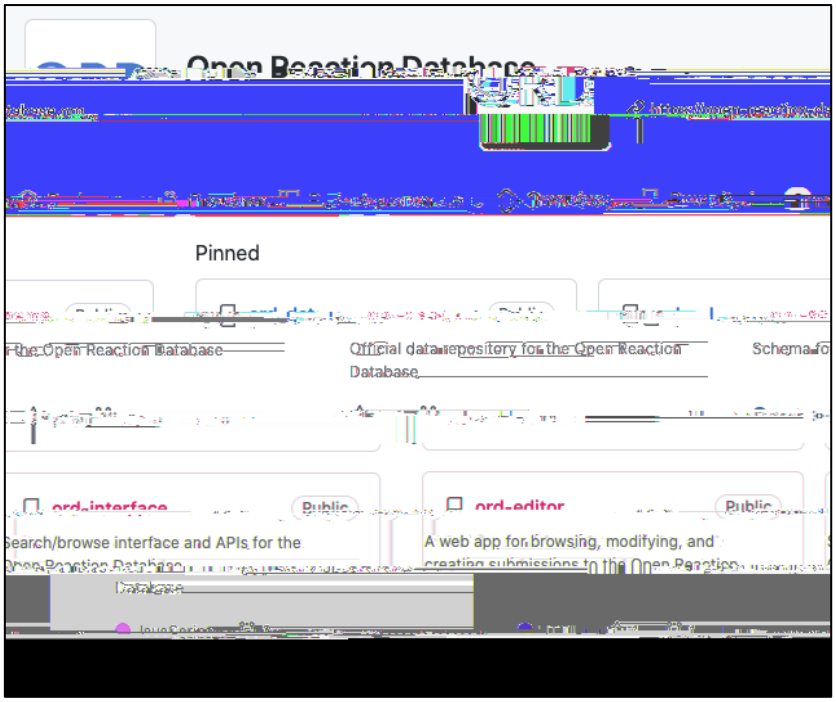
3Ï Ì ÛÈÉÓÌ É Ì ÓÖÞ ×ÙÖÝÐË Ì Ú È ÚÜÔ ÔÈÙà ÖÍ "(" &ȚÚ ×ÓÈÕÕÌ Ë ÈÕË ×ÙÖ×ÖÚÌ Ë ÍÜÛÜÙÌ ÚÊÐÌ ÕÛÐÍÐÊ ÈÕË educational Ô Ì Ì ÛÐÕ Î Ú ß %ÖÙ ÔÖÙÌ ÐÕÍÖÙÔÈÛÐÖÕÛ×Ó Ì ÈÚÌ ÊÖÕÛÈÊÛ "(" &ȚÚ " Ï ÈÐÙÛ # Ù " Ï ÙÐÚ 2ÞÈÐÕ

Meeting	Date	Location	Further Information
Open-Source Software Workshops	2022	Virtual	An ongoing series of workshops.



The infrastructure for the effort is organised into four parts: the schema that describes how chemical reactions are defined, a graphical editor for defining and downloading new reactions, a client for searching/viewing data with a database deployment, and the data itself. All four exist on GitHub under various open-source licences (Apache 2.0 for code, CC BY-SA for the data). All project development is done on GitHub and has been supported by volunteer effort so far, including generous contributions of time and computing resources from Google and Relay Therapeutics.

At present, most procedural details about chemical reactions are reported in unstructured supporting information documents as Word files exported to a PDF. That information, which includes even basic details of quantitative amounts of reactants, is not fully captured by current databasing efforts to our knowledge. The ORD schema is designed to capture the most important fields of a chemical reaction to support downstream , +ш ÛÈÚÒÚß (Ûш ÐÚш Þ ÙÐÛÛÌÕш ÜÚÐÕÎш &ÖÖÎÓÌ¿Úш ×ÙÖÛÖÊÖÓш ÉÜÍÍÌÙш ÛÌÊÏÕÖÓÖÎàÓш ÚÖш ÛÏÈÛш ÛÏÌш ÚÊÏÌÔÈш ÐÛÚÌÓÍш ÊÈÕш ÉÌш ÌÈÚÐÓàш understood by users with little programming familiarity. In developing the schema, our goal was to address the most common types of contributions we expect to receive (based on a survey of roughly 170 users in late 2019). These types include single-step batch reactions as performed by hand or in a high-throughput experimentation workflow.

A snippet from the protocol buffer defining a reaction.

One of the ways to define data is through the graphical user interface, where users may provide all details of their reaction (includin Î ÌÈÛÜÙÌÚÒÒÌÚÛÙÜÊÛÜÙÌËÙÈÞÐÕÎÈÕËÕÈÔÌÓÖÖÒÜ×ÚÆ(ÛÇÚÈÓÚÖÕÈÙÙÈÔÍÖÙÛÏ Ì ËÐÎÐÛÈÓËÈÛÙÌÊÖÙËÛÖÉÌ?ÉÖÙÕËÐÎÐÛÈÓ?ÈÕËÉÌËÌÍÐÕÌËׯÙÖÎÙÈÔÔÈÛÐÊÈÓÓàÅ3ÏÌÚÊÏÌÔÈÊÈÕÉÌÊÖÔ×ÐÓÌËÛÖ / àÛÏÖÕÊÓÈÚÚÌÚ which in combination with numerous helper functions we have written, makes it easy to define hundreds or thousands of results (e.g., from high-throughput screening) programmatically. An enumeration script allows users to combine a template reaction with a spreadsheet to produce a full dataset of many reactions. Example Jupyter notebooks show2 reW00950204≥30110(ugh)6(p)-12(ut)-61(scre)-4(e)shlly. An enundred( )-50(o)4(w)-8(0(n30.0017Qc

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

This article is the second 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

We launched ChemDoodle 2D v11.4 on 2 April 2021. A new chemical image recovery function was included for automatically rebuilding chemical drawings from an image, which this article discusses in detail.

