Design, synthesis and biological evaluation of rhomboid protease specific inhibitors

Inaugural dissertation

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University Duesseldorf

presented by

Parul Goel
from Gohana, India

Duesseldorf, August 2016

from the Institute for Neuropathology, University Hospital
at the Heinrich Heine University Duesseldorf

Published by permission of the
Faculty of Mathematics and Natural Sciences at
Heinrich Heine University Duesseldorf

Supervisor: Prof Dr. Sascha Weggen
Co-supervisor: Prof. Dr. Jörg Pietruzska

Date of the oral examination: 24.11.16

Summary

Rhomboids are intramembrane serine proteases present in prokaryotic, archaeal and eukaryotic organisms. Rhomboids are composed of 6 core transmembrane helices and feature a serine-histidine catalytic dyad. They hydrolyze the peptide-bond of substrate membrane proteins within the lipid bilayer and control diverse biological processes; e.g. EGF-receptor signaling in *Drosophila melanogaster*, quorum sensing in *Providencia stuartii*, host cell invasion of the malaria parasite *Plasmodium falciparum*, and mitochondrial integrity in mammals. Due to this remarkable range of biological functions, rhomboids might have great potential as drug targets. Previously, small molecules such as isocoumarins, fluorophosphate, β-lactams, β-lactones and chloromethylketones have been found to inhibit rhomboid proteases but most of them had low potency or insufficient selectivity over soluble serine proteases such as chymotrypsin. Hence, the objective of this thesis was the design, synthesis and biological evaluation of new and improved rhomboid inhibitors. Based on computer-aided approaches, candidate-based screening and large molecular library virtual screening, suitable drug-like candidates (peptidic and non-peptidic) were selected and screened against rhomboid proteases. Molecular modeling experiments were performed with the Molecular Operating Environment (MOE) software and a co-crystal structure of the *E. coli* rhomboid GlpG and the phosphate inhibitor Cbz-AlaP(OiPr)F or CAPF (PDB: 3UBB). Subsequently, the best compounds were selected for synthesis or purchasing and subjected to biological analysis. With the candidate-based screening, we discovered two new classes of rhomboid inhibitors: aryl acid substituted N-methylated saccharins and 2-styryl benzoxazinones. In an *in vitro* activity assay, 12 saccharin derived inhibitors were found to inhibit the *E. coli* rhomboid GlpG with IC50 values below 1 µM. The IC50 of the best derivative (BSc5195) was 0.2 µM. All saccharin inhibitors had at least a 20-fold window of selectivity for inhibition of the rhomboid GlpG over the soluble serine protease α-chymotrypsin. Furthermore, we were able to demonstrate that the saccharin inhibitors did not indiscriminately inhibit all rhomboids opening up the possibility to develop inhibitors selective for the evolutionary older bacterial and archaeal rhomboids. With respect to the mechanism of inhibition, we confirmed the release of the aryl acid leaving group during the enzymatic reaction. However, the predicted formation of a cross-linked enzyme-inhibitor complex was not observed. Indeed, our findings suggest that only the active site serine is covalently modified by the saccharin inhibitors. The 2-styryl benzoxazinones were also found active against GlpG but their potency was comparatively low (best compound BSc5172 with IC50 ≈ 25 µM). Finally, we performed a virtual screen of a large molecular library using the DOCKTITE workflow and identified a new phosphate inhibitor (VS5). VS5 was slightly more potent than known phosphate inhibitors with good selectivity over α-chymotrypsin proving that a virtual screening strategy can be successfully employed to discover novel protease inhibitors. In summary, N-methylated saccharins are a novel and promising class of mechanism-based inhibitors

against rhomboid proteases with inhibitory potency in the submicromolar range and clear potential for further improvement.

Zusammenfassung

Rhomboide sind Intramembranproteasen und kommen sowohl in Archaeen und Prokaryoten als auch in Eukaryoten vor. Rhomboide weisen ein Grundgerüst aus 6 Transmembranhelices auf und besitzen im aktiven Zentrum eine katalytische Dyade aus einem Serin und einem Histidin. Rhomboide hydrolysieren die Peptidbindung ihrer Substrate innerhalb der Lipiddoppelschicht einer biologischen Membran und kontrollieren auf diese Weise vielfältige biologische Prozesse. Dazu gehören z.B. der EGF-Rezeptor vermittelte Signalweg in *Drosophila melanogaster*, das Quorum Sensing in *Providencia stuartii*, die Zellinvasion durch den Malaria Parasiten *Plasmodium falciparum* und die Aufrechterhaltung der mitochondrialen Integrität in Säugerzellen. Aufgrund dieser beeindruckenden Vielfalt an Aufgaben, stellen Rhomboide ein interessantes Ziel für die Entwicklung neuartiger Wirkstoffe dar. Zu den bisher identifizierten Wirkstoffklassen, die als Rhomboid Inhibitoren fungieren, zählen Isocoumarine, Fluorophosphonate, β-Lactame, β-Lactone und Chloromethylketone. Diese Substanzklassen weisen allerdings entweder eine niedrige Potenz auf oder sind im Hinblick auf lösliche Serin-Proteasen wie z.B. Chymotrypsin wenig selektiv. Daher bestand das Ziel dieser Arbeit im Design, der Synthese und der funktionalen biologischen Charakterisierung neuer und verbesserter Rhomboid-Inhibitoren. Auf der Basis Computer-gestützter Ansätze, Kandidaten-basiertem Screening und dem virtuellen Screening einer großen Molekül-Datenbank wurden vielversprechende Wirkstoff-ähnliche Kandidaten (Peptid-basierte und nicht-Peptid basierte) ausgewählt und ihre Aktivität im Hinblick auf die Inhibition von Rhomboiden untersucht. Die Experimente zum Molekularen Modeling wurden mit der Software Molecular Operating Environment (MOE) durchgeführt. Als Ausgangsbasis wurde eine Ko-Kristallstruktur des *E. coli* Rhomboids GlpG im Komplex mit den Phosphat-Inhibitoren Cbz-AlaP(OiPr)F oder CAPF (PDB: 3UBB) verwendet. Im Anschluss wurden die für aussichtsreich befundenen Substanzen entweder synthetisiert oder käuflich erworben und in einem biologischen Assay auf ihre Aktivität untersucht. Beim Kandidaten-basierten Screening wurden auf diese Weise zwei neue Klassen von Rhomboid-Inhibitoren identifiziert: Arylsäure-substituierte N-methylierte Saccharine und 2-Styryl Benzoxazinone. In einem *in vitro* Aktivitätsassay konnten für 12 Saccharin-basierte Inhibitoren IC50-Werte im Bereich von unter 1 µM für die Inhibition von GlpG identifiziert werden. Der IC50-Wert der potentesten Substanz lag bei 200 nM (BSc5195). Alle Saccharin-basierten Inhibitoren wiesen ein mindestens 20-faches Selektivitätsfenster gegenüber Chymotrypsin auf. Weiterhin konnten gezeigt werden, dass Saccharin-basierte Inhibitoren nicht alle Rhomboide gleichermaßen inibieren. Dies eröffnet die Möglichkeit Inhibitoren darzustellen, die selektiv für die evolutiv älteren Rhomboide aus Bakterien und Archaeen sind. Der zugrundeliegende Mechanismus der Inhibition von Rhomboid Proteasen basiert auf der Arylsäure-Abgangsgruppe und der damit verbundenen Modifikation des Enzyms während der Katalysereaktion. Der vorhergesagte kovalent quervernetzte Komplex aus Enzym und Inhibitor konnte jedoch nicht bestätigt werden. Die vorliegenden Ergebnisse weisen allerdings darauf hin, dass die Serin-Seitenkette im aktiven Zentrum

von Saccharin-basierten Inhibitoren kovalent modifiziert wird. Für die 2-Styryl Benzoxazinone konnte ebenfalls eine inhibitorische Aktivität in Bezug auf GlpG nachgewiesen werden. Allerdings hatten diese Substanzen eine vergleichbar geringe Potenz – die beste Substanz BSc5172 wies einen IC50-Wert von ca. 25 µM auf. Abschließend wurde ein virtuelles Screening einer großen Molekülbibliothek mit Hilfe des DOCKTITE Algorithmus durchgeführt. Das Screening identifizierte den Phosphat-Inhibitor VS5, der potenter als bisher beschriebene Phosphatinhibitoren war, und eine gute Selektivität gegenüber Chymotrypsin aufwies. Somit konnte gezeigt werden, dass die Strategie, neue Protease-Inhibitoren über ein virtuelles Screening zu entdecken, durchaus vielversprechend sein kann. Zusammenfassend lässt sich festhalten, dass die identifizierten N-methylierten Saccharine eine neue und aussichtsreiche Substanzklasse Mechanismus-basierter Inhibitoren für Rhomboid Proteasen sind. deren Potenz schon zum jetzigen Zeitpunkt im submikromolaren Bereich liegt und die deutliches Potential für weitere Verbesserungen aufweisen.

Contents

LIST OF FIGURES ..... **I**

LIST OF TABLES ..... **IV**

ABBREVIATIONS ..... **V**

1 INTRODUCTION **1**

1.1 PROTEASES 1

1.1.1 Proteases classification *1*

1.1.2 Intramembrane proteases *2*

1.2 RHOMBOID PROTEASES 3

1.2.1 Rhomboid protease functions and their importance in diseases *4*

1.2.2 Rhomboid like super-family classification *7*

1.2.3 Rhomboid Structure *7*

1.2.4 Substrate specificity of Rhomboid proteases *15*

1.3 INHIBITION OF SERINE PROTEASES 17

1.3.1 Peptidic serine proteases inhibitors *17*

1.3.2 Non-peptidic serine proteases inhibitors *19*

1.4 INHIBITOR DESIGN 22

1.5 RHOMBOID PROTEASE INHIBITORS 24

1.6 OBJECTIVE 27

2 MATERIAL **28**

2.1 BACTERIAL STRAINS AND CELL LINES 28

2.1.1 Bacterial strains *28*

2.1.2 Cell lines *28*

2.2 VECTORS AND PRIMERS 28

2.2.1 Vectors *28*

2.2.2 Primers *29*

2.3 ANTIBODIES 29

2.3.1 Primary antibodies *29*

2.3.2 Secondary antibodies *29*

2.4 REAGENTS 29

2.4.1 Chemicals *29*

2.4.2 Antibiotics *31*

2.4.3 Inhibitors *31*

2.4.4 Size Standards *35*

2.4.5 Enzymes *35*

2.4.6 Kits *35*

2.5 Laboratory hardware and appliances 35
2.6 Consumables 36
2.7 Softwares 37

3 METHODS 38

3.1 Molecular Biology 38
3.1.1 Plasmid isolation 38
3.1.2 Preparation of chemically competent E. coli cells 38
3.1.3 Transformation of E.coli 39
3.1.4 Sequencing of plasmids 40
3.2 Protein Biochemistry 40
3.2.1 Cryocultures 40
3.2.2 Bacterial preculture 40
3.2.3 Protein expression 40
3.2.4 Cell disruption 41
3.2.5 Fractional centrifugation 42
3.2.6 Protein purification: Immobilized metal ion affinity chromatography (IMAC) 42
3.2.7 Protein purification: Affinity chromatography using GST resin 43
3.2.8 Protein concentration 44
3.2.9 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE) 44
3.2.10 Coomassie staining 46
3.2.11 Western blot 47
3.2.12 EcGlpG in vitro activity assay 49
3.2.13 AaROM in vitro activity assay 49
3.2.14 α-Chymotrypsin Assay 50
3.3 Cell culture 50
3.3.1 Passaging of adherent cell lines (HEK293T) 51
3.3.2 Thawing of frozen cell stocks 51
3.3.3 Determination of cell number 52
3.3.4 Cell seeding and Transfection of HEK293T cells 52
3.3.5 Harvesting of cell culture supernatants 53
3.3.6 Preparation of cell lysates 53
3.3.7 Bicinchonic acid protein assay (BCA) 54
3.3.8 Inhibitor treatment 55
3.3.9 Cell toxicity assay 55
3.4 Molecular Docking 56
3.4.1 Preparation of the Receptor 57
3.4.2 Preparation of the Molecule database 57
3.4.3 Docking and scoring 57

3.4.4 Ranking of ligands *61*
3.5 CHEMICAL COMPOUND SYNTHESIS 62
3.5.1 General information *62*
3.5.2 Synthesis of Ketobenzothiazole derivatives *63*
3.5.3 Synthesis of oxathiazol-2-one derivatives *66*
3.5.4 Synthesis of Saccharin derivatives *68*
3.5.5 Synthesis of Benzoxazinone derivatives *76*
4 RESULTS 81
4.1 THESIS WORKFLOW 81
4.2 PURIFICATION OF INTRAMEMBRANE PROTEINS GLPG, GURKEN SUBSTRATE AND AQRHO 82
4.2.1 Recombinant expression of the rhomboid GlpG *83*
4.2.2 Recombinant expression of the Gurken substrate *83*
4.2.3 Purification of recombinant proteins by immobilized metal ion affinity chromatography (IMAC) *84*
4.2.4 Recombinant expression and purification of the Aquifex aeolicus rhomboid (AaROM) *84*
4.3 GLPG *IN VITRO* ACTIVITY ASSAY 85
4.3.1 Optimization of Gurken substrate concentration for the GlpG in vitro activity assay *85*
4.3.2 Optimization of the reaction time of the GlpG in vitro activity assay *86*
4.3.3 Validation of GlpG in vitro activity assay with known inhibitors *87*
4.4 VALIDATION OF THE MOE DOCKING METHODS 89
4.5 DESIGN, SYNTHESIS AND BIOLOGICAL EVALUATION OF PEPTIDIC INHIBITORS 89
4.5.1 Keto-benzothiazole based peptide-mimetic inhibitors *90*
4.5.2 Oxathiazol-2-one derivatives *103*
4.6 DESIGN, SYNTHESIS AND BIOLOGICAL EVALUATION OF NON-PEPTIDIC RHOMBOID INHIBITORS 105
4.6.1 Rhodanine-3-acetic acid (RAA) inhibitors *105*
4.6.2 Saccharin-based inhibitors *110*
4.6.3 Benzoxazinone-based inhibitors *125*
4.7 VIRTUAL SCREENING FOR NOVEL RHOMBOID INHIBITORS 130
4.7.1 Virtual screening based on DOCKTITE in MOE *130*
4.7.2 Biological evaluation of virtual screening compounds *134*
4.7.3 Cytotoxicity of the virtual screening compound VS5 ..... *136*
4.8 SELECTIVITY OF NEW NOVEL RHOMBOID INHIBITORS OVER SOLUBLE SERINE PROTEASES (A-CHYMOTRYPSIN) 137
4.8.1 Validation of the α-chymotrypsin assay with the known inhibitor DCI. *138*
4.8.2 Activity of novel rhomboid inhibitors (saccharins, benzoxazinones, VS5) against α-chymotrypsin.. *138*
4.9 ACTIVITY OF SACCHARIN BASED INHIBITORS WITH DIFFERENT RHOMBOID PROTEASES AND SUBSTRATES 140
4.9.1 Inhibitory activity of saccharin based inhibitors in a modified GlpG in vitro activity assay using the LacY substrate. *140*
4.9.2 Inhibitory activity of saccharin-based inhibitors against the AaROM rhomboid from Aquifex aeolicus. *141*

4.9.3 Inhibitory activity of saccharin-based inhibitors in a cell-based assay for the mammalian rhomboid protease RHBDL2. ..... *142*

4.10 PRELIMINARY STUDIES ON THE MECHANISM OF RHOMBOID INHIBITION BY SACCHARIN-BASED INHIBITORS. 144

5 DISCUSSION 146

3.7 RHOMBOID PROTEASE ACTIVITY ASSAYS 146

3.8 DESIGN, SYNTHESIS AND BIOLOGICAL EVALUATION OF PEPTIDIC INHIBITORS 148

3.9 DESIGN, SYNTHESIS AND BIOLOGICAL EVALUATION OF NON-PEPTIDIC INHIBITORS 149

3.9.1 Rhodanine-based inhibitors *149*

3.9.2 N-methylated saccharin based inhibitors *150*

3.9.3 Selectivity of N-methylated saccharin based rhomboid inhibitors *152*

3.9.4 Mechanism of N-methylated saccharin based rhomboid inhibitors *154*

3.9.5 Benzoxazinone based inhibitors *155*

3.10 VIRTUAL SCREENING FOR NOVEL RHOMBOID INHIBITORS 155

6 REFERENCES 157

7 APPENDIX 169

7.1 MOLECULAR DOCKING DATA 169

7.1.1 List of aryl acid derived saccharin molecules with variable ring substitutions *169*

7.1.2 A molecular docking model of the saccharin inhibitor with LG38 in the GlpG pocket *170*

7.1.3 A molecular docking model of the saccharin derivatives ACA30 and ACA92 in the GlpG pocket *170*

7.1.4 List of virtual screening top ranked ligands *172*

7.2 BIOLOGICAL EXPERIMENTS: *IN VITRO* GLPG ACTIVITY ASSAY WITH SACCHARIN INHIBITORS 173

7.3 CURRICULUM VITAE 175

7.4 CONTRIBUTIONS TO INTERNATIONAL MEETINGS 177

7.5 ACKNOWLEDGEMENTS 178

7.6 DECLARATION 179

List of Figures

FIGURE 1.1: MECHANISM OF THE FIRST STEP OF PEPTIDE HYDROLYSIS IN THE SIX GENERAL PROTEASE FAMILIES. 1

FIGURE 1.2: SCHEMATIC REPRESENTATION OF BASIC TOPOLOGIES OF THE THREE FAMILIES OF INTRAMEMBRANE PROTEASES (IMPS). . 3

FIGURE 1.3: CATALYTIC RESIDUES OF SERINE PROTEASES. 4

FIGURE 1.4: RHOMBOID REGULATED EGFR SIGNALING PATHWAY IN *DROSOPHILA MELANOGASTER*. 4

FIGURE 1.5: FOUR BIOLOGICAL FUNCTIONS OF RHOMBOID PROTEASES. 5

FIGURE 1.6: RHOMBOID-LIKE SUPERFAMILY. 7

FIGURE 1.7: THREE TOPOLOGICAL FORMS OF RHOMBOID PROTEINS. 8

FIGURE 1.8: THE NOMENCLATURE OF THE SUBSTRATE SEQUENCE AND THE PROTEASE SUBSITES. 9

FIGURE 1.9: CRYSTAL STRUCTURE OF THE RHOMBOID GLPG FROM E. COLI (PDB CODE: 2XOV). 9

FIGURE 1.10: GATING MECHANISM OF THE *E. COLI* RHOMBOID PROTEASE GLPG. 10

FIGURE 1.11: STRUCTURE OF THE OXYANION HOLE IN THE GLPG RHOMBOID PROTEASE IN COMPLEX WITH A COVALENTLY BOUND ISOCOUMARIN INHIBITOR (JLK-6). 12

FIGURE 1.12: SUBSTRATE ENTRY REGULATED BY THE L5 CAP. 13

FIGURE 1.13: STRUCTURAL COMPARISON OF THREE ECGLPG ACTIVE SITES DERIVED FROM DIFFERENT CRYSTAL STRUCTURES. 14

FIGURE 1.14: DIAGRAMMATIC REPRESENTATION FOR DISPLACEMENT OF RESIDUES ON L5 CAP AND TMD 2 AND TMD5. 15

FIGURE 1.15: CO-CRYSTALLIZED STRUCTURE OF GLPG IN COMPLEX WITH THE AC-IATA-CMK INHIBITOR (PDB: 4QO2). 15

FIGURE 1.16: MODEL OF SUBSTRATE RECOGNITION AND PROTEOLYSIS BY RHOMBOIDS. 17

FIGURE 1.17: SCHEMATIC REPRESENTATION OF PEPTIDOMIMETIC COVALENT INHIBITORS. 18

FIGURE 1.18: CHEMICAL STRUCTURE OF KETOBENZOTHIAZOLE-CONTAINING PEPTIDOMIMETIC INHIBITORS OF MATRIPTASE. 19

FIGURE 1.19: EXAMPLES OF ALKYLATING AGENTS. 20

FIGURE 1.20: EXAMPLES OF ENZYME-ACTIVATED INHIBITORS AND THEIR INHIBITION MECHANISMS. 21

FIGURE 1.21: MECHANISM OF INHIBITION OF BENZOXAZINONE DERIVATIVES. 22

FIGURE 1.22: SCHEMATIC REPRESENTATION OF TWO APPROACHES OF INHIBITOR DESIGNING. 23

FIGURE 1.23: STRUCTURES AND MECHANISMS OF KNOWN CLASSES OF RHOMBOID PROTEASE INHIBITORS. 25

FIGURE 3.1: BLOTTING STACK OF THE WESTERN BLOT FOR WET TRANSFER. 47

FIGURE 3.2: BLOTTING STACK OF THE WESTERN BLOT FOR SEMI-DRY TRANSFER. 48

FIGURE 3.3: OVERVIEW OF THE DOCKTITE WORKFLOW. 59

FIGURE 3.4: ELECTROPHILIC WARHEADS IMPLEMENTED IN DOCKTITE. 60

FIGURE 4.1: SCHEMATIC REPRESENTATION OF DESIGN STRATEGIES FOR RHOMBOID PROTEASE INHIBITORS IN THIS THESIS. 81

FIGURE 4.2: WESTERN BLOT ANALYSIS OF RECOMBINANT GLPG EXPRESSION. 83

FIGURE 4.3: WESTERN BLOT ANALYSIS OF RECOMBINANT GURKEN AND RECOMBINANT LACY SUBSTRATE EXPRESSION. 83

FIGURE 4.4: SDS-PAGE ANALYSIS OF THE IMAC ELUTION FRACTIONS. 84

FIGURE 4.5: EXPRESSION AND PURIFICATION OF AAROM 85

FIGURE 4.6: OPTIMIZATION OF GURKEN SUBSTRATE CONCENTRATION FOR THE *IN VITRO* ACTIVITY ASSAY. 86

FIGURE 4.7: OPTIMIZATION OF THE REACTION TIME OF THE *IN VITRO* ACTIVITY ASSAY. 87

FIGURE 4.8: GLPG *IN VITRO* ACTIVITY ASSAY WITH A BETA-LACTAM INHIBITOR (L16) 88

FIGURE 4.9: VALIDATION OF THE MOE DOCKING METHODS. 89
FIGURE 4.10: A TETRA-PEPTIDE AS A LEAD COMPOUND FOR PEPTIDOMIMETIC RHOMBOID INHIBITORS. 90
FIGURE 4.11: TOP VIEW OF THE RHOMBOID GLPG-CAPF COMPLEX (PDB: 3UBB). 91
FIGURE 4.12: THE STARTING POINT DIPEPTIDE SCAFFOLD USED FOR THE MOLECULAR MODELING EXPERIMENTS IN THE RHOMBOID POCKET 92
FIGURE 4.13: TOP VIEW OF THE BEST SCORING STRUCTURES OF THE P1 SEARCH. 94
FIGURE 4.14: COVALENTLY DOCKED DIPEPTIDE INHIBITORS IN THE SEARCH FOR THE P2 RESIDUE. 95
FIGURE 4.15: STRUCTURES OF THE TRIPETIDE (A) AND THE TETRAPETIDE (B) USED IN THE SEARCH FOR THE BEST AMINO ACID AT THE P3 POSITION. 96
FIGURE 4.16: 2D INTERACTIONS OF THE FOUR BEST CONFORMATIONS OBSERVED DURING THE P3 SEARCH. 97
FIGURE 4.17: COVALENTLY DOCKED TETRAPEPTIDE INHIBITOR IN THE SEARCH FOR THE P4 RESIDUE. 98
FIGURE 4.18: COVALENT DOCKING OF A DIPEPTIDE INHIBITOR WITH THE BENZOTHIAZOLE SERINE TRAP MOIETY. 100
FIGURE 4.19: SYNTHETIC SCHEMES OF KETOBENZOTHIAZOLE-BASED PEPTIDIC INHIBITORS. 101
FIGURE 4.20: RHOMBOID *IN VITRO* ACTIVITY ASSAY WITH KETOBENZOTHIAZOLE-BASED PEPTIDOMIMETIC COMPOUNDS. 102
FIGURE 4.21: THE OXATHIAZOL-2-ONE BASED SCAFFOLD SELECTED TO DESIGN RHOMBOID INHIBITORS. 103
FIGURE 4.22: SYNTHETIC SCHEME OF OXATHIAZOLE-2-ONE-BASED PEPTIDE INHIBITORS. 104
FIGURE 4.23: RHOMBOID *IN VITRO* ACTIVITY ASSAY WITH OXATHIAZOLE-2-ONE BASED PEPTIDOMIMETIC COMPOUNDS. 104
FIGURE 4.24: SYNTHESIS SCHEMES OF RHODANINE-3-ACETIC ACID (RAA) DERIVATIVES. 106
FIGURE 4.25: MOLECULAR DOCKING MODEL OF THE RHOMBOID PROTEASE GLPG IN COMPLEX WITH A RHODANINE-3-ACETIC ACID (RAA) DERIVATIVE. 108
FIGURE 4.26: RHOMBOID *IN VITRO* ACTIVITY ASSAY WITH RHODANINE-3-ACETIC ACIDS (RAA) DERIVATIVES. 109
FIGURE 4.27: NOVEL SACCHARIN RING DERIVED INHIBITORS. 110
FIGURE 4.28: N-METHYLATED SUBSTITUTED SACCHARIN-DERIVED INHIBITORS OF HLE. 112
FIGURE 4.29: PUTATIVE MECHANISM OF INHIBITION BY N-METHYLATED SUBSTITUTED SACCHARINS. 113
FIGURE 4.30: MOLECULAR MODEL SHOWING THE END PRODUCT OF THE ENZYMATIC REACTION BETWEEN N-METHYLATED SACCHARINS AND THE RHOMBOID PROTEASE (PDB: 3UBB). 114
FIGURE 4.31: COMPARISON OF THE MOLECULAR DOCKING RESULTS FOR A SACCHARIN DERIVATIVE AND THE CAPF INHIBITOR (PDB:3UBB). 116
FIGURE 4.32: MOLECULAR MODEL OF DOCKED N-METHYLATED SACCHARIN IN GLPG CLEFT BEFORE AND AFTER ENZYMATIC REACTION (PDB: 3UBB). 117
FIGURE 4.33: A COMPARATIVE MOLECULAR DOCKING MODEL OF THE SACCHARIN DERIVATIVES LG47 AND LG38 IN THE GLPG POCKET. 118
FIGURE 4.34: MOLECULAR MODEL OF THE RHOMBOID PROTEASE GLPG DOCKED WITH SACCHARIN DERIVED ARYL SULFIDES AND SULFONES. 119
FIGURE 4.35: SYNTHESIS SCHEMES OF N-METHYLATED SUBSTITUTED SACCHARIN-DERIVED INHIBITORS. 121
FIGURE 4.36: *IN VITRO* ACTIVITY ASSAY WITH PRELIMINARY BATCH OF N-METHYLATED SACCHARIN DERIVATIVES. 123
FIGURE 4.37: GLPG *IN VITRO* ACTIVITY ASSAY WITH THE ARYL CARBOXYLATE SACCHARIN INHIBITOR BSC5188 (N=2). 124
FIGURE 4.38: MECHANISM OF INHIBITION OF SERINE PROTEASES BY 2-SUBSTITUTED BENZOXAZINONE DERIVATIVES. 126

FIGURE 4.39: MOLECULAR DOCKING MODEL OF THE RHOMBOID PROTEASE GLPG IN COMPLEX WITH A BENZOXAZINONE-BASED INHIBITOR. 127
FIGURE 4.40: SYNTHESIS OF 2-SUBSTITUTED BENZOXAZIN-4-ONES INHIBITORS. 128
FIGURE 4.41: GLPG *IN VITRO* ACTIVITY ASSAY WITH BENZOXAZINONE-BASED COMPOUNDS. 129
FIGURE 4.42: GLPG *IN VITRO* ACTIVITY ASSAY WITH THE BENZOXAZINONE-BASED INHIBITOR BSC5172 (N=1). 129
FIGURE 4.43: EXAMPLES OF SACCHARIN BASED INHIBITORS FOLLOWING DIFFERENT REACTION RULES. 131
FIGURE 4.44: WARHEAD DATASET OF 34 VARIABLE WARHEADS DISTINGUISHED BY THEIR CATALYTIC MECHANISM. 132
FIGURE 4.45: GLPG *IN VITRO* ACTIVITY ASSAY WITH THE VIRTUAL SCREENING COMPOUNDS VS1-VS5. 135
FIGURE 4.46: GLPG *IN VITRO* ACTIVITY ASSAY WITH THE VIRTUAL SCREENING COMPOUND VS5 (N=1). 136
FIGURE 4.47: CELL TOXICITY ASSAY WITH THE VIRTUAL SCREENING COMPOUND VS5 (N=2). 137
FIGURE 4.48: A-CHYMOTRYPSIN *IN VITRO* ASSAY. 137
FIGURE 4.49: VALIDATION OF THE A-CHYMOTRYPSIN ASSAY WITH THE ISOCOUMARIN INHIBITOR DCI (N=2). 138
FIGURE 4.50: A-CHYMOTRYPSIN ASSAY WITH THE N-METHYLATED SACCHARIN INHIBITOR BSC5195 (N=2). 139
FIGURE 4.51: A-CHYMOTRYPSIN ASSAY WITH THE 2-STYRYL BENZOXAZINO-4-ONE INHIBITOR BSC5174 AND THE VIRTUAL SCREENING INHIBITOR VS5 (N=1). 140
FIGURE 4.52: GLPG *IN VITRO* ACTIVITY ASSAY USING THE LACY SUBSTRATE WITH THE SACCHARIN-BASED INHIBITOR BSC5204 (N=1). 141
FIGURE 4.53: *IN VITRO* ACTIVITY ASSAY FOR THE *AQUIFEX AEOLICUS* RHOMBOID (AAROM) WITH THE SACCHARIN-BASED INHIBITOR BSC5188 (N=1). 142
FIGURE 4.54: INHIBITORY ACTIVITY OF THE SACCHARIN-BASED INHIBITOR BSC5156 IN A CELL-BASED ASSAY FOR THE MURINE RHOMBOID RHBDL2 (N=3). 143
FIGURE 4.55: A FLUORESCENCE-BASED ASSAY TO STUDY THE REACTION MECHANISM OF GLPG WITH THE SACCHARIN-BASED INHIBITOR BSC5193 (N=2). 144
FIGURE 4.56: *IN VITRO* GLPG ACTIVITY ASSAY WITH RHOMBOID ISOCOUMARIN INHIBITOR (JLK-6) AND ARYL CARBOXYLATE SACCHARIN INHIBITOR BSC5188. 145
FIGURE 5.1: HEAT MAP REPRESENTATION OF INHIBITION OF 9 RHOMBOID PROTEASES BY SACCHARIN INHIBITORS. 153
FIGURE 7.1 A MOLECULAR DOCKING MODEL OF THE SACCHARIN DERIVATIVE WITH LG38 IN GLPG POCKET. 170
FIGURE 7.2: A MOLECULAR DOCKING MODEL OF THE SACCHARIN DERIVATIVES ACA30 AND ACA92 IN THE GLPG POCKET. 170
FIGURE 7.3: IN VITRO GLPG ASSAY OF SACCHARIN INHIBITORS (BSC5187-BSC5206) AT 250 μM CONCENTRATION. 173
FIGURE 7.4: *IN VITRO* GLPG ASSAY OF SACCHARIN INHIBITORS. 174

List of Tables

TABLE 1.1: RHOMBOID CRYSTAL STRUCTURES 11

TABLE 4.1: DOCKING SCORES OF DIFFERENT AMINO ACIDS IN THE P1 POSITION. 93

TABLE 4.2: SCORES OF THE MOST FAVORABLE AMINO ACIDS IN THE P3 POSITION WITHIN THE TRIPEPTIDE. 96

TABLE 4.3: SCORES OF THE MOST FAVOURABLE AMINO ACIDS IN THE P3 POSITION WITHIN THE TETRAPEPTIDE. 97

TABLE 4.4: RHODANINE-3-ACETIC ACID (RAA) DERIVATIVES. 107

TABLE 4.5: DOCKING RESULTS OF RHODANINE-3-ACETIC ACID DERIVATIVES. 108

TABLE 4.6: TEST SET OF SACCHARINS WITH 50 DIFFERENT LEAVING GROUPS. 115

TABLE 4.7: DOCKING RESULTS OF THE TEST SET SACCHARINS, DISPLAYING THE FIVE BEST LEAVING GROUPS. 117

TABLE 4.8: PRELIMINARY BATCH OF N-METHYLATED SACCHARIN DERIVATIVES. 120

TABLE 4.9: N-METHYLATED SACCHARIN-BASED ACID DERIVATIVES. 122

TABLE 4.10: IC50 VALUES FOR RHOMBOID PROTEASE (GLPG) INHIBITION BY NOVEL SACCHARIN-BASED ACID DERIVATIVES 125

TABLE 4.11: TEMPLATE OF 2-SUBSTITUTED BENZOXAZIN-4-ONES AND ITS DERIVATIVES, SELECTED FOR THE DESIGN OF RHOMBOID INHIBITORS. 126

TABLE 4.12: DOCKING RESULTS OF THREE BEST 2-SUBSTITUTED BENZOXAZIN-4-ONES DERIVATIVES. 128

TABLE 4.13: IC50 VALUES FOR OF RHOMBOID PROTEASE (GLPG) INHIBITION BY NOVEL BENZOXAZINONE-BASED COMPOUNDS (N=1). 130

TABLE 4.14: SELECTED MOLECULAR LIBRARY TO SELECTED FOR RHOMBOID INHIBITORS. 131

TABLE 4.15: VIRTUAL SCREENING COMPOUNDS SELECTED AS PUTATIVE RHOMBOID PROTEASE INHIBITORS. 134

TABLE 4.16: IC50 VALUES FOR INHIBITION OF A-CHYMOTRYPSIN BY NOVEL SACCHARIN-BASED ACID DERIVATIVES. 139

TABLE 4.17: IC50 VALUES OF N-METHYLATED SACCHARIN INHIBITORS IN THE GLPG *IN VITRO* ACTIVITY ASSAY WITH THE LACY SUBSTRATE. 141

TABLE 4.18: IC50 VALUES FOR INHIBITION OF *AQUIFEX AEOLICUS* RHOMBOID (AAROM) BY N-METHYLATED SACCHARIN INHIBITORS (N=1). 142

TABLE 5.1: IC50 VALUES OF KNOWN RHOMBOID PROTEASE INHIBITORS IN A GLPG ACTIVITY ASSAY. 147

Abbreviations

α	alpha	DMF	N,N-dimethyl formamide
Å	Ångström	DMSO	dimethylsulfoxide
Ac	acyl	*E.coli*	*Escherichia coli*
Ar	aryl	ee	enantiomeric excess
aq.	aqueous	Enz	enzyme
Bn	benzyl	eq	equivalent
β	beta	Et	ethyl
BOC	t-butoxycarbonyl	Et_3N	triethylamine
BSA	bovine serum albumin	EtOAc	ethyl acetate
BSc	Boris Schmidt substance database code	EDAC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
Bu	butyl	EGF	Epidermal growth factor
^{o}C	degree centigrade	EGFR	epidermal growth factor receptor
cbz	benzyloxycarbonyl	EI	electron impact
conc.	concentration	ESI-MS	electron spray ionization mass spectrometry
C-terminal	carboxy-terminal	EWG	electron withdrawing group
CAPF	phosphonofluoridate inhibitor, Cbz-Ala^{P}(O-iPr)F	FCS	fetal calf serum
CMK	chloromethyl ketone	FRET	fluorescence resonance energy transfer
CTF	carboxy-terminal fragment	g	gram
d	doublet	GC-MS	Gas Chromatography-Mass Spectrometry
dd	doublet of doublets	h	hour
ddd	doublet of doublets of doublets	HEK	Human embryonic kidney
dt	doublet of triplets	HOBt	Hydroxybenzotriazole
dr	diastereomeric ratio	HPLC	high performance liquid chromatography
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene	HTS	high throughput screening
DCC	N,N'-Dicyclohexylcarbodiimide	Hz	Hertz
DCM	dichloromethane	IC	isocoumarin
DFP	diisopropyl fluorophosphonate	IC50	the half maximal inhibitory concentration
DIPEA	diisopropylethylamine	I-CLiPs	intermembrane-cleaving proteases
DMAP	4-dimethylaminopyridine	J	coupling constant
DMEM	Dulbecco's Modified Eagle Medium complete	kDa	Kilo Dalton
Ki	binding affinity	PBS	phosphate-buffered saline

L	loop	q	quartet
LG	leaving group	®	registered sign
m	multiplet	rpm	revolutions per minute
m	milli	Rho	rhomboid
mg	miligram	RT	room temperature
min	minutes	s	singlet
M	molar (solution)	Ser	serine
M^+	molecular ion (molecular mass)	S2P	site-2 protease family
Me	methyl	SAR	structure activity relationship
Mp.	melting point	SPP	signal peptide peptidases
MOE	Molecular operating environment	T	temperature
MS	mass spectroscopy	t	triplet
m/z	mass to charge ratio	t	tertiary
n	nano	TBS	tris-buffered saline
N	Normal (solution)	TFA	trifluoroacetic acid
NaCl	sodium chloride	THF	tetrahydrofuran
NaH	sodium hydride	TLCK	tosyllysine chloromethyl ketone hydrochloride
Na_2SO_4	sodium sulfate	TMEDA	N,N,N,N-tetramethylethylenediamine
NH_4Cl	ammonium chloride	TM	transmembrane
Nu	nucleophile	TMD	transmembrane domains
NMR	nuclear magnetic resonance	TMS	trimethylsilyl
NTF	amino-terminal fragment	TPCK	tosyl phenylalanyl chloromethyl ketone
o	ortho	TSAP6	tumor suppressor activated pathway-6
p	para	µM	micromolar
P. stuartii	*Providencia stuartii*	v/v	volume to volume ratio
PARL	presenilin associated rhomboid-like	w/w	weight to weight ratio

Amino Acid	Three letter code	One letter code
Alanine	Ala	A
arginine	Arg	R
asparagine	Asn	N
aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

1 Introduction

1.1 Proteases

Proteases are ubiquitously distributed in all types of organisms. It is the largest class of enzymes in most eukaryotes and they catalyse the breakdown of proteins by hydrolysis of peptide bonds. According to bioinformatic analysis, the human genome comprises at least 500-600 proteases, which corresponds to about 2% of human genes [1, 2]. They are responsible for a diverse number of biological processes such as neuronal outgrowth, antigen presentation, cell-cycle regulation [3, 4], blood coagulation [5], immune and inflammatory cell migration and activation, and apoptosis [6]. Dysregulated protease functions are known to be involved in a variety of diseases like cancer [7], arthritis [8], osteoporosis, neurodegenerative disorders [9] and cardiovascular diseases [10, 11], which makes them attractive targets for drug development [12-14].

1.1.1 Proteases classification

On the basis of residues involved in the enzyme catalytic mechanism, proteases are classified into six major groups: serine, cysteine, threonine, aspartate, metallo, and glutamic acid proteases [15]. The general mechanism of proteolytic action involves the activation of a nucleophile, which attacks the carbonyl carbon of the scissile peptide bond. This results in the formation of a negatively charged tetrahedral intermediate. Later, this covalent intermediate hydrolyses and releases two peptide fragments [16] (Figure 1.1).

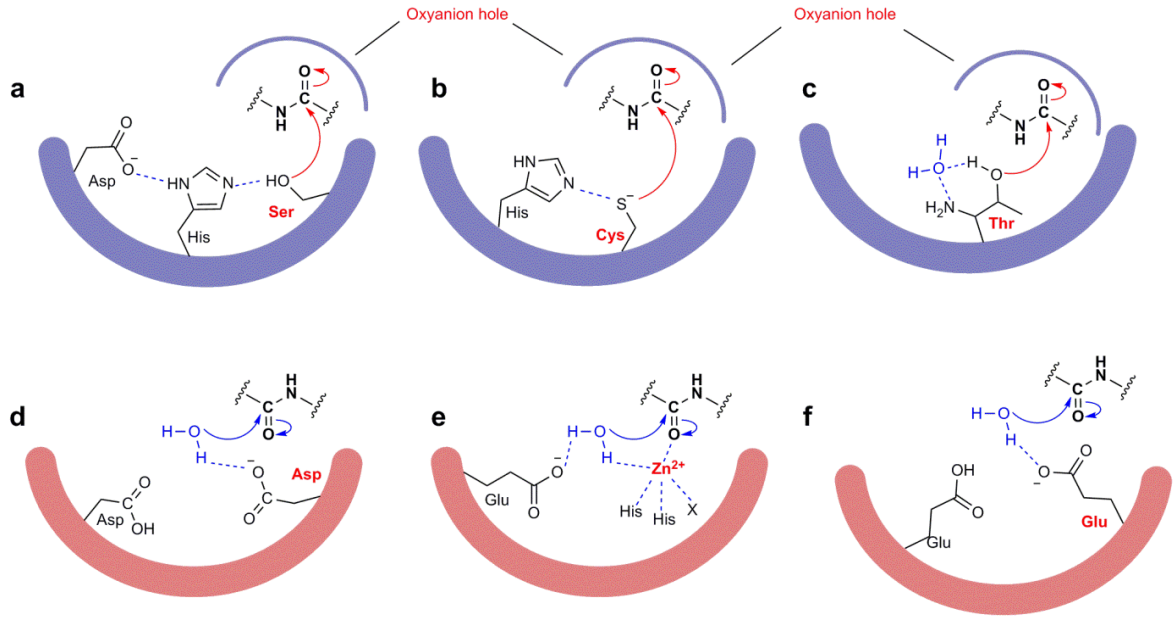


Figure 1.1: Mechanism of the first step of peptide hydrolysis in the six general protease families. (a) Serine, (b) cysteine, and (c) threonine proteases promote nucleophilic attack of the peptide carbonyl carbon by serine, cysteine, or threonine through activating interactions with a proximal histidine side chain. The oxyanion hole stabilizes the tetrahedral intermediate formed after initial nucleophilic attack of the peptide carbonyl and during hydrolysis of the acyl enzyme intermediate. (d) The aspartyl and (f) glutamic acid proteases coordinate a water molecule that is activated through abstraction of a proton, while (e) the metalloproteases use glutamate and a coordinated metal ion (often zinc) to activate a water molecule to attack the peptide carbonyl carbon. Arrows indicate the movement of electrons and dotted lines indicate hydrogen bonding or electrostatic interactions. The peptide substrate is shown in bold.

Aspartate, metallo, and glutamic acid proteases use an activated water molecule as a nucleophile to attack the carbonyl carbon of the peptide bond and to form a tetrahedral intermediate (Figure 1.1) [17]. In contrast, in the other three classes, the nucleophile involved in hydrolysis is a catalytic amino acid residue (Ser, Cys or Thr respectively). This catalytic residue is located in the enzyme active site and maintained in a nucleophilic state by interactions with an activating base, generally histidine [18, 19], while serine proteases consist of a catalytic triad (Ser-His-Asp). The catalytic process further involves a tetrahedral transition state, which is stabilized by the oxyanion-binding hole of the enzyme (Figure 1.1).

1.1.2 Intramembrane proteases

Over the past fifteen years, the most surprising and interesting class of proteases discovered have been the intramembrane proteases (IMPs), which demonstrated that proteolytic activity is not limited to soluble proteases and that IMPs can cleave other transmembrane proteins [20]. All known intramembrane proteases are themselves integral membrane proteins with multiple transmembrane domains (TMDs), and their active sites are buried within the lipid bilayer of cellular membranes (Figure 1.2). IMPs have been considered unique since they are able to hydrolyse their transmembrane substrates despite the water-poor environment of the lipid bilayer. Therefore, they are also called intermembrane-cleaving proteases (I-CLiPs). The cleavage reaction leads to the liberation of N- or C-terminal substrate fragments (or both) from the membrane. Due to the presence of common proteolytic features, IMPs were initially linked to soluble proteases regarding their evolutionary origin. However, it is now understood that intramembrane proteases are not evolutionarily related to classical soluble proteases and that their catalytic sites originated as a result of convergent evolution [16].

Based on distinct hydrolytic mechanisms, three major families of IMPs have been described: 1) metalloproteases, for example the site-2 protease family (S2P) [21]; 2) aspartyl proteases, which include γ-secretase and the signal peptide peptidases (SPP) [20, 22]; 3) rhomboid-like proteases, which are serine proteases. The first IMP identified was mammalian S2P, This IMP is involved in human cholesterol metabolism and liberates the N-terminal fragment of sterol regulatory element-binding protein (SREBP), which acts as a transcription factor, from the membrane [21, 23, 24]. It was also found that the helical membrane system of S2P contains a conserved HExxH sequence, which is also present in water-soluble zinc metalloproteases and is considered to have an important role in coordination bond formation with the zinc metal [25]. The aspartyl IMPs feature two aspartate residues in their active sites. γ-secretase is a multi-subunit complex consisting of four different proteins, with presenilin being the catalytic subunit. One substrate of γ-secretase is the amyloid precursor protein (APP). Cleavage of APP by γ-secretase produces amyloid-β (Aβ) peptides, which aggregate and form senile plaques in the brain, a process considered to be a major cause for the degeneration of nerve cell in Alzheimer's disease [26-28]. Signal peptide peptidases (SPP) cleave

signal peptides, regulate viral reproduction rate (Hepatitis C virus), and have key roles in immune surveillance [22, 29-31].

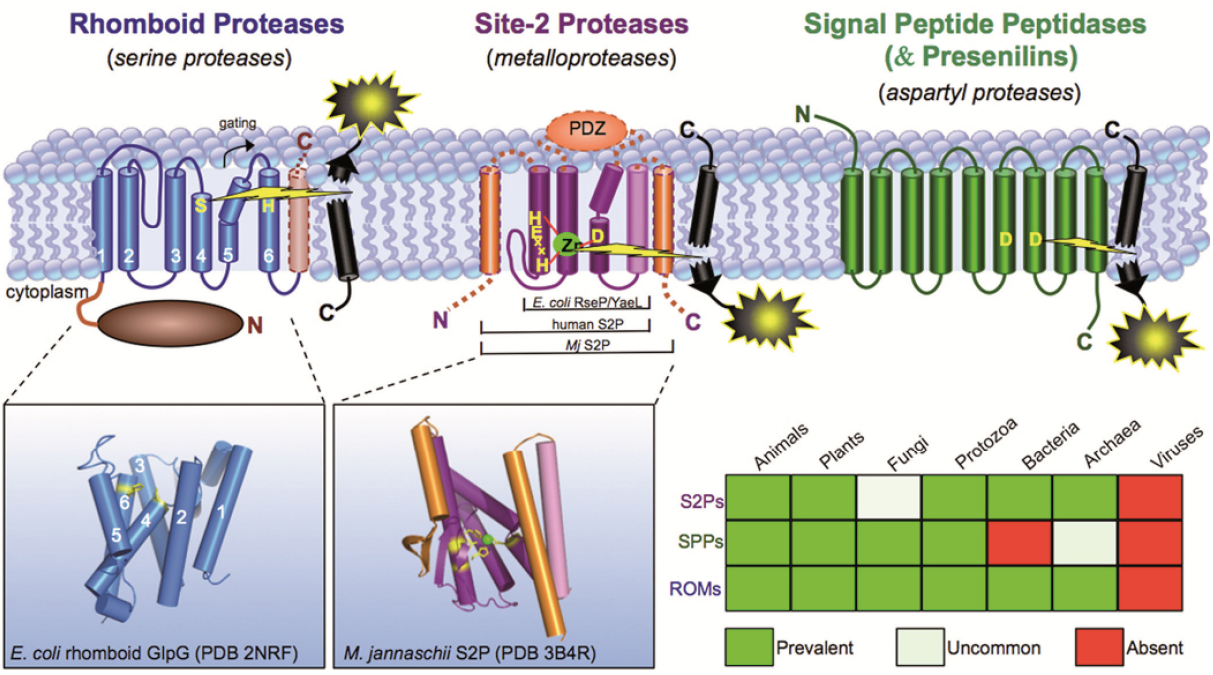


Figure 1.2: Schematic representation of basic topologies of the three families of intramembrane proteases (IMPs). Active site residues are highlighted in yellow, and the metal zinc ion of site 2 protease (S2P) is shown in green. Protein substrate are indicated in grey, the yellow flash depicts the proteolytic event, and the yellow star indicates the released substrate domain with a known biological function. The core six transmembrane domains (TMD) of rhomboids are shown in blue, and the additional TMD and the cytosolic domain is indicated in brown color. The conserved three TMDs of S2P are shown in dark purple, and the rest of the protein in light purple and orange color with an extramembrane domain (frequently a PDZ domain) also in orange. Signal peptide peptidases (SPP) with nine TMDs are shown in green. Presenilins also consist of nine TMDs but with opposite orientation. The additional structures below show the membrane core domains in crystal structures of the *Escherichia coli* rhomboid GlpG (PDB : 2NRF) and *Methanocaldococcus jannaschii* S2P (PDB : 3B4R). The lower right figure illustrates the conservation of IMPs in all life forms. Figure adopted from [32].

The most recently and structurally best characterized class of IMPs is the rhomboid–like family, which is the focus of this thesis. Rhomboids are intramembrane serine proteases conserved in all domains of life (Figure 1.2) and carry out a wide variety of biological functions [33, 34]. Differentiating their proteolytic mechanism from the other two types of IMPs, rhomboids do not require preprocessing of their substrates by other enzymes and they release factors to the outside of cells rather than to the cytosol [16, 35].

1.2 Rhomboid proteases

In general, rhomboid proteases follow the same catalytic mechanism described for soluble serine proteases. However, rhomboid proteases use a catalytic dyad (Ser-His) instead of a catalytic triad (Ser-His- Asp) (Figure 1.3).

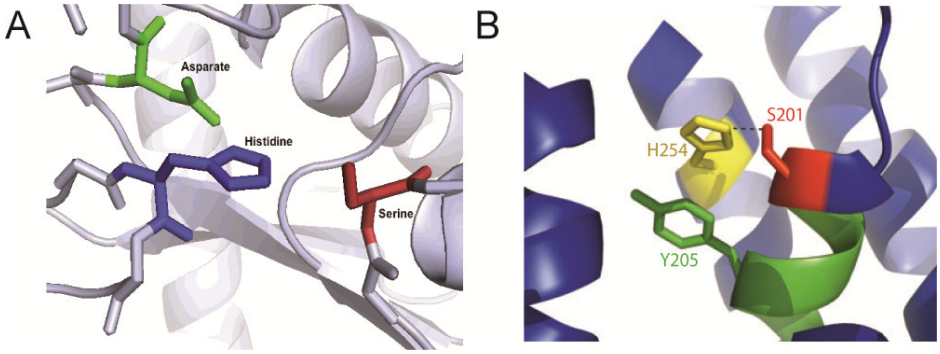


Figure 1.3: Catalytic residues of serine proteases. A) Soluble serine protease (SC-PEP) with Ser- His- Asp catalytic triad. B) Rhomboid protease (GlpG) with Ser201-His254 catalytic dyad. The interaction between Ser and His is indicated by a dashed line. The catalytic His254 is aligned with Tyr205 (not conserved in all rhomboids), potentially stabilizing the former's orientation and interaction with the catalytic Ser201. Figure adopted from http://2011.igem.org/Team:Washington/Celiacs/Background and [36].

The first rhomboid protease was discovered through a mutation in the rhomboid-1 (Rho-1) gene of *Drosophila*, which caused a rhomboid-shaped head skeleton [37]. Further studies revealed that the *Drosophila* Rho-1 is involved in the activation of the epidermal growth factor receptor (EGFR) signalling pathway (Figure 1.4) [38]. Rho-1 cleaves the transmembrane substrate Spitz, which leads to the release of an EGF domain from the cell surface into the extracellular space. This activated ligand binds the EGF receptor on neighboring cells and triggers downstream responses that influence the embryonic development of *Drosophila* [39, 40].

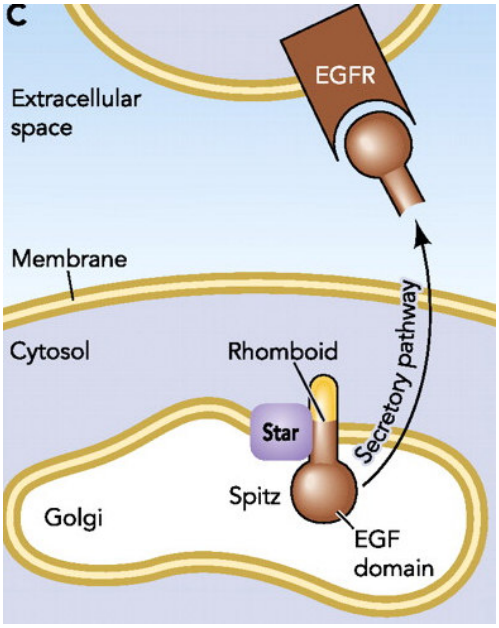


Figure 1.4: Rhomboid regulated EGFR signaling pathway in *Drosophila melanogaster*. The rhomboid proteolytically cleaves its substrate Spitz within the TMD of Golgi apparatus and releases EGF receptor ligand. This can activate the EGF receptors of adjacent cells in extracellular space. Figure adopted from [41].

1.2.1 Rhomboid protease functions and their importance in diseases

Rhomboid proteases are present in all branches of life, bacteria, archaea, and eukaryotes [33], and by cleaving other transmembrane proteins exert a variety of biological functions, some with potential biomedical relevance [36, 42, 43]. As described above, rhomboids were first discovered to trigger

EGFR signalling in *Drosophila* (Figure 1.4 and Figure 1.5). *Providencia stuartii*, a pathogenic bacterium that causes urinary tract infections in humans [44], contains a rhomboid named AarA. This rhomboid was found to be involved in bacterial quorum sensing (Figure 1.5), a chemical form of communication that allows single cells to measure the density of a cell population [39, 45, 46]. AarA cleaves the N-terminal domain of the transmembrane protein TatA, which leads to the release of a signal for quorum sensing. TatA of *P. stuartii* (consisting of a transmembrane helix and a short N-terminal domain) is a subunit of the twin-arginine translocase, which is located in the bacterial cell membrane and known to be involved in protein transport [46, 47]. This implies that, with the use of inhibitors of the AarA rhomboid, quorum sensing and urinary tract infections due to *P. stuartii* might be controlled.

Another example are rhomboids in the unicellular eukaryotic apicomplexan parasites, which were found to be involved in the proteolysis of adhesin (Figure 1.5). Adhesins are cell surface proteins expressed in parasites that bind to receptors on the host cell and are required for host invasion [48-51]. In *Toxoplasma* the rhomboid TgROM5 has been identified to cleave the MIC adhesin [51-53], and in *Plasmodium* the rhomboid PfROM4 is responsible for the proteolysis of the adhesin EBA -175. PfROM4 has also been demonstrated to influence the parasitic life cycle; it has been shown that a mutation in the TMD of the rhomboid prevents the growth of the parasites [50, 51, 54, 55]. Based on these observations it was proposed that the parasitic rhomboids have functions that go beyond host cell invasion and affect many other aspects of their life cycle. Hence, specific inhibitors against these rhomboids could provide potential drugs to treat these parasitic diseases.

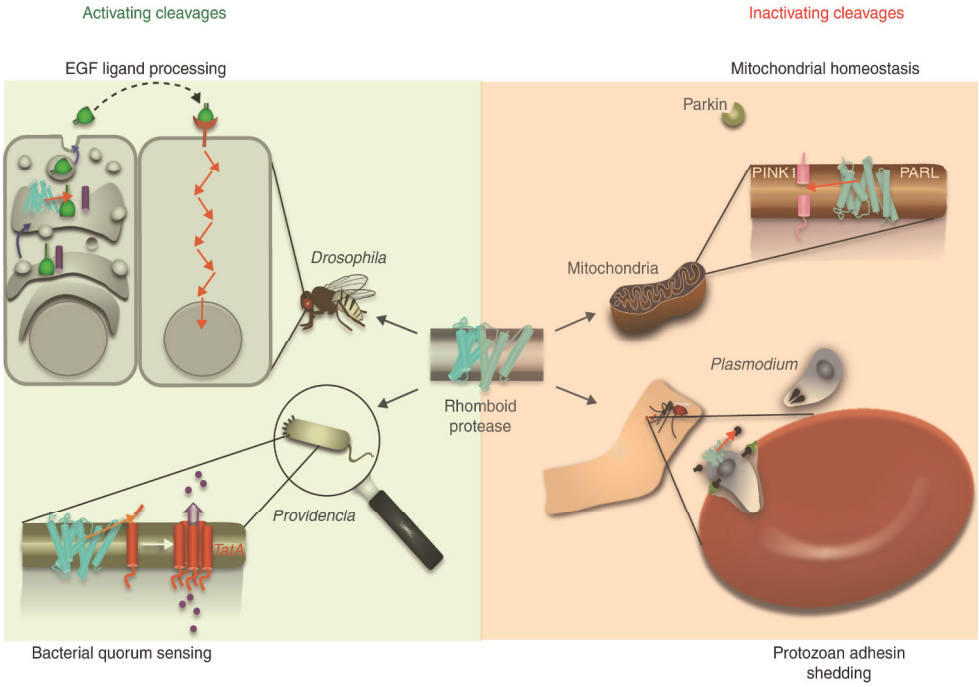


Figure 1.5: Four biological functions of rhomboid proteases. Top left: Rhomboid-1 (blue) is localized in the Golgi apparatus and cleaves the substrate Spitz (green), which leads to the secretion of a soluble EGF domain and stimulation of the EGF receptor signaling pathway in neighboring cells. **Bottom left:** The bacterial rhomboid protease AarA (blue) from *Providencia stuartii* cleaves an aminoterminal extension from its substrate TatA (orange). TatA is a subunit of the machinery responsible for quorum sensing. **Bottom right:** Adhesins are cleaved by rhomboids (blue) in the parasites *Plasmodium* and

Toxoplasma, a process important for efficient invasion of the host cell. **Top right:** Pink1 (pink) is cleaved by the rhomboid Parl (blue), which allows Pink1 to recruit Parkin to damaged mitochondria. Figure adopted from [56].

Rhomboids have also been identified in mitochodria. The first mitochondrial rhomboid Rbd1 was found in the yeast *Saccharomyces cerevisiae* and it is known to cleave two substrates, cytochrome c peroxidase (Ccp1) and Mgm1 (dynamin-like GTPase). Mgm1 regulates the mitochondrial membrane dynamics in conjunction with other GTPases. Additional investigations showed that Mgm1 hydrolysis by Rbd1 promotes remodeling of the mitochondrial membrane, and that loss of either Mgm1 or Rbd1 leads to similar phenotypic alterations with mitochondrial fragmentation. In contrast to Mgm1, loss of Ccp1 caused no changes in the mitochondria and is therefore assumed to be less significant, whereas Mgm1 appears to be an essential substrate for Rbd1 [57-61]. Further experiments and sequence comparisons have shown that all eukaryotes have mitochondrial rhomboids [36].

The mammalian mitochondrial rhomboid, presenilin associated rhomboid-like (PARL), was initially found to cleave OPA1 (a homologue of Mgm1), and early investigations indicated that PARL regulates mitochondrial function as well as their morphology. Later, it was shown that PARL promotes proteolytic degradation of the kinase PINK1. PINK1 acts as a sensor for damaged mitochondria, which can be removed from the cell in a process known as mitophagy (Figure 1.5), and mutations of PINK1 lead to a recessively inherited form of Parkinson's disease (PD)[62-64]. Another substrate of PARL is PGAM5 (phosphoglycerate mutase 5), whose cleavage can promote apoptosis [65], and PARL has also been associated with type 2 diabetes [66-68]. Moreover, *Drosophila* mitochondrial *rhomboid-7* was found to cleave OPA1 and to be involved in processing of the PD-associated proteins PINK1 and HtrA2 [64]. These findings illustrate that mitochondrial rhomboids play important roles in mitochondrial homeostasis and pathophysiological processes.

In addition to mitochondrial rhomboids, four mammalian rhomboid proteases (RHBDL1-4) were also discovered that appear to have an important role in the release of secreted proteins [69]. Several substrates have been suggested for RHBDL2 including thrombomodulin, ephrinB3, EGF, and EGFR [70-73] and it might participate in wound healing [74]. Recent advances indicate that RHBDL4 is structurally different from the other three secretase-type rhomboids, RHBDL1-3, and that it is the only active rhomboid in the endoplasmic reticulum (ER). RHBDL4 has been demonstrated to regulate apoptosis and exosome trafficking through proteolysis of the pro-apoptotic protein BIK and the polytopic membrane protein TSAP6 [75-77]. Additionally, it has been shown that RHBDL4 might contribute to ER-associated degradation (ERAD) of proteins [71]. However, overall very little is known about the physiological functions and substrates of the mamalian secretase-type rhomboids and more detailed investigations of these rhomboids are needed.

1.2.2 Rhomboid like super-family classification

A systemic classification of rhomboids was presented by Freeman in 2008 with a subdivision into three groups: active rhomboids, inactive rhomboids (iRhoms) and rhomboid-like proteins [36]. More recently, Freeman designated this class of IMPs as the 'Rhomboid-like superfamily', excluding the evolutionary relationships (Figure 1.6) [43]. It is believed that all rhomboids had a common ancestor and that this was probably an active intramembrane protease. The proteolytically inactive iRhoms are highly conserved in the metazoan kingdom, Only some iRhoms lack active site residues but all feature a proline at -1 position to the active serine, which is predicted to abrogate the catalytic activity [36].

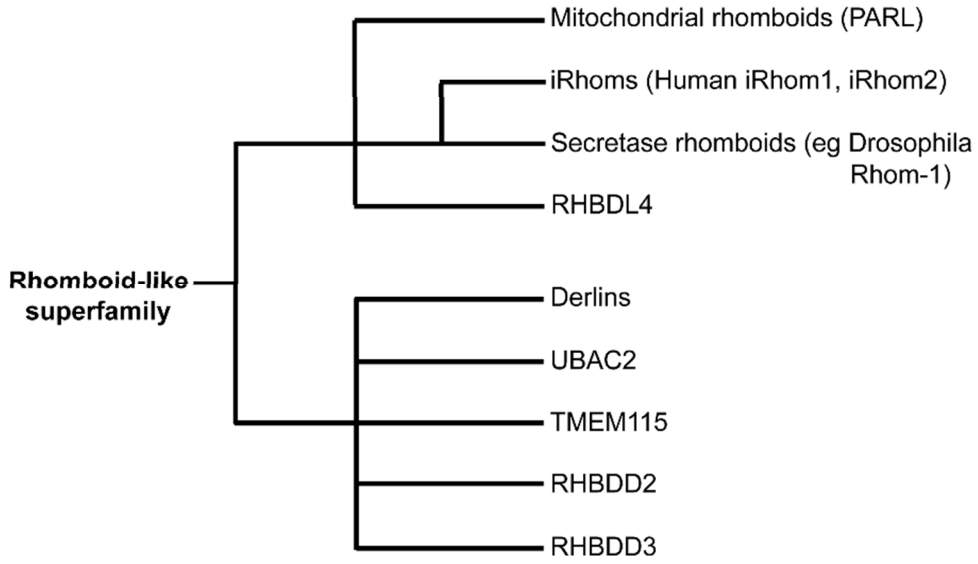


Figure 1.6: Rhomboid-like superfamily. Adopted and modified from [43].

Besides active rhomboids and iRhoms, many more branches of rhomboids (TMEM115, UBAC2, RHBDD2, RHBDD3) have been found, which have diverged extensively but clearly belong to the rhomboid-like superfamily [43].

1.2.3 Rhomboid Structure

1.2.3.1 General Features

Based on sequence analysis rhomboids have been divided into three different topological classes (Figure 1.7). The first class is the simplest and smallest with six transmembrane (TM) helices, representing the catalytically active core of rhomboid proteases. It is more common among prokaryotic rhomboids with some examples in eukaryotic organisms such as RHBDL4 [56]. The other two classes consist of 7 TM topologies with an additional TM segment adjoining the 6 TM core either at the C-terminus (6+1 TM form) or at the N-terminus (1+6 TM form). These predominate in eukaryotic rhomboids [33, 56, 78], but a few examples of the 7 TM topology (6+1) have been found in prokaryotic rhomboids, for example AarA and YqgP. In contrast, the 1+6 TM topology is confined to endosymbiotic organelles like mitochondria and plastids. Among the non-catalytic rhomboid proteins,

iRhoms acquired the 6+1 TM topology with large domains at the N-terminus as well as between TM1 and TM2, whereas derlins consist of the 6TM core form [79].

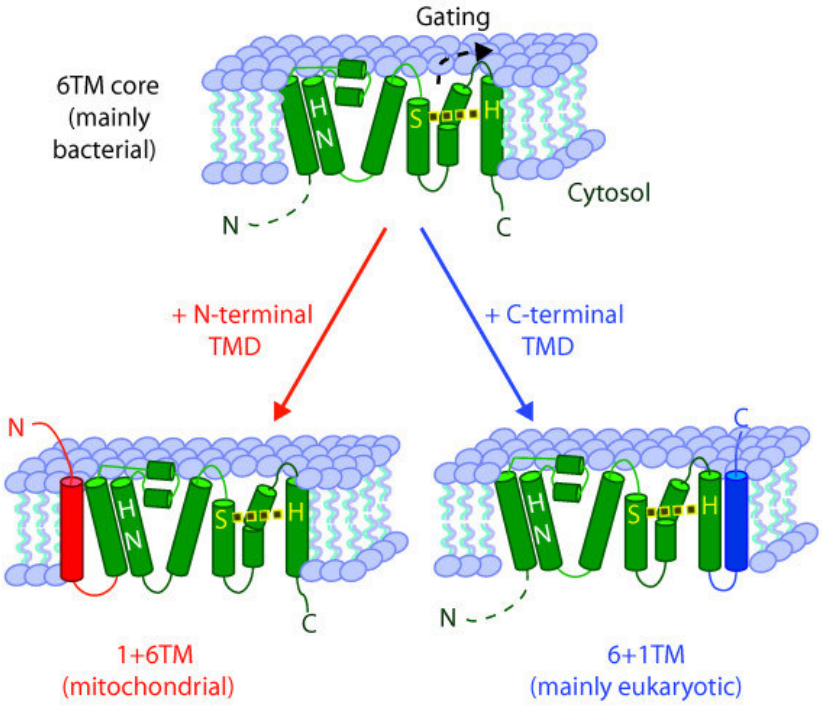


Figure 1.7: Three topological forms of rhomboid proteins. The simplest catalytically active form of the rhomboid proteases consists of a core of 6 TM, with variable N-termini represented by dashed lines. Mainly mitochondrial and eukaryotic rhomboid proteases display the 7 TM form with an additional TM segment added to the 6 TM core either amino-terminally (1+6) or carboxy-terminally (6+1). Catalytic residues involved in hydrolysis are depicted in yellow and electrophilic residues supporting the oxyanion transition state are depicted in white. Figure adopted from [56]

1.2.3.2 Rhomboid crystal structure and catalytic mechanism

Rhomboid proteases cleave other TM proteins and, according to the general structural nomenclature for all proteases, residues flanking the cleavage bond are referred to P1 on the N-terminal side and P1' on the C-terminal side with subsequent numbering from the inside to the ouside (Figure 1.8). The corresponding subsites of the enzyme accommodating these residues are called S1 and S1' with similar outward numbering [80]. Although topology predictions and biochemical analysis of specific residues have contributed initial insights into the rhomboid structure, detailed insights into the catalytic mechanism were only later provided by X-ray crystallography studies.

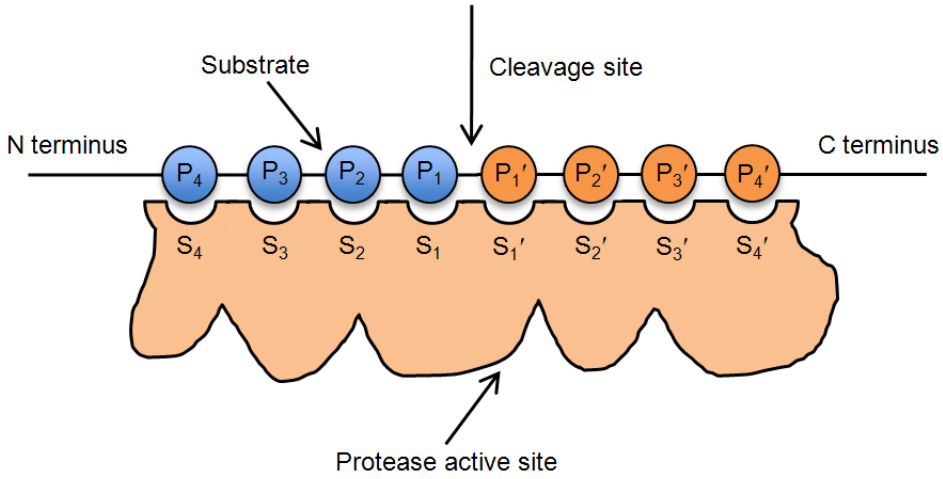


Figure 1.8: The nomenclature of the substrate sequence and the protease subsites. The cleavage site is located between P1 and P1' (indicated by an arrow). The amino acids residues in the substrate sequence are named P1-P2-P3-P4 oriented towards the N-terminus from the scissile bond (shown in blue color) and P1'-P2'-P3'-P4'- towards the C-terminus (shown in orange color), while the corresponding subsites of the protease are labelled as -S4-S3-S2-S1-S1'-S2'-S3'-S4'- (adopted from https://prosper.erc.monash.edu.au/nomenclature.png).

In 2006, Wang *et al*. presented the first intramembrane protease crystal structure of native *E. coli* rhomboid (EcGlpG) [81]. This structure showed that the proteolytic core of rhomboids consists of six helical TMs, arranged in form of a bundle. The catalytic residue Ser201 was found to be located in the short TM4 at a depth of ~ 10 Å within the cell membrane, further surrounded by the other five TMDs. The catalytic residue His254 was located in TM6, and it was described that the active site residues along with their surrounding residues create a hydrophilic pocket which opens to the periplasm through a flexible cap called loop 5 (L5). L5 was the connecting chain between TM5 and TM6 and proposed to play a major role in the catalytic gate mechanism of rhomboids (Figure 1.9).

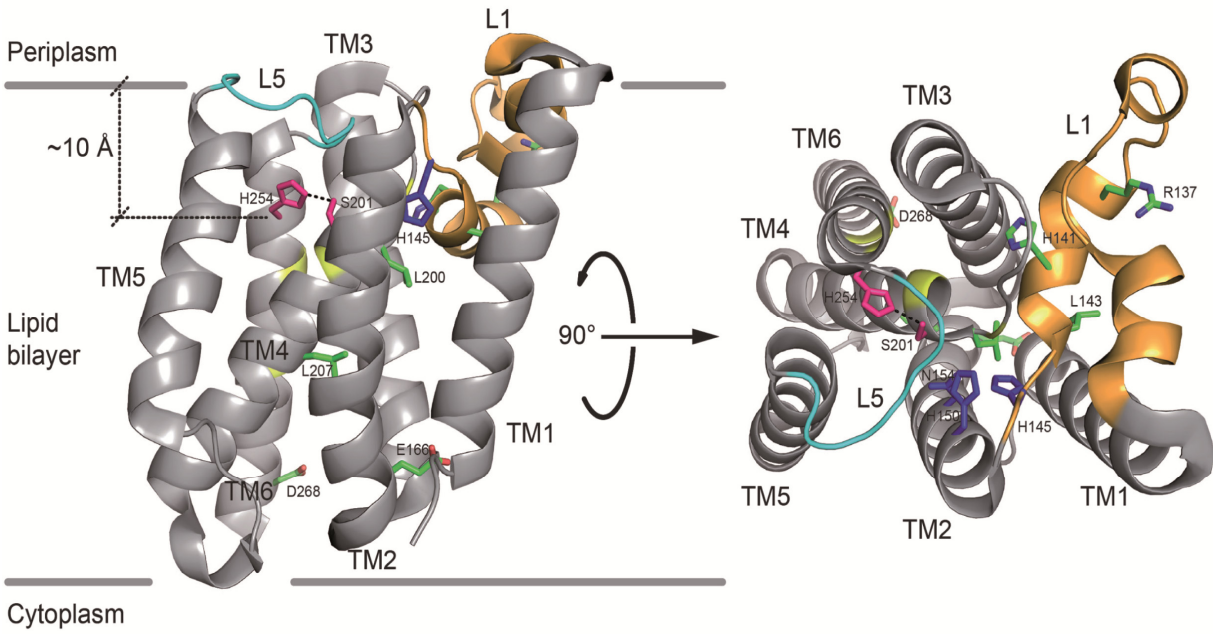


Figure 1.9: Crystal structure of the rhomboid GlpG from E. coli (PDB code: 2XOV). Left figure represents the side view and right figure shows the top view. The catalytic residues serine 201 and histidine 254 are located ~ 10 Å below the surface of the cell membrane (pink). The flexible loop L5 (blue) between TM5 and TM6 in the displayed conformation closes the access to the active site. The L1 loop between TM1 and TM2 (orange) is important for the stabilization of the conformation of the GlpG. Figure adopted from [82].

These structural features and insights into the catalytic mechanism were subsequently supported by many additional crystal structures of native EcGlpG and native *H. influenzae* GlpG [83-85], and considerable further advancements were made by solving inhibitor-rhomboid complex structures [86-91] (all known crystal structures are listed in Table 1.1). In fact, it has been substantiated that a gating mechanism controls substrate binding and proteolysis by rhomboids. This gating involves structural rearrangements for substrate access to the rhomboid active site, and until now two main theories have been proposed. The "lateral gating" theory involves the movement of TM5 and results in opening of a lateral gate between TM2 and TM5. This leads to dislocation of L5 and eventually permits the substrate lateral access to the rhomboid active site. This theory was further supported by biochemical experiments in which mutations resulting in weak TM2/TM5 packing enhanced the protease activity [85, 92, 93].

