Chemical Information

BULLETIN

Spring 2010 Volume 62 No. 1

Division of Chemical Information

www.acscinf.org

Chemical Information Bulletin

A Publication of the Division of Chemical Information of the ACS

Volume 62 No. 1 (Spring) 2010

Svetla Baykoucheva, Editor
White Memorial Chemistry Library
1526 Chemistry Building, University of Maryland
College Park, Maryland 20742 USA
sbaykouc@umd.edu

IN THIS ISSUE

Message from the Chair	3
Letter from the Editor	4
CINF sponsors for the ACS Spring 2010 National Meeting	5
Technical Program Highlights	6
CINF Division meetings and social events	7
Technical Program Schedule	8
Awards and Scholarships	15
Future ACS national meetings	16
From ISI to NFAIS: Interview with Bonnie Lawlor	17
Past and present of FIZ CHEMIE Berlin: Interview with René Deplanque	24
On the periodic table of the elements and its significance: Interview with Eric Scerri	27
Book Review	33
Abstracts	34
CINF 2010 Officers	76

Photo on cover by Svetla Baykoucheva

CINF 2010 EXECUTIVE COMMITTEE

Chair:	Ms. Carmen Nitsche
Chair Elect:	Dr. Gregory M. Banik
Past Chair:	Ms. Svetlana Korolev
Secretary:	Ms. Leah R Solla
Treasurer:	Ms. Meghan Lafferty (2009-2010)
Program Chair:	Dr. Rajarshi Guha (2009-2010)
Membership Chair:	Ms. Jan Carver (2009-2011)
Councilors:	Ms. Bonnie Lawlor (2010-2012)
	Ms. Andrea Twiss-Brooks (2009-2011)
Alternate	Dr. Guenter Grethe (2010-2012)
Councilors:	Mr. Charles F. Huber (2009-2011)

ISSN: 0364-1910

Chemical Information Bulletin, © Copyright 2010 by the Division of Chemical Information of the American Chemical Society.

MESSAGE FROM THE CHAIR

Carmen Nitsche

Dear Colleagues—

Welcome to the first edition of the CINF bulletin that is exclusively electronic. It is part of our continuing efforts to embrace electronic rather than print activities and comes on the heels of another big step for our Division last year – electronic balloting. The CINF publications team spent 2009 preparing for this transition. Thanks to their efforts we will be able to continue to share the valuable information captured in our *Chemical Information Bulletin* while at the same time reducing our impact on the environment. The timing could not be better, given that the theme for the Spring National meeting is "Chemistry for a Sustainable World."

As I begin my tenure as Chair of CINF, I want to extend a warm thank you to Svetlana Korolev, our immediate past chair, for all her efforts last year, and to Dave Martinsen, who has come to the end of his CINF Chair rotation. I look forward to working with the CINF membership to advance the interests of our discipline, with particular focus on our interdisciplinary activities and member outreach and mentoring.

Our division exists to support our members, providing technical programming that spans the breadth of our discipline and a network of colleagues who are an invaluable resource. The planned CINF programming at the Spring National Meeting in San Francisco exemplifies our commitment to interdisciplinary approach, with CINF either sponsoring or co-sponsoring sessions with the INORG, MEDI, COMP, ANYL, CHED and CHAL Divisions of the ACS, the CSA Trust and the ACS Green Chemistry institute.

If you have never come to CINF committee meetings, make 2010 the year you get more involved in CINF activities. I invite you to join us at the CINF Long Range Planning meeting and breakfast on Saturday, March 20th. This is immediately followed by committee meetings until noon, a luncheon at noon, and the executive committee meeting in the afternoon. We also will have a CINF luncheon on Tuesday, and three receptions, including our Welcoming reception and Scholarship for Scientific Excellence Poster session on Sunday, the traditional Harry's party on Monday, and the CINF Reception on Tuesday.

See you in SF!

Carmen Nitsche
Chair, ACS Chemical Information Division
Carmen.Nitsche@symyx.com

LETTER FROM THE EDITOR

Svetla Baykoucheva

This first issue of the *Chemical Information Bulletin* (*CIB*), which will be published online, is still "a work in progress." Many readers of *CIB* have told us that they will miss the print, as they could carry it to the ACS National Meeting, check the Technical Program sessions, the times for the social events, and read the interviews. The print version of *CIB* has existed since 1949—we celebrated its 60th anniversary last year. In this first online issue we have preserved the format of the print, so that this change could be experienced more gradually and those who needed to carry it to the meeting could just print it out. The *Bulletin* will be adopting a more "online look" in the next issues.

Having the journal published online will allow us to provide more current information, as we will not be bound by the restrictions of a printing process. Also, the *Bulletin* will be available as an open-access publication that anybody can access. There will be more content, as we will not be concerned about the high cost of printing and mailing. We are starting a new column, **Book Reviews**, and I would like to thank our colleague **Bob Buntrock** for proposing it and for volunteering to manage it. **Rajarshi Guha**, the Chair of the Technical Programming Committee, has provided highlights of the technical program for the San Francisco meeting.

In this issue we have three interviews that reflect the views of three people with very different careers. In her interview, **Bonnie Lawlor** tells us about her role in the information industry and gives us an insight into the inner workings of the Institute for Scientific Information. ACS and CINF have established close relationships with German organizations involved in chemical information. Readers of this issue will learn not only about one of these organizations, FIZ CHEMIE Berlin, from an interview with its managing director, **René Deplanque**, but also about the past and the present of the chemical information field in Germany and how he feels about the unification of the country.

Some of you may have noticed that I am particularly interested in the history of chemistry. For this issue, I did an interview with **Eric Scerri**, who presents a philosopher's view on the Periodic Table of the Elements and talks about its significance.

I would like to mention that the interview with David Lide, former editor-in-chief of the *Handbook of Chemistry and Physics* that was included in the previous (Fall, 2009) issue of the *Bulletin* was also published in the November, 2009, issue of *Against the Grain* and can be viewed at www.against-the-grain.com/2010/01/atg-interviews-david-lide/

As we embark on a new venture with this issue, we would like to hear from you. Tell us what content you would like to see in the *Bulletin*, and what format would be most convenient to use and pleasing to read.

Enjoy the issue!

Svetla Baykoucheva, Editor
sbaykouc@umd.edu

CINF Sponsors for the ACS Spring 2010 National Meeting

The American Chemical Society Division of Chemical Information (CINF) is very fortunate to receive generous financial support from our sponsors to maintain the high quality of the Division's programming at the ACS Spring 2010 National Meeting in San Francisco to promote communication between members at social functions, and to support other divisional activities during the year, including scholarships to graduate students in Chemical Information.

The Division gratefully acknowledges contribution from the following sponsors:

Platinum Level:

Gold Level:

Bio-Rad Laboratories
Elsevier
Thieme Publishing Group

Silver Level:

FIZ CHEMIE Berlin

Bronze Level:

ACS Publications
Cambridge Crystallographic Data Centre
CambridgeSoft
Symyx Technologies

Contributor Level:

InfoChem

Opportunities are available to sponsor Division of Chemical Information events, speakers, and material. Our sponsors are acknowledged on the CINF web site, in the *Chemical Information Bulletin*, on printed meeting materials, and at any events for which we use your contribution.

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

Graham Douglas
Chair, Fundraising Committee
Email: Graham_C_Douglas@hotmail.com
Tel: 510-407-0769

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

ACS Chemical Information Division (CINF)
Spring 2010 ACS National Meeting
San Francisco, CA (March 21-25)
Technical Program Highlights

Rajarshi Guha
Chair, Technical Programming

With the Spring ACS coming up, it's my pleasure to highlight some of the technical programming that has been planned for this meeting. This year we have a packed program – approximately 140 papers in 24 sessions. One of the features of this meeting is the great diversity of topics being covered, which complements the diversity of the CINF membership!

Rajarshi Guha

We start off with a packed Sunday. The morning starts with Christoph Steinbeck who has organized a symposium on metabolomics focusing on infrastructure, analytical tools and methods. Concurrently, Jeremy Garritano and Judith Currano have put together a symposium focusing on faculty-library collaborations. On Sunday afternoon we will have the second instance of the Best Presentation Award Symposium. This will be organized by Roger Schenck and will focus on green chemistry and chemical information. The winner will receive an invitation to the CINF luncheon, and a plaque as well as $1,000 towards registration and expenses.

This meeting sees a number of new symposium organizers. On Sunday and Monday, Maciej Haranczyk and Berend Smit have put together a symposium on cheminformatics applications in materials chemistry, focusing on the use of cheminformatics techniques in the design of materials for hydrogen storage and solar energy. From Monday to Tuesday, Noel O'Boyle, Andrew Lang and Jean Claude Bradley have organized a symposium on visualization of chemical information. With the deluge of data from high throughput experiments and various databases, visualization is one of the ways in which one can gain insight from the raw data. This symposium will highlight new developments in visualizations of various chemical data types – ranging from molecular structure to SARs and documents.

As with previous Spring meetings, CINF hosts the CINF-CSA Trust symposium, organized this year by Bill Town, David Martinsen and Wendy Warr, focusing on scholarly publishing. It will run from Monday till Wednesday and will cover topics ranging from business models to emerging technologies and Web 2.0. Bill and Dave have put together an excellent speaker list and it's the place to be if you want to know where publishing is going. Wednesday sees Rachelle Bienstock and Alex Tropsha running a symposium on fragment based drug design, which received a great response the first time it ran last Spring. The meeting ends with four sessions of General Papers. I'm quite excited about these sessions as they cover a very diverse range of topics ranging from toolkits and databases to novel algorithms and infrastructures. If you can make it till Thursday, I highly recommend sitting in on some of these sessions.

Obviously, this doesn't cover all the symposia we have scheduled for San Francisco, but should whet your intellectual appetites. It goes without saying that this program could not have come together without the help and contributions from the members of the Program Committee and the symposium organizers themselves. Thanks to everybody and I look forward to meeting you in San Francisco.

Spring 2010 ACS National Meeting
San Francisco, CA (March 21-25)

CINF DIVISION MEETINGS AND SOCIAL EVENTS

The CINF Executive Meeting is a closed meeting; if you wish to attend it, contact the division chair. CINF members are **WELCOME AND ENCOURAGED** to attend any of the other committee meetings and all social functions.

Saturday, March 20, 2010

7:30 AM - 9:00 AM CINF: Long Range Planning & Breakfast Meeting/206-210 East, Moscone Center
9:00 AM - 10:00 AM CINF: Awards Committee Meeting/212 East, Moscone Center
10:00 AM - 11:00 AM CINF: Fundraising Committee Meeting/212 East, Moscone Center
11:00 AM - 12:00 PM CINF: Finance Committee Meeting/212 East, Moscone Center
9:00 AM - 10:30 AM CINF: Membership Committee Meeting/202-204 East, Moscone Center
10:30 AM - 12:00 PM CINF: Careers Committee Meeting/202-204 East, Moscone Center
9:00 AM - 12:00 PM CINF: Publications Committee Meeting/200 East, Moscone Center
9:00 AM - 12:00 PM CINF: Education Committee Meeting/220 East, Moscone Center
9:00 AM - 12:00 PM CINF: Program Committee Meeting/206-210 East, Moscone Center
12:00 PM - 1:00 PM CINF: Functionary Luncheon/206-210 East, Moscone Center
1:00 PM - 5:30 PM CINF: Executive Committee Meeting/206-210 East, Moscone Center

Sunday, March 21, 2010

12:00 PM - 2:00 PM CINF: CSA Trust Group Meeting
202-204 East, Moscone Center

6:30 PM - 8:30 PM CINF: Welcoming Reception & Scholarship for Scientific Excellence Posters
110-111 North, Moscone Center

Monday, March 22, 2010

5:30 PM - 7:30 PM CINF: Harry's Party/Presidential Suite, Palace Hotel/2 New Montgomery St.

Tuesday, March 23, 2010

12:00 PM - 1:30 PM CINF: Luncheon/Esplanade Brm 304, Moscone Center
6:30 PM - 8:30 PM CINF: Reception/Esplanade Brm 304, Moscone Center

Wednesday, March 24, 2010

12:00 PM - 5:00 PM CINF: CIC Collaborative Working Group Meeting/200 East, Moscone Center

ACS Chemical Information Division (CINF)
Spring 2010 ACS National Meeting
San Francisco, CA (March 21-25)
Technical Program Schedule

Rajarshi Guha
Chair, Technical Programming

OTHER SYMPOSIA OF INTEREST:
Medicinal Chemistry in Rare, Orphan and Neglected Diseases (see *MEDI*, Sun)
Advances in Conformational Sampling (see *COMP*, Sun - Mon)
New Drug Targets (see *BIOL*, Tue)

SUNDAY MORNING

Section A/The Moscone Center/Room 212 East

Sustainable Chemical Information Education through Faculty and Librarian Partnerships
J. Currano and J. Garritano, *Organizers*

8:00 Introductory Remarks.
8:05-1. "Chemical Information Instructor" feature: Telling the story of successful faculty-librarian partnerships in the Journal of Chemical Education. **A. Twiss-Brooks**
8:35-2. Teaching students to use REAXYS in order to find synthetic methods for making benzylic acid from benzaldehyde. **S. K. Cardinal**, W. D. Jones
9:05-3. In depth chemical information instruction in a second year organic chemistry course at Indiana University-Purdue University Indianapolis. **E. Snajdr**
9:35-4. Qualitative analysis in the library. **J. N. Currano**
10:05 Intermission.
10:20-5. Integrating library instruction into SAGES seminars at Case Western Reserve University for undergraduate chemistry students: A strategic plan. **S. Guo**
10:50-6. USC Chemistry OIL program. **N. Xiao**
11:20-7. Beyond Google: Integrating chemical information into the undergraduate chemistry curriculum. **M. C. Peters**
11:50-8. No chemistry librarian: How chemistry faculty can participate in collection development and its increased use. **A. W. Kozlowski**

Section B/The Moscone Center/Room 206/210

Metabolomics: A Field at the Boundaries between Chemistry and BiologyApplications and Infrastructures
Cosponsored by ANYL and MEDI
C. Steinbeck, *Organizer, Presiding*

9:00 Introductory Remarks.
9:05-9. Beyond Metlin: Extending metabolite databases to enhance identifications in metabolomics. **G. Suizdak**
9:45-10. Stable isotope-resolved metabolomics analysis of UDP-GlcNAc UDP-GalNAc: Computational resolution and modeling of their converging biosynthetic pathways. **H. N. B. Moseley**, A. C. Belshoff, R. M. Higashi, T. W.-M. Fan, A. N. Lane
10:10-11. Functional and structural network modeling of metabolomics datasets. **D. K. Barupal**, G. Wohlgemuth, O. Fiehn
10:35 Intermission.
10:50-12. Computational aspects of metabolomic Quantitative Trait Locus (mQTL) mapping and metabolome-wide genome-wide association studies. **M.-E. Dumas**, J.-B. Cazier, J. K. Nicholson, D. Gauguier
11:15-13. Metabolite identification pipeline based on MS fragmentation. **M. Rojas-Chertó**, **J. E. Peironcely**, P. T. Kasper, A. Bender, J.-L. Faulon, T. Reijmers, L. Coulier, R. Vreeken, T. Hankemeier
11:40-14. Consensus-phenotype integration of metabolomic and transcriptomic data enhances the detection of pathways associated with drug response. R. Cavill, A. Kamburov, M. S. C. Blagrove, T. J. Athersuch, J. K. Ellis, R. Herwig, **T. M. D. Ebbels**, H. C. Keun
12:05-15. Designing a cyberinfrastructure for metabolomics. **M. Arita**

Section C/The Moscone Center/Room 200 East

Cheminformatics Tools and High-throughput Approaches for the Discovery of New Materials

Microporous and Composite Materials
Cosponsored by COMP, INOR and PHYS
M. Haranczyk, *Organizer*
B. Smit, *Organizer, Presiding*
9:00 Introductory Remarks.
9:05-16. Computational discovery of new zeolite-like materials. **M. W. Deem**, D. J. Earl
9:35-17. Using multiple levels of computer modeling to

accelerate development of metal organic frameworks materials for membrane applications. **D. S. Sholl**, S. Keskin, T. Watanabe, E. Haldoupis, S. Nair
10:05-18. Informatics approaches for analyzing zeolites. **E. Blaisten-Barojas**
10:35 Intermission.
10:45-19. Screening tools for identification of porous materials for CO_2 separation. **M. Haranczyk**, J. Kloke, K. Theisen, B. Liu, G. Carlsson, B. Smit
11:15-20. 3D-Characterization and modeling of composite materials. **O. Wirjadi**

SUNDAY AFTERNOON

Section A/The Moscone Center/Room 212 East

Green Chemistry: Multidisciplinary Use of Chemical Information Resources

R. Schenck, *Organizer, Presiding*

2:00 Introductory Remarks.
2:10-21. Growing focus on green chemistry: What the CAS databases reveal. **R. Schenck**
2:40-22. California's green chemistry initiative: Application of life cycle assessment in public policy. **B. Boughton**
3:10 -23. ChemSpider: How an online resource of chemical compounds, reaction syntheses, and property data can support green chemistry. **A. J. Williams**, V. Tkachenko
3:40 Intermission.
3:55-24. From biochemical pathways to retrobiosynthesis. **C. H. Schwab**, O. Sacher, B. Bienfait, J. Gasteiger
4:25-25. Green chemistry database and analytical compliance suite (GC-DACS) . **G. R. Thompson**
4:55-26. Importance of benchmarking Green Chemistry. G. Gurau, **R. D. Rogers**
5:25 Concluding Remarks.

Section B/The Moscone Center/Room 206/210

Metabolomics: A Field at the Boundaries between Chemistry and Biology

Analytical Aspects

Cosponsored by ANYL and MEDI
C. Steinbeck, *Organizer, Presider*
2:00-27. Software advancements for performing compound identification QC on large metabolomic datasets. **C. DeHaven**
2:25-28. Fold-change analysis and visualization of multispectral datasets in NMR-based metabolomics. **S. L. Robinette**, A. S. Edison
2:50-29. Data mining tool for automated metabolite identification and quantification using *J*-resolved NMR spectroscopy. S. He, C. Ludwig, J. M. Easton, H. Chen, S. Tiziani, A. Lodi, S. Manzoor, A. D. Southam, T. N. Arvanitis, U. L. Guenther, **M. R. Viant**
3:15 Intermission.
3:30 30. Identification of parent-fragment pairs via rigorous statistical modeling of LC-MS metabolomic data. **A. Ipsen**, E. Want, J. Lindon, T. Ebbels
3:55 31. Generation of in-silico MS/MS mass spectra using combinatorial algorithms and reaction prediction expert systems. **T. Kind**, K.-H. Liu, D. Y. Lee, O. Fiehn

SUNDAY EVENING

Section A/The Moscone Center/Room 110 – 111 North

2010 CINF Scholarship for Scientific Excellence

G. Grethe, *Organizer*

6:30 - 8:30
32. Using aggregative Web Services for drug discovery. **Q. Zhu**, M. S. Lajiness, D. J. Wild
33. Semantifying polymer science using ontologies. **E. O. Cannon**, N. Adams, P. Murray-Rust
34. PDZ domain: A computational insight into protein similarity and ligand specificity. **C. S. D. Sisu**, G. H. Grant
35. Homology modeling and molecular dynamics of the APJ receptor, a co-receptor for cellular entry of HIV. **N. J. M. Macaluso**, R. C. Glen
36. Assigning stereochemistry using GIAO NMR shift calculation. **S. G. Smith**, J. M. Goodman
37. Combination of target-specific pose filters and force field-based scoring functions to improve the structure-based virtual screening. **J.-H. Hsieh**, S. Yin, X. S. Wang, S. Liu, N. V. Dokholyan, A. Tropsha
38. Toxicity reference database (ToxRefDB) to develop predictive toxicity models and prioritize compounds for future toxicity testing. **H. Tang**, H. Zhu, L. Zhang, A. Sedykh, A. Richard, I. Rusyn, A. Tropsha
39. Chem2bio2RDF: Semantic systems chemical biology. **B. Chen**, X. Dong, D. Jiao, H. Wang, Y. Ding, D. J. Wild
40. Application of novel data mining techniques to improve chemical array design decisions during lead optimisation. **G. Papadatos**, M. Alkarouri, V. Gillet, P. Willett, V. Kadirkamanathan, C. Luscombe, G. Bravi, N. Richmond, S. Pickett, J. Pritchard, A. Cooper, S. Macdonald
41. Biological and chemical network construction from linked open data. **D. Jiao**
42. Novel approach to drug discovery integrating chemogenomics and QSAR modeling: Applications to anti-Alzheimer's agents . **R. Hajjo**, S. Wang, B. L.

Roth, A. Tropsha
43. Use of QSAR models to validate the HTS assay data in PubChem: Application to the serotonin receptor 5-HT1A ligands. **M. Luo**, S. X. Wang, A. Golbraikh, A. Tropsha
44. Speeding up chemical database searches. **R. J. Nasr**, P. Baldi
45. OrbDB: A database of molecular orbital interactions. **M. A. Kayala**, C. A. Azencott, J. H. Chen, P. F. Baldi
46. PropOrb: A frontier molecular orbital interaction proposer. **C.-A. Azencott**, M. A. Kayala, P. Baldi

MONDAY MORNING

Section A/The Moscone Center/Room 212 East

Visual Analysis of Chemical Data

Data Integration
Cosponsored by COMP
J.-C. Bradley, A. Lang, and N. O'Boyle, *Organizers*

8:45 Introductory Remarks.
8:50-47. Dataviz explosion: Considering effectiveness in classrooms and across disciplines. **E. Dorland**
9:30-48. ASAP: Emphasizing multidimensional drug discovery . **W. P. Walters**, Y. Bennani, T. Kramer
9:55-49. Visual analyses for guiding compound selection and design. **E. Champness**
10:20-50. Contextual visualization of biological and chemical networks using linked open life science data. **D. Jiao**, Y. Ding, D. J. Wild
10:45 Intermission.
11:00-51. iTunes for chemistry and biology: Addressing usability in cheminformatics. D.-T. Nguyen, Y. Wang, R. Guha, **N. Southall**, R. Huang, **A. Jadhav**
11:25-52. Interaction Maps: Understanding compound pathway space. **Y. Tanrikulu**, M. Weisel, R. Kondru, H. M. Bitter
11:50-53. PubChem: A public cheminformatics system for bioactivity analysis. **Y. Wang**

Section B/The Moscone Center/Room 206/210

Cheminformatics Tools and High-throughput Approaches for the Discovery of New Materials

Electronic and Optical Materials
Cosponsored by COMP, INOR and PHYS
B. Smit, *Organizer*
M. Haranczyk, *Organizer, Presiding*

9:00 Introductory Remarks.
9:05-54. High throughput synthesis and screening system for discovery of improved materials for rechargeable batteries. B. Li, F. Matsumoto, B. Howard, R. Olugbile, D. Greenburg, C. Riley, C. O'Neill, M. S. Bailey, **S. S. Kaye**
9:35-55. The Clean Energy Project: Finding new renewable energy materials, one screensaver at a time. **A. Aspuru-Guzik**
10:05-56. Compound discovery and design via materials informatics: An "omics" approach to materials science. **K. Rajan**
10:35 Intermission.
10:45-57. Organic electronic materials by design: Finding a needle through the haystack. **G. R. Hutchison**, N. O'Boyle
11:05-58. Searching chemical space by inverse design. **B. C. Rinderspacher**, J. Andzelm, A. Rawlett, J. Dougherty, R. Lambeth
11:25-59. High-throughout quantum chemistry and virtual screening for materials solutions. **M. D. Halls**

Beyond the Bench: Non-Traditional Careers in Chemistry

CHAL's 25th Anniversary Series
Sponsored by CHAL, Cosponsored by CINF

MONDAY AFTERNOON

Section A/The Moscone Center/Room 212 East

What Happened to My Library?: Managing Organizational and Space-related Challenges
A. Twiss-Brooks, *Organizer*
E. Kajosalo, *Organizer, Presiding*

1:45 Introductory Remarks.
1:50-60. Impacts of closing academic branch libraries. **S. J. Redalje**
2:15-61. e-Books and e-references collection in academic libraries: A case study. **N. Xiao**
2:40-62. Future of the liaison in academic libraries. **M. Lafferty**
3:05-63. New workspaces: From library facility to online service. **L. Solla**
3:30 Panel Discussion.

Section B/The Moscone Center/Room 206/210

The Future of Scholarly Communication

Evolving Business Models

Cosponsored by CHED
D. Martinsen, W. G. Town, and W. A.Warr, *Organizers*
1:00 Introductory Remarks.
1:05-64. PLoS ONE: A new model for online journal publishing. **P. Binfield**
1:35-65. Impact of emerging economies on publishing. **J. Bhate**
2:05-66. Ensuring sustainability of scientific data resources: Technologies and business models for the future. **C. R. Groom**
2:35 Intermission.
2:45-67. Vision and strategy for the future of RSC Publishing. **R. J. Parker**, J. Milne
3:15-68. Challenges facing the STM industry. **M. O'Malley**
3:45-69. Vision and strategy for scholarly communication at the ACS. **S. King**
4:15 Intermission.
4:20 Open Meeting. CINF Division.
4:30 Intermission.
4:35 Open Meeting. Committees on Publications and Chemical Abstracts Service.

MONDAY EVENING
Section A/The Moscone Center/Hall D
Sci-Mix
R. Guha, *Organizer*
8:00 - 10:00
7. See previous listings
72, 102, 122. See subsequent listings.

TUESDAY MORNING

Section A/The Moscone Center/Room 212 East

Visual Analysis of Chemical Data

QSAR

Cosponsored by COMP
J.-C. Bradley, A. Lang, and N. O'Boyle, *Organizers*

8:15-70. Network visualization of structure activity landscapes. **R. Guha**
8:40-71. Visualization and analysis of bioisosteric analogs generated with BROOD. **J. J. Corkery**, A. G. Skillman, K. E. Schmidt, B. Kelley
9:05-72. Visualizing polypharmacology in chemical libraries. **D. Marcus**, D. Barasch, A. Rayan, A. Goldblum
9:30-73. Linear scaling mapping of chemical space. **J. S. Delaney**
9:55 Intermission.
10:10-74. Integrated visualization for the interpretation of QSPR and QSAR models. **G. D. Purvis**, W. D. Laidig, D. T. Stanton
10:35-75. VTK: The visualization toolkit and its applications to the analysis of chemical data. **M. D. Hanwell**
11:00-76. Molecular similarity characterization of ADME landscapes. **R. R. Gupta**, **B. Chen**, E. M. Gifford
11:25-77. Chemical information usage patterns under the microscope: The prospects of GIS-based visualization for libraries. **M. P. Brändle**

Section B/The Moscone Center/Room 206/210

The Future of Scholarly Communication

Towards Web 2.0

Cosponsored by CHED
D. Martinsen, W. G. Town, and W A.Warr, *Organizers*

8:15 Introductory Remarks.
8:20-78. Scholarly communication in the 21st Century: Changing paradigms of communication and access. **R. Schwarzwalder**
8:50-79. oreChem project: Semantic infrastructure and applications for chemistry scholarship. **C. Lagoze**
9:20-80. Chemistry: Computation, synthesis and communication. **J. M. Goodman**
9:50-81. Primary data for chemistry. I. Sens, S. Haak, J. Brase, **G. F. Herrmann**
10:20 Intermission.
10:35-82. Papers to papers: The future of reading, storing, and producing chemical literature in an academic environment. **J. N. Currano**
11:05-83. Socialized medicine: Medical publishing in the era of the read-write web. **K. R. Anderson**
11:35-84. Scholarly communication between chemists and their uptake and use of Web 2.0 and other new technologies. **W. G. Town**

TUESDAY AFTERNOON

Section A/The Moscone Center/Room 212 East

Visual Analysis of Chemical Data

Molecular Visualization

Cosponsored by COMP
J.-C. Bradley, A. Lang, and N. O'Boyle, *Organizers*

2:00-85. ChemSci3D: An integrated 2D and 3D

desktop chemistry interface. **T. J. O'Donnell**
2:25-86. Real-time ray-tracing in molecular graphics. **M. Keil**
2:50-87. Avogadro: Framework for chemical structure analysis and visualization. **M. D. Hanwell**
3:15-88. 2D graphical depiction of zeolite voidspace topology. **K. J. Theisen**, B. Smit, M. Haranczyk
3:40 Intermission.
3:55-89. Computational methods for 2D-visualization of molecular interaction patterns. K. Stierand, **M. Rarey**, C. Lemmen
4:20-90. Visualizing combinatorial molecular materials research using Avogadro. **G. R. Hutchison**, M. D. Hanwell
4:45-91. Visualizing chemistry in Second Life. **A. S. I. D. Lang**, **J.-C. Bradley**
5:10-92. Next-generation visualization technologies: How close are we to the Science 2.0?. **O. Isayev**

Section B/The Moscone Center/Room 206/210

The Future of Scholarly Communication

Application of Emerging Technologies
Cosponsored by CHED
D. Martinsen, W. G. Town and W. A. Warr, *Organizers*

2:00 Introductory Remarks.
2:05-93. CAS and ACS publications: Together, providing new pathways to chemical research. **J. Morgan**, **M. Dennis**
2:35-94. Adding structure to publishing chemistry. **J. N. Wilde**
3:05-95. Under the hood: A technical view on publishing chemistry in the future. **R. Kidd**
3:35 Intermission.
3:50-96. Interactive publishing of crystal structure data. **B. McMahon**, P. R. Strickland
4:20-97. Science video journal to increase efficiency and standardization in experimental research. **M. Pritsker**
4:50-98. Nano-Publication. **J. J. J. M. Velterop**

TUESDAY EVENING

Section D/The Moscone Center/Hall D

General Posters
R. Guha, *Organizer*

6:00 - 8:00
99. Speeding up chemical database searches. **R. J. Nasr**, P. Baldi
100. Comparison of backfiles offered by major publishers of chemical journals. **A. D. Bolek**
101. Marketing chemical research with custom web databases. **B. Brandys**
102. Advanced materials/processes for information technology: Microfluidic synthesis, nanofabrication, and soft lithography. **K. Choi**
103. New Knovel interface. **S. Gurke**
104. Expanding and understanding metabolite space. **J. E. Peironcely**, A. Bender, M. Rojas-Chertó, T. Reijmers, L. Coulier, T. Hankemeier
105. Metabolite identification based on MS fragmentation. **M. Rojas-Chertó**, P. T. Kasper, J. E. Peironcely, T. Reijmers, R. J. Vreeken, T. Hankemeier

WEDNESDAY MORNING

Section A/The Moscone Center/Room 212 East

Fragment Based Drug Design: Success Stories due to Novel Computational Methods Applications
Cosponsored by COMP and MEDI
R. Bienstock, *Organizer, Presiding*

8:30 Introductory Remarks.
8:40-106. Fragment database analysis using molecular shape fingerprints. **J. D. MacCuish**, N. E. MacCuish, M. Hawrylycz, M. Chapman
9:10-107. Qsearch: Pharmacophore-based search in fragment spaces. **T. Lippert**, M. Rarey
9:40-108. Fragment-based lead discovery through chemotype evolution. **D. Erlanson**
10:10 Intermission.
10:25-109. Leading fragments to lead structures: Fragment evolution, merging and core replacement, and…docking. **C. Detering**, M. Gastreich, C. Lemmen
10:55-110. Fragment-based drug design using PASS approach. O. A. Filz, A. A. Lagunin, D. A. Filimonov, **V. V. Poroikov**
11:25-111. Use of virtual fragment screening for lead modification. **Y. Xu**, H. Jansen, E. Martin

Section B/The Moscone Center/Room 206/210

The Future of Scholarly Communication

Authoring and Discovery Tools
Cosponsored by CHED
D. Martinsen, W. G. Town, and W. A.Warr, *Organizers*

8:15 Introductory Remarks.
8:20 112. IUPAC InChI project: A status report. **S. R. Heller**, A. McNaught
8:50-113. Tools for the scholarly communication lifecycle. **A. D. Wade**, L. Dirks
9:20-114. Integrated language for chemical publication.

