co-authored with Hans Raabe and Brian Jones. Policy changes, especially in the regulatory requirements for the safety of new products, are often impeded because decision makers in national regulatory bodies are unaware of the science supporting new methodologies. This is not entirely unexpected since such individuals may be more exposed to political concerns on a daily basis than scientific ones. A current example is the area of non-animal methods for toxicity testing where significant international differences in acceptance exist. Europe and the U.S., for example, are quickly moving to using human-derived cells and tissues rather than whole animal based models. Other countries, such as China, may be reluctant to make a change because their scientists have not had sufficient time to develop sound databases of information. The authors have found that providing specific hands-on training and education on standard methods directly to regulators and scientists in these countries has significantly improved the recognition and acceptance of new approaches.

The next speaker was Julie Jones, who highlighted “Policy divergence in the absence of science: The case of e-cigarettes,” a presentation co-authored by David Lawson. Over the past five years electronic cigarettes (e-cigarettes) have emerged as a new consumer product that is being used by an increasing number of smokers who are seeking less risky alternatives to conventional cigarettes. E-cigarettes tend to be designed to look and feel similar to conventional cigarettes, but they do not contain tobacco. They are battery-powered devices that produce an aerosol usually containing nicotine. Currently, there is significant inconsistency in the way that e-cigarettes are being regulated:

e-cigarettes are banned in some countries or are being regulated either as medicinal, tobacco or general consumer products in others. There is also a diversity of views regarding the potential role that e-cigarettes could play in helping to reduce the public health impacts of tobacco use. In fact, the science to support this emerging category of products is still under development, and there are many gaps. E-cigarettes represent a timely case study on policy development for regulation of a new product category in the absence of a solid scientific foundation. Julie presented her views on how the development of such a scientific foundation might be accelerated to help inform development of an appropriate regulatory framework for e-cigarettes.

In a related paper, Christopher J. Proctor discussed the “Role of regulatory science in reducing the public health impact of tobacco use,” co-authored by Chuan Liu. The U.S. FDA, through the 2009 U.S. Family Smoking and Prevention Tobacco Control Act, is introducing a variety of regulations aimed at reducing the public health impact of tobacco use. These include considering the levels of harmful and potentially harmful constituents of tobacco products and regulations governing modified risk tobacco products. The FDA has set out a series of research questions that it believes are needed to underpin its regulatory proposals and has initiated a large research funding program, in association with NIH. Other scientific advisory groups, including the World Health Organization’s Scientific Advisory Committee on Tobacco Product Regulation have also listed research needed to assist the development of science-based public policy on tobacco. Christopher summarized the research questions being framed by regulators as related to product regulation, and provided some views on how the development of regulatory science in tobacco might be accelerated.

The final speaker, David Richardson, described “Systematic and structural risk analysis approaches for establishing maximum levels of essential nutrients and other bioactive substances in fortified foods and food supplements.” Nutritional risk analysis addresses the essential nutrients and other substances with nutritional and physiological effects and the risk to health from their inadequate and/or excessive intake. David reviewed the principles of risk management in order to underpin regulatory developments around the world to establish maximum amounts of vitamins and minerals and other substances in fortified foods and food supplements. Proposed science-based risk management models for public health decision-making take into account international risk assessments and (1) the tolerable upper intake levels (ULs) for vitamins and minerals, (2) the highest observed intakes (HOIs) for bioactive substances for which no adverse effects have been identified, and (3) the contributions to total intake from conventional foods, fortified foods and food supplements. These models propose the allocation of nutrient substances into three categories of risk and maximum levels in order to protect consumers, both adults and children, from excessive intakes.

William Town, Symposium Organizer

Herman Skolnik Award Symposium 2013

Honoring Richard D. (Dick) Cramer

Introduction

Dick Cramer is best known as the inventor of the technique of Comparative Molecular Field Analysis (CoMFA) and its introduction to the molecular and drug design fields. Early in his career, in the research group of E.J. Corey, Dick was involved with the first artificial intelligence methods to predict chemical synthesis, coining the acronym “LHASA” (Logic and Heuristics Applied to Synthetic Analysis) for the project. Dick has remained active in research and publishing at the forefront of his field. His work on “topomeric” descriptors, which allows CoMFA without tedious alignment of ligands, is proving a very successful tool in drug discovery. He currently serves as Senior Vice President, Science, and Chief Scientific Officer for Tripos, a Certara Company. Dick has also made major contributions to another entirely different field: baseball. He became interested in applying computers to baseball statistics and developed a program to feed detailed baseball statistics into the commentators’ box. He consulted with a number of major league teams, and is featured in the book *Moneyball* by Michael Lewis (recently made into a major motion picture). The award symposium covered all Dick’s fields of endeavor.

CoMFA

Since Dick is best known as the inventor CoMFA¹, it was fitting that the opening talk, by Bob Clark of Simulations Plus (bob@simulations-plus.com), outlined the history of CoMFA, citing eight articles in which he himself was a co-author.¹⁻⁸ CoMFA required the identification of the “bioactive conformer” and this was difficult in the era of combinatorial chemistry, so Dick and other colleagues came up with topomers and rules for conformer alignment, while Bob concentrated on traditional CoMFA.⁹⁻¹⁷

“The alignment problem” has many dimensions. One is aligning ligands to themselves (conformation), i.e., studying the relationships between substructures within an individual molecule. Another is aligning ligands to each other (“alignment”), i.e., studying the relationships between substructures in *different* ligand molecules. Finding an appropriate protein conformation and aligning the protein to the ligands are further dimensions. Bob favors the ligands’-eye view of protein binding¹⁵ over the protein’s-eye view; given a basic pose obtained by docking or pharmacophore alignment, he likes to refine the alignment based on common substructures in the ligands and see how the protein adjusts to accommodate ligand variation.

At the spring 1998 ACS meeting, Bob spoke about making 3D QSAR both simple and robust. The literature background included seminal CoMFA publications,^{1,18} papers on region selection methods,^{19,20} and articles on descriptor transforms.^{21,22} At that time there were concerns about “out of the box” CoMFA: the sensitivity of q^2 to changes in conformation and lattice alignment, and reproducibility from published applications. Approaches to dealing with the variability included avoiding alignment and grids altogether; better tempered molecular fields; “preventive medicines” such as inertial template alignment (which is somewhat related to topomers) and simple modified grid designs; and region focusing (weighting).

Adding grid points can reduce aliasing for unsampled field points, while removing grid points optimizes covariance between grid points. Bob discussed how to strike a reasonable balance using

anisotropic spacing, and a face-centered cubic lattice to make CoMFA much less sensitive to alignment. He presented some plots of the effect of rotation on q^2 for the different lattices, and concluded that the sensitivity to positioning was less for the face-centered cubic grid, and the average performance was better as well.

Yvonne Martin (yvonnecmartin@comcast.net) presented a different perspective on the history of CoMFA. A molecule can be represented in 3D using shape, or electrostatic potential on a van der Waals surface or quantum chemical regions of high and low electron density, for example, but how do you convert these lovely 3D colored images into relevant descriptors for 3D QSAR? This is the problem that Dick started to address while he was working at Smith Kline & French (SK&F). He and Margaret Wise²³ described molecules by coarse steric and electrostatic energy maps calculated from the Boltzmann-weighted sum of the conformers of the compound. They derived descriptors using principal components analysis of the fields of the various molecules. The use of partial least squares (PLS) in solving underdetermined matrices was instrumental in helping Dick develop CoMFA. Svante Wold suggested this solution to Dick at the 1981 QSAR Conference. Most QSAR practitioners at that time did not know about PLS²⁴ or understand its power. CoMFA^{1,25} was a descendent of DYLOMMS²³ combined with PLS, after Dick had left SK&F and associated with Garland Marshall, who had just founded Tripos. Dick's insight into the choice of fields for CoMFA is validated²⁶ by the observation that it well describes the traditional linear free energy descriptions, Hammett sigma constant and Taft E_s values.

Yvonne listed some key elements leading to innovation, each of which contributed to Dick's success. The four factors are: recognition that there is a problem, persistence in searching for a solution, creativity and insight in the search for a solution, and chance. The program GRID²⁷ is one example of an innovation, but its author Peter Goodford did not go on to invent CoMFA. He was aware of QSAR and the use of statistics in QSAR, but he did not focus on the problem of correlating the 3D properties of ligands with their biological potency. He failed to recognize the problem.

Yvonne's own team also missed the opportunity to invent CoMFA. Abbott had tested some compounds for diuretic activity and explained the progressive decrease in potency of these compounds as they occupy more and more new space compared to the most potent compounds. Extrapolating from the linear free energy relationship (LFER) explanation that the Taft E_s values are a function of the radius of the atom, her team wrote a program that generated 96 descriptors of shape as the length of vectors emanating from the first moment of inertia of the aligned molecules, and used statistics to derive the QSAR, but they did not find a good relationship with this dataset or others. What they missed was the correct description of molecules. Because they relied too much on the traditions from LFER, they failed on creativity and insight in the search for a solution.

Corwin Hansch's work leading to the invention of QSAR^{28,29} started in 1948 with his collaboration with Robert Muir, a botanist who happened to have an office in the chemistry building. Hansch and Muir emphasized the Hammett sigma constant in their work on plant growth regulators. After a decade of struggling with the Hammett relationship, Hansch decided to investigate a possible relationship to partitioning into the cell. He found precedents in the work of Runar Collander,³⁰ and others. At this point, he hired Toshio Fujita: the second bit of luck (after the chance of meeting Muir) that led to QSAR. Neither the Hammett constant nor $\log P$ describes the SAR, but Fujita suggested that perhaps

both properties contribute to the SAR. He also recognized the additive nature of $\log P$. Hansch suggested a parabolic function in $\log P$ to account for an optimum value. There now was the problem of how to fit the data to the proposed equation. Fortunately, there was a faculty member of the geology department, Donald McIntyre, who was fascinated by the possible influence of computers on research. He not only convinced a donor to give a computer to the chemistry department, but he also coded up the multiple regression equation. Chance was, however not the only factor in the invention of QSAR: there were 15 years of persistence behind the innovation.

Yvonne discussed a few examples of prominent scientists who could have invented QSAR, but did not. The Fieser group had evidence for the additive and constitutive nature of lipophilicity but they seemed to be unaware of earlier work on partitioning. What the Fieser group did not do was to recognize that there is a general problem in structure-activity relationships, that calculating lipophilicity would be a valuable exercise, and that multiple factors might contribute to potency. Brodie and Schanker studied drug absorption in 1960 but missed inventing QSAR mainly because they did not realize the general nature of the problem, because they did not know about Collander's work on octanol, and because they did not follow the LFER field, but especially they failed because they did not think to apply statistics to their relationships. So they failed on both persistence and insight.

Another example of the role of chance in innovation comes from Yvonne's own team.³¹ They knew that they could not do CoMFA unless they knew how to choose conformations and how to align a diverse set of molecules. The only literature solutions required choosing the atoms to match. By chance, Yvonne read a paper by Brint and Willett³² and realized that a pharmacophore is just a 3D maximal common substructure, but one in which the points are not atoms, but pharmacophore features. Chance, rather than persistence was the innovation factor here. Two other groups^{33,34} worked on the alignment problem. Both provided means to select corresponding conformations, but as input they required the atoms or features that correspond in the various molecules, and this is not always obvious. They ignored part of the problem: recognition was the failure point in this case.

In conclusion, the fact that invention requires so many elements to coalesce does not negate the powerful role of persistent focus on attempting various solutions to the problem.

More on QSAR

The next speaker should have been Tony Hopfinger (hopfingr@gmail.com), but on the morning of the symposium he was taken ill. I had a copy of his slides and Dick Cramer valiantly attempted to present the paper in Tony's absence. Clearly anything I write in this article will be a poor reflection of what Tony might have said had he been there in person.

Tony worked with Dick, while Dick was at SK&F, to provide the structure generator eventually commercialized as ChemLab.³⁵ The two of them had an argument at an ACS meeting in Houston, Texas before the first CoMFA publication appeared. The issues were field *versus* overlap volume descriptors, conformation and alignment. Tony and Dick agreed to continue to disagree. Dick went on to gain fame from fields and CoMFA. Tony went on to develop Molecular Shape Analysis³⁶ and found it to be a dead-end, but then, in an epiphany, 4D-QSAR analysis was born.³⁷ The fourth "dimension" in the paradigm is *sampling* and includes the sampling of conformation, alignment, pharmacophore sites and entropy. The composite information coming from each of these sampled property sets is embedded in the resulting QSAR model.

The descriptors in 4D-QSAR analysis are the grid cell (spatial) occupancy measures of the atoms composing each molecule in the training set realized from the sampling of conformation and alignment spaces. A single “active” conformation can be postulated for each compound in the training set and combined with the optimal alignment for use in other molecular design applications including other 3D-QSAR methods. The influence of the conformational entropy of each compound on its activity can be estimated. Serial use of PLS, regression and a genetic algorithm (GA) is used to perform data reduction and identify the manifold of top 3D-QSAR models for a training set. The unique manifold of 3D-QSAR models is arrived at by computing the extent of orthogonality in the residuals of error among the most significant 3D-QSAR models in the general GA population. The models can be graphically represented by plotting the significant 3D-QSAR grid cells in space along with their descriptor attributes.

4D-QSAR is used to create and screen against 3D-pharmacophore QSAR models and can be used in receptor-independent or receptor-dependent modes. More recently Tony introduced a pseudo structure-based method, Membrane-Interaction QSAR analysis,^{38,39} to estimate a wide range of ADME and toxicity endpoints based on interaction of test compounds with models of cellular membranes and a set of unique property descriptors.

The n-dimensional QSAR themes used in Tony’ slides were conformation, alignment, spatial descriptors, the pharmacophore, whether or not to include the receptor, and what to do with conflicting or weird results. He reckons that he is now in a position to identify, probe, and think meaningfully about, but not perhaps solve, how to handle the many obstacles that have long plagued nD-QSAR analysis. Examples with respect to conformation are:

- How to completely explore conformations (MD, MC, or brute-force).
- How to handle receptor-independent and receptor-dependent searches.
- How to set limits on upper energies of ligand conformations and ligand receptor complexes.
- How to model large geometric changes in receptor geometry.

He wondered whether we should we let X-ray and NMR do the “heavy lifting” and let modeling come in for clean-up and refinement.

