

CHEMICAL

Information

BULLETIN

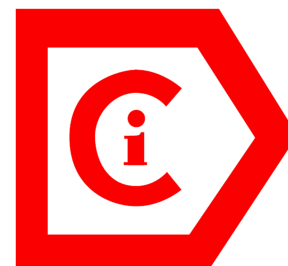


ACS
Chemistry for Life[®]

Winter 2017 — Vol. 69, No. 4

A Publication of the
Division of Chemical Information
of the ACS

ISSN: 0364-1910



Chemical Information Bulletin

A Publication of the Division of Chemical Information of the ACS

Winter 2017 — Vol. 69, No. 4

David S. Shobe, Editor
avidshobe@yahoo.com

Message from the Chair	3
Letter from the Editor	4
Awards and Scholarships	5
2018 Herman Skolnik Award Announced	5
2019 Herman Skolnik Award: Call for Nominations	7
CINF Scholarship for Scientific Excellence	8
Applications Invited for CSA Trust Grants for 2018	11
Interview	17
Technical Program	24
Informatics and Chemical Biology: Identifying Targets and Biological Pathways	24
Professional Skills Development for Undergraduates, Graduates, and Post-Docs	26
What do synthetic chemists want from their reaction systems?	29
Herman Skolnik Award Symposium 2017	37
Committee Reports	64
Book Review	69
Sponsor Announcements	73
ChemRxiv	73
<i>Journal of Chemical Information and Modeling</i>	74
Simulations Plus ADMET Predictor 8.5	75
Royal Society of Chemistry news update	76
A Global Collaborative ELN: Signals Notebook	78
CINF Officers and Functionaries	80
Contributors to This Issue	84

ISSN: 0364-1910

Chemical Information Bulletin

Message from the Chair

Erin Davis, CINF Division Chair



Hello CINF!

I've always enjoyed the location-relevant and thematic programming that CINF embraces for national meetings. This past meeting in Washington, DC was no exception, with heavy focus around government funded data sources and feverish discussion around Open Data. As someone who has always been a consumer of data, I'm fascinated to hear about the philosophical issues behind the curation of said data. I'm glad that there are such passionate people in CINF looking to ensure that data remain both high-quality and accessible, on both sides of the debate.

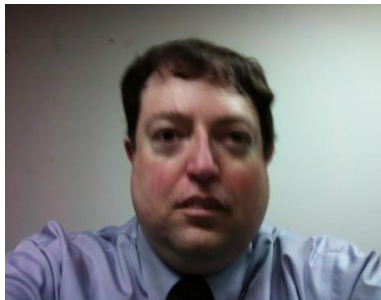
We look forward to the spring national meeting in New Orleans. I hope that more members will be able to join us, as this is usually a more affordable location. We also have a diverse array of symposia lined up, so please have a look and see if there are any for you to present in!

As always, I would love to continue hearing from CINF members about what they'd like to get from the division, especially those who cannot attend national meetings. Please reach out to me if you have any thoughts on the subject.

Thanks for being a member!

Letter from the Editor

David Shobe, Editor of the Winter *Chemical Information Bulletin*



I actually have had the privilege of visiting the Washington, DC area four years ago. I even have some pictures to show for it, although unfortunately they are not suitable for the cover photo of this publication. (They *literally* prove that I was there, in the sense that some of them show my blurry fingertip alongside iconic DC buildings).

Unfortunately, I was unable to return to Washington this August 20-24, 2017 for the ACS meeting. I have come to think of reading the summer and winter issues of the *Chemical Information Bulletin* as the next best thing, with their summaries of technical program symposia and committee reports. In fact, I would have liked more submissions in those areas, but I realize that people are busy and it takes work to take notes during the meeting and write them up for the bulletin. Given that fact, I would like to thank those who did contribute content, as your hard work is appreciated.

In addition to the technical program and committee reports, this issue includes an interview of former division chair Carmen Nitsche by Svetlana Korolev and a book review of *Bibliometrics and Research Evaluation: Uses and Abuses* from Bob Buntrock. Also be sure to read our sponsor announcements, as our sponsors keep the Chemical Information Division financially afloat. Finally, congratulations to Gisbert Schneider, as the winner of the 2018 Herman Skolnik award; his profile, by David Evans, follows on the next page.

Do you want to relive (or experience for the first time) key moments from this fall's meeting?

Wendy Warr has published her photos from the ACS National Meeting in Washington, DC on her Flickr stream. Visit <https://www.flickr.com/photos/cinf/albums> to access the photos. They appear in six albums:

- ACS Washington 2017 Skolnik symposium
- ACS Washington 2017 Skolnik reception
- ACS Washington 2017 Schrödinger reception
- ACS Washington 2017 International reception
- ACS Washington 2017 COMP reception
- ACS Washington 2017 CINF welcome reception

While you're there, check out the photos from other past national meetings and special CINF events!

Awards and Scholarships

2018 Herman Skolnik Award Announced

David Evans, Chair, ACS Division of Chemical Information Awards Committee



The American Chemical Society Division of Chemical Information is pleased to announce that Gisbert Schneider, ETH Zurich, Switzerland has been selected to receive the 2018 Herman Skolnik Award for his seminal contributions to *de novo* design of bioactive compounds and the application of these innovative design concepts in both academia and industry. The award recognizes outstanding contributions to and achievements in the theory and practice of chemical information science and related disciplines. The prize consists of a \$3,000 honorarium and a plaque. Prof Dr Schneider will also be invited to present an award symposium at the fall 2018 ACS national meeting to be held in Boston.

Prof. Dr. Schneider is a full professor at ETH Zurich, holding the Chair for Computer-Assisted Drug Design. Over the last 25 years he has worked in a variety of areas in cheminformatics and computational molecular design. He is recognized as being a pioneer in the integration of machine-learning methods into practical medicinal chemistry, and for his coining the phrases "scaffold-hopping" and "frequent hitter". His career has led him from the pharmaceuticals division at Roche to academia, initially to the Goethe-University in Frankfurt where he held the Beilstein Endowed Chair for Chem- and Bioinformatics, and then to his current position at ETH in Zurich. He is an elected Fellow of the University of Tokyo, and an Adjunct Professor at Goethe-University. He has co-founded several start-up companies, including inSili.com GmbH, AlloCyte Pharmaceuticals AG, and Endogena Therapeutics Inc.

His current research interests focus on the development of methods for adaptive autonomous systems in drug research. Current projects include developing, implementing, and experimentally validating these innovative concepts in applied settings, including:

- Virtual screening by active learning
- Constructive *de novo* molecular design methods
- Macromolecular target profiling and polypharmacology
- Self-organizing systems for molecular pattern recognition.

His research group develops and applies these often nature-inspired algorithms to virtual compound screening, drug repurposing and deorphaning, *in silico* polypharmacology and chemogenomics projects, protein structure analysis, and the design of allosteric and natural-product-derived ligands.

The awarding of the 2018 Herman Skolnik Award to Schneider recognizes his significant contributions to the fields of cheminformatics and *in silico* molecular design methods. His nomination discussed his

broad contributions to the field including:

- Automated reaction-based *de novo* design for rapid lead prototyping
- Conception of multi-objective adaptive fitness landscapes for library design
- Development of ligand- and structure-based methods for identifying macromolecular targets of pharmaceutically active compounds.

Schneider's early research resulted in the invention and coining of computational peptide design by adaptive optimization, using neural networks and evolutionary algorithms. These seminal studies laid the foundation for "artificially intelligent" molecular design. His more recent work pursues innovative machine-learning models for target and similarity prediction in automated hit and lead discovery, bridging the chemical and biological worlds. Schneider's name is found on the Thomson Reuters list of the "World's Most Influential Scientific Minds". His inspirational studies are documented in over 400 publications and six books.

Schneider is also cited for his contributions to our field. He was a founding editor of the journal *Molecular Informatics*, and serves as a reviewer for many top-ranking journals including *Nature*, *Science*, and *Angewandte Chemie*. He has also been very active in teaching, and his efforts have been recognized by the students at ETH: Schneider has won the "Golden Owl" award for outstanding faculty teaching.

2019 Herman Skolnik Award: Call for Nominations

David Evans, Chair, ACS Division of Chemical Information Awards Committee

The ACS Division of Chemical Information established this award to recognize outstanding contributions to and achievements in the theory and practice of chemical information science. The award is named in honor of the first recipient, Herman Skolnik.

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

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

The award consists of a \$3,000 honorarium and a plaque. The recipient is expected to give an address at the time of the Award presentation. In recent years, an award symposium has been organized by the recipient.

Nominations for the Herman Skolnik Award should describe the nominee's contributions to the field of chemical information and should include supportive materials such as a biographical sketch and a list of publications and presentations. Three seconding letters are also required. Nominations and supporting material should be sent by email to awards@acscinf.org. Paper submissions will not be accepted. The deadline for nominations for the 2019 Herman Skolnik Award is June 1, 2018.

CINF Scholarship for Scientific Excellence

Purpose

The scholarship program of the Division of Chemical Information (CINF) of the American Chemical Society (ACS) is designed to reward graduate and postdoctoral students in chemical information and related sciences for scientific excellence and to foster their involvement in CINF.

Scope

Scholarships valued at \$1,000 each will be awarded at the ACS National Meetings.

Eligibility & Application

Applicants must:

- be enrolled at a certified college or university.
- present a poster during the CINF Welcoming Reception and the Sci-Mix session at the National Meeting.
- Abstracts for the poster must be submitted electronically through the [Meeting Abstracts Programming System](#) (MAPS) according to ACS rules approximately three months in advance of the meeting in question.
- send in electronic form a 2,000-word long abstract describing the work to be presented:
 - submitted to: [Stuart Chalk](#)
 - due by mid-June for presentation at the Fall Conference of that year.
 - due by mid-January for presentation at the Spring Conference of the upcoming year.

Any questions related to applying for one of the scholarships should also be directed to [Stuart Chalk](#)

Selection

Winners will be chosen based on content, presentation, and relevance of the poster, and they will be announced at the meeting. The content shall reflect upon the student's work and describe research in the field of cheminformatics and related sciences. At the Sci-Mix session, winning posters will be marked as "Winner of ACS Publications CINF Scholarship Award for Scientific Excellence"

Recent Sponsors and Recipients

ACS Nat'l	Sponsor	Recipient/Poster Title
#254, Fall, 2017	ACS Publications	Phyo Phyo Kyaw Zin , <i>Department of Chemistry, Bioinformatics Research Center, North Carolina State University, Raleigh, NC 27695, USA</i> “PKS Enumerator Software to Explore the Chemical Space of Macrolides”
#254, Fall, 2017	ACS Publications	Mohammad Atif Faiz Afzal , <i>Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Buffalo, New York, USA</i> “Deep learning approach for the fast and accurate prediction of optical properties of organic molecules”
#254, Fall, 2017	ACS Publications	Jeremy R. Ash , <i>Department of Chemistry, Bioinformatics Research Center, North Carolina State University, Raleigh, NC 27695, USA</i> “Cheminformatics Approach to Exploring and Modeling Trait-Associated Metabolic Profiles”
#253, Spring, 2017	ACS Publications	Andrew McEachran , <i>National Center for Computational Toxicology, Environmental Protection Agency, Research Triangle Park, North Carolina, USA</i> “Mobilizing EPA’s Comptox Chemistry Dashboard data on mobile devices”
#253, Spring, 2017	ACS Publications	Matthew Seddon , <i>Information School, University of Sheffield, Regent Court, 211 Portobello, Sheffield S1 4DP, UK</i> “Global spectral and diffusion geometry descriptors of 3D molecular shape for virtual screening”
#253, Spring, 2017	ACS Publications	Christopher T. Lee , <i>Department of Chemistry and Biochemistry, University of California - San Diego, La Jolla CA 92093, USA</i> “Investigating transport properties with multiscale computable mesh models from heterogeneous structural datasets”
#252, Fall, 2016	ACS Publications	Mojtaba Haghightlari , <i>Department of Chemical and Biological Engineering, University at Buffalo</i> “ChemML: A Machine Learning and Informatics Program Suite for the Chemical and Materials Sciences”
#252, Fall, 2016	ACS Publications	George Van Den Driessche , <i>Department of Chemistry, Bioinformatics Research Center, North Carolina State University</i> “Forecasting Adverse Drug Reactions Triggered by the Common HLA-B*57:01 Variant”

#252, Fall, 2016	ACS Publications	Nathanael Kazmierczak , <i>Department of Chemistry & Biochemistry, Calvin College</i> “Modeling spectrophotometric titration data: tracking error from the measurement, through the model, and to the targeted output parameters”
#251, Spring, 2016	Springer & InfoChem	Wiliam Augusto Cortopassi , <i>Chemistry Research Laboratory, University of Oxford, Oxford, UK</i> “Prediction and quantification of cation- π interactions in ligand-bromodomain binding: Using quantum chemistry to capture electronic effects”
#251, Spring, 2016	Springer & InfoChem	Iva Lukac , <i>School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool, UK</i> “Quantifying the effect that chemical environment exerts upon changes in property in matched molecular pairs analysis”
#251, Spring, 2016	Springer & InfoChem	Yu-Chen Lo , <i>Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California, USA</i> “CSNAP: A new chemoinformatics approach for target identification using chemical similarity networks”



Chemical Structure Association Trust

Applications Invited for CSA Trust Grants for 2018.

The Chemical Structure Association (CSA) Trust is an internationally recognized organization established to promote the critical importance of chemical information to advances in chemical research. In support of its charter, the Trust has created a unique Grant Program and is now inviting the submission of grant applications for 2018.

Purpose of the Grants

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

Who is Eligible?

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

Which Activities are Eligible?

Grants may be awarded to acquire the experience and education necessary to support research activities; e.g. for travel to collaborate with research groups, to attend a conference relevant to one's area of research (including the presentation of an already-accepted research paper), to gain access to special computational facilities, or to acquire unique research techniques in support of one's research. Grants will not be given for activities completed prior to the grant award date.

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 and a recent photograph;
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. A copy of the completed application document must be supplied for distribution to the Grants Committee and can be submitted via regular mail or e-mail to the Committee Chair (see contact information below).

Deadline for Applications:

Application deadline for the 2018 Grant is March 30, 2018. Successful applicants will be notified no later than May 9, 2018.

Address for Submission of Applications

The application documentation can be mailed via post or emailed to: Bonnie Lawlor, CSA Trust Grant Committee Chair, 276 Upper Gulph Road, Radnor, PA 19087, USA. If you wish to enter your application by e-mail, please contact Bonnie Lawlor at chescot@aol.com prior to submission so that she can contact you if the e-mail does not arrive.

Chemical Structure Association Trust: Recent Grant Awardees

2017

Jesus Calvo-Castro: University of Hertfordshire, England, was awarded a Grant to cover travel to present his work at the Fifth International Conference on Novel Psychoactive Substances to be held in Vienna, Austria from August 23-23, 2017. He works on the development of novel methodologies for the in-the-field detection of novel psychoactive substances (NPS), where chemical structure and information play a crucial role.

Jessica Holien: St. Vincent's Institute of Medical Research, Fitzroy, Victoria, Australia, was awarded a Grant to cover travel to present her work at the 2017 Computer-Aided Drug Design (CADD) Gordon

Research Conference scheduled to take place July 16-21, 2017 in Mount Snow, VT, USA. She is a Postdoctoral researcher at St. Vincent's and is responsible for a range of computational molecular modelling including; compound database development, virtual screening, docking, homology modelling, dynamic simulations, and drug design.

2016

Thomas Coudrat: Monash University, Australia, was awarded a Grant to cover travel to present his work at three meetings in the United States: the Open Eye Scientific CUP XVI, The American Chemical Society Spring Meeting, and the Molsoft ICM User Group Meeting. His work is in ligand directed modeling.

Clarisse Pean: Chimie Paris Tech, France, was awarded a Grant to cover travel to give an invited presentation at the 2016 Pacific Rim Meeting on Electrochemical and Solid State Science later this year.

Qian Peng: University of Oxford, England, was awarded a Grant to attend the 23rd IUPAC Conference on Physical Organic Chemistry. His research is in the development of new ligands for asymmetric catalysis.

Petteri Vainikka: University of Turku, Finland, was awarded a Grant to spend the summer developing and testing new methods for modelling organic solvents in organic solutions with Dr. David Palmer and his group at the University of Strathclyde, Glasgow, Scotland.

Qi Zhang: Fudan University, China, was awarded a Grant to attend a Gordon Conference on Enzymes, coenzymes and metabolic pathways. His research is in enzymatic reactions.

2015

Dr. Marta Encisco: Molecular Modeling Group, Department of Chemistry, La Trobe Institute for Molecular Science, La Trobe University, Australia was awarded a Grant to cover travel costs to visit collaborators at universities in Spain and Germany and to present her work at the European Biophysical Societies Association Conference in Dresden, Germany in July 2015.

Jack Evans: School of Physical Science, University of Adelaide, Australia was awarded a grant to spend two weeks collaborating with the research group of Dr. Francois-Xavaier Coudert (CNRS, Chimie Paris Tech).

Dr. Oxelandr Isayer: Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, was awarded a Grant to attend summer classes at the Deep Learning Summer School 2015 (University of Montreal) to expand his knowledge of machine learning to include Deep Learning (DL). His goal is to apply DL to chemical systems to improve predictive models of chemical bioactivity.

Aleix Gimeno Vives: Cheminformatics and Nutrition Research Group, Biochemistry and Biotechnology Dept., Universitat Rovira I Virgili was awarded a Grant to attend the Cresset European User Group Meeting in June 2015 in order to improve his knowledge of the software that he is using to determine what makes an inhibitor selective for PTP1B.

2014

Dr. Adam Madarasz: Institute of Organic Chemistry, Research Centre for Natural Sciences, Hungarian Academy of Sciences. He was awarded a Grant for travel to study at the University of Oxford with Dr. Robert S. Paton, a 2013 CSA Trust Grant winner, in order to increase his experience in the development of computational methodology which is able to accurately model realistic and flexible transition states in chemical and biochemical reactions.

Maria José Ojeda Montes: Department of Biochemistry and Biotechnology, University Rovira i Virgili, Spain. She was awarded a Grant for travel expenses to study for four months at the Freie University of Berlin to enhance her experience and knowledge regarding virtual screening workflows for predicting therapeutic uses of Natural molecules in the field of functional food design.

Dr. David Palmer: Department of Chemistry, University of Strathclyde, Scotland. He was awarded a Grant to present a paper at the fall 2014 meeting of the American Chemical Society on a new approach to representing molecular structures in computers based upon ideas from the Integral Equation Theory of Molecular Liquids.

Sona B. Warriar: Departments of Pharmaceutical Chemistry, Pharmaceutical Biotechnology, and Pharmaceutical Analysis, NMIMS University, Mumbai. She was awarded a Grant to attend the International Conference on Pure and Applied Chemistry to present a poster on her research on inverse virtual screening in drug repositioning.

2013

Dr. Johannes Hachmann: Department of Chemistry and Chemical Biology at Harvard University, Cambridge, MA. He was awarded the Grant for travel to speak on "Structure-property relationships of molecular precursors to organic electronics at a workshop sponsored by the Centre Européen de Calcul Atomique et Moléculaire (CECAM) that will take place October 22 – 25, 2013 in Lausanne, Switzerland. .

Dr. Robert S. Paton: University of Oxford, UK. He was awarded the Grant to speak at the Sixth Asian Pacific Conference of Theoretical and Computational Chemistry in Korea on July 11, 2013. Receiving the invitation for this meeting has provided Dr. Paton with an opportunity to further his career as a Principal Investigator.

Dr. Aaron Thornton: Material Science and Engineering at CSIRO in Victoria, Australia. He was awarded the Grant to attend the 2014 International Conference on Molecular and Materials Informatics at Iowa

State University with the objective of expanding his knowledge of web semantics, chemical mark-up language, resource description frameworks and other on-line sharing tools. He will also visit Dr. Maciej Haranczyk, a prior CSA Trust Grant recipient, who is one of the world leaders in virtual screening.

2012

Tu C. Le: CSIRO Division of Materials Science & Engineering, Clayton, VIV, Australia. She was awarded the Grant for travel to attend a Cheminformatics course at Sheffield University and to visit the Membrane Biophysics group of the Department of Chemistry at imperial College London.

2011

J. B. Brown: Kyoto University, Kyoto, Japan. J.B. was awarded the Grant for travel to work with Professor Ernst Walter-Knappat the Freie University of Berlin and Professor Jean-Phillipe Vert of the Paris MinesTech to continue his work on the development of atomic partial charge kernels

2010

Noel O'Boyle: University College Cork, Ireland. Noel was awarded the grant to both network and present his work on open source software for pharmacophore discovery and searching at the 2010 German Conference on Cheminformatics.

2009

Laura Guasch Pamies: University Rovira & Virgili, Catalonia, Spain. Laura was awarded the Grant to do three months of research at the University of Innsbruck, Austria.

2008

Maciej Haranczyk: University of Gdansk, Poland. Maciej was awarded the Grant to travel to Sheffield University, Sheffield, UK, for a 6-week visit for research purposes.

2007

Rajarshi Guha: Indiana University, Bloomington, IN, USA. Rajarshi was awarded the Grant to attend the Gordon Research Conference on Computer Aided Design in August 2007.

2006

Krisztina Boda: University of Erlangen, Erlangen, Germany. Krisztina was awarded the Grant to attend the 2006 spring National Meeting of the American Chemical Society in Atlanta, GA, USA.

2005

Dr. Val Gillet and **Professor Peter Willett**: University of Sheffield, Sheffield, UK. They were awarded the Grant for student travel costs to the 2005 Chemical Structures Conference held in Noordwijkerhout, the Netherlands.

2004

Dr. Sandra Saunders: University of Western Australia, Perth, Australia. Sandra was awarded the Grant to purchase equipment needed for her research.

2003

Prashant S. Kharkar: Institute of Chemical Technology, University of Mumbai, Matunga, Mumbai. Prashant was awarded the Grant to attend the conference, Bioactive Discovery in the New Millennium, in, Lorne, Victoria, Australia (February 2003) to present a paper, The Docking Analysis of 5-Deazapteridine Inhibitors of Mycobacterium avium complex (MAC) Dihydrofolate reductase (DHFR).

2001

Georgios Gkoutos: Imperial College of Science, Technology and Medicine, Dep. of Chemistry. London, UK. Georgios was awarded the Grant to attend the conference, Computational Methods in Toxicology and Pharmacology Integrating Internet Resources, (CMTPI-2001) in Bordeaux, France, to present part of his work on internet-based molecular resource discovery tools.

Interview

An interview with Carmen Nitsche, a highly effective collaborator in chemical information

Svetlana Korolev, interviewer



This article continues a series of interviews with functionaries of the ACS Division of Chemical Information (CINF) compiled at:

<http://www.acscinf.org/content/interviews>. Carmen Nitsche is well-recognized in CINF for her leadership role as the division chair 2010, webinars coordinator 2014-16, alternate councilor 2016-18, and a frequent speaker at ACS national meetings. Bio: Carmen Nitsche earned a B.A. with honors in chemistry from the University of Minnesota at Minneapolis and an M.S. from the University of California at Berkeley. She held laboratory positions at ARCO and Los Alamos National Laboratory before entering the field of chemical information in 1987, joining Nalco

Chemical Company as an Information Scientist. During her 12 years with the Library and Information Services group, Carmen was responsible for technical and chemical searching and end-user search services. She left her Staff Scientist position in 1999 to join Nalco sales, working with water treatment customers in the greater San Antonio area.

Carmen returned to chemical information in 2001, when she accepted a position in the Business Development group at MDL Information Systems, an Elsevier Company. She continued there through Symyx's purchase of MDL, and Accelrys' purchase of Symyx, holding positions of Vice President Content and Vice President Corporate Development. In 2013 Carmen founded CINforma Consulting, dedicated to business development in the scientific software and content arenas. Her largest client is the Pistoia Alliance, where she helps develop their project portfolio, and their membership base. Carmen's professional memberships include Sigma Xi and ACS. She continues to publish and speak, most recently at the spring 2017 ACS meeting in San Francisco and the Dassault "Science in the Age of Experience" conference in May.

