Establishing an antibody mimetic applicable to targeting in the gastrointestinal tract



Niels Sören Wicke

Department of Biochemistry

St Catherine's College

University of Oxford

Trinity 2021

This thesis is submitted to the Department of Biochemistry, University of Oxford, in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Abstract

Proteins are a valuable class of therapeutics because proteins, like antibodies or non-antibody scaffolds, can bind their targets with high selectivity and affinity. The upper gastrointestinal tract, however, rapidly inactivates protein-based targeting reagents. Compared to injections, oral administration of protein therapies is favourable for treating gastrointestinal disease. In this thesis, I have developed a protein scaffold termed a gastrobody, that is suited to applications in the gastrointestinal tract. Gastrobodies are derived from Kunitz-type soybean trypsin inhibitor (SBTI). SBTI is resilient to digestion with stomach-like concentrations of pepsin and pH 2. Furthermore, SBTI is stable in the presence of intestinal proteases and bile acids, as well as being resilient to high temperatures. I used the protein modelling software Rosetta to assist design of a binding surface on SBTI. I also established phage display of SBTI to enable directed evolution. As proof of concept, gastrobody binders to domains of two toxins of the enteric pathogen *Clostridium difficile* were developed. After optimisation of the binding surface and affinity maturation, I identified gastrobodies that bound their target with nanomolar affinity. Gastrobodies bind their target after incubation with physiological concentrations of gastrointestinal proteases, including pepsin at pH 2. Lastly, I explored alternative gastrobody backbones using structural homologues of SBTI. The work presented in this thesis may provide a foundation for generating binding reagents for therapeutic and diagnostic use in the gastrointestinal tract.

Authorship declaration

I declare that this thesis is entirely my own work, except for cases described below. This thesis has not been submitted for any other degree at the University of Oxford or any other institution.

Experiments performed by others:

Differential scanning calorimetry (chapter 3, 4 and 5) was performed by Dr David Staunton.

Electrospray ionisation mass spectrometry (chapter 3, 4 and 5) was performed by Dr Anthony Tumber.

Publication arising from this thesis:

Wicke, N., Bedford, M.R. & Howarth, M. Gastrobodies are engineered antibody mimetics resilient to pepsin and hydrochloric acid. *Commun Biol* **4,** 960 (2021). https://doi.org/10.1038/s42003-021-02487-2.

Acknowledgements

I am grateful to my academic supervisor Professor Mark Howarth for four instructive and formative years. Mark inspired me with his goal-oriented and realistic approach to producing translatable scientific research. I would also like to thank my industrial supervisor Dr Mike Bedford for his unwavering support and his enthusiasm for the project.

For providing the financial support that enabled my study, I would like to acknowledge the Biotechnology and Biological Sciences Research Council, the Interdisciplinary Bioscience Doctoral Training Partnership and AB Vista.

I would like to thank the Sherratt, Whitby and Uphoff labs for a friendly and productive working environment. To past and present Howarth lab members, thank you for offering a supportive, positive, and wonderfully ambitious atmosphere. More than our fruitful scientific conversations, I will remember sharing with you the realities of research: both the exciting and the disappointing moments.

Thank you to my friends for providing welcome distractions from the lab. To my family, thank you for your support and all you have taught me. A special thanks to my mother Michaela. I admire your strength, resilience, and kindness. And lastly, to Clare – thank you for your love and support. I am inspired by your determination and courage. Here's to our next adventure together.

Abbreviations

BBI	Bowman-Birk inhibitor (soybean)
BBKI	*Bauhinia bauhinioides* plasma kallikrein inhibitor
BLA	β-lactamase
BSA	Bovine serum albumin
CDR	Complementarity determining region
CROP	*Clostridium difficile* toxin A combined repetitive oligopeptide domain
DSC	Differential scanning calorimetry
DTT	1,4-dithiothreitol
ECTI	*Erythrina caffra* trypsin inhibitor
ELISA	Enzyme linked immunosorbent assay
ESI-MS	Electrospray ionisation mass spectrometry
ETEC	Enterotoxigenic *Escherichia coli*
FCR	Feed conversion ratio
FDA	Food and Drug Administration
GDR	Gastrobody determining region
GI	Gastrointestinal
GST	Glutathione S-transferase
GTD	*Clostridium difficile* toxin B glucosyltransferase domain
HEL	Hen egg lysozyme
HRP	Horseradish peroxidase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth
MBP	Maltose binding protein

mEGFP	Monomeric enhanced green fluorescent protein
MWCO	Molecular weight cut-off
NGS	Next generation sequencing
NTA	Nitrilotriacetic acid
NPase	Non-starch polysaccharidase
OSF	Osmotic shock fluid
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG8000	Polyethylene glycol 8000
pIII	M13 minor coat protein 3
PMSF	Phenylmethylsulfonyl fluoride
POI	Protein of interest
POPS	Parameter optimised surfaces (webserver)
QB	Quantum Blue phytase
Q-score	Quality of alignment score
REU	Rosetta energy units
RMSD	Root mean square deviation
SA	Streptavidin
SBTI	Kunitz-type soybean trypsin inhibitor
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SHIME	Simulator of the human intestinal microbial ecosystem
SPR	Surface plasmon resonance

TcdA	*Clostridium difficile* toxin A
TcdB	*Clostridium difficile* toxin B
TEA	Triethylamine
TIM	Netherlands Organisation for Applied Scientific Research (TNO) intestinal model
TIMagc	TNO advanced gastric intestinal model
TNFα	Tumour necrosis factor alpha
WBCI	Winged bean chymotrypsin inhibitor
WT	Wild-type

Table of contents

Abstract ..... **i**
Authorship declaration ..... **ii**
Acknowledgements ..... **iii**
Abbreviations ..... **iv**
Table of contents ..... **vii**
List of figures ..... **xiii**
Chapter 1: Introduction **1**
1.1. The Gastrointestinal (GI) tract is a hostile environment for proteins 2
1.1.1. Proteases in the GI tract 2
1.1.2. pH profile of the GI tract 4
1.1.3. Bile acids in the small intestine 5
1.1.4. In vitro models of the GI tract 6
1.2. Phage display 7
1.2.1. Principles of phage display 7
1.2.2. Phage display in GI tract-like conditions 8
1.3. Protein targeting reagents 9
1.3.1. Nanobodies 10
1.3.2. Nanofitins 12
1.4. Soybean trypsin inhibitor 14
1.5. What makes a protein protease-resistant? 15
1.6. Advantages of orally administered protein therapeutics 18
1.7. Enzymes in the animal feed industry 20
1.8. Thesis aims 21
Chapter 2: Methods **23**
2.1. Cloning 24
2.1.1. PCR 24
2.1.2. Agarose gel electrophoresis 24
2.1.3. PCR product purification 25
2.1.4. Gibson assembly 25
2.1.5. Heat-shock transformation 26
2.1.6. Plasmid purification 26
2.1.7. Library cloning 26
2.1.8. Phage display 27

2.2. Protein expression and purification 28
2.2.1. Media 28
2.2.2. Expression constructs and cell strains 28
2.2.3. Protein expression conditions 30
2.2.4. Whole cell lysis 30
2.2.5. Periplasmic extraction 31
2.2.6. Ni-NTA purification 32
2.2.7. Size-exclusion chromatography 33
2.2.8. Other proteins 33
2.3. SDS-PAGE 34
2.4. Quantification of protein concentration 35
2.5. Site-specific biotinylation 35
2.6. Mass spectrometry 35
2.6.1. Intact protein ESI-MS 36
2.6.2. Mapping protease cut sites 36
2.7. Rosetta 37
2.7.1. Modelling missing density in SBTI crystal structure 39
2.7.2. Generating an ensemble of SBTI structures 39
2.7.3. In silico mutagenesis of solvent accessible residues 40
2.7.4. Rosetta data processing and visualisation 41
2.8. Gastric fluid proteolysis assay 41
2.9. Protease resistance assays 42
2.9.1. SDS-PAGE pepsin digestion assay 42
2.9.2. SDS-PAGE pancreatin or elastase digestion assay 43
2.9.3. Mouse gastric fluid digestion assay 43
2.9.4. ELISA on binding GTD after protease digestion 43
2.10. Temperature resistance assays 45
2.10.1. Temperature-dependent solubility assay 45
2.10.2. Heat-resilience ELISA 45
2.10.3. Differential scanning calorimetry (DSC) 46
2.11. Surface plasmon resonance (SPR) 46
2.11.1. Multi-cycle SPR 47
2.11.2. Single-cycle SPR 48
2.12. Plate assays on GT01 R63A binding 48

2.13. ELISA on binding specificity 49
2.14. Phytase assay 50
2.15. Phage display 50
2.15.1. Phage production 50
2.15.2. Phage purification 51
2.15.3. Gastrobody selection against CROP 52
2.15.4. Initial gastrobody selection against GTD 54
2.15.5. Affinity maturation selection of GT_S1_01 55
2.15.6. Affinity maturation selection of GT01 56
2.16. Phage retention on trypsin-agarose beads 56
2.17. SBTI retention on trypsin-agarose beads 57
2.18. GTD inhibition assay 57
2.19. Trypsin inhibition assay 58
2.20. Western blot 59
2.20.1. Arabinose titration of pIII expression 59
2.20.2. SBTI display on full-length pIII 59
2.21. PDBeFold 60
2.22. Structure representation 61
Chapter 3: Exploring scaffolds suitable for the GI tract 62
3.1. Proteolytic activity of mouse and chicken gastric fluid 64
3.2. Gastrointestinal stability of scaffold proteins and SBTI 65
3.2.1. MS to show correct expression of SBTI, nanobody and nanofitin 66
3.2.2. Pepsin stability of nanofitin, nanobody and SBTI 67
3.2.3. Pancreatin stability of nanofitin, nanobody and SBTI 68
3.2.4. Elastase stability of nanofitin, nanobody and SBTI 69
3.2.5. Protease cleavage sites in SBTI 70
3.2.6. SBTI inhibits trypsin in the presence of bile acids 71
3.3. SBTI is a thermo-stable protein 72
3.3.1. Temperature dependent solubility of SBTI 73
3.3.2. Binding activity of SBTI after heating 73
3.3.3. Thermal stability of SBTI at low pH 73
3.3.4. SBTI is a compatible fusion partner to phytase 74
3.3.5. MS shows SBTI-QB can be expressed in E. coli 75
3.3.6. SBTI fusion to QB does not compromise QB phytase activity 76

3.3.7. SBTI-QB is retained on trypsin-agarose resin 77
3.4. Discussion 77
Chapter 4: Engineering SBTI to become a gastrobody 81
4.1. Rosetta to predict evolvability 84
4.1.1. Generating an ensemble of structures with relax 84
4.1.2. pmutscan of surface-accessible residues 85
4.1.3. Choosing a versatile binding surface 89
4.2. Experimental validation of Rosetta predictions 90
4.2.1. MS confirms expression of alanine mutants 90
4.2.2. Temperature dependent solubility of alanine mutants 90
4.2.3. Pepsin resistance of alanine mutants 91
4.3. Establishing phage display of SBTI 93
4.3.1. Optimising phagemid construct and conditions to ensure genetic stability 93
4.3.2. Retention of SBTI-M13 on trypsin-agarose beads 93
4.3.3. Western blot verifies display of SBTI on M13 94
4.4. Selecting an SBTI binder to CROP 95
4.4.1. Expression, purification and biotinylation of CROP 95
4.4.2. Cloning a library of SBTI for phage display 97
4.4.3. Panning to select an SBTI binder to CROP 97
4.4.4. ELISA to screen individual clones for CROP binding 98
4.4.5. Sequences of anti-CROP SBTI clones 99
4.5. Characterisation of anti-CDTA gastrobodies 99
4.5.1. Expression and MS validation of anti-CROP hits 99
4.5.2. Binding affinity of anti-CROP gastrobodies 101
4.5.3. ELISA to test specificity of anti-CROP gastrobodies 103
4.5.4. Pepsin resistance anti-CROP gastrobodies 104
4.5.5. Protease cleavage sites in T12 105
4.5.6. Thermal stability of anti-CROP gastrobodies 105
4.6. Discussion 106
Chapter 5: Improving gastrobody affinity 112
5.1. Selecting a gastrobody binder to GTD 115
5.1.1. Expression and purification of GTD 115
5.1.2. Cloning a library of SBTI for display on full length pIII with extended loops 116

5.1.3. Panning to select a binder to GTD 116
5.1.4. Affinity maturation of anti-GTD gastrobody 119
5.2. Characterisation of anti-GTD gastrobodies 120
5.2.1. Expression and MS validation of anti-GTD hits 120
5.2.2. Binding kinetics of anti-GTD gastrobodies by SPR 121
5.2.3. Effect of GT01 on GTD hydrolytic activity 123
5.2.4. Thermal stability of anti-GTD gastrobodies 125
5.2.5. Pepsin-resistance of anti-GTD gastrobodies 126
5.2.6. Stability of GT01 in mouse gastric fluid 128
5.2.7. Protease cleavage sites in GT01 128
5.3. Expanding the gastrobody binding surface 129
5.3.1. MS spectra confirm expression of GDR3/4 alanine mutants 130
5.3.2. Pepsin resistance of GDR3/4 alanine mutants 130
5.3.3. Phage display for affinity maturation of GT01 with randomised GDR3/4 131
5.3.4. Binding kinetics of GT44 and GT47 binding to GTD 133
5.3.5. Heat-resilience of GT44 and GT47 133
5.3.6. Pepsin-resistance of GT44 and GT47 134
5.4. Removing trypsin binding of GT01 135
5.4.1. Trypsin binding of GT01 R63A 136
5.4.2. GT01 R63A binding kinetics to GTD 136
5.4.3. GT01 R63A binding activity after protease digestion 137
5.4.4. Protease cleavage sites in GT01 R63A 137
5.5. Alternative gastrobody backbones 138
5.5.1. Expression, purification and MS of BBKI, ECTI and WBCI 139
5.5.2. Pepsin-resistance of structural homologs 140
5.5.3. Temperature-dependent solubility of structural homologues 141
5.5.4. Thermal stability of structural homologues at low pH 142
5.6. Discussion 143
Chapter 6: Summary and general discussion 150
Bibliography 159
Appendix 181

List of figures

Figure 1.1. Schematic of protease site nomenclature. 3
Figure 1.2. Protein scaffold structures. 13
Figure 1.3. Schematic of proteolysis dynamics. 16
Figure 2.1. Workflow for in silico analysis of SBTI mutability. 38
Figure 3.1. Proteolytic activity in gastric fluid and protein scaffolds for oral delivery. 66
Figure 3.2. Stability of candidate scaffolds in gastric conditions. 68
Figure 3.3. Protease cut-sites in SBTI. 70
Figure 3.4. SBTI stability in the presence of bile acids. 72
Figure 3.5. Temperature stability of SBTI. 74
Figure 3.6. SBTI is a compatible fusion partner with phytase. 76
Figure 4.1. Rosetta-guided prediction of evolvable regions in SBTI. 85
Figure 4.2. Designing a binding surface on SBTI. 89
Figure 4.3. Experimental alanine scan of GDR residues. 92
Figure 4.4. Establishing phage display of SBTI. 94
Figure 4.5. Expression, purification and biotinylation of CROP. 96
Figure 4.6. Diagram of gastrobody selection strategy against CROP. 98
Figure 4.7. Selecting a gastrobody binder against CROP. 100
Figure 4.8. Expression and purification of anti-CROP gastrobodies. 101
Figure 4.9. Binding characterisation of anti-CROP gastrobodies. 102
Figure 4.10. Pepsin resistance of anti-CROP gastrobodies. 103
Figure 4.11. Protease cut sites in T12 gastrobody. 104
Figure 4.12. Temperature stability of anti-CROP gastrobodies. 106
Figure 5.1. Expression and biotinylation of GTD. 115
Figure 5.2. Diagram of gastrobody selection strategy against GTD. 117
Figure 5.3. Initial selections for a gastrobody binder to GTD. 118
Figure 5.4. Clones from affinity maturation against GTD. 121
Figure 5.5. Analysis of anti-GTD gastrobody binding to GTD. 123
Figure 5.6. Effect of GT01 on GTD hydrolytic activity 124
Figure 5.7. Temperature stability of GT01. 126
Figure 5.8. Stability of GT01 to stomach-like conditions 127
Figure 5.9. Protease cut sites in GT01 gastrobody. 129
Figure 5.10. Extending the gastrobody binding surface. 131

Figure 5.11. Exploring mutability of GDR3/4. 132
Figure 5.12. Pepsin- and heat-stability of GT44 and GT47. 134
Figure 5.13. Removing trypsin-binding of GT01. 135
Figure 5.14. Protease-stability of GT01 R63A. 136
Figure 5.15. Protease cut sites in GT01 R63A. 138
Figure 5.16. Expression and purification of structural homologues of SBTI. 140
Figure 5.17. Pepsin- and heat-resilience of β-trefoil proteins. 141
Figure 5.18. Stability of β-trefoil proteins at low pH. 142
Table 1. Summary of the structural homologues of SBTI. 143

Chapter 1: Introduction

Chapter 1: Introduction

1.1. The Gastrointestinal (GI) tract is a hostile environment for proteins

The human GI tract is a series of hollow organs specialized in the breakdown of ingested foods. Up to 9 L of fluid including digestive enzymes, acid and bile acids are secreted by the GI tract per day to facilitate digestion and absorption of ingested material. (Greenwood-Van Meerveld et al., 2017). Nutrients released from digested foods are absorbed from the lumen of the GI tract through the mucosal epithelial surface. Mucus protects the absorptive and secretory epithelial cell layers. An attached inner mucus layer and unattached less dense outer mucus layer protect the stomach and colon, while a single unattached mucus layer is found in the small intestine (Johansson et al., 2013).

This section of the introduction will provide an overview of the challenges for proteins in the GI tract, with particular focus on the stomach.

1.1.1. Proteases in the GI tract

The stomach is the first major organ in the GI tract that initiates digestion and protects against foodborne pathogens (Hunt et al., 2015). In the stomach, ingested material mixes with hydrochloric acid secreted by parietal cells and pepsinogens secreted by chief cells (Greenwood-Van Meerveld et al., 2017; Hunt et al., 2015).

There are five families of pepsinogens: prochymosin, progastricsin and pepsinogen A, B and F (Kageyama, 2014). Pepsinogen A and progastricsin are expressed in adult vertebrates, while prochymosin and pepsinogen F are specific to infants (Kageyama, 2014; Kageyama et al., 1990). Pepsinogen B is specific to pigs (Ryle, 1970).

Acidic pH triggers the autocatalytic activation of pepsinogen to form pepsin (Kageyama, 2014). Pepsins are aspartic endopeptidases with broad cleavage specificity (Hamuro et al., 2008; Kageyama, 2014). Residues in the P1 or P1' positions influence pepsin cleavage probability the most (Fig. 1.1). Hydrophobic amino acids in P1 or P1' generally favour cleavage, but cleavage occurs with a wide range of amino acids at P1 or P1'. The highest pepsin cleavage probability was observed when Phe, Leu or Met are at P1 and Phe, Trp and Tyr are at P1'. Positively charged amino acids like His, Lys or Arg at P1 disfavour pepsin cleavage (Hamuro et al., 2008).

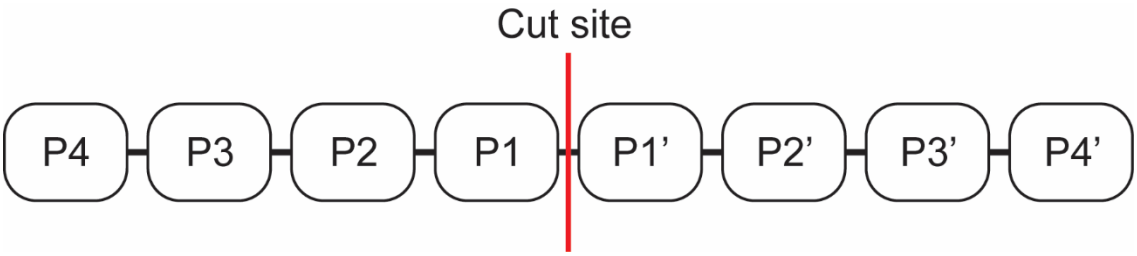


Figure 1.1. Schematic of protease site nomenclature. Proteases cut C-terminal of P1 and N-terminal of P1' amino acids. The nomenclature was introduced by Schechter and Berger (Schechter and Berger, 1967).

In the human stomach the concentration of pepsin is 0.5 – 1 mg/mL, where 1 mg has ≥ 2,500 U pepsin activity (Brodkorb et al., 2019; Zhu et al., 2006). Pepsin activity is defined as one unit producing "ΔA280 of 0.001 per min at pH 2.0 at 37 °C, measured as trichloroacetic acid-soluble products using hemoglobin substrate" (Pharmacopeia, 2015; Ryle, 1970). The pH optimum for pepsin is pH 1.5 – 2.5, with ~ 70% activity at pH 4.5 and almost no activity at pH 5.5 (Piper and Fenton, 1965). Pepsin is irreversibly denatured above pH 8 (Piper and Fenton, 1965). Pepsin can be inhibited by pepstatin (Marciniszyn et al., 1977) and pepsin inhibitor-3 from the intestinal parasite *Ascaris suum* (Ng et al., 2000).

Ingested material mixed with gastric secretions exiting the stomach is called chyme. In the small intestine, secretions from exocrine cells of the pancreas, termed pancreatin, are mixed with chyme (Lee and Yamamoto, 1989; Mahato et al., 2003). Pancreatin includes the peptidases trypsin, chymotrypsin, elastase, carboxypeptidase A and carboxypeptidase B (Lee and Yamamoto, 1989; Mahato et al., 2003). Trypsin is expressed as a zymogen (trypsinogen) and is the most abundant protease in pancreatic juice (Rinderknecht et al., 1979). There are three trypsin isoforms in humans, all of which cut C-terminal of Arg or Lys with a slight preference for Ala or Gly in the P1' position (Schilling et al., 2018). Chymotrypsin is the second most abundant protease in pancreatin (Carrère et al., 1986). Chymotrypsin cuts C-terminal of Tyr, Phe and Trp (Appel, 1986).

1.1.2. pH profile of the GI tract

Ingested material is rapidly acidified in the stomach. The pH in various segments of the human GI tract is highly variable between individuals and depends on whether food has been consumed (Koziolek et al., 2015). Early studies with healthy young men and women reported median fasted gastric pH to be 1.7 (Dressman et al., 1990; Fallingborg, 1999). More recent studies with the wireless drug delivery and pH-sensing capsule IntelliCap® system report mean fasted gastric pH to be 2.7 ± 0.8, though variation between 1.4 and 4.6 were observed (Koziolek et al., 2015). The liquid volume in the lumen of the stomach is 10 – 50 mL, and thus entry of food into the stomach will transiently buffer gastric pH. For example, a meal can result in a temporary median peak pH 6.7 in the stomach, before drifting back to pH 1.7 (Dressman et al., 1990). In the study by Dressman *et al*. gastric pH returned to pH 4 in 28 ± 24 min and to pH 2

in 107 ± 70 min (error is 1 standard deviation) (Dressman et al., 1990). Ingestion of water has little effect on pH due to the low buffering capacity and rapid gastric emptying (Koziolek et al., 2015).

The time to gastric emptying is strongly dependent on the ingested meals (Hansrod et al., 2020). Depending on the individual and the meal, the half-emptying time can vary from 18 to 166 min (Hansrod et al., 2020).

The pH rises rapidly after exiting the stomach to pH 6 in the duodenum and continues to increase slowly to pH 7.4 in the terminal ileum (Fallingborg, 1999; Koziolek et al., 2015; Riethorst et al., 2016). The rapid rise in pH is due to pancreatic and duodenal bicarbonate secretion. In the presence of food, the duodenal pH is 7.5 (Riethorst et al., 2016). A decrease to pH 6 is observed in the colon as a result of fermentation of the colonic microbiota, but the pH rises to 7 again in the rectum (Fallingborg, 1999; Koziolek et al., 2015).

1.1.3. Bile acids in the small intestine

Bile acids play a key role in lipid (Hu et al., 2018) and protein digestion (Calvert and Legg, 1977; Gass et al., 2007; Robic et al., 2011; Sarkany and Moreland, 1985) in the small intestine. Bile acids are synthesized in the liver and stored in the gallbladder until secretion into the small intestine (Dawson and Karpen, 2015). In the fasted state, the concentration of free bile acids in the small intestine is 4 mM, rising to 10 – 15 mM following a meal (Minnekus et al., 1995; Northfield and McColl, 1973; Riethorst et al., 2016). Bile acids can denature proteins, thereby facilitating intestinal proteolysis (Cremers et al., 2014). Trypsin and chymotrypsin activity is potentiated by the

presence of bile acids (Calvert and Legg, 1977; Gass et al., 2007; Robic et al., 2011; Sarkany and Moreland, 1985).

1.1.4. In vitro models of the GI tract

Numerous *in vitro* models of digestion have been developed, but it is difficult to capture all aspects of digestion in one model (Dupont et al., 2018). Basic components to be modelled include the activity of digestive enzymes, pH, temperature, and bile salts. However, digestion is a dynamic process whereby the conditions are constantly changing; ingested particles move through the GI tract compartments, mix with various digestive juices, diffuse through mucus and breakdown products are absorbed. Mechanical digestion is a further layer of complexity. *In vitro* models to study the fate of ingested materials can broadly be divided into mono- and multi-compartmental models (Dupont et al., 2018).

The most basic approach to modelling digestion *in vitro* uses static conditions to represent the different segments of the GI tract. For example, the stomach is modelled at a fixed pH with calibrated activity levels of pepsin and gastric lipase (Brodkorb et al., 2019; Mulet-Cabero et al., 2020). Addition of stock solutions can convert the stomach phase into the intestinal phase (semi-dynamic) or intestinal digestion can be modelled in isolation (static). This approach is unified in the INFOGEST 2.0 protocol and benefits from ease of set up and analysis (Brodkorb et al., 2019).

One example of a multi-compartment model for digestion is the TIM model developed at the Netherlands Organisation for Applied Scientific Research. The first model (TIM-1) was developed to model dynamic digestion of the stomach and small intestine (Minekus et al., 1999); while TIM-2 models the large intestine (Minekus et al., 1999).

TIM systems consist of a series of compartments each modelling a GI tract segment, connected by peristaltic valves and fitted with semi-permeable membranes. Parameters like peristaltic mixing, temperature, transit times, pH and secretion of digestive fluids are controlled by a computer program. The latest advanced gastric TIM model (TIMagc) enables modelling of gastric motility and pressure forces (Bellmann et al., 2016). TIM models are designed and validated against *in vivo* data. TIM models appropriately model the dynamics of digestion, but a drawback is that the models do not include mucus (Dupont et al., 2018). Another multi-compartment model worth mentioning is the Simulator of the Human Intestinal Microbial Ecosystem (SHIME) (Molly et al., 1993; Van de Wiele et al., 2015). As the name suggests, SHIME was developed for modelling the impact of ingested foods on the composition of the microbiota.

1.2. Phage display

1.2.1. Principles of phage display

First developed in 1985, phage display is a powerful tool for protein engineering and directed evolution (Clackson and Lowman, 2004; Shim, 2017; Smith, 1985). The utility of phage display for engineering proteins, and in particular antibodies, was recognised with the Nobel prize in 2018 (Smith, 2019; Winter, 2019). Phage display relies on the physical link between the phenotype and genotype of viruses: the protein of interest (POI) is fused to a coat protein of a virus and the genetic information encoding the POI is packaged within the virus (Clackson and Lowman, 2004). Phage that are displaying variants of the POI (termed a library) are usually selected for novel or improved properties through protein-protein interactions. Libraries of up to 10^{10} variants of a POI

can be included from in a single round of panning, thus enabling efficient high-throughput selection (Clackson and Lowman, 2004; Shim, 2017). Iterative rounds of selection are performed to enrich for clones with the desired properties. Lysogenic filamentous phage (e.g. M13) are the most popular phage display virus. Each M13 particle contains five copies of minor coat protein 3 (pIII). POI are commonly fused to pIII because pIII-fusion allows efficient monovalent display of the POI (Clackson and Lowman, 2004).

Selection of strong non-covalent binding of antibodies or protein scaffolds to novel targets has been a key application of phage display (Frenzel et al., 2016; Shim, 2017; Vazquez-Lombardi et al., 2015). Phage display has accelerated the discovery and engineering of fully human antibodies by replacing time-consuming hybridoma technology. By 2016, six antibodies developed at least in part by phage display were approved for human therapy, and many more are in clinical development (Frenzel et al., 2016).

1.2.2. Phage display in GI tract-like conditions

The selection conditions during a phage display campaign influence the properties of the resulting proteins. For example, Kong *et al.* incubated a phage library displaying double-bridged peptides with pancreatin for 30 min at 37 °C, before selecting for binding to interleukin-23 or coagulation Factor XIa (Kong et al., 2020). Peptide hits from selection with pancreatin were more resistant to digestion by pancreatin than clones identified from campaigns where proteolytic pressure was not applied (Kong et al., 2020).

Selection for resistance to stomach-like conditions is difficult, because commonly used phage vectors like T7 are deactivated by acid (Nobrega et al., 2016; Teesalu et al., 2009), and M13 is deactivated by pepsin (Fitza, 2002; Tóthová et al., 2012). Inactivation of the phage prevents the iterative rounds of selection usually required for identifying the best binders from a large library. Nevertheless, there are reports of *in vivo* phage display where peptide-displaying M13 phage libraries were administered orally to identify peptides that drive transcytosis across the intestinal barrier in mice (Duerr et al., 2004; Kang et al., 2008). Higgins *et al.* administered a phage library into the GI tract by intestinal loop incision to identify peptides binding to M cells or Peyer's patch tissue (Higgins et al., 2004).

1.3. Protein targeting reagents

Antibodies are a valuable class of molecules, with nearly 100 antibodies approved for therapy (Kaplon and Reichert, 2021). However, some properties of antibodies also restrict their applicability (Vazquez-Lombardi et al., 2015). Antibodies are usually ~ 150 kDa, consisting of multiple domains and chains with important disulfide bonds and post-translational glycosylation (Rouet et al., 2014). The combination of these parameters complicates production of antibodies and precludes facile expression in bacterial systems (Birch and Racher, 2006; Rouet et al., 2012). In addition, the large size limits the tissue penetration of antibodies which can necessitate large doses (Chames et al., 2009). Antibodies are rapidly inactivated in the mammalian GI tract (Anselmo et al., 2018; Yadav et al., 2016). To overcome size, production and stability restrictions of antibodies a larger number of small protein scaffolds have been developed (Škrlec et al., 2015; Vazquez-Lombardi et al., 2015).

1.3.1. Nanobodies

Camelidae have single variable domain, heavy chain-only antibodies (Hamers-Casterman et al., 1993; Hamers et al., 1993). The variable domain of camelid antibodies can be expressed solubly as an antibody mimetic called a nanobody. The proof-of-principle for nanobodies was reported in 1997 (Fig. 1.2A) (Arbabi Ghahroudi et al., 1997). Arbabi Ghahroudi *et al.* immunised camels with a tetanus toxoid or lysozyme and cloned the heavy-chain antibody variable domain repertoire into *E. coli.* Phage display was used to further select for nanobodies binding to lysozyme or the tetanus toxoid from the variable domain repertoire. Nanobodies with low nanomolar binding affinities to their targets were identified (Arbabi Ghahroudi et al., 1997).

Since the seminal work in the 1990s, nanobodies have become an important class of therapeutics in their own right (Steeland et al., 2016). In 2019, caplacizumab for treatment of acquired thrombotic thrombocytopenic purpura was the first nanobody to be approved by the U.S. FDA (Morrison, 2019). Nanobodies are small (15 kDa) and express well in microbial systems (Fig. 1.2A). Like antibodies, three hypervariable complementary determining regions (CDRs) participate in binding target proteins, with CDR3 usually contributing the majority of the antigen-contacting surface (de Genst et al., 2006). The convex paratope of nanobodies is well suited to binding pockets inaccessible to larger antibodies (Steeland et al., 2016). Engineering efforts and deliberate selection strategies have produced nanobodies with notable stability profiles: nanobodies resilient to high temperatures (Ewert et al., 2002; Kunz et al., 2018; van der Linden et al., 1999), to low pH and to proteases (Crowe et al., 2018; Hussack et al., 2011a, 2014; Virdi et al., 2013) have been described.

Indeed, the stability profile of nanobodies has led to research efforts toward oral administration of nanobodies. Orally administered nanobodies that bind to key effector proteins of enteric pathogens may protect against disease. For example, binders to *Campylobacter jejuni* flagella (Hussack et al., 2014; Riazi et al., 2013) or *Escherichia coli* F4 fimbriae (Harmsen et al., 2005, 2006; Virdi et al., 2013) have been developed. Anti-F4 fimbriae nanobodies, however, failed to protect pigs from enterotoxigenic *E. coli* (ETEC) induced diarrhoea (Harmsen et al., 2005), citing pepsin degradation as a challenge for orally administered nanobodies (Harmsen et al., 2006). Multimerization of nanobodies in an IgA-like format may confer additional stability in the GI tract. Fusion of anti-F4 fimbriae nanobodies to the constant (Fc) stalk of porcine IgA creates a pentameric secretory IgA-like (Virdi et al., 2013). Inclusion of anti-F4 nanobody IgA in feed protects against and reduces the burden of ETEC infection (Virdi et al., 2013, 2019; Wallace et al., 2020). Similarly, nanobodies binding to *C. jejuni* flagella pentamerised with verotoxin B homopentamer reduced *C. jejuni* colonisation of chickens (Riazi et al., 2013).

Other orally administered nanobodies target human proteins. VHsquared has progressed to clinical trials oral anti-tumour necrosis factor alpha (TNFα) nanobody variants (termed Vorabodies™) for the treatment of ulcerative colitis (Crowe et al., 2018, 2019). Crowe *et al.* show survival of Vorabodies in the small intestine, but a gastro-protective vehicle that includes milk and sodium bicarbonate (Crowe et al., 2018) or an enteric coating (Crowe et al., 2019) was used, suggesting that Vorabodies are not stable in the stomach.

Another immunoglobulin worth noting is IgY, the functional equivalent of human IgG in birds, amphibians and reptiles (Warr et al., 1995). IgY antibodies accumulate in

chicken egg yolk following immunization of a hen with an antigen of interest (Rahman et al., 2013). Oral administration of IgY has been trialled as therapy for treatment of gastrointestinal pathogens, including *E. coli*, *Helicobacter pylori,* and *Pseudomonas aeruginosa* (Horie et al., 2004; Nilsson et al., 2008; Rahman et al., 2013) Indeed, IgY is relatively resistant to digestion by chymotrypsin and trypsin (Hatta et al., 1993). Although the egg yolk components may confer some protection, IgY is rapidly inactivated by pepsin which poses a barrier for oral administration of therapeutic IgY (Hatta et al., 1993).

1.3.2. Nanofitins

Another class of protein scaffolds with good biophysical stability are nanofitins, also known as small artificial binding proteins (affitins) (Fig. 1.2B) (Pecorari and Alzari, 2006). Nanofitins are derived from the DNA-binding Sul7d protein family found in the extremophilic archaeal genera *Sulfolobus, Acidianus* and *Metallosphaerea* (Kalichuk et al., 2016, 2018). Sul7d proteins have a single SH3-like domain comprised of five or six β-strands arranged as antiparallel sheets with a C-terminal α-helix (Fig. 1.2B). The first nanofitin was a variant of the Sac7d protein, evolved by ribosome display to bind to *Klebsiella oxytoca* virulence factor PulD (Mouratou et al., 2007). Fourteen residues of Sac7d known to bind DNA were randomised to engineer binders with sub-nanomolar binding affinity (Mouratou et al., 2007). By solving the crystal structure of the nanofitin, the authors showed that the native fold is preserved despite the mutations.

Subsequent nanofitin research has produced binders derived from other Sul7d proteins, like the more thermostable Sso7d (Béhar et al., 2014; Gera et al., 2011) or

the smaller Aho7c (Kalichuk et al., 2018) backbones. Furthermore, the paratope has been modified to suit various epitopes. For example, Correa *et al.* showed that extending a binding loop was critical for identifying nanofitins binding to *Clostridium thermocellum* glycosidase CelD with a deep binding cavity (Correa et al., 2014). Nanofitins are usually evolved by ribosome display, but evolution by yeast display (Gera et al., 2011) and phage display (Zhao et al., 2016, 2017) has been reported.

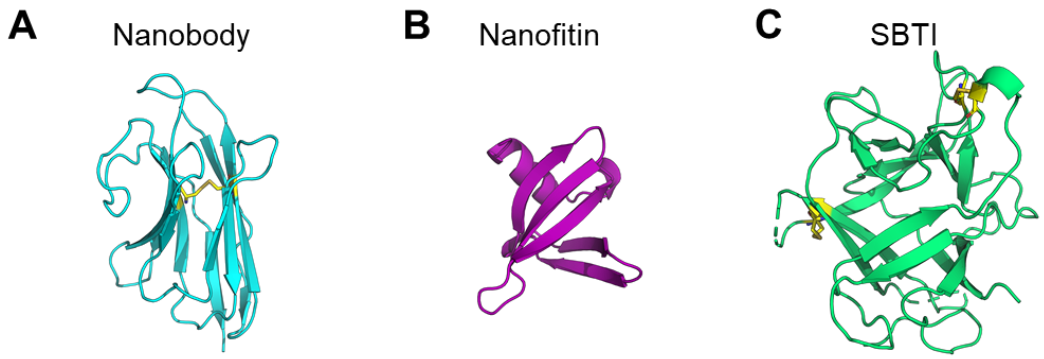


Figure 1.2. Protein scaffold structures. PyMOL structures of **(A)** nanobody (PDB ID: 1QD0), **(B)** nanofitin (PDB ID: 1BNZ), and **(C)** soybean trypsin inhibitor (SBTI, PDB ID: 1AVU). The nanofitin shown here binds to DNA. The nanobody shown here binds to the hapten azo-dye Reactive Red. Disulfide bonds are represented in stick format and coloured yellow.

Nanofitins are resilient to high temperatures and a wide range of pH. Melting temperatures (T_m) as high as 100.2 °C have been reported (Béhar et al., 2014; Kalichuk et al., 2016). Circular dichroism experiments have shown that nanofitins remain folded between pH 0.33 and pH 12.5 (Béhar et al., 2014). As a result, there has been some interest in exploring the oral administration of nanofitins. Affilogic, a company commercialising nanofitins, has presented an anti-TNFα nanofitin for the treatment of inflammatory bowel disease (Labouebe et al., 2016). The anti-TNFα nanofitin was stable in simulated gastric fluid (pH 1.6, 0.1 mg/mL pepsin), intestinal fluid (pH 6.5, 0.1 mg/mL pancreatin) and colon fluid (pH 5.8) for 60 min as assessed

by size exclusion chromatography (Labouebe et al., 2016). To my knowledge, no further data on the anti-TNFα nanofitin has been released since the first report in 2016. Recent reports suggest that nanofitins are not stable in the presence of pepsin, despite some engineering toward pepsin-resistance (Loussouarn et al., 2020).

1.4. Soybean trypsin inhibitor

Plants accumulate protease inhibitors in tissues as a mechanism of defence against ingestion by insects and large herbivores (Murdock and Shade, 2002). For example, Kunitz-type soybean trypsin inhibitor (SBTI) is a trypsin inhibitor found in *Glycine max* (soybeans) (Kim et al., 1985). In food digestibility assessments, SBTI was shown to be resistant to digestion by pepsin and stable in the low pH in the stomach (Foster et al., 2013; Roychaudhuri et al., 2004; Sung et al., 2014; Wang et al., 2017).

SBTI has a compact, roughly spherical fold called a β-trefoil (Fig. 1.2C) (McLachlan, 1979; Song and Suh, 1998; Sweet et al., 1974). The β-trefoil fold is composed of twelve anti-parallel β-strands connected by long loops (Guerra et al., 2016; Murzin et al., 1992). Six of the β-strands form a short anti-parallel β-barrel, while the other six β-strands form a lid. SBTI has 181 amino acids (20 kDa) and two disulfide bonds (Fig. 1.2C) (McLachlan, 1979; Song and Suh, 1998; Sweet et al., 1974). Soybeans also contain another, smaller protease inhibitor: the 71 amino acid Bowman-Birk inhibitor (BBI) (Birk, 1985). BBI can inhibit trypsin and chymotrypsin independently or simultaneously. Other plant protease inhibitors, like the trypsin inhibitor from *Erythrina caffra* seeds (Onesti et al., 1991), adopt a similar β-trefoil to SBTI but share low sequence identity with SBTI (Guerra et al., 2016).

SBTI belongs to the family of substrate-like competitive inhibitors. Trypsin-binding of SBTI is mediated by twelve residues, five of which form a reactive site loop that binds the active site of trypsin (Song and Suh, 1998). The reactive site loop is stabilized in a substrate-like conformation through a network of hydrogen bonds. The scissile R63 of SBTI makes five hydrogen bond contacts with trypsin (Song and Suh, 1998). The trypsin inhibition activity of SBTI has been used to protect orally administered protein therapeutics, such as insulin (Fay et al., 2020; Kidron, 2007; Yamamoto et al., 1994).

1.5. What makes a protein protease-resistant?

Limited proteolysis has been used to probe protein conformations (Fontana et al., 2004), because cleavage susceptibility by non-specific proteases is largely determined by protein dynamics (Hubbard, 1998). For proteolysis to occur, a polypeptide chain must bind to the protease's active site and conform to the specific stereochemistry of the protease. A protein in its native, rigid fold will act as a poor substrate because it may not fit into the active site (Fontana et al., 2004). However, the native conformation of a protein is in dynamic equilibrium with locally or globally unfolded structures (Fig. 1.3) (Bai et al., 1995; Chamberlain et al., 1996; Chang and Park, 2009). Local or global unfolding events then present good substrates for proteolysis. Indeed, regions with higher conformational flexibility, as determined by hydrogen-deuterium exchange experiments, correlate with proteinase K (Wang and Kallenbach, 1998) or thermolysin cleavage sites (Park and Marqusee, 2004).

The degree of unfolding required for cleavage is dependent on the protease. For example, thermolysin digestion of maltose binding protein (MBP) occurs at the same rate as the unfolding rate by urea denaturation. However, trypsin digestion of MBP

occurs at a faster rate than by urea denaturation, suggesting that trypsin only requires local unfolding while thermolysin requires global unfolding (Park et al., 2007).

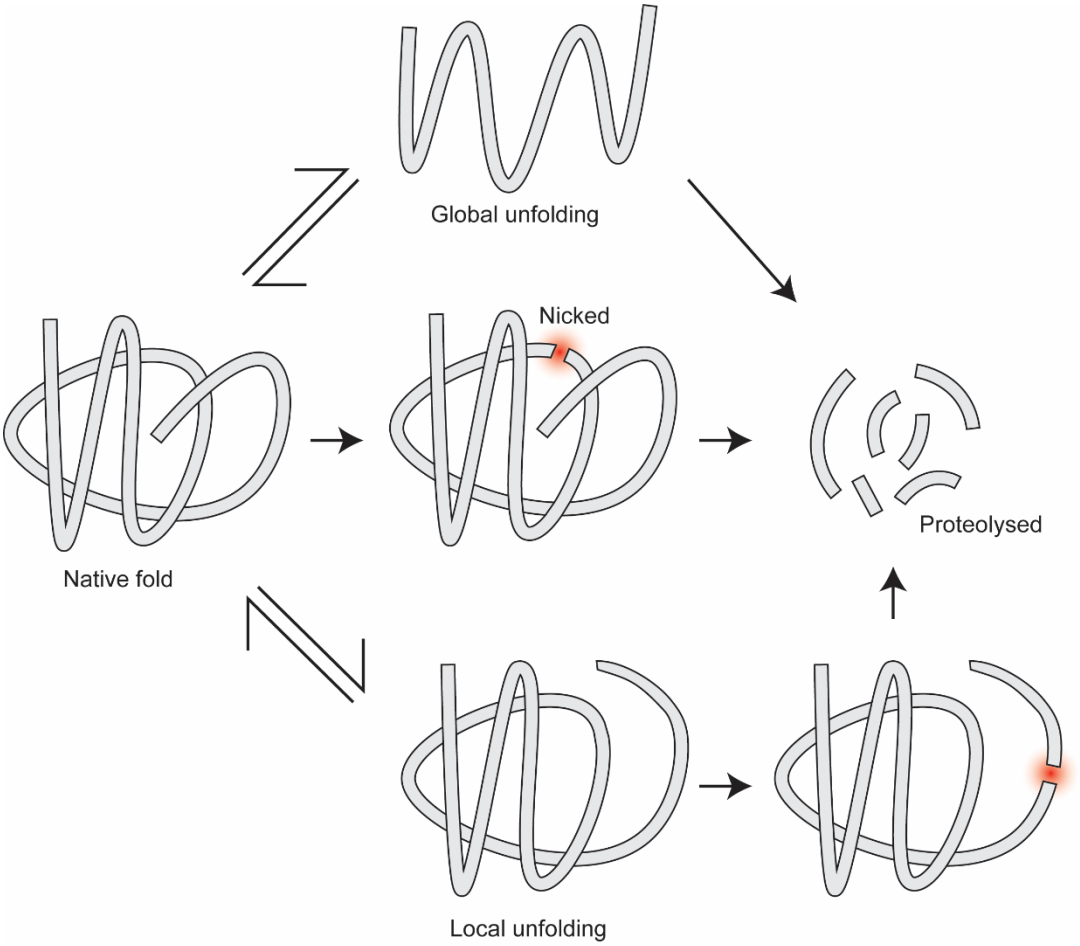


Figure 1.3. Schematic of proteolysis dynamics. Local or global unfolding of proteins exposes sites with higher probability of cleavage than sites in the native fold. External factors, like temperature, pH or bile acids, may contribute to unfolding events.

Loops and termini on the protein surface are more likely to be accessible to a protease and tend to be flexible (Fontana et al., 2004; Hubbard, 1998). The importance of loops in protease susceptibility was highlighted in a study screening a site-saturation mutagenesis library of loop residues at 86 positions in *Bacillus subtilis* lipase (Ahmad et al., 2012). The authors reported a positive correlation between thermostability (apparent melting temperature) and resistance to digestion by the non-specific protease subtilisin A (Ahmad et al., 2012). The importance of the global stability of the

protein fold was also highlighted by Daniel *et al.,* who showed that the chymotrypsin, papain or trypsin susceptibility of L-asparaginase and β-galactosidase correlated with thermal stability (Daniel et al., 1982). Reduction of conformational flexibility thus benefits thermal stability and protease-resistance (Fontana et al., 2004; Hubbard, 1998). Nevertheless, conformational flexibility, thermal stability and protease-resistance are not always correlated (LeMaster et al., 2005).

Protein digestion follows a cooperative “all-or-nothing” mechanism (Chang and Park, 2009; Linderstrøm-Lang, 1950). Cleaved protein fragments are likely to lack tertiary structure and therefore have a high probability of subsequent proteolysis events. Hence, it is usually difficult to detect cleavage fragments for determining the initial cut site, because the subsequent cleavage events occur rapidly (Linderstrøm-Lang, 1950).

Protein dynamics are also important in the context of digestibility of foods. Evenepoel *et al.* fed chickens cooked or uncooked radiolabelled eggs. Ileal sampling showed that cooked egg protein was better digested than uncooked egg protein (Evenepoel et al., 1998). In the GI tract, bile acids and acid contribute to protein digestion by promoting local or global unfolding (Cremers et al., 2014; Fink et al., 1994; Gass et al., 2007). For example, intramolecular repulsion is driven by protonation of amino acids in acidic conditions, as found in the stomach (Fink et al., 1994). Amino acids will be protonated when pH < pK_a. The precise pK_a of the ionisable residues depends on the chemical environment. The lowest average pK_a of side chains in wild-type proteins are of aspartic acid (pK_a = 3.43) and glutamic acid (pK_a = 4.14) (Pahari et al., 2019). Hence, the average pK_a values of amino acids in proteins are higher than the median fasted gastric pH in the human stomach (pH 1.7) (Dressman et al., 1990; Fallingborg, 1999;

Koziolek et al., 2015; Pahari et al., 2019). The lowest experimentally determined amino acid pK_a (0.5) is of a buried aspartic acid in Ribonuclease T1 or T7 lysozyme, where the aspartic acid is stabilised by hydrogen bonds or a salt bridge (Anderson et al., 2002; Giletto and Pace, 1999).

1.6. Advantages of orally administered protein therapeutics

Fear of injections reduces patient quality of life and can negatively impact treatment outcomes, compared to oral administration of medicines (Abramson et al., 2019; Fu et al., 2009; Nir et al., 2003). The reduction in quality of life from injections is particularly pronounced for chronic conditions like diabetes (Fu et al., 2009). Furthermore, the fear of injections and associated pain can be particularly problematic in young adults and children (Nir et al., 2003; Schechter et al., 2007). It is well known that sharing of needles for injecting drugs contributes to disease transmission (Sawangjit et al., 2017), but use of multi-dose vials in hospital settings can also enable disease transmission (Lanini et al., 2009). For example, a meta-analysis of patient-to-patient hepatitis B virus outbreaks in the EU and U.S. reported 30.3% of the transmission pathways to be as a result of multi-dose vials (Lanini et al., 2009).

In farm animals injections may not be an economically feasible method of administering medicines (Shahid and Daniell, 2016). Commercially farmed chicken have an approximately six-week life span, which would result in a large number of injections. Higher purity requirements also increase the cost of injected therapeutics which could be prohibitive in low margin livestock industries (Lei et al., 2013; Shahid and Daniell, 2016). Vaccines or therapeutics that can be included in the feed could

address both issues of cost and ease of administration (Shahid and Daniell, 2016; Virdi et al., 2013, 2019).

Unlike small molecule drugs, orally administered proteins are likely to be confined to the GI tract and thus not act systematically. Local action can be advantageous to reduce side effects, for example with anti-TNFα therapies against inflammatory bowel disease (Crowe et al., 2018; Gareb et al., 2019). Systemic immunosuppression by intravenous delivery of anti-TNFα antibodies can increase risk of cancer (Lemaitre et al., 2017), opportunistic infections (Bonovas et al., 2016) and demyelinating disease (Kunchok et al., 2020). Nevertheless, hijacking intestinal transcytosis pathways is a clever means of combining the ease of oral administration with gut-restricted, local delivery beyond epithelial barrier (Fay et al., 2020; Li et al., 2021; Taverner et al., 2020). For example, Fay *et al.* fused human interleukein-10 (hIL-10) to the *Vibrio cholerae* exotoxin cholix, thereby enabling oral delivery of hIL-10 to the lamina propria (Fay et al., 2020). Subcutaneous delivery of hIL-10 for the treatment of Crohn's disease leads to dose-related side effects (Colombel et al., 2001; Fedorak et al., 2000; Schreiber et al., 2000).

Higher variability in treatment efficacy is likely to be a disadvantage of oral administration of protein therapeutics compared to injections (Tyagi et al., 2018). As discussed above, the GI tract is a dynamic and variable environment which is likely to impact the efficacy of medicines. Orally administered proteins for systemic distribution may also be more expensive than injected versions, because higher doses are required to reach similar levels of bioavailability (Abramson et al., 2019; Smart et al., 2014; Tyagi et al., 2018).

As early as 1923, research was directed at an oral formulation of insulin for the treatment of diabetes (Harrison, 1923), but currently there is no such approved product. Stability in the GI tract and bioavailability are major barriers for oral administration of protein therapeutics (Anselmo et al., 2018; Wang et al., 2015a). Despite considerable research, by 2016 there were only six approved protein-based therapies that are administered orally (Moroz et al., 2016). A notable approval in 2019 is the oral formulation of semaglutide for the treatment of obesity with one weight-related condition, such as high blood pressure or type 2 diabetes (Mullard, 2021). Semaglutide is a human glucagon-like peptide-1 (GLP-1) mimetic that stimulates insulin secretion. Oral administration of protein therapeutics may be particularly effective for targets in the intestinal lumen, like bacterial exotoxins (Moroz et al., 2016).

1.7. Enzymes in the animal feed industry

Enzymes are included in the feed of livestock to improve the growth rate and health of animals (Choct, 2006). For example, phytases are included in feed to catalyse the release of phosphate from phytic acid (*myo*-inositol hexakisphosphate, IP_6), the predominant storage form of phosphate in plants (Lei et al., 2013; Singh, 2008). The key mineral phosphate in phytic acid is not bioavailable to mono-gastric animals because these animals lack the endogenous enzymes to degrade phytic acid. Other than improving the health of the animal, inclusion of phytase improves the feed-conversion-ratio (FCR) (Onyango et al., 2005; Sommerfeld et al., 2018; Walk et al., 2014). The FCR quantifies the mass gained by animal to the mass of feed consumed. Another important class of feed enzymes are non-starch polysaccharidases (NSPase)

which aid in digestion of viscous fibres like arabinoxylans found in wheat and barley (Aftab and Bedford, 2018).

In chickens phytase is active in the gizzard (Elkhalil et al., 2007; Sommerfeld et al., 2018; Walk et al., 2014), a muscular organ immediately following the glandular stomach with an average pH 1.1 – 2.5 and pepsin (Lee et al., 2017, 2018; Svihus, 2011). Evidence for the gizzard being the major site of exogenous phytase action in chickens comes from tracking the degradation products of phytic acid in different segments of the GI tract (Sommerfeld et al., 2018; Walk et al., 2014). Administering three-fold higher doses than the standard dose of phytase in commercial feeds leads to linear increases in FCR, but such high doses may not be commercially viable (Walk et al., 2014). Fusion of a protein to phytase that anchors the enzyme in the gizzard could be a means of improving the efficacy of the administered enzyme without increasing the dose.

1.8. Thesis aims

The primary aim of this thesis was to develop a protein scaffold for targeting in the GI tract. First, the focus was identifying a protein that is resilient to the challenges of the GI tract, particularly to low pH and pepsin. Further criteria included that the protein should express well in *E. coli*, be small, be thermostable and tolerant to engineering and directed evolution. I found SBTI to be resistant to digestion with physiological concentrations of gastrointestinal proteases. Indeed, SBTI was substantially more resistant to digestion with pepsin than established protein scaffolds for oral delivery; nanobodies and nanofitins were rapidly digested at 100-fold lower concentration of

pepsin than found in the stomach, while little degradation of SBTI was observed in stomach-like conditions.

Subsequent work focused on identifying mutable regions in SBTI and establishing phage display for directed evolution, to turn SBTI into a protein scaffold that we termed a gastrobody. I used Rosetta to assist designing the gastrobody binding surface, which was then validated experimentally with an alanine scan. The proof-of-concept was evolving gastrobody binders to two targets, without compromising the stability of the parent protein. I developed binders to two domains of toxins of the enteric pathogen *Clostridium difficile*. Anti-toxin gastrobodies bound their target after digestion with gastrointestinal proteases, but the binding affinity was weak. Hence, I expanded the gastrobody binding surface, extended the length of the binding loops and performed affinity maturation selections. Further work included testing the pepsin-resistance of structural homologues of SBTI, to better understand the structural features contributing to pepsin-resistance.

Chapter 2: Methods

Chapter 2: Methods

2.1. Cloning

Cloning was performed by Gibson assembly (Gibson et al., 2009) using standard methods. Thermo-cycling for polymerase chain reaction (PCR) or other cloning steps was performed in a C1000 Thermal Cycler (Bio-Rad). All primers were synthesized by Integrated DNA Technologies (IDT). Codon optimised sequences of genes were ordered as gBlocks from IDT. Sanger sequencing, performed by Source BioScience (Nottingham, UK), was used to confirm correct cloning of constructs. An ND-1000 NanoDrop (Thermo Scientific) was used to quantify DNA during cloning. Amino acid sequences of selected protein constructs are listed in the Appendix.

2.1.1. PCR

PCR was performed with Q5® High-Fidelity 2× Master mix (NEB). Typical PCRs contained 1× Q5 High-Fidelity master mix, 0.5 µM forward and reverse primers, and 5 ng of template DNA made up to 50 µL with deionised water. Typical cycling conditions included 95 °C for 3 min, followed by 30 cycles of denaturing at 95 °C for 15 s, annealing at 68 °C for 15 s and elongating at 72 °C for 90 s. A typical cycle finished with 5 min at 72 °C followed by cooling to 12 °C. For some PCRs, the annealing temperature was changed, ranging from 60 to 72 °C.

2.1.2. Agarose gel electrophoresis

PCR products were analysed on 0.7% - 1.4% agarose gels depending on the size of the PCR product. 1× SYBR Safe stain (Thermo Scientific) was included in gels to

visualize DNA products. PCR products were loaded with 1× gel loading dye (0.2 M sucrose, 125 µM bromophenol blue, 155 µM xylene cyanol FF). DNA fragments were separated for 45 min at 130 V in TAE buffer (40 mM Tris base, 20 mM acetic acid, 1 mM EDTA, pH 8.0). Imaging was performed on a XcitaBlue Conversion Screen and Filter (Bio-Rad) in a ChemiDoc XRS+ (Bio-Rad). Images were analysed using ImageLab 6.0.1 (Bio-Rad).

2.1.3. PCR product purification

Desired PCR products were excised from the agarose gel. DNA was purified from gel slices using the Wizard® SV Gel and PCR Clean-Up kit (Promega) according to the manufacturer's instructions. Fragments were eluted in deionised water.

2.1.4. Gibson assembly

Home-made mastermix was used for Gibson assembly according to the original protocol (Gibson et al., 2009). Phusion-like DNA polymerase (Wang et al., 2003) and Taq DNA ligase were cloned, expressed and purified by Karl Brune in the Howarth Lab. T5 Exonuclease was purchased from New England Biolabs (NEB). 2× Gibson mix was stored in 10 µL aliquots at -20 °C.

