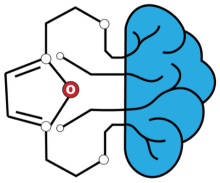
The Explainable AI for Molecules (AiChemist) Project

Newsletter #2, August 2024



The Explainable AI for Molecules (AiChemist) project is a Marie-Skłodowska-Curie Doctoral Network (DN) funded by the European Commission under the Horizon Europe Programme, Horizon-MSCA-2022 grant agreement number 101120466, which started on 1st September 2023. The network brings together sixteen academic and industry partners from 8 European countries in addition to the Korean Institute of Toxicology (South Korea) to train fourteen doctoral candidates (DCs) in close collaboration with associated partners based in Europe and the USA.

Follow us on Twitter: https://x.com/aichemist_dn

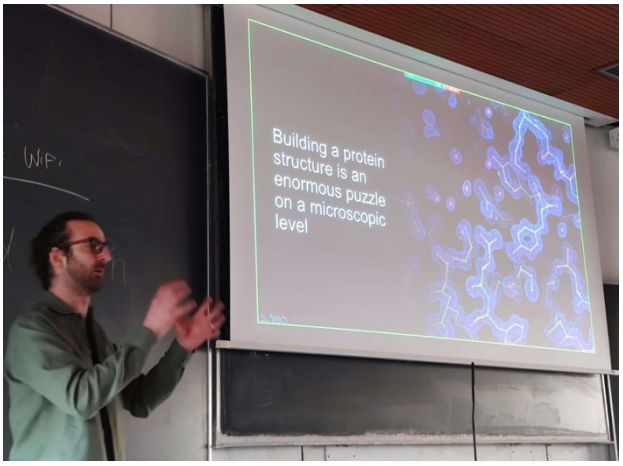
First AiChemist School – Spring School on Chemoinformatics and Machine Learning in Berlin

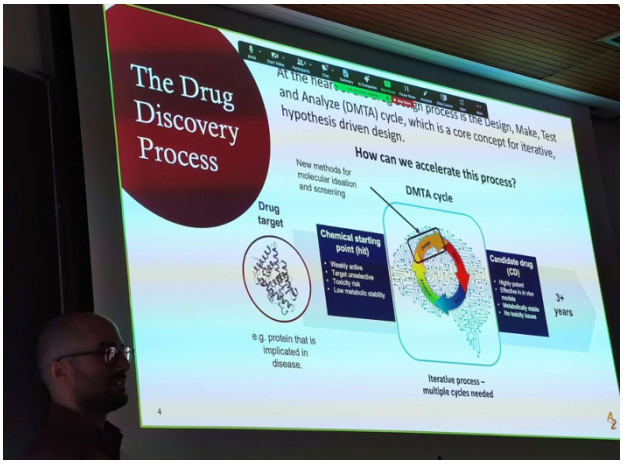
The first AiChemist School took place from March 4th to 15th at the Institute of Mathematics, Freie Universität Berlin, coinciding with the final AIDD School (https://ai-dd.eu). The combined Schools were jointly organised by Bayer and Pfizer – who are beneficiaries of both projects. The Institute of Mathematics of FUB is situated in the peaceful, residential Dahlem district in southwestern Berlin, which has been a center for high calibre research since the beginning of the 20th century, at one time being referred to as the "German Oxford".

The Spring School was attended by fellows of both the AiChemist and AIDD projects, as well as PIs, associated partners and guest speakers from various European countries, the US, Asia and Australia. The programme of the Spring School covered a diverse range of topics, such as chemoinformatics, various ML/AI methods and applications, GPU programming and also some contemporary experimental methods used in drug discovery pipelines. The vast majority of the lecture programme was delivered in person and most of the lectures were run open to the public and broadcasted over Zoom, attracting many participants outside of the AiChemist and AIDD consortia. Many of the lectures and workshops were delivered by speakers from Bayer and Pfizer, in addition to specially selected guest speakers from outside of Europe. The PIs of the AiChemist project formally introduced themselves to the fellows and the AIDD PIs, and a handful of PIs and Associated Partners from both projects gave lectures summarising their most recent work.

