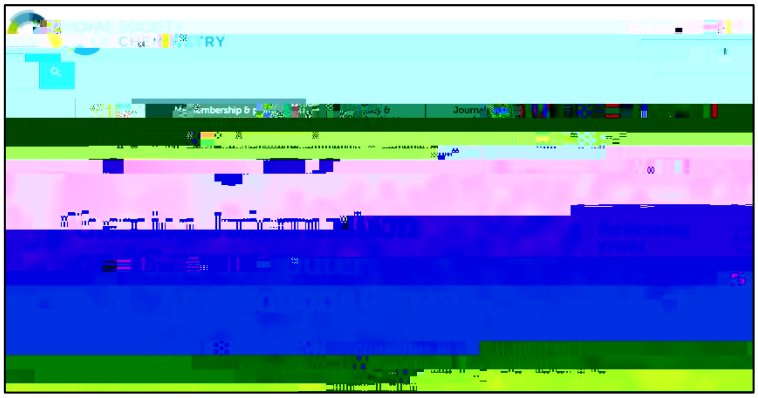
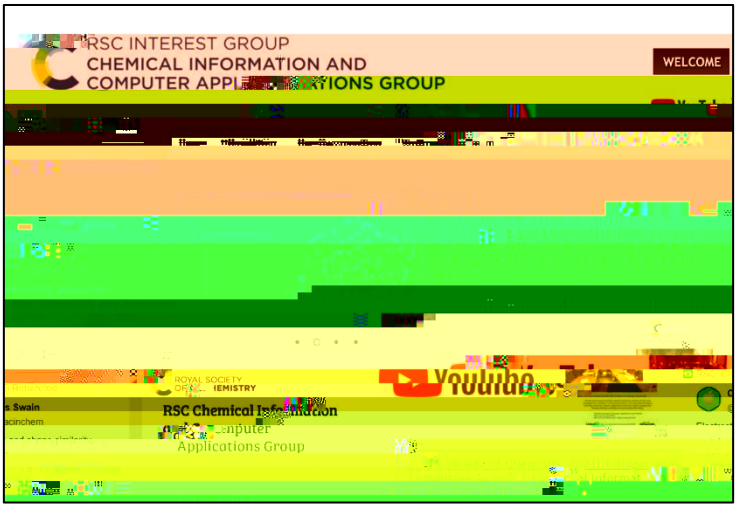
NEWSLETTER

Summer 2022

CICAG Websites and Social Media



http://www.rsc.org/CICAG



http://www.rsccicag.org

https://www.youtube.com/c/RSCCICAG

https://www.linkedin.com/groups/1989945/

@RSC_CICAG

https://twitter.com/RSC_CICAG

Contents

Your CICAG Committee - Introducing Our New Members

CICAG Committee members present at our Zoom meeting on 24 June 2022 (Nathan Brown and Diana Leitch were unable to attend on this occasion).

In June 2022 we were delighted to welcome three new members to the CICAG Committee.

Nessa Carson was born in Warrington, UK. She received her MChem degree from Oxford University, before completing postgraduate studies in catalysis and organic methodology at the University of Illinois at Urbana-Champaign. After her studies, she started in industry at AMRI, initially as a synthetic chemist for AMRI, and then running the high-throughput automation facility on behalf of Eli Lilly in Windlesham, UK, working across both discovery and process chemistry. After this, she moved to process development using high-throughput reaction optimisation and other automation-based experimentation at Pfizer. Nessa started at Syngenta in 2020, where she works with automation, reaction optimisation, and data management, currently holding the title of Automated Data Workflow Specialist. In 2021, she was the recipient of the Salters' Institute Centenary Award for early career chemists with the potential to make a long-term contribution to indur7-8(ith)-5( )uti(s)-9(TQq0.000008871 0 595.32 8

Willem van Hoorn had a misfire at the beginning of his career when he realised halfway through his chemical engineering degree in the Netherlands that he had chosen the wrong subject, at which time switching was no longer possible. After a PhD in computational chemistry studying organic molecules in organic solvents and a postdoc doing Monte Carlo simulations on similar systems, he joined Pfizer Sandwich in 1999 as a computational chemist not knowing the 20 natural amino acids. At that time data sets generated by combinatorial library design and high-throughput screening became large enough to overwhelm Excel but cheminformatics in the guise of Pipeline Pilot came to the rescue. This finally was the beginning of a career. After ten years at Pfizer and a brief period at Accelrys/Biovia he joined Exscientia in 2013 working on cheminformatics-based tools and active learning. Willem still has a cheat sheet with the 20 amino acids!

- - - - - - - - - - -

Free Workshops on Open-Source Tools for Chemistry

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

All workshops are recorded and can be viewed on YouTube.

PDBe Knowledge Base

David Armstrong, EMBL-EBI, Cambridge, UK

This workshop explores the Protein Data Bank in Europe Knowledge Base resource and its tools for the investigation, analysis, and interpretation of biomacromolecular structures. PDBe-KB brings together data from all PDB entries and displays this as aggregated information for individual proteins, including ligand binding sites, macromolecular interactions and more. Furthermore, this community-led resource brings together structural and functional information from a host of other related resources.

In this workshop, you will learn how to use the PDBe-KB aggregated views for proteins to investigate structural and function information for proteins and their associated ligands. We will also demonstrate effective use of novel visualisation components of large-scale structural data on these pages, including 3D visualisation of superposed protein structures with their bound ligands.