Svetlana Korolev: Carmen, I first met you at a CINF executive committee meeting after your winning the Chair-Elect 2009 position. Since then you have been very active in division governance with continuous commitment to "strengthening interdisciplinary approach, leveraging new technologies, and reaching out to new and existing members", to quote from your statement of goals in the 2015 division elections. Please share your insights into CINF and ACS in this context. What can the division and the society do to better serve its members?

Carmen Nitsche: The most exciting changes tend to happen at the interfaces of disciplines. Colleagues with different perspectives inject new thinking and new ideas into the mix, and so one can make real progress in tackling challenges. CINF is a small division, but our discipline affects absolutely every

member of the chemical community, and we can have the appropriate impact if we seek out those partnerships with other disciplines in other divisions. I would highlight the CINF/CHAS partnership around safety, where there are several projects underway where the librarians are contributing their deep knowledge about information management, handling and access, and the EH&S (Environmental Health and Safety) colleagues bring their safety and laboratory management experience and expertise to bear. Together, we will come up with much more robust and sound solutions and recommendations, and we have a much better chance for wider adoption, as more stakeholders have been involved.

As far as technology is concerned, I have always felt that information professionals are the ideal candidates to help all of us navigate the tremendous changes we see in information access and volume. We should never fear our obsolescence with every new technology change. At any point in time, the medium and tools at hand are less important than the underlying principles around data validation, data handling, information dissemination and the like, and these are the core of our discipline.

I do feel our division needs to create more tangible benefits for its members and prospective members, and we should be reaching out more to all manner of chemists. This is why I was so involved in the webinar series. Our virtual events were a way to provide idea exchange and connection outside the national meetings, which many of our members cannot afford to attend. This is where the national organization could help as well: providing the infrastructure to support small division ambitions, so we can focus on inviting interesting speakers, or making useful materials readily available, rather than scrambling to find a webinar platform or workable document sharing tool.

SK: Reflecting on your third goal for “reaching out to new and existing members”, I would like to acknowledge again generous funding of the CINF Scholarship for Scientific Excellence program sponsored by your former companies, Symyx and Accelrys. While speaking at the “Careers in Chemical Information and Cheminformatics” discussion panel (a [summary](#) of which was published in the Chemical Information Bulletin), you advised to younger scientists: “It is important to network and stay in touch with what is going on. Actively look for mentors who have an interest in you and offer a reality check”. Have you benefited from mentor relationships at any point of your career? What are your personal favorite tools for staying abreast in the chemical information field?

CN: I have been fortunate over the years to have several trusted mentors who have helped me navigate new terrain. When I look back, most every job I have taken was a leap into the unknown. But each leap was less intimidating because of mentors who helped me find my strengths and my voice.

I would call out a couple of people in particular. When I first started in chemical information, I knew next to nothing about the job I had just accepted. But Steve Boyle at Nalco hired me anyway. He became my most valuable teacher and mentor, and ultimately my friend. He let me shadow him through all the tasks, and he seemed to trust me when I was not sure I trusted myself. He made it clear I could always ask for help, which made it so much easier to try new things. He also encouraged me to engage outside our organization, allowing me to develop professionally and bring back ideas we could try to implement

in-house. This is how Nalco became one of the beta test sites for SciFinder, and how I first became involved in CINF.

I would also call out John Regazzi, the former CEO of Engineering Information, whom I met during our Elsevier days. He has had a rich and varied professional life after leaving Elsevier, including a stint as the Dean of the College of Information and Computer Science at Long Island University and as a Managing Director at Akoya Capital Partners LLC, where he leads their Professional Information Services Sector. I had long admired his embrace of change, and reached out to him as I was contemplating starting my own business. John has always been generous with his time, and we touch base regularly. And now and then I am able to give back. One must always remember, a successful mentor/mentee relationship is a two-way street.

Regarding staying in touch with the chemical information field, I still depend heavily on the CHMINF-L listserv hosted at Indiana University. I know and trust the people on this list and know that if an important topic is trending, someone on the list will keep the rest of us abreast.

SK: What was the most interesting item in your schedule during the latest ACS national meeting in Washington, DC?

CN: As you know, I am involved in a safety project at the Pistoia Alliance. At this last ACS meeting I was able to meet so many folks at ACS that are deeply committed to safety culture, and I have very high hopes that we will soon see many more concrete, useful actions from the society that embody the new ACS core value of safety. The most interesting session I attended was hosted by Allison Campbell, President of the ACS, on Safety Culture, top down and bottom up. The room really should have been overflowing, this is such an important issue and the panel was world class.

SK: How is the ACS national meeting experience different from the Bio-IT World Conference & Expo? What were your highlights at the last Bio-IT World Conference, May 23-25, 2017?

CN: The Bio-IT universe is different from ACS. The purpose is quite focused – bringing together life sciences IT professionals to “advance science, technology and patient care”, and the attendance is smaller: 3,400 attended Bio-IT this year, and it attracts mostly corporate participants. There were definitely aspects that would be of interest to CINF members. For example, there was [a hackathon dedicated to FAIR data](#), (data that are findable, accessible, interoperable, and reusable) and the expo attracted many companies delivering information-related tools and services data management, visualization and analysis that don’t tend to exhibit at ACS. I was especially proud this year that the Pistoia Alliance had 18 of its members involved in the expo or sponsoring, and that does not include those involved in the technical program.

SK: The Chemical Safety Library (CSL) project was launched by Pistoia Alliance on March 15, 2017 with “hot off the press” announcements in many scientific magazines such as Chemical & Engineering News (C&EN) and Chemistry World. You spoke about CSL at the Division of Chemical Safety symposium earlier

this year during the spring ACS national meeting in San Francisco. What was the “reaction” of the audience to the launch? Were there any concerns about disclosing trade secrets requiring protection? Is the required registration for using CSL aimed at specific restrictions? What was your most surprising finding in this project?

CN: The CSL is one of the most exciting projects I have ever worked on. The Pistoia Alliance began this initiative last year, to collect and disseminate information on laboratory chemical reaction incidents, including what components were combined, what the unexpected outcome was, and suggested warnings. It became clear early on that this initiative was going to be of interest not just to Pistoia Alliance members, but rather to all chemistry lab practitioners. First, we had to build a simple prototype tool to collect and return this type of information. Then we began our grand community experiment to determine 1) is there a need for experiential insights from the lab about reactions gone awry and 2) is the community ready to share such information.

To solicit the broadest range of participants, we embarked on an extensive publicity campaign that garnered coverage from *C&EN* and *Chemistry World*; we enlisted help from CINF and CHAS divisions, both of whom have advisors on our project team; and, as you note, I presented the launch report at a CHAS session in San Francisco in the spring.

I guess what most surprised me was the instant, overwhelming interest. By the time I gave my talk two weeks after launch we had 600 people signed up. I just checked, and today we have 871 registered users, and every week we have a few more signups. All the feedback we have received at the session and beyond suggests there is an interest and a need for this type of information to supplement current safety resources. So to our first experiment question I believe the answer is overwhelmingly yes.

On the submission side, we have seen modest increases. We started with 27 entries from our members at launch. We now have 108 entries, of which 10% are coming from non-Pistoia Alliance members, which is terrific. But we are looking for more. We are announcing a CSLDatathon for end of October, targeting librarians, students, EH&S practitioners, and others, to promote submissions. We will be providing training and prizes, and expect the event will help populate the database substantially. And for those who wonder how they might be able to use the database, we will be holding a CSLHackathon to demonstrate how the collection of data might be put to use.

You mention a few points I want to address specifically. Regarding trade secrets, we have made it clear that there is no need (or even, at this point, no ability) to share proprietary information in our prototype. And yes, we do require registration of the submitters, because our small curation team needs to be able to follow up on entries should there be any questions. However, based on the feedback in San Francisco, we have removed the public display of the submitter’s name and institution, as personal embarrassment seems to be a barrier to submission. I would suggest, however, that we should embrace and thank any submitter who is passing along their hard-learned experience, and that they should not feel shame, but rather pride in helping their fellow lab mates.

We realize that some fear legal or regulatory reprisals, but it is not clear whether these are warranted concerns or not. To that end, I am organizing a CINF session for New Orleans, which will include a panel discussion, where we will discuss safety-data-sharing fears, and evaluate how to address them with experts in the field. We will at minimum be hosting this in conjunction with CHAS and CHAL, but I hope to secure additional divisional co-sponsorship.

SK: ACS instituted safety as one of its core value last December. For one thing, safety is currently underlined in many presidential initiatives such as a symposium “Building a Safety Culture across the Chemistry Enterprise” cosponsored by 20 ACS committees and divisions, including CINF, at the fall national meeting in Washington, DC, planning for a symposium “RAMPing up the Culture of Safety” at regional meetings, and creating a ChemLuminary Award for Leadership in Safety Culture. Please talk about other projects occurring in collaboration with ACS committees and divisions in support of safety. How does your work with the [Chemical Safety Library \(CSL\)](#) project relate to them?

CN: I think it is terrific that ACS is taking a greater leadership role around safety. We see this for example with the new ACS Publications policy around novel or significant hazards reporting by authors, and with the new Task Force for Safety Education Guidelines (TFSEG). I know that CHAS, together with CINF and the Committee on Chemical Safety have a variety of other ideas underway, and there are InChI Trust and Research Data Alliance ties as well. I could even imagine partnership with ACS around the CSL, which would be most exciting.

SK: What are the other exciting developments for the chemical information community by Pistoia Alliance going on right now, in your view?

CN: It would have to be [HELM](#), the Hierarchical Editing Language for Macromolecules. HELM is both a notation, and a set of open-source tools and applications that implement the notation. The notation enables representation of a wide variety of biomolecules, from proteins, to nucleotides, to antibody drug conjugates, allowing for easy exchange of data. This representation has become the de-facto standard. It is supported by a variety of software vendors, and is available in both the ChEMBL and PubChem database records.

I should also mention that our members are about to release into the public domain, via the Protein Data Bank (PDB), a set of previously internal antibody crystal structures as part of the Pistoia Alliance [AbVance](#) project. The ultimate goal of this project is to improve antibody predictive modeling, and data sharing is one plank of that effort.

SK: Let’s move on to the subject of Open Access. In 2014, Steven Bachrach and you wrote a chapter, “Tying It All Together: Information Management for Practicing Chemists”, for an ACS symposium book, [The Future of the History of Chemical Information](#), where you pointed out key problems regarding difficulty in reusing data content from supporting materials archived as PDF files and quality curation of data depositions at open access repositories. Has much changed since that publication? Are there any exemplary endeavors?

CN: I am not fully up to date on this topic, but certainly we are seeing growing efforts to deal with the data deposition challenge. Many journals and funding organizations are now requiring data deposition along with paper publication, so we are definitely beyond the “should we do this” stage. Progress also has been made in assigning persistent identifiers to data sets, which is a prerequisite to successful data sharing efforts. In fact, the RDA plenary meeting is this week (Sept 19-21, 2017), during which the [Persistent Identifier Interest \(PID\) Group](#) will be discussing the progress, the gaps, and the growing community initiatives around PID. And of course the commercial efforts like [FigShare](#) are maturing. I guess I am a bit disappointed that ChemSpider has not taken more of a leadership role here.

SK: Carmen, I remember vividly your leadership role for the Division of Chemical Information becoming a supporter of the InChI Trust in 2010. You invited a guest presenter to the CINF executive committee, collaborated with ACS for their clearance, attended the InChI Trust board meetings, and wrote updates of the InChI projects for Chemical Information Bulletin. What made you want to focus on that? Are there specific contributions or recent activities you would like to discuss?

CN: Data exchange and reuse is key to scientific collaboration and discovery. Within CINF we understand what needs to be done and we have the obligation to take a leadership role in advancing initiatives that promote such data exchange and collaboration. InChI was one such initiative, and at the time I thought it was important for us to support this open community effort. InChI is now over ten years old, and has become the open standard for small molecule data exchange. This year we saw the release of the first version of the reaction InChI, known as RInChI, which is very exciting. We would like to adopt the RInChI for the Chemical Safety Library, because this would be a perfect application.

SK: Another outstanding contribution to the Division was your organizing a series of [webinars](#) during 2014-16. In collaboration with Belinda Hurley for hosting, you invited over a dozen prominent guest speakers from beyond the CINF scientific information community to share their expertise for the benefits of division members. Let's imagine that ANY guest speaker could agree to participate. Who would you be interested in hosting?

CN: I am glad to see that we are reviving the webinar series, because it is an excellent current awareness vehicle to support our members beyond the national meetings. I guess if I could invite anyone, I would want to do a series on data access and sharing where we brought in senior leaders of leading information institutions and corporations: the Librarian of Congress, the head of the National Library of Medicine, the head of the USPTO, CIO of Google, folks like that. Maybe we should try that!

SK: In conclusion, let me ask a couple of personal questions. What does a typical day in the office of Carmen Nitsche look like?

CN: I have been working remotely since 2001. So my days either start with that long 30 foot walk to my home office, or a ride to the airport. Since many of my clients are in Europe, I do find myself on early

calls frequently (I did myself a favor by moving from Texas to New Jersey, which cut down on those 6 am calls). I spend most days in GoToMeetings, or on the phone.

SK: Please tell us what you like to do when you aren't working?

CN: I love to cook and entertain, and am always looking for that next new recipe to add to the party rotation. I also love traveling. We moved to the Jersey Shore last year. This is our first time living on the East Coast. So we are enjoying exploring our new locale, heading to NYC and Philly frequently, and taking in the beautiful New Jersey countryside.

SK: Thank you, Carmen, for fostering collaborations in developing useful tools in the interest of increased safety in the chemical enterprise. We will look forward to learning more at your symposium titled, "Community Sharing of Chemical Safety Data: Yes, No, Maybe?" being organized for the next ACS national meeting in New Orleans, LA, March 18-22, 2018.

Carmen Nitsche's presentations at recent ACS national meetings:

- (1) Nitsche, C; Whittick, G.; Manfredi, M. Reaction Safety Information: Engaging the Community in Collecting and Sharing of Safety Learnings. Spring 2017, San Francisco, CA; CHAS-32. ([slides at CHAS; symposium](#))
- (2) Nitsche, C. Data Sharing and beyond: Lessons Learned from the Life Sciences Industry. Fall 2016, Philadelphia, PA; COMP-79.
- (3) Nitsche, C. Data Sharing in Life Sciences R&D: Pre-competitive Collaboration through the Pistoia Alliance. Spring 2016, San Diego, CA; CINF-68. ([Presentation on demand](#))
- (4) Nitsche, C. Pre-competitive Collaboration to Advance Laboratory Safety. Spring 2016, San Diego, CA; CHAS-41.
- (5) Bachrach, S.; Nitsche, C. Tying it all together: Information management for practicing chemists. In *Future of the History of Chemical Information*; McEwen, L., Buntrock R., Eds.; ACS Symposium Series 1164; American Chemical Society: Washington, DC, 2014; pp 255-268.
- (6) Nitsche, C.; Taylor, K. InChI Names and Keys: Do They Add Value to Commercial Software and Databases. Spring 2012, San Diego, CA; CINF-102.
- (7) Taylor K.; Hassan M.; Foss D.; Nitsche C. Interactive Prediction of Biological Activity. Fall 2011, Denver, CO; CINF Flash - Lightning Talks.
- (8) Nitsche, C. One Search, Many Answers: Bringing Together Results from Multiple Databases through the DiscoveryGate Platform. Fall 2009, Washington, DC; CINF-61. ([slides](#))

Technical Program

Informatics and Chemical Biology: Identifying Targets and Biological Pathways

CINF symposium at the Fall 2017 ACS Meeting in Washington, DC

Rachelle J Bienstock

With the increasing availability of genomic information and biological expression data, one of the current challenges in drug discovery is linking biological pathway data with small-molecule drug data. How can drug pathway target information and metabolic pathway information be linked to small ligand information? These are some of the issues and questions addressed by the CINF symposium “Informatics and Chemical Biology: Identifying Targets and Biological Pathways”, at the Fall 2017 ACS Meeting in Washington, DC.

David Sheen, NIST, (National Institute of Standards and Technology), started off the symposium by discussing incompatibilities with metabolic data reported by different groups, and the need to have improved databases and data harmonization methods to address varying uncertainties in reported experimental metabolic data. This will enable comparison of data from different laboratories and sources. NIST maintains <http://qmet.nist.gov>, the quality assurance program in metabolomics, to encourage exchange of spectral data and comparison of uncertainties in measurement, and is conducting literature interlibrary comparison studies. Reproducibility analysis for spectral data is an issue as well.

Dr. Karina Martinez Mayorga, Instituto de Quimica, UNAM, reported using the PLIF (Chemical Computing Group Protein Ligand Interaction Fingerprints) method for screening biased ligands for opioid receptors. There are approximately 800 opioid receptors that are members of the G-protein coupled receptor (GPCR) family and they are significant targets for pain management. Databases of these interaction fingerprints, combined with methodologies to identify structural traits for selective agonists, will lead to the successful development of drugs with fewer side effects.

Dr. Doug Selinger, Plex (www.plexresearch.com), discussed development of a search engine for chemical biology and drug discovery. The search engine begins with a query molecule and expands to more compounds with similar chemical structures and biological transcriptional profiles. Compound-compound and compound-target relationships are used in search algorithms to rank compounds and targets. Data sources include: Open Targets, PubChem, Entrez Gene, chemical similarity, and ChEMBL bioactivities. Plex as a search engine, searches data (1.7 billion rows of data), not Web pages. The search engine can search compounds, targets, or pathways; InChIs, SMILES, and structures can be drawn directly into the search bar. The more datasets included in the search engine, the better the search engine gets at providing answers.

Dr. Anne Wassermann, Merck Informatics, discussed the chemical probe databases: libraries of small molecules with known targets, which permit the development of correlations between chemical and mechanistic properties. She discussed generating target hypotheses for molecules through the use of biologically annotated libraries. The Chemical Probes Portal (<http://chemical.probes.org>) is one example of a publicly available database of probes. Merck is working on Web applications that can be used to relate phenotypes and protein targets and biological pathways.

Way2drug (www.way2drug.com), a cheminformatics platform for drug repurposing, was discussed by **Vladimir Poroikov**, Institute Biomed Chemistry, Moscow. This platform provides drug-target interaction predictions, toxicity predictions, and predictions of the effects of drugs on gene expression. The PASS (Prediction of Activity Spectra for Substances) dataset includes information on predicted biological activity spectra, MPDS (<http://mpds.osdd.net>) provides molecular property predictors, and MetaTox gives metabolic predictions. Way2drug has links to the *Kyoto Encyclopedia of Genes and Genomes* (KEGG), PDB, and Thompson Reuters Integrity databases.

Safety and toxicity are among the most significant drug development issues. **Matthew Clark**, Elsevier, discussed the development of bioassays as predictors of adverse events in clinical trials. FDA submissions, a large number of journals, and Open PHACTS were used as data, looking for relationships between bioactivity and toxicity, the goal being the development of methods for corroborating evidence from pathway analysis for prediction of important targets.

The session concluded with presentations by two groups on deep learning neural network (DNN) applications for small molecule drug discovery. **Dr. Abraham Heifets**, Atomwise, <http://www.atomwise.com/>, gave a presentation on developing predictive models for drug mechanism of actions using deep convolutional neural networks. Deep neural networks are constrained neural networks. AtomNet is a structure-based DNN for molecule bioactivity prediction, which uses a nearest-neighbor structure-based binding algorithm. Atomwise is working on developing these methods, and Abraham presented some results and benchmarks based on their efforts. **Antonio de la Vega de Leon**, The University of Sheffield, gave a presentation on deep neural networks to predict the activity in a specific screen, and to suggest which target hits the compound. The machine learning algorithm was based on assay description, biological pathway data, and ChEMBL bioactivity data.

Neural network methods show significant promise in their ability to make extensions and predictions based on learning sets and data. With more data available, improved learning sets, and better algorithms to develop correlations, predictions of phenotype, pathways, and targets with small molecule structure and chemical properties data will greatly improve.

Collaborating for Success:

Professional Skills Development for Undergraduates, Graduates, and Post-Docs

CINF symposium at the Fall 2017 ACS Meeting in Washington, DC

Jeremy Garritano and Elsa Alvaro, symposium organizers

Employers in every sector seek to hire employees that have a variety of skills and talents. Though there are not always standard definitions, effective communication, critical thinking, creativity, initiative, and adaptability continually rank high in surveys related to desirable skills employers seek. However, these skills are not always part of the academic experience. With only a small percentage of STEM graduates securing tenure-track positions, expanding the training to cover these areas can have a great impact on their careers as future STEM professionals.

The symposium “Collaborating for Success: Professional Skills Development for Undergraduates, Graduates and Post-Docs” took place Monday, August 21, 2017. It explored the professional development needs of undergraduates, graduate students, and postdoctoral researchers in chemistry and other STEM fields. It also revolved around ways that institutions, graduate programs, funders, professional societies, and libraries are contributing to their success.

Several talks focused on the professional development needs of graduate students from a programmatic approach. **Laura Regassa** and **Nadeene Riddick** of the National Science Foundation (NSF) spoke about the NSF Research Traineeship (NRT) program, which encourages new models in STEM graduate education. They discussed the professional development skills that have been more often addressed in NRT projects: science communication, including oral, written, and digital communication; mentoring (both faculty and student mentoring); career preparation, such as internships, networking, and career paths; and research ethics, including responsible conduct of research and ethics of data acquisition and management.

Also from a broad perspective, **David Zwicky** described a needs assessment project aimed at understanding how to support graduate students at Purdue University. The overall needs that surfaced were: 1) professional development, including teaching, building professional identity, coding, communicating professionally, and project management; 2) spaces, such as spaces for collaboration and research; and 3) information resources, data services, and software.

Rigoberto Hernandez of Johns Hopkins University addressed the topic of diversity and equity in chemistry departments, and talked about the Open Chemistry Collaborative in Diversity and Equity (OXIDE). OXIDE works with department chairs to reduce inequitable diversity barriers to career advancement through National Diversity Equity Workshops (NDEW) that facilitate discussion between department chairs, federal agency representatives, and diversity policy leaders.

Danielle Watt of the Chemistry at the Space-Time Limit (CASTL) Center, one of the NSF Centers for Chemical Innovation (CCI), discussed how CCIs train STEM students in leadership. Danielle described professional development needs identified by trainees and how they are addressing those needs. Specific examples include training in innovation, collaboration, and effective communication.

The symposium also focused on the professional development needs of undergraduate students. **Thomas Wenzel** of Bates College, who chairs the ACS Committee on Professional Training (CPT), addressed the importance of skills development on the ACS certified bachelor's degree in chemistry. The [2015 guidelines](#) state that "programs must provide experiences that go beyond chemistry content knowledge to develop competence in other critical skills necessary for a professional chemist". These skills include problem solving, chemical literature and information management, laboratory safety, teamwork, communication, and ethics.

The symposium balanced these broad, programmatic perspectives by including talks that described examples of developing specific skills. **Donna Wrublewski** of Caltech discussed her experience organizing Data Carpentry workshops to teach programming skills to scientists and engineers. **Ron Kaminecki** focused on the development of courses on patent information research and analysis to equip students who have a scientific background with practical skills in patent research. **Svetla Baykoucheva** of the University of Maryland described the implementation and assessment of a program aimed at helping students develop information literacy skills, including finding, managing, and sharing scientific information. From a database provider's perspective, **Mindy Pozenel** from Chemical Abstracts Service described their recently created Chemical Class Advantage (CCA) modules for instructors to use in organic chemistry courses. These modules encourage students to use SciFinder to discover the scientific literature, as well as provide opportunities for students to demonstrate their ability to read those articles while quizzing them on the content of the articles. **Megan Sheffield** (Clemson University) and **Marguerite Savidakis-Dunn** (Shippensburg University) spoke about developing data management skills in chemists. **Rachel Borchardt** of American University described the major metrics available to chemists, including journal Impact Factors, citation distributions, and altmetrics, and discussed the importance of mastering those metrics to influence the research evaluation narrative. Along the same lines, **Antony Williams** of EPA talked about the importance of creating an online presence, and the free tools available for that purpose; specific examples include LinkedIn; Slideshare and Google Scholar to track publications and citations; ResearchGate for networking and citation tracking; Publons for getting credit for reviewing papers; Kudos; Figshare; and altmetrics tools such as ImpactStory and Altmetric scores.

