Structural and Functional Investigation of a Type III CRISPR Immune Pathway and a Bacterial Sialic Acid TRAP Transporter

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Niels Schneberger

aus

Bergisch Gladbach

Bonn, 2024

Angefertigt am Institut für Strukturbiologie
der Rheinischen Friedrich-Wilhelms-Universität Bonn
mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Gutachter / Betreuer: PD Dr. Gregor Hagelüken
Gutachter: Prof. Dr. Ulrich Kubitscheck

Tag der Promotion: 14.11.2024

Erscheinungsjahr: 2024

Table of Contents

Preface **1**

PART I **3**

Abstract **5**

1. Introduction Part I **7**
1.1 Bacterial pathogens 7
1.2 TRAP transporters mediate the sialic acid import in *H. influenzae* and *V. cholerae* 12
1.3 Substrate binding proteins 15
1.4 Of Antibodies and Nanobodies 19

2. Results Part I **25**
2.1 SiaP specific VHH antibodies and their specific antigen complexes 25
2.2 VHHs inhibit Neu5Ac binding by SiaP 37

3. Discussion Part I **47**
3.1 Similarities and differences between the presented VHH antibodies 47
3.2 Biochemical and structural analyses of the VcSiaP/NbS001/NbS002 complex 49
3.3 Mechanistical relevance of the identified conformation-specific surface cavities 50
3.4 Conformation-specific cavities as starting point for drug discovery 52
3.5 Outlook 55

PART II **57**

Abstract **59**

4. Introduction Part II **61**
4.1 Bacterial immunity 61
4.2 CRISPR provides adaptive immunity against bacteriophages and plasmids 61
4.3 Lon proteases are conserved elements to react on cellular stress 70
4.4 Bacterial regulation of transcriptional initiation incorporates sigma factors 72

5. Results Part II **77**
5.1 CalpL is a cyclic nucleotide activated Lon protease 77
5.2 CalpL is activated by cA_4-induced oligomerization 84
5.3 Site specific cleavage requires CalpT–L complex formation 90
5.4 Elucidating the role of CalpT in a phage defense signaling cascade 95

6. Discussion Part II **103**
6.1 Analysis of the gene neighborhood conservation of CalpL 103
6.2 Mechanistic details of the CalpS-T-L mediated immune response 109
6.3 Outlook 115

7. Methods **119**
7.1 Molecular genetics and cloning 119
7.2 General methods for protein biochemistry 122
7.3 Recombinant protein expression and purification 124
7.4 Analytical methods 128
7.5 Crystallization and Structure Determination 135
7.6 Design of a VcSiaP mutant that can be locked in its closed conformation 137

8. Materials 139
8.1 Chemicals 139
8.2 Consumables 140
8.3 Enzymes, buffers and more 140
8.4 Buffers and media 141
8.5 Bacterial strains 142
8.6 Plasmids and Oligonucleotides 143
8.7 Instruments and Columns for protein purification and analysis 145
8.8 Software 146
9. Additional Material 147
9.1 Additional Figures (Part I) 147
9.2 Additional Tables (Part I) 148
9.3 Additional Figures (Part II) 151
9.4 Additional Tables (Part II) 154
References 157
List of Abbreviations 180
List of Tables 181
List of Figures 182

Acknowledgements

Many thanks to PD. Dr. Gregor Hagelüken for the excellent supervision, the valuable ideas and the exciting discussions. Furthermore, I would like to thank Prof. Dr. Ulrich Kubitscheck, Prof. Dr. Arne Lützen and Prof. Dr. Jonathan Schmid-Burgk for being part of my PhD committee. I would also like to acknowledge Prof. Dr. Matthias Geyer for welcoming me in his group, letting me use all equipment and laboratories, and for his suggestions and opinions during discussions.

I am grateful for the opportunity to meet and collaborate with many scientists over the past four years. Thank you for the interesting meetings and the willingness to solve scientific problems together.

Also, thank you to my colleagues at the Institute of Structural Biology for the great working atmosphere, the good exchange and the mostly result-oriented discussions.

I wish my grandmothers, who would have been very proud, could have seen the day of my thesis defense. And I am all the more grateful that my parents, [redacted], and my brother, [redacted], were able to experience this phase of my life with me and supported me throughout that time.

Most of all I would like to thank my loving wife [redacted] and our lovely children [redacted] for all the love, support and understanding they have shown.

Preface

In nature, a very general scenario prevails: living organisms must protect themselves, adapt, or die. This holds true from the simplest to the most advanced life forms, resulting in a never-ending evolutionary war between viruses, bacteria and eukaryotes. Multiple mechanisms have evolved to protect organisms from invaders, and these immune strategies vary widely in complexity, specificity, response time, and long-term protection. While all types of immune systems continuously optimize and adapt for different applications, invaders also developed strategies to evade, or even actively counteract, the host immune response, ending up in a highly dynamic arms race of increasingly sophisticated attack and defense strategies.

Recent findings reveal striking similarities between eukaryotic and procaryotic cell-autonomous immune mechanisms, suggesting a common ancestry. Beyond its evolutionary relevance, the significant conservation of essential cellular mechanisms among prokaryotes and eukaryotes may help to unravel human defense strategies by characterizing their bacterial counterparts. On the other hand, studying bacterial strategies to evade host immune mechanisms can help to fight pathogens more specifically and effectively.

PART I

Bacterial strategies to survive in the host environment.

Abstract

Some bacteria have evolved sophisticated strategies to survive in the host environment. For example, the pathogenic bacteria *Vibrio cholerae* and *Haemophilus influenzae* rely on the import of sialic acid by tripartite ATP-independent periplasmic (TRAP) transporters. In *V. cholerae* the availability of this sugar enhances pathogenicity, whereas *H. influenzae* uses it to modify its lipopolysaccharide coat and thereby actively evade the host immune response. A better understanding of the TRAP transporter-mediated sialic acid transport mechanism may help to combat these pathogens more efficiently. TRAP transporters employ dedicated substrate binding proteins (SBPs) that contribute to a selective and efficient transport process by binding the substrate and delivering it to the transporter. It is known that camelid-derived VHH antibodies are able to inhibit SBPs of ABC transporters and can serve as tools to investigate their mechanistic and structural details. We selected a set of 11 VHH antibodies that specifically target the sialic acid TRAP transporter SBPs (also called P-domains) from *H. influenzae* and *V. cholerae*. Two of these nanobodies were able to completely inhibit substrate binding to the SBP.

A thorough structural and biophysical characterization of the VHH/SBP complexes revealed an allosteric mechanism that does not only inhibit the high-affinity binding of sialic acid, but also triggers the release of already bound sialic acid from the binding pocket. Structure-guided mutagenesis revealed a previously unnoticed hydrophobic surface cavity as a key element in the conformational rearrangement of the SBP upon sialic acid binding. Our results shed new light on the structural mechanism of TRAP transporters, and provided evidence for a novel substrate release strategy. In addition, the analysis of SBP-VHH antibody complexes generated interesting starting points for the development of drugs to fight the serious diseases caused by *H. influenzae* and *V. cholerae*.

1. Introduction Part I

1.1 Bacterial pathogens

Bacteria are the most ubiquitous form of life and form the group of prokaryotes along with archaea. In contrast to eukaryotes, bacteria have a simpler structure and their genetic material is not enclosed in the nucleus, but is free in the cytoplasm. The cytoplasmic space is also used for all metabolic pathways and ribosomal protein expression. The cytoplasm is surrounded by a plasma membrane and a protective peptidoglycan layer. This murein layer is substantially thicker in Gram-positive bacteria than in Gram-negative bacteria, and in the latter, this structural strength-providing layer is wrapped by an outer-membrane made of glycolipids (Püschel et al., 2011).

As bacteria occur in almost every conceivable habitat on Earth, they also naturally colonize the human body, and they do so by an enormous number of 10^{13}-10^{14} (Sender et al., 2016). Some bacteria are highly beneficial for eukaryotic organisms, but others can cause severe infectious diseases. Since the discovery of the antibiotic effect of penicillin by Sir Alexander Fleming in 1929, many bacterial infections were successfully treated. However, frequent usage of antibiotics causes an adaption of many bacteria to develop antibiotic resistances. Unfortunately, this effect is rapidly increasing and simultaneously, the development of new antibiotic substances decreased drastically. Together, these troubling global trends are responsible for the deaths of patients from bacterial infections that were once treatable (Luepke et al., 2017; Ventola, 2015). To combat the emerging crisis of antibiotic resistances, a more specific treatment of bacterial infections is of high interest. However, to fight the cause of an infection more specifically and effectively, an extensive knowledge on the bacterial pathways which are crucial for pathogenicity is essential.

1.1.1 Pathogenic bacteria benefit from sialic acid uptake

In 2016, the World Health Organization (WHO) was asked to compile a priority list of antibiotic-resistant bacteria. Based on several criteria, such as mortality, prevalence of resistance, treatability, and more, research groups selected 20 bacterial pathogens and argued for their special importance in drug discovery. One of these pathogens was the Gram-negative bacterium *Haemophilus influenzae* (Tacconelli et al., 2018). *H. influenzae* is defined as a class 2 pathogen and causes infectious diseases such as meningitis, otitis media, conjunctivitis, and chronic obstructive pulmonary disease (COPD) (Bakaletz & Novotny, 2018; Casadevall & Pirofski, 1999). Not only due to the development of Ampicillin resistances by *Haemophilus influenzae* tybe b (Hib) and despite significant vaccination activities in western countries, the pathogen causes high mortality rates in the group of children aged less than five years (Peltola, 2000). This work focuses on the two class 2 pathogens *H. influenzae* and *V. cholerae* which were shown to rely on the import of sialic acid by tripartite ATP-independent periplasmic (TRAP) transporters to survive in the host environment (Almagro-Moreno & Boyd, 2009; Hood et al., 1999; E. Vimr et al., 2000). *V. cholerae* encodes the cholera toxin and is responsible for an acute diarrhea that primarily affects populations in the Third World. Especially for people with an impaired immune system, the disease can quickly become life-threatening (Vanden Broeck et al., 2007).

1.1.2 Sialic acids

The family of sialic acids comprises of more than 50 distinct acidic α-keto monosaccharides which share a nine-carbon backbone (Deng et al., 2013). Unlike many other monosaccharides that are ubiquitous in nature, sialic acids are predominantly found in vertebrates and a few higher non-vertebrates. However, most bacteria lack a biosynthesis pathway for these products of a condensation reaction of a neutral six-carbon unit and pyruvate (Angata & Varki, 2002). The high structural diversity among the different family members is reflected in the wide variety of important biological processes where they participate. As important components of glycoproteins, gangliosides and polysaccharides, sialic acids are known to play an important role in neural transmission, stabilization of glycoconjugates and the cell membrane, cancer metastasis and in diverse cellular processes including cell adhesion, signal transduction, and more (X. Chen & Varki, 2010; Ohtsubo & Marth, 2006; Schauer, 1985; B. Wang & Brand-Miller, 2003).

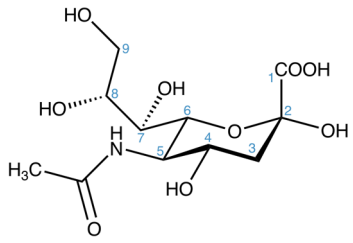


Figure 1-1: Chemical structure of sialic acid. *The most prominent member, N-acetylneuraminic acid (Neu5Ac), of the family of sialic acids is often, albeit confusingly, named sialic acid itself.*

At the vertebratal cell surface glycocalyx, sialic acids often occupy the terminal position where they are glycosidically linked to the underlying glycans via the C-2 carbon. This position makes them highly suitable for cell-cell interactions, intercellular signaling and immune recognition. In mammals, the most abundant representatives of sialic acids are N-acetylneuraminic acid (Neu5Ac) **(Figure 1-1)**, from which all other sialic acids are formed, and N-glycolylneuraminic acid (Neu5Gc) (Deng et al., 2013; Schauer, 1985).

1.1.3 Sialic acid utilization by the human pathogens *H. influenzae* and *V. cholerae*

Many commensal and pathogenic bacteria have evolved strategies to utilize sialic acids. For this, only few bacteria can rely on their own *de novo* pathways for sialic acid biosynthesis, whereas most others depend on scavenging of the sugar molecule or precursors of it from their mammalian host (E. R. Vimr et al., 2004). The pathogens discussed in this work use the latter route to benefit from sialic acids. However, free sialic acids are scarce in the cellular environment because the molecule is mostly attached to glycoconjugates. To overcome this, many bacteria, albeit not all, secrete a sialidase to free the sugar and make it available for import (Corfield, 1992). Once free and monomeric sialic acid is present, it must cross the outer membrane of Gram-negative bacteria via outer membrane transporters, such as porins (**Figure 1-2**.1). For the transport from the periplasm across the cytoplasmic membrane into the cytosol, a variety of transporters can be used in different bacterial species. They belong either to the family of primary ATP-binding cassette (ABC) transporters, TRAP transporters or other secondary active transporters (Bell et al., 2023). These different classes of transporters will be described and distinguished later in more detail. As mentioned above, *H. influenzae* and *V. cholerae* exclusively use TRAP transporters to accomplish this task (**Figure 1-2**.2). However, the utilization of cytoplasmic sialic acid differs between the two pathogens. While *V. cholerae* can only metabolize the sugar (**Figure 1-2**.3), *H. influenzae* can additionally

incorporate it into its lipopolysaccharide (LPS)-decorated cell surface to evade the human immune response (**Figure 1-2**.4/5) (Almagro-Moreno & Boyd, 2009; Bouchet et al., 2003; Hood et al., 2001; E. R. Vimr et al., 2004).

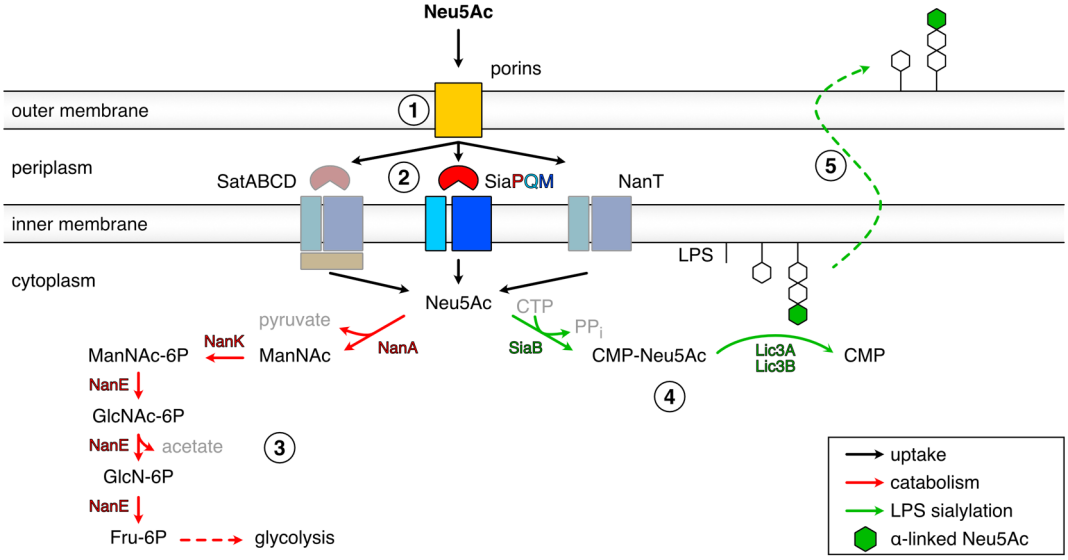


Figure 1-2: Overview of important key elements of the sialic acid utilization in bacteria.
To be transported into the cytoplasm of Gram-negative bacteria, Neu5Ac enters via porins into the periplasmic space (1). Further transport into the cytoplasm can be achieved by different transporter classes. The ABC transporter SatABCD, the TRAP transporter SiaPQM, and the secondary active transporter NanT are shown as examples (2). The uptake route is marked by black arrows. V. cholerae *can metabolize Neu5Ac after uptake (red arrows) (3).* H. influenzae *can additionally use sialic acid to evade the mammalian immune response by LPS sialylation (green arrows).*
This figure was modified from (Severi et al., 2007) and only shows pathways which are important for this work, but does not represent all known prokaryotic mechanisms for sialic acid utilization.

1.1.3.1 LPS sialylation by *H. influenzae* to evade the eukaryotic immune response

Recognition of bacterial pathogens by the eukaryotic immune system

Since bacteria are common invaders of eukaryotic organisms, pathways to combat harmful prokaryotes and distinguish them from commensal bacteria have evolved to be rapid and highly effective. The secret of this effective differentiation and control probably lies in the intricate interplay of many different signaling pathways as well as a successful cooperation between the innate and adaptive immune systems. (Gross et al., 2009; Srinivasan, 2010). One important component of the first line of defense against pathogens, which is provided by the innate immune system, is the recognition of pathogen associated molecular patterns (PAMPs) by pattern recognition receptors (PRRs). Such receptors are primarily expressed on the surface of immune cells, but can also be part of the cytosol. They particularly sense conserved molecular patterns of bacterial cell wall components, such as Lipoteichoic acid (LTA) and lipopolysaccharides (LPS), or fungi by their specific β-glucans. PAMP-recognition then leads to the induction of pro-inflammatory cytokines and antimicrobial effectors (Janeway Jr. & Medzhitov, 2002).

Early detection of Gram-negative bacteria is achieved by the recognition of picomolar LPS levels by the Toll-like receptor (TLR) 4 mediated immune response. In this multistep process, LPS is first bound by the LPS-binding protein (LBP) and transferred to the cluster of differentiation 14 (CD14) (**Figure 1-3**.1). The TLR4/myeloid differentiation-2 (MD-2) complex binds LPS which induces dimerization of the ternary complex

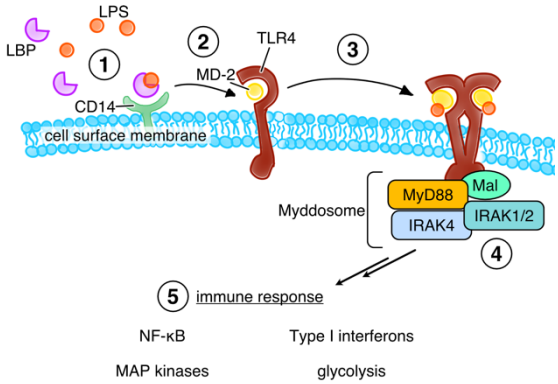


Figure 1-3: LPS induced TLR4 pathway.
LPS is bound by the LPS-binding protein (LBP) and transferred to the coreceptor CD14 (1). Upon recognition and binding to the TLR4 MD-2 complex (2) induces dimerization (3). This in turn leads to a recruitment of different adaptor proteins that form the myddosome (4) leading to the activation of several downstream pathways to cause an immune response (5). This figure was inspired by (Aboudounya & Heads, 2021).

(**Figure 1-3**.2/3). While the hydrophobic lipid chains are buried in a large binding pocket formed by MD-2, the glucosamine backbone of the LPS interacts with the TLR4. The exact orientation of the LPS seems to be dependent of the length and composition of lipid chains and might be one component to distinguish pathogens and commensals (Park et al., 2009). The assembly results in the recruitment of particular adaptor protein complexes, also termed the myddosome (Motshwene et al., 2009), to activate transcription factors such as NF-kB, followed by the expression of pro-inflammatory cytokines and interferon genes, activation of MAP kinases as well as initiation of downstream effects (**Figure 1-3**.4/5**)** (Zamyatina & Heine, 2020). Additional to this so-called canonical and TLR4 mediated pathway, a caspase mediated immune response was discovered in 2013. This non-canonical and TLR4-independent inflammatory pathway senses cytosolic LPS to finally trigger inflammasome activation in combination with pyroptosis and the release of IL-1β (Hagar et al., 2013). However, the combination of both pathways provides a robust and rapid protection of Gram-negative bacterial pathogens.

Strategies to prevent recognition by LPS-triggered antibacterial signaling pathways

For some bacterial pathogens that colonize the mucosa of their host organism, peculiarities in the architecture and composition of their LPS have been discovered. LPS synthesis as well as their sialylation are cytosolic processes. For the sialylation of LPS fragments, imported sialic acid is linked to cytidine monophosphate (CMP) by the synthetase SiaB. This activated form of sialic acid is then attached to the LPS. For *H. influenzae*, it was shown that the latter process is mostly catalyzed by the two sialyltransferases Lic3A and Lic3B **(Figure 1-2**.4**)**, with Lic3A being responsible for a monosialylation of the LPS and Lic3B for the attachment of mono- as well as disialic acid (Fox et al., 2006). After completion of the LPS fragments, these are exported and incorporated to the outer membrane of the Gram-negative bacterium (Sukupolvi-Petty et al., 2006) **(Figure 1-2**.5**)**. *H. influenzae* frequently forms short-chain LPS, some of which contain specific sialylated glycoforms. Hood and coworkers could observe for two different strains of the bacterium that for an increased resistance in human serum the availability of sialic acid was crucial. Furthermore, by mass spectrometric analysis they could identify the detailed composition of the LPSs and found significant differences between those of cells for which sialic acid was available and those for which it was not available. For both *H. influenzae* strains examined in the studies, Neu5Ac was observed at the very end of an additional tetrasaccharide unit that was not attached to the LPS in

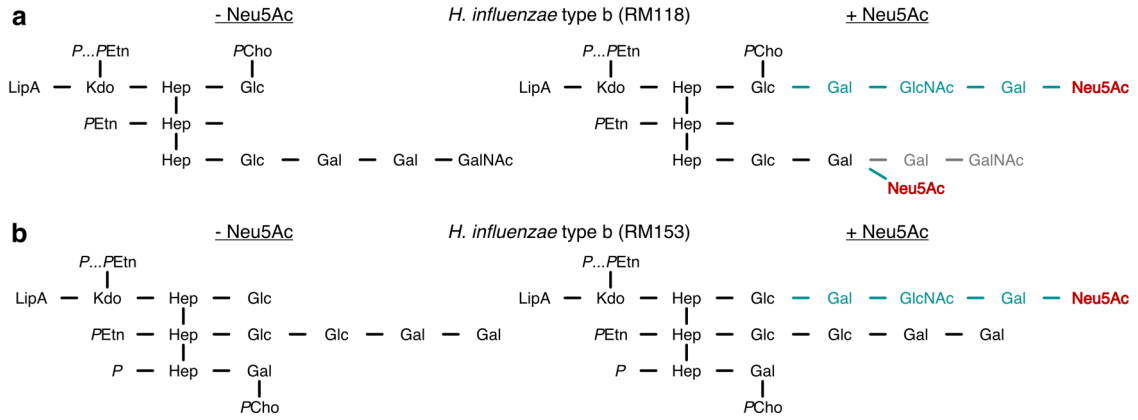


Figure 1-4: LPS composition of different* H. influenzae *strains dependent on Neu5Ac availability. ***a)*** *Schematic representation of the LPS composition of strain RM118 (type d) after colonization without (left) or with (right) sialic acid.* ***b)*** *The same as in a) but* H. influenzae *type b (strain RM153) Used abbreviations: Kdo, 2-keto-3-deoxyoctulosonic acid; Hep, L-glycero-D-manno-heptose; Glc, D-glucose; Gal, D-galactose; GalNAc, N-acetylgalactosamine;* P*Etn, phosphoethanolamine;* P*, phosphate;* P*Cho, phosphocholine; LipA, lipid A. This figure was adapted from (Hood et al., 2004).*

absence of sialic acid (Hood et al., 2001, 2004) (**Figure 1-4**). Within the past decades, LPS sialylation could be shown to be a major virulence factor of invasive *H. influenzae* strains (Bouchet et al., 2003; Hallström & Riesbeck, 2010; Hood et al., 2001). Decoration of the outermost layer of the bacterial cell surface with sialic acid actively provides resistance for human serum by masking the highly immunogenic LPS layer.

1.1.3.2 Sialic acid uptake increases pathogenicity of *Vibrio cholerae*

Commensals and pathogenic bacteria that colonialize the human gastrointestinal (GI) tract can benefit from the ability to catabolize sialic acids. The GI system itself essentially contributes to mammalian immunoregulation as it represents a large area that is steadily in contact with external substances (Vighi et al., 2008). One important component to protect this interfacial area is the mucus layer which is composed of highly glycosylated proteins (Lamont, 1992). These so-called mucins do not only protect the intestine, but are also important for regulatory processes and function as an attachment site and nutrient source for commensal bacteria as well as for pathogens (Arike & Hansson, 2016; Juge, 2012). Sialic acids, and especially Neu5Ac, are common components of the high molecular weight glycoproteins and their specific distribution increases from the ileum to the colon (Robbe et al., 2004; Robbe-Masselot et al., 2009). Considering the abundance of sialic acids in the gut, it is not surprising that the ability to catabolize sialic acid provides an evolutionary advantage in this particular habitat. This beneficial effect is reflected by significant correlations between a diet rich in sialic acid and the improved growth of bacteria that can catabolize the sugar (Coker et al., 2021).

The pathogen *V. cholerae* encodes a sialidase to liberate Neu5Ac from the mucins. The next step in the sialic acid scavenging pathway is the import across the outer- and inner membrane of the Gram-negative bacterium. For the latter transport process the TRAP transporter genes *vc1777-vc1779* are responsible. These genes are localized within the *Vibrio* pathogenicity island-2 (VPI-2) and are in close proximity to the *nan* gene cluster that encodes most of the enzymes required for the conversion of Neu5Ac to fructose 6-phosphate that serves as substrate for glycolysis (see **Figure 1-2**.3) (Chowdhury et al., 2012; Jermyn & Boyd, 2002). These important genes are the sialidase

gene (*nanH*) and those encoding the sialic acid catabolic enzymes NanA (Neu5Ac aldolase), NanK (N-acetylmannosamine kinase), NanE (ManNAc-6-phosphate epimerase) (E. R. Vimr et al., 2004). The fact that only pathogenic strains of *V. cholerae* contain the VPI-2 highlights the important role of sialic acid import and utilization in virulence and host colonialization which was also experimentally confirmed (Almagro-Moreno & Boyd, 2009).

1.2 TRAP transporters mediate the sialic acid import in *H. influenzae* and *V. cholerae*

The importance of molecular transport across biological membranes is reflected by the fact that transport proteins constitute roughly 10 % of most proteomes (Quick & Javitch, 2007). More than 40 different transporter families control the uptake of ions, nutrients, signaling molecules and many more solutes (Hediger et al., 2004). The two pathogens *H. influenzae* and *V. cholerae* essentially rely on the import of sialic acid, as they lack pathways for *de novo* synthesis of the sugar. Therefore, they employ so called tripartite ATP-independent periplasmic (TRAP) transporters to scavenge sialic acid from host tissues (Mulligan et al., 2011; Rosa et al., 2018). For both bacteria it could be shown, that the sialic acid-mediated increased virulence strongly depends on the presence and integrity of the transporter (Chowdhury et al., 2012; Severi et al., 2005).

While characterizing a C4 dicarboxylate system from *Rhodobacter capsulatus*, the group of David J. Kelly identified a substrate binding protein (SBP), called DctP, and its essentiality for transport, similar to ATP-binding cassette (ABC) transporter SBPs (Shaw et al., 1991). Sequence analysis to investigate this similarity revealed only two integral membrane protein encoding genes (*dctQ* and *dctM*) and no genes characteristic for nucleotide-binding domains (NBDs) which are typically associated with ABC transporters. Furthermore, functional studies showed that the DctPQM system was sensitive to inhibitors of secondary transporters, suggesting a membrane potential-dependent but ATP-independent transport mechanism (Forward et al., 1997). Furthermore, the transport mechanism was found to be dependent on a sodium ion gradient (Mulligan et al., 2009). Thus, these tripartite ATP-independent periplasmic (TRAP) transporters (**Figure 1-5 b**) can be considered as a structural and functional mix of ABC importers (**Figure 1-5 a**) and secondary active transporters (**Figure 1-5 c**). The ion gradient driven substrate transport (similar to secondary transporters) is initiated by dedicated SBPs (a hallmark of ABC transporters). The SBP contributes to a selective and effective transport process by binding the substrate in the periplasm of Gram-negative bacteria and delivering it to the transmembrane domains (termed Q- and M-domain) for translocation (Fischer et al., 2010; Forward et al., 1997; Peter, Ruland, et al., 2022).

Subsequent searches for TRAP transporter related genes in the genomes of various archaeal and bacterial organisms demonstrated the widespread use of these transporter family in microbial organisms (Kelly, 2001; Mulligan et al., 2007; Rabus et al., 1999). Furthermore, it was found that TRAP transporters are absent in eukaryotes, a feature shared with SBP-dependent ABC importers (Davidson et al., 2008). And as diverse and abundant as the different TRAP transporters are among the various prokaryotic organisms, so are their substrates. These range from C4-dicarboxylates over α-keto acids to sugar molecules (Mulligan et al., 2011). Unlike in classical secondary transporters but analogous to ABC transporters, the substrate specificity of TRAP transporters seems to be outsourced to their SBP (Maqbool et al., 2015).