It is clear that there are still differences of opinion between Dick and Tony, but Tony concluded with a very fitting tribute to Dick as a colleague. This was in the form of the last two lines of a poem by William Butler Yeats, somewhat paraphrased: *“When I think where man's glory most begins and ends, I say my glory is to have such a good and questioning friend.”*

Ajay Jain (ajain@jainlab.org) of the University of California San Francisco has developed a family of 3D QSAR and docking approaches. Ajay showed Figure 6 of Dick’s much-cited CoMFA paper.¹ It shows the major steric features of the QSAR for steroid binding to testosterone-binding globulin (TeBG). In this work Dick illuminated a new and exciting path for our field: his model predicted the right thing for the right reasons. Ajay’s initial work with Compass^{40,41} created a linkage between model and molecular pose. Compass involved a new representational scheme for capturing the 3D surface properties of small molecules that made it possible to address systematically the choice of the relative alignment and conformation (or pose) of competitive ligands including the detailed relationship of their hydrophobic shapes. A key insight was that the choice of pose should be directly governed by the

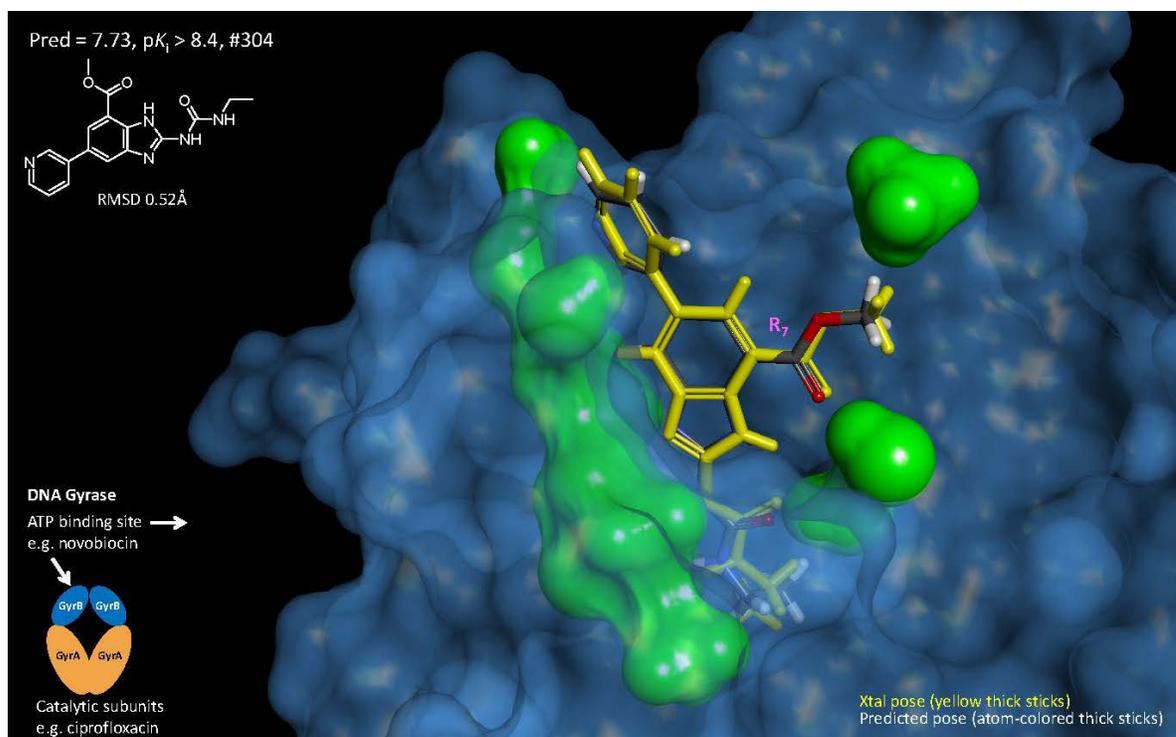
function being used to predict binding affinity (essentially a direct analogy to physics where the lowest energy state is sought). The difficulty was that the function to predict activity was being induced at the same time as the pose choice. The Compass method overcame this problem, and was one of the foundational methods in establishing the field of multiple-instance learning.

Ajay showed a model of dihydrotestosterone (1D2S) binding to TeBG. If you make a small change to the steroid, to 1LHO, the alignment shifts a little. If you use estradiol (ILHV) the alignment flips. The protein moves, too. These bidirectional relationships must be modeled in 3D QSAR. Because substituent modifications affect molecular pose, effects on activity will often be non-additive. Jain believes that it is vital to address the basic physical realities of protein-ligand binding.

For QSAR as physical modeling, there must be a direct linkage between the model and molecular pose: if the model changes, the poses will as well; if substituents are changed, alignments will as well. Details of molecular shape and electrostatic properties have to matter to the model; non-additive behavior should be a natural consequence; and the models should have a direct relationship to physical protein binding pockets.

The QMOD approach⁴² takes QSAR to a new level, by transforming the problem into one of molecular docking. A protein binding site is induced given SAR data using the multiple-instance machine learning paradigm developed for Compass. A skin is built around a small molecule pose, inducing a binding pocket that explains the data, so that you can predict the activity and geometry of new ligands. Model construction is fully automated. The agnostic Surflex-QMOD hypothesis for TeBG cares about the surfaces, not the atoms.

Ajay's student, Rocco Varela, has applied QMOD to 426 Vertex gyrase inhibitors.⁴³ He performed an iterative, temporal lead optimization exercise. A series of gyrase inhibitors with known synthetic order formed the set of molecules that could be selected for "synthesis." Beginning with a small number of molecules, based only on structures and activities, a model was constructed. Compound selection was done computationally, each time making five selections based on confident predictions of high activity and five selections based on a quantitative measure of three-dimensional structural novelty. Compound selection was followed by model refinement using the new data. Iterative computational candidate selection produced rapid improvements in selected compound activity, and incorporation of explicitly novel compounds uncovered much more diverse active inhibitors than strategies lacking active novelty selection. One of Rocco's models was chosen for the cover of the *Journal of Medicinal Chemistry*.



The pocket model actually looks like the experimentally determined gyrase pocket. QMOD is predicting the right thing for the right reason, just as Dick's CoMFA model predicted the right thing for the right reasons in 1988.

Drug Discovery

Bobby Glen of the University of Cambridge (rcg28@cam.ac.uk) moved on from adventures in "CoMFA-land" to adventures in drug discovery. He started by contrasting computation with reality. We do a calculation, but we do not know the correctness of our prediction until we see the results of the experiment, and the "experiment" may be the patient who takes the drug. Our objective is to mimic the real world of the patient as closely as possible, but describing molecules is difficult⁴⁴ so we make approximations, but when we do our calculations we need to think about what happens in the real world. We are interested in the properties of molecules, not so much in what they *are*, but in what they *do*.⁴⁵ In the real world a drug does lots of things, especially to sick people. A drug tested in a 25-year old male Olympic rower will have very different effects on a 67-year female patient with multiple chronic conditions.⁴⁶ Compounds show different physiological effects for many different reasons and the multiple mechanisms are hard to model in a single structure-activity relationship. Toxicity is often unexpected and is discovered in the clinic.

Computational methods are evolving to address the complexity of the process; the nature of drug discovery is now multivariate, and more and more data are becoming available. It is now possible to construct bioprints of molecules and their effects on multiple receptor systems. We can also introduce the effects of other biological systems such as transport and metabolism.^{47,48}

Bobby's team has developed MetaPrint2D software (<http://www-metaprint2d.ch.cam.ac.uk/metaprint2d>) to predict the sites and products of metabolism.^{49,50} In an

example, Bobby input the SMILES for a partial agonist which has a main metabolite that is a full agonist. So, as the drug concentration lowers in blood, the remaining compound becomes more potent. In another example, the toxicity of acetaminophen (paracetamol) is predicted and the two relevant metabolic pathways are displayed. The primary pathway is glucuronidation which yields a relatively non-toxic metabolite, but at higher doses this pathway is saturated, and N-acetyl-p-benzoquinone is produced, causing liver damage.

A drug's activity can be modified by metabolism. Bobby showed the predicted metabolic pathways of promazine, and the predicted activities of some metabolites. It is also possible to predict "in reverse" and identify prodrugs. Bobby showed some biological effects of a promazine metabolite: effects which may possibly relate to phenotypic changes. The terminal metabolite thiodiphenylamine was predicted to be active against amine oxidase, cyclooxygenase 1 and 2, and the sodium-dependent noradrenaline transporter.

Bobby concluded that drug discovery is developing holistic tendencies, driven by access to "Big Data," faster processing and, most of all, more complete algorithms, and, of course, more experimental validation. He paid tribute to Dick for being at the forefront of this revolution: he was one of the first to use multivariate data in CoMFA, and before that he was using multi-dimensional property visualization.

Tudor Oprea (toprea@salud.unm.edu) acknowledged that CoMFA had a huge influence in his own career. The lesson he learned from Dick Cramer and Dave Patterson was "If you can't be right, be consistent." In his talk entitled "Think Local, Act Global" he said that in a Newtonian Universe it would be possible to predict the future, but we do not live in one. In chemical space, as in geography, maps need to be consistent. Tudor's chemical global positioning system, ChemGPS, makes a drugspace map by systematically applying conventions when examining chemical space, in a manner similar to the Mercator convention in geography. Chemography is the art of navigating in chemical space.^{51,52} Rules are equivalent to dimensions (e.g., longitude and latitude), while structures are equivalent to objects (e.g., cities and countries). Selected rules include size, lipophilicity, polarizability, charge, flexibility, rigidity, and hydrogen bond capacity. Core structures include most marketed drugs with good oral permeability, as well as other biologically active compounds, while "satellites" are intentionally placed outside the chemical space of drugs, and include molecules having extreme values in one or more of the dimensions of interest. The map coordinates are *t*-scores extracted by principal component analysis (PCA) from 72 descriptors that evaluate the rules on a total set of 423 satellite and core structures. The PCA method, and ChemGPS, were inspired by Dick Cramer's BC(DEF) work.⁵³

By successfully combining virtual and biomolecular screening, Tudor's team at the University of New Mexico discovered G-1, the first GPR30-specific agonist, capable of activating GPR30 in a complex environment of classical and new estrogen receptors.⁵⁴ They used a composite approach. 2D fingerprint technologies are really fast, but they can lead you into a local trap: if you use a steroid as a query, the high-similarity hits will almost all be steroids. 3D technologies are not as fast as 2D, and require a choice of conformers, but if you submit a rigid steroid as query, the chances are that you will find fewer steroids. 3D approaches include the ROCS shape-based method (<http://www.eyesopen.com/rocs>) and ALMOND (http://www.moldiscovery.com/soft_almond.php)

based on pharmacophores. Cristian Bologa used a weighting scheme of 40% 2D (MDL and Daylight fingerprints), 40% shape, and 20% ALMOND with the intention of screening the top 100 hits, analyzing the primary hits and then fine-tuning the weighting scheme. In practice he got lucky: hits number 58 and 65 bound to ER α and ER β , hit number 95 was G-1, and the hits were active in secondary assays. The team went on to identify a potent GPR30 antagonist.⁵⁵

Data reliability can be a problem, but often goes unrecognized. In one previously used⁵⁶ dataset for human intestinal absorption (HIA), sulfasalazine was wrong because bacterial azo bond reduction occurs in the intestine and the measured HIA value was that of a metabolite. After removing two azo-containing drugs, as well as two drugs absorbed by paracellular mechanism, the bottom end of the sigmoidal curve describing Caco-2 absorption was removed, with little or no sigmoidal effect left.

The Biopharmaceutics Drug Disposition Classification System (BDDCS)⁵⁷ has four categories: class 1 high solubility and extensive metabolism, class 2 low solubility and extensive metabolism, class 3 high solubility and poor metabolism, and class 4 low solubility and poor metabolism. Tudor and his colleagues have compiled the BDDCS classification for 927 drugs.⁵⁸ They have also reported a computational procedure for predicting BDDCS class from molecular structures.⁵⁹ Transporter effects in the intestine and the liver are not clinically relevant for BDDCS class 1 drugs, but potentially can have a high impact for BDDCS class 2 (efflux in the gut, and efflux and uptake in the liver) and class 3 (uptake and efflux in both gut and liver) drugs. A combination of high dose and low solubility is likely to cause BDDCS class 4 to be under-populated in terms of approved drugs.⁵⁹ The model reported by Tudor and co-workers showed highest accuracy in predicting classes 2 and 3 with respect to the most populated class 1. For class 4 drugs a general lack of predictability was observed.

BDDCS has also been used to improve blood brain barrier predictions of oral drugs.⁶⁰ BDDCS class membership was integrated with *in vitro* P-gp efflux and *in silico* permeability data to create a classification tree that accurately predicted CNS disposition for more than 90% of 153 drugs in the dataset. Medicinal chemists are often taught that second generation antihistamines are successful due to log*P* optimization, which supposedly leads to little or no blood brain barrier (BBB) penetration. Tudor's team has shown that this is not true, since neither log*P* nor log*D* distribution differ between first generation (BBB penetrating) and second generation antihistamines.⁶¹ They compared 64 H1R antagonists the log*P* and log*D* profiles of which overlap. The nine that are *effluxed* by P-gp include all second generation antihistamines. For these, P-gp becomes, *de facto*, a drug target.

In some work with Scott Boyer, Tudor examined the CEREP BioPrint dataset (<http://www.cerep.fr/Cerep/Users/pages/ProductsServices/BioPrintServices.asp>). The total number of potential activities was 371,448, whereas the total number of observed activities was 31,264, leading to a probability of 8.41% for observing bioactivity. They defined "biased targets" as those that exceed the 8.41% probability, and noticed that biased targets account for 76.81% of all activities in the CEREP dataset. Tudor and co-workers further looked at 871 chemicals measured in 131 DrugMatrix assays (<http://ntp-server.niehs.nih.gov/?objectid=72016020-BDB7-CEBA-F3E5A7965617C1C1>) and found that biased targets account for 83.34% of the activities in DrugMatrix. Tudor's pie-charts showed only partial overlap of chemicals and targets in CEREP and DrugMatrix; data-by-data comparison has revealed several molecular-target sets for which the overlap of bioactives in CEREP

and DrugMatrix is zero, substantiating the need for accurate assay annotation and proper bioassay ontologies such as the work done by Stefan Schurer (<http://bioassayontology.org>).

If you have an assay, you have information; if you have two assays for the same target, you may have confirmation, or confusion. Do we really need big data when we often cannot handle small data? Human curation and attention to detail are needed before decision-making is well-served by experiment and model.

Sabermetrics

And so on to something completely different: David Smith (dwsmith@retrosheet.org) focused on the philosophy of science and how this relates to many different kinds of inquiry, including baseball research. Science is a procedure for study, largely independent of the topic under investigation. Nowadays this definition is being blurred in discussions of STEM disciplines, missing the point that science is special because of how questions are analyzed, not because of what is studied. Louis Pasteur said “*There are no such things as applied sciences, only applications of science.*” Economics is an example of the application of scientific methods to an important area that is not a natural science. Since we are not defining science by what is studied, we need to define it in terms of key features: definition of questions and proper criteria for evaluation. This is seen in the classical formulation of hypothesis, experiment, and conclusion.

Not all areas of traditional science fit neatly into this paradigm. Astronomy, for example, is a scientific discipline, but Copernicus depended primarily on observation, not on manipulation. Evolutionary biology, David’s own discipline, is another example. In astronomy, the Copernican proposal of a heliocentric solar system made sense of a number of phenomena at a single stroke; there is no single observation that “proves” the theory. Before Darwin, biology was almost entirely descriptive and very fragmented. Darwin’s proposal of natural selection provided the same sort of satisfying and unifying explanation that Copernicus did.

Natural history is the starting point for almost every scientific discipline: its observations eventually became organized and lent themselves to questions. At this point the study became scientific. Carl Linnaeus created classification systems for thousands of species of organisms, a feat of great organization, but little analysis. The naturalist Alexander von Humboldt explored South America 30 years before Darwin, but his observations went further than mere cataloguing. The transition from natural history to science is perhaps best seen in the person of Charles Darwin who began his voyage on HMS Beagle as a naturalist charged with collecting samples and making observations, but who after his return to England spent 20 years organizing the material he had collected. He began to ask why certain patterns existed.

Evolution is often described as a historical science and so is baseball research. Evolutionary hypotheses and predictions are not about the future, but about an unknown past. In baseball, for decades the conventional wisdom was that the best batters were those who had the highest batting average, that is, the most hits per opportunity (at bat). Detailed study of modern events led to the hypothesis that reaching base by any means was of greater significance than base hits considered alone. Furthermore, advancing runners with extra base hits was historically undervalued. These two measures: reaching base (on base average) and advancing runners (slugging percentage) were combined to a single measure called OPS (on-base plus slugging) that was then used to examine

baseball from 1901 through 2012. The results show a stronger correlation between runs per game and OPS than between runs per game and batting average. Note also that differences such as these are much easier to demonstrate when large datasets are available. Here we have a scientific result plus retrospective prediction.