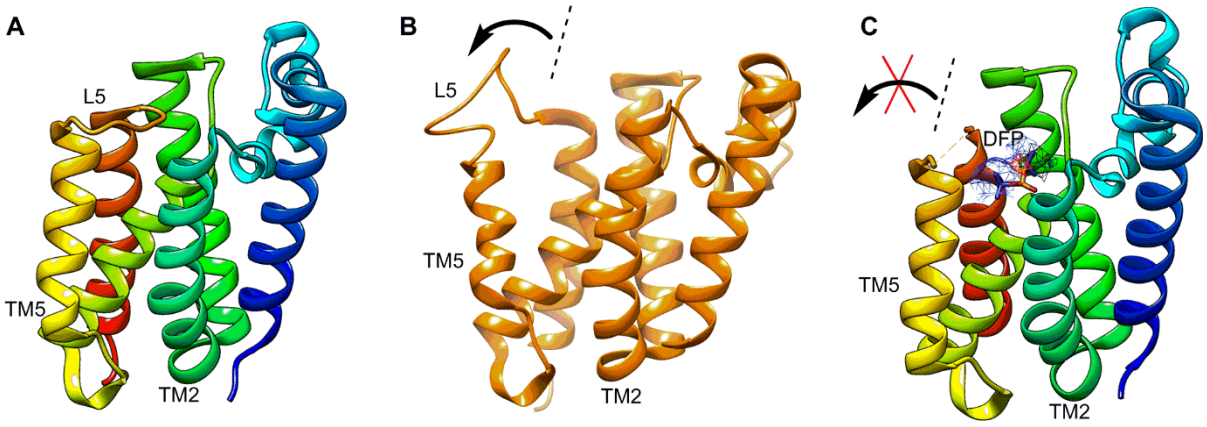


Figure 1.10: Gating mechanism of the *E. coli* rhomboid protease GlpG. A) Closed conformation of native GlpG (PDB : 2IC8) [81]. B) Open conformation of GlpG (PDB : 2NRF) [85]. TM5 helix rotates away from the main body of the GlpG (indicated by an *arrow*) C) Cocrystal structure of GlpG with DFP indicating only a small lateral movement of TM5 [94] . The *dashed line* indicates the orientation of TM5 in the closed conformation.

On the contrary, the second theory postulated that the L5 cap of rhomboid alone controls the substrate access to the active site [95]. This idea was supported by experiments showing no significant decrease in substrate proteolysis by GlpG after chemical cross-linking of TM2 and TM5 restricting any lateral movement [94]. In addition, cocrystallization studies of GlpG in complex with diisopropyl fluorophosphonate (DFP) inhibitor did not demonstrate a significant movement of TM5 (Figure 1.10C) [94]. However, both theories support that proteolysis can take place within the cell membrane, with the rearrangement of the helices leading to water accessiblity of the active site and the gating mechanism controling the overall catalytic reaction kinetics [96].

Table 1.1: Rhomboid crystal structures

Structures	PDB ID	Space group, resolution (Å)	Features	Ref
Native structures				
EcGlpG	2IC8	R32, 2.10	Native	[81]
EcGlpG	2IRV	$P2_1$, 2.30	Native	[83]
EcGlpG	2XOV	R32, 1.65	native	[87]
EcGlpG	3B45	R32, 1.90	Native	[97]
EcGlpG	2O7L	R32, 2.50	open cap structure	[95]
EcGlpG	2NRF	$P3_1$, 2.60	lateral gate open structure	[85]
EcGlpG	4NJN	H32, 2.40	Crystal Structure of *E. coli* GlpG at pH 4.5	[96]
EcGlpG	4NJP	H32, 2.40	rate-governed reaction not Driven by substrate affinity	[96]
hiGlpG	2NR9	C2, 2.20	Native	[84]
hiGlpG	3ODJ	C2, 2.84	Disordered loop 4, TM5 and loop5	[98]
Inhibitor structures				
EcGlpG: isocoumarin	2XOW	H32, 2.09	ISM mechanism-based inhibitor	[87]
EcGlpG: isocoumarin	3ZEB	H32, 2.20	S106, mechanism-based inhibitor	[88]
EcGlpG: DFP	3TXT	H32, 2.30	GlpG in complex with a mechanism-based inhibitor, DFP	[90]
EcGlpG: DFP	4H1D	H32, 2.90	Large lateral movement of transmembrane helix S5 is not required	[94]
EcGlpG: CAPF	3UBB	H32, 2.60	GlpG in complex with a phosphonofluoridate inhibitor	[89]
EcGlpG: β-lactam	3ZMH	H32, 2.30	GlpG in complex with monobactam L62	[86]
EcGlpG: β-lactam	3ZMI	H32, 2.20	GlpG in complex with monobactam L29	[86]
EcGlpG: β-lactam	3ZMJ	H32, 2.30	GlpG in complex with monobactam L61	[86]
EcGlpG: β-lactam	3ZOT	H32, 2.40	GlpG in complex with monobactam L29	[86]
EcGlpG: cmk	4QO0	$P6_3$, 2.9	GlpG in complex with peptide derived inhibitor Ac-FATA-cmk	[91]
EcGlpG: cmk	4QO2	$P6_3$, 2.1	GlpG in complex with Ac-IATA-cmk	[91]

EcGlpG: cmk	4QNZ	$P6_3$, 2.55	GlpG F146I in complex with peptide derived inhibitor Ac-FATA-cmk	[91]

Further insights into the catalytic mechanism of rhomboids were achieved through co-crystal structures with various inhibitors (Table 1.1). The first structure with an isocoumarin inhibitor (JLK-6,Figure 1.23) displayed the formation of two covalent bonds, one with the active site Ser201 and other with the active site His254 [87]. The catalytic mechanism involved the nucleophilic attack of the catalytic serine residue on the carbonyl group of the inhibitor resulting in ring opening. This is followed by rearrangements leading to generation of a new electrophile with which His254 reacts forming a second covalent bond. Another important feature noticed in the structure was the formation of an 'oxyanion hole', consisting of several neighbouring residues like His150 and Asn154, which support the carbonyl group of the inhibitor with hydrogen bonding (Figure 1.11). This oxyanion hole formation was also observed in other co-crystal structures of GlpG with DFP [90], CAPF [89], isocoumarin IC 16 [88], N-carbonated β-lactams [86], and a tetrapeptide chloromethyl ketone [91], in which the carbonyl or the phosphoryl oxygen atom interacts with oxyanion hole residues. A notable exception, the carbonyl oxygen points away from these residues in co-crystal structures with β-lactams [86]. Nevertheless, all inhibitor complex structures also indicated that a lift of the L5 loop permits access to the active site with partial dislocation of TM5.

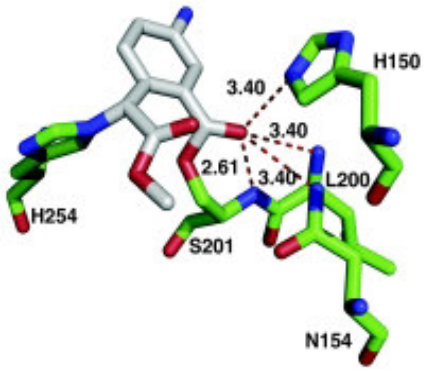


Figure 1.11: Structure of the oxyanion hole in the GlpG rhomboid protease in complex with a covalently bound isocoumarin inhibitor (JLK-6). The most important amino acids participating in binding or hydrogen bonding are labelled. The length of the hydrogen bonds is also annotated. The hydrogen bonds stabilizing the oxyanion are formed by the side chains of histidine 150 (H150) and asparagine 154 (N154) as well as the backbone of serine 201 (S201) and leucine 200 (L200). Figure adopted from [87].

Another interesting discovery was that similar to JLK-6, the isocoumarin IC 16 also formed a second covalent bond but with His150 rather than His254 [88]. This indicated that the role of His150 is not confined to the oxyanion hole and it has been also proposed that the new electrophile formed after rearrangements following the first nucleophilic attack by Ser201 displaced away from His254 because of significant hydrophobic interactions with the side chains of Met149 and Phe153 in TM2 and the side chains of Phe245 and Met247 in L5 [88]. Movement of Phe245 was also observed in structures with CAPF (Figure 1.12), N-carbanoylated β-lactams, and peptidomimetic chloromethyl ketone

inhbitors, resulting in cap rotation and uncovering of the binding site. Hence, this could represent a key feature for substrate entry to the enzyme active site cleft [86, 89, 91].

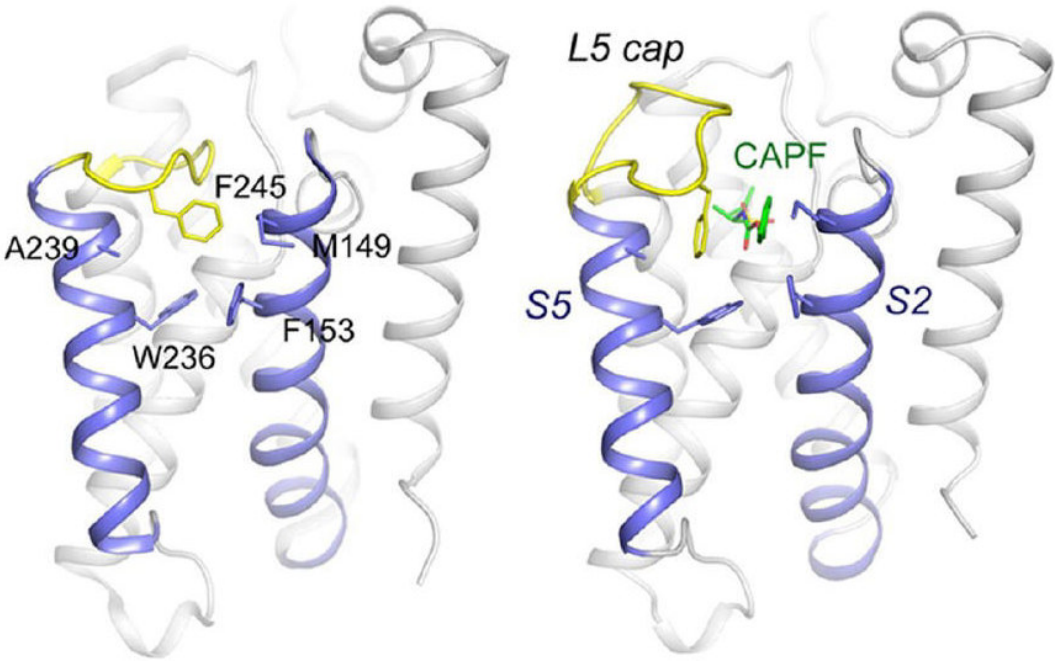


Figure 1.12: Substrate entry regulated by the L5 cap. a) L5 rests upon the active site in the absence of CAPF inhibitor. b) Significant movement of the Phe245 (F245) residue of L5 results in lateral gate opening in the bounded state of GlpG-CAPF. The L5 cap is highlighted in yellow color, and TM2 and TM5 in blue color. CAPF inhibitor colored in green. Figure adopted from [89]

Another key feature observed in the co-crystallized EcGlpG-inhibitor structures was the formation of a substrate recognition site (S' subsite) after interaction with the ligands, accommodating side chains of the inhibitors and, putatively rhomboid protein substrates. Figure 1.13 shows three different crystal structures of EcGlpG with a focus on the active site binding cleft. Figure 1.13A shows the isoform of the rhomboid protease that was used in this work (PDB:3UBB), with EcGlpG bound to CAPF (inhibitor not shown) [89]. Figure 1.13B shows the co-crystal structure of EcGlpG with an isocoumarin inhibitor (inhibitor not shown, PDB:XOW) [87], and Figure 1.13C shows the wildtype rhomboid protease without any bound inhibitor (PDB:2IC8). By comparing these three different structures the formation of a S2' pocket is observed in the two inhibitor-bound structures while the S2' pocket is absent in the structure of wildtype EcGlpG. The inhibitors occupy the S2' subsite of the enzyme and the enzyme adapts a different shape compared to native EcGlpG. Co-crystal structures of EcGlpG with β-lactam and peptidomimetic chloromethyl ketone inhbitors also indicated the formation of a S2' subsite [86, 91].

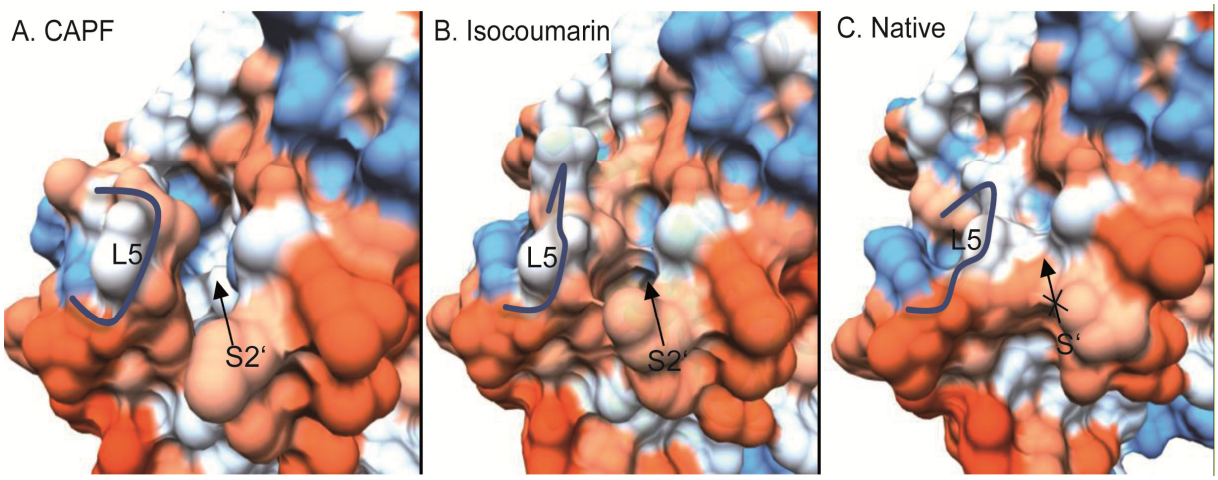


Figure 1.13: Structural comparison of three EcGlpG active sites derived from different crystal structures. A) EcGlpG in complex with CAPF (PDB: 3UBB) B) EcGlpG in complex with isocoumarin inhibitor (JLK-6, PDB : 2XOW).. C) native EcGlpG (PDB : 2IC8). Arrow indicates the location of the S' subsite. Surface colouring according to hydrophobic scale of Kyte and Doolittle: orange = hydrophobic; blue = hydrophilic. The form of the active site changes completely in the different structures, which shows the high flexibility of the GlpG protease. The loop connecting TM5 and TM6 (shown as blue curve) acts as a cap thus shielding the active site.

Other important structural rearrangements were observed in co-crystal strucures of EcGlpG with a tetrapeptide chloromethyl ketone, which mimics the P1-P4 positions of the bacterial rhomboid substrate TatA [91]: 1) The formation of a 'water retention site', which contributes to the S1 subsite and involves residue Gln189 (Figure 1.14 and Figure 1.15). Moreover, displacement of residues Met249, Ala250 and Met247 of the L5 loop appears to unlock the water channel to facilitate proteolysis (Figure 1.14C). 2) Rotation of residue Trp236 residue in TM5 with movement of the L5 loop (Figure 1.14A). 3) The S1 subsite is relatively smaller than the large S2 and S3 subsites, and it is the proximal part of the active site [91] (Figure 1.15). 4) Three side chain residues of loop 1 (Phe146, Met120 and Met144) with little contribution from loop 3 interact with the P4 substituent of the substrate and define a hydrophobic S4 subsite of rhomboid [91]. Overall, these findings indicated that the S1 and S4 subsites of EcGlpG are particularly important for substrate interaction and specificity (see below).

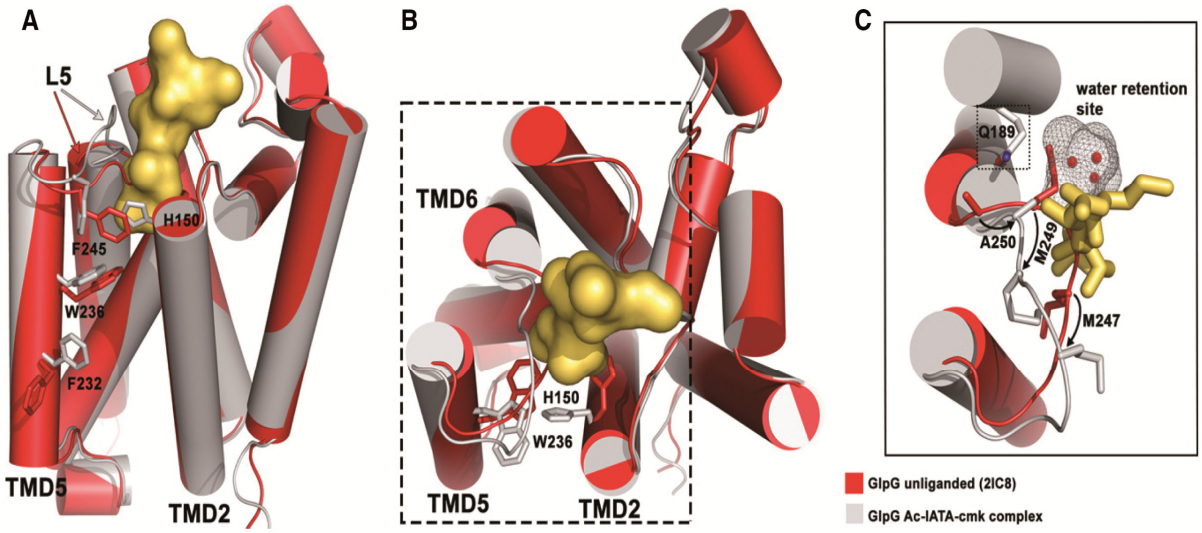


Figure 1.14: Diagrammatic representation for displacement of residues on L5 cap and TMD 2 and TMD5. A) Side view of structurally aligned native GlpG (PDB: 2IC8, shown in red) and ligated GlpG (PDB: 4QO2, shown in grey and yellow) with a peptidomimetic chloromethyl ketone inhibitor (Ac-IATA-cmk). This shows the displacement of residues in L5, TMD2 and TMD5. B) Top view of the same aligned GlpG structures. C) Movement of L5 allows displacement of residues M249, A250 and M247 from their original positions and results in formation of a water retention site. Figure adopted from [91].

1.2.4 Substrate specificity of Rhomboid proteases

Generally, the proteolytic activity and specificity of soluble proteases relies on short binding motif surrounding the scissile bond of the substrate and the interaction of these substrate residues with enzyme subsites [99-101]. Cleavage of substrates by rhomboids within the membrane appears to be governed by the sequence of the substrate but also by additional mechanisms [43]. Up to now, only transmembrane proteins of type I and type III (proteins with a transmembrane α helix and an extracellular N- terminus but without a signal peptide) were identified as natural substrates of rhomboids [82]. As an initial step in substrate recognition prior to the proteolytic cleavage, an interaction of the transmembrane α-helix of the substrate with an exosite region on the surface of the rhomboid might be required but this remains controversial [82] (Figure 1.16). The proteolytic cleavage site of the substrate can be located both inside and outside of a TMD. It has also been demonstrated that different rhomboids from different organisms were able to process the same substrates [102]. It has been tested that short synthetic peptides derived from natural rhomboid substrates *P. stuartii* TatA and *Drosophilla* Gurken were cleaved by bacterial rhomboid proteases AarA and GlpG [103]. Furthermore, LacYTM2-derived synthetic protein was also found to be cleaved by *E. coli* GlpG [104]. From these observations, it has been concluded that certain recognition sequences are essential for the proteolysis and must have been conserved in the rhomboid substrates. Experiments based on x-ray crystallography and mutagenesis revealed some sequence determinants for rhomboid substrates, including small P1/P1' residues and large, hydrophobic chains at P4/ P2' positions [105-107].

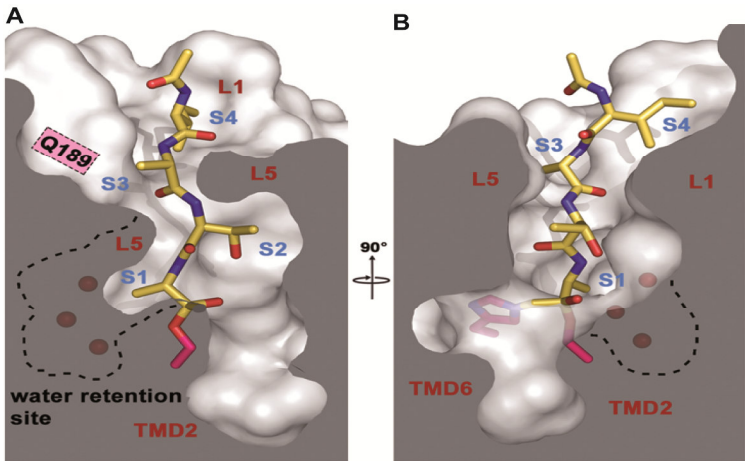


Figure 1.15: Co-crystallized structure of GlpG in complex with the Ac-IATA-cmk inhibitor (PDB: 4QO2). A) and B) The structure indicates the location of the S1-S4 subsites in GlpG with a view from different angles and the water retention site, which is in continuation with S1 subsite. Surface colouring according to the hydrophobic scale of Kyte and Doolittle. The inhibitor is shown in yellow and water molecules are presented as red spheres. Figure adopted from [91].

However, particularly the GlpG-complex structures with small mechanism based inhibitors, which provided important insights into the proteolysis mechanism, were of limited use to study substrate specificity as these inhibitors are relatively small and structurally very different from peptidic substrate. The goal is to solve high-resolution structures of rhomboid in complex with natural protein substrates, but so far nobody has succeeded. One major hurdle for the generation of such enzyme-substrate complex structures is likely that rhomboids possess relatively low affinity for their substrates [96]. To overcome this issue, Zoll *et al*. designed a substrate derived peptidyl-chloromethyl ketone (CMKs) [91], based on previous findings that tosyllysine chloromethyl ketone hydrochloride (TLCK) and tosyl phenylalanyl chloromethyl ketone (TPCK) were weak inhibitors of rhomboids [40, 102]. They first synthesized tetrapeptidyl-CMK Ac-Ile-Ala-Ala-$COCH_2Cl$ (or Ac-IAAA-cmk), which was derived from the bacterial rhomboid substrate TatA [91, 106]. Subsequently, scanning mutagenesis was used to systematically alter the P1-P5 positions of a TatA model substrate, and to correlate the efficacy of proteolytic cleavage of these model substrates with the inhibitor potency of CMK peptides harboring the same amino acid exchanges in the P1-P4 positions. This showed that, for both the proteolytic cleavage of the model TatA substrates and for the potency of the CMK inhibitors, the P1 and P4 positions were the most critical. P1 prefered small amino acid chains such as alanine and P4 prefered hydrophobic residues, consistent with previous results [105-107]. Another important observation was that P2, P3 and P5 were the least restrictive positions and that P2 could accept all amino acids. Finally, the co-crystal structures of GlpG with several peptidyl-CMKs inhibitors were determined [91]. These co-crystal structures defined the location of the S1-S4 subsites in rhomboid for a peptide mimicing TatA and confirmed the important role of the S1 and S4 subsites for substrate recognition and proteolysis (Figure 1.15).

Intriguingly, Urban and Moin have recently questioned whether the primary sequence of rhomboid substrates is at all relevant for substrate selection and cleavage site specificity [99]. They studied the dynamics of TMDs using circular dichorism and mutagenesis and concluded that the key feature of rhomboid substrates is that they have transmembrane α -helices, which receive their stability only through the hydrophobic environment of the membrane and which contain helix-destabilizing amino acid residues such as glycine. These permissive transmembrane α-helices are unstable in the hydrophilic environment of the rhomboid active site so that the catalytic residues can attack the peptide bond and the substrate is proteolytically cleaved (Figure 1.16). Importantly, it was demonstrated that, by introducing helix-destabilizing amino acid residues into the TMD, non-substrates could be converted into rhomboid substrates. For example, while APP is not cleaved by rhomboid in cellular and *in vitro* assays, the introduction of a single proline residue into its TMD is sufficent to transform APP into a highly permissible rhomboid substrate [99].

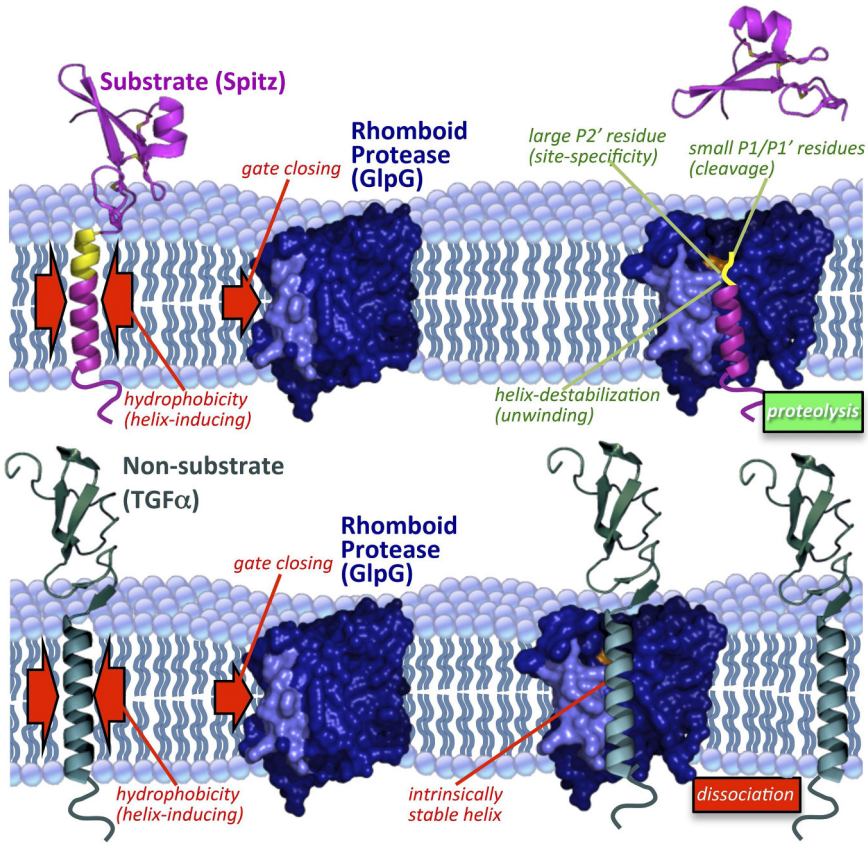


Figure 1.16: Model of substrate recognition and proteolysis by rhomboids. The lipid bilayer is a hydrophobic environment with stabilizing effects on transmembrane α-helices of substrate proteins. α-Helical TMDs are poor substrates for proteases. When a substrate passes through the opening of the rhomboid into the hydrophilic environment, the substrate partially unfolds and binds to the catalytic residues in the active site (orange), with subsequent proteolysis of the scissle bond. In contrast, interaction of the rhomboid with a non-substrate protein (below) with a stable α-helical TMD does not permit proteolysis and the non-substrate dissociates from the rhomboid. Figure adopted from [99].

1.3 Inhibition of serine proteases

Soluble and intramembrane serine proteases play vital roles in various biological processess such as extracellular matrix (ECM) remodelling, angiogenesis, wound healing, tumor invasion, parasitic invasion, and bacterial quorum sensing [36, 108]. Proteolytic dysfunction can cause a variety of diseases (section 1.1), and serine proteases in particular are associated with inflammation, emphysema, chronic bronchitis, arthritis, hypertension and clotting abnormalities [36, 108]. Additionally, serine proteases have been implicated in bacterial, viral and parasitic infections [36, 108]. Hence, serine protease inhibitors could not only be beneficial to control a variety of pathogenic diseases but are also useful tools to elucidate the precise role of serine proteases in a multitude of biological processess. Based on design patterns, serine protease inhibitors have been classified into two categories: peptidic inhibitors and non-peptidic inhibitors.

1.3.1 Peptidic serine proteases inhibitors

Peptide-based inhibitors are small protein-like compounds, whose amino acid residues mimic a natural peptide with tailored molecular properties such as enhanced stability or biological activity. They are also named peptidomimetic inhibitors, and peptides are generally a rational starting point for designing

protein inhibitors as they have the ability to interact with the enzyme similar to their natural substrates [109]. Other major advantage of peptide-based inhibitors is their high selectivity towards the enzyme, since these conformationally restrained structures hold minimal chances for non-target binding. In general, depending on the binding mode, peptide-derived inhibitors of serine proteases can be covalent or non-covalent binders. Covalent peptidomimetic inhibitors consist of two moieties, a serine trap and a recognition motif (Figure 1.17) [110]. The serine trap is a reactive motif attached C-terminal to the scissile bond, which normally reacts with the active site serine residue of the enzyme resulting in covalent bond formation.

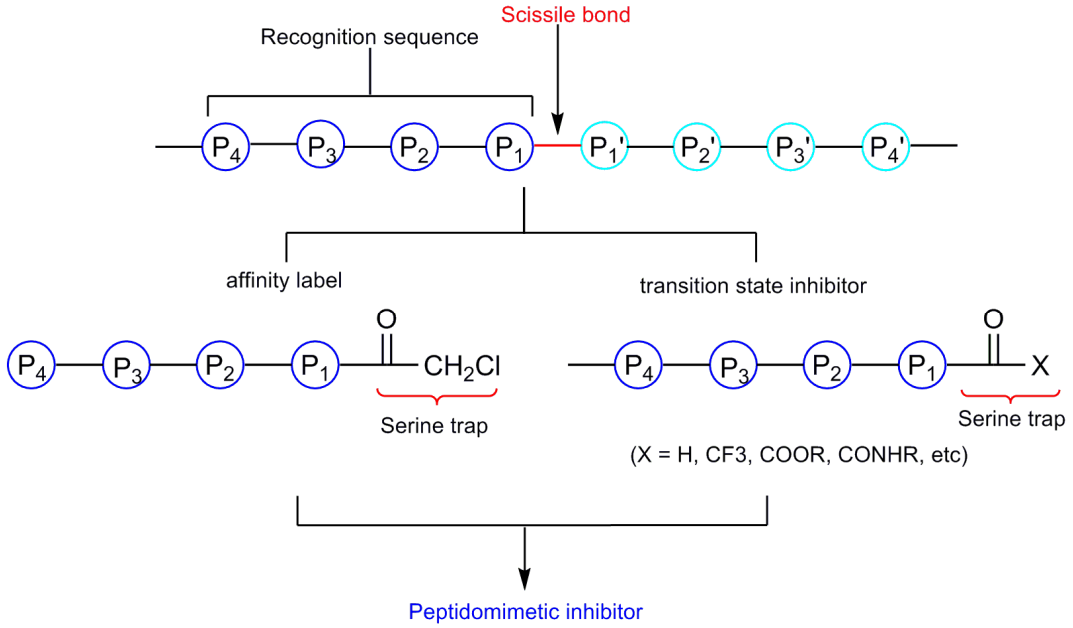


Figure 1.17: Schematic representation of peptidomimetic covalent inhibitors. The peptide chain is the recognition sequence and the serine trap is the reactive motif, adopted and modified from [110].

The functional groups used as serine traps can be customized, for example chloromethyl ketone (cmk, $-COCH_2Cl$) as an affinity label, aldehydes(-CHO), trifluoromethylketone ($-COCF_3$), or boronic acids ($-B(OH)_2$) as transition state labels (Figure 1.17). The recognition motif is an amino acid sequence derived from the natural substrate of the enzyme or or one that mimics the original sequence of the substrate closely and results in the same biological effects (Figure 1.17). Besides the reactive motifs mentioned above, medicinal chemists have also introduced other novel functionalities such as α-keto heterocycles. These incorporate C-terminal electron-withdrawing groups and have been designed to result in enhanced susceptibilty towards the nucleophilic serine residue. Recently, a ketobenzothiazole-containing peptidomimetic compound (Figure 1.18) was presented as a very potent matriptase inhibitor with high selectivity [111], matriptase is a type II membrane-bound serine protease (but not a transmembrane protein). These compounds were covalent and reversible inhibitors, in which Arg-Gln-Ala-Arg (RQAR) mimiced the substrate sequence and acted as a recognition sequence for matriptase and the ketobenzothiazole moiety contributed the serine trap (Figure 1.18). In the present study, ketobenzothiazole-derived peptidomimetic inhibitors were used as a starting point for designing peptide-based rhomboid protease inhibitors (section 4.5.1).

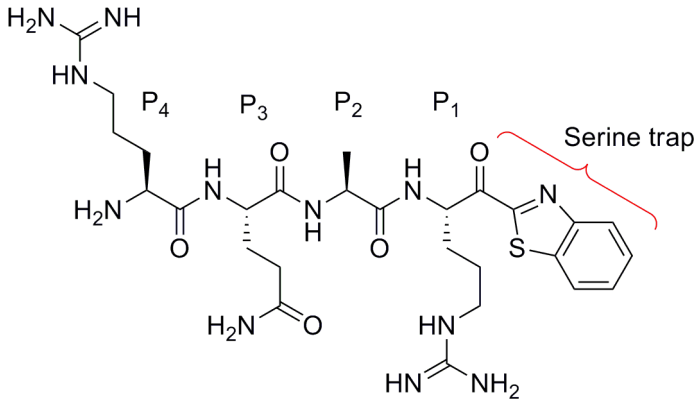


Figure 1.18: Chemical structure of ketobenzothiazole-containing peptidomimetic inhibitors of matriptase. Arg-Gln-Ala-Arg (P4-P1) represents the recognition sequence (black) and ketobenzothiazole functions as serine trap (red). Adopted from [111].

Non-covalent peptidomimetic inhibitors are also available and consist of a non-reactive terminal fragment such as a 3-sulfonylaminopyridinone group attached to a peptidyl chain [112]. In contrast to a serine trap, this fragment must be chemically inert towards the nucleophilic serine but it should efficiently contribute non-covalent binding interactions such as H-bonds, hydrophobic bonds, and pi-pi interatctions. Besides their high selectivity, peptidomimetic inhibitor also possess significant pharmacological drawbacks such as limited proteolytic stability, bioavailability, solubility and toxic effects. Hence, through the design of nonpeptidic inhibitors, medicinal chemist have attempted to overcome the limitations of peptidomimetic inihibtors.

1.3.2 Non-peptidic serine proteases inhibitors

Based on their molecular template as well as their mechanism of inhibition, non-peptidic serine protease inhibitors have been categorised into three major classes: Alkylating agents, enzyme-activated inhibitors, and alternate-substrate inhibitors.

Alkylating agents

Alkylating agents are mechanistically the simplest class of serine protease inhibitors, which lead to permanent deactivation of the enzyme. The inhibition mechanism involves the attack of the nucleophilic serine residue on an electron-deficient atom of the chemical agent with the release of a leaving group (Figure 1.19). Some of most commonly used alkylating agents are organophosphorus and organosulfur based compounds (Figure 1.19). Dialkylfluorophosphates and sulfonyl fluorides are known to be selective for serine proteases [113], and diisopropylfluorophosphate (DFP) and p-Toluenesulfonylfluoride have been extensively used as diagnostic agents for serine proteases. Generally, phosphoryl fluoride inhibitors are known to be highly reactive and hydrolytically unstable, while sulfonyl fluoride inhibitors are less reactive and hydrolytically stable. However, modifications of the scaffold of these inhibitors have lead to the development of inhibitor classes with improved enzyme selectivity and hydrolytic stability [113].

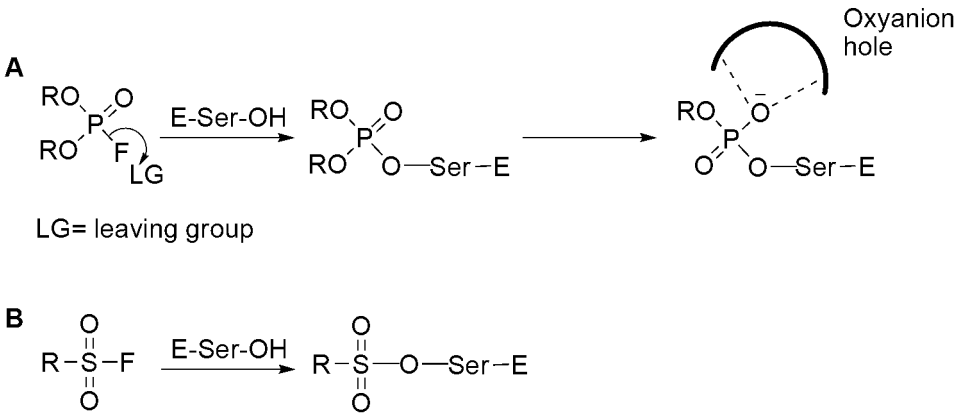


Figure 1.19: Examples of alkylating agents. a) fluorophosphate compounds b) sulfonylfluoride compounds. Adopted from [113].

Enzyme-activated inhibitors

Enzyme-activated inhibitors are inherently unreactive compounds that contain a latent reactive functionality (electrophilic species), which is unmasked during the catalytic cycle of the enzyme. The corresponding reaction involves the recognition of the inhibitor by the enzyme as a natural substrate and leads to enzyme inactivation by covalent bond formation with the nucleophilic serine residue of the catalytic site [114] (Figure 1.20). Since the reactive species are generated within the active site of the target enzyme, a high degree of selectivity could principally be achieved with this class of inhibitors. A large number of templates have been employed to generate enzyme-activated serine protease inhibitors and some examples, isocoumarin, beta-lactam, phthalimide, and succinimide with their inhibition mechanisms are summarised in Figure 1.20 [108]. The general mechanism of inhibition involves initially the formation of a covalent bond between an electrophilic atom in the heterocycle and the nucleophilic serine residue, resulting in a mono-covalent adduct. This is followed by a rearrangement reaction within the heterocyclic ring and unmasking of another electrophilic center, which is attacked by the catalytic histidine residue in the active site of serine proteases. This results into formation of a second covalent bond and stable and irrversible inhibition of the serine protease. These compounds are also called mechanism-based inhibitors or suicide inhibitors because the enzyme itself triggers the inhibitory reaction by exposing a latent functionality in the compound leading to irreversible inhibition [108, 110, 113].

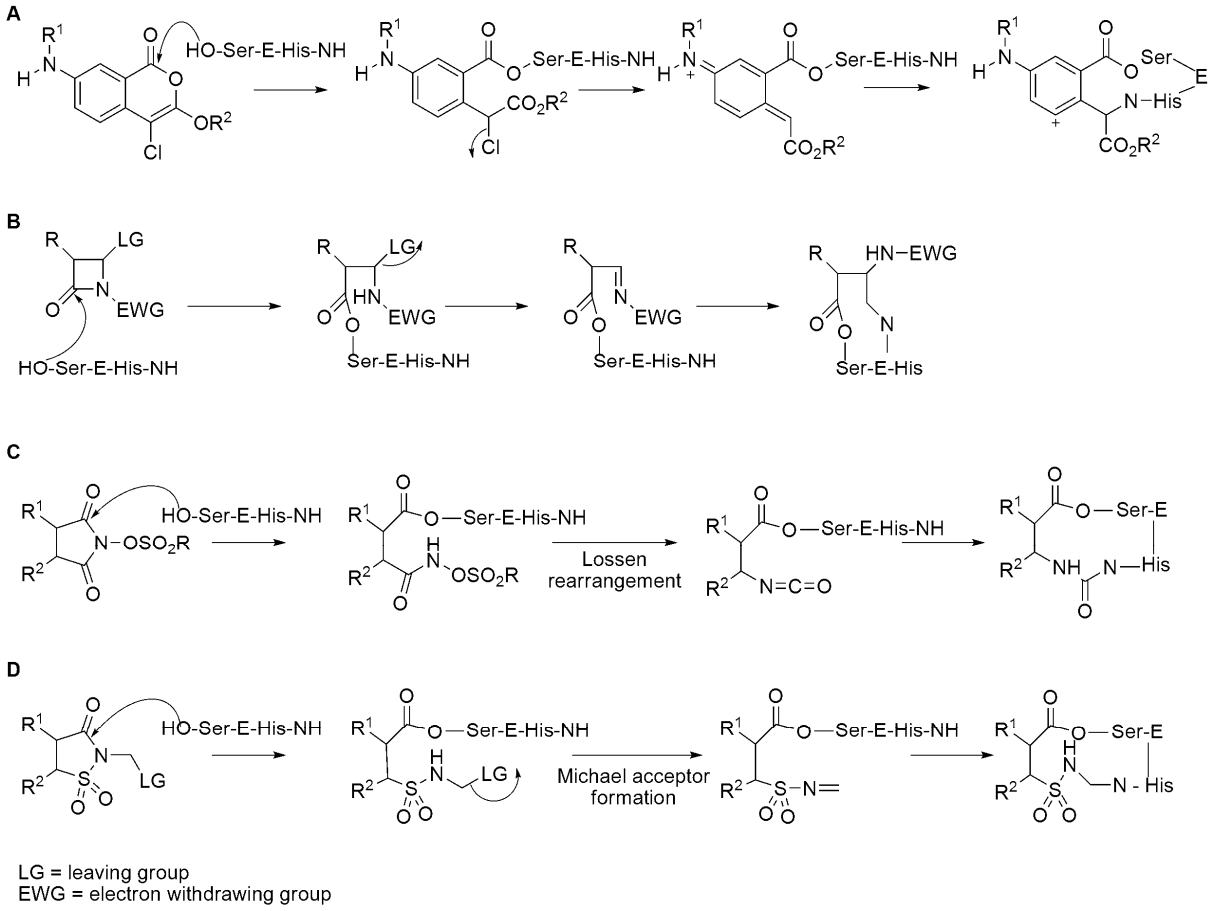


Figure 1.20: Examples of enzyme-activated inhibitors and their inhibition mechanisms. A) 7-substituted isocoumarin template B) b-lactam template C) phthalimide template D) saccharin template. Adopted from [108]

Alternate-substrate inhibitors

Alternate-substrate inhibitors are also known as acylating agents, which inhibit the enzyme by formation of a covalent and stable acyl enzyme [110]. However, unlike enzyme-activated inhibitors they do not bind irrversibly because they lack a second electrophilic species, which can react with the active site histidine residue. Some of the acylating agent templates for inhibiting serine proteases are: pyrolidine trans-lactams and lactones, 4H-3,1-benzoxazin-4-one, 1,3-diazetidine-2,4-dione and their variants [110]. The general mechanism of inhibition of the 4H-3,1-benzoxazin-4-one template and the reaction products are summarised in Figure 1.21. Benzooxazinones inhibit serine proteases by the formation of a stable acyl enzyme (Figure 1.21). The enzyme selectivity and the stability of the acyl enzyme depend on the substitution of the aromatic ring and the C-2 position (R_2 substituent). Indeed, the acyl enzyme is slowly deacylated. However, the parent inhibitor is not generated again but turned over as a substrate, which explains the name alternate-substrate inhibitors (Figure 1.21).

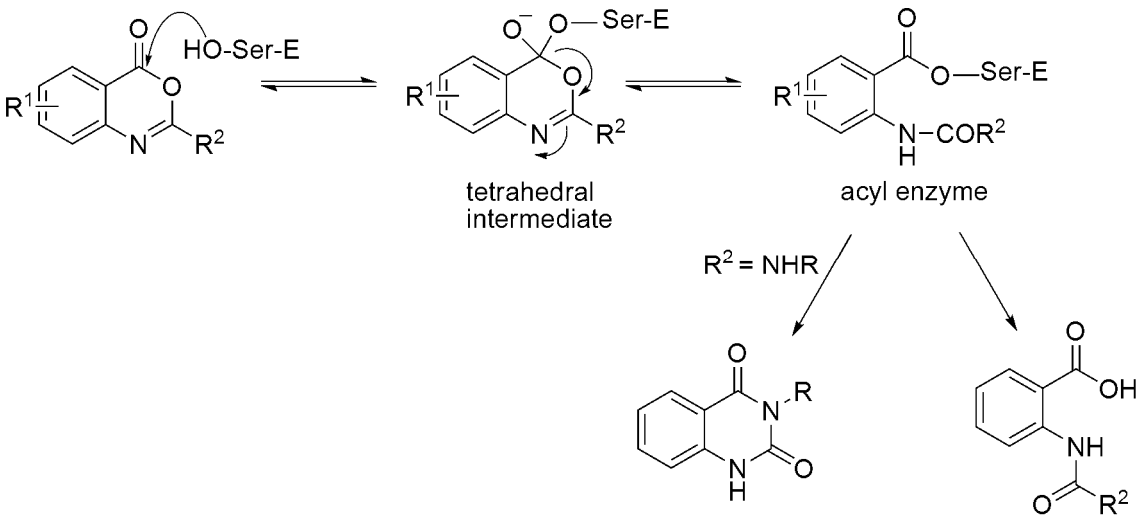


Figure 1.21: Mechanism of inhibition of benzoxazinone derivatives. Adopted from [110].

1.4 Inhibitor Design

Inhibitor design (or more broadly drug design) involves the design of a molecule that can bind tightly to its biological target such as enzymatic proteins. Enzyme inhibitors can have therapeutic potential and should posess specific properties such as adequate potency, bioavailability, hydrolytic and metabolic stability. These properties can be achieved by considering the Lipinski 'rule of five' for drug design [115, 116]. Inhibitors are generally small organic molecules, which have the ability to interact with protein molecules and fit well into a molecular cleft. The search for inhibitors of a specific target can be achieved by either a rational or a computer-aided approach (Figure 1.22).

The rational approach comprises the screening of potential drug like candidates or big chemical libraries by performing small scale or high-throughput (HTS) screening assays *in vitro*. For instance, previous findings of rhomboid inhibitors such as isocoumarins and β-lactams resulted from rational approaches [87, 102, 103]. β-lactams were discovered as rhomboid inhibitors through a HTS screen of ≈ 60,000 compounds including two small libraries of protease and kinase inhibitors and a collection of 2,000 natural compounds [103]. In this case, the compound libraries were screened using a fluorescence-based *in vitro* assay of rhomboid activity [103]. If sucessful, the initial screening results in a lead-like compound, which can be further optimized based on structure activity relationship (SAR) studies (Figure 1.22), eventually leading to the generation of potent and selective enzyme inhibitors. Today, in addition to HTS screening, drug design frequently relies on computational techniques for virtually screening of suitable candidate molecules prior to any *in vitro* or *in vivo* activity assays and this method of inhibitor design is called computer-aided approach [117]. Figure 1.22 shows the basic steps in the computer-aided approach. First, suitable candidate molecules or online available molecular databases are selected. Subsequently, these molecular libraries are virtually screened based on molecular modeling experiments using suitable docking software programs such as molecular operating environment (MOE), Gold, or Autodock. The modeling experiments not only

provide information about binding affinities of the ligand to the receptor but also predict conformational changes. Based on the modeling experiments, the best suited ligands are selected as 'hits', followed by their biological evaluation. Finally, positive hits or lead compounds are further optimized considering SAR.

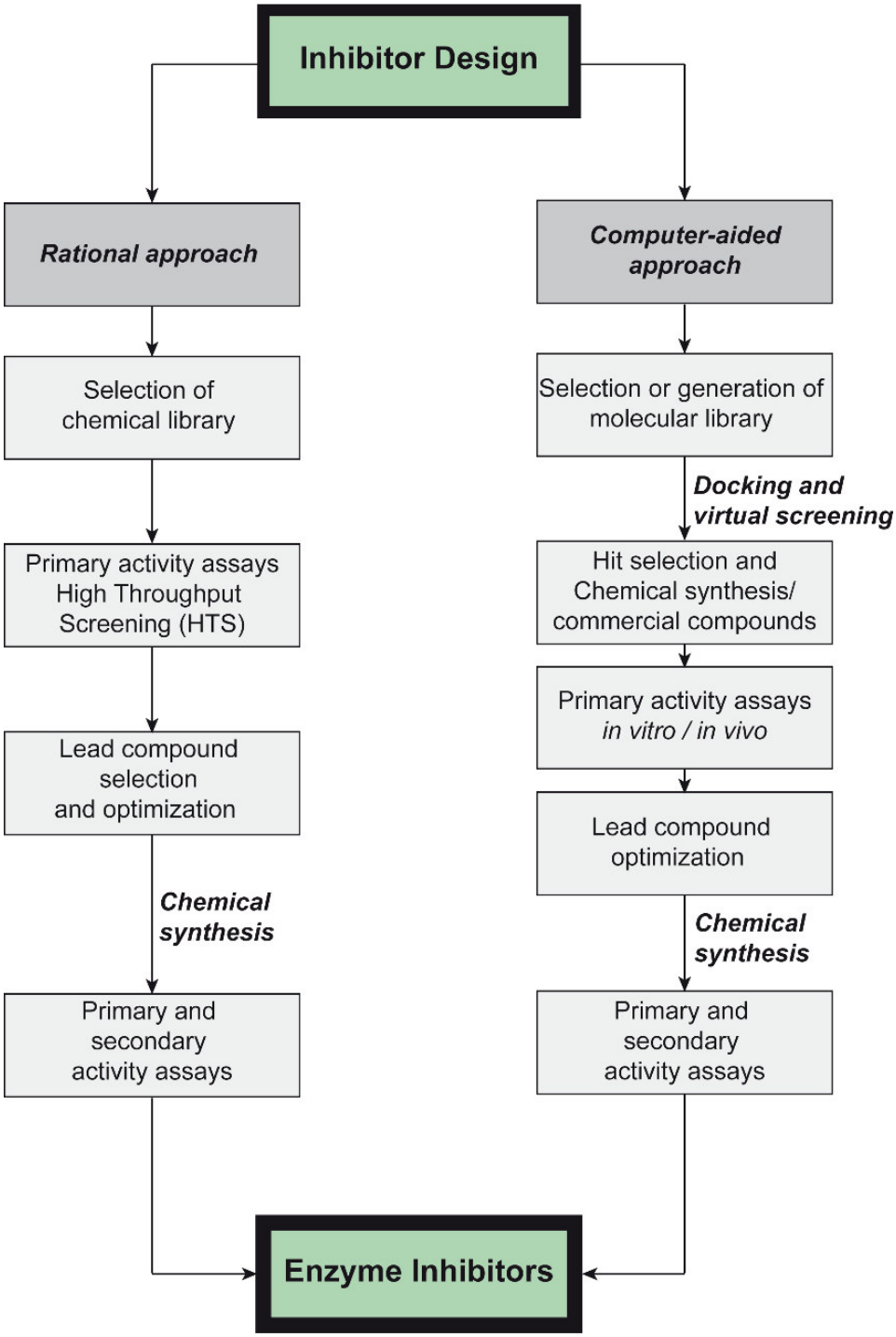


Figure 1.22: Schematic representation of two approaches of inhibitor designing.

Computer aided virtual screening can be classified into two types: structure-based screening and ligand-based screening. Structure-based virtual screening entails the docking of a number of ligands

into a target protein followed by their ranking based on ligand efficiencies, scoring function and ligand-protein interaction (section 3.4) [118, 119]. In contrast, ligand-based virtual screening utilizes the structural information from a collection of structurally diverse molecules that can bind the receptor in order to generate a pharmacophore (ph4) model. Based on this ph4 model, candidate ligands can be analysed for their binding capabilities. [120, 121]. In summary, advances in computational techniques have enabled new *in silico* methods for virtual screening of candidate molecules that greatly fasten the process of lead identification and lead optimization[122].

1.5 Rhomboid protease inhibitors

Early studies revealed that rhomboids are not susceptible to most of the broad-spectrum serine protease inhibitors [40, 102]. However, continuous efforts have led to the discovery of a few inhibitor classes for rhomboid proteases, which are shown in Figure 1.23. All of them exhibit irreversible-covalent inhibition mechanism (Figure 1.23), with a nuclophilic attack of the active site serine residue on the electrophilic site of the inhibitor. The potency of these covalent binders as indicated by their IC50 values which dependent on the preincubation time of enzyme and inhibitor prior to the addition of the enzyme substrate, and dependent on the read-out assay. Hence, in this section, while comparing the potency of different rhomboid inhibitors, preincubation timed and read-out assays have also been taken into consideration.

The first rhomboid inhibitor discovered was 3,4-dichloroisocoumarin (DCI) (**1**), which initially was found to inhibit DmRho-1 [40] but was later established as a 'pan inhibitor' for rhomboids [123, 124]. Pan inhibitor means that DCI has the ability to inhibit various rhomboids from different organisms, including *E. coli* GlpG, *P. stuartii* AarA, *B. subtilis* YqgP, *T. gondii* ROM5, *A. Aeolicus* rhomboid, *H. sapiens* PARL and RHBDL2. However, the potency of DCI for the different rhomboids is variable and rather low [123]. Isocoumarins are very well known classical inhibitors for serine proteases and, consequently, DCI lacked selectivity for rhomboids. In order to study the inhibition mechanism, co-crystallization experiments were attempted with *E. coli* GlpG (EcGlpG) and DCI. However, these were unsuccesful due to the unstable covalent bond between the catalytic serine and the carbonyl group of DCI (Figure 1.23A). Subsequently, by further exploring the isocoumarin structure, two analogues carrying a 7-amino group, JLK-6 (**2**) and IC 16 (**3**) were discovered. These showed improved potency against GlpG with IC50 values of 6 µM and 0.7 µM with preincubation for 30 min [87, 88]. Interestingly, the 7-amino group carrying analogue JLK-6 displayed a different reaction mechanism compared to DCI, involving scaffold opening after the nucleophilic attack by the active site Ser201 and successive rearrangements resulting in formation of an additional electrophile. This further reacted with the active site His254 leading to a double-bonded enzyme-inhibitor complex (Figure 1.23B). The proposed mechanism of catalysis and the geometry of the binding site bound to the isocoumarin inhibitor were confirmed by structural experiments, gel-shift assays and X-ray

crystallography [87, 88]. The crystallographic analysis also proved the formation of the characteristic oxyanion hole, which is a key structural element of serine proteases and plays a major role in the rhomboid catalytic mechanism (section 1.2.3.2). Eventually, crystallization studies of the IC 16 inhibitor in complex with GlpG confirmed that this compound similar to JLK-6 formed an adduct with two covalent bonds. Surprisingly, these studies showed that unlike JLK-6 the second nucleophilic attack was by His150 instead of His254, which was previously only known to contribute to the oxyanion hole. However, these findings suggested that the role of His150 is not limited to the oxyanion hole.

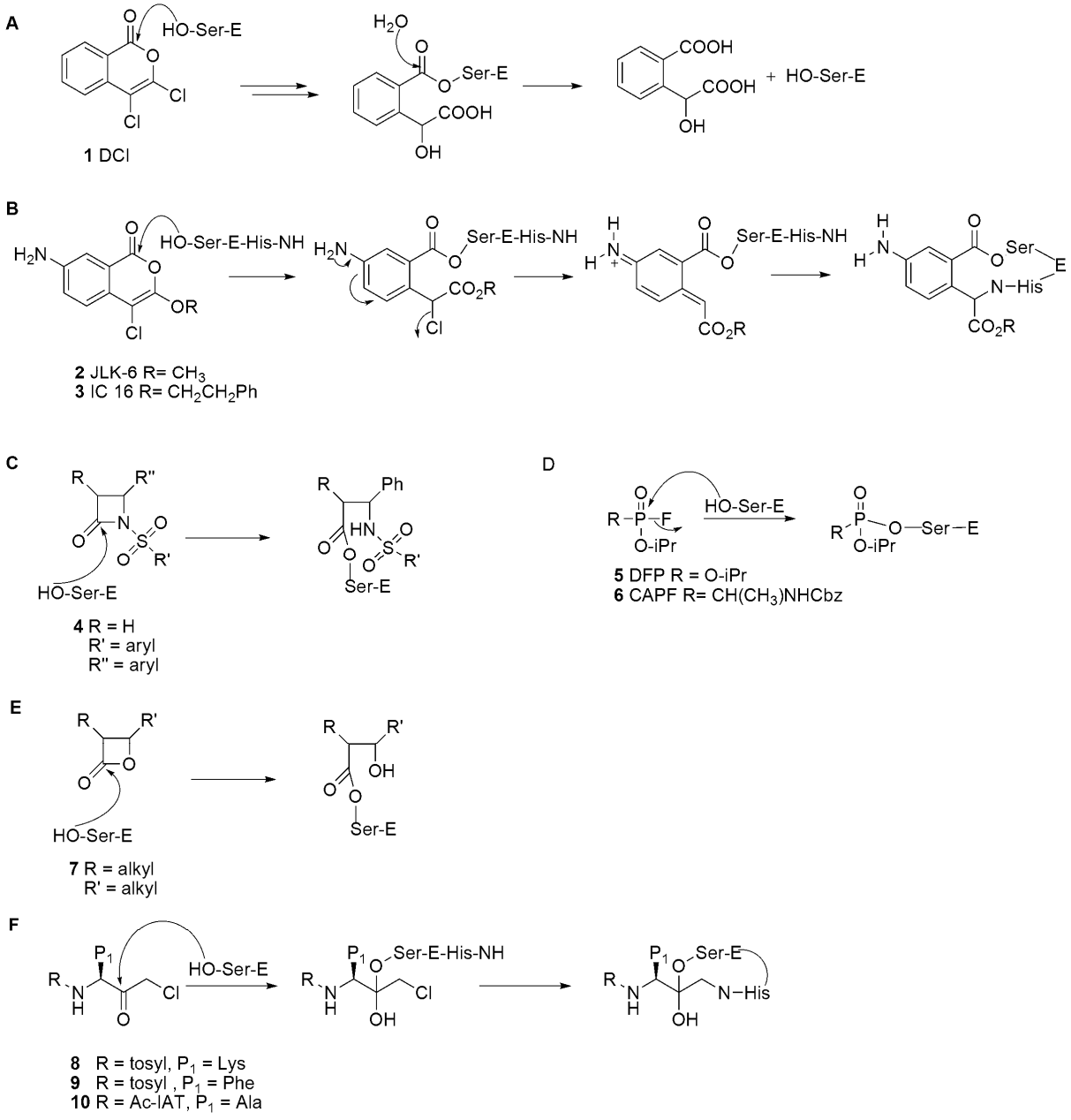


Figure 1.23: Structures and mechanisms of known classes of rhomboid protease inhibitors. A) DCI, B) 7-amino-4-chloro isocoumarins, JLK-6 and IC 16, C) β-lactams, D) Fluorophosphate, DFP and CAPF E) β-lactones, F) Chloromethyl ketones. Adopted and modified from [123].

In 2011, Pierrat *et al*. decribed a new class of rhomboid inhibitors, namely monocyclic β-lactams (Figure 1.23C). Their were discovered by HTS screening of ≈60,000 substances against rhomboid AarA using a fluorescence based assay [103]. β-lactams are also irreversible inhibitors of serine proteases and form a covalent bond between the indispensable β-lactam ring and the active site serine (Figure 1.23C) [86]. N-aryl sulfonyl substituted β-lactam (**4**) was the most potent inhibitor with an IC50 of 0.4 µM for GlpG (preincubation 30 min). Importantly, some of the derivatives were also able to inhibit the eukaryotic rhomboid RHBDL2 in a cell-based assay [103]. The advantages of β-lactams were that they have less cytoxicity compared to isocoumarins and that they are a class of drugs already used in pharmaceuticals making them potentially suitable candidates for further development. Despite that, β-lactams were shown to have poor selectivity over soluble serine protease such as chymotrypsin, which is also an important criteria for inhibitor development. In addition, Ha et. al discovered diisopropylfluorophosphate (DFP) (**5**) as an irreversible-covalent inhibitor of GlpG [90] (Figure 1.23D) Later, a peptide derived analogue of a phosphonofluoridate inhibitor, Cbz-Ala^{P}(O-iPr)F or CAPF (**6**) was also found [89] (Figure 1.23D). Like isocoumarins and β-lactams, they also formed a covalent bond with the active site Ser201 but both displayed only low potency against GlpG (IC50 ~50 µM, with preincubation 30 min). However, enzyme-inhibitor complex crystal structures with DFP and CAPF have provided insights into the catalytic mechanism of rhomboids [89, 90, 94] (section 1.2.3.2).

In 2013, Wolf *et al*. presented β-lactones (**7**) as new class of rhomboid inhibitors with IC50 values of 26-44 µM (preincubation for 30 min) [125]. Similar to β-lactams, the reaction mechanism of β-lactones also involves opening of the four-membered ring after the serine attack, resulting in covalent bond formation (Figure 1.23E) [124]. With respect to peptidic compounds, the commercially available chloro methyl ketones (CMKs) tosyllysine chloromethyl ketone hydrochloride (TLCK) (**8**) and tosyl phenylalanyl chloromethyl ketone (TPCK) (**9**) have been reported as weak inhibitors for YqgP and DmRho-1[40, 102]. The weak inhibition was explained by improper P1 residues (Lys or Phe) as the P1 position was found to be not suitable for larger amino acids [87, 106]. Based on these compounds, Zoll *et al*. subsequently reported the first peptido-mimetic inhibitor (**10**) for rhomboids [91]. Peptidyl chloromethyl ketone analogues were synthesised and the peptide chain residues were selected according to the bacterial rhomboid substrate TatA. X-ray crystallography studies and gel-based assays verified that they are also mechanism-based irreversible inhibitors, forming a double-bonded covalent adduct with the catalytic dyad and resulting in cross-linking of TM4 and TM6 [91]. The reaction mechanism involved the formation of a hemi-ketal after attack by the Ser201 and later alkylation of the active site His254 (Figure 1.23F). However, these peptidic compounds were also weak inhibitors of rhomboids with IC50 values of about 100 µM even with 180 min of preincubation time [91]. In conclusion, all of the above described classes of rhomboid inhibitors lack sufficient potency and selectivity, demonstrating a definite need for substantially improved compounds.

1.6 Objective

With respect to structure and catalytic mechanism, the rhomboids are the best understood family of intramembrane proteases [82]. Rhomboids are evolutionarily conserved and ubiquitously distributed in all types of life forms [43]. Because of their diverse biological functions including EGF signaling, mitochondrial integrity and wound healing, and their association with several diseases such as Parkinson's disease and type-2 diabetes, rhomboids might have considerable medical potential.

The functions of rhomboids in many organism are still unknown. For example, the plant *Arabidopsis* contains at least 13 genes encoding for active rhomboid proteases for which no physiological functions have been yet assigned [126]. Although functional genetics has played a major role in understanding rhomboids, this approach is complex and laborious because of the high redundancy of rhomboid genes in many organisms. Therefore, potent and specific chemical inhibitors would be extremely valuable tools to probe rhomboid functions. Moreover, inhibition studies would also provide further information with regard to substrate specificity and the catalytic mechanism of rhomboids. Surprisingly, rhomboids were shown to be insensitive to almost all classes of inhibitors for soluble serine proteases [5]. Some irrversible covalent binders of rhomboids have been discovered, for example isocoumarin and monocyclic β-lactam derivatives (section 1.5), but these mostly displayed low potency and poor selectivity [123]. Enzyme selectivity is of paramount importance in order to prevent toxic side-effects. For instance, β-lactams are known to be broad-spectrum irreversible inhibitors of several hydrolytic enzymes like the proteasome, which could result in adverse events. Consequently, the main focus of this thesis was to obtain more potent and selective rhomboid inhibitors, with the following specific aims:

- Design of novel small molecule inhibitors against rhomboids based on molecular modeling predictions.
- Chemical synthesis of molecular modeling hits.
- Screening of novel compounds for bioactivity against prokaryotic and/or eukaryotic rhomboids in cell-free and in cell-based assays.
- Selectivity analysis of new novel rhomboid inhibitors within the family of rhomboid proteases and over soluble serine proteases like chymotrypsin.
- Mechanistic studies of novel rhomboid inhibitors.

2 Material

2.1 Bacterial strains and Cell lines

2.1.1 Bacterial strains

Strain	Genotype	Reference
E. coli BL21 (DE3) C43	F^- ompT gal dcm $hsdS_B(r_B^- m_B^-)$(DE3)	DSMZ
E. coli DH5α	Cloning of plasmid DNA F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1	DSMZ

2.1.2 Cell lines

Name	Description	Reference
HEK293	Human embryonic kidney cells	ATCC

2.2 Vectors and Primers

2.2.1 Vectors

Name	Description	Reference
pET41b(+)-GlpG	T7-Promoter, T7-Transcriptionsstart, His6-Tag, MCS (*Psh*AI - *Xho*I), T7-Terminator, lacI, KanR	[106]
pMALp2E-Gurken	MBP-TMD-Trx-His6, Gurken-TMD-AAA28598, residues Gln239–Gln288, T7-Promoter	[106]
pMALp2E- LacY	MBP-TMD-Trx-His6, LacYTM2- AP_000995.1, TM2, residues His39–Lys74, T7-Promoter	[106]
pcDNA3-RHBDL2	RHBDL2-Sequence, HA-Tag, NeoR	[106]
pcDNA3-Gurken	Gurken-construct-Sequence, 3xFLAG-Tag, NeoR	[106]
pGEX-6P-1-AaROM	AaROM-construct, N-terminal-GST-tagged fusion protein	[102]

2.2.2 Primers

Name	Sequence 5' → 3'	Supplier
T7-promotor	TAATACGACTCACTATAGGG	Invitrogen

2.3 Antibodies

2.3.1 Primary antibodies

Name	Species	Type	Working dilutions	Supplier
anit-His-*probe* (H-15)	Rabbit	polyclonal	1:1000	Santa Cruz Biotechnology
anti-FLAG® BioM2	Mouse	monoclonal	1:1000	Sigma aldrich
anti-HA-*probe* (F-7)	Mouse	monoclonal	1:1000	Santa Cruz Biotechnology
anti-MBP	Mouse	monoclonal	1:1000	BioLabs

2.3.2 Secondary antibodies

Name	Species	Type	Working dilutions	Supplier
anti-Rabbit	goat	polyclonal	1:10000	LI-COR
anti-Mouse	goat	polyclonal	1:10000	LI-COR

2.4 Reagents

2.4.1 Chemicals

Name	Supplier
30% Acrylamid, 37.5:1, Bis-Acrylamide	Sigma-Aldrich, Steinheim
Agar-Agar	Roth, Karlsruhe
Agarose peqGOLD *low melt*	PEQLAB, Erlangen
AlamarBlue	Invitrogen, Karlsruhe
Ammoniumpersulfat (APS)	Sigma-Aldrich, Steinheim
Bicine	Merck, Darmstadt
Bis-Tris	FLUKA, Steinheim
Boric acid	Sigma-Aldrich, Steinheim
Bromphenol Blue	Sigma-Aldrich, Steinheim

BSA	Thermo Scientific
Calciumchloride	Merck, Darmstadt
Coomassie Brillantblue	Merck, Darmstadt
Desoxynucleotide-tri-phosphate (dNTP)	New England Biolabs
Dimethylsulfoxid (DMSO)	Sigma-Aldrich, Steinheim
Dithiothreitol (DTT)	Sigma-Aldrich, Steinheim
DMEM	Invitrogen, Karlsruhe
Ethanol	Merck, Darmstadt
Ethylendiamintetraessigsäure (EDTA)	Amresco, USA
Fetal Calf Serum (FCS)	Gibco, USA
Gelatine (*cold water fish skin*)	Sigma-Aldrich, Steinheim
GeneJuice® Transfection Reagent	Merck, Darmstadt
Glacial acetic acid	Merck, Darmstadt
Glycerin	Roth, Karlsruhe
Hefeextrakt	Sigma-Aldrich, Steinheim
HEPES	Merck, Darmstadt
Hydrochloric acid (HCl)	Merck, Darmstadt
Imidazol	Sigma-Aldrich, Steinheim
Isopropanol	Roth, Karlsruhe
Isopropyl-β-D-thiogalactopyranosid (IPTG)	Invitrogen, Karlsruhe
Magnesiumchloride - Hexahydrat	Roth, Karlsruhe
MES	AppliChem, USA
Methanol	Roth, Karlsruhe
Midori Green Advance	NIPPON Genetics, Japan
Milchpulver	Roth, Karlsruhe
N,N,N',N'-Tetramethylethylendiamine (TEMED)	Roth, Karlsruhe
n-Dodecyl-β-Maltosid	Glykon, Luckenwalde
n-Nonyl-β-D-Glucopyranosid	Glykon, Luckenwalde
NP40	Sigma-Aldrich, Steinheim
Opti-MEM	Invitrogen, Karlsruhe

Poly-L-Lysin-Lösung	Sigma-Aldrich, Steinheim
Protease Inhibitor	Roche, Mannheim
Sodium azide	Merck, Darmstadt
Sodium chloride	Roth, Karlsruhe
Sodium dodecyl sulfate (SDS)	Roth, Karlsruhe
Sodium hydroxide	Merck, Darmstadt
Spdium pyruvate	Invitrogen, Karlsruhe
Tris-hydrochloride	Roth, Karlsruhe
Tris-base	Roth, Karlsruhe
Trypan blue	Merck, Darmstadt
Trypsin/EDTA	Invitrogen, Karlsruhe
Trypton/Pepton from Casein	Roth, Karlsruhe
Tween® 20	Roth, Karlsruhe

2.4.2 Antibiotics

Name	Supplier
Ampicillin	AppliChem, Darmstadt
Kanamycin	Sigma-Aldrich, Steinheim
Chloroamphenicol	Sigma-Aldrich, Steinheim
Streptomycin	Gibco, USA
Tetracycline	Sigma-Aldrich, Steinheim

2.4.3 Inhibitors

2.4.3.1 Reference inhibitors

Name	Reference
3,4-Dichloroisocoumarin	Calbiochem®, Darmstadt
JLK-6	BSc5091, BSc archive, Schmidt lab, TU Darmstadt
L16	Strisovsky's lab (Academy of Sciences of the Czech Republic, Prague)

2.4.3.2 keto-benzothiazole inhibitors

Name	Reference
BSc5123	This study, PG029, Schmidt lab, TU Darmstadt
BSc5124	This study, PG039, Schmidt lab, TU Darmstadt
BSc5125	This study, PG044, Schmidt lab, TU Darmstadt

2.4.3.3 oxathiazole-2-one inhibitors

Name	Reference
BSc5138	This study, PG058, Schmidt lab, TU Darmstadt
BSc5139	This study, PG055, Schmidt lab, TU Darmstadt
BSc5140	This study, PG070, Schmidt lab, TU Darmstadt

2.4.3.4 Rhodanine inhibitors

Name	Reference
BSc4764	BSc archive, Schmidt lab, TU Darmstadt
BSc4820	BSc archive, Schmidt lab, TU Darmstadt
BSc4821	BSc archive, Schmidt lab, TU Darmstadt
BSc4824	BSc archive, Schmidt lab, TU Darmstadt
BSc4825	BSc archive, Schmidt lab, TU Darmstadt
BSc4826	BSc archive, Schmidt lab, TU Darmstadt
BSc4828	BSc archive, Schmidt lab, TU Darmstadt
BSc4882	BSc archive, Schmidt lab, TU Darmstadt
BSc4884	BSc archive, Schmidt lab, TU Darmstadt
BSc4885	BSc archive, Schmidt lab, TU Darmstadt
BSc4886	BSc archive, Schmidt lab, TU Darmstadt
BSc4994	BSc archive, Schmidt lab, TU Darmstadt
BSc4445	BSc archive, Schmidt lab, TU Darmstadt

2.4.3.5 Saccharin inhibitors

Name	Reference
BSc5152	This study, PG087, Schmidt lab, TU Darmstadt
BSc5153	This study, PG102, Schmidt lab, TU Darmstadt
BSc5154	This study, PG090, Schmidt lab, TU Darmstadt
BSc5155	This study, PG095, Schmidt lab, TU Darmstadt
BSc5156	This study, PG099/PG168, Schmidt lab, TU Darmstadt
BSc5157	This study, PG101, Schmidt lab, TU Darmstadt
BSc5158	This study, PG089, Schmidt lab, TU Darmstadt
BSc5159	This study, PG091, Schmidt lab, TU Darmstadt
BSc5160	This study, PG093, Schmidt lab, TU Darmstadt
BSc5161	This study, PG094/PG108, Schmidt lab, TU Darmstadt
BSc5187	This study,PG127, Schmidt lab, TU Darmstadt
BSc5188	This study, PG128, Schmidt lab, TU Darmstadt
BSc5189	This study, PG129, Schmidt lab, TU Darmstadt
BSc5190	This study, PG130, Schmidt lab, TU Darmstadt
BSc5191	This study, PG131, Schmidt lab, TU Darmstadt
BSc5192	This study, PG132, Schmidt lab, TU Darmstadt
BSc5193	This study, PG134, Schmidt lab, TU Darmstadt
BSc5194	This study, PG135, Schmidt lab, TU Darmstadt
BSc5195	This study, PG136, Schmidt lab, TU Darmstadt
BSc5196	This study, PG137, Schmidt lab, TU Darmstadt
BSc5197	This study, PG138, Schmidt lab, TU Darmstadt
BSc5198	This study, PG139, Schmidt lab, TU Darmstadt
BSc5199	This study, PG141, Schmidt lab, TU Darmstadt
BSc5200	This study, PG143, Schmidt lab, TU Darmstadt
BSc5201	This study, PG148/PG175, Schmidt lab, TU Darmstadt
BSc5202	This study, PG144, Schmidt lab, TU Darmstadt
BSc5203	This study, PG149, Schmidt lab, TU Darmstadt

BSc5204	This study, PG159, Schmidt lab, TU Darmstadt
BSc5205	This study, PG169, Schmidt lab, TU Darmstadt
BSc5206	This study, PG171, Schmidt lab, TU Darmstadt

2.4.3.6 Benzooxazinones inhibitors

Name	Reference
BSc5170	This study, PG112, Schmidt lab, TU Darmstadt
BSc5171	This study, PG113, Schmidt lab, TU Darmstadt
BSc5172	This study, PG114, Schmidt lab, TU Darmstadt
BSc5173	This study, PG115, Schmidt lab, TU Darmstadt
BSc5174	This study, PG117, Schmidt lab, TU Darmstadt
BSc5175	This study, PG118, Schmidt lab, TU Darmstadt
BSc5176	This study, PG121, Schmidt lab, TU Darmstadt
BSc5177	This study, PG122, Schmidt lab, TU Darmstadt
BSc5178	This study, PG123, Schmidt lab, TU Darmstadt
BSc5179	This study, PG124, Schmidt lab, TU Darmstadt
BSc5207	This study, PG152, Schmidt lab, TU Darmstadt
BSc5208	This study, PG161, Schmidt lab, TU Darmstadt
BSc5209	This study, PG174, Schmidt lab, TU Darmstadt

2.4.3.7 Virtual screening inhibitors

Name	Supplier
ZINC02809614	AKOS000714802, AKos GmbH, Steinen
ZINC33751351	AKOS003268417, AKos GmbH, Steinen
ZINC05442813	AKOS004937630, AKos GmbH, Steinen
ZINC006000122	AKOS022133268, AKos GmbH, Steinen
ZINC02573666	Sigma-aldrich

2.4.4 Size Standards

Name	Supplier
PageRuler™ Prestained Protein Ladder	Thermo Scientific, USA
Quick-Load® 2-Log DNA Ladder	New England Biolabs, Frankfurt

2.4.5 Enzymes

Name	Supplier
Dnase I (4056,6 U/mg)	AppliChem, Darmstadt
Lysozym (25674 U/mg)	AppliChem, Darmstadt
α-Chymotrypsin (1675,4 U/mg)	AppliChem, Darmstadt

2.4.6 Kits

Name	Supplier
Gel Extraction Kit	Qiagen, Hilden
Genopure Plasmid Maxi Kit	Roche, Schweiz
Pierce BCA Protein Assay	Thermo Scientific, USA
Plasmid Mini I	Omega Bio-Tek, USA
Silver staining kit	Invitrogen

2.5 Laboratory hardware and appliances

Name	Supplier
Analytical balance	Denver Instrument, Göttingen
Centrifuges	Eppendorf, Hamburg
	Hettich, Tuttingen
Electrophoresis voltage	Consort, Belgium
Flat bed shaker	Heidolph, Kelheim
Freezers and fridges	4°C Liebherr, Schweiz
	-20°C Liebherr, Schweiz
	-80°C Heraeus, Hanau
Glassware	Schott, Mainz

Heating block	HLC BioTech, Bovenden
Infrared detection system	LI-COR, Bad Homburg
Incubator	Binder, Tuttingen
Laminar flow	Gelaire, Australien
Light optical microscope	Olympus, Hamburg
Magnetic stirrer	Heidolph, Kelheim
Microplate Readers	Beckman Coulter, Krefeld
Multichannel pipette	Brand,UK
Nitrogen cavitation bomb	Parr instrument Co., USA
Pasteur pipettes	Roth, Karlsruhe
pH meter	inoLab, Weinheim
Pipettes 0,1 - 1 ml	Gilson, USA
Pipetten Accu Jet	Brand, UK
Platform shakers	Heidolph, Kelheim
Scales	Kern, Staufen
Semidry-Blot system	Hoefer, USA
Spectrophotometer	PEQLAB, Erlangen
Tank blotter	C.B.S. Scientific
Thermoblock	Eppendorf, Hamburg
T3 Thermocycler	Biometra, Göttingen
Vortex	IKA, Staufen

2.6 Consumables

Name	**Supplier**
0.2 /1.5 /2 mL reaction tubes	Eppendorf, Hamburg
10 cm Petri dishes (cell culture)	Thermo Scientific
6 /12-well plate	Thermo Scientific
15 /50 mL centrifuge tubes	Thermo Scientific
Aluminum foil	Aro, Metro
Disposable gloves	Micro-Touch® Nitra-Tex®

Gel casettes	Invitrogen, Karlsruhe
Immobilion-P Transfer membrane	Immobilion®-FL
Pipet tips	Starlab, Ahrensburg
Pipettes (2 mL – 25 mL)	Thermo Scientific
Whatman gel blot GB003	Sigma-Aldrich, Steinheim

2.7 Softwares

Adobe Design standard CS3
ChemBioDraw Ultra 14.0
Chimera
CLC workbench
Graphpad Prism 5
Image J
Image Studio Software (LI-COR)
MestReNova
Molecular Operating Environment (MOE) 2013 & 2014
Microsoft office 2010
Serial Cloner

3 Methods

3.1 Molecular Biology

3.1.1 Plasmid isolation

Isolation of plasmid DNA from bacteria is performed by a combination of alkaline lysis and ion-exchange columns, which is usually basis for all DNA extraction kits. In the present study, the Genopure Plasmid Maxi kit was used for the large scale plasmid extraction from 250 ml bacterial culture and Plasmid mini I kit was used for small scale DNA extraction from 5 ml of bacterial culture. All the necessary buffers and solutions as well as protocols were provided by the manufacturer and by following them plasmids were isolated and resuspended in an appropriate amount of RNase-free ddH_2O to obtain adequate concentrations. The concentration of the isolated DNA was measured with the NanoDrop device.

