Inside this issue:

Message from the CINF Chair
Message from the Editor
Message from the Program Chairs
Announcements
  Judith Currano to be inducted as a 2020 ACS Honorary Fellow
  Dr. Stephen R. Heller to be inducted as a 2020 ACS Honorary Fellow
Awards and Scholarships
Reports
  Report on the Council Agenda for August 19, 2020
Book Reviews
  The Misinformation Age; How False Beliefs Spread
  3D Printing in Chemical Sciences: Applications across Chemistry.
Twenty-five Years Ago in Chicago
Science and Popular Culture, 5
Sponsor Announcements
2020 CINF Officers and Functionaries
Fall 2020 CIB Contributors
Fall 2020 National Meeting—Technical Program Broadcast Schedule
Fall 2020 National Meeting—Technical Program On-Demand Schedule
Fall 2020 National Meeting—Technical Program With Abstracts
Message from the Chair

Since last I wrote, the world has significantly changed. Priorities have shifted, attentions have been split, as for many of us, and our personal and professional worlds have combined during various stages of quarantine. Routines have been disrupted, while connections and communications have been harder to maintain or diminished due to a reliance on virtual meetings. CINF’s work is only accomplished by volunteers, and I appreciate all of those that have continued to move CINF forward during these difficult times. Our efforts have slowed, but not stopped. I intend to work with Chair-Elect, Donna Wrublewski, to reprioritize the CINF Strategic Plan and resume some of the implementation teams.

While the virtual ACS National Meeting is not free, I hope that you can join us if possible. The Meeting will be an experiment and the more of our members that participate, the more feedback we can collect in order to adapt should there continue to be virtual-only Meetings or a move to hybrid Meetings.

Our Program Chairs, Sue Cardinal and Ye Li, have shown extreme patience, dedication and flexibility when it comes to the CINF program this fall. They have juggled keeping symposium organizers and speakers informed as over the last several months ACS deliberated with making implementation decisions and communicating out options for the National Meeting. Besides our regular programming, we will have additional opportunities for networking, a vendor roundtable, and some informal roundtables for sharing information about remote instructional services.

We had to make the tough decision in postponing our Skolnik symposium in honor of Wendy Warr, moving it to the ACS National Meeting next fall scheduled for Atlanta, GA. Other technical divisions have also struggled with how to appropriately honor awardees. Given Wendy’s dedication to both the profession and CINF, our Awards Committee agreed to move the Skolnik symposium and delay the awarding of another Skolnik award until 2022.

CINF members: I hope that you took the opportunity to vote in our latest election. If you are ever interested in running for an office or would like more information about what various positions do or what opportunities are available, please reach out to me.

Thank you again for being involved with CINF, even if it only means you read this Bulletin. Networking is more important than ever, and I welcome your feedback on what CINF can do to make your membership more valuable.

Jeremy Garritano, CINF Chair
2020….

I am back to editing my regular issue of the *Chemical Information Bulletin*, in time for a very different meeting than what we were all expecting at the beginning of 2020. But there will be a meeting this time, thanks to our presenters, session organizers, and programming chairs Sue Cardinal and Ye Li. Please review their message on the next page summarizing the meeting schedule and events. Some of these sessions were rescheduled from the cancelled spring meeting, with others being deferred to next year.

And a special congratulations to newly inducted ACS Fellows, honored for their service to the profession and to ACS (https://cen.acs.org/acs-news programs/2020-ACS-fellows/98/i29).

- Stephen Heller (InChI Trust and NIH/PubChem)
- Judith Currano (University of Pennsylvania)

In this issue, you will find information about the upcoming meeting, including the broadcast schedule, the on-demand recorded sessions, and the abstracts. We also have two book reviews, Wendy Warr’s memories of the ACS meeting from 25 years ago, and a list of some recent articles, videos, and podcasts about science and popular culture.

If you are looking for another way to get involved in CINF, Judith, David and I welcome articles, book reviews, interviews and other features that would be of interest to CINF members. There is also an opening for a *CIB* editor for the spring issue that comes out before the national meeting. Please contact one of us if you have questions or want to find out more about what is involved in putting together an issue.

Hopefully I’ll “see” you at the national meeting.

---

Teri M. Vogel, Fall *CIB* Editor  
UC San Diego  
tmvogel@ucsd.edu
Please register and attend our first virtual CINF Division program using a computer near you, Aug 17 - 20th. You will have your choice of ~60 CINF speakers. There will be two broadcast symposia with live Q&A every day, Monday - Thursday at 10-12 and 1-3 pm pacific time. Additionally, there will be ~20 “on-demand” presentations that can be viewed any time during the conference and for a week afterwards. If you can’t register, several of the talks may be available after the conference via SciMeetings, so check that out. Topics for our symposia include Machine Learning and Artificial Intelligence, Making Chemistry FAIRer, Leveraging Artificial Intelligence & Advanced Computational Methods to Solve Hard Problems, Reaction Prediction & Synthesis Planning, Moving Chemistry from the Lab into the Open, and Cheminformatic Methods, Tools and Models. Details are in this Chemical Information Bulletin (CIB) with updates in the online planner - see https://plan.core-apps.com/acs_sf20/customScreen/aboutShow.

New this meeting are some Zoom socials & roundtables. Follow the direction from the conference platform to join technical division networking rooms and choose CINF at the times listed below.

CINF Networking Lunch
Casual lunch "reception" Bring your own lunch/ snack and say hi.
Monday, Aug 17 12 - 1 pm Pacific Time

Remote Instructional Services Roundtable 1
Opportunity to exchange tips, ideas and best practices for delivering remote chemical information instruction and instructional support services.
Tuesday, Aug 18 7-8 am Pacific Time

Vendor Roundtable
Vendors and spring meeting sponsors to give a 10 mins talk with Q&A
Wednesday, Aug 19 12 - 1 pm Pacific Time

Remote Instructional Services Roundtable 2
Opportunity to exchange tips, ideas and best practices for delivering remote chemical information instruction and instructional support services.
Thursday, Aug 20 9-10 am Pacific Time

A special Thank You to organizers, speakers and program committee members. This was an especially tough planning cycle as the rules (from in-person to online) kept changing and sessions kept getting revised, some rescheduled and some ended. Fortunately, ACS CINF is not a lifetime commitment. Join us when you can because it takes many hands to develop an excellent program. The next programming meeting will be on August 16 (Sunday) at 1 pm eastern time to discuss the future program. If you are interested, I’ll send you the zoom link.

Want to speak? The Call for Papers for the Spring 2021 (San Antonio?) meeting will be open for abstract submission on ~ September 21 - October 26th. Please don’t wait until the last minute. The Spring 2021 meeting with the theme “Macromolecular Chemistry: The Second Century” is scheduled for March 21 - 25, 2021.

Do you have an idea for a great program? Symposium and organizer ideas can be sent any time via email to committee members or via this Google form at https://forms.gle/UUR1m5kBtbXBkkCH6. Thank you to all that have suggested topical symposia in past. Would you like to organize a program or recommend someone else? The next opportunities for these programs and organizers will be for the 2021 meetings.

If your program is ready to be submitted to the Call for Papers for the Spring 2021 meeting, please use this Google Form at https://forms.gle/RDAvLZXWhkEQdA8e7. These will be due very soon.
(mid to late August) for Spring in late November 2021 for Fall. Would you like to join the Programming Committee? Some Committee members are listed on our website (it may need updating) at https://acscinf.org/content/program. We are always looking for people to sort through the ideas, to organize, or assist organizers, and to manage the logistics of putting the program together. It is a great way to network and to keep up with the trends in our field.

“See” you at the conference!

Ye Li, 2020-2021 Program Planner
yel@mit.edu

Sue Cardinal, 2019-2020 Program Planner
scardinal@library.rochester.edu
Announcements

Judith Currano to be inducted as a 2020 ACS Honorary Fellow

It is our pleasure to share the news that our friend and colleague Ms. Judith N. Currano will be honored as an ACS Fellow this year. Judith’s significant contributions to both chemical information education, and its close association with scientific ethics, have been inspiring to a wide range of people involved in the chemistry enterprise. Her dedication to her profession, and to, in her own words, “improving methods of communication between current scientists,”(1) is an inspiration to all those who are fortunate enough to work with her.

Judith is the Head of the Chemistry Library at the University of Pennsylvania, a position she has held since 1999. She is responsible for chemical resource collection development, as well as reference and instruction around chemical databases, information sources and other bibliographic tools. In addition to her library duties, she serves as an advisor to Penn’s Women in Chemistry group.

Judith teaches “CHEM 601 Chemical Information: Organization and Retrieval,” a credit course which is required for all chemistry PhD students (https://www.chem.upenn.edu/node/9352). Many chemistry librarians would consider a mandatory class for graduate students as something of a holy grail, a guaranteed venue to interact and train students in a consistent way that could potentially address student questions in a timely and organized way, rather than the more scattered approach of interacting with individual students as they need assistance. Chemistry librarians are fortunate that she has taken the time to share her experiences involving this class with the community.(2,3)

Judith has been a member of ACS since 1999, and has been active both in her Philadelphia Local Section as well as Technical Divisions and Governance Committees. We have had the privilege of working with Judith in the Chemical Information Division, where she has served (and continues to serve!) several leadership roles. She served as Secretary in 2008-2009 and Division Chair in 2014, and has had continuing involvement with both the Education Committee and the Chemical Information Bulletin editorial team.

Judith’s national service includes time as a member on the Committee on Chemical Abstracts Service and the Committee on Committees. She is probably most recently recognized for her work in scientific ethics, and is the current Chair of the ACS Committee on Ethics. In addition to comments highlighting the work of the Ethics Committee to the entire chemical community and beyond in Chemical & Engineering News,(4-6) she also co-edited and contributed to the 2018 ACS Symposium Series volume Credit Where Credit Is Due: Respecting Authorship And Intellectual Property.(7) In a time where it seems to many that there are more questions (and questionable practices) than answers, Judith’s steady voice and dedication to the Committee’s vision of “leading a culture of ethics in chemistry”(3) is a welcome relief.

In addition to ACS, Judith has been an active member in the Special Libraries Association (SLA) Chemistry Division since 1999, serving as Division Chair, Professional Development Chair, and ACS CINF Division Liaison. Applying her expertise in teaching the credit class at Penn, she has also developed and taught several classes, both in-person and webinars, for librarians through SLA’s Continuing Education program. Topics have ranged from the introductory “Chemistry for the Non-Chemist Librarian” to advanced structure searching in common chemistry databases. In 2016, her teaching and mentoring excellence was recognized by SLA with the Rose L. Vormelker Award.(8)

Judith has given over two dozen talks at ACS national meetings on topics ranging from teaching chemical structure searching to chemical information integrity, and has also contributed several chapters to ACS Symposium Series volumes. This also made her uniquely qualified to co-edit RSC’s 2014 book volume Chemical Information for Chemists with Dana L. Roth, another well-known and highly respected figure in the field of chemical librarianship.

Her demonstrated expertise with many chemical information products and databases had led to her
Announcements

Judith Currano to be inducted as a 2020 ACS Honorary Fellow

appointment on several advisory committees, including serving as the Chair of the CCDC’s Board of Trustees, the first librarian to do so.(9)

Judith’s contributions to our field cannot be understated, and we are incredibly fortunate that she is both a member of our community and a dedicated, ethical teacher, mentor and friend to many of us. Congratulations on this well-deserved honor!

Svetlana N. Korolev, CINF Councillor (2016-2019)

Donna T. Wrublewski, CINF Chair-Elect (2020)

References (all accessed August 9, 2020)

(1) American Chemical Society. Philadelphia Local Section Councillor Statements https://phillyacs.org/councillor/


(9) Judith Currano. Trustees. Cambridge Crystallographic Data Centre. https://www.ccdc.cam.ac.uk/ theccdcprofile/trustees/?id=fd933eda-7fd7-4955-bc52-9c45d967f222
Announcements

Dr. Stephen R. Heller to be inducted as a 2020 ACS Honorary Fellow

I am pleased to report that Dr. Stephen R. Heller will be inducted this year as an ACS Honorary Fellow. I have known and worked with Steve for decades, both as a business person and as a volunteer for scientific societies. Based upon his expertise in building chemical information databases, Eugene Garfield often called upon Steve as an advisor when we were about to embark on a new venture at the then Institute for Scientific Information (ISI) where I was the Director of Chemical Information Services. Steve was and remains an invaluable and creative resource - a straight shooter who gives honest advice. Working with him taught me a lot and I am very grateful. I have also worked with him in CINF well as in other scientific societies such as the International Union of Pure and Applied Chemistry (IUPAC) and the Chemical Structure Association Trust. He has always proven to be innovative, tenacious, and vocal in moving things forward, and tireless in his volunteer efforts to serve the chemistry profession. Around the globe his many accomplishments are an inspiration to those new to the field of chemical information/cheminformatics.

Today, Steve is probably most known for his work in the development of the InChI (https://www.inchi-trust.org/) code, an acronym for IUPAC International Chemical Identifier. The code provides a string of characters representing a unique digital signature for a chemical substance. It is a free, nonproprietary identifier that is currently used in printed and electronic data sources utilized by professional chemical database providers, publishers, chemistry software vendors, librarians, patent attorneys, and information specialists around the world. Virtually every Pharma company in the world now uses InChI internally, as does Chemical Abstracts Service (CAS). Sigma Aldrich is in the process of putting the InChI on every chemical bottle it sells. Steve continues his InChI development work as Chair of IUPAC Division VIII’s Subcommittee on the IUPAC Chemical Identifier (InChI).

Steve also initiated the NIH/EPA/NIST mass spectrometry dataset that is the primary tool used by the EPA for the identification of pollutants. Today, with more than 200,000 entries, the database is distributed by NIST to more than 3,000 users each year. He is also recognized for his crossover work into bioinformatics, where he developed computer-based plant genome mapping and sequencing databases at the U.S. Department of Agriculture (USDA). He continues to be active in this area where he served as Founder and Chairman of the Plant & Animal Genome (PAG) Conference (1992-2017).

His ACS volunteer efforts include service to the ACS Division of Computers in Chemistry (COMP) and to CINF. He was a Councilor for COMP from 1988-1995 and chair of its nomination committee from 1986-1990. In parallel, Steve was equally involved with CINF, organizing and speaking at symposia; his service in both Divisions facilitated improved communication between the two units, resulting in increased programming collaboration which continues to this day. In the early days of desktop computing he identified the need for software reviews and initiated the software review section of the ACS Journal of Chemical Information and Computer Sciences (JCICS) and its successor, the Journal of Chemical Information and Modeling (JCIM). For almost 20 years (1987-2015) he served as the volunteer, unpaid software review editor, soliciting and writing software reviews on computer programs for and about chemistry. In addition to other awards, Steve has been recognized by CINF via the Herman Skolnik Award in 2000 (https://acscinf.org/content/herman-skolnik-award), and in 2015 by the ACS Dayton and Columbus, Ohio local sections with their Patterson-Crane Award (https://www.acs.org/content/acs/en/funding-and-awards/awards/acs-local-section-awards/dayton-local-section/dayton-patterson-crane-award.html).

I could go on, but I will stop here and offer a round of congratulations to Steve!!!!!!

Bonnie Lawlor,
ACS Honorary Fellow/CINF Councilor
Applications Invited for CSA Trust Grants for 2021

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 2021.

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. Proposals from those who have not received a Grant in the past will be given preference. 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 2021 Grant is April 16, 2021. Successful applicants will be notified no later than May 24, 2021.

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.
Recent Grant Awardees

2020

Daniel Csókás, a member of Professor Imre Pápai’s research team at the Research Centre for Natural Sciences, Budapest, Hungary was awarded $4,000 to travel to the University of Bristol (UK) to expand the scope of his experience in computational chemistry and to acquire new skills and research techniques in the area of data-led catalyst design. The project will involve the creation of a ligand knowledge base for tridentate ligands using calculated descriptors. The database will then be processed to retrieve structural and reactivity information about tridentate ligands and their transition metal complexes. The award is pending the lifting of travel restrictions due to the pandemic.

Andrew Tarzia, a Research Associate at the Imperial College London, was awarded $3,500 to visit Asst. Prof. Cory Simon at Oregon State University for three weeks in 2021 to initiate a collaboration in the use of machine learning algorithms to predict host-guest finding affinities based upon molecular shapes. The award is pending the lifting of travel restrictions due to the pandemic.

Nicola Knight, an Enterprise Research Fellow - Physical Sciences Data-Science Service (PSDS) at the University of Southampton (UK) where she works with the newly-established national research facility to provide access to chemistry and physical sciences data at a national scale and to increase not only the breadth of the data, but also the ways in which the data can be used by the scientific community. She was awarded $2,500 to fund efforts related to a knowledge sharing retreat that will involve four early-career researchers (ECRs) from a cross section of research domains to participate in a 3-day workshop on the depiction of chemical information using the FAIR principles.

2019

Vinicius Alves, University of North Carolina Eshelman School of Pharmacy, Chapel Hill (U.S.A.), was awarded $2,572 to present his research paper entitled “Multi-Descriptor Read Across (MuDRA) as a novel computational approach for Chemical Toxicity Prediction" at the 10th International Symposium on Computational Methods in Toxicology and Pharmacology Integrating Internet Resources that was held in Ionnina, Greece, from June 23-27, 2019.

Guilian Luchini, Colorado State University, Fort Collins, CO (U.S.A.), was awarded $1,399.00 to attend the American Chemical Society meeting that was held from August 24-29 in San Diego, CA, where he presented his research in applying often-overlooked corrections to DFT frequency calculations in an automated fashion.

Roi Rutenberg, Chemistry Department, Stanford University, Stanford, CA (U.S.A.), was awarded $2,072 for travel to visit the University of Illinois, Chicago in order to model molecular dynamic (MD) simulations at the Kral group as part of his research related to retrieving information about pEtN cellulose’s chemical structure as an individual compound and as a partner in future chemical reactions.

Monika Szabo, Monash Institute of Pharmaceutical Sciences, Monash University, Victoria, Australia, was awarded $2,000.00 for travel to attend two conferences at which she presented her research on drugs for myelofibrosis. The conferences were EFMC-ASMC’19 International Symposium on Advances in Synthetic and Medicinal Chemistry, held in Athens, Greece, from September 1-5, 2019, and the 20th SCI/RSC Medicinal Chemistry Symposium, held in Cambridge, U.K., from September 8-11, 2019.