The name given to this sort of work is Sabermetrics, a term based on “SABR”, the acronym of the Society for American Baseball Research (<http://sabr.org/>), a national group of some 6000 members founded in 1971. When Dick and others began working on Sabermetrics in the late 1970s, detailed data were not readily available. Bill James then started to collect them and interesting studies could be carried out. For example, a stolen base is a valuable play that increases the chance of scoring, but the counterpart, a caught stealing, has a negative effect. A study showed that a stolen base attempt must be successful in at least two thirds of cases to be worth the risk. The importance of a first pitch strike has also been studied. If that first pitch is a strike because of a swing and a miss, then the pitcher usually has a good outcome, but if the first pitch is a foul ball, then the batter does better, and if that first pitch is hit into play, the batter does extremely well. A third example is clutch hitting: the assertion that some hitters increase their performance in tight situations. Dick did a sophisticated analysis (<http://cyrilmorong.com/CramerClutch2.htm>) to show that clutch hitting is an illusion. In addition to writing analysis software, Dick founded STATS, Inc., which began by gathering data for baseball studies, but now covers other sports. Many baseball teams now use Sabermetrics and measures such as slugging are displayed on scoreboards.

David’s group, Retrosheet (<http://www.retrosheet.org>), has gathered play-by-play data for 165,000 of 185,000 games played since 1901 and has made it freely available on the Internet. Collection, digitization, and publication of such data have led to a variety of Sabermetric analyses. Dick is one of the volunteers who examines images of old scorecards and converts them to digital form using specialized software. Baseball research offers unique opportunities to ask meaningful questions in a scientifically rigorous way and this is why professional scientists such as Dick and David are attracted to it.

Synthesis planning

In recognition of Dick’s early work in the LHASA project, Todd Wipke (wipke@ucsc.edu) who published a seminal paper with Corey,⁶² addressed the subject of synthesis planning at the award symposium. A much earlier paper by Corey’s team⁶³ had a section entitled “Synthesis Plan” that discussed alternative disconnection plans: the reviewers did not like that section. In a later paper⁶⁴ Corey said “*the first task ... should be an exhaustive analysis of the topological properties of the carbon network to define the range of possible precursors.*” At that time Todd was generating all possible isomers of undecane and learning about 3D molecules and NMR. Corey and Wipke had different skill sets.

In 1967 the PDP-1 computer available at Harvard used 24K 18-bit words, drum storage, DECtape and paper tape, and the DECAL assembly language. It had three cathode ray tubes, a Rand tablet, and a joystick, and a Calcomp plotter for graphic output. The first synthesis planning program, Organic Chemical Synthesis Simulation (OCSS), used *ab initio* mechanistic reactions in several steps to make a “name” reaction. Functional groups, rings, ring junctures, aromaticity, conjugation, and atom and bond classes were perceived. The logic-oriented approach uses clues in the target to predict a

precursor: clues such as relationships of functional groups, functional group appendage relationships, ring sizes and ring junctures, and functional groups and rings. It was thus necessary to represent these entities. Chemists were excited by the Corey and Wipke publication;⁶² there must be logic to synthesis planning because a computer can do it. Dick Cramer and Jeff Howe then joined the team^{65,66} and Todd moved to Princeton.

Synthesis planning needed large programs and long term projects. Other problems included capturing reaction knowledge, granularity and consistency of the knowledge base, planning *versus* experimental detail, the shortage of trained, interested people, and the emergence of drug design as the new shiny toy. Chemists were turned off by predictions known to fail; they measured plans against empirical knowledge. Reaction databases were non-existent, but chemists really wanted automated reaction retrieval. There was a drive toward specific representation, leading to a large number of transforms and to large synthesis trees.

The CGL computer at Princeton in 1969 used 64K 36-bit words and had a 5MB disk; it was a multi-user system. This was used for the Simulation and Evaluation of Chemical Synthesis (SECS) program⁶⁷ which featured interactive 3D energy minimization; an acoustic tablet for drawing and control; prediction of steric and electronic control; trigonal, tetrahedral, and trigonal bipyramidal stereochemistry; heterocyclic chemistry; metabolic reactions, and the ALCHEM language for transforms.

MDL's REACCS program was launched in 1980. It allowed reaction databases to be created. Classic reaction collections such as Theilheimer were digitized. These could be searched by structure, substructure, reaction centers, and even stereochemistry. Multi-step sequences were handled and full literature references were stored. REACCS enabled manual synthesis planning. Using selective databases such as Current Synthetic Methodology and Current Chemical Reactions it was found⁶⁸ that citation analysis with reaction substructure search allowed retrieval of reactions not even in the computer. Todd's team also worked on mining a large reaction database^{69,70} to automate the building of the SECS knowledge base. Nowadays synthesis planning from a large reaction database is also available in Reaxys (<http://www.elsevier.com/online-tools/reaxys>).

Simply knowing the rules in chess does not make you a good chess player and the same can be said for SECS, where strategic control is necessary. A transform is the inverse of a synthetic reaction. Strategy is the problem solving method referring only to molecular structures. A goal is the result of applying a particular strategy to a particular problem, and refers to structure. Character is a type of structural change resulting from a transform. Wipke showed a symmetry example: three goals for breaking bonds in the retrosynthesis of beta-carotene. He also showed some QED predicate calculus of a strategy, and a topological goals chart. An important innovation was the separation of strategy from transforms.

Many companies used SECS. It was converted to a timesharing application and given a graphical GUI. A patent attorney wondered whether a synthesis produced by SECS were patentable. The program technology was adopted for other uses too. Students learned the logic of synthesis planning and synthesis papers included planning, just as Corey had anticipated in the 1960s.

Topomers

The final two papers in the symposium brought us up to date with Dick's current research interests. Bernd Wendt (Bernd.Wendt@certara.com) gave an overview of a wide array of topomer applications. To avoid duplication, I am leaving a detailed description of topomer technology until later. The first topomer application, in 1996 was ChemSpace,^{6,71} used in library design for general screening. It was followed by the DBTOP shape similarity search tool for activity mining, topomer CoMFA,⁷²⁻⁷⁵ AllChem,⁷⁶ a library of 10²⁰ synthesizable structures, Quantitative Series Enrichment Analysis (QSEA)⁷⁷ for SAR mining and, in 2013, Whole-Template CoMFA (WTC) to compare X-ray with template-based alignments.

Topomer shape similarity searching is very fast and increases the probability of finding active compounds. DBTOP for prospective selection of screening candidates by topomeric similarity was implemented as an automated workflow at Tripos Discovery Research leading to 308 selected compounds, and 11 successful "lead hops" in 13 assays.⁸

More recently, Bernd and his colleagues⁷⁸ identified a series of potent toluidinesulfonamide HIF-1 inhibitors, but the series was threatened by a potential liability to inhibit CYP2C9 which could cause dangerous drug-drug interactions. They then used structure-activity data from PubChem to develop a topomer CoMFA model that guided the design of novel sulfonamides with high selectivity for HIF-1 over CYP2C9 inhibition.

With Dick Cramer, Bernd examined the composition of 16 published QSAR datasets using Quantitative Series Enrichment Analysis (QSEA),⁷⁷ a procedure based on topomer technologies. QSEA allows the extraction of structure-activity relationships from large chemogenomic spaces starting from a single chemical structure. A heat map display in combination with topomer CoMFA and a novel series trajectory analysis revealed information for the assembly of structures into meaningful series. Global and local centroid structures can be determined from a similarity distance matrix and they build the origins for stepwise model building by increasing the similarity radius around the centroid nucleus. Bernd and Dick were able to determine whether compounds belonged to an emerging structure-activity relationship, and which compounds can be predicted within reliable limits.

QSEA has also been used in modeling off-target effects.⁷⁹ Queries were taken from the Jain set of marketed drugs to mine PubChem, ChemBank, and ChEMBL. SAR tables were constructed by assembling similar structures around each query structure that have an activity record for a particular target. QSEA was applied to these SAR tables to identify trends and to transform these trends into topomer CoMFA models. These models were able to highlight the structural trends associated with various off-target effects of marketed drugs, including cases where other structural similarity metrics would not have detected an off-target effect. One SAR trend identified was that fentanyl is inactive on hERG.

WTC is a current research project. Bernd and Dick took three datasets published by Brown and Muchmore (75 compounds tested against urokinase, 110 PTP-1B compounds and 123 Chk1-kinase compounds)⁸⁰ and aimed to develop CoMFA and CoMSIA⁸¹ models for X-ray ligand poses and multi-template aligned ligand poses, and then compare model robustness and interpretability and examine fluctuations of grid point interaction energies. In the three datasets having an experimental X-ray structure for every tested molecule, WTC alignment yielded CoMFA models which, compared to the

“all-X-ray aligned” CoMFA models, provided equal or better statistical quality and seemingly superior interpretability and utility.

Dick Cramer’s (Richard.Cramer@certara.com) award address homed in on Whole Template CoMFA. In theory, the primary cause of potency differences among ligands is steric and electrostatic field differences. Dick noted that when the goal is an informative comparison of ligand field differences, increasing ligand shape similarity is at least as productive as increasing physicochemical precision. As Tudor had observed earlier, if you cannot be sure of physical models, you can at least try to be consistent. Whole template CoMFA achieves ligand shape similarity by “copying” coordinates from any atom within a template ligand that “matches” a candidate’s atom, and by using the topomer protocol to generate coordinates for the remaining “non-matching” atoms.

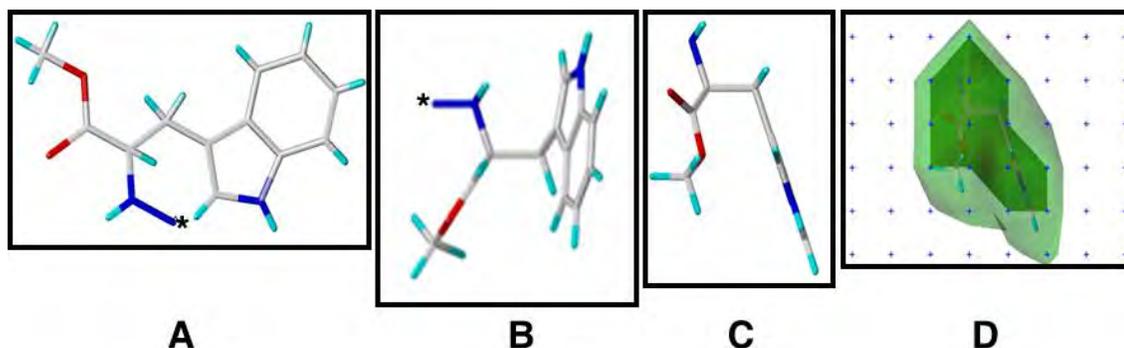
Dick has published four prospective “make and test” outcomes from topomer CoMFA.⁸² Shape similarity is highly productive for a number of reasons. All QSARs seek to explain differences in training set potencies. For example, the difference could be down to the substitution of fluorine for hydrogen. The gridded fields of 3D-QSAR’s descriptors directly and predictably express this: the differences in the local fields caused by changing hydrogen to fluorine may cause substantial change in the ligands’ and the receptor’s binding geometries. Furthermore, the frequency of chance correlation using PLS⁸³ is much lower than that for stepwise multiple regression, but perfect correlations involving descriptor subsets are not detected by PLS if the number of irrelevant descriptors is excessive. In CoMFA applications, the probability of chance correlation is usually negligible. Docking a small library moves the core around, producing field variation that is noise, because an invariant core cannot have caused changes in biological activity. With PLS, such noise tends to obscure the direct, certain, and causative field variation adjacent to the hydrogen or fluorine. Topomer generation rules were developed to produce alignments that are identical wherever the structures being compared are identical, or similar wherever the structural differences are slight. Topomer CoMFA focuses field variation and the resulting 3D-QSAR onto those direct, certain, and causative effects of 2D structure variation.

In WTC you identify the best matching “anchor bond” in the “candidate” (the test or training set structure to be aligned) and orient the candidate by overlay of its anchor bond onto that of the template. Anchor bond identification can be entirely automatic; template manual plus candidate automatic; or entirely manual. The best matching anchor bond includes all the candidate atoms that match a template atom. You then copy the coordinates of the matched template atoms to the matching candidate atoms and position the unmatched candidate atoms, by attaching their CONCORD-generated fragments and applying the topomer protocol.

To identify the candidate bond that best matches any template bond, the software considers, in order: every template, both “directions” of any bond, similarity in eight “localized” bond properties (or, within identical Murcko skeletons, identical location), and fraction of heavy atoms that match template atoms. Fully automatic identification involves combinatorial comparison of all pairings of plausible candidate and template bonds, where “plausible” means that one of the atoms defining the bond must not be carbon, or the bond type must be double or triple, or one of the atoms defining the bond must be in a ring and attached to at least three non-hydrogen atoms.

Atom matching uses breadth-first traversal, starting from a possible pairing of anchor bonds, in two passes: exact matching of atom and bond types (match score = 2), and skeleton matching only (the default, with match score = 1). Coordinate copying, using depth-first traversal, occurs if the atom is alicyclic or in rings whose atoms completely match, and hybridization agrees. It does not occur if the atom is in a ring and there are non-matching atoms in that ring. To modify or extend an outcome, a user can add templates.

A topomer is a single 3D model of a monovalent fragment constructed by a “black-box.” The only input is the “2D structure” of a single fragment (**A** below) embedded in 3D space by superposing the open valence (**B**), using valence geometries (bonds, angles, and rings) from CONCORD (**B**), and torsions, stereochemistry, and ring flips from canonical rules (**C**). The resulting strain energy is ignored. Several series can be combined in WTC to give a single 3D-QSAR, objectively based on all data, and X-ray interpretable.



Dick presented some initial WTC results that indeed combine diverse structures into a single predictive 3D-QSAR model, and are derived automatically. He used all the Factor Xa inhibitors in Bindingdb and showed that a combined WTC model was better than the single series WTC models. For example, the q^2 value ranged from -0.843, for 15 compounds that binding with PDB code NFX, to 0.602 for 21 1FJS structures; the combined model had q^2 0.616. Twelve, mainly poor datasets were combined into one good one. Results were even better for MAP kinase P38 alpha inhibitors.

The q^2 values for the combined models are probably to some extent artifacts since Bindingdb ligands are subsets, probably chosen to provide the most docking challenges for the least computation, and leave-one-out q^2 is too pessimistic when a unique structural change produces a strong effect on potency. Nevertheless, as it turns out, this contrast in results requires that the potency effect of a field at any particular lattice point be uniform, regardless of the great diversity of training set structures that produce different field intensities at that point. Thus the combined WTC models worked “for the right reason.”

One application area suggested for WTC is off-target prediction. Topomer applications in that field have already been published.^{78,79,84} WTC allows any scientist to carry out 3D-QSAR modeling. Different project team members receive different benefits. Synthetic chemists can simultaneously consider the tradeoffs between synthetic costs and likelihood of therapeutic benefit. For the computer-

aided molecular design practitioner an automatic protocol allows more attention on the most important issues such as training set composition and assessing validity of project-critical predictions. For project leaders, WTC allows more complete consideration of the dozens of relevant biological endpoints and the astronomical numbers of possible structural modifications.

In summary, WTC is a ligand alignment protocol for classical CoMFA that uses as input only 3D template(s) and a 2D SAR table, thus providing fast and convenient throughput; objectively determined models; application of crystallographic and/or pharmacophoric constraints; and structurally unlimited applicability. As output, it enables rapid, objective, structurally unlimited potency predictions that so far are reasonably accurate; contour maps that are more structurally informative; 3D database searching with potency predictions; and *de novo* design constrained by potency prediction. Its 3D-QSAR models can combine multiple series within a single model and be generated completely automatically.

