



Medicinal Chemistry and Chemical Biology Highlights

Division of Medicinal Chemistry and Chemical Biology

A Division of the Swiss Chemical Society

DMCCB Basel Symposium 2024: Therapeutics by Computational Design: Innovations in Drug Discovery and AI

Simon R. Williams*

*Correspondence: Dr. S. R. Williams, E-mail: simon.williams@syngenta.com
Syngenta Crop Protection AG, Schaffhauserstr. 101, CH-4332 Stein

2023 was the year of ChatGPT and artificial intelligence is becoming increasingly ubiquitous in all walks of life. In scientific research AI is now being applied to almost every aspect of chemical discovery and optimisation from the generation of new molecules, to predicting their synthetic routes and modelling complex pharmacology, not to mention the more mundane tasks such as writing up results (disclaimer, this report was written entirely by human intelligence!).

AI in drug discovery was therefore a timely topic for this year's DMCCB Basel symposium, held at the University of Basel's BioZentrum on February 12th. The meeting was attended by approximately 120 participants from across the academic and industrial sectors with specialisms spanning drug discovery, chemical biology, computational chemistry, and machine learning.

At the early stage of the discovery process generative AI can be used as a tool to discover new chemical matter for screening against a target. A generative AI workflow requires a generative model that 'invents' molecules with a particular set of properties, coupled with a scoring algorithm that ranks the generated molecules based on relevant parameters. The results for the scoring algorithm then feed back into the generative model to inform the design of further compounds. Developing models that can generate structures that are novel but chemically and synthetically reasonable is an ongoing challenge.

To this end **Prof. Hongming Chen** (Guangzhou Laboratory) presented his groups research on the 'Development of Novel Generative Models for Molecule Design'. Prof. Chen's group has developed 'Tree-Invent',^[1] a molecular generative model capable of producing structures that adhere to specific topological constraints. The model's adaptability extends to various molecule design applications, including scaffold decoration, scaffold hopping, and linker generation, showcasing its utility in diverse design scenarios. Prof Chen also presented EC-Conf, a novel diffusion method for rapid generation of low energy conformers.^[2] Conformer generation is necessary for many computational methods such as pharmacophore modelling, virtual screening and QSAR modelling and often represents a bottleneck in high throughput computational workflows. EC-Conf demonstrated up to two orders of magnitude improvement in efficiency over traditional diffusion models.

Prof. Gerard J. P. Van Westen (Leiden University) also presented a new approach to generative modelling called 'UnCorrupt SMILES'. This model corrects any invalid smile outputs of a generative model to improve the coverage of chemical space.^[3]

Prof. Van Westen also presented several other workflows and models designed to facilitate virtual screening. QligFEP^[4] and QresFEP^[5] are automated workflows for binding free energy calculations. QligFEP uses a dual topology strategy to provide a clear perturbation pathway between the ligands being studied in a free-energy perturbation experiment and QresFEP is designed to specifically explore the impact of point mutations in the target protein. DrugEx is a multi-objective reinforcement learning-based, open-source software package for *de novo* drug design.^[6] Integrated within the GenUI software platform, DrugEx provides a versatile tool for users ranging from those seeking quick experiment setup and model building through graphical user interfaces, to those desiring deeper workflow customization via its new Python API.



Fig. 1. Clockwise from top-left: Prof. Hongming Chen, Prof. Andrea Volkamer, Prof. Gerard JP van Westen, Dr. Nadine Schneider.

Another approach to generating new leads is mining open-source data such as that stored in databases such as ChEMBL or PDB. **Prof. Andrea Volkamer** (Universität des Saarlandes) presented her group's work on the development of tools to aid the design and discovery of new kinase inhibitors, a class particularly rich in relevant data. Prof. Volkamer presented methods for templated docking of kinase inhibitors to increase the size of datasets used to train ML models for generative chemistry^[7] and a method called KiSSim that can be used to predict possible off-targets within the kinome.^[8]

One of the key challenges with the rapid development of AI models to assist drug design is how to ensure that the next generations of medicinal chemists are appropriately skilled in their use, especially as a lot of software is either proprietary or requires a degree of familiarity with programming languages to implement. The Volkamer group have developed 'TeachOpenCADD', an ex-

Can you show us your Medicinal Chemistry and Chemical Biology Highlight?

