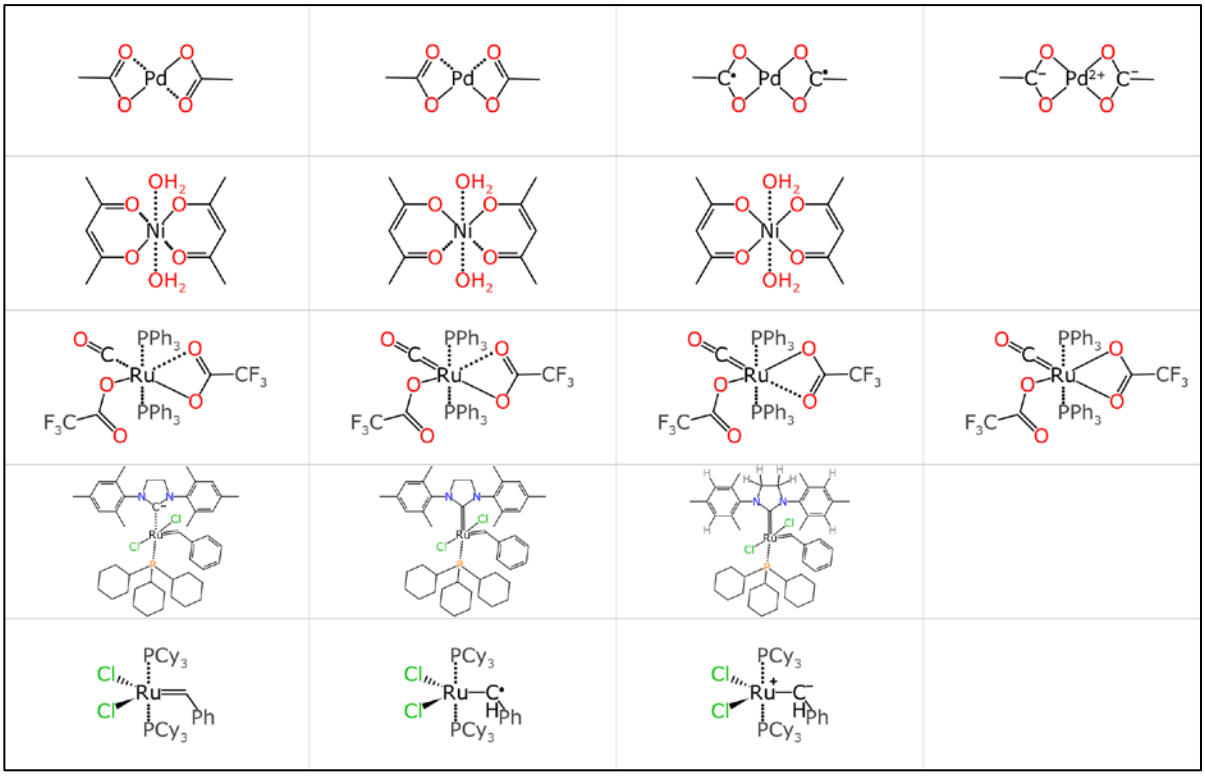
RSC INTEREST GROUP
CHEMICAL INFORMATION AND
COMPUTER APPLICATIONS GROUP

NEWSLETTER winter 2019-20



Above: A representation of some coordination complexes - see '*Extending the InChI Identifier to Include Inorganic and Organometallic Compounds*' on page 14

CICAG aims to keep its members abreast of the latest activities, services, and developments in all aspects of chemical information, from generation through to archiving, and in the computer applications used in this rapidly changing area through meetings, newsletters and professional networking.

Chemical Information & Computer Applications Group Websites:

http://www.rsccicag.org
http://www.rsc.org/CICAG

LinkedIn http://www.linkedin.com/groups?gid=1989945

MyRSC http://my.rsc.org/groups/cicag

https://twitter.com/RSC_CICAG

QR Code

Table of Contents

Chemical Information & Computer Applications Group Chair's Report 3
CICAG Planned and Proposed Future Meetings 4
Periodic Table brought to life at Catalyst! 5
From Nantwich to Oxygen: Joseph Priestley's Journey of Discovery 6
Dr Andy Vinter 1943 - 2019 8
Open Force Field Initiative on Energetics for Molecular Modelling 9
Dr William Geoffrey Town 12
Extending InChI to Include Inorganic & Organometallic Compounds 14
Wendy Warr named 2020 Herman Skolnik Award Winner 17
A Practical Case for WebAssembly 18
Tony Kent Strix Award and Annual Lecture 2019 20
Meeting Report: 20 Years of the Rule of Five 21
Meeting Report: 2nd RSC BMCS & CICAG AI in Chemistry Meeting 28
Chemical Information / Cheminformatics and Related Books 30
News from CAS 31
News from the RSC 34
News from AI3SD 34
Other Chemical Information Related News 35

Contributions to the CICAG Newsletter are welcome from all sources - please send to the Newsletter Editor: Stuart Newbold, FRSC, email: stuart@psandim.com

Chemical Information & Computer Applications Group Chair's Report

Contributed by RSC CICAG Chair Dr Chris Swain, email: swain@mac.com

Whilst RSC members can join up to 3 interest groups for free, in practice many members do not take up this opportunity. In an effort to highlight this, CICAG is again running an advertising campaign to coincide with annual subscription renewals at the start of the year. Whilst the renewal letter is the traditional time for members to join interest groups, RSC members can make a request to join a group via email (membership@rsc.org) or telephone (01223 432141).

Social media is becoming an increasingly active way for communicating with members, with both Twitter with its nearly 1000 followers and LinkedIn with 400 followers gaining in popularity. Importantly, the social media feeds provide an opportunity for communication with both RSC and non-RSC members around the world. Our CICAG website is now active and we would be very interested to hear suggestions for additional content.

The April meeting "Workshop on Computational Tools for Drug Discovery" in Birmingham, co-organised with the SCI Fine Chemicals Group, was a great success. The event ran at full capacity and the feedback from both delegates and software vendors has been excellent. Planning for the next meeting is already underway, and it will be taking place on 19 May 2020 at the Studio, Leeds .

The 2nd RSC-CICAG/RSC-BMCS Artificial Intelligence in Chemistry Meeting was held at Fitzwilliam College in Cambridge, on 2-3 September. We had a really outstanding line-up of speakers from around the world and the meeting was a great success. The feedback both during the meeting and from the feedback forms was very positive. Many delegates enjoyed the "Gordon Conference" atmosphere of the meeting. The event also had a great online following with the Twitter hashtag #AIChem19 being used 600 times. There were three poster prize winners:

- Ya Chen, (University of Hamburg, Germany)
  NP-Scout: Machine learning approach for the identification of natural products and natural product-like compounds in large molecular databases
- Adam Green, (University of Leeds, UK)
  Activity-directed discovery of inhibitors of the p53/MDM2 interaction: towards autonomous functional molecule discovery
- Jenke Scheen, (The University of Edinburgh, UK)
  Improving the accuracy of alchemical free energy methods by learning correction terms for binding energy estimates

Planning for the 3rd RSC-CICAG/RSC-BMCS Artificial Intelligence in Chemistry Meeting has already started (28-29 September 2020, Churchill College, Cambridge, UK) so please watch this space.

Another joint meeting between RSC-CICAG and RSC-BMCS was to acknowledge 20 years of the Rule of 5, and this was also very well received. It was held on Wednesday, 20 November 2019, at Sygnature Discovery, BioCity, Nottingham, UK, and was generously sponsored by Sygnature. The meeting had been preceded by a series of online polls regarding the use of metrics in Drug Discovery, the results of which were then shared at the meeting. These, together with the talks and panel discussions certainly provoked much animated debate, and there was enthusiastic support for holding another meeting in this format. A meeting report is on page 21.

CICAG co-sponsored the 'In Silico Toxicology' Network Meeting 2019' in Cambridge alongside the British Toxicology Society (BTS), the Cambridge Alliance on Medicines Safety (CAMS), and Lhasa Ltd. This

inaugural event had 100 attendees and 10 scientific presentations. Drug and Food safety is an increasingly important area of science, and with the move away from animal testing, in-silico predictions will be of an increasing focus. This is clearly an area of importance for cheminformatics, and we hope to organise more joint meetings in the future.

We have just started planning a 3 day event on Open Chemical Science, 11-13 November 2020, at Burlington House, London, UK, which will explore the benefits, risks and likely future developments which will enable the opportunities provided by open chemistry to be maximised.

CICAG is committed to supporting attendance at our meetings in as many ways as possible, and we always ensure that the venue supports wheelchair access and that any meals accommodate any dietary requirements that have been gathered on registration. We offer bursaries to help cover registration, travel, and accommodation if required, and together with partners CICAG has already provided 19 bursaries this year. There are also travel grants, assistance grants, and support for carers, and if members need any additional help please contact the conference organisers.

This newsletter also includes articles from Prof David Mobley, Dr Wendy Warr, Dr Alex M Clark, Jane List, Dr Rajarshi Guha and Dr Dan Ormsby, and CICAG is very appreciative of these external contributions to this widely-read publication. If readers have any suggestions for contributions that would be of interest to the CICAG community, please get in touch with myself or Stuart.

Finally, to streamline the editorial process CICAG committee has decided that the newsletter will now revert to its previous spring- and autumn-publication timetable. Look out for the next issue once summer holidays are completed in September.

CICAG Planned and Proposed Future Meetings

The table below provides a summary of CICAG's planned and proposed future scientific and educational meetings. For more information, please contact CICAG's Chair, Dr Chris Swain.

Meeting	Date	Location	Further Information
Workshop on Computational Tools for Drug Discovery	19 May 2020	The Studio, Leeds	SCI Events page
3rd Artificial Intelligence in Chemistry Meeting	28-29 Sep 2020	Churchill College, Cambridge	Joint event from RSC-CICAG and RSC-BMCS division
Open Chemical Science event	11-13 Nov 2020	Burlington House, London	Three 1-Day events, exploring sharing and collaborative developments
Open Source Software for Medicinal Chemists	TBD	EBI, Hinxton, Cambridge	Proposed training workshop (Joint workshop with the SCI)
Big Data	TBD	TBD	Proposed joint meeting with the SCI)

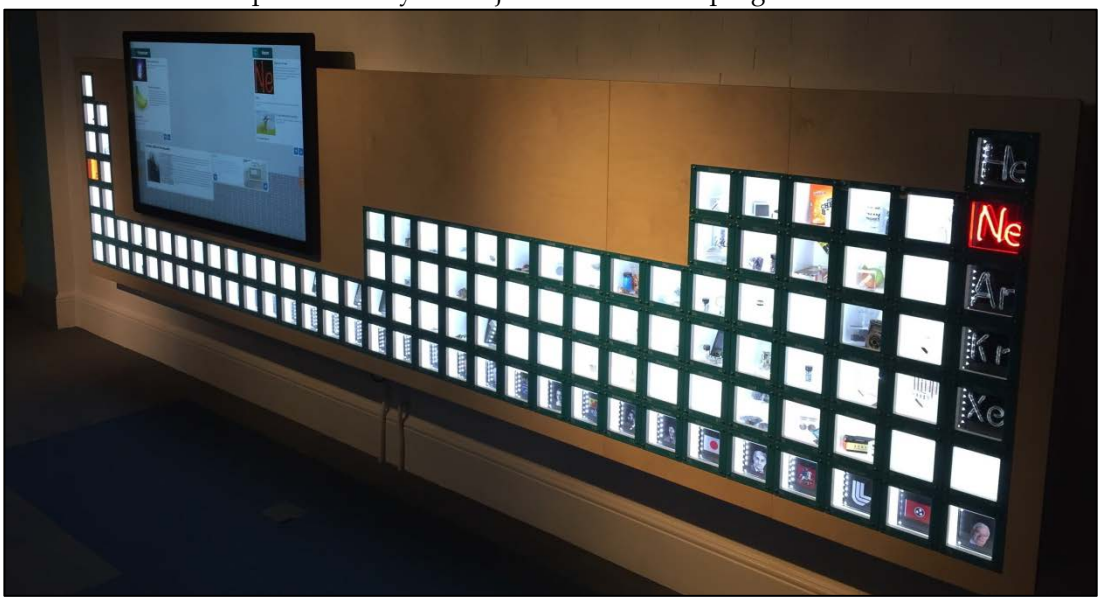
****

Two CICAG Committee members, Dr Helen Cooke and Dr Diana Leitch, are very grateful for funding from the special RSC IYPT Outreach Fund during 2019; the funds enabled organisation of events at museums in Cheshire where they are Trustees of Nantwich Museum and the Catalyst Science Discovery Centre respectively, to celebrate the International Year of the Periodic Table (IYPT 2019). In both cases this money was supplemented by financial support from CICAG (and in the case of Nantwich Museum also from the North Staffordshire Local Section).

Periodic Table brought to life at Catalyst!

Contribution from CICAG Committee member Dr Helen Cooke, email: helen.cooke100@gmail.com

A new and unique interactive Periodic Table has been installed at the Catalyst Science Discovery Centre in Widnes, Cheshire, funded by an RSC International Year of the Periodic Table grant and by CICAG. It is the first exhibit to be installed as part of Catalyst's major refurbishment programme.



Catalyst's Periodic Table's official launch took place on 22 October 2019. Professor Neil Berry and Dr Helen Cooke represented CICAG, and invited guests included Catalyst trustees and patrons, pupils from local schools and representatives from Runcorn WI Ferry Girls who contributed artwork for a special Periodic Table exhibition. It was unveiled by Professor David Phillips CBE, FRS, Past President of the RSC and Patron of Catalyst, who said: "*Catalyst is an inspiration. This is exactly the type of thing that we, the Royal Society of Chemistry should be supporting. It demonstrates the utility of chemistry for the world and it is in the site where so much of the chemical industry was founded and still continues today. It is a very, very special place.*"

Chair of the Catalyst Trustees, Dr Diana Leitch said "*The official opening of the new interactive Periodic Table exhibit was the culmination of a dream for Catalyst to be the first place in the UK to have such an exhibit and entirely appropriate for it to be based in the only science discovery centre in the UK specialising in the chemical sciences and also the only museum of the chemical industry in the UK. For it to be sited in a building which already existed as part of Gossages great soap works in 1869, when Mendeleev worked out the layout of the Periodic Table, is an added bonus. We cannot thank the Royal Society of Chemistry enough for co-funding this new exhibit which is part of our Inspiring Science Project funded by the Wellcome Trust and UKRI.*"

Left: Martin Pearson (CEO of Catalyst), Dr Diana Leitch MBE (Chair of Catalyst Trustees), and Professor David Phillips CBE, FRS (RSC)

Catalyst CEO, Martin Pearson said "*This is a really exciting time for Catalyst, the Inspiring Science Fund project is enabling us to redevelop some of the physical spaces within the building and also expand our audience to work with more groups in the local community. The interactive Periodic Table is the first of 25 new exhibits that Catalyst will take delivery of over the next six months…we are very much looking forward to a positive future as this transformation progresses.*"

The Periodic Table exhibit will be a permanent feature at Catalyst and at the thoroughly enjoyable launch event was already proving popular amongst visitors, young and old. It will clearly be an excellent mechanism for communicating chemical science and engaging people – an area CICAG is proud to be involved with.

More information about the Catalyst Science Discovery Centre can be obtained from Meryl Jameson, Marketing Manager, on 0151 420 1121, or via email at meryl@catalyst.org.uk.

Catalyst can also be found on Facebook: https://www.facebook.com/CatalystScienceDiscoveryCentre/ and Twitter: @CatalystSDC.

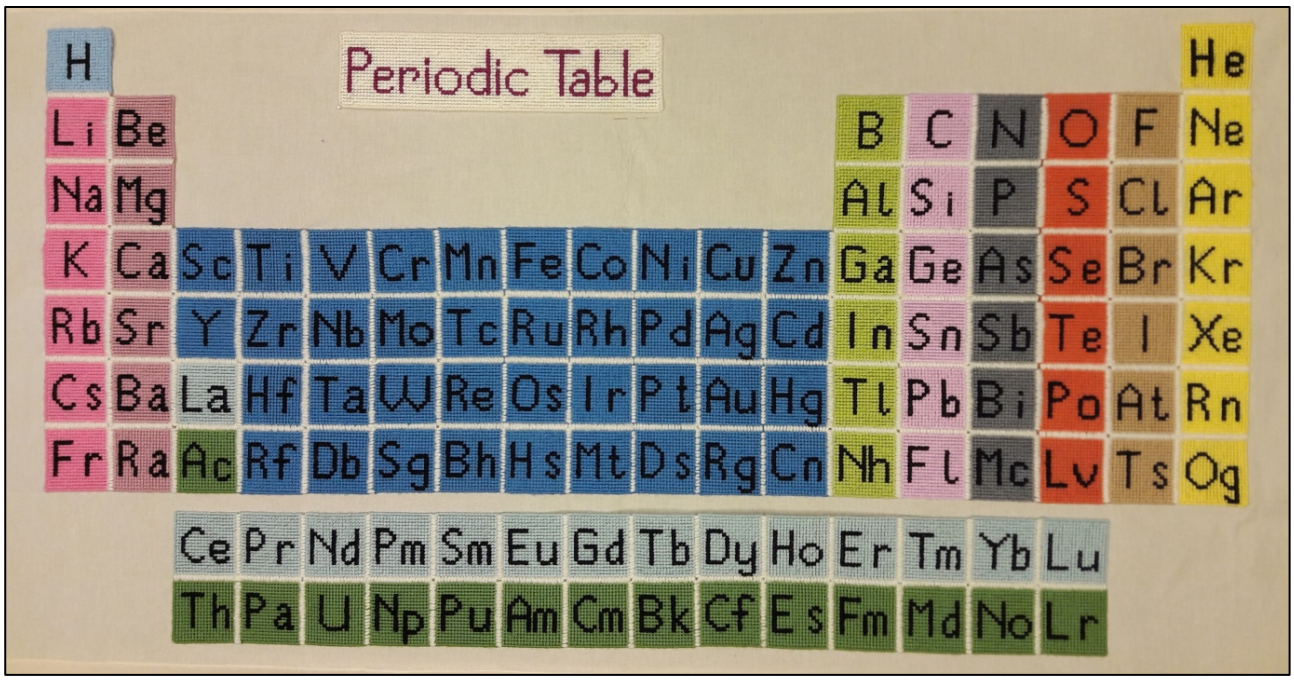
From Nantwich to Oxygen: Joseph Priestley's Journey of Discovery

Contribution from CICAG Committee member Dr Helen Cooke, email: helen.cooke100@gmail.com

In CICAG's summer 2019 Newsletter I reported on preparations for Nantwich Museum's "*From Nantwich to Oxygen: Joseph Priestley's Journey of Discovery*" exhibition which took place from 14 August – 26 October 2019. The exhibition successfully raised Priestley's profile and, at the same time, celebrated the Periodic Table's 150th anniversary.