For routine cloning, purified PCR products or gBlocks were mixed in equimolar ratio with the highest possible DNA concentration in a final volume of 20 µl with 1× Gibson mix. Gibson reactions were incubated for 1 – 3 h at 50 °C.

2.1.5. Heat-shock transformation

1 µL of purified supercoiled plasmid or 8 µL of Gibson reaction was transformed into 100 µL aliquots of chemically competent *Escherichia coli* NEB Turbo cells, T7 SHuffle® (NEB) (Lobstein et al., 2012), T7 Express *lysY/I^q* (NEB) or BL21 (DE3) RIPL (Agilent). DNA and cells were incubated for 5 min on ice, before heat-shocking for 45 s at 42 °C. Transformed cells were mixed with 300 µL of SOC media and recovered for 1 h at 37 °C and 200 RPM before plating onto selective media. Lysogeny broth (LB) agar plates containing 100 µg/mL ampicillin or 50 µg/mL kanamycin were used to select for transformants.

2.1.6. Plasmid purification

10 mL LB broth with 100 µg/mL ampicillin or 50 µg/mL kanamycin was inoculated with a relevant colony and grown for 4 – 16 h at 37 °C and 200 RPM. Plasmid was purified from the cells using the GeneJET Plasmid Miniprep kit (Thermo Fisher) according to the manufacturer's instructions.

2.1.7. Library cloning

Phage display libraries were constructed by Gibson assembly (Gibson et al., 2009) with home-made Gibson mix (original library, first loop extended library) or NEBuilder HiFi DNA Assembly Master Mix (NEB) (affinity maturation libraries). Degenerate oligonucleotides with NNK codons in the gastrobody determining regions (GDRs) were ordered from IDT. Randomized SBTI residues were 6 – 9 (GDR3), 22 – 25 (GDR1), 47 – 50 (GDR2) and 124 – 128 (GDR4). NNK codons were inserted in GDRs where

loops were extended. Where libraries contained different loop sizes, separate Gibson reactions were set up to create each combination of loop length. 0.2 pmol of each DNA fragment was mixed with Gibson master mix to create a 20 µL reaction. After 2 h incubation at 50 °C the reactions were pooled and purified by Wizard SV Gel and PCR clean up kit (Promega). Assembled Gibson products were eluted in deionised water. For each library, eight 25 µL aliquots of electrocompetent *E. coli* TG1 or ER2738 (Lucigen) were each transformed with 300 ng purified Gibson assembly in a maximum of 2 µL DNA. 0.2 mm cuvettes (Bio-Rad) were used to electroporate cells with a single 2.5 kV pulse in a MicroPulser (165-2100, Bio-Rad). Electroporated cells were immediately recovered in 1 mL recovery medium (Lucigen). Transformed cells were incubated at 37 °C and 200 RPM for 1 h. Recovered cells were pooled and an aliquot serially diluted in recovery medium (Lucigen) before plating onto LB + 100 µg/mL ampicillin to titer transformants. Library size was calculated from the number of colonies multiplied by the dilution factor.

The remaining cells were plated onto LB + 0.8% (w/v) glucose + 100 µg/mL carbenicillin. Plated cells were grown for 16 h at 37 °C. Library cells were scraped from plates and pooled in 2×TY and centrifuged for 15 min at 16,900 *g* at 4 °C. The library cell pellet was resuspended in 2×TY + 20% (v/v) glycerol and stored at -80 °C. OD_{600} of a 1:100 dilution of the glycerol stock was measured to inform the inoculum volume (at least 10× coverage of library size) for phage production.

2.1.8. Phage display

M13KO7 (New England Biolabs, NEB) was modified to enhance stability following (Fitza, 2002) and is referred to as M13 in this thesis. The M13 helper phage plasmid

was transformed into *E. coli* XL1 by heat-shock transformation and used to produce M13 phage.

Phagemid vector pFAB5c-His_6 (Engberg et al., 1996) was a kind gift from Dr Anthony Keeble in the Howarth lab. With pFAB5c-His_6, POI are expressed fused to pIII(216-425) under control of P_{lacZ}. The promoter system in pFAB5c-His_6 was replaced with P_{ara} and *araC* to create the arabinose-inducible phagemid vector pWPAT. P_{ara} and *araC* were derived from MP6 (Addgene plasmid ID 69669, a kind gift from David Liu) (Badran and Liu, 2015). The phagemid pWPAF was created from pWPAT by replacing pIII(216-425) with pIII from M13.

2.2. Protein expression and purification

2.2.1. Media

Media were prepared by the Media Glassware Sterilization Laboratory of the University of Oxford Department of Biochemistry. *E. coli* was grown in LB, auto-induction media (AIMLB0205, Formedium) or 2×TY. All media were sterilised by autoclaving for 20 min at 121 °C before storage at room temperature in sealed bottles. Cultures were grown in a INFORS Multitron shaker (Infors UK Ltd) with 2.5 cm throw at 200 RPM.

Single colonies were inoculated into 10 mL LB + 50 µg/mL kanamycin and grown for 16 h at 37 °C and 200 RPM to use as starter culture.

2.2.2. Expression constructs and cell strains

Soybean trypsin inhibitor (SBTI), winged bean chymotrypsin inhibitor (WBCI), *Erythrina caffra* trypsin inhibitor (ECTI), SBTI alanine mutants and gastrobodies were expressed in *E. coli* T7 SHuffle® (NEB) (Lobstein et al., 2012) to facilitate disulfide bond formation in the cytoplasm. T7 SHuffle cells express the disulfide isomerase DsbC from a chromosomal copy. SBTI was expressed in cells transformed with pET28a-SBTI-His_6 or pET28a-His_6-Thrombin site-SpyTag003-SBTI. Gastrobody constructs were in the format pET28a-His_6-Thrombin site-SpyTag003-gastrobody. ECTI and WBCI were expressed with C-terminal His_6.

Clostridium difficile toxin A combined repetitive oligopeptide domain (CROP, residues 2304 – 2710) was in expressed *E. coli* BL21 (DE3) RIPL (Agilent) transformed with pET28a-AviTag-His_6-CROP (GenBank MZ337399). *C. difficile* toxin B glucosyltransferase domain (GTD, residues 2 – 543) was expressed in T7 Express *lysY/I^q* (NEB) transformed with pET28a-AviTag-His_6-GTD (GenBank MZ337400).

Anti-CDTA (CROP) nanobody A4.2m (Hussack et al., 2011a) (pET28a-anti-CROP-His_6), anti-IgG Sso7d nanofitin (Béhar et al., 2014) (pET28a-anti-IgG-Sso7d-His_6), plasma kallikrein inhibitor from *Bauhinia bauhinioides* (BBKI-His_6), SBTI-QuantumBlue (Nguyen and Winter, 2006)(pET28a-ss-SBTI-QB) and monomeric enhanced green fluorescent protein (mEGFP, pET28a-His_6-Thrombin site-mEGFP) were expressed in *E. coli* BL21 (DE3) RIPL (Agilent). The plasmid pET28a-His_6-Thrombin site-mEGFP was generated by Dr Robert Wieduwild in the Howarth lab. The SBTI-QB construct was expressed with a periplasmic signal sequence, where the backbone of the plasmid was a kind gift from Dr Jisoo Jean in the Howarth lab. The nanobody was also expressed in *E. coli* T7SHuffle.

2.2.3. Protein expression conditions

His_6-Thrombin site-SpyTag003-SBTI, His_6-Thrombin site-SpyTag003-gastrobody, SBTI GDR3/4 alanine mutants-His_6, AviTag-His_6-GTD, WBCI-His_6, ECTI-His_6 and -anti-CROP-His_6 was produced in 2×TY + 0.5% (v/v) glycerol + 50 µg/mL kanamycin. 1 L cultures were grown at 37 °C and 200 RPM until OD_{600} = 0.5, when cells were induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) (Fluorochem) to 0.42 mM. Growth was continued for 16 h at 18 °C and 200 RPM.

Starter cultures of AviTag-His_6-CROP and His_6-Thrombin site-mEGFP were diluted into 1 L auto-induction medium + 50 µg/mL kanamycin and grown for 22 h at 30 °C and 200 RPM. Starter cultures of SBTI GDR1/2 alanine mutants-His_6 were inoculated into 1 L auto-induction medium + 50 µg/mL kanamycin and grown for 6 h at 37 °C followed by 16 h at 18 °C. A colony of *E. coli* BL21 (DE3) RIPL transformed with pET28a-ss-SBTI-His_6-QB was inoculated into 1 L auto-induction medium + 50 µg/mL kanamycin and grown at 37 °C until OD_{600} = 0.5, when the temperature was reduced to 18 °C and growth was continued for 20 h.

Anti-IgG-Sso7d-His_6 and BBKI-His_6 were produced in LB + 50 µg/mL kanamycin at 18 °C (BBKI) or 30 °C (nanofitin) for 16 h. Cultures were grown at 37 °C and 200 RPM until OD_{600} = 0.5 before induction with IPTG and lowering of the temperature.

2.2.4. Whole cell lysis

Cultures were harvested by centrifugation at 4,000 *g* for 15 min at 4 °C. Cell pellets were resuspended in phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10

mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4). The cell suspension was centrifuged again at 4,000 *g* for 15 min at 4 °C and the supernatant discarded.

Washed cell pellets of WBCI-His_6, BBKI-His_6, ECTI-His_6, His_6-Thrombin site-mEGFP, AviTag-His_6-CROP, AviTag-His_6-GTD, anti-IgG-Sso7d-His_6, and anti-CROP-His_6 were resuspended in 1× Ni-NTA buffer (50 mM Tris-HCl, 300 mM NaCl, pH 7.8) supplemented with cOmplete Mini EDTA-free Protease Inhibitor Cocktail and 1 mM phenylmethylsulfonyl fluoride (PMSF). Cell suspensions were incubated with 100 µg/mL lysozyme for 30 min at 25 °C on a roller to initiate cell lysis. Lysate was sonicated on ice four times for 1 min with 1 min rest periods on a 50% duty cycle using a Sonics Vibra-Cell sonicator (Ultrasonic Processor).

His_6-Thrombin site-SpyTag003-SBTI, His_6-Thrombin site-SpyTag003-gastrobody, SBTI alanine mutants-His_6, and SBTI-His_6 washed cell pellets were resuspended in 3 mL 10× BugBuster protein extraction reagent (Merck) per g pellet. 2 U/mL benzonase, 100 µg/mL lysozyme, cOmplete Mini EDTA-free Protease Inhibitor Cocktail and 1 mM PMSF was added to resuspended cells and incubated for 30 min at 25 °C on a roller. 2-mercaptoethanol was added to 10 mM to lysed cultures of SBTI-His_6 and His_6-Thrombin site-SpyTag003-anti-CROP gastrobodies prior to clarification of lysate.

Cell lysates were clarified by centrifugation at 16,900 *g* for 30 min at 4 °C (sonicated samples) or at 30,000 *g* for 45 min at 4 °C (BugBuster lysed samples). The protein-containing supernatant was transferred to a clean polypropylene tube and kept on ice before Ni-NTA purification.

2.2.5. Periplasmic extraction

SBTI-His_6-QB was purified from the periplasm by osmotic shock. Cells were pelleted by centrifugation at 4,000 *g* at 4 °C before resuspending in 250 mL per L culture ice-cold TES buffer (50 mM Tris-HCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 20% (w/v) sucrose, pH 8.0). The cell suspension in TES buffer was incubated for 15 min at 4 °C with stirring, before pelleting by centrifugation at 4,000 *g* for 15 min at 4 °C. The cell pellet was resuspended in ice-cold 100 mL per L culture 5 mM $MgCl_2$ with 10 mM 2-mercaptoethanol and incubated for 15 min with stirring at 4 °C to generate the osmotic shock fluid (OSF). OSF was clarified by centrifugation at 4,000 *g* at 4 °C for 20 min and filtered through a 0.2 µm filter. SBTI-His_6-QB was purified from OSF by Ni-NTA as described below.

2.2.6. Ni-NTA purification

Ni-NTA purification was performed in the cold room at 4 °C. Clarified cell lysate or OSF was incubated with Ni-NTA beads (Qiagen) on a rotary shaker for 15 – 45 min at 4 °C before applying to a Polyprep gravity column (Bio-Rad). Ni-NTA beads were washed with ten column volumes of 1× Ni-NTA buffer + 10 mM imidazole (wash 1). 10 mM 2-mercaptoethanol was added to wash 1 for SBTI-His_6 and His_6-Thrombin site-SpyTag003-anti-CROP gastrobodies. Next, Ni-NTA beads were washed with five column volumes of 1× Ni-NTA buffer + 30 mM imidazole. Bound protein was eluted using 1× Ni-NTA buffer + 200 mM imidazole. Flow-through and wash samples were collected for later analysis. Protein content of the elution fractions was monitored by A_{280} measurements with a NanoDrop. Elution fractions were pooled for dialysis.

All proteins were dialyzed with 3 kDa molecular weight cut-off (MWCO) dialysis membrane into a buffer in three steps at 4 °C, where each step was at least 3 h. SBTI-

His_6, SBTI GDR1/2 alanine mutants-His_6, and His_6-Thrombin site-SpyTag003-anti-CROP gastrobodies were dialyzed against 50 mM Tris-HCl pH 8.0 + 100 mM NaCl. SBTI-His_6-QB was dialyzed against 100 mM sodium acetate pH 4.5. All other proteins were dialyzed against PBS pH 7.4. Post dialysis, proteins were stored at -80 °C.

Typical yields of purified protein per L culture were 0.5 – 2 mg for His_6-Thrombin site-SpyTag003-gastrobodies, 2.9 mg for SBTI-His_6-QB, 3 mg for SBTI-His_6, 6.8 mg for WBCI-His_6, 10.7 mg for ECTI-His_6, 5.2 mg for BBKI-His_6, 1.8 mg for AviTag-His_6-GTD, 2 mg for AviTag-His_6-CROP, 23 mg for anti-CROP-His_6 (nanobody), 12 mg for anti-IgG-Sso7d-His_6 (nanofitin), 30 mg for His_6-Thrombin site-mEGFP.

2.2.7. Size-exclusion chromatography

His_6-Thrombin site-SpyTag003-SBTI, His_6-Thrombin site-SpyTag003-gastrobodies and AviTag-His_6-GTD (post-biotinylation) were purified by size-exclusion chromatography (SEC). After dialysis, samples were concentrated to < 1 mL using Vivaspin 6 10 kDa MWCO (Sartorius) spin columns. Concentrated proteins were loaded onto a HiLoad 16/600 Superdex 75 pg column controlled by an ÄKTA Pure 25 (GE Healthcare). SEC was run at a 1 mL/min flow rate of PBS pH 7.4 at 4 °C. AviTag-His_6-GTD was purified on a HiLoad 16/600 Superdex 200 pg column. Fractions containing the desired protein were pooled and concentrated using Vivaspin 20 5 kDa MWCO (Sartorius) to < 2 mL.

2.2.8. Other proteins

β-lactamase (BLA) was a kind gift from Dr Jisoo Jean in the Howarth lab (Schoene et al., 2014). BLA was expressed in *E. coli* BL21 (DE3) RIPL (Agilent) transformed with pET28a-BLA-His_6. BLA was expressed in LB + 0.8% (w/v) glucose at 18 °C for 16 h and purified with Ni-NTA (Qiagen). BLA-His_6 was dialyzed into PBS pH 7.4.

ss-His_6-QuantumBlue was a kind gift from Dr Jisoo Jean in the Howarth group. ss-His_6-QuantumBlue was purified by periplasmic extraction and Ni-NTA as described above, before dialysis into 100 mM sodium acetate pH 4.5.

GST-BirA was produced from pGEX-2T (Fairhead and Howarth, 2015). pGEX-2T was a kind gift from Chris O'Callaghan at the University of Oxford.

2.3. SDS-PAGE

6× SDS-PAGE loading buffer [0.23 M Tris-HCl pH 6.8, 24% (v/v) glycerol, 120 µM bromophenol blue, 0.23 M sodium dodecyl sulfate] was mixed with samples and boiled for 3 min at 99 °C. For reducing gels, 66 µM 1,4-dithiothreitol (DTT) was included in the sample before boiling. Samples were loaded onto 12 or 16% Tris-glycine gels depending on the molecular weight of proteins being analysed. Gels were run in an XCell SureLock system (Thermo Fisher) at 190 V for 60 min or until the dye front reached the bottom of the gel. Gels were stained with InsantBlue Coomassie stain (Expedeon) at 25 °C on a shaker for 16 h and destained with deionised water for 16 h at 25 °C. Gels were imaged using a ChemiDoc XRS+ imaging system (Bio-Rad).

Image analysis and densitometry was performed with ImageLab 6.0.1 (Bio-Rad). ImageLab intensity measurements were performed with low band sensitivity band detection and background subtraction at 2 mm disk size.

2.4. Quantification of protein concentration

Protein concentration was calculated from A_{280} measurements with a ND-1000 NanoDrop and the extinction coefficient, calculated with ExPASy ProtParam (Gasteiger et al., 2003).

2.5. Site-specific biotinylation

AviTag-His_6-GTD and AviTag-His_6-CROP were mono-biotinylated using GST-BirA as described previously (Fairhead and Howarth, 2015). 30-fold excess of AviTag proteins in PBS pH 7.4 were incubated with GST-BirA and 5 mM $MgCl_2$, 2 mM ATP and 100 µM biotin for 1 h at 30 °C. The same quantity of GST-BirA and biotin was added again and the incubation continued for 1 h at 30 °C. Excess biotin was removed by three 3 h dialysis steps against PBS pH 7.4 with a 3 kDa MWCO dialysis membrane at 4 °C.

Biotinylation was tested by mass spectrometry (see below) and a streptavidin gel shift assay. Varying concentrations of biotinylated protein were incubated with 0.2 mg/mL streptavidin (NEB) for 10 min at 25 °C. Samples were mixed with SDS-loading buffer and analysed by SDS-PAGE without boiling (Fairhead and Howarth, 2015).

2.6. Mass spectrometry

Electrospray ionisation mass spectrometry (ESI-MS) was performed with the Mass Spectrometry Research Facility, Department of Chemistry at the University of Oxford. I prepared all samples and analysed data. The samples were run on a RapidFire

Agilent 6550 quadrupole-time of flight mass spectrometer by Dr Anthony Tumber. Data were analysed with the MassHunter software platform Qualitative Analysis B.07.00 (Agilent).

2.6.1. Intact protein ESI-MS

SBTI GDR1/2 alanine mutants in 50 mM Tris-HCl pH 8.0 100 mM NaCl were heated to 75 °C for 10 min to aggregate contaminants. After cooling, proteins were centrifuged at 16,900 *g* for 30 min at 4 °C and the supernatant was diluted to 10 µM. All other proteins were diluted to 10 µM in PBS pH 7.4 without heating.

Prior to loading onto a C4 solid-phase extraction cartridge, samples were acidified to 0.9% (v/v) with formic acid and aspirated under vacuum for 0.3 s. The main elution fraction with 85% (v/v) acetonitrile, 15% (v/v) deionized water and 0.1% (v/v) acetic acid was analysed. Spectra were deconvoluted using the MassHunter software platform with the following settings: Deconvolute (protein), maximum entropy deconvolution algorithm, mass range 10,000.00 – 50,000.00 Da, mass step 1.0000 Da, m/z limited range 500.0000 – 2400.0000 m/z, baseline subtraction factor 7.00, proton adduct and automatic isotope width. Expected protein masses were calculated with ExPASy ProtParam (Gasteiger et al., 2003), assuming formation of all disulfide bonds and removal of N-terminal formylmethionine. Gluconoylation (+178 Da) or phosphogluconoylation (+258 Da) is a common post-translational modification of His-tagged proteins expressed in *E. coli* (Geoghegan et al., 1999).

2.6.2. Mapping protease cut sites

To map protease cut sites, gastrobodies were digested with physiological concentrations of proteases before ESI-MS. Digestion was performed for 15 min at 37 °C. SBTI-His_6, His_6-Thrombin site-SpyTag003-T12, His_6-Thrombin site-SpyTag003-GT01 and His_6-Thrombin site-SpyTag003-GT01 R63A were digested with 1 mg/mL porcine pepsin A in 50 mM glycine-HCl pH 2.2. Pepsin digestion was stopped by addition of 2.5 M Tris-HCl pH 8.8. Digestion with 100 TAME U/mL trypsin from bovine pancreas (T1426, Merck) [1 unit hydrolyses 1 µM of p-toluene-sulfonyl-L-arginine methyl ester (TAME) per minute at 25 °C at pH 8.2 in the presence of 1 mM Ca^{2+}], 25 BTEE U/mL α-chymotrypsin from bovine pancreas (C3142, Merck) [1 unit hydrolyses 1 µM of benzoyl-L-tyrosine ethyl ester (BTEE) per minute at pH 7.8 and 25 °C], or 10 U/mL elastase from porcine pancreas (E1250, Merck) was performed in 50 mM Tris-HCl pH 6.8 + 10 mM $CaCl_2$ and stopped by addition of PMSF to 1 mM.

After stopping digestion, formic acid was added to 0.9% (v/v) before aspirating under vacuum for 0.3 s. Aspirated acidified proteins were loaded onto a C4 solid-phase extraction cartridge. Fractions in the first wash step [0.1% (v/v) formic acid in deionized water] after the main elution [85% (v/v) acetonitrile, 15% (v/v) deionized water and 0.1% (v/v) acetic acid] were analysed using the MassHunter software platform with the settings describe above. Numbering of residues in ESI-MS figures was based on their position in the construct, not on PDB ID: 1AVU.

2.7. Rosetta

Rosetta3 (Leaver-Fay et al., 2011) (Release Version 2018.09.60072) was used to model SBTI (PDB ID:1AVU) (Song and Suh, 1998) (Fig. 2.1). Calculations were performed on my personal computer (Lenovo Yoga 920 with Intel® Core™ i5-8250U

CPU). Output files were processed with Bash and Python scripts and visualised in Microsoft Excel. Python and Rosetta commands were run via the Ubuntu application for Windows (version 2004.2021.825.0). Annotated scripts and commands are listed in the Appendix. Structures were manually inspected in PyMOL 2.3.4 (Schrödinger).

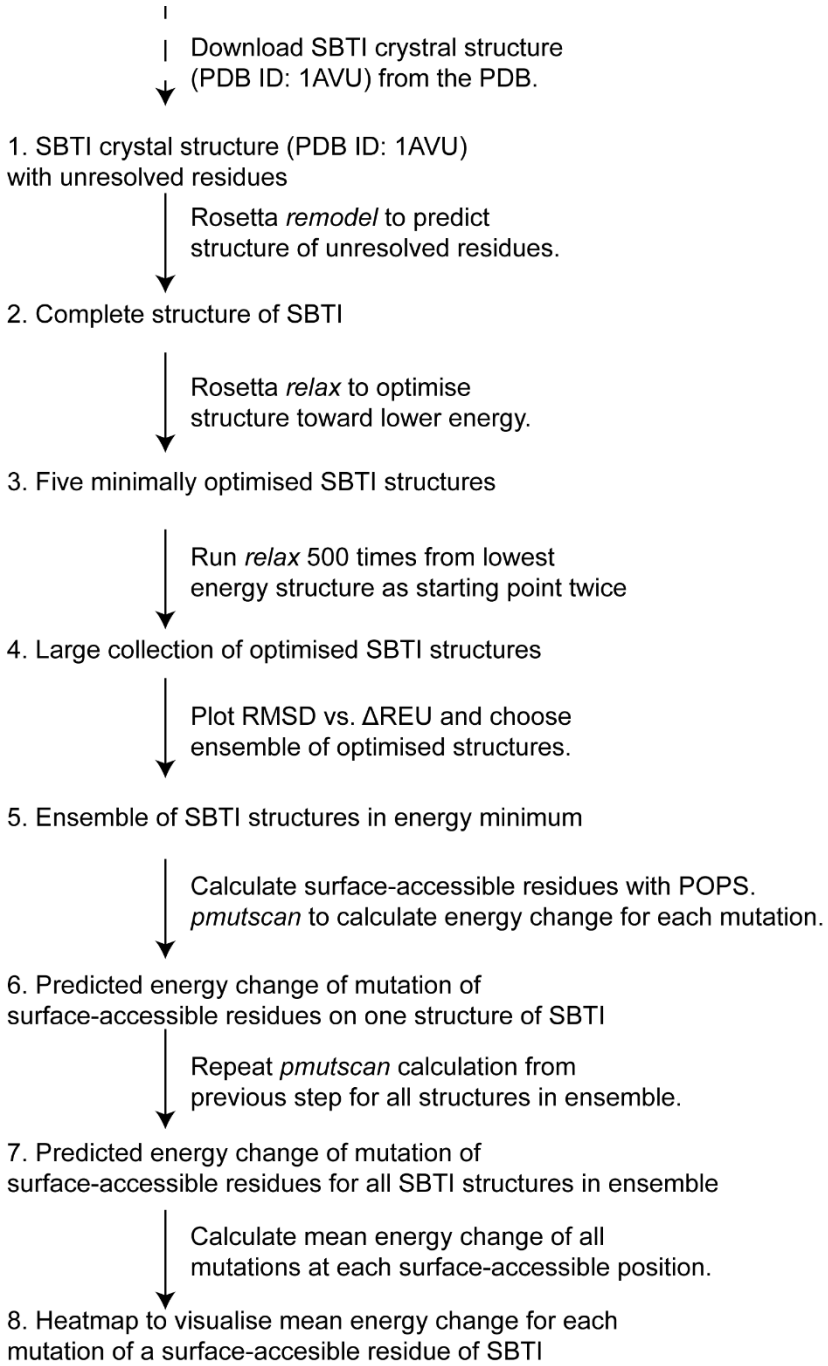


Figure 2.1. Workflow for in silico analysis of SBTI mutability. Mutability analysis was performed using the Rosetta3 modelling software. Steps in mutability analysis are shown as a flow chart.

2.7.1. Modelling missing density in SBTI crystal structure

The crystal structure of SBTI (PDB ID: 1AVU) (Song and Suh, 1998) was downloaded from the PDB. Density for D125, D126, A140, E141 and D142 is missing in the crystal structure of SBTI. It is not possible to analyse the structure of SBTI with Rosetta when there are unresolved residues. The *remodel* application can be used to design loops (Huang et al., 2011). In this case, *remodel* was used to model the unresolved residues. First, the 'getBlueprintFromCoords.pl' script was used to generate a blueprint file from the 1AVU PDB file. The blueprint file is used to describe what *remodel* should do. The blueprint file was manually edited to include the missing residues. Two residues either side of the missing residues being modelled were allowed to be redesigned to avoid steric clashes. Because design calculations in Rosetta are stochastic, the *remodel* protocol was with three trajectories (starting points) to produce thirty predicted structures of SBTI including D125, D126, A140, E141 and D142.

2.7.2. Generating an ensemble of SBTI structures

The Rosetta *pmutscan* protocol mutates residues *in silico* and predicts the change in folding energy, represented as rosetta energy units (REU). The ΔREU scores calculated by *pmutscan* are influenced by backbone and side chain geometry. Since protein structure is dynamic, it is important to sample a protein's conformational space when analysing the impact of mutations using *pmutscan*. I used the *relax* protocol to generate an ensemble of structures, representative of the SBTI conformational space (Conway et al., 2014).

Relax uses a stochastic process to “wiggle” a starting structure into into a lower energy state. The output structure from *remodel* had a high REU, so I initially ran 5 iterations of the *relax* protocol to generate a starting structure. The lowest energy structure was used as starting point for another 500 iterations (run 1) of *relax*. I used Rosetta’s *score.static* application to calculate the root mean square deviation (RMSD) of all the structures in run 1 relative to the lowest energy structure. It is expected that RMSD and REU will converge onto an energy minimum. RMSD was plotted against REU in Microsoft Excel. Since convergence of RMSD and REU was not observed, the lowest energy structure of run 1 was used for a second instance of 500 iterations of *relax* (run 2), before plotting RMSD against REU. Structures within an energy minimum (-503 < REU < -497 and 0.119 Å < RMSD < 0.238 Å) were picked as the ensemble (467 structures).

2.7.3. In silico *mutagenesis of solvent accessible residues*

The Parameter Optimised Surfaces (POPS) webserver (Cavallo et al., 2003) was used to define solvent accessible residues from a representative structure in the ensemble. Solvent accessibility of residues (Q) was scored by POPS as the quotient of surface accessible area and the surface area of isolated atoms (Cavallo et al., 2003). Residues where Q > 0.2 were defined as surface-accessible. A bash script was used to extract the residue ID, residue number and Q score from the ouput file generated by the POPS server.

Rosetta *pmutscan* was performed on all surface-accessible residues (except cysteine) of all structures in the ensemble. *Pmutscan* requires an input PDB structure and file specifying the individual mutations to be scored. A Python script was used to create a

list of all the surface accessible residues where Q > 0.2 and to specify each to be mutated to all natural amino acids except cysteine. Cysteine was excluded to avoid disruption of native disulfide bond formation. All the structures in the ensemble were listed in a text file. A bash script was used to execute the *pmutscan* command for each structure in the ensemble by iterating through the list of ensemble structures.

2.7.4. Rosetta data processing and visualisation

For each structure, *pmutscan* generates a file listing the ΔREU of each predicted mutation. A bash script was used to calculate the mean ΔREU and standard deviation for each individual mutation from all ensemble structures. A Python script was used to reshape the data into a matrix, where the x-axis is the residue that was mutated to and the y-axis is the residue position. The matrix was viewed in Microsoft Excel and coloured according to the mean ΔREU score. Mean ΔREU for a residue in the SBTI structure was calculated by averaging the mean ΔREU of all 18 amino acid mutations considered at that position. Mutation to proline was excluded from the calculation because it leads to disproportionately high REU scores. Mean ΔREU of residues was visualized using PyMOL putty and plotted as a bar graph in Microsoft Excel.

2.8. Gastric fluid proteolysis assay

Gastric fluid of female, non-medicated, non-immunized, non-fasted BALB/c mice was purchased from BioIVT. Gastric fluid from *ad libitum*-fed day 22 Ross 308 broilers at Drayton Animal Health (UK) was collected post-mortem from gizzards. Chicken gastric fluid was kept on ice for 4 h after collection (during transport) and centrifuged for

16,900 *g* for 30 min at 4 °C to remove large particulates. Clarified chicken gastric fluid and mouse gastric fluid were stored at -80 °C.

Proteolytic activity in gastric fluid was assayed by digestion of mEGFP (Malik et al., 2005). 1 volume of mEGFP in 50 mM glycine-HCl pH 2.2 (final concentration 14 µM) was added to 4 parts 1 mg/mL pepsin from pig gastric mucosa (final concentration 3,028 U/mL) (P6887, Merck) in 50 mM glycine-HCl pH 2.2 and incubated at 37 °C. Alternatively, 1 volume of mEGFP in 50 mM glycine-HCl pH 2.2 (final concentration 14 µM) was mixed with 4 parts mouse or chicken gastric fluid and incubated at 37 °C. Addition of 1 M Tris-HCl pH 8.8 was used to stop digestion. Neutralised digests were incubated for 5 min to allow mEGFP to re-gain fluorescence. Fluorescence was measured at 528 nm (excitation 488 nm) with a ClarioSTAR plate-reader (BMG Labtech). T = 0 wells were neutralised to inactivate pepsin before addition of mEGFP. Fluorescence at t = 0 was set to 100%.

2.9. Protease resistance assays

2.9.1. SDS-PAGE pepsin digestion assay

Proteins of interest were incubated with 1 mg/mL porcine pepsin A (final concentration 3,028 U/mL) (P6887, Merck) at 37 °C. 1 part protein of interest was mixed with 4 parts pepsin. Pepsin was in 50 mM glycine-HCl pH 2.2 Proteins of interest were in PBS pH 7.4. SBTI GDR1/2 alanine mutants were in 50 mM Tris pH 8 + 100 mM NaCl. Nanobody, nanofitin and SBTI pepsin stability was compared in serial dilutions of pepsin in 50 mM glycine-HCl pH 2.2. To stop digestion, SDS-loading buffer was added to 1× and samples boiled immediately at 99 °C for 3 min.

Boiled samples were run on SDS-PAGE and stained with Coomassie. Digestion was measured by quantifying the band intensity of the intact protein using ImageLab 6.0.1 (Bio-Rad). Remaining protein was defined as 100× [target band intensity / mean band intensity at 0 min]. Digests were performed in triplicate.

2.9.2. SDS-PAGE pancreatin or elastase digestion assay

Scaffold proteins (SBTI, nanobody, nanofitin) were incubated with 0, 0.1, 1 or 10 mg/mL pancreatin from pig pancreas (P1750, Merck) or 0, 0.1, 1 or 10 U/mL elastase from pig pancreas (E7885, Merck) in 50 mM Tris-HCl pH 6.8 + 10 mM $CaCl_2$ at 37 °C for 30 min. Samples were boiled for 3 min at 99 °C with 1× SDS-loading buffer to stop digestion before running on SDS-PAGE.

2.9.3. Mouse gastric fluid digestion assay

4 volumes of female non-medicated, non-immunized, non-fasted BALB/c mouse gastric fluid (BioIVT) was mixed with 1 volume of His_6-Thrombin site-SpyTag003-GT01 in PBS pH 7.4. The reaction was incubated at 37 °C for 0, 10, 30 or 60 min in triplicate. Samples were taken at each time point and digestion stopped by boiling in 1× SDS-loading buffer at 99 °C for 3 min. Samples with only His_6-Thrombin site-SpyTag003-GT01 or mouse gastric fluid were included as controls.

2.9.4. ELISA on binding GTD after protease digestion

Wells of a 96-well Nunc Maxisorp plate were coated with 1 µg/mL GTD in PBS pH 7.4 for 16 h at 4 °C. After washing once with PBS + 0.05% Tween-20 (PBS-T), wells were blocked for 1 h with 3% (w/v) skim milk in PBS pH 7.4. Gastrobodies were digested with proteases for 0 or 30 min at 37 °C. 0 min digests were stopped immediately after addition of gastrobody. Pepsin digests were with 1 mg/mL porcine pepsin A (3,028 U/mL, P6887, Merck) in 50 mM glycine-HCl pH 2.2. Pepsin digests were stopped by dilution into PBS pH 7.4. Intestinal protease digests were in 50 mM Tris-HCl pH 6.8 + 10 mM $CaCl_2$ with 100 or 10 TAME U/mL trypsin from bovine pancreas (T1426, Merck), 25 BTEE U/mL α-chymotrypsin from bovine pancreas (C3142, Merck) or 10 U/mL elastase from porcine pancreas (E1250, Merck). PMSF was added to 5 mM to stop trypsin, chymotrypsin and elastase digestion. To minimise trypsin-bound SBTI, trypsin digests were incubated for 30 min at 25 °C after addition of PMSF. Gastrobody controls for trypsin digests were incubated in 50 mM Tris-HCl pH 6.8 + 10 mM $CaCl_2$. Gastrobody controls for pepsin, chymotrypsin or elastase digests were incubated in PBS pH 7.4.

Digest samples were serially diluted 1:3 in PBS-T + 1% (w/v) skim milk and transferred to GTD-coated wells. After 1 h at 25 °C, bound gastrobodies were detected with 1:5,000 diluted rabbit anti-trypsin inhibitor (Rockland, 200-401-079, lot #36926) followed by 1:15,000 diluted anti-rabbit IgG(H+L):horseradish peroxidase (HRP) (65-6120, Invitrogen). Antibodies were in PBS-T + 1% (w/v) bovine serum albumin (BSA) and incubated for 1 h at 25 °C. Between each gastrobody or antibody incubation the wells were washed three times with PBS-T. Following the last three washes, 1-Step Ultra TMB-ELISA substrate solution (34029, Thermo Fisher) was added to the wells. Development of the ELISA was monitored at 652 nm with a FLUOStar Omega (BMG

Labtech) until the highest absorbance value was approximately 1. Colour development was stopped by addition of HCl to 0.5M and absorbance was read at 450 nm.

2.10. Temperature resistance assays

2.10.1. Temperature-dependent solubility assay

SBTI GDR1/2 alanine mutants were heated to 25, 75, 90 or 100 °C in 50 mM Tris-HCl pH 8.0 + 100 mM NaCl for 10 min. His_6-Thrombin site-SpyTag003-WT SBTI or His_6-Thrombin site-SpyTag003-gastrobodies were heated to 37, 55, 75 or 100 °C for 10 min in PBS pH 7.4. Cooled proteins were centrifuged for 30 min at 16,900 *g* at 4 °C to pellet aggregates. Soluble protein in the supernatant was mixed with 1× SDS-loading dye and run on SDS-PAGE. Gels were stained with Coomassie. Band intensity was quantified with ImageLab 6.0.1 (Bio-Rad). Soluble protein was defined as 100 × [target band intensity / mean intensity of control band]. Controls were samples kept at 25 °C for GDR1/2 alanine mutants or 37 °C for gastrobodies. The experiment was performed in triplicate.

2.10.2. Heat-resilience ELISA

Wells of a 96-well Nunc Maxisorp plate were coated with 10 µg/mL trypsin (T1426, Merck) in PBS pH 7.4 for 16 h at 4 °C and blocked with PBS-T + 3% (w/v) BSA for 1 h at 25 °C. His_6-Thrombin site-SpyTag003-WT SBTI or His_6-Thrombin site-SpyTag003-gastrobodies were heated to 37, 55, 75 or 100 °C for 10 min in PBS pH 7.4, and aggregates removed as described in ‘Temperature-dependent solubility assay’. Soluble proteins were serially diluted in PBS-T + 1% (w/v) BSA and incubated

in trypsin-coated wells for 1 h at 25 °C. Bound SBTI or gastrobody was detected as described above for ‘ELISA on binding GTD after protease digestion’.

2.10.3. Differential scanning calorimetry (DSC)

DSC was performed on a MicroCal PEAQ-DSC (Malvern). The machine was operated by Dr David Staunton of the Department of Biochemistry, University of Oxford. Dr David Staunton performed raw data processing. I prepared samples and analysed the data.

SBTI-His_6 was dialyzed against 50 mM Na_2HPO_4 adjusted to pH 2.0 or pH 7.4 with orthophosphoric acid (chapter 3). In all other DSC experiments, proteins were dialyzed against 10 mM KH_2PO_4 (pH 1.2), 50 mM KH_2PO_4 (pH 2.0), or 50 mM K_2HPO_4 (pH 7.4) where the phosphate buffer pH was adjusted with HCl.

Thermal transitions were monitored from 20 to 100 °C at 3 °C/min and 3 atm pressure. MicroCal PEAQ-DSC analysis software (version 1.22) was used to analyse data. Blank buffer signal from a reference cell was subtracted from the experimental sample. Baseline subtraction was included. MicroCal PEAQ-DSC analysis software (version 1.22) was used to fit the thermal transition to a two-state model to obtain a melting temperature (T_m).

2.11. Surface plasmon resonance (SPR)

A Biacore T200 (Cytiva Life Sciences) was used for SPR experiments. Data were analysed with the Biacore T200 evaluation software (Cytiva Life Sciences). Biotinylated AviTag-His_6-CROP or biotinylated AviTag-His_6-GTD in PBS pH 7.4 +

0.05% Tween 20 (PBS-T) was flowed over a Sensor Chip CAP coated in Biotin CAPture reagent (Cytiva Life Sciences) to create the binding surface. The reference cell contained Biotin CAPture reagent but no biotinylated target protein.

2.11.1. Multi-cycle SPR

For multi-cycle experiments, serial dilutions of His_6-Thrombin Site-SpyTag003-gastrobody were injected in PBS-T at a flow rate of 60 µL/min for 200 s, followed by 200 s dissociation time. Gastrobody dilutions were injected in order of increasing concentration. 6 M guanidine + 0.25 M NaOH (pH ~ 13.5) was used to regenerate the binding surface between cycles. All experiments were performed at 25 °C with double reference subtraction. Double reference subtraction was performed by subtracting signal from an injection of running buffer (PBS-T) and signal from injection into the reference cell without biotinylated bait.

Binders in the micromolar range (His_6-Thrombin Site-SpyTag003-GT_S1_01 and His_6-Thrombin Site-SpyTag003-anti-CROP gastrobody) were analysed using equilibrium analysis assuming a 1:1 binding model to obtain the dissociation constant (K_d). Equilibrium binding values obtained from Biacore T200 Evaluation software (Cytiva Life Sciences) were fit to a nonlinear regression in GraphPad Prism (version 9.0.0) to determine K_d. The specific 1:1 binding model describes a rectangular hyperbola:

$$Y = \frac{B_{max} \times [L]}{K_d \times [L]}$$

Where [L] is the concentration of ligand (gastrobody), B_{max} is the maximum binding, K_d is the dissociation constant, and Y is the observed binding. The standard error reported represents the uncertainty of fit. Binding of His_6-Thrombin Site-SpyTag003-

GT_S1_01 and His_6-Thrombin Site-SpyTag003-anti-CROP gastrobody was analysed once with one duplicate concentration.

Kinetic analysis with a specific 1:1 binding model in the Biacore T200 Evaluation software (Cytiva Life Sciences) was used to obtain k_{off} (dissociation rate constant) and k_{on} (association rate constant) of nanomolar range binders (anti-GTD gastrobodies). His_6-Thrombin Site-SpyTag003-GT01 binding was analysed in three separate experiments. Other anti-GTD gastrobodies binders were analysed once. His_6-Thrombin Site-SpyTag003-GT01 R63A K_d was determined by equilibrium analysis from triplicate measurements (mean ± 1 s.d.).

2.11.2. Single-cycle SPR

For single-cycle experiments, increasing concentrations of His_6-Thrombin Site-SpyTag003-GT01 or His_6-Thrombin Site-SpyTag003-WT SBTI (25, 50, 100, 200, 400 nM) in 20 mM MES pH 6.5 150 mM NaCl + 0.05% Tween 20 were injected for 200 s at 60 µl/s at 25 °C. Kinetic analysis with a specific 1:1 binding model in the Biacore T200 Evaluation software (Cytiva Life Sciences) was used to obtain k_{off} (dissociation rate constant) and k_{on} (association rate constant). Three separate single-cycle experiments were performed for His_6-Thrombin Site-SpyTag003-GT01 and His_6-Thrombin Site-SpyTag003-WT SBTI.

2.12. Plate assays on GT01 R63A binding

Wells of a 96-well Nunc Maxisorp plate were coated with 10 µg/mL trypsin from bovine pancreas (T1426, Merck) or 1 µg/mL non-biotinylated AviTag-His_6-GTD in PBS pH 7.4

for 16 h at 4 °C. Antigen-coated wells were washed with PBS-T before blocking for 1 h with PBS pH 7.4 + 3% (w/v) BSA at 25 °C. His_6-Thrombin site-SpyTag003-GT01 or His_6-Thrombin site-SpyTag003-GT01 R63A were serially diluted in PBS-T and incubated in antigen-coated wells for 1 h at 25 °C.

Trypsin-bound gastrobodies were incubated with 20 nM biotinylated AviTag-His_6-GTD in PBS-T + 1% (w/v) BSA for 1 h at 25 °C. Bound biotinylated AviTag-His_6-GTD was detected by incubation with 1:5,000 diluted streptavidin-HRP (SA10001, Thermo Fisher) in PBS-T + 1% (w/v) BSA for 1 h at 25 °C.

GTD-bound gastrobodies were incubated with primary antibody 1:5,000 diluted rabbit anti-trypsin inhibitor (34549, Abcam) and secondary antibody 1:15,000 diluted anti-rabbit IgG(H+L):HRP (65-6120, Invitrogen). Antibodies were in PBS-T + 1% (w/v) BSA and incubated for 1 h at 25 °C.

Between incubations, wells were washed three times with PBS-T. Following three final washes with PBS-T, the plate assays were developed with 1-Step Ultra TMB-ELISA substrate solution (34029, Thermo Fisher). Development was monitored using a FLUOStar Omega (BMG Labtech) at 652 nm. Data of binding to GTD were normalized to the absorbance at the highest gastrobody concentration.

2.13. ELISA on binding specificity

Nunc Maxisorp plates were coated with 5 µg/mL hen egg lysozyme (HEL, L6876, Merck), BLA-His_6, trypsin from bovine pancreas (T1426, Merck), non-biotinylated AviTag-His_6-GTD or AviTag-His_6-CROP in PBS pH 7.4 for 16 h at 4 °C. Wells were blocked with PBS pH 7.4 + 5% (w/v) skim milk for 2 h at 25 °C, after washing antigen-coated wells once with PBS-T. His_6-Thrombin Site-SpyTag003-gastrobody were diluted to 500 nM in PBS pH 7.4 and incubated in antigen-coated wells for 30 min at

25 °C. Gastrobody binding was detected with primary antibody 1:5,000 diluted rabbit anti-trypsin inhibitor (34549, Abcam). The secondary antibody was 1:7,000 diluted anti-rabbit IgG(H+L):HRP (65-6120, Invitrogen). Antibodies were in PBS-T + 1% (w/v) skim milk and allowed to bind for 45 min at 25 °C. Wells were washed three times with PBS-T between incubations. The ELISA was developed with 1-Step Ultra TMB-ELISA substrate solution (34029, Thermo Fisher) and colour change was measured with a FLUOStar Omega (BMG Labtech) at 652 nm.

2.14. Phytase assay

The phytase assay was modified from (Buldun et al., 2018). 96-well, flat-bottom, polystyrene plates were blocked with 100 mM sodium acetate pH 4.5 + 3% (w/v) BSA for 2 h at 37 °C. Blocked wells were washed twice with 100 mM sodium acetate pH 4.5. 2 nM of ss-His_6-QuantumBlue or ss-SBTI-His_6-QuantumBlue was incubated with 5 µM phytic acid sodium salt (Merck) at 50 °C in 100 mM sodium acetate pH 4.5. Reactions were incubated for 0, 0.5, 1, 3, 5, 10 or 15 min, but were set up such that all reactions finished at the same time. Biomol Green (Enzo Life Sciences) was added to stop the reactions. Plates were incubated for 20 min at 25 °C. Absorbance at 620 nm was measured using a SpectraMax M3 microplate reader (Molecular Devices). The experiment was performed in triplicate.

2.15. Phage display

2.15.1. Phage production

E. coli TG1 cells (Lucigen) were grown in 200 mL 2×TY + 2% (w/v) glucose at 37 °C and 200 RPM shaking before phage infection. If TG1 cells were transformed with a phagemid, 100 µg/mL ampicillin was included in the growth media. Cultures were inoculated from glycerol stocks. When the *E. coli* TG1 culture reached OD_{600} = 0.5, M13 was added at a multiplicity of infection of 10:1 (phage:bacteria). Phage were incubated with bacteria for 45 min at 37 °C and 80 RPM to enable infection. The infected culture was centrifuged at 2,500 *g* for 10 min at 4 °C to pellet bacterial cells. The supernatant was discarded and the cell pellet was resuspended in induction medium comprising 200 mL 2×TY + 0.2% (w/v) L-arabinose + 0.2% (v/v) glycerol + 100 µg/mL ampicillin + 50 µg/mL kanamycin. Different concentrations of L-arabinose for induction of recombinant pIII production were tested. Phage growth was continued for 16 h at 18 °C and 200 RPM.

In early pilot experiments, *E. coli* ER2738 was used in place of *E. coli* TG1 and media were supplemented with 10 µg/mL tetracycline. Recombinant pIII production was induced with IPTG added to 0.42 mM instead of L-arabinose.

2.15.2. Phage purification

Cultures were incubated with 5% (w/v) polyethylene glycol 8000 (PEG8000, Thermo Fisher) + 0.5 M NaCl for at least 1 h at 4 °C on ice to precipitate phage. Precipitated phage were pelleted by centrifugation at 15,000 *g* at 4 °C for 30 min and the supernatant discarded. The phage pellet was resuspended in PBS pH 7.4, before pelleting bacterial cells by centrifugation at 15,000 *g* at 4 °C for 10 min. Purification and concentration of phage by precipitation was repeated for a total of three rounds. Resulting purified phage were stored in PBS pH 7.4 + 15% (v/v) glycerol at -80 °C.

Stocks of phage were titered by quantitative PCR (qPCR). qPCR primers Fwd2 (5'-GTCTGACCTGCCTCAACCTC-3') and Rev2 (5'-TCACCGGAACCAGAGCCAC-3') anneal to *gIII* of M13 phage. Phage were diluted at least 10^3-fold in PBS pH 7.4 and mixed with 2× SensiMix (Bioline) master mix and primers Fwd2/Rev2 (250 nM final concentration). Samples were titered in duplicate relative to a dilution series of M13KO7 (NEB) ranging from 10^{10} to 10^6 PFU/mL. qPCR was performed on a Mx3000P qPCR machine (Agilent). Data were analysed using MxPro qPCR software (Agilent).

2.15.3. Gastrobody selection against CROP

Selections against biotinylated AviTag-His_6-CROP were performed with M13 phage displaying a library of SBTI on truncated pIII(216-425). SBTI-pIII expression during phage production was induced with 0.2% (w/v) L-arabinose. In the naïve library, GDR1 (amino acid residues 22 – 25) and GDR2 (amino acid residues 47 – 50) had been randomised with NNK codons. Two selection campaigns were performed in parallel, where bound phage was eluted either at pH 2 or at pH 11.

In the first round, 10^{13} phage were incubated with 500 nM biotinylated AviTag-His_6-CROP in PBS pH 7.4 + 3% (w/v) BSA at 25 °C for 3 h. The phage-bait mixture was incubated in a microfuge tubed blocked with 3% (w/v) BSA to reduce phage sticking to the tube. Biotin Binder Dynabeads (Thermo Scientific) were used to capture bait-bound phage. Beads were incubated with the phage-bait mixture at 25 °C for 1 h. Non-specifically bound phage were removed by four washes with PBS-T, one wash with 50 mM Tris-HCl pH 7.5 + 0.5 M NaCl, and two washes with PBS pH 7.4. Phage were eluted with 50 mM glycine-HCl pH 2.2 or 100 mM triethylamine pH 11. Elutions were

neutralised with 2.5 M Tris-HCl pH 8.8 (following acid elution) or 1 M Tris-HCl pH 7.5 (following alkaline elution).

Recovered neutralised phage were incubated with *E. coli* TG1 cells for 1 h at 37 °C and 200 RPM. Infected *E. coli* TG1 cells were serially diluted in SOC media and plated onto LB + 100 µg/mL ampicillin plates to titre the eluted phage. Remaining infected cells were transferred to 50 mL 2×TY + 2% (w/v) glucose + 100 µg/mL ampicillin and grown for 16 h at 37 °C and 200 RPM. Bacterial cells were pelleted by centrifugation at 4,000 *g* and 4 °C for 15 min. The cell pellet was resuspended in 2×TY + 20% (w/v) glycerol and stored at -80 °C. Glycerol stocks were used as inoculum for producing phage for the next round of selection. Phage production cultures were inoculated with at least 10^{10} *E. coli* cells to a starting OD_{600} < 0.05, to ensure complete library coverage.

The second and third round of selections against biotinylated AviTag-His_6-CROP were performed as for the first round, but with the following modifications to increase the stringency of selection. Binding to streptavidin or the beads was selected against before round 2 and 3 by incubating phage with Biotin Binder Dynabeads in PBS pH 7.4 + 3% (w/v) BSA for 90 min at 25 °C. Biotin Binder Dynabeads were settled by centrifugation in a mini centrifuge (~2,000 *g*) for 1 min at 25 °C. The supernatant containing unbound phage was used as input for the selection steps. In round 2 and 3, phage input was reduced to 10^{11}. The concentration of biotinylated AviTag-His_6-CROP was reduced to 300 nM (round 2) or 200 nM (round 3), while incubation of phage with bait was reduced to 1 h at 25 °C. Two washes (PBS-T and 50 mM Tris-HCl pH 7.5 + 0.5 M NaCl) in round 3 were extended to 10 min at 25 °C.

2.15.4. Initial gastrobody selection against GTD

Selections against GTD were performed with the original library as described in 'Gastrobody selection against CROP' and separately with a loop extended library. The loop extended library had the following features: a) SBTI had any combination of five, six or seven amino acids in GDR1/2 randomised with NNK codons, b) SBTI was fused to full length pIII and expression was induced with 0.2% (v/v) L-arabinose, and c) the library was produced in *E. coli* TG1 cells. The loop extended library contained 4.8 x 10^6 variants.

In round 1, 10^{10} naïve phage were incubated with 500 nM biotinylated AviTag-His_6-GTD in PBS-T + 1.5% (w/v) BSA for 2 h at 25 °C. Selection was performed in a 96-well cell-culture plate (655161, Greiner Bio) blocked with PBS pH 7.4 + 3% (w/v) BSA to reduce non-specific binding. Biotin Binder Dynabeads (Thermo Scientific) were incubated with phage-bait mixture at 25 °C for 1 h. Non-specifically bound phage were removed by four washes with PBS-T, one wash with 50 mM Tris-HCl pH 7.5 + 0.5 M NaCl, and two washes with PBS pH 7.4. Beads were incubated with 100 mM triethylamine pH 11 at 25 °C for 3 min to elute bound phage. The elution was neutralised by addition of 1 M Tris-HCl pH 7.5 and recovered by infecting log-phase *E. coli* TG1 cells as described above.

In parallel with the standard selection, I introduced protease pressure in round 2 and 3. In round 2, 10^8 phage amplified from round 1 were incubated with 0.1 mg/mL pepsin in 50 mM glycine-HCl pH 2.2 at 37 °C for 10 min. Digestion was stopped by addition of 2.5 M Tris-HCl pH 8.8. In round 3, pepsin-digested phage were then incubated with 10 mg/mL pancreatin from bovine pancreas (T1750, Merck) in 50 mM Tris-HCl pH 6.8 + 10 mM $CaCl_2$ at 37 °C for 10 min.

To remove proteases, phage were re-purified by precipitation. Phage were precipitated by addition of phage precipitation buffer [final concentration 4% (w/v) PEG8000 + 0.5 M NaCl] and incubating at 4 °C on ice for 1 h. Phage were pelleted by centrifugation for 30 min at 16,900 *g* and 4 °C. The pellet was washed twice in 4% (w/v) PEG8000 + 0.5 M NaCl before resuspending in PBS-T + 1.5% (w/v) BSA and continuing with the selection as for round 1 with the following modifications. In round 2 and 3, AviTag-His_6-GTD concentration was reduced to 250 nM. Three of the PBS-T washes were extended to 10 min at 25 °C. A negative selection step against the Biotin Binder Dynabeads was performed in round 3, as described for the selections against CROP above.

2.15.5. Affinity maturation selection of GT_S1_01

A new library with GT_S1_01 as template was cloned for affinity maturation. In the affinity maturation library, diversity was introduced into GDR1/2 separately (keeping one GDR sequence fixed). GDRs contained five, six or seven NNK codons. The affinity maturation library contained 3.6 x 10^9 variants. The selection procedure was modified from the initial panning protocol against GTD described above.

The concentration of AviTag-His_6-GTD was gradually reduced from 200 nM (round 1) to 100 nM (round 2) or 50 nM (round 3). Phage input was reduced from 10^{11} in round 1 and 2 to 10^{10} in round 3. Off-rate selection was driven by addition of excess (2 µM) non-biotinylated AviTag-His_6-GTD. Round 1 included one 20 min off-rate wash at 25 °C. Round 2 and 3 included two off-rate washes for 1 h (round 2) or 2 h (round 3) at 37 °C. The on-phase was also controlled by addition of excess non-biotinylated AviTag-His_6-GTD before capture on Biotin Binder beads. The on-phase was shortened

to 30 min in round 2 and 10 min in round 3 at 25 °C. As a separate selection campaign, pepsin pressure was included in round 2 and 3 as described above in the selections against CROP.

2.15.6. Affinity maturation selection of GT01

GDR3/4 were randomised with NNK codons for affinity maturation of GT01. The phagemid containing GT01 displayed on pIII (purified during sequencing of hit clones) was used as template. GDR3 contained four or five amino acids. GDR4 contained five, seven or nine amino acids. The library contained 3.3 x 10^8 variants.

The methodology for the second affinity maturation campaign was as described for the first affinity maturation campaign. Four rounds of selection were performed. In round 1, 2 and 3, 10^{10} phage were used as input and SBTI-pIII production was induced with 0.05% (w/v) L-arabinose. Phage input was reduced to 10^9 in round 4 and SBTI-pIII was induced with 0.02% (w/v) L-arabinose. The on-phase was 30 min in round 1 and 2, and shortened to 10 min in round 3 and 4. Biotinylated AviTag-His_6-GTD concentration was 50 nM (round 1), 25 nM (round 2), 15 nM (round 3) or 1 nM (round 4). Washes included one 1 h off-rate step at 37 °C in round 1 and 2, two 1 h off-rate steps at 37 °C in round 3, and three 1 h off-rate steps at 37 °C in round 4.

2.16. Phage retention on trypsin-agarose beads

10^{13} M13 or M13-SBTI in 50 mM Tris-HCl pH 8.0 + 0.05% (v/v) Tween-20 were incubated with trypsin-agarose beads (20230, Thermo Scientific) in microfuge spin columns for 20 min at 25 °C. Both phage preparations were produced in *E. coli*

ER2738. SBTI was fused to truncated pIII(216-425). SBTI-pIII(216-425) expression was induced by 0.42 mM IPTG. Trypsin-agarose beads were washed six times with 50 mM Tris-HCl pH 8.0 + 0.05% (v/v) Tween-20 before eluting bound phage with 50 mM glycine-HCl pH 2.2. Eluted phage were diluted 1:100 in deionised water and titered by qPCR as described in 'Phage purification'. The experiment was performed in triplicate.