Figure 1. In this laboratory setting, an android is using its ability to see and understand molecule drawings and communicate with a scientist. New chemical image recovery features in ChemDoodle 2D make this future a possibility.

When we communicate as chemists, we often use images of molecules because a picture is the most effective way to communicate information to visual creatures such as us. For well over a century, images of molecules

Chemical image recovery (CIR) is the process of taking an image of a chemical drawing, with no provided information other than the defined pixels, and using a computer to recreate the original chemical data to be used or edited further. For instance, take the following image of galanthamine. The image on the left is the input image, and the image on the right is the result of the CIR function in ChemDoodle 2D.

Figure 2a. An image of the molecule galanthamine.

Figure 2b. The recovered chemical drawing of galanthamine.

The first impression may be, "Great! I now have a less blurry image." Yet, the result is much more significant. The actual chemical data, the arrangement of atoms and bonds, is digitised. We can now further process this information. For instance, we can produce a molecular formula, resolve the CIP stereochemical configurations, and change the graphical style to ACS 1996. We may even optimise the molecular structure in 3D and calculate a distance for the hydrogen bond. All of this output is easily produced from the result of the CIR function on the input image. Without the CIR function, a person would be required to redraw this molecular structure in a program to perform any computational chemistry task. One image is work by itself, imagine having to transcribe thousands of chemical drawings.

Figure 3a. The recovered chemical drawing is further processed; we are able to change styles, resolve stereochemical configurations and produce a molecular formula.

By recreating the chemical data originally lost in images, CIR makes it possible to produce many solutions for scientists. You can have a program automatically catalogue drawings from laboratory notebooks. Students can simply point their camera at a chemical structure on a poster and get the associated IUPAC name. An assistive tool can be produced to help vision-impaired chemists. A researcher can snap a picture of a molecule from a publication he/she is reading and immediately find more information from a chemical search engine. We may even be able to produce androids for our labs with the ability to observe and understand chemical drawings and then complement scientists so they can get their work done faster. The android could also protect the chemist if safety becomes a concern.

Figure 4. The input image of the molecule galanthamine.

The first step is to categorise and normalise the image, to recognise what must be done to understand the

All procedural CIR algorithms perform these steps, with some level of success. The remainder of the algorithm is the most important and ChemDoodle CIR excels here. The interpretation of the shapes is not trivial. Take a look at the two shapes pointed out by arrows. The top one looks like a fork coming off of a complex ring system and the bottom one looks like bent arm with a hand. How are these to be perceived? Our goal was to produce an algorithm to match how a humanζ

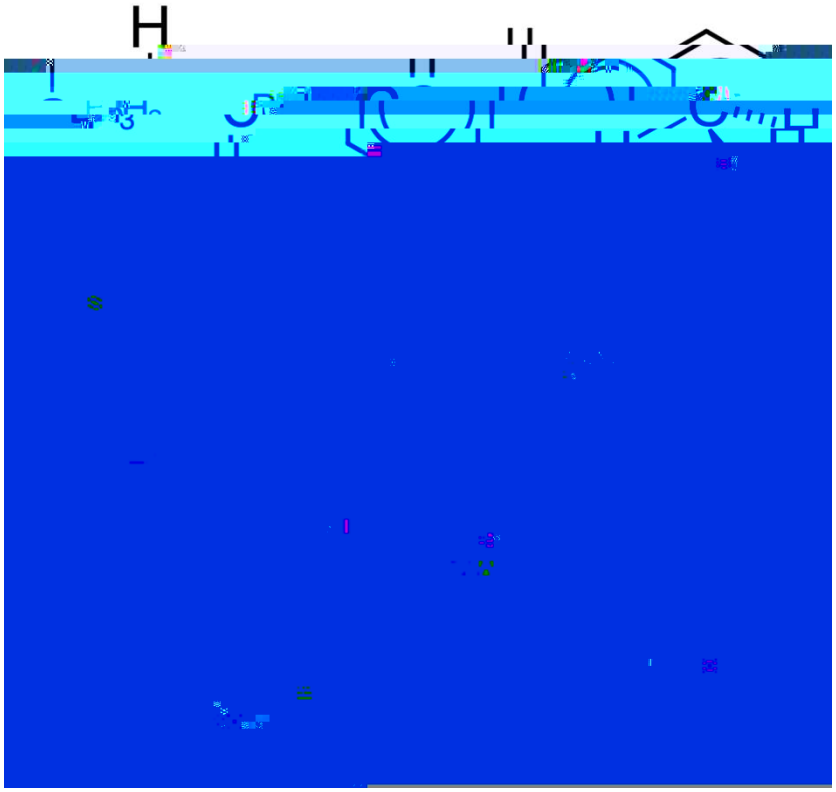


Figure 10. Simplistic tests often overlooked by other CIR algorithms.