During his time in Nantwich Priestley laid the foundations for his future scientific work and achievements. He founded a school, where he purchased some "philosophical instruments" (which today we'd call scientific apparatus) and used these to teach his older students and entertain their parents and friends with his experiments.

Nantwich Museum's Research Group developed information panels which introduced Museum visitors to Priestley's life and achievements before, during and after Nantwich, culminating in his oxygen discovery which formed the link to the Periodic Table. A number of exhibits were loaned to the Museum for the exhibition, including two pairs of Priestley's spectacles, some of his written works and an embroidered Periodic Table (handmade by the Museum's Craft Group).



Above: Tapestry Periodic Table embroidered by Nantwich Museum's Craft Group

For children, throughout the exhibition there were books, games and quizzes. Also, an element hunt around the Museum developed by Heidi Dobbs, RSC Midlands education coordinator, enabled visitors young and old to learn about elements present in the Museum's historical exhibits. Heidi, supported by a team of volunteers from the RSC's North Staffordshire Local Section, also led a drop-in family workshop, "Priestley's Element", featuring hands-on experiments. Over 70 children and adults learnt about carbon dioxide, the properties of metals, spectroscopy and more.

A celebratory reception was held with entertainment ranging from flute music (Priestley learnt to play the flute while in Nantwich) to live chemistry. The keynote speaker was Professor Mark Ormerod, Deputy Vice Chancellor and Provost at Keele University. A demonstration by Dr Katherine Haxton (also of Keele) featured the decomposition of hydrogen peroxide, producing "elephant's toothpaste". Community events included a Periodic Table themed coffee morning and "Joseph Priestley's Footsteps" walking tours around Nantwich.

A sold-out programme of talks, "Joseph Priestley in his Element", featured Priestley's time in Nantwich (Helen Cooke), his alliance with Josiah Wedgwood and Thomas Bentley (Gaye Blake-Roberts), his membership of the Lunar Society (Cameron Arthur), and exciting demonstrations bringing Priestley's work to life (Fabio Parmeggiani). As well as Friends of Nantwich Museum and local residents, the event attracted RSC members from as far away as Cumbria and Hereford.

Key to the exhibition's success was ensuring that local history (the main interest of the Museum's visitors), as well as science, was a theme. The range of activities for all age groups engaged a diversity of visitors and the partnership with the RSC enabled the Museum to reach a broad audience. A Museum member commented: "*At last we've had a proper tribute to Joseph Priestley in Nantwich*".

Most of the information panels are available online via the Museum's website. The exhibition will also be featured in the RSC's *Voice* and is available on the RSC website News and Events page.

The exhibition would not have been possible without the financial support received from CICAG, the North Staffordshire Local Section, and an IYPT grant. The exhibition was also generously supported by Keele University, the University of Manchester, RSC Historical Chemistry and Environmental Chemistry Interest Groups, RSC Library, Priestley House Museum (Pennsylvania, USA), Harris Manchester College Oxford, Warrington Archives and Wedgwood Museum Stoke-on-Trent.

Above: "Joseph Priestley in his element" speakers: (left to right) Dr Fabio Parmeggiani (University of Manchester), Cameron Arthur (Soho House Museum, Birmingham), Professor Dame Janet Finch (event Chair), Gaye Blake-Roberts (Wedgwood Museum, Stoke-on-Trent), Dr Helen Cooke (Nantwich Museum)

Dr Andy Vinter 1943 - 2019

Contribution from Dr Rob Scoffin (rob@cresset-group.com)

It is with sadness that I share the news that Dr Andy Vinter passed away on Monday 16th December 2019 after a short illness. Andy was a great scientist, colleague and friend and will be hugely missed by all those who met him and whose lives were enriched by his warmth, passion for life and scientific integrity.

One of Andy's favourite quotes was from Hamlet:

"*This above all: to thine own self be true*", and I think it is very fitting to apply it to Andy himself, and especially how the quote continues: "*And it must follow, as the night the day, Thou canst not then be false to any man*." We will miss him greatly and remember him fondly. Our thoughts and prayers are with his wife, Elaine and his children, Francesca and James at this very sad time.

Andy was a founder member of the Molecular Graphics and Modelling Society (MGMS) in 1981 and served as the first chairman and remained on the committee for several years.

Open Force Field: A new, open Initiative Focusing on Energetics for Molecular Modelling

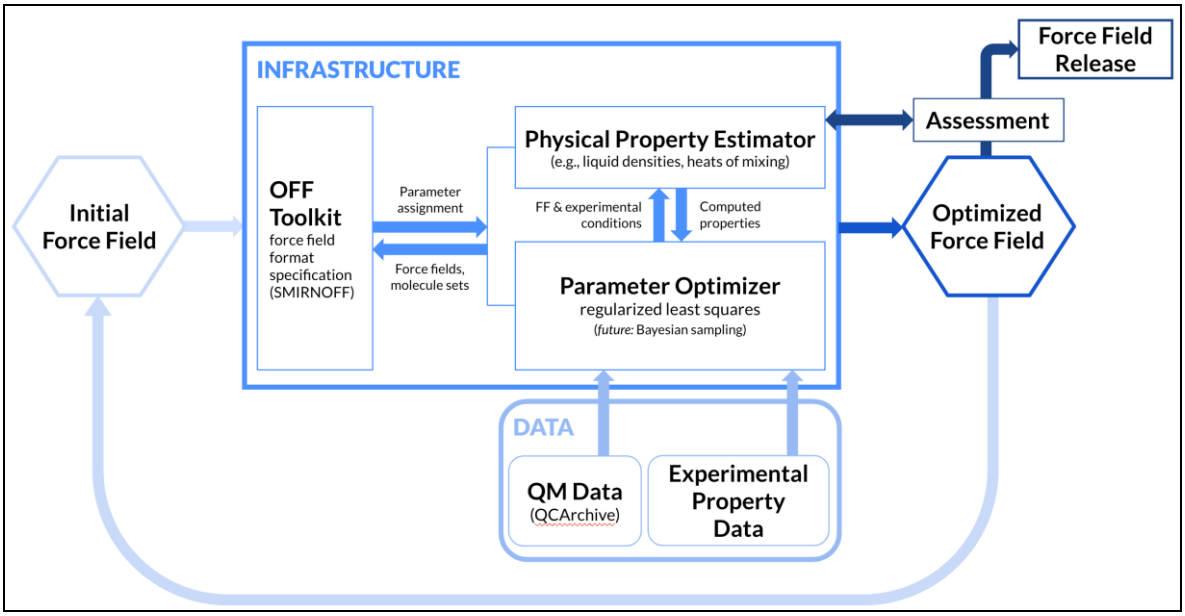
Contribution from David Mobley, Associate Professor, University of California Irvine, email: dmobley@uci.edu

Molecular modelling is widely used in diverse pharmaceutical discovery applications, but its utility and predictive power is limited by the accuracy of the underlying molecular mechanics force field used to compute the energetics of biomolecular systems. The Open Force Field (OpenFF) initiative focuses building iteratively more accurate force fields to improve predictive design, along with the necessary infrastructure to make these force fields easier to build and use. In October, the initiative produced its first major new force field, OpenFF-1.0 (codenamed Parsley), which is freely available to the community. The initiative is supported in part by the OpenFF Consortium, a pre-competitive industry alliance.

Why a new force field? Current molecular mechanics force fields are built on early modelling work largely performed in the 1980s and 1990s, and the software infrastructure for developing, refining, and applying them has remained largely unchanged since. While molecular modelling based on these force fields currently plays a key role in the pharmaceutical discovery pipeline, improved accuracy is needed to more effectively guide pharmaceutical discovery and design and to reduce costs and time to market. Also a modern software infrastructure is needed to allow these force fields to be rapidly built, updated and applied. The OpenFF effort breaks free of constraints imposed by legacy force field efforts while maintaining compatibility with all major molecular simulation packages used in today's workflows.

The effort aims to take advantage of improved theoretical methods, modern maintainable software infrastructure and vast increases in machine-readable physical property and quantum chemical data. This will enable development of new, more accurate and easily extensible force fields that can be steadily improved to meet the needs of modern pharmaceutical R&D.

Aims: The OpenFF initiative aims to: (1) engineer a modern, open, sustainable, extensible and well-supported framework for automated force field improvement and application; (2) use this to release rapid iteratively improved versions of an AMBER-compatible small molecule force field developed to take advantage of modern cheminformatics; (3) produce entirely new comprehensive force fields that break free of legacy accuracy limitations while maintaining compatibility with existing simulation software, providing dramatically improved accuracy for modelling predictions in diverse applications ranging from predictions of binding affinity, selectivity and drug resistance, to partitioning, solubility, kinetics and other properties; and (4) work closely with industry partners to ensure the development path follows that most relevant to R&D needs.

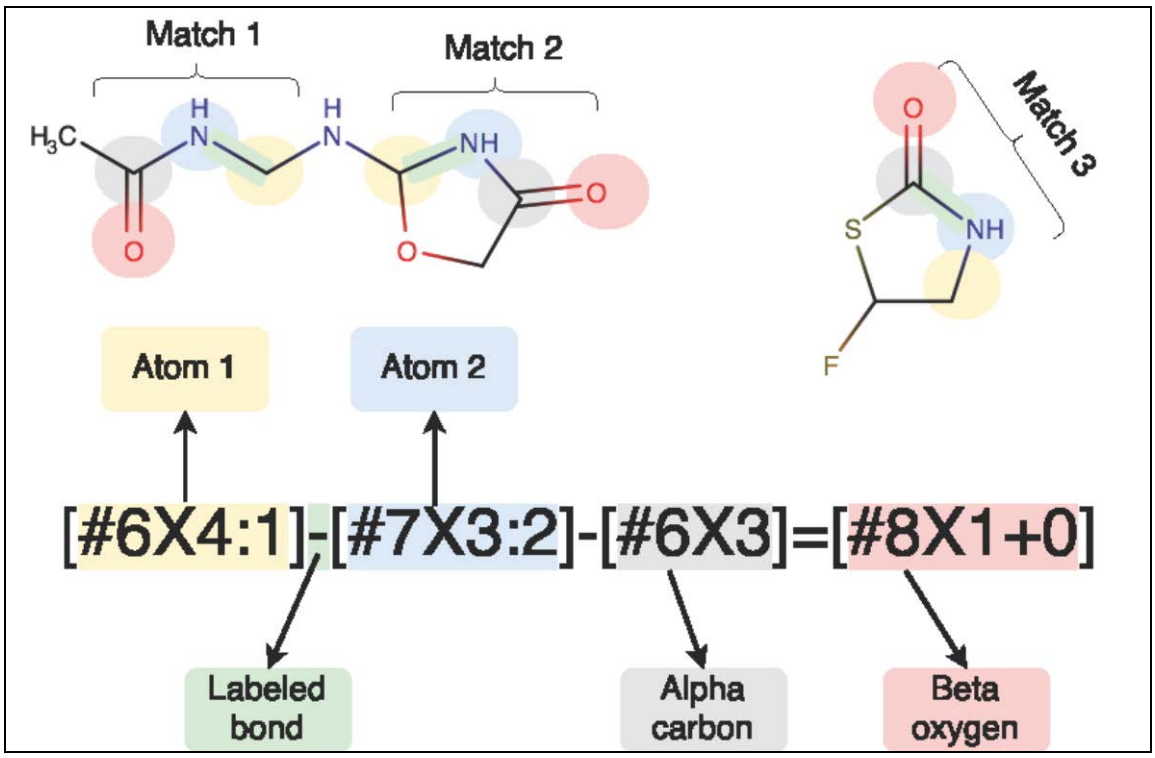


Above: OpenFF's efforts place an emphasis on automated infrastructure for force field refitting. The process begins with an initial force field, left, and the OFF Toolkit uses this to assign parameters. The initiative's property estimator provides a general framework for computing condensed phase properties (e.g. densities, heats of mixing, dielectric constants and others) for use in force field optimization, and the automated infrastructure can also draw on quantum chemical data (via MolSSI's QCArchive project). Our parameter optimizer takes an input force field, and a choice of data (quantum mechanical and experimental) and automatically optimizes the input force field to improve agreement with this data. The resulting optimized force field is then assessed, and if adequately improved, it is released. Otherwise, it can be the basis of further optimizations. This infrastructure is extensible to various types of force fields and functional forms, and can be used to assist in developing force fields both inside and outside the formal initiative.

Open source, open data, open science: All software, code, data and force fields are open and freely available under Open Source initiative and Creative Commons approved licenses, providing a foundation for further science beyond the scope and timescale of the formal initiative.

Industry partners will be enabled to extend the force fields by adding proprietary data using open tools developed as part of this initiative, or build their own workflows based on these tools providing a pathway to sustainability.

The initiative's Parsley release: The initiative recently announced the release of "Parsley", the Open Force Field 1.0 small molecule force field - the first in a series of iteratively-improved small molecule force fields for biomolecular simulation funded in part by the Open Force Field Consortium. This is the first optimized force field to use the SMIRNOFF force field specification for atom type-free direct chemical perception, and provides substantially improved valence (bond, angle, and torsion) parameters relative to its predecessor, the AMBER-lineage SMIRNOFF99Frosst. This force field was optimized to improve agreement with quantum chemical geometries, energetics, and vibrational frequencies, and will likely provide improved accuracy (relative to its predecessor) for a wide variety of properties, especially energetics and geometries relative to gas phase quantum chemical calculations. The release was also accompanied by a set of benchmarks comparing energetics to quantum chemical data. In some sense, this represents a proof-of-principle release, representing the first major use of the initiative's automation infrastructure – but there are substantial plans for progress from here.



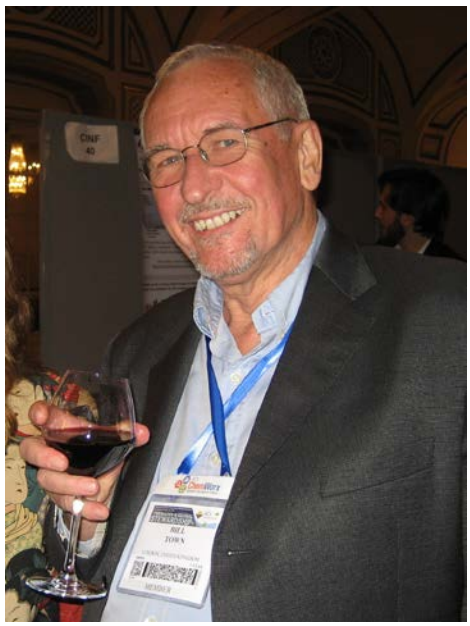
Above: The SMIRKS chemical query language enables direct chemical perception by substructure matching. Here, a SMIRKS pattern recognizes the carbon-nitrogen bond in an amide group, so that a bond-stretch parameter can be assigned to it.
In the examples, the pattern finds three matches (Matches 1, 2 and 3) in in two different molecules, as indicated by colour coding of the atoms and SMIRKS patterns. The relevant pattern is a carbon with four connections ([#6X4:1]) (yellow) single bonded (-) to a trivalent nitrogen ([#7X3:2], blue) which is single-bonded to a trivalent carbon ([#6X3], grey) which itself is double bonded (=) to a neutral oxygen with a single connected atom ([#8X1+0], red). The first carbon and nitrogen in the pattern are singled out for special treatment by having numerical atom labels (:1 and :2) assigned to them, because the SMIRKS pattern in this case is used to assign a bond parameter to the bond connecting the two labeled atoms.

Outlook: The Open Force Field Consortium is an exciting pre-competitive collaboration between academic and industry partners, working to build a shared open infrastructure and ecosystem that will aid discovery.

For further information email David at dmobley@uci.edu or info@openforcefield.org, or visit the website consortium page.

Dr William Geoffrey Town

Contribution by Bill's longtime friend Wendy Wendy A. Warr, email: wendy@warr.com. The article is reproduced here courtesy of Chemical Information Bulletin, where it appears in the winter 2019 edition of CIB. CICAG is very grateful for the opportunity for reproduction.



Bill Town: March 31, 1943 to June 24, 2019

Our dear friend Bill Town passed away suddenly, of a cardiac arrest, on June 24, 2019, at the age of 76. Bill was the son of William Henry Town and Amy Town (née Morton). He grew up in Dagenham, Essex, U.K. He obtained a B.Sc. in chemistry from the University of Birmingham in 1964, and a Ph.D. from the University of Lancaster in 1967. His specialism was crystallography. Later, when he founded his own company, he achieved accreditation from the Chartered Association of Certified Accountants.

In 1967, Bill moved to Sheffield University, where he conducted research on chemical structure searching systems, as part of the pre-eminent research group led by Michael Lynch. In 1968, he moved to the University of Cambridge, where he worked under Olga Kennard in the Cambridge Crystallographic Data Centre, another world-leading group. He was an author on one of the first publications about the Cambridge Structural Database (Kennard, O.; Watson, D. G.; Town, W. G. Cambridge Crystallographic Data Centre. I. Bibliographic file. J. Chem. Doc. 1972, 12 (1), 14-19). CINF members will be interested to hear that Bill was attached to the school of librarianship.

In 1971, Bill moved to Ispra, in Italy, to work for the European Community Joint Research Centre. The family settled on the shores of Lake Maggiore, with views across the lake and the Alps in the distance (Bill loved sunshine), but they also had an apartment across the border in Switzerland. Bill became fluent in Italian: I loved to hear him speak this most musical of languages, but he also spoke French well, he spoke more German than I do, and later in life he also took up Spanish. Bill also got interested in skiing during his years in Italy.

At the Joint Research Centre, Bill led a team building the Environmental Chemical Data and Information Network (ECDIN). The team was also involved in the preparation of the European Inventory of Existing Commercial Chemical Substances (EINECS), which was to be published in seven languages. Bill was always an environmental crusader, but even more so in his seventies when he unfortunately developed chronic obstructive pulmonary disease. He was passionate about trying to reduce air pollution in London.
In 1983, the family returned to the United Kingdom, where Bill and his colleague David Proctor set up Hampden Data Services Ltd. (HDS). The HDS years were the height of Bill's career. It was at HDS that STN Express was developed, and I had the pleasure of leading one of the key industrial teams (at ICI Pharmaceuticals) driving its development. The HDS chemical structure system, PSIDOM, was one of the first chemical structure editors, and an early way of building personal chemical structure databases and using structure entry as a front end to online searching systems.

