Abstract:

Structural basis of rhomboid protease substrate recognition and transmembrane helix arrangement by cryo-EM and integrative modeling

Co-supervisor 1: Daniel RODERER, Leibniz-Forschungsinstitut für Molekulare Pharmakologie (FMP)

Co-supervisor 2: Adam LANGE, FMP and HU Berlin

Co-supervisor 3: Han SUN, FMP and TU Berlin

Rhomboid proteases (RPs) are highly conserved intramembrane proteases with physiological and pathological relevance in signaling processes, neurodegenerative disorders, and cancer. Their exact substrate recognition and proteolysis mechanisms are unknown due to the lack of structural data of native-like RP – substrate complexes, which impedes efficient therapies against RP-dependent diseases. Here, we will elucidate the structural basis for substrate recognition and intramembrane proteolysis of RPs at highest possible resolution, using the RP model system GlpG together with its substrate TatA. We will apply cryogenic electron microscopy (cryo-EM) and integrative modeling to solve structures of full-length GlpG in lipid environment, alone and in a covalent complex with a TatA-derived peptide that comprises its cleavage site. Together with molecular dynamics simulations, the obtained structures will illuminate substrate binding process and proteolysis. Ultimately, we will use analogous strategies to determine the structure of human mitochondrial PARL in complex with a peptide-derived substrate – based on the known substrates PINK1 and PGAM5.

Extended description version of the project:

Structural basis of rhomboid protease substrate recognition and transmembrane helix arrangement by cryo-EM and integrative modeling

Co-supervisor 1: Daniel RODERER, FMP

Co-supervisor 2: Adam LANGE, FMP and HU Berlin

Co-supervisor 3: Han SUN, FMP and TU Berlin

1. Overall goal of the project

Rhomboid proteases (RPs) are intra-membrane serine proteases that often initiate signaling pathways through cleavage of membrane proteins within the lipid bilayer and release of signal proteins. Despite their broad physiological and pathological relevance, exact substrate recognition and proteolysis mechanisms of RPs are unknown due to the lack of structural data of native-like RP-substrate complexes. The aim of this project is to decipher the structural basis for substrate recognition and intramembrane proteolysis of RPs at near-atomic resolution, using *E. coli* GlpG as a well-understood model system and its substrate TatA. We will apply cryogenic electron microscopy (cryo-EM) and integrative modeling to solve structures of: (1) full-length GlpG (including the N-terminal regulatory domain) in lipid environment, and (2) the covalent complex of GlpG with an α-helical peptide that comprises the TatA cleavage site. Together with molecular dynamics (MD) simulations, the structures will illustrate molecular re-arrangements during substrate binding and proteolysis. Based on the data for the GlpG-substrate complex, we plan to extend the project towards structural analysis of the clinically relevant human rhomboid protease PARL and its substrates PINK1 and PGAM5 using the same strategy.

2. State of the art

The protein superfamily of RPs is highly conserved in all branches of life. Their members fulfill important regulatory processes, such as EGF receptor signaling activation, quorum sensing in bacteria, and membrane protein quality control. Aberration of RP function can result in neurodegenerative disorders and cancer. The proteins are intramembrane serine proteases and cleave substrate membrane proteins within the lipid bilayer, resulting in release of peptides or proteins relevant for signaling reactions. Despite their physiological and pathological relevance and in contrast to soluble proteases, RPs are highly understudied. The exact substrate recognition/binding and cleavage mechanisms are unknown. Only the model rhomboid protease, GlpG of *E. coli*, has been structurally characterized to a high resolution[1,2]. The C-terminal, catalytic domain of GlpG (residues 87–276) consists of 6 transmembrane (TM) helices, where 5 helices surround one central helix (TM4) that harbors the catalytic Ser 201. All GlpG structures so far lack the N-terminal, cytosolic domain (residues 1–86) that fulfills a regulatory function[3]. To date, no structural data of native-like substrate-protease complexes are available for GlpG and other RPs. The closest homolog with known structure is a complex with a tetrapeptide aldehyde inhibitor[4], which displays catalytic snapshots without the context of a full substrate TM domain and thereby without information on substrate recognition and release during catalysis. There is a desperate need to understand the substrate recognition mechanism for RPs as clinically important proteases, both to understand how this is affected in diseases and for developing specific inhibitors. Previously, **Lange**

identified a hitherto undescribed gating mechanism of GlpG that involves α–helix 5 using solid-state NMR spectroscopy[5]. By combining NMR with MD simulations, **Lange** and **Sun** clarified the binding modes of both reversible and irreversible rhomboid protease inhibitors under physiological conditions[6]. However, both works were carried out without the N-terminal GlpG domain, which fulfills important regulatory functions, and without a native-like, helical peptide substrate. We propose to assess the full-length GlpG structure with a native-like substrate in lipid environment via cryo-EM and single particle analysis (SPA). To assess the applicability of cryo-EM to visualize full-length GlpG (31 kDa), the purified protein was embedded in nanodiscs (**Lange**), vitrified, and analyzed via cryo-EM/SPA (**Roderer**). In 2D classification of cryo-EM data, we observed disk-shaped particles with ~15 nm diameter, matching the size of GlpG integrated in the used polymer nanodiscs. Further 3D analysis is hampered by the small size of GlpG and the lack of an identifiable extramembraneous localization marker, which we plan to overcome by the legobody strategy (see below).