2.17. SBTI retention on trypsin-agarose beads

Trypsin-agarose beads (20230, Thermo Scientific) were washed three times with 50 mM Tris-HCl pH 7.8 in microfuge spin columns. ss-SBTI-His_6-QuantumBlue was incubated with trypsin-agarose beads in 50 mM Tris-HCl pH 7.8 for 15 min at 25 °C in the spin columns. The liquid fraction was eluted and beads were washed three times with 50 mM Tris-HCl pH 7.8. Trypsin-agarose bound proteins were eluted in 200 mM glycine-HCl pH 2.2. Flow-through, wash and elution fractions were collected. Samples were boiled for 3 min at 99 °C in 1× SDS-loading buffer and run on SDS-PAGE, before Coomassie staining. A diluted sample representative of the input protein was included as control.

2.18. GTD inhibition assay

UDP-Glo glycosyltransferase assay kit (V6991, Promega) was used for detection of UDP-Glucose hydrolysis. The assay was performed in a black 96-well half area no-binding plate (3993, Corning). 1:2 serial dilutions of His_6-Thrombin site-SpyTag003-GT01 or His_6-Thrombin site-SpyTag003-SBTI (5 µM – 10 nM) were incubated with

500 nM non-biotinylated AviTag-His_6-GTD for 15 min at 25 °C in PBS pH 7.4. After addition of 25 µM UDP-Glucose, the reaction was incubated for 1 h at 25 °C. The reaction was stopped by addition of an equal volume of nucleotide detection reagent from the UDP-Glo glycosyltransferase assay kit (V6991, Promega) to the reaction mixture. The UDP released by GTD is converted into ATP by the nucleotide detection reagent. The stopped reaction was incubated for 1 h at 25 °C. Released ATP was used by luciferase in the nucleotide detection reagent to generate bioluminescence. A FLUOStar Omega (BMG Labtech) was used to record bioluminescence at 520 nM. The experiment was performed in triplicate. A sample without gastrobody was included as a positive control.

2.19. Trypsin inhibition assay

Inhibition of trypsin by SBTI was tested in the presence of 10 mM of the four most abundant bile acids in human bile (Hu et al., 2018). Each of the following bile acids was tested individually: sodium glycocholate hydrate (G7132, Merck), sodium glycodeoxycholate (G9910, Merck), sodium taurocholate hydrate (86339, Merck) or sodium taurodeoxycholate hydrate (T0875, Merck). 0, 0.042, 0.42 or 4.2 µM SBTI-His_6 was incubated with 80 U/mL trypsin from bovine pancreas (T1426, Merck) and 10 mM bile acid in 50 mM Tris-HCl pH 8.0 for 20 min at 25 °C. The colorimetric synthetic analogue N_α-Benzoyl-L-arginine-4-nitroanilide-hydrochloride (L-BAPA, B3133, Merck) was added to 32 µM and the reaction was incubated at 37 °C with shaking at 400 RPM (double orbital shaking). The L-BAPA stock was dissolved in dimethyl sulfoxide (DMSO). A_{405} was measured using a FLUOStar Omega plate reader (BMG Labtech) to monitor trypsin cleavage of L-BAPA. The experiment was performed in triplicate.

2.20. Western blot

2.20.1. Arabinose titration of pIII expression

M13-SBTI phage were produced and purified as described above, where SBTI-pIII(216-425) expression was induced with 0, 0.002, 0.02 or 0.2% (w/v) L-arabinose. M13-SBTI produced with 0.42 mM IPTG induction (pFAB5c phagemid) was used as control.

10^{12} phage were mixed with 1× SDS-loading buffer and boiled for 3 min at 99 °C before running on SDS-PAGE. Phage proteins were transferred onto a nitrocellulose membrane pre-soaked in tris-glycine transfer buffer [20% (v/v) methanol, 10 mM Tris base, 100 mM glycine] kept at 4 °C. Transfer was performed in a XCell II blot module (Thermo Scientific) for 2 h at 30 V at 25 °C. The nitrocellulose membrane was blocked for 16 h at 4 °C in PBS pH 7.4 + 5% (w/v) skim milk. pIII was blotted with 1:1,000 diluted anti-pIII (PSKAN3, MobiTec) in PBS-T + 5% (w/v) skim milk. The secondary antibody was 1:5,000 diluted anti-mouse IgG:HRP (A4416, Merck). The membrane was incubated with each antibody for 1 h at 25 °C. Between antibody incubations the membrane was washed four times with PBS-T. After four final washes, the blot was developed with SuperSignal West Pico (34577, Thermo Scientific) and imaged on a ChemiDoc XRS+ (Bio-Rad). A brightfield image was taken to align the ladder.

2.20.2. SBTI display on full-length pIII

M13 or M13-SBTI phage were produced as described above. Phage were produced in *E. coli* TG1 cells and expression of SBTI-pIII was induced with 0.2% (w/v) L-

arabinose. 10^{10} phage were mixed with 1× SDS-loading buffer and boiled for 3 min at 99 °C before separating on SDS-PAGE in duplicate. Phage proteins were transferred onto a nitrocellulose membrane (IB23001, Thermo Scientific). Transfer was performed with the iBlot2 Dry Blotting system (Thermo Scientific), running default program P0. The nitrocellulose membrane was blocked for 16 h at 4 °C in PBS pH 7.4 + 5% (w/v) skim milk. The membrane was cut in half and either blotted against SBTI or against pIII. The primary antibodies were 1:1,000 rabbit anti-trypsin inhibitor (1 mg/mL stock, 34549, Abcam) or mouse anti-pIII (0.5 mg/mL, stock PSKAN3, MobiTec) in PBS-T + 5% (w/v) skim milk. The secondary antibodies were 1:5,000 anti-rabbit IgG(H+L) HRP (1 mg/mL stock, 65-6120, Thermo Fisher) or anti-mouse IgG:HRP (A4416, Merck) in PBS-T + 5% (w/v) skim milk. The membranes were incubated with each antibody for 1 h at 25 °C. Between antibody incubations the membrane was washed four times with PBS-T. After four final washes, the blot was developed with SuperSignal West Pico (34577, Thermo Scientific) and imaged on a ChemiDoc XRS+ (Bio-Rad). A brightfield image was taken to align the ladder.

2.21. PDBeFold

PDBeFold (Krissinel and Henrick, 2004) was used to identify structural homologues of SBTI (PDB ID: 1AVU). Chain A (SBTI domain) of PDB ID 1AVU was queried against the whole PDB archive and sorted by the quality of alignment score (Q-score). The Q-score represents the quality of Cα-alignment, taking into account the length of the alignment and the root mean square deviation (RMSD) to correct for artificially high or low RMSD scores with short or long alignments. The expression for the Q-score is shown below:

$$Q\ score = \frac{N_{algn} \times N_{algn}}{(1 + (\frac{RMSD}{R_0})^2 \times N_{res1} \times N_{res2}}$$

where N_{algn} = number of aligned residues, R_0 is an empirical parameter set to 3Å, N_{res1} and N_{res2} is the number of residues in the aligned proteins, and RSMD is calculated from the matched Cα-atoms in Å. The results from the PDBeFold search were inspected manually for proteins originating in plants with protease inhibitor function and high Q-scores.

2.22. Structure representation

Protein structures were visualized with PyMOL 2.3.4 (Schrödinger). Structures shown are SBTI (PDB ID: 1AVU), SBTI in complex with porcine trypsin (PDB ID: 1AVX) (Sweet et al., 1974), Sso7d nanofitin (PDB ID: 1BNZ) (Gao et al., 1998), nanobody (PDB ID: 1QD0) (Spinelli et al., 2000), WBCI (PDB ID: 1EYL) (Ravichandran et al., 2001), ECTI (PDB ID: 1TIE) (Onesti et al., 1991), and BBKI (PDB ID: 4ZOT) (Zhou et al., 2015). The WT Sso7d (nanofitin) depicted was crystallized with DNA. The nanobody depicted lacks the second engineered disulfide bond. The nanobody binds to hapten azo-dye Reactive Red. Some structures of SBTI shown were generated during Rosetta simulations from PDB ID: 1AVU.

Chapter 3: Exploring scaffolds suitable for the GI tract

Chapter 3: Exploring scaffolds suitable for the GI tract

Oral administration of drugs is generally preferred to injections due to improved patient quality of life (Fu et al., 2009). Unlike orally administered drugs, injections also create hazardous needle waste (Ishtiaq et al., 2012; Olowokure et al., 2003). Low uptake of orally administered proteins into the bloodstream prevents systemic exposure which may lead to side effects (Bonovas et al., 2016; Lemaitre et al., 2017). The hostile conditions of the gastrointestinal (GI) tract, however, pose a strong barrier to oral administration of protein-based targeting reagents (Moroz et al., 2016).

In the gastric phase, orally administered proteins will encounter low pH and the aspartic protease pepsin. The median pH in the stomach of fasted young adults is 1.7 (Dressman et al., 1990), and concentrations of pepsin in the human stomach are 0.5 – 1 mg/mL (Brodkorb et al., 2019; Zhu et al., 2006). The intestinal phase features further digestive proteases including trypsin, chymotrypsin and elastase (Lee and Yamamoto, 1989; Mahato et al., 2003; Wen et al., 2002; Yadav et al., 2016). Proteins are also denatured by the presence of bile acids secreted from the pancreas (Calvert and Legg, 1977; Robic et al., 2011; Sarkany and Moreland, 1985).

Scaffold proteins are a promising alternative therapeutic class to antibodies (English et al., 2020; Steeland et al., 2016; Vazquez-Lombardi et al., 2015). Designed to overcome the limitations of conventional antibodies, scaffold proteins tend to be small, easy to express in microbial fermentation, thermostable and can be evolved *in vitro* (Vazquez-Lombardi et al., 2015). Scaffold proteins are rarely tested for their stability to GI tract conditions. Nanobodies are heavy-chain antibody variable domains from *Camelidae* (Hamers-Casterman et al., 1993; Hamers et al., 1993). Toward oral delivery of nanobodies, Hussack *et al.* engineered an anti-*Clostridium difficile* toxin A

nanobody to form a second disulfide bond to improve protease resistance (Hussack et al., 2011a, 2011b). Others have studied orally available nanobodies against enterotoxigenic *E. coli* (anti-F4 fimbriae) (Harmsen et al., 2005, 2006), *C. jejuni* (anti-flagella) (Hussack et al., 2014; Riazi et al., 2013) or tumour necrosis factor alpha (TNFα) (Crowe et al., 2018, 2019). Nanofitins are a class of protein scaffold derived from Sac7d, a DNA-binding protein from an acidophilic archaeon (Pecorari and Alzari, 2006). Nanofitins are stable at low pH (Béhar et al., 2014) and anti-TNFα nanofitins have been investigated for oral delivery (Labouebe et al., 2016).

Here I show that an anti-IgG nanofitin (Béhar et al., 2014) and an anti-*C. difficile* toxin A nanobody (Hussack et al., 2011a) that was developed for oral administration are rapidly digested by gastrointestinal proteases at concentrations that match proteolytic activity *in vivo.* However, I demonstrate that Kunitz-type soybean trypsin inhibitor (SBTI) is stable in the presence of pepsin and intestinal proteases (trypsin, chymotrypsin, elastase). Differential scanning calorimetry (DSC) and enzyme linked immunosorbent assay (ELISA) were used to analyse the thermostability of SBTI at neutral and acidic pH. SBTI is easy to express in *E. coli* shake flask culture and is a compatible fusion partner to the industrial feed enzyme phytase. I propose that SBTI is a promising candidate for development of a novel protein scaffold suited to the GI tract.

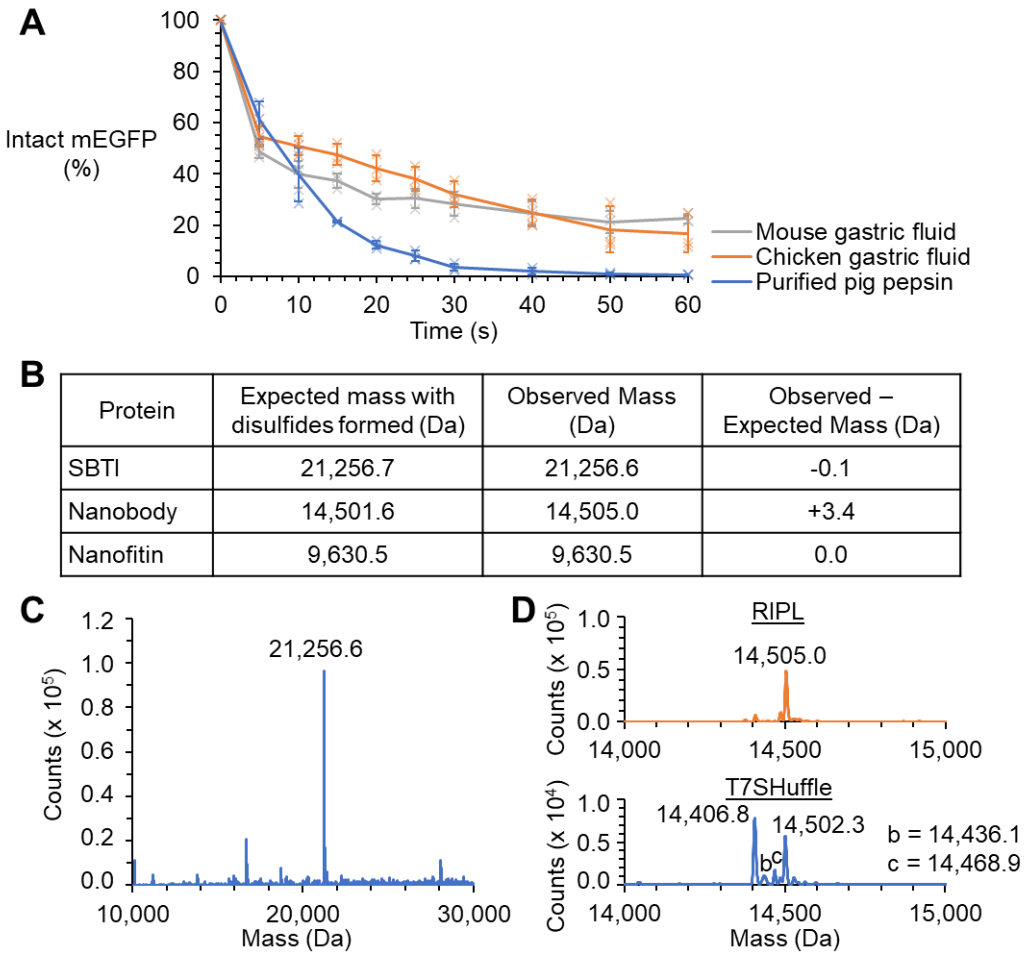
3.1. Proteolytic activity of mouse and chicken gastric fluid

Pepsin is a promiscuous protease (Hamuro et al., 2008) in the acidic environment of the human stomach, present at 0.5 – 1 mg/mL, where 1 mg comprises ≥ 2,500 U pepsin activity (Brodkorb et al., 2019; Zhu et al., 2006). Pepsin activity is measured by

digestion of haemoglobin and monitoring the increase in trichloroacetic acid soluble products, where one unit produces “ΔA280 of 0.001 per min at pH 2.0 at 37 °C” (Ryle, 1970). First, I sought to establish a standard gastric challenge against which to test protein scaffold candidates. I compared the digestion of monomeric enhanced green fluorescent protein (mEGFP) in mouse or chicken gastric fluid against 1 mg/mL purified pig pepsin (3,028 U/mL) at pH 2.2 (Fig. 3.1A). The digestion rate of mEGFP in the gastric fluids tested was comparable to the rate in 1 mg/mL pepsin (Fig. 3.1A). Therefore, I set 1 mg/mL (3,028 U/mL) pepsin at pH 2.2 at 37 °C as benchmark challenge that a protein scaffold should resist to transit or be active in the human stomach.

3.2. Gastrointestinal stability of scaffold proteins and SBTI

Non-antibody scaffold proteins are in clinical development for a range of applications including therapy and imaging (Vazquez-Lombardi et al., 2015), but the stability of these proteins in gastrointestinal conditions is usually not tested. Previous studies have reported SBTI to resist digestion by pepsin (Roychaudhuri et al., 2004; Wang et al., 2017), so I chose to test SBTI in my system alongside a nanofitin and nanobody developed for oral administration.



Protein	Expected mass with disulfides formed (Da)	Observed Mass (Da)	Observed – Expected Mass (Da)
SBTI	21,256.7	21,256.6	-0.1
Nanobody	14,501.6	14,505.0	+3.4
Nanofitin	9,630.5	9,630.5	0.0

Figure 3.1. Proteolytic activity in gastric fluid and protein scaffolds for oral delivery. (A) Proteolytic activity in mouse or chicken gastric fluid compared to porcine pepsin. mEGFP was incubated with mouse or chicken gastric fluid, or an equal volume of purified pig pepsin (1 mg/mL, 3,028 U/mL final concentration), at 37 °C. Digestion was stopped by neutralisation of the solution. Proteolysis activity was monitored by loss of fluorescence due to digestion of mEGFP (mean ±1 s.d., n = 3 with individual data-points as crosses). **(B)** Electrospray ionization mass spectrometry (ESI-MS) of candidate scaffold proteins. **(C)** Representative ESI-MS spectra of SBTI. **(D)** ESI-MS spectra of nanobody expressed in *E. coli* RIPL or T7SHuffle cells. Modification of cysteine to dehydroalanine corresponds to -34 Da change.

3.2.1. MS to show correct expression of SBTI, nanobody and nanofitin

I expressed SBTI, the nanobody and nanofitin in *E. coli* and purified using Ni-NTA. I chose a nanobody engineered for pepsin-resistance (A4.2m) (Hussack et al., 2011a) and an anti-IgG-Sso7d (nanofitin) known to be stable at low pH (Béhar et al., 2014). SBTI was expressed in *E. coli* T7SHuffle (Lobstein et al., 2012) to promote efficient

disulfide bond formation. The identity of the proteins was confirmed using electrospray ionisation mass spectrometry (ESI-MS) (Fig. 3.1B,C). SBTI formed both disulfide bonds (Fig. 3.1B,C). The nanobody expressed in *E. coli* BL21 (DE3) RIPL did not form the two expected disulfide bonds (Fig. 3.1B,D). Some species of the nanobody when produced in T7SHuffle formed two disulfide bonds (peak at 14,502.3 Da) (Fig. 3.1D). However, this preparation also included peaks which may correspond to species with one (peak at 14,468.9 Da) or two (peak at 14,436.1 Da) cysteines modified to dehydroalanine, and one peak I could not assign (peak at 14,406.8 Da) (Fig. 3.1D). Therefore, I continued with the nanobody produced in BL21 (DE3) RIPL cells which was more homogenous (Fig. 3.1D).

3.2.2. Pepsin stability of nanofitin, nanobody and SBTI

I compared the pepsin stability of the nanobody, nanofitin and SBTI (Fig. 3.2A). The nanofitin and nanobody were nearly undetectable after 10 min in the presence of 100-fold lower concentration of pepsin than found in the human stomach (Fig. 3.2A). SBTI, on the other hand, showed little to no signs of digestion after 10 min at pH 2.2 with 1 mg/mL pepsin, the benchmark pepsin concentration (Fig. 3.2A). Compared to the nanobody and nanofitin tested here, SBTI appears extremely stable in stomach-like conditions.

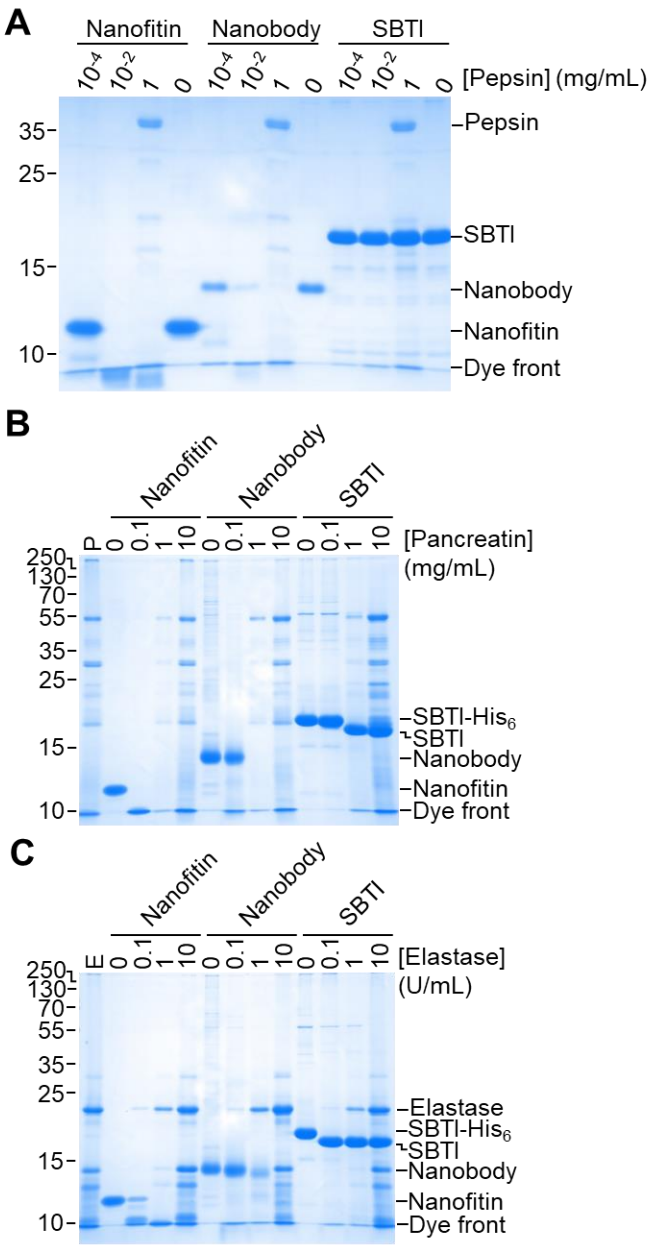


Figure 3.2. Stability of candidate scaffolds in gastric conditions. (A) Pepsin stability of scaffolds. SBTI, nanobody and nanofitin were incubated with serial dilutions of 1 mg/mL (3,028 U/mL) pepsin at pH 2.2 at 37 °C for 10 min. Digestion was stopped by boiling in SDS-loading dye. **(B)** Pancreatin stability of scaffolds. SBTI, nanobody and nanofitin were incubated with serial dilutions of pancreatin at pH 6.8 at 37 °C for 30 min. **(C)** Elastase stability of scaffolds. SBTI, nanobody and nanofitin were incubated with serial dilutions of elastase at pH 6.8 at 37 °C for 30 min. Digested proteins were separated on SDS-PAGE and stained with Coomassie.

3.2.3. Pancreatin stability of nanofitin, nanobody and SBTI

Pancreatin is a secretion of digestive enzymes from the exocrine cells of the pancreas (Lee and Yamamoto, 1989; Mahato et al., 2003; Wen et al., 2002). Thus, after the gastric phase, proteins will encounter further proteases from pancreatin in the small intestine including trypsin, chymotrypsin and elastase. As for pepsin, I tested the stability of the candidate scaffold proteins in serial dilutions of pancreatin where the highest concentration tested (10 mg/mL) represents the proteolytic activity in the small intestine. Some digestion of SBTI was observed in the presence of 10 mg/mL pancreatin, but the nanobody and nanofitin were not detectable by Coomassie staining after 30 min in the presence of 10-fold less pancreatin (Fig. 3.2B). The shift of the SBTI band at 1 and 10 mg/mL pancreatin is consistent with loss of the His_6-tag (Fig. 3.2B).

3.2.4. Elastase stability of nanofitin, nanobody and SBTI

Elastase inactivates the clinically approved anti-TNFα monoclonal antibodies Adalimumab and Infliximab (Yadav et al., 2016). 10 U/mL elastase is representative of its activity in the small intestine (Yadav et al., 2016). I observed minimal digestion of SBTI when incubated with 10 U/mL elastase for 30 min (Fig. 3.2C). The nanofitin and nanobody were fully digested at intestinal concentrations of elastase (Fig. 3.2C). Taken together, in my tests, SBTI appears to be more stable to gastric (Fig. 3.2A) and intestinal (Fig. 3.2B,C) proteases than the nanobody and nanofitin.

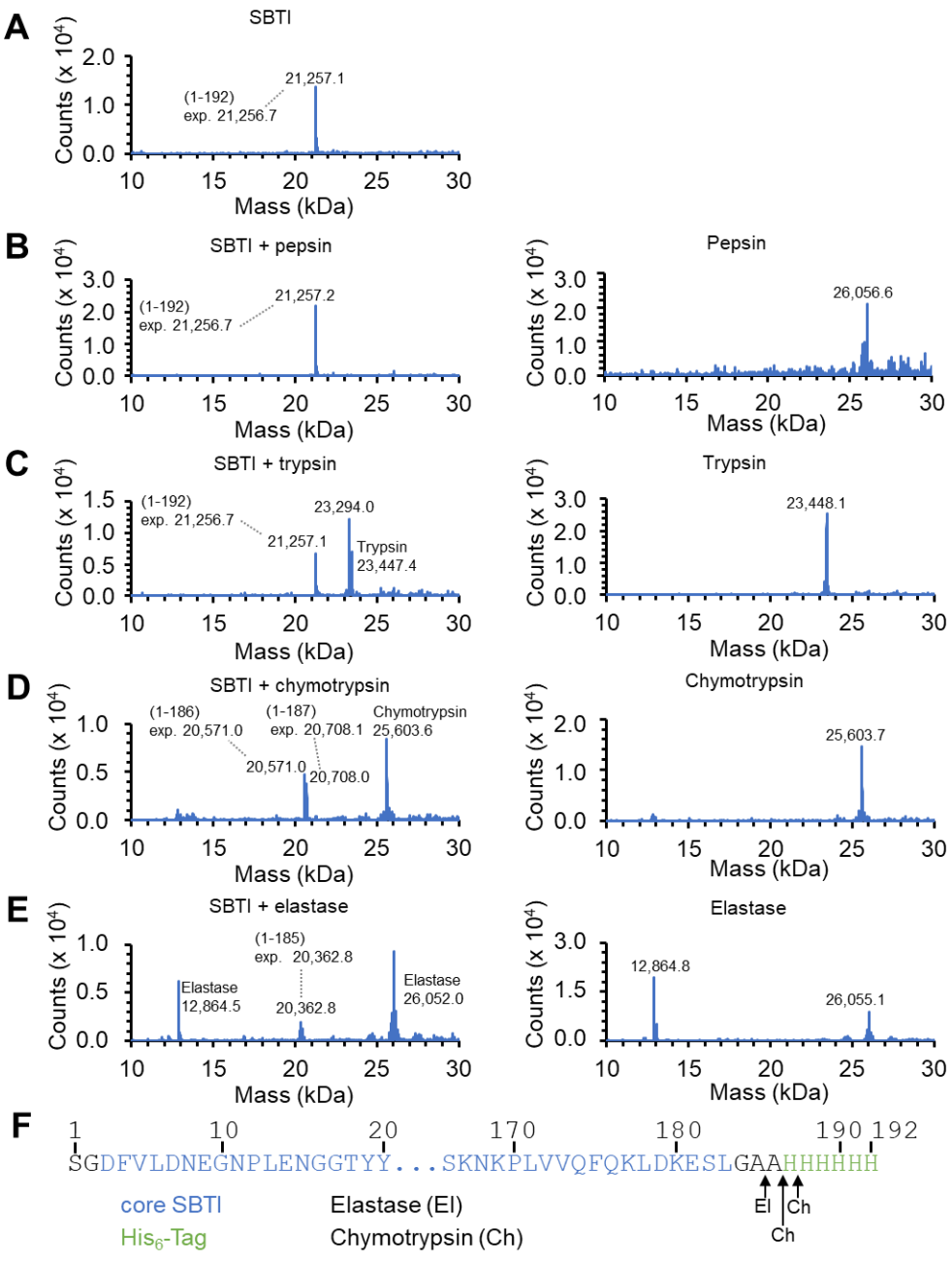


Figure 3.3. Protease cut-sites in SBTI. SBTI-His6 **(A)** construct was incubated with 1 mg/mL pepsin **(B)**, 100 U/mL trypsin **(C)**, 25 U/mL chymotrypsin **(D)** or 10 U/mL elastase **(E)** for 15 min at 37 °C. Digestion products were analysed by ESI-MS. Expected masses of construct fragments are marked. Control spectra without SBTI are shown in the right-hand column. **(F)** The observed cut sites in the amino acid sequence of the WT SBTI construct are marked with arrows. I performed sample preparation and data analysis. Anthony Tumber loaded and ran the mass spectrometer.

3.2.5. Protease cleavage sites in SBTI

SDS-PAGE based experiments do not directly inform where proteases in SBTI are cutting. I digested SBTI with gastrointestinal concentrations of proteases before analysing the products with ESI-MS to identify potential cleavage sites (Fig. 3.3). Purified proteases were used rather than pancreatin to allow assignment of cleavage sites to each protease. Pepsin (Fig. 3.3B) and trypsin (Fig. 3.3C) did not cut SBTI-His_6. Consistent with the shift seen on SDS-PAGE gels, chymotrypsin (Fig. 3.3D) and elastase (Fig. 3.3E) removed the C-terminal His_6-tag of the construct (Fig. 3.3F). I did not identify cleavage sites in the SBTI domain.

3.2.6. SBTI inhibits trypsin in the presence of bile acids

Bile acids potentiate the activity of digestive proteases and can promote denaturation of proteins (Calvert and Legg, 1977; Gass et al., 2007). In humans, following a meal the total free bile acid concentrations in the upper ileum reach 10 mM and 2 mM in the lower ileum (Northfield and McColl, 1973). To test the resilience of SBTI to bile acids, I pre-incubated SBTI with 10 mM of the four most common bile acids in human bile (Hu et al., 2018), before measuring the trypsin inhibition activity of SBTI (Fig. 3.4). Trypsin activity increased in the presence of bile acids, but there was no clear difference in inhibition of trypsin activity by SBTI in the presence or absence of bile acids (Fig. 3.4).

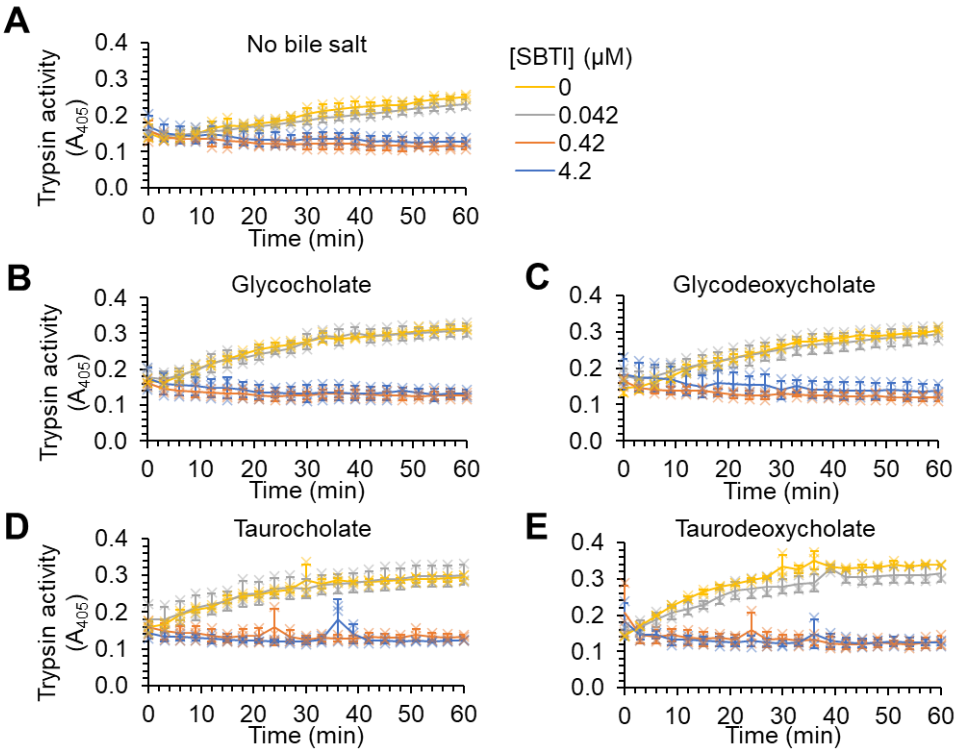


Figure 3.4. SBTI stability in the presence of bile acids. Trypsin inhibition activity of SBTI was tested without bile salts **(A)** or in the presence of 10 mM glycocholate **(B)**, glycodeoxycholate **(C)**, taurocholate **(D)**, or taurodeoxycholate **(E)** at pH 8.0 at 37 °C. Increase in A_{405} due to cleavage of a chromogenic substrate was used to monitor trypsin activity (mean ±1 s.d., n = 3 with individual data-points as crosses).

3.3. SBTI is a thermo-stable protein

Thermostable proteins provide a robust starting point for evolution or engineering (Bloom et al., 2006). Resilience to high temperatures is also important for certain applications, like animal feed supplementation because the proteins need to withstand the steam sterilisation process (Rebello et al., 2017; Woyengo and Nyachoti, 2011). Similarly, thermal stability is important for withstanding stresses during fill/finish steps in biologics manufacturing (Rathore and Rajan, 2008). For example, heat may be used as part of viral clearance steps of recombinantly produced therapeutic proteins (Cipriano et al., 2012).

3.3.1. Temperature dependent solubility of SBTI

Aggregation is a common mechanism for loss of protein after heating. I tested the thermal resilience of SBTI by heating to 25, 75, 90 or 100 °C for 10 min before cooling and pelleting aggregates by centrifugation. The remaining soluble fraction was analysed by SDS-PAGE coupled with gel densitometry (Fig. 3.5A,B). I used β-lactamase (BLA) as a mesophilic control protein (Schoene et al., 2014). >90% of SBTI remained soluble after heating to 100 °C compared to 25 °C (Fig. 3.5A,B). The majority of BLA (~90%), however, was aggregated after heating to 75 °C (Fig. 3.5A).

3.3.2. Binding activity of SBTI after heating

Despite remaining soluble, SBTI may misfold during heating which could compromise binding affinity or functionality. I heated SBTI to 37, 55, 75 or 100 °C for 10 min before measuring binding to trypsin in an enzyme-linked immunosorbent assay (ELISA) to test functionality after exposure to high temperatures (Fig. 3.5C). Compared to 37 °C, there was an approximately three-fold loss of bound protein after heating to 100 °C. However, heating to 55 or 75 °C resulted in minimal loss of bound protein compared to 37 °C (Fig. 3.5C).

3.3.3. Thermal stability of SBTI at low pH

Protonation of amino acids in acidic conditions leads to intramolecular repulsion, promoting denaturation of proteins (Fink et al., 1994). Hence, digestion of proteins in the stomach is facilitated by acid. I compared the stability of SBTI at neutral and gastric

pH using DSC (Fig. 3.5D). The melting temperature (T_m) of SBTI decreased by 6.9 °C from 67.2 °C at pH 7.4 to 60.3 °C at pH 2.0 (Fig. 3.5D) indicating that SBTI retains good stability at a pH characteristic of the gastric environment.

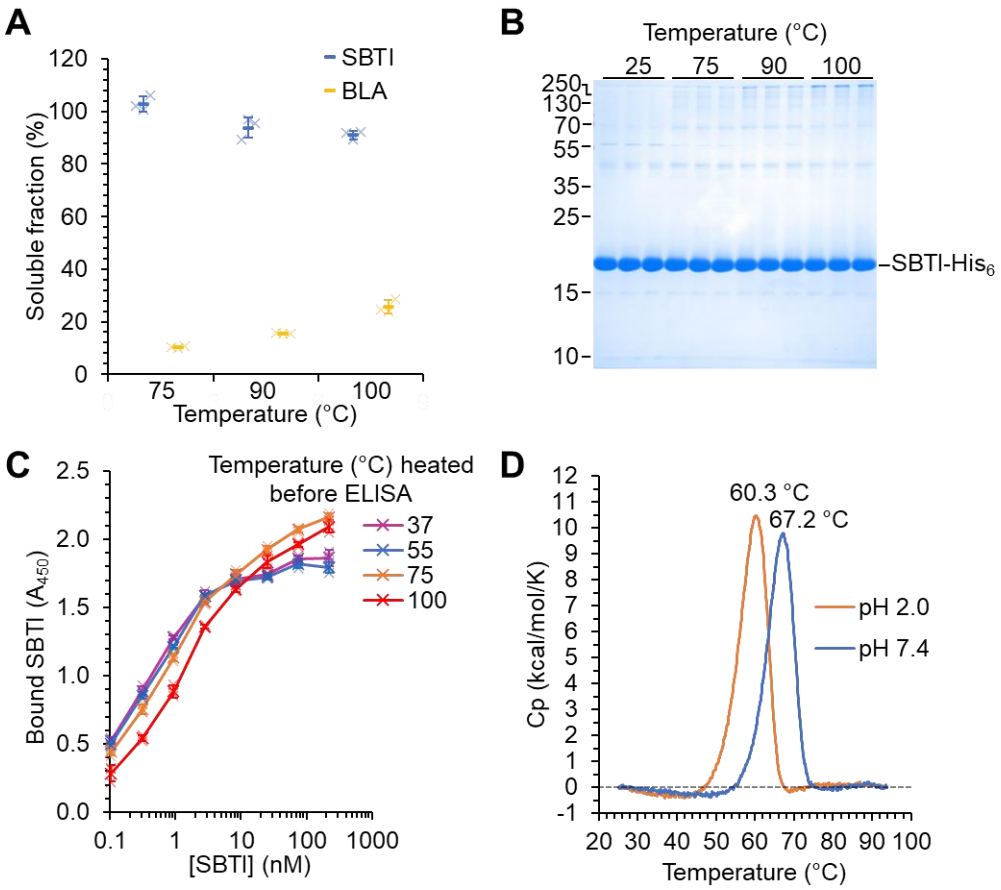


Figure 3.5. Temperature stability of SBTI. (A) Temperature dependent solubility of SBTI. SBTI-His_6 or BLA was heated to 75, 90 or 100 °C for 10 min before cooling and pelleting aggregates by centrifugation. The soluble fraction was run on SDS-PAGE and analysed by gel densitometry. Samples kept at 25 °C were set to 100% (mean ±1 s.d., n = 3 with individual data-points as crosses). **(B)** Soluble fraction of SBTI after heating to indicated temperatures for 10 min. Soluble protein was run on SDS-PAGE and stained with Coomassie. Band intensity is quantified in (A). **(C)** Functional test of binding after heating. SBTI was heated to 37, 55, 75 or 100 °C for 10 min before serial dilution and measuring binding to immobilised trypsin in an ELISA (mean ±1 s.d., n = 3 with individual data-points as crosses). **(D)** Temperature stability at pH 7.4 and pH 2.0. DSC was performed in phosphate buffer at pH 7.4 or pH 2.0. The melting temperature at each pH is marked.

3.3.4. SBTI is a compatible fusion partner to phytase

Enzymes are a common additive in livestock feed to improve animal growth and health (Choct, 2006). Phytase removes phosphate groups from phytic acid (*myo*-inositol hexakisphopshate), the predominant storage form of phosphate in plant-based feed, and from lower inositol-phosphate species (Wodzinski and Ullah, 1996). Phytase improves growth of non-ruminants because these monogastric animals lack the enzymes to digest phytase, thereby rendering the phosphate in phytic acid non-bioavailable (Dersjant-Li et al., 2014; Onyango et al., 2005; Singh, 2008; Sommerfeld et al., 2018; Walk et al., 2014). Research into 'superdosing' phytase, where higher doses are administered than is currently standard, suggests further benefits can be achieved but higher doses may not be economically viable (Lee et al., 2018; Walk et al., 2014). Targeting and retaining feed enzymes like phytase to their site of action might increase efficacy. It is important for targeting domains to have no detrimental effect on the normal function of the fusion partner.

3.3.5. MS shows SBTI-QB can be expressed in E. coli

I expressed a genetic fusion of SBTI to Quantum Blue (SBTI-QB), a leading commercial phytase enzyme developed by AB Vista (Nguyen and Winter, 2006). The SBTI and QB domains were joined with a $(GS)_3$-His_6-GSG linker. I expressed SBTI-QB in *E. coli* BL21 (DE3) RIPL with a periplasmic signal sequence and purified the His-tagged fusion protein by osmotic shock with Ni-NTA. The yield of SBTI-QB was 2.9 mg/L culture. I confirmed the identity of SBTI-QB with ESI-MS (Fig. 3.6A). SBTI has two disulfide bonds and QB has four. The difference between the observed and expected mass (+1.3 Da) suggests that all or all but one disulfide bond were formed.

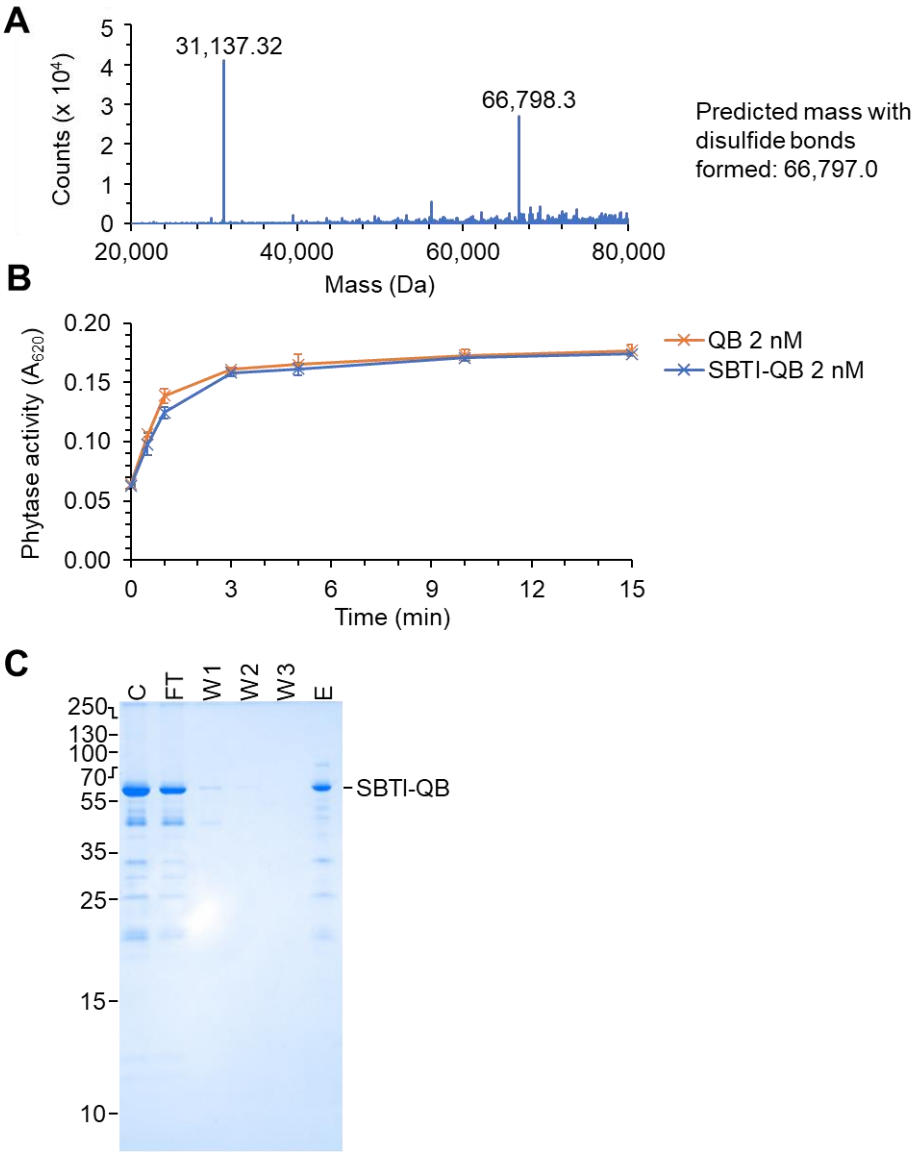


Figure 3.6. SBTI is a compatible fusion partner with phytase. (A) ESI-MS of SBTI-QB. **(B)** Phytase activity of SBIT-QB compared to QB. 2 nM of phytase constructs were incubated with 5 µM phytic acid at pH 4.5. Release of phosphate was measured with a molybdate/malachite green-based reagent at A_{620} (mean ±1 s.d., n = 3). **(C)** Retention of SBTI-QB on trypsin-agarose resin. SBTI-QB was incubated with trypsin-agarose resin at pH 7.8 for 15 min at 25 °C before washing three times followed by elution at pH 2.2. Fractions collected at each step were run on SDS-PAGE and stained with Coomassie. Stock (control, C), flow-through (FT), washes (W1 – W3) and elution (E) fractions.

3.3.6. SBTI fusion to QB does not compromise QB phytase activity

Next, I compared the phytase activity of SBTI-QB to that of QB alone (Fig. 3.6B). 2 nM of each protein was incubated with phytic acid and release of free phosphate

monitored. There was no substantial difference in phytase activity between SBTI-QB and QB (Fig. 3.6B).

3.3.7. SBTI-QB is retained on trypsin-agarose resin

I incubated SBTI-QB with trypsin-agarose resin to test the functionality of the SBTI domain in the fusion protein. Following a pre-incubation step, SBTI-QB was retained on trypsin-agarose during washes at pH 7.8 but eluted at pH 2.2, which denatures the trypsin (Fig. 3.6C). Taken together, both the SBTI and QB domains retain some functionality when expressed as a fusion protein, but further experiments are required to confirm full functionality of the construct.

3.4. Discussion

Current protein targeting reagents are rapidly inactivated in the GI tract, especially in the stomach (Moroz et al., 2016). Generalisable strategies for improving the stability of protein scaffolds in the stomach or transiting the stomach include formulations to protect the protein cargo from the gastric environment or neutralising the stomach pH (Anselmo et al., 2018). Formulations complicate the pharmacokinetics of drugs. Neutralising the stomach pH could lead to a range of complications including higher susceptibility to infection or indigestion (Tariq et al., 2017). A different strategy could be to develop a protein scaffold that is itself stable in the stomach and wider GI tract. Toward that goal, I have shown that SBTI is stable in the presence of physiological concentrations of gastrointestinal proteases. SBTI maintains effective inhibition of trypsin in the presence of bile acids.

First, I set a benchmark test for scaffolds that mimics the proteolytic activity in the stomach by comparing the digestion of mEGFP in gastric fluid with a known concentration of pepsin (1 mg/mL or 3,028 U/mL). Degradation of SBTI was minimal after 10 min in the presence of 1 mg/mL pepsin at pH 2.2, while the nanofitin and nanobody were completely digested. The surprising susceptibility of the nanofitin and nanobody to pepsin relative to SBTI might be explained by the strength of the proteolytic pressure used in past studies. Previous experiments exploring the pepsin stability of the nanobody used 46 U/mL pepsin at pH 2.0 for 37 °C for 1 hour (Hussack et al., 2011a). Simulated gastric fluid containing 0.1 mg/mL pepsin, but with unknown pepsin activity, showed stability of a nanofitin (Labouebe et al., 2016). Subsequent efforts to engineer pepsin stability into the backbone of nanofitins were not successful (Loussouarn et al., 2020). These findings highlight the importance of a well-defined *in vitro* model that reflects the conditions *in vivo*. More sophisticated models like INFOGEST 2.0 (Brodkorb et al., 2019) or SHIME® (Van de Wiele et al., 2015) would further improve the quality of the data but are more labour-intensive to set up. It stands to reason that the inclusion of other proteins competing for digestion, as would be the case after a meal, would further affect the rate of protein digestion. Similarly, other foods particles or mucus are likely modulate digestion (Wen et al., 2002).

Protease inhibitors accumulate in plant tissues to deter consumption by large herbivores or act as defence against insects (Murdock and Shade, 2002). The protease resistance of SBTI has likely evolved to protect the soybean from digestion in the GI tract. Pepsin exhibits a poorly defined cleavage specificity (Hamuro et al., 2008), so it is difficult to define why SBTI is particularly resistant to digestion by pepsin. Initial tests for protease resistance were based on SDS-PAGE, where a small shift of the SBTI occurred in the presence of pancreatin and elastase. I performed ESI-MS

experiments to confirm that SBTI was not being cut internally by gastrointestinal proteases. A weakness of the ESI-MS approach here is that it did not identify internal cleavage sites which could be used to engineer protease resistance. Samples were centrifuged to pellet aggregates before loading the soluble fraction on the mass spectrometer. Fragments of SBTI produced from internal cleavage sites would likely have been removed in this process. In addition, a single internal cleavage event might lead to unfolding of the protein and higher probability of subsequent cleavage events of that molecule. This “all or nothing” pathway for protease-mediated protein degradation was first proposed in 1950 by Linderstrøm-Lang (Linderstrøm-Lang, 1950). A more controlled approach with lower protease concentrations (limited proteolysis) might help identify the first cleavage site.

The two disulfide bridges in SBTI might contribute to tight packing of the structure (Roychaudhuri et al., 2004). Tightly packed loops likely present a worse substrate for proteases because of steric clashes with the active site and lower propensity for local unfolding (Hubbard, 1998). The nanobody construct in this study had previously been engineered to contain an additional disulfide bond to improve pepsin resistance (Hussack et al., 2011a). In the original work by Hussack *et al.* (2011a), the formation of the disulfide bonds was confirmed by tryptic MS. I was unable to produce a homogeneous preparation where the nanobody forms completely the two expected disulfide bonds. Expression of the nanobody in T7SHuffle cells engineered to promote disulfide bond formation in the cytosol (Lobstein et al., 2012), did not yield a homogeneous preparation. Hussack *et al.* (2011a) showed approximately 80% of the nanobody remained after 1 hour in the presence of 46 U/mL pepsin at pH 2.0. I observed some residual nanobody after 10 min in the presence of 30 U/mL pepsin at pH 2.2, indicating lower stability than expected. Nevertheless, these conditions are

100-fold below the pepsin concentration found in the stomach. Anti-TNFα nanobodies developed by a different group are orally administered with a commercially available enteric coating (Crowe et al., 2019) or gastro-protective formulation consisting of sodium bicarbonate and milk (Crowe et al., 2018). Lastly, Hussack *et al.* (2011a) reported that the increase in pepsin resistance of the engineered nanobody was a trade-off for reduced the trypsin and chymotrypsin resistance.

DSC analysis revealed that SBTI remains thermostable at pH 2.0 with a T_m substantially higher than physiological temperatures (60.3 °C). Further experiments like circular dichroism could add confidence that SBTI remains well folded in gastric conditions. Paradoxically, SBTI was soluble following heating to 100 °C for 10 min, despite DSC showing unfolding of SBTI at 65 °C. The preference for refolding into a soluble protein rather than aggregation on cooling, however, has been observed for a variety of proteins (Schoene et al., 2014, 2016). Analysis of SBTI following heating using an ELISA showed that heating to temperatures (e.g. 75 or 100 °C) above the T_m (67.2 °C) reduced the proportion of functional protein. Indeed, heating SBTI to temperatures below the T_m (e.g. 55 °C) did not reduce the proportion of SBTI binding to trypsin compared to samples kept at 37 °C.

Chapter 4: Engineering SBTI to become a gastrobody

Chapter 4: Engineering SBTI to become a gastrobody

12% of all healthcare-associated infections and 71% of gastrointestinal infections in the U.S. are caused by the Gram-positive pathogen *Clostridium difficile* (Magill et al., 2014). *C. difficile* toxin A (TcdA) and toxin B (TcdB) are large multi-domain proteins that enter intestinal epithelial cells and disrupt the cytoskeleton (Chen et al., 2016; Kuehne et al., 2010; Poutanen and Simor, 2004). Disruption of the cytoskeleton compromises the epithelial barrier enabling pathogen invasion. Antibodies generated by passive immunisation with TcdA/B protect against *C. difficile* challenge (Babcock et al., 2006; Kink and Williams, 1998; Torres et al., 1995). In fact, Actoxumab (anti-TcdA) and bezlotoxumab (anti-TcdB) were evaluated for protection against recurring *C. difficile* infection in phase III clinical trials (Wilcox et al., 2017). The fully human neutralising antibody bezlotoxumab was subsequently approved by the U.S. Food and Drug Administration (FDA). Bezlotoxumab targets the combined repetitive oligopeptide domain (CROP) of TcdB that is involved in epithelial cell binding (Babcock et al., 2006; Orth et al., 2014; Yang et al., 2015). Actoxumab and bezlotoxumab are administered intravenously (Wilcox et al., 2017). Given that *C. difficile* toxins initially act from the intestinal lumen, there is scope for oral administration of neutralising molecules. Indeed, neutralizing nanobodies against TcdA/B have been developed to enable oral administration (Hussack et al., 2011a, 2011b, 2018), though the efficacy *in vivo* is unknown.

Antibody mimetics provide a backbone for engineering binders to novel targets. Non-antibody protein scaffolds benefit from ease of expression and high stability (Gebauer and Skerra, 2020). It is important that scaffolds tolerate mutation of variable regions while keeping the stability of the parent protein. Building on the results presented in

Chapter 3, I have proposed SBTI as the basis for a new protein scaffold, because SBTI is stable in gastrointestinal conditions (Foster et al., 2013; Recoules et al., 2017; Wang et al., 2017), unlike leading protein scaffolds.

SBTI belongs to the structural family of β-trefoil proteins (McLachlan, 1979; Song and Suh, 1998; Sweet et al., 1974). Although some inward-facing hydrophobic residues are conserved, β-trefoil proteins share low sequence identity despite striking structural homology (Guerra et al., 2016; Murzin et al., 1992). I hypothesize that a lack of sequence identity, but high structural homology, is indicative of a robust scaffold for evolution and engineering. In this chapter, I engineer SBTI to become an evolvable protein scaffold, termed a gastrobody. Computational prediction of mutability is complemented by experimental alanine scanning, to define a surface-accessible binding surface which we call gastrobody determining region (GDR). I chose TcdA combined repetitive oligopeptide domain (CROP) (Hussack et al., 2011b) as model gastrobody target. Gastrobody binders to CROP were evolved by phage display.

4.1. Rosetta to predict evolvability

The goal of this work was to design a binding surface comprising more than one evolvable region, to enable targeting of more diverse epitope types than a single peptide-like loop (Diem et al., 2014). Mutation of residues in the binding surface should not compromise pepsin stability of the protein. The cleavage pattern of pepsin is difficult to define by the linear amino acid sequence (Hamuro et al., 2008). I reasoned that preserving a rigid, well-folded structure is important for retaining the protease resistance of SBTI. Hence, I used the Rosetta protein modelling suite (Alford et al., 2017; Kuhlman et al., 2003; Leaver-Fay et al., 2011) to guide the selection of mutable regions in SBTI.

4.1.1. Generating an ensemble of structures with relax

First, I used the *remodel* function (Huang et al., 2011) to model missing density in the crystal structure of SBTI (PDB ID: 1AVU). Analysing a pool of SBTI structures will improve confidence in later Rosetta calculations. I used *relax* (Conway et al., 2014) to generate an ensemble of SBTI structures (Fig. 4.1A). After an initial 500 instances of *relax*, I chose the lowest energy structure as the starting point for a second 500 instances. When plotting the root mean square deviation (RMSD) against Rosetta energy units (REU) the structures produced by the second instance of *relax* converged (Fig. 4.1A). 467 structures were included in an ensemble that was taken forward for analysis.

