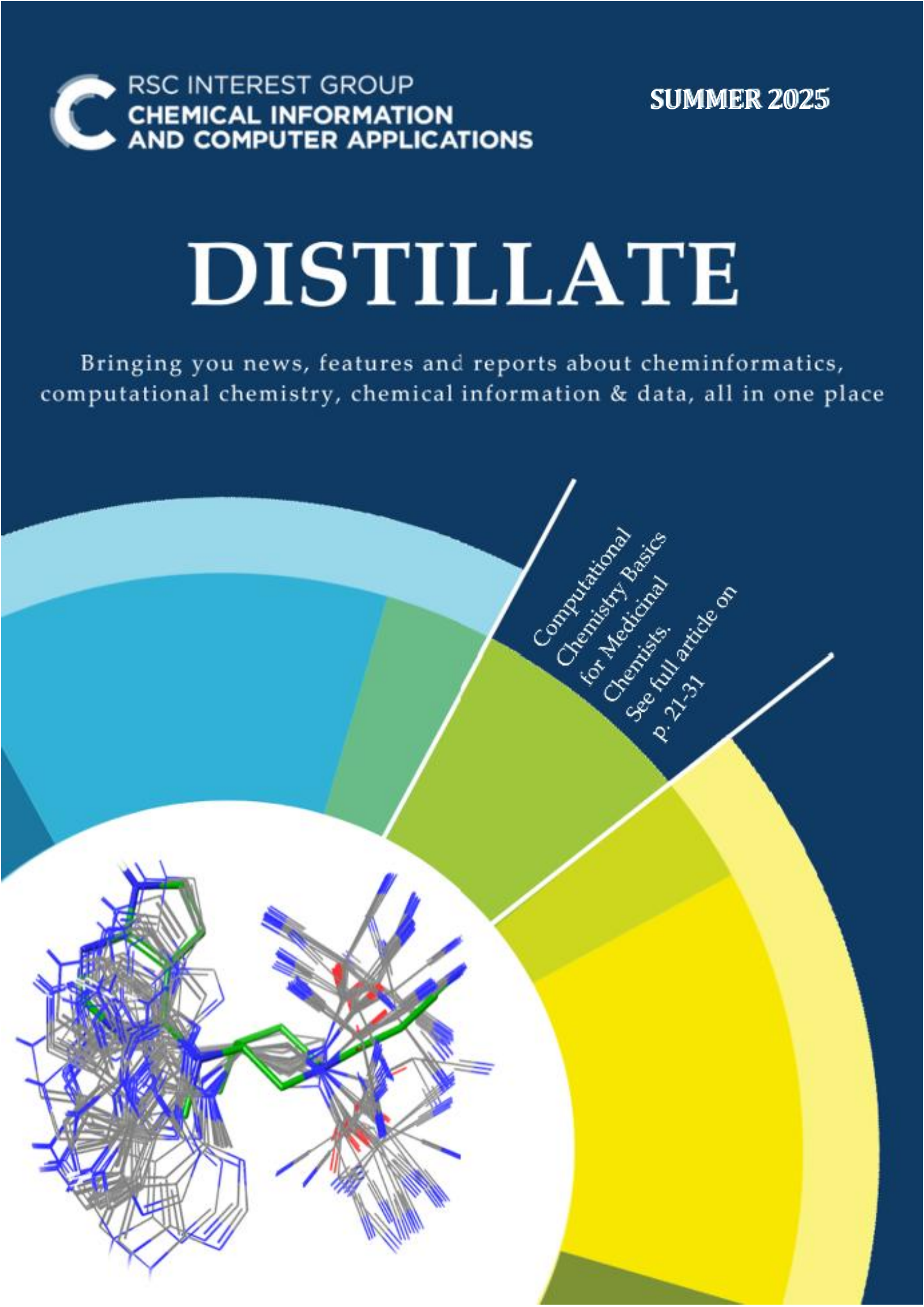
RSC INTEREST GROUP
CHEMICAL INFORMATION
AND COMPUTER APPLICATIONS

SUMMER 2025

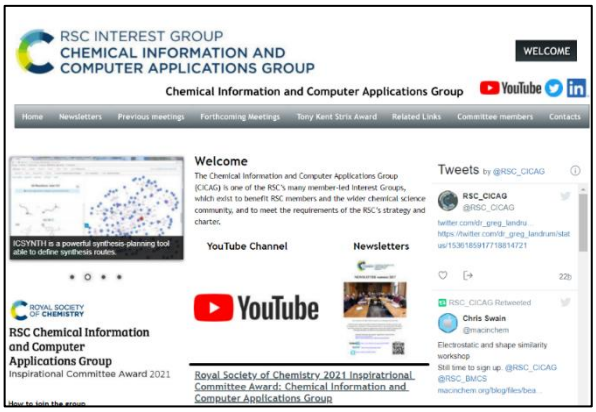
DISTILLATE

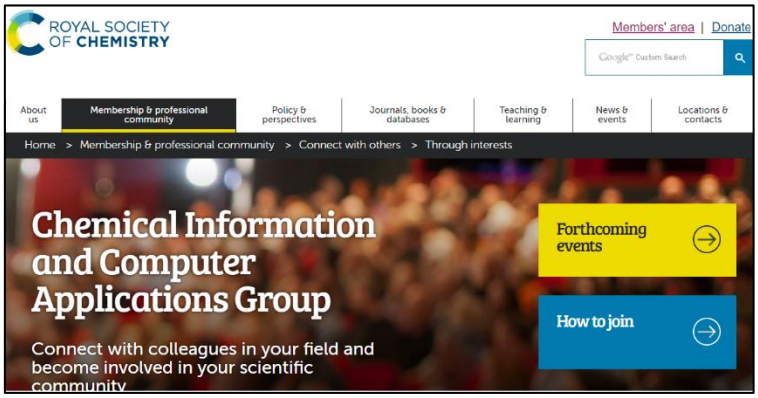
Bringing you news, features and reports about cheminformatics, computational chemistry, chemical information & data, all in one place

Computational Chemistry Basics for Medicinal Chemists.
See full article on p. 21-31



CICAG Websites and Social Media





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

rsc-cicag page
rsc-cicag group

@rsc-cicag.bsky.social

@rsccicag

@RSCCICAG

Contents

Chemical Information and Computer Applications Group Chair's Report 3
CICAG Planned and Proposed Future Meetings 3
Copyright for the *CICAG Distillate* 4
Collaboration with *Chemistry International* 5
Professor Michael Felix Lynch 5
DUST Conference 2025: Advancing Scientific Terminology & Standards for the Digital Era 8
CICAG Members' Publications: Opportunity to Publicise your Research! 9
Mike Lynch Award 2025: Professor Val Gillet 12
Meeting Report: ChEMBL@15 Symposium 13
Meeting Report: SCI-RSC Workshop on Computational Tools for Drug Discovery 2025 14
Modernising the InChI and Bridging the age-old Gap to Inorganic Chemistry 15
Computational Chemistry Basics for Medicinal Chemists. Part 3: Conformational Analysis Worked Example 21
Our Gem in Widnes: The Catalyst Science Discovery Centre & Museum 30
Communicating Chemistry: 'The Wonders of Colour' Exhibition and Activities 35
NFDI4Chem: Tools for the whole Research Data Lifecycle in Chemistry 35
Professor Graham Richards 38
Book Reviews 40
Chat GPT – Friend or Foe? 43
Hot Topics: Robotics and Automation Meeting Report 45
Celebrating the Life and Legacy of Dr Olga Kennard 46
Cheminformatics and Chemical Information Books 62
Careers and Skills for Data-driven Research Network+ News 63
Advancing Open Hardware for Autonomous Discovery 67
Martin Karplus: a Pioneering Theoretical Chemist who Brought Molecular Dynamics Simulations to Biology 68
RSC Chemistry Databases Update 70
A CICAG FAIR Cookery Recipe Book for Xmas 71
Physical Sciences Data Infrastructure (PSDI) News 71
Cambridge Structural Database (CSD) Updates 79
Update from the AIchemy Hub 81
Catalyst Science and Discovery Centre and Museum News 84
News from CAS 87
The Open Reaction Database 89
Molecular and Materials Modelling at Newcastle University (2015-2025) 92
UKeIG: 2025 Tony Kent Strix Award 98
Other Chemical Information News 99

Contributions to the CICAG *Distillate* are welcome from all sources – please contact the editor Dr Helen Cooke FRSC: email helen.cooke100@gmail.com

Chemical Information and Computer Applications Group Chair's Report

Contribution from RSC CICAG Chair Dr Chris Swain, email: swain@mac.com

The current CICAG membership now stands at 864, up from 387 in 2018. The increase in membership is due to the increased importance of cheminformatics and computational chemistry and also to the continued hard work of the CICAG committee in organising conferences, workshops, webinars and reaching out to the chemistry community. In our discussions with other interest groups and communities it is clear there are opportunities for joint meetings and requests for more workshops. To meet this increased demand, we initiated the process to recruit new members to the committee. We were delighted to receive 14 really excellent applications, the RSC Networks team duly organised an election and Carmen Domene, Garrett M. Morris and Alex Porter were elected at the start of June, we look forward to working with them. Many thanks to all applicants and to everyone who voted in the election.

In April we held the annual SCI-RSC Workshop on Computational Tools for Drug Discovery in Leeds. This meeting gives users the opportunity to try out a range of commercial tools.

The next meeting is our flagship event, the 8th CICAG/BMCS Artificial Intelligence in Chemistry. This meeting increases in size every year and has a stella lineup of speakers. In addition, it will build on 2024's popular "Introduction to AI in Chemistry" workshop on Day 1, this time with an additional introduction to Python/Jupyter notebooks. Last year CICAG and BMCS gave a total of 11 bursaries to attend this meeting and applications for 2025 are open.

The 9th AI in Chemistry meeting is planned for September 2026.

AI has generated a huge amount of interest, and I'm delighted that one of our committee members has contributed a thought-provoking article on the ethics of AI (page 43), it is easy to be seduced by the latest new toy and Sami Pearman-Kanza's article is a welcome reality check.

As I mentioned, discussions with other groups have identified several possibilities for joint meetings; these are detailed in the Planned and Proposed Meetings section on p.3-4.

The *Distillate* continues to add unique content, and we now have a number of regular contributors which are greatly appreciated. We welcome contributions that would be of interest to the CICAG community. Please contact the *Distillate* Editor, Helen Cooke, or me to discuss your ideas.

- -

CICAG Planned and Proposed Future Meetings

The table below provides a summary of planned scientific and educational meetings which CICAG is currently organising or contributing to. For more information, please contact CICAG's Chair, Dr Chris Swain.

Meeting	Date	Location	Further Information
8th RSC CICAG-BMCS AI in Chemistry meeting	22-24 Sept 2025	Churchill College, Cambridge	Meeting web page

The 9th AI in Chemistry meeting is planned for Sept 2026.

 © RSC Chemical Information & Computer Applications Special Interest Group

CICAG is also considering the following meetings and events in late 2025 or 2026, some of which may be arranged jointly with other groups or organisations (details to be confirmed):

- Python for Chemists
- Open Chemical Sciences
- 2nd AI4proteins
- AI for Materials Chemistry
- 2nd Molecular Simulations
- Nomenclature for the Physical Sciences
- Centenary of Markush Structures
- CICAG AGM

Copyright for the CICAG *Distillate*

Contribution from Helen Cooke, CICAG Distillate editor, email: helen.cooke100@gmail.com

The CICAG Committee has been considering re-categorising the copyright status for the CICAG *Distillate*, to clarify the rights of authors and readers. As it stands at present, there is no particular type of licence attached to the CICAG newsletters, either on the RSC's website or CICAG's own website. Articles in the *Distillate* (and its predecessor the CICAG *Newsletter*) are currently free to read, but are not free to reuse.

To clarify the situation, we have investigated copyright options and propose to use the CC BY-NC-ND 4.0 categorisation for the *Distillate*, which states that readers are free to:

Share – copy and redistribute the material in any medium or format
The licensor cannot revoke these freedoms as long as you follow the licence terms.

Under the following terms:

Attribution – You must give appropriate credit, provide a link to the licence, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

Non-commercial – You may not use the material for commercial purposes.

No Derivatives – If you remix, transform, or build upon the material, you may not distribute the modified material.

No additional restrictions – You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

The CICAG Committee agreed that the current copyright statement and terms will be retained for the Summer 2025 *Distillate* issue. The new licence will be applied to the Winter 2025/26 issue.

Collaboration with *Chemistry International*

Contribution from Helen Cooke, email: helen.cooke100@gmail.com, Chris Swain, RSC-CICAG Chair, Jeremy Frey
School of Chemistry and Chemical Engineering, University of Southampton



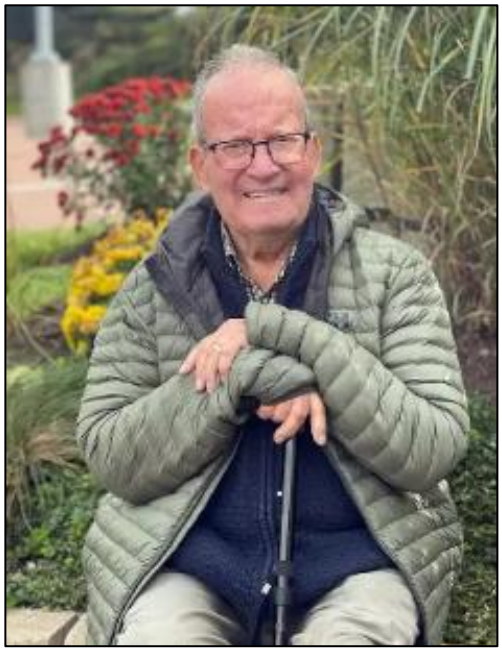
The CICAG Committee is very pleased to have been contacted by Dr Fabienne Meyers, managing editor of *Chemistry International* (*CI*), the news magazine of the International Union of Pure and Applied Chemistry (IUPAC). *CI* includes news about IUPAC, its chemists, publications, recommendations, conferences and the work of its Divisions and Committees along with special feature articles. Published quarterly, *CI* is an online-only publication freely accessible to all readers.

We have agreed that *CI* will from time to time repurpose some content from the CICAG *Distillate* by republishing selected articles, along with links and references. CICAG will seek agreement from the authors before articles first published in the *Distillate* are included in *CI*.

The CICAG Committee sees this as an excellent opportunity and it is planned that the arrangement will start with the October 2025 issue of *Chemistry International*.

Professor Michael Felix Lynch
21 February 1932 – 15 November 2024

Contribution from Dr Wendy Warr, Wendy Warr & Associates, email: wendy@warr.com and Professor Peter Willett, email: pwillettbjmt@gmail.com



One of the pioneers of chemical information, Michael (Mike) Felix Lynch, has died at the age of 92. Mike obtained BSc and PhD degrees in chemistry at University College, Dublin before conducting postdoctoral research at ETH Zürich with Vladimir Prelog, the recipient of the 1975 Nobel Prize for chemistry. He then worked for two years with Ciba-Geigy in Cambridge UK, before moving to the United States in 1961 to work for Chemical Abstracts Service (CAS). At this time, CAS was conducting some of the first research anywhere on the use of computers to build searchable databases of both journal abstracts and of the structures of chemical compounds, and Mike directed much of this work when he became Head of the Basic Research Department.[1,2]

On returning to the UK in 1965, he was awarded a research grant by the Office of Scientific and Technical Information to conduct studies of the automatic indexing of textual documents. He joined the University of Sheffield's Postgraduate School of Librarianship (as it was then named, now the Information School) to conduct this research, and spent the rest of his career there until his retirement in 1997. Mike can truly be regarded as the founding father of information science research at Sheffield: starting as a Senior Research Fellow, he joined the permanent staff in 1968, and in 1974 he

was appointed to the personal chair which made him the first professor of information science in the United Kingdom.

Mike's initial research in Sheffield involved the development of methods for the automated production of articulated subject indexes such as those used in the CAS subject indexes.[3] This work was followed by initial studies of techniques for the indexing, storage, and retrieval of chemical reaction information. These studies, which focused on the identification of the reaction centre involved in a transformation, continued for over a decade before the development of graph-matching techniques that, after extensive subsequent refinement, underlie modern reaction retrieval systems.[4,5] While this work was under way he started to develop and test systematic approaches for the selection of fragment substructures[6] for screening chemical substructure searches, and was also responsible for probably the first book anywhere on the computer handling of chemical structures.[7] In addition to their use in chemical information systems, the selection techniques developed for screening subsequently found application in several types of textual processing, for example, for compression, for sorting, and for searching online catalogues.[8]

Much of the second half of Mike's Sheffield career, from 1980 onwards, was devoted to the development of techniques for the representation, storage, and retrieval of the generic (Markush) structures that characterise many chemical patents. This work extended over some 15 years and was carried out in collaboration with CAS, Derwent, and International Documentation for Chemistry (IDC).[9,10] Other areas of interest in this period included the use of parallel computer architectures for database searching[11] and the application of natural language processing methods to the textual components of chemical patents.[12]

When Mike arrived in Sheffield in the mid-1960s, there was very little computer-related research being undertaken in library schools in the United Kingdom. His work established a tradition that has enabled the Sheffield school to play a leading role over the years in the development not just of chemical information but also of information science more generally, with several of the researchers that he supervised subsequently following academic careers in Sheffield.

During his career he received many awards, perhaps most notably the 1989 Herman Skolnik Award of the ACS Division of Chemical Information for his "pioneering research of more than two decades on the development of methods for the storage, manipulation, and retrieval of chemical structures and reactions as well as related bibliographic information". A further mark of recognition was the 2002 creation by the Chemical Structure Association Trust of an award in his name to recognise and encourage "outstanding accomplishments in education, research, and development activities that are related to the systems and methods used to store, process, and retrieve information about chemical structures, reactions, and properties".

Mike had a huge impact not just on our own careers, but upon the whole profession, yet he will be remembered not only for his outstanding professional achievements, but also for his remarkable personal qualities. He was a practising Roman Catholic and a devoted family man. He loved classical chamber music. He was kind and gently spoken. To journal editors who needed papers reviewing, Mike was known as a real 'softie': he looked for the value in everyone's work and was never excessively critical. These qualities are well illustrated by the comments of one of his doctoral students, John Barnard: "Mike was generous in the independence he gave his students, without in any way stinting in the support he offered. When he was invited to talk about the Sheffield Generic Chemical Structures project at a conference, just six months after its inception, he immediately suggested that my fellow student on the project, Steve Welford, and I should do the talking instead. That immediately projected Steve and me into the centre of the international chemical information community and provided us with contacts that directly benefited our subsequent careers. Mike involved us in all aspects of the development of the project, including discussions with potential funders, recruitment of additional staff, and

as shareholders in its commercial exploitation. Mike's easy-going and gentle Irish character made him an ideal companion on social occasions, whether in the bar at conferences, for pub lunches, or at the informal dinners at his home so often given for scientists visiting our project. I literally owe my entire career in chemical information systems to Mike, and it was a privilege and a pleasure to have been one of his students".

Michael's first wife, Mary, died of a heart condition in 1993. In 1995, he married Mary Dykstra, a professor at Dalhousie University in Halifax, Nova Scotia, Canada who had studied in Michael's department at the University of Sheffield. They had a long and happy retirement, travelling extensively and living in both Sheffield and Nova Scotia. Mary survives him. Our sympathies also go to Catherine and Kevin, his children by his first wife; his stepsons, Mark and Jeffery Dykstra; his eight grandchildren; and other members of his extended family worldwide.

References

(1) Dyson, G. M.; Lynch, M. F. Chemical-biological activities: a computer-produced express digest. *J. Chem. Doc.* **1963,** *3*, 81-85.

(2) Cossum, W. E.; Krakiwsky, M. L.; Lynch, M. F. Advances in automatic chemical substructure searching techniques. *J. Chem. Doc.* **1965,** *5* (1), 33-35.

(3) Armitage, J. E.; Lynch, M. F. Articulation in the generation of subject indexes by computer. *J. Chem. Doc.* **1967,** *7* (3), 170-178.

(4) Armitage, J. E.; Lynch, M. F. Automatic detection of structural similarities among chemical compounds. *J. Chem. Soc. C* **1967**, 521-528.

(5) Lynch, M. F.; Willett, P. The automatic detection of chemical reaction sites. *J. Chem. Inf. Comput. Sci.* **1978,** *18* (3), 154-159.

(6) Adamson, G. W.; Cowell, J.; Lynch, M. F.; McLure, A. H. W.; Town, W. G.; Yapp, A. M. Strategic considerations in the design of a screening system for substructure of chemical structure files. *J. Chem. Doc.* **1973,** *13* (3), 153-157.

(7) Lynch, M.; Harrison, J. M.; Town, W. G.; Ash, J. *Computer Handling of Chemical Structure Information*, Macdonald, London and American Elsevier, New York, 1971. http://openlibrary.org/books/OL4770313M/Computer_handling_of_chemical_structure_information.

(8) Lynch, M. F. Variety generation – a reinterpretation of Shannon's mathematical theory of communication, and its implications for information science. *J. Am. Soc. Inf. Sci.* **1977,** *28* (1), 19-25.

(9) Lynch, M. F.; Barnard, J. M.; Welford, S. M. Computer storage and retrieval of generic chemical structures in patents. 1. Introduction and general strategy. *J. Chem. Inf. Comput. Sci.* **1981,** *21* (3), 148-150.

(10) Lynch, M. F.; Holliday, J. D. The Sheffield generic structures project-a retrospective review. *J. Chem. Inf. Comput. Sci.* **1996,** *36* (5), 930-936.

(11) Brint, A. T.; Gillet, V. J.; Lynch, M. F.; Willett, P.; Manson, G. A.; Wilson, G. A. Chemical graph matching using transputer networks. *Parallel Comput.* **1988,** *8* (1), 295-300.

(12) Lawson, M.; Kemp, N.; Lynch, M. F.; Chowdhury, G. G. Automatic extraction of citations from the text of English-language patents - an example of template mining. *J. Inf. Sci.* **1996,** *22* (6), 423-436.

- -

DUST Conference 2025: Advancing Scientific Terminology & Standards for the Digital Era

Contribution from Colin L. Bird and Jeremy Frey, School of Chemistry and Chemical Engineering, University of Southampton, email: j.g.frey@soton.ac.uk, and Agnes Jasinska, Digital Curation Centre, University of Edinburgh

The Units, Symbols, and Terminology in the Physical Sciences in and for the Digital Era (DUST) Conference 2025 took place on 13–14 March 2025 at the Royal Society of Chemistry in London. Co-organised by IUPAC, CODATA, the Physical Sciences Data Infrastructure (PSDI) project, and hosted by the Royal Society of Chemistry, the conference brought together an international cohort of scientists, data experts, philosophers, and policy makers to address a common goal: ensuring scientific terminology, units, and symbols are fit for purpose in a predominantly digital and interdisciplinary research landscape.

The conference focused on harmonising scientific expression across digital platforms, enhancing machine-readability, and supporting data interoperability – key enablers for FAIR data practices and global collaboration. The programme featured expert-led sessions, interactive project tables, the IUPAC Green and Gold books, and in-depth discussion on emerging challenges and technologies.

Key highlights included:

- Dr Vanessa Seifert (Universities of Bristol & Athens) explored the implications of artificial intelligence on scientific progress, questioning how opacity in AI systems challenges traditional scientific virtues such as explanation, prediction, and truth.
- Dr Blair Hall (Measurement Standards Laboratory of New Zealand) introduced the M-layer, a novel framework enabling unambiguous expression and transformation of measurement data by linking quantitative values to defined aspects and scales.
- Dr Cerys Willoughby (University of Southampton) presented user-experience research on improving metadata practices in electronic lab notebooks, emphasising cognitive prompting techniques to enhance data quality and reusability.
- György Gyomai (OECD) examined the organisation of data using counting-units, proposing nuanced approaches for treating discrete quantities – such as population data – within measurement systems.
- Philip Dunn (LGC Group) provided a critical overview of ongoing IUPAC projects on isotopic analysis, highlighting inconsistencies in symbols, legacy definitions, and the importance of clear guidance for data comparability.
- Joseph Wright (University of East Anglia) illustrated the practical challenges of Unicode symbol representation in scientific computing, advocating for coordinated community input to encoding new concepts.
- Professor Simon Coles (University of Southampton) offered a crystallographer's perspective on FAIR data, detailing how information frameworks like CIF have supported reproducibility and transparency.
- Dr Samantha Pearman-Kanza (University of Southampton) provided a pragmatic perspective on the Semantic Web, encouraging the use of modular, standards-based ontologies for scientific data exchange.

Parallel sessions included strategic planning for the Digital Green Book, updates on the Gold Book revision, exploration of CODATA's Cross Domain Interoperability Framework (CDIF), and discussions on digital tools for PSDI to support unit representation and metadata interoperability.

For members of CICAG, DUST offered valuable insights into the future of scientific communication, data interoperability, and infrastructure – particularly how consistent terminology and machine-readable formats

underpin open science and interdisciplinary collaboration. These themes align closely with CICAG's mission to advance chemical information.



In closing, Professor Jeremy Frey (University of Southampton) emphasised the value of interdisciplinary dialogue and philosophical insight in guiding future developments. Discussions at DUST 2025 are expected to catalyse new partnerships and contribute meaningfully to the ongoing work by PSDI, CODATA/DRUM and IUPAC in supporting open, accurate, and interoperable science.

The talks can be found on YouTube.

The timestamps for the different talks are :
2:46 – Chemical progress in the age of AI – Vanessa A. Seifert
45:47 – Stuff we don't know we know: unpacking tacit assumptions about the semantics of units and dimensions – Blair Hall
1:34:40 – Human aspects of Units, Symbols, Terminology – Cerys Willoughby
2:18:00 – Counting Units, are they different from first class units? – Gyorgy Gyomai
2:53:10 – 3rd time lucky - past and current IUPAC projects on terminology for isotopic analyses – Phil Dunn
3:18:09 – Unicode for Scientific Symbols – Joseph Wright
3:34:51 – IUPAC, UNDRR/ISC Hazard Information Profiles and Disaster Risk Reduction – Virginia Murray
3:53:46 – RSC-IUPAC Committee – Andrew Shore
4:06:00 – FAIR Data Standards – A Crystallographers Perspective – Simon Coles
4:42:53 – A Pragmatic view of the Semantic Web for the Physical Sciences – Samantha Pearman-Kanza
5:21:40 – Information on the digital Green Book and how to represent units and systems – Stuart Chalk
5:55:53 – Where next for DRUM, CDIF and WorldFAIR+? – Simon Hodson
6:26:15 – Discussion round: Units and Names

The presentations can be found here. (These will shortly be moved to a more permanent archive with PSDI.)

- -

CICAG Members' Publications: Opportunity to Publicise your Research!

Contribution from Dr Helen Cooke, CICAG Distillate Editor, email: helen.cooke100@gmail.com

This is the first in a new series of articles for the *CICAG Distillate*. To start the series, we have invited CICAG Committee members to contribute references to their recent publications, see below. Going forward, we are inviting all CICAG members to send through references for their recently published articles which might be of interest to the *Distillate* readership. If you have recent publications (abstracts are not essential, and will only be included if provided) you would like us to mention in the next issue (Winter 2025-26), please send bibliographic

details to Helen Cooke, confirming that you are a member of CICAG. Please include a URL and/or information about to how to find the publications.

Publications to be featured should contain some content relating to chemical information and computer applications, but can be to journal articles, books, patents, published conference proceedings, and published reports as long as they are in the public domain.

Design and Mechanistic Analysis of a Potent Bivalent Inhibitor of Transthyretin Amyloid Fibrillogenesis
Mangione, P.P.; Verona, G.; Cantarutti, C.; Nocerino, P.; Mimmi, C.M.; Swain, C.J.; Canetti, D.; Giorgetti, S.; Uings, I.; Gillmore, J.D; Taylor, G.W; Pepys, M.B.; Bellotti, V.; Corazza, A.
Journal of Medicinal Chemistry **2025**. https://pubs.acs.org/doi/10.1021/acs.jmedchem.5c00430
Transthyretin amyloidosis (ATTR) is a systemic disease that primarily affects the heart and the peripheral nervous system. Despite available therapeutic options, advanced ATTR amyloidosis still presents unmet medical needs. We have therefore focused on the design of bivalent small molecules starting from our prototype palindromic ligand mds84, whose binding by transthyretin (TTR) greatly improves stability of the native structure by overcoming the negative cooperativity which is typical of monovalent stabilizers. Among the newly designed compounds here, we present B26, which is pseudoirreversibly bound by native TTR with faster entry kinetics into the protein molecule compared to mds84. It retains the ability to inhibit fibril formation *in vitro*, together with improved solubility. Using solution NMR, we show that B26 occupies both TTR binding sites simultaneously, leading to conformational effects distant from the binding site, including the proteolytic cleavage site involved in fibril formation by the mechano-enzymatic mechanism.

Return to Flatland
Churcher, I.; Newbold, S.; Murray, C.W.
Nature Reviews Chemistry **2025**, *9* (3), 140-141. https://doi.org/10.1038/s41570-025-00688-5
Fifteen years ago, an analysis of drug molecules moving through the stages of clinical development suggested that increased three-dimensional character was a marker of greater success. Now, we perform a similar analysis to evaluate the impact of that landmark publication and to see if the trends observed then still hold today.

Pills, Potions, and Poisons: Interdisciplinary Public Engagement Opportunities Inspired by a 17th Century Apothecary
Cooke, H.; Parmeggiani, F.
Journal of Chemical Education **2025**, *102* (5), 1848–1855. https://doi.org/10.1021/acs.jchemed.4c00925
The transcript of the inventory of the contents of a 17th century apothecary's shop was discovered during research for a museum exhibition on health and disease through the ages. This led to research into the remedies dispensed by the apothecary, which provided the impetus for development of public-engagement deliverables, including interdisciplinary talks, chemistry demonstrations, a searchable record of the contents of the inventory, and production of an accessible local-history booklet. The approach taken used history as a gateway to introduce chemistry and inspire people of all ages to engage with science. Detailed instructions are provided for four striking chemical demonstrations centered around apothecaries and exploiting some of the substances present in the inventory.

Zombie Cheminformatics: Extraction and Conversion of Wiswesser Line Notation (WLN) from Chemical Documents
Blakey, M.; Pearman-Kanza, S.; Frey, J.G. *J. Cheminform.* **2024**, *16*, 42. https://doi.org/10.1186/s13321-024-00831-2
Wiswesser Line Notation (WLN) is a old line notation for encoding chemical compounds for storage and processing by computers. Whilst the notation itself has long since been surpassed by SMILES and InChI, distribution of WLN during its active years was extensive. In the context of modernising chemical data, we

present a comprehensive WLN parser developed using the OpenBabel toolkit, capable of translating WLN strings into various formats supported by the library. Furthermore, we have devised a specialised Finite State Machine l, constructed from the rules of WLN, enabling the recognition and extraction of chemical strings out of large bodies of text. Available open-access WLN data with corresponding SMILES or InChI notation is rare, however ChEMBL, ChemSpider and PubChem all contain WLN records which were used for conversion scoring. Our investigation revealed a notable proportion of inaccuracies within the database entries, and we have taken steps to rectify these errors whenever feasible.

Blockchain Technology: Driving change in the scientific research workflow

Lawlor, B.; Chalk, S.; Frey, J.; Hayashi, K.; Kochalko, D.; Shute, R.; Sopek, M. *Pure and Applied Chemistry* **2025**, *97* (4), 279-330. https://doi.org/10.1515/pac-2023-1204

The goal of this white paper is to present an objective overview of the current use of blockchain technology along the scientific research workflow and in related areas such as chemical/drug supply chains and education. It represents the culmination of three years of data gathering, including input from multiple interviews with pioneer users of the technology, as well as from more recent adopters around the globe, and recent industry technology analysts' reports. Within these pages are descriptions of successful applications of the technology at each step of the scientific research workflow – from the timestamping of ideas to funding, to actual experimentation, to the analysis of research results, and ultimately to the sharing of information and the publication of results. However, not all blockchain use cases have such a successful conclusion. In this white paper you will learn where the technology has not worked – and why – thanks to those interviewed who discussed in detail the lessons that they themselves learned during their own blockchain journey. In addition, the paper highlights the potential future uses of the technology; the pitfalls to avoid when considering its use; when and how legislation and regulatory policies come into play; and how the technology is evolving and growing stronger (some say that the fourth generation of the blockchain evolution is on the horizon!). The paper also discusses parallel developments in quantum computing, its potential impact on blockchain technology, and what developments are in progress to ensure a stable and provably secure, quantum safe alternative to the existing blockchain approaches.

Who was Amy Davies (1864-1937)?

Leitch, D.M.

Liverpool Family Historian, **2025**, *47* (2), 88-93. (Magazine of Liverpool & South West Lancashire Family History Society, ISSN 0260-759X.) The story of a Lancastrian woman who was the ward of Sir Henry Tate, the great sugar magnate and philanthropist, related to members of the Gossage family of soap industry fame in Widnes, and wife of the man who saved Newton Heath FC and renamed it Manchester United.

For more information contact the Liverpool and S.W. Lancashire Family History Society.

Reusing Chemical Data Across Disciplines: Initiatives and Common Challenges

Mustafa, F.; Lynch, I.; Theunis, J.; Elapavalore, A.; Taha, H.M.; Frey, J.; Bach, F.; Minella, C.B.; McEwen, L. *Chemistry International*. **2025,** *47* (2), 12-16. https://doi.org/10.1515/ci-2025-0203

This work discusses reuse of chemical data across disciplines and the role of various data initiatives and projects including PARC, NORMAN-SLE, MassBank, WorldFAIR, PSDI and NFDI4Chem to facilitate increased data sharing. Improved machine-readable chemical data supports global research and interdisciplinary methodologies crucial for sustainable development and achievement of UNESCO's Open Science priorities and the UN Sustainability Development Goals. Examples of success and ongoing approaches include integrating toxicology and chemical exposure data using ontologies, linking specialised chemical data collections with larger repositories such as PubChem, and developing IUPAC International Chemicals Identifier (InChI) extensions for nanomaterials and mixtures. National data infrastructure projects in the UK and Germany focus on digitising and standardising chemical research data management workflows, aiding scientists in data

collection, storage, processing, analysis, disclosure, and reuse. These global initiatives aim to enhance chemical data interoperability to solve real-world problems, foster collaboration, and promote innovation while considering sustainable data resources beyond individual projects

- -

Mike Lynch Award 2025: Professor Val Gillet

Reproduced from the CSA Trust website



In recognition of her pioneering research in chemical structure analysis, visionary educational leadership, and enduring dedication to advancing scientific collaboration, Professor Val Gillet has been awarded the 2025 Mike Lynch Award by the Trustees of the CSA Trust.

The Trustees are proud to honour Professor Gillet for her outstanding and innovative contributions to the field of chemoinformatics. Her influential body of work – widely cited and highly regarded – has shaped the chemoinformatics landscape and driven progress in pharmaceutical research and development.

In parallel with her research achievements, Professor Gillet has played a central role in building and refining the chemoinformatics educational program at the University of Sheffield. Under her supervision, the program has earned international recognition as a hub of excellence for research and training. She is also co-author of a foundational textbook that remains a key reference in the field.

Beyond her academic accomplishments, Professor Gillet has been instrumental in fostering collaboration and visibility within chemoinformatics. Her longstanding support of the Cheminformatics Conference in Sheffield and the International Conference on Chemical Structures in Noordwijkerhout has helped elevate them to the field's most prominent international gatherings.

Val Gillet is a Professor of Chemoinformatics at the University of Sheffield, where she leads the Chemoinformatics Research Group within School of Information, Journalism and Communication.

The CSA Trust is an internationally recognised, registered charity that promotes education, research, and development in fields related to chemical structures and cheminformatics.

- -

Meeting Report: ChEMBL@15 Symposium

Contribution from Dilhara Manawadu, Research Assistant at the University of Colombo, Sri Lanka, email: dilharamanawadu@gmail.com



I am Dilhara Manawadu, and I had the wonderful opportunity to attend the ChEMBL 15-Year Symposium, held on the 1st and 2nd of October 2024 at EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridge. I am very grateful to have received a bursary from the Royal Society of Chemistry's Chemical Information and Computer Applications Group which covered my registration, accommodation, and travel expenses, making it possible for me to attend this event.

The two-day symposium was a new experience for me, as it was the first international conference I had ever attended. I enjoyed every minute of the symposium, from the workshops and poster presentations to the inspiring talks by experts from both academia and industry.

On the first day, I attended the workshop entitled *Extracting bioactivity data for drug-like compounds from ChEMBL*, which was both informative and interactive. The workshop provided an introduction to ChEMBL, covering all the data types and molecular structures, as well as how to navigate the ChEMBL database through interactive examples. The next workshop I attended was *Introduction to the ChEMBL load process*. This session introduced me to various ways we can structure the data and also what to expect from the ChEMBL team during data deposition.

Day Two was equally exciting, with a series of talks by invited speakers, flash poster presentations, a nature walk in the Hinxton Wetlands, and a celebratory cake-cutting to mark the 15-year milestone of ChEMBL. The day started with a warm welcome by Dr Barbara Zdrazil, followed by informative talks from Dr Andrew Leach and Dr Nicolas Bosc on the evolution of ChEMBL and SureChEMBL, respectively. A particularly interesting talk by Dr Greg Landrum from ETH Zürich highlighted the importance of data quality when working with large datasets. Dr Anna Gaulton from Exscientia gave a brief history of the ChEMBL database, followed by a fascinating talk on AI-driven drug discovery. Prof. Brian Marsden from the University of Oxford shared insights about the EUbOPEN project and its importance in providing an open-access chemogenomic library with chemical probes.

The poster sessions during the coffee breaks were an exciting opportunity to talk with other ChEMBL users. I made connections with individuals who shared similar interests from across the globe, which is something I am very grateful for. I distinctly remember the excitement of walking into a room full of experts in the field and forming new connections from both academia and industry.

At the time of the symposium, I was in the final year of my Bachelor's degree in Biochemistry and Molecular Biology at the University of Colombo, Sri Lanka. My final year research project focused on the development of

a novel biosensor to detect fluoride levels in drinking water, and I was thrilled to have this work accepted for a poster presentation at the symposium. Since then, I have completed my degree and am now preparing to begin my doctoral studies in biological chemistry at the University of Manchester in September 2025.

I would like to express my heartfelt thanks to the RSC CICAG group for the generous bursary, without which attending the ChEMBL 15-Year Anniversary Symposium would not have been possible. A special thanks goes to the ChEMBL committee for organising such a wonderful event, especially Dr Barbara Zdrazil and Ms Clare Kim Palferman, who guided me from the beginning of the abstract submission process right through to the end – even after the symposium. I would also like to sincerely thank Professor Neil Berry and Dr Helen Cooke for their kind support and guidance throughout the bursary process.

I believe ChEMBL is not just a database, but a thriving community that brings people together to drive meaningful progress in drug discovery. I look forward to seeing many more impactful years of ChEMBL ahead, filled with new ideas and exciting collaborations.

- -

Meeting Report: SCI-RSC Workshop on Computational Tools for Drug Discovery 2025

Contribution from Professor Neil Berry, Department of Chemistry, University of Liverpool, email: ngberry@liv.ac.uk

The SCI-RSC Workshop on Computational Tools for Drug Discovery 2025 was held in The Studio, Leeds, on 8 April 2025.

This workshop gave a unique opportunity for the >30 delegates to try out a range of software packages for themselves, with expert tuition in different aspects of pre-clinical drug discovery. Attendees were able to choose from sessions covering data processing and visualisation; ligand and structure-based design, or ADMET prediction run by the software providers.

The organising committee are very grateful for contributions from:

- Medchemica – MedChemica's MCPairs Online – drug discovery analysis and design tools based on matched molecular pair analysis
- OpenEye – Practical drug discovery in the cloud using ORION
- CDD Vault – Maximising drug discovery: data management, analysis, and visualisation with CDD Vault
- CCDC – CSD tools for drug discovery
- Schrödinger – Accelerating drug discovery through computational design
- Pharmacelera – Explore ultra-large chemical spaces with accurate 3D descriptors

All software and training materials required for the workshop were provided for attendees to install and run on their own laptops and they had use of for a limited period afterwards. The event ended with a networking event enabling attendees to talk with each other and the presenters.