Nuclear fission was discovered in this building, at that time known as the Kaiser Wilhelm Institute for Physics. It is currently part of the FUB Dahlem Campus and was renamed the Hahn-Meitner Building in 2010. It is now used by biochemists. Image Credit: Bernd Wannenmacher. Image source: https://www.fu-berlin.de/presse/service/download/hahn-meitner-bau.jpg

The first day of the School (Monday the 4th March) kicked off with some short introductory presentations from AiChemist PIs, followed by a lecture from John O'Donnel, Head of the Structural Biology Laboratory at Bayer AG (photo on left), who gave an overview of the integrated approaches for structure determination used by his team, in the context of his previous work on the elucidation of the architecture and function of endoplasmic reticulum membrane protein complexes. The final lecture of the day focused on the computational chemistry workflow manager, *maize*, developed and used by Thomas Löhr (photo on right) and his team at AstraZeneca.





John O'Donnel (left) and Thomas Löhr (right) lecturing at the podium

At the end of the day, the students and PIs gathered for a drinks reception at Biermeisterei by Lemke, a beer hall and garden located in the iconic Alexanderplatz.

Lively scientific exchanges over local Berlin beers

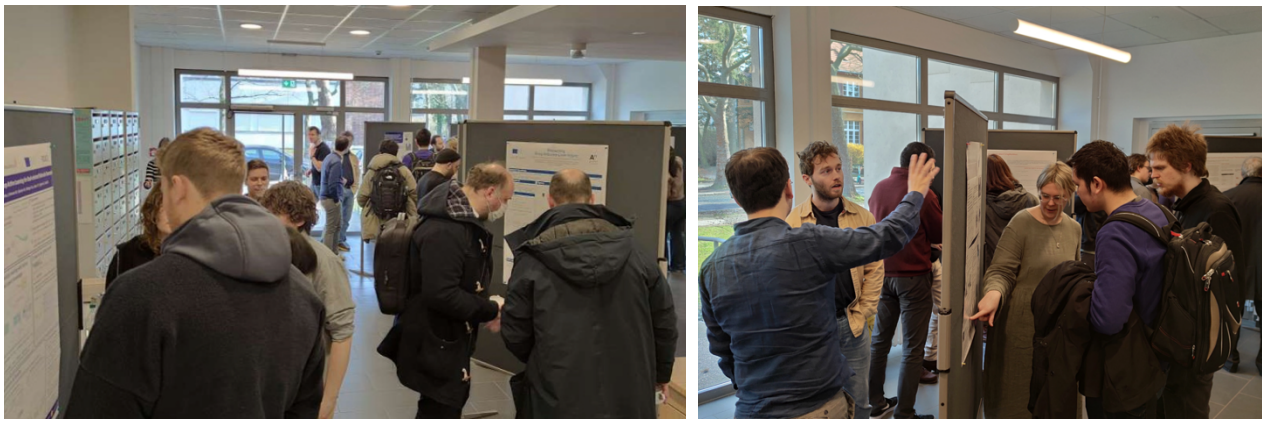
The second day consisted of a visit to Bayer AG's Berlin site in Müllerstraße, the global corporate headquarters of Bayer's pharmaceutical division. With around 5000 employees, the Bayer site in Berlin is one of the largest private employers in the city. The fellows and PIs were received and welcomed onsite by Dr. Matteo Aldeghi, Director of ML Research at Bayer, and his colleagues Dr. Marina Garcia de Lomana and Dr. Santiago Villalba. In addition to a series of lectures on computer aided drug design, toxicology, targeted radiotherapies, CRISPR and other life science technologies delivered by the Bayer team, the AiChemist students and PIs were given a historical tour of Bayer, and had an exclusive opportunity to visit the automated medicinal chemistry, compound logistics and targeted radiotherapies laboratories.

Students and PIs from the AiChemist and AIDD projects during their visit to Bayer

On the third day, the school continued back at FUB, with workshops on QSAR and diverse applications of ML/AI on the agenda. The first workshop, delivered by Dr. Gilles Marcou from the University of Strasbourg, covered the recent developments in KNIME Analytics Platform in the context of QSAR modelling, with a focus on standardisation of molecules. During the second workshop, Prof. David Winkler from La Trobe University illustrated strategies for building and refining ML/DL models to solve a diverse array of problems in the life sciences: from studying the origins of SARS-CoV-2, repurposing drugs to treat SARS-CoV-2, to designing novel antibiotics to address the antibiotic resistance crisis, to designing novel polymeric substrates for growing stem cells.