Runtime is also an important consideration. ChemDoodle's CIR algorithm processes the galanthamine image with an average runtime of 76.4ms. The most time-intensive part of the algorithm is optical character recognition. The longest runtime we have found is for an image we created with 13 complex atom labels resulting in an average runtime of 115.1ms. So, expect the performance to scale with the amount of text in the image requiring recognition. All benchmarks were performed on a 2017 iMac running macOS 11.2.2 with a 4.2 GHz Quad-Core Intel Core i7 CPU. Each image was recovered 20 times, with the first iteration disregarded as a warm-up. The remaining 19 iterations were averaged. Java version 11.0.8 was used to compile and run the tests.

Finally, there are some limitations, as to be expected with a CIR project. CIR will work on computer-generated, skeletal images of chemical structures. Hand-drawn images may not work well. Clearer images at a crisp resolution will have the best results. The messier or blurrier the image is, the more ChemDoodle will have to use intuition to resolve the chemical structure, similar to a human looking at an unclear image. ChemDoodle may interpret graphics differently than you may. Our goal is to have the CIR features perfectly match what you perceive in the image, but you should expect to perform some level of post-editing on some images.

I hope this discussion has provided an in-depth view into our current CIR work in ChemDoodle. The initial results are excellent and ChemDoodle's CIR out-performs many competing CIR solutions. We really hope this feature will help eliminate the effort you spend transcribing chemical structures from images. If you are not happy with the results, please send us the image so we may improve the algorithm. We will continuously develop it. As always, ChemDoodle subscribers, Site and Lifetime licensees are entitled to our latest

and I (all former Presidents of the IIS) compiled a history of the Institute in the course of the last three years, one of our objectives was to compile profiles of all the Presidents. However, we could find very little information about Dyson and his role in the establishment of the IIS. After his period as President, he seemed to play no further role in the ((2ζÚ‱activities. The only public source of information was a short entry in the German-language version of Wikipedia.

With the work on the history complete, I decided to research his life and achievements hoping to gain a better understanding of why he was invited to be President. Six months of research has resulted in an 18,000 word/140 citation profile of Dyson which will be released early in 2022 to mark the 20th anniversary of the formation of CILIP. This pre-×ÜÉ⁰ÊÈÙÞÖÕ‱ÚÜÔÔÈÙà‱ÖÕóà‱ÊÖÕÚÞËÌÙÚ‱#àÚÖÕζÚ‱ÞÖÙk relating to chemical information science and hopefully illustrates how appropriate his achievements made him the ideal choice to be the first President of the Institute.

Malcom Dyson (he never used his initial given name of George) was born on 5 April 1902 in Plumstead, South London. Dyson went up to Jesus College, Oxford, on a full scholarship in October 1921. Jesus College was one of six Oxford colleges to have its own chemistry laboratory, with David Leonard Chapman FRS as Senior Fellow. Dyson not only gained a First in 1925 but records show that he also gained a First in chemistry in 1923, and then in 1925 a PhD from the University of London as anJesus College, Oxford

From his authorship of *The Chemistry of Chemotherapy*, his decade as an academic and his research interests, Dyson was well aware of the challenges of creating logical and consistent names and structural diagrams for organic compounds. In addition, he had a very good understanding of the use of the chemical literature and of the needs of research chemists.

It is important to appreciate that Dyson regarded his work on notation as means of giving each chemical ÊÖÔ×ÖÜÕËÈÜÕÐØÜÌÊÐ×ÏÌÙÇÕÈÕÌßÈÔ×ÓÌÏÌÎÐÝÌÚÐÚÛÏÈÛ"2"Ó4 could be described as:

Perchloromethyl mecapatan
Thiocarbonyl tetrachloride
Trichloromethyl sulfur chloride
Tetrachloromethyl thiol
Trichloromethyl sulfenyl chloride

To make matters worse there were British, French, German and American naming conventions.

The first announcement of what would become known as the Dyson Notation was a letter by Dyson dated 24 June 1944 and published in *Nature* on 22 July 1944. In the letter he mentions that he would be publishing a book on the systematic notation that he was developing. He stated the objective as establishing a database (though he did not use this term) of codes, each of which represented the structure of a unique chemical entity. The notation was based around determining and then supplementing the longest carbon chain.

The first public presentation by Dyson of his notation for organic compounds was at a meeting of the Royal Institute of Chemistry in 1946. The Institute was so impressed it circulated a copy of his lecture to its members. The first edition of his book *A New Notation and Enumeration System for Organic Compounds* was published by Longmans in 1947. Then on 3 February 1948 he gave a lecture to the British Society for International Bibliography that was reprinted in the inaugural issue of *Aslib Proceedings*

At the c

science: Dyson through his vision of what the emergent technology could mean for a chemical information ËÈÙÈÉÈÚÌ㫹×ÈÕÕÐÕÎ㫹ÈÓÓ㫹ÖÍ㫹ÛÏÌ㫹àÌÈÙÚ㫹ÈÕËã3ÈÛÌ¿Ú㫹×ÙÌÖÊÊÜ×ÈÙÐÖÕ㫹ÉÌÐÕÎ㫹ÞÐÛÏ㫹ÜÚÐÕÎ㫹ÛÏÌ㫹ÛÌÊÏÕÖÓÖÎà㫹ÛÖ㫹ÎÌÛ㫹ÛÏÌ㫹ÚÏÖÞ㫹ÖÕ㫹ÛÏÌ㫹ÙÖÈËÛ㫹ÛÖ㫹ÊÖÕÛÐÕÜÌ㫹ÛÖ㫹×ÙÖÝÐËÌÚ㫹ÚÌÙÝÐÊÌÚ㫹Þ ÏÐÊÏ㫹 ÏÈË㫹ÖÜÛÎÙÖÞÕ㫹×ÙÐÕÛÌ¿Ú㫹ÐÕÒ㫹ÈÕËã leading in time to a cumulative database. Each was right in his own vision of the future, differing mainly about the means by which they were to be achieved. Dyson and Tate stood above the technology of their time ł the fact that they started with an 8K IBM 1401 did not limit their thinking ł rather, they were confident that when the need arose the ÛÌÊÏÕÖÓÖÎà㫹ÞÖÜÓË㫹ÈÙÐÚÌ㫹ÛÖ㫹ÛÏÌ㫹ÖÊÊÈÚÐÖÕ?