3.1.2 Preparation of chemically competent *E. coli* cells

1. Prepare inoue transformation buffer:

- 0.5 M PIPES (pH 6.7), piperazine-1,2-bis[2-ethanesulfonic acid] – Stock:
  To 15.1 g of PIPES 100 mL ddH_2O was added (pH adjusted with 5 M KOH) and the solution was sterilized by filtration through a pre-rinsed sterile filter (0.45 µM) and stored at -20°C.
- 55 mM $MnCl_2 \bullet 4H_2O$ (10.88 g/l)
  15 mM $CaCl2 \bullet 2H_2O$ (2.2 g/l)
  250 mM KCl (18.65 g/l)
  10 mM PIPES (20 ml stock)
  Add to 1 l ddH_2O
- Sterilized by filtration through a pre-rinsed sterile filter (0.45 µM) and stored at -20°C.

2. Prepare SOB medium:

- Mix the following in 900 ml ddH_2O:
  - 20 g Trypton
  - 5 g Hefe extract
  - 0.5 g NaCl
  - 2.5 ml 1M KCl
- pH was adjusted to 7.0 with 10 M NaOH (~100 µl) and ddH_2O was added to 990 ml.
- The resulting solution was sterilized by autoclaving and stored at room temperature.
- Before use, added 10 ml sterilized 1M of $MgCl_2$.

1. A single colony of *E. coli* was picked from a fresh agar-plate incubated for 16 h at 37°C and inoculated into 25 ml SOB medium in a 250 ml flask. Then the culture was incubated at 37°C with vigorous aeration for 6 - 8 h (250 - 300 rpm in a rotary shaker) until 0.4 OD_{600} ~ 10^8 bacteria/ml was reached and was subsequently used as starter-culture.

2. Then three 1 l flasks, each containing 250 ml of SOB were inoculated with 10, 4 and 2 ml starter-culture respectively. Cultures were incubated at 18-22°C for overnight with moderate shaking (180-200 rpm).

3. When OD_{600} of one of the cultures reached 0.6-0.8, the flask was rapidly transferred to an ice-water bath for 10-15 min. The remaining two cultures were discarded.

4. The culture was transferred to ice-cold centrifuge tubes and centrifuged at 4000 rpm for 10 min at 4°C. Supernatant was decanted and pellet was dried by keeping the tube inverted for 2 min.

5. Then the pellet was resuspended in 80 ml ice-cold Inoue transformation buffer and harvested by centrifugation at 4000 rpm for 10 min at 4°C.

6. Finally, the pellet was resuspended in 20 ml ice-cold Inoue transformation buffer and 1.5 ml DMSO were added. The resulting solution was stored on ice for 10 min.

7. By dispensing 50 µl of cell suspension, aliquots were prepared in eppendorf tubes, dropped into a bath of liquid nitrogen and stored at -80°C.

3.1.3 Transformation of *E.coli*

For transformation, 40 µl of chemically competent cells (*E. coli* DH5α) were allowed to thaw on ice and 20-100 ng of plasmid DNA was added. The resulting solution was mixed by gentle pipetting and incubated on ice for 30 min. Then the mixture was incubated in a heating block at 42°C for 90 sec, followed by addition of 500 µl LB medium and incubation on ice for 1 min. Then the sample was incubated for regeneration (37°C, 250 rpm, and 1 h). Subsequently, the transformed bacteria were streaked on LB/kan or LB/amp agar plates and incubated at 37°C overnight. Plates were transferred to 4°C when colonies became visible.

LB medium: 10 g/l Trypton/Pepton

5 g/l Yeast extract

5 g/l NaCl

Autoclaved immediately and stored at 4°C

LB-agar 10 g/l Trypton/Pepton

5 g/l Yeast extract

5 g/l NaCl

15 g/l agar

Autoclaved immediately, poured into plates, and stored at 4°C. To make LB - agar plates with antibiotics:

Antibiotic stock used:

Ampicillin	50 mg/ml
Kanamycin	50 mg/ml

Sterile filter and stored at -20°C.

50 µg/ml of Kanamycin or Ampicillin were added to autoclaved LB-agar after cooling down to 50°C and then poured into plates and stored at 4°C.

3.1.4 Sequencing of plasmids

All DNA constructs were sequenced at StarSEQ®GmbH (Mainz) and samples were prepared according to company instructions. In each case, 400-700 ng of plasmid DNA and 1 µl of a 10 µM primer solution were added to a PCR reaction tube and RNase free ddH_2O was added to a total volume of 7 µl.

3.2 Protein Biochemistry

3.2.1 Cryocultures

Escherichia coli (*E.coli*) cells were inoculated in LB/antibiotic medium and incubated overnight at 37°C and 250 rpm. 500 µl overnight culture and 500 µl sterile 50 % glycerol were mixed in a sterile cryo-tube and stored at - 80°C.

3.2.2 Bacterial preculture

A 50 mL plastic tube with 20 ml LB (containing 50 µg/ml of the appropriate antibiotic) was inoculated with colonies from LB-agar plates or from the cryo-culture tubes. These preculture tubes were incubated overnight (14-16 h) at 37°C and 200 rpm.

3.2.3 Protein expression

1 l flasks containing LB media (500 ml), supplemented with antiobiotic (50 µg/ml), were inoculated with 10 ml of preculture and incubated (37°C, 225 rpm) until an OD_{600} of 0.6-0.8 was reached. After induction with IPTG (1 mM), the cultures were again incubated either at 18°C and 225 rpm overnight

or at 37°C and 225 rpm for 4 h; depending on the protein being expressed. Harvesting of bacteria was carried out by centrifugation of the bacterial cultures for 20 min at 4°C and 5000 rpm. Supernatant was discarded and the pellet was stored at -20°C until further use.

3.2.4 Cell disruption

The sedimented bacterial pellet was resuspended in suspension buffer (up to 20 ml) and the cells were lysed by enzymatic cell disruption and cell-decompression.

Enzymatic cell disruption

For the enzymatic cell disruption, the bacterial suspension was treated with lysozyme and DNase, and incubated on a rotator for 30 min in the 4°C cold room. Lysozyme cleaves β-1,4-glucosidic linkages between N-acetylmuramic acid and N-acetylglucosamine in the bacterial peptidoglycan. This results in destabilisation of the outer cell membrane of gram-negative bacteria and facilitates cell decompression [127]. Moreover, DNase I digests the DNA of the bacteria by cleaving the phosphodiester bonds of the DNA backbone.

Cell-decompression

The mechanical cell disruption was carried out by nitrogen decompression, nitrogen diffuses into the cells under high pressure within a pressure vessel, called nitrogen cavitation bomb. Then, the rapid release of the pressure from the pressure vessel leads to cell membrane rupture by the gas flowing out [128]. The advantage in using this method is that nitrogen decompression is more protective for enzymes and organalles than ultrasonic and mechanical homogenizing methods. These methods depend upon friction or a mechanical shearing action that generate heat, while the nitrogen decompression procedure is accompanied by an adiabatic expansion that cools the sample instead of heating it.

For this purpose, after enzymatic disruption the cell suspension was transferred into the nitrogen cavitation bomb. The pressure vessel was filled with N_2 to a pressure of 2000 psi (about 138 bar), and then incubated at 4°C for 10 min. After incubation the pressure was released through a valve and the cell lysate was collected. This procedure was repeated three times and was followed by fractional centrifugation.

Suspension buffer:	50 mM HEPES-NaOH, pH 7.4
	200 mM NaCl
	1 mM $MgCl_2$
	1 mM $CaCl_2$

10 % Glycerin (w/v)

3.2.5 Fractional centrifugation

Fractional centrifugation allows differential clearing of a cell lysate, involving separation of organalles, cell membranes and soluble proteins based on their different sedimentation velocities. This is consecutively dependent on size and density of the corresponding component, with larger and denser particles pelleting at lower centrifugal forces.

In this work, fractional centrifugation was divided in three steps for cell membrane preparation: First, the cell lysate in 50 ml falcon was centrifuged at 1780 x g and 4°C for 15 min, the resulting pellet contained unlysed cells and nuclei. Supernatant was then centrifuged further at 12000 x g and 4°C for 30 min to remove cell-components such as low-density sedimented mitochondria.. The resulting supernatant from the second centrifugation step was then used for ultracentrifugation at 50,000 x g and 4°C for 60 min, resulting in sedimentation of cell membranes. The supernatant was discarded and the pelleted membranes were stored at -80°C until further use.

3.2.6 Protein purification: Immobilized metal ion affinity chromatography (IMAC)

Immobilized metal ion affinity chromatography (IMAC) is based on the distribution of a protein sample between a mobile and a stationary phase, which involves a specific coordinate covalent bond formation of amino acids, particularly histidine, to immobilized metal ions such as cobalt, nickel, or copper which are retained at the column matrix (agarose and metal chelator) [129]. The target protein with His-tag binds specifically and reversibly to metal ions, while non or weakly binding proteins are washed from the column. Further, initially non-specifically bound protein was eluted from the column by using imidazole, since imidazole competitively displaces bound histidines from metal ions. At higher concentrations of imidazole (100-200 mM) specifically bound target proteins were eluted.

Preparation of IMAC

The cell membrane pellets were solubilized in loading buffer (for IMAC) with 1.5% DDM and incubated on a rotator at 4°C for 2 h. The suspension was centrifuged at 50,000 x g and 4°C for 60 min. The supernatant was collected and pursued for IMAC.

The purification of GlpG enzyme and Gurken or LacY substrates was carried out with a AKTAprime system using a TALON®Crude 1 ml HiTrap column (GE Healthcare). Normally, the columns were stored in 20% (v/v) ethanol, hence it was first washed with 10 times column volume of ddH_2O and then equiliberated with 10 times column volume of loading buffer. Before use, the AKTAprime system along with tubes A and B, which are used to accomodate the buffer, were rinsed with ddH_2O. Then the system was equiliberated first with elution buffer and then with loading buffer with a flow rate of 5 ml/min. After equilibrating the system the flow rate was kept at 0.5 ml/min and the column

was placed into the system. The protein sample was diluted with loading buffer (1: 5) and loaded onto the column via tube A (flow rate 1 ml/ml). After the sample was loaded the column was washed with loading buffer until the UV signal came back to baseline signal intensity. During the washing step the impurities and non or weakly binding proteins were washed out of column. The absorpstion of proteins at 280 nm was detected by a UV detector, which is located downstream of the column. After the washing step an imidazole gradient was used to elute the specifically bound proteins (gradient length 60 ml, final imidazole concentration of 500 mM, flow rate 1 ml/ml, fraction size 3 ml). The elution fractions were collected with an autosampler and later stored on ice. In addition to the UV spectrum a SDS-PAGE was run to confirm the fractions containing the desired protein. All the fractions containing the desired protein were pooled and pursued for protein concentration.

Loading buffer for IMAC:	50 mM HEPES-NaOH, pH 7.4
	200 mM NaCl
	1 mM $MgCl_2$
	1 mM $CaCl_2$
	10 % Glycerin (w/v)
	0.05 % DDM
Elution buffer for IMAC:	50 mM HEPES-HCl, pH 7.4
	200 mM NaCl
	1 mM $MgCl_2$
	1 mM $CaCl_2$
	10 % Glycerin (w/v)
	500 mM Imidazol
	0.05 % DDM

3.2.7 Protein purification: Affinity chromatography using GST resin

Another common method used to purify recombinant proteins is affinity chromatography using resins. In the present work rhomboid protease from Aquifex aeolicus (AaROM) was purified using this method. This method involves genetically modified proteins, so-called fusion proteins, bearing tags such as glutathione-S-transferase (GST) or maltose binding protein (MBP). The basic principle behind this method is that these tagged proteins possess a high affinity towards a corresponding chelator. For example, GST has an affinity for glutathione, which is commercially available as glutathione sepharose. The AaROM recombinant protein was GST tagged and purified using this technique.

Cell membrane pellets were solubilized in suspension buffer with 1.5% DDM and incubated on a rotator at 4°C for 2 h. The suspension was centrifuged at 50,000 x g and 4°C for 60 min. The supernatant containing the GST-tagged rhomboid was collected and pursued for affinity chromatography using glutathione sepharose resin. The GST-tagged rhomboid was incubated with 1-2 ml of glutathione sepharose for 2 h at room temperature or overnight at 4°C on a rotator. Then the resin was washed with 10 column volumes of suspension buffer with 0.25% DDM, and the bound rhomboid was eluted by PreScission (Amersham Pharmacia Biotech) protease cleavage according to the manufacturer's instructions. The eluted rhomboid was again incubated with fresh 0.5 ml of glutathione-sepharose in order to get rid of unbound GST-PreScission, followed by centrifugation and collection of supernatant containing purified rhomboid.

3.2.8 Protein concentration

After IMAC, imidazole gradient purified proteins are in a buffer system with high concentrations of imidazole (100-200 mM) and hence the buffer of the pooled fractions must be exchanged before use. Centrifugal filter devices with defined pore sizes were used to concentrate the pooled fraction and exchange the buffer at the same time..In this case loading buffer was used dilute the sample after concentration. The dilution and concentration procedure was repeated 3 times.

In this study, concentration of the GlpG (MW ≈ 33 kDa), AaROM (MW ≈ 20 kDa) and Gurken or LacY (MW ≈ 66 kDa) fractions were carried out with Amicon® Ultra Centrifugal filters with a 10 kDa and 30 kDa molecular weight cut-off (Millipore).

3.2.9 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-polyacrylamide gel electrophoresis allows the separation of denatured proteins according to their electrophoretic mobility, which is generally based on size, conformation and charge of the molecule. A polyacrylamide gel is formed by the free radical crosslinking of acrylamide with N,N-methylene-bisacrylamide. The sample loading buffer contains the anionic detergent SDS (Sodium dodecyl sulfate) which denatures proteins and results in an overall negative charge. Therefore, separation of the proteins is not influenced by any intrinsic charges and mainly depends on the molecular weight.

NuPage Novex Gels

- Prepare resolving gel mixture:

	10 %		12 %	
	1 gel	**2 gels**	**1 gel**	**2 gels**
30% Acrylamide (37.5:1)	2.20 ml	4.40 ml	2.64 ml	5.28 ml
1.6 M Bis-Tris, pH 6.4	1.65ml	3.30 ml	1.65ml	3.30 ml
MilliQ-water	2.70 ml	5.40 ml	2.26 ml	4.52 ml

10% APS stock solution	33 µl	66 µl	33 µl	66 µl
TEMED	11 µl	22 µl	11 µl	22 µl

- Gently mix the resolving gel mixture and fill ¾ of the Novex gel cassette.
- Carefully overlay the gel mixture with isopropanol and wait until the polymerization is complete (≈15-20 min).
- Pour off the isopropanol and wash the gel surface with MilliQ water. Carefully drain all the water out of the gel cassette and dry the inside with a small Whatman paper. Do not touch the gel surface!
- Prepare the stacking gel mixture:

	4 %	
	1 gel	**2 gels**
30% Acrylamide (37.5:1)	260 µl	520 µl
1.6 M Bis-Tris, pH 6.4	500 µl	1.00 ml
MilliQ-water	1.23 ml	2.46 ml
10% APS stock solution	20 µl	40 µl
TEMED	5 µl	10 µl

- Gently mix the stacking gel mixture and pour it onto the gel cassette up to the top.
- Place the comb into the gel cassette and wait until the gel is polymerized.
- Gels can be used immediately or wrapped in wet paper towels and stored at 4°C.
- Remove the comb and wash each slot with MilliQ water.
- Place gel in into running chamber and fill with 1 x MES running buffer.
- Load samples, as a size standard PageRulerTM prestained protein ladder (10-170 kDa, Thermoscientific) was used, run the gel at 100 V for initial 20 min and then at 150 V for 30 min.

Bis-Tris gel-buffer: 1.6 M Bis-Tris, pH 6.4

Store at RT

10 % APS: 10 % Ammoniumpersulfate in MilliQ water

Store at -20°C

20 x MES: 1 M MES

1 M Tris-Base

69.3 mM SDS

20.5 mM EDTA

	Store at RT
1 x MES:	1:20 dilution of 20 x stock in MilliQ water
4 x SDS sample buffer:	1.44 M Bis-Tris
	0.64 M Bicine
	4 % SDS
	100 mM DTT
	0.05% Bromophenoleblue
	Store at -20°C

3.2.10 Coomassie staining

Coomassie brilliant blue staining

Coomassie brilliant blue is a staining dye that was used to detect proteins in acrylamide gels. The Coomassie dyes (R-250 and G-250) bind to proteins through ionic interactions between dye sulfonic acid groups and positive protein amine groups as well as through Van der Waals attractions.

- Place gel in prewarmed Coomassie staining solution and incubate for 30 min with gentle shaking
- Remove staining solution (do not discard, can be reused multiple times) and rinse gel once with water.
- Add warm destain solution and incubate at RT with shaking. Discard the destaining solution and repeat it multiple times until the background staining is removed and the proteins bands are clearly visible.

Coomassie staining solution:	10 % glacial acetic acid
	25 % ethanol
	2.5 % coomassie brilliant blue
Destain solution:	10 % glacial acetic acid
	12.5 % ethanol

InstantBlue™ staining

The staining of polyacrylamide gels with InstantBlue™ (Expedeon) is also a Coomasie-based staining method. The SDS gels were incubated with InstantBlue™ at RT for 15 min with gentle shaking. Once used, the staining solution should be discarded and the SDS gel was washed with MilliQ water for 20 min.

3.2.11 Western blot

Western blot is an immunological detection method for proteins. In general, a SDS-PAGE is performed first and the proteins that were separated within the SDS gel are transferred on a membrane by migration in the electric field. The membrane is then incubated with specific antibodies which bind to the target protein. In this work, the proteins were transferred on polyvinylidene fluoride (PVDF) membranes (LI-COR), which are suitable to use with infrared dye coupled antibodies (FL-PVDF).

Western blot protocol (wet transfer)

- First cut out 2 pieces of Whatman paper and one piece of PVDF membrane (0.2 µm pore size) per gel, size should correspond to gel size.
- Then soak Whatman paper along with two blotting sponges in 1 x transfer buffer, activate the PVDF membrane in 100 % ethanol.
- Prepare blotting stack as shown in figure 3.1.

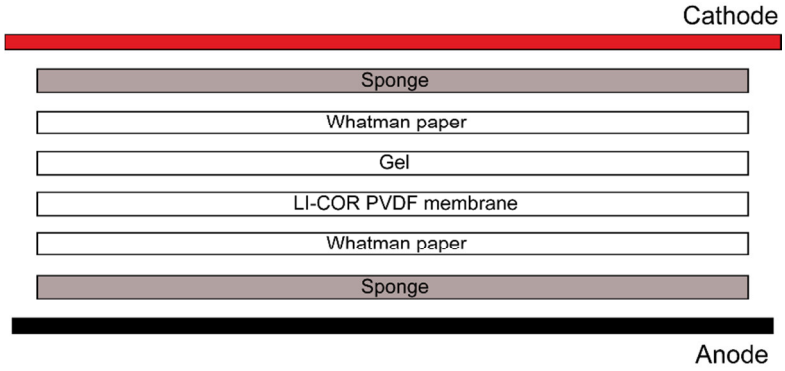


Figure 3.1: Blotting stack of the western blot for wet transfer.

- Fill blotting tank with 1 x transfer buffer and place stack in the tank.
- Blot for 2 h at 200 mA or overnight at 30 V.

10 x transfer buffer: 250 mM Tris-Base

1920 mM Glycin

Store at RT

Western blot protocol (semi-dry transfer):

- Cut out four pieces of Whatman paper and one PVDF membrane (0.2 µm pore size) per gel, the size should correspond to the gel size.
- Pour blotting buffers into corresponding containers and equilibrate Whatman paper. Place one piece in blotting buffer A, one piece in blotting buffer B, and two pieces of paper in blotting buffer C, ensure soaking for at least 15 min.

- Soak PVDF membrane in methanol for activation, followed by washing briefly with MilliQ water and then placing in blotting buffer B.
- Equilibrate gel as well by soaking in blotting buffer C and assemble the blotting stack as shown in figure 3.2.
- The transfer was carried out by applying constant current of 50 mA per blot for 60 min.
- After blotting, rinse membrane in MilliQ water once.

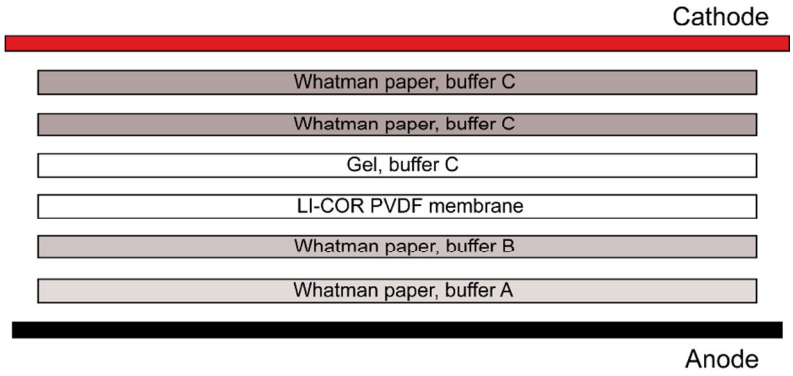


Figure 3.2: Blotting stack of the western blot for semi-dry transfer.

Blotting buffer A:	210 mM Tris-Base, pH 10.4
	30 % Methanol
Blotting buffer B:	25 mM Tris-Base, pH 10.4
	30 % Methanol
Blotting buffer C:	25 mM Tris-Base
	0.025 % SDS
	pH to 9.0 with 5 M boric acid

Immunostaining of PVDF membranes

- After completion of the transfer, block membrane in 5% dry-milk/1x TBS-T for 1 h at RT while shaking.
- Wash blocking solution of the membrane, 1x 15 min with 1x TBS-T.
- Dilute primary antibody in 1x TBS-T and add 0.02% NaN_3. Incubate for 2 h at RT or 4°C overnight with gentle shaking. Dilution depends on the antibody.
- Remove primary antibody solution (do not discard, can be reused multiple times when stored at 4°C) and wash membrane three times with 1 x TBST buffer at RT for 20 min each.
- Incubate membrane with secondary antibody solution (diluted 1:10000) for 1 h at RT with gentle shaking (in the dark!). For the anti-His antibody dilute secondary antibody in blocking solution.
- The secondary antibody solution was discarded and the membrane was washed at least twice with 1 x TBS-T and once with 1 x TBS for 20 min each (RT, in the dark!).

- All incubations were carried out on an orbital shaker and the blot membrane was exposed on an Odyssey® CLx (LI-COR) for analysis. The protein standard was detected at a wavelength of 700 nm and the IRDye®-infrared dyes of the secondary antibodies at 800 nm.

10x TBS, pH 7.4:	1.37 M NaCl
	27 mM KCl
	0.25 M Tris-Base
1x TBS:	1:10 dilution of 10x TBS
1x TBS-T:	1:10 dilution of 10x TBS
	0.01 % Tween-20

3.2.12 EcGlpG *in vitro* activity assay

All activity assays were performed in a 50 or 20 µl reaction mixture. The purified and concentrated protein samples of rhomboid GlpG (0.35 µM) and Gurken or LacY substrate (1.8 µM) were diluted into reaction buffer. For the activity assay, inhibitors with defined concentrations were first added in reaction buffer with GlpG and then preincubated for 30 min at 37°C and 600 rpm. Thereafter, the substrate was added and the reaction was continued for 90 min at 37°C and 600 rpm. After completion of reaction, a SDS-PAGE was run to analyse the reaction products. As additional controls only rhomboid (reaction buffer and GlpG), only substrate (reaction buffer and Gurken) and reference (reaction buffer, substrate and rhomboid) were used in the activity assays.

Reaction buffer:	50 mM HEPES-NaOH, pH 7.5
	5 mM EDTA
	0.4 M NaCl
	10 % Glycerol (v/v)
	0.05 % DDM

3.2.13 AaROM *in vitro* activity assay

All activity assays of AaROM were performed in 20 µl reaction mixture containing the purified and concentrated protein samples of rhomboid AaROM (0.094 µM) and Gurken substrate (0.4 µM). Activity assay buffer used here was the same as for the GlpG *in vitro* activity assay. For the determination of IC50 values, inhibitors with defined concentrations were first added in reaction buffer with AaROM and then preincubated for 30 min at 37°C and 600 rpm. Thereafter, the cleavage reaction was initiated by addition of substrate and the cleavage reaction was continued for 90 min at 37°C and 600 rpm. After completion of reaction, a SDS-PAGE was run to analyse the reaction

products. Similar to the GlpG *in vitro* activity assay, control reactions containing only rhomboid (reaction buffer and AaROM), only substrate (reaction buffer and AaROM) and reference (reaction buffer, substrate and rhomboid) were used in the activity assays.

3.2.14 α-Chymotrypsin Assay

N-Succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (Suc-AAPF-pNA) is a known substrate of bovine α-chymotrypsin and has been used for determination of chymotrypsin activity [130]. Enzymatic cleavage of the 4-nitroanilide substrate results in 4-nitroaniline which has a yellow color under alkaline conditions. The rate of hydrolysis of 4-nitroanilide was measured in chymotrypsin assay buffer at pH 7.5 at 25°C. The increase in absorbance at 410 nm was detected with a microtiter plate reader (Beckman Coulter). For the chymotrypsin activity assay a 50 µl reaction containing 4 µM chymotrypsin with or with out inhibitor was preincubated for 30 min at 25°C. For evaluation of the IC50 of an unknown compound against chymotrypsin a defined range of concentrations (0 µM – 1 mM) were used and as additional controls only chymotrypsin, only substrate and reference (substrate and chymotrypsin) were used in the activity assays. After this 2.5 µl of pre-incubated reactions was added to 100 µl of 0.12 mM Suc-AAPF-pNA substrate in HEPES buffer. As a result Suc-AAPF-pNA gets hydrolysed and release 4-nitroaniline with in approximately 5 min, and the initial and final absorbance values were recorded at 410 nm. Absorbance values for different concentrations were used to calculate IC50 values by a nonlinear regression curve fit of the log (inhibitor) vs. response (variable slope) in Graphpad prism software.

α-Chymotrypsin assay buffer: 50 mM HEPES-NaOH, pH 7.5
5 mM EDTA
0.4 M NaCl
10 % Glycerol (v/v)
10 % DMSO

Chymotrypsin: Stock solution of 4 µM in assay buffer.

N-Succinyl-Ala-Ala-Pro-Phe p-nitroanilide: stock solution of 0.12 mM in assay buffer (gentle warming for dissolution required).

3.3 Cell culture

Cell culture experiments were performed using consumables such as glasswares, pittetes, dishes, filter tips, etc. under sterile working conditions. All solutions and media were kept at 4°C and prewarmed in a water bath at 37°C before use.

Cell culture media and buffer:

DMEM - Dulbecco's Modified Eagle Medium complete (Gibco)

(+ 4.5 g/l Glucose, + L-Glutamin, -Pyruvate)

10 % FCS

2 % Penicillin/Streptomycin (10000 U/ml Penicillin/ 10000 µg/ml Streptomycin)

1 % Sodium pyruvate

DPBS - Dulbecco's phosphate-buffered saline (Gibco)

-$CaCl_2$, -$MgCl_2$

3.3.1 Passaging of adherent cell lines (HEK293T)

Eukaryotic cells were cultured in sterile cell culture dishes (Ø = 10 cm) with 10 ml of cell culture medium (DMEM) at 37°C and 5 % CO_2. The cells were cultured until they were confluent. Medium was changed if the color of the medium changed before cells reached confluency.

Passaging of adherent cells:

- Remove the cell culture medium from dish.
- Add 10 ml PBS (Gibco, Invitrogen) to wash cells, sway dish and remove buffer using suction pump.
- Subsequently, the cells were treated with 1 ml trypsin, sway dish and incubate at 37°C for 2-3 min to detach the cells from the bottom of the cell culture dish.
- Add 10 ml of DMEM and pipette up and down to separate cells.
- Transfer 1 ml cell suspension into a new 10 cm dish and add 9 ml fresh cell culture medium, sway dish carefully to ensure equal cell distribution.
- Place dish into cell culture incubator at 37°C and 5 % CO_2.

3.3.2 Thawing of frozen cell stocks

- Prewarm the culture medium (DMEM) in a waterbath at 37°C.
- Pour 10 ml DMEM into a new sterile 10 cm dish.
- Thaw cells by placing cryo-culture tube in water bath (37°C) for 2-3 min.
- Transfer cells into the 10 cm dish containing DMEM using 2 ml pipette and place in incubator at 37°C and 5 % CO_2.

3.3.3 Determination of cell number

The determination of the cell number of a cell suspension for seeding a certain number of cells was performed with a Vi-CELL Cell Viability Analyzer (Beckman Coulter). This automated method is based on trypan blue staining and in addition to the total number of cells and it also determines the number of viable cells.

3.3.4 Cell seeding and Transfection of HEK293T cells

Day 1: Cell seeding

- Discard old cell culture medium.
- Add 10 ml 1 x PBS to wash cells, sway dish and discard buffer.
- Subsequently, the cells were treated with 1 ml trypsin, sway dish and incubate at 37°C for 2-3 min, to detach the cells from the bottom of cell culture dish.
- Add 10 ml of DMEM and pipette up and down to separate cells.
- Transfer the cell suspension in a 15 ml tube and put 1 ml into a cell counter vial and perform cell count with Vi-CELL.
- Coat dishes with poly-L-lysine (5 µg/ml in 1 x PBS; sterile filtered) for 30 min at 37°C and wash once with 1 x PBS.
- Seed appropriate number of cells/well on a new plate and add fresh media.
    - **6 well dish:** $0.5x10^6$ cells
    - **10 cm dish:** $4.0x10^6$ cells
    - **15 cm dish:** $9.0x10^6$ cells
- Carefully sway plate to acheive even distribution of cells.
- Place plate in incubator (37°C, 5% CO_2).

Day 2: Transfection

Transfection was performed using GeneJuice® or PEI as transfection reagents according to the manufacturer's protocol for transfection of adherent cells.

- For transfection cell density should be 60-80%
- Prior to transfection bring all reagents to room temperature.
- Prepare transfection mixture: In a sterile tube dilute plasmid DNA in OptiMEM w/o phenol red (volume of media is 10% of final volume in culture dish) and incubate for 5 min at RT.
    - **6 well dish:** 200 µl + 0.5 -1 µg of total DNA
    - **10cm dish:** 1 mL + 6-7 µg of total DNA
    - **15cm dish:** 2 mL + 10-12 µg of total DNA

- Add GeneJuice® or PEI (1µg/µL) to the diluted DNA. Mix immediately by vortexing or pipeting and incubate for 15-20 min at RT. The volume of transfecting reagent used is based on a 3:1 ratio of GeneJuice® or PEI (µl): total DNA (µg).
    - **6 well dish:** 1.5 µl of PEI(1µg/µl) = 1.5 µg
    - **10cm dish:** 18 µl of PEI (1µg/µl) = 18 µg
    - **15cm dish:** 30 µl of PEI(1µg/µl) = 30 µg
- Carefully change the medium on HEK293T cell.
- Add the transfection mix in a dropwise manner onto the cells and carefully sway the plate.
- Place in incubator (37°C, 5% CO_2) in order to express the target proteins in HEK293T cells and harvest transfected cells 48 hours post-transfection.

Reagents:

PEI (1µg/ul) – PEI is Polyethylenimine, 25kD, linear from Polysciences (cat# 23966-2). To prepare a PEI stock solution:

- Dissolve PEI in endotoxin-free ddH_2O that has been heated to ~80°C.
- Let the mixture cool to room temperature.
- Titrate to pH 7.0, filter sterilize (0.22 µm), aliquot and store at -20°C; a working stock can be kept at 4°C.

3.3.5 Harvesting of cell culture supernatants

Cell culture supernatant was harvested to investigate secreted proteins:

- Label two 1.5 ml tubes for each sample.
- Add 25 x Protease Inhibitor stock solution (20 µl per 500 µl of supernatant).
- Transfer supernatant from a cell culture dish (keept on ice) into the tube containing the protease inhibitor solution.
- Spin samples for 3 min at 4°C and 12000 x g.
- Transfer supernatant to fresh tube and store at -80°C.

25 x Complete Protease Inhibitor (EDTA-free; Roche): one tablet in 2 ml Milli-Q water.

3.3.6 Preparation of cell lysates

- Label two 1.5 ml tubes for each sample.
- Keep 1 x PBS on ice.
- Dilute the Complete Protease Inhibitor stock solution (25 x) to a final concentration of 1 x using NP40 lysis buffer.
- Place cell culture dish on ice.
- Discard supernatant and add cold 1 x PBS to wash cells.

- Scrape off cells from plate by adding 1 ml 1 x PBS.
- Pipette cells into 1.5 ml tube and centrifuge for 3 min at 12000 x g.
- Discard supernatant and add appropriate amount of NP40 lysis buffer with protease inhibitor.
- Incubate for 20 min on ice and mix every 5 min by vortexing.
- Centrifuge for 15 min at 4°C and 12000 x g.
- Transfer supernatant into new tube.
- Store at -20°C.

NP40 Lysis Buffer:	50 mM Tris-HCl, pH 7.8
	150 mM NaCl
	1 % NP40
	Store at 4°C.
10 x PBS, pH 7.4:	137 mM NaCl
	2.7 mM KCl
	10 mM Na_2HPO_4
	10 mM KH_2PO_4
	Store at RT.

3.3.7 Bicinchonic acid protein assay (BCA)

In order to use a specific amount of protein for experiments and also for loading equal amounts of protein onto SDS gels, the concentration of the samples was determined by bicinchoninic acid (BCA) protein assay. This was performed by using the Pierce BCA protein assay kit according to the manufaturer's instructions. The BCA protein assay involves the reduction of Cu^{2+} ions to Cu^{+} ions by proteins in an alkaline solution. Subsequently, BCA forms a color complex with the Cu^{+} ions and has an absorption maximum at 562 nm.

Determine protein concentrations in cell lysates:

- Prepare one set of 1.5 ml tubes for the protein standard.
- Set heating-block to 60 °C.
- Prepare dilutions of BSA stock solution (1 mg/ml) for the standard curve:

Final concentration [µg/ml]	Volume NP40 buffer [µl]	Volume BSA standard [µl]
0	100	0
100	90	10
200	80	20
300	70	30

400	60	40
500	50	50
600	40	60

- Transfer 25 µl of each standard mixture into two wells of a 96-well plate.
- Add 22.5 µl NP40 buffer per sample into the 96-well plate and add 2.5 µl of sample.
- Prepare BCA Reagent mix with reagent A and B in a 50:1 ratio.
- Add 200 µl of BCA reagent mix to each sample and standard.
- Incubate for 20-30 min at 60 °C.
- Measure OD at 562 nm with a microtiter plate reader (Beckman Coulter) and calculate the protein concentration of cell lysate samples using a calibration curve of the standards (absorbance as a function of protein concentration).

3.3.8 Inhibitor treatment

To evaluate the effects of the rhomboid inhibitors in a cell culture-based system, cells were treated for 24 h with the respective compounds and vehicle control and cell lysates were analyzed.

- Cell were seeded into a 6-well plate and incubated until they reached 50-80% confluency before transfection (3.3.5).
- After transfection the cells were incubated for 2 h (37°C, 5% CO_2).
- Dilute compounds and vehicle control in pre-warmed culture medium. Make sure that after addition of inhibitors at the desired final concentrations the DMSO concentration must not exceed 1 % to prevent toxicity. As a control, cells were incubated without inhibitor and DMSO only.
- 24 h after transfection cell lysates were prepared (3.3.6), followed by protein concentration determination (3.3.7), SDS-PAGE and Western blot (3.2.8 and 3.2.10) to evaluate the inhibitor effect.

3.3.9 Cell toxicity assay

Day 1: Cell seeding

- Prewarm the culture medium (DMEM), PBS buffer, and trypsin in a waterbath at 37°C.
- Discard cell culture medium from cells.
- Add 10 ml 1 x PBS to wash cells, sway dish and discard buffer.
- Subsequently, the cells were treated with 1 ml trypsin, sway dish and incubate at 37°C for 2-3 min, so that cells can detach from the bottom of the cell culture dish.
- Add 10 ml of DMEM and pipette up and down to achieve cell separation.

- Transfer the cell suspension in a 15 ml tube and add 1 ml into a cell counter vial and perform cell count with Vi-CELL.
- Seed 30000 cells/well by into a 96-well plate and keep the total volume of media equal to 200 µl in each well.
- Place plate in incubator (37°C, 5% CO_2).

Day 2 : Compound treatment

- Prewarm the culture medium (DMEM) in a waterbath at 37°C.
- Prepare the stock solutions of the inhibitors at ten different concentrations (0 µM – X µM) in such a way that the total volume of the media equals to 650 µl and % of DMSO ≤ 1, vortex well.
- Take the 96-well plate with seeded cells from the incubator and discard the medium, add 200 µl of the prepared inhibitor solutions with diffrerent concentrations in triplicates.
- Place the plate back into the incubator (37°C, 5% CO_2).

Day 3 : Treatment with alamarBlue®

alamarBlue® is a cell viability indicator which relies on the antioxidant/reducing capacity of living cells. The cells therefore convert resazurin into the fluorescent molecule resorufin. Hence, by addition of alamar blue to cells treated with an unknown compound the cytotoxicity of a substances can be assessed.

- Take a 96-well plate containing cells treated with compound.
- Add 20 µl alamarBlue® to each well treated with compound using a multichannel pipette.
- Place the 96-well plate back into incubator (37°C, 5% CO_2).
- Keep observing time to time change in the color of the media from violet blue to pink which indicates cell viability and can be measured by performing fluorescence readout.

3.4 Molecular Docking

Molecular docking is a computational technique in drug design which is used for predicting the binding mode of small drug like molecules within the active site of an enzyme or cell surface receptor binding pocket. Molecular modeling allows us to visualize different orientations and conformations of ligands within the protein active site. Moreover, it also enables us to rank various ligands based on their binding affinity, which is rated according to optimal binding geometries and energies of the respective molecules. In the present study, all docking experiments were performed using the Molecular Operating Environment (MOE) software.

3.4.1 Preparation of the Receptor

Molecular modeling experiments were based on a co-crystal structure of the *E. coli* rhomboid GlpG and the phosphate inhibitor Cbz-AlaP(OiPr)F or CAPF (PDB-ID: 3UBB. The original X-ray structure of the receptor was prepared with the LigX interface in MOE window. LigX is a combination of several procedures that include the deletion of distant water molecules (4.5 Å farther from the ligand), addition of hydrogens, calculation of charges, molecular geometry optimization and system refinement ($RMSD_{gradient}$ = 0.1 kcal/mol). Most protein structures lack hydrogen atoms, therefore addition of hydrogen in the LigX interface can be performed with Protonate 3D. Along with this, terminal amides as well as sulfonamide and imidazole groups can be allowed to flip to optimize the hydrogen bond network by using the ASN/GLN/HIS-"Flips" option in Protonate 3D. For energy minimization, forcefield AMBER12: EHT was used which is parameterized for both proteins and small molecules. After all those changes the prepared receptor was saved as a new moe file.

3.4.2 Preparation of the Molecule database

A new database was generated using the builder module in MOE. This involved drawing the molecule in MOE window. In a following step hydrogens were added and charges were assigned using the potential setup interface. Subsequently, a distance-dependent dielectric solvation treatment and energy minimization were performed in MMFF94x-forcefield. After completion of these tasks the prepared ligand was added into a newly created database using the panel "add new entry".

3.4.3 Docking and scoring

The molecular modeling in MOE can be performed in two ways: non covalently by using the DOCK workflow and covalently by using the Conflexdock workflow or the DOCKTITE workflow. Both these methods provide the favorable binding modes of ligands within the target protein active site. Furthermore, docking allows us to generate a number of different spatial arrangements (poses) for each ligand which are subsequently energy minimized and scored. The scoring function can either be based on (i) the free energy of binding (which includes among others solvation and entropy terms), (ii) is enthalpy based and includes polar interaction terms as well as metal ion coordination, or (iii) results in a qualitative shaped-based numerical value.

3.4.3.1 Non-covalent docking module

The DOCK application in MOE consists of five stages: (1) conformation generation, (2) placement, (3) scoring, (4) pose refinement, and (5) rescoring.

To date, only covalent inhibitors in complex with a rhomboid protease are known. Hence, before performing the docking experiments using DOCK module in MOE an important step was to break the

covalent bond between the ligand and the receptor (rhomboid) using the builder tool which allows the ligand to be free from all restrictions due to covalent binding to its target. The GlpG complex with the CAPF inhibitor consists of a covalent bond between a phosphorus of CAPF and an oxygen atom of serine 201 of GlpG. This bond was removed before commencing the DOCK workflow. DOCK workflow can be applied for the generation of numerous conformations of a given ligand by applying a collection of preferred torsion angles to rotatable bonds within the ligand molecule. A systematic search is carried out keeping unaltered bond lengths and bond angles and producing all combinations of angles on a grid. Next, by using the triangle matcher placement method, a number of poses were generated from the collection of ligand conformations. The generated poses were rescored by using London dG as scoring function. It estimates the free energy of binding of the ligand from a given pose to the target structure and comprises seven components: c (average loss/gain of rotational and translational entropy), E_{flex} (energy due to loss of ligand flexibility), f_{HB} (geometric imperfections of hydrogen bonds) c_{HB} (energy of an ideal hydrogen bond), f_M (geometric imperfections of metal ligations), c_M (energy of an ideal metal ligation) and D_i (desolvation energy of atoms). The top thirty poses were further refined by energy minimization inside the active site using the MMFF94x forcefield or AMBER12: EHT-forcefield to a RMSD-gradient of 0.01 kcal/mol and keeping tether side chains. The poses were rescored in a similar manner using London dG or Affinity dG which resulted into the ten highest rated conformations that were collected in the output database.

3.4.3.2 Covalent docking : ConflexDock in MOE

In most modeling computational softwares, including MOE, molecular docking is mainly based on docking noncovalent modulators to a target protein. However, the formation of a covalent bond within the active site of the target protein is necessary for covalently binding inhibitors or drugs. Hence, to obtain significant information about covalent binders, covalent docking is an essential and advanced computational tool in MOE and one possible way of implementation is by following the ConflexDock workflow.

The ConflexDock wizard in MOE allow us to examine a covalent inhibitor in its regular binding mode and includes the covalent bond formation between ligand and receptor. Hence, using this application in MOE it was possible to analyse the protein/ligand docked structure in its physiologically relevant form along with all neighbouring molecular interactions. The most important step in covalent docking is labeling the atom that forms the covalent bond ANC, which defines the anchoring point. The rhomboid protease GlpG has an active site serine residue which forms a covalent bond with irreversible inhibitors such as CAPF. This covalent bond is formed between the phosphorus of CAPF and the oxygen of Ser201 of GlpG. To perform the covalent docking experiment the phosphorous of CAPF labeled ANC and the oxygen of Ser201 have to be kept fixed. During the ConflexDock workflow the atoms of the active site amino acid side-chains can be permitted to move but backbone

atoms are fixed in their original position. Similar to the docking workflow described earlier the ConflexDock workflow also results in the generation of a defined number of poses and is followed by gradient based energy minimization. The final highest scoring poses, along with their scores and conformation energies, are exported to a database for further analysis.

3.4.3.3 Covalent docking : DOCKTITE in MOE

DOCKTITE is a step-by-step, time-efficient and user friendly workflow for covalent docking in MOE that has recently been designed using SVL-scripts by Scholz *et al.* [131]. It is a highly versatile script adopted from the non-covalent docking wizard DOCK in MOE and is compatible for large-scale docking-based virtual screening experiments. Moreover, DOCKTITE helps to overcome the drawbacks from other covalent docking methods (DOCKovalent, CovDock) such as distortion of the docked ligand, warhead limitations, etc [131]. The receptor and ligand database were prepared by following the method explained in section 3.4.1 and 3.4.2. The DOCKTITE workflow (figure 3.3) includes four major steps: 1) Warhead screening 2) Side chain attachment 3) Pharmacophore-guided docking 4) Side chain cleavage and pose rescoring.

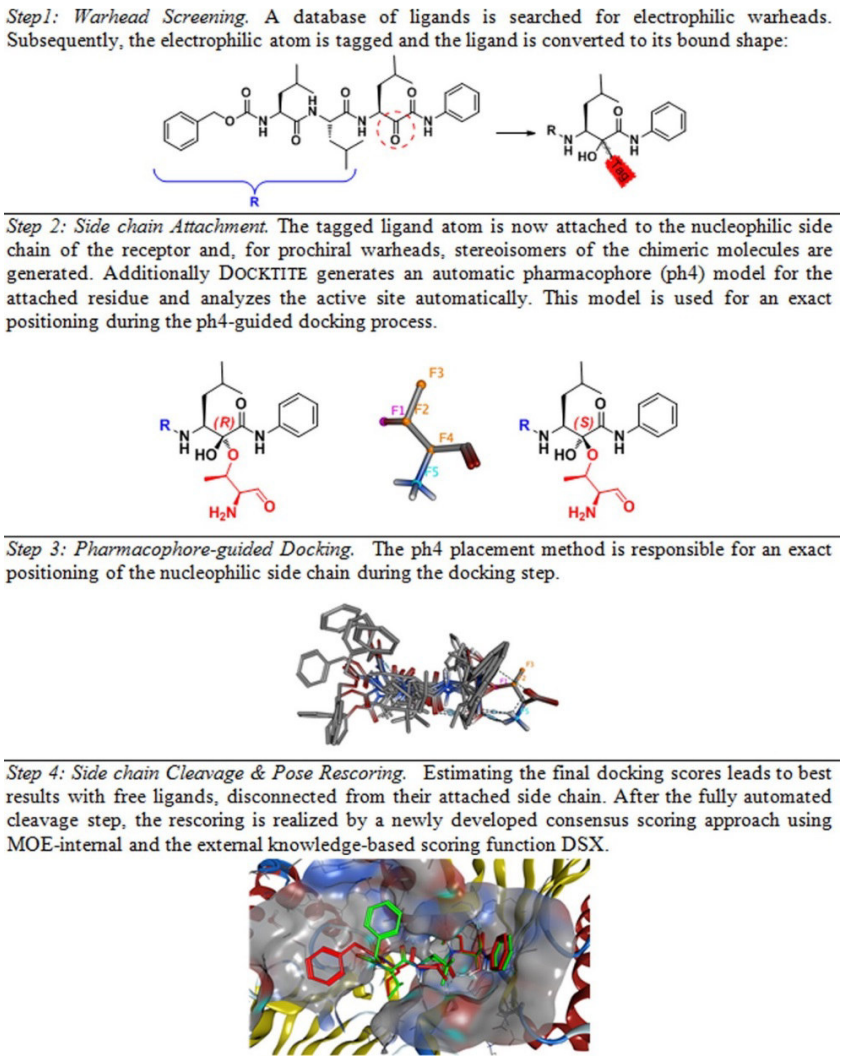


Figure 3.3: Overview of the DOCKTITE workflow. Figure adopted from [131].

1) Warhead screening: The first step included the screening of a large database of drug-like molecules via user-defined electrophilic/covalent warheads. The electrophilic atom of the warheads was transiently tagged with a dummy atom which leads to dissociation of the leaving group and finally results in a structure which remsembles that of a bound-state ligand. All these warhead reactions are perfomed automatically by a modified MedChem Transformations algorithm (MCT) of MOE, specialized for customizing the binding reactions. Reactions were drawn in Chemdraw software and added to set of SVL-based MCT which was implemented within the drug-like molecules database. This involved the search of substructures that match with the MCT query set. These substructures were modified in such a way that a potential leaving group is split off – this models the modified structure of the ligand after reaction with the target enzyme. Afterwards, the structure was temporarily tagged by an unique element such Tantalum (Ta) which is normally absent in drug-like candidates. Besides this, two other tagging elements were also employed: Yttrium (Y) for leaving groups and Germanium (Ge) for mimicking a tetravalent boron atom; two properties which were intrinsically absent in MCT. This warhead screening helps us to recognise different covalent inhibitors in MOE and warhead designing is based on the binding mechanisms encompassing rearrangements and also release of leaving groups. For example, some saccharin-based inhibitor binding mechanism involves the covalent bond formation with receptor and afterwards featuring release of leaving group [110]. In this study, we have used a dataset of 34 warhead classes (section 4.7) and some of them are shown in figure 3.4.

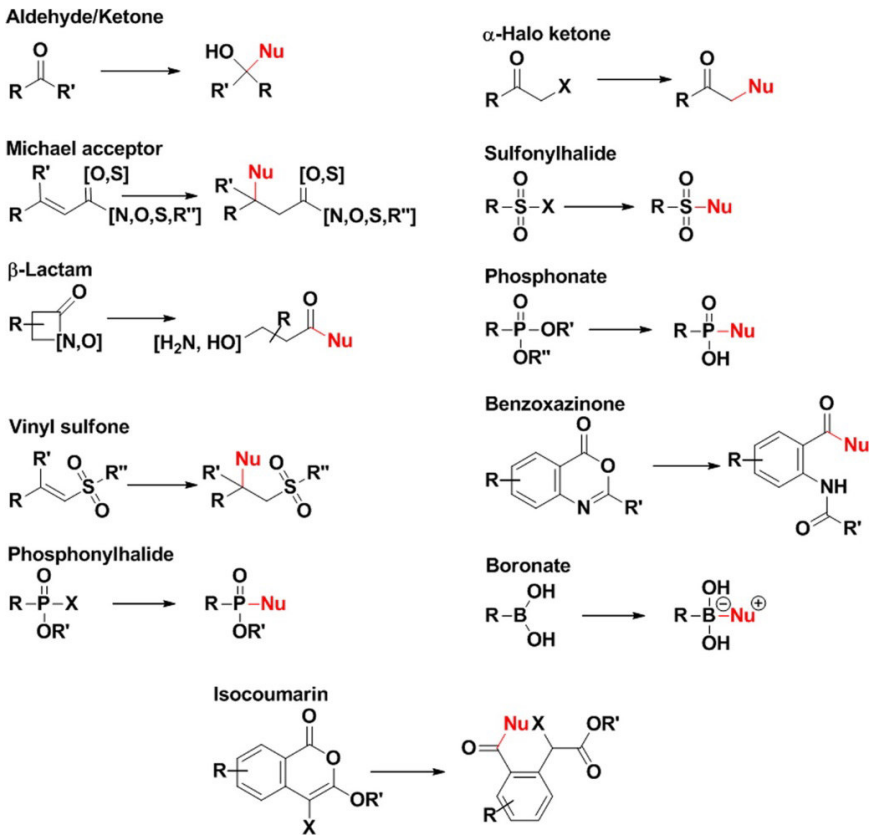


Figure 3.4: Electrophilic warheads implemented in DOCKTITE. Nu: nucleophilic attachment point of receptor. R, R', R": substituents with terminal carbon. X: halogen. The covalent bond is shown in red. Figure adopted from [131].

2) Side chain attachment: This step is a collection of various modifications and preparations for the final covalent docking. First, the active site residue bearing the nucleophilic side chain (Ser201 in rhomboid protease GlpG) of the prepared receptor (3.4.1) was detached from rest of the receptor molecule and its native position was fixed using pharmacophore (ph4) constraints.The tagged ligands from the first step were used as input database for the current step and side chain attachment was performed. This key step involved the replacement of the Ta-tag with the nucleophilic residue, exchange of the Ge-tag to boron and deletion of the Y-tagged part of the ligand, which represents the leaving group. Furthermore, the resulting modified ligand was energy minimized and ligands featuring prochiral warheads generated stereoisomers which were considered as separate ligands during the next steps.

3) Pharmacophore-guided docking: This is the final docking step that was performed using DOCK wizard in MOE. As described in section 3.4.3.1 dock module is a compilation of five steps. The resulting ligand conformations are obtained by calculating the preferred torsion angle to rotational bonds orientation. Using the pharmacophore placement method nucleophilic residues could be precisely positioned, which was further guided by subsequent refinement steps. The generated poses were ranked using the emperical scoring function Affinity dG which is based on counting the number of various types of interactions between the two binding partners. Affinity dG represents enthalpic contributions to the free energy of binding by a linear function accounting for favorable and unfavorable hydrophobic contacts, hydrogen bonds, ionic interactions, and metal ligations. Similar to the non-covalent docking algorithm of the DOCK module the 100 best scored poses were refined by energy minimzation using grid minimization in AMBER12:EHT forcefield. The refined poses were rescored using Affinity dG. The binding affinities of the ligand here also depends on the attached nucleophilic residue and even small conformational alterations can have a huge impact on the calculated results. Hence, the next step in the pharmacophore-guided docking workflow is to disconnect the nucleophilic residue from the chimeric molecule and perform a pose rescoring.

4) Side chain cleavage and pose rescoring: This last DOCKTITE workflow step includes automated detachment of the nucleophilic residue from the ligand and a scoring algorithm using affinity dG which results in only one best scored pose.

3.4.4 Ranking of ligands

In this study, molecular modeling was performed using DOCK, ConflexDock and DOCKTITE modules in MOE. The resulting binding affinities of the ligands were based on docking forcefield and empirical scoring functions (London dG or Affinity dG). Generally, binding affinity (scoring function) is the first criteria for hit selection and optimization to obtain a lead molecule. But affinity often relies on molecular size. Hence, to obtain high affinity compounds one may has to compromise with its physicochemical properties according to Lipinski's rule of five [115]. As a result, taking scoring

function as the first ranking aspect can be misleading. Hence, Hopkins *et al.* [132] introduced a new ranking criteria called ligand efficiency (LE) which corresponds to the binding affnity in relation to the number of heavy atoms or non-hydrogen atoms in a compound. This ranking factor accounts for the individual contribution of every atom rather than the whole molecule to the binding affinity and therefore leads towards molecules that use their atoms most efficiently to bind to the assigned target structure. Hit selection based on LE may result in compounds with lower scores but can be attractive for optimization. Mathematically, ligand efficiency (LE) is defined as the ratio of Gibbs free energy (ΔG) to the number of heavy atoms or non-hydrogen atoms of the ligand (HA).

$$LE = \frac{\Delta G}{HA} = \frac{1.4\ (-logIC50)}{HA}$$

Where $\Delta G = RTlnKi = \ 1.4\ (-logIC50)$

Therefore LE was calculated for every docked compounds in MOE by dividing docking score (S-value) by number of haevy atoms (HA) and called as calculated ligand efficiency.

$$cLE = \frac{Docking\ score}{HA}$$

cLE was considered as the first criteria for ligand ranking in this study, followed by scoring function. Besides these two factors, another important criteria are ligand interactions (LI). This allows us to explicitly observe the placement of a ligand within the receptor pocket and compare the resulting binding interactions.

3.5 Chemical compound synthesis

All chemical compounds listed in Table 2.1-2.6 have been synthesized by following the scheme explained in section 4.5 and 4.6 and their detailed synthesis and spectral analysis are desribed below.

3.5.1 General information

All commercial chemicals, reagents, solvents were purchased from Sigma–Aldrich and were used without any further purification. All reactions were performed under argon atmosphere using dry solvents unless otherwise specified. The progress of all the reaction was monitored with thin layer chromatography (TLC) using aluminium sheets precoated with silica gel 60 F254 (0.2 mm, Merck) in ethyl acetate/n-hexane as solvent system. Chromatographic spots were visualized by UV light. High performance liquid chromatography (HPLC) was carried out using an Agilent 1100 (Column: reverse phase, Zorbax Eclipse XDB-C8, 4.6 x 150 mm, Wavelength: 254 nm, Solvent System: Acetonitrile/Water (gradient), Method: 12 min run time. 70% water: 30% ACN for initial minute, then linear gradient until it reaches to 90% ACN by 12 min controlled by Chemstation software. All the

solvents used for HPLC were HPLC-grade. Column chromatography was performed as flash column chromatography using Merck silica gel 60 (15–40 mm) and a TELEDYNE ISCO Combiflash Rf 4x system (Flow rate: 5-200 ml/min, maximum pressure: 200 psi, column size: 4 gm to 40 gm: UV range: 254 to 360 nm). ^{1}H NMR spectra were recorded on a Bruker AC300, ARX300 and DRX 500 spectrometer at 300 MHz and 500 MHz respectively. The ^{13}C NMR spectra were recorded on a Bruker AC300, ARX300 and DRX 500 spectrometer at 75 MHz and 126 MHz respectively. Chemical shifts values were reported as % values (ppm) downfield from Me_4Si. Mass spectrometry was performed on a MAT 95 double focusing sector field MS and purity of all the compounds was determined by HPLC.

3.5.2 Synthesis of Ketobenzothiazole derivatives

3.5.2.1 Tert-butyl 1-(methoxy(methyl)amino)-1-oxopropan-2-ylcarbamate (11):

To a solution of Boc-Ala-OH (500 mg, 2.6 mmol) in dry DCM (20 ml) were added PyBOP (2.6 mmol, 1 eq), HN(Me)OMe•HCl (5.2 mmol, 2 eq) and 1.3 ml DIPEA (3eq) under argon atmosphere. The reaction mixture was stirred at room temperature for 2h. After the completion of reaction, the product was extracted in ethyl acetate and the organic layer was washed with 0.1 N HCl (2x), saturated $NaHCO_3$ (2x) and brine (1x). The product was dried with Na_2SO_4 and *in vacuo*. The Weinreb amide 11 was obtained as colorless solid (498 mg, 82%). **HPLC**: R_t = 2.56 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 5.23 (s, 1H), 4.66 (s, 1H), 3.75 (s, 3H), 3.19 (s, 3H), 1.42 (s, 9H), 1.29 (d, J = 6.9 Hz, 3H).

3.5.2.2 Tert-butyl 1-(benzo[d]thiazol-2-yl)-1-hydroxypropan-2-ylcarbamate (12):

A solution of *n*-butyllithium (1.6 M in hexane, 36 ml, 17.5 eq), was added dropwise to a stirred solution of benzothiazole (4.1 ml, 38.6 mmol, 18 eq) in 120 ml anhydrous THF cooled to -78 ^{o}C was under argon atmosphere. After complete addition, a dark red reaction mixture was obtained and continued to stir for 30 min at -78 ^{o}C. To this reaction mixture was added dropwise a solution of Weinreb amide 11 (498 mg, 2.1 mmol, 1 eq) in THF over 15 min at -78^{o}C. After the complete addition, the temperature of reaction mixture was immediately raised to -20^{o}C and continued to stir for another 2h at -20^{o}C. The reaction was quenched with saturated aqueous NH_4Cl (50 ml) and extracted with EtOAc (3 x 50 ml). The combined organic layers were washed with water (2x), brine (2x), dried with Na_2SO_4, filtered and concentrated *in vacuo*. The residue was dissolved in anhydrous MeOH (30 ml), cooled to -20^{o}C. To this stirred solution was added $NaBH_4$ (12.84 mmol, 6 eq) at -20^{o}C and after 1 h, reaction was quenched by addition of acetone (40 ml). The reaction mixture was stirred for another 15 min and then brought to room temperature. The mixture was concentrated *in vacuo* and partitioned between water and EtOAc. The aqueous layer was extracted with EtOAc (3x) and combined organic layers were washed with brine (2x), dried with Na_2SO_4, filtered and concentrated *in vacuo*. The crude residue was purified (EtOAc/Hexane 1:2) to give intermediate 12 as a yellow solid (271 mg, 40%). **HPLC**: R_t = 7.57 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 7.98 (t, J = 8.9 Hz, 1H), 7.93 – 7.85 (m,

1H), 7.50 – 7.42 (m, 1H), 7.41 – 7.34 (m, 1H), 5.17 (d, J = 2.6 Hz, 1H), 4.26 (s, 1H), 3.74 – 3.61 (m, 1H), 1.44 (s, 9H), 1.09 (dd, J = 19.9, 5.9 Hz, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 173.55, 155.76, 152.85, 134.91, 125.88, 124.82, 122.81, 121.65, 80.29, 75.77, 74.33, 52.14, 50.49, 28.30, 28.22, 28.10.

3.5.2.3 2-Amino-1-(benzo[d]thiazol-2-yl)propan-1-ol (13):

The intermediate 12 (250 mg, 0.81 mmol) was dissolved in 25 ml of TFA/DCM (1:4) and stirred for 2-3 h. After reaction completion, the mixture was concentrated *in vacuo*. The residue was diluted in DCM and concentrated *in vacuo* and under high vacuum to give the trifluoroacetate salt of amino alcohol 13 (80%). **HPLC**: R_t = 6.88 min. **^{1}H NMR** (500 MHz, MeOD) δ [ppm] = 8.06 – 7.93 (m, 2H), 7.53 (tdd, J = 8.4, 2.4, 1.2 Hz, 1H), 7.47 – 7.42 (m, 1H), 3.69 – 3.55 (m, 1H), 3.33 (dt, J = 3.2, 1.6 Hz, 1H), 1.34 – 1.25 (m, 2H), 1.20 (t, J = 6.2 Hz, 2H).**^{13}C NMR** (126 MHz, MeOD) δ [ppm] = 177.01, 176.47, 163.39, 163.11, 154.56, 154.25, 136.02, 135.77, 127.37, 126.47, 126.34, 123.66, 123.53, 123.08, 123.02, 119.11, 116.91, 75.07, 74.29, 53.06, 52.68, 17.21, 14.99. **MS** (EI, 70 eV): *m/z* = 332 [M].

3.5.2.4 Benzyl (S)-1-((S)-1-(benzo[d]thiazol-2-yl)-1-hydroxypropan-2-ylamino)-1-oxopropan-2-ylcarbamate 14a:

To a solution of Z-Ala-OH (86.2 mg, 0.36 mmol, 1 eq) was added EDAC (1 eq), HOBt (1.2 eq) in anhydrous DMF (3 ml) and stirred for 15 min. To this stirring solution was added Et_3N (1.5 eq) and residue 13 (1 eq) and reaction was continued overnight. After completion, reaction mixture was diluted with DCM (20 ml) and washed with 0.1 N HCl (3x), saturated $NaHCO_3$ (3x) and brine (1x). The organic layer was dried with Na_2SO_4, filtered and concentrated *in vacuo* to give compound 14a as a colorless solid (90 mg, 60%). **HPLC**: 91% R_t = 5.59 min (> 90% diastereomeric purity). **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.03 – 7.93 (m, 3H), 7.86 (dd, J = 15.7, 6.6 Hz, 1H), 7.48 – 7.36 (m, 1H), 7.35 – 7.27 (m, 5H), 5.21 – 4.92 (m, 3H), 4.64 – 4.41 (m, 1H), 4.27 – 4.09 (m, 1H), 2.93 (s, 3H), 2.86 (s, 3H).

3.5.2.5 Benzyl (S)-1-((S)-1-(benzo[d]thiazol-2-yl)-1-hydroxypropan-2-ylamino)-4-(methylthio)-1-oxobutan-2-ylcarbamate 14b:

To a solution of Z-Met-OH (41 mg, 1 eq) was added EDAC (1 eq), HOBt (1.2 eq) in anhydrous DMF (2 ml) and stirred for 15 min. To this stirring solution was added Et_3N (1.5 eq) and residue 13 (1 eq) and the reaction was continued overnight. After completion, reaction mixture was diluted with DCM (20 ml) and washed with 0.1 N HCl (3x), saturated $NaHCO_3$ (3x) and brine (1x). The organic layer was dried with Na_2SO_4, filtered and concentrated *in vacuo* to give compound 14b as a light brown solid (42 mg, 62 %). **HPLC**: R_t = 6.48 min (> 90% diastereomeric purity). **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 7.98 (d, J = 8.0 Hz, 1H), 7.86 (d, J = 7.9 Hz, 1H), 7.48 – 7.42 (m, 1H), 7.38 (dd, J = 11.2, 4.0

Hz, 1H), 7.31 (d, J = 18.5 Hz, 5H), 5.68 (d, J = 7.6 Hz, 1H), 5.22 – 4.94 (m, 3H), 4.54 (s, 1H), 4.32 (d, J = 5.7 Hz, 1H), 2.45 – 2.20 (m, 2H), 2.01 – 1.89 (m, 3H), 1.31 (dd, J = 15.2, 9.6 Hz, 3H), 1.25 (d, J = 9.2 Hz, 2H), 0.10 – 0.04 (m, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 174.34, 171.34, 155.91, 152.19, 135.98, 134.40, 128.38, 128.07, 127.92, 126.17, 125.09, 122.65, 121.76, 73.92, 66.98, 54.02, 50.55, 31.71, 29.65, 16.25, 14.85. **MS** (ESI, 70 eV): *m/z* = 496.2 [M+Na].

3.5.2.6 Benzyl (S)-1-((S)-1-(benzo[d]thiazol-2-yl)-1-oxopropan-2-ylamino)-1-oxopropan-2-ylcarbamate 15 (BSc5125):

A solution of compound 14a (0.2 mmol, 1 eq) in 2 ml of DMSO was treated with IBX (1.5 eq). The reaction mixture was stirred overnight and after completion diluted with 10 ml DCM. The product was extracted in organic layer and washed with with water (3x), saturated $NaHCO_3$ (3x) and brine (1x). The organic layer was dried with Na_2SO_4, filtered and concentrated *in vacuo* to give crude product which was purified by flash chromatography (EtOAc/Hexane 50:50 to 70:30) to give compound 15 was obtained as a offwhite solid (65 mg, 73 %). **HPLC**: R_t = 6.9 min, (> 90% diastereomeric purity). Mp. 175°C. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.22 – 8.16 (m, 1H), 8.01 – 7.95 (m, 1H), 7.62 – 7.51 (m, 2H), 7.39 – 7.28 (m, 5H), 5.84 – 5.73 (m, 1H), 5.43 (d, J = 20.1 Hz, 1H), 5.18 – 5.05 (m, 2H), 1.65 – 1.51 (m, 3H), 1.41 (dt, J = 8.6, 4.3 Hz, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 193.19, 171.64, 163.23, 153.33, 137.09, 136.09, 128.41, 128.35, 128.04, 127.96, 127.93, 127.83, 127.08, 125.68, 122.24, 66.99, 51.51, 50.42, 18.79, 18.70, 17.39. **MS** (ESI, 70 eV): *m/z* = 434.2 [M+Na].

3.5.2.7 Benzyl (2S)-1-((S)-1-(benzo[d]thiazol-2-yl)-1-oxopropan-2-ylamino)-4-(methylsulfinyl)-1-oxobutan-2-ylcarbamate 16 (BSc5123):

A solution of compound 14b (1 eq) in 2 ml of DMSO was treated with IBX (1.5 eq). The reaction mixture was stirred overnight and after completion diluted with 10 ml DCM. The product was extracted in organic layer and washed with water (3x), saturated $NaHCO_3$ (3x) and brine (1x). The organic layer was dried with Na_2SO_4, filtered and concentrated *in vacuo* to give crude product which was purified by flash chromatography (EtOAc/Hexane 50:50 to 70:30) to give compound 16 as a offwhite solid (75 %). Mp. 175°C. **HPLC**: R_t = 7.76 min (> 90% diastereomeric purity). **^{1}H NMR** (300 MHz, $CDCl_3$) δ [ppm] = 8.25 – 8.17 (m, 1H), 7.98 (dt, J = 6.3, 3.6 Hz, 1H), 7.65 – 7.49 (m, 2H), 7.37 – 7.32 (m, 4H), 5.86 – 5.57 (m, 2H), 5.13 (s, 2H), 2.59 (dt, J = 9.7, 5.2 Hz, 2H), 2.13 – 2.07 (m, 3H), 2.00 (dd, J = 14.2, 7.1 Hz, 1H), 1.62 (d, J = 7.2 Hz, 3H), 1.38 – 1.08 (m, 1H). **^{13}C NMR** (75 MHz, $CDCl_3$) δ [ppm] = 192.86, 170.39, 162.95, 155.74, 153.18, 136.90, 135.89, 128.22, 127.87, 126.89, 125.47, 121.96, 66.39, 53.30, 51.63, 31.60, 29.82, 18.18, 14.33. **MS** (ESI, 70 eV): *m/z* = 510.2 [M+Na].

3.5.2.8 (S)-Benzyl 1-(methoxy(methyl)amino)-1-oxopropan-2-ylcarbamate 17:

To a solution of cbz-Ala-OH (100 mg, 0.45 mmol, 1 eq) in dry DCM (5 ml) were added PyBOP (0.45 mmol, 1 eq), HN(Me)OMe•HCl (0.89 mmol, 2 eq) and 0.2 ml DIPEA (3 eq) under argon atmosphere. The reaction mixture was stirred for 2h at room temperature. After the completion of reaction, the product was extracted in ethyl acetate and organic layer was washed with 0.1 N HCl (2x), saturated $NaHCO_3$ (2x) and brine (1x). The product was dried with Na_2SO_4 and *in vacuo*. The Weinreb amide 17 was obtained as colorless solid (114 mg, 95%). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 173.46, 155.81, 136.57, 128.61, 128.19, 128.11, 66.85, 61.86, 60.62, 47.26, 46.66, 32.29, 26.47, 18.77.

3.5.2.9 (S)-benzyl 1-(benzo[d]thiazol-2-yl)-1-oxopropan-2-ylcarbamate 18 (BSc5124):

A solution of *n*-butyllithium (1.6 M in hexane, 17.5 eq), was added dropwise to a stirred solution of benzothiazole (18 eq) in 40 ml anhydrous THF cooled to -78°C was under argon atmosphere. After complete addition, a dark red reaction mixture was obtained and continued to stir for 30 min at -78°C. To this reaction mixture was added dropwise a solution of Weinreb amide 17 (100 mg, 0.37 mmol, 1 eq) in THF over 15 min at -78°C. After the complete addition, the temperature of reaction mixture was immediately raised to -20°C and continued to stir for another 2h at -20°C. The reaction was quenched with saturated aqueous NH_4Cl (50 ml) and extracted with EtOAc (3 x 50 ml). The combined organic layers were washed with water (2x), brine (2x), dried with Na_2SO_4, filtered and concentrated *in vacuo*. The residue was dissolved in anhydrous MeOH (10 ml), cooled to -20°C. To this stirred solution was added $NaBH_4$ (6 eq) at -20°C and after 1 h, reaction was quenched by addition of acetone (5 ml). The reaction mixture was stirred for another 15 min and then brought to room temperature. The mixture was concentrated *in vacuo* and partitioned between water and EtOAc. The aqueous layer was extracted with EtOAc (3x) and combined organic layers were washed with brine (2x), dried with Na_2SO_4, filtered and concentrated *in vacuo*. The crude residue was purified (EtOAc/Hexane 1:2) to give compound 18 as a brown solid (40%). Mp. 114°C. **HPLC**: R_t = 6.31 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.21 (d, J = 8.1 Hz, 1H), 7.99 (dd, J = 7.3, 0.8 Hz, 1H), 7.58 (dtd, J = 15.0, 7.2, 1.2 Hz, 2H), 7.34 (dd, J = 14.7, 5.8 Hz, 5H), 5.21 – 5.05 (m, 2H), 1.63 (d, J = 6.7 Hz, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 194.17, 163.81, 155.94, 153.94, 137.64, 136.72, 128.91, 128.53, 128.44, 127.57, 126.19, 122.77, 67.38, 53.45, 30.09, 19.78. **MS** (ESI, 70 eV): *m/z* = 363.2 [M+Na].

3.5.3 Synthesis of oxathiazol-2-one derivatives

3.5.3.1 Tert-butyl 1-amino-1-oxopropan-2-ylcarbamate 19:

To a solution of sodium carbonate (105 mg) in 1 ml of water was added alaninamide hydrochloride (0.4 mmol, 1 eq). In a separate flask, Di-tert-butyl dicarbonate (0.80 mmol, 2 eq) was dissolved in 1,4-dioxanes (0.8 ml) and this organic solution was added to the aqueous solution and allowed to stir for 15-16 h. After completion, the reaction mixture diluted with 1.75 ml of 1N HCl and extracted with EtOAc (4 x 5 ml). The organic layers were combined, washed with brine (1x), dried over Na_2SO_4, and

concentrated *in vacuo* to give a colorless solid BOC-protected carboxamide residue 19 (56%). **HPLC**: R_t = 1.06 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 1.53 – 1.50 (m, 3H), 1.44 (s, 9H).