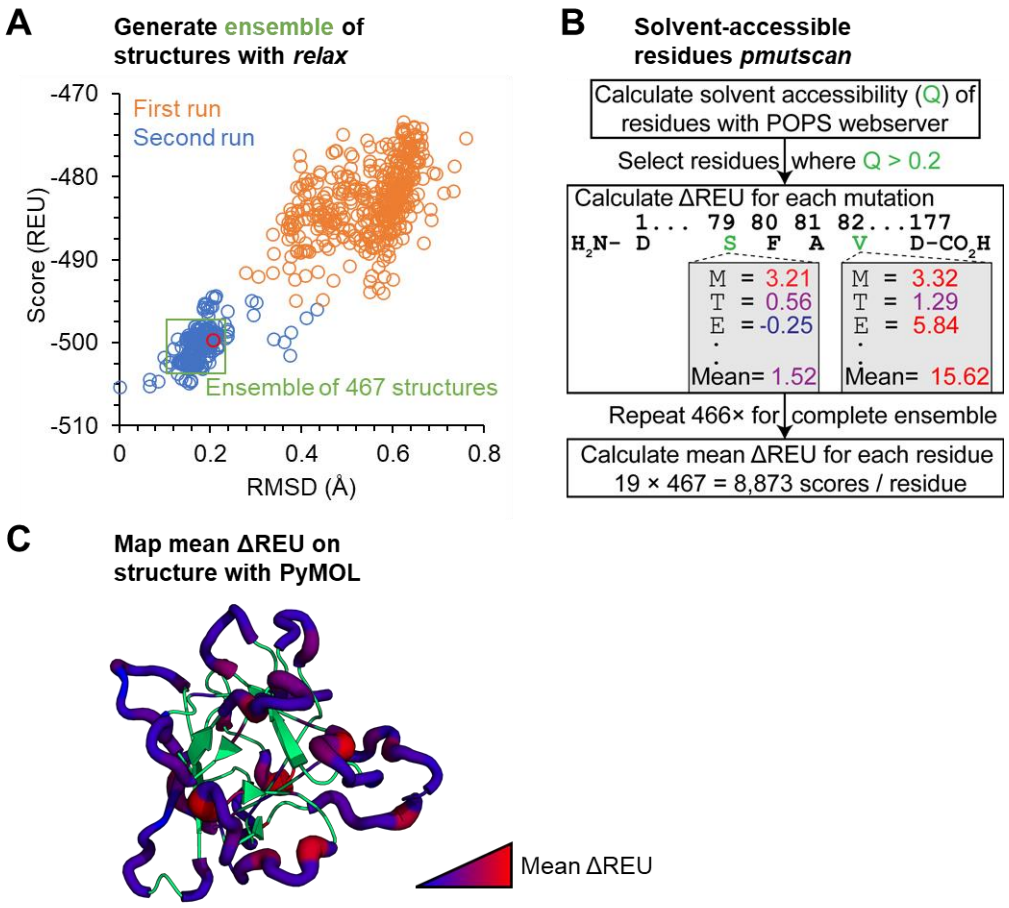


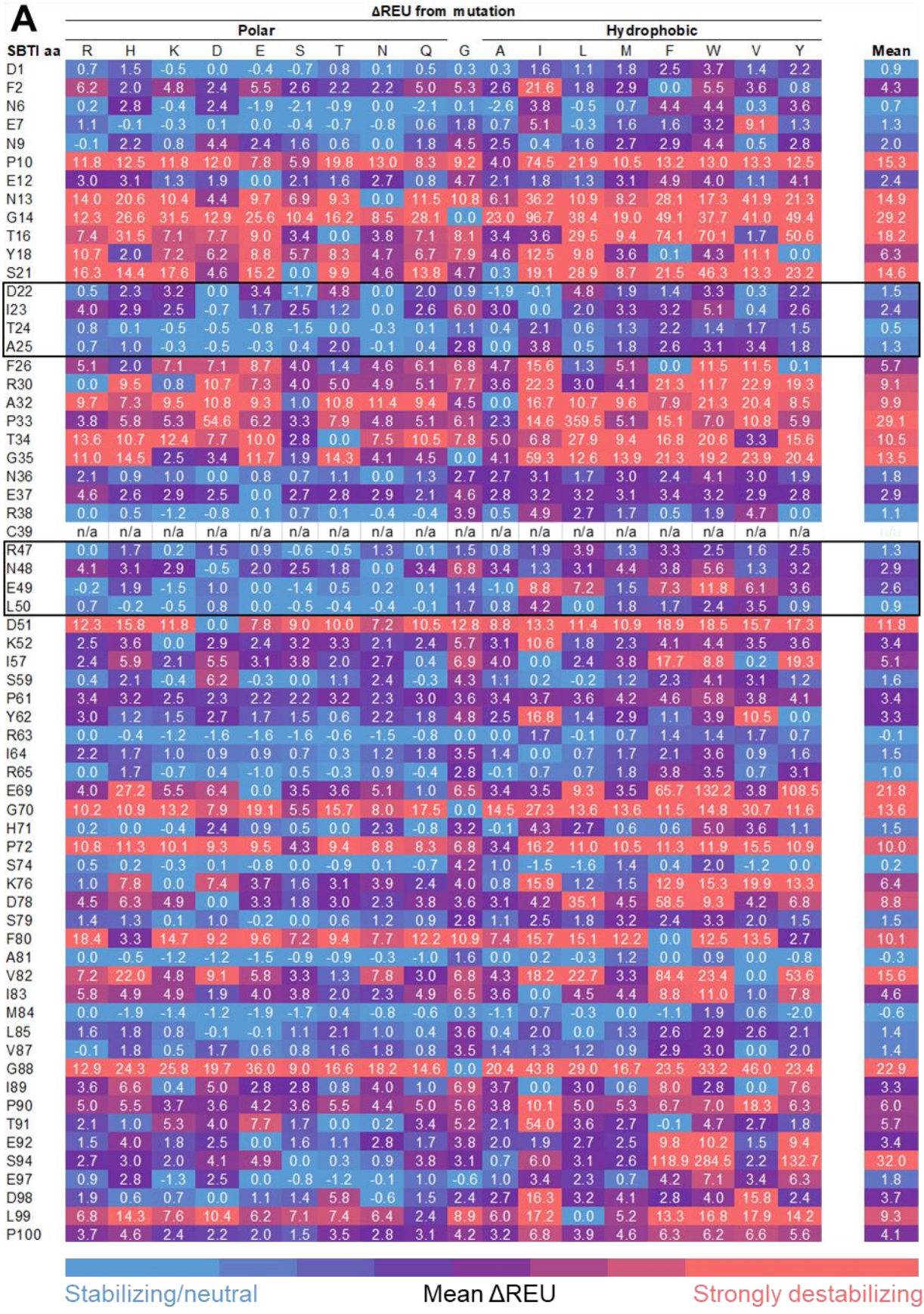
Figure 4.1. Rosetta-guided prediction of evolvable regions in SBTI. (A) Rosetta *relax* to create an ensemble of SBTI structures. *Remodel* was used to model missing density in the crystal structure of SBTI (PDB ID: 1AVU). Two iterations of *relax* with 500 instances were run consecutively to produce a convergent ensemble of structures (green rectangle; -503 < REU < -497 and 0.119 Å < RMSD < 0.238 Å). The lowest energy structure of the first run (red) was the starting point for the second run. **(B)** Rosetta *pmutscan* to predict surface-exposed mutable residues in SBTI. *pmutscan* was performed on surface-accessible residues (excluding cysteines) in each structure in the ensemble. **(C)** Visualising stability changes. Mean ΔREU of each residue (excluding cysteine and proline) was mapped onto the SBTI PyMOL structure using putty format. Low ΔREU values (blue) are stabilising, large ΔREU values (red) are destabilising. Non surface-accessible residues are represented in green in cartoon format because they were excluded from mutability calculations. The underlying structure was generated from PDB ID: 1AVU during Rosetta calculations.

4.1.2. pmutscan of surface-accessible residues

The Parameter Optimised Surfaces (POPS) webserver (Cavallo et al., 2003) was used to calculate the solvent-accessibility of SBTI residues. Residues were defined as

surface-accessible if the quotient of surface accessible area and surface area of isolated atoms (Q) was greater than 0.2 (Cavallo et al., 2003). Only residues defined as surface-accessible were analysed in subsequent steps because the mutable regions should bind other proteins.

Rather than modelling large new loops, I focused on the impact of single mutations to reduce the computational power required. The impact on folding energy of mutating residues with Q > 0.2 (except cysteine) was analysed using *pmutscan* (Kuhlman et al., 2003; Leaver-Fay et al., 2011). I mutated all surface-accessible residues to all natural amino acids (cysteine and proline) *in silico* (Fig. 4.1B). I excluded proline because it resulted in large destabilising ΔREU values. Cysteine was excluded to avoid disrupting native disulfide bond formation or creating the potential for dimerization. Mean ΔREU was calculated for each individual mutation from the scores of each structure in the ensemble. The mean ΔREU scores for each mutation of a specific residue (e.g. A25) were then averaged to generate a general mutability score for that residue. I visualized the results of the *pmutscan* results by mapping the change in Rosetta energy for mutations (mean ΔREU) at each residue on the structure of SBTI with PyMOL putty (Fig. 4.1C). The mean ΔREU for each mutation and general mutability score at each residue is tabulated in Fig. 4.2A.


A
ΔREU from mutation
Polar
Hydrophobic
SBTI aa R H K D E S T N Q G A I L M F W V Y Mean
D1 0.7 1.5 -0.5 0.0 -0.4 -0.7 0.8 0.1 0.5 0.3 0.3 1.6 1.1 1.8 2.5 3.7 1.4 2.2 0.9
F2 6.2 2.0 4.8 2.4 5.5 2.6 2.2 2.2 5.0 5.3 2.6 21.6 1.8 2.9 0.0 5.5 3.6 0.8 4.3
N6 0.2 2.8 -0.4 2.4 -1.9 -2.1 -0.9 0.0 -2.1 0.1 -2.6 3.8 -0.5 0.7 4.4 4.4 0.3 3.6 0.7
E7 1.1 -0.1 -0.3 0.1 0.0 -0.4 -0.7 -0.8 0.6 1.8 0.7 5.1 -0.3 1.6 1.6 3.2 9.1 1.3 1.3
N9 -0.1 2.2 0.8 4.4 2.4 1.6 0.6 0.0 1.8 4.5 2.5 0.4 1.6 2.7 2.9 4.4 0.5 2.8 2.0
P10 11.8 12.5 11.8 12.0 7.8 5.9 19.8 13.0 8.3 9.2 4.0 74.5 21.9 10.5 13.2 13.0 13.3 12.5 15.3
E12 3.0 3.1 1.3 1.9 0.0 2.1 1.6 2.7 0.8 4.7 2.1 1.8 1.3 3.1 4.9 4.0 1.1 4.1 2.4
N13 14.0 20.6 10.4 4.4 9.7 6.9 9.3 0.0 11.5 10.8 6.1 36.2 10.9 8.2 28.1 17.3 41.9 21.3 14.9
G14 12.3 26.6 31.5 12.9 25.6 10.4 16.2 8.5 28.1 0.0 23.0 96.7 38.4 19.0 49.1 37.7 41.0 49.4 29.2
T16 7.4 31.5 7.1 7.7 9.0 3.4 0.0 3.8 7.1 8.1 3.4 3.6 29.5 9.4 74.1 70.1 1.7 50.6 18.2
Y18 10.7 2.0 7.2 6.2 8.8 5.7 8.3 4.7 6.7 7.9 4.6 12.5 9.8 3.6 0.1 4.3 11.1 0.0 6.3
S21 16.3 14.4 17.6 4.6 15.2 0.0 9.9 4.6 13.8 4.7 0.3 19.1 28.9 8.7 21.5 46.3 13.3 23.2 14.6
D22 0.5 2.3 3.2 0.0 3.4 -1.7 4.8 0.0 2.0 0.9 -1.9 -0.1 4.8 1.9 1.4 3.3 0.3 2.2 1.5
I23 4.0 2.9 2.5 -0.7 1.7 2.5 1.2 0.0 2.6 6.0 3.0 0.0 2.0 3.3 3.2 5.1 0.4 2.6 2.4
T24 0.8 0.1 -0.5 -0.5 -0.8 -1.5 0.0 -0.3 0.1 1.1 0.4 2.1 0.6 1.3 2.2 1.4 1.7 1.5 0.5
A25 0.7 1.0 -0.3 -0.5 -0.3 0.4 2.0 -0.1 0.4 2.8 0.0 3.8 0.5 1.8 2.6 3.1 3.4 1.8 1.3
GDR1
F26 5.1 2.0 7.1 7.1 8.7 4.0 1.4 4.6 6.1 6.8 4.7 15.6 1.3 5.1 0.0 11.5 11.5 0.1 5.7
R30 0.0 9.5 0.8 10.7 7.3 4.0 5.0 4.9 5.1 7.7 3.6 22.3 3.0 4.1 21.3 11.7 22.9 19.3 9.1
A32 9.7 7.3 9.5 10.8 9.3 1.0 10.8 11.4 9.4 4.5 0.0 16.7 10.7 9.6 7.9 21.3 20.4 8.5 9.9
P33 3.8 5.8 5.3 54.6 6.2 3.3 7.9 4.8 5.1 6.1 2.3 14.6 359.5 5.1 15.1 7.0 10.8 5.9 29.1
T34 13.6 10.7 12.4 7.7 10.0 2.8 0.0 7.5 10.5 7.8 5.0 6.8 27.9 9.4 16.8 20.6 3.3 15.6 10.5
G35 11.0 14.5 2.5 3.4 11.7 1.9 14.3 4.1 4.5 0.0 4.1 59.3 12.6 13.9 21.3 19.2 23.9 20.4 13.5
N36 2.1 0.9 1.0 0.0 0.8 0.7 1.1 0.0 1.3 2.7 2.7 3.1 1.7 3.0 2.4 4.1 3.0 1.9 1.8
E37 4.6 2.6 2.9 2.5 0.0 2.7 2.8 2.9 2.1 4.6 2.8 3.2 3.2 3.1 3.4 3.2 2.9 2.8 2.9
R38 0.0 0.5 -1.2 -0.8 0.1 0.7 0.1 -0.4 -0.4 3.9 0.5 4.9 2.7 1.7 0.5 1.9 4.7 0.0 1.1
C39 n/a n/a n/a n/a n/a n/a n/a n/a n/a n/a n/a n/a n/a n/a n/a n/a n/a n/a
R47 0.0 1.7 0.2 1.5 0.9 -0.6 -0.5 1.3 0.1 1.5 0.8 1.9 3.9 1.3 3.3 2.5 1.6 2.5 1.3
N48 4.1 3.1 2.9 -0.5 2.0 2.5 1.8 0.0 3.4 6.8 3.4 1.3 3.1 4.4 3.8 5.6 1.3 3.2 2.9
E49 -0.2 1.9 -1.5 1.0 0.0 -1.4 0.5 0.2 0.1 1.4 -1.0 8.8 7.2 1.5 7.3 11.8 6.1 3.6 2.6
L50 0.7 -0.2 -0.5 0.8 0.0 -0.5 -0.4 -0.4 -0.1 1.7 0.8 4.2 0.0 1.8 1.7 2.4 3.5 0.9 0.9
GDR2
D51 12.3 15.8 11.8 0.0 7.8 9.0 10.0 7.2 10.5 12.8 8.8 13.3 11.4 10.9 18.9 18.5 15.7 17.3 11.8
K52 2.5 3.6 0.0 2.9 2.4 3.2 3.3 2.1 2.4 5.7 3.1 10.6 1.8 2.3 4.1 4.4 3.5 3.6 3.4
I57 2.4 5.9 2.1 5.5 3.1 3.8 2.0 2.7 0.4 6.9 4.0 0.0 2.4 3.8 17.7 8.8 0.2 19.3 5.1
S59 0.4 2.1 -0.4 6.2 -0.3 0.0 1.1 2.4 -0.3 4.3 1.1 0.2 -0.2 1.2 2.3 4.1 3.1 1.2 1.6
P61 3.4 3.2 2.5 2.3 2.2 2.2 3.2 2.3 3.0 3.6 3.4 3.7 3.6 4.2 4.6 5.8 3.8 4.1 3.4
Y62 3.0 1.2 1.5 2.7 1.7 1.5 0.6 2.2 1.8 4.8 2.5 16.8 1.4 2.9 1.1 3.9 10.5 0.0 3.3
R63 0.0 -0.4 -1.2 -1.6 -1.6 -1.6 -0.6 -1.5 -0.8 0.0 0.0 1.7 -0.1 0.7 1.4 1.4 1.7 0.7 -0.1
I64 2.2 1.7 1.0 0.9 0.9 0.7 0.3 1.2 1.8 3.5 1.4 0.0 0.7 1.7 2.1 3.6 0.9 1.6 1.5
R65 0.0 1.7 -0.7 0.4 -1.0 0.5 -0.3 0.9 -0.4 2.8 -0.1 0.7 0.7 1.8 3.8 3.5 0.7 3.1 1.0
E69 4.0 27.2 5.5 6.4 0.0 3.5 3.6 5.1 1.0 6.5 3.4 3.5 9.3 3.5 65.7 132.2 3.8 108.5 21.8
G70 10.2 10.9 13.2 7.9 19.1 5.5 15.7 8.0 17.5 0.0 14.5 27.3 13.6 13.6 11.5 14.8 30.7 11.6 13.6
H71 0.2 0.0 -0.4 2.4 0.9 0.5 0.0 2.3 -0.8 3.2 -0.1 4.3 2.7 0.6 0.6 5.0 3.6 1.1 1.5
P72 10.8 11.3 10.1 9.3 9.5 4.3 9.4 8.8 8.3 6.8 3.4 16.2 11.0 10.5 11.3 11.9 15.5 10.9 10.0
S74 0.5 0.2 -0.3 0.1 -0.8 0.0 -0.9 0.1 -0.7 4.2 1.0 -1.5 -1.6 1.4 0.4 2.0 -1.2 0.0 0.2
K76 1.0 7.8 0.0 7.4 3.7 1.6 3.1 3.9 2.4 4.0 0.8 15.9 1.2 1.5 12.9 15.3 19.9 13.3 6.4
D78 4.5 6.3 4.9 0.0 3.3 1.8 3.0 2.3 3.8 3.6 3.1 4.2 35.1 4.5 58.5 9.3 4.2 6.8 8.8
S79 1.4 1.3 0.1 1.0 -0.2 0.0 0.6 1.2 0.9 2.8 1.1 2.5 1.8 3.2 2.4 3.3 2.0 1.5 1.5
F80 18.4 3.3 14.7 9.2 9.6 7.2 9.4 7.7 12.2 10.9 7.4 15.7 15.1 12.2 0.0 12.5 13.5 2.7 10.1
A81 0.0 -0.5 -1.2 -1.2 -1.5 -0.9 -0.9 -0.3 -1.0 1.6 0.0 0.2 -0.3 1.2 0.0 0.9 0.0 -0.8 -0.3
V82 7.2 22.0 4.8 9.1 5.8 3.3 1.3 7.8 3.0 6.8 4.3 18.2 22.7 3.3 84.4 23.4 0.0 53.6 15.6
I83 5.8 4.9 4.9 1.9 4.0 3.8 2.0 2.3 4.9 6.5 3.6 0.0 4.5 4.4 8.8 11.0 1.0 7.8 4.6
M84 0.0 -1.9 -1.4 -1.2 -1.9 -1.7 0.4 -0.8 -0.6 0.3 -1.1 0.7 -0.3 0.0 -1.1 1.9 0.6 -2.0 -0.6
L85 1.6 1.8 0.8 -0.1 -0.1 1.1 2.1 1.0 0.4 3.6 0.4 2.0 0.0 1.3 2.6 2.9 2.6 2.1 1.4
V87 -0.1 1.8 0.5 1.7 0.6 0.8 1.6 1.8 0.8 3.5 1.4 1.3 1.2 0.9 2.9 3.0 0.0 2.0 1.4
G88 12.9 24.3 25.8 19.7 36.0 9.0 16.6 18.2 14.6 0.0 20.4 43.8 29.0 16.7 23.5 33.2 46.0 23.4 22.9
I89 3.6 6.6 0.4 5.0 2.8 2.8 0.8 4.0 1.0 6.9 3.7 0.0 3.0 0.6 8.0 2.8 0.0 7.6 3.3
P90 5.0 5.5 3.7 3.6 4.2 3.6 5.5 4.4 5.0 5.6 3.8 10.1 5.0 5.3 6.7 7.0 18.3 6.3 6.0
T91 2.1 1.0 5.3 4.0 7.7 1.7 0.0 0.2 3.4 5.2 2.1 54.0 3.6 2.7 -0.1 4.7 2.7 1.8 5.7
E92 1.5 4.0 1.8 2.5 0.0 1.6 1.1 2.8 1.7 3.8 2.0 1.9 2.7 2.5 9.8 10.2 1.5 9.4 3.4
S94 2.7 3.0 2.0 4.1 4.9 0.0 0.3 0.9 3.8 3.1 0.7 6.0 3.1 2.6 118.9 284.5 2.2 132.7 32.0
E97 0.9 2.8 -1.3 2.5 0.0 -0.8 -1.2 -0.1 1.0 -0.6 1.0 3.4 2.3 0.7 4.2 7.1 3.4 6.3 1.8
D98 1.9 0.6 0.7 0.0 1.1 1.4 5.8 -0.6 1.5 2.4 2.7 16.3 3.2 4.1 2.8 4.0 15.8 2.4 3.7
L99 6.8 14.3 7.6 10.4 6.2 7.1 7.4 6.4 2.4 8.9 6.0 17.2 0.0 5.2 13.3 16.8 17.9 14.2 9.3
P100 3.7 4.6 2.4 2.2 2.0 1.5 3.5 2.8 3.1 4.2 3.2 6.8 3.9 4.6 6.3 6.2 6.6 5.6 4.1
Stabilizing/neutral
Mean ΔREU
Strongly destabilizing

	ΔREU from mutation																		
	Polar										Hydrophobic								
SBTI aa	R	H	K	D	E	S	T	N	Q	G	A	I	L	M	F	W	V	Y	Mean
E101	2.4	0.7	2.0	1.6	0.0	1.7	1.7	0.5	0.3	3.9	2.7	3.9	2.4	3.9	2.3	5.4	2.2	2.2	2.2
G102	19.7	19.0	17.7	20.7	21.5	18.1	20.0	20.8	17.7	0.0	17.6	118.0	32.6	21.7	20.7	22.4	131.3	19.9	31.1
P103	10.7	15.2	13.4	10.9	12.5	5.0	25.0	10.8	12.1	6.7	3.8	49.8	13.1	13.0	15.1	14.2	72.8	11.5	17.5
K106	2.7	4.5	0.0	7.3	6.0	5.3	3.0	5.3	3.3	8.0	4.4	16.2	5.2	3.5	39.4	15.3	2.3	38.8	9.5
G108	206.7	190.0	207.8	169.0	111.8	26.4	176.8	144.6	105.9	0.0	24.9	247.2	190.0	90.3	230.5	404.1	208.0	260.7	166.4
E109	2.1	3.2	1.3	1.2	0.0	1.8	2.4	1.1	1.7	5.5	2.7	1.5	2.2	2.5	3.6	4.7	1.1	3.2	2.3
N110	19.1	33.6	31.6	13.2	10.5	3.7	7.5	0.0	8.2	10.2	4.3	16.1	7.0	15.4	83.0	144.7	12.5	46.3	25.9
K111	0.7	2.8	0.0	1.6	2.4	2.3	2.3	2.7	2.8	4.1	2.9	4.2	3.6	4.8	5.4	6.0	4.0	4.9	3.2
D112	2.8	1.8	1.6	0.0	3.2	3.5	5.6	1.6	3.3	4.8	4.1	11.5	2.2	3.7	2.8	4.5	11.9	2.8	4.0
A113	2.4	4.6	2.3	17.6	3.6	1.6	5.8	7.0	3.1	3.2	0.0	54.5	92.5	3.7	24.6	55.5	29.2	31.3	19.0
M114	4.6	6.8	3.6	4.2	4.2	0.9	7.1	3.6	1.6	3.5	1.0	23.1	9.4	0.0	9.0	8.1	15.8	9.0	6.4
D115	1.7	1.9	0.4	0.0	0.3	0.4	2.0	1.1	1.5	2.9	0.3	4.4	1.2	2.2	3.3	3.5	4.6	2.9	1.9
G116	195.4	105.3	175.5	100.4	165.2	90.5	121.3	91.3	191.6	0.0	85.2	678.0	97.7	201.4	136.6	284.6	369.4	164.1	180.7
W117	3.7	1.8	4.7	6.5	4.9	2.2	3.5	3.6	4.1	5.3	3.4	7.3	6.9	3.4	-0.5	0.0	4.4	-1.0	3.6
E121	7.4	8.8	6.8	6.0	0.0	5.6	5.4	6.0	4.8	9.8	6.2	3.6	3.2	5.4	12.7	12.0	3.6	11.9	6.6
R122	0.0	3.9	1.3	8.9	4.5	2.0	4.3	4.4	1.8	4.5	2.3	4.3	6.3	2.9	59.6	4.2	11.4	11.3	7.7
V123	5.7	5.4	5.1	5.7	4.2	5.1	4.6	5.4	3.4	8.7	4.7	27.8	12.7	5.4	5.0	8.5	0.0	3.7	6.7
S124	0.5	0.6	-0.5	-1.4	-0.5	0.0	1.9	-0.9	-0.7	1.3	-0.1	3.3	0.3	0.6	1.8	3.0	2.9	1.6	0.8
D125	0.9	0.5	0.1	0.0	0.9	0.2	1.9	0.2	0.4	7.8	2.2	3.5	0.5	3.2	2.1	3.1	6.2	2.1	2.0
D126	10.5	9.0	1.0	0.0	10.3	-1.1	13.3	0.8	1.0	-1.3	0.7	38.4	14.1	10.7	14.4	13.3	31.7	13.5	10.0
E127	0.3	-0.5	-0.1	-0.8	0.0	1.5	1.1	-1.2	0.6	5.4	1.4	1.9	0.9	3.4	0.5	3.7	1.2	0.7	1.1
F128	0.8	0.5	-1.7	4.5	0.8	-0.4	-2.0	0.3	-0.2	2.2	0.7	12.9	-0.6	1.4	0.0	6.0	-1.6	0.8	1.3
N129	4.6	3.4	12.5	4.8	6.5	4.4	3.2	0.0	3.6	6.0	3.6	21.2	44.7	4.4	4.1	7.4	12.6	10.4	8.7
K132	3.7	26.2	0.0	13.4	9.6	6.7	7.8	9.1	5.6	9.4	6.8	21.8	15.6	7.1	72.6	50.2	24.1	71.9	20.1
P137	5.1	11.6	5.3	9.0	6.9	4.7	7.3	8.3	5.7	6.9	3.2	11.1	7.9	7.4	14.8	16.9	11.5	15.0	8.8
Q138	0.4	1.8	-0.6	2.1	0.2	1.1	4.2	1.3	0.0	2.5	-1.1	3.0	1.4	1.0	2.6	3.4	3.6	2.0	1.6
Q139	0.9	1.0	-0.3	-0.5	-0.8	0.2	0.6	-0.3	0.0	2.1	1.2	1.9	0.3	1.8	3.6	2.2	1.6	5.3	1.1
E141	1.5	2.1	0.4	1.6	0.0	0.7	2.1	1.2	1.0	2.5	1.1	3.6	1.7	2.0	3.0	2.5	3.2	2.4	1.8
D142	3.6	3.2	2.4	0.0	1.6	2.0	2.8	2.1	2.3	3.7	2.9	5.5	3.2	4.4	4.7	5.6	4.6	4.0	3.3
D143	3.2	2.4	2.0	0.0	2.9	3.2	7.4	1.0	2.3	2.2	3.8	17.0	3.7	4.6	4.2	5.2	16.1	3.6	4.7
K144	0.2	1.1	0.0	2.3	2.0	0.1	0.1	1.2	0.0	2.8	1.0	2.0	1.1	2.4	1.7	3.4	2.0	1.2	1.4
C145	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
G146	0.0	-1.6	-1.6	-1.6	-2.0	-0.7	0.1	-1.9	-1.3	0.0	-0.8	13.4	-2.4	-2.3	-1.8	-1.6	0.3	-2.1	-0.4
D147	12.7	7.5	13.1	0.0	5.0	5.7	8.5	4.7	9.7	8.6	5.3	10.4	8.6	7.1	11.7	8.8	11.8	13.6	8.5
S151	1.0	1.5	0.5	2.3	1.0	0.0	1.8	1.7	-0.1	5.2	1.3	1.7	-0.2	0.3	1.9	4.0	2.5	0.9	1.5
I152	1.6	9.1	0.7	2.5	0.3	1.9	0.4	2.9	1.0	4.4	2.5	0.0	2.3	2.9	14.9	14.1	-0.4	14.6	4.2
D153	11.8	19.0	13.6	0.0	9.0	3.1	6.2	2.3	9.1	4.3	2.1	49.3	8.7	7.0	42.1	39.0	100.2	44.3	20.6
H154	-0.6	0.0	0.1	0.7	0.7	0.6	0.7	0.9	1.5	3.0	1.8	0.9	1.0	3.3	2.2	4.5	1.0	2.1	1.4
D155	3.3	2.9	2.3	0.0	1.2	2.7	3.2	2.2	2.4	5.0	3.5	3.6	2.6	4.5	3.7	4.6	3.3	3.2	3.0
D156	5.0	11.8	4.9	0.0	5.5	4.2	3.5	3.4	6.2	8.6	6.5	14.5	19.3	7.2	23.2	17.6	14.9	22.5	9.9
G157	10.5	9.0	7.4	6.0	13.0	5.6	15.1	6.0	10.2	0.0	12.4	25.3	11.4	14.1	10.6	12.6	28.2	10.1	11.5
T158	2.6	7.1	2.9	3.1	1.9	1.2	0.0	3.8	3.0	4.4	0.1	9.9	8.7	-0.3	5.5	5.2	9.6	4.9	4.1
R160	0.0	21.7	0.5	5.9	2.9	3.8	2.0	4.0	1.3	6.4	2.9	2.2	12.2	2.2	58.6	51.5	2.2	51.2	12.9
V163	5.0	2.7	5.5	3.9	2.0	3.7	3.3	3.5	2.6	9.6	4.4	-1.5	0.1	5.8	4.0	4.4	0.0	3.6	3.5
S164	10.2	7.0	7.0	2.2	10.9	0.0	-0.1	2.6	8.1	3.1	3.9	20.1	12.7	11.2	7.7	7.1	8.1	8.2	7.2
K165	1.1	1.4	0.0	-0.3	-0.6	0.4	0.7	0.5	0.7	1.8	0.5	1.5	1.3	2.7	2.4	2.5	1.5	1.8	1.1
N166	3.4	2.6	2.6	2.3	2.9	2.5	4.6	0.0	2.5	4.1	2.0	4.4	2.4	4.2	3.8	3.9	4.2	3.7	3.1
K167	1.3	3.8	0.0	8.0	3.0	2.2	7.0	3.2	1.3	2.3	-7.2	42.5	-2.0	3.5	4.0	11.4	35.1	3.3	6.8
P168	-7.4	-6.8	-8.7	-9.4	-9.0	-8.8	-4.9	-8.6	-7.5	-6.2	-8.5	-3.2	-7.8	-6.3	-5.5	-7.0	-4.1	-6.4	-7.0
V170	2.1	2.4	1.4	1.4	2.9	3.0	1.8	1.3	3.7	6.9	3.1	0.3	2.4	4.1	5.0	3.7	0.0	5.4	2.8
Q172	0.9	10.1	2.5	7.5	2.9	2.4	0.6	3.9	0.0	3.3	3.2	-0.5	2.0	2.6	43.4	17.8	-1.0	49.3	8.4
Q174	3.5	2.5	3.3	4.6	-0.9	1.9	2.8	3.9	0.0	7.2	2.9	6.9	1.7	2.4	4.2	6.8	2.2	3.2	3.3
K175	1.4	6.4	0.0	9.0	4.9	4.8	5.6	6.4	4.3	7.0	3.9	5.0	3.9	4.3	6.0	8.5	5.0	5.4	5.1
L176	1.5	2.7	0.9	4.1	1.0	1.2	2.5	2.5	0.3	6.5	1.4	23.1	0.0	1.9	3.3	5.7	0.4	2.5	3.4
D177	4.1	5.0	2.9	0.0	2.0	2.2	3.6	2.7	3.4	2.9	3.0	4.9	4.6	4.7	6.7	8.4	4.2	5.7	3.9

Stabilizing/neutral — Mean ΔREU — Strongly destabilizing

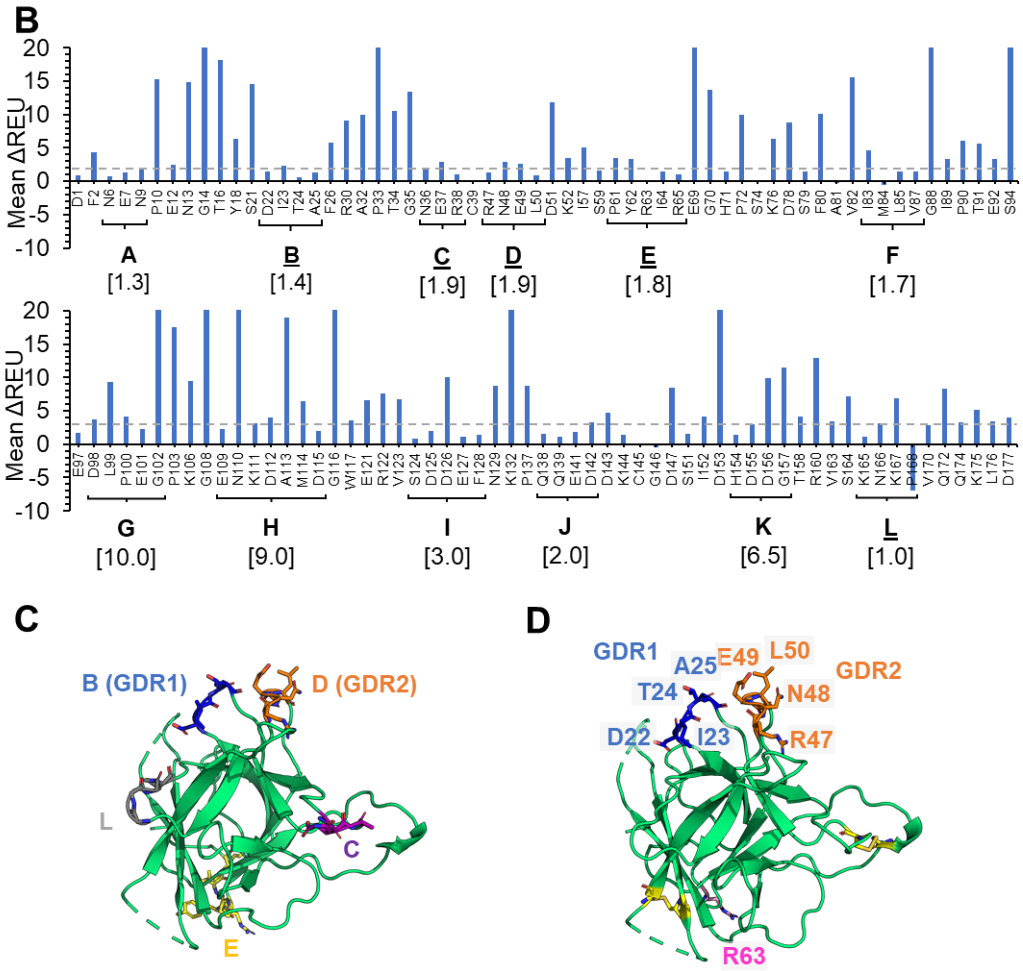


Figure 4.2. Designing a binding surface on SBTI. (A) Tabulated results of *pmutscan* results. Mean ΔREU values for each mutation originating from the ensemble structures are shown. The resulting mean ΔREU at each SBTI residue is shown in the right-hand column. Stabilising or neutral ΔREU values are coloured blue, destabilizing values are coloured red. **(B)** Candidate evolvable loops in SBTI. Candidate loops are marked alphabetically. Average of ΔREU of residues in a loop is shown in square brackets. Loops are underlined if two criteria are met: i) average ΔREU of the loop is < 2, and ii) the loop comprises ≥3 consecutive residues. **(C)** Assessing three-dimensional proximity of candidate loops. Loops meeting the criteria set out in (B) were mapped onto SBTI in PyMOL (PDB ID: 1AVU). **(D)** Gastrobody determining regions (GDRs). Two loops close in space (GDR1, GDR2) were chosen as the binding surface. The scissile R63 is marked in pink. Disulfide bonds are shown in yellow.

4.1.3. Choosing a versatile binding surface

Together with visualizing the mean ΔREU in PyMOL (Fig. 4.1C), I plotted the general mutability score on a bar chart (Fig. 4.2B). The mutability of loops was assessed by

averaging the mutability scores of the residues comprising each loop. Some loops were missing scores because not all residues in the loop were surface-accessible. I selected loops as candidates for the binding surface if they consisted of at least three consecutive residues with an average ΔREU less than 2 (Fig. 4.2B). Candidate loops close in three-dimensional space were identified from manual inspection in PyMOL (Fig. 4.2C). I chose two four-amino acid loops close in space, termed gastrobody determining regions (GDR), as binding surface (Fig. 4.2C,D). GDR1 consists of D22, I23, T24 and A25. GDR2 consists of R47, N48, E49 and L50.

4.2. Experimental validation of Rosetta predictions

4.2.1. MS confirms expression of alanine mutants

I sought to validate the computational results experimentally by using alanine scanning mutagenesis. Each residue in the GDRs was mutated to alanine (except A25, which was mutated to glycine). I expressed the alanine mutants in *E. coli* T7SHuffle with a C-terminal His_6 tag. Yields of the alanine mutants ranged from 3.9 to 9 mg /L culture. Correct expression of each protein and formation of the disulfide bond was confirmed by ESI-MS (Fig. 4.3A). All alanine mutants expressed successfully and formed the two expected disulfide bonds (Fig. 4.3A).

4.2.2. Temperature dependent solubility of alanine mutants

Heat resilience of the alanine mutants was tested with the temperature dependent solubility assay as described in Chapter 3. Alanine mutants, WT SBTI and BLA were heated to 25, 75, 90 or 100 °C for 10 min. BLA was included as a representative

thermolabile mesophilic protein (Schoene et al., 2014). Aggregated protein was removed by centrifugation and the remaining soluble fraction was analysed by SDS-PAGE (Fig. 4.3B). I did not observe catastrophic loss of heat resilience of any alanine mutant, compared to WT SBTI. The most thermolabile alanine mutants were T24A and D22A, but substantial fractions of D22A (32%) and T24A (44%) remained soluble after heating to 100 °C (Fig. 4.3B).

4.2.3. Pepsin resistance of alanine mutants

Next, I tested the pepsin resistance of the alanine mutants. Alanine mutants and WT SBTI were incubated with 1 mg/mL pepsin at pH 2.2 and 37 °C. Intact protein was detected by SDS-PAGE and Coomassie staining (Fig. 4.3C). In line with the temperature resilience, no SBTI GDR1/2 alanine mutant completely lost the pepsin resistance of the parent protein. SBTI N48A was the most susceptible mutant, but 38% remained after 100 min with 1 mg/mL pepsin, compared to 80% of WT SBTI. Some alanine mutants showed better pepsin resistance than the WT. For example, 95% of D22A or 91% of L50A remained at 100 min (Fig. 4.3C).

Taken together, individual mutation of GDR residues to alanine does not appear to completely remove the heat and pepsin resistance of SBTI.

A

Clone	Expected mass with disulfides formed (Da)	Observed mass (Da)	Observed mass – expected mass
D22A	21,212.7	21,212.5	-0.2
I23A	21,214.6	21,215.1	+0.5
T24A	21,226.7	21,226.9	+0.2
A25G	21,242.7	21,243.1	+0.4
R47A	21,171.6	21,172.1	+0.5
N48A	21,213.6	21,214.1	+0.5
E49A	21,198.7	21,199.3	+0.6
L50A	21,214.6	21,215.0	+0.4

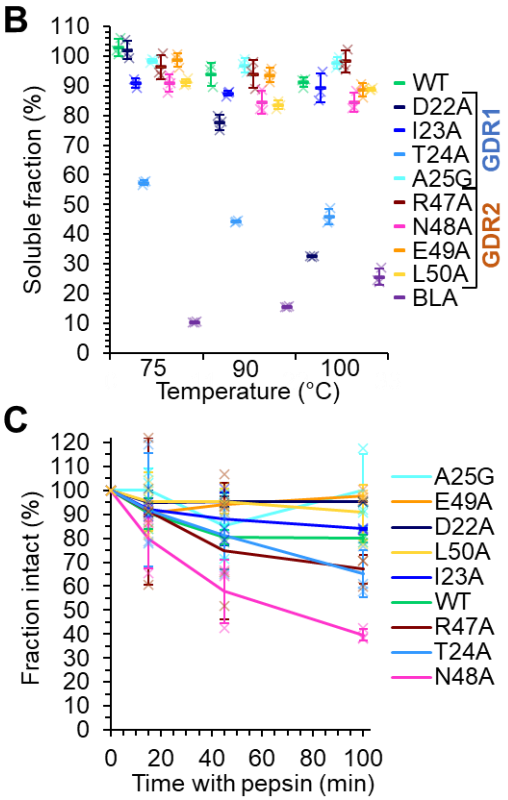


Figure 4.3. Experimental alanine scan of GDR residues. (A) ESI-MS confirms alanine mutant expression. **(B)** Thermal resilience of alanine mutants. SBTI, alanine mutants or BLA were heated to 75, 90 or 100 °C for 10 min before cooling and pelleting aggregates. Remaining soluble protein was analysed on SDS-PAGE and quantified by gel densitometry. Wild-type SBTI (WT) and β-lactamase (BLA) were included as controls. Samples kept at 25 °C were set to 100% (mean ±1 s.d., n = 3 with individual data-points as crosses). **(C)** Pepsin stability of alanine mutants. WT SBTI and alanine mutants were digested with 1 mg/mL pepsin at pH 2.2 at 37 °C. Digestion was stopped by boiling in SDS-loading dye for 3 min. Remaining intact protein was analysed by SDS-PAGE and quantified by gel densitometry. Samples at t = 0 were set to 100% (mean ±1 s.d., n = 3 with individual data-points as crosses).

4.3. Establishing phage display of SBTI

4.3.1. Optimising phagemid construct and conditions to ensure genetic stability

Phage display is a useful technique for evolving novel binding interactions (Frenzel et al., 2016; Shim, 2017; Smith, 1985). Previous studies employing M13 phage display of protease inhibitors have reported genetic instability of phagemid constructs, possibly due to toxic effects of overexpression of protease inhibitors during phage effects (Beekwilder et al., 1999). Optimisation parameters for successful phage display include the host cells, promoter, leader sequence and what phage coat protein to fuse the POI to. My initial phagemid construct was based on pFAB5c (Engberg et al., 1996) (Fig. 4.4A). pFAB5c expresses SBTI fused to truncated minor coat protein III (pIII, residued 216-425) and a DsbA leader sequence [DsbA(ss)] under control of P_{lacZ} (Fig 4.4A). DsbA(ss) targets proteins to the periplasm by the co-translational signal recognition particle (SRP)-mediated pathway, which improves display of fast-folding thermo-stable proteins (Steiner et al., 2006). I chose DsbA(ss) as leader sequence because disulfide bridge-containing proteins tend to fold better in the oxidizing periplasm (de Marco, 2009).

4.3.2. Retention of SBTI-M13 on trypsin-agarose beads

M13 phage and M13 phage displaying SBTI (M13-SBTI) were produced in *E. coli* ER2738 at 18 °C. I tested functional display of SBTI by incubating phage with trypsin agarose beads (Fig. 4.4B). M13-SBTI was retained on trypsin-agarose resin whilst the WT phage control was not (Fig. 4.4B). However, I was unable to perform selections with these phages. After selection rounds, all hits contained either frameshift mutations or deletions of the coding region of SBTI.

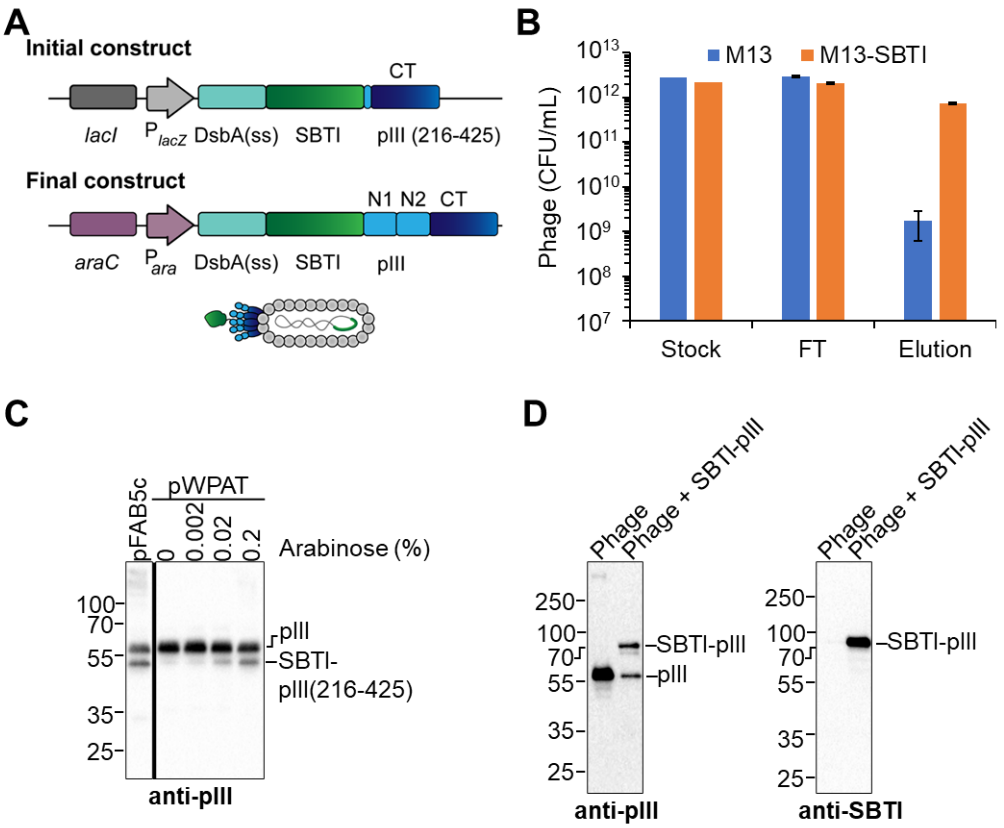


Figure 4.4. Establishing phage display of SBTI. (A) Phagemid constructs for phage display. A cartoon of a phage is drawn below. **(B)** Functional assay of SBTI display. WT M13 phage or M13-SBTI phage produced with the initial construct (A) in *E. coli* ER2738 was incubated with trypsin-agarose resin. Trypsin-agarose beads were washed with Tris-buffer pH 8 before elution with pH 2.2. Input phage (stock), flow-through (FT) and elution were titered by qPCR (mean ±1 s.d., n = 3). **(C)** Titratable induction of SBTI-pIII(216-425) expression. M13-SBTI phage with SBTI-pIII(216-425) under control of an arabinose-inducible promoter (P_{ara}) were produced with varying induction concentrations of arabinose. Phage were blotted with anti-pIII to detect expression of SBTI-pIII(216-425). Phage produced with IPTG induction (pFAB5c vector) was included as control. **(D)** Display of SBTI fused to full-length pIII. M13 phage displaying SBTI-pIII was produced in *E. coli* TG1 cells with arabinose induction. Display was confirmed by parallel blotting with anti-pIII or anti-SBTI. M13 SR phage was included as control.

4.3.3. Western blot verifies display of SBTI on M13

Genetic stability of the phage system was improved by replacing P_{lacZ} with an arabinose inducible promoter, P_{ara} (Guzman et al., 1995), creating the phagemid vector pWPAT (Huang et al., 2000) (Fig. 4.4A). I also switched to recombination deficient *E. coli* TG1 cells for phage production. Control over expression of SBTI-

pIII(216-425) was confirmed by Western blot. The intensity of the SBTI-pIII(216-425) band increased with higher arabinose concentrations in a dose-dependent manner (Fig. 4.4C). Five copies of pIII are present in each phage particle. Efficient incorporation of SBTI-pIII(216-425) is evidenced by the relative intensity of the recombinant pIII band to the WT pIII band (Fig. 4.4C). In samples induced with 0.2% arabinose, the band intensity of SBTI-pIII(216-425) was approximately one fifth of the wild-type pIII band intensity in the same sample.

In parallel, I tested the display of SBTI on full length pIII (Fig. 4.4D). I cloned the N-terminal domains (N1, N2) from pIII from the M13 helper phage genome into the pWPAT vector to create pWPAF. As previously, strong band intensity of the recombinant SBTI-pIII band relative to WT pIII showed efficient display of SBTI in phage particles. Display of SBTI was also confirmed by blotting with anti-SBTI antibody (Fig. 4.4D).

4.4. Selecting an SBTI binder to CROP

4.4.1. Expression, purification and biotinylation of CROP

C. difficile toxin A combined repetitive oligopeptide domain (CROP) was chosen as first target for gastrobodies. TcdA and TcdB disrupt the intestinal epithelium (Kuehne et al., 2010). As such, the toxins are key effectors in *C. difficile* pathogenesis and neutralisation of the toxins may protect against *C. difficile* challenge (Babcock et al., 2006).

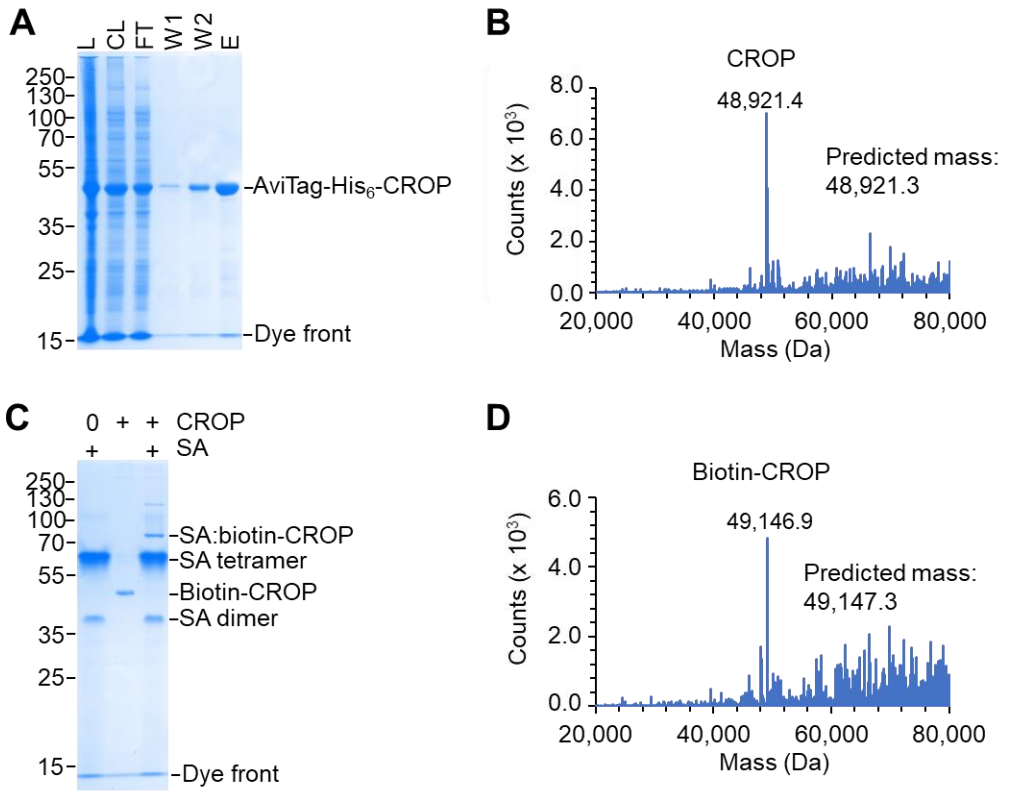


Figure 4.5. Expression, purification and biotinylation of CROP. (A) AviTag-His$_6$-CROP was expressed in *E. coli* BL21 (DE3) RIPL at 25 °C and purified using Ni-NTA. Lysate (L), cleared lysate (CL), flow-through (FT), washes (W1, W2) and elution (E) fractions were run on SDS-PAGE. **(B)** ESI-MS confirming identity of purified CROP. **(C)** Gel shift assay to test biotinylation of CROP. Purified AviTag-His$_6$-CROP was biotinylated with GST-BirA for use in phage display. Biotin-CROP was incubated with streptavidin (SA) for 15 min before running on SDS-PAGE. **(D)** ESI-MS confirms biotinylation of CROP.

I cloned and expressed CROP (TcdA residues 2304 – 2710) (Hussack et al., 2011b) with a N-terminal AviTag (Fairhead and Howarth, 2015; Schatz, 1993) and His$_6$-tag (Fig. 4.5A). Correct expression of AviTag-His$_6$-CROP was confirmed by ESI-MS (Fig. 4.5B). Biotinylated bait is often used in phage display to capture bait on magnetic beads decorated with streptavidin. The strong biotin-streptavidin interaction allows facile, reproducible capture following an in-solution binding phase of the bait and phage. I used BirA for site-specific biotinylation of CROP at the AviTag (Fairhead and Howarth, 2015). Successful production of mono-biotinylated CROP was confirmed by

a streptavidin gel-shift assay (Fig. 4.5C) and ESI-MS (Fig. 4.5D). In the presence of excess streptavidin, the biotin-CROP band shifted to a higher molecular weight indicative of complete biotinylation (Fig. 4.5C). In line with addition of one biotinyl species, the CROP ESI-MS peak was 226 Da heavier after biotinylation (Fig. 4.5D).

4.4.2. Cloning a library of SBTI for phage display

Following the optimisation of the phagemid and phage production conditions, I cloned a library of SBTI that was randomised in GDR1 (residues 22 – 25) and GDR2 (residues 47 – 50). I used degenerate oligonucleotides and Gibson assembly (Gibson et al., 2009) to assemble the library in pWPAT, where SBTI is fused to pIII(216-425). The four amino acid binding regions were randomised with NNK codons, allowing all twenty natural amino acids and the amber stop codon (TAG). The library was cloned in TG1 cells which suppress the amber stop with a glutamine. Correct assembly of the library was confirmed by sequencing ten clones. The library contained 6.3×10^7 clones.

4.4.3. Panning to select an SBTI binder to CROP

I performed two separate selection campaigns, with three rounds of biopanning each (Fig. 4.6). In both campaigns I incubated the library of M13-SBTI with biotinylated CROP. Bait-bound phage were captured on magnetic streptavidin beads. Beads were washed to remove non-specific or weak binders. The first campaign eluted bound phage with pH 2; the second campaign eluted with pH 11 (Fig. 4.6). To increase the stringency of selection rounds, I reduced the number of input phage from 10^{13} in the first round to 10^{11} in the second and third round.

I observed consistent enrichment of the recovered phage titre between rounds (Fig. 4.7A). The enrichment between the first and third round was approximately 2.8 x 10^4 for the glycine campaign and 3.4 x 10^3 for the triethylamine elution campaign (Fig. 4.7A).

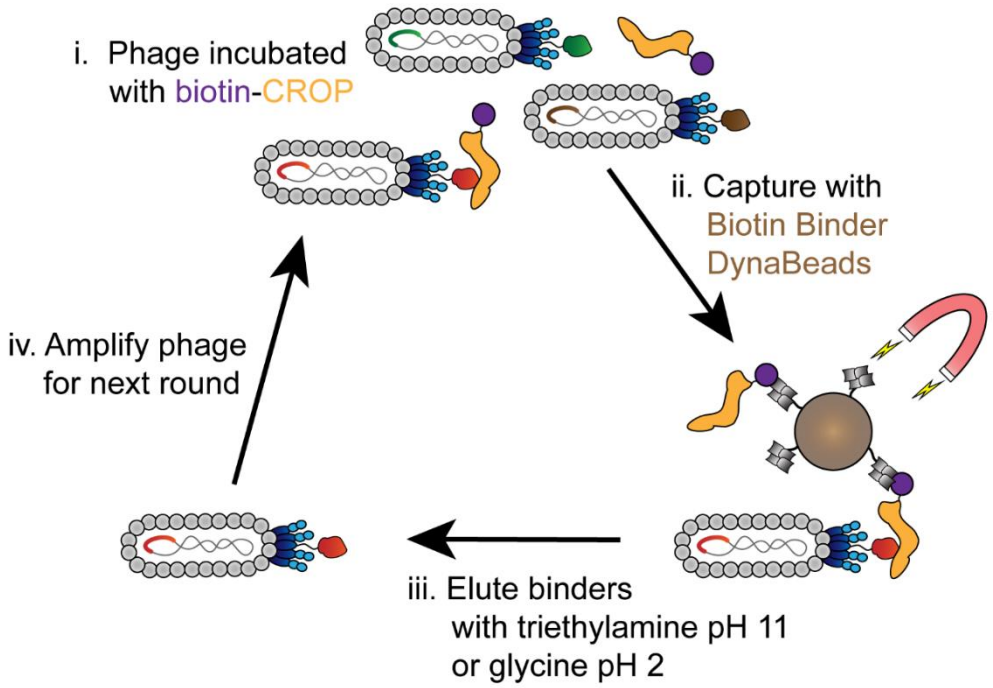


Figure 4.6. Diagram of gastrobody selection strategy against CROP. M13 SR phage displaying a library of SBTI on truncated pIII (residues 216-425) were incubated with biotinylated CROP (orange). Phage bound to bait were captured on magnetic streptavidin beads (brown). Washes were used to reduce unbound or non-specifically binding phage. Bound phage were eluted with triethylamine (pH 11) or glycine (pH 2). Recovered phage are amplified in *E. coli* TG1 for subsequent rounds of selection.

4.4.4. ELISA to screen individual clones for CROP binding

Encouraged by the enrichment across selections campaigns, I screened recovered phage for binding to CROP using a monoclonal phage ELISA (Fig. 4.7B). Antigens were coated on wells of 96-well plates and incubated with phage-containing supernatant. I included anti-M13 antibody as a control antigen to confirm phage

production. Trypsin was included as a control antigen to confirm display of functional SBTI on phage. Lysozyme served as negative control antigen. Phage production was relatively consistent across the screened clones, but not all clones displayed functional SBTI (Fig. 4.7B). Several CROP-binding clones were identified (Fig. 4.7B).

4.4.5. Sequences of anti-CROP SBTI clones

Ten hits identified in the monoclonal phage ELISA were sequenced (Fig. 4.7C). Three clones had amber stop codons (G22, T18, T42), while three other clones included frameshift mutations (G28, T23, T29). Two sequences were observed multiple times in separate clones (G2/G27/T12 and T18/T42) (Fig. 4.7C).