Feedback from the event was very positive praising the venue, organisation, the ability to try out a diversity of advanced software packages and learn about new approaches to drug discovery.

The committee has started to plan for the next SCI-RSC Workshop on Computational Tools for Drug Discovery event. If you are interested in contributing to this event as a software provider please contact Professor Neil Berry.

- -

Modernising the InChI and Bridging the age-old Gap to Inorganic Chemistry

Contribution from Prof. Dr Sonja Herres-Pawlis, Lehrstuhl für Bioanorganische Chemie, RWTH Aachen University, Germany, email: sonja.herres-pawlis@ac.rwth-aachen.de, and Dr Gerd Blanke, StructurePendium Technologies GmbH, Germany, email: gerd.blanke@structurependium.com

Abstract

The International Chemical Identifier (InChI) has long served as a pivotal tool in chemical informatics, providing a standardised textual representation of chemical substances. Developed under the auspices of IUPAC, InChI facilitates the unambiguous identification and exchange of chemical information across databases and publications. However, as the chemical sciences evolve, so too must the tools we use to describe them. Recent efforts have focused on modernising the InChI to better accommodate inorganic compounds, ensuring the system remains FAIR (Findable, Accessible, Interoperable, and Reusable) and sustainable for future applications. As its latest development, the InChI enables the correct description of molecular inorganic compounds which is a major step in the future evolution.

Introduction

Chemistry, as a discipline, thrives on the precise and consistent communication of molecular information. Whether one is designing a new drug, analysing pollutants, or exploring novel materials, the ability to unambiguously describe a chemical substance is essential. One of the most important tools for this purpose is the International Chemical Identifier, or InChI.

Developed by the InChI Trust and safeguarded by the International Union of Pure and Applied Chemistry (IUPAC), the InChI system was launched in 2005 as a means to convert chemical structures into a unique, textual code. This code – compact, digital, and universally readable – makes it easier to search, compare, and link chemical information across software tools, scientific literature, and databases. InChIs are now used by major repositories like PubChem, Reaxys, and ChemSpider, and form the backbone of chemical data exchange across disciplines.

However, like any technological solution, the original InChI system has its limitations. It was designed primarily with organic molecules in mind – relatively small, well-defined compounds with clear bonding rules. In contrast, molecular inorganic compounds often exhibit complex bonding situations, variable oxidation states, coordination geometries, and a level of structural diversity that challenges the rules underpinning early InChI algorithms. As a result, many inorganic compounds are either poorly represented or omitted entirely in systems relying on InChI.

At the same time, the scientific landscape is shifting. Inorganic chemistry plays a critical role in fields such as energy conversion, catalysis, environmental chemistry, and materials science. These domains generate vast and diverse datasets, often involving compounds that defy simple structural conventions. To keep pace with these developments and ensure interoperability of data, the InChI system must evolve.

The legacy and limitations of InChI

Since its inception, InChI has revolutionised the way chemists share and search for molecular information. By converting complex chemical structures into unique strings, it enables efficient data retrieval and integration. However, the original design of InChI was predominantly tailored to organic molecules with normal covalent bonds. This focus led to challenges when representing molecular inorganic compounds, particularly those involving haptic bonds in organometallics. In such cases, the standard InChI often fails to capture the bonding and structure, rendering it less effective for inorganic chemistry applications.

Driving the need for modernisation

The impetus for updating the InChI codes stems from the growing recognition of its limitations in handling inorganic compounds in general and especially stereochemically active inorganic compounds. As the field of inorganic chemistry expands, the demand for accurate and consistent representation of these substances becomes paramount. Without a robust system, data interoperability suffers, hindering research and collaboration. Moreover, aligning InChI with FAIR principles ensures that chemical data remains accessible and usable across various platforms and disciplines.

Key enhancements in InChI Version 1.07

The release of InChI version 1.07 marks a significant milestone in addressing previous shortcomings. This update introduces several critical improvements:

- **Comprehensive code documentation**: Previously, the InChI algorithm operated as a 'black box', with limited insight into its inner workings. The new version offers detailed documentation, enhancing transparency and facilitating user understanding.
- **Bug and security fixes**: Over 3,000 bugs and security vulnerabilities have been addressed, bolstering the system's reliability. Additionally, nearly 60 issues identified by Google's OSS-Fuzz have been resolved, ensuring robustness against potential threats.
- **Enhanced testing framework**: The implementation of new test systems allows users to directly access code developments, promoting confidence in the system's accuracy and performance.
- **Transition to open-source development**: By moving the InChI codebase to GitHub, the project invites collaboration from the broader scientific community. This openness fosters innovation and ensures the system evolves in line with user needs.

Addressing inorganic chemistry challenges

One of the most significant advancements in the upcoming version is the improved handling of inorganic compounds. It can be tested in the webdemo as development version named '1.07.3 with molecular inorganics'. Historically, InChI has struggled with representing coordination complexes and metal-containing structures due to its treatment of metal-ligand bonds as disconnected. This approach often led to ambiguous or incomplete representations. The upcoming versions introduce new routines that better capture the connectivity and geometry of inorganic molecules, providing more meaningful and accurate identifiers.

To adapt the InChI for the molecular inorganics without interfering with the calculation of InChIs for organic molecules an additional structure normalisation step is introduced at the very first step within the InChI workflow. The metal atoms of molecular inorganics are characterised by their coordination numbers and bond orders. Double, triple or higher bonds to metal atoms are kept, bonds with lower bond orders are treated according to the following rules:

- If the coordination number of the metal atom does typically not occur in the context of salts the bonds pointing to the metal atoms are kept. The 'salt coordination numbers' are administered in an internal array.
- For the disconnection of metal bonds to its ligands the idea of Pauling is adapted to transfer single bonds into ionic bonds in case the difference of the electronegativity (ΔEN) of two atoms linked by the bond is higher than 1.7. Within the InChI program any difference between two elements is handled in a binary table where 0 (no bond break) corresponds to ΔEN <= 1.7 while 1 stands for transfer into ionic bond according to ΔEN > 1.7.

Figure 1 illustrates the handling of single bonds in molecular inorganics and Figure 2 shows examples.

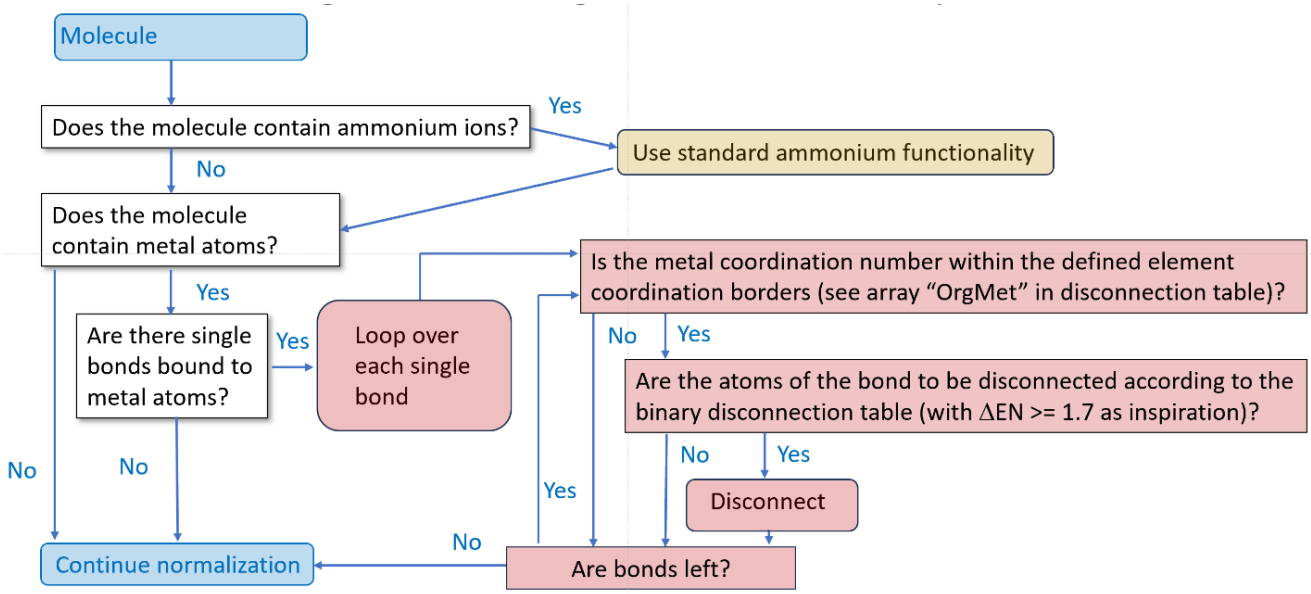


Figure 1. Workflow for the preprocessing of molecular inorganic compounds.

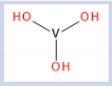
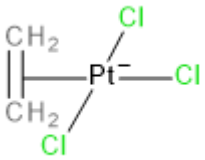
Structure	Standard InChI	Transformation	Mol. Inorganics InChI	Comments
	InChI=1S/3H2O.V/h3*1H2;/q;;;+3/p-3		InChI=1/3H2O.V/h3*1H2;/q;;;+3/p-3	Coord# = 3 in OrgMet4 for Vanadium ΔEN (O-V) >= 1.7
	InChI=1S/3FH.V/h3*1H;/q;;;+3/p-3		InChI=1/3FH.V/h3*1H;/q;;;+3/p-3	Coord# = 3 in OrgMet4 for Vanadium ΔEN (F-V) >= 1.7
	InChI=1S/3ClH.V/h3*1H;/q;;;+3/p-3		InChI=1/Cl3V/c1-4(2)3	Coord# = 3 in OrgMet4 for Vanadium ΔEN (Cl-V) < 1.7
	InChI=1S/FH.O.V/h1H;;/q;;+1/p-1		InChI=1/FH.OV/c;1-2/h1H;/q;+1/p-1	Coord# = 3 in OrgMet4 for Vanadium ΔEN (F-V) >= 1.7 but V=O
	InChI=1S/ClH.O.V/h1H;;/q;;+1/p-1		InChI=1/ClOV/c1-3-2	Coord# = 3 in OrgMet4 for Vanadium ΔEN (Cl-V) < 1.7 but V=O

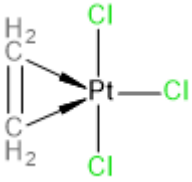
Figure 2. Examples of inorganic molecules and their representation in the standard InChI and the new InChI version.

Within molecular inorganics the same compound can be represented by different bond types. Coordinative and haptic bonds are typical bond types used in this context. As an identifier InChI must be able to transform these different representation types into one representation style. Because InChI only knows single, double, and triple bonds internally, coordinative or haptic bonds are automatically transferred into single bond systems while the valences of non-metal bonds are adapted accordingly to keep the hydrogen counts of the related atoms as needed. The following example of Zeise salt demonstrates the bond adaption process.

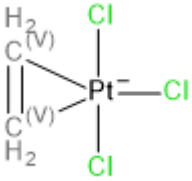
- In the Zeise salt the ethylene is located perpendicular to the PtCl3 plane:



- Using coordinative bonds, the Zeise salt can be represented in the following way:



- Internally, InChI automatically modifies any 'unknown' bond type like coordinative bonds into single bonds. To keep the hydrogen count at the C atoms with II, the valence of the C atoms must be increased by 1 to a value of V while the metal atom Pt is not related to any internally predefined valence count and does not need to be changed.



- The related molecular inorganic InChI is given by InChI=1/C2H4Cl3Pt/c3-6(4,5)1-2-6/h1-2H2/q-1 while the standard InChI of version 1.07.3 is calculated to InChI=1S/C2H4.3ClH.Pt/c1-2;;;;/h1-2H2;3*1H;/q;;;;+2 /p-3 because the standard InChI sees the structure to consist out of 5 single fragments after the bond breaks between the metal atom and its neighbours.

- In a similar but more complex approach haptic bonds are made available to the InChI algorithm.

Continuous test facility

The InChI Web demo has been extended to allow the direct test of the current developments. Under 'versions' the development branch 'InChI 1.07.3 Molecular Inorganics' can be found. This temporary test version is based on the official release 1.07.3 and extended by the most recent programming efforts to handle the enhancements described in this article.

These development versions change quite regularly. Not all of the intended functionality may already be available at the time of a test. Especially the error message handling is adapted in late stages of the development and may be misleading while working with early development versions.

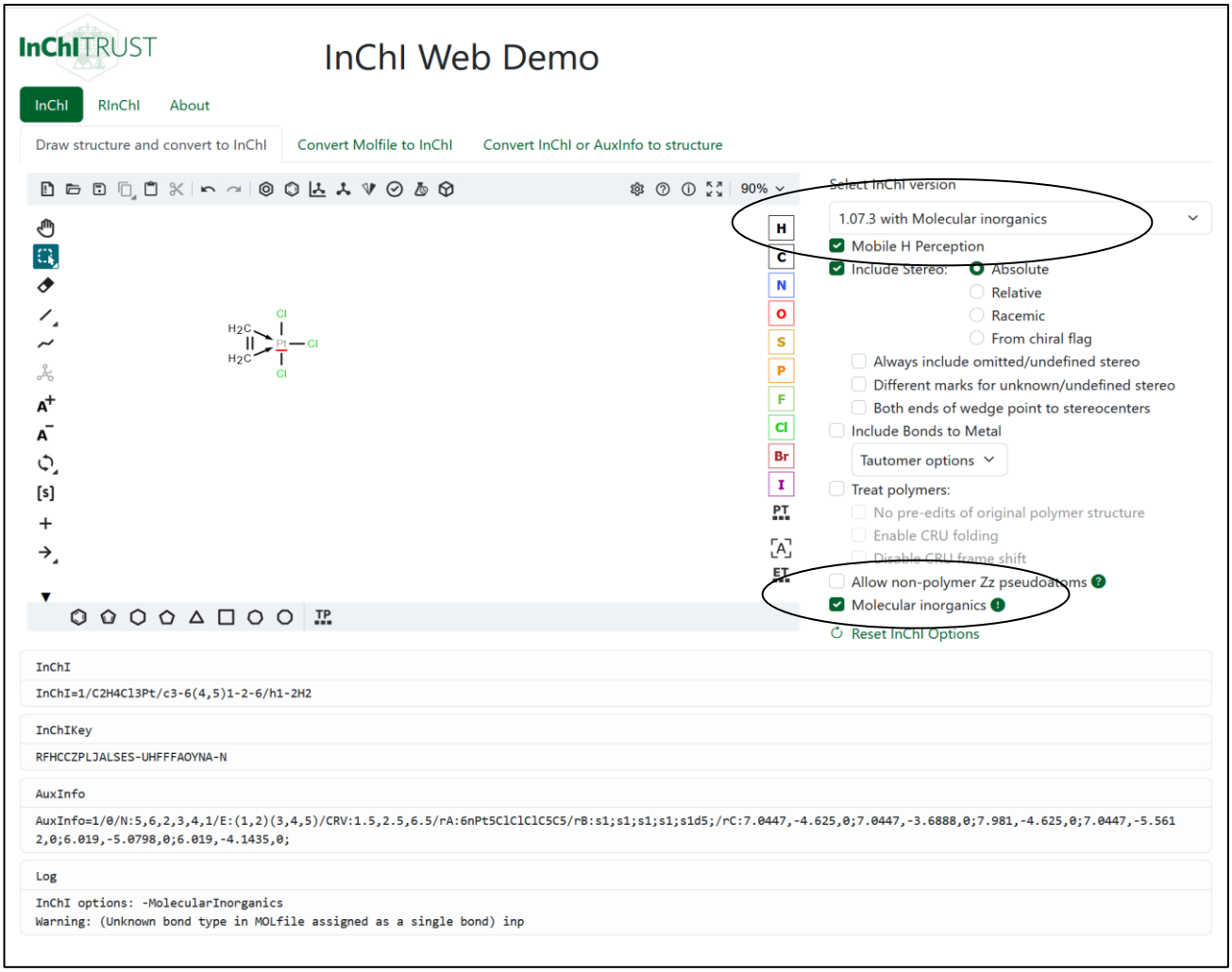


Figure 3. InChI Demo with highlighted menus and choice of InChI version.

'Hallucinating' hydrogen atoms

The first 'testathon' of the molecular inorganic developments in March 2025 in Aachen showed a general issue of working with molecular inorganics. Like most of the chemical drawing editors InChI was developed for organic molecules. Especially the valence definitions vary between the organic and inorganic world; e.g. a phosphorus atom with a coordination of four other atoms is interpreted having a valence of V (because the valence of III is exceeded and IV is not defined in the context or organic chemistry). One additional H atom is added to the P atom to reach the valence of 5 (see Figure 4).

To avoid InChI showing hallucinating Hydrogen atoms the display of implicit H atoms in the drawing editor must be set 'on' to see how the system interprets the actual depiction. Because InChI needs a correct hydrogen count for each atom, the valence counter of atoms with a false number of hydrogens must be adapted to reach the intended H atom count. In the example the upper P atom gets an explicit valence of IV so that the organic valence rules are disabled for this atom.

While the next InChI release enhances the bond handling of molecular inorganics the upcoming work will primarily focus to the stereochemistry of these substance classes beside addressing issues like the introduction of the 'MDL enhanced stereochemistry and the handling of atropisomers for organic molecules.

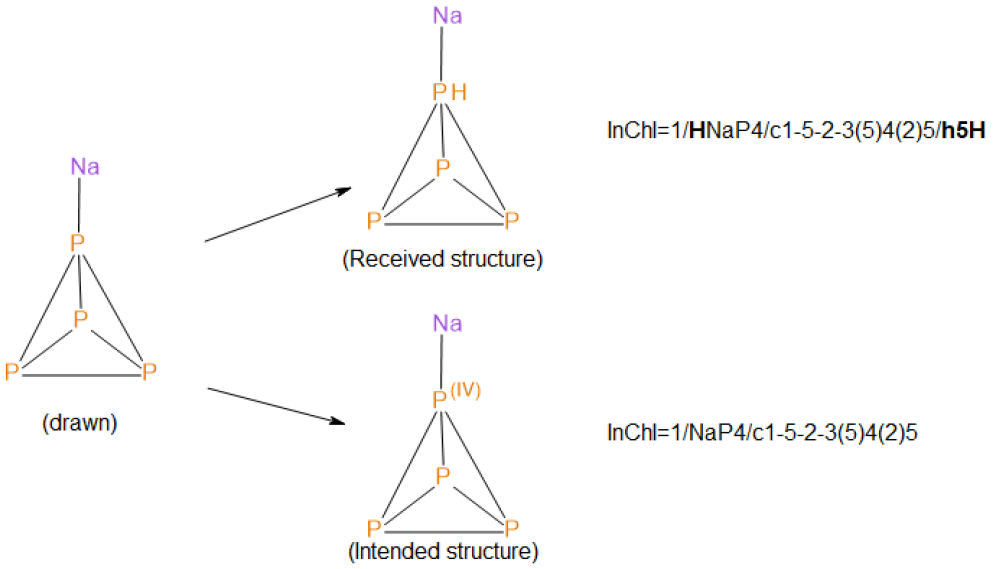


Figure 4. Hallucinated H atom in NaP_4 if drawn with standard organic valences.

Implications for the chemical community

The modernisation of InChI holds profound implications for chemists across various disciplines:

- **Enhanced data interoperability**: With improved representations, researchers can more effectively share and integrate inorganic chemical data across databases and publications.
- **Facilitated research and discovery**: Accurate identifiers enable more efficient literature searches and data mining, accelerating the pace of discovery in inorganic chemistry, potentially enhanced by AI tools.
- **Promotion of FAIR principles**: By aligning with FAIR guidelines, the updated InChI ensures that chemical data remains accessible and reusable, supporting open science initiatives as well as in-house data systems of industry.
- **Community engagement**: The open-source nature of the project encourages contributions from chemists worldwide, fostering a collaborative environment for ongoing development.

Conclusion

The evolution of the InChI system reflects the dynamic nature of the chemical sciences. By addressing previous limitations and embracing modern principles of openness and interoperability, the upcoming InChI version stands poised to serve a broader spectrum of the chemical community. As inorganic chemistry continues to grow in importance, having a reliable and accurate system for representing these compounds is essential. Through continued collaboration and innovation, the InChI project exemplifies how the scientific community can adapt and thrive in an ever-changing landscape. The next steps will focus on the stereochemistry, stay tuned!

Acknowledgement

We would like to acknowledge the InChI Trust and its members and coworkers as well as the VolkswagenStiftung, NFDI4Chem, the federal ministry of research of research, technology and space (BMFTR, DALIA project) and the Beilstein Institute for generous support.

References

Heller, S.; McNaught, A.; Stein, S.; Tchekhovskoi, D.; Pletnev, I. InChI – the worldwide chemical structure identifier standard. *J Cheminform.* **2013**, *5*, 7. https://doi.org/10.1186/1758-2946-5-7.

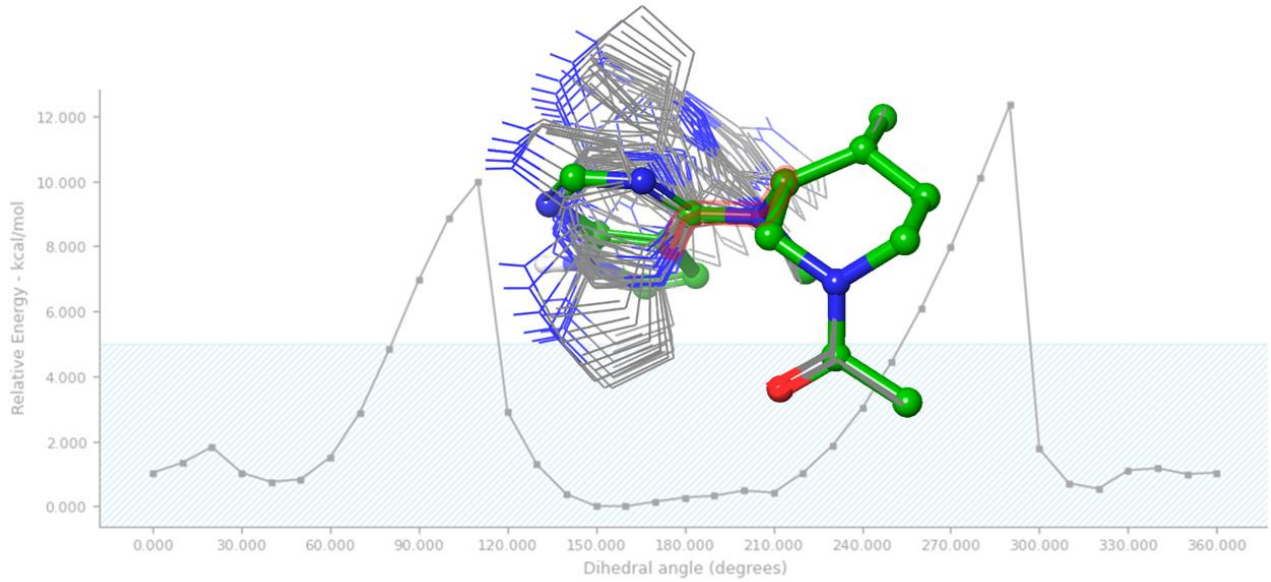
International Chemical Identifier: https://www.inchi-trust.org/

- -

Computational Chemistry Basics for Medicinal Chemists. Part 3: Conformational Analysis Worked Example

Contribution from Andreas Ritzén, Senior Principal Scientist at Novo Nordisk, email: auze@novonordisk.com

Part 1 of the CICAG Newsletter/Distillate series is in the Summer 2024 issue, Part 2 in the Winter 2024-25 issue. The full series is available on LinkedIn.



In the previous articles, we discussed compchem concepts and methods that I have found useful as a practising medicinal chemist. But how do we apply these methods to our real-world drug discovery projects? In this and the following article (coming soon), we will go through a typical workflow consisting of conformational analysis of a known ligand and docking of some analogs we design. Specifically, we will be working through the tofacitinib example discussed in part 1. I will be using the Schrödinger Maestro software, but if you are using something else, the principles will be very similar. For the record, I am not affiliated with Schrödinger and I do not receive any compensation from them.

Conformational analysis of the JAK inhibitor tofacitinib

In this example, we are going to perform a conformational analysis of tofacitinib (Figure 1) and compare the results with the known bioactive conformation from an X-ray co-crystal structure of tofacitinib in complex with the kinase JAK1. The goal of this analysis is to discover if tofacitinib has an inherent preference for the bioactive conformation, and if not, to understand why. This will enable designing analogues that are pre-disposed to adopt the bioactive conformation. We will

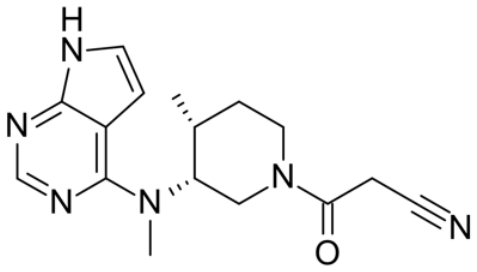


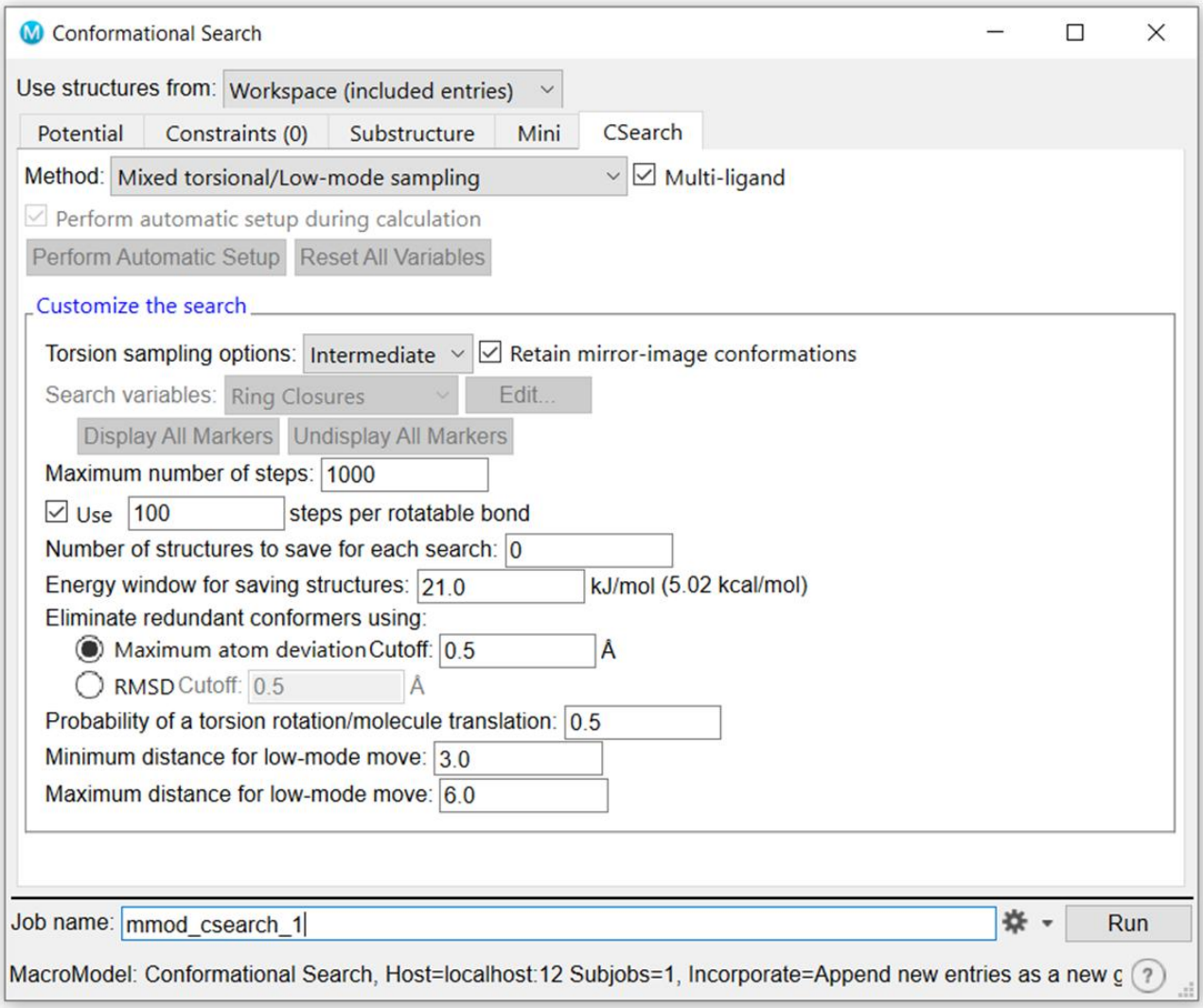
Figure 1. The structure of the JAK inhibitor tofacitinib.

 © RSC Chemical Information & Computer Applications Special Interest Group

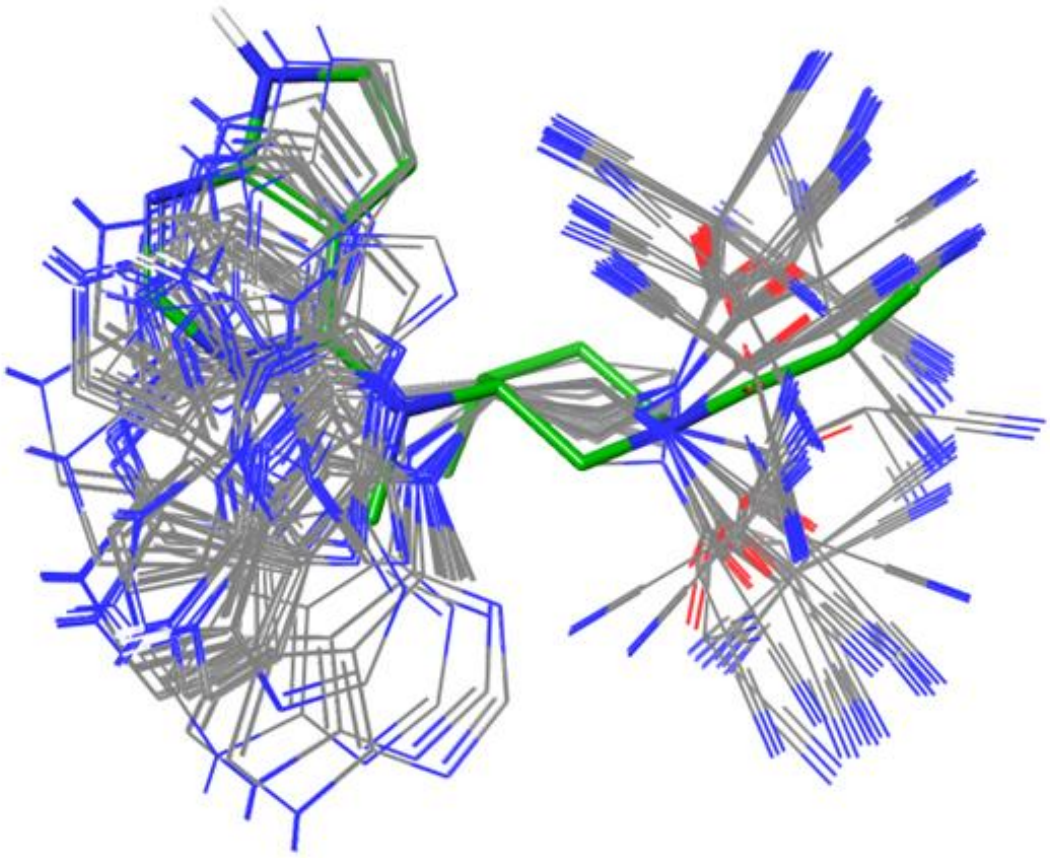
start with molecular mechanics (MM) methods (fast but less accurate) and then refine our analysis using quantum mechanics (QM) methods (much slower but more accurate). As we will see, tofacitinib is a fairly flexible molecule that can adopt many low-energy conformations. This adds some complexity, but don't be alarmed: many real-world problems are easier to work with than the one I am discussing here. I picked it to be able to highlight as many interesting details as possible. The discovery of tofacitinib (originally known as CP-690,550) was reported in Flanagan et al., J. Med. Chem. 2010, 53, 8468–8484. The stereochemistry implications and a limited conformational analysis using MM methods have been reported in Jiang et al., J. Med. Chem. 2008, 51, 8012–8018.

Molecular mechanics (MM) conformational search

The first step is to map the conformational ensemble of the molecule. To do this, we will run a conformational search using molecular mechanics (MM) methods. If you want to follow along yourself, start by drawing (building) the molecule with the correct stereochemistry and make sure it exists as a correct, all-atom, 3D model before you start. The starting conformation doesn't matter as long as the connectivity is correct. The Schrödinger MM engine is called MacroModel, and the tool we want is called Conformational Search. We will use the default settings: OPLS4 force field with implicit water solvation, saving conformers within 5 kcal/mol of the one with the lowest energy found. You can check the force field details under the tab Potential and the method and energy window settings under the tab CSearch shown below.

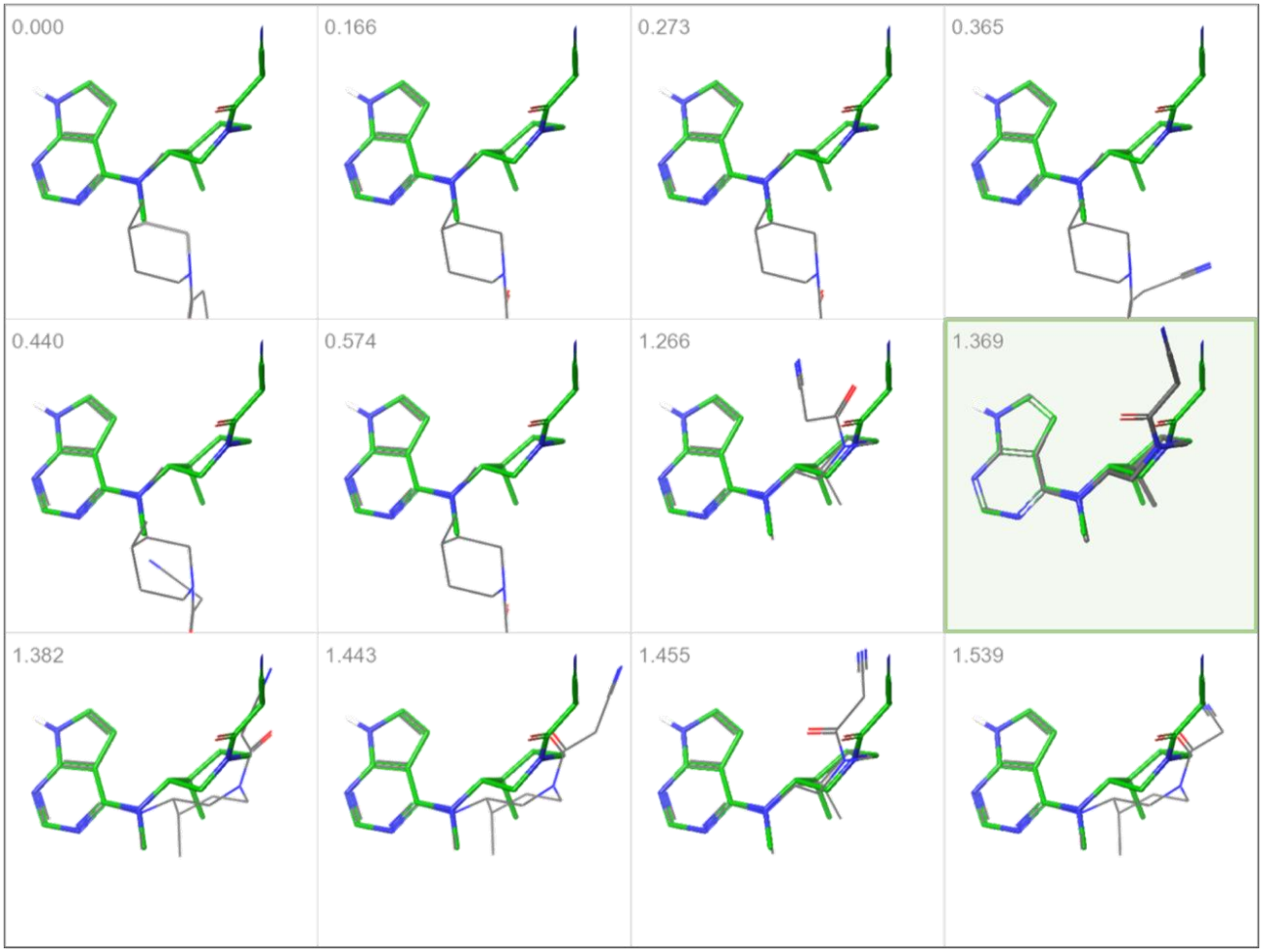


I got 120 conformations using this method, see Figure 2. In the top panel, I aligned all the conformers to the piperidine ring of the bioactive conformer from the X-ray crystal structure of tofacitinib bound to JAK1 (PBD: 3EYG) shown overlaid in green sticks. There is some conformational flexibility in the piperidine ring, but the big picture is dominated by rotations around the heterocycle-nitrogen-piperidine bonds and the cyanoacetamide moiety. The bottom panel shows the 12 lowest-energy conformations found during the search (aligned to the heterocycle, no the piperidine ring) with their respective relative energies in kcal/mol.



All conformers

Interestingly, the conformer closest to the bioactive conformer from the X-ray structure (highlighted in Figure 2) is about 1.4 kcal/mol above the global minimum. The main difference is a rotation around the heterocycle-nitrogen bond. This difference would be interesting to examine further using QM methods. But there are too many conformations with similar energy for QM calculations (assuming you run them on your own desktop or laptop computer). If we take a closer look at the cloud of conformers in Figure 2 (top panel), we can see that there are clusters of rotamers around bonds in the cyanoacetamide moiety. These contribute many conformers with similar energies and the rest of the molecule being in the same conformation. To reduce the number of conformers we need to analyse, we can simplify the molecule a bit by deleting the nitrile moiety. This should result in much fewer conformations while still capturing the heterocycle-nitrogen-piperidine part of the molecule. It can be quite useful to analyse a **model system** if your ligand of interest has many rotatable bonds. Removing atoms that do not contribute to the key conformational questions we want to answer will speed up QM calculations a lot: There are fewer atoms per conformer, and fewer conformations.



Low-energy conformers

*Figure 2. Conformational search results for tofacitinib using MM (Schrödinger MacroModel, OPLS4, implicit water). The bioactive conformation from the X-ray crystal structure (**PBD: 3EYG**) is shown in green sticks and the 120 generated conformers are shown in gray (top panel). The 12 lowest-energy conformations are shown individually overlaid with the bioactive conformation, and MM energies are shown in kcal/mol relative to the lowest-energy conformer (bottom panel, reproduced from **part 1**). The conformer most closely matching the bioactive is highlighted in the green box.*

Model system MM conformational search

Let's make a model system with an acetamide capping the pyridine ring nitrogen. Using this model system with the same conformational search tool as above, I got 49 conformers – still a lot, but more manageable than the 120 we started with. I looked through those 49, and I noticed that many of the high-energy ones (3-5 kcal/mol) had piperidine rings with axial heterocycle chair conformations, or non-chair conformations not worthy of further consideration. Knowing that MM methods are usually accurate to within ~3 kcal/mol, I focused on the 15 conformers within that energy. Because the C-N bond of the acetamide contributes two rotamers of similar energy (unless the heterocycle interferes with the carbonyl), several pairs of rotamers could be reduced to just one conformer (I picked the one with lowest energy in the pair). I selected eight conformers (tofacitinib_model_c1 –tofacitinib_model_c8) for further analysis, including the lowest-energy piperidine chair with an axial heterocycle (tofacitinib_model_c8), see Figure 3.