Chemical Information Bulletin, 2020, 72 (3)
Recent Grant Awardees

2018

Stephen Capuzzi, Division of Chemical Biology and Medicinal Chemistry at the University of North Carolina Eshelman School of Pharmacy, Chapel Hill (USA), was awarded a Grant to attend the 31th ICAR in Porto, Portugal from 06/11/2018 to 06/15/2018, where he presented his research entitled “ComputerAided Discovery and Characterization of Novel Ebola Virus Inhibitors.”

Christopher Cooper, Cavendish Laboratory, University of Cambridge, UK, was awarded a Grant to present his current research on systematic, high-throughput screening of organic dyes for co-sensitized dye-sensitized solar cells. He presented his work at the Solar Energy Conversion Gordon Research Conference and Seminar held June 16-22, 2018 in Hong Kong.

Mark Driver, Chemistry Department, University of Cambridge, UK, was awarded a Grant to offset costs to attend the 7th EUCheMS conference where he will present a poster on his research that focuses on the development and applications of a theoretical approach to model hydrogen bonding.

Geqing Wang, La Trobe Institute for Molecular Sciences, La Trobe University, Australia, was awarded a Grant to present his work at the Fragment-Based Lead Discovery Conference (FBLD2018) in San Diego, USA in October 2018. The current focus of his work is the development of novel antivirulence drugs which potentially overcome the problems of antibiotic resistance of Gram-negative bacteria.

Roshan Singh, University of Oxford, UK, was awarded a Grant to conduct research within Dr. Marcus Lundberg’s Group at Uppsala University, Sweden, as part of a collaboration that he has set up between them and Professor Edward Solomon’s Group at Stanford University, California. He conducts research within Professor John McGrady’s group at the University of Oxford. The collaboration will look to consolidate the experiments studies on heme Fe (IV)=O complexes currently being studied by Solomon’s Group with future multi-reference calculations to be conducted within Lundberg’s Group.

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.
Recent Grant Awardees

**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.
Reports

Report on the Council Agenda for August 19, 2020

The Council of the American Chemical Society will meet virtually on Wednesday, August 19, 2020 from 11:00am until approximately 1:00pm. Up for Council vote are four routine action items, one urgent action item, and another action item calling for the dissolution of the Committee on Chemical Abstracts.

Nominations and Elections

The Committee on Nominations and Elections has announced the list of ten candidates for membership on the Council Policy Committee (CPC). These are: Gary D. Anderson; Elizabeth M. Howson; Brian B. Brady; James M. Landis; Mitchell R. M. Bruce; Zaida C. Morales-Martinez; James C. Carver; Margaret J. Schooler; Mark D. Frishberg; and Jeanette M. Van Emon.

Council must elect five individuals. The four candidates receiving the highest numbers of votes will be declared elected for the 2021-2023 term, and the candidate receiving the fifth highest vote will be declared elected for a one-year term for 2021.

The slate of candidates for membership on the Committee on Nominations and Elections has also been announced. Candidates are: Thomas R. Beattie; Michael J. Kenney; Mark A. Benvenuto; R. Daniel Libby; Jetty L. Duffy-Matzner; Robert A. Pribush; Kevin J. Edgar; Julianne M. D. Smist; Neil D. Jespersen; and Linette M. Watkins. The Council must elect five individuals: The five candidates receiving the highest numbers of votes will be declared elected for the 2021-2023 term. And finally, the slate of candidates for membership on the Committee on Committees has been announced. Candidates are: Catherine E. Costello; Michael D. Mosher; Debbie M. Decker; Sarah M. Mullins; Harry J. Elston; Susan J. Olesik; Martha G. Hollomon; Andrea B. Twiss-Brooks; Diane Krone; and Javier Vela. The Council must elect five individuals. The five candidates receiving the highest numbers of votes will be declared elected for the 2021-2023 term.

Petitions for Vote

Petition to Clarify Amendments to the Standing Rules Bylaw XI

Explanatory: Standing Rule V states that the Committee on Constitution and Bylaws (C&B) is responsible for the Governing Documents Function. The duties of C&B are listed in Sec. 1, b of this Standing Rule and include the following:

a. review provisions of the Constitution, Bylaws, and Standing Rules and initiate such action as may seem appropriate; and

b. interpret and initiate such action as may seem appropriate to eliminate conflicts in the Constitution, Bylaws, or Standing Rules.

However, the procedure in Bylaw XI for making amendments to the Standing Rules contains no provision for C&B’s review before action by Council. This could lead to problems and inconsistencies within the Standing Rules.

The Standing Rules include many of the provisions that formerly were in the Bylaws. C&B’s review was required for Constitution and Bylaw changes in the former version of the ACS Governing Documents and is still required for amendments to the Constitution (Article XIII, Sec. 2) and amendments to the Bylaws (Bylaw X, Sec. 2, a). To ensure consistency within all documents in the ACS Governing Documents, including the Standing Rules, it is imperative that Bylaw XI be amended and that Council approve this urgent action petition.
Report on the Council Agenda for March 25, 2020

The Society Committee on Budget and Finance has examined this petition and concludes that it will have no impact on the finances of the Society ($0).

Petition to charter a new International Chemical Sciences Chapter

One legal application has been received for the formation of a new international chemical sciences chapter to be known as the Israel International Chemical Sciences Chapter. The Israel International Chemical Sciences Chapter will consist of the country of Israel and is not part of any other Chapter or Local Section of the Society.

The petition was initiated and signed by ACS members in good standing and residing in the territory. The application meets all of the requirements of Bylaw VI and Standing Rule VII of the Governing Documents, and includes a statement that the applicants are familiar with and will abide by all governing documents of the Society including specifically Standing Rule VII Section 3(d), which states that the Chapter and its officers as representatives of the Chapter shall not engage in political activity, shall avoid any activities that may adversely affect the interests and/or public and professional image of the Society, and shall assure that all activities of the Chapter shall be open to all members of the Society. The application includes a proposed budget for the operation of the Chapter, which includes no allotment of funds from the Society. The petition has been reviewed by the ACS Joint-Board Committee on International Activities (IAC). This action seeks the approval of the Council and is contingent on the approval from the ACS Board of Directors, after which, the Chapter will begin operation.

Petitions for Consideration

There are no petitions for consideration listed in the Council Agenda Book.

Recommendation on the Dissolution of the Committee on ACTION Chemical Abstracts Service

ConC voted in October 2019 to recommend to Council that it disband the Joint Board-Council Committee on Chemical Abstracts Service (CCAS). ConC reviewed CCAS for multiple cycles, and after numerous and thoughtful discussions came to the conclusion that most, if not all, of the responsibilities defined in the CCAS charter are already being performed by Chemical Abstracts Service (CAS) itself. CAS already has in place many tools to regularly communicate with ACS members, Society governance, and customers. These include: 1) quarterly meetings with the Governing Board for Publishing; 2) quarterly meetings with the ACS Board of Directors; 3) routine staff participation in ACS local section meeting and events; 4) daily CAS customer service inquiries; 5) customer service field representatives who meet regularly with customers; 6) global event participation with representatives fielding questions, engaging in product demonstrations, etc.; 7) formal and informal customer surveys; 8) proactively working with the Membership Division to understand ways CAS can interact with them to add value to membership; and 9) in collaboration with ACS Membership, SciFinder® Member Benefit program outreach. CAS plays a key role in supporting the overall financial operations of ACS. Cognizant of the importance to ACS while aware of threats from competitive organizations, CAS has become very dynamic in its market analysis of its customer base that has led to new products, including those that are responsive to individual user needs. The ACS Governing Board for Publishing, a Board Committee, comprised of ACS members, experts, and business leaders also provides oversight and guidance to CAS. CAS has outpaced the need for committee support and the need for CCAS, and therefore the need for this committee has gone away. ConC values the time, talents, and energy of ACS volunteers invest in our committees and believes these investments should be impactful. With the continued growth and success of CAS over the years, a valuable and meaningful role for CCAS could not be found. This was not an easy decision for ConC,
Report on the Council Agenda for March 25, 2020

and we are making every effort to ensure that this recommendation to disband CCAS is managed with transparency and respect. If Council approves the recommendation, it will then go to the ACS Board of Directors for ratification.

Councilors and other interested members are invited to attend the committee's virtual open meeting on Monday, August 17, 2020, from 11:00 a.m.-12:00 p.m. to offer their views on any topics on the agenda, or other matters of interest.

Note: The Council Agenda Book can be accessed at: https://www.acs.org/content/acs/en/about/governance/councilors.html. It is the first item listed under “Fall Meeting.”

Respectfully submitted, August 6, 2020

Bonnie Lawlor
Andrea Twiss-Brooks
CINF Councillors
This is a timely book on an important topic. The authors are both professors of logic and philosophy of science at the University of California Irvine. “Fake news” is nothing new except the channels of dissemination of all information have grown rapidly, especially in this era of social networks. The introduction summarizes the rest of the book. Beliefs, in truth and knowledge, science and evidence, are discussed: how they are formed, how they persist, why they spread, why they are intransigent in the face of contrary evidence, and how they can be changed.

“Belief [def. 3, Webster’s New Collegiate Dictionary]: conviction of the truth of some statement or the reality of some being or phenomenon, especially when based on examination of evidence.” This definition works when applied in scientific discovery, research, or communication, but can have the overtones of religious tenets for non-scientists, especially when they are sparring with scientists. Making accusations that scientists are exhibiting “believing” in a religion when discussing evolution or climate change is typical. That’s why I avoid the use of “belief” when discussing these topics and use “acceptance” instead.

False beliefs are those inconsistent with available evidence. They can be the result of cognitive biases or blind spots, quirks of human psychology, lack of experience, or poor education, or “too stupid” (this last is to be avoided in discussion even in retaliation when accused of being the same). Beliefs tend to originate in what is told to us by others. The information can be both bad and good. Misinformation is influenced by the media and propaganda, both political and industrial. Extreme partisan stands are becoming more common, and the battles over the fluoridation of municipal water systems are exemplary. Currently, disregard for evidence, expert knowledge, and logical coherence is on the rise. Group beliefs rather than just individual beliefs are increasing. Echo chambers” do not necessarily lead to inaccurate and possibly fraudulent results, but they do influence beliefs. Even scientists can persist in false beliefs, due to social factors. The ways that science and scientific research are publicized and spread is key. “Fair” reporting of two conflicting sides of a scientific debate may lead to equal credence for both sides even if unwarranted. Scientific debate can bias what the public sees. More information doesn’t necessarily solve the problem.

Chapter 1, “What is Truth”, describes definitions of truth throughout the years beginning with Pilate vs. Christ and concentrating on examples within science and involving discussions on what is reality, Bayes analysis, and the philosophies of Kuhn and Popper.

The discovery of the Antarctic ozone hole is an example. The announcement by Rowland and Molina led to a public debate between them and DuPont who stressed a “wait and see” approach since they required certainty. Further examples have led to the Johnson/McNamara escalation of the Vietnam War, the infamous “alternative facts” propositions of Rove and Conway, and accusations that the scientists Rowland and Molina - were politically motivated. Scientists are definitely influenced by their current culture, but those who decide to make recommendations for public action are being responsible citizens and therefore political. Acceptance of new evidence depends on Bayes’ Rule: the ability to calculate what your degree of belief or credibility should be in learning of new evidence depending on what your previous beliefs were and the likelihood of the validity of the evidence. Kuhn carried this further, that adherence to “normal science” often led to revolution. However, he was referring to the revolution with paradigms which often leads to changes in what is considered to be good and applicable evidence as well as theory.

Subsequent discussions led to the era of “science studies” in which the interactions of science, politics and culture were linked. Industrialists were prone to accuse Rowland and Molina as political agents, influenced by their background views, especially on environmentalism. Previous situations tended to confirm this since Galton’s views on Eugenics influenced his founding of modern statistical analysis.

The eruption of Laki in Iceland (in 1783) which spewed large amounts of acidic gases with disastrous
results on animal life, preceded the concept of acid rain in 1859 from industrial sources. (Note: the authors make an error in chemistry in stating that one of the gases was fluorine, not hydrogen fluoride). Those studies in Britain came home to the US when in 1974 Likens and Borman showed that acid rain was common in the Northeast US from power plants further west. In 1981, the EPA considered the evidence “unimpeachable”. Europe and Canada instituted control measures, but Keyworth, a Reagan appointee, managed to derail the definitive report on the need for control begun by the Carter administration. Fred Singer (who comes up later), along with conservatives Sietz and Jastrow, managed to get the government to delay action since more definitive data was needed. In addition, “science wars” broke out between scientists that academics in the humanities were trying to undermine science. This quote from the final paragraph of the chapter sums it up: “The real threat to science is not from the way it is influenced by its cultural context … [but] from those people who would manipulate scientific knowledge for their own interests or obscure it from the policy makers who might act on it.”

Chapter 2, “Polarization and Conformity”, discusses arguments over mercury poisoning, causation of stomach ulcers, Lyme Disease, and political polarization, as well as confirmation bias, abortion, and climate change. Begun by the use of models by Crick and Watson to determine the structure of DNA [derisively termed “Tinkertoys” by the authors], the use of models to study these issues dominate subsequent analyses throughout the book. The Bala/Goyal mathematical model, introduced in 1998, in which individuals learn about the world both by observation and listening to others, is used. The math becomes obvious when communication networks, composed of the connections or lack of connections between nodes, are diagrammed. These networks are essential to the growth of knowledge within a community. Evidence is best assessed or judged on its own merits rather on than the beliefs of those who present it.

Estimation of the size of audiences came into prominence when discussing the size of the audience for President Trump’s inauguration. These discrepancies can be attributed to “conformity bias” (based on psychology studies by S. Asch), where decisions are influenced by choices made by others (even when reviewing the same evidence) rather than by veracity. “Information cascades” are different even though they describe the spread of beliefs through a group in spite of false information.

Chapter 3, “Evangelization of Peoples,” discuses smoking and cancer, Revelle vs. Singer on climate change, and the weaponization of reputation, especially by industry propaganda. The length of time it took between the bombshell declaration by the Readers Digest of the link between smoking and cancer and subsequent action, including the Surgeon General’s report in 1964 and restriction on sales of tobacco even later, was influenced by the “tobacco strategy” used by an industry organization which offered contrary data and other propaganda denying certainty. This saga was documented by Oreskes and Conway in “Merchants of Doubt.” The use of propaganda has a long history including disputes between religious bodies, the influencing of the U.S. entry into World War I, endless information conflicts between countries, and the climate change debates.

Chapter 4 covers social networks in depth and Bala/Goyal models are used for analysis throughout. Examples include the cause of the sinking of the USS Maine in the Havana Harbor (possible “fake news” via newspapers), Clinton’s prostitution ring out of a pizza parlor (leading to a shooting), debates over the Affordable Care Act/Obamacare, and Russian hacks of the Democratic National Committee. The proliferation of modern social networks led to an explosion of “fake news.” Journalists try to avoid the social spread of false beliefs, but by presenting both sides of a debate in the course of “fairness,” often give credence to the false side. The Wikipedia “proper weighting” standard includes presentation of all sides of a discussion, but the weight/space given to each side is proportional to the veracity of the information, determined by the number of articles published in high-impact journals [note the use of journal impact, a hot item in our profession].

The chapter (and book) conclude with a controversial proposal based on books from Philip Kitcher: “reimagining” democracy. “Vulgar democracy,, the misuse of majority rule is to be avoided and policy
based on the best available evidence should not always be up for a vote.

Notes, bibliography, and an index are included. I recommended this book for all audiences.

R. E. Buntrock
Buntrock Associates
This book is one few (and there are not many) books to explore 3D printing applications across the chemical sciences in detail. The authors took a well-balanced approach and included examples on several topics relating to the chemical sciences. 3D printing is an exciting and emerging technology, gaining attention around the world for its uses in teaching, research, and industry. The impact of this technology in automotive, health, and engineering sectors are already well established; its application potential within the chemical sciences is evident as dozens of new publications appear every week. The book will be a great resource for students, teachers, and researchers in the chemical sciences. This 250-page book is organized in eight easy-to-follow chapters. Each chapter includes an introduction, an outlook, and a value-added summary of the presented work, followed by citations.

The first chapter provides a historical perspective, compiling 3D printing history concisely and contrasting available techniques in detail (Table 1.1). One helpful feature is a compilation of commonly used 3D printers in the chemical sciences (Table 1.2) with the models’ highest resolution and approximate cost; this is indeed very useful information for the reader and a strong point of this chapter. An outlook of the 3D revolution, supported with market projection data is also provided.

Chapter two describes examples of 3D printing used in micro- and macrofluidic devices to eliminate the time-consuming and labor-intensive multistep production process for engineers. 3D printing offers a specific advantage in terms of microfluid prep. This chapter also covers the use of 3D printing for the fabrication of complex fluid control devices, with potential use in replacing conventional microfabrication techniques. Table 2.2 presents a nice view of 3D technology used for micro- and macrofluidic devices.

In the third chapter, the author explains developmental applications and examples of 3D printed analytical detectors, as well as optical and electrochemical detector designs. This chapter is especially fascinating because it includes examples of 3D printed smartphone interfaces used in a biomolecule and chemical-detection essays using smartphone apps. They deliver sensitive luminescence tests even with a low-resolution smartphone camera. This chapter also includes the applications of 3D printing in electrochemical detector cells, electrodes, and electrodes arrays.

Chapter four covers the applications of 3D printing in analytical chemistry and the use of 3D printing technology in fabrication and extraction devices. It covers solid-phase extraction using 3D printed microfluid devices as sorbents and bioreactors and explains the applications of 3D printing in analytical detectors for several liquid-chromatographic techniques. The chapter could have been combined with chapter two because of its overlap in coverage.

Chapter five describes the various 3D printing technologies used in the pharmaceutical industry, from oral to transdermal drug delivery, for customized formulations using various 3D printing techniques. This chapter explores how 3D printing technology can be used to develop new and personalized medicine to support patient compliance. Examples of interesting research, such as the successful replication of conventional dosages using control-release methods and adjusting printing parameters, are included.