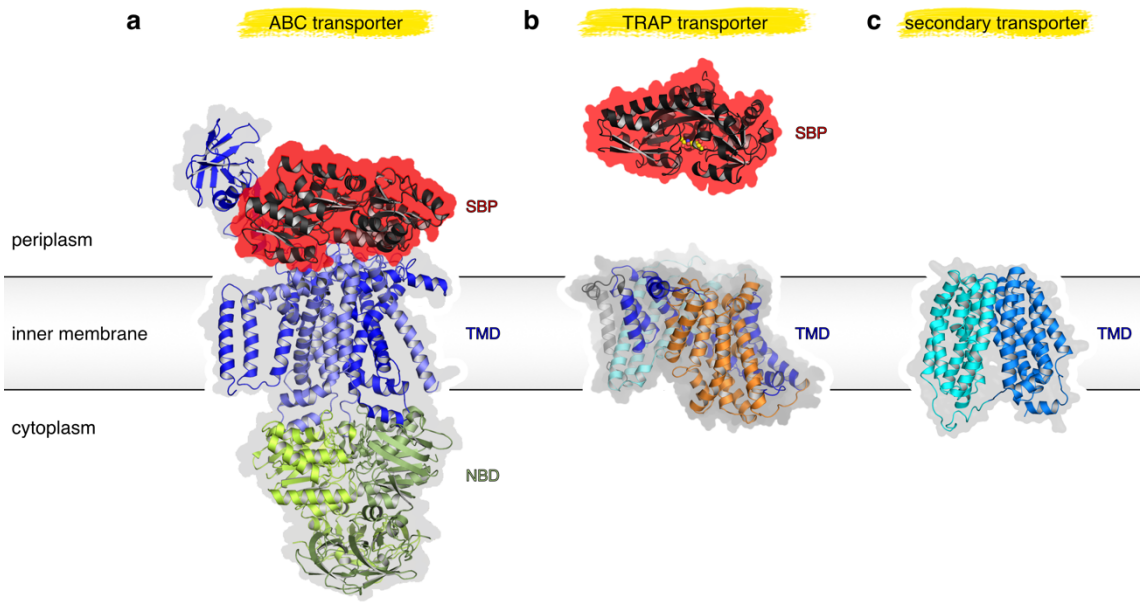


Figure 1-5: Comparing overview of three different transporter families.
a) *Primary ABC-transporter represented by the maltose importer MalEFGK$_2$ (PDB ID:2r6g). The SBP is depicted in black and red, the transmembrane domain (TMD) in blue and the nucleotide-binding domain (NBD) in green.* ***b)*** *TRAP transporter from* H. influenzae *composed of an SBP (black/red, PDB ID: 3b50) and the transmembrane domain formed by the QM proteins (blue/cyan/orange/gray, PDB ID: 7qe5).* ***c)*** *Secondary transporter represented by the lactose permease LacY that only consists of a TMD (cyan/blue, PDB ID: 1pv6). This figure was modified from (Mulligan et al., 2011).*

1.2.1 Architecture and transport mechanism of TRAP transporters

As the name implies, TRAP transporters consist of three individual domains that have been named according to the genes *dctP*, *dctQ* and *dctM* from *Rhodobacter capsulatus* encoding for the first discovered member of the family (Forward et al., 1997). The smaller Q- and larger M- transmembrane domains were analyzed according to their hydrophobic character and hypothesized to be built up of four and twelve helices, respectively (Forward et al., 1997). Sequence similarities to the 0160membrane domains of secondary transporters suggested the M-domain to form the substrate translocation channel (Forward et al., 1997; Rabus et al., 1999; Y. Shi, 2013). Both domains are essential for a functional transport mechanism and in some organisms, they are even fused into a single peptide chain (for example in HiSiaQM) (Kelly, 2001). In other organisms, like in *V. cholerae*, the two proteins are expressed individually but still build a stable complex (Mulligan et al., 2012). As no experimental structures existed, the function of the Q-domain was highly speculative for a long time. It was thought to act as a kind of landing platform for the mobile P-domain (Kelly, 2001) or might have chaperone-like properties (Mulligan et al., 2011). When the structure of the QM-domains was finally solved in 2022, this shed light onto the transport mechanism that was believed to work by a “Rocker Switch” mechanism (Mulligan et al., 2009) (Figure 1-6 **a**). The fused Q- and M-domains from *H. influenzae* consist of 15 transmembrane helices and two helical hairpins that do not cross the lipid bilayer. The M-domain is built from two similar repeat units that are wrapped on one side by the long and inclined α-helices of the Q-domain (Q1-Q4) (Peter, Ruland, et al., 2022). This untypical architecture might enable TRAP transporters to

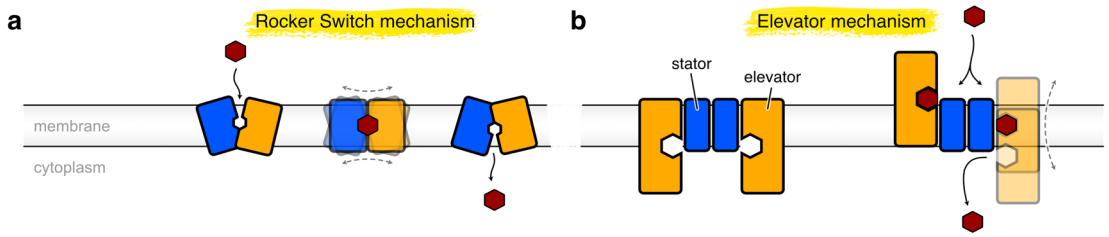


Figure 1-6: Different types of transport mechanisms across the cytoplasmic membrane.
a) *Representation of the "Rocker Switch" mechanism. The substrate binds to the outward open conformation leading to conformational changes towards the inward open state in which the substrate is released into the cytoplasm.* ***b)*** *Representation of an elevator-type mechanism. The stator (blue) is fixed inside the membrane, while the elevator (orange) can move up and down to translocate the substrate. The inward facing state (left) is considered as the resting state. To bind extracytoplasmic substrate, the elevator domain moves upwards to adopt the outward facing conformation. After substrate binding, the elevator moves down again to release the molecule into the cytosol. Black arrows show the translocation path of the substrate and dashed gray arrows the movement of the mobile domains.*

function by a monomeric elevator type mechanism with the extended Q-domain helices providing a rigid anchor for the stator domain (Figure 1-6 **b**), whereas elevator type transporters are normally known to form multimers inside the membrane (Nie et al., 2017; Verdon et al., 2014). Structural similarities of the HiSiaQM transporter domains to the dimeric elevator-type carboxylate transporter VcINDY reveal that one part of HiSiaM forms a bundle of α-helices that built up the mobile elevator domain including the two helical hairpins. The second part of the M-domain superimposes well onto one stator subunit of the VcINDY dimer while the Q-domain structurally mimics the oligomerization domain of the second VcINDY chain (Nie et al., 2017; Peter, Ruland, et al., 2022) (mapped onto the most left part of **Figure 1-7**). Using AlphaFold predictions, the tripartite complex of the soluble P-domain and the QM-transmembrane domains could be modeled in two independent groups. Sequence conservation in the P–QM interface across several TRAP transporters supports the prediction that the N-terminal lobe of the SBP interacts with the fixed stator domain while the C-terminal lobe binds to and moves along with the elevator domain during the transport cycle (J. S. Davies et al., 2023; Peter, Ruland, et al., 2022). The results suggested a conformational coupling of the soluble and membrane incorporated domains which was recently confirmed experimentally (Peter et al., 2024). According to this findings the transport mechanism of TRAP transporters was updated. In a first step, the P-domain specifically binds its extracytoplasmic substrate which induces a conformational rearrangement from the open- to the closed state. The membrane domains are unaffected by this event and remain in their inward-open state (C_i) which is most probably their resting state (**Figure 1-7**.1). In its closed, substrate-bound conformation, the SBP can bind to the inward-facing transporter domains (**Figure 1-7**.2). This binding event somehow triggers the upward movement of the elevator accompanied by the closed-to-open transition of the SBP which might allosterically cause a release of the substrate into the membrane translocation channel (Marinelli et al., 2011). Along with the substrate hand-over, two sodium ions are presumably bound to the translocation channel to be transported into the cytoplasm along with the substrate (**Figure 1-7**.3/4). The substrate-free P-domain dissociates in its open state from the transporter and the elevator 'falls' back into its resting inward open state (**Figure 1-7**.5) and the substrate as well as the Na^+ ions are released into the cytoplasm (**Figure 1-7**.6). After one transport cycle, all individual domains have returned their original, substrate-

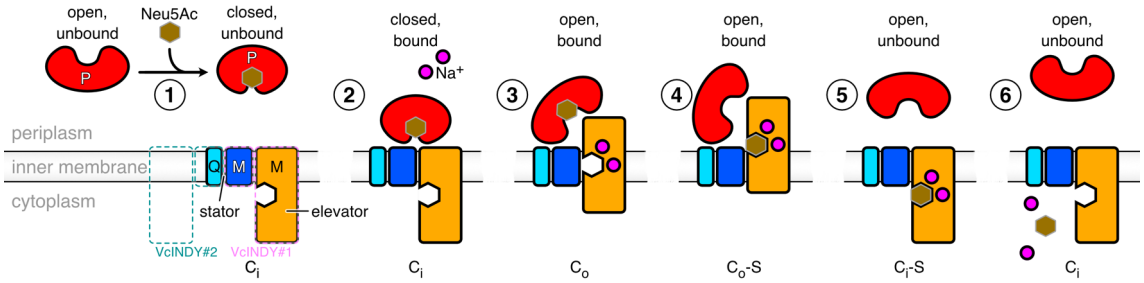


Figure 1-7: Working model for the mechanism of transport of TRAP transporters.
Schematic representation of the proposed transport mechanism showing the individual components P-domain (red), Q-domain (cyan) and M-domain (blue, orange) in their different conformational states according to the different steps (numbered 1-6). The stator and elevator subunits are indicated and their counterparts in the structurally related dimeric VcINDY carboxylate transporter are mapped by dashed boxes in the left panel. Furthermore, sialic acid (Neu5Ac, brown hexagon) and sodium ions (magenta spheres) are included in the sketch. A detailed and step-wise description of the mechanism is written in the main text. This figure was modified from (Peter, Ruland, et al., 2022).

free, configuration and are ready for another translocation cycle (J. S. Davies et al., 2023; Peter, Ruland, et al., 2022). Since the affinity of the closed P-domain to the QM-domains in its resting state is significantly higher than that of the open SBP and the closed conformation is only stable when sialic acid is bound, empty transport cycles are avoided (Peter et al., 2021, 2024). Especially in an environment where substrate is sparse, transporters benefit from the presence of an SBP, as it increases the effective substrate concentration around the translocating membrane domains and can serve as substrate store (Bosdriesz et al., 2015). Moreover, the directionality of TRAP transporter-mediated substrate translocation was shown to be imposed by the P-domain (Mulligan et al., 2009).

1.3 Substrate binding proteins

In 2016, an updated structural classification of SBPs found more than 500 SBP-type structures that can be divided into seven different classes (Scheepers et al., 2016). This large number not only represents the research interest in SBP-dependent processes, but also demonstrates their biological importance. Substrate binding proteins were first reported in the early 1970s as essential components of ABC transporter-mediated substrate import (Berger & Heppel, 1974; Higgins et al., 1990). More recently, these extracytoplasmic proteins, which specifically bind their substrates with submicromolar affinities, have been shown to be part of tripartite tricarboxylate transporters (TTTs) and TRAP transporters. In all these different transporter families, SBPs are responsible for specifically binding the substrate and delivering it to their cognate transmembrane domains for translocation across the plasma membrane into the cytoplasm. In Gram-negative bacteria the soluble binding proteins freely diffuse in the periplasm, whereas in Gram-positive bacteria they are anchored to the cell wall (Berntsson et al., 2010; Fischer et al., 2010; Maqbool et al., 2015). Despite the wide range of substrates and diverse functions that SBPs serve, they share a highly conserved three-dimensional structural fold, although sequence identities are often low (Scheepers et al., 2016). In general, SBPs (also called P-domains for TRAPs) consist of two αβ-domains with a central β-sheet flanked by α-helices. These two domains, that are also referred to as the N- and C-terminal lobes, respectively, are connected by an extended hinge helix (Berntsson et al., 2010; Müller et al., 2006; Trakhanov et al., 2005) (compare **Figure 1-9 a, b**). The substrate binding cleft is located in between the two αβ-domains which close around the ligand like

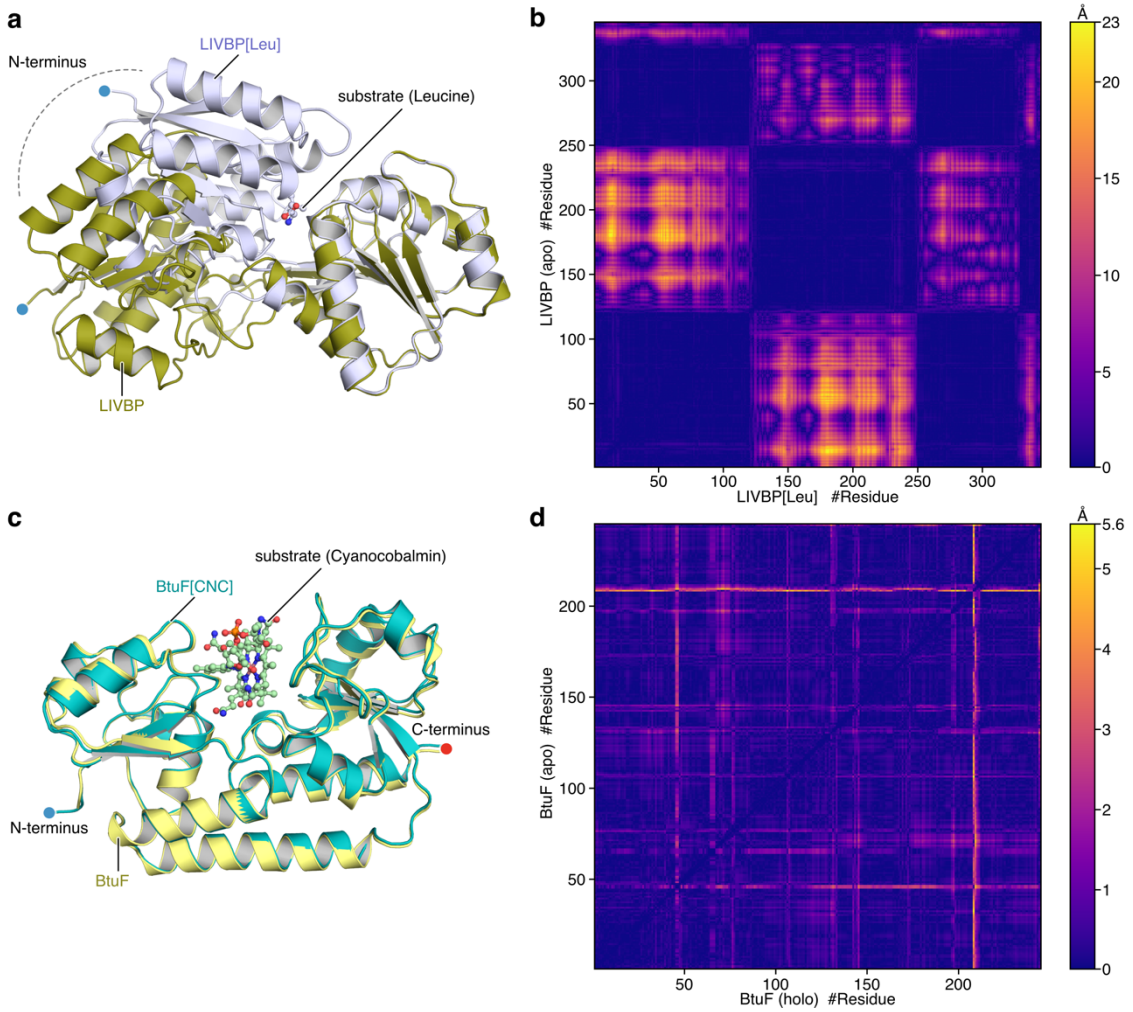


Figure 1-8: Conformational changes of ABC transporter substrate binding proteins.
a) *Superposition of crystal structures of the leucine/isoleucine/valine-binding protein (LIVBP or LivJ) in its substrate free (olive, PDB ID: 1z15) and leucine bound conformation (lightblue, PDB ID: 1z16). The structures have been aligned to their C-terminal lobe. The N-terminus is indicated by a blue sphere.* ***b)*** *Difference distance map to visualize the conformational differences between the apo and holo structures. Purple regions indicate low relative conformational changes and yellow regions show high differences. The scale for minimal and maximal values is shown on the right-hand side. The map was created using MtsslWizard (Hagelueken et al., 2012)* ***c)*** *Superposition of the crystal structures of the vitamin B12 binding protein (BtuF) in its substrate free (yellow, PDB ID: 1n4d) and substrate bound conformation (teal, PDB ID: 1n2z). The structures have been aligned to their N-terminal lobe. The N-terminus and C-terminus are indicated by a blue and red sphere, respectively.* ***d)*** *The same as in b) but for BtuF and using a different scale as indicated on the right-hand side.*

a Venus flytrap (Mao et al., 1982). However, the extent of conformational changes can vary drastically between different SBPs. While a relative structural rearrangement of the lobes by angles as large as 60° was observed for the leucine/ isoleucine/ valine binding protein (LIVBP or LivJ) (**Figure 1-8 a, b**) (Trakhanov et al., 2005), the lobe movements of the vitamin B12 ABC importer SBP, BtuF, show only small rearrangements (**Figure 1-8 c, d**) (Karpowich et al., 2003).

1.3.1 Sialic acid binding TRAP transporter SBPs

The substrate binding proteins of bacterial TRAP transporters have been biochemically and structurally characterized in detail and are the best known components of this transporter family. Like their homologs from ABC transporters, TRAP transporter

SBPs consist of N- and a C-terminal lobes connected by a hinge region (**Figure 1-9 a, b**) and undergo similar structural rearrangements upon substrate binding (**Figure 1-9 c**). The two distinct states were first reported for the sialic acid TRAP transporter P-domain from *H. influenzae* (Müller et al., 2006). These observations could be validated in (frozen) solution by EPR spectroscopy and FRET experiments for the closely related SBP of the VcSiaPQM TRAP transporter from *V. cholerae* (Glaenzer et al., 2017). Furthermore, no intermediate states were observed and the open-to-closed transition was shown to be strictly substrate-dependent. At low substrate concentrations, single molecule FRET (smFRET) experiments showed that the SBP spontaneously reopens to release the substrate without any external force (Peter et al., 2021). More sialic acid-specific TRAP transporter SBPs have been identified in *Fusobacterium nucleatum* (Fn) and *Pasteurella multocida* (Pm), and their binding affinities were observed in the same range (15 nM-50 nM) as for HiSiaP (P-domain of the sialic acid TRAP transporter from *H. influenzae*). However, VcSiaP shows a significantly lower affinity (200-300 nM) for its substrate, which is in the range of the P-domain of the sialic acid TRAP transporter from *Photobacterium profundum* (Pp) (J. S. Davies et al., 2023; Johnston et al., 2008; Peter et al., 2021; Setty et al., 2014). All of these Neu5Ac-specific SBPs share a highly conserved binding site and some were shown to also bind the closely related Neu5Gc, albeit with a reduced affinity (Setty et al., 2014). In 2019, Darby et al. found that the substrate affinity is not only affected by directly interacting amino acid residues, but also by more distant side chains through a fragile, highly defined and closely interconnected water network (Darby et al., 2019). For the direct interactions between the SBP and its substrate an important arginine residue (R147 in HiSiaP ≙ R145 in VcSiaP) stands out. It essentially coordinates the carboxylic acid moiety of Neu5Ac and is conserved between different organisms. Interestingly, through a mutational approach they were able to engineer the SBP to bind a non-cognate ligand sialylamide, in which the carboxylate functional group

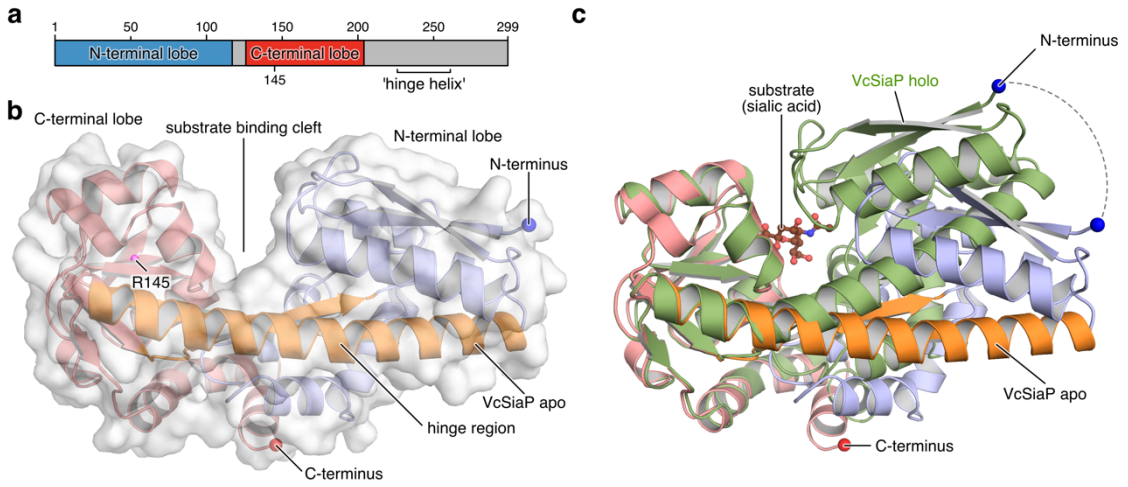


Figure 1-9: Domain architecture and closing transition of sialic acid TRAP transporter SBPs. ***a)*** *Schematic representation of the VcSiaP primary sequence with annotated structural motifs. The N-terminal lobe (blue) and the C-terminal lobe (red) are highlighted and the important R145 amino acid as well as the position of the hinge α-helix is indicated below.* ***b)*** *Structural overview of the open state of VcSiaP. N- and C-terminus are indicated by blue and red spheres, respectively. The N- and C-terminal lobes that form the substrate binding cleft are labeled and color coded as in a). The hinge region formed by the extended α-helix together with two beta strands is colored in orange. The position of R145 is annotated and higlighted by a magenta sphere.* ***c)*** *Superposition of the the substrate-free (PDB ID: 4mag) and -bound (PDB ID: 7a5q) structure of VcSiaP to visualize the conformational transition. The color coding of the apo state is similar as in b) and the holo state structure is colored in green. Sialic acid is depicted as brown ball and stick model. The structures have been aligned according to their C-lobe. The closing movement of the N-lobe is highligted by a dashed line.*

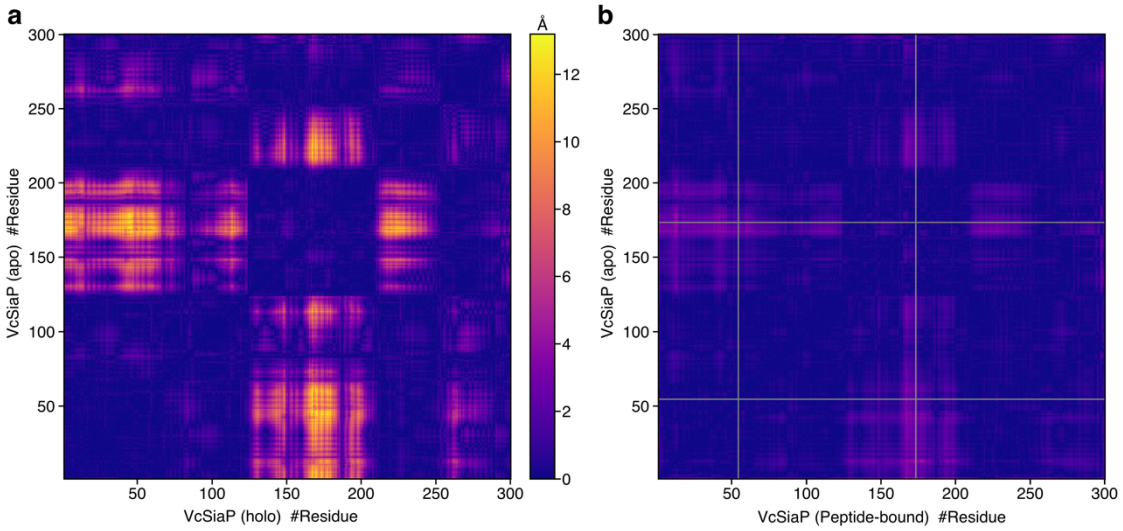


Figure 1-10: Natural or artificial substrates trigger the same structural rearrangements in VcSiaP. ***a)*** *Difference distance map between the substrate-free (PDB ID: 4mag) and -bound (PDB ID: 7a5q) conformation of the TRAP transporter SBP, VcSiaP. Relative changes of the distances between individual Cα atoms are indicated by a color gradient (purple (no changes) over magenta (medium changes) to yellow (maximum changes of 13.2 Å). The color gradient and the respective distances are explained in the legend on the right-hand side.* ***b)*** *The same as in a) but showing the conformational changes triggered by binding of an artificial peptide sequence (PDB ID: 7a5c). The same scale and color gradient was used for both maps. The maps were created using MtsslWizard (Hagelueken et al., 2012).*

of Neu5Ac is exchanged to an amide (Fischer et al., 2015). In 2021, our group found that the closing-mechanism appears to be triggered by physically bridging the gap between the N-lobe and the C-lobe. This induction of the conformational rearrangement by stabilization of the closed state is not only possible by the natural substrate Neu5Ac and structural homologs, but can also be induced by an artificial peptide sequence. Notably, the peptide did not even bind at the same position as sialic acid, but rather at the upper edge of the binding pocket (Peter et al., 2021). Nonetheless, the same conformational rearrangements were triggered, albeit to a lesser extent (**Figure 1-10**). The possibility of TRAP transporter SBPs to bind to artificial peptide sequences is particularly interesting with respect to the development of TRAP transporter inhibitors. For practical applications the affinity needs to be increased significantly. This could be achieved by artificially optimizing the peptide sequence in a structure-guided approach, however such approaches are often time-intense and not straight forward. At that time, we remembered a publication by Mireku and coworkers that described a VHH antibody binding to the substrate binding cleft of the vitamin B12 ABC transporter SBP BtuF with high affinity, leading to an inhibition of transport (Mireku et al., 2017).

1.4 Of Antibodies and Nanobodies

Immunoglobulins (Igs), also known as antibodies, are components of the eukaryotic adaptive immune system and play a pivotal role in the specific recognition and neutralization of pathogens. There are five primary classes of immunoglobulins present in the human blood plasma (IgG, IgA, IgM, IgD and IgE) and their expression in specialized white blood cells, the B-lymphocytes, is regulated by regulatory T-cells and T-helper cells. All immunoglobulins contain two identical heavy (H) chains and two identical light (L) chains (Lehninger, 1979; Püschel et al., 2011). The individual classes can be distinguished by the structure and composition of their heavy chains which are called α-chains (IgA), γ-chains (IgG), δ-chains (IgD), ε-chains(IgE), or μ-chains(IgM). The heavy chains contain an N-terminal variable domain (VH) and multiple constant domains (CH1-CH3 for IgG and IgA, CH1-CH3 for IgM and IgE), whereas each light chains comprises one N-terminal variable (VL) and one constant (CL) domain. Depending on the type of heavy chain, the distinct immunoglobulins function in different types and at particular stages of the immune response. IgD, IgE and IgG are very similar in their overall structure. IgM can either be bound as a monomer to the plasma membrane or is secreted to form a pentameric complex. IgA is also secreted and can be found as monomer, dimer or trimer (Püschel et al., 2011; Ridley, 1988).

1.4.1 Conventional IgGs

The most common antibodies in secondary immune responses are immunoglobulin-γ (IgG) antibodies which are produced in so called memory B-cells. Like other immunoglobulins, IgGs are linked by multiple conserved disulfide bonds that provide a balance of stability and flexibility to the characteristic Y-shaped structure. Typically, at least two disulfide bonds bridge the two heavy chains in the hinge region between the "trunk" and the "arms" of the macromolecule (H. Liu & May, 2012; Püschel et al., 2011). The "trunk" is formed by dimerization of the CH2 and CH3 of two individual heavy chains and is called the crystallizable fragment (Fc). This region is often glycosylated in the CH2 domain and is particularly important for the recruitment of immune cells (Lehninger, 1979; Muyldermans, 2013). Each "arm" of the macromolecule is built of two heavy chain domains (VH and CH1) which are covalently linked to the constant region of the light chain (CL) by one disulfide bridge. This antigen-binding fragment (Fab) can be cleaved off from the heavy chain domains CH2 and CH3 by either papain or pepsin (**Figure 1-11 a**). While incubation with pepsin results in a bridged Fab_2' dimer and an Fc' fragment, proteolytic cleavage with papain occurs N-terminally of the disulfide bonds in the hinge region and results in 2 single Fab fragments (**Figure 1-11 b**) and one Fc (Püschel et al., 2011). The antigen binding site is located in the N-terminal variable domains, VH and VL, of the Fab fragment. Six hypervariable loops, three located in the VH and three in the VL domain, built up the complementary determining regions (CDRs) which are responsible for antigen recognition and specificity. It could be shown that the Fab fragment is stable on its own and retains the ability to bind antigens (Poljak et al., 1972). Even smaller antigen-binding fragments can be produced by artificially linking the variable VH and VL domains yielding single-chain Fv (scFv) antibodies (**Figure 1-11 b**). Both variable chains are structurally similar and their backbone is exclusively built up of β-sheets. Conserved disulfide bonds between two individual β-sheets provide a rigid backbone. The CDRs are flexible loops on the N-terminal side of the subdomains (CDRH1-3 and CDRL1-3) (Asaadi et al., 2021) (**Figure 1-11 c, d**). Besides their

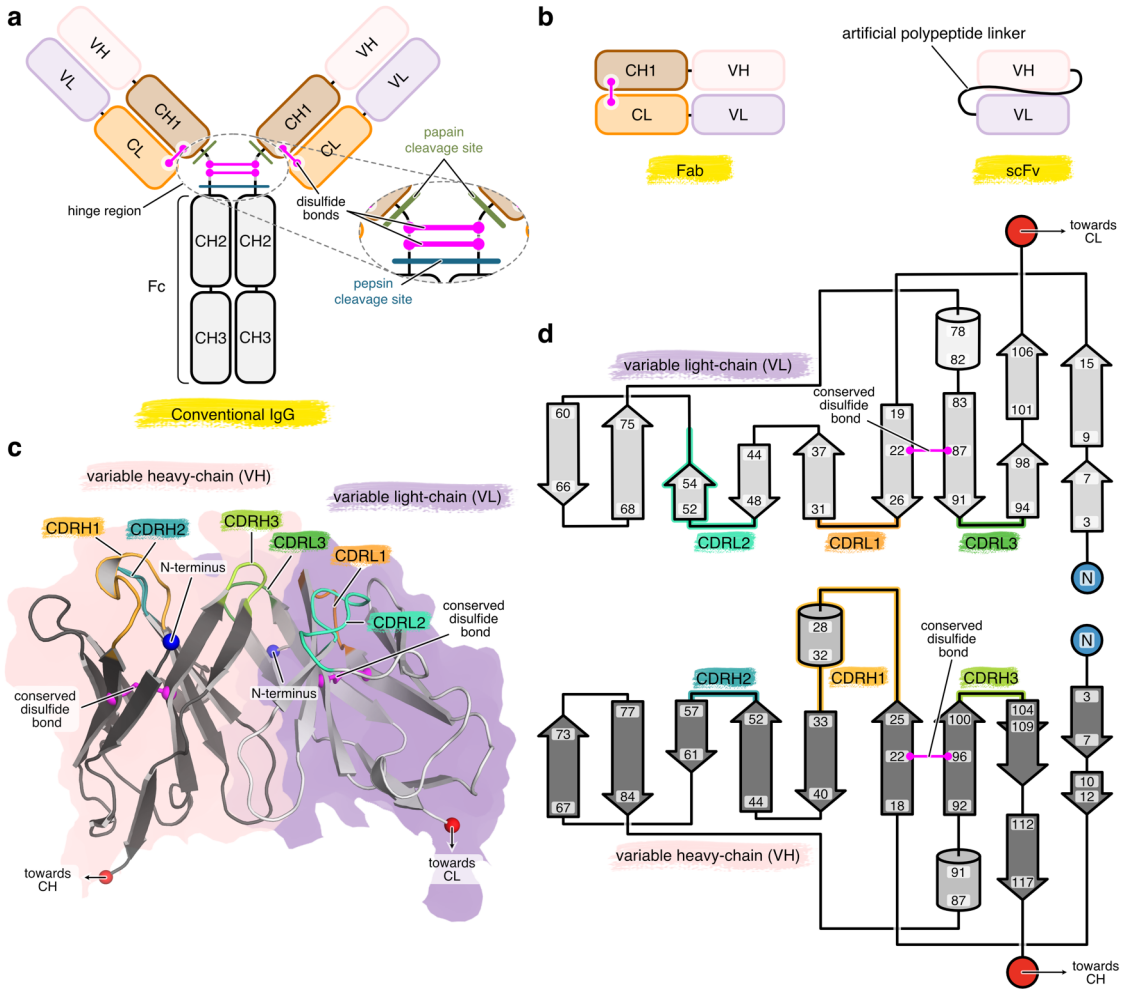


Figure 1-11: Overview on IgG antibodies and stable antigen binding fragments thereof. ***a)*** *Schematic representation of a conventional antibody built from two heavy (H) chains and two light (L) chains. Disulfide bonds that connect different polypeptide chains are indicated by magenta spheres and sticks. The hinge region is indicated and a magnification is shown as an inset where the papain cleavage site and the pepsin cleavage site is annotated.* ***b)*** *Sketch of a Fab fragment (left) and a single-chain variable fragment antibody which is linked by an artificial polypeptide chain (right).* ***c)*** *Structural overview of one variable fragment of an IgG (PDB ID: 2fbj). The complementary determining regions are colored and labeled as CDRH1-3 and CDRL1-3 for the heavy chain and the light chain, respectively. The conserved intramolecular disulfide bonds are indicated as magenta ball and stick representation and the N-terminus is highlighted by a blue sphere. The C-terminal end of the displayed sequence is indicated by a red sphere, note that the original structure continues at this point and merges into the CH1 and CL domain, respectively.* ***d)*** *Topology diagrams for the partial structures shown in c). The numbers describe the first and last amino acid residue of the particular structural motif. All positions and regions indicated in c) are also represented in the topology diagram. The color scheme was kept constant from a) to d).*

contribution to antigen binding, which is different for each CDR (Wilson & Stanfield, 1994), these loops play an essential role in the heterodimerization of VH and VL (Padlan, 1994). Unfortunately, scFv antibodies often exhibit reduced protein stability and antigen affinity and their production in prokaryotic expression systems is often non-trivial (Harmsen & Haard, 2007; Ward et al., 1989). Up to date, several biotechnological and therapeutic applications of antibodies and the antigen-binding fragments derived thereof have been reported. The development of antibody (-fragments) with optimized properties, such as enhanced analytical or diagnostic performance and improved pharmacokinetic

profiles, has been of great interest ever since (Holliger & Hudson, 2005; Sifniotis et al., 2019; Ward et al., 1989).

