

**“There are millions of proteins in nature that we still don’t know,”
says bioinformatician Monika Rosinská**

Only a few people can truly navigate the intersection of two such different — and equally demanding — worlds as computer science and the life sciences. It requires analytical thinking, deep knowledge of both fields, and above all, the ability to uncover new connections.

Trained as a back-end developer, Monika Rosinská began studying environmental sciences two years ago, and she couldn’t be happier about combining these two disciplines. She works on developing tools for protein analysis and design, and this year she was among the awardees of Ph.D. Talent, a program of JCMF that supports young scientific talents in the South Moravian Region.



You’re working on a tool that analyses proteins as part of your doctoral research. Can you describe what your work involves?

I focus on enzymes, which are a type of protein. These molecules are found everywhere in nature and in the human body — and they do an incredible amount of work. Many people know proteins are important for building muscle, but they’re essential for many other processes as well, such as digestion. The activity of enzymes is also behind the functioning of many laundry detergents.

Most enzymes that naturally occur in the environment are adapted to natural conditions. But if we want to use them in products, chemical processes, and similar applications, we need to modify them to function as we want. In the lab, we change them — we mutate them. Recently, it has also been confirmed that nature contains many proteins and enzymes we don’t yet know,

and some of them are adapted to much harsher conditions. These may be proteins from organisms living in hot springs, for example.

There are databases of such proteins that have been sequenced from various organisms, but nobody has analysed them, so we don't know their properties or the environments in which they function. The tool EnzymeMiner, which I work on as part of my PhD, searches these databases and tries to find the most suitable candidates. It focuses on specific protein regions that are important to the user. It then suggests proteins that may show the desired activity and provides additional annotations, such as the temperature or pH at which they function.

How many of these unexplored proteins might exist? Hundreds, thousands?

We're talking about millions! Metagenomic databases are currently being created: a researcher takes a soil or other environmental sample and starts sequencing the molecules found there. Sometimes these are full proteins, sometimes DNA, sometimes just fragments. And hidden in these databases may be the "diamonds" — proteins with exactly the properties we need.

Once the database finds suitable proteins, how long does experimental verification take to confirm they actually work as expected?

At least several months. You need to create a gene for the protein, insert the gene into cells, extract the protein from the cells, and only then begin verification, which is a long process in itself. Along the way, we may discover that the protein cannot be produced under our conditions or in the organism we're using.

In which organisms do you produce proteins?

Most commonly in *E. coli* cells or yeast. You can also buy a synthetically produced protein, but that's significantly more expensive, so it depends on the budget. Verification is very individual — sometimes it can take years.

How do you use artificial intelligence and machine learning in your work?

The version of EnzymeMiner I helped develop uses many machine-learning tools for protein annotation. It takes a protein sequence and tries to predict its properties — for example, the temperature at which the protein will function. These things are quite difficult to do statistically. For some properties, it's completely impossible, so machine learning can provide a useful perspective. One must keep in mind that it's not 100 % accurate, but it can filter out interesting cases worth further testing.

You started studying Environmental Health, which examines chemical processes in nature and how various chemicals affect humans and ecosystems. Why did you choose this field after studying computer science?

My specialization is Mathematical Biology, Bioinformatics, and Modelling, which includes analysing enzymes, proteins, and DNA in organisms. It was an opportunity to stay connected to computer science, which I studied, while diving deeper into bioinformatics. The program also allows me to study and work simultaneously. If I worked in a typical IT corporation,

combining that with a PhD would be very difficult. I enjoy computer science, and at the same time, I'd like to help nature and human health.

It's quite a unique combination — not many people have this range. And as a woman, you must have been in the minority among computer scientists!

I was. (laughs) In recent years, the number of women in the field has been increasing. But when I started, we were one woman for every fifty students.

How did you find studying in such an environment?

I felt very comfortable. I grew up among boys, and I get along with them. Sometimes it was a bit difficult to fight for my place in team projects — many classmates preferred to work on their own, in their own way. I had to assert myself to be involved in the important parts of the work. They often imagined I would write documentation and do testing, while they handled the algorithmic parts.

What do you enjoy most about your work?

I really love solving problems. I enjoy gradually discovering solutions — in computer science, you can try many different approaches. There's always something new to explore.

It's an incredibly fast-moving field. Is it even possible to keep up with the pace and all the new developments?

At the beginning, it was very demanding — after all, I only encountered bioinformatics during my master's studies. Understanding all the theory and the broad foundations isn't easy. I realised you really have to focus on a very narrow and specific part of the field and keep up with that, while expanding your horizons in your free time and learning the broader context. For me, at first, it was impossible to grasp the broader context of protein-focused bioinformatics.

You say many proteins — or at least their functions and potential — remain undiscovered. How large is that part of the spectrum?

The general opinion is that we've already discovered most reactions. I don't think we'll find a completely new reaction that's fundamentally different from all others. But there may be enzymes with mechanisms that differ from those we currently know.

And if we already know a protein's function, can we infer all the biological processes happening within it?

No. We can see how it reacts — that a particular protein performs the function we want — but studying the mechanism itself is extremely challenging. Observing a protein "in action" is very difficult. Simulations involving quantum mechanics are used, and it's a very complex process. Several people may work on a single protein for years, if it succeeds at all. It's work with a very uncertain outcome.

Where do you think this field — your work or the discipline as a whole — will be in five or ten years?

That's a very difficult question. We don't even know where it will be in a year. It would be amazing if we could create a machine-learning tool that helps either with protein design or with annotation — meaning that from a sequence alone, we could determine a protein's function. It might happen soon, or it might not happen even in ten years... Hard to say.

When you're not working, how do you clear your mind?

I do sports — cycling or exercising. I'd like to find time to start horseback riding again. I also read books, from detective stories and novels to science fiction.

martina.kopecka@fnusa.cz