Chapter six delves into the biochemical applications of 3D printing in cellular analysis, rapid diagnostics pharmacokinetics, and the pharmacodynamic profiling of drugs, and how 3D printing aids the greater understanding of complex biological and biochemical systems. This technique offers extensive design freedom, customization, and economic value to the applied models. This chapter also describes the biotoxicity of 3D polymers and resins that are currently used for 3D printing. The strong measures that are suggested to minimize the biotoxicity are the highlight of the chapter. A few methods are discussed as to how to increase material biocompatibility.
Book Review

Chapter seven explains the applications of 3D printing in synthetic and physical chemistry. It is interesting to read that the 3D printed batch reactors are one of the most explored applications, which provides the flexibility to modulate the reaction outcome just by altering the reactor configuration. It significantly reduced the requirement for chemical handling and special equipment, but provided similar products during tested reactions. The author also explains the use of a 3D printed reactor for nano-electrospray ionization mass spectrometry (nano-ESI) in spectroscopic studies. The chapter also discusses in detail the thermal stability and solvent-resistance of most commonly-available 3D printing materials. Limitations of 3D printing materials can be critical, but we hope this challenge can be overcome by more stable and solvent-resistant material inventions in the future.

The last chapter of the book covers use cases of 3D printing in chemical education. 3D printing has made it simpler to create customized educational material of complex theories and to understand difficult concepts in chemistry, like molecular complexes, crystal structures, and conformation, by enhancing students’ hands-on learning experiences. The authors provide a nice collection of resources for 3D-printable chemistry models, such as the MolPrint3D. This chapter also includes software available to convert CIF and PDB files directly to STL,(1) such as the NIH 3D Print Exchange, an open-access interactive resource. Safety, an important aspect of 3D printing technology, is also covered; the authors provide literature resources for safer printing practices.

In conclusion, this book would be a great addition to any chemistry-centered makerspace. This book is a timely resource that provides a comprehensive review of the applications of 3D printing across chemical sciences. The entire book is well-illustrated with 3D models discussed in the chapter text. I would highly recommend this for students and chemists interested in experimenting with 3D printing technology inclusion in the chemical science education or research.

Neelam Bharti
Carnegie Mellon University

Cover image provided by the Royal Society of Chemistry and used with permission

Twenty-five Years Ago in Chicago

Continuing our trip down memory lane, we come to fall 1995 in Chicago. In my meeting report the Herman Skolnik Award Symposium honoring Reiner Luckenbach and Clemens Jochum is covered in detail. Sadly, Reiner had a stroke on Friday, August 18 and was taken to Northwestern University hospital. He was unable to attend the award symposium on August 22, but his award was presented to him in the hospital and Guenter Grethe gave the award address on his behalf.

We were saddened by the death of Lucille Wert the week before the meeting; her paper was still presented in the CINF program, thanks to her husband. CINF’s Lucille M. Wert Scholarship continues in her memory.

SciFinder had been launched in spring 1995 and, at the fall meeting, Brian Cannan arranged for me to have a private session using SciFinder in a suite in Chicago. I had been very involved in the development of STN Express and I really needed to give SciFinder a serious trial. What I believe was the first published review of SciFinder (my own) was appended to my meeting report. The review was published in Information World Review in October 1995. I scored the software 4 out of 5 on value for money, 5 out of 5 on ease of use, and 4 out of 5 on features. It is amusing to look back and see how few features it had compared with today’s SciFinder. No wonder it was so easy to use! I reported that “Unlimited use of SciFinder is available for a minimum subscription of approximately $60,000, serving a maximum of 20 scientists. New pricing options will enable organizations to access SciFinder for under $10,000, which will purchase a specific number of SciFinder tasks.” I will let readers draw their own conclusions about value for money. Back in 1995, though, SciFinder was truly revolutionary and I was very impressed.

Very few of the companies in my news section are still extant. MDL Information Systems, Tripos, Biosym/MSI, Oxford Molecular, Chemical Design, Molecular Applications Group, Daylight Chemical Information Systems, Synopsys Scientific Systems, Interactive Simulations, SoftShell International, and PSI International are all distant memories now. Some 1995 scientific themes are still of interest, though. Volume 7 of Reviews in Computational Chemistry, edited by Kenny B. Lipkowitz and Donald B. Boyd, and published by VCH, contained chapters addressing important “hot” areas of research, including molecular diversity, hybrid QM/MM methods, and density functional theory (DFT). The September 1995 issue of NetSci online was devoted to 3D databases. I noted that people who had no access to the World Wide Web could get a copy by email. COMP had a session on synthetic feasibility, a topic which is really hot nowadays. There was a CINF symposium on the challenges of large databases. Could we have anticipated Enamine REAL, GDB-17, and the like? We could never have anticipated that 25 years later we would be having a virtual meeting, let alone looking at virtual libraries of chemicals.

I am very much hoping that we can have “real” meetings again in 2021, but whether or not that happens, I ought to produce my next “25 years” article, featuring the spring 1996 meeting in New Orleans.

Wendy Warr
Wendy Warr & Associates
As usual, I spent the last year collecting interesting articles, videos, and podcasts about science and popular culture. However, I had no firm plans to write up another installment in this issue. I usually have less free time in late July, as I'm either preparing for Comic-Con (SDCC), actually at the con for the four to five days, or recovering from it. Then I catch up on work and focus getting CIB content from everyone else, waiting for copyediting, and assembling the issue by early August. The cancellation of SDCC this year left me with a few free days to work on some projects, like this.

Nothing highlights the intersection of science and popular culture like an article in Science on Godzilla. In honor of the 2019 release of the latest Godzilla film, two Dartmouth researchers wrote about Godzilla’s evolution, both in size and as a reflection of our own fears and anxieties about impending disasters. Starting as a response to the U.S. atomic testing, in the latest film Godzilla’s fellow titans are released to counter the threat of mankind over the environment. “We’re the infection,” the scientist declares as she launches her plan that in hindsight was not good one.

Through an Ars Technica piece from Jennifer Ouellette, I found this article in an undergraduate physics journal that answered one of those all-important, burning questions: just how long would it take for the USS Enterprise to fill up with tribbles? The authors’ answer, assuming Spock was correct in assuming that there were about 1.77 million already on board after three days, was that it would take four and a half days for the 18.4 x 10⁹ tribbles to fill the ship completely. Film critic Keith Phipps pondered other questions. What would it be like to spend a thousand years being slowly digested by the Sarlacc from Return of the Jedi? So he reached out to biologists and ecologists for their thoughts on the plausibility of the Sarlacc’s digestive system. They agreed it that it was not plausible, that anyone who was unfortunate enough to get thrown into the pit would face the quick (but still painful) death that one would expect. One could also ask how Jabba gained his knowledge of Sarlaccian physiology.

Rhett Allain at Wired goes deep into the physics of Star Wars, taking on the acceleration rate of the Millennium Falcon when it goes into hyperdrive, and then analyzing all of the Jedi jumps from the films. Staying on the hyper/warp drive theme, Chanda Prescod-Weinstein wrote about the antimatter that fuels the warp engines on Star Trek for New Scientist, and wishing that more people at conventions would ask her about that rather than when the engines themselves will be a reality.

Kevin Shanks (aka @forensictoxguy on Twitter) launched a podcast last year: Dose Makes the Poison: The Toxcast. The third episode covers one of my favorite films from last year, Knives Out, with a review of the role that toxicology plays in the story. As for how the toxicology was presented in the file, Shanks says it was “pretty dang good.”

Raychelle Burks (2020 recipient of the ACS Grady-Stack Award) gave a talk earlier this summer on science and The Avengers, where she touched on triboluminescence that occurs when a certain hammer makes contact with a certain shield. Burks also namechecks Sibrina Collins, who was cited in an earlier installment of this series for her work on incorporating vibranium into a class activity with the periodic table.

Of all the shows I’m woefully behind on, my favorite is The Expanse. Stephen Humphrey interviewed showrunner Naren Shankar for Science, where he talked about how he uses his academic background in applied physics and electrical engineering to find drama in keeping the show more grounded, like using communication delays and making spaceship turns that don’t look like WWII fighter planes. It is brief, but entertaining. Shankar also spoke with Allain at Wired about their efforts to base elements of the show in physics, including something as simple as pouring a drink in space.
I also revisited The Science Of, a website covering ways to use popular culture to communicate and teach STEM subjects. This time around, I found interesting articles outside the usual collection of major media franchises such as Marvel and Star Wars. The first is about geoengineering and climate change in the television series Snowpiercer, though much of this would apply to Bong Joon Ho’s film as well.(13) The second is about Joe Bang’s gummy bear bomb from Logan Lucky,(14) The scene where Joe explains the science to the Logan brothers is one of the funniest because it is so unexpected. The author points out the challenge of portraying science (particularly chemistry) on screen accurately, while at the same time not providing all of the materials and procedures so someone at home cannot replicate what they are seeing on screen, whether it is an explosive device, or meth.

For something more scholarly, I discovered a submitted manuscript for an article subsequently published in the Journal of Science and Popular Culture, on the depiction of scientists in the James Bond films, as the villains, “Bond girls,” and Quartermasters.(15) I had never thought about Holly in Moonraker to be the first of the Bond girls with a skillset that Bond himself does not have since she is an astronaut as well as CIA agent, or Q as a counterbalance to the mad scientist villains that Bond defeats. However, I had to stop and chuckle a bit when the author mentioned nuclear physicist/Bond girl Christmas Jones.

Not surprisingly, there has not been much covering works released this year, since most of the major films have been postponed. One of the few I did come across, in Wired, focuses on the power (or superpower) that Wonder Woman would need to smash a bullet in midair, as seen in the trailer.(16) For Slate, film critic Sam Adams interviewed Clifford Johnson about the “temporal rut” of the Hulu film Palm Springs, his interest in accurate portrayals of scientists and the scientific method over the accuracy in the science itself, and how a casting recommendation led to his own cameo in the film as a scientist.(17)

Any articles or videos on the time-bending science behind Tenet will have to wait until next year. Happy reading and watching.

Teri Vogel
UC San Diego

References (all accessed August 3, 2020)


Opportunities are available to sponsor Division of Chemical Information events, speakers, and material. Our sponsors are acknowledged on the CINF website, in the Chemical Information Bulletin, on printed meeting materials, and at any events for which we use your contribution. For more information, please review the Sponsorship Brochure at http://www.acscinf.org/PDF/CINF_Sponsorship_Brochure.pdf.

Please feel free to contact me if you would like more information about supporting the ACS Division of Chemical Information.

Tina Qin
Chair, Fundraising Committee 2020
Email: Sponsorship@acscinf.org
Tel: 617-998-1789

The ACS CINF Division is a non-profit tax-exempt organization with taxpayer ID no. 52-6054220.
2020 marks the 20th anniversary of *Science of Synthesis* (SOS), the online synthetic methodology review compendium used by chemists worldwide. SOS, the successor to the well-known *Houben-Weyl* series was established in 2000 by an esteemed Editorial Board of international chemists, including Nobel Prize winner Ryoji Noyori. Today, under the guidance of Editor-in-Chief, Alois Fürstner, a team of eminent editors commission quality content from expert authors and ensure the selection of useful and practical methods. Chemists therefore have quick access to thorough and quality overviews on the entire range of organic synthesis topics, saving them hours of searching and literature research. SOS is considered the place to begin when writing a thesis, preparing a talk, writing a paper, starting out in a new area of chemistry, or preparing consultancy work.

The SOS Editorial Board comprises chemists from academia and industry: E. M. Carreira (ETH Zurich), M. Faul (Amgen Inc.), A. Fürstner (Editor-in-Chief, MPI/Mülheim), S. Kobayashi (University of Tokyo), G. Koch (Topadur Pharma AG), G. A. Molander (University of Pennsylvania), C. Nevado (University of Zurich), B. M. Trost (Stanford University), and S.-L. You (Shanghai Institute of Organic Chemistry). The overarching goal of the Editorial Board is to make the suite of SOS resources the first and foremost focal point for critically evaluated information on chemical transformations for anyone involved in the design and synthesis of organic molecules. Over 100 volume editors and 2,500 authors worldwide have contributed to the series over the last 20 years.

The SOS Advisory Board comprises experts who have significant experience of chemical information systems in both industry and academia: G. Baysinger (Stanford University), L. Betschart (ETH Zurich), J. Currano (University of Pennsylvania), J. Goodman (University of Cambridge, UK), C. Keil (Pfizer), Ye Li (MIT), Xiaoxia Li (CAS, China), Y. Sevryugina (University of Michigan, USA), and D. Wrublewski (Caltech). They regularly contribute to discussions regarding the development of the electronic product and chemical information in general. We would like to say sincere thanks for their contribution over the years.

Over 20 years, some 96,000 "pages" of evaluated information have been published, including over 2 million molecules, 425,000 selected reactions, and no less than 54,000 experimental procedures. SOS has proven to be the perfect research companion for the synthetic organic chemist and will continue to provide the chemistry community with valuable synthetic methodology reviews in the years to come! Anyone interested in having free trial access to the online product ([sos.thieme.com](https://sos.thieme.com)) at their institution should contact the publisher at esales@thieme.com or sign up for a free 14-day trial at [www.thieme.de/sos-trial](https://www.thieme.de/sos-trial).

**SoS as a Teaching Resource**
Find links to *Science of Synthesis* chapters that will be useful as a resource for the preparation and teaching of advanced organic chemistry courses.
[https://www.thieme.de/en/thieme-chemistry/teaching-resources-156692.htm](https://www.thieme.de/en/thieme-chemistry/teaching-resources-156692.htm)

**Open access SOS articles** and information about the 20th anniversary:

**Videos from chemists** about the SOS project: [https://www.youtube.com/watch?v=CRHlF5peYdI&list=PLTvWnM-9-1HyI1ubWzLhc8YHy18Srozx](https://www.youtube.com/watch?v=CRHlF5peYdI&list=PLTvWnM-9-1HyI1ubWzLhc8YHy18Srozx)
Call for nominations - 2nd Dr. Margaret Faul Award for Women in Chemistry 2021, sponsored by SoS

In addition to providing an important editorial contribution for the chemistry community, the dynamic and forward-thinking SOS Editorial Board founded the first major international synthetic chemistry award for Women in Chemistry. As a result, the first award was presented to Sarah Reisman (Caltech) at the 21st ESOC in Vienna in 2019. The second award will be presented at the ESOC in Ghent, Belgium in 2021 and nominations can be submitted by e-mail to Dr. Marcus White (marcus.white@thieme.de) until December 11th, 2020.


Select Crowd Review Shortlisted for the ALPSP Awards for Innovation in Publishing 2020

Select Crowd Review uses the mechanisms of social media communication to make the review process much faster than classical peer review, but still with the same or even better quality. https://www.thieme.de/en/thieme-chemistry/select-crowd-review-136859.htm

This innovative technology, that is available for SYNLETT and SynOpen authors, has been shortlisted for the ALPSP Awards for Innovation in Publishing 2020: https://www.alpsp.org/News/alpsp-awards-for-innovation-shortlist-jun-2020

Free Access to Covid-19 Relevant Active Pharmaceutical Ingredients in Pharmaceutical Substances

Earlier this year, Thieme Chemistry released free PDFs for download covering synthetic routes to 15 active pharmaceutical ingredients that have been investigated for repurposing as Covid-19 treatments. https://www.thieme.de/de/thieme-chemistry/15-apis-of-interest-covid-19-treatment-156090.htm

Contact
Science of Synthesis/Thieme Chemistry
Thieme Group
Ruedigerstrasse 14
70469 Stuttgart
Germany

Editorial/Technical Queries:
SOS_techsupport@thieme.com

Access/Marketing:
marketing@thieme-chemistry.com

Website: sos.thieme.com
2020 CINF Officers and Functionaries

Chair
Jeremy Garritano
University of Virginia
jg9jh@virginia.edu

Chair-Elect
Donna Wrublewski
Caltech Library
dtwrub@caltech.edu

Past-Chair
Elsa Alvaro
Northwestern University
elsa.alvaro@northwestern.edu

Secretary
Nick Ruhs
Florida State University
nruhs@fsu.edu

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

CINF Councilors
Bonnie Lawlor
chescot@aol.com

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

CINF Alternate Councilors
Rachelle Bienstock
RJB Computational Modeling LLC
rachelleb1@gmail.com

Charles Huber
University of California, Santa Barbara
huber@library.ucsb.edu

Archivist/Historian
Bonnie Lawlor
chescot@aol.com

Awards Committee Chair
Rajarshi Guha
Vertex Pharmaceuticals
rajarshi.guha@gmail.com

Careers Committee Chair
Neelam Bharti
Carnegie Mellon University Libraries
nbharti@andrew.cmu.edu

Communications and Publications Committee Chair
Graham Douglas (Interim)
communications@acscinf.org

Procedures Chair
Bonnie Lawlor
See Councilor

Education Committee Chair
Grace Baysinger
Stanford University
graceb@stanford.edu

Finance Committee Chair
Stuart Chalk
University of North Florida
schalk@unf.edu

Fundraising Committee Chair
Tina Qin
Harvard University
qinnamsu@gmail.com

Membership Committee Chair
Donna Wrublewski (interim)
Caltech Library
dtwrub@caltech.edu
Nominating Committee Chair
Elsa Alvaro
Northwestern University
elsa.alvaro@northwestern.edu

2019–2020 Program Committee Chair
Sue Cardinal
University of Rochester
scardinal@library.rochester.edu

2020–2021 Program Committee Chair
Ye Li
MIT
yel@mit.edu

Chemical Information Bulletin Editor, Spring
TBD

Chemical Information Bulletin Editor, Summer
David Shobe
Patent Information Agent
avidshobe@yahoo.com

Chemical Information Bulletin Editor, Fall
Teri Vogel
UC San Diego
tmvogel@ucsd.edu

Chemical Information Bulletin Editor, Winter
Judith Currano
University of Pennsylvania
currano@pobox.upenn.edu

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

Stuart Chalk
University of North Florida
schalk@unf.edu

Articles and Features
Svetlana Korolev
Bonnie Lawlor
Teri M. Vogel
Wendy Warr
Donna Wrublewski

Committee Reports
Bonnie Lawlor

Book Reviews
Neelam Bharti
Robert E. Buntrock

Sponsor Information
Tina Qin

Production
Bonnie Lawlor
David Shobe
Teri M. Vogel
# Technical Program Broadcast Schedule