In 1991, Bill relinquished his interest in Hampden Data Services (which still survives as a part of Chemical Abstracts Service), and founded William Town Associates, which he ran briefly, before joining Derwent Information (later Thomson Corporation) from 1992 to 1997, as Business Development Manager, Scientific Information. In 1997, he became Managing Director at ChemWeb, (later owned by Elsevier), and some readers may remember the Boston tea party reception that ChemWeb sponsored at a Boston ACS meeting. In 2002, Bill set up Kilmorie Consulting. This would later become Kilmorie Clarke, after he met Maggie Clarke and they decided to combine their businesses.

Bill was also a visiting professor in the Department of Information Studies at the University of Sheffield until 2006, and he was Chair of the Board of Governors at the Cambridge Crystallographic Data Centre, where his career had started. He was a member of the American Chemical Society for over 30 years, and he regularly

attended ACS national meetings. He was a member of the CINF Publications Committee from 1993-1996, and chaired the Awards Committee from 2002-2005, and the Nominating Committee in 2001. He served the division as Chair-Elect in 1999, Chair in 2000, and Past-Chair in 2001, and was an alternate councilor from 2006-2008. He was the recipient of the 2008 CINF Meritorious Service Award.

Svetla Baykoucheva reports that Bill played an important role in the transition of the Chemical Information Bulletin (CIB) from print to digital. In 2009, he became chair of the CINF Publications Committee, and Svetla was editor of CIB. They organized the digitization of all print issues of the Bulletin, from its inception in 1949 to summer 2010. Bill secured a grant from ACS to pay for the scanning and the maintenance of the archive. He negotiated the price and conditions for access to the archive with the University of North Texas.

In the mid-1980s, Bill was secretary to what is now the RSC Chemical Information and Computer Applications Group. He and Ian Tarr did all the organizational work involved in setting up the Chemical Structure Association Trust (CSA Trust), and Bill was a signatory to the Declaration of Trust on December 5, 1988. Bill was the first chairman (sic) of the trust.

Many people have spoken about the time and effort Bill put in helping to mentor them, providing them with his quiet influence, which helped them grow with the confidence to progress at the start of their careers. He was a great friend to many in this respect, and was always ready to help, and listen, particularly over a good meal and some fine wine and coffee. He was always full of new ideas and was constantly thinking of new things to do and pushing things in new directions.

In an interview that he gave to the CINF Chemical Information Bulletin in 2009, Bill talked about his passion for exotic travels. He was always keen on travel, and, at the age of only 15, he cycled all the way from London to Cornwall. While at Cambridge he attended a conference in Moscow and drove there with a friend in a hastily purchased and unreliable minivan, at a time when the cold war was in full swing and traveling behind the iron curtain was ill-advised. In 1999, he became an eclipse chaser after witnessing his first total solar eclipse in France. His love of the moment steered him to visit new and interesting places, including Botswana, Spain, Libya, Siberia, and Easter Island, to view more total eclipses. On a trip to Zimbabwe and Zambia, his canoe overturned and pitched him into the Zambezi on the first day, very shortly after he had sighted crocodiles. In 2007, he also toured Thailand and Cambodia, this time not chasing eclipses.

In later life, Bill became a political activist. He became a founding member of the Social Democratic Party (SDP) in 1981. Coincidentally, the announcement of the formation of the SDP was at St. Ermine's Hotel in London, where large numbers of us were gathered for a major conference on chemical nomenclature. Bill followed the evolution of the SDP into the Liberal Democrats, as they are today. He was a passionate European and was committed to the European Union. He became a pivotal part of the Lewisham Liberal Democrats, and gave a lot of his time to serving on committees and helping to deliver leaflets, man stalls or help with IT issues. In 2013, he even stood as a Liberal Democrat candidate for one ward. He was a keen marcher and took part in the many rallies and marches against Brexit. One of Bill's happiest moments was the recent Lib Dem victories in the European elections, in areas which had never previously voted Lib Dem.

I knew Bill for more than forty years. We missed our first opportunity to meet in Sheffield in 1968, when he would have tutored me on a course on computer handling of chemical structural information. For some reason I did not attend; maybe the course was over-subscribed. So, we first met, briefly, at a Chemical Notation Association conference in Kent in 1979. Our friendship blossomed after we both walked at the same time into the lobby of a hotel in Columbus, Ohio, in April 1982. Bill had come from Italy to visit Chemical Abstracts Service (CAS) on EINECS business, and, coincidentally, I was being given a tour of CAS on the same day. Our CAS hosts arranged for us to have lunch together.

Bill spoke about EINECS at the Fall 1982 ACS National Meeting in Kansas City, and I persuaded him to organize a symposium on EINECS at the fall meeting the next year, in Washington, DC. It was there that Bill showed me pictures of the house he was buying in Oxfordshire, England, and told me about the founding of HDS. That company played a big part of both our lives for the next eight years. During this period, Bill and I were also co-organizers of the first international chemical structures meeting, held in Noordwijkerhout in 1987. I reviewed Derwent CD ROMs when Bill worked for Derwent, and, from 1997 until 2002, I had a contract writing for ChemWeb.com. Space and time will not allow me to list all the projects we worked on together.

Bill was a very big part of my life for over 30 years. He was a source of business inspiration, a confidante who could be trusted with my secrets, an advisor when big decisions had to be made, and a shoulder to cry on when things were not going well. He was one of the cleverest people I knew. He was always able to stay calm, and do exactly the right thing in times of crisis. I loved to hear his voice.

As his son Matthew said, Bill was a gentleman, a scholar, modest, kind, sensitive, level-headed, quietly knowledgeable, always willing to help, a good man. Dear Bill, may you rest in peace. We offer our sincere sympathies to Maggie, Bill's wife; to Anne, his first wife, and their children Helen, Matthew, and Amy; to his sister Pat Bollens; and to his grandchildren, Joe, Ella, Emi, Tom and Casper.

Wendy Warr, October 30, 2019

Extending the InChI Identifier to Include Inorganic and Organometallic Compounds

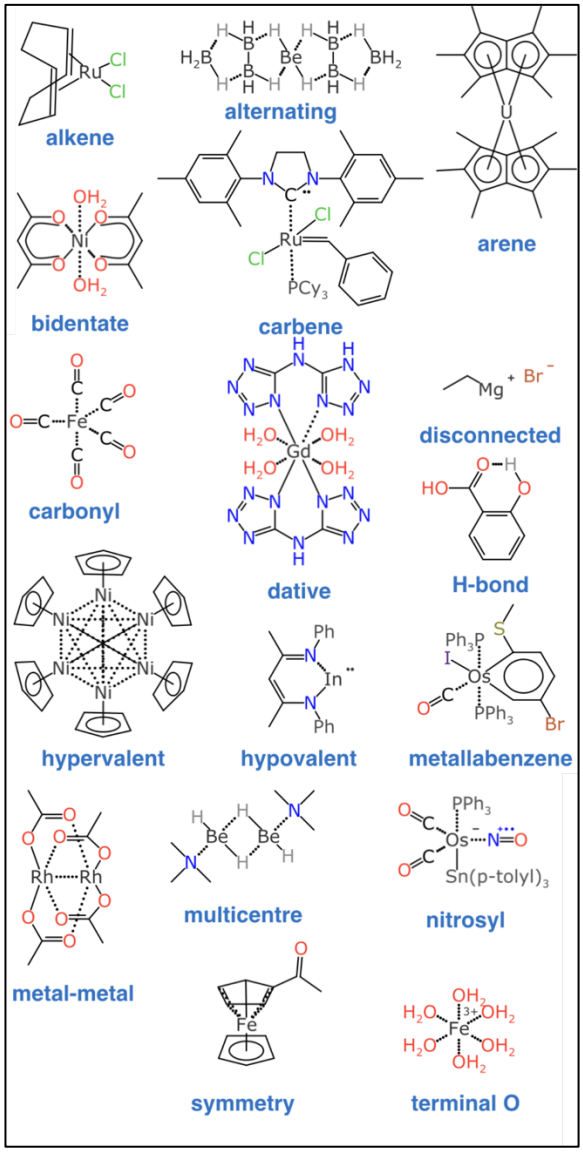
Contribution from Dr Alex M Clark, founder of Molecular Materials Informatics: email aclark.xyz@gmail.com

Readers will be familiar with the InChI identifier, which has come to be thought of as the definitive standard way of turning a molecular representation into a canonical string. The general idea is that whenever two drawings of a compound represent the same thing, they produce the same identifier, even if they were drawn differently or as different tautomers. The value to information systems is immense: these identifiers can be used as an essentially instant way to disambiguate chemicals, or to index them in databases that are themselves not in possession of any chemical awareness.

One of the weaknesses of the InChI identifier is shared by most of the cheminformatics industry: it was designed and validated primarily for drug-like organic compounds that have bonding arrangements that fit the Lewis octet rule. Outside of this domain lies the rest of the periodic table with its vast diversity of exotic bond types, many of which seem to exist solely for the purpose of defying classification. The standard InChI generation algorithm partially solves this problem by simply disconnecting any bonds that involve metals. While this entails a significant amount of destruction of information (i.e. loss of reversibility), in principle it could have only a small impact on the primary use case for InChI (i.e. indexing of chemicals). In practice, though, bond deletion is not compatible with multiple equally valid valence models, and interferes with downstream normalisation algorithms. The result is that generating standard InChI identifiers for inorganic/organometallic structures frequently gives rise to nonsensical descriptions which are mutually incompatible.

In spring 2019, the InChI Trust commissioned an effort to gauge the scope of the problem and chart a course toward augmenting the InChI identifier so that it can play nice with all manner of different coordination complexes that are found throughout the inorganic/organometallic realm. The results of this initial step can be found in the public GitHub repository, which also includes a detailed write-up of the technology.

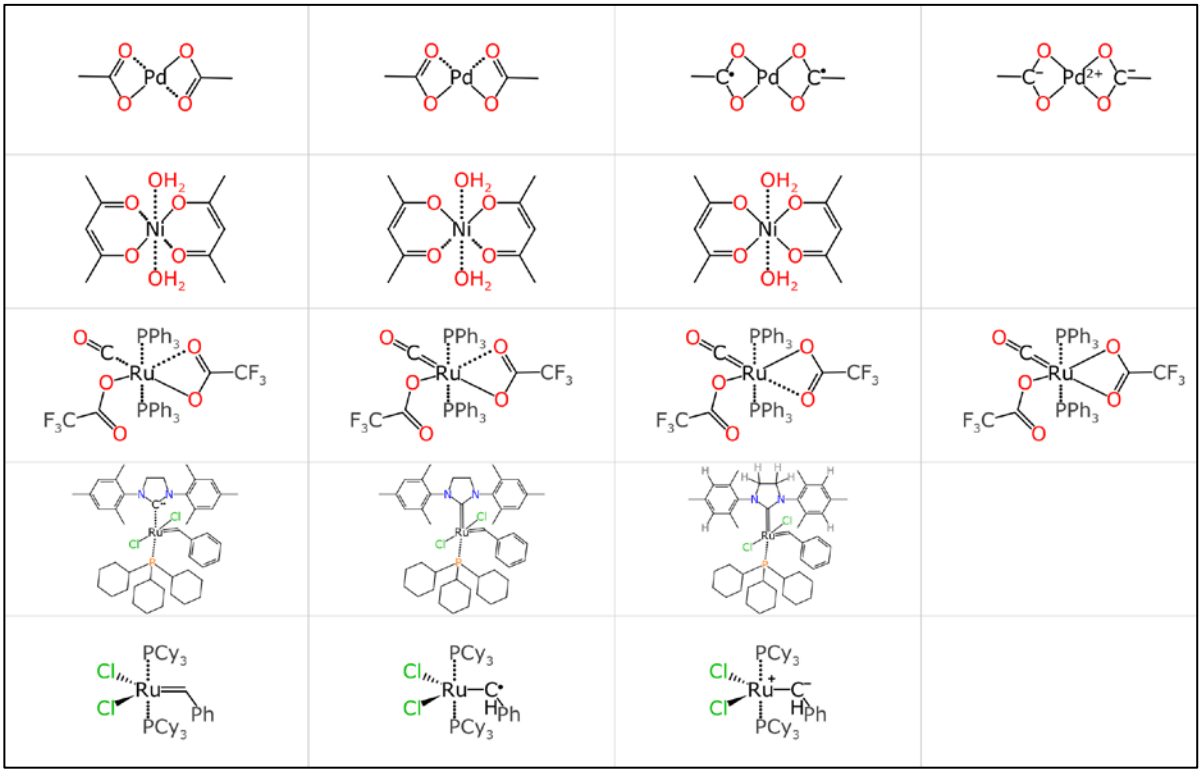
The first order of business was to gather data. As many are aware, there is very little software support for representing coordination compounds, and consequently even less useful data. One of the starting points was a set of 500 or so complexes that I had curated myself over the years. Another source was PubChem, but unfortunately the vast majority of coordination complexes were either drawn incorrectly at the time of submission or broken beyond recognition by the autocorrect algorithms. The most useful source turned out to be the Cambridge Structural Database which was made available for this project, along with some secret information that unlocked valuable cheminformatics properties, and which will hopefully be unveiled at a later date. About half of the million structures are made of non-polymeric coordination complexes, which are fair game for this task.



From these collections, a diverse set of coordination complexes was assembled, which were determined to be correct – and which for this purpose is defined as:

1. Complete bond graph, i.e. no disconnections between bonded atoms
2. All hydrogens accounted for correctly, whether virtual or actual
3. Charges correctly localised onto their connected components

Note that the above criteria are quite loose. Condition (1) says nothing about how the bond types are labelled; condition (2) can be satisfied by creating an atom for each hydrogen, or defining a formula for H-counts, or storing the number as an explicit property, or some combination thereof; and (3) allows charges to be placed anywhere, as long as they add up to the right value without jumping across space. This means that almost all coordination complexes have a huge number of valid ways to draw, and most of them have more than one variant which might be chosen by at least one chemist within the limitations of a particular drawing tool. For the dedicated training set, I went through 500 diverse structures and ensured that each one had several different sketches, each of them valid in its own way.



The primary goal of the coordination-enhanced InChI identifier is to be able to create the same output string for every structure that describes a particular molecular entity correctly. At this stage of exploration, we are ignoring stereochemistry (which is fortunately a separate problem), and are also not too worried about being able to reverse the process (i.e. get back a meaningful drawing). The question is: can we devise a canonical labelling method that is invariant to the multitude of different ways that coordination complexes can be drawn?

The short answer is yes: there is at least one way - as demonstrated - which works well and has an acceptably small set of known edge cases, with supporting evidence of this claim drawn from real world structures (i.e. the Cambridge Structural Database). The approach that turned out to be productive involves using the bond type metadata to perceive electron delocalisation subgraphs, within which electrons (and charges) are considered to be mobile and thus smeared across that proportion of the graph. This includes the familiar concept of aromaticity and double/single bond resonance, as well as ligands coordinated to metal centres. Once this treatment has been applied to the input structure, it is possible to compose a method for labelling each of the atoms & bonds, and selecting a canonical ordering for the atoms and bonds. This can be used to generate a line notation string, much the same as is done for the standard InChI and various implementations of canonical SMILES. The technical details (and the known edge cases) are described in detail on the GitHub site.

For the future of InChI, this alternate method of labelling represents an integration challenge, since it cannot coexist with the normalisation procedures that the standard InChI algorithm uses. The first official implementation may be done by grafting on an additional layer to the standard InChI, which can be used as an alternate option when coordination bonds are part of a particular use case. The proof of concept approach for coordination complexes also does not attempt to address tautomerisation (which may not even be viable for inorganic compounds) and defers the issue of stereochemistry (which has several extra geometries that simple organic compounds do not).

In conclusion, we now have an openly available demonstration that an InChI-like canonical identifier is possible for coordination complexes, and it does not rely on introducing any specific drawing conventions besides within reason. And we also have a validation set of real world coordination complex examples that can be used to verify correctness for any further refinements.

Editor's note: InChIs for Organometallic Compounds is not the only project the InchI Trust is sponsoring. InChI-based identifiers for reactions, mixtures, QR-codes, inorganics, and variable structures are amongst the things being developed. Do other areas to be explored too? The CICAG committee includes the secretary of the IUPAC InChI subcommittee, Prof Jonathan Goodman (jonathan@inchi-trust.org), who would be delighted to hear your suggestions.

Wendy Warr named 2020 Herman Skolnik Award Winner

Contribution via the Skolnik Award press release and courtesy of Dr Rajarshi Guha, Chair, ACS CINF Awards Committee: email rajarshi.guha@gmail.com

The American Chemical Society Division of Chemical Information is pleased to announce that Dr. Wendy Warr has been selected to receive the 2020 Herman Skolnik Award for her contributions to the fields of chemical information and a number of related fields that impinge on chemical information including chemical structure representation, substructure searching, retrosynthesis, and reaction prediction. The prize consists of a $3000 honorarium and a plaque. Dr. Warr will also be invited to organize an Award Symposium at the Fall 2020 ACS National Meeting to be held in San Francisco.

For more than 40 years, Dr Warr has had a global influence on chemical information and cheminformatics. She provides services to the chemical information, cheminformatics, and computational chemistry communities worldwide and has evolved into a key opinion leader and trend watcher. These trends have included: combinatorial chemistry, chemistry and the Internet, and intranet and ethernet in industry (these three in the early 1990s), outsourcing, changing pharma R&D strategies (these two in the early 2000s), and continue to today's AI, cloud computing, and blockchain.

Wendy formed her current company, Wendy Warr & Associates, in 1992, and since that time she has been successfully supplying business and competitive intelligence services to a broad spectrum of clients across the world. Her success stems from her extensive network, incredible energy, deep curiosity, and her specialized market knowledge of chemical information, computational chemistry, drug discovery, cheminformatics, STM publishing, and scientific communication. Wendy's clients have included at least 15 major pharmaceutical and chemical companies, venture capitalists, financial analysts, all of the well-known chemistry publishers, software companies, and many cheminformatics and analytical chemistry companies, as well as many smaller commercial and not-for-profit or academic organizations. Scientific database producers have benefited from her expert counsel and services in recent years.