Conclusion

The symposium was ably chaired by Brian Masek and Terry Stouch. After Dick's award address, Antony Williams, Chair of the ACS Division of Chemical Information, formally presented the Herman Skolnik Award:



References

- (1) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110* (18), 5959-5967.
- (2) Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. Neighborhood behavior: a useful concept for validation of "molecular diversity" descriptors. *J Med Chem* **1996**, *39* (16), 3049-3059.
- (3) Cramer, R. D.; Clark, R. D.; Patterson, D. E.; Ferguson, A. M. Bioisosterism as a Molecular Diversity Descriptor: Steric Fields of Single "Topomeric" Conformers. *J. Med. Chem.* **1996**, *39* (16), 3060-3069.
- (4) Clark, R. D.; Cramer, R. D. Taming the combinatorial centipede. *CHEMTECH* **1997**, *27* (5), 24-31.
- (5) Clark, R. D.; Ferguson, A. M.; Cramer, R. D. Bioisosterism and molecular diversity. *Perspect. Drug Discovery Des.* **1998**, *9/10/11* (3D QSAR in Drug Design: Ligand/Protein Interactions and Molecular Similarity), 213-224.
- (6) Cramer, R. D.; Patterson, D. E.; Clark, R. D.; Soltanshahi, F.; Lawless, M. S. Virtual Compound Libraries: A New Approach to Decision Making in Molecular Discovery Research. *J. Chem. Inf. Comput. Sci.* **1998**, *38* (6), 1010-1023.
- (7) Clark, R. D.; Brusati, M.; Jilek, R.; Heritage, T.; Cramer, R. D. Validating novel QSAR descriptors for use in diversity analysis. In *Molecular Modeling and Prediction of Bioactivity, Proceedings of the European Symposium on Quantitative Structure-Activity Relationships: Molecular Modeling and Prediction of Bioactivity*, 12th, Copenhagen, Denmark, Aug. 23-28, 1998; Gundertofte, K.; Jorgensen, F. S., Eds.; Kluwer Academic/Plenum Publishers: New York, NY, 2000; pp 95-100.
- (8) Cramer, R. D.; Jilek, R. J.; Guessregen, S.; Clark, S. J.; Wendt, B.; Clark, R. D. "Lead Hopping". Validation of Topomer Similarity as a Superior Predictor of Similar Biological Activities. *J. Med. Chem.* **2004**, *47* (27), 6777-6791.
- (9) Clark, R. D. Synthesis and QSAR of herbicidal 3-pyrazolyl α,α,α -trifluorotolyl ethers. *J. Agric. Food Chem.* **1996**, *44* (11), 3643-3652.
- (10) Clark, R. D.; Leonard, J. M.; Strizhev, A. Pharmacophore models and comparative molecular field analysis (CoMFA). In *Pharmacophore Perception, Development, and Use in Drug Design*; Güner, O. F., Ed.; International University Line: La Jolla, CA, 1999; pp 153-167.
- (11) Clark, R. D.; Sprous, D. G.; Leonard, J. M. Validating models based on large data sets. In *Rational Approaches to Drug Design. (Proceedings of the 13th European Symposium on Quantitative Structure-Activity Relationships, held 27 August-1 September 2000, in Dusseldorf, Germany.)*; Holtje, H. D.; Sippl, W., Eds.; Prous Science: Barcelona, Spain, 2001; pp 475-485.
- (12) Wolohan, P. R. N.; Clark, R. D. Predicting drug pharmacokinetic properties using molecular interaction fields and SIMCA. *J. Comput.-Aided Mol. Des.* **2003**, *17* (1), 65-76.
- (13) Clark, R. D. Boosted leave-many-out cross-validation: the effect of training and test set diversity on PLS statistics. *J. Comput.-Aided Mol. Des.* **2003**, *17* (2-4), 265-275.
- (14) Clark, R. D.; Fox, P. C. Statistical variation in progressive scrambling. *J. Comput.-Aided Mol. Des.* **2004**, *18* (7-9), 563-576.
- (15) Clark, R. D. A ligand's-eye view of protein binding. *J. Comput.-Aided Mol. Des.* **2008**, *22* (6-7), 507-521.
- (16) Clark, R. D. DPRESS: Localizing estimates of predictive uncertainty. *J Cheminform* **2009**, *1* (1), 11.
- (17) Clark, R. D. Prospective ligand- and target-based 3D QSAR: state of the art 2008. *Curr. Top. Med. Chem. (Sharjah, United Arab Emirates)* **2009**, *9* (9), 791-810.
- (18) Clark, M.; Cramer, R. D., III; Jones, D. M.; Patterson, D. E.; Simeroth, P. E. Comparative molecular field analysis (CoMFA). 2. Toward its use with 3D-structural databases. *Tetrahedron Comput. Methodol.* **1990**, *3* (1), 47-59.

- (19) Cho, S. J.; Tropsha, A. Cross-Validated R2-Guided Region Selection for Comparative Molecular Field Analysis: A Simple Method To Achieve Consistent Results. *J. Med. Chem.* **1995**, *38* (7), 1060-1066.
- (20) Norinder, U. Single and domain mode variables selection in 3D QSAR applications. *J. Chemom.* **1996**, *10* (2), 95-105.
- (21) Kroemer, R. T.; Hecht, P. Replacement of steric 6-12 potential-derived interaction energies by atom-based indicator variables in CoMFA leads to models of higher consistency. *J. Comput.-Aided Mol. Des.* **1995**, *9* (3), 205-212.
- (22) Lindgren, F.; Geladi, P.; Wold, S. Kernel-based pls regression; cross-validation and applications to spectral data. *J. Chemom.* **1994**, *8* (6), 377-389.
- (23) Wise, M.; Cramer, R. D.; Smith, D.; Exman, I. Progress in three-dimensional drug design: the use of real time colour graphic and computer postulation of bioactive molecules in DYLOMMS. In *Pharmacochemistry Library, Vol. 6: Quantitative Approaches to Drug Design*; Dearden, J. C., Ed.; Elsevier: Amsterdam, The Netherlands, 1983; pp 145-146.
- (24) Wold, S.; Martens, S.; Wold, H. The Multivariate Calibration Problem in Chemistry Solved by the PLS Method. In *Matrix Pencils: Proceedings of a Conference Held at Pite Havsbad, Sweden, March 22-24, 1982 (Lecture Notes in Mathematics)*; Kagström, B.; Ruhe, A., Eds.; Springer Verlag: Heidelberg, Germany, 1983; pp 286-293.
- (25) Cramer, R. D., III; Wold, S. B. Comparative molecular field analysis (CoMFA). US5025388A, 1991.
- (26) Kim, K. H.; Martin, Y. C. Evaluation of electrostatic and steric descriptors for 3D-QSAR: the hydrogen ion and methyl group probes using comparative molecular field analysis (CoMFA) and the modified partial least squares method. *Pharmacochem. Libr.* **1991**, *16* (QSAR: Ration. Approaches Des. Bioact. Compd.), 151-154.
- (27) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28* (7), 849-857.
- (28) Hansch, C.; Fujita, T. ρ - σ - π Analysis; method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* **1964**, *86* (8), 1616-1626.
- (29) Fujita, T.; Iwasa, J.; Hansch, C. A new substituent constant, π , derived from partition coefficients. *J. Am. Chem. Soc.* **1964**, *86* (23), 5175-5180.
- (30) Collander, R. Partition of organic compounds between higher alcohols and water. *Acta Chem. Scand.* **1951**, *5*, 774-780.
- (31) Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J. Comput.-Aided Mol. Des.* **1993**, *7* (1), 83-102.
- (32) Brint, A. T.; Willett, P. Algorithms for the identification of three-dimensional maximal common substructures. *J. Chem. Inf. Comput. Sci.* **1987**, *27* (4), 152-158.
- (33) Sheridan, R. P.; Nilakantan, R.; Dixon, J. S.; Venkataraghavan, R. The ensemble approach to distance geometry: application to the nicotinic pharmacophore. *J. Med. Chem.* **1986**, *29* (6), 899-906.
- (34) Dammkoehler, R. A.; Karasek, S. F.; Shands, E. F. B.; Marshall, G. R. Constrained search of conformational hyperspace. *J. Comput.-Aided Mol. Des.* **1989**, *3* (1), 3-21.
- (35) Pearlstein, R. A.; Malhotra, D.; Orchard, B. J.; Tripathy, S. K.; Potenzzone, R., Jr.; Grigoras, S.; Koehler, M.; Mabilia, M.; Walters, D. E.; Doherty, D.; Harr, R.; Hopfinger, A. J. Three-dimensional structure modeling and quantitative molecular design using CHEMLAB-II. *New Methods Drug Res.* **1988**, *2*, 147-174.
- (36) Rhyu, K. B.; Patel, H. C.; Hopfinger, A. J. A 3D-QSAR Study of Anticoccidial Triazines Using Molecular Shape Analysis. *J. Chem. Inf. Comput. Sci.* **1995**, *35* (4), 771-778.
- (37) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR Models Using the 4D-QSAR Analysis Formalism. *J. Am. Chem. Soc.* **1997**, *119* (43), 10509-10524.

- (38) Iyer, M.; Tseng, Y. J.; Senese, C. L.; Liu, J.; Hopfinger, A. J. Prediction and mechanistic interpretation of human oral drug absorption using MI-QSAR analysis. *Mol. Pharm.* **2007**, *4* (2), 218-231.
- (39) Santos-Filho, O. A.; Hopfinger, A. J. Combined 4D-fingerprint and clustering based membrane-interaction QSAR analyses for constructing consensus Caco-2 cell permeation virtual screens. *J. Pharm. Sci.* **2008**, *97* (1), 566-583.
- (40) Jain, A. N.; Koile, K.; Chapman, D. Compass: Predicting Biological Activities from Molecular Surface Properties. Performance Comparisons on a Steroid Benchmark. *J. Med. Chem.* **1994**, *37* (15), 2315-2327.
- (41) Jain, A. N.; Dietterich, T. G.; Lathrop, R. H.; Chapman, D.; Critchlow, R. E., Jr.; Bauer, B. E.; Webster, T. A.; Lozano-Perez, T. Compass: a shape-based machine learning tool for drug design. *J. Comput.-Aided Mol. Des.* **1994**, *8* (6), 635-652.
- (42) Jain, A. N. QMOD: physically meaningful QSAR. *J. Comput.-Aided Mol. Des.* **2010**, *24* (10), 865-878.
- (43) Varela, R.; Walters, W. P.; Goldman, B. B.; Jain, A. N. Iterative Refinement of a Binding Pocket Model: Active Computational Steering of Lead Optimization. *J. Med. Chem.* **2012**, *55* (20), 8926-8942.
- (44) Glen, R. C. Connecting the virtual world of computers to the real world of medicinal chemistry. *Future Med. Chem.* **2011**, *3* (4), 399-403.
- (45) Orchard, S.; Al-Lazikani, B.; Bryant, S.; Clark, D.; Calder, E.; Dix, I.; Engkvist, O.; Forster, M.; Gaulton, A.; Gilson, M.; Glen, R.; Grigorov, M.; Hammond-Kosack, K.; Harland, L.; Hopkins, A.; Larminie, C.; Lynch, N.; Mann, R. K.; Murray-Rust, P.; Lo, P. E.; Southan, C.; Steinbeck, C.; Wishart, D.; Hermjakob, H.; Overington, J.; Thornton, J. Minimum information about a bioactive entity (MIABE). *Nat. Rev. Drug Discovery* **2011**, *10* (9), 661-669.
- (46) Gleeson, M. P.; Modi, S.; Bender, A.; Robinson, R. L. M.; Kirchmair, J.; Promkatkaew, M.; Hannongbua, S.; Glen, R. C. The challenges involved in modeling toxicity data in silico: a review. *Curr. Pharm. Des.* **2012**, *18* (9), 1266-1291.
- (47) Koutsoukas, A.; Simms, B.; Kirchmair, J.; Bond, P. J.; Whitmore, A. V.; Zimmer, S.; Young, M. P.; Jenkins, J. L.; Glick, M.; Glen, R. C.; Bender, A. From in silico target prediction to multi-target drug design: Current databases, methods and applications. *J. Proteomics* **2011**, *74* (12), 2554-2574.
- (48) Koutsoukas, A.; Lowe, R.; KalantarMotamedi, Y.; Mussa, H. Y.; Klaffke, W.; Mitchell, J. B. O.; Glen, R. C.; Bender, A. In Silico Target Predictions: Defining a Benchmarking Data Set and Comparison of Performance of the Multiclass Naive Bayes and Parzen-Rosenblatt Window. *J. Chem. Inf. Model.* **2013**, *53* (8), 1957-1966.
- (49) Kirchmair, J.; Williamson, M. J.; Tyzack, J. D.; Tan, L.; Bond, P. J.; Bender, A.; Glen, R. C. Computational Prediction of Metabolism: Sites, Products, SAR, P450 Enzyme Dynamics, and Mechanisms. *J. Chem. Inf. Model.* **2012**, *52* (3), 617-648.
- (50) Kirchmair, J.; Howlett, A.; Peironcely, J. E.; Murrell, D. S.; Williamson, M. J.; Adams, S. E.; Hankemeier, T.; van, B. L.; Duchateau, G.; Klaffke, W.; Glen, R. C. How Do Metabolites Differ from Their Parent Molecules and How Are They Excreted? *J. Chem. Inf. Model.* **2013**, *53* (2), 354-367.
- (51) Oprea, T. I.; Gottfries, J. Chemography: The Art of Navigating in Chemical Space. *J. Comb. Chem.* **2001**, *3* (2), 157-166.
- (52) Oprea, T. I. Chemical space navigation in lead discovery. *Curr. Opin. Chem. Biol.* **2002**, *6* (3), 384-389.
- (53) Cramer, R. D., III BC(DEF) parameters. 1. The intrinsic dimensionality of intermolecular interactions in the liquid state. *J. Am. Chem. Soc.* **1980**, *102* (6), 1837-1849.
- (54) Bologna, C. G.; Revankar, C. M.; Young, S. M.; Edwards, B. S.; Arterburn, J. B.; Kiselyov, A. S.; Parker, M. A.; Tkachenko, S. E.; Savchuck, N. P.; Sklar, L. A.; Oprea, T. I.; Prossnitz, E. R. Virtual and biomolecular screening converge on a selective agonist for GPR30. *Nat. Chem. Biol.* **2006**, *2* (4), 207-212.