P. Murray-Rust, J. A. Townsend, L. Hawizy
9:50-115. Chem4Word. **J. A. Townsend**, P. Muray-Rust, J. Downing, T. Haughton, A. Wade, L. Dirks
10:20 Intermission.
10:35-116. Capturing and publishing chemistry from the bench to paper: Does Google Wave offer a solution?. **C. Neylon**
11:05-117. CAS REGISTRY: Maintaining quality standards as scientific discovery accelerates. **M. J. Toussant**
11:35-118. Chempedia: A social medium for chemical information. **R. L. Apodaca**
12:05-119. Enhancing discoverability across Royal Society of Chemistry content by integrating to ChemSpider, an online database of chemical structures. **A. J. Williams**, V. Tkachenko, S. Shevelev, R. Kidd

WEDNESDAY AFTERNOON

Section A/The Moscone Center/Room 212 East

Fragment Based Drug Design: Success Stories due to Novel Computational Methods Applications

Cosponsored by COMP and MEDI
R. Bienstock, *Organizer, Presiding*

2:00-120. Fragment-based screening of stabilized G protein-coupled receptors. **M. Congreve**
2:30-121. Novel histamine GPCR family antagonists by fragment screening and molecular modeling. **R. J. Law**, T. Hesterkamp, A. Kahrs, M. Whittaker, S. Pal, A. Heifetz
3:00 Intermission.
3:15-122. Mitotic kinesin Eg5 inhibitors generation by computational MED-portion based drug design at PDB scale. K. Oguievetskaia, L. Martin-Chanas, A. Vorotyntsev, O. Doppelt-Azeroual, X. Brotel, S. Adcock, A. de Brevern, **F. Delfaud**, **F. Moriaud**
3:45 123. Starting small and staying small: Fragment-based lead discovery in CNS disease. **V. L. Nienaber**
4:15 124. Finding druggable sites in protein-protein interfaces by computational fragment mapping. **D. Kozakov**, S. Vajda
4:45 Concluding Remarks.

Section B/The Moscone Center/Room 206/210

The Future of Scholarly Communication

Peer Review and Impact Metrics

Cosponsored by CHED
D. Martinsen, W. G. Town, and W. A. Warr, *Organizers, Presiding*

2:00-125. Peer review is not perfect but are the alternatives worse?. **W. A. Warr**
2:30-126. Interactive open access publishing and public peer review: Perspectives and effectiveness of transparency and self-regulation in scientific communication and quality assurance. **U. Pöschl**
3:00-127. Impact factors, post-publication peer review and other metrics. **R. P. Grant**
3:30 Intermission.
3:45-128. Eigenfactor: Ranking and mapping the scholarly literature. **J. D. West**
4:15-129. Applying domain expertise assessment to compound annotation for enhanced collaborations between chemists and biologists. **Y. Pouliot**
4:45 Concluding Remarks.

Competition Law and E-publishing in Scientific Fields

Sponsored by CHAL, Cosponsored by CINF

THURSDAY MORNING

Section A/The Moscone Center/Room 212 East

General Papers

Chemical Data Mining

R. Guha, *Organizer*

8:30-130. Exploring molecular networks for polypharmacology analysis. **S. Zhang**
8:55-131. Picking novel actives from large databases using a combined ranking method. **D. Marcus**, D. Barasch, A. Rayan, A. Goldblum
9:20-132. Human proteome in the context of pathways, diseases, drugs, and tool compounds. **F. Nigsch**, J. Jenkins
9:45-133. Atom type preferences, constitutional diversity, and property profiles of known drugs: A knowledge-intensive, comparative assessment of drugability. **V. N. Viswanadhan**, H. Rajesh, V. N. Balaji
10:10 Intermission.
10:25-134. Improving the consistency of data fusion in virtual screening. I. Mott, P. Gedeck, **V. J. Gillet**
10:50-135. Extending interaction fingerprints: A novel approach to characterizing protein binding sites. **C. Higgs**, W. Sherman, J. Blaney
11:15-136. Molecular signatures of promiscuous enzyme-substrate interactions. **J.-L. Faulon**, P. Carbonell
11:40-137. Metabolic regioselectivity models for nine CYP P450 isozymes. **J. M. Zaretzki**, T.-W. Huang, C. M. Breneman, C. Bergeron, K. P. Bennett

Section B/The Moscone Center/Room 206/210

General Papers

Algorithms and Toolkits

R. Guha, *Organizer*

8:15-138. Novel topological molecular key for cheminformatics. **P. Liu**, D. Agrafiotis
8:40-139. Classification of enzyme function based on similarities in reaction mechanisms and common substrate substructures. **D. E. Almonacid**, P. C. Babbitt
9:05-140. Chem_BLAST: A rule-based method to develop advanced structural ontologies for chemical bioinformatics and the PDB, the PubChem. **T. N. Bhat**
9:30-141. Chemical entity extraction and interpretation. **D. M. Lowe**, P. T. Corbett, P. Murray-Rust, R. C. Glen
9:55 Intermission.
10:10-142. PolyGWT: A Google Web Toolkit application to assist polymer research. **E. O. Cannon**, P. Murray-Rust
10:35-143. OpenTox: An open source predictive toxicology software framework. **B. Hardy**, **D. A. Gallagher**, S. Chawla
11:00-144. Combining disparate cheminformatics resources into a single toolkit: The Cinfony library. **N. M. O'Boyle**, G. R. Hutchison

THURSDAY AFTERNOON

Section A/The Moscone Center/Room 212 EastGeneral Papers

Platforms and Databases

R. Guha, *Organizer*

1:30-145. Role of Rh(III) as negative catalyst in N-chlorosuccinimide oxidation of glycine in alkaline medium: A kinetic and mechanistic study. **A. K. Singh**, M. Singh, R. Srivastava, J. Srivastava, S. Rahmani
1:55-146. Psychological barriers to good decision-making: How addressing cognitive biases could improve drug discovery. **M. D. Segall**, A. Chadwick
2:20-147. NCI/CADD: Open-access chemical structure web platform. **M. Sitzmann**, W.-D. Ihlenfeldt, M. C. Nicklaus
2:45-148. OrbDB: A database of molecular orbital interactions. **M. A. Kayala**, C. A. Azencott, J. H. Chen, P. F. Baldi
3:10 Intermission.
3:25-149. Application of crowdsourcing for metadata curation of digitized texts. **A. J. Williams**, R. Kidd, V. Tkachenko, S. Shevelev
3:50-150. Five years of collaborative drug discovery in the cloud. **B. A. Bunin**, S. Ekins, M. Hohman, S. Ernst, K. Gregory
4:15-151. CWM Global Search: An Internet search engine for the chemist. **A. J. Kos**, H.-J. Himmler
4:40-152. www.emolecules.com: The comprehensive source of commercially available, in stock chemicals. **K. Gubernator**

Section B/The Moscone Center/Room 206/210

General Papers

Small Systems and Large Calculations

R. Guha, *Organizer*

2:00-153. Petascale lattice-Boltzmann simulations of dynamical processes in ternary amphiphilic liquid crystalline systems. R. S. Saksena, **P. V. Coveney**
2:25-154. Novel, accurate high-throughput 3D pharmacophore screening algorithm. **G. Wolber**, T. Seidel, F. Bendix, P. Markt
2:50-155. Structure – toxicity relationship study for organophosphorus compounds: QSAR analysis. **Y. Paukku**, D. Magers, G. Hill
3:15 Intermission.
3:30-156. EEM-Hückel model for partial atomic charges. **M. Waldman**, R. Fraczkiewicz, W. S. Woltosz
3:55-157. PropOrb: A frontier molecular orbital interaction proposer. **C.-A. Azencott**, M. A. Kayala, P. Baldi
4:20-158. Fast and accurate prediction of the 3D structure of small molecules. **A. Andronico**, A. Randall, P. Baldi

AWARDS AND SCHOLARSHIPS

Chemical Structure Association Trust

Applications Invited for CSA Trust Jacques-Émile Dubois Grants for 2011

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, renamed in honor of Professor Jacques-Émile Dubois who made significant contributions to the field of cheminformatics. The Trust is currently inviting the submission of grant applications for 2011. (N*ote: the cutoff date for the 2010 Grants is March 15, 2010).*

***Purpose of the Grants*:**

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

***Who is Eligible*?**

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

What Activities are Eligible?

Grants may be awarded to acquire the 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 document must be supplied for distribution to the Grants Committee.

***Deadline for Applications**:*

Applications must be received no later than March 14, 2011. Successful applicants will be notified no later than May 2, 2011.

***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 blawlor@nfais.org prior to submission so that she can contact you if the e-mail does not arrive.

2010 CINF Scholarship for Scientific Excellence Sponsored by FIZ CHEMIE Berlin

The scholarship program of the Division of Chemical Information (CINF) of the American Chemical Society (ACS), funded by FIZ CHEMIE Berlin, is designed to reward graduate and postdoctoral students in chemical information and related sciences for scientific excellence and to foster their involvement in CINF. Up to three scholarships valued at $1,000 each will be presented at the 240th ACS National Meeting in Boston, MA, August 22 – 26, 2010. Applicants must be enrolled at a certified college or university, and they will present a poster during the Welcoming Reception of the division on Sunday evening at the National Meeting. Additionally, they will have the option to also show their poster at the Sci-Mix session on Monday night. Abstracts for the poster must be submitted electronically through PACS, the new abstract submission system of ACS.

To apply, please inform the Chair of the selection committee, Guenter Grethe at ggrethe@comcast.net, that you are applying for a scholarship. Submit your abstract to http://abstracts.acs.org using your ACS ID. If you do not have an ACS ID, follow the registration instructions and submit your abstract for "CINF Scholarship for Scientific Excellence." The deadline for submitting an abstract to PACS is March 28, 2010. Additionally, please send a 2,000-word abstract describing the work to be presented in electronic form to the Chair of the selection committee by June 15, 2010. Any questions related to applying for one of the scholarships should be directed to the same e-mail address.

Winners will be chosen based on contents, presentation and relevance of the poster and they will be announced during the reception. The contents shall reflect upon the student's work and describe research in the field of cheminformatics and related sciences. Winning posters will be marked "Winner of FIZ CHEMIE-CINF Scholarship for Scientific Excellence" at the poster session.

Guenter Grethe

SCHEDULE OF FUTURE ACS NATIONAL MEETINGS

239th	Spring 2010	March 21-25	San Francisco, California
240th	Fall 2010	August 22-26	Boston, Massachusetts
241st	Spring 2011	March 27-31	Anaheim, California
242nd	Fall 2011	August 28 - September 1	Denver, Colorado
243rd	Spring 2012	March 25-29	San Diego, California
244th	Fall 2012	September 9-13	New York, New York

From the Institute for Scientific Information (ISI) to the National Federation of Advanced Information Services (NFAIS): Interview with Bonnie Lawlor

Svetla Baykoucheva

Bonnie Lawlor

Bonnie Lawlor is Executive Director of NFAIS, a membership association for organizations that aggregate, organize and facilitate access to authoritative information. Prior to NFAIS, Bonnie was Senior Vice President and General Manager of UMI's Library Division (now ProQuest Information and Learning) where she was responsible for the development and worldwide sales and marketing of their products to academic, public, and government libraries.

Before UMI, Bonnie was Executive Vice President of the Database Publishing Division at the Institute for Scientific Information (ISI – now Thomson Reuters, Healthcare & Science) where she was responsible for product development, production, publisher relations, editorial content, and worldwide sales and marketing of all of ISI's products and services.

Bonnie is a very active member of the American Chemical Society. She is currently a Councilor for the Division of Chemical Information and an elected member of the Council Policy Committee. She is a past chair of the American Chemical Society's Committee on Copyrights and the ACS Committee on Divisional Activities, and has also served on the ACS Committee on Budget and Finance and the Committee on Nominations and Elections. She is currently a Trustee and Secretary of the Chemical Structure Association (CSA) Trust, an internationally recognized organization established to promote the critical importance of chemical information to advances in chemical research. She also serves on the Board of LYRASIS (formerly PALINET) and on the Board of the Philosopher's Information Center.

Bonnie has also served as a Board and Executive Committee Member of the Information Industry Association (IIA) and a Board Member of the American Society for Information Science (ASIS). Ms. Lawlor earned a BS in Chemistry from Chestnut Hill College (Philadelphia), an MS in chemistry from St. Joseph's University (Philadelphia), and an MBA from the Wharton School (University of Pennsylvania). In 1992 Dr. Garfield wrote about Bonnie's accomplishments at ISI that was published in the *Current Contents* and can be viewed at http://www.garfield.library.upenn.edu/essays/v15p280y1992-93.pdf

Svetla Baykoucheva: You have held a number of executive positions in different companies and non-profit organizations and you have served as an elected official in the American Chemical Society (ACS). It seems that all the organizations that you have been affiliated with professionally have something in common—they are all related to scientific information and scientific publishing. How did you come to this field, what triggered your interest in it, and what were the main factors that have influenced your career (e.g., education, chance, timing, etc.)?

Bonnie Lawlor: Svetla, I fell into the field of scientific publishing quite unintentionally. Immediately after college I went to the University of Pennsylvania to study for my Ph.D. Upon completion of my coursework I left to find a job as I had become engaged to a Vietnam War veteran who wanted to complete his college degree. With only a Bachelor's degree in chemistry the opportunities were less than exciting, plus I was uncertain as to whether or not a laboratory career was really for me. I saw an advertisement for a chemical indexer in the now defunct *Philadelphia Bulletin*. I had no idea what being a "chemical indexer" actually entailed, but I interviewed, was tested, and was offered the position at the Institute for Scientific Information (ISI). After two years I was hooked. ISI was, at that time, small, entrepreneurial and very interesting. Plus I was able to use my education and love of the theory of chemistry without having to spill chemicals (which I had been known to do!). Ultimately I became involved with other areas of the company - *Current Contents*, the citation indexes, etc. – and was caught up in the industry transition from print to electronic publications. An exciting era only made more so by the introduction and evolution of the Web!

SB: You are currently Executive Director of NFAIS. What does this acronym stand for and what does this organization do?

BL: NFAIS is short for the National Federation of Advanced Information Services. It is a non-profit organization that was founded in 1958. At that time President Eisenhower directed the National Science Foundation to ensure the provision of indexing, abstracting, translation, and other information services that would lead to a more effective dissemination of scientific information. He believed that science had won WWII and that science would keep the peace. As the U.S. mobilized to create a new information infrastructure for the promotion of scientific innovation, G. Miles Conrad, Director of Biological Abstracts (later BIOSIS and now part of Thomson Reuters), called an meeting of leading not-for-profit and government scientific Abstracting & Indexing services. He encouraged the group to join forces, cooperate, and interact so that as a unified force they could make rapid progress in achieving national priorities while simultaneously promoting the international advancement of science. As a result of his efforts a new organization - the National Federation of Science Abstracting and Indexing Services (*NFSAIS*) - was formed with the charter membership of fourteen information services, including *Chemical Abstracts*, *Engineering Index*, *AGRICOLA*, *Current List of Medical Literature* (NLM), etc. The organization has since expanded beyond science to include all scholarly disciplines. Membership is now available to for-profit organizations and is no longer limited to A&I services. NFAIS currently serves all those who create, aggregate, organize, and otherwise provide ease of access to and effective navigation and use of *authoritative* information and our Member organizations represent a global cross-section of content and technology providers, including database creators, publishers, libraries, host systems, information technology developers, content management providers, and other related groups. Despite diverse interests, all NFAIS members embrace the philosophy underlying the organization's original motto, *Promotion through Cooperation*, and work together to facilitate collaboration and communication throughout the Information Community. The work of NFAIS is to:

- Facilitate the exchange of information among NFAIS members
- Promote *NFAIS* members and their essential role within the Information Community
- Encourage discussion, understanding and cooperation across all Information Community sectors
- Sponsor topical conferences, seminars and educational courses
- Publish newsletters, current awareness alerts, books and reports
- Develop Codes of Practice, Guiding Principles and White Papers on Information Policy and New Technologies

SB: Being Executive Vice President of the Database Publishing Division of the Institute for Scientific Information (ISI is now Thomson Reuters, Healthcare & Science) and being responsible for so many areas (product development, production, publisher relations, editorial content, and worldwide sales and marketing of all of ISI's products and services) could be a daunting responsibility. What imprint, do you think, your work has made on ISI's success and image?

BL: Over the twenty-eight year span that I spent at ISI, I would perhaps choose a few "turning points" where I know that I had an impact on the outcome and the ultimate shaping of the company. The first is regarding ISI's chemical information products. *Index Chemicus*, a weekly alert to new chemical compounds, was launched by Dr. Garfield in the early 1960's before I joined the company. It was not a popular move and three vice presidents even left the company, partially due to this initiative that they perceived as being risky. In 1982 the entire chemistry product line was made a separate division under my leadership, with the directive to make it work. We were responsible for product development, production, sales and marketing. We had a great team and many in CINF may remember them – Judy Sarkisian, Jack Coulson, Kerry Louiso – and the indexing and encoding staff, some of who are still with the company – Pat Rosso, Maria Gonzalez, Josie Ortega, Shelly Rahman, Dave Jordan, etc. We believed in the importance of reaction indexing and wanted to create a database of new chemical reactions in organic chemistry, but had no funding. We were given approval to see if we could obtain seed money from interested chemical and pharmaceutical companies. So we launched a Charter Club in which those organizations who provided funding would have a say in the development of the reaction product. We were able to obtain the funding, develop the product and, through a partnership with Molecular Design Ltd. (MDL) offer a graphic interface to the reactions. It was one of the first of its kind and was quite successful. As a result the chemistry product line became financially viable and grew, and it remains a source of viable product offerings from Thomson Reuters. I am extremely proud of everyone who had a part in making that happen as the odds were not in our favor.

By the mid to late 1980's the entire abstracting and indexing community faced another challenge – how to adapt its print products and services to the newly emerging digital environment sparked in 1981 by the launch of personal computers and fueled by the emergence of the CD-ROM and diskette distribution media. We were very fortunate. We had been creating electronic versions of all of our citation indexes, *Current Contents* and the chemical products as a by-product of computerized production that most major A&I services had adopted in the 1960's. The issue was to take the data already available on magnetic tape and make it compatible with the new platforms. Change is not easy and it took some doing to convince staff (and in some cases management) that digital was the future. Again, my staff rose to the occasion – Theresa Rosen on the citation index side and Beverly Bartolomeo on the *Current Contents* side and together with the assistance of programmers, editorial staff, sales and marketing, we made it happen. Within two years 20% of our print base had converted to the new format.

This shift to digital products and services was coincidental with another major change that was specific to ISI, for we had caught the attention of JPT - a publishing company owned by Ted Cross, Joe Pallazolo and Paul Neuthaler. They were interested in acquiring the company – which they did in 1988. And over the next four years they helped us grow the business. During that period there were two decisions that I was able to shape that were to have a significant long-term impact. First, was the pricing of electronic products. JPT believed that they should be priced lower than print because there was no printing involved, shipping was cheaper, etc. I was just as convinced that they should be priced higher because of factors that were unique to digital products – ongoing investment in technology and software, training (digital information products were still relatively new), support via help desk activities and the fact that initial purchases would be by existing customers migrating from the print. It took a lot of meetings, presentations and analyses, but they ultimately agreed to launch *Current Contents on Diskette* at a price higher than the print. A good move if I say so myself! Remember, in the 1980's computer literacy was not the norm. Customer training and support was not limited to the product that was being sold, but spilled over to the technology as well.
In 1988, help desk phone activity grew 72% over prior year and 1989 grew 105% over 1988. By 1990 there was a 581% increase in phone activity. And the staff handling that activity had to be both computer and product literate. In addition, my guess was right – initial customers of electronic products were due to cannibalization of the print. And, as noted earlier, within just two years 20% of the print base was gone. (These stats appeared in a report that I wrote for NFAIS in 1991that was published in *Information Distribution Issues for the 90's*: copies available upon request).

The second decision that I was able to get approved was to add English language author abstracts to ISI products. Up until this time they were only included in the print issues of *Index Chemicus* and *Current Chemical Reactions* and I believed that they were an essential addition to our new electronic offerings. In addition, many of our competitors already had abstracts in their products. Again, many discussions and meetings – and outreach to publishers. In the end approval was won and the announcement was celebrated amid much fanfare at a customer party during the online Information meeting in New York in May, 1991. JPT funded a number of innovations that made ISI very attractive to much larger content providers. After four years - and many presentations to competing suitors - the company was sold to Thomson (now Thomson Reuters) in 1992.

Svetla, your question made me think of specific instances where a visible and long-lasting impact was made. In general, I would say that the combination of my fiscal responsibility and love of ISI together was a great foil to Dr. Garfield's creativity and drive. Throw in the unbelievable genius of people such as Irv Sher, George Vladutz and Henry Small, and the work ethic and loyalty of hundreds of employees who were devoted to the company – ISI became a major force in the Information community. I was just one of many and I am grateful that I had the opportunity to be part of the unique ISI family.

SB: What did it take to work and succeed in an environment (such as the one at ISI at that time) that was so innovative, dynamic and competitive—and dominated by a mythological figure such as Eugene Garfield? Could you tell us what your first encounter with Dr. Garfield was?

BL: As I mentioned earlier, when I joined ISI it was relatively small and very entrepreneurial. We all were made to feel that we were part of the creation of something of value. When a customer wrote to tell Dr. Garfield that a product or service solved a problem, he let us know (of course, we also heard all of the complaints). It was truly nourishing environment. In the early days I did not observe biases of any kind. No matter what your gender, color or educational status – if you had an idea, Dr. Garfield was willing to hear it. It was an environment that offered great opportunity if you were creative and willing to work hard. It was also a crazy place to work – perhaps due to the culture of the late 60's and early 70's. People parked their

motorcycles by their desks. The work dress ranged from normal to eccentric. I remember one person wore baby doll pajamas to the office and one executive always wore a small teddy bear on his belt (these same two people "streaked" at one of the company parties!). When my boss complained about the length (or lack thereof) of miniskirts, the corporate (unofficial) response was that the only dress code requirement was shoes! The examples are endless. But when I went to UMI in the 1990's I heard similar stories from their staff. I suspect the ISI environment was a combination of the times and the personality of our corporate leader.

I still smile about my first encounter with Dr. Garfield. Every day the coffee shop in the lobby of our building sent a cart to each floor in mid-morning and afternoon so that everyone could get a snack. While I waited in line by the elevators to get my caffeine fix in the early days of my employment, a rather strange vision emerged from the elevators wearing a gray jacket with a fur collar and wild hair reminiscent of Albert Einstein. I asked the person behind me who it was (I thought perhaps he was a handyman). When the laughter subsided I was told the vision in question was Dr. Garfield. Ultimately I came to know, respect, and occasionally fear him. I learned so very much from him – the importance of such things as quality, responsiveness to customers, innovation – and being a professional. Even though we competed with the American Chemical Society, he made sure that we were active in the ACS - particularly in what is now the Division of Chemical Information. He said that we were chemists and should actively promote the profession. He encouraged us to get involved and to have good working relationships with CAS staff. It is due to him that I and many others at ISI became active. In retrospect, I could not have had a better mentor. We still keep in touch and I treasure our relationship.

SB: The Science Citation Index has provided a new approach to information retrieval. Web of Science, which is based on the Science Citation Index, does not use topical indexing—it heavily relies on words used in titles of documents. What will happen if a particular term has been misspelled in the title of an article? Is Web of Science going to miss this article?

BL: I cannot address ISI's current processing system, but I can briefly talk about the "unique word dictionary (UWD)" and the process that was in place for providing accurate index terms for the citation indexes while I was there. Rather than use a controlled thesaurus for creating index terms, the decision very early in ISI's history was to use the natural language of science that would evolve over time. Simply put, we used the title words from each article processed. To minimize errors, every title was separately keyed by two different staff and the results were compared. In addition, the words were checked against the master dictionary file compiled to date and new terms were flagged. These were checked to see if they were simply author misspellings, keying errors that had gotten through, or real new terms being introduced for the first time. All terms were standardized to American spelling. The unique word dictionary was not a dictionary in the traditional sense of the term. It was a compilation of unique words that had been taken from titles and checked as thoroughly as is humanly possible and it grew in size over time. It allowed us to identify when new terms or phrases were coined and to track changes in science from a unique perspective, including the frequency by which a certain term was used during a given time period. This is a very simplistic description of the UWD. It actually was made up of several files: a file of words having 12 or fewer letters; a file of words containing 13-30 letters; and a cross-reference file that included variant-to-preferred spellings of words. In addition, there was a file of two-word "terms" created by the editors if they believed it was necessary for accurate search and retrieval. As I mentioned earlier, quality in all of its manifestations was an ISI goal and information scientists such as Irv Sher and George Vladutz were unbelievably innovative in developing systems that would provide accurate search and retrieval. Did errors get through? Yes, and the systems immediately were modified so that the probability of the same error happening again would be pretty low. I should note that the indexing process for the chemistry products was quite different. The indexes were created by chemists who would apply standard nomenclature rules to create the names of the new compounds that had been indexed, along with other terms that would identify relevant biological activities, new synthetic reactions, etc.

SB: How did the ISI decide which journals to cover?

BL: There was a set of criteria by which a journal was measured before being added to a specific product line. Journal evaluation was a never-ending process that was used not only to review the new journals under consideration, but also to review those currently covered to see if such coverage remained appropriate. The criteria included the timeliness of the journal - did it have and meet a regular publication schedule; were the articles written in English; were author-abstracts included; did it conform to standards for article publishing (e.g. have descriptive titles, author names and addresses, full references to cited materials, funding information for the research, etc.); were the articles peer-reviewed; was the publisher known and respected. If it was not a brand new journal, we would look at the

citations to the journal as a measure of acceptance in the market and the quality of research that it published.
Respected abstracting and indexing services serve as a "marketing" arm for publishers. They offer a unique distribution channel – exposing journals before the eyes of thousands of scholars and researchers around the globe. Therefore, it was very important to have published criteria and to strictly adhere to those criteria so that one could clearly justify exclusion of a title to a journal editor or publisher and still maintain a good relationship with him or her.

I took a quick look at the current selection criteria posted on the Thomson Reuters site. It is pretty much the same, with the addition of criteria for electronic journals, international diversity for global markets and specific criteria for material focused on region-specific products (http://isiwebofknowledge.com/benefits/essays/journalselection/).

SB: You have been involved in database publishing for a long time. How do you see the future of the secondary publishers? How will models such as Google Scholar that rely on parsing the full text of documents affect the commercial databases and in what respect? How will services such as PubChem affect the commercial vendors of chemical property information?

BL: I believe that the current climate of change in scholarly communication will impact *all* publishers, both primary and secondary. A 2008 blog entry by Clay Shirky (http://us.penguingroup.com/static/html/blogs/tools-and-transformations-clay-shirky) says it all. The Internet, like the printing press before it, has created an information revolution that is generating new forms of scholarly communication and publishing. That said, I will focus my comments on the Abstracting and Indexing (A&I) world. The concept of an A&I service was first noted in 1665 with the creation of *The Journal Des Scavans.* The journal's primary purpose was to catalog and provide a brief description of the principal books then being printed in Europe, as well as to provide readable and critical accounts of current scholarly writings. Its goal was to facilitate information discovery and to minimize information overload. A&I services as we know them began to emerge in the early 1800's when there were approximately 300 scientific journals. Since then their purpose has never changed: They play an essential role in allowing scholars to navigate masses of information with relative ease. The bibliographic pointers such as keywords, subject indexes, authors, titles, etc. facilitate the discovery of information; abstracts allow the evaluation of a document's relevance to one's research; and links – either a bibliographic reference, or in today's world, an electronic link, allow retrieval of the full text. And as over the years these services build a body of information, they serve as the continuum between past, current and future scholarly thinking upon which all human knowledge is built. This is the essential role that organizations such as CAS and ISI play even today. They began when scholarly communication was print-based and they have adapted; we now progress through a transition consisting of both print and digital media.

You have raised two issues, the first dealing with Google Scholar (and this can be extended to all free information on the Web) and the second dealing with scholarly information services that are available from the government or have been established using an open access business model such as the Public Library of Science (PLoS).

Based on survey results that I have heard NFAIS members quote, researchers use Google 100% of the time for concept searches and to obtain ideas. Who doesn't use it? John Regazzi reported on this trend almost six years ago (http://www.nfais.org/page/42-john-j-regazzi-2004). But when researchers become involved in a specific project they turn to the more traditional services offered by their libraries or information centers in order to obtain their information, and they do so for two reasons: 1) they know that these services cover the source material in which the vast majority of scientists and scholars publish (Google Scholar does not); and, 2) they know that these services provide authoritative, reliable content (all Google content is not reliable).
To the extent that a free A&I service such as PubMed offers the same authoritative content and comparable coverage as a fee-based service, researchers will use the free service if it meets their requirements; if not, they will use a fee-based version if one is available to them. Fee-based products based on MedLine are a good example. There have been many competing variations of MedLine over the years and they have done well based upon the features and functionalities that their creators built around the content. They created "value" that could be measured by the user. A&I services need to continue their never-ending investment in the creation of measurable value.

Open Access journals are covered by most A&I services. I view such journals as an alternative to the traditional *primary* publishing model. Even the venerable publisher Springer Verlag has moved into the open access arena, acquiring BioMed Central in 2008. Open Access journals are not head-to-head competitors to A&I services.

Having said that, I do believe that the well-established A&I services are vulnerable if they do not pay attention to the new forms of scholarly communication. Their charter is to facilitate the discovery of and access to scholarly and scientific information. As the primary basis of that communication (journals) evolves into a more dynamic, online, collaborative "conversation," they must adapt their services to capture and preserve the content of the conversation. Not easy, as to do so one must deal with issues of authority (credible content), privacy, ownership (copyright), etc. But they must ensure that they deliver products offering ease of access to *all* the available information that is needed by their particular user base – no matter what the source. Traditional A&I services have the knowledge and expertise to be the A&I services of the future. But they must embrace the new forms of scholarly communication today, not ignore them, and not "wait and see."

I see the biggest hurdles to their future being the fact that to offer high quality A&I services requires a significant ongoing investment. During the journal explosion of the 1960's and 70's many questioned their ability to survive. With the help of technology, many did; others were ultimately acquired by stronger organizations. The information explosion sparked by the Web is having a similar impact on the growth of information. This, combined with the constant struggle to identify new business models, makes them vulnerable in the long term to new, creative competitors who can freely experiment with business models as they have no "baggage' (existing revenue streams) that could be threatened.

But if the A&I community is aggressive in creating new value-added products by leveraging their well-honed skills on the growing body of Web-based literature – creating the "A&I seal of approval" for scholarly users of the Web – their future could be secured. Bottom line, survival for all traditional content providers - including libraries – is to insure that they are providing value as measured by the user.

SB: You have held many elected positions at ACS, and you have also served as editor of the Bulletin. What was your role as editor and how did you put the issues together? Of the many roles that you have played in ACS, in general, and in the Chemical Information Division (CINF), which one was most interesting and satisfying to you and which one, in your opinion, has made a difference for ACS and CINF?

BL: Putting the *Chemical Information Bulletin* (CIB) together was a manual labor intensive process when I was editor (1977-1983). There were three printed issues per year. I had to solicit articles and advertisements and create the actual typewritten materials (there were no personal computers). I was fortunate, though. My boss, Gabrielle (Gaby) Revesz had been editor before me and was very active in CINF. I was permitted to use the talents of the "paste-up" artists that put together the print editions of *Index Chemicus*. They did the actual copy and layout work and prepared the final copy for the printer on huge sheets of paper. We used the same printer used for ISI's chemistry products (CINF paid the cost of printing). The only pain process was the mailing. We would get the division mailing labels in zip code order from ACS Headquarters. We then had to manually apply the labels to the printed Bulletins, bundle them by zip code (bundles had to consist of 10 or more *CIB*'s going to the same zip code), and put them in mail bags from the post office (supplied by ISI's wonderful mail room staff). The bags were then hauled (not by me) to the 30th Street Post Office a few blocks away where we had a non-profit license to mail the copies. I have to say that putting CIB together in those days was interesting and fun and many of the indexing staff participated.

You can see their names on the masthead at: http://digital.library.unt.edu/permalink/meta-dc-5684:2 including that of Marge Matthews who was an Assistant Editor and who eventually took over as editor when I stepped down. The ISI artists created the hand drawn cover designs for each issue (see an example at http://digital.library.unt.edu/permalink/meta-dc-5694:1) and even cartoons on occasion – it became a tradition since the *CIB* was edited by a series of ISI staff members over a long period of time.

The most satisfying role that I have played in ACS in general was when I served on the Committee on Nominations and Elections (N&E). N&E is the recipient of complaints about the nominations and elections process and has seriously spearheaded changes over the last decade. I served for six years (2000-2005) and played a role in making the election process more equitable and in making sure that Divisions were being adequately looked at to fill elected positions. The committee is often criticized because its work is confidential. It identifies potential candidates for elected committees (except for N&E, that is done by the Council Policy Committee (CPC)), for ACS Directors and for ACS President-elect. Behind closed doors there is much discussion about the proposed nominees' qualifications and the development of a rank-ordered list of names takes place. N&E attempts to ensure that the most qualified people are asked, that diversity is achieved and maintained, and that Divisions and Local Sections are treated equally. Hence, confidentiality is an

absolute requirement of the process. I am proud of what the committee has accomplished in getting the ACS Bylaws changed to make the process more equitable and glad that I was able to be a part of that process. CINF is fortunate that Andrea Twiss-Brooks is now representing Division needs on N&E and being a part of the change process.