1.4.2 VHH antibodies

In 1993, scientists from the group of Raymond Hamers reported of the discovery of "naturally occurring antibodies devoid of light chains" in the serum of camels (Hamers-Casterman et al., 1993). This special type of antibodies is composed only of heavy-chain dimers and was therefore named as heavy chain only antibodies (HCAbs). Functional heavy-chain only immunoglobulins have been found exclusively in camelids and cartilaginous fish, such as sharks and rays (Greenberg et al., 1995). In both of these animal families, HCAbs and conventional Ig antibodies are produced by the adaptive immune system, but the exact ratio varies between the different species (Blanc et al., 2009; Muyldermans, 2013). While camelid HCAbs contain a typical IgG Fc region built of CH2 and CH3 domains and one variable domain (VHH) (**Figure 1-12 a**), Ig new antigen receptors (IgNARs) from fish are formed by two identical heavy chains composed of five constant subdomains ($C1_{NAR}$- $C5_{NAR}$) and one variable domain (V-NAR) (Hamers-Casterman et al., 1993; Roux et al., 1998). In camelids, the length of the hinge region can be either similar to that of conventional IgGs or significantly extended N-terminally of the conserved interchain disulfide bridge (**Figure 1-12 a**). The long hinge region contains a 12-fold repeated Pro-X motif (X: Gln, Glu, Lys) which provides structural rigidity to this region and may compensate for the lack of a CH1 region (Hamers-Casterman et al.,

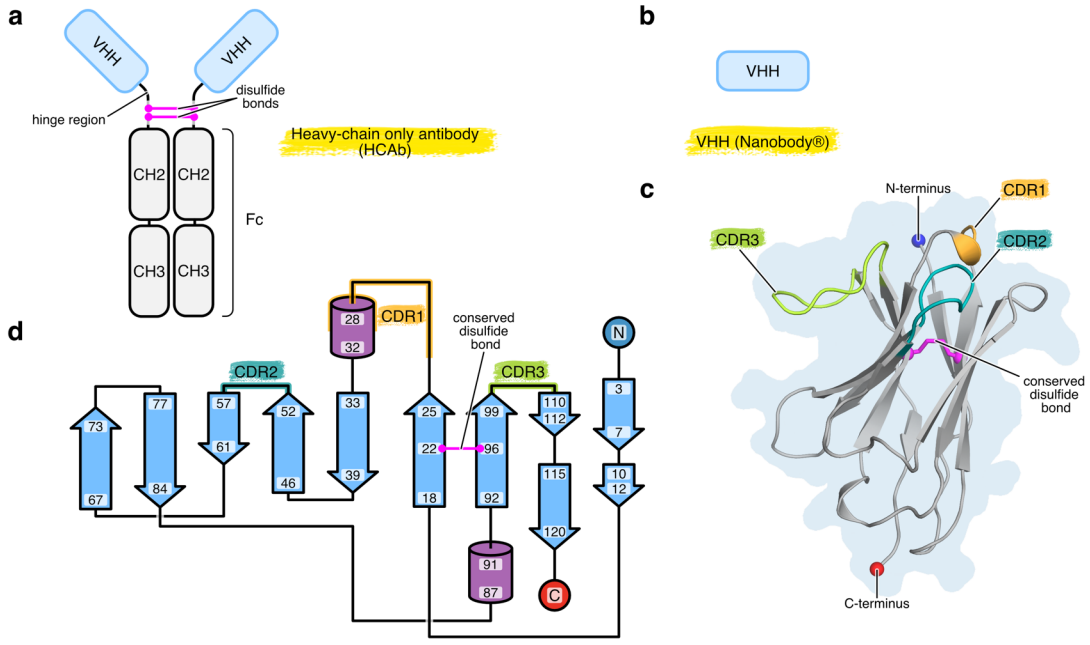


Figure 1-12: Overview on heavy chain only antibodies and VHH antibodies.
a) *Schematic representation of a heavy chain only antibody (HCAb) built from two heavy (H) chains. Disulfide bonds that connect different polypeptide chains are indicated by magenta spheres and sticks. The hinge region and the crystallizable fragment (Fc) region is indicated.* ***b)*** *Sketch of the heavy chain only variable fragment derived from a camelid HCAb.* ***c)*** *Structural overview of a VHH antibody. The complementary determining regions are colored and labeled as CDR1-3. The conserved intramolecular disulfide bonds are indicated as magenta ball and stick representation. The N-terminus and the C-terminus are highlighted by a blue and red sphere, respectively.* ***d)*** *Topology diagram for the structure shown in c). The numbers describe the first and last amino acid residue of the particular structural motif. All positions and regions indicated in c) are also represented in the topology diagram. The color scheme was kept constant from a) to d).*

1993). Just like conventional VH domains, the N-terminal variable domain of the heavy chain (VHH) contains three hypervariable loops, the CDRs, which are involved in antigen binding (Desmyter et al., 1996; Muyldermans et al., 1994) (**Figure 1-12 c, d**). Interestingly, the variable domains of shark heavy-chain only antibodies have only a rudimentary CDR2 that does not contribute to antigen binding (Streltsov et al., 2004). The complementary determining region 3 (CDR3) of HCAbs is often significantly longer than that of conventional immunoglobulins and, in addition to its essential role in antigen binding, is critical for the stability of the VHH. The hydrophobic VH-VL interface consists of conserved amino acid residues whose counterparts in VHHs are replaced by more hydrophilic amino acids to increase the overall hydrophilicity of the surface. In addition to these changes on sequence level, the CDR3 folds over the region that would be covered by the VL domain in IgGs (Padlan, 1994; Vu et al., 1997). When it comes designing single domain fragments, single domain VHHs (also called nanobodies) (**Figure 1-12 b**) essentially benefit from these subtle changes. This is reflected in a drastically increased solubility, stability and antigen affinity compared to the single chain variable fragments (scFvs) derived from IgGs (J. Davies & Riechmann, 1994; Muyldermans et al., 2001). Like IgGs, VHHs contain a highly conserved intramolecular disulfide bond between residues C20 and C96 (**Figure 1-12 c, d**), but unlike IgGs, many dromedary VHHs and also some llama VHHs have an additional disulfide bond between residue ~50 that provides additional rigidity to the extended (Harmsen & Haard, 2007).

1.4.2.1 The potential of nanobodies in biotechnology, pharmacy and therapy.

The number and range of possible applications of single-domain antibody fragments, derived from camelid heavy-chain only antibodies, has grown rapidly since their discovery in the early 1990ies (Jovčevska & Muyldermans, 2020). VHHs are about half the size (~15 kDa) of the smallest antigen-binding fragments that can be constructed from conventional IgGs (~30 kDa) and have improved solubility (Glockshuber et al., 1990; Hamers-Casterman et al., 1993; Muyldermans, 2013). The small size results in a fast tissue penetration and rapid clearance from the blood. Together with a high sequence similarity to human VHs these properties lead to a low immunogenicity of VHHs (Cortez-Retamozo et al., 2004; Harmsen & Haard, 2007). For application in clinical therapies, where an extended half-life in the blood is advantegous, this feature can be increased by the construction of fusion proteins (Kontermann, 2009). In addition, nanobodies tolerate a wide range of pH values and temperatures, as well as high concentrations of chemical denaturants, and can be engineered to be insensitive to proteolytic degradation (Dumoulin et al., 2002; Hussack et al., 2011; van der Linden et al., 1999). Due to their single-domain nature, nanobodies are much easier to select from display libraries than single-chain variable fragments designed by artificially linking the individual VH and VL domains from conventional antibodies. When scFv libraries are generated, the VH and VL domains must be amplified separately by PCR, resulting in a high number of dysfunctional VH-VL combinations upon random assembly. Therefore, VHH libraries can be several orders smaller while covering the same immune repertoire as VH-VL libraries (Gonzalez-Sapienza et al., 2017; Harmsen & Haard, 2007). Another advantage of the single-domain architecture and monomeric behavior of VHHs is the suitability for the development of multidomain constructs (Saerens et al., 2008). Different approaches to engineer dimeric nanobody constructs have been reported where either two identical VHHs, or two VHHs that target different epitopes of the same antigen were combined to

increase the functional affinity towards a specific antigen (Emmerson et al., 2011). It is even possible to tether two independent antigens by the combination of two nanobodies with each targeting a different antigen (Chames & Baty, 2009; Conrath et al., 2001). Furthermore, the production of VHHs in *E. coli* and the subsequent purification is simpler than that of scFvs which is often non-trivial and requires complex refolding steps (Arbabi-Ghahroudi et al., 1997, 2005).

All these properties in combination with a nanomolar antigen affinity and high specificity prone them for the use in biotechnological, diagnostic and therapeutic applications. Nanobodies are used as biosensors for proteins and small molecules (Bever et al., 2016; Pleschberger et al., 2004), as crystallization chaperones in macromolecular crystallography (Manglik et al., 2017; Nguyen et al., 2018), and even as additives in shampoos for the prevention of dandruff (Dolk et al., 2005). Several nanobody-based diagnostic *in vitro* tests have been developed and are used, for example, to screen for prostate cancer (Huang et al., 2005) or viral infections (Gelkop et al., 2018). The enormous potential in therapy is reflected by several VHH-based drugs in clinical and preclinical studies (Harmsen & Haard, 2007; Holliger & Hudson, 2005). The functionality and efficacy of such macromolecular substances in the inhibition of enzymes, cellular transport mechanisms, as well as in the neutralization of viruses have been demonstrated both *in vitro* and *in vivo* (Desmyter et al., 2002; Kang-Pettinger et al., 2023; Koenig et al., 2021; Peter, Ruland, et al., 2022). Monoclonal antibodies (also referred to as ‘first generation antibodies’) and antibody fragments (‘second generation antibodies’) have led to an overwhelming number of drugs approved by the United States Food and Drug Administration (FDA). In the literature, VHH antibodies are now referred to as ‘third generation antibodies’ and the first commercial medical applications are eagerly awaited (Arbabi-Ghahroudi, 2017).

2. Results Part I

Parts of this section have been submitted to Nature Communications Biology and are currently under review. The generation, panning and hit identification process for VcSiaP specific VHH antibodies was done by the Core Facility Nanobodies of the University of Bonn. The same holds for repanning of the thereby created Phage library with HiSiaP as bait protein. Erik Gehrke characterized the resulting NbS003 during his master thesis in 2021-2022. Panning and hit identification of HiSiaP specific nanobodies after a second immunization campaign as well as their biochemical characterization was done by Philipp Hendricks as one part of his master thesis in 2022. Sophie Binder constructed and cloned a pET-28a based expression vector for VHHs, containing a cleavable N-terminal His_6 tag and a pelB signal sequence during her work as research assistant. Some of the elucidated protein structures were made available on the Protein Data Bank (PDB), accession codes: (9fvb, 9fvc, 9fve).

2.1 SiaP specific VHH antibodies and their specific antigen complexes

As VHH antibodies were reported to be helpful tools for investigating mechanistical details of proteins (Manglik et al., 2017; Mireku et al., 2017; Valenciano-Bellido et al., 2023) and can be helpful for structural analyses (Pardon et al., 2014; Peter, Ruland, et al., 2022) we aimed to generate VHH antibodies against the sialic acid TRAP transporter P-domains from *Vibrio cholerae* and *Haemophilus influenzae*. The next paragraph shortly describes the antigen preparation and the immunization strategy using VcSiaP. For the

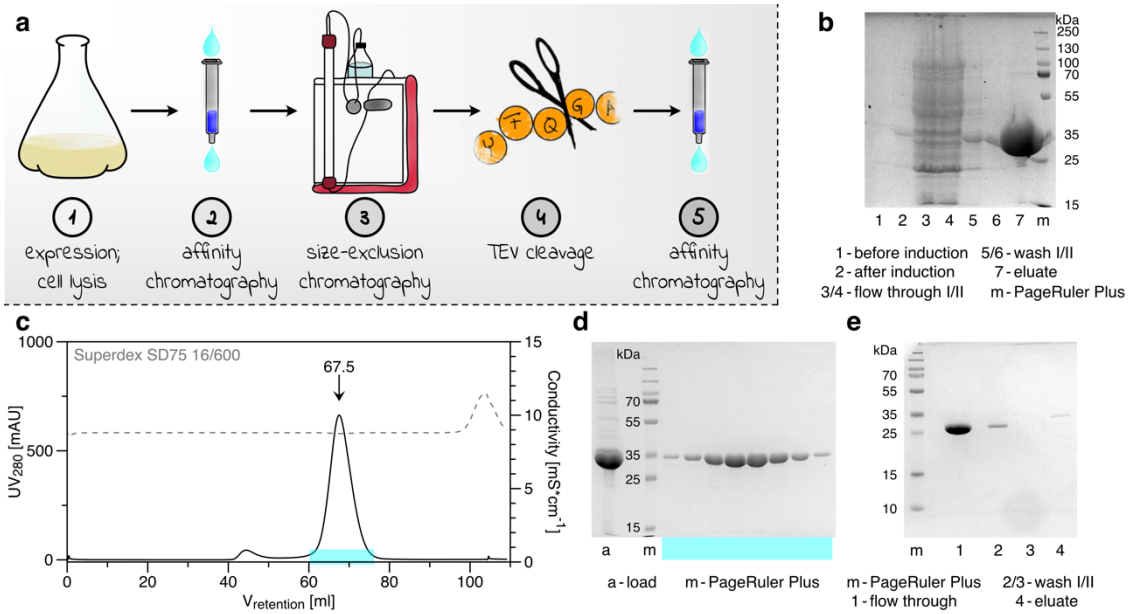


Figure 2-1: Purification of VcSiaP.
a) *Schematic overview of the expression and purification procedure of VcSiaP. Expression in M9-minimal media and full cell lysis (1) was followed by Ni^{2+} affinity chromatography (2) and size-exclusion chromatography (3). Subsequently, protein containing fractions were pooled, concentrated and supplemented with TEV protease to remove the His_6 tag (4). Another affinity chromatography was done to remove His-tagged TEV protease and non-cleaved protein yielding highly pure VcSiaP.* ***b)*** *Polyacrylamide gel analysis including samples from expression and from affinity chromatography.* ***c)*** *Chromatogram from size-exclusion chromatography indicating one separated peak in UV absorption at λ=280 nm at a retention volume of 67.5 ml. Additionally, the conductivity was monitored and shows a peak at V≈105 ml. Fractions from gel filtration as well as from affinity chromatography after TEV protease digestion were analyzed by SDS-PAGE. The resulting Polyacrylamide gels are shown in* ***d)*** *and* ***e)****, respectively.*

generation of HiSiaP specific nanobodies, the procedure was done in a similar way but using HiSiaP.

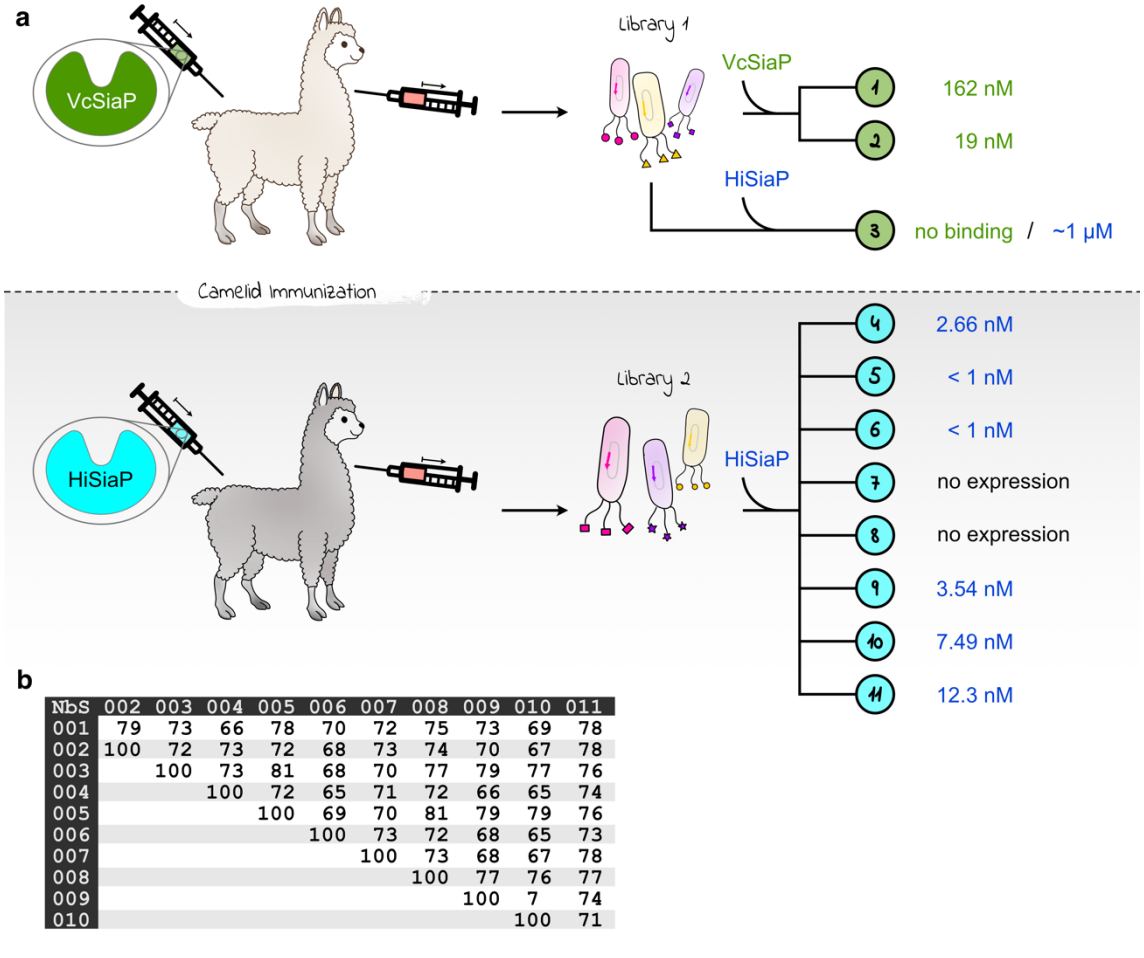
2.1.1 Preparations for nanobody generation and classification

2.1.1.1 Expression and purification of VcSiaP for Alpaca immunization

Immunization of camelids is the first step in the nanobody generation process. Therefore, recombinant VcSiaP was expressed in *E. coli* C43 cells using M9-minimal media to avoid the presence of sialic acid (Glaenzer et al., 2017). However, sialic acid cannot be excluded during the immunization as it is ubiquitous in vertebrates such as alpacas. After harvesting and cell lysis, the His_6 tagged protein was purified by Ni^{2+} affinity chromatography and size-exclusion chromatography (**Figure 2-1 a 1-3 & b, c, d**). Tag removal by TEV protease and subsequent affinity chromatography led to a highly pure protein solution (**Figure 2-1 a 4-5 & e)**. For one part of the final yield the buffer was exchanged to 20 mM HEPES, pH=7.4, 150 mM NaCl to be used for immunization of an alpaca (Vicugna pacos) by six subcutaneous injections over 12 weeks with 200 µg protein. The process was authorized by the Landesuntersuchungsamt Rheinland-Pfalz (23 177- 07/A 17-20-005 HP) and supervised by the Core Facility Nanobodies of the University of Bonn. The remaining volume of protein solution was concentrated to ~40 mg/ml, flash frozen in liquid nitrogen and stored at -80 °C.

2.1.1.2 Selection and classification of SiaP specific nanobodies

The two independent immunizations with VcSiaP and HiSiaP, respectively, gave rise to 11 VHH antibodies in total. While panning of the VcSiaP based phage library with biotinylated VcSiaP Q245C immobilized to magnetic streptavidin beads yielded two high affinity binding nanobodies (NbS001 and NbS002). A repanning of the same library using the closely related and biotinylated HiSiaP K254C as bait protein resulted in one HiSiaP binder (NbS003) (50.17 % sequence identity for wild-type VcSiaP and HiSiaP). However, this antibody did show only a micromolar affinity towards its antigen, but more interesting, no binding to VcSiaP could be observed at all even though the library originated from an immunization with VcSiaP (**Figure 2-2 a** top). To gain VHH antibodies with higher affinities the immunization process was repeated with HiSiaP as described earlier. The generated phage library resulting from the second immunization campaign using a different individual alpaca was panned against immobilized HiSiaP. This strategy yielded eight different nanobodies (NbS004-NbS011) that showed nanomolar affinities against their specific antigen (**Figure 2-2 a** bottom). As typical for nanobodies all eleven proteins share a highly conserved backbone and variable complementary determining regions (CDRs) (**Figure 2-2 c**). Their overall sequence identities vary from 64.9 % to 80.91 %. Interestingly, some of the VHHs share a high sequence identity even though they bind to different antigens. While others binding the same P-domain have a lower sequence identity (**Figure 2-2 b**). However, this might be a result of the high total similarity of the proteins that cause a low range of diversity. Additionally, even small alterations within the CDRs can lead to a completely different binding epitope.



b

NbS	002	003	004	005	006	007	008	009	010	011
001	79	73	66	78	70	72	75	73	69	78
002	100	72	73	72	68	73	74	70	67	78
003		100	73	81	68	70	77	79	77	76
004			100	72	65	71	72	66	65	74
005				100	69	70	81	79	79	76
006					100	73	72	68	65	73
007						100	73	68	67	78
008							100	77	76	77
009								100	7	74
010									100	71

c

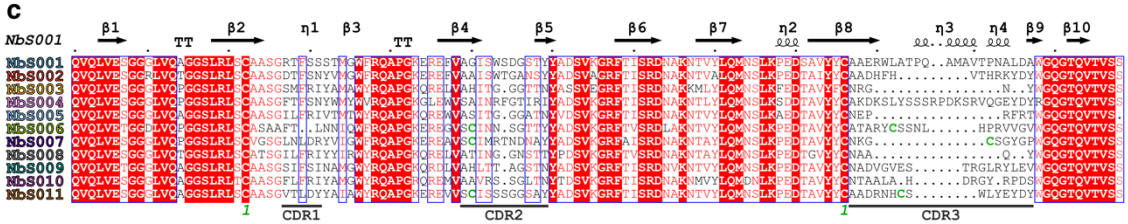


Figure 2-2: From immunization to hit identification and sequence analysis of VHH antibodies. ***a)*** *Sketch of the general procedure of two independent Camelid immunizations with the TRAP transporter P-domains from* V. cholerae *(i) and* H. influenzae *(ii) which allowed the generation of phage library 1 and library 2, respectively. The first library was used for panning with VcSiaP and subsequently also with HiSiaP as bait protein yielding in two high affinity VHH antibodies for VcSiaP and one Nanobody that was able to exclusively but weakly bind HiSiaP. Panning of phage library 2 with immobilized HiSiaP gave rise to eight individual antibodies of which six could be expressed and identified as selective and high affinity binding nanobodies for HiSiaP. Affinity constants determined by isothermal titration calorimetry are shown* ***b)*** *Sequence identity matrix of all 11 VHH antibodies showing the pairwise amino acid sequence identity of all possible combinations in percent.* ***c)*** *Primary amino acid sequence alignment of all 11 VHH antibodies, including secondary structure elements from NbS001 annotated above. Conserved regions are highlighted in red and CDRs are indicated below. The highly conserved disulfide bond is marked by a "1" below the sequences.*

2.1.2 Characterization of selected VHH-antibodies

All nanobodies were expressed and purified in the same way according to an established protocol (see **Section 7.3.3.2**). To enable formation of the disulfide bond between the highly conserved cysteine residues in β2 and β8, respectively (see (**Figure 2-2 c**), the nanobody expression vector encodes for a pelB leader sequence N-terminally of the protein which ensures export out of the reductive cytoplasm. Furthermore, export

into the periplasm allows osmotic lysis of the cells and subsequently, use the periplasmic extract for further purification steps (**Figure 2-3**). As the generation of VcSiaP specific nanobodies was done almost a year before that for HiSiaP, results of the biochemical characterization of VHHs targeting VcSiaP will be described first.

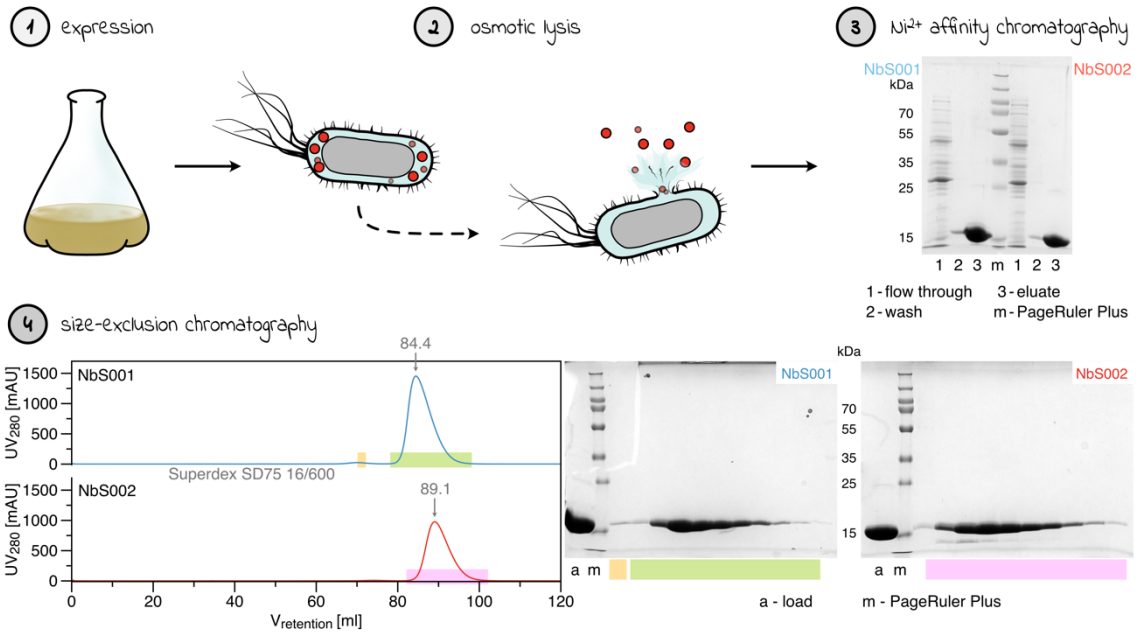


Figure 2-3: Overview on expression and purification of VHH antibodies.
(1) Expression of a VHH antibody construct including a C-terminal His_6 tag and an N-terminal pelB leader sequence that ensures export to the periplasm was carried out in Terrific-Broth *(TB) media. (2) Schematic representation of osmotic lysis upon destabilization of the outer cell membrane and drastically decreasing sucrose concentration. (3) SDS-PAGE analysis of samples from Ni^{2+} affinity chromatography of NbS001 and NbS002. (4) Size exclusion chromatography chromatograms and the respective SDS-PAGE analysis.*

2.1.2.1 Biochemical and biophysical characterization of VcSiaP specific nanobodies

Expression and purification of VcSiaP specific VHH antibodies resulted in a high amount of >10 mg of protein per liter of expression culture. However, in every single approach the yield for NbS002 was lower than that for NbS001 which is reflected in a less intense band in the SDS-PAGE analysis after Ni^{2+} affinity chromatography, as well as in the lower peak height of the UV absorption at λ=280 nm in gel filtration for NbS002 (**Figure 2-3**.3/4). Moreover, NbS001 eluted slightly earlier during SEC than NbS002 (84.4 ml compared to 89.1 ml). Retention volumes are in accordance with monomeric proteins as stated by the column manufacturer (MW_{NbS001}=15.8 kDa, MW_{NbS002}=15.4 kDa).

SEC-MALS experiments did not only show reasonable values for the MW of the individual proteins, but also for the heterodimeric complexes consisting of one VHH and VcSiaP. The analysis of a 1:1:1 molar mixture of VcSiaP:NbS001:NbS002 did show a further peak shift than the two individual 1:1 complexes and simultaneously the experimental molecular weight of 65 kDa was in good proximity to a possible heterotrimeric complex of all three proteins. It was also observed, that over the whole peak widths for all individual runs, the MALS data resulted in an almost horizontal line indicating a monodisperse sample (**Figure 2-4 a**). Quantitative binding analysis by isothermal titration calorimetry (ITC) confirmed the VHH–antigen interaction and revealed nanomolar affinities of NbS001 towards VcSiaP (162 nM) and an even roughly

10-fold higher affinity for NbS002 towards its target (13 nM). Interestingly, while NbS002 exhibited a large ΔH value thus, an enthalpic contribution, NbS001 barely showed a thermal response upon binding, indicating a highly entropy driven reaction mechanism which showed up in a large term for ΔS (**Figure 2-4 b, c).**

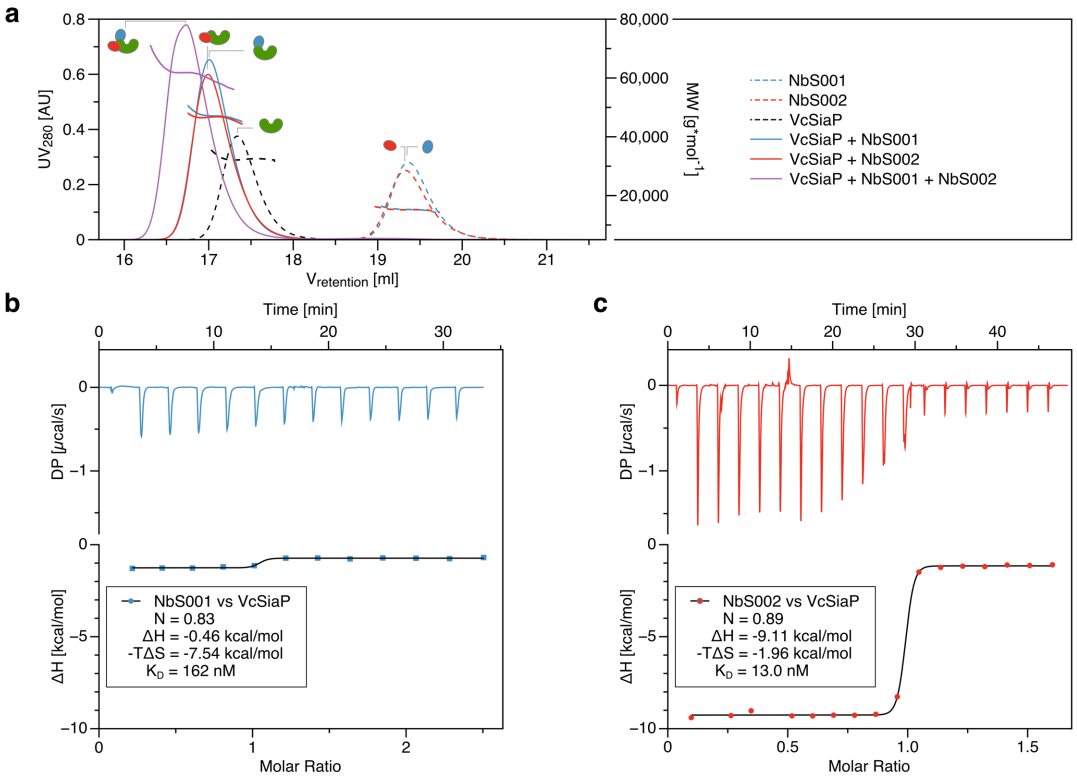


Figure 2-4: Analysis of 1:1 complex formation of VcSiaP and NbS001 and NbS002, respectively. **a)** *SEC-MALS analysis of VcSiaP with NbS001 and NbS002. The x-axis shows the retention volume, the left y-axis the UV absorption at λ=280 nm and the right y-axis the experimentally determined molecular weight. The graphs are color coded according to the legend on the right-hand side and tiny pictograms further illustrate which peak belongs to which protein (-complex).* ***b)*** *ITC thermogram for the titration of NbS001 versus VcSiaP (top) and the resulting binding curve (bottom) the obtained thermodynamic values and the affinity are shown within the graph.* ***c)*** *same as b) but for NbS002.*

2.1.2.2 Biochemical and biophysical characterization of HiSiaP specific nanobodies

To get HiSiaP binding nanobodies it was first tried to isolate them by panning of the already existing VcSiaP phage library (library 1) but using HiSiaP as bait protein. This idea came up because of the high sequence identity of the two TRAP transporter P-domains from different species. The much lower costs, as well as the much faster process were good arguments to test this. Unfortunately, only one hit could be identified but at least the resulting nanobody, NbS003, showed a slight peak shift in SEC MALS experiments when mixed to HiSiaP but no changes after incubation with VcSiaP (**Figure 2-5 a**). Dynamic light scattering (DLS) revealed significantly larger particles for the sample containing HiSiaP and NbS003 at the same time compared to all single protein samples as well as to a 1:1 molar mixture of VcSiaP and the VHH (**Figure 2-5 b**). Results from nanoDSF experiments did also support these findings, as addition of NbS003 to HiSiaP lead to a significant thermal stabilization which was not observed upon addition of the VHH to VcSiaP (**Figure 2-5 c**). Quantitative investigation of the interaction by ITC showed an affinity of ~1 µM for HiSiaP (**Figure 2-5 d**). This rather low affinity fitted

the observation of a small peak shift in SEC as diffusion of the weak complex during gel filtration occurs in a higher degree than for a more stable one.