- (55) Dennis, M. K.; Burai, R.; Ramesh, C.; Petrie, W. K.; Alcon, S. N.; Nayak, T. K.; Bologna, C. G.; Leitao, A.; Brailoiu, E.; Deliu, E.; Dun, N. J.; Sklar, L. A.; Hathaway, H. J.; Arterburn, J. B.; Oprea, T. I.; Prossnitz, E. R. In vivo effects of a GPR30 antagonist. *Nat. Chem. Biol.* **2009**, *5* (6), 421-427.
- (56) Oprea, T. I.; Gottfries, J. Toward minimalistic modeling of oral drug absorption¹. *J. Mol. Graphics Modell.* **2000**, *17* (5/6), 261-274.
- (57) Wu, C.-Y.; Benet, L. Z. Predicting Drug Disposition via Application of BCS: Transport/Absorption/ Elimination Interplay and Development of a Biopharmaceutics Drug Disposition Classification System. *Pharm. Res.* **2005**, *22* (1), 11-23.
- (58) Benet, L. Z.; Broccatelli, F.; Oprea, T. I. BDDCS Applied to Over 900 Drugs. *AAPS J.* **2011**, *13* (4), 519-547.
- (59) Broccatelli, F.; Cruciani, G.; Benet, L. Z.; Oprea, T. I. BDDCS Class Prediction for New Molecular Entities. *Mol. Pharmaceutics* **2012**, *9* (3), 570-580.
- (60) Broccatelli, F.; Larregieu, C. A.; Cruciani, G.; Oprea, T. I.; Benet, L. Z. Improving the prediction of the brain disposition for orally administered drugs using BDDCS. *Adv. Drug Delivery Rev.* **2012**, *64* (1), 95-109.
- (61) Broccatelli, F.; Carosati, E.; Cruciani, G.; Oprea, T. I. Transporter-mediated efflux influences CNS side effects: ABCB1, from antitarget to target. *Mol. Inf.* **2010**, *29* (1-2), 16-26.
- (62) Corey, E. J.; Wipke, W. T. Computer-assisted design of complex organic syntheses. *Science* **1969**, *166* (3902), 178-192.
- (63) Corey, E. J.; Ohno, M.; Mitra, R. B.; Vatakencherry, P. A. Total synthesis of longifolene. *J. Am. Chem. Soc.* **1964**, *86* (3), 478-485.
- (64) Corey, E. J. General methods for the construction of complex molecules. *Pure Appl. Chem.* **1967**, *14* (1), 19-37.
- (65) Corey, E. J.; Wipke, W. T.; Cramer, R. D., III; Howe, W. J. Computer-assisted synthetic analysis. Facile man-machine communication of chemical structure by interactive computer graphics. *J. Am. Chem. Soc.* **1972**, *94* (2), 421-430.
- (66) Corey, E. J.; Wipke, W. T.; Cramer, R. D., III; Howe, W. J. Techniques for perception by a computer of synthetically significant structural features in complex molecules. *J. Am. Chem. Soc.* **1972**, *94* (2), 431-439.
- (67) Wipke, W. T.; Whetstone, P. Graphic digitizing in 3-D. In *Computer Graphics*; ACM: New York, NY, 1971; Vol. 5, p 10.
- (68) Wipke, W. T.; Vladutz, G. An alternative view of reaction similarity: citation analysis. *Tetrahedron Comput. Methodol.* **1990**, *3* (2), 83-107.
- (69) Yanaka, M.; Nakaura, K.; Kurumisawa, A.; Wipke, W. T. Automatic knowledge base building for the organic synthesis design program (SECS). *Prog. Clin. Biol. Res.* **1989**, *291* (QSAR: Quant. Struct.-Act. Relat. Drug Des.), 147-150.
- (70) Yanaka, M.; Nakamura, K.; Kurumisawa, A.; Wipke, W. T. Automatic knowledge base building for the organic synthesis design program (SECS). *Tetrahedron Comput. Methodol.* **1990**, *3* (6A), 359-375.
- (71) Cramer, R. D.; Poss, M. A.; Hermsmeier, M. A.; Caulfield, T. J.; Kowala, M. C.; Valentine, M. T. Prospective Identification of Biologically Active Structures by Topomer Shape Similarity Searching. *J. Med. Chem.* **1999**, *42* (19), 3919-3933.
- (72) Cramer, R. D. Topomer CoMFA: A Design Methodology for Rapid Lead Optimization. *J. Med. Chem.* **2003**, *46* (3), 374-388.
- (73) Jilek, R. J.; Cramer, R. D. Topomers: A Validated Protocol for Their Self-Consistent Generation. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (4), 1221-1227.
- (74) Cramer, R. D.; Cruz, P.; Stahl, G.; Curtiss, W. C.; Campbell, B.; Masek, B. B.; Soltanshahi, F. Virtual Screening for R-Groups, including Predicted pIC₅₀ Contributions, within Large Structural Databases, Using Topomer CoMFA. *J. Chem. Inf. Model.* **2008**, *48* (11), 2180-2195.
- (75) Cramer, R. D. R-group template CoMFA combines benefits of "ad hoc" and topomer alignments using 3D-QSAR for lead optimization. *J. Comput.-Aided Mol. Des.* **2012**, *26* (7), 805-819.

- (76) Cramer, R. D.; Soltanshahi, F.; Jilek, R.; Campbell, B. AllChem: Generating and searching 10^{20} synthetically accessible structures. *J. Comput.-Aided Mol. Des.* **2007**, *21* (6), 341-350.
- (77) Wendt, B.; Cramer, R. D. Quantitative Series Enrichment Analysis (QSEA): a novel procedure for 3D-QSAR analysis. *J. Comput.-Aided Mol. Des.* **2008**, *22* (8), 541-551.
- (78) Wendt, B.; Mulbaier, M.; Wawro, S.; Schultes, C.; Alonso, J.; Janssen, B.; Lewis, J. Toluidinesulfonamide Hypoxia-Induced Factor 1 Inhibitors: Alleviating Drug-Drug Interactions through Use of PubChem Data and Comparative Molecular Field Analysis Guided Synthesis. *J. Med. Chem.* **2011**, *54* (11), 3982-3986.
- (79) Wendt, B.; Uhrig, U.; Bos, F. Capturing structure-activity relationships from chemogenomic spaces. *J. Chem. Inf. Model.* **2011**, *51* (4), 843-851.
- (80) Brown, S. P.; Muchmore, S. W. Large-Scale Application of High-Throughput Molecular Mechanics with Poisson-Boltzmann Surface Area for Routine Physics-Based Scoring of Protein-Ligand Complexes. *J. Med. Chem.* **2009**, *52* (10), 3159-3165.
- (81) Klebe, G.; Abraham, U.; Mietzner, T. Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict Their Biological Activity. *J. Med. Chem.* **1994**, *37* (24), 4130-4146.
- (82) Cramer, R. D. Rethinking 3D-QSAR. *J. Comput.-Aided Mol. Des.* **2011**, *25* (3), 197-201.
- (83) Clark, M.; Cramer, R. D., III The probability of chance correlation using partial least squares (PLS). *Quant. Struct.-Act. Relat.* **1993**, *12* (2), 137-145.
- (84) Nisius, B.; Goeller, A. H. Similarity-Based Classifier Using Topomers to Provide a Knowledge Base for hERG Channel Inhibition. *J. Chem. Inf. Model.* **2009**, *49* (2), 247-256.

Wendy Warr, Reporter, 2013 Herman Skolnik Award Symposium



Symposium speakers: Bobby Glen, Brian Masek, Bernd Wendt, Dick Cramer, Tudor Oprea, Todd Wipke, Bob Clark, Yvonne Martin

Before and After Lab: Instructing Students in 'Non-Chemical' Research Skills

Continuing a theme put together at a symposium at the 2012 Biennial Conference on Chemical Education (a [report](#) was published in *Chemical Information Bulletin*, Winter 2012), the presenters at the recent CINF symposium at the 2013 Fall ACS National Meeting discussed a variety of topics useful for chemistry students above and beyond the basic skills of chemical research.

To open the session, Teri Vogel of the University of California – San Diego library presented “Chemical information across San Diego County: a community college and university library collaboration for an independent synthesis project” co-authored with Cynthia Gilley of Palomar College. Vogel and Gilley joined forces to introduce Gilley’s organic synthesis lab students to major resources for locating syntheses in the primary literature. After introducing them to the concept of the flow of information, and searching techniques, Vogel obtained guest access to SciFinder for the 13 students, and they were directed to use SciFinder and/or Reaxys to find a reference for a two-step synthesis of their designated compound. Most students had success with the databases, though in some cases, both steps were not found in the same document. The collaborators have not yet decided whether to repeat the experiment with the coming year’s class.

Unfortunately, the second scheduled paper, “Teaching chemical information literacy through an undergraduate laboratory project” by Martin Walker of the State University of New York at Potsdam, had to be withdrawn due to a family emergency.

Shu Guo, science reference librarian at Central Michigan University, offered “Integrating citations as a teaching element into chemistry information literacy training methods.” CMU’s organic chemistry lab course had previously offered four embedded chemical information instruction sessions, covering general searching, Web of Science, SciFinder and Reaxys. Most recently, Shu has added an element focusing on citations: reading citations, and what, when, why and how to cite, including introduction to cited reference searching in Web of Science and the ACS citation style. Students reported increased confidence in dealing with the literature, and ability to apply it in their lab course.

The next paper, “Designing instruction activities to guide students through the research lifecycle: a science librarian approach,” described assignments designed to show the students the role of information in each stage of the research cycle: creating a proposal, planning and carrying out the experiment, sharing the results and application of the results. Ye Li, chemistry librarian at the University of Michigan at Ann Arbor, introduced the students to methods to find, organize, manage and evaluate scientific information.

“I can just copy this, right?” discussed aspects of copyright that students need to know, first as users, then as producers of copyrighted material. Charles Huber of the University of California at Santa Barbara touched on copyright as relevant to undergraduates: the basic meaning of copyright, what “fair use” allows them to do...and does not, and the distinctions between copyright violation and plagiarism. As producers of publications, graduate students need to know a lot more, both the permissions needed to reuse others’ copyrighted materials and what their rights as authors are. Key topics include work-for-hire rules at institutions, transfer of copyright to publishers, the various meanings of “open access” and the opportunities offered by Creative Commons licensing.

Donna Wrublewski, currently a science librarian at Caltech, described some of her collaboration with faculty in her previous job in “Anything BUT overlooked: librarians teaching scientific communication skills at the University of Florida” co-authored with Sara Gonzalez and Margeaux Johnson of the University of Florida Libraries. Summarizing the material covered as “what I wish I’d known when I

started grad school,” Donna described an honors program course offered to a group of about twenty students, mostly freshmen. Topics included evaluating scientific literature, creating an annotated bibliography, preparing and presenting a poster, and writing abstracts and papers. The program included faculty guest lecturers introducing research opportunities for the undergraduates. The course evolved from one session to another in response to student feedback.

Electronic laboratory notebooks have made great inroads in industry, but so far have not become widespread in academia. Svetla Baykoucheva described some of the efforts to do so in “Introducing electronic laboratory notebooks (ELNs) to students and researchers at the University of Maryland – College Park.” The benefits of ELNs are many (they can save time, preserve data, establish priority for intellectual property purposes, and facilitate data management plans now frequently required by funding agencies), but academics have often found them expensive and difficult to implement. The University of Maryland evaluated both LabArchives Classroom Edition and a “light” version of the Contur ELN from Accelrys, deciding on the former in 2011. The library partnered with instructors in 2013 to develop a project for an instrumental lab course. Students were to use the ELN system to access lab protocols, create and submit lab reports, and share files. This required a great deal of effort from the librarian: assigning materials to both students and TAs, and grading for 45 students. Students were able to search PubMed from within the ELN, and used customized calculators to analyze their data. ELN use improved communication among students and with instructors. Key problems encountered include: undergraduates do not generate enough data to make effective use of the ELNs, students did not like bringing laptops to the laboratory, and the ELN was perceived as more time-consuming than standard lab notebooks. Future plans include broadening the use of ELN across more courses for chemistry majors and graduate students.

Dealing with chemical information instruction in large laboratory classes was the subject of Judith Currano’s “Teaching chemical information in bulk.” Previous attempts to incorporate chemical information instruction in the University of Pennsylvania’s organic chemistry lab course had run afoul of lack of time during the quarter, and a tendency for students to skip the lecture. However, a new approach, teaching small groups during the lab check-in week proved more successful. This format allowed a full 90-minute session, with opportunities for discussion and hands-on practice with electronic resources. Topics covered included “the anatomy of a handbook” and identifying substances. Handouts compared resources on their ease of use and fee-based vs. free. Both instructors and students deemed the sessions successful, with the students asking good questions about the material.

Antony Williams of the Royal Society of Chemistry (RSC) discussed “Social profile of a chemist online: potential profits of participation.” In a scholarly environment where online presence and influence is measurable, altmetrics will increasingly supplement, and perhaps supplant, such traditional estimators of scientific stature as citation statistics and the impact factors of the journals in which one publishes. A researcher can help craft his or her own scholarly profile in a variety of ways: creating an ORCID identifier to help ensure proper attribution of published work; micropublishing through tools like ChemSpider and ChemSpider Synthetic Pages to preserve and disseminate research that might never make it into a traditional paper; sharing your work freely on the Web using repositories, as well as tools like SlideShare, YouTube, and SciVee; and blogging and tweeting. Antony recommended maintaining separate “identities”/accounts for purely social and personal networking vs. professional and scholarly networking, and maintaining a single spot where all your professional networking sites can be found. Sources like ImpactStory and Plum Analytics can help researchers track their own altmetrics.

“Safety outreach to the academic chemistry community” was the topic of Ralph Stuart’s presentation. Recent accidents in academic laboratories have highlighted the need for the development of a “safety culture” in academic institutions. He noted that personal safety is not the same thing as system safety, pointing to the Deep Water Horizon oil platform disaster as an example. His position in the Department of Environmental Health and Safety at Cornell University has involved him directly in trying to develop safety culture. The key concept is RAMP: Recognize, Assess, Manage, Prepare. One traditional chemical safety resource, the Materials Safety Data Sheet, with its lack of standardization, is “dead,” being replaced by the Globally Harmonized System for the Classification of Chemicals. Stuart recommended “Laboratory Safety for Chemistry Students” by Hill and Finster (Wiley, 2010) and the [website](#) of the ACS Division of Chemical Health and Safety as good starting points for resources.

Pamela Scott of Pfizer concluded the session with “Other skills for post-graduates,” enumerating many of the “soft skills” that can be as important for professional success as the technical knowledge and laboratory skills which students traditionally learn. Self-assessment is extremely important, and Pamela commended the Meyers-Briggs personality assessment as a useful tool to assess creativity and innovation, and motivation and commitment. One’s non-job interests and social interactions can be important to professional success. Time management, priority setting, problem solving, negotiation and team skills are all vital in any organization. Budgeting, contracts and other fiscal skills are important, and can often be developed through volunteer work in non-profit organizations, as well as skills in dealing with clients. Communication skills, including written and oral presentations, can also be developed both inside and outside the academic environment. Making the habit of continuous learning is vital to keeping all of these skills honed.

Charles Huber, Symposium Co-Organizer



2014 Biennial Conference on Chemical Education

August 3-7, 2014

*GRAND VALLEY STATE UNIVERSITY
ALLENDALE, MICHIGAN*

<http://www.bcce2014.org/>

Call for abstracts begins January 1, 2014

Exchangeable Molecular and Analytical Data Formats and their Importance in Facilitating Data Exchange

During the morning session on molecular data formats, Keith Taylor (Accelrys) and Roger Sayle (NextMove Software) both noted that while a small number of molecular graphic formats were in common use (like the ubiquitous molfile), some users did not conform to either the Mol V2000 or V3000 published standards. Roger noted that for a data set that was created with a range of different element and charge types and then tested with 24 different “mol” file reader packages the failures and errors were disturbingly large.

Geoffrey Hutchinson (University of Pittsburgh) then gave a description of the OpenBabel project (<http://openbabel.org>) that has produced a toolbox to read, write and convert over 110 chemical file formats, and the difficulties that have been created by the non-conformity with formats.

In his presentation, Phil McHale noted that Perkin-Elmer was working on an Open XML format for export of data from Electronic Laboratory Notebooks (ELN). At present formats were generally proprietary. He reviewed the CDX and CDXML formats as well, both of which have been widely accepted and utilized.

www.inchi-trust.org

Too many “standards” actually slow things down and make getting to the information you want and need take a lot longer time and effort than it would take with InChI

InChI TRUST

Slide courtesy of Stephen Heller

videos have been produced to attempt to explain their application.

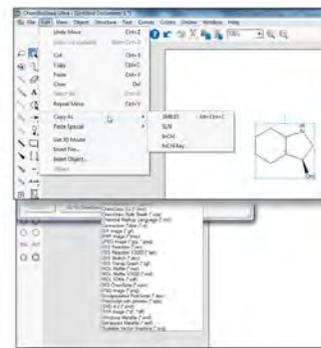
Evan Bolton (National Center for Biotechnology Information, National Institutes of Health) spoke about the new features in the PubChem data submission portal that support a wide range of user-defined data, and about the need for data standards.

Barry Bunin (Collaborative Drug Discovery) noted that there was no standard approach for a computer-based way of managing large molecules such as: peptides, antibodies, therapeutic proteins or vaccines. HELM (Hierarchical Editing Language for Macromolecules) was being introduced as an Open Source approach by Pfizer and was released into production in 2008. He then introduced the CDD (Collaborative Drug Discovery) vault as a hosted database solution for secure management and sharing of chemical and biological data.

CDX and CXML in Use



- ▶ Standard chemical file format for USPTO
- ▶ Chemical file format accepted by J.Amer.Chem.Soc.
- ▶ Interconvertible with molfile, SMILES, InChI
- ▶ Compatible with SciFinder, Reaxys, ChemSpider, Spotfire, eMolecules, KRMS, Vortex ...



These structures may change without notice as we respond to customer requirements.