KLIFS a ki

Albert Je Kooistra, Copen n University, *and Andrea Volkamer, Charité-Universitätsmedizin, Berlin*

KLIFS is hase database that dissects erimen str f cata se domains kinase inhibitors act with them. The KL structu lignmen les the parison all structu and ligands to ea over, the K FS resi numberin heme cap ng the catalytic cleft 85 residues enables the comp of the raction tterns of ki e inhibitors r examp cial interactions determining kina nhibit electivi The works will be in segm 1) an intro ion to KLIF d 2) programmati cess a pplica s of KLIFS

ESP-Sim shap ostatics map g

Esther Heid, Technical University of Vienna

Electrostatic effects along with volume restrictions play a major role in enzyme and receptor recognition. Evaluating electrostatic and shape similarities of pairs of molecules such as proposed versus known ligands can therefore be valuable indicators of prospective binding affinities. This workshop will demonstrate how to compute electrostatic and shape similarities using the open-source tool ESP-Sim. Available options for comparing electrostatics 0(st)3(a)-5(tic)-302(a)-5(n)4(d)-15( )-303(sh)6(a)-5(pe)-3( )-303(si)4(m)3(il)4(a)-5(rit)-11(ies)-302(usi

With $1M of philanthropic donations and bootstrapped academic grants, and in 18 months, COVID Moonshot built a community resource comprising:

>500 ligand-bound X-ray structures
>10,000 assay measurements
>2,400 synthesised compounds
Preclinical candidates that are in IND-enabling studies funded by a $11M grant from the Wellcome Trust

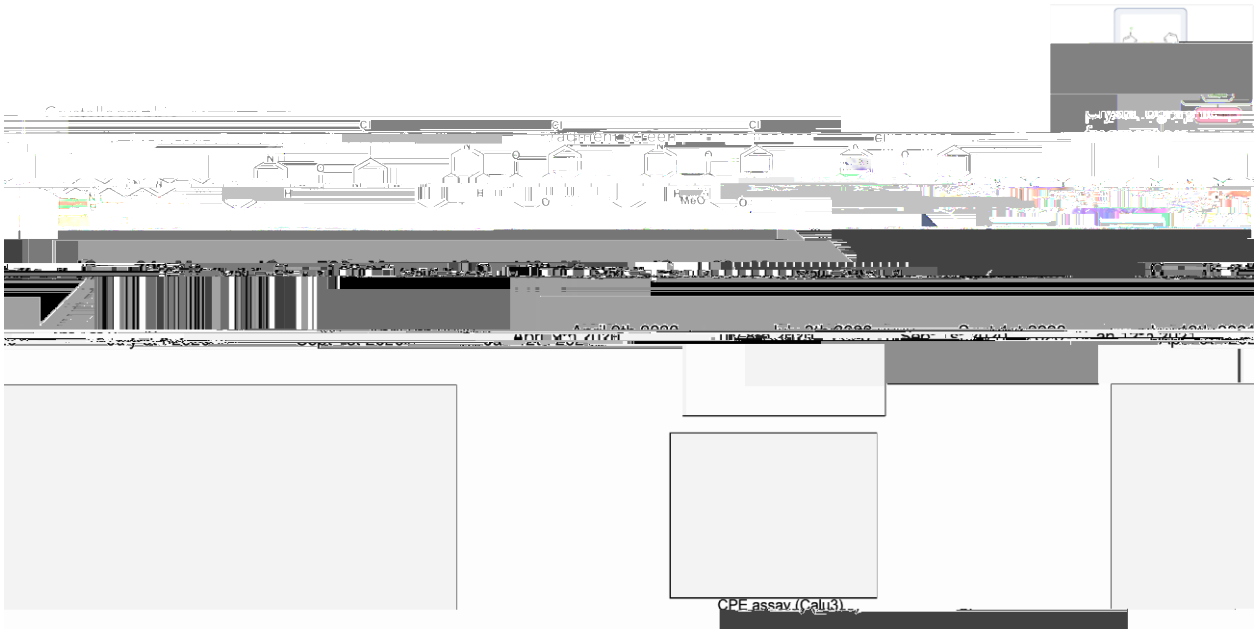
Our starting point was a rapid crystallographic fragment screen that assessed 1,495 fragment-soaked crystals screened within weeks to identify 78 hits that densely populated the active site. This dataset was posted online on 18 March 2020, only days after the screen was completed. The non-covalent fragment hits did not show detectable inhibition. However, they provided a high-resolution map of key interactions that optimised compounds may exploit to inhibit Mpro.

Crowd sourcing drug discovery

We launched an online crowd-sourcing platform on 18 March 2020 inviting participants to submit compounds designed based on the fragment hits. Data from biochemical assays and X-ray crystallography were released rapidly on the same platform, enabling contributing designers to build on all available data, as well as designs contributed by others.

To ensure there would be no delays in ultimately delivering potential drug candidates straight to generics manufacture due to IP licensing issues, all designers were asked to contribute their designs directly into the public domain. Every design and related experimental data were immediately disclosed online and made openly available explicitly free of IP restrictions.