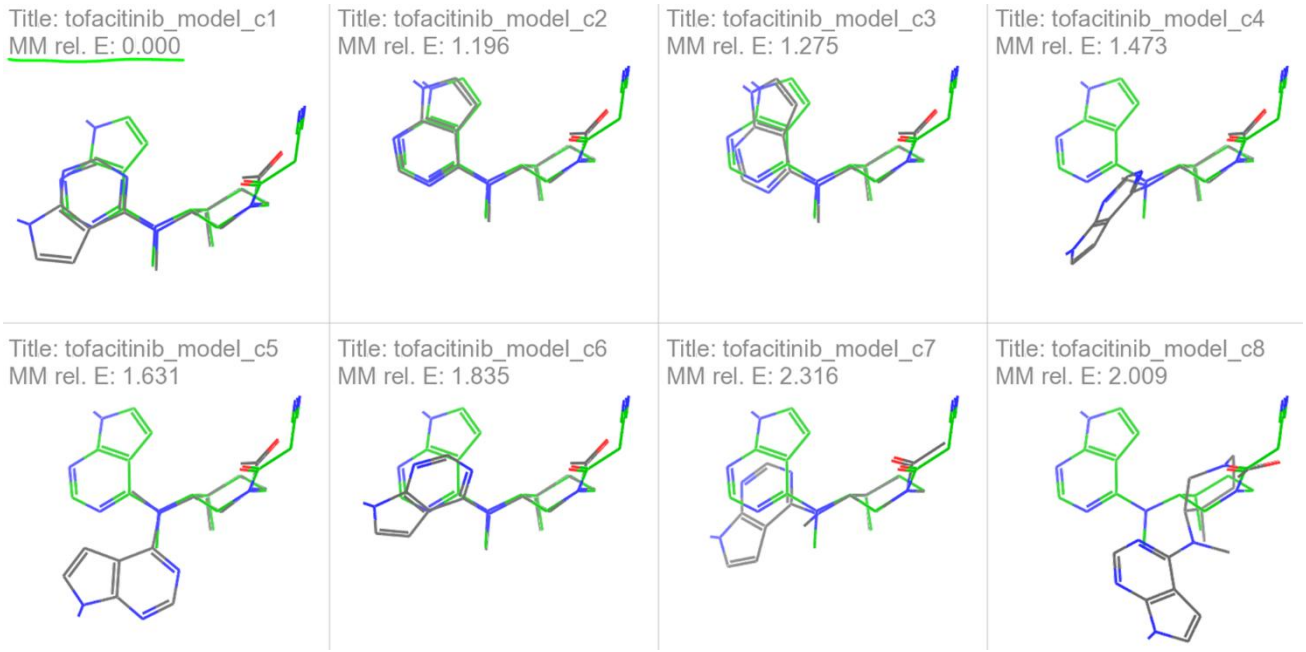
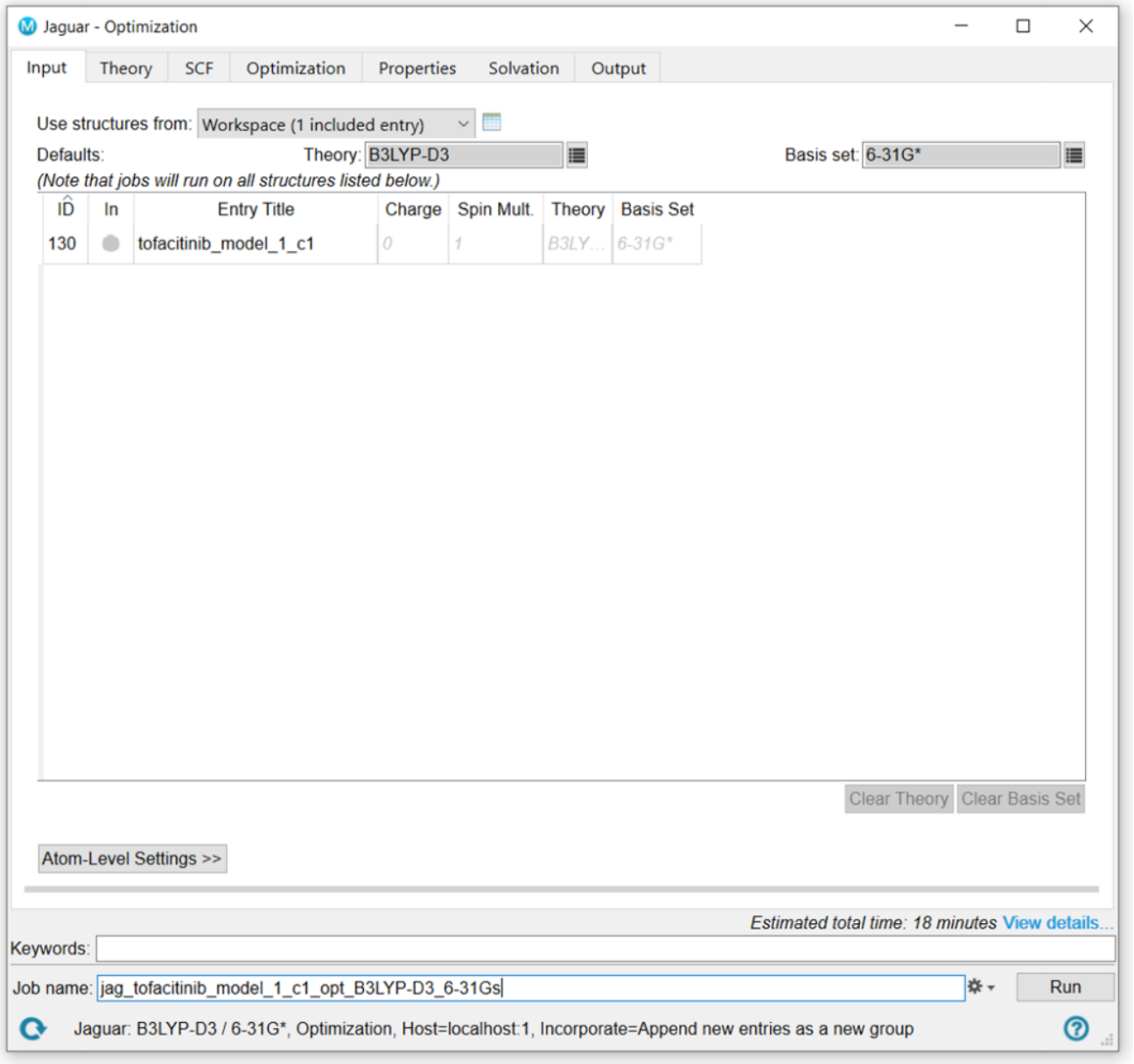


Figure 3. Selected conformers from the conformational search of the model system of tofacitinib. Each generated conformer (grey) is shown overlaid with the bioactive conformation (green) and the relative energy in kcal/mol is shown in the top left corner of each panel.

Notice that the model system also shows that the conformers closest to the bioactive one (tofacitinib_model_c2 and tofacitinib_model_c3) are 1.2-1.3 kcal/mol above the global minimum. If this is true, it means that tofacitinib will have <10% bioactive conformation in solution. In other words, the enthalpic cost of binding (ΔH) will be 1.2-1.3 kcal/mol, because MM energies are potential energies, which are roughly the same as ΔH. Because locking the molecule in a minor conformer upon binding will remove a lot of the conformational ensemble, there is also an unfavourable contribution to conformational entropy (ΔS) as discussed in part 2 of this series. If the MM energy landscape is correct, the binding affinity might be improved if the molecule could be tweaked to prefer the bioactive conformation instead of the global minimum we got from our MM conformational search.

Quantum mechanics (QM) geometry optimisation

QM methods are more accurate than MM, so the next step is to calculate QM energies. Maybe this will change the picture a bit? After all, tofacitinib is a marketed drug with high potency and affinity for its target. As we discussed in the previous article, our QM workflow consists of a geometry optimisation using a computationally less expensive, but also less accurate method, followed by a calculation of the energy of the optimised structure using a more accurate method. Starting with the geometry optimisation, we will run this using the B3LYP-D3 functional with the 6-31G* basis set. This is a relatively small basis set, so it will not give very accurate energies, but is enough to get reliable geometries. The Schrödinger Maestro tool is called Jaguar - Optimization, and on the Input tab, we can set this as 'Theory: B3LYP-D3' and 'Basis set: 6-31G*', see below. I left the rest of the settings at their default values and ran this for each of the eight conformers in Figure 3.



QM conformer energy calculation

For the eight selected conformers in Figure 3, there were no major changes in conformation from the starting geometries generated in the MM conformational search. Next, to give reasonably accurate conformational energies, I calculated the energies of each conformer with both the B3LYP-D3 and the more modern M06-2X-D3 functional using the larger 6-311G** basis set. I used the Schrödinger Maestro tool Jaguar – Single Point Energy with default settings. It is very similar in appearance to the geometry optimisation tool above, except it does not include the Optimization tab. There were no major difference in the energies between the two functionals, so I chose the M06-2X-D3 values for the following discussion. Figure 4 shows the optimised geometries and their QM energies (QM vac. rel. E) compared with the MM energies from Figure 3 (MM rel. E).

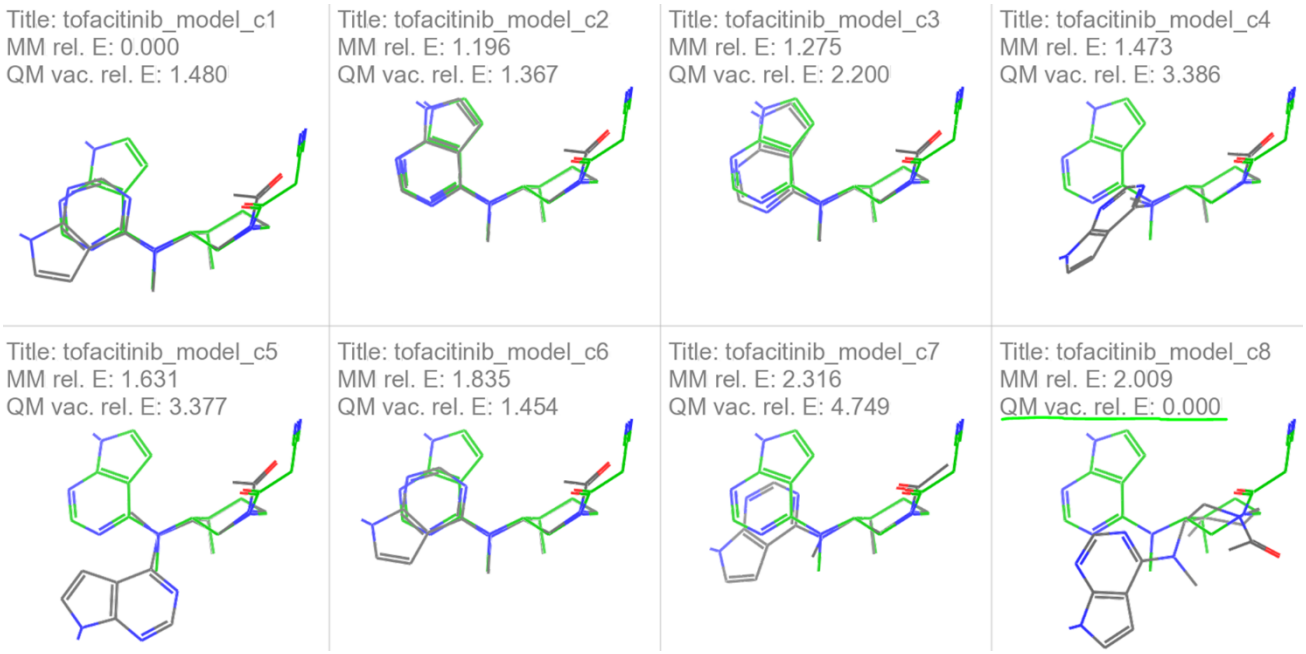


*Figure 4. MM energies from Figure 3 compared with QM (M06-2X-D3/6-311G**) in vacuum after QM geometry optimisation (B3LYP-D3/6-31G*). Each geometry optimised conformer (gray) is shown overlaid with the bioactive conformation (green).*

The results from QM are different from the MM energies, and the rank ordering has also changed quite a bit. The MM results are within 1-2 kcal/mol of the QM results, so MM is still very useful to narrow down the most relevant conformers. Note that the lowest energy conformer according to QM is tofacitinib_model_8, the only conformer with an axial heterocycle-amine and equatorial methyl group. What could be the reason for this? One possible explanation is the gauche effect, i.e., the preference for two electronegative atoms connected through two sp^3 carbons to adopt a gauche geometry (Figure 5). This is a stereoelectronic effect that is captured by the QM calculation.

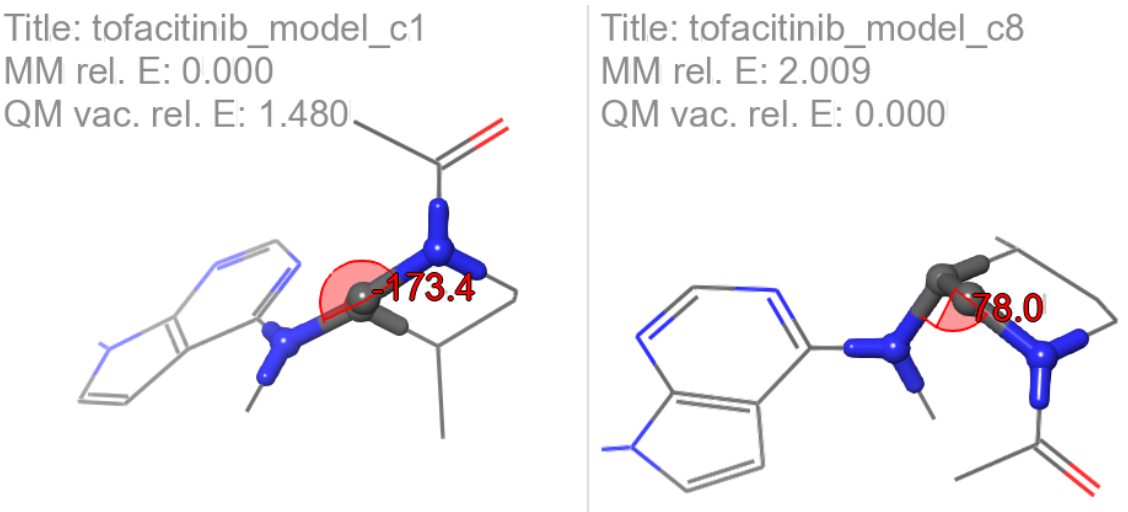


Figure 5. Equatorial (left) vs. axial (right) conformation of the heterocycle-amine moiety illustrating the gauche effect: The two nitrogen atoms are trans in the equatorial conformer and gauche in the axial conformer. The dihedral angles across the N-C-C-N bonds are shown in red in degrees.

QM conformer energy calculation with solvation

The bioactive conformers (tofacitinib_model_c2 and tofacitinib_model_c3) are 1.4-2.2 kcal/mol above the global minimum according to our QM calculation – even higher than what we found with MM! Can this really be true? We know that QM is more accurate than MM, and the level of theory I used here (M06-2X-D3 functional with the 6-311G** basis set) is a medium-level method that we can expect to give fairly accurate rank ordering and energies to within 0.5 kcal/mol or so. But the calculation of the energy was done for one molecule in isolation, i.e., in vacuum without considering solvation effects. This can have a dramatic influence on conformer energies, and there are several solvation models that can be added to the QM energy calculation to give a better estimate of solution energies. These models do not take individual solvent molecules into account, but rather use bulk solvent properties.

In Figure 6, I added the PBF (Poisson Boltzmann Finite element method) water solvation model to the M06-2X-D3/6-311G** calculation we discussed in Figure 4. The 'QM sol. rel. E' is the result of this calculation. With this calculation, tofacitinib_model_c2 is the global minimum, and this is very close to the bioactive conformation shown in green. But the energy difference between the first three conformers is negligible suggesting that they are all populated. We also see that the 'axial' conformer tofacitinib_model_8 is now 1.2 kcal above the global minimum – a pronounced difference from the previous calculation in vacuum! With solvation, the energies of the tofacitinib_model_c1 and tofacitinib_model_c8 are reversed.

To conclude, the two major conformers of our model system in aqueous solution are tofacitinib_model_c1 (almost identical to tofacitinib_model_c6) and tofacitinib_model_c2 (almost identical to tofacitnib_model_c3). They differ by a ~180 degree rotation around the heterocycle-nitrogen bond, and are of similar energies.

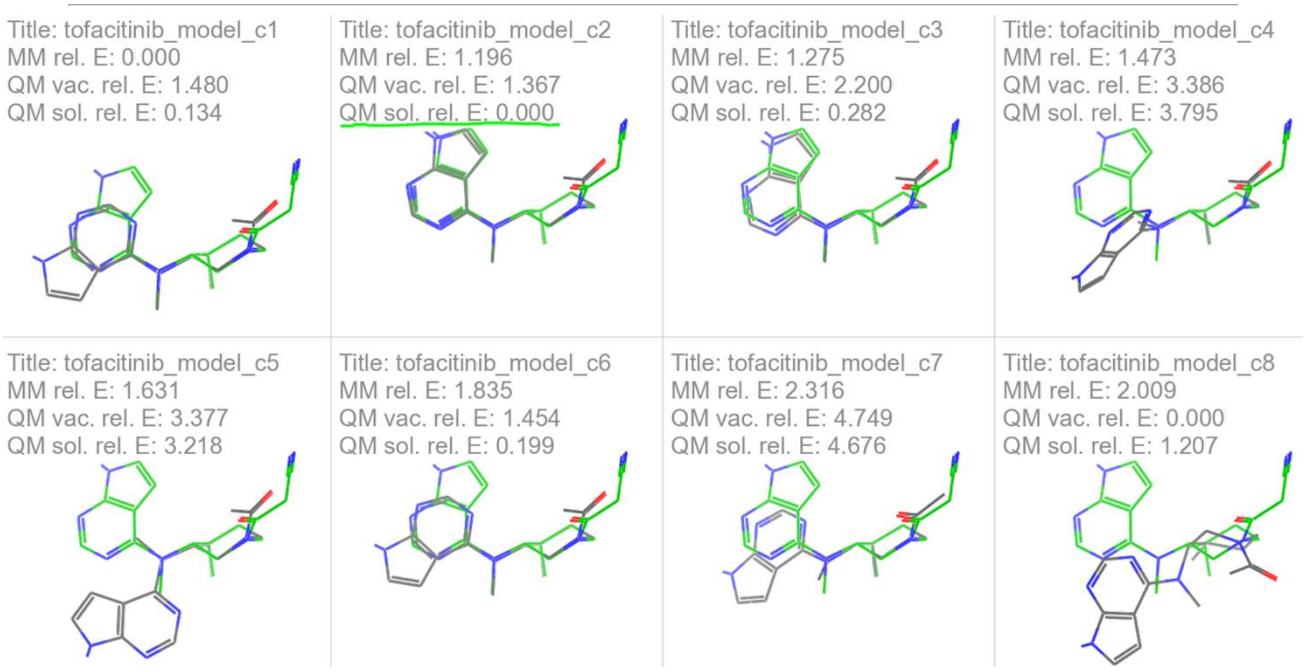


*Figure 6. Energies from Figure 4 with the addition of QM energies (M06-2X-D3/6-311G**) calculated using the PBF water solvation model. Each geometry optimised conformer (gray) is shown overlaid with the bioactive conformation (green).*

QM geometry optimisation and energies for tofacitinib

Now that we have thoroughly analysed our model system, let's go back to tofacitinib itself. The analysis above shows that we only need to consider a very small number of conformers out of the 120 we found. We will select

conformers that are similar to the ones from the model system tofacitinib_model_c1 and tofacitinib_model_c2, geometry optimise them using QM, and calculated the energies using QM with solvation. I picked four conformers for this, see Figure 7: in addition to tofacitinib_c1 and tofacitinib_c2, I added a cyanoacetamide rotamer of tofacitnib_c1 (tofacitinib_c1a) and the 'axial' conformer tofacitinib_c8. Geometry optimisation (B3LYP-D3/6-31G*) and energy calculation (M06-2X-D3/6-311G** without and with PBF water solvation) was done as described for the model system above.

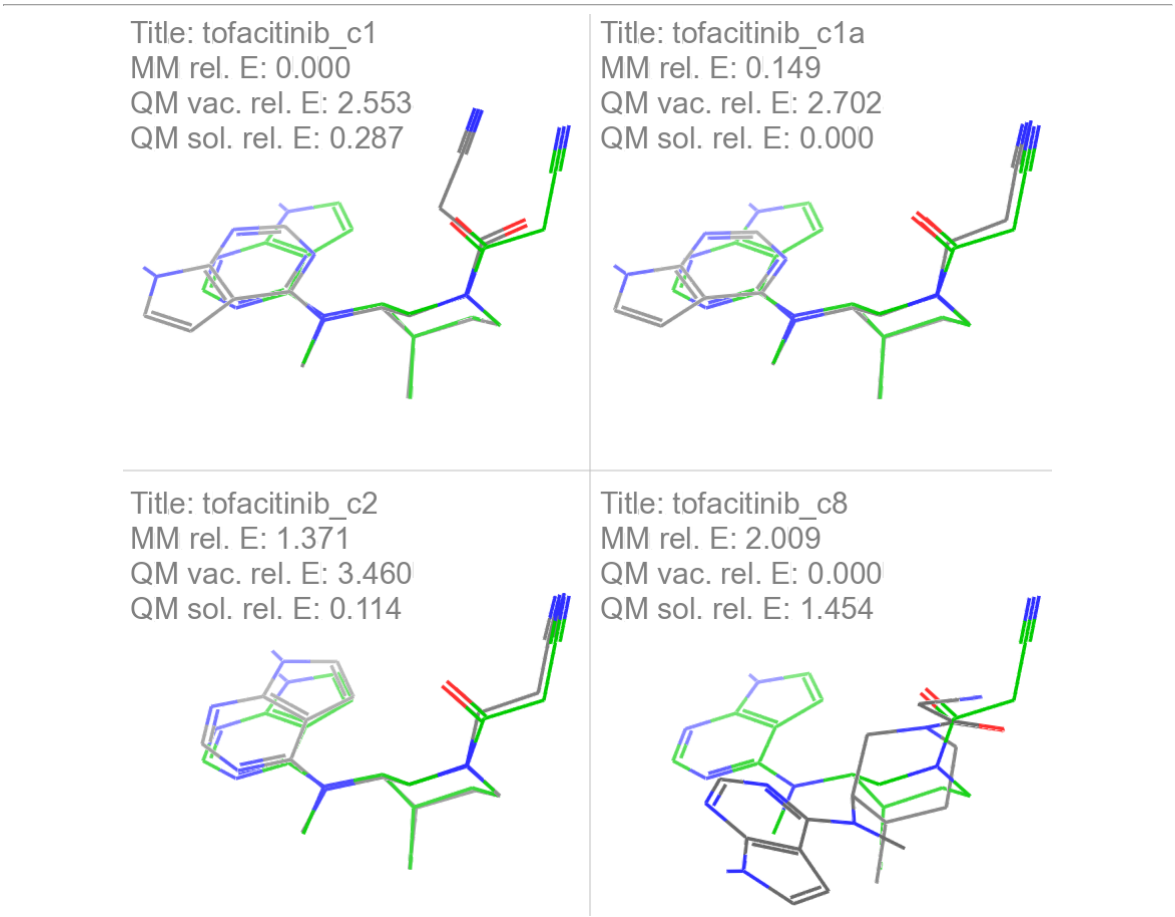


*Figure 7. Tofacitinib conformers. QM energies (M06-2X-D3/6-311G**) without solvation, and with the PBF water solvation model shown. Each geometry optimised conformer (gray) is shown overlaid with the bioactive conformation (green).*

We see the same picture as we did for the model system: The lowest energy conformers (tofacitinib_c1, its cyanoacetamide rotamer tofacitinib_c1a, and tofacitinib_c2) have the piperidine-amine-heterocycle in the bioactive conformation, with two heterocycle-amine rotamers tofacitinib_c1/tofacitinib_c1a and tofacitinib_c2 of almost the same energy. The axial conformer tofacitinib_c8 is 1.4 kcal/mol higher in energy using the solvated QM calculation and is probably not present in significant amounts.

Conclusion

The two low-energy conformers tofacitinib_c1/tofacitinib_c1a and tofacitinib_c2 are probably both present in solution, but only tofacitinib_c2 is capable of binding to JAK1. In fact, a crystal structure of an analog of tofacitinib (alone; not in complex with JAK1) was reported in Flanagan et al., *J. Med. Chem.* 2010, 53, 8468–8484 and it shows the molecule in a conformation similar to tofacitinib_c1, i.e., **not** in the bioactive conformation. Thus, binding will reduce the conformational space of the molecule, which is entropically disfavoured, and may also require a conformation of slightly higher energy than the global minimum, which is enthalpically

disfavoured. If we could restrict the conformational space so that it is biased towards the bioactive conformation, we might improve the affinity. We already discussed one such design in part 1, see Kulagowski et al., *J. Med. Chem.* 2012, 55, 5901–5921. In the next article we will take a closer look at docking some similar designs from that paper.

Literature examples

The following papers show some more advanced applications of the methods we discussed above, and how they are used in drug discovery.

Novartis ReSCoSS: *J. Comput. Aided Mol. Des.* 2021, 35, 399–415

Schrödinger E-sol to predict Kp,uu: J. *Chem. Inf. Model.* 2023, 63, 3786−3798

In this article, we went through a typical conformational analysis in small-molecule drug discovery. In the next article, we will continue with docking some compounds designed based on what we learned about tofacitinib.

- -

Our Gem in Widnes: The Catalyst Science Discovery Centre & Museum

Contribution from Dr David Nellist, Trustee at Middlewich Heritage Trust and Chair of Winsford History Society, email: dankalmia@gmail.com



Catalyst Museum entrance (Catalyst Museum photograph).

Recently I was asked why do I enjoy visiting the Catalyst Science and Discovery Centre in Widnes? For me, it's simple, it's a perfect blend of education and entertainment but then I am a chemical engineer and I need entertainment! I can't recommend it highly enough, especially if you're interested in science, technology, engineering and/or local history. It offers interactive exhibits focused on chemistry, physics, biology, chemical and process engineering, electricity and magnetism, and the history of the region's industrial heritage, with particular emphasis on the chemical industry. The Museum is next to Spike Island, a public park, located between the River Mersey and the Sankey Canal, an area once littered with chemical plants now transformed into woodlands, wetlands, footpaths and an area of industrial archaeological history. Widnes was the centre of the Northwest chemical manufacturing industry and has been heavily influenced by it for more than 150 years.

The Catalyst Museum building and surroundings

The building housing the Museum, originally known as the Tower Building, was constructed around 1860 by John Hutchinson as the head office for his alkali business. It later became the head office of the Gossage soap company, which at one time was the largest soap works in the world. Gossages was absorbed into Lever Bros in 1923. A glass lift and an enclosed glazed roof-top observation deck offer a panoramic view of the local area, from here you can see views of the River Mersey, the new Mersey Gateway Bridge, the Silver Jubilee Bridge, the area's industrial heritage and Widnes Dock. The site also boasts a children's play area.



Roof-top observation deck (Catalyst Museum photograph).

A visit to Catalyst is ideal for families, with exhibits designed to engage young visitors, making it a great choice for a day out with kids. We've always found it difficult to tear the grandchildren away! If you're into science, industrial history, or simply looking for something educational to do, it is a great visit. It's not massive, so if you're looking for a full day out, you might want to pair it with other nearby attractions such as the Lion Saltworks in Northwich, Nantwich Museum and the recently restored Murgatroyd's brine pumping station at Middlewich.



The Silver Jubilee Bridge (image source: Wikimedia Commons, photographer Geoff Wynne).

Interactive exhibits

The Museum has many hands-on interactive exhibits on the ground floor, it's a children's paradise, it makes learning about science and technology fun for all ages. On the upper floors the Museum showcases the town's industrial past and its impact on science and industry, demonstrating why Widnes has been such a key location in the chemical industry for centuries.



The Gallery (Catalyst Museum photograph).

The interactive exhibits are fascinating and include a regular 'Chemistry in Action' offering which allows visitors of all ages to engage with chemical reactions through safe, hands-on experiments, designed to showcase the role chemistry plays in everyday life – topics such as interactive demonstrations of common chemical reactions, like combustion, colour changes, or the creation of gases. Exhibits highlight how chemistry is applied in industries, including pharmaceuticals, textile dyeing, glass making, soap making, environmental science, petroleum refining and materials engineering. Regular hands-on practical experiments allow visitors to explore concepts such as acid-base reactions and polymers.

Demonstrations featuring how chemistry can help address global challenges such as energy sustainability and pollution control, including how chemistry is central to green technologies, are regularly organised.

The Museum also offers virtual reality (VR) experiences where visitors can explore scientific concepts in immersive digital environments, such as virtual tours of chemical plants or even exploring molecular structures. Potential VR experiences may include science simulations which allow visitors to simulate chemical reactions, physics experiments, or even molecular interactions, helping them better understand scientific principles in a visually engaging and memorable way. Other VR experiences may also offer a glimpse into the future of technology and industry, providing a vision of how emerging technologies might shape our world. Occasionally educational VR games related to chemistry, physics, engineering and technology may also be available, offering an entertaining way for visitors to learn while having fun.

There are regular interactive displays on industrial chemistry, allowing visitors to interact with machines and simulations to understand the industrial processes that have shaped modern life, such as the production of chemicals and materials and how they are used in many industries, with particular emphasis on local industries. Some interactive displays also explain how and why the large number of chemical and associated industries came to be centred around the Mersey estuary.

The importance of Cheshire's vast brine reserves and how brine played a key role in the establishment of soap making, glass making, paper making, pharmaceutical manufacturing and crude oil refining in the area are featured. Interactive timelines and digital displays showcase the evolution of chemical processes, including the discovery of catalysts and significant industrial chemical reactions.

Visitors can interact with models and simulations of early chemical plants, learning how the industrial revolution changed chemical manufacturing with the introduction of unit operations and a new area of study, that of chemical engineering.

It would be remiss of me, particularly whilst writing about the Catalyst Museum, not to mention the important role catalysts play in chemical reactions. Generally, this has been amply covered in the past with virtual experiments or touchscreen interfaces where visitors can explore how catalysts work to increase the speed of reactions and reduce energy consumption. There are displays showing how catalysts are used in processes such as refining petroleum, producing fertilizers, and creating sustainable energy solutions.

Chemical engineering principles are explained with hands-on exhibits that allow visitors to experiment with concepts like chemical reaction rates, mixing, filtration, or distillation. VR experiences are often available where users can simulate running a chemical plant, balancing reactions, or optimising processes. Interactive stations illustrate how modern industrial chemistry can be used to reduce environmental harm, such as carbon capture and recycling and simulations that challenge visitors to balance the benefits of industrial chemistry with the need for sustainable practices and minimal ecological impact are often available.

Exhibits also show the real-world applications of industrial chemistry, such as in the creation of plastics, pharmaceuticals, soaps and detergents, glass making, paper making, and materials for renewable energy systems. Visitors can also learn about innovations like hydrogen fuel cells or biofuels and test their knowledge.

Hands-On Lab

The Hands-On Lab is a particularly informative unit for all ages. This space often offers opportunities for visitors to perform simple science experiments, making abstract scientific concepts more tangible and accessible. For example, you could participate in experiments that demonstrate the principles of catalysis, speeding up chemical reactions, or reducing energy usage in processes.



Hands-on science (Catalyst Museum photographs).

Examples include hands-on demonstrations where visitors might observe catalysts in action, like how certain substances can alter the speed or efficiency of chemical reactions. There are regular organised activities led by experts to help participants learn about specific topics in chemistry, including the role of catalysts in sustainable practices, medicine, or energy production.

Displays and instruments help visitors understand complex scientific concepts in a fun, accessible way, such as use of 3D models or digital simulations.

Overall, the Hands-On Lab at the Catalyst Museum is an engaging and educational experience, providing participants with a deeper understanding of the science behind catalysis and its importance in modern technologies.

The Catalyst theatre

The theatre in a multi-functional venue, and is used for educational presentations to school groups and students, which can include interactive science demonstrations, workshops, and lectures. Such events allow

visitors to ask questions and interact with scientists, providing a space for teachers or guest speakers to present topics related to STEM (Science, Technology, Engineering, and Mathematics).

The theatre often hosts talks and lectures aimed at the general public, addressing topics in science, technology, and innovation. These may feature guest speakers, experts in various fields, or special themed events. It is also used for hands-on educational workshops or seminars that are part of temporary exhibitions, ongoing museum programs, celebrations and events that reach out to the local community. It is a versatile space designed to support educational outreach and public learning, fostering greater interest in science and technology. These activities aim to make complex scientific concepts engaging and fun for all ages, encouraging learning through active participation.

The Catalyst archives

If you are looking for something to do whilst the grandchildren are wrapped up in the interactive exhibits, there's the Catalyst archives! Here you can while away hours if you have the time and are fascinated with the history of the chemical industry. The Museum has a multitude of records related to the chemical industry, particularly the growth of the chemical sector in Widnes including company records, technical papers and industry reports. The archives contain many photos documenting the development of the local chemical industries, the people who worked in them, and the scientific innovations that took place in the area over time. Countless objects related to early industrial processes, scientific instruments, and machinery used in the chemical and engineering fields are held in the archive, highlighting the region's importance to the global chemical industry. The Museum itself holds recorded recollections from individuals who worked in local industries or contributed to scientific advancements in the area. These personal stories can provide unique insights into the region's industrial heritage.

If, by then, you've not had your fill, there are dozens of periodicals, scientific papers and books that help explain the development of chemistry, engineering, and other sciences in Widnes, and the broader industrial impact of the region. A more thorough account of these extensive and growing archives was given by Judith Wilde some months ago in the *CICAG Newsletter*, Summer 2022, p. 12-14.

And finally, don't leave hungry after hours of museum activity…visit the in-house café!



The café (*Catalyst Museum photograph).*

Plan your visit

As you will have seen in the foregoing, the Museum has plenty to offer; but don't expect every one of the activities noted above to be available on your visit, check before you visit, activities are detailed on the website.

I'm sure you won't be disappointed. Plan your visit and you'll have a great time.

- -

Communicating Chemistry: 'The Wonders of Colour' Exhibition and Activities

Contribution from Dr Helen Cooke, CICAG Distillate Editor and Nantwich Museum trustee and researcher, email: helen.cooke100@gmail.com



Nantwich Museum is delighted to have been awarded an Outreach grant from the RSC to develop a range of activities and events with a chemistry of colour theme. These will take place at the Museum in the summer and autumn of 2025. Throughout, links will be made with the local history of the area. The history of dyes and pigments will feature in an exhibition, a trail around the Museum's most colourful objects, a series of talks, live chemistry demonstrations, and activities for families.

Talks and live chemistry events suitable for older children and adults, presented by Professor Fabio Parmeggiani, Politecnico di Milano, will take place on 11th and 12th August, featuring exciting colourful demonstrations. The first event will be entitled "Chemistry to dye for: the fascinating history of the quest for colourful chemicals", the second "The palette on your plate: xxploring the rainbow of foods". In addition, during the school summer holidays there will be a range of family-focused workshops including tie-dyeing, simple chromatography experiments, and use of natural indicators.

The exhibition, developed by members of the Museum's Research Group (many of whom are scientists), will be displayed from 14 October – 22 November, during which period there will also be a series of talks, including "Dyeing to know: how salt brings colour to life" and "Dyeing wool and other fibres with garden and hedgerow plants".

Admission to Nantwich Museum and children's events is free; small charges will be made for talks. Check the Museum's website, Facebook, LinkedIn and Bluesky accounts for more information as our plans progress, or email Helen Cooke.

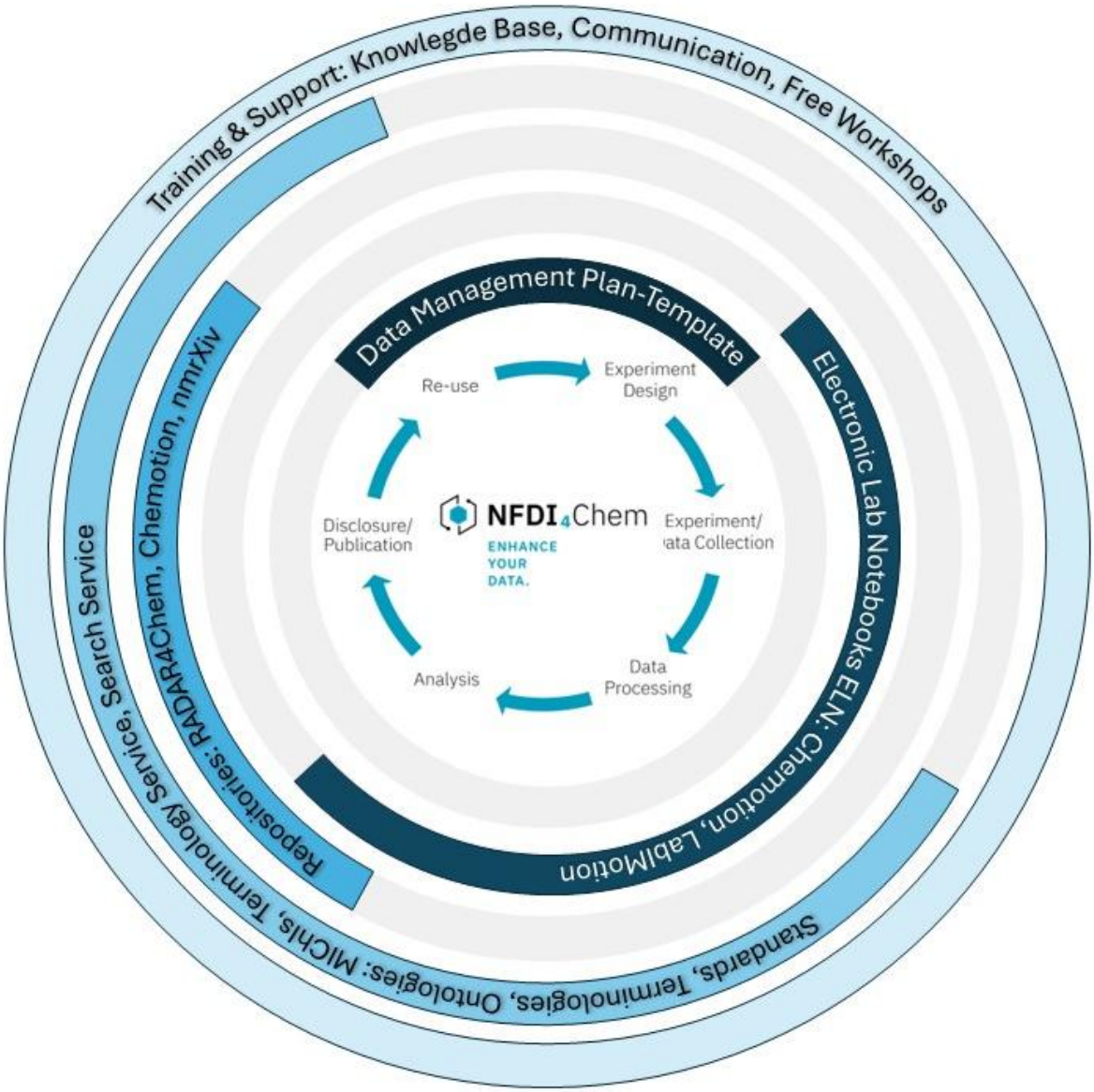
NFDI4Chem: Tools for the Whole Research Data Lifecycle in Chemistry

Contribution from Christoph Steinbeck, Felix Bach, Oliver Koepler, Theo Bende, German National Research Data Infrastructure for Chemistry, NFDI4Chem, email: contact@nfdi4chem.de

As in most areas of science, digitalisation is also becoming increasingly widespread in chemistry. As part of the National Research Data Infrastructure (NFDI), the NFDI4Chem consortium has set itself the goal of facilitating the planning, collection, storage, processing, analysis and publication of research data. Over the past five years, NFDI4Chem has established a solid foundation for research data management (RDM) in chemistry by fostering the FAIR principles (Findable, Accessible, Interoperable, and Reusable) and creating a network of integrated

services and infrastructures.[1] The vision of NFDI4Chem is to develop seamless data workflows from data creation to data reuse.