Workshops by Gilles Marcou (left) and David Winkler (right)

On Thursday morning, Dr. Alessandra Roncaglioni, head of the computational toxicology research unit at IRFMN (AIDD Associated Partner and AiChemist Beneficiary) spoke about some of her recent work applying ML to predict drug cardiotoxicity within the AOP framework. The second morning lecture was given by Prof. Alexandre Tkatchenko of the University of Luxembourg, who discussed hybrid QM/ML -based approaches for inverse molecular design. For the afternoon session, Prof. Aixia Yan of the Beijing University of Chemical Technology (Associated Partner of AIDD) gave a lecture on applications of ML for predicting the bioactivities of BTK, CDK4 and FLT3 inhibitors, as well as her groups' recent efforts to design novel EGFR inhibitors with an ML-driven computational screening cascade pipeline. The last lecture of the day, given by Dr. Robin Winter, a senior ML researcher at Pfizer, focused on multi-objective optimization in continuous latent spaces. The day was rounded off with a poster session, during which the AIDD fellows discussed the recent advances within their projects with PIs, invited speakers and AiChemist fellows.



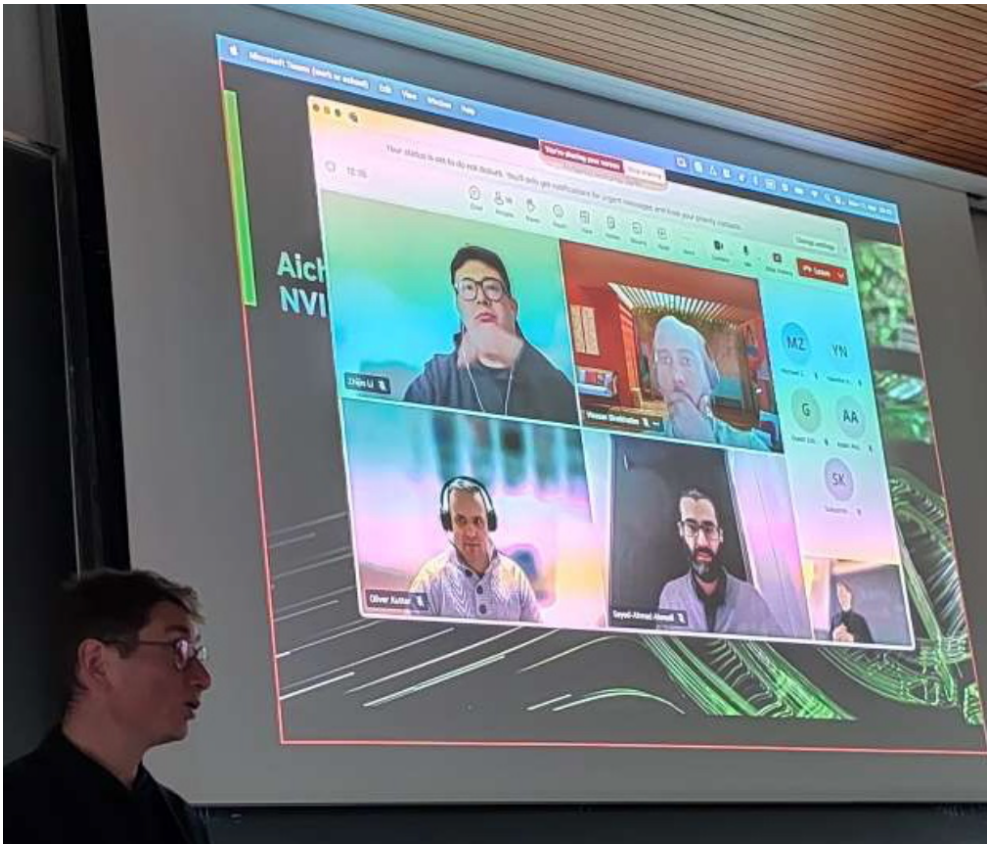
The AIDD poster session on Thursday afternoon

The programme continued on Saturday, starting with a morning round-table discussion between the AiChemist and AIDD students at The Social Hub, Alexanderplatz, during which the AiChemist fellows gave some flash talks introducing themselves and their plans for their research projects, and AIDD students shared their experiences as MSCA fellows with the new cohort of fellows, giving them some practical advice, wisdom and helpful tips for getting started with their projects, collaborating effectively and overcoming hurdles along the way. After lunch, the fellows set off to Teufelsberg Street Art Gallery together, where, in celebration of International Women's Day, they got to experience some truly unique feminist performance art in addition to an extensive collection of street art.