This aggressive open-science policy enabled contributors from multiple fields in both academia and industry to share their ideas freely. Within the first week, we received over 2,000 submissions, representing a diverse set of design strategies. Our initial hit-to-lead strategy focused on compact ligand-efficient designs, heavily triaging based on synthetic complexity forecasted by machine-learning algorithms (detailed below).



Gratifyingly, many submissions exploited spatially overlapping fragment hits. The submission TRY-UNI-714a760b-6 was inspired by five overlapping fragments, furnishing a non-covalent inhibitor with a SARS-CoV-2 Mpro enzymatic IC_{50} of 24 µM. The next potency jump was a design based on substituting the pyridine with an isoquinoline. Following that, rigidifying the scaffold with a benzopyran further increased potency, wit 379.1

An example Manifold search of an initial hit on the Moonshot project.

AI-driven Structure-enabled Antiviral Platform (ASAP)

Extending the impact of COVID Moonshot, we are delighted to launch the AI-driven Structure-enabled Antiviral Platform (ASAP). The ASAP pandemic-preparedness platform aims to discover globally and equitably accessible antivirals against coronaviruses, picornaviruses (including potentially debilitating enteroviruses as well other cold-causing viruses), and flaviviruses (responsible for endemic diseases such as dengue and zika). Launched with $68M of initial funding from the US National Institutes of Health over three years, this five-year programme aims to deliver three IND-ready assets, six optimised leads, and chemical probes for at least nine viral targets.

ASAP aims to tackle pandemics by targeting two salient processes: mutations, by which viruses mutate into more virulent and pathogenic strains, either in human or animal reservoirs, and transmission, whereby dominant strains spread uncontrollably within human populations. These two forces drive a vicious cycle: increased transmission drives more infections, creating a greater risk of mutations due to intrinsic viral mutation rates, thereby leading to the emergence of more fit and virulent strains.

To address mutations, we identify viral targets that cannot easily evolve resistance to small molecule antivirals generated by our platform, using three complementary strategies: (1) phylogenetic analysis of circulating strains to identify conserved sites across members of a viral family; (2) deep mutational scanning (DMS) to interrogate the fitness cost of mutations in a druggable site; and (3) mechanistic analysis of the viral lifecycle to identify 'dominant targets' where the liganded targets disrupt viral growth and cannot be rescued by mutations arising within the same cell.

Rapid discovery of suitable inhibitors against these targets is enabled by high-throughput structural biology, which illuminates druggable sites with dense fragment maps that can be synthesised into novel potent chemotypes where target engagement is restricted to regions where mutations would compromise fitness. Machine learning is then used to execute structure-based design to drive lead optimisation and rapidly arrive at candidate molecules, whilst simultaneously delivering chemical probes to validate target biology and verify the fitness costs of engaging each target.

To address transmission, we will use an open-science approach to drug discovery, developing new antivirals in a manner that enabl

Chemical Data Recovery 3: Legacy Chemical Data Recovery

Contribution from Kevin Theisen, President, iChemLabs, email: kevin@ichemlabs.com

This article is the third part of a three-part series on chemical data recovery written by Kevin Theisen, President of iChemLabs:

1. Embedded Chemical Data Recovery
2. Chemical Image Recovery
3. Legacy Chemical Data Recovery

Figure 1. A chem-archaeologist has discovered an ancient library filled with long-lost molecular secrets. She transcribes some of the information using a chemist's triangle.

Introduction

Cheminformatics solutions can be incredibly challenging to implement. What this really means is cheminformatics problems are incredibly rewarding to solve. While cheminformatics work is difficult, such solutions are very important to our scientists and our society. As we discover more about the universe we exist in, the already impressive work created by those in the cheminformatics field only grows in usefulness.

Of core importance in cheminformatics is the actual storage and communication of chemical information. The

novel substructure searching based on WLN strings. The ISI was heavily invested in the WLN protocol and provided several programs and searching solutions, including the Index Chemicus Registry System for recording chemical data in the Index Chemicus as WLN codes and the popular CROSSBOW program for handling WLN strings as computer data structures.

The International Union of Pure and Applied Chemistry (IUPAC) even considered WLN as their standard line notation, before choosing the competing Dyson notation instead. The decision was very controversial and resulted in a lot of protest. Bonnie Lawlor from the Chemical Structure Association (CSA) Trust (the CSA Trust evolved from the CNA) summarises the importance of WLN and IUPAC's choice in a 2016 article. Neither notation is maintained by IUPAC today.

In a 1982 publication, Wiswesser postulated what might become of WLN in the future, "Soon online computer and word-processing terminals will be as commonplace as IBM

It may seem logical that the introduction of the much more readily implementable Daylight SMILES line notation in 1988 led to the replacement of WLN, but in reality, WLN had already fallen out of favour as more practical methods for storing chemical information were developed for the many computer systems introduced. The MDL connection tables, *circa* 1979, are ASCII formats many programmers could easily use, and Mike Elder produced the DARING software to help aid in the conversion from WLN to MDL connection tables. New developments in computer algorithms were also complicit in the demise of WLN. J. R. Ullman published an algorithm for graph isomorphism in 1976, enabling cheminformatics applications to directly and efficiently match parts of chemical structures based on the constituent atoms and bonds. Granito's Chemical Substructure Index was no longer the optimal solution and WLN was losing popularity by the time CIMI published the 3rd edition; Granito would soon change business directions.

