

The core competence of the research group of Prof. Schwaneberg (<http://www.biotec.rwth-aachen.de/>) is the evolutionary and rational design of proteins. Our projects range from basic research to elucidate structure-function relationships and their modeling, methodological developments for Directed Evolution and optimization of biocatalysts for industrial production and peptides as materials for medical applications.

Master/Bachelor Thesis or internship Data driven protein engineering using Machine Learning

Description:

Enzymes are biologically produced proteins that catalyse chemical processes and speed up chemical reactions (up to millions of times faster), with many applications in food and beverages, detergents, animal feed, pulp and paper, fine chemicals, pharmaceutical, biofuel industry and many more. As we move from a petrochemical to a more sustainable, green bioeconomy we must ensure that our ability to design and engineer biocatalysts is not restricting the industrialization of biology. However, in the setting where we know only relatively little about the enzyme and its molecular mechanism, one frequently targets positions throughout the entire protein sequence for optimization, resulting in a search-space that is too vast to be explored systematically – here, a data-driven approach will act beneficial by predicting desired protein variants. Beyond that, high-throughput technologies have turned protein engineering into a data-intensive discipline, requiring bioinformaticians carry out data management and analysis tasks on large scale by using high-performance computing resources. This project involves the analysis of large-scale biotechnological data through the application of bioinformatics workflows and the development of innovative methods and platforms as required. The position is most suitable for highly motivated students, ideally with expertise in the analysis of large data and basic knowledge in any programming language such as Python, R and/or C. The individuals will collaborate very closely with experimental researchers to work on joint initiatives.

This project aims to explore **state-of-the art artificial intelligence methods using machine learning** to model protein sequence–function relationships and guide the protein engineering by rapidly identifying beneficial diversity for recombination.

Qualifications

We are looking for ambitious MSc or BSc students from biology, biotechnology, chemistry, or computer/data sciences that are interested in developing their skills in machine learning, bioinformatics and statistics that

- have basic knowledge about bioinformatics, (bio)statistics, or computer science
- have an interest in life science applications such as structural biology, biochemistry, or protein engineering
- ideally are proficient in Python (and/or R or any programming language)
- show high motivation, flexibility, creativity, team-, organizational- and communicational skills

What we offer /Techniques you will learn:

We offer an opportunity to get experience in applied machine learning using unique, large high-quality datasets. You will work together with a team of enthusiastic specialists in protein design and engineering and have access to a local compute cluster. Expect to get plenty of experience and make real contributions to an important field of research using data you won't find anywhere else.

Estimated time: According to “Studienordnung”.

Start: Applications will be considered until the position is filled.

Contact person:

Dr. Mehdi D. Davari

Head of Computational Biology Division, Lehrstuhl für Biotechnologie, Worringer Weg 3
52074 Aachen

Phone: +49-241-80-20676

E-Mail: m.davari@biotec.rwth-aachen.de