Please contact: Prof. Dr. Kathrin Lang, Dept. of Chemistry and Applied Bioscience, ETH Zurich, Vladimir-Prelog-Weg 1-5/10, CH-8093 Zurich, E-mail: kathrin.lang@org.chem.ethz.ch

tensive training toolkit for teaching computer aided drug design that provides simple tutorials and workflows, using open-source software, for common tasks in CADD.^[9] This is a valuable resource both for students of organic and medicinal chemistry but also for established practitioners looking to keep their skills up to date.

It is not just small molecule discovery that can benefit from AI techniques, and this was elegantly illustrated by **Prof. Patrick Barth** (EPFL, Laboratory of Protein and Cell Engineering). Proteins are flexible molecules and Prof. Barth presented his group's work on using molecular dynamics and deep learning models to help understand the allosteric behaviour of membrane receptor proteins in response to different stimuli. This has allowed the group to design new sensors, demonstrated with the development of an ultrasensitive chemotactic sensor based on the chemokine receptor CXCR4 and a modified SDF-1 peptide.^[10] Incorporation of conformational dynamics into the design workflow was essential to develop potent receptor agonists without switching into antagonism.

In addition to the keynote speakers there were three short talks. **Marvin Alberts** (IBM / University of Zürich) presented his PhD studies on the development of methods for automated structure determination using machine learning models for the interpretation of NMR and IR spectra.^[11] **Jennifer Müller** (ETH Zürich) developed novel inhibitors of the amino-acid transporter LAT-1 as potential antitumor agents.^[12] Using the clinical candidate JPH203 as a lead compound and supported by cryo-EM data, she used a cyclisation strategy to rigidify the compound and improve potency. **Lukas Schneider** (University of Basel) presented a new method for DNA-encoded library screening based on fusing a nucleotidyl transferase enzyme to the target protein.^[13] In the presence of ATP this polyadenylates the DNA tag of any compound that binds to the target allowing for simple isolation of binders using poly-dT functionalised beads. This method overcomes a number of limitations with conventional affinity selection and improves the resolution of the selection step. Importantly it is fully back compatible with existing DNA encoded libraries. Jennifer Müller and Lukas Schneider have received the DMCCB PhD prize for their work at the symposium.

Best poster prizes were won by **Pinwen Cai** (University of Basel) for his work on 'Selecting DNA encoded small molecules to direct protein ubiquitination' and **David Kreutter** (University of Bern) who is developing an open-source tool for 'chemoenzymatic multistep retrosynthesis with transformer loops'.

On the other side of the coin to the development of new methods is their implementation and use. The final two talks of the day gave an industrial perspective on the progress of AI in industry and examples of how it is being used to support projects and generate new leads.

Nadine Schneider (Novartis) gave an overview of the integration of machine learning models in medicinal chemistry at Novartis and demonstrated their impact on all stages of the design-make-test-analyse cycle.^[14] Applications include retrosynthesis prediction, generative chemistry, protein structure prediction and activity prediction both on- and off-target. Successful integration into daily medicinal chemistry activities is both cultural and technical as many medicinal chemists are not computational experts. Connected platforms with integrated ML capabilities make the use of AI straightforward and no-code platforms for developing new ML models allow 'citizen developers' to experiment with developing their own tools. Above all team members must be encouraged to experiment with AI tools and collaborate across different domains of experience. She also highlighted the importance of both industry and academia contributing to the development of open-source tools to democratise access to AI for all and enable collaboration.