Wiswesser would pass away in 1989, leaving one of the most impactful and impressive chemistry protocols ever created as his legacy. To date, WLN is still the most concise, lossless, string representation of chemical information. The WLN protocol is a passion project of a talented group of cheminformatics experts, and a work I hold in very high regard.

A breakdown of Wiswesser Line Notation

WLN is a substructure-based, canonical, line notation for molecular structure(s). The characters in a WLN string define the atoms and bonds in the molecular structure(s). The entirety of the periodic table of elements is supported, and single, double, triple and dative bond types are available. Any type of complex ring system is compatible, including polycyclic fused, perifused, spiros, bridged and pseudo-bridged structures. There is no

I have to admit, implementing WLN is much more difficult than I anticipated. There is a significant level of detail concerning each aspect of chemical structures.

why SMILES is not as simple as it appears. InChI, while very complex, is an open standard and the software to handle it is funded and open source. IUPAC naming is the most massive undertaking, by far, but that is a whole other complex discussion.

Regardless, I enjoyed my time with WLN. It is a unique perspective on chemistry data and I hope to return to find a better algorithm for the complex WLN ring systems. If you have the time, please learn about it, try to implement it, and teach it to your students and colleagues. If you are interested in cheminformatics, writing a WLN parser or writer is an incredibly challenging project providing you with very thorough experience into the concepts of chemical structure and graph theory. It will certainly be an impressive statement in your portfolio.

Preservation

We may now understand why such an important chemistry protocol, which was widely adopted in the mid-1900s, is no longer known today. The importance of preserving the

Documentation, published its first articles in 1961.[5] For the purposes of this article I shall take the turn of the century as the cut-off point since by then the discipline had become sufficiently well established for the introduction in 2000 of the first specialist masters course in cheminformatics at the University of Sheffield, with others at the University of Manchester Institute of Science and Technology and at Indiana University starting in the following year.[6]

Research into cheminformatics started in Sheffield in 1965 with the appointment of Mike Lynch to a position in what was then called the Postgraduate School of Librarianship. Most academic research in cheminformatics over the years has, not surprisingly, been carried out in chemistry (or,

automatic indexing of chemical reactions and for suggesting novel synthetic pathways. During the year he was in Sheffield (after which he went to work for the Institute for Scientific Information in the USA) he returned to this early work and devised an approach to reaction indexing based on a maximal common subgraph

patterns of atoms and bonds provided an efficient screening mechanism in 2D substructure searching, followed by a subgraph isomorphism check for an exact match with the atoms and bonds comprising the query substructure. Focussing on the distances separating atoms, rather than the bonds linking them together, provided a natural way of adapting these 2D techniques to permit the introduction of operational systems for 3D searching, with subsequent developments based on smoothed bounded distances permitting the extension of these ideas to encompass the conformational flexibility that characterises many small molecules of biological interest. Other research in the 1980s and 1990s involved work on molecular diversity, genetic algorithms for molecular docking and pharmacophore mapping, and the application of small-molecule graph matching algorithms to the analysis of biological macromolecules *inter alia*.[19,20]

In conclusion, it is perhaps not unreasonable to suggest that the research in Sheffield has played a significant role in the development of cheminformatics – most obviously in ways of processing the databases that increasingly underlie so much research in modern chemistry – and these studies have continued[21] since the early days that have been reviewed here.

(1) Brown, F.K. Chemoinformatics: what is it and how does it impact drug discovery? *Annual Reports in Medicinal Chemistry.* 1998, *33*, 375-384.
(2) Hann, M.; Green, R. Chemoinformatics: a new name for an old problem? *Current Opinion in Chemical Biology.* 1999, *3*, 379-383.
(3) Ray, L.C.; Kirsch, R.A. Finding chemical records by digital computers. *Science.* 1957, *126*, 814-819.
(4) Hansch, C. *et al*. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature.* 1962*, 194*, 178-180.
(5) Willett, P. From chemical documentation to chemoinformatics: 50 years of chemical information science. *Journal of Information Science.* 2008, *34*, 477-499.
(6) Schofield, H. *et al.* Recent developments in chemoinformatics education. *Drug Discovery Today.* 2001, *6*, 931-934.
(7) Lynch, M.F. *et al. Computer Handling of Chemical Structure Information*. Macdonald, 1971.
(8) Lynch, M.F.; Willett, P. Information retrieval research in the Department of Information Studies, University of Sheffield: 1965-1985. *Journal of Information Science*. 1987, *13*, 221-234.
(9) Lynch, M.F. Variety generation—a reinterpretation of Shannon's mathematical theory of communication, and its implications for information science. *Journal of the American Society for Information Science*. 1977, *28* , 19-25.
(10) Lynch, M.F.; Holliday, J.D. The Sheffield generic structures project: a retrospective review. *Journal of Chemical Information and Computer Sciences.* 1996, *36*, 930-936.
(11) Lynch, M.F.; Willett, P. George Vladutz, 1928-1990. *Journal of Chemical Information and Computer Sciences.* 1990*, 30*, 349.
(12) Willett, P. Textual and chemical information processing: different domains but similar algorithms.
.
(13) Willett, P. *Similarity and Clustering in Chemical Information Systems.* Research Studies Press, 1987.
(14) Robertson, S.E.; Sparck Jones, K. Relevance weighting of search terms*. Journal of the American Society for Information Science and Technology.*

UKeiG Call for Nominations for the Prestigious Tony Kent Strix Award 2022

Contribution from Gary Horrocks, UKeiG, CILIP,

Nominations for the 2022 award must reach the judges by 6 pm GMT on Friday 30 September 2022. Please email to: John Wickenden secretary.ukeig@cilip.org.uk (Hon. Secretary UkeiG), and copy in Gary Horrocks info.ukeig@cilip.org.uk (UKeiG administrator) and Sue Silcocks treasurer.ukeig@cilip.org.uk (Hon. Treasurer UKeiG).