3.5.3.2 Tert-butyl 1-(2-oxo-1,3,4-oxathiazol-5-yl)ethylcarbamate 20 (BSc5139):

A reaction mixture of BOC-protected carboxamide 19 (0.8 mmol, 1 eq) and chlorocarbonylsulfenyl chloride (1 mmol, 1.2 eq) in 40 ml of anhydrous toluene was refluxed for 2 h under argon atmosphere. The solvents were removed and crude residue was purified flash chromatography (Hexane/EtOAc, 80:20) to give intermediate 20 (BSc5139) as oily liquid (60 %). **HPLC**: R_t = 3.54 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 5.02 (s, 1H), 4.67 (d, J = 71.4 Hz, 1H), 1.49 (d, J = 7.1 Hz, 3H), 1.44 (s, 9H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 174.08, 161.96, 155.05, 81.03, 47.09, 28.62, 19.14.

3.5.3.3 5-(1-aminoethyl)-1,3,4-oxathiazol-2-one 21:

The intermediate 20 (0.15 mmol) was dissolved in 5 ml of TFA/DCM (1:4) and stirred for 2-3 h. After reaction completion, the mixture was concentrated *in vacuo*. The residue was diluted in DCM and concentrated *in vacuo* and under high vacuum to give trifluoroacetate salt of carboxamide 21 (84%). **HPLC**: R_t = 0.78 min. **^{1}H NMR** (500 MHz, MeOD) δ [ppm] = 4.20 – 4.04 (m, 1H), 1.65 (dd, J = 7.0, 0.6 Hz, 3H).

3.5.3.4 Benzyl (S)-1-oxo-1-((S)-1-(2-oxo-1,3,4-oxathiazol-5-yl)ethylamino)propan-2-ylcarbamate 22 (BSc5140):

To a solution of Z-Ala-OH (1 eq) was added EDAC (1 eq), HOBt (1.2 eq) in anhydrous DMF (2 ml) and stirred for 15 min. To this stirring solution was added Et_3N (1.5 eq) and residue 21 (1 eq) and reaction was continued for overnight. After completion, reaction mixture was diluted with DCM (20 ml) and washed with 0.1 N HCl (3x), saturated $NaHCO_3$ (3x) and brine (1x). The organic layer was dried with Na_2SO_4, filtered and concentrated *in vacuo* to give compound 22 (BSc5140) as a colorless solid (71 %). Mp. 113-115°C. **HPLC**: R_t = 5.01 min (> 90% diastereomeric purity). **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 7.38 – 7.29 (m, 5H), 5.32 (d, J = 7.0 Hz, 1H), 5.15 – 5.06 (m, 2H), 5.04 – 4.81 (m, 1H), 4.35 – 4.09 (m, 1H), 1.52 – 1.44 (m, 3H), 1.42 – 1.38 (m, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 174.06, 173.81, 172.56, 161.14, 136.08, 128.99, 128.93, 128.75, 128.61, 128.49, 77.16, 67.45, 57.00, 45.97, 18.75, 18.43. **MS** (ESI, 70 eV): *m/z* = 374.2 [M+Na].

3.5.3.5 Benzyl(S)-4-(methylthio)-1-oxo-1-((S)-1-(2-oxo-1,3,4-oxathiazol-5-yl)ethylamino)butan-2-ylcarbamate 23 (BSc5138):

To a solution of Z-Met-OH (1 eq) was added EDAC (1 eq), HOBt (1.2 eq) in anhydrous DMF (2 ml) and stirred for 15 min. To this stirring solution was added Et_3N (1.5 eq) and residue 21 (1 eq) and reaction was continued for overnight. After completion, reaction mixture was diluted with DCM (20

ml) and washed with 0.1 N HCl (3x), saturated $NaHCO_3$ (3x) and brine (1x). The organic layer was dried with Na_2SO_4, filtered and concentrated *in vacuo* to give compound 23 (BSc5138) as a colorless solid (70 %). Mp. 118°C. **HPLC**: R_t = 6.45 min (> 90% diastereomeric purity). **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 7.39 – 7.30 (m, 5H), 5.53 (d, J = 7.7 Hz, 1H), 5.11 (s, 2H), 5.05 – 4.95 (m, 1H), 4.41 (d, J = 6.5 Hz, 1H), 2.69 – 2.53 (m, 2H), 2.13 – 2.06 (m, 3H), 2.05 – 1.93 (m, 2H), 1.56 – 1.41 (m, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 172.78, 170.39, 160.09, 155.65, 135.48, 128.06, 127.80, 127.56, 66.67, 53.04, 44.84, 30.77, 29.48, 17.80, 14.63. **MS** (ESI, 70 eV): *m/z* = 845.2 [2M+Na].

3.5.4 Synthesis of Saccharin derivatives

3.5.4.1 N-chloromethyl saccharin or (chloromethyl)benzo[d]isothiazol-3(2H)-one 1,1-dioxide 37 (BSc5152) [133]

N-Chloromethyl saccharin 37 (BSc5152) was prepared from saccharin by reacting it first with formaldehyde to obtain N-(hydromethyl) saccharin and then reacting it with thionyl chloride at 0°C. A colorless crystalline solid (62%). Mp. 147°C. **HPLC**: 99%, R_t = 5.29 min. **^{1}H NMR** (300 MHz, $CDCl_3$) δ [ppm] = 8.15 – 8.10 (m, 1H), 7.99 – 7.85 (m, 3H), 5.58 (s, 2H). **^{13}C NMR** (75 MHz, $CDCl_3$) δ [ppm] = 158.11, 138.23, 136.24, 135.34, 127.02, 126.36, 121.89, 45.93. **MS** (EI, 70 eV): *m/z* = 232 [M].

3.5.4.2 N-cyanomethyl saccharin or 2-(1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)acetonitrile 38 (BSc5153) [134]

To a stirring solution of saccharin (500 mg, 1.0 eq) in 2.5 ml of DMF was added 415 mg of potassium carbonate and 45 mg of potassium iodide (10 mol %). Then, to this mixture 0.7 mL of chloroacetonitrile (4.0 eq) were added dropwise and the reaction stirred at 80°C for 48 h. After completion, the mixture was poured on ice and the resulting suspension was filtered to give crude residue which was purified by flash chromatography (Hexane/EtOAc, 50:50) to give compound 38 as a yellow solid (57%). Mp. 137°C. **HPLC**: R_t = 4.3 min. **^{1}H NMR** (500 MHz, DMSO) δ [ppm] = 8.40 – 8.38 (m, 1H), 8.20 – 8.17 (m, 1H), 8.12 (td, J = 7.7, 1.2 Hz, 1H), 8.05 (td, J = 7.6, 1.0 Hz, 1H), 5.09 (s, 2H).**^{13}C NMR** (126 MHz, DMSO) δ [ppm] = 158.08, 136.85, 136.58, 135.76, 125.86, 125.73, 122.11, 114.98, 25.58. **MS** (EI, 70 eV): *m/z* = 222 [M].

3.5.4.3 4-(((1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl)thio)benzoic acid 43 (BSc5158) [135]

A reaction mixture was prepared by dissolving N-chloromethyl saccharin 37 (1.09 mmol, 1 eq) and 4-mercaptobenzoic acid (1 eq) in anhydrous acetonitrile. This stirring solution was treated with DBU (2

eq) and reaction continued for overnight at room temperature. After reaction completion, the solvent was removed *in vacuo* and residue was taken up in 5 % sodium bicarbonate (10 ml) and extracted with EtOAc (2x). The aqueous layer was then acidified to pH 1 using 6 N HCl and precipitated solid was collected by filtration, gave compound 43 (BSc5158, 65%). Mp. 250°C**. HPLC**: R_t = 5.77 min. **^{1}H NMR** (500 MHz, DMSO) δ [ppm] = 8.31 (d, J = 7.7 Hz, 1H), 8.14 – 7.86 (m, 6H), 7.69 – 7.63 (m, 3H), 5.41 (s, 2H). **^{13}C NMR** (126 MHz, DMSO) δ [ppm] = 166.94, 158.29, 140.91, 139.38, 136.97, 136.36, 135.56, 130.44, 130.02, 129.93, 126.30, 125.86, 125.50, 121.78, 42.11. **MS** (EI, 70 eV): *m/z* = 349 [M].

3.5.4.4 Ethyl 4-(((1,1-dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl)thio)benzoate 44 (BSc5159) [135]

The sulfide 43 (0.29 mmol, 1 eq) was esterified by preparing a reaction mixture with dry ethanol (0.15 ml), DMF (0.36 ml), N,N'-Dicyclohexylcarbodiimide (DCC, 1eq) and 4-dimethylamino pyridine (4-DMAP, 0.13 eq). The mixture was stirred overnight and after completion, the reaction was poured into 5 ml of water and extracted with EtOAc (3 x 5 ml). The organic layers were combined, dried over Na_2SO_4 and concentrated *in vacuo* to give crude residue. This was purified by flash chromatography (Hexane/EtOAc, 80:20) to give compound 44 as an oily liquid (46%). Mp. 154°C. **HPLC**: R_t = 7.74 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.05 – 8.02 (m, 1H), 8.01 – 7.98 (m, 2H), 7.92 – 7.80 (m, 3H), 7.63 – 7.59 (m, 2H), 5.24 (s, 2H), 4.36 (q, J = 7.1 Hz, 2H), 1.38 (dd, J = 8.8, 5.5 Hz, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 166.41, 158.80, 139.42, 138.19, 135.58, 134.90, 131.24, 130.58, 130.16, 127.15, 125.89, 121.48, 61.47, 43.30, 14.70. **MS** (EI, 70 eV): *m/z* = 377 [M].

3.5.4.5 Ethyl 4-(((1,1-dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl)sulfonyl)benzoate 45 (BSc5160) [135]

The esterified sulfide 44 (0.08 mmol, 1 eq) was dissolved in dry DMF (2 ml) and treated with *m*-chloroperbenzoic acid (3 eq). The reaction mixture was stirred overnight at room temperature and after completion diluted with DCM (10 ml). The organic layer was washed with 5% aqueous $NaHCO_3$ (3x) and solvent was evaporated to give the compound 45 as a colorless solid (81%). Mp. 160°C. **HPLC**: R_t = 6.90 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.23 – 8.18 (m, 2H), 8.09 – 8.04 (m, 2H), 8.01 (dt, J = 7.6, 0.9 Hz, 1H), 7.97 – 7.83 (m, 3H), 5.08 (s, 2H), 4.41 (q, J = 7.1 Hz, 2H), 1.44 – 1.38 (m, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 164.79, 158.18, 140.53, 137.57, 135.89, 135.72, 134.72, 134.29, 130.22, 129.69, 129.27, 128.12, 125.83, 121.37, 61.75, 59.14, 14.10. **MS** (ESI, 70 eV): *m/z* = 432.2 [M+Na].

3.5.4.6 4-(((1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl)sulfonyl)benzoic acid 46 (BSc5161)[135]

The sulfide 43 (0.1 mmol, 1 eq) was dissolved in dry DMF (2 ml) and treated with *m*-chloroperbenzoic acid (3 eq). The reaction mixture was stirred overnight at room temperature and after completion diluted with DCM (10 ml). The solvent was evaporated to give the compound 46 as a colorless solid (90 %). Mp. 250-250°C. **HPLC**: R_t = 4.82 min. **^{1}H NMR** (500 MHz, DMSO) δ [ppm] = 8.35 (d, J = 7.7 Hz, 1H), 8.19 – 8.15 (m, 2H), 8.14 – 8.06 (m, 4H), 8.05 – 8.00 (m, 1H), 7.96 (s, 1H), 5.45 (s, 2H). **^{13}C NMR** (126 MHz, DMSO) δ [ppm] = 166.04, 162.33, 140.90, 136.67, 136.23, 136.07, 135.66, 130.39, 130.10, 129.20, 127.47, 125.70, 125.19, 121.97, 59.00. **MS** (ESI, 70 eV): *m/z* = 404.1 [M+Na].

General procedure for the synthesis of aryl acid derived N-methylated saccharins

A solution of N-chloromethyl saccharin (1eq.), carboxylic acid (2eq.) and triethylamine (2eq.) in 3 mL acetonitrile was refluxed for 3 h. The reaction mixture was cooled to room temperature and diluted with dichloromethane (5 mL). The solution was washed with water (2 X 5 mL) and saturated sodium bicarbonate (2 X 5 mL). The solution was dried with sodium sulfate and evaporated *in vacuo* to a crude product which was purified by flash column chromatography.

3.5.4.7 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl benzoate (39, BSc5154) [136]

A colorless crystalline solid (yield = 55%). Mp. 103-105°C. **HPLC:** 85%, R_t = 6.94 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.15 – 8.11 (m, 1H), 8.08 – 8.05 (m, 2H), 7.98 – 7.85 (m, 3H), 7.59 – 7.55 (m, 1H), 7.45 – 7.40 (m, 2H), 6.10 (s, 2H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 165.12, 158.09, 137.73, 135.41, 134.63, 134.47, 133.52, 129.95, 128.70, 128.36, 126.48, 125.69, 121.20, 121.13, 61.43. **MS** (ESI, 70 eV): *m/z* = 340.1 [M+Na].

3.5.4.8 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 3-methoxybenzoate 40 (BSc5155)

A colorless crystalline solid (yield = 74.3%). Mp. 140°C. **HPLC:** R_t = 7.17min. **^{1}H NMR** (300 MHz, $CDCl_3$) δ [ppm] = 8.13 (dd, J = 6.7, 1.3 Hz, 1H), 8.00 – 7.84 (m, 3H), 7.70 – 7.63 (m, 1H), 7.59 (dd, J = 2.6, 1.5 Hz, 1H), 7.38 – 7.29 (m, 1H), 7.11 (ddd, J = 8.3, 2.7, 1.0 Hz, 1H), 6.09 (s, 2H), 3.83 (s, 3H).**^{13}C NMR** (75 MHz, $CDCl_3$) δ [ppm] = 164.31, 158.89, 157.48, 137.10, 134.81, 133.87, 129.32, 128.79, 125.85, 125.09, 121.80, 120.53, 119.69, 113.56, 60.85, 54.75. **MS** (EI, 70 eV): *m/z* = 347 [M].

3.5.4.9 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 3-bromobenzoate 41 (BSc5156)

A colorless crystalline solid (yield = 97.4%). Mp. 144°C. **HPLC:** R_t = 7.7 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.20 (t, J = 1.7 Hz, 1H), 8.16 – 8.12 (m, 1H), 8.04 – 7.86 (m, 5H), 7.70 (ddd, J =

8.0, 2.0, 1.1 Hz, 1H), 7.32 (t, J = 7.9 Hz, 1H), 6.09 (s, 2H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 164.39, 158.59, 138.22, 137.03, 136.00, 135.05, 133.37, 131.14, 130.46, 129.05, 126.94, 126.27, 122.98, 121.69, 62.10. **MS** (EI, 70 eV): *m/z* = 396 [M].

3.5.4.10 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 4-methoxybenzoate 42 (BSc5157) [137].

An offwhite solid (yield = 84.6%). Mp. 138°C. **HPLC:** R_t = 7.1 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.13 (ddd, *J* = 7.5, 1.1, 0.7 Hz, 1H), 8.04 – 8.01 (m, 2H), 7.96 (ddd, *J* = 7.7, 1.2, 0.7 Hz, 1H), 7.92 (td, *J* = 7.5, 1.2 Hz, 1H), 7.87 (td, *J* = 7.5, 1.3 Hz, 1H), 6.92 – 6.89 (m, 2H), 6.07 (s, 2H), 3.85 (s, 3H).**^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 165.26, 164.78, 163.85, 158.11, 137.77, 135.36, 135.28, 134.44, 132.11, 126.54, 125.59, 121.11, 121.05, 113.67, 60.87, 55.11. **MS** (EI, 70 eV): *m/z* = 347 [M].

3.5.4.11 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 4-iodobenzoate 47 (BSc5187)

A colorless solid (yield = 70.1%). Mp. 167°C. **HPLC:** R_t = 8.04 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.14 (d, *J* = 7.5 Hz, 1H), 7.99 – 7.86 (m, 4H), 7.82 – 7.75 (m, 3H), 6.08 (s, 2H).**^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 165.20, 158.56, 138.29, 138.19, 136.04, 135.99, 135.14, 135.04, 131.79, 128.68, 126.92, 126.24, 121.67, 102.12, 62.02. **MS** (EI, 70 eV): *m/z* = 443 [M].

3.5.4.12 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 2-iodobenzoate 48 (BSc5188)

A colorless crystalline solid (yield = 64.8%). Mp. 125°C. **HPLC:** R_t = 7.67 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.16 – 8.11 (m, 1H), 8.02 – 7.85 (m, 5H), 7.38 (td, *J* = 7.7, 1.1 Hz, 1H), 7.15 (td, *J* = 7.7, 1.7 Hz, 1H), 6.09 (s, 2H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 164.67, 157.96, 141.51, 137.69, 135.54, 135.47, 134.53, 133.19, 131.65, 127.87, 126.43, 125.75, 121.17, 94.48, 61.75. **MS** (EI, 70 eV): *m/z* = 443 [M].

3.5.4.13 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 3-methylbenzoate 49 (BSc5189)

A colorless crystalline solid (yield = 57.4%). Mp. 105-110°C. **HPLC:** R_t = 7.39 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.13 (ddd, *J* = 9.5, 2.1, 1.7 Hz, 1H), 7.98 – 7.84 (m, 5H), 7.37 (dd, *J* = 7.6, 0.4 Hz, 1H), 7.31 (t, *J* = 7.6 Hz, 1H), 6.09 (s, 2H), 2.37 (s, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 165.29, 157.99, 138.17, 137.70, 135.39, 134.45, 134.29, 130.39, 128.58, 128.33, 127.10, 126.46, 125.65, 121.11, 61.35, 21.27. **MS** (EI, 70 eV): *m/z* = 331 [M].

3.5.4.14 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 3-iodobenzoate 50 (BSc5190)

A colorless crystalline solid (yield = 77.4%). Mp. 144°C. **HPLC:** R_t = 7.96 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.39 (t, *J* = 1.6 Hz, 1H), 8.14 (dd, *J* = 10.1, 3.7 Hz, 1H), 8.05 – 7.86 (m, 5H), 7.17 (t, *J* = 7.9 Hz, 1H), 6.08 (s, 2H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 164.21, 158.59, 142.90, 139.18, 138.19, 136.00, 135.04, 131.07, 130.54, 129.61, 126.86, 126.25, 121.68, 94.34, 61.86. **MS** (EI, 70 eV): *m/z* = 443 [M].

3.5.4.15 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 2-methylbenzoate 51 (BSc5191)

A colorless solid (yield = 61.6%). Mp. 106°C. **HPLC:** R_t = 7.38 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.13 (d, *J* = 7.6 Hz, 1H), 7.99 – 7.85 (m, 4H), 7.40 (t, *J* = 7.1 Hz, 1H), 7.25 – 7.20 (m, 2H), 6.07 (s, 2H), 2.63 (d, *J* = 5.5 Hz, 3H).**^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 166.28, 158.62, 141.49, 138.26, 136.05, 135.90, 135.14, 134.97, 133.08, 132.13, 128.38, 127.00, 126.20, 121.63, 61.37, 22.27. **MS** (EI, 70 eV): *m/z* = 331 [M].

3.5.4.16 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 4-methylbenzoate 52 (BSc5192) [137]

An offwhite solid (yield = 71.4%). Mp. 107°C. **HPLC:** R_t = 7.36 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.13 (dd, *J* = 7.6, 0.5 Hz, 1H), 7.99 – 7.85 (m, 5H), 7.24 – 7.20 (m, 2H), 6.08 (s, 2H), 2.39 (s, 3H).**^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 165.17, 158.11, 144.37, 143.92, 137.75, 135.38, 134.63, 134.45, 130.01, 129.08, 126.51, 125.95, 125.67, 121.12, 61.33, 21.79. **MS** (EI, 70 eV): *m/z* = 331 [M].

3.5.4.17 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 2-aminobenzoate 53 (BSc5193)

A yellow brown solid (yield = 31.1%). Mp. 110-115°C. **HPLC:** R_t = 6.68 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.13 (d, *J* = 7.5 Hz, 1H), 7.99 – 7.84 (m, 4H), 7.27 (d, *J* = 1.5 Hz, 1H), 7.24 (d, *J* = 1.5 Hz, 1H), 6.66 – 6.59 (m, 1H), 6.06 (s, 2H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 166.35, 158.12, 150.85, 137.73, 135.36, 134.74, 134.44, 131.49, 126.51, 125.66, 121.10, 116.52, 116.30, 109.12, 60.78. **MS** (EI, 70 eV): *m/z* = 332 [M].

3.5.4.18 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 3-chlorobenzoate 54 (BSc5194)

A light yellow solid (yield = 90.8%). Mp. 126°C. **HPLC:** R_t = 7.58 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.14 (d, *J* = 7.5 Hz, 1H), 8.04 (t, *J* = 1.8 Hz, 1H), 7.99 – 7.91 (m, 3H), 7.88 (td, *J* = 7.5, 1.3 Hz, 1H), 7.54 (ddd, *J* = 8.0, 2.1, 1.1 Hz, 1H), 7.38 (t, *J* = 7.9 Hz, 1H), 6.09 (s, 2H).**^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 163.99, 158.07, 137.68, 135.49, 134.58, 134.54, 133.59, 130.42, 129.93, 129.71, 128.08, 126.41, 125.74, 121.17, 61.33. **MS** (ESI, 70 eV): *m/z* = 374.2 [M+Na].

3.5.4.19 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 2-chlorobenzoate 55 (BSc5195) [137]

A colorless crystalline solid (yield = 54.2%). Mp. 123°C. **HPLC:** R_t = 7.25 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.13 (d, *J* = 7.5 Hz, 1H), 7.99 – 7.85 (m, 4H), 7.47 – 7.40 (m, 2H), 7.29 (ddt, *J* = 10.1, 4.8, 2.5 Hz, 1H), 6.09 (s, 2H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 163.78, 158.00, 137.67, 135.47, 134.52, 134.39, 133.15, 131.91, 131.12, 128.20, 126.51, 126.41, 125.72, 121.14, 61.69. **MS** (ESI, 70 eV): *m/z* = 374.2 [M+ Na].

3.5.4.20 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 2-(3-benzoylphenyl)propanoate 56 (BSc5196) [138]

A yellow compound (yield = 92%). Mp. 123-125°C. **HPLC:** R_t = 8.10 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.21 – 8.01 (m, 1H), 7.99 – 7.83 (m, 3H), 7.83 – 7.71 (m, 3H), 7.67 (ddd, *J* = 11.4, 6.5, 4.9 Hz, 1H), 7.63 – 7.53 (m, 2H), 7.52 – 7.39 (m, 3H), 5.93 – 5.78 (m, 2H), 3.86 (q, *J* = 7.2 Hz, 1H), 1.60 – 1.54 (m, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 196.78, 173.19, 158.48, 140.12, 138.39, 138.15, 137.90, 135.91, 134.95, 132.84, 132.04, 130.48, 129.74, 129.52, 129.01, 128.68, 126.87, 126.18, 121.59, 61.62, 45.46, 18.67**. MS** (ESI, 70 eV): *m/z* = 472.3 [M+ Na].

3.5.4.21 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 5-bromo-2-fluorobenzoate 57 (BSc5197)

An offwhite crystalline solid (yield = 80.17%). Mp. 140°C. **HPLC:** R_t = 7.60 min. **^{1}H NMR** (300 MHz, $CDCl_3$) δ [ppm] = 8.17 – 8.12 (m, 1H), 8.07 (dd, *J* = 6.3, 2.6 Hz, 1H), 8.00 – 7.85 (m, 3H), 7.63 (ddd, *J* = 8.8, 4.2, 2.6 Hz, 1H), 7.03 (dd, *J* = 10.0, 8.8 Hz, 1H), 6.08 (s, 2H). **^{13}C NMR** (75 MHz, $CDCl_3$) δ [ppm] = 163.02, 158.70, 138.66, 138.54, 138.32, 136.20, 135.40, 135.24, 127.05, 126.45, 121.87, 119.69, 119.38, 117.11, 62.27. **MS** (ESI, 70 eV): *m/z* = 438 [M+ Na].

3.5.4.22 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 3,4-dichlorobenzoate 58 (BSc5198)

A colorless solid (yield = 81.9%). Mp. 134°C. **HPLC:** R_t = 8.10 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.17 – 8.11 (m, 2H), 7.99 – 7.86 (m, 4H), 7.53 – 7.49 (m, 1H), 6.08 (s, 2H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 163.85, 158.55, 138.84, 138.15, 136.04, 135.08, 133.55, 132.28, 131.08, 129.45, 129.01, 126.86, 126.27, 121.69, 62.15. **MS** (ESI, 70 eV): *m/z* = 408.1[M+ Na].

3.5.4.23 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 3-cyanobenzoate 59 (BSc5199)

An offwhite solid (yield = 92.5%). Mp. 138°C. **HPLC:** R_t = 6.80 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.33 (d, *J* = 1.4 Hz, 1H), 8.31 – 8.26 (m, 1H), 8.15 (d, *J* = 7.5 Hz, 1H), 8.00 – 7.83 (m, 4H), 7.58 (t, *J* = 7.9 Hz, 1H), 6.11 (s, 2H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 163.75, 158.54, 138.12, 137.01, 136.11, 135.13, 134.41, 134.00, 130.57, 129.98, 126.83, 126.30, 121.72, 118.00, 113.60, 62.29 **MS** (ESI, 70 eV): *m/z* = 365 [M+ Na].

3.5.4.24 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 2-amino-3-methylbenzoate 60 (BSc5200)

A brown solid (yield = 66.9%). Mp. 120°C. **HPLC:** R_t = 7.17 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.13 (t, *J* = 6.5 Hz, 1H), 7.99 – 7.83 (m, 3H), 7.79 (dd, *J* = 8.2, 1.1 Hz, 1H), 7.19 (t, *J* = 7.2 Hz, 1H), 6.59 (dd, *J* = 15.3, 7.8 Hz, 2H), 6.06 (d, *J* = 6.3 Hz, 2H), 5.58 (s, 1H), 2.20 (d, *J* = 22.6 Hz, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 167.30, 158.64, 149.35, 138.24, 136.03, 135.86, 134.94, 129.96, 128.95, 126.16, 123.71, 121.60, 116.59, 109.49, 61.32, 17.78. **MS** (ESI, 70 eV): *m/z* = 369.2 [M+ Na].

3.5.4.25 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 3-bromo-5-nitrobenzoate 61 (BSc5201)

The corresponding 3-bromo-5-nitrobenzoic acid, required as an adduct was synthesized, as follows 3-nitrobenzoic acid (3 mmol, 1 eq) was added to 1.5 ml of concentrated sulfuric acid and heated to 60°C with stirring. To this solution, NBS (3.6 mmol, 1.2 eq) was added in three portions each in 15 min. After reaction completion, the mixture was poured on ice and precipitated solid was collected by filtration. The residue was washed with water (10 ml) and *n*-hexane (5ml), dried to get 3-bromo-5-nitrobenzoic acid (yield = 91%). **HPLC:** R_t = 4.94 min. **^{1}H NMR** (300 MHz, $CDCl_3$) δ [ppm] = 8.88 (dd, J = 2.1, 1.4 Hz, 1H), 8.65 – 8.62 (m, 1H), 8.58 – 8.55 (m, 1H). The title compound was synthesized by following procedure mentioned above aryl acid derived N-methylated saccharins. A light yellow solid (yield = 42%). Mp. 194-196°C. **HPLC:** R_t = 3.89 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.80 (dd, *J* = 2.1, 1.5 Hz, 1H), 8.57 (t, *J* = 2.0 Hz, 1H), 8.52 – 8.47 (m, 1H), 8.19 – 8.14 (m, 1H), 8.01 – 7.88 (m, 3H), 6.14 (s, 2H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 162.59, 158.54, 149.23, 138.88, 138.12, 136.18, 135.19, 132.37, 131.53, 126.80, 126.39, 123.96, 123.69, 121.77, 62.54. **MS** (EI, 70 eV): *m/z* = 441.1 [M].

3.5.4.26 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 3-(trifluoromethyl)benzoate 62 (BSc5202)

The corresponding 3-(trifluoromethyl)benzoic acid required as an adduct was synthesized by the oxidation of the respective aldehyde. The aldehyde (1 eq) was dissolved in DMF (0.2 M), and oxone (1 eq) was added. The mixture stirred at room temperature for 3 h and after completion EtOAc was added to extract the products and 1N HCl was used the dissolve the salts. The organic layer was washed with 1N HCl (3x) and brine(1x), dried over Na_2SO_4 and the solvent was removed under reduced pressure to obtain the 3-(trifluoromethyl)benzoic acid (yield = 85.3%). **HPLC:** R_t = 4.46 min. **^{1}H NMR** (300 MHz, $CDCl_3$) δ [ppm] = 8.39 (s, 1H), 8.31 (d, *J* = 7.9 Hz, 1H), 7.88 (d, *J* = 7.8 Hz, 1H), 7.64 (t, *J* = 7.8 Hz, 1H). **^{13}C NMR** (75 MHz, $CDCl_3$) δ [ppm] = 170.57, 133.52, 131.69, 131.25, 130.51, 130.33, 129.39, 127.34. The title compound was synthesized by following procedure mentioned above aryl acid derived N-methylated saccharins. An offwhite solid (yield = 71%). Mp. 140°C. **HPLC:** R_t = 7.6 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = ^{1}H NMR (500 MHz, CDCl3) δ 8.33 (s, 1H), 8.26 (d, *J* = 7.9 Hz, 1H), 8.17 – 8.13 (m, 1H), 8.00 – 7.81 (m, 4H), 7.59 (t, *J* = 7.8 Hz, 1H), 6.12 (s, 2H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 164.42, 158.54, 138.16, 135.99, 135.06, 133.59, 131.77, 131.45, 130.51, 130.10, 129.57, 127.28, 126.79, 126.24, 121.66, 61.80. **MS** (ESI, 70 eV): *m/z* = 408.3 [M+ Na].

3.5.4.27 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 3-nitrobenzoate 63 (BSc5203)

An offwhite solid (yield = 83%). Mp. 168°C. **HPLC:** R_t = 7.23 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.91 – 8.83 (m, 1H), 8.46 – 8.35 (m, 2H), 8.15 (dd, *J* = 4.2, 3.8 Hz, 1H), 8.01 – 7.86 (m, 3H), 7.66 (t, *J* = 8.0 Hz, 1H), 6.14 (s, 2H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 162.79, 157.65, 147.87, 137.24, 135.21, 134.24, 130.11, 129.31, 127.56, 125.95, 125.43, 124.51, 120.83, 61.46, 29.18. MS (ESI, 70 eV): *m/z* = 385.1[M+ Na].

3.5.4.28 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 3-(dimethylamino)benzoateate 64 (BSc5204)

A light brown solid (yield = 75.1%). Mp. 125°C. **HPLC:** R_t = 5.02 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.17 – 8.12 (m, 1H), 8.01 – 7.86 (m, 3H), 7.49 – 7.39 (m, 2H), 7.32 – 7.25 (m, 1H), 6.93 (dd, *J* = 8.3, 2.5 Hz, 1H), 6.11 (s, 2H), 3.01 (s, 6H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 165.81, 158.08, 150.34, 137.74, 135.34, 134.42, 129.27, 128.97, 126.50, 125.65, 121.09, 117.89, 117.34, 113.47, 61.39, 40.36, 40.36. **MS** (EI, 70 eV): *m/z* = 360 [M].

3.5.4.29 (1,1-Dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 2-naphthoate 65 (BSc5205)

A colorless crystalline solid (yield = 84%). Mp. 155-158°C. **HPLC:** R_t = 8.01 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.65 (s, 1H), 8.18 – 8.12 (m, 1H), 8.10 – 8.04 (m, 1H), 8.01 – 7.82 (m, 6H), 7.56 (dddd, *J* = 18.4, 8.1, 6.9, 1.3 Hz, 2H), 6.16 (s, 2H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 165.55, 158.41, 137.93, 135.97, 135.67, 134.73, 132.51, 132.04, 129.64, 128.73, 128.45, 127.90, 126.87, 126.71, 126.11, 125.94, 125.39, 121.39, 61.73. **MS** (EI, 70 eV): *m/z* = 367 [M].

3.5.4.30 (1,1-dioxido-3-oxobenzo[d]isothiazol-2(3H)-yl)methyl 3-aminobenzoate 66 (BSc5206)

A light brown solid (yield = 80.9%). Mp. 118-120°C. **HPLC:** R_t = 3.23 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.13 (d, J = 7.5 Hz, 1H), 7.91 (qdd, J = 8.7, 7.3, 2.6 Hz, 3H), 7.50 – 7.40 (m, 1H), 7.36 – 7.30 (m, 1H), 7.18 (dt, J = 12.6, 7.9 Hz, 1H), 6.91 – 6.79 (m, 1H), 6.14 – 6.00 (m, 2H), 5.17 – 4.78 (m, 1H), 3.61 – 3.33 (m, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 165.80, 158.62, 146.94, 138.18, 135.90, 134.97, 130.10, 129.67, 126.98, 126.18, 121.63, 120.62, 120.50, 116.44, 61.88. **MS** (EI, 70 eV): *m/z* = 332 [M].

3.5.5 Synthesis of Benzoxazinone derivatives

General procedure for the synthesis of 2-substituted benzoxazinones: Under argon atmosphere, anthranilic acid (1 eq) was dissolved in 10 ml of dry pyridine at room temperature and then the solution was cooled to 0°C. To this solution was slowly added the respective acyl/benzoyl chloride (2 eq) dissolved in pyridine. After the complete addition, the reaction mixture was stirred at 0°C for 1 h and then at room temperature for 2 h. After completion, the mixture was neutralized with $NaHCO_3$ solution and then poured onto ice water. The precipitate formed was collected by filtration and dried to give compounds 67 and 69-79.

Further, a reaction mixture was prepared with anthranilic acid (1.46 mmol, 1 eq) and acetic anhydride (6 eq). Then this was subjected to a microwave reaction at 130°C for 3 min. After the reaction completion, remaining acetic anhydride was evaporated *in vacuo* and under high vacuum to give crude residue. This was recrystallized using EtOAc/MeOH (5:1) to obtain compound 68.

3.5.5.1 2-Phenyl-4H-benzo[d][1,3]oxazin-4-one 67 (BSc5170) [139]

An offwhite solid (yield = 73.2%). Mp. 122-123°C. **HPLC:** 99%, R_t = 7.01 min. **^{1}H NMR** (300 MHz, $CDCl_3$) δ [ppm] = 8.37 – 8.28 (m, 2H), 8.24 (dd, J = 7.9, 1.4 Hz, 1H), 7.88 – 7.77 (m, 1H), 7.69 (d, J = 8.0 Hz, 1H), 7.62 – 7.45 (m, 4H). **^{13}C NMR** (75 MHz, $CDCl_3$) δ [ppm] = 159.69, 157.28, 147.14,

136.68, 132.75, 130.40, 128.88, 128.74, 128.62, 128.46, 128.38, 127.37, 117.18. **MS** (EI, 70 eV): *m/z* = 223 [M].

3.5.5.2 2-Methyl-4H-benzo[d][1,3]oxazin-4-one 68 (BSc5171) [139]

A light brown crystalline solid (yield = 74%). Mp. **HPLC:** R_t = 2.1 min. **^{1}H NMR** (300 MHz, MeOD) δ [ppm] = 8.50 (dd, J = 8.4, 0.8 Hz, 1H), 8.04 (dd, J = 8.0, 1.6 Hz, 1H), 7.54 – 7.47 (m, 1H), 7.23 – 7.06 (m, 1H), 2.17 (s, 3H). **^{13}C NMR** (75 MHz, MeOD) δ [ppm] = 171.41, 171.26, 142.43, 135.13, 132.46, 123.96, 121.41, 116.70, 25.17. **MS** (EI, 70 eV): *m/z* = 161 [M].

3.5.5.3 (E)-2-(2-chlorostyryl)-4H-benzo[d][1,3]oxazin-4-one 69 (BSc5172) [140]

A light yellow solid (yield = 37.3%). Mp. 145°C. **HPLC:** R_t = 7.9 min. **^{1}H NMR** (300 MHz, $CDCl_3$) δ [ppm] = 8.29 – 8.18 (m, 2H), 7.87 – 7.75 (m, 1H), 7.69 (dt, J = 8.2, 3.3 Hz, 1H), 7.62 (d, J = 8.0 Hz, 1H), 7.56 – 7.47 (m, 1H), 7.44 (dt, J = 5.8, 3.3 Hz, 1H), 7.37 – 7.27 (m, 2H), 6.80 (d, J = 16.1 Hz, 1H). **^{13}C NMR** (75 MHz, $CDCl_3$) δ [ppm] = 159.67, 157.61, 147.47, 138.30, 136.97, 135.55, 133.55, 131.59, 130.87, 129.28, 129.01, 128.06, 127.75, 127.63, 122.15, 117.68. **MS** (EI, 70 eV): *m/z* = 283 [M].

3.5.5.4 2-(4-Methoxyphenyl)-4H-benzo[d][1,3]oxazin-4-one 70 (BSc5173) [141]

An offwhite solid (yield = 25%). Mp. 151°C. **HPLC:** R_t = 7.14 min. **^{1}H NMR** (300 MHz, $CDCl_3$) δ [ppm] = 8.24 (ddd, J = 13.1, 7.5, 1.6 Hz, 3H), 7.87 – 7.74 (m, 1H), 7.64 (d, J = 7.8 Hz, 1H), 7.50 – 7.41 (m, 1H), 7.00 (d, J = 9.0 Hz, 2H), 3.89 (s, 3H). **^{13}C NMR** (75 MHz, $CDCl_3$) δ [ppm] = 163.44, 159.91, 157.33, 147.35, 136.45, 130.43, 128.69, 127.83, 127.06, 122.64, 116.79, 114.30, 55.50. **MS** (EI, 70 eV): *m/z* = 253 [M].

3.5.5.5 (E)-2-Styryl-4H-benzo[d][1,3]oxazin-4-one 71 (BSc5174) [139]

A light yellow solid (yield = 83%). Mp. 149°C. **HPLC:** R_t = 7.4 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.22 (dd, J = 7.9, 1.3 Hz, 1H), 7.86 (d, J = 16.1 Hz, 1H), 7.83 – 7.77 (m, 1H), 7.65 – 7.56 (m, 3H), 7.49 (tt, J = 7.5, 3.8 Hz, 1H), 7.46 – 7.36 (m, 3H), 6.80 (d, J = 16.1 Hz, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = δ 159.66, 157.84, 147.50, 142.56, 136.98, 135.10, 130.77, 129.44, 129.08, 128.61, 128.45, 127.30, 119.24, 117.38. **MS** (ESI, 70 eV): *m/z* = 272.1 [M+Na].

3.5.5.6 2-(Furan-2-yl)-4H-benzo[d][1,3]oxazin-4-one 72 (BSc5175) [139]

A light yellow solid (yield = 58.7%). Mp. 108-110°C. **HPLC:** R_t = 5.34 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 8.21 (dd, J = 7.9, 1.3 Hz, 1H), 7.86 – 7.77 (m, 1H), 7.76 – 7.66 (m, 2H), 7.54 – 7.45 (m, 1H), 7.36 (dd, J = 3.5, 0.6 Hz, 1H), 6.62 (dd, J = 3.5, 1.7 Hz, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = 158.46, 149.70, 146.93, 146.61, 144.36, 136.65, 128.65, 128.14, 127.05, 117.07, 116.85, 112.43. **MS** (ESI, 70 eV): *m/z* = 236 [M+Na].

3.5.5.7 2-(3-Bromophenyl)-4H-benzo[d][1,3]oxazin-4-one 73 (BSc5176) [142]

A light yellow solid (yield = 71.4%). Mp. 158°C. **HPLC:** R_t = 8.05 min. **^{1}H NMR** (300 MHz, $CDCl_3$) δ [ppm] = 8.46 (t, J = 1.8 Hz, 1H), 8.28 – 8.19 (m, 2H), 7.89 – 7.79 (m, 1H), 7.74 – 7.65 (m, 2H), 7.59 – 7.49 (m, 1H), 7.39 (t, J = 7.9 Hz, 1H). **^{13}C NMR** (75 MHz, $CDCl_3$) δ [ppm] = 159.88, 156.51, 147.48, 137.54, 136.33, 133.07, 132.05, 131.09, 129.54, 129.50, 128.20, 127.64, 123.77, 117.84. **MS** (EI, 70 eV): *m/z* = 301[M].

3.5.5.8 2-(4-Fluorophenyl)-4H-benzo[d][1,3]oxazin-4-one 74 (BSc5177) [143]

An offwhite solid (yield = 71.5%). Mp. 177°C. **HPLC:** 98% R_t = 7.15 min. **^{1}H NMR** (300 MHz, $CDCl_3$) δ [ppm] = 8.36 – 8.26 (m, 2H), 8.22 (dd, J = 7.9, 1.4 Hz, 1H), 7.86 – 7.77 (m, 1H), 7.66 (d, J = 8.1 Hz, 1H), 7.56 – 7.46 (m, 1H), 7.24 – 7.11 (m, 2H). **^{13}C NMR** (75 MHz, $CDCl_3$) δ [ppm] = 166.98, 163.61, 159.03, 155.94, 146.59, 136.32, 130.49, 130.37, 128.34, 127.98, 126.85, 116.57, 116.47, 115.84. **MS** (EI, 70 eV): *m/z* = 241 [M].

3.5.5.9 2-(2-Fluorophenyl)-4H-benzo[d][1,3]oxazin-4-one 75 (BSc5178) [144]

A colorless solid (yield = 52.7%). Mp. 113-115°C. **HPLC:** R_t = 6.70 min. **^{1}H NMR** (300 MHz, $CDCl_3$) δ [ppm] = 8.26 (dd, J = 7.9, 1.2 Hz, 1H), 8.12 (td, J = 7.6, 1.8 Hz, 1H), 7.89 – 7.78 (m, 1H), 7.72 (dd, J = 8.1, 0.6 Hz, 1H), 7.61 – 7.49 (m, 2H), 7.29 (dt, J = 6.5, 3.3 Hz, 1H), 7.23 – 7.17 (m, 1H). **^{13}C NMR** (75 MHz, $CDCl_3$) δ [ppm] = 163.27, 159.82, 159.31, 155.04, 146.84, 136.73, 134.18, 134.06, 131.28, 128.89, 128.51, 127.58, 124.47, 124.42, 119.27, 119.15, 117.55, 117.26, 117.16. **MS** (EI, 70 eV): *m/z* = 241 [M].

3.5.5.10 2-(3-Fluorophenyl)-4H-benzo[d][1,3]oxazin-4-one 76 (BSc5179) [145]

A colorless solid (yield = 68.9%). Mp. 150°C. **HPLC:** R_t = 7.25 min. **^{1}H NMR** (300 MHz, $CDCl_3$) δ [ppm] = 8.28 – 8.21 (m, 1H), 8.13 – 8.06 (m, 1H), 8.01 (ddd, J = 9.7, 2.5, 1.6 Hz, 1H), 7.84 (ddd, J = 8.2, 7.3, 1.6 Hz, 1H), 7.70 (dd, J = 8.1, 0.6 Hz, 1H), 7.59 – 7.43 (m, 2H), 7.26 (s, 1H). **^{13}C NMR** (75 MHz, $CDCl_3$) δ [ppm] = 164.44, 161.24, 159.10, 146.67, 136.67, 132.54, 132.43, 130.43, 130.33,

128.68, 128.62, 127.35, 124.01, 123.97, 119.78, 119.50, 117.08, 115.39, 115.08. **MS** (EI, 70 eV): *m/z* = 241 [M].

3.5.5.11 (E)-2-(4-Methoxystyryl)-4H-benzo[d][1,3]oxazin-4-one 77 (BSc5207) [146]

The corresponding cinnamic acid chloride, (E)-4-Methoxycinnamic acid chloride, required as an adduct was synthesized. (E)-4-Methoxycinnamic acid (1 eq) was dissolved in MeCN and treated with $SOCl_2$ (3 eq) to yield the corresponding acid chloride (93%). **^{1}H NMR** (300 MHz, MeOD) δ [ppm] = 5.91 – 5.70 (m, 2H), 5.23 – 5.10 (m, 2H), 4.66 – 4.47 (m, 1H), 2.12 – 1.95 (m, 3H). Then the title compound was synthesized by following procedure mentioned above in *general procedure* for the synthesis of 2-substituted benzoxazinones. A colorless solid (yield = 46%). **HPLC:** R_t = 6.60 min. **^{1}H NMR** (500 MHz, $CDCl_3$) δ [ppm] = 7.66 (dd, J = 20.9, 12.4 Hz, 1H), 7.53 – 7.41 (m, 2H), 6.97 – 6.80 (m, 2H), 6.31 (d, J = 15.9 Hz, 1H), 3.83 (d, J = 4.4 Hz, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ [ppm] = δ 167.22, 161.24, 145.37, 144.12, 131.95, 129.57, 127.21, 115.71, 114.23, 113.37, 55.27. **MS** (EI, 70 eV): *m/z* = 279 [M].

3.5.5.12 (E)-2-(3,4-difluorostyryl)-4H-benzo[d][1,3]oxazin-4-one 78 (BSc5208)

The corresponding acid chloride, (E)-3-(3,4-difluorophenyl)acryloyl chloride, required as an intermediate was synthesized, as follows: (E)-3-(3,4-Difluorophenyl)acrylic acid (1 eq) was dissolved in MeCN and treated with $SOCl_2$ (3 eq) to yield the corresponding acid chloride (93%). **^{1}H NMR** (300 MHz, $CDCl_3$) δ [ppm] = 7.78 (dd, J = 15.8, 4.9 Hz, 1H), 7.49 – 7.27 (m, 3H), 6.60 (d, J = 15.6 Hz, 1H). Then the title compound was synthesized by following procedure mentioned above in *general procedure* for the synthesis of 2-substituted benzoxazinones**.** A colorless solid (yield = 57%). Mp. 150-152^{o}C. **HPLC:** R_t = 6.68 min. **^{1}H NMR** (300 MHz, MeOD) δ [ppm] = 8.60 (dd, J = 8.3, 1.0 Hz, 1H), 8.14 – 8.04 (m, 1H), 7.66 – 7.51 (m, 2H), 7.43 (ddd, J = 8.4, 6.1, 1.7 Hz, 2H), 7.35 – 7.22 (m, 1H), 7.15 – 6.99 (m, 1H), 6.73 (d, J = 15.7 Hz, 1H). **^{13}C NMR** (75 MHz, MeOD) δ [ppm] = 174.35, 166.44, 154.70, 154.52, 154.10, 153.92, 151.21, 150.86, 150.65, 142.12, 140.60, 134.50, 133.41, 133.11, 126.68, 125.61, 124.43, 124.14, 121.41, 119.60, 119.30, 118.00, 117.77. **MS** (EI, 70 eV): *m/z* = 285 [M].

3.5.5.13 (E)-2-methoxy-5-(2-(4-oxo-4H-benzo[d][1,3]oxazin-2-yl)vinyl)phenyl propionate 79 (BSc5209)

The corresponding caffeic acid chloride, (E)-5-(3-chloro-3-oxoprop-1-en-1-yl)-2-methoxyphenyl propionate, required as an intermediate was synthesized, as follows: A solution was prepared by dissolving 100 mg of 4-methoxy caffeic acid (1 eq) in 1 ml of tetrahydrofuran, and then the solution was placed in an ice bath. This was followed by addition of triethylamine (3 eq) and acetyl chloride (3

eq) to this solution. The reaction mixture was stirred for overnight and after completion, the solvent was removed *in vacuo*. The residue was dissolved in DCM (20 ml), washed with water (3x), dried with Na_2SO_4 and concentrated in vacuo to obtain intermediate acid. The intermediate compound (1 eq) was dissolved in MeCN and treated with $SOCl_2$ (3 eq) to yield the corresponding acid chloride (73.8%). **^{1}H NMR** (300 MHz, MeOD) δ [ppm] = 7.56 – 7.23 (m, 2H), 7.06 – 6.81 (m, 2H), 6.34 – 6.13 (m, 1H), 3.84 – 3.69 (m, 3H), 3.21 (dt, J = 3.3, 1.6 Hz, 2H), 1.19 – 0.94 (m, 3H). Then the title compound was synthesized following the procedure mentioned above in *general procedure* for the synthesis of 2-substituted benzoxazinones**.** A colorless solid (yield = 42%). Mp. 140-145°C. **HPLC:** R_t = 6.88 min. **^{1}H NMR** (300 MHz, $CDCl_3$) δ [ppm] = 8.87 (d, J = 8.0 Hz, 1H), 8.13 (dd, J = 8.0, 1.5 Hz, 1H), 7.85 – 7.51 (m, 2H), 7.47 – 7.29 (m, 2H), 7.12 (t, J = 7.2 Hz, 1H), 7.03 – 6.91 (m, 2H), 6.59 – 6.28 (m, 1H), 3.86 (dd, J = 7.6, 4.1 Hz, 3H), 2.72 – 2.55 (m, 2H), 1.25 (t, J = 5.0 Hz, 3H). **^{13}C NMR** (75 MHz, $CDCl_3$) δ [ppm] = 172.68, 164.84, 153.52, 152.85, 146.31, 146.08, 142.45, 141.59, 140.27, 135.58, 56.18, 27.51, 9.37. **MS** (EI, 70 eV): *m/z* = 351 [M].

4 Results

4.1 Thesis workflow

As outlined in section 1.6, the objective of this thesis was to identify novel small molecule inhibitors for rhomboid proteases, which would overcome the drawbacks of already known inhibitors with respect to potency and selectivity. Rational drug design (section 1.4) is a well-established method in the search of new enzyme inhibitors. However, in the present work, a computer-aided approach was applied in order to identify novel rhomboid inhibitors (Figure 4.1).

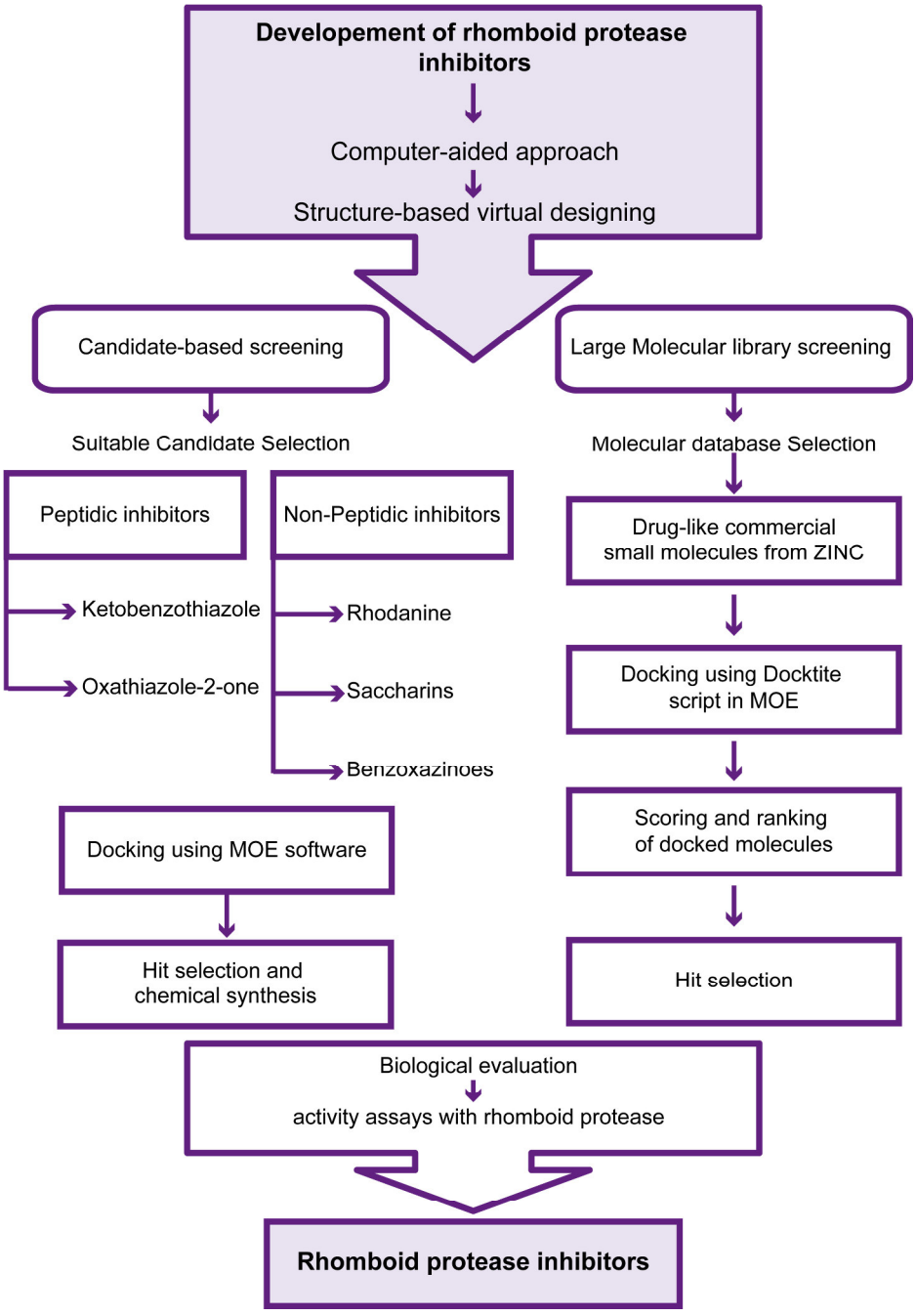


Figure 4.1: Schematic representation of design strategies for rhomboid protease inhibitors in this thesis.

Virtual screening is an automated computer-aided technique, which helps to screen drug-like candidate molecules against specific target proteins. All modeling experiments employed structure-based virtual screening (section 1.4) with the software Molecular Operating Environment (MOE), and the cocrystal structure of the *E. coli* rhomboid protease GlpG in complex with the peptidic phosphonofluoridate inhibitor CAPF (PDB: 3UBB). In this study, virtual screening was implemented in two ways: 1) candidate-based virtual screening, and 2) large molecular library screening (Figure 4.1). In the candidate approach, suitable candidates were selected based on the available literature about serine proteases inhibitors, and both peptidic and non-peptidic inhibitors were designed. For the peptidic inhibitors, the scaffolds ketobenzothiazole and oxathiazole-2-one were chosen. For the non-peptidic inhibitors, the small molecular candidates rhodanines, saccharins and benzooxazinones were selected. The detailed strategies for the design of these inhibitors are reported in section 4.5 and 4.6. Subsequently, a molecular library with all relevant compounds was prepared using MOE followed by docking experiments into the rhomboid protease pocket. The docked molecules were scored and ranked based on their ligand efficiency, scoring function and ligand interactions (section 3.4.4). Finally, the best candidates were selected as hits and chemical synthesis was performed.

Large molecular library virtual screening involves the automatic evaluation of very large libraries of molecules using molecular modeling programs. In present study, a huge compound library of commercially available small drug like molecules (size = 2,419,472) was screened against the rhomboid GlpG. Modelling experiments were executed using DOCKTITE script in MOE [131]. The method is described in section 3.4.3.3 and all strategical details are provided in section 4.7. Docking experiments were followed by scoring and ranking similar to the candidate-based approach, and the best molecules were selected as hits. This was followed by the biological evaluation, in which all hit molecules from both approaches were tested in cell-free and cell-based rhomboid activity assays to confirm their activity as rhomboid protease inhibitors.

4.2 Purification of intramembrane proteins GlpG, Gurken substrate and AqRho

In this study, the inhibitor activity of all chemical compounds was determined by an *in vitro* rhomboid activity assay. This assay consisted of the *E. coli* rhomboid GlpG and chimeric substrate proteins incorporating either the transmembrane domain of the natural rhomboid substrate Gurken or of the non-natural substrate LacY (section 1.5) [103]. Besides GlpG, a second rhomboid protease from *Aquifex aeolicus* (AaROM) was employed to confirm the activity of novel inhibitors and to investigate their selectivity [102]. These experiments required sufficient amounts of pure enzyme and substrate. For this purpose, different molecular biology techniques and methods were used as detailed below.

4.2.1 Recombinant expression of the rhomboid GlpG

For the production of the recombinant rhomboid, the *E. coli* strain BL21 (DE3) C43 was transformed with the expression vector pET41b (+) – GlpG (3.1.3) and precultured overnight in 50 ml LB medium at 37°C (3.2.2). On the following day, the preculture was added into 2L of LB medium, and the culture was grown at 37°C until it reached an OD_{600} of 0.5-0.8. Subsequently, recombinant protein expression was induced with 1 mM IPTG for 16-18 h at 18°C (3.2.3). To check for the expression of the recombinant protein, samples before and after induction were collected and analysed by Western blotting using an antibody against the C-terminally fused His-tag (3.2.11). The Western blot analysis (Figure 4.2) showed the appearance of a protein band of approximately 33 kDa after IPTG induction, confirming the successful expression of the GlpG rhomboid.

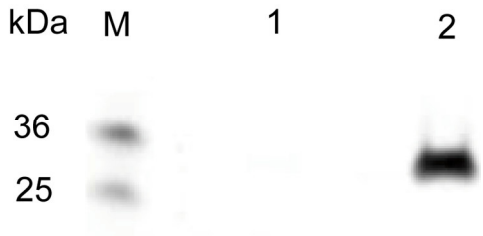


Figure 4.2: Western blot analysis of recombinant GlpG expression. Anti-His immunodetection of GlpG expression. Lane 1: before induction of recombinant protein expression with IPTG. Lane 2: after induction a protein band of approximately 33 kDa was observed that corresponded to the predicted molecular weight of the GlpG rhomboid.

4.2.2 Recombinant expression of the Gurken substrate

The Gurken or LacY derived chimeric substrates with an N-terminal MBP-tag and a C-terminal His-tag (MBP-TMD-TrX-His6) were expressed similarly to the rhomboid GlpG. However, these substrate proteins were transformed into a GlpG knockout strain of *E. coli* to ensure that the endogenous GlpG would not cleave the recombinant substrate proteins. Expression of the substrate proteins was induced by adding 1 mM IPTG at 37°C for 4 h. Western blotting analysis with an antibody against the MBP-tag for Gurken and the His-tag for LacY were used to confirm successful expression of the substrate proteins (Figure 4.3).

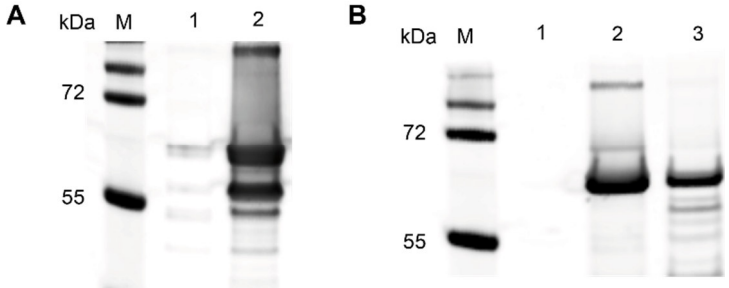


Figure 4.3: Western blot analysis of recombinant Gurken and recombinant LacY substrate expression. A) Anti-MBP immunodetection of Gurken expression. Lane 1: before induction of recombinant protein expression with IPTG. Lane 2: after induction a protein band of approximately 66 kDa was observed that corresponded to the predicted molecular weight of the Gurken substrate. B) A) Anti-His immunodetection of LacY expression. Lane 1: before induction of recombinant protein expression with IPTG. Lane 2: Purified Gurken substrate. Lane 3: after induction a protein band of approximately 66 kDa was observed similar to Gurken corresponded to the LacY substrate. Smaller proteins or protein fragments were also observed but were not co-purified in the subsequent affinity chromatography step (see below).

4.2.3 Purification of recombinant proteins by immobilized metal ion affinity chromatography (IMAC)

In the next step, the recombinantly expressed GlpG, Gurken and LacY proteins were purified by IMAC. Bacteria were harvested, lysed by three passes through a nitrogen bomb (3.2.4), and cellular membranes were prepared by differential centrifugation (3.2.5). Subsequently, the membranes were solubilized in 1.5% DDM (3.2.6), and the solution was subjected to IMAC and loaded on a column. Column-bound proteins were eluted using an imidazole gradient (3.2.6). The collected IMAC elution fractions were analysed by SDS-PAGE (3.2.9) and the results are summarised in Figure 4.4. All the fractions containing the desired proteins, GlpG (17-23), Gurken (17-23) and LacY (14-17) were pooled and pursued for protein concentration (3.2.8). As a result, reasonably pure enzyme and substrate proteins were generated to perform activity assays.

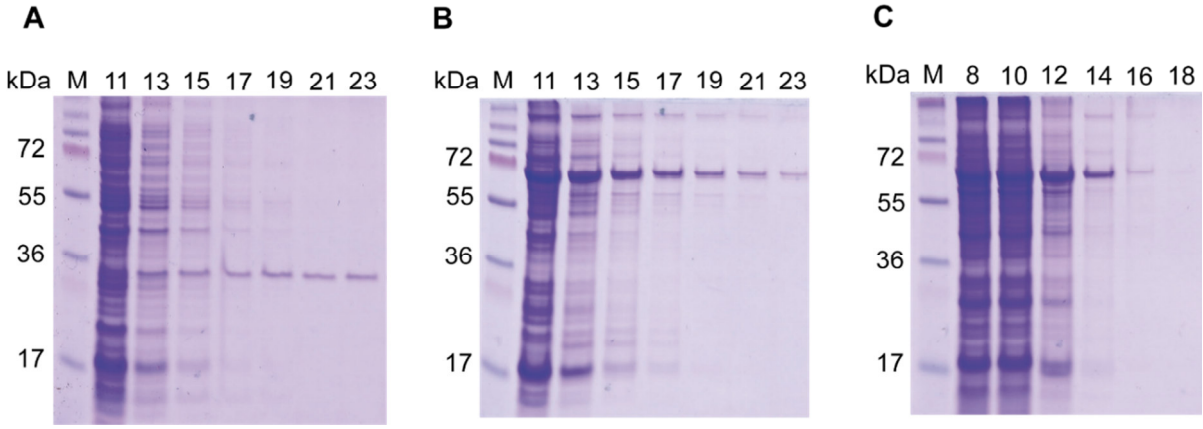


Figure 4.4: SDS-PAGE analysis of the IMAC elution fractions. A) Elution fractions of the GlpG rhomboid protease. B) Elution fractions of the Gurken substrate. C) Elution fractions of the LacY substrate. M: molecular weight protein marker. Fractions containing the desired proteins, GlpG (17-23), Gurken (17-23) and LacY (14-17) were pooled and pursued for protein concentration.

4.2.4 Recombinant expression and purification of the *Aquifex aeolicus* rhomboid (AaROM)

For the production of the N-terminal-GST-tagged fusion protein AaROM, the *E. coli* strain BL21 (DE3) C43 was transformed with the expression vector pGEX-6P-1 – AaROM and bacteria were cultured as described. Recombinant protein expression was induced with 50 µM IPTG (3.2.3), and AaROM was expressed for 16-18 h at 22°C. AaROM protein expression was analysed with an anti-GST antibody by western blotting (3.2.11). After induction with IPTG, a protein band of approximately 40 kDa was observed that corresponded to the molecular weight of the GST-fused AaROM (Figure 4.5A, lane 2). Afterwards bacterial membranes were prepared and solubilized in DDM as described above (3.2.4 and 3.2.5), followed by affinity chromatography to purify the GST-AaROM (3.2.7). For this, solubilized membrane preparations were incubated with 2 ml of glutathione-sepharose overnight at 4°C. The following day, the resin was washed with 10-column volumes of buffer. The Western blot in Figure 4.5A (lanes 4-6) and the silver stained SDS-gel in Figure 4.5B (lanes 3-7) show the flow-through fractions containing unbound GST-AaROM. Next, the bound GST-

AaROM was eluted using GST-tagged PreScission protease (Amersham Pharmacia Biotech) (3.2.7). Lane 7 in Figure 4.5A and lane 8 in Figure 4.5B display the GST-PreScission bound to the column. The eluted rhomboid was again incubated with 0.5 ml of glutathione-sepharose in order to completely eliminate unbound GST-PreScission (Figure 4.5A, lane 8; and Figure 4.5B, lane 9). Lane 9 of the Western blot in Figure 4.5A contained the purified AaROM but the band is not visible as the AaROM is no attached to the GST-tag. However, the purification of AaROM was verified through silver staining (Figure 4.5B, lane 10) with a single protein band of approximately 20 kDa corresponding to the expected molecular weight of AaROM.

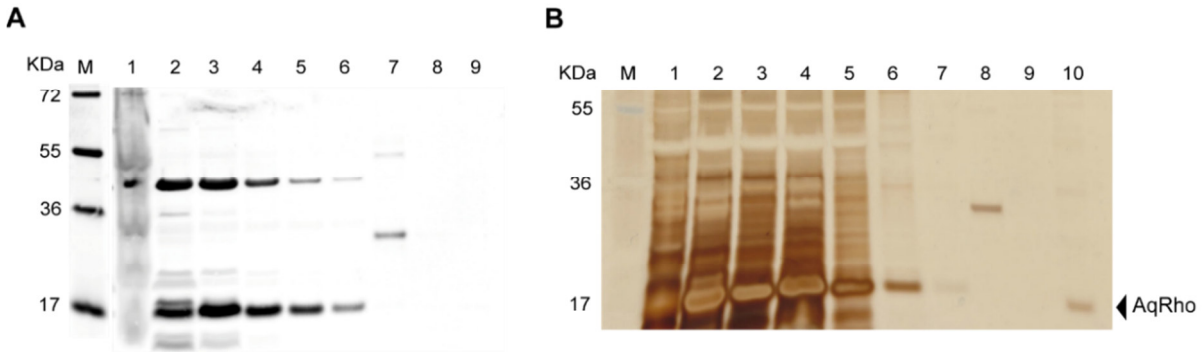


Figure 4.5: Expression and purification of AaROM. A) Western blot analysis with an anti-GST antibody Lane 1: before induction of recombinant protein expression with IPTG. Lane 2: after induction. Lane 3: solubilized membrane fraction containing GST-AaROM (~ 40 kDa). Lanes 4-6: flow-through of column washing steps. Lane 7: GST-PreScission protease bound to the column. Lane 8: new glutathione-sepharose column showing no remaining GST-PreScission protease. Lane 9: purified AaROM band is not visible in the Western blot since eluted AaROM is not tagged with GST anymore. B) Silver stained SDS-gel: Lane 1: after induction of recombinant protein expression with IPTG. Lane 2: solubilized membrane fraction containing GST-AaROM (~ 40 kDa). Lanes 3-7: flow-through of column washing steps. Lane 8: GST-PreScission protease bound to the column. Lane 9: new glutathione-sepharose column showing no remaining GST-PreScission protease. Lane 10: a single silver stained protein band corresponded to the purified AaROM (~ 20 kDa).

4.3 GlpG *in vitro* activity assay

To analyse the activity of the purified bacterial rhomboid GlpG, approximately 0.35 µM of the enzyme were incubated with the Gurken substrate as described in section EcGlpG *in vitro* activity assay 3.2.11. As a result, the Gurken substrate was cleaved by GlpG and an N-terminal cleavage fragment (NTF) and a C-terminal cleavage fragment (CTF) were observed in the SDS-PAGE gel (Figure 4.6A). In further experiments, the substrate concentration and the reaction time were optimized with respect to the fixed GlpG concentration.

4.3.1 Optimization of Gurken substrate concentration for the GlpG *in vitro* activity assay

To optimize the Gurken substrate concentration for the GlpG *in vitro* activity assay, increasing concentrations of Gurken from 0-3 µM (Figure 4.6) were incubated with 0.35 µM GlpG. The reaction mixture was incubated for 2 h, the reaction products were analysed by SDS-PAGE, and the protein bands were visualized by coomassie staining. Figure 4.6A shows the full-length Gurken band and the resulting NTF and CTF bands after the catalytic reaction. Since the NTF band was more clearly visible

and more easy to quantify, the optimization of substrate concentration for further experiments was based on the NTF band intensity. The NTF and full-length Gurken bands were quantified using ImageJ, and percent of the maximal NTF band intensity was plotted versus the Gurken concentration in GraphPad Prism (Figure 4.6B). This showed that the band intensity of the NTF reached a plateau at a concentration of 1.8 µM Gurken and did not increase further with higher Gurken concentrations. Hence, a concentration of 1.8 µM Gurken substrate was selected for further experiments. Similar calculations for the CTF showed that substrate turnover was not saturated at this concentration of Gurken substrate (Figure 4.6C).

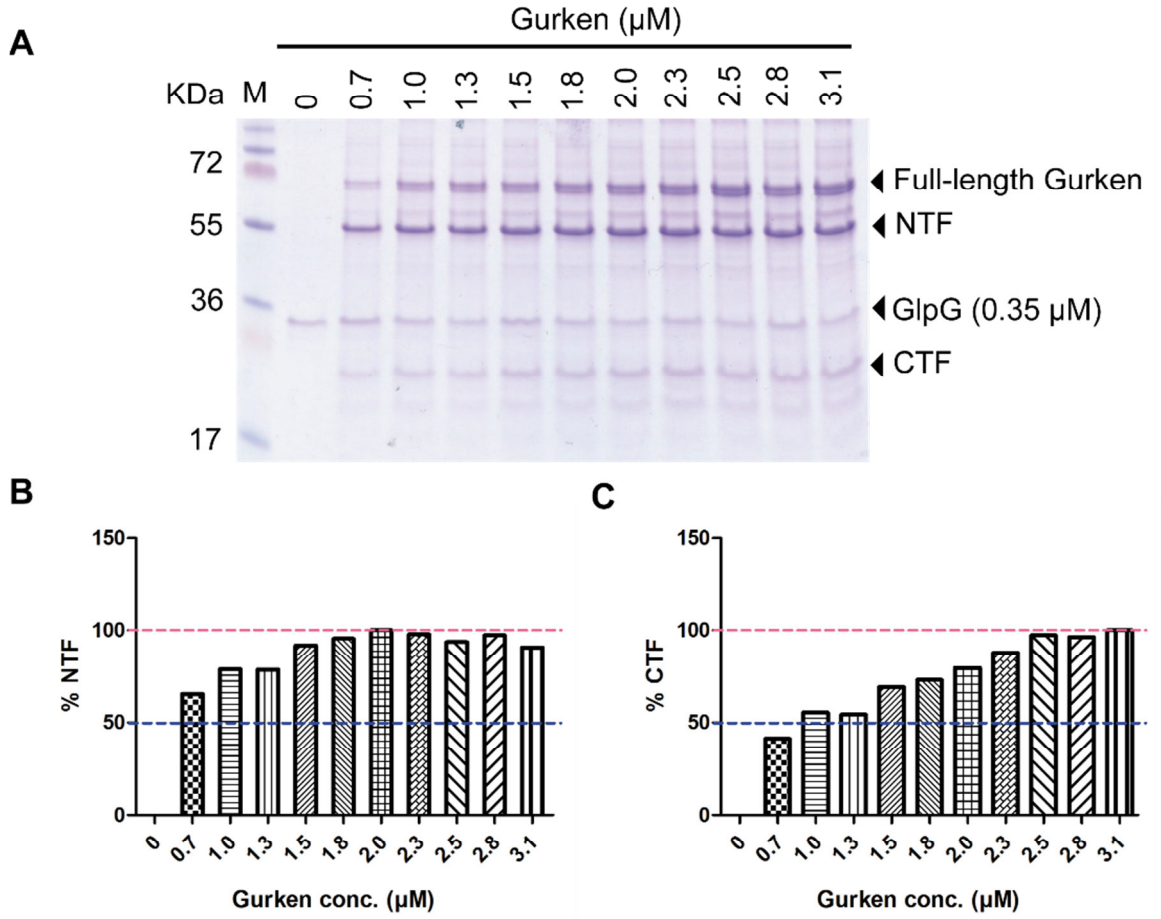


Figure 4.6: Optimization of Gurken substrate concentration for the *in vitro* activity assay. A) SDS-PAGE showing increasing Gurken concentrations (0-3.1 µM) incubated with the GlpG enzyme (0.35 µM). B) Plot showing % N-terminal fragment (NTF) versus Gurken concentration. Gel bands were quantified using ImageJ and plotted in GraphPad Prism. This showed that the NTF band intensity reached a maximum with a Gurken concentration of 1.8 µM. The dashed line in red indicates 100% NTF band intensity and the blue line indicates 50% NTF band intensity. The gel was stained with coomassie brilliant blue. C) Similar calculations were performed for the CTF and showed that CTF generation was not saturated at a Gurken concentration of 1.8 µM.

4.3.2 Optimization of the reaction time of the GlpG *in vitro* activity assay

Optimization of the reaction time of the *in vitro* activity assay was also important to ensure that turnover of the substrate as measured by the band intensity of the NTF was not saturated. Thus, a reaction mixture was prepared containing the optimized concentrations of Gurken substrate (1.8 µM)

and GlpG (0.35 µM). Each reaction mixture (volume 15 µl) was incubated at 37°C, and the reaction samples were collected at different time points from 15 min to overnight incubation (ON). The collected reaction samples were analysed by SDS-Page gel electrophoresis (Figure 4.7A). The intensities of the NTF and the remaining full-length Gurken bands were quantified using ImageJ and plotted in GraphPad Prism. The plot in Figure 4.7C shows the % remaining full-length Gurken substrate versus the reaction time. A gradual decrease in the intensity of the full-length Gurken band with increasing incubation time from 15 min to overnight was observed. The second plot in Figure 4.7B shows the % NTF versus the reaction time, and here a gradual increase in the reaction product was observed for increasing incubation time from 15 min to overnight. The dashed red line indicates the reaction saturation phase defined as the maximal NTF band intensity (100%) after overnight incubation. Overall, these experiments showed a near linear increase in NTF band intensity between 1-5 h of reaction time with no saturation. 1.5 h was selected as the optimized reaction time for further activity assays.