With regard to the various roles I have played for the Division – Chair, Secretary, *CIB* editor , and Councilor– each had or has its own fulfilling rewards. When I was Secretary the Division won the award for best annual report from a medium-size Division. That was exciting for me and for CINF. There were no report forms to fill out at that time, it was all free form and you could include as much or as little as you wanted. I recollect that our winning report was in a 2" binder – and it was a pure marketing tool for CINF (I was at Wharton at the time and I think I let my MBA mindset take over - I wrote it almost as a business report. My own secretary helped and chided me on my verbosity). To this day, when we get together for lunch, that report finds its way into the conversation.

I think that I have had a chance to most effectively serve the Division in my role as Councilor over the past seventeen years. My longevity has given me visibility and as a result I have been appointed or elected to committees where I could impact how Divisions were perceived, recognized and rewarded. Most recently I have been asked to participate on a Task Force on the electronic dissemination of meeting content that will have its first meeting on February 2010. Hopefully, that will lead to something for CINF.

SB: This is the first issue of the *Bulletin* that will be produced only online. What would you like to see in future issues? How could we make it more interesting and relevant to the chemical information community?

BL: I thoroughly enjoy reading your interviews. The personal history is fascinating. I would like to see articles on information industry trends in general as well as how those trends impact the flow of scientific and scholarly communication. This could include articles on information policy and copyright legislation. Perhaps we could include a summary at year-end - sort of a look back at the highlights of the year with regards to technology changes, mergers and acquisitions, new products, meeting highlights, etc. Not the Division annual report, (although it could include a link to the report) – I mean a much more global, industry wide overview with links to relevant sites. Specific initiatives or technologies discussed at the ACS meetings could be summarized with links to podcasts. The electronic format opens up a lot of possibilities that CINF can pursue. As an aside, and not really related to the *CIB*, we could use Val Metanomski's CINF history as the foundation of a wiki-like history of Chemical Information with links to items from the Chemical Heritage Foundation and other sources. Perhaps pieces of the history could be "reprinted" in *CIB* and expanded upon using links to relevant sources as an ongoing serial. Or we can highlight the history of specific technologies or companies of interest to CINF members. Best I shut up unless I am willing to contribute!

Past and Present of FIZ CHEMIE Berlin: Interview with René Deplanque, Managing Director

Svetla Baykoucheva

René Deplanque

Dr. René Deplanque was born in 1950 in Lübeck, northern Germany. After completing his „Diplom“ studies as a chemical engineer in Hamburg, he obtained a Master's of Science degree in electrochemistry and a Masters of Philosophy degree in physical chemistry at Sir John Cass College, Guild Hall University, London. After subsequently gaining his Doctor of Philosophy degree in engineering sciences, in addition to his research activities, he taught electrochemistry and physical chemistry at the University of Cambridge. After returning to Germany in 1983, he worked for Brown Boveri in Mannheim where he was Head of the Electrochemistry Laboratory and Head of the Computer Department for the company's German operations, and also served as assistant to the Board. In 1988 he moved to the Gmelin Institute of the Max Planck Gesellschaft and was Head of the Gmelin-Online Project. In 1994 he became Managing Director of the Fachinformationszentrum CHEMIE GmbH in Berlin (FIZ CHEMIE Berlin; the Chemistry Information Center); a position which he holds to the present day. In 2000 René Deplanque was appointed Professor of the Technical University Berlin. He is a member of various national and international committees. He acts as adviser to several organizations and governments.

More information on FITZ CHEMIE Berlin is available at *www.fiz-chemie.de/en/center/fiz-chemie/company.html*

Svetla Baykoucheva: Many readers of the *Chemical Information Bulletin* are familiar with FIZ CHEMIE, and some members of the ACS Chemical Information Division (CINF) have even been directly affiliated with it through their employment. It would be interesting to hear from you about some important moments in the history of the organization, what its philosophy is, and what distinguishes it from other players in the chemical information field. What are the most important things you would like our readers to know about FIZ CHEMIE?

René Deplanque: Our history goes as far back as1830 when we were founded as "pharmaceutical abstracting service." In 1840 the name was changed to "Chemisches Zentralblatt," which existed until 1969. We continued as a Department of the German Chemical Society and produced the first ChemInform Reaction Handbook in print. In 1981, the German Center for Chemical Information, FIZ CHEMIE Berlin, was founded, and the ChemInform editorial board formed the basis of this institute.

FIZ CHEMIE Berlin (FCH) is now the leading German information center for chemistry. As an Institute of the Leibniz community, it cooperates actively in the organization of the information society in Germany. The major task of the Institute is to provide scientists, teachers and the industry with high quality information services for general chemistry, chemical technology and adjacent areas. The FCH mission statement is: "We want to make information understandable and usable." This is done along this line: research—development—production—distribution—support and happens by: (1) Providing standards, (2) Creating networks, (3). Establishing processes in information science, and (4) Establishing and following editorial processes.

FIZ CHEMIE Berlin is a non-profit company whose partners are the federal government, the counties, and many private institutions.

SB: Which organizations and individuals are the main users of your products?

RD: Our products service a large number of customers in research, industry and in the teaching process. We are servicing large research networks as well as practically all major companies of the chemical industry. We advise governments and governmental bodies. This advice is given nationally as well as internationally and, with our very large e-learning systems, we supply all German schools, Universities and the industry with high quality teaching material. Just within the teaching area we had in 2009 nearly six million single users. Therefore it is very difficult to pinpoint individuals or organizations.

SB: Could you tell us something about your personal life and interests? What made you switch from electrochemistry to chemical information? What do you do outside of your professional work?

RD: During my study times I collected degrees as other people are collecting stamps and this happened in a variety of subject areas such as chemical engineering, physical chemistry, electrochemistry and corrosion engineering. Only within chemical information could I combine all those areas. What do I do outside my professional work? I play the guitar, classic and modern blues. I am a member of the Magic Circle in Berlin, and an associate of the Magic Club in Nice. Being a magician is very helpful—if I can't solve a problem the normal way, I have always something to fall back on. I also love to cook, and I am producing ancient ship models.

SB: Several years ago, I attended the First German Conference on Cheminformatics, which took place in the beautiful town of Goslar. The audience at the conference was mainly German, with a few foreign participants. I enjoyed very much both the conference and the social program, but I was struck to see so few women there. If I remember it correctly, there were 125 attendees, and I could count no more than 10 women. I have attended many conferences in Europe and the United States (not only in chemical information, but also, as a lab-bench researcher, conferences in such fields as chemistry of natural products, biochemistry, microbiology, and molecular biology), but I have never seen such a small percentage of women at any conference that I have ever attended. Is the field of Chemical Information mostly male-dominated, or this is typical of other areas of science in Germany?

RD: If you would come to Goslar now, you will find that it is still a nice town; the conference is still very interesting and it became fully international, doubling in size. You will find that there are many more women now than there were in the first Goslar conference. I do not think that we have a gender problem in the sciences. But in the past there were by far more male than female students in the sciences. Luckily, this changed dramatically and now we have nearly an equal percentage of male and female students. Coming back to the area of Cheminformatics, you will find that this is still a specialized area and that it is taught at only very few German universities. It is, indeed, still male-dominated, but in recent years the number of women is increasing.

SB: In the academic institutions in Germany, what is the role that librarians play in supporting research and education? What is the status of librarians in these institutions? What education is required for such positions? What does a generic job description for a chemistry librarian in a German academic institution look like?

RD: There is quite a difference in the position of an American librarian and a German librarian. In Germany, the librarian has still a very traditional role. They are the ones to whom you go only to get information. Unfortunately, the librarian is seldom in the position of partner to the researcher. This is unfortunate because in most cases, the librarians have to have an academic degree in librarianship to work in a scientific library. Only very rarely do scientists decide to go into a career as a librarian. I think that one of the problems that we have to cope with is that universities depend on reputation, and reputation depends on research; librarians in Germany are not part of the research process. Therefore, the librarians do not have their true position within this framework. In recent years, this very traditional positioning of the librarian is weakening a bit, but it will still take many years before the librarian is accepted as a valued partner for the scientist.

SB: Many chemistry resources require a steep learning curve. In the United States, chemistry librarians do a lot of instructional classes for end users. How is the training of researchers and students organized in Germany?

RD: The librarian in Germany is highly educated but only very seldom introduced into the education process, especially not in the sciences. Very few universities have courses on teaching the use of information systems that are actually led by a librarian. It is very sad that, within the educational system in

Germany, the true value of the librarian is underestimated, and because of this librarians are not used to their full potential. The training of researchers and students is organized by the research professionals within the framework of their department. Lectures are also given by scientific assistants and professors. This is true for both regular universities and universities in the applied sciences.

SB: Are there any specific cheminformatics areas that are predominant in Germany? Which are the institutions in Germany where research in this field is done? Where do researchers in this area publish their papers?

RD: The centers of cheminformatics in Germany have changed. In the past it was always the University of Erlangen that played this role. Now, the universities of Bonn and Hamburg are also strong in this field. Cheminformatics is still pretty small compared to Bioinformatics. The researchers mainly publish in the ACS' *Journal of Chemical Information and Modeling*, as well as in the *Journal of Cheminformatics*.

SB: FIZ CHEMIE has been very generous in its support for the ACS Chemical Information Division (CINF), in general, and the Chemical Information Bulletin (CIB), in particular, and we are very grateful for this support. What is the relationship between the German Chemical Society (GDCh) and ACS and between the Chemistry Information Center in Germany and CINF? What are the main past and current initiatives and what are the plans for the future?

RD: There are many consultations and close relations between the ACS and the GDCh. One example of this cooperation is the formation of a working group between the divisions of Computer in Chemistry (CIC) of the GDCh and CINF of the ACS. The task of this working group is the development of a repository of teaching materials for chemical information. The content will be provided by the major universities of both countries and will be open access. The first version of this repository will be launched within this year. This kind of interdivisional working group will be extended into other areas.

SB: In the late eighties, while attending a molecular biology conference in the Harz Mountains, I had the opportunity to visit East Berlin. Two years ago I went back to Berlin for a conference on scientometrics and was thrilled by the new spirit and the infectious mood of the city. One evening, while crossing a bridge on the Spree River, I saw people dancing on a boat. What happened in Berlin after the reunification that it now competes with Paris in atmosphere and liveliness? How did the unification of Germany affect you personally and professionally? What was its impact on the chemical information field?

RD: Berlin developed into a very young city, as after the reunification many artists and young people moved there. It came to a total remake of a capital. Practically all famous architects took part in building new city centers and exploring ways to redesign and rebuild the old and ancient parts of this very large city. If Berlin is having a party (and there are many of these during a year), it is expected that at least one million people will attend. When this drops under 700,000, it will be seen as too small and will be dropped.

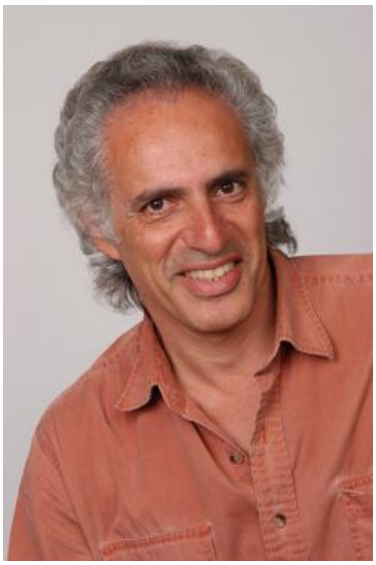
I moved to Berlin just after the reunification and experienced the tremendous spirit which was developing here. The merger of two completely different cultures and the renewal of a totally run down system is something that can only be experienced and would be very difficult to explain. What we had was a real revolution, but no one was dead and nobody was hurt (at least, not physically).

West Germany became more East and East Germany became West, with all the problems and difficulties which are inherent to such a process. The impact which this had on the field of chemical information was that we were able to employ many new and experienced scientists, who came to us with a different background and with new and different ideas. Also, new and different collections of data were found. They are normally hidden in the treasure chests of the chemical industry, like our now large collections of thermodynamic data. More than 25% of our staff are from former East Germany and are highly respected for their knowledge, their achievements, their work, and their humanity. In the sciences we are already united. As far as the society as a whole is concerned, there is still some work to be done.

SB: René, thank you for sharing with us your interesting perspective on cheminformatics in Germany.

A Philosopher's View on the Periodic Table of the Elements and Its Significance: Interview with Eric Scerri

Svetla Baykoucheva



Eric Scerri

"Imagine the sight of the snow under a ski chairlift on the first sunny day after a storm, when untracked powder tempts you from below. You know that no matter what, once you hit the snow, it's going to be a great day. Some runs will be steep and full of bumps, some will be easy cruisers, and some will be tricky routes through trees. But even if you take the occasional wrong turn, most of the day will be wonderfully rewarding."

Lisa Randall (2006) Warped Passages: Unraveling the Mysteries of the Universe's Hidden Dimensions, p. 8.

Eric Scerri is originally from Malta. He obtained all his degrees in the UK. His bachelors and masters degrees were in chemistry from the universities of London and Southampton. His PhD was in history and philosophy of science from Kings College, London. He went to the US as a postdoctoral fellow at Caltech fifteen years ago and has remained in the US ever since. For the past 10 years he has been teaching chemistry and philosophy of science at the University of California at Los Angeles.

Scerri is one of the founders of the field of philosophy of chemistry as well as the founder and editor of the journal *Foundations of Chemistry*, now in its twelfth year of publication. Dr. Scerri has written what reviewers have called "the definitive book on the periodic table," *The Periodic Table: Its Story and Its Significance*. He is also the author of two books with Imperial College Press in London. He is currently writing two more books with Oxford University Press.

Scerri is also the author of over 100 journal articles in chemistry, chemical education and history and philosophy of science. In recent years he is turning increasingly towards working in the TV and radio media and to presenting his work to the general public.

Svetla Baykoucheva: In an interview for *Nature Chymist*, a *Nature Chemistry* blog, when asked which historical figure you would like to have dinner with, you said this would be Dmitrii Mendeleev (http://blogs.nature.com/thescepticalchymist/2008/01/reactions_eric_scerri.html). In your book, out of 10 chapters two are devoted to Mendeleev. What is it about him that has struck you most? If you could really have dinner with him, what questions would you ask him?

Eric Scerri: I am struck with the fact that he was so far ahead of the other chemists who groped their way towards the periodic table. I am also amazed by the confidence with which he believed in the periodic law which he formulated and the even greater confidence with which he predicted the existence of new elements. Allow me to quote from his paper which appeared in German in 1871, just two years after his initial periodic table of 1869: "…we can see, by referring to tables I and II, in which the periodic relations are shown, that many elements are missing, and we can confidently predict their discovery. I am therefore going to describe the properties of several as yet undiscovered elements. In this way I hope to demonstrate in an accurate and perfectly clear manner the adequacy of the law, although confirmation of these proofs is reserved for the future."

I would ask him whether he really had not seen the rudimentary periodic tables of De Chancourtois, Newlands and Lothar Meyer, before publishing his own table, as he later claimed. I would also ask him about his philosophical views on the nature of 'elements' as this has become a much discussed topic within contemporary philosophy of

chemistry. Mendeleev certainly had profound views on the issue and referred to them even in the first of the two volumes of the first edition of his famous book. He did this even before formulating the periodic table, which he did while considering how to make a transition to the remaining elements that he wanted to discuss after those in volume one.

Dmitrii Mendeleev
AIP Emilio Segre Visual Archives

SB: An excellent book on Mendeleev was written by Michael Gordin (Gordin, M.D. (2004) *A Well-ordered Thing: Dmitrii Mendeleev and the Shadow of the Periodic Table.* Basic Books, New York). Your book is the most comprehensive work entirely devoted to the periodic table. I can admit that I read it like a novel. How did you do the research for it and what is generally your way of writing? What resources did you use?

ES: Thanks very much for saying so. It is the single most intensive piece of work I have ever done and it involved a fair share of trials and tribulations. The research for the book stretches back for a period of about twenty years to my PhD thesis in history and philosophy of science, which was on the reduction of chemistry to physics. But rather than trying to deal with the whole of chemistry, I decided to concentrate on one very central aspect of chemistry, which embodies so much chemical knowledge and information—the Periodic Table—and to ask how, or to what extent, it has been reduced to physics, in general, and to quantum mechanics, in particular. As for my sources, I would like to think that I have read everything published on the subject, in many cases a number of times.

The other 'source,' of inspiration at least, goes back to my childhood and the fact that I have always admired the orderliness and compactness of the periodic table, the feeling that all of chemistry was somehow contained in this elegant chart.

SB: I read somewhere that you had been involved with some TV productions. What did you do for TV and what are the projects you have been working on now? What is in your future plans?

ES: I have been a consultant and interviewee in a couple of series on the periodic table, one in the UK and the other one in the US, both of which are yet to air. I was also part of a one-hour radio show made by the Canadian Broadcasting Corporation, CBC, called "The Music of Matter." It consisted of extended interviews with John Emsley, Oliver Sacks and with me on the subject of the periodic table. I am very interested in getting more involved in this kind of work, perhaps even on the production side, and I invite anyone reading this interview to get in touch with offers or suggestions. As you know there is far too little on chemistry in popular science TV and radio, and of course not enough popular science books centered on chemistry. I hope to do my bit to redress this imbalance as much as possible.

SB: You are evidently very interested in predictions in chemistry, in general, and in the development of the periodic system, in particular. You have a sub-chapter in your book, "Mendeleev's less successful predictions." How do philosophers of science view the failed predictions of scientists, and how are such predictions treated, especially if they have never been published? Should they be counted as failures?

ES: Thank you for raising the issue of predictions because I want to clear the record on this. When an article describing my book appeared in the *New York Times*, the author claimed that I was trying to minimize the achievements of Mendeleev and was claiming that many others had discovered the periodic table before him, especially the Frenchman De Chancourtois. In fact I have the utmost respect for Mendeleev's work, although a historian is obliged to lay-out the record of discoveries as he sees it. My reason for devoting a great deal of attention to predictions is because of a long-standing debate in the philosophy of science which seeks to establish whether scientific theories and developments are accepted mainly because of successful novel predictions or perhaps for their successful accommodations of already known facts.

One of the pioneers in this question has been Professor Stephen Brush who you may have known before he recently retired from the University of Maryland at College Park. Brush argued that in a number of key theories in physics, for example, it was successful accommodation of already known phenomena that was in fact more decisive in the acceptance of the theories in question. For example, he and others have claimed that Einstein's theory of general relativity was primarily accepted for correctly explaining the long-known advance of the perihelion of the planet Mercury and not for the dramatic and successful prediction that Einstein made of the bending of starlight by massive gravitational objects. But when it came to chemistry, Brush appeared to vacillate and claimed that this was one case in which successful predictions had indeed been more decisive than accommodation of already known elements in the acceptance of the periodic table.

I wrote a long paper with the philosopher John Worrall in which we argued that accommodation of elements by Mendeleev's table may also have been more decisive, or at least as decisive, as the predictions in the acceptance of the periodic system. This is why we examined his predictions so closely, including his failed predictions. The point is that if predictions had been regarded as being so decisive, one would expect the failed predictions would have counted against the acceptance of the periodic table. But there were several failed predictions and yet the system was accepted, perhaps because the outcome of predictions was not all that crucial contrary to the popular accounts.

SB: With my previous question, I wanted to find out whether, when talking about Mendeleev's unsuccessful predictions in your book and later in one of your papers (Scerri, E. R. (2008) The past and future of the periodic table. *American Scientist* 96 (1), 52-58), you have counted as his failures even predictions that had not been published officially. Making predictions and having hypotheses is what scientists do all the time. This is how science is done—speculate and prove whether you were right. Should such failed predictions, if they have not been published officially, be taken into account when evaluating how successful a scientist has been in making predictions?

ES: I agree with you that unpublished predictions should not count in the assessment of a scientist's work. When I claimed in my book that nine out of eighteen predictions made by Mendeleev had been successful, I was referring to predictions that he actually published at one time or another. This includes his prediction of two elements that he believed to be lighter than hydrogen and his prediction of ether as an element, which he made in 1905 a couple of years before he passed away. Some authors regard these predictions as the speculative outpourings of an elderly and established scientist with nothing to lose and I suppose one must agree with that, although these ideas appear to have been well thought-out and appeared in a short-book form.

SB: Has Mendeleev betrayed some of his predictions that he based on the atomic weight of the elements for the purpose of accommodating their chemistry, and why did he do that (for example, putting tellurium (Te) before iodine (I)?

ES: This is another very subtle and interesting issue. You are referring to the pair-reversals. In the case of Te and I, adhering to the then known atomic weights should have placed these elements in chemically incorrect groups since iodine had a lower atomic weight. Mendeleev, and incidentally, Newlands and Odling before him, all placed Te before I because of the chemical similarities which each of these elements had shown with the oxygen group and the halogens respectively. This much is true, but Mendeleev did not really betray his belief that the ordering principle lays in using atomic weight. Instead, he repeatedly insisted that either the weight of tellurium or that of iodine had been incorrectly determined. Contrary to the popular account, we might say he did not simply put more emphasis on chemical properties, but instead maintained that the atomic weight ordering was the exceptionless criterion for ordering the elements. But throughout his life, the repeated attempts that he made to establish that Te has a lower atomic weight have failed. Of course the correct ordering principle is atomic number, and not weight, but this would only emerge in the second decade of the 20th century following the work of van den Broek and Moseley.

SB: The main theme in your book is the relationship between chemistry and physics and to what extent quantum mechanics explains the periodic system of elements. From the point of view of a philosopher of chemistry, can chemistry be reduced to physics?

ES: First let me say that I do not try to answer the 'in principle' question of whether chemistry *can* be reduced to quantum mechanics. My emphasis has always been on the extent to which chemistry has in fact *been* reduced to quantum mechanics. After all, who really knows what is possible in principle or what might be possible via a future development in quantum mechanics? On the other hand, one can carefully examine whether present day quantum mechanics does a good job of explaining in a more fundamental manner chemical phenomena such as perhaps the form of the periodic table. I have claimed that although quantum mechanics gives a perfectly good *ab initio* explanation of the lengths of periods of elements it does not fully explain the precise order of shell filling which is summarized by the $n + \ell$ or Madelung rule which is familiar to anyone who has studied college or even high school chemistry. And let me repeat—I am not claiming that such an explanation is not possible in principle but just that nobody has yet succeeded in deriving this rule from the general principles of quantum mechanics.

SB: How do scientists working in individual disciplines relate to and treat each other? Are there disciplines that are considered superior to other disciplines? When I was in college, I remember that physical chemists were kind of elite. I can admit that this might have played a role for me choosing to do a Master's degree in physical chemistry. I later spent most of my scientific career working in areas closer to organic chemistry (lipids and fatty acids).

ES: Yes I agree with you, physical chemists and even more so theoretical chemists consider themselves somehow superior to other specialists within chemistry. And theoretical physicists consider themselves even more superior. Some of this one-upmanship is perhaps inevitable since the more theoretical the discipline, or the more one moves towards physics, the more the scientist is operating at a more fundamental level. But more fundamental in the literal sense should not be mistaken to mean superior since work at the less fundamental levels is just as essential and in most cases remains indispensable. This is the paradox of reduction in all the sciences. We gain knowledge of larger objects in terms of their microscopic components and the way they are organized and yet the detailed knowledge must still be acquired by painstaking work at each of the appropriate levels in the scientific hierarchy.

SB: Another major theme that runs through your book concerns the nature of the element. This question has puzzled Mendeleev and has significantly influenced his views. I would like to quote something from your book: "Mendeleev held a dual view on the nature of elements, where they could be regarded as unobservable basic substances and also as Lavoisier's simple substances at the same time. Mendeleev thus acknowledged one of the central mysteries running throughout the long history of chemistry, which is the question of how, if at all, the elements survive in the compounds they form when they are combined together. For example, how can it be claimed that a poisonous gray metal like sodium is still present when it combines with a green poisonous gas chlorine, given that the compound formed, sodium chloride, or common table salt, is white and not only nonpoisonous but also essential for life?" How does the philosophy of chemistry deal with this question and what diverging views exist in this area?

ES: There is a sense in which the question you are alluding to now lies at the very heart of philosophy of chemistry. Let me use a well-worn example. When iron filings are mixed with sulfur, we obtain a physical mixture of two elements. But when sodium reacts with chlorine, we obtain a compound in which the component elements are not as easily recovered and also seem to have been radically transformed. To me the chemical change is far more magical and far more mysterious. Understanding such chemical changes requires its own philosophical approach. Mendeleev and others provided the beginnings of such an approach. We need to regard elements in two senses, or maybe even three senses. First there is the element as a simple substance such as grey sodium. Then there is the sodium, which is combined as in the case of sodium chloride. And thirdly there is what is common to both of these forms of the element. The third option has been called "the element as a 'basic substance'", a somewhat unfortunate terminology in view of the unintended reference to acid-base properties. Mendeleev believed that it was this more fundamental, more philosophical sense of the elements that were represented on the periodic table and not the other two senses. This is not so hard to appreciate. For example, the halogens, as simple isolated substances, are rather different and they even display all three states of matter at room temperature. They are not grouped together because of their similarities as simple substances. On the other hand, their compounds with sodium, for instance, are all very similar. The properties of the combined elements appear to be more decisive. Philosophers of chemistry are currently trying to clarify these distinctions, especially given the centrality of the concept of 'element' in chemistry. There is a good deal of disagreement with some wanting to regard the most fundamental sense of element as metaphysical, others—as merely abstract— and so on.

SB: Although it took you seven years to finish this book, it has been on your mind for 20 years. Why is the periodic table philosophically important?

ES: Quite simply put the periodic table embodies the heart of chemistry and the relationships among the elements. It is by far the biggest idea in modern chemistry, equaled only by the notion of chemical bonding to which it is of course strongly related. The nature of the periodic table defies the usual ways in which scientific concepts are classified. It is not a theory or a model but, in a sense, it is just a representation and yet one that is capable of establishing far-reaching correlations. Attempts to explain the periodic table by people like Bohr and Pauli have produced important concepts and have helped to shape quantum mechanics itself. Instead of thinking that quantum mechanics explains the periodic table, I prefer to think of the periodic table as having helped to give birth to quantum mechanics.

SB: The development of the periodic table has involved interplay between theory and experiment. How have the philosophical deliberations of Mendeleev about the nature of the element influenced his predictions?

ES: This brings us back to the philosophical understanding of the concept of an 'element'. In my book, and some articles, I have claimed that Mendeleev's willingness to go beyond the facts, to make predictions and to correct the atomic weights of many elements, was based on his belief that the periodic system was primarily a classification of the elements in their more abstract and fundamental sense. Had he been more tied to the properties of the elements as simple substances, I don't believe that he would have been able to go beyond the known facts on so many occasions. He had a deeper appreciation of the nature of the elements than his competitors and, like all truly great scientists, he was as much a philosopher as a technical scientist.

SB: As you have mentioned in a recent article, through the years, many people have tried to create their own models of the periodic table. This somehow reminds me of a situation in France, where even today there are people who are proposing ideas about how to calculate the face of the circle. They send their ideas to the French Academy of Sciences, where each new proposal is registered (in the Archives of the Academy). You also have proposed your own version of the periodic table (Fig. 1), in which you have put H in the first column/group, together with the halogens. Did you do it this way because you consider the atomic weight to be more important than the chemical properties of the elements?

ES: Actually my proposal was based first of all on atomic numbers and secondly on the desire to create new atomic number triads. Triads using atomic weights provided what I have called the first hint that there were numerical regularities among the elements. Since atomic number is now known to be a superior ordering principle, it is worthwhile considering atomic number triads. If we place hydrogen above fluorine and chlorine, we obtain a new perfect atomic number triad in the sense that the atomic number of the middle element of these three, fluorine (9), is the exact average of the atomic numbers of hydrogen (1) added to that of chlorine (17). If hydrogen is left among the alkali metals where it most usually placed we have nowhere near to an atomic number triad H (1), Li (3) and Na (11). But I still have work to do in convincing the scientific community that atomic number triads are fundamental and not merely a form of numerology. I am in the process of publishing an article in which I argue that electronic structure is governed by the nucleus, which if true would strengthen the notion that atomic numbers and relations among them have a fundamental role to play in the question of the placement of elements in the periodic table.

SB: What are the hot areas of research in the philosophy of chemistry today?

ES: The question of reduction is still important as are the notions of natural kinds. Some of my colleagues are interested in whether chemical bonding should be regarded structurally or in a more realistic sense. There is also work being conducted on oscillating reactions, the philosophy of organic chemistry and of biochemistry. You can get some better idea by consulting copies of the journal *Foundations of Chemistry* which is now in its twelfth year of publication.

H	He	Li	Be											B	C	N	O
F	Ne	Na	Mg											Al	Si	P	S
Cl	Ar	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se
Br	Kr	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te
I	Xe	Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po
At	Rn	Fr	Ra	Lr	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Uub	Uut	Uuq	Uup	Uuh
Uus	Uuo																
		La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb		
		Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No		

Figure 1. This version of Eric Scerri's Periodic Table places H with the halogens. The table was featured in an article in American Scientist (Scerri, E.R. The past and future of the periodic table. *American Scientist* **96**(1), 2008, 52-58.

SB: In one of your papers you pondered on whether "our reliance on the two-dimensional forms of the periodic table is due to the predominance, until recently, of the two-dimensional textbook page surface and the two-dimensional nature of the walls of lecture theaters?" And you were wondering whether, "with the development of the new technologies, Mendeleev's famous icon might be transformed into something that even he might not recognize if he were here to see it?" In her brilliantly written book, "Warped passages: Unraveling the mysteries of the Universe's hidden dimensions," the theoretical physicist Lisa Randal talks about the possible existence of parallel universes, warped geometry, and an extra-dimensional world. How do you see the future development of the periodic table and the possible role that philosophy could play in it?

ES: I think that the most interesting question is whether it makes any sense to seek an optimal periodic table or whether most representations are ultimately just as worthwhile. I am a realist about chemical periodicity, that is to say I believe that the approximate recurrence of the elements after certain intervals is an objective fact of nature. It is not a convention that we impose on nature. Consequently, there is no question in my mind that some periodic tables are better than others and indeed that it is worth seeking an optimal table. I am also a little puzzled that many chemists, who usually espouse realist views on most scientific matters, often deny the notion of one best table. Let me stress that I don't mean what shape the periodic table should be presented in whether it be rectangular, circular, elliptical or pyramidal. I am referring to differences concerning the placement of certain troublesome elements such as hydrogen, helium. I cannot agree with textbook authors and periodic table designers who show hydrogen, or even hydrogen coupled with helium, alone hovering above the main body of the periodic table. This amounts to the claim that these two elements are not subject to the periodic law which I find distasteful. There is also a long-standing debate about whether group 3 should consist of Sc, Y, La, Ac or perhaps Sc, Y, Lu, Lr. I am in favor of the latter arrangement incidentally, again because of atomic number triads. Some of these topics do not involve technical scientific information but are to some extent meta-scientific. This is where philosophy of chemistry can play an active role. Thank you for allowing me to voice my views on my favorite topic.

Eric Scerri teaches chemistry as well as history & philosophy of science at UCLA
http://faculty.chemistry.ucla.edu/institution/personnel?personnel_id=294334

Books written by Scerri can be found at
www.amazon.com/Read-more/lm/R2ALNIIFUZIUCY/ref=cm_srch_res_rpli_alt_1

Book Review

Debons, Anthony. Information Science 101; Scarecrow Press: Lanham, MD, 2008, $50.00 (Paperback). 241 pp. ISBN: 0-8108-5289-6.

To begin, a disclaimer: this reviewer has never taken a course in either library or information science. However, over his multi-decade career in technical information searching, he has taken many training courses on database usage and related topics and has published in these areas. Therefore, this review is written with searchers, users, and subject specialists in mind. At least two reviews have appeared (1, 2), one favorable, one not.

As could be expected from the title, the book is aimed at students, both high school and college freshmen, as an introduction to the very broad field of information science (IS) and with an eye to a broad description of the jobs and careers available within the field. Chapter titles include introduction to information; professional identities and opportunities; IS: nature and function; the ADIK system (defined below); how do we acquire data, information, and knowledge (DIK); movement of DIK, transmission; the computer; decision making—problem solving; communications; ergonomics and IS; value and quality of DIK; security, privacy, ethics; future of IS/knowledge sciences.