For more information about UKeiG, the Tony Kent Strix Award and previous winners of this prestigious international award is available on the CILIP website.

A video of UKeiG's 7th Tony Kent Strix Annual Memorial Lecture 2021 – delivered by the 2020 Strix award winner Ian Ruthven, Professor of Information Seeking and Retrieval at the Department of Computer and Information Sciences, University of Strathclyde – is available. The Award was presented in recognition Professor Ruthven's outstanding practical innovation and achievements in the field of information retrieval.

Professor Ruthven's lecture was entitled: Go

In a sense, this scientific knowledge cloud has been a big success, a0ktbit

of the best system of 0.8672, and then the recognised chemicals are linked to the corresponding formal identifiers, with the best F-score being 0.8136:

A very rough back-of-the-envelope calculation can give us an estimate of the quality of mining such entire relations:

An overall F-score of 0.40, as

entitled

that way, but they should also come with representations in formal logic for anyone or anything that knows how to deal with that. For that special issue, we chose the journal Data Science, of which I am an editor-in-chief.

We also had to make a practical concession though: while the whole setup *could* be used to publish novel findings, we restricted ourselves to findings from existing publications. For that, we introduced the concept of a 'formalisation paper' whose novel contribution is the formalisation of an existing finding. So, authors of a formalisation paper take credit for the formalisation of the finding, but not for the finding itself.

We ended up with 15 formalisation papers in our special issue, as summarised by this table:

Authors	CONTEXT	SUBJECT

Greg Landrum Receives the Mike Lynch Award

Contribution from Professor Jonathan Goodman, Yusuf Hamied Department of Chemistry, University of Cambridge, email: jmg11@cam.ac.uk

The Mike Lynch Award recognises and encourages 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.

The Trustees of the CSA Trust awarded the 2022 Award to Greg Landrum in recognition of his work on the development of RDKit and his fostering of the community around it, a transformative software resource for cheminformatics and machine learning. The Award was presented at the 12th International Conference on Chemical Structures (ICCS), Noordwijkerhout, in June 2022.

Greg is a senior scientist in Sereina Riniker's group at the ETH Zurich, Founder and Managing Director of T5 Informatics GmbH, a Senior Advisor to Knime, and the primary developer for the RDKit.

Jonathan Goodman, chair of the CSA Trust, commented: "I am delighted that Greg Landrum has accepted this award. His work on RDKit has made chemical informatics techniques more accessible to scientists worldwide both in industry and academia. When introducing students to cheminformatics, becoming familiar with RDKit is a key part of the learning process, and makes it possible to explore new ideas in chemical information rapidly and reliably."

Greg Landrum, who gave a keynote address at the ICCS, said: "I am really honoured to have been selected for this award; it's especially meaningful to me because of the foundational importance of Mike Lynch and the '

it has become commonplace to use statistical approaches exploiting large databases of evolutionary information to predict which pairs of amino acids were in contact. However, one thing was clear: except for some special cases, we were not very good at predicting the structure of proteins.

Then, out of the blue, the algorithm presented by DeepMind at the fourteenth edition of CASP beat every other group, and then some. Their methodology introduced a large number of novel ideas, working in unison to produce predictions of unprecedented accuracy. For example, rather than using a costly exploratory algorithm, like simulated annealing, AlphaFold 2 predicts a structure 'end-to-end' on a single shot. The model takes a protein sequence (well, a multiple sequence alignment) on one end, and outputs a list of coordinates on the other. Borrowing from the well-established ideas in the field, they also used their machine-learning expertise to extract as much information as possible from evolutionary databases. Finally, and what perhaps has had

vast differences in binding affinity, a phenomenon known as ‘activity cliff’. In contrast, AlphaFold 2 is not even very good at predicting the effect of a single amino acid mutation.

There are important industrial players forming in this quest. Unless you have lived blind to the world for the past years, you must be aware that Demis Hassabis, the CEO of DeepMind, has embarked on a new venture, Isomorphic Labs. The project – “reimagining the entire drug discovery process from first principles with an AI-first approach” – is fazed with the standard secretism, but few doubt that predicting protein-ligand figures amongst their objectives. Another competitor, Charm Therapeutics, featuring David Baker, recently raised a $50M series A round to take Isomorphic Labs in what may be one of the most interesting technological competitions of this decade.

My second bet is that the battle over protein-ligand interactions will not be fought over computational methods, but the availability of data. The AlphaFold breakthrough was possible because over 170k protein structures had been painstakingly solved by structural biologists, at an estimated aggregated cost of ~$10 billion. Unfortunately, data on protein-ligand interactions is much more scarce – and, for the increased complexity discussed above, probably several more orders of magnitude would be required to train a model that rivals AlphaFold’s complexity. Furthermore, much of this data is held in the internal databases of pharmaceutical companies which, for good reasons, will protect them fiercely to guarantee their intellectual property.

