The Advanced Machine Learning for Innovative Drug Discovery (AIDD) project

Newsletter 8, 22nd April 2024

The **Advanced Machine Learning for Innovative Drug Discovery** (AIDD) project is a Marie-Skłodowska-Curie Innovative Training Network (ITN) for Early Stage Researchers (ESRs) funded by the European Commission under the Horizon 2020 Programme, Marie Skłodowska-Curie grant agreement No 956832. The project brings together fifteen academic and industry beneficiaries from ten European countries and the University of British Columbia (Canada) to train sixteen PhD students in close collaboration with partners from the USA, Australia, China, Israel, and other countries.

The AIDD project started on 01.01.2022 and will finish at the end of March 2025.

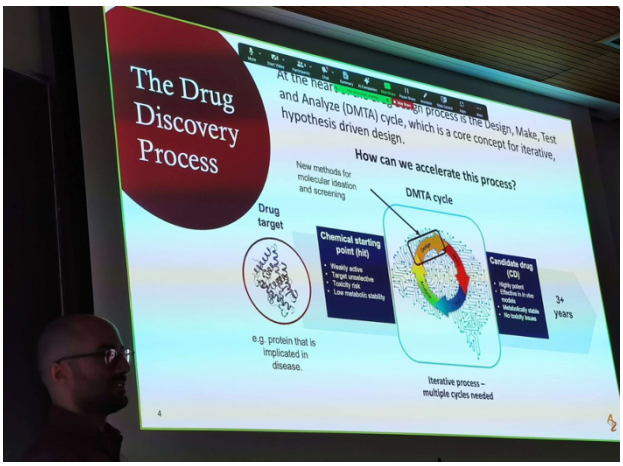
Final AIDD School – Spring School on Advanced Machine Learning in Berlin

The final AIDD School took place from March 4th to 12th at the Institute of Mathematics, Freie Universität Berlin, coinciding with the first AiChemist School (https://aichemist.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 AIDD and AiChemist projects , and well as PIs, associated partners and guest speakers from various european countries, the US, Asia and Australia. The program of the AIDD portion of the Spring School was dedicated to advanced ML/AI methods, in addition to contemporary and cutting-edge experimental methods used in drug discovery pipelines. The vast majority of the lecture program 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 AIDD and AiChemist 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 introduced themselves to the fellows and the AIDD PIs, and 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 architechture 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

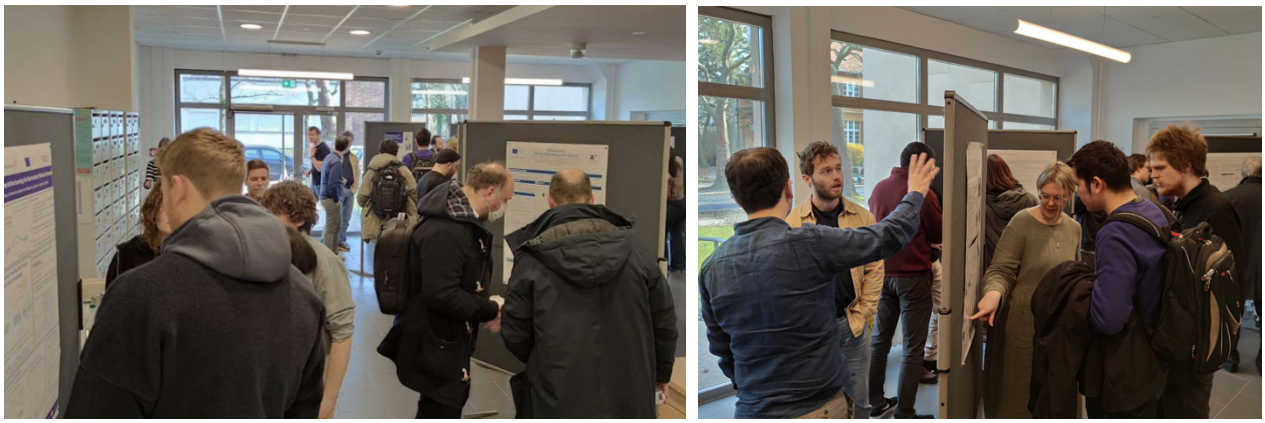
The second day consisted of a visit to Bayer AG's Berlin site in Müllerstraße, the global corporate headquarters of Bayer's pharmaceutical division. With around 5000 employees, the Bayer site in Berlin is one of the largest private employers in the city. The fellows and PIs were received and welcomed 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 AIDD 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 AIDD and AiChemist projects during their visit to Bayer

On the third day, the school continued at 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 bioactivies of BTK, CDK4 and FLT3 inhibitors, as well as her groups recent effors 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.