4.5. Characterisation of anti-CDTA gastrobodies

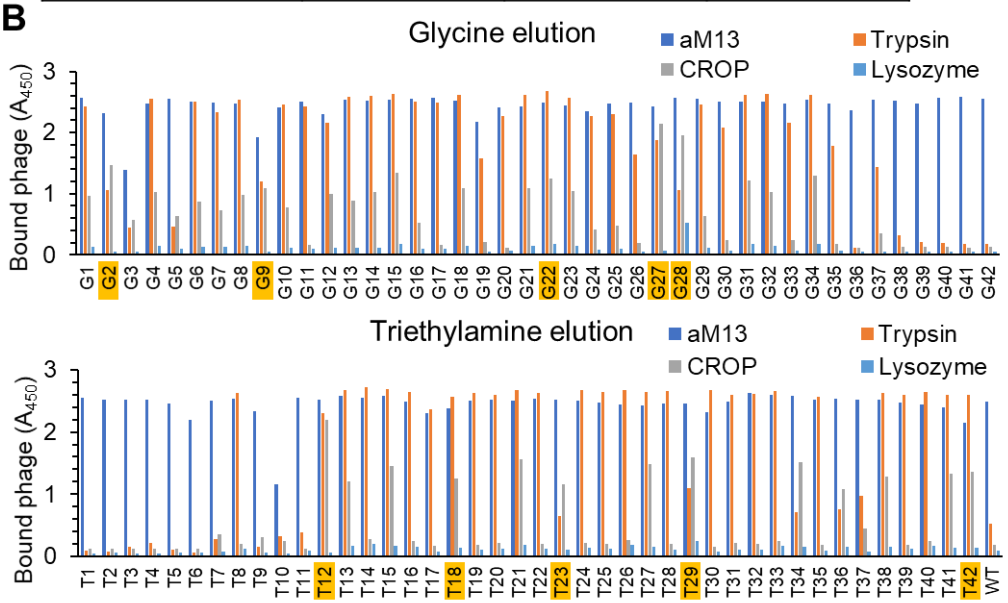
4.5.1. Expression and MS validation of anti-CROP hits

Unique clones were cloned into pET28a for expression, after frameshift or amber stop codons had been corrected. Hits were cloned in the format His_6-SpyTag003-gastrobody. I expressed the anti-CROP gastrobodies in *E. coli* T7SHuffle and purified with Ni-NTA, followed by size exclusion chromatography (SEC) (Fig. 4.8A). The anti-CROP gastrobodies formed monodisperse peaks in SEC and eluted at a similar volume to WT SBTI (Fig. 4.8A). The yields of protein post-SEC ranged from 0.2 mg/L to 1.2 mg/L culture.

A

Elution	Eluted phage titer (CFU)		
	R1	R2	R3
Glycine pH 2.2	3.5×10^3	5.5×10^5	9.9×10^7
Triethylamine pH 11	9.7×10^3	5.5×10^5	3.3×10^7

B



C

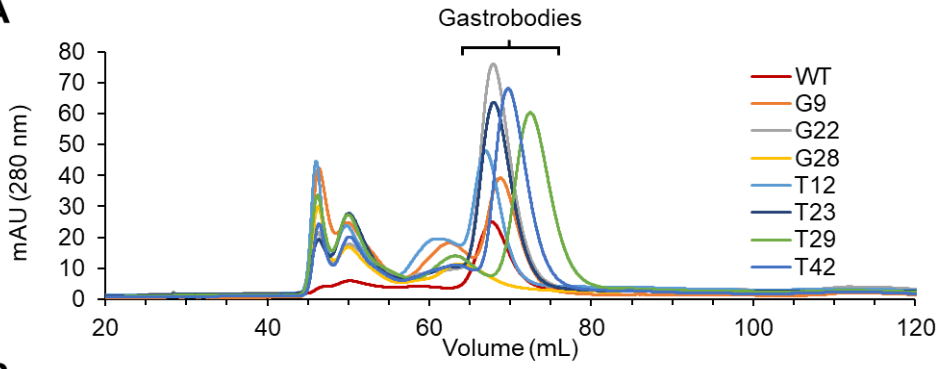
Clone	Sequence		Elution
	GDR1	GDR2	
WT	DITA	RNEL	n/a
G2	IAGT	LVFP	Glycine
G9	HPDL	LYLF	Glycine
G22	L*NR	HDFM	Glycine
G27	IAGT	LVFP	Glycine
G28‡	PYRY	WLHR	Glycine
T12	IAGT	LVFP	TEA
T18	VPH*	RNKW	TEA
T23‡	SFRG	FRLE	TEA
T29‡	RPPQ	TWGW	TEA
T42	VPH*	RNKW	TEA

‡frameshift mutation
*amber stop codon

Figure 4.7. Selecting a gastrobody binder against CROP. (A) Elution titres from three rounds of selection. Bound phage were eluted with glycine pH 2.2 or triethylamine pH 11. Phage recovered from each round were titered by serial dilution and plating of infected log-phase *E. coli* TG1 cells on selective media. **(B)** Monoclonal phage ELISA to screen for CROP binders. Phage from the third round of selection were grown in a 96-well plate. Cells were pelleted and the phage-containing supernatant was incubated in plates coated with anti-M13 antibody (aM13), trypsin, CROP or lysozyme. Phage bound to antigens were detected with anti-M13:HRP. Clones chosen for sequencing are highlighted in yellow. **(C)** Anti-CROP hit GDR sequences. 10 hits from anti-CROP selections were sequenced. GDR sequences for hits from glycine or triethylamine selections are tabulated. Identical sequences appearing twice are highlighted in grey (T18, T42). Hits appearing thrice are highlighted in green (G2, G27, T12). TG1 cells repress amber codons (TAG, indicated by *) as Gln. Mutations leading to frameshifts are indicated by ǂ. TEA = triethylamine, n/a = not applicable.

I analysed the anti-CROP gastrobodies by ESI-MS (Fig. 4.8B). All hits expressed correctly and formed the two expected disulfide bonds. In some spectra there were peaks indicating gluconoylation or phosphogluconoylation (Fig. 4.8B), spontaneous post-translational modifications previously observed for His-tagged proteins expressed in *E. coli* (Geoghegan et al., 1999).

A



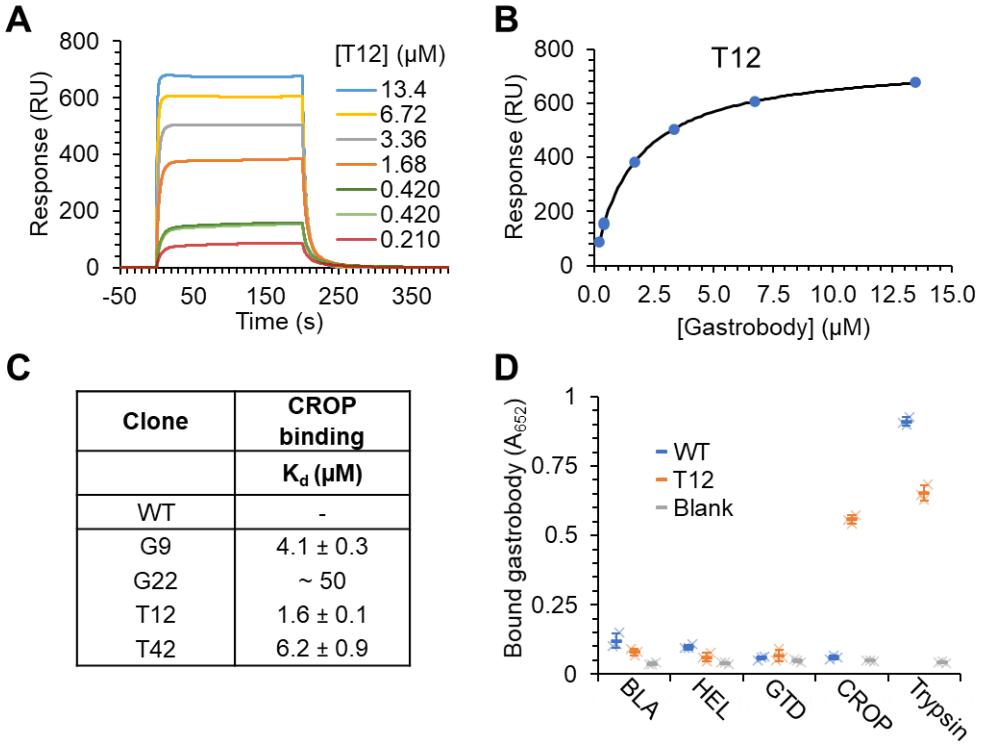
B

Clone	Expected mass with disulfides formed (Da)	Observed mass (Da)	Observed mass – expected mass	Gluconoylation	Phospho-gluconoylation
WT	24,374.4	24,375.0	+0.6	x	
G9	24,460.6	24,461.1	+0.5	x	x
G22	24,503.6	24,505.1	+1.5	x	x
T12	24260.4	24,261.3	+0.9	x	x
T23	24,454.5	24,455.1	+0.6	x	
T29	24,470.5	24,471.1	+0.6	x	
T42	24,507.6	24,508.1	+0.5	x	x

Figure 4.8. Expression and purification of anti-CROP gastrobodies. (A) Size-exclusion chromatography (SEC) of anti-CROP gastrobodies. Anti-CROP gastrobodies were expressed in the format His_6-SpyTag003-gastrobody in *E. coli* T7SHuffle. Anti-CROP gastrobodies were purified by Ni-NTA and SEC. Fractions containing gastrobodies are marked. mAU = milli absorbance units. **(B)** ESI-MS of anti-CROP gastrobodies. Samples with peaks indicating gluconoylation (+178 Da) or phosphogluconoylation (+258 Da) are marked with x.

4.5.2. Binding affinity of anti-CROP gastrobodies

Next, I sought to determine whether the anti-CROP hits bind their target when expressed without phage. Surface plasmon resonance (SPR) was used to determine the binding affinity of anti-CROP gastrobodies (Fig. 4.9A-C). Hits with frameshift mutations when clones were sequenced initially (G28, T23, T29; Fig. 4.7C) did not bind CROP. T12, also identified as G2 and G27, was the tightest binder with a K_d of 1.6 ± 0.1 µM (Fig. 4.9A,B). The remaining hits bound with low micromolar affinity (Fig. 4.9C).



Clone	CROP binding
	K_d (µM)
WT	-
G9	4.1 ± 0.3
G22	~ 50
T12	1.6 ± 0.1
T42	6.2 ± 0.9

Figure 4.9. Binding characterisation of anti-CROP gastrobodies. (A) Binding of T12 to immobilized CROP. Multi-cycle SPR trace of T12 binding to CROP at pH 7.4 and 25 °C. **(B)** Equilibrium analysis of T12 binding affinity to CROP. **(C)** Binding affinity (K_d) of anti-CROP gastrobodies. K_d of gastrobodies binding to CROP was determined by SPR and equilibrium analysis in 1:1 binding model. Error represents the uncertainty of fit to the binding curve. **(D)** Binding specificity of T12. Binding of T12 to wells coated with β-lactamase (BLA), hen egg lysozyme (HEL), *C. difficile* toxin B glucosyltransferase domain (GTD), CROP or trypsin was analysed in an ELISA. Bound T12 or WT SBTI was detected with polyclonal anti-SBTI antibody. Buffer without gastrobodies (blank) was included as control. (mean ±1 s.d., n = 3 with individual data-points as crosses).

4.5.3. ELISA to test specificity of anti-CROP gastrobodies

Specificity of T12 binding to CROP was assessed in an ELISA (Fig. 4.9D). T12 showed binding to CROP and trypsin, but no binding to control antigens β-lactamase (BLA), lysozyme (HEL) or *C. difficile* toxin B glycosyltransferase domain (GTD). Interestingly, the binding of T12 to trypsin appeared weaker than the binding of WT SBTI to trypsin (Fig. 4.9D).

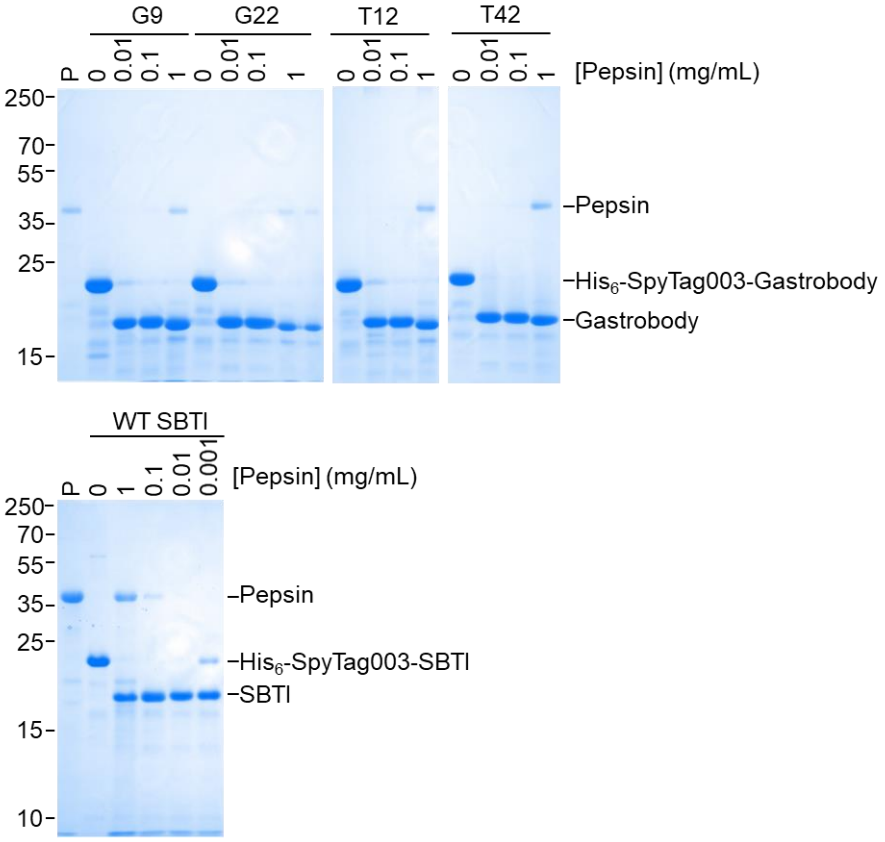


Figure 4.10. Pepsin resistance of anti-CROP gastrobodies. Anti-CROP gastrobodies or WT SBTI were incubated with serial dilutions of pepsin at pH 2.2 at 37 °C for 10 min. Digestion was stopped by boiling in SDS loading dye. Remaining protein was separated by SDS-PAGE and stained with Coomassie. Pepsin alone without gastrobody (P) was run as control.

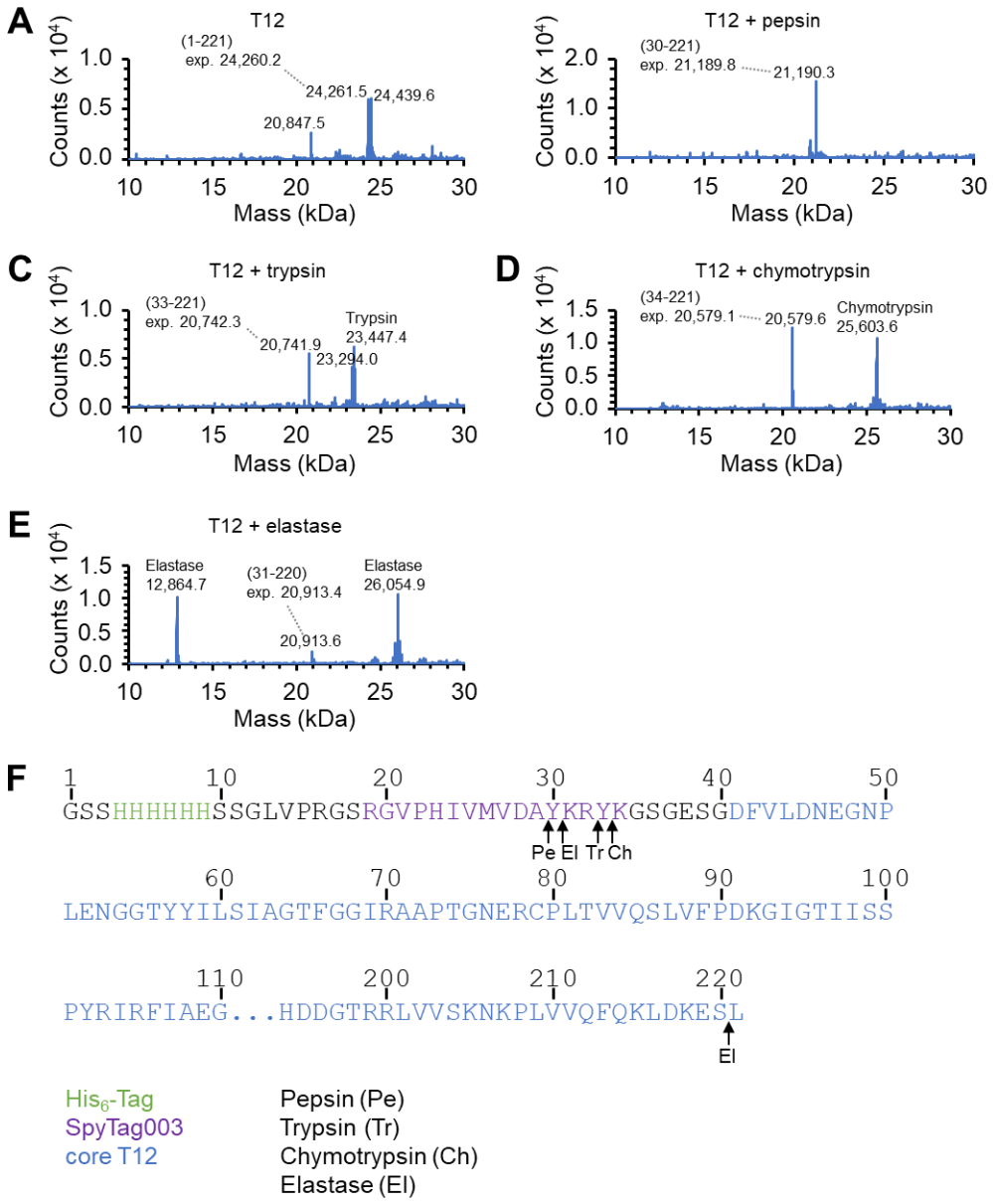


Figure 4.11. Protease cut sites in T12 gastrobody. His_6-SpyTag003-T12 construct was incubated in PBS **(A)** or with 1 mg/mL pepsin **(B)**, 100 U/mL trypsin **(C)**, 25 U/mL chymotrypsin **(D)** or 10 U/mL elastase **(E)** for 15 min at 37 °C. Digest products were analysed by ESI-MS. Expected mass of construct fragments are marked. **(F)** The observed cut sites in the amino acid sequence of the His_6-SpyTag003-T12 construct are indicated with arrows. I performed sample preparation and data analysis. Dr Anthony Tumber loaded and ran the mass spectrometer.

4.5.4. Pepsin resistance anti-CROP gastrobodies

A key goal of the project was to ensure that evolved gastrobodies retain the pepsin resistance of SBTI. Each of the hits has eight amino acid mutations relative to WT

SBTI. CROP-binding gastrobodies were incubated with serial dilutions of 1 mg/mL pepsin at pH 2.2 and 37 °C to determine pepsin stability (Fig. 4.10). In all conditions tested with pepsin, the gastrobody band shifted down by about 5 kDa, likely consistent with loss of the N-terminal tags (Fig. 4.10). The band shift was also observed for WT SBTI with the same N-terminal tags. Following the shift, little degradation of the gastrobody bands was evident, even in the presence of 1 mg/mL pepsin, indicative of similar pepsin stability to WT SBTI (Fig. 4.10).

4.5.5. Protease cleavage sites in T12

I digested anti-CROP gastrobody T12 with gastrointestinal proteases and analysed the digest products with ESI-MS, to determine where in the construct the proteases cut (Fig. 4.11). Pepsin, trypsin, chymotrypsin and elastase cut in the N-terminal SpyTag003 (Fig. 4.11). Elastase also cleaved the C-terminal Leu (Fig. 4.11E,F). Cleavage sites in the main gastrobody domain were not identified.

4.5.6. Thermal stability of anti-CROP gastrobodies

Lastly, I sought to characterize the heat stability of anti-CROP gastrobodies. DSC was performed to compare the T_m of the gastrobodies at pH 7.4, pH 2.0 and pH 1.2 (Fig. 4.12A,B). Interestingly, at pH 7.4 the T_m values of G9 (69.9 °C), G22 (75.3 °C) and T42 (78.6 °C) were higher than WT SBTI (67.3 °C), while the T_m of T12 was slightly lower (66.8 °C) (Fig. 4.12A,B). At pH 2.0 or pH 1.2 the T_m of all gastrobodies was lower than at pH 7.4. T42 was most stable at pH 2.0 with a T_m of 71.0 °C. Both T12 and WT SBTI appeared to remain reasonably stable at pH 1.2, with melting temperatures above physiological temperatures (Fig. 4.12A).

Trypsin-binding was used as a functional test for heat resilience of T12 (Fig. 4.12C). There was little difference in T12's binding to trypsin after heating to 55 °C for 10 min compared to 37 °C. After heating to 75 or 100 °C, however, less than 5% of bound T12 remained, relative to samples heated to 37 °C (Fig. 4.12C).

A

Clone	T_m (°C) from DSC		
	pH 7.4	pH 2.0	pH 1.2
WT	67.3	60.4	56.2
G9	69.9	55.2	n.a.
G22	75.3	61.0	n.a.
T12	66.8	53.6	47.1
T42	78.6	71.0	n.a.

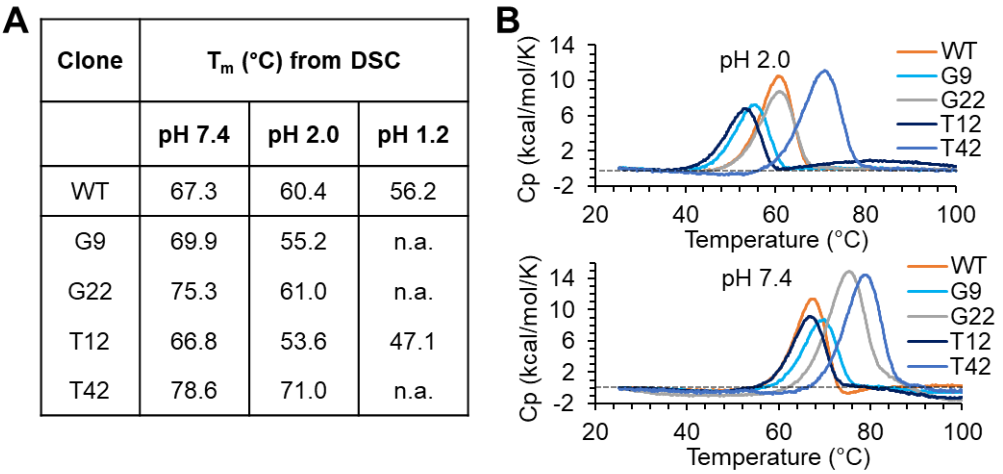


Figure 4.12. Temperature stability of anti-CROP gastrobodies. (A) Melting temperatures (T_m) of anti-CROP gastrobodies compared to WT SBTI. DSC was performed in phosphate buffer at pH 7.4, 2.0 or 1.2. All constructs were in the format His_6-Thrombin site-SpyTag003-gastrobody/SBTI. G9, G22 and T42 were not analysed (n.a.) at pH 1.2. **(B)** DSC traces of anti-CROP gastrobodies at pH 2.0 and pH 7.4. I performed sample preparation. Dr David Staunton loaded the DSC machine and processed raw data. **(C)** Functional test of heat resilience of T12. Trypsin binding of soluble T12 after heating to indicated temperature for 10 min was analysed by ELISA at 25 °C (mean ±1 s.d., n = 3 with individual data-points as crosses).

4.6. Discussion

SBTI is relatively stable in gastrointestinal conditions compared to other protein-based targeting modalities, but SBTI has not been used as a scaffold. Kunitz domain proteins have previously been engineered to bind new targets for intravenous administration (Dunlevy et al., 2012; Williams and Baird, 2003), but to my knowledge there are no reports of engineering Kunitz-type inhibitors like SBTI (Gebauer and Skerra, 2020; Vazquez-Lombardi et al., 2015). Hence, the protease stability of engineered β-trefoil proteins was likely unknown. In this chapter, I show that gastrobodies can retain pepsin resistance when two four-amino acid loops are randomised to bind to new targets.

The binding surfaces of many protein scaffolds like nanobodies (Arbabi Ghahroudi et al., 1997), nanofitins (Mouratou et al., 2007) or affibodies (Nord et al., 1997) are derived from the residues involved in binding the cognate target of the parent protein. Rather than trying to repurpose the trypsin-binding residues (Song and Suh, 1998), I designed a new binding surface. Binding surface design is important for diversifying the types of epitopes that can be targeted (Koide et al., 2012; Zimmermann et al., 2018). For example, long flexible loops may be suitable for targets with a cleft or groove, while a flat or concave binding surface may be better suited to flat epitopes. Kunitz domains, like bovine pancreatic trypsin inhibitor (BPTI), were previously engineered to bind proteases similar to their cognate protease targets with a binding surface that is a single loop (Dennis et al., 1995; Markland et al., 1996a, 1996b; Roberts et al., 1992; Röttgen and Collins, 1995; Szardenings et al., 1995; Williams and Baird, 2003). Knottins are plant-derived peptides constrained by three disulfide bonds arranged in a knot with remarkable stability (Weidmann and Craik, 2016). For example, kalata B1 was discovered because it remains active after boiling during preparation of a traditional uterotonic medicine (Craik et al., 1999). Knottins and

disulfide-rich peptides have been engineered and evolved by phage display to bind new targets for therapeutic applications (Barkan et al., 2016; Craik et al., 1999; Smith et al., 1998; Stoop and Craik, 2003). Some cysteine-rich peptides have been engineered for use with oral administration. For example, a bradykinin B_1-binding peptide was grafted onto sunflower trypsin inhibitor and shown to be active when administered orally (Qiu et al., 2017). However, SBTI was chosen as template for gastrobodies because SBTI may be more amenable to mutation and design of versatile binding surfaces than Kunitz domains or cysteine-rich peptides. Specifically, SBTI has more loops that could be engineered without being too constrained by the small size and geometry of the backbone scaffold. Furthermore, due to a less complicated fold, SBTI may be easier and cheaper to produce at industrial scale for feed applications.

I used Rosetta to predict mutable residues in SBTI and combined consecutive mutable residues into binding loops (termed GDRs). Individual mutations were analysed for simplicity and reducing computational power required. The weakness of this approach is that it does not account for compensatory mutations when multiple mutations are made simultaneously. Similarly, the backbone conformation dependence of running *pmutscan* after *relax* was reduced by analysing an ensemble of structures. The mutability of GDR residues was then validated experimentally. Again, the focus was on studying the contribution of individual residues to stability of the protein because pepsin resistance is not fully understood. An alternative approach could be to replace possible binding loops with serine-threonine rich loops, which are commonly found at protein-protein interfaces (Zimmermann et al., 2018).

Initial attempts at phage display selections with gastrobodies were unsuccessful, only yielding deletion or frameshift mutants. A study including phage display of potato protease inhibitor (PI2) describes similar complications due to toxicity caused by the metabolic burden on the cell during protein overexpression and phage infection (Beekwilder et al., 1999). I overcame these issues by using a titratable induction system (P_{ara}) and *E. coli* TG1 host cells. Nevertheless, some hits identified in the phage selections against CROP featured frameshift mutations. G28, T23 and T29 bound CROP when displayed on phage, but not when expressed solubly. The disconnect between observed binding on phage and as pure protein is likely due to stickiness or non-specific binding of the truncated protein on phage. More stringent selection conditions might reduce the incidence of false hits.

The low micromolar affinity of gastrobodies to CROP is inadequate for practical applications, where affinities in the double or single digit nanomolar range are required. Length of the current GDRs may restrict the affinity of gastrobodies. The current design randomises eight amino acids in total, while the CDR3 of antibodies in clinical development usually consist of twelve to fifteen residues (Raybould et al., 2019). Furthermore, this particular study was likely limited by the library size. Phage display libraries are limited by the transformation efficiency of *E. coli* allowing libraries of up to 10^9 or 10^{10} variants (Frei and Lai, 2016). To fully sample the sequence space of the gastrobody design, a library of more than 2.6×10^{10} variants ($20^8 = 2.6 \times 10^{10}$) would need to be constructed, not accounting for codon redundancy. The library here, however, only contained 6.3×10^7 clones. Thus, constructing a larger library, perhaps with restricted diversity (Bhattacharjee and Biswas, 2010; Seeger et al., 2013; Zimmermann et al., 2018), could improve the hit rate of high affinity of binders. Alternatively, an affinity maturation approach where diversity is introduced into the

GDRs of current binders should optimise binders toward better affinity (Thie et al., 2009). Further targets should also be explored. CROP is a flexible protein which might make achieving high affinity binding more difficult (Chumbler et al., 2016).

Despite disappointing binding affinities, the anti-CROP gastrobodies remained resistant to digestion by pepsin. ESI-MS analysis showed that pepsin cut in the N-terminal SpyTag003, C-terminal of an Ala and N-terminal of a Tyr. The pepsin cleavage pattern in SpyTag003 is consistent with the preference of pepsin to cut N-terminal of hydrophobic residues like Phe, Trp or Tyr (Hamuro et al., 2008). No internal cut sites in T12 were identified but as discussed in Chapter 3, it seems unlikely that internal sites would be identified due to experimental set-up. Cleavage fragments are unlikely to be detected because the probability of further rapid cleavage events is higher for the fragments than the well folded SBTI domain. Further experiments should include functional validation that T12 retains its binding affinity following protease digestion.

Indeed, the heat resilience of T12 followed a similar pattern to WT SBTI (presented in chapter 3). Heating of T12 above the T_m resulted in substantial loss of functional protein. Surprisingly, several gastrobodies were more stable than WT SBTI, demonstrating that mutation of the GDRs does not necessarily compromise the biophysical properties of the protein. In fact, T42 was substantially more stable than WT SBTI at neutral and acidic pH. The tightest binder (T12) was the least stable gastrobody suggesting there could be a trade-off between stability and affinity. I did not include conditions specifically selecting for improved stability, but this could be useful in future campaigns to ensure that the evolved gastrobody retains the desired properties. For example, including protease pressure would be good to only select

gastrobodies that bind tightly and resist digestion (Kong et al., 2020). Success in such a campaign, however, would rely on having a phage display system that sufficiently resists digestion. Wild-type M13 phage is inactivated in gastric conditions (Tóthová et al., 2012), so a mutant of M13 stable in the presence of pepsin and at low pH would be needed (Fitza, 2002). It will be important to show that other components of the system are pepsin resistant. For instance, the linker between pIII and gastrobody must be pepsin resistant to ensure the genotype-phenotype link is not severed. Furthermore, the protease will need to be inactivated or removed from the phage to prevent degradation of the bait protein. Protease separation can be achieved by PEG precipitation of the phage (Kong et al., 2020).

Chapter 5: Improving gastrobody affinity

Chapter 5: Improving gastrobody affinity

In Chapter 4, I showed that SBTI can be engineered into an antibody mimetic, termed a gastrobody. Gastrobodies selected against CROP bound with low micromolar affinity. Single digit nanomolar or sub-nanomolar affinity is usually required for therapeutic applications (Vazquez-Lombardi et al., 2015). A key aim of Chapter 5 is to improve the binding affinity of gastrobodies without losing the protease resistance and heat stability of SBTI.

The GDRs of anti-CROP gastrobodies only contain eight amino acids. Clinical antibody CDR3s alone usually consist of twelve to fifteen residues (Raybould et al., 2019). Extending the GDRs might be a strategy to increase the potential affinity of gastrobodies. Longer loops could also be better suited to targets with deep clefts (Correa et al., 2014; Koide et al., 2012; Zimmermann et al., 2018). It will be important to ensure that amino acid insertions or extensions of the GDRs do not compromise gastrobody stability. Long flexible loops are likely good substrates for protease cleavage. In addition, insertions could destabilise the core structure. Complementary to extending the existing loops, other loops could be randomised to increase the binding surface. Previous studies with monobodies showed that randomising additional residues to the canonical binding loop enables binding to more diverse epitopes (Koide et al., 2012).

Successful selections also rely on large naïve libraries for maximal sampling of the structural space. Where more residues are randomised than is possible to sample in a single phage display library (maximum size ~ 10^{10} clones), an affinity maturation approach can be employed. Affinity maturation takes initial hits and introduces new diversity to improve the affinity or stability of the clone (Thie et al., 2009).

In this chapter, I optimise library cloning to improve my chances of identifying high affinity binders. I use an affinity maturation approach to select gastrobody binders with extended loops against the *C. difficile* TcdB glucosyltransferase domain (GTD) (Chaves-Olarte et al., 1997; Ciesla and Bobak, 1998). GTD is required for TcdB-mediated pathogenesis *in vivo* (Bilverstone et al., 2020). Anti-GTD gastrobodies with loop extensions and randomisation of additional GDRs bind with low nanomolar affinity. Despite additional mutations, anti-GTD gastrobodies bind their target after protease digestion or heating.

5.1. Selecting a gastrobody binder to GTD

5.1.1. Expression and purification of GTD

TcdB glucosyltransferase domain (GTD, residues 2 – 543) was chosen as second gastrobody target. I expressed GTD with a N-terminal AviTag-His_6 in *E. coli* and purified by Ni-NTA (Fig. 5.1A). GST-BirA was used to mono-biotinylate AviTag-His_6-GTD prior to SEC purification (Fig. 5.1B). I confirmed the biotinylation of GTD with a streptavidin gel-shift assay (Fig. 5.1C). In the presence of streptavidin, the biotin-GTD band shifted, indicative of complete biotinylation (Fig. 5.1C).

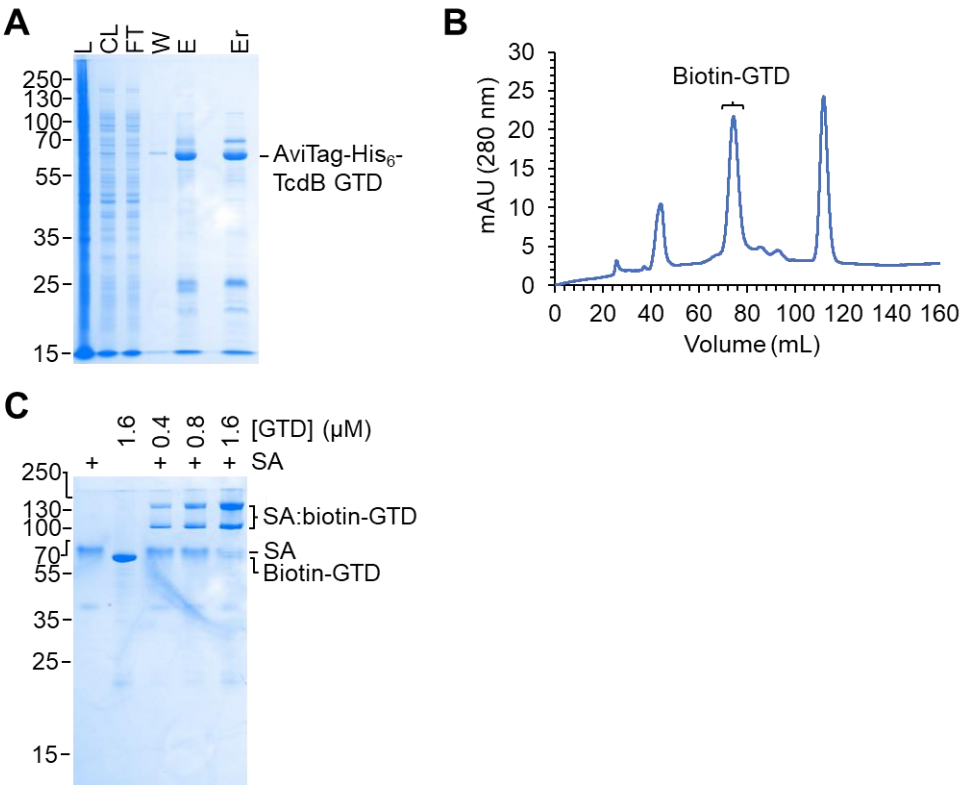


Figure 5.1. Expression and biotinylation of GTD. (A) Purification gel of AviTag-His_6-GTD. AviTag-His_6-GTD (TcdB residues 2 – 543) was expressed in *E. coli* T7 Express *lysY/I*^{q} and purified by Ni-NTA. Fractions from cell lysate (L), cleared lysate (CL), flow-through (FT), washes (W), elution (E) and the elution in reducing conditions (Er) were run on SDS-PAGE and stained with Coomassie. **(B)** Size exclusion chromatography of biotin-GTD. After Ni-NTA and biotinylation, biotin-GTD was run on a HiLoad 16/600 Superdex 200 pg column. The fractions containing GTD are marked. mAU = milli absorbance units. **(C)** Streptavidin gel-shift assay to confirm biotinylation of GTD. Various concentrations of biotin-GTD were incubated with 0.2 mg/mL steptravidin (SA) for 10 min at 25 °C, before running on SDS-PAGE without boiling.

5.1.2. Cloning a library of SBTI for display on full length pIII with extended loops

Increasing the length of the gastrobody binding loops may improve the affinity of binders. Although longer loops may improve affinity, the extended loops may also destabilise the fold or provide a good substrate for proteases. As a compromise, I cloned a new gastrobody library where GDR1 and GDR2 consisted of any combination of either five, six or seven randomised amino acids compared to the natural length of four residues. Degenerate oligonucleotides with NNK codons in GDRs were used to construct the library by Gibson assembly. In the loop extended library, SBTI was fused to full length pIII. The library contained 4.8×10^6 variants.

5.1.3. Panning to select a binder to GTD

I performed three rounds of selection against GTD with the new loop-extended library and the original gastrobody library in parallel (Fig. 5.2). In the second round of selections, I included a 10 min incubation with 0.1 mg/mL pepsin at pH 2.2 and 37 °C to select for pepsin-resistant gastrobodies. In the third round, I included a 10 min incubation with 10 mg/mL pancreatin at pH 6.8 after the pepsin challenge. In parallel, I performed the selections without protease pressure (Fig. 5.2). After each selection round I titered the recovered phage (Fig. 5.3A). I observed 59-fold enrichment between the third and first round in selections with the loop-extended library without protease pressure. Little to no enrichment was observed in the other selection series with protease pressure or with the original gastrobody library (Fig. 5.3A).

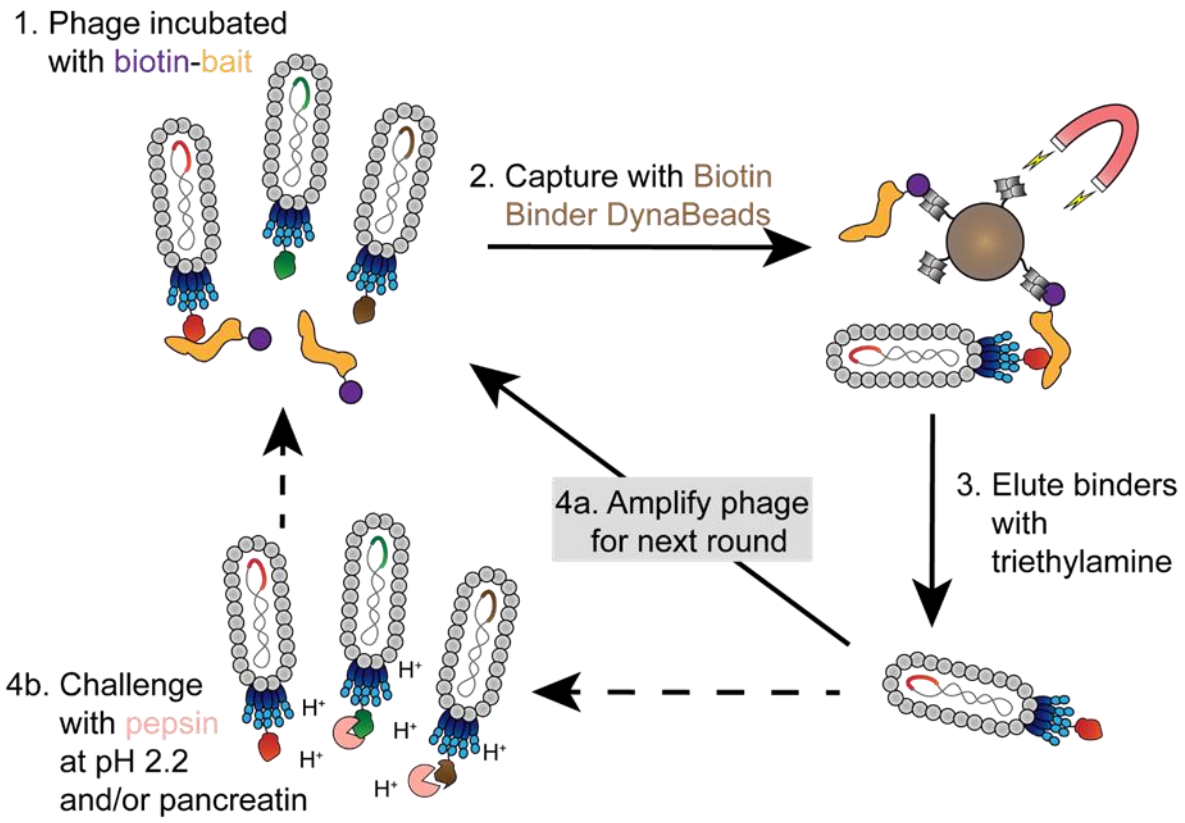
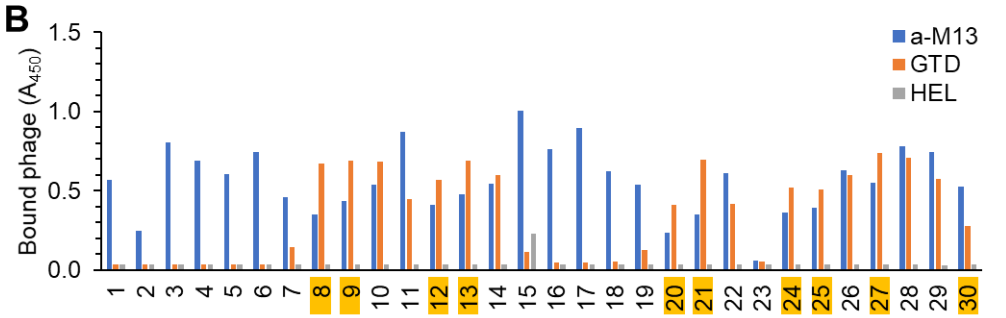


Figure 5.2. Diagram of gastrobody selection strategy against GTD. M13 phage displaying a library of SBTI on pIII or truncated pIII (residues 216 – 425) were incubated with biotinylated GTD (orange). Time of the binding phase was controlled by adding excess non-biotinylated bait. Bait-bound phage were captured on magnetic streptavidin-coated beads (brown). Beads were washed to remove non-specifically bound phage. Incubations with excess non-biotinylated bait at 25 or 37 °C were used to select for slow off-rates. In some campaigns phage were challenged with 0.1 mg/mL pepsin at pH 2.2 and 37 °C for 10 min, followed by 10 mg/mL pancreatin at pH 6.8 and 37 °C for 10 min later rounds. Bait-bound phage were eluted with triethylamine (pH 11). Recovered phage are amplified in *E. coli* TG1 for subsequent rounds of selection.

Clones from the third round of selection with the loop extended library were screened for binding to GTD in a monoclonal phage ELISA (Fig. 5.3B). Several clones binding GTD but not the negative control antigen HEL were identified (Fig. 5.3B). Sequencing of ten phage clones that bound GTD revealed that all hits had the same sequence, with six amino acids each in GDR1/2 (Fig. 5.3C).

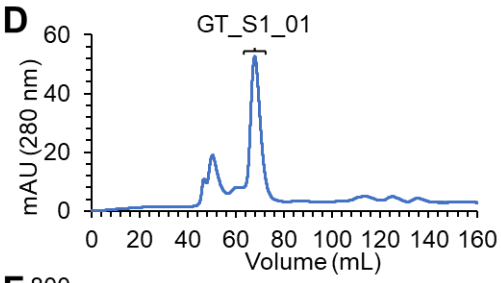
A

Campaign	Eluted phage titer (CFU)		
	R1	R2	R3
Original	9.3×10^3	1.6×10^4	2.9×10^4
Original + protease		7.1×10^3	5.0×10^3
Extended	8.4×10^3	3.2×10^4	5.0×10^5
Extended + protease		8.0×10^3	1.4×10^4



C

Clone	GDR1	GDR2
WT	DITA	RNEL
GT_S1_01	EFGIGL	LFRNHR



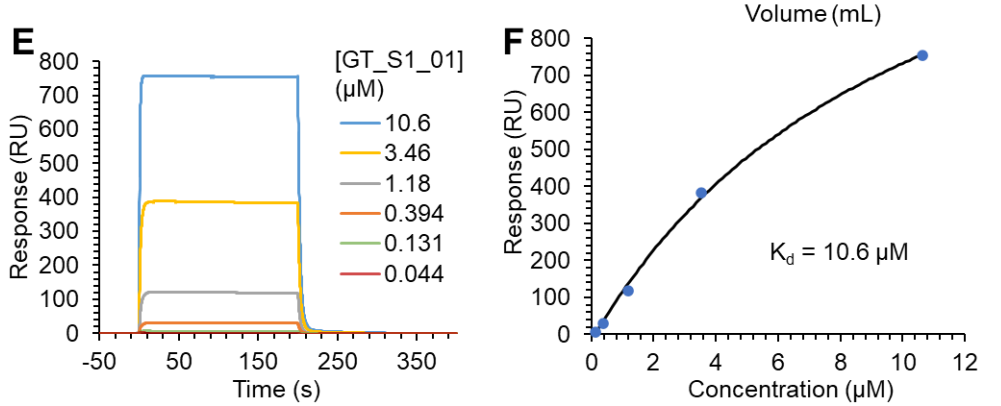


Figure 5.3. Initial selections for a gastrobody binder to GTD. (A) Elution titers from initial selections against GTD. Selections were performed with libraries of SBTI randomised to contain four amino acids in each GDR (original library) or any combination five, six or seven amino acids in each GDR (extended library). After the first round of selection (R1), separate selection campaigns were started to include protease pressure. **(B)** Monoclonal phage ELISA to screen for binding to GTD. Individual phage clones from R3 were grown in plates and the supernatant was incubated in wells coated with anti-M13 antibody (a-M13), GTD or lysozyme (HEL). Clones highlighted in orange were sequenced. **(C)** GDR1/2 sequence of hit in selection against GTD. **(D)** Size exclusion chromatography (SEC) trace of GT_S1_01. mAU = milli absorbance units. **(E)** Surface plasmon resonance trace of GT_S1_01 binding to immobilized GTD at pH 7.4 and 25 °C. **(F)** Equilibrium analysis of GT_S1_01 binding to GTD in a 1:1 binding model.

I cloned the consensus clone, termed GT_S1_01, into pET28a with a N-terminal His_6-SpyTag003 for expression. GT_S1_01 was expressed in T7SHuffle and purified by Ni-NTA. The post-SEC yield of GT_S1_01 was 0.5 mg/L culture (Fig. 5.3D). Preliminary analysis of GT_S1_01 binding to GTD by SPR revealed a dissociation constant (K_d) of approximately 10.6 µM (Fig. 5.3E,F).

5.1.4. Affinity maturation of anti-GTD gastrobody

I sought to improve gastrobody affinity into the nanomolar range to enable practical use. Hence, I decided to use GT_S1_01 as basis for affinity maturation. A larger library size will enable sampling more diverse biochemical space and increase the probability of identifying high affinity binders. I first optimised the conditions for library construction by Gibson assembly to maximize the library size. I found that using 0.2 pmol of each PCR fragment in a 20 µl reaction with 1× NEBuilder HiFi DNA Assembly Master Mix (NEB) worked well for assembling the phagemid. 300 ng of purified Gibson product was transformed into aliquots of electrocompetent *E. coli* TG1 cells.

As previously, I used degenerate oligonucleotides with NNK codons in the GDRs to assemble the library. For the affinity maturation library, I used GT_S1_01 as template and introduced diversity either into GDR1 or GDR2 separately. GDR1/2 were randomised separately with NNK codons to contain five, six or seven amino acids. Focused randomisation with NNK was chosen over the commonly used error-prone PCR (Unkauf et al., 2018) to avoid potentially destabilising mutations in the main gastrobody domain. Restricting the error-prone PCR to the GDRs would be difficult practically because the PCR products would be very short. Furthermore, as less than eight residues are randomised, all permutations can be sampled in a library containing

10^9 variants. The affinity maturation library was constructed with the optimised protocol and contained 3.6 x 10^9 variants. Correct assembly was verified by sequencing ten clones from the naïve library.

Three rounds of selections were performed in the affinity maturation campaign. Selection stringency was increased in later rounds by reducing phage input, decreasing the bait concentration, and reducing the time of the binding phase. In parallel with the normal selection, I included a 10 min incubation with 0.1 mg/mL pepsin at pH 2 and 37 °C to favour pepsin-resistant clones. In addition, I incubated bait-bound phage with excess non-biotinylated bait, to select for slow off-rates. Five clones from the campaign with pepsin and five clones from the campaign without pepsin were sequenced after the third biopanning round (Fig. 5.4A).

All hits mutated GDR1 and kept the GDR2 sequence of GT_S1_01 (Fig. 5.4A). The hits all featured an amber stop, repressed as Gln, at position five in GDR1. The position three Gly and position six Leu of GT_S1_01 were conserved across the unique clones. One sequence (clone GT01) appeared five times in the ten sequenced clones, in selections with and without pepsin (Fig. 5.4A). The clones generally only differed by one or two amino acids from one another. There was no obvious difference between clones where pepsin was present or absent in the selection conditions. The length of GDR1 was six amino acids for all hits (Fig. 5.4A).

5.2. Characterisation of anti-GTD gastrobodies

5.2.1. Expression and MS validation of anti-GTD hits

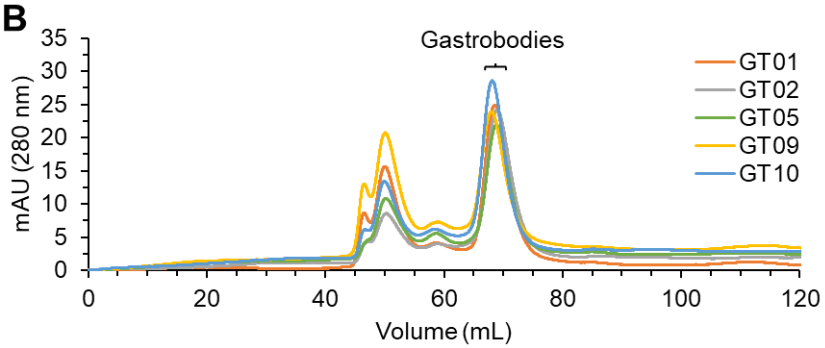
Hits from the affinity maturation selections were cloned into pET28a with a N-terminal His_6-SpyTag003 and the amber codon corrected. Anti-GTD hits were expressed in *E. coli* T7SHuffle and purified by Ni-NTA followed by SEC (Fig. 5.4B). All hits expressed with yields 0.2 – 0.8 mg/L culture after SEC. I confirmed correct expression and disulfide bond formation of anti-GTD gastrobodies by ESI-MS (Fig. 5.4C).

A

Clone	Sequence		Number of times identified	Pepsin in selection
	GDR1	GDR2		
WT	DITA	RNEL	n/a	n/a
GT_S1_01	EFGIGL	LFRNHR	n/a	n/a
GT01	DYGR*L	LFRNHR	5	x*
GT02	NYGR*L	LFRNHR	2	x*
GT05	SYGR*L	LFRNHR	1	
GT09	DFGL*L	LFRNHR	1	x
GT10	NYGL*L	LFRNHR	1	x

*Identified in selection with and without pepsin

B



C

Protein	Expected mass with disulfides formed (Da)	Observed Mass (Da)	Observed – Expected Mass (Da)
GT01	25,018.0	25,017.9	-0.1
GT02	25,017.0	25,017.6	+0.6
GT05	24,990.0	24,990.0	0.0
GT09	24,959.0	24,958.8	-0.1
GT10	24,974.0	24,974.1	+0.1

Figure 5.4. Clones from affinity maturation against GTD. (A) GDR sequences of unique clones from affinity maturation selection against GTD. Consensus residues are highlighted in green (GDR1) or grey (GDR2). The number of times a unique sequence was identified from the ten sequenced clones is indicated. Clones are marked from selections where pepsin challenge was included. Some hits were identified in selections with and without pepsin challenge. n/a = not applicable. **(B)** SEC trace of anti-GTD gastrobodies. mAU = milli absorbance units. **(C)** Masses of main peaks from ESI-MS of anti-GTD gastrobodies.

5.2.2. Binding kinetics of anti-GTD gastrobodies by SPR

I tested the binding kinetics of gastrobodies to immobilised GTD at pH 7.4 by SPR (Fig. 5.5A,B). Consensus clone GT01 bound GTD the tightest with a K_d of 85 ± 2.3 nM, a substantial improvement on the original GT_S1_01 clone (Fig. 5.5A,B). The other hits also showed improved affinities with dissociation constants in the hundreds of nanomolar range.

Binding specificity is an important parameter for translatability of binders. I tested the binding specificity of GT01 in an ELISA (Fig. 5.5C). GT01 bound to GTD and trypsin, but not to the control antigens CROP, BLA or HEL (Fig. 5.5C).

The pH in the small intestine rises from pH 6 in proximal segments to pH 7.4 in distal segments, while the pH in the colon is between pH 5 and pH 8 (Koziolek et al., 2015; Maqbool et al., 2009). As a preliminary test of the pH dependence of gastrobody binding, I analysed GT01 binding to GTD at pH 6.5 (Fig. 5.5D). GT01 bound GTD with a similar K_d at pH 6.5 (45.9 ± 0.1 nM) as at pH 7.4 (85 ± 2.3 nM), possibly due to a slower off-rate at pH 6.5 [(2.0 ± 0.1) x 10^{-2} s^{-1}] than at pH 7.4 [(3.6 ± 0.2) x 10^{-2} s^{-1}] (Fig. 5.5A,B,D). There was no detectable binding of WT SBTI to GTD at pH 6.5 (Fig. 5.5E).

A

Protein	k_{on} ($M^{-1}s^{-1}$)	k_{off} (s^{-1})	K_d (nM)
GT01	(4.2 ± 0.3) x 10^5	(3.6 ± 0.2) x 10^{-2}	85 ± 2.3
GT02	3.0 x 10^5	6.3 x 10^{-2}	215
GT05	2.7 x 10^5	1.1 x 10^{-1}	421
GT09	1.4 x 10^5	8.9 x 10^{-2}	631
GT10	2.2 x 10^5	1.0 x 10^{-1}	458

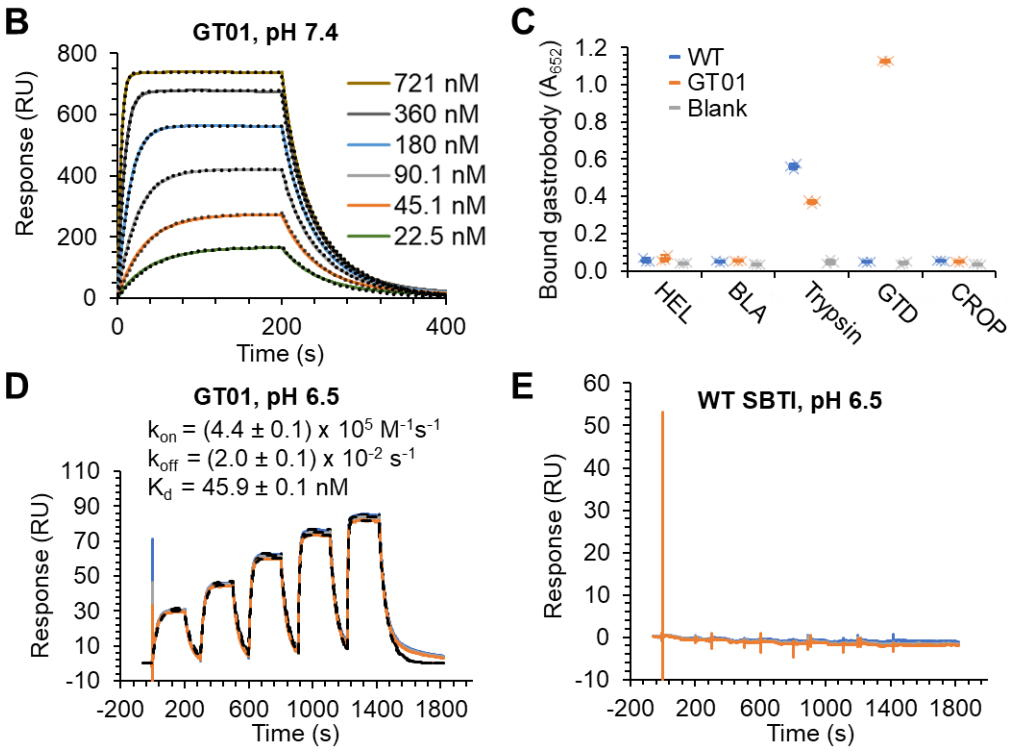


Figure 5.5. Analysis of anti-GTD gastrobody binding to GTD. (A) Binding kinetics of gastrobodies binding to GTD. Binding of gastrobodies to immobilised GTD was analysed by multi-cycle SPR at pH 7.4 and 25 °C. Traces were fit to a 1:1 binding model to obtain a dissociation rate constant (k_{off}) and an association rate constant (k_{on}). Experiments involved six concentrations of each gastrobody. All gastrobodies except GT01 were analysed once. Data for GT01 are mean ± 1 s.d. from three separate experiments. **(B)** Representative SPR trace of GT01 binding to GTD at pH 7.4 and 25 °C. Data were fit to a 1:1 binding model (dotted line). **(C)** Binding specificity of GT01. GT01, WT SBTI or blank buffer (PBS pH 7.4) were incubated in wells coated with lysozyme (HEL), β-lactamase, trypsin, GTD or CROP. Antigen-bound gastrobodies after washes were detected with anti-SBTI antibody (mean ± 1 s.d, n = 3 with individual data points as crosses). **(D)** GT01 binding to GTD at pH 6.5 and 25 °C. GT01 binding to GTD was analysed by single-cycle SPR. Data were fit to a 1:1 binding model (striped line). Dissociation and rate constants are marked (mean ± 1 s.d, n = 3). **(E)** Single cycle SPR of WT SBTI binding to GTD at pH 6.5 and 25 °C. N = 3.