There is a deeper problem: drug-discovery datasets are plagued by biases. Think about it: most of our best data comes from the hands of incredibly talented medicinal chemists who have to work hard for every datapoint, and thus invest significant thought to design molecules with a serious chance of binding. These chemis

function." Easy structure prediction has been a powerful step in this direction: by predicting the structure of a protein and comparing against known structures, it is possible to elucidate function in some cases. In a way, searching structure databases using predictions may well become a substitute to sequence search methods like the BLAST family.

What is certain is that we are likely to witness an era of flourishing computational tools that enable deeper insight into interesting problems, as well as exciting applications throughout the life sciences.

4. What needs to happen (the rant section)

Many good things have come from AlphaFold 2 beyond scientific progress and new ideas. Perhaps the most important one is openness. In computational biology, a field, where advances were sometimes kept private until the following CASP exercise, it now seems like the floodgates have truly been opened. Ideas have flown

Diana Leitch –

At the age of 16, my love of the world of chemical information started. I spent my summer vacation helping out in the Information Department at ICI's R&D Division at the Runcorn Heath complex and whilst there came under the influence of the work of John Wales, the Information Manager, and Angela Haygarth-Jackson, at ICI Pharmaceuticals Division, who became a lifelong inspiration to me of what women could achieve in a very male-dominated world. Chemical information was almost the only way that women could become senior managers in ICI.

When I was 18 in 1965, the magic of academic chemistry beckoned and I went to study chemistry at Edinburgh University.

On 6 July 1973 I turned up for work in the Christie Science Library. I was given a manual typewriter and told to catalogue and classify books. Not quite what I expected as I had never classified a book in my life or done a library training course but it was another formative experience! I took over from another chemist, Dr Alan Neville (who sadly died in 2020), who had been moved to the Medical Library to work on the first experiments on the use of digitised *Index Medicus* (the Medlars system). I found myself in good company. Also working in the John Rylands University Library (JRUL) of Manchester at the time were Bill Simpson, Chris Hunt, Reg Carr, John Hall, John Henshall, John Lancaster, John Tuck and Ian Lovecy, who all went on to become University Librarians and senior ambassadors for the library profession in the UK and abroad. 'Information' was not a word that was allowed to be used in the overall Library at that time and neither was training in information usage for students and staff. I spent the next four years as a cataloguer and then in 1977 I found that I was pregnant. I asked to take maternity leave but was told this would not be possible. I needed a job as my husband, also a research chemist from Edinburgh, had been made redundant from ICI Dyestuffs Division at Blackley in

these could be moved to storage in a cotton mill in Stockport. I oversaw the whole operation. Over 50 miles of materials were moved around the whole JRUL system in the two-and-a-half years to provide space to maintain

OneThree Biotech is a start-up spun out of Cornell by Neel Madhukar and three other co-founders in 2017. Neel explained how his company is using AI in their pipeline to de-risk the earliest stage of the drug-discovery pipe that often hampers clinical success, the discovery of novel disease biology and target identification.

Our last talk was by Daniel of Biorelate, a company which he spun out of the University of Manchester in 2013. Daniel's talk titled *Using Cause-and-Effect to Empower Drug Discovery* described Galactic AITM, a supercomputing platform that automatically curates biomedical research literature to distinguish simple correlation from casual therapeutic pathways. By connecting obfuscated evidence, Biorelate's goal is to accelerate development of important new therapies.

Start-up panel

I knew this section of the conference would be the one of the best the second we all joined the private Zoom lobby before going live; I hardly needed to encourage the Founders to bounce ideas off of each other and the mood, even though virtual, was electric. In attendance were Rabia, Neel and Dan from before as well as another panelist, Laksh Aithani. Laksh is the CEO of CHARM Therapeutics, a start-up he co-founded along with the world-famous protein designer David Baker.

During the panel, we touched on all the usual pillars of entrepreneurship, such as finding the right co-founder, raising funds and recruiting the right people into your team early on. Of particular interest was the discussion about how there has been a recent shift of AI in drug-discovery start-ups away from simply being the AI partner of a traditional pharma company and instead developing their own 'end-to-end' drug-discovery capabilities. Indeed, many of our AI co-founders were actively building out their own wet lab capabilities. This trend is exemplified by companies like Exscientia and BenevolentAI going public with very successful IPOs based on their AI-driven discovery platform and current in-house drug-development portfolio.

Since the event, the Founders have been leading their companies from success to success. A recent example is with CHARM Therapeutics, which officially came out of stealth mode and announced a $50M Series A round co-led by F-Prime Capital and OrbiMed at the time of writing. Finally, all our panellists were keen to point out that they are all actively hiring!

Science panel

The day ended with a final panel discussion on the *Scientific Frontiers of AI and Drug Discovery*. The discussion was aimed at being highly interdisciplinary and our panellists reflected this. In attendance was Michael Bronstein and two excellent complementary guests. The first was Nathan Benaich who is the Founder and General Partner of Air Street Capital, a London-based VC firm specialising in making investments into AI and Life Science companies (including LabGenius, Exscientia and Valence Discovery) and is also the Founder of London.AI, Spinout.fyi and the RAAIS Foundation. He is not just a businessman however, Nathan holds a PhD in computational biology from the University of Cambridge, is co-author in the acclaimed *State of AI Report* and moves fluently between discussions on biology, AI, business and geopolitics.