Developing safety skills was addressed in presentations by **Joseph Pickel** of Oak Ridge National Laboratory and **Samuella Sigmann** of Appalachian State University. Joe described his experience transitioning to a safety officer position after a career as a scientist, and discussed the skills required in a research operations position. Sammie spoke about embedding safety professionals to engage with faculty and help educate undergraduate students and develop their critical thinking skills.

The importance of communication skills also featured prominently in some of the presentations. **Christin Monroe** of Princeton University discussed the Science Communication Education Network

(SCENe). A collaboration with the NSF-funded communication program Portal to the Public (PoP) National Network, this workshop seeks to develop communication skills of scientists, including their ability to engage with different audiences, and build confidence as communicators. **Kiyomi Deards** of the University of Nebraska Lincoln reported several ways to engage with a wide audience and facilitate broader impacts; high commitment examples that Kiyomi described include Sci Pop talks and partnering with the Undergraduate Research Council to showcase undergraduates' research and creative work.

Finally, expanding career opportunities for STEM graduates was also a recurrent theme of the sessions. **Amy Clobes** of the University of Virginia and **Natalie Lundsteen** of The University of Texas Southwestern Medical Center described the work of the Graduate Career Consortium (GCC) organization, which helps members provide career and professional development for doctoral students and postdoctoral scholars. They discussed several specific resources for occupational exploration, and also ways to incorporate these tools into the work of librarians, including engaging with campus partners or collaborating with GCC. There were also two talks from professional societies with a careers focus. **Shannon O'Reilly** of the ACS discussed how the ACS on Campus program has evolved since 2010 to meet the career and professional development needs of students and faculty in the chemical sciences. The program has become increasingly modularized as well as expanding out to international audiences. **Scott Nichols** of AAAS gave an overview of AAAS Professional Development & Career Services, including myIDP, which is an individual development plan to help explore career possibilities and set goals, and the AAAS Career Development Center resources.

To conclude, the symposium offered a nearly comprehensive overview of the many approaches to contribute to the professional development of STEM students and graduates. The combination of programmatic approaches and case studies focused on specific skills was particularly enriching, and encouraged a very positive engagement among the different speakers and the audience.

What do synthetic chemists want from their reaction systems?

CINF symposium at the Fall 2017 ACS Meeting in Washington, DC

Wendy Warr, symposium organizer

David Evans and I organized a CINF symposium at the fall 2017 ACS national meeting. We had sought talks on progress in reaction searching, reaction planning, synthesis design, retrosynthesis, and reaction prediction. We would really have liked contributions from practicing synthetic chemists on their current needs, both met and unmet, and their frustrations with current systems, but no end users volunteered. Nevertheless, it was an interesting symposium and I have received positive feedback.

Academic research

Connor Coley of MIT was the first speaker. A critical challenge for computer-assisted synthesis design is that the reaction steps proposed may fail when attempted in the laboratory. The true measure of success for any synthesis program is whether the predicted outcome matches what is observed experimentally. Connor and his co-workers have trained a neural network model on experimental data from the USPTO and Reaxys to provide qualitative predictions of organic reaction outcomes *in silico*. In this method reaction databases are supplemented with chemically plausible negative reaction examples to overcome the literature bias towards successful reactions. Traditional reaction templates are used to generate a list of candidate outcomes for the machine learning model to score, so reactivity rules are implicitly learned rather than encoded. A new, edit-based reaction representation has been developed to focus on the fundamental transformation at the reaction site. In a 5-fold cross-validation, the trained model assigns the major product rank 1 in 71.8% of cases, rank ≤ 3 in 86.7% of cases, and rank ≤ 5 in 90.8% of cases.¹ Connor presented some correct and incorrect predictions. Mispredictions are often chemically reasonable or attributable to data quality issues. Extension of the method to condition-dependent predictions achieves similar performance, but conditions are rarely necessary to make the prediction. Multi-step pathway planning remains challenging.

Mark Waller of the University of Muenster and Shanghai University has also used neural networks, but in this case deep neural networks, in both retrosynthesis and reaction prediction.^{2,3} The machine is trained with essentially the complete published knowledge of organic chemistry (more than 3.5 million reactions acquired from the Reaxys database). Circular fingerprints are used to represent the structures. Training can be carried out overnight with GPUs, and retraining can be carried out weekly. The approach has a higher than 95% accuracy when allowed to suggest up to 10 different routes for a target molecule on a test set of around one million reactions. Deep learning is 150 times faster than a rule-based approach, so handling multistep syntheses becomes feasible. Furthermore, preliminary studies indicate that coupling the neural networks with Monte Carlo tree search techniques outperforms traditional computational synthesis planning with hand-coded transformations.^{4,5}

The international chemical identifier for reactions

The next two talks concerned the International Chemical Identifier for Reactions (RInChI). **Gerd Blanke** of StructurePendium Technologies explained that RInChI is a single string providing a unique representation of a reaction, independent of how the reaction has been drawn. The Long-RInChIKey is calculated from the IUPAC International Chemical Identifiers (InChIs) of each reactant, product and reagent. The Short-RInChIKey is a fixed-length hash over all reagents, products and agents. Web-RInChIKey is a fixed-length hash developed from the reaction components, but ignoring the specific role within the reaction.

Long-RInChIKeys are valuable for the storage of reactions. They allow uniqueness checks, and the identification of each reaction component by simple text searches based on Standard InChIKeys, but they do not have a fixed length. Short-RInChIKey has a fixed length of 55 letters, plus 8 hyphens as separators. The fixed length of Short-RInChIKey makes it suitable for exact searches of reactions in databases (and on the Web), indexing reactions in databases, and linking identical reactions in different databases. Web-RInChIKey allows for the fact that the depiction of a chemical reaction is not uniquely defined. For Web-RInChIKey, all InChIs of the reaction components are ordered alphabetically. Roles of the components are ignored. The Web-RInChIKey has a fixed length of 47 characters, with 17 letters in the major layer, and 15 letters in the minor layer. It is used for searches over reaction databases with an unknown drawing model, and comparison of reaction databases with different data models. The longer string sets for the major and minor layers make searches over the Web more precise. The first RInChI release was in March 2017. The InChI and RInChI formats and algorithms are non-proprietary, and the software is [open source](#). RInChIs for 4.5 million reactions from the SPRESI database have been generated by InfoChem: only 239 reactions could not be converted.

Jonathan Goodman of the University of Cambridge started his talk with an example of an *in silico* inspired⁶ total synthesis of (-)-Dolabriferol.⁷ Synthetic chemists want data that are accessible, comprehensive, and reliable. InChIs are successful because people use them. Can RInChI be useful too? A good synthesis uses cheap, sustainable, and reproducible starting materials; has low hazards; produces low waste products; uses familiar reactions, and chemists' expertise; has no inseparable by-products; gives high yields and high stereoselectivity; uses convenient processes; makes a product quickly, cheaply, and reproducibly; and is suitable for making analogues.

Jonathan believes that to achieve a good synthesis, we need to understand our reactions, to make best use of our analytical data, to search the literature effectively, and to store our results, so we, and others, can make best use of this knowledge for the next project and the next molecule. The contributions of Jonathan's team to experimental chemistry, computational chemistry, and chemical informatics have helped advance all of these areas. Jonathan presented some examples of work that his team has done on the automatic generation of diastereomers using InChI strings,⁸ prediction of stereochemistry,⁹ the conformational properties of a polypeptide,¹⁰ and the risk assessment of chemicals.¹¹ It is desirable to bring these disparate fields together, so that a single reaction system can enable users to benefit from

them all. Using RInChI, we can connect diverse data to individual reactions. Jonathan concluded with an amusing vision of the future synthesis machine.

Search and faceting of large reaction databases

The next talk was by **John Mayfield** of NextMove Software. Synthetic chemists want data, diagrams, classification and search for their reaction systems. Workers at NextMove have previously described the extraction of reactions from patents. [LeadMine](#) and Chemical Tagger convert unstructured text to a structured reaction table. NextMove have also assembled over six million extracted reaction details consisting of the connection tables, procedure, quantities, solvents, catalysts and yields into a [searchable ELN](#) for multiple pharmaceutical companies. Good reaction diagrams are essential in communicating synthetic chemistry: NextMove has also done work in this field. In the area of classification, [NameRXN](#) software allows the recognition and categorization of reactions from their connection tables. Using a large rule-base of known reaction mechanisms and transformations, NameRXN is able to categorize reactions to a NameRXN code.¹² Reactions are classified and assigned to leaves in the RXNO ontology. The ontologies are used to provide organization, faceting, and filtering of results. [Pistachio](#) is a reaction dataset interface providing loading, querying, and analytics of chemical reactions. NextMove's Arthor technology is reportedly up to 100 times faster than other "fast search" systems.

The history of chemical reactivity

Guillermo Restrepo of the University of Leipzig showed that a computational approach to the history of chemical reactions sheds light on the patterns behind the development and use of substances and reaction conditions along two centuries. He and his co-workers have explored more than 45 million reactions in Reaxys and revealed historical patterns for substances, types of substances, catalysts, solvents, temperatures, and pressures of those reactions. Reaxys was treated as a graph database. Despite the exponential growth of substances and reactions, little variation of catalysts, solvents, and reactants is observed throughout time. The vast majority of reactions fall into a narrow domain of temperature and pressure. World wars caused a drop in chemical novelty for substances and reactions. The First World War took production back around 30 years and the Second around 15. After the Second World War, the use of organic solvents skyrocketed. Guillermo anticipates that this study, and especially its methodological approach, will be the starting point for the history of chemical reactivity, where social and economic contexts are integrated.

SciFinderⁿ and ChemPlanner

The next two papers concerned work that CAS is doing to enhance Wiley's ChemPlanner^{13,14} with additional reaction content and associated references, including reactions from patents. A new version of ChemPlanner, including stereoselective retrosynthetic prediction and customizable relevance ranking, will be delivered exclusively in SciFinderⁿ. **Orr Ravitz** spoke first, largely concentrating on ChemPlanner itself. Chemists use ChemPlanner to boost creativity, overcome biases, and cover more options.

Previous perceptions of retrosynthesis have been skepticism, fear of overload of information, and concerns about the coverage and the currency of the reaction database, and about accuracy and selectivity. Orr discussed automatic rule generation. Deriving selectivity from data requires statistical power, which is not always sufficient with a database such as CIRX. Literature examples, sorted by similarity to predictions, provide insight into experimental conditions, and enhance user confidence. Greater coverage is expected by using *Chemical Abstracts* data instead of CIRX. A nearly exhaustive reaction source will have many variations on the same reaction, or the same reaction with very similar reactants and products. Growth of the rule set will be significantly sublinear. Adding examples to existing rules will address functional group tolerance, give more statistical power for regioselectivity calculations, more automation for stereoselective rules, and improved yield prediction. There will be some consolidation of rules.

Jonathan Taylor of CAS started his talk with an introduction to SciFinderⁿ. Everything about SciFinderⁿ is new: the interface, the application, the search architecture and the data model. User feedback and usability testing were critical in the design. Layout and first surface information were users' main priorities. The final design balances surfaced information, aesthetics, and browsability and filter options. In the past, synthetic chemists wanted reaction finding tools, today they have synthetic planning tools, and in future they will have help with predictive synthetic routes: SciFinderⁿ will deliver new predictive synthesis planning capabilities by integration of an enhanced ChemPlanner. Having ten times the reaction content will provide ChemPlanner with more synthetic options to build pathways and improve prediction quality. Jonathan concluded with some screen mockups of user input, of how SciFinderⁿ will propose potential synthetic routes, and of how users will know how the prediction was constructed.

Reaction classification

Next, **Valentina Eigner-Pitto** of InfoChem spoke about the renaissance of reaction classification and visualization. InfoChem's *ICMAP* reaction mapping software identifies reaction centers. The CLASSIFY¹⁵ software automatically categorizes a reaction according to the type of chemical transformation, and it can be used for organization of large reaction databases and hit lists. It provides unique identifiers (ClassCodes) that can be used in reaction database analysis. This allows companies to study the kind of chemistry performed in-house, to examine the evolution of chemistry over time, and to compare in-house content with other repositories. Classification can also be used in network graphs, which can be used as visualization tools for reaction content. Workers at Merck KGaA, in collaboration with BioSolveIT and InfoChem, have demonstrated a workflow which exploits the chemist's electronic laboratory notebook (ELN) in order to obtain and refine transforms for existing and novel chemical transforms,¹⁶ which in turn are used to enrich existing virtual libraries. The novelty of the added chemical space is assessed through a multitude of descriptors with a particular focus on three-dimensionality, scaffold diversity, and fingerprint enrichment. Additionally, each added transform is evaluated for its propensity to reconstitute known drugs and chemical probes. Computer-aided synthesis design programs include ChemPlanner and InfoChem's¹⁷ *ICSYNTH*. Prediction of chemical space (forward reaction prediction) is also illustrated in the Merck poster.¹⁶

Use of Reaxys and ReaxysTree

Two papers followed from experts at Elsevier. **Juergen Swienty Busch** discussed ReaxysTree and the taxonomies used in Reaxys. He began with an exposition of new Reaxys, before turning to the taxonomies. Reaxys has information on documents, substances, reactions, and substance properties, and on bioactivities and targets in Reaxys Medicinal Chemistry Index. For documents, terms from ReaxysTree, Embase, Compendex, and Geobase make search and analysis possible on ReaxysTree. For substances, analysis is possible on substance classes and available properties. For reactions, search and analysis is possible on reaction classes, catalyst classes, and solvent classes. For targets, search and analysis is possible on gene and protein taxonomy, organisms, cell lines, and administrative route. Substances have been curated by Richter classes, rings, and functional groups. Solvents, reagents, and catalysts have been curated for reactions. ReaxysTree allows concepts and synonyms to be used for search, filtering, analysis, and indexing. ReaxysTree concepts for reactions include name reactions, and classes and types such as cyclization, condensation, and addition. Juergen next outlined how reaction mapping is carried out, a transition state is assigned, and the transform is coded. In searching reactions with ReaxysTree, taxonomy terms are connected with actual Reaxys queries using transforms, and other appropriate search terms such as product substructures.

Matt Clark of Elsevier thinks that medicinal chemists themselves only want to find transformation details for chosen steps in synthesis, while management wants to lower the cost of making compounds, and wants reliable reaction schemes that can be sent to a contract research organization (CRO) for fast turnaround. Reaxys is a treasury of reported chemistry, with a built-in synthesis planning tool and display of experimental procedures. The API allows you to use similarity for compounds and reactions, access some data elements not visible in the user interface, and create your own analytics and reaction networks. Pipeline Pilot and KNIME offer an easy way to use the API, and offer interoperability with other software products.

Matt discussed a reaction graph analysis application to address questions around a specific potential CDK8 inhibitor. What chemistries are known about compounds like this? What conditions and solvents were used by different chemists? Where is this chemistry reported? Ultimately, what are the most efficient and flexible methods to make compounds like this? The application involves searching for reactions with the target compound as product, and similarity search for very similar compounds, and then searching for reactions using the reactants as the product, and then repeating for the desired graph depth. An interesting finding was that for very similar compounds, different chemistries and starting materials have been used. One tree showed a set of compounds that used a common set of starting materials. Using Cytoscape you can drill down to references for each edge. You can compare intermediates for similar compounds made by different groups and, by accessing Scopus, examine a network of institutions publishing a specific chemistry.

Using Reaxys you can also analyze reaction conditions, grouping known transformations at different levels of detail to get the best conditions. Grouping uses reaction similarity, based on Reaxys

transformation codes. Searching for “Buchwald-Hartwig Aminations” by keyword produced 4,179 results. These were grouped by transformation codes, from general to specific: level 0 had one group with 4,179 members, level 1 had 99 groups, level 2 had 160 groups, and so on. A summary of solvent and conditions for level 0 showed that toluene is a popular solvent, a temperature of around 110°C is common, reaction time is not very long, and inert atmosphere and microwave use were mentioned. These conditions can be selected based on membership in one of the other groupings.

An expert searcher’s viewpoint

The final speaker was **Judith Currano** of the University of Pennsylvania. Introducing variable substituents during a reaction search is challenging. A researcher may not have a definite substituent in mind, instead suggesting that a site can be occupied by “any aryl group” or, still worse, “any electron withdrawing group”. Even a researcher who generates an R-group and populates it with specific substituents can run into problems because atom mapping from reactant to product is prohibited within R-group fragments. Judith used the term “specific ambiguity” to talk about a type of attachment without specifying exactly what it is. This includes general classes of attachments, user-defined groups of attachments (variables or R-groups), and stereocenters where you do not care about the identity of all of the attachments. She presented case studies based on troublesome requests from synthetic chemists.

The first examples concerned functional group transformations (plus mapping from reactant to product), and sensitive functional groups. Searchers should understand that sometimes a review source is worth a thousand searches. (*Science of Synthesis* was good for one example.) Searchers should also use caution when employing mapping. Database vendors should perhaps give users the ability to make mapping less atom-specific and more atom-type-specific. Structure search algorithms should have a way of manually grouping fragments that appear in the same reactant or product, allowing the searcher to specify multiple fragments in one substance while allowing additional substances on that side of the equation. (Old Beilstein Crossfire worked well in one of Judith’s examples.)

The second set of examples involved specific ambiguity of stereocenters or variables. Judith recommends searchers make use of system-defined generics whenever possible. In the case of user-defined generics, it may be necessary to run multiple searches if your generic does not exist. Vendors should note that all structure search algorithms should permit stereocenters containing system- or user-defined variables, and all search algorithms should permit stereo-specific reaction searches.

Finally Judith discussed reaction searches involving both specific transformations and specific ambiguity (mapping R-groups, mapping variables, and including the elusive electron-withdrawing group). She warns users that if it is essential that they map a user-defined R-group from reactant to product, they should be prepared to do multiple searches for the various substances represented. Database vendors should note that adding generics like electron withdrawing groups would make users very, very happy.

Acknowledgments

My thanks to all the speakers for their interesting contributions, and for providing me with copies of their slides, allowing me to study the talks in more depth, and, ultimately to include more detailed summaries in my meeting report. My thanks also to Matt Clark for handling all the PC and projector issues so that I could concentrate on introducing speakers, on handling questions, and, above all, on being stimulated by the interesting science.

References

- (1) Coley, C. W.; Barzilay, R.; Jaakkola, T. S.; Green, W. H.; Jensen, K. F. Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Cent. Sci.* **2017**, *3* (5), 434-443.
- (2) Segler, M. H. S.; Waller, M. P. Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. *Chem. - Eur. J.* **2017**, *23* (25), 5966-5971.
- (3) Segler, M. H. S.; Waller, M. P. Modelling Chemical Reasoning to Predict and Invent Reactions. *Chem. - Eur. J.* **2017**, *23* (25), 6118-6128.
- (4) Segler, M. H. S.; Preuss, M.; Waller, M. P. Learning to Plan Chemical Syntheses. 2017, arXiv.org e-Print archive. <https://arxiv.org/abs/1708.04202> (accessed September 22, 2017).
- (5) Segler, M. H. S.; Preuss, M.; Waller, M. P. Towards "AlphaChem": Chemical Synthesis Planning with Tree Search and Deep Neural Network Policies. 2017, arXiv.org e-Print archive <https://arxiv.org/abs/1702.00020> (accessed September 22, 2017).
- (6) Socorro, I. M.; Goodman, J. M. The ROBIA Program for Predicting Organic Reactivity. *J. Chem. Inf. Model.* **2006**, *46* (2), 606-614.
- (7) Currie, R. H.; Goodman, J. M. In Silico Inspired Total Synthesis of (-)-Dolabriferol. *Angew. Chem., Int. Ed.* **2012**, *51* (19), 4695-4697.
- (8) Ermanis, K.; Parkes, K. E. B.; Agback, T.; Goodman, J. M. Expanding DP4: application to drug compounds and automation. *Org. Biomol. Chem.* **2016**, *14* (16), 3943-3949.
- (9) Reid, J. P.; Simon, L.; Goodman, J. M. A Practical Guide for Predicting the Stereochemistry of Bifunctional Phosphoric Acid Catalyzed Reactions of Imines. *Acc. Chem. Res.* **2016**, *49* (5), 1029-1041.
- (10) Fedorov, M. V.; Goodman, J. M.; Schumm, S. To Switch or Not To Switch: The Effects of Potassium and Sodium Ions on α -Poly-L-glutamate Conformations in Aqueous Solutions. *J. Am. Chem. Soc.* **2009**, *131* (31), 10854-10856.
- (11) Allen, T. E. H.; Goodman, J. M.; Gutsell, S.; Russell, P. J. A History of the Molecular Initiating Event. *Chem. Res. Toxicol.* **2016**, *29* (12), 2060-2070.
- (12) Schneider, N.; Lowe, D. M.; Sayle, R. A.; Landrum, G. A. Development of a Novel Fingerprint for Chemical Reactions and Its Application to Large-Scale Reaction Classification and Similarity. *J. Chem. Inf. Model.* **2015**, *55* (1), 39-53.
- (13) Law, J.; Zsoldos, Z.; Simon, A.; Reid, D.; Liu, Y.; Khew, S. Y.; Johnson, A. P.; Major, S.; Wade, R. A.; Ando, H. Y. Route Designer: A Retrosynthetic Analysis Tool Utilizing Automated Retrosynthetic Rule Generation. *J. Chem. Inf. Model.* **2009**, *49* (3), 593-602.
- (14) Cook, A.; Johnson, A. P.; Law, J.; Mirzazadeh, M.; Ravitz, O.; Simon, A. Computer-aided synthesis design. 40 years on. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2* (1), 79-107.

- (15) Kraut, H.; Eiblmaier, J.; Grethe, G.; Loew, P.; Matuszczyk, H.; Saller, H. Algorithm for reaction classification. *J. Chem. Inf. Model.* **2013**, *53* (11), 2884-2895.
- (16) Knehans, T.; Klingler, F.-M.; Kraut, H.; Saller, H.; Herrmann, A.; Rippmann, F.; Eiblmaier, J.; Lemmen, C.; Krier, M. *Merck Accessible Inventory (MASSIV): In silico synthesis guided by chemical transforms obtained through bootstrapping reaction databases*, Abstracts of Papers, 254th ACS National Meeting & Exposition, Washington, DC, USA, August 20-24, 2017; American Chemical Society: Washington, DC, 2017; COMP 283.
- (17) Bøgevig, A.; Federsel, H.-J.; Huerta, F.; Hutchings, M. G.; Kraut, H.; Langer, T.; Löw, P.; Oppawsky, C.; Rein, T.; Saller, H. Route Design in the 21st Century: The ICSYNTH Software Tool as an Idea Generator for Synthesis Prediction. *Org. Process Res. Dev.* **2015**, *19* (2), 357-368.