In the last two decades of his life Dyson provided CAS with a vision of how computers could be used to create indexes of organic chemicals, developed the first computer-based current-awareness service, wrote two books on searching the literature of chemistry, made a major contribution to the success of the 1948 Royal Society Conference, supported Jason Farradane in setting up the Institute of Information Scientists and (albeit indirectly) led to the establishment of what would become a world-respected cheminformatics research group at the University of Sheffield.

In his transition from a highly regarded research chemist to these achievements in chemical information management Dyson was arguably the first şinformation scientistz.

Robin Darwell-Smith (Archivist, Jesus College, Oxford), Evan Hepler-Smith (Andrew W. Mellon Assistant Professor of History at Duke University, Durham, North Carolina), David Allen (Librarian, Royal Society of Chemistry), Andrew Dalke (Dalke Scientific Software AB) and Peter Willett (Emeritus Professor, Information School, University of Sheffield) have made substantial contributions to my research.

This list includes only papers authored or co-authored by Dyson in the area of what might now be referred to as cheminformatics. There are also research papers, patents and a number of chemistry monographs, most of which can be found on Google Scholar under GM Dyson.

1944	G.M. Dyson A notation for organic compounds. *Nature* 1944*, 154,* 144.
1946	G.M. Dyson Lecture on a new notation for organic chemistry and its application to library and indexing problems. Delivered at a joint meeting of the Chemical Society, the Royal Institute of Chemistry, the Society of Chemical Industry and the Bureau of Abstracts at the London School of Hygiene and Tropical Medicine, 21 October 1946.
1947	G.M. Dyson A New Notation and Enumeration System for Organic Compounds, 1st ed

1952	G.M. Dyson. The preservation and availability of chemical knowledge. *J. Chem. Educ.* 1952, *29* (5), 239. (Presented at the XII International Congress of Pure and Applied Chemistry, New York, September 1951.)
1955	G.M. Dyson. Advances in classification. *J.Doc.* 1955, *11* (1) 12-18. (Presentation to an Aslib meeting 17 December 1954.)
1958	G.M. Dyson. A Short Guide to the Chemical Literature, 2nd ed. Longmans, Green & Co., 1958.
1961	G.M. Dyson Current research at Chemical Abstracts. *J. Chem. Doc*. 1961, *1* (1), 24‡28. (13 September 1960.)
1961	G.M. Dyson. Searching the older chemical literature. *Advances in Chemistry 4*, Chapter 15, 96-103.
1961	G.M. Dyson Indexing scientific progress by computer. *New Scientist,* 1961 (30 March), *2283,* 817-819.
1962	G.M. Dyson & E.F. Riley. Use of machine methods at Chemical Abstracts Service. *J. Chem. Doc.* 1962*, 2* (1), 19-22.
1962	C.L. Bernier, G.M. Dyson & H.J. Friedman. Correlative Indexes VII. Trope vocabularies and trope indexes for chemistry. *J. Chem. Doc*., 1962, *2* (2), 93

A report by Dr Wendy A Warr, https://www.warr.com/

3Ï Ì ¢?ÝÐÙÛÜÈÓ?Æ ÚàÔ×ÖÚÐÜÔ Þ È Ú Ö Ù Î È Õ Ð Ú Ì Ë É à Û Ï Ì 1 Ö à È Ó 2 Ö Ê Ð Ì Û à Ö Í " Ï Ì Ô Ð Ú Ú Ù à ¿ Ú ! Ð Ö Ó Ö Î Ð Ê Ê È Ó È Õ Ë , Ì Ë Ð Ê Ð Õ È Ó

Nevertheless, AI alone cannot transform drug design. High-throughput data generation, automatisation in the DMTA cycle, and combining AI with physics (e.g., to predict physicochemical properties and estimate binding affinity) can add value to AI approaches. Combining AI with big data can transform synthesis prediction.[8] In AstraZeneca, chemists have access to data on 20 million reactions in the ReactionConnect database, from which predictive models can be built and used to automate synthesis. ReactionConnect is populated with data from AstraZeneca reaction sources and ELNs, a USPTO database, and Reaxys and Pistachio flat files.[9] AiZynthFinder can be used in retrosynthetic planning. The algorithm is based on a Monte Carlo tree search that recursively breaks down a molecule to purchasable precursors. The tree search is guided by an artificial neural network policy that suggests possible precursors by using a library of known reaction templates.[10] ▯?1ÐÕÎ▯!ÙÌÈÒÌÙ?▯ algorithm[11] improves the route-finding. It uses a data-driven approach to enable the prediction of ring-forming reactions, useful in establishing the synthetic accessibility of unprecedented ring systems. Another improvement, RAscore,[12] is an ML-based method able to classify whether a synthetic route can be identified or purchasable

(core and R-groups) resulting from exploration of the chemical space, showing some simple structures but with different R-groups and complexity starting to appear.