The primary subject and associated processes discussed is the ADIK system: augmented data, information, knowledge. This system bears a strong resemblance to the numbers/ data/ information/ knowledge/ wisdom continuum often used by the reviewer in presentations on searching and database usage (and not original with him). Relations of all of these concepts with each other and their ubiquity in society and daily living are particularly good, especially the distinctions between information and knowledge.

Since this is a broadly based introduction, some of the many subjects are covered in a more elementary fashion than might be desirable at a higher level of education, background, or training. The index is not as thorough as it could be for effective use. Abbreviations are often neither defined in the text nor indexed. A few pioneers in the field are identified along with brief biographies but many notables are not. Gerard Salton and his work are notably absent. Citation searching is briefly described but Eugene Garfield is only mentioned with a footnote and, unlike the others, is not indexed.

Several related topics are discussed superficially and are not particularly up to date. For example, the basics of data transmission are discussed, but satellite transmission is inadequately described only in terms of GPS while DSL is not even mentioned. Artificial Intelligence is briefly described but could use more elaboration. Value and quality of DIK are discussed but not in terms of relevance, recall, or sheer volume. The last relates to Google and other search engines which do not appear at all. Costs are discussed only in terms of provision of information from libraries, and Open Access and Publishing do not appear. Thesauri and concordances are absent. End user vs. mediated searching needs more material, especially in these days of SciFinder, Google, etc.

Granted, this is an introductory text, but in addition to the above, several aspects of information retrieval of interest to those of us in the chemical information field are either absent or inadequately described. Meaningful descriptions are lacking for databases, their construction and use; computer based retrieval; relevance vs. recall; text, graphics, and structures; indexed vs. text or search engine searching; and of course, the 800 pound gorilla, Google.

The author, Anthony Debons, is emeritus professor at the School of Information Sciences, Univ. of Pittsburg. It can be assumed that this book is compiled lecture notes for one of his courses. Educators can probably find portions of this book usable in their course and other instruction. However, I believe that many of us would like to see more depth on a many of the topics. Information Science 102, 502, anyone?

References

1) Zins, C. *JASIS&T*; 2009, Dec2009, Vol. 60 Issue 12, p2600-2602, 3p

2) Hutchinson, A. R. *Choice: Current Reviews for Academic Libraries*; Oct2009, Vol. 47 Issue 2, p278-278, 1/6p

Robert E. Buntrock
Buntrock Associates
16 Willow Drive
Orono, ME 04473
207-866-7930
buntrock16@myfairpoint.net

ACS Chemical Information Division (CINF)
Spring 2010 ACS National Meeting
San Francisco, CA (March 21-25)

ABSTRACTS

CINF 1"Chemical Information Instructor" feature: Telling the story of successful faculty-librarian partnerships in the Journal of Chemical Education

A. Twiss-Brooks, atbrooks@uchicago.edu. John Crerar Library, University of Chicago, Chicago, IL, United States

Since 1992 the Chemical Information Instructor (CII) feature has been providing instructors with practical information on a wide range of topics related to teaching information-searching skills to undergraduates, graduate students, and other researchers. The Editors of CII have encouraged the submission of articles featuring creative and successful collaborations between librarians and teaching faculty. Some of these collaborations have included integration of information instruction into one or more chemistry courses, incorporating instruction using specific information tools, and addressing scientific information retrieval in non-science major courses. Information on CII, examples of interesting collaborations reported in the feature, and possible future directions will be presented.

CINF 2 Teaching students to use REAXYS in order to find synthetic methods for making benzylic acid from benzaldehyde

S. K. Cardinal[1], scardinal@library.rochester.edu, and W. D. Jones[2]. [1]Carlson Science and Engineering Library, University of Rochester, Rochester, NY, United States, [2]Department of Chemistry, University of Rochester, Rochester, NY, United States

In Honors Organic Chemistry, second semester sophomores search the chemical literature to find their own experimental procedures: one for oxidizing benzoin to benzil and a second for the conversion of benzil to benzilic acid. Because the literature on these reactions is extensive, they need to evaluate the methods using parameters that are mutually agreed upon in class. Together, a librarian and a faculty member have created a hands-on tutorial that demonstrates how Reaxys (formerly Beilstein CrossFire) may be used to identify relevant articles.

CINF 3 In depth chemical information instruction in a second year organic chemistry course at Indiana University-Purdue University Indianapolis

E. Snajdr, esnajdr@hotmail.com. University Library, Indiana University-Purdue University Indianapolis, Indianapolis, IN, United States

We are embedded librarians in a second semester organic chemistry course at Indiana University-Purdue University Indianapolis. Librarian involvement in this course consists of approximately six hours of instruction. The chemistry librarians lead the students in six separate modules which encompass the following topics: searching for analytical data, searching for a chemical structure and synthetic procedure, searching for a reagent, searching for an experimental procedure, and finally, searching for scientific papers by a known author. The course was developed in collaboration between the chemistry librarians and instructors of the course and has continued to evolve along with technology and information tools.

CINF 4 Qualitative analysis in the library

J. N. Currano, currano@pobox.upenn.edu. Chemistry Library, University of Pennsylvania, Philadelphia, PA, United States

This paper describes a long-standing collaboration between the University of Pennsylvania's chemistry librarian and the instructors of the advanced undergraduate laboratory course. The chemistry librarian teaches the students two intense lectures, complete with credit-bearing assignments, that teach them to use the chemical information resources. The first lecture and assignment review the structure of the chemical literature and teach the students to find physical properties and preparations of substances. The second lecture teaches students techniques of text and structure searching using truncation, preparing them to use the chemical information resources to answer open-ended questions, exemplified in the accompanying assignment.

CINF 5 Integrating library instruction into SAGES seminars at Case Western Reserve University for undergraduate chemistry students: A strategic plan

S. Guo, shu.guo@case.edu. Kelvin Smith Library, Case Western Reserve University, Cleveland, Ohio, United States

Chemistry is a unique science discipline which requires special searching techniques to locate its enormous collection of information. Systematic library instruction on the chemical literature and specialized search techniques required for chemistry and the related scientific disciplines have long been recognized as important. Case Western Reserve University has had its SAGES program in place for years, offering undergraduate students with a series of seminars to enhance their skills on learning, writing, and scholarly inquiry. Currently our librarians have been teaching general library collection/resources and searching skills in most SAGES First and University seminars. The strategic plan is to integrate library instruction sessions on chemistry resources and searching skills into SAGES's Departmental and Capstone Seminars. Through the joint efforts of science librarians and teaching faculty, undergraduate chemistry students will be systematically exposed to chemical information resources and specialized search techniques. Therefore, the students will be well prepared for their academic classes and professional careers.

CINF 6 USC Chemistry OIL program

N. Xiao, nxiao@usc.edu. Science and Engineering Library, University of Southern California, Los Angeles, CA, United States

With more information available on-line, more technologies available for users to access information virtually, and more interdisciplinary research shaping our collection, librarians are facing more challenges than before. For instance, how can we penetrate users' information search life and help them find what they want in numerous information resources that have ever-changing search interfaces and features? How do we shape our users' chemical information literacy in the e-resources environment? On the other hand, to what degree can we partner with users, especially faculty members, to provide the customized information services and collections that they need? The USC Chemistry Outreach and Information Literacy Program (OIL) serves to reach out to users and to collaborate with faculty members to provide customized, systemic, broad, and in-depth information literacy training (e.g. research group, course-related, project-related). This presentation will answer the questions posed above and offer some best practices in sustainable chemical information education.

CINF 7 Beyond Google: Integrating chemical information into the undergraduate chemistry curriculum

M. C. Peters, mpeters@library.ucla.edu. Science Engineering Library, UCLA, Los Angeles, CA, United States

At UCLA the focus is on partnering with faculty and teaching assistants in integrating the basics of chemical information into three sequential courses required for chemistry and biochemistry majors. Resources "beyond Google" include many recommended in "Information Competencies for Chemistry Undergraduates: the elements of information literacy" (http://units.sla.org/division/dche/il/cheminfolit.pdf). A multi-faceted approach includes primarily hands-on sessions in small laboratory sections, occasional large group lectures, and a comprehensive Web site, http://www.library.ucla.edu/libraries/sel/12451.cfm. The Instructional Resources Web site, available 24/7, features "Library resources for chemistry and biochemistry, chemical and biomolecular engineering, and materials science and engineering." Web pages cover hands-on, guided library exercises including the organic "chemystery," homework assignments, lecture notes, and resources for specific courses, plus general reference sources. Many UCLA sites feature links to "Library Resources" and its pages, including departmental sites, course management systems, the undergraduate Science Learning Center workstations' "Welcome" screen, and individual course pages.

CINF 8 No chemistry librarian: How chemistry faculty can participate in collection development and its increased use

A. W. Kozlowski, kozlowskia@ccsu.edu. Department of Chemistry, Central Connecticut State University, New Britain, CT, United States

Designating an interested chemistry faculty member as a library liaison can benefit both the department and the library. This person can help faculty make the best use of resources that are already available, assist with instruction, and articulate to librarians the highest needs for the curriculum. This approach helps get the most value for the money available.

CINF 9 Beyond Metlin: Extending metabolite databases to enhance identifications in metabolomics

G. Suizdak, siuzdak@scripps.edu. Center for Mass Spectrometry, The Scripps Research Institute, La Jolla, California, United States

Global (untargeted) metabolite profiling using mass spectrometry has a history dating to 1966 (*Biochemical Journal*, **1966**, 792) and was then limited by both the gas chromatography mass spectrometry technology and the tools used to perform post-experimental analysis. These limitations included comparative analysis capabilities and database searching, largely because neither existed. Over the course of the last 40 years GC/MS has evolved into a very mature technology for performing metabolomics experiments, however, the more comprehensive LC/MS experiments are experiencing the same limitations as did its predecessor, namely a lack of differential analysis tools and a very limited number of known metabolites. The METLIN database currently contains ;24,000 compounds and over 10,000 high resolution (ESI-QTOF) MS/MS spectra. Data mining and machine learning techniques have been developed to not only identify known metabolites but also to characterize and classify unknowns. Based on known structures, the fragment assignment to MS/MS data of known compounds, and novel fragmentation prediction software, we have been able to identify the structural characteristics for unknown metabolites. Machine learning methods are used to predict functional groups from MS/MS data and subsequently assign unknown metabolites to compound classes. We are developing a web based application where 1) users can upload MS/MS data to perform an automatic database search to identify metabolites, 2) MS/MS based similarity search can be conducted to show structural similarities for unknown metabolites and 3) functional groups are predicted and unknown metabolites are assigned to compound classes.

CINF 10 Stable isotope-resolved metabolomics analysis of UDP-GlcNAc, UDP-GalNAc: Computational resolution and modeling of their converging biosynthetic pathways

H. N. B. Moseley[1,3], hunter.moseley@louisville.edu, A. C. Belshoff[2,3], R. M. Higashi[1,3], T. W.-M. Fan[1,2,3], and A. N. Lane[2,3]. [1]Department of Chemistry, University of Louisville, Louisville, Kentucky, United States, [2]JG Brown Cancer Center, University of Louisville, Louisville, Kentucky, United States, [3]Center for Regulatory and Environmental Analytical Metabolomics (CREAM), University of Louisville, Louisville, Kentucky, United States

Regulation of uridine diphospho-N-acetylglucosamine (UDP-GlcNAc) and uridine diphospho-N-acetylgalactosamine (UDP-GalNAc) levels underlies the control of many biological processes via O-linked glycosylation of proteins, including nutrient sensing, protein degradation, and gene expression. We used FT-ICR-MS to make rapid, high-confidence measurements of ^{13}C incorporation into both UDP-GlcNAc and UDP-GalNAc from [U- ^{13}C]-glucose, in LNCaP-LN3 prostate cancer cells. This incorporation involves the flow of different ^{13}C-labeled functional moieties via parallel, converging metabolic pathways. We developed a simulated annealing and genetic algorithms (SAGA) method to parse all 17 sets of combined UDP-GlcNAc/UDP-GalNAc mass-equivalent isotopomers (isotopologues) into these functional moieties. This revealed the relative contribution of specific metabolic pathways to ^{13}C incorporation and allowed direct confirmation using NMR positional isotopomer data. The rates of synthesis and utilization of UDP-GlcNAc/UDP-GalNAc were substantially decreased by methyl seleninic acid (MSA), suggesting MSA effects homeostatic regulation of these sugar nucleotides.

CINF 11 Functional and structural network modeling of metabolomics datasets

D. K. Barupal, dinkumar@ucdavis.edu, G. Wohlgemuth, and O. Fiehn. Genome Center, University of California, Davis, Davis, California, United States

Network approaches facilitate metabolic studies. For mass spectrometry based metabolomics datasets, the BinBase database generated lists of known and unknown metabolites detected in biological matrix. We compared two types of network models, functional and structural. Functional network models were less comprehensive than structural models because biochemical databases such as KEGG and MetaCyc lack pathway and reaction annotation for several known metabolites. Only PubChem has entries for all the known metabolites. By applying chemical similarity, mass spectral similarity and atomic mapping of reaction pairs, structural networks were constructed which included all known and unknown metabolites. Cytoscape visualization of these models yielded a modular network graph resembling known metabolic modules. Statistical differences in ER+/- breast cancer comparisons were imposed by color grades to highlight differences in biochemical pathways.

CINF 12 Computational aspects of metabolomic Quantitative Trait Locus (mQTL) mapping and metabolome-wide genome-wide association studies

M.-E. Dumas[1], m.dumas@imperial.ac.uk, J.-B. Cazier[2], J. K. Nicholson[1], and D. Gauguier[2,3]. [1]Department of Surgery and Cancer, Biomolecular medicine, Imperial College London, London, United Kingdom, [2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom, [3]INSERM U872, Centre de Recherche des Cordeliers, Paris, France

The study of human multifactorial diseases, like diabetes, cardiovascular disease, insulin resistance, or complex biological processes such as ageing, represent a major healthcare challenge for the western and developing world. Metabolomic Quantitative Trait Locus (mQTL) mapping consists of the robust and accurate statistical integration of genome-wide genotyping and metabolome-wide profiling, to identify candidate biomarkers and susceptibility genes for disease. Like other QTL mapping approaches, mQTL, relies on the extended Haley-Knott method. To improve the initial mQTL technology, we present a range of computational developments in dataset normalisation, co-factor corrections, automatic peak alignment, statistical recoupling and automatic peak detection and integration, aiming at optimising QTL analysis algorithm efficiency on ^{1}H NMR variables and mQTL discovery rate. We apply this approach to biofluids, aqueous and lipid extracts from both rats and mice, to identify new candidate metabolite biomarkers and loci associated to disease.

CINF 13 Metabolite identification pipeline based on MS fragmentation

M. Rojas-Chertó[1,3], m.rojas@lacdr.leidenuniv.nl, **J. E. Peironcely**[1,2], peironcely@chem.leidenuniv.nl, P. T. Kasper[1,3], A. Bender[4], J.-L. Faulon[5], T. Reijmers[1,3], L. Coulier[2], R. Vreeken[1,3], and T. Hankemeier[1,3]. [1]Department of Netherlands Metabolomics Centre, Leiden University, Leiden, Zuid Holland, The Netherlands, [2]Quality of Life, TNO, Zeist, Utrecht, The Netherlands, [3]Department of Analytical Biosciences, Leiden University, Leiden, Zuid Holland, The Netherlands, [4]Department of Medicinal Chemistry, Leiden University, Leiden, Zuid Holland, The Netherlands, [5]Department of Biology, Evry University, Evry, Île-de-France, France

Here we will show different parts of the metabolite identification pipeline based on MS fragmentation data. **MS^{n}** spectral trees were acquired on the LTQ-Orbitrap (Thermo) equipped with a Triversa nanoMate (Advion) nanoelectrospray ion source. By using in-house developed software, integrating among others the Chemistry Development Kit (CDK) and XCMS libraries, the spectral data was processed. Application of the Multi-stage Elemental Formula (MEF) tool resolved the elemental composition of the parent compound and its fragment ions. A database containing fragmentation trees of metabolite standards is created, enabling the characterization of fragments from an unknown compound by querying and matching subtrees from known compounds. To be able to elucidate the structure of an unknown compound a structure generator is developed that uses as input the elemental formula and fragments. The generated structures These structures are evaluated and ranked with chemoinformatic models predicting 'metabolite likeness'.

CINF 14 Consensus-phenotype integration of metabolomic and transcriptomic data enhances the detection of pathways associated with drug response

R. Cavill[1], A. Kamburov[2], M. S. C. Blagrove[1], T. J. Athersuch[1], J. K. Ellis[1], R. Herwig[2], **T. M. D. Ebbels**[1], t.ebbels@imperial.ac.uk, and H. C. Keun[1]. [1]Biomolecular Medicine, Department of Surgery and Cancer, Imperial College London, London, United Kingdom, [2]Max Planck Institute for Molecular Genetics, Berlin, Germany

The successful combination of different types of biological data is a critical challenge in systems biology. In this work, we aim to define pathways predictive of drug sensitivity using publically available drug response, transcriptomic and metabolomic measurements from the NCI60 cell line panel, together with a novel approach to data integration. Initially focusing on platinum-based chemotherapeutics we show that the sets of pathways associated with sensitivity to four platinum drugs are highly coincident. Many significantly altered pathways, covering wide range of metabolic and signaling processes, were consistent with the previously reported molecular determinants of platinum response. The sensitivity for the detection of predictive pathways was improved by □ ;85% by combining "-omics" data at the point of pathway analysis. Extending the approach across a wide panel of compounds, we show how consensus-phenotype integration can help to reveal the complex relationships between biological pathways and drug response.

CINF 15 Designing a cyberinfrastructure for metabolomics

M. Arita, arita@k.u-tokyo.ac.jp. Department of Computational Biology, The University of Tokyo, Kashiwa, Chiba, Japan

Most Wiki-based databases do not provide a measure to control vocabulary (ontology) of, or extract statistical information from input data. This is the often overlooked side of community-driven projects: terminology in a wiki environment tends to be heterogeneous and it will be extremely hard to 'manage' data. To overcome this disadvantage, we present an extension to the MediaWiki system, the most popular wiki-based platform. As a proof of principle, a collection of more than 7000 flavonoid structures, over 1700 plant species information, and the information for basic metabolism are linked together using latin names and structural IDs as database keys. In this talk, I will introduce functionality of this approach and describe the difference from conventional systems. The system is available at http://metabolomics.jp/. This mechanism will be supported by several academic societies in Japan to integrate their academic information into a single portal site.

CINF 16 Computational discovery of new zeolite-like materials

M. W. Deem, mwdeem@rice.edu, and D. J. Earl, dearl@pitt.edu. Bioengineering and Physics Astronomy, Rice University, Houston, TX, United States

We present a database of computationally predicted zeolite-like materials. The materials were identified by a Monte Carlo search of Si atom positions as the number of unique atoms, density, space group, and unit cell of the crystalline material was systematically explored. Over 2.7M unique structures were identified, with roughly 10% within the +30 kJ/mol,Si energetic band above alpha-quartz in which the known zeolites lie. Predicted structures within this band have geometric and topological characteristics similar to that of the known zeolites. Known zeolites are shown to lie on the low-density edge of the distribution of predicted structures. Dielectric constants and X-ray powder diffraction patterns are calculated. Strategies for chemical synthesis of these materials are discussed, a low-density subset of the materials is identified as particularly interesting, and the complementarity of these materials to high-throughput methods is discussed. These structures have been deposited in two publicly available databases.

CINF 17 Using multiple levels of computer modeling to accelerate development of metal organic frameworks materials for membrane applications

D. S. Sholl, david.sholl@chbe.gatech.edu, S. Keskin, T. Watanabe, E. Haldoupis, and S. Nair. School of Chemical Biomolecular Engineering, Georgia Institute of Technology, United States

Metal-organic frameworks (MOFs) are crystalline nanoporous materials with a variety of potential applications in gas separations and storage. Because thousands of distinct MOFs are known, approaches that can rapidly identify structures with particular promise for specific applications have considerable value. We will describe a hierarchical modeling approach to screen MOFs for membrane-based gas separations, an application that is challenging because it requires information on both adsorption affinity and molecular diffusivities. Our approach uses geometric characterization to screen hundreds of materials, then molecular modeling and quantum chemistry calculations to make detailed predictions about strong candidate materials. We will show examples of using this approach to develop materials for efficient separation of methane and carbon dioxide. We will also describe our progress in fabricating and testing MOF-based membranes based on materials identified via modeling.

CINF 18 Informatics approaches for analyzing zeolites

E. Blaisten-Barojas, blaisten@gmu.edu. Computational Materials Science Center, George Mason University, Fairfax, Virginia, United States

Zeolites are natural and synthetic crystalline materials with interesting microporous framework structures. The framework type of zeolites is traditionally determined from information of coordination sequences, vertex symbols, tiling and transitivity. We developed the Framework-Type-Predictor (FTP), a novel informatics-based model that is very efficient for classifying zeolites according to their framework type. The FTP is built on topological, physical, and chemical features, is able to precisely classify into 41 framework types, achieves accuracies of up to 100%, and is superior for zeolites with crystalline distortions. Additionally, we present an unsupervised clustering model for grouping zeolites into four attractor-classes. Trained with 1300 zeolites from the Inorganic Crystal Structure Database, with an optimal 9-dimensional feature vector, this model yields classification into four attractors with 97% accuracy. Such attractors should be useful for material designers producing computationally

generated zeolitic compounds. *(Supported by National Science Foundation CHE-0626111, ICSD data courtesy of NIST).*

CINF 19 Screening tools for identification of porous materials for CO_2 separation

M. Haranczyk[1], mharanczyk@lbl.gov, J. Kloke[2], K. Theisen[3], B. Liu[3], G. Carlsson[2], and B. Smit[3]. [1]Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States, [2]Department of Mathematics, Stanford University, Palo Alto, CA, United States, [3]Department of Chemical Engineering, University of California, Berkeley, Berkeley, CA, United States

Porous materials, e.g. zeolites, have many applications in the chemical industry. They can also be used for gas separations. The number of possible zeolite structures has been estimated to be larger than 2.5 millions. Databases of hypothetical zeolite structures are being developed and they could in principle be screened for best zeolites for the task of CO_2 capture. The current state-of-the-art molecular simulations allow for accurate prediction of zeolite properties but the computational cost of such calculations prohibits their application in the characterization of the entire database of hypothetical structures, which would be required to perform brute-force creening for novel structures with useful properties. Our work focuses on the development of an efficient screening technique that requires such expensive characterization only for carefully selected and statistically relevant subset of a database. Then, the database is screened employing the similarity principle. The developed screening technique, structural descriptors and similarity measures as well as first results will be presented.

This work is supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

CINF 20 3D-Characterization and modeling of composite materials

O. Wirjadi, wirjadi@itwm.fraunhofer.de. Image Processing Group, Fraunhofer ITWM, Kaiserslautern, Germany

Composites form an important class of functional materials in many industries. Examples include open metal foams used as heat exchangers and carbon fiber-reinforced polymers to reduced the weight of fuselages in aircraft. An important factor that determines these materials' macroscopic properties is their geometry at the micro- and meso-scale. Combining tomographic imaging with image analysis allows to describe their 3D-geometry. We present 3D-image analysis algorithms and stochastic models for automatic characterization and modeling of large batches of specimens. In a first step, we characterize the microstructure in terms of mean values and empirical distributions (e.g. strut lengths and cell geometry for foams, fiber densities and orientation distributions for fiber-reinforced polymers). From these measurements, a second step is to model these materials. Such models allow for "virtual design" of composites, i.e., when combined with finite element-simulation of physical or mechanical properties, geometries with desirable properties can be identified.

CINF 21 Growing focus on green chemistry: What the CAS databases reveal

R. Schenck, rschenck@cas.org. Chemical Abstracts Service, Columbus, OHIO, United States

Green chemistry is a trend that has spiked during recent years. This presentation will focus on specific areas of interest and the nature of the research going on in those areas, as seen from the perspective of the CAS databases. For instance, the fermentation process that yields alcohol from corn is well understood, but the CAS databases contain a wealth of articles on catalysts for optimizing this process in addition to research on the genetic engineering of corn, or the biomass, so that it may be more efficiently converted to ethanol. Features in SciFinder for identifying green reactions will be illustrated, as well as content in the CAS databases that support REACH reporting requirements.

CINF 22 California's green chemistry initiative: Application of life cycle assessment in public policy

B. Boughton, bboughto@dtsc.ca.gov. Department of Toxic Substances Control, California EPA, Sacramento, California, United States

The California Green Chemistry initiative comes at a time of growing concern that the federal Toxic Substances Control Act, passed three decades ago, has failed to control an explosion of hazardous materials. California law passed last fall has jump-started the program. AB 1879 requires the state to identify "chemicals of concern" and to evaluate safer alternatives by implementing a policy option, "Accelerate the Quest for Safer Products, creating a systematic, science-based process to evaluate chemicals of concern and alternatives to ensure product safety and reduce or eliminate the need for chemical-by-chemical bans". The adopted rules will lay out the process for products with chemicals of concern to be compared to

alternatives by assessing the "environmental and economic footprint" of each using a life-cycle perspective. This application of life-cycle thinking for assessing environmental impacts and for assessing total costs will be described. The background information on the initiative is at www.dtsc.ca.gov

CINF 23 ChemSpider: How an online resource of chemical compounds, reaction syntheses, and property data can support green chemistry

A. J. Williams, antony.williams@chemspider.com, and V. Tkachenko. ChemSpider, Royal Society of Chemistry, Wake Forest, NC, United States

ChemSpider is an online database containing in excess of 20 million chemical compounds and associated experimental and predicted physicochemical data, reaction synthesis details and analytical data. A significant amount of the data contained within the database has been harvested and collated from a number of inventory systems and integrated to provide a centralized resource for the community. The ChemSpider database has the added benefit of being available for community deposition, annotation and curation. As a result it offers the potential for researchers to share their latest research with the public and participate in the creation of a rich resource of chemistry related information for the Green Chemistry community. This presentation will provide an overview of present capabilities and discuss the future vision for the platform.

CINF 24 From biochemical pathways to retrobiosynthesis

C. H. Schwab, schwab@molecular-networks.com, O. Sacher, sacher@molecular-networks.com, B. Bienfait, bienfait@molecular-networks.com, and J. Gasteiger, gasteiger@molecular-networks.com. Molecular Networks GmbH, Erlangen, Bavaria, Germany

Intracellular transformations as outlined by the Biochemical Pathways wall charts form a high-dimensional space that interconnects many biochemical compounds by a multitude of enzyme-catalyzed reactions. In order to explore (and exploit) this vast information to address challenges in the areas of metabolomics, drug discovery, modeling of biological systems or green chemistry and metabolic engineering, the wall charts (plus recent literature data) have been converted into the database BioPath of cross-linked chemical structures, reactions and pathways [1]. The paper introduces the BioPath database and its application in the program system BioPath.Design that supports the planning of biotechnological manufacturing of bulk and fine chemicals. Starting with the full target structure, BioPath.Design systematically searches in the BioPath database and suggests new biochemical reaction steps and synthesis routes based on similar known enzyme-catalyzed reactions, even if the target is not present in the database. The method can be used either in a step-wise, interactive or multi-step, fully automatic retrosynthetic manner. Some examples will be demonstrated.

[1] Reitz, M.; Sacher, O.; Tarkhov, A.; Trümbach, D.; Gasteiger, J. Enabling the exploration of biochemical pathways. Org. Biomol. Chem. 2004, 2, 3226-3237.

CINF 25 Green chemistry database and analytical compliance suite (GC-DACS)

G. R. Thompson, georgethompson@chemply.com. Chemical Compliance Systems, Inc., Lake Hopatcong, NJ, United States

Application of "the twelve principles of green chemistry" to complex formulations requires comparison of both efficacy and toxicity data (i.e., standardized criteria) for multiple chemical constituents and/or their alternatives. Thus, "reducing toxicity" or identifying "hazardous substances" must encompass broad ecological, health, and safety measures. How can these parameters be integrated into a quantitative set of "green" criteria for application to chemicals, processes, products, and waste across the entire product lifecycle (i.e., cradle to cradle)? In 2003, Chemical Compliance Systems, Inc. (CCS) developed four web-based, analytical compliance modules for their GC-DACS to help improve product design, manufacturing process comparisons, and "greening" the waste stream while assuring compliance with the national Standard for Green Chemical Products and Processes. Ultimately, these analytical tools enhance the green principles by providing a broad definition and a depth of quantification not previously available in other web-based capabilities.

CINF 26 Importance of benchmarking Green Chemistry

G. Gurau, and **R. D. Rogers**, rdrogers@as.ua.edu. Department of Chemistry, Center for Green Manufacturing, The University of Alabama, Tuscaloosa, AL, United States

Readily searchable and accessible data with which to benchmark 'improvements' to chemical or chemical technology are critical to the field of Green Chemistry. Green Chemistry strives to develop chemicals and chemical processes which are non toxic, not harmful to the environment, and sustainable. It is however, often difficult to determine whether improvements in one aspect of a process (e.g., elimination of volatile solvents), might actually be more harmful in another (e.g., greater energy use). Green Chemistry is really about careful decision making in order to balance many, often competing goals. Here we will discuss how extracting benchmarking information from the literature can help in these decision making processes.

CINF 27 Software advancements for performing compound identification QC on large metabolomic datasets

C. DeHaven, lgosselin@metabolon.com. Department of Metabolomics, Metabolon, Research Triangle Park, NC, United States

Traditional collection, sorting and analysis of metabolomic data has generally involved single data files each corresponding to a single biological sample. Data is then individually compared to spectral libraries of known metabolites in order to identify compounds contained in each biological sample. This labor-intensive approach does not lend itself to industrialization of the process; it also presents a problem correlating the analysis of multiple biological samples, including meta-data, with one another to determine trends and population differences. The ability to analyze and perform QC on large-scale, multi-sample metabolomics data on an industrial scale is an important step in the evolution of metabolomics technology as a whole. This presentation will provide an overview of software advancements which provide options for quickly performing rapid quality-control of metabolomic data.

CINF 28 Fold-change analysis and visualization of multispectral datasets in NMR-based metabolomics

S. L. Robinette, slrobin@ufl.edu, and A. S. Edison. Biochemistry and Molecular Biology, University of Florida, Gainesville, FL, United States

Large dataset visualization is a critical component in the analysis and interpretation of metabolomics data. Nuclear magnetic resonance offers particular challenges in terms of visualization as datasets are generally composed of multiple highly complex spectra with hundreds of overlapping signals arising from many small molecules. Here, we explore methods of fold-change based visualization using one- and two-dimensional NMR spectra in order to identify and interpret patterns of differential expression of metabolites. We show that cluster analysis of peak fold-change matrices identifies both structurally related peaks and metabolic coregulation using statistical relationships between signals in 1D ^{1}H NMR spectra and that fold-change analysis of individually aligned two-dimensional spectra provides complementary information. We demonstrate the utility of this methodology using tissue extract and media samples relevant to metabolomics.

1. Robinette, S.L.; Veselkov, K.A.; Bohus, E.; Coen, M.; Keun, H.C.; Ebbels, T.M.D.; Beckonert, O.; Holmes, E.C.; Lindon, J.C.; Nicholson, J.K. Anal. Chem. 2009, 81, 6581-6589.
2. Schroeder, F.C.; Gibson, D.M.; Churchill, A.C.; Sojikul, P.; Wursthorn, E.J.; Krasnoff, S.B.; Clardy, J. Angew. Chem. Int. Ed. 2007, 46, 901-904.

CINF 29 Data mining tool for automated metabolite identification and quantification using *J*-resolved NMR spectroscopy

S. He[1], C. Ludwig[2], J. M. Easton[3], H. Chen[1], S. Tiziani[2], A. Lodi[2], S. Manzoor[4], A. D. Southam[4], T. N. Arvanitis[3], U. L. Guenther[2], and **M. R. Viant**[4], m.viant@bham.ac.uk. [1]School of Computer Science, University of Birmingham, Birmingham, West Midlands, United Kingdom, [2]CR UK Institute for Cancer Studies, Henry Wellcome Building for Biomolecular NMR Spectroscopy (HWB-NMR), University of Birmingham, Birmingham, West Midlands, United Kingdom, [3]School of Engineering, University of Birmingham, Birmingham, West Midlands, United Kingdom, [4]School of Biosciences, University of Birmingham, Birmingham, West Midlands, United Kingdom

Although one-dimensional (1D) nuclear magnetic resonance (NMR) spectroscopy remains one of the leading analytical technologies in metabolomics, it suffers from severe spectral overlap which limits its ability to identify and quantify metabolites. 2D *J*-resolved (JRES) NMR spectroscopy is rapidly gaining in popularity, and benefits from a dispersion of peaks into a second dimension, improving metabolite specificity and potentially the accuracy of

quantification. Here we report a freely-available data mining software package for quantitative metabolic profiling based on 2D JRES NMR spectroscopy. It incorporates a novel 2-step screening/quantification algorithm along with a 500 MHz NMR library containing ca. 220 JRES spectra of pure metabolites. The NMR library is housed in a MySQL database, with web accessibility, and is compliant with the Metabolomics Standards Initiative endorsed data standards. Validation of the data mining tool will be demonstrated using simulated as well as real metabolomics datasets comprising of chemically-defined metabolite mixtures.