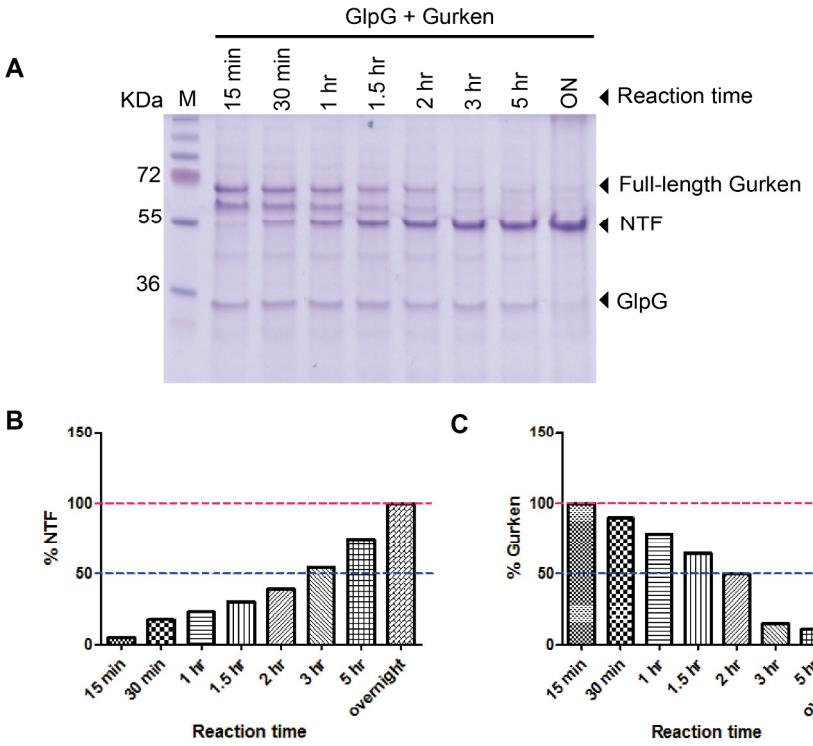


Figure 4.7: Optimization of the reaction time of the *in vitro* activity assay. A) SDS-PAGE gel showing the reaction products with variable incubation time from 15 min to overnight incubation. The assay contained 1.8 µM Gurken substrate and 0.35 µM GlpG. B) Plot with % NTF band intensity versus the reaction time displaying a gradual increase in the amount of NTF with increasing incubation time. The red dashed line indicates the maximal (100%) and blue dashed line half maximal (50%) NTF band intensity after overnight incubation. C) Plot with % remaining full-length Gurken substrate versus the reaction time. Gel bands were quantified using ImageJ and plotted in GraphPad Prism. The gel was stained with coomassie brilliant blue.

4.3.3 Validation of GlpG *in vitro* activity assay with known inhibitors

In order to validate the GlpG *in vitro* activity assay with the optimized substrate concentration and reaction time, activity assays were performed with three known inhibitors: a beta-lactam inhibitor (L16) [103] and two isocoumarin inhibitors (DCI and JLK-6) (Figure 1.23) [40, 87]. For this, GlpG

(0.35 µM) was preincubated with defined concentrations of inhibitors (0.1 µM - 250 µM) for 30 min at 37°C (section 3.2.12). This was followed by the addition of Gurken substrate (1.8 µM), and the reaction was continued for 90 min. As explained in method section 3.2.12, the reaction was set up to evaluate the percentage activity of GlpG in the presence and absence of inhibitors. After completion of the reactions, an SDS-PAGE was run to analyse the reaction products. Figure 4.8 shows the activity assay performed with inhibitor L16 and the quantification of the gel using ImageJ. The percentage of inhibition of rhomboid by L16 is shown in the plot of L16 concentration versus % NTF (Figure 4.8B). 100 % in Figure 4.8B indicates the NTF band intensity at maximal enzyme activity without inhibitor (DMSO vehicle control), and the dashed green line indicates 50 % inhibition. A nonlinear regression curve to fit the data of the log (inhibitor concentration) versus the NTF band intensity (variable slope) was generated in GraphPad Prism to determine the IC50 values (Figure 4.8C). The IC50 value for L16 was determined to be approximately 0.9 µM, which was comparable to data in the literature [103]. IC50 values of the two isocoumarin inhibitors were 6 µM for JLK-6 and 19 µM for DCI (data not shown). Subsequently, the GlpG activity assay was used to evaluate the inhibition potency of unknown inhibitors (see below).

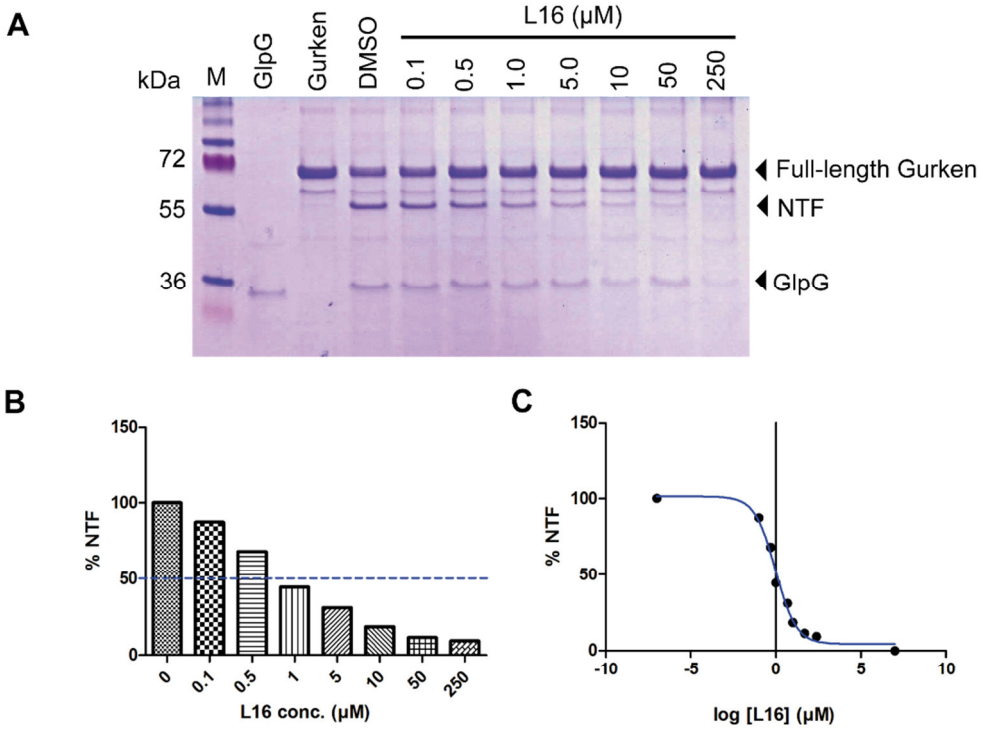


Figure 4.8: GlpG *in vitro* activity assay with a beta-lactam inhibitor (L16). A) SDS-PAGE of the reaction products after incubation of GlpG and Gurken substrate with the beta-lactam inhibitor L16 in increasing concentrations (0.1 µM - 250 µM). The arrows indicate the full-length Gurken substrate (~ 66 kDa), the N-terminal Gurken fragment (NTF, ~ 55 kDa) and GlpG (~ 33 kDa). M: Molecular weight marker. B) Plot with L16 concentration versus % NTF band intensity. The dashed blue line indicates 50% enzyme inhibition. C) Nonlinear regression curve of the log (L16 concentration) versus % NTF band intensity. The gel was stained with coomassie brilliant blue, quantified with ImageJ, and the values were plotted in GraphPad Prism.

4.4 Validation of the MOE docking methods

As explained in section 3.4, molecular docking in the present studies was performed using the X-ray co-ordinates of the co-crystallized structure of the *E. coli* rhomboid protease GlpG with the phosphatase inhibitor CAPF (PDB: 3UBB). To validate the DOCK and ConflexDock methods in MOE, the inhibitor CAPF was redocked into GlpG. The methods were performed as explained in section 3.4: this involved receptor preparation (section 3.4.1) and docking using the DOCK (section 3.4.3.1) and ConflexDOCK workflows (section 3.4.3.2). The results of the redocking experiments (Figure 4.9) for both workflows showed that the redocked CAPF in its best energy minimized conformation aligned well with the position of CAPF in the GlpG pocket as found in the co-crystallized structure (PDB: 3UBB).

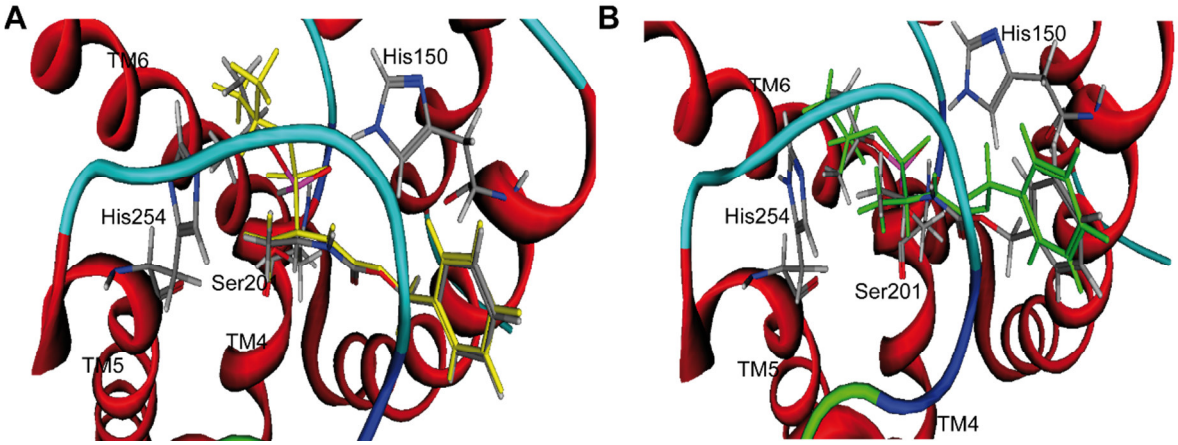


Figure 4.9: Validation of the MOE docking methods. A) CAPF redocked into the co-crystallized structure of the *E. coli* rhomboid protease GlpG with CAPF (PDB: 3UBB) using the DOCK workflow. Energy minimized redocked CAPF (yellow) aligned in position with the original CAPF in the co-crystallized structure (grey colour). B) CAPF redocked with the ConflexDock workflow. Again, energy minimized redocked CAPF (green) aligned in position with the original CAPF (grey colour).

4.5 Design, synthesis and biological evaluation of peptidic inhibitors

As explained in section 1.3.1, peptides are a rational starting point in the design of enzyme inhibitors. One of the proven strategies to create low molecular weight peptide inhibitors is mimicking the natural substrate sequence, so-called peptidomimetic compounds. Generally, pharmacophores based on the natural substrate of the enzyme will be selected due to their high specificity, affinity and ability to interact with the target enzyme. In our study, the same strategy was employed along with a molecular modeling approach. The design process began by developing structure-activity relationships (SAR) that would define a minimal active sequence or a major pharmacophore element, and by identifying the key residues responsible for the biological effect.

4.5.1 Keto-benzothiazole based peptide-mimetic inhibitors

4.5.1.1 Design and docking studies of ketobenzothiazole derivatives

Hypothesis

A peptide chain coupled to a C-terminal benzothiazole (Figure 4.10) was chosen as a lead structure for the first peptidomimetic series. The hypothesis behind this lead structure was that the benzothiazole moiety would act as a serine trap towards the rhomboid protease [111]. As described above in section 1.3.1, ketobenzothiazole bearing peptidomimetics were previously reported as slow, tight binding inhibitors of matriptase, with the Arg-Gln-Ala-Arg (RQAR) peptide sequence mimicking the P1–P4 substrate recognition sequence of the enzyme (Figure 1.18) [111]. Generally, these inhibitors are reversible, covalent binders of the enzyme active site. However, the tetra-peptide amino acid sequence of the matriptase inhibitor cannot be adopted for rhomboids as the natural substrates of these proteases diverge from each other. Figure 4.10 shows the basic structure of a ketobenzothiazole inhibitor with a peptidyl chain coupled to the benzothiazole serine trap at the C-terminus and to a carboxybenzyl (cbz) protecting group at the N-terminus.

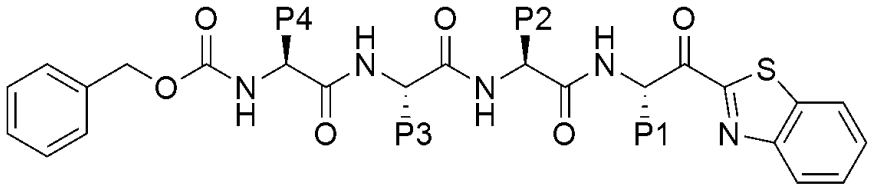


Figure 4.10: A tetra-peptide as a lead compound for peptidomimetic rhomboid inhibitors.

The major drawback for the design of a peptidemimetic inhibitors was that there were no previously published peptide-based rhomboid inhibitors until 2014 and, consequently, no co-crystal structures. Hence, the major problems in designing these inhibitors were: (1) to locate the position of the substrate sequence in the rhomboid subsites, (2) to find an adequate length of the peptide chain which would contribute to rhomboid inhibition, and (3) to identify the ideal amino acid sequence that would fit best to the respective positions and to ensure selectivity over other proteases. Nevertheless, some previous findings supported the design of peptidomimetic inhibitors. First, the 'oxyanion hole' of the enzyme active site (mentioned in section 1.2.3.2 and shown in Figure 1.11) is known to play a key role in inhibitor binding and in ligand positioning as demonstrated in several known GlpG-inhibitor complexes [86, 87, 89]. Secondly, the location of S1 and S1' subsites had been determined from co-crystal structures of rhomboid in complex with the Cbz-AlaP(OiPr)F or CAPF inhibitors (Figure 4.11) [89]. Third, the pattern of the tetra peptide-benzothiazole inhibitor interacting with the S1-S4 binding sites of matriptase [111] was also a critical clue for the design of peptidomimetic rhomboid inhibitors. Hence, these were all important features to consider while analysing the position of peptidic ligands in the rhomboid binding cleft.

One surprising observation from the co-crystal GlpG-CAPF complex structure was that the amino acid residue alanine of the CAPF inhibitor was found to point towards the S1' pocket [89]. However, considering the orientation of the peptide residues P1-P4 in the case of matriptase [111], these peptide residues should have pointed towards the S pocket, in the opposite direction to what has been observed with CAPF. Hence, I needed a lead peptidic compound present within the rhomboid pocket for our modeling experiments. Based on above described facts, my starting point scaffold (Figure 4.12) consisted of a di-peptide with L-Ala at P1 and L-Met at P2 position adopted from Gurken, which is a natural protein substrate for the *Drosophila* rhomboid and also known to be cleaved by the *E. coli* rhomboid GlpG [103]. This construct was used as basis to find peptidomimetic inhibitors.

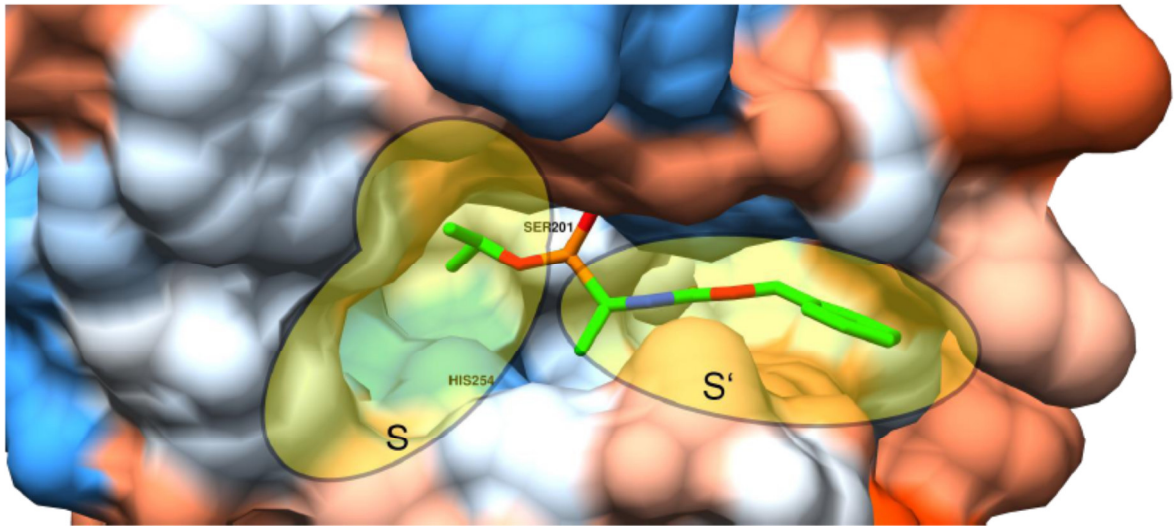


Figure 4.11: Top view of the rhomboid GlpG-CAPF complex (PDB: 3UBB). The structure shows the subsites S (N-terminal) and S' (C-terminal from the cleavage site) depicted in yellow, and the orientation of the phosphate inhibitor CAPF in the enzyme's binding pocket. The isopropyl moiety points into the S binding pocket whereas the rest of the inhibitor occupies the S' site.

Creating a starting point

In this study, the molecular modeling experiments were based on the co-crystal structure of the *E. coli* rhomboid GlpG and the phosphate inhibitor Cbz-AlaP(OiPr)F or CAPF (PDB: 3UBB). This ligand inhibits the protease irreversibly by covalent binding to the catalytically active Ser201 in the active site [89]. The receptor was prepared using the LigX interface (section 3.4.1). Then, to create a starting point, CAPF was manually exchanged with the L-Ala-L-Met dipeptide sequence using the builder tool in MOE and positioned roughly into the S1 and S2 sites. Since the major goal of the docking experiments was to find the best peptide sequence, the benzothiazole group was removed to minimize the computational costs. This decision was also supported by the fact that the only possibility for the benzothiazole group to rotate was around one single bond and that there were only weak degrees of freedom in this part of the molecule. In addition, the N-terminal protecting group was also not included in the starting molecule as it is dispensable after the synthesis.

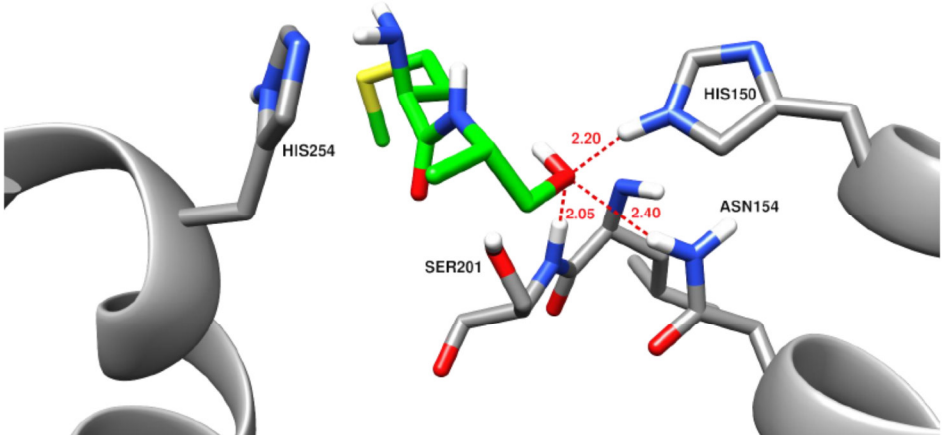


Figure 4.12: The starting point dipeptide scaffold used for the molecular modeling experiments in the rhomboid pocket. Energy minimized conformation of the dipeptide inhibitor showing the oxyanion hole structure at the hemiacetalic oxygen. The ligand is depicted in green and the receptor in grey. Hydrogen bonds (red dotted lines) are formed by the imidazole ring system of His150, the amide group of Asn154, and the backbone nitrogen of Ser201. Bond lengths were automatically calculated by the Chimera software.

Afterwards, the covalent bond between the enzyme and the peptide was cleaved, and a search for the best conformation was performed. Default parameters of the LowModeMD method were chosen for the conformational search. The selection of the best conformation was based on the assumption that the hemiacetalic oxygen of the inhibitor (formed through nucleophilic attack by Ser201) should have been positioned in a way that it may engage in any of the three important hydrogen bonds of the oxyanion hole (Figure 1.11). Water molecules were deleted from the best docking result with a score of S = - 7.05. The molecule conformation depicted in Figure 4.12 shows the final dipeptide-benzothiazole. This was followed by the evaluation of different amino acids for the P1 to P4 positions, and by exploring whether an inhibitor with four amino acids or less/more would fit best into the rhomboid cavity.

Docking strategy

In this thesis, a procedure for the covalent docking of a peptidomimetic inhibitor to the bacterial rhomboid protease GlpG from *E. coli* was developed. A splitting strategy was used for the inhibitor to reduce the computational costs. Moreover, the procedure ensured a step-by-step optimization of each single peptide amino acid as well as the optimization of the peptide length. All molecule databases were prepared using the Builder module, followed by energy minimization in MOE (section 3.4.2). Since ketobenzothiazole peptidomimetic derivatives are likely to form a covalent bond with the receptor molecule, the docking experiments were performed with the ConflexDock script in MOE 2013 using the AMBER12: EHT forcefield (section 3.4.3). The ConflexDock script in MOE was designed for covalent inhibitors (section 3.4.3.2). The carbon atom of the inhibitor and the oxygen atom of Ser201 forming the covalent bond were fixed, and the carbon atom of the inhibitor was labelled ANC, while the rest of the ligand was allowed to move. The options for the script were set as follows: poses = 1; bond rotation = 0°. Other parameters were set to default values. The

conflexdock.svl script computes an energy minimization for each conformation that is docked into the active site of the enzyme. However, a major concern with this script was that it was not able to keep planar systems as rigid bodies. Peptide bonds as well as sp2 hybridized atoms in the ring systems were often found bent out of their plane after docking. To avoid this problem a modified version of the conflexdock script that lacked the energy minimization step was used (named ConflexDock_womm). In order to find the best fitting residue in the P1 position, a set of four amino acids was docked with Met derived from the Gurken peptide in the P2 position to gain some insight into the positioning of the main chain. The same was done for the P2 position, with the best amino acid derived from the docking results in the P1 position. Once an energy minimized conformation was found, the chain length was raised to four. A search for the third amino acid was performed by keeping P2 and P1 fixed. To finish the sequence optimization of the tetrapeptide, the first three positions were kept constant according to their best results and the P4 position was varied. At last, the protecting group as well as the benzothiazole group were added to the best di-, tri and tetrapeptides and redocked to analyse the overall interactions with the enzyme.

Search for P1

After the creation of an initial scaffold, the next step was to perform the search for a suitable amino acid in the P1 position of the peptide inhibitor. The Gurken substrate exhibits an Ala at this position and sequence specific experiments have shown that only small amino acids are tolerated at the first position such as Ala, Ser, Gly and Cys [106]. Hence, a database was prepared with all four possible small amino acid residues at the P1 position, with Met at the P2 position. A conformation search for all four molecules was performed, leading to four result databases. These four databases were imported into one combined set of conformations, and docking was carried out using the ConflexDock_womm script. The first output database showed that Ser at P1 possessed the best score (S = -9.3) of all amino acids, followed by Ala (S = -8.5) and Cys (S = -7.9). All of these conformations were energy minimized after connecting them to the catalytically active Ser. During the minimization step at this stage, the enzyme was always fixed except for the Ser201 side chain so that the ligand could move freely. Table 4.1 summarizes the results obtained before and after the energy minimization step (MM).

Table 4.1: Docking scores of different amino acids in the P1 position. Docking scores (S-values) and corresponding calculated ligand efficiencies (cLE) are listed. The MM values indicate the energy minimized and rescored structures with London dG.

P1 residue	S-value	cLE	S-value (MM)	cLE (MM)
Ser	-9.3	-0.66	-9.2	-0.66
Ala	-8.5	-0.65	-8.8	-0.68
Cys	-7.9	-0.56	-8.5	-0.61

As mentioned in section 3.4.4, the ranking of ligands primarily depends on their ligand efficiency, which is the docking score divided by the number of heavy atoms. Although Ser had the best overall scores, its calculated ligand efficiency (cLE) was lower than that of Ala because it has one extra heavy atom. None of the three amino acids displayed significant interactions with the active site (Figure 4.13). Since peptide inhibitors are innately very hydrophilic, the use of hydrophobic residues such as Ala might improve resorption in the body. In addition, Ala is the amino acid present in the P1 position of the Gurken peptide. Hence, Ala in P1 was used for the subsequent optimization of the P2 position.

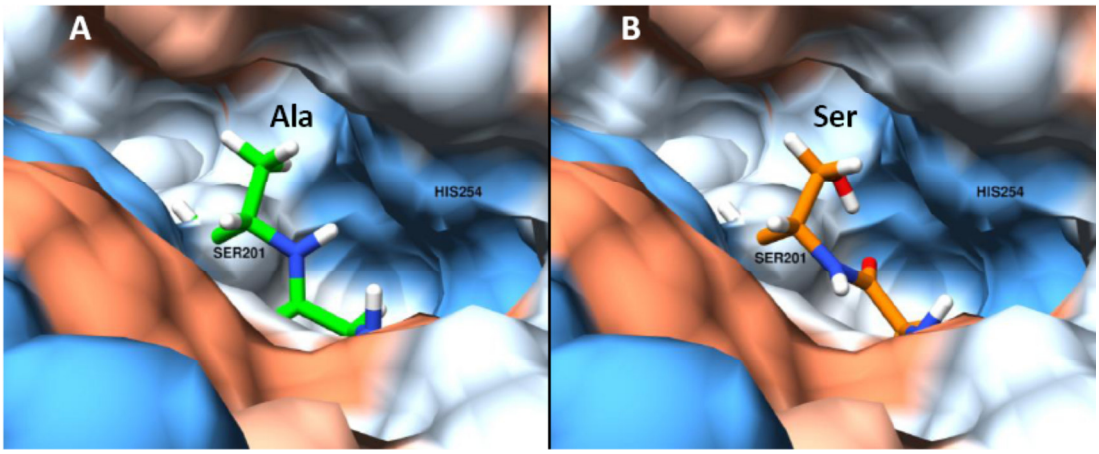


Figure 4.13: Top view of the best scoring structures of the P1 search. Surface colouring according to the hydrophobicity scale of Kyte and Doolittle: orange = hydrophobic; blue = hydrophilic. **A)** Ala displayed the best calculated ligand efficiency (- 0.68) compared to all other tested amino acids. **B)** Ser showed the highest score of all amino acids, but did not reach the cLE of Ala. No specific interactions of either Ala or Ser with the enzyme could be observed. The carbon atom that was covalently bound to the enzyme disappeared below the surface of the enzyme, which is an artefact of the dockings in MOE.

Search for P2

It has been shown previously that the P2 position in rhomboid substrates is non-specific and can feature almost all amino acids since mutations at this site are generally well tolerated with respect to substrate specificity [106]. Additionally, the rhomboid substrates Gurken and TatA exhibit Met and Thr residues at the P2 position, which have very different properties. To find the best-fitting residue through our modelling experiments, a database with a variety of amino acids at P2 was prepared: Met (hydrophobic), Thr (polar and uncharged), Lys (positive), and Glu (negative). To gain specificity, small residues were omitted for this position. A combined conformational database was created as described above and covalently docked with the previously mentioned settings. In order to restrict the degrees of freedom for the conformational search, the ligand was fixed at the Cα-atom of Ala but without the hydroxyl group (Figure 4.14).

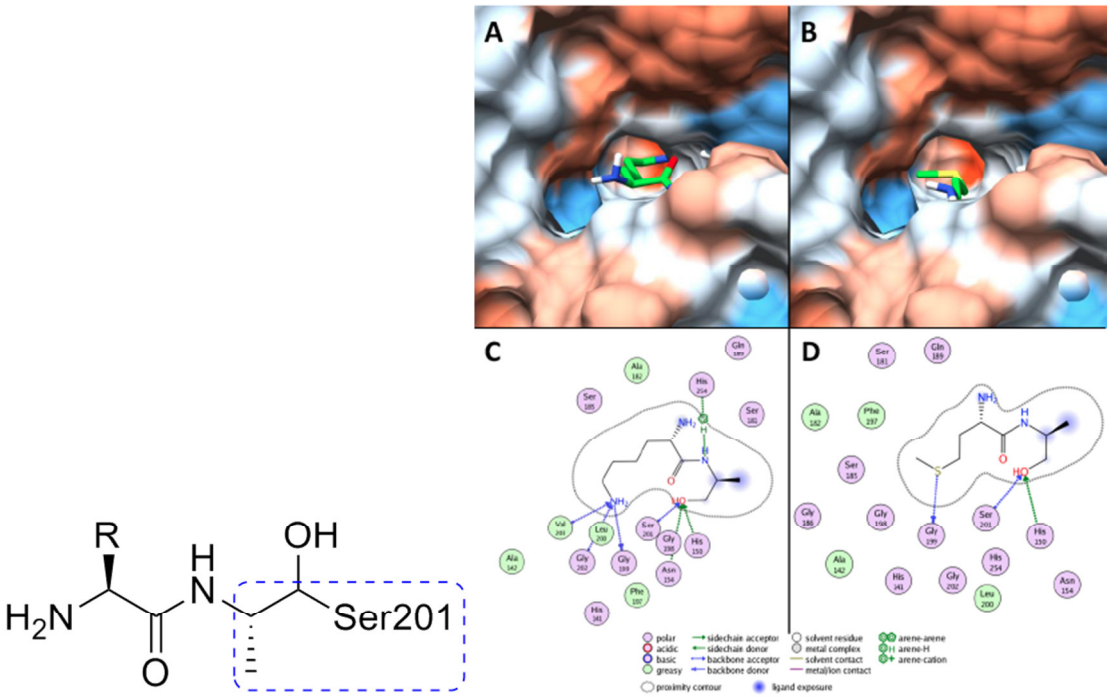


Figure 4.14: Covalently docked dipeptide inhibitors in the search for the P2 residue. The peptide chain was kept fixed at the Cα-atom of the alanine residue (P1) during the docking procedure. Subsequently, an energy minimization with completely free molecules was performed. The blue box indicates the part of the molecule that was fixed for the conformational search. Right: Molecular modeling results. In the upper panel, the active site of the rhomboid protease GlpG is shown with Lys (A) or Met (B) in the P2 position. In the lower panel, the 2D ligand interactions for the dipeptide inhibitors with Lys (C) and Met (D) at the P2 position are displayed.

Lys showed the 12 best docking scores (S = -9.3 to -8.3), followed by Met on the 13th rank (S = -8.3) and Thr (S = -8.0). Thr had the best cLE followed by Lys and Met with values of -0.67, -0.66 and -0.64, respectively. In Figure 4.14, the rhomboid pocket with the binding interactions for Lys and Met is displayed. It was observed that the Lys residue contributed the maximum number of interactions with the enzyme (Figure 4.14, A and C) followed by Met (Figure 4.14, B and D), which fitted very well into the deep pocket. Taking specificity into account, Thr with a comparably small side chain was not considered further. Energy minimization of the best conformation of Lys at the P2 position even increased its score and cLE to - 10.3 and - 0.74. Therefore, Lys at the P2 position was used for the subsequent optimization of the P3/P4 positions. However, since Met is present in the natural substrate Gurken in P2, this residue was also considered in the final docking experiments with the complete molecule incorporating the benzothiazole group (see below).

Search for P3

Similar to the P2 position, the P3 position has also been reported to be non-restrictive [106]. Therefore, several different residues were tested: Arg (naturally occurring at P3 in the Gurken peptide), Lys, Tyr, Leu, Glu, Gln, Met, Ser and Ala. Based on the scaffold with Lys at P2 and Ala at P1, the conformational search was conducted after fixing the atoms as shown in Figure 4.15A to ensure a limited degree of freedom.

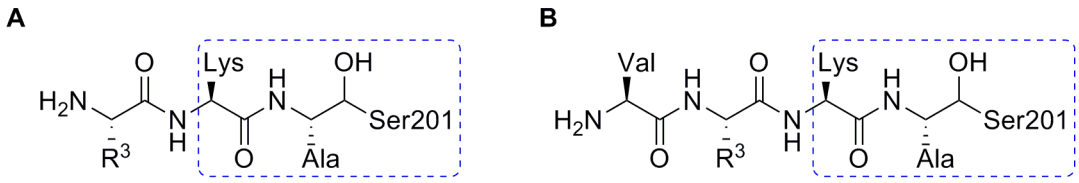


Figure 4.15: Structures of the tripetide (A) and the tetrapetide (B) used in the search for the best amino acid at the P3 position. The blue boxes indicate the parts of the molecules that were fixed for the conformational search.

Afterwards, docking was performed as described above, which revealed Ser as the best amino acid at P3 with both the highest score and the highest cLE. The raw ligand structures of the docking experiments showed sterical clashes with the enzyme surface. However, after energy minimization of the best conformations these clashes disappeared, resulting in the best cLE for serine at P3. The scores and calculated ligand efficiencies of the five best residues after the minimization step are shown in Table 4.2.

Table 4.2: Scores of the most favorable amino acids in the P3 position within the tripeptide. S-values and corresponding calculated ligand efficiencies (cLE) of the dockings after energy minimization (MM) are listed. Parameters were measured by rescoring the energy minimized structures with London dG.

P3 residue	S-value (MM)	cLE (MM)
Ser	-12.2	-0.61
Arg	-12.2	-0.49
Glu	-12.4	-0.54
Lys	-12.4	-0.53
Gln	-12.1	-0.52

The scores for all five amino acids were in a similar range. In addition, the interaction maps of the best-scored ligands showed that there were no significant interactions with the enzyme contributed by the P3 residues (Figure 4.16). Indeed, compared to the dipeptide, the calculated ligand efficiency dropped dramatically to a value of - 0.6, with Ser at P3. To exclude negative sterical influences on these results, another P3 search was performed using a tetrapeptide approach. For this purpose, a new construct with an extended backbone up to a tetrapeptide was prepared (Figure 4.15B). To keep interactions as minimal as possible, Ala was chosen for P3 and Val for the P4 position as this residue is present in the Gurken substrate. This construct was energy minimized with the ligand and the Ser201 side chain free to move, and the structure was saved for the P3 and P4 search. Next, a conformational database with all the amino acids mentioned above for the third position was generated and subsequently docked. Surprisingly, a completely different order of preference for P3 was obtained (Table 4.3). Leu (S = - 14.2, LE= - 0.49) showed the best results before the energy minimization step. In contrast, Met had the best S-value and cLE after the energy minimization step (S = - 14.6, LE = - 0.50).

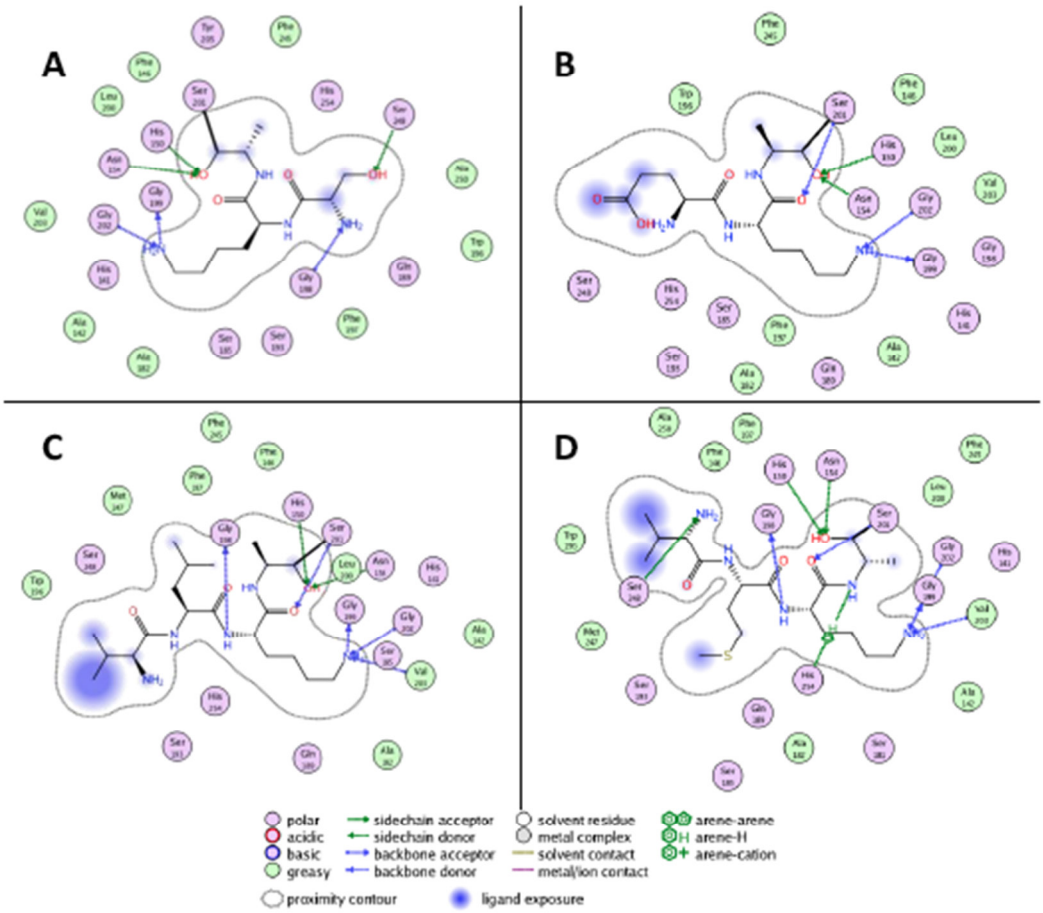


Figure 4.16: 2D interactions of the four best conformations observed during the P3 search. Upper two panels: best conformations of the dockings performed with a tripeptide. A) Ser displayed the best calculated ligand efficiency and showed two weak hydrogen bonds with the enzyme. B) Glu was the second best residue in cLE, but it did not exhibit interactions with the enzyme and showed high solvent exposure. Lower two panels: best conformations of the dockings performed with a tetrapeptide. Leu (C) as well as Met (D) did not form any interactions with the enzyme. The residue in the P4 position points out of GlpG pocket and is greatly exposed to the solvent.

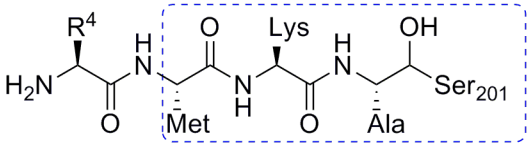
Table 4.3: Scores of the most favourable amino acids in the P3 position within the tetrapeptide. S-values and corresponding calculated ligand efficiencies (cLE) of the dockings before and after energy minimization (MM) are listed. Parameters after MM were measured by rescoring the energy minimized structures with London dG.

P3 residue	S-value	cLE	S-value (MM)	cLE (MM)
Leu	-14.2	-0.49	-13.4	-0.46
Met	-13.5	-0.47	-14.6	-0.50
Gln	-13.4	-0.45	-12.8	-0.43
Ala	-11.4	-0.44	-	-
Lys	-12.9	-0.43	-	-

Interestingly, Ser was not even in the top ranked residues with the tetrapeptide approach. These results appear difficult to explain. However, the common observation with both the tri- and the tetrapeptide was a drop in cLE compared to the dipeptide, and the residues at P3 did not show any specific interactions with the enzyme pocket. Only Ser in the tripeptide formed hydrogen bonds whereas all other residues did not form any interactions. The 2-D ligand interaction results are summarized in Figure 4.16, which shows the two best amino acids at P3 with their energy-minimized conformations for both the tripeptide and the tetrapeptide approaches.

Search for P4

According to the literature, the P4 position is quite important for substrate recognition by rhomboid proteases, and hydrophobic and bulky residues in P4 are preferred for activity [106]. Hence, a conformational search was performed with all possible amino acids: Val (present in the Gurken peptide), Phe, Leu, Ile, Met, Tyr and Trp. As before, atoms of the ligand were fixed to the Cα of the Met at P3 (Figure 4.17). The docking results before and after the energy minimization step identified the best four residues (Phe, Val, Met, Trp) for the P4 position, and all possessed similar calculated ligand efficiencies after the energy minimization. In contrast to the biochemical evidence in the literature, all amino acids at P4 showed no significant interactions with the enzyme, similar to the P3 position. The positioning of the tetrapeptide in the rhomboid active site as well as the ligand interactions with the amino acids of the receptor are shown in Figure 4.17. Trp appeared to contribute to ligand interactions despite its surface exposure (Figure 4.16) and was therefore the best candidate for the P4 position.



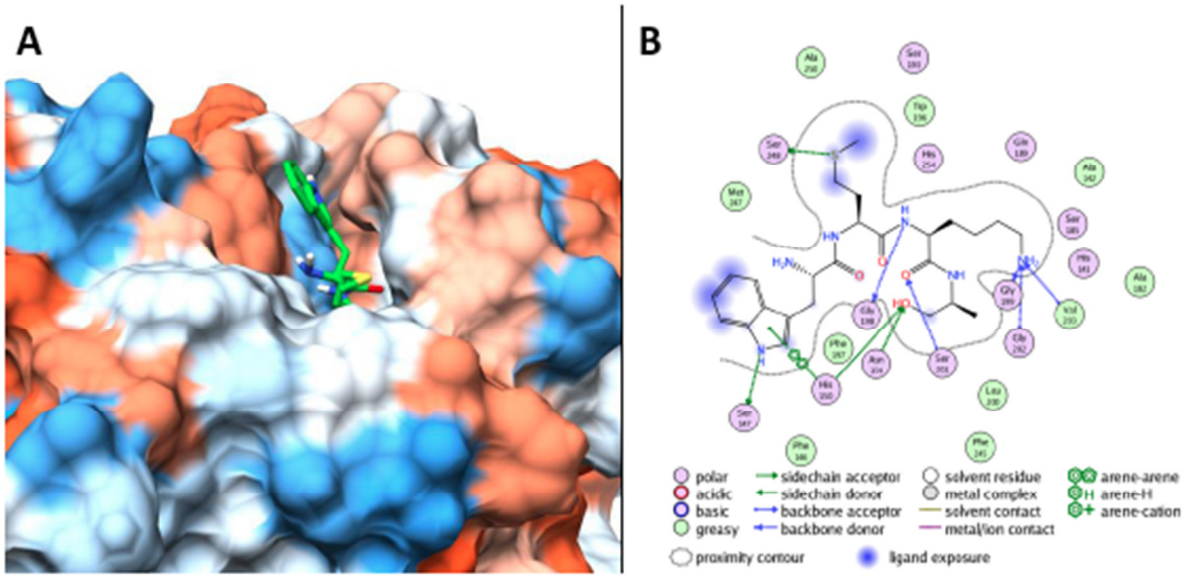


Figure 4.17: Covalently docked tetrapeptide inhibitor in the search for the P4 residue. Top: Structure of the tetrapeptide used in the search for the best amino acid in the P4 position. The blue box indicates the part of the molecule that was fixed for the conformational search. Bottom: Molecular modeling results with Ala at P1, lysine at P2, Met at P3, and Trp at P4. A) The S aperture of the active site of the rhomboid protease GlpG with Trp at P4, which is exposed to the surface. B) 2D ligand interactions for the tetrapeptide inhibitor.

As mentioned above, only weak interactions were contributed by the Met residue at P3 accompanied by high ligand exposure. The addition of another amino acid to the N-terminus resulted in a further

drop in cLE. Furthermore, according to our docking results, the P4 residues protruded out of the receptor pocket and was exposed to the surrounding media, which is illustrated in Figure 4.17 for Trp at P4. Overall, the docking results indicated that it was not promising to extend the peptidic component beyond a dipeptide for high affinity inhibition. Nevertheless, the docking experiments performed for P3 and P4 might later be useful to improve the selectivity of the peptidomimetic inhibitors.

Docking of the di-peptide inhibitor with benzothiazole group

Based on the docking results, the ideal length for a peptidomimetic inhibitor against rhomboids was determined to be a dipeptide. The final step was to add the benzothiazole serine trap to a dipeptide with either Lys or Met in the P2 position to explore the binding affinities of the whole molecule. The N-terminal cbz protecting group was not included as it can be cleaved off if needed. Moreover, it did not contribute much to interactions with the enzyme (data not shown). The best energy minimized conformations with Lys or Met at the P2 position were selected, and the benzothiazole group was added at the C-terminus using the builder module. Then, a conformational search was performed for both scaffolds and the best conformations of the subsequent dockings were energy minimized.

An improvement in the docking scores but a decline in the cLE were observed for both scaffolds. Surprisingly, it was also observed that the energy-minimized conformation of the Lys scaffold with the benzothiazole did not display the three significant H-bond interactions with the binding site (Figure 4.18B), which were present without the benzothiazole moiety (Figure 4.18A). However, the Met scaffold fitted perfectly into a putative S2 binding pocket (Figure 4.18C) after addition of the benzothiazole moiety. Hence, Met was selected as the best suited residue for the P2 position instead of Lys. In summary, the docking analysis of protein/ligands indicated that a dipeptide is the adequate length for the design of peptidomimetic inhibitors against the rhomboid protease GlpG, and the best suited residues appeared to be Ala for the P1 position and Met for the P2 position.

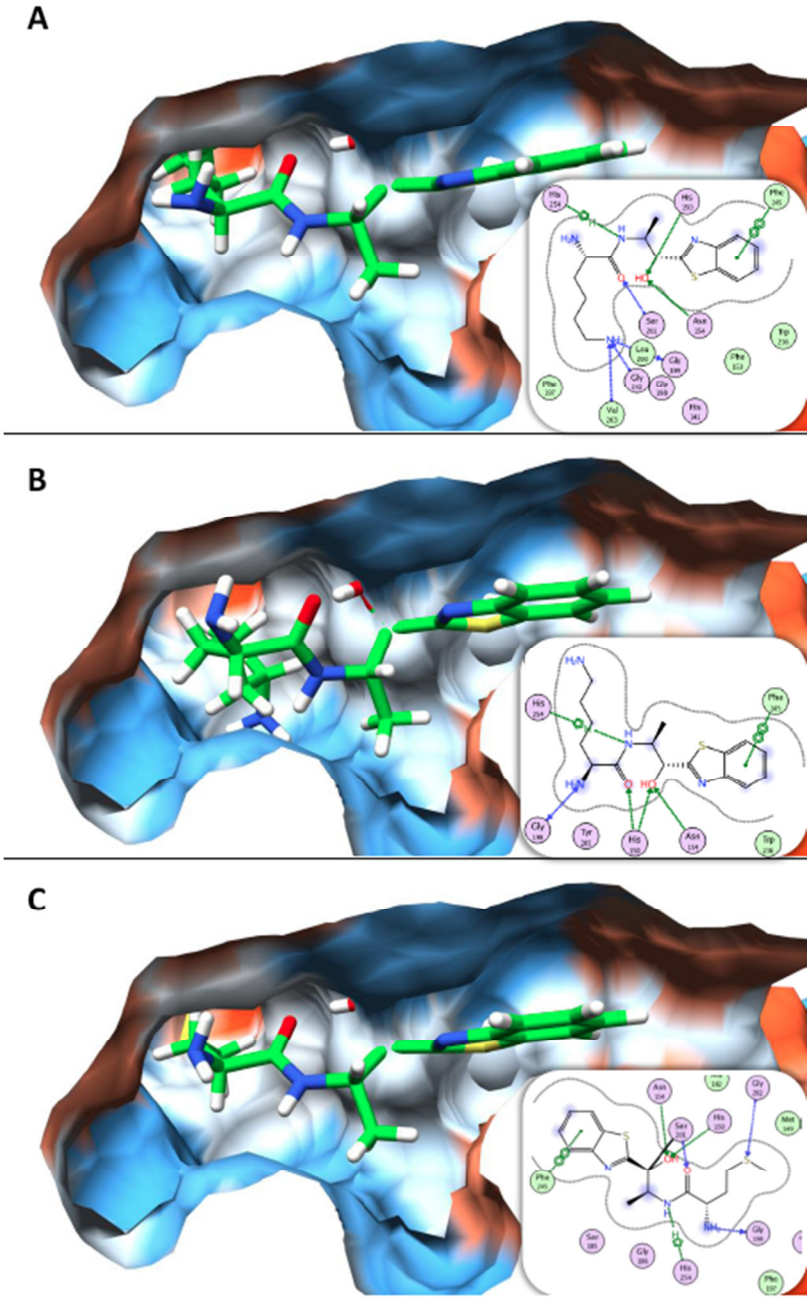


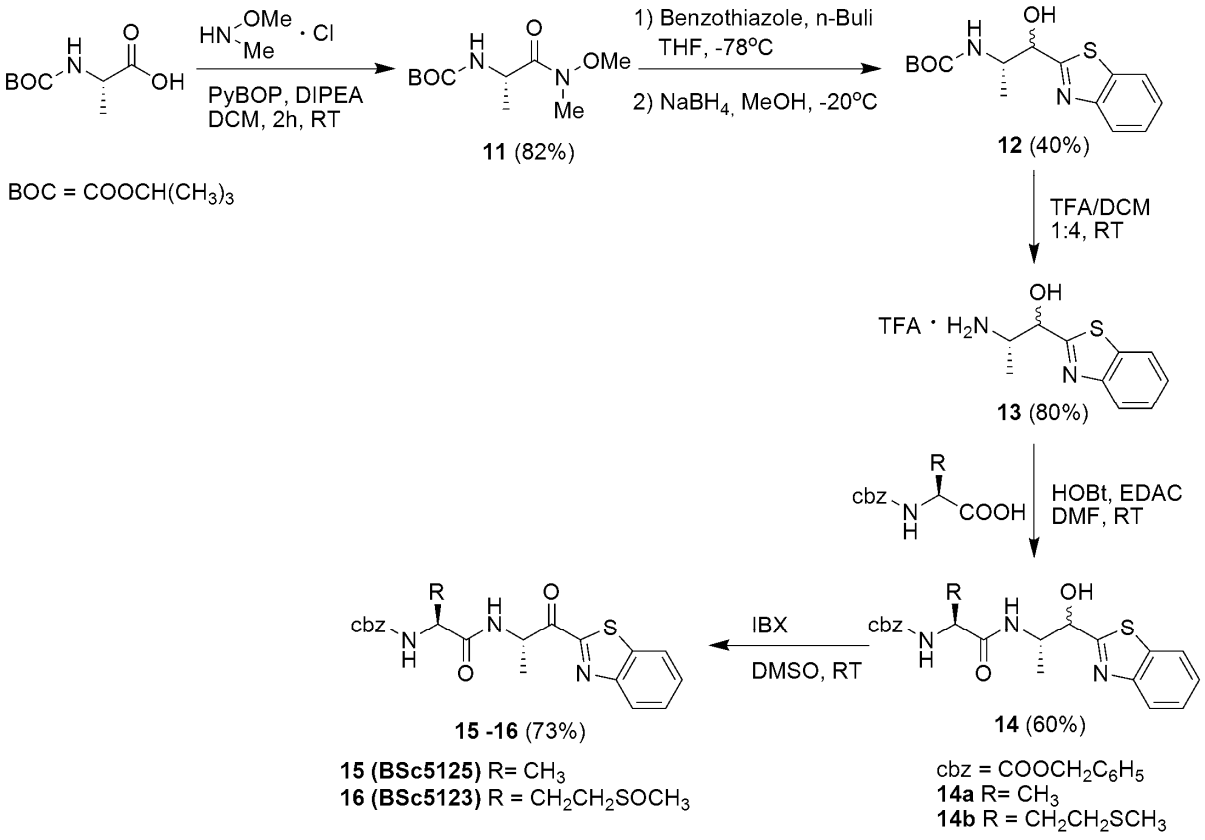
Figure 4.18: Covalent docking of a dipeptide inhibitor with the benzothiazole serine trap moiety. 3D illustrations and 2D interaction maps are shown for each of the three best conformations. A) The best Lys conformation from the P2 search (Figure 4.14) with the benzothiazole group added. The ligand interaction map shows that all three hydrogen bonds were still present. B) Another round of conformational search and energy minimization was performed with Lys at P2 and the benzothiazole moiety. Afterwards, the Lys residue did not enter the postulated S2 pocket and was therefore unable to form the hydrogen bonds. C) The dipeptide with Met at the P2 position and the benzothiazole moiety after another round of conformational search and energy minimization. The side chain reached into the pocket and formed a hydrogen bond with Gly202.

4.5.1.2 Synthesis of ketobenzothiazole derivatives

Based on the docking results, the synthesis of peptidomimetic ketobenzothiazole based inhibitors consisting of either the monopeptide sequence cbz-Ala (BSc5124) or the dipeptide sequences cbz-Met-Ala (BSc5123) and cbz-Ala-Ala (BSc5125) with a terminal ketobenzothiazole serine trap was performed. The inhibitors BSc5123 and BSc5125 were obtained by a method based on a classic

peptide coupling strategy via the Weinreb amide route (Figure 4.19, scheme 1), reported by Costanzo *et al*. [147]. In the first step, fragment 12 carrying the serine trap was generated by addition of in situ produced 2-lithiobenzothiazole to the Weinreb amide 11. The resulting ketobenzothiazole was further reduced in the same step using $NaBH_4$ in order to protect the electrophilic keto group. This was followed by BOC group deprotection through acidolysis, which lead to the functionalization of P1 carrying fragment 13. This fragment was further coupled to the respective amino acid of the P2 position (cbz-Ala or cbz-Met), resulting in the dipeptide scaffold 14. Then oxidation was performed using 2-iodoxybenzoic acid (IBX) to obtain the desired inhibitors 15 (BSc5125) and 16 (BSc5123).

Scheme 1



Scheme 2

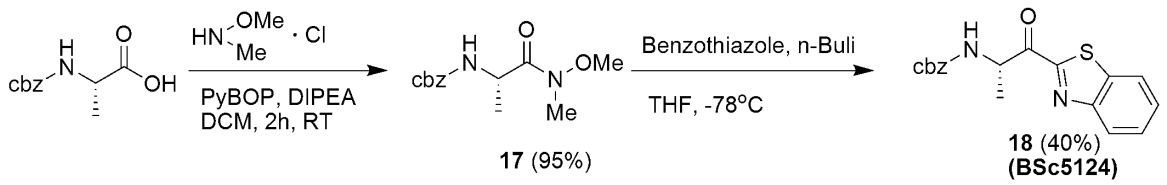


Figure 4.19: Synthetic schemes of ketobenzothiazole-based peptidic inhibitors. Scheme 1 shows the general synthesis of the ketobenzothiazole-based dipeptide inhibitors (BSc5125 and BSc5123). Scheme 2 shows the synthesis of the monopeptide inhibitor BSc5124.

Importantly, during the last oxidation step for Met at P2 both the sulfur and the hydroxyl group get oxidized. Hence, the inhibitor BSc5123 contained oxidized methionine at the P2 position. The inhibitor BSc5124 was synthesized through a similar strategy. The analogous Weinreb amide 17 was

obtained by general peptide coupling, and was then reacted with 2-lithiobenzothiazole in situ to produce inhibitor 18 (BSc5124). For detailed synthesis information and spectral analysis of the corresponding intermediate molecules and inhibitors, please refer to the method section 3.5.2 of this thesis.

4.5.1.3 Biological Evaluation: *In vitro* activity assay with ketobenzothiazole derivatives

The synthesized ketobenzothiazole-based peptidomimetic compounds BSc5123, BSc5124 and BSc5125 were tested for their activities in the GlpG *in vitro* activity assay (described in section 3.2.12). All inhibitors were first examined at high concentrations, 250 µM and 500 µM. 50 µL reaction mixtures containing 0.35 µM GlpG were pre-incubated at 37˚C with the inhibitors for 30 min. This was followed by the addition of 1.8 µM Gurken substrate, and the reaction was continued for 90 min at 37˚C. Several controls were employed: a DMSO vehicle control (maximal enzyme activity), and reaction mixtures that contained only either the GlpG enzyme or the Gurken substrate. The reaction products were analysed by SDS-PAGE (Figure 4.20).

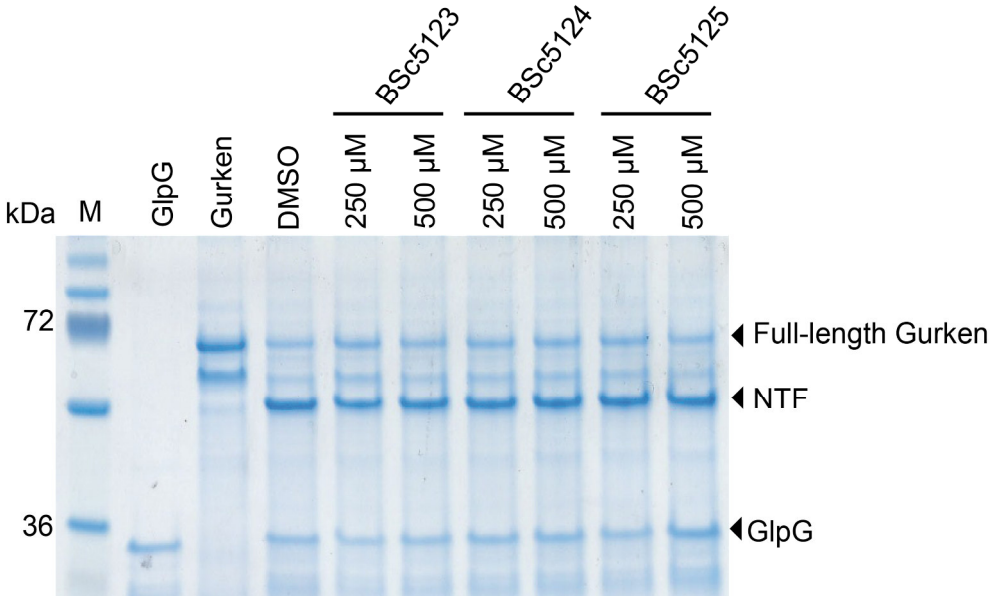


Figure 4.20: Rhomboid *in vitro* activity assay with ketobenzothiazole-based peptidomimetic compounds. SDS-PAGE analysis of activity assays with derivatives BSc5123, BSc5124 and BSc5125. The arrows indicate the full-length Gurken substrate (~ 66 kDa), the N-terminal fragment (NTF, ~ 55 kDa) and the GlpG enzyme (~ 33 kDa). M: Molecular weight marker. The gel was stained with Instant Blue™.

The full-length substrate and the N-terminal Gurken fragment (NTF) were visible around 66 kDa and 55 kDa. To analyse the effectiveness of the inhibitors, the intensity of the NTF of the DMSO vehicle control reaction was compared to the reaction mixtures containing inhibitors (Figure 4.20). This demonstrated that none of the three peptidomimetic inhibitors showed any activity in the tested concentration range.

4.5.2 Oxathiazol-2-one derivatives

4.5.2.1 Design and docking studies of oxathiazol-2-one derivatives

Another interesting scaffold for the design of peptidic inhibitors are oxathiazol-2-one derivatives. This scaffold was discovered to inhibit the proteasome, with the terminal oxathiazol-2-one ring acting as the reactive motif attached to a recognition peptide sequence (Figure 4.21) [148]. The mechanism of inhibition involves the reaction of the electrophilic ketone center in the reactive motif with the threonine (Thr) nucleophilic active site residue [148]. Since serine proteases have a similar inhibition mechanism, I selected oxathiazol-2-one as a second peptomimetic template to generate novel rhomboid protease inhibitors. Considering the docking results for peptide-based inhibitors described above (section 4.5.1.1), I kept the dipeptide as the target length for the peptidic component and three derivatives were synthesized.

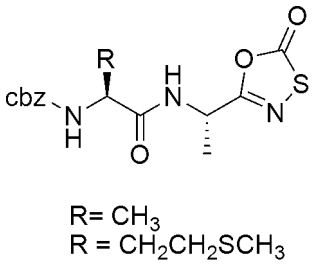


Figure 4.21: The oxathiazol-2-one based scaffold selected to design rhomboid inhibitors.

4.5.2.2 Synthesis of oxathiazol-2-one derivatives

The synthesized peptidomimetic oxathiaozle-2-one derivatives consisted either of the monopeptide sequence BOC-Ala (BSc5139) or the dipeptide sequences cbz-Met-Ala (BSc5138) and cbz-Ala-Ala (BSc5140) with a terminal oxathiaozle-2-one. The potential inhibitors were obtained by a method reported by Gezginci *et al*. [149] and respective BOC-proteccted carboxamide were prepared by method reported Mathews *et al*. [150](Figure 4.22, scheme 3). This involved the treatment of the BOC-protected carboxamide fragment 19 with chlorocarbonylsulfenyl chloride, which resulted in compound 20 (BSc5139). This compound was further activated by BOC-deprotection and resulted in fragment 21, which was then coupled with a cbz-protected amino acid, either cbz-Ala or cbz-Met to yield the intended inhibitors 22 (BSc5140) and 23 (BSc5138). For detailed synthesis information and spectral analysis of the corresponding intermediate molecules and inhibitors, please refer to the method section 3.5.3 of this thesis.

Scheme 3

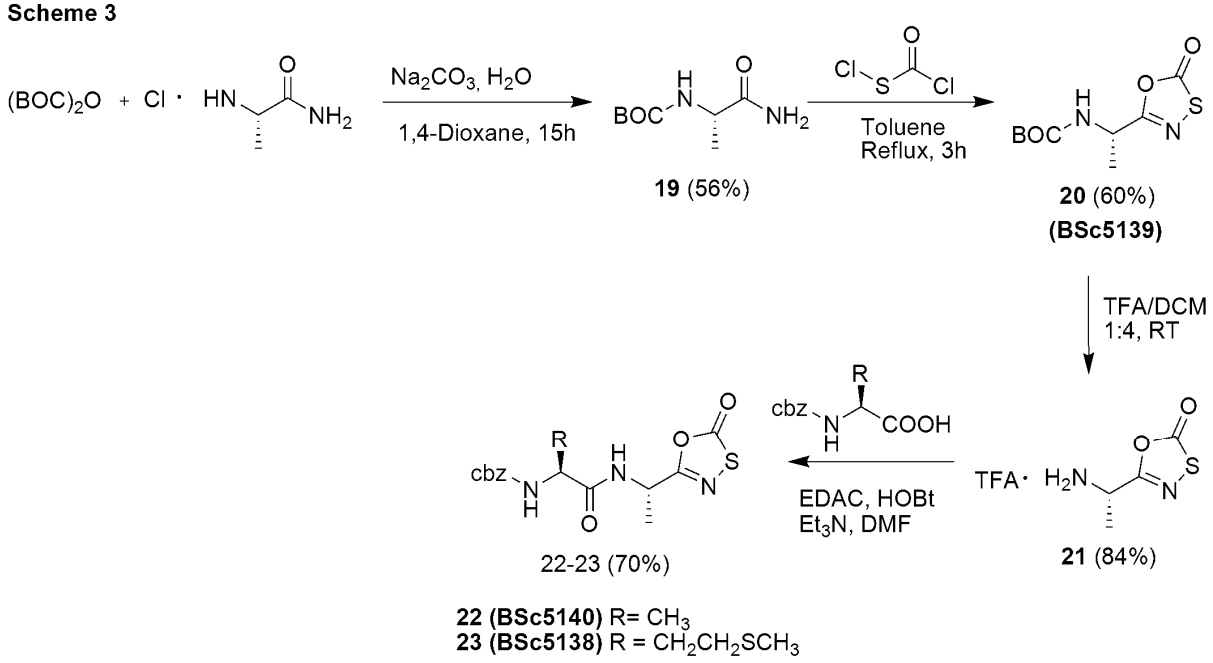


Figure 4.22: Synthetic scheme of oxathiazole-2-one-based peptide inhibitors. Scheme 3 shows the general synthesis of oxathiaozle-2-one-based monopeptide (BSc5139) and dipeptide inhibitors (BSc5140 and BSc5138).

4.5.2.3 Biological evaluation of oxathiazol-2-one derivatives

Similarly, to the ketobenzothiazole-based peptidomimetic compounds described in section 4.5.1.3, the newly synthesized oxathiaozle-2-one based peptidomimetic compounds BSc5138, BSc5139 and BSc5140 were tested for their activities against the rhomboid GlpG in the *in vitro* activity assay. 50 µL reaction mixtures containing 0.35 µM GlpG were pre-incubated at 37˚C with the inhibitors. Subsequently, 1.8 µM of Gurken substrate was added and the reaction was continued for 90 min at 37˚C.

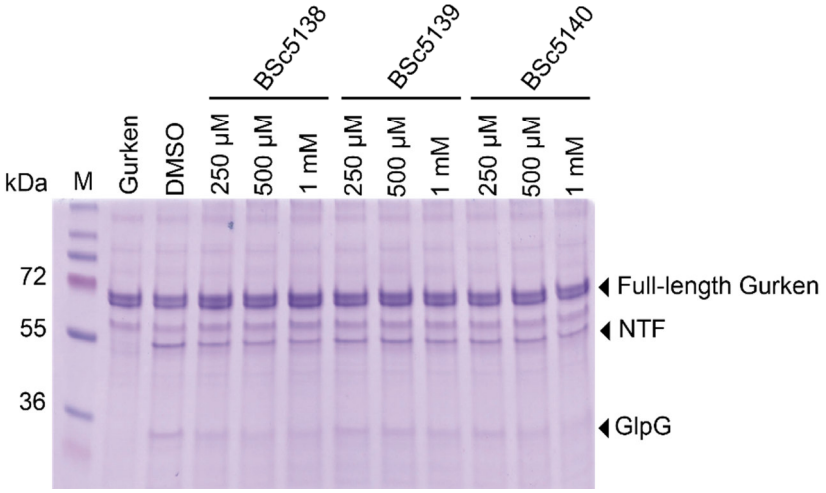


Figure 4.23: Rhomboid *in vitro* activity assay with oxathiazole-2-one based peptidomimetic compounds. SDS-PAGE analysis of activity assays with derivatives BSc5138, BSc5139 and BSc5140. The arrows indicate the full-length Gurken substrate (~ 66 kDa), the N-terminal fragment (NTF, ~ 55 kDa), and the GlpG enzyme (~ 33 kDa). M: Molecular weight marker. The gel was stained with coomassie brilliant blue.

All inhibitors were initially tested at high concentrations of 250 µM, 500 µM and 1 mM, and the reaction products were analysed by SDS-PAGE (Figure 4.23). Again, a comparison of the NTF band intensity in the DMSO vehicle control reaction with the NTF band intensities of the reaction mixtures containing inhibitors showed that none of the three peptidomimetic compounds had any inhibitory activity in the tested concentration range.

4.6 Design, synthesis and biological evaluation of non-peptidic rhomboid inhibitors

Highly potent peptidomimetic inhibitors that mimic natural substrates have been frequently applied to investigate the roles of proteases in biological systems [109]. But peptidomimetic inhibitors suffer from serious pharmacological drawbacks including poor solubility and metabolic stability, poor oral bioavailability and toxic effects [151]. To overcome these problems, mechanism based serine protease inhibitors including small molecules with heterocyclic and non-peptidic scaffolds have been designed (section 1.3.2) [108, 110]. Some of these broad-spectrum mechanism-based inhibitors such as isocoumarins, β-lactams, and β-lactones were already known to inhibit rhomboid proteases (section 1.5) [123]. However, except β-lactam derivatives, none of them displayed high potency for rhomboids. β-lactams also showed only modest selectivity for rhomboids over the soluble serine protease chymotrypsin [103]. Hence, our aim was to find new classes of non-peptidic small molecule inhibitors with better potency and specificity compared to the already existing rhomboid protease inhibitors.

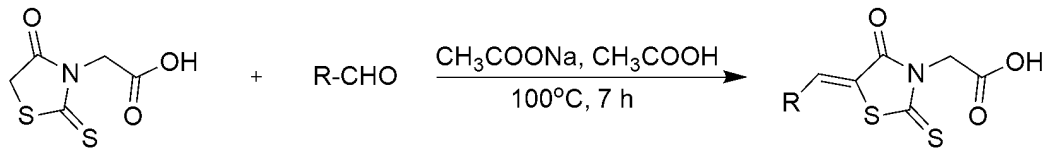
4.6.1 Rhodanine-3-acetic acid (RAA) inhibitors

4.6.1.1 Design, synthesis and docking studies of rhodanine-3-acetic acid derivatives.

Rhodanine containing compounds are five-membered multiheterocyclic (FMMH) ring inhibitors and are known as ‘privileged scaffolds’ or ‘promiscuous binders’ [152]. These compounds bind numerous distinct targets, which probably originate from a particular high potential to form intermolecular interactions but is not related to aggregation or other nonspecific inhibition mechanisms. In 1997, a study showed that despite a wide variety of bioactivities the pharmaceutical utility of rhodanines was limited because of poor solubility [153, 154]. But this problem has been overcome by using suitable salts of rhodanine-3-acetic acids (RAA) prepared by Körner in 1908, and Knoevenagel condensation products of the acid with various aldehydes, namely, [(5*Z*)-(5-benzylidene-4-oxo-2-thioxo-1,3-thiazolidin-3-yl)]acetic acids [153]. Since then, rhodanines have been further examined and were shown to exhibit potential anti-mycobacterial, anti-fungal, pesticidal, anti-hypertensive, and anti-neoplastic activities. [153, 154]. They are known to inhibit numerous targets such as HCV serine proteinase [155, 156], cathepsin D [157], aldose reductase [158, 159], C class beta-lactamase [160], UDP-N-acetylmuramate/L-alanine ligase [161], they encompass antidiabetic agents [162], and they were also proposed as potential inhibitors of bacterial toxins namely of the protease anthrax lethal

factor and the botulinum neurotoxin type A [163] [164]. Hence, considering their potential, RAA probes were selected as candidate inhibitors for rhomboids, and this section describes the docking, the synthesis and the biological effects of rhodanine-3-acetic acids probes. Recently, a series of RAA fluorescent probes were synthesized by Anumala *et al*. in the Schmidt lab (TU Darmstadt) and employed to visualize neurofibrillary tangles in Alzheimer's disease brains [165]. These RAA derivatives were obtained by published procedures through a Knoevenagel condensation between RAA and corresponding aldehydes using sodium acetate in acetic acid (Figure 4.24, scheme 4) [166]. Most of the aldehydes used here were commercially available but a few aldehydes had to be synthesized by known methods (Figure 4.24, scheme 5).

Scheme 4



Scheme 5

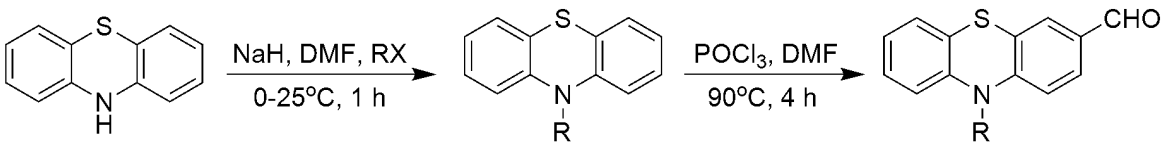
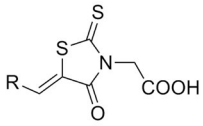
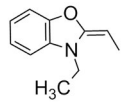
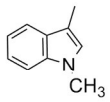
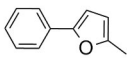
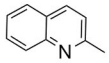
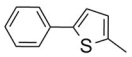
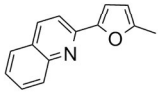
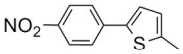
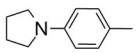
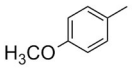
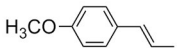
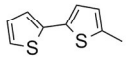
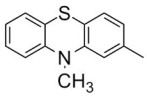
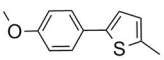


Figure 4.24: Synthesis schemes of rhodanine-3-acetic acid (RAA) derivatives. Scheme 4 shows the general synthesis of RAA derivatives used in the present study, and scheme 5 shows the synthesis of the aldehyde used for the synthesis of probe BSc4994. The R groups are listed in table 4.4.

Probe BSc4994 was synthesized in a three-step procedure. In the first step, alkylation of phenothiazine in the presence of sodium hydride and dimethyl formamide was performed. The N-alkyl phenothiazine was formylated in the second step employing the Vilsmeier–Haack reaction of phosphorous oxychloride and dimethylformamide. In the final step, a Knoevenagel condensation was carried out between the RAA derivative and N-alkyl phenothiazine aldehyde to obtain the compound. The yields of the Knoevenagel condensation for probes varied from 61% to 85% depending on the aldehydes. The presence of electron donating groups increased the reactivity whereas electron-withdrawing groups decreased the reactivity and the yield of derivative formation. These synthesized RAA probes available in the BSc archive of the Schmidt lab are listed in Table 4.4 and were selected to test against rhomboids. It was decided to follow the computer-aided approach and first dock all these probes into the rhomboid protease pocket before commencing with the biological activity assays.

Table 4.4: Rhodanine-3-acetic acid (RAA) derivatives.



Entry No.	BSc No.	R	Entry No.	BSc No.	R
24	4445		31	4828	
25	4764		32	4882	
26	4820		33	4884	
27	4821		34	4885	
28	4824		35	4886	
29	4825		36	4994	
30	4826				

Molecular docking studies were performed using the X-ray co-ordinates of the co-crystal structure of the *E. coli* rhomboid protease GlpG with the phosphate inhibitor CAPF (PDB: 3UBB). Similar to previous docking experiments, the phosphate inhibitor CAPF was selected as a control molecule for comparative data analysis. By following the protocol in MOE, the docking receptor and a molecular database with BSc rhodanine derivatives along with the reference molecule CAPF were prepared (section 3.4). The docking was performed using the DOCK workflow in MOE with the AMBER12: EHT-forcefield (3.4.3.1). All resulting conformations were first scored with London dG and the 30 best conformations were energy minimized inside the pocket. All these steps were performed for each molecule in the database, which produced an output database with the ten highest rated poses for each molecule. Docking results were evaluated based on three major criteria: calculated ligand efficiency, scoring function and ligand interactions (section 3.4.4). The analysis of the protein/ligand docking results indicated that the RAA derivatives listed in Table 4.4 were nicely positioned in the binding pocket of the enzyme. As shown in Figure 4.25, the RAA derivative BSc4820 occupied the rhomboid cavity similar to CAPF and was not exposed to the external environment. Both the exo-sulfur and the oxygen atom of the rhodanine ring and the terminal carboxylate group were in close proximity to the rhomboid active site and displayed close interactions with neighbouring residues.