Our final panellist for the evening was Sir Tom Blundell. Sir Tom is a highly interdisciplinary biochemist, structural biologist, computational biologist and science administrator. He is currently Professor Emeritus of Biochemistry at Cambridge and holds an impressive list of achievements. These included: being part of the team that solved the first structure of insulin, founding Astex pharmaceuticals (which successfully commercialised fragment-based drug d

The discussion started with each of the guests describing what they felt was the main challenge for making AI have a real-world impact on drug discovery and biology more generally. Answers given ranged from data availability and quality, current model architecture and slow organisational embracement of AI technology. Tom was also keen to point out that there is currently a parallel revolution going on within drug discovery alongside AI and that is Cryo Electron Microscopy (CryoEM). He went on to describe his (at the time) radical adoption of CryoEM into Astex and that better integration of ML and CyroEM is an increasingly important area of research.

Conclusion

The majority of talks are available online via the CuAI YouTube channel, however, the panel discussions were not recorded to encourage openness in the discussions. If you were not able to make it, however, there is no need to worry! We are already planning a second conference of the same name next year which will likely be in person and have an even greater emphasis on interdisciplinary discussions between science, business and governmental experts. More information will be on the CamAIDD websit-5(ve)-3( )-62(a)-5(n)4( )-62(ev)-3(en)-61(gre)-3(a)-

Upcoming events

We are running two hybrid events in July 2022.

Failed it to Nailed it: Nailing your data visualisation: a hands-on training workshop – 13-14 July 2022

Where: This is a hybrid meeting, the physical sessions will take place at the Chilworth Manor Hotel in Southampton. Virtual attendees can join via Zoom.

What: This event forms part of the *Failed it to Nailed it* series run by the Artificial Intelligence for Scientific Discovery Network+ (AI3SD), the *Cell Press Patterns Journal* and the Physical Sciences Data-Science Service (PSDS). This event is the second in our 2022 edition of the *Failed it to Nailed it* series and is designed to t3.96 Tf 4 0 0 1 400.78 121.34 Tm0.133 g0.133 G[(a)-5(n)4(d)-2( )-50(th)4(e)-53(P)-9(h)-7(ys)hu.pflu0.1

that delve into some of the tools and techniques that can be used for this and a number of challenges to choose from. Feel free to come in a team or come solo and we will match you up. There will be helpers on hand if anyone needs advice.

Practical requirements: This is a hands-on workshop so after the talks there will be some team-based practical activity. You will be expected to bring your own device. If this is not possible please let us know when you register.

AI4SD & Directed Assembly Early Career Researchers Event – 25-26 July 2022

Where: This is a hybrid meeting, the physical sessions will take place at the Chilworth Manor Hotel in Southampton. Virtual attendees can attend via Zoom.

What: This is a joint event between the AI4SD and Directed Assembly networks. It is specifically designed to inform, upskill and facilitate networking between early career researchers. The event will contain talks on scientific publishing, ED&I (equality, diversity & inclusion), grant and fellowship applications, networking and much more. There will also be dedicated time for networking.

About AI4SD: The AI4SD Network+ is funded by EPSRC and hosted by the University of Southampton. It aims to bring together researchers looking to show how cutting-edge artificial and augmented intelligence technologies can be used to push the boundaries of scientific discovery.

About Directed Assembly: We aim to gain unprecedented control of the assembly of molecules and structures that are the building blocks of many functional materials, consumer and industrial products.

possible about the content of the Green Book. The 4th edition which has been updated with the new SI will be published soon along with an abridged edition but we are considering more major changes for the 5th edition.

Please fill in our survey to help us gather community views on what should be included in the 5th edition. The survey will remain open until the end of the summer but early replies will be very useful in planning the 5th edition project.

We have received full ethics approval from the University of Southampton Ethics and Research Governance Team to run this survey under ERGO No. 72139. Please read the following participant information to make sure that you understand and agree to the terms of this study.

To contribute to this discussion please go to the Survey.

RSC Historical Group – Women in Chemistry Symposium

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

200 years ago there were few opportunities for women to study or practise any form of science. By the late 19th century, the door was beginning to open for women to study for a university degree. Since the beginning of the 20th century there has been a huge change in the career opportunities available to women.

This one-day symposium organised by the RSC's Historical Group will take place on 13 October 2022

ACS CINF Report for July 2022

Contribution from Sue Cardinal, 2022 CINF Chair

CAS Formulus

2022 is off to a great start with CAS Formulus. The year started by focusing on improving the search experience for our users. This included making enhancements to autosuggest, improving controlled vocabulary handling, and incorporating supplier trade names as searchable ingredient identifiers.

In addition to improving the search experience, CAS is investing in expanding our formulations content. CAS Formulus now includes expanded surfactant formulations, covering products such as detergents, hard surface cleaners, and consumer products. Throughout this year, we will also include formulations from coatings, inks, and paints.

Looking ahead for the rest of 2022, CAS Formulus will offer the ability to save items of interest as well as setting alerts on search queries. These key features will offer better workflow support to our formulators. Another upcoming aspect of workflow support is expanded exporting capabilities, including the ability to export information such as *Commonly Formulated With…*, *Commonly Used As…*, and formulation-centric regulatory data.