An essential building block is the training programme, which includes introductory RDM courses, Chemotion Electronic Laboratory Notebook (ELN) and repository training courses, and individual workshops for research groups. To date, these training courses have reached over 1,000 chemists. A template for data management plans supports the first step in the data lifecycle, the planning. The 'Smart Lab', which provides methods, concepts and software solutions for efficient data acquisition, documentation and publication, plays a central role in data generation. Its core component, the ELN, serves as an essential entry point for researchers and can be seamlessly integrated with other NFDI4Chem tools, lab devices and services directly in the laboratory. With the Smart Lab, chemists can load data from instruments into the ELN for analysis and evaluation.[2,3]



Across all stages of the digital data lifecycle, NFDI4Chem offers technical solutions, infrastructure and support.

Structured and annotated data can then be uploaded to the federated repositories of NFDI4Chem, which are trusted by a growing community of chemists, recommended by more than 30 chemistry journals and cover all major data types used in chemistry. This enables data to be stored, archived, and published together with rich contextual metadata, allowing it to be shared with and reused by other scientists.

Both the Smart Lab and the federated repositories are based on a foundation of ontologies, metadata and standards. Ontologies formally describe chemical entities, their properties, reactions, and relevant metadata. NFDI4Chem has curated a comprehensive collection of 39 chemistry-specific ontologies,[4] some of which were developed by the consortium to facilitate semantic annotation, linking, and detailed description of research data. The NFDI4Chem Terminology Service (TS) enables the search, access and maintenance of these ontologies for both human users and software applications such as ELNs and repositories. Ontologies are further applied in NFDI4Chem's work on the creation of Minimum Information for Chemical Investigations (MIChI), which standardises data and metadata for various chemical resources. The NFDI4Chem services continuously integrate and refine these standards, thereby gradually improving the services.

In the end MIChIs support the generation of machine-actionable chemical data to train AI models and AI applications in chemistry.

The NFDI4Chem Search Service harvests and indexes semantically annotated data from repositories, currently providing access to approximately 140,000 datasets.

In addition to the infrastructure, NFDI4Chem supports the chemical community with extensive resources. The knowledge base comprises over 70 articles and videos covering all aspects of data management in chemical research, and a dedicated help desk has already addressed approximately 600 queries from the community.

While tools and infrastructure covering the entire data lifecycle, from initial planning to reuse, are well underway, challenges around cultural change remain. Therefore, it is an ongoing effort of NFDI4Chem to facilitate research data management for a better digital future of chemical research by addressing an 'error culture', communicating existing opportunities, and further refining its tools and services.

References

(1) Neumann, S., et.al. **2024** Interim Report NFDI4Chem 2023. Research Ideas and Outcomes 10: e124977. https://doi.org/10.3897/rio.10.e124977

(2) Tremouilhac, P., et al. Chemotion ELN: an Open Source electronic lab notebook for chemists in academia. *J. Cheminform.* **2017**, *9*, 54. https://doi.org/10.1186/s13321-017-0240-0

(3) Herrmann, D., et al. Enhancing FAIRdata by providing digital workflows from data generation to the publication of data: an open source approach described for cyclic voltammetry. *Chem. Sci.*, **2025**, *16*, 4430-4441. https://doi.org/10.1039/D4SC08620A

(4) Strömert, P., et al. Ontologies4Chem: the landscape of ontologies in chemistry. *Pure and Applied Chemistry*, **2022**, *94* (6), 605-622. https://doi.org/10.1515/pac-2021-2007

- -

Professor Graham Richards

Contribution from Professor Garrett M. Morris, Associate Professor of Systems Approaches to Biomedicine, University of Oxford, email: garrett.morris@dtc.ox.ac.uk

On 11 February 2025, we lost a towering figure in the field of computational chemistry: one of the pioneers of computer-aided drug discovery, Professor W. Graham Richards, CBE, FRS, and an Emeritus Fellow of Brasenose College, Oxford.

Graham was born on the Wirral[1] with strong Welsh roots, in October 1939. He suffered with polio as a child, and had to be treated in an iron lung; it was this that in later life drove him to become an ardent athlete.

Graham won a scholarship to a top school in Birkenhead where, he said he had a "wonderful physics teacher", and later a place at Brasenose College, Oxford. While he'd applied to study physics, I think it's fair to say our field can be very grateful that he was admitted (somewhat to Graham's surprise) to read chemistry. Graham himself said, "I was too naïf to question this and to this day don't know if it was a mistake or they saw something I did not. But it was a wonderful piece of luck".[2]

Graham's journey into computational chemistry didn't take long. He began his research in 1961, working in a group of diatomic molecular spectroscopists. He was tasked with trying to explain halogen dissociation energies – but unlike most chemists of the day, he realised that digital computers could calculate the otherwise very difficult integrals numerically. It's hard to believe today, but the University of Oxford's computer back then (yes, singular) had 25K of memory. As Graham recounted, "From this, I rapidly became involved in doing quantum mechanical calculations on the computer, despite most of my senior colleagues believing that this had no future." Graham continued to work "very much at the physics end of chemistry", successfully predicting the radio frequency to detect CH in interstellar space.

He received his MA in chemistry in 1962, and his DPhil in 1964. It was probably because of his then visionary efforts using computers in chemistry that he became a Junior Research Fellow at Balliol College from 1964-1966, returning to Brasenose in 1966 as a Fellow and Tutor in Physical Chemistry: roles he held from a relatively young age (27), until 2008.

Oxford in the 1960s had mostly all-male and a small number of all-female colleges. While it seemed inevitable that all-male Oxford colleges would admit women, it required a two-thirds majority of the Governing Body to change statutes. Graham had done his postdoctoral fellowship in France and found the ambiance at the mixed Cité Universitaire "vastly superior to life in an all-male college". So in 1967, Graham succeeded in producing the required majority at Brasenose (an "unlikely pioneer" in Graham's words) – but they were blocked from enacting the change by the principals of the five women's colleges, fearing that any women who applied to Brasenose would not be considered by them. Despite this setback, they obtained a promise that if any college could go mixed, Brasenose could. In 1974, five all-male colleges achieved the required two-thirds majority, including Graham's Brasenose (and coincidentally, my future college, Jesus).

Perhaps, the turning point in Graham's journey to the life sciences can be traced back to a letter and paper he received from Jim Black (later Sir James Black and Nobel Laureate). Earlier, while at ICI Pharmaceuticals, Black had discovered β-blockers, but now Black was at Smith Kline and French and trying to find blockers of histamine in the gut. Graham thought the paper was, quote, "crap" – but Black invited Graham to discuss the problem, together ultimately showing that having a methyl at the 4-position of the imidazole ring blocked binding to the H1 receptor.

As Graham once said, he was now "hooked" and decided to focus on pharmacological applications of computational chemistry. He published an influential book, *Quantum Pharmacology*, that attracted the attention of pharmaceutical companies, prompting them to send their own researchers to Graham's group to learn more about his research and obtain copies of his software.

Graham was also very important in my own academic life. He stood out as one of my favourite lecturers as an undergraduate, demonstrating his brilliance in understanding the fundamentals of physical chemistry and his star-like quality as an engaging communicator.

In 1987, as I approached my fourth year of my undergraduate Chemistry degree – at Oxford it is known as "Part II" and consists of a year-long research project, at the end of which you write a dissertation and have a *viva* – I just couldn't wait to work with computers. By chance, I'd heard through a friend about the (then) amazing molecular graphics workstations Graham had in his lab. I applied and was fortunate enough to be accepted by Graham to do my Part II with him from 1987-88, and it was during this that I wrote what was possibly the first graphical protein informatics software, on a stunning, $50,000 Silicon Graphics workstation. I published a paper on it in the *Journal of Molecular Graphics* — a journal Graham had helped to found. One of its unusual features was that they would publish full colour figures with no additional cost to the authors. I loved my time in Graham's lab so much that I decided to stay on and do a DPhil with him, from 1988-1991.

It was during this time that very sadly, Graham's wife, Jessamine, passed away. Graham was clearly devastated by this loss, and it was perhaps because of this that he decided to throw himself into launching a spin-out, with a couple of other Oxford professors and one of his former students, Tony Marchington. Thus, Oxford Molecular was born. This was in an era when 'profit' was a dirty word. I remember giving demos of my software to visitors, who looking back, were probably potential investors. Oxford Molecular launched in 1989 with Graham as its Founding Scientist and Director, and five pieces of cleaned up, well-tested and well-documented academic software, including mine which was marketed as 'Cameleon'. It was thanks to Oxford Molecular and their inviting me to demonstrate it in 1990 at one of the American Chemical Society's national meetings in Washington, DC, and also the Protein Society meeting in San Diego, USA, that I was able to meet and demonstrate my software to my next academic mentor, Professor Arthur J. Olson, at what was then known as the Scripps Clinic and Research Foundation (now Scripps Research, in La Jolla, California). Graham also insisted that the graduate students be properly recognised, making sure they received any royalties from the sale of their commercialised research software.

Oxford Molecular plc went from strength to strength, being floated on the London Stock Exchange, ultimately being sold in 1999, and undergoing seven takeovers in the USA. Graham also became the Director of Isis Innovation Ltd in 1994 and held this role until 2014. Graham said, "Getting involved with industry and later setting up my own company has been the most rewarding part of my research career."

Graham's impact on the University of Oxford and its Department of Chemistry, and the wider higher education sector in the UK are undeniable. Today, it is almost fashionable for an academic to have their own start-up. But Graham had a significant hand in changing our attitudes to commercialising academic research: he negotiated the deal that helped raise the £64 million that funded Oxford's impressive Chemistry Research Laboratory on South Parks Road. Graham made a novel deal with David Norwood who gave £20 million in return for half the university's equity in any spin-outs from the Department of Chemistry for 15 years. This idea was very successful, and also spawned a new kind of industry; that one deal yielded more than £100 million for the University of Oxford alone, and created major companies such as IP Group, which Graham was also at one time Chairman. I recall Graham glowing with pride as he gave me a guided tour of the CRL one time I was over from California.

Graham was awarded the title of Professor of Chemistry in 1996, and became the first Chairman of the Department of Chemistry – arguably the largest in the English-speaking world – from 1997 until 2006.

Graham was made an Honorary Fellow of Balliol College in 2005, an honour he held until his death in 2025. He was also granted the title of Emeritus Fellow, Brasenose College in 2008. Graham's significant contributions to chemistry were finally recognised in 2018 when he was elected a Fellow of the Royal Society.

Graham has launched the careers of so many in the fields of computer-aided drug discovery, structural bioinformatics, protein informatics and cheminformatics. What is striking reading the messages of condolence on Brasenose's website, is how *kind* Graham was, how he looked out for his former students long after they'd left his lab. He was an inspiring scientist and mentor who has left Oxford a better place. He will be deeply missed.

References

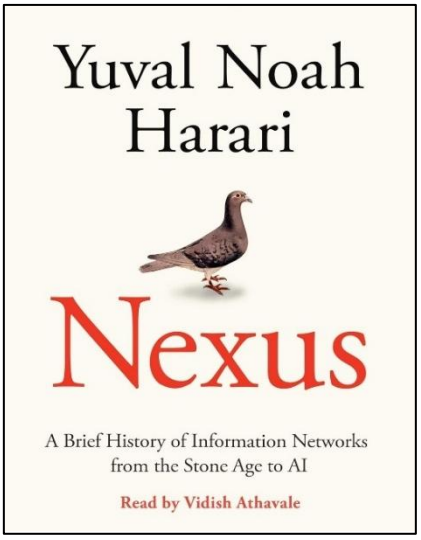
(1) Richards, G. *A Scientific Life*. **2021**. Bloomington, IN, USA; AuthorHouse UK.
(2) Greed, S. Secrets from a spin-out success story. *Nat. Rev. Chem.* **2024**, *8* (10), 719-720.

- -

Book Reviews

Contribution from Robert E. (Bob) Buntrock, Buntrock Associates, Orono, ME, USA, email: buntrock16@roadrunner.com

Nexus: A Brief History of Information Networks from the Stone Age to AI
Yuval Noah Harari, Random House, 2024, 518 pp + xxxii, ISBN 9780593734223



At first glance, *Nexus* seems aimed at a relativity small audience of librarians and information specialists. However, it's important this book should be read by a much broader audience concerned with history, technical information, networks, public information, journalism, politics, psychology, and even philosophy and theology. The author (history, Hebrew Univ. of Jerusalem) addresses the way we accept and assimilate the mountains of information (and disinformation) accessible to us, what or whom we trust, and our capacity for critical thought.

Nexus is actually two books in one, as described in the 20-page Prologue, which serves as an abstract for the book. The first three chapters describe information, documents, and networks. The succeeding eight chapters discuss the use of these resources, infallibility, democracies, totalitarianism, and finally the 'Silicon Curtain'.

Humans have acquired power through discoveries, inventions, and conquest, but power does not lead to wisdom. Information is assimilated both as individuals and as groups, which lead to networks. The author cites recent history when at least two totalitarian leaders failed in their attempts to gain power. We may be witnessing increased attempts by populists to further their power via information.

The naïve view of information is that it leads to truth, which leads to both wisdom and power. Valid truth comes from valid information. The exploding tsunami of information, with increasing availability of access, also allows the proliferation of invalid data and information. This perversion of information leads to ignorance. As Orwell stated, ignorance is strength, and it is the source of power that politicians and autocrats use to justify their actions and philosophies.

Power becomes the only reality of populist powers, and information—*their* information—becomes a weapon. Scientific research has become labelled as not seeking objective truth, but using power to determine what is truth in the service of elites. This view becomes the epitome of projection since this is the philosophy of populist power, determining truth in the service of elites.

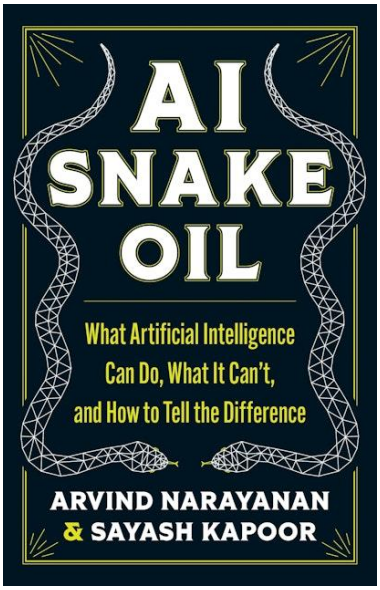
The subsequent chapters elaborate the concepts outlined in the Prologue. Chapter titles include: What is information; Stories: Unlimited Connections; Documents; Errors: the fantasy of infallibility; Decision: a brief history of democracy and totalitarianism; How computers are different than printing presses; The network is always on; The network is often wrong; Democracies; Totalitarianism, all power to the algorithms;, and The Silicon Curtain: global empire or global split?

In the epilogue, the author discusses his views on the history, evolution, and uses of AI and networks and his hope for control, good decisions, and wise uses for these programs. Chemistry is not mentioned *per se* but aspects of science in general are discussed in the chapters on Documents and Errors.

Quality publication, timely considering the current state of politics and information, domestic and global. Recommended to all those interested in STEM, history, information, computing, AI, politics, journalism, psychology, philosophy, theology, and general audiences. Includes a notes/bibliography section and index.

A longer, more detailed review will appear in the publications of ALA/CHOICE, the Magazine and the LibTech Insights Newsletter

AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference
Arvind Narayanan & Sayash Kapoor. Princeton University Press, 2024. 360p + x, ISBN 9780691249131



An important and timely book by two experts in the field (computer science Professors, Princeton). The book is described as "a guide to identifying AI snake oil and AI hype". The Introduction sets the stage and provides much of the essential information.

AI or Artificial Intelligence has been present for some time with several and often subtle uses but has gained public notoriety and use with the release of Chat-GPT from Open AI in November 2023 and use, publicity, and writing on it has erupted. Generative AI creates written documents based on a 'library' of the written word in a typically human style. Some of the first usage results exhibited growing pains, often quite humorous. The program had more than a million users in the first two months after it was issued. Open AI was surprised by the sudden rush in usage and initially didn't have the necessary computing capacity to support it. Computer programmers loved it since it could write good code based on only the requirements asked for. Microsoft licensed the technology and incorporated it into its Bing search

engine, a ‘chatbot’, that could answer questions based on search results. Google followed suit with its own search chatbot first named Bard, then Gemini.

Misuse became widespread. Chatbots often have difficulty with factual information since they’re built from statistical patterns of data from the internet from which remixed text is then written. These programs have costs and risks with their use which are not always apparent. Their use in education is very controversial often leading to them being banned them or with difficulties in identifying their use in student writing.

Another species of AI is Predictive AI which has been used for some time. Unlike generative AI, it often doesn’t work at all. Developers of applications use this technology to predict outcomes for people including predicting recidivism/future crimes of defendants or predicting if applicants for jobs would do well on the job. Predictions of needed stays in hospitals for Medicare patients are particularity erroneous, often with tragic results.

It’s often difficult to determine whether or not AI is being used. Three questions may help. (1) “Does the task to be performed require creative effort or training for a human to perform?” If yes, it may be AI for a computer to do so. (2) Is the system behaviour due to programmer code or does it come from ‘learning’ through database analysis? (3) Are the resulting decisions autonomous and environmentally adaptable? If the answers are yes, the program may be considered AI but labelling is complicated by other factors.

Other ‘sins’ of AI usage are described. Facial recognition is not infallible but even if it works well it can be harmful and use can be abused by governmental agencies. Textbook errors from the use of AI are very common as are fallacious research results, especially in biomedical studies. The authors have found it difficult to find publishers for their studies on incorrect predictive AI. Although science claims to be self-correcting, it’s extremely difficult to publish corrections, a common occurrence in other fields as well.

The 35 page Introduction (Chapter 1) is a good abstract and outline of the rest of the book and may be sufficient for many readers. The goal of the book is to determine AI snake oil (many dubious medical remedies were promoted by traveling salesmen a century ago, often literally containing ‘snake oil’). Cases are cited where even if it works well, it can be abused, including if its intent is to replace human expertise rather than enhance it. Figure 1.2 is a graph of AI snake oil, hype and harm, where it works, hype and snake oil are plotted on the vertical axis and benign to harmful are plotted on the horizontal axis. Nine uses are charted and discussed further in succeeding chapters.

Chapter 2 discusses automated decision-making increasing the use of Predictive AI. Chapter 3 discusses why predicting the future is so difficult. Generative AI is even more complicated than predictive AI and is analysed in Chapter 4, with both beneficial and harmful applications. Often false claims are made for the accurate labelling of essays as AI generated. However, the potential valid use of generative AI is also discussed.

The existential risk of the use of AI is discussed in Chapter 5. Will the advanced growth of AI become uncontrollable? The authors think not due to the measured evolution of AI (‘ladder of generality’) but acknowledge the seriousness of the issues.

Social media are discussed in Chapter 6. The controversial uses of content moderation, free speech vs. deletion of harmful content is analysed. Regulation of free speech already exists without AI and accountability is often lacking.

Chapter 7 discusses the pervasive myths about AI and the lesser ability of academia and the press to counteract misleading and fallacious claims made by the producers.

Chapter 8 is titled "Where do we go from here?". Three directions for change are: (1) set ground rules for the construction and advertising of products, (2) paths to integrate AI into society, (3) concentrate on the demand for AI rather than the supply. Use in education is especially critical. The authors are confident that uses of AI can be beneficial but close scrutiny is needed for possible and probable misuse.

Recommended to a wide variety of audiences including the general pubic. Consulting the related website and substack, AISnakeOil.com, is recommended especially for use of AI in courses.

A longer, more detailed review will appear in the publications of ALA/CHOICE, the Magazine and the LibTech Insights Newsletter.

Chat GPT – Friend or Foe?

Contribution from Dr Samantha Pearman-Kanza, School of Chemistry & Chemical Engineering, University of Southampton, email: S.Pearman-Kanza@soton.ac.uk

It seems crazy to think that ChatGPT was only actually launched in November 2022. That is less than three years ago, and yet the changes it has brought about, and the explosion of Generative AI / Large Language Model (LLMs) tools has taken our world by storm.



Researchers are using it to help answer questions, identify relevant research papers, provide suggestions on what to write, fix coding bugs, and of course, as is expected with the internet, to produce fun pictures of cats!

This all sounds great, right? But is it? Is there a more insidious undertone to these new tools that needs to be considered?

Coming from my lofty position on the Faculty ethics committee, I have spent a lot of time considering the ethical aspects of different technology, and more specifically the ethical issues of AI.

Any AI system needs to be considered with respect to classical 'Ethics of AI' issues such as:

- **Data Bias:** Any algorithm is only as good as the data it is fed, if we feed it biased data (e.g. facial recognition systems that are only trained on one nationality), then through no fault of its own, that algorithm will produce biased results.
- **Transparency/Explainable AI:** Understanding how decisions are made and what information these are being derived from.
- **Responsible AI:** Considering whether these systems are actually going to make a positive difference, and/or if there is the potential for subversion with ill intent.

- **Sustainable AI:** How much power do these solutions require vs the benefits they are going to produce.
- **Human Intervention:** Ensuring that there is human validation for vital components, e.g. medical diagnosis.

However, these concerns are exacerbated with these new AI models, none more so than the last point of human intervention. There is a real concern that there are many situations where humans may get written out of the loop entirely. Consider this quite likely scenario:

"A busy researcher uses ChatGPT, or Co-Pilot, or whatever their favourite flavour of LLM is, to produce a research paper, and on the surface, it looks and sounds relatively viable, they don't check it too rigorously, and then submit it…

A busy reviewer, also uses ChatGPT to provide a review, and again, this review sounds fairly sensible, they tweak a few words and submit it.

The paper gets through…it gets cited by various other researchers…until one day someone realises that a lot of the information presented in it is scientifically invalid, but by that point it has been cited hundreds of times, and this misinformation has been perpetuated across dozens of studies"

Now perhaps you are reading this thinking, surely that won't happen, but it already seems to be. There are projects dedicated to highlighting potential misuse of AI technology in academic literature, and at the time of writing this article, they have identified a potential 757 papers in this sphere. If you search Google Scholar for "GPT generated papers" you will already find dozens of results of researchers writing on this topic, ranging from considering the capabilities of these tools to determining how to identify their usage.

This in itself, is a great ethical concern, as the reliance on these tools, which leads to a reduction (or indeed complete removal) of critical human intervention, could cause widespread mis-information, which could have all manner of consequences.

However, this isn't the only ethical concern we need to think about. I mentioned data bias as a key issue; we currently have plenty of biased datasets, such as medical studies that have only been conducted on men, or specific ethnic groups. If we continue to feed these algorithms with biased data, then they will themselves become inherently more biased.

We also need to consider that there are many ethical questions and issues pertaining to technology development, that we ourselves don't even have answers to. Take the classic self-driving car example, in the event of a potential crash, who should be prioritised? If we can't answer that, how can technology? Fans of the popular 2004 film I, Robot, will remember Will Smiths anger at the chilling decision of a robot to save him over a young child, because statistically he had the better chance of survival. If we allow technology to take the reins on this, those are the potential decisions that we may have to face. This poses an even more sinister minefield when you consider that it is now very common to use LLMs to write code, debug code, and test code.

Another key issue to think about is sustainability and environmental impact. Research has shown that the use of LLMs produces a substantial carbon footprint. There are of course counter arguments that purport that the intelligent use of these tools to replace certain human tasks could ultimately lead to more sustainable practices (which sadly probably doesn't include using these tools to draw cute cats or answer simple questions). Although, this then in turn feeds into the widespread concern that various human-centric tasks will become obsolete.

I am also deeply concerned that we are becoming far too reliant on these tools. We need to start seeing them as just that, tools, not some magic all singing all dancing solution that doesn't require human intervention. If you asked your friend or colleague to write something for you, would you just put your name on it and pass it on without checking it? I'd hope not, so why on earth are we blindly trusting LLMs to essentially take on this role. Furthermore, the strong reliance on them to answer and question or solve any problem is likely to rob us entirely of our critical thinking skills.

I'm not saying we should stop using them and ultimately, we can't stop others using them, but given that case, we should be educating our young minds to treat these tools with appropriate suspicion. After all, when you are taught history at school, you are taught to critique your sources: When was this written – and how could this affect its accuracy? Who wrote it - do they have any likelihood for bias? Does this match up with the facts as I am aware of them? Does it match up with facts in certified texts? We should be asking all of these questions and more. So I urge all of you to treat these tools much like you would a brainy but somewhat haphazard intern, they may do a great job with some things, but ultimately shouldn't ever be trusted to be left to their own devices.

- -

Hot Topics: Robotics and Automation Meeting Report

Contribution from Nessa Carson, Early Chemical Development, Pharmaceutical Sciences, R&D, AstraZeneca, Macclesfield, UK, email: nessa.carson@astrazeneca.com

In November 2024, I had the pleasure of both helping to organise and speaking at the Hot Topics: Robotics and Automation meeting, jointly organised by the RSC Biological and Medicinal Chemistry Sector (BMCS) and CICAG. This is part of the series of BMCS Hot Topics meetings, which are run a few times a year as fully virtual, half-day symposia on topics of interest to practitioners of pharmaceutical and biological chemistry.

I was glad to be able to open the meeting with a half-hour presentation on the crucial importance of apt data management and streamlined workflows, demonstrating the need to fully capitalise on expensive robotic systems rather than waste their potential. Following this, Blake Baker provided an excellent talk about democratisation of automation at Astex, where automation experts have used such workflows to enable non-specialist chemists to use high-end robotics for rapid library synthesis. In the second session, Andy Cooper from the University of Liverpool introduced fantastic mobile robots that work autonomously to discover and optimise in multiple fields including catalysis, formulation, and materials chemistry. Julie Fournier then discussed the use of automation in nanoscale, high-throughput chemistry at GSK, exemplifying its value in direct-to-biology applications where purification-free methods allow crude mixtures to be tested directly in bioassays. The third and final session featured a talk from Ian Fairlamb at the University of York, who neatly illustrated how miniaturisation and parallel reactions can be used in combination with careful analysis to investigate mechanistic hypotheses. The day's final talk was from Marty Burke at the University of Illinois at Urbana-Champaign, discussing "a world where anyone can make molecules": the power of robot-enabled synthesis to expand the toolbox of modular, synthetically-useful molecule building. Altogether, the afternoon was extremely enjoyable and worthwhile, and the Q&A sessions particularly represented a valuable chance for attendees to grill experts in many topics related to both robotics and software workflows in chemistry.

The next BMCS Hot Topics meeting is on Antibody-Drug Conjugates, held online in November 2025.

- -

Celebrating the Life and Legacy of Dr Olga Kennard

Contribution from Dr Wendy Warr, Wendy Warr & Associates, email: wendy@warr.com

Report reproduced with permission.

Extract from *Chemical Information and Computation* 2024, No.1, ACS National Meeting, New Orleans, 17-21 March 17-21, 2024.

Dr Olga Kennard (23 March 1924 – 2 March 2023) was a pioneer in structural science and a visionary leader who saw the potential for the collective use of data to accelerate the advancement of science. To recognise her many contributions to science and society at large, a symposium was held in her honour at the Spring ACS Meeting in New Orleans in March 2024.

In 1965, Olga Kennard established the Cambridge Structural Database (CSD) and she was the founding Executive Director of the Cambridge Crystallographic Data Centre (CCDC) until her retirement in 1997. During that time, the CSD evolved to provide a platform for knowledge-based software applications that today inform research challenges in structural chemistry across industry and academia. As well as her pivotal role in establishing the CSD, she made significant contributions as a crystallographer, solving the first structure of ATP and various nucleotide structures. Olga excelled as a successful woman in science during a time when such accomplishments were uncommon. She became a Fellow of the Royal Society and received several awards recognising her scientific contributions throughout her lifetime. Beyond her scientific endeavours, she nurtured a keen interest in art, architecture, and design, and served as a Trustee of the British Museum.

Life of Dr Olga Kennard: a visionary leader

***Suzanna Ward, Ian Bruno**, The Cambridge Crystallographic Data Centre, Cambridge, UK*

Dr Olga Kennard was a true pioneer. Born in Hungary, she moved to the United Kingdom in 1939 just before the outbreak of the second world war. After graduating from the University of Cambridge in 1944 she joined the Cavendish Laboratory under Lawrence Bragg as an assistant to Max Perutz.[1,2]

Olga became a prominent and active member of the crystallographic community. She was a member of the IUCr Commission on Crystallographic Data from 1957-1966 and chair from 1966-1969. She was a founder member of the British Crystallographic Association and the UK's representative to the European Crystallographic Committee in 1973. In 1975 she was elected President.

Olga and John Bernal "had a passionate belief that the collective use of data would lead to the discovery of new knowledge which transcends the results of individual experiments".[3] Following this vision, Olga established the Cambridge Structural Database (CSD) in 1965. She helped set up the Protein Databank (PDB) and was a guiding influence for the establishment of the Nucleic Acid Sequence Database.

Early CSD creation involved punch cards and knitting needles in an elaborate set of checks so that the database could be used with confidence. During her time leading the CCDC, Olga oversaw the initial delivery of the data in book form, the addition of computer-produced diagrams, the development of a database,[4] new ways to search and analyse the data, and the establishment of a knowledge base derived from the database.[5]

In 1988 she was awarded the OBE "For services to Scientific Research on the Structure of Biological Molecules". Stepping down from her role as director of CCDC in 1997 gave her more opportunity to pursue her interests in art, architecture, and design. At the age of 91, she gave the opening lecture at the 50th Anniversary of the CSD, in 2015 at Downing College, Cambridge.

Protein Data Bank: from two epidemics to the global pandemic to mRNA vaccines and Paxlovid

Stephen K. Burley,[1,2,3] *(1) RCSB Protein Data Bank, Piscataway, New Jersey, USA (2) Institute for Quantitative Biomedicine and Chemistry & Chemical Biology, Rutgers the State University of New Jersey, New Brunswick, New Jersey, USA (3) Cancer Pharmacology Research Program, Rutgers Cancer Institute of New Jersey, New Brunswick, New Jersey, USA*

Stephen described how the PDB has evolved following Olga's guiding hand. Structural biologists and the Protein Data Bank (PDB) played decisive roles in combating the COVID-19 pandemic. Global three-dimensional (3D) biostructure data was turned into global knowledge, allowing scientists and engineers around the world to understand the inner workings of coronaviruses and develop effective countermeasures against SARS-CoV-2.

Despite the SARS and MERS virus crises, no one was prepared for COVID-19, but the PDB came to the rescue and the PDB now offers structures on open access for all three viruses. State-of-the-art mRNA vaccines, initially designed with guidance from single-particle cryo-electron microscopy structures of the SARS-CoV and MERS-CoV spike proteins, benefited more than five billion individuals around the world by preventing viral infections entirely or significantly reducing morbidity and mortality.

Structure-guided drug discovery efforts at Pfizer, first initiated in the 2000s in response to the SARS-CoV epidemic and reactivated in 2020 early in the global pandemic, yielded nirmatrelvir, a potent, orally bioavailable, covalently acting, peptidomimetic inhibitor of the SARS-CoV-2 main protease.[6] This targeted anti-viral drug received emergency use authorisation from the United States Food and Drug Administration in December 2021, less than two years following public release of the viral genome sequence. It is used clinically for the treatment of acute SARS-CoV-2 infections in a fixed dose combination with ritonavir and sold under the brand name Paxlovid.

Bolstered by open access to research data generated with public and private monies, particularly 3D structures of coronavirus proteins and their complexes with one another, with antibodies, and with small-molecule inhibitors archived in the PDB, basic and applied researchers made a difference when the world desperately needed them to succeed.

There is a sobering postscript.[7] The market failed us: if only Pfizer had been incentivised by public money in the 2000s. Millions of lives could have been saved if discussions on public investment had taken place after the MERS epidemic. Fortunately, Pfizer invested huge sums of money into COVID-19 research in 2020-2021. Underscoring the importance of these contributions, Dr Anthony Fauci, former head of the National Institute of Allergy and Infectious Disease, said "Show me a person who's vaccinated, got infected, took Paxlovid and died. I can't find anybody."

Olga Kennard as an icon of interdisciplinarity

Rivka L. Isaacson*, Chemistry, King's College London, London, UK*

Rivka orientated Olga Kennard within the interdisciplinary narrative of X-ray crystallography, archiving molecular structures, creative mentorship, visual design and appreciation, and Rivka's own arts–science exploits at the National Portrait Gallery and beyond. She weaved her memories and experiences with Olga into an exploration of Olga's legacy that inspired the PDB. A variation of this talk has already been published.[8]

In 2017, Rivka was invited to lecture at the National Portrait Gallery (NPG) for International Women's Day. Olga was the only surviving one of the five female structural biologists in the NPG's collection, so Rivka wrote to Olga in preparation for the talk and received an invitation to Olga's house in Cambridge. This meeting led to an interesting talk at the NPG, a *Times Higher Education* article and subsequent visits to Olga's house. Rivka

shared some of the many wonderful things she had seen and learnt through Olga including a menu from Dorothy Hodgkin's 80th birthday celebration with the signatures of many distinguished scientists.

At the age of 80, Olga was appointed by the Royal Society as a trustee of the British Museum and served from 2004 until 2012. In the liaison for a British Museum new building documenting 20th century Europe, she was a moral voice for the museum's purpose and brought out the best in everyone. She was a peacemaker and valued everyone's opinion.

Rivka herself works on translation on the ribosome. She showed two structures she has been involved in solving: a novel RNA polymerase-binding protein that interacts with a sigma-factor docking site[9] ('Fin') and CsfB,[10] also called "Gin". Structure solution by NMR of the latter was featured in a PDBe article "Gin with a twist".

As part of the PDB Art project, the Protein Data Bank in Europe (PDBe) works with school art departments from around the world to create artworks inspired by the molecules of life. Rivka showed examples of the stunning outwork including "Snow leopard with myoglobin".

"Viewing the invisible", an event at the Arcade, Bush House in 2019, was an interdisciplinary collaboration between King's College London, London Fine Art Studios, and the National Portrait Gallery, bringing together scientists and artists to explore the similarities in their working methods. The project used painting and film to highlight how scientists and artists alike seek to demystify human identity. Rivka concluded with photographs taken at that exhibition.

The chemist and the architect

***Dirk Trauner**, Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania, USA*

To imagine a structure and then express it in material form is one of the most satisfying human activities. It is pervasive throughout the arts and crafts, and it is one of the defining features of architecture. It is also at the heart of synthetic chemistry.[11] Dirk never knew Olga but he loves the CSD. Do architects know about the CSD?

Kekulé is one of the fathers of chemical structure theory. He famously "saw" carbon atoms joining in a "giddy dance" in a daydream. Later he saw dancing strings of carbon atoms, but this time he saw one that, snakelike, took its tail into its mouth, which gave him the idea for the ring form of the benzene molecule. His student Jacobus van't Hoff (the first Nobel Laureate in chemistry) greatly valued the power of imagination in scientific work. He proposed the development of 3-dimensional chemical structural formulas. Richard Willstätter (who noted that cyclooctatetraene was not aromatic) and Eugene van Tamelen (who synthesised bicyclo[2.2.0]hexa-2,5-diene, Dewar benzene) were both interested in architecture. σ-Bishomoaromaticity is an extreme case of the chemical bond and aromaticity. It was found by Horst Prinzbach's group when they synthesised dodecahedrane via the pagodane (CSD VABNAN) route.

Both chemists and architects like to sketch. Dirk showed sketches by the Danish architect Jan Utzon and the chemist Samuel J. Danishefsky. Both chemists and architects like clean line drawings; Dirk showed various publication-quality structures of morphine and Andrea Palladio's plan of the Villa in *I quattro libri dell'architettura*, 1570.

Chemists and architects like models; Dirk showed photos of the architect Zaha Hadid with a model and Robert B. Woodward with a model of vitamin B12. He also showed a model of high-rise buildings and models from the CSD, all 3D-printed.

Both chemists and architects also like to build. Dirk compared stages in the building of the Eiffel Tower with the total synthesis of temuline.[12] The constructive process is very similar both in timelines and in auxiliary construction. Buildings and molecules occasionally collapse. The eastern semidome of the Hagia Sophia (designed by the Greek geometers Isidore of Miletus and Anthemius of Tralles) collapsed in 558 but was reconstructed by 562. Tetracyclo[4.4.1.0[3,11].0[8,11]]undecan-4-one (CAS RN 1824494-08-7) collapses to 2-(3-Buten-1-yl)phenol (CAS RN 14003-84-0) on heating. Chemists like copies, but architects may not like rows of identical houses. Dibefurin has "copies" in its structure (CSD POVBIF).[13]

Chemists and architects are dreamers. Étienne-Louis Boullée proposed a huge, spherical cenotaph for Isaac Newton, claiming that the most beautiful and perfect natural body is the sphere. The use of light in the building's design was intended to create the effect of day and night. The first attempt at synthesis of dodecahedrane (CSD DOPSIK) was initiated in 1964 by Robert B. Woodward who had hoped that triquinacene might to dimerise to dodecahedrane: it did not. Dodecahedrane (CSD DOPSIK) was synthesised by Leo Paquette in 1982. The pagodane of route of Horst Prinzbach was published in 1987.

Trauner and his colleagues have proposed and analysed in detail the contiguous linear extension of twistane with ethano (ethane-1,2-diyl) bridges to create a new chiral, hydrocarbon nanotube called polytwistane.[14] Shockingly, there is no X-ray structure of twistane in CSD. Carbon nanothreads have since been produced by formed by a high-pressure, solid-state reaction of benzene.[15] Dirk finished his talk with some pictures of carbon nanothreads compared to high-rise buildings.

In his talk, Dirk highlighted Olga's passion for architecture and design. He used molecular models from the CCDC and PDB databases to reveal the similarities between synthetic chemistry and architecture, between molecules and buildings. He showed how chemists and architects share experiences and character traits that would make them equally successful in either field, helping to explain how Olga had managed to use her skills to make an impact in the world of architecture too.

Olga Kennard as research mentor in the mid-1970s: a DNA fragment, dipeptides, an early use of disorder tools, and cannabidiol

Lawrence R. Falvello,[1,2] *(1) Universidad de Zaragoza Departamento de Quimica Inorganica, Zaragoza, Spain (2) Instituto de Nanociencia y Materiales de Aragón, Universidad de Zaragoza Facultad de Ciencias, Zaragoza, Spain*

Lawrence ("Larry") was a first-year graduate student supervised by Olga Kennard in the mid-1970s, in what was a small but productive group studying the structural properties of biologically interesting compounds. Olga's ability to undertake challenging, interesting problems led to the study of a variety of structures involving different difficulties with different, novel solutions.

Olga's "signature" structure (adenosine triphosphate, CSD ADENTP)[16] was hard to get: the compound was hard to crystallise. Larry showed a list of Olga's 151 structures, including vancomycin, and her 24 entries in PDB. Larry recounted his first meeting with Olga. In addition to presenting her vision of the research that Larry would be doing, she showed him a room where there were 41,000 photocopies of crystal structure reports: an early glimpse into the database that was then taking shape under her leadership. On the research side, the first task she assigned him was to build a density gradient column. Beyond that, his first major project was the separation and purification of a DNA fragment that was expected to yield a double helical structure.[17] Olga's approach to the DNA project illustrates the most important thing he learned from her. Steve Salisbury had synthesised the compound, and Larry was tasked with isolating the compound using column chromatography. This compound was not easy to synthesise or to purify, but the successful conclusion of this project made a lasting impression on Larry, so much so that when he set up his own first laboratory it was in fact a synthesis laboratory.

The crystal coordinates of the DNA double helical fragment were published four years later,[18] long after Larry had left the research group. The publication mentions ion exchange chromatography and says that "structure determination was not straightforward". This was the first structure of a DNA double helical fragment at atomic resolution. Olga had a talent for bringing people together to solve difficult problems. Pictures from PLUTO were informative but nowadays you can colour the fragments and make visualisation clearer. [Note, PLUTO was later replaced by Mercury.]