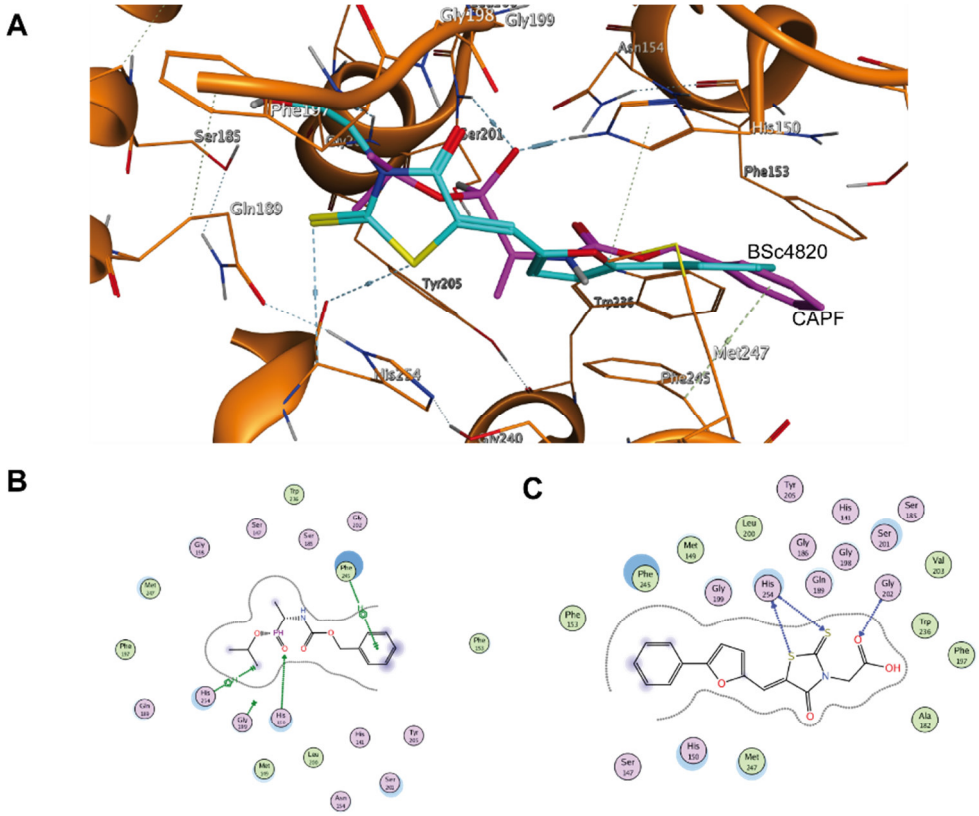


Figure 4.25: Molecular docking model of the rhomboid protease GlpG in complex with a rhodanine-3-acetic acid (RAA) derivative. A) GlpG docked with RAA derivative BSc4820 and CAPF. BSc4820 is shown in cyan blue color and CAPF in pink. Both compounds occupy a similar space in the rhomboid pocket and are not exposed to the external surface. B) Ligand interaction map of CAPF in the GlpG pocket. C) Ligand interaction map of BSc4820 in the GlpG pocket, showing significant interactions with neighbouring residues. Green: hydrophobic residues, pink: polar, red: exposed

Compared to the re-docked reference inhibitor CAPF, the calculated ligand efficiencies and scoring functions of the rhodanine derivatives appeared superior (Table 4.5), which provided an indication that rhodanines might be potential inhibitors for rhomboids. Subsequently, these compounds were evaluated for their ability to inhibit the *E. coli* GlpG in the *in vitro* activity assay.

Table 4.5: Docking results of rhodanine-3-acetic acid derivatives. S-values and corresponding calculated ligand efficiencies (cLE) of the dockings are listed.

Inhibitor	S-value	cLE
CAPF	-7.0393	-0.3521
BSc4445	-8.2195	-0.3435
BSc4764	-9.2351	-0.4015
BSc4820	-8.7293	-0.3795
BSc4821	-10.2170	-0.3931
BSc4824	-7.1285	-0.3564
BSc4825	-8.0992	-0.3681
BSc4826	-9.4967	-0.3799
BSc4828	-7.6733	-0.3488
BSc4882	-6.8699	-0.3123
BSc4884	-9.9043	-0.3668
BSc4885	-8.3360	-0.3624
BSc4886	-7.6217	-0.3424
BSc4894	-9.5548	-0.3539

4.6.1.2 Biological evaluation of rhodanine-3-acetic acid (RAA) inhibitors

All RAA inhibitors were examined at a concentration of 250 µM, and the assay was performed in the same manner as described in section 4.5.1.3. Protein bands around 66 kDa and 55 kDa corresponded to the full-length Gurken substrate and N-terminal proteolytic cleavage product (Figure 4.26). In these assays, the isocoumarin derivative JLK-6, a known GlpG inhibitor [87], was included as a positive control. With JLK-6 in the reaction mixture, 100 % enzyme inhibition was observed as the intensity of the full-length Gurken substrate was comparable to a reaction mixture containing only Gurken substrate but no GlpG enzyme. In addition, no N-terminal fragment (NTF) generation was observed in the presence of JLK-6 (Figure 4.26). In contrast, in the reaction mixtures containing RAA derivatives at a concentration of 250 µM the NTF band intensities were not different from the DMSO vehicle control indicating that no inhibition of the GlpG enzyme was achieved. Hence, this class of RAA derivatives was considered to be inactive against rhomboid proteases.

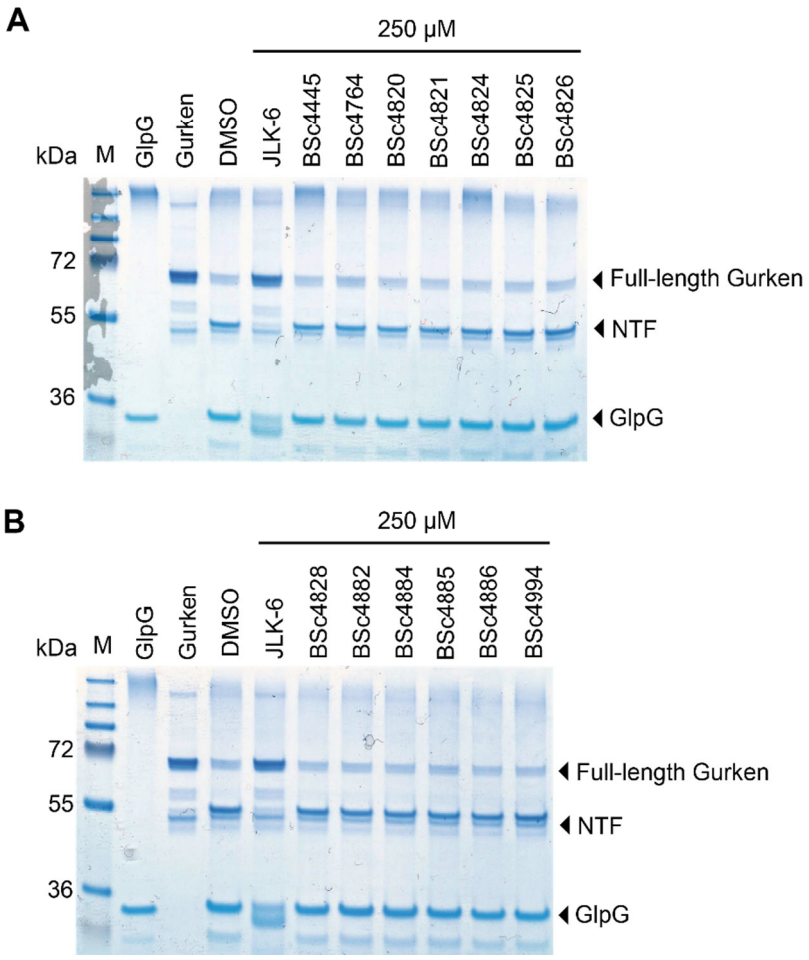


Figure 4.26: Rhomboid *in vitro* activity assay with rhodanine-3-acetic acids (RAA) derivatives. SDS-PAGE analysis of activity assays with RAA derivatives. The arrows indicate the full-length Gurken substrate (~ 66 kDa), the N-terminal fragment (NTF, ~ 55 kDa), and the GlpG enzyme (~ 33 kDa). No inhibitory effects of the RAA derivatives were observed at 250 µM. In contrast, the isocoumarin derivative JLK-6 reduced the NTF band intensity and caused accumulation of the full-length substrate, indicating inhibition of GlpG enzyme activity. M: Molecular weight marker. The gel was stained with Instant Blue™.

4.6.2 Saccharin-based inhibitors

In this section, saccharin or the 1,2-benzisothiazol-3-one 1,1-dioxide template (Figure 4.27c) was employed to design new enzyme-activated heterocyclic inhibitors against rhomboids. The basic structure of saccharin consists of a five-membered ring sulfimide with lactam and cyclic sulfonamide moieties, attached to an aromatic ring system (Figure 4.27a). The major assets of saccharin inhibitors are that they combine a broad spectrum of activity with the ability to differentiate between various serine proteases, presumably because of variable stability of the presumptive acyl-enzyme formed during the reaction of saccharins with the enzyme active site [110, 113]. Saccharin derived compounds were first discovered as irreversible acylating agents of serine proteases such as human leukocyte elastase (HLE) and cathepsin in 1980 by Zimmerman *et al* [167], and were further explored by testing *in vitro* and *in vivo* against HLE [168].

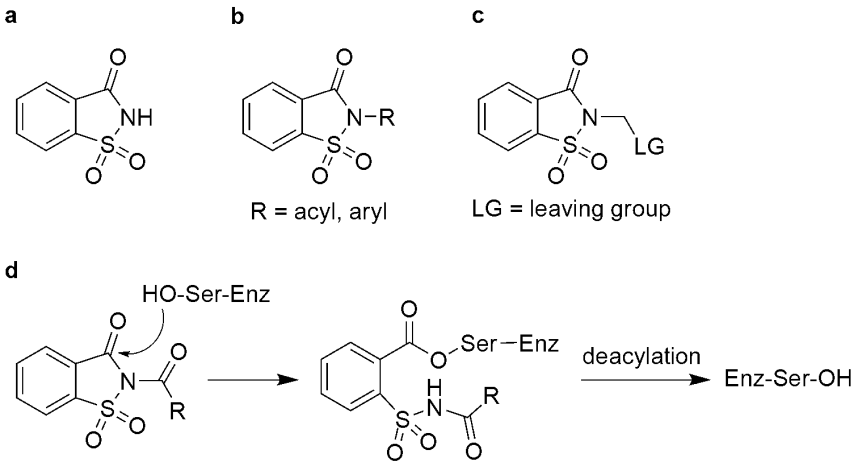


Figure 4.27: Novel saccharin ring derived inhibitors. a) Saccharin basic structure. b) General structures of N-acylsaccharins and N-arylsaccharins. c) Novel N-methylated substituted saccharin scaffold selelcted for the design of rhomboid protease inhibitors. d) Mechanism of inhibition of serine proteases by N-acyl saccharins, which act as alternate substrate inhibitors.

Two scaffolds of saccharin-based acylating agents that were initially employed, N-acylsaccharins and N-arylsaccharins (Figure 4.27b), both act as alternate substrate inhibitors (Figure 4.27d). This involves the formation of an acyl-enzyme via attack of the active site serine on the carbonyl group of the inhibitor, resulting in ring opening and formation of a sulphonamide (Figure 4.27d) [113]. It was observed that N-acylsaccharins were more biased towards HLE, pancreatic elastase, cathepsin G, and trypsin while *N*-arylsaccharins were only able to inhibit HLE and chymotrypsin [113]. Later, Groutas *et al* extensively investigated saccharins and a possible mechanism of inactivation based on their known chemistry and behaviour towards nucleophiles. They discovered a new class of saccharin inhibitors with properties of suicide inhibitors by attachment of a leaving group (LG) to the nitrogen atom (Figure 4.27c) [169-173], which are called next generation N-methylated substituted saccharin inhibitors. These N-methylated substituted saccharins also form an acyl-enzyme through the attack of the active site serine. However, unlike previous classes of saccharin inhibitors these N-methylated

substituted saccharins have the potential to generate a reactive intermediate that could covalently crosslink to the second active site nucleophilic histidine residue to form an irreversible enzyme-inhibitor adduct [113] (Figure 4.29). This new class of saccharin-based enzyme-activated inhibitors was reported to be active against HNE [151, 168, 174]. Another important advantage over previously known saccharin inhibitors was their potential for further derivatization, through changing the leaving group or substitution of the phenyl ring of the heterocyclic core structure. Moreover, their considerable utilisation in medicinal chemistry made them more interesting. For example, certain benzisothiazolone derivatives are key structural elements of orally bioavailable inhibitors of HLE [168] (Figure 4.28) and were also employed as intermediates in the production of significant anti-inflammatory drugs (oxicam) [175]. Additionally, saccharins being weak acid readily form salts (saccharinates) with different basic active pharmaceutical ingredients (API) thus making them highly soluble, and their sweetener properties are able to overcome the bitter taste of drugs [175]. In conclusion, N-methylated substituted saccharin inhibitors could be considered a privileged framework for the design of suicide or mechanism-based inhibitors against serine proteases.

4.6.2.1 Design and docking studies of saccharin-derived inhibitors

Hypothesis

Based on the observations described above, the N-methylated substituted saccharin ring system was explored as a starting scaffold for the design of mechanism-based inhibitors against rhomboid proteases. It has been reported that the potency of N-methylated substituted saccharins varied depending on the modifications of the saccharin ring system and the electron withdrawing properties of the LGs, and many of the SAR studies focused on both properties [168, 171, 176]. For example, Hlasta *et al*. reported that the potency of saccharin inhibitors for HLE depended on the C-4 and C-6 substitutions and also on the type of the LG [177, 178] (Figure 4.28). Various LGs attached to the N-methylated moiety have been studied in detail for HLE. Initially, Hlasta *et al.* identified 1-phenyl mercapto tetrazole as a suitable LG [177], and later SAR studies revealed many more favourable LGs such as aryl ethers, beta-dicarbonyl systems [174], phosphate esters [179], and aryl carboxymates [180] (some examples are shown in Figure 4.28). The derivatives with 2,6-dichlorobenzoate LG such as WIN 63394, WIN 63759, and WIN 64733 were the most efficient and specific inhibitors for HLE [113, 168] (Figure 4.29b-d). In addition, docking studies revealed that the aryl ring of saccharins acts as a recognition motif for HLE, with C-4 substituents pointing towards the S1 subsite and C-6 substitutions such as an alkoxy group leading to improved interactions with the enzyme pocket. This also provided enhanced *in vivo* stability which is the major drawback of mechanism-based inhibitors of serine proteases [113]. It was also found that the LG attached to nitrogen pointed towards the S' region of HLE, that interactions between the inhibitor and the S' region contributed to molecule

recognition, and that modification of the LG allowed the generation of trypsin specific inhibitors [113]. Hence, substitutions around the aryl ring of the heterocycle and the nature of the LG both influenced inhibitor specificity.

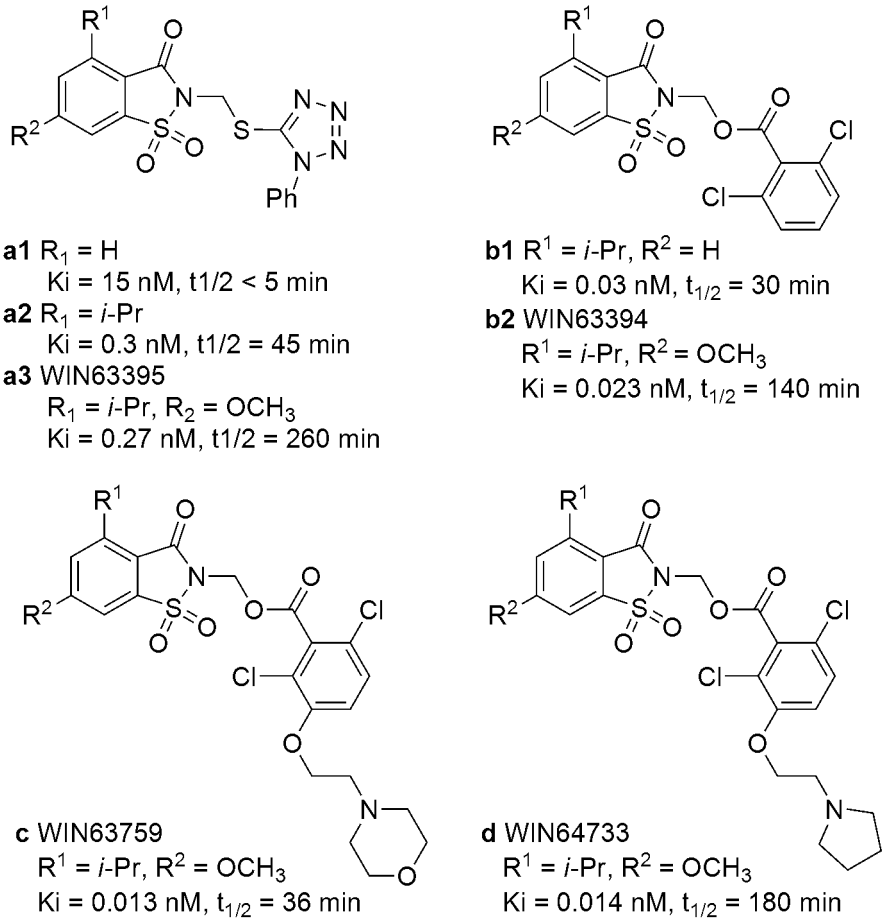


Figure 4.28: N-methylated substituted saccharin-derived inhibitors of HLE. The potency of the N-methylated substituted saccharin inhibitors shown in a-d against HLE depends on the nature of the leaving group as well as on the C-4 and C-6 substitutions. K_i refers to their inhibitory potency; the $t_{1/2}$ value provides the stability of the compounds in human blood.

However, functionalization of the heterocycle has proven difficult because electrophilic substitution is hindered by the presence of carbonyl and sulfonyl electron withdrawing groups. In contrast, the bond angle between C-S-N is fairly small (92.2°) resulting in a strained ring system, and the amide group can easily react with electrophilic reagents [175]. Hence, the derivatization of the N-functional group of the saccharin template was preferred over the aromatic ring system in most studies, and was also in the design of saccharin-based rhomboid inhibitors. Consequently, 50 saccharin derivatives with the most favourably LGs for serine proteases were selected from the literature (listed in Table 4.6) [175] and docked into the GlpG enzyme pocket, in order to get an initial idea which LG would be most suitable for rhomboid proteases.

Putative mechanism of inhibition by N-methylated substituted saccharins

Based on the literature [110], a possible mechanism of inhibition of rhomboid protease by N-methylated substituted saccharins involves the enzyme-induced ring opening by the attack of Ser201 similar to N-acylsaccharins (Figure 4.27d). Previously, Zimmerman *et al*. also clarified that the alkaline hydrolysis results into opening of the saccharin heterocyclic ring by the cleavage of the carbonyl-nitrogen bond rather than sulfonyl-nitrogen bond [181]. Since there was no sulfonic acid derivative reported after hydrolysis and they also showed that sulfonamides were not enzymatically cleaved, both in linear molecules or ring structures. Moreover, the tetrahedral sulfonamide resembles the transition state of the nucleophilic attack, thus there is no trajectory available for attack by the serine-OH. This suggested that the amide bond in the saccharin ring is the feasible site for enzyme attack. The acylated enzyme forms a tetrahedral intermediate followed by a prototropic shift and the release of the leaving group (Figure 4.29). This might result in the generation of a new electrophilic center, the Michael acceptor (Figure 4.29), which permits further enzyme inactivation through nucleophilic attack by His254 and the formation of an end product that is covalently crosslinked to the rhomboid.

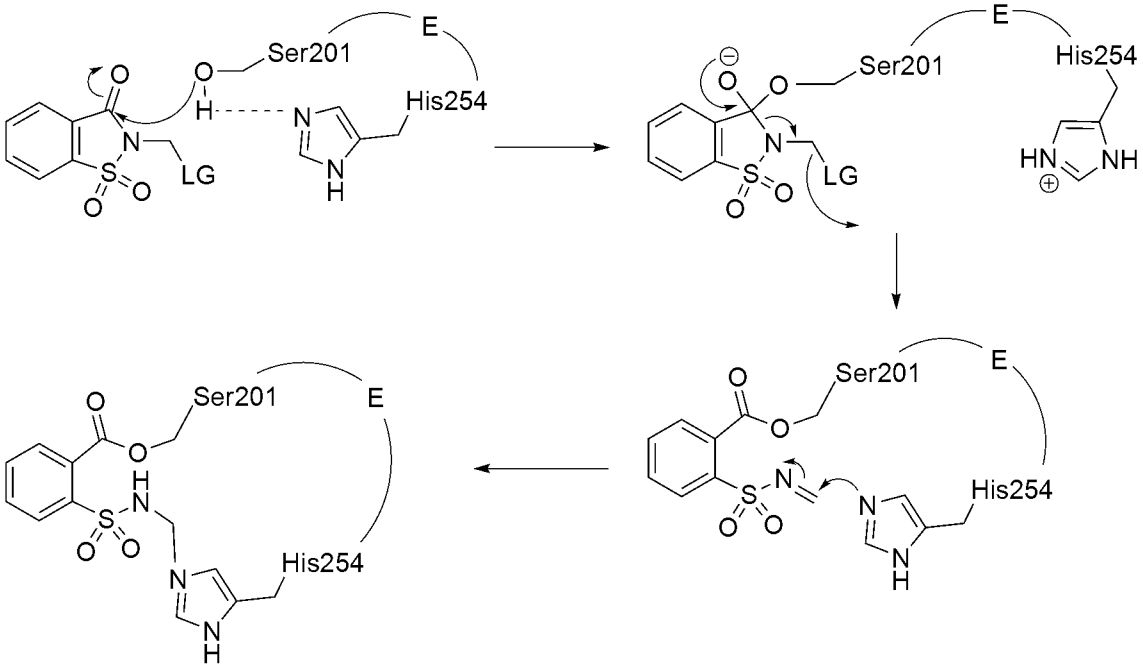


Figure 4.29: Putative mechanism of inhibition by N-methylated substituted saccharins. The possible mechanism involves the enzyme-induced formation of a Michael acceptor and results in an end product that is covalently crosslinked to the enzyme.

Docking strategy

As described above, the proposed inhibition mechanism of N-methylated substituted saccharins would lead to the release of the LG. As a consequence, reactions of the rhomboid enzyme with saccharin-

based inhibitors featuring different LGs would always result in the same double bonded end product (Figure 4.29 and Figure 4.30). Therefore, in the docking studies of saccharin derivatives we focused on their initial interactions with the rhomboid cavity rather than on the end product.

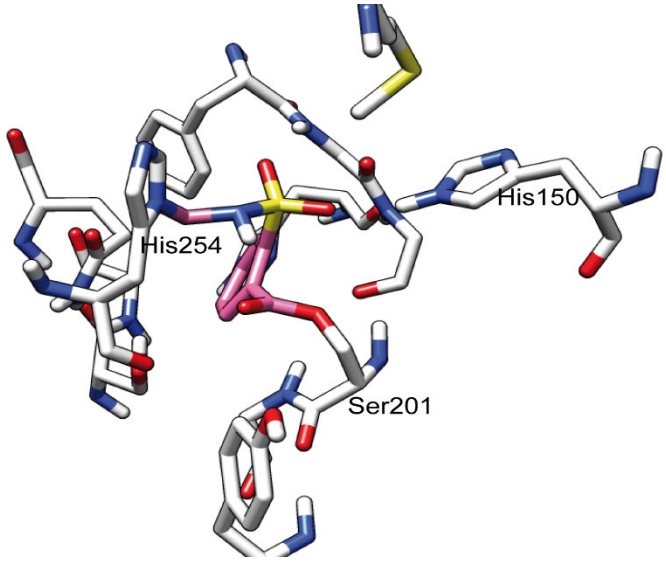
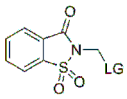
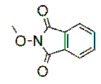
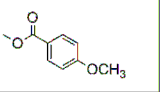
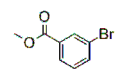
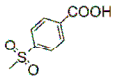
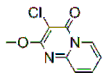
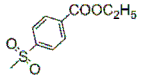
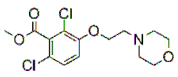
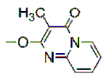
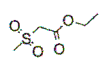
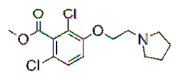
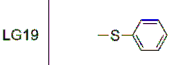
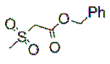
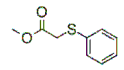
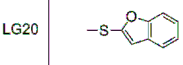
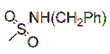
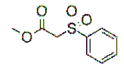
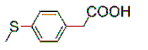
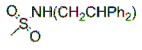
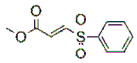
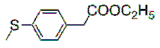
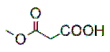
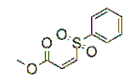
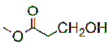


Figure 4.30: Molecular model showing the end product of the enzymatic reaction between N-methylated saccharins and the rhomboid protease (PDB: 3UBB). Two covalent bonds are formed between the saccharin and the rhomboid active site residues Ser201 and His254. The saccharin is shown in pink and the rhomboid cavity in white. The image was drawn using chimera.

A database called Test Set (Table 4.6) was prepared in MOE (section 3.4.2) consisting of saccharins with 50 different LGs. Since the initial interactions of the inhibitor with the rhomboid pocket was the focus of the docking experiments, the molecular modeling was performed using the non-covalent docking module DOCK in MOE (section 3.4.3.1). Similar to previous docking experiments in this study, the X-ray crystal structure of the rhomboid protease GlpG was obtained from the Protein Data Bank (PDB: 3UBB) and the receptor was prepared for docking (3.4.1). 3UBB is a co-crystal structure of GlpG with the CAPF inhibitor, in which CAPF is covalently bound to the active site Ser201. Since no rhomboid structure with a non-covalent inhibitor is known, in order to perform the docking, the covalent bond between the phosphorus of CAPF and the oxygen of Ser201 was cleaved using the builder module and the structure was energy minimized. Then, according to the DOCK module in MOE the prepared Test set of saccharin derivatives was docked (3.4.3.1) using placement as triangle matcher and the AMBER12: EHT-forcefield.

Table 4.6: Test set of saccharins with 50 different leaving groups.



ID	leaving group (LG)	ID	leaving group (LG)	ID	leaving group (LG)	ID	leaving group (LG)
LG1		LG14		LG27		LG39	
LG2		LG15		LG28		LG40	
LG3		LG16		LG29		LG41	
LG4		LG17		LG30		LG42	
LG5		LG18		LG31		LG43	
LG6		LG19		LG32		LG44	
LG7		LG20		LG33		LG45	
LG8		LG21		LG34		LG46	
LG9		LG22		LG35		LG47	
LG10		LG23		LG36		LG48	
LG11		LG24		LG37		LG49	
LG12		LG25		LG38		LG50	
LG13		LG26					

Docking Results

The docking results showed that the saccharin derivatives interacted with the rhomboid pocket in a similar way to the CAPF inhibitor, with the saccharin ring occupying the S region and the LG extending towards the S' region (Figure 4.31A and Figure 4.31B). This may indicate that in relation to CAPF the LGs occupy the S' cavity and thus play a major role in inhibitor recognition, as shown in Figure 4.31A for the aryl carboxylate LG38. A significant interaction was assigned to the H-bond formation between the sulfoxide moiety of the saccharin scaffold and His150 and Gly199, displayed in the ligand interaction map of LG38 and the GlpG pocket in Figure 4.31D. An additional pi-pi interaction was observed between the aryl ring of LGs with aromatic carboxylic acids and Phe245 of

the rhomboid (Figure 4.31D). The distance between the aromatic rings involved in pi-pi interaction was 3.86 Å and the angle between the plane was 27° from being parallel (Figure 7.1). The generated poses for all Test Set molecules were scored by London dG and Affinity dG followed by energy minimization inside the enzyme pocket. The generated output database included all the molecules with their 10 best poses along with their scoring function (S). Using the descriptor module in database viewer, the number of heavy atoms (HA) was calculated for all molecules. Subsequently, the corresponding calculated ligand efficiencies (cLE) were calculated, which is the scoring function divided by the number of heavy atoms. As explained before, the ranking of the resulting poses was based on LE, followed by scoring function and ligand interactions (section 3.4.4).

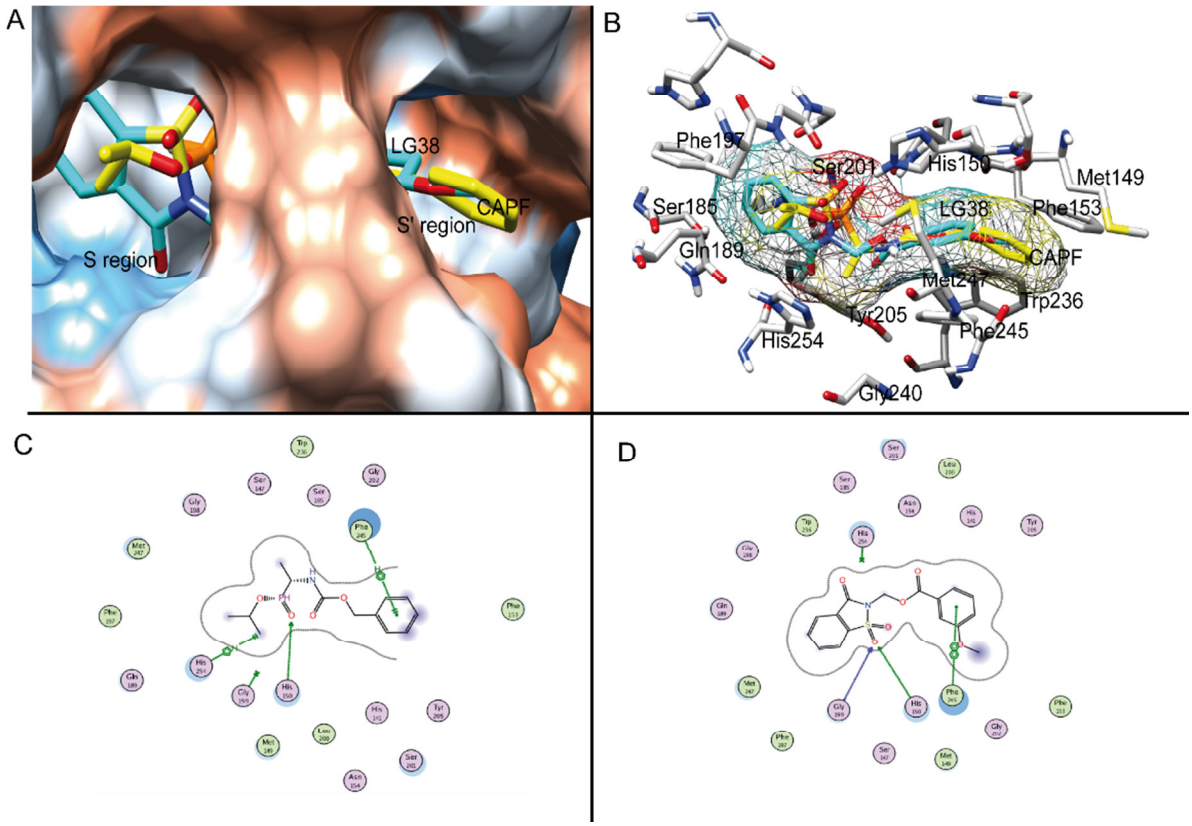


Figure 4.31: Comparison of the molecular docking results for a saccharin derivative and the CAPF inhibitor (PDB:3UBB). A) Molecular docking model of GlpG with CAPF and the saccharin derivative LG38 along the surface. The saccharin derivative resides in the GlpG pocket similar to CAPF, with the heterocyclic ring occupying the S region while the LG protrudes towards the S' region. B) Docking model showing the initial interactions of the inhibitors CAPF and LG38 surrounded by the neighbouring residues of the GlpG cleft. C) Ligand interaction map of GlpG with CAPF displaying interactions between CAPF and GlpG residues His150, His254 and Phe245. D) Ligand interaction map of GlpG with the saccharin derivative LG38. The map shows the formation of two H-bonds between the sulfoxide moiety of LG38 and His150 and Gly199, and a significant pi-pi interaction between the aryl ring of LG38 and Phe245. CAPF is shown in yellow color and the saccharin LG38 in cyan blue color. Surface colouring according to the hydrophobicity scale of Kyte and Doolittle: orange = hydrophobic; blue = hydrophilic. The image was prepared using Chimera.

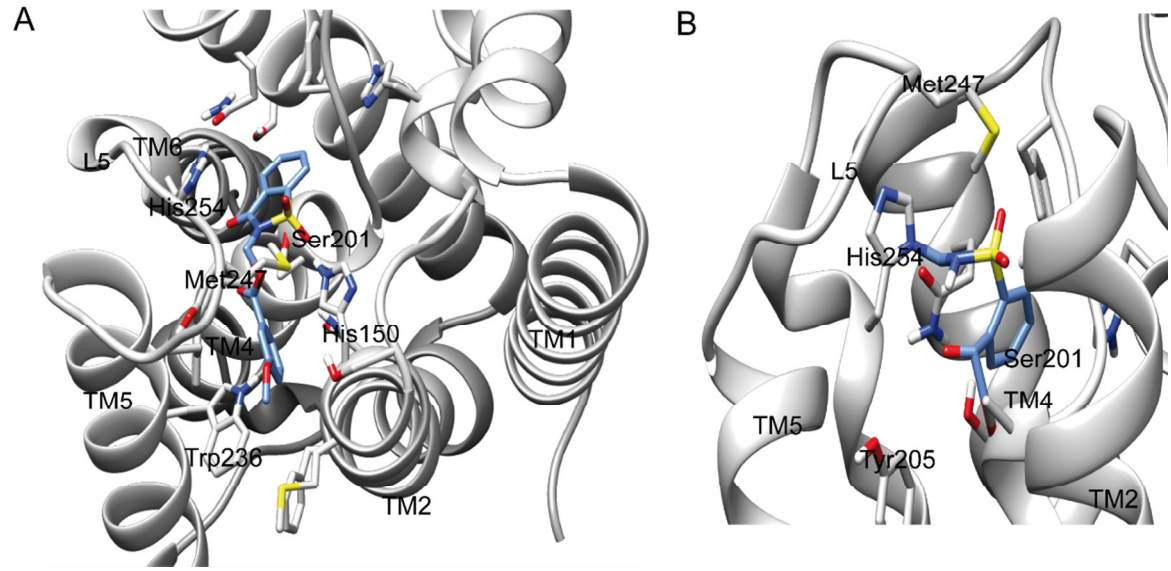
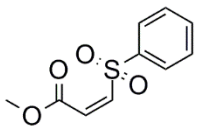
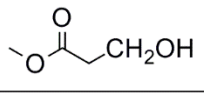
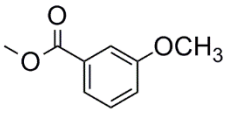
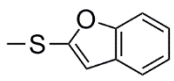


Figure 4.32: Molecular model of docked N-methylated saccharin in GlpG cleft before and after enzymatic reaction (PDB: 3UBB). A) This model represents the full structure of one Test Set saccharin inhibitor (LG 38) docked into GlpG, indicating the initial enzyme-inhibitor interaction, i.e. before the enzymatic reaction. B) This model indicates the end product of the enzymatic reaction between the saccharin LG 38 and GlpG based on the proposed reaction mechanism, with two covalent bonds formed between the saccharin and the GlpG active site residues Ser201 and His254. The saccharin LG38 is shown in blue color and the image was prepared using Chimera.

Table 4.7: Docking results of the Test set saccharins, displaying the five best leaving groups.

ID	LG	S-value	HA	cLE
LG47		-8.51	25	-0.34
LG36		-6.12	18	-0.34
LG38		-8.08	24	-0.33
LG1	−Cl	-4.68	14	-0.33
LG20		-7.59	23	-0.33

The best five leaving groups were ranked based on their cLE summarised in Table 4.7, and the results indicated that the carboxylates (LG47, LG36 and LG38) were the most favourable leaving groups. All the top ranked LGs possessed similar cLE due to a variable number of heavy atoms. Hence, considering the scoring function as a second ranking factor, the carboxylate leaving groups with an aryl ring in the extending chain (LG47 and LG38) were best. Moreover, to gain specificity for rhomboids, small carboxylate molecules like LG1 were ignored. It was also found that the energy minimized best pose for the aliphatic carboxylate chain of LG36 pointed into the S region of the

rhomboid pocket, opposite to the expected direction. LG47 and LG38 both had similar cLE and S-values (Table 4.7). However, by comparing the ligand interactions, it was observed that both ligands placed their saccharin core ring at the same position in the rhomboid pocket but that the extended leaving group chain of LG38 acquired an additional pi-pi interaction with the Phe245 residue of the GlpG pocket (Figure 4.33). Therefore, in consideration of all these factors, it was decided to synthesize saccharins with aryl carboxylate leaving groups for a preliminary test batch.

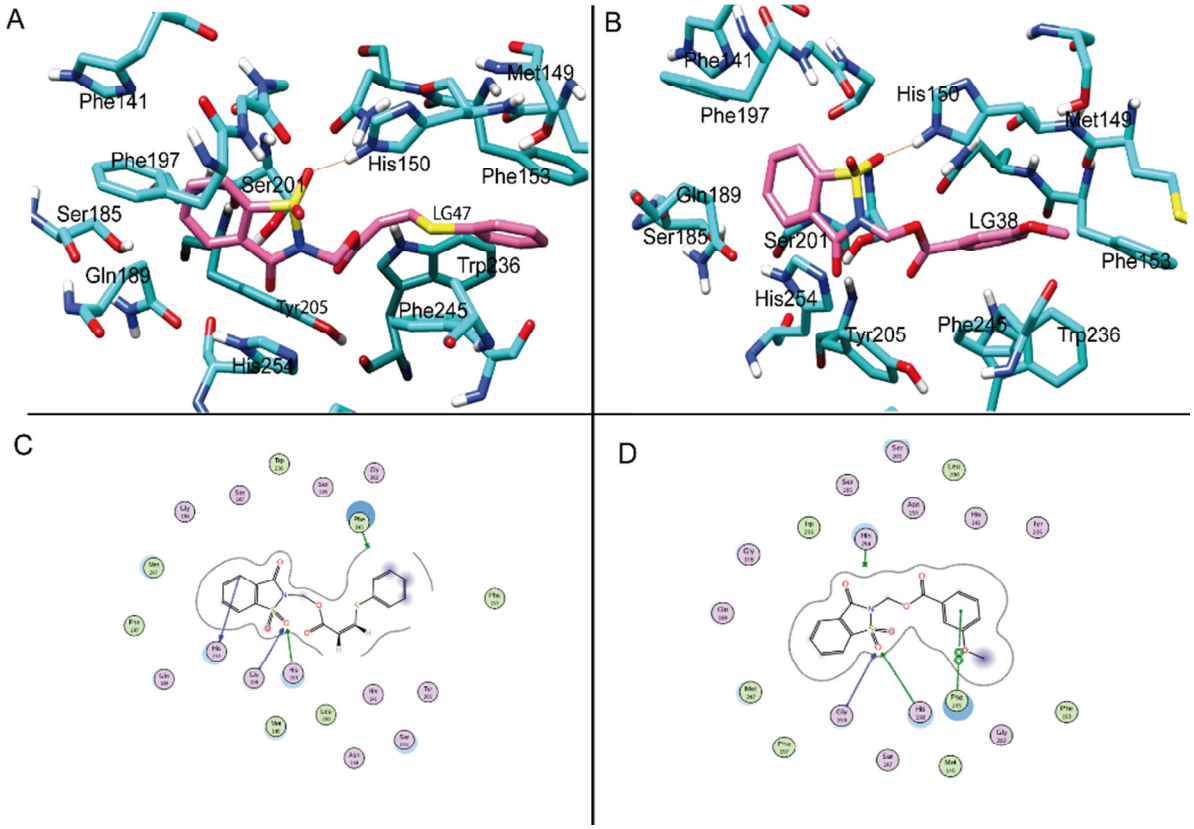


Figure 4.33: A comparative molecular docking model of the saccharin derivatives LG47 and LG38 in the GlpG pocket. A) Docking model of LG47. B) Docking model of LG38. C) Ligand interaction map of GlpG with LG47. The maps show the formation of two H-bonds between the sulfoxide moiety of LG47 with the GlpG residues His150 and Gly199, D) Ligand interaction map of GlpG with LG38. Besides two H-bonds with His150 and Gly199, LG38 also showed an additional pi-pi interaction between the extended aryl ring and Phe245 of GlpG.

Another class of leaving groups consisting of aryl sulfides and sulfones (LG20, LG21, LG22, LG29 and LG30) also contributed ligand interactions to the rhomboid pocket comparable to the aryl carboxylates (Figure 4.34). However, they lacked in scoring function (S-value). Nevertheless, to correlate the MOE docking results with the biological activities, some other saccharin derivatives aside from the aryl carboxylates such as LG1, LG3, LG21, LG22, LG29 and LG30 were also selected for the initial synthesis batch. Overall, ten compounds were selected and synthesized for the preliminary test batch, listed in Table 4.8.

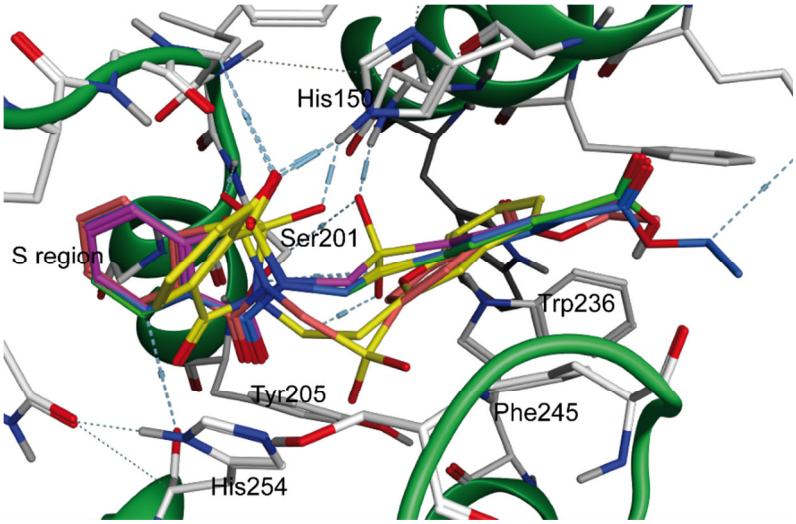
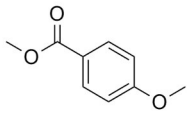
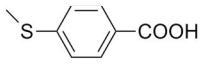
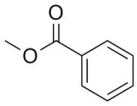
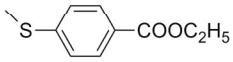
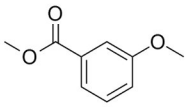
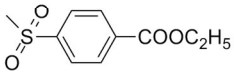
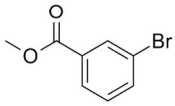
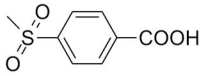


Figure 4.34: Molecular model of the rhomboid protease GlpG docked with saccharin derived aryl sulfides and sulfones. LG20 (yellow), LG21 (green), LG22 (blue), LG29 (pink) and LG30 (orange).

This first batch of saccharins was evaluated for biological activity in the GlpG *in vitro* activity assay. This showed that saccharins with the aryl carboxylate leaving groups were active against the rhomboid protease GlpG. Moreover, a difference in the potency of these aryl carboxylate derivatives was observed (the results of the activity assays are described in more detail in section 4.6.2.3). Therefore, a N-methylated saccharin with an aryl carboxylic acid as leaving group was selected as the lead molecule and further docking experiments were performed with derivatives of the aryl ring of the leaving group to gain insight into the best possible extensions as well as the inhibition efficacy of saccharin inhibitor with reformed electron withdrawing property of the leaving group. Since stronger electron withdrawing groups activate the saccharin core for attack by the active site serine. For this, a molecular database of 103 saccharin derived aryl carboxylates with variable extensions on the aryl ring was prepared (appendix 7.1.1). Docking experiments were performed using the DOCK module in MOE. From the docking results, it was observed that there were no significant interactions between variable aryl ring substituents and their neighbouring residues (appendix 7.1.2) except for the pi-pi interaction between the extended aromatic ring and Phe245. Moreover, there is an interesting feature in MOE called electrostatic surface map, which enables us to identify high value hydrophobic, and polar regions by using the non-linear Poisson-Boltzmann equation. These electrostatic regions are designated by variable colours where blue stands for positive regions (H-donor), red for negative (H-acceptor) and green for neutral (hydrophobic) regions (appendix 7.1.2). This was employed for analysing saccharin docking results and it was observed that the both aryl rings, one in the saccharin core structure and other in the LG fit perfectly into the favored hydrophobic region (appendix 7.1.3). According to electrostatic map, the para position might favor hydrophobic extensions and the ortho position may engage in proton donor/acceptor interactions. Small substitutions looked quite feasible at these positions while larger extensions might lead to molecular clashes or external exposure. Summarizing the data from the comparative study and all the facts described above, the N-methylated

saccharin-based acid derivatives with variable extensions (listed in Table 4.9, BSc5187-BSc5206) were synthesized and assayed for their biological activity.

Table 4.8: Preliminary batch of N-methylated saccharin derivatives.

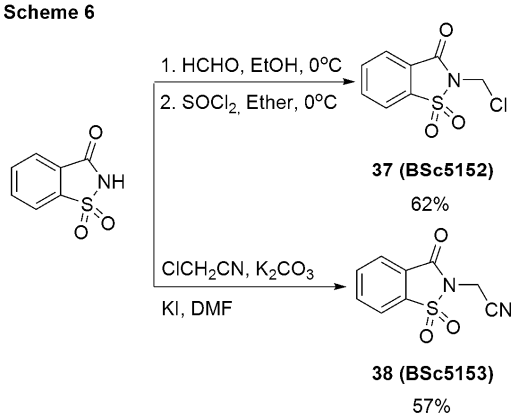
Entry no.	BSc no.	R	Entry no.	BSc no.	R
37	5152	—Cl	42	5157	
38	5153	—CN	43	5158	
39	5154		44	5159	
40	5155		45	5160	
41	5156		46	5161	

4.6.2.2 Synthesis of saccharin-derived inhibitors

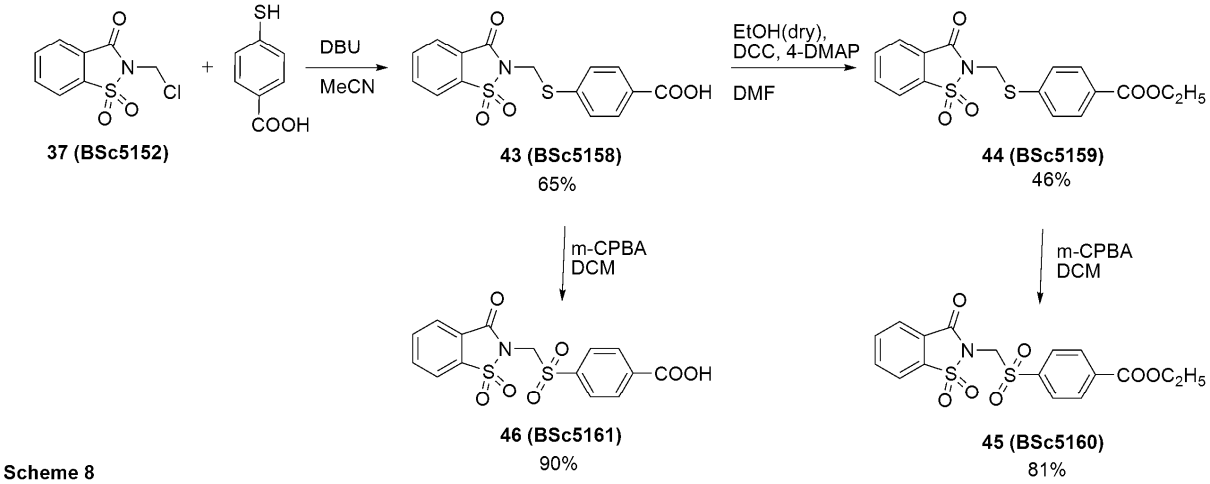
N-(chloromethyl) saccharin 37 (BSc5152) was prepared from saccharin by reacting it first with formaldehyde to obtain N-(hydromethyl) saccharin and then with thionyl chloride (Figure 4.35, scheme 6) [133]. N-(cyano methyl) saccharin 38 (BSc5153) was prepared by deprotonating saccharin using anhydrous potassium carbonate. The resulting salt was then reacted with chloroacetonitrile (Figure 4.35, scheme 6) [134]. Compounds 43-46 were prepared using the reported methods (Figure 4.35, scheme 7) [135]. Compound 43 (BSc5158) was obtained by alkylating the 4-mercaptobenzoic acid with N-chloromethyl saccharin in the presence of 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) and acetonitrile, followed by peracid oxidation of the resulting purified sulfide 43 to get compound 46 (BSc5161). Reaction product 43 was esterified to prepare compound 44 (BSc5159), which was further oxidised using peracid yielding compound 45 (BSc5160).

The desired aryl carboxylates (39-42 and 47-66) were synthesized by the known reaction of N-chloromethyl saccharin (37) with the appropriate carboxylic acids (Figure 4.35, scheme 8) in the presence of the base triethyl amine [136, 138]. The product yield depended on the particular acid used in the reactions. All synthesized saccharin-based acid derivatives are listed in Table 4.9. For detailed synthesis information and spectral analysis of the corresponding intermediate molecules and inhibitors, please refer to the method section 3.5.4 of this thesis.

Scheme 6



Scheme 7



Scheme 8

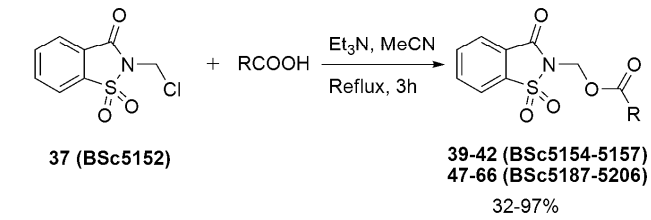
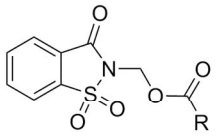
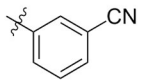
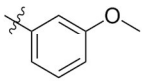
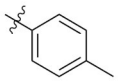
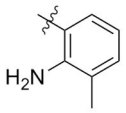
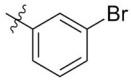
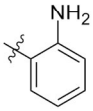
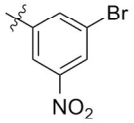
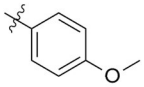
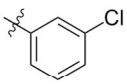
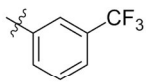
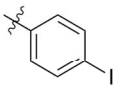
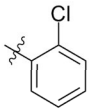
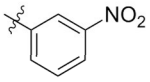
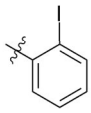
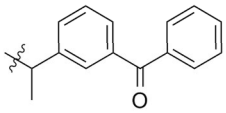
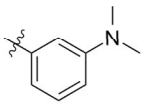
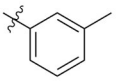
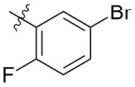
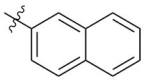
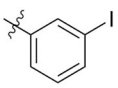
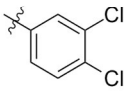
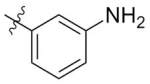


Figure 4.35: Synthesis schemes of N-methylated substituted saccharin-derived inhibitors. Scheme 6 shows the synthesis of N-(chloromethyl) saccharin (BSc5152) used as starting material for the synthesis of other saccharin derivatives, and the synthesis of N-(cyanomethyl) saccharin (BSc5153). Scheme 7 displays the general synthesis of saccharin derived aryl sulfides and sulfones (BSc5158-BSc5161). Scheme 8 shows the general synthesis of the N-methylated saccharin-based acid derivatives (BSc5154-5157 and BSc5187-BSc5206). The R groups are listed in Table 4.9.

Table 4.9: N-methylated saccharin-based acid derivatives.



Entry no.	BSc no.	R	Entry no.	BSc no.	R	Entry no.	BSc no.	R
39	5154		51	5191		59	5199	
40	5155		52	5192		60	5200	
41	5156		53	5193		61	5201	
42	5157		54	5194		62	5202	
47	5187		55	5195		63	5203	
48	5188		56	5196		64	5204	
49	5189		57	5197		65	5205	
50	5190		58	5198		66	5206	

4.6.2.3 Biological evaluation of N-methylated saccharin-derived inhibitors

Inhibitory activity of the preliminary batch of N-methylated saccharin derivatives

In order to analyse the inhibitory potency of the preliminary batch of saccharin derivatives, the compounds were evaluated in the GlpG *in vitro* activity assay (section 3.2.12). Initially, this class of inhibitors was also tested at a high concentration (250 µM), and the assay was performed with 50 µl reaction mixtures containing 0.35 µM GlpG. All inhibitors were pre-incubated with the enzyme for 30 min at 37˚C, followed by addition of the Gurken substrate (1.8 µM) and another 90 min of incubation. The reaction samples were then analysed by Western blotting using an antibody against the His-tag fused to the C-terminus of the chimeric Gurken substrate (Figure 4.36).

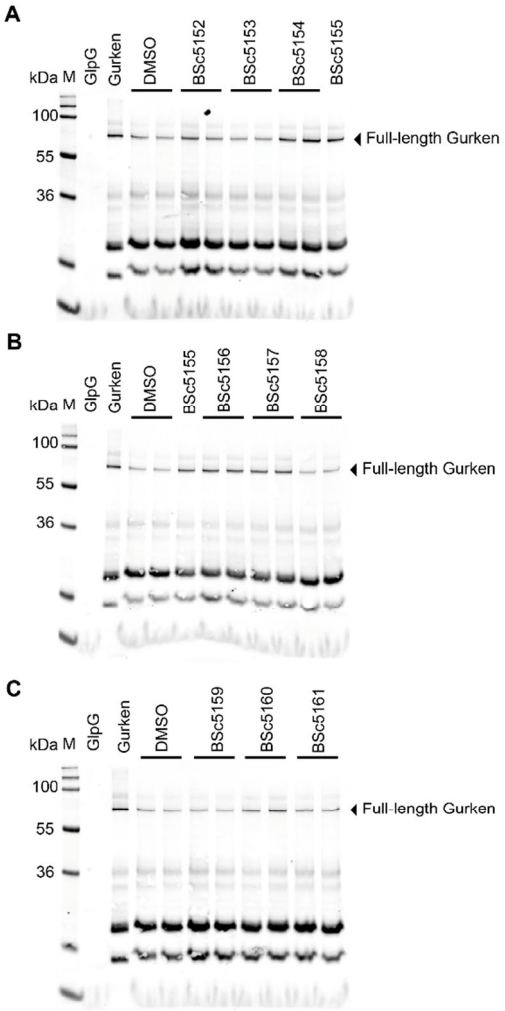


Figure 4.36: ***In vitro*** **activity assay with preliminary batch of N-methylated saccharin derivatives.** Western blot of activity assays with BSc5152 - BSc5161 at 250 µM and anti-His immunodetection of reaction samples dipicting the full length Gurken indicated by an arrow (~ 66 kDa). Inhibitor treated reactions containing BSc5154, BSc5155, BSc5156 and BSc5157 were able to retain the same signal intensity of the full-length Gurken as in the control reaction where only Gurken is present and are hence able to inhibit GlpG. M: Molecular weight marker.

In contrast to the GlpG activity assays described above, the readout for enzyme inhibition was based on the band intensity of the full-length Gurken substrate visible at around 66 kDa and not on the N-terminal Gurken reaction product. As shown in Figure 4.36A, incubation of the Gurken substrate with GlpG (DMSO vehicle control, lanes 4 and 5) reduced the band intensity of the full-length substrate as compared to a reaction mixture containing only Gurken (lane 3), indicating turnover of the substrate by the GlpG enzyme. However, in some reactions containing inhibitors the band intensity of the full-length Gurken substrate was retained as compared to the control reaction containing only Gurken (e.g., BSc5154, lanes 10 and 11), indicating that these inhibitors were able to inhibit GlpG. Overall, these experiments demonstrated that saccharins with aryl acid LGs (BSc5154-BSc5157) were active at 250 µM while all other leaving groups were inactive against GlpG (Figure 4.36). In order to analyse the inhibitory activity of the aryl carboxylates in more detail, these compounds were tested at lower concentrations down to 5 µM and it was found that the activity was dependent on the substitution of the aryl ring (data not shown). Hence, saccharin aryl carboxylates were selected as lead structures for

rhomboid inhibitors, and a second batch of 20 compounds bearing different substitutions on the aryl ring of the leaving group were synthesized (BSc5187-BSc5206, Table 4.9).

Inhibitory activity of the second batch of saccharin derivatives

The second batch of saccharins was initially also analyzed in the GlpG *in vitro* activity assay at a single concentration of 250 µM. In these assays, the band intensity of the Gurken NTF was again chosen as the readout for the inhibitory activity of the saccharin derivatives (Figure 4.37). On the SDS-PAGE, the full-length Gurken substrate and the NTF were visible around 66 kDa and 55 kDa, respectively. By quantification of the band intensity of the Gurken NTF, it was observed that all aryl carboxylates were able to inhibit GlpG at 250 µM (appendix 8.3.1). Therefore, all derivatives were further analysed at lower concentrations from 1-10 µM, and aryl carboxylates active at 10 µM were selected for determination of IC50 values.

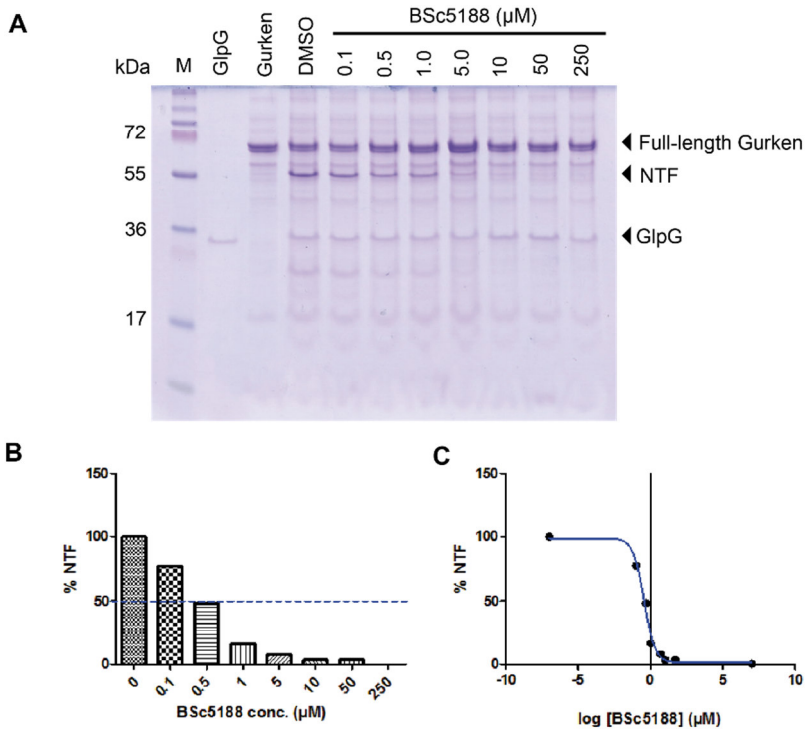


Figure 4.37: GlpG *in vitro* activity assay with the aryl carboxylate saccharin inhibitor BSc5188 (N=2). A) SDS-PAGE of the reaction products after incubation of GlpG and Gurken substrate with BSc5188 in defined concentrations (0.1 µM - 250 µM). The arrows indicate the full length Gurken substrate (~ 66 kDa), the N-terminal Gurken fragment (NTF, ~ 55 kDa), and GlpG (~ 33 kDa). M: Molecular weight marker. B) Plot with BSc5188 concentration versus % NTF band intensity. The dashed blue line indicates 50% enzyme inhibition. C) Nonlinear regression curve of the log (BSc5188 concentration) versus % NTF band intensity. The gel was stained with coomassie brilliant blue, quantified with ImageJ, and the values were plotted in GraphPad Prism. N= number of experiments.

To determine the IC50 values of the thirteen aryl acid saccharin inhibitors active at 10 µM, defined concentrations of the saccharin derivatives (0.1 µM - 250 µM) were evaluated in the GlpG *in vitro* activity assay as described in section 4.3.3. As an example, Figure 4.37 shows the GlpG activity assay

performed with the inhibitor BSc5188. IC50 values were calculated by fitting a nonlinear regression curve to a plot of log (BSc5188 concentration) versus the % NTF band intensity (Figure 4.37C). The IC50 value for BSc5188 was 0.3 µM and number of experiments (N) were 2 to ensure the inhibition potential in the GlpG assay. Table 4.10 shows the IC50 values for all aryl acid saccharin inhibitors. To confirm the IC50 value, similar to BSc5188, the experiment was performed two times for some other saccharin inhibitors, BSc5156, BSc5187, BSc5195 and BSc5203 and obtained similar results. All thirteen saccharin derivatives had IC50 values around or lower than 1 µM.

Table 4.10: IC50 values for rhomboid protease (GlpG) inhibition by novel saccharin-based acid derivatives.

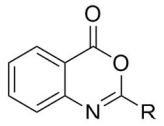
Entry No.	BSc No.	IC50 (µM)
41	5156	0.61
47	5187	0.54
48	5188	0.36
52	5192	0.78
54	5194	0.64
55	5195	0.20
56	5196	0.56
57	5197	0.64
58	5198	0.80
62	5202	1.26
63	5203	0.33
64	5204	0.41
65	5205	0.61

4.6.3 Benzoxazinone-based inhibitors

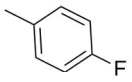
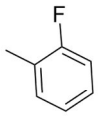
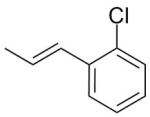
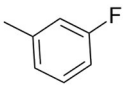
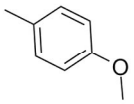
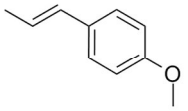
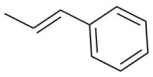
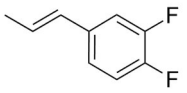
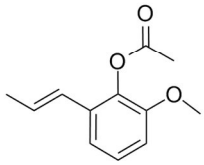
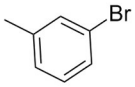
4.6.3.1 Design and docking studies of benzoxazinone-based inhibitors

Benzoxazin-4-ones have been identified as alternate substrate inhibitors (section 1.3.2), and 2-substituted benzoxazin-4-ones have been effectively used as heterocyclic acylating agents against serine proteases such as HLE, α-chymotrypsin, and cathepsin G [113]. The mechanism of inhibition involves the formation of an o-acyl enzyme intermediate (Figure 4.38). The nucleophilic serine reacts with the C-4 carbonyl of benzoxazinone, which results in opening of the heterocyclic ring and formation of the o-acyl enzyme intermediate (Figure 4.38) [113]. Formation of the o-acyl enzyme intermediate has been verified by spectral analysis and X-ray crystallography studies [113]. The enzyme selectivity and potency of acylating agents is promoted by fast acylation and slow deacylation, which is dependent on the substitution of the aromatic ring and the C-2 position in case of benzoxazinones [110]. The advantage of benzoxazin-4-ones is that the core structure consists of two fused aromatic rings, which allows for broad structural variations and optimization with respect to the target enzyme. For the design of rhomboid inhibitors, the 2-alkyl or aryl substituted benzoxazinones template and its derivatives (shown in Table 4.11) were considered as a starting point for the docking study and synthesis.

Table 4.11: Template of 2-substituted benzoxazin-4-ones and its derivatives, selected for the design of rhomboid inhibitors.



R = alkyl or aryl group

Entry no.	BSc no.	—R	Entry no.	BSc no.	—R
67	5170		74	5177	
68	5171	—CH_3	75	5178	
69	5172		76	5179	
70	5173		77	5207	
71	5174		78	5208	
72	5175		79	5209	
73	5176				

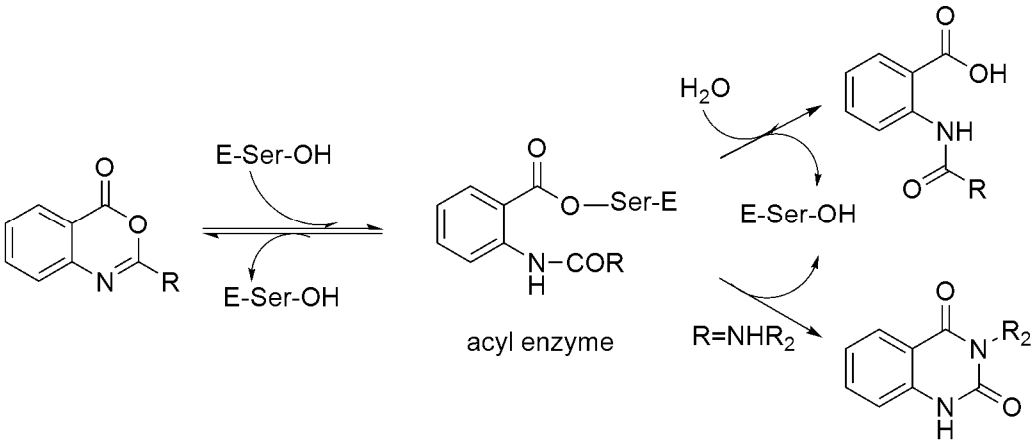


Figure 4.38: Mechanism of inhibition of serine proteases by 2-substituted benzoxazinone derivatives. Adopted from [110].

Prior to chemical synthesis, the benzoxazinones derivatives listed in Table 4.11 were docked into the rhomboid protease pocket. Molecular docking studies were performed using the DOCK module of MOE (section 3.4). According to the reaction mechanism shown in Figure 4.38, the acyl enzyme intermediate could deacylate into two different products and the final reaction product is uncertain. Therefore, similar to the docking studies of saccharin derivatives, we focused on the initial interactions of benzoxazinones with the rhomboid cavity rather than on the final product.

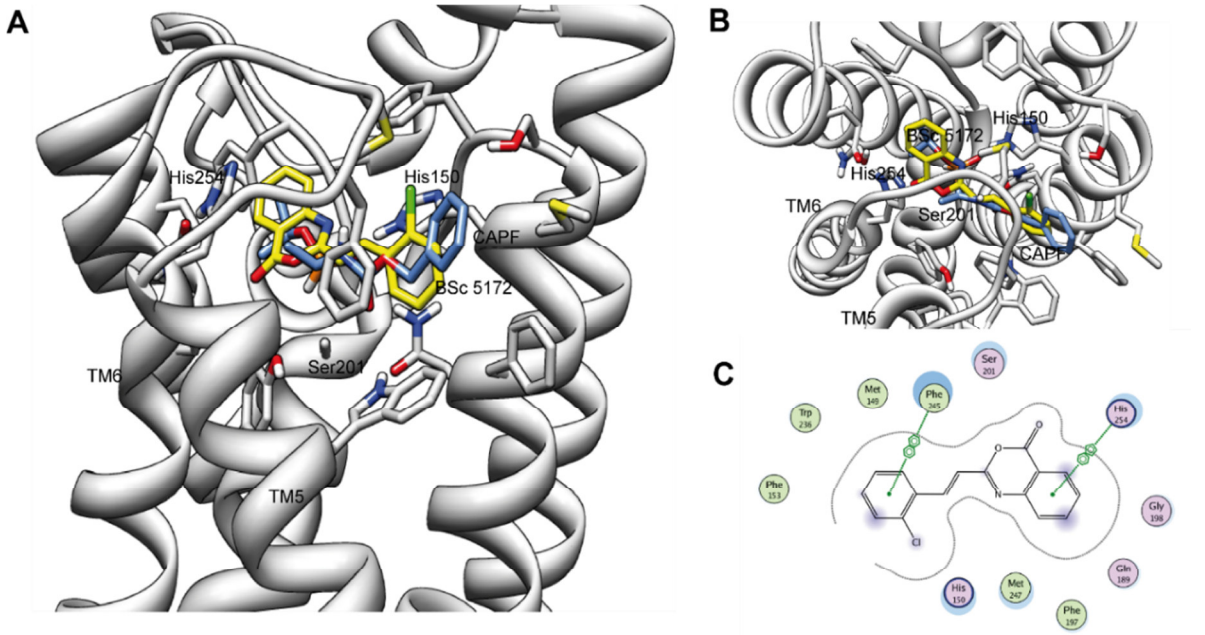


Figure 4.39: Molecular docking model of the rhomboid protease GlpG in complex with a benzoxazinone-based inhibitor. A) Front view of GlpG docked with the benzoxazinone-based inhibitor BSc5172 and with CAPF. BSc5172 is shown in yellow and CAPF in blue. The model indicated that both compounds occupied a similar space in the rhomboid pocket and were not exposed to the external surface. B) Top view of BSc5172 and CAPF in the GlpG pocket. C) Ligand interaction map of BSc5172 in the GlpG pocket. Significant interactions were observed with neighbouring residues His254 and Phe245. Green: hydrophobic, pink: polar, red: exposed.

Using the co-crystal structure of the *E. coli* rhomboid protease GlpG with the phosphonofluoridate inhibitor CAPF (PDB: 3UBB) docking experiments were performed, which involved the preparation of the molecule database, preparation of the receptor, docking with the DOCK workflow using the MMFF94x forcefield, and scoring with London dG and Affinity dG. Output data was then ranked based on ligand efficiency (section 3.4.4) and the best three compounds are listed in Table 4.12. This revealed derivative BSc5172 as the most favourable of all 2-substituted benzoxazinones shown in Table 4.11. A comparative analysis of the protein/ligand docking results of BSc5172 and CAPF indicated that BSc5172 was adequately fitting into the binding pocket of the enzyme and was not exposed to the external environment (Figure 4.39). Moreover, close interactions of BSc5172 with the neighbouring residues His254 and Phe245 as shown in the ligand interaction map suggested to explore this scaffold further (Figure 4.39C). However, for a comparative study, all the derivatives listed in Table 4.11 were synthesized and evaluated for their biological activity.

Table 4.12: Docking results of three best 2-substituted benzoxazin-4-ones derivatives.

Inhibitor	S-value	LE
BSc5172	-6.3280	-0.3164
BSc5171	-3.7418	-0.3118
BSc5208	-6.2264	-0.2965

4.6.3.2 Synthesis of benzoxazinone-based inhibitors

The benzoxazinone based inhibitors listed in Table 4.11 were obtained by methods shown in schemes 9 and 10 (Figure 4.40). For 2-substituted benzoxazinones, a solution of anthranilic acid in dry pyridine was reacted with the corresponding acyl/benzoyl chloride at 0°C (scheme 9) [182, 183], followed by neutralization with aqueous sodium bicarbonate solution to remove the unreacted acid. The deposited solid product after completion of the reaction or after pouring onto ice-water was filtered out to obtain compounds 67 and 69-79. Compound 68, 2-methyl benzooxazin-4-ones (BSc5171), was obtained by performing a microwave reaction of anthranilic acid with acetic anhydride (scheme 10).

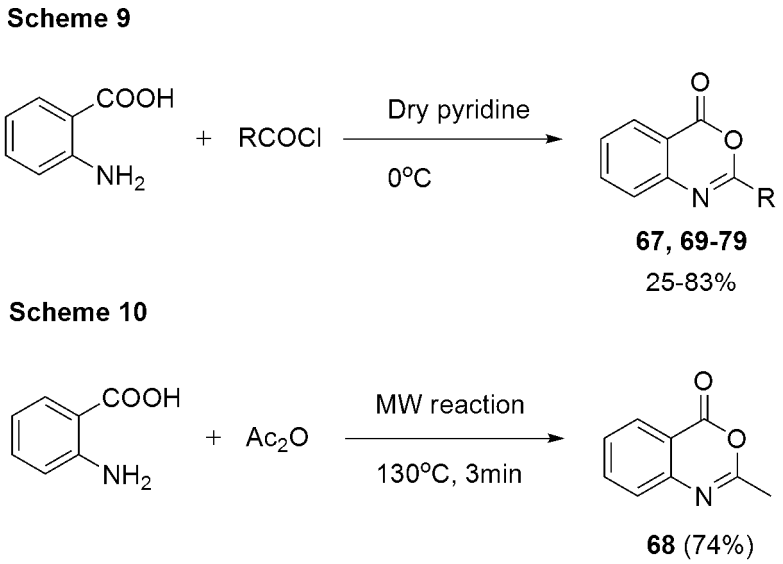


Figure 4.40: Synthesis of 2-substituted benzoxazin-4-ones inhibitors. Scheme 9 shows the synthesis of 2-alkyl/aryl substituted benzoxazinones (67, 69-79) obtained by the reaction of corresponding acyl/benzoyl chlorides with anthranilic acid. Scheme 10 shows the synthesis of 2-methyl benzooxazin-4-ones (68, BSc5171). The R groups are listed in Table 4.11.

For detailed synthesis information and spectral analysis of the corresponding intermediate molecules and inhibitors, please refer to the method section 3.5.5 of this thesis.

4.6.3.3 Biological evaluation of benzoxazinone-based inhibitors

In order to analyse the inhibitory potency of the benzoxazinone-based derivatives, the compounds were evaluated in the GlpG *in vitro* activity assay as described in section 3.2.12 at a single concentration of 250 µM. By analysing the reaction samples on an SDS-PAGE and comparing the protein band intensities of the full-length Gurken substrate (66 kDa) and the NTF (55 kDa), it was

concluded that BSc5172, BSc5174 and BSc5207 displayed some inhibitory activity at 250 µM (Figure 4.41). Hence, these compounds were further selected to determine their IC50 values.

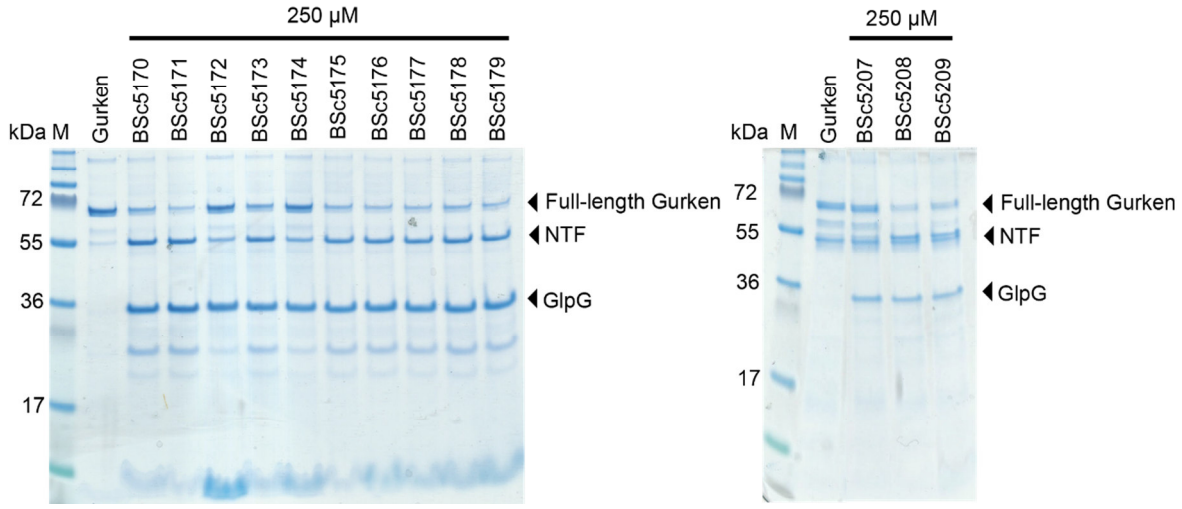


Figure 4.41: GlpG *in vitro* activity assay with benzoxazinone-based compounds. SDS-PAGE gel showing inhibitory activity of BSc5172, BSc5174 and BSc5207 at 250 µM. The band intensities of the full-length Gurken substrate for these three compounds were comparable to the control reaction containing only Gurken substrate (lane 2). In addition, the band intensity of the NTF was reduced in the reaction samples containing these three derivatives as compared to all other derivatives, indicating inhibitory activity. However, note that these experiments did not include a vehicle control reaction precluding a definitive assessment of the NTF band intensities. The arrows indicate the full length Gurken substrate (~ 66 kDa), the N-terminal Gurken fragment (NTF, ~ 55 kDa), and GlpG (~ 33 kDa). The gel was stained with coomassie brilliant blue.