Dr. Warr obtained her doctoral degree (D. Phil.) from Oxford University in 1971. She subsequently joined ICI Pharmaceuticals, where she held multiple positions, eventually becoming head of the Information Services department. Wendy established Wendy Warr & Associates in 1992, to provide business and competitive intelligence services to a broad spectrum of clients in the United States, Europe, Australia, the Middle East, and Asia. She has played key roles in several professional organisations including the American Chemical Society, RSC, the German Chemical Society, the Society of Chemical Industry, the Chemical Structure Association Trust, and the Institute of Information Scientists and, in many cases, has been instrumental in shaping their activities. She has published over 80 articles in academic journals and over 100 commercial reports, along with numerous invited lectures at venues such as NIST, Washington D.C., and University of Strasbourg, France. She has been an Associate Editor for the Journal of Chemical Information and Modeling (as well as its predecessor: the Journal of Chemical Information and Computer Sciences). Wendy has received numerous awards and honours including the Ernie Hyde Award of the

Chemical Structure Association (1984), and is a Fellow of the Royal Society of Chemistry and a Fellow of the Chartered Institute of Library and Information Professionals.

A Practical Case for WebAssembly

Contribution from Dr Dan Ormsby, email: do@dotmatics.com

Dan Ormsby works for Dotmatics Ltd in Bishop's Stortford, England. He is one of the authors of the *Dotmatics Chemistry Toolkit*.

Dotmatics provides a software platform for scientific research organisations. End users connect to the platform using a web browser and on the cloud side there is an application server and a database. Features include an electronic lab notebook (ELN), biological and chemical registration, reaction querying and reporting, document sharing, biological assay management, analysis and visualisation.

There is a big push from our customers for web based technologies where end users can bring any devices like tablets, PCs, and Macs and use the Dotmatics platform via a web browser. Web pages are great for "write once run anywhere" where the same code - HTML, CSS, JavaScript can be used on all the devices. This is quite different to compiling code for desktop PCs where Mac/PCs applications have to be built quite separately. The recent acceptance of WebAssembly as a new fourth language of web pages [1] has pushed Dotmatics to adopt this language for some Cheminformatics operations.

Cheminformatics tasks, such as chemical depiction, substructure searching, and matched pair analysis, previously would have been done within the database using a cartridge, on the application server or in a "fat client" code where a native application is executed on the client machine. At Dotmatics we have now compiled our chemistry engine into WebAssembly, meaning all cheminformatics tasks can now all be performed in the web browser.

Why?

Web deployments really mean "zero install on client, so the users don't have to install a full desktop application. The Dotmatics platform is optimised for speed of presentation of data in the browser. One product we recently created is called Clarity (Figure 1) which defaults to presenting users grids of structures. If rendering one structure on the server takes 10ms but the application is used by 100 simultaneous users who want to present 12 structures each, the total CPU required to draw the structures is approximately 12 CPU seconds if we leave the structure rendering server-side. This doesn't sound like much of a practical issue, but at Dotmatics we want everything to feel instantaneous (which isn't actually instant but what human perceive as instant and defined as 0.1s) [2].

The use case for WebAssembly here is to move the depiction algorithm to the client side, execute the same code (this time compiled to WebAssembly) and use the client side CPU to perform the structure rendering (via a WebWorker). Suddenly the 12 seconds of rendering time is divided over 100 client side CPUs, and each user sees the structures rendering instantaneously.

Figure 1: Screenshot of the Dotmatics web based application Clarity

At Dotmatics we maintain two internal chemistry toolkit codes, one is written in C and the other in JavaScript. Our C implementation is used most frequently, and as C is a fundamental language the code can be compiled into libraries for other technologies written in Java, C++, Oracle and Python. When we call our Pinpoint Oracle cartridge (via OCI), our chemistry in Java (via JNI), our Dotmatics for Office plugin (.Net library), a command line call under Browser, or property calculation on our iPad app, it is the same C code being run.

In each case the C code is compiled to the framework and platform architecture (ARM on iPad, x86 32-bit or x86_64 Windows/Mac/Linux). In 2009 we invested in our JavaScript chemistry engine to enable chemistry in web-powered devices like tablets and phones which were beginning to rise in popularity. This included our *Elemental* app, a drawing tool for depicting small molecules in web browsers.

WebAssembly brought us something new and meant we no longer had to compile to the native architectures. The Emscripten compiler (others are available like Cheerp) enables C code to be directly compiled to WebAssembly (Wasm files) which execute directly in a web page. This has removed any platform dependencies which come from having code on the server side or in a desktop application, as all the code is deployed on the client side. We can compile once, run anywhere.

The initial WebAssembly prototypes we created simply blew us away; a proof-of-concept exercise showed that compiling the Dotmatics C Chemistry Toolkit into WebAssembly showed remarkable performance within a few percent of native compiled code in our hands. The compilation could all be done with zero code change to the core toolkit. A small interface C code was written (a few dozen lines) and everything else "just worked".

Brute force substructure search tests were observed to complete in an average 0.24s for 50K matches (from SMILES strings). With a simple lightweight fingerprint this time fell to under 0.01s on all platforms (iPad/iPhone/Android Tablet/Safari on Mac/Edge on Windows).

This technology enables a stackless solution to a chemical catalogue search. No more do we need a HTML front end, an application server and a database to store the substructure search algorithm. We can instead present a single web page that contains the WebAssembly compiled algorithm and a text list of SMILES strings (or MOL files). All substructure searching can be done client side "instantly" for 50K-100K lists of compounds.

The current WebAssembly implementations run like they are on a 32-bit Linux. There is a 2GB of RAM per tab maximum limit over modern browsers (NB on an iPhone 6s you get about 1.8GB).

So in conclusion the future from Dotmatics software at least will be one containing a lot more WebAssembly.

[1] https://www.w3.org/TR/wasm-core-1/

[2] Instantaneous was redefined as within 100ms in a US Military standard. The use case would be when an operator presses a "Fire" button it should light within 100ms (or the operator may press it twice). The human operator perceives the illumination as occurring instantaneously. MIL_STD_1472D – March 1989 HUMAN ENGINEERING DESIGN CRITERIA FOR MILITARY SYSTEMS, EQUIPMENT AND FACILITIES - section 5.4.3.2.2.1.2 - Dynamic characteristics - page 107 - http://everyspec.com/MIL-STD/MIL-STD-1400-1499/MIL_STD_1472D_1209/

Editor's note: Interested readers may also like to read the *WebAssembly* article contributed from Richard L. Apodaca in the CICAG Summer 2019 Newsletter (see http://www.rsccicag.org/newsletters.htm)

Tony Kent Strix Award and Annual Lecture 2019

Contributed by RSC CICAG Member Jane List, email jane@extractinfo.info

Right: Prof Pia Borlund giving her lecture. A link to the 2019 Strix Award review and presentation recordings can be found here.

At the end of November a small group of information scientists and information retrieval enthusiasts met at the Geological Society in London to hear the announcement of the 2019 Tony Kent Strix Award and to hear a lecture presented by last year's winner, Professor Pia Borlund, from the Department of Archivists, Library and Information Science at Oslo Metropolitan University.

Professor Borlund's acceptance talk focussed on interactive information retrieval (IIR) research by reviewing what *not* to do when conducting 'live' research in this field. Her research attempts to approach information retrieval research by using information seeking activities which are close to real information tasks, conducted by real people. Examples given were based on students searching for holiday destinations, and similar activities. Today everyone uses digital information to search for answers to their everyday questions. Professor Borlund described methodological issues in designing studies to assess IR systems. Her research concerns were regarding the scope of the study itself, and also regarding the influence by which the people who carry out IR can influence the test results. Factors such as how much the task mattered to the participants and their level of knowledge on the subject in the first place could bias study results. A lively Q&A session followed Borlund's talk.

Prior to the main lecture Dr Andrew Macfarlane gave a talk entitled '*Sender vs Recipient Orientated Information Systems Revisited*'. Dr McFarlane looked back to 1976 and, in particular, Belkin and Robertson's paper on ethical implications of theoretical research in information science, which 40 years ago foresaw some of the issues which need addressing today. These include issues concerning information bias, misinformation, and the power of the sender of information to deliver the message through hyper-personalisation of results listings, particularly in the politics and advertising worlds. Dr MacFarlane argued for more transparency in information systems, and a return of control to the user. Another lively discussion ensued, with arguments aired on the side of greater control, and on the opposite side of having a free public sphere where to para-phrase Milton 'truth and falsehood grapple'. Both sides agreed that education in information seeking skills and greater information literacy are needed. It was agreed that the recipients of all the information now available 24/7 on any subject under the sun, both sought and uninvited, need to be better able to judge its validity for themselves.

2019 Tony Kent Strix Memorial Award Winner

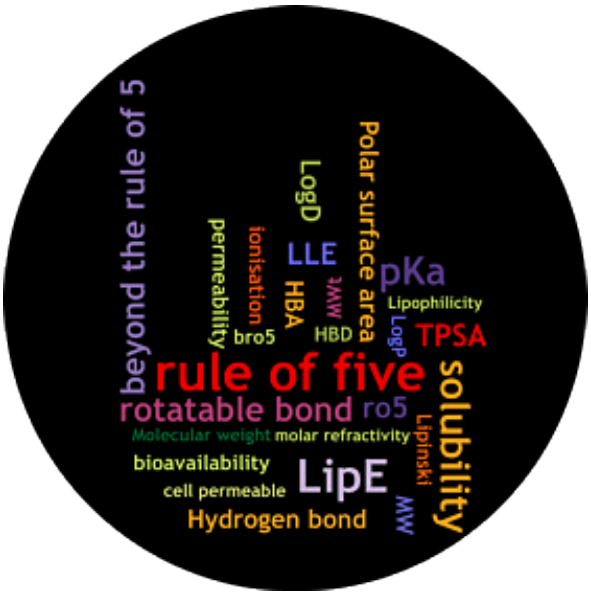
Finally, Professor Ingemar J. Cox was announced as the 2019 Tony Kent Strix Memorial Award winner. Professor Cox has been invited to deliver a lecture in November 2020, as he has been recognised in particular for his work on visual information retrieval across multiple disciplines, including in the medical and watermarking areas. Prof Cox is currently head of the Media Futures Research Group at University College London and Professor at Department of Computer Science University of Copenhagen.

Meeting Report: 20 Years of the Rule of Five

Contributed by RSC CICAG Chair Dr Chris Swain, email: swain@mac.com

Joint RSC CICAG-BMCS Meeting, 20th November 2019, Sygnature Discovery, BioCity, Nottingham, UK

It has been over twenty years since Lipinski published his work determining the properties of drug molecules associated with good solubility and permeability [1]. The "Rule of 5" rapidly became synonymous with drug-like properties though-out the Pharma industry (16,000 citations). Since then, there have been a number of additions and expansions to these "rules". There has also been keen interest in the application of these guidelines in the drug discovery process and how these apply to new emerging chemical structures "beyond the rule of five" (bro5) such as macrocycles.



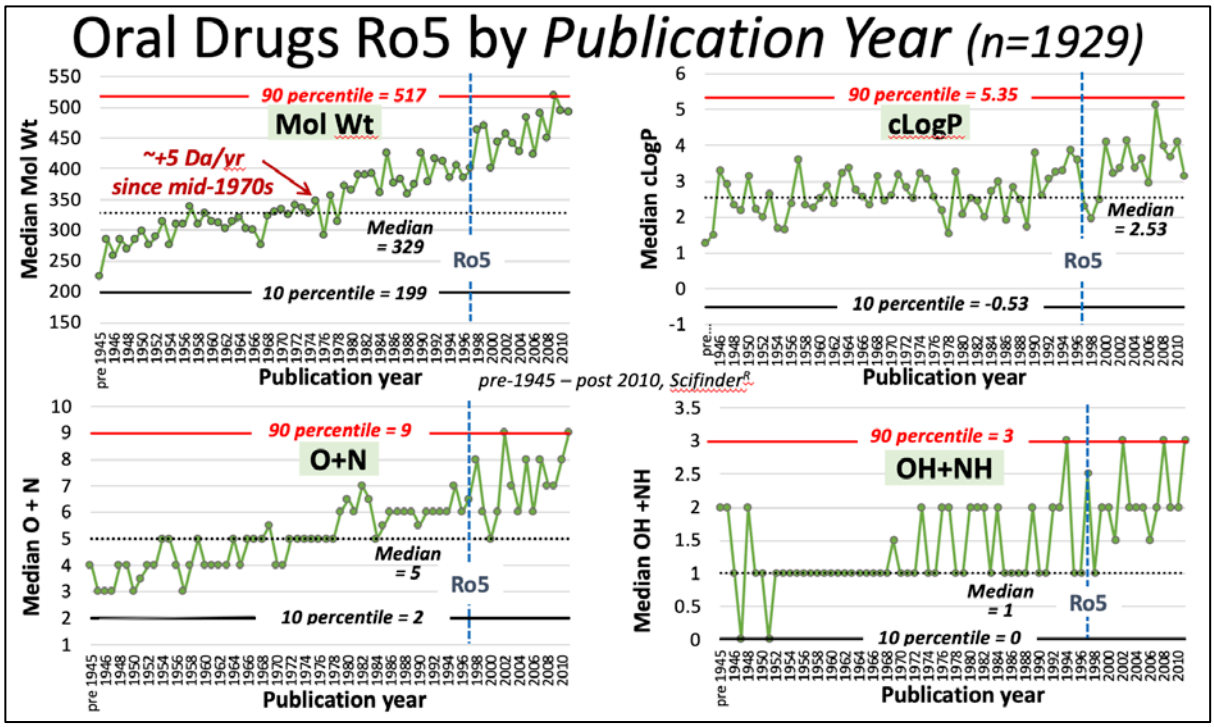
The *20 Years of the Rule of Five* Meeting brought together researchers from a number of different areas of drug discovery and provided both a historical overview of the use of Lipinski's rules, as well as looking to the future and how these rules might evolve in the changing drug compound landscape. The meeting had a capacity attendance of over 100, with Sygnature kindly providing the venue. The audience was a nice mix of industry "veterans", students, and those new to the industry. The meeting format was a morning session giving a historical viewpoint followed by a panel discussion, and the afternoon was consisted of a more forward looking session, again followed by a panel discussion.

Paul Leeson (Paul Leeson Consulting) started the meeting by reminding everyone that the original paper came about as a response to hits from high-throughput screening that had unfavourable physicochemical properties.

> *"Experimental and computational approaches to estimate solubility and permeability in discovery and development settings are described. In the discovery setting `the rule of 5' predicts that poor absorption or permeation is more likely when there are more than 5 H-bond donors, 10 H-bond acceptors, the molecular weight (MWT) is greater than 500 and the calculated Log P (CLogP) is greater than 5 (or MlogP>4.15)."*

The calculated properties can be categorised as "size, lipophilicity and polarity", whilst molecular weight is used to describe size, perhaps this overestimates the contributions of some atoms e.g. fluorine. Calculated LogP is used as a measure of lipophilicity but it was pointed out that different algorithms may give different results. The number of H-bond donors and acceptors in the original paper is actually a count of O and N or OH and NH and not a true measure of all atoms and their ability to form hydrogen bonds. In the panel discussion it was suggested that the use of the use of H-bonds scales such as those proposed by Abraham [2] might be interesting.

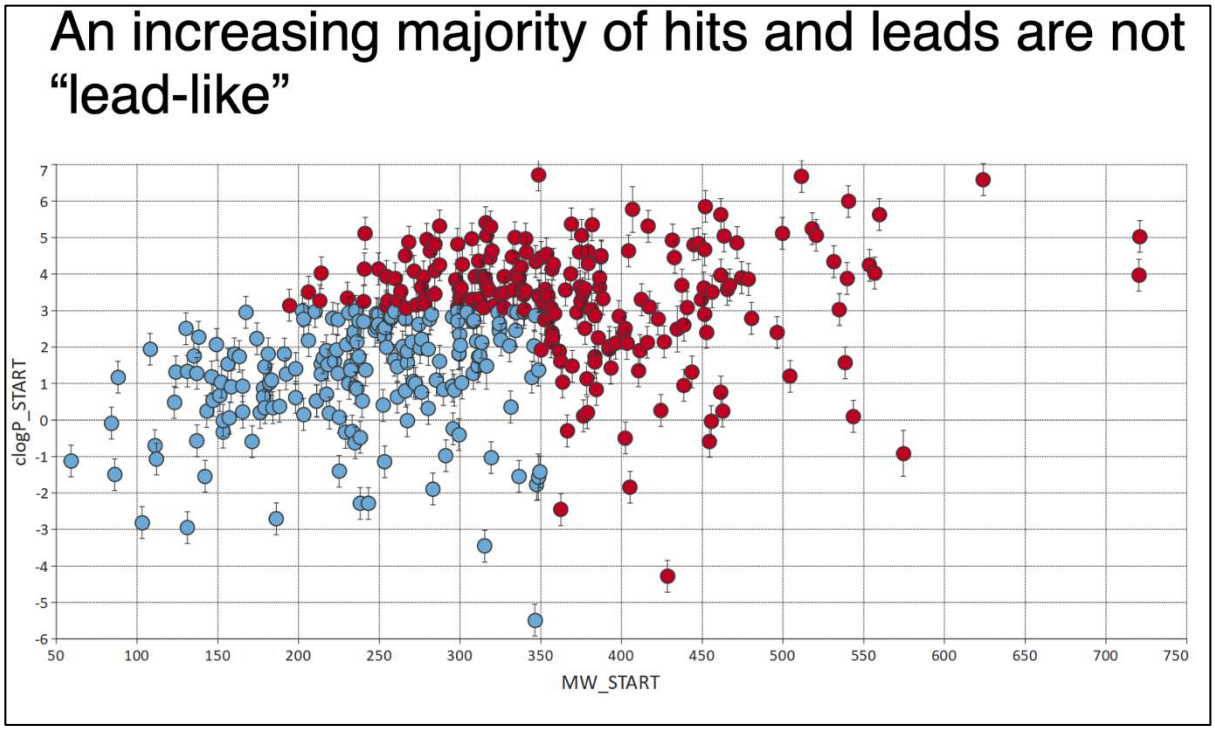
Looking at the properties of published drugs (we of course don't know about all failures) since the publication of the rule of 5 paper, molecular weight has continued to increase, cLogP appears to have plateaued around 4, and whilst there is an increase in the number of HBA, there is only a marginal increase in HBD.



It was suggested, and later elaborated in the discussions, that much of the increase is probably driven by the targets now being addressed, for example those with extended and/or shallow binding sites, protein-protein

interactions, HIV, and HCV. It was also noted that many of the high molecular weight drugs had extremely high affinity (sub-nanomolar) and thus required low doses.