3. Specific aims and how they may be reached:

We propose to solve the structure of membrane-embedded, full-length GlpG, both alone and in a covalently trapped complex with a substrate peptide that comprises the single TM helix of TatA, using cryo-EM and SPA. A covalent and stable complex between an inactivated GlpG variant and the TatA peptide will be achieved by introducing cysteines to form an intermolecular disulfide bond, analogous to structure determination of γ-sectretase with its substrate APP[7]. With 31 kDa, GlpG alone is too small for high-resolution cryo-EM, and the small N-terminal cytosolic domain (residues 1-86) outside the membrane is not sufficient to unambiguously orient the protein in 3D analysis. Therefore, we will apply a strategy termed legobody[8] to increase the overall size and gain a large and identifiable extramembraneous localization marker for 3D analysis. The legobody is a rigid, ~120 kDa ternary construct of a custom-made nanobody against GlpG (commercially produced for this project, Hybrigenics), a Fab against the nanobody, and a ProteinG/maltose binding protein fusion (Addgene #176075, 176076, 176077; already obtained and produced by **Roderer**). The legobody proved successful for high resolution structural analysis of the 23 kDa membrane protein KDELR and the 22 kDa SARS-CoV2 spike protein[8], both smaller than GlpG.

WP1: Structure of full-length GlpG in lipid nanodiscs. GlpG (full-length; 1-278) will be expressed in *E. coli* and purified in the **Lange** group as previously established[6]. The protein will subsequently be reconstituted in nanodiscs (NDs; **Lange** / **Roderer**) using MSP1D1ΔH5, resulting in 8 nm discs. With this small scaffold protein, we minimize multiple GlpG insertions per ND to obtain a homogeneous cryo-EM sample[9]. ND-embedded GlpG will be used in a synthetic library screen to obtain a high-affinity nanobody against the N-terminal domain. With this nanobody, we will assemble the legobody, bind it to nanodisc-embedded GlpG, and analyze the entire complex (ca. 190 kDa) by cryo-EM and SPA (**Roderer**). The resulting cryo-EM density map (ideally at a resolution of ~3 Å in both GlpG domains) will be combined with existing NMR data (**Lange**) for integrative modeling (**Sun**).

WP2: Structure of GlpG-substrate complex. We will produce an inactivated, full-length GlpG mutant with a Cys introduced at the substrate binding site to facilitate the formation of a disulfide bond with a TM helix Cys mutant of TatA (1 – 21). The TatA peptide will be synthesized by the FMP peptide facility and reconstituted in NDs together with GlpG under oxidizing conditions (**Lange**). The following steps of legobody complex assembly, cryo-EM and SPA will be carried out analogous to WP1 (**Roderer**), focusing on optimizing resolution of the GlpG-substrate binding site by 3D sorting in SPA. After atomic modeling (**Roderer** / **Sun**), we will assess intermediate conformations of substrate binding and cleavage via MD simulations (**Sun**).

Dependent on the outcomes of WPs 1 and 2, we will apply the same strategy for structural analysis of the clinically relevant, human mitochondrial rhomboid protease PARL, for which production in yeast and alternatively by cell-free expression is currently optimized by **Lange**.

Resources. The wet labs of **Roderer** and **Lange** are equipped to carry out the required biochemical work. Cryo-EM data will be collected at the Charité Cryo-EM core facility at Campus Berlin Buch, where **Roderer** has high-priority access to a Titan Krios equipped with a K3 direct electron detector and energy filter. For cryo-EM data analysis, modeling, and MD simulations, **Roderer** and **Sun** have several high-end CPU and GPU servers (most of them from 2022) and access to multiple super-computing centers such as HLRN.

References:

1 Wang, Y., Zhang, Y. & Ha, Y. Crystal structure of a rhomboid family intramembrane protease. *Nature* **444**, 179-180, doi:10.1038/nature05255 (2006).

2 Vinothkumar, K. R. Structure of rhomboid protease in a lipid environment. *J Mol Biol* **407**, 232-247, doi:10.1016/j.jmb.2011.01.029 (2011).

3 Sherratt, A. R., Blais, D. R., Ghasriani, H., Pezacki, J. P. & Goto, N. K. Activity-based protein profiling of the Escherichia coli GlpG rhomboid protein delineates the catalytic core. *Biochemistry* **51**, 7794-7803, doi:10.1021/bi301087c (2012).

4 Cho, S., Baker, R. P., Ji, M. & Urban, S. Ten catalytic snapshots of rhomboid intramembrane proteolysis from gate opening to peptide release. *Nat Struct Mol Biol* **26**, 910-918, doi:10.1038/s41594-019-0296-9 (2019).

5 Shi, C. *et al.* Structure and Dynamics of the Rhomboid Protease GlpG in Liposomes Studied by Solid-State NMR. *J Am Chem Soc* **141**, 17314-17321, doi:10.1021/jacs.9b08952 (2019).

6 Bohg, C. *et al.* A combination of solid-state NMR and MD simulations reveals the binding mode of a rhomboid protease inhibitor. *Chem Sci* **12**, 12754-12762, doi:10.1039/d1sc02146j (2021).

7 Zhou, R. *et al.* Recognition of the amyloid precursor protein by human gamma-secretase. *Science* **363**, doi:10.1126/science.aaw0930 (2019).

8 Wu, X. & Rapoport, T. A. Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies). *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2115001118 (2021).

9 Roderer, D., Hofnagel, O., Benz, R. & Raunser, S. in *Proceedings of the National Academy of Sciences of the United States of America* Vol. 116 23083-23090 (2019).