ÜÚÌËÉà#ÜÈÙÛÌ¿ÚÛÌÈÔ starts from a small number of randomly selected points in the configuration space, from which active learning training of intra- and inter-molecular components of the energy and forces is carried out. The CUR algorithm[27,30] is applied.

Splitting the database into training and test sets and using a standard retrospective validation strategy is not practical in the current application so a temporal cumulative error metric was used based on the time required

Another

forward model captures enzymatic reaction rules based on the EC number. The model mimics the expert-curated reaction rules in automated retrosynthesis.

A resident chemist has tried the system out and has been able to replace a traditional reaction with an enzyme-catalysed one. Anyone can try the system for free at IBM RXN for Chemistry. Stereochemistry is included for all reactions. The enzymatic data and the trained models are available through the RXN for Chemistry network and on GitHub.

Alexander Pritzel, DeepMind, London, UK

A central part of #Ì Ì× , ÐÕËչÚ Ô ÐÚÚÐÖÕ Ð Ú Û Ö ÚÖÓÝÌ ÍÜÕËÈÔÌÕÛÈÓ ÚÊÐÌÕÛÐ ŗÊ ×ÙÖÉÓÌÔÚ ÞÐÛÌ (ð

modelling the backbone and also builds the side chains by predicting torsion angles. The AlphaFold architecture can be trained to high accuracy using only supervised learning on PDB data, but accuracy can be further enhanced using an approach similar to noisy student self-distillation.[68] This is the way AlphaFold makes use of unlabelled sequences. The

excited to see what others are building on top of the AlphaFold database. There is great potential in AI for science as a whole.

Directed evolution (for which Frances H. Arnold won half a Nobel Prize in 2018) depends on generating a large gene library, needing lots of costly effort. Rational, computer-aided design techniques might never be able to ÚÈÔ×óÌ‱ÛÏÙÖÜÎÏ‱ÛÏÌ‱ÌÕÛÐÙÌ‱×ÙÖÛÌÐÕ‱ÚÌØÜÌÕÊÌ‱ÚÊÌ‱ÈÕËÉÌÕÌÍÐÛ‱ÍÙÖÔ‱ÕÈÛÜÙÌ¿Ú‱ÍÜÓÓ‱×ÖÛÌÕÛÐÈÓ‱ÍÖÙ‱ÛÏÌ generation of better enzymes. There is a clear trend to combine the rational design and directed evolution approaches. Semi-rational design generates small, functionally rich, mutant libraries using rationally pre-selected target sites. Knowledge-driven approaches navigate sequence space intelligently. Recently, ML methods have been increasingly applied to find patterns in data that help predict protein structures, improve enzyme stability, solubility, and function, predict substrate specificity, and guide rational protein design.[71-74]

In evolutionary biology, fitness landscapes are used to understand the relationship between genotypes and reproductive success. It is assumed that every genotype has a well-defined replication rate (fitness). This fitness ÐÚ‱ÛÏÌ‱Ì?ÏÌÐÎÏÛ?‱ÖÍ‱ÛÏÌ‱ÚÈÕËÚÊÈ×Ì‰&ÌÕÖÛà×ÌÚ‱Þ ÏÐÊÏ‱ÈÙÌ‱ÚÐÔÐÓÈÙ‱ÈÙÌ‱ÚÈÐË‱ÛÖ‱ÉÌ‱ÊÓÖÚÌ‱ÛÖ‱ÌÈÊÏ‱ÖÛÏÌÙÛ‱Þ ÏÐÓÌ‱ÛÏÖÚÌ‱ÛÏÈÚ‱ are very different are far from each other. The set of all possible genotypes, their degree of similarity, and their related fitness values is then called a fitness landscape. The size of the protein sequence space is huge and the fitness landscape is complex. Current challenges are screening throughput (leading to limited exploration, information gaps and local maxima); the combinatorial problem of epistasis (a phenomenon in which the effect of a gene mutation is dependent on the presence or absence of mutations in one or more other genes); and cost and time.

Combining next generation sequencing (high-throughput analysis of DNA and RNA sequences) with high-throughput screening of 10^4-10^8 variants per day is a powerful strategy (deep mutational scanning) for comprehensively analysing sequence-function relationships.[72] ML-guided directed evolution reduces experimental effort and mutates multiple positions simultaneously, combining directed evolution and rational design (Figure 7).[71,73]

Pythonic Protein Engineering Framework (PyPEF‰ÐÎÜÙÌ‱ÙÆ‱ÐÚ‱È‱ÎÌÕÌÙÈÓ×ÜÙ×ÖÚÌ‱ÍÙÈÔÌÞÖÙÒ‱ÍÖÙ‱ËÈÛÈ±ËÙÐÝÌÕ‱ protein engineering by combining ML methods (partial least squares (PLS), RF, support vector regression

increasingly available) but structure-based design is not biased by prior ligand knowledge. G Protein Coupled Receptors (GPCRs) are a particular target class where structural data can have a significant impact.[77]

Thomas and his co-workers[78] have assessed the use of molecular docking *via* Glide (a structure-based approach) as a scoring function to guide the deep generative model REINVENT[5,79] and compare model performance and behaviour to a ligand-based scoring function. The case study involved dopamine receptor D2 (DRD2). The approach taken is depicted in Figure 9, where data sources are coloured blue and scoring functions orange. The REINVENT framework (in grey) consists of two recurrent neural networks, a prior and an agent. The main steps in the current work are (1) removing known DRD2 active molecules from the ZINC training data; (2) training the prior model on druglike molecules from ZINC; (3) initializing the agents as a copy of the prior; (4) preparing the scoring functions to evaluate *de novo* molecules; (5) iteratively training both agents *via*

