

Bachelor Thesis Project: Developing a Sample-Specific Protein Database Workflow Using Nextflow

Are you interested in biomedical data analysis and workflow automation? The Medical Bioinformatics group at Ruhr University Bochum is offering an exciting **bachelor thesis project** focused on improving **protein identification in proteomics** using **Nextflow**.

Project Overview

In proteomics, proteins are analyzed by breaking them into smaller peptides and matching them against a reference database. However, standard protein databases are often incomplete, leading to unidentified peptides. This project aims to enhance protein identification by incorporating sample-specific sequencing data (DNA/RNA) into a dynamically updated protein database.

Your Role

As part of this project, you will:

- ✓ Gain an understanding of **proteomics, bioinformatics, and computational biology**
- ✓ Identify and integrate **existing bioinformatics tools** into a structured pipeline
- ✓ Design and implement a **Nextflow-based workflow** for automated database generation
- ✓ Optimize performance and ensure scalability
- ✓ Document your work and contribute to scientific discussions

What You'll Learn

- **Nextflow** for scalable workflow automation
- **Python/R** scripting for bioinformatics
- **Containerization** (e.g., Docker, Singularity) for reproducibility
- **Proteomics & Genomics** data analysis, Scientific **documentation and reporting**

Who Should Apply?

This project is ideal for **bachelor students in bioinformatics, informatics, computer science, data science, or related fields** with an interest in:

- ♦ Biomedical data analysis (Proteomics/Genomics/Transcriptomics)
- ♦ Scripting in **Python** (and optionally R), **Linux/macOS/WSL** environments
- ♦ Scientific research and problem-solving

Project organization



Duration: flexible



Location: Ruhr University Bochum, Main Campus, on-site, hybrid- possible

- ♦ **Supervisors:** Prof. Julian Uszkoreit, Sven Berres (medbioinf[[@](mailto:medbioinf@rub.de)][rub.de](mailto:medbioinf@rub.de))