Street art, feminist performance art and breathtaking Berlin views on the Saturday

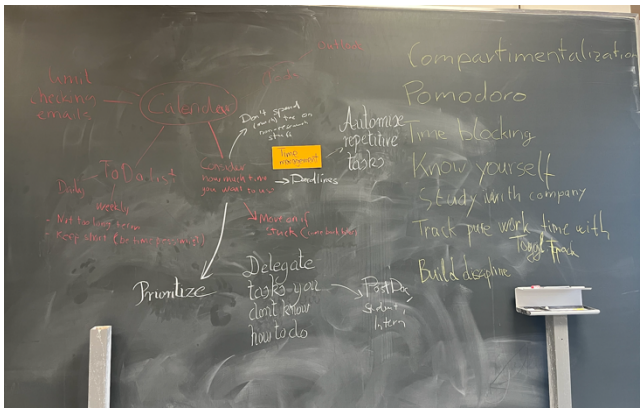
On Monday 11th March, the fellows and PIs reconvened back at FUB for a high-level GPU programming workshop organised by NVIDIA. David Ruau, Head of Strategic Alliances, Drug Discovery AI EMA at NVIDIA, who attended in person, gave a fantastic introduction to NVIDIAs current suite of AI-powered drug discovery platforms and data science libraries, and the capabilities of their hardware. David's remote team, dialing in from various parts of Europe, gave some in-depth demonstrations of NVIDIAs BioNeMo Framework, the RAPIDS cuML suite of libraries for implementing GPU-accelerated ML algorithms through a scikit-learn API, and finally, accelerating model inference with TensorRT.

The workshop attracted 15 online participants outside of the AIDD and AiChemist Consortia and was very well received by early-career and senior researchers alike.



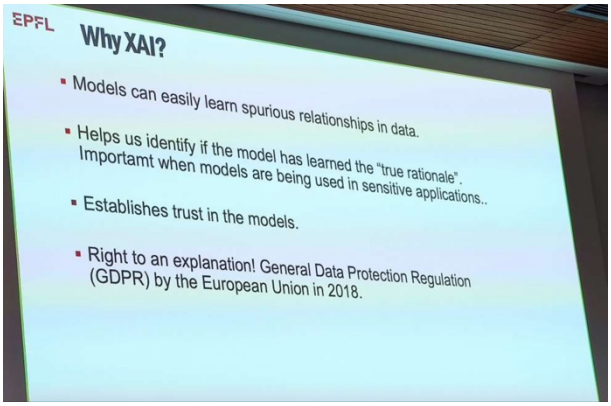
David Ruau introduces his team – Oliver Kutter (bottom left), Zhijin Li (top left), Hassan Sirelkhatim, Sayeh-Ahmad Ahmadi.

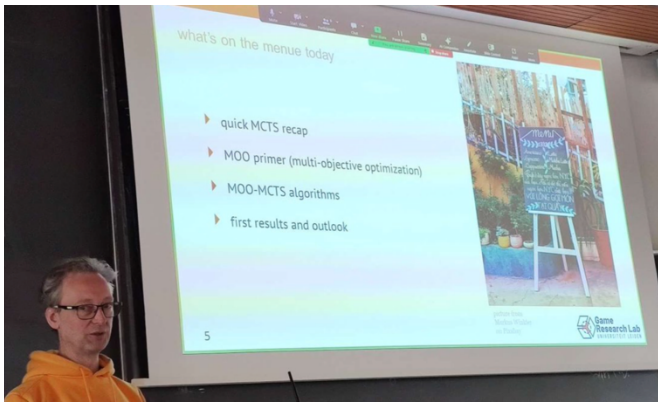
The Monday afternoon session was dedicated to project management for researchers. Dr. Alexander Egeling gave the fellows an overview of the key tools and basic principles of project management, in particular Agile Project Management, and explained they are so valuable for young researchers in particular. The fellows took some time to think about which project management principles they were already unknowingly applying in their daily work, and how they could streamline them better through dedicated tools, methodologies and software solutions.