Prediction of rat PK is a stepping stone towards modelling human PK. An AI model predicts rat PK parameters from chemical structure and measured *in vitro* ADME properties. The chemical structure is encoded by a graph convolutional neural network (GCN). Properties used as input features are solubility, Caco2 (colorectal adenocarcinoma cell) intrinsic permeability and efflux, intrinsic clearance (CL_{int}) in human liver microsomes (HLM), rat hepatocytes, intrinsic clearance and fraction unbound, and rat and human plasma protein binding (PPB). Properties predicted are clearance (CL), bioavailability (%F, the fraction of an oral administered drug that reaches systemic circulation), C_{max} (the maximum serum concentration that a drug achieves in a specified test area of the body after the drug has been administrated and before the administration of a second dose), $t_{1/2}$ (elimination half-life, the time required for the concentration of the drug in the plasma to reach half of its original value) and V_{ss} (volume of distribution at steady-state).

3ÏÌÔÌÛÏÖËÜÚÌÚÔÌÚÈÎÌ×ÈÚÐÕÎ - - ÚÍÖÔÍÊÜÓÈÙ×ÙÖ×ÌÙÛà×ÙÌËÐÊÛÐÖÕ?chemprop?ÍÙÖÔÛÏÌ Machine Learning for Pharmaceutical Discovery and Synthesis (MLPDS) consortium, a collaboration between industry and the Massachusetts Institute of Technology. The rat PK model achieved good accuracy on key PK parameters (Table 2). CL was predicted within 2-fold error for 75% of compounds and within 3-fold for 90% of compounds.

(34) Sushko, I.; Salmi

(55)

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

Save the date for our Network Conference: 1-3 March 2022.

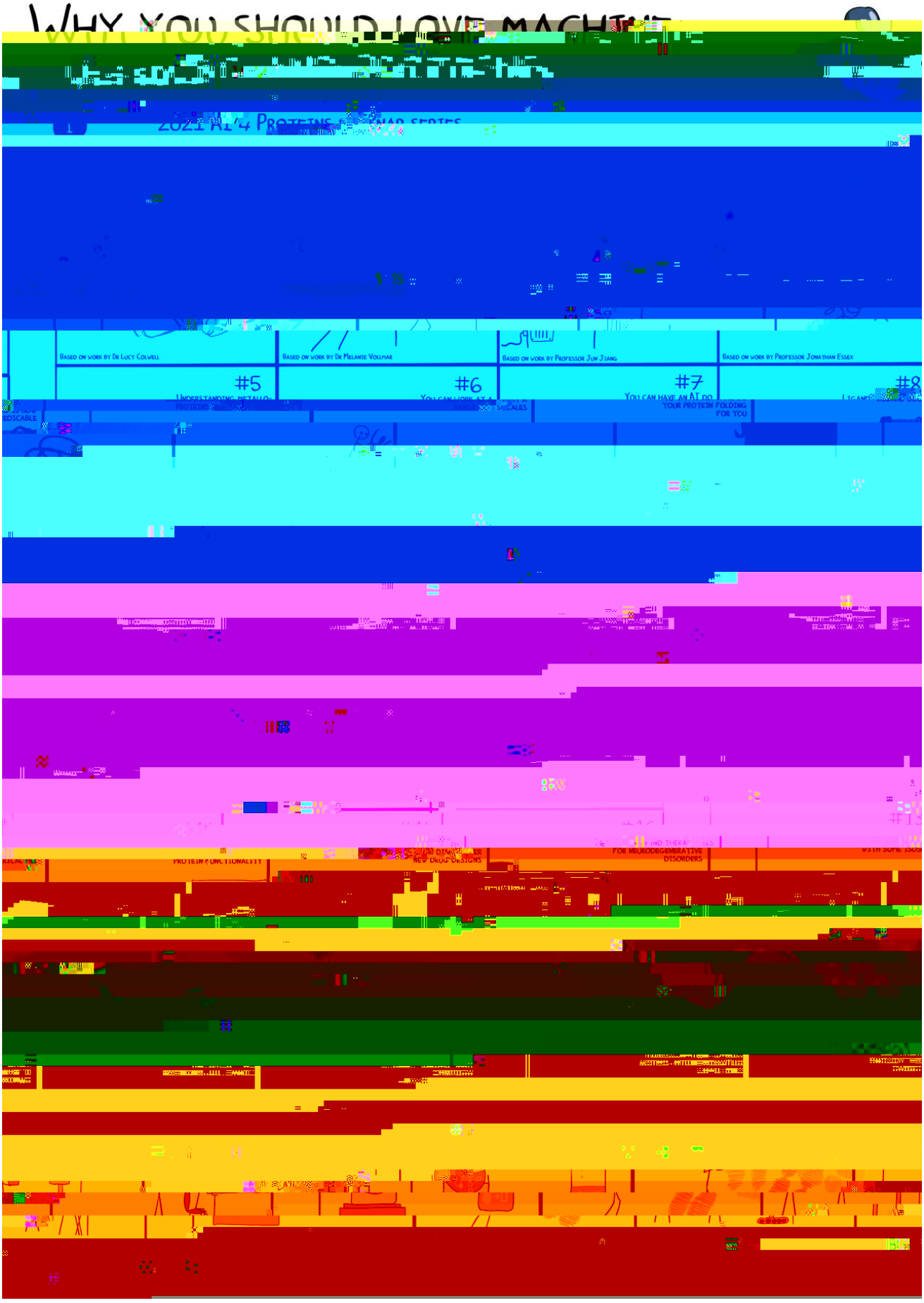
We would be delighted if you would join us for our AI3SD Network Conference on 1-3 March 2022. We are hoping to run this as a hybrid event at the Best Western Chilworth Manor, although it may be moved to fully online depending on the Covid situation. The conference will be a mixture of keynote talks, discussion sessions and networking opportunities.