Herman Skolnik Award Symposium 2017

Honoring David Winkler

Wendy Warr (wendy@warr.com) for the ACS CINF *Chemical Information Bulletin*

Introduction

David Winkler, CSIRO Fellow, and professor at Latrobe Institute for Molecular Science, and Monash Institute of Pharmaceutical Sciences, Melbourne, Australia, received the 2017 Herman Skolnik Award for his seminal contributions to chemical information in the development of optimally sparse, robust machine learning methods for QSAR, and in leading the application of cheminformatics methods to biomaterials, nanomaterials, and regenerative medicine. A [summary of his achievements](#) has been published in the *Chemical Information Bulletin*. David was invited to present an award symposium at the Fall 2017 ACS National Meeting in Washington, DC. He invited six speakers:



L to R: Alex Tropsha, Johnny Gasteiger, Yoram Cohen; Tim Clark, David Winkler, Ceyda Oksel, Tudor Oprea

Approaching reality: simulating electronic devices

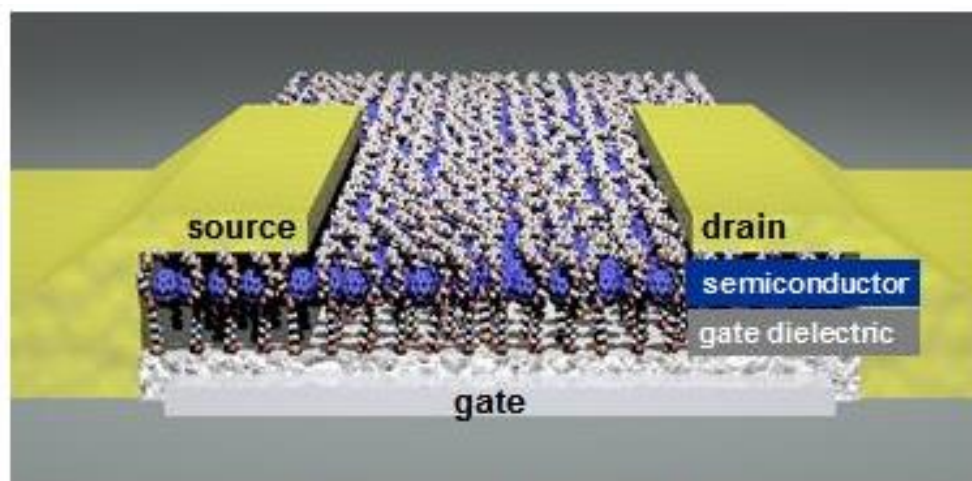


Tim Clark, of the University of Erlangen-Nürnberg, was the first speaker. The impact of modern hardware and software on simulations has not been an issue of doing things faster and faster, but rather one of doing calculations that we could not do before. *Ab initio* calculations can now be done on compounds with several hundred atoms, density functional theory calculations on a few thousand atoms, and semiempirical molecular orbital (MO) calculations on 100,000 atoms. Simulations of several microseconds are now standard.

Semiempirical (neglect of diatomic differential overlap, NDDO) molecular orbital (MO) calculations without local approximations are now possible for 100,000 atoms or more with the massively parallel semiEMPIRical molEcular-Orbital Program (EMPIRE) program,¹⁻³ which is [freely available](#) to academic groups. Calculation scales with approximately $N^{2.5}$. We are no longer limited to small or homogeneous, perfect systems, but can now include defects, dopants, impurities or domain boundaries in the calculations, or even calculate amorphous systems.

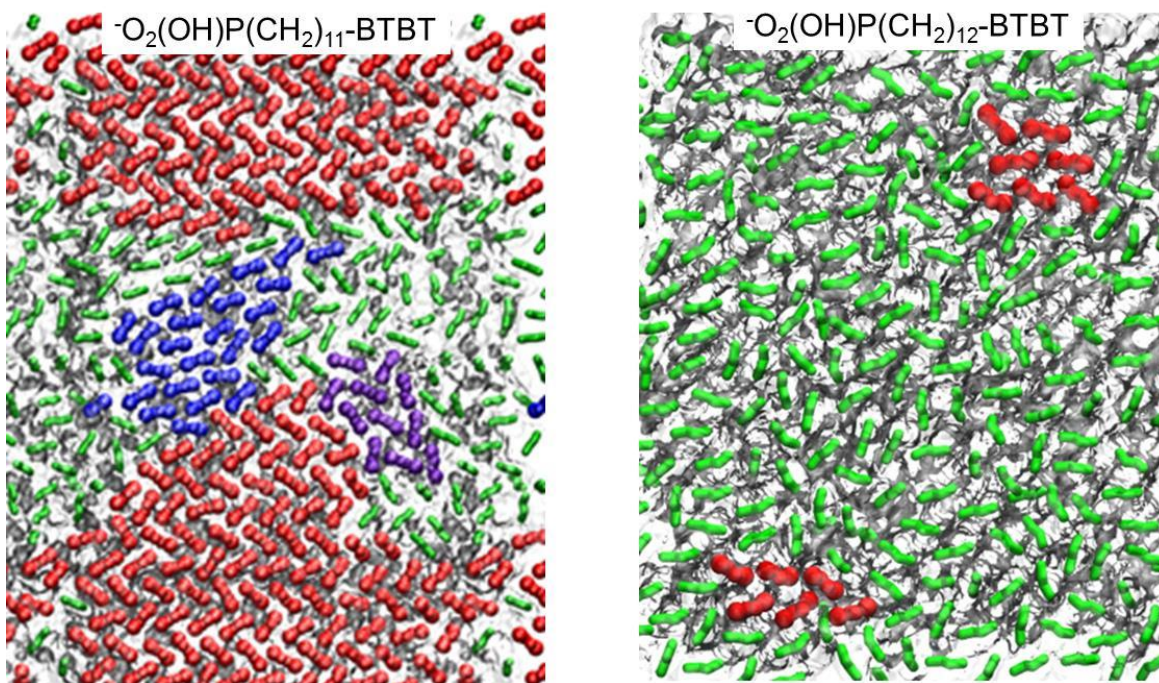
The results of such calculations can be used to simulate charge-transport through disordered monolayers. Clark's team has studied self-assembled monolayer field-effect transistors (SAMFETs) handling conformational freedom using classical atomistic molecular-dynamics (MD) simulations, electronic properties using very large scale semiempirical MO theory, and conductance by propagating single electrons or using diffusion quantum Monte-Carlo (DQMC) charge-transport simulations.⁴⁻⁹

The molecules that comprise the SAM contain insulating and semiconducting moieties, so that they serve as both gate dielectric and the active transistor channel in a device:



Tim's team has used simulations to describe and optimize complex systems of self-assembled monolayers on surfaces, not only to explain their morphology, but also to predict molecular compositions and arrangements favorable for improved charge transport.⁷ In more recent work,¹⁰ they have constructed transistors based on SAMs of two molecules that consist of the organic p-type

semiconductor benzothieno[3,2-b][1]benzothiophene (BTBT), linked to a C₁₁ or C₁₂ alkylphosphonic acid. Both molecules form ordered SAMs, but the experiments show that the size of the crystalline domains and the charge-transport properties vary considerably in the two systems. Because of the angle of the head groups one can form crystalline domains and the other cannot. This can be reproduced with simple force field calculations.



The procedure for charge transfer simulations is as follows:

- Calculate the neutral system and use local properties as external potentials:
 - Local electron affinity^{11,12} for electrons, local ionization energy¹³ for holes
 - Cluster model or periodic-boundary conditions
- Monte-Carlo search for conductance paths
- DQMC simulations¹⁴ for many electrons
- Propagate single charge carriers on these potentials to determine time scales.

Tim showed an MD simulation of the charge transport paths. For the transport calculations, the team employed a fully quantum mechanical description, namely Landauer transport theory.⁹ In accord with experiment, they found an improved charge transport across BTBT-C₁₁-PA SAMs compared to BTBT-C₁₂-PA SAMs.

DQMC reproduces voltage/current curves (assuming that the number of Monte Carlo steps correlates with time) and reproduces experimentally observed hysteresis. It also revealed dimeric fullerene electron traps.¹⁵ Density functional theory calculations indicate that van der Waals fullerene oligomers can form interstitial electron traps in which the electrons are even more strongly bound than in isolated fullerene radical anions. Spectroelectrochemical measurements on a bis-fullerene-substituted peptide

provide experimental support. The proposed deep electron traps are relevant for all organic electronics applications in which non-covalently linked fullerenes in van der Waals contact with one another serve as n-type semiconductors.

Finally Tim showed the results of simulations of hole-transport through a self-assembled monolayer substituted with a p-type organic semiconductor and with crystalline domains (see the work above on BTBT linked to a C₁₁ or C₁₂ alkylphosphonic acid). He illustrated hole transport through the monolayers. Hysteresis is not observed in this case. Tim also illustrated well-defined paths through the crystalline domains of the O₂(OH)P(CH₂)₁₁-BTBT material. The researchers have shown that structural order is particularly important for the electronic properties of semiconducting self-assembled monolayers, and they predict that semiconducting SAMs with a higher degree of crystallinity and larger crystalline regions will exhibit superior performance.

Applications of machine learning to materials and chemical property prediction



Alex Tropsha, of the University of North Carolina Chapel Hill, UNC Eshelman School of Pharmacy, is benefiting from the explosive growth of materials data. There are 160,000 entries in the Inorganic Crystal Structure Database (ICSD). There are numerous commercial and open experimental databases (NIST, MatWeb, MatBase etc.), and huge databases such as AFLOWLIB, Materials Project, and Harvard Clean Energy. The chemical space of possible materials is *huge* : about 10¹⁰⁰ candidates.¹⁶ The US government's Materials Genome Initiative recognizes the need for new high performance materials. The growth of materials databases and emerging informatics approaches offers the opportunity to transform materials discovery into data- and knowledge-driven rational

design.

[AFLOW](#) is a globally available database of 1,688,245 material compounds, with over 167,136,255 calculated properties. The optimized geometries, symmetries, band structures, and densities of states available in the AFLOWLIB consortium databases have been converted into two distinct types of fingerprints: Band structure fingerprints (B- fingerprints), and Density of States fingerprints (D- fingerprints).¹⁷ The framework is employed to query large databases of materials using similarity concepts, to map the connectivity of materials space (as a materials cartogram) for rapidly identifying regions with unique organizations and properties, and to develop predictive quantitative materials structure–property relationship (QMSPR) models for guiding materials design.

To represent the library of materials as a network (a material cartogram), the researchers considered each material, encoded by its fingerprint, as a node. Edges exist between nodes with similarities above certain thresholds (in this case, Tanimoto similarity and a threshold of 0.7). A materials map from B-fingerprints was made from 15,000 materials from ICSD, using DFT PBE calculations from [AFLOWLIB](#). Four big clusters were observed: insulators, ceramics, and complex oxides; bimetals and polymetals; metallic and nonmetallic combinations; and small band gap semiconductors.

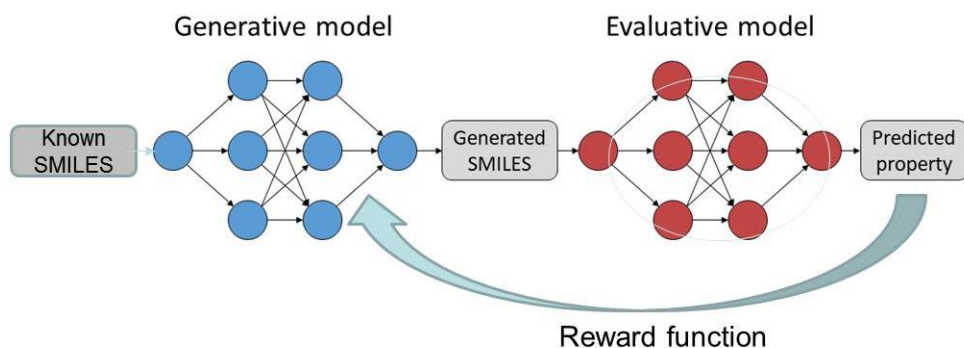
Novel descriptors (property-labeled materials fragments) not requiring prior DFT calculations have also been developed by Voronoi tessellation and neighbors search of crystal structures, followed by infinite periodic graph construction and property labeling, and generation of circular fingerprints.¹⁸ Starting from only a crystal structure, regression models can be built to predict band gap energy, and thus electronic properties, or to predict thermo-mechanical properties such as bulk modulus, shear modulus, thermal expansion, heat capacity, and thermal conductivity. All the models are trained based on DFT-computed properties. Heuristic design rules can be extracted.

Material informatics has also been applied to the design of a novel photocathode material for dye-sensitized solar cells (DSSCs).¹⁹ By conducting a virtual screening of 50,000 known inorganic compounds, the researchers have identified lead titanate (PbTiO_3), as the most promising photocathode material. Notably, lead titanate is significantly different from the traditional base elements or crystal structures used for photocathodes. In experimental validation, the fabricated lead titanate DSSC devices exhibited the best performance in aqueous solution, showing remarkably high fill factors compared to typical photocathode systems. Currently, device performance is low, but it might be improved by designing a new dye.

Next Alex discussed applications of machine learning to designing chemicals with the desired physical and biological properties where compound structure is described only by its SMILES notation, and no other conventional chemical descriptors are used. The new approach developed in his lab is based on concepts from text mining that rely on neural networks to solve the problem of semantic similarity of texts.

The British linguist J. R. Firth is noted for drawing attention to the context-dependent nature of meaning. In particular, he is known for the 1957 quotation: “You shall know a word by the company it keeps”. To define the semantic similarity between two entities, Alex and his colleagues have made use of approaches embedded in Word2Vec, a neural-network-based approach to describe linguistic context of words developed at Google.²⁰ With Word2Vec, a network is trained using each word of a corpus of text and some configurable number of surrounding words. The model can be trained to either predict the surrounding context based on the current word, or to predict the current word from the context. Elena Tutubalina and Alex (manuscript in preparation) have performed drug clustering in semantic similarity space, using webmd.com, patient.info, drugs.com, amazon.com askapatient.com, and dailystrength.org as sources of user comments, and showed that drugs with similar pharmaceutical action do cluster together in the semantic similarity space.

Alex’s team has also experimented with *de novo* design of molecules with the desired properties using SMILES in Deep Reinforcement Learning:



Structural bias, physical properties, and biological activity have been used in proof of concept case studies of user-biased molecular design. In summary, Alex cited Confucius who said, “Without knowing the force of words, it is impossible to know more”. Alex quipped “And remember: anything you say can, and will be used ... for text mining!”.

A nanoinformatics platform for environmental impact assessment of manufactured nanomaterials

Yoram Cohen of the University of California Center for Environmental Implications of Nanotechnology



gave a talk co-authored by [colleagues](#) at the University of California.

[Nanoinfo.org](#) is a nanoinformatics platform that supports the environmental impact assessment of engineered nanomaterials (ENMs) with a central database of ENM safety data and a toolkit for various exploration and analysis methods.²¹ These methods include the estimation of environmental exposure levels of ENMs (MendNano), evaluation of environmental releases of ENMs (LearNano), analysis of high throughput toxicity data of ENMs (ToxNano), and predictive toxicity models, and analysis of the environmental impact of ENMs *via* Bayesian inference (NanoEIA).

[NanoDatabank](#) is a data repository of ENM properties, and experimental and simulation datasets of ENM toxicity and environmental fate and transport (F&T). It contains databases that include physicochemical properties; toxicological properties; experimental datasets of ENM toxicity and F&T; and results of model simulations and estimation of ENM toxicity and F&T behavior, and physicochemical properties. It includes data for over 300 nanomaterials, and toxicity data for various cell lines, zebrafish and bacterial strains, from 325 publications. [ToxNano](#) is a high-content data analysis tool (HDAT)^{22,23} offering QSARs using random forest and Bayesian network toxicity models; analysis of knowledge evidence, and data visualization. [MendNano](#) (multimedia environmental distribution of nanomaterials) is a Web-based modeling platform.^{24,25} Nanoinf.org has 400 users from more than 50 countries.

As an example of work on the toxicity of nanomaterials, Yoram presented unpublished results on evaluating the body of evidence on quantum dots (QDs) *via* meta-analysis. QDs are very small semiconductor particles, only several nanometers in size, so small that their optical and electronic properties differ from those of larger particles. Many types of quantum dot will emit light of specific

frequencies if electricity or light is applied to them, and these frequencies can be precisely tuned by changing the dots' size, shape, and material.

QD data were collected from 448 publications, reporting 2,703 samples, with 7 core types, 12 shell types, 13 surface modifications, 14 surface ligands, and 20 assay types. In the predictive toxicity model R^2 was about 0.81 for cell viability, and about 0.83 for IC_{50} . Yoram and his colleagues studied cause-effect relationships between cellular bioactivity and QD attributes. Median IC_{50} was ≤ 10 mg/L, for the surface ligands of type amphiphilic polymer, lipid, other hydrophobic, aminothiols, and other amphiphilic. It was uniformly distributed for silica. There was no correlation between surface charge and IC_{50} . The sensitivity distribution of IC_{50} for cell anatomical type suggests that more differentiated cells are more adversely affected by exposure to QDs. Toxicity is not governed by QD size alone: there is a wide range of IC_{50} for a given size, and toxicity can be high or low irrespective of the size. Core type affects toxicity, but the wide range of IC_{50} for a given core type suggests that there are other important attributes.

Bayesian network models can be useful for handling uncertainties, mixed attributes, and hidden conditional relationships since they provide rigorous and simple mathematical means of handling data uncertainty; they integrate graphical representation of the problem with probabilistic evaluation of variable relationships; they can incorporate prior knowledge based on data as well as expert opinion in a convenient representation of probability distributions; and they calculate the likelihood of specific scenarios based on prior knowledge.

Bayesian network model sensitivity analysis showed that QD toxicity is correlated with the most relevant (or significant) attributes tabulated below. The QD attributes identified in this study were consistent with previous analysis *via* random forest.²⁶

Bayesian network for IC_{50}	Random forest for IC_{50}
Surface ligand	QD diameter
Shell	Surface ligand
QD diameter	Shell
Assay type	Assay type
Exposure time	Exposure time
Surface modification	Surface modification
Surface charge	Surface charge

Bayesian network for cell viability	Random forest for cell viability
Surface ligand	QD diameter
QD diameter	QD concentration
QD concentration	Surface ligand
Exposure time	Exposure time
Shell	Surface modification
Assay type	Assay type
Surface modification	Surface charge
Surface charge	

Bayesian networks for new explorations of association rules among various biological responses as a result of exposure to manufactured nanomaterials have also been demonstrated in zebrafish toxicity studies. Yoram and his co-workers used a nanomaterial biological interaction knowledge base of zebrafish phenotype data with 1,147 samples, and 11 biological responses (including mortality). The data included exposure to seven material types (carbon, cellulose, dendrimer, metal, (metal) oxide, polymeric, and semiconductor) of 0.8–250 nm average primary size; concentration; number of embryos per experiment; and responses recorded for each exposure scenario.

The Bayesian network model for zebrafish mortality (percentage of dead embryos) had an R^2 of about 0.79. Sensitivity analysis of the key material properties and exposure conditions that correlate with zebrafish mortality was carried out, and cause-effect relationships between zebrafish phenotypes and material properties and exposure conditions were investigated. Attribute significance was determined by exhaustive search of 13 attributes using bootstrapping. Mortality at 120 hours post-fertilization correlated with concentration used, core atomic composition, outermost surface, average particle size, surface charge, shell composition and purity. The significant attributes at 24 hours post-fertilization were the same but the ranking of the top four differed slightly.

The responsible development of beneficial manufactured nanomaterials requires a thorough understanding of their potential adverse environmental and human health impacts. This requires predicting the biological response of various receptors when exposed to these materials, along with an understanding of their fate and transport, and their range of likely exposure concentrations. Yoram's work helps to rank various nanomaterials with respect to their potential environmental impact.

Accurate and interpretable nano-QSAR models from genetic programming-based decision tree construction approaches



Ceyda Oksel of Imperial College London reported on the PhD work²⁷ she had done at the University of Leeds in collaboration with Xue Wang and David Winkler. Given the ever-increasing use of ENMs, it is essential to assess properly all potential risks that may occur as a result of exposure to ENMs. The distinctive characteristics of ENMs that have made them superior to bulk materials for particular applications might also have a substantial impact on the level of risk they pose. Despite the clear benefits that nanotechnology can bring, there are serious concerns about the potential health risks associated with the production and use of ENMs, intensified by the limited understanding of what makes ENMs toxic and how to make them safe.

The involvement of computational specialists in nano-safety research has become more prominent since Registration, Evaluation, Authorization and restriction of Chemicals (the European Union's [REACH](#) regulation) promoted the use of *in silico* techniques such as QSAR for toxicity assessment. Data-driven models that decode the relationships between the biological activities of ENMs and their physicochemical characteristics provide an attractive means of maximizing the value of scarce, and expensive, experimental nanotoxicity data.

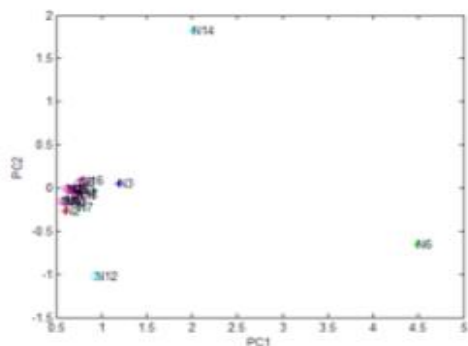
Nano-QSAR models can be used to predict the properties of new materials and to design safer materials. Leeds-based genetic programming-based decision tree (GPTree) approach²⁷ applies decision tree learning algorithms to identify the best combination of physicochemical properties to predict biological activity of ENMs. The trees are automatically constructed from the data. Decision trees have several advantages. They are able to deal with small, large and noisy datasets; they can detect nonlinear relationships (as well as linear ones); they allow input variables to be selected automatically; they are transparent; and they represent knowledge clearly (i.e., the models are interpretable).

GPTree begins with a random population of solutions and repeatedly attempts to find better solutions by applying genetic operators such as mutation and crossover. The first step is to construct a user-specified number of trees (usually a large number) starting from a random compound and a randomly chosen descriptor. Once the initial population is generated, tournament selection is performed to identify the best tree to be used as a parent tree for genetic operators such as crossover. The best tree from the subset of trees is chosen by its fitness (e.g., accuracy). Genetic operators such as crossover and mutation are used to form the next generation of trees that are added or replace the current generation. These steps are repeated until the user-specified number of generations has been created. The decision tree model with the highest accuracy of classification for the training set is selected as the optimal decision tree model.

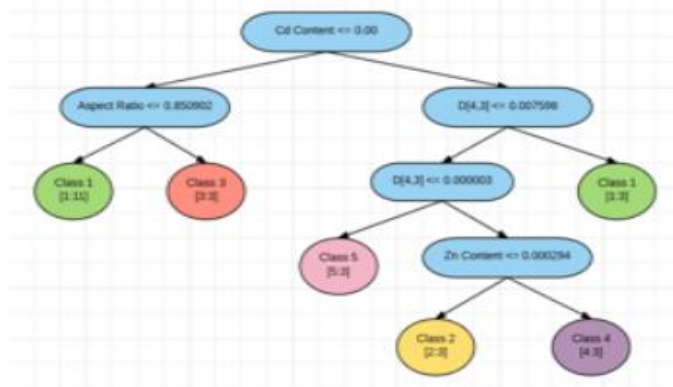
Ceyda demonstrated the application of genetic-programming-based decision tree construction algorithms to QSAR modeling of ENM toxicity by five case studies. The accuracy of the model predictions was satisfactorily high and clearly highly statistically significant relative to the classification rate due to chance.

In the first case study, a large set of in-house *in vitro* data (obtained in collaboration with Edinburgh University) was used. The dataset included a panel of 18 ENMs with varying structures (e.g., carbon-based materials and metal oxides), a set of *in vitro* cytotoxicity assays (e.g., LDH release, apoptosis, necrosis, viability, MTT and hemolytic effects), and several experimentally measured physicochemical properties (e.g., particle size and size distribution, surface area, morphology, metal content, reactivity and free radical generation). After a set of data preparation and scaling steps, a heat map of toxicity data combined with hierarchical clustering was constructed. As a second step, C-Visual Explorer (CVE) was used as a tool to create a parallel coordinate plot of the multivariate toxicity data. Similar to the heat map visualization results, the parallel coordinate plot showed that the aminated polystyrene latex beads and zinc oxide had the highest toxicity values in nearly all assays, followed by nanotubes that had medium to high toxicity values in viability and MTT assays.

Then, a dimensionality reduction technique, principal component analysis, was performed on all the toxicity data and the ENMs were divided into five categories according to their toxicity values. GPTree was used to identify potential descriptors contributing to the toxicity of four particular ENMs that were clearly separated from the main cluster formed by low-toxicity ENMs. It was concluded that high aspect ratio contributed to the toxicity of nanotubes, while the most likely factor driving the toxicity of zinc oxide was its high zinc content.