Alexander Egeling leading the project management workshop

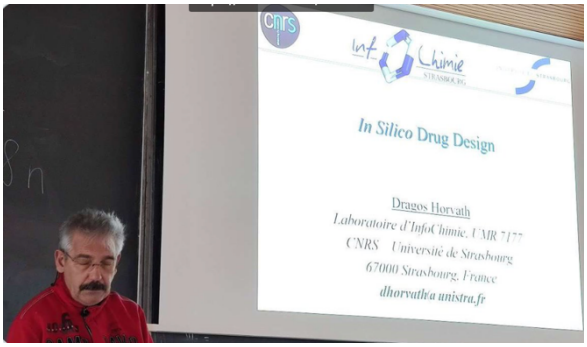
On Tuesday morning, the program continued with stimulating lectures on eXplainable AI for Chemistry and Monte Carlo tree search and its multiobjective variants, which drew a particularly large and highly engaged online audience over Zoom. For the afternoon session, the AIDD fellows took to the podium and gave 10 minute flash talks summarising their PhD projects.

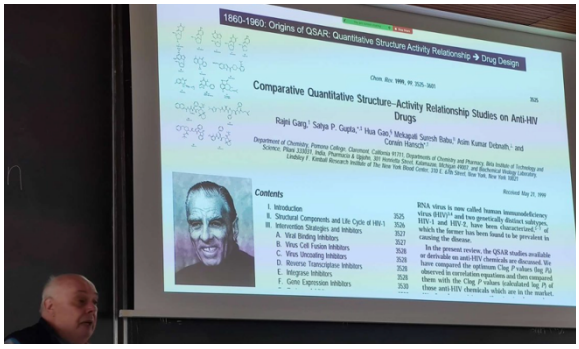




Dr. Geemi Wellawate, postdoctoral researcher at EPFL, spoke about XAI for Chemistry, before Associate Prof. Mike Preuss from Leiden University delivered his lecture on Monte Carlo tree search and multiobjective optimization

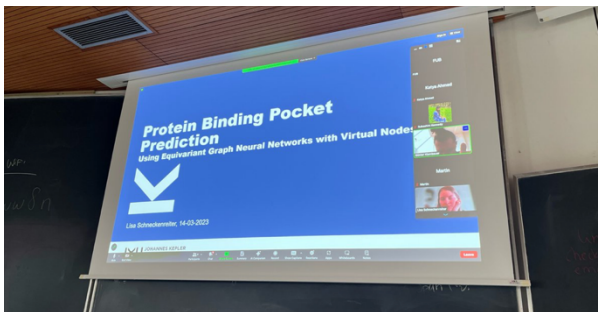
The following day's lectures focused on the foundations of chemoinformatics and ML for drug discovery. Dragos Horvath from the University of Strasbourg started with an overview of the most commonly used methods for in silico drug discovery and the importance of descriptor choice and model validation in QSAR. This was followed on with a recap of the history of QSAR, an overview of linear regression models and a practical guide to best practice in machine learning, delivered by Mark Embrechts, guest professor at Helmholtz Munich. To round off the day, Sven Giese of Bayer led a hands-on tutorial on unit testing, code structuring and documentation.





Dr. Dragos Horvath (left) gave an overview of in silico drug discovery, while Prof. Mike Embrechts (right) went over the history of QSAR and tips for best practice in ML.

The penultimate day of the school kicked off with virtual talks by Günter Klambauer and Lisa Schneckenreiter from Johannes Kepler University Linz, on the theory behind graph neural networks and they can be applied to binding pocket prediction problems, with a special focus on equivariant graph neural networks. After a short break, Dragos Horvath was welcomed back on to the podium to give a lecture on various methods for navigating chemical space, such as the use of generative topographic maps and autoencoders. After lunch, AiChemist PIs Marc Bianciotto and Hyun-Kil Shin gave short lectures introducing themselves and some of their most recent work to the fellows, and the final session of the day was dedicated to a tutorial on deep learning with Pytorch, delivered by Bayer colleague Adrien Bitton.



Prof. Günter Klambauer and Lisa Schneckenreiter gave a virtual lecture on graph neural networks, which attracted ~12 attendees outside of the consortium, before AiChemist partners Dr. Hyun Kil Shin (Korea Institute of Toxicology, top right) and Dr. Marc Bianciotto (Sanofi, bottom right) introduced themselves and their research activities to the consortium.

Dr. Adrien Bitton (left) and Dr. Gregory Landrum (right) rounded off the programme with technical workshops

The final day of the school was devoted to an in-depth exploration into the functionality of Rdkit, led by its leading developer Gregory Landrum, who dialed in from Zürich to deliver a workshop on the widely used chemoinformatics toolkit. After one last group lunch together, the AiChemist School came to a close.