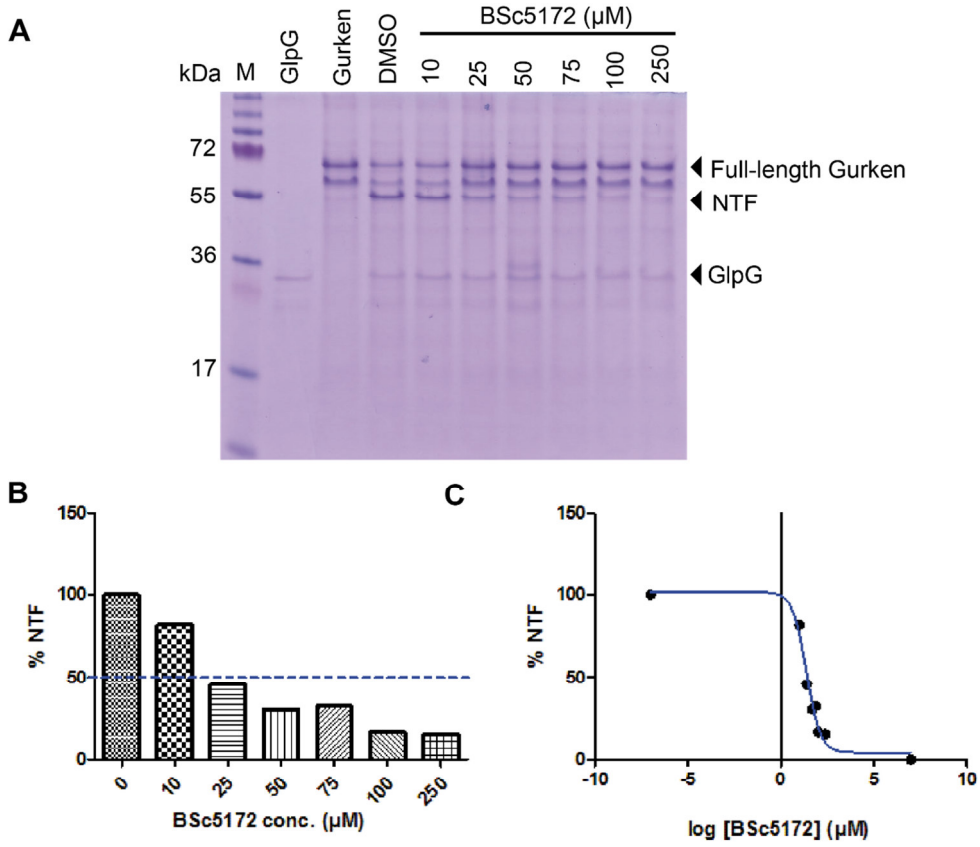


Figure 4.42: GlpG *in vitro* activity assay with the benzoxazinone-based inhibitor BSc5172 (N=1). A) SDS-PAGE of the reaction products after incubation of GlpG and Gurken substrate with BSc5172 in defined concentrations (10 µM - 250 µM).

The arrows indicate the full length Gurken substrate (~ 66 kDa), the N-terminal Gurken fragment (NTF, ~ 55 kDa), and GlpG (~ 33 kDa). M: Molecular weight marker. B) Plot with BSc5172 concentration versus % NTF band intensity. The dashed blue line indicates 50% enzyme inhibition. C) Nonlinear regression curve of the log (BSc5172 concentration) versus % NTF band intensity. The calculated IC50 value was 25 µM. The gel was stained with coomassie brilliant blue, quantified with ImageJ, and values were plotted in GraphPad Prism. N= number of experiments.

To determine the IC50 values of the three benzoxazinone-based inhibitors, defined concentrations (10 µM - 250 µM) of the compounds were evaluated in the GlpG *in vitro* activity assay as described in section 4.3.3. As an example, Figure 4.42 shows the results for the derivative BSc5172. IC50 values were calculated by fitting a nonlinear regression curve to a plot of log (BSc5172 concentration) versus the % NTF band intensity (Figure 4.42C). The IC50 value for BSc5172 was 25 µM, the values for the other two inhibitors were ≥ 50 µM (listed in Table 4.13). In conclusion, some benzoxazinone-based compounds were found to be active GlpG inhibitors, but their potency was rather low compared to the saccharin derivatives described in this thesis.

Table 4.13: IC50 values for of rhomboid protease (GlpG) inhibition by novel benzoxazinone-based compounds (N=1).

Entry No.	BSc No.	IC50 (µM)
41	5172	25
47	5174	47
48	5207	80

4.7 Virtual screening for novel rhomboid inhibitors

4.7.1 Virtual screening based on DOCKTITE in MOE

Virtual screening (VS) is a computational technique, which allows to screen large libraries of small molecules (section 1.4) to identify the most favorable candidates for a protein receptor. Besides the candidate-based approach described above, virtual screening was also used in the search for novel rhomboid protease inhibitors in the present study. This was performed with the DOCKTITE module for covalent docking in MOE (section 3.4.3.3). As explained before, DOCKTITE involves several steps:

1) Warhead dataset preparation: the design of warheads accomplished via MOE's MedChem transformations (MCT) application (section 3.4.3.3) was based on the binding mode of the corresponding ligand with consideration of possible rearrangements and release of leaving groups. Due to usage restrictions of MCT it was required to design different warheads for the same class of compounds, which followed variable binding mechanisms. For example, binding mechanism of saccharin based inhibitors involves the covalent bond formation with receptor and afterwards featuring release of leaving group (Figure 4.43) and because of restricted usage it was not possible to use the one warhead reaction set for all class of saccharin inhibitors.

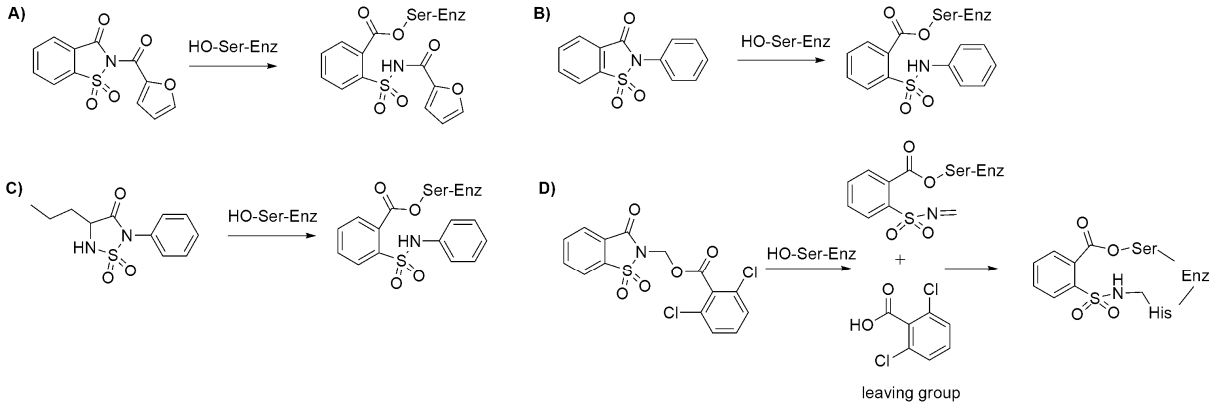


Figure 4.43: Examples of saccharin based inhibitors following different reaction rules. A) AcylSaccharin_LG_to_Tantal. B) ArylSaccharin_LG_to_Tantal C) ArylSaccharin2_LG_to_Tantal D) Saccharin_LG_to_Tantal.

Therefore, in the present study, we have employed 34 different warheads (Figure 4.44) to screen the extensive molecular library. The possible binding mode is called 'reaction rule' here which was followed by the corresponding ligand such as Phosphan_to_Tantal or Arylsaccharin_LG_to_Tantal (Figure 4.44). Moreover, to validate the warhead screen (Wh_screen) set, a benchmark database consisting of true and false hits was prepared with 112 molecules. Subsequently, this benchmark database was allowed to pass through the warhead filter (section 3.4.3.3), and it was confirmed that all reactions were running accurately as all positive hits passed through the warhead filter and all false hits did not.

2) Selection of the molecular library: the molecular library used in the search for rhomboid protease inhibitors was adopted from the ZINC database. The ZINC database is an online library of commercially available lead-like compounds for virtual screening. Table 4.14 indicates the features of the adopted library from the ZINC database (ref http://www.ncbi.nlm.nih.gov/pubmed/10649345).

Table 4.14: Selected molecular library to selected for rhomboid inhibitors.

Molecular library Features	Data
Source	ZINC database
Database size	2,419,472
Category	Leads now
Criteria	p.mwt <= 350 and p.mwt >= 250 p.xlogp <= 3.5 and p.rb <= 7
Subset ID	21

3) Preparation of the molecular database: the downloaded molecular library consisted of 21 subsets in the 'sdf' file format. All 'sdf' files were converted to 'mdb' files using MOE.

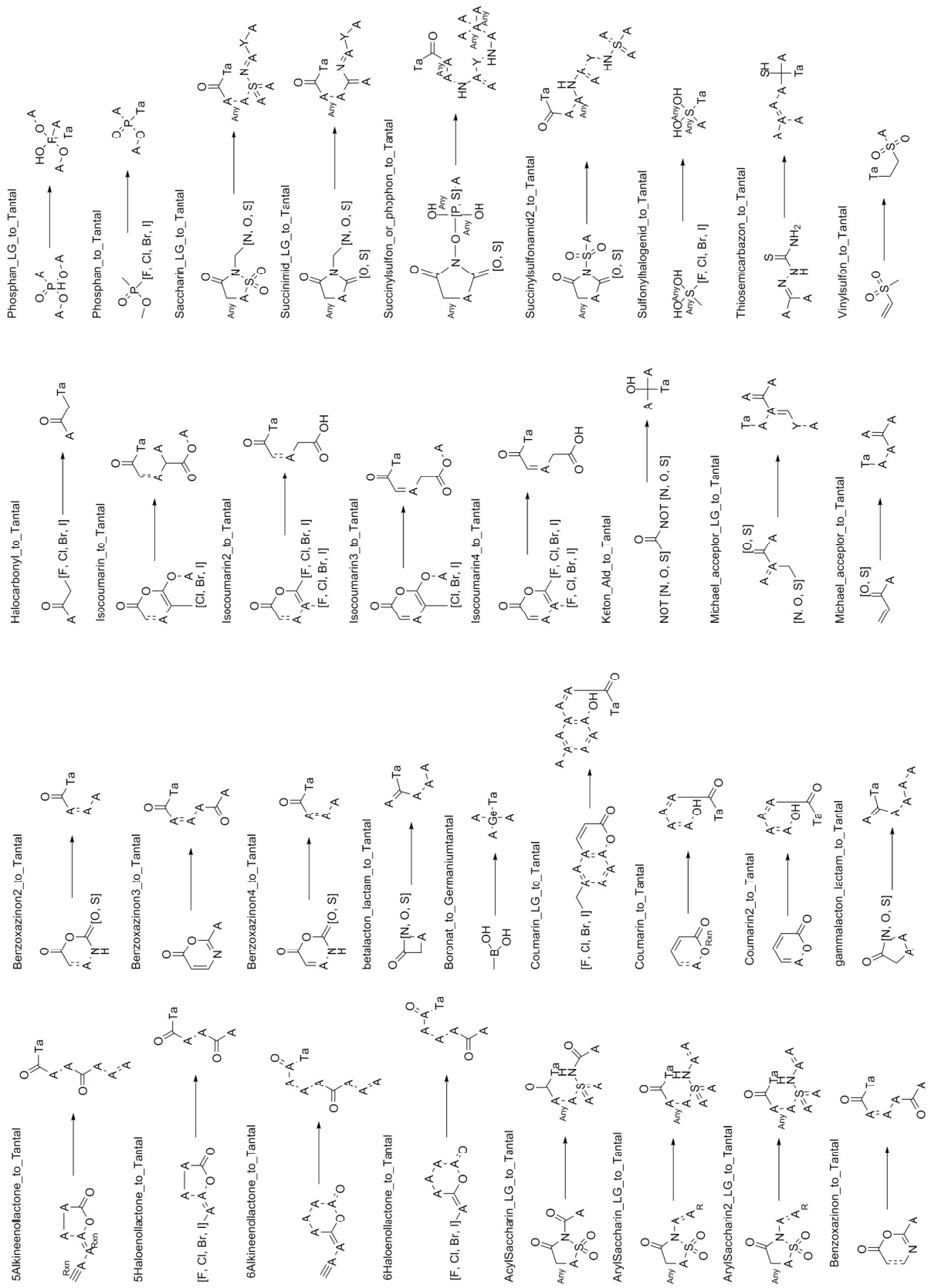
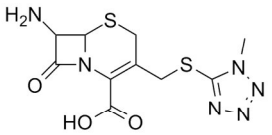
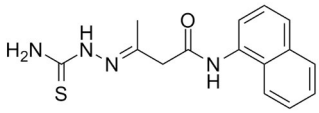
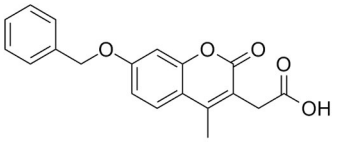
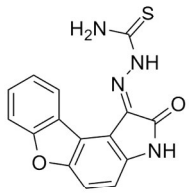
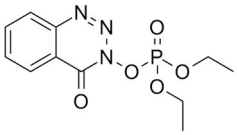


Figure 4.44: Warhead dataset of 34 variable warheads distinguished by their catalytic mechanism. These warheads were used to filter the molecular library in the search of novel rhomboid inhibitors.

4) Preparation of the receptor protein: the DOCKTITE module involved the use of two types of receptor proteins. First, the receptor (3UBB_prep) was prepared by protonation and energy minimization in MOE as explained in section 3.4.1. Second, the active site residue bearing the nucleophilic side chain of the prepared receptor (Ser201 in the rhomboid protease GlpG) was detached from the rest of the receptor molecule (3UBB_Ser_del) and its native position was fixed using pharmacophore (ph4) constraints.

5) DOCKTITE: for docking in MOE using DOCKTITE, all 21 molecular data subsets were first screened using the Wh_filter dataset containing 34 variable warheads (Figure 3.3 and section 3.4.3.3). This was followed by side chain attachment resulting into modified ligands, a step that included the attachment of the nucleophilic Ser side chain of the receptor to the Ta-tagged electrophilic center of the ligand, the conversion of Ge-tags to boron, and the removal of Yttrium (Y)-tagged leaving groups. During this step, stereoisomer (R or S) databases called attached_ligands_r and attached_ligands_s was generated for prochiral ligands. These were treated as individual molecular databases in the later steps. A pharmacophore (ph4) query was generated based on the original inhibitor CAPF in rhomboid pocket and used for docking. After the side chain attachment, a pharmacophore-guided docking was performed using the DOCK module in MOE (section 3.4.3.3). This involved five steps: generation of conformations, placement, scoring, pose refinement and rescoring. Every entry was docked using ph4-placement and Affinity dG scoring, which resulted in 100 poses. These poses were refined using GridMin because of its faster speed compared to forcefield and the force was kept 1e+100, to ensure fixation despite of clashes. The resulted poses were rescored using Affinity dG. In the next step, the Ser residue was disconnected from the chimeric molecule, and the rescoring of the poses was performed using Affinity dG with consideration of only one output conformation per ligand. The resulting best scored poses were then sorted by their calculated ligand efficiency (cLE = S-value / HA) (section 3.4.4). Then all docked and rescored databases of all the initial 21 datasets were merged, maintaining the different databases for stereoisomers as merged_r and merged_s.

Table 4.15: Virtual screening compounds selected as putative rhomboid protease inhibitors.

ID	Mol name	Structure
VS1	ZINC02809614	
VS2	ZINC33751351	
VS3	ZINC05442813	
VS4	ZINC006000122	
VS5	ZINC02573666	

Furthermore, since aldehyde-, ketone- and halocarbonyl-based inhibitors are known to be highly unspecific and reactive, these classes of inhibitors were omitted from further refinement and were all deleted from the "merged" databases. Additionally, to avoid overrepresentation of particular warhead classes, only 100 (20%) of the ligands of a certain class were considered for refinement and the others were deleted. This resulted in a total of 10,000 ligands to be refined, and redocking was performed with an optimized ph4 (boundary of ph4 was limited to 0.4 Å), forcefield refinement (MMFF94x, force constant: 1e+9) and Affinity dG. A maximum of 10 poses were refined per ligand, and this resulted in separate R-steroisomer and S-steroisomer result databases (rescored_final_r and rescored_final_s). Based on cLE, the top ten ligands were selected from the rescored_final_r and the top five ligands from the rescored_final_s databases (appendix 7.1.4). After further consideration of their drug-like properties, only five out of these shortlisted and commercially available compounds were selected and purchased (VS1-VS5, listed in Table 4.15). These compounds were evaluated for their inhibitory potency in the GlpG *in vitro* activity assay.

4.7.2 Biological evaluation of virtual screening compounds

The putative rhomboid inhibitors VS1-VS5 resulting from the virtual screening were evaluated in the GlpG *in vitro* activity assay as described in section 3.2.12 at a single concentration of 250 µM. In this

assay, only compound VS5 displayed inhibitory activity against GlpG (Figure 4.45). Incubation of VS5 with GlpG resulted in accumulation of the full-length Gurken substrate and reduced generation of the Gurken NTF as compared to the DMSO vehicle control reaction. This effect of VS5 was comparable to the isocoumarin JLK-6, which was used as a positive control. Compounds VS1-VS4 did not show any activity in the assay (Figure 4.45). Subsequently, the IC50 value of VS5 was determined as described in section 4.3.3 and found to be approximately 15 µM (Figure 4.46). This value was based on one experiment and we would need many more concentrations between 10-30 µM and replications to provide an exact number. However, two phosphate inhibitors of rhomboid proteases are already known, diisopropyl fluorophosphate (DFP) and the phosphonofluoridate inhibitor CAPF, which have displayed IC50 ~50 µM [89] but here we identified the novel and more potent phosphate inhibitor VS5.

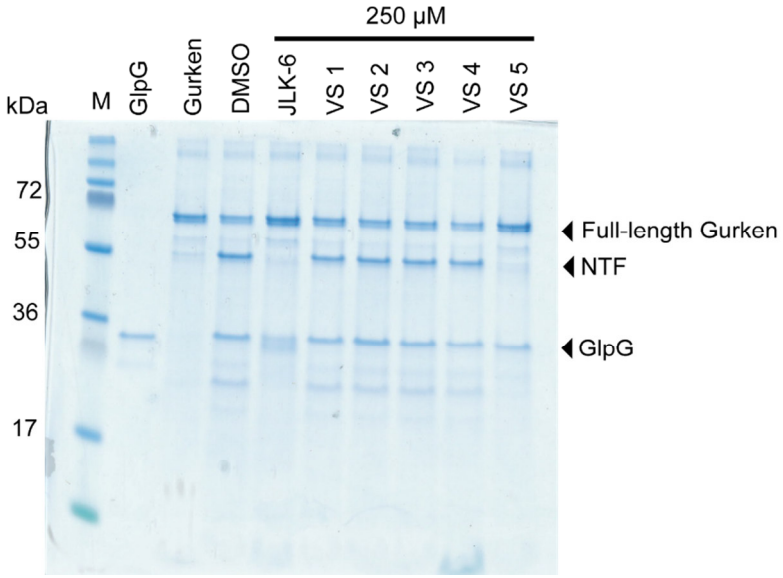


Figure 4.45: GlpG *in vitro* activity assay with the virtual screening compounds VS1-VS5. The SDS-PAGE gel shows that only VS5 inhibits GlpG. In contrast, the compounds VS1-VS4 displayed no inhibitory activity at 250 µM. After incubation of GlpG with VS5, the band intensity of the full-length Gurken substrate was increased and the band intensity of the Gurken NTF was reduced as compared to the DMSO vehicle control indicating enzyme inhibition. Moreover, the band intensities of the full-length Gurken and the NTF after VS5 treatment were comparable to the positive control JLK-6, which represented 100% enzyme inhibition at 250 µM. The arrows indicate the full-length Gurken substrate (~ 66 kDa), the N-terminal Gurken fragment (NTF, ~ 55 kDa) and GlpG (~ 33 kDa). M: Molecular weight marker. M: Molecular weight marker. The gel was stained with coomassie brilliant blue.

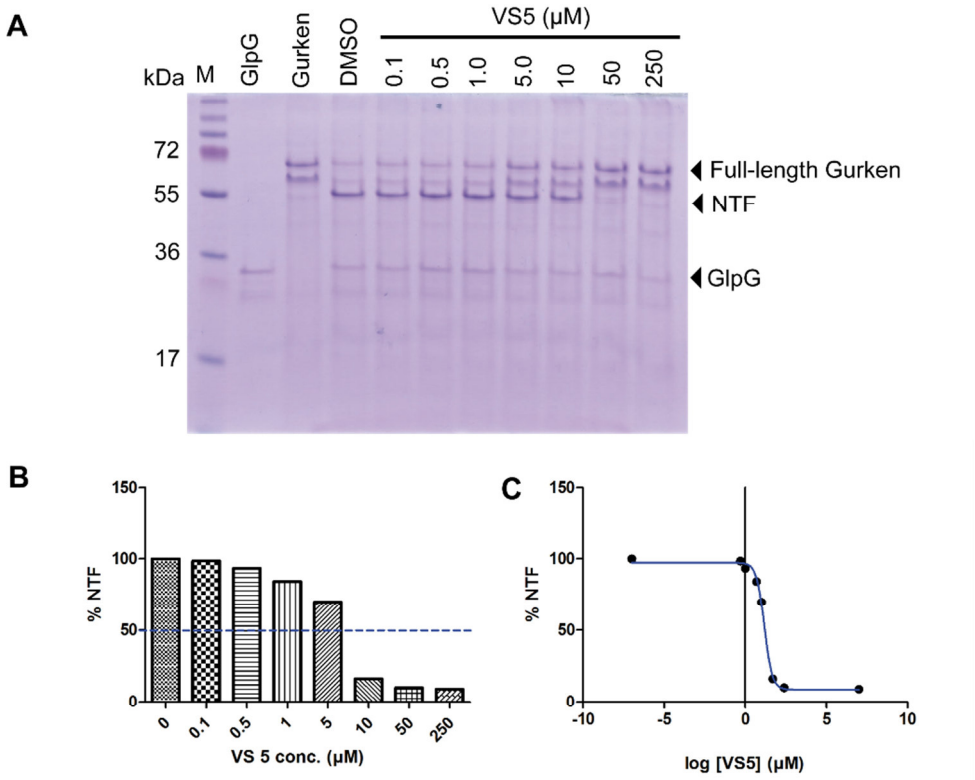


Figure 4.46: GlpG *in vitro* activity assay with the virtual screening compound VS5 (N=1). A) SDS-PAGE of the reaction products after incubation of GlpG and Gurken substrate with VS5 in defined concentrations (0.1 µM - 250 µM). The arrows indicate the full-length Gurken substrate (~ 66 kDa), the N-terminal Gurken fragment (NTF, ~ 55 kDa) and GlpG (~ 33 kDa). M: Molecular weight marker. B) Plot with VS5 concentration versus % NTF band intensity. The dashed blue line indicates 50% enzyme inhibition. C) Nonlinear regression curve of the log (VS5 concentration) versus % NTF band intensity. The IC50 value was calculated to be approximately 15 µM. The gel was stained with coomassie brilliant blue, quantified with ImageJ, and values were plotted in GraphPad Prism. N= number of experiments.

4.7.3 Cytotoxicity of the virtual screening compound VS5

The virtual screening compound VS5 was able to inhibit GlpG in the *in vitro* activity assay. However, this compound belongs to the phosphate class of inhibitors, which frequently cause cellular toxicity because of their highly reactive and unspecific nature [184]. Hence, a cell toxicity assay was performed with compound VS5 as explained in section 3.3.9. On the first day, human embryonic kidney cells (HEK293T) were seeded in a 96-well plate (30000 cells/well). The following day, the cells were treated with compound VS5 in ten increasing concentrations (0 µM – 60 µM) in triplicates. On the third day, the alamarBlue® reagent was added to the cells. alamarBlue® (resazurin) is a cell viability indicator, which relies on the antioxidant/reducing capacity of living cells. Live cells convert resazurin into the fluorescent molecule resorufin, which can be quantified with a multiplate reader. The toxicity assay did not show any reduction in cell viability up to a concentration of 60 µM VS5 (Figure 4.47). The same experiment was repeated once and obtained the same result with no reduction in cell viability up to 60 µM of VS5. Hence, VS5 may be considered a non-toxic compound in the concentration range that effectively inhibited the enzymatic activity of the rhomboid GlpG *in vitro*.

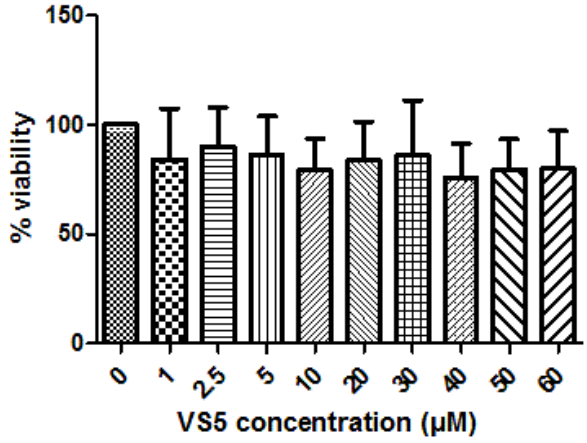


Figure 4.47: Cell toxicity assay with the virtual screening compound VS5 (N=2). HEK293T cells were treated with increasing concentrations of VS5 and cell viability was determined with the alamarBlue® reagent. The plot shows % cell viability as a function of the VS5 concentration. No cellular toxicity was observed up to a concentration of 60 µM VS5. Values were plotted in GraphPad Prism. N= number of experiments.

4.8 Selectivity of new novel rhomboid inhibitors over soluble serine proteases (α-chymotrypsin)

To investigate the selectivity of the novel rhomboid inhibitors over soluble serine proteases, the newly identified inhibitors (saccharins, benzoxazinones and VS5) were examined for their activity in a α-chymotrypsin *in vitro* assay as explained in section 3.2.14. This assay is based on a known peptidic substrate of α-chymotrypsin; N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (Suc-AAPF-pNA), which is cleaved under alkaline condition resulting in the production of p-nitroaniline, a yellow substance that can be quantified by measuring its absorbance at 410 nm (Figure 4.48) [130].

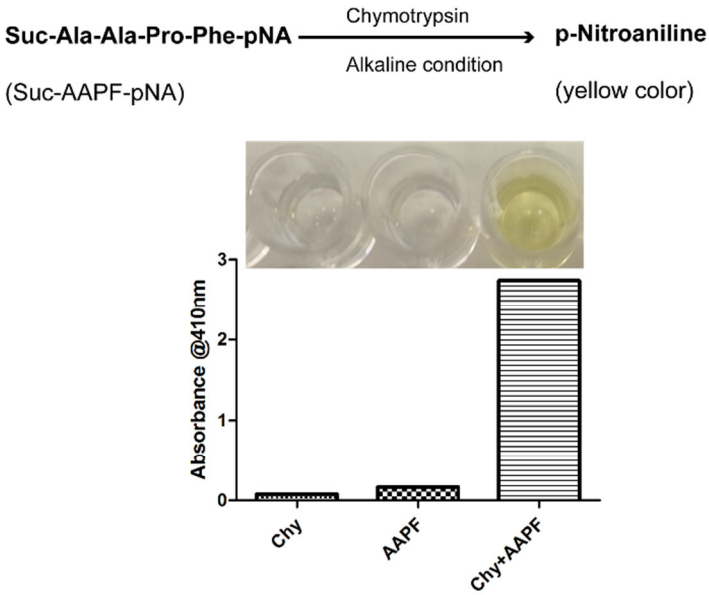


Figure 4.48: α-chymotrypsin *in vitro* assay. The peptidic substrate Suc-AAPF-pNA is cleaved by α-chymotrypsin under alkaline conditions resulting in the yellow colored p-nitroaniline. No yellow color is observed in the control reactions containing only α-chymotrypsin (Chy) or the substrate (AAPF).

4.8.1 Validation of the α-chymotrypsin assay with the known inhibitor DCI.

To validate the α-chymotrypsin assay, the assay was performed with a known chymotrypsin inhibitor, the isocoumarin DCI (Figure 4.49). 50 µl reaction mixtures containing 4 µM of chymotrypsin were pre-incubated with defined concentrations of DCI (0 µM - 250 µM) at 25°C for 30 min. Each sample was then added to 100 µl of the substrate solution (0.12 mM Suc-AAPF-pNA), and the absorbance at 410 nm was recorded within 5 min in order to quantify the concentration of hydrolysis product. The resultant bar graph (Figure 4.49A) is the average of three replicates and the IC50 value was then calculated by fitting a nonlinear regression curve to a plot of log (DCI concentration) versus the % chymotrypsin activity (Figure 4.49B). The IC50 value for DCI was 3.5 µM, which was comparable to a previously reported value in the literature (2.8 µM) [103].

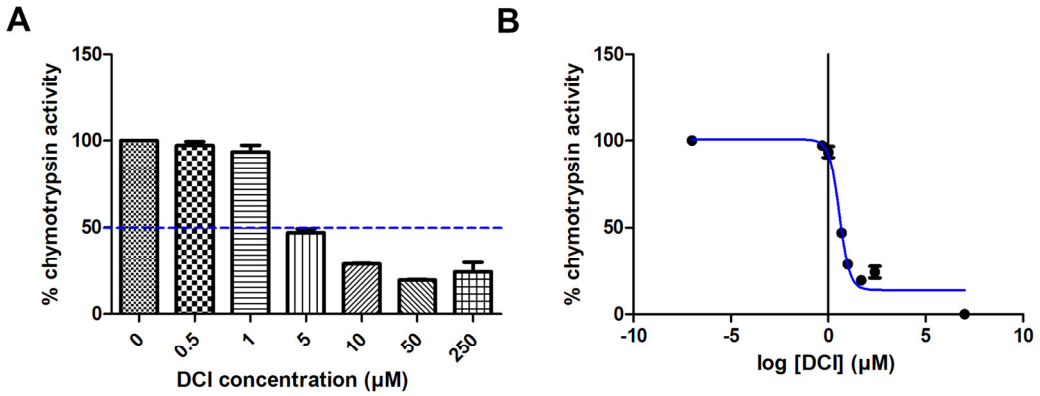


Figure 4.49: Validation of the α-chymotrypsin assay with the isocoumarin inhibitor DCI (N=2). A) Plot of the DCI concentration versus percentage of α-chymotrypsin activity. The dashed blue line indicates 50% enzyme inhibition. B) Nonlinear regression curve of the log (DCI concentration) versus % chymotrypsin activity. The IC50 value for DCI was calculated equal to 3.5 µM. N= number of experiments.

4.8.2 Activity of novel rhomboid inhibitors (saccharins, benzoxazinones, VS5) against α-chymotrypsin.

We observed that some N-methylated saccharin based acid derivatives were able to inhibit the rhomboid protease GlpG in the submicromolar range (section 4.6.2.3). To gain some insight into their selectivity 18 saccharin derivatives were tested for their activity in the α-chymotrypsin assay, and IC50 values were determined as described above. As an example, Figure 4.50 displays the assay results for the saccharin derivative BSc5195. The IC50 for α-chymotrypsin inhibition was calculated to be 10 µM, the values for the other derivatives are listed in Table 4.16. To ensure the IC50 value, the assay was performed one more time for BSc5188 and BSc5195 but for all other saccharin derivatives the tested IC50 value is based on one experiment. Overall, the results showed that the saccharin derivatives have at least a 20-fold window of selectivity for inhibition of the rhomboid GlpG over the soluble serine protease α-chymotrypsin.

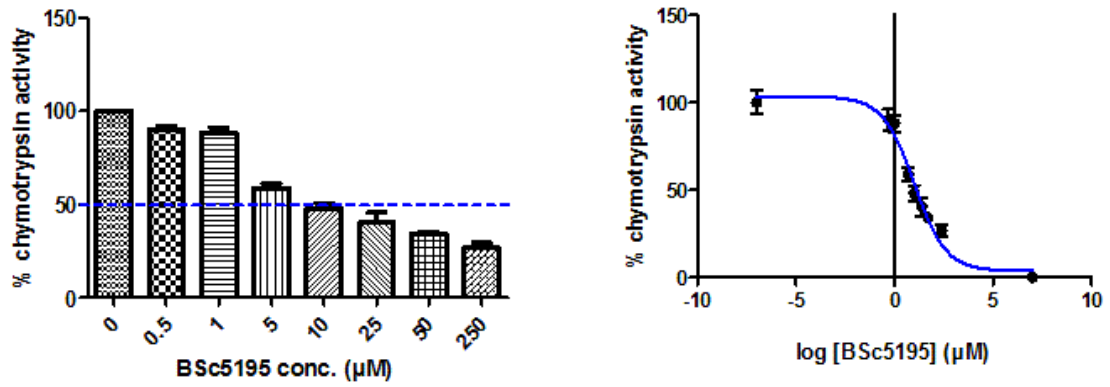


Figure 4.50: α-Chymotrypsin assay with the N-methylated saccharin inhibitor BSc5195 (N=2). A) Plot of BSc5195 concentration versus percentage of chymotrypsin activity. The dashed blue line indicates 50% enzyme inhibition. B) Nonlinear regression curve of the log (BSc5195 concentration) versus % chymotrypsin activity. The IC50 value for BSc5195 was 10 µM. N= number of experiments.

Table 4.16: IC50 values for inhibition of α-chymotrypsin by novel saccharin-based acid derivatives.

Entry No.	BSc No.	IC50 (µM)
41	5156	33
47	5187	50
48	5188	9
49	5189	29
50	5190	27
51	5191	14
52	5192	22
53	5193	54
54	5194	19
55	5195	10
56	5196	208
57	5197	68
58	5198	56
59	5199	39
60	5200	120
63	5203	40
64	5204	12
65	5205	18

Similarly, the active benzoxazinone based inhibitors and the virtual screening derived compound VS5 were tested for activity in the α-chymotrypsin assay. The 2-styryl benzoxazinone based inhibitor BSc5174 did not display any inhibitory activity at the highest concentration of 250 µM (Figure 4.51), a concentration 5-fold above its IC50 value in the GlpG *in vitro* activity assay (Table 4.13). A similar negative result was obtained for the compound BSc5172 (data not shown). Likewise, the virtual screening derived compound VS5 was inactive in the α-chymotrypsin assay up to a concentration of 1 mM (Figure 4.51), which was approximately 60-fold above its IC50 value for GlpG inhibition (Figure 4.46).

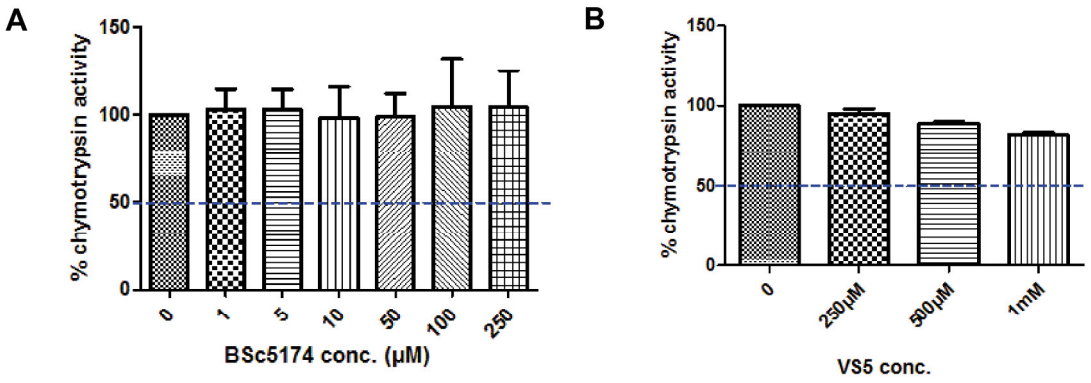


Figure 4.51: α-Chymotrypsin assay with the 2-styryl benzoxazino-4-one inhibitor BSc5174 and the virtual screening inhibitor VS5 (N=1). A) No chymotrypsin inhibition was observed in the concentration range of 1 µM - 250 µM BSc5174. B) No chymotrypsin inhibition was observed in the concentration range of 250 µM – 1 mM VS5. N= number of experiments.

4.9 Activity of saccharin based inhibitors with different rhomboid proteases and substrates

4.9.1 Inhibitory activity of saccharin based inhibitors in a modified GlpG *in vitro* activity assay using the LacY substrate.

To confirm the inhibitory potency of the saccharin inhibitors for *E. coli* GlpG with a different substrate, the GlpG *in vitro* activity assay was performed with a chimeric substrate incorporating the TMD sequence of the *E. coli* lactose permease (LacY) instead of the Gurken TMD [103]. The assay with LacY was performed under identical conditions as the GlpG *in vitro* activity assay with the Gurken chimeric substrate (section 3.2.12). 0.35 µM GlpG were pre-incubated at 37˚C with the saccharin inhibitors. Subsequently, 1.8 µM of LacY substrate was added and the reaction was continued for 90 min at 37˚C. The reaction samples were analysed by SDS-PAGE and the band intensity of the N-terminal LacY cleavage product (NTF) was quantified with ImageJ. Figure 4.52 shows the assay results for BSc5204. The calculated IC50 value was 0.45 µM. This value was very similar to the IC50 of BSc5204 in the GlpG *in vitro* activity assay with the Gurken chimeric substrate (0.41 µM, Table 4.10). The IC50 values for three other saccharin derivatives in the LacY assay are listed in Table 4.17, and were also similar to the values of the same compounds tested before in combination with the Gurken substrate (Table 4.10).

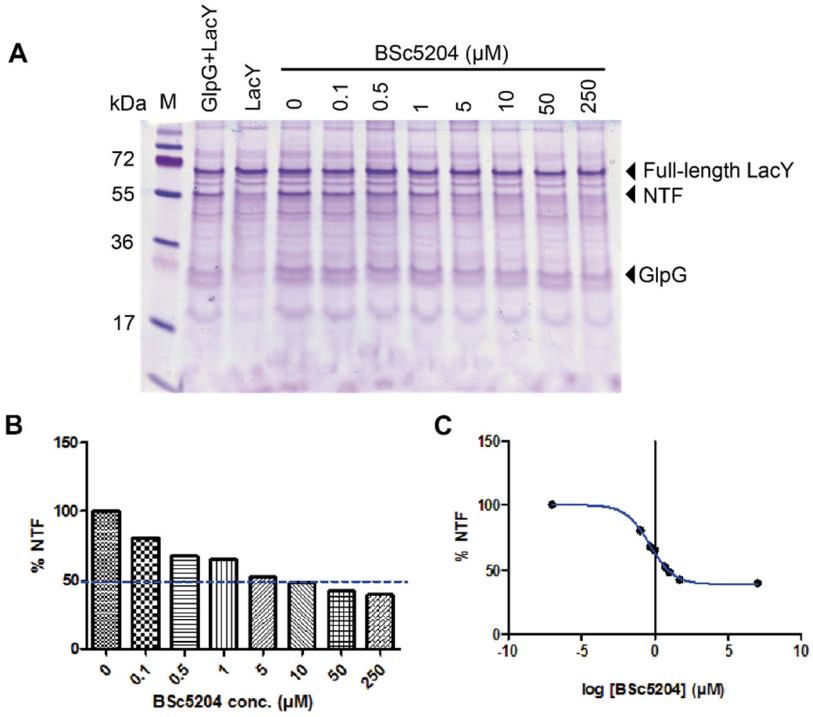


Figure 4.52: GlpG *in vitro* activity assay using the LacY substrate with the saccharin-based inhibitor BSc5204 (N=1). A) SDS-PAGE of the reaction products after incubation of GlpG and LacY substrate with BSc5204 in defined concentrations (0.1 µM - 250 µM). The arrows indicate the full-length LacY substrate (~ 66 kDa), the N-terminal LacY fragment (NTF, ~ 55 kDa), and GlpG (~ 33 kDa). M: Molecular weight marker. B) Plot with BSc5204 concentration versus % NTF band intensity. The dashed blue line indicates 50% enzyme inhibition. C) Nonlinear regression curve of the log (BSc5204 concentration) versus % NTF band intensity. The gel was stained with coomassie brilliant blue, quantified with ImageJ, and the values were plotted in GraphPad Prism. N= number of experiments.

Table 4.17: IC50 values of N-methylated saccharin inhibitors in the GlpG *in vitro* activity assay with the LacY substrate.

Entry No.	BSc No.	IC50 (µM)
48	5188	0.23
55	5195	0.53
63	5203	0.33
64	5204	0.45

4.9.2 Inhibitory activity of saccharin-based inhibitors against the AaROM rhomboid from *Aquifex aeolicus*.

To gain some insight into the selective inhibition of N-methylated saccharins for different rhomboids, the three most potent derivatives (BSc5188, BSc5195 and BSc5204) were first tested with another prokaryotic rhomboid from the thermophilic bacteria *Aquifex aeolicus* (AaROM). As explained in section 3.2.13, the *in vitro* activity assay for AaROM was performed in a similar manner as the GlpG activity assay. Reaction mixtures containing 0.094 µM AaROM were incubated with the inhibitors at a single concentration of 250 µM concentration. Then Gurken substrate (0.4 µM) was added, the incubation was continued for another 90 min, and the reaction products were analysed by SDS-PAGE.

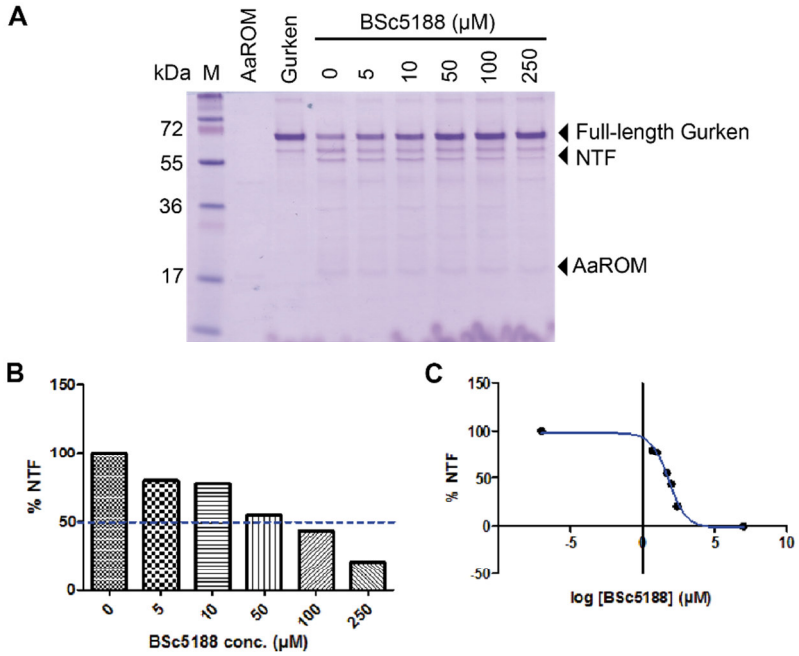


Figure 4.53: *In vitro* activity assay for the *Aquifex aeolicus* rhomboid (AaROM) with the saccharin-based inhibitor BSc5188 (N=1). A) SDS-PAGE of the reaction products after incubation of AaROM and Gurken substrate with BSc5188 in defined concentrations (5 µM - 250 µM). The arrows indicate the full length Gurken substrate (~ 66 kDa), the N-terminal Gurken fragment (NTF, ~ 55 kDa), and the AaROM enzyme (~ 17 kDa). M: Molecular weight marker. B) Plot with BSc5188 concentration versus % NTF band intensity indicating percentage of inhibition. The dashed blue line shows 50% enzyme inhibition. C) Nonlinear regression curve of the log (BSc5188 concentration) versus % NTF band intensity. The IC50 value was calculated to be approximately 62 µM. The gel was stained with coomassie brilliant blue, quantified with ImageJ, and values were plotted in GraphPad Prism. N= number of experiments.

This showed that all three inhibitors were able to partially inhibit AaROM activity at 250 µM (data not shown). Subsequently, IC50 values of the three derivatives for inhibition of AaROM were determined from a single experiment. Figure 4.53 shows the results for the compound BSc5188. The IC50 values were 62 µM for BSc5188, 89 µM for BSc5195, and 61 µM for BSc5204 (Table 4.18). This indicated that the inhibitory potency of the saccharin-based inhibitors was at least a 100-fold lower for the *Aquifex* AaROM compared to the *E. coli* GlpG.

Table 4.18: IC50 values for inhibition of *Aquifex aeolicus* rhomboid (AaROM) by N-methylated saccharin inhibitors (N=1).

Entry No.	BSc No.	IC50 (µM)
48	5188	62
55	5195	89
64	5204	61

4.9.3 Inhibitory activity of saccharin-based inhibitors in a cell-based assay for the mammalian rhomboid protease RHBDL2.

In order to investigate whether the saccharin-based compounds might also inhibit the enzymatic activity of a more distant member of the rhomboid protease family, a previously established cell-based assay for the murine rhomboid RHBDL2 was employed [103]. HEK 293T cells were grown to 50-

80% confluency in 6-well plates and then transiently transfected with 0.5 µg of an expression plasmid encoding HA-tagged RHBDL2 (pcDNA3-RHBDL2) and 1 µg of an expression plasmid encoding FLAG-tagged Gurken substrate (pcDNA3-Gurken) (section 3.3.4). The saccharin inhibitors (dissolved in DMSO) were added 2h post transfection to a final concentration of 50, 100 or 250 µM (section 3.3.8). After further incubation for 48 h, the cells were harvested and lysed (sections 3.3.5 and 3.3.6). The protein concentrations of the extracts were determined (section 3.3.7) and equal amounts of protein were separated by SDS-PAGE followed by Western blotting and immunodetection with anti-HA and anti-FLAG antibodies. Figure 4.54 shows the results for the compound BSc5156. Upon co-transfection of RHBDL2 and the Gurken substrate, an N-terminal Gurken cleavage product (NTF) appeared at around ~ 55 kDa (Figure 4.54, lanes 3-6), which was not detectable in cells only transfected with the full-length Gurken substrate (lane 2). The rhomboid RHBDL2 was detected at ~ 40 kDa (lane 1). The band intensity of the Gurken NTF was not reduced after treatment with increasing concentrations of BSc5156 (lanes 4-6) as compared to the DMSO vehicle control (lane 3), indicating that BSc5156 did not inhibit the activity of the murine rhomboid RHBDL2. Similar negative results were obtained with the compounds BSc5205 and BSc5192 (data not shown).

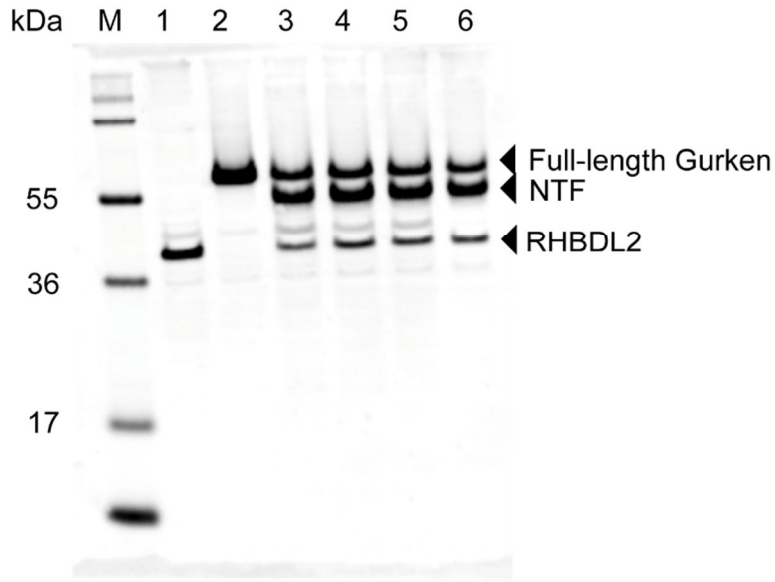


Figure 4.54: Inhibitory activity of the saccharin-based inhibitor BSc5156 in a cell-based assay for the murine rhomboid RHBDL2 (N=3). HEK293T cells were transiently co-transfected with expression plasmids encoding HA-tagged RHBDL2 and FLAG-tagged Gurken substrate. Subsequently, cells were treated for 48 h with increasing concentrations of BSc5156 or DMSO vehicle. Expression of RHBDL2 and Gurken was analysed by Western blotting and immunodetection with a mixture of anti-HA and anti-FLAG antibodies. Arrows indicate the full-length Gurken substrate (~ 60 kDa), the N-terminal Gurken cleavage fragment (NTF,~ 55 kDa), and RHBDL2 (~ 40 kDa). Lane-1: cells transfected with murine RHBDL2, Lane-2: cells transfected with full-length Gurken substrate, Lane-3: cells co-transfected with RHBDL2 and full-length Gurken and treated with DMSO vehicle. An additional N-terminal Gurken cleavage fragment (NTF) was detected. Lane 4-6: cells co-transfected with RHBDL2 and full-length Gurken and treated with 50, 100 or 250 µM BSc5156. The intensity of the NTF band was unchanged compared to the DMSO control (lane 3) indicating that BSc5156 was not able to inhibit the activity of RHBDL2. N= number of experiments.

4.10 Preliminary studies on the mechanism of rhomboid inhibition by saccharin-based inhibitors.

According to the proposed mechanism of inhibition of soluble serine proteases by N-methylated substituted saccharins (Figure 4.29) [110], the reaction with the nucleophilic serine of the rhomboid active site and electronic rearrangements should lead to the release of the leaving group (LG). In the case of the saccharin-based acid derivatives, the corresponding acid would be released after the enzymatic reaction (Figure 4.55). Compound BSc5193 is an N-methylated saccharin derivative with *o*-amino benzoic acid (anthranilic acid) as leaving group (Figure 4.55). This compound did inhibit GlpG in the *in vitro* activity assay with an IC50 of approximately 10 µM (appendix 7.2). Anthranilic acid is highly fluorescent with excitation and emission wavelengths of 310 nm and 400 nm [185]. Hence, we investigated whether the reaction of BSc5193 with GlpG leads to an increase in fluorescence signal indicating the enzymatic release of the anthranilic acid LG (Figure 4.55). GlpG (0.35 µM) and BSc5193 (5 µM) were incubated for 30 min and the fluorescence signal was measured with a multiplate reader. Compound BSc5195 (IC50 = 0.2 µM) bearing a non- fluorescent leaving group, *o*-chloro benzoic acid was used as a control. Three samples containing only GlpG, only BSc5193 and only BSc5195 were measured to control for background fluorescence. Reaction sample containing GlpG (0.35 µM) incubated with BSc5195 (5 µM) for 30 min showed no increase in fluorescent signal compared to control samples. Whereas, approximately 4-fold increase in the fluorescent signal was observed after the incubation of GlpG with BSc5193 as compared to the control samples (Figure 4.55). These results support the idea that the reaction mechanism of N-methylated saccharin-based acid derivatives with the rhomboid active site involves the release of the acidic LG.

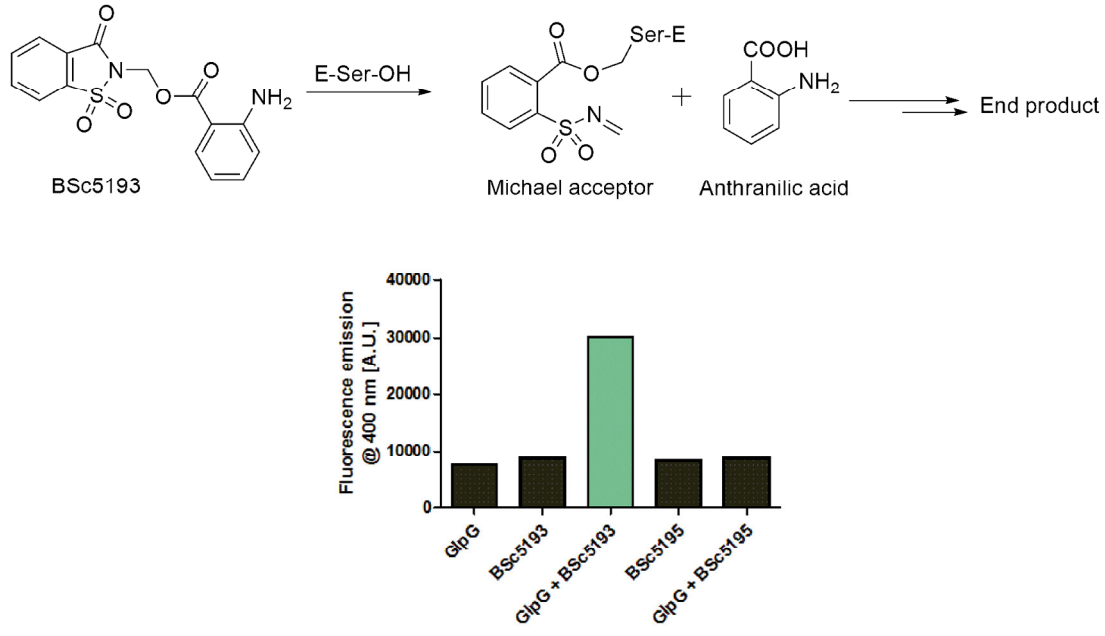


Figure 4.55: A fluorescence-based assay to study the reaction mechanism of GlpG with the saccharin-based inhibitor BSc5193 (N=2). Upper panel: The putative reaction mechanism of BSc5193 with the rhomboid active site involves the formation of a Michael acceptor and the release of the acidic leaving group. Lower panel: GlpG (0.35 µM) and BSc5193 (5 µM) were incubated for 30 min and the fluorescent signal was measured by excitation/emission at 310/400 nm.

Approximately 4-fold increase in the fluorescent signal was observed as compared to the control samples containing only either enzyme or inhibitor, indicating the enzymatic release of the fluorescent anthranilic acid as leaving group. Whereas no increase in the fluorescent signal was observed in case of GlpG incubated with BSc5195.

According to the proposed mechanism, inhibition of soluble serine proteases by N-methylated substituted saccharins (Figure 4.29) [110], might result in a doubly-bonded enzyme-inhibitor end product. Earlier, Vinothkumar *et al*. used crystallization to describe that the rhomboid inhibitor 7-amino-4-chloro-3-methoxy-isocoumarin (JLK-6) forms a double-bonded enzyme-inhibitor complex [87]. Later, Vosyka *et al*. showed that this doubly-bonded end product could result in intramolecular cross-linking of the enzyme which would eventually alter the mobility of GlpG during SDS-gel electrophoresis [88]. The cross-linked GlpG runs at a lower apparent molecular weight than the singly-bonded GlpG. In the case of the saccharin-based acid derivatives, the corresponding doubly-bonded end product should also result in a cross-linked GlpG. Hence, JLK-6 (250 µM) and saccharin inhibitor BSc5188 (0.1-250 µM) were analysed in the GlpG *in vitro* activity assay. On the SDS-PAGE, JLK-6 reaction resulted in 100% inhibition of GlpG and a double band for GlpG was observed where the lower band indicated the cross-linked GlpG with JLK-6 inhibitor (Figure 4.56). The saccharin inhibitor (BSc5188) also resulted in 100% inhibition of GlpG at 250 µM but there was no double band for GlpG visible (Figure 4.56). Therefore, in contrast to the isocoumarin inhibitor (JLK-6), N-methylated substituted saccharins do not facilitate intramolecular cross-linking within the enzyme and might inhibit GlpG by modifying only one of the catalytically active amino acids within the active side cleft. In conclusion, the underlying mechanism of rhomboid protease inhibition by N-methylated saccharin-based acid derivatives involves the release of the acidic LG but not the formation of a doubly-bonded enzyme-inhibitor end product.

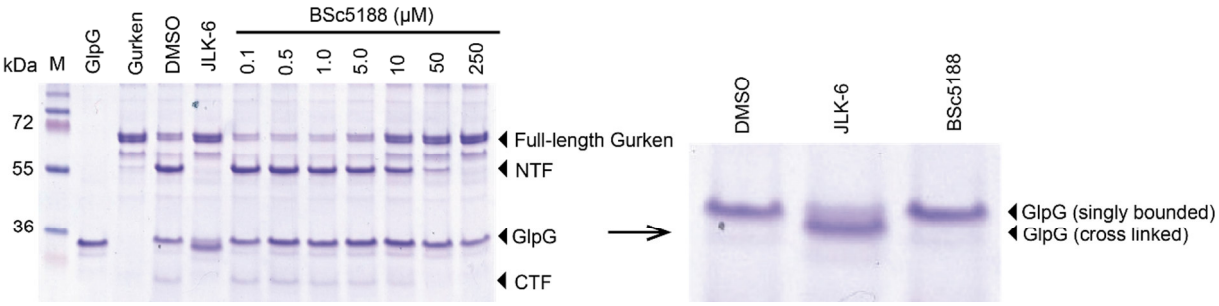


Figure 4.56: *In vitro* GlpG activity assay with rhomboid isocoumarin inhibitor (JLK-6) and aryl carboxylate saccharin inhibitor BSc5188. SDS-PAGE of the reaction products after incubation of GlpG and Gurken substrate with JLK-6 (250 µM) and BSc5188 in defined concentrations (0.1 µM - 250 µM). The arrows indicate the full length Gurken substrate (~ 66 kDa), the N-terminal Gurken fragment (NTF, ~ 55 kDa), and GlpG (~ 33 kDa). A double band of GlpG upon JLK-6 treatment is visible, the lower band corresponds to the intramolecular cross-linked GlpG. M: Molecular weight marker. The gel was stained with coomassie brilliant blue.

5 Discussion

The rhomboid proteases belong to a family of ubiquitously distributed intramembrane proteases and perform a variety of biological functions. They exist in both prokaryotes and eukaryotes [36, 42]. Rhomboids cleave their substrates within the transmembrane domain in a process called regulated intramembrane proteolysis (RIP), and the substrates of the rhomboids are type I or type III membrane proteins [82]. Rhomboid proteases function in the activation of the EGF signaling pathway in *Drosophila*, in quorum sensing in *Providencia stuartii*, during host cell invasion by *Plasmodium* and *Toxoplasma*, in the regulation of mitochondrial morphology and dynamics in yeast, and in the process of apoptosis in mammalian cells [36]. Because of their diverse biological functions, rhomboids might also have great potential as drug targets. For example, the mitochondrial rhomboid PARL is involved in the proteolytic degradation of the kinase PINK1, and mutations in the human PINK1 gene result in a recessively inherited form of Parkinson's disease [186]. Consequently, one focus of activity in the rhomboid research field in recent years has been the search for potent and selective inhibitors [36]. A few small molecules such as isocoumarins and β-lactams have been previously reported to inhibit rhomboid proteases. However, while β-lactams were found to possess good potency in the low micromolar range, they displayed insufficient selectivity over soluble serine proteases such as chymotrypsin [103]. Hence, the present work was focused on the design, the synthesis and the biological evaluation of new and improved rhomboid inhibitors. Taking advantage of structural information about rhomboid proteases available in the literature, we utilized a computer-aided approach that involved the virtual screening of drug-like candidates against the rhomboid protease. All molecular modeling experiments were performed using a co-crystal structure of the *E. coli* rhomboid GlpG and the phosphate inhibitor Cbz-AlaP(OiPr)F or CAPF (PDB-ID: 3UBB) [89] and the Molecular Operating Environment (MOE) software package. Two different approaches were explored: candidate-based screening and large molecular library virtual screening. In the candidate-based approach, we chose peptidic and non-peptidic inhibitors, which were docked into the GlpG pocket. Subsequently, the best candidates were selected for synthesis and biological analysis. In contrast, in the large molecular library screening, a vast compound library of commercially available small drug like compounds was screened against the rhomboid GlpG. Similar to the candidate-based approach, the best candidate compounds were selected for purchase followed by activity analysis in a cell-free rhomboid protease assay.

3.7 Rhomboid protease activity assays

In order to quantify the inhibitory potency of new inhibitors, both cell-free and cell-based activity assays were established for pro- and eukaryotic rhomboids based on published protocols [103]. As a first step, the GlpG rhomboid and the rhomboid substrate Gurken were expressed in *E. coli* and purified through a His-tag by IMAC affinity chromatography. In addition, a second rhomboid,

AaROM from *Aquifex aeolicus*, was expressed as a GST fusion protein in *E. coli* and purified by affinity chromatography with glutathione-sepharose resin [102]. Both rhomboids and the Gurken substrate were successfully purified in sufficient amounts. Purified and concentrated preparations of enzymes and substrate were subsequently used for the cell-free activity assays. In this assay, the Gurken substrate is cleaved by the rhomboid, either GlpG or AaROM, and the resulting N- and C-terminal proteolytic Gurken fragments (NTF and CTF) were analysed by SDS-PAGE gel or Western blotting [87]. Substrate concentration and reaction time were optimized to ensure that the substrate was present in excess amounts and that the enzymatic reaction was in the linear phase without saturation. The obtained values (substrate concentration 1.8 µM for ≈ 0.35 µM GlpG, and reaction time 1.5h) were comparable to previously reported results[103]. Finally, the optimized GlpG assay was successfully validated with three previously known rhomboid inhibitors, the beta-lactam inhibitor (L16) and two isocoumarins (DCI and JLK-6) [103]. The obtained IC50 values for these known inhibitors were also comparable to values previously reported in the literature [87, 103].

Table 5.1: IC50 values of known rhomboid protease inhibitors in a GlpG activity assay.

Inhibitor	Evaluated IC50 (µM)	IC50 (µM) from literature
L16	0.9	0.4
DCI	19	11
JLK-6	6	6

One bottleneck in the development of pharmacological inhibitors is the identification of substances that can pass through the cell membrane and are effective in intact cells. Cell-free activity assays are widely used to optimize the potency and selectivity of inhibitors, but to select candidate molecules for further development it is important to test compounds in cell-based assays. Therefore, a cell-based assay for the murine rhomboid RHBDL2 was also established based on a previous study [103]. In this assay, HEK293T cells were transiently co-transfected with expression vectors encoding the mammalian rhomboid RHBDL2 and the Gurken substrate. Co-expression of RHBDL2 and the Gurken substrate led to the formation of an N-terminal proteolytic cleavage product as confirmed by Western blotting (Figure 4.54). In summary, cell-free assays for the bacterial rhomboids GlpG and AaROM were successfully established and benchmarked with known rhomboid inhibitors, which was a prerequisite to identify and characterize new rhomboid inhibitors. In addition, a cell-based assay for the murine rhomboid RHBDL2 was established to screen new compounds for activity against a mammalian rhomboid and to identify potential cell-permeable rhomboid inhibitors.

3.8 Design, synthesis and biological evaluation of peptidic inhibitors

The mechanism and the specificity of substrate binding and cleavage by rhomboid proteases are not fully understood due to the lack of structural data for substrate-enzyme complexes. Several co-crystal structures of rhomboids with small molecule mechanism-based inhibitors have been previously reported [123]. However, the small size and the chemical variability of the reported inhibitors in comparison to a polypeptide restrict their use as models for substrate binding. Hence, one of the reasons for us to develop peptidic inhibitors was that the design of peptidomimetic inhibitors might help to gain insight into the substrate binding mechanism. Moreover, numerous studies have shown that peptides are a rational starting point to design enzyme inhibitors as peptide sequences are able to mimic the natural substrate sequence [109, 113] and can provide high specificity and affinity. Usually, covalent peptidomimetic inhibitors for serine proteases consist of two moieties, a serine trap and a recognition motif [110]. The serine trap is a reactive motif, which reacts with the active site serine residue of the enzyme resulting in covalent bond formation, and the amino acid sequence attached to the serine trap acts as a recognition motif. Two covalent peptidomimetic templates, ketobenzothiazole-based [187] and oxathiazol-2-one [148] based inhibitors, were selected and optimized through a computer-aided approach. A ketobenzothiazole-containing peptidomimetic compound was shown to be a very potent and selective inhibitor of the type II membrane-bound serine protease matriptase [111], and these inhibitors are usually reversible, covalent binders of the enzyme active site. The design process included the search for the minimal length of the peptidic component using an optimized covalent docking module 'ConflexDock' in MOE. Based on the docking results, the optimal length of a peptidic inhibitor for rhomboids was determined to be a dipeptide, and the best suitable residues were Ala and Met for the P1-P2 positions. Extending the peptidic component beyond a dipeptide did not look promising. However, additional docking experiments that we performed for the P3 and P4 positions might later be useful to improve the selectivity of peptidomimetic inhibitors. Peptidic oxathiazol-2-one based inhibitors were initially discovered to inhibit the proteasome [148]. These are also covalent binders and the reaction of the electrophilic ketone center in the reactive motif with the threonine nucleophilic active site residue results in inhibition of the enzyme. Because serine proteases have a similar inhibition mechanism, we selected oxathiazol-2-one as a second peptomimetic template and, considering the docking results, we kept the dipeptide as the target length for the peptidic component. Overall, three derivatives each were successfully synthesized for the ketobenzothiazole (BSc5123, BSc5124 and BSc5125) and the oxathiazol-2-one (BSc5138, BSc5139 and BSc5140) based templates. All compounds were tested at high concentrations (250 µM – 1mM) in the GlpG *in vitro* activity assay. Unfortunately, none of the peptidomimetic compounds showed any activity in this concentration range. As a result, peptidic inhibitors were not further explored in this thesis. We would consider three possible reasons for the failure of these peptidic compounds to inhibit the rhomboid protease. The first explanation is a non-suitable serine trap, meaning that the ketobenzothiazole and the oxathiazol-2-one functional groups were not able to react with the active

site serine of the rhomboid. Alternatively, it has been proposed by Dickey *et al*. that intramembrane proteolysis by rhomboids is an unusually slow process, which is kinetically controlled rather than by the protein-protein affinity of enzyme and substrate in the membrane [96]. Hence, a proteolytic event within the membrane might take minutes to complete, which could provide an explanation for the poor affinity of our synthesized peptidic inhibitors for the rhomboid. The third explanation is non-suitable recognition motif, in which either the length of the amino acid sequence or the amino acid residues selected based on the docking results were inappropriate to inhibit the rhomboid. Two of these explanations are partially supported by a recent publication from Zoll *et al*., in which peptidyl-chloromethylketones were described as the first peptidic inhibitors against rhomboids. The tetrapeptide recognition sequence in these inhibitors (Ac-IATA-cmk) was derived from a natural bacterial rhomboid substrate, *P.stuartii* TatA [91]. Based on co-crystal X-ray structures they demonstrated that subsites for the P1–P4 residues are vital for substrate recognition and efficient catalysis. They also reported that the substrate residues P1 and P4 were most crucial for the binding affinity and that only residues beyond P4 were completely solvent-exposed. Hence, unlike suggested by our docking studies, a dipeptidyl chain was not sufficient for recognition by the rhomboid. They also found that the P1 and P4 positions were the most restrictive compared to the P2, P3 and P5 positions. Similar to our docking results they considered Ala to be the best residue for the P1 position, and the P2 position was able to accept almost any amino acid. Overall, it appears likely that the failure of ketobenzothiazole and oxathiazol-2-one based peptidic compounds to inhibit the rhomboid protease was either due to an insufficient length of the peptidyl chain or due to an inappropriate reactive motif. If the latter were true, extending the peptidyl length up to a tetrapeptide might lead to active inhibitors, a possibility that could be explored in future studies.

3.9 Design, synthesis and biological evaluation of non-peptidic inhibitors

A few broad-spectrum mechanism-based inhibitors for rhomboids including isocoumarins, β-lactams, and β-lactones have been previously reported [123]. However, except β-lactam derivatives, none of these displayed high potency for rhomboids, and β-lactams showed only modest selectivity for rhomboids over the soluble serine protease chymotrypsin [103]. Hence, our objective was to find new classes of non-peptidic small molecule inhibitors for rhomboids with better potency and specificity. Three different small molecule templates were selected and refined through docking studies rhodanines, saccharins and benzoxazinones.

3.9.1 Rhodanine-based inhibitors

Rhodanines are heterocyclic ring inhibitors, which are known to bind numerous distinct targets such as HCV serine proteinase and cathepsin D [153, 154]. These compounds have a high potential to form intermolecular interactions that are not due to aggregation or other nonspecific inhibition mechanisms.

Recently, Anumala *et al*. synthesized a series of rhodanine-3-acetic acid (RAA) fluorescent probes to visualize neurofibrillary tangles in Alzheimer's disease brains [165]. These RAA probes were available in the BSc archive of the Schmidt lab at the TU Darmstadt. Therefore, rhodanine derivatives were selected as the first non-peptidic template, and the potential interaction of these probes with the rhomboid protease pocket was investigated using the DOCK script in MOE. The phosphate inhibitor CAPF was employed as a control molecule for comparative studies. The docked probes were scored and ranked based on their calculated ligand efficiency, scoring function and ligand interactions. The docking indicated that, similar to CAPF, the RAA probes were nicely positioned in the rhomboid pocket and were not exposed to the external environment. This suggested that the RAA probes were likely to inhibit rhomboid proteases. However, in the GlpG *in vitro* activity assay, none of the RAA probes had any inhibitory activity against the rhomboid protease at a concentration of 250 µM. Consequently, rhodanines were not further investigated.

3.9.2 N-methylated saccharin based inhibitors

N-methylated saccharins were selected as the second non-peptidic template to develop novel rhomboid protease inhibitors. N-methylated saccharins, in which a leaving group (LG) is attached to the nitrogen atom of the saccharin ring, were introduced by Groutas *et al*. as mechanism-based inhibitors for serine proteases such as human leukocyte elastase (HLE) [169]. It was reported that the mechanism of inhibition by N-methylated saccharins involves two steps. The first step is the formation of an acyl enzyme through attack of the active site serine. In a second step, further electronic rearrangement leads to the formation of another electrophilic site, which could be attacked by the histidine active site residue. This might facilitate the formation of a double-bonded enzyme-inhibitor complex [110, 188]. In addition, N-methylated saccharins possess substantial potential for further derivatization, either through alteration of the LG or through substitution of the phenyl ring of the saccharin core structure [113]. Such modifications were successfully used as key structural elements in orally bioavailable inhibitors of HLE [178] and in intermediates for the production of anti-inflammatory drugs (oxicam) [175]. Since the derivatization of the N-functional group was easier compared to the heterocyclic ring of the saccharin core, we intended to identify the most favourable LGs for the rhomboid protease. Hence, a list of known saccharin based serine protease inhibitors with suitable LGs were docked into the rhomboid pocket using the DOCK module in MOE. The docking was based on the initial interactions of the compounds with the rhomboid cavity rather than on the end product of the reaction because, according to the putative mechanism of inhibition, the covalently bonded end product would be identical in each case after the release of the LG (Figure 4.29 and Figure 4.30). The docking studies revealed that aryl carboxylate LGs were the most favourable of all docked LGs. However, for comparative studies, a preliminary batch of 10 N-methylated saccharin derivatives with variable LGs (Table 4.8) was selected and successfully synthesized. The biological evaluation of this preliminary batch in the GlpG *in vitro* activity assay showed that saccharins with aryl carboxylate LGs were able

to inhibit the rhomboid protease at a concentration of 10 µM while saccharins with other LGs were inactive at 250 µM. Importantly, these results were fully consistent with our docking findings and indicated that aryl carboxylate LGs might act as a recognition motif for the rhomboid protease. Consequently, N-methylated saccharin with an aryl carboxylic acid LG was selected as a lead molecule. Next, docking experiments were performed with a molecular database of 103 saccharin derivatives to investigate the best possible extensions on the aryl ring of the LG and the influence of substitutions on the electron withdrawing properties of the LG. No significantly improved interactions between the aryl ring substituents and their neighbouring residues were found. However, to analyse the influence of the reformed aryl acid with electron donor/acceptor groups on the inhibition potential, N-methylated saccharin acid derivatives with variable extensions (listed in Table 4.9) were synthesized and their biological activity was determined. All saccharin derivatives were able to inhibit GlpG in the *in vitro* activity assay at 250 µM, and compounds with activity at 10 µM were further selected for determination of IC50 values. 12 saccharin derivatives had IC50 values below 1 µM, the IC50 of the best derivative (BSc5195) was 0.2 µM. One interesting question was which effect the electron withdrawing properties of the aryl acid LG would have on the activity of the compounds and inhibition of the rhomboid. Electron withdrawing groups such as halogen substitutions on the aryl ring should result in an increased acylation rate while the addition of electron donating groups such as an amino group should decrease the acylation rate. The activity data showed that aryl acid derivatives substituted with halogens at the ortho position (BSc5188, BSc5195, BSc5197) had IC50 values below 1 µM while the amino substituted inhibitor BSc5193 was unable to completely inhibit GlpG at 10 µM (appendix 7.2). Therefore, electron withdrawing groups at the ortho position of the aryl carboxylate LG appeared to have a positive effect on inhibitor potency. However, with this limited range of derivatization it is not possible to definitively conclude how the electron withdrawing properties of the leaving group affect the inhibitor potency. Nevertheless, as a scaffold, N-methylated saccharins with an aryl carboxylic acid LG are clearly amendable to further derivatization with the goal to improve inhibitor potency. It is possible to improve the electron withdrawing properties of aryl carboxylic acid LG, for example by 2,6-di-substitution with EWG similar to HLE saccharin inhibitors such as WIN 63394 [180] or by ortho substitution with stronger EWG rather than halogen. Besides varying the electron withdrawing properties of the LGs, modifications of the saccharin core structure also has the potential for enhanced enzyme inhibition. Previously, Hlasta *et al*. predicted that substituents at C-4 and C-6 positions influence the enzyme inhibition where incorporation of an iso-propyl group at position 4 and methoxy groups at position 6 resulted in enhanced HLE inhibition (Figure 4.28) [176, 177]. It has been also reported from docking studies of saccharin compounds with HLE that the heterocyclic core ring fits into the S region of HLE where the C-4 substituent would interact with the S1 pocket and C-6 extension resulting into improved interactions with the enzyme pocket [176]. According to our docking results, the saccharin core ring also located to the S region of the rhomboid

pocket (Figure 4.31) and modifications at the saccharin heterocyclic ring similar to HLE inhibitors might influence rhomboid inhibition.