## 2020 Fall Meeting (all times Pacific)

### MONDAY MORNING

#### Machine Learning & Artificial Intelligence in Computational Chemistry

*T. Robertson, Organizer*

<table>
<thead>
<tr>
<th>Time</th>
<th>Session No.</th>
<th>Title</th>
<th>Presenter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00</td>
<td>CINF 22</td>
<td>Quantum descriptors real-time predictions and applications to reactivity predictions.</td>
<td><strong>Y. Guan</strong></td>
</tr>
<tr>
<td>10:20</td>
<td>CINF 23</td>
<td>Synthetic feasibility and de novo molecular generation and optimization.</td>
<td><strong>C. Coley</strong></td>
</tr>
<tr>
<td>10:40</td>
<td>CINF 24</td>
<td>Big errors in big data: when automated data curation misses the mark.</td>
<td><strong>R. Clark</strong></td>
</tr>
<tr>
<td>11:00</td>
<td>CINF 25</td>
<td>Conformal regression for Profile-QSAR: Adding confidence intervals to point predictions.</td>
<td><strong>X. Zhu</strong></td>
</tr>
<tr>
<td>11:20</td>
<td>CINF 29</td>
<td>Discovering novel fluorescent molecules via global optimization of molecular properties.</td>
<td><strong>J. Lee</strong></td>
</tr>
<tr>
<td>11:40</td>
<td>CINF 30</td>
<td>Predicting diffusion in Lennard Jones fluids and beyond using machine learning.</td>
<td><strong>J. Allers</strong></td>
</tr>
</tbody>
</table>

Q&A will take place at 2:20

### MONDAY AFTERNOON

#### Machine Learning & Artificial Intelligence in Computational Chemistry

*T. Robertson, Organizer*

<table>
<thead>
<tr>
<th>Time</th>
<th>Session No.</th>
<th>Title</th>
<th>Presenter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:00</td>
<td>CINF 31</td>
<td>Heuristic Global Optimization in Chemical Compound Space.</td>
<td><strong>C. Rinderspacher</strong></td>
</tr>
<tr>
<td>1:20</td>
<td>CINF 32</td>
<td>Clustering and predicting properties of general polymers based on their monomer unit structure.</td>
<td><strong>H. Yamano</strong></td>
</tr>
<tr>
<td>1:40</td>
<td>CINF 33</td>
<td>Performance and scope of similarity-based and machine learning approaches for predicting the macromolecular targets of small molecules.</td>
<td><strong>N. Mattai</strong></td>
</tr>
<tr>
<td>2:00</td>
<td>CINF 35</td>
<td>Deep learning showed increased phenotypic assay sensitivity in quantifying inhibitors of α-synuclein inclusion formation.</td>
<td><strong>L. Akella</strong></td>
</tr>
</tbody>
</table>

2:20 (40 minutes) Q&A with all AM and PM session speakers.
# TUESDAY MORNING

## Making Chemistry FAIRer

*I. Bruno, S. Chalk, L. McEwen, N. Ruhs, Organizers*

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Title</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00</td>
<td>CINF 59</td>
<td>Advancing FAIR: Moving toward the Internet of FAIR data and services.</td>
<td>M. Cragin</td>
</tr>
<tr>
<td>10:15</td>
<td>CINF 60</td>
<td>FAIR IUPAC: Advancing pure and applied chemistry worldwide.</td>
<td>L. McEwan</td>
</tr>
<tr>
<td>10:45</td>
<td>CINF 62</td>
<td>GO-FAIR chemistry implementation network (ChIN).</td>
<td>S. Chalk</td>
</tr>
<tr>
<td>11:00</td>
<td></td>
<td>(60 minutes) Q&amp;A with all AM session speakers</td>
<td></td>
</tr>
</tbody>
</table>

# TUESDAY AFTERNOON

## From Bench to Market: Leveraging AI & Advanced Computational Methods to Solve Hard Problems

*A. Handzel & H. Jooya, Organizers*

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Title</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:00</td>
<td>CINF 37</td>
<td>From High-Throughput Computational Chemistry and Molecular Pattern Recognition to the Targeted Design of Novel Chemistry.</td>
<td>J. Hachmann</td>
</tr>
<tr>
<td>1:20</td>
<td>CINF 38</td>
<td>Machine-learning correlation energies using hybrid structural-energetic descriptors.</td>
<td>R. DeStasio</td>
</tr>
<tr>
<td>1:40</td>
<td>CINF 39</td>
<td>Machine learning augmented quantum chemistry: From deep learning of wave functions to ML/QM-tandem methods.</td>
<td>R. Maurer</td>
</tr>
<tr>
<td>2:00</td>
<td>CINF 41</td>
<td>Accurate Gaussian process models for high-dimensional molecular systems.</td>
<td>R. Krems</td>
</tr>
<tr>
<td>2:20</td>
<td>CINF 40</td>
<td>BioHarmony across semantic drug discovery data streams.</td>
<td>B. Bunin</td>
</tr>
<tr>
<td>2:40</td>
<td></td>
<td>(20 minutes) Q&amp;A with all PM session speakers</td>
<td></td>
</tr>
</tbody>
</table>
Technical Program Broadcast Schedule
2020 Fall Meeting (all times Pacific)

WEDNESDAY MORNING

Reaction Prediction & Synthesis Planning

G. Blanke, J. Goodman, G. Grethe & C. Keil, Organizers

10:00 CINF 9  Predicting reaction sequences: Deep neural networks and reaxys data bases. F. Kroll
10:20 CINF 6  Chemistry puppeteer: Enhancing the diversity of retrosynthetic predictions. A. Toniato
10:40 CINF 12 Data preparation for reaction prediction: the quest for "AI-ready" data. F. van den Broek
11:00 CINF 13 Reaction Transformers for Fingerprints, Classification and Atom-Mapping. P. Schwaller
11:20 CINF 15 Evolutionary computing strategies and feedback control for directed execution and optimisation of chemical reactions. H. Makatsoris
11:40 CINF 4 SAVI a la carte: Moving toward molecules on demand by AI. The development of the SLICE (Smarts and Logic In ChEmistry) language. V. Delannee

Q&A will take place at 2:40.

WEDNESDAY AFTERNOON

Reaction Prediction & Synthesis Planning

G. Blanke, J. Goodman, G. Grethe & C. Keil, Organizers

1:00 CINF 5 Retrosynthetic software for practicing chemists: Novel and efficient in silico pathway design validated at the bench. L. Rickenshauser
1:20 CINF 7 Molecule-Edit Graph Attention Network: Modeling retrosynthesis prediction as a sequence of edits. M. Sacha
1:40 CINF 8 Computer-aided synthesis planning & ASKCOS. C. Coley
2:00 CINF 11 Overcoming conflicts and dilemmas in computer-aided synthesis design. O. Ravitz
2:20 (40 minutes) Q&A with all AM and PM session speakers.
Technical Program Broadcast Schedule
2020 Fall Meeting (all times Pacific)

THURSDAY MORNING

Moving Chemistry from the Lab into the Open

*M. Hicks, C. Huber, & W. Patterson, Organizers*

10:00  CINF 18  Publishing crystal structure data – keeping up with the times. *I. Bruno*
10:40  CINF 20  Integration of chemistry with everything else. *I. Wetherbee*
11:00  CINF 42  Enhancing data validation and summarization in PubChem via automated link analysis. *L. Zaslavsky*

11:20  (40 minutes) Q&A with all AM session speakers

THURSDAY AFTERNOON

Novel Cheminformatic Methods, Tools & Models for Predictive Toxicology & Read-Across

*A. Williams, B. Zdraził, Organizers*

1:00  CINF 1  New CSRML-based features to categorize and fingerprint PFAS structure lists for cheminformatics analysis and read-across. *R. Lougee*
1:20  CINF 2  New, extensible set of molecular substructures for interpretable QSARs and read-across applications. *V. Gombar*
1:40  CINF 3  Data-enrichment approach to prioritize chemical structures associated with binding to the thyrotropin-releasing hormone receptor (TRHR). *M. Shobair*
2:00  CINF 50  Machine learning for Scent: Learning generalizable perceptual representations of small molecules. *A. Wiltschko*

2:20  (40 minutes) Q&A with all PM session speakers.
Technical Program On-Demand Schedule
2020 Fall Meeting (all times Pacific)

**MONDAY AFTERNOON -- THURSDAY AFTERNOON**
** indicates talk is prerecorded only and not part of the broadcast schedule

CINF 1  New CSRML-based features to categorize and fingerprint PFAS structure lists for cheminformatics analysis and read-across. R. Lougee
CINF 2  New, extensible set of molecular substructures for interpretable QSARs and read-across applications. V. Gombar
CINF 3  Data-enrichment approach to prioritize chemical structures associated with binding to the thyrotropin-releasing hormone receptor (TRHR). M. Shobair
CINF 4  SAVI a la carte: Moving toward molecules on demand by AI. The development of the SLICE (Smarts and Logic In ChEmistry) language. V. Delannee
CINF 5  Retrosynthetic software for practicing chemists: Novel and efficient in silico pathway design validated at the bench. L. Rickenshauser
CINF 6  Chemistry puppeteer: Enhancing the diversity of retrosynthetic predictions. A. Toniato
CINF 7  Molecule-Edit Graph Attention Network: Modeling retrosynthesis prediction as a sequence of edits. M. Sacha
CINF 8  Computer-aided synthesis planning & ASKCOS. C. Coley
CINF 9  Predicting reaction sequences: Deep neural networks and reaxys data bases. F. Kroll
CINF 10** Combining artificial intelligence with structured high quality data in chemistry – delivering outstanding predictive chemistry applications. J. Swenty-Busch
CINF 11  Overcoming conflicts and dilemmas in computer-aided synthesis design. O. Ravitz
CINF 12  Data preparation for reaction prediction: the quest for "AI-ready" data. F. van den Broek
CINF 13  Reaction Transformers for Fingerprints, Classification and Atom-Mapping. P. Schwaller
CINF 14** Comprehensive search for compounds and chemical reactions in big query. S. Boyer, L. Weber
CINF 15  Evolutionary computing strategies and feedback control for directed execution and optimisation of chemical reactions. H. Makatsoris
CINF 16** Use of conditional generative adversarial networks (cGAN) with molecular orbital fields to study chlorination sites. M. Clark
CINF 17** Expanding the synthetically feasible chemical space of amides to a billion of compounds. O. Isayev
CINF 18  Publishing crystal structure data – keeping up with the times. I. Bruno
CINF 19  Is Web Assembly the future of visualizing molecules in the browser? J. Fine.
CINF 20  Integration of chemistry with everything else. I. Wetherbee
CINF 22  Quantum descriptors real-time predictions and applications to reactivity predictions. Y. Guan
CINF 23  Synthetic feasibility and de novo molecular generation and optimization. C. Coley
CINF 24  Big errors in big data: when automated data curation misses the mark. R. Clark
CINF 25  Conformal regression for Profile-QSAR: Adding confidence intervals to point predictions. X. Zhu
CINF 27** Guiding Anti-Malarial Optimisation Using Deep Learning Imputation and Compound Generation. M.D. Segall
Technical Program On-Demand Schedule
2020 Fall Meeting (all times Pacific)

MONDAY AFTERNOON -- THURSDAY AFTERNOON
** indicates talk is prerecorded only and not part of the broadcast schedule

CINF 26** ~ Revealing hidden determinants of molecular recognition in hit-to-lead optimization with graph convolutional attention mechanisms. **H. van den Bodem**

CINF 28** ~ Board Game AI in Continuous Action Spaces: A Monte Carlo Tree Search Algorithm for Material Application. **T. Loeffler**

CINF 29 ~ Discovering novel fluorescent molecules via global optimization of molecular properties. **J. Lee**

CINF 30 ~ Predicting diffusion in Lennard Jones fluids and beyond using machine learning. **J. Allers**

CINF 31 ~ Heuristic Global Optimization in Chemical Compound Space. **C. Rinderspacher**

CINF 32 ~ Clustering and predicting properties of general polymers based on their monomer unit structure. **H. Yamano**

CINF 33 ~ Performance and scope of similarity-based and machine learning approaches for predicting the macromolecular targets of small molecules. **N. Mattai**

CINF 35 ~ Deep learning showed increased phenotypic assay sensitivity in quantifying inhibitors of α-synuclein inclusion formation. **L. Akella**

CINF 36** ~ ChEMU shared task: chemical entity recognition and event extraction of chemical reactions from patents. **C. Thorne**

CINF 37 ~ From High-Throughput Computational Chemistry and Molecular Pattern Recognition to the Targeted Design of Novel Chemistry. **J. Hachmann**

CINF 38 ~ Machine-learning correlation energies using hybrid structural-energetic descriptors. **R. DeStasio**

CINF 39 ~ Machine learning augmented quantum chemistry: From deep learning of wave functions to ML/QM-tandem methods. **R. Maurer**

CINF 40 ~ BioHarmony across semantic drug discovery data streams. **B. Bunin**

CINF 41 ~ Accurate Gaussian process models for high-dimensional molecular systems. **R. Krems**

CINF 42 ~ Enhancing data validation and summarization in PubChem via automated link analysis. **L. Zaslavsky**

CINF 43** ~ Development of a unified platform for registration of all entities from small molecules to biologics and conjugates using HELM. **J. Buttrick**

CINF 44** ~ Template based modeling of protein assemblies in the ClusPro server. **D. Padhorny, D. Kozakov**

CINF 45** ~ Adapting evolutionary algorithms for autonomous machine learning in chemistry. **G. Vishwakarma**

CINF 46** ~ Insights into therapeutic fusion protein R&D from an analysis of the CAS databases. **Y. Li**

CINF 48** ~ Reproducible Data Analysis and Publishing in Chemistry with R: Creating a workshop experience during the ACS National Meeting. **Y. Li, D.T. Wrublewski**

CINF 49** ~ Chemists’ Data Needs for Machine Learning Research in Academia. **Y. Li**

CINF 50 ~ Machine learning for Scent: Learning generalizable perceptual representations of small molecules. **A. Wiltschko**
Technical Program On-Demand Schedule
2020 Fall Meeting (all times Pacific)

MONDAY AFTERNOON -- THURSDAY AFTERNOON
** indicates talk is prerecorded only and not part of the broadcast schedule

CINF 51**  
R&D digitalization and data intelligence strategies for flavors and fragrances applications. **M. Petersen**

CINF 52**  
Communicating computation: an automated, reproducible workflow for computational organic chemistry studies and their analysis. **J.V.A. Requena**

CINF 53**  
Overcoming complications in informatics for materials and devices. **P. Son**

CINF 54**  
Exploring steric activity and molecular descriptors for building statistical models. **G. Luchini**

CINF 56**  
Molecular modeling expeditions in the chemical space. **S. Forli**

CINF 57**  

CINF 58**  
Quality matters in chemotype discovery. **J.M. Jansen**

CINF 59  
Advancing FAIR: Moving toward the Internet of FAIR data and services. **M. Cragin**

CINF 60  
FAIR IUPAC: Advancing pure and applied chemistry worldwide. **L. McEwan**

CINF 61  
PubChem COVID-19 response: Opportunities and challenges in delivering timely research data. **E. Bolton**

CINF 62  
GO-FAIR chemistry implementation network (ChIN). **S. Chalk**

CINF 63**  
Open conversation about FAIR in Chemistry. **I. Bruno, S. Chalk, L. McEwan, N. Ruhs**
CINF 1

New CSRML-based features to categorize and fingerprint PFAS structure lists for cheminformatics analysis and read-across

Ryan Lougee1, Lougee.Ryan@gmail.com, Grace Patlewicz4, Ann Richard2, Christopher Grulke3, Antony J. Williams4, Chihae Yang5, James F. Rathman6, Tomasz Magdziarz5. (1) EPA / ORISE, Alton Bay, New Hampshire, United States (2) MD 205-01, US EPA, Research Triangle Park, North Carolina, United States (3) National Center of Computational Toxicology, US EPA, New Hill, North Carolina, United States (4) National Center for Computational Toxicology, Environmental Protection Agency, Wake Forest, North Carolina, United States (5) Molecular Networks, GmbH, Erlangen, Germany (6) Ohio State University, Columbus, Ohio, United States

Per- and polyfluoroalkyl substances (PFAS) are of high public interest due to widespread production, environmental persistence, and adverse ecological and health impacts. Recently, both the Organisation for Economic Co-operation and Development (OECD) and the United States Environmental Protection Agency’s Comptox Chemicals Dashboard have published extensive lists of over 4500 and 6000 curated PFAS structures. Whereas most studies to-date have focused on the health effects of a small number of PFAS compounds, such as PFOA and PFOS, relatively little is known about the health effects of the vast majority of PFAS and their byproducts. Methods for profiling the PFAS chemical structure space are needed to support modeling and structure-based categorization efforts. However, the unique PFAS structural space is ill-suited for publicly available molecular fingerprinting methods. Furthermore, expert-defined PFAS chemical category terms are limited to simpler categories (e.g., perfluorocarboxylic acids) and often lack clear structure definition. With the publicly available CSRML (Chemical Subgraphs and Reactions Markup Language), we developed more than 130 fingerprints to capture unique aspects of PFAS structures, including perfluoro chains, polyfluoro substructures, fluorinated rings, and various perfluoro branching patterns. These CSRML PFAS categories and features can be used with the public Chemotyper, provide comprehensive coverage of available PFAS lists, and are being used to profile and describe PFAS chemical lists currently undergoing testing within EPA.

CINF 2

New, extensible set of molecular substructures for interpretable QSARs and read-across applications

Vijay K. Gombar1, vijay.gombar@sciome.com, Ruchir R. Shah1, Nicole Kleinstreuer2, Scott S. Auerbach2, Alex Merrick2, Alexander Sedykh1. (1) Sciome LLC, Research Triangle Park, North Carolina, United States (2) National Toxicology Program (NTP), National Institute of Environmental Health Sciences (NIEHS), Research Triangle Park, North Carolina, United States

Predictions from QSAR models and read-across can gain wider acceptance if these are interpretable in terms of chemistry-based reasoning. Focusing on prediction interpretability, we present here a newly compiled set of ~1000 molecular substructures, Saagar. For chemistry-based read-across applications, we compared Saagar with publicly available MACCS, Pub-
Chem, ECFP4, and ToxPrint fingerprints for early active enrichment from ~150,000 compounds spanning 119 molecular targets. For the top 1%, 2%, 5%, and 10% false discovery rates, Saagar returned a higher enrichment factor for active probes in 18 of 20 comparisons, thus, testifying that Saagar features support efficient retrieval of bioactive chemicals from diverse datasets.

We also evaluated the utility of Saagar features in developing interpretable quantitative QSAR models using recursive partitioning (RP) for all (n=41) Tox21 assays at 10uM and 100uM thresholds. The ROC AUC in 5-fold cross-validation was between 0.58 – 0.94 (average of 0.72) for the 10uM threshold, and between 0.55 – 0.91 (average of 0.70) for the 100uM threshold models. The average absolute difference in ROC AUCs between Saagar-RP models and those developed with the well-known, but less interpretable, Mordred descriptors is 0.02, suggesting comparable performance of Saagar substructures and Mordred descriptors.