STN IP Protection Suite

CAS provides IP solutions and services while seeking innovative ways to identify and navigate the IP landscape. Our CAS STNext has AI Aating sup

BioWorld by Clarivate Identifies the Top Six Global Biopharmaceutical Stories and Trends of 2021
Advances in Alzheimer's, innovation in DNA vaccines, increased regulatory collaboration and the evolving impact of Artificial Intelligence, join Covid-19 as key stories and trends of last year.
https://clarivate.com/news/bioworld-by-clarivate-identifies-the-top-six-global-biopharmaceutical-stories-and-trends-of-2021/
Source: Clarivate

Benefits of Semantic Enrichment Across the Drug Development Pipeline
Discussing practical use cases using semantic enrichment technology in various stages of drug development to show the benefits that leveraging semantic search can bring to the table.
https://www.copyright.com/wp-content/uploads/2021/09/CCC-SciBite-White-Paper-Benefits-of-Semantic-Enrichment-Across-the-Drug-Development-Pipeline.pdf
Source: CCC

Nature Masterclasses Online Training made free to access for Researchers in Lower Income Countries
Nature Masterclasses offers highly targeted courses designed to enhance the skills and boost the confidence of early career researchers, with over 60,000 researchers globally having benefited from the platform to date. The world-class training is delivered by editors from across the Nature Portfolio of journals as well as experienced researchers, funders and professionals via a subscription.
https://www.stm-publishing.com/nature-masterclasses-online-training-made-free-to-access426.94 482.83 Tm0 0 1 RG[(-)] T

Improving International Research Collaboration

CHORUS and ChemRxiv sign MOU to pilot Preprint Dashboard Service

CHORUS and ChemRxiv have signed a one-year Memorandum of Understanding (MOU) to pilot a preprint dashboard service. By using persistent identifiers, CHORUS will create a dashboard for *ChemRxiv* that connects preprints to funders and datasets as well as information related to public accessibility and other key metadata to be added later.

https://www.knowledgespeak.com/news/chorus-and-chemrxiv-sign-mou-to-pilot-preprint-dashboard-service/

Source: Knowledgespeak

Small-Molecule Antiviral could be used in Anti-COVID-19 Nasal Spray

N-0385 is a peptidomimetic compound with the sequence Ms-Gln-Phe-Arg-kbt (Ms = mesyl, Gln = glutamine, Phe = phenylalanine, Arg = arginine, kbt = ketobenzothiazol). The drug targets the enzyme TMPRSS2 (transmembrane protease serine 2), which is important for cell entry.

https://www.chemistryviews.org/details/news/11347216/Small-Molecule_Antiviral_Could_Be_Used_in_Anti-COVID-19_Nasal_Spray/

Source: ChemistryViews

Prototype for Game-Changing Ammonia Plant

Engineers are designing a green ammonia plant that can reliably and efficiently generate ammonia using only intermittent renewable energy as the source of power.

https://www.ukri.org/news/prototype-for-game-changing-ammonia-plant/

Source: UKRI

Elsevier's Reaxys becomes the leading source of Curated Chemistry Patents in Collaboration with LexisNexis' PatentSight

Elsevier has announced its market-leading position in chemistry patent coverage and the extension of its collaboration with *LexisNexis® PatentSight®*. In March 2021, Elsevier launched its initiative to strengthen the existing patent coverage in Reaxys®. The content expansion resulted in a 15-fold increase in patent coverage and ensures pharma and chemical companies and their researchers do not miss key competitive intelligence insights.

https://www.knowledgespeak.com/news/elseviers-reaxys-becomes-the-leading-source-of-curated-chemistry-patents-in-collaboration-with-lexisnexis-patentsight/

Source: Knowledgespeak

How Institutions can better Support their Postdoc Fellows

Competition for fellowships can be fierce, but experiences will vary.

https://www.natureindex.com/news-blog/how-institutions-can-better-support-their-postdoc-fellows

Source: Nature

CAS SciFinder includes unique Biosequence data Collection and Capabilities

CAS has announced the launch of a major expansion of the CAS SciFinder Discovery Platform into life sciences. The enhanced platform includes over 2 million modified biosequences, 60 years of patent literature, and one of the largest collections of journal information including PubMed's biomedical and life science data.

https://www.knowledgespeak.com/news/cas-scifinder-discovery-platform-includes-unique-biosequence-data-collection-and-capabilities/

Source: Knowledgespeak

Institutions partner with ACS to advance first California-wide Transformative OA Agreement

Three California consortia, representing nearly 60 academic and research institutions, and the ACS have announced the first-ever California-wide transformative open access agreement. It is also ACS' first "read and publish" agreement in the U.S. composed of multiple consortia. Through a partnership with the 10-campus University of California (UC) system, the 23-campus California State University (CSU) system, and 25 subscribing institutions represented by the Statewide California Electronic Library Consortium (SCELC), readers and researchers at dozens of California research institutions will be able to benefit from full access to subscription content.

https://www.knowledgespeak.com/news/institutions-partner-with-acs-to-advance-first-california-wide-transformative-oa-agreement/

Source: Knowledgespeak

All UKRI Councils now on one, Integrated Website

UKRI has launched a new integrated website delivering a simpler, more efficient user experience. They have brought all seven research councils, *Innovate UK*, and *Research England* websites into one site, providing a strong