3.9.3 Selectivity of N-methylated saccharin based rhomboid inhibitors

The most potent saccharin inhibitors were further characterized in additional biological assays to investigate their ability to inhibit other rhomboids besides the *E. coli* GlpG and to determine their selectivity over the soluble serine protease α-chymotrypsin. First, the three most potent derivatives were tested for activity against the *Aquifex aeolicus* rhomboid (AaROM) using an *in vitro* activity assay with almost identical conditions compared to the GlpG assay. All three compounds (BSc5188, 5195 and 5204) were able to inhibit the AaROM. However, the IC50 values were between 60-90 µM, indicating 100-fold lower potency towards AaROM as compared to GlpG. In addition, three other potent saccharins (BSc5156, 5192, 5205) with IC50 values below 1 µM in the GlpG assay were tested in the cell based assay for the murine rhomboid RHBDL2. None of the compounds displayed any inhibitory activity up to a concentration of 250 µM. Two potential explanations could account for these negative results. First, the tested saccharin compounds were not able to inhibit the mammalian rhomboid RHBDL2 as opposed to the bacterial GlpG. Second, the compounds were not able to cross the membrane of living cells and were unable to reach the target protein. Although we do not have data to exclude the second explanation this appears rather unlikely given the small size of the saccharin derivatives and moreover, saccharin inhibitors have been known to inhibit elastase *in vivo* [168]. Therefore, to further address this question, we started a collaboration with the laboratory of Dr. Steven Verhelst (University of Leuven, Belgium). Recently, this group had developed a method called *activity-based protein profiling* to screen potential rhomboid inhibitors against a large panel of bacterial, archaeal and mammalian rhomboids [124]. This assay takes advantage of small molecule activity-based probes (ABP) that specifically label the active site of recombinantly expressed and proteolytically active rhomboids. The ABP incorporates a fluorescent group that allows to visualize the labelled enzyme by SDS page and fluorescent imaging.` Potential inhibitors are preincubated with the enzyme prior to labelling with the ABP. Reduced labelling of the enzyme by the ABP indicates an active inhibitor that blocks binding of the ABP to the active site in a competitive fashion [124]. Three potent saccharins (BSc5195, 5196 and 5205) were profiled against 9 different rhomboids. All saccharins were screened at a concentration of 50 µM and the known rhomboid inhibitor DCI was added as a positive control. Figure 5.1 shows the profiling results represented in a heat map. The isocoumarin DCI inhibited 80-100% of the activity of all 9 rhomboids as previously reported [124]. Hence, DCI can be classified as a pan rhomboid inhibitor. Only five out of nine rhomboids were inhibited by all three saccharin inhibitors. These included four bacterial rhomboids (*E. coli* GlpG, *Haemophilus influenzae* HiGlpG, *Vibrio cholerae* VcROM, *Providencia stuartii* PsAarA) and one archaeal rhomboid (*Methanocaldococcus jannaschi* MjROM). The partially or non-inhibited rhomboids included two bacterial rhomboids (*Aquifex aeolicus* AaROM, *Thermotoga maritima*

TmROM) and two eukaryotic rhomboids (*Drosophila melanogaster* DmRho1, *Mus musculus* RHBDL3). Several conclusions can be drawn from these experiments. First, in contrast to DCI, the saccharin inhibitors did not indiscriminately inhibit all rhomboids and they might have a preference for evolutionary older bacterial and archaeal rhomboids. Second, the results confirmed the low inhibitory activity of the saccharin derivative BSc5195 against the *Aquifex aeolicus* AaROM. Third, while the murine rhomboid RHBDL2 was not included in the experiments, all three saccharins did not show any activity against the close murine homolog RHBDL3. This appears to indicate that saccharins are unable to inhibit murine rhomboids, consistent with the lack of activity in the cell-based RHBDL2 assay.

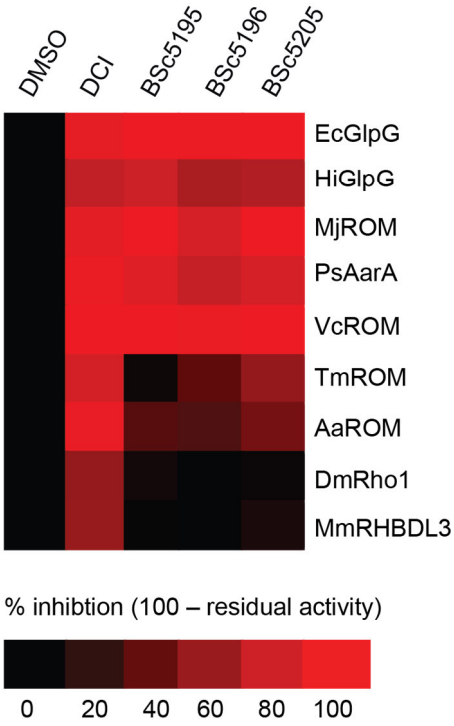


Figure 5.1: Heat map representation of inhibition of 9 rhomboid proteases by saccharin inhibitors. In duplicate, nine rhomboids were incubated with DCI and three saccharin inhibitors (50 µM) or DMSO, and the residual active protease was then labeled by FP-Rh and analyzed by SDS-PAGE. Quantification was performed by gel band densitometry, and the values were converted into a heat map.

Finally, we wanted to examine the selectivity of the novel saccharin-based rhomboid inhibitors over the soluble serine protease α-chymotrypsin. In the colorimetric α-chymotrypsin *in vitro* assay, a peptidic substrate is cleaved by recombinant α-chymotrypsin leading to the production of p-nitroaniline, which can be quantified through its absorbance at 410 nm. The IC50 for the isocoumarin DCI in the α-chymotrypsin assay was 3.5 µM, comparable to a reported value in the literature (2.8 µM) [103] and below the IC50 for DCI in the GlpG *in vitro* activity assay (19 µM). This result was consistent with DCI being a pan serine protease inhibitor. Subsequently, 18 saccharin derivatives were tested in the α-chymotrypsin assay and IC50 values were determined. For each compound the IC50 values were between 20-300 fold higher as compared to the IC50 values in the GlpG *in vitro* activity assay, indicating that the saccharin inhibitors provide at least a 20-fold window of selectivity for

inhibition of the rhomboid GlpG over the soluble serine protease α-chymotrypsin. However, it is important to notice that the IC50 values for α-chymotrypsin and GlpG were determined in entirely different assays and are not directly comparable. In addition, aside from α-chymotrypsin, potential off-targets of the aryl acid substituted N-methylated saccharins would have to be excluded for many other soluble and membrane-anchored serine proteases.

3.9.4 Mechanism of N-methylated saccharin based rhomboid inhibitors

Based on the literature [110, 188], we proposed that aryl acid substituted N-methylated saccharins would irreversibly inhibit rhomboids through the formation of two covalent bonds with the active site serine and histidine residues accompanied by release of the acidic LG. While, at this point we have not formally disproven this reaction mechanism, our current results only partially support it. To investigate whether the LG would be released during the reaction of aryl acid substituted N-methylated saccharins with the rhomboid, we took advantage of a particular saccharin derivative BSc5193, which incorporates ortho-amino benzoic acid (anthranilic acid) as the LG. Anthranilic acid is highly fluorescent and approximately 4-fold increase in fluorescent signal was observed after incubation of GlpG with BSc5193 but not in control reactions containing only GlpG or BSc5193, supporting that the reaction mechanism involves the release of the acidic LG. Concerning the formation of a double-bonded enzyme-inhibitor end product, our findings indicate that this might not be the case and that the saccharin inhibitors simply form a stable acyl-enzyme. Previously, it has been shown by co-crystallization that the rhomboid inhibitor 7-amino-4-chloro-3-methoxy-isocoumarin forms a double-bonded enzyme-inhibitor end product [87], and that this alters the mobility of GlpG during SDS-gel electrophoresis [88]. After incubation of the same compound (JLK-6) with GlpG, we have confirmed the appearance of a double band in SDS-gels, with the lower, faster running protein species likely representing GlpG cross-linked by the amino-methoxy-isocoumarin. In contrast, we never observed double bands after incubation of GlpG with any of the saccharin inhibitors (Figure 4.56). This appears to rule out formation of a double-bonded enzyme-inhibitor end product. In addition, in collaboration with Dr. Kvido Strisovsky's laboratory (Academy of Sciences of the Czech Republic, Prague), we have received preliminary confirmation by mass spectrometry that only the active site serine is covalently modified by the saccharin inhibitors. In these experiments, GlpG was incubated with the saccharin derivatives BSc5195 or 5196 and analyzed by MALDI-TOF mass spectrometry. A mass shift corresponding to N-methylated saccharins after release of the LG was observed after incubation of the compounds with wild type GlpG but not with a S201A GlpG mutant (data not shown). In summary, we successfully identified aryl acid substituted N-methylated saccharins as a new class of mechanism-based inhibitors for rhomboids, with inhibitory potency in the submicromolar range and clear potential for further improvement.

3.9.5 Benzoxazinone based inhibitors

2-substituted benzoxazin-4-ones have been known as effective heterocyclic acylating agents against serine proteases such as HLE [113]. The inhibition mechanism involves the formation of a benozyl acyl enzyme intermediate, which has been confirmed by X-ray crystallography studies [113]. We selected this template because of its large potential for structural variations. Hence, enzyme selectivity can be easily modulated. A set of 2-alkyl or 2-aryl substituted benzoxazinones shown in Table 4.11 was selected as a starting point for docking studies and synthesis. The docking studies revealed that out of all 2-substituted benzoxazinones the derivative BSc5172 fitted best into the rhomboid pocket. Subsequently, all 13 synthesized benzoxazin-4-one derivatives were screened for activity in the GlpG *in vitro* activity assay. Three derivatives BSc5172, 5174 and 5204 displayed inhibitory activity against GlpG with IC50 values of 25 µM, 47 µM and 80 µM. Interestingly, the docking results correlated well with the results from the biological activity assays with BSc5172 being the best compound in both cases. In addition, the 2-styryl benzoxazinone based inhibitors BSc5172 and BSc5174 were tested in the α-chymotrypsin *in vitro* assay and did not show any inhibition at 250 µM, a concentration 10-fold and 5-fold above their IC50 values in the GlpG *in vitro* activity assay. In conclusion, benzooxazinone based inhibitors were found to be active against GlpG but their potency was low compared to the saccharin derivatives described in this thesis. Therefore, the benzoxazinone template was not further pursued in this work. However, the 2-styryl benzoxazinone scaffold might be used to design potent rhomboid inhibitors in the future.

3.10 Virtual screening for novel rhomboid inhibitors

In addition to the candidate approach, we also performed a virtual screen of a large molecular library to discover novel rhomboid inhibitors. We wanted to search covalent binders for rhomboids and for this it was more appropriate to perform covalent docking inside the rhomboid protease pocket. In principle, this could be performed with the CovDock module in MOE but this software is neither time efficient nor user friendly for the screening of large libraries. Recently, Scholz *et al*. developed a DOCKTITE module in MOE, which is a step-by-step and time efficient workflow for covalent docking in MOE [131] and has been used for the virtual screening in this thesis. A large molecular library (size = 2,419,472) of commercially available compounds was selected from the ZINC database and docking was performed with the DOCKTITE workflow. Virtual screening results were ranked based on their LE and, considering their drug-like properties, only five commercially available compounds were shortlisted and purchased (Table 4.15). All five compounds were initially tested at a concentration of 250 µM in the GlpG *in vitro* activity assay. Only one compound (VS5) showed inhibitory activity against GlpG with an approximate IC50 value of 15 µM. VS5 belongs to the phosphate class of inhibitors, which frequently cause cellular toxicity. Therefore, a toxicity assay with human HEK293T cells was performed. No cellular toxicity was observed up to the highest

concentration of 60 µM indicating that VS5 was non-toxic in the concentration range that effectively inhibited the rhomboid GlpG in the *in vitro* activity assay. In addition, VS5 was found inactive in the α-chymotrypsin *in vitro* activity assay up to a concentration of 1 mM, approximately 60-fold above its IC50 value for GlpG inhibition. Previously, two phosphate inhibitors of rhomboid proteases have been described, diisopropyl fluorophosphate (DFP) and the phosphonofluoridate inhibitor CAPF with IC50 values around 50 µM [89]. With VS5 we have discovered a new phosphate inhibitor that is slightly more potent with good selectivity over α-chymotrypsin. This result validates our approach to combine docking studies with the screening of a large molecular library and confirms that this strategy can be successfully employed to discover novel protease inhibitors.

6 References

1. Puente, X.S., et al., *Human and mouse proteases: a comparative genomic approach.* Nat Rev Genet, 2003. **4**(7): p. 544-58.

2. Rawlings, N.D., A.J. Barrett, and A. Bateman, *MEROPS: the peptidase database.* Nucleic Acids Res, 2010. **38**(Database issue): p. D227-33.

3. Vermeulen, K., Z.N. Berneman, and D.R. Van Bockstaele, *Cell cycle and apoptosis.* Cell Prolif, 2003. **36**(3): p. 165-75.

4. Vermeulen, K., D.R. Van Bockstaele, and Z.N. Berneman, *The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer.* Cell Prolif, 2003. **36**(3): p. 131-49.

5. Riewald, M. and W. Ruf, *Science review: role of coagulation protease cascades in sepsis.* Crit Care, 2003. **7**(2): p. 123-9.

6. Patel, T., G.J. Gores, and S.H. Kaufmann, *The role of proteases during apoptosis.* FASEB J, 1996. **10**(5): p. 587-97.

7. Egeblad, M. and Z. Werb, *New functions for the matrix metalloproteinases in cancer progression.* Nat Rev Cancer, 2002. **2**(3): p. 161-74.

8. Krane, S.M., *Elucidation of the potential roles of matrix metalloproteinases in skeletal biology.* Arthritis Res Ther, 2003. **5**(1): p. 2-4.

9. Lichtenthaler, S.F., C. Haass, and H. Steiner, *Regulated intramembrane proteolysis--lessons from amyloid precursor protein processing.* J Neurochem, 2011. **117**(5): p. 779-96.

10. Siefert, S.A. and R. Sarkar, *Matrix metalloproteinases in vascular physiology and disease.* Vascular, 2012. **20**(4): p. 210-6.

11. Luttun, A., et al., *The role of proteinases in angiogenesis, heart development, restenosis, atherosclerosis, myocardial ischemia, and stroke: insights from genetic studies.* Curr Atheroscler Rep, 2000. **2**(5): p. 407-16.

12. Mittl, P.R. and M.G. Grutter, *Opportunities for structure-based design of protease-directed drugs.* Curr Opin Struct Biol, 2006. **16**(6): p. 769-75.

13. Esler, W.P. and M.S. Wolfe, *A portrait of Alzheimer secretases--new features and familiar faces.* Science, 2001. **293**(5534): p. 1449-54.

14. Turk, B., *Targeting proteases: successes, failures and future prospects.* Nat Rev Drug Discov, 2006. **5**(9): p. 785-99.

15. Rawlings, N.D. and A.J. Barrett, *Evolutionary families of peptidases.* Biochem J, 1993. **290 (Pt 1)**: p. 205-18.

16. Erez, E., D. Fass, and E. Bibi, *How intramembrane proteases bury hydrolytic reactions in the membrane.* Nature, 2009. **459**(7245): p. 371-8.

17. Polgar, L., *The mechanism of action of aspartic proteases involves 'push-pull' catalysis.* FEBS Lett, 1987. **219**(1): p. 1-4.

18. Storer, A.C. and R. Menard, *Catalytic mechanism in papain family of cysteine peptidases.* Methods Enzymol, 1994. **244**: p. 486-500.

19. Hedstrom, L., *Serine protease mechanism and specificity.* Chem Rev, 2002. **102**(12): p. 4501-24.

20. Wolfe, M.S., *Intramembrane-cleaving proteases.* J Biol Chem, 2009. **284**(21): p. 13969-73.

21. Rawson, R.B., et al., *Complementation cloning of S2P, a gene encoding a putative metalloprotease required for intramembrane cleavage of SREBPs.* Mol Cell, 1997. **1**(1): p. 47-57.

22. Fluhrer, R. and C. Haass, *Signal peptide peptidases and gamma-secretase: cousins of the same protease family?* Neurodegener Dis, 2007. **4**(2-3): p. 112-6.

23. Alba, B.M., et al., *DegS and YaeL participate sequentially in the cleavage of RseA to activate the sigma(E)-dependent extracytoplasmic stress response.* Genes Dev, 2002. **16**(16): p. 2156-68.

24. Kanehara, K., K. Ito, and Y. Akiyama, *YaeL (EcfE) activates the sigma(E) pathway of stress response through a site-2 cleavage of anti-sigma(E), RseA.* Genes Dev, 2002. **16**(16): p. 2147-55.

25. Kinch, L.N., K. Ginalski, and N.V. Grishin, *Site-2 protease regulated intramembrane proteolysis: sequence homologs suggest an ancient signaling cascade.* Protein Sci, 2006. **15**(1): p. 84-93.

26. De Strooper, B., et al., *Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein.* Nature, 1998. **391**(6665): p. 387-90.

27. Hardy, J. and D.J. Selkoe, *The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics.* Science, 2002. **297**(5580): p. 353-6.

28. Wolfe, M.S., *gamma-Secretase in biology and medicine.* Semin Cell Dev Biol, 2009. **20**(2): p. 219-24.

29. Narayanan, S., T. Sato, and M.S. Wolfe, *A C-terminal region of signal peptide peptidase defines a functional domain for intramembrane aspartic protease catalysis.* J Biol Chem, 2007. **282**(28): p. 20172-9.

30. Sato, T., et al., *Signal peptide peptidase: biochemical properties and modulation by nonsteroidal antiinflammatory drugs.* Biochemistry, 2006. **45**(28): p. 8649-56.

31. Weihofen, A., et al., *Identification of signal peptide peptidase, a presenilin-type aspartic protease.* Science, 2002. **296**(5576): p. 2215-8.

32. Urban, S., *Making the cut: central roles of intramembrane proteolysis in pathogenic microorganisms.* Nat Rev Microbiol, 2009. **7**(6): p. 411-23.

33. Koonin, E.V., et al., *The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers.* Genome Biol, 2003. **4**(3): p. R19.

34. Urban, S., D. Schlieper, and M. Freeman, *Conservation of intramembrane proteolytic activity and substrate specificity in prokaryotic and eukaryotic rhomboids.* Curr Biol, 2002. **12**(17): p. 1507-12.

35. Wolfe, M.S. and R. Kopan, *Intramembrane proteolysis: theme and variations.* Science, 2004. **305**(5687): p. 1119-23.

36. Freeman, M., *Rhomboid proteases and their biological functions.* Annu Rev Genet, 2008. **42**: p. 191-210.

37. Mayer, U. and C. Nusslein-Volhard, *A group of genes required for pattern formation in the ventral ectoderm of the Drosophila embryo.* Genes Dev, 1988. **2**(11): p. 1496-511.

38. Sturtevant, M.A., M. Roark, and E. Bier, *The Drosophila rhomboid gene mediates the localized formation of wing veins and interacts genetically with components of the EGF-R signaling pathway.* Genes Dev, 1993. **7**(6): p. 961-73.

39. Lee, J.R., et al., *Regulated intracellular ligand transport and proteolysis control EGF signal activation in Drosophila.* Cell, 2001. **107**(2): p. 161-71.

40. Urban, S., J.R. Lee, and M. Freeman, *Drosophila rhomboid-1 defines a family of putative intramembrane serine proteases.* Cell, 2001. **107**(2): p. 173-82.

41. Lal, M. and M. Caplan, *Regulated intramembrane proteolysis: signaling pathways and biological functions.* Physiology (Bethesda), 2011. **26**(1): p. 34-44.

42. Freeman, M., *Rhomboids: 7 years of a new protease family.* Semin Cell Dev Biol, 2009. **20**(2): p. 231-9.

43. Freeman, M., *The rhomboid-like superfamily: molecular mechanisms and biological roles.* Annu Rev Cell Dev Biol, 2014. **30**: p. 235-54.

44. Dealler, S.F., P.M. Hawkey, and M.R. Millar, *Enzymatic degradation of urinary indoxyl sulfate by Providencia stuartii and Klebsiella pneumoniae causes the purple urine bag syndrome.* J Clin Microbiol, 1988. **26**(10): p. 2152-6.

45. Gallio, M., et al., *A conserved mechanism for extracellular signaling in eukaryotes and prokaryotes.* Proc Natl Acad Sci U S A, 2002. **99**(19): p. 12208-13.

46. Stevenson, L.G., et al., *Rhomboid protease AarA mediates quorum-sensing in Providencia stuartii by activating TatA of the twin-arginine translocase.* Proc Natl Acad Sci U S A, 2007. **104**(3): p. 1003-8.

47. Lee, P.A., D. Tullman-Ercek, and G. Georgiou, *The bacterial twin-arginine translocation pathway.* Annu Rev Microbiol, 2006. **60**: p. 373-95.

48. Ejigiri, I., et al., *Shedding of TRAP by a rhomboid protease from the malaria sporozoite surface is essential for gliding motility and sporozoite infectivity.* PLoS Pathog, 2012. **8**(7): p. e1002725.

49. Howell, S.A., et al., *Distinct mechanisms govern proteolytic shedding of a key invasion protein in apicomplexan pathogens.* Mol Microbiol, 2005. **57**(5): p. 1342-56.

50. Sibley, L.D., *Intracellular parasite invasion strategies.* Science, 2004. **304**(5668): p. 248-53.

51. O'Donnell, R.A., et al., *Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite.* J Cell Biol, 2006. **174**(7): p. 1023-33.

52. Brossier, F., et al., *A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by Toxoplasma.* Proc Natl Acad Sci U S A, 2005. **102**(11): p. 4146-51.

53. Opitz, C., et al., *Intramembrane cleavage of microneme proteins at the surface of the apicomplexan parasite Toxoplasma gondii.* EMBO J, 2002. **21**(7): p. 1577-85.

54. Baker, R.P., R. Wijetilaka, and S. Urban, *Two Plasmodium rhomboid proteases preferentially cleave different adhesins implicated in all invasive stages of malaria.* PLoS Pathog, 2006. **2**(10): p. e113.

55. Singh, S., et al., *Mononeme: a new secretory organelle in Plasmodium falciparum merozoites identified by localization of rhomboid-1 protease.* Proc Natl Acad Sci U S A, 2007. **104**(50): p. 20043-8.

56. Urban, S. and S.W. Dickey, *The rhomboid protease family: a decade of progress on function and mechanism.* Genome Biol, 2011. **12**(10): p. 231.

57. Herlan, M., et al., *Alternative topogenesis of Mgm1 and mitochondrial morphology depend on ATP and a functional import motor.* J Cell Biol, 2004. **165**(2): p. 167-73.

58. Tatsuta, T., et al., *m-AAA protease-driven membrane dislocation allows intramembrane cleavage by rhomboid in mitochondria.* EMBO J, 2007. **26**(2): p. 325-35.

59. Herlan, M., et al., *Processing of Mgm1 by the rhomboid-type protease Pcp1 is required for maintenance of mitochondrial morphology and of mitochondrial DNA.* J Biol Chem, 2003. **278**(30): p. 27781-8.

60. McQuibban, G.A., S. Saurya, and M. Freeman, *Mitochondrial membrane remodelling regulated by a conserved rhomboid protease.* Nature, 2003. **423**(6939): p. 537-41.

61. Sesaki, H., et al., *Cells lacking Pcp1p/Ugo2p, a rhomboid-like protease required for Mgm1p processing, lose mtDNA and mitochondrial structure in a Dnm1p-dependent manner, but remain competent for mitochondrial fusion.* Biochem Biophys Res Commun, 2003. **308**(2): p. 276-83.

62. Meissner, C., et al., *The mitochondrial intramembrane protease PARL cleaves human Pink1 to regulate Pink1 trafficking.* J Neurochem, 2011. **117**(5): p. 856-67.

63. Shi, G., et al., *Functional alteration of PARL contributes to mitochondrial dysregulation in Parkinson's disease.* Hum Mol Genet, 2011. **20**(10): p. 1966-74.

64. McQuibban, G.A., et al., *Normal mitochondrial dynamics requires rhomboid-7 and affects Drosophila lifespan and neuronal function.* Curr Biol, 2006. **16**(10): p. 982-9.

65. Sekine, S., et al., *Rhomboid protease PARL mediates the mitochondrial membrane potential loss-induced cleavage of PGAM5.* J Biol Chem, 2012. **287**(41): p. 34635-45.

66. Civitarese, A.E., et al., *Regulation of skeletal muscle oxidative capacity and insulin signaling by the mitochondrial rhomboid protease PARL.* Cell Metab, 2010. **11**(5): p. 412-26.

67. Hatunic, M., et al., *The Leu262Val polymorphism of presenilin associated rhomboid like protein (PARL) is associated with earlier onset of type 2 diabetes and increased urinary microalbumin creatinine ratio in an Irish case-control population.* Diabetes Res Clin Pract, 2009. **83**(3): p. 316-9.

68. Walder, K., et al., *The mitochondrial rhomboid protease PSARL is a new candidate gene for type 2 diabetes.* Diabetologia, 2005. **48**(3): p. 459-68.

69. Lemberg, M.K. and M. Freeman, *Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases.* Genome Res, 2007. **17**(11): p. 1634-46.

70. Adrain, C., et al., *Mammalian EGF receptor activation by the rhomboid protease RHBDL2.* EMBO Rep, 2011. **12**(5): p. 421-7.

71. Fleig, L., et al., *Ubiquitin-dependent intramembrane rhomboid protease promotes ERAD of membrane proteins.* Mol Cell, 2012. **47**(4): p. 558-69.

72. Liao, H.J. and G. Carpenter, *Regulated intramembrane cleavage of the EGF receptor.* Traffic, 2012. **13**(8): p. 1106-12.

73. Pascall, J.C. and K.D. Brown, *Intramembrane cleavage of ephrinB3 by the human rhomboid family protease, RHBDL2.* Biochem Biophys Res Commun, 2004. **317**(1): p. 244-52.

74. Cheng, T.L., et al., *Functions of rhomboid family protease RHBDL2 and thrombomodulin in wound healing.* J Invest Dermatol, 2011. **131**(12): p. 2486-94.

75. Ren, X., et al., *Rhomboid domain containing 1 inhibits cell apoptosis by upregulating AP-1 activity and its downstream target Bcl-3.* FEBS Lett, 2013. **587**(12): p. 1793-8.

76. Wan, C., et al., *Exosome-related multi-pass transmembrane protein TSAP6 is a target of rhomboid protease RHBDD1-induced proteolysis.* PLoS One, 2012. **7**(5): p. e37452.

77. Wang, Y., et al., *A novel member of the Rhomboid family, RHBDD1, regulates BIK-mediated apoptosis.* Cell Mol Life Sci, 2008. **65**(23): p. 3822-9.

78. Brooks, C.L. and M.J. Lemieux, *Untangling structure-function relationships in the rhomboid family of intramembrane proteases.* Biochim Biophys Acta, 2013. **1828**(12): p. 2862-72.

79. of, U.s.f.r.i.t.r.f., M.K. intramembrane proteaseLemberg, and M. Freeman, *Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases.* Genome Res, 2007. **17**(11): p. 1634-46.

80. Schechter, I. and A. Berger, *On the size of the active site in proteases. I. Papain.* Biochem Biophys Res Commun, 1967. **27**(2): p. 157-62.

81. Wang, Y., Y. Zhang, and Y. Ha, *Crystal structure of a rhomboid family intramembrane protease.* Nature, 2006. **444**(7116): p. 179-80.

82. Strisovsky, K., *Structural and mechanistic principles of intramembrane proteolysis--lessons from rhomboids.* FEBS J, 2013. **280**(7): p. 1579-603.

83. Ben-Shem, A., D. Fass, and E. Bibi, *Structural basis for intramembrane proteolysis by rhomboid serine proteases.* Proc Natl Acad Sci U S A, 2007. **104**(2): p. 462-6.

84. Lemieux, M.J., et al., *The crystal structure of the rhomboid peptidase from Haemophilus influenzae provides insight into intramembrane proteolysis.* Proc Natl Acad Sci U S A, 2007. **104**(3): p. 750-4.

85. Wu, Z., et al., *Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry.* Nat Struct Mol Biol, 2006. **13**(12): p. 1084-91.

86. Vinothkumar, K.R., et al., *Structure of rhomboid protease in complex with beta-lactam inhibitors defines the S2' cavity.* Structure, 2013. **21**(6): p. 1051-8.

87. Vinothkumar, K.R., et al., *The structural basis for catalysis and substrate specificity of a rhomboid protease.* EMBO J, 2010. **29**(22): p. 3797-809.

88. Vosyka, O., et al., *Activity-based probes for rhomboid proteases discovered in a mass spectrometry-based assay.* Proc Natl Acad Sci U S A, 2013. **110**(7): p. 2472-7.

89. Xue, Y., et al., *Conformational change in rhomboid protease GlpG induced by inhibitor binding to its S' subsites.* Biochemistry, 2012. **51**(18): p. 3723-31.

90. Xue, Y. and Y. Ha, *Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate.* J Biol Chem, 2012. **287**(5): p. 3099-107.

91. Zoll, S., et al., *Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures.* EMBO J, 2014. **33**(20): p. 2408-21.

92. Baker, R.P. and S. Urban, *Cytosolic extensions directly regulate a rhomboid protease by modulating substrate gating.* Nature, 2015. **523**(7558): p. 101-5.

93. Urban, S. and R.P. Baker, *In vivo analysis reveals substrate-gating mutants of a rhomboid intramembrane protease display increased activity in living cells.* Biol Chem, 2008. **389**(8): p. 1107-15.

94. Xue, Y. and Y. Ha, *Large lateral movement of transmembrane helix S5 is not required for substrate access to the active site of rhomboid intramembrane protease.* J Biol Chem, 2013. **288**(23): p. 16645-54.

95. Wang, Y. and Y. Ha, *Open-cap conformation of intramembrane protease GlpG.* Proc Natl Acad Sci U S A, 2007. **104**(7): p. 2098-102.

96. Dickey, S.W., et al., *Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity.* Cell, 2013. **155**(6): p. 1270-81.

97. Wang, Y., et al., *The role of L1 loop in the mechanism of rhomboid intramembrane protease GlpG.* J Mol Biol, 2007. **374**(4): p. 1104-13.

98. Brooks, C.L., et al., *Insights into substrate gating in H. influenzae rhomboid.* J Mol Biol, 2011. **407**(5): p. 687-97.

99. Moin, S.M. and S. Urban, *Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics.* Elife, 2012. **1**: p. e00173.

100. Ng, N.M., R.N. Pike, and S.E. Boyd, *Subsite cooperativity in protease specificity.* Biol Chem, 2009. **390**(5-6): p. 401-7.

101. Schechter, I., *Mapping of the active site of proteases in the 1960s and rational design of inhibitors/drugs in the 1990s.* Curr Protein Pept Sci, 2005. **6**(6): p. 501-12.

102. Urban, S. and M.S. Wolfe, *Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity.* Proc Natl Acad Sci U S A, 2005. **102**(6): p. 1883-8.

103. Pierrat, O.A., et al., *Monocyclic beta-lactams are selective, mechanism-based inhibitors of rhomboid intramembrane proteases.* ACS Chem Biol, 2011. **6**(4): p. 325-35.

104. Maegawa, S., K. Ito, and Y. Akiyama, *Proteolytic action of GlpG, a rhomboid protease in the Escherichia coli cytoplasmic membrane.* Biochemistry, 2005. **44**(41): p. 13543-52.

105. Akiyama, Y. and S. Maegawa, *Sequence features of substrates required for cleavage by GlpG, an Escherichia coli rhomboid protease.* Mol Microbiol, 2007. **64**(4): p. 1028-37.

106. Strisovsky, K., H.J. Sharpe, and M. Freeman, *Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates.* Mol Cell, 2009. **36**(6): p. 1048-59.

107. Urban, S. and M. Freeman, *Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain.* Mol Cell, 2003. **11**(6): p. 1425-34.

108. Walker, B. and J.F. Lynas, *Strategies for the inhibition of serine proteases.* Cell Mol Life Sci, 2001. **58**(4): p. 596-624.

109. Vagner, J., H. Qu, and V.J. Hruby, *Peptidomimetics, a synthetic tool of drug discovery.* Curr Opin Chem Biol, 2008. **12**(3): p. 292-6.

110. Zhong, J. and W.C. Groutas, *Recent developments in the design of mechanism-based and alternate substrate inhibitors of serine proteases.* Curr Top Med Chem, 2004. **4**(12): p. 1203-16.

111. Colombo, E., et al., *Design and synthesis of potent, selective inhibitors of matriptase.* ACS Med Chem Lett, 2012. **3**(7): p. 530-4.

112. Sanderson, P.E., *Small, noncovalent serine protease inhibitors.* Med Res Rev, 1999. **19**(2): p. 179-97.

113. Powers, J.C., et al., *Irreversible inhibitors of serine, cysteine, and threonine proteases.* Chem Rev, 2002. **102**(12): p. 4639-750.

114. Silverman, R.B., *Mechanism-based enzyme inactivators.* Methods Enzymol, 1995. **249**: p. 240-83.

115. Lipinski, C.A., *Lead- and drug-like compounds: the rule-of-five revolution.* Drug Discov Today Technol, 2004. **1**(4): p. 337-41.

116. Lipinski, C.A., et al., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.* Adv Drug Deliv Rev, 2001. **46**(1-3): p. 3-26.

117. Klebe, G., *Virtual ligand screening: strategies, perspectives and limitations.* Drug Discov Today, 2006. **11**(13-14): p. 580-94.

118. Cavasotto, C.N. and A.J. Orry, *Ligand docking and structure-based virtual screening in drug discovery.* Curr Top Med Chem, 2007. **7**(10): p. 1006-14.

119. Kroemer, R.T., *Structure-based drug design: docking and scoring.* Curr Protein Pept Sci, 2007. **8**(4): p. 312-28.

120. Sun, H., *Pharmacophore-based virtual screening.* Curr Med Chem, 2008. **15**(10): p. 1018-24.

121. Lengauer, T., et al., *Novel technologies for virtual screening.* Drug Discov Today, 2004. **9**(1): p. 27-34.

122. Rester, U., *From virtuality to reality - Virtual screening in lead discovery and lead optimization: a medicinal chemistry perspective.* Curr Opin Drug Discov Devel, 2008. **11**(4): p. 559-68.

123. Wolf, E.V. and S.H. Verhelst, *Inhibitors of rhomboid proteases.* Biochimie, 2015.

124. Wolf, E.V., A. Zeissler, and S.H. Verhelst, *Inhibitor Fingerprinting of Rhomboid Proteases by Activity-Based Protein Profiling Reveals Inhibitor Selectivity and Rhomboid Autoprocessing.* ACS Chem Biol, 2015. **10**(10): p. 2325-33.

125. Wolf, E.V., et al., *A new class of rhomboid protease inhibitors discovered by activity-based fluorescence polarization.* PLoS One, 2013. **8**(8): p. e72307.

126. Knopf, R.R. and Z. Adam, *Rhomboid proteases in plants - still in square one?* Physiol Plant, 2012. **145**(1): p. 41-51.

127. Salazar, O. and J.A. Asenjo, *Enzymatic lysis of microbial cells.* Biotechnol Lett, 2007. **29**(7): p. 985-94.

128. Simpson, R.J., *Disruption of cultured cells by nitrogen cavitation.* Cold Spring Harb Protoc, 2010. **2010**(11): p. pdb prot5513.

129. Cheung, R.C., J.H. Wong, and T.B. Ng, *Immobilized metal ion affinity chromatography: a review on its applications.* Appl Microbiol Biotechnol, 2012. **96**(6): p. 1411-20.

130. DelMar, E.G., et al., *A sensitive new substrate for chymotrypsin.* Anal Biochem, 1979. **99**(2): p. 316-20.

131. Scholz, C., et al., *DOCKTITE-a highly versatile step-by-step workflow for covalent docking and virtual screening in the molecular operating environment.* J Chem Inf Model, 2015. **55**(2): p. 398-406.

132. Hopkins, A.L., C.R. Groom, and A. Alex, *Ligand efficiency: a useful metric for lead selection.* Drug Discov Today, 2004. **9**(10): p. 430-1.

133. Guzel, O. and A. Salman, *Synthesis, antimycobacterial and antitumor activities of new (1,1-dioxido-3-oxo-1,2-benzisothiazol-2(3H)-yl)methyl N,N-disubstituted dithiocarbamate/O-alkyldithiocarbonate derivatives.* Bioorg Med Chem, 2006. **14**(23): p. 7804-15.

134. D'Ascenzio, M., et al., *Design, synthesis and evaluation of N-substituted saccharin derivatives as selective inhibitors of tumor-associated carbonic anhydrase XII.* Bioorg Med Chem, 2014. **22**(6): p. 1821-31.

135. Groutas, W.C., et al., *Design, synthesis, and in vitro inhibitory activity toward human leukocyte elastase, cathepsin G, and proteinase 3 of saccharin-derived sulfones and congeners.* Bioorg Med Chem, 1996. **4**(9): p. 1393-400.

136. Subramanyam, C., et al., *2,6-Disubstituted aryl carboxylic acids, leaving groups "par excellence" for benzisothiazolone inhibitors of human leukocyte elastase.* J Med Chem, 1994. **37**(17): p. 2623-6.

137. Kim, S.-K., Moon, J.-G., Lee, S.-G., Choi, S.-Y., Cho, S.-D., Choi, W.-Y. and Yoon, Y.-J., *Retro-ene reaction I: reaction of N-(hydroxymethyl)saccharin with benzoyl chlorides and alkyl halides.* Journal of Heterocyclic Chemistry, 1993. **30**(2): p. 353–356.

138. William C. Groutas, H.H., Radhika Venkataraman, Nadene Houser-Archield, Jeffrey B. Epp, *Dual-Action Inhibitors of Proteolytic Enzymes: Potential Therapeutic Agents for Cystic Fibrosis and Related Ailments.* Bioorganic & medicinal chemistry letters, 1993. **1**(4): p. 273-277.

139. Krantz, A., et al., *Design and synthesis of 4H-3,1-benzoxazin-4-ones as potent alternate substrate inhibitors of human leukocyte elastase.* J Med Chem, 1990. **33**(2): p. 464-79.

140. Anwar, M. and Anwar, *Reactions of some 4(3H)quinazolinones.* Revue roumaine de chimie, 1981. **26**(4): p. 639-645.

141. A., K.Z., et al., *Synthesis and antimicrobial activity of 2-aryl-4H-3,1-benzoxazin-4-ones.* Asian J. Chem. , 2013. **25**(1): p. 152-156.

142. Mujeeb Ur, R., et al., *Synthesis and characterization of quinazoline derivatives: search for hybrid molecule as diuretic and antihypertensive agents.* J Enzyme Inhib Med Chem, 2014. **29**(5): p. 733-43.

143. Bozdag, M., et al., *Benzenesulfonamides incorporating bulky aromatic/heterocyclic tails with potent carbonic anhydrase inhibitory activity.* Bioorg Med Chem, 2015. **23**(24): p. 7751-64.

144. Khan, Z., *Synthesis of 2-Aryl-4H-3,1-Benzoxazin-4-ones: A Class of a-Chymotrypsin Inhibitors.* Asian Journal of Chemistry, 2014. **26**(15): p. 4561-4565.

145. Lu, W., et al., *Synthesis, crystal structure and biological evaluation of substituted quinazolinone benzoates as novel antituberculosis agents targeting acetohydroxyacid synthase.* Eur J Med Chem, 2015. **94**: p. 298-305.

146. Jagani, C., et al., *Microwave promoted synthesis and antimicrobial activity of 3-thiazole substituted 2-styryl-4(3H)-quinazolinone derivatives.* Journal of Saudi Chemical Society, 2012. **16**(4): p. 363-369.

147. Costanzo, M.J., et al., *In-depth study of tripeptide-based alpha-ketoheterocycles as inhibitors of thrombin. Effective utilization of the S1' subsite and its implications to structure-based drug design.* J Med Chem, 2005. **48**(6): p. 1984-2008.

148. Lin, G., et al., *Inhibitors selective for mycobacterial versus human proteasomes.* Nature, 2009. **461**(7264): p. 621-6.

149. Gezginci, M.H., A.R. Martin, and S.G. Franzblau, *Antimycobacterial activity of substituted isosteres of pyridine- and pyrazinecarboxylic acids. 2.* J Med Chem, 2001. **44**(10): p. 1560-3.

150. Mathews, T.P., et al., *Discovery, biological evaluation, and structure-activity relationship of amidine based sphingosine kinase inhibitors.* J Med Chem, 2010. **53**(7): p. 2766-78.

151. Martyn, D.C., M.J. Moore, and A.D. Abell, *Succinimide and saccharin-based enzyme-activated inhibitors of serine proteases.* Curr Pharm Des, 1999. **5**(6): p. 405-15.

152. Mendgen, T., C. Steuer, and C.D. Klein, *Privileged scaffolds or promiscuous binders: a comparative study on rhodanines and related heterocycles in medicinal chemistry.* J Med Chem, 2012. **55**(2): p. 743-53.

153. Bhatti, R.S., et al., *Recent pharmacological developments on rhodanines and 2,4-thiazolidinediones.* Int J Med Chem, 2013. **2013**: p. 793260.

154. Dolezel, J., et al., *Rhodanineacetic acid derivatives as potential drugs: preparation, hydrophobic properties and antifungal activity of (5-arylalkylidene-4-oxo-2-thioxo-1,3-thiazolidin-3-yl)acetic acids.* Molecules, 2009. **14**(10): p. 4197-212.

155. Kakiuchi, N., et al., *Non-peptide inhibitors of HCV serine proteinase.* FEBS Lett, 1998. **421**(3): p. 217-20.

156. Sing, W.T., et al., *Arylalkylidene rhodanine with bulky and hydrophobic functional group as selective HCV NS3 protease inhibitor.* Bioorg Med Chem Lett, 2001. **11**(2): p. 91-4.

157. Leung, D., G. Abbenante, and D.P. Fairlie, *Protease inhibitors: current status and future prospects.* J Med Chem, 2000. **43**(3): p. 305-41.

158. Bruno, G., et al., *Synthesis and aldose reductase inhibitory activity of 5-arylidene-2,4-thiazolidinediones.* Bioorg Med Chem, 2002. **10**(4): p. 1077-84.

159. Fujishima, H. and K. Tsubota, *Improvement of corneal fluorescein staining in post cataract surgery of diabetic patients by an oral aldose reductase inhibitor, ONO-2235.* Br J Ophthalmol, 2002. **86**(8): p. 860-3.

160. Grant, E.B., et al., *The synthesis and SAR of rhodanines as novel class C beta-lactamase inhibitors.* Bioorg Med Chem Lett, 2000. **10**(19): p. 2179-82.

161. Sim, M.M., et al., *Benzylidene rhodanines as novel inhibitors of UDP-N-acetylmuramate/L-alanine ligase.* Bioorg Med Chem Lett, 2002. **12**(4): p. 697-9.

162. Momose, Y., et al., *Studies on antidiabetic agents. X. Synthesis and biological activities of pioglitazone and related compounds.* Chem Pharm Bull (Tokyo), 1991. **39**(6): p. 1440-5.

163. Johnson, S.L., et al., *Anthrax lethal factor protease inhibitors: synthesis, SAR, and structure-based 3D QSAR studies.* J Med Chem, 2006. **49**(1): p. 27-30.

164. Johnson, S.L., et al., *Rhodanine derivatives as selective protease inhibitors against bacterial toxins.* Chem Biol Drug Des, 2008. **71**(2): p. 131-9.

165. Anumala, U.R., et al., *Fluorescent rhodanine-3-acetic acids visualize neurofibrillary tangles in Alzheimer's disease brains.* Bioorg Med Chem, 2013. **21**(17): p. 5139-44.

166. Ohishi, Y., et al., *Preparations of 5-alkylmethylidene-3-carboxymethylrhodanine derivatives and their aldose reductase inhibitory activity.* Chem Pharm Bull (Tokyo), 1990. **38**(7): p. 1911-9.

167. Zimmerman, M., et al., *Inhibition of elastase and other serine proteases by heterocyclic acylating agents.* J Biol Chem, 1980. **255**(20): p. 9848-51.

168. Hlasta, D.J., et al., *Orally bioavailable benzisothiazolone inhibitors of human leukocyte elastase.* J Med Chem, 1995. **38**(5): p. 739-44.

169. William C. Groutas, M.J.B., Radhika Venkataraman, Jeffrey B. Epp, Nadene Houser-Archield, Lee S. Chong, Jerald J. McClenahan, *POTENTIAL MECHANISM-BASED INHIBITORS OF PROTEOLYI'IC ENZYMES.* Bioorganic & medicinal chemistry letters, 1992. **2**(2): p. 175-180.

170. William C.Groutas, N.-A., LeeS.Chong, RadhikaVenkataraman, JeffreyB.Epp, He Huang, and Jerald J. McClenahan, *Efficient Inhibition of Human Leukocyte Elastase and Cathepsin G by Saccharin Derivatives.* Journal of medicinal chemistry, 1993. **36**: p. 3178-3181.

171. William C. Groutas, J.B.E., Radhika Venkataraman, Rongze Kuang, Tien My Truong, Jerry J. McClenahan, and Om Prakash, *Design, Synthesis, and In Vitro Inhibitory Activity Toward Human Leukocyte Elastase, Cathepsin G, and Proteinase 3 of Saccharin-Derived Sulfones and Congeners.* Bioorganic & Medicinal Chemistry, 1996. **4**(9): p. 1393-1400.

172. William C. Groutas, L.S.C., Radhika Venkataraman, Rongze Kuang, Jeff B. Epp, Nadene Houser-Archield, He Huang, and John R. Hoidal, *Amino Acid-Derived Phthalimide and Saccharin Derivatives as Inhibitors of Human Leukocyte Elastase, Cathepsin G, and Proteinase 3.* Archives of biochemistry and biophysics, 1996. **332**(2): p. 335-340.

173. William C. Groutas, R.K., Sumei Ruan, Jeffrey B. Epp, Radhika Venkataraman and Tien M. Truong, *Potent and Specific Inhibition of Human Leukocyte Elastase, Cathepsin G and Proteinase 3 by Sulfone Derivatives Employing the 1,2,5-Thiadiazolidin-3-one 1,1 D.* Bioorganic & Medicinal Chemistry, 1998. **6**: p. 661-671.

174. Hlasta, D.J., et al., *A novel class of cyclic beta-dicarbonyl leaving groups and their use in the design of benzisothiazolone human leukocyte elastase inhibitors.* J Med Chem, 1995. **38**(23): p. 4687-92.

175. Jakopin, Z. and M.S. Dolenc, *Advances in the chemistry of saccharins: from synthetic novelties towards biologically active compounds.* Curr Med Chem, 2010. **17**(7): p. 651-71.

176. Dennis J. Hlasta, J.J.C., Ranjit C. Desai, Timothy G. Talomie, and Jian Shen, *A COMPARATIVE SAR AND COMPUTER MODELING STUDY OF BENZISOTHIAZOLONE, MECHANISM-BASED INHIBITORS WITH PORCINE PANCREATIC AND HUMAN LEUKOCYTE ELASTASE.* Bioorganic & medicinal chemistry letters, 1996. **6**(24): p. 2941-2946.

177. Dennis J. Hlasta, M.R.B., Neil W. Boax, John J. Court, Ranjit C. Desai, Catherine A. Franke,t Albert J. Mum, Chakmpani Subramanyam, and Richard P. Dunlapt, *A benzothiszolone class of potent, selective mechanism-based inhibitors of human leukocyte elastase.* Bioorganic & medicinal chemistry, 1994.

178. Dennis J. Hlasta, ' Chakrapani Subramanyam,?Malcolm R. Bell,t Philip M. Carabateas,t John J. Court,? Ranjit C. Desai,t Marion L. Drozd,$W. Mark Eickhoff,t Edward W. Ferguson,l Robert J. Gordon,ll, V.K. Judith A. Johnson, ?Alan L. Maycock,l Karl R. Mueller,t Edward D. Pagani,", and M.T.S. David T. Robinson, v Paul J . Silver," and Sandhya Subramanian, *Orally Bioavailable Benzisothiazolone Inhibitors of Human Leukocyte Elastase.* Journal of medicinal chemistry, 1995. **38**(5).

179. Ranjit C. Desai, S.J.C.C., Edward Ferguson, Robert J. Gordon, and Dennis J. Hlasta, *Phosphonates and Phosphinates: Novel Leaving Groups for Benzisothiazolone Inhibitors of Human Leukocyte Elastase.* Journal of medicinal chemistry, 1995. **38**: p. 1571-1574.

180. Chakrapani Subramanyam, M.R.B., Philip Carabateas, John J. Court, John A. Dority, Jr., Edward Ferguson, Robert Gordon, Dennis J. Hlasta, Virendra Kumar, and Manohar Saindane, *2,6-DisubstitutedAryl Carboxylic Acids, Leaving Groups "Par Excellence" for Benzisothiazolone Inhibitors of Human Leukocyte Elastase.* Journal of medicinal chemistry, 1994.

181. Morris Zimmerman, H.M., Dennis Mulvey, HowardJonesl, Robert Frankshun, and Bonnie M. Ashe, *Inhibition of Elastase and Other Serine Proteasesby Heterocyclic Acylating Agent.* The Journal of biological chemistry, 1980. **255**(20): p. 9848-9851.

182. Gutschow, M., et al., *Studies on 2-benzyloxy-4H-3,1-benzoxazin-4-ones as serine protease inhibitors.* Pharm Acta Helv, 1998. **73**(2): p. 95-103.

183. Srivastava, V., et al., *Disubstituted 4(3H) quinazolones: a novel class of antitumor agents.* Chem Biol Drug Des, 2009. **74**(3): p. 297-301.

184. Tripathi, H.L. and W.L. Dewey, *Comparison of the effects of diisopropylfluorophosphate, sarin, soman, and tabun on toxicity and brain acetylcholinesterase activity in mice.* J Toxicol Environ Health, 1989. **26**(4): p. 437-46.

185. Ito, A.S., et al., *Fluorescent properties of amino acids labeled with ortho-aminobenzoic acid.* Biospectroscopy, 1998. **4**(6): p. 395-402.

186. Narendra, D., J.E. Walker, and R. Youle, *Mitochondrial quality control mediated by PINK1 and Parkin: links to parkinsonism.* Cold Spring Harb Perspect Biol, 2012. **4**(11).

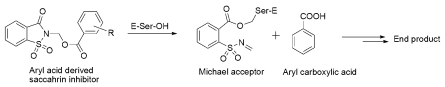
187. Colombo, É., et al., *Design and Synthesis of Potent, Selective Inhibitors of Matriptase.* ACS Medicinal Chemistry Letters, 2012. **3**(7): p. 530-534.

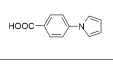
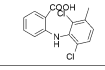
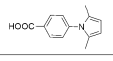
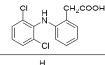
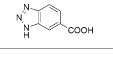
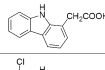
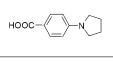
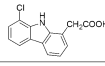
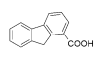
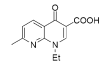
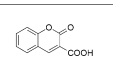
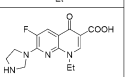
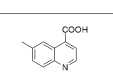
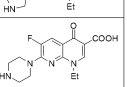
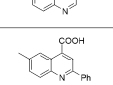
188. Groutas, W.C., et al., *Efficient inhibition of human leukocyte elastase and cathepsin G by saccharin derivatives.* J Med Chem, 1993. **36**(21): p. 3178-81.

7 Appendix

7.1 Molecular docking data

7.1.1 List of aryl acid derived saccharin molecules with variable ring substitutions



ID	Leaving group (LG)	ID	Leaving group (LG)	ID	Leaving group (LG)	ID	Leaving group (LG)	ID	Leaving group (LG)
ACA1		ACA24		ACA47		ACA70		ACA93	
ACA2		ACA25		ACA48		ACA71		ACA94	
ACA3		ACA26		ACA49		ACA72		ACA95	
ACA4		ACA27		ACA50		ACA73		ACA96	
ACA5		ACA28		ACA51		ACA74		ACA97	
ACA6		ACA29		ACA52		ACA75		ACA98	
ACA7		ACA30		ACA53		ACA76		ACA99	
ACA8		ACA31		ACA54		ACA77		ACA100	
ACA9		ACA32		ACA55		ACA78		ACA101	
ACA10		ACA33		ACA56		ACA79		ACA102	
ACA11		ACA34		ACA57		ACA80		ACA103	
ACA12		ACA35		ACA58		ACA81			
ACA13		ACA36		ACA59		ACA82			
ACA14		ACA37		ACA60		ACA83			
ACA15		ACA38		ACA61		ACA84			
ACA16		ACA39		ACA62		ACA85			
ACA17		ACA40		ACA63		ACA86			
ACA18		ACA41		ACA64		ACA87			
ACA19		ACA42		ACA65		ACA88			
ACA20		ACA43		ACA66		ACA89			
ACA21		ACA44		ACA67		ACA90			
ACA22		ACA45		ACA68		ACA91			
ACA23		ACA46		ACA69		ACA92			

where leaving group = Aryl carboxylic acid (ACA)

7.1.2 A molecular docking model of the saccharin inhibitor with LG38 in the GlpG pocket

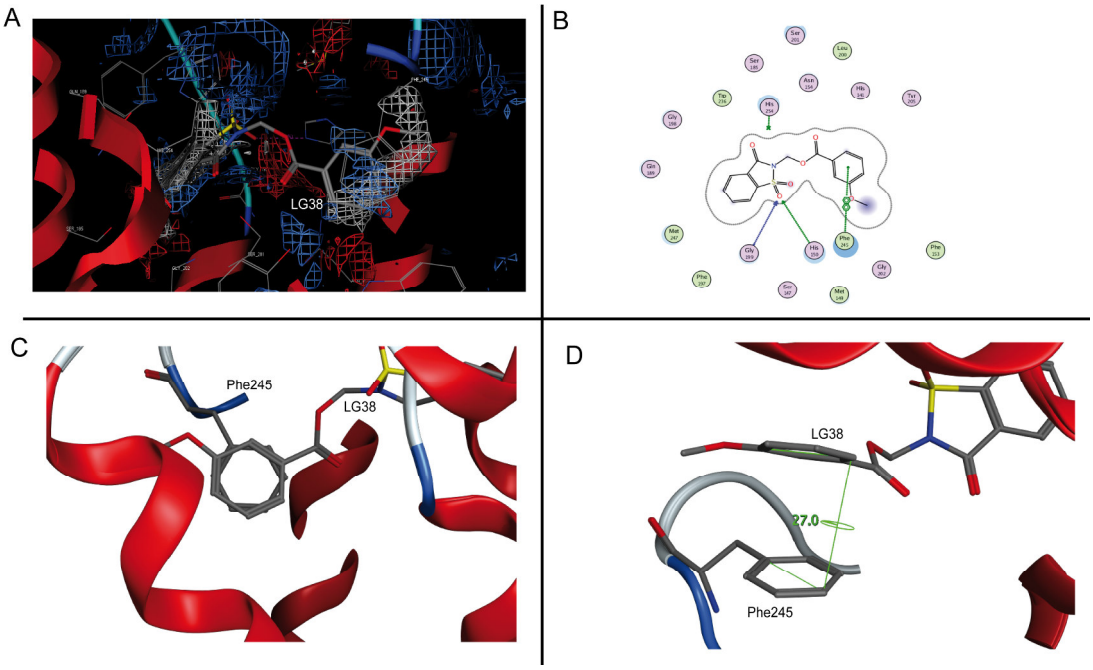


Figure 7.1 A molecular docking model of the saccharin derivative with LG38 in GlpG pocket. A) Docking model of LG38 with electrostatic surface map where blue color indicates positive regions (H-donor), red color indicates negative regions (H-acceptor) and green color signifies neutral (hydrophobic) regions. B) Ligand interaction map of LG38 in glpG pocket. C) pi-pi interaction between aryl ring of LG38 and Phe245 (distance = 3.86 Å). D) The angle between the plane of aryl rings involved in pi-pi interaction was 27°.

7.1.3 A molecular docking model of the saccharin derivatives ACA30 and ACA92 in the GlpG pocket

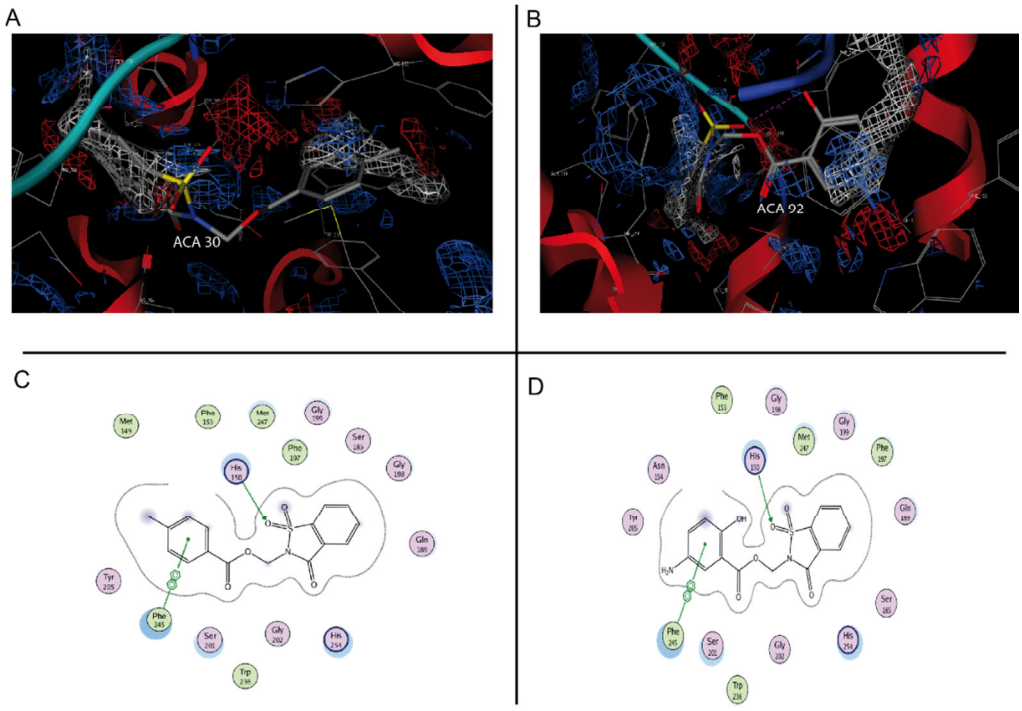
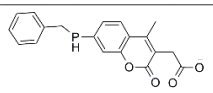
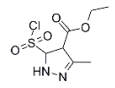
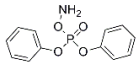
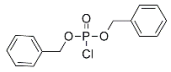
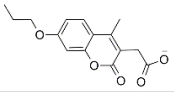
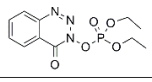
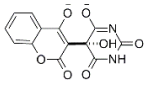
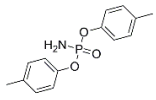
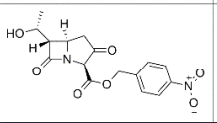
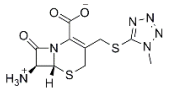


Figure 7.2: A molecular docking model of the saccharin derivatives ACA30 and ACA92 in the GlpG pocket. A) Docking model of ACA30 with electrostatic surface map where blue color indicates positive regions (H-donor), red color

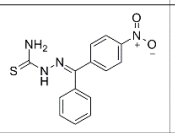
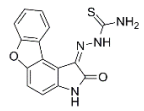
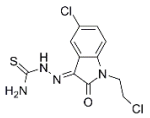
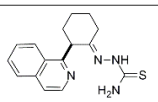
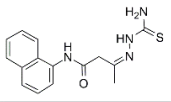
indicates negative regions (H-acceptor) and green color signifies neutral (hydrophobic) regions. B) Docking model of ACA92 with electrostatic surface map. Both A) and B) indicates that both core aryl rings of inhibitors fit perfectly into the favoured hydrophobic region but except this there was no significant interactions observed, C) Ligand interaction map of GlpG with ACA30. D) Ligand interaction map of GlpG with ACA92. Both the maps show the formation of two H-bonds between the sulfoxide moiety of ACA30 and ACA92 with the GlpG residues His150 and pi-pi interaction between the aryl ring and Phe245 of GlpG. But no improved interactions with extended aryl extensions.

7.1.4 List of virtual screening top ranked ligands

Top ten ligands of rescored_final_r database

Mol Name	Entry no.	Data base	Structure	Reaction rule	calculated ligand efficiency (cLE)
ZINC05442813	16	13		Coumarin_to_Tantal.rxn	-0.0969
ZINC8113220	20	14		Sulfonylhalogenid_to_Tantal.rxn	-0.0928
ZINC55168505	25	6		Phosphan_LG_to_Tantal.rxn	-0.0902
ZINC72399674	37	3		Phosphan_LG_to_Tantal.rxn	-0.0753
ZINC02092745	43	3		Coumarin_to_Tantal.rxn	-0.0726
ZINC02573666	58	3		Phosphan_LG_to_Tantal.rxn	-0.0627
ZINC08764859	59	15		Coumarin_to_Tantal.rxn	-0.0615
ZINC03177070	69	6		Phosphan_LG_to_Tantal.rxn	-0.0555
ZINC16696721	70	2		betalacton lactum to Tantal.rxn	-0.0540
ZINC02809614	79	9		betalacton_lactum_to_Tantal.rxn	-0.0500

Top five ligands of rescored_final_s_database

Mol Name	Entry no.	Data base	Structure	Reaction rule	calculated ligand efficiency (cLE)
ZINC05772291	3	14		Thiosemicarbazon_to_Tantal.rxn	-0.0914
ZINC06000122	4	14		Thiosemicarbazon_to_Tantal.rxn	-0.0871
ZINC13570979	6	3		Thiosemicarbazon_to_Tantal.rxn	-0.0749
ZINC05244012	22	7		Thiosemicarbazon_to_Tantal.rxn	-0.0500
ZINC33751351	49	2		Thiosemicarbazon_to_Tantal.rxn	-0.0260

7.2 Biological experiments: *in vitro* GlpG activity assay with Saccharin inhibitors

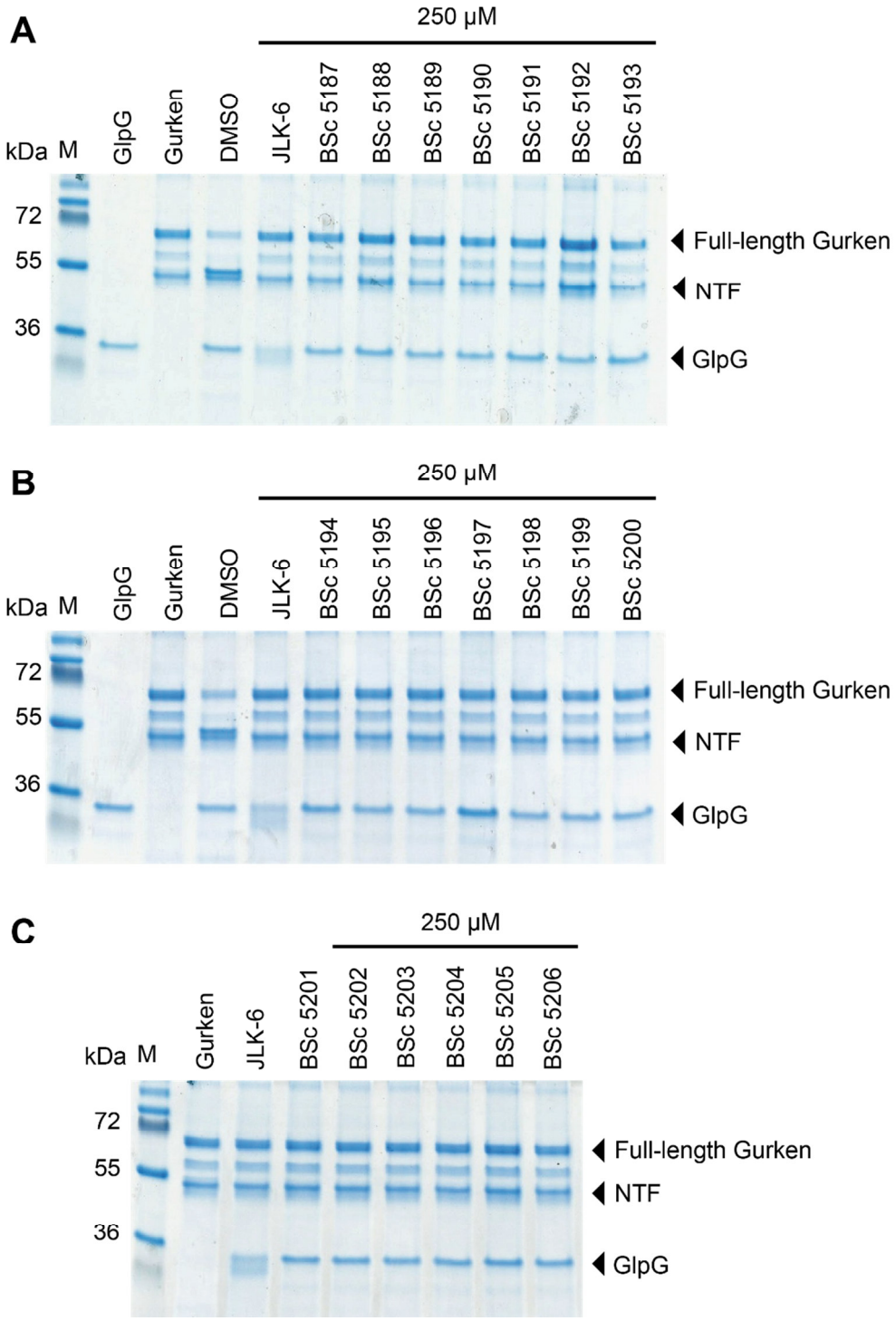


Figure 7.3: in vitro GlpG assay of saccharin inhibitors (BSc5187-BSc5206) at 250 µM concentration. All aryl acid derived N-methylated saccharins inhibits GlpG at tested concentration. The arrows indicate the full-length Gurken substrate (~ 66 kDa), the N-terminal Gurken fragment (NTF, ~ 55 kDa) and GlpG (~ 33 kDa). M: Molecular weight marker. JLK-6: reference isocoumarin inhibitor.

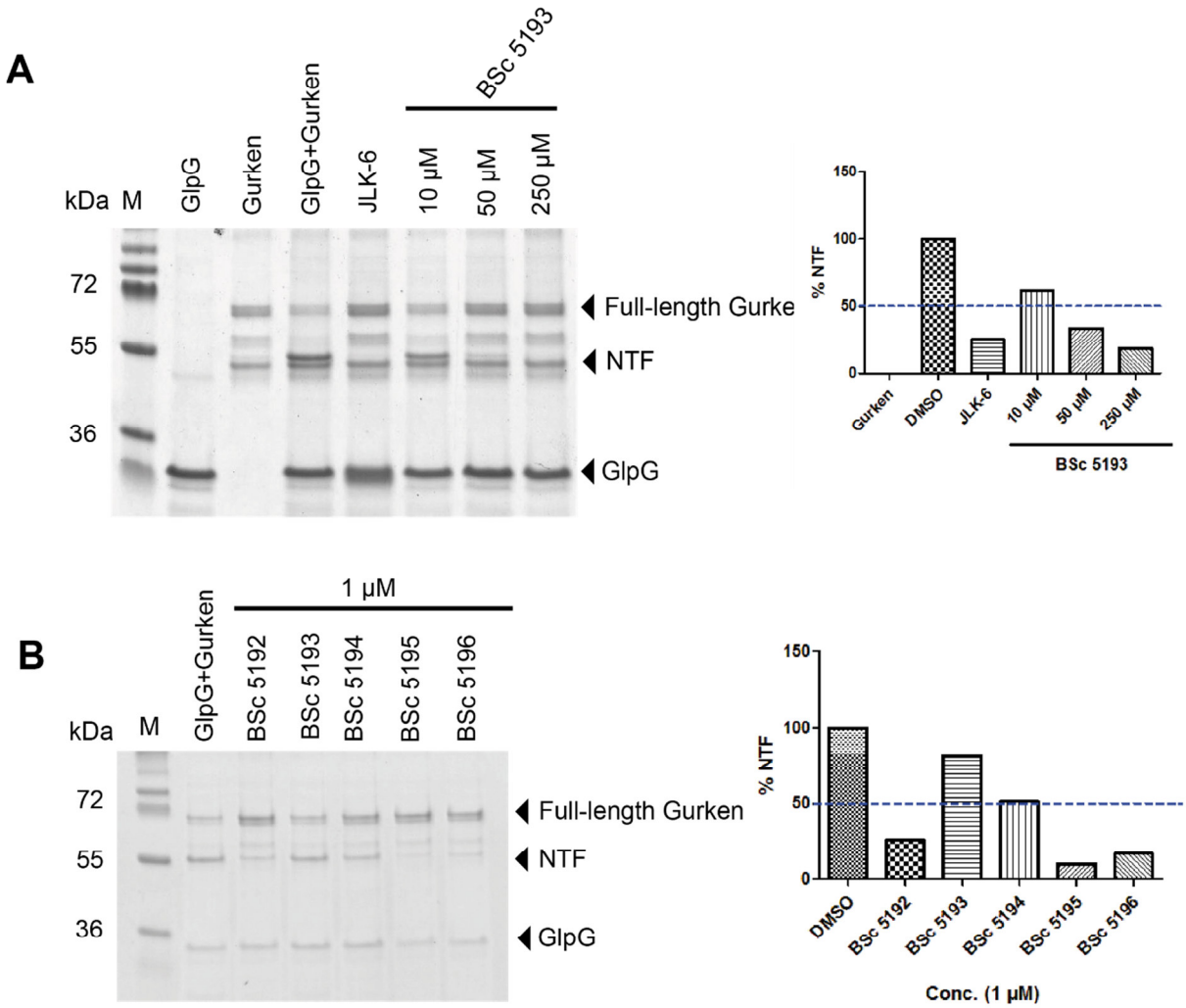


Figure 7.4: *in vitro* GlpG assay of saccharin inhibitors. A) BSc5193 with about 50% inhibition at 10 µM concentration. B) BSc5192 -BSc5196 tested at 1 µM. The arrows indicate the full-length Gurken substrate (~ 66 kDa), the N-terminal Gurken fragment (NTF, ~ 55 kDa) and GlpG (~ 33 kDa). M: Molecular weight marker. JLK-6: reference isocoumarin inhibitor.

7.3 Curriculum Vitae

Personal Information

Name	Parul Goel
Date and place of birth	August 28, 1986 in Gohana
Citizenship	Indian
Correspondence address	Robens str. 24, 52070 Aachen
Permanent address	113-C, Sector-5, Part-3, Gurgaon, Haryana-122001

Research experience

09/2014 – 12/2015	PhD studies at Department of Neuropathology, University Hospital, Heinrich Heine University (HHU) Duesseldorf, Germany Supervisors: Prof. Dr. Sascha Weggen and Prof. Dr. Boris Schmidt
11/2012 – 08/2014	PhD studies at Department of Organic Chemistry, TU Darmstadt, Germany Supervisors: Prof. Dr. Sascha Weggen and Prof. Dr. Boris Schmidt
09/2011 – 05/2012	Master Thesis at Department of Biotechnology, RWTH Aachen University, Germany Supervisor: Prof. Dr. Ulrich Schwaneberg
05/2011 – 08/2011	Minor project, Biophysical lab, Indian Institute of Technology (IIT) Delhi, India Supervisor: Associate Prof. Dr. Shashank Deep

Qualifications

06/2010 – 06/2012	Master of Technology (M. Tech.) in Molecular Engineering: chemical synthesis and analysis (CGPA = 9.5/10) Indian Institute of Technology (IIT) Delhi, India
06/2008 – 05/2010	Master of Science (M. Sc.) in Organic Chemistry (7.5/10) Jamia Millia Islamia Delhi, India
07/2004 – 05/2007	Bachelor of Science (B. Sc.) in Life Sciences (7.4/10) Gargi College, Delhi University, New Delhi, India

Awards and Scholarships

07/2015	Grant from Federation of European Biochemical Societies (FEBS) to attend 15th Young Scientists' Forum (YSF) and 40th FEBS Congress

11/2012 – 05/2015	iGRASPseed scholarship for doctoral research, Heinrich Heine University, Duesseldorf.
07/2004 – 05/2007	Indian oil corporation award for Best M.Tech. thesis Department of chemistry, IIT Delhi, India
09/2011 – 05/2012	Deutscher Akademischer Austausch Dienst (DAAD) fellowship for Master thesis at RWTH Aachen University, Germany
12/2010	Junior Research Fellowship (JRF) Council of Scientific and Industrial Research (CSIR), Delhi, India
02/2010	Graduate Aptitude Test of Engineering (GATE) – scholarship Ministry of Human Resources Development (MHRD), Government of India

7.4 Contributions to International meetings

15th Young Scientists' Forum (YSF) and 40th Federal of European Societies (FEBS) Congress, Berlin, Germany

Poster: Design, Synthesis and biological evaluation of specific rhomboid protease inhibitors

Parul Goel[1], Thorsten Jumpertz[1], Anne Harbig[1], Sascha Weggen[1], Boris Schmidt[2]

[1]Department of Neuropathology, Heinrich-Heine-University, 40225 Duesseldorf, Germany.

[2]Clemens Schöpf-Institute of Organic Chemistry and Biochemistry, Technische Universität Darmstadt Alarich-Weiss-Strasse 4, 64287 Darmstadt, Germany.

7.5 Acknowledgements

Foremost, my heartiest gratitude to my research supervisor Prof. Dr. Sascha Weggen for giving me a very interesting subject of PhD thesis and providing me with a constant support and a very kind supervision.

I would like to thank my research supervisor Prof. Dr. Boris Schmidt for his guidance, cooperation and opportunity to work in a highly productive environment in his work group. His enthusiasm and passion towards research has inspired and motivated me at every step in my research tenure.

I want to thank my second supervisor Prof. Dr. Jörg Pietruszka for his approachability during the research project.

Many thanks to all members of Weggen lab for their support and friendly working environment. I thank especially to Dr. Thorsten Jumpertz for his continuous help, advice, discussion and encouragement, which greatly contributed to solve problems I faced in my research work at HHU.

I would like to extend my thanks to all members of Schmidt lab and especially to Dr. Christoph Scholz for his guidance and discussion during molecular docking experiments for my research work.

I would like to thank Heinrich-Heine International Graduate School of Protein Science and Technology (iGRASPseed) for providing the financial support during my PhD project.

I thank Prof. Dr. Steven Verhelst (University of Leuven, Belgium) for collaborating with us in profiling rhomboid inhibitors.

I thank Prof. Dr. Matthew Freeman (University of Oxford) and Prof. Dr. Sinisa Urban (Johns Hopkins University, Baltimore) for providing expression plasmids.

I also want to thank Prof. Dr. Kvido Stříšovský (Academy of Sciences of the Czech Republic, Prague) for providing L16 inhibitor.

Finally, I would like to thank my friends and family for their unconditional love and support during my research work and life in general. I would specially like to thank my mother Mrs. Rajbala Goel who always prayed for my success and constantly encouraged me throughout in my career. I also thank my husband Mr. Harshal Patel for motivating throughout writing this thesis and providing me with a constant sense of reliance.

7.6 Declaration

I hereby declare that the work presented here is reflection of my own independent efforts and had been conducted without any unauthorized assistance. Wherever contributions and consultation of other sources are involved, every effort is made to indicate this clearly with due reference to the literature and acknowledgement of collaborative research and discussions, if any. Moreover the dissertation has never been submitted in any form to any other institution.

Duesseldorf, August 2016 (Parul Goel)

The present work was conducted during the time period from November 2012 to August 2014, at Department of Organic Chemistry, TU Darmstadt and from September 2014 to December 2015, at the Department of Neuropathology, University Hospital, Heinrich-Heine University, Duesseldorf under the supervision of Prof. Dr. Sascha Weggen (HHU) and Prof. Dr. Boris Schmidt (TU Darmstadt).