While delivering comparable performance, Saagar-RP models provide the added advantage of relating structural chemical motifs directly to the predicted assay outcome via a decision-tree format, providing clear paths to interpretability. For the aryl hydrocarbon receptor (AhR) activation model the top split with Saagar feature SGR10001 (>8 ring atoms) explains endogenous L-Kynurenine and indole derivatives as agonists of AhR. Similarly, the Saagar-RP model of the PPARδ agonist assay points to the binding of large molecules with a sp\(^3\)C-COOH group (like fatty acids), which forms the basis of lipid homeostasis. In conclusion, Saagar features, in combination with recursive partitioning techniques, provide transparency and interpretability to chemical profiling by read-across and QSAR predictions.

CINF 3

Data-enrichment approach to prioritize chemical structures associated with binding to the thyrotropin-releasing hormone receptor (TRHR)

Mahmoud Shobair, shobair.mahmoud@epa.gov, Daniel T. Chang, Christopher Gruulke, Katie Paul-Friedman, Ann Richard. Center for Computational Toxicology and Exposure, United States Environmental Protection Agency, Research Triangle Park, North Carolina, United States

Application of high-throughput screening (HTS) to toxicology, as in the cross-federal Tox21 partnership or EPA’s ToxCast program, faces the challenge of relating HTS data to molecular initiating events (MIEs) and adverse outcome pathways (AOPs) when the relationship between the output and MIE is indirect. Thyrotropin-Release Hormone Receptor (TRHR) activation is a MIE in the AOP network for thyroid hormone (TH) disruption, but the ability of environmental chemicals to perturb TRHR signaling is unknown. A Tox21 HTS biochemical assay is available, but no orthogonal or confirmatory Tox21 or ToxCast assays can be used to differentiate TRHR true positive responses. To test the hypothesis that environmentally-relevant chemicals can directly interact with TRHR, we developed an in silico cheminformatic workflow to identify chemicals predicted to interfere with TRH binding. The tiered approach has the following steps: 1) analyze assay hits using structure-based chemotype enrichment analysis; 2) eliminate noise from cytotoxicity or assay interference; 3) identify other sources of interference such as perturbation in calcium signaling; and 4) prioritize potential TRHR modulators by likelihood of binding. For in silico binding model training, we curated a reference structure-activity dataset from literature reports of chemical concentrations required to displace binding of radiolabeled TRH in competitive binding assays. The dataset is balanced between binders and non-binders and mainly contains TRH derivatives and psychoactive drugs. We used 3D pharmacophore modeling to discriminate between binders and non-binders. Candidate binders were visualized in a 3D homology model. Preliminary results suggest that < 5% of the actives in the Tox21 biochemical assay contain TRH-like binding features. Chemicals predicted
to bind likely conserve the TRHR ligand binding site and exhibit structural similarity to known
TRHR modulators, like benzodiazepines and neuropeptides. Our tiered in silico workflow in-
creases the value of in vitro data for chemical prioritization by grounding the results in physi-
cal determinants that are directly related to experimental target binding results. Our findings
suggest that combining structure-based methods and data enrichment analysis can be ap-
plied more broadly to increase confidence in HTS results and inform further screening and
hazard characterization.

CINF 4

SAVI a la carte: Moving toward molecules on demand by AI. The development of the
SLICE (Smarts and Logic In ChEmistry) language

Victorien Delannee¹, victorien.delannee@nih.gov, Marc C. Nicklaus². (1) National Cancer
Institute, Frederick, Maryland, United States (2) NCI-Frederick Bldg 376 RM 207, Natl Inst
Health NCI Ft Detrick, Frederick, Maryland, United States

The current version of SAVI (Synthetically Accessible Virtual Inventory) is an expert system
for the generation of very large libraries of easily synthesizable molecules, utilizing the LHA-
SA project's language pair CHMTRN/PATRAN for synthetic knowledge rules. One unique
strength of CHMTRN is the use of a logic validated by chemists to predict both possible fail-
ure and success of reactions. However, CHMTRN is an old unstructured and non-
standardized language working retrosynthetically, which presents challenges in a forward-
synthetic context. To overcome these limitations, we have developed SLICE (Smarts and
Logic In ChEmistry), which combines SMARTS with a logic language. SMARTS describes the
molecular patterns, while the logic allows reasoning by defining rules such as "IF statement
THEN action." This new language is highly inter-operable, suitable in both a forward and retro-
synthetic context, easily readable and extended SMARTS capabilities. We show that this lan-
guage can be used in the context of SAVI to speed up the first-step reaction prediction by
making it forward-synthetic and can also be used in a retro-synthetic context to retrieve the
synthetic road predicted by AI.

CINF 5

Retrosynthetic software for practicing chemists: Novel and efficient in
silico pathway design validated at the bench

Lindsey Rickershauser, lindey.rickershauser@milliporesigma.com. Cheminformatics
Technologies,
MilliporeSigma, Burlington, Massachusetts, United States

In a continuously evolving landscape of in silico chemical intelligence and machine learning,
computer assisted synthetic planning has come to the forefront of discussion in the
cheminformatics space. Herein, we describe the use of SYNTHIA™, a retrosynthetic design
software in drug discovery, industrial, and academic laboratories all over the world. As a prod-
uct of over 15 years of research, this unique tool is poised to not only get better with time, but
also revolutionize the way chemists approach designing pathways to their complex tar-
gets. SYNTHIA™’s unique approach to building our expert database of known reactions by
hand coding each transformation has allowed this tool to become a bench chemist’s ally by
‘learning’ chemistry much like a chemist would themselves, and suggesting diverse pathways
towards their targets, thus generating ideas and providing cost effective routes based on each
user’s unique needs. As a product of over 15 years of research, this unique tool is poised to not only get better with time, but also revolutionize the way chemists approach designing pathways to their complex targets

CINF 6

Chemistry puppeteer: Enhancing the diversity of retrosynthetic predictions

Alessandra Toniato¹, ato@zurich.ibm.com, Philippe Schwaller¹, Antonio Cardinale¹, Teodoro Laino². (1) IBM Research Zurich, Zurich, Switzerland (2) IBM Research GmbH, Rueschlikon, Switzerland

One of the main issues of AI-based retrosynthesis planning algorithms is that, usually, the proposed disconnection strategies lack in diversity. When the goal is to find a suitable set of precursors for a given target molecule, the generated precursors typically fall in the same chemical macro class (ex. protection, deprotection or same C-C bond formation with a slightly different set of reagents) and the automatic synthesis planning tools might get stuck. Most of the previous approaches do not allow a machine learning model to have a broader exploration and are focused on the top single-step predictions, which can be detrimental for the multi-step strategy. To enhance diversity in our approach, we introduced tokens of macro classes in the training inputs. The learned embeddings of the given sample partly codify some characteristics of the reactions belonging to that class. At test time, the macro classes allow us to stir the model towards different kinds of disconnection strategies. In this work, we show with results on a set of patent data that the diversity of the predictions can improve consistently.

While the use of excessively specific groupings can decrease the model performances in terms of valid proposed set of precursors, the use of chemically relevant policies to construct smaller macro groups allows to recover the quality of the predictions without the loss of the found diversity.

CINF 7

Molecule-Edit Graph Attention Network: Modeling retrosynthesis prediction as a sequence of edits

Mikolaj Sacha¹, mikolajsacha@gmail.com, Mikolaj Blaz¹, Piotr Byrski¹, piotr@molecule.one, Pawel Wlodarczyk-Prusynski¹, Stanislaw Jastrzebski¹,², staszek.jastrzebski@gmail.com. (1) Molecule.one, Warsaw, Poland (2) New York University, New York, New York, United States

Retrosynthesis - the task of finding a set of reactants that can be used to synthesize a target molecule, is one of the fundamental problems in organic chemistry. Recently, there have been substantial developments in the field of computer-aided retrosynthesis. Most of such approaches employ reaction templates, which are able to give highly accurate and interpretable predictions but suffer from low generalization due to the limited coverage. On the other hand, template-free methods that view retrosynthesis simply as a translation problem lack interpretability and are prone to making trivial mistakes.

In this work, we present the Molecule-Edit Graph Attention Network (MEGAN) - a template-free model that encodes reactions as series of operations on molecular graphs. Our model is highly scalable and does not require any manual rule encoding. Moreover, the sequence of edits generated by MEGAN provides mapping of atoms between product and substrates.
MEGAN surpasses all published template-free and template-based models on a popular USPTO-50k retrosynthesis benchmark in terms of top-10 accuracy while achieving above 95% coverage of the held-out validation reaction dataset.

CINF 8

Computer-aided synthesis planning & ASKCOS

Connor W. Coley, ccoley@mit.edu. Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

Machine learning and artificial intelligence have enabled new data-driven approaches to CASP where statistical models are trained directly on published experimental data. A group of researchers at MIT developed several of these tools in an integrated software suite, ASKCOS, that is capable of proposing retrosynthetic routes to new molecules, proposing reaction conditions for each step, and assessing the likelihood of experimental success. This talk will summarize the development of ASKCOS, algorithmic contributions, user adoption, user experiences, and outstanding challenges. I will emphasize a new template-free, graph-based approach to proposing retrosynthetic disconnections.

CINF 9

Predicting reaction sequences: Deep neural networks and reaxys data bases

Friedrich Kroll, f.kroll@elsevier.com. Elsevier, Frankfurt, Deutschland (DEU), Germany

Effective synthetic organic chemistry is based on several principal skills: i) envisioning all feasible synthetic steps and selecting the most suited ii) finding applicable procedures to run the experiments iii) mastering instruments and equipment iv) analysing and drawing the right conclusions from reaction outcomes. The complexity of such tasks explains why it takes several years to educate and enable chemists in this discipline successfully. Even then, non-yielding synthesis steps are part of the reasons why the amount of time to be invested is considerable higher than often expected - particularly in cases of complex molecules.

Reaxys is a software tool that contains well curated FAIR (Findable, Accessible, Interoperable, Reusable) scientific information and teaches chemists how to find and apply the right literature for chemical transformations, identify analytical information and most importantly educates, how to develop retrosynthesis schemes swiftly. The possibilities to evaluate and compare different routes, yields and conditions to molecules are exceptional and allow educators to illustrate retrosynthesis in organic chemistry classes on a day to day basis. Furthermore, computer-aided retrosynthesis tools, which can predict reactions to novel molecules correctly, are part of Reaxys. The ‘deep learning’, neuronal networks produce schemes of sequences of reactions to the desired compounds. This novel AI tool has processed nearly every reaction ever published (> 15 million) and has the potential to transform the way organic chemists work in the future. In an assessment, various synthesis routes generated were tested in a double-blinded trial with 45 organic chemists from two institutes in China and Germany and the routes have proven to be scientifically sound and robust. This unique approach to retrosynthesis will lead to an increase of the success rate in synthetic organic chemistry and should have an enormous benefit in terms of discovering sustainable chemical solutions and
minimizing expenditure.

CINF 10

Combining artificial intelligence with structured high quality data in chemistry – delivering outstanding predictive chemistry applications

Juergen Swienty-Busch, J.Swienty-Busch@elsevier.com, Abhinav Kumar, Elena Herzog, Ivan Krstic. Elsevier, Frankfurt, Germany

Reaxys and its predecessors Beilstein, Gmelin and Patent Chemistry have served the chemistry community well with scrutinized, high quality chemical information over the last 150 years. Reaxys is typically used by bench chemists whose focus is on synthetic chemistry, however, Reaxys Medicinal Chemistry, a companion of Reaxys, is also becoming an essential tool for medicinal chemists in drug discovery workflows. With advancements in machine learning technologies and improvements in computing power, new applications are envisioned, and some are already in place, which leverage the vast amount of Reaxys chemistry data readily available in a structured machine-readable format. This talk will discuss new ways on how to provide developers, data scientists and computational chemists with access to this huge knowledge base and presents Reaxys Retrosynthesis Engine (Pending AI) as application in the field of predictive retrosynthesis with examples and recent developments.

CINF 11

Overcoming conflicts and dilemmas in computer-aided synthesis design

Orr Ravitz², ORavitz@cas.org, Gary Gustafson¹, Paul Peters¹. (1) Customer Success, CAS, Columbus, Ohio, United States (2) Product Management, CAS, Columbus, Ohio, United States

Computer-aided synthesis design (CASD) tools are now routinely used by synthetic chemists across different fields and industries. In SciFinder-n alone, thousands of chemists are using predictive retrosynthesis on a daily basis as a means of boosting creativity, and as a way to expose and thoroughly and efficiently explore alternative strategies and methods for making new and known molecules. The coming of age of CASD has been enabled by decades of research, the increase in the amount of usable data, advances in computational capabilities, and, as importantly, by the shift of the mind-set of users. Old skepticism and even hostility have been replaced by realistic expectations and the omnipresence of computational tools in every aspect of modern life.

With many hurdles behind, it is remarkable how many of the remaining challenges of the field are echoes of issues that trailed the discipline from inception: e.g., finding the balance between accuracy and novelty, addressing selectivity, applying strategic planning and identifying the most productive solutions. In this talk, we will share observations and insights from over a year of usage of retrosynthesis in SciFinder-n: the frequency of usage, the main use-cases and the typical user experience. We will provide an overview of the ongoing research and development programs at CAS. In stereo-selectivity and regio-selectivity we will discuss machine-learning vs statistical approaches, as well as the benefits and limitations of deriving the knowledge from the entire corpus vs learning the selectivity for individual reactions. The relationship between solution-set size and diversity and the search algorithm, as well as the dependency on scoring and distance functions will be demonstrated. Finally we
will discuss the impact of the data-set on the scope and accuracy of solutions, and the criticality of evidence on the user-experience deriving the knowledge from the entire corpus vs learning the selectivity for individual reactions. The relationship between solution-set size and diversity and the search algorithm, as well as the dependency on scoring and distance functions will be demonstrated. Finally we will discuss the impact of the data-set on the scope and accuracy of solutions, and the criticality of evidence on the user-experience.

CINF 12

Data preparation for reaction prediction: the quest for “AI-ready” data

Frederik van den Broek¹, f.broek@elsevier.com, Markus Fischer², Marcus Stamm², Gerd Blanke³, Benoit Pasquereau⁴, Michael Collingsworth⁴. (1) R&D Solutions, Elsevier, Amsterdam, Netherlands (2) Information Systems GmbH, Elsevier, Frankfurt, Netherlands (3) StructurePendium Technologies GmbH, Essen, Germany (4) RELX Group, Elsevier, London, United Kingdom

Many believe that Artificial Intelligence has the potential to revolutionise not only life sciences and healthcare, but also chemistry and chemical synthesis, which is why there has recently been a great increase in applying Machine Learning and Big Data-handling techniques to chemical synthesis. Data preparation for synthesis routes’ predictions is often tedious and time-consuming due to variations in how the reactions data is stored in electronic laboratory notebooks or extracted from scientific literature, how reactions are annotated, stored and transformed to be usable for the modelling exercise.

This talk will give an overview of the data preparation, normalisation and de-duplication challenges cheminformaticians and data scientists often face when preparing the AI/ML models. It will also focus on how one could prepare the ever-sought-after “AI-ready” data sets, using the right tools, conventions and chemical rule-sets.

CINF 13

Reaction Transformers for Fingerprints, Classification and Atom-Mapping

Philippe Schwaller¹,² phischwaller@gmail.com, Daniel Probst², Ben Hoover³, Alain Vaucher⁴, Vishnu Nair⁵, David Kreutter⁶, Hendrik Strobelt³, Teodoro Laino¹, Jean-Louis Reymond⁶. (1) IBM Research -- Zurich, Rueschlikon, Switzerland (2) Department of Chemistry and Biochemistry, University of Bern, Bern, Switzerland (3) IBM Research -- Cambridge / MIT-IBM Lab, Cambridge, Massachusetts, United States

Self-supervised language models called transformers have recently revolutionized natural language processing and show tremendous potential when applied to text-based representations of chemical reactions. The patterns in chemical reactions are learned by predicting masked parts of reaction SMILES. The pretrained models can then be specialized on a task like reaction classification, where they reach unprecedented accuracies. Not only can specific outputs of the transformer models serve as fingerprints to map the chemical reaction space without the need of knowing the reaction center or distinguishing between reactants and reagents, but they can also be used to recover the rearrangement between reactant and product atoms. By opening the black-box using detailed visual analysis, we discovered that the transformer models learned atom-mapping without supervision. Atom-mapping, known to be an NP-hard problem, is necessary for making chemical reaction data...
better machine-accessible and crucial for graph- and template-based reaction prediction and synthesis planning approaches. Here, we present an attention-guided reaction mapper that shows remarkable performance in terms of speed and accuracy, even for strongly imbalanced reactions as typically found in patents. This work is the first demonstration of knowledge extraction from a self-supervised language model with a direct practical application in the chemical reaction domain.

CINF 14
Comprehensive search for compounds and chemical reactions in big query

Stephen Boyer¹, skboyer@gmail.com, Ian Wetherbee², Timo Böhme³, Matthias Irmer³, Christoph Ruttkies³, Lutz Weber⁴, lutz.weber@ontochem.com. (1) Collabra Inc., San Jose, California, United States (2) Google, Mountain View, California, United States (3) OntoChem GmbH, Halle/Saale, Germany (4) IT, OntoChem, Germering, Germany

Google BigQuery (BQ) provides a platform for a growing amount of scientific information in various open access data collections. To take advantage of the data in such a large and diverse aggregation, we have enabled structure- and ontology-based searches. The SciWalker Open Data project maintains BQ chemistry-related tables of compounds (currently numbering 131 million) that are updated daily from mentions found in the text or images of patents. These tables include compound SMILES, InChIs, and InChI-Keys as well as unique public ontology concept identifiers, OCIDs. Similarly, we register chemical reactions from patents into a table containing reaction SMILES, RInChIs, the short RInChI-Keys, as well as OCIDs. Both compounds and reactions are classified into respective ontology classes available in BQ for searching.