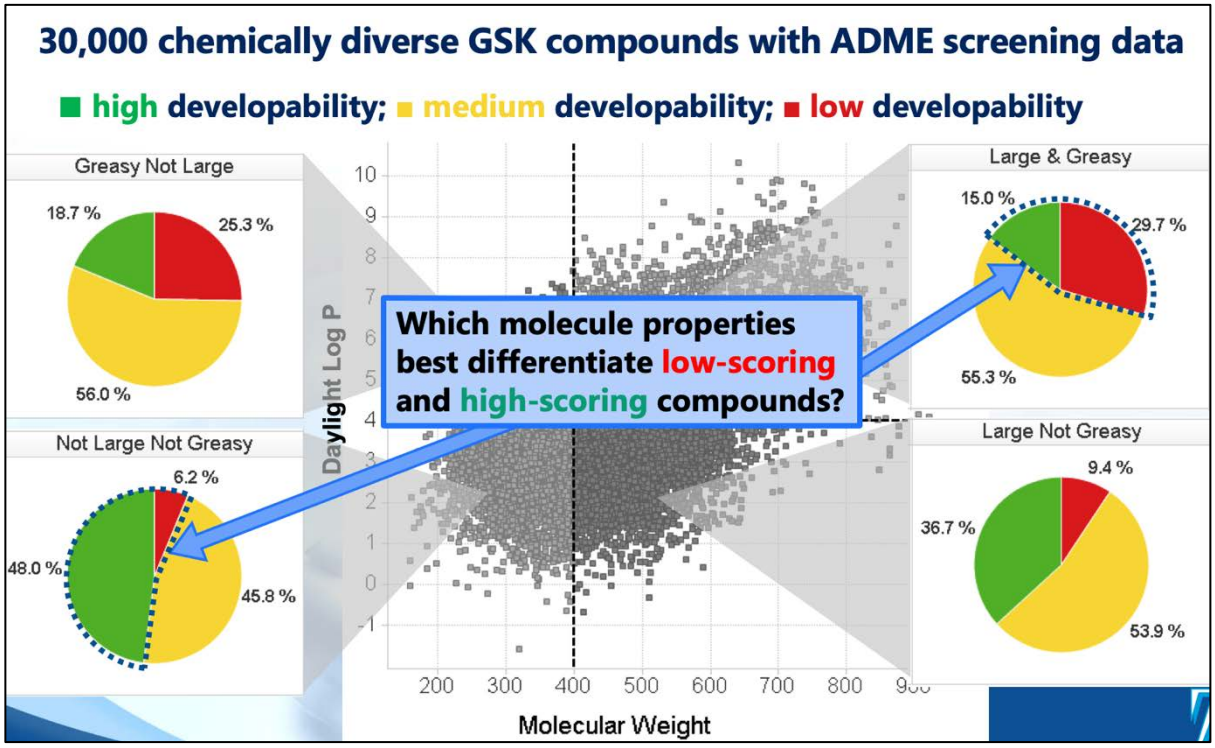
Michael Shultz (Novartis) continued the review of properties used to predict "drug-likeness" by introducing polar surface area (TPSA), rotatable bonds (RotB), fraction of sp3 carbons (Fsp3) and aromatic ring counts (#ArRNG). He also highlighted the difference between, hits, leads and drugs, and the suggestion that lead-like structures have MWt <350 and cLogP <3. He also noted that historically in drug discovery optimisation processes there are significant increases to MWt and cLogP [3]. However, using data from a recent publication [4] he suggested that leads are becoming less lead-like with increasing MWt and cLogP in particular. In fact, only 47% of recent leads have MWt <350 and cLogP <3. This was followed up in the panel discussion with the suggestion that a good strategy might be to first prune down any hits from screening, rather than seeking to immediately add to the hit structure.



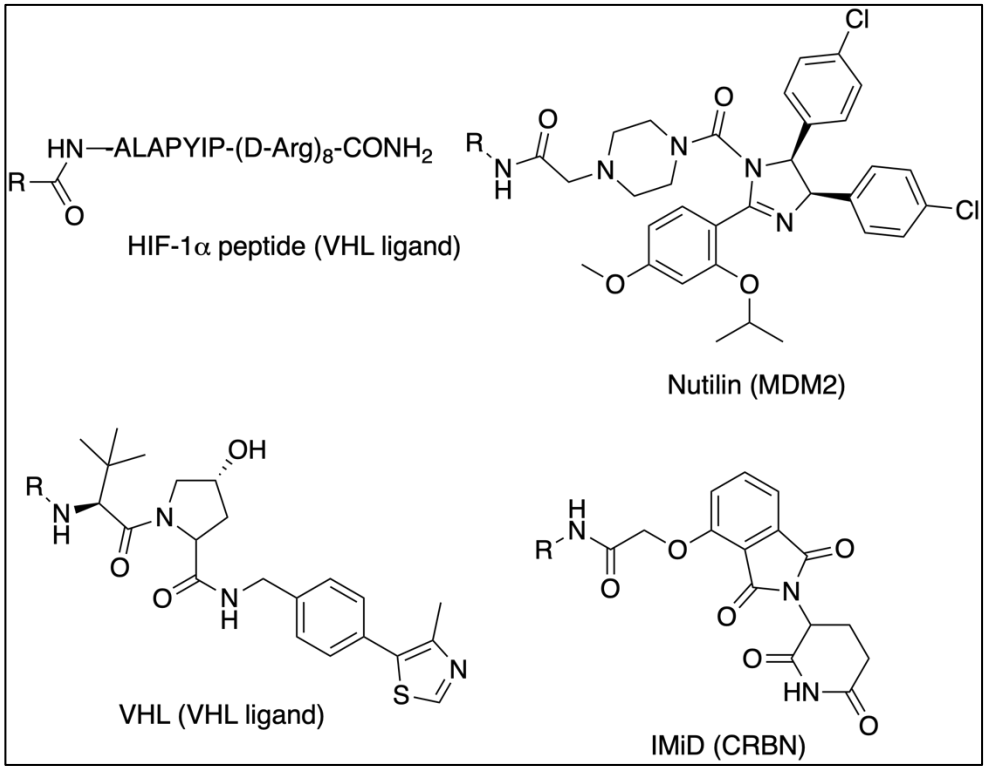
Michael also highlighted the issues of experimental measurement of LogP and solubility of highly lipophilic (>6) and poorly water-soluble compounds.

Tim Ritchie (Zerlavanz Consulting) continued the discussion on descriptors by focussing on aromatic and aliphatic descriptors and the impact on 'developability' [4,5]. Increasing carboaromatic ring count decreases solubility, and in addition also leads to increased plasma protein binding, increased CYP450 inhibition and HERG issues. Replacing carboaromatic rings with heteroaromatic or heteroaliphatic rings can lead to improvements in all properties. Interestingly, whilst MWt and cLogP have increased, the aromatic-aliphatic atom balance has not changed in oral drugs over time.

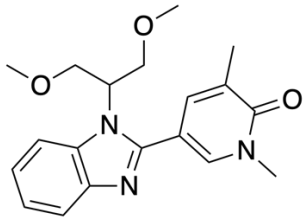
Tim also described the developability score [7] for distinct cLog P/molecular weight regions that define optimal and sub-optimal chemical space, and a developability score derived from regression models using solubility, permeability, protein binding and 3A4 inhibition screening data. Whilst the sector MWt <400, cLogP <4 suggested the greatest chance of success, it was noted that even the MWt >400, cLogP >4 sector included some developable molecules, albeit with a much lower chance of success.



John Priestner (GSK, Univ of Strathclyde) described their work investigating modifications to the linker functionality of BET PROTACs and the effect on physicochemical properties. PROTACs [8] are bifunctional molecules that bind to the target protein and an E3 ligase, the simultaneous PROTAC binding of two proteins brings the target protein in close enough proximity for polyubiquitination by the E2 enzyme associated to the E3 ligase, which flags the target protein for degradation through the proteasome. Despite having properties well outside those described by the rule of 5, the first PROTAC entered phase I clinical trials in 2019. Whilst the first PROTACs used peptidic E3 ligase ligands, more recent PROTACs use small molecule ligands.



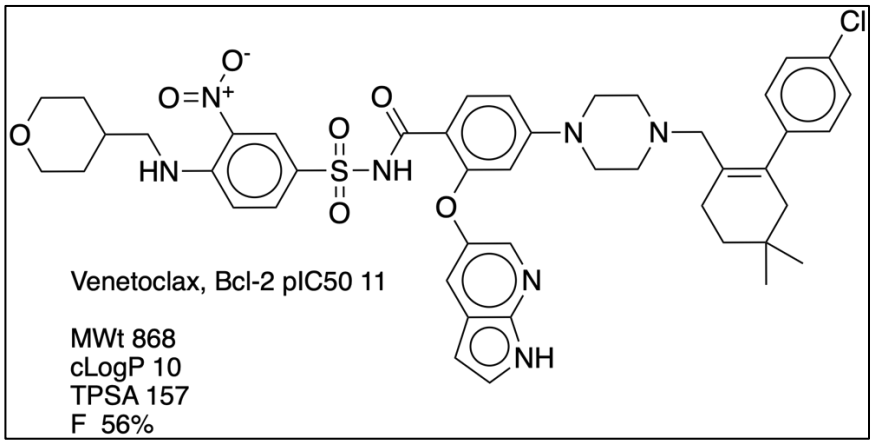
In this work the VHL ligand was chosen, and a pan-BET ligand (GSK001) identified from an encoded library screen.



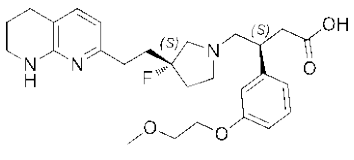
GSK0001

A variety of synthetic strategies were investigated for linker incorporation, and amide coupling proved to be the most robust. The physicochemical properties of the resulting PROTACs were determined and shown to cover a wide area of property space. Several examples have been shown to degrade the Bromodomain-containing protein 4 BDR4 in vitro. A more detailed investigation of linker properties is ongoing.

Elisabetta Chiarparin (AZ) described their work investigating the impact of conformational constraints on drug-like properties. In a series of elegant NMR studies they showed that macrocycle formation can preorganise ligands into the bioactive conformation, leading to significant potency increases. The NMR studies also highlight reduced flexibility (NMR Rot lower), but also masking of hydrogen bond donor/acceptors. This polarity masking led to the exposed polar surface area (ePSA) descriptor, which correlates with Caco-2 permeability. Elisabetta then made the case that we need to use 3D descriptors (3D Ro5) to better describe the properties of drugs, particularly high molecular weight drugs. In the case of Venetoclax, despite its very high molecular weight, cLogP and TPSA, it has acceptable oral bioavailability. Conformational analysis and NMR studies suggest it adopts a folded structure driven by pi-stacking and intramolecular hydrogen bonds.



James Thompson (GSK) reported their work investigating the chameleonic properties of $\alpha v\beta_6$ integrin antagonists for the treatment of IPF. The binding site is highly polar, and the two key binding interactions are salt bridges to a metal ion an aspartic acid residue in the receptor. Small molecule ligands for RGD integrins are zwitterions containing polar acidic and basic functionality [9].



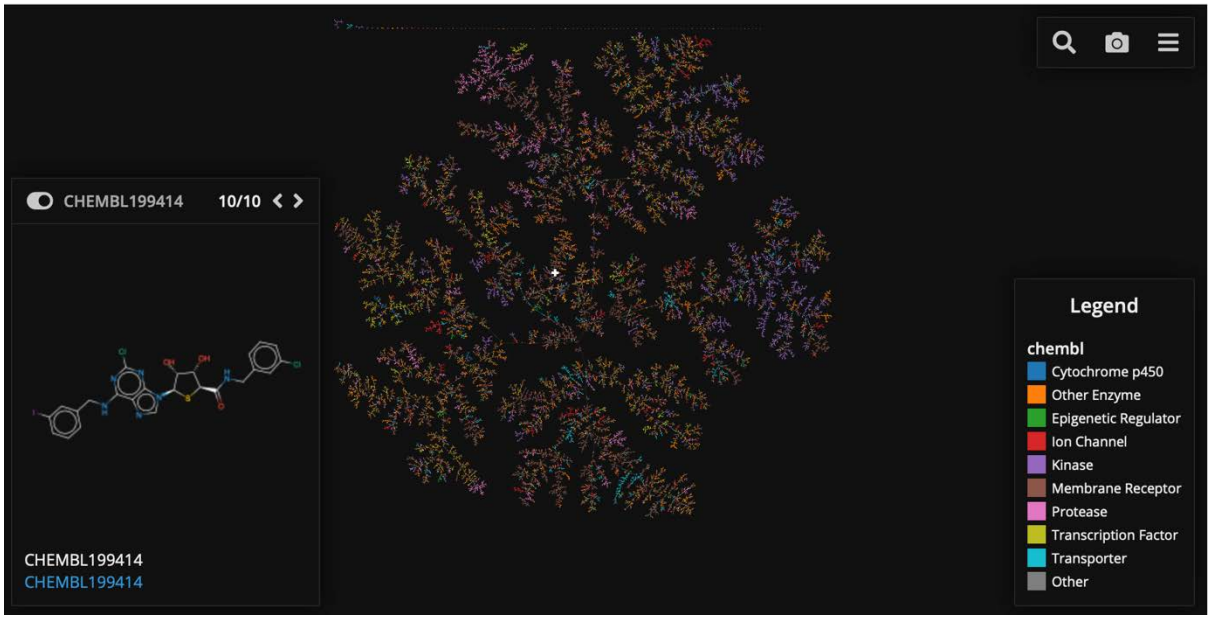
1

Interestingly the diastereomer **1** was found to be more permeable than the other diastereomers despite identical pK_a and LogD. It was hypothesised that this was due to an intramolecular hydrogen bond between the acid and the tetrahydro-1,8-naphthyridine. A series of analogues were prepared and ^{15}N NMR used to

investigate conformations and presence of the intramolecular hydrogen bond and the impact on permeability. These methods have subsequently been successfully used to rationalise the permeability of several more recent series of $\alpha_v\beta_6$ antagonists.

Whilst at times on a discovery programme drug space can seem a little crowded, Jean-Louis Reymond gave an insight into how vast chemical space really is. The GB-17 database contains 166.4 billion possible molecules up to 17 atoms of C, N, O, S and halogens following simple chemical stability and synthetic feasibility rules [10]. Using a set of MedChem rules and sampling this was filtered to produce GDBMedChem, a compact collection of 10 million small molecules [11].

Visualisation of very large datasets is a significant issue and they have also been developing novel fingerprint and visualisation tools. Tmap is a very fast visualization library for large, high-dimensional datasets. A demonstration of Tmap displaying the 1.1 million ChEMBL dataset can be found here.

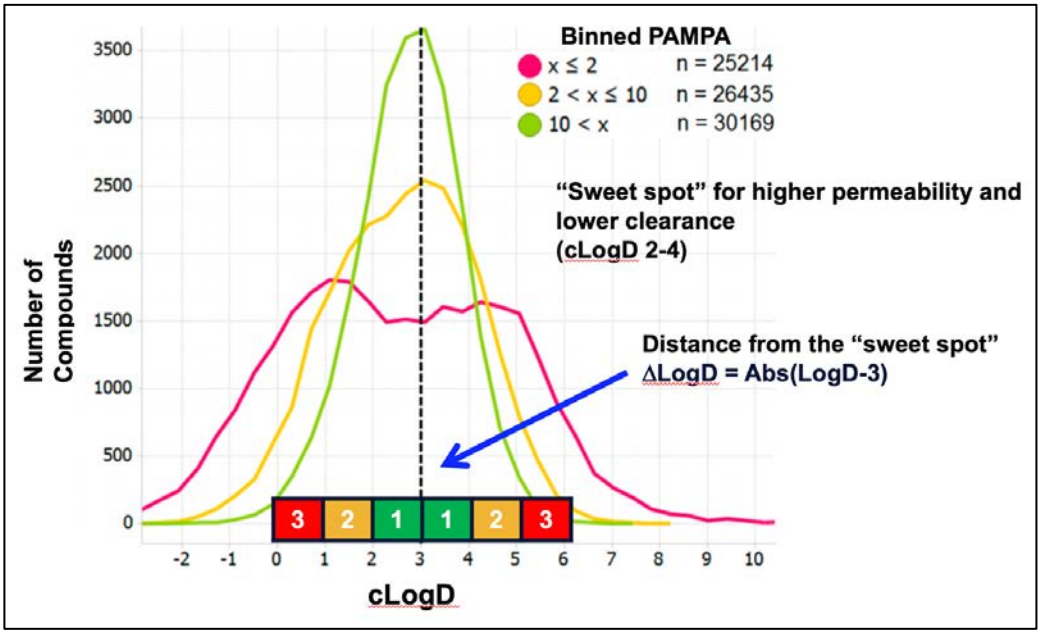


Most of the descriptors are quick and easy to calculate, however Jóhannes Reynisson (Keele Univ) has chosen Density Functional Theory (DFT) to try and describe chemical space. Several properties were determined for known drugs such as dipole moments, polarisability, ionisation potentials, and electron affinity. Higher polarisability tended to be associated with poor oral availability, and lower dipole moment with higher bioavailability, but there was considerable variation.

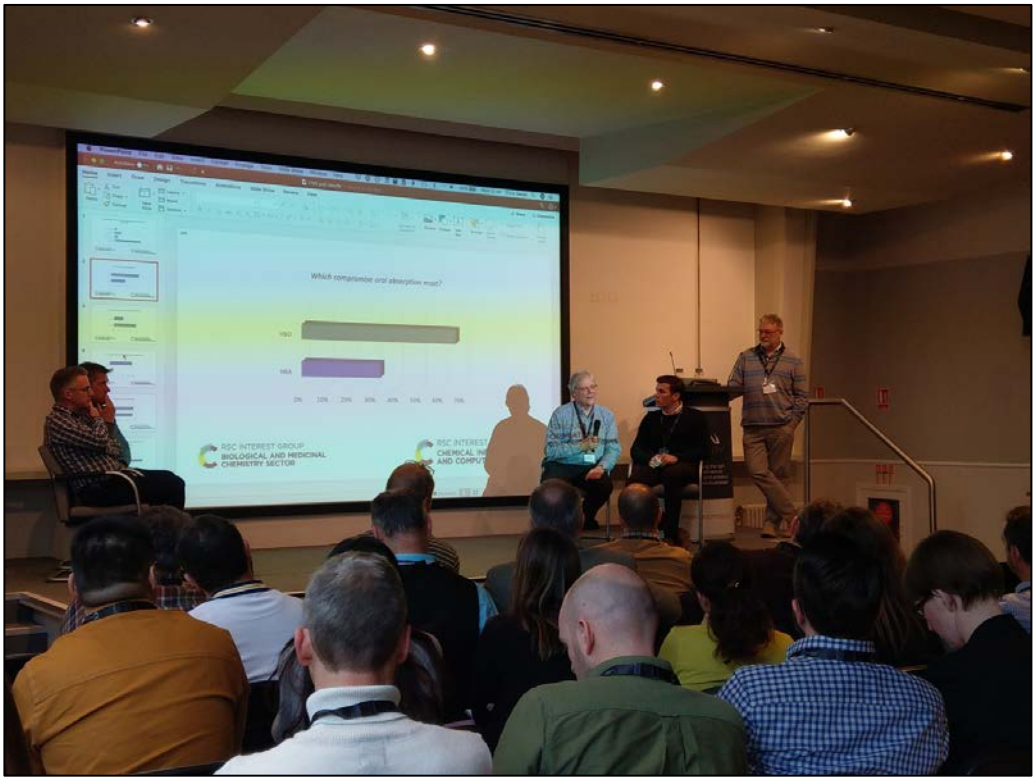
Phil Cox (AbbVie) finished the afternoon session by describing some of the lessons learned from AbbVie's work in the Beyond the Ro5 space (Bro5) [12]. Whilst many people have attempted to describe conventional drug space (Ro5), there are many difficult targets that require exploration of chemical space beyond the Ro5. In addition, whilst there is increasing amount of experimental data for compounds within the Ro5 space, there is very little for Bro5. An analysis of the AbbVie DMPK database looking specifically at compounds Bro5 using a wide variety of calculated properties identified three properties that correlated: Number of aromatic rings (NAR), number of rotatable bonds (NRB), and the difference in cLogD from the mean cLogD of all compounds (LogD-3).