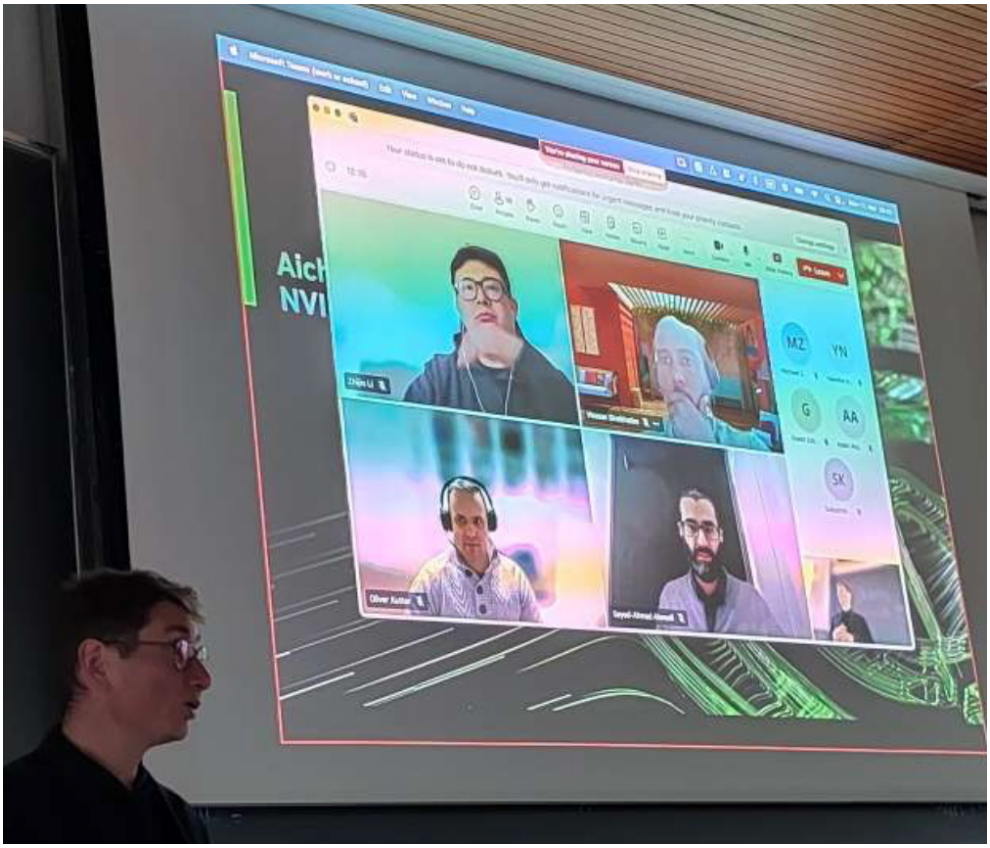
The AIDD poster session on Thursday afternoon

The programme continued on Saturday, starting with a morning round-table discussion between the AIDD and AiChemist 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 Womens 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

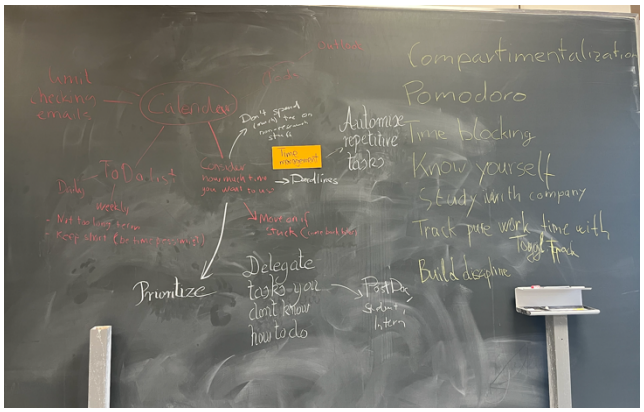
On Monday 11th March, the fellows and PIs reconvened back at FUB for a high-level GPU programming workshop organised by NVIDIA. David Ruau, Head of Strategic Aliiances, Drug Discovert AI EMA at NVIDIA, who attended in person, gave a fantastic introduction to NVIDIAs current suite of AI-powered drug discovery 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.



