AI 4 Science Discovery Network+

AI4SD Interview with Dr Barbara Zdrazil
02/12/2021
Online Interview

Michelle Pauli
Michelle Pauli Ltd

30/06/2022

Humans-of-AI4SD:Interview-23

AI4SD Interview with Dr Barbara Zdrazil
Humans-of-AI4SD:Interview-23
30/06/2022
DOI: 10.5258/SOTON/AI3SD0224
Published by University of Southampton

Network: Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery
This Network+ is EPSRC Funded under Grant No: EP/S000356/1

Principal Investigator: *Professor Jeremy Frey*
Co-Investigator: *Professor Mahesan Niranjan*
Network+ Coordinator: *Dr Samantha Kanza*

Contents

1 Interview Details **1**

2 Biography **1**

3 Interview **2**

1 Interview Details

Title	AI4SD Interview with Dr Barbara Zdrazil
Interviewer	MP: Michelle Pauli - MichellePauli Ltd
Interviewee	AC: Dr Barbara Zdrazil - EMBL-EBI & University of Vienna
Interview Location	Online Interview
Dates	02/12/2021

2 Biography

Figure 1: Dr Barbara Zdrazil

Dr Barbara Zdrazil: 'I have always believed in the power of AI, but I was surprised that it led to such a high success rate in virtual screening"

Barbara Zdrazil works as ChEMBL Coordinator at the European Bioinformatics Institute (EMBL-EBI). Prior to this role she was a group leader at the University of Vienna focusing on integrating data science approaches into the computational molecular design process. Barbara also contributes to Open Targets, a project which aims to enable systematic target identification and prioritisation and is Co-Editor in Chief of the Journal of Cheminformatics.

In this Humans of AI4SD interview she discusses how chemists can generate new knowledge from data, moving away from the "one drug, one target" paradigm, how Covid has boosted collaboration, and what she considers important to pursue a career in research.

3 Interview

MP: What's been your path to where you are today?

BZ: After finishing school, I knew I wanted to do a degree in natural sciences, but I couldn't decide which discipline. I wanted to study something that offers a broad variety of different disciplines, which is how I discovered pharmacy. Later, I realised it was pharmaceutical chemistry and pharmacology that I found most interesting.

I did my Master thesis on an Erasmus year in Spain where I did chemical synthesis in the lab. I wanted to do a PhD in the chemistry field as well and I was told about Gerhard Ecker, a 'young professor doing innovative things on the computer'. I started my PhD in his lab doing computational chemistry, mainly focusing on ligand-based methods, like QSAR analyses, applied to ABC transporters. After that, I went for my postdoc to Germany, where I stepped into structure-based modelling: molecular dynamic simulations involving enzymes, not transporters, which I worked on in my PhD. It was very important for me to get this additional knowledge from structure-based modelling as well. After this first postdoc, I was recruited back to the University of Vienna, where I could apply all my skills in ligand- and structure-based modelling.

Later, I got involved in two large EU-funded IMI projects. The first one was Open PHACTS, which was a five-year-long public-private partnership that the University of Vienna was leading as academic coordinator. The aim of this project was to create a semantically integrated hub for life science data sources: a platform that could lower the barrier to access drug discovery data. This is when my passion for data-driven approaches began. The second project was EU-ToxRisk, where we focused on toxicity testing and predictions. We wanted to make a difference by replacing animal methods with in vitro-based and in silico-based test methods.

In 2016, I got my own project funded by the Austrian Science Fund (FWF), and that's when my independent career really started. In this project we were focusing on the modelling of hepatic organic anion-transporting polypeptides. These are important transporters located at the basolateral membrane of hepatocytes. It's important to understand how they work and what's their molecular basis for ligand binding and selectivity, which both influence the pharmacokinetics of drugs.

Due to my knowledge in computational toxicology, I got interested in a position at the European Bioinformatics Institute focusing on target safety. Since January 2021, I've been working for EBI in the Open Targets project which is a public-private partnership using human genetics and genomics to prioritise and identify drug targets. In the project, I was collating target safety information to be integrated into the platform. There's a first version already online showing information about safety liabilities of more than 500 different protein targets.

This month (May 2022) I started to work as ChEMBL Coordinator at EBI. I am very excited about this new role, especially about the possibility to work in such a great team of talented people. ChEMBL is one of the most heavily used open bioactivity databases globally. Being able to serve the community as part of this role I envisage to be very rewarding.

This is also true for my role as Co-Editor in Chief of the Journal of Cheminformatics. The editor role also allows me to identify current trends but also needs in the field of

Cheminformatics. We are currently working on new initiatives to address diversification in the field of Cheminformatics which are very much needed in order to drive the field forward.

MP: What drew you to data science in particular?

BZ: I was always fascinated by the possibilities that computation offers given a data set of small molecules and associated bioactivity values. Since my PhD, I have been working with ligand-based drug design approaches, initially focusing on chemical similarity. The basic similarity principle – stating that chemically similar molecules will likely exert similar biological effects – still has many applications in drug discovery and toxicology.

One of my research foci involves extracting time trends in data available in the corpus of medicinal chemistry literature. For instance, we have looked at scaffold trends and target trends over time. The ultimate goal here is to learn from the trends in published data and how can we use this data to predict future trends in drug discovery.