CINF 30 Identification of parent-fragment pairs via rigorous statistical modeling of LC-MS metabolomic data

A. Ipsen, andreas.ipsen06@imperial.ac.uk, E. Want, J. Lindon, and T. Ebbels. Biomolecular Medicine, Department of Surgery. Cancer, Imperial College London, United Kingdom

The untargeted analysis of metabolic samples by Liquid Chromatography-Mass Spectrometry often generates a large number of signals that are due to unknown compounds and whose identification forms an important challenge in many metabolomic studies. The analysis of a metabolite's fragmentation pattern following collision induced dissociation provides an important tool for its identification, but one that can be severely impeded by the close chromatographic coelution of distinct metabolites. We propose and validate a statistically rigorous algorithm for identifying clusters of related fragments and for distinguishing these from unrelated compounds. Unlike current techniques the algorithm presented here is derived directly from the statistical distribution of ion arrivals and moreover provides a rigorous probabilistic measure of the uncertainty of the assignment. We argue that many of the central problems that are encountered in the analysis of LC-MS data could potentially be addressed by means of similar 'ab initio' methods.

CINF 31 Generation of in-silico MS/MS mass spectra using combinatorial algorithms and reaction prediction expert systems

T. Kind, tkind@ucdavis.edu, K.-H. Liu, D. Y. Lee, and O. Fiehn. Metabolomics, UC Davis Genome Center, Davis, CA, United States

Metabolomic profiling using mass spectrometry suffers from the large number of unidentified compounds and small mass spectral libraries. For certain substance classes like lipids and oligosaccharides that are built from consisting building blocks it is possible to generate fragmentation patterns and to model peak abundances. We discuss methods for mass spectrometric fragmentation prediction. Using combinatorial library enumeration tools like ModLab SMILIB and Markush enumeration tools like ChemAxon Reactor we developed in-silico compound libraries for lipid classes which are not yet found in CAS, PubChem or ChemSpider databases but were already analyzed using mass spectrometric techniques. We used HighChem Mass Frontier and own fragmentation and abundance modeling software to simulate accurate mass tandem mass spectra. We show successful compound identifications using experimental tandem mass spectra from different types of mass analyzers including ion traps, quadrupoles, FT-MS, time-of-flight as well as hybrid ion mobility analyzers and discuss further applications in workflows.

CINF 32 Using aggregative Web Services for drug discovery

Q. Zhu[1], qianzhu@indiana.edu, M. S. Lajiness[2], and D. J. Wild[1]. [1]school of informatics and computing, Indiana Unversity, Bloomington, In, United States, [2]Eli Lilly and Company, Indianapolis, In, United States

Recent years have seen a huge increase in the amount of publicly-available information pertinent to drug discovery, including online databases of compound and bioassay information; scholarly publications linking compounds with genes, targets and diseases; and predictive models that can suggest new links between compounds, genes, targets and diseases. However, there is a distinct lack of data mining tools available to harness this information, and in particular to look for information across multiple sources. At Indiana University we are developing an aggregative web service framework to solve this kind of problems. It offers a new approach to data mining that crosses information source types to look at the "big picture" and to identify corroborating or conflicting information from models, assays, databases and publications.

CINF 33 Semantifying polymer science using ontologies

E. O. Cannon, eoc21@cam.ac.uk, N. Adams, and P. Murray-Rust. Department of Chemistry, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom

Ontologies are graph based, formal representations of information in a domain. Currently, there is a large interest in ontologies for biology and medicine, though little effort has been concentrated in the field of

chemistry, let alone polymer science. We have developed a number of ontologies for polymer science: properties, measurement techniques and measurement conditions, using the Web Ontology Language. These ontologies will help facilitate the standardization of data exchange formats in polymer science by providing a common domain of knowledge. The properties ontology contains over 150 properties and has been integrated with the measurement techniques and conditions ontology, to give information on how a property is measured and under what conditions. The ontologies will be of use to polymer scientists wishing to reach a consensus in this area of knowledge. The ontologies also have the advantage that they can be integrated into software applications to leverage this knowledge.

CINF 34 PDZ domain: A computational insight into protein similarity and ligand specificity

C. S. D. Sisu, csds2@cam.ac.uk, and G. H. Grant. Unilever Centre for Molecular Science Informatics, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom

PDZ domain is one of the most common protein interaction modules. We explored its peculiar properties from two perspectives: protein similarity and ligand specificity. Firstly, the protein similarity was studied using reduced dimensionality representations. A two dimensional map was generated for every protein using a Monte Carlo like technique. The similarity between proteins was assessed based on the similarity of their respective maps. The proteins were compared at the level of amino acid side chain conformation and binding induced dynamics. Based on our findings we propose a novel classification of PDZ domains. Secondly, the promiscuous binding properties of PDZ domains were investigated using molecular dynamics simulations. MMPBSA as implemented in Amber 10 was used to calculate the thermodynamic parameters of protein-peptide interactions. We show that PDZ domain proteins capture the subtle differences between potential binding partners dynamically. Even more there is a direct correlation between binding affinity and protein dynamics.

CINF 35 Homology modeling and molecular dynamics of the APJ receptor, a co-receptor for cellular entry of HIV

N. J. M. Macaluso, max.macaluso2@gmail.com, and R. C. Glen, rcg28@cam.ac.uk. Unilever Centre for Molecular Sciences Informatics, Department of Chemistry, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom

The APJ receptor is a class A GPCR and binds the endogenous ligand apelin-13 (NH_2-QRPRLSHKGPMPF-COOH). This signaling pathway is linked to cardiovascular disease, tumor neoangiogenesis, and HIV infection. The absence of a crystal structure has hampered the identification of a binding site and subsequent design of small molecule antagonists. We have constructed the first homology model of APJ using the backbone coordinates of the recent □2-adrenergic receptor crystal structure. This model was initially refined through extensive molecular dynamics in a solvated POPC lipid bilayer and further simulations were performed with the receptor in complex with the apelin peptide. Sequence information and structure-activity data derived from synthesis and binding affinity evaluation of novel cyclic peptides guided our initial placement of apelin. The resulting simulation is an important step in elucidating the binding mode of apelin.

CINF 36 Assigning stereochemistry using GIAO NMR shift calculation

S. G. Smith, sgs30@cam.ac.uk, and J. M. Goodman. Department of Chemistry, University of Cambridge, Cambridge, United Kingdom

GIAO NMR shift calculation has been applied to the challenging task of assigning stereochemistry to complex molecules such as natural products. Two new approaches, CP3 and DP4, have been developed for quantifying the agreement between calculated and experimental shifts in two different situations. These approaches are significantly more successful at making correct assignments with high confidence than are those based on the correlation coefficient and mean absolute error.

CINF 37 Combination of target-specific pose filters and force field-based scoring functions to improve the structure-based virtual screening

J.-H. Hsieh[1], juihua_hsieh@unc.edu, S. Yin[1], syin@email.unc.edu, X. S. Wang[1], xswang@email.unc.edu, S. Liu[2], shubin@email.unc.edu, N. V. Dokholyan[1], dokh@med.unc.edu, and A. Tropsha[1], alex_tropsha@unc.edu. [1]University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [2]Renaissance Computing Institute at UNC-CH, United States

Poor performance of scoring functions is currently a bottleneck in structure-based virtual screening (VS). Knowledge-based scoring functions capture subtle physical interactions that are otherwise difficult to model explicitly while force field-based scoring functions are less dependent on training sets. Thus these two approaches are intrinsically complementary to each other. We have combined a cheminformatics-based, target-specific pose filters with a force field-based scoring function, MedusaScore, to score presumed ligand poses during VS. Our results demonstrate that the novel pose filter improves the hit recovery in VS, especially for top-ranked ligands, for 5 out of 6 benchmarking sets retrieved from the Database of Useful Decoys (DUD). The further analysis shows that the improvements result from the elimination of decoy poses with high MedusaScore so these decoys are not ranked. Moreover, the top 10 hits include chemically diverse ligands, which is a highly desired outcome of any biological screening effort.

CINF 38 Toxicity reference database (ToxRefDB) to develop predictive toxicity models and prioritize compounds for future toxicity testing

H. Tang[1,2], tangh@email.unc.edu, H. Zhu[1], L. Zhang[1], A. Sedykh[1], A. Richard[3], I. Rusyn[4], and A. Tropsha[1]. [1]Division of Medicinal Chemistry and Natural Products, School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [2]Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [3]National Center for Computational Toxicology, Office of Research Development, U.S. Environmental Protection Agency, Chapel Hill, NC, United States, [4]Department of Environmental Sciences and Engineering, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

EPA's ToxCast program aims to use in vitro assays to predict chemical hazards and prioritize chemicals for toxicity testing. We employed the predictive QSAR workflow to develop computational toxicity models for ToxCast compounds with historical animal testing results available from ToxRefDB. To ensure model stability and robustness, multiple classifiers and 5-fold external cross-validation were applied. Results show that for three of the 78 toxicity endpoints, including one chronic and two reproductive endpoints, the Correct Classification Rate for external validation datasets was above 0.6 for all types of QSAR models. Our studies suggest that it is feasible to develop QSAR models for some endpoints, which could be further augmented by in vitro assay measures. The validated toxicity models were used for virtual screening of 50,000 chemicals compiled for the REACH program. The compounds predicted as toxic could be regarded as candidates for future toxicity testing. Abstract does not reflect EPA policy.

CINF 39 Chem2bio2RDF: Semantic systems chemical biology

B. Chen[1], binchen@indiana.edu, X. Dong[1], D. Jiao[1], H. Wang[1], Y. Ding[2], and D. J. Wild[1]. [1]School of Informatics and Computing, Indiana University, Bloomington, IN, United States, [2]School of Liberary and Information Science, Indiana University, Bloomington, IN, United States

Recent advances in high throughput techniques, predictive modeling, and data generation have contributed to the emergence of new chemical biological disciplines. In particular, systems biology adopts a comprehensive approach to the study of biological systems; chemogenomics studies the impact of small molecules towards biological systems and carries datum description about interaction among chemical entities and protein molecules; and the integration between cheminformatics and bioinformatics within the realm of systems biology leads to a new related discipline of systems chemical biology. These disciplines all involve the processing and data mining of very large amounts of information that come from diverse sources and disciplines, and aggregating and integrating these information sources can be very challenging. The work reported here addresses this problem using Semantic Web technology. Our focus is on the aggregation of public chemogenomic data distributed around the Web into RDF formats, which enables linking with other biological Semantic Web information resources such as Bio2RDF and LODD. We examine the integrated use of this information in addressing problems in polypharmacology and adverse drug reaction using a variety of cheminformatics and bioinformatics techniques.

CINF 40 Application of novel data mining techniques to improve chemical array design decisions during lead optimisation

G. Papadatos[1], g.papadatos@shef.ac.uk, M. Alkarouri[1], V. Gillet[1], P. Willett[1], V. Kadirkamanathan[2], C. Luscombe[3], G. Bravi[3], N. Richmond[3], S. Pickett[3], J. Pritchard[3], A. Cooper[3], and S. Macdonald[3]. [1]Department of Information Studies, University of Sheffield, Sheffield, United Kingdom, [2]Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, United Kingdom, [3]GlaxoSmithKline Medicines Research Centre, Stevenage, United Kingdom

Matched molecular pair analysis (MMPA) studies the effect of specific structural modifications on properties of interest. In modern drug discovery this analytical tool becomes very important, especially during the lead optimisation phase, when medicinal chemists synthesise and test iteratively hundreds of compounds in chemical arrays. In this study, we have developed novel data mining algorithms to assess the extent to which MMPA is valid and informative. To this end, a set of both established and bespoke 2D and 3D descriptors and multiple assay data from several GlaxoSmithKline lead optimisation projects have been used. We have investigated a number of important drug properties such as hERG inhibition, solubility, lipophilicity as well as biological activity. The outcomes are useful as part of a decision support system which enables the medicinal chemists to mine relevant information and to explore systematically the chemical space in the vicinity of the seed structure in each round of chemical array design.

CINF 41 Biological and chemical network construction from linked open data

D. Jiao, djiao@indiana.edu. School of Informatics, Indiana University at Bloomington, Bloomington, IN, United States

Linked open data (LOD) is a joint effort in the Semantic Web community to provide linkages among open data. LOD data sets in the life science domain, such as Bio2RDF and Linked Open Drug Data (LODD), provide a great resource for systematic study of life science. However, the sheer size and heterogeneity of the LOD data sets is daunting, suggesting that studies based on these data sets could benefit substantially from methods that facilitate the retrieval of targeted information. In this poster, we propose a workflow-based method for constructing biological and chemical networks from the underlying RDF graphs of Linked Open Data, using semantic web technologies and scientific workflow tools. It uses implicit SPARQL queries and generic rule-based reasoning for retrieval and construction of the networks. Using this method, biological and chemical networks, such as protein-proteininteraction networks, drug-target networks, or metabolic networks, could be retrieved from Bio2RDF or LODD.

CINF 42 Novel approach to drug discovery integrating chemogenomics and QSAR modeling: Applications to anti-Alzheimer's agents

R. Hajjo[1], hajjo@email.unc.edu, S. Wang[1], B. L. Roth[2], and A. Tropsha[1]. [1]Department of Medicinal Chemistry and Natural Products, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [2]Department of Pharmacology, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Chemogenomics is an emerging interdisciplinary field relating the receptorome-wide biological screening to functional or clinical effects of chemicals. We have developed a novel chemogenomics approach combining QSAR modeling, virtual screening (VS), and gene expression profiling for drug discovery. Gene signatures for the Alzheimer's disease (AD) were used to query the Connectivity Map (cmap,http://www.broad.mit.edu/cmap/) to identify potential anti-AD agents. Concurrently, QSAR models were developed for the serotonin, dopamine, muscarinic and sigma receptor families implicated in the AD. The models were used for VS of the World Drug Index database to identify putative ligands. 12 common hits from QSAR/VS and cmap studies were subjected to parallel binding assays against a panel of GPCRs. All compounds were found to bind to at least one receptor with binding affinities between 1.7 - 9000 nM. Thus, our approach afforded novel experimentally confirmed GPCR ligands that may be implied as putative treatments for the AD.

CINF 43 Use of QSAR models to validate the HTS assay data in PubChem: Application to the serotonin receptor 5-HT1A ligands

M. Luo, mluo@email.unc.edu, S. X. Wang, xswang@email.unc.edu, A. Golbraikh, golbraik@email.unc.edu, and A. Tropsha, alex_tropsha@unc.edu. Department of Medicinal Chemistry, University of North Carolina, Chapel Hill, Chapel Hill, NC, United States

PubChem is an open-access data repository system, acting as the portal site for NIH Molecular Libraries and Imaging initiative. To help evaluate the quality of some biological activities deposited in PubChem, we

conducted *in silico* modeling studies on 5-Hydroxytryptamine Receptor Subtype 1A (5-HT1A) ligands (PubChem AID:613, 718, 755). The dataset of 105 binders/61 non-binders were retrieved from the NIMH Psychoactive Drug Screening Program Ki database, and three methods, *k*-Nearest Neighbor, Random Forest and Support Vector Machines, were employed for model building using Dragon descriptors. Models were further validated by predicting 69 additional 5-HT1A ligands from the WOMBAT database with Correct Classification Rate of 0.94. However, 25 out of the 46 agonists/antagonists in PubChem bioassays were predicted to be non-binders with high confidence, suggesting that they may be false positives. We propose that rigorously developed QSAR models could be used to identify both false positives and false negatives reported in PubChem.

CINF 44 Speeding up chemical database searches

R. J. Nasr, rnasr@ics.uci.edu, and P. Baldi, pfbaldi@ics.uci.edu. Department of Computer Science, University of California, Irvine, Irvine, CA, United States

In many large chemoinformatics database systems, molecules are represented by long binary fingerprint vectors whose components record the presence or absence of particular functional groups or combinatorial features. To speedup database searches, we develop and validate two new approaches. In the first approach, we compute for each fingerprint a corresponding small count vector containing the sums of hashed subsets of components. From these count vectors, we derive tight bounds on fingerprint similarity measures, such as the Tanimoto measure. During a database search, every time these bounds are unfavorable, the corresponding molecule can be discarded with no need for further inspection. The second approach adapts techniques from text search based on inverted indices to chemical fingerprints. For both approaches we construct probabilistic models that allow us to estimate the level of pruning under different conditions. These theoretical results are corroborated by experimental results obtained on a large set of molecules.

CINF 45 OrbDB: A database of molecular orbital interactions

M. A. Kayala, mkayala@ics.uci.edu, C. A. Azencott, J. H. Chen, and P. F. Baldi. Department of Computer Science, University of California - Irvine, Irvine, CA, United States

The ability to anticipate the course of a reaction is essential to the practice of chemistry. This aptitude relies on the understanding of elementary mechanistic steps, which can be described as the interaction of filled and unfilled molecular orbitals. Here, we create a database of mechanistic steps from previous work on a rule-based expert system (ReactionExplorer). We derive 21,000 priority ordered favorable elementary steps for 7800 distinct reactants or intermediates. All other filled to unfilled molecular orbital interactions yield 106 million unfavorable elementary steps. To predict the course of reactions, one must recover the relative priority of these elementary steps. Initial cross-validated results for a neural network on several stratified samples indicate we are able to retrieve this ordering with a precision of 98.9%. The quality of our database makes it an invaluable resource for the prediction of elementary reactions, and therefore of full chemical processes.

CINF 46 PropOrb: A frontier molecular orbital interaction proposer

C.-A. Azencott, cazencot@ics.uci.edu, M. A. Kayala, mkayala@ics.uci.edu, and P. Baldi, pfbaldi@ics.uci.edu. Department of Information and Computer Science, UC Irvine, Irvine, CA, United States

Chemoinformatics methods to assist chemists in the essential problem of reaction prediction are highly desirable. Brute force examination of all potentially interacting molecular orbitals (MOs) is computationally intractable. For example, concurrent work to develop a database of MO interactions yields more than 100 million interactions for a few thousand systems. Inspired by human chemists, we leverage the principles of frontier orbital theory to propose a tool that rapidly identifies the most reactive molecular orbitals in a given system. Due to the simplicity of our approach, we aim to identify only a few top candidates. Further resource intensive refinement, such as quantum mechanical or statistical approaches, might then be necessary to differentiate between the proposer's suggestions. However, our approach significantly reduces the number of possibilities for downstream analysis. The speed, simplicity, and efficacy of our MO interaction proposer make it an essential component of a reaction prediction system.

CINF 47 Dataviz explosion: Considering effectiveness in classrooms and across disciplines

E. Dorland, dorland@wustl.edu. PARC Communications, Washington University, St Louis, MO, United States

Data visualization is exploding. Newspapers and businesses use complex visualizations in stories and reports. Visualizations are valued for their beauty as well as clarity, and the best "go viral" in Facebook, Twitter, and the blogosphere. The popular website "Visual Complexity" has hundreds of examples. But how effective are visualizations? What types are best in which situations? How do we know? The interdisciplinary community of the Gordon Research Conference on Visualization in Science and Education focuses on how students learn from visualizations. When is 3D or video best? When is a simple diagram better than a more complex one? Is 3D always better than 2D? I will discuss findings from collaborative research among scientists, mathematicians, engineers, cognitive psychologists, graphic designers, web and video developers, and the learning sciences. Examples will come from undergraduate science topics and will include the presentation and display of data in 3D Virtual Worlds.

CINF 48 ASAP: Emphasizing multidimensional drug discovery

W. P. Walters, pat_walters@vrtx.com, Y. Bennani, and T. Kramer. Vertex Pharmaceuticals, Cambridge, MA, United States

One of the biggest challenges facing drug discovery teams is extracting information from the large volume of data generated in the course of a lead optimization effort, and using this information to make decisions. A typical lead optimization project can track between 15 and 30 assays (enzyme, cell, properties, PK, etc), and identifying trends in this data can be difficult. In an effort to address this problem and help scientists to make better decisions, we have developed a new software platform called ASAP. ASAP provides an intuitive overview of the data that also allows scientists to easily "drill down" and examine the details of particular experiments. A combination of "filters" and heat maps allows teams to focus on aspects of the data while remaining aware of the "big picture".

CINF 49 Visual analyses for guiding compound selection and design

E. Champness, ed.champness@optibrium.com. Optibrium Ltd, Cambridge, United Kingdom

Decision-making in drug discovery is challenging due to the need to simultaneously optimise multiple, often conflicting properties for which the available data have varying degrees of confidence. In this presentation we will present a workflow, supported by appropriate visualisations, to guide decisions on selection and design of compounds in the context of a drug discovery project. This includes approaches to explicitly consider the ability to confidently distinguish between molecules, visualise the trade off between quality and diversity and interpret the relationship between structure and properties for individual molecules. We will demonstrate the integration of these into a dynamic interface that can be intuitively used by discovery scientists of all disciplines. This can significantly reduce the time to factor all information into the decision-making process as a drug discovery project progresses and the volume of data increases.

CINF 50 Contextual visualization of biological and chemical networks using linked open life science data

D. Jiao[1], djiao@indiana.edu, Y. Ding[2], dingying@indiana.edu, and D. J. Wild[1], djwild@indiana.edu. [1]School of Informatics, Indiana University, Bloomington, IN, United States, [2]School of Library and Information Science, Indiana University, Bloomington, IN, United States

Semantic web technologies, such as ontologies and RDF (Resource Description Framework), have played important roles in life science data representation, interchange, and integration. Recently several groups have built RDF networks to integrate open data in life science. The results, which are part of the open linked data, provide great resources for research in life science. For example, Bio2RDF and LODD (Linked Open Drug Data) could be used in the area of systems chemical biology, since they contain structured information in almost every aspect that systems chemical biology requires. However, the sheer size and heterogeneity of the open linked data is daunting, suggesting that studies based on these data could benefit substantially from targeted technologies of visualization and information retrieval. In this talk we will discuss a visualization tool to facilitate the systematic study of chemistry and biology based on open linked data such as Bio2RDF or LODD. This tool aims at the capability of implicit retrieval of biological/chemical networks from Bio2RDF and LODD and integrated nonlinear and contextual visualization of these networks. For example, using this tool, researchers would be able to extract and visualize networks such as drug-target interaction networks, or metabolic networks in the context of Bio2RDF or

LODD, which means that they could obtain detailed information, such as biological pathways that contains certain targets or properties of a drug, while viewing and analyzing the networks. This tool is built using Jena, a semantic web framework, and integrated to Cytoscape, a widely used system biology network visualization and analysis software. The many types of view of information are achieved by assembling various semantic lenses, which are tools for context-specific display of data.

CINF 51 iTunes for chemistry and biology: Addressing usability in cheminformatics

D.-T. Nguyen, Y. Wang, R. Guha, **N. Southall**, southalln@mail.nih.gov, R. Huang, and **A. Jadhav**, ajadhav@mail.nih.gov. NIH Chemical Genomics Center, National Institutes of Health, Bethesda, MD, United States

Giving scientists access to data through a well-established framework can empower naïve users to perform very complicated tasks. The analysis and interpretation of large datasets from high throughput screening (HTS) is one such task. It involves differentiating bioactives from assay artifacts, rapid identification of chemical series, interpretation of structure-activity-relationships, and the assessment of activity profiles of series in context of related biological assays. The proliferation of available chemical and biological databases in the public domain makes the task even more complicated. Such analysis is often performed in ad-hoc manner using various combinations of commercial and in-house tools. We have been developing an open-source system that integrates over 70 public databases and makes them available in a user-friendly way, which enables rapid analysis and visualization of chemical libraries and HTS data. In this talk, we highlight some key features of the current platform: data visualization, fragment-based automated analysis, mining of activity databases such as PubChem and StARLite, and the integration of compounds, fragments, genes, targets, pathways, and diseases with minimal user effort.

CINF 52 Interaction Maps: Understanding compound pathway space

Y. Tanrikulu[1], yusuf.tanrikulu@roche.com, M. Weisel[1], R. Kondru[2], and H. M. Bitter[1]. [1]In Silico Sciences, Hoffmann-LaRoche, Nutley, New Jersey, United States, [2]Discovery Chemistry, Hoffmann-LaRoche, Nutley, New Jersey, United States

Traditional efforts in bio- and cheminformatics have utilized sequence identity and structural homology as a way to relate proteins to gain a better understanding of biological pathways and processes. Here, we unveil for the first time an internally developed data representation which allows for linking targets through chemical space, that is by their ligands. Initial efforts have been conducted in academia and thus have been limited by the relatively sparse data available in the public domain. In-depth analysis of our in-house data enabled us to generate mappings of compounds to targets, which may reveal molecules which are being profiled for one target but show greater potency for another. Therefore, our method may provide a fresh perspective on mechanisms of action, potential repurposing of drugs as well as identification of off-target events and novel insights of well-known diseases by integrating chemical and biological space or "ligand-target profiles".

CINF 53 PubChem: A public cheminformatics system for bioactivity analysis

Y. Wang, ywang@ncbi.nlm.nih.gov. NIH, NCBI/NLM, Bethesda, MD, United States

PubChem (http://pubchem.ncbi.nlm.nih.gov) is a public repository for chemical structures and biological test results hosted by the U.S. National Institutes of Health(NIH). It was created as part of the research program under the NIH Molecular Libraries Program (MLP). PubChem's contents are currently derived from the voluntary contributions of over 90 commercial, academic and government organizations, with the majority of biological test results contributed by the MLP funded laboratories. PubChem currently contains 61 million chemical substance records, representing more than 25 million unique chemical structures. The PubChem BioAssay component contains 1700 bioassay depositions with more than 50 million bioactivity outcomes involving over 3000 molecular targets. The goal of the PubChem BioAssay system is to optimize the utility of such rich information, and to provide services to make the biological activity data of this large scale collection easily accessible to the public. PubChem can be accessed and searched through the NCBI Entrez information retrieval system. Furthermore, PubChem provides an integrated cheminformatics system to facilitate the utilization of the information. Web-based summary services are provided allowing one to review the deposited substance and bioassay record in detail. A suite of exploratory bioactivity analysis tools are developed to integrate the chemical and biological activity information, and to support the navigation and in-depth data analysis that facilitates identification of chemical probes and biological targets contained within PubChem.

CINF 54 High throughput synthesis and screening system for discovery of improved materials for rechargeable batteries

B. Li, F. Matsumoto, B. Howard, R. Olugbile, D. Greenburg, C. Riley, C. O'Neill, M. S. Bailey, and **S. S. Kaye**, skaye@wildcatdiscovery.com. Wildcat Discovery Technologies, San Diego, CA, United States

Substantial improvement in the energy density of batteries is required to meet the future needs for vehicle electrification, energy storage, consumer electronics, and medical devices. Recent trends highlight the importance of simultaneous optimization of many components, including active material microstructure, electrolyte, binder, conductive diluents and porosity. Historically, these were optimized independently and often years apart in separate laboratories. Wildcat Discovery Technologies has developed a platform for high throughput discovery of battery materials that enables evaluation of thousands of cells per week. Wildcat's system produces materials in bulk form, enabling high throughput formulation of the electrochemically active material into an electrode and evaluation of its properties in a complete cell. This allows parallel optimization of all aspects of the cell, including active materials, binders, separator, electrolyte and additives. Here, I will discuss the development and validation of Wildcat's high throughput discovery platform and our initial results from our first discovery libraries.

CINF 55 The Clean Energy Project: Finding new renewable energy materials, one screensaver at a time

A. Aspuru-Guzik, aspuru@chemistry.harvard.edu. Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, United States

In collaboration with IBM's World Community Grid, my reserach group and I have embarked on "The Clean Energy Project", an endeavour to scan the electronic properties of tens of thousands of molecular candidates for organic photovoltaic materials using computer time from voluntary donors around the world. The users download a distributed computing client that performs calculations to aid in the development of a database of compounds. The project encompasses the combinatorial generation of candidates, the distributed molecular dynamics and electronic structure calculations, and a user-friendly property database. New candidates for high hole and electron mobilities, good light-absorption properties and photoresistance are sought. In this talk, I will provide an overview of the project and discuss our preliminary results.

CINF 56 Compound discovery and design via materials informatics: An "omics" approach to materials science

K. Rajan, krajan@iastate.edu. Department of Materials Science and Engineering, Iowa State University, Ames, IA, United States

One of the fundamental questions in materials crystal chemistry is why do atoms arrange themselves in the way they do? Historically, there have been two primary approaches for addressing this challenge. One is to search, organize and classify homologous compounds from known data. Another approach has been to take advantage of advances in first principles computational modeling and high performance computing methods to explore large numbers of combinations of atomic configurations to search and identify potentially new structures. In this presentation, we describe a new alternative strategy, based on statistical learning. It systematically integrates diverse attributes of chemical and electronic structure descriptors of atoms with descriptors associated with crystal structure of compounds to capture complexity in crystal geometry and bonding. Using informatics, we have been able to discover new compounds with targeted functionalities, the chemical design rules governing the stability of these compounds as well new structure-property relationships.

CINF 57 Organic electronic materials by design: Finding a needle through the haystack

G. R. Hutchison[1], geoffh@pitt.edu, and N. O'Boyle[2]. [1]Department of Chemistry, University of Pittsburgh, Pittsburgh, PA, United States, [2]Analytical Biological Chemistry Research Facility, University College Cork, Ireland

Organic electronic materials show promise for exciting applications ranging from energy-efficient displays to inexpensive, flexible solar cells. We discuss how to use cheminformatics techniques to efficiently search molecular structures in silico, including developing significant diversity sets, finding efficient, accurate screening methods for electronic structure and charge transport, and using genetic algorithms to find novel photovoltaics.

CINF 58 Searching chemical space by inverse design

B. C. Rinderspacher, berend.rinderspacher@arl.army.mil, J. Andzelm, A. Rawlett, J. Dougherty, and R. Lambeth. Army Research Laboratory, Aberdeen Proving Ground, Maryland, United States

Finding optimal architectures of functional molecules that contain numerous substituents in various conformations is a daunting task. The number of accessible compounds with molecular weight below 850 a.m.u. (a typical size, e.g., of electro-optical chromophores) is staggering at about 10^{200}. More than 20 years of research have established many rules for the optimal design of chromophores, effectively reducing the design space. Nevertheless, surprising new architectures of chromophores are appearing at increased rates with computational chemistry playing an important partnership role in these discoveries. Efficient exploration of enormous molecular spaces requires new optimization schemes that search directly for the property optimum, and in the process discover new molecular architectures. Often the objective is subject to multiple constraints, further complicating the search. We introduce a general optimization algorithm based on an interpolation of property values on a hypercube. The resultant algorithm is related to branch and bound/tree search methods. To demonstrate the broad applicability, we have applied the framework to optimizing electro-optic chromophores under spectral constraints as well as to optimizing organo-metallic clusters for tensile strength.

CINF 59 High-throughout quantum chemistry and virtual screening for materials solutions

M. D. Halls, mhalls@mhalls.com. Materials Science Division, Accelrys, Inc., San Diego, CA, United States

The use of virtual structure libraries for computational screening to identify lead systems for further investigation has become a standard approach in drug discovery. Transferring this paradigm to challenges in material science is a recent possibility due to advances in the speed of computational resources and the efficiency and stability of materials modeling packages. This makes it possible for individual calculation steps to be executed in sequence comprising a high-throughput quantum chemistry workflow, in which material systems of varying structure and composition are analyzed in an automated fashion with the results collected in a growing data record. This record can then be sorted and mined to identify lead candidates and establish critical structure-property limits within a given chemical design space. To-date, only a small number of studies have been reported in which quantum chemical calculations are used in a high-throughput fashion to compute properties and screen for optimal materials solutions. However, with time high-throughput computational screening will become central to advanced materials research. In this presentation, the use of high-throughput quantum chemistry to analyze and screen a materials structure library, representing a well defined chemical design space, is demonstrated for organic light-emitting diode (OLED) materials and Li-ion battery additives. The OLED case focuses on the chemical design space around the archetype electron-transport and emitting material tris(8-hydroxyquinoline) aluminum(III) (Alq3). The battery additive example explores the property space for derivatives of ethylene carbonate (EC).