To close the meeting, **Chris Baker** (Syngenta) gave an overview of a recent collaboration between Syngenta and *in silico*

medicine. The goal of the collaboration was to discover new leads for weed-control using an inverse design strategy driven by *in silico* medicine's AI platform PHARMA.AI. The collaboration explored different strategies for generating and scoring compounds and were ultimately able to identify *in silico* generated compounds with *in vitro* and *in vivo* activity. The conclusions to this talk summarised some of the learnings from the collaboration and which represent some of the current challenges in the area. Sufficient, high-quality FAIR data is key to building high quality models. When this is available excellent ML models can be built, however for complex endpoints like off-target toxicity data availability is a major limitation. Scoring compounds for synthetic accessibility was also noted as a challenge and this is a key feature to enable generative chemistry algorithms to propose 'reasonable' compounds.

These talks concluded an excellent and broad ranging day's science that highlighted developments and challenges in the technical aspects of applying AI to drug development but also the educational and cultural changes that are required for its successful implementation. On this aspect it seems fitting to conclude with a point that has been made before and was made again at this meeting: 'AI will not replace chemists but chemists using AI will replace those not using AI'.

Received: May 3, 2024

- [1] M. Xu and H. Chen, *J. Chem. Inf. Model.* **2023**, *63*, 7067, <https://doi.org/10.1021/acs.jcim.3c01626>.
- [2] Z. Fan, Y. Yang, M. Xu, H. Chen, arXiv: 2308.00237, <https://doi.org/10.21203/rs.3.rs-4005080/v1>.
- [3] L. Schoenmaker, O. J. M. Béguignon, W. Jespers, G. J. P. van Westen, *J. Cheminform.* **2023**, *15*, 22, <https://doi.org/10.1186/s13321-023-00696-x>.
- [4] W. Jespers, M. Esguerra, J. Åqvist, H. Gutiérrez-de-Terán *J. Cheminform.* **2019**, *11*, 26, <https://doi.org/10.1186/s13321-019-0348-5>.
- [5] W. Jespers, G. V. Isaksen, T. A. H. Andberg, S. Vasile, A. van Veen, J. Åqvist, B. Brandsdal, H. Gutiérrez-de-Terán, *J. Chem. Theory Comput.* **2019**, *15*, 5461, <https://doi.org/10.1021/acs.jctc.9b00538>.
- [6] M. Šicho, S. Luukkonen, H. W. van den Maagdenberg, L. Schoenmaker, O. J. M. Béguignon, G. J. P. van Westen, *J. Chem. Inf. Model.* **2023**, *63*, 3629, <https://doi.org/10.1021/acs.jcim.3c00434>.
- [7] M. Backenköhler, J. Groß, V. Wolf, A. Volkamer, <https://doi.org/10.26434/chemrxiv-2023-prk53>.
- [8] D. Sydow, E. Aßmann, A. J. Kooistra, F. Rippmann, A. Volkamer, *J. Chem. Inf. Model.* **2022**, *62*, 2600, <https://doi.org/10.1021/acs.jcim.2c00050>.
- [9] <https://projects.volkamerlab.org/teachopencadd/>
- [10] R. E. Jefferson, A. Oggier, A. Füglistaler, N. Camviel, M. Hijazi, A. R. Villarreal, C. Arber, P. Barth, *Nat. Commun.* **2023**, *14*, 2875, <https://doi.org/10.1038/s41467-023-38491-9>.
- [11] M. Alberts, F. Zipoli, A. C. Vaucher, <https://doi.org/10.26434/chemrxiv-2023-8wxcz>
- [12] R. Yan, Y. Li, J. Müller, Y. Zhang, S. Singer, L. Xia, X. Zhong, J. Gertsch, K.-H. Altmann, Q. Zhou, *Cell Discov.* **2021**, *7*, 16, <https://doi.org/10.1038/s41421-021-00247-4>.
- [13] L. A. Schneider, B. Sauter, K. Dagher, D. Gillingham, *J. Am. Chem. Soc.* **2023**, *145*, 20874, <https://doi.org/10.1021/jacs.3c05961>.
- [14] A. Volkamer, S. Riniker, E. Nittinger, J. Lanini, F. Grisoni, E. Evertsson, R. Rodríguez-Pérez, N. Schneider, *Artificial Intelligence in the Life Sciences*, **2023**, *3*, 100056, <https://doi.org/10.1016/j.aiisci.2022.100056>.