They proposed a simple multiparametric scoring function (AB-MPS) that correlated preclinical PK results with cLogD, the number of rotatable bonds, and the number of aromatic rings.

AB-MPS = Abs(LogD-3) + NAR + NRB, is the mnemonic for approximating odds of oral absorption for compounds bRo5.



In conclusion, this was a very popular meeting, with a key piece of feedback from attendees being the panel discussions which were particularly enjoyable. CICAG will endeavor to repeat this format in future planned events.



Above: One of the panel discussion sessions, reviewing the history of the Rule of Five

References

[1] Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Advanced Drug Delivery Reviews. 46 (1–3): 3–26 https://doi.org/10.1016/S0169-409X(00)00129-0
[2] Hydrogen bonding. Part 9. Solute proton donor and proton acceptor scales for use in drug design. J Chem Soc Perkin Trans 2:1355–1375 (1989) https://doi.org/10.1039/p29890001355
[3] Is there a difference between leads and drugs? A historical perspective. J Chem Inf Comput Sci. 2001 Sep-Oct;41(5):1308-15 https://doi.org/10.1021/ci010366a

[4] Where Do Recent Small Molecule Clinical Development Candidates Come From? J. Med. Chem. 2018 Nov 8;61(21):9442-9468, https://doi.org/10.1021/acs.jmedchem.8b00675
[5] The impact of aromatic ring count on compound developability – are too many aromatic rings a liability in drug design?, Drug Discovery Today Vol 14 (21-22) November 2009, Pages 1011-1020, https://doi.org/10.1016/j.drudis.2009.07.014
[6] Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success, J. Med. Chem. 2009, 52, 21, 6752-6756, https://doi.org/10.1021/jm901241e
[7] Increasing small molecule drug developability in sub-optimal chemical space, Med. Chem. Commun., 2013,4, 673-680, http://dx.doi.org/10.1039/C3MD00003F
[8] Protac-Induced Protein Degradation in Drug Discovery: Breaking the Rules or Just Making New Ones? J. Med. Chem. 2018, 61, 2, 444-452, https://doi.org/10.1021/acs.jmedchem.7b0127
[9] WO/2016/046225 https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2016046225
[10] Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17, J. Chem. Inf. Model. 2012, 52, 11, 2864-2875, https://doi.org/10.1021/ci300415d
[11] Medicinal Chemistry Aware Database GDBMedChem, Mol. Inf. 2019, 38, 1900031 https://doi.org/10.1002/minf.201900031
[12] Beyond the Rule of 5: Lessons Learned from AbbVie's Drugs and Compound Collection, J. Med. Chem. 2018, 61, 7, 2636-2651, https://doi.org/10.1021/acs.jmedchem.7b00717

Meeting Report: 2nd RSC BMCS & CICAG AI in Chemistry Meeting

Right: Josh Meyers from BenevolentAI makes his presentation at the meeting

Contribution from CICAG Committee member Dr Nathan Brown, nathan.brown@benevolent.ai

2nd-3rd September 2019, Fitzwilliam College, Cambridge, UK

Organising Committee: Nathan Brown (Chair), Maggi Churchouse, Samantha Hughes, Phil Jones (Treasurer), Chris Swain

In 2019 we held the 2nd iteration of our Artificial Intelligence in Chemistry meeting. Given the overwhelming success of the first meeting, the organising committee decided to extend the meeting over two days and use a much larger venue to cope with the huge demand for delegate places in the first year. We extended the programme to cover sixteen speakers, with eight on each day. Of the eight speakers each day, two were keynote speeches from thought leaders in the field.

As an organising committee, we encouraged open discussion and this was demonstrated by the active Twitter engagement on the hashtag #AIChem19, in addition to extended networking lunches to facilitate discussions after the speaker and flash poster presentations.

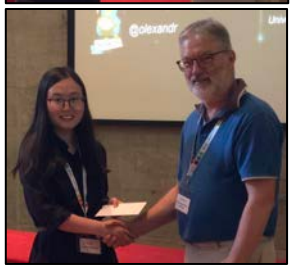
One of my personal highlights was a robust and honest presentation from Anthony Nicholls (CEO, OpenEye Scientific Software), entitled: *Is your machine learning telling you anything you didn't already know*? Ant's

presentation moved through the ability (or lack thereof) for machine learning to tell us what we didn't already know through, to the ability to tell us things that are surprising or unexpected and how these relate to the real world challenges of drug discovery. Ant's talk, just before the excellent conference dinner on Monday evening, gave all of the delegates additional food for thought and led to a great deal of excited discussion.

The 2019 Poster winners (presentation from CICAG Chair Dr Chris Swain) were:

No 17: Jenke Scheen of the University of Edinburgh: "*Improving the accuracy of alchemical free energy methods by learning correction terms for binding energy estimates*"

No 6: Adam Green of the University of Leeds: "*Activity-directed discovery of inhibitors of the p53/MDM2 interaction: towards autonomous functional molecule discovery*"



No. 3: Ya Chen of the University of Hamburg: "*NP-Scout: machine learning approach for the identification of natural products and natural product-like compounds in large molecular databases*"

The content of the conference was very strong and we are delighted to be able to share a large number of the speaker and poster presentations on the CICAG website.

2nd RSC BMCS & CICAG AI in Chemistry Meeting: Event statistics -

Twitter Mentions (#AIChem19)	*597*
Unique Twitter Accounts	*169*
Delegates	*163 (53 from universities, 26 students)*
Speakers	*12*
Keynotes	*4*
Posters	*19*
Session Chairs	*6*
Days	*2*
Organisers	*5*

Delegates from Australia, Belgium, China, Croatia, Denmark, France, Germany, Hungary, Italy, Japan, Korea, Netherlands, Poland, Saudi Arabia, Spain, Sweden, Switzerland, USA
Thank you to our sponsors: 3 Sponsors Catapult, Charles River and AstraZeneca
And to our exhibitors: IKTOS, MedChemica Ltd, OpenEye Scientific Software, Optibrium, Mcule

Given the continued success of the meeting, the organising committee is already planning the third meeting which will take place at Churchill College, Cambridge, UK on 28-29 September 2020. The organising committee is seeking speaker and poster abstracts by 1st May and 17th August, respectively. Submit your abstracts here!

Chemical Information / Cheminformatics and Related Books

Contributed by RSC CICAG Newsletter Editor Stuart Newbold, email: stuart@psandim.com

Recording Science in the Digital Era

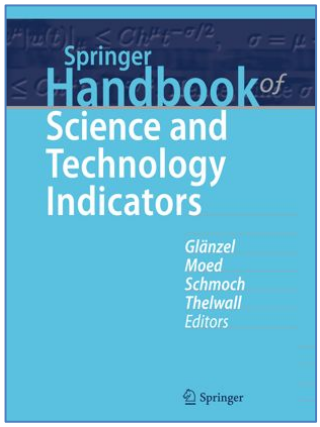
For most of the history of scientific endeavour, science has been recorded on paper. In this digital era, however, there is increasing pressure to abandon paper in favour of digital tools. Despite the benefits, there are barriers to the adoption of such tools, not least their usability. As the relentless development of technology changes the way we work, we need to ensure that the design of technology not only overcomes these barriers, but facilitates us as scientists and supports better practice within science. This book examines the importance of record-keeping in science, current record-keeping practices, and the role of technology for enabling the effective capture, reuse, sharing, and preservation of scientific data.

Covering the essential areas of electronic laboratory notebooks (ELNs) and digital tools for recording scientific data, including an overview of the current data management technology available and the benefits and pitfalls of using these technologies, this book is a useful tool for those interested in implementing digital data solutions within their research groups or departments. This book also provides insight into important factors to consider in the design of digital tools such as ELNs for those interested in producing their own tools. Finally, it looks at the role of current technology and then considers how that technology might develop in the future to better support scientists in their work, and in capturing and sharing the scientific record.

RSC Publishing

Author: Cerys Willoughby

Springer Handbook of Science and Technology Indicators



This handbook presents the state of the art of quantitative methods and models to understand and assess the science and technology system. Focusing on various aspects of the development and application of indicators derived from data on scholarly publications, patents and electronic communications, the individual chapters, written by leading experts, discuss theoretical and methodological issues, illustrate applications, highlight their policy context and relevance, and point to future research directions.

A substantial portion of the book is dedicated to detailed descriptions and analyses of data sources, presenting both traditional and advanced approaches. It addresses the main bibliographic metrics and indexes, such as the journal impact factor and the h-index, as well as altmetric and webometric indicators and science mapping techniques on different levels of aggregation and in the context of their value for the assessment of research performance as well as their impact on research policy and society. It also presents and critically discusses various national research evaluation systems.

Complementing the sections reflecting on the science system, the technology section includes multiple chapters that explain different aspects of patent statistics, patent classification and database search methods to retrieve patent-related information. In addition, it examines

the relevance of trademarks and standards as additional technological indicators.

The Springer Handbook of Science and Technology Indicators is an invaluable resource for practitioners, scientists and policy makers wanting a systematic and thorough analysis of the potential and limitations of the various approaches to assess research and research performance.
Editors: Wolfgang Glänzel, Henk F. Moed, Ulrich Schmoch and Mike Thelwall

---

News from CAS

Contribution from Dr Anne Jones, email: ajones2@acs-i.org

The second half of 2019 was a busy one for the UK CAS team, with some new staff changes and additions as well as supporting the launch of new features and solutions from CAS.

There are now an additional four new UK team members. We have Amar Rauf, new manager of the EMEA West area, Lee Walsh our Customer Success Manager for UK South, Mourad Yahiaoui our Customer Success Manager for UK North and Ireland and lastly Shahid Qazi our new Solutions Engineer. No doubt over the course of 2020 many readers will get to meet our new team members and I am sure they will be happy to say hello and have a chat. All of our new colleagues are fully active in the field so if you want to discuss anything around your existing CAS products or other ways in which we can help, then please let us know.

Yvonne Pope and Joelle Gola have been promoted into new roles as Solutions Engineer and Senior Customer Success Manger respectively, so congratulations to them both for their success. Jonathan Agbenyega remains our Business Development Manager and does a wonderful job covering the whole of the UK. In addition to our growth in staff numbers in 2019, CAS also launched of a brand new feature in SciFinder^n, a new database in STNext, and a completely new standalone product. Early in 2020, we have also been able to announce an exciting collaboration with Scilligence. Read on for further details.

Predictive Retrosynthesis in SciFinder^n

In December 2019, CAS launched the predictive component of the retrosynthesis functionality in SciFinder^n.

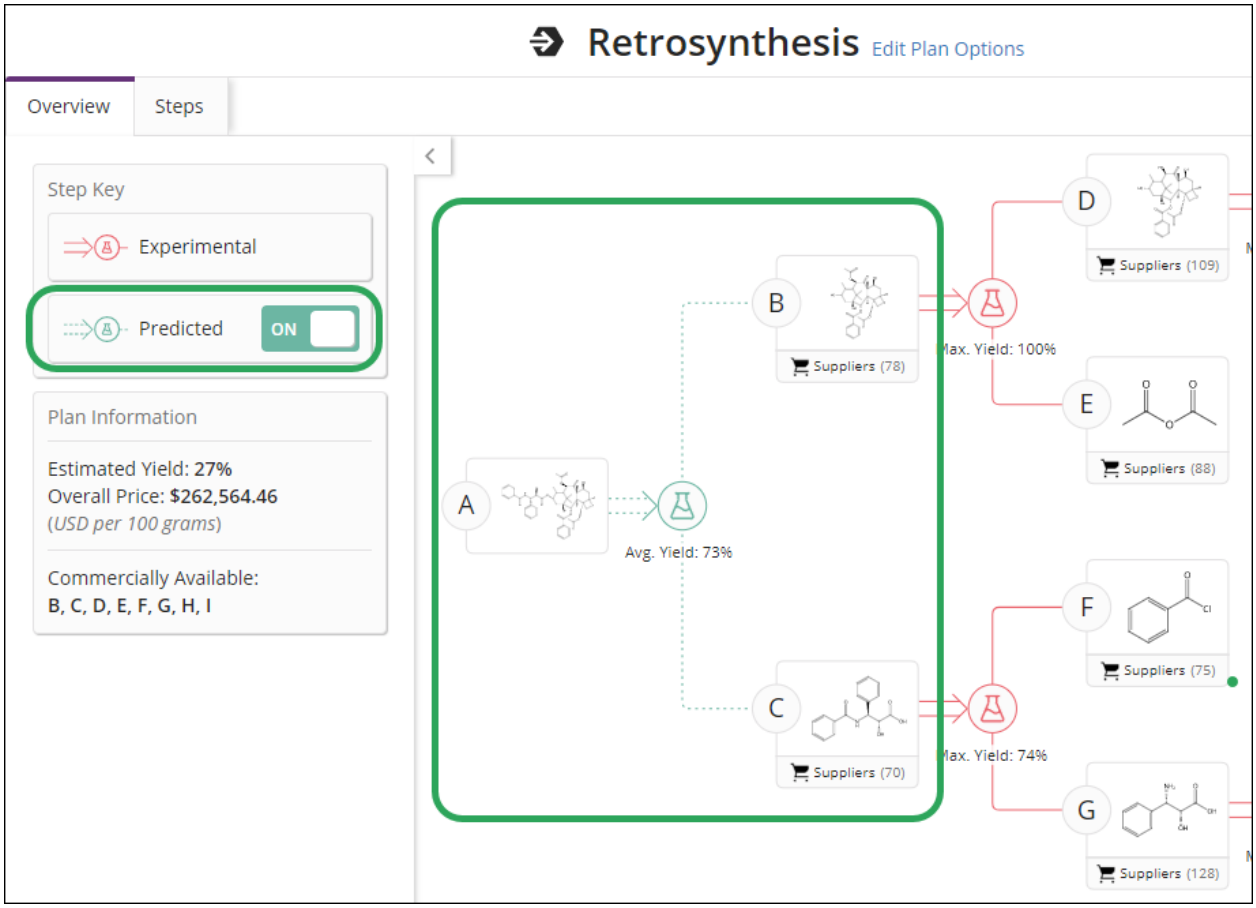
The experimental part of the retrosynthesis has been in place since June and this addition adds further power and functionality to a search. This computer-aided synthetic design (CASD) solution utilizes AI technology, powered by CAS's unmatched collection of scientist-curated reaction content and leverages John Wiley and Sons, Inc.'s award-winning ChemPlanner technology to identify predicted retrosynthetic routes for both known and novel compounds.

"*Synthetic planning is a critical step in the R&D pipeline and is often a significant bottleneck impeding speed to market*" says Tim Wahlberg, Vice President, Product Management at CAS. "*Customers have responded enthusiastically to the new retrosynthetic capabilities we launched in SciFinder^n last year, and we are excited to extend this technology with predictive enhancements that allow chemists to be significantly more innovative, more confident, and more efficient*."

The SciFinder^n retrosynthesis planner uses an advanced retrosynthetic engine to construct routes to desired compounds consisting of experimental and predicted reaction steps from 121 million reactions in the CAS

collection amassed over 110 years of chemistry research. CAS human-curated scientific content is recognised as the most comprehensive in the world. Pairing AI with this unique, high-quality dataset maximizes the power of CASD technology to provide new insights to help solve vexing synthesis challenges.

The dynamic, interactive plans developed in SciFinder^{n} allow chemists to review alternative reaction steps, providing an intuitive means to inspire new thinking, evaluate alternative synthetic strategies and compare tactical approaches. SciFinder's retrosynthetic capabilities support critical chemistry R&D workflows including synthesizing new molecular entities innovations, reaction scale up, and identifying opportunities for new breakthroughs in methods development. Expediting these key steps in the R&D pipeline will allow SciFinder^{n} customers to get more innovations to the market more quickly.



CAS FORMULATIONS™ on STNext®

CAS FORMULATIONS™ is now available exclusively on STNext. The CAS FORMULATIONS database (file label CASFORM) focuses on the chemistry content of formulations. This database was developed in response to the frequently discussed topic – that comprehensive IP searching in this area is time consuming and difficult. The CAS FORMULATIONS content collection leverages human curation to provide precise and efficient searching of formulations IP.

Our indexers revisited and reanalysed selective CAS content (patents and journals) for unique formulation information. Detailed indexing that describes substances used in product formulations means that the database includes brand new fields such as component or group optionality, component functions and forms, just to name a few. Each record represents one formulation, but also includes the bibliographic information and the location of the formulation (claims 1, 2, 3; Example 4).

If present in the record, experimental activities and processes for making the formulation are also included. An additional source of formulations also comes from the drug product inserts from the DailyMed.

The core subject areas include pharmaceuticals, agricultural chemistry and cosmetics, but there is also coverage of coatings, consumer goods food and materials.

FORMULUS

In addition to formulations content now searchable in STNext® and SciFinder^n, a new unique solution has been created with the formulators themselves in mind.

Formulus delivers the world's largest collection of formulations sourced from journals, patents and product inserts. Integrated with the unmatched CAS substance collection, detailed supplier specifications and regulatory resources, Formulus is a one-stop shop for formulators. Formulus enables users to quickly explore the formulations landscape, identify relevant formulations approaches and source ingredients.