The main purpose of these tables is to link information from numerous BigQuery tables via standard SQL queries. For example, we can ask ontology-based questions like “which sesquiterpenes have been used in clinical trials as treatments?” We have also used the Chemistry Development Kit (CDK) to create fingerprints for the compounds and reactions, implementing structure and substructure search capability on top of BigQuery.
CINF 15

Evolutionary computing strategies and feedback control for directed execution and optimisation of chemical reactions

Harris Makatsoris\textsuperscript{1,2}, harris.makatsoris@kcl.ac.uk. (1) Engineering, King's College London, London, United Kingdom (2) Centillion Technology Limited, MK43 0TR, Cranfield, United Kingdom

The evolution of a chemical system towards a desired property within a fitness landscape, is a very attractive strategy for reaction optimisation experimentally. It allows the efficient exploration within complex parameter spaces that may contain multiple maxima or minima but without any detailed knowledge of the structure of the space. Unlike other approaches, it does not impose a requirement to collect information necessary to calculate gradients. Experimental design approaches that have been recently reported although have demonstrated good performance, they employ search strategies along a single steepest ascent or descent pathway with some cases requiring gradient calculation. This prevents them from discovering better designs within complex spaces as they get confined (trapped) very quickly within a particular region of the space as they only explore around a single extremum. In contrast, evolutionary approaches avoid this as they sample points from across the whole search space. Furthermore, evolutionary strategies are robust and resilient to experimental and measurement errors and can be applied in manual or fully automated experimental scenarios. However, automated techniques platforms rely on feedback mechanisms that require the integration of Process Analytical Tools (PAT) and methodologies to pre-process the data from observables before determining the next experimental design of the next iteration. This talk demonstrates the application of these techniques with the use of a fully automated flow system.
Directed reaction planning and control via a closed loop evolutionary computation strategy

**CINF 16**

**Use of conditional generative adversarial networks (cGAN) with molecular orbital fields to study chlorination sites**

*Matthew Clark, mclark@drmatthewclark.com. Life Sciences, Elsevier, Wayne, Pennsylvania, United States*

We have been studying use of 3D molecular orbitals of products and reactants as input to
conditional generative adversarial networks (cGAN) to model organic chlorinations. Our previous work with electron densities has been extended to include nuclear density. The resulting cGAN then takes a 3D computed density and generates an image of the electronic structure and nuclear positions of the chlorinated product. The current status, successes, and limitations will be discussed.

Example chlorination

CINF 17

Expanding the synthetically feasible chemical space of amides to a billion of compounds

Olexandr Isayev, olexandrisayev.com. Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States

Machine learning (ML) has got considerable attention during the last years and is currently at the peak of the hype cycle. A number of applications of ML have been reported including drug discovery and synthesis planning. We have decided to apply ML to the available experimental data aiming to expand the REAL chemical space offered by Enamine. Amide coupling is one of the widely-utilized transformations for the last stage functionalization. Herein, we report our results on the application of ML methods and extensive experimental validation one of the protocols of amide formation.

CINF 18

Publishing crystal structure data – keeping up with the times

Ian Bruno, bruno@ccdc.cam.ac.uk. Cambridge Crystallographic Data Centre, Cambridge, United Kingdom

Over many decades, communities of crystallographers have worked to establish the standards, services and workflows that enable tens of thousands of crystal structures to be published each year. The widespread use of crystallographic data resources in chemistry and
biology by both industry and academia is testament to the successes of these initiatives at channelling publicly funded research outputs into commercial innovation. A key enabler of this has been the partnership between data repositories and journal publishers who together have encouraged authors to publish their crystal structures in ways that enable scientific validation and data reuse. However, publication paradigms are changing, expectations surrounding data publication are evolving and machines are becoming increasingly involved in the execution of science. How are data sharing practices relating to crystal structures keeping up?

This presentation will look at how the sharing of crystallographic data may – or may not – need to adapt to a publishing environment shaped by an increasing uptake of preprints and growing pressure from initiatives such as Plan-S. It will go on to look at how data repositories can enable publication of datasets independently of journal articles and consider how doing so might help researchers satisfy policy requirements. Finally, we will reflect on the increasing adoption of digital practices in structural chemistry across industry and academia and look at how this is generating requirements for a wider range of high-quality experimental datasets and new reporting standards.

**CINF 19**

**Is Web Assembly the future of visualizing molecules in the browser?**

*Jonathan Fine, jonathanafine@gmail.com, Gaurav Chopra, gchopra@purdue.edu. Chemistry, Purdue University, West Lafayette, Indiana, United States*

As the internet has evolved, so has the techniques used to share chemical information over the world wide web. One of the earliest methods for visualizing molecules on the web was a browser plugin called Chime. It allowed chemists to create websites to explain chemistry in a three-dimensional manner. As internet technologies moved towards Java plugins, the Jmol applet was developed to provide a similar experience that could be distributed by the website itself without the need for a specialized browser plugin. With the introduction of WebGL and the HTML5 canvas element, developers moved towards applications written in Javascript to render molecules in 3D dimensions. Examples of these Javascript applications include 3Dmol, JSmol, NGL viewer, GLmol, and LiteMol. Recently, a new technology called Web Assembly (WASM) has emerged and is widely heralded as an important leap forward for the development of new applications that run in a web browser. WASM allows one to use compiled computer languages such as C and C++ to be deployed by users without the use of any plugins, but the compiled nature of the applications allows for significant gains in performance across all applications. This technology is already supported by 95% of all web users, and many companies have reported success in porting traditional desktop applications to the web using this technology. The following question remains, will developers of chemical visualization software adopt this new technology? In this talk, I will discuss the process of creating a new 3D molecule viewer using this technology in the hopes of inspiring a discussion on what the adoption of WASM means for the feature of visualizing molecules on the web as the adoption of this new platform could make chemistry available to wider audience. Specific examples will include the visualization of molecular orbitals and vibration frequencies and their potential association with measurements in the laboratory and how these technologies can be deployed to users worldwide.
CINF 20

Integration of chemistry with everything else

Ian Wetherbee\(^3\), wetherbeei@google.com, Lutz Weber\(^1\), Evan Bolton\(^2\), Steve S. Walker\(^3\), Stephen Boyer\(^4\), skboyer@gmail.com, Jane Frommer\(^4\). (1) IT, OntoChem, Germering, Germany (2) National Center for Biotechnology Information, Bethesda, Maryland, United States (3) Google, Mountain View, California, United States (4) Collabra Inc., San Jose, California, United States

The discovery and association of molecules and their attributes are important for strengthening scientific awareness in commercial and societal settings. With increasing demand for molecular information as input for machine learning - leading to yet further discovery - the identification, organization and availability of the world’s molecular content are all in demand. Together Google Patents and OntoChem are making a significant contribution of computer-curated molecular data to NIH PubChem, upholding the FAIR data principles of Findable, Accessible, Interoperable, and Reusable. This donation to NIH will make available for the first time machine-curated data derived from text and images of previously uncurated patents from around the world. Additionally, Google is providing a BigQuery platform for the integration of molecular data with content from other worldwide resources. Areas to benefit from this platform include medicine, economics, agriculture, climate change, pharma, chemistry and the law.

CINF 21

Withdrawn

CINF 22

Quantum descriptors real-time predictions and applications to reaction planning

Yanfei Guan\(^1\), yanfeig@mit.edu, Haoyang Wu\(^1\), Duminda S. Ranasinghe\(^2\), Esther Heid\(^1\), Thomas J. Struble\(^3\), Lagnajit Pattanaik\(^1\), Connor W. Coley\(^3\), William H. Green\(^4\), Klavs F. Jensen\(^5\). (1) Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States (2) department of Chemistry, Wesleyan University, Middletown, CT, Middletown, Connecticut, United States (3) Chemical Engineering, Massachusetts Institute of Technology, Belmont, Massachusetts, United States (4) Rm E17-504, MIT, Cambridge, Massachusetts, United States (5) Dept of Chem Eng Rm 66 350, MIT, Cambridge, Massachusetts, United States

Through use of domain knowledge, feature engineering has created various of descriptors that enable machine learning algorithms to predict molecular properties or activity and
selectivity for a given organic reactions. However, such feature engineering methods usually require expert knowledge about the predicting target, and are difficult to be transferred to other tasks. On the other hand, feature learning methods, such as graph neural network (GNN), automatically discover the representation of atoms and bonds through 'end-to-end' learning on molecular structures. Those deep neural networks usually require complex models, which significantly hinders its application on a relatively small dataset. Herein, we propose a prototype to combine the feature engineering and feature learning techniques by using computed common atomic/bond descriptors, e.g. Fukui indices and bond orders, as input for a GNN model to improve its performance. The computed descriptors provide the model with physical and chemical information that the model is unable to capture by its own. We applied the proposed prototype to the selectivity predictions for regio-selective reactions selected from the Pistachio database. Our results demonstrated that quantum descriptors enable training a GNN model on a sparse dataset with improved accuracy and extrapolated performance. A real time predicting model is also developed to allow the on-the-fly prediction of those descriptors.

CINF 23

Synthetic feasibility and de novo molecular generation and optimization

Wenhao Gao, Connor W. Coley, ccoley@mit.edu. Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

There is substantial interest in de novo molecular generation and optimization techniques as a way to propose new molecules during early-stage drug discovery. Deep generative models and other inverse design techniques allow for the multi-objective optimization of surrogate models (e.g., for activity, druglikeness) without relying on brute-force screening of a virtual chemical library. However, their utility is limited by an ignorance of synthesizability. This talk will describe an evaluation of several state-of-the-art methods in terms of their abilities to generate synthesizable molecules as judged by a data-driven chemical synthesis planning program. I will summarize emerging methods that explicitly integrate synthetic feasibility into their generative processes, which is an important step toward overcoming this limitation.
Big errors in big data: when automated data curation misses the mark

Robert D. Clark\textsuperscript{1}, bob@simulations-plus.com, Pankaj R. Daga\textsuperscript{2}, Marvin Waldman\textsuperscript{3}. \textsuperscript{(1)} Simulations Plus, Inc., Lancaster, California, United States \textsuperscript{(2)} Simulations-plus Inc, Lancaster, California, United States \textsuperscript{(3)} Simulations Plus, Inc., San Diego, California, United States

Cheminformatics modeling has always drawn heavily on compilations as a source of data. The most desirable ones were assembled by experts in a particular field, typically as literature. Compilations assembled by cheminformatics groups specifically for model building purposes tended to be less carefully curated but were attractive because they were typically large - sometimes because they were compilations of compilations. The advent of sophisticated data mining tools has made it much easier to automate compilation, and the need for big data to drive deep learning has brought us into the age of Big Data. The sheer size of the data sets involved has driven development of automated curation techniques, particularly ones focused on standardizing structures. Unfortunately, the tools involved sometimes introduce systematic errors or over-standardize data; this talk will explore examples of the kinds of errors that can be inadvertently introduced in the process.

Conformal regression for Profile-QSAR: Adding confidence intervals to point predictions

Xiangwei Zhu\textsuperscript{1}, xwzhunc@gmail.com, Eric J. Martin\textsuperscript{2}. \textsuperscript{(1)} Novartis Institutes for Biomedical Research, Emeryville, California, United States \textsuperscript{(2)} Computational Chemistry, Novartis, El Cerrito, California, United States

Profile-QSAR (pQSAR), a massive multi-task machine learning platform that enables transfer learning among thousands of assays, has been applied to over 100 Novartis drug discovery projects. pQSAR overall correlation between prediction and experiment is comparable to 4-concentration IC\textsubscript{50}s on novel compounds for over 8,000 assays. Until now, however, pQSAR did not estimate confidence intervals for individual predictions. In this study, conformal regression was adapted to profile-QSAR to deliver prediction intervals guaranteed correct at a given confidence level as far as the assumption of exchangeability with the calibration set holds. Finding the right nonconformity measure and calibration set is the key to conformal regression. Conformal analysis is complicated for pQSAR, because it uses a non-random test set of novel compounds which complicates exchangeability, it does not give an ensemble of predictions to estimate non-conformity, and distance to the training set is complicated by the contributions from helper models. Our efforts have been focusing on three questions: (1) which ensemble algorithms that could provide standard deviation as a nonconformity measure generate predictions as accurate as our \textit{de facto} standard of partial least squared regression (PLSR) at the second step of profile-QSAR; (2) are there residual models that work well as an alternative nonconformity measure; and (3) what is the best strategy in trading-off prediction accuracy, prediction intervals, and computational cost for such a big data problem. We conclude that splitting the “realistic” test set allows for exchangeability, ensemble support vector machines are comparable with PLSR in accuracy and produced fairly good conformal prediction intervals, but at high computational cost, pQSAR residual models that include the profile-QSAR predictions themselves as a descriptor gave good conformal prediction intervals at low computational cost, and weight optimization of the
nonconformity score improved the prediction intervals and reduced prediction interval outliers.

**CINF 26**

**Revealing hidden determinants of molecular recognition in hit-to-lead optimization with graph convolutional attention mechanisms**

Brandon Anderson¹, Stefan Schrödl¹, Henry van den Bedem¹², vbedem@stanford.edu. (1) Atomwise, San Francisco, California, United States (2) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, California, United States

Virtual high throughput screening is a common first step in structure based drug discovery for identifying hits against a therapeutic target from vast libraries of compounds. Virtual screening generally yields low-affinity binders that serve as a starting point for lead generation. However, optimizing a hit’s pharmacokinetic and toxicological properties while maintaining or improving its potency and selectivity remains a formidable challenge. Small chemical changes can result in dramatic, unforeseen losses in activity, often referred to as an ‘activity cliff’. Here, we address that challenge by explicitly visualizing the chemical determinants of molecular recognition between a compound and target. We trained our graph convolutional network, AtomNet Graphite, to predict affinities of a large number of diverse compounds on a class of kinases. AtomNet Graphite includes an attention mechanism to distinguish the contribution of distinct molecular interactions to binding affinity. We visualized the interaction weights imposed by the attention mechanism, and analyzed their distributional entropies. We found that our attention mechanism improved AtomNet Graphite’s predictive power, and helped direct hyperparameter optimization. A retrospective analysis of several compound series revealed that attention weights of chemical groups that increased generally corresponded to more potent compounds, while decreasing weights corresponded to increased selectivity. AtomNet Graphite’s attention network analysis is a powerful new tool in our hit-to-lead optimization toolbox.

**CINF 27**

**Guiding Anti-Malarial Optimisation Using Deep Learning Imputation and Compound Generation**

Benedict Irwin¹, Alexander Wade², Thomas Whitehead³, Gareth Conduit³², Matthew D. Segall¹, matthew.d.segall@gmail.com. (1) R&D, Optibrium Limited, Cambridgeshire, United Kingdom (2) Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom (3) Intellegens Limited, Cambridge, United Kingdom

We describe a combination of deep learning imputation, compound generation and multi-parameter optimisation (MPO) methods to guide compound optimisation, resulting in the discovery of an active compound targeting a novel anti-malarial mechanism of action. Deep learning imputation was applied to a sparse data set provided by the Open Source Malaria (OSM) project, including activities measured in multiple assays for activity against the pfATP4 target and the Plasmodium falciparum parasite. The resulting model accurately predicted activities across this panel of assays and provides an uncertainty estimate for each predicted value.

This model was applied to generate virtual compounds using medicinal chemistry
transformations applied to a chemical series previously explored against the pfATP4 target. The resulting compounds were prioritised using the Probabilistic Scoring method for MPO to identify those with the best balance of activities across multiple assays and other desirable properties such as solubility. The most confidently predicted high-quality suggestion was synthesised and tested by the OSM project and verified to be active, with an IC$_{50}$ of 640 nM.

The integration of these methods, combines state-of-the-art machine learning technologies to explore new strategies for optimisation and confidently focus on those with the best chance of success.

CINF 28

Board Game AI in Continous Action Spaces: A Monte Carlo Tree Search Algorithm for Material Application

Troy Loeffler$^1$, tloeffler@anl.gov, Sukriti Manna$^{2,4}$, Rohit Batra$^{2,4}$, Srilok Srinivasan$^2$, Henry Chan$^{2,4}$, Subramanian Sankaranarayanan$^{3,4}$. (1) Chemistry, Louisiana State University, Baton Rouge, Louisiana, United States (2) Argonne National Laboratory, Schaumburg, Illinois, United States (3) Argonne National Lab, Naperville, Illinois, United States (4) University of Illinois at Chicago, Chicago, Illinois, United States

From forcefield development to structural optimization, optimizing a continous set of problem parameters is a commonly reoccurring family of problems in computational sciences. While excellent local optimization approaches exist, the field of global optimization algorithms is still very young field with much open space to explore. Researchers are always in need of highly efficient global search algorithms that are capable of handing very high dimensional search spaces without a loss of efficiency. In this presentation we present a new optimization algorithm based on the popular Monte Carlo Tree Search algorithm. Monte Carlo Tree Search is an algorithm traditionally designed to play board games such as Chess, Go, and other games with a high branch factor. It has been found that this algorithm's formalism can be adapted to continous parameter spaces in a way that can produce a highly efficient search algorithm that is capable of handing high dimensional search spaces ranging from 13 parameters and extremely far beyond.
Discovering novel fluorescent molecules via global optimization of molecular properties

Beomchang Kang, Chaok Seok, Juyong Lee, juyong.lee@kangwon.ac.kr. (1) Department of Chemistry, Kangwon National University, Chuncheon, Gangwon-do, Korea (the Republic of) (3) Department of Chemistry, Seoul National University, Seoul, Korea (the Republic of)

Fluorescent molecules are widely used for bio-imaging. They are attached to specific cell organelles and/or proteins, enabling observation of detailed structure and dynamics in the cell. Efficient fluorescent molecules must have a high quantum yield for effective bio-imaging. Here, we present a systematic approach to discovering novel fluorescent molecules that combines machine-learning and global optimization algorithms. We recast the problem of discovering novel fluorescent molecules with high-intensity emission light into a global optimization problem by using the oscillator strength of a molecule as an objective function for optimization.

A statistical machine that predicts excitation energies and associated oscillator strengths, the probability of absorption or emission of light in transitions between different energy states, of a molecule were trained using the random forest algorithm. The PubchemQC database, which contains TD-DFT calculation results of 3.8 million known molecules, was used as a training set. The extended connectivity fingerprints of molecules were used as input vectors. To optimize the oscillator strength of a molecule, a highly efficient global optimization algorithm called CSA was used. For CSA global optimization, SMILES representation of a molecule was mapped to a 200-dimensional integer vector by using Natural Language Toolkit. After CSA global optimization calculation converged, we assessed the validity of our approach by performing quantum mechanical calculations. TD-DFT calculations were carried out to verify whether novel molecules obtained by this procedure actually have high oscillator strength.