Slide courtesy of Phil McHale

Stephen Heller gave an update on the InChI representation and the InChI Trust. Like barcodes, and QR codes, InChIs are not designed to be interpreted by humans, but are produced by computer from structures drawn on-screen with existing structure drawing software. The original structure can be regenerated from an InChI with appropriate software. Steve noted that a number of

For the afternoon session on spectroscopic data, the first presentation was a joint paper from Tony Davies (AkzoNobel Chemicals) and Robert Lancashire (University of the West Indies) who gave some history on the JCAMP-DX data formats. Recognition was given to Paul Wilks, Bob McDonald and Jeannette Grasselli-Brown as pioneers in the publication of JCAMP-DX standards. Since 1988 the standards for a wide range of techniques have been published and in 1995 they became the responsibility of IUPAC.

Michael Boruta (Advanced Chemistry Development) followed by showing the transition from hand written annotations on chart paper copies of spectra to electronic equivalents that could be stored in "knowledgebases." For example, ACD/Labs Spectrus Process includes separate knowledgebases for IR and Raman. The assignments can be exported as part of JCAMP-DX files, but no standard for this exists.

Where is NMR today



The stored data should be complete.

It has to be possible to recreate or rerun the data collection with the information stored

This also applies to exchangeable data formats

For NMR this means:

Pulseprogram, parameters, extras such as decoupling sequences, lists, shaped pulse etc

Slide courtesy of Clemens Anklin

Clemens Anklin (Bruker Biospin) identified the common data formats used for various techniques. In the case of NMR this was predominantly JCAMP-DX. He lamented the fact that whilst 2D NMR had existed before any JCAMP-DX standards were published, the latest accepted standard for NMR was 5.01 published in 1999 and this only covered 1D. The version 6 format for 2D has been in draft since 2002 and has been implemented by vendors who could not wait any longer.

Stuart Chalk (University of North Florida) introduced the AnIML specification and highlighted the features and benefits of using an XML protocol that could be fully validated. He noted that from 2003 it was designed to be a (backwards compatible) replacement for JCAMP-DX. The task group guiding the process set its charter: "to develop an analytical data standard that can be used to store data from any analytical instrument" and holds virtual meetings on a monthly basis to develop the specification. The first set of specifications is targeted to go through ASTM balloting in early 2014

Bob Hanson (St. Olaf College) finished this session with a proposal to have an extension to the JCAMP-DX standard whereby a single file could contain the molecular graphics data as well as the spectrum, together with annotations linking the two. This would allow interaction with cloud services such that a molfile could be passed to a server and a simulated spectrum returned with sufficient information to apply all the required annotations to identify the peaks.

The full symposium program is listed in *Chemical Information Bulletin*, 2013, 65(3) at: <http://bulletin.acscinf.org/node/486#THa>.

InChI videos: What on Earth is InChI? <http://www.youtube.com/watch?v=rAnJ5toz26c>
The Birth of the InChI <http://www.youtube.com/watch?v=X9c0PHXPfso>
The Googable InChIKey <http://www.youtube.com/watch?v=UxSNOtv8Rjw>
InChI and the Islands <http://www.youtube.com/watch?v=qrCqJ0o4jGs>

Robert Lancashire and Antony Williams, Symposium Organizers

Multidisciplinary Program Planning Group (MPPG)



The general theme of the 246th ACS National Meeting in Indianapolis, September 8 -12, 2013, was “Chemistry in Motion” as a nod to Indianapolis as the site of the “Indy 500.” ACS staff did a great job in advertising the theme with posters, flyers, and inserts alerting attendees to the thematic highlights of the meeting and showing the meeting logo with a racing-themed periodic table. Thematic program organizer, Professor Robert Weiss, The University of Akron, broadened the theme by including the concept of “Driving Innovation.” Sixteen divisions, including CINF, participated with theme-related symposia. The Indianapolis Local Section of ACS held an event at the Speedway and PMSE, POLY, ENFL, and SOCED organized a joint symposium on “The Chemistry of Racing” with the keynote address by retired race car driver Stephan Gregoire.

MPPG was also involved in the selection of the speakers for “The Kavli Foundation Innovation in Chemistry Lecture” and the “The Kavli Foundation Emerging Leader in Chemistry Lecture.” Both outstanding lectures were presented to a full house with the majority being young chemists. Harry B. Gray, Arnold O. Beckman Professor of Chemistry, California Institute of Technology, talked about “Powering the planet with solar fuel” and Martin D. Burke, Associate Professor of Chemistry, Howard Hughes Medical Institute, University of Illinois at Urbana-Champaign, gave a lecture on “Making molecular prosthetics with a small molecule synthesizer.” The Kavli Foundation will continue the Emerging Leader in Chemistry Lecture for researchers under the age of 40 through 2016. Divisions have been solicited to send speaker nominations for the Emerging Leader Lecture Series at the Dallas meeting to MPPG. Each Division can nominate two candidates and any nominations from CINF certainly will help the Division’s visibility.

The Plenary Session organized by Robert Weiss again attracted a very large audience. It consisted of presentations by three eminent scientists addressing the broadened theme: Naomi J. Halas, Rice University, “Solar Steam: Discovery, mechanism, and applications in energy,” Daniel R. Kittle, Dow Agrosciences, LLC, “From lab bench to table top – science serving the needs of a growing world,” and Bret E. Huff, Eli Lilly and Company, “Continuous processing in the pharmaceutical industry.”

At the General Meeting of MPPG in Indianapolis, Michelle Buchanan, Oak Ridge National Lab, and Nitash P. Balsara, UC Berkeley, organizers for the upcoming ACS meeting in Dallas in March, outlined the thematic program “Chemistry and Materials for Energy.” Theme areas include: catalysis, harnessing solar energy, materials under extremes, materials in nuclear systems, electrical storage, new materials and systems for the grid, materials for energy efficiency, enhanced oil recovery and unconventional oil and gas, and CO₂ capture, utilization and storage. They also presented a slate of plenary speakers and candidates for the Kavli lectures. The organizers plan a local event, “Fuel Up!” at the Perot Museum with nine hands-on stations to experience and learn about alternative energies. Based on the information provided, we can look forward to a very interesting program. Don’t miss the program announcements in *C&EN*!

The future of thematic programming at ACS meetings looks bright. More and more technical divisions organize symposia related to the theme of a meeting, often cosponsored by other divisions indicating the interdisciplinary nature of chemistry. Also, local sections become more and more involved. We definitely have seen a strong upwards trend in the last few meetings. As per charter, themes for the next three years have been approved and organizers are in place for 2014 and 2015. The Program

Committee of CINF should look closely at the themes and available synopses to work together with the thematic program chairs to organize companion symposia. Any symposium within a given theme will provide extra and valuable publicity to the Division.

Here are the themes for future meetings:

S2014 Dallas, TX: **Chemistry and Materials for Energy**. Thematic program chairs: Michelle Buchanan, Oak Ridge National Lab buchananmv@ornl.gov and Nitash Balsara, UC Berkeley nbalsara@berkeley.edu

F2014 San Francisco, CA: **Chemistry and Global Stewardship**. Thematic program chair: Robin Rogers, University of Alabama rdrogers@as.ua.edu

S2015 Denver, CO: **Chemistry of Natural Resources**. Thematic program chair: TBD

F2015 Boston, MA: **Innovation from Discovery to Application**.

S2016 San Diego, CA: **Computers in Chemistry**.

F2016 Philadelphia, PA: **Chemistry of the People, by the People and for the People**.

S2017 San Francisco, CA: **Water and Chemistry** (proposed).

F2017: Washington, DC: **Chemistry and Globalization** (proposed).

The CINF Program Chair will be notified about the details of these themes as soon as soon as they become available.

Guenter Grethe, CINF representative to MPPG

Join Us Again in Dallas



Image credit: <http://www.acs.org/content/acs/en/meetings/spring-2014.html>

Registration & Housing will open mid-December, 2013

Interview

Evolving roles of librarians in data management and curation:



An Interview with Kristin Briney, 2013 Lucille Wert Scholarship Winner

Bio: Kristin Briney has recently obtained a M.A. in library and information science from the University of Wisconsin-Madison (May 2013) and shortly after started in a temporary position as Data Services Librarian at the University of Wisconsin-Milwaukee. She is a holder of a PhD in physical chemistry from the University of Wisconsin-Madison (2010) and a B.A. in chemistry and computer science from DePauw University (2005).

Svetlana Korolev: Kristin, please accept my sincere congratulations on being the recipient of the 2013 Lucille Wert Scholarship awarded to you by Division of Chemical Information. The award announcement says that you are planning to combine your advanced scientific background in physical chemistry with knowledge of library and information science to tackle challenges in the burgeoning field of data curation. What has brought you to the realization of such professional interest? Has it been triggered by a person, an event, or something else?

Kristin Briney: It has long been my goal to do something at the intersection of science and technology but it wasn't until recently that I figured out this was data curation. I was never a great laboratory scientist and much preferred hacking at my data, thinking about e-lab notebooks, and engaging other scientists to handle solvents. Happily, all of these interests align with issues in data management.

I found out about data curation when I was looking into career options after finishing my PhD. I was speaking with Emily Wixson, the now retired chemistry librarian at UW-Madison, about science librarianship and she mentioned this new growing area called data curation. At that point, everything started to fall into place. I really credit Emily for introducing me to this field and for giving me a push into this career by connecting me with the right people.

SK: Speaking of the "burgeoning field of data curation" and, coincidentally, cheering for Antony Williams, 2013 CINF Chair, on his recognition with the 2012 [Jim Gray eScience Award](#) for making chemistry publically available through collective action via ChemSpider, let me ask you about the terminology of the field. Since this is an evolving methodology, which terms do you use when talking to chemical scientists for describing your services?

KB: This is a good question. The area in which I work is a fairly new space so the terminology is still settling and being spread outside of the community. I consider my work to be in "data management," which I think is a more accessible way of saying "data curation."

Data curation is related to eScience, in that they both focus on digital data, but there are distinct differences. Data curation is about the organization, management, and preservation of research data, while eScience concerns the new modes of scientific research that can be done such as data mining because data are now digital. Good data management enables better eScience, but I don't actually do any scientific research. I've recently read an article [Thoughts on "eResearch": a Scientist's Perspective](#) by Amanda Whitmire that discusses the etymology of "eScience" and "eResearch." It attributes the term "eScience" to John Taylor. If you are interested in eScience, I recommend the

books [*Reinventing Discovery: The New Era of Networked Science*](#) by Michael Nielsen and [*The Fourth Paradigm: Data-Intensive Scientific Discovery*](#) edited by Tony Hey, Stewart Tansley, and Kristin Tolle.

Complicating this terminology is that there are other newish fields related to data curation, such as digital preservation, open data, and open notebook science. Digital preservation focuses on the retention of digital information, research data included, over the long term. Open data is the growing movement to disseminate datasets along with their published articles. Open notebook science takes this concept further by putting laboratory notebooks online in real time. The latter two terms fall under the umbrella of “open science.”

Part of what I like about data management is that it borrows so heavily from other fields, not limited to the ones listed here. Things are still in flux, which makes it an exciting time to be in this field.

SK: Kristin, you are the most recent graduate from the School of Library and Information Studies, University of Wisconsin – Madison (May 2013). What is the current state of the curriculum supporting campus research needs and someone’s specialization in the data management processes, especially focusing on natural sciences? Have you benefited from taking any courses on this subject? What advice could you give to individuals wishing to get educated on this matter?

KB: Data curation has been around long enough that there is some established coursework in the area, but very few whole programs. I know of two universities that offer MLIS degrees specifically in data curation: the University of Illinois at Urbana-Champaign and the University of North Carolina at Chapel Hill. Other universities offer one, some, or no data curation classes, depending on the program.

I went to a more traditional library school and had to create my own path through the curriculum. I had only one class in digital curation, but was able to piece together a useful program from classes both inside and outside of the department. Having a flexible advisor really helped.

Honestly, the most important part of my library school experience was working on data curation projects outside of the classroom. I spent a year embedded in a microscopy laboratory focusing on data management issues, a semester teaching information literacy and good data management practices, and another semester on a team building a data management system for a virtual reality laboratory. Each of these experiences added to my understanding of data management at a deeper level than I could ever get in the classroom. So if I had one piece of advice for anyone going into data curation, it would be to get as much hands-on experience as you can.

SK: Based on your job hunting experience have you noticed whether many libraries are recruiting for a similar position as Data Services Librarian? What are the common titles for such positions and the skills required of librarians in this new area? Were you able to find electronic discussion groups or other networking opportunities? Which professional societies could be relevant for your specialization?

KB: It is a really good time to be in data management because many universities are looking to establish data services. These jobs are listed under many names: Data Services Librarian, Data Librarian, E-Science Librarian, Research Data Librarian, etc. I have also seen quite a few “science librarian” positions in which a portion of the job responsibilities pertain to data curation.

Data curation jobs come with a variety of required skills, most often: an MLS, data curation experience, research experience or an advanced degree, various technical skills, communication skills, and project management experience. Beyond matching up skills in a job post, I also look to work in an environment that is open in to new initiatives. There is no one standard way to address data problems, meaning that data managers must try new things (risking the occasional failure) and build incremental progress. Being in a supportive environment makes a big difference to the ultimate success of data services.

Now that I'm in a data curation position, it's important for me to keep up with the field because new things are always coming up. I can't recommend Twitter enough here. Twitter lets me network with peers in my field (many of whom are heavy Twitter users), keep up with the latest articles, and get immediate feedback when I'm stuck on a data problem. Besides Twitter, I subscribe to several listserves (acr-idgc-l, asis-l, sts-l, and chminf-l) and am looking forward to going to the [Research Data Access and Preservation Summit](#) this spring. There is really no one place to talk about data curation at the moment, so I'm always on the lookout for new forums for discussion.

SK: What are the major trends and opportunities for librarians to be involved in the lifecycle of the scholarly creation and management of data? At what stage of the research process do you think the librarians could contribute to this process? Could you envision possible new activities evolving in five years?

KB: I think that academic libraries are going to become much more involved with research data as data dissemination becomes a regular part of the scholarly process. I anticipate that in the near future we will be helping patrons find and cite research datasets in the way that we currently help them find and cite the data's corresponding journal articles.

But to be the leaders in finding and citing data, we need to take part in solving the current data problem: disorganized and mismanaged data. It's not a role that we have traditionally played, but it is an area where we have important skills: organization, documentation via metadata, and preservation. We risk losing our relevance as experts in these areas if we ignore the data problem entirely. Additionally, being involved early in the data management process lets us shape the data dissemination systems that we will be soon helping our patrons use.

SK: How are you embracing the first tasks in the newly created Data Services Librarian position at the University of Wisconsin - Milwaukee? What projects have you been working on recently?

KB: One of my big focuses is on outreach and education. In particular, I don't think that we are adequately preparing students to manage data well once they become working scientists. I remember being incredibly frustrated by dealing with data when I was a grad student in chemistry and that's an experience that I don't want other students to have.

So data management training is one of the two services I'm starting up in my new position at UWM (the other is data management plan consultations). In addition to the formal sessions I'm planning, I'm hoping to have a lot of informal discussions about data management with students and faculty on campus. I have found that once you start talking with researchers they intrinsically understand the problems around data management, but they don't necessarily know the solutions. That's where I come in. It's my goal to engage with and help as many researchers along as many avenues as I possibly can.

SK: Tell us something about yourself. Do you enjoy living in Wisconsin? What hobbies do you have?

KB: I have been living in Wisconsin for eight years now, but only just moved to Milwaukee. I have really enjoyed learning the culture of this state and the rich history of Milwaukee. At this point, I have been on more brewery tours than I can count and have come to believe that one should always live near a lake of some sort.

When I'm not thinking about data, I like to go out biking. That is wonderful for about half the year in Wisconsin and for the other half I have wool and knitting. My other big hobby is blogging. My knitting blog has been around for over 5 years and I just started a blog called "Data Ab Initio" that aims to demystify data management for researchers. The data blog has been a very fun project and has really helped my own understanding of many data management issues.