Formulations play a critical role in bringing safe and effective new products to market. Formulus empowers innovation and ensures that formulators start the development phase on the most efficient path. Over 70% of formulators that tested Formulus confirmed that it helped them narrow focus more quickly. This reduces the overall time and resources required to optimize each formulation from early-phase clinical trials to the final marketable product.

Scilligence and CAS Collaborate to Integrate SciFinder^n with ELN

CAS has now collaborated with Scilligence to integrate SciFinder^n with Scilligence ELN (Electronic Laboratory Notebook). This integration helps scientists simplify their workflow by allowing easy access to vital research information in SciFinder^n and maintaining traceability of their ideas. Structure and reaction searches in the SciFinder^n content collection can be initiated directly from Scilligence ELN.

Integration of SciFinder^n with Scilligence ELN provides various search options directly from Scilligence ELN. Researchers can search by reaction, patent Markush, reference, substance, and supplier. Advanced algorithms developed by CAS scientists and technologists allow researchers to search smarter and find answers faster, accelerating research. The application allows users to open multiple result sets at once, and extensive filtering and Boolean search options in SciFinder^n help users quickly find the most relevant results.

With the SciFinder^n integration, scientists will have access to all the relevant information associated with their experiment within Scilligence ELN, which includes a variety of experimental templates. Users can quickly access all of the data within SciFinder^n with a click of a button directly from Scilligence ELN.

"*CAS is committed to the continuous enhancement of scientists' experience with SciFinder^n and CAS content. We welcome opportunities to integrate our solutions into our customers' workflows to improve the effectiveness of their daily activities. The integration of SciFinder^n with Scilligence ELN accomplishes this by eliminating additional steps in our users' workflow, providing actionable information more quickly and efficiently*", said Tim Wahlberg, Vice President of Product Management at CAS.

Maximize the strategic impact of scientific information via CAS Services

Innovators must navigate a more complex and interconnected information landscape than ever before. Whether drowning in an ocean of information or scouring for specialized data, customized solutions from CAS maximize the value of your digital assets, power data-driven decisions and enable you to innovate more efficiently. Capitalize on more than 110 years of experience in scientific information by talking with CAS on how we can assist with your projects from AI and Machine Learning to Workflow Integration to Advanced Analytics, Further information and Case Studies can be found on the CAS Services website. Attentively please contact Yvonne Pope (ypope@acs-i.org) for a more detailed discussion.

News from the RSC

Contribution from Jill Davidson, email: DavidsonJ@rsc.org

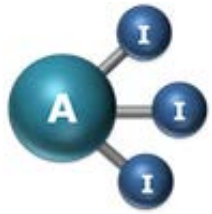
Article Submission with Time to Spare
Our submission system will save 33,000 research hours by 2023. What could you achieve with that much time? Find out more here.

Royal Society of Chemistry agrees Landmark Journals Deal with Netherlands Consortium
The Royal Society of Chemistry has agreed a new deal with SURFmarket, a Dutch consortium representing research universities across the Netherlands. The fresh, two-year agreement grants access to the full range of Royal Society of Chemistry content and services, in 2019 and 2020, while supporting the transition to open access; a shared goal between RSC and the scientific community in the Netherlands. Read the full article.

Introducing RSC Chemical Biology
We are delighted to announce that our new, gold open access journal, dedicated to publishing exceptionally significant findings in chemical biology, is now welcoming your breakthrough findings and high quality Reviews. Find out more.

News from AI3SD

Contribution from AI3SD Network+ Coordinator Dr Samantha Kanza, email: s.kanza@ai3sd.org



On 17 October 2019 AI3SD, IoFT (The Internet of Food Things Network+) and members of the Food Water and Waste Research Group at the University of Nottingham teamed up to host an event on AI for Allergen Detection and Smart Cleaning within Food Production. If this is of interest there is an excellent blog post on the event written by Michelle Pauli.

AI3SD also has interim reports from their first funding call up on the website: http://www.ai3sd.org/reports.

AI for Reaction Outcome and Synthetic Route Prediction Meeting on the 9th-11th March 2020

Registration for the *AI for Reaction Outcome and Synthetic Route Prediction Meeting* on the 9th-11th March 2020 is now closed. This is a joint meeting between AI3SD, Dial-a-Molecule & Directed Assembly! The meeting will examine the state of the art and future opportunities in the use of Artificial Intelligence to predict the outcome of unknown chemical reactions, and consequently design optimum synthetic routes to desired molecules. A wide variety of AI approaches will be illustrated, including expert systems, statistical methods, mechanism based and Machine Learning.

The meeting will also consider:
- Data sourcing, sharing, and quality
- Automated experimentation to generate reaction knowledge
- Theoretical calculations to enrich or replace experimental data

There will be talks to introduce the breadth of the area to all participants and plenary lectures from Pierre Baldi, Bartosz Grzybowski and Lee Cronin. The event will also include ample discussion sessions and opportunities to develop collaborations will be a key aspect. The full list of speakers can be found here.

The conference includes poster presentations. Anyone interested in showcasing their work through a poster, or indeed already have a relevant poster please apply to present at the conference. This is a great opportunity to showcase your research, especially if you are looking for potential collaborators. If you wish to submit a poster, please fill in the AI3SD Poster Submission Form. IKTOS have generously agreed to sponsor our poster session, and there will be a £300 prize for the top poster, and there will be prizes for the top 5 posters.

Running from 10:30 on Monday 9th March through to 4pm on Wednesday 11th March, this 3-day event is residential and registration includes 2 nights' accommodation and all meals at the De Vere Tortworth Court Hotel.

Exhibiting / Sponsoring:
This meeting has been generously sponsored so far by Reaxys, IKTOS, Molecule.one & NextMove Software. If you are interested in exhibiting/sponsoring this event please contact us.

Other Chemical Information Related News

Contributed by RSC CICAG Member Dr Keith White and RSC CICAG Newsletter Editor Stuart Newbold

[*All hyperlinks correct & working as of 28 January 2020*]

Alexa and your Phone are Getting Schooled in Chemistry
Software developers are prepping standard voice-activated assistants to work in chemical and pharmaceutical research.
https://cen.acs.org/business/informatics/Alexa-phone-getting-schooled-chemistry/97/i36
Source: Chemical & Engineering News

GSK puts Faith in AI to Make More Successful Drugs More Quickly
GlaxoSmithKline is ramping up its use of artificial intelligence and recruiting 80 AI specialists by the end of 2020 as it turns to cutting-edge computing to develop medicines of the future.
https://www.theguardian.com/business/2019/dec/25/glaxosmithkline-gsk-artificial-intellienge-ai-drug-development
Source: The Guardian

Listen to the Latest Research with Improved Accessibility on Taylor & Francis Online

Taylor & Francis Online has introduced a new feature offering an innovative text to audio option for all journal content, enriching their content for all online users, while making it more accessible to a wider range of readers. The introduction of *Readspeaker* allows logged-in users on Taylor & Francis Online to select the journal article they are looking for and listen to it via audio, simply by highlighting or hovering over specific sections, or by
pressing play to listen to the entire article from start to finish. The tool even enables users to download an MP3 recording of the journal article to listen to on the go or choose from fifteen different languages to hear it in – without the need to download any extra programs.
https://newsroom.taylorandfrancisgroup.com/listen-to-the-latest-research-with-improved-accessibility-on-taylor-francis-online/

Source: Taylor & Francis

From Database to Knowledge Graph — Using Data in Chemistry
Over the last couple of decades, the scientific community has made large efforts to process and store experimental and computational chemical data and information on the World Wide Web. This review summarizes several databases and ontologies available on the web for researchers to use. We also discuss briefly the categories of chemistry data that are stored, its main usage and how it can be accessed and understood in the framework of the Semantic Web.
https://www.sciencedirect.com/science/article/pii/S2211339819300322
Source: Current Opinion in Chemical Engineering, ScienceDirect

Notation System Allows Scientists to Communicate Polymers More Easily
In BigSMILES, polymeric fragments are represented by a list of repeating units enclosed by curly brackets. The chemical structures of the repeating units are encoded using normal SMILES syntax, but with additional bonding descriptors that specify how different repeating units are connected to form polymers. This simple design of syntax would enable the encoding of macromolecules over a wide range of chemistries.
http://news.mit.edu/2019/bigsmiles-notation-system-allows-scientists-communicate-polymers-more-easily-0918
Source: MIT News

75% of European Spending on Scientific Journals goes to 'big five' Publishers
The European Universities Association (EUA) found that overall expenditure by 26 European countries was €597 million (£515 million) in 2017. But 75% of that – some €451 million – was spent on subscriptions to journals published by the 'big five': Elsevier, Springer Nature, Wiley, Taylor & Francis and the American Chemical Society (ACS). Together they accounted for 56% of articles published.
https://www.chemistryworld.com/news/75-of-european-spending-on-scientific-journals-goes-to-big-five-publishers/4010616.article
Source: Chemistry World

5 Features of a Highly Cited Article
Mohamed Elgendi outlines key differences between highly cited and lowly cited papers.
https://www.natureindex.com/news-blog/five-features-highly-cited-scientific-article
Source: Nature Index

Machine Learning Predicts Electron Densities with DFT Accuracy
Non-covalent interactions and electron densities can be explored quickly without the need for expensive and time-consuming quantum chemical calculations.
https://www.chemistryworld.com/news/machine-learning-predicts-electron-densities-with-dft-accuracy/4010448.article
Source: Chemistry World

Abstract Images: Producing the Perfect Image to Capture your Audience
A common misconception is that the abstract image should capture the details or key results of a study; rather, it should capture an audience. Communicating and interpreting the results of your work is the job description of the paper itself. The abstract image should be the eye-catching advertisement that stops someone from scrolling away and invites them to ask questions, to look at your paper and to learn more.
https://www.advancedsciencenews.com/abstract-images-producing-the-perfect-image-to-capture-your-papers-audience/
Source: Advanced Science News

UKRI-Funded Pioneering Data Research Centres to Enable Cutting-Edge Research and Innovation to Benefit UK Patients

- Seven new data hubs to be rolled out across the UK to speed up research for new medicines, treatments, and technologies that support quicker diagnoses and save lives
- The hubs will use the latest advances in technology to connect and analyse health data from existing locations and in partnership with the NHS, ensuring data is kept safe and secure
- Patients and the public will be involved in decisions about how their data is used and accessed

https://www.ukri.org/news/ukri-funded-pioneering-data-research-centres-to-enable-cutting-edge-research-and-innovation-to-benefit-uk-patients/
Source: UK Research and Innovation

RSC Report finds Publishing Pipeline Hinders Women
Peer reviewers were more likely to reject papers from female authors, especially if the reviewer was male.
https://www.chemistryworld.com/news/rsc-report-finds-publishing-pipeline-hinders-women/4010608.article
Source: Chemistry World

New Deals Could Help Scientific Societies Survive Open Access
In the push to make the scientific literature open access, small scientific societies have feared they could be collateral damage. Many rely on subscription revenue from their journals—often among the most highly cited in their disciplines—to fund other activities, such as scholarships. And whereas big commercial publishers have the scale to absorb financial losses in some of their journals, many scientific societies operate at most a handful of journals.
https://www.sciencemag.org/news/2019/09/new-deals-could-help-scientific-societies-survive-open-access
Source: Science

With little Training, Machine-learning Algorithms can Uncover Hidden Scientific Knowledge
Researchers have shown that an algorithm with no training in materials science can scan the text of millions of papers and uncover new scientific knowledge. They collected 3.3 million abstracts of published materials science papers and fed them into an algorithm called Word2vec. By analyzing relationships between words the algorithm was able to predict discoveries of new thermoelectric materials years in advance and suggest as-yet unknown materials as candidates for thermoelectric materials.
https://www.sciencedaily.com/releases/2019/07/190703134059.htm
Source: ScienceDaily

Open Data Linked to Higher Citations for Journal Articles
Studies that provide access to underlying data are cited 25% more often than those that don't.
https://www.chemistryworld.com/news/open-data-linked-to-higher-citations-for-journal-articles/3010723.article
Source: Chemistry World

Elsevier Investigates Hundreds of Peer Reviewers for Manipulating Citations
The publisher is scrutinizing researchers who might be inappropriately using the review process to promote their own work.
https://www.nature.com/articles/d41586-019-02639-9
Source: Nature

Education Publisher Pearson to Phase out Print Textbooks
The world's largest education publisher has taken the first step towards phasing out print books by making all its learning resources "digital first". Pearson said students would only be able to rent physical textbooks from now on, and they would be updated much less frequently. The British firm hopes the move will make more students buy its e-textbooks which are updated continually.
https://www.bbc.co.uk/news/business-48998789
Source: BBC News

AI Learns the Language of Chemistry to predict how to make Medicines
Researchers have designed a machine learning algorithm that predicts the outcome of chemical reactions with much higher accuracy than trained chemists and suggests ways to make complex molecules, removing a significant hurdle in drug discovery.
https://www.sciencedaily.com/releases/2019/09/190903111250.htm
Source: ScienceDaily

The Plan to Mine the World's Research Papers
A technologist who has spent decades publishing copyrighted legal documents and then arguing that such texts should be legally within the public domain has turned his eye to paywalled scientific literature. Carl Malamud and his colleagues are building a cache of text and images extracted from 73 million journal articles dating from 1847. The goal is to allow researchers to pull insights from the papers using computerized analysis, without actually reading them — meaning that they won't be breaching publishers' copyright.
https://www.nature.com/articles/d41586-019-02142-1
Source: Nature

World Health Organization joins Controversial Open Access Plan S
The Who becomes first UN agency to support the push for open access science publishing.
https://www.chemistryworld.com/news/world-health-organization-joins-controversial-open-access-plan-s/3010925.article
Source: Chemistry World

RSC Report: The State of Open Data 2019
The Royal Society of Chemistry recently convened a panel to react to the updated guidance for Plan S – an initiative supported by research funders to accelerate the transition to full open access scientific publishing. The RSC's panel discussion revealed a tug of war between two concepts:

- Plan S is a great idea and a move in the right direction; all published scientific research should be publicly available and not profit-driven
- The implementation of Plan S – particularly in restricting where researchers will be able to publish – leaves significant potential for unintended consequences that could hurt scientific collaborations and the career development of early career researchers

See: https://www.rsc.org/news-events/articles/2019/jun/plan-s-livestream/, which include a video of the discussion.
Source: RSC

Mapping Science Using Microsoft Academic Data
One of the most exciting developments in the past few years in the field of bibliometrics is the emergence of a number of important new data sources. Dimensions, created by Digital Science and made openly available for research purposes, is a prominent example. Other examples are Crossref and OpenCitations, which provide data that is fully open. The launch of Microsoft Academic in 2016 also represents a significant development. In this blog post, we discuss the data made available by Microsoft Academic and we show how the most recent version of our VOSviewer software can be used to create science maps based on this data.
https://www.cwts.nl/blog?article=n-r2x284
Source: University of Leiden

A Journal Club to Fix Science
ReproducibiliTea can build up open science without top-down initiatives.
https://www.nature.com/articles/d41586-019-02842-8
Source: Nature

How to Manage Research Data
Managing research data is becoming increasingly complex. New concepts and standards are needed to share data from experiments or computations and to reuse them.
https://www.chemistryviews.org/details/ezine/11185826/How_to_Manage_Research_Data.html
Source: ChemViews Magazine

New eReader Trial Launched for Taylor & Francis Journal Articles
The ability to read journal articles in EPUB format using the new eReader is being trialled on 22 selected Taylor & Francis journals. The eReader introduces a new way of reading, discovering, sharing and interacting with journal articles on tandfonline.com. It provides immersive, uncomplicated reading in any browser and on any device, online or offline, and is particularly beneficial for mobile users due to the dynamic and responsive nature of the EPUB file format.
https://newsroom.taylorandfrancisgroup.com/new-ereader-trial-launched-for-taylor-francis-journal-articles/
Source: Taylor & Francis

From Artificial Intelligence and AI Scientists to Pharmaceutical Analytics – 2019 Catalyst Grant Winners
BPT Analytics, Intoolab and MLprior, three projects aiming to disrupt the academic space, are the latest recipients of the Catalyst Grant award for innovative start-ups.
https://www.digital-science.com/press-releases/from-artificial-intelligence-and-ai-scientists-to-pharmaceutical-analytics-2019-catalyst-grant-winners/
Source: Digital Science

Clarivate Analytics Acquires the SequenceBase Business to expand its Search and analysis Capabilities for the Life Sciences IP Market
The acquisition will deliver an expanded solution to the fast-growing biologics market. It complements Derwent's GENESEQ™ database and leading patent research platform Derwent Innovation™.
https://clarivate.com/news/clarivate-analytics-acquires-the-sequencebase-business-to-expand-its-search-and-analysis-capabilities-for-the-life-sciences-ip-market/
Source: Clarivate Analytics

Turing AI Fellowships to keep UK at Forefront of Artificial Intelligence Revolution
World-leading researchers will develop pioneering new approaches to Artificial Intelligence (AI) through a new set of fellowships announced by UK Research and Innovation (UKRI). UK and international researchers from a diverse range of backgrounds are encouraged to apply to the next wave of Turing AI Fellowships, named after AI pioneer Alan Turing, which will support the very best with £36 million of funding.
The first five Turing AI Fellows, funded through the AI Sector Deal, have been announced by The Alan Turing Institute, the UK's national institute for AI and data science. Their research will have direct relevance to areas ranging from mental health and the aerospace industry to astrophysics. The fellowship investment is part of a comprehensive AI talent initiative that also includes 16 UKRI AI Centres for Doctoral Training, established with a £200 million UKRI, industry and university investment at 14 universities with 300 partners across industry. The first 200 students have started at the centres this autumn.
https://www.ukri.org/news/turing-ai-fellowships-to-keep-uk-at-forefront-of-artificial-intelligence-revolution1/
Source: UK Research and Innovation

Improving the Odds of Synthetic Chemistry Success
In a new publication in Nature, University of Utah chemists Jolene Reid and Matthew Sigman show how analyzing previously published chemical reaction data can predict how hypothetical reactions may proceed, narrowing the range of conditions chemists need to explore. Their algorithmic prediction process, which includes aspects of machine learning, can save valuable time and resources in chemical research.
https://www.sciencedaily.com/releases/2019/07/190717132752.htm
Source: ScienceDaily

