# Structural & Computational Exploration of Serine Protease Off-Target Landscapes

## Aims and Description

The aim of this master thesis is to comprehensively characterize therapeutically relevant serine proteases from the complement or coagulation system using computational and structural approaches, as part of an effort to expand PanScreen, an in-house toxicity and off-target screening platform. This inter-group collaborative project bridges structural biology, cheminformatics, computational drug discovery, and machine learning, contributing to both method development and biological interpretation.

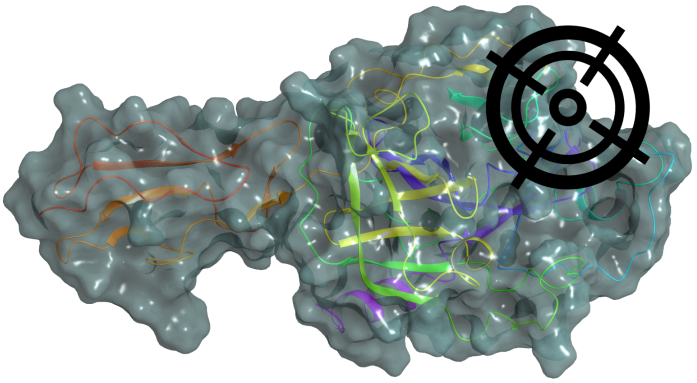

Fig. 1: Targeting Serine Proteases

In this project, you will:

- Conduct comprehensive analysis and curation of available crystal structures (quality, binding site identity, mutation assessment) and systematically prepare protein structures (selection, curation, and preprocessing)
- Perform cheminformatics analyses to curate and characterize binding pockets and ligand properties, explore binding site similarity to identify potential protein ensembles reflecting binding site diversity, and conduct ligand data mining and curation
- Conduct molecular docking into protein ensembles adopting validated solvation states and structural variants to evaluate binding and interactions involved, while developing ensembles for improved receptor flexibility consideration

- Integrate the new target to the machine learning models already developed in PanScreen for binding affinity prediction and validate through redocking, pose RMSD assessment, and machine learning model evaluation
- (Bonus) Contribute to the automation and integration of the PanScreen Python Package

Depending on interest and time, advanced methods such as molecular dynamics, machine learning interpretation, or enhanced sampling can be explored to capture detailed biophysical insights and further refine predictions. By the end of the project, the candidate will have created a reproducible workflow, generated new insights into serine protease off-target interactions, and contributed methodological improvements to PanScreen, directly enabling safer drug discovery campaigns.

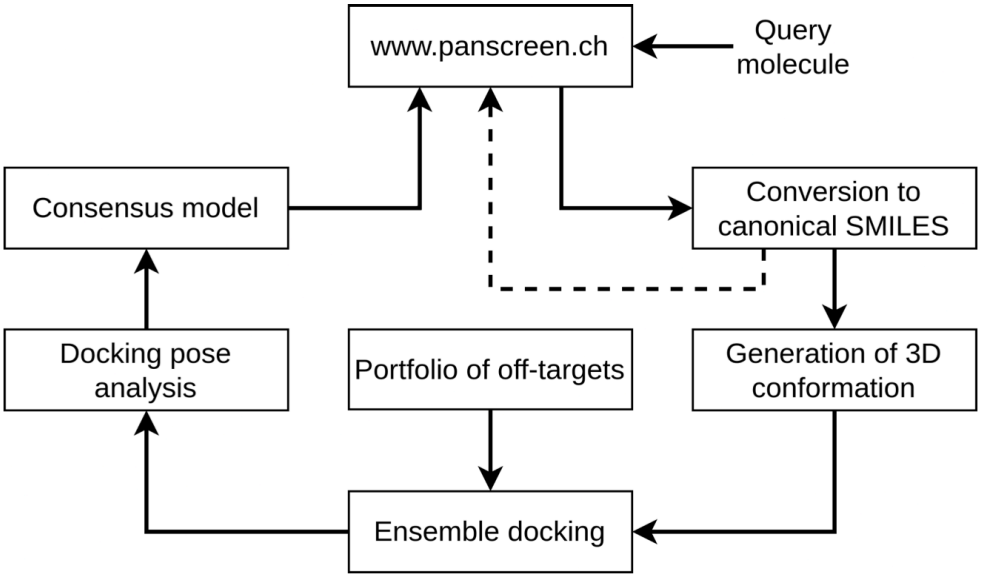


Fig. 2: Flow of data in PanScreen

# Requirements

- Open to students in Pharmacy and Drug Science as well as other UniBas students (external students may also apply)
- Interest in proteins, biophysics, chemistry and computational biology
- Strong motivation for computational tasks (docking, analysis, coding)
- Willingness to develop or extend skills in Python programming, which will be fundamental for some computational methods applied in the project
- Comfort with scientific reading, data analysis, and documentation

# Methods Used

- Molecular Docking: Predicting binding orientations and affinities of inhibitors to protein targets to uncover interaction patterns and off-target risks
- Molecular Dynamics (optional/advanced): Simulating atomistic movements for dynamic insight into protein-ligand interactions
- Rational Design: Iteratively refining hypotheses and experiments based on computational predictions and experimental validation data
- Cheminformatics: Ligand curation, descriptor calculation, diversity analysis
- Structural Bioinformatics: Structure alignment, superposition, comparative modeling, RMSD calculation, binding site/ensemble design
- Machine Learning: Application or extension of ML models (for affinity/tox prediction, pose scoring, etc.) as integrated in PanScreen or related software

# Software Used

- Maestro (Schrödinger software suite): Protein preparation, docking, analysis
- PyMOL: Visualization and inspection of protein-ligand complexes and structural features
- Python: Workflow scripting, data preprocessing, computational analysis (NumPy, Pandas, Matplotlib)
- PyTorch: Machine learning and deep learning model implementation/evaluation (leveraged by gnina or PanScreen)
- PanScreen Python package: Core platform for ensemble docking, off-target prediction, and workflow integration
- ProteInsight, BSDiff, RMSDockSelect: Auxiliary tools for structure curation, site comparison, solvation state/ensemble optimization
- (Bonus/optional) High Performance Computing (Linux/Unix environment), Jupyter notebooks for reproducible workflows, Git/GitHub for version control

# Duration

Possible durations: 5 – 10 months

Start date: Jan 2026 – Mai 2026

# Supervisors

**Dr. Peter Rüthemann**
PostDoc
Molecular & Computational Pharmacy
peter.ruethemann@unibas.ch

**Sofiene Khiari**
PhD Candidate
Computational Pharmacy
sofiene.khiari@unibas.ch

**Ziyu She**
PhD Candidate
Computational Pharmacy
ziyu.she@unibas.ch

**Prof. Markus A. Lill**
Head of Research Group
Computational Pharmacy
markus.lill@unibas.ch

# Application

Please apply with a *curriculum vitae* and a motivation letter to Dr. Peter Rüthemann