SK: Kristin, thank you very much for an inspiring discussion of the opportunities for librarians to get involved in data management and curation. Once again, congratulations on being the recipient of the 2013 Lucille Wert Scholarship and best wishes on your endeavors!

Additional Information:

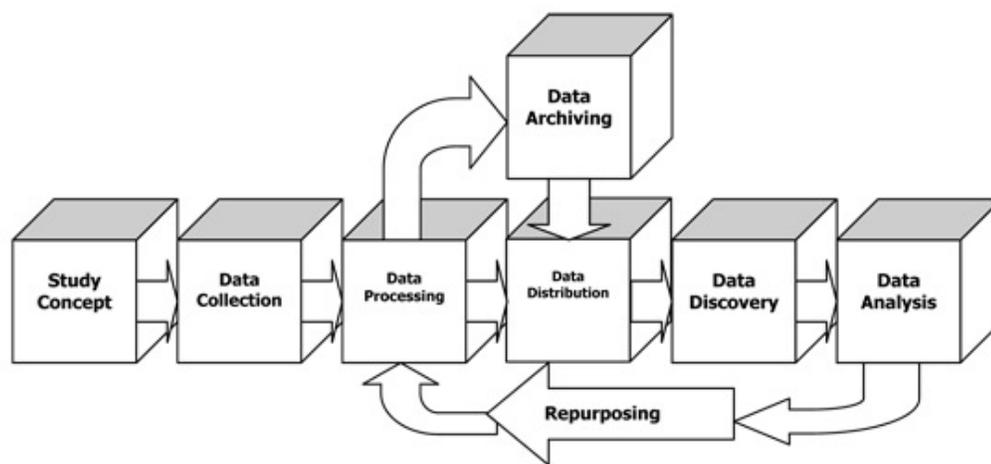
New England E-Science Portal <http://esciencelibrary.umassmed.edu/index>
Contains information on data management and links to many data management resources

E-Science Thesaurus <http://esciencelibrary.umassmed.edu/thesaurus>
Defines many terms related to data management

MANTRA <http://datalib.edina.ac.uk/mantra/>
A set of online tutorials on a range of data management topics

Federal Research Data Requirements Set to Change <http://crln.acrl.org/content/74/8/421.short>
This paper examines current and future US policies about data

DMPTool <https://dmp.cdlib.org/>
A useful tool for creating data management plans for several different funding agencies



Data Life Cycle Model <http://esciencelibrary.umassmed.edu/thesaurus/data-lifecycle>

Book Reviews

CRC Handbook of Chemistry and Physics Celebrates its 100th Anniversary

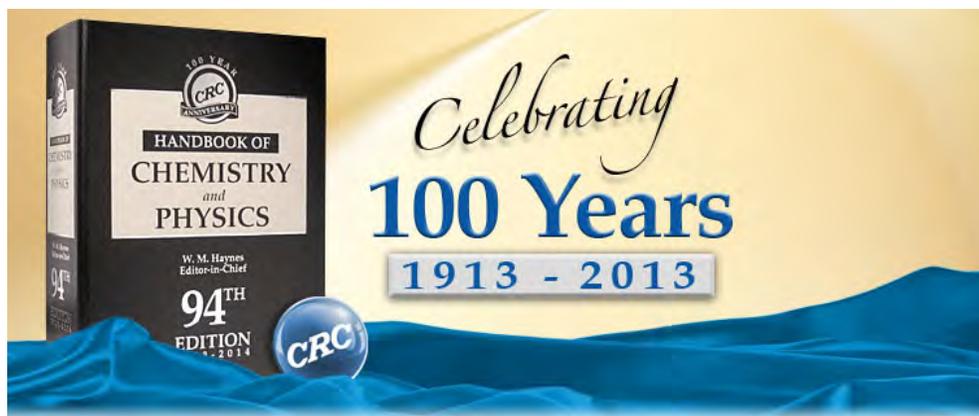


Image credit: <http://www.crcpress.com>

With publication of the 94th (2013-2014) Edition of the *CRC Handbook of Chemistry and Physics* this year, the Handbook celebrates its 100th Anniversary. The first edition of the Handbook was published in 1913. It was a small, pocket-size document of 116 pages. The publisher of the first volume was a family owned Cleveland company that sold chemical laboratory supplies and that remained in control of the Handbook until 1987. The Preface in the first volume stated:

“In compliance with the requests of hundreds of our friends for a small but comprehensive book of reference on chemical and physical topics, we have designed and compiled this pocket manual of Chemistry and Physics.”

The Handbook came to be commonly known as the “Rubber Bible” because as Sir William Wakeham stated in the Foreword of the 92nd Edition, “it seemed to us to contain the collated data of science.” The Handbook has continued to be published and updated on an annual basis except for missing some editions during the world wars of the past century.

The Handbook has had only five editors over its lifetime. The first Editor was William R. Veazey, a professor of chemistry at the Case School of Applied Science, now Case Western Reserve University. He was succeeded by Charles D. Hodgman, professor of physics at Case, who led the Handbook for almost 50 years from 1915 to 1963. Next, Robert C. Weast, professor of chemistry at Case, served as editor for 25 years from 1964 to 1989, and David R. Lide of the National Institute of Standards and Technology (NIST) served for 20 years from 1990 to 2009. W. M. Haynes of NIST has served as Editor since 2010. The relative stability and broad interests of the editors have played key roles in the successful transition of the Handbook over a number of generations of students and scientists. Another important factor to the success of the Handbook has been the contributions of a worldwide network of outstanding scientists who have provided input of the highest quality in their own areas of expertise on a continuing basis.

After the publication of the Handbook was controlled by a single company for almost 75 years, the company was acquired by Time Mirror Co. in 1987. Subsequent owners, through the present publisher, Taylor and Francis Group, have recognized the Handbook as a flagship publication and have provided the strong support necessary to maintain its high standards and enable it to evolve into the internet age while meeting the changing needs of its users. The current hardcopy edition of the *CRC Handbook of Chemistry and Physics* comprises 2,600 pages of critically evaluated data; it is also available in eBook and interactive online formats.

The *CRC Handbook of Chemistry and Physics* has become an essential resource for students and scientists worldwide for physical and chemical data and information on related topics, such as biochemistry, geophysics, astronomy, and environmental science. It has attained a position as the first source for technical data, especially for those seeking information in areas outside their own area of expertise, but which they need to bridge a scientific gap. A high standard of quality has been a hallmark for the Handbook over the past century in providing critically evaluated data with reliable sources of documentation and in continually updating and expanding the coverage of the diverse subject matter in the Handbook consistent with advances in science and technology. Few could argue with the claim that the *CRC Handbook of Chemistry and Physics* has been the single most widely used source of physical and chemical data over the past century.

As the *CRC Handbook of Chemistry and Physics* enters its second century, it is expected that, based on its current reputation and its long-term history, it will continue to serve as the single most trusted source for chemical and physical data. Efforts will continue to make it more user friendly in terms of delivery mechanisms, to expand its coverage to meet the needs of the next generations of students and scientists, and to maintain the highest standards of quality control of the data in terms of the reliability of the information and its documentation.

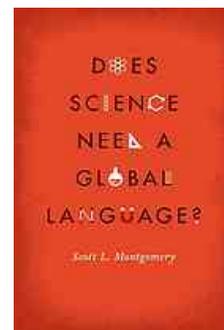
Mickey Haynes, Editor-in-Chief, CRC Handbook of Chemistry and Physics

Handbook of Chemistry and Physics. A ready-reference pocket book of chemical and physical data. Fifth edition. Compiled from the most recent authoritative sources. The Chemical Rubber Co., Cleveland, Ohio. 414 pp. Price, \$2.00.

This handy book contains data in part of the same sort as those which are given in "Chemiker-Kalender," including Five-place Logarithms, Logarithms of the Trigonometrical Functions, Wire Tables, Comparison of Customary and Metric Units, Physical constants of inorganic and organic compound, Gravimetric factors, Densities of acids, bases and alcohol, Definitions and formulas of physical chemistry and physics and many other valuable tables of physical and chemical properties. It has the advantage over the "Chemiker-Kalender" that all is bound with a single cover. As the book is sold to Clubs of students at \$1.00 it should have a very wide sale in chemical laboratories. W. A. NOYES.

NEW BOOKS. *J. Am. Chem. Soc.*, 1917, 39 (4), pp 837–840

Montgomery, Scott L. *Does Science Need a Global Language?* University of Chicago Press, Chicago, 2013; pp. 1-226 + xiii, ISBN 978-0-226-53503-6 (hardcover). \$22.50.



The obvious rise of English as the predominant language in the communication of science has generated quite a bit of commentary in recent years. Beginning with the editorials and publishing practices of Gene Garfield in the '60s (he not only observed the rise of English, but championed it), the first books on the subject appeared in the 90's. The predominance of English in the sciences parallels its predominance in other areas, especially in the arts. In fact, many non-native speakers and authors of English cite movies and TV as their initial exposure to English. Language is indeed power and English has become the near universal language of international scientific conferences even when held in areas where English is not the primary language. Written publications in English, especially scientific journals, have lagged somewhat, but the number of both journals and articles in English is on the increase. As a result, most scientific communication in English is between non-native speakers. The history of the languages of scientific communication is discussed and the rise of English into predominance has been relatively rapid, similar to the rise of Arabic in the first millennium. Has English become the *Lingua Franca* of science, similar to Latin and Arabic in the past? It's not there yet, but may be on the way. Data and discussion are given for the number of English speakers and the number of countries represented as well as data and discussion on global education in and of English. Unfortunately, scientists who are native language speakers of English lag behind the rest of the world in bi- or multilingual capabilities.

Pros and cons of global scientific English are discussed, including Brain Drains, supplantation and even suppression of local languages, perceived hegemony, etc. The latter is not deemed to be happening and communication in local languages is actually encouraged, although global publication of results in English is not only preferred but essential. Occasional fears of non-English cultural suppression are probably unfounded. Translation, especially machine translation, has an effect on the trend, but the author deems the latter to not yet be accurate enough to facilitate scientific publication. Scientific publication globally has shifted more toward for-profit publishers. Anglo-American English predominates in publication, especially at the hands of editors, many of whom have market and profit motives, and tends to suppress the omnipresent non-standard forms of English. Those versions, being far more common, may have to be accommodated in the future.

Previous *Lingua Francas* or other dominant languages of science have been supplanted, so what is the future of English in that exalted state? Chinese is currently touted as a possibility, but that probably will not happen given the trend for the Chinese to publish in English.

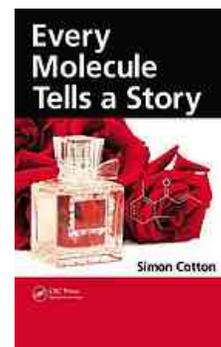
This reviewer found some topics lacking in discussion, including the rise and effects of open publishing and the hindrance of differing alphabets in learning any new language. However, the book is an excellent treatment of topics very important to scientific research, communication, and education in general. Highly recommended. In the last section, the answer to the title question is given and the answer is "yes."

Bob Buntrock, Member, CINF Communications and Publications Committee

Cotton, Simon, *Every Molecule Tells a Story*, CRC Press, Boca Raton FL, 2012; pp. 266 + x, ISBN 978-1-4398-00-0073-6 (hardcover). \$62.95.

The author describes this book neither as a textbook nor as a collection of reviews. It is a collection of essays on more than 200 chemical compounds. It is probably aimed at the lay public, but would also find value in schools and colleges.

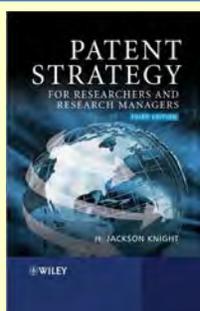
The chemistry is reasonably sophisticated with many structures. The chemical and topical essays are grouped in fourteen chapters. Titles include atmosphere and water, carbohydrates and artificial sweeteners, hydrocarbons, acids and alkalis, steroids and sex, the senses, cosmetics and perfumes, natural killers, unnatural killers, explosives, pleasure molecules (alcohol, nicotine, designer drugs), natural healers, unnatural healers, and synthetic polymers.



The terminology is British and in some cases misleading. “Killers” would better be described as toxins, especially since at least one, thalidomide, is not a “killer” but a notorious fetotoxin. “Healers” are better described as medicinals or pharmaceuticals. Several (in)famous compounds are conspicuous by their absence, including BPA and dioxin (TCDD). Dimethylmercury is described, but methylmercuric cation is not.

The bibliography lists additional resources on general topics discussed, as well as resources for each chapter and some specific chemicals. Recommended for high school and college libraries, and educators.

Bob Buntrock, Member, CINF Communications and Publications Committee



[Patent Strategy for Researchers and Research Managers](#)

Knight, Jackson. 3rd Edition (February 2013), Hoboken, NJ: Wiley. ISBN: 978-0-470-05774-2 (hardcover, also available paperback and ebook), 256 pages, US \$90.00.

“The book is designed to be a ‘how to’ manual rather than a guide to patent law. Aimed at the researcher or technology manager, it explains how to use the patent system to best advantage for commercial gain. There is a strong focus on business throughout the book – explanations of legal concepts are pragmatic rather than academic, and the insightful advice evidently draws on personal experience... While there is no substitute for experience, this book is possibly the next best thing.”
(*Chemistry & Industry*, April 2013)

Good book reviews were brought to our attention by Wendy Warr

Committee Reports

ACS Council Meeting

The Council of the American Chemical Society met in Indianapolis, IN on Wednesday, September 11, 2013. The following report is adapted from the ACS Office of the Secretary's Councilor Talking Points. Items of particular interest to members of the Division of Chemical Information have been selected for inclusion in this report.

ACTIONS OF THE COUNCIL

Election Results

Elected to the **Committee on Committees** beginning in 2014: Janet L. Bryant, Dee Ann Casteel, Amber S. Hinkle, Wayne E. Jones, Jr., and V. Michael Mautino for the 2014-2016 term.

Elected to the **Council Policy Committee** beginning in 2014: Harmon B. Abrahamson, Judith H. Cohen, Alan M. Ehrlich, and Angela K. Wilson for the 2014-2016 term.

Elected to the **Committee on Nominations and Elections** beginning in 2014: Lisa M. Balbes, Jeannette E. Brown, Martha L. Casey, D. Richard Cobb, and Lissa Dulany for the 2014-2016 term.

Candidates for President-Elect and Board of Directors

The candidates for the fall 2013 ACS national election were announced as follows:

Candidates for President-Elect, 2014

Dr. G. Bryan Balazs, Associate Program Leader, Lawrence Livermore National Lab, Livermore, CA
Dr. Charles E. Kolb, Jr., President and CEO, Aerodyne Research Inc., Billerica, MA
Dr. Diane Grob Schmidt, Section Head R&D, The Procter & Gamble Company, Cincinnati, OH

Candidates for Directors-at-Large, 2014-2016 (two will be elected)

Dr. Susan B. Butts, Independent Consultant, Susan Butts Consulting, Midland, MI
Dr. Thom H. Dunning, Jr., Director, National Center for Supercomputing Applications and Professor, Distinguished Chair for Research, University of Illinois at Urbana-Champaign, Urbana, IL
Dr. Dorothy J. Phillips, Retired, Waters Corporation, Milford, MA
Dr. Kathleen M. Schulz, President, Business Results, Inc., Albuquerque, NM

Candidates for District II Director, 2014-2016

Dr. George M. Bodner, Arthur Kelly Distinguished Professor of Chemistry Education and Engineering, Purdue University, West Lafayette, IN
Dr. Alan A. Hazari, Director of Chemistry Labs and Lecturer, University of Tennessee, Knoxville, TN

Candidates for District IV Director, 2014-2016

Dr. Rigoberto Hernandez, Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA

Dr. Larry K. Krannich, Professor Emeritus of Chemistry, University of Alabama, Birmingham, AL

Committee Reviews and Committee Charters

Council voted to continue the joint Board-Council Committee on International Activities and the Committee on Nomenclature, Terminology and Symbols. Continuation of the Committee on International Activities also requires Board of Directors concurrence.