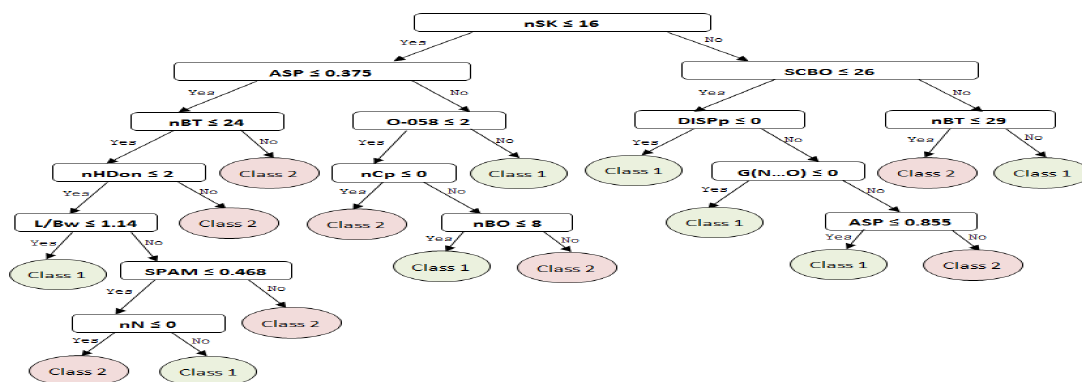


Class 1: 14 non-toxic NPs
 Class 2: NP12 (Nickel Oxide)
 Class 3: NP3 (Carbon Nanotubes)
 Class 4: NP14 (Zinc Oxide)
 Class 5: NP6 (Aminated Polystyrene Latex Beads)



GPTree Conclusions
 •Important descriptors: aspect ratio, size measurements, Cd content and Zn content
 •Class 3 (Nanotube): aspect ratio
 •Class 4 (Zinc oxide): Zn content

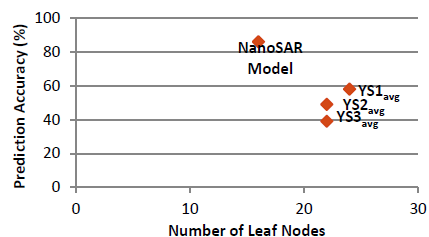
In the second case study, the cellular uptake of nanoparticles, 13 descriptors representing the hydrogen-bonding characteristics, functional group counts, molecular shape, composition and polarizability were found to be significant among a larger set of 147 chemically interpretable descriptors. The findings of GPTree analysis regarding the large contribution of lipophilicity, hydrogen bonding and molecular shape descriptors in the cellular uptake behavior of nanoparticles is consistent with earlier studies.



Internal and External Validation

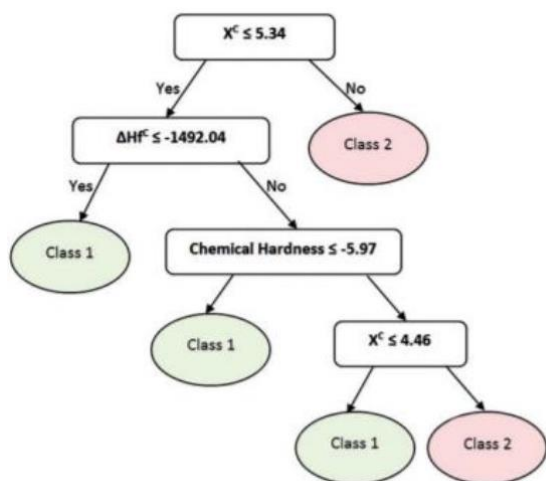
Internal Validation			External Validation		
Training Set	Predicted Class		Test Set	Predicted Class	
Actual Class	Nontoxic	Toxic	Actual Class	Nontoxic	Toxic
Nontoxic	39	1	Nontoxic	7	3
Toxic	0	47	Toxic	0	11
Sensitivity	98%		Sensitivity	79%	
Specificity	100%		Specificity	100%	
Accuracy	99%		Accuracy	86%	

Y-randomization



For a cytotoxicity to human keratinocytes dataset (the third case study),²⁸ the descriptors selected by GPTree were the enthalpy of formation of metal oxide nanocluster representing a fragment of the surface (ΔH_f^C), the Mulliken's electronegativity of the cluster, X^C , and the chemical hardness, η . The former two descriptors are consistent with the properties reported to be important for cytotoxicity of

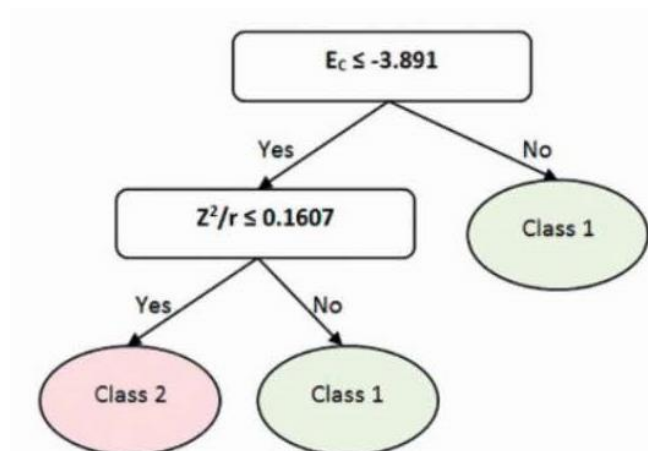
metal oxide nanoparticles. In addition, the chemical hardness corresponding to the reactivity was found to be an influential parameter on the cytotoxicity of nanoparticles.



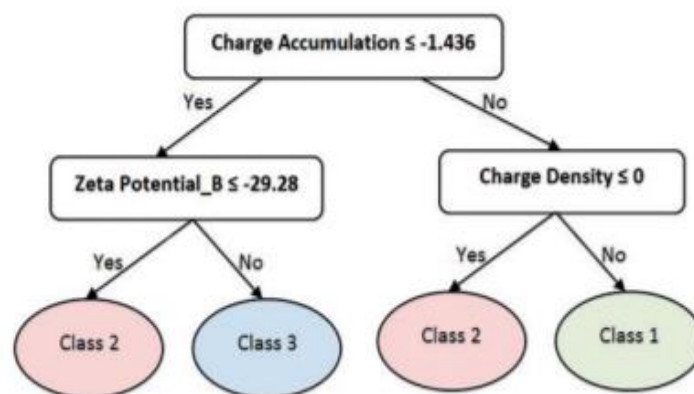
Training Set		Test Set			
	Predicted Class			Predicted Class	
Actual Class	Nontoxic	Toxic	Actual Class	Nontoxic	Toxic
Nontoxic	5	0	Nontoxic	4	0
Toxic	0	5	Toxic	0	4
Sensitivity	100%		Sensitivity	100%	
Specificity	100%		Specificity	100%	
Accuracy	100%		Accuracy	100%	

The descriptors selected by GPTree were used to develop a regression model which was statistically significant and had good predictivity ($R^2 = 0.92$, $Q^2 = 0.72$). A variable importance plot showed that X^c was twice as important as ΔH_f^c , which was a little more important than η .

The data used in the fourth case study included a set of 27 descriptors, 23 ENMs, and a set of multi- and single-parameter toxicity screening assays. The descriptors selected by the GPTree model included nanoparticle conduction band energy, E_c , and ionic index of metal cation, Z^2/r . This finding is very consistent with past studies that identified these two descriptors as being important for the toxicity of metal oxide nanoparticles.



In the last case study, exocytosis of gold nanoparticles in macrophages, the optimal descriptors for predicting the exocytosis were the charge accumulation, zeta potential and charge density. These findings are in line with previous studies revealing an association between surface characteristics of gold nanoparticles, especially high positive surface charge, and their exocytosis patterns in macrophages.



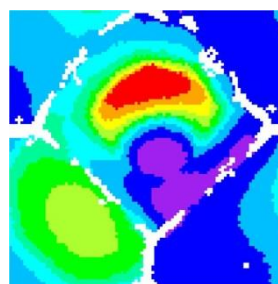
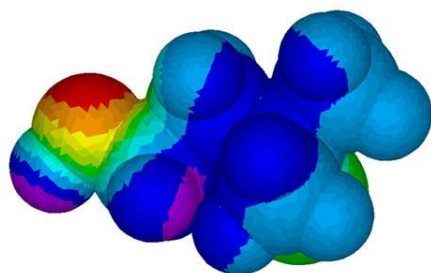
Ceyda concludes that the genetic-programming-based decision tree construction algorithm shows considerable promise in its ability to identify the relationship between molecular descriptors and biological effects of ENMs. Selected decision tree models yielded (external) prediction accuracies of 86-100%. Another statistical test (Y-randomization) was also performed to demonstrate the robustness of the selected models. This work is a first step in the implementation of a genetic programming based decision tree construction algorithm to nano-QSAR studies.

Self-organizing neural networks in chemistry



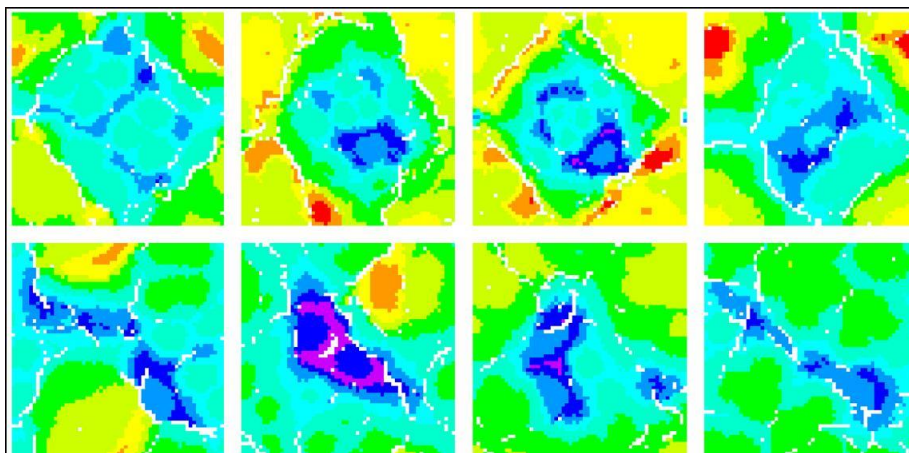
Johnny Gasteiger of the University of Erlangen-Nürnberg is skeptical about deep neural networks: they are good for getting funding, but they are yet to be proven. Johnny illustrated some of the useful applications of shallow neural networks. Much like the human brain generates two-dimensional sensory maps of the environment, a Kohonen network (a self-organizing map) can generate two-dimensional maps of high-dimensional chemical data. Crucial for the success of the study of chemical problems by a self-organizing neural network is the representation of the chemical data.

The shape and surface of molecules are very important: the entire electrostatic potential can be seen in a colored 3D model. Johnny has projected the 3D Cartesian coordinates of, for example, 2-chloro-4-hydroxy-2-methylbutane onto a Kohonen net to get a 2D map:



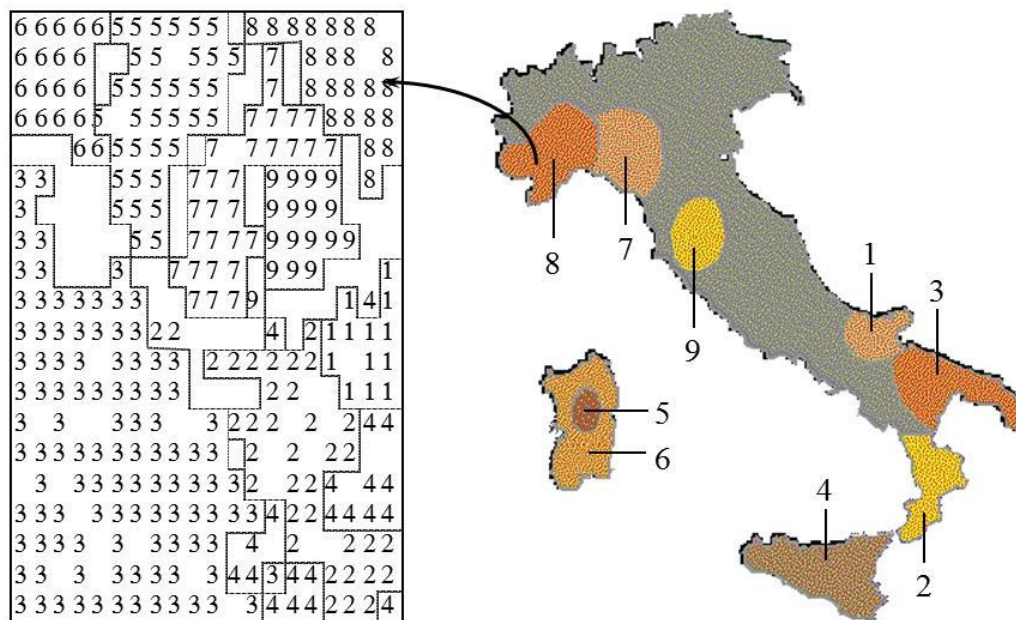
The neurotransmitter acetylcholine binds to two types of receptors, the muscarinic and the nicotinic receptor. Kohonen maps of the van der Waals surface of muscarinic agonists (muscarine, atropine,

scopolamine, pilocarpine) and nicotinic agonists (nicotine, (+)-anatoxin a, mecamlamine, pempidine) have also been produced by projecting points of the 3D surface on a 2D space.²⁹ Such maps allowed the total molecular electrostatic potential (MEP) of a compound to be represented in a single picture, instead of requiring a series of pictures as formerly. Johnny showed the maps of the MEPs of the eight compounds with muscarinic agonists in the top row and nicotinic agonists below.



The results showed that the MEP is important for the binding of these compounds to their receptors. The Kohonen maps reflect significant characteristics of the MEPs and can therefore be used in the search for biologically active compounds.

In analytical chemistry, neural networks have been used in the classification of Italian olive oils.³⁰ The classification was performed on a set of 572 Italian olive oils, from nine different regions, on the basis of an analysis of eight fatty acids. Kohonen learning was superior to a network using the back-propagation of errors. There were 250 oils in the training set and 322 in the test set; 312 of the 322 were correctly predicted. The nine Italian regions were nicely differentiated in the Kohonen map. What is, however, even more interesting is that the Kohonen map is reflecting the map of Italy. This emphasizes the power of unsupervised learning, discovering information that is hidden in the data. In this case, clearly, the different climates and the different soils are responsible for the separation of the regions of Italy in the self-organizing map:



Kohonen networks use unsupervised learning. Johnny next discussed examples of supervised learning. In one experiment the electronic properties located on the atoms of a molecule such as partial atomic charge, and electronegativity and polarizability values were encoded by an autocorrelation vector accounting for the constitution of a molecule.³¹ Using the 49-dimensional vector of seven properties and seven distances, it is possible to distinguish between 112 dopamine agonists and 60 benzodiazepine receptor agonists even after projection into a Kohonen map. The two types of compounds can still be distinguished if they are buried in a dataset of 8,323 compounds of a chemical supplier catalog comprising a wide structural variety. The method can be used for searching for structural similarity, and, in particular, for finding new lead structures with biological activity.

Gasteiger's team has also worked on simulation of infrared spectra.³² They developed an empirical approach to the modeling of the relationships between the 3D structure of a molecule and its IR spectrum based on a novel 3D structure representation, and a counterpropagation (CPG) neural network. The 3D coordinates of the atoms of a molecule are transformed into a structure code that has a fixed number of descriptors irrespective of the size of a molecule. The structure coding technique is referred to as radial distribution function (RDF) code.³³ 3D structures were transformed into radial codes (128 values) and put into a CPG network. IR spectra (128 absorbance values) were also input, and the network was trained. When IR spectra are simulated the fingerprint region is predicted well because of the representation of the 3D structure. A CPG network can be operated in reverse mode,³³ enabling the prediction of a structure code. The input of a query infrared spectrum into a trained CPG network provides a structure code vector, which represents the radial distribution function with 128 discrete values. This RDF code is then decoded to provide the Cartesian coordinates of a 3D structure.

Johnny concluded by mentioning his recent collaboration with David Winkler on dye solubility in carbon dioxide.³⁴ David has also worked on melting points of ionic liquids, fibrinogen adsorption to polymeric

surfaces, and normalized metabolic activity of polymeric biomaterials. Johnny encouraged David to continue to do good science.

Understudied proteins. Time to shift the paradigm



Tudor Oprea of the University of New Mexico believes that identifying novel targets as a precompetitive endeavor can lead to new therapeutic opportunities if academia and industry work together. Most protein classification schemes are based on structural and functional criteria. For therapeutic development, it is useful to understand how many data and what types of data are available for a given protein, thereby highlighting well-studied and understudied targets. [Tudor and his co-workers](#) classify proteins annotated as drug targets as “Tclin”; proteins for which *potent* small molecules are known as “Tchem”; proteins for which biology is better understood as “Tbio”; and proteins that lack antibodies, publications or National Center for Biotechnology Information (NCBI) Gene References Into Function (GeneRIFs) as “Tdark”.

Tclin proteins are associated with drug mechanism of action (MoA). Tchem proteins have bioactivities in [ChEMBL](#) and [DrugCentral](#), plus human curation for some targets. A Tbio protein lacks small molecule annotation, and is above the cutoff criteria for Tdark, or is annotated with a Gene Ontology (GO) molecular function or biological process leaf term(s) with an experimental evidence code, or has confirmed Online Mendelian Inheritance in Man ([OMIM](#)) phenotype(s). Tudor and his colleagues used name entity recognition software³⁵ from L. J. Jensen’s lab to evaluate nearly 27 million abstracts to derive a publication score per protein. Tdark proteins (“understudied proteins”) have little information available, and meet two of the following three criteria: a PubMed text mining score of less than five, three or fewer GeneRIFs, and 50 or fewer antibodies available according to [antibodypedia](#). As external validation, Tdark proteins have statistically significantly lower values compared to the other three target development levels (TDLs) in terms of fewer GO terms, fewer patents, fewer National Institutes of Health (NIH) R01 grants, and fewer searches of the [STRING-db](#) database.

Tudor’s first “take home message” was that there is a knowledge deficit: over 37% of the proteins remain understudied (the Tdark ones) and only about 10% of the proteome (Tclin and Tchem) can be targeted by potent small molecules. Are Tdark proteins underfunded because there is no scientific interest in this category, or is the lack of knowledge perpetuated by lack of funding? It is possible that the absence of high-quality, well-characterized molecular tools (i.e., antibodies or chemical probes) may be a root cause for this situation, but lack of tools leads to lack of interest, and lack of interest diminishes the probability of such tools being developed.

The patent literature is also of interest. Almost half of patent bioactivity data are never published elsewhere, and compounds may appear in patents two to four years before they appear in the literature. The [SureChEMBL](#) team has annotated the SureChEMBL patent corpus with gene and disease terms. Looking at patents between 2001 and 2013, they processed a set of 99 approved patents of interest to the Illuminating the Druggable Genome ([IDG](#)) consortium. These bioactivity data from 99

patents were manually extracted: 20,941 activity measurements for 11,358 compounds, and 1,134 assays. These data are already uploaded into ChEMBL 23. Data for seven IDG Phase 2 targets were uncovered by this patent data extraction exercise, data which progress TDLs of two targets (GPR6 and HCAR1) from Tbio to Tchem.

Anne Hersey of ChEMBL has estimated that more than 50% of the data from patents do not end up in peer-reviewed papers. [IDG](#), [Open Targets](#), [BindingDB](#), and others could collectively, in a precompetitive manner, mine data from patents (if necessary, for only terminated projects, or out-of-patent drugs) and upload these data into [ChEMBL](#) and [Pharos](#). Pharos³⁶ is the user interface to the [Knowledge Management Center](#) (KMC) for the IDG program funded by the NIH.

Approximately one-third of all mammalian genes are essential for life. Phenotypes resulting from knockouts of these genes in mice have provided insight into gene function and congenital disorders. The International Mouse Phenotyping Consortium ([IMPC](#)) has published research on the high-throughput discovery of novel developmental phenotypes.³⁷ They identified 2,788 genes with 8,241 significant phenotype calls in 25 major categories. The promise of the IMPC annotations is illustrated by examining the definite and clear links between human neurological and behavioral disorders (191 human genes) and the corresponding gene knockout mouse neurological and behavioral phenotypes. The majority of these links are for schizophrenia, Alzheimer's disease, epilepsy, and amyotrophic lateral sclerosis. Several rare diseases are also associated with these genes.

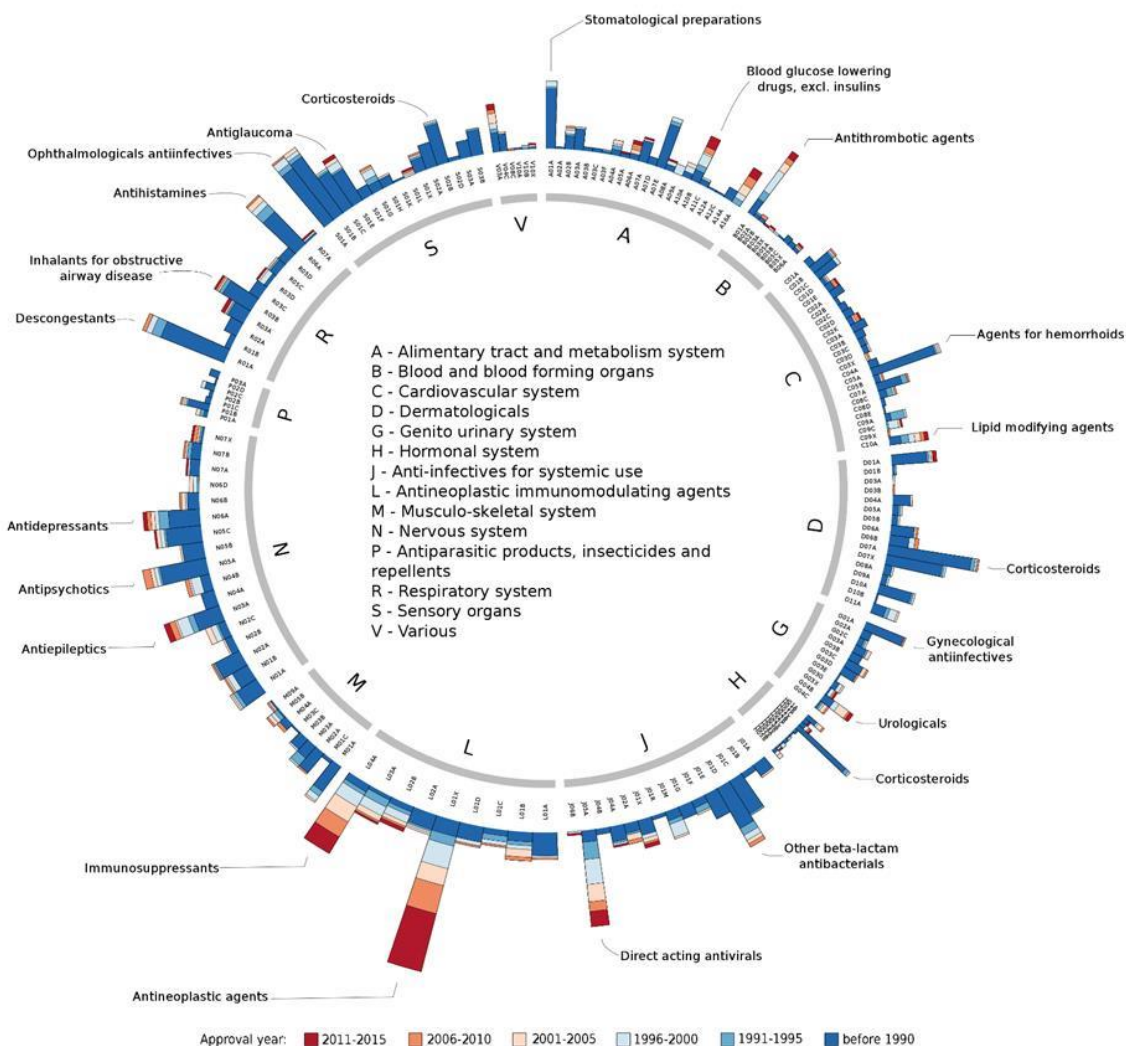
Of 119 Tdark genes prioritized by [KMC](#) to IMPC, 45 mouse lines were produced, with 41 phenotypes observed. Knockouts of the Tdark kinase [Alpk3](#) have increased embryonic and perinatal lethality, with the surviving adults displaying severe heart defects. Of 482 Tbio genes submitted by KMC, 184 mouse lines were produced, with 145 phenotypes observed. Knockouts of the Tbio GPCR [Adgrd1](#) display reproductive defects. (These are Tdark and Tbio statistics as of April 2017.) Tudor commented: "If you don't know very much to begin with, don't expect to learn a lot quickly."