Predicting diffusion in Lennard Jones fluids and beyond using machine learning

Joshua Allers, jpaller@sandia.gov, Jacob Harvey, Fernando Garzon, Todd M. Alam. (1) Organic Materials Science, Sandia National Laboratory, Albuquerque, New Mexico, United States (2) MS 0886, Org. 1816, Sandia National Labs, Albuquerque, New Mexico, United States (3) Geochemistry, Sandia National Laboratory, Albuquerque, New Mexico, United States (4) Center of Micro-Engineered Materials, University of New Mexico, Albuquerque, New Mexico, United States (5) Power Sources, Sandia National Laboratory, Albuquerque, New Mexico, United States

Accurately predicting the molecular self-diffusion in solutions is important for the development and understanding of many processes and chemical reactions. The self-diffusion is affected by many parameters including operating conditions and molecular properties; however, the relationships between those parameters and the diffusion constant can be difficult to extract. To improve these relationships, a better understanding of how specific properties influence the self-diffusion needs to be explored. The development of a general predictive model that can accurately estimate self-diffusion is therefore crucial for understanding which properties are important and which are correlated. Machine Learning (ML) methods such as Artificial Neural Networks (ANN) and Random Forests (RF) have proven to be powerful tools for
predicting chemical and material properties. They also have the potential to provide information about the properties that most impact diffusion. In this presentation, we describe studies on a test system involving the well-defined Lennard Jones (LJ) fluid to assess the performance of ANN and RF models and gain insights into methods of improving the diffusion prediction process. Feature engineering is employed to generate a vast feature dimensionality, allowing us to assess the impact of using different functional forms of base features to improve ML model performance. This simple example paves the way for modeling of diffusion in real fluid systems, including mixtures and fluid diffusion in porous materials, using molecular properties as features during the training of ML models.

CINF 31

Heuristic Global Optimization in Chemical Compound Space

Christopher B. Rinderspacher, crinders@gmail.com. Army Research Laboratory, Baltimore, Maryland, United States

We outline developments in combinatorial optimization under constraints, in particular in chemical space. Presented is the analysis of periodic kernels for efficient global optimization (EGO) compared to heuristic local searches and global sampling: the modified Dirichlet kernels $D_{n}(x) = \sum_{i=1}^{n} \left[ \cos(2\pi ix) - (-1)^{i} \right]/i$, $K_{L}(x) = \cos(\pi x/L)^{2} + \delta_{x/2}$, and the periodic Gaussian kernel $G_{L,\sigma}(x) = \exp(-\sum|x/L-1/2|^{2}/2\sigma^{2})$, where $n$ is the order of the Dirichlet kernel, $L$ is the number of choices on a combinatorial site, $|x|$ is the periodic absolute value function over [-1/2,1/2], and $\sigma$ is a broadening factor. Three problem types were considered. The first are random functions. The second is a constrained optimization over a database of electrochromic molecules with 2 constraints. Finally, a large search space of $10^{12}$ electrochromic molecules was searched for trifunctionality, introducing 6 constraints. We find that, on average, EGO outperforms a local search for random functions but is less competitive with sequential heuristic next-neighbor search (HNNS) on the first electrochromic problem. HNNS finds the global optimum (GO) ~90% of the time within 8% of the search space. With the right kernel and hyperparameters EGO performs similarly. EGO is most effective in the large search-space scenario where it finds viable candidates within only a 100 evaluations out of the $10^{12}$ molecules.
Clustering and predicting properties of general polymers based on their monomer unit structure

Hitoshi Yamano\textsuperscript{1,2}, yamano.hitoshi.ya8@is.naist.jp, Hiroaki Shimizu\textsuperscript{2}, Shigehiko Kanaya\textsuperscript{1,3}, Tomoyuki Miyao\textsuperscript{4,3}, Aki Morita\textsuperscript{1}, Naoki Ono\textsuperscript{1,3}. (1) Division of Information Sciences, Computational Systems Biology Lab., Nara Institute of Science and Technology (NAIST), Ikoma, Nara, Japan (2) Tokyo Ohka Kogyo Co., Ltd., Koza-gun, Kanagawa, Japan (3) Data Science Center, Nara Institute of Science and Technology (NAIST), Ikoma, Nara, Japan (4) Division of Materials Science, Data Driven Chemistry Laboratory, Nara Institute of Science and Technology (NAIST), Ikoma, Nara, Japan

In polymer material development, multiple physical and chemical properties need to be simultaneously optimized. However, polymer properties are usually more difficult to predict than those of small molecules due to them forming superstructures. In this work, we aimed at finding a versatile approach to predict multiple polymer properties using imperfect data with missing values.

We prepared a data set consisting of 50 polymers with 45 properties along with their monomer unit structures. The data set was hierarchically clustered on the basis of two independent factors: polymer properties and polymer structures. In polymer property-based clustering, visualizing relations of polymers was found to be an effective way of estimating the difficulty of polymer property prediction. In polymer structure-based clustering, each cluster could be formed based on the structural features. Thus, the clustering contributed to understanding structural characteristics of monomer unit structures.

In addition to analyzing the data set in an unsupervised manner, we constructed polymer properties prediction models based solely on the information of monomer unit structures. Partial least squared regression models could predict density, glass transition temperature, and dissolution parameter with high accuracy. Values of $R^2$ for test data sets were between 0.88 and 0.97. In our presentation, we will discuss the possibility of predicting various polymer properties in a way similar to ours.

Performance and scope of similarity-based and machine learning approaches for predicting the macromolecular targets of small molecules

Neann Mathai\textsuperscript{1,2}, neannm@gmail.com, Johannes Kirchmair\textsuperscript{1,2,3}. (1) Computational Biology Unit, University of Bergen, Bergen, Norway (2) Department of Chemistry, University of Bergen, Bergen, Norway (3) Department of Pharmaceutical Chemistry, University of Vienna, Vienna, Austria

Spurred by the growing amount of available biological data, the increasing use of phenotypic assays in small-molecule drug discovery, and a shift from the “one drug one target” paradigm to “polypharmacology”, a large number of target prediction methods have been developed and applied in the last few years. Majority of the studies on in-silico target prediction present the performance of a method as an averaged performance over all query compounds from retrospective validation (testing the methods’ ability to predict known interactions). This leaves the user unclear about how the performance of a method is related to the relationship between a query and the underlying knowledge base used by the approach. Additionally,
machine learning approaches are typically based on a series of binary classification models, and studies based on such approaches tend to present performance with respect to the average classification performance of the individual models themselves and not of the approach as a whole. Here, we present the results of our work on quantifying the performance of two popular target prediction approaches: similarity-based and random forest machine learning models, with respect to a query’s similarity to the underlying reference data used by the approach. We test the approaches under three different testing scenarios: a standard testing scenario with external data, a time-split testing scenario and a “close-to-real-world” testing scenario where targets which are not represented by an approach’s knowledge base are taken into account. We show that the similarity-based approach performed better than the machine learning approach under all scenarios, ranking at least one known target among the top-5 proteins for 88%, 63% and 59% of all test compounds under the standard external testing scenario, the time-split testing scenario, and the close-to-real-world testing scenario, respectively. The respective success rates for the machine learning approach were 86%, 57%, 52%.

CINF 34
Withdrawn

CINF 35
Deep learning showed increased phenotypic assay sensitivity in quantifying inhibitors of α-synuclein inclusion formation

Lakshmi B. Akella, lakshmbakella@gmail.com, Melissa Bennion, Patrick Faloon, Andreas Weihofen, Richard Nelson. Biogen, Cambridge, Massachusetts, United States

Formation of α-synuclein inclusions is linked to pathogenesis of Parkinson’s disease. A high-content imaging screen was conducted to probe Biogen’s chemogenomic set for inhibitors of α-synuclein inclusion formation in M17 neuroblastoma cell line stably expressing α-syn 3K-GFP. Cellular phenotypes in image-based assays are complex and the richness of high-content data collected during phenotypic screens are often underutilized. New methods are needed to fully leverage this high-content data. An approach assisted by machine learning could boost the robustness and sensitivity of phenotypic assays used to characterize small-molecule modulators. We will present the comparison of single-feature analysis relative to the models obtained from high content features towards the sensitivity of inhibitor potency.

CINF 36
ChEMU shared task: chemical entity recognition and event extraction of chemical reactions from patents

Christian Druckenbrodt¹, Saber Akhondi¹, Karin Verspoor², Dat Nguyen², Camilo Thorne¹, c.thorne.1@elsevier.com, Zenan Zhai². (1) Elsevier, Frankfurt, Germany (2) University of Melbourne, Melbourne, New South Wales, Australia

We introduce a new chemical information extraction shared task, named ChEMU, part of the 11th Conference and Labs of the Evaluation Forum (CLEF-2020). ChEMU proposes two key information extraction tasks over chemical reactions from patents. Task 1 --- named entity
recognition (NER) --- is to identify specific types of chemical compounds, i.e. to assign the label of a chemical compound according to the role for which the chemical compound plays within a chemical reaction. Task 2 --- event extraction over chemical reactions (EE) --- involves on the other hand event trigger detection and argument recognition. We will publicly release reaction-specific gold standards --- that we describe in this presentation --- derived from patent literature and annotated by chemists (for NER and EE, resp.) in early 2020. Thereafter academic or industrial teams working in the field will be encouraged to participate in a shared evaluation campaign, by developing and contributing NER and EE models. The models will be then evaluated (by measuring their recall, precision and F1 scores), compared and jointly discussed at CLEF-2020.

CINF 37

From High-Throughput Computational Chemistry and Molecular Pattern Recognition to the Targeted Design of Novel Chemistry

Johannes Hachmann, hachmann@buffalo.edu. Dept of Chemical and Biological Engineering, University at Buffalo, SUNY, Buffalo, New York, United States

The process of developing new compounds and materials is increasingly driven by computational modeling and simulation, which allow us to characterize candidates before pursuing them in the laboratory. In this work, we employ virtual high-throughput screenings to generate extensive chemical data sets, employ novel pattern recognition techniques on these data sets, and then employ the gained insights to generate new compounds with targeted property profiles.

CINF 38

Machine-learning correlation energies using hybrid structural-energetic descriptors

Robert A. Distasio, distasio@cornell.edu. Department of Chemistry and Chemical Biology, Cornell University, Freeville, New York, United States

In this work, we will describe a novel set of machine-learning (ML) descriptors that incorporate both structural and energetic information. Such hybrid structural-energetic ML descriptors will be employed for learning the following two fundamental quantities in quantum chemistry: molecular correlation energies and basis-set incompleteness error.

CINF 39

Machine learning augmented quantum chemistry: From deep learning of wave functions to ML/QM-tandem methods

Reinhard J. Maurer, r.maurer@warwick.ac.uk. Department of Chemistry, University of Warwick, Coventry, United Kingdom

Atomistic simulation based on quantum mechanics (QM) is currently being revolutionized by machine-learning (ML) methods. Many existing approaches use ML to predict molecular properties from quantum chemical calculations. This has enabled molecular property prediction within vast chemical compound spaces and high-dimensional parametrization of
energy landscapes for the efficient molecular simulation of measurable observables.

However, as all properties derive from the QM wave function, an ML model that is able to predict the wave function also has the potential to predict all other molecular properties. In this talk, I will explore ML approaches that directly represent wave functions and QM Hamiltonians and their derivatives for developing methods that use ML and QM in synergy. Using examples from molecular chemistry and heterogeneous catalysis, I will discuss the challenges associated with encoding physical symmetries and invariance properties into deep learning models. Upon overcoming these challenges, integrated ML-QM methods offer the combined benefits of big-data-driven parametrization and first-principles-based methods. I will discuss several opportunities associated with building ML-augmented quantum chemical methods, including Inverse Chemical Design based on ML-predicted wave functions and the development of efficient and accurate semi-empirical methods to study hybrid metal-organic materials.

CINF 40

BioHarmony across semantic drug discovery data streams

Barry Bunin, bbunin@collaborativedrug.com, Hande K. McGinty. Collaborative Drug Discovery, Alexandria, Virginia, United States

The recently developed CDD’s BioHarmony (BH) platform is geared to leverage machines to accelerate all stages of drug development and approval (preclinical, clinical, post approval, repurposing). It uses structured semantic data and meta-data to allow researchers to have greater confidence in their data quality for machine learning applications. The online BioAssay Express technology automatically extracts unstructured free text data (from online databases, offline publications, and electronic laboratory notebooks entries) into structured databases.

The technology combines aspects of manual quality curation with automated algorithms to give the optimal tradeoff per effort. The end result is the conversion of human-readable text descriptions into a machine-readable database format. BioHarmony uses public semantic standards (ontologies) to markup drug development and approval data that could previously only be organized by crude text searching.

One of several annotation-support strategies within BH platform is the use of machine learning models to provide statistically backed suggestions. We will describe our efforts to complement these models by applying ontology derived text mining, association rules mining based on existing annotations, and axioms that are embedded within the underlying ontologies. BioAssay Express includes the BioAssay Ontology (BAO), Gene Ontology (GO), Drug Target Ontology (DTO), and Cell Line Ontology (CLO). It can also be extended to handle other, proprietary ontologies, given it is a general technical solution. Datasets are constantly increasing, with the volume already exceeding the ability of individual scientists to manage productively, necessitating technical solutions to state ahead of the curve.
Accurate Gaussian process models for high-dimensional molecular systems

**Roman Krems**, rkrems@chem.ubc.ca. Chemistry, The University of British Columbia, Vancouver, British Columbia, Canada

High-dimensional chemistry and physics problems are generally modelled by neural networks (NNs). However, NNs require large data and are generally problem-specific. An alternative approach is probabilistic modelling based on Gaussian processes (GPs), which can be made problem-agnostic, can be fully automated and can be used for Bayesian optimization. However, GPs suffer from the numerical complexity and have been limited to low-dimensional problems for applications that require high accuracy. In this talk I will illustrate that it is possible to build accurate GP models of a 51-dimensional physical system based on 5000 inputs, using fragmentation of the input space and automated model complexity enhancement. I will illustrate the algorithms by GP models of the global potential energy surfaces (PES) for molecular systems with up to 19 atoms. I will show that GP models thus constructed have generalization power, allowing one to extrapolate PESs from low energies to high energies. I will argue that this work opens the prospect for building machine-learning models that can be used to search for new physics, such as phase transitions, in high-dimensional physics and chemistry problems with unknown property landscapes by extrapolation.

Enhancing data validation and summarization in PubChem via automated link analysis

**Leonid Zaslavsky**, zaslavsk@ncbi.nlm.nih.gov, Tiejun Cheng, Asta Gindulyte, Paul Thiessen, Bo Yu, Jian Zhang, Evan Bolton. National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, United States

PubChem and other biomedical data repositories contain a wealth of assertions about relationships by data contributors based on their experiments, clinical, or theoretical work. Additionally, such relationships can be derived by analyzing biomedical literature, patent literature, sequence or structural data, epidemiological reports, clinical trials data, and other sources.

In order to uncover important relationships between chemicals, genes, proteins, and diseases, the relevant pieces of data in numerous aforementioned information sources are being organized into a knowledge graph, and a functionality to identify and efficiently retrieve relevant links is being created. We also plan to develop heuristics for a scoring system that can assess the link quality from all available sources and develop techniques for automated summarization of link data.

The results of this project will enable better knowledge representation in PubChem pages, allow to verify submitter-provided links in our databases helping to automate bio-curation efforts, as well as uncover and present new links calculated from the complex data.
Development of a unified platform for registration of all entities from small molecules to biologics and conjugates using HELM

Jonathan Buttrick, jonathancbuttrick@gmail.com. Scilligence, South San Francisco, California, United States

The diversity of entities involved in the drug discovery process makes the development of a unified registration system particularly challenging. The growing interest in complex systems such as antibody-drug conjugates and DNA encoded libraries makes the need for such a unified system vital. The collaborative development of HELM and implementation of it as an industry standard has helped make this dream a reality. The speaker will demonstrate the types of challenges HELM has helped to solve for biologics registration as well as discuss the current limitations.

Template based modeling of protein assemblies in the ClusPro server

Dzmitry Padhorny¹, dnpodgorny@gmail.com, Dima Kozakov¹, midas@laufercenter.org, Sandor Vajda², Ignatov Mikhail¹, Kathryn Porter³, Sergey Kotelnikov¹, Ryota Ashizawa¹, Andrey Alekseenko¹. (1) Applied Mathematics and Statistics, Stony Brook University, Boston, Massachusetts, United States (2) Boston University, Boston, Massachusetts, United States (3) Boston University, Brookline, Massachusetts, United States

ClusPro is a web server designed to make the structural modeling of protein assemblies accessible to the wide research community. Historically, the functionality of ClusPro was focused on ab initio protein docking, where the structure of the binary protein complex is predicted based on the structures of the components provided as inputs. Over the years, this basic functionality was expanded with symmetry-guided docking, protein-peptide docking, use of small angle X-ray scattering data and geometric restraints. Together, this arsenal of tools repeatedly allowed ClusPro to excel at the CAPRI competition for blind prediction of protein complexes. However, our analysis of server performance results showed that, when good templates were available, a purely template-based approach for multimer structure prediction produced models of substantially higher-quality compared to ab initio docking. Motivated by this observation, we have enhanced ClusPro with template-based modeling capabilities, improving the server's ability to produce higher accuracy predictions, both for homomeric and heteromeric targets. Here we review this latest addition to the ClusPro web server and discuss our predictions for recent CASP and CAPRI targets that benefited from this development. We also describe ongoing work which is not yet added to the server, including the development of methods of improving template-based models and the use of co-evolutionary information in protein assembly modeling.
Adapting evolutionary algorithms for autonomous machine learning in chemistry

Gaurav Vishwakarma¹, gvishwak@buffalo.edu, Johannes Hachmann². (1) Chemical and Biological Engineering, University at Buffalo, Buffalo, New York, United States (2) Dept of Chemical and Biological Engineering, University at Buffalo, SUNY, Buffalo, New York, United States

Over the past few years, there has been a rising trend in the use of data-driven approaches for the discovery and design of new chemistry. These strategies encompass virtual high-throughput exploration of chemical space and analysis of the generated data using data mining techniques, for which our group has developed a cyberinfrastructure to help accomplish these tasks. However, in order to accelerate this process, it is critical to identify and optimize several steps in our research paradigm and we therefore employ a Genetic Algorithm (GA) for generating high quality solutions to search and optimization problems.

Rational model selection and hyperparameter optimization are important concerns for the efficient and successful use of machine learning but have so far largely remained unexplored by this community. To address these issues, we use GA to automatically perform these tasks and present results for training an artificial neural network to predict refractive indices of one hundred thousand small organic molecules. Further, we benchmark its performance against other algorithms popular in the data science community and the results show that our implementation outperforms these methods both in terms of time and accuracy. The effectiveness of our implementation is further demonstrated via a scenario involving multi-objective optimization for model selection.