If you would be interested in submitting a short talk abstract please complete our AI3SD Conference 2022 Abstract Submission Form. There will also be some musical entertainment in the evening! We will be opening up registration for this conference closer to the time. Further details can be found on our website.

If you would be interested in finding out how to sponsor this event please contact us info@ai3sd.org and we can tell you more.

AI3SD has been working as part of the Network of Networks group to combine our shared knowledge on how to run a Network+. This resource has been produced by a group of diverse research management professionals, representing different disciplines and organisations to aid network managers and investigators in the creation and management of research communities. Please check it out here and share with your contacts.

Back in April, AI3SD and RSC-CICAG launched our AI4Proteins Seminar Series. The videos from this series can be found on our AI4Proteins Playlist. This entire series has been captured in an incredibly detailed report


#5
#6
#7
#8
FOR NEURODEGENERATIVE DISORDERS

As detailed in "(" &ź Ú summer newsletter, AI3SD teamed up with the Physical Sciences Data-science Service (PSDS) to create a Skills4Scientists series for our summer interns. This included events to educate scientists on: research data management, Python, version control, LaTeX, creating posters and presentations, ethics and valuable careers skills, including a two-day virtual event run in conjunction with RSC-CICAG on careers and posters. All of the material from these sessions are now available on our YouTube Channel: Skills4Scientists Playlist.

Here is a list of our intern projects with links to their reports:

A deep neural network for generation of functional organic molecules		

Introduction to Cheminformatics and Machine Learning

The CAS blog features deep insights into a variety of issues impacting the scientific community. Recently covered topics include bioorthogonal reactions, applications, and trends in the CAS Content Collection™; emerging trends in targeting "undruggable" RAS proteins for cancer treatment; assessing structural novelty of

Contributed by Stuart Newbold, email: stuart@psandim.com

The latest volume in the *Reviews in Computational Chemistry* series, the invaluable reference to metho

Chemistry Entrepreneurship is a step-by-step guide that is specifically devoted to understanding what it takes to start and grow a new company in the chemistry sector. Comprehensive in scope, the book covers the various aspects of the creation of a new chemical enterprise including: the protection of the invention, the business plan, the transfer from the research center or university, the financing, the legal setup, the launching of the company and its growth and exit strategies.

This hands-on book contains the information needed to help to determine if you have what it takes to be a chemistry entrepreneur, explains how to take an ideas out of the lab and into the real world, reveals how to develop your burgeoning business, and shows how to sustain and grow your business. This much-needed resource also includes interviews with founding scientists who created their own successful chemical companies. This important book:

- Provides the practical information on how to start a company based on a scientific breakthrough
- Offers information on the mindset it takes to become, and remain, successful in the marketplace
- Presents case studies from world-renowned and highly experienced professionals who have successfully started a company

Written for chemists in industry, chemists, materials scientists, chemical engineers, Chemistry Entrepreneurshipis a guide for becoming a founder of a successful chemical company.

A new journal concept from Cambridge University Press will bring researchers from different fields together around the fundamental questions that cut across traditional disciplines. By focussing research on finding answers to such questions, this unique approach will speed discovery by fostering collaboration and knowledge sharing between subject communities. It will also provide opportunities to publish research from areas that are not well served by traditional, discipline-specific journals.
https://www.cambridge.org/news-and-insights/news/Taking-the-research-journal-in-a-new-

As part of the efforts of the National Library of Medicine (NLM) to transform and accelerate biomedical

Artificial intelligence can be used to scour the crowdsourced encyclopedia for contradictory information and flag it to human editors.
https://www.newscientist.com/article/2298169-wikipedia-tests-ai-for-spotting-contradictory-claims-in-articles/#ixzz7H7DtnG1S
Source: NewScientist

Digital Science has announced a new partnership with OntoChem GmbH. The partnership allows OntoChem and Digital Science to join forces for mutual clients, particularly in the Life Sciences industry, through . ÕÛÖ " Ï Ì Ô¿Úщ×ÖÞ ÌÙÍÜóщÚÌ ÔÈÕÛÞÊщÐÕËÌßÞÕÎщÊÈ×ÈÉÐóÞÛÞÌÚõщ . ÕÛÖ " Ï Ì Ôщ Ï ÈÚщÔÖÙÌщÛ Ï ÈÕщhuKщàÌÈÙÚ¿щÌß×ÌÙÞÌÕÊÌщÊÙÌÈÛÞÕÎщ innovative technologies. The German-based life sciences company develops cognitive computing solutions, indexing intranet and internet data and applying semantic search solutions for pharmaceutical, material science and technology-driven businesses.
https://www.digital-science.com/press-release/digital-science-partners-with-ontochem/
Source: Digital Science

Karger Publishers has started offering its authors an automated proofreading service from the company Writefull. Writefull uses AI-based language models to suggest language edits, enabling authors to improve the language of their manuscript before submission.
https://www.stm-publishing.com/karger-publishers-launches-trial-with-writefull-language-check/
Source: STM Publishing News

Karger Publishers is a globally active independent publisher dedicated to serving the information needs of the

Ex Libris has announced the publication of its annual study on the challenges that academic researchers face, the priorities of research office leaders, and key opportunities for libraries and research offices to advance scholarship at their institution. Commissioned by Ex Libris, the study was conducted by Alterline, an independent research agency. The report presents findings from a survey of more than 400 researchers and research office leaders across a range of disciplines in the USA, the UK and Australia.