5.2.3. Effect of GT01 on GTD hydrolytic activity

C. difficile toxins enter mammalian cells by receptor-mediated endocytosis (LaFrance et al., 2015). Following entry into the cytosol, the glucosyltransferase domain of TcdA/B glucosylate Rho GTPases, thereby disrupting the cytoskeleton (Fig. 5.6A) (Chaves-Olarte et al., 1997; Ciesla and Bobak, 1998; Voth and Ballard, 2005). Disruption of the cytoskeleton leads to cell death and break down of the intestinal epithelial barrier (Voth and Ballard, 2005). TcdB activity can be measured *in vitro,* because in the absence of an acceptor protein, GTD will hydrolyse UDP-glucose (Fig. 5.6A) (Ciesla and Bobak, 1998). I tested the effect of GT01 binding on the hydrolytic activity of GTD (Fig. 5.6B). Serial dilutions of GT01 were pre-incubated with GTD and UDP-glucose, before measuring release of UDP from UDP-glucose in a luminescent assay (Fig. 5.6B). I observed a dose-dependent reduction of GTD hydrolytic activity with increasing concentrations of GT01, but no effect of different concentrations of WT SBTI (Fig. 5.6B).

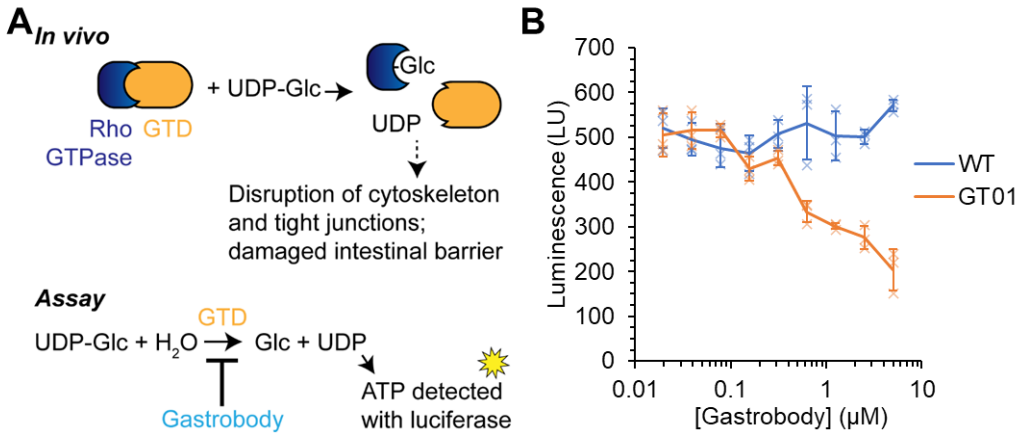


Figure 5.6. Effect of GT01 on GTD hydrolytic activity. (A) Schematic of GTD activity *in vivo* and an *in vitro* assay. TcdB GTD (orange) monoglucosylates Rho GTPases (blue), contributing to *C. difficile* pathogenesis. In the absence of an acceptor protein GTD hydrolyses UDP-glucose. Hydrolytic activity of GTD can be assayed by monitoring release of free UDP. **(B)** Inhibition of GTD hydrolytic activity by GT01. Serial dilutions of GT01 and WT SBTI were incubated with GTD and UDP-glucose. Release of free UDP was monitored by luminescence (mean ± 1 s.d., n = 3 with individual data-points as crosses).

5.2.4. Thermal stability of anti-GTD gastrobodies

GT01 has two amino acid insertions in each GDR compared to WT SBTI or the anti-CROP gastrobodies. I tested the thermal stability of GT01 in a functional assay (Fig. 5.7A) and by DSC (Fig. 5.7C,D). GT01 was heated to 37, 55, 75 or 100 °C for 10 min and aggregates were pelleted by centrifugation after cooling. There was little loss of bound GT01 after heating to 55 °C compared to 37 °C, but about 3-fold less GT01 was bound to GTD after heating to 75 °C (Fig. 5.7A). Only some of the loss of bound GT01 was due to aggregation, because 57% of GT01 remained soluble after heating to 75 °C compared to 37 °C (Fig. 5.7B).

GT01 had a lower T_m at pH 7.4, 2.0 and 1.2 compared to WT SBTI (Fig. 5.7C,D). The largest difference was at pH 7.4 where the T_m of WT SBTI was 67.3 °C (Fig. 5.7C) and the T_m of GT01 was 59.3 °C (Fig. 5.7D). Both GT01 and WT SBTI were less stable at lower pH, with the T_m shifting down several degrees Celsius. There was only a very small downward shift (0.3 °C) in melting temperature of GT01 from pH 2.0 to pH 1.2 (Fig. 5.7D), while WT SBTI shifted by 4.2 °C from 60.4 °C (pH 2.0) to 56.2 °C (pH 1.2) (Fig. 5.7C).

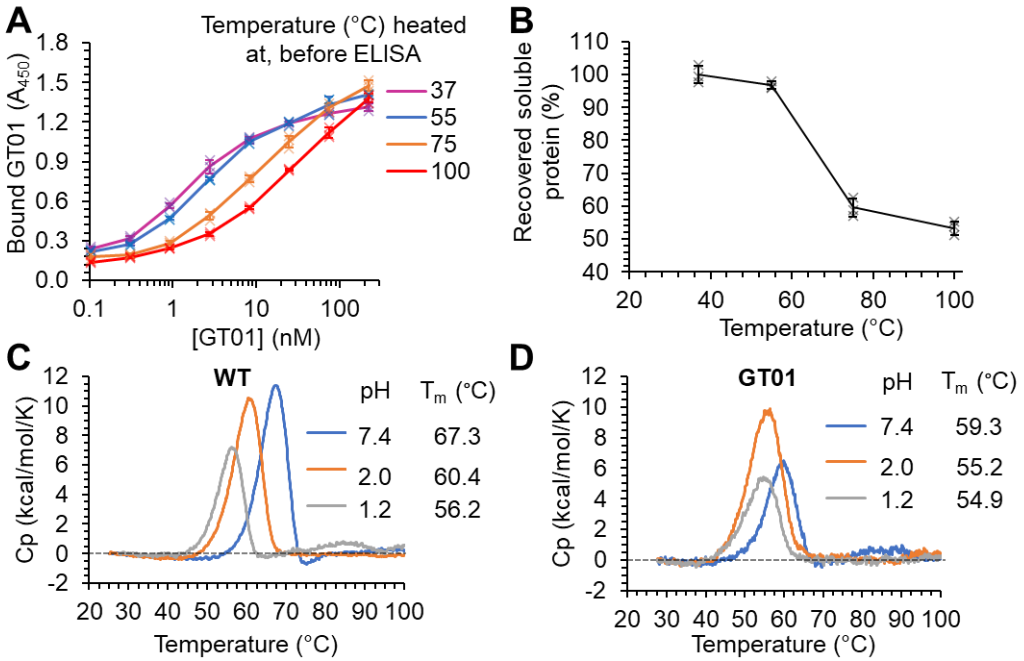


Figure 5.7. Temperature stability of GT01. (A) Temperature resilience of GT01. GT01 was heated to 37, 55, 75 or 100 °C for 10 min before cooling and pelleting aggregates by centrifugation. Binding to trypsin at 25 °C was analysed by ELISA (mean ±1 s.d., n = 3 with individual data-points as crosses). **(B)** Heat-induced aggregation of GT01. GT01 was heated and cooled as in (A). Soluble protein was separated by SDS-PAGE, stained with Coomassie, and quantified by gel densitometry (mean ±1 s.d., n = 3 with individual data-points as crosses). **(C,D)** Thermostability of gastrobodies at low pH. WT SBTI (C) or GT01 (D) was analysed by DSC in phosphate buffer at pH 7.4, 2.0 or 1.2. Melting temperatures (T_m) are indicated. N = 1.

5.2.5. Pepsin-resistance of anti-GTD gastrobodies

Considering that GT01 was reasonably heat stable at low pH, I tested the pepsin-resistance of GT01 (Fig. 5.8). First, I incubated GT01 with 1 mg/mL pepsin for 30 min at pH 2.2 and 37 °C (Fig. 5.8A,B). Intact GT01 was quantified from Coomassie-stained SDS-PAGE. GT01 was less stable than WT SBTI, but 39% of GT01 remained after 30 min with 1 mg/mL pepsin (Fig. 5.8A,B).

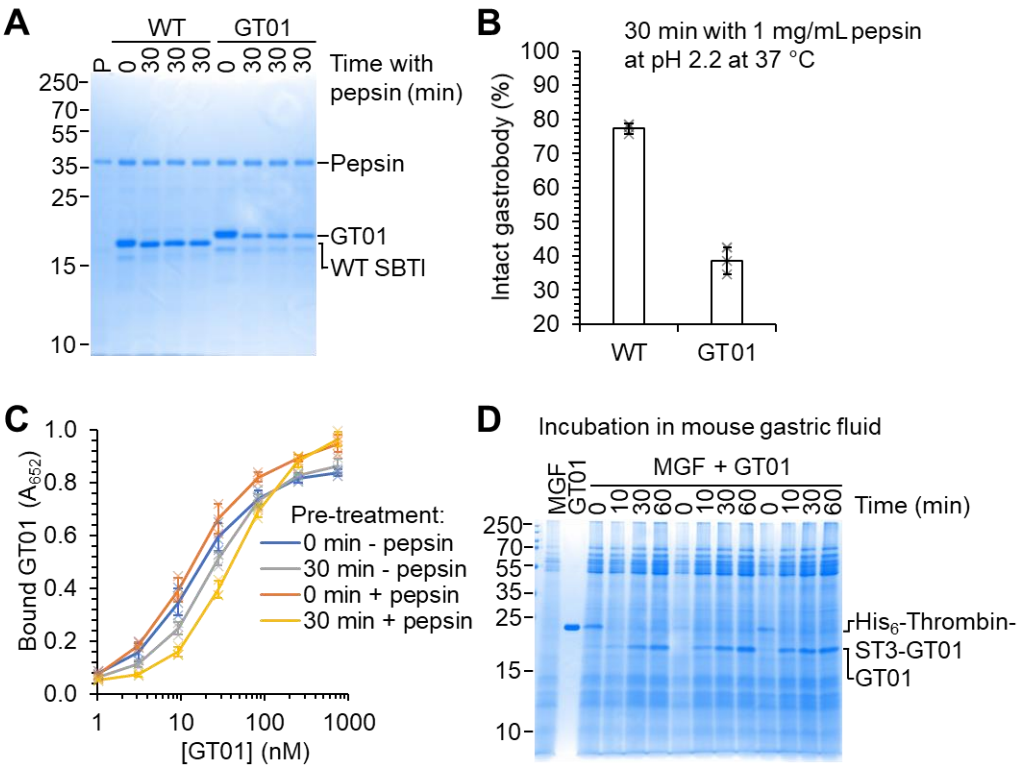


Figure 5.8. Stability of GT01 to stomach-like conditions. (A) Pepsin stability of GT01. WT SBTI and GT01 were incubated for 30 min with 1 mg/mL pepsin at pH 2.2 and 37 °C. Digestion was stopped by boiling in SDS-loading dye, before running SDS-PAGE and staining with Coomassie. P = pepsin alone. **(B)** Quantification of GT01 pepsin resistance from gel in (A). Intensity of WT SBTI or GT01 bands from 30 min with pepsin were quantified by gel densitometry. Intensity at 0 min (immediate stopping of digestion) was set to 100% (mean ±1 s.d., n = 3 with individual data-points as crosses). **(C)** GT01 binding after pepsin digestion. GT01 was digested with 0 or 1 mg/mL pepsin at pH 2.2 and 37 °C for 30 min before neutralising to stop digestion. Binding of digested GT01 to GTD at pH 7.4 was analysed by ELISA (mean ±1 s.d., n = 3 with individual data-points as crosses). **(D)** Stability of GT01 in mouse gastric fluid (MGF). His_6-SpyTag003-GT01 was incubated with mouse gastric fluid at 37 °C. At each time-point a sample was taken and boiled in SDS-loading dye to stop digestion.

Resistance of GT01 to pepsin digestion by SDS-PAGE was corroborated by a functional test (Fig. 5.8C). Binding of GT01 to pepsin after exposure to pH 2.2 with and without pepsin was measured in an ELISA (Fig. 5.8C). 3-fold less GT01 bound to GTD after 30 min in the presence of 1 mg/mL pepsin compared to 0 min (immediate quenching of digestion). Some loss of binding of GT01 binding was also observed

when incubating GT01 at pH 2.2 and 37 °C but without pepsin before analysis by ELISA at pH 7.4 (Fig. 5.8C).

5.2.6. Stability of GT01 in mouse gastric fluid

Digestion of proteins in the GI tract is a variable and complex process, influenced by fluctuations in pH, presence of mucus and protein load (Wen et al., 2002). Hence, I tested the stability of GT01 in mouse gastric fluid (Fig. 5.8D). GT01 underwent a band shift but then formed a stable product, likely due to loss of the N-terminal tags (Fig. 5.8D).

5.2.7. Protease cleavage sites in GT01

I analysed the cleavage sites in GT01 by ESI-MS (Fig. 5.9). Consistent with the cleavage pattern observed for T12, pepsin, trypsin and chymotrypsin cut in the N-terminal SpyTag003 of the GT01 construct. The peak of the trypsin digest product was 18 Da heavier than expected, indicative of an additional cleavage site (Fig. 5.9C). No additional peak was found that indicates where the second trypsin cleavage site occurs. GT01 has one Arg in GDR1 and two Arg in GDR2, which are potential cleavage sites for trypsin. WT SBTI has one Arg in GDR2. No peak matching a GT01 digest product was identified in the ESI-MS spectrum after incubation with elastase (Fig. 5.9E).

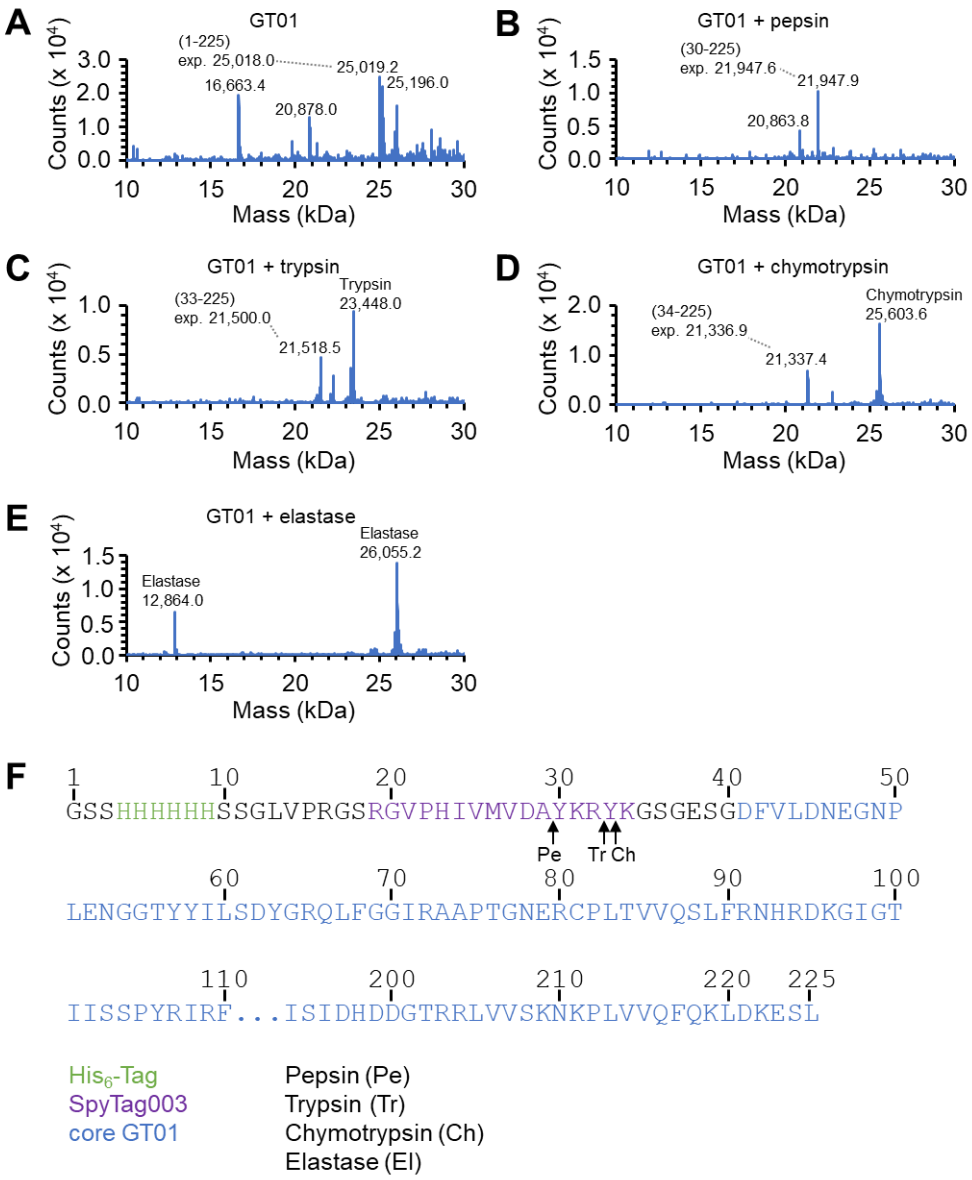


Figure 5.9. Protease cut sites in GT01 gastrobody. His_6-SpyTag003-GT01 construct was incubated in PBS **(A)** or with 1 mg/mL pepsin **(B)**, 100 U/mL trypsin **(C)**, 25 U/mL chymotrypsin **(D)** or 10 U/mL elastase **(E)** for 15 min at 37 °C. Digest products were analysed by ESI-MS. Expected mass of construct fragments are marked. **(F)** The observed cut sites in the amino acid sequence of the His_6-SpyTag003-GT01 construct are marked with arrows. I performed sample preparation and data analysis. Dr Anthony Tumber loaded and ran the mass spectrometer.

5.3. Expanding the gastrobody binding surface

Engineered antibodies and protein scaffolds frequently achieve single digit nanomolar and sub-nanomolar affinities (Hackel et al., 2008; Vazquez-Lombardi et al., 2015;

Zahnd et al., 2007). As an alternative to further expanding the length of GDR1/2, I decided to explore additional candidate loops originally identified in the computational analysis of SBTI. Loop A (N6 – N9) had a favourable mutability score (1.3, Fig. 4.2B) but was missing a mean ΔREU score for G8 because the residue was not solvent accessible. Loop I (S124 – F128) consists of five consecutive residues but the mean ΔREU was 3.0, above the criteria of < 2.0. The variance of the mean ΔREU of loop I, however, was high: ΔREU of D126 was 10 while the remaining four residues were below 2.0 (Fig. 4.2B). I decided to explore the mutability of loop A (termed GDR3) and loop I (termed GDR4) experimentally. GDR3/4 are close to GDR1/2 in space (Fig. 5.10A,B).

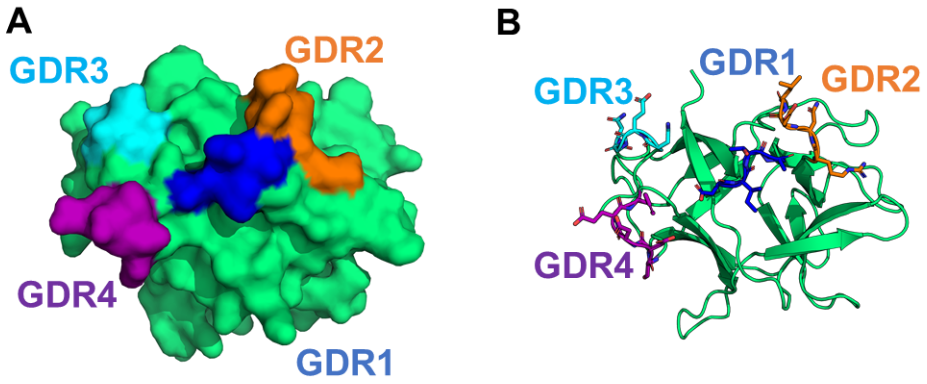
5.3.1. MS spectra confirm expression of GDR3/4 alanine mutants

I cloned alanine mutants of the GDR3/4 residues in pET28a with a C-terminal His_6-tag. The GDR3/4 alanine mutants were expressed in T7SHuffle and purified by Ni-NTA. All mutants expressed correctly and formed the two expected disulfide bonds (Fig. 5.10C).

5.3.2. Pepsin resistance of GDR3/4 alanine mutants

Alanine mutants in GDR3/4 were generally stable in the presence of 1 mg/mL pepsin (Fig. 5.11A). I incubated the alanine mutants with 1 mg/mL pepsin for 30 min at pH 2.2 and 37 °C, before measuring intact protein by SDS-PAGE with Coomassie staining (Fig. 5.11A). D125A was the most susceptible mutant with 33% remaining after 30

min, compared to 79% of WT SBTI, or 97% of the most stable mutant (S124A). Alanine mutants in GDR4 were more stable than mutants in GDR3 (Fig. 5.11A).



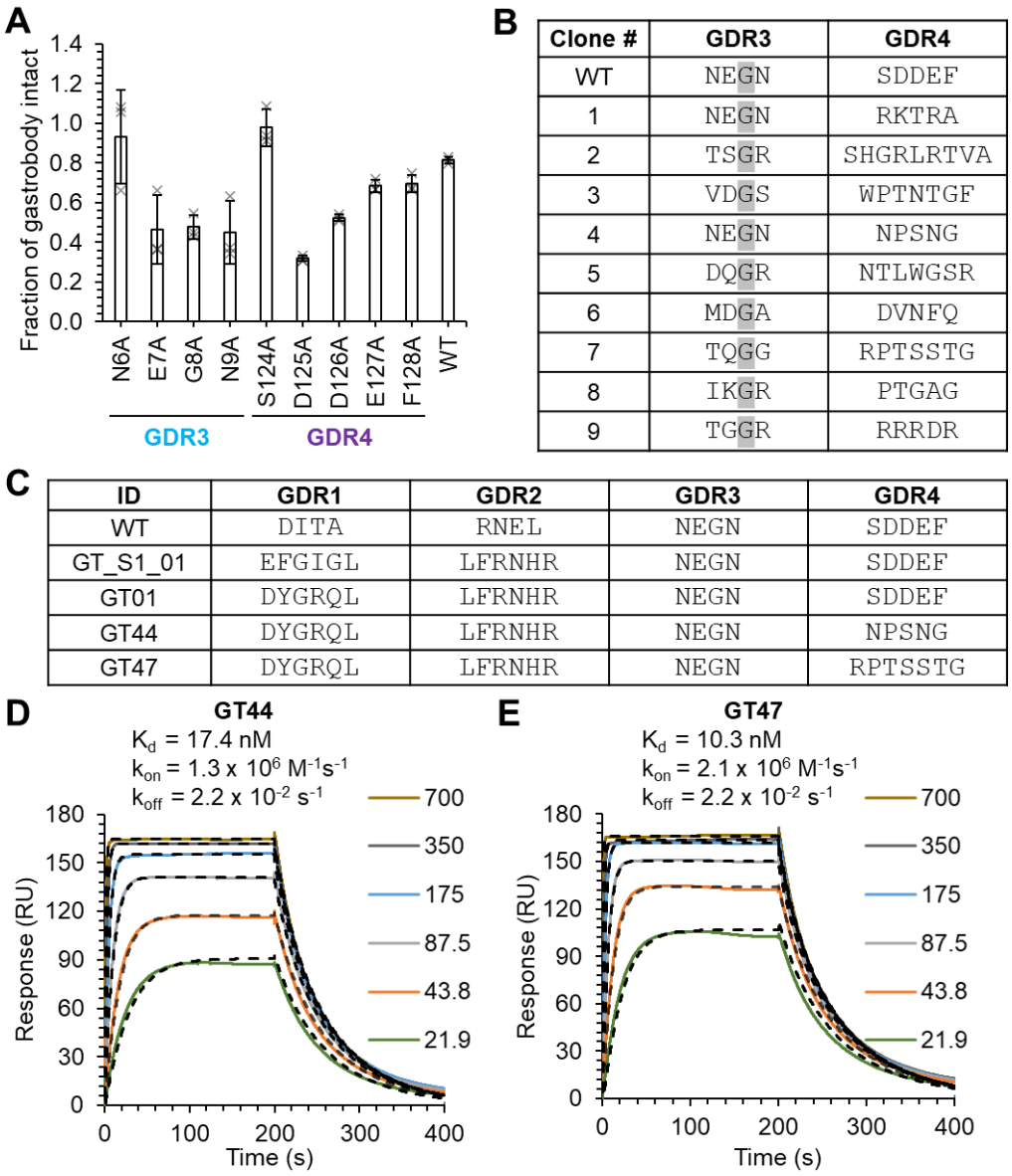
C

Protein	Expected mass with disulfides formed (Da)	Observed Mass (Da)	Observed – Expected Mass (Da)
SBTI N6A	21,213.7	21,214.7	+1.0
SBTI E7A	21,198.7	21,198.7	0.0
SBTI G8A	21,270.7	21,269.3	-1.4
SBTI N9A	21,213.7	21,214.7	+1.0
SBTI S124A	21,240.7	21,241.0	+0.3
SBTI D125A	21,212.7	21,211.9	-0.8
SBTI D126A	21,212.7	21,211.9	-0.8
SBTI E127A	21,198.7	21,198.8	+0.1
SBTI F128A	21,180.6	21,180.6	0.0

Figure 5.10. Extending the gastrobody binding surface. (A) Surface view of SBTI (PDB ID: 1AVU). GDR1 (blue), GDR2 (orange), GDR3 (cyan) and GDR4 (purple) are marked. **(B)** Cartoon view of SBTI. GDRs are shown as sticks. Colouring is as for (A). **(C)** ESI-MS of GDR3/4 alanine mutants.

5.3.3. Phage display for affinity maturation of GT01 with randomised GDR3/4

Encouraged that mutating individual residues in GDR3/4 to alanine does not lead to complete loss of pepsin-resistance, I explored mutability of GDR3/4 by phage display. I cloned a new naïve library with NNK-randomised GDR3/4, using GT01 fused to full-length pIII as template. Any combination of four or five residues in GDR3 and five, seven or nine residues in GDR4 were randomised. The library contained 3.3 x 10^8 variants and construction was validated by sequencing ten clones.



Clone #	GDR3	GDR4
WT	NEGN	SDDEF
1	NEGN	RKTRA
2	TSGR	SHGRLRTVA
3	VDGS	WPTNTGF
4	NEGN	NPSNG
5	DQGR	NTLWGSR
6	MDGA	DVNFQ
7	TQGG	RPTSSTG
8	IKGR	PTGAG
9	TGGR	RRRDR

ID	GDR1	GDR2	GDR3	GDR4
WT	DITA	RNEL	NEGN	SDDEF
GT_S1_01	EFGIGL	LFRNHR	NEGN	SDDEF
GT01	DYGRQL	LFRNHR	NEGN	SDDEF
GT44	DYGRQL	LFRNHR	NEGN	NPSNG
GT47	DYGRQL	LFRNHR	NEGN	RPTSSTG

Figure 5.11. Exploring mutability of GDR3/4. (A) Pepsin-resistance of GDR3/4 alanine mutants. Alanine mutants were incubated with 1 mg/mL pepsin at pH 2.2 and 37 °C for 30 min. Digestion was stopped by boiling in SDS-loading dye. Intact protein was quantified from Coomassie-stained SDS-PAGE with gel densitometry. Samples where digestion was immediately stopped (t = 0 min) were set to 1 (mean ±1 s.d., n = 3 with individual data-points as crosses). **(B)** GDR3/4 sequences from hits of selection against GTD. Nine clones from the fourth round of selection were sequenced. Consensus residues are highlighted in grey. **(C)** Summary of GDR sequences of anti-GTD gastrobodies. **(D,E)** GT44 (D) and GT47 (E) binding to immobilized GTD. Multi-cycle SPR analysis was performed at pH 7.4 and 25 °C. Data were fit to a 1:1 binding model (striped line) to obtain rate and dissociation constants. N = 1.

I selected binders against GTD through four rounds of affinity maturation, with increasing stringency each round. In the final round the bait concentration was 1 nM,

the on-phase 10 min, and included three 1 h off-rate washes with excess non biotinylated bait at 37 °C. After four rounds of affinity maturation selection with the GDR3/4 library against biotinylated GTD, I sequenced nine clones (Fig. 5.11B). Two clones had the WT GDR3 sequence (NEGN) and all clones had Gly in position three at GDR3. Four of the nine clones had Arg in position four of GDR3. All the hit GDR3 sequences consisted of four amino acids - no five amino acid GDR3 was found (Fig. 5.11B). Clones with five, seven and nine amino acids in GDR4 were identified. No obvious consensus motif emerged in the selections (Fig. 5.11B).

5.3.4. Binding kinetics of GT44 and GT47 binding to GTD

I cloned the GDR4 sequence of clone 4 (GT44) and clone 7 (GT47) from the selections with the GDR3/4 library into His_6-SpyTag003-GT01 (Fig. 5.11C). After expression and purification by Ni-NTA and SEC, I analysed the binding kinetics of GT44 and GT47 to immobilised GTD by SPR (Fig. 5.11D,E). The affinity of GT47 (K_d = 10.3 nM) was approximately 8-fold better than GT01 (K_d = 85 ± 2.3 nM). Driven by improvements of on- and off-rate, both GT44 and GT47 bound more tightly than GT01 (Fig. 5.11C,D; Fig. 5.5A,B).

5.3.5. Heat-resilience of GT44 and GT47

The heat resilience of GT44 and GT47 was tested with an ELISA (Fig. 5.12A,B). Binding to GTD was analysed after heating to 37, 55, 75 or 100 °C for 10 min. Both GT44 and GT47 showed minimal loss of GTD-bound protein after heating to 55 °C

compared to 37 °C (Fig. 5.12A,B). However, GT47 was more to susceptible heat-induced loss of binding than GT44 when heating to 75 °C (Fig. 5.12A,B).

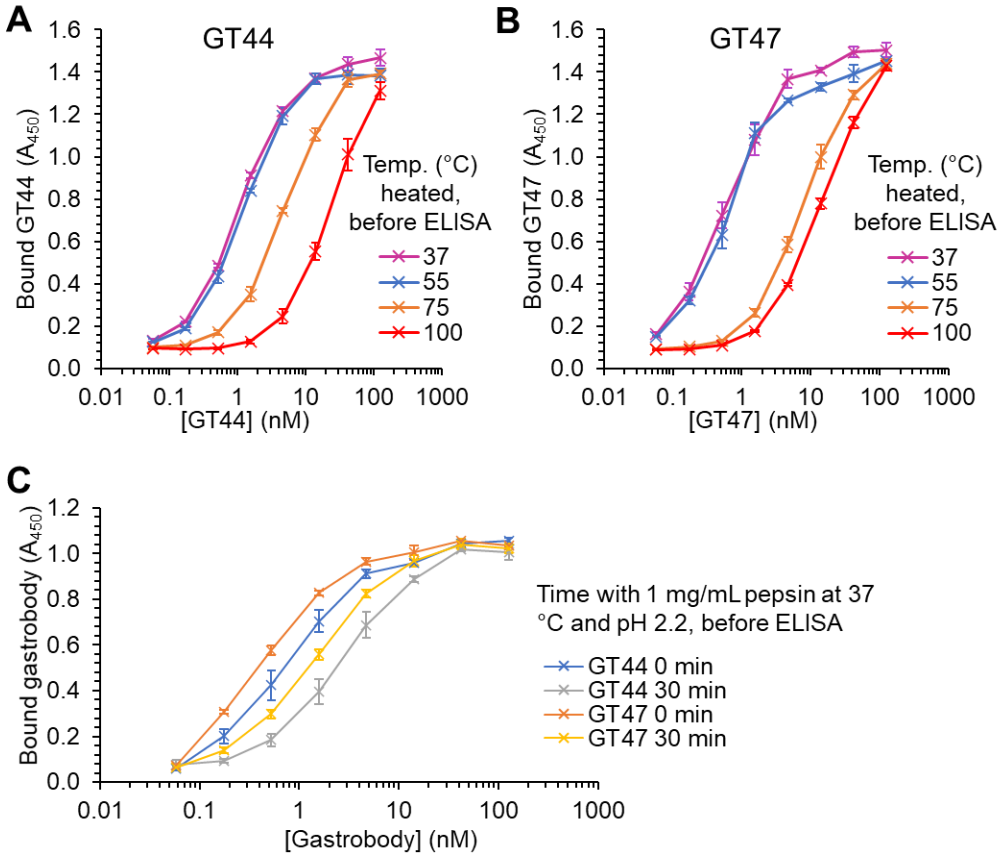


Figure 5.12. Pepsin- and heat-stability of GT44 and GT47. (A,B) Binding of GT44 (A) or GT47 (B) to GTD after heating to 37, 55, 75 or 100 °C for 10 min. Aggregates were pelleted after heating. Binding of soluble GT44 or GT47 to GTD was analysed by ELISA at 25 °C (mean ±1 s.d., n = 3). **(C)** Pepsin-stability of GT44 and GT47. Gastrobodies were incubated with 1 mg/mL pepsin at pH 2.2 and 37 °C for 0 or 30 min. Digestion was stopped by neutralisation. Binding to GTD was measured by ELISA at 25 °C (mean ±1 s.d., n = 3).

5.3.6. Pepsin-resistance of GT44 and GT47

Lastly, I subjected GT44 and GT47 to a test of pepsin-resistance (Fig. 5.12C). I incubated the anti-GTD gastrobodies with 1 mg/mL pepsin at pH 2.2 and 37 °C for 30 min, before neutralising and measuring binding at pH 7.4. 3-fold less GT44 and GT47 bound GTD after 30 min with 1 mg/mL pepsin, consistent with findings for GT01 (Fig. 5.12C; Fig. 5.8C).

5.4. Removing trypsin binding of GT01

SBTI's trypsin inhibition activity might be detrimental to some applications. The scissile R63 of SBTI is a key residue for trypsin binding (Song and Suh, 1998; Sweet et al., 1974). The GDRs are on the opposite face of SBTI to R63 (Fig. 5.13A). I cloned GT01 R63A to remove trypsin binding of GT01.

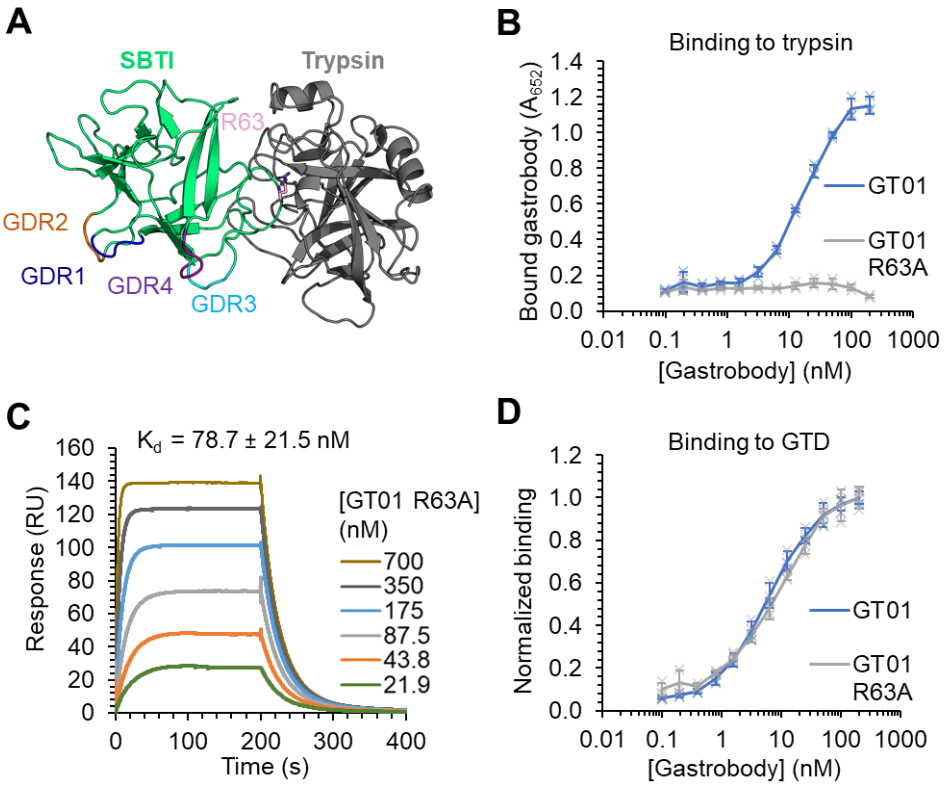


Figure 5.13. Removing trypsin-binding of GT01. (A) SBTI in complex with porcine trypsin (PDB ID: 1AVX). The scissile R63 is marked in stick format in pink. GDR1 (blue), GDR2 (orange), GDR3 (cyan) and GDR4 (purple) are coloured. **(B)** Binding of GT01 and GT01 R63A to trypsin. Gastrobodies bound to trypsin-coated wells were detected with biotinylated GTD and streptavidin-HRP in a plate assay (mean ±1 s.d., n = 3 with individual data-points as crosses). **(C)** Representative SPR trace of GT01 R63A binding to immobilized GTD at pH 7.4 and 25 °C. Data were fit to a 1:1 equilibrium binding model. Dissociation constant (K_d) was calculated from three separate experiments (mean ±1 s.d., n = 3). **(D)** GT01 and GT01 R63A binding to GTD. Binding of gastrobodies to GTD was analysed by ELISA. Bound gastrobody was detected by anti-SBTI antibody. Absorbance values were normalized to the absorbance at the highest concentration of each gastrobody (mean ±1 s.d., n = 3 with individual data-points as crosses).

5.4.1. Trypsin binding of GT01 R63A

I also compared GT01 R63A and GT01 trypsin binding (Fig. 5.13B). GT01 bound to trypsin with an apparent dissociation constant in the nanomolar range, but no binding of GT01 R63A to trypsin was observed (Fig. 5.13B).

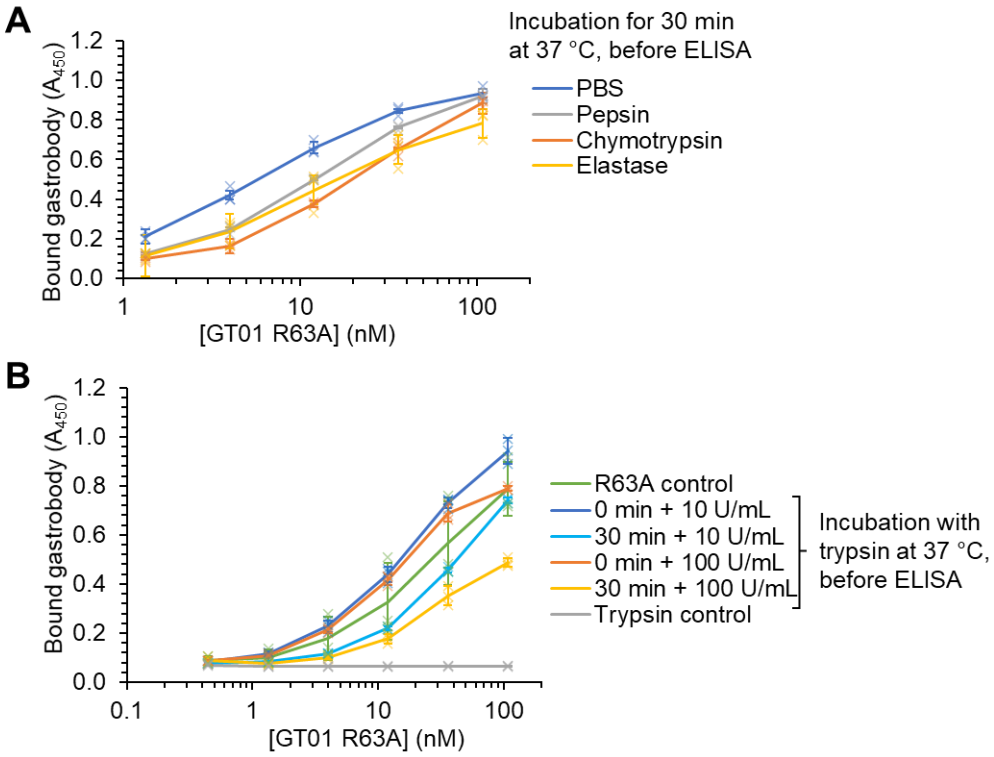


Figure 5.14. Protease-stability of GT01 R63A. (A) Binding of GT01 R63A to GTD after pepsin, chymotrypsin and elastase digestion. GT01 R63A binding to GTD was analysed by ELISA at pH 7.4 following digestion at 37 °C for 30 min with 1 mg/mL pepsin at pH 2.2, or with 25 U/mL chymotrypsin or 10 U/mL elastase at pH 6.8. GT01 R63A incubation in PBS without protease was included as control (mean ±1 s.d., n = 3 with individual data-points as crosses). **(B)** GT01 R63A binding to GTD after trypsin exposure. GT01 R63A was incubated with 10 or 100 U/mL trypsin at pH 6.8 and 37 °C, before analysing binding to GTD at 25 °C by ELISA. Controls of GT01 R63A only (green) or trypsin only (grey) were incubated for 30 min at 37 °C (mean ±1 s.d., n = 3 with individual data-points as crosses).

5.4.2. GT01 R63A binding kinetics to GTD

I tested the binding kinetics of GT01 R63A to GTD (Fig. 5.13C,D). The K_d of GT01 R63A binding to GTD (78.7 ± 21.5 nM, Fig. 5.13C) was not substantially different from

the K_d of GT01 binding to GTD (85 ± 2.3 nM, Fig. 5.5A,B). Consistent with the SPR analysis, no difference in apparent binding affinity was observed in an ELISA comparing GT01 and GT01 R63 binding to GTD (Fig. 5.13D).

5.4.3. GT01 R63A binding activity after protease digestion

The R63A mutation may impact the protease stability of GT01. I pre-incubated GT01 R63A with physiological concentrations of pepsin (1 mg/mL) at pH 2.2, or with chymotrypsin (25 U/mL) or elastase (10 U/mL) at pH 6.8 and 37 °C for 30 min before testing binding to GTD at pH 7.4 (Fig. 5.14A). After incubation with each protease, approximately 3-fold less GT01 R63A bound to GTD compared to a control kept in PBS (Fig. 5.14A). GT01 R63A was most susceptible to trypsin of all the proteases tested: pre-incubation with intestinal concentrations of trypsin (100 U/mL) for 30 min at pH 6.8 and 37 °C caused 4-fold loss of bound GT01 R63A compared to the 0 min time point (Fig. 5.14B). Pre-incubation with 10 U/mL trypsin for 30 min reduced bound GT01 R63A 3-fold compared to the 0 min time point (Fig. 5.14B).

5.4.4. Protease cleavage sites in GT01 R63A

As for SBTI, T12 and GT01, I used ESI-MS to map the protease cleavage sites in GT01 R63A (Fig. 5.15). Consistent with data for GT01, the main trypsin digestion peak of GT01 R63A was 18 Da heavier than expected indicative of a second cleavage event (Fig. 5.15C). I identified a second peak corresponding to Q66 – L225 of GT01 R63A, but this peak, too, was 18 Da heavier than expected. I was unable to identify where the additional trypsin cut sites were. Pepsin, chymotrypsin, and elastase cut in the N-

terminal SpyTag003 (Fig. 5.15B,D-G). Elastase also removed the C-terminal Leu, previously observed for T12 (Fig. 5.15E,F; Fig. 4.11A).

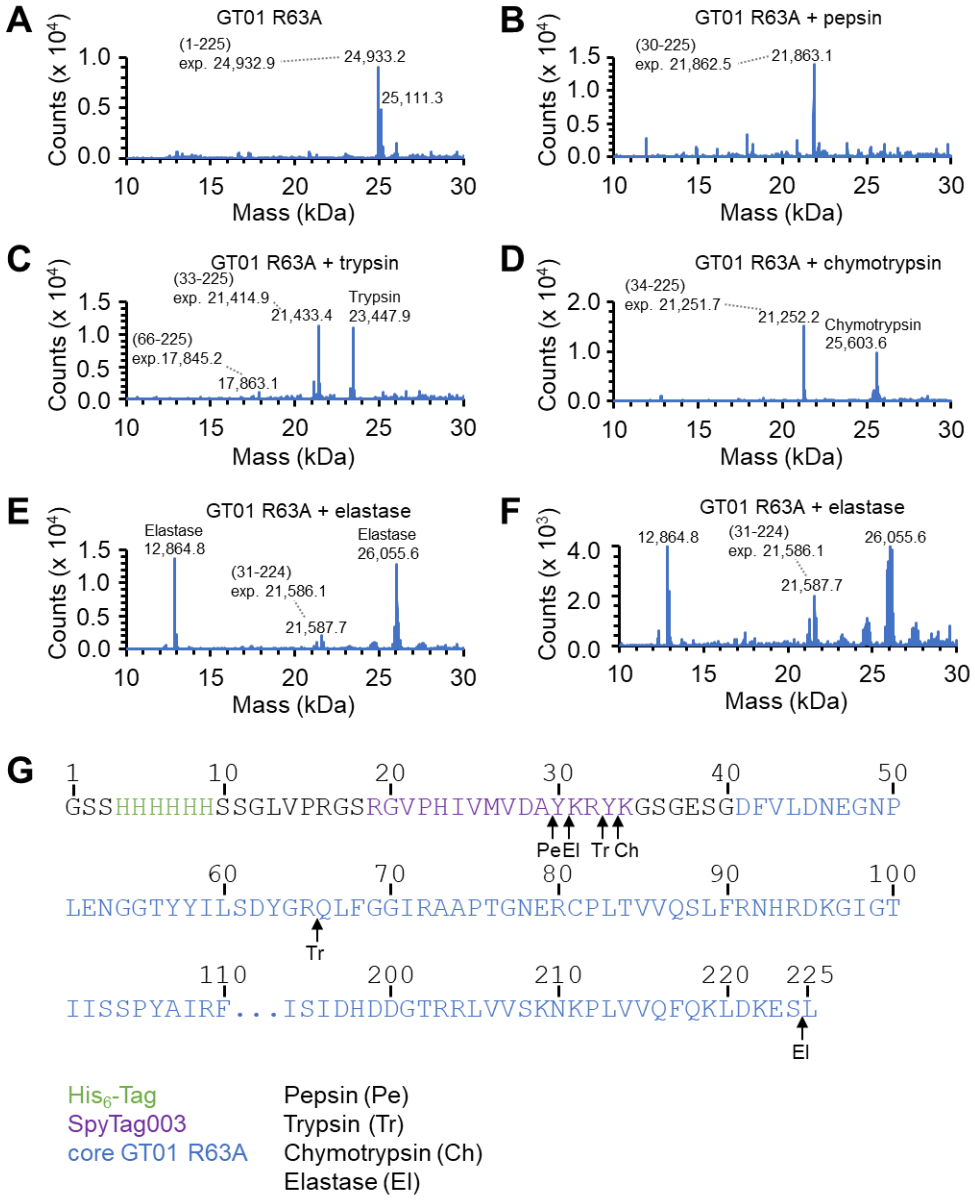


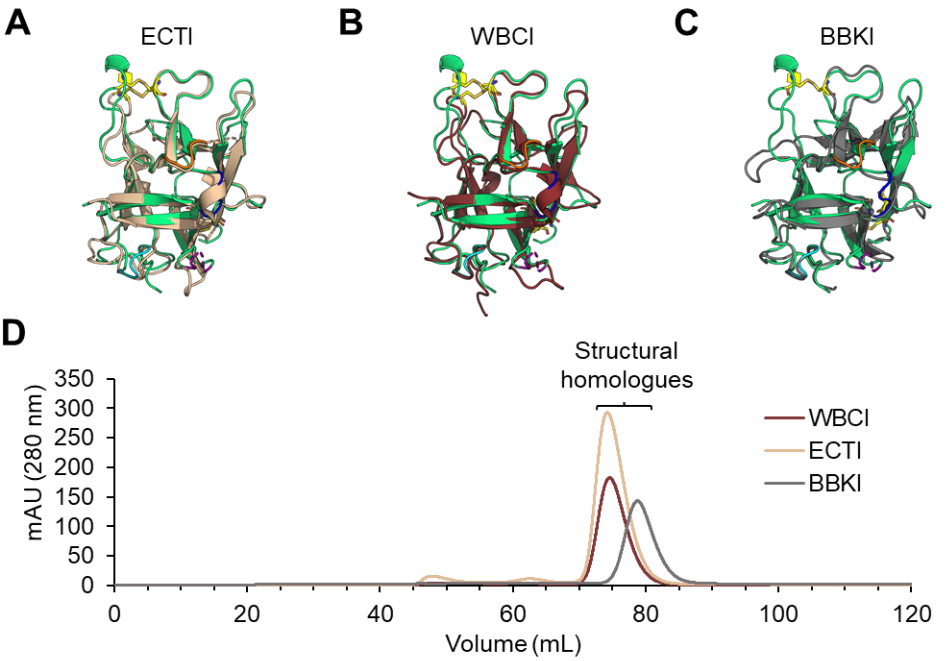
Figure 5.15. Protease cut sites in GT01 R63A. His_6-SpyTag003-GT01 R63A construct was incubated in PBS **(A)** or with 1 mg/mL pepsin **(B)**, 100 U/mL trypsin **(C)**, 25 U/mL chymotrypsin **(D)** or 10 U/mL elastase **(E,F)** for 15 min at 37 °C. Digest products were analysed by ESI-MS. Expected mass of construct fragments are marked. (F) is a zoom-in of data in (E). **(G)** The observed cut sites in the amino acid sequence of the His_6-SpyTag003-GT01 R63A construct are marked with arrows. I performed sample preparation and data analysis. Dr Anthony Tumber loaded and ran the mass spectrometer.

5.5. Alternative gastrobody backbones

Proteins in the β-trefoil family share substantial structural homology but little sequence identity (Guerra et al., 2016) (Fig. 5.16A-C). To better understand the properties of the β-trefoil fold, I decided to explore the pepsin- and heat-resilience of structural homologues of SBTI. PDBe Fold (Krissinel and Henrick, 2004) was used to identify three structural homologues of SBTI (PDB ID: 1AVU) (Sweet et al., 1974): the *Erythrina caffra* seed Kunitz-type trypsin inhibitor (ECTI, PDB ID: 1TIE) (Onesti et al., 1991), the *Bauhinia bauhinioides* plasma kallikrein inhibitor (BBKI, PDB ID: 4ZOT) (Zhou et al., 2015), and the winged bean chymotrypsin inhibitor (WBCI, PDB ID: 1EYL) (Ravichandran et al., 2001). The sequence identity of SBTI to ECTI is 42%, 29% to BBKI, and 44% to WBCI (Fig. 5.16A-C). ECTI (Fig. 5.16A) and WBCI (Fig. 5.16B) have two disulfide bonds in analogous positions to SBTI. BBKI has one unpaired cysteine (Fig. 5.16C).

5.5.1. Expression, purification and MS of BBKI, ECTI and WBCI

I cloned ECTI, WBCI and BBKI into pET28a with a C-terminal His_6-tag for expression and purification. WBCI and ECTI were expressed in *E. coli* T7SHuffle to promote disulfide bond formation (Lobstein et al., 2012), while BBKI was expressed in *E. coli* BL21 (DE3) RIPL. The structural homologues were purified by Ni-NTA followed by SEC (Fig. 5.16D). All homologues formed monodisperse peaks by SEC (Fig. 5.16D). ESI-MS confirmed the expression and disulfide bond formation of the proteins (Fig. 5.16E). Yields of the structural homologues were 6.8 (WBCI), 10.7 (ECTI) and 5.2 (BBKI) mg/L culture.



E

Protein	Expected mass with disulfides formed (Da)	Observed Mass (Da)	Observed – Expected Mass (Da)
WBCI	21,403.9	21,404.23	+0.3
ECTI	20,437.7	20,437.78	+0.1
BBKI	19,138.5	19,138.34	-0.2

Figure 5.16. Expression and purification of structural homologues of SBTI. (A-C) PyMOL cartoon alignment of ECTI (PDB ID: 1TIE, beige)(A), WBCI (PDB ID: 1EYL, brown)(B) or BBKI (PDB ID: 4ZOT, grey)(C) with SBTI (PDB ID: 1AVU). SBTI is coloured green. GDR1 (blue), GDR2 (orange), GDR3 (cyan) and GDR4 (purple) are coloured. Cysteines are represented as sticks in yellow. **(D)** SEC trace of WBCI, ECTI and BBKI. Fractions containing the structural homologues are marked. mAU = milli absorbance units. **(E)** ESI-MS of WBCI, ECTI and BBKI.

5.5.2. Pepsin-resistance of structural homologs

I tested the pepsin stability of BBKI, ECTI and WBCI with the benchmark challenge with 1 mg/mL pepsin at pH 2.2 and 37 °C (Fig. 5.17A). In this preliminary test of pepsin stability, the structural homologues were more resistant to digestion than SBTI. WBCI was the most stable with 91% remaining after 120 min with 1 mg/mL pepsin, compared to 59% of SBTI (Fig. 5.17A).

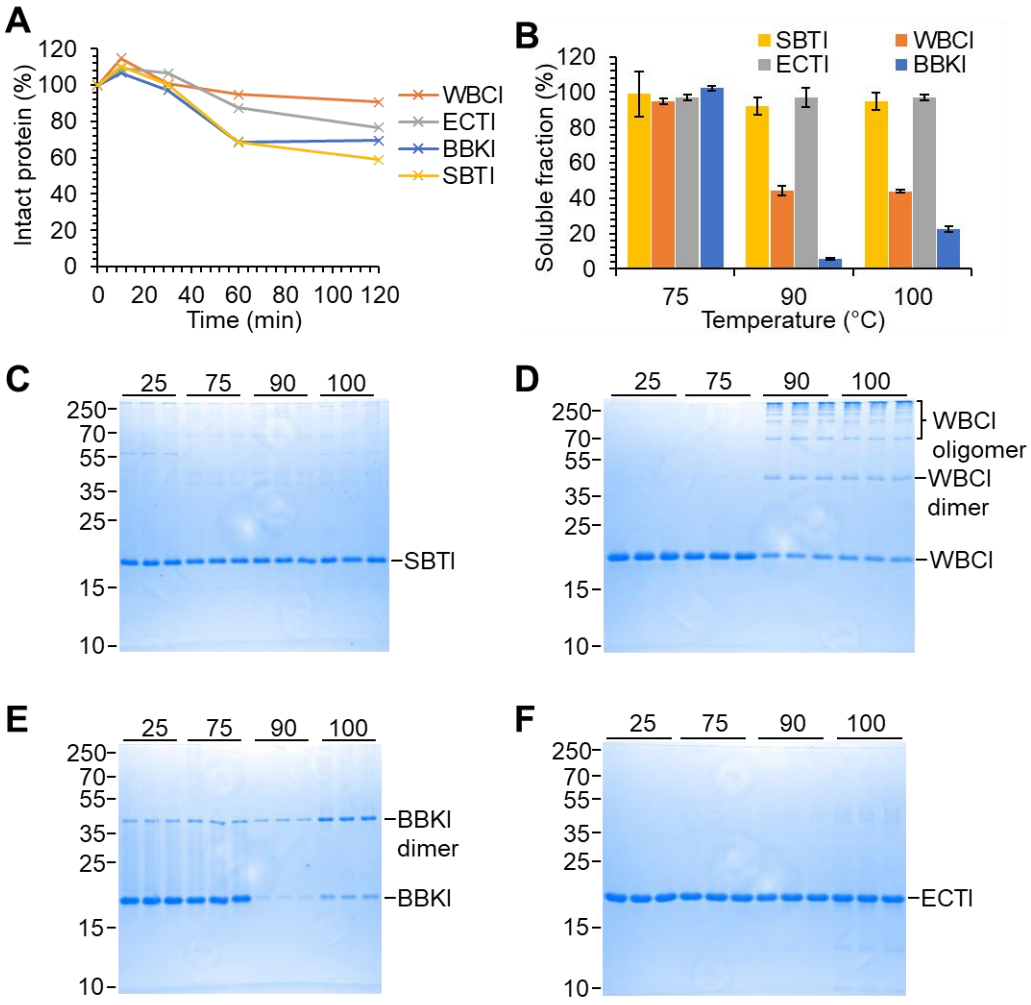
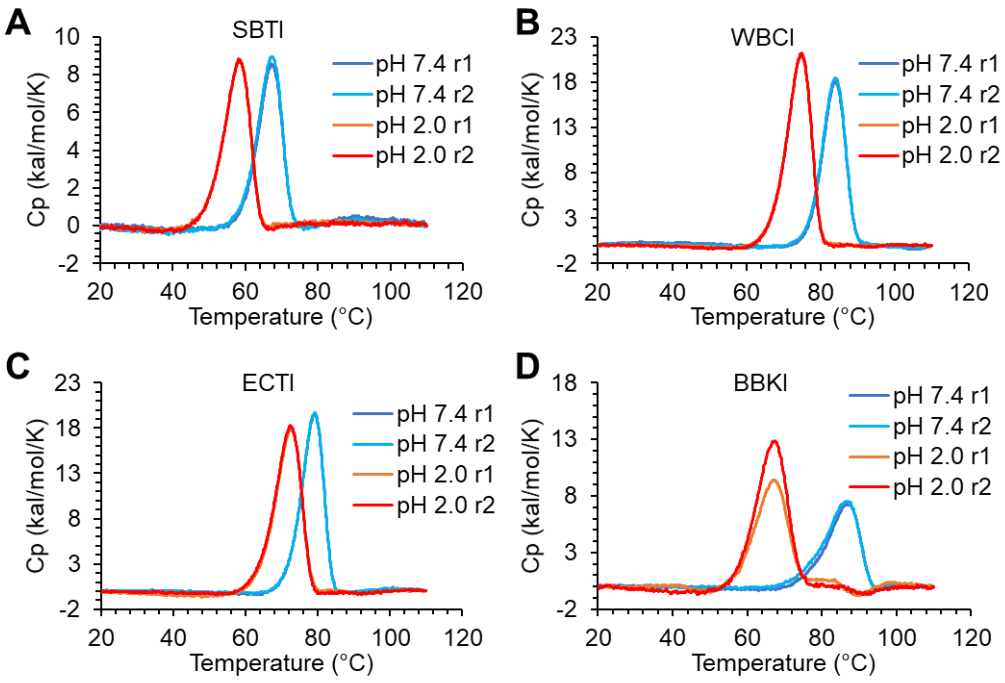


Figure 5.17. Pepsin- and heat-resilience of β-trefoil proteins. (A) Pepsin resistance of structural homologues. SBTI, ECTI, BBKI and WBCI were incubated with 1 mg/mL pepsin at pH 2.2 and 37 °C. Digestion was stopped by boiling in SDS-loading dye. Digested proteins were run on SDS-PAGE, stained with Coomassie, and intact bands were quantified by gel densitometry. Band intensity at t = 0 was set to 100%. N = 1. **(B-F)** Temperature dependent solubility of structural homologues. SBTI (C), WBCI (D), BBKI (E) or ECTI (F) was heated to 25, 75, 90 or 100 °C for 10 min. Aggregated protein was pelleted and the soluble fraction was separated on SDS-PAGE (C-F). Band intensity of monomers was quantified by densitometry (B). Intensity of bands of samples at 25 °C was set to 100% (mean ±1 s.d., n = 3).