CINF 60 Impacts of closing academic branch libraries

S. J. Redalje, curie@u.washington.edu. University of Washington, United States

How do you complete a year-long celebration of your branch library's 20 years in its current location? By closing it, of course. This is what happened to the University of Washington's Chemistry Library. It, along with two other branches, were closed and merged with the main library at the end of the 2008-2009 school year. Nationally, greatly reduced budgets and the desire to take advantage of new technologies and approaches to better serve the users 24/7 are some of the factors driving the closures and mergers of academic branch libraries. What are the impacts on the students and faculty? What about the librarians whose lives have been turned upside down? These, and other factors will be discussed in the paper.

CINF 61 e-Books and e-references collection in academic libraries: A case study

N. Xiao, nxiao@usc.edu. Science and Engineering Library, University of Southern California, Los Angeles, CA, United States

The nature of publishing and communication in science and technology fields has changed dramatically, and more and more collection are moving on-line and/or digitized. How will librarians select users-need center e-collection to support institution's education and research mission, especially under challenging economic environment? This presentation will attempt to answer this question. It will overview e-books and e-references collection change at USC Science and Engineering Library, which has space shortage issue

and supports a very large SE information users population, including more and more users from engineering distance education programs (http://mapp.usc.edu/distanceeducation/). Issues of pricing models, e-collection accessibility and flexibility, and benefits to users and the institution in short and long term will be discussed as well.

CINF 62 Future of the liaison in academic libraries

M. Lafferty, mlaffert@umn.edu. University Libraries, University of Minnesota, Minneapolis, Minnesota, United States

The role of the departmental library liaison in colleges and universities has been in flux over the last decade and continues to shift, largely as a result of changes in technology. Over the last couple of years, the University of Minnesota Libraries created a position description framework to capture and describe the new roles liaisons are taking on and clarify what is expected of them. Among other changes, the emphasis of our positions is moving more towards supporting the process of research and away from the end products. I will address where we are currently and where we, as well as liaisons in general, might be headed.

CINF 63 New workspaces: From library facility to online service

L. Solla, LRM1@cornell.edu. Physical Sciences Library, Cornell University, Ithaca, NY, United States

Local users of the former Physical Sciences Library at Cornell once had a one-stop place for library services: access to journals and books, librarians, computers and quiet group study. Now the journals and books are online, the librarians and computers are roving and the quiet group study options are changing with new building spaces. Although the physical facility closed in December 2009, the library's presence continues as a portal for scholarly resources, a virtual collection and service through research specialists. The Edna McConnell Clark Physical Sciences Virtual Library provides 24/7 electronic access to scholarly literature and information sources in astronomy, chemistry, and physics for the Cornell community. Print materials in these subject areas are available at a number of Cornell library facilities and subject librarians are on-site in the physical sciences building cluster to provide research services and support for faculty, staff and students. 9 months of strategic planning by several teams of faculty and students in the physical sciences working with library staff went into gathering input and mapping out the transition process from a facility to an online service model. At press time, anticipated results include enhanced online access to journal backfiles, books, reference databases, online browsing, document delivery, and workflow services delivered from a new local taxonomy based website. This presentation will address the status of these results, the effectiveness for the users and continuing future strategies.

CINF 64 PLoS ONE: A new model for online journal publishing

P. Binfield, pbinfield@plos.org. Public Library of Science, San Francisco, CA, United States

PLoS ONE (www.plosone.org) is an innovative, online-only, Open Access, peer-reviewed journal from the Public Library of Science (www.plos.org). Since launch in December 2006 *PLoS ONE* has grown to be the third largest journal in the world (having published 4,600 articles in 2009) and was the recipient of the 2009 ALPSP Award for Publishing Innovation. This presentation will detail the innovative editorial criteria of the journal (which do not make any subjective judgments about the potential 'importance' of any submission); PLoS's development of post-publication evaluation tools at the article-level (so called 'article-level metrics, which record citations, online usage, comments, ratings, social bookmark data and so on, on each article); some of the more interesting functionality in the journal; and the use of 'web 2.0' social media tools to effectively promote the journal in a cost effective manner.

The presentation will be made by the Publisher of *PLoS ONE*.

CINF 65 Impact of emerging economies on publishing

J. Bhate, jignesh@molecularconnections.com. Corporate Development Centre, Molecular Connections Private Limited, Basavanagudi, Bangalore, India

Over the next few years, developing economies such as China and India, are expected to continue their dramatic growth. These developments are influencing global companies to realign their businesses and operational planning in order to establish increased presence in these developing markets. This trend is already visible among leading publishers. This talk highlights economic performance indicators as well as other significant market drivers such as the availability of a highly skilled and educated workforce capable of fulfilling a variety of professional jobs, including information and publishing roles. Predictions for the

future shape of the primary and secondary publishing industry in the next five to ten years will be made.

CINF 66 Ensuring sustainability of scientific data resources: Technologies and business models for the future

C. R. Groom, edir@ccdc.cam.ac.uk. Cambridge Crystallographic Data Centre, United Kingdom

The Cambridge Crystallographic Data Centre provides the Cambridge Structural Database to scientists in well over 1000 institution, spread over 69 countries. To do this we must balance a number of conflicting requirements; the need for complex processing and curation contrasts with the need to handle large volumes of data. Let's see how 'technology' will help the CCDC. We must also deliver knowledge and data in a way that appeals to users – we can learn a great deal from other communities. We are changing our organisation to exploit technological advances, but must also respond to the continued challenge to established publishing business models. We are adapting to our new environment, but can also shape this environment ourselves. This presentation will discuss whether the technology we use and business model we have serve us for today. Most importantly we'll look to the years ahead – what must we change?

CINF 67 Vision and strategy for the future of RSC Publishing

R. J. Parker, parkerr@rsc.org, and J. Milne. Royal Society of Chemistry, Cambridge, United Kingdom

RSC Publishing is currently a large small publisher: large compared with many scientific society publishers, but small compared with most commercial publishers. What might the future hold for such an animal, especially during and after a global economic downturn combined with a period of unprecedented disruptive change? How will it evolve and adapt to take advantage of the challenges and opportunities that it faces?

CINF 68 Challenges facing the STM industry

M. O'Malley, m.omalley@elsevier.com. Corporate Markets, Elsevier, Inc., New York, NY, United States

The broad trends of technological advance, globalisation, and information proliferation present threats and opportunities for the STM customer base and the STM industry. In this presentation we discuss the evolution of the STM publishing industry, against the backdrop of historical and near-term change. We conclude that to meet the challenges of a continually evolving market, both large and small players need to transition their role from traditional publishers to information solution providers. This transformation requires a close collaboration with customers, a willingness to partner openly with technology providers and competitors, and new transparent approaches to solution development and deployment.

CINF 69 Vision and strategy for scholarly communication at the ACS

S. King, s_king@acs.org. American Chemical Society, Washington, DC, United States

ACS is uniquely positioned to serve the interdisciplinary communications needs of research professionals as well as foster the education and practice of chemistry and allied sciences. The Society's mission-driven support of scholarly communication encompasses the application of innovative publishing technologies that serve the needs of authors and readers of ACS peer-reviewed journals in traditional and evolving areas and that are among the most highly cited publications in the world; the scientific reporting excellence contained in ÇEN, the leading weekly news magazine for the chemical industry; the comprehensive indexing by CAS of the global research and patent literature across chemistry and allied fields; and the overall coordination of the world's largest scientific society, including local chapters and technical divisions and the programming for two national meetings per year. Building on integrated content linking achieved among CAS *SciFinder*® resources and the Web Editions of ACS Journals, as well as the social networking features of the ACS Network, the ACS has just begun to tap the power of Web 2.0 and evolving information technologies to further foster scholarly communication. ACS welcomes its members, authors, readers, customers, and the global scientific community to join with us as we move beyond traditional boundaries, while preserving the quality and integrity of scholarly communication.

CINF 70 Network visualization of structure activity landscapes

R. Guha, guhar@mail.nih.gov. NIH Chemical Genomics Center, Rockville, MD, United States

Lead optimization is an integral part of the drug discovery process. During this stage synthetic programs will generate a number of compounds, generally based on a few chemotypes. A key feature of such programs is identifying small structural changes that lead to significant changes in measured activity - so called

'activity cliffs'. Such cliffs represent interesting regions of the SAR landscape and can be suggestive of specific receptor binding features, as well as suggest new directions for synthesis. In this paper we present a visualization of SAR landscapes, which allows users to highlight activity cliffs. The approach is based on a network representation of a molecular dataset (structures and activities), and employs the Structure Activity Landscape Index (SALI). This metric allows one to quantitate the degree to which a pair of ocmpounds represent an activity cliff. By using the SALI value as a cutoff one can generate multiple network representations that highlight activity cliffs of varying degree. We will show that the structure of the network allows one to explore synthetic programs to identify modifications that lead to improvements in activity. While useful as a retrospective visual summary of a synthetic program we will also highlight the use of this approach in quantitatively charactering the quality of QSAR models that attempt to encode a SAR landscape. This methodology has been implemented in a Java-based tool, which is freely available, allowing users to generate SALI networks with their own data.

CINF 71 Visualization and analysis of bioisosteric analogs generated with BROOD

J. J. Corkery[1], jcorkery@eyesopen.com, A. G. Skillman[2], K. E. Schmidt[1], and B. Kelley[1]. [1]OpenEye Scientific Software, Cambridge, MA, United States, [2]OpenEye Scientific Software, Santa Fe, NM, United States

BROOD is a program developed by OpenEye Scientific Software to generate bioisosteres based on shape and electrostatic similarity. A single BROOD run may generate 1000s of potential ideas for a specified query. Not surprisingly, visualization and analysis of this data can be particularly challenging. To solve this problem, a custom Python extension to facilitate BROOD visualization was developed for VIDA (OpenEye's visualization and data analysis platform). This extension enables the ideas to be easily organized into "Idea Groups" which can quickly and easily be viewed in 2D, 3D, or both simultaneously. In addition, all of the non-structural data generated by BROOD (e.g. synthetic accessibility, similarity to query, etc) can be viewed side-by-side with the structures or in the embedded chemically-aware spreadsheet which can be used to further filter/organize the results. The techniques utilized in this talk can easily be extrapolated to facilitate visualization and analysis of a large variety of other result types including but not limited to clustering output and a wide variety of virtual screening results.

CINF 72 Visualizing polypharmacology in chemical libraries

D. Marcus, david.marcus1@mail.huji.ac.il, D. Barasch, A. Rayan, and A. Goldblum. Institute of Drug Research, The Hebrew University of Jerusalem, Jerusalem, Israel

Modern drug discovery increasingly focuses on drugs that should interact with multiple targets to elicit the desired pharmacological response. Visualizing several SAR models at once may be helpful for easier interpretation when comparing different chemical libraries or for picking new actives. A colored matrix is commonly used to visualize each compound's predictions, however this visualization is highly complex when analyzing large libraries. We developed a ranking method to prioritize compounds by both their activity and their reliability using a single index value. We display these indexes by tag clouds that show each target as a tag and vary by size, color and position, which clarifies better the polypharmacology contents of each library. Visually comparing clouds from different libraries makes comparisons easier to evaluate. We also demonstrate how we use this visualization with huge libraries such as ZINC, to pick molecules that have the desired polypharmacology nature in several cases.

CINF 73 Linear scaling mapping of chemical space

J. S. Delaney, john.delaney@syngenta.com. Chemistry, Syngenta, Bracknell, Berkshire, United Kingdom

A novel method for producing 2D maps of chemical space is presented. Each compound is represented by standard binary substructural fingerprint (Daylight or SciTegic) which is used to produce an integer number in a defined range. This is used to define a 2D coordinate for each compound by mapping the number onto a Hilbert space-filling curve. The key advantages of the method are that it scales linearly with the number of compounds to be mapped (allowing large datasets to be processed) and it produces an absolute coordinate in a well-defined range for each compound, allowing maps of different compound sets to be compared on a like-for-like basis. Examples of the method applied to large compound sets (300k+) and comparisons between maps of sets with different properties (structural and biological activity) will be presented.

CINF 74 Integrated visualization for the interpretation of QSPR and QSAR models

G. D. Purvis[1], gpurvis@us.fujitsu.com, W. D. Laidig[2], and D. T. Stanton[2]. [1]Biosciences Group, Fujitsu America Inc, Beaverton, Oregon, United States, [2]Modeling Simulation, Miami Valley Innovation Center, The Procter Gamble Company, Cincinnati, OH, United States

Molecular structure visualization integrated with conventional 2D plots of quantitative structure property relationships (QSPR) and quantative structure activity relationships (QSAR) is shown to speed the analyis and interpertation of QSPR/QSAR models. Identification of outliers is simplified and relationships of specific molecular descriptors to molecular structure are more easily comprehended. Integrated visualization is demonstrated for least squares models, partial least squares (PLS) models and robust regression modeling.

CINF 75 VTK: The visualization toolkit and its applications to the analysis of chemical data

M. D. Hanwell, marcus.hanwell@kitware.com. Department of Scientific Visualization, Kitware, Inc., Clifton Park, NY, United States

The visualization toolkit (VTK) is an open source, cross platform, freely available software library providing a framework for visualization ranging from 3D computer graphics, image processing and data filtering to information visualization and standard 2D plotting. It is written in C++, tested daily on Linux, Mac and Windows, and has bindings for Python, Tcl and Java. VTK has been used in chemical visualization applications such as Molekel (C++) and the CCP1 GUI (Python). Related applications such as ParaView use VTK to process and visualize very large data sets. Many of the same algorithms and techniques can be used by chemists to create new visual representations of large data sets. This presentation gives an overview of the functionality available in VTK, and ways in which it can be leveraged to rapidly analyze and visualize chemical data.

CINF 76 Molecular similarity characterization of ADME landscapes

R. R. Gupta[1], rishi.gupta@pfizer.com, **B. Chen**[2], binchen@indiana.edu, and E. M. Gifford[1]. [1]Computational Sciences Center of Emphasis, Pfizer Global Research and Development, Groton, CT, United States, [2]School of Informatics and Computing, Indiana University, Bloomington, IN, United States

Activity landscape characterization has been demonstrated to be a valuable tool in screening and lead optimization. In this work, we applied various molecular similarity methods to characterize chemical landscapes for 9 ADME (Absorption, Distribution, Metabolism, and Excretion) endpoints. Our results show that landscapes behave differently among endpoints and the observation can be quantified to prioritize compounds while transforming compound from HIGH risk class to LOW risk class.

CINF 77 Chemical information usage patterns under the microscope: The prospects of GIS-based visualization for libraries

M. P. Brändle, braendle@chem.ethz.ch. Informationszentrum Chemie Biologie Pharmazie, ETH Zürich, Zürich, Switzerland

The utility of geographical information systems (GIS) as decision support tools that provide analysis of geocoded demographic data on maps had been recognized in particular by public libraries in the UK and US. However, GIS can also be used for in-house analysis on library floor plans, as J. Xia had shown in preliminary work (e.g. J. Xia, Library Hi Tech 22(2), 209(2004)). The topic has recently received broad attention since we published interactive visualizations of usage and holding patterns on the Web (www.infochembio.ethz.ch/en/holdings.html). In this presentation, technical and organizational prerequisites for library space-based visualization of library data, its benefits for collection management and user guidance, and novel descriptors for library collection analysis will be discussed.

CINF 78 Scholarly communication in the 21st Century: Changing paradigms of communication and access

R. Schwarzwalder, rns@stanford.edu. Cecil H. Green Library, Stanford University, Stanford, CA, United States

The world of scholarly communication is experiencing a growing identity crisis. While the structure of academic publishing and its role in establishing intellectual property and academic credentials continues to be traditionally defined, the nature of information access and usage is being revolutionized by technical, business and social forces. This presentation will present an overview on the factors that have preserved the traditional model of academic publication and the changing environment that threatens the continued viability of that approach. Ideas will be discussed as to how traditional publishers could

respond to the changing nature of scholarship without abandoning their core mission and focus.

CINF 79 oreChem project: Semantic infrastructure and applications for chemistry scholarship

C. Lagoze, clagoze@gmail.com. Information Science Program, Cornell University, Ithaca, NY, United States

The oreChem Project is an international collaboration between chemists and information/computer scientists to design, prototype, and deploy a semantics-aware information infrastructure for chemistry scholarship and applications to demonstrate the utility of that infrastructure. A major component of this work is the development of ontologies to represent in machine-readable form information about scholarly publications, molecular compounds, experiments, and reactions. These ontologies provide the foundation for the representation of information retrospectively extracted from legacy publications and information proactively accumulated in electronic laboratories notebooks and from instruments. A subsequent phase of this work will leverage this data as the basis for innovative applications for searching, browsing, and mining chemical information.

CINF 80 Chemistry: Computation, synthesis and communication

J. M. Goodman, jmg11@cam.ac.uk. Department of Chemistry, University of Cambridge, Cambridge, United Kingdom

My research group has projects in both experimental and computational chemistry. What form of publishing would best help us to communicate our disparate results in a form that is accessible to all our readers without prejudicing the interests of our sponsors? We try to balance between reporting enough to be valuable without eroding the intellectual property we share with everyone involved in the project, using a variety of forms of scholarly publishing. I will explain how I would like to see scholarly publishing develop in the fields of computational and synthetic organic chemistry.

CINF 81 Primary data for chemistry

I. Sens[1], irina.sens@tib.uni-hannover.de, S. Haak[2], Susanne.Haak@thieme.de, J. Brase[1], Jan.Brase@tib-hannover.de, and **G. F. Herrmann**[2], Guido.Herrmann@thieme.de. [1]German National Library of Science and Technology (TIB), Hannover, Germany, [2]Georg Thieme Verlag KG, Stuttgart, Germany

Primary data is scientific data gathered from experimental measurements and predominately available in electronic formats. In the field of chemistry, such data is accumulated by a variety of analytical, spectroscopic or computer simulation methods. Analytical data, from various experiments, is the foundation of research work and scientific papers. Thus far, the vast amount of data lies scattered on the computers of scientists, who have produced the information. As no central repository exists, no accessible archival storage is possible at the moment. Scientific results are solely published in journals – but not the primary data from which those results originate. The German National Library of Science and Technology (TIB) and Thieme Publishers have addressed this problem in collaboration. In this talk we will present a system and a process to systematically make primary chemistry data accessible worldwide. Thieme and TIB have developed a uniform structure for publishing primary data. Through structuring and central data registration, a Germany-wide unique service of TIB, valuable knowledge will be harnessed. From now on, primary data will be registered and made available online via the Thieme eJournals website (www.thieme-connect.com/ejournals) using digital object recognition in the form of Digital Object Identifiers (DOI). The publication of primary data will be implemented into the existing system of publications as well as their citations. The data will be permanently saved and, by assigning them a DOI, made accessible and searchable, as well as citable and linkable. This enhancement to the chemistry publishing process will enable scientists to easily locate research articles, including accompanying data, and make more productive use of the scientific content. Peer review will benefit from access to the original primary data by potential improvements to the process and better validation of the presented research. Primary Data Sample (FREE access):

K. Jarowicki, C. Kilner, P. J. Kocienski*, Z. Komsta, J. E. Milne, A. Wojtasiewicz, V. Coombs Synthesis 2008, 2747-2763:

http://dx.doi.org/10.1055/s-2008-1067226

CINF 82 Papers to papers: The future of reading, storing, and producing chemical literature in an academic environment

J. N. Currano, currano@pobox.upenn.edu. Chemistry Library, University of Pennsylvania, Philadelphia, PA, United States

The proliferation of primary literature in chemistry has created a conundrum for the well-read, well-published chemist. Managing the flow of information from the outside world is challenging, but keeping the useful articles that one finds at one's fingertips is even more daunting. This paper reviews techniques currently used by academic chemists to assemble and manage their own electronic file cabinets, focusing on the needs and desires of the researchers, and closing with a suggested model system extrapolated from requests articulated by scientists during the course of library reference encounters.

CINF 83 Socialized medicine: Medical publishing in the era of the read-write web

K. R. Anderson, kanderson@nejm.org. New England Journal of Medicine, Waltham, MA, United States

Publishers have long been experts in distribution, formatting, packaging, and selling content for distribution. However, as Clay Shirky stated in 2009, "The Internet is the first media that has native capacity to support both groups and conversations at the same time." This digital publishing revolution is forcing publishers from a broadcast distribution paradigm into a conversation paradigm. For established brands, embracing this shift requires a retooling and rethinking, and a test of skills and goals. At the New England Journal of Medicine, the Publishing Division is undertaking an ambitious social media publishing project called CardioExchange, which combines semantically tagged content, editorial curation, community management, and social media tools to create a new value equation around an important and dynamic medical specialty. This presentation will provide an overview of the smaller community experiments that led to this more ambitious pilot, and will cover the rapid, agile development process used to build the pilot, considerations in community creation and definition, measurements and goals, staff roles and work definition changes, and marketing/engagement issues, as well as presenting results from the experiment thus far.

CINF 84 Scholarly communication between chemists and their uptake and use of Web 2.0 and other new technologies

W. G. Town, bill.town@kilmorie.com. Kilmorie Clarke Ltd, London, United Kingdom

Web 2.0 and other new technologies potentially can bring beneficial changes in the access to, use and re-use of journal and database content by chemists. However, it is widely believed that chemists are slow in adopting change although so far there has been little factual evidence to support this. This project set out to make a snapshot of scholarly communication practices of chemists by undertaking a series of individual interviews with key decision makers and then conducting an online survey to identify the specific requirements of chemistry researchers and teachers. The defined requirements were then matched to an up-to-date list (also developed as part of the project) of new and recent developments in methods of scholarly communication used by chemists. This presentation will give a summary of the research findings and present ideas for improving the situation. Comparison will be made with a parallel study of economists.

CINF 85 ChemSci3D: An integrated 2D and 3D desktop chemistry interface

T. J. O'Donnell, tjo@gnova.com. gNova, Inc., San Diego, CA, United States

There are many types of 2D and 3D models of chemical information. Some computer applications concentrate on the display of 3D structural information. Others excel at representation of 2D structures, tables and plots. I present here a new application that can display of a variety of 2D and 3D models in a single desktop interface. This is accomplished by integrating a general 3D graphics and animation system (GRAMPS) with tools to display 2D structures, tables of data, plots and other diagrams. Currently under development are techniques to allow interactions among these various graphical components. I will show examples of how this integrated chemistry interface can be used and discuss the underlying software architecture components. I will also show ways in which ChemSci3D can be used to integrate novel chemical models.

CINF 86 Real-time ray-tracing in molecular graphics

M. Keil, mkeil@chemcomp.com. Chemical Computing Group, Inc., Montreal, Quebec, Canada

In recent years a wide variety of new features were added to modern graphics cards. One major enhancement is the ability to directly program parts of the graphics pipeline with the new OpenGL Shading language. It is even possible to perform computations on the graphic processing unit (GPU). This talk will present how one can enhance molecular graphics by ray-tracing basic geometric objects like spheres and cylinders. Due to the highly parallel nature of modern GPUs this ray-tracing can be done in real-time and even outperform traditional triangle based drawing methods. Other molecular graphics improvements using modern graphics card features will be discussed as well.

CINF 87 Avogadro: Framework for chemical structure analysis and visualization

M. D. Hanwell, marcus.hanwell@kitware.com. Department of Scientific Visualization, Kitware, Inc., Clifton Park, NY, United States

Avogadro is an open source, cross platform visualization framework with a well defined C++ API, and Python bindings. It makes use of OpenBabel, and so it is possible to use the functionality of OpenBabel and PyBel within Avogadro's API. The library is plugin based, and can be extended at runtime by both compiled C++ plugins and Python extension scripts. This means that researchers can easily tailor Avogadro to their specialized area by writing a few small scripts and/or plugin classes. Details will be presented on creating custom applications that leverage the Avogadro library, and how to write plugins that can visualize structures using custom techniques without the need to deal with the low level APIs and boilerplate code usually necessary. Several examples of Avogadro's use in industry and research will be presented.

CINF 88 2D graphical depiction of zeolite voidspace topology

K. J. Theisen[1,2], kevin@tfam.us, B. Smit[1], and M. Haranczyk[2]. [1]Department of Chemistry, University of California: Berkeley, Berkeley, CA, United States, [2]Department of Computational Research, Lawrence Berkeley National Laboratory, Berkeley, CA, United States

Zeolite materials contain complex networks of nanopores and channels that can be exploited for industrial applications. Nearly 200 unique embeddings have been synthesized with millions predicted. Researchers perform lengthy simulations on these structures to identify optimal choices for tasks such as adsorbate separation and catalytic conversion. Researchers usually rely on 3D visualization to gain insight about structural features and choose embeddings to investigate, a task that is often hindered by the structure's sheer complexity, especially if an orthographic perspective view is used. We have developed an alternative method to depict the important features of these complex structures, facilitating their visual analysis. Our approach involves reducing these structures based on symmetry and perceived building blocks to a compressed two-dimensional graphical representation that allows for easy and efficient structure analysis and comparison.

CINF 89 Computational methods for 2D-visualization of molecular interaction patterns

K. Stierand[1], **M. Rarey**[1], rarey@zbh.uni-hamburg.de, and C. Lemmen[2]. [1]Center for Bioinformatics, University of Hamburg, Hamburg, Germany, [2]BioSolveIT GmbH, Sankt Augustin, Germany

Chemists are well trained in perceiving 2D molecular sketches. On the side of computer assistance, the automated generation of such sketches becomes very difficult when it comes to multi-molecular arrangements such as protein-ligand complexes in a drug design context. During the last few years we have developed PoseView,[1,2] a tool which displays molecular complexes incorporating a simple, easy-to-perceive arrangement of the ligand and the amino acids to which it forms interactions. Resulting in atomic resolution diagrams, PoseView operates on a fast tree re-arrangement algorithm to minimize crossing lines in the sketches. Due to a de-coupling of interaction perception and the drawing engine, PoseView can draw any interactions, such as hydrogen bonds, metal interactions, pi interactions and undirected hydrophobic contacts, determined by either distance-based rules or the FlexX interaction model. Owing to the small-molecule drawing engine 2Ddraw[3], molecules are drawn in a textbook-like manner following the IUPAC regulations. Besides the novel underlying interaction models, we will present new algorithmic approaches, assess usability issues and a large-scale validation study on the PDB.

[1] K. Stierand, M. Rarey, ChemMedChem, **2007**, 2: 853. [2] K. Stierand, P. Maaß, M. Rarey, Bioinformatics **2006**, 22, 1710. [3] P.C. Fricker, M.

Gastreich, M. Rarey, J. Chem. Inf. Comput. Sci. **2004**, 44(3): 1065.

CINF 90 Visualizing combinatorial molecular materials research using Avogadro

G. R. Hutchison, geoffh@pitt.edu, and M. D. Hanwell. Department of Chemistry, University of Pittsburgh, Pittsburgh, PA, United States

The open source molecular visualization package Avogadro can be used interactively, or as a scripted viewer. We outline methods to explore combinatorial data sets for organic electronic materials using the Python scripting interface, particularly correlating multi-dimensional data on screen. We explore the use of color to focus on particular molecular structure motifs. In particular, the plugin interfaces in Avogadro allow users to easily change rendering style or color to reflect correlated data.

CINF 91 Visualizing chemistry in Second Life

A. S. I. D. Lang[1], alang@oru.edu, and **J.-C. Bradley**[2], bradlejc@drexel.edu. [1]Department of Computer Science and Mathematics, Oral Roberts University, Tulsa, OK, United States, [2]Department of Chemistry, Drexel University, Philadelphia, PA, United States

This talk will focus on the current level on chemistry research, education, and visualization possible within the multi-user virtual environment of Second Life. We will discuss how Second Life has been used as a platform for the interactive and collaborative visualization of data from molecules and proteins to spectra and experimental data. We then review how these visualizations can be scripted for immersive educational activities and real-life collaborative research. We also discuss the benefits of the social networking within Second Life for both chemists and chemistry students.

CINF 92 Next-generation visualization technologies: How close are we to the Science 2.0?

O. Isayev, olexandr.isayev@case.edu. Department of Chemistry, Case Western Reserve University, Cleveland, OH, United States

A set of technologies called Web 2.0 is transformed the Internet. Web sites such as YouTube, and Facebook, in addition to various blogs and wikis attract hundreds of millions of people. Yet this Web 2.0 transformation of science and chemical visualization in particular is just in process. How close we are to so called "Science 2.0"? Since its inception, the Internet has changed the ways scientists work, communicate, collaborate, and educate. Successful scientific collaboratories among researchers, engineering innovations through open-source software, and community-based participation in projects are all indicators of the transformative nature of collaboration. However, what is the influence on chemical visualization? What are the novel ways to display graphical information in the web? What are the possible workflows? With this presentation, I would like to overview some of the latest trends and rich internet applications that aim to bring us a step closer to the Science 2.0.

CINF 93 CAS and ACS publications: Together, providing new pathways to chemical research

J. Morgan[1], j_morgan@acs.org, and **M. Dennis**[2], mdennis@acs.org. [1]Publications Division, American Chemical Society, Washington, DC, United States, [2]Chemical Abstracts Service (CAS) Division, American Chemical Society, Columbus, OH, United States

ACS Publications' newest delivery platform and CAS' latest SciFinder enhancements continue to help scientists locate quality chemical information with efficiency, while promoting electronic collaboration. This talk demonstrates these two technologies, and how when used in concert, aid in the process of scientific research and discovery. The presentation concludes with a discussion to garner feedback regarding concepts for future collaborative enhancements and its potential usefulness to scholarly communication and chemical research.

CINF 94 Adding structure to publishing chemistry

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

The way chemical research is published is moving at an ever-increasing speed. From new systems of indexing chemical structures to new methods of searching for chemical information, development in this field is fast and furious. The challenge for both publishers and authors is to ensure that the research being published is understandable not only to chemists but also to computers. Machine readable research means that content can be linked into related articles, databases, protocols and products, as well as ensuring that the content can be internally (by journal, publisher, Institutional Archive etc...) indexed and searchable. Nature Publishing Group, over the past few years, have been developing new tools and systems to reduce the requirement for authors to submit ever increasing amounts of additional information, whilst also developing automated workflows to take author files

(text and chemical structures) and link them into the information world beyond the paper.

CINF 95 Under the hood: A technical view on publishing chemistry in the future

R. Kidd, kiddr@rsc.org. Informatics Department, Royal Society of Chemistry, Cambridge, United Kingdom

The Royal Society of Chemistry's developments and experiences in semantic publishing will be summarised, along with the platforms for delivering our published content with additional functionality. The development and application of open standards for linking publications and data with the end user offers new opportunities in the ecosystem of the chemical sciences, and several projects with relevance to scholarly communication will be discussed with their implications.

CINF 96 Interactive publishing of crystal structure data

B. McMahon, bm@iucr.org, and P. R. Strickland. International Union of Crystallography, Chester, United Kingdom

Crystal structure determinations published in IUCr journals include structural (geometry, atomic displacements) and experimental data (structure factors). Nowadays these are in the Crystallographic Information Framework (CIF) standard. CIF has become the vehicle for submission of short structure reports, allowing the development of software that improves consistency and integrity of the publication, and permits semi-automated assessment of data quality during peer review. New authoring tools increase the coupling between data and article. For macromolecules, an article can be built from a Protein Data Bank deposition, incorporating deposited data, prompting for additional data items, and providing wizards to retrieve and insert data from other online resources. Another toolkit allows authors with no prior experience to create interactive three-dimensional visualizations using Jmol. Expert Jmol scripts are also supported. The resultant figure is an integral part of the publication, allowing readers to manipulate the structural model independently or according to the author's preferred views.

CINF 97 Science video journal to increase efficiency and standardization in experimental research

M. Pritsker, moshe.pritsker@jove.com. JoVE, Somerville, MA, United States

Biological and medical sciences chronically suffer from the low reproducibility of experiments, as they are typically described in the Materials and Methods sections of scientific articles. This is due to the fact that the traditional text format cannot adequately describe complex experimental procedures, creating a critical problem for academic research and drug discovery. Journal of Visualized Experiments (JoVE) applies video online to create visualized publication of experimental studies. Visualization greatly enables unambiguous presentation of experimental techniques and methods, and therefore increases the overall efficiency and standardization of biological research. During three years of its operations, JoVE has published 33 monthly issues including over 450 video-articles on experimental approaches in neuroscience, immunology, developmental biology, bioengineering and other fields. To facilitate integration of video into scientific publishing, JoVE has developed an organizational and technological structure to conduct production of scientific videos in research labs in the USA, Europe and Japan. This video-based approach to scientific publishing can be expanded to increase efficiency in other areas of experimental research including chemistry and engineering.