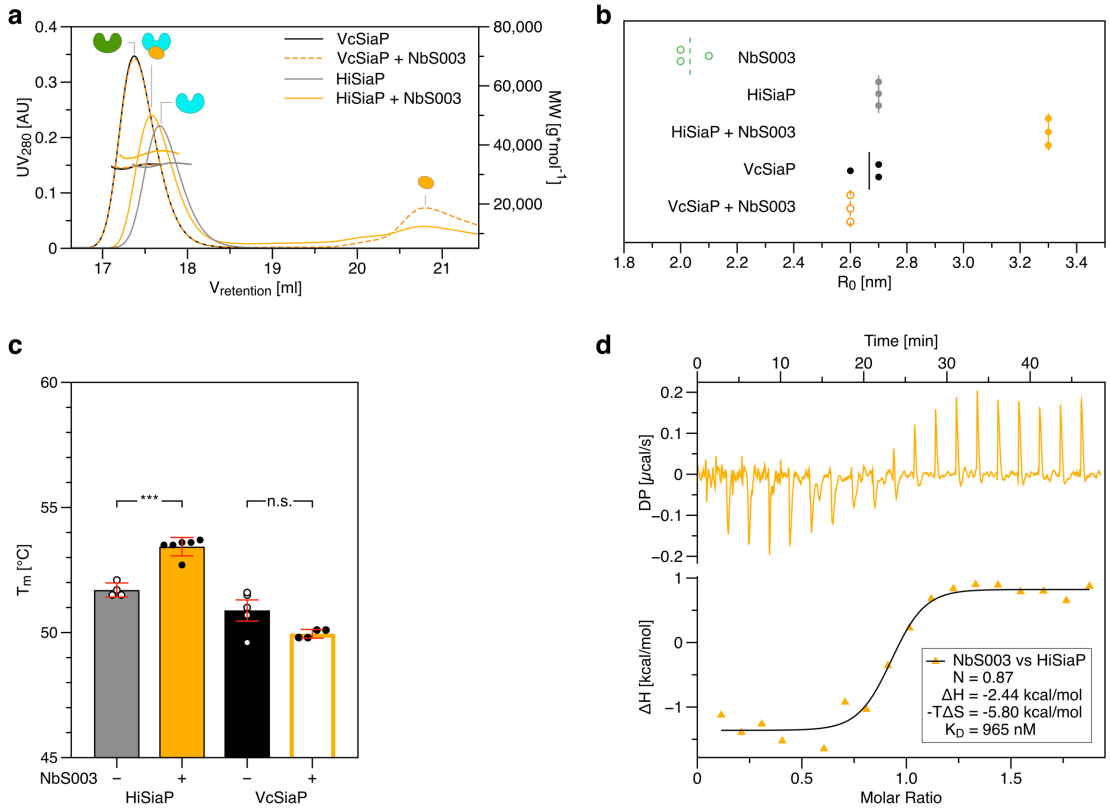


Figure 2-5: Biochemical characterization of HiSiaP binding VHH antibody NbS003.
a) *SEC-MALS analysis of NbS004-NbS006 (top) and NbS009-NbS011 (bottom) and their individual 1:1 complexes with HiSiaP. The x-axis shows the retention volume, the left y-axis the UV_{280} absorption and the right y-axis the molecular weight. The individual coloring of the curves is explained on the right-hand side.* ***b)*** *Hydrodynamic radii of the samples investigated in a) determined by DLS. The size in nm is shown on the x-axis, each dot represents one single measurement and the vertical lines the resulting mean value.* ***c)*** *Bar diagram of denaturation temperateures derived from nanoDSF experiments of HiSiaP (left) and VcSiaP (right), each without (-) and with (+) NbS003. The y-axis depicts the temperature in °C and results from two-sided t-tests are annotated above the bars ($p=4.38*10^{-5}$ for HiSiaP, $p=0.0599$ for VcSiaP). The bar length describes the averaged temperature value from multiple individual experiments and the resulting standard deviation is given as error bars.* ***d)*** *ITC thermogram for the titration of NbS003 versus HiSiaP (top) and the resulting binding curve (bottom) the obtained thermodynamic values and the affinity are shown within the graph.*

To yield more nanobodies against HiSiaP, that hopefully also show higher binding affinities, a second immunization campaign was done using HiSiaP as antigen. The generation and selection as well as the expression and purification procedure were done in the same way as described for VcSiaP. However, the nanobody genes were provided within another expression vector with chloramphenicol resistance and a L(+) -Arabinose inducible promoter but yet contained an N-terminal pelB signal sequence and C-terminal His_6 tag so that the expression and purification protocol did only contain minor changes (see **Section 7.3.3.2**). Two out of eight resulting VHHs (NbS004 - NbS011) could not be expressed and purified, namely NbS007 and NbS008. All others behaved similar as the previously described VHHs, but with higher variations within the overall yield (0.75 mg/l - 37 mg/l). SEC-MALS experiments showed significant peak shifts for all individual 1:1 mixtures of the six VHHs with HiSiaP compared to the P-domain peak. Also, the molecular weight determination fitted quiet well to the calculated values for the

complexes (**Figure 2-6 a**). By isothermal titration calorimetry, all 1:1 complexes were characterized quantitatively, but only the data for NbS004 and NbS011 are depicted representatively (**Figure 2-6 c, d**). Details of all binding parameters are listed below.

Table 2-1: Binding parameters of the 1:1 complex formation of SiaPs and their specific VHHs.

Measurements which are denoted by a '' were only performed once, the given error was determined by the data analysis software. All other measurements were performed at least twice and the given error values correspond to the standard deviation.*

Experiment	**# sites N**	**Error N**	**K_D** [nM]	**Error K_D**	**ΔH** [kcal/mol]	**Error ΔH**	**-TΔS** [kcal/mol*K]	**Error -TΔS**
VcSiaP vs NbS001	0.830	$4.3e^{-2}$	162	24.8	-0.46	$7.0e^{-2}$	-7.54	$4.5e^{-1}$
VcSiaP vs NbS002	0.890	$3.9e^{-3}$	13.0	2.31	-9.11	$1.6e^{-1}$	-1.96	$1.4e^{-1}$
HiSiaP vs NbS003	0.866	$9.9e^{-3}$	965	305	-3.15	0.11	-5.80	0.38
HiSiaP vs NbS004*	0.957	$1.5e^{-3}$	2.66	1.24	-9.34	$4.4e^{-2}$	-2.36	
HiSiaP vs NbS005*	0.745	$1.1e^{-3}$	< 1	3.35	-14.5	$6.0e^{-2}$	1.21	
HiSiaP vs NbS006*	0.842	$5.6e^{-4}$	< 1	2.38	-21.8	$5.3e^{-2}$	5.41	
HiSiaP vs NbS009*	0.789	$1.7e^{-3}$	3.54	1.76	-14.5	$9.2e^{-2}$	3.01	
HiSiaP vs NbS010*	0.961	$1.0e^{-3}$	7.49	5.82	-10.9	$4.9e^{-2}$	-0.175	
HiSiaP vs NbS011*	0.903	$1.8e^{-3}$	12.3	5.80	-14.4	$9.2e^{-2}$	3.61	

All of the characterized VHHs bind with nanomolar, or lower affinities and enthalpy driven complex formation could be observed. To test the maximal possible amount of simultaneously bound VHHs a master mix consisting of all six expressible nanobodies and HiSiaP (1:1:1:1:1:1:1 molar ratio) was run for SEC-MALS analysis. This experiment showed a significantly earlier eluting peak compared to the 1:1 complexes and a MW of 75 kDa suggesting three non-overlapping individual epitopes and a heterotetrameric complex (**Figure 2-6 b**). By running all possible heterotrimeric combinations, we found out, that NbS004-NbS006 and NbS010 could not bind at once but each of them in combination with NbS009 and NbS011 formed stable complexes indicated by a peak shift and increased molecular weight. NbS009 and NbS011 also showed this behavior when simultaneously run with HiSiaP. These observations were in line with a maximal tetrameric complex size suggesting individual binding epitopes for NbS009, NbS011 and the group of NbS004-NbS006 and NbS010.

SEC-MALS analysis of VcSiaP specific nanobodies already suggested the presence of a heterotrimeric complex as explained (see **Figure 2-4 a**). To further demonstrate the existence of higher order complexes DLS, nano differential scanning fluorometry (nanoDSF) and ITC was used. The hydrodynamic radii of the antibody fragments alone could be determined to R(NbS001)=2.0 nm and R(NbS002)=2.1 nm and that of VcSiaP to R(VcSiaP)=2.5 nm. Heterodimeric complex formation resulted in significantly larger

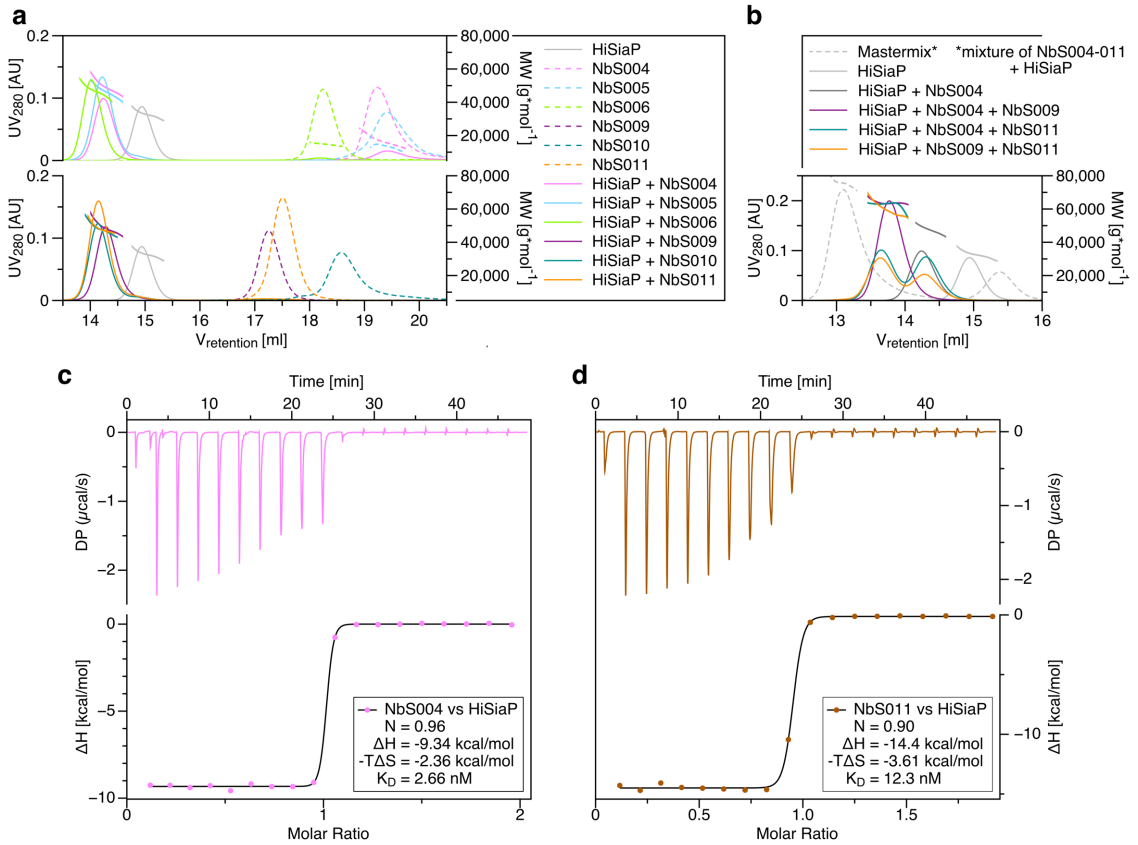


Figure 2-6: Biochemical analysis of HiSiaP–VHH complexes and determination of epitopes. ***a)*** *SEC-MALS analysis of NbS004-NbS006 (top) and NbS009-NbS011 (bottom) and their individual 1:1 complexes with HiSiaP. The x-axis shows the retention volume, the left y-axis the UV_{280} absorption and the right y-axis the molecular weight. The individual coloring of the curves is explained on the right-hand side.* ***b)*** *Same as a) but for investigation of 1:1:1 complexes and using a mastemix to specify the maximal complex size.* ***c)*** *ITC thermogram for the titration of NbS004 versus HiSiaP (top) and the resulting binding curve (bottom) the obtained thermodynamic values and the affinity are shown within the graph.* ***d)*** *same as c) but for NbS011.*

particles (R(VcSiaP/NbS001)=3.2 nm, R(VcSiaP/NbS002)=3.1 nm) and a combination of all three proteins at 1:1:1 molar ratio exhibited a further increase radius of 3.8 nm (**Figure 2-7 a**). For the 1:1 complexes, a thermal stabilization of ΔT≈6 °C was observed in nanoDSF experiments compared to VcSiaP only. The stabilizing effect increased to ΔT≈16 °C when adding both nanobodies at once to VcSiaP (**Figure 2-7 b**). Calorimetric isothermal titration of NbS002 to VcSiaP preincubated with NbS001 finally proved that all three proteins form a thermodynamically stable 1:1:1 complex (**Figure 2-7 c**).

In combination, the described experiments did not only exhibit the presence of interactions but further characterized them in terms of size, stability and revealed thermodynamic binding data. Furthermore, we could show that HiSiaP can be targeted by three - and VcSiaP by two VHHs at once (**Figure 2-7 d**).

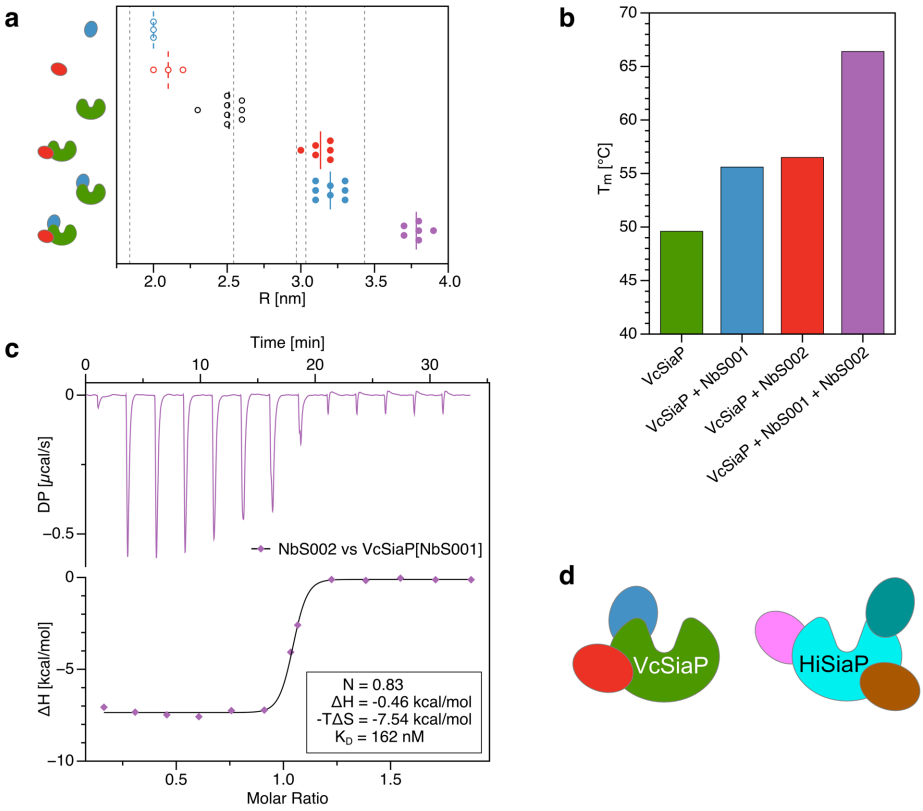


Figure 2-7: Epitope binning experiments for VcSiaP nanobodies.
a) *Hydrodynamic radii of VcSiaP/VHH complexes determined by DLS. The size in nm is shown on the x-axis, each dot represents one single measurement and the vertical lines the resulting mean value. Small pictograms on the left-hand side visualize the sample contents (blue: NbS001, red: NbS002, green: VcSiaP).* ***b)*** *Thermal shift assays to investigate the effects of VHH-VcSiaP complex formation on the thermal denaturation of the proteins. The determined denaturation temperatures of different protein (-complexes) are depicted as bars.* ***c)*** *ITC measurement of NbS002 against a preincubated VcSiaP-NbS001 complex. The thermogram of the titration and the resulting binding curve are depicted. The resulting data are shown in the lower right corner.* ***d)*** *Pictograms of VcSiaP and HiSiaP, respectively, with the maximum number of VHH antibodies that ca be bound simultaneously.*

2.1.3 Structural analysis of VcSiaP targeting VHHs

To further understand the different properties of the VHHs and their antigen assemblies, crystallization experiments were performed. After 2 days of incubation at 20 °C, single crystals of both different VcSiaP/VHH heterodimers were obtained and yielded diffraction data. The structures could be solved by molecular replacement (McCoy et al., 2007) at (2.64 Å VcSiaP/NbS001) and 2.05 Å (VcSiaP/NbS002) resolution using VcSiaP (PDB ID: 4mag (Setty et al., 2014)) and a BtuF specific VHH (PDB ID: 5ovw_H (Mireku et al., 2017)) as search models. All VcSiaP residues (1-299) could be detected within the electron density, whereas the C-terminal HA-His_6 tag was not resolved. Both nanobodies were identified to bind to different concave surface regions located at the N-terminal lobe of the P-domain (**Figure 2-8 a, b**).

The interaction interface of the VcSiaP/NbS001 structure was analyzed using the PDBePISA online tool (Krissinel & Henrick, 2007) and amounts to an area of 556.5 $Å^2$. The most striking feature of the interface is the side chain W101 of NbS001, which lies flat on the VcSiaP surface and forms hydrophobic interactions with residues A24, L37, A38, and L39 of VcSiaP. In addition, NbS001 interacts with VcSiaP by polar interactions involving residues D116 (K21 (2.9 Å)), Q99 (Y17 (2.7 Å), K21 (2.4 Å)), R100 (E28 (2.6,

3.3 Å)), W101 (D25 (2.9 Å)) T111 (V12 (2.9 Å) and T104 (Q44 (2.4 Å)) (see **Figure 2-8 d**). Several hydrophobic interactions are also involved in the binding (**Figure 2-8 e**).

The NbS002-VcSiaP structure has a slightly larger (+15%) interaction area of 639.3 Å^2. The interface is centered around F101 of the nanobody, which penetrates into a hydrophobic cluster at the surface of VcSiaP, formed by residues R49, Q53, W73, F112 and W114 (**Figure 2-8 d**). A rotation of W73 forms a distinct pocket in the surface of VcSiaP. Also notable is a cation-π interaction between VcSiaP R49 and NbS002 W53 with a distance of 3.6 Å. Further polar and ionic interactions were identified and are summarized in **Figure 2-8 f**. Interestingly, the NbS002/VcSiaP interface has fewer electrostatic and polar interactions than found for NbS001 despite the higher affinity and drastically larger binding enthalpy.

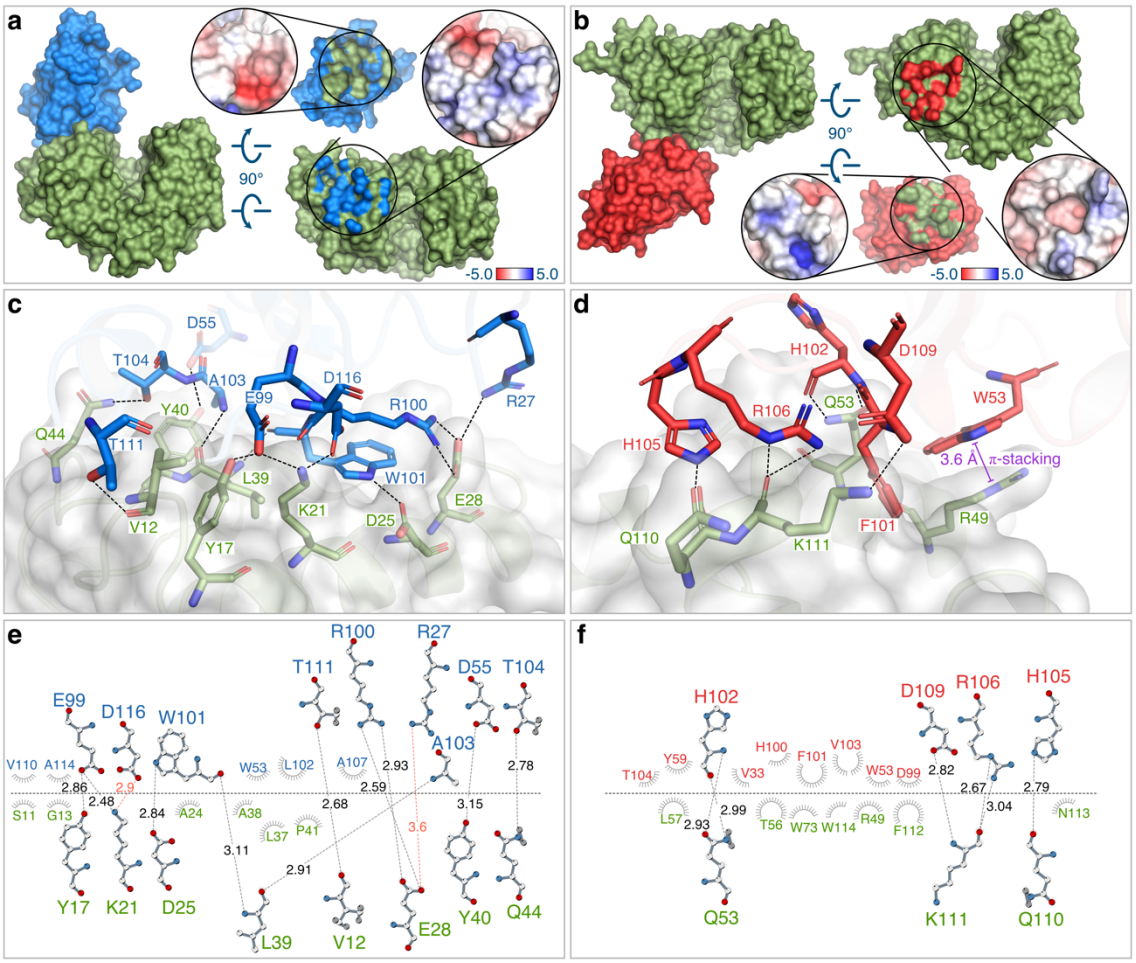


Figure 2-8: Structural analysis of VcSiaP/VHH 1:1 complexes.
a) *Structural overview of the NbS001/VcSiaP complex. The proteins are shown as surface models in green (VcSiaP) and blue (NbS001), respectively. On the right-hand side, both proteins are rotated by 90° against each other to visualize the interaction surface which is highlighted in the color of the respective binding partner. Magnifications of the two surface regions are shown as patches colored by their electrostatic surface potential.* ***b)*** *same as in a) but for the complex of VcSiaP (green) and NbS002 (red).* ***c)*** *Magnification of the interaction area. The surface of VcSiaP is shown in translucent gray. Both proteins are visualized as cartoon representation, and interacting side chains as stick model. All residues are labeled and bonds are indicated by dashed lines. The color scheme was adopted from a).* ***d)*** *Same as c) but for the VcSiaP/NbS002 complex. The colors are the same as in b).* ***e)*** *Detailed visualization of all direct interactions of the complex using LigPlot+ (Laskowski & Swindells, 2011) and the same colors as in c). Electrostatic interactions (black) and salt bridges (red) are indicated by dashed lines and the respective bond lengths are given in Ångström (Å). Residues that undergo hydrophobic interactions are labeled in one letter code and their orientation is depicted.* ***f)*** *LigPlot+ figure of the VcSiaP/NbS002 complex analogous to e).*

2.1.3.1 Investigating the heterotrimeric complex of VcSiaP and its VHHs

To structurally verify the presence of a heterotrimeric complex of VcSiaP and both nanobodies, crystallization with a 1:1:1 mixture of the three proteins was performed. After optimizing the expression vector to yield a VHH construct with a TEV protease cleavable N-terminal His_6 tag, crystallization was successful. The resulting crystal structure was solved at 2.28 Å resolution by molecular replacement (McCoy et al., 2007) using the individual structures as search models. Conspicuously, the structure revealed an interesting crystal packing. Upon homodimer formation of the heterotrimer within the asymmetric unit, NbS001 (NbS001-2) bound to one P-domain (VcSiaP-2), was squeezed in between NbS001 (NbS001-1) and NbS002 (NbS002-1) bound to the other SBP (VcSiaP-1) (**Figure 2-9**). This phenomenon resulted in a slightly shifted orientation of NbS001 on the VcSiaP surface compared to the heterodimeric VcSiaP/NbS001 structure (**Figure 2-9 c**). Whereas the overall fold of NbS001 remains mostly unaffected by the crystal packing and thus, the VHH structure in the heterotrimeric VcSiaP/NbS001/NbS002 complex aligns with a r.m.s.d of 0.369 (over 116 Cα atoms) onto the VHH from the heterodimer VcSiaP/NbS001 complex. VcSiaP/NbS001/NbS002 complex 1 showed intermolecular interactions to NbS001-2 of the second complexNbS001 (**Figure 2-9 b**). NbS002-1 and NbS001-2 share an H-bond in between A56 and R19 and several hydrophobic interactions involving NbS002-1 residues N57, S58, Y59, Y60, A65 and G66 and NbS001-2 residues V5, E6, S7, G8, S17 and L18. NbS001-1 and NbS001-2 interact via H-bond in between T58 and T58, as well as between K65 and Y60, D62. Additional hydrophobic interactions were observed between NbS001-1 residues D55, G56, S57, Y60 and D62 and NbS001-2 residues D55, G56, S57 and K65. Interestingly, the two NbS001 proteins formed complementary interactions, since both shared an identical interaction surface. As the only possible connection between NbS001-2 and VcSiaP-1 a 3.1 Å distant H-bond was identified. However, if the formation of a dimer of the trimeric subcomplex is just a crystal packing artefact, or does also exist, at least partially, in solution could not clearly been identified.

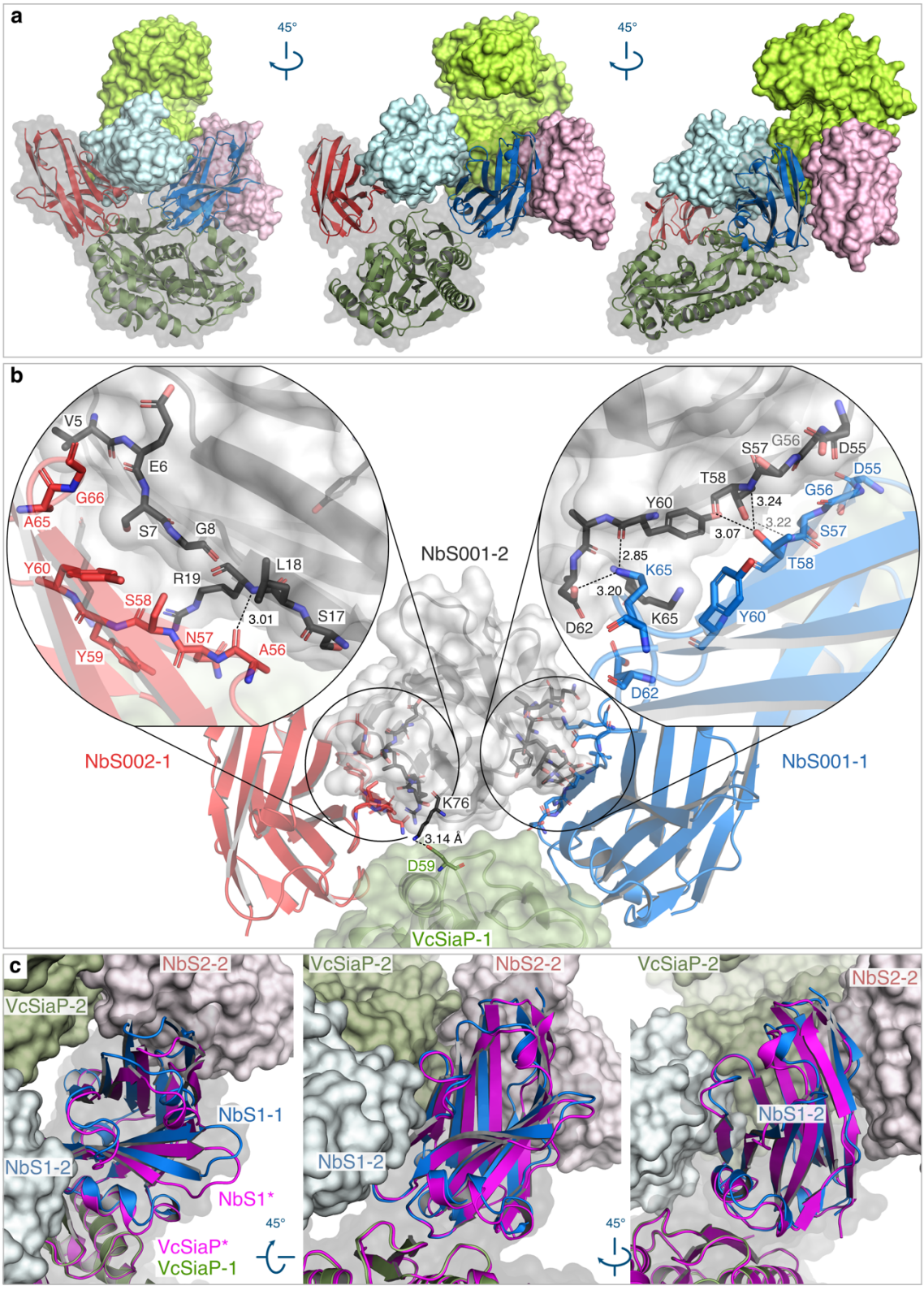


Figure 2-9: Structural analysis of the heterotrimeric VcSiaP/NbS001/NbS002 complex.
a) *Overview of the crystallographic homodimeric complex of the VcSiaP/NbS001/NbS002 heterotrimer depicted from different view angles rotated by 45° as indicated.* ***b)*** *Visualization of the interactions between two different subcomplexes. Interacting residues are shown as sticks and are labeled in a magnificated image patch.* ***c)*** *Comparison of the tertiary structure of NbS001 (blue) from the homodimeric complex structure and the individual VHH/SBP structure (magenta and annotated by a "*") elucidated before. The general color coding of the whole figure was kept consistent: VcSiaP (green), NbS001 (blue), NbS002 (red) with differentiating subcomplex 1 (dark shades) and subcomplex 2 (lighter shades), exceptions are indicated.*

2.2 VHHs inhibit Neu5Ac binding by SiaP

For mechanistical studies of the substrate binding protein, SiaP, identification of a nanobody that affect the substrate binding would be of high interest. Despite the knowledge that VcSiaP targeting VHHs bind to epitopes distant from the substrate binding cleft, the effect of sialic acid on the binding behavior of the nanobodies, and vice versa, was investigated. The characteristic conformational rearrangement of the TRAP transporter P-domain (Berntsson et al., 2010; Glaenzer et al., 2017; Peter et al., 2021) could affect the surface morphology and thus, result in alterations within the interaction areas. Two different sets of ITC experiments were performed to analyze this. In the first set, VcSiaP was loaded with sialic acid VcSiaP[Neu5Ac] and then titrated with either NbS001 or NbS002 (**Figure 2-10 a**). For this type of experiment, one could imagine three different results: either the VHH binds (i) or does not bind to $VcSiaP_{holo}$ (ii), or the competitive one, where binding of the VHH leads to a release of sialic acid from the binding site (iii). In the second set of experiments, the SBP was preincubated with one of the VHHs and then titrated with Neu5Ac (**Figure 2-10 b**). Similar scenarios as described for the first experiment set would be possible (see **Figure 2-10 b** (i-iii)). All experiments were also done utilizing HiSiaP and its specific set of VHHs, NbS004-NbS011. In the most cases no effect of sialic acid on the nanobody binding was observed (scenario "a) i" or "b) i" shown in **Figure 2-10**) except for NbS002 and NbS011. However, since only VcSiaP/VHH complexes were studied structurally, only results for VcSiaP specific VHHs will be shown in detail in the following.