Student Voices from the Berlin School

"I enjoyed my experience at this joint AiChemist/AIDD PhD school hosted by FUB in Berlin. As a last year PhD student in the AIDD project, I found it really good to have the opportunity to interact with the newly hired students from the AiChemist project. Their fresh perspectives gave a new energy to our discussions. Moreover, the poster presentations were a fantastic opportunity to showcase our research endeavours and exchange ideas with peers and industry professionals. It was nice to see the diverse array of projects being pursued within our project, underscoring the breadth and depth of talent within the AIDD consortium. Our visit to Bayer Pharma was also a highlight of the school, offering insights into real-world applications of our research. I also really enjoyed the visit to the Teufelsberg Street Art Festival, one of the coolest things I have seen in Berlin. In summary, the school was well-planned, offering a good blend of academic and industrial engagement, as well as networking opportunities."

- Alessio Fallani, AIDD fellow, Janssen

"I thoroughly enjoyed my time at the Berlin Spring School. As a Ph.D. candidate in the AiChemist programme who is new to cheminformatics and drug development, the school served as a valuable introduction to the field's key methodologies. Meeting fellow participants was particularly rewarding, laying the groundwork for academic connections I am excited to develop further. The city of Berlin was a great backdrop to the event, allowing for exploratory trips around the city. A highlight was our visit to Bayer's headquarters, where we saw some of their research facilities and heard about the company's history. I visited the radiotherapy lab, which was a very cool and educational experience I won't soon forget. The combination of lectures, engaging tours, Berlin, and meeting my new colleagues made this a great experience."

- Karoline Schjelde, AiChemist fellow

“This year's AIDD spring school was a valuable experience. Numerous interesting lectures provided important impulses for my future research. Two presentations that I found especially interesting were: the talk on Multimodal and Multiobjective Optimisation by Robin Winter and "Choosing an algorithm, descriptors, and approach for diverse applications of AI and ML" by David Winkler. The highlight of the spring school was definitely the poster session, providing unique insights into several areas of research.”

- Vincent Alexander Scholtz, PhD student, University of Vienna

Meet AiChemist consortium members at the following conferences (July - December 2024)

- IEEE Conference on Games, *5th - 8th August*, Milan, Italy, Mike Preuss (ULEI).
- Scandinavian Community for Quantum Chemistry Meeting, 12th – 14th August, Odense, Denkmark, Jan Jensen (UCPH).
- KDD 2024, *25th – 20th August*, Barcelona, Spain, Andrea Hunklinger (CRG).
- Interspeech 2024, 1st – 5th September, Kos, Greece, Michael Wand (USI).
- EUROTOX 2024, 8th – 11th September, Copenhagen, Denmark, Igor Tetko (HMGU).
- ETH – Microsoft Workshop on High-Throughput Computational Chemistry, 11th September, Zurich, Switzerland, Jan Jensen (UCPH).
- ICANN2024, 17th – 20th September, Lugano, Switzerland. The AiChemist consortium will be co-hosting a workshop alongside the AIDD project : AI in Drug Discovery!
- SETAC Asia-Pacific 14th Biennial Meeting, 21st – 25th September, Tianjin, China, Igor Tetko (HMGU).
- EuroQSAR, 22nd – 26th September, Barcelona, Spain, Ola Engkvist (AstraZeneca).
- MATHIAS2024, 23rd – 26th September, Paris, France, Gilles Marcou (UNISTRA).
- Elrig Drug Discovery, 2nd – 3rd October, London, UK, Olga Engkvist (AstraZeneca).

Publications from the AiChemist consortium

(1) Hartog, P.; Krüger, F.; Genheden, S.; Tetko, I.V. Using test-time augmentation to investigate explainable AI: inconsistencies between method, model and human intuition. *Journal of Cheminformatics*. 2024. https://doi.org/10.1186/s13321-024-00824-1

(2) Kopp, A.; Hartog, P.; Šícho, M.; Godin, G.; Tetko, I. The openOCHEM consensus model is the best-performing open-source predictive model in the First EUOS/SLAS Joint Compound Solubility Challenge. *SLAS Discovery*. 2024. https://doi.org/10.1016/j.slasd.2024.01.005