What I think is a little underestimated to date is pure data analysis – what you can learn from the data by examining the distribution of certain features in your data, by extracting trends in your data, and so on. Of course, data analysis doesn't appear as fancy as generating high-performing predictive models, but I was always compelled by the amount you can learn from just looking at your data more closely.

MP: What does success look like with the Open Targets project?

BZ: Open Targets aims to systematically identify and organise drug target information. The ChEMBL team also helps to improve the platform further. One example is information on target tractability, which is now being included, as well as a target safety information.

Finally, having a platform that allows you to start from a target and distil disparate types of information about that target, like information about related diseases, but also drug information and safety information and so on, is a very handy thing to have. When you start a drug discovery campaign you can first check all this information about your target of interest without having to look for it at various different places.

MP: What challenges are you facing with this work?

BZ: With respect to our work on target safety, the challenges we're facing are best understood if we think of this problem as a triangle connecting drugs-targets-adverse effects. If a drug or small molecule in general triggers a safety event, then this is often happening via the interaction with one or multiple protein targets. In this project we are interested in establishing the connections between protein targets and adverse events, because these are often unknown. More often we just know about the side effects or adverse effects that a drug is causing but not the biological mechanism underlying these events.

Moreover, we need to move away from the "one drug, one target" paradigm, in order to understand the biological event as a network of interacting biological entities. Rarely is a drug only hitting one target, it often shows activity on many targets. We need to differentiate between the therapeutic target and the "off-target", which often leads to a safety liability. Sometimes the "on-target" can also lead to safety liabilities, like in the case of the hERG channel. Here, the drugs that are acting as therapeutics can also lead to side effects caused by the interaction with the hERG channel. The challenge is to interconnect and gather all the

disparate information properly. How do we differentiate "on" and "off" targets? There isn't really a good way to do that at the moment.

MP: What has surprised you in your work?

BZ: In a successful collaboration with Gerard van Westen's group at Leiden University, we combined AI-based methods with molecular docking to generate a virtual screening pipeline in order to detect new inhibitors for hepatic organic anion transporting polypeptides in a big vendor library. What surprised me was that the combination of proteochemometric modelling, together with molecular docking, led to a selection of six compounds which, when we tested them, were all active in the micromolar to nanomolar range. One showed the potency of 40 nanomolar, which is comparable to the most potent compound currently known for OATP2B1. I have always believed in the power of AI, but I was still surprised that this led to such a high success rate!

MP: How is AI changing how we do science?

BZ: I think that the rise of AI-based methods has led to a change in how efficiently we can make use of openly available data. Using such methods as part of your research will enable you to establish hypotheses faster. Yet it is unclear to date if AI-based methods will really accelerate the drug discovery process to a significant extent – it is still too early to confidently assess this.

The danger in my opinion is that we now have such sophisticated machine learning models that some of them will have a tendency to produce a good model while interpretability is poor. What use is a good model if you don't understand what your model is doing? My advice would be to go for a simpler method if you can, even if its predictive power is a bit lower, as long as it's still interpretable. These days researchers start exploring explainable AI, which in contrast to "black box" models aims to understand why a specific decision is reached by a model.

MP: Where would you say the Open Science movement is at the moment?

BZ: In the Journal of Cheminformatics, where I'm an editor, we only publish research articles that report reproducible science. This means that code and data needed to reach the conclusions of the paper have to be made openly available.

People have mostly accepted the need for reproducible open science. Despite our strict requirements, we get more and more article submissions every month. My hypothesis is that through the open science requirement we attract even higher quality work.

However, in general the open science movement is a slow process still which requires not only efforts undertaken by journals, but also by research organisations, and the whole community as such. As long as we measure research impact mainly by journal impact factors and citation metrics, making data openly available for free isn't an attractive thing to do. We need to think about reward systems that directly benefit the data depositors' careers and reputation.

MP: How has the Covid-19 pandemic affected this aspect of the work?

BZ: During this crisis, people have started to be more collaborative, and to share their data and models with greater willingness and ease. For instance, in the Covid Moonshot project,

300 or so researchers are collaborating to accelerate the design-make-test cycle in search for Covid-19 therapeutics. The pandemic has sparked collaboration and open science efforts. I'm just hoping that when this pandemic is over, researchers will continue to collaborate in this effective manner.

Another positive aspect is that many companies and universities now tend to accept more flexible working arrangements. I think such a change in working policies was urgently needed. Such greater flexibility helps, for instance, parents who are facing difficulties during their careers when trying to balance out a family life with their jobs.

My co-workers have been working just as effectively during these past two years as they did before the pandemic started. I don't see any problems of declining productivity during home office days – not for myself and not for team members.

MP: What advice would you give to those looking to pursue a career in research today?

BZ: I met many students who felt pressured because of the competition they are facing. They need to apply for many PhD positions before they actually get selected. This is a quite frustrating process for many of them. The advice I usually give is to believe in yourself, as cheesy as this sounds. Also, I think a career path is never linear, so you can't plan it all out. Be open about new or unexpected possibilities and when the right moment comes, just go for it!

Also, don't let other people decide what is the right path for yourself. If you are passionate about and want to work in a certain scientific domain, you should focus on your goal and go for it. I would hire a person that is passionate about what they are doing. It's about commitment in the end — these are the people you want to employ and support further. So, if you really like what you're doing, you will be successful — just don't think there's a linear path to a career, it's a multi-step process that can't always be planned out linearly.

Disclaimer: The opinions raised by Barbara Zdrazil are her own and do not necessarily reflect the opinions of her employers.