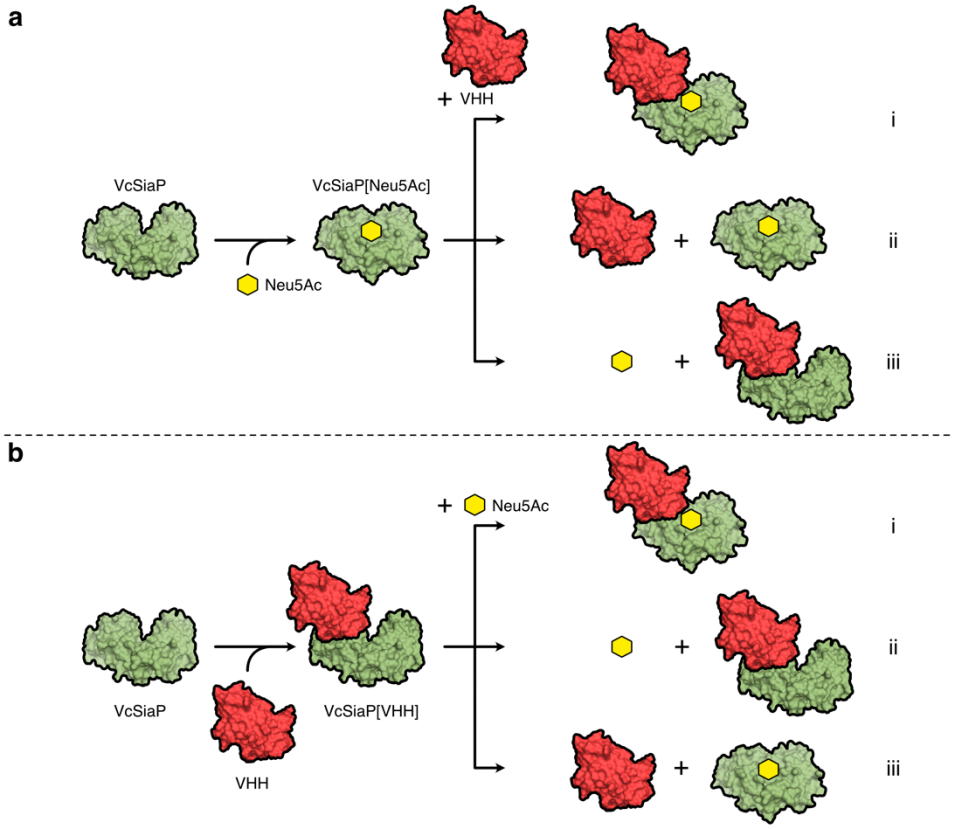


Figure 2-10: Visualization of ITC experiments to analyze VHH binding in presence of sialic acid. ***a)*** *VcSiaP (green) was loaded with Neu5Ac (yellow), the resulting complex, VcSiaP[Neu5Ac] was then titrated with a VHH. Three different imaginable scenarios are depicted on the right-hand side (i-iii)* ***b)*** *Same as in a) but to investigate the effect of preincubation of SiaP with one of the VHHs onto a subsequent titration with sialic acid. Again three different scenarios are diplayed (i-iii).*

2.2.1 Allosteric inhibition of VcSiaP substrate binding by NbS002

While preincubation of VcSiaP with NbS001 did not affect the substrate binding of the SBP (**Figure 2-11 a**), the VcSiaP/NbS002 complex was unable to bind sialic acid (**Figure 2-11 b**). To test if the inhibitory effect was exclusively caused by the antibody fragment, another experiment was done where VcSiaP was mixed with NbS002 in a 1:0.5 molar ratio previously to the titration experiment. Strikingly, this resulted in the same inhibition, but the half of all available SBP (which was not complexed with VHH) showed binding to Neu5Ac reflected by an inflection point of the binding isotherm at a molar ratio of ~0.5 (**Figure 2-11 b** brown dashed line).

The observation, that NbS001 does not affect substrate binding could also be validated by structural analysis. Crystallization trials of the VcSiaP[Neu5Ac]/NbS001 complex were successful and the structure could be solved at 1.96 Å resolution by molecular replacement (McCoy et al., 2007) using sialic acid bound VcSiaP (PDB ID: 7a5q (Peter et al., 2021)) and the previously shown structure of NbS001 as search models. Comparison of that structure to the structures of $VcSiaP_{holo}$ and of the previously shown VcSiaP/NbS001 complex did not show significant changes in neither the coordination of sialic acid (**Figure 9-1 a**), nor the interactions between the SBP and its VHH (**Figure 9-1 b**). A detailed look into the conformational rearrangements by MtsslWizard (Hagelueken et al., 2012) did also reveal no changes in the C-α movement (**Figure 9-1 c**).

To test whether the second VcSiaP specific VHH can only bind to the P-domain by

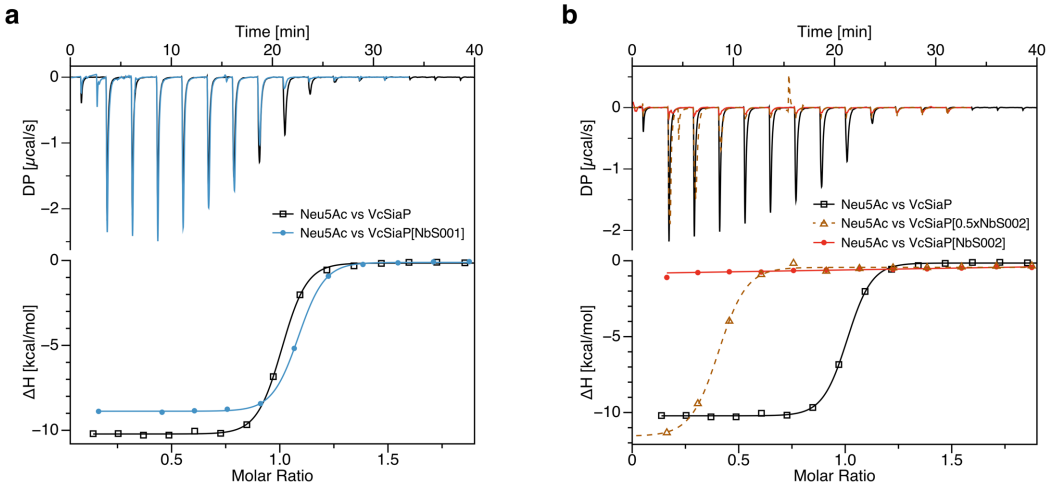


Figure 2-11: Substrate binding of VcSiaP is inhibited by preincubation with NbS002.
a) *ITC result from the titration of Neu5Ac to VcSiaP[NbS001]. The upper part shows the thermogram with the differential power on the ordinate and the experiment time on the abscissa. The lower part contains the integrated values fitted by the binding isotherm with the enthalpy (ΔH) on the y-axis and the molar ratio on the x-axis.* ***b)*** *Same as in a) but for the titration of Neu5Ac to VcSiaP[NbS002]. An additional experiment with preincubation of VcSiaP with 0.5x VHH was included. The different graphs were observed by individual measurements. Information on which graph refers to which experiment can be found in the corresponding legend. The same binding curve of sialic acid to VcSiaP was included in both graphs as reference.*

inducing a conformational change to the open state, different experiments with a double cysteine mutant of VcSiaP were conducted. This mutant can be locked in the substrate bound state upon formation of a disulfide bond bridging the N-terminal and C-terminal lobe at the upper rim of the substrate binding cleft. A proof of function of this method utilizing HiSiaP was recently published by our group (Peter et al., 2024). First, analytical size exclusion chromatography was conducted to quickly get an overview on the conformer specificity of VHH binding. As expected, NbS001 binding was not dependent

on the conformational state of SiaP and a peak shift for VcSiaP and VcSiaP S14C/T192C (VcSiaP CC) could be observed. For NbS002, addition of VcSiaP led to a peak shift while preincubation with VcSiaP CC gave rise to a double peak. The retention volume of the earlier eluting peak was comparable to that of the heterodimeric complex and the second peak corresponded to free VcSiaP. For a sample containing both VHHs and VcSiaP CC a peak referring to the 1:1 complex with a shoulder towards lower retention volumes was observed (**Figure 2-12 a**). A consecutive ITC experiment verified in the first titration (Neu5Ac vs VcSiaP and VcSiaP S14C/T192C (VcSiaP CC), respectively) that the double cysteine mutant could not bind further sialic acid after preincubation with the sugar and subsequent oxidation with H_2O_2 to force disulfide bond formation (VcSiaP CC*). Wild type VcSiaP in turn behaved as expected and bound Neu5Ac. Subsequently, 36.4 µl from the sample cell were taken out (this equals the volume that have been added during the titration) and the remaining volume was used for a second titration experiment with NbS002. For wild-type VcSiaP, a binding curve was observed that could be fitted by a competitive reaction model (**Figure 2-12 b**, red). VcSiaP CC* in turn, did not reveal significant alterations in the peak heights resulting from ligand injections and thus,

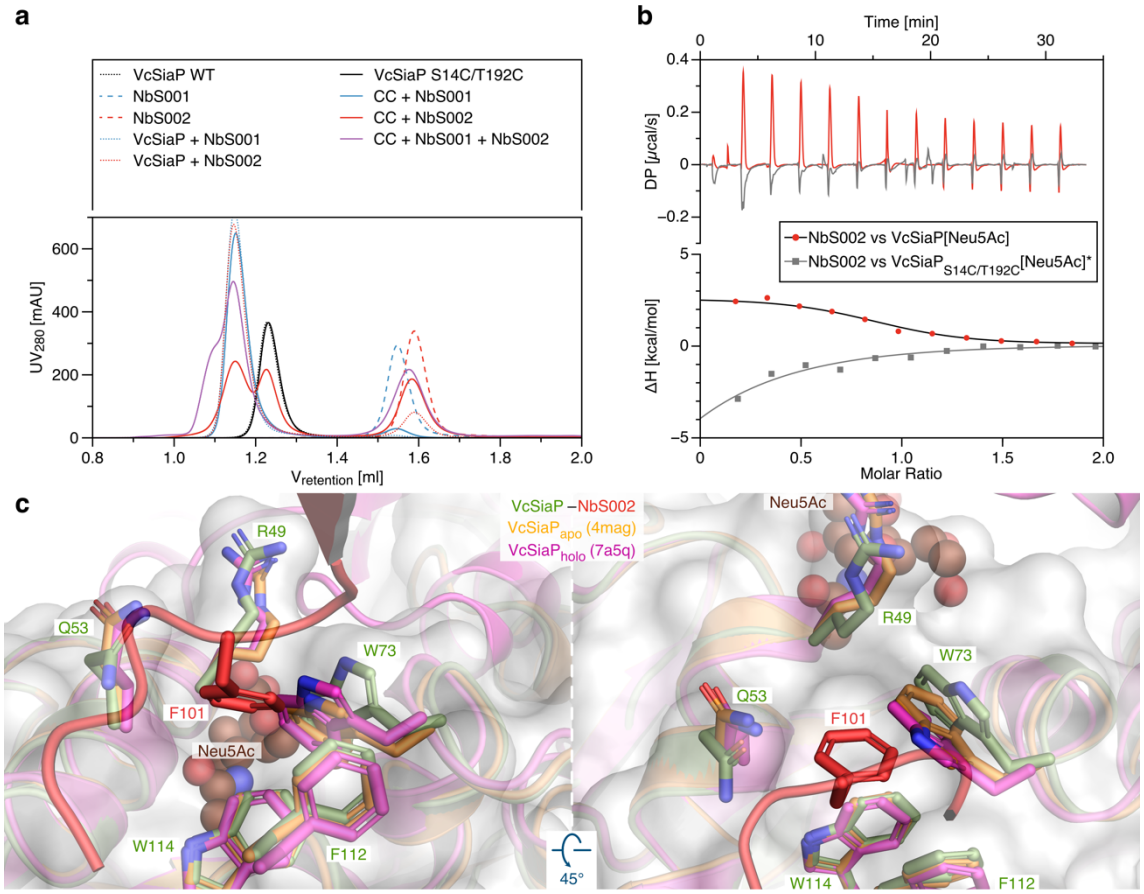


Figure 2-12: Allosteric inhibition by preventing conformational changes.
a) *Analytical SEC data showing that the formation of the NbS002/VcSiaP complex is dependent on the conformational state of the P-domain. UV_{280} absorbance is shown on the y-axis and the retention volume on the abscissa. Results from multiple seperate gel filtration runs are included, the color coding is explained in the legend above the graph.* ***b)*** *ITC thermograms of two independent experiments (top) and the corresponding binding curves (bottom). One experiment utilized wild type VcSiaP bound to sialic acid and the other one VcSiaP S14C/T192C trapped in its closed conformation.* ***c)*** *Structural overview of the surface cavity of VcSiaP intruded by F101 of NbS002. The structures of VcSiaP in its substrate free (orange, 4mag) and -bound (magenta, 7a5q) state were aligned onto the VHH bound structure (green, red). Interesting side chains are shown as sticks and labeled. The left-hand view and that on the right side are rotated by 45° as indicated.*

showed no binding to NbS002 (**Figure 2-12 b**, gray). Taken together, these results demonstrated that for a competitive binding of NbS002 towards VcSiaP[Neu5Ac] the opening of the P-domain needs to be induced. We wondered what mechanism causes the allosteric inhibition and forces the SBP to adopt its open conformation. To study this, the structures of the P-domain in its different conformations were aligned to that of VcSiaP/NbS002. **Figure 2-12 c** shows the surface cavity of VcSiaP intruded by F101 of the VHH. The comparison of the surrounding residues protruded tryptophane at position 73 (W73) that undergoes a minor but yet distinct conformational rearrangement during the closing movement. This motion might lead to a tightening of the cavity that in turn makes it unsuitable for F101 of the nanobody. The crucial role of F101 for the antigen binding could be demonstrated by gel filtration of an F to A mutant of the nanobody together with VcSiaP, both proteins eluted individually. However, a strongly weakened interaction disqualified this mutant for further mechanistical analyses.

2.2.2 Tryptophane 73 of VcSiaP plays a key role in the allosteric inhibition by NbS002

To check, whether the altered conformation of W73 is indeed key to the allosteric inhibition, we mutated the tryptophane to an alanine, with the intention to create more space and allow the structural transition in the presence of NbS002. The protein behaved as the wild type during the whole expression and purification process and was still recognized by the VHH albeit with a slightly reduced affinity (~40 nM for mutant vs 13 nM for WT). And indeed, an initial thermal shift assay demonstrated, that the sialic acid induced thermal stabilization of VcSiaP could be recovered using the mutant bound to NbS002, while wild type VcSiaP could not be stabilized by sialic acid in presence of the nanobody. The stabilizing effect of Neu5Ac was not affected by NbS001 (**Figure 2-13 a**). Furthermore, binding analysis by ITC showed a clear binding curve for the titration of Neu5Ac to VcSiaP[NbS002] (**Figure 2-13 a**). However, a weaker binding affinity compared to the binding of sialic acid to VcSiaP WT was observed.

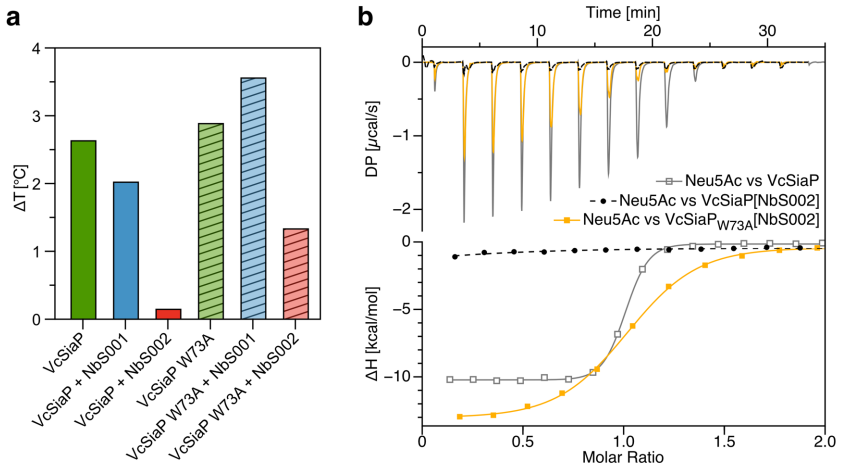


Figure 2-13: VcSiaP mutant W73A is able to bind Neu5Ac in presence of NbS002.
a) *Bar diagram of the thermal stabilization of VcSiaP and its VHH complexes by sialic acid. The temperature difference between $VcSiaP_{holo}$ and $VcSiaP_{apo}$ is denoted on the y-axis and the sample composition on the x-axis.* ***b)*** *ITC result from the titration of Neu5Ac to $VcSiaP_{W73A}$[NbS002] (orange). The upper part shows the thermogram with the differential power on the ordinate and the experiment time on the abscissa. The lower part contains the integrated values fitted by the binding isotherm with the enthalpy (ΔH) on the y-axis and the molar ratio on the x-axis. Titrations of Neu5Ac vs VcSiaP (gray) and $VcSiaP_{WT}$[NbS002] (black), respectively, were included as references.*

To further characterize the competitive binding effect that has been observed when VcSiaP[Neu5Ac] was titrated with NbS002, a sequential titration experiment was designed. Since previous data did not explicitly show if sialic acid is repelled from the substrate binding site upon VHH binding we aimed to sense free sialic acid in solution after the experiment. Therefore, a first titration added sialic acid to the P-domain **(Figure 2-14 i**), followed by calorimetrically controlled injections of NbS002 to the SBP/substrate complex (**Figure 2-14 ii**). In the last step, titration of HiSiaP to the solution would lead to a binding of sialic acid. The total amount of sialic acid that is available for HiSiaP should yield information if VHH binding released the sugar from VcSiaP (**Figure 2-14 iii**). In practice. this procedure was performed using wild type VcSiaP and VcSiaP W73A, respectively, in independent experiments. During the first part, an almost identical binding behavior of wild type and mutant towards Neu5Ac was observed (**Figure 2-15 a, d**). Addition of VHH to the complex of SBP and its natural substrate led to the characteristic competitive effect in the case of wild type protein (**Figure 2-15 b**). For the mutant, this step resulted in a binding curve comparable to that of NbS002 to substrate free VcSiaP (**Figure 2-15 e**). During the subsequent titration of HiSiaP to the solution, binding of sialic acid to that closely related P-domain was observed for both the experimental setup using VcSiaP WT and VcSiaP W73A. Stinkingly, the inflection point of the binding curve was significantly shifted in terms of its molar ratio. While HiSiaP was able to bind Neu5Ac in a molar ratio of 1.7-2.0 for the wild type approach (**Figure 2-15 c**), only about one half of the sialic acid amount could be detected when VcSiaP W73A was utilized (**Figure 2-15 f**). Meaning that upon addition of NbS002, sialic acid remains bound for the mutant while it was released from the binding pocket of the wild type protein.

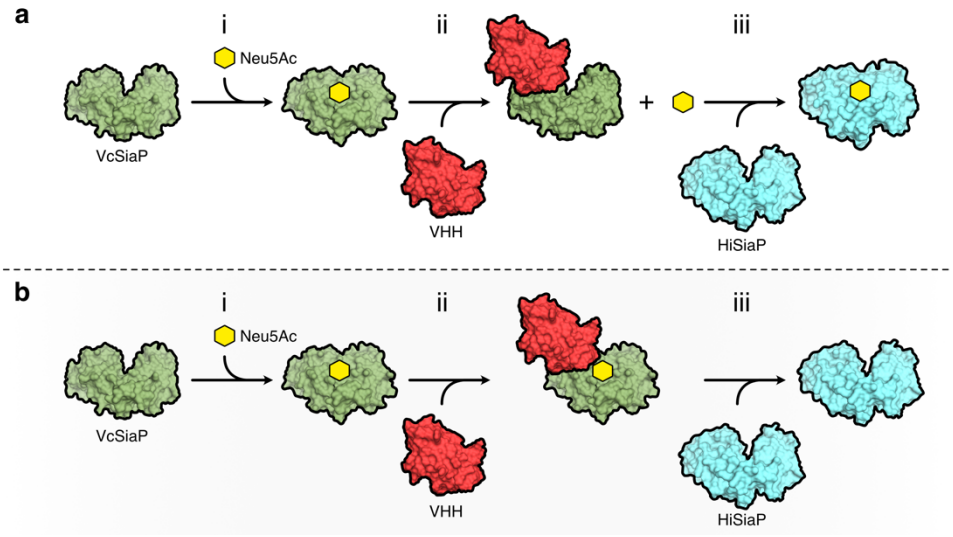


Figure 2-14: Schematic representation of the events during the sequential ITC analysis.
a) *The first scenario includes binding of sialic acid to VcSiaP (i) and binding of NbS002 towards the preformed complex (ii). Upon binding of NbS002, Neu5Ac is released from the substrate binding site and can be bound by HiSiaP in the next titration step (iii).* ***b)*** *The second scenario includes the same experimental steps as in a). However, is binding of NbS002 to VcSiaP[Neu5Ac] did not result in a release of sialic acid from the binding site, titration of HiSiaP in the last step would not result in sialic acid binding (iii). Note, that in the real experiment all titrants were added in a ~2-fold excess, thus, sialic acid will be detected in the last step of both scenarios, but the molar ratio would differ significantly.*

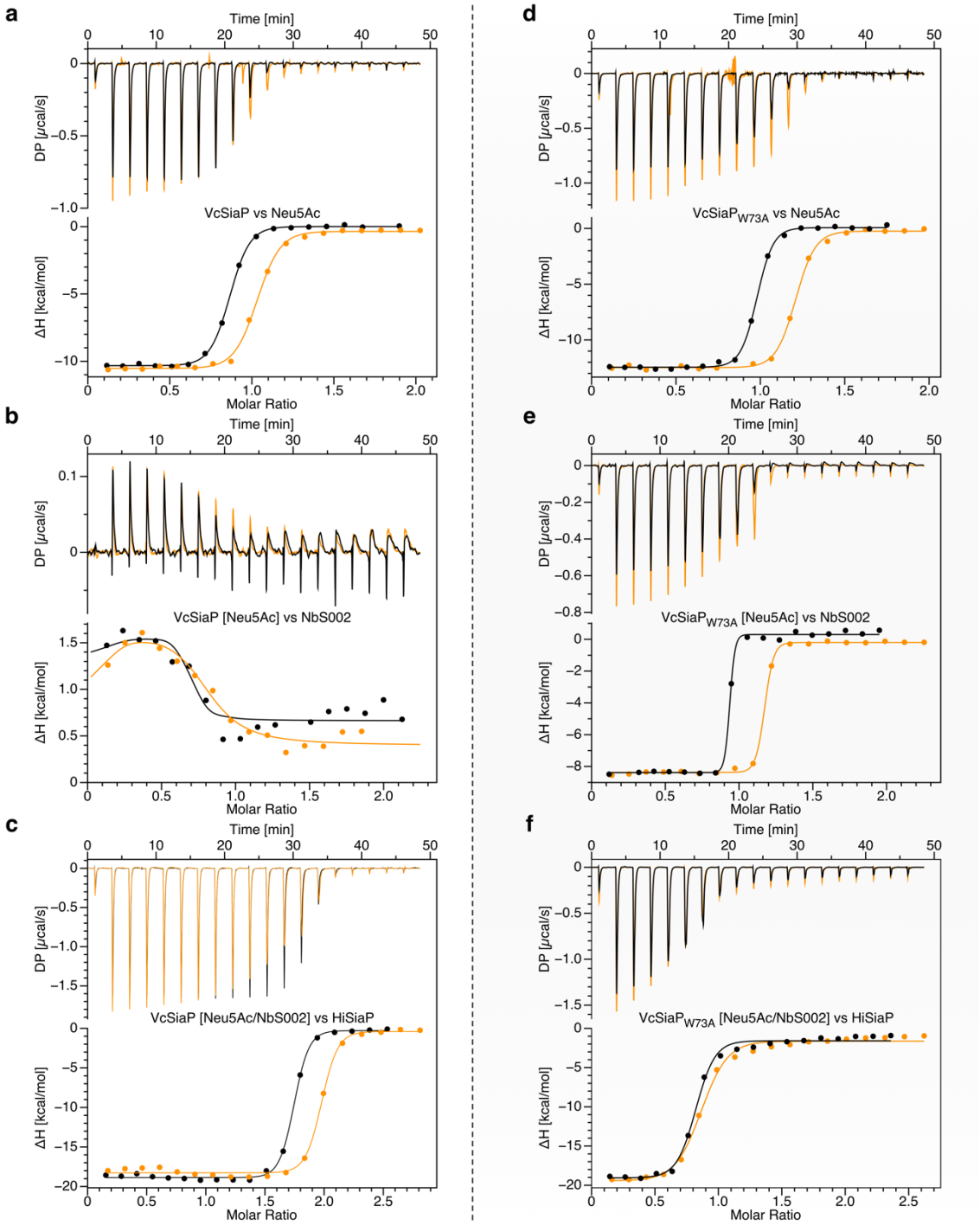


Figure 2-15: Results from sequential ITC analysis of wild type VcSiaP and W73A.
a) - c) *Results from the individual titrations of the sequential ITC experiment using wild type VcSiaP as indicated inside the graphs For the thermogram, the differential power was plotted against the experiment time. The resulting binding isotherm is shown below, with the enthalpy (ΔH) on the y-axis and the molar ratio on the x-axis. The black and orange curves depict the data of two individual replicates.* ***d)-f)*** *The same as a)-c) but for the titration series using VcSiaP W73A. The individual titration experiments are explained in the schematic representation in Figure 2-14.*

2.2.2.1 Structure data prove the recovery of conformational flexibility in VcSIaP$_{W73A}$

Biochemical analysis not only demonstrated the importance of the hydrophobic surface cavity for the conformational flexibility of the sialic acid TRAP transporter P-domain, but also showed that the allosteric inhibition induced by VHH binding to VcSiaP

could be neglected by reducing the steric strength inside the pocket of the SBP. To provide structural evidence that the W73A mutant of VcSiaP can bind sialic acid and undergo the substrate-induced conformational rearrangement even in complex with NbS002, X-ray crystallography was once again used. After several attempts, the protein complex crystallized with a 5-fold excess of sialic acid after more than 30 days of incubation at 20 °C (ProPlex G3 (Molecular Dimension)). Although the crystal was heavily intergrown (**Figure 2-16 a**), it was possible to collect acceptable diffraction data, and the structure was solved at a resolution of 2.8 Å. By molecular replacement, using the structure of holo VcSiaP (PDB ID: 7a5q) and the previously shown structure of NbS002 as search models, a structural model was obtained that contains 12 times the heterodimeric VHH-P-domain complex within the asymmetric unit (**Figure 2-16 b**). Nonetheless, in each SBP, sialic acid could be modeled into electron density within the substrate binding cleft. Compared to the VcSiaP[NbS002] structure described in **Section 2.1.3**, the sialic acid induced conformational change of the P-domain was clearly visible (**Figure 2-16 c**). While NbS002 prevented the closing mechanism of wild type VcSiaP, mutation of tryptophane 73 to alanine, indeed fully recovered the ability to undergo the conformational rearrangement by reducing the steric strength. Even though the VHH was still bound at the same position as observed for the apo structure, all residues rearranged in the same way as observed for VcSiaP in absence of NbS002. **Figure 2-16 d** shows a difference distance map to visualize the conformational rearrangement of the individual residues. The upper-left triangle shows the result of the comparison for apo and holo VcSiaP in absence of VHH (PDB ID: 4mag, 7a5q) and the lower-right part shows the same but using $VcSiaP_{W73A}$[Neu5Ac/NbS002] as holo structure. Obviously, both triangles look the same, and thus the same residues move in the same manner. The structure clearly demonstrated that upon decreasing the steric strength, the helices αC and αD are no longer prevented from moving inwards when NbS002 is bound to the SBP (**Figure 2-16 e, f**). Taken together, the biochemical and structural investigation demonstrated the importance of W73 in the allosteric inhibition of the sialic acid TRAP-transporter SBP from *V. cholerae* by the VHH antibody, NbS002.