Larry also worked on crystallisation of dipeptides.[19] This led to an example of Olga's talents with hardware and aligning crystals. Another example from this era in the Cambridge laboratory was the early use of disorder tools. In the crystal structure of dicyclohexylammonium arsonomethylphosphonate, the anions are disordered, with P and As atoms interchanged in the two orientations.[20] The disorder tools using constraints and restraints in the program SHELX-76 were found to be key to successful refinement. George Sheldrick wrote SHELX-76, the first of the SHELX programs for small molecule refinement.[21]

Larry's first successful structure analysis, in 1977, was that of cannabidiol,[22] for which there was evidence of anti-epileptic activity. Structure solution by direct methods encountered difficulties caused by pseudo-symmetry. SHELX did many jobs in one program and Larry finally got good refinement. He also did some energy calculations with the program EENY, written by Sam Motherwell. It was not until 2017 that there were other structure determinations of cannabidiol. More recently, there have been new polymorph publications and cannabidiol has been approved for treatment of seizures.

Years later, when Larry was leaving Cambridge, Olga gave him a book of poetry by the crystallographer Alan Mackay. Crystallographers call this the "Richards Box" which was an indispensable tool[23] used by crystallography laboratories around the world for about 10 years, when it was finally replaced by computer graphics systems.

Concomitant with the activities of Olga's research group were those of the much larger staff of the Cambridge Crystallographic Data Centre. The research group had access to batch-mode searches in the database, using BibSer and ConnSer, a luxury that presaged today's widespread use of the CSD as an integral component of research programs in structural chemistry.

Olga Kennard and the Cambridge Crystallographic Data Centre: an inspiration for modern scientific data science. An outsider's view

***John Rumble**, R&R Data Services, Gaithersburg, Maryland, USA*

Over the last five decades, every area of science has become data-oriented, an incredible achievement that was anticipated by a few visionaries, such as Vannevar Bush (Science, the endless frontier: a report to the President, in 1945), David Lide,[24] and Olga Kennard.[25] As an outsider, John reflected on Olga's work and that of the Cambridge Crystallographic Data Centre that has heavily influenced today's scientific data science. These contributions include:

- The importance of metadata and metadata standards in reporting a scientific measurement
- Building a community-driven data repository and addressing its long-term support
- Mandatory data deposition as a requirement for scientific publication
- International cooperation on scientific data work
- Professionalism of scientific data work
- Support of quality assessment of scientific data

John first addressed the issue of standards. Prior to the establishment of the CCDC, X-ray diffraction studies were very difficult. There were few if any standard procedures, instruments were home-made, everybody used their own unique conditions and specialised software, and results were reported iconoclastically. Today, there are standards for what must be controlled and reported and for data and metadata formats. These standards ensure that diffraction studies are published with appropriate amounts of information. Because standards are used only if developed by the measurement community, CCDC's leadership in involving the community and collaborating with IUCr resulted in widespread adoption of the Crystallographic Information Framework (CIF).[26] Work on metadata and standards was an international team effort. John listed the names of many contributors.

Another CCDC contribution was the CSD repository and deposition of data. Although in 1960 CAS identified about 5 million papers related to chemistry, very few of those papers made their data available. Crystallography was the first field to impose mandatory data deposition, working with IUCr and other publishers. Large quantities of data have been saved. Unfortunately, this model has yet to be fully replicated for other chemistry disciplines.

Next, John turned to international cooperation on scientific data. In the past, there was no coordination in terms of collecting and preserving measurement results among instrument manufacturers, academic scientists, data users, collectors, and evaluators, and industrial researchers. Today, cooperation has emerged. Changes and improvements are driven by community needs. There is recognition, though reluctant, of the need for long-term financial support, a need the CCDC was the first to recognise. Today, there is a clear delineation among crystallographic data centres for inorganics, metals, and proteins, and powder diffraction and single crystal studies, each utilising a different model for long term support. Nowadays, scientists from many countries work together to create a data resource of the highest quality and impact because of community-driven data repositories, mandatory data deposition, and metadata and standards. CCDC, through its partnerships with crystallographic data groups worldwide, demonstrated how such cooperation benefits the scientific community.

Professionalism in scientific data work is relatively new. In the past, people interested in scientific data were data zealots. Astronomers collaborated for centuries. Beilstein and Gmelin were launched in Germany in the late 1800s. The National Bureau of Standards (NBS) worked on atomic spectra from 1945. Other fields had their zealots. The CCDC under Olga's leadership was one of the first stand-alone entities that focused on providing a professional atmosphere for scientific work, from junior staff and non-scientific assistants to senior scientists. Prestige has been established for scientific discoveries through scientific data science. Paths of promotion were possible, and professional data publications were recognised as important scientific contributions. Data are making an increased impact in academic research and industrial development.

Finally, John addressed data quality. In the past, evaluation activities were done on many specialised chemistry topics (e.g., by Landolt-Börnstein, Gmelin, NBS, and many others). These had different editors, and no metadata standards. Comprehensiveness was elusive, and data collection and extraction were carried out by individual evaluators. Now, the information revolution has made comprehensiveness possible. Collection is facilitated by author deposition. Metadata standards have allowed development of software tools. Databases of evaluated data allow meta-analysis (i.e., knowledge discovery). In conjunction with other crystallographic data centres, CCDC has been a leader in developing tools to facilitate quality assessment of scientific data.

The CCDC has been a pioneer in establishing modern scientific data work. Olga Kennard, through her innovation and leadership, has been an inspiration for modern scientific data science.

Olga Kennard: the CSD and wider contributions to the community

***Jeremy Sanders**, Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK*

Jeremy knew Olga personally and professionally for more than 40 years. The Cambridge Jewish Reform Group was founded in 1976. Its first social events and services were held in peoples' homes on Friday night and on festivals. In 1981, the Reform Group became Beth Shalom Reform Synagogue but remained without its own building until 2015. Several early events were held in Olga's home, designed in 1960 by Eric Christian Sørensen and now a nationally listed building. The furniture and decorations designed by Sørensen are unchanged, over 60 years later. The study of the house is featured in the National Portrait Gallery portrait of Olga, dated 1988.

Jeremy and Olga were also colleagues in the Cambridge Chemistry Department, although Olga and her unit were actually employed by the Medical Research Council. Jeremy and Olga were both close to the Head of the Department of Organic, Inorganic and Theoretical Chemistry at Cambridge, Ralph Raphael, who was very supportive. Olga was also a pioneer in employing mothers part-time and at home.

Jeremy's first experience of the CSD (in 1981) concerned an unexpected conformational preference detected in solution by NMR.[27] The same preference is found in numerous structures in the CSD but the correlation had never been looked for or noticed. The CSD contains all the answers to chemistry if only we knew the right questions to ask. The CCDC building in Cambridge (1991) was designed by Eric Christian Sørensen. It is wrapped around the eastern end of the Chemistry Department. The adjacent Unilever Centre for Molecular Informatics and the library (1998) were also designed by Eric Christian Sørensen.

Two of Olga's retirement activities were being a Trustee of the British Museum (2004–2012) and serving on the committee choosing the architect for Beth Shalom Reform Synagogue. The synagogue was designed by Cowper Griffiths and opened in 2015. A "Tree of Life", dedicated on March 10, 2024, was donated by Olga's family in her memory. Each leaf is donated by members to commemorate the life of loved ones.

There was much more to Olga than crystallography. She made great contributions to Cambridge architecture and community life through her house, the CCDC and a synagogue, and was a Trustee of the British Museum.

Approximate symmetry in organic crystal structures having $Z' > 1$

***Carolyn P. Brock**, Chemistry, University of Kentucky, Lexington, Kentucky, USA*

Carolyn first visited the CCDC in 1981. Her first study based on the CSD was published in 1989. She talked about seven of her 14 CSD-related studies. Unusual crystal structures used to be attributed to "crystal packing effects". Biphenyl, which is (on average) planar in the solid state is the classic example of a conformation being stabilised by crystal packing. Carolyn's first study showed that systematic packing effects are possible but unlikely.[28] This is not like a Boltzmann distribution. Twenty-three years later, a CCDC team showed that systematic conformational bias in small-molecule crystal structures (including 690 biphenyls) is rare and explicable.[29] Their work showed that concerns, expressed by some drug discovery chemists and other scientists, that conformations observed in crystal structures are systematically biased due to crystal packing effects is overstated.

Wallach's rule,[30] that racemic crystals tend to be denser than their chiral counterparts. has been substantiated by CSD.[31] Wallach's 1895 rule was based on eight pairs of densities, Carolyn's 1991 paper had 129 pairs. Wallach's Rule is true but its predictive value (if any) cannot be proved by comparing pairs of structures. Carolyn realised that one- and two-component systems (i.e., whether the enantiomers interconvert or not) must be considered separately. If the enantiomers cannot interconvert the racemic phase is unlikely to be found unless it is more stable than the enantiopure phase. The two phases melt to give different liquids so the racemic phase must be compared to the eutectic of the enantiomers.

WEJDOE (P21, Z'=3) was a random crystal structure that led to an interesting generalisation.[32] Carolyn knew that Z'=3 structures were unusual and realised that the Z'=3 allowed the formation of an OH-bonded helix. A survey of the CSD showed that hydrocarbons with a single OH substituent commonly crystallise with more than one molecule in the asymmetric unit, or in space groups with three or fourfold axes. Fifty-five structures were studied[33] in 1994. In March 2024 the study was repeated.[34] The authors could not reproduce the 1994 criteria, but the 2024 criteria are at least as strict. The 2024 survey took only a few hours.

Carolyn became interested in high-Z' structures. Approximate symmetry is very common[35] if Z'>4. Approximate symmetry is pervasive in organic substances that crystallise in low-symmetry space groups.[34,36,37] 3D approximate symmetry is dominated by approximate translations and inversion mimicry, which are about equally important. 3D approximate symmetry is often translational; an example is KULYIC (C2, Z'=3). Robin Taylor wrote an algorithm for Carolyn in that study. Mimicry of non-Sohncke packing by enantiopure material is very common [e.g., ICIYUP (C2, Z'=2; C2/c, Z'=1)]. Inversion is so favourable for crystal packing that achiral parts of the molecule may mimic inversion even if the whole structure cannot [e.g., RUNQIC (C2, Z'=2)].

There are, however, some examples of other types of 3D approximate symmetry. An example is VENRUC02 (C2, Z'=2 at 300 K) and Iba2, Z'=1 (achiral molecule). Crystals were grown by slow evaporation at room temperature in the dark. The structure is the same at 100, 150, 200, 250, 300, and 323 K; it seems to decompose irreversibly above 353 K. In Iba2, gamma would be 89.7° at 300 K. The approximate symmetry is obvious but so are the deviations from crystallographic symmetry.

Layers quite often have higher approximate symmetry than the full 3D structure: a example is PEXKEI (C2, Z'=2; p2122, z'=1). Structures having obvious layers related by a 2x axis are surprisingly common; in LAFCOL (C2, Z'=2) adjacent c211 layers are offset by 0.16a (3.6 Å). These structures are not rare for space group 5 (C2, I2, A2): 50 structures (9% of all considered). The layers have symmetry c211 and are very well related by a twofold rotation but are offset by a non-simple amount.

Worries about crystallisation changing the molecular conformation are mostly in the past. We now know that conformations of independent molecules are usually very similar. The CCDC program Mercury was of fundamental importance in the research work reported and Carolyn acknowledged the contributions of several CCDC scientists. Above all, she paid tribute to Olga for her vision for the possibilities of the CSD and her tenacity in driving its development.

From vision through data to knowledge: the insights of Dr Olga Kennard

***Ian Bruno**, The Cambridge Crystallographic Data Centre, Cambridge, UK*

Ian used as inspiration for his talk the lecture Olga gave in 1995 about J.D. Bernal and also used her contribution in 1997 to a workshop on electronic publishing.[3] Regarding the CSD, she said "The database was established in 1965 to fulfil a dream of myself and a great scientist, the polymath J.D. Bernal. We had a passionate belief that the collective use of data would lead to the discovery of new knowledge which transcends the results of individual experiments".

As far back as the 1940s, Bernal foresaw the coming information explosion and the valuable knowledge that could be derived from this, given the "long and tedious process of capturing information and developing tools". In 2024, we are living in an age of big data, AI, and misinformation. Olga pointed out in 1995 that visualisation is particularly important since few people are able to think about structures in three dimensions. She also recognised that "there is also a vast amount of information, produced by scientific workers in the past, which has not, as yet, been fully exploited".

She pointed out the challenges posed by restrictions on fundings, mixed popular reactions to science, and the lucrative business model of some publishers. While the way in which the results of scientific projects were disseminated had not as yet changed radically, there was a risk of uncontrolled publication of results on the web and a flood of unreviewed research papers. Olga said "we were determined that our database would have numbers that could be relied on. So, a most elaborate set of checks was introduced, including the recomputing of the distances from coordinates ... about 20% of the published structures contained errors and inconsistencies which had to be corrected, often by correspondence with the authors". CCDC's experience was that data centres could play a significant role in the reviewing process. In 2024, CCDC found that 900 structures known to be associated with ChemRxiv preprints were not known to be published in a peer-reviewed journal.

In 2024, CCDC has an online deposition process with embedded data integrity checks and internal duplicate checks (automated and manual). More than 90% of the data reflects data in associated peer reviewed scientific articles. CCDC has strong relationships with major publishers and a reviewer request service. A team of expert PhD-level editors and scientists at CCDC, including a data integrity scientist, support the curation of CSD data. CCDC has workflows to identify some instances of fraud in CSD-deposited datasets. It is a member of the Committee on Publication Ethics and has strong connections with IUCr, individuals, and some publishers, and it uses other external resources. It has a clear retraction policy and workflow.

At the 50th anniversary of CCDC, Olga said that they decided "not to use the atom coordinates as listed in the paper, but to use the atoms and the symmetry-related atoms transformed to orthogonal coordinates": a simple idea but unorthodox when CSD was in its infancy. She further said that correlating the 3D structures of the molecules with their 2D chemical structures made it possible to search for a combination of chemical and structural attributes. These ideas contributed greatly to the rapid take-up of the system by communities outside of crystallography and the re-use of CSD data.

In 1995, Olga said "It now seems that the understanding of many diverse biological events depends critically on a knowledge of the three-dimensional structure of the molecular components involved". About 3000 protein structures had been determined at that time but only a small percentage of the hundreds of thousands of proteins could be analysed in crystals. Olga said "It would give a revolutionary new direction to our understanding of the detailed function of proteins if three-dimensional structures could be derived from sequence alone." In 2021, AlphaFold[38] was able to do that highly accurately. Nevertheless, while AlphaFold predictions are valuable hypotheses, they accelerate rather than replace experimental structure determination.[39]

An important function of a data bank is the reduction of a collection of individual datasets to generalised sets of constants and rules. This is illustrated by work to build knowledge bases of bond lengths and intermolecular forces, as recognised by Olga in her 1997 lecture. In 1995, CCDC was collaborating with Bob Pearlman to evaluate and improve CONCORD, then one of the best-known 3D structure generators.

In 1997, Olga said, "Free access to validated and enhanced data worldwide is a beautiful dream. The reality, however, is more complex…". Data input needed to be standardised (CCDC was heavily involved in CIF26 specification) but a "priority problem" was that some journals were formulating rules that deposition on the web counted as prior publication, meaning the work would not be accepted for publication.

In 1997, users wishing to access individual data items were well-served by electronic networks but the ability to install the database in individual laboratories was needed for in-depth research. CCDC used a human network of affiliated centres in 40 countries to distribute data and software on CD-ROM because each national centre knew the needs of its local community and could channel funds adjusted to the size of the country's scientific population. Olga introduced this policy in the 1970s when there was no concept of actually paying for

information, in order to ensure that young scientists, in particular, had access to the database. Olga put the CSD on a sound financial footing. She gained income from industrial companies so that CCDC did not have to rely on special grants from government sources, unlike some data centres, whose data were accessed, free of charge, and whose electronic existence could well be threatened if the special grant were withdrawn. In 2024, CCDC enables and engages with research initiatives that connect industry and academia. The CCDC Crystal Form Consortium, for example, has led to the development of tools that use knowledge from the CSD to aid drug product development.

Reflecting Olga's 1997 visions, CSD technological developments have kept pace with the exponential growth of information; validation tools have been developed to ensure reliability; data have been enhanced to enable reuse; knowledge bases have been developed so as to reduce labor; and by channeling resources, data banks have ensured that information once gained is not lost.

Olga delivered her Ewald Prize Lecture in Prague in 2021. In interviews telling the IUCr about her work and being awarded the 12th Ewald Prize she made the following comments:

"How do I see the future of the Cambridge Structural Database? Unstoppable probably … At present I cannot anticipate the time when such data could become superfluous".

"Even though we know a great deal about the formation and structure of crystals, … purely theoretical structural predictions are not likely to take over from good experimental data".

"I am now in my late 90s, an avid user of the internet and enjoying the power of computing. I am too old alas to keep up with new publications, but I can still appreciate, from the sidelines, the enormous expansion of scientific knowledge and mourn the fact that it is not being made full use of to ensure the future viability of our planet and civilization".

Integration of CCDC tools into pharmaceutical solid form development

***Rajni Bhardwaj**, Material Sciences, Drug Product Design, Pfizer Global Research and Development, Groton, Connecticut, USA*

Over the past decade, the pharmaceutical industry has witnessed an ever-increasing utilisation of CCDC tools to complement experimental crystal and particle engineering approaches during pharmaceutical development. It is essential to control polymorphism.[40] There are two well-known cautionary tales. In 1998, manufacturing problems hit Abbott's HIV drug ritonavir.[41] Abbott had to withdraw ritonavir and this had severe financial implications. UCB's rotigotine is a Parkinson's disease administered in patches. A previously undiscovered polymorph of rotigotine emerged with an altered 3D conformation and subsequently a different packing arrangement of rotigotine molecules. When rotigotine crystallised in the patch, UCB's share price fell. The PCT patent for rotigotine filed in 2003 states that no polymorphism was observed during development.

It has been shown[42] that application of informatics-based assessment of compounds complements experimental studies and provides deeper understanding. Information from structural analyses is incorporated into risk assessments. Informatics techniques are quick to apply and are straightforward to use. Full interaction maps and hydrogen bond propensities generated using the CSD can be used to assess polymorph stability.

The CCDC Crystal Form Consortium (CFC) is a collaboration between industry and the CCDC to provide structural informatics tools and approaches for the rational design of the solid form. Solid form "health checks" provide digital risk assessment,[42] looking for unusual conformations and optimal H-bonding and investigating if there is a better way of crystal packing.

Rajni presented some case studies. The first[43] demonstrates how crystal structure prediction, hydrogen bonding propensity analyses, and rational solvent selection for stable form screening were used in in silico risk assessment of solid form selection for crizotinib (CAS RN 877399-52-5) and axitinib (CAS RN 319460-85-0).

The second study[44] started when the solid form screening of galunisertib (CAS RN 700874-72-2) produced many solvates. Rajni and her co-workers then investigated possible risks to the development of the favoured monohydrate form, using full interaction maps analysis plus an unusual range of experiments, including melt crystallisation under pressure. The structure with a global minimum in an intermolecular probability map was kinetically inaccessible. Ten unsolvated polymorphs of galunisertib were found but the structure predicted to be the most stable has yet to be obtained.

Another application of CCDC tools concerns paritaprevir (CAS RN 1216941-48-8)[45] which has a 15-membered macrocyclic ring and high conformational flexibility. Rajni and her colleagues used a combination of computational and experimental methods, focusing on molecular chameleonicity and crystal form complexity. The ability of the molecule to toggle between high and low 3D polar surface area conformations is underpinned by intramolecular hydrogen bonding interactions and intramolecular steric effects. Crystal packing and protein binding resulting in strong intermolecular interactions disrupt these intramolecular interactions. Crystalline form I benefits from strong intermolecular interactions. The weaker interactions in form II are partially compensated for by the energetic advantage of intramolecular hydrogen bonding. Strategies for derisking the choice of metastable form II are reported.

Paritaprevir and ombitasvir (CAS RN 1258226-87-7) are part of a triple combination, antiviral regimen (CAS RN 1799419-19-4) as a fixed-dose tablet. Form II of paritaprevir and form I of ombitasvir are high stoichiometry crystalline hydrates. The soft viscoelastic nature of the dominant surface of paritaprevir Form II and the slightly softer and more elastic dominant surface of ombitasvir form I have been studied.[46] The elastic features of the dominant surface of ombitasvir in combination with well-designed process controls allow crystals to heal from or avoid the loss of mechanical integrity when undergoing two solid-solid phase conversions in the commercial isolation process.

Rajni's case studies demonstrate how the structural knowledge derived from thousands of crystal structures present in CSD can be used to get a better understanding and to inform key decisions in pharmaceutical solid form development. Olga Kennard had a pioneering vision that has led to valuable computational advances in polymorphism and "right-sizing" of polymorph screening.

Teaching crystallography to chemistry students using the Cambridge Structural Database

***Diane A. Dickie**, Department of Chemistry, University of Virginia, Charlottesville, Virginia, USA*

X-ray diffraction (and small-molecule, single-crystal X-ray diffraction in particular) is widely used in academic and industrial research for the characterisation of new molecules. It is disappointing, therefore, that crystallography rarely appears in chemistry courses even though other characterisation methods such as NMR or IR spectroscopy are introduced to virtually all students. It is also surprising how few students are aware that using a specialised scientific database like the CSD will frequently turn up a far more useful and relevant set of responses than entering a query into the search bar of their favourite web-browser or artificial-intelligence chatbot. Diane shared examples of how she uses the CSD and other tools created and maintained by the CCDC to teach key crystallographic concepts to chemistry students at the graduate and undergraduate level.

CellCheckCSD is a freely available, automated tool for checking unit cells against the CSD. In a crystallography class, Diane uses it to teach students about identifying the components of a mixture. The mixture chosen is dextrose, citric acid, and ascorbic acid. It is used in fruit preservation and is safe for students to use since it is

not toxic. Students consider if their crystal is the same as others. This prompts questions about crystallography and students are asked whether anyone has picked out a citric acid crystal.

Another example is confirming the identity of a product made in a synthetic chemistry class. The exercise requires access to a diffractometer but only three or four minutes of data collection. Students are typically thrilled to see crystals under a microscope and to discover that they have made the right molecule.

The next example does not require access to a diffractometer. CSD tutorials on the CCDC website are useful for learning skills independently if a tutor is absent or if a diffractometer is in use. Students learn how to use ConQuest for searching 3D structures in the CSD and are asked why they used it and what they liked least or best about the tool.

They can learn in more detail about ConQuest (Mercury to view 3D chemistry and produce graphics) and Mogul (to assess molecular conformations). Since the mid-2010s, coordinates plus data in CSD have been an abundant source of practice material. Over four years, at least one student per class claimed to discover a "problem" structure.

So, students are taught how to find problematic structures and fix them. Diane showed a histogram from Mogul and Mercury. Most bond lengths were clustered in one place but there were a few outliers. One is probably a case of disorder, but another had R = 49% (poor) and a student ran checkCIF on the structure. The student found lots of alerts that included an indication of voids which CCDC had also noticed. The student tried to refine unassigned electron density. The R value improved significantly.

Another search was carried out for high R values. Some of the structures found had justifiably high R values but two (with R of 66% and 41%) were useful for teaching purposes. In one of these, there were actually two structures and applying symmetry led to a decrease in R value. The structure in the publication was correct but the one deposited in the CSD was wrong. In the other high R value example, lots of bond lengths were outside the norm. The researchers may have used an older algorithm, and they used the wrong tautomer. The students were thus able to correct an entry in the CSD.

The CSD is not just a database, it also has lots of useful tools. Diane showed the impact of structural data in teaching and how students have gone on to use the resource to learn about the importance of critical evaluation in science.

Future of structural science: building on Olga Kennard's enduring legacy
***Jürgen Harter**, The Cambridge Crystallographic Data Centre, Cambridge, UK*
Jürgen began his talk by outlining the history of the CSD but mostly he talked about the present and future, with six main themes:

- Strategic priorities (vision and mission)
- Data
- Chemistry
- Technological progress
- Functional and legal
- New trends

CCDC's strategic priorities are to provide a unified data experience connecting molecular structures and properties (and add extensive metadata); to be the world leader in enabling structural-data-inspired research;

to enable individuals from all walks of life to contribute to structural science and develop professionally; and to create a thriving ecosystem delivering increasing value for all stakeholders.

In 2018-2023, the community grew, CCDC underwent a turnaround and doubled its revenue, and there was investment in infrastructure, and products, as well as a threefold investment in science. In the five-year plan since then, CCDC plans a fourfold increase in charitable impact, delivering impactful data, science, and education. Enablers will be a unified data experience, an innovation pipeline, efficient processes, professional services investment, and diversified income (commercial revenues, grant income, commercial funding, and donations). Changes such as user behavior, technology (e.g., cloud and AI), open science and FAIR data, increasing complexity, new modalities, and geopolitics (fragmentation) will have to be navigated.

Data present multiple challenges and opportunities. The growth in the volume of data demands more and higher levels of automation and quality control and assurance at scale. The lack of negative data and of metadata, the use of new data sources (e.g., crystal structure prediction), and data dissemination need to be addressed. Data provenance can mitigate against fraudulent data. CCDC has expanded structural data connections with key global data repositories such as the PDB, ICDD, and ICSD. The joint web deposition service for the CSD and the Inorganic Crystal Structure Database (ICSD) is a multistage process to help researchers check and enhance their data. It enables and encourages depositors to comply with community standards and joins up with journal publisher workflows to ensure timely publication of data. Depositors themselves do not have to decide upon which repository is the more relevant.

Across the wider ecosystem, there are many adjacent data sources that can potentially provide deeper understanding when considered in tandem with the CSD:

- Specialty databases
  - DrugBank
  - FooDB
  - Pesticide Properties Database (PPDB)
  - Veterinary Substances Database (VSDB)
- General data resources such as Wikidata
- Biological resources
  - Protein Data Bank (PDB)
  - ChEMBL
  - EMBL-EBI databases
  - WikiPathways
- General chemistry resources
  - Chemical Abstracts Service (CAS)
  - Reaxys
  - SureChEMBL
  - PubChem
  - ChemSpider
  - Chemical supplier catalogs
- Inorganics and materials databases
  - ICSD
  - NOMAD
- Analytical chemistry resources
  - NMRShiftDB
  - Chemotion ELN and repository

- Mass Spectral Database, MassBank
- International Centre for Diffraction Data (ICDD)

Jürgen speculated on the future of search and presented a stylised CSD knowledge graph relating CSD REFCODES, crystal data, elements, molecules, authors, journals, and articles. CSD is in the very early stages of a project to find out how people would use such a graph. This could be a way to improve CCDC's compound registration system.

Chemistry also presents multiple challenges and opportunities, for example in both structural types and searching. CSD molecules are getting bigger and more complex. They include noncovalent assemblies (e.g., rotaxanes), and non-atomistic models of crystal structures (e.g., Möbius strip molecules and primary, secondary, ternary, and quaternary structures of biological molecules). LLMs, GNNs, and foundational models offer opportunities. Good metal-organic cheminformatics methods are lacking. Structural data could be made more useful, for example in solubility prediction.

Jürgen next considered technological challenges and opportunities. There has been an explosion in the volume of data (both experimental and now theoretically calculated) and metadata which CCDC distributes globally. Augmented reality and virtual reality present opportunities in visualisation. CCDC's software portfolios are marching toward the cloud. New scientific innovations and new functionalities must be delivered quickly in new platforms (e.g., SaaS and software ecosystems).

The CSD portfolio today comprises CSD-Core, CSD-Discovery, CSD-Materials, CSD-Particle, CSD-Theory, and CSD-Community. The foundational driver was Olga's "passionate belief that the collective use of data will lead to the discovery of new knowledge, which transcends the results of individual experiments". The CSD knowledge base is used to produce insights. An example is the CSD powering data-inspired medicines manufacturing where structural science has opportunities for analyses of hydrate formation, hygroscopicity, melting properties, solubility properties, stability, dissolution properties, morphology, and disproportionation.

Functional, operational, and legal issues include scalability of operations; sustainability and openness; intellectual property; geopolitics (fragmentation); global coverage, access, and deposition; FAIRness;[47] access for AI and ML (where many new models are coming); and partnerships and alliances that go beyond organic business growth.

Finally, Jürgen summarised some new trends. The CSD has been sustainable for over 58 years and is a trusted CoreTrustSeal repository. Trust principles for digital repositories have been published.[48] Curlew Research has carried out interviews on behalf of CCDC to determine industry's needs for FAIR-enabled R&D. A key interoperability need is to see 3D structural data alongside other physical properties. Also, the physical properties must be linked to a specific sample as well as the identities of chemical entities involved. A key reusability issue is the need for high quality FAIR data. The CSD is recognised globally as the gold standard database for 3D structural data: entries are thoroughly curated and checked for completeness, accuracy, integrity, etc. CCDC needs to build more quality metrics and reflect these in a more objective way as metadata for 3D structural data.

AI and ML techniques are evolving rapidly but the consequences of using poor-quality data can be far reaching: incorrect scientific conclusions, wasted investment and effort, a loss of trust, and ultimately, poor business or research decisions. CCDC has published an e-book entitled "AI for Chemical Discovery" including strategies to advance the vision of AI in chemistry in seven key areas: political will, talent pipeline, the profile of AI in

chemistry, data, infrastructure, ethics, and sustainability. As a sector we need to advertise successes and how they address global priorities. The white paper also includes a look toward the future.

Jürgen concluded with some key messages. Solutions need to work across all of chemistry, not just organic compounds, but inorganics and materials too. Data standards, standard identifiers, and common protocols are essential for enabling discovery and reuse of data across systems and domains. Standards need to be collaboratively developed; scientific unions provide an important focal point. Science is becoming ever more data-driven and multidisciplinary; computation and experiment are commonly used together, and standards need to evolve. There is a vital need for well-described property data that can be linked to structures to enable development of predictive models. Those data must reflect failure as well as success. Data publication is just the beginning: curation and quality assurance are vital for enabling meaningful use in the context of AI and ML. Finally, as well as FAIR we need to assure trust and in particular sustainability.

References

(1) Perutz, M. F.; Weisz, O. Crystal structure of tribromo(trimethylphosphine)gold. *J. Chem. Soc.* **1946**, 438.

(2) Perutz, M. F.; Weisz, O. Crystal structure of human carboxyhemoglobin. Nature (London, UK) **1947**, *160*, 786.

(3) Kennard, O. From private data to public knowledge. In The Impact of Electronic Publishing on the Academic Community; Butterworth, I., Ed.; Portland Press: London, **1997**; pp 159-166.

(4) Allen, F. H. et al. The development of versions 3 and 4 of the Cambridge Structural Database System. J. Chem. Inf. Comput. Sci. **1991**, *31* (2), 187-204.

(5) Bruno, I. J. et al. Isostar: a library of information about nonbonded interactions. J. Comput.-Aided Mol. Des. **1997**, *11* (6), 525-537.

(6) Hoffman, R. L. et al. Discovery of ketone-based covalent inhibitors of coronavirus 3CL proteases for the potential therapeutic treatment of COVID-19. *J. Med. Chem.* **2020**, *63* (21), 12725-12747.

(7) Burley, S. K. How to help the free market fight coronavirus. *Nature* **2020**, *580* (7802), 167.

(8) Martinez-Lumbreras, S. et al. Structural and functional Insights into Bacillus subtilis sigma factor inhibitor, CsfB. *Structure* **2018**, *26* (4), 640-648.e5.

(9) Isaacson, R. L. Olga Kennard as an Icon of Interdisciplinarity. *Crystal Growth & Design* **2024**, *24* (8), 3095-3099.

(10) Wang Erickson, A. F. et al. A novel RNA polymerase-binding protein that interacts with a sigma-factor docking site. Molecular Microbiology **2017**, *105* (4), 652-662.

(11) Trauner, D. The chemist and the architect. *Angew. Chem., Int. Ed.* **2018**, *57* (16), 4177-4191.

(12) Cakmak, M. ; Mayer, P.; Trauner, D. An efficient synthesis of loline alkaloids. *Nat. Chem.* **2011**, *3* (7), 543-545.

(13) Ellerbrock, P. ; Armanino, N.; Trauner, D. Biomimetic synthesis of the calcineurin phosphatase inhibitor dibefurin. *Angew. Chem., Int. Ed.* **2014**, *53* (49), 13414-13418.

(14) Barua, S. R. et al. *Polytwistane. Chem. - Eur. J.* **2014**, *20* (6), 1638-1645.

(15) Fitzgibbons, T. C. et al. Benzene-derived carbon nanothreads. *Nat. Mater.* **2015**, *14* (1), 43-47.

(16) Kennard, O. et al. The crystal and molecular structure of adenosine triphosphate. *Proc. R. Soc. London, Ser. A* **1971**, *325* (1562), 401-436.

(17) Viswamitra, M. A. et al. DNA double helical fragment at atomic resolution. *Nature* **1978**, *273* (5664), 687.

(18) Viswamitra, M. A. et al. Structure of the deoxytetranucleotide d-pApTpApT and a sequence-dependent model for poly(dA-dT). *Biopolymers* **1982**, *21* (3), 513-533.

(19) Jones, P. G.; Falvello, L.; Kennard, O. The crystal structure of L-alanyl-L-serine. *Acta Crystallogr., Sect. B* **1978,** *B34* (6), 1939-1942.

(20) Falvello, L. et al. Dicyclohexylammonium arsonomethylphosphonate. *Acta Crystallogr., Sect. B* **1977**, B33 (10), 3207.

(21) Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2008**, *64* (1), 112-122.
(22) Jones, P. G. et al. Cannabidiol. *Acta Crystallogr., Sect. B* **1977**, *B33* (10), 3211.
(23) Fox, R. O. Frederic Richards (1925–2009). *Nature* **2009**, *457* (7232), 976.
(24) Lide, D. R., Jr. Critical data for critical needs. Science **1981**, *212* (4501), 1343.
(25) Allen, F. H.; Kennard, O.; Taylor, R. Systematic analysis of structural data as a research technique in organic chemistry. *Acc. Chem. Res.* **1983**, *16* (5), 146.
(26) Hall, S. R.; Allen, F. H.; Brown, I. D. The crystallographic information file (CIF): a new standard archive file for crystallography. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1991**, 47 (6), 655-685.
(27) Mersh, J. D.; Sanders, J. K. M. Preferred conformations of enol ethers in solution: detection of an equilibrium by kinetic nuclear Overhauser effects. *Tetrahedron Lett.* **1981**, *22* (40), 4029.
(28) Brock, C. P.; Minton, R. P. Systematic effects of crystal-packing forces: biphenyl fragments with hydrogen atoms in all four ortho positions. *J. Am. Chem. Soc.* **1989**, *111* (13), 4586-4593.
(29) Cruz-Cabeza, A. J.; Liebeschuetz, J. W.; Allen, F. H. Systematic conformational bias in small-molecule crystal structures is rare and explicable. *CrystEngComm* **2012**, *14* (20), 6797-6811.
(30) Wallach, O. Terpenes and ethereal oils. *Justus Liebigs Ann. Chem.* **1895,** *286*, 90-118.
(31) Brock, C. P.; Schweizer, W. B.; Dunitz, J. D. On the validity of Wallach's rule: on the density and stability of racemic crystals compared with their chiral counterparts. *J. Am. Chem. Soc.* **1991**, *113* (26), 9811-9820.
(32) Brock, C. P.; Stoilov, I.; Watt, D. S. A steroid derivative that crystallizes with three molecules in the asymmetric unit. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* **1994**, *C50* (3), 434-438.
(33) Brock, C. P.; Duncan, L. L. Anomalous Space-Group Frequencies for Monoalcohols CnHmOH. *Chem. Mater.* **1994**, *6* (8), 1307-1312.
(34) Brock, C. P. Approximate symmetry in low-symmetry space groups: P2 and C2. *Cryst. Growth Des.* **2024**, *24* (15), 6211-6217.
(35) Brock, C. P. High-*Z*′ structures of organic molecules: their diversity and organizing principles. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016,** *72* (6), 807-821.
(36) Brock, C. P. Pervasive approximate periodic symmetry in organic P1 structures. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2022**, *78* (4), 576-588.
(37) Pratt Brock, C. Prevalence and Significance of Approximate Symmetry in Organic Pc Structures. *Helv. Chim. Acta* **2023**, *106* (2), e202200170.
(38) Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-589.
(39) Terwilliger, T. C. et al. AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nat. Methods* **2024**, *21* (1), 110-116.
(40) Braga, D.; Grepioni, F.; Maini, L.; Polito, M. Crystal polymorphism and multiple crystal forms. *Struct. Bonding (Berlin, Ger.)* **2009**, *132*, 25-50.
(41) Bauer, J. et al. Ritonavir: an extraordinary example of conformational polymorphism. *Pharm. Res.* **2001**, *18* (6), 859-866.
(42) Feeder, N. et al. The integration of solid-form informatics into solid-form selection. *J. Pharm. Pharmacol.* **2015**, *67* (6), 857-868.
(43) Abramov, Y. A. Current computational approaches to support pharmaceutical solid form selection. *Org. Process Res. Dev.* **2013**, *17* (3), 472-485.
(44) Bhardwaj, R. M. et al. A prolific solvate former, galunisertib, under the pressure of crystal structure prediction, produces ten diverse polymorphs. *J. Am. Chem. Soc.* **2019**, *141* (35), 13887-13897.
(45) Sheikh, A. Y. et al. Implications of the conformationally flexible, macrocyclic structure of the first-generation, direct-acting anti-viral Paritaprevir on its solid form complexity and chameleonic behavior. *J. Am. Chem. Soc.* **2021**, *143* (42), 17479-17491.

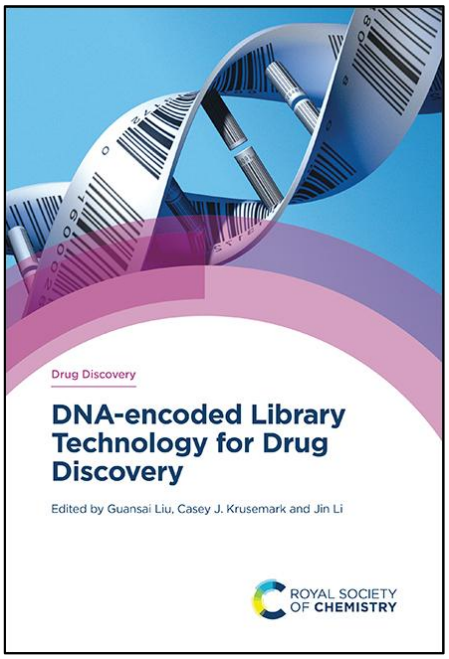
(46) Miglani Bhardwaj, R. Origins and implications of extraordinarily soft crystals in fixed-dose combination hepatitis C regimen. *Cryst. Growth Des.* **2022**, *22* (7), 4250-4259.
(47) Wilkinson, M. D. et al. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018.
(48) Lin, D. The TRUST principles for digital repositories. *Sci. Data* **2020**, *7* (1), 144.

Cheminformatics and Chemical Information Books

Contribution from Helen Cooke, CICAG Newsletter Editor, email: helen.cooke100@gmail.com

Descriptions are as provided by the publishers and are not necessarily the view of the contributor or CICAG.

DNA-encoded Library Technology for Drug Discovery

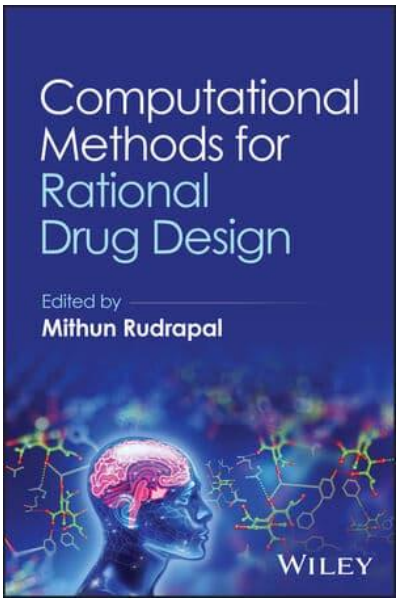


DNA-encoded libraries have numerous advantages over traditional screening methods, including easy identification of compounds and the large quantity of compounds that can be screened simultaneously.