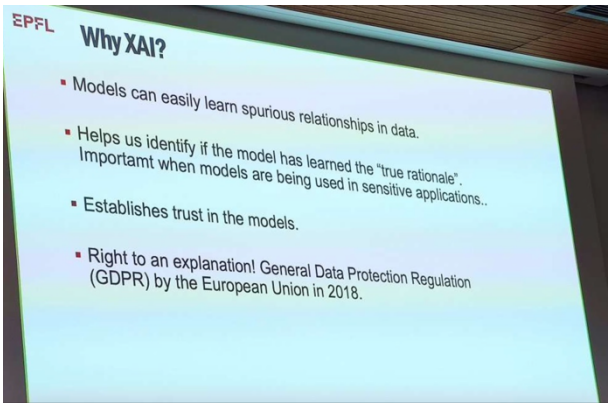
David Ruau introduces his team – Oliver Kutter (bottom left), Zhijin Li (top left), Hassan Sirelkhatim, Sayeh-Ahmad Ahmadi

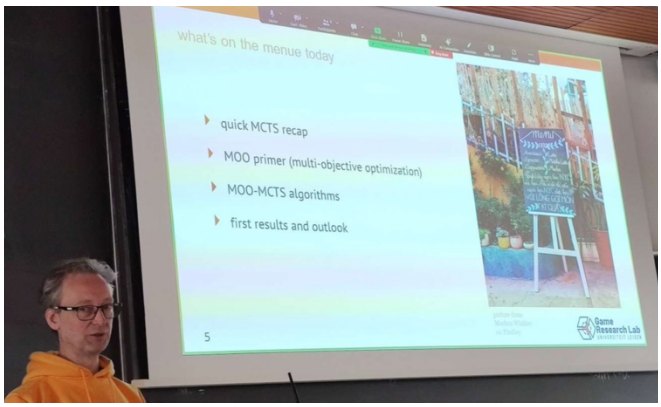
The Monday afternoon session was dedicated to project management for researchers. Dr. Alexander Egeling gave the fellows an overview of the key tools and basic principles of project management, in particular Agile Projecty Management, and explained they are so valuable for young researchers in particular. The fellows took some time to think about which project management principles they were already unknowlingly 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

The AIDD school came to an end on Tuesday 12th March, rounded off nicely 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. To wrap up the day, the AIDD fellows took to the podium and gave 10 minute flash talks summarising their PhD projects, and with that, the AIDD Berlin School came to a close.





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

Student Voices from the Berlin School

"I enjoyed my experience at this last 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 program 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 starting in July 2024

“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 Wrinkler. 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

Conferences attended by AIDD consortium members (March - September 2024)

- International Leopoldina Symposium on Molecular Machine Learning, *4th - 6th March*, Halle, Germany, Ola Engkvist
- Workshop: NAMs - Use and application of QSAR and read-across, 22nd March, Paris, France.
- ACS Spring, *17th - 24th March*, New Orleans, USA, Varvara Voinarovska.
- Spring Meeting of the Swiss Society of Pharmacology and Toxicology, *18th April*, Bern, Switzerland, Igor Tetko.
- ICLR 2024, *6th - 12th May*, Vienna, Austria, Julian Cremer.
- The third International Conference on Hybrid Human-Artificial Intelligence, *10th-14th June,* Malmö, Sweden, Yasmine Nahal.
- CVPR 2024, *17th - 21st June*, Seattle WA, USA, Adam Arany.
- Strasbourg Chemoinformatics Summer School, *24th - 28th June*, Strasbourg, France, Yasmine Nahal, Vincenzo Palmacci, Igor Tetko.
- ICCS 2024, 2nd July, Malaga, Spain, Vincenzo Palmacci.
- ICML 2024, *21st - 27th July*, Vienna, Austria, Vincenzo Palmacci, Mike Preuss, Adam Arany, Rosa Friesacher, Emma Svensson.
- IEEE Conference on Games, *5th - 8th August*, Milan, Italy, Mike Preuss.
- ICANN2024, 17th – 20th September, Lugano, Switzerland. The AIDD consortium will be hosting their own workshop: AI in Drug Discovery!