Data from Cristian Bologa suggest that on average it takes 15-20 years for Tdark to bear fruit. The leptin receptor was Tdark in 1995, but led to an approved drug in 2014. The smoothed receptor was Tdark in 1997, and a drug was launched in 2012. Tudor gave several other examples. There is room for improvement in research funding. Text mining of all NIH grants for the period 2000-2015 suggests that 8,858 proteins received zero NIH funding. Of these, 6,051 are Tdark, and 2,616 are Tbio. This is to be expected, but 119 are Tchem and 72 are Tclin. Possible explanations could be old drug targets or research funded elsewhere. (Data from funding sources other than NIH are not available.) Pharma and academia could pay more attention to these 8,858 underfunded proteins.

Tudor's second take home message was that just because something is ignored it does not mean it lacks importance. Understudied proteins need funding and patience. Based on current evidence, IMPC has the most concerted Tdark exploration approach.

DrugCentral (<http://drugcentral.org>) is an open access online drug compendium³⁸ integrating structure, bioactivity, regulatory information, pharmacologic actions, and indications for active pharmaceutical ingredients approved by regulatory agencies. It integrates content for active ingredients with

pharmaceutical formulations, indexing drugs and drug label annotations, and complementing similar resources available online. Tudor's team used it initially to find how many drugs there are, but they also wanted to know how many drug targets there are. They have studied innovation patterns per therapeutic area:³⁹



Drugs distributed by Anatomical Therapeutic Chemical (ATC) codes (levels 1-2). Concentric rings indicate ATC levels. Histograms represent the number of drugs distributed per year of first approval.

They have also examined the [commercial impact of target classes](#) by evaluating data from IMS Health on drug sales from 75 countries, aggregated over a five-year period (2011–2015). After excluding categories such as homeopathic medicines, they identified 51,095 unique products, and mapped them to 1,069 active pharmaceutical ingredients from DrugCentral, corrected by the number of active pharmaceutical ingredients (APIs) per product, then by the number of Tclin targets per API. The most lucrative target class from a therapeutic perspective was G-protein coupled receptors (GPCR, 27.42% market share).

Tudor also tabulated the top 20 targets by revenue. His third take home message was that there are many unexplored opportunities. By his conservative estimate (about 15,000 disease concepts, and about 2500 unique drug indications), we address about 15% of human diseases with therapeutic agents.

It has been said that the absence of a quantitative language is the flaw of biological research⁴⁰ or “the more facts we learn the less we understand”. Again, when little is known, we should not expect knowledge to accumulate quickly. Separation by organ and cell is a conceptual fallacy. Medicine maintains this separation for necessity: by organ (e.g., cardiology or ophthalmology), and by disease category (e.g., oncology or infection). NIH Institutes are organized in a similar way. Many pharmaceutical companies are organized by therapeutic area. Yet genes, proteins and pathways do not observe such separation. The impact of this “mental divide” in science has yet to be understood.

A. B. Jensen *et al.* have studied disease correlations and temporal disease progression (trajectories)⁴¹ on a large scale over 15 years, and grouped 1,171 significant trajectories into temporal patterns centered on a small number of early diagnoses that are central to disease progression. Hence it is important to focus on early diagnoses in order to mitigate the risk of adverse patient outcomes. The authors suggest such trajectory analyses may be useful for predicting and preventing future diseases of individual patients. Using data from the Cerner HealthFacts database, Tudor’s team has found that the top diseases prior to Alzheimer’s (over 5 years or more) are essential hypertension, hyperlipidemia, Type 2 diabetes mellitus, hypercholesterolemia, and coronary atherosclerosis. For renal failure, diseases over the previous five years are essential hypertension, heart failure, angina pectoris, chronic heart disease, and diabetes mellitus.

Diseases are concepts. They lack physical manifestation outside patients, so the search for cures has to be patient-centered.⁴² Animal models should be combined with mining of patient data. We ought to use electronic health record data to prioritize targets for further drug discovery. For example, we should get genes associated with diseases that *precede* Alzheimer’s to investigate possible causality. Such priorities could be disease-specific, or phenotype-specific.

It is time to acknowledge that target prioritization for drug discovery is precompetitive knowledge. The pharmaceutical industry reward system is based on patents, which are awarded for *drugs*, not targets. Finding a good target leads to the “me-too” phenomenon. It is time to pool resources together on targets, team up with [Open Targets](#) and create a Target Selection Consortium, partnering industry with academia. “Double blind” studies could be cosponsored, to avoid the reproducibility crisis. [IDG KMC](#) is seeking new knowledge.

Sparse QSAR modeling methods for therapeutic and regenerative medicine



David Winkler's award address was co-authored by his colleague Frank Burden, now retired from CSIRO, and by co-workers at Imperial College London, King's College London, and the University of Nottingham, whose work is acknowledged in the literature references.

David's research concerns computational chemistry applied to a molecular level understanding of interactions of molecules and materials with biology. He has a strong interdisciplinary, translational research focus. His modeling, design, and optimization of bioactive materials focus on testing model predictions by subsequent experiments. He employs a range of computational tools including quantum chemistry, molecular dynamics and mechanics, molecular graphics, pharmacophore models, protein docking, and, in the case of this talk, quantitative structure-property relationship modeling. He is interested in the design of drugs and materials for therapeutic and regenerative medicine, especially control of stem cell fate, with a particular focus on the application of artificial intelligence (AI), machine learning, pattern recognition, complex systems science, evolutionary algorithms, and adaptive learning.

His work has had commercial impact, including the transfer of neural network modeling technology to BioRAD Corporation; several field trials candidates with Du Pont and Schering Plough; and clinical trials of a radioprotectant drug for cancer radiotherapy patients (with Sirtex and the Peter Mac Cancer Institute). He developed core intellectual property (a novel antibacterial target in bacterial replisome) for the Betabiotics company spinoff, and discovered a new mechanism for strontium biomaterial-induced differentiation of mesenchymal stem cells to bone. He carried out a large project with Air Liquide Santé on using *in silico* methods to understand the surprisingly rich biological properties of noble gases. He discovered new antifibrotic and antihypertensive agents for Vectus Biosystems (allowing them to float on the stock market) and a first in class drug lead for myelofibrosis, which will be further developed by a new spin off company soon.

Winkler's research thinking was greatly influenced by complex systems science, which finds deep mechanistic similarities between areas of science that appear to have nothing in common. Concepts include nonlinear dynamical behavior, networks and their attractor states, self-organized criticality, chaos, and emergent properties. Complex systems science stimulates substantial lateral thinking and novel problem solving. Methods from other areas of science can provide novel solutions to problems in drug discovery; and methods developed for drug discovery can provide novel solutions to problems in other areas of science, such as biomaterials, gene expression, non-biological materials, and regenerative medicine.

QSAR was invented by Toshio Fujita (very recently deceased) and Corwin Hansch, and rapidly evolved into a method for optimization of drugs and agrochemicals. David and Toshio published a recent paper⁴³ on the two forms of QSAR: "explain" and "predict". Graham Richards' and Peter Andrews' seminal commercialization ventures influenced David to make translation a strong focus in his research.

The research for which David received the Skolnik award involved the application of modern computational and mathematical methods to optimizing the QSAR modeling process.⁴⁴ The first operation is to generate descriptors. Model quality is critically dependent on descriptors. Descriptors with low or no relevance to the property modeled degrade the model. Bad descriptors were a problem in early QSAR work, and there is still a major research need for good descriptors for materials. Next a subset of descriptors is chosen for the model in a context-dependent way. Choosing too many subsets can give chance correlations. In generating the relationship between the descriptors and the target property, model quality is less dependent on the modeling algorithm than on the descriptors, but there can be issues in overfitting, overtraining, ambiguity in network architecture, and subjective choices. The next operation is validating the performance of the model in predicting properties of new data. Here, cross validation and bootstrapping generate optimistic measures of performance, and an independent test set not used in training is best. The final operation is making new predictions from the model and synthesizing and testing new materials.

Descriptors are the last major research problem for QSAR. Many (such as DRAGON descriptors) are arcane; efficient, interpretable descriptors are needed. Descriptors specific to complex materials are essential, but the field is embryonic. High throughput characterization data can augment computed descriptors.

There are advantages in removing irrelevant features. Least squares in multiple linear regression (MLR) has a Gaussian prior. This can be replaced with a Laplacian prior which effects the removal of uninformative weights by driving them to zero. Sparse Bayesian feature selection methods (feature selection using expectation maximization) identify a small number of relevant features very efficiently.⁴⁵

There are many methods of varying sophistication in finding structure-activity relationships,⁴⁴ including simple linear statistical regression methods such as multiple linear regression; nonlinear regression methods using polynomials or nonlinear kernels, and nonlinear machine learning; bioinspired methods such as neural nets; support vector machines; and random forests. These have new applications in materials, nanotechnology, and regenerative medicine.

The universal approximation theorem states that neural networks can model any complex relationship given sufficient training data. Neural networks are very well suited to modeling of complex data, but they have problems such as overfitting and overtraining. They raise an ill-posed problem in statistics (instability), and optimum network architecture is ambiguous. The contribution of David and his co-workers is to develop very robust, self-optimizing sparse feature selection and neural network methods that overcome all these problems.⁴⁶ These methods have been shown to have performance similar to that of deep neural networks.

Sparse Bayesian modeling and feature selection, replacing the Gaussian prior with the Laplacian prior, is a general nonlinear modeling method^{45,47-49} that automatically optimizes model complexity, prunes neural network weights to avoid overfitting, and prunes irrelevant descriptors to optimize the predictivity of a model. A sparsity-inducing Laplacian prior (LP) was introduced into Winkler's Bayesian

Regularized Artificial Neural Network algorithm (BRANN) creating BRANNLP.^{47,49} Low relevance weights are set to zero, and descriptors are also pruned from the model if all weights are zero.

From selection and mapping, David turned to validation. Cross validation, bootstrapping, and other methods give an overly optimistic estimate of predictive power because the test set is not independent of the training set. An independent test set never seen by the model is the gold standard. Many measures of predictivity have been proposed. Test set validation is actually a simple problem in statistics; standard error of prediction, test set (SEP) is preferred over r^2 as it is less dependent on dataset size and model complexity.^{46,50}

Methods from other areas of science can provide novel solutions to problems in drug discovery, and methods developed for drug discovery can provide novel solutions to problems in other areas of science. Implantable medical devices are an example. Bacterial adhesion and growth on biomaterial surfaces of joint prostheses, heart valves, shunts, vascular and urinary catheters, and intraocular lenses are serious problems in health care. There is a major unmet medical need for new coating materials for implantable and indwelling medical devices. David and his co-workers from Morgan Alexander's research team at the University of Nottingham have used machine learning methods to derive quantitative models relating the molecular structure of a polymer to the attachment of the bacteria to that polymer surface. These models can be used to screen large databases of new materials for those with low pathogen attachment.

Hook *et al.* have detected the attachment of selected bacterial species to 576 polymeric materials in a high-throughput microarray format.⁵¹ In work by David and his colleagues, data from a large polymer microarray exposed to three clinical pathogens were used to derive robust and predictive machine learning models of pathogen attachment.⁵² The BRANN models can predict pathogen attachment for the polymer library quantitatively. The models also successfully predict pathogen attachment for a second-generation library, and identify polymer surface chemistries that enhance or diminish pathogen attachment. A manuscript on work on multiple pathogen attachment models has been submitted.

Sparse feature selection methods have also identified a new mechanism for strontium biomaterial-induced differentiation of mesenchymal stem cells to bone. Strontium ranelate (Protelos) is a drug approved in the European Union for the treatment and prevention of osteoporosis. It reduces risk of vertebral and non-vertebral fractures in post-menopausal women. Although controversial, it is reported to have an anabolic *and* anti-catabolic effect on bone. Strontium ion's mechanism of action is not fully understood, but it is thought to up-regulate differentiation of osteoprogenitors or stimulate bone formation.⁵³⁻⁵⁵

David and his Imperial College co-workers,⁵⁶ Molly Stevens, Eileen Gentleman, and H elene Autefage, have evaluated the global response of human mesenchymal stem cells to strontium-substituted bioactive glasses using a combination of unsupervised biological and physical science techniques. Their objective analyses of whole gene-expression profiles, confirmed by standard molecular biology techniques, revealed that strontium-substituted bioactive glasses up-regulated the isoprenoid pathway, suggesting an influence on both sterol metabolite synthesis and protein prenylation processes.

In future, David hopes to see exploitation of new AI methods such as deep learning; improved descriptors for molecules that are effective and interpretable; exploitation of evolutionary methods of discovery aided by robotics; synergy of AI and evolutionary methods for adaptive evolution; adoption of *in silico* methods from drug discovery for materials and regeneration; development of autonomous or semiautonomous “closed loop” design methods; and more effective exploration of vast molecular or materials spaces.

Deep learning was predicted to be a breakthrough technology in 2013. Deep neural networks are not necessarily magic. According to the universal approximation theorem, a feed-forward network with a single hidden layer containing a finite number of neurons can approximate any continuous function, under mild assumptions on the activation function. This was first proved by Cybenko in 1989 for sigmoid activation functions. Hornik showed in 1991 that it is not the choice of the activation function, but the multilayer architecture itself which gives neural networks the potential of universal approximators.⁴⁶

Deep learning methods have generated impressive improvements in image and voice recognition, and are now being applied to QSAR and QSAR modeling. A recent publication⁴⁶ describes the differences in approach between deep and shallow neural networks, compares their abilities to predict the properties of test sets for 15 large drug datasets, discusses the results in terms of the universal approximation theorem for neural networks, and describes how deep neural networks may ameliorate or remove troublesome “activity cliffs” in QSAR datasets. Materials space is vast and at least in some of its many dimensions, the fitness landscape is smooth. This allows adaptation, one step (one mutation) at a time. Evolution and machine learning can be combined in adaptive learning (the [Baldwin effect](#)).

A recent review discusses the problems of large materials spaces, the types of evolutionary algorithms employed to identify or optimize materials, and how materials can be represented mathematically as genomes.⁵⁷ It describes fitness landscapes and mutation operators commonly employed in materials evolution, and provides a comprehensive summary of published research on the use of evolutionary methods to generate new catalysts, phosphors, and a range of other materials. Another recent paper describes the materials genome in action.⁵⁸

Machine learning methods have achieved wide applicability: for example, in aqueous solubility of drugs;⁵⁹ polymers for stem cell growth;⁶⁰ cubane as a benzene isostere;⁶¹ benign organic corrosion inhibitors;⁶² markers for stem cell division;⁶³ materials for stem cell factories;⁶⁴ adverse effects of nanomaterials;⁶⁵ anticancer farnesyltransferase inhibitors;⁶⁶ and prediction of materials properties.⁴⁴

In summary, AI tools developed for therapeutic medicine also work well for regenerative medicine. Neural networks are machine learning methods that are very applicable to (bio)materials design. The universal approximation theorem means that deep learning methods should not be superior to shallow neural networks for molecular design. Bayesian regularized neural networks can generate robust, predictive models of many types of materials and properties. Sparse Bayesian feature selection methods can reduce the dimensionality of problems, improve interpretability, and generate robust models with better predictivity. Evolutionary methods, combined with machine learning (adaptive evolution) can find effective materials quickly and efficiently.

Conclusion

Erin Davis, chair of the ACS Division of Chemical Information, formally presented the Herman Skolnik Award to David Winkler at a reception held in honor of David, following the symposium.



Erin Davis and David Winkler

References

- (1) Hennemann, M.; Clark, T. EMPIRE: a highly parallel semiempirical molecular orbital program: 1: self-consistent field calculations. *J. Mol. Model.* **2014**, *20* (7), 2331.
- (2) Margraf, J. T.; Hennemann, M.; Meyer, B.; Clark, T. EMPIRE: a highly parallel semiempirical molecular orbital program: 2: periodic boundary conditions. *J. Mol. Model.* **2015**, *21* (6), 144.
- (3) Wick, C. R.; Hennemann, M.; Stewart, J. J. P.; Clark, T. Self-consistent field convergence for proteins: a comparison of full and localized-molecular-orbital schemes. *J. Mol. Model.* **2014**, *20* (3), 2159.
- (4) Novak, M.; Jaeger, C. M.; Rumpel, A.; Kropp, H.; Peukert, W.; Clark, T.; Halik, M. The morphology of integrated self-assembled monolayers and their impact on devices - A computational and experimental approach. *Org. Electron.* **2010**, *11* (8), 1476-1482.
- (5) Jedaa, A.; Salinas, M.; Jaeger, C. M.; Clark, T.; Ebel, A.; Hirsch, A.; Halik, M. Mixed self-assembled monolayer of molecules with dipolar and acceptor character. Influence on hysteresis and threshold voltage in organic thin-film transistors. *Appl. Phys. Lett.* **2012**, *100* (6), 063302/1-063302/4.

- (6) Salinas, M.; Jaeger, C. M.; Amin, A. Y.; Dral, P. O.; Meyer-Friedrichsen, T.; Hirsch, A.; Clark, T.; Halik, M. The Relationship between Threshold Voltage and Dipolar Character of Self-Assembled Monolayers in Organic Thin-Film Transistors. *J. Am. Chem. Soc.* **2012**, *134* (30), 12648-12652.
- (7) Jaeger, C. M.; Schmaltz, T.; Novak, M.; Khassanov, A.; Vorobiev, A.; Hennemann, M.; Krause, A.; Dietrich, H.; Zahn, D.; Hirsch, A.; Halik, M.; Clark, T. Improving the Charge Transport in Self-Assembled Monolayer Field-Effect Transistors: From Theory to Devices. *J. Am. Chem. Soc.* **2013**, *135* (12), 4893-4900.
- (8) Bauer, T.; Schmaltz, T.; Lenz, T.; Halik, M.; Meyer, B.; Clark, T. Phosphonate- and Carboxylate-Based Self-Assembled Monolayers for Organic Devices: A Theoretical Study of Surface Binding on Aluminum Oxide with Experimental Support. *ACS Appl. Mater. Interfaces* **2013**, *5* (13), 6073-6080.
- (9) Leitherer, S.; Jaeger, C. M.; Halik, M.; Clark, T.; Thoss, M. Modeling charge transport in C60-based self-assembled monolayers for applications in field-effect transistors. *J. Chem. Phys.* **2014**, *140* (20), 204702/1-204702/10.
- (10) Schmaltz, T.; Gothe, B.; Krause, A.; Leitherer, S.; Steinrueck, H.-G.; Thoss, M.; Clark, T.; Halik, M. Effect of Structure and Disorder on the Charge Transport in Defined Self-Assembled Monolayers of Organic Semiconductors. *ACS Nano* **2017**, Ahead of Print.
- (11) Ehresmann, B.; Martin, B.; Horn, A. H. C.; Clark, T. Local molecular properties and their use in predicting reactivity. *J. Mol. Model.* **2003**, *9* (5), 342-347.
- (12) Clark, T. The local electron affinity for non-minimal basis sets. *J. Mol. Model.* **2010**, *16* (7), 1231-1238.
- (13) Sjoberg, P.; Murray, J. S.; Brinck, T.; Politzer, P. Average local ionization energies on the molecular surfaces of aromatic systems as guides to chemical reactivity. *Can. J. Chem.* **1990**, *68* (8), 1440-1443.
- (14) Bauer, T.; Jaeger, C. M.; Jordan, M. J. T.; Clark, T. A multi-agent quantum Monte Carlo model for charge transport: Application to organic field-effect transistors. *J. Chem. Phys.* **2015**, *143* (4), 044114/1-044114/9.
- (15) Shubina, T. E.; Sharapa, D. I.; Schubert, C.; Zahn, D.; Halik, M.; Keller, P. A.; Pyne, S. G.; Jennepalli, S.; Guldi, D. M.; Clark, T. Fullerene Van der Waals Oligomers as Electron Traps. *J. Am. Chem. Soc.* **2014**, *136* (31), 10890-10893.
- (16) Walsh, A. Inorganic materials: The quest for new functionality. *Nat. Chem.* **2015**, *7* (4), 274-275.
- (17) Isayev, O.; Fouches, D.; Muratov, E. N.; Oses, C.; Rasch, K.; Tropsha, A.; Curtarolo, S. Materials Cartography: Representing and Mining Materials Space Using Structural and Electronic Fingerprints. *Chem. Mater.* **2015**, *27* (3), 735-743.
- (18) Isayev, O.; Oses, C.; Toher, C.; Gossett, E.; Curtarolo, S.; Tropsha, A. Universal fragment descriptors for predicting properties of inorganic crystals. *Nat. Commun.* **2017**, *8*, 15679.
- (19) Moot, T.; Isayev, O.; Call, R. W.; McCullough, S. M.; Zemaitis, M.; Lopez, R.; Cahoon, J. F.; Tropsha, A. Material informatics driven design and experimental validation of lead titanate as an aqueous solar photocathode. *Mater. Discovery* **2016**, *6*, 9-16.
- (20) Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. 2013, arXiv.org e-Print archive. <https://arxiv.org/abs/1301.3781> (accessed September 4, 2017).
- (21) Cohen, Y.; Rallo, R.; Liu, R.; Liu, H. H. In Silico Analysis of Nanomaterials Hazard and Risk. *Acc. Chem. Res.* **2013**, *46* (3), 802-812.
- (22) Liu, R.; Jiang, W.; Walkey, C. D.; Chan, W. C. W.; Cohen, Y. Prediction of nanoparticles-cell association based on corona proteins and physicochemical properties. *Nanoscale* **2015**, *7* (21), 9664-9675.