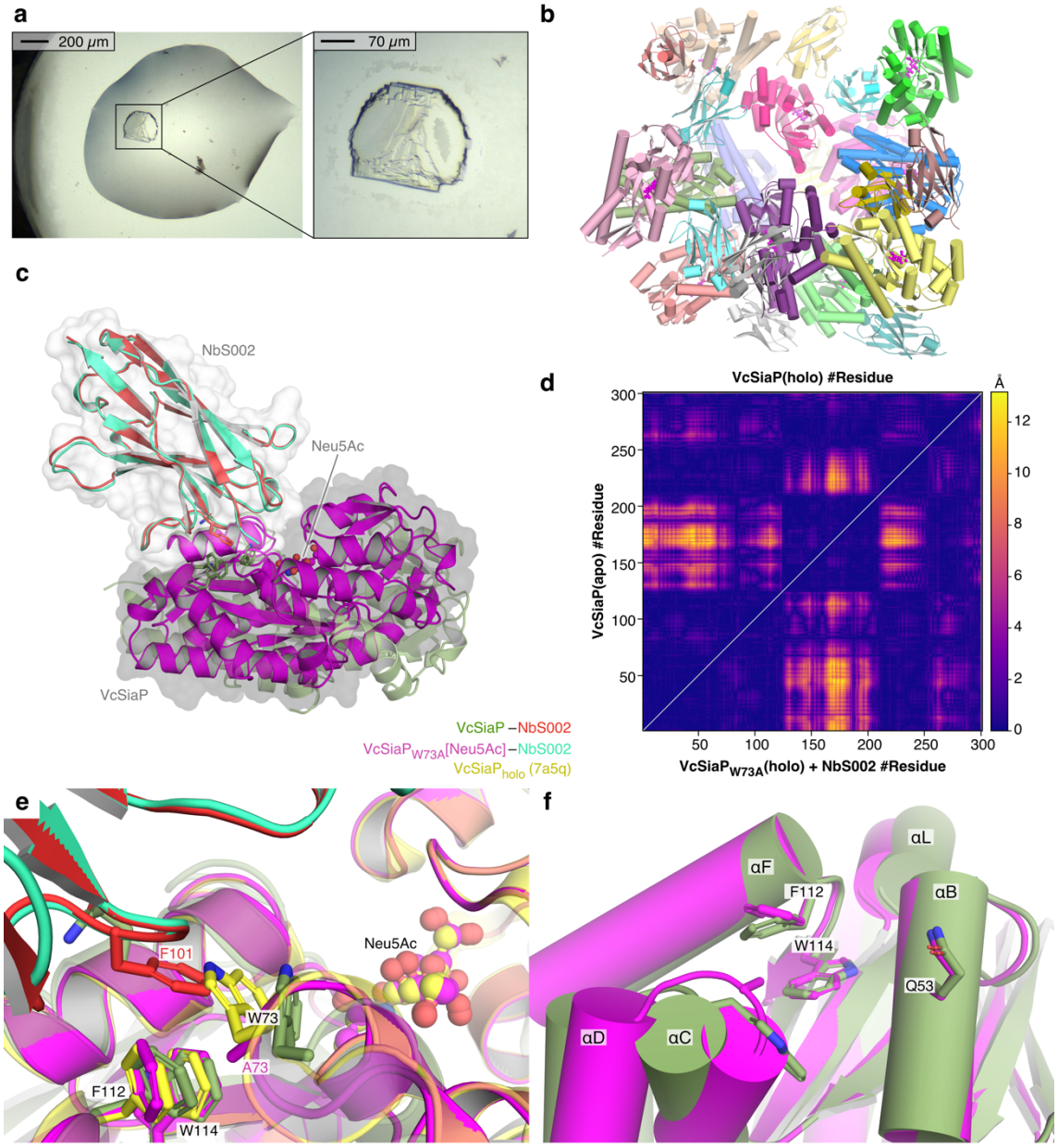


Figure 2-16: Structural analysis of VcSiaP$_{W73A}$ bound to sialic acid and NbS002.
a) *Image of the protein complex crystal taken after 48 days of incubation at 20 °C. A magnification is shown on the right-hand side.* ***b)*** *Overview on the asymmetric unit of the structure showing all 24 protein chains (colored by chain). For each P-domain, sialic acid (magenta ball and stick model) could be modeled into the electron density. Loops were smoothened to improve the visibility.* ***c)*** *Superposition of chains A and B of the VcSiaP$_{W73A}$[Neu5Ac/NbS002] structure (magenta, cyan) onto the VcSiaP$_{WT}$[NbS002] structure (green, red) to visualize the conformational changes.* ***d)*** *Comparison of the conformational changes between apo VcSiaP (PDB ID: 4mag) and holo VcSiaP (PDB ID: 7a5q) or VcSiaP$_{W73A}$[Neu5Ac/NbS002]. The difference distance matrices were generated with MtsslWizard (Hagelueken et al., 2012).* ***e)*** *Magnified view on the VHH binding region with superpositions of the structures mentioned above.* ***f)*** *Same as in c) but showing details of the VHH binding region to highlight the conformational changes of helices αC and αD. The color coding of panels c), e) and f) are kept consistent and are explained in the legend.*

3. Discussion Part I

3.1 Similarities and differences between the presented VHH antibodies

3.1.1 Comparison of the characterized VHHs based on their sequence

After two individual immunizations of two different alpacas (*Vicugna pacos*) with the P-domain of the sialic acid TRAP transporter from either *Vibrio cholerae* (VcSiaP) or *Haemophilus influenzae* (HiSiaP), a total of 11 VHH antibodies was selected. Two of these VHHs target VcSiaP and nine are HiSiaP specific binders. A high overall sequence conservation (65-81 %) was observed (see **Figure 2-2 b**). The most conserved regions are the framework regions that make up the backbone structure (Muyldermans, 2013) while the CDRs are highly variable among the different antibodies (see **Figure 2-2 b**). Interestingly, NbS003 and NbS005, which are both HiSiaP specific but were obtained from the two different libraries, share the highest sequence identity. Unfortunately, the high variance in the CDRs does not allow prediction of antigen specificity, nor whether overlapping or exclusive epitopes are being targeted. However, these observations demonstrate the importance of shape complementarity at the structural level (Akiba et al., 2019) rather than sequence complementarity. Inspecting the sequences of the framework regions reveals that the non-expressable VHHs (NbS007 and NbS008) contain destabilizing amino acid substitutions, such as an M to I substitution at position 34 (Dingus et al., 2022). The two VcSiaP specific VHHs are very similar in their backbone sequences with NbS002 having an unfavorable substitution (S58 instead of T58) which explains their similar stabilities and the slightly lower expression yields for NbS002. The variance within the framework composition of the HiSiaP specific antibody fragments is significantly higher, which is in line with the very different expression yields observed. It is puzzling that NbS003 binds only HiSiaP but not VcSiaP, although it was derived from the library obtained by the immunization with VcSiaP. Identification of this VHH was most likely a coincidence and the underlying immunoglobulin was a dysfunctional result of the complex antibody diversification process of the immune system (Alberts et al., 2002; Püschel et al., 2011), or even a functional antibody against a completely different target. As the antigen affinity is varying dependent on the stage of selection and maturation of the antibody during the immune response, NbS003 was most likely a result of an early stage and did not pass the affinity maturation process (Frank, 2002; Lavoie et al., 1999). The high specificity of antibodies against their distinct antigens seems to prevent cross-reactivity even for structurally similar homologs (50.17 % sequence identity between VcSiaP and HiSiaP). Considering the large repertoire of antibodies in the mammalian body it is astonishing that additional selectivity to host antigens is rather rare. Such phenomena of cross-reactivity or polyspecificity would in turn be impaired with autoimmune reactions (Keitel et al., 1997; Levin et al., 2002). Unfortunately, this seems to preclude the approach of re-panning an existing VHH library for specific binders against a similar protein. On the other hand, it was reported that antibodies raised against a particular epitope can sometimes bind to mutated epitopes with higher affinity and can even bind more strongly to related antigens (Van Regenmortel, 1998).

3.1.2 Comparing VcSiaP specific VHHs on a structural basis

The detailed structural analysis of VcSiaP bound to one of its specific VHH antibodies, revealed relatively small surface areas of interaction (557 Å^2 and 639 Å^2 for NbS001 and NbS002, respectively (Krissinel & Henrick, 2007)) compared to reports on VHH covered surface areas of 850 Å^2 to 1150 Å^2 (Desmyter et al., 2002). Nevertheless, high affinity binding in the nanomolar range could be shown (see **Figure 2-4**). For NbS002, a CDR3-stabilizing salt bridge was observed between residue D99 and R106 (**Figure 3-1 a**) which is similar to that described by Chien et al. to be crucial for the VHH stability (Chien et al., 1989). Interestingly, NbS001 lacks such a salt bridge, but rather the CDR3 region seems to be stabilized by partially folding into an α-helix that lies on the β-sheet backbone (**Figure 3-1 b**). But since antigen-binding affects the structural organization of the CDR loops, discussing the inherent stability of VHHs based on complex structures alone might not be an ideal approach.

Both nanobodies bind to concave surface sites matching reports on shape-complementary antigen recognition by VHH antibodies (see **Figure 2-8**) (Akiba et al., 2019). The binding epitopes targeted by both of the antibodies are of discontinuous nature, meaning that no primary sequence motif is recognized but rather the tertiary structure is essential to build up the three-dimensional epitope (Benjamin et al., 1984). Epitopes of that type are often more unique and less tolerant to mutations within the roughly 15 residues that build up the surface region (Benjamin & Perdue, 1996). The resulting interaction is more resistant to competitive effects by small peptide sequences as the probability for the peptide to attack in the right conformation is comparably low and thus, peptides normally need to be present in a large excess to increase this probability (Hodges et al., 1988; Keitel et al., 1997).

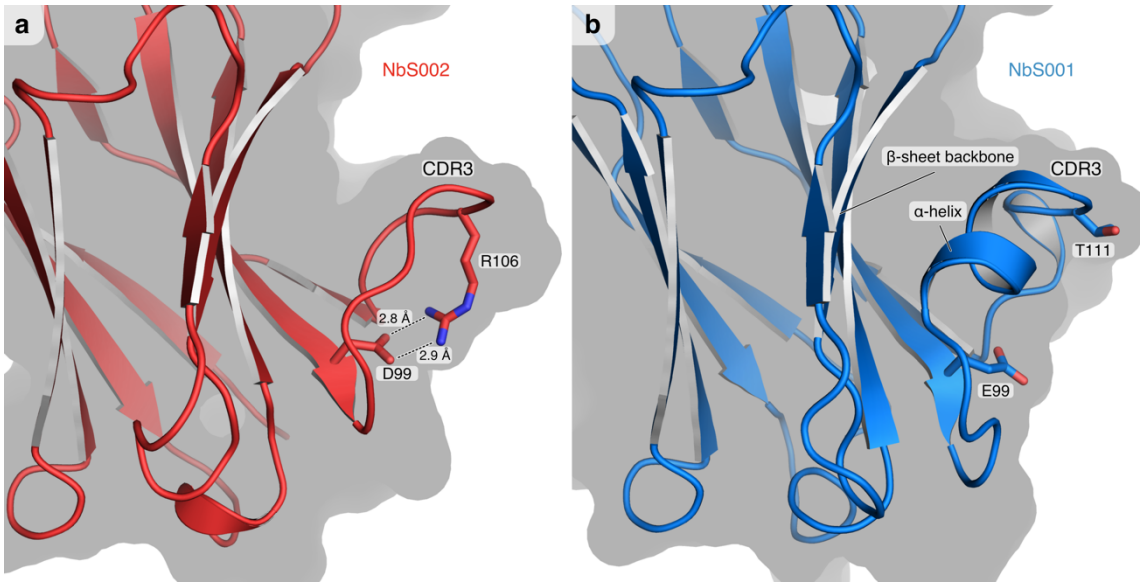


Figure 3-1: Different types of stabilization for the CDR3 region of VcSiaP specific Nanobodies. ***a)*** *The crystal structure of NbS002 reveals a salt bridge between CDR3 residues D99 and R106. The bond lengths are given in Ångström.* ***b)*** *In the crystal structure of NbS001, no interaction between the residues in this position was observed, but the CDR3 partially folds into an α-helix that lies flat on the β-sheet backbone of the VHH.*

3.2 Biochemical and structural analyses of the VcSiaP/NbS001/NbS002 complex

The structural analysis of VcSiaP specific VHH antibodies, NbS001 and NbS002, is consistent with the results of the biochemical characterization. Both VHHs bind to individual epitopes and binding of one does not interfere with binding of the other one (see **Figure 2-7**). For the heterotrimeric complex of both VHHs and their antigen, VcSiaP, the presented analyses show evidence for an existing homodimer, so does the crystal structure. The hydrodynamic radius was observed to be slightly larger than expected (R_{calc}=3.44 nm, R_{obs}=3.78 nm) but did not reach the calculated value for a dimer of the heterotrimeric complex ($R_{calc,dim}$=4.39 nm) (Fleming & Fleming, 2018) (see **Figure 2-7 a**). The thermal stabilization of VcSiaP in complex with both VHHs was determined to $\Delta T_{[NbS001\ \&\ NbS002]}$=16 °C which is more than twice the stabilizing effect of one single VHH ($\Delta T_{[NbS001\ /\ NbS002]}$≈6 °C).

The described observations could be explained by a weak interaction between two VcSiaP[NbS001, NbS002] complexes (see **Figure 2-7 b**). The crystal structure of the complex reveals such interactions in between NbS001-1 and NbS001-2 ($A_{interface}$≈300 Å^2), NbS002-1 and NbS001-2 ($A_{interface}$≈318 Å^2), and VcSiaP-1 and NbS001-2 ($A_{interface}$≈70 Å^2) (Krissinel & Henrick, 2007) (**Figure 3-2** and **Figure 2-9**). Even though the total interface area of A_{total}≈688 Å^2 is of the same range as the stable direct antibody antigen interaction interfaces, the described dimer interface is most likely a crystal packing artefact (Janin & Rodier, 1995) or at least a weak interaction with an affinity weaker than K_D=1 mM, as no evidence for a stable complex in solution was observed during SEC-MALS analysis (Stevens & Schiffer, 1981) (see **Figure 2-4 a**). The slightly increased radius in DLS experiments might be explained by such a weak dimeric assembly of the heterotrimeric complex, but neither biochemical data, nor the structural data are sufficient to state the existence of such a complex in solution.

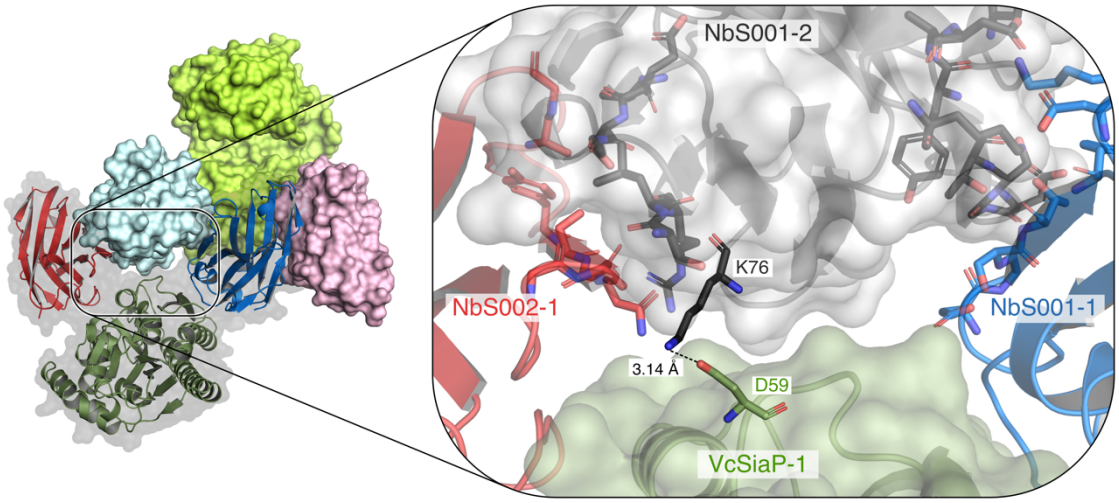
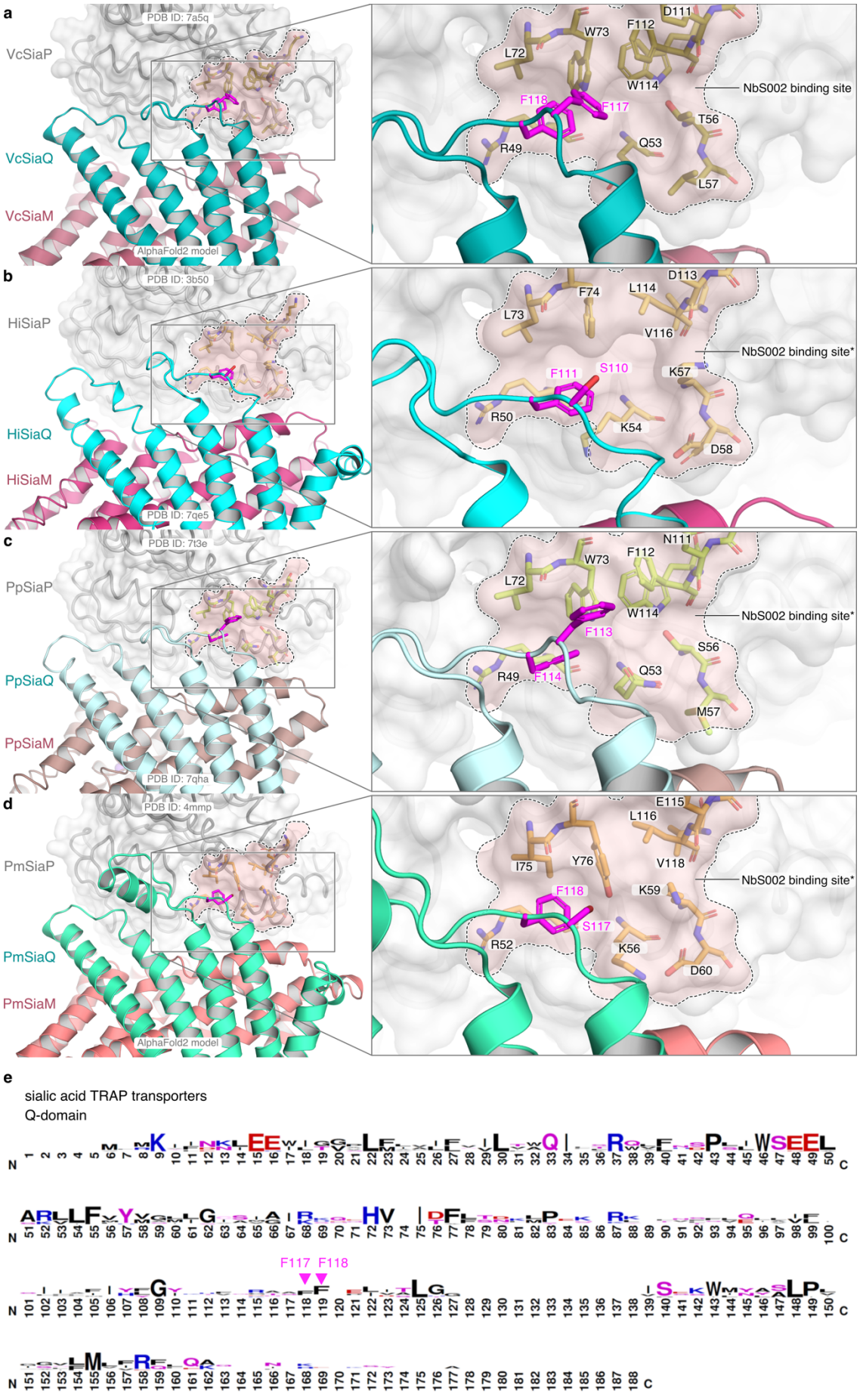


Figure 3-2: Overview on the dimeric assembly of the VcSiaP[NbS001, NbS002] complex.
Homodimeric assembly of the heterotrimeric VcSiaP[NbS001, NbS002] could explain an increased hydrodynamic radius and a high thermal stability. However, this complex may result from crystal packing artifacts and has not been unambiguously proven in solution. For more details on this crystal structure see Figure 2-9.

3.3 Mechanistical relevance of the identified conformation-specific surface cavities

Conformation-specific surface cavities, as the one targeted by NbS002, do not only exhibit a potential for drug design, but also highlight the fragile interplay of multiple residues during the conformational rearrangement of TRAP transporter P-domains upon substrate binding (Glaenzer et al., 2017; Peter et al., 2021). We could show that binding of NbS002 to VcSiaP and insertion of a phenylalanine residue into this cavity leads to a release of sialic acid from the high affinity binding site. By mutating W73 of VcSiaP to an alanine, the steric pressure could be reduced which resulted in a loss of this allosteric push-to-open mechanism. Modeling the tripartite PQM-complex, based on experimental structures of the TRAP transporter membrane domains, revealed an interesting periplasmic loop of the Q-domain (J. S. Davies et al., 2023; Peter, Ruland, et al., 2022). This mainly hydrophobic loop contains conserved phenylalanine residues in close proximity to the hydrophobic cavity where NbS002 binds to (**Figure 3-3**). This observation could be a hint towards an allosteric mechanism to 'crack up' the shell-like substrate bound P-domain once it binds to the transporter QM-domains. Mechanisms to allosterically induce conformational changes by steric pressure have not been reported for SBP-dependent transporters, but are known to activate ion channels (Kapsalis et al., 2019; Y. Wang et al., 2021), and also from a G-protein-coupled receptor (GPCRs) where agonist binding pushes against specific residues to cause conformational changes in the ligand binding pocket (Katritch & Abagyan, 2011). A push-to-open mechanism would not exclude but rather support an allosteric basis of substrate hand-over by forcing a conformational rearrangement that leads to a decreased affinity of the SBP towards its substrate (Marinelli et al., 2011). In this case, the upwards motion of the elevator subdomain forces the bound P-domain to open and release the substrate into the membrane translocation channel (Peter, Ruland, et al., 2022; Peter et al., 2024). As the SBP-substrate interaction is of high affinity and the enthalpic contribution to the binding process is quite large ($\Delta H(VcSiaP)=10.8$ kJ*mol^{-1}*K^{-1}, $\Delta H(HiSiaP)=18.7$ kJ*mol^{-1}*K^{-1}), 'cracking up' the P-domain just by the upwards movement of the elevator would most likely be energy intensive. A cooperative effect of the elevator motion and a push-to-open mechanism could drastically decrease the energy consumption of the transport process. The existence of such a process might explain why evolution did not yield an ultimate binder, as we were able to increase the substrate affinity of VcSiaP by mutating W73 to an alanine side chain. Highlighting the fragile interplay between optimized affinity and selectivity for substrate uptake and the need of an energetically efficient substrate release mechanism.

Figure 3-3: Structural modelling reveals a hydrophobic loop next to the NbS002 binding site. ***a)*** *Model of the tripartite SiaPQM complex from* Vibrio cholerae*. The model was built up from AlphaFold2 predictions and experimental structures. The binding area of VHHVcP #2 is indicated and labelled. The Q (teal) and M(red) transmembrane domains are depicted as cartoon model and the P-domain is shown as surface representation. On the right-hand-side, a magnification of the region of interest is shown and the highlighted amino acid residues are labeled.* ***b) – d)*** *Same as a) but for the tripartite SiaPQM complex from* Haemophilus influenzae *(b),* Photobacterium profundum *(c), and* Pasteurella multocida *(d), respectively.* ***e)*** *Sequence logo to visualize the hydrophobic loop of the Q-domain that is conserved among sialic acid TRAP transporters and includes the conserved phenylalanine residue 118 (111 in HiSiaQ). This image is an excerpt from (Peter, Ruland, et al., 2022) and was created with WebLogo (Crooks et al., 2004).*


a
PDB ID: 7a5q
VcSiaP
VcSiaQ
VcSiaM
AlphaFold2 model
D111
W73
F112
L72
W114
NbS002 binding site
F118
F117
T56
R49
Q53
L57
b
PDB ID: 3b50
HiSiaP
HiSiaQ
HiSiaM
PDB ID: 7qe5
D113
F74
L114
L73
V116
NbS002 binding site*
F111
S110
K57
R50
K54
D58
c
PDB ID: 7t3e
PpSiaP
PpSiaQ
PpSiaM
PDB ID: 7qha
N111
W73
F112
L72
W114
NbS002 binding site*
F113
S56
R49
F114
Q53
M57
d
PDB ID: 4mmp
PmSiaP
PmSiaQ
PmSiaM
AlphaFold2 model
E115
L116
I75
Y76
V118
NbS002 binding site*
F118
K59
S117
R52
K56
D60
e
sialic acid TRAP transporters
Q-domain
F117 F118

The importance of an interaction between P-domain and QM was demonstrated by *in vivo* transport assays showing decreased cell growth for mutants in the potential P–QM interface (Peter, Ruland, et al., 2022). This study also investigated the role of the conserved Q-loop phenylalanine (F118 in VcSiaQ, F111 in HiSiaQM) but did not reveal any effects for a Phe to Ala mutant. But it should be noted that these *in vivo* experiments are not well suited to characterize subtle effects. Even P-domain mutants that showed no sialic acid binding in biochemical experiments (Fischer et al., 2015), did not abolish bacterial growth (Peter, Ruland, et al., 2022). On the other hand, these observations might be a hint, that blocking the interaction could be more efficient than inhibiting substrate binding by the SBP. The essential role of membrane transporter proteins, which represent a bottleneck in many cellular processes, is reflected in the fact that numerous drugs target such proteins ("Cellular Gatekeepers," 2016; Overington et al., 2006).

Although the current state of research does not allow to definitively propose mechanistic details, the observations discussed could affect the proposed elevator mechanism (Peter, Ruland, et al., 2022) in the following ways. When the substrate-bound SBP binds to the transmembrane domains, the partially conserved Q-domain loop binds to the allosteric pocket (**Figure 3-4 a**). This interaction forces the P-domain to open (**Figure 3-4 b**), weakening the substrate affinity (**Figure 3-4 c**) and decreasing the energy to move the elevator upwards (**Figure 3-4 d**).

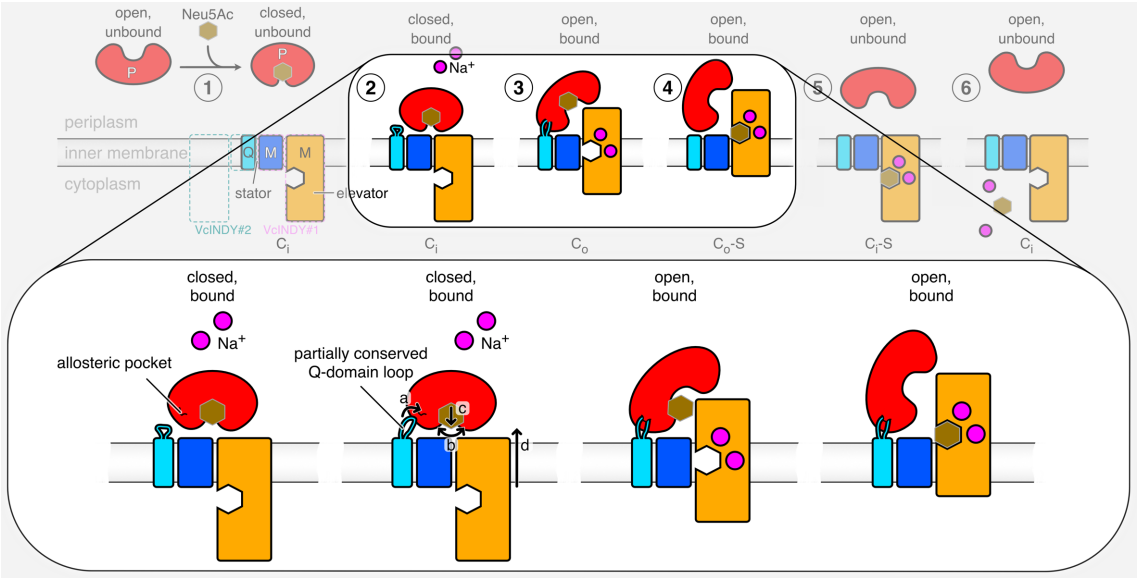


Figure 3-4: Elevator mechanism including an allosterical triggered SBP opening mechanism. *The substrate-bound P-domain (red) binds to the QM-domains (blue, orange) and the partially conserved Q-domain loop binds to the allosteric pocket of the SBP (**a**). This interaction allosterically triggers the opening of the P-domain (**b**), simultaneously weakens the substrate affinity (**c**) and decreases the energy needed for the elevator upwards-movement (**d**).*

3.4 Conformation-specific cavities as starting point for drug discovery

The detailed characterization of the P-domain interactions with specific VHH antibodies demonstrated the importance of concave surface areas for the open-closed transition of the SBP. Simultaneously, the identified surface cavities were found to be capable of binding molecules of the size of amino acid residues. The surface cavity targeted by NbS002 could be detected by *in silico* methods (Hussein et al., 2015; S. Wang

et al., 2023; Xu et al., 2018) only for the open state structure of VcSiaP (PDB ID: 4mag, (Setty et al., 2014)) but not the closed conformation (PDB ID: 7a5q (Peter et al., 2021)). The NbS001 binding site was not predicted, neither for the closed- nor for the open conformation. This approach yielded several further surface cavities that could be potential binding sites for small molecule compounds (**Figure 3-5**). Some of them were found for both conformational states (B/B', E/E', G/G', and H/H') but the absolute geometry and size varied. The potential small molecule binding sites that were predicted for one conformational state exclusively might be of special interest, as they potentially inhibit the open-closed transition in a similar way as NbS002.

The design and identification of small molecules that specifically target surface cavities as those described above could be addressed by fragment-based drug design (FBDD). This screening technique has grown constantly within the last 20 years to become a frequently used alternative to traditional screening of compound libraries (Erlanson et al., 2016). To get a handle on the huge estimated number of more than 10^{60} possible drug-like molecules (Bohacek et al., 1996) traditional compound libraries are often coupled to high-throughput screening (HTS) techniques (Erlanson et al., 2016). On the other hand, fragment libraries that cover a similar chemical space can be orders of magnitude smaller. This beneficial characteristic is a result of the idea that not drug-sized molecules are screened to identify lead compounds, but smaller molecules with fewer atoms (typically less than 15 non-hydrogen atoms) and functional groups (a maximum of

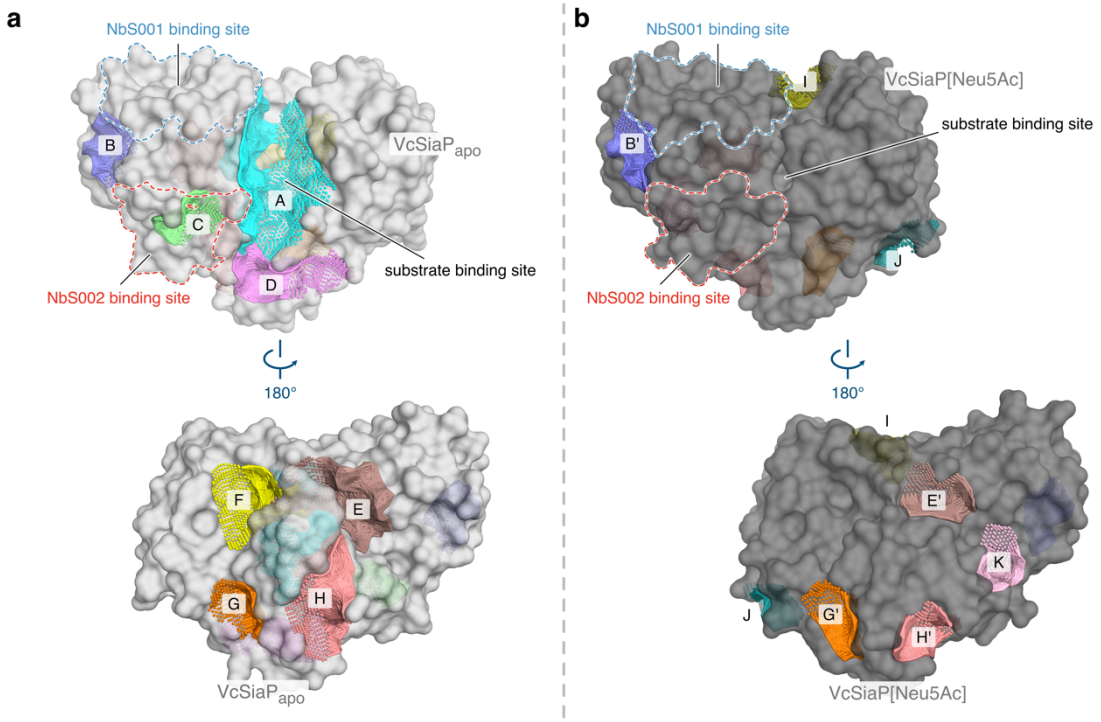


Figure 3-5: Overview of predicted druggable surface cavities of VcSiaP.
a) *Predicted druggable surface cavities of VcSiaP in its open conformation (labelled A-H) mapped onto the surface representation of the structure (PDB ID: 4mag). The two view angles are rotated by 180° as indicated. For the top view (upper image), the regions that interact with the two VHHs are indicated and the position of the substrate binding site is annotated.* ***b)*** *Surface cavities that are potentially druggable of VcSiaP in its substrate-bound conformation mapped onto the surface representation of the structure (PDB ID: 7a5q). The two view angles are rotated by 180° as indicated. For the top view (upper image), the regions of interacting amino acid residues with the two VHHs are indicated and the position of the substrate binding site is annotated. The cavities that are simila to those shown in a) are labelled B', E', G',and H'. Cavities that were found for the closed conformation only are labelled I and J. The prediction was done using the CavityPlus server (S. Wang et al., 2023; Xu et al., 2018).*

two functional groups per fragment) (Ruddigkeit et al., 2012). These ‘simple’ chemical building blocks often show only weak binding to the selected target, but can be optimized on a structural basis to yield high affinity binders in the end. Nowadays, structurally diverse fragment libraries are available for screening by co-crystallization with the target protein or soaking of native protein crystals (Barthel et al., 2022; Wollenhaupt et al., 2020). Since the surface pocket of interest is penetrated only by a phenylalanine residue of the VHH antibody, small molecule fragments may be suitable for mimicking this binding mode. Of course, such an approach requires reliable crystallization conditions but given the gained expertise of crystallizing VcSiaP in both conformational states with and without nanobodies we have a good basis to develop optimization screens that serve these prerequisites.