This book provides a comprehensive guide to the implementation of DNA-encoded library technology (DELT) in drug discovery for screening and hit validation, illustrated with successful case studies.

A valuable resource for researchers in drug discovery, this book is complete with successful case studies to illustrate the best practice in implementation and operation of DELT.

Edited by Guansai Liu; Casey J Krusemark; Jin Li. Publication date: 21 February 2025. Royal Society of Chemistry.
DOI: https://doi.org/10.1039/9781788016032



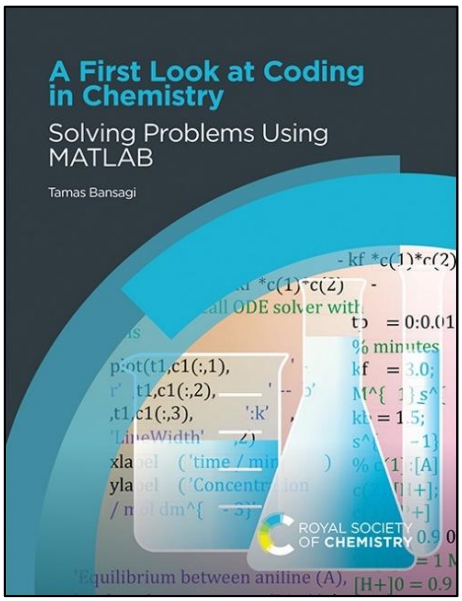
Computational Methods for Rational Drug Design

Comprehensive resource covering computational tools and techniques for the development of cost-effective drugs to combat diseases, with specific disease examples.

Computational Methods for Rational Drug Design covers the tools and techniques of drug design with applications to the discovery of small molecule-based therapeutics, detailing methodologies and practical applications and addressing the challenges of techniques like AI/ML and drug design for unknown receptor structures. Divided into 23 chapters, the contributors address various cutting-edge areas of therapeutic importance such as neurodegenerative disorders, cancer, multi-drug resistant bacterial infections, inflammatory diseases, and viral infections.

Edited by Mithun Rudrapal. Publication date November 2024. Wiley. ISBN: 978-1-394-24916-9

A First Look at Coding in Chemistry: Solving Problems Using MATLAB



Do you have a project that requires some coding skills and don't know where to start? Are you fitting a curve to your experimental data and need tools to help? Trying coding to solve maths problems in chemistry for the first time? This book will uncover all the help you need to apply these skills.

Focusing on developing basic coding skills in chemistry, this concise introductory text takes a problem-based approach and is organised in a workbook style for helping those new to programming. Discussions on coding are viewed from a chemistry perspective and embedded in solving problems familiar to most first-year undergraduate chemistry students. While primarily building programming and broadly-applicable related skills, some topics in data analysis and presentation, uncertainties in measurements, and areas of mathematics relevant to these and chemistry are also briefly surveyed.

By Tamas Bansagi. Publication date 25 July 2025. Royal Society of Chemistry.
DOI: https://doi.org/10.1039/9781837677351

Careers and Skills for Data-driven Research Network+ News

Contributions from Dr Samantha Pearman-Kanza, Professor James Baker, Professor Simon Coles, Professor Simon Hettrick, and Ms Isobel Stark, University of Southampton, email: casdar@soton.ac.uk

CaSDaR details

http://www.casdar.ac.uk

@CaSDaRNetwork

http://linkedin.com/company/casdarnetwork

@casdarnetwork.bsky.social

@CaSDaRNetwork@mstdn.social

https://www.youtube.com/@CaSDaRnetwork

Mailing List: https://www.jiscmail.ac.uk/CASDAR

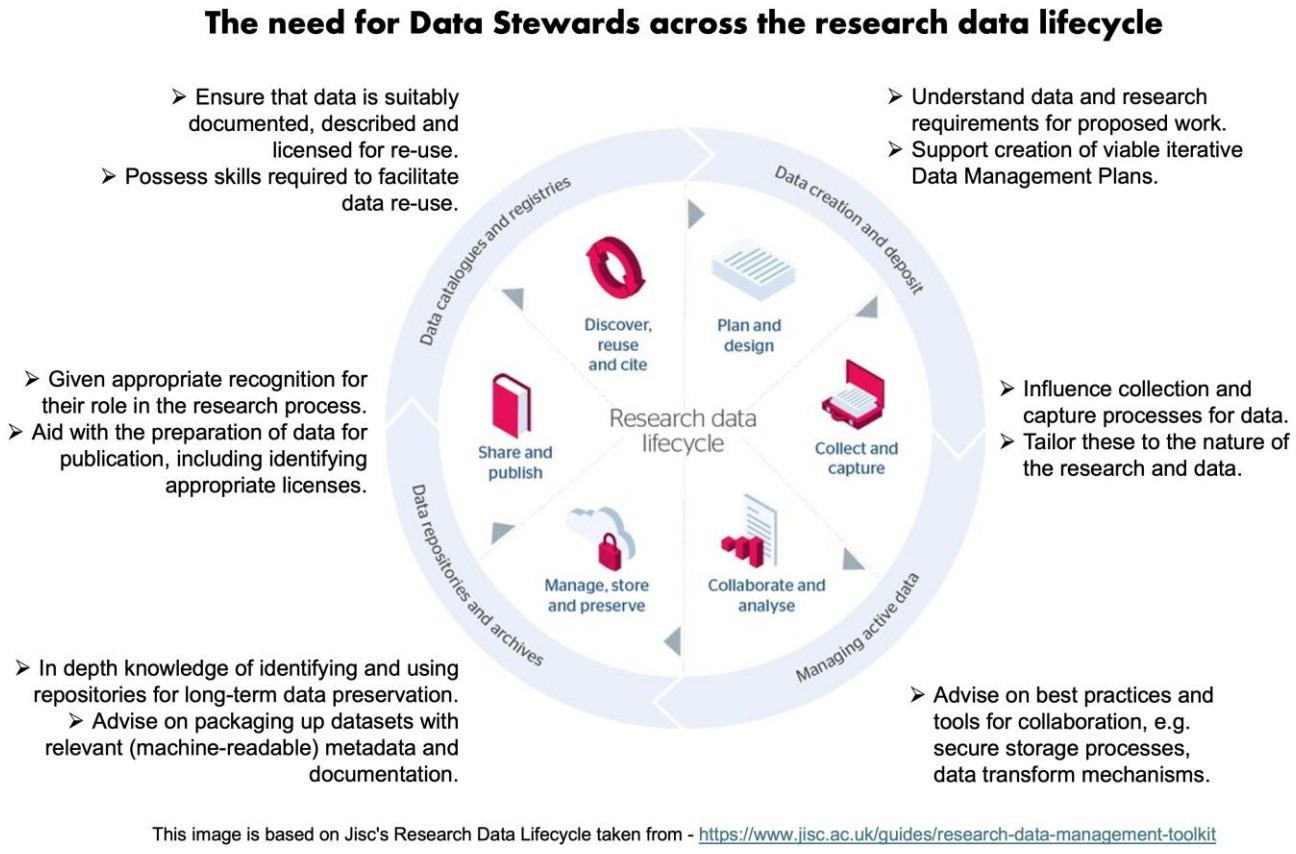
We are absolutely delighted to introduce CaSDaR. The brand-new Careers and Skills for Data-driven Research Network+. This Network+ was funded in response to the UKRI Research Technical Professional Call, and our mission is to build a supportive, diverse, inclusive and self-sustaining Network of Data Stewards.

The Careers and Skills for Data-driven Research (CaSDaR) Network+ exists to transform how research data is managed, valued, and reused across all disciplines. Our vision is a thriving, inclusive community of Data Stewards, empowered with clear career paths and recognised for their crucial role in driving high-quality research. CaSDaR aims to bridge the gap between data creation and reuse, championing best practices and sustainable approaches to unlock the full potential of research data. Through community-driven initiatives, training, and collaboration, CaSDaR is dedicated to shaping a future where data stewardship accelerates discovery, fosters innovation, and benefits society as a whole.

What are data stewards and why are they important?

The amount of data generated by research, is growing at an exponential rate. And yet, so much of this data is unusable, due to the lack of expertise, tools, and resources for effective data management. Data Stewards are the key to bridging the gap between data generation and reuse, as they have a fundamental role that ensures the quality, accuracy, accessibility and longevity of data across the entire data lifecycle. We place great value in data, but the current investment in the time and resources to drive forward data excellence is sorely lacking, and best practice like FAIR (Findable, Accessible, Interoperable, Reusable) cannot be implemented without investing in data stewards.

This demonstrates the sheer level of skills and expertise that are required by this role, which we cannot possibly expect researchers to take upon themselves in addition to their own work. Ultimately we need people power in the form of data stewards to work hand in hand with the researchers to ensure that we can produce data that can actually be re-used and built upon to further research.



This image is based on Jisc's Research Data Lifecycle taken from - https://www.jisc.ac.uk/guides/research-data-management-toolkit

The CaSDaR team

Our Network+ will be led out of the University of Southampton, and is running from 1 April 2025 until 31 March 2029.



Dr Samantha Pearman-Kanza
Principal Investigator
Senior Enterprise Fellow
School of Chemistry & Chemical Engineering
University of Southampton



Professor James Baker
Co-Investigator
Professor & Director of Digital Humanities
School of Humanities
University of Southampton



Professor Simon Coles
Co-Investigator
Professor of Structural Chemistry
School of Chemistry & Chemical Engineering
University of Southampton



Professor Simon Hettrick
Co-Investigator
Professorial Enterprise Fellow
School of Electronics and Computer Science
University of Southampton



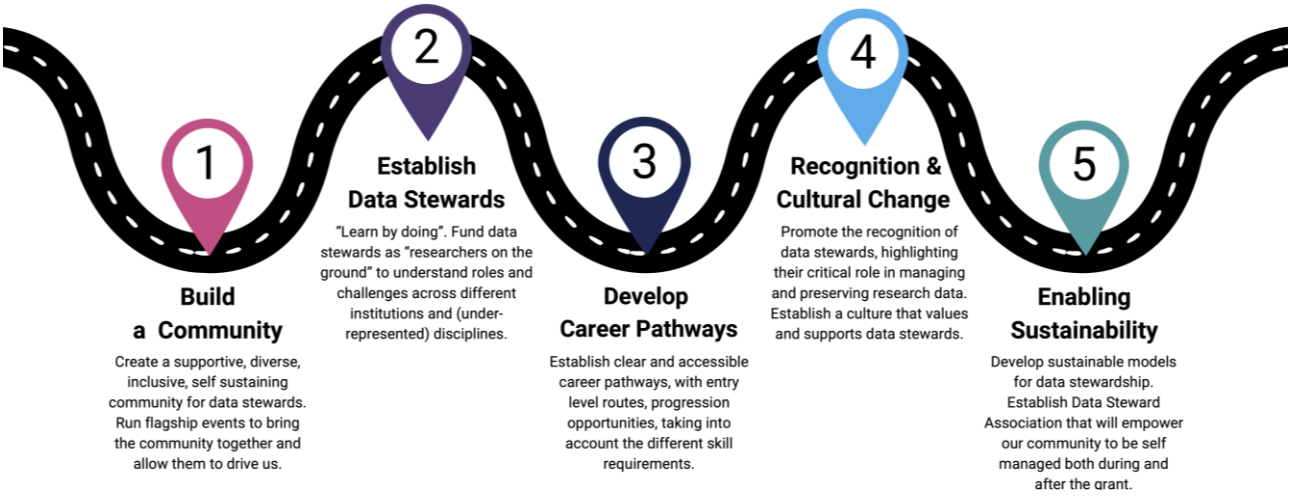
Ms Isobel Stark
Co-Investigator
Head of Research Data Management
University Library
University of Southampton

Advisory board

We have an expert, diverse, multidisciplinary advisory board: Professor Christian Bokhove (University of Southampton & UKRN), Dr Sarah Callaghan (University of Oxford), Tao Chang (Arts & Humanities Research Council), Professor Jeremy Frey (University of Southampton & AIChemy Hub), Professor Tony Hey (STFC), Dr Guy Jones (Springer Nature), Dr Emma Karoune (Alan Turing Institute), Dr William Nixon (Research Libraries UK), Professor Andy Stanford-Clark (IBM UK), Dr Marta Teperek (University of Cambridge & Open Science NL), Dr Becca Wilson (University of Liverpool).

Project partners
CaSDaR has the following project partners:

CaSDaRs roadmap & aims



CaSDaR deliverables
CaSDaR has six key deliverables:

1. **Institutional Framework:** Models for roles, responsibilities, competencies, and career progression.
2. **Resource Repository:** Online database of job descriptions, career resources, and best practices.
3. **Skills Assessment:** Report on UK data stewardship skills gaps to inform and enable tailored training.
4. **CaSDaR Analysis Report:** Impact analysis of CaSDaR activities on data stewardship and the community.
5. **Grant Data Stewardship Awards:** Support Data Steward positions across different domains & UK research institutions.
6. **Data Stewardship Association:** Formal association with mission, governance, and membership criteria.

CaSDaR activities (events)
We are planning a number of different types of events over the next four years:

- **Townhall launch:** Launch event with interdisciplinary experts, to seed initial ideas, plans and to help us shape our roadmap, first funding call and workshop plans.
- **Conference series:** Large scale Data Steward Conferences to facilitate knowledge sharing and networking on an international scale.
- **Community workshops:** Bi-annual workshops to bring the community together for discussion, networking and knowledge sharing.
- **Training workshops:** Targeted training workshops to address specific data stewardship needs and skill development.

- **Hackathons and datathons:** Interactive activities to foster collaboration and the ability to work on real-world problems that require data stewardship.

CaSDaR activities (funding)
We will be running a number of open funding calls throughout the duration of the Network. These will support data steward positions across different UK Research Institutions and domains, both at a fellow and internship level. This is part of our "learn by doing" approach to understand the common and differing requirements of data stewards across different domains/institutions.

CaSDaR values and philosophy
At CaSDaR our core values encompass inclusivity, collaboration, integrity, sustainability and recognition. We believe that diversity and inclusion are fundamental to creating a robust and effective data stewardship community. By fostering a diverse range of perspectives, we can better understand the complex needs of Data Stewards and ensure that our data accurately reflects the world it represents. We are committed to embedding Equality, Diversity and Inclusion principles into all aspects of CaSDaR, from membership to funding opportunities. This includes creating inclusive spaces, providing accessible resources, and offering flexible support and funding options to accommodate diverse needs and circumstances, such as childcare costs for events, budget allocated to support neurodiverse people to complete administrative tasks, and allowable extensions to funding deadlines for extenuating circumstances.

Through our Network, we want to engender a sense of belonging for anyone involved in working with research data, in any discipline, at any stage of their career, with extra support for early career stage researchers to ensure they can progress on their optimum data-driven journey. We will actively promote teamwork and knowledge sharing amongst the data stewardship community to foster collaboration and contribution and strive to ensure that our research data is accurate, of high quality and managed in an ethical and responsible manner. Our intent is to create a Network and develop practices that will far outlive the remit of this grant to ensure the long-term preservation and accessibility of research data. Finally, we firmly believe that all types of contribution to research data should be recognised and accredited with the value they deserve.

Advancing Open Hardware for Autonomous Discovery

Contribution from Anna Rulka, Executive Editor, Digital Discovery,
Royal Society of Chemistry, email: rulkaa@rsc.org, DigitalDiscovery-rsc@rsc.org

As science undergoes a digital transformation, the need for collaborative, transparent, and adaptable research tools is more pressing than ever. Open hardware sits at the heart of this shift, empowering researchers to design, modify, share, and build physical devices using openly licensed specifications. By making key design files, such as schematic diagrams, CAD drawings, and logic designs, freely accessible and legally reusable, open hardware removes barriers to innovation, fosters creativity, and encourages continuous improvement through global collaboration.

This approach is especially critical for the development of autonomous laboratories, where progress depends on seamless integration, advanced instrumentation, and often, bespoke hardware. Yet, traditional chemistry and materials science journals rarely prioritise hardware-focused research, risking the loss of critical innovations in broader scientific narratives.

Digital Discovery aims to fill this gap by providing a peer-reviewed platform specifically for discovery-enabling instrumentation that advances automation and autonomy in chemical, materials, and biotechnology research. We actively encourage submissions related to open-source hardware, provided they meet rigorous criteria around relevance, usability, safety, reproducibility, and innovation, as outlined in our Editorial published earlier this year.[1]

With this commitment, Digital Discovery joins the ranks of journals such as Journal of Open Hardware, HardwareX, and Review of Scientific Instruments, while offering a distinct focus on accelerating digitally driven research in the molecular sciences.

As this exciting field continues to evolve, we invite CICAG members and the broader scientific community to consider how open hardware can enhance their work, whether by contributing new designs, building on existing tools, or engaging with the growing body of open hardware literature.

Explore standout examples of open hardware in action through the following freely available articles:

(1) Hein, J. E.; Shrier, J. Guidelines for hardware-focused articles. *Digital Discovery*. **2024**, *3*, 447-448. https://doi.org/10.1039/D4DD90009J
(2) Pence, M. A. et al. An automated electrochemistry platform for studying pH-dependent molecular electrocatalysis. *Digital Discovery*. **2024**, *3*, 1812-1821. https://doi.org/10.1039/D4DD00186A
(3) Yoshikawa, N. et al. Digital pipette: open hardware for liquid transfer in self-driving laboratories. *Digital Discovery*. **2023**, *2*, 1745-1751. https://doi.org/10.1039/D3DD00115F
(4) Jones, R. J. R. Accelerated screening of gas diffusion electrodes for carbon dioxide reduction. *Digital Discovery.* **2024**, *3*, 1144-1149. https://doi.org/10.1039/D4DD00061G

Martin Karplus: a Pioneering Theoretical Chemist who Brought Molecular Dynamics Simulations to Biology

Contribution from Professor Adrian Mulholland, Centre for Computational Chemistry, School of Chemistry, University of Bristol, email: Adrian.Mulholland@bristol.ac.uk



Professor Martin Karplus.
Image source: Wikipedia, CC BY 2.0.

Martin Karplus was one of those rare scientists who makes notable contributions to several different fields. A theoretical chemist, he developed methods for interpreting nuclear magnetic resonance (NMR) results and predicting molecular structure (the Karplus equation is named after him); pioneered detailed simulations of reaction dynamics, showing how chemical reactions happen at the atomic level; and changed understanding of biology through simulations of biomolecules, revealing their mechanisms and showing dynamics to be crucial to protein function.

He was born in Vienna before the Second World War and came as a refugee to the United States in 1938 with his family, fleeing the Nazis. He said that this experience shaped his view of the world and also influenced his approach to science, giving him the courage to change research areas, which he did

several times during his career. As a boy, he developed a passion for nature, particularly birdwatching, and as a 17 year old published a (single author) paper on bird nesting in the Arctic, which is still being cited today. This fascination with biology was to stay with him all his life. He intended to become a doctor, starting his studies at Harvard, but from his ornithology work, decided that he wanted to do research into biology at a fundamental level, so switched to study chemistry and physics. Following advice from Oppenheimer, he moved to Caltech for graduate studies. Always a man of wide interests, there he organised showings of classic films, and visited Hollywood, where a chance meeting with Charlie Chaplin was a highlight. He worked for his PhD with Linus Pauling, developing practical approaches for quantum mechanical calculations on hydrogen bonds.

Following his PhD, he worked in Oxford, England, on a National Science Foundation (NSF) Fellowship. He took advantage of Oxford's short terms to travel widely in Europe. His photographs from these travels were later made into an exhibition and book, and provide striking images of post-war Europe. He developed a love of France in particular, of the country and of its cuisine (already a fine cook, he later trained as a chef in Paris). In this time, he was thinking of how quantum mechanics could be used to interpret chemical experiments, NMR in particular. He moved to the University of Illinois at Urbana-Champaign (UIUC), where NMR experiments were being pioneered. There, he developed a theoretical relationship relating NMR coupling constants to dihedral angles. This relationship is now called the Karplus equation, is included in standard organic chemistry textbooks and is still used in analyses of molecular structure.

In 1960, he joined the IBM Watson Scientific Laboratory at Columbia University, New York City. This gave him use of an early supercomputer (featuring the then-advanced technology of punch-card input, which made scientific programming much easier). This provided the computational power for his simulations of the $H + H_2$ reaction. He saw that this would allow calculation of enough trajectories to achieve statistically meaningful results, having first developed an approximate but realistic potential energy function to represent the reaction. The results showed the reaction in atomic detail, and provided predictions of its temperature dependence, allowing tests of theories of chemical reactivity.

After another five years, again following his desire to explore new areas, he returned to Harvard as a Professor in 1966. There, he continued his theoretical chemistry research, but returned to his first love, biology, beginning with quantum mechanical prediction (later proved correct) of the structure of retinal, which is involved in vision. Then, with Arieh Warshel, he developed multiscale methods that combined quantum mechanics (QM) with simpler, empirical 'molecular mechanics' (MM) to calculate spectra of retinal and similar conjugated molecules. They used this QM/MM method to investigate how light is absorbed by retinal, driving a change in its shape, as the first event in light sensing and vision. QM/MM methods allow quantum chemical calculations in large systems, which would otherwise be intractable or very time consuming. They have proved to be a practical way to simulate various properties and processes, particularly chemical reactions, in large molecules such as enzymes. It was for the development of multiscale methods that Karplus and Warshel, with Michael Levitt, were awarded the 2013 Nobel Prize in Chemistry.

Martin also developed important theories to explain cooperativity of oxygen binding in haemoglobin, and protein folding. Perhaps the work he is most widely known for is on molecular simulations of protein dynamics. He saw that atomically detailed simulations of molecular dynamics, as he had studied for the simple $H + H_2$ reaction, could reveal how proteins moved, realising the dream of Feynman of explaining and understanding biology from the jiggling and wiggling of atoms. He published the first molecular dynamics simulation of a protein in 1977; while only a few picoseconds long, and not including water, it showed how the prevailing conceptual picture of proteins (e.g. from crystal structures) as static was wrong. It heralded a fundamental change in understanding, leading to the now widely accepted view that dynamics of proteins and other biomolecules are inherent to their biological functions. Molecular dynamics (MD) simulations are now very

widely used in chemical and biochemical research, complementing and as part of experiments on biological macromolecules. They can be used to calculate physically measurable quantities, to identify molecular movements and conformational changes, and to analyse their interactions in exquisite detail. MD simulations are practically useful, and now routinely applied, in the design of drugs and vaccines and in protein design and engineering. Martin continued to develop simulation methods, and to apply them, throughout his career.

Martin worked with more than 250 PhD students and postdocs, and there is not space here to note all of their contributions to his work. His eloquent and personal scientific autobiography[1] provides more details of his life, work and family.

Martin died at home in Cambridge, Massachusetts, on 28 December 2024, aged 94.

(1) Karplus, M. Spinach on the Ceiling: A Theoretical Chemist's Return to Biology. *Annual Review of Biophysics* **2006**, *35*, 1-47. https://doi.org/10.1146/annurev.biophys.33.110502.133350

- -

RSC Chemistry Databases Update

Contribution from Tamara Hughes, email: hughest@rsc.org

RSC Chemistry Databases – announcing the 2025 ChemSpider webinars

In 2023, ChemSpider and the RSC hosted a three-part webinar pilot focused on helping our community to embrace digital data. Close to 700 attendees joined us and our panel members to dive into topics such as "Data explosion in chemistry: what are we going to do with all the data, and what will it do to us?" (Lynn Kamerlin, Georgia Tech), "How to initiate the cultural change towards digital chemistry" (Sonja Herres-Pawlis, RWTH Aachen), "Wanted – standard notation for reusable chemistry data" (Leah McEwen, Cornell University) and "How can we combat heterogeneous, unfair and disparate data in digital chemistry?" (Samantha Pearman-Kanza, University of Southampton).

Recordings of all our 2023 webinars are available on youtube.

Following this successful mini-series, we are excited to announce that ChemSpider and the RSC will be hosting further webinars towards the end of 2025. This year we are partnering with panel members interested in exploring "Chemistry data and AI", "Data standards in chemistry", and "Data management". We are hoping to showcase current and planned initiatives to develop standards and tools, research infrastructures, and developing cultures to support FAIR chemistry data preparation, publication, and reuse.

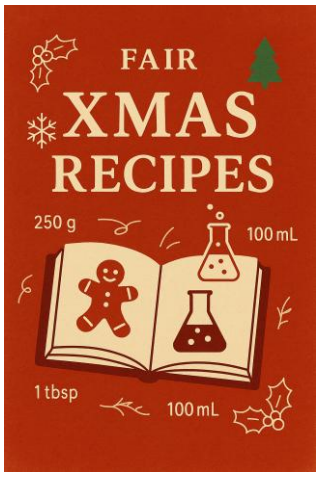
More information about our speakers and webinar dates will be available on our blog soon, so make sure to keep an eye out for further details!

Please also feel free to reach out to us via email if you have any questions or should you be interested in becoming a panel member or sponsoring our events.

- -

A CICAG FAIR Cookery Recipe Book for Xmas

Contribution from Jeremy Frey, School of Chemistry and Chemical Engineering, University of Southampton, email: j.g.frey@soton.ac.uk



Following on from the very successful IUPAC (WorldFAIR) FAIR Chemistry Cookbook we are hoping to launch a festive CICAG Cookery Book in time for Xmas, which will of course meet the FAIR data standards and, rich semantic metadata and well specified units!

Contributions will be very welcome – we will post a template on the CICAG website and circulate this through CICAG's social media: LinkedIn (rsc-cicag page), Bluesky @rsc-cicag.bsky.social and Mastodon @rsccicag). Or send your ideas to Jeremy Frey.

The recipes don't have to be seasonal and if there are any interesting chemical comments to be made about them we can include those too.

- -

Physical Sciences Data Infrastructure (PSDI) News

Contribution from Dr Samantha Pearman-Kanza, Dr Nicola Knight, Professor Jeremy Frey and Professor Simon Coles, University of Southampton, email: psdi@soton.ac.uk

We have been very busy at PSDI over the last few months, and the biggest piece of news is that we have launched the first version of our services. Read all about our wide range of offerings, in addition to all our latest activities. If you want to be kept up to date more regularly, sign up to our mailing list or follow us on social media.

PSDI details

www.psdi.ac.uk

@PSDI_UK

linkedin.com/company/psdiuk

@psdi-uk.bsky.social

@PSDI@mstdn.science

Mailing List: https://www.jiscmail.ac.uk/PSDI



PSDI launch

We are delighted to announce the release of the first wave of PSDI resources, which have been designed to help physical scientists with their research. Please browse our website and explore what we have to offer. Some resources require login (currently available to UK academics), but most resources are available to all.

PSDI showcase

If you want to hear first-hand about our first release, then save the date for our hybrid showcase on 17 September 2025.

Data sources

PSDI provides access to over 20 different databases and repositories of physical sciences data. You can perform a search of these data sources using the PSDI Cross Data Search Service. There are also a number of additional datasets available via PSDI that can be accessed individually, including data on physical chemical properties, critical micelle concentration and biomolecular simulations.

Data services & tools

PSDI provides a number of software services and tools that enable you to access, transform, and process your data:

PSDI Data Services	
	PSDI Data Conversion Service – Your one-stop shop for file format conversion, PSDI's Data Conversion Service enables users to seamlessly convert between different file formats and assess the quality of proposed conversions. We provide a web-based service, and some additional tools (a python library, a command line application, and a python/flask application) to suit a wide range of user needs.
DataRevival	PSDI Data Revival – Through PSDI we offer a free version of the Data Revival Service. This is an AI-driven service that can seamlessly convert your handwritten lab book pages into machine-readable data, unlocking the potential of your chemistry data. Operating with a high degree of accuracy, Data Revival creates relational links across your chemical data and enables users to run semantic searches across potentially millions of data points, including complex chemical structures.
CatalysisHub CDI	Catalysis Data Infrastructure (CDI) Thematic Portal and Bibliographic Data Web Service – The Catalysis Data Infrastructure (CDI) is an example of a thematic portal (which offers summary reports on publications and impact within a specific scientific domain), illustrating how such portals can organise and present catalysis research. Additionally, the CDI Bibliographic Data Web Service enables programmatic access to this publication data via an API.
CCP-NC Collaborative Computational Project for NMR Crystallography	Collaborative Computational Project for NMR Crystallography (CCP-NC) 2.0 MagresView and MAGRES Database Search – MagresView is a visualisation that serves as an online interface for CCP-NC and facilitates crystal structure and NMR tensor visualisation from .magres files. The database search service enables a user to search the MAGRES database by a number of different parameters.
Galaxy	Materials Galaxy – This service (built off the open-source scientific workflow management system Galaxy), provides access to tools for tackling computational challenges in muon spectroscopy and X-ray absorption spectroscopy.

PSDI Data Tools			
abcd, Data Management for Machine Learning Interatomic Potentials 	CCP-NC Magres View 2.0 Library 	janus-core, Machine Learning Interatomic Potentials Tool 	Soprano NMR Crystallography Python Library 
aiida-mlip, Workflow Management for Machine Learning Interatomic Potentials 	Demo BioSim Database With Local Web Interface 	PSDI Data Conversion Command Line Application 	Catalysis Data Infrastructure (CDI) Thematic Portal Code 
Amber Plugin for AiiDA 	DSync (Data Transfer & Synchronisation) Container 	PSDI Data Conversion Python Library 	DL_MONTE Container Image 
	GROMACS Plugin for AiiDA 	PSDI Data Conversion Python/Flask Application 	

Guidance, case studies & training

Through our extensive Knowledgebase, PSDI offers guidance and training materials for physical scientists and those who support them, including guidance on working with scientific research data and using the technologies and resources provided by PSDI. Self-paced training supplements in-person learning opportunities to broaden researchers' abilities beyond their specialised fields, focusing on areas like programming, machine-learning, research data management, and other transferable skills.

Guidance

Our knowledgebase provides a range of Guidance for physical scientists on a range of topics related to PSDI, including data management, data sharing, electronic notebooks, and data science, and hosts the documentation for the technologies and services provided by PSDI.

Training

PSDI provides training for physical scientists to learn a variety of skills and technologies through several different mechanisms, including our interactive learning platform that hosts the Skills4Scientists courses, a number of tutorials, and a range of in-person training workshops and summer schools.

Community efforts

PSDI is not just about the technology, we are also dedicated to bringing together and providing support to communities, desperately needed in the physical sciences, through several different activities.

Events

PSDI run a rich range of events, including webinars, workshops, community townhalls and training events. All of our upcoming and past events can be found here.



PSDI Team at the Research Data Forum in Edinburgh (organised with the DCC): Agnes Jasinska, Cerys Willoughby, Nicola Knight, Samantha Pearman-Kanza, Linda Freimane, Ian Bruno.

Data curation & management

We have run four community data workshops across 2024/2025 with the Digital Curation Centre, to explore the challenges and opportunities surrounding research data in the physical sciences, and to establish what support, tools and community activities researchers need to improve their data. Most recently this included a Research Data Forum at the University of Edinburgh, and a workshop on Creating communities around best practices and common challenges in data at the International Data Curation Conference. In addition to this we are running several roadshows about data curation and management across different areas of the UK, starting with Cambridge and Southampton.

Electronic research notebooks

PSDI are working to re-establish the Electronic Research Notebook community defined by this mailing list: RESEARCHNOTEBOOKS@jiscmail.ac.uk, which is a community for anyone, and everyone interested in or trying to implement Electronic Research Notebooks. We re-launched this community with a hybrid meeting at UCL on 20 May to bring together interested parties and to discuss how we can best serve the community moving forward.

ELNFinder

PSDI are also proud to contribute to community-led projects. ELNFinder assists researchers and staff from research institutions to select the most suitable Electronic Lab Notebook (ELN) software for their needs. Users can easily filter the extensive list of ELNs by key criteria such as import/export capabilities, search functionality, delivery model, data storage, and pricing options.

Collaboration

PSDI isn't just a single institution or organisation, but a group of individuals, organisations and institutions. The main partners in PSDI are the University of Southampton and the Science and Technology Facilities Council. We also have a growing number of partners. If you are interested in partnering with PSDI then please get in touch with us.

Upcoming events featuring PSDI

The PSDI team will be featuring at (and in some cases co-organising) some international conferences over the next few months.

CORDI (Conference on Research Data Infrastructure)

Dates: 26-28 August 2025

Location: RWTH Aachen University, Germany

Details: The CoRDI stands for more knowledge through better use of research data and the resulting added value for society. It boasts a number of specific disciplinary tracks to address infrastructure aspects relevant for the individual differences, and a cross-disciplinary track focusing on more specific topics regarding the delivery of Research Data Management Infrastructure. Dr Nicola Knight will be presenting at this conference on "The Physical Sciences Data Infrastructure (PSDI) in the UK: Resources to accelerate physical sciences research".

International Data Week

Dates: 13-16 October 2025

Location: Brisbane, Australia

Details: IDW takes place every two years and is the leading international conference on research data. It is jointly organised by the International Science Council's Committee on Data (CODATA) and World Data System (WDS), and the Research Data Alliance (RDA). It combines the RDA Plenary Meeting, the biannual meeting of this international member organisation working to develop and support global infrastructure facilitating data sharing and reuse, with SciDataCon, the scientific conference addressing the frontiers of data in research organised by CODATA and WDS. Several members of the PSDI team will be attending IDW; Jeremy Frey will be running a session on "Units, Symbols and Terminology for Data Driven Science: Philosophy to Pragmatism" and Ian Bruno will be running a session on "Enabling effective data curation in the long tail of the Physical Sciences".

MADICES 3 (Machine-Actionable Data Interoperability for the Chemical Sciences)

Dates: 20-25 October 2025

Location: PSI, Villigen, Switzerland

Details: This interactive workshop strives to foster interdisciplinary collaboration and innovative use of emerging technologies to realise the full potential of FAIR data principles in modern research. PSDI is taking an active role in the third iteration of this conference. Dr Samantha Pearman-Kanza now sits on the organising committee, and several PSDI colleagues will be attending to take part in this event to discuss key issues in the physical sciences and take part in the hackathon to try and identify new tools and services that can aid in these issues.

Lab Innovations

Dates: 29-20 October 2025

Location: NEC, Birmingham

Details: Lab Innovations brings together a lab expo with five conference theatres to cover a range of different themes including sustainability, robotics and automation, biotech innovation, quality, accreditation and compliance, professional development and technical product advice and demos. Dr Samantha Pearman-Kanza will be presenting at this conference.

Ontologies4Chem
Dates: 11-13 November 2025
Location: Limburg an der Lahn
Details: Jointly organised by NFDI4Chem, NFDI4Cat, the Beilstein-Institut, and PSDI, this two-day event will focus on the development and application of ontologies in the chemical sciences. It will bring together researchers, data specialists, and infrastructure providers to exchange ideas, share progress, and foster collaboration across disciplines.

Recent webinars
Since our last update, we have held three more webinars and have added to the recordings available on our YouTube channel. We are busy planning many more webinars for the rest of the year, if you are interested in presenting your work at a PSDI webinar do get in contact.

Previous webinars
Catch up on our previous webinar recordings on YouTube (with more published as recordings become available)

- Introduction to NOMAD
- KnowLedger: An Open Ecosystem for Research Data Management

Recent PSDI events
Units, Symbols, and Terminology in the Physical Sciences in and for the Digital Era
This meeting brought together several groups who are trying to combat similar problems – how to bring the standards for scientific nomenclature and terminology into the digital age. There were representatives from: PSDI – Physical Sciences Data Infrastructure, CODATA DRUM Task Group – Digital Representation of Units of Measure, IUPAC Green Book – Quantities, Units, and Symbols in Physical Chemistry and the IUPAC Gold Book – Compendium of Chemical Terminology. This meeting boasted plenary talks covering ideas from the Philosophy of Chemistry, through Metadata, to the use of the Semantic Web in Science. The presentations and YouTube video of this conference are available on our YouTube Channel.



The attendees of the Units, Symbols and Terminology in the Physical Sciences in and for the Digital Era.

Chemical and Materials Machine Learning School (CaMMLS) 2025
The CaMMLS 2025 workshop at Daresbury brought together 45 attendees over five days, featuring engaging discussions, hands-on tutorials, and valuable networking opportunities. Highlights included a lively social BBQ night, a poster presentation session with three student winners, and a successful blog contest with high-quality

entries to be published soon. Attendees praised the breadth of topics covered, with one commenting, "It was a great way to understand the options for applying machine learning to scientific research."

Recent events & presentations featuring PSDI

The PSDI team has been busy over the last couple of months attending a variety of conferences, workshops and events. We are updating our zenodo community with many of the presentations and publications made by the team.

Highlights from the last few months include:

Future Labs Live, Basel 2025

In late May, Dr Samantha Pearman-Kanza and Mr Samuel Munday attended Future Labs Live 2025. This was a really interesting meeting with a rich range of presentations and panel discussions. Samantha presented on PSDIs work on Electronic Lab Notebooks: "One does not simply implement an Electronic Lab Notebook: The epic saga of ELN implementation". At Future Lab Live 2025. Samantha also chaired a panel session on "FAIR data and publishing – why should scientific publishers and the lab community care about FAIR data". This was a lively discussion with a strong focus on the importance of metadata and the use of common vocabularies/taxonomies/ontologies to standardise our metadata, the importance of incentives for the community to rigorously review papers not just for their scientific papers, but also their supplementary data, and the responsibility of publishers to set higher standards for this. Samantha also chaired a panel session on "Addressing the gender bias to achieve ethical AI", which was a very informative and insightful discussion that highlighted the importance of diversifying our datasets, both with respect to legacy data, but also in terms of our practices moving forward to ensure that new data is created in an unbiased fashion. Large Language Models and Generative AI are the latest developments in the rapidly accelerating field of AI, and with such a fast-moving field, we must be constantly mindful that we are really considering the data we generate, lest we fall foul to self-perpetuating the bias that is already embedded in existing datasets and systems.

London Lab Live 2025



Mentimeter Word Cloud from London Lab Live Panel – Has FAIR failed us, or have we failed FAIR?

In early May, Dr Samantha Pearman-Kanza, Dr Philip Leadbitter and Dr Cerys Willoughby attended London Lab Live 2025. Samantha presented on "The Physical Sciences Data Infrastructure: What can we do for you?" to inform the community about all our offerings. Samantha also chaired and took part in the provocatively titled panel on "Has FAIR failed us, or have we failed FAIR?". This was a very interesting discussion, that kicked off with a Mentimeter fuelled debate on the barriers and challenges of FAIR, and then a deep dive into the age-old question of "Is it ever truly possible to be FAIR?". A lot of the

discussions focused on the notion of "FAIR enough" and trying to future proof outputs as much as possible and consider when revamping/updating is necessary as "What is FAIR today won't necessarily be FAIR tomorrow". The team also wanted to give a shout out to Suchitra Ramaswamy of Zifo – with her incredibly insightful presentation on "Digital Transformation: do we still know what it means?" For the most salient point of the event, tech does not equal transformation! If we are going to implement technical solutions for so called transformation, then we need to ensure that they are actually bringing added value and that alongside the tech, comes a genuine improvement in working practice.

Fellowship of the data

In early April, Dr Samantha Pearman-Kanza and Dr Nicola Knight attended the Fellowship of the Data meeting in Jena, Germany, which aims to bring together RDM enthusiastic communities! Leaning into the Lord of the Rings themed title, Samantha presented on "I can't record it for you, but I can provide you with a Digital Notebook – The daring journey of the pre-requisites and hurdles of Digital Notetaking".

Paperless Labs Europe

In March, Dr Samantha Pearman-Kanza and Mr Samuel Munday attended Paperless Labs Europe. Samantha presented on: "Not all heroes wear capes! Data Stewards: The unsung heroes of the Research Data Lifecycle" to highlight the importance of data stewardship to the lab community.

PSDI case studies and reports

Check out our latest case studies and reports:

Electronic Supplementary Information Case Study

This case study explores the challenges of losing valuable raw data when it is embedded within PDFs as part of supplementary information (ESIs). To address this, data was extracted from the supplementary information of the paper titled "Towards the Prediction of Antimicrobial Efficacy for Hydrogen Bonded, Self-Associating Amphiphiles", and converted into JSON format. The process was carefully documented to showcase how such data can be transformed into a more machine-readable format, improving accessibility and usability for further analysis.