Next, we use the GA-optimized machine learning model for small organic molecules to ‘inverse-engineer’ new molecules that will have a high refractive index. Since chemical space is infinite and its exhaustive exploration is impractical for any real application, it is desirable that this exploration proceeds in a direction biased towards a set of target properties. In principle, a fitness function for GA can be designed such that it penalizes sub-structures that suppress the desired property in a molecule. However, the biggest challenge in using GA for high-throughput screening is the encoding of molecules into a GA chromosome. We modify our implementation of GA to not only handle this by using a node-network of python objects for every molecule, but also keep track of the connections between molecules which can later be used for creating structural descriptors in a machine learning model. Using GA, we propose new molecules with high refractive index and validate the predicted values using Density Functional Theory.

Insights into therapeutic fusion protein R&D from an analysis of the CAS databases

Yingzhu Li, yingzhuli@cas.org, Yi Deng, Cynthia Y. Liu. CAS, Columbus, Ohio, United States

Therapeutic fusion proteins have emerged as a promising class of biologics. Since the first therapeutic fusion protein, etanercept (Enbrel), was approved by the U.S. FDA in 1998, total global sales has been growing at a rate of ~6 percent annually, with the market expected to reach an estimated $24 billion by 2025. In collaboration with the National Science Library of the Chinese Academy of Science, CAS conducted multifaceted analyses of 30 years’ worth of
curated therapeutic fusion protein information in the CAS reference and substance collections. Our analyses not only revealed an interesting global R&D landscape and patent flow pattern, but also covered fusion protein classifications based on half-life extending and activity components, target analysis, mechanism of action and disease indications, among others. Our big data analysis showcased a wealth of biologics information in CAS databases and suggested promising alternative applications of fusion proteins in vaccines, as well as gene and cell therapies.

CINF 47
Withdrawn

CINF 48
Reproducible Data Analysis and Publishing in Chemistry with R: Creating a workshop experience during the ACS National Meeting

Ye Li², yel@mit.edu, Steven P. Wathen³, Donna T Wrublewski¹, dtwrub@caltech.edu. (1) Caltech Library MC 1-43, California Institute of Technology, Pasadena, California, United States (2) MIT Libraries, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States (3) Chemistry, Siena Heights University, Adrian, Michigan, United States

At the Fall 2019 American Chemical Society National Meeting, two chemistry librarians partnered with a chemistry faculty member to offer a hands-on computer programming workshop. The workshop used the R programming environment and was structured for beginners with no previous experience with it. Three main topics were covered during the five-hour session: introduction to R, working with QSAR data, and creating a report using R Markdown. This report will cover the workshop’s logistics, instructional design and content, and discuss feedback received from participants. Suggestions for future improvements will also be discussed.

CINF 49
Chemists’ Data Needs for Machine Learning Research in Academia

Ye Li, yel@mit.edu. MIT Libraries, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

Applied data science and machine learning (ML) are making differences in the chemistry domain including molecular design of drugs and materials as well as prediction and optimization of synthesis routes and production processes. To train effective ML models for these emerging areas, a large amount and a variety of well-curated experimental datasets are essential. Among such datasets needed in chemistry, the structured and curated datasets are oftentimes locked down in proprietary indexing databases, corporate in-house databases, or reference works, either in print or digitized; while the unstructured data scattered in the published research articles. In this presentation, we will first analyze the recent publications reporting ML research in chemistry and examine some ML research projects at MIT to highlight the types and extent of chemistry datasets needed for ML. Then, we will discuss the availability of these datasets and texts as well as the economic, legal and technical challenges in obtaining access to these datasets for data and text mining purposes. To tackle
these challenges, we will share some strategies on exploring and developing collaborations among researchers, librarians, publishers, and database providers.

CINF 50

Machine learning for scent: learning generalizable perceptual representations of small molecules

Alexander B. Wiltschko\textsuperscript{1}, alexbw@google.com, Benjamin Sanchez-Lengeling\textsuperscript{1}, Jennifer Wei\textsuperscript{1}, Brian Lee\textsuperscript{1}, Richard Gerkin\textsuperscript{3}, Alan Aspuru-Guzik\textsuperscript{2}. (1) Google, Cambridge, Massachusetts, United States (2) University of Toronto, Toronto, Ontario, Canada (3) ASU, Phoenix, Arizona, United States

Predicting the relationship between a molecule's structure and its odor remains a difficult, decades-old task. This problem, termed structure-odor relationship (SOR) modeling, is an important challenge in chemistry, impacting human nutrition, manufacture of synthetic fragrance, the environment, and sensory neuroscience. We propose the use of graph neural networks for SOR, and show they significantly out-perform prior methods on a novel data set labeled by olfactory experts. Additional analysis shows that the learned embeddings from graph neural networks capture a meaningful odor space representation of the underlying relationship between structure and odor, as demonstrated by strong performance on two challenging transfer learning tasks. Machine learning has already had a large impact on the senses of sight and sound. Based on these early results with graph neural networks for molecular properties, we hope machine learning can eventually do for olfaction what it has already done for vision and hearing.

CINF 51

R&D digitalization and data intelligence strategies for flavors and fragrances applications

Max Petersen, max.petersen@dotmatics.com. Dotmatics, San Diego, California, United States

R&D digitalization has promise to greatly accelerate innovation rates. Especially data analytics and data-driven approaches to supplement traditional lab experimentation are attracting lots of interest. A major obstacle to this vision is the general lack of availability of data in a curated and contextualized form. This is despite many organizations investing into technologies to digitalize experimental data capture and move traditional paper-based lab notebooks to their digital counterparts.

Data intelligence fills the gap between R&D IT infrastructure designed for data capture and the scientist’s need to extract these data in a meaningful way. This talk covers an approach that combines data federation, contextualization and out-of-the-box scientific domain models to solve various complex data extraction use cases.
Communicating computation: an automated, reproducible workflow for computational organic chemistry studies and their analysis

Juan Vicente Alegre Requena, juanvi89@hotmail.com, Shree Sowndarya, Robert S. Paton. Chemistry, Colorado State University, Fort Collins, Colorado, United States

Computational studies should be carried out in reproducible and transparent ways that allow other scientists to obtain the same results easily. However, modern workflows use multiple software programs and heterogeneous file types and require manual intervention at multiple steps. The standard reporting formats of journal articles and supporting information make replication and error-checking of complex computational investigations hard to perform. In this work, we present a Python-based workflow that has been designed to make the collection, analysis and publication of computational chemistry as transparent as possible. Our approach, DBGEN, links conformational analysis with standard toolkits (such as RDKit, xTB or ANI1) starting from easy-to-make inputs (SMILES strings, XYZ, COM/GJF and SDF files, etc) to quantum chemistry software and eventual postprocessing to generate figures and tables. The resulting output files are automatically analyzed by DBGEN and resubmitted if they have imaginary frequencies or they did not end normally. When this iterative process to solve errors finishes, GoodVibes is used to post-process the output files and generate thermochemistry data and extract molecular coordinates ready to be inserted into a supporting information document. Additionally, this software also creates publication-quality figures with energy profiles including multiple reaction pathways.

Overcoming complications in informatics for materials and d

Pillhun Son, pillhun@gmail.com. Solutions Engineering, CAS, Columbus, Ohio, United States

Chemical Informatics has been contributing more and more into various fields of research, and in the past few years there were great strides in assisting innovation in materials and devices. In terms of Informatics and the techniques, content and models that are being used, there are unique challenges for Informatics specifically in materials and devices. We will go over the main tactics that could be applied to overcome these challenges in the perspective of incorporating internal data for training sets and also in incorporating external content to evolve the predictive models. We will also delve into strategies in overcoming the unique challenges in this space moving forward and how Informatics could play a stronger role in the development of new materials and devices.

Exploring steric activity and molecular descriptors for building statistical models

Guilian Luchini, Guilian.Luchini@colostate.edu, Robert S. Paton. Chemistry, Colorado State University, Fort Collins, Colorado, United States

Quantifying chemical space and the differences between molecular characteristics is key to the development of meaningful and predictive structure-reactivity relationships. These descriptors capture the spatial, electronic, or energetic features of molecules, and may be
obtained in numerous ways, from empirical rules to full quantum mechanical treatments. Open source and commercial tools exist for this purpose.

Traditionally, molecular size is captured by Taft and Charton steric parameters. These descriptors attempt to reduce molecular shape (which formally depends upon the electron density at each point in space) to a singular value. The multidimensional Sterimol parameter uses three values along different axes to account for length, minimum and maximum widths of molecules. While this multidimensional family of parameters does provide more information on molecular shape, the Sterimol parameter set fails to account for changes in molecular shape along a specified length, such as proximal or distal steric effects.

Inspired by Grid-QSAR techniques, I have worked to develop ways to describe steric behavior more holistically using molecular slices for improved descriptions of contours along a molecule’s length. With these improved parameters used in conjunction with other chemical descriptors, improved relationships between molecular properties can be made with molecular activity descriptions such as experimental yield, selectivity or reaction rate. In addition to quantitative relationships, these new descriptors aid in mechanistic interpretation based on the significance of the position of steric influence, indicating the presence of proximal or remote interactions in the system under study.

CINF 55
Withdrawn

CINF 56
Molecular modeling expeditions in the chemical space

Stefano Forli, forli@scripps.edu. Integrative Computational and Structural Biology, The Scripps Research Institute, La Jolla, California, United States

Despite the vastness of the chemical space, chemical diversity of molecules screened for a given biological activity is still relatively low. The reasons for this are a combination of modeling methodological limitations, and the tendency of commercial vendors to rely on well-established synthetic pathways that are easy to automate and scale up. On one hand, it has been demonstrated that it is still possible to find very active molecules while searching in these somehow bounded and limited regions of the chemical space. On the other, exploring new uncharted chemical territories will likely provide molecules with new and interesting combinations of physico-chemical properties. We will present an overview of potential promising sources, which we explored for identifying new chemical matter, and the methodological improvements we implemented in the AutoDock Suite to overcome some of the modeling limitations associated with such new chemotypes. Examples include the use of high-throughput derivatization using new click-chemistry approaches; enumeration of large libraries of synthesizable macrocycles and functionalized natural compound scaffolds; the generation of covalent binders using a variety of warheads. We will discuss the methods for the libraries preparation, the modeling of the macrocycle flexibility for docking, and the application of the Reactive Docking protocol for high-throughput screening of covalent binders.

CINF 57
Journey from Hits to Leads and Clinical Candidates: A Modern Industry Perspective

Dean G. Brown, deanbrown0506@gmail.com. Jnana Therapeutics, Boston, Massachusetts, United States

One of the most critical decisions that a drug discovery team will face for any new project is how to identify chemical starting points. A variety of approaches exist such as high-throughput screening, fragment-based lead generation, DNA-encoded library (DEL) screening, virtual screening and fast-follower methods. An analysis has been done on 66 recently published success stories of clinical candidates along with the source and properties of the original hit scaffold. Additionally, a review of FDA approved drugs from 2010-2019 will be discussed in the context of emerging chemical modalities, therapeutic areas, and how this will likely impact the future of hit-to-lead discovery.

CINF 58

Quality matters in chemotype discovery

Johanna M. Jansen¹, johanna.jansen@novartis.com, Charles A. Wartchow², Eric J. Martin³. (1) Novartis, Emeryville, California, United States (2) Novartis, Lafayette, California, United States (3) Computational Chemistry, Novartis, El Cerrito, California, United States

The ability to discover high quality chemotypes in the hit generation stages of our programs drives the ability to deliver high quality drugs to the clinic. At the Novartis Institutes for BioMedical Research, our hit generation strategies are tailored to each project to address multi-objective quality criteria. Experimental and virtual screening strategies are designed to effectively explore relevant chemical space.

For projects where there are no applicable virtual screening models, we make deliberate choices on compound sets for experimental screening. This includes consideration of screening our whole archive, carefully designed diversity subsets of the archive, fragment collections, or a screening set named “Scope” where we apply newly defined “Fragment-Plus” principles, which will be described in this talk.

Chemotype discovery is especially effective when experimental screening is informed by computational approaches that leverage all available knowledge and that can complement that knowledge with a “biased diversity” component. We leverage our extensive knowledgebase through our profile-QSAR multitask machine-learning platform and incorporate experiences from many successful structure- and ligand-based lead optimization programs to articulate quality criteria and project-specific biases. Several examples and our general workflow will be presented.

It is important to recognize that an initial hit-list, whether experimental or virtual, typically undergoes a process of triaging, through which the team focuses their resources on those hits that are deemed most likely to succeed. In cases where a project needs to differentiate from known chemotypes, this triaging will try to bias away from that known chemical space. There are several approaches to assess whether a chemotype is novel compared to known chemotypes. A hit generation case study will be presented that highlights this challenge for a program that ultimately delivered a compound that progressed into human clinical trials.

Chemical Information Bulletin, 2020, 72 (3)
Advancing FAIR: Moving toward the Internet of FAIR data and services

Melissa Cragin, mcragin@sdsc.edu, Christine Kirkpatrick. San Diego Supercomputing Center, University of California, San Diego, San Diego, California, United States

Improving transparency of the research process and access to results of publicly funded research can increase trust in the scientific enterprise and facilitate the reuse of data. The FAIR data principles guide foundational actions needed to make data and metadata “Findable, Accessible, Interoperable, and Reusable.” Production of data that are FAIR facilitates scientific discovery, enhancing societal and economic benefit of investments in research and development (R&D).

GO (Global Open) FAIR is a ‘bottom up’ initiative driven by people and organizations from across the globe dedicated to implementing the FAIR principles. GO FAIR works through “Implementation Networks” to change culture toward open science, provide training, and build new tools and data resources. One of several regional coordination entities, the U.S.
FAIR office at the San Diego Supercomputer Center (SDSC) is working to build national capacity around FAIR data, and to increase the availability of machine-readable data and metadata for use in R&D and education.

The U.S. Office is leading efforts to develop a national pool of experts who will available for training activities and data “FAIRification” events. Partnering with GO FAIR leaders in the EU, we delivered the first GO FAIR Data Stewardship training session in the U.S., and in early 2020 held professional development workshop. In April, we produced a webinar series on COVID-19, as part of GO FAIR’s new VODAN (Virus Outbreak Data Network) Implementation Network.

This talk will introduce the FAIR Data Principles and the international GO FAIR initiative, and report on FAIR activities in the U.S.

CINF 60

FAIR IUPAC: Advancing pure and applied chemistry worldwide

Leah R. McEwen\textsuperscript{1,3}, lrm1@cornell.edu, Ian Bruno\textsuperscript{2,3}. (1) Clark Library, Cornell University, Ithaca, New York, United States (2) Cambridge Crystallographic Data Centre, Cambridge, United Kingdom (3) Committee on Publications and Cheminformatics Data Standards, International Union of Pure and Applied Chemistry (IUPAC), Durham, North Carolina, United States

Data and metadata standards for communicating scientific concepts consistently and accurately in the digital context are critical for applying the criteria articulated in the FAIR guidelines, particularly at the domain level. The International Union of Pure and Applied Chemistry (IUPAC) supports scientific communication with a century of experience in systematically defining terms and criteria for describing chemical measurements and data. IUPAC expert groups engage in nuanced meta-analyses of experimental data and best quality data are reported in standard layouts that encapsulate the rules for evaluation based on contextual information about the measurement such as method of determination, estimates of error and assessment of relevant experimental conditions. IUPAC is undertaking several projects to formalize standard metadata profiles to translate these expressions into rules for machine processing.

Availability of recommended data values, nomenclature and terminology enables more accurate models for chemical systems. Breaking down this challenge into discrete interoperable functionalities that are essential for accurate digital expression will show-case the application of IUPAC assets to global problems in digital science and enable broader utility. IUPAC has launched a FAIR Data initiative to assess the role and suitability of IUPAC outputs to support FAIRer exchange of Chemical Data. This presentation will highlight digital standard activities in several areas of Chemistry, including spectroscopy, solubility, isotopic abundance, and molecular representation and report on the progress to establish FAIR-related guidance for the wider community.

CINF 61

PubChem COVID-19 response: Opportunities and challenges in delivering timely research data
Providing timely update of content is important to provide scientists with the most current research data. Now add in an imperative that involves a topic that is, for all intents and purposes, brand new and of critical importance. In many ways, this is what PubChem faced with COVID-19 as the disease spread around the globe. Seemingly everyone wanted to know everything possible about SARS-CoV-2 virus and COVID-19 disease, yet there was very little if anything to show. The protein sequences were only just being made available and research data was being fast tracked for immediate delivery within generalist repositories, preprint servers, along with repositories designed for the data, such as GenBank, the Protein DataBank (PDB), and beyond. This talk will give a brief overview of PubChem efforts to provide COVID-19 related content and ponder the opportunities and challenges in delivering timely research data.

CINF 62

GO-FAIR chemistry implementation network (ChIN)

Stuart J. Chalk1, schalk@unf.edu, Simon J. Coles2, Jeremy G. Frey2, Egon L. Willighagen3. (1) Department of Chemistry, University of North Florida, Jacksonville, Florida, United States (2) University of Southampton, Southampton, United Kingdom (3) Maastricht University, Maastricht, Netherlands

The Chemistry Implementation Network (ChIN) comes out of the GO FAIR project, where there are a range of networks covering both technical/political and discipline engagement aspects of FAIR. The principles and objectives of the ChIN have been endorsed by IUPAC and the two will work together, along with Research Data Alliance led initiatives, to embed FAIR principles into chemistry practice. This talk will present the current activities of the ChIN and seed ideas about additional needs for FAIR related standards, tools and services for chemical data.

CINF 63

Open conversation about FAIR in Chemistry

Ian Bruno1, bruno@ccdc.cam.ac.uk, Stuart J. Chalk2, schalk@unf.edu, Leah R. McEwen3, lrn1@cornell.edu, Nicholas Ruhs4, nruhs@fsu.edu. (1) Cambridge Crystallographic Data Centre, Cambridge, United Kingdom (2) Department of Chemistry, University of North Florida, Jacksonville, Florida, United States (3) Clark Library, Cornell University, Ithaca, New York, United States (4) Florida State University, Tallahassee, Florida, United States

In this session we intend to conduct a mediated workshop with the aim of exploring current projects and practices that have the potential to support FAIR Chemistry. We will look specifically at the building blocks we have - and those we need - in order to move forward and reflect on what individuals can do to make a difference. We will conduct an open community discussion with the goal of developing a snapshot of the status of FAIR in chemistry as well as identification of action items that will make Chemistry FAIRer.