- (23) Liu, R.; Rallo, R.; Bilal, M.; Cohen, Y. Quantitative Structure-Activity Relationships for Cellular Uptake of Surface-Modified Nanoparticles. *Comb. Chem. High Throughput Screening* **2015**, *18* (4), 365-375.
- (24) Liu, H. H.; Bilal, M.; Lazareva, A.; Keller, A.; Cohen, Y. Simulation tool for assessing the release and environmental distribution of nanomaterials. *Beilstein J. Nanotechnol.* **2015**, *6*, 938-951.
- (25) Liu, H. H.; Cohen, Y. Multimedia Environmental Distribution of Engineered Nanomaterials. *Environ. Sci. Technol.* **2014**, *48* (6), 3281-3292.
- (26) Oh, E.; Liu, R.; Nel, A.; Gemill, K. B.; Bilal, M.; Cohen, Y.; Medintz, I. L. Meta-analysis of cellular toxicity for cadmium-containing quantum dots. *Nat. Nanotechnol.* **2016**, *11* (5), 479-486.
- (27) Oksef, C.; Winkler, D. A.; Ma, C. Y.; Wilkins, T.; Wang, X. Z. Accurate and interpretable nanoSAR models from genetic programming-based decision tree construction approaches. *Nanotoxicology* **2016**, *10* (7), 1001-1012.
- (28) Gajewicz, A.; Schaeublin, N.; Rasulev, B.; Hussain, S.; Leszczynska, D.; Puzyn, T.; Leszczynski, J. Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. *Nanotoxicology* **2015**, *9* (3), 313-325.
- (29) Gasteiger, J.; Li, X. Representation of the electrostatic potentials of muscarinic and nicotinic agonists with artificial neuronal nets. *Angew. Chem., Int. Ed. Engl.* **1994**, *33* (6), 643-646.
- (30) Zupan, J.; Novic, M.; Li, X.; Gasteiger, J. Classification of multicomponent analytical data of olive oils using different neural networks. *Anal. Chim. Acta* **1994**, *292* (3), 219-34.
- (31) Bauknecht, H.; Zell, A.; Bayer, H.; Levi, P.; Wagener, M.; Sadowski, J.; Gasteiger, J. Locating Biologically Active Compounds in Medium-Sized Heterogeneous Datasets by Topological Autocorrelation Vectors: Dopamine and Benzodiazepine Agonists. *J. Chem. Inf. Comput. Sci.* **1996**, *36* (6), 1205-1213.
- (32) Schuur, J.; Gasteiger, J. Infrared Spectra Simulation of Substituted Benzene Derivatives on the Basis of a 3D Structure Representation. *Anal. Chem.* **1997**, *69* (13), 2398-2405.
- (33) Hemmer, M. C.; Steinhauer, V.; Gasteiger, J. Deriving the 3D structure of organic molecules from their infrared spectra. *Vib. Spectrosc.* **1999**, *19* (1), 151-164.
- (34) Tarasova, A.; Burden, F.; Gasteiger, J.; Winkler, D. A. Robust modelling of solubility in supercritical carbon dioxide using Bayesian methods. *J. Mol. Graphics Modell.* **2010**, *28* (7), 593-597.
- (35) Pletscher-Frankild, S.; Palleja, A.; Tsafou, K.; Binder, J. X.; Jensen, L. J. DISEASES: Text mining and data integration of disease-gene associations. *Methods (Amsterdam, Neth.)* **2015**, *74*, 83-89.
- (36) Nguyen, D.-T.; Mandava, G.; Sheils, T.; Simeonov, A.; Southall, N.; Jadhav, A.; Guha, R.; Mathias, S.; Bologna, C.; Holmes, J.; Liu, G.; Mani, S.; Patel, J.; Sklar, L. A.; Ursu, O.; Waller, A.; Yang, J.; Oprea, T. I.; Brunak, S.; Jensen, L. J.; Fernandez, N.; Ma'ayan, A.; Rouillard, A. D.; Gaulton, A.; Hersey, A.; Karlsson, A.; Overington, J.; Liu, G.; Mehta, S.; Schurer, S.; Vidovic, D.; Mehta, S.; Patel, J.; Schurer, S.; Vidovic, D.; Sklar, L. A.; Waller, A. Pharos: Collating protein information to shed light on the druggable genome. *Nucleic Acids Res.* **2017**, *45* (D1), D995-D1002.
- (37) Dickinson, M. E.; Flenniken, A. M.; Ji, X.; Teboul, L.; Wong, M. D.; White, J. K.; Meehan, T. F.; Weninger, W. J.; Westerberg, H.; Adissu, H.; Baker, C. N.; Bower, L.; Brown, J. M.; Caddle, L. B.; Chiani, F.; Clary, D.; Cleak, J.; Daly, M. J.; Denegre, J. M.; Doe, B.; Dolan, M. E.; Edie, S. M.; Fuchs, H.; Gailus-Durner, V.; Galli, A.; Gambadoro, A.; Gallegos, J.; Guo, S.; Horner, N. R.; Hsu, C.-W.; Johnson, S. J.; Kalaga, S.; Keith, L. C.; Lanoue, L.; Lawson, T. N.; Lek, M.; Mark, M.; Marschall, S.; Mason, J.; McElwee, M. L.; Newbigging, S.; Nutter, L. M. J.; Peterson, K. A.; Ramirez-Solis, R.; Rowland, D. J.; Ryder, E.; Samocha, K. E.; Seavitt, J. R.; Selloum, M.; Szoke-Kovacs, Z.; Tamura, M.; Trainor, A. G.; Tudose, I.; Wakana, S.; Warren, J.; Wendling, O.; West, D. B.; Wong, L.; Yoshiki, A.; McKay, M.; Urban, B.; Lund, C.; Froeter, E.; LaCasse, T.; Mehalow, A.; Gordon, E.; Donahue, L. R.; Taft, R.; Kutney, P.; Dion, S.; Goodwin, L.; Kales, S.; Urban, R.; Palmer, K.; Pertuy, F.; Bitz, D.; Weber, B.; Goetz-Reiner, P.; Jacobs, H.; Le Marchand, E.; El Amri, A.; El Fertak, L.; Ennah, H.; Ali-Hadji, D.; Ayadi, A.; Wattenhofer-Donze, M.; Jacquot, S.; Andre, P.; Birling, M.-C.; Pavlovic, G.; Sorg, T.; Morse, I.; Benso, F.; Stewart, M. E.; Copley, C.; Harrison, J.; Joynson,

- S.; Guo, R.; Qu, D.; Spring, S.; Yu, L.; Ellegood, J.; Morikawa, L.; Shang, X.; Feugas, P.; Creighton, A.; Castellanos Penton, P.; Danisment, O.; Griggs, N.; Tudor, C. L.; Green, A. L.; Icoresi Mazzeo, C.; Siragher, E.; Lillistone, C.; Tuck, E.; Gleeson, D.; Sethi, D.; Bayzatinova, T.; Burvill, J.; Habib, B.; Weavers, L.; Maswood, R.; Miklejewska, E.; Woods, M.; Grau, E.; Newman, S.; Sinclair, C.; Brown, E.; Ayabe, S.; Iwama, M.; Murakami, A.; MacArthur, D. G.; Tocchini-Valentini, G. P.; Gao, X.; Flicek, P.; Bradley, A.; Skarnes, W. C.; Justice, M. J.; Parkinson, H. E.; Moore, M.; Wells, S.; Braun, R. E.; Svenson, K. L.; de Angelis, M. H.; Herault, Y.; Mohun, T.; Mallon, A.-M.; Henkelman, R. M.; Brown, S. D. M.; Adams, D. J.; et al. High-throughput discovery of novel developmental phenotypes. *Nature (London, U. K.)* **2016**, *537* (7621), 508-514.
- (38) Ursu, O.; Holmes, J.; Bologa, C. G.; Yang, J. J.; Mathias, S. L.; Nelson, S. J.; Oprea, T. I.; Knockel, J. DrugCentral: online drug compendium. *Nucleic Acids Res.* **2017**, *45* (D1), D932-D939.
- (39) Santos, R.; Ursu, O.; Gaulton, A.; Bento, A. P.; Donadi, R. S.; Bologa, C. G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T. I.; Overington, J. P. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discovery* **2017**, *16* (1), 19-34.
- (40) Lazebnik, Y. Can a biologist fix a radio? Or, what I learned while studying apoptosis. *Cancer Cell* **2002**, *2* (3), 179-182.
- (41) Jensen, A. B.; Moseley, P. L.; Oprea, T. I.; Ellesoe, S. G.; Eriksson, R.; Schmock, H.; Jensen, P. B.; Jensen, L. J.; Brunak, S. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nat. Commun.* **2014**, *5*, 4022.
- (42) Horrobin, D. F. Opinion: Modern biomedical research: an internally self-consistent universe with little contact with medical reality? *Nat. Rev. Drug Discovery* **2003**, *2* (2), 151-154.
- (43) Fujita, T.; Winkler, D. A. Understanding the Roles of the "Two QSARs". *J. Chem. Inf. Model.* **2016**, *56* (2), 269-274.
- (44) Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. Quantitative Structure-Property Relationship Modeling of Diverse Materials Properties. *Chem. Rev. (Washington, DC, U. S.)* **2012**, *112* (5), 2889-2919.
- (45) Figueiredo, M. A. T. Adaptive sparseness for supervised learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25* (9), 1150-1159.
- (46) Winkler, D. A.; Le, T. C. Performance of Deep and Shallow Neural Networks, the Universal Approximation Theorem, Activity Cliffs, and QSAR. *Mol. Inf.* **2017**, *36* (1-2), 1600118.
- (47) Burden, F. R.; Winkler, D. A. Robust QSAR models using Bayesian regularized neural networks. *J. Med. Chem.* **1999**, *42* (16), 3183-3187.
- (48) Burden, F. R.; Winkler, D. A. An Optimal Self-Pruning Neural Network and Nonlinear Descriptor Selection in QSAR. *QSAR Comb. Sci.* **2009**, *28* (10), 1092-1097.
- (49) Burden, F. R.; Winkler, D. A. Optimal sparse descriptor selection for QSAR using Bayesian methods. *QSAR Comb. Sci.* **2009**, *28* (6-7), 645-653.
- (50) Alexander, D. L. J.; Tropsha, A.; Winkler, D. A. Beware of R2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. *J. Chem. Inf. Model.* **2015**, *55* (7), 1316-1322.
- (51) Hook, A. L.; Chang, C.-Y.; Yang, J.; Lockett, J.; Cockayne, A.; Atkinson, S.; Mei, Y.; Bayston, R.; Irvine, D. J.; Langer, R.; Anderson, D. G.; Williams, P.; Davies, M. C.; Alexander, M. R. Combinatorial discovery of polymers resistant to bacterial attachment. *Nat. Biotechnol.* **2012**, *30* (9), 868-875.
- (52) Epa, V. C.; Hook, A. L.; Chang, C.; Yang, J.; Langer, R.; Anderson, D. G.; Williams, P.; Davies, M. C.; Alexander, M. R.; Winkler, D. A. Modeling and prediction of bacterial attachment to polymers. *Adv. Funct. Mater.* **2014**, *24* (14), 2085-2093.
- (53) Reginster, J. Y.; Seeman, E.; De Vernejoul, M. C.; Adami, S.; Compston, J.; Phenekos, C.; Devogelaer, J. P.; Curiel, M. D.; Sawicki, A.; Goemaere, S.; Sorensen, O. H.; Felsenberg, D.; Meunier, P. J. Strontium Ranelate Reduces the Risk of Nonvertebral Fractures in Postmenopausal Women with Osteoporosis: Treatment of Peripheral Osteoporosis (TROPOS) Study. *J. Clin. Endocrinol. Metab.* **2005**, *90* (5), 2816-2822.

- (54) Meunier, P. J.; Roux, C.; Seeman, E.; Ortolani, S.; Badurski, J. E.; Spector, T. D.; Cannata, J.; Balogh, A.; Lemmel, E.-M.; Pors-Nielsen, S.; Rizzoli, R.; Genant, H. K.; Reginster, J.-Y.; Graham, J.; Ng, K. W.; Prince, R.; Prins, J.; Seeman, E.; Wark, J.; Reginster, J. Y.; Devogelaer, J. P.; Kaufman, J. M.; Raeman, F.; Ziekenhuis, J. P.; Walravens, M.; Pors-Nielsen, S.; Beck-Nielsen, H.; Charles, P.; Sorensen, O. H.; Meunier, P. J.; Aquino, J. P.; Benhamou, C.; Blotman, F.; Bonidan, O.; Bourgeois, P.; De Vernejoul, M. C.; Dehais, J.; Fardellone, P.; Kahan, A.; Kuntz, J. L.; Marcelli, C.; Prost, A.; Vellas, B.; Weryha, G.; Lemmel, E. M.; Felsenberg, D.; Hensen, J.; Kruse, H. P.; Schmidt, W.; Semler, J.; Strucki, G.; Phenekos, C.; Balogh, A.; De Chatel, R.; Ortolani, S.; Adami, S.; Bianchi, G.; Brandi, M. L.; Cucinotta, D.; Fiore, C.; Gennari, C.; Isaia, G.; Luisetto, G.; Passariello, R.; Passeri, M.; Rovetta, G.; Tessari, L.; Badurski, J. E.; Hozzowski, K.; Lorenc, R. S.; Sawicki, A.; Diez, A.; Cannata, J. B.; Diaz Curiel, M.; Rapado, A.; Gijon, J.; Torrijos, A.; Padrino, J. M.; Roces Varela, A.; Bonjour, J. P.; Rizzoli, R.; Spector, T. D.; Clements, M.; Doyle, D. V.; Ryan, P.; Smith, I. G.; Smith, R. The effects of strontium ranelate on the risk of vertebral fracture in women with postmenopausal osteoporosis. *N. Engl. J. Med.* **2004**, *350* (5), 459-468.
- (55) Meunier, P. J. Postmenopausal osteoporosis and strontium ranelate. Reply. *N. Engl. J. Med.* **2004**, *350* (19), 2002-2003.
- (56) Autefage, H.; Gentleman, E.; Littmann, E.; Hedegaard, M. A. B.; Von Erlach, T.; O'Donnell, M.; Burden, F. R.; Winkler, D. A.; Stevens, M. M. Sparse feature selection methods identify unexpected global cellular response to strontium-containing materials. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (14), 4280-4285.
- (57) Le, T. C.; Winkler, D. A. Discovery and Optimization of Materials Using Evolutionary Approaches. *Chem. Rev. (Washington, DC, U. S.)* **2016**, *116* (10), 6107-6132.
- (58) Thornton, A. W.; Simon, C. M.; Kim, J.; Kwon, O.; Deeg, K. S.; Konstas, K.; Pas, S. J.; Hill, M. R.; Winkler, D. A.; Haranczyk, M.; Smit, B. Materials Genome in Action: Identifying the Performance Limits of Physical Hydrogen Storage. *Chem. Mater.* **2017**, *29* (7), 2844-2854.
- (59) Salahinejad, M.; Le, T. C.; Winkler, D. A. Aqueous Solubility Prediction: Do Crystal Lattice Interactions Help? *Mol. Pharm.* **2013**, *10* (7), 2757-2766.
- (60) Epa, V. C.; Yang, J.; Mei, Y.; Hook, A. L.; Langer, R.; Anderson, D. G.; Davies, M. C.; Alexander, M. R.; Winkler, D. A. Modelling human embryoid body cell adhesion to a combinatorial library of polymer surfaces. *J. Mater. Chem.* **2012**, *22* (39), 20902-20906.
- (61) Chalmers, B. A.; Xing, H.; Houston, S.; Clark, C.; Ghassabian, S.; Kuo, A.; Cao, B.; Reitsma, A.; Murray, C.-E. P.; Stok, J. E.; Boyle, G. M.; Pierce, C. J.; Littler, S. W.; Winkler, D. A.; Bernhardt, P. V.; Pasay, C.; De Voss, J. J.; McCarthy, J.; Parsons, P. G.; Walter, G. H.; Smith, M. T.; Cooper, H. M.; Nilsson, S. K.; Tsanaktsidis, J.; Savage, G. P.; Williams, C. M. Validating Eaton's Hypothesis: Cubane as a Benzene Bioisostere. *Angew. Chem., Int. Ed.* **2016**, *55* (11), 3580-3585.
- (62) Winkler, D. A.; Breedon, M.; Hughes, A. E.; Burden, F. R.; Barnard, A. S.; Harvey, T. G.; Cole, I. Towards chromate-free corrosion inhibitors: structure-property models for organic alternatives. *Green Chem.* **2014**, *16* (6), 3349-3357.
- (63) Huh, Y. H.; Noh, M.; Burden, F. R.; Chen, J. C.; Winkler, D. A.; Sherley, J. L. Sparse feature selection identifies H2A.Z as a novel, pattern-specific biomarker for asymmetrically self-renewing distributed stem cells. *Stem Cell Res.* **2015**, *14* (2), 144-154.
- (64) Celiz, A. D.; Smith, J. G. W.; Langer, R.; Anderson, D. G.; Winkler, D. A.; Barrett, D. A.; Davies, M. C.; Young, L. E.; Denning, C.; Alexander, M. R. Materials for stem cell factories of the future. *Nat. Mater.* **2014**, *13* (6), 570-579.
- (65) Epa, V. C.; Burden, F. R.; Tassa, C.; Weissleder, R.; Shaw, S.; Winkler, D. A. Modeling Biological Activities of Nanoparticles. *Nano Lett.* **2012**, *12* (11), 5808-5812.
- (66) Polley, M. J.; Winkler, D. A.; Burden, F. R. Broad-Based Quantitative Structure-Activity Relationship Modeling of Potency and Selectivity of Farnesyltransferase Inhibitors Using a Bayesian Regularized Neural Network. *J. Med. Chem.* **2004**, *47* (25), 6230-6238.

Committee Reports

ACS Council

Svetlana Korolev, Bonnie Lawlor, and Andrea Twiss-Brooks, CINF Councilors

The Council of the American Chemical Society met in Washington, DC on Wednesday, August 23, 2017 from 8:00 am until approximately 12:00 pm in the Marriott Ballrooms 1-6 of the Marriott Marquis Washington, DC hotel. There were four items for council action and they are summarized below.

Nominations and Elections

By electronic ballot, the council voted to fill the 2018-20 terms on the following three committees:

Committee on Committees: The Council elected Mitchell R. M. Bruce, Jetty Duffy-Matzner, Martha G. Hollomon, Diane Krone, Robert A. Pribush.

CINF note: [Bonnie Lawlor was recognized for her length of service 2012-17 on the Committee on Committees.](#)

Council Policy Committee: The Council elected Karl S. Booksh, Mark D. Frishberg, Zaida C. Morales-Martinez, and Linette M. Watkins for a three-year term (2018-20), and Ella L. Davis for a one year term (2018).

Committee on Nominations and Elections: The Council elected Michael Appell, Neil D. Jespersen, Mamie W. Moy, Eleanor D. Siebert, and Julianne M.D. Smist.

Petitions

On the recommendation of the Committee on Membership Affairs (MAC), a petition on *International Chemical Sciences Chapters* proposing a change to the ACS's Bylaws to allow an International Chemical Sciences Chapter to receive financial support from the Society (Bylaw IX, Section 4) failed to achieve the two-thirds majority required to amend the bylaws.

On the recommendation of the Committee on Divisional Activities (DAC), the Council defeated a proposal to establish a probationary *Division of Space Chemistry* on January 1, 2018.

On the recommendation of the Committee on Local Section Activities (LSAC), the Council approved a request by the South Jersey Local Section for annexation of the unassigned territory of Ocean County, New Jersey.

Reports of Society Committees and Committee on Science

Budget and Finance (B&F)

The Society's 2017 probable year-end projects a net from operations of \$25.3 million. This is \$2.1 million favorable to the approved budget and \$1.6 million higher than 2016. Total revenues are projected to be \$553.0 million, which is \$2.4 million unfavorable to the budget, but 5.0% higher than the prior year. Total expenses are projected at \$527.6 million, which is \$4.5 million favorable to the budget, and 4.9% higher than 2016.

On the recommendation of B&F, the board voted to approve the advance member registration fee for national meetings in 2018 at \$475; and to authorize two new program funding requests: an *ACS Online Course in Laboratory Safety*, and a *New Faculty Workshop Series*.

Education (SOCED)

The U.S. team received the outstanding record of four gold medals at the International Chemistry Olympiad in Nakhon Pathom, Thailand, July 6-15, 2017.

The American Association of Chemistry Teachers (AACT) ended May 2017 with a total of 4,314 members. Of this total, 88% are K-12 teachers of chemistry. In June, the fourth Dow & AACT Teacher Summit was held in the Philadelphia, PA area.

Since its launch in 2010, *Middle School Chemistry* (middleschoolchemistry.com), has received nearly 11 million visits from 234 countries.

ChemIDP (<https://chemidp.acs.org>) had over 1,850 logins; 11 workshops were held during spring-early summer to increase use and understanding of the tool directed toward graduate students and postdoctoral scholars.

There are currently 19,433 undergraduate student members, compared to 19,645 in June 2016.

CINF note: [Jeremy Garritano is an associate member of SOCED.](#)

Science (ComSci)

ComSci took actions on three policy statements: *A Competitive U.S. Business Climate: Innovation, Chemistry*, on *Jobs and Scientific Integrity in Public Policy*, and on *Sustainability and the Chemistry Enterprise*; and organized a symposium on *Sustaining Water Resources: Environmental and Economic Impact*, in Washington, DC.

Reports of Council Standing Committees

Membership Affairs (MAC)

MAC continues its efforts to attract and retain members. More than 23,000 new members joined the Society in 2016, including 3,685 members recruited through the Member-Get-a-Member campaign. In 2013 MAC began a series of market data collection tests including the recent incentives offering a discount for reinstating regular members (576 members rejoined the Society as a result of this program in 2016), a discount for a multi-year renewal (710 members used this auto-renewal offer), an approval of a multi-year membership for undergraduate students, and a new package to include ACS memberships as part of an institutional bundle with professional education courses, ACS Publications and CAS products.

Meetings and Expositions (M&E)

The theme of the 254th ACS National Meeting was *Chemistry's Impact on the Global Economy*. As of Tuesday evening, August 22, attendance was:

Attendees	7,938
Students	2,997
Exhibitors	1,068
Expo only	475
Guest	426
Total	12,904

Attendance at the fall national meetings since 2005 is as follows:

2005: Washington, DC	13,148
2006: San Francisco, CA	15,714
2007: Boston, MA	15,554
2008: Philadelphia, PA	13,805
2009: Washington, DC	14,129
2010: Boston, MA	14,151
2011: Denver, CO	10,076
2012: Philadelphia, PA	13,251
2013: Indianapolis, IN	10,840
2014: San Francisco, CA	15,761
2015: Boston, MA	13,888
2016: Philadelphia, PA	12,800
2017: Washington, DC	12,904

For the next spring national meeting in New Orleans, M&E will conduct an experiment with no technical programming on Thursday. For the fall meeting in Boston, there will be a discounted Expo-only pass priced at \$10.

Divisional Activities (DAC)

DAC is revising the allocation formula with an attempt to integrate additional incentives for divisions to engage more actively both at ACS regional meetings and internationally. The revised formula will be presented for council action at the 2018 national meeting in New Orleans. In order to help divisions increase their membership, DAC is investigating the feasibility of increasing the number of free one-year division memberships to new members from one to three. DAC is encouraging all 32 divisions to take advantage of the Innovative Projects Grants (up to \$7,500/project), especially for strategic planning that divisions are recommended to do every five years.

Constitution and Bylaws (C&B)

Constitution, Bylaws, and Regulations (Bulletin 5) was updated on June 1, 2017. There were two petitions for consideration at this meeting: *Election of Committee Chairs* and *the Composition of Society Committees*, for council consideration at this meeting. The deadline for new petitions to amend the bylaws must be received by the Executive Director (bylaws@acs.org) by November 29 to be included in the council agenda for the spring 2018 meeting. C&B has certified 15 bylaws since January 2017. In response to a frequently asked question, C&B clarified that graduate students and postdoctoral scholars are “regular, full” members and may be involved in all activities of the Society.

CINF note: [Svetlana Korolev is a member of C&B. Andrea Twiss-Brooks is a member of the Council Policy Committee.](#)

Reports of Other Committees

Chemical Abstracts Service (CCAS)

CCAS pursued a number of goodwill initiatives for ACS members and the chemistry community: participation in the *ACS on Campus* program, establishment of the ACS Member SciFinder benefit with 25 free searches per year, establishment of the *SciFinder Future Leaders* program, and creation with Wikipedia of the free resource *Common Chemistry* (commonchemistry.org), which includes names and CAS Registry Numbers for approximately 7,900 common chemicals. CCAS invites input regarding ways in which CCAS and CAS could further serve ACS members. Wendy Cornell, CCAS chair, wdcornell@yahoo.com.

CINF note: [Grace Baysinger and Rachelle Bienstock are members of CCAS.](#)

Chemical Safety (CCS)

In 2016 the ACS strategic plan included safety as one of its core values and the ACS journals added a new requirement for authors to emphasize any hazards or risks associated with the reported work. The society is dedicating a full-time staff member to work on chemical safety matters. The 8th (2017) edition of *Safety in Academic Chemistry Laboratories. Best Practices for the First- and Second- Year University Students* (www.acs.org/SACL) includes new sections on safety culture, changes in the OSHA Hazard Communication Standard to reflect the use of the Globally Harmonized System, and additional reorganization of the document. Two public policy statements, *Safety in the Chemistry Enterprise* and *Safety Guidelines for the Chemistry Professional – Understanding Your Role and Responsibilities*, were

developed by members of CCS, the Division of Chemical Health and Safety, and ACS public policy staff. CCS welcomes comments at safety@acs.org.

CINF note: Leah McEwen is an associate member of CCS.

Community Activities (CCA)

National Chemistry Week celebrates its 30th anniversary in 2017, October 22-26, with a theme *Chemistry Rocks!* In 2018, the program *Chemists Celebrate Earth Day* becomes *Chemists Celebrate Earth Week* during the week of April 22-28 with a theme *Dive into Marine Chemistry*.

Ethics (ETHX)

The authors from a symposium on *Ethical Consideration in Authorship* organized jointly by ETHX and CINF at the spring national meeting in San Francisco submitted a book proposal to the *ACS Symposium Series*.

CINF note: a summary of the symposium was published in [Chemical Information Bulletin](#).