5.5.3. Temperature-dependent solubility of structural homologues

SBTI, ECTI, WBCI and BBKI were heated to 75, 90 or 100 °C for 10 min to investigate the heat-induced aggregation tendency of the proteins (Fig. 5.17B-F). After heating,

aggregates were pelleted, and the soluble fraction of the protein monomer was quantified by SDS-PAGE with Coomassie staining and gel densitometry. SBTI and ECTI remained almost completely soluble after 10 min at 100 °C (Fig. 5.17B,C,F). Substantial fractions of soluble monomer of BBKI and WBCI disappeared upon heating above 75 °C (Fig. 5.1B,D,E). Bands indicative of dimer formation were observed for BBKI at all temperatures, increasing in intensity after heating to 100 °C (Fig. 5.17E). WBCI formed dimers and oligomers after heating to 90 or 100 °C (Fig. 5.17D).



E

Protein	pH 7.4		pH 2.0	
	r1	r2	r1	r2
BBKI	86.6	86.8	66.8	67.5
ECTI	79.2	79.2	72.6	72.3
WBCI	84.0	83.9	74.8	74.9
SBTI	67.1	67.1	58.6	58.1

Figure 5.18. Stability of β-trefoil proteins at low pH. SBTI **(A)**, WBCI **(B)**, ECTI **(C)** and BBKI **(D)** were analysed by DSC in phosphate buffer at pH 7.4 or pH 2.0. Two replicates (r1, r2) of each condition were tested (N = 2). **(E)** Melting temperatures (T_m) from DSC traces (°C). T_m from the two replicates are shown (r1, r2).

5.5.4. Thermal stability of structural homologues at low pH

Following the aggregation tests, I used DSC to compare the melting temperatures of SBTI, WBCI, ECTI and BBKI at pH 7.4 and at pH 2.0 (Fig. 5.18). All structural homologues were more stable than SBTI at both pH 7.4 and pH 2.0. BBKI had the highest melting temperature at pH 7.4 (~86 °C, Fig. 5.18D) followed by WBCI (~84 °C, Fig. 5.18B). However, BBKI was also most destabilized by acid, with the T_m shifting by 20 °C to 68 °C at pH 2.0 (Fig. 5.18D). The T_m of ECTI, by contrast, only shifted 7 °C from 79 °C at pH 7.4 to 72 °C at pH 2.0 (Fig. 5.18C).

The features of the structural homologues studied here are summarised in Table 1.

Protein	PDB ID	Q-score	MW (kDa)	Disulfide bonds	Sequence identity to SBTI (%)	Mean T_m (°C)	
						pH 7.4	pH 2.0
BBKI	1EYL	0.63	19.1	0*	29	86.6	67.2
ECTI	1TIE	0.66	20.4	2	42	79.2	72.5
WBCI	4ZOT	0.63	21.4	2	44	84.0	74.9
SBTI	1AVU	1.00	21.3	2	100	67.1	58.4

Table 1. Summary of the structural homologues of SBTI. The quality of alignment score (Q-score) was calculated by PDBeFold taking into account the alignment length and root mean square deviation (RMSD) of aligned Cα-atoms (Krissinel and Henrick, 2004). MW = molecular weight. * = BBKI has one unpaired cysteine. Melting temperature (T_m) was calculated from DSC experiments.

5.6. Discussion

In this chapter, I developed and optimised gastrobody binders to GTD. I hypothesized that it might be easier to identify high affinity binders to GTD, which has an active site cleft (Alvin and Lacy, 2017), unlike CROP which has a relatively flexible, flat, and elongated structure (Kroh et al., 2017; Murase et al., 2014). No gastrobody binders to GTD were identified from the original library with four randomised amino acids in GDR1/2. Selection with the loop extended gastrobody library, where GDR1/2 contains

any combination of five, six or seven amino acids, identified one hit. The consensus hit GT_S1_01 had six amino acids in each loop. Identifying hits with the loop-extended but not the original library may be due the short loop length of the original library preventing access into the active site for high affinity binding. Binding to sites other than the active site is also possible, but a larger binding surface is likely beneficial for bind to flat epitopes, too. Nonetheless, the low micromolar affinity of GT_S1_01 for GTD was still inadequate for practical applications.

No binders were found when selecting after challenge with 0.1 mg/mL pepsin, indicating that further optimisation is required. Indeed, there was no obvious difference in clones selected with 0.1 mg/mL pepsin during affinity maturation selections. This likely reflects that another component of the phage display system, such as the linker between pIII and gastrobody, is more susceptible to pepsin cleavage than the gastrobody. Ideally, selections will be performed with the physiologically relevant pepsin concentration of 1 mg/mL. It will be important to develop a phage system that is as pepsin-resistant as the desired gastrobody, to ensure that the pepsin pressure applied in selection is directed toward the gastrobody.

Affinity maturation of GT_S1_01 resulted in hits with nearly 100-fold improved affinity. GT01 was the tightest binder, consistent with being identified the highest number of times in the selection. The hits in affinity maturation all differed in GDR1 and kept the six amino acid length of GDR1. It is tempting to suggest that six amino acids are optimal for GDR1/2 but the sample size studied here is too small. In addition, the hits here are from affinity maturation, so also reflect the optimal size of binding GTD at this epitope. Furthermore, two residues in GDR1 were retained in all affinity maturation hits and the original GT_S1_01 clone, which may indicate optimisation of binding to a

particular epitope on GTD. GT01 did not bind to the control antigens CROP, BLA or lysozyme. GT01 binding affinity to GTD was not substantially different at pH 6.5 and pH 7.4.

Increasing the binding surface further by selecting with randomised GDR3/4 enabled better binding affinities with GDR1/2 alone. Hits with diversity in both GDR3/4 were identified but I only tested binders with mutated GDR4. Hits with five, seven or nine amino acids in GDR4 were identified, indicating that substantial insertions into GDR4 might be tolerated. A binder with seven amino acids in GDR4 (GT47, K_d = 10.3 nM) bound slightly more tightly than one with five amino acids (GT44, K_d = 17.4 nM). Despite four rounds of selection with off-rate selection steps, the campaign with randomised GDR3/4 did not converge on a consensus binder. Testing of the other clones from round four is needed to determine if all hits bind tightly, or if the stringency of the selections should have been increased further. A clear motif, however, was conservation of the Gly in position three (G8) in GDR3 indicating that mutation of this residue is detrimental to gastrobody folding. SBTI G8A expressed well and kept pepsin resistance. G8 is not solvent accessible: G8 faces inward as a loop folds back on itself close to the N-terminus (Song and Suh, 1998; Sweet et al., 1974). Mutation of G8 to an amino acid with a larger side chain than Gly or Ala could therefore cause misfolding due to steric clashes. Misfolding of SBTI G8 mutants could be exacerbated by the proximity to the N-terminus. The hypothesis that GDR3 is less amenable to mutation than the other GDRs is consistent with the lack of identifying hits with five amino acids (one amino acid extension) in GDR3. Conclusions on the mutability on GDR3 should be validated by separate selections against a different target with a completely naïve library. Next-generation sequencing (NGS) is commonly performed to identify clones with strong increases in relative abundance during selections and enrichment for

motifs at particular residues (Rouet et al., 2018; Villequey et al., 2017). Future selection campaigns should include NGS to better understand the tolerability of extensions and mutations of particular positions in the GDRs.

The resilience of GT01 to heat was reduced compared to WT SBTI. Reduction in heat stability was shown by a reduced T_m at neutral and acidic pH, as well as higher heat-induced aggregation tendency at temperatures above the T_m. Reduced GT01 binding after heating to temperatures above the T_m is consistent with misfolding on cooling. The T_m of some anti-CROP gastrobodies was higher than WT SBTI (see chapter 4). The obvious difference is that anti-CROP gastrobodies did not have amino acid insertions in the GDRs, which could affect the compact structure of SBTI. Despite a higher T_m than GT01, anti-CROP T12 displayed worse refolding after heat-induced unfolding. Both GT01 and T12 lost more trypsin binding than WT SBTI after heat-induced unfolding. These results show that there is a complex interplay of melting temperature, aggregation tendency and refolding capacity. Only one gastrobody (GT01) with extended loops has been analysed for temperature stability, so any conclusions require validation with additional clones. Nanobodies have comparable T_m (60 – 66 °C) compared to gastrobodies but refold well after heat-induced unfolding (Ewert et al., 2002; van der Linden et al., 1999). Aggregation is the main mechanism of nanobody inactivation at high temperatures, which appears to be independent of CDR3 length (Kunz et al., 2018). Nanofitins have a rigid structure with few long, flexible regions but exceptionally high melting temperatures (70 – 100 °C) (Béhar et al., 2014; Kalichuk et al., 2016, 2018). High temperature resilience is beneficial for sterile fill/finish manufacturing procedures of recombinant biologics (Cipriano et al., 2012; Rathore and Rajan, 2008) and feed enzymes (Rebello et al., 2017; Woyengo and Nyachoti, 2011).

Mutation and extension of the GDRs reduced the pepsin-resistance of GT01 compared to WT SBTI. Nonetheless, 39% of GT01 was intact after 30 min with 1 mg/mL pepsin compared to 75% of WT SBTI. The apparent pepsin-resistance of GT01 was higher in mouse gastric fluid. There was no obvious discrepancy between pepsin-resistance observed by SDS-PAGE or ELISA experiments, consistent with DSC experiments indicating that GT01 is folded at physiological temperatures and pH 2. Interestingly, no increase in pepsin susceptibility was observed when GDR4 of GT01 was mutated to generate GT44 or GT47.

ESI-MS analysis identified at least one trypsin cleavage site in the main GT01/GT01 R63A domain. However, not all binding of GT01 R63A to GTD was lost after trypsin digestion. A limited proteolysis approach, as discussed in chapter 4, or mutation of the Arg in GDR1/2 should enable identification of the trypsin cleavage site. Future work should consider the use of biased libraries that exclude Arg or Lys in GDRs.

SBTI (Yamamoto et al., 1994) and trypsin inhibitors (Cristina Oliveira de Lima et al., 2019) have been explored for therapeutic benefits, but the trypsin inhibition activity may not be appropriate for some gastrobody applications. I found that mutating R63A blocks trypsin binding. Mutation of R63A in GT01 did not substantially impact binding to GTD. Indeed, I showed that GT01 can bind trypsin and GTD simultaneously. Blocking of trypsin inhibition did not compromise the protease stability of GT01 R63A: binding to GTD was observed after digestion with pepsin, trypsin, chymotrypsin or elastase. The trypsin, chymotrypsin and elastase resistance of GT01 was not tested. Side-by-side comparison of multiple clones with and without R63A is needed to determine if reducing trypsin binding impairs resistance to proteases.

I found some β-trefoil fold homologues of SBTI to be more pepsin resistant than SBTI. Like SBTI, the structural homologues are all plant protease inhibitors and likely exposed to similar evolutionary pressures as SBTI. I was surprised to find that BBKI had the highest T_m of all the proteins tested at pH 7.4, despite a lack of disulfide bonds. Despite having the highest T_m at neutral pH, BBKI was most destabilized by acid. Unlike the other three β-trefoil proteins, BBKI does not have disulfide bonds. Disulfide bonds may help to resist denaturation at low pH. Protein denaturation in acidic conditions is caused by intramolecular repulsion of protonated amino acids (Fink et al., 1994). The unpaired cysteine of BBKI does, however, lead to homo-dimerization which could be impractical for most applications. *Bauhinia bauhinioides* cruzipain inhibitor (BBCI, PDB ID: 2GZB), a close homolog of BBKI, has a Phe in place of the Cys of BBKI, indicating that this could be favourable mutation (Hansen et al., 2007). The disulfide bonds in ECTI, WBCI and SBTI are in homologous positions. Interestingly, I observed reshuffling of disulfide bonds after unfolding of WBCI, but not ECTI and SBTI. Heat-induced disulfide bond re-arrangement has been reported previously (Creamer et al., 2004; Lowe et al., 2004; Rombouts et al., 2015; Surroca et al., 2002), but further experiments are needed to understand why reshuffling occurred with WBCI, but not SBTI and ECTI.

The position of the GDRs in SBTI align well with the corresponding residues on ECTI and WBCI. It may be possible to graft the GDRs identified in selections with the SBTI backbone onto ECTI or WBCI. Grafting of binding loops has previously been reported with a variety of scaffolds including knottins (Moore et al., 2013), nanofitins (Béhar et al., 2014) and single domain antibodies (Griffiths et al., 2016). Grafting GDRs onto ECTI may be favourable because WT ECTI is more temperature-stable at pH 7.4 and pH 2.0 than WT SBTI. Fusion of two different gastrobody backbones might be useful

for generating bi-specific gastrobodies to increase potency (Boersma et al., 2011; Labrijn et al., 2019).

Chapter 6: Summary and general discussion

Chapter 6: Summary and general discussion

Protein-based targeting reagents, such as antibodies or protein scaffolds, are valuable therapeutic and diagnostic molecules, owing to their binding selectivity and binding affinity (Vazquez-Lombardi et al., 2015). Indeed, nearly 100 antibody products have been approved as therapeutics, with monoclonal antibodies (mAbs) constituting a major fraction of new approved drugs (Kaplon and Reichert, 2021). However, proteins are rapidly inactivated in the mammalian stomach (Kelly et al., 1997), making oral administration of protein-based targeting reagents very challenging (Anselmo et al., 2018; Mrsny, 2012; Smart et al., 2014; Wang et al., 2015b). Here, I sought to engineer a protein scaffold that is resistant to the challenges of the GI tract to enable protein-based targeting in the gut.

Upon entry into the stomach, proteins are exposed to acid and 1 mg/mL pepsin (Brodkorb et al., 2019; Zhu et al., 2006), followed by other proteases and bile acids in the intestinal phase (Lee and Yamamoto, 1989; Mahato et al., 2003; Northfield and McColl, 1973). Nanobodies against enteric pathogens (Harmsen et al., 2005, 2006; Hussack et al., 2011b, 2014, 2018; Sulea et al., 2018) and TNFα (Crowe et al., 2018, 2019) have been engineered for improved stability in the small intestines, however, these proteins are not stable in the stomach. Generalisable strategies for oral delivery of protein biologics have focused on formulations that protect the protein during transit of the stomach (Anselmo et al., 2018; Liu and Basit, 2010; Smart et al., 2014), for example, by encapsulation in plant cells (Shaaltiel et al., 2015), neutralising stomach acid or lyophilisation (Virdi et al., 2019).

In this thesis, I have shown that SBTI can be engineered into an evolvable protein scaffold without complete loss of the gastrointestinal stability of the parent protein. Gastrobodies are stable at pH 1.2 and retain some binding after exposure to

physiological concentrations of gastrointestinal proteases or high temperatures. After affinity maturation, low nanomolar binding affinities can be achieved for some gastrobodies.

The resistance to digestion in the presence of gastric concentrations of pepsin was not unique to SBTI: other proteins with a homologous β-trefoil fold like WBCI, ECTI and BBKI were also stable in the presence of 1 mg/mL pepsin. The low sequence identity between the structural homologs suggests that the resistance to pepsin is, at least in part, governed by the tertiary structure of the proteins. Pepsin exhibits very broad cleavage specificity (Hamuro et al., 2008). Pepsin is present and most active around pH 2 (Piper and Fenton, 1965), so it is important for proteins to remain well-folded in these acidic conditions to be pepsin stable. I observed melting temperatures substantially above physiological temperatures for all β-trefoil proteins studied here, even at pH 2.0. Little conformational change between pH 7.4 and pH 3.0 has previously been observed for bi-functional potato inhibitor of the Kunitz soybean trypsin inhibitor family (Guerra et al., 2016). Solving the crystal structure of SBTI or gastrobodies at different pH would further validate the hypothesis that maintaining the native fold at low pH is critical for pepsin resilience. Circular dichroism (CD), as done for nanofitins (Béhar et al., 2014), may also offer insights into the extent of conformational changes at different pH. Alternatively, hydrogen-deuterium exchange experiments could offer insight into the conformational landscape of SBTI or gastrobodies, to identify flexible regions susceptible to proteolysis (Bai et al., 1995; Chamberlain et al., 1996; Chang and Park, 2009; Wang and Kallenbach, 1998).

All the structural homologs that I studied are protease inhibitors originating in plants and may therefore have experienced similar evolutionary pressures (Murdock and

Shade, 2002). It would be interesting to test whether proteins with a similar fold from non-plant origin or with a non-protease inhibitor function are similarly stable. For example, winged bean albumin (PDB ID: 1WBA) (Dayan et al., 1987) is a seed storage protein with substantial structural homology to SBTI.

At the outset of the project, it was unclear whether mutation of SBTI would compromise pepsin stability. I hypothesized that there may be a trade-off between affinity and protease stability. The number of gastrobodies studied in this thesis is too small to make firm conclusions, but generally the extension of the loops appeared to increase pepsin susceptibility. Increased protease susceptibility with longer loops is to be expected because longer loops provide more potential for local denaturation and thus a good protease substrate (Fontana et al., 2004). Extension and mutation of GDR4 did not increase the pepsin susceptibility of GT44 or GT47 compared to GT01, but yielded substantial improvements in affinity. Little impact of GDR4 mutation on pepsin resistance might be explained by the GDR4 sequence and/or the impact of mutating GDR4 on the overall gastrobody fold. It would be difficult to separate the influence on pepsin resistance of the loop length *per se* from the impact of longer loops on the overall structural stability. Testing different length loops of $(Gly\text{-}Ser)_n$ or $(Ser\text{-}Thr)_n$ at different GDRs might provide some insight on the tolerance of the different GDRs and subsequent pepsin resistance. However, the caveat here would be that the residues in the loops would also influence the structure.

An alternative approach to advance the understanding of pepsin resistance of gastrobodies would be to create a large naïve library with varying loop lengths and perform selections for pepsin resistant variants. Pepsin-digested phage could be captured by binding to trypsin so that the GDRs can be mutated freely. Analysing the

relative abundance of recovered clones NGS on each round (Kong et al., 2020; Rouet et al., 2018; Villequey et al., 2017) might allow the identification of motifs or design rules for pepsin resistance. A complementary approach could include selecting for more pepsin-resistant gastrobodies using a library randomising the complete gastrobody domain constructed with error-prone PCR (Unkauf et al., 2018). NGS studies may be more informative than sequencing clones identified by ELISA because clones can be tracked across selection rounds and in response to different selection pressures like pepsin or heat. Tracking of clones provides insight into changes in relative abundance as opposed to absolute abundance at the end of a selection campaign (Kong et al., 2020; Rouet et al., 2018; Villequey et al., 2017). It will be important to sample a sufficiently large sequence space to be confident in the conclusions of the NGS studies. Design rules might inform GDR length and codon bias for future naïve libraries. Such studies, however, will also rely on a phage platform that is at least as pepsin-resistant as SBTI.

The protease resistance tests in this thesis are simple *in vitro* assays with purified protease and protein of interest. However, digestion *in vivo* will be influenced by pH fluctuations (Fallingborg, 1999; Maqbool et al., 2009), mucus (Wen et al., 2002), the presence of other proteins or food particles, and gastric retention times (Hunt et al., 2015). Improved recovery of an orally administered antibody was observed when the protein was administered to subjects in a fasted state (shortening gastric retention) (Kelly et al., 1997). Co-formulated or ingested food ‘decoy’ proteins would act as competitive substrate inhibitors and influence survival of the active ingredient (e.g. a gastrobody). Milk proteins have been used to shield antibodies from pepsin digestion, but it should be noted that some milk proteins (caseins) delay gastric emptying because caseins aggregate in the stomach to form gastric curd (Crowe et al., 2018;

Huppertz and Chia, 2021; Jones and Martino, 2016). Clearly, it will be important to test the stability of gastrobodies in more complex environments. For example, gastrobodies could be digested together with different foods or cell lysate in a simulated *in vitro* process like INFOGEST 2.0 (Brodkorb et al., 2019) or TIMagc (Bellmann et al., 2016) and binding of gastrobodies would be tested at various time-points using an ELISA. Beyond the protective effect of other proteins, digestion might also produce sticky peptides that could contribute to aggregation or inactivation of gastrobodies.

Considerable research efforts have developed formulations to protect from the acid and pepsin in the stomach to deliver proteinaceous drugs to the intestines (Anselmo et al., 2018; Liu and Basit, 2010; Smart et al., 2014; Wang et al., 2015a). The active ingredient is filled into capsules and coated with pH-responsive polymers like poly(methyl methacrylate) or poly(ethyl acrylate), commercially known as Eudragit® (Dvořáková et al., 2011; Thakral et al., 2013). However, pH-triggered drug delivery systems for delivery into the small intestines can show erratic dissolution behaviour *in vivo* due to variable pH and residence times (Ibekwe et al., 2008). Due to the low intestinal stability of the active ingredients, it is important that the coating disintegrates at the correct site in the GI tract. Gastrobodies may not require protective formulation for delivery into the small intestine and thus benefit from more predictable pharmacokinetics. Conventional enteric coatings do not allow delivery to the proximal small intestine because dissolution is too slow, but double-coatings improve the dissolution rate (Liu et al., 2009). Nevertheless, these technologies do not enable targeting in the stomach, which has been described as “gateway to the GI tract” and plays a key role in health and disease (Hunt et al., 2015). The pepsin and pH stability of gastrobodies may allow targeting features in the stomach. Lastly, formulations add

cost and complexity to the manufacturing process. In low margin industries like the animal feed industry, enteric coatings to protect feed additives or therapies may not be economically feasible (Lei et al., 2013; Paloheimo et al., 2016).

Neutralising stomach acid can improve small intestinal delivery of antibodies (Warny et al., 1999). However, gastric acid suppression, for example by using proton pump inhibitors, has been linked to increased risk of *C. difficile* infection (Tariq et al., 2017).

The pepsin resistance of SBTI has been reported as part of food allergy assessments (Foster et al., 2013; Roychaudhuri et al., 2004; Sung et al., 2014; Wang et al., 2017), which raises the possibility of allergy concerns with gastrobodies. However, SBTI is not on the official list of allergens of the World Health Organisation. 13.3% of soybean-sensitive patients react to SBTI (Arshad et al., 2001; Sung et al., 2014). Soybean is consumed worldwide: infants on soy formula consume up to 10 mg of SBTI per day (Brandon et al., 1991) and some soy flour varieties contain 9 – 11 mg SBTI per g (Kumar et al., 2019). 1.5 or 10 mg/mL SBTI has been co-formulated with insulin to improve absorption of orally-administered insulin in rats (Taverner et al., 2020; Yamamoto et al., 1994). The protective effect of trypsin inhibitors for oral insulin is being explored by Oramed Pharmaceuticals (Kidron, 2007).

Similarly, the inhibition of trypsin by SBTI has been investigated in the treatment of obesity by increasing feelings of satiety (Cristina Oliveira de Lima et al., 2019). Reseland *et al.* have suggested that the presence of protease inhibitors like SBTI induces overstimulation of trypsin and chymotrypsin secretion (Reseland et al., 1996). Soybean trypsin inhibitors confer anti-nutritive properties to soybean feed in poultry diets (Andrade et al., 2016; Perez-Maldonado et al., 2003). I found that the R63A mutation in the gastrobody removed detectable trypsin binding, which reduces

concerns relating to the trypsin-inhibition function of SBTI. The effect of the R63A mutation should be validated in a trypsin inhibition assay.

Oral protein therapies benefit from improved patient compliance and reduced need for medical supervision compared to injections (Fu et al., 2009). A key step in the future development of therapeutic gastrobodies will be choosing suitable targets. Due to the size, gastrobodies are likely confined to the lumen of the GI tract unless there is breakdown of epithelial barrier. For some therapies, being confined to the GI tract and not reaching systemic circulation could be an advantage (Crowe et al., 2018, 2019; Jones and Martino, 2016). For example, intravenous administration of anti-TNFα in the treatment of inflammatory bowel disease leads to systemic immunosuppression, thereby increasing the risk of cancer (Lemaitre et al., 2017), demyelinating disease (Kunchok et al., 2020) and opportunistic infections (Bonovas et al., 2016). Nevertheless, innovative research projects have shown that hijacking epithelial transcytosis pathways (Fay et al., 2020; Li et al., 2021; Taverner et al., 2020), use of ionic liquids (Banerjee et al., 2018) or physical methods (Luo et al., 2021) can enable systemic delivery of orally administered macromolecules. Gastrobodies may synergise well with some of these technologies.

Future targets for gastrobodies in the lumen of the GI tract could include further exotoxins as explored with *C. difficile* TcdA/B in this thesis. Neutralisation of toxins might reduce the burden of increasing concerns over antibiotic resistance. Targets of previously studied oral antibodies are reviewed in (Jones and Martino, 2016). Other than *C. difficile ,* human relevant targets might include *Helicobacter pylori* urease (den Hoed et al., 2011; Horie et al., 2004) or outer membrane proteins of *H. pylori* which colonises the gastric epithelium (Oleastro and Ménard, 2013). Veterinary targets could

include enterotoxigenic *E. coli* fimbrial tip adhesin FaeG (Virdi et al., 2013, 2019) or *Clostridium perfringens* NetB (Fernandes da Costa et al., 2013).

The composition of the heavily glycosylated, negatively charged mucus barrier differs in each segment of the GI tract, becoming up to 800 µm thick in the colon, with average pores 100 – 200 nm in diameter (Pearson et al., 2016). Interaction and diffusion of gastrobodies through mucus will need to be evaluated to determine access to various therapeutic or diagnostic targets.

Diffusion through the mucus layer would also enable targeting of cytokines involved in inflammatory bowel disease (Neurath and Travis, 2012) like TNFα (Crowe et al., 2018, 2019; Yadav et al., 2016) or IL-23p19 (Aschenbrenner et al., 2021). Alternatively, gastrobodies could be conjugated to imaging probes, to enable rapid molecular imaging. Binding to specific molecular features of the GI tract could be helpful in early diagnosis of gastrointestinal disease. For example, lectins specific to glycosylation patterns have been used for endoscopic diagnosis of dysplasia in Barrett's oesophagus (Bird-Lieberman et al., 2012).

Taken together, I have presented basic proof-of-concept for a novel protein scaffold specifically designed for therapeutic or diagnostic action in the GI tract. Further development of gastrobodies might help to address diverse unmet medical or veterinary needs in gastrointestinal disease.

Bibliography

Abramson, A., Halperin, F., Kim, J., and Traverso, G. (2019). Quantifying the Value of Orally Delivered Biologic Therapies: A Cost-Effectiveness Analysis of Oral Semaglutide. J. Pharm. Sci. *108*, 3138–3145.

Aftab, U., and Bedford, M.R. (2018). The use of NSP enzymes in poultry nutrition: myths and realities. Worlds. Poult. Sci. J. *74*, 277–286.

Ahmad, S., Kumar, V., Ramanand, K.B., and Rao, N.M. (2012). Probing protein stability and proteolytic resistance by loop scanning: A comprehensive mutational analysis. Protein Sci. *21*, 433–446.

Alford, R.F., Leaver-Fay, A., Jeliazkov, J.R., O'Meara, M.J., DiMaio, F.P., Park, H., Shapovalov, M. V., Renfrew, P.D., Mulligan, V.K., Kappel, K., et al. (2017). The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. J. Chem. Theory Comput. *13*, 3031–3048.

Alvin, J.W., and Lacy, D.B. (2017). Clostridium difficile toxin glucosyltransferase domains in complex with a non-hydrolyzable UDP-glucose analogue. J. Struct. Biol. *198*, 203–209.

Anderson, D.E., Becktel, W.J., and Dahlquist, F.W. (2002). pH-Induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. Biochemistry *29*, 2403–2408.

Andrade, J.C., Mandarino, J.M.G., Kurozawa, L.E., and Ida, E.I. (2016). The effect of thermal treatment of whole soybean flour on the conversion of isoflavones and inactivation of trypsin inhibitors. Food Chem. *194*, 1095–1101.

Anselmo, A.C., Gokarn, Y., and Mitragotri, S. (2018). Non-invasive delivery strategies for biologics. Nat. Rev. Drug Discov. *18*, 19–40.

Appel, W. (1986). Chymotrypsin: Molecular and catalytic properties. Clin. Biochem. *19*, 317–322.

Arbabi Ghahroudi, M., Desmyter, A., Wyns, L., Hamers, R., and Muyldermans, S. (1997). Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. FEBS Lett. *414*, 521–526.

Arshad, S., Tariq, S., Matthews, S., and Hakim, E. (2001). Sensitization to common allergens and its association with allergic disorders at age 4 years: a whole population birth cohort study. Pediatrics *108*.

Aschenbrenner, D., Quaranta, M., Banerjee, S., Ilott, N., Jansen, J., Steere, B., Chen, Y.-H., Ho, S., Cox, K., Arancibia-Cárcamo, C. V, et al. (2021). Deconvolution of monocyte responses in inflammatory bowel disease reveals an IL-1 cytokine network that regulates IL-23 in genetic and acquired IL-10 resistance. Gut *70*, 1023–1036.

Babcock, G.J., Broering, T.J., Hernandez, H.J., Mandell, R.B., Donahue, K., Boatright, N., Stack, A.M., Lowy, I., Graziano, R., Molrine, D., et al. (2006). Human monoclonal antibodies directed against toxins A and B prevent Clostridium difficile-

induced mortality in hamsters. Infect. Immun. *74*, 6339–6347.

Badran, A.H., and Liu, D.R. (2015). Development of potent in vivo mutagenesis plasmids with broad mutational spectra. Nat. Commun. *6*, 8425.

Bai, Y., Sosnick, T.R., Mayne, L., and Englander, S.W. (1995). Protein folding intermediates: Native-state hydrogen exchange. Science *269*, 193–197.

Banerjee, A., Ibsen, K., Brown, T., Chen, R., Agatemor, C., and Mitragotri, S. (2018). Ionic liquids for oral insulin delivery. Proc. Natl. Acad. Sci. *115*, 7296–7301.

Barkan, D.T., Cheng, X., Celino, H., Tran, T.T., Bhandari, A., Craik, C.S., Sali, A., and Smythe, M.L. (2016). Clustering of disulfide-rich peptides provides scaffolds for hit discovery by phage display: application to interleukin-23. BMC Bioinformatics *17*, 481.

Beekwilder, J., Rakonjac, J., Jongsma, M., and Bosch, D. (1999). A phagemid vector using the E. coli phage shock promoter facilitates phage display of toxic proteins. Gene *228*, 23–31.

Béhar, G., Pacheco, S., Maillasson, M., Mouratou, B., and Pecorari, F. (2014). Switching an anti-IgG binding site between archaeal extremophilic proteins results in Affitins with enhanced pH stability. J. Biotechnol. *192*, 123–129.

Bellmann, S., Lelieveld, J., Gorissen, T., Minekus, M., and Havenaar, R. (2016). Development of an advanced in vitro model of the stomach and its evaluation versus human gastric physiology. Food Res. Int. *88*, 191–198.

Bhattacharjee, N., and Biswas, P. (2010). Position-specific propensities of amino acids in the β-strand. BMC Struct. Biol. *10*, 1–10.

Bilverstone, T.W., Garland, M., Cave, R.J., Kelly, M.L., Tholen, M., Bouley, D.M., Kaye, P., Minton, N.P., Bogyo, M., Kuehne, S.A., et al. (2020). The glucosyltransferase activity of C. difficile Toxin B is required for disease pathogenesis. PLOS Pathog. *16*, e1008852.

Birch, J.R., and Racher, A.J. (2006). Antibody production. Adv. Drug Deliv. Rev. *58*, 671–685.

Bird-Lieberman, E.L., Neves, A.A., Lao-Sirieix, P., O'Donovan, M., Novelli, M., Lovat, L.B., Eng, W.S., Mahal, L.K., Brindle, K.M., and Fitzgerald, R.C. (2012). Molecular imaging using fluorescent lectins permits rapid endoscopic identification of dysplasia in Barrett's esophagus. Nat. Med. *18*, 315–321.

Birk, Y. (1985). The Bowman-Birk inhibitor. Trypsin- and chymotrypsin-inhibitor from soybeans. Int. J. Pept. Protein Res. *25*, 113–131.

Bloom, J.D., Labthavikul, S.T., Otey, C.R., and Arnold, F.H. (2006). Protein stability promotes evolvability. Proc. Natl. Acad. Sci. U. S. A. *103*, 5869–5874.

Boersma, Y., Chao, G., Steiner, D., Wittrup, K., and Plückthun, A. (2011). Bispecific designed ankyrin repeat proteins (DARPins) targeting epidermal growth factor receptor inhibit A431 cell proliferation and receptor recycling. J. Biol. Chem. *286*, 41273–41285.

Bonovas, S., Fiorino, G., Allocca, M., Lytras, T., Nikolopoulos, G.K., Peyrin-Biroulet,

L., and Danese, S. (2016). Biologic Therapies and Risk of Infection and Malignancy in Patients With Inflammatory Bowel Disease: A Systematic Review and Network Meta-analysis. Clin. Gastroenterol. Hepatol. *14*, 1385–1397.

Brandon, D., Bates, A., and Friedman, M. (1991). ELISA analysis of soybean trypsin inhibitors in processed foods. Adv. Exp. Med. Biol. *289*, 321–337.

Brodkorb, A., Egger, L., Alminger, M., Alvito, P., Assunção, R., Ballance, S., Bohn, T., Bourlieu-Lacanal, C., Boutrou, R., Carrière, F., et al. (2019). INFOGEST static in vitro simulation of gastrointestinal food digestion. Nat. Protoc. *14*, 991–1014.

Buldun, C.M., Jean, J.X., Bedford, M.R., and Howarth, M. (2018). SnoopLigase Catalyzes Peptide–Peptide Locking and Enables Solid-Phase Conjugate Isolation. J. Am. Chem. Soc. *140*, 3008–3018.

Calvert, G.D., and Legg, E.F. (1977). The effect of bile salts on the esterolytic assay of trypsin. Clin. Chim. Acta *80*, 535–541.

Carrère, J., Figarella, C., Guy, O., and Thouvenot, J.P. (1986). Human pancreatic chymotrypsinogen A: a non-competitive enzyme immunoassay, and molecular forms in serum and amniotic fluid. Biochim. Biophys. Acta - Gen. Subj. *883*, 46–53.

Cavallo, L., Kleinjung, J., and Fraternali, F. (2003). POPS: A fast algorithm for solvent accessible surface areas at atomic and residue level. Nucleic Acids Res. *31*, 3364–3366.

Chamberlain, A.K., Handel, T.M., and Marqusee, S. (1996). Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. Nat. Struct. Biol. *3*, 782–787.

Chames, P., Regenmortel, M. Van, Weiss, E., and Baty, D. (2009). Therapeutic antibodies: successes, limitations and hopes for the future. Br. J. Pharmacol. *157*, 220–233.

Chang, Y., and Park, C. (2009). Mapping Transient Partial Unfolding by Protein Engineering and Native-State Proteolysis. J. Mol. Biol. *393*, 543–556.

Chaves-Olarte, E., Weidmann, M., Von Eichel-Streiber, C., and Thelestam, M. (1997). Toxins A and B from Clostridium difficile differ with respect to enzymatic potencies, cellular substrate specificities, and surface binding to cultured cells. J. Clin. Invest. *100*, 1734–1741.

Chen, S., Wang, H., Gu, H., Sun, C., Li, S., Feng, H., and Wang, J. (2016). Identification of an essential region for translocation of clostridium difficile toxin B. Toxins (Basel). *8*, 241.

Choct, M. (2006). Enzymes for the feed industry: past, present and future. Worlds. Poult. Sci. J. *62*, 5–16.

Chumbler, N.M., Rutherford, S.A., Zhang, Z., Farrow, M.A., Lisher, J.P., Farquhar, E., Giedroc, D.P., Spiller, B.W., Melnyk, R.A., and Lacy, D.B. (2016). Crystal structure of Clostridium difficile toxin A. Nat. Microbiol. *1*, 15002.

Ciesla, W.P., and Bobak, D.A. (1998). Clostridium difficile toxins A and B are cation-dependent UDP-glucose hydrolases with differing catalytic activities. J. Biol. Chem.

273, 16021–16026.

Cipriano, D., Burnham, M., and Hughes, J. V. (2012). Effectiveness of Various Processing Steps for Viral Clearance of Therapeutic Proteins: Database Analyses of Commonly Used Steps. Methods Mol. Biol. *899*, 277–292.

Clackson, T., and Lowman, H.B. (2004). Phage Display — A Practical Approach (Oxford University Press).

Colombel, J.-F., Rutgeerts, P., Malchow, H., Jacyna, M., Nielsen, O.H., Rask-Madsen, J., Deventer, S. Van, Ferguson, A., Desreumaux, P., Forbes, A., et al. (2001). Interleukin 10 (Tenovil) in the prevention of postoperative recurrence of Crohn's disease. Gut *49*, 42–46.

Conway, P., Tyka, M.D., DiMaio, F., Konerding, D.E., and Baker, D. (2014). Relaxation of backbone bond geometry improves protein energy landscape modeling. Protein Sci. *23*, 47–55.

Correa, A., Pacheco, S., Mechaly, A.E., Obal, G., Bé Har, G., Mouratou, B., Oppezzo, P., Alzari, P.M., Dé, F., and Pecorari, R. (2014). Potent and Specific Inhibition of Glycosidases by Small Artificial Binding Proteins (Affitins). PLoS One *9*.

Craik, D.J., Daly, N.L., Bond, T., and Waine, C. (1999). Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. J. Mol. Biol. *294*, 1327–1336.

Creamer, L., Bienvenue, A., Nilsson, H., Paulsson, M., van Wanroij, M., Lowe, E., Anema, S., Boland, M., and Jiménez-Flores, R. (2004). Heat-Induced Redistribution of Disulfide Bonds in Milk Proteins. 1. Bovine β-Lactoglobulin. J. Agric. Food Chem. *52*, 7660–7668.

Cremers, C.M., Knoefler, D., Vitvitsky, V., Banerjee, R., and Jakob, U. (2014). Bile salts act as effective protein-unfolding agents and instigators of disulfide stress in vivo. Proc. Natl. Acad. Sci. *111*, 1610–1619.

Cristina Oliveira de Lima, V., Piuvezam, G., Leal Lima Maciel, B., and Heloneida de Araújo Morais, A. (2019). Trypsin inhibitors: promising candidate satietogenic proteins as complementary treatment for obesity and metabolic disorders? J. Enzyme Inhib. Med. Chem. *34*, 405–419.

Crowe, J.S., Roberts, K.J., Carlton, T.M., Maggiore, L., Cubitt, M.F., Clare, S., Harcourt, K., Reckless, J., MacDonald, T.T., Ray, K.P., et al. (2018). Preclinical Development of a Novel, Orally-Administered Anti-Tumour Necrosis Factor Domain Antibody for the Treatment of Inflammatory Bowel Disease. Sci. Rep. *8*, 4941.

Crowe, J.S., Roberts, K.J., Carlton, T.M., Maggiore, L., Cubitt, M.F., Ray, K.P., Donnelly, M.C., Wahlich, J.C., Humphreys, J.I., Robinson, J.R., et al. (2019). Oral delivery of the anti-tumor necrosis factor α domain antibody, V565, results in high intestinal and fecal concentrations with minimal systemic exposure in cynomolgus monkeys. Drug Dev. Ind. Pharm. *45*, 387–394.

Daniel, R.M., Cowan, D.A., Morgan, H.W., and Curran, M.P. (1982). A correlation between protein thermostability and resistance to proteolysis. Biochem. J. *207*, 641–644.

Dawson, P.A., and Karpen, S.J. (2015). Intestinal transport and metabolism of bile acids. J. Lipid Res. *56*, 1085–1099.

Dayan, S., Van Donkelaar, A., and Kortt, A. (1987). Crystallization and preliminary X-ray studies of the basic lectin from winged bean (Psophocarpus tetragonolobus). J. Biol. Chem. *262*, 10287–10289.

Dennis, M.S., Herzka, A., and Lazarus, R.A. (1995). Potent and selective Kunitz domain inhibitors of plasma kallikrein designed by phage display. J. Biol. Chem. *270*, 25411–25417.

Dersjant-Li, Y., Awati, A., Schulze, H., and Partridge, G. (2014). Phytase in non-ruminant animal nutrition: a critical review on phytase activities in the gastrointestinal tract and influencing factors. J. Sci. Food Agric. *95*, 878–896.

Diem, M.D., Hyun, L., Yi, F., Hippensteel, R., Kuhar, E., Lowenstein, C., Swift, E.J., O'Neil, K.T., and Jacobs, S.A. (2014). Selection of high-affinity Centyrin FN3 domains from a simple library diversified at a combination of strand and loop positions. Protein Eng. Des. Sel. *27*, 419–429.

Dressman, J.B., Berardi, R.R., Dermentzoglou, L.C., Russell, T.L., Schmaltz, S.P., Barnett, J.L., and Jarvenpaa, K.M. (1990). Upper Gastrointestinal (GI) pH in Young, Healthy Men and Women. Pharm. Res. *7*, 756–761.

Duerr, D., White, S., and Schluesener, H. (2004). Identification of peptide sequences that induce transport of phage across the gastrointestinal mucosal barrier. J. Virol. Methods *116*, 177–180.

Dunlevy, F.K., Martin, S.L., de Courcey, F., Elborn, J.S., and Ennis, M. (2012). Anti-inflammatory effects of DX-890, a human neutrophil elastase inhibitor. J. Cyst. Fibros. *11*, 300–304.

Dupont, D., Alric, M., Blanquet-Diot, S., Bornhorst, G., Cueva, C., Deglaire, A., Denis, S., Ferrua, M., Havenaar, R., Lelieveld, J., et al. (2018). Can dynamic in vitro digestion systems mimic the physiological reality? Food Sci. Nutr. *59*, 1546–1562.

Dvořáková, K., Rabišková, M., Muselík, J., Gajdziok, J., and Bajerová, M. (2011). Coated hard capsules as the pH-dependent drug transport systems to ileo-colonic compartment. Drug Dev. Ind. Pharm. *37*, 1131–1140.

Elkhalil, E.A.I., Männer, K., Borriss, R., Simon, O., Elkhalily, E.A.I., Nner, K.M., and Simon, A.O. (2007). In vitro and in vivo characteristics of bacterial phytases and their efficacy in broiler chickens In vitro and in vivo characteristics of bacterial phytases and their efficacy in broiler chickens. Br. Poult. Sci. *48*, 64–70.

Engberg, J., Andersen, P.S., Nielsen, L.K., Dziegiel, M., Johansen, L.K., and Albrechtsen, B. (1996). Phage-Display Libraries of Murine and Human Antibody Fab Fragments. Appl. Biochem. Biotechnol. *6*, 287–310.

English, H., Hong, J., and Ho, M. (2020). Ancient species offers contemporary therapeutics: an update on shark VNAR single domain antibody sequences, phage libraries and potential clinical applications. Antib. Ther. *3*, 1–9.

Evenepoel, P., Geypens, B., Luypaerts, A., Hiele, M., Ghoos, Y., and Rutgeerts, P. (1998). Digestibility of Cooked and Raw Egg Protein in Humans as Assessed by

Stable Isotope Techniques. J. Nutr. *128*, 1716–1722.

Ewert, S., Cambillau, C., Conrath, K., and Plückthun, A. (2002). Biophysical properties of camelid V(HH) domains compared to those of human V(H)3 domains. Biochemistry *41*, 3628–3636.

Fairhead, M., and Howarth, M. (2015). Site-specific biotinylation of purified proteins using BirA. Methods Mol. Biol. *1266*, 171–184.

Fallingborg, J. (1999). Intraluminal pH of the human gastrointestinal tract. Dan. Med. Bull. *46*, 183–196.

Fay, N.C., Muthusamy, B.-P., Nyugen, L.P., Desai, R.C., Taverner, A., MacKay, J., Seung, M., Hunter, T., Liu, K., Chandalia, A., et al. (2020). A Novel Fusion of IL-10 Engineered to Traffic across Intestinal Epithelium to Treat Colitis. J. Immunol. *205*, 3191–3204.

Fedorak, R.N., Gangl, A., Elson, C.O., Rutgeerts, P., Schreiber, S., Wild, G., Hanauer, S.B., Kilian, A., Cohard, M., LeBeaut, A., et al. (2000). Recombinant human interleukin 10 in the treatment of patients with mild to moderately active Crohn's disease. Gastroenterology *119*, 1473–1482.

Fernandes da Costa, S., Mot, D., Bokori-Brown, M., Savva, C., Basak, A., Van Immerseel, F., and Titball, R. (2013). Protection against avian necrotic enteritis after immunisation with NetB genetic or formaldehyde toxoids. Vaccine *31*, 4003–4008.

Fink, A.L., Calciano, L.J., Goto, Y., Kurotsu, T., and Falleros, D.R. (1994). Classification of Acid Denaturation of Proteins: Intermediates and Unfolded States. Biochemistry *33*, 12504–12511.

Fitza, M.A. (2002). Evolving Phage for Protein Display in Gastric Juice by Evolving Phage for Protein Display in Gastric Juice. Masters thesis. University of Texas, Austin.

Fontana, A., de Laureto, P., Spolaore, B., Frare, E., Picotti, P., and Zambonin, M. (2004). Probing protein structure by limited proteolysis. Acta Biochim. Pol. *51*, 299–321.

Foster, E.S., Kimber, I., and Dearman, R.J. (2013). Relationship between protein digestibility and allergenicity: Comparisons of pepsin and cathepsin. Toxicology *309*, 30–38.

Frei, J.C., and Lai, J.R. (2016). Protein and Antibody Engineering by Phage Display. In Methods in Enzymology, (Academic Press Inc.), pp. 45–87.

Frenzel, A., Schirrmann, T., and Hust, M. (2016). Phage display-derived human antibodies in clinical development and therapy. MAbs *8*, 1177–1194.

Fu, A.Z., Qiu, Y., and Radican, L. (2009). Impact of fear of insulin or fear of injection on treatment outcomes of patients with diabetes. Curr. Med. Res. Opin. *25*, 1413–1420.

Gao, Y.G., Su, S.Y., Robinson, H., Padmanabhan, S., Lim, L., McCrary, B.S., Edmondson, S.P., Shriver, J.W., and Wang, A.H.J. (1998). The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. Nat. Struct. Biol. *5*,

782–786.

Gareb, B., Posthumus, S., Beugeling, M., Koopmans, P., Touw, D.J., Dijkstra, G., Kosterink, J.G.W., and Frijlink, H.W. (2019). Towards the Oral Treatment of Ileo-Colonic Inflammatory Bowel Disease with Infliximab Tablets: Development and Validation of the Production Process. Pharmaceutics *11*, 428.

Gass, J., Vora, H., Hofmann, A.F., Gray, G.M., and Khosla, C. (2007). Enhancement of Dietary Protein Digestion by Conjugated Bile Acids. Gastroenterology *133*, 16–23.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R., and Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. *31*, 3784–3788.

Gebauer, M., and Skerra, A. (2020). Engineered Protein Scaffolds as Next-Generation Therapeutics. Annu. Rev. Pharmacol. Toxicol. *60*, 391–415.

de Genst, E., Silence, K., Decanniere, K., Conrath, K., Loris, R., Kinne, J., Muyldermans, S., and Wyns, L. (2006). Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. Proc. Natl. Acad. Sci. *103*, 4586–4591.

Geoghegan, K.F., Dixon, H.B.F., Rosner, P.J., Hoth, L.R., Lanzetti, A.J., Borzilleri, K.A., Marr, E.S., Pezzullo, L.H., Martin, L.B., Lemotte, P.K., et al. (1999). Spontaneous α-N-6-phosphogluconoylation of a “His tag” in Escherichia coli: The cause of extra mass of 258 or 178 Da in fusion proteins. Anal. Biochem. *267*, 169–184.

Gera, N., Hussain, M., Wright, R.C., and Rao, B.M. (2011). Highly stable binding proteins derived from the hyperthermophilic Sso7d scaffold. J. Mol. Biol. *409*, 601–616.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Giletto, A., and Pace, C.N. (1999). Buried, Charged, Non-Ion-Paired Aspartic Acid 76 Contributes Favorably to the Conformational Stability of Ribonuclease T1. Biochemistry *38*, 13379–13384.

Greenwood-Van Meerveld, B., Johnson, A.C., and Grundy, D. (2017). Gastrointestinal Physiology and Function. In Handbook of Experimental Pharmacology, (Springer), pp. 1–16.

Griffiths, K., Dolezal, O., Cao, B., Nilsson, S.K., See, H.B., Pfleger, K.D.G., Roche, M., Gorry, P.R., Pow, A., Viduka, K., et al. (2016). i-bodies, human single domain antibodies that antagonize chemokine receptor CXCR4. J. Biol. Chem. *291*, 12641–12657.

Guerra, Y., Valiente, P.A., Pons, T., Berry, C., and Rudiño-Piñera, E. (2016). Structures of a bi-functional Kunitz-type STI family inhibitor of serine and aspartic proteases: Could the aspartic protease inhibition have evolved from a canonical serine protease-binding loop? J. Struct. Biol. *195*, 259–271.

Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995). Tight regulation,

modulation, and high-level expression by vectors containing the arabinose P(BAD) promoter. J. Bacteriol. *177*, 4121–4130.

Hackel, B.J., Kapila, A., and Dane Wittrup, K. (2008). Picomolar Affinity Fibronectin Domains Engineered Utilizing Loop Length Diversity, Recursive Mutagenesis, and Loop Shuffling. J. Mol. Biol. *381*, 1238–1252.

Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hammers, C., Songa, E.B., Bendahman, N., and Hammers, R. (1993). Naturally occurring antibodies devoid of light chains.

Hamers, R., Hamers-Casterman, C., Muyldermans, S.V.M., Frenken, L.G.J., and Verrips, C.T. (1993). Production of antibodies or (functionalized) fragments thereof derived from heavy chain immunoglobulins of camelidae (European Patent Office: EP93201239).

Hamuro, Y., Coales, S.J., Molnar, K.S., Tuske, S.J., and Morrow, J.A. (2008). Specificity of immobilized porcine pepsin in H/D exchange compatible conditions. Rapid Commun. Mass Spectrom. *22*, 1041–1046.

Hansen, D., Macedo-Ribeiro, S., Veríssimo, P., Yoo Im, S., Sampaio, M.U., and Oliva, M.L.V. (2007). Crystal structure of a novel cysteinless plant Kunitz-type protease inhibitor. Biochem. Biophys. Res. Commun. *360*, 735–740.

Hansrod, S., James, G., Notghi, A., Croasdale, J., O'Brien, J., and Thomson, W.H. (2020). Gastric emptying: Methodology and normal ranges for two commonly used meals in the UK. Nucl. Med. Commun. 636–650.

Harmsen, M.M., Van Solt, C.B., Hoogendoorn, A., Van Zijderveld, F.G., Niewold, T.A., and Van Der Meulen, J. (2005). Escherichia coli F4 fimbriae specific llama single-domain antibody fragments effectively inhibit bacterial adhesion in vitro but poorly protect against diarrhoea. Vet. Microbiol. *111*, 89–98.

Harmsen, M.M., Van Solt, C.B., Van Zijderveld-Van Bemmel, A.M., Niewold, T.A., and Van Zijderveld, F.G. (2006). Selection and optimization of proteolytically stable llama single-domain antibody fragments for oral immunotherapy. Appl. Microbiol. Biotechnol. *72*, 544–551.

Harrison, G.A. (1923). Insulin in alcoholic solution by the mouth. Br. Med. J. *2*, 1204–1205.

Hatta, H., Tsuda, K., Akachi, S., Kim, M., Yamamoto, T., and Ebina, T. (1993). Oral Passive Immunization Effect of Anti-Human Rotavirus IgY and Its Behavior against Proteolytic Enzymes. Biosci. Biotechnol. Biochem. *57*, 1077–1081.

Higgins, L.M., Lambkin, I., Donnelly, G., Byrne, D., Wilson, C., Dee, J., Smith, M., and O 'mahony, D.J. (2004). In Vivo Phage Display To Identify M Cell–Targeting Ligands. Pharm. Res. *21*.

den Hoed, C., de Vries, A., Mensink, P., Dierikx, C., Suzuki, H., Capelle, L., van Dekken, H., Ouwendijk, R., and Kuipers, E. (2011). Bovine antibody-based oral immunotherapy for reduction of intragastric Helicobacter pylori colonization: A randomized clinical trial. Can. J. Gastroenterol. *25*, 207.

Horie, K., Horie, N., Abdou, A.M., Yang, J.O., Yun, S.S., Chun, H.N., Park, C.K.,

Kim, M., and Hatta, H. (2004). Suppressive Effect of Functional Drinking Yogurt Containing Specific Egg Yolk Immunoglobulin on Helicobacter pylori in Humans. J. Dairy Sci. *87*, 4073–4079.

Hu, P.L., Yuan, Y.H., Yue, T.L., and Guo, C.F. (2018). Bile acid patterns in commercially available oxgall powders used for the evaluation of the bile tolerance ability of potential probiotics. PLoS One *13*, e0192964.

Huang, P., Ban, Y., Richter, F., Andre, I., Vernon, R., Schief, W., and Baker, D. (2011). RosettaRemodel: a generalized framework for flexible backbone protein design. PLoS One *6*.

Huang, W., McKevitt, M., and Palzkill, T. (2000). Use of the arabinose p(bad) promoter for tightly regulated display of proteins on bacteriophage. Gene *251*, 187–197.

Hubbard, S.J. (1998). The structural aspects of limited proteolysis of native proteins. Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. *1382*, 191–206.

Hunt, R., Camilleri, M., Crowe, S., El-Omar, E., Fox, J., Kuipers, E., Malfertheiner, P., McColl, K., Pritchard, D., Rugge, M., et al. (2015). The stomach in health and disease. Gut *64*, 1650–1668.

Huppertz, T., and Chia, L.W. (2021). Milk protein coagulation under gastric conditions: A review. Int. Dairy J. *113*, 104882.

Hussack, G., Hirama, T., Ding, W., MacKenzie, R., and Tanha, J. (2011a). Engineered Single-Domain Antibodies with High Protease Resistance and Thermal Stability. PLoS One *6*, e28218.

Hussack, G., Arbabi-Ghahroudi, M., Van Faassen, H., Songer, J.G., Ng, K.K.S., MacKenzie, R., and Tanha, J. (2011b). Neutralization of Clostridium difficile toxin A with single-domain antibodies targeting the cell receptor binding domain. J. Biol. Chem. *286*, 8961–8976.

Hussack, G., Riazi, A., Ryan, S., van Faassen, H., MacKenzie, R., Tanha, J., and Arbabi-Ghahroudi, M. (2014). Protease-resistant single-domain antibodies inhibit Campylobacter jejuni motility. Protein Eng. Des. Sel. *27*, 191–198.

Hussack, G., Ryan, S., van Faassen, H., Rossotti, M., MacKenzie, C.R., and Tanha, J. (2018). Neutralization of Clostridium difficile toxin B with VHH-Fc fusions targeting the delivery and CROPs domains. PLoS One *13*, e0208978.

Ibekwe, V.C., Fadda, H.M., McConnell, E.L., Khela, M.K., Evans, D.F., and Basit, A.W. (2008). Interplay Between Intestinal pH, Transit Time and Feed Status on the In Vivo Performance of pH Responsive Ileo-Colonic Release Systems. Pharm. Res. *25*, 1828–1835.

Ishtiaq, O., Qadri, A.M., Mehar, S., Gondal, G.M., Iqbal, T., Ali, S., Mati-ur-Rahman, and Janjua, N.Z. (2012). Disposal of syringes, needles, and lancets used by diabetic patients in Pakistan. J. Infect. Public Health *5*, 182–188.

Johansson, M.E. V., Sjövall, H., and Hansson, G.C. (2013). The gastrointestinal mucus system in health and disease. Nat. Rev. Gastroenterol. Hepatol. *10*, 352–361.

Jones, R.G.A., and Martino, A. (2016). Targeted localized use of therapeutic antibodies: A review of non-systemic, topical and oral applications. Crit. Rev. Biotechnol. *36*, 506–520.