The Council voted to approve amendments to the charters of the committees on International Activities and on Nomenclature, Terminology and Symbols.

Selected Committee Reports (Highlights)

Society Committee on Education

SOCED recognized the achievement that is present in the Next Generation Science Standards, including their basis in research on teaching and learning, their formulation as performance standards, and their basis in the NRC framework and its dimensions of Science and Engineering Practices, Disciplinary Core Ideas, and Cross Cutting Concepts. SOCED supported the Standards as a document that is broadly applicable as a basis for K-12 science instruction and called upon the Society to develop innovative programming to support the implementation of the Standards.

Standing Committee on Economic and Professional Affairs (CEPA)

Committee on Economic and Professional Affairs reported that employment is up and unemployment is down for ACS chemists. The complete review of the Comprehensive Salary Survey will appear in the September 23 issue of *C&EN*.

Realignment of Electoral Districts

ACS Bylaws require that the six electoral districts – from which six directors are elected to the ACS Board of Directors – be balanced in their total member populations. The Council voted to approve a proposal by the Committee on Nominations and Elections (N&E) to realign these districts. The realignment meets the specified criteria for redistricting as required by Bylaw V, Section 4a and brings all six districts within permissible population range. This change takes place in 2014 and does not affect the 2013 national ACS elections. Councilors and others may visit the N&E website to look at the actual proposal and its impact.

Meeting Registration Report

As of the morning of September 11, 2013, the ACS fall national meeting had attracted 10,840 registrants, including 6,630 regular attendees and 2,584 students. The history of attendance at ACS fall national meetings since 2004 is as follows:

2004 Philadelphia, PA	14,025
2005 Washington, DC	13,148
2006 San Francisco, CA	15,714
2007 Boston, MA	15,554
2008 Philadelphia, PA	13,805

2009 Washington, DC	14,129
2010 Boston, MA	14,151
2011 Denver, CO	10,076
2012 Philadelphia, PA	13,251
2013 Indianapolis, IN	10,840

Local Sections

The Council voted, on the recommendation of the Committee on Local Section Activities (LSAC), to approve a request from the Syracuse Local Section to change its name to the Central New York Local Section. Council also approved a recommendation from LSAC that the Monmouth County Local Section (in New Jersey) be dissolved, effective January 1, 2014, due to a decline in activity over the last several years. The North Jersey Local Section has contacted LSAC and will submit a petition in 2014 to annex the Monmouth County territory.

Divisions

After much debate, a proposed name change for the Division of Colloid and Surface Chemistry to the Division of Colloids, Surfaces, and Nanomaterials was defeated by the Council in a close vote.

Special Discussion Item

ACS President Marinda Wu presented and moderated a discussion on “What can we – as the Society and as individual citizens – do to help create jobs or demand for chemists?” She shared five recommendations from the presidential task force “Vision 2025: Helping ACS Members to Thrive in the Global Chemistry Enterprise” and what they might imply for our efforts to help create jobs: discover and share information about the skills and competencies that a wide range of employers will need; continue to expand resources which help our members to position themselves for successful careers in the global chemistry enterprise; enable entrepreneurs to create and strengthen their startups that hire chemistry professionals; advocate for policies that improve the business climate and promote the creation of chemistry jobs; and work with other stakeholders to understand and influence the supply and demand of chemists and jobs. Following the presentation, numerous Councilors engaged in a discussion of this focused topic on possibilities to encourage jobs creation and offered several suggestions.

ACTIONS OF THE BOARD OF DIRECTORS

At this meeting, the ACS Board of Directors considered a number of key strategic issues and responded with several actions.

The members of the Division of Chemical Information may be particularly interested in the specific bullets below regarding the formation of a National Association of Chemistry Teachers and the appointment of Manny Guzman as the President of Chemical Abstracts Service (CAS).

The Board’s Committees and Working Groups

The Board held a discussion on the topic “Connecting Chemists with Each Other.” It considered what the role of ACS should be in helping chemists develop relationships with other chemists and the strategies that enable those relationships; how these strategies encourage and support younger and

international members; and how these strategies develop relationships to leverage the world renowned chemists/innovators who comprise our membership.

On the recommendation of the Committee on Grants and Awards, the Board voted to approve Society nominations for the National Science Board's Public Service Award and the National Science Foundation's (NSF) Alan T. Waterman Award. The National Science Board's Public Service Award honors individuals and groups who have made substantial contributions toward increasing public understanding of science and engineering in the US. The Alan T. Waterman Award recognizes an outstanding young researcher in any field of science or engineering supported by the NSF.

The working group on Society Program Portfolio Management briefed the Board on its activities. The working group is charged with delivering a process for portfolio management of Society programs in the divisions of Membership and Scientific Advancement, Education, and the Office of the Secretary and General Counsel (Office of Public Affairs) and pilot programs.

The Board received a briefing and approved a recommendation from its Committee on Executive Compensation. The compensation of the Society's executive staff receives regular review from the Board.

On the recommendation of the Committee on Budget and Finance (B&F), the Board voted to approve an advance member registration fee of \$380 for national meetings held in 2014. The Board also voted to reauthorize funding in next year's proposed budget for the ACS International Center, and the ACS Entrepreneurial Initiative, and to authorize funding for a new initiative, the National Association of Chemistry Teachers (NACT). This association will be an ACS program to provide teachers a professional home. Through NACT they will have access to specialized resources and the broader ACS community.

The Board confirmed the recommendation of the ACS Executive Director/CEO of the new President of Chemical Abstracts Service (CAS). He is Manuel (Manny) Guzman, most recently Executive Vice President of Learning and Research Solutions of Cengage Learning. Mr. Guzman succeeds Robert J. Massie, who is retiring in March 2014 after leading CAS for 21 years. Mr. Guzman will begin September 30. Mr. Massie will assist in the transition when he returns from medical leave.

The Executive Director/CEO Report

The Executive Director/CEO and her direct reports updated the Board on the following: highlights and high-level recommendations on the ACS global presence; and the activities of CAS (Chemical Abstracts Service) and the ACS Publications Division. As a follow-up to the Publications report, the Board voted to approve one journal editor appointment and several editor re-appointments.

Other Society Business

The Board also voted to hold the December 2015 Board of Directors meeting in Honolulu, Hawaii, in conjunction with the 2015 International Chemical Congress of Pacific Basin Societies (Pacifichem). The ACS is the host society for the 2015 Pacifichem meeting, and co-location will allow Board members to participate in this very successful Pacifichem meeting.

Andrea Twiss-Brooks and Bonnie Lawlor, CINF Councilors

Highlights from the Joint Board-Council Committee on Publications

The open session of the ACS Joint Board-Council Committee on Publications (JBCCP) is usually scheduled on Friday afternoon before each National Meeting and is open to any society members. The President of the ACS Publications Division, Brian Crawford, provides an update on the activities of the Division over the past 6 months to the Committee and to other ACS members in attendance. Highlights from the Indianapolis presentation are posted [here](#). Questions regarding the Committee may be directed to the Committee Chair, Stephanie Brock (sbrock@chem.wayne.edu). Leah McEwen, CINF Secretary, has ended her term on the Committee. The following summary and slides were kindly provided by Debra Davis, ACS Staff Secretary of the Committee (d_davis@acs.org).

It was announced that three Editor Search Committees had recommended, and the ACS Board of Directors had approved the following appointments:

- Dr. Harry A. Atwater (California Institute of Technology) to serve as the inaugural Editor of the new journal, *ACS Photonics*.
- Dr. Vincent M. Rotello (University of Massachusetts – Amherst) to serve as the new Editor of *Bioconjugate Chemistry*.
- Dr. Phillip E. Savage (University of Michigan) to serve as the new Editor of *Industrial & Engineering Chemistry Research*.

The staff report demonstrated that *Chemical & Engineering News (C&EN)* continues to fulfill its mission. *C&EN's* Strategic Plan for 2013-2017 was presented and discussed. *C&EN* marked its 90th anniversary during the Indianapolis meeting with a variety of activities, including a special issue on how chemistry has changed the world (September 9, 2013), a webinar on “Food Fraud: How Scientists Detect It & What You Should Know,” and a performance by celebrity chef Alton Brown. Print subscribers received a poster supplement depicting highlights in the recent history of the chemical enterprise.

The Division President provided informational updates on recent allegations concerning ethical violations by authors published in ACS journals, open access publishing plans approved by the Society's Governing Board for Publishing, and guidance provided to ACS Editors regarding sanctions by the US Office of Foreign Assets Control (OFAC) that affect the peer review of manuscripts submitted by authors who are employees of the Government of Iran.

Debra Davis, Staff Secretary, Joint Board-Council Committee on Publications

The following is a selected list of URLs presented on slides at the ACS Council meeting.

www.my.acs.org Showcases stories and photos submitted by members describing what best defines their ACS membership experience. If your story is selected, you receive a T-shirt.

www.acs.org/getinvolved ACS offers many ways to get involved at the local, regional, and national levels. There are opportunities for everyone, whether you are a student or a seasoned professional.

www.ACS.org/ChemistryAmbassadors Visit the Chemistry Ambassadors website for ideas and resources to engage your community with positive messages about chemistry.

CINF Social Networking Events

Once again, the Division of Chemical Information put together a set of great social events, starting with the Welcoming Reception on Sunday. The reception was attended by about 80 members and friends. Helped along by a lovely selection of food and drinks, it was great to see everybody mingling and chatting. As before, the Welcoming Reception hosted the CINF Scholarship for Scientific Excellence poster session (sponsored by the [Royal Society of Chemistry](#)) and this meeting saw two posters being presented. The Division is grateful to the sponsors for this event: [Bio-Rad Laboratories](#), [InfoChem](#), [Optibrium](#), [PerkinElmer](#), [Thieme Chemistry](#), and [ACS Symposium Series](#).

Monday saw two events on the CINF social calendar. During the afternoon RSC eScience team hosted a workshop as part of the “Social Networking” symposium organized by Antony Williams and Jennifer Maclachlan. In the evening the Division hosted Harry’s Party, sponsored exclusively by [ACS Publications](#). As in the tradition of this famous Party (see an [article](#) about Harry’s Party in *Chemical Information Bulletin*, Spring 2013, and a [page](#) about Harry’s Party at CINF website), it was arranged in a small suite with a good view, excellent snacks and drinks, and stimulating conversations among about 60 attendees.

Tuesday featured the Division of Chemical Information Luncheon held at the Convention Center, exclusively sponsored by the [Royal Society of Chemistry](#). A superb lunch menu was accompanied by a very well received talk by the invited speaker, Prof. Katy Borner of Indiana University, who spoke about “Multi-Scale Maps of Scholarly Activity” ([slides](#)). During the Luncheon, the CINF Scholarship for Excellence awards were presented to the winners: Abhik Seal and Johannes Hachmann. Congratulations to the winners!

Tuesday evening saw the Herman Skolnik Reception held at the JW Marriott, honoring the 2013 awardee, Dr. Richard Cramer. Attended by more than 80 people, this was a great opportunity for guests to meet Dr. Cramer and interact with the speakers from the Symposium held earlier in the day. The Symposium and Reception were generously co-sponsored by [Certara](#), [Infochem](#), [Novartis](#), and the [Journal of Cheminformatics](#).

As always, many thanks to our Division colleagues, Leah McEwen and Judith Currano, who put in effort and time for making room arrangements and menu orders for our social events. And, of course, we all greatly appreciate our generous sponsors, without whom we would not have been able to put on such a great line-up of events in Indianapolis.

I anticipate another set of great social networking events at the next Spring 2014 ACS Meeting and hope to see you all there.

Rajarshi Guha, Chair, CINF Fundraising Committee

CINF photos from the Fall 2013 ACS National Meeting are at:

<http://www.flickr.com/photos/cinf/>

Photos by Wendy Warr

CINF Officers and Functionaries

Chair:

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

Chair Elect:

Ms. Judith Currano
University of Pennsylvania Chemistry Library
231 S. 34th St. 5th Floor
Philadelphia, PA 19104
215-746-5886 (voice)
215-898-0741 (fax)
currano@pobox.upenn.edu

Past Chair/Nominating 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

Secretary:

Ms. Leah McEwen
Cornell University Clark Library
283 Clark Hall
Ithaca, NY 14853
607-796-6217 (voice)
607-255-5288 (fax)
leah.solla@cornell.edu

Treasurer:

Dr. Rob McFarland
Washington University
Louderman Hall Room 549
1 Brookings Drive
Saint Louis, MO 63130
314-935-4818 (voice)
314-935-4778 (fax)
mcfarland@wustl.edu

Councilor:

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

Councilor:

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

Alternate Councilor:

Mr. Charles 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
510-865-5152 (voice and fax)
ggrethe@att.net

Program Committee Chair:

Mr. Jeremy Garritano
Purdue University
M. G. Mellon Library of Chemistry
504 West State Street
West Lafayette, IN 47907
765-496-7279 (voice)
jgarrita@purdue.edu

Membership Committee Chair:

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

Archivist/Historian:

Ms. Bonnie Lawlor
See Councilor

Audit Committee Chair: TBD**Awards Committee Chair:**

Ms. Andrea Twiss-Brooks
See Councilor

Careers Committee Chair: TBD**Chemical Information Bulletin Editors:**

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

Dr. Vincent Scalfani (Fall 2013)
University of Alabama
109 Rodgers Library
Tuscaloosa, AL 35487
205-348-5806 (voice)
vfscalfani@ua.edu

Communications and Publications Committee Chair:

Dr. David Martinsen
American Chemical Society
1155 16th St NW
Washington, DC 20036
202-452-2110 (voice)
d_martinsen@acs.org

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:

Ms. Grace Baysinger
Stanford University
Swain Library of Chemistry & Chem. Engin.
364 Lomita Dr., Stanford, CA 94305
650-725-1039 (voice)
650-725-2274 (fax)
graceb@stanford.edu

Finance Committee Chair:

Dr. Rob McFarland
See Treasurer

Fundraising Committee Chair:

Dr. Rajarshi Guha
See Past Chair/Nominating Chair

Tellers Committee Chair:

Ms. Susan Cardinal
University of Rochester
Carlson Library
Rochester, NY 14627
585-275-9007 (voice)
585-273-4656 (fax)
scardinal@library.rochester.edu

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

Contributors to this issue of *Chemical Information Bulletin*

Articles & special features

Martin Brändle
Kristin Briney
Bob Buntrock
Paul Clemons
Eric Dawson
Jeremy Garritano
Guenter Grethe
Rajarshi Guha
Mickey Haynes
Charles Huber
Svetlana Korolev
Robert Lancashire
Jennifer Maclachlan
David Martinsen
Tudor Oprea
William Town
Wendy Warr
Antony Williams

Awards & calls for nominations

Guenter Grethe
Bonnie Lawlor
Marge Matthews
Andrea Twiss-Brooks

Council & committee reports

Debra Davis
Rajarshi Guha
Bonnie Lawlor
Andrea Twiss-Brooks

Editing & production

Danielle Dennie
Svetlana Korolev
Bonnie Lawlor
Wendy Warr

Division of Chemical Information gratefully
acknowledges contribution from its sponsors
for the Fall 2013 ACS National Meeting

CERTARATM
Implementing Translational Science

 ACS Publications
MOST TRUSTED. MOST CITED. MOST READ.

BIO-RAD

 **REAXYS**[®]

 ROYAL SOCIETY
OF CHEMISTRY

 **accelrys**[®]

ACS
Symposium
SERIES

 **InfoChem**

 Journal of
Cheminformatics

 **KILMORIE**
CLARKE LTD

 **NOVARTIS**

 **optibrium**

 **PerkinElmer**
For the Better

 **Thieme Chemistry**

 **PID**
ANALYZERS