3.5 Outlook

The results demonstrate that the substrate-induced conformational changes of the TRAP transporter P-domains, depend not only on the hinge helix and residues inside the binding pocket, but also on the orientation and environment of surface-exposed amino acid residues. We identified an open-state specific surface cavity that provides insights into a possible substrate-release mechanism. We showed that this allosteric pocket can be used to inhibit substrate binding by the P-domain making this pocket an interesting target for fragment-based drug design. The aim is to find small molecule binders for this surface area that can perform this function, similar to NbS002.

The mechanistic relevance of the allosteric pocket and the partially conserved hydrophobic Q-domain loop should be further investigated. To study the effect of mutations within these regions on the transporter function, a transport assay using radioactively labeled Neu5Ac may be appropriate. Such assays have previously been done for TRAP transporters (Mulligan et al., 2009). Radioactive transport assays are often very sensitive and could therefore capture a wider range of transport efficiency, in contrast to the *in vivo* growth assay, which only shows full growth or no growth with few gradations in between (Peter, Ruland, et al., 2022). Instead of sensing imported sialic acid by radioactivity, it might also be possible to design a fluorophore-based readout system. For example, by attaching a FRET pair to HiSiaP, as it was done previously (Peter, Gebhardt, et al., 2022), the substrate-dependent conformational changes would lead to a difference in the FRET signal intensity. A similar sensor could be constructed using a single fluorophore, if an environmentally sensitive fluorescent dye such as IANBD is used. Preliminary experiments showed, that the fluorescence intensity is indeed dependent on the sialic acid concentration. In all cases, the transporter domains must be reconstituted into liposomes and a signal is generated if sialic acid is transported into the liposomes.

In another approach, conformation-specific nanobodies could be used as a readout tool. A closed-state-specific VHH antibody would be perfect for this, as it binds to the SBP only in presence of substrate. This binding could be converted into a signal by FRET. If one label is attached to the VHH and another to the SBP, there is a very broad distribution of FRET distances in the unbound state, whereas it is defined when the two proteins interact.

PART II

Bacterial immune strategies to survive viral attacks.

Abstract

Like higher organisms, also bacteria have evolved immune strategies to defend against mobile genetic elements, such as phages or plasmids. CRISPR systems represent one of the most prominent groups of prokaryotic immune systems and are classified into two classes with a large variety of types and subtypes. Type III CRISPR systems enable a highly complex and versatile immune response via the synthesis of cyclic oligoadenylate (cOA) second messengers, which are known to bind to the CARF domain of specific effector proteins and thereby activate them. Numerous effector domains with diverse functions have been identified, ranging from RNases to transcriptional regulators to DNases. Bioinformatics studies have reported an unusual effector protein, a membrane-bound "CRISPR-associated Lon protease" (CalpL). Here, we present the structural and mechanistic analysis of this cA_4 activated protease and the associated phage defense system. Unlike predicted, we found that CalpL is a soluble monomeric protein that is activated by cA_4-induced oligomerization. The activated protease cleaves the CalpT/S complex, which consists of two small proteins that are encoded in the same operon, directly upstream of CalpL. Several approaches to investigate the function of CalpT and CalpS finally led us to identify a bacterial sigma/anti-sigma factor system. We found that upon viral RNA recognition, cA_4 activates CalpL to cleave the anti-sigma factor CalpT, releasing CalpS from the complex and allowing it to bind to RNA polymerase to regulate cellular transcription. This finding directly links a type III CRISPR system to the transcription machinery of the cell. Our results show similarities to the CBASS system, recent reports on CRISPR-activated caspases (Craspases), and even to mammalian systems such as the cGAS-STING pathway.

4. Introduction Part II

4.1 Bacterial immunity

Like eukaryotes, also prokaryotes are under intense evolutionary pressure from multiple threats. In addition to physical stress such as heat shock, bacteria also face infection by bacteriophages, viruses that hijack the bacterial cell machinery. The existence of phages was first reported during World War I by Frederick Twort and Félix d'Hérelle. Shortly after the discovery, d'Hérelle also recognized the potential of using phages to treat bacterial infections (Chanishvili, 2012; Duckworth, 1976). With an enormous number of estimated 10^{31} phages in the biosphere, they are the most abundant biological entity, outnumbering bacteria by a factor of more than 10 (Brüssow & Hendrix, 2002; Fortier & Sekulovic, 2013). To combat the high rate of phage infections, bacteria have evolved a broad repertoire of sophisticated active defense strategies. The large number and diversity of immunity-conferring mechanisms is a result of phage-host co-evolution and faces an equally wide range of offensive strategies (Dy et al., 2014; Labrie et al., 2010; Stern & Sorek, 2011). The vast arsenal of known antiphage strategies has grown in recent years, largely due to the development of culture-independent sequencing techniques (Riesenfeld et al., 2004). Among the best-known bacterial defense strategies, besides restriction-modification (R-M) and CRISPR-Cas, are abortive infection systems (Abi), which lead to cell death or dormancy, and other less characterized mechanisms such as prokaryotic Argonautes (pAgos), BREX and DISARM (Doron et al., 2018).

For an effective protection against bacteriophages and other mobile genetic elements (MGEs), most bacterial genomes encode several different defense systems in a unique and strain-specific manner. The genes are often clustered and encode nucleases, helicases, proteases and much more (Georjon & Bernheim, 2023; Makarova, Wolf, et al., 2011). To ensure an effective protection without harming the host, most systems combine a sensor and an effector element to specifically sense and then target an invader (Georjon & Bernheim, 2023).

4.2 CRISPR provides adaptive immunity against bacteriophages and plasmids

In the late 1990s and early 2000s, different groups studied peculiar repetitive DNA sequences in the prokaryotic genomes. *In silico* analyses revealed that the short (~30 bp) and partially palindromic repeat sequences are regularly interspaced by unique sequences of similar length. These motifs were observed to be clustered and often flanked on one side by a leader sequence of 300-500 base pairs (Jansen et al., 2002; Mojica et al., 1995, 2000). Ensuing from those characteristic features, this family of repetitive sequences was termed "clustered regularly interspaced short palindromic repeats" (CRISPR) (Jansen et al., 2002). The repeat sequences of closely related species appeared to be highly conserved and even share equivalent motifs in very distant phylogenetic groups (Jansen et al., 2002; Mojica et al., 2000). In contrast to that, the unique spacer sequences were found to differ even in very close relatives (Pourcel et al., 2005). Mojica and coworkers not only reported that those sequences are identical to fragments of phage and plasmid DNA, but also showed that the cells containing a specific spacer were not infected by the corresponding virus (Mojica et al., 2005). Along with similar findings by other groups, this suggested CRISPR to function as a prokaryotic defense system against mobile genetic elements. Repurposing the function of this highly effective adaptive immune system to develop a tool for RNA-guided DNA manipulation, increased the popularity of CRISPR

and demonstrated once more the potential of prokaryotic systems for biotechnology (F. Jiang & Doudna, 2015).

4.2.1 Classification and general organization of CRISPR Cas systems

Roughly 40 % of bacterial and most archaeal genomes contain CRISPR systems which are organized in so called CRISPR loci. These contain the CRISPR array comprised of the clustered repeats and spacers, and variable CRISPR associated sequences (Cas) arranged in operons (Georjon & Bernheim, 2023; Jansen et al., 2002).

For different CRISPR variants, the locus architecture and the actual composition of Cas genes show a high diversity, which makes classification a difficult task (Makarova et al., 2015; Takeuchi et al., 2012). Earlier classification was based on the phylogeny of the conserved *cas1* gene. However, due to the rapid discovery of more and more new CRISPR systems, this type of organization became problematic (Makarova et al., 2018). To solve this problem, a new approach was developed. At the first stage, with the strongest basis of differentiation, CRISPR systems are divided into class 1, containing multiple subunit effector complexes, and class 2, harboring single subunit effector complexes. The next stage separates the six different types based on the mechanism of action, with type I, III and IV being part of class 1 systems and type II, V and VI of class 2 systems. Furthermore, each type is characterized by a unique signature gene. This links type I to *cas3*, type III to *cas10*, and type IV to *csf1* (Makarova et al., 2015). For class 2 systems, the well-known *cas9* gene is associated with type II, *cas12* with type V, and *cas13* with type VI (S. Shmakov et al., 2017). Separation of the individual types into subtypes becomes complicated. While for some subtypes, specific *cas* gene variants can be identified, other subtypes are mainly characterized by their locus organization. This results in subdividing the six types into several sub-types ranging from two to eight or more per CRISPR type (Makarova et al., 2018). **Figure 4-1** shows an overview on the different classes, types and sub-types.

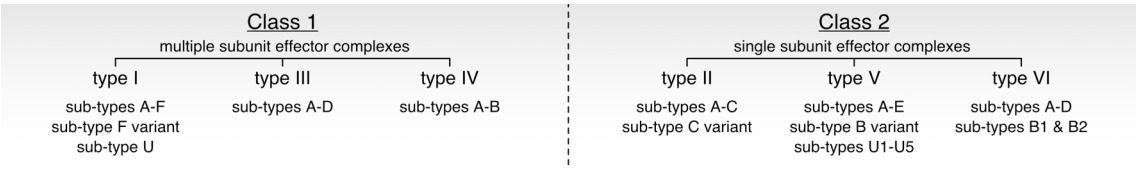


Figure 4-1: Classification of CRISPR systems.
Class 1 and class 2 are each divided into three different types, which in turn can be separated into multiple subtypes.

Despite the diversity and variability, all systems contain similar core elements needed for the immune response. Most systems rely on a metal-dependent DNase, Cas1, and a metal-dependent endoribonuclease, Cas2, for spacer adaption. But often also other *cas* genes are involved in the insertion of proto-spacers. Furthermore, the characteristic CRISPR array, as well as the leader sequence are essential components of CRISPR immunity. When it comes to the interference stage, things become much more individual and specific. **Figure 4-2** shows the Cas9-encoding gene as well as the tracrRNA which are essential for the type 2 specific processing of crRNA (Barrangou, 2013; Chaudhuri et al., 2022).

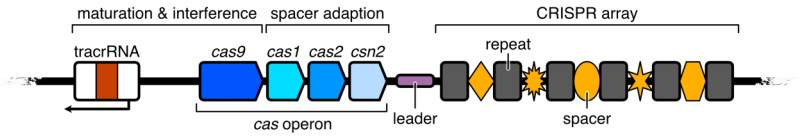


Figure 4-2: Representation of an exemplary CRISPR locus.
A typical CRISPR locus architecture containing elements which are required for the different stages of a CRISPR immune response. The CRISPR array contains the name-giving repeat-spacer sequences, and the cas operon contains the individual cas *genes. The* cas1 *and* cas2 *genes are highly conserved among most CRISPR systems, as well as the presence of a leader sequence. The presence of a tracrRNA element and the* cas9 *gene, in contrast, is specific for class 2, type II CRISPR systems. This figure was modified from Chaudhuri et al., 2022.*

4.2.2 Similarities and differences in CRISPR Cas mediated immune responses

The similar yet distinct structure of CRISPR loci of different classes and types is reflected in their characteristic functionality. The highly conserved Cas1 and Cas2 proteins form the core elements of CRISPR Cas systems and are involved in spacer acquisition and insertion during the adaption stage. The two proteins assemble to a heterohexameric integrase complex with one Cas2 homodimer in the middle, sandwiched by two Cas1 homodimers (Barrangou, 2013; Nuñez et al., 2014). A conserved E-H-E motif within one of the four Cas1 active sites of the hexameric assembly is essential for catalyzing the spacer integration at the leader end of the CRISPR array. Thereby, the histidine functions as base to activate the 3'-OH of the prespacer for nucleophilic attack at the 5' phosphate of the terminal repeat (Xiao et al., 2017). Along with each spacer integration event, a new repeat is generated. The detailed structural assembly of the integrase complex differs in the respective orientation of the dimeric subunits as well as in the presence or absence of auxiliary Cas proteins. These additional components are thought to essentially contribute to the protospacer selection (Sasnauskas & Siksnys, 2020). Notably, some type III systems are able to acquire new spacers from viral RNA by employing a reverse transcriptase which is encoded in the CRISPR locus and is often fused to Cas1 (Silas et al., 2016). All these differences are reflected by different substrate preferences as well as varying spacer and repeat lengths (Sasnauskas & Siksnys, 2020). Also, the position of a protospacer-adjacent motif (PAM) within the recognized foreign nucleic acid sequence, or even the complete absence of such a sequence feature, highlights the individuality of the different CRISPR variants (**Figure 4-3**.1). PAMs are short nucleotide sequences that enable an effective selection of viral DNA by helping to discriminate between self and non-self, and they also seem to play essential roles during interference (Marraffini & Sontheimer, 2010; Mojica et al., 2009). However, the details of proto-spacer selection and the certain role of PAMs during this mechanism remain only partially understood (Gleditzsch et al., 2019; Leenay et al., 2016).

The collection of spacer-repeat units represents a library for adaptive immunity which is transcribed to a long pre-crRNA in the expression stage (**Figure 4-3**.2). The non-coding, A/T rich leader sequence upstream of the CRISPR array acts as a promoter for this cellular mechanism (Barrangou, 2013). The subsequent processing of the primary transcript into mature short crRNAs differs for the different CRISPR types. While for type I systems a CRISPR associated complex for antiviral defense (*Cascade*) containing the endoribonuclease Cas6 is essential for the specific cleavage of the pre-crRNA, type II systems employ the housekeeping RNase III in combination with Cas9 and a *trans*-encoded small RNA (tracrRNA). This unique element is also included in the CRISPR locus (see **Figure 4-2**), contains a repeat-complementary sequence and activates pre-

crRNA cleavage (Barrangou, 2013; Makarova, Haft, et al., 2011). Processing of pre-crRNA in type III CRISPR systems occurs in two steps. First, the primary transcript is sequence-specifically diced by the endoribonuclease Cas6, leading to a set of full spacer sequence which are flanked by 8 nts of the repeat at the 5' end and the rest of the repeat sequence (~30-40 nts) at the 3'end. Subsequently, this RNA fragement is handed over to a *Cascade*-like protein complex, built of Csm- (type III-A) or Cmr-proteins (type III-B), to allow further trimming of the 3' end by exonucleases (Carte et al., 2008; Hale et al., 2009). The short repeat-derived tag remains at the 5' end of the mature crRNA and was

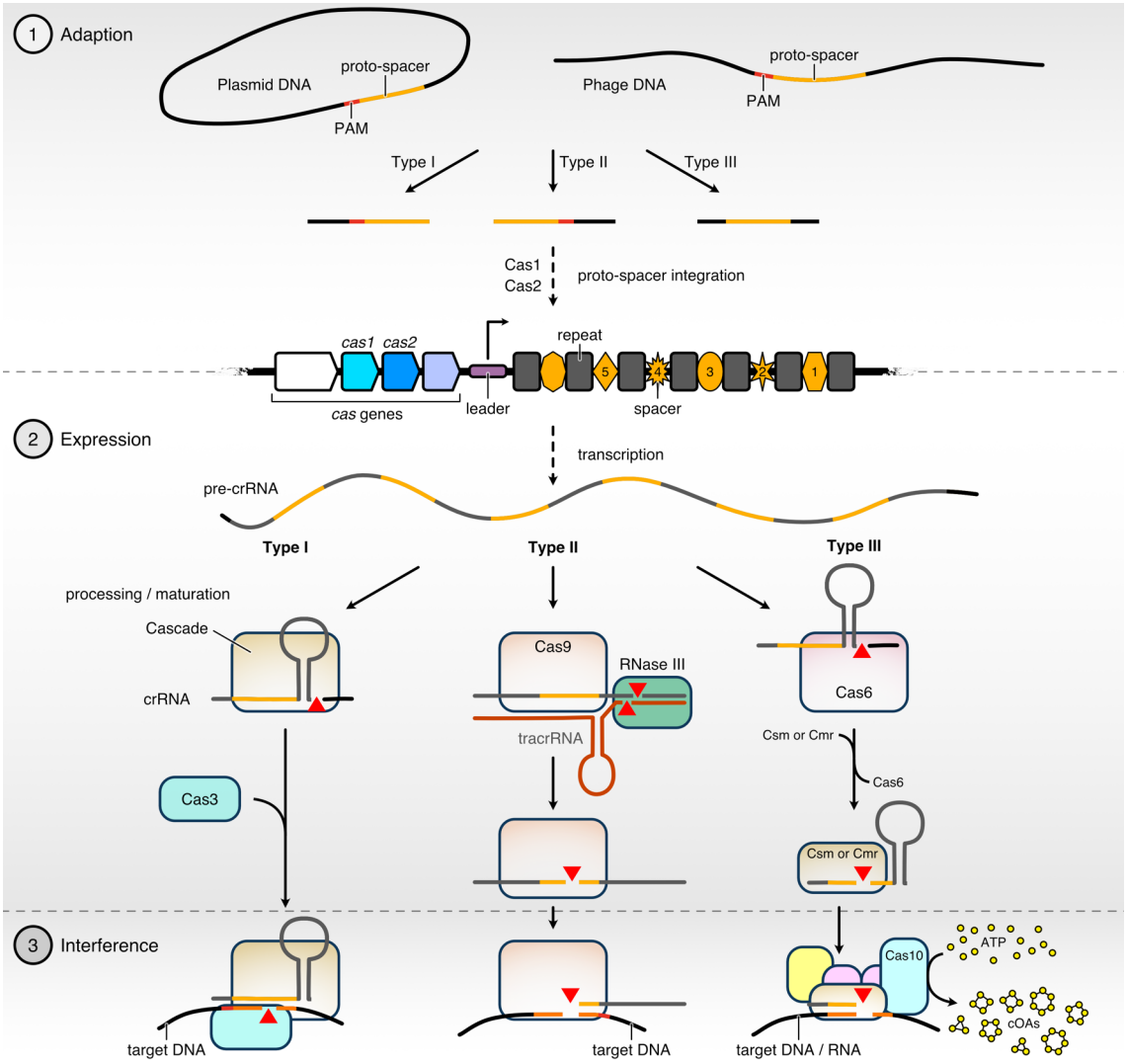


Figure 4-3: Similarities and differences in the mode of action of different CRISPR types.
*The three stages adaption (1), expression (2) and interference (3) of a CRISPR-Cas mediated immune response. Adaptive immunity is achieved by sensing of invasive nucleic acids and integration of fragments as proto-spacers into the bacterial CRISPR array. For type I and type II systems, but not for type III systems, so-called proto-spacer-adjascent motifs (PAMs) are key features for the recognition and incorporation of foreign nucleic acids. Spacer acquisition is catalyzed by the conserved proteins Cas1 and Cas2 and occurs at the leader end (**1**). The CRISPR array is transcribed during the expression stage, resulting in a long pre-crRNA. This primary transcript is further processed to yield mature crRNAs which form a ribonucleoprotein complex with various Cas effector proteins (**2**). During the interference, the crRNAs are essential guides to pair with reverse complementary sequences of the target DNA or RNA which is subsequently cleaved by CRISPR nucleases. In type III CRISPR systems, Cas10 is a key component of the effector complex and additionally generates cyclic oligoadenylates (cOAs) of different sizes to function as second messengers (**3**). All steps are shown for different examples of type I, type II and type III CRISPR systems, respectively. This figure was adapted from Makarova et al., 2011.*

shown to be essential for interference. While type III systems lack a PAM sequence that help to discriminate between self and non-self, a chromosomal type III CRISPR locus and target DNA or RNA can be distinguished by the presence or absence of a sequence reverse complementary to the 5' repeat fragment. Thus, homology of a potential target nucleic acid with the crRNA tag would lead to base pairing and prevents interference (Marraffini & Sontheimer, 2010; Zhang et al., 2012).

During the interference stage, the full spectrum of the CRISPR mediated immunity becomes apparent (**Figure 4-3**.3). For all different types, the target specificity of the effector complex is dictated by the mature crRNAs resulting in cleavage of the invading DNA or RNA within the spacer-complementary sequence. However, the composition of the effector complexes varies drastically. Besides *Cascade*, Cas3 is an essential component of the multidomain effector complex of type I CRISPR systems. Cas3 harbors an histidine-aspartate (HD) domain and a helicase domain which collaborate in the unwinding and cleavage of a double-stranded DNA (dsDNA) target, respectively (Brouns et al., 2008; Jore et al., 2011). In contrast, type II systems exhibit a single domain effector complex by definition. Here, the multifunctional and tracrRNA activated Cas9 protein provides crRNA-guided cleavage of a dsDNA target (Jinek et al., 2012). Cas9 itself is a multidomain protein containing two different nuclease domains which both are required during the interference (Sapranauskas et al., 2011). As type I systems, also type III systems form multi-protein effector complexes. Type III-A effector complexes are built of six different Csm proteins and were thought to target foreign DNA (Marraffini & Sontheimer, 2008; Rouillon et al., 2013), whereas type III-B effector complexes were reported to contain six or seven different Cmr proteins and degrade RNA (Hale et al., 2009; Zhang et al., 2012). However, this previously assumed subtype-specificity for DNA or RNA targets is now outdated (Tamulaitis et al., 2017). In both subtypes the multidomain-protein Cas10 represents the largest subunit of the ribonucleoprotein effector complex. Interference of these effector complexes was shown to be transcription-dependent (G. W. Goldberg et al., 2014; Samai et al., 2015). Accordingly, target RNA binding triggers sequence-specific RNase activity as well as non-specific ssDNA degradation. DNase activity is provided by the HD nuclease domain of Cas10 which is present in most of the type III variants (Makarova et al., 2015). Cas10 was also reported to contains two small α-helical domains and two Palm domains. The latter include a conserved GGDD motif that was demonstrated to be responsible for metal-dependent conversion of ATP to cyclic oligoadenylates (cOAs) (Kazlauskiene et al., 2017; Niewoehner et al., 2017).

4.2.3 Type III CRISPR systems use second messenger signaling

Nucleotide based second messengers – an evolutionary success story

Cyclic nucleotides are well-known components of immune signaling pathways in eukaryotes as well as in prokaryotes. For example, the human cGAS-STING pathway uses the cyclic dinucleotide cGAMP for signaling in antiviral immunity (**Figure 4-4**.1). The stimulator of interferon genes (STING) is a transmembrane protein conserved among eukaryotes (Kranzusch et al., 2015). The protein links recognition of viral nucleic acids to the initiation of antiviral immune responses (Ishikawa & Barber, 2008; Zhong et al., 2008). However, STING does not sense DNA by itself, but is activated by the second messenger cGAMP (cyclic 2',3'-GMP–AMP) (Burdette et al., 2011). This cyclic

dinucleotide is synthesized by the cyclic GMP–AMP synthase (cGAS) which is activated by sensing and binding to cytosolic DNA (Sun et al., 2013; Wu et al., 2013).

A similar system was discovered in bacteria, here, a phage infection disrupts the inhibition of a bacterial cGAS-like protein by metabolites and cyclic dinucleotides are produced (**Figure 4-4**.2). Beyond 2',3'-cGAMP also the synthesis of other cyclic dinucleotides was observed, for example c-di-GMP, c-di-AMP and 3', 3'-cGAMP.

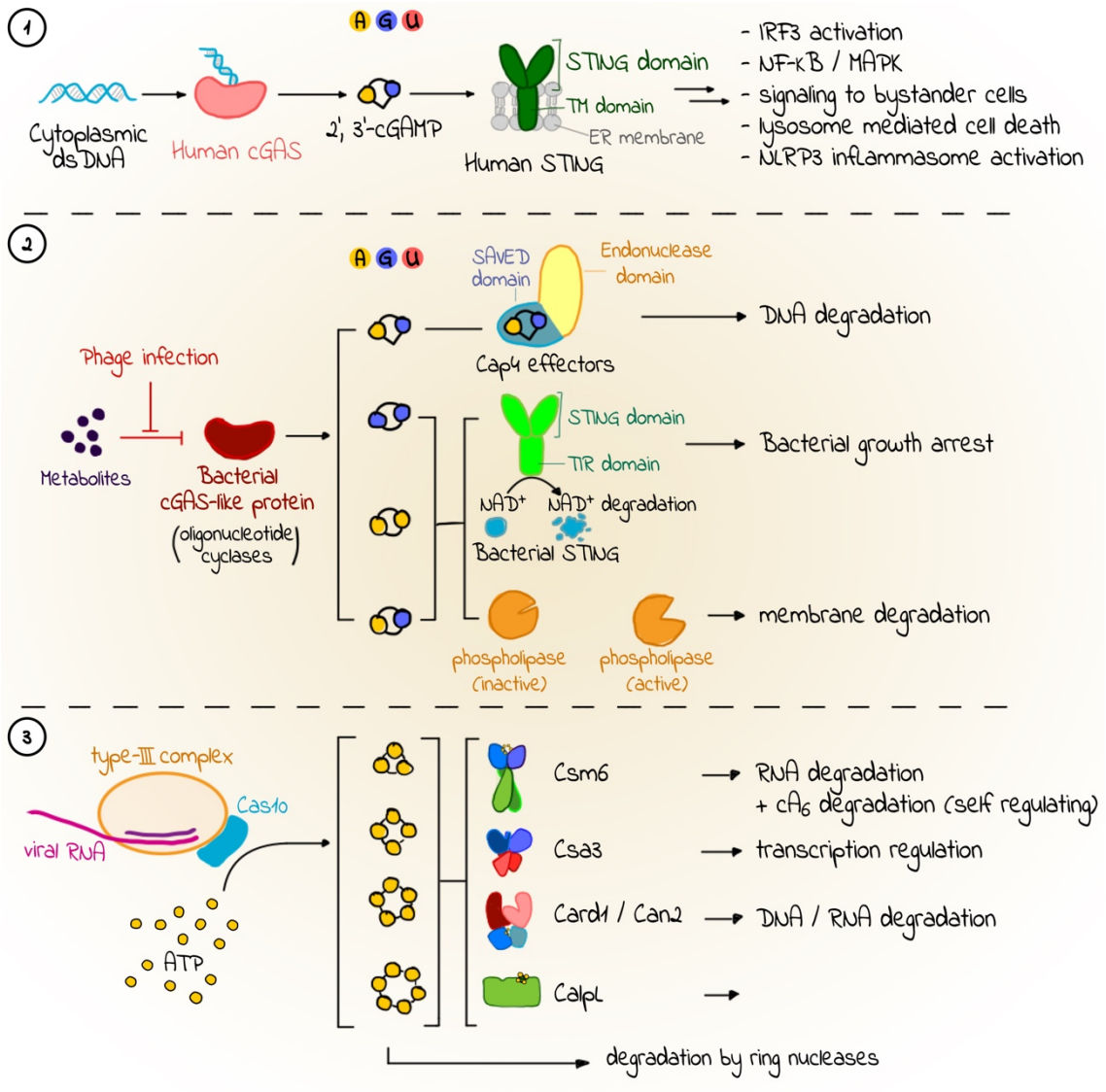


Figure 4-4: Immune signaling by cyclic nucleotides – an overview.
1) *The human cyclic GMP–AMP synthase (cGAS) recognizes DNA ligands and dimerizes to produce the 2', 3'-linked cyclic GMP–AMP (cGAMP) which binds to stimulator of interferon genes (STING). Downstream signaling leads to activation of interferon regulatory factor 3 (IRF3) with subsequent transcriptional regulation, and to the expression of pro-inflammatory cytokines.* ***2)*** *The bacterial cGAS-like protein is inhibited by metabolites in the uninfected state. Upon phage infection this inhibition is relieved and different cyclic dinucleotides are synthesized (2', 3'-cGAMP; c-di-GMP; c-di-AMP; 3', 3'-cGAMP). These activate different effector proteins like the SAVED domain containing endonuclease Cap4, bacterial STING, or even phospholipases resulting in DNA degradation, bacterial growth arrest and membrane degradation, respectively.* ***3)*** *Type III CRISPR effector complexes sense invading nucleic acids which triggers the Cas10 subunit to convert ATP to cyclic oligoadenylates of different sizes (cA_3, cA_4, cA_5, cA_6). These oligonucleotides activate CARF- or SAVED domain containing effector proteins resulting in RNA and/or DNA depletion, transcriptional regulation, or proteolytic activity. The cOA induced stimulus can be downregulated by specific ring nucleases. This figure was assembled with the help of the following reviews: (Athukoralage & White, 2021; Hopfner & Hornung, 2020; N. Liu et al., 2022)*

These bacterial second messengers activate an arsenal of immune effector proteins which often contain a SAVED domain to sense the nucleotides (SAVED is an acronym for: SMODS-associated and fused to various effector domains; with SMODS being an acronym for: second messenger oligonucleotide or dinucleotide synthetase) (Lowey et al., 2020; Morehouse et al., 2020). The immune responses mediated by these effector proteins range from DNA and/or RNA cleavage over bacterial growth arrest to membrane degradation (N. Liu et al., 2022). These bacterial systems that provide immunity against phages, were termed cyclic oligonucleotide-based antiphage signaling systems (CBASS) (Cohen et al., 2019).

In 2017, two independent research groups discovered a cyclic nucleotide based signaling pathway in type III CRISPR systems (Kazlauskiene et al., 2017; Niewoehner et al., 2017) (**Figure 4-4**.3). The conserved GGDD motif of the Palm domain of Cas10 were found to be responsible for the synthesis of cyclic oligoadenylates (coAs) after recognition of target RNA by the interference complex. These cyclic second messengers consist of three to six 3', 5'-linked AMP subunits (cA_3, cA_4, cA_5, cA_6) and its synthesis is deactivated directly after target RNA cleavage (Kazlauskiene et al., 2017; Rouillon et al., 2018). To sense the second messenger, the effector proteins often contain a CRISPR-associated Rossmann fold (CARF) domain, a WYL (for: tryptophan, tyrosine, leucine which are conserved in these domains) domain or a SAVED domain (Athukoralage & White, 2021; Makarova et al., 2014). Second messenger binding then, allosterically activates the effector domain. Within the last years many different cOA-activated effector proteins were described. These cOA activated enzymes range from DNAses and RNAses over proteases to transcription regulating proteins (Charbonneau et al., 2021; Lau et al., 2020; Rostøl et al., 2021; Strecker et al., 2022).

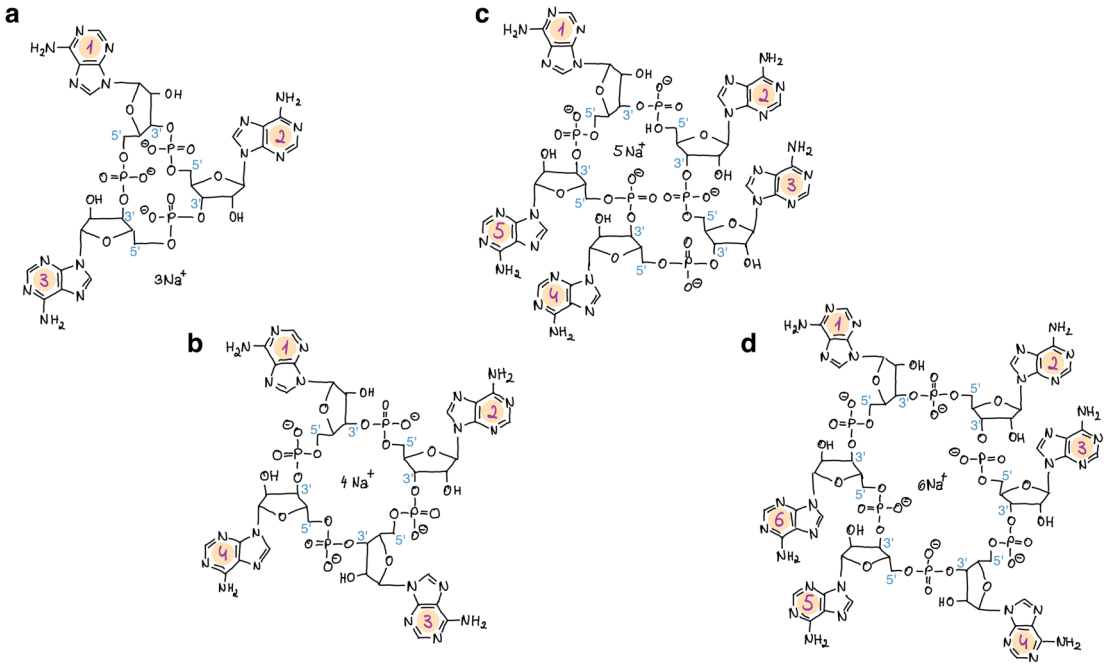


Figure 4-5: Chemical structures of different 3',5'-linked cyclic oligoadenylates.
a) *cyclic triadenylate (cA_3);* ***b)*** *cyclic tetraadenylate (cA_4);* ***c)*** *cyclic pentaadenylate (cA_5);* ***d)*** *cyclic hexaadenylate (cA_6). For all molecules, the 3',5' linkage is indicated and the adenosine subunits are numbered.*

Regulation of cyclic oligoadenylate signaling in type III CRISPR systems

As cOAs activate downstream effector proteins that often degrade nucleic acids nonspecifically, this harbors a high potential for self-destruction. While fighting the viral attack, such mechanisms might slow down phage propagation by setting the cells into a dormancy state. However, an active escape from this cell dormancy is crucial thus, a tight regulation of the cellular level of signaling molecules is essential. The first regulating step is the deactivation of cOA synthesis immediately after complete degradation of viral transcripts, and as a second step an active cleavage of the second messenger occurs (Athukoralage et al., 2018). Elimination of extant cOA ensures a complete escape from the dormant state. This active cleavage was observed to be either catalyzed by standalone ring nucleases or by enzymes with dual activity, acting as effector and ring nuclease simultaneously (Athukoralage et al., 2018; Brown et al., 2020; Garcia-Doval et al., 2020; Jia et al., 2019).