Resurrecting Second Harmonic Generation (SHG) Data Case Study

The preservation, accessibility, and reuse of scientific data are increasingly vital in contemporary research. Historical datasets, often termed legacy data, provide significant opportunities for new discoveries, validation of past results, and insights through modern analytical methods. Despite their potential value, legacy datasets frequently suffer from poor documentation, outdated formats, and technological obsolescence, complicating their reuse. This case study addresses these challenges through the reconstruction of databases originating from Second Harmonic Generation (SHG) experiments, conducted as part of a doctoral research project completed in 2011 at the University of Southampton.

PSDI DCC Community Workshop Report: Edinburgh and Southampton 2024

The PSDI DCC Community Workshop Report, authored by Agnes Jasinska, Clara Lines Diaz, and Kevin Ashley from the Digital Curation Centre, summarises the findings and recommendations from two workshops held in Edinburgh and Southampton in 2024. These workshops aimed to engage researchers and stakeholders in the physical sciences to understand the challenges and opportunities around data sharing and reuse. Key challenges identified include the lack of knowledge and support for data sharing, insufficient documentation standards, and the high cost of data management. The report recommends the development of an integrated platform to address these issues, providing technical solutions, standards, and community-building resources to facilitate data sharing and collaboration in the physical sciences.

PSDI articles

Check out our latest articles:

How do you implement an Electronic Lab Notebook? - The Southampton ELN trial and research

This article with Dr Samantha Pearman-Kanza explores the work that PSDI has been undertaking around process recording, including the University of Southampton ELN Trial and the University of Nottingham AI4Green ELN Case Study.

An interview with Dr James Gebbie-Rayet on the creation of BioSimDB

Dr James Gebbie-Rayet and his research team are poised to revolutionise the Biomolecular Simulation Research field by addressing long-standing, systemic challenges. This interview explores the work they are undertaking in this area for PSDI.

Cambridge Structural Database (CSD) Updates

Contribution from Ana Machado, Marketing Executive at CCDC, email: hello@ccdc.cam.ac.uk

CSD data and software update (May 2025)

The latest CSD update, CSD 6.00, brings significant advancements, including a new data format that is extendable, faster to search, contains more data and is smaller to download. It includes 1,371,757 entries, with 45% organic and 55% metal-organic structures. Notably, over 165,000 CSD entries now feature fully calculated disorder models, allowing for improved visualisation and analysis of disordered structures.

Additionally, existing entries have been enriched with new searchable fields to enhance their scientific value and analytical potential.

This release also includes:

- Covalent docking enhancements in GOLD and ligand overlay in the CSD Python API.
- New powder X-ray diffraction functionality
- CLP force field added to VisualHabit in CSD-Particle
- Four new CSD Subsets: Generally Recognised as Safe (GRAS) substances, retracted entries, semiconductors and entries with associated raw data DOIs

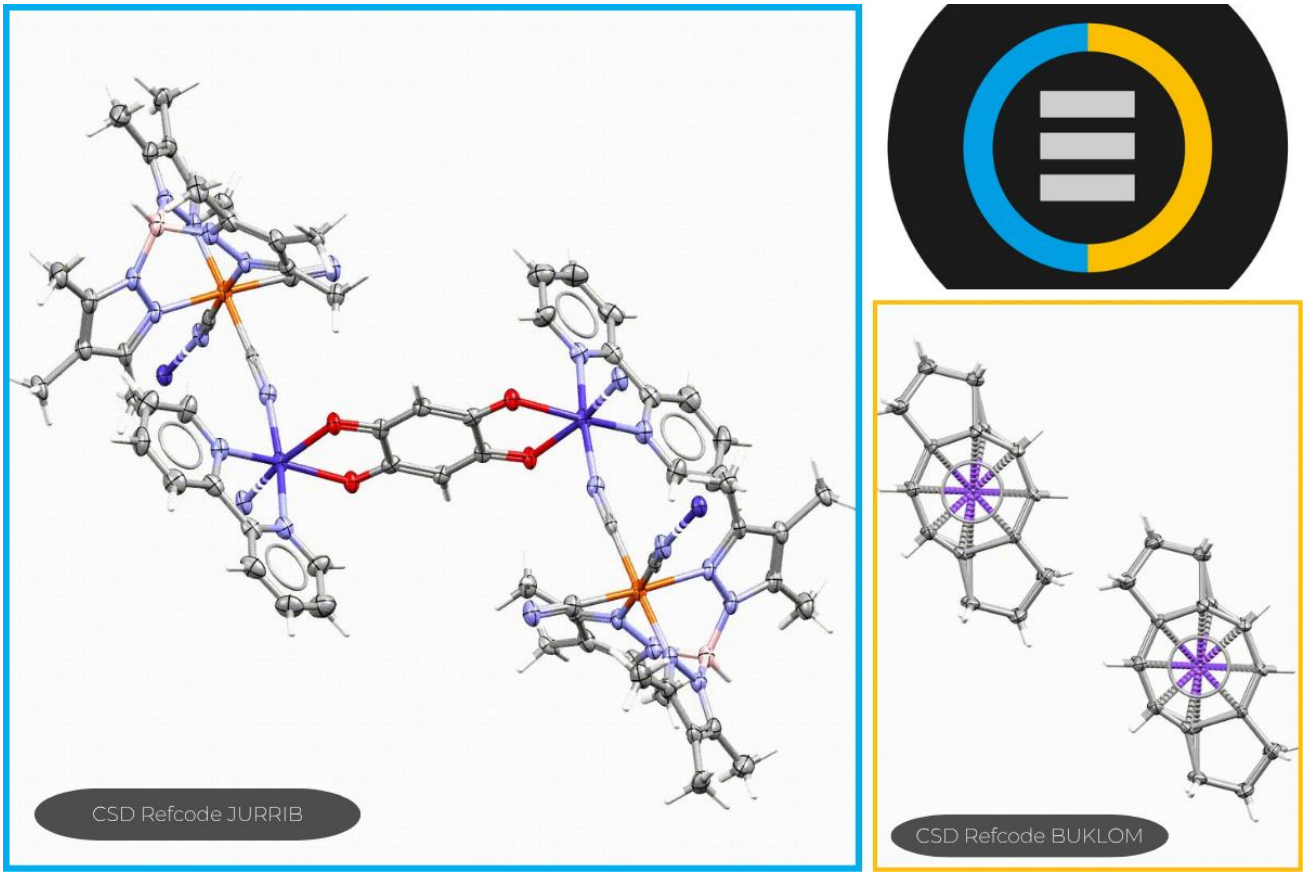
Latest blogs and videos

Enriching the CSD: new data fields for deeper structural insights

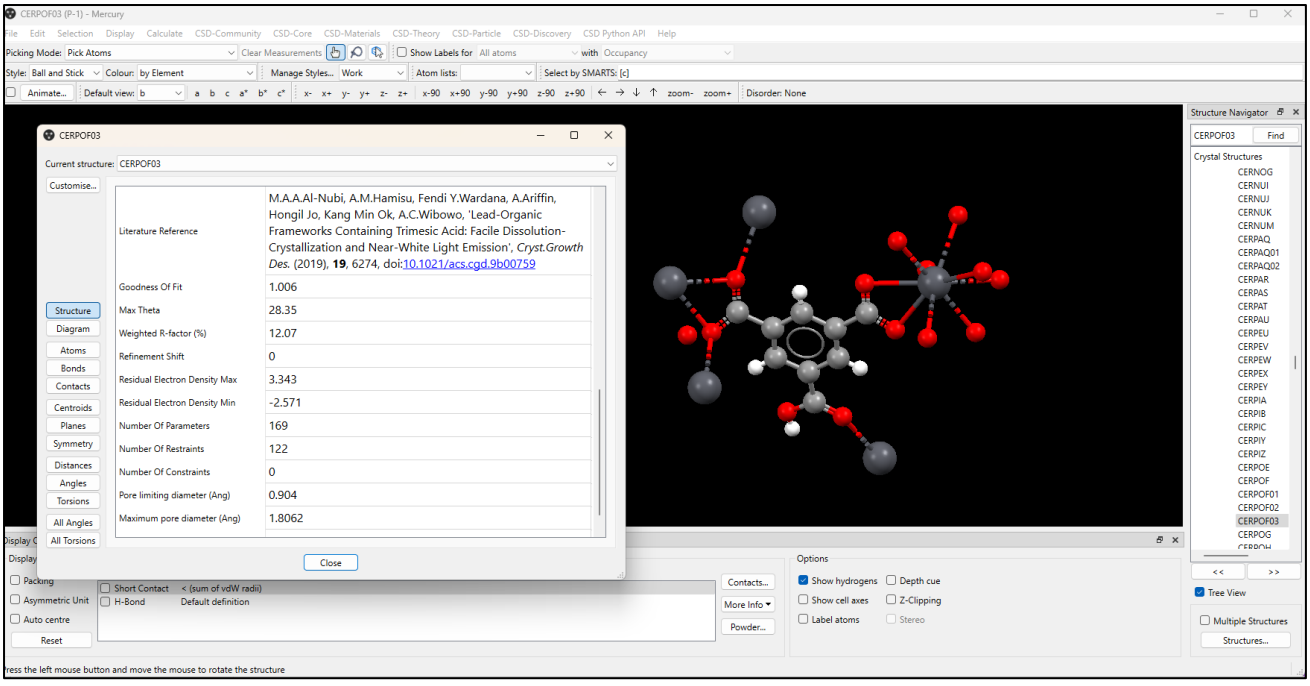
The Cambridge Structural Database (CSD) 6.00 features new searchable fields that enhance existing entries, providing greater scientific value. You can now identify relevant structures more easily, with new data integrity fields added for over one million entries from submitted CIFs.

Bringing Order to Disorder: Expanding Structural Detail in the Cambridge Structural Database

We are improving how disorder is represented in the Cambridge Structural Database (CSD). Read on for the enhancements, challenges faced, and the significance of accurately capturing disorder for structural chemistry and data-driven research.



New examples of functional metal-organic frameworks are included in this release.



New data fields are visible in the Structure Information pane in Mercury.

How to: Search Scientific Literature with the CSD

New to the CSD? About to teach students or guide undergraduate research projects? This video introduces the database, its contents, and how to search and download data to help enrich scientific studies and inspire new directions of research.

Upcoming webinar

Frontiers of Chemistry Education Panel Webinar: Preparing Students for Emerging Career Paths

Technology is evolving faster than ever before, and with it, science career paths change too. How can educators, institutions, and students best prepare for career paths that may not yet exist? Join our panel to explore key questions for chemistry students and educators, to learn how to best prepare for the future of work and careers.

Thursday 4 September 2025

- 3pm BST (London, Dublin, Edinburgh, Lisbon)
- 10am ET (Boston, New York, Ontario)
- 4pm CEST (Berlin, Brussels, Rome, Warsaw)
- 7am PT (San Francisco, Seattle, Vancouver)

Register now

Free virtual workshops

We are excited to announce that a new series of free virtual workshops is scheduled for October and November 2025. We are currently finalising the topics, dates, and times, which will be announced soon.

The free virtual workshop series is divided into three sessions. Each workshop will follow the format "Show one, Try one, Explore more!":

- **Show one:** a guided demo of the software by the CCDC tutors.
- **Try one:** a hands-on example for you to try during the session.
- **Explore more:** summary of learning outcomes, challenges, and quizzes.

All the sessions include presentations and demonstrations by CCDC expert tutors, along with a hands-on part for you to try the software, with the tutors available to answer any questions.

Please check our events page for updates as we finalise the Autumn series programme.

Update from the AIchemy Hub

Contribution from: Dr Chris Mellor, email c.mellor@imperial.ac.uk and Dr Ben Alston email: ben.alston@liverpool.ac.uk, AIchemy Hub Managers

Introduction

AIchemy aims to transform the chemistry–AI interface by conducting world-leading research and catalysing the wider adoption of AI in experimental and computational chemistry communities. As we move into the second year of the Hub, our forerunner research projects are now up and running and we're ramping up our community events and activities. Having discussed the launch of our Pump Priming Awards in the last edition of the *CICAG Distillate*, we can now give an outline of what has been funded and provide details our next funding calls. Keep reading to find out more.

Funding

Outcome of Pump Priming call

We received an incredible 42 applications for our first pump priming call, and we're thrilled to announce that ten projects have been awarded a total of £200K in funding. These innovative, early-stage ideas span nine institutions and highlight the exciting diversity of research happening where AI meets chemistry. The awarded projects are now underway, with summaries of the projects on our website, along with details of the outcomes and scoring.

New! AIchemy Frontier Fund

We've recently opened the call for our AIchemy Frontier Fund. This Fund is for projects of 1-2 years duration and up to £1.25 million. We have £2.5 million in total available to support ambitious, high-impact research at the cutting edge of AI and chemistry.

The call has a two-stage application process with the first stage requiring a short application, closing date 12 September 2025, the second stage will be a panel interview in December 2025. All details are available via the funding page of our website, with drop-in sessions ongoing throughout the summer for people to ask questions and clarify any points before submitting their application.

New! AIchemy Collaborative Travel Fund

We've also launched a new Collaborative Travel Fund, offering up to £7,500 to support collaborative activities that bridge the gap between AI and chemistry. Whether you're looking to spark a new partnership or strengthen an existing one, this fund is designed to help researchers connect, share expertise and build lasting collaborations across sectors and disciplines.

The Fund will open twice a year with Round 1 currently open for applications, Round 2 will open later in the year. All details can be found on the funding page on our website.

Events

Annual conference

The inaugural AIchemy Annual Conference took place at Imperial College on 17 March 2025. We welcomed over 150 delegates on the day, bringing together a range of speakers to discuss the challenges and opportunities for AI in chemistry across four themes: Optimisation, GenerativeAI/LLMs, Human in the Loop/Robotics and AI tools for Chemistry. We closed the day with a thought-provoking discussion about the responsible use of AI. Key takeaways from the day are available on our website.



2025 Annual Conference poster prize winners.

Workshops

Our community workshops launch this summer with a series of events designed to inspire and upskill researchers and educators at the interface of AI in Chemistry. Further details of all events can be found via the events page of our website.

- **Generative AI in Chemistry Education** — 21 July, Imperial College
- **Inorganic Informatics: From Rules to Embeddings** – 31 July, Imperial College
- **RAG Models & AI Agents in Chemical Discovery** — 19 September, University of Leeds
- **Introduction to Open Chemistry Robotics and Automation (OChRA)** — 24 September, University of Liverpool

Our workshop topics are proposed by the community, therefore if you have an idea for a workshop you'd like to run in 2026, do get in touch. Submit your idea via our expression of interest form.

Webinar series

Our webinar series has been running monthly since October 2024 and we've welcomed a wide range of high-profile speakers, whilst also giving a platform for postdocs and PhD students to present. Our April webinar provided industry insights with speakers from AstraZeneca (Nessa Carson) and the Imperial spin-out SOLVE Chemistry (Linden Schrecker). The webinars are recorded, and you can catch our previous speakers' presentations on our YouTube channel.

Confirmed speakers for the rest of the year are below. You can sign up for these webinars via our website.

- 23 July – Prof. Jason Hein, University of British Columbia
- 17 September – Prof. Rafael Gomez-Bombarelli, MIT
- 22 October – Prof. Bingqing Cheng, University of Berkeley
- 19 November – Dr Lauren Ye Seol, University College London
- 10 December – Prof. Natalie Fey, University of Bristol

If you would like to present at a future webinar, or to nominate a speaker, you can do so via this online form. We welcome speakers from all career stages across academia and industry.



Participants at the Machine Learning Training School.

Training and community building

Machine Learning Training School

We ran our first Chemistry and Materials Machine Learning School (CaMMLS) from 31 March to 4 April 2025, in partnership with PSDI (Physical Sciences Data Infrastructure) and STFC (Science and Technology Facilities Council). Over 40 students came together from across Europe to gain practical skills and theoretical insights through expert-led talks and tutorials. There were also research seminars from Prof. Kim Jelfs, Imperial College London, and Dr Nong Artrith, University of Utrecht. A round-up of the training school is available on our website.

Due to the success of the training school, we're already in the planning stages for the 2026 edition!

Undergraduate internships

Our 2025 internships have all been allocated, with 12 undergraduate students taking up eight-week research projects across eight different UK universities this summer. These opportunities are helping to expand the next generation of talent at the exciting intersection of AI and chemistry.

Stay in touch

Website: https://www.aichemy.ac.uk
Twitter (X): @aichemyhub
LinkedIn: @aichemy-ukhub
YouTube: @AIchemyhubUK
Bluesky: @aichemyhub.bsky.social
Contact us at: info@aichemy.ac.uk

Catalyst Science and Discovery Centre and Museum News

Contribution from Dr Diana Leitch, Trustee, Catalyst Science and Discovery Centre and Museum, and CICAG Committee Member, email: diana.leitch@gmail.com

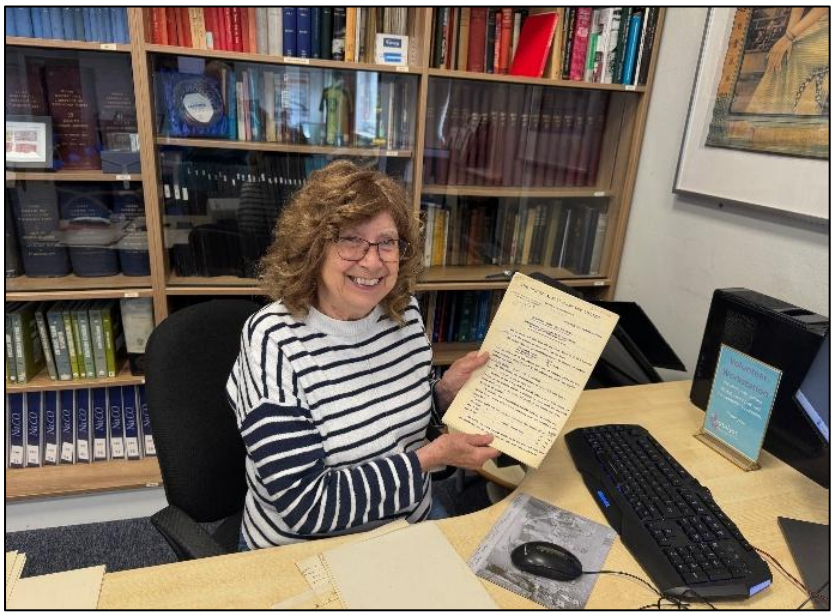
In the Winter 2024 *Distillate* I wrote about the Synergy Project in which Catalyst has been involved now for the last two years. We had heard in October 2024 that we had been successful in our application to NHLF for over £1 million and in early January 2025 we started to put our plans into action for the delivery phase of the project which will take over two years. We hoped the actual work would start in April 2025 but we are still waiting for permission to start from NHLF which hopefully will come in July 2025. As any of you who have run such an NHLF project will know there are many hoops to go through and fences to jump over regarding finances, legalities, contracts, etc before you can start any work and no contracts can be offered to potential new staff and contractors before NHLF agree. We have been allowed to interview and know who we would like to offer the work to, but we have to wait.

In the meantime we have continued our education work with schools and had numerous sleepovers both with uniformed groups (Cubs, Brownies, Boys Brigade, etc) at the weekends and schools during the week. We draw these groups from all over the north west of England and north Wales and have many pupils from diverse backgrounds coming to our workshops and shows.

Chemistry at Work Week in January 2025 was very successful and we had a waiting list of both secondary schools and industrial firms wanting to attend. The IchemE, who sponsored it, were delighted and all participants have asked if we can run a second one each year. As a lot of admin work is involved in putting on such a week our Education Manager, Lucinda, is trying to work out how we could do that alongside all the other education work which is planned.



Our Heritage and Collections Manager/Archivist, Judith Wilde, continues to receive new material for our chemical industry archives virtually every day from individuals who are sorting out lofts or family bequests and are looking for a good home for material. It is not just ICI material but also from other NW chemical industry firms. She also gets contacted to retrieve material as happened recently. Over the past few weeks doing two trips using two cars and a van she has rescued from a derelict building about 150,000 photographs in their individual folders which had been taken by the ICI photographic team in Runcorn before ICI disappeared in the 1980s/1990s. These photos match up with a large quantity of photos we have had at Catalyst for nearly 30 years and are an amazing addition to collective history. It was fortunate they had survived. The 15 volunteers who she organises are already digitising these photographs and trying to provide information about people and processes in the photos while they are doing this. One find last week was the photo of the former Research Director of ICI Mond Division, Dr John Beacham CBE, FRSC, in his younger days presenting skills awards in 1993. John has been a Catalyst Patron for many years.



Digitisation of our extensive collection of ICI General Chemicals Group Technical Reports continues mainly in the hands of Sandra Spruce, who used to be a senior manager for ICI at the Heath before retirement. In this picture you will see her displaying a United Alkali Company Ltd Research Report she is digitising. Up to Friday 5 June 2025 she had personally digitised 50% of these reports and has set herself the target of doing the rest by Christmas 2025! This work is very important as we get still get requests from all over the world from academics and firms for these reports and if they were only in paper format it would involve someone climbing up ladders in our archive store to get each one from their individual boxes. Sandra's dedication is amazing. This digitisation is true 'chemical information and computer applications' at work!

Catalyst is determined to become more inclusive in its provision and to bring under-represented groups in to attend its workshops and shows at the weekends and in the holidays. I reported in the last *Distillate* that my application for a Large Grant to the RSC Outreach Fund for money to support a Project called IDHHP: Interpreters for Deaf and Hard of Hearing People was successful**.** We have been working with the local Deafness Resource Centre to encourage and enable deaf people to come to our workshops and shows but with a BSL Interpreter present. The six workshops which were funded by the RSC were completed by March 2025 and the evaluation of this project has gone to the RSC Outreach Team including a frank report on what works and what doesn't. A lot has been learnt by participating in these. We thought we would have to stop after the initial six months as it is expensive to pay for BSL interpreters but the Chemical Business Association were so impressed by what we were trying to do that they stepped in immediately and provided money for the work to continue until September 2025. In the photograph one of our weekend presenters, Halcyon, is doing a show with the BSL interpreter in the background. Halcyon is a postgraduate student at the Liverpool School of Tropical Medicine and one of our weekend team.



On 18 January 2025 it was 200 years since a renowned Victorian chemist, Sir Edward Frankland FRS, KCB was born at the little village of Caterall near Garstang in Lancashire. Not only is he known as the 'Father of Organometallic Chemistry', the namer of the element helium and the person who defined the theory of valence, but is also renowned for his work on water analysis and water purification from the 1870s onwards. He is credited with saving the lives of thousands of people by his work. He was also the first Professor of Chemistry at Owens College in Manchester when it opened in 1851. My small exhibition about him ran for the month of January 2025 in our Community Exhibition area and two groups of his descendants came to see the display. We had hoped to show it elsewhere and I still have all the display materials if we can find a venue.

A new project will be launched on Friday 11 July which is funded by a local Widnes businessman and the Granada Foundation. It is an example of optical illusion pavement art outside the front door of Catalyst created by a local group of artists called Urban Canvas. It will be the first chemical reaction anamorphic image they have created so I am waiting to see it. We have invited our Patron, Professor Gill Reid, to come to this event. I will send a photo in the next issue of the *Distillate*.

So lots of educational and heritage activities are happening at Catalyst.

We have also had the roof of our Grade II listed building, built in 1855, repaired to keep the rain out, thanks to an Arts Council MEND 3 grant. In the meantime fund raising continues relentlessly for operational activities and new ventures. I continue to give talks in the north-west to adult groups spreading the word about chemistry, chemicals and Catalyst.

Do come and visit us if you are in the area.

- -

News from CAS

Contribution from Dr Anne Jones, Senior Customer Success Specialist, email: ajones2@acs-i.org

CAS SciFinder Discovery Platform™

The CAS SciFinder Discovery Platform consists of multiple solutions that keep you connected to the world's scientific knowledge. Stay aware of global developments and find answers to your scientific questions to help advance your research. Below are notable updates related to these solutions:

- With the advantage of AI-driven query interpretation, text searches for reactions provide more focused, relevant results. This includes query elements such as substance names, molecular formulas, substance classes, and functional groups.
- The substance results display is now customisable. Users can now select up to six experimental or predicted properties that are most important to them to be included in the substance results display.
- The retrosynthesis experience in CAS SciFinder now features improved modelling of multicomponent substances, increasing consistency within a plan.
- An increased breadth of NMR data improves the predicted spectra for substances, with Fluorine-19, Nitrogen-15, and Phosphorus-31 predicted NMR data now present where available.
- It is now possible to use SMILES and InChI (keys or strings) when creating searches for substances, reactions, and suppliers.
- Users in commercial organisations now have access to CAS Chemical Compliance Index™ as part of their subscription. This solution, which can be accessed directly from CAS SciFinder®, provides a single source for global, up-to-date, and accurate compliance and regulatory information sourced from nearly 150 global inventories and lists, covering more than 400,000 regulated substances.
- It is now possible to navigate directly from the Inventory List accordion in CAS Formulus® to the ingredient's detail page in CAS Chemical Compliance Index, giving quicker access to more detailed information within the existing workflow.

More details on the latest enhancements can be found by viewing the "What's New?" section in CAS SciFinder and CAS Formulus. In addition, keep track of significant developments planned for later this year. You can also contact us for more information about using the CAS SciFinder Discovery Platform.

STN IP Protection Suite™

The STN IP Protection Suite consists of multiple solutions, including CAS STNext®, designed to help IP searchers uncover comprehensive insights and minimise risk. CAS continues to enhance these solutions to meet the growing search needs of our users. Recent notable enhancements for CAS STNext include:

- ***Ultimate Owner (/UO) and Ultimate Owner Standardized (/UOS) expanding into additional patent databases***
  UO and UOS data are now available in the US patent databases, USPATFULL and USPAT2, adding to the 10+ databases with UO and UOS information, including CAplus and INPADOC. These new fields provide the current owner of the IP. Include these fields in your search when you need to understand the full scope of an organisation's current IP portfolio.
- ***Evaluation improvement in sequence databases GENESEQ, USGENE®, and PATGENE***
  A new display format (D ALIGNC) in the sequence databases GENESEQ (Derwent Geneseq™), USGENE, and PATGENE enables users to evaluate short sequence stretches in search results of BLAST or FASTA searches quickly by combining them with a motif search (GETSEQ). This format is also useful

for CDR or antibody searches. In addition, a new ALIGN display for GETSEQ searches focuses on alignment, reduces display size for large sequences, and improves display performance.

- ***Pharmaceutical Substances (PS) database expands active pharmaceutical ingredients (API) content***
  Pharmaceutical Substances (PS) has been reloaded with over 100 new active pharmaceutical ingredients (APIs). The new APIs cover kinase inhibitors for cancer treatment, anti-obesity drugs, and rare disease drugs. PS, produced by Thieme, now includes essential information for about 2,900 APIs, including preparation methods and trade data.

Find more details on the latest enhancements to CAS STNext on our product help site. You can also reach out to us directly with any questions or training needs.

CAS BioFinder™

The CAS BioFinder Discovery Platform™ is our newest solution designed by and for drug discovery scientists. With a foundation of interconnected biology and chemistry data, the CAS BioFinder search capabilities, analytical visualisations, and predictive models help those involved with early-stage drug discovery identify targets, understand disease states, and prioritise drug candidates quickly and confidently.

The latest enhancements to CAS BioFinder enable researchers to:

- Search proteins with pharmacological data, pathways, and 3D structures.
- Explore antibodies, ADCs, and metabolites.
- Consult regulatory and clinical information on drugs, including clinical trial status, indications, and targets.
- Identify key targets with scaffold-based matched molecular pair analysis (MMPA) with heat maps.

Screen novel compounds with CAS BioFinder predictive models to move beyond published data and anticipate interactions for more confident lead selection and optimisation. Seamless integration with CAS SciFinder workflows ensures full coverage of your drug discovery process.

CAS Insights™

Actionable insights from emerging research trends across industries

CAS Insights is a free, open resource for actionable perspectives on the latest developments across science, technology, and innovation powered by CAS human-curated data collection and the expertise of our science team.

Stay ahead of the latest trends in emerging areas:

- Lipid-based drug discovery: The latest research and design innovations.
- Materials science breakthroughs: 2025 trends to watch, including metamaterials, industrial decarbonisation, and consumer goods.
- Biomarkers and autism: Objective measurements to improve diagnosis and treatment of autism spectrum disorder (ASD).

Want to make sure you don't miss anything? Subscribe to CAS Insights and get new articles, reports, and webinars delivered to your inbox.

CAS Future Leaders™

On 20 March, CAS announced the selectees for the 2025 CAS Future Leaders program. The new cohort features 35 early-career scientists from 12 countries, conducting research in fields such as nanotechnology,

computational chemistry, materials science, green chemistry, and more. During the in-person program, which will take place in August, participants will attend leadership and career development workshops, learn about mentorship from industry leaders, participate in science discourse, and forge connections with fellow scientists.

The two-week program will also offer participants exclusive insight into CAS operations and how the organisation provides leading innovators with access to and analysis of published scientific information from around the globe. Additionally, participants will have the opportunity to present their research at the ACS Fall 2025 conference in Washington, DC, and will receive a three-year ACS membership as part of the program. To learn more about the program and 2025 selectees, visit our Future Leaders web page.

The Open Reaction Database

Contribution from Benjamin J. Deadman, email: ben@bjdeadman.co.uk

The Open Reaction Database (ORD) is an open-access schema and infrastructure for structuring and sharing organic reaction data, including a centralised data repository. The ORD schema supports conventional and emerging technologies, from benchtop reactions to automated high-throughput experiments and flow chemistry. Our vision is that a consistent data representation and infrastructure to support data sharing will enable downstream applications that will greatly improve the state of the art with respect to computer-aided synthesis planning, reaction prediction, and other predictive chemistry tasks.

Since the ORD coalition formed in October 2019, the database has grown to include 2M reactions (including a large dataset of reactions extracted from USPTO sources) and received contributions from academic and industrial users, both from published and unpublished work. To keep updated with the latest ORD news please follow on LinkedIn, and subscribe to our LinkedIn Newsletter.

The database can be quickly explored using the online graphical browse and search interface, which includes chemical structure searching functionality. Datasets and search lists can be downloaded as an ORD protocol buffer file from the interface and the open source ord-schema package allows users to work with these files in Python. The full ORD database is also available on GitHub as the ord-data repository as the source of truth, and users are encouraged to clone the repo for more complex querying and applications. The schema also includes an Object Relational Mapper (ORM) which allows the serialised protocol buffer data files to be converted into a PostgreSQL relational database.

Release of the improved ORD reaction editor

The new and improved ORD reaction editor was recently released. The new interface streamlines the documentation of chemical reactions, using a tabular view which is intuitive to experimental chemists, while encouraging the detailed recording of procedure details (Figure 1).

The introduction of group management and sharing features encourages collaboration, and the editor platform has been designed to support reaction documentation during the early planning and execution of experiments. The platform could also supplement other Electronic Research Notebook systems, by providing a structured, readable reaction data file which can be uploaded or linked to, adding chemical intelligence to your ERN. The open source code behind the editor is also freely available, and re-use of the code is encouraged. Please contact

us by email if you would like to learn more about how ORD tools could add advanced chemical reaction data support to your ERN system.

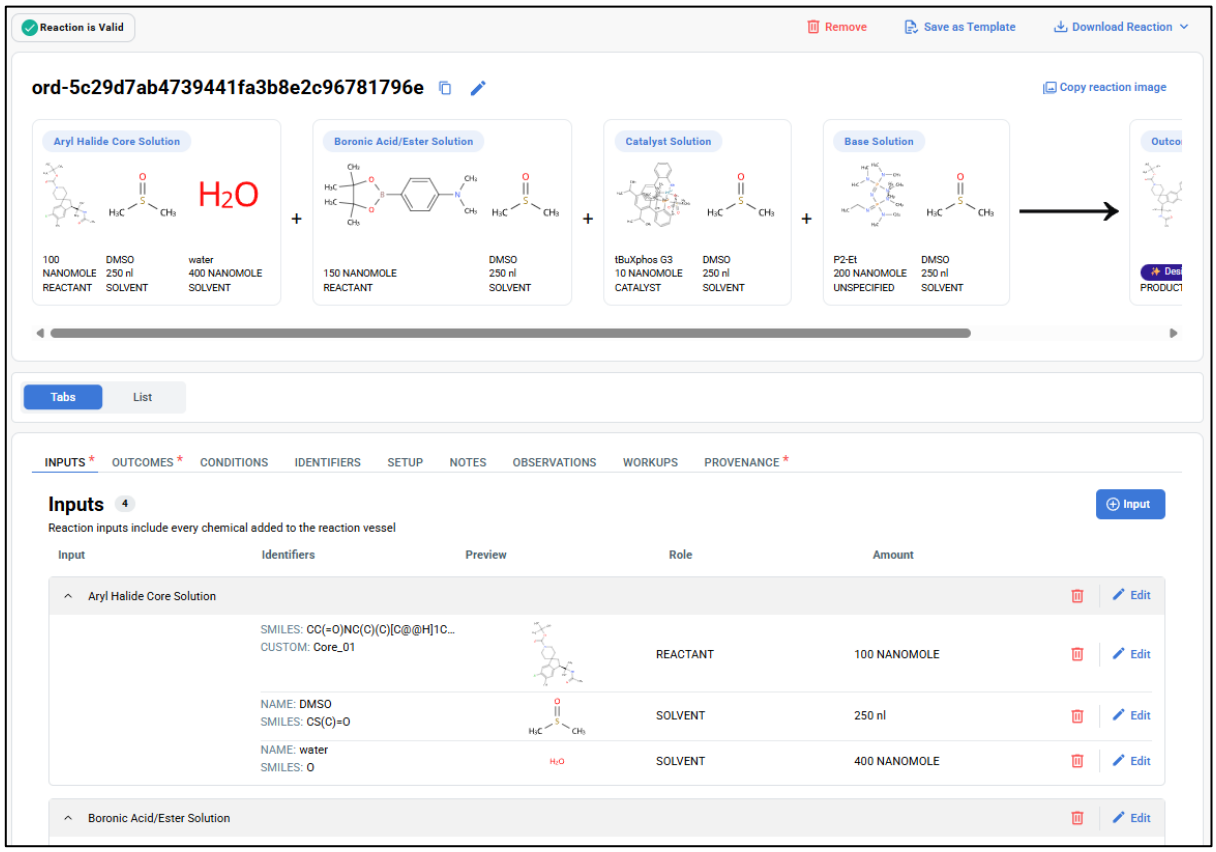


Figure 1. The reaction inputs view of the new improved ORD reaction editor.

Does your group have re-use standard descriptions for analytical methods? The new editor allows users to copy/paste chunks of structured data for re-use in reaction records. The data enumeration workflow has also been streamlined, with a fully graphical interface now simplifying the conversion of tabular reaction data into an ORD dataset.

Front line support is available to help you get started with the ORD, both preparing datasets for submission, and also using the ORD in downstream applications. To request support please submit an issue on GitHub, or contact us by email.

Announcement of the 2024 ORD HTE Center Grant Recipients

In April we announced the recipients of the inaugural round of their ORD HTE Center Grants. These grants are intended to support the creation of large, open-access reaction datasets which will enable the community to develop machine learning models and pursue related efforts in reaction prediction, chemical synthesis planning, and experiment design. The combined efforts of the recipients are anticipated to create almost 100,000 new quantitative reaction data points which will be published under a CC BY-SA licence in the structured ORD format, ready for machine learning by academic and commercial users alike.

Recipients of the first round of ORD HTE Center Grants:

Tim Cernak, The Cernak Lab, University of Michigan
The Cernak Lab will employ ultrahigh-throughput experimentation resources to generate large amounts of chemical reaction data. New ways to couple amines to carboxylic acids in a miniaturised format will be explored. The data available will enable machine learning research into diverse substrates, reaction conditions, and reaction mechanisms.

Shannon Stahl, Daniel Weix, Spencer Heins, UW-Madison Synthesis & Catalysis Center, Department of Chemistry, University of Wisconsin-Madison
The groups of Profs. Shannon Stahl and Daniel Weix at the University of Wisconsin-Madison (UW-Madison) in partnership with the UW-Madison Synthesis & Catalysis Center, Merck and GSK will generate high-throughput experimentation data on cross-electrophile coupling reactions. The project will also demonstrate the potential for partnerships between academia, industrial HTE labs, and funders to accelerate the generation of open reaction data.

Marisa Kozlowski, UPenn HTE Center, University of Pennsylvania
The group of Prof. Marisa Kozlowski will exploit their recently published direct (LC-)MS method for the analysis of regioisomers (ACS Central Science, DOI: 10.1021/acscentsci.4c01765) to generate a large regioselectivity dataset for the Minisci reaction. The dataset will have potential applications in site selectivity prediction and will be a rich source of orthogonal analytical data for machine learning applications beyond traditional summary statistics of yield or conversion.

The 2025 ORD HTE Center Grants call for proposals opened in June, with a submission deadline of Friday 18 July.

The ORD library
We now have a public Zotero Group and library where we keep track of publications relevant to the ORD, including:

- Citations of the Open Reaction Database citations, including:
  - ML applications that have utilised ORD data
  - Published software applications and services which use ORD data, schema or other tools
- The journal publications which have reactions in the ORD database
  - It does not include the individual USPTO patents (there are over 13,000 of them) for clarity

Training opportunity
The ORD Trainee Program has also been recently revised to make use of the new reaction editor interface, and provide a more structured experience. The program, which is open to students and early-career research chemists in the UK and USA, sees trainees working with the ORD Program Manager to prepare 1-2 datasets in the ORD format, from previously published research. It is expected that applicants spend up to 10 hours per dataset over a 2-3 month period and there is some flexibility to adapt this to the applicant's situation.

Participants gain valuable knowledge and experience of the ORD toolset to define reaction data in a structured format. All dataset submissions to the ORD receive formal recognition and a persistent identifier which can be included in CVs, professional web pages and professional development records. Time spent on pre-approved ORD activities is also reimbursed at a rate of up to 30 USD per hour, with a maximum limit of paid hours per dataset.

If you are interested in joining the program then please submit an Expression of Interest through our Google Form. Trainees are inducted in batches, after selection from our waiting list.

How you can get involved

The Open Reaction Database is a community effort and we welcome you to get involved. Some areas where we are seeking the community's support are:

- Please use the ORD data to do great things, and if possible tell us about it. The data is all published under a CC-BY-SA licence to permit both academic and commercial use.
- Do you have some reaction data that could be made open access? It could be some historical data, or a new paper that you are publishing soon. Please get in touch so we can support you in preparing your first ORD dataset.
- Do you want more or different reaction data than what we have in the ORD already? Please work with us as we connect experimental data generators with the machine learning community. We think there is huge potential in our community and we are working together on collaborative funding bids for further dataset creation.
- We can help software vendors to implement ord-schema input and output tools in their software. While we don't have the resources to develop translators for every ELN or reaction data management software, we can usually provide advice on how the ord-schema relates to your data structures.
- Finally, show your support for the ORD with one of our laptop stickers. Look for Ben at conferences, or get in touch to request some by post.

Upcoming events

The ORD will be presented at the following events:

- ACS Fall 2025, 17-21 August, Washington DC, United States
- ELRIG Drug Discovery 2025, 21-22 October, Liverpool, United Kingdom
- More to be announced.

Reference

Kearnes, S. M. et al. The Open Reaction Database. *J. Am. Chem. Soc.* **2021**, *143* (45), 18820-18826. https://doi.org/10.1021/jacs.1c09820

Molecular and Materials Modelling at Newcastle University (2015-2025)

Contribution from Tom Penfold, Agnieszka K. Bronowska, Danny Cole, James Dawson & Ioan-Bogdan Magdau. School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, email: daniel.cole@ncl.ac.uk

We have been asked to give a general overview of computational chemistry at Newcastle University. The University was established in 1834, and the College of Physical Science was founded later in 1871. Notable alumni in theoretical chemistry include Dr Ian Cooper, who worked on statistical and quantum mechanics, and Prof. Jeremy Burdett, who studied the electronic structure and properties of inorganic solids whilst working as a senior research officer at Newcastle University. The Schools of Chemistry and Physics have always worked closely together, and indeed they briefly merged in the early 2000s, which led to Prof. Patrick Briddon (author of the AIMPRO modelling software), Dr Jerry Hagon and Prof Werner Hofer bringing in expertise from condensed matter physics and quantum theory. What follows is an informal set of personal perspectives of our

current staff members on their research and teaching, as well as our engagement with the wider computational chemistry community, covering the period from 2015 to the present day.