Publications by AIDD fellows

(1) Svensson, E., Hoedt, P.-J., Hochreiter, S., Klambauer, G. Task-conditioned modeling of drug-target interactions. In ELLIS Machine Learning for Molecules Discovery Workshop. November 28, 2022. https://moleculediscovery.github.io/workshop2022/
(2) Hassen, A. K.; Torren-Peraire, P.; Genheden, S.; Verhoeven, J.; Preuss, M.; Tetko, I. V. Mind the Retrosynthesis Gap: Bridging the Divide between Single-Step and Multi-Step Retrosynthesis Prediction; 2022. https://openreview.net/forum?id=LjdtY0hM7tf
(3) Andronov, M.; Voinarovska, V.; Andronova, N.; Wand, M.; Clevert, D.-A.; Schmidhuber, J. Reagent Prediction with a Molecular Transformer Improves Reaction Data Quality. *Chem. Sci.* 2023. https://doi.org/10.1039/D2SC06798F
(4) Radaeva, M.; Ban, F.; Zhang, F.; LeBlanc, E.; Lallous, N.; Rennie, P.S.; Gleave, M.E.; Cherkasov, A. Development of Novel Inhibitors Targeting the D-Box of the DNA Binding Domain of Androgen Receptor. *Int. J. Mol. Sci.* 2021, *22*, 2493, https://doi.org/10.3390/ijms22052493
(5) Cremer, J.; Medrano Sandonas, L.; Tkatchenko, A.; Clevert, D.A.: De Fabritiis, G. Equivariant Graph Neural Networks for Toxicity Prediction. *Chem. Res. Toxicol.* 2023. https://doi.org/10.1021/acs.chemrestox.3c00032
(6) Sarkis, M.; Fallani, A.; Tkatchenko, A. Modeling Non-Covalent Interatomic Interactions on a Photonic Quantum Computer. *Physical Review Research*. 2023. 10.1103/PhysRevResearch.5.043072
(7) Sanchez-Fernandez, A.; Rumetshofer, E.; Hochreiter, S.; Klambauer, G. CLOOME:contrastive learning unlocks bioimaging databases for queries with chemical structures. *Nature Communications*. 2023. https://doi.org/10.1038/s41467-023-42328-w
(8) Le, T.; Cremer, J.; Noé, F.; Clevert, D-A.; Schütt, K. Navigating the design space of equivariant diffusion-based generative models for de novo 3D molecule generation. arXiv. 2023. https://doi.org/10.48550/arXiv.2309.17296
(9) Voinarovska, V.; Kabeshov, M.; Dudenko, D.; Genheden, S.; Tetko, I.V. When Yield Prediction Does Not Yield Prediction: An Overview of the Current Challenges. *J. Chem. Inf. Model.* 2024. https://doi.org/10.1021/acs.jcim.3c01524
(10) Andronov, M.; Andronova, N., Wand, M., Schmidhuber, J., Clevert, D-A. A reagent-driven visual method for analyzing chemical reaction data. arXiv. 2024. https://doi.org/10.26434/chemrxiv-2024-q9tc4
(11) 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
(12) Torren Peraire, P.; Hassen, A.K.; Genheden, S.; Verhoeven, J., Clevert, D-A.; Preuss, M.; Tetko, I.V. Models Matter: the impact of single-step retrosynthesis on synthesis planning. *Digital Discovery*. 2024. https://doi.org/10.1039/D3DD00252G
(13) 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 (14) Bernatavicius, A.; Šícho, M.; Janssen, A.; Hassen, A.K.; Preuss, M.; van

Westen, G. AlphaFold meets de novo drug design: leveraging structural protein information in multi-target molecular generative models. ChemRxiv. 2024. https://doi.org/10.26434/chemrxiv-2024-60tc7
(15) Ha, S.; Leuschner, L.; Czodrowski, P. FSL-CP: a benchmark for small molecule activity few-shot prediction using cell microscopy images. *Digital Discovery.* 2024. DOI: 10.1039/d3dd00205e
(16) Svensson, E., Hoedt, P.-J., Hochreiter, S., Klambauer, G. HyperPCM: Robust Task-Conditioned Modeling of Drug-Target Interactions. *J. Chem. Inf. Model.* 2024. *https://doi.org/10.1021/acs.jcim.3c01417*
(17) Hassen, A.K.; Šícho, M.; van Aalst, Y.J.; Huizenga, M.C.W.; Reynolds, D.N.R.; Luukkonen, S.; Bernatavicius, A.; Clevert, D-A.; Janssen, A.P.A.; van Westen, G.J.P.; Preuss, M. Generate What You Can Make: Achieving in-house synthesizability with readily available resources in de novo drug design. ChemRvix. 2024. https://doi.org/10.26434/chemrxiv-2024-wtjt6
(18) Tan, L; Hirte, S.; Palmacci, V.; Stork, C.; Kirchmair, J. Tackling assay interference associated with small molecules. *Nature Reviews Chemistry*. 2024. https://doi.org/10.1038/s41570-024-00593-3

Additional information

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Actions, grant agreement No 956832.

Follow us at Twitter: https://twitter.com/AiddOne