Interestingly, ring nucleases that rapidly degrade cyclic oligoadenylates were identified recently. The viral anti-CRISPR (Acr) ring nuclease, AcrIII-1, was reported to efficiently cleave cA_4 and by this helps the virus to evade the cyclic tetraadenylate-mediated immune response of type III CRISPR systems (Athukoralage et al., 2020). Besides this specific counteraction against cOA-mediated immunity, numerous Acr with different modes of action have been discovered in the last years as reviewed in: (Jia & Patel, 2021; Pawluk et al., 2018). The presence of anti-CRISPR systems shows once more the close co-evolution of bacteria and phages.

4.2.3.1 Cyclic oligonucleotide sensor domains are conserved in CBASS and CRISPR

The sensing of cyclic oligonucleotide based second messengers often employs CRISPR-associated Rossman fold (CARF) or SMODS-associated and fused to various effector domains (SAVED) domains. These key components are conserved among CBASS and CRISPR systems (Makarova et al., 2020). CARF domains are built from six characteristic β-sheets forming a Rossman-like fold surrounded by two α-helices on both sides. The highest sequence conservation is obtained in the two loops located directly after β1 and β4 (**Figure 4-6 a**). These loops have been reported to play key roles in binding and in ribonucleolytic cleavage of the ligand. Typically, homomeric assembly of two CARF domains creates a distinct binding pocket and binding of second messenger leads to a conformational rearrangement of the adjacent fused effector domain, resulting in allosteric activation (Makarova et al., 2014; Xia et al., 2022). As dimerization of two identical CARF domains results in a two-fold symmetry, CARF domain containing effector proteins are well-suited for symmetric signaling molecules, such as cA_4 or cA_6 which are most common in type III CRISPR systems. However, in CBASS signaling pathways most cGAS/DncV-like nucleotidyltransferase (CD-NTase) enzymes produce asymmetric second messengers (Whiteley et al., 2019). The fusion of two CARF-like subunits into a single SAVED domain enables the sensor domain to recognize an expanded range of nucleotide signals. Due to this, CBASS immunity can employ different nucleobases and linkages as well as alternative ring sizes. **Figure 4-6 b** shows a comparison of structural core elements of CARF and SAVED, for example, a β-sheet surrounding α-helix (**Figure 4-6 b**, α3 in CARF corresponds to α2, α6 in SAVED) and two central helices in SAVED (α3, α7) which correspond to the central helix at the dimerization interface of CARF (α4) (Lowey et al., 2020). Despite some structural and

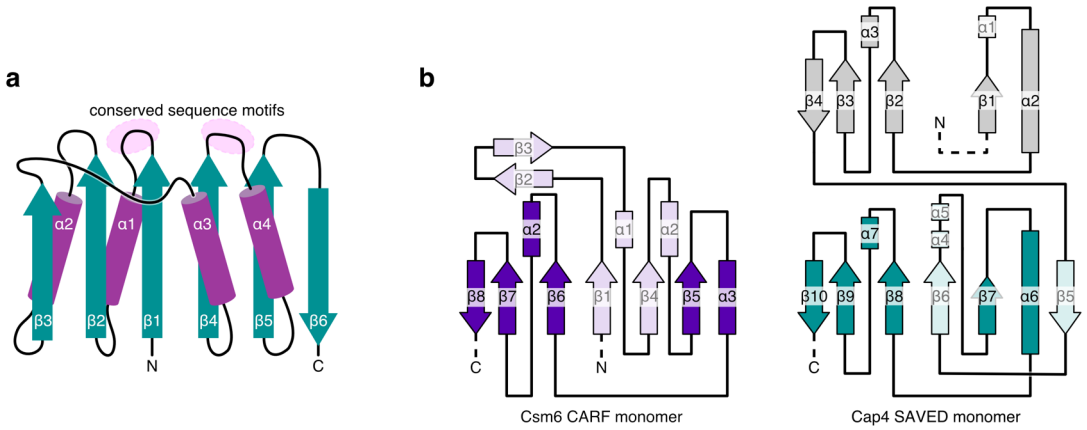


Figure 4-6: Schematic representation of the architecture of CARF- and SAVED domains. ***a)*** *The conserved core of a CARF domain consists of six β-sheets surrounded by two α-helices on each side. The conserved sequences in the loops after β1 and β4 are involved in binding and degradation of cOAs. This figure was adapted from (Makarova et al., 2020).* ***b)*** *Comparison of the topology diagrams of a Csm6 CARF domain (left) and of the Cap4 SAVED domain (right). Similar secondary structure elements are highlighted. This figure was adapted from (Lowey et al., 2020).*

mechanistical similarities, CARF and SAVED differ in the way of activating the fused effector domains. While activation of CARF domain containing effector proteins was reported to be a consequence of second messenger induced significant conformational changes, such structural rearrangements could not be observed for SAVED domain containing effector proteins. Rather, it was observed that the latter oligomerize upon specific binding of the signaling molecule, and this oligomerization somehow activates the effector domain (Lowey et al., 2020; Makarova et al., 2020). Within the last few years scientists found out that SAVED domains do not only occur in CBASS but also in CRISPR systems sharing the same activation mechanism and partially using asymmetric cOAs (Hogrel et al., 2022; Rouillon*, Schneberger* et al., 2023; Steens et al., 2023).

4.2.4 Identification of previously unknown mechanisms in CRISPR immunity

The success of utilizing Cas proteins as genome and transcriptome editing tools has certainly contributed to the increasing popularity of CRISPR immunity over the past few decades. However, Cas9, Cas12 and Cas13 represent only a fraction of the diverse mechanisms associated with clustered regularly interspaced short palindromic repeat sequences. In recent years, a plethora of additional mechanisms and components contributing to CRISPR-mediated antiviral responses have been uncovered, demonstrating the richness and complexity of these mechanisms.

4.2.4.1 Systematic prediction of genes that are likely to be linked with CRISPR-Cas

Several groups have developed specialized pipelines for mining of new CRISPR systems, or genetic components that could be related to CRISPR mediated immune responses. Therefore, protein encoding genes in the vicinity of CRISPR arrays from sequenced bacterial and archaeal genomes were scored according to several parameters which describe the strength of CRISPR association. There is now a large number of predictions of genes that are functionally related to CRISPR-Cas systems. One of these studies specifically searched for membrane-associated proteins and found several strong candidates, most of which were located in type III CRISPR loci (S. A. Shmakov et al., 2018). One of these candidates particularly caught our attention, a CRISPR-associated Lon protease (CalpL, originally named Lon-CARF) from *Sulfurihydrogenibium* sp.

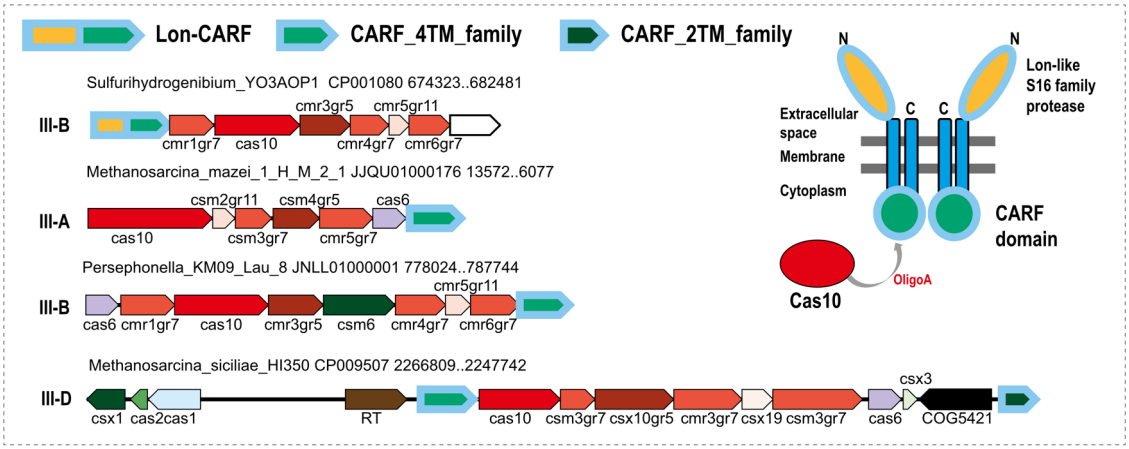


Figure 4-7: CRISPR loci of type III systems containing predicted CRISPR-associated genes. *Predicted CRISPR-linked and conserved genes that encode potential CARF domain containing membrane proteins. A hypothetical domain organization of the protein from* Sulfurihydrogenibium sp Y03A0P1 *is shown on the right-hand side. The predicted protein contains an N-terminal Lon-like protease domain, two transmembrane helices and a cOA sensing CARF domain facing the cytoplasm. This figure is an excerpt from (S. A. Shmakov et al., 2018)*

YO3AOP1 (**Figure 4-7**). This protein was listed in a group of type III-associated proteins that contain a CARF domain and two predicted transmembrane helices, for some of them an additional Lon-like protease domain was identified. This is also the case for CalpL, in which the N-terminal Lon-like protease domain was predicted to face the extracellular space and a cOA-sensing CARF domain protrudes into the cytoplasm. As common for CARF domain containing effector proteins, CalpL most likely forms a dimer and is activated by cytosolic cOAs produces by Cas10. However, a detailed mechanism and the role of an extracellular protease remains unclear (S. A. Shmakov et al., 2018).

4.3 Lon proteases are conserved elements to react on cellular stress

While most characterized CRISPR systems use nucleases to degrade foreign genetic material, a proteolytic enzyme would be an intriguing new element in the vast array of type III-mediated immunity. Lon proteases are common components of various regulatory pathways in prokaryotes as well as in eukaryotes. In bacteria, the variety of functions include general stress response, regulation of toxin–antitoxin systems, degradation of damaged proteins and replication and repair of DNA. Moreover, Lon proteases have been reported to be essential virulence factors in pathogenic bacteria (Kirthika et al., 2023; Tsilibaris et al., 2006). In CRISPR-mediated immune responses, Lon proteases may be particularly involved in stress response and transcriptional regulation.

4.3.1 Lon proteases are serine proteases with broad substrate specificity

Most Lon proteases were found to form homohexameric assemblies with the proteolytic active Ser-Lys catalytic dyad facing towards an internal degradation chamber (Amerik et al., 1991; Kirthika et al., 2023). Although substrate recognition by these proteolytically active enzymes is not fully understood, no specific sequence motif appears to be recognized. Rather, substrate specificity appears to be achieved structurally, for example, the N-terminal domain of Lon proteases is involved in substrate recognition (Tsilibaris et al., 2006). While the barrel-shaped hexameric structure of Lon prevents most

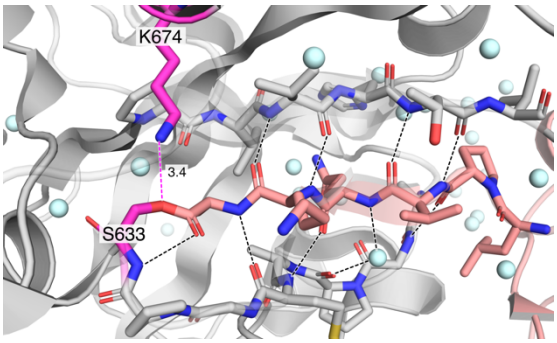


Figure 4-8: The active site of a Lon protease. *Acyl-enzyme complex of a viral Lon protease (gray) and its substrate (salmon) (PDB-ID: 4izj). The catalytic residues are shown in magenta and the distance of K674 (magenta dashed line) is given in Å. Substrate and enzyme built a β-sheet, their peptide backbones form an extensive network of hydrogen bonds (black dashed lines).*

folded proteins from degradation (Kirthika et al., 2023), in the viral Lon protease, VP4, the target peptide is stabilized within the binding groove by forming a β-sheet with the adjacent loops (**Figure 4-8**) (Chung & Paetzel, 2013). Another interesting feature is the interplay of all different domains in substrate recognition, providing a highly organism specific substrate recognition (He et al., 2018). Proteolytic activity of the enzyme often requires multiple steps to prepare the substrate for presentation the protease active site. However, experiments with isolated protease domains showed that these were capable of degrading unstructured peptides. In these cases, the catalytic dyad residues were found to be located within a hydrogen bond distance in between Ser-Oγ and Lys-Nζ, thus a distinct three-dimensional arrangement of these residues seems to be important (Wlodawer et al., 2022). For the proteolytic cleavage, the active site lysine acts as general base and the catalytic serine nucleophilically attacks the carboxyl carbon (Chung & Paetzel, 2013). In addition to the nucleophilic serine, a conserved threonine or serine is commonly observed in the vicinity of the general base and might contribute in the not yet fully understood proteolytic cleavage mechanism (Chung & Paetzel, 2013; Paetzel et al., 2002).

4.4 Bacterial regulation of transcriptional initiation incorporates sigma factors

To understand the role of the CRISPR-associated Lon protease from *Sulfurihydrogenibium* sp. YO3AOP1 in an antiviral immune response, we took advantage of the fact that functionally linked prokaryotic genes are often organized in operons (Rogozin et al., 2002). We therefore searched for conserved genes within the direct neighborhood of *calpL* employing the webFlaGs online tool (Saha et al., 2020). Besides genes as *cas1*, *cas2*, *cas6* and *cas10* which directly prove the association to type III CRISPR systems, three highly conserved open reading frames (ORFs) upstream of *calpL* could be identified (**Figure 4-9**). Further investigation of the proteins encoded by two of these ORFs lead us to name them CalpT (red, T for target) and CalpS (purple, S for sigma factor). The third conserved gene (gray) was not yet characterized. Both the sequence and the predicted structures show parallels to transcription-regulating σ factor–anti-σ factor systems, led us to investigate these in more detail.

Sigma factors (σ factors) are multi-domain proteins that are essential components of bacterial RNA polymerase (RNAP) complexes (Burgess et al., 1969). The core complex of the bacterial RNAP is built of two α-subunits, one β-subunit, one β'-subunit and one ω-subunit, the resulting complex (~380 kDa) is catalytically active and able to non-specifically bind DNA. However, for specific recognition of distinct promoters and subsequent transcriptional initiation, the core RNAP relies on dissociable σ factors. Besides their role in identification of promoter regions, σ factors are important stabilizing elements for the separated DNA strands around the transcription site and stimulate early steps in the RNA synthesis (Paget & Helmann, 2003). The proteome of most bacteria contains multiple sigma factors with different DNA binding specificities and different affinities for RNAP. Thus, exchange of the sigma factor directly affects the gene expression of RNAP (Losick & Pero, 1981). Tight regulation of the availability of distinct σ factors by complex regulatory pathways, enables a highly efficient transcriptional

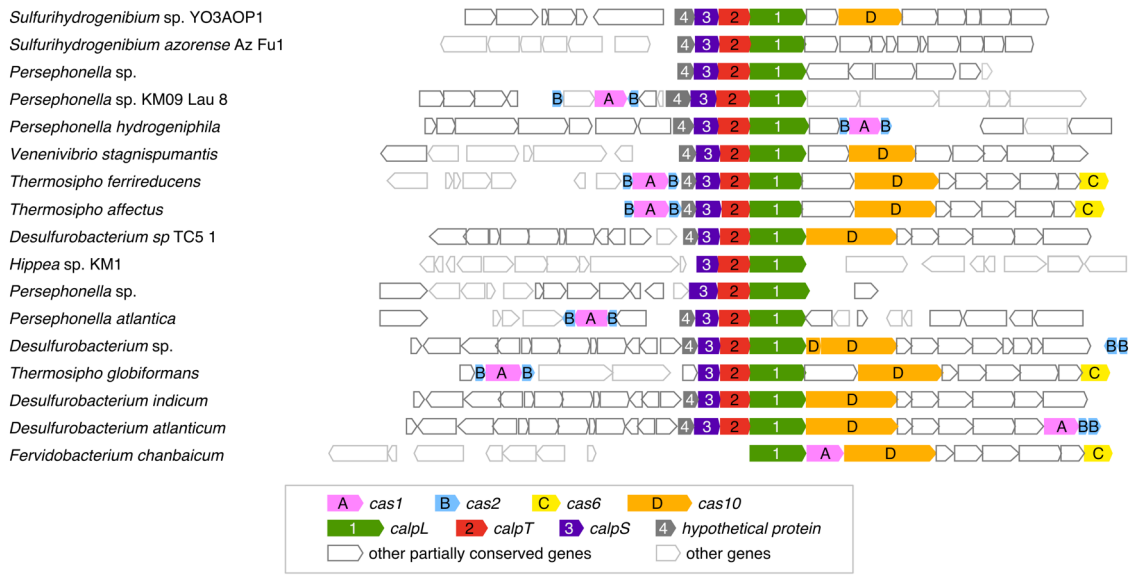


Figure 4-9: Analysis of the gene neighborhood of calpL reveals functionally linked genes.
The analysis of the conservation of the genomic context of calpL was done using the webFlaGs online tool (Saha et al., 2020). The original organism is shown on top and the CalpL encoding gene is shown in green (1). Three conserved open reading frames (ORFs) were identified directly upstream of calpL indicated in red (2), purple (3) and gray (4). Additionally, conserved genes among type III CRISPR systems were highlighted in magenta (cas1), blue (cas2), yellow (cas6) and orange (cas10).

response on environmental changes. Sigma factors can be classified into two structurally unrelated families, the enhancer and ATP dependent σ^{54} sigma factors are often employed in the response to environmental signals. The more abundant σ^{70} family regulate the transcription of housekeeping genes and are also employed in various stress response mechanisms (Gruber & Gross, 2003; Treviño-Quintanilla et al., 2013).

4.4.1 Classification of σ^{70} sigma factors

Members of the σ^{70} family share two domains with high structural similarity and sequence conservation. These two domains, σ_2 and σ_4, specifically interact with distinct promoter elements centered -10 bp and -35 bp upstream of the transcription start site, respectively. Furthermore, these domains mediate most of the σ factor–RNAP interactions (Murakami et al., 2002). In addition to their specialized functions, the individual groups 1 to 4 differ by the presence or absence of additional regions, that are called $\sigma_{1.1}$, $\sigma_{1.2}$ (which is part of the σ_2 region) and σ_3. The largest members of the σ^{70} family are found in group 1, which consists of the essential primary sigma factors and often contain an additional non-conserved region (NCR) in between $\sigma_{1.2}$ and $\sigma_{2.1}$ (Paget, 2015) (**Figure 4-10**). Furthermore, members of group 1 are the only sigma factors that contain an N-terminal $\sigma_{1.1}$ domain which is involved in autoinhibition (Gruber & Gross, 2003). The optional $\sigma_{1.2}$ domain is an important factor in the response to changes in the nutritional environment by interacting with the so-called discriminator element of the promoter (Haugen et al., 2006). The presence of a σ_3 domain can stabilize transcription

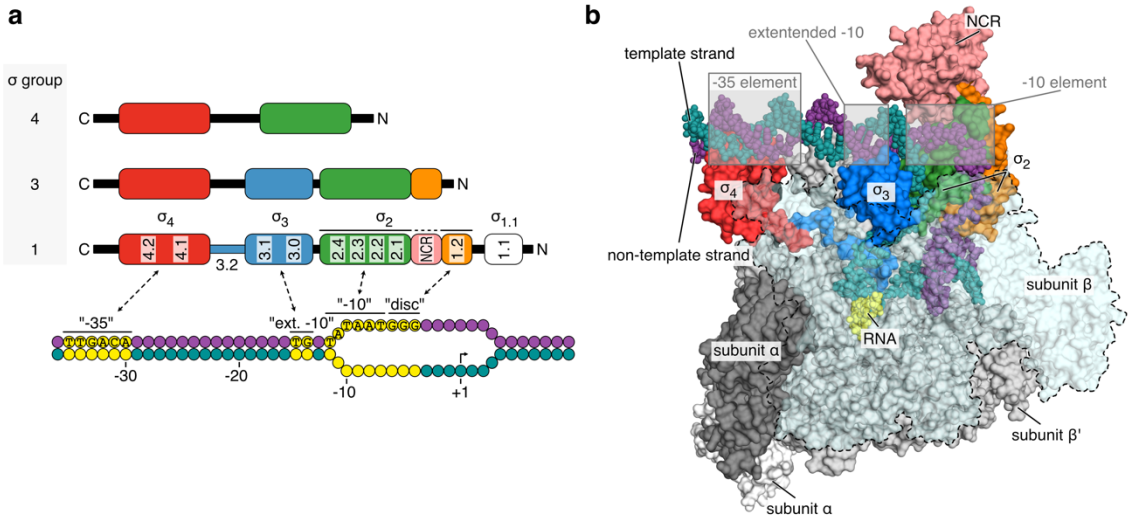


Figure 4-10: Domain architecture and structural arrangement of bacterial sigma factors.
a) *The domain architecture of σ^{70} sigma factors contains the conserved σ_2 (green) and σ_4 (red) domains. Members of group 3 contain additional σ_3 and $\sigma_{1.2}$ domains (blue and orange, respectively) that mediate interactions with the extended -10 ("ext. -10") region and the discriminator element ("disc") of the promoter, respectively. In contrast to the alternate sigma factors of groups 2-4, the housekeeping sigma factor (group 1) include an additional non-conserved region (NCR, salmon) in between the $\sigma_{1.2}$ and $\sigma_{2.1}$ domains. The interactions of the individual domains with a consensus* E. coli *promoter DNA are illustrated below. The non-template strand is depicted in purple and the template strand is colored in teal. The conserved sequence elements "-35", extended -10, "-10" and discriminator are colored in yellow.* ***b)*** *Transcription initiation complex of* E. coli *RNA polymerase (different shades of gray) bound to a group 1 σ^{70} sigma factor (colored according to a)) and target DNA (purple/teal). The individual subunits of RNA polymerase and of the sigma factor are labeled in the figure and RNAP subunit β is shown transparent with a dashed outline to provide a better view. The different elements of the DNA (-35 element, extended -10, -10 element) as well as the template and non-template strand are indicated (PDB ID: 4yln). This figure was adapted from (Paget, 2015).*

initiation complexes by interacting with the extended -10 region of the promoter, rendering the otherwise critical presence of the -35 element superfluous (Mitchell et al., 2003). The members of the groups 2 to 4 are also referred to as alternative sigma factors, completely lack the $\sigma_{1.1}$ and NCR domains and their range of functions varies enormously. Group 2 sigma factors are frequently involved in the cellular adaption to stress caused by extreme temperatures or nutrient limitation (Paget, 2015). The influence of these sigma factors on transcription is directly connected to their growth-rate dependent cellular expression level (Ihssen & Egli, 2004). Members of group 3 are structurally more divergent from group 1 sigma factors and often regulate gene expression correlated with heat shock, flagellar biosynthesis or sporulation (Gruber & Gross, 2003). The numerically largest group of sigma factors is also called the extracytoplasmic function (ECF) group and consists of more than 40 phylogenetically distinct subgroups that fulfill important roles in bacterial signal transduction (Staroń et al., 2009).

4.4.1.1 ECF sigma factors

ECF sigma factors only contain the conserved σ_2 and σ_4 domains and their encoding genes are often found in close proximity to their regulons (Paget, 2015). Their name was derived from the fact that initially found members of this group were associated with important roles in the response to extracytoplasmic stimuli (Lonetto et al., 1994). Promoter sequences that are recognized by this group of proteins often include a conserved -35 “AAC” motif while the -10 element is more diverged (Staroń et al., 2009). Structural and biochemical analysis of an ECF sigma factor from *E. coli* with its -10 element DNA sequence revealed that the specific recognition is mediated by a flexible loop of the σ_2 sub-domain. However, alterations in this loop completely altered the promoter specificity. Moreover, it was found that, unlike the melting of dsDNA by group 1 sigmas, only a single base at position -10 upstream of the transcription start site was flipped out to interact with multiple amino acid residues of the loop. These findings suggest that a finely tuned interplay of the -10 sequence and the σ_2 domain (or more specifically the variable loop in region 2.3 of this domain) is responsible for a highly specific detection (Campagne et al., 2014). To ensure that these transcriptional regulatory proteins are activated only when a specific signal is detected, tight control mechanisms are required that can respond quickly to this very signal.

4.4.1.2 Regulation of ECF sigma factors by anti-sigma proteins

Regulation of group 4 sigma factors is frequently achieved by co-transcription with anti-sigma proteins that bind to the sigma factor to inhibit its binding to the core RNAP. Thus, the specific regulons of a sigma factor are silenced until signal induced dissociation of its anti-σ partner. Anti-σ factors can either be soluble or anchored to the inner membrane via their variable C-terminal domain. In contrast to the C-terminal domain the N-terminal domain, that is also called anti-sigma domain (ASD), exhibits a certain structural conservation and is responsible for the binding of the sigma factor (Campbell et al., 2007; Treviño-Quintanilla et al., 2013). Binding to the sigma factor is often achieved by inserting a bundle of helices in between the σ_2 and σ_4 domains. Thereby, the sigma factor is kept in an inactive conformation and important regions for interactions with both the RNAP and the promoter are blocked (Campbell et al., 2003, 2007). Alternatively, a different group of anti-σ factors use parts of their ASD to wrap around

the σ_2 and σ_4 domains to inhibit binding of the sigma factor to RNAP and the promoter DNA (Maillard et al., 2014). To initiate transcription in response to a cellular signal, the inhibited σ factor needs to be released from its cognate anti-σ protein. There exist numerous different mechanisms to accomplish sigma factor release of which some are poorly understood (Hughes & Mathee, 1998; Treviño-Quintanilla et al., 2013). However, the most frequently used mechanisms can be classified into direct sensing, partner-switching and regulated proteolysis. In some σ/anti-σ factor complexes, the coordination of Zn^{2+} by the ASD is essential for complex formation. Upon direct sensing of oxidative stress, the Zn^{2+} coordinating region undergoes changes to repel the metal ion which causes dissociation of the σ factor/anti-σ factor complex (Campbell et al., 2007). On the other hand, partner-switching mechanisms utilize an external sensing system that often regulates a phosphatase. If this enzyme is activated, an anti-anti-σ factor is dephosphorylated to enable complex formation with the anti-σ protein which in turn releases the sigma factor (Hecker et al., 2007). Mechanisms based on regulated proteolysis also involve external sensors which regulate the activity of proteases. After proteolytic cleavage of the anti-σ factor by a serine protease the complex of sigma factor and the ASD domain of the sigma factor is released into the cytoplasm (Ades et al., 1999). This complex still prevents transcription initiation, however, the N-terminal anti-σ domain can be rapidly degraded by ClpX proteases to release the fully active sigma factor. These ClpX protease systems often depend on specific adaptors that act as recognition motif for effective targeted degradation (Flynn et al., 2004). As also both proteolytically active enzymes are regulated by different mechanisms, two independent signals are necessary to release the transcription-regulating sigma factor. Such a two-factor authentication ensures a tightly regulated response on specific signals which are robust to fluctuations to a certain extend.

5. Results Part II

This chapter focuses on the experimental findings related to CalpL, a Lon protease activated by cyclic oligoadenylates. At the beginning of our investigation, CalpL was a relatively enigmatic protein with limited information available. The basis of our assumptions largely relied on computational models, especially those put forth by Shmakov et al. in 2017. After a short time, we were puzzled by the consistent detection of a predicted membrane protein in the soluble fraction after lysis. To unravel this unexpected phenomenon and gain mechanistic insights, we aimed for structural elucidation. However, the journey proved to be full of unforeseen discoveries and challenges. Throughout this research, excellent collaborations helped shed light on this story. SAXS data acquisition and analysis was done in cooperation with S. Da Vela and D. Svergun (EMBL, Hamburg), EPR spectroscopy was performed by K. Ackermann and B. E. Bode (University of St Andrews, Scotland). K. Blumenstock and J. Schmid-Burgk (University of Bonn, Bonn) as well as H. Chi and M. White (University of St. Andrews, Scotland) helped to investigate the role of CalpT. Parts of our scientific findings have been published (Rouillon*, Schneberger* et al. 2023). I would also like to thank Stella Arau Jakubzik. During her master thesis from 2022 to 2023, she contributed to answer some of the many questions that arose. Last but not least, I would like to thank Christoph Winterberg for the buffer receipt of 'Buffer W', which was crucial for the purification of CalpT.

5.1 CalpL is a cyclic nucleotide activated Lon protease

Prior to this work, a codon-optimized variant of the CalpL-encoding gene (*calpL*) from *Sulfurihydrogenibium* sp. YO3AOP1 (UniProt identifier B2V8L9) was cloned into the IPTG inducible pET-11a vector. The construct was designed to contain an N-terminal His_{10} tag and a TEV protease cleavage sequence. Recombinant protein expression was done in *E. coli* BL21 (DE3) cells and reliable protein yields were observed.

5.1.1 Preparations for successful structure elucidation

5.1.1.1 Optimization of purification procedure

Although large amounts of protein could be expressed and extracted from the cell lysate by Ni^{2+} affinity chromatography, subsequent gel filtration could not provide an adequate degree of purity. Before starting the crystallization of CalpL, the purification process was optimized. During the first purification trials, the cell lysate was loaded to a HisTrap affinity chromatography column (**Figure 5-1 a**, **e** 1) but over time Ni^{2+}-NTA beads on a gravity flow column turned out to provide higher purity. The protein eluted as a single peak from size exclusion chromatography, with a tiny shoulder towards lower retention volumes (**Figure 5-1 b**, **e** 2). A comparison of the main peak retention volume with molecular weight references suggested a monomeric protein. With the help of the ProtParam online tool (Gasteiger et al., 2005), the theoretical pI value of CalpL was computed to pI=6.16. Thus, at pH 8.0 the polypeptide should carry a negative net charge and be suited for anion-exchange chromatography (AIEX). This method was tested as the next purification step and a significant decrease of impurities could be observed in the following SDS-PAGE analysis (**Figure 5-1 c**, **e** 3). After this, the decahistidine tag was cleaved off using a His_6 tagged TEV protease construct (**Figure 5-1 d**, **e** 4). CalpL was separated from TEV protease, non-cleaved protein and the affinity tag by another Ni^{2+}

affinity chromatography. This optimized procedure yielded protein of high purity that was used for biochemical and structural investigation.

To verify a successful His tag removal and demonstrate that the purified protein is indeed CalpL we did a Western Blot experiment (**Figure 5-2**). After binding of a horseradish peroxidase (HRP) conjugated secondary antibody to a His tag specific primary one, the blot was developed with a commercially-available enhanced chemiluminescence kit. For the His tagged protein sample, the immunoblot showed an intense band slightly larger than 55 kDa (**Figure 5-2 b** +). A much weaker, but yet distinct, signal at the same position was observed for the sample without His tag (**Figure 5-2 b** -). However, the significant intensity difference indicates that the main