The Electrochemical Society (ECS), together with IOP Publishing, is launching two new, fully open access (OA) journals. *ECS Advances* and *ECS Sensors Plus* add to the SoÊÐÌÛàzÚ journal family and provide the research community with a diverse suite of interconnected journals sharing impactful research across the world. *ECS Advances* delivers a platform for research across all areas of electrochemical and solid-state science and technology research with the broadest dissemination of all journals in the field. *ECS Sensors Plus* offers a specialised outlet for all content related to sensors technology. The journal will lead and promote scholarly communication and interactions among scientists, engineers, and technologists whose primary interests focus on materials, structures, properties, performance, and characterisation of sensing and detection devices and systems, including sensor arrays and networks.
https://www.knowledgespeak.co8(k.c)-2(o8(k.c)-2(o8(k.c)-2(o8(k.ctw)-11(o)4()-8(k.6ciGrheo)4(f(l)] TJ16(n)4(d)1(eaem4(d)-n)

https://www.scientific-computing.com/news/ibm-research-europe-and-thieme-chemistry-collaboration-accelerates-discovery-organic-chemistry

Source: Scientific Computing World

The American Association for Cancer Research has announced the opening of the submission site for its new open access journal, *Cancer Research Communications*, signalling the official call for papers. This new journal is ÛÏÌ₩ƴÛÏ₩ÐÕ₩ÛÏÌ₩ "1ʐÚ₩×ÖÙÛÍÖÓÐÖ₩ÖÍ₩ÚÊÐÌÕÛÐÍÐÊ₩×ÜÉÓÐÊÈÛÐÖÕÚÛ₩ÈËËÐÕÎ₩ÈÕ₩Ö×ÌÕ₩ÈÊÊÌÚÚ₩×ÜÉÓÐÚÏÐÕÎ₩Ö×ÛÐÖÕ₩ÛÖ₩ÛÏÐÚ₩ÏÐÎÏ- quality, trusted journal collection. *Cancer Research Communications* welcomes research spanning the full breadth of cÈÕÊÌÙ₩ÚÊÐÌÕÊÌ₩ÈÕË₩ÔÌËÐÊÐÕÌð₩(Õ₩ÈËËÐÛÐÖÕ₩ÛÖ₩Ì ß×ÈÕËÐÕÎ₩ÛÏÌ₩ "1ʐÚ₩×ÖÙÛÍÖÓÐÖ₩ÖÍ₩×ÌÌÙ-reviewed publications, the new journal will further stimulate the exchange of innovative ideas and approaches in cancer research and will provide a rapid publication outlet that serves the cancer field.
https://www.knowledgespeak.com/news/aacr-issues-call-for-papers-for-new-oa-journal-cancer-research-communications/
Source: Knowledgespeak

3ÏÌ₩ş5ÖÐÊÌ₩ÖÍ₩ÛÏÌ₩ . ÕÓÐÕÌ₩+ÌÈÙÕÌÙʐ1Ì×ÖÙÛ₩%ÐÕËÚ₩/ÖÖÓ₩ÖÍ₩ . ÕÓÐÕÌ₩+ÌÈÙÕÌÙÚ₩$ß×ÈÕËÌË₩#ÜÙÐÕÎ₩ÛÏÌ₩/ÈÕËÌÔÐÊÛ₩ÈÕË₩
New Cohort Emerged that Skews Younger, More Likely to Pursue Online Undergraduate Degrees.
https://newsroom.wiley.com/press-releases/press-release-details/2021/Pandemic-Has-Cultivated-New

Dialog Solutions announced it was adding the ClinicalTrials.gov database to its platform. Since ClinicalTrials.gov is a free database on the web, what is the advantage to searching it on Dialog?
https://www.infotoday.eu/Articles/News/Featured-News/Dialog-Solutions-Adds-ClinicalTrials-database-149298.aspx
Source: Information Today

The Bioinfogate OFF-X a- portal is a cutting-edge safety intelligence solution aimed at empowering pharmaceutical organisations to identify toxicology and safety signals, mitigate safety liabilities and de-risk early-stage assets. It is one of the largest translational safety and toxicity portals, featuring over 1,200,000 safety alerts corresponding to over 23,000 drugs and biologics and more than 15,000 targets of pharmacological interest. As a leading provider of trusted information and insights to accelerate innovation, Clarivate offerings include a comprehensive suite of research intelligence solutions coupled with deep domain expertise. The acquisition of Bioinfogate will fill a critical need for drug toxicity data and translational safety intelligence across all stages of drug R&D. This follows a previ(o)-7(h)4(a)-5(t is th)3(e)-5(a)-5(d)-264-wsgh thPro(teli(o)-7(hf(a)6(t)-32(o

aT/F3 Wc Exl -3pi9((n)4(ee)-5380.G(en nBT4(ee)-5380.G(en nBT4(ee)-5380.7sm0 pg0.0000088a)-5()-5(n)] TJETN.f26 u0o).e)-4(8ETN.f26 u