CINF 98 Nano-Publication

J. J. J. M. Velterop, velterop@conceptweballiance.org. Concept Web Alliance, Nijmegen, The Netherlands

What if we could freely get to all the essential knowledge contained in scientific articles, without those articles themselves necessarily being open access? And what if such free access to the essential knowledge is actually in the interest of publishers as well? For that, we need to split the roles of papers as 'scientific record keeping' (the 'interface with officialdom') on the one hand, and 'knowledge dissemination' on the other. This is made possible by what we call "nano-publication". The essence of scientifically meaningful content, especially in the natural sciences, can to a very large degree – often completely – be expressed in statements such as 'DMD interacts with SNT1', which are virtually always 'triples' of the form subject > predicate > object. Such triples can be extracted from the literature, and, particularly when semantically disambiguated, lend themselves optimally for machine reading on a large

scale and for efficiently conveying knowledge and facilitating analyses, even for quasi-mathematical 'reasoning'. Triples can be seen – and published – as 'nano-publications'. And they would naturally be open access. Whilst in principle nano-publications extracted from classical papers would be subject to copyright, this could in practice only be used to ensure proper acknowledgement (science is an 'acknowledge economy' after all). Nano-publications that are rich semantic triples are in essence references, and wide and open availability of references to the content they publish is what most publishers crave. Nano-publications are therefore necessarily open access. And this open access is actually beneficial not just to scientists, but to publishers as well, be they open access publishers or traditional ones. In addition, one can imagine business models that financially support open access nano-publications. A concrete example will be presented.

CINF 99 Speeding up chemical database searches

R. J. Nasr, rnasr@ics.uci.edu, and P. Baldi, pfbaldi@ics.uci.edu. Department of Computer Science, University of California, Irvine, Irvine, CA, United States

In many large chemoinformatics database systems, molecules are represented by long binary fingerprint vectors whose components record the presence or absence of particular functional groups or combinatorial features. To speedup database searches, we develop and validate two new approaches. In the first approach, we compute for each fingerprint a corresponding small count vector containing the sums of hashed subsets of components. From these count vectors, we derive tight bounds on fingerprint similarity measures, such as the Tanimoto measure. During a database search, every time these bounds are unfavorable, the corresponding molecule can be discarded with no need for further inspection. The second approach adapts techniques from text search based on inverted indices to chemical fingerprints. For both approaches we construct probabilistic models that allow us to estimate the level of pruning under different conditions. These theoretical results are corroborated by experimental results obtained on a large set of molecules.

CINF 100 Comparison of backfiles offered by major publishers of chemical journals

A. D. Bolek, bolek@uakron.edu. Science-Technology Library, The University of Akron, Akron, OH, United States

Most of the major publishers of chemical journals offer their backfiles in packages of several titles, including the American Chemical Society, the Royal Society of Chemistry, Elsevier, Springer, and Taylor. Francis. At the beginning of 2009, Wiley permitted organizations to purchase titles individually, or in the previously offered packages. Some publishers allow you to pay a one-time fee for the backfile with permanent access, while others require an additional access fee each year. Some allow you to lease the backfile each year, while others include the backfile with a current subscription. Most publishers allow you to purchase an article on a pay-per-view basis, with the major exceptions being JSTOR and Portico. A few publishers make their backfiles free, as does J-STAGE's Journal@rchive, the Japan Science and Technology Information Aggregator, Electronic. This poster will compare what is offered by the major publishers of chemical journals at the beginning of 2010.

CINF 101 Marketing chemical research with custom web databases

B. Brandys, brandysb@mail.nih.gov. Department of Library Services, National Institutes of Health, Bethesda, Maryland, United States

This poster is a flow chart of the development of web-accessible databases of research compounds synthesized by NIH laboratories. Two databases containing Image Probes and Nitric Oxide generating compounds are used as examples. These databases enable public access to the laboratories' research, and provide an easily accessible medium for marketing and collaboration for further development.

CINF 102 Advanced materials/processes for information technology: Microfluidic synthesis, nanofabrication, and soft lithography

K. Choi, choikm@uci.edu. Chemistry, University of California, Irvine, CA, United States

Nanotechnology has been widely investigated due to specific advantages arise from their small scale in sizes. Chemists/materials scientists have shown intensive challenges on the development of new materials/processes to improve performances of information technologies. For chemist, the overall object of nanotechnology is to carry out the entire operations normally performed in a chemical laboratory including synthesis, processing, mixing, purification and analysis efficiently using small amounts of reagents. For example, microfluidic synthesis has taken a great attention due to advantages, which can't be achieved by conventional wet chemistry. Patterns

fabricated using functional polymers is also promising for applications in information technology to integrate high performance optoelectronic devices. We demonstrate nanofabrication generated using functional polymers. For example, functional polymers can be produced by "molecular imprinting technique", which is a general protocol for "synthetic receptor sites" with specific molecular recognitions. To seek for advanced process, we also develop a new class of PDMS stamps for nano-resolution soft lithography to overcome the limitations and thus to extend current soft lithography to advanced level.

CINF 103 New Knovel interface

S. Gurke, sgurke@knovel.com. Knovel Corp., New York, New York, United States

In the past year Knovel interface has undergone several facelifts and some more serious changes. This paper discusses user centric approach to design and implementation of new GUI. New features, including Knovel Lab, GET Search, autocomplete, spell checker, browse interface and Knovel Math are discussed. Current work and plans for the future enhancements, including expansion of content coverage, are discussed as well.

CINF 104 Expanding and understanding metabolite space

J. E. Peironcely[1,2], peironcely@chem.leidenuniv.nl, A. Bender[3], M. Rojas-Chertó[2,4], T. Reijmers[2,4], L. Coulier[1], and T. Hankemeier[2,4]. [1]Quality of Life, TNO, Zeist, The Netherlands, [2]Department of Netherlands Metabolomics Centre, Leiden University, Leiden, The Netherlands, [3]Department of Medicinal Chemistry, Leiden University, Leiden, The Netherlands, [4]Department of Analytical Biosciences, Leiden University, Leiden, Zuid Holland, The Netherlands

The 'Metabolite Space' is the total chemical universe of metabolites present in all compartments and in all states from any organism. We build models based on discriminative features that predict 'metabolite likeness'. Here, we focus on the human metabolite space, including both endogenous and exogenous metabolites. Metabolites are compared with other molecules representative from different areas of the whole chemical space. We also explore the regions that lay outside the known metabolite space. In order to expand 'Metabolite Space' we employ RECAP to generate new molecules that possess features similar to those present in metabolites. We study how discernible these new molecules are from real metabolites. The new chemistry present in the expanded metabolite space is analyzed by looking at Murcko assemblies, ring systems, and other chemical properties. By combining all the above analyses we expect to characterize better the metabolite space, and to predict the metabolite-likeness of a molecule.

CINF 105 Metabolite identification based on MS fragmentation

M. Rojas-Chertó[1,2], m.rojas@lacdr.leidenuniv.nl, P. T. Kasper[1,2], J. E. Peironcely[1,3], T. Reijmers[1,2], R. J. Vreeken[1,2], and T. Hankemeier[1,2]. [1]Department of Netherlands Metabolomics Centre, Leiden University, Leiden, ZH, The Netherlands, [2]Department of Analytical Biosciences, Leiden University, Leiden, ZH, The Netherlands, [3]Department of Quality of Life, TNO, Zeist, Utrecht, The Netherlands

Identification of metabolites in studies of biological samples can be in principle achieved using high resolution multistage mass spectrometry (MS^n) because it provides a feature rich fingerprint of the precursor structure. However, neither general methodology for the identification nor extensive databases of metabolites with multistage mass spectrometric data are available at the moment. High resolution multi stage MS experiments were performed on LTQ-Orbitrap (Thermo) equipped with Triversa nanoMate (Advion) nanoelectrospray ion source. A defined protocol was established for the reliable and reproducible acquisition of mass spectral trees. An in-house developed software, the Multi-stage Elemental Formula (MEF) tool, integrating among others: Chemistry Development Kit (CDK) and XCMS libraries, was used for spectral data processing. MEF uses a method to resolve the elemental composition of the compound and fragment ions derived from MS^n data using a cyclic constraining process. Resulting fragmentation trees were stored in a database. Our results provide firm basis for developing a generic, multi stage mass spectrometry based platform for efficient identification of metabolites. We also present a powerful tool to discriminate metabolite isomers with the same elemental composition and an only slightly different structure.

CINF 106 Fragment database analysis using molecular shape fingerprints

J. D. MacCuish[1], john.maccuish@mesaac.com, N. E. MacCuish[1], M. Hawrylycz[2], and M. Chapman[1]. [1]Mesa Analytics, Computing, Inc., Santa Fe, New Mexico, United States, [2]Allen Institute for Brain Science, Seattle, Washington, United States

Shape clustering of fragment databases using both 3D shape fingerprints (generated via Quasi-Monte Carlo

integration) and 2D structure fingerprints will be described. Individual clusters are then analyzed with 3D shape fingerprints incorporating substructure information, akin to substructure commonality programs with 2D fingerprints, such as Stigmata and ChemTattoo.

CINF 107 Qsearch: Pharmacophore-based search in fragment spaces

T. Lippert, lippert@zbh.uni-hamburg.de, and M. Rarey. Zentrum für Bioinformatik, Universität Hamburg, Hamburg, Germany

Fragment spaces have proven to be a valuable source of molecules that are biologically active and synthetically feasible. A fragment space consists of a set of molecular fragments with defined linking positions and a set of rules to combine fragments to new molecules. We have developed an expert system for medicinal chemists to allow to search fragment spaces for molecules that can fulfill a chosen three dimensional pharmacophore. The fragment space is searched with an evolutionary approach, where partial solutions evolve to fit the posed query by adding, deleting of replacing fragments. The fitness of a partial solution is calculated by its ability to obey to the constraints of the pharmacophore. We tested the program by searching several focused fragment spaces with pharmacophores for common drug targets. The resulting molecules obey to the input pharmacophore and look chemically sound.

CINF 108 Fragment-based lead discovery through chemotype evolution

D. Erlanson, derlanson@carmot.us. Carmot Therapeutics, Inc., San Francisco, CA, United States

Fragment-based ligand discovery has established itself as a powerful method to identify drug leads. The two primary practical challenges in applying fragment-based methods have been finding fragments and linking (or growing) them. The first challenge has recently become much easier. Advancing fragments to leads, however, remains a significant hurdle. We have developed a technology, Chemotype Evolution, which uses rapid in-situ chemistry to expand a fragment into a diverse range of hits. Chemotype Evolution is amenable to elaborating fragments found using any method: a starting fragment can be a privileged pharmacophore derived from a known inhibitor, substrate, or cofactor, or a fragment identified through a previous screen. Through iterative application of Chemotype Evolution, the starting fragment can be transformed into novel, varied chemotypes, while desired properties can be enhanced by incorporating counter screens. Chemotype Evolution provides a general solution to the question of how to turn promising fragments into promising leads.

CINF 109 Leading fragments to lead structures: Fragment evolution, merging and core replacement, and…docking

C. Detering, detering@biosolveit.de, M. Gastreich, and C. Lemmen. BioSolveIT, Sankt Augustin, Germany

Lead discovery often starts from small fragment binders for which experimental evidence has been found in an active site. Development into a lead structure can involve three possible scenarios: a) to grow from these 'needles' into the depth of the pocket; b) linkage of two or more fragments into one compound with optimized potency; or c) merging two or more fragments in regions of mutual overlap. These tasks can now be accomplished with a novel software tool**,** which comprises the interactive fragment based software ***ReCore*** and the well established docking engine ***FlexX***. With ***ReCore***, synthetically accessible compounds can be generated in seconds by using an indexed 3D fragment library on fragments or compounds that should be altered. The results can then be validated by docking without leaving the software environment. We will elucidate the basic principles and give examples which map onto experimental data and evolve into novel lead ideas.

CINF 110 Fragment-based drug design using PASS approach

O. A. Filz, fioland@yandex.ru, A. A. Lagunin, alexey.lagunin@ibmc.msk.ru, D. A. Filimonov, dmitry.filimonov@ibmc.msk.ru, and **V. V. Poroikov**, vladimir.poroikov@ibmc.msk.ru. Department for Bioinformatics, Institute of Biomedical Chemistry of Rus. Acad. Med. Sci., Moscow, Russian Federation

Fragments are considered as important entities in medicinal chemistry, because they are (1) the determinants of presence/absence of certain biological activity, (2) key elements of structural modifications for obtaining new molecules with the required properties. The role of structural fragments in chemical series is analyzed by experimental methods (X-ray crystallography, NMR, surface plasmon resonance) and by molecular modelling or (Q)SAR. We have developed computer program PASS that predicts more than 3000 biological activities on the basis of structural formula of chemical compounds (http://www.ibmc.msk.ru/PASS; Filimonov and Poroikov, in: Chemoinformatics Approaches to Virtual Screening. A. Varnek and A. Tropsha, Eds. RSC Publishing, 2008, 182-216). Prediction is based on SAR

analysis of the training set containing over 200000 biologically active compounds collected from different sources. PASS calculates the impact of each of atoms in a molecule into a certain activity using MNA descriptors for particular atom and its immediate neighbours. These estimations could be used for identification of fragments responsible for binding chemical compounds with a specific target, and for further computer-aided design or generation of new "candidates" with the required biological activity. The proposed approach: (1) does not require information about 3D structure of the target, (2) could be applied for finding multitargeted ligands, and (3) provides medicinal chemist to make a design *in silico* of synthesizable molecules. Examples of fragment-based design for endothelin receptor antagonists, cyclin-dependent kinase inhibitors, ligands of 5HT and D2 receptors will be discussed.

CINF 111 Use of virtual fragment screening for lead modification

Y. Xu, yongjin.xu@novartis.com, H. Jansen, and E. Martin. Global Discovery Chemistry/Oncology. Exploratory Chemistry, Novartis Institutes for Biomedical Research, Emeryville, CA, United States

In many drug discovery programs, lead structures often need to be modified while maintaining biological activity. With binding site information, modifications can be proposed using two conceptually different approaches, named "cut and fit" and "fit and cut". The "cut and fit" approach, which is under development, starts by fragmenting a compound database and then replacing a portion of the lead molecule with the fragments. The "fit and cut" approach starts with whole molecules from a compound database, evaluates whether any part of these molecules provides a good fit for the region under consideration. Selected fragments are separated and merged with relevant parts of the lead molecule. Finally, the fitting of the new modifications are confirmed with docking method. This method has produced interesting ideas in multiple kinase projects. As a validation of the method, a case study with the P38 ATP pocket and the MDDR database will be described.

CINF 112 IUPAC InChI project: A status report

S. R. Heller, steve@hellers.com, and A. McNaught, mcnaught@ntlworld.com. CBRD, MS-8320, NIST, Gaithersburg, MD, United States

The current status of the IUPAC InChI project will be described. The relationship with and the activities of the related InChI Trust will be presented and show how the InChI and InChIKey are being used in scientific publications and scientific databases to enhance the linking and connecting of chemical information and data from a variety of sources and resources. The cooperation between IUPAC, the major organizations in the scientific publishing community, and chemical software companies will be presented to show how this widely diverse group of national and international organizations are collaborating.

CINF 113 Tools for the scholarly communication lifecycle

A. D. Wade, alex.wade@microsoft.com, and L. Dirks. Microsoft Research, Redmond, WA, United States

Microsoft External Research strongly supports the process of research and its role in the innovation ecosystem. We are dedicated to supporting research efforts and to developing new tools to enhance the broadest goals of the scholarly communication lifecycle. Through collaborations with universities, national libraries, publishers, and governmental organizations, we seek to develop software and services that will advance the process of scholarly information capture and dissemination. Such collaborations demonstrate our ongoing work towards producing next-generation documents that increase productivity and empower authors to increase the discoverability and appropriate re-use of their work. Throughout the research process, software can and should assist us in the process of information capture, organization, analysis, collaboration, authoring, dissemination and long-term information preservation. This session will highlight several freely available and open source efforts from Microsoft External Research, and will demonstrate how these can help to enhance and evolve the process of scholarly communication.

CINF 114 Integrated language for chemical publication

P. Murray-Rust, pm286@cam.ac.uk, J. A. Townsend, and L. Hawizy. Department of Chemistry, University of Cambridge, Cambridge, Greater East Anglia within England, United Kingdom

Chemical communication involves many languages: text, speech, graphics, numerical information, chemical structures and computations. Many concepts are confined to one sublanguage ("lower alkyls" is linguistic , "dotted bond" is graphical and "1.3kcal" is computational). Normal publications reduce these to a single medium such as PDF, computer logs, or audio but this introduces loss and semantic ambiguity. We describe the systematisation and integration of chemical

sublanguages through markup, ontologies and knowledgebases. The combination of information (graphical, linguistic) can resolve ambiguity and leads to an adaptive machine-learning process. We have developed tools to support authoring of semantic documents (datuments) and to read current chemical documents (text, graphics, computation) into this infrastructure. Humans and machines can now communicate in a natural manner (speech, text, gestures) leading to a simple level of articficial intelligence.We have a complete Open toolchain and propose this approach for formal scholarly communication.

CINF 115 Chem4Word

J. A. Townsend[1], jat45@cam.ac.uk, P. Muray-Rust[1], J. Downing[1], T. Haughton[2], A. Wade[3], and L. Dirks[3]. [1]Department of Chemistry, University of Cambridge, Cambridge, CAMBS, United Kingdom, [2]Microsoft Research, Cambridge, CAMBS, United Kingdom, [3]Microsoft Research, Redmond, Washington, United States

Chem4Word is an Open Source Add-in for Word2007 developed in a collaborative project (Cambridge/Microsoft) for enhancing scholarly communication in chemistry. Version 1 supports a complete datument (text, formulae, chemical structures, navigation and indexing) and is adaptive, learning from the author's previous actions and updates through the web. All components are semantic so that chemical structures or other graphics cannot be misinterpreted by machines or humans. Multi-interface tools (speech, gestures, PDAs, etc.) are supported. The system captures the complete semantic history of chemistry in the document and Chemical Styles allow consistent look-and-feel to be applied to the entire document. Version 2 is being developed to support scholarly communication to and from undergraduates and introduces peaklists, spectral data and reactions. C4W manages many of the components in a mainstream chemical publication and our Open governance model and deliberately extensible add-in architecture invites those interested to develop community components for Open re-use.

CINF 116 Capturing and publishing chemistry from the bench to paper: Does Google Wave offer a solution?

C. Neylon, cameron.neylon@stfc.ac.uk. ISIS, Science and Technology Facilities Council, Didcot, OXON, United Kingdom

Google Wave is a new communication and document management platform that takes advantage of advances in web technology to provide a wide range of functionality to the user in an integrated fashion. Documents or "waves" can be shared in real time between multiple participants, which can include automated web-based agents ("Robots") that can read, parse, and modify the wave. Rich visualisation and control functionality can be added javascript "gadgets" that might include three dimensional chemical models, instrument control, or data visualisation. The combinaton of these features offers an enormous potential to deliver a rich chemical authoring experience including automated and user-guided markup of a chemical record as well as publishing that record in different forms to the web. This talk will discuss progress on developing tools within the Wave framework for chemical authoring and publishing and the prospects for using Wave as an integrating technology for chemistry on the web.

CINF 117 CAS REGISTRY: Maintaining quality standards as scientific discovery accelerates

M. J. Toussant, mtoussant@cas.org. Chemical Abstracts Service (CAS) Division, American Chemical Society, Columbus, OH, United States

CAS REGISTRY is the world's most comprehensive and fastest growing compendium of publicly disclosed chemical substance information, comprising more than 50 million organic and inorganic substances (as of 9/09) and 61 million sequences. In recent years, the pace of discovery of new substances has accelerated dramatically and further growth is anticipated. This talk will focus on the strategies, policies, technology and organizational approaches employed by CAS to meet its mission (to organize and make available for research publicly disclosed chemical information) in light of this growth, and in particular to ensure the quality of the substance information in REGISTRY.

CINF 118 Chempedia: A social medium for chemical information

R. L. Apodaca, rapodaca@metamolecular.com. Metamolecular, LLC, La Jolla, CA, United States

Chempedia (http://chempedia.com) is a free service designed to solve the problem of uniquely identifying and naming chemical substances. With the Web's ever-increasing role in chemistry comes the need for a free substance registry as a platform on which to build the chemical information systems of the future. Chempedia addresses this need by offering a peer-reviewed, real-time platform that can be read from, written to, and republished by the global chemistry community free of charge. Created with the recognition that chemistry is an inherently social activity, Chempedia adapts many of the best practices and technologies used in modern social media. This talk will discuss the importance of a free substance registry to the future of chemical comminication, the principles and technologies applied to Chempedia, and the ways Chempedia is now being used.

CINF 119 Enhancing discoverability across Royal Society of Chemistry content by integrating to ChemSpider, an online database of chemical structures

A. J. Williams, antony.williams@chemspider.com, V. Tkachenko, S. Shevelev, and R. Kidd. Royal Society of Chemistry, Wake Forest, NC, United States

The ability to query across a chemistry publishers content using chemical structure searching can dramatically enhance discoverability. RSC has been applying a number of procedures to integrate RSC's ChemSpider community resource with our published content and databases. These include: 1) entity extraction procedures 2) chemical name conversion procedures using software algorithms and curated dictionaries 3) semantic markup and 4) a crowdsourced curation processes. This presentation will provide an overview of the processes we have utilized in order to provide structure-based integration to RSC content. We will discuss our ongoing efforts to extend the approaches to the mining of data from the rich supplementary information sections of many RSC publications. Our intention is to provide access to synthesis procedures and analytical data and further enrich the ChemSpider database for the benefit of the chemistry community.

CINF 120 Fragment-based screening of stabilized G protein-coupled receptors

M. Congreve, miles.congreve@heptares.com. Chemistry, Heptares Therapeutics Ltd, Welwyn Garden City, Hertfordshire, United Kingdom

GPCRs as a target class are notoriously difficult proteins to work with because conformational flexibility, heterogeneity and instability outside of the cell membrane makes them particularly hard to crystallize and limits their use in screening applications. Heptares is developing a technology that expedites the study of GPCRs by dramatically stabilizing these important receptors outside of the cell membrane. The new stabilized receptors (StaRs) are much more robust than the corresponding wild type proteins making them amenable to crystallography and to biophysical screening techniques. Screening results from both TINS-NMR and Surface Plasmon Resonance (SPR) binding studies for representative StaR proteins using libraries of fragments will be outlined. Computational approaches to interpret the screening results will be discussed, and the potential for in-silco design using fragments as start points for lead identification presented.

CINF 121 Novel histamine GPCR family antagonists by fragment screening and molecular modeling

R. J. Law[1], richard.law@evotec.com, T. Hesterkamp[2], A. Kahrs[2], M. Whittaker[3], S. Pal[1], and A. Heifetz[1]. [1]Computational Chemistry, Evotec, Abingdon, Oxfordshire, United Kingdom, [2]Evotec AG, Hamburg, Germany, [3]Evotec UK, Abingdon, United Kingdom

We have applied a small fragment collection to the screening of three histamine receptors with the goal to identify subtype specific antagonists. The rationale for doing this is to cover as much chemical diversity as possible in a small screening effort and to meet the chemical space of the natural ligand, histamine. We have followed up on the resulting fragment hits by building H3 and H4 receptor models based on similarity to known GPCR crystal structures andoptimized using a series of molecular dynamics procedures. These models were used for a docking procedure to reveal the bioactive conformation of the bound ligands, with a view to structure-guided fragment-to-lead expansion. A subsequent shape-based analogue search provided a short list of hits from which we have been able to obtain novel submicromolar and lead-like H3 and H4 antagonists.

CINF 122 Mitotic kinesin Eg5 inhibitors generation by computational MED-portion based drug design at PDB scale

K. Oguievetskaia[1], L. Martin-Chanas[1], lmartin@medit.fr, A. Vorotyntsev[1], O. Doppelt-Azeroual[1], X. Brotel[1], S. Adcock[1], A. de Brevern[2], **F. Delfaud**[1], fdelfaud@medit.fr, and **F. Moriaud**[1], fmoriaud@medit.fr. [1]MEDIT SA, Palaiseau, Idf, France, [2]Equipe DSIMB, Dynamique des Structures et Interactions des Macromolécules Biologiques, INSERM UMR-S 665, Paris, Idf, France

Eg5, a mitotic kinesin exclusively involved in the formation and function of the mitotic spindle has attracted interest as an anticancer drug target. Eg5 is co-crystallized with several inhibitors bound to its allosteric binding pocket. Each of these occupies a pocket formed by loop5/helix 2. Recently designed inhibitors additionally occupy a hydrophobic pocket of this site. The goal of the present study was to identify new fragments which fill this hydrophobic pocket and might be interesting chemical moieties to design new inhibitors.

CINF 123 Starting small and staying small: Fragment-based lead discovery in CNS disease

V. L. Nienaber, vicki@zenobiatherapeutics.com. Drug Discovery, Zenobia Therapeutics, La Jolla, CA, United States

Diseases of the central nervous system (CNS) are among the most devastating to patients and their families. Despite this, treatments have lagged behind other therapeutic areas in part because of the challenge for compounds that cross the blood brain barrier. Recent analyses of successful drugs have shown that their chemical properties have not changed substantially over the past 40 years while the properties of compounds entering the clinic have become inflated. This property inflation has only exacerbated the challenges of CNS drug discovery as the requirements for delivery to the brain are even more stringent than those for other tissues. In this presentation, we will discuss the merits of fragment based lead discovery and how we have adapted our approach to address the chemical property challenges of CNS drug discovery. We include results from our internal programs targeting protein kinases and other enzyme classes in neurodegenerative diseases.

CINF 124 Finding druggable sites in protein-protein interfaces by computational fragment mapping

D. Kozakov, midas@bu.edu, and S. Vajda. Department of Biomedical Engineering, Boston University, Boston, MA, United States

Development of small-molecule inhibitors that disrupt protein–protein interfaces has received considerable attention in recent years. We describe a method based on computational fragment mapping to identify “hot spot” regions in protein-protein interfaces that are most likely capable of binding small molecules. The method places molecular probes – small molecules or functional groups – on the protein surface, and finds and clusters the energetically favorable sites for a variety of probes. Application of mapping to a number of PPI targets has shown that the method always finds at least some part of the appropriate sites. To account for protein plasticity we then use an algorithm combining statistical analysis and energy minimization to identify the nearby “moveable” side chains, and generate and then map the alternative protein structures. Results are presented for a number of protein-protein interaction targets, including interleukin-2, Bcl-xL, MDM2, HPV-11 E2, ZipA, TNF-alpha, and NEMO.

CINF 125 Peer review is not perfect but are the alternatives worse?

W. A. Warr, wendy@warr.com. Wendy Warr & Associates, Holmes Chapel, Cheshire, United Kingdom

Scientists are generally agreed that publication of learned articles should not be a free-for-all, and that some form of quality control is necessary. On the other hand anyone who has studied peer review even cursorily has to admit that the traditional system is far from perfect. It succeeds as a compromise for the quality journals because of the way it is administered by journal editors. Some of the alternatives that have been proposed, such as double blind peer review or open peer review are simply modifications of the traditional process. Other options (usually in addition to the traditional method) are post-publication review and services such as Faculty of 1000. More progressive are concepts such as computer-assisted peer review; reputation; moderated peer review and systems such as Roy Rustum’s “super peer review”; volunteer review, voting and systematic aggregation of reviews; and use of crowdsourcing and the blogosphere. This author (a journal editor) is skeptical about some of the alternatives, but nevertheless, the presentation will concentrate more on revolution than on evolution.

CINF 126 Interactive open access publishing and public peer review: Perspectives and effectiveness of transparency and self-regulation in scientific communication and quality assurance

U. Pöschl, poeschl@mpch-mainzh.mpg.de. Biogeochemistry Department, Max Planck Institute for Chemistry, Mainz, Germany

The traditional forms of scientific publishing and peer review do not live up to the demands of efficient communication and quality assurance in today's highly diverse and rapidly evolving world of science. They need to be advanced and complemented by interactive and transparent forms of review, publication, and discussion that are open to the scientific community and to the public. The advantages of open access, public peer review and interactive discussion can be efficiently and flexibly combined with the strengths of traditional publishing and peer review. Since 2001 the benefits and viability of this approach are clearly demonstrated by the highly successful interactive open access journal Atmospheric Chemistry and Physics (ACP, www.atmos-chem-phys.net) and a growing number of sister journals launched by the publisher Copernicus (www.copernicus.org) and the European Geosciences Union (EGU, www.egu.eu). These journals are practicing a two-stage process of publication and peer review combined with interactive public discussion, which effectively resolves the dilemma between rapid scientific exchange and thorough quality assurance. The same or similar concepts have recently also been adopted in other disciplines, including the life sciences and economics. Note, however, that alternative approaches where interactive commenting and public discussion are not fully integrated with formal peer review by designated referees tend to be less successful. The principles, key aspects and achievements of interactive open access publishing (top quality. impact, efficient self-regulation . low rejection rates, little waste. low cost) will be presented and discussed. Moreover, perspectives for future developments in scientific communication and and quality assurance will be oulined. For more information see: www.atmospheric-chemistry-and-physics.net/general_information/public_relations.html

CINF 127 Impact factors, post-publication peer review and other metrics

R. P. Grant, richard.grant@f1000.com. Faculty of 1000, London, United Kingdom

The ISI Impact Factor (IF) is the de facto standard for assessing 'quality' and importance of published research. It suffers many problems: it is slow, at least two years out of date; it depends on accurate and relevant citation; it is unduly skewed by review articles; it cannot reliably predict quality; it is irrelevant to non-citable research or non-citing communities; and important research does get published in low IF journals. Alternative indicators include usage data, network analysis and comment aggregation. We have developed a metric that harnesses expert review and the power of the social web, and that can furthermore predict citation rates. We have used this to rank journals and institutions according to research quality rather than IF. We also propose that it is a more relevant measure of an individual scientist's output than one based on citation rates.

CINF 128 Eigenfactor: Ranking and mapping the scholarly literature

J. D. West, jevinw@u.washington.edu. Department of Biology, University of Washington, Seattle, WA, United States

Each year, tens of thousands of scholarly journals publish hundreds of thousands of scholarly papers, collectively containing tens of millions of citations. As De Solla Price recognized in 1965, these citations form a vast network linking up the collective research output of the scholarly community. In our lab, we have developed a set of tools to mine the wealth of information contained within the full structure of these citation networks. In this talk, I will give an overview of the Eigenfactor project and how we are using centrality approaches to better evaluate and navigate the ever-expanding, scholarly literature.

CINF 129 Applying domain expertise assessment to compound annotation for enhanced collaborations between chemists and biologists

Y. Pouliot, ypouliot@researchscorecard.com. ResearchScorecard Inc., Belmont, California, United States

Medicinal chemistry and other fields require close collaboration between chemists and biologists. Because of their cross-disciplinary nature, the success of these collaborations is particularly dependent on making an optimal choice of collaborators. Unfortunately, finding and selecting potential collaborators remains a haphazard process. ResearchScorecard addresses this issue by applying data mining algorithms to comprehensive descriptors of academic researchers. By facilitating the assessment of the quality and volume of research data conveyed by databases, we believe that the combination of such "expertise data" with compound properties can enhance researchers'

understanding of these compounds. This is enabled by ResearchScorecard's use of data gathered automatically from public sources, followed by heavy cleansing and semantic integration. Such an objective approach enables the approximate ranking of researchers based on their degree of domain expertise and their overall prominence. Some of the applications of automated domain expertise assessment to compound annotation will be discussed.

CINF 130 Exploring molecular networks for polypharmacology analysis

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

In silico prediction of molecular polypharmacology is of great importance as it helps to design safer drugs with higher efficacy. We recently have embarked on the construction of chemical and biological databases with annotation of molecular biological functions and 2D/3D structures. The information was then employed to analyze the relationships among drugs and targets based on drug chemical similarity search and their interactions with multiple targets. The derived molecular networks were visualized in 2D graphs, and the prediction of potential drug off-target properties and toxicities was conducted. Based on known biological and pharmacological functions, about 3933 out of 5059 drugs were found interacting with multiple targets. On the other hand, chemical similarity search demonstrated that 2528 drugs share >85% structure similarities with some other drugs, indicating these drugs might have similar bioactivities. With this implementation, we provided an efficient way to evaluate the polypharmacology of chemical compounds.