ETHX will commence a ChemLuminary award for *Outstanding Local Section Programming Related to the Promotion of Ethics in Chemistry* in 2018.

CINF note: Judith Currano is a member and incoming 2018 chair of ETHX. Bonnie Lawlor has been the Committee on Committees (ConC) liaison to ETHX for the past six years.

Public Relations and Communications (CPRC)

A webinar on *Improving your Social Chemistry* on July 13, 2017 was attended by a record number of over a hundred participants. Its video recording is available for ACS members at www.acs.org/acswebinars.

CPRC co-sponsored a presidential symposium on *Science Communications: The Art of Developing a Clear Message*, and organized a joint symposium with Division of Small Chemical Businesses on *Social Media for Science Advocacy in Public Policy* at the Washington, DC meeting.

CINF note: Raychelle Burks is a member of CPRC.

Younger Chemists (YCC)

The inaugural meeting of the International Younger Chemists Network was held at the Chemistry Congress of IUPAC in July 2017 in São Paulo, Brazil. YCC members participated in the *ACS Chemistry on the Hill Advocacy Workshop* organized by President Allison Campbell at the Washington, DC meeting. The *Advocacy Toolkit* (www.acs.org/advocacy) under presidential initiative for science advocacy is rolling out this fall. YCC will host its annual *Catalyze the Vote* virtual Town Hall with the 2017 ACS President candidates, Bonnie A. Charpentier and Willie E. May, on September 28, 2017, 7:30 pm EDT, at http://bit.ly/CtV_2017a.

The Board Open Session

The board held an open session on Sunday, August 20, 2017, which featured a discussion by Glenn Ruskin and Anthony Pitagno, ACS External Affairs & Communications, on the role ACS and its members play in advocating for adoption of public policy priorities to foster scientific advancement and innovation.

Book Review

Book Review: Bibliometrics and Research Evaluation: Uses and Abuses

Robert E. (Bob) Buntrock, Buntrock Associates

Bibliometrics and Research Evaluation: Uses and Abuses; Gingras, Yves; NIT Press, Cambridge, 2016. 119p. xxii, ISBN 978-0-262-03512-5. Hardcover \$25.99.

The title of this book is a mini-abstract. The author is Professor of the History and Sociology of Science at the University of Quebec, Montreal. In his words, “An opinionated essay, not a survey of the field” and “rankings have no scientific validity”. The book is an author-translation and updated version of the original publication in French. Reviews of that version (1) and this English version (2) have been published. The book concludes with chapter notes and an index. Several of the references in the former are in French.

The introduction begins with “Since the first decade of the new millennium, the words ranking, evaluation, metrics, h-index, and Impact Factors have wreaked havoc in the world of higher education and research” (the footnote cites several of many books on the subject). The rest of the book is outlined and the history of the book provided. The book is aimed at researchers and research managers and bibliometric experts will not encounter much on technical details, except for the author’s criteria for evaluating the indicator validity.

However, along with definitions, Chapter 1, Origins, presents a concise history of citations, citation indexes, and bibliometrics. The value of citations precedes the developments of Eugene Garfield 50 to 60 years ago, but of course Garfield’s Science Citation Index solidified the field as not only a useful searching tool, but also a field of study. Scientometrics, for journal evaluation, appeared in 1978, the Journal Research Evaluation began in 1991, and extension to individuals/researchers began in the early 21st century.

Chapter 2 begins with more history, tracing the development and attributes of the Web of Science (WOS; née SCI), Elsevier’s Scopus in 2004, and eventually Google Scholar (GS). Although fee-based, the first two databases are superior to GS with author addresses and countries, bibliographies of papers cited in the article, and subfield classification. Citation searching has been extended to include patents, although use in that area remains controversial. Deficiencies and myths are discussed, including the impact of self-citation, author bylines only for first named author, Impact Factors (IFs) only for journals not books (penalizes social science and humanities), and myths like “only papers in the last five years are cited”. Types of citations are discussed (affirmative, negative, and perfunctory), and the differing value of each for evaluation is noted. Use as a value indicator for possible commercialization of research is not necessarily appropriate.

Eugene Garfield was on record that extension of Impact Factors and the *Science Citation Index* (SCI) beyond evaluation of journals was not advisable and that journal editors were encouraged to require complete citation records for publication of manuscripts, and that more data were needed for good citations than just for retrieval.

Chapter 3, Proliferation of Research Evaluation, intensifies the critique of misuse of Impact Factors and other bibliometrics. Although researchers have been evaluated for about 350 years, extension of bibliometrics as a supplement to peer review for evaluation of scholarly publications and communications, grant applications, teaching, promotions, departments and research centers, graduate programs, and universities, is a feature of the last few decades. Peer review for hiring researchers goes back two centuries, but bibliometrics began to be used in the 1970s. Citation counts are not always objective. Garfield also recommended that Nobel Prizes should not be awarded on citation counts alone, but that citation searching should be used to access articles followed by evaluation of the relevance of the citation to the evaluation of the researcher. Lysenko, the discredited, fraudulent Soviet biologist is presented as an example. He was highly cited, but mostly in negative mode. The development of the h-index in 2005 is described as well as its worth and deficiencies. It does not necessarily measure both production and the quality of that production. Index developer Hirsch maintains that the index is more democratic than others, but Gingras (and others) say not so. Normalization is difficult and the numerical value can never decrease.

The Impact Factor also comes under critical scrutiny. For 70 years, it has been touted as the measurement not only of the quality of the journal but also of the papers published within. The main source of data since 1975 is the Journal Citation Report (JCR), now based on WOS data. However, the two year “window” of data is too short. The “half-life” of publications varies by discipline and a longer period is needed for validity. The basic version also uses self-citations, so IFs are also published by deleting self-citations. However, not all self-citations are bad, but often are necessary. JCR does “blacklist” some journals for manipulation of their data. Some individuals also manipulate their data via peer review by friends. False precision is also generated by listing IFs to three decimal places, since that precision is not warranted. The Nature Index ranks countries and organizations on the publications they publish in “high-quality journals”. However, the number of journals is only 68, and these rankings put pressure on organizations to publish in these journals. Once again, individual articles are not evaluated, just the journal.

Chapter 4, Evaluation of Research Evaluation, covers just that. Metric indicators do not make evaluations that determine real value. For example, academic and non-academic organizations have different criteria. WOS and Scopus cover only peer-reviewed publications. GS has no such limit, but is coverage of Web sources really “democratization”? The h-index can be manipulated as demonstrated by 100 articles submitted under a fictitious name which yielded an h-index of 94. The evaluation market is growing and at least two private organizations have appeared, typically with contracts to universities. Such an arrangement generated controversy at Rutgers and elsewhere. Market forces have increased competition between WOS and Scopus, and coverage data is given for both. Neither indexes books, but

of course books are cited in articles in both resources. There is an English language bias, but English is becoming the international language of science.

There has been much criticism published on the unintended consequences of misapplication of metric indicators, but there is little interest in improving the meaning and accuracy of the measurements. Even the Berlin Principles demonstrate that evaluation is not equivalent to valid ranking. Both require valid indicators. Gingras lists three criteria for valid indicators: (1) adequacy of the indicator for the property or object measured, (2) sensitivity to the “inertia” or lifetime of the object, and (3) homogeneity of the dimensions of the indicator. Production is more easily evaluated than “quality” and “impact” of research, and the latter are better analyzed by surveys. The number of Nobel Prize winners associated with a university is not a good indicator, nor is presence on the Web. A good indicator varies in concert with the variability of the object being measured. Annual rankings of organizations exhibiting large variance are meaningless, and longer intervals are recommended. They tend to be only useful in marketing strategies. Combination of indicators like that done in generating the h-index becomes heterogeneous and it is difficult to assign reasons for any changes. If the values of concepts measured increase, so must a valid indicator increase. For example, there is a limit to the value of having foreign students and professors increase. These criteria were used to determine validity of both the Shanghai ranking and the h-index. For the latter, mixing the number of publications with the number of citations leads to invalidity. A lower index often obscures better researchers.

So, why are invalid indicators used? Mainly for marketing, but political reasons for funding are also involved. Such abuse is not limited to administrators, since scientists have also embraced the use of indicators, especially the h-index. Examples are given for boosting the ranking of universities using misleading figures in relative ranking. Other manipulation is also common, including “dummy” affiliations with other organizations to demonstrate international cooperation. These and other shady actions can lead to fraud.

The emphasis of the book is on abuse of the metrics by universities and administrations, but such abuses also occur at the personal and individual level. The conclusion compares abuse by using invalid indicators to the famous story of The Emperor’s New Clothes. Invalid metrics are often used in ranking of institutions and research, as well as for promotions and hiring decisions.

I have some criticisms of the book. Metrics other than the h-index are alluded to but not addressed, including the g-index, the h(2)-index, and the w-index, which admittedly have similar anomalous behavior as well as being more complicated scoring indexes (3). For comprehensive searching, citation searching should not be the only method used, especially for chemistry with excellent indexed databases. Also, with the use of citations to determine value, the citation is often to a concept not related to the subject of the paper in question. However, these are just quibbles, and the book is a valuable summation of the problems of misuse of bibliometrics, of interest to researchers, librarians, and administrators.

References

- (1) Book Review of *Les dérives de l'évaluation de la recherche*; Zitt, Michel; *J. Am. Soc. Inf. Soc. Technol.*, **2015**, *66*, 2171-2176.
- (2) Book Review of *Bibliometrics and Research Evaluation: Uses and Abuses*; Bar-Ilan, J.; *J. Am. Soc. Inf. Soc. Technol.*, **2017**, *68*, 2290-2292
- (3) *Scholarly Metrics Under the Microscope: From Citation Analysis to Academic Auditing*; Cronin, B. Sugimoto, C.R., Eds., ASIS&T/Information Today, Medford, NJ, 2015. pp. 522-524.
Reviewed in CIB, **2016**, *68* (2).

Sponsor Announcements



ChemRxiv is open for business... What you need to know

On behalf of the chemical science community, we are pleased to introduce [ChemRxiv](#), the open preprint server for the global chemistry community. You can put your research immediately out on the Web and share it with other scientists and colleagues, prior to formal peer review. ChemRxiv is openly accessible, with no subscription fees for readers and no submission charges for authors.

However, you might have a few additional questions...

But what is a preprint?

Generally speaking, a preprint is a freely accessible preliminary communication that contains new research findings and data not yet published in a peer-reviewed outlet, such as a journal.

What do my colleagues and I get out of ChemRxiv?

- ChemRxiv allows you freely to share your initial research findings with scientists around the world.
- ChemRxiv research is indexed in Chemical Abstracts and Google Scholar enhancing discovery of your research.
- ChemRxiv articles are assigned a Digital Object Identifier (DOI) upon publication allowing your preprint article to be fully citable.
- The original submitted file will be available for future download, preserving the original file.

Are there other features that ChemRxiv offers?

- Authors can easily submit preprints to ChemRxiv via a drag-and-drop Web upload.

- A link will be established between the preprint and the final published article, directing interested users to the final version of record.
- Multiple file formats are available including downloadable and shareable PDFs.
- ChemRxiv has an open API available to developers.

These can't be peer-reviewed, can they?

ChemRxiv preprints are not peer-reviewed, but will be checked for plagiarism, offensive, dangerous, highly controversial, and non-scientific content.

If I submit to ChemRxiv, can I submit the same research to a journal?

Different journals have different policies regarding the publication of research already on a preprint server. If you plan to submit research from ChemRxiv to an ACS journal, please review our [policies on prior publication](#).

Who runs ChemRxiv?

ChemRxiv is supported by strategic input from the American Chemical Society, the Royal Society of Chemistry, the German Chemical Society, and other not-for-profit organizations, as well as other scientific publishers and preprint services. ChemRxiv is managed on behalf of the chemical science community by ACS and is powered by Figshare, an online digital repository for academic research.

Where can I learn more about ChemRxiv?

View other [frequently asked questions](#) about ChemRxiv.

What do I do next?

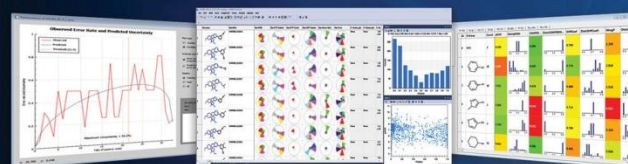
Start sharing your research today at [ChemRxiv.org](#)!

JOURNAL OF CHEMICAL INFORMATION AND MODELING

Kennie Merz, Editor-in-Chief

Journal of Chemical Information and Modeling is excited to be sponsoring the Division of Chemical Information and our team looks forward to working with many of its members in the coming months. In the past I mentioned changes at the journal which involved creating two new manuscript types: reviews and application notes. We have published several application notes describing software appropriate to chemical information and other areas and are in the process of publishing several reviews. If you have an idea for a review please contact me at eic@j cim.acs.org and we can discuss your idea further. We also are expanding our support in the area of molecular simulation and materials informatics, so please consider sending your manuscripts in these areas to the journal. All the best in 2017!

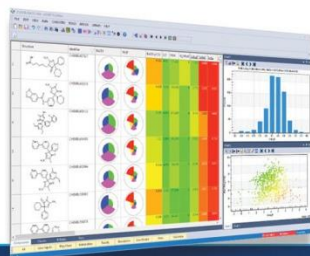
ADMET Predictor 8.5



Predict with Confidence!

With the #1 ranked ADMET property prediction software

ADMET Predictor is a software package that quickly and accurately predicts over 140 absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties. The graphical user interface is intuitive, easy to use, and contains stunning graphics that allows one to quickly assess ADMET properties. Cheminformatic features such as substructure and text queries, duplicate checking, structure standardization, and other tools are also included.



ADMET Predictor Modules

- Physicochemical and Biopharmaceutical
- Metabolism
- Toxicity
- MedChem Studio
- ADMET Modeler
- Simulation

Simulations Plus:
+1-661-723-7723
info@simulations-plus.com

Cognigen:
+1-716-633-3463
info@cognigencorp.com

DILIsym Services:
+1-919-558-1323
info@dilisym.com

The Simulation Module - New in ADMET Predictor 8.5

Predict intestinal absorption (Fa%), oral bioavailability (F%), and the oral dose required to achieve a specified plasma concentration. The algorithm uses our Advanced Compartmental Absorption and Transit (ACAT) model to simulate dissolution, transit, and absorption in the GI tract. Simulations can be performed using rat or human physiology. The liver clearance used in the algorithm can come from our *in silico* intrinsic clearance models or an *in vitro* experiment. This method can be used to filter compounds based on predicted exposure.

What else can it do?

- The MedChem Studio Module contains tools to cluster compounds containing the same scaffold, perform R-group deconvolution, matched molecular pair analysis (MMPA), and much more.
- The ADMET Modeler Module can be used to create highly accurate QSAR/QSPR models from any in-house data sets you have.
- ADMET liabilities can be determined quickly using the ADMET Risk score that evaluates each molecule based on physicochemical, biopharmaceutical, metabolism, and toxicity properties. The property values are then compared to thresholds obtained by analysis of a subset of orally available compounds in the World Drug Index (WDI).

Royal Society of Chemistry news update

The Royal Society of Chemistry is the world's leading chemistry community, advancing excellence in the chemical sciences. With over 50,000 members and a knowledge business that spans the globe, we are the United Kingdom's professional body for chemical scientists; a not-for-profit organization with 175 years of history and an international vision for the future. We promote, support and celebrate chemistry. We work to shape the future of the chemical sciences, for the benefit of science and humanity.

Our publishing portfolio publishes 44 peer-reviewed journals, more than 1,500 books, and a collection of online databases and literature updating services. For more information about our portfolio, please visit pubs.rsc.org.

Here is the latest news from our publishing portfolio.

Introducing Molecular Omics

In 2018 *Molecular BioSystems* is refocusing its scope and relaunching as *Molecular Omics* which will focus on the –omics sciences.

Molecular Omics will publish molecular-level experimental and bioinformatics research in the –omics sciences, including genomics, proteomics, transcriptomics and metabolomics. We will also welcome multidisciplinary papers presenting studies combining different types of omics, or the interface of omics and other fields such as systems biology or chemical biology.

All papers published in *Molecular BioSystems* will remain permanently available on our publishing platform. *Molecular BioSystems* will publish all 12 issues in 2017, but closed for submissions on August 16, 2017. We are now accepting articles for submissions to *Molecular Omics*.

We will continue to serve the chemical biology community with a portfolio of high quality journals: *Chemical Science*, *Chemical Society Reviews*, *Chemical Communications* and *Organic & Biomolecular Chemistry*. These journals are high quality, offering great service and very fast publication times. All four are available in PubMed and have leading associate editors.

From 2018, *Molecular Omics* will be added to our Gold package and will be available for individual subscriptions. Please contact sales@rsc.org for more information.

Read & Publish

We have launched a new pilot scheme to support the transition to open access (OA) publishing.

Corresponding authors at institutions with Read & Publish can publish gold OA in all hybrid Royal Society of Chemistry journals. At the same time, the institution will gain access to our hybrid journal portfolio.

How does it work?

1. Pay a tailored publishing fee to publish 100% gold open access.

We calculate an institution's publishing fee by analyzing the last full year's publishing output from corresponding authors. Unless the author chooses to opt out, every accepted article from the corresponding authors will be published gold open access.

2. Pay a set reading fee to unlock access to every article.

An institution then pays a set reading fee which gives the library perpetual access rights to the content published in our hybrid portfolio during the term of the contract.

Max Planck Digital Library was one of our first customers to sign up to this new agreement. Dr. Ralf Schimmer, Head of Information at the Max Planck Digital Library says:

“We regard this new agreement with the Royal Society of Chemistry as another practical step in the transition from subscription to open access as envisioned in the OA2020 initiative. With this new approach, we shift our payments and workflows in a way to make open access the default of publishing for our researchers.”

For more information on Read & Publish please contact your account manager or email sales@rsc.org.

A Global Collaborative ELN: Signals Notebook

Perkin-Elmer

The ELN Evolves to Scientific Research Data Management & Decision Support

While ELNs were initially introduced primarily for IP compliance purposes, the benefits of electronic lab notebooks have extended and embraced global collaboration and workflow support.

Today, ELN systems are positioned to manage substantial portions of R&D data in a research organization. It is a significant shift, and it means that a properly developed and implemented ELN can play a fundamental role in decision support.

Expanding Role of Today's Lab Notebook Integral to Decision-Making Process

Michael Swartz, VP of Business Development at PerkinElmer Informatics recently commented in the May issue of Lab Manager's [Informatics Resource Guide](#):

“A key objective of our approach is to provide functionality that firstly, is very adaptable to different research purposes, and secondly, makes it natural for scientists to structure their data in the course of their work. This enables the data in an ELN to be easily transferred to an analytics platform where researchers can decide what to do next. Enabling this natural transition from collaboration and workflow to decision support is probably the most important requirement for ELN's today.”

This workflow represents a fundamental transformation from ELN systems in the past which served more like document repositories. Such systems, while searchable, did not structure the data to enable decision-making and analytics.

The world of science is changing faster than ever, and scientists need an electronic lab notebook with the power and flexibility to change with it.

PerkinElmer's new Signals Notebook Web-based ELN ([watch the video](#)) also delivers an effective scientific research data management solution. Write up your research data in notebooks and experiments, then drag and drop, store, organize, share, find and filter data with ease. All scientific data are electronically recorded and stored, making it simple to be more effective, reproducible, and accurate in your scientific endeavors.



It's Not a Notebook...It's a *Place*.

Signals Notebook is a centralized, secure Web-based ecosystem that allows your team access anytime, from anywhere. It is a place where people, ideas and data come together to collaborate, seamlessly share data and discover crucial insights.

Signals Notebook is More Than an ELN

As mentioned above, ELNs are nothing new. In fact, PerkinElmer has been an industry leader in the space for years. But Signals Notebook reinvents the ELN for today's science...and beyond.

We have incorporated the market-leading chemical drawing platform, ChemDraw, available without any additional install. We've also fully integrated Signals Notebook with Microsoft Office and Microsoft Office Online. You can now effortlessly create or attach your Office documents with your experiments and update them.

Signals Notebook – The Global ELN for a Collaborative Scientific World

Connecting and sharing with colleagues and collaborators around the world has never been easier. Signals Notebook lets you start discussions, provide feedback, coordinate follow-on experiments, and stay in sync, anywhere and anytime.

Scientific “Eureka” Moments Happen Faster Than Ever Before

With the support of PerkinElmer, you can be up and running in minutes! There is no software to install, since the platform is 100% Web-based. With no downloads needed, no hardware to buy, and no IT assets to maintain, Signals Notebook provides immediate Return on Investment (ROI) for budget-minded science teams.

Achieve meaningful scientific breakthroughs with PerkinElmer Signals Notebook. [Watch a short video](#) or [learn more about PerkinElmer's powerful new Web based ELN - Signals Notebook](#).

CINF Officers and Functionaries

Chair and Chair-Elect

Erin Davis,
Schrödinger, Inc.
erindavis@gmail.com

Past-Chair

Rachelle Bienstock,
RJB Computational Modeling LLC
rachelleb1@gmail.com

Secretary

Tina Qin,
Vanderbilt University
ginna@mail.lib.msu.edu

Treasurer

Rob McFarland,
Washington University
rmcfarland@wustl.edu

CINF Councilors

Bonnie Lawlor,
chescot@aol.com
Andrea Twiss-Brooks,
University of Chicago
atbrooks@uchicago.edu
Svetlana N. Korolev,
University of Wisconsin, Milwaukee
skorolev@uwm.edu

CINF Alternate Councilors

Carmen Nitsche,
carmen@cinformaconsulting.com
Charles Huber,
University of California, Santa Barbara
huber@library.ucsb.edu
Jeremy Ross Garritano,
University of Virginia
jg9jh@virginia.edu

Archivist/Historian

Bonnie Lawlor,
chescot@aol.com

Audit Committee Chair

TBD

Awards Committee Chair

David Evans,
RELX
david.evans@relx.ch

Careers Committee Co-Chairs

Pamela Scott,
Pfizer
pamela.j.scott@pfizer.com
Sue Cardinal,
University of Rochester
scardinal@library.rochester.edu

Communications and Publications Committee Chair

Graham Douglas,
communications@acscinf.org

Procedures Chair

Bonnie Lawlor,
chescot@aol.com

Education Committee Chair

Grace Baysinger,
Stanford University
graceb@stanford.edu

Finance Committee Chair

Rob McFarland,
Washington University
rmcfarland@wustl.edu

Fundraising Interim Committee Chair

Graham Douglas,
communications@acscinf.org

Membership Committee Chair

Donna Wrublewski,
Caltech Library
dtwrub@caltech.edu

Nomination Committee Chair

Rachelle Bienstock,
RJB Computational Modeling LLC
rachelleb1@gmail.com

Program Committee Chair

Elsa Alvaro,
Northwestern University
elsa.alvaro@northwestern.edu

Tellers Committee Chair

Sue Cardinal,
University of Rochester
scardinal@library.rochester.edu

***Chemical Information Bulletin* Editor Spring**

Vincent F. Scalfani,
The University of Alabama
vfscalfani@ua.edu

***Chemical Information Bulletin* Editor Summer**

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

***Chemical Information Bulletin* Editor Fall**

Teri Vogel,
UC San Diego Library
tmvogel@ucsd.edu

***Chemical Information Bulletin* Editor Winter**

David Shobe,
Patent Information Agent
avidshobe@yahoo.com

Webmaster

Stuart Chalk,
University of North Florida
schalk@unf.edu

Contributors to This Issue

Articles and Features

Elsa Alvaro
Rachelle Bienstock
Robert E. Buntrock
Erin Davis
Jeremy Garritano
Svetlana Korolev
Carmen Nitsche
David Shobe
Wendy Warr

Awards announcements

David Evans

Committee Report

Svetlana Korolev
Bonnie Lawlor
Andrea Twiss-Brooks

Sponsor Information

Graham Douglas

Cover photo

Ken Lund, under [Creative Commons License](#)

Production

Stuart Chalk
Judith Currano
Svetlana Korolev
Bonnie Lawlor
David Shobe
Wendy Warr