Automated Research Data Management
Many scientists are asking themselves how to effectively organize research data. In the light of increasingly extensive, complex, and heterogeneous data from experiments or simulations, the question gets more urgent. According to good scientific practice, data should be stored for ten years. However, it is unclear in which data formats and where data can be stored in the long term. Publishing data as part of articles or supplements, only to a certain extent meets the requirements and chances of data re-use. The availability of machine-readable data and metadata for emerging research fields in the context of big data and AI algorithms is becoming increasingly important.
https://www.chemistryviews.org/details/ezine/11167842/Automated_Research_Data_Management.html
Source: ChemViews Magazine

KAUST signs Read & Publish Agreement with Royal Society of Chemistry
The Royal Society of Chemistry has agreed a new Read & Publish deal with the King Abdullah University of Science & Technology (KAUST), in Saudi Arabia, becoming the 10th country to have agreed such a deal.
https://www.rsc.org/news-events/articles/2019/jul/kaust-agreement-2019/
Source: RSC

Machine-learning Mendeleevs have Rediscovered the Periodic Table
Since Dmitri Mendeleev (and others) first sketched out the periodic relationships between the elements in the 1860s, it has been estimated that around a thousand different tables have appeared in print – and that's before considering all those on the internet. Mostly these tables embody careful deliberation about what to put where, which information to prioritise, which message to convey. But two recent papers outlining machine learning techniques have shown that it is now possible to rediscover the table empirically, from the way it is implicitly embedded within the milieu of chemistry.
https://www.chemistryworld.com/opinion/machine-learning-mendeleevs-have-rediscovered-the-periodic-table/3010720.article
Source: Chemistry World

Springer Nature and Allen Institute for Artificial Intelligence (AI2) Expand Collaboration to cover Chemistry, Earth Sciences, Public Health and More
Semantic Scholar, AI2's free AI-based academic search engine used by over 4 million researchers each month, partners with Springer Nature to power advanced discovery of academic research content.
https://group.springernature.com/gp/group/media/press-releases/springer-nature-and-allen-institute-for-artificial-intelligence/17017996
Source: Springer-Nature

DOE to Provide $27.6 Million for Data Science Research in Chemical and Materials Sciences
The 19 awards, 14 to universities and 5 to U.S. Department of Energy national laboratories, will advance the application of modern data science techniques such as artificial intelligence and machine learning to develop new materials and chemical processes.
https://www.eurekalert.org/pub_releases/2019-08/ddoe-dtp081519.php
Source:EurekAlert! Science News

Artificial Intelligence Promises to Enhance Sustainable Investing
Computers and data scientists can unearth key themes missed by traditional research. While some companies might not report on their carbon usage, for example, there are other ways to see what they are doing, Jennifer Wu, head of ESG investing at JPMorgan Asset Management said. By using AI to scrape for keywords in patent filings, JPMorgan can now identify companies linked to new low-carbon technology.
Restricted link: https://www.ft.com/content/7c40cdfc-b528-11e9-bec9-fdcab53d6959
Source: Financial Times

Machine learning's next Frontier: Epigenetic Drug Discovery: Scientists create a Machine-Learning Algorithm that Automates High-Throughput Screens of Epigenetic Medicines
Scientists have developed a machine-learning algorithm that gleans information from microscope images, allowing for high-throughput epigenetic drug screens that could unlock new treatments for cancer, heart disease, mental illness and more.
https://www.sciencedaily.com/releases/2019/10/191022080734.htm
Source: ScienceDaily

Growth in Data and Questions on Quality are Increasing Researcher Workload, Finds New Study from Elsevier and Sense about Science
Data proliferation and a lack of confidence in the multitude of research outputs is adding to researcher workload and is likely to impact public confidence in science, according to a survey of the global research community on trust in research – one of the largest surveys known to date.
https://www.prnewswire.co.uk/news-releases/growth-in-data-and-questions-on-quality-are-increasing-researcher-workload-finds-new-study-from-elsevier-and-sense-about-science-811849124.html
Source: Cision PR Newswire

Hundreds of Extreme Self-citing Scientists Revealed in New Database
Some highly cited academics seem to be heavy self-promoters — but researchers warn against policing self-citation.
https://www.nature.com/articles/d41586-019-02479-7
Source: Nature

New Co-Owners for Chemistry Preprint Server ChemRxiv
The Gesellschaft Deutscher Chemiker (GDCh, German Chemical Society), the American Chemical Society (ACS), and the Royal Society of Chemistry (RSC) have announced a partnership with the Chinese Chemical Society (CCS) and the Chemical Society of Japan (CSJ) as co-owners to support the strategic and financial development of ChemRxiv, a preprint server for the global chemistry community.
https://www.chemistryviews.org/details/news/11179556/New_Co-Owners_for_Chemistry_Preprint_Server_ChemRxiv.html
Source: ChemistryViews

Retrosynthetic Algorithm Broadened to Design Similar, but Different, Molecules
An algorithm for performing challenging multistep retrosynthesis and predicting the most efficient synthetic pathway has been extended to libraries of compounds with different substitutions or isotope labels. This broadens the utility of the computer program Chematica, which interweaves mechanistic rules, quantum calculations and artificial intelligence to predict the best synthesic routes.
https://www.chemistryworld.com/news/retrosynthetic-algorithm-broadened-to-design-similar-but-different-molecules/3010882.article
Source: Chemistry World

Data- and AI-Related Topics Grew more than 50 Percent from Last Year on O'Reilly's Learning Platform
Data Scientists are Upskilling and Reskilling to Better Understand and Use AI and Machine Learning Technologies.
https://www.oreilly.com/pub/pr/3283
Source: O'Reilly

Industry Leaders' Urgent call for Government to keep the UK Chemistry Talent Pipeline Flowing
Industry leaders have called on the government and education bodies to help school pupils discover chemistry and its many and varied career paths as UK applications to chemistry courses fall by over 20%.
https://www.rsc.org/news-events/articles/2019/sep/aspires-industry-skills-pipeline/
Source: RSC

Open Letter from Industry Leaders
The UK is one of the most important countries in the world for chemical science, but that reputation is intrinsically tied to having a strong pipeline of home grown talent. That this pipeline looks to be slowing because pupils don't know what career options come from studying chemistry is a matter of significant concern for a number of sectors deemed crucial to the country's economy.
https://www.rsc.org/news-events/articles/2019/sep/open-letter-from-industry-leaders/
Source: RSC

Fears for Britain's Standing in World of Science as Students Shun Chemistry Degrees
Business leaders urge government to take action after university applications fall 21% in three years.
https://www.theguardian.com/education/2019/aug/31/students-shun-chemistry-degrees-university-applications-fall
Source: The Guardian

Chemical Safety Database gets American Chemical Society and IUPAC Backing
Library of hazardous reactions and lab near misses has found partners to develop it and help keep it afloat.
https://www.chemistryworld.com/news/chemical-safety-database-gets-american-chemical-society-and-iupac-backing/3010932.article
Source: Chemistry World

China to Boost Domestic Science Journals
The agencies vowed to support locally published English-language journals to compete with international journal publishers, boost Chinese-language journals, increase the weight of Chinese-language articles in scientist promotion assessments, encourage the formation of domestic journal publishing groups, increase the digitization of journal content, and recruit globally for journal management.
https://pubs.acs.org/doi/abs/10.1021/cen-09735-polcon3
Source: Chemical & Engineering News

Doctoral Training in Artificial Intelligence for Healthcare
On 1st October 2019, the UKRI Centre for Doctoral Training in Artificial Intelligence for Healthcare opened its doors to their first cohort.
https://www.imperial.ac.uk/news/192847/doctoral-training-artificial-intelligence-healthcare/
Source: Imperial College London

Clarivate Analytics Partners with Hitachi to launch new Solution to Improve the Efficiency of Global Patent Research, Review and Evaluation Processes in Japan
Customers of Hitachi's Shareresearch, a market-leading Japanese patent information search platform for Intellectual Property (IP) and Research & Development (R&D) teams, will now have the option to access Derwent World Patents Index (DWPI) content translated into Japanese within the platform. Shareresearch users can now streamline the patent review process when working with large numbers of overseas patent documents in foreign languages. Used by more than 40,000 users and 40 patent offices around the world, DWPI is a trusted source of global patent data, correcting and summarizing the novelty, use and advantage of an invention for more than 84 million worldwide patents, covering 41 million patent families in more than 26 languages.
https://clarivate.com/blog/news/clarivate-analytics-partners-with-hitachi-to-launch-new-solution-to-improve-the-efficiency-of-global-patent-research-review-and-evaluation-processes-in-japan/
Source: Clarivate Analytics

Human Biases Cause Problems for Machines trying to Learn Chemistry
Including 'unpopular' reagents and reaction conditions into datasets could lead to better machine-learning models
https://www.chemistryworld.com/news/human-biases-cause-problems-for-machines-trying-to-learn-chemistry/3010970.article

Source: Chemistry World

AI to Enhance Manufacturing, Energy, and Healthcare Thanks to £6.5m EPSRC Grant
The grant will help scientists, engineers, and clinicians train algorithms to solve energy, manufacturing, and healthcare problems.
https://www.imperial.ac.uk/news/192947/ai-enhance-manufacturing-energy-healthcare-thanks/
Source: Imperial College London

Chemistry Team Leverages 'Data Revolution' to Solve Current Issues in Chemistry
Part of three projects chosen nationally this year by the National Science Foundation, Matthew Sigman and his team will help create a new generation of data chemists.
https://www.eurekalert.org/pub_releases/2019-09/uou-ctl091319.php
Source: EurekAlert! Science News

CAS Accelerates Product Development with Launch of Specialised Formulations Solutions
Formulus®, a new solution aimed at getting innovations to market faster by addressing the unique information needs of formulations scientists.
https://www.cas.org/resources/press-releases/formulus-product-launch
Source: CAS

Novartis and Microsoft Announce Collaboration to Transform Medicine with Artificial Intelligence

- The multiyear alliance underpins the Novartis commitment to leverage data & artificial intelligence (AI) to transform how medicines are discovered, developed and commercialised.
- Novartis to establish AI innovation lab to empower its associates to use AI across the business
- Joint research activities will include co-working environments on Novartis Campus (Switzerland), at Novartis Global Service Center in Dublin, and at Microsoft Research Lab (UK) – starting with tackling personalised therapies for macular degeneration; cell & gene therapy; and drug design

https://news.microsoft.com/2019/10/01/novartis-and-microsoft-announce-collaboration-to-transform-medicine-with-artificial-intelligence/
Source: Microsoft.com

Universities Unite in High-Level Strategy Group to Govern Negotiations with Publishers
Universities UK has convened a new group to support and enhance the mandate of Jisc in negotiating publishing agreements with major academic publishers. The formation of this new, high-level group recognises the importance of involving senior leaders from UK institutions in order to achieve sustainable, affordable and innovative agreements with major publishers.
https://www.jisc.ac.uk/news/universities-unite-in-high-level-strategy-group-to-govern-negotiations-with-publishers-25-nov-2019
Source: Jisc

Language-based Software's Accurate Predictions Translate to Benefits for Chemists
The state-of-the-art design for computer language processing, results in improved models for predicting chemistry.
https://www.chemistryworld.com/news/language-based-softwares-accurate-predictions-translate-to-benefits-for-chemists/4010437.article
Source: Chemistry World

Web of Science Group Relaunches Master Journal List of 21k Journals
The relaunch introduces a new way to browse, search, and explore the only complete authorised list of journals in the Web of Science index.
https://clarivate.com/webofsciencegroup/news/web-of-science-group-relaunches-master-journal-list-of-21k-journals/
Source: Web of Science Group

International Collaboration Generates High Quality Clusters of Pharmaceutical Patents
The worldwide active patent portfolio has nearly doubled in numbers and strength since 2000. The number of active pharmaceutical patent families has tripled in the same time period. The quantitative growth results mostly from a surge of patents from China.
https://www.future-science.com/doi/full/10.4155/ppa-2019-0017
Source: Future Science

A Machine Learning Framework for the Analysis and Prediction of Catalytic Activity from Experimental Data
The paper describes a machine learning framework to explore the predictability limits of catalytic activity from experimental descriptor data (which characterizes catalyst formulations and reaction conditions). Artificial neural networks are used to fuse descriptor data to predict activity and we use principal component analysis (PCA) and sparse PCA to project the experimental data into an information space and with this identify regions that exhibit low- and high-predictability.
https://www.sciencedirect.com/science/article/abs/pii/S0926337319310045
Source: Applied Catalysis B: Environmental, ScienceDirect

Cray named as Hardware Provider for UK's Most Powerful Supercomputer
Cray has been awarded the contract to supply the hardware for the next national supercomputer, ARCHER2, following a procurement exercise delivered by UK Research and Innovation (UKRI). ARCHER2 represents a significant step forward in capability for the UK's science community, with the system expected to sit among the fastest fully general purpose (CPU only) systems in the world when it comes into service in May 2020. It will be able to achieve performance levels that are eleven times higher (on average, across a range of benchmarking codes) than the current ARCHER system. The £79 million new resource will be housed at the University of Edinburgh's Advanced Computing Facility. The project is being delivered and supported by UKRI's Engineering and Physical Sciences Research Council (EPSRC) and Natural Environment Research Council (NERC).
https://www.ukri.org/news/cray-named-as-hardware-provider-for-uks-most-powerful-supercomputer/
Source: UK Research and Innovation

Careless Citations don't just Spread Scientific Myths – they can make them Stronger
How misconceptions persist and proliferate within the scientific literature.
https://www.natureindex.com/news-blog/misciting-scientific-myths-spread-strengthen-hawthorne-effect
Source: Nature Index

How long do Scientists Spend Formatting Manuscripts for Publication Every Year?
Study puts a number on the cost in time and money.
https://www.chemistryworld.com/news/how-long-do-scientists-spend-formatting-manuscripts-for-publication-every-year/4010582.article
Source: Chemistry World

Springer Nature first Publisher to Implement SeamlessAccess.org Service
As part of its commitment to finding solutions to problems researchers are facing, Springer Nature has been working with others in the industry and in collaboration with SeamlessAccess.org to address the current sometimes confusing and frustrating experience researchers face trying to access research subscribed to by their university when outside of the institution's network.
https://group.springernature.com/gp/group/media/press-releases/springer-nature-first-publisher-to-implement-seamlessaccess-org/17372856
Source: Springer-Nature

Scholastica Announces Automated Crossref DOI Registration for Publishing Platform Users
Scholastica, the peer review and publishing platform for academic journals, has enabled those journals using Scholastica for open access publishing to now set up automated DOI registration and metadata deposits via Crossref.
https://www.eurekalert.org/pub_releases/2019-11/s-saa110619.php
Source: EurekAlert! Science News

Gender Diversity Quotas to be introduced for Royal Society of Chemistry Events
A third of speakers and chairs at conferences funded by the society will have to be women in 2020.
https://www.chemistryworld.com/news/gender-diversity-quotas-to-be-introduced-for-royal-society-of-chemistry-events/4010665.article
Source: Chemistry World

New ACS Guide to Scholarly Communication
In light of a rapidly changing publishing landscape and an expanding definition of communications within the scholarly research community, the Publications Division of the American Chemical Society (ACS) is proud to announce the new ACS Guide to Scholarly Communication. It is a significant expansion of the highly renowned ACS Style Guide, updated to include essential assistance to researchers at various stages in their career — from students who are learning about scholarly publishing and communications to seasoned research professionals, educators and librarians.
https://pubs.acs.org/doi/book/10.1021/acsguide
Source: ACS Publications

Global Highly Cited Researchers 2019 List Reveals Top Talent in the Sciences and Social Sciences
Researchers in the US continue to dominate the list; however mainland China is now home to the second largest concentration of highly cited researchers, superseding the UK.
https://clarivate.com/news/global-highly-cited-researchers-2019-list-reveals-top-talent-in-the-sciences-and-social-sciences/
Source: Clarivate Analytics

The Science Institutions Hiring Integrity Inspectors to Vet their Papers
Some researchers have their manuscripts screened for errors before they go to journals.
https://www.nature.com/articles/d41586-019-03529-w
Source: Nature

Google Health-Data Scandal Spooks Researchers
Google and one of the largest health-care networks in the US are embroiled in a data-privacy controversy that researchers fear could jeopardise public trust in data-sharing practices and, potentially, academic studies. At issue is a project dubbed Nightingale that gives Google access to the health-care information, including names and other identifiable data, of tens of millions of people without their knowledge. The people were treated at facilities run by the health network Ascension, which is based in St Louis, Missouri.
https://www.nature.com/articles/d41586-019-03574-5
Source: Nature

To fix Research Assessment, Swap Slogans for Definitions
Evaluation reforms will go round in circles without conceptual clarity.
https://www.nature.com/articles/d41586-019-03696-w
Source: Nature

CCS Chemistry has been collected by the CAS database
The flagship journal of the Chinese Chemical Society has been formally collected by the Chemical Abstracts Service, CAS database. The Index will include all papers published since CCS Chemistry launched.
https://www.chinesechemsoc.org/do/10.5555/d8993f37-2020-40a8-9dc6-37f89fe393c6/full/

Source: Chinese Chemical Society

AIs Direct Search for Materials Breakthroughs
Artificial intelligence (AI) and robotics are combining to tackle a big problem in the discovery of new materials. With more than 100 elements in the periodic table and the ability to combine them in virtually limitless ways, the number of possible materials is daunting.
https://science.sciencemag.org/content/366/6471/1295
Source: Science

National Science Review, published by Oxford University Press, moves to fully Open Access
Oxford University Press and China Science Press are delighted to announce National Science Review (NSR) now publishes as a fully Open Access journal. Launched in 2014, NSR is a scientific journal published in English under the auspices of the Chinese Academy of Sciences and aimed at reporting cutting-edge developments across science and technology in China and around the world. The latest Impact Factor for NSR is 13.222, according to the latest Journal Citation Reports (Clarivate Analytics, 2019).
https://academic.oup.com/journals/pages/announcements_from_oup/national_science_review
Source: Oxford Academic Journals

Articles in 'Predatory' Journals Receive few or no Citations
Many scientists have decried the rise of "predatory" journals—open-access publications that charge authors to publish but offer little or no peer review or other quality controls and use aggressive marketing tactics. Scholars have worried that the resulting articles have contaminated the literature with mediocre, flawed, or intentionally misleading findings. But a new study finds that 60% of articles published in a sample of "predatory journals" attracted not a single citation over a 5-year period. And the articles that received citations did so at a rate much lower than papers in conventional peer-reviewed journals.
https://science.sciencemag.org/content/367/6474/129
Source: Science

Eleven tips for Working with Large Data Sets
Big data are difficult to handle. These tips and tricks can smooth the way.
https://www.nature.com/articles/d41586-020-00062-z
Source: Nature