CINF 131 Picking novel actives from large databases using a combined ranking method

D. Marcus, david.marcus1@mail.huji.ac.il, D. Barasch, A. Rayan, and A. Goldblum. Institute of Drug Research, The Hebrew University of Jerusalem, Jerusalem, Israel

Early stages of modern drug discovery often involve screening of small molecular libraries to find "hits". In order to construct highly enriched libraries, we developed a ranking method that combines predictions from several models to supply an index for each compound and swiftly rank large compound databases. The method is based on our Iterative Stochastic Elimination algorithm that can locate the best model sets out of a huge number of possible solutions. Models are assigned to distinguish between multiple activity classes (inactive, modestly active, highly active) and taken together with different weights for each model and class, a combined ranking index is produced. We found that this index has a higher reliability and correlativity to experimental values than the common approach based solely on classification of two activities. This method was applied successfully in a few test cases (Estrogen receptor, Acetylcholine Esterase) and is able to pick highly enriched libraries.

CINF 132 Human proteome in the context of pathways, diseases, drugs, and tool compounds

F. Nigsch, florian.nigsch@novartis.com, and J. Jenkins. Center for Proteomic Chemistry, Lead Discovery Informatics, Novartis Institutes for Biomedical Research, Cambridge, Massachusetts, United States

To get an overview of chemical matter currently known to be biologically active, we accumulated several sources of compound-target activity information into a repository comprising 20 million activity records of 3 million distinct compounds. Extensive target normalization and the use of a commercial database of curated pathways, regulatory network, disease and other relevant information for drug discovery (GeneGo Metabase) was used to place compound-target associations into a rich biological context. Analysis of this wealth of data reveals pharmacological hotspots and also provides information on the global coverage of the human proteome. Apart from the amount of data, however, a prime focus was put on target annotation in an effort to be able to relate the effects of compounds to the biological networks they act upon. The workflows that are made possible in this way span methods from both bioinformatics and cheminformatics and provide novel methodology to enable drug discovery research.

CINF 133 Atom type preferences, constitutional diversity, and property profiles of known drugs: A knowledge-intensive, comparative assessment of drugability

V. N. Viswanadhan, vellarkad_viswanadhan@jubilantbiosys.com, H. Rajesh, and V. N. Balaji. Department of Computational Chemistry, Jubilant Biosys Ltd, Bangalore, Karnataka, India

A novel characterization of known drugs and an improved drugability score are presented based on an analysis of physicochemical and topological properties at different levels: atom types, functional groups, and the whole molecule. A new united atom classification, UALOGP, is proposed and validated, extending the

earlier ALOGP method. UALOGP was used to analyze atom type distributions and identify atom types preferred in drugs, relative to a non-drug database, with a similar MW distribution. Using a new quantification of atom type diversity, it is shown that drug molecules are intrinsically more *diverse*. A critical re-examination of well-known characterizations of drugability, such as the Ro5 and GVW criteria, is accomplished based on an analysis of physicochemical property distributions. A comparative analysis of the present and older drug property ranges leads to an improved drugability filter and scores, correctly identifying the chemical space occupied by orally bioavailable drugs.

CINF 134 Improving the consistency of data fusion in virtual screening

I. Mott[1], P. Gedeck[2], and **V. J. Gillet**[1], v.gillet@sheffield.ac.uk. [1]Department of Information Studies, University of Sheffield, Sheffield, United Kingdom, [2]Novartis Institutes for BioMedical Research, CH-4002 Basel, Switzerland

It is well known that different virtual screening methods capture different aspects of structure-activity relationships and this has led to the use of data fusion or consensus scoring methods to combine the output from different methods. For example, it has been shown that enrichments in docking can be improved by combining the results obtained with different scoring functions. Data fusion involves combining the ranks produced by the individual methods typically using either the minimum rank (the MIN rule) or the average rank (the SUM rule). The relative abilities of these two rules to improve virtual screening results depend on the distributions of pairwise similarity values in the dataset which is difficult to predict upfront. Here we investigate the use of Pareto ranking for data fusion and show that greater consistency can be achieved over different datasets compared to the use of either the MIN or the SUM rule.

CINF 135 Extending interaction fingerprints: A novel approach to characterizing protein binding sites

C. Higgs[1], chris.higgs@schrodinger.com, W. Sherman[1], and J. Blaney[2]. [1]Department of Application Science, Schrodinger, New York, NY, United States, [2]Department of Discovery Chemistry, Genentech, South San Francisco, California, United States

The characterization of the protein binding sites based on protein-ligand interactions can be valuable in structure-based drug design and successful applications using interaction fingerprints have been published. However, most methods are limited to characterizing binding sites of the same or related proteins due to the fingerprints being dependent on a consistent numbering of the atoms in the binding site. Here we present a new type of interaction fingerprint based on geometric profiling that is independent of the atom/residue numbering and can be compared between any targets. In addition to using this fingerprint method as a tool to post-process docking results, the fingerprints can be used to find other complexes with similar protein-ligand interaction motifs, cluster complexes based on interaction profiles, and perform substructure searches to find regions of the binding site that may be similar to unrelated proteins.

CINF 136 Molecular signatures of promiscuous enzyme-substrate interactions

J.-L. Faulon, jfaulon@gmail.com, and P. Carbonell, pcarbonell@epigenomique.genopole.fr. Department of Biology, University of Evry, Evry, France

Catalytic promiscuity, the ability of enzymes to catalyze more than one reaction, has been used to engineer enzymes with altered activities. However, the plasticity mechanism that makes a promiscuous catalytic active site able to accommodate different substrates is at present not fully understood. In this work, we use a graph-based representation known as molecular signatures to characterize both sides of the promiscuous enzyme-substrate binding process. In particular, we dissect the structural determinants of promiscuity at catalytic sites, and how chemical similarity is involved in substrate recognition. Performing this study for the entire KEGG database allows us to get insights about evolutionary relationships between different metabolic pathways, and how new catalytic functions are acquired. We show here an enzyme engineering application of this method to the directed evolution of natural occurring protein scaffolds with latent catalytic activities.

CINF 137 Metabolic regioselectivity models for nine CYP P450 isozymes

J. M. Zaretzki[1], zaretj@gmail.com, T.-W. Huang[1], C. M. Breneman[1], C. Bergeron[2], and K. P. Bennett[2]. [1]Department of Chemistry, Rensselaer Polytechnic Institute, Troy, NY, United States, [2]Department of Mathematics, Rensselaer Polytechnic Institute, Troy, NY, United States

Computational models for rapidly predicting the susceptibility of sites on small drug-like molecules to oxidative metabolism have been extended to include cytochrome isozymes 1A2, 2A6, 2B6, 2C19, 2C8, 2C9,

2D6, 2E1, as well as 3A4. Within this methodology, topologically distinct regions of each ligand are quantified using 540 2D and QM-based electronic descriptors, and ranked according to their susceptibility as metabolic sites. A quick and accurate tool for making these predictions on substrate databases is now available online.

CINF 138 Novel topological molecular key for cheminformatics

P. Liu, pliu24@its.jnj.com, and D. Agrafiotis. Johnson Johnson PRD, Exton, PA, United States

Representing molecules in the form of numerical or symbolic keys is increasingly employed to capture the structural characteristics and chemical properties in chemoinformatics research. Here we present a novel topological molecular key to encode the connecting atoms and bonds into two separate components, which allows the convenient comparison of each individual component. We have applied this new key for the screening stage of substructure search on a relational database for a dataset of about 1 million compounds. The screening efficiency with this new key is greater than 99.9%. More importantly, this set of new keys can uniquely determine the correct hits for certain generic queries in a fraction of seconds, which is usually a challenging task for other substructure search methods. It also has potential in similarity search and clustering.

CINF 139 Classification of enzyme function based on similarities in reaction mechanisms and common substrate substructures

D. E. Almonacid, daniel.almonacid@ucsf.edu, and P. C. Babbitt. Departments of Bioengineering and Therapeutic Sciences, and Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences, University of California San Francisco, San Francisco, California, United States

Historically proteins have been studied and classified in terms of their sequence and structure, and then often independently associated with function. Sophisticated mathematical methods have been developed to measure similarities between protein sequences and structures, generating robust classification schemes. In contrast, for protein function, most, if not all, classification systems are based on qualitative conceptual frameworks rather than on quantitative measures. Here, we classify enzyme molecular functions based on two quantitative measures: (1) similarity of enzyme reaction mechanisms and (2) common substrate substructures between reactions. We apply this new classification to families of evolutionarily related enzymes in our Structure-Function Linkage Database. The results indicate that similarities in reaction mechanisms and substrate substructures are orthogonal and thus complementary to classifications based on sequence and structure. We discuss the value of quantification of function similarity for functional prediction, annotation and engineering of enzyme functions.

CINF 140 Chem_BLAST: A rule-based method to develop advanced structural ontologies for chemical bioinformatics and the PDB, the PubChem

T. N. Bhat, bhat@nist.gov. CSTL, NIST, Gaithersburg, MD, United States

Today's Chemical Bioinformatics community must interact with a variety of information standalone applications and ontologies. This limitation promotes the need to define and develop rule-based stringent ontologies for information processing and sharing. Chemical Block Layered Alignment of Substructure Technique (Chem-BLAST) first recursively dissects chemical structures into blocks of substructures using rules that operate on atomic connectivity and then aligns them one against another to develop first Chemical Resource Description Framework (RDF) and then chemical ontologies in the form of a 'tree' made up of 'hub-and-spoke'. The technique was applied for (a) both 2-D and 3-D structural data for AIDS (http://bioinfo.nist.gov/SemanticWeb_pr2d/chemblast.do); (b) to □ ;60000 structures from the PDB which is now available from the RCSB/PDB Web site (http://www.rcsb.org/pdb/explore/externalReferences.do?structureId=3GGT) and advanced features at http://xpdb.nist.gov/chemblast/pdb.html . Full description of the Chem_BLAST along with recent results and illustrations including those for approximately a million compounds from the PDB and PubChem will be presented.

CINF 141 Chemical entity extraction and interpretation

D. M. Lowe[1], dl387@cam.ac.uk, P. T. Corbett[2], P. Murray-Rust[1], and R. C. Glen[1]. [1]Chemistry, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom, [2]Linguamatics, Cambridge, Cambridgeshire, United Kingdom

OSCAR is an extensible Open-source system for chemical entity recognition in text which has recently been re-factored through UK eScience (OMII). Using pattern based and machine learning techniques (with interchangeable tools) OSCAR recognises chemical compounds, reactions, enzyme names and other chemical terms. To create structure searchable corpora

and to enhance viewing of documents chemical names can be resolved to structures. This is achieved either through lookup (ChEBI, Pubchem) or complete parsing of systematic names using the OPSIN module. Using a regular grammar, several chemical dialects, and a largely complete list of IUPAC name components OPSIN employs finite-state parsing to recognise IUPAC morphemes and disambiguate their meaning. A connection table, including atom numbering, is then built and output as CML (Chemical Markup Language) and hence other formats such as SMILES. Additions to OPSIN's vocabulary can be made without code modification. Metrics showing OPSIN's current performance on chemical names will be presented.

CINF 142 PolyGWT: A Google Web Toolkit application to assist polymer research

E. O. Cannon, eoc21@cam.ac.uk, and P. Murray-Rust. Department of Chemistry, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom

As we move into the 21st century, demand for open source software for polymer scientists has never been greater. We introduce PolyGWT, an open source state of the art three-tier architecture semantic web application for polymer science, based on the Google Web Toolkit (GWT) software framework, Restlet and an Apache Derby relational database. The application addresses a number of issues: building and displaying polymers, searching for repeat units by substructure and by property using ontology browsing. It also allows repeat unit property calculation on the fly and a suite of machine learning tools for classification and clustering of new repeat units. We expect this application to be of primary use to polymer chemists wishing to develop a strategy for a laboratory synthesis. It will also be of interest to software engineers to develop or extend, as an application programming interface is provided.

CINF 143 OpenTox: An open source predictive toxicology software framework

B. Hardy[1], barry.hardy@douglasconnect.com, **D. A. Gallagher**[2], gallagher.da@gmail.com, and S. Chawla[3], sunil@seascapelearning.com. [1]Douglas Connect, Switzerland, [2]CAChe Research LLC, United States, [3]Seascape Learning, India

Recent (Q)SAR developments allow a much more accurate prediction of complex toxicological endpoints than a few years ago. This progress has been caused by (i) the development of improved (Q)SAR algorithms and (ii) by the availability of larger and better curated public databases. The routine application of these new generation models is however still rare, because of incomplete, incompatible toxicity databases and QSAR models. The European Commission has hence seed funded OpenTox (www.opentox.org) under the FP7 RD Framework that provides a unified access to toxicity data, (Q)SAR models; and procedures supporting validation and reporting requirements of REACH. Now, at the halfway stage of the initial three-year project, the technical developments including architecture design, interfaces, ontologies, and approach to web services will be presented and discussed. The experiences with developing initial software prototypes based on the OpenTox approach will be reviewed.

CINF 144 Combining disparate cheminformatics resources into a single toolkit: The Cinfony library

N. M. O'Boyle[1], n.oboyle@ucc.ie, and G. R. Hutchison[2]. [1]School of Pharmacy, University College Cork, Cork, Co. Cork, Ireland, [2]Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania, United States

Open Source cheminformatics toolkits such as OpenBabel, the CDK and the RDKit share the same core functionality but complement each other in their support for different file formats and forcefields, and molecular fingerprints and descriptors. However using these toolkits in the same program is difficult as they are implemented in different languages (C++ versus Java), have different underlying chemical models and have different application programming interfaces (APIs). We describe Cinfony, a Python module that presents a common interface to all three of these toolkits, allowing the user to easily combine methods and results from any of the toolkits.

CINF 145 Role of Rh(III) as negative catalyst in N-chlorosuccinimide oxidation of glycine in alkaline medium: A kinetic and mechanistic study

A. K. Singh, ashokeks@rediffmail.com, M. Singh, dr.manjulasngh@rediffmail.com, R. Srivastava, rashmisri11@rediffmail.com, J. Srivastava, sri_jayasri@rediffmail.com, and S. Rahmani, shahlaau@rediffmail.com. Chemistry, University of Allahabad, Allahabad, Uttar Pradesh, India

Kinetics and mechanism of N-chlorosuccinimide (NCS) oxidation of glycine in aqueous alkaline medium using Rh(III) as negative catalyst have been studied at 35^0C. The observed kinetic data suggest that order with respect to [NCS] is unity throughout its ten-fold variation. Making basis to pseudo-first-order rate constant (k_1) values for varying concentration of glycine and OH^- ions, it is concluded that the first -

order kinetics observed at low concentration of each reactant changes to zero-order at its higher concentrations. The significant feature of the reaction under investigation is that there is inverse proportionality between pseudo-first-order rate constant (k_1) and [Rh(III)]. Inverse fractional order in [NHS] and nil effect in [Cl^-] were observed throughout their variations. The reaction rates remain unaffected by the change in ionic strength (m) and dielectric constant (D) of the medium. The reaction has been studied at four different temperatures and with the help of observed k_1 values, various activation parameters including the entropy of activation ($\Delta S^{\#}$) were calculated. Spectral studies for the verification of the formation of complex species during the course of reaction have been made by the help of Varian Cary 300 Bio UV-Vis spectrophotometer. Suitable mechanism, consistent with the observed kinetic data together with reactive species of NCS, Rh(III) and glycine in alkaline medium and spectrophotometric evidence for the formation of complex species, has been proposed for oxidation of glycine by NCS in presence of Rh(III) as negative catalyst.

CINF 146 Psychological barriers to good decision-making: How addressing cognitive biases could improve drug discovery

M. D. Segall[1], matt.segall@optibrium.com, and A. Chadwick[2]. [1]Optibrium, Cambridge, Cambridgeshire, United Kingdom, [2]Tessella, Burton upon Trent, Staffordshire, United Kingdom

Better individual and team decision-making could enhance R̤D performance. Reproducible biases effecting human decision making, known as cognitive biases, have been understood by psychologists for at least half a century. These threaten objectivity and balance and so are credible causes for continuing unpleasant surprises in late development and high operating costs of compound discovery. For four of the most common and insidious cognitive biases we will consider the risks to R̤D decision-making. We will suggest approaches for overcoming these, such as strategies adapted from evidence-based medicine and computational tools that seek to guide the decision making process, encouraging objective consideration of all of the available information and explicit consideration of the impact of uncertainty in drug discovery.

CINF 147 NCI/CADD: Open-access chemical structure web platform

M. Sitzmann[1], sitzmann@helix.nih.gov, W.-D. Ihlenfeldt[2], and M. C. Nicklaus[1]. [1]Chemical Biology Laboratory, Center for Cancer Research, National Cancer Institute, National Institutes of Health, DHHS, Frederick, MD, United States, [2]Xemistry GmbH, Lahntal, Germany

We present an update on the Chemical Structure Lookup Service, the Chemical Identifier Resolver and further services available at the web server (http://cactus.nci.nih.gov) of the NCI/CADD Group. They all have been implemented on the basis our new open-access chemical structure web platform. We will give an overview on how this resources can be integrated and linked in into one's own web applications. The chemical structure web platform also includes open access to the underlying chemical structure database which currently indexes approx. 103 million chemical structure records representing about approx. 70 million unique chemical structures.

CINF 148 OrbDB: A database of molecular orbital interactions

M. A. Kayala, mkayala@ics.uci.edu, C. A. Azencott, J. H. Chen, and P. F. Baldi. Department of Computer Science, University of California - Irvine, Irvine, CA, United States

The ability to anticipate the course of a reaction is essential to the practice of chemistry. This aptitude relies on the understanding of elementary mechanistic steps, which can be described as the interaction of filled and unfilled molecular orbitals. Here, we create a database of mechanistic steps from previous work on a rule-based expert system (ReactionExplorer). We derive 21,000 priority ordered favorable elementary steps for 7800 distinct reactants or intermediates. All other filled to unfilled molecular orbital interactions yield 106 million unfavorable elementary steps. To predict the course of reactions, one must recover the relative priority of these elementary steps. Initial cross-validated results for a neural network on several stratified samples indicate we are able to retrieve this ordering with a precision of 98.9%. The quality of our database makes it an invaluable resource for the prediction of elementary reactions, and therefore of full chemical processes.

CINF 149 Application of crowdsourcing for metadata curation of digitized texts

A. J. Williams, antony.williams@chemspider.com, R. Kidd, V. Tkachenko, and S. Shevelev. ChemSpider, Royal Society of Chemistry, Wake Forest, NC, United States

The Royal Society of Chemistry has tens of thousands of digitized texts in their archive representing over a hundred years of chemistry research. ChemSpider is the RSC's internet based community resource of structure-based chemistry, serving resources including chemical compounds, synthesis procedures and associated meta-data. We have initiated a project to facilitate structure searching of the RSC archive by integrating the chemical structures associated with the RSC publications into the ChemSpider database. This has involved chemical name extraction and conversion from the publication and manual validation of the resulting data. The resulting data collection demands a crowdsourced approach for data curation. This presentation will discuss the progress of our project, will provide an overview of the associated workflows and provide an overview of how to utilize crowdsourcing for the purpose of curating the metadata associated with digitized texts.

CINF 150 Five years of collaborative drug discovery in the cloud

B. A. Bunin, bbunin@collaborativedrug.com, S. Ekins, sekins@collaborativedrug.com, M. Hohman, moses@collaborativedrug.com, S. Ernst, sylvia@collaborativedrug.com, and K. Gregory, kgregory@collaborativedrug.com. Collaborative Drug Discovery (CDD), Burlingame, CA, United States

Collaborative Drug Discovery hosts a widely used drug discovery data cloud platform. The CDD Vault is a secure, private industrial-strength database combining traditional drug discovery informatics (registration and SAR) with social networking capabilities. CDD Collaborate enables real-time collaboration by securely exchanging selected confidential data. Traditional drug discovery capabilities include the ability to import/export to Excel™ and sdfiles, Boolean queries for potency, selectively, and therapeutic windows for small molecule enzyme, cell, and animal data, substructure and Tanimoto similarity search, physical chemical property search, as well as IC50 calculation/curve generation, heat-maps, and Z/Z' statistics for archived data (protocols, molecules, plates, hyperlinked files). CDD Public has unique, constantly growing drug discovery SAR content. The CDD Vault, Collaborate, and Public together allow, for example, literature Tuberculosis, gene-family-wide, and private SAR data to be simultaneously queried with commercial compounds for screening (□ ;3M). Case studies from neglected infectious disease (Malaria, TB, African Sleeping Sickness, etc) and secure traditional commercial drug discovery with global teams of researchers in virtual pharmaceutical companies and between companies (i.e. PI3 Kinase inhibitors for NSCLC) will be presented.

CINF 151 CWM Global Search: An Internet search engine for the chemist

A. J. Kos, software@akosgmbh.de, and H.-J. Himmler. AKos GmbH, Steinen, Baden-Württemberg, Germany

The Internet is a rich source of data and information for chemist. There are numerous multidisciplinary databases available for free on the Internet. Some examples of such data repositories are: PubChem, ChemSpider, eMolecules, Drugbank, KEGG, NIST, ChemSynthesis, PharmGKB, Free patents online ... It should be obvious that an end user is a) not aware of all the resources, and b) has not the time to learn every user interface and is unable to search over all of them. We provide CWM Global Search as an application that enables to search by structure, CAS Registry Number and free text over all these sources. Presently CWM Global Search performs searches in 30 databases and search engines accessing more than 100 million pages that associate data with structures.

CINF 152 www.emolecules.com: The comprehensive source of commercially available, in stock chemicals

K. Gubernator, klaus@emolecules.com. eMolecules, Inc., Del Mar, CA, United States

Samples of research chemicals are at the core of life science and chemical research organizations. Synthetic chemists use them as building blocks in their synthesis, and biological screening experts use them in their high throughput assays. In order to make the selection and procurement of these chemicals a deterministic process, the catalog data used in the selection process must be reflective of actual in stock availability. eMolecules uses electronic data uploads from most of the top suppliers to achieve 96% reliability of it's catalog data. The www.emolecules.com web site offers sophisticated data import, exact and substructure searches, and methods for refining searches on an easy-to-use website. Compounds selected can be purchased online on the eCommerce system. As a result, procurement of research chemicals becomes a seamless and predictable process.

CINF 153 Petascale lattice-Boltzmann simulations of dynamical processes in ternary amphiphilic liquid crystalline systems

R. S. Saksena, r.saksena@ucl.ac.uk, and **P. V. Coveney**, p.v.coveney@ucl.ac.uk. Department of Chemistry, University College London, London, United Kingdom

We present large-scale kinetic lattice-Boltzmann simulations of the spontaneous self-assembly of periodic lamellar, hexagonal and cubic (gyroid, diamond and primitive) amphiphilic mesophases in a mixture of oil, water and amphiphiles.. Different types of periodic mesophases are formed as pressure and relative concentration of the component species is varied, in agreement with experiments. We use state-of-the-art grid middleware to explore the mesophase diagram. The Application Hosting Environment facilitates construction of complex workflows involving deployment of simulations on geographically distributed supercomputers and access to terascale computational storage and visualization resources. Our lattice-Boltzmann code LB3D exhibits excellent parallel scaling performance on petascale supercomputers, allowing us to efficiently run large-scale simulations free of finite-size effects. This computational infrastructure has allowed analysis of multi-scale dynamical processes in these nanomaterials at high temporal resolution which was not feasible previously in a production sense.

CINF 154 Novel, accurate high-throughput 3D pharmacophore screening algorithm

G. Wolber[1,2], wolber@inteligand.com, T. Seidel[1], F. Bendix[1], and P. Markt[2]. [1]Science, Technology, Inte:Ligand GmbH, Vienna, Austria, [2]Centre for Molecular Biosciences Innsbruck, Institute of pharmacy, Dptmt. pharm chemistry, University of Innsbruck, Innsbruck, Austria

Virtual screening using three-dimensional arrangements of chemical features (3D pharmacophores) has become an important method in computer-aided drug design. Although frequently used, considerable differences exist in the interpretation of these chemical features and their corresponding 3D overlay algorithms. We have recently developed an efficient and accurate 3D alignment algorithm based on a pattern recognition technique [1]. In the presented work, we extend this algorithm to be used for high-performance virtual database screening and investigate, whether applying this geometrically more accurate 3D alignment algorithm improves virtual screening results over conventional incremental n-point distance matching approaches. [1] G. Wolber, A. Dornhofer, and T. Langer. Efficient overlay of small molecules using 3-D pharmacophores. J. Comput.-Aided Mol. Design, 20(12): 773-788 (2006)

CINF 155 Structure – toxicity relationship study for organophosphorus compounds: QSAR analysis

Y. Paukku[1], yuliya@ccmsi.us, D. Magers[2], and G. Hill[1]. [1]Department of Chemistry, Jackson State University, Jackson, MS, United States, [2]Department of Chemistry and Biochemistry, Mississippi College, Clinton, MS, United States

A QSAR study is performed on a set of organophosphorus compounds in order to reveal structural and quantum-chemical features responsible for toxic effect. The development of QSAR equations which will accurately predict toxicity for organophosphorus compounds is in high demand. Current study is devoted to the investigation of a relationship between structure and acute toxicity of organophosphates by application of quantum-chemical techniques and QSAR approach followed by subsequent validation of the results. Obtained QSAR models are beneficial for virtual screening of toxicity for new compounds of interest. All experimental data have been collected from literature sources. Toxicity of organophosphorus compounds is dependent on conformational properties. Therefore, in order to find the lowest-energy structures, conformational search has been performed prior to optimization of geometries. Quantum-chemical calculations are carried out at DFT/B3LYP level of theory with 6-311++G(d,p) basis set for global minima optimization with further evaluation of electronic properties of target molecules.

CINF 156 EEM-Hückel model for partial atomic charges

M. Waldman, marv@simulations-plus.com, R. Fraczkiewicz, and W. S. Woltosz. Simulations Plus, Inc., Lancaster, CA, United States

Partial atomic charges are employed in a variety of molecular modeling and cheminformatics applications including force fields and as descriptors in modeling numerous biopharmaceutical properties such as pKa, log P, activity, toxicity, and metabolism. We present a new model for partial atomic charges covering an extremely broad range of chemistries. The model was parameterized using an *ab initio* dataset of almost 1000 molecules and over 19,000 atoms. The partial charges were separated into sigma and pi components using Natural Atomic Orbitals (NAO) and Natural Bond Orbitals (NBO) analysis. The sigma charges were

modeled using an Electronegativity Equalization Method (EEM) and pi charges using a Huckel model approach. Performance of the model on both training and external test sets achieved root mean squared errors (RMSE) below 0.05 a.u. Performance on various complex molecules will be discussed along with an analysis of parameter coverage on several large databases.

CINF 157 PropOrb: A frontier molecular orbital interaction proposer

C.-A. Azencott, cazencot@ics.uci.edu, M. A. Kayala, mkayala@ics.uci.edu, and P. Baldi, pfbaldi@ics.uci.edu. Department of Information and Computer Science, UC Irvine, Irvine, CA, United States

Chemoinformatics methods to assist chemists in the essential problem of reaction prediction are highly desirable. Brute force examination of all potentially interacting molecular orbitals (MOs) is computationally intractable. For example, concurrent work to develop a database of MO interactions yields more than 100 million interactions for a few thousand systems. Inspired by human chemists, we leverage the principles of frontier orbital theory to propose a tool that rapidly identifies the most reactive molecular orbitals in a given system. Due to the simplicity of our approach, we aim to identify only a few top candidates. Further resource intensive refinement, such as quantum mechanical or statistical approaches, might then be necessary to differentiate between the proposer's suggestions. However, our approach significantly reduces the number of possibilities for downstream analysis. The speed, simplicity, and efficacy of our MO interaction proposer make it an essential component of a reaction prediction system.

CINF 158 Fast and accurate prediction of the 3D structure of small molecules

A. Andronico, aandroni@uci.edu, A. Randall, and P. Baldi. Institute for Genomics and Bioinformatics, University of California Irvine, Irvine, CA, United States

A large variety of chemical and biological properties of a molecule are related to the spatial arrangement if its atoms (its 3D structure). The positioning of atoms in space can be accurately predicted via quantum mechanics, but, when dealing with large numbers of molecules, it is necessary to resort to less time-consuming methods. In this paper, we present a fast method for generating the 3D structure of small molecules. The prediction is obtained by fragmenting the molecule into rigid segments of known 3D structure: these fragments, stored in a database, are then joined back together using torsions information also obtained from experimental data. Whenever a fragment is not found in the database, different steps are taken, depending on the complexity of its structure. By comparison with other conformation predicting programs, we show that our method produces high-quality 3D structures of organic and organometallic molecules with sub-second prediction time.

2010 CINF OFFICERS AND FUNCTIONARIES

Chair:
Ms. Carmen Nitsche
Symyx Technologies, Inc.
254 Rockhill Drive
San Antonio, TX 78209
510-589-3555 (mobile phone)
210-820-3459 (office and fax)
Carmen.Nitsche@symyx.com

Chair Elect:
Dr. Gregory Banik
Bio-Rad Laboratories, Inc.
Informatics Division
2 Penn Center Plaza, Suite 800
1500 John F Kennedy Blvd
Philadelphia, PA 19102-1721
267.322.6952(voice)
267.322.6953 (fax)
gregory_banik@bio-rad.co

Past Chair/Nominating Chair:
Ms. Svetlana Korolev
University of Wisconsin, Milwaukee
2311 E. Hartford Avenue
Milwaukee, WI 53211
414.229.5045(voice);414.229.6791(fax)
skorolev@uwm.edu

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

Treasurer:
Ms. Meghan Lafferty
Science & Engineering Library
University of Minnesota
108 Walter Library
117 Pleasant St SE
Minneapolis MN 55455
612.624.9399 (voice)
612.625.5583 (fax)
mlaffert@umn.edu

Councilor:
Ms. Bonnie Lawlor
National Federation of Advanced
Information Services (NFAIS)
276 Upper Gulph Road
Radnor, PA 19087-2400
215.893.1561 (voice)
215.893.1564 (fax)
blawlor@nfais.org

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

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

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

Archivist/Historian:
Ms. Bonnie Lawlor
National Federation of Advanced
Information Services (NFAIS)
276 Upper Gulph Road
Radnor, PA 19087-2400
215.893.1561 (voice)
215.893.1564 (fax)
blawlor@nfais.org

Audit Chair:
Ms. Jody Kempf
Science & Engineering Library
University of Minnesota
108 Walter Library
117 Pleasant St SE
Minneapolis MN 55455
612.624.9399 (voice)
612.625.5583 (fax)
j-kemp@umn.edu

Awards Committee Chair:
Dr. Phil McHale
CambridgeSoft Corporation
375 Hedge Rd
Menlo Park, CA 94025-1713
650.235.6169 (voice)
650.362.2104 (fax)
pmchale@cambridgesoft.com

Careers Committee Chair:
Ms. Patricia Meindl
University of Toronto
A. D. Allen Chemistry Library
80 St George Street, Rm 480
Toronto, ON M5S 3H6
416.978.3587 (voice)
416.946.8059 (fax)
pmeindl@chem.utoronto.ca

Constitution, Bylaws & Procedures:
Ms. Susanne Redalje
University of Washington
Chemistry Library
BOX 351700
Seattle, WA 98195
206. 543.2070 (voice)
curie@u.washington.edu

Education Committee Chair:
Mr. Jeremy Garritano
Purdue University Libraries—CHEM,
504 W. State Street,
West Lafayette, IN 47907
765. 496.7279 (voice)
765. 494.1579 (fax)
jgarrita@purdue.edu

Finance Committee Chair:
See Treasurer

Fund Raising Committee Chair:
Mr. Graham Douglas
Scientific Information Consulting
1804 Chula Vista Drive
Belmont, CA 94002
510.407.0769 (voice)
510.217.3853 (fax)
Graham_C_Douglas@hotmail.com

Membership Committee Chair:
Ms. Jan Carver
University of Kentucky
Chemistry Physics Library
150 Chem Phys Bldg
Lexington, KY 40506-0001
859. 257.4074 (voice)
859. 323.4988 (fax)
jbcarv1@email.uky.edu

Program Committee 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

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

Chemical Information Bulletin Editor
Dr. Svetla Baykoucheva
White Memorial Chemistry Library
University of Maryland
College Park, MD 20742
301.405.9080 (voice)
301. 314.5910 (fax)
sbaykouc@umd.edu

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

Webmaster:
Mr. Richard Williams
Vertex Pharmaceuticals
P.O. Box 290718
Boston, MA 02129
617.444.6325 (voice)
RickWilliams@wis-llc.org