Image from the 2024 CCPBioSim conference held at Newcastle University.

Tom Penfold joined Newcastle University in September 2015 as Lecturer in Theoretical and Computational Chemistry, following a PhD (2010, University of Birmingham, under Graham Worth), and postdoctoral appointments at EPFL Lausanne and the Paul Scherrer Institute working on ultrafast spectroscopy and X-ray free electron laser methods. His group focuses on theoretical and computational investigations of electronically excited-state dynamics – emphasising quantum nuclear dynamics and non-adiabatic processes – to understand and predict time-resolved spectroscopic signatures in molecules and materials.

The core motivation of his research is simulating and understanding excited state processes, often using dynamical methodologies (wavepacket and trajectory-based techniques) combined with first-principles electronic structure, aimed at elucidating electronic relaxation pathways such as internal conversion and intersystem crossing. Tom established a significant research programme in the area of molecular emitters for thermally activated delayed fluorescence (TADF) in organic light-emitting diodes (OLEDs). His work revealed the 'spin-vibronic mechanism' for TADF, demonstrating that triplet harvesting involves not only charge-transfer states but also coupling to states of different electronic character activated by vibrational modes. This mechanism has become a cornerstone of the field, influencing the understanding of intersystem crossing in a wide range of functional organic and metal-organic systems. Subsequent studies have further elucidated the role of host environments and solid-state solvation in tuning the energetics and emission properties of TADF molecules, addressing key challenges in the design of efficient emitters.

Tom's research has also combined excited state dynamics simulations with ultrafast spectroscopic experiments – including pump–probe and core-level X-ray spectroscopy – leveraging high-performance computing to interpret and guide experimental results. His work on ultrafast X-ray absorption spectroscopy, provided the foundational principles for interpreting femtosecond time-resolved X-ray absorption near-edge structure (TR-XANES) experiments. This approach enabled the direct tracking of wavepacket dynamics in transition metal

complexes and was directly responsible for the experimental findings. This work and the computational approaches developed have been extended to disentangle the interplay between electronic and nuclear degrees of freedom in single molecule magnets.

Recognising the computational challenges of simulating dynamic observables, Tom has spearheaded the application of machine learning to computational spectroscopy. His algorithm, XANESNET and datasets, provides affordable, accurate, and generalisable predictions of X-ray absorption spectra, achieving an average accuracy of ±2%. This open-source tool has democratised access to high-fidelity spectroscopic predictions, paving the way for data-driven approaches to complement traditional methods in computational spectroscopy.

Tom holds a EPSRC Open Fellowship grant entitled "Deep Neural Networks for Real-Time Spectroscopic Analysis", to develop and subsequently equip researchers with easy-to-use, computationally inexpensive, and accessible tools for the fast and automated analysis and prediction of X-ray spectra using deep neural networks. He is a member of the COSMOS Programme grant which seeks to achieve the creation and widespread adoption of advanced universal software for quantum dynamics simulations.

Recognising the ever increasing role of computational chemistry in molecular design, Agnieszka Bronowska and Danny Cole joined in 2016 to establish computational medicinal chemistry as a new research area at Newcastle. Newcastle has had a strong track record of early-stage drug discovery and translational research, including for example the development of the first Newcastle drug, rucaparib, for treatment of BRCA-deficient cancers.

Agnieszka has a background in computational chemistry and computational structural biology, during her postdoctoral period she gained expertise in experimental techniques such as recombinant protein production and binding assays (biophysical and cell-based). Her research interests revolve around structure-based drug design as a strategy to target individual proteins and protein-protein complexes with small organic molecules and short peptides. She is particularly interested in development of allosteric inhibitors and in mapping of novel, cryptic and metastable 'druggable' sites. She is also interested in modelling assembly and regulation of ultra-large protein complexes such as inflammasomes and autophagy receptors, and developing multiscale workflows to reliably capture structure and dynamics of those challenging systems.

Agnieszka's group is pursuing multiscale molecular dynamics simulations and molecular docking calculations to study relationships between structure, dynamics, and biological function of relevant protein targets. After her appointment, she engaged in several successful collaborations with experimental cell and molecular biology groups, e.g. with Prof. Viktor Korolchuk (Newcastle University), using multiscale modelling techniques to understand modulation of autophagy receptors, Prof. Tim Blower (Durham University), deconvoluting mechanism of regulation of mycobacterial toxin-antitoxin systems, Prof. Haike Antelmann (FU Berlin) on modelling allosteric regulation in bacterial proteins, and Dr Rebecca Coll (University of Belfast), discovering the role of NLRP3 inflammasome as an intracellular thermosensor, and provided molecular simulation support to multidisciplinary projects within Newcastle University (e.g. works on novel inhibitors of cancer-linked protein kinases such as NIK and EGFR, led by Prof. Mike Waring).

Agnieszka's research, at the interface between chemistry, biology and physics, has been funded by EPSRC, BBSRC, industrial partners, and charities (e.g. Alzheimer Research UK). Over the last years, she has received several Impact Accelerator Awards – small grants awarded to enable commercialisation of research ideas and fostering engagement with different stakeholders, including industrial partners. These required working alongside multiple partners from different sectors in joint development of ideas, discussing the outputs and planning future directions. Agnieszka's ongoing engagement with the industrial partners resulted in three

iCASE PhD studentships (co-funded by AstraZeneca and RxCelerate), two IAFRI-FERA PhD studentships, a BBSRC-funded grant (with Fontus Environmental as a Project Partner), and ongoing collaboration with RxCelerate (since 2022), which resulted in the recently awarded KTP (InnovateUK), which aims to deliver methodology to computationally predict and subsequently experimentally validate 'druggable' states in disease-linked proteins.

Agnieszka's experience in research, computational as well as experimental techniques, and training of early-career researchers enables her to "look at the bigger picture", yet remaining technically focussed, which is crucial for research and innovation. For example, this includes identifying gaps in the current research funding landscape, priority areas in rapidly changing fields such as drug discovery, benefits provided by the rapid development of AI methods in chemistry and structural biology and their intrinsic challenges, skills and training.

On the other hand, Danny arrived with a background in Physics. His PhD in condensed matter physics focussed on impurity segregation and water structuring at oxidised silicon surfaces. As is typical for many career paths, his journey to medicinal chemistry was a gradual transition. He took up a postdoctoral position with Prof. Mike Payne in Cambridge to work on linear-scaling density functional theory, which was just becoming mature enough to give access to typical length scales in biology. Danny showed, for example, that such methods can be used to study the binding of small molecules by metalloproteins and the parameterisation of model Hamiltonians to describe energy transfer in light-harvesting complexes.

However, a down-side of working exclusively with quantum methods is that there's a lot of sitting around waiting for simulations to finish! Around 2010, Danny saw a talk by Prof. Bill Jorgensen (similar to this one) about the use of formally rigorous free energy methods for calculating the binding of small molecules to therapeutic targets, in theory, enabling the design of new drugs on the computer. Of course in practice this is much more difficult than it first sounds, and Danny was fortunate enough to spend some time in Bill's lab at Yale University to learn about the importance of molecular mechanics force fields and sampling of biomolecular dynamics in the accurate determination of protein-ligand binding affinities.

Much of this previous experience has inevitably shaped the research direction of Danny's group at Newcastle. Today, he holds a UKRI Future Leaders Fellowship that focuses primarily on the development of accurate force fields to describe the dynamics and interactions of molecules at the atomic scale, and their application to computer-aided drug design. He is passionate about open science, and pleased to collaborate with the Open Force Field initiative, and wider Open Molecular Software Foundation. He has contributed to one of the mainline small molecule force fields ('Sage'), and led the development of the OpenFF-BespokeFit package for fitting molecule-specific torsion parameters to quantum mechanics. As an example proof-of-concept work, he experimented with using the Open Force Field software stack to replace the Lennard-Jones functional form with a more flexible double exponential function.

Danny's group also contributes open software packages for de novo drug design and absolute binding free energy calculations. He works with the Newcastle University Centres for Cancer and Rare Diseases to apply these methods for prospective drug design, an example of which was participation in the CACHE challenge to design new small molecule inhibitors of the SARS-CoV-2 NSP13 helicase.

After a JSPS Postdoctoral Fellowship at Kyoto University and postdoctoral positions at Cambridge and Bath, James Dawson started his independent research career at Newcastle in January 2020 and now leads the Ion Transport and Interfaces Group, with over 20 members. His expertise lies in the atomistic modelling (primarily density functional theory and molecular dynamics) of energy materials.

Interfaces and ion transport are central to the performance and stability of energy materials and devices, including batteries, fuel cells and photovoltaics. However, the stunning heterogeneity, complex ion diffusion mechanisms and interfacial processes that occur in energy materials are often not fully understood, particularly at the atomic scale. The mastering of interfaces and ion transport in energy materials directly contributes to the development of next-generation devices to aid our transition towards electric mobility and net zero and fundamental research advances in chemistry, materials science, physics and engineering. On this basis, James' research is focused on understanding and designing energy (storage, conversion and generation) materials and their interfaces for current and future applications to combat climate change and the ever-evolving energy crisis. Specifically, his group employs both atomistic modelling and machine learning techniques to improve the fundamental comprehension and performance of material structures, interfaces and properties for a variety of applications.

Much of James' research is focused on solid-state batteries, which represent a prime example of a next-generation battery technology with the potential to revolutionise energy storage. Solid-state batteries are based on solid electrolytes and are often touted to have two main advantages over conventional Li-ion batteries. The first is the increased safety that results from isolating the electrodes with the solid electrolyte to avoid Li build-up that can trigger short-circuiting. The second is the energy density gains resulting from the use of an energy dense Li-metal anode. However, these advantages have yet to be fully realised and there are many fundamental challenges to overcome, including electrochemical stability and interfacial resistance. Solid-state batteries have been described as the battery technology of the 2030s but the research challenge of the 2020s.

James holds a UKRI Frontier Research grant entitled "Amplifying Ion Transport at the Interfaces of Solid-State Batteries (AMPed)", which seeks to revolutionise the understanding and design of the ion transport and interfaces within solid-state battery architectures. AMPed utilises state-of-the-art classical, quantum mechanical, structure prediction and machine learning approaches to develop battery materials with improved performance, stability and sustainability. James is also actively involved in the Faraday Institution, the UK's flagship battery research programme, and is a Co-Investigator on both the Next-Generation Solid-State Batteries (SOLBAT) and Battery Degradation projects.

Ioan joined the Computational Chemistry group at Newcastle in April 2023. His research focuses on method development and new applications of Machine Learning Interatomic Potentials (MLIPs), a novel technology transforming the field.

Foundational MLIPs are revolutionising ground-state chemistry and materials science by extending ab initio accuracy to larger length and time scales. These models excel in regions where potential energy surfaces are smooth. But what about excited states? In collaboration with Tom, Ioan is exploring whether the same success can be achieved for excited-state chemistry, where the energy landscapes are far more complex. Around conical intersections, the surfaces become non-smooth and the forces discontinuous. Moreover, non-adiabatic dynamics requires not only energies and forces but also non-adiabatic coupling vectors, vector fields that cannot be expressed as gradients of scalar functions, making them much harder to model with current MLIP approaches. Could these challenges be overcome? Could we one day build a foundational MLIP for excited states? Imagine the possibilities: an atlas of high-throughput photochemistry, large-scale non-adiabatic dynamics in extended biological systems, and more.

Speaking of biological systems, MLIPs are not dissimilar from classical molecular mechanics force fields. Traditional force fields rely on simple, fixed functional forms with hand-tuned parameters based on bonding and non-bonding terms. MLIPs, by contrast, generate highly expressive, data-driven functional forms parameterised automatically from atomic environments. They identify atom types and local features through

learned similarity measures rather than human-defined categories. While classical force fields are fast and ideal for large biological systems, they struggle to model reactions or strong local interactions. MLIPs offer much higher accuracy but at a significantly greater computational cost. In collaboration with Danny, Ioan is investigating the middle ground: can we sparsify MLIP representations to reduce their computational burden while preserving accuracy? Conversely, can we use insights from MLIPs to identify missing atom types or many-body interactions in classical force fields, enhancing their accuracy at modest cost?

Another of Ioan's research directions, in collaboration with James, involves applying foundational models to study cathode-electrolyte interfaces in battery materials. Degradation at these interfaces is a key factor limiting battery lifetime. Understanding the molecular processes that reduce capacitance, and performance could guide better battery design. With foundational MLIPs, it is now possible to model such complex, condensed-phase systems, thousands of atoms over nanoseconds, with unprecedented fidelity. Ioan and James are pushing the boundaries of these models at the interface between layered NMC cathodes and multi-component carbonate electrolyte mixtures. Key questions include: How well do foundational MLIPs perform in these chemically diverse environments? Can they reproduce ab initio accuracy? And can they explain, or even predict, experimental observations?

Teaching

As the role of computers in chemistry has evolved, so has our teaching of computational chemistry at Newcastle. A curriculum refresh in 2020 allowed us to embed scientific computing (using python taught interactively through Jupyter notebooks) in the first two years of the undergraduate degree. This is taught alongside core material in quantum mechanics, and is supplemented by molecular modelling workshops in physical chemistry in the third year.

Agnieszka drove the major restructuring of the Drug Chemistry MSc programme, specifically updating the admission criteria and novel Research Dissertation Project Module (launched in 2024/25 academic year; 80 credits with extended skills training). The new module was shaped after the consultations with industrial partners, and in response to gaps in key skills they identified as detrimental to employability of our graduates (e.g. computational and analytical chemistry). The module involves streamlined training for the entire cohort with the same key areas/tasks aligned with modern drug discovery workflows (data-driven discovery, prioritisation, synthesis, analytical chemistry and validation). The pilot, launched in 2022/23, received very good feedback from students and academics alike, and the external examiners deemed the programme "sector-leading". Following this, several other programmes (e.g. Chemistry MSc) are currently going through the consultations regarding adopting a similar model of the project work for their students.

Agnieszka leads a pilot on the use of generative AI in student assessment. Since its launch, generative AI gained significant attention in higher education, and many universities banned AI tools over misinformation and plagiarism concerns. This pilot explores the ways in which we can safely and effectively use AI for learning and collecting reliable information. If successful, the new assignment model can be transferred to other modules and replace more traditional assignments which are at higher risk of plagiarism using generative AI tools.

Community

Newcastle Chemistry is proud to host the renowned CCP5 summer school. Founded over 40 years ago, CCP5 continues to attract PhD students and postdocs from across the UK and beyond. This is the second consecutive year that Newcastle has hosted the school. Ioan, James, Agnieszka, and Danny have all been involved in teaching and/or hosting and are delighted to see the program thriving at Newcastle.

Agnieszka and Danny are members of the Management Group of CCPBioSim, and involved in organising annual CCPBioSim conferences (Newcastle 2024, Southampton 2025). Agnieszka has been co-organising bi-annual Manchester Multiscale Meetings for a number of years, and this year she is organising the joint CCPBioSim-CECAM Flagship Workshop "Entropy of Soft Matter" in Vienna. Tom also leads the HPC-CONEXS network, which is part of the EPSRC high-end computing consortium.
Computation has played an important role in two centres for doctoral training, with which Newcastle Chemistry has recently been involved: "Molecular Sciences for Medicine" (MoSMed) and "Renewable Energy Northeast Universities" (ReNU), both aiming to train the next generation of scientists and produce cutting-edge research at the interface between computation and experiments.

UKeIG: 2025 Tony Kent Strix Award

Contribution from Gary Horrocks, UKeiG, CILIP, email: info.ukeig@cilip.org.uk

A call for nominations for the 2025 Tony Kent Strix Award
The Tony Kent Strix Award was inaugurated in 1998 by the Institute of Information Scientists. It is presented by UKeiG in partnership with the International Society for Knowledge Organisation UK (ISKO UK), the Royal Society of Chemistry Chemical Information and Computer Applications Group (RSC CICAG) and the British Computer Society Information Retrieval Specialist Group (BCS IRSG) and awarded in recognition of an outstanding practical innovation or achievement in the field of information retrieval and search. Nominations must be for a major, sustained or influential achievement that meets one or more of the criteria listed below.

Science
The advancement of our understanding of information retrieval methods, experimentation and evaluation, at either the theoretical or the practical level. The scope includes approaches as diverse as linguistic, probabilistic, fact-checking or artificial intelligence applied to search.

Service delivery
The development and management of systems, networks or services:

- Enhancement of the mechanisms/technology/standards underpinning information products or services
- Establishing an innovative information resource or service
- Innovations leading to improved accessibility/usability of information resources

Education and organisational infrastructure
The provision of leadership in education, training, community development and/or collaboration to advance information retrieval at local, national or international level.

Submission guidelines
Nominations should take the form of a succinct justification (maximum 2 pages) for the individual or team nomination showing clearly which of the award criteria the nominee meets and how they are met. Please include full contact details, job title and qualifications of both the nominee and the nominator emailed to Richard Bridgen – Hon. Secretary UKeiG secretary.ukeig@cilip.org.uk and copied to: Gary Horrocks - info.ukeig@cilip.org.uk and Elizabeth Roberts – Hon. Treasurer UKeiG treasurer.ukeig@cilip.org.uk.
Please clarify in the email subject heading - 'UKeiG 2025 Strix award nomination'.

Self-nominations will not be accepted.
Please include testimonials, letters of support, references, a selective bibliography relevant to the nomination, or links to supplementary information including professional profiles or social media, for example.
Please email info.ukeig@cilip.org.uk for more information and guidance.

Deadline
Nominations must reach the UKeiG by **6 pm GMT on Friday September 12th, 2025.**

Further information
Further information about the Strix award is available here, and includes previous winners.

Please share the news
The library and information science community is encouraged to raise awareness of these awards with individuals and groups in their networks and help encourage nomination submissions.

The winners will be invited to present at an online awards ceremony in November 2025.

- -

Other Chemical Information News

Contribution from Stuart Newbold, email: stuart@psandim.com

CAS BioFinder accelerates Drug Discovery
Customers' use of the tool reveals a hunger for structured, curated data.
https://cen.acs.org/acs-news/CAS-BioFinder-accelerates-drug-discovery/103/web/2025/06
Source: Chemical & Engineering News

Biochemical Society Journals transition to full Open Access in 2025
The Biochemical Society has confirmed that all five of its journals will be published open access (OA) in 2025 under the Subscribe to Open (S2O) model. This transition was unanimously endorsed by the Portland Press Board and made possible through the renewed commitments of libraries under the S2O model. As a result, all research published in these journals will be freely accessible worldwide with no author fees. The following journals will be fully OA:

- Biochemical Journal (Volume 482)
- Clinical Science (Volume 139)
- Biochemical Society Transactions (Volume 53)
- Essays in Biochemistry (Volume 69)
- Emerging Topics in Life Sciences (Volume 9)

https://www.knowledgespeak.com/news/biochemical-society-journals-transition-to-full-open-access-in-2025/
Source: Knowledgespeak

AlphaFold is running out of data — so drug firms are building their own version
Thousands of 3D protein structures locked up in big-pharma vaults will be used to create a new AI tool that won't be open to academics.
https://www.nature.com/articles/d41586-025-00868-9
Source: Nature

How to find the most important research amid the growing number of publications?

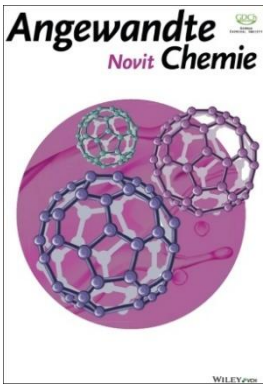


Image courtesy of Angewandte Chemie

The German Chemical Society (GDCh) and Wiley-VCH launch *Angewandte Chemie Novit*, a highly selective journal with a unique selection process to give the best research extra visibility. *Angewandte Chemie Novit* is a highly selective journal dedicated to elevating the most significant chemistry research, inspiring discourse, and stimulating further research in important and emerging areas of chemistry.
Instead of receiving manuscripts through direct submissions, *Angewandte Chemie Novit* will draw on the wide body of high-quality research submitted to *Angewandte Chemie*. We have designed a unique multi-step manuscript selection process that identifies articles for their scientific merit and expected long-term impact.
https://www.chemistryviews.org/how-to-find-the-most-important-research-amid-the-growing-number-of-publications/
Source: Chemistry Views

I told AI to make me a protein. Here's what it came up with
A new crop of artificial-intelligence models allows users to create, manipulate and learn about biology using ordinary language.
https://www.nature.com/articles/d41586-025-01586-y
Source: Nature

bioRxiv and medRxiv transition to independent governance under openRxiv
bioRxiv and medRxiv, the leading preprint platforms for life sciences and medical research, will now operate under the openRxiv governance model, ensuring long-term sustainability through a community-driven nonprofit framework.
https://www.knowledgespeak.com/news/biorxiv-and-medrxiv-transition-to-independent-governance-under-openrxiv/
Source: Knowledgespeak

AI linked to explosion of low-quality biomedical research papers
Analysis flags hundreds of studies that seem to follow a template, reporting correlations between complex health conditions and single variables based on publicly available data sets.
https://www.nature.com/articles/d41586-025-01592-0
Source: Nature

67 Bricks develops new agentic AI research assistant for proprietary content
For publishers wanting to integrate natural language querying without reducing control over their content, 67 Bricks are now offering a compelling solution. Their new tool, developed in-house, can be seamlessly integrated into all existing bespoke or third-party platforms, delivering a research tool that answers users' queries with responses grounded in paywalled and proprietary content.
https://www.stm-publishing.com/67-bricks-develops-new-agentic-ai-research-assistant-for-proprietary-content/
Source: STM Publishing News

OpenAI's 'deep research' tool: is it useful for scientists?
The model produces cited, pages-long reports that might be helpful for generating literature reviews.
https://www.nature.com/articles/d41586-025-00377-9
Source: Nature

Artificial intelligence is revolutionising everything from workflows to networking - so how can you bring AI into your research practice?
https://www.chemistryworld.com/research/how-ai-is-transforming-chemistry-research/4020650.article
Source: Chemistry World

CAS introduces CAS Search Services to support scientific teams with limited capacity
CAS has announced the launch of CAS Search Services℠, a new offering designed to assist CAS SciFinder® users with literature searches. The service aims to support research teams facing capacity constraints by providing access to expert search execution and analysis, enabling scientists to focus more fully on their core research activities.
https://www.knowledgespeak.com/news/cas-introduces-cas-search-services-to-support-scientific-teams-with-limited-capacity/
Source: Knowledgespeak

Web of Science Research Intelligence available for early adopters in August
Software solution aimed at enabling research institutions to accelerate innovation and showcase societal impact.
https://www.researchinformation.info/news/web-of-science-research-intelligence-available-for-early-adopters-in-august/
Source: Research Information

What are the best AI tools for research? Nature's guide
There are many large language models to choose from; some excel at coding, whereas others are better for synthesising information.
https://www.nature.com/articles/d41586-025-00437-0
Source: Nature

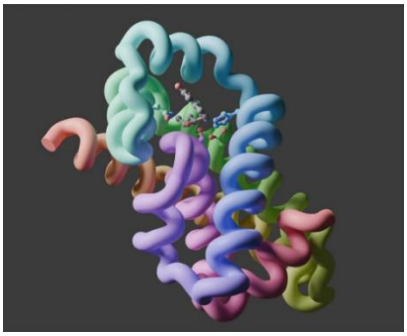
Making AI-generated code more accurate in any language
Researchers developed a more efficient way to control the outputs of a large language model, guiding it to generate text that adheres to a certain structure, like a programming language, and remains error free.
https://www.sciencedaily.com/releases/2025/04/250424121658.htm
Source: Science Daily

Elsevier adds half a million records from ClinicalTrials.gov to Embase, enabling a seamless search experience in the world's most comprehensive biomedical database
Data from studies conducted in more than 200 countries supports greater access to research and facilitates systematic reviews, clinical trial design, medical device approvals and improved drug safety.
https://www.elsevier.com/about/press-releases/elsevier-adds-half-a-million-records-from-clinicaltrials-gov-to-embase
Source: Elsevier

Case Study: Facilitating AI/ML in chemistry R&D - data science and high throughput experimentation
High throughput experimentation (HTE) enables efficient AI/ML dataset production and benefits from AI-driven experiment design. This case study presents how leading pharma firms are using Katalyst D2D software to streamline workflows, generate high-quality datasets, and apply integrated machine learning to reduce experimental workload while enhancing data-driven insights for discovery and process chemistry.
Link
Source: Scientific Computing World

Copyright compliance and AI awareness "more important than ever"
Latest content usage trends "reaffirm the importance of having a solid copyright policy in place".
https://www.researchinformation.info/news/awareness-of-copyright-compliance-and-ai-tools-more-important-than-ever/
Source: Research Information

Scientists use AI to design life-like enzymes from scratch



Combined approach takes AI-engineered enzymes one step closer to practical applications.
Researchers have used artificial intelligence (AI) to design brand-new enzymes that can go through multi-step reactions, a key feature of natural enzymes. The structures they made accelerated a four-step chemical reaction pivotal to many biological and industrial processes, including plastic recycling. Earlier efforts to design enzymes from scratch using AI have had limited success, often producing ones that stall after the first step of a reaction.
https://www.nature.com/articles/d41586-025-00488-3
Source: Nature

Everyone talks about making a first impression, but what about your last one?
Whether you're moving on or being moved out, how you leave can shape your reputation more than how you led.
https://www.biospace.com/career-advice/everyone-talks-about-making-a-first-impression-but-what-about-your-last-one
Source: BioSpace

'WithdrarXiv' database of 14,000 retracted preprints launches
A trove of data is providing insights into the main reasons studies are pulled from the arXiv preprint platform.
https://www.nature.com/articles/d41586-025-00011-8
Source: Nature

Springer Nature donates AI tool to publishing community
Tool designed to identify artificially generated nonsense text will be available via the STM Integrity Hub.
https://www.researchinformation.info/news/springer-nature-donates-ai-tool-to-publishing-community/
Source: Research Information

Agentic AI and AI Librarians: Future roles for the profession
The newest wrinkle in AI technologies involves agents, AI that independently takes action, executes tasks, solves multi-step problems, and makes decision. Will Agentic AI open up new job opportunities for AI Librarians? It's possible.
https://www.infotoday.eu/Articles/Editorial/Featured-Articles/Agentic-AI-and-AI-Librarians-Future-roles-for-the-profession-167436.aspx
Source: Information Today

We need to build the data that chemistry deserves - here's how we can do it
Creating a purpose-built repository of standardised reaction data is a tall order, but the reward would be huge.
https://www.chemistryworld.com/opinion/we-need-to-build-the-data-that-chemistry-deserves-heres-how-we-can-do-it/4021420.article
Source: Chemistry World

Clarivate launches DRG Fusion – a new life sciences analytics product powered by real-world data
Clarivate has announced the launch of DRG Fusion, an innovative platform to support commercial analytics in life sciences. Powered by integrated real-world data and built by clinical and data science experts, Fusion provides biopharma and medtech organisations with the tools to understand and act in varied disease and competitive landscapes — eliminating the complexities of raw data management.
https://clarivate.com/news/clarivate-launches-drg-fusion-a-new-life-sciences-analytics-product-powered-by-real-world-data/
Source: Clarivate

Clarivate's EndNote 2025 features AI support
Features "streamline tedious tasks and provide insights and recommendations that improve the research journey".
https://www.researchinformation.info/news/clarivates-endnote-2025-features-ai-support/
Source: Research Information

CCC launches RightFind software bundle to enhance literature management for Life Science companies
CCC has introduced its RightFind Suite Growth Bundle, a comprehensive software solution designed to help small- to mid-sized life science companies streamline their scientific literature workflows. The bundle allows companies to efficiently access, organise, and collaborate around scientific literature through a single enterprise software subscription.
https://www.knowledgespeak.com/news/ccc-launches-rightfind-software-bundle-to-enhance-literature-management-for-life-science-companies/
Source: Knowledgespeak

AI "will play significant role in research process within two years"
Majority in Wiley study believe AI currently outperforms humans in more than half of the 43 use cases tested.
https://www.researchinformation.info/news/ai-will-play-significant-role-in-research-process-within-two-years/
Source: Research Information

Science sleuths flag hundreds of papers that use AI without disclosing it
Telltale signs of chatbot use are scattered through the scholarly literature — and, in some cases, have disappeared without a trace.
https://www.nature.com/articles/d41586-025-01180-2
Source: Nature

Free PDF of Condensed Version of IUPAC's Green Book
IUPAC's official guide on quantities, units, and symbols (Green Book – abridged version) is available as a free PDF.
https://www.chemistryviews.org/free-pdf-of-condensed-version-of-iupacs-green-book/
Source: Chemistry Views

New Atom-Swapping method applied to Complex Organic Structures
Chemists have developed an efficient skeletal editing method for frequently used heteroaromatic structures. The technique could serve as a means to chemically modify biologically active compounds.
https://www.sciencedaily.com/releases/2025/05/250522125415.htm
Source: Science Daily

EBSCO launches new AI features
New capabilities are part of "ongoing commitment to enhancing the researcher journey and user experience".
https://www.researchinformation.info/news/ebsco-launches-new-ai-features/
Source: Research Information

Elsevier launches ScienceDirect AI to transform research with rapid mission-critical insights from trusted content
New generative AI tool on ScienceDirect, the world's largest platform for peer-reviewed research used by millions of researchers each day, saves half of literature research time, supports better outcomes.
https://www.elsevier.com/about/press-releases/elsevier-launches-sciencedirect-ai-to-transform-research-with-rapid-mission
Source: Elsevier

Clarivate Enhances OFF-X with Pathway Maps to Accelerate Target Safety Assessments
Clarivate has announced the incorporation of Pathway Maps in OFF-X, a cutting-edge translational safety intelligence solution. With over 30% of drug development failures attributed to safety issues, preclinical and translational research teams face increasing pressure to identify potential liabilities earlier in the target selection process. Pathway Maps provide powerful visual tools that offer deeper insights into the biological mechanisms underlying diseases and target interactions, streamlining target safety assessments and enabling faster, data-driven decisions.
https://clarivate.com/news/clarivate-enhances-off-x-with-pathway-maps-to-accelerate-target-safety-assessments/
Source: Clarivate

AI tools are spotting errors in research papers: inside a growing movement
Large language models are being used to check papers, but researchers warn they come with risks.
https://www.nature.com/articles/d41586-025-00648-5
Source: Nature

LIBER releases action plan to address barriers in accessing and reusing research outputs
LIBER, the Association of European Research Libraries, has published a detailed Action Plan aimed at improving access to and reuse of research results by addressing legal and regulatory challenges surrounding copyright and data rights.
https://www.knowledgespeak.com/news/liber-releases-action-plan-to-address-barriers-in-accessing-and-reusing-research-outputs/
Source: Knowledgespeak

Chemists create matchmaking app for biocatalysis
Machine learning model Catnip helps chemists find biocatalysts to try in synthesis.
https://cen.acs.org/acs-news/acs-meeting-news/Chemists-create-matchmaking-app-biocatalysis/103/web/2025/03
Source: Chemical & Engineering News

AI, Copyright & Licensing Insights
Copyright is central to high quality outcomes as copyrighted material is the fuel for AI systems. Licensing is an effective solution enabling the use of copyrighted material as society realises the benefits promised by AI systems. Read the latest news and insights.
https://www.copyright.com/resource-library/communities/ai-copyright-licensing
Source: CCC

Springer Nature adds AI tool to identify irrelevant references in manuscripts
Springer Nature has introduced a new AI-based tool to detect irrelevant references in submitted manuscripts, expanding its suite of in-house research integrity technologies. The tool is being deployed by the publisher's Research Integrity Group (RIG) across nearly all Springer Nature journals and books. It automatically analyses the relevance of cited references in submissions and flags manuscripts with a high volume of irrelevant citations for further manual review by RIG. If confirmed, the submission may be withdrawn.
https://www.knowledgespeak.com/news/springer-nature-adds-ai-tool-to-identify-irrelevant-references-in-manuscripts/
Source: Knowledgespeak

Meta's AI memorised books verbatim – that could cost it billions
Many AI models were trained on the text of books, but a new test found at least one model has directly memorised nearly the entirety of some books, including *Harry Potter and the Philosopher's Stone*, which could complicate ongoing legal battles over copyright infringement.
https://www.newscientist.com/article/2483352-metas-ai-memorised-books-verbatim-that-could-cost-it-billions/
Source: New Scientist

AI inspired by AlphaFold can predict chromatin structures found in chromosomes
System could help unravel how genome folding influences genes.
https://www.chemistryworld.com/news/ai-inspired-by-alphafold-can-predict-chromatin-structures-found-in-chromosomes/4020964.article
Source: Chemistry World

ProQuest launches AI-powered Ebook Central Research Assistant
ProQuest, part of Clarivate, has introduced the Ebook Central Research Assistant, a new AI tool designed to enhance user engagement with scholarly ebooks.
https://www.knowledgespeak.com/news/proquest-launches-ai-powered-ebook-central-research-assistant/
Source: Knowledgespeak

Web-scraping AI bots cause disruption for Scientific Databases and Journals
Automated programs gathering training data for artificial-intelligence tools are overwhelming academic websites.
https://www.nature.com/articles/d41586-025-01661-4
Source: Nature

CCC launches AI systems training license
Collective license designed to help organisations comply with copyright laws for the use of third-party content.
https://www.researchinformation.info/news/ccc-launches-ai-systems-training-license/
Source: Research Information

How AI is transforming Chemistry Research

New RSC website coming soon

The RSC's main website will be re-launched in mid-June, giving it a fresh look and feel. Some of the content usually accessed will be in a different place, but the overall experience will be simpler and more streamlined.

https://www.rsc.org/news-events/articles/2025/05-may/new-royal-society-of-chemistry-website-coming/

Source: RSC News

Digital Science adds AI-powered summaries to Symplectic Elements to drive research discoverability

Digital Science has announced new enhancements to Symplectic Elements, which will now offer the ability to embed AI-generated summaries for publication abstracts within a researcher's public profile. Symplectic Elements enables the creation of comprehensive public profiles. These profiles are hosted on sleek, modern, and intuitive online portals that offer advanced search and discovery capabilities while ensuring alignment with organisational branding. Profiles can be made available not only for researchers and faculty but also for equipment, services, and other institutional resources.

https://www.digital-science.com/news/digital-science-adds-ai-powered-summaries-to-symplectic-elements/

Source: Digital Science

Wiley AI agent to launch on Amazon

Tool "will demonstrate how researchers can conduct comprehensive full-text scientific literature search".

https://www.researchinformation.info/news/wiley-ai-agent-to-launch-on-amazon/

Source: Research Information

Choosing a flexible Digital Platform for Drug Discovery

At a recent roundtable discussion, hosted by Scientific Computing World, leading drug discovery experts discussed the data impact of working with these multiple modalities, covering areas such as data standards, ontologies, cultural challenges and software needs.

Link

Source: Scientific Computing World

DeepMind unveils 'spectacular' general-purpose science AI

System improves chip designs and tackles unsolved maths problems but has not been rolled out to researchers outside the company.

https://www.nature.com/articles/d41586-025-01523-z

Source: Nature

CCC named recipient of 2025 Artificial Intelligence Excellence Award for copyright licensing framework

CCC has been selected as a winner in the 2025 Artificial Intelligence Excellence Awards for its work on the Annual Copyright License (ACL). The ACL provides enterprise-level licensing for internal use of legally obtained copyrighted materials within AI systems and facilitates rights clearance for using such materials in conjunction with AI technologies. The license framework aggregates internal-use AI rights from a wide spectrum of rightsholders into a unified offering.

https://www.knowledgespeak.com/news/ccc-named-recipient-of-2025-artificial-intelligence-excellence-award-for-copyright-licensing-framework/

Source: Knowledgespeak

ASCO and Google Cloud launch AI tool for faster access to oncology guidelines
The American Society of Clinical Oncology (ASCO) and Google Cloud have announced the launch of the ASCO Guidelines Assistant, an artificial intelligence–powered tool designed to streamline clinician access to ASCO's full library of evidence-based clinical practice guidelines. Built using Google Cloud's Vertex AI platform and Gemini models, the new assistant allows oncology professionals to retrieve guideline recommendations more quickly and interactively.
https://www.knowledgespeak.com/news/asco-and-google-cloud-launch-ai-tool-for-faster-access-to-oncology-guidelines/
Source: Knowledgespeak

AI race in 2025 is tighter than ever before
State of the industry report also shows that 2024 was a breakthrough year for small, sleek models to rival the behemoths.
https://www.nature.com/articles/d41586-025-01033-y
Source: Nature

Speaking their language: Bridging the gap between libraries and users
Stories about engaging and collaborating with library users to communicate the value of the library provides several action ideas.
https://www.infotoday.eu/Articles/Editorial/Featured-Articles/Speaking-their-language-Bridging-the-gap-between-libraries-and-users-%c2%a0-169320.aspx
Source: Information Today

CCC integrates 'Get It Now' article delivery with EBSCO Discovery Service
CCC has expanded its strategic agreement with EBSCO Information Services to integrate its 'Get It Now' article delivery service into the EBSCO Discovery Service™ (EDS) platform.
https://www.knowledgespeak.com/news/ccc-integrates-get-it-now-article-delivery-with-ebsco-discovery-service/
Source: Knowledgespeak

'OMG, did PubMed go dark?' Blackout stokes fears about database's future
A brief outage has focused attention on scientists' reliance on the US-government-funded website.
https://www.nature.com/articles/d41586-025-00674-3
Source: Nature

AI-powered Research Assistance now available in ProQuest Central
ProQuest has announced the integration of its AI-powered tool, ProQuest Research Assistant, into ProQuest Central, one of the world's largest and most widely used research databases.
https://www.knowledgespeak.com/news/ai-powered-research-assistance-now-available-in-proquest-central/
Source: Knowledgespeak

Symplectic Elements adds AI-powered summaries
Enhancement has ability to embed AI-generated summaries for publication abstracts within a researcher's public profile.
https://www.researchinformation.info/news/symplectic-elements-adds-ai-powered-summaries/
Source: Research Information

Insilico Medicine launches Nach01 Foundation model on AWS Marketplace to help accelerate Generative Chemistry Innovation

The biotechnology company driven by generative artificial intelligence, has announced the recent release of its Nach01 foundation model on Amazon Web Services (AWS), available via AWS Marketplace, a curated storefront helping companies of all sizes find, buy, and deploy solutions from AWS Partners. The launch represents a major step in the ongoing collaboration between AWS and Insilico Medicine's Pharma.AI Platform.

https://www.eurekalert.org/news-releases/1086967

Source: AAAS EurekAlert

Invasion of the 'journal snatchers': the firms that buy science publications and turn them rogue

Study finds dozens of journals that have hiked their fees and started churning out papers after being acquired by small, recently formed companies.

https://www.nature.com/articles/d41586-025-01198-6

Source: Nature

Wiley issues position statement on AI content "scraping"

Developers and companies must obtain authorisation before using publisher's content for AI development, training, or implementation.

https://www.researchinformation.info/news/wiley-issues-position-statement-on-ai-content-scraping/

Source: Research Information

Biggest-ever AI Biology Model writes DNA on Demand

An artificial-intelligence network trained on a vast trove of sequence data is a step towards designing completely new genomes.

https://www.nature.com/articles/d41586-025-00531-3

Source: Nature

Clarivate will stop counting citations to retracted articles in JIF (Journal Impact Factor)

Starting with the 2025 Journal Citation Reports release, Clarivate will exclude citations to and from retracted content when calculating the JIF.

https://www.infotoday.eu/Articles/Editorial/Featured-Articles/Clarivate-will-stop-counting-citations-to-retracted-articles-in-JIF-Journal-Impact-Factor-169778.aspx

Source: Information Today