Kageyama, T. (2014). Pepsinogens, progastricsins, and prochymosins: structure, function, evolution, and development. Cell. Mol. Life Sci. *59*, 288–306.

Kageyama, T., Tanabe, K., and Koiwai, O. (1990). Structure and development of rabbit pepsinogens. Stage-specific zymogens, nucleotide sequences of cDNAs, molecular evolution, and gene expression during development. J. Biol. Chem. *265*, 17031–17038.

Kalichuk, V., Béhar, G., Renodon-Cornière, A., Danovski, G., Obal, G., Barbet, J., Mouratou, B., and Pecorari, F. (2016). The archaeal 7 kDa DNA-binding proteins: extended characterization of an old gifted family. Sci. Rep. *6*.

Kalichuk, V., Renodon-Cornière, A., Béhar, G., Carrión, F., Obal, G., Maillasson, M., Mouratou, B., Préat, V., and Pecorari, F. (2018). A novel, smaller scaffold for Affitins: Showcase with binders specific for EpCAM. Biotechnol. Bioeng. *115*, 290–299.

Kang, S.K., Woo, J.H., Kim, M.K., Woo, S.S., Choi, J.H., Lee, H.G., Lee, N.K., and Choi, Y.J. (2008). Identification of a peptide sequence that improves transport of macromolecules across the intestinal mucosal barrier targeting goblet cells. J. Biotechnol. *135*, 210–216.

Kaplon, H., and Reichert, J.M. (2021). Antibodies to watch in 2021. MAbs *13*.

Kelly, C., Chetham, S., Keates, S., Bostwick, E., Roush, A., Castagliuolo, I., LaMont, J., and Pothoulakis, C. (1997). Survival of anti-Clostridium difficile bovine immunoglobulin concentrate in the human gastrointestinal tract. Antimicrob. Agents Chemother. *41*, 236–241.

Kidron, M. (2007). Methods and compositions for oral administration of proteins (United States Patent Office: US20070087957A1).

Kim, S.H., Hara, S., Hase, S., Ikenaka, T., Toda, H., Kitamura, K., and Kaizuma, N. (1985). Comparative Study on Amino Acid Sequences of Kunitz-Type Soybean Trypsin Inhibitors, Tia, Tib, and Tic1. J. Biochem. *98*, 435–448.

Kink, J.A., and Williams, J.A. (1998). Antibodies to Recombinant Clostridium difficile Toxins A and B Are an Effective Treatment and Prevent Relapse of C. difficile-Associated Disease in a Hamster Model of Infection. Infect. Immun. *66*, 2018–2025.

Koide, A., Wojcik, J., Gilbreth, R.N., Hoey, R.J., and Koide, S. (2012). Teaching an old scaffold new tricks: Monobodies constructed using alternative surfaces of the FN3 scaffold. J. Mol. Biol. *415*, 393–405.

Kong, X.D., Moriya, J., Carle, V., Pojer, F., Abriata, L.A., Deyle, K., and Heinis, C. (2020). De novo development of proteolytically resistant therapeutic peptides for oral administration. Nat. Biomed. Eng. *4*, 560–571.

Koziolek, M., Grimm, M., Becker, D., Iordanov, V., Zou, H., Shimizu, J., Wanke, C., Garbacz, G., and Weitschies, W. (2015). Investigation of pH and Temperature Profiles in the GI Tract of Fasted Human Subjects Using the Intellicap® System. J. Pharm. Sci. *104*, 2855–2863.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D. Biol. Crystallogr. *60*, 2256–2268.

Kroh, H.K., Chandrasekaran, R., Rosenthal, K., Woods, R., Jin, X., Ohi, M.D., Nyborg, A.C., Rainey, G.J., Warrener, P., Spiller, B.W., et al. (2017). Use of a neutralizing antibody helps identify structural features critical for binding of Clostridium difficile toxin TcdA to the host cell surface. J. Biol. Chem. *292*, 14401–14412.

Kuehne, S.A., Cartman, S.T., Heap, J.T., Kelly, M.L., Cockayne, A., and Minton, N.P. (2010). The role of toxin A and toxin B in Clostridium difficile infection. Nature *467*, 711–713.

Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L., and Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. Science *302*, 1364–1368.

Kumar, V., Rani, A., Mittal, P., and Shuaib, M. (2019). Kunitz trypsin inhibitor in soybean: contribution to total trypsin inhibitor activity as a function of genotype and fate during processing. J. Food Meas. Charact. *13*, 1583–1590.

Kunchok, A., Aksamit, A.J., Davis, J.M., Kantarci, O.H., Keegan, B.M., Pittock, S.J., Weinshenker, B.G., and McKeon, A. (2020). Association between tumor necrosis factor inhibitor exposure and inflammatory central nervous system events. JAMA Neurol. *77*, 937–946.

Kunz, P., Zinner, K., Mücke, N., Bartoschik, T., Muyldermans, S., and Hoheisel, J.D. (2018). The structural basis of nanobody unfolding reversibility and thermoresistance. Sci. Rep. *8*, 1–10.

Labouebe, M.Z., Cinier, M., Rousseaux, C., Castro, R., Cunha, A., Tan, N.T., Kitten, O., Gurny, R., and Scapozza, L. (2016). SADEL Project: Scaffolds for Alternative Delivery. In ECCO Congress, p.

Labrijn, A.F., Janmaat, M.L., Reichert, J.M., and Parren, P.W.H.I. (2019). Bispecific antibodies: a mechanistic review of the pipeline. Nat. Rev. Drug Discov. *18*, 585–608.

LaFrance, M.E., Farrow, M.A., Chandrasekaran, R., Sheng, J., Rubin, D.H., and Lacy, D.B. (2015). Identification of an epithelial cell receptor responsible for Clostridium difficile TcdB-induced cytotoxicity. Proc. Natl. Acad. Sci. *112*, 7073–7078.

Lanini, S., Puro, V., Lauria, F.N., Fusco, F.M., Nisii, C., and Ippolito, G. (2009). Patient to patient transmission of hepatitis B virus: a systematic review of reports on outbreaks between 1992 and 2007. BMC Med. *7*, 1–9.

Leaver-Fay, A., Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak, R., Kaufman, K.W., Renfrew, P.D., Smith, C.A., Sheffler, W., et al. (2011). Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. Methods Enzymol. *487*, 545–574.

Lee, V.H.L., and Yamamoto, A. (1989). Penetration and enzymatic barriers to

peptide and protein absorption. Adv. Drug Deliv. Rev. *4*, 171–207.

Lee, S.A., Dunne, J., Mottram, T., and Bedford, M.R. (2017). British Poultry Science Effect of diet phase change, dietary Ca and P level and phytase on bird performance and real-time gizzard pH measurements Effect of diet phase change, dietary Ca and P level and phytase on bird performance and real-time gizzard pH. Br. Poult. Sci. *58*, 290–297.

Lee, S.A., Dunne, J., Febery, E., Wilcock, P., Mottram, T., and Bedford, M.R. (2018). Superdosing phytase reduces real-time gastric pH in broilers and weaned piglets. Br. Poult. Sci. *59*, 330–339.

Lei, X.G., Weaver, J.D., Mullaney, E., Ullah, A.H., and Azain, M.J. (2013). Phytase, a New Life for an “Old” Enzyme. Annu. Rev. Anim. Biosci. *1*, 283–309.

Lemaitre, M., Kirchgesner, J., Rudnichi, A., Carrat, F., Zureik, M., Carbonnel, F., and Dray-Spira, R. (2017). Association between use of thiopurines or tumor necrosis factor antagonists alone or in combination and risk of lymphoma in patients with inflammatory bowel disease. J. Am. Med. Assoc. *318*, 1679–1686.

LeMaster, D.M., Tang, J., Paredes, D.I., and Hernández, G. (2005). Enhanced thermal stability achieved without increased conformational rigidity at physiological temperatures: Spatial propagation of differential flexibility in rubredoxin hybrids. Proteins Struct. Funct. Bioinforma. *61*, 608–616.

Li, R., Laurent, F., Taverner, A., Mackay, J., Bank, P.A. De, and Mrsny, R.J. (2021). Intestinal Transcytosis of a Protein Cargo and Nanoparticles Mediated by a Non-Toxic Form of Pseudomonas aeruginosa Exotoxin A. Pharmaceutics *13*, 1171.

van der Linden, R., Frenken, L., de Geus, B., Harmsen, M., Ruuls, R., Stok, W., de Ron, L., Wilson, S., Davis, P., and Verrips, C. (1999). Comparison of physical chemical properties of llama VHH antibody fragments and mouse monoclonal antibodies. Biochim. Biophys. Acta *1431*, 37–46.

Linderstrøm-Lang, K. (1950). Structure and enzymatic break-down of proteins. Cold Spring Harb. Symp. Quantiative Biol. *14*, 117–126.

Liu, F., and Basit, A. (2010). A paradigm shift in enteric coating: achieving rapid release in the proximal small intestine of man. J. Control. Release *147*, 242–245.

Liu, F., Lizio, R., Meier, C., Petereit, H., Blakey, P., and Basit, A. (2009). A novel concept in enteric coating: a double-coating system providing rapid drug release in the proximal small intestine. J. Control. Release *133*, 119–124.

Lobstein, J., Emrich, C.A., Jeans, C., Faulkner, M., Riggs, P., and Berkmen, M. (2012). SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microb. Cell Fact. *11*, 753.

Loussouarn, A., Béhar, G., Pecorari, F., Croyal, M., and Renodon-Cornière, A. (2020). Characterization of Affitin proteolytic digestion in biorelevant media and improvement of their stabilities via protein engineering. Sci. Rep. *10*, 1–13.

Lowe, E., Anema, S., Bienvenue, A., Boland, M., Creamer, L., and Jiménez-Flores, R. (2004). Heat-Induced Redistribution of Disulfide Bonds in Milk Proteins. 2.

Disulfide Bonding Patterns between Bovine β-Lactoglobulin and κ-Casein. J. Agric. Food Chem. *52*, 7669–7680.

Luo, Z., Paunović, N., and Leroux, J.C. (2021). Physical methods for enhancing drug absorption from the gastrointestinal tract. Adv. Drug Deliv. Rev. *175*, 113814.

Magill, S.S., Edwards, J.R., Bamberg, W., Beldavs, Z.G., Dumyati, G., Kainer, M.A., Lynfield, R., Maloney, M., McAllister-Hollod, L., Nadle, J., et al. (2014). Multistate point-prevalence survey of health care-associated infections. N. Engl. J. Med. *370*, 1198–1208.

Mahato, R.I., Narang, A.S., Thoma, L., and Miller, D.D. (2003). Emerging trends in oral delivery of peptide and protein drugs. Crit. Rev. Ther. Drug Carrier Syst. *20*, 153–214.

Malik, A., Rudolph, R., and Söhling, B. (2005). Use of enhanced green fluorescent protein to determine pepsin at high sensitivity. Anal. Biochem. *340*, 252–258.

Maqbool, S., Parkman, H.P., and Friedenberg, F.K. (2009). Wireless Capsule Motility: Comparison of the SmartPill® GI Monitoring System with Scintigraphy for Measuring Whole Gut Transit. Dig. Dis. Sci. *54*, 2167–2174.

Marciniszyn, J., Hartsuck, J., and Tang, J. (1977). Pepstatin inhibition mechanism. Adv. Exp. Med. Biol. *95*, 199–210.

de Marco, A. (2009). Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. Microb. Cell Fact. *8*, 26.

Markland, W., Ley, A., and Ladner, R. (1996a). Iterative optimization of high-affinity protease inhibitors using phage display. 2. Plasma kallikrein and thrombin. Biochemistry *35*, 8058–8067.

Markland, W., Ley, A., Lee, S., and Ladner, R. (1996b). Iterative optimization of high-affinity proteases inhibitors using phage display. 1. Plasmin. Biochemistry *35*, 8045–8057.

McLachlan, A.D. (1979). Three-fold structural pattern in the soybean trypsin inhibitor (Kunitz). J. Mol. Biol. *133*, 557–563.

Minekus, M., Smeets-Peeters, M., Bernalier, A., Marol-Bonnin, S., Havenaar, R., Marteau, P., Alric, M., Fonty, G., and Huis in't Veld, J.H.J. (1999). A computer-controlled system to simulate conditions of the large intestine with peristaltic mixing, water absorption and absorption of fermentation products. Appl. Microbiol. Biotechnol. *53*, 108–114.

Minnekus, M., Marteau, P., Havenaar, R., and Veld, J. (1995). A multicompartmental dynamic computer-controlled model simulating the stomach and small intestine. ATLA *23*, 197–209.

Molly, K., Vande Woestyne, M., and Verstraete, W. (1993). Development of a 5-step multi-chamber reactor as a simulation of the human intestinal microbial ecosystem. Appl. Microbiol. Biotechnol. *39*, 254–258.

Moore, S.J., Leung, C.L., Norton, H.K., and Cochran, J.R. (2013). Engineering Agatoxin, a Cystine-Knot Peptide from Spider Venom, as a Molecular Probe for In

Vivo Tumor Imaging. PLoS One *8*, e60498.

Moroz, E., Matoori, S., and Leroux, J.C. (2016). Oral delivery of macromolecular drugs: Where we are after almost 100 years of attempts. Adv. Drug Deliv. Rev. *101*, 108–121.

Morrison, C. (2019). Nanobody approval gives domain antibodies a boost. Nat. Eeviews Drug Discov. *18*, 485–487.

Mouratou, B., Schaeffer, F., Guilvout, I., Tello-Manigne, D., Pugsley, A.P., Alzari, P.M., Dé, F., and Pecorari, R. (2007). Remodeling a DNA-binding protein as a specific in vivo inhibitor of bacterial secretin PulD. PNAS *104*, 17983–17988.

Mrsny, R.J. (2012). Oral drug delivery research in Europe. J. Control. Release *161*, 247–253.

Mulet-Cabero, A.I., Egger, L., Portmann, R., Ménard, O., Marze, S., Minekus, M., Le Feunteun, S., Sarkar, A., Grundy, M.M.L., Carrière, F., et al. (2020). A standardised semi-dynamic: in vitro digestion method suitable for food-an international consensus. Food Funct. *11*, 1702–1720.

Mullard, A. (2021). New hope for anti-obesity drugs. Nat. Rev. Drug Discov. *20*, 575.

Murase, T., Eugenio, L., Schorr, M., Hussack, G., Tanha, J., Kitova, E.N., Klassen, J.S., and Ng, K.K.S. (2014). Structural Basis for Antibody Recognition in the Receptor-binding Domains of Toxins A and B from Clostridium difficile. J. Biol. Chem. *289*, 2331–2343.

Murdock, L.L., and Shade, R.E. (2002). Lectins and protease inhibitors as plant defenses against insects. J. Agric. Food Chem. *50*, 6605–6611.

Murzin, A.G., Lesk, A.M., and Chothia, C. (1992). β-Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1β and 1α and fibroblast growth factors. J. Mol. Biol. *223*, 531–543.

Neurath, M.F., and Travis, S.P.L. (2012). Mucosal healing in inflammatory bowel diseases: A systematic review. Gut *61*, 1619–1635.

Ng, K.K.S., Petersen, J.F.W., Cherney, M.M., Garen, C., Zalatoris, J.J., Rao-Naik, C., Dunn, B.M., Martzen, M.R., Peanasky, R.J., and James, M.N.G. (2000). Structural basis for the inhibition of porcine pepsin by Ascaris pepsin inhibitor-3. Nat. Struct. Biol. *7*, 653–657.

Nguyen, K.Q., and Winter, B. (2006). Polypeptide having a phytase activity and nucleotide seuence coding therefor (United States Patent Office: US8313937B2).

Nilsson, E., Larsson, A., Olesen, H. V., Wejåker, P.E., and Kollberg, H. (2008). Good effect of IgY against Pseudomonas aeruginosa infections in cystic fibrosis patients. Pediatr. Pulmonol. *43*, 892–899.

Nir, Y., Paz, A., Sabo, E., and Potasman, I. (2003). Fear of injections in young adults: prevalence and associations. Am. J. Trop. Med. Hyg. *68*, 341–344.

Nobrega, F.L., Costa, A.R., Santos, J.F., Siliakus, M.F., van Lent, J.W.M., Kengen, S.W.M., Azeredo, J., and Kluskens, L.D. (2016). Genetically manipulated phages with improved pH resistance for oral administration in veterinary medicine. Sci. Rep.

6, 39235.

Nord, K., Gunneriusson, E., Ringdahl, J., Ståhl, S., Uhlén, M., and Nygren, P. (1997). Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. Nat. Biotechnol. *15*, 772–777.

Northfield, T.C., and McColl, I. (1973). Postprandial concentrations of free and conjugated bile acids down the length of the normal human small intestine. Gut *14*, 513–518.

Oleastro, M., and Ménard, A. (2013). The Role of Helicobacter pylori Outer Membrane Proteins in Adherence and Pathogenesis. Biology (Basel). *2*, 1110–1134.

Olowokure, B., Duggal, H. V., and Armitage, L.E. (2003). The disposal of used sharps by diabetic patients living at home. Int. J. Environ. Health Res. *13*, 117–123.

Onesti, S., Brick, P., and Blow, D.M. (1991). Crystal structure of a Kunitz-type trypsin inhibitor from Erythrina coffra seeds. J. Mol. Biol. *217*, 153–176.

Onyango, E.M., Bedford, M.R., and Adeola, O. (2005). Efficacy of an Evolved Escherichia coli Phytase in Diets of Broiler Chicks. Poult. Sci. *84*, 248–255.

Orth, P., Xiao, L., Hernandez, L., Reichert, P., Sheth, P., Beaumont, M., Yang, X., Murgolo, N., Ermakov, G., DiNunzio, E., et al. (2014). Mechanism of action and epitopes of Clostridium difficile toxin B-neutralizing antibody bezlotoxumab revealed by X-ray crystallography. J. Biol. Chem. *289*, 18008–18021.

Pahari, S., Sun, L., and Alexov, E. (2019). PKAD: a database of experimentally measured pKa values of ionizable groups in proteins. Database *2019*.

Paloheimo, M., Haarmann, T., Mäkinen, S., and Vehmaanperä, J. (2016). Production of Industrial Enzymes in Trichoderma reesei. In Gene Expression Systems in Fungi: Advancements and Applications, (Springer), pp. 23–57.

Park, C., and Marqusee, S. (2004). Probing the High Energy States in Proteins by Proteolysis. J. Mol. Biol. *343*, 1467–1476.

Park, C., Zhou, S., Gilmore, J., and Marqusee, S. (2007). Energetics-based Protein Profiling on a Proteomic Scale: Identification of Proteins Resistant to Proteolysis. J. Mol. Biol. *368*, 1426–1437.

Pearson, J., Chater, P., and Wilcox, M. (2016). The properties of the mucus barrier, a unique gel--how can nanoparticles cross it? Ther. Deliv. *7*, 229–244.

Pecorari, F., and Alzari, P. (2006). Ob-fold used as scaffold for engineering new specific binders (European Patent Office: EP1930342B1).

Perez-Maldonado, R.A., Mannion, P.F., and Farrell, D.J. (2003). Effects of heat treatment on the nutritional value of raw soybean selected for low trypsin inhibitor activity. Br. Poult. Sci. *44*, 299–308.

Pharmacopeia, U.S. (2015). Pepsin Activity. In Ninth Edition of the Food Chemicals Codex (FCC9), (Rockville, MD), pp. 1410–1411.

Piper, D.W., and Fenton, B.H. (1965). pH stability and activity curves of pepsin with special reference to their clinical importance. Gut *6*, 506–508.

Poutanen, S.M., and Simor, A.E. (2004). Clostridium difficile-associated diarrhea in adults. CMAJ *171*, 51–58.

Qiu, Y., Taichi, M., Wei, N., Yang, H., Luo, K.Q., and Tam, J.P. (2017). An Orally Active Bradykinin B1 Receptor Antagonist Engineered as a Bifunctional Chimera of Sunflower Trypsin Inhibitor. J. Med. Chem. *60*, 504–510.

Rahman, S., Van Nguyen, S., Icatlo, F.C., Umeda, K., and Kodama, Y. (2013). Oral passive IgY-based immunotherapeutics: A novel solution for prevention and treatment of alimentary tract diseases. Hum. Vaccines Immunother. *9*, 1039–1048.

Rathore, N., and Rajan, R.S. (2008). Current perspectives on stability of protein drug products during formulation, fill and finish operations. Biotechnol. Prog. *24*, 504–514.

Ravichandran, S., Dasgupta, J., Chakrabarti, C., Ghosh, S., Singh, M., and Dattagupta, J.K. (2001). The role of Asn14 in the stability and conformation of the reactive-site loop of winged bean chymotrypsin inhibitor: crystal structures of two point mutants Asn14→Lys and Asn14→Asp. Protein Eng. Des. Sel. *14*, 349–357.

Raybould, M.I.J., Marks, C., Krawczyk, K., Taddese, B., Nowak, J., Lewis, A.P., Bujotzek, A., Shi, J., and Deane, C.M. (2019). Five computational developability guidelines for therapeutic antibody profiling. Proc. Natl. Acad. Sci. U. S. A. *116*, 4025–4030.

Rebello, S., Jose, L., Sindhu, R., and Aneesh, E.M. (2017). Molecular advancements in the development of thermostable phytases. Appl. Microbiol. Biotechnol. *101*, 2677–2689.

Recoules, E., Sabboh-Jourdan, H., Narcy, A., Lessire, M., Harichaux, G., Labas, V., Duclos, M.J., and Réhault-Godbert, S. (2017). Exploring the in vivo digestion of plant proteins in broiler chickens. Poult. Sci. *96*, 1735–1747.

Reseland, J., Holm, H., Jacobsen, M., Jenssen, T., and Hanssen, L. (1996). Proteinase inhibitors induce selective stimulation of human trypsin and chymotrypsin secretion. J. Nutr. *126*, 634–642.

Riazi, A., Strong, P.C.R., Coleman, R., Chen, W., Hirama, T., van Faassen, H., Henry, M., Logan, S.M., Szymanski, C.M., MacKenzie, R., et al. (2013). Pentavalent Single-Domain Antibodies Reduce Campylobacter jejuni Motility and Colonization in Chickens. PLoS One *8*, e83928.

Riethorst, D., Mols, R., Duchateau, G., Tack, J., Brouwers, J., and Augustijns, P. (2016). Characterization of Human Duodenal Fluids in Fasted and Fed State Conditions. J. Pharm. Sci. *105*, 673–681.

Rinderknecht, H., Renner, I.G., and Carmack, C. (1979). Trypsinogen variants in pancreatic juice of healthy volunteers, chronic alcoholics, and patients with pancreatitis and cancer of the pancreas. Gut *20*, 886–891.

Roberts, B.L., Markland, W., Ley, A.C., Kent, R.B., White, D.W., Guterman, S.K., and Ladner, R.C. (1992). Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. Proc. Natl. Acad. Sci. U. S. A. *89*, 2429.

Robic, S., Linscott, K.B., Aseem, M., Humphreys, E.A., and McCartha, S.R. (2011).

Bile acids as modulators of enzyme activity and stability. Protein J. *30*, 539–545.

Rombouts, I., Lagrain, B., Scherf, K.A., Lambrecht, M.A., Koehler, P., and Delcour, J.A. (2015). Formation and reshuffling of disulfide bonds in bovine serum albumin demonstrated using tandem mass spectrometry with collision-induced and electron-transfer dissociation. Sci. Reports 2015 51 *5*, 1–12.

Röttgen, P., and Collins, J. (1995). A human pancreatic secretory trypsin inhibitor presenting a hypervariable highly constrained epitope via monovalent phagemid display. Gene *164*, 243–250.

Rouet, R., Lowe, D., Dudgeon, K., Roome, B., Schofield, P., Langley, D., Andrews, J., Whitfeld, P., Jermutus, L., and Christ, D. (2012). Expression of high-affinity human antibody fragments in bacteria. Nat. Protoc. 2012 72 *7*, 364–373.

Rouet, R., Lowe, D., and Christ, D. (2014). Stability engineering of the human antibody repertoire. FEBS Lett. *588*, 269–277.

Rouet, R., Jackson, K.J.L., Langley, D.B., and Christ, D. (2018). Next-Generation Sequencing of Antibody Display Repertoires. Front. Immunol. *0*, 118.

Roychaudhuri, R., Sarath, G., Zeece, M., and Markwell, J. (2004). Stability of the allergenic soybean Kunitz trypsin inhibitor. Biochim. Biophys. Acta - Proteins Proteomics *1699*, 207–212.

Ryle, A.P. (1970). The Porcine Pepsins and Pepsinogens. Methods Enzymol. *19*, 316–336.

Sarkany, R.P.E., and Moreland, B.H. (1985). Enhancement of the autocatalytic activation of trypsinogen to trypsin by bile and bile acids. BBA - Gen. Subj. *839*, 262–267.

Sawangjit, R., Khan, T.M., and Chaiyakunapruk, N. (2017). Effectiveness of pharmacy-based needle/syringe exchange programme for people who inject drugs: a systematic review and meta-analysis. Addiction *112*, 236–247.

Schatz, P. (1993). Use of Peptide Libraries to Map the Substrate Specificity of a Peptide-Modifying Enzyme: A 13 Residue Consensus Peptide Specifies Biotinylation in Escherichia coli. Nat. Biotechnol. *11*, 1138–1143.

Schechter, I., and Berger, A. (1967). On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. *27*, 157–162.

Schechter, N.L., Zempsky, W.T., Cohen, L.L., McGrath, P.J., McMurtry, C.M., and Bright, N.S. (2007). Pain Reduction During Pediatric Immunizations: Evidence-Based Review and Recommendations. Pediatrics *119*, e1184–e1198.

Schilling, O., Biniossek, M.L., Mayer, B., Elsässer, B., Brandstetter, H., Goettig, P., Stenman, U.-H., and Koistinen, H. (2018). Specificity profiling of human trypsin-isoenzymes. Biol. Chem. *399*, 997–1007.

Schoene, C., Fierer, J.O., Bennett, S.P., and Howarth, M. (2014). SpyTag/SpyCatcher cyclization confers resilience to boiling on a mesophilic enzyme. Angew. Chem. Int. Ed *53*, 6101–6104.

Schoene, C., Bennett, S.P., and Howarth, M. (2016). SpyRing interrogation:

Analyzing how enzyme resilience can be achieved with phytase and distinct cyclization chemistries. Sci. Rep. *6*, 21151.

Schreiber, S., Fedorak, R.N., Nielsen, O.H., Wild, G., Williams, C.N., Nikolaus, S., Jacyna, M., Lashner, B.A., Gangl, A., Rutgeerts, P., et al. (2000). Safety and efficacy of recombinant human interleukin 10 in chronic active Crohn's disease. Gastroenterology *119*, 1461–1472.

Seeger, M.A., Zbinden, R., Flütsch, A., Gutte, P.G.M., Engeler, S., Roschitzki-Voser, H., and Grütter, M.G. (2013). Design, construction, and characterization of a second-generation DARPin library with reduced hydrophobicity. Protein Sci. *22*, 1239–1257.

Shaaltiel, Y., Gingis-Velitski, S., Tzaban, S., Fiks, N., Tekoah, Y., and Aviezer, D. (2015). Plant-based oral delivery of β-glucocerebrosidase as an enzyme replacement therapy for Gaucher's disease. Plant Biotechnol. J. *13*, 1033–1040.

Shahid, N., and Daniell, H. (2016). Plant-based oral vaccines against zoonotic and non-zoonotic diseases. Plant Biotechnol. J. *14*, 2079–2099.

Shim, H. (2017). Antibody Phage Display. (Springer), pp. 21–34.

Singh, P.K. (2008). Significance of phytic acid and supplemental phytase in chicken nutrition: a review. Worlds. Poult. Sci. J. *64*, 553–580.

Škrlec, K., Štrukelj, B., and Berlec, A. (2015). Non-immunoglobulin scaffolds: a focus on their targets. Trends Biotechnol. *33*, 408–418.

Smart, A.L., Gaisford, S., and Basit, A.W. (2014). Oral peptide and protein delivery: Intestinal obstacles and commercial prospects. Expert Opin. Drug Deliv. *11*, 1323–1335.

Smith, G.P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science *228*, 1315–1317.

Smith, G.P. (2019). Phage Display: Simple Evolution in a Petri Dish (Nobel Lecture). Angew. Chemie Int. Ed. *58*, 14428–14437.

Smith, G.P., Patel, S.U., Windass, J.D., Thornton, J.M., Winter, G., and Griffiths, A.D. (1998). Small binding proteins selected from a combinatorial repertoire of knottins displayed on phage. J. Mol. Biol. *277*, 317–332.

Sommerfeld, V., Künzel, S., Schollenberger, M., Kühn, I., and Rodehutscord, M. (2018). Influence of phytase or myo-inositol supplements on performance and phytate degradation products in the crop, ileum, and blood of broiler chickens. Poult. Sci. *97*, 920–929.

Song, H.K., and Suh, S.W. (1998). Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from Erythrina caffra and tissue-type plasminogen activator. J. Mol. Biol. *275*, 347–363.

Spinelli, S., Frenken, L.G.J., Hermans, P., Verrips, T., Brown, K., Tegoni, M., and Cambillau, C. (2000). Camelid heavy-chain variable domains provide efficient combining sites to haptens. Biochemistry *39*, 1217–1222.

Steeland, S., Vandenbroucke, R.E., and Libert, C. (2016). Nanobodies as

therapeutics: Big opportunities for small antibodies. Drug Discov. Today *21*, 1076–1113.

Steiner, D., Forrer, P., Stumpp, M.T., and Plückthun, A. (2006). Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. Nat. Biotechnol. *24*, 823–831.

Stoop, A.A., and Craik, C.S. (2003). Engineering of a macromolecular scaffold to develop specific protease inhibitors. Nat. Biotechnol. *21*, 1063–1068.

Sulea, T., Hussack, G., Ryan, S., Tanha, J., and Purisima, E.O. (2018). Application of Assisted Design of Antibody and Protein Therapeutics (ADAPT) improves efficacy of a Clostridium difficile toxin A single-domain antibody. Sci. Rep. *8*, 1–11.

Sung, D., Ahn, K.M., Lim, S.-Y., and Oh, S. (2014). Allergenicity of an enzymatic hydrolysate of soybean 2S protein. J. Sci. Food Agric. *94*, 2482–2487.

Surroca, Y., Haverkamp, J., and Heck, A.J.R. (2002). Towards the understanding of molecular mechanisms in the early stages of heat-induced aggregation of β-lactoglobulin AB. J. Chromatogr. A *970*, 275–285.

Svihus, B. (2011). The gizzard: function, influence of diet structure and effects on nutrient availability. Worlds. Poult. Sci. J. *67*, 207–224.

Sweet, R.M., Wright, H.T., Janin, J., Chothia, C.H., and Blow, D.M. (1974). Crystal structure of the complex of porcine trypsin with soybean trypsin inhibitor (kunitz) at 2.6 Å resolution. Biochemistry *13*, 4212–4228.

Szardenings, M., Vasel, B., Hecht, H., Collins, J., and Schomburg, D. (1995). Highly effective protease inhibitors from variants of human pancreatic secretory trypsin inhibitor (hPSTI): an assessment of 3-D structure-based protein design. Protein Eng. *8*, 45–52.

Tariq, R., Singh, S., Gupta, A., Pardi, D.S., and Khanna, S. (2017). Association of gastric acid suppression with recurrent clostridium difficile infection: A systematic review and meta-analysis. JAMA Intern. Med. *177*, 784–791.

Taverner, A., MacKay, J., Laurent, F., Hunter, T., Liu, K., Mangat, K., Song, L., Seto, E., Postlethwaite, S., Alam, A., et al. (2020). Cholix protein domain I functions as a carrier element for efficient apical to basal epithelial transcytosis. Tissue Barriers *8*, 1710429.

Teesalu, T., Sugahara, K.N., Kotamraju, V.R., and Ruoslahti, E. (2009). C-end rule peptides mediate neuropilin-1-dependent cell, vascular, and tissue penetration. Proc. Natl. Acad. Sci. *106*, 16157–16162.

Thakral, S., Thakral, N.K., and Majumdar, D.K. (2013). Eudragit®: a technology evaluation. Expert Opin. Drug Deliv. *10*, 131–149.

Thie, H., Voedisch, B., Dübel, S., Hust, M., and Schirrmann, T. (2009). Affinity maturation by phage display. Methods Mol. Biol. *525*.

Torres, J.F., Lyerly, D.M., Hill, J.E., and Monath, T.P. (1995). Evaluation of formalin-inactivated Clostridium difficile vaccines administered by parenteral and mucosal routes of immunization in hamsters. Infect. Immun. *63*, 4619–4627.

Tóthová, L., Bábíčková, J., and Celec, P. (2012). Phage survival: The biodegradability of M13 phage display library in vitro. Biotechnol. Appl. Biochem. *59*, 490–494.

Tyagi, P., Pechenov, S., and Anand Subramony, J. (2018). Oral peptide delivery: Translational challenges due to physiological effects. J. Control. Release *287*, 167–176.

Unkauf, T., Hust, M., and Frenzel, A. (2018). Antibody Affinity and Stability Maturation by Error-Prone PCR. In Phage Display. Methods in Molecular Biology, M. Hust, and T. Lim, eds. (Humana Press, New York, NY), pp. 393–407.

Vazquez-Lombardi, R., Phan, T.G., Zimmermann, C., Lowe, D., Jermutus, L., and Christ, D. (2015). Challenges and opportunities for non-antibody scaffold drugs. Drug Discov. Today *20*, 1271–1283.

Villequey, C., Kong, X.-D., and Heinis, C. (2017). Bypassing bacterial infection in phage display by sequencing DNA released from phage particles. Protein Eng. Des. Sel. *30*, 761–768.

Virdi, V., Coddens, A., De Buck, S., Millet, S., Goddeeris, B.M., Cox, E., De Greve, H., and Depicker, A. (2013). Orally fed seeds producing designer igas protect weaned piglets against enterotoxigenic escherichia coli infection. Proc. Natl. Acad. Sci. U. S. A. *110*, 11809–11814.

Virdi, V., Palaci, J., Laukens, B., Ryckaert, S., Cox, E., Vanderbeke, E., Depicker, A., and Callewaert, N. (2019). Yeast-secreted, dried and food-admixed monomeric IgA prevents gastrointestinal infection in a piglet model. Nat. Biotechnol. *37*, 527–530.

Voth, D.E., and Ballard, J.D. (2005). Clostridium difficile toxins: Mechanism of action and role in disease. Clin. Microbiol. Rev. *18*, 247–263.

Walk, C.L., Santos, T.T., and Bedford, M.R. (2014). Influence of superdoses of a novel microbial phytase on growth performance, tibia ash, and gizzard phytate and inositol in young broilers. Poult. Sci. *93*, 1172–1177.

Wallace, A.L., Schneider, M.I., Toomey, J.R., Schneider, R.M., Klempner, M.S., Wang, Y., and Cavacini, L.A. (2020). IgA as a potential candidate for enteric monoclonal antibody therapeutics with improved gastrointestinal stability. Vaccine *38*, 7490–7497.

Wang, L., and Kallenbach, N.R. (1998). Proteolysis as a measure of the free energy difference between cytochrome c and its derivatives. Protein Sci. *7*, 2460–2464.

Wang, J., Yadav, V., Smart, A.L., Tajiri, S., and Basit, A.W. (2015a). Toward Oral Delivery of Biopharmaceuticals: An Assessment of the Gastrointestinal Stability of 17 Peptide Drugs. Mol. Pharm. *12*, 966–973.

Wang, Q., Euler, C.W., Delaune, A., and Fischetti, V.A. (2015b). Using a Novel Lysin To Help Control Clostridium difficile Infections. Antimicrob. Agents Chemother. *59*, 7447–7457.

Wang, R., Edrington, T.C., Storrs, S.B., Crowley, K.S., Ward, J.M., Lee, T.C., Liu, Z.L., Li, B., and Glenn, K.C. (2017). Analyzing pepsin degradation assay conditions used for allergenicity assessments to ensure that pepsin susceptible and pepsin

resistant dietary proteins are distinguishable. PLoS One *12*, e0171926.

Wang, Y., Vander Horn, P., and Xi, L. (2003). Methods of using improved polymerases (United States Patent Office: US8232078B2).

Warny, M., Fatimi, A., Bostwick, E., Laine, D., Lebel, F., LaMont, J., Pothoulakis, C., and Kelly, C. (1999). Bovine immunoglobulin concentrate-clostridium difficile retains C difficile toxin neutralising activity after passage through the human stomach and small intestine. Gut *44*, 212–217.

Warr, G.W., Magor, K.E., and Higgins, D.A. (1995). IgY: clues to the origins of modern antibodies. Immunol. Today *16*, 392–398.

Weidmann, J., and Craik, D.J. (2016). Discovery, structure, function, and applications of cyclotides: circular proteins from plants. J. Exp. Bot. *67*, 4801–4812.

Wen, J.-Y., Ledger, A.R., Mcleod, A.B.J., Davies, A.N.M., Butt, A.G., and Tucker, A.I.G. (2002). Protein and peptide degradation in the intestine of the common brushtail possum (Trichosurus vulpecula). J. Comp. Physiol. *172*, 553–559.

Van de Wiele, T., Van den Abbeele, P., Ossieur, W., Possemiers, S., and Marzorati, M. (2015). The Simulator of the Human Intestinal Microbial Ecosystem (SHIME ®). In The Impact of Food Bioactives on Health: In Vitro and Ex Vivo Models, (Springer International Publishing), pp. 305–317.

Wilcox, M.H., Gerding, D.N., Poxton, I.R., Kelly, C., Nathan, R., Birch, T., Cornely, O.A., Rahav, G., Bouza, E., Lee, C., et al. (2017). Bezlotoxumab for Prevention of Recurrent *Clostridium difficile* Infection. N. Engl. J. Med. *376*, 305–317.

Williams, A., and Baird, L.G. (2003). DX-88 and HAE: A developmental perspective. In Transfusion and Apheresis Science, (Elsevier Ltd), pp. 255–258.

Winter, G. (2019). Harnessing Evolution to Make Medicines (Nobel Lecture). Angew. Chemie Int. Ed. *58*, 14438–14445.

Wodzinski, R.J., and Ullah, A.H. (1996). Phytase. Adv. Appl. Microbiol. *42*, 263–302.

Woyengo, T.A., and Nyachoti, C.M. (2011). Review: Supplementation of phytase and carbohydrases to diets for poultry. Can. J. Anim. Sci. *91*, 177–192.

Yadav, V., Varum, F., Bravo, R., Furrer, E., and Basit, A.W. (2016). Gastrointestinal stability of therapeutic anti-TNF α IgG1 monoclonal antibodies. Int. J. Pharm. *502*, 181–187.

Yamamoto, A., Taniguchi, T., Rikyuu, K., Tsuji, T., Fujita, T., Murakami, M., and Muranishi, S. (1994). Effects of Various Protease Inhibitors on the Intestinal Absorption and Degradation of Insulin in Rats. Pharm. Res. *11*, 1496–1500.

Yang, Z., Ramsey, J., Hamza, T., Zhang, Y., Li, S., Yfantis, H., Lee, D., Hernandez, L., Seghezzi, W., Furneisen, J., et al. (2015). Mechanisms of protection against Clostridium difficile infection by the monoclonal antitoxin antibodies actoxumab and bezlotoxumab. Infect. Immun. *83*, 822–831.

Zahnd, C., Wyler, E., Schwenk, J., Steiner, D., Lawrence, M., McKern, N., Pecorari, F., Ward, C., Joos, T., and Plückthun, A. (2007). A designed ankyrin repeat protein evolved to picomolar affinity to Her2. J. Mol. Biol. *369*, 1015–1028.

Zhao, N., Schmitt, M.A., and Fisk, J.D. (2016). Phage display selection of tight specific binding variants from a hyperthermostable Sso7d scaffold protein library. FEBS J. *283*, 1351–1367.

Zhao, N., Spencer, J., Schmitt, M.A., and Fisk, J.D. (2017). Hyperthermostable binding molecules on phage: Assay components for point-of-care diagnostics for active tuberculosis infection. Anal. Biochem. *521*, 59–71.

Zhou, D., Hansen, D., Shabalin, I.G., Gustchina, A., Vieira, D.F., de Brito, M.V., Araújo, A.P.U., Oliva, M.L.V., and Wlodawer, A. (2015). Structure of BbKI, a disulfide-free plasma kallikrein inhibitor. Acta Crystallogr. Sect. F *71*, 1055–1062.

Zhu, H., Hart, C.A., Sales, D., and Roberts, N.B. (2006). Bacterial killing in gastric juice - Effect of pH and pepsin on Escherichia coli and Helicobacter pylori. J. Med. Microbiol. *55*, 1265–1270.

Zimmermann, I., Egloff, P., Hutter, C.A.J., Arnold, F.M., Stohler, P., Bocquet, N., Hug, M.N., Huber, S., Siegrist, M., Hetemann, L., et al. (2018). Synthetic single domain antibodies for the conformational trapping of membrane proteins. Elife *7*, e34317.

Appendix

Amino acid sequences of proteins employed in this thesis

SBTI alanine mutants are not listed separately. Only some gastrobody sequences are listed. Proteins are listed in order of appearance.

SBTI-His_6

MSGDFVLDNEGNPLENGGTYYILSDITAFGGIRAAPTGNERCPLTVVQSRNELDKGIGTIIS
SPYRIRFIAEGHPLSLKFDSFAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAMDGWFRLE
RVSDDEFNNYKLVFCPQQAEDDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQKLDKESLGA
AHHHHHH

Anti-IgG-Sso7d-His_6

MATVKFKYKGEEKEVDISKIKKVWRDRLAAVFTYDEGGGKTGYGWVFTKDAPKELLQMLEKQ
KKAAAGSARGSHHHHHHEYSNGGAS

Anti-CROP-His_6 (nanobody)

QVKLEESGGGLVQAGGSLRLSCAASGRTFNTLSMGWFRQAPGKEREFVCAVSRSGGSTYYAD
SVKGRFTCSRDNAKNTVYLQMNSLKPEDTAVYYCAAAATKSNTTAYRLSFDYWGQGTQVTVS
SGAAHHHHHH

His_6- Thrombin site-SpyTag003-SBTI

MGSSHHHHHHSSGLVPRGSRGVPHIVMVDAYKRYKGSGESGDFVLDNEGNPLENGGTYYILS
DITAFGGIRAAPTGNERCPLTVVQSRNELDKGIGTIISSPYRIRFIAEGHPLSLKFDSFAVI
MLCVGIPTEWSVVEDLPEGPAVKIGENKDAMDGWFRLERVSDDEFNNYKLVFCPQQAEDDKC
GDIGISIDHDDGTRRLVVSKNKPLVVQFQKLDKESL

AviTag-His_6-CROP

MSGLNDIFEAQKIEWHEGSGHHHHHHGSGYQNKFLTLNGKKYYFDNDSKAVTGWQTIDGKKY
YFNLNTAEAATGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFNTNTFIASTGYTSINGKH
FYFNTDGIMQIGVFKGPNGFEYFAPANTDANNIEGQAILYQNKFLTLNGKKYYFGSDSKAVT
GLRTIDGKKYYFNTNTAVAVTGWQTINGKKYYFNTNTSIASTGYTIISGKHFYFNTDGIMQI
GVFKGPDGFEYFAPANTDANNIEGQAIRYQNRFLYLHDNIYYFGNNSKAATGWVTIDGNRYY
FEPNTAMGANGYKTIDNKNFYFRNGLPQIGVFKGSNGFEYFAPANTDANNIEGQAIRYQNRF
LHLLGKIYYFGNNSKAVTGWQTINGKVYYFMPDTAMAAAGGLFEIDGVIYFFGVDGVKAPGI
YG

His_6- Thrombin site-SpyTag003-T12

MGSSHHHHHHSSGLVPRGSRGVPHIVMVDAYKRYKGSGESGDFVLDNEGNPLENGGTYYILS
IAGTFGGIRAAPTGNERCPLTVVQSLVFPDKGIGTIISSPYRIRFIAEGHPLSLKFDSFAVI
MLCVGIPTEWSVVEDLPEGPAVKIGENKDAMDGWFRLERVSDDEFNNYKLVFCPQQAEDDKC
GDIGISIDHDDGTRRLVVSKNKPLVVQFQKLDKESL

AviTag-His$_6$-GTD

MSGLNDIFEAQKIEWHEGSGHHHHHHGSGSLVNRKQLEKMANVRFRTQEDEYVAILDALEEY
HNMSENTVVEKYLKLKDINSLTDIYIDTYKKSGRNKALKKFKEYLVTEVLELKNNNLTPVEK
NLHFVWIGGQINDTAINYINQWKDVNSDYNVNVFYDSNAFLINTLKKTVVESAINDTLESFR
ENLNDPRFDYNKFFRKRMEIIYDKQKNFINYYKAQREENPELIIDDIVKTYLSNEYSKEIDE
LNTYIEESLNKITQNSGNDVRNFEEFKNGESFNLYEQELVERWNLAAASDILRISALKEIGG
MYLDVDMLPGIQPDLFESIEKPSSVTVDFWEMTKLEAIMKYKEYIPEYTSEHFDMLDEEVQS
SFESVLASKSDKSEIFSSLGDMEASPLEVKIAFNSKGIINQGLISVKDSYCSNLIVKQIENR
YKILNNSLNPAISEDNDFNTTTNTFIDSIMAEANADNGRFMMELGKYLRVGFFPDVKTTINL
SGPEAYAAAYQDLLMFKEGSMNIHLIEADLRNFEISKTNISQSTEQEMASLWSFDDARAKAQ
FEEYKRNYFEGSL

His$_6$- Thrombin site-SpyTag003-GT_S1_01

MGSSHHHHHHSSGLVPRGSRGVPHIVMVDAYKRYKGSGESGDFVLDNEGNPLENGGTYYILS
EFGIGLFGGIRAAPTGNERCPLTVVQSLFRNHRDKGIGTIISSPYRIRFIAEGHPLSLKFDS
FAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAMDGWFRLERVSDDEFNNYKLVFCPQQAE
DDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQKLDKESL

His$_6$- Thrombin site-SpyTag003-GT01

MGSSHHHHHHSSGLVPRGSRGVPHIVMVDAYKRYKGSGESGDFVLDNEGNPLENGGTYYILS
DYGRQLFGGIRAAPTGNERCPLTVVQSLFRNHRDKGIGTIISSPYRIRFIAEGHPLSLKFDS
FAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAMDGWFRLERVSDDEFNNYKLVFCPQQAE
DDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQKLDKESL

His$_6$- Thrombin site-SpyTag003-GT01 R63A

MGSSHHHHHHSSGLVPRGSRGVPHIVMVDAYKRYKGSGESGDFVLDNEGNPLENGGTYYILS
DYGRQLFGGIRAAPTGNERCPLTVVQSLFRNHRDKGIGTIISSPYAIRFIAEGHPLSLKFDS
FAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAMDGWFRLERVSDDEFNNYKLVFCPQQAE
DDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQKLDKESL

His$_6$- Thrombin site-SpyTag003-GT47

MGSSHHHHHHSSGLVPRGSRGVPHIVMVDAYKRYKGSGESGDFVLDNEGNPLENGGTYYILS
DYGRQLFGGIRAAPTGNERCPLTVVQSLFRNHRDKGIGTIISSPYRIRFIAEGHPLSLKFDS
FAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAMDGWFRLERVRPTSSTGNNYKLVFCPQQ
AEDDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQKLDKESL

BBKI-His$_6$

MSGSVVVDTNGQPVSNGADAYYLVPVSHGHAGLALAKIGNEAEPRAVVLDPHHRPGLPVRFE
SPLRINIIKESYFLNIKFGPSSSDSGVWDVIQQDPIGLAVKVTDTKSLLGPFKVEKEGEGYK
IVYYPERGQTGLDIGLVHRNDKYYLAVKDGEPCVFKIRKATDEGAAHHHHHH

ECTI-His$_6$

MSGVLLDGNGEVVQNGGTYYLLPQVWAQGGGVQLAKTGEETCPLTVVQSPNELSDGKPIRIE
SRLRSAFIPDDDKVRIGFAYAPKCAPSPWWTVVEDEQEGLSVKLSEDESTQFDYPFKFEQVS
DQLHSYKLLYCEGKHEKCASIGINRDQKGYRRLVVTEDYPLTVVLKKDESSGAAHHHHHH

WBCI-His$_6$

MSGDDDLVDAEGNLVENGGTYYLLPHIWAHGGGIETAKTGNEPCPLTVVRSPNEVSKGEPIR
ISSQFLSLFIPRGSLVALGFANPPSCAASPWWTVVDSPQGPAVKLSQQKLPEKDILVFKFEK
VSHSNIHVYKLLYCQHDEEDVKCDQYIGIHRDRNGNRRLVVTEENPLELVLLKAKSETASSH
GAAHHHHHH

Scripts used during computational analysis

Rosetta

Command to generate a blueprint file called '1avu_missing_loops.remodel' for use in *remodel* calculations.

```
/mnt/c/Users/niels/Documents/1-
Rosetta/rosetta_bin_linux_2018.09.60072_bundle/tools/remodel/getBluePrin
tFromCoords.pl -pdbfile 1AVU.pdb -chain A > 1avu missing loops.remodel
```

Script to run Rosetta *remodel* to model missing density in PDB ID: 1AVU. I used a bash loop to allow running the same Rosetta commands on multiple structures. The script is run from an executable bash file. The for-loop reads a batch file called 'list.txt' which includes the names of PDB files that I would like to analyse. In this instance, 'list.txt' only contains one structure which is '1avu.pdb'. I re-use this batch structure for future Rosetta commands.

```
#!/bin/bash


### remove previous files.

rm *.sc

### link paths for rosetta to call on

ln -s /mnt/c/Users/niels/Documents/1-
Rosetta/rosetta_bin_linux_2018.09.60072_bundle/main/source/bin bin

ln -s /mnt/c/Users/niels/Documents/1-
Rosetta/rosetta_bin_linux_2018.09.60072_bundle/main/database database

ln -s /mnt/c/Users/niels/Documents/1-
Rosetta/rosetta_bin_linux_2018.09.60072_bundle/demos demos


### read batch list.

for i in $(cat list.txt); do

      name=$(echo $i | sed 's/.pdb//g')


      ../rosetta_bin_linux_2018.09.60072_bundle/main/source/bin/remodel.
static.linuxgccrelease -database database -in:file:s $name.pdb -
out:file:silent 1avu_out.o -remodel:blueprint 1avu_missing_loops.remodel
-run::chain A -num_trajectory 3 -nstruct 30 -ex1 -ex2 -overwrite


done

exit
```

Script to extract the REU scores from the output file ‘1avu_out.o’ generated from the *remodel* calculations. Scores are sorted from highest to lowest energy. Sorted scores are placed in a file called ‘sorted_scores.sc’.

```
grep '^SCORE' 1avu_relax_out.o > scores.sc

sort -k1,1 -k2n scores.sc > sorted_scores.sc
```

Script to extract the names of the top 10 lowest energy structures from the ‘sorted_scores.sc’ file.

```
#Take top 10 lowest energy structures and list their names

awk '{ print $23 }' sorted_scores.sc > top10.tag



#Run extract_pdbs

../rosetta_bin_linux_2018.09.60072_bundle/main/source/bin/extract_pdbs.s
tatic.linuxgccrelease -database database -in:file:silent 1avu_out.o -
in:file:tagfile top10.tag
```

Script for running relax for five iterations. It is used with the same batch structure described above for *remodel*.

```
../rosetta_bin_linux_2018.09.60072_bundle/main/source/bin/relax.static.l
inuxgccrelease -database database -in:file:s $name.pdb -nstruct 5
```

Script for running relax for 500 iterations in the batch structure. A flag is added to make the output silent and write it into a file called ‘1avu_relax_out.o’.

```
../rosetta_bin_linux_2018.09.60072_bundle/main/source/bin/relax.static.l
inuxgccrelease -database database -in:file:s $name.pdb -out:file:silent
1avu_relax_out.o -nstruct 500
```

Script for calling on *score.static* for calculating the RMSD of all the output structures in ‘1avu_relax_out.o’ relative to the lowest energy structure ‘1avu_loops_relax1_0131.pdb’. The output is written into ‘rms_scores.sc’.

```
../rosetta_bin_linux_2018.09.60072_bundle/main/source/bin/score.static.l
inuxgccrelease -database database -in:file:silent 1avu_relax_out.o -
in:file:native 1avu_loops_relax1_0131.pdb -out:file:scorefile
```

Script for extracting the residue name, residue number and Q-score from the score file generated from the POPS server (‘1AVU_SASA.csv’) and writing it into a new file called ‘SASA_Q.txt’.

```
awk '{print $1, $3, $8}' 1AVU_SASA.csv > SASA_Q.txt
```

Python script for generating the mutant list file for *pmutscan*. The output file is called ‘surface_residues.txt’. The output file contains a list of all combinations of surface accessible residues to be mutated to all other possible amino acids.

```
#import libraries

import sys

import pandas as pd


#Load the files and rename the columns

sasa = pd.read_csv('./SASA_Q.txt',sep=' ', header=None)

sasa.columns = ['res', 'ID', 'QS']


#Creating a dictionary for three letter code to 1 letter code

d = {'CYS': 'C', 'ASP': 'D', 'SER': 'S', 'GLN': 'Q', 'LYS': 'K',

     'ILE': 'I', 'PRO': 'P', 'THR': 'T', 'PHE': 'F', 'ASN': 'N',

     'GLY': 'G', 'HIS': 'H', 'LEU': 'L', 'ARG': 'R', 'TRP': 'W',

     'ALA': 'A', 'VAL':'V', 'GLU': 'E', 'TYR': 'Y', 'MET': 'M'}


#Write all the residue IDs where QS > 0.2 to a new file with according
to the mutants_list syntax for rosetta. It should be chain_ID
original_residue residue_position mutant_residue. Mutations to cysteine
will not be made.


with open("surface_residues.txt", "w") as f:

    for i in range(0,177):                                #Iterate through
the positions in structure

        if sasa.loc[sasa.index[i],'QS'] > 0.2:            #Check if SASA >
0.2

                resi = sasa.loc[sasa.index[i],'res']    #Set variables

                resi = d[resi]

                ID = sasa.loc[sasa.index[i],'ID']

                for x in d:                              #Iterate through
dictionary of amino acids

                    if d[x] != 'C':                      #Not for
cysteine, print mutations in rosetta syntax

                        f.write("A "+str(resi)+" "+str(ID)+"
"+str(d[x])+"\n")
```

Script for running *pmutscan* on all structures of the ensemble. Here the batch structure is used again to enable running *pmutscan* on each of the structures in the chosen ensemble. The ‘pmut_list.txt’ file contains the list of all structures in the ensemble. The loop iterates to run the *pmutscan* command with the mutant list ‘surface_residues.txt’ on all structures in ‘pmut_list.txt’. For each structure a score file is generated and labelled with the name of the structure and the ‘.sc’ file extension. The score file lists each mutation made and the predicted ΔREU.

```
#!/bin/bash


### remove previous files.

rm *.sc

### link to rosetta

ln -s /mnt/c/Users/niels/Documents/1-
Rosetta/rosetta_bin_linux_2018.09.60072_bundle/main/source/bin bin

ln -s /mnt/c/Users/niels/Documents/1-
Rosetta/rosetta_bin_linux_2018.09.60072_bundle/main/database database

ln -s /mnt/c/Users/niels/Documents/1-
Rosetta/rosetta_bin_linux_2018.09.60072_bundle/demos demos


### read batch list.

for i in $(cat pmut_list.txt); do

      name=$i


      ../rosetta_bin_linux_2018.09.60072_bundle/main/source/bin/pmut_sca
n_parallel.static.linuxgccrelease -database database -in:file:s $name -
mutants_list surface_residues.txt -DDG_cutoff 999 -
ignore_unrecognized_res -no_his_his_pairE -multi_cool_annealer 10 -mute
basic core -ex1 -ex2 > $name.sc


done


exit
```

Rosetta output data processing

This script calculates the average and standard deviation of all the ΔΔG (ΔREU) scores for each individual mutation (e.g. A25G) from the ensemble. The output is stored in ‘aveSDdata.csv’.

```
lines=$(wc -l < list.txt)
for i in $(cat list2.txt) ; do
      grep $i *.sc | awk '{print $4}' > numlist.tx
      nums=$(cat numlist.tx)
      #sum=$(grep $i *.sc | awk '{print $4}' | paste -s -d+ | bc)
      sum=$( paste -s -d+ numlist.tx | bc)
      ave=$(echo "scale=4; $sum / $lines" | bc)
      standardDeviation=$(
    echo "$nums" |
        awk '{sum+=$1; sumsq+=$1*$1}END{print sqrt(sumsq/NR -
(sum/NR)**2)}'
)
      echo $i $ave $standardDeviation | sed 's/ /,/g'  >> aveSDdata.csv
done
```

This script shapes the data in ‘aveSDdata.csv’ into a matrix where the x-axis is the residue that was mutated to, and the y-axis is the residue position. The output is stored in ‘ave_data.csv’ (averages) and ‘stdev_data.csv’ (standard deviations).

```
#Import numpy for use in the script
import numpy as np
import csv
#Read in the data files
stdev = np.genfromtxt('aveSDdata.csv', delimiter=',', usecols=3)
ave = np.genfromtxt('aveSDdata.csv', delimiter=',', usecols=2)
#Re-shape into appropriate matrix
stdev = np.reshape(stdev, (119,19))
ave = np.reshape(ave, (119,19))

#Write arrays to csv files
np.savetxt("ave_data.csv", ave, delimiter=",")
np.savetxt("stdev_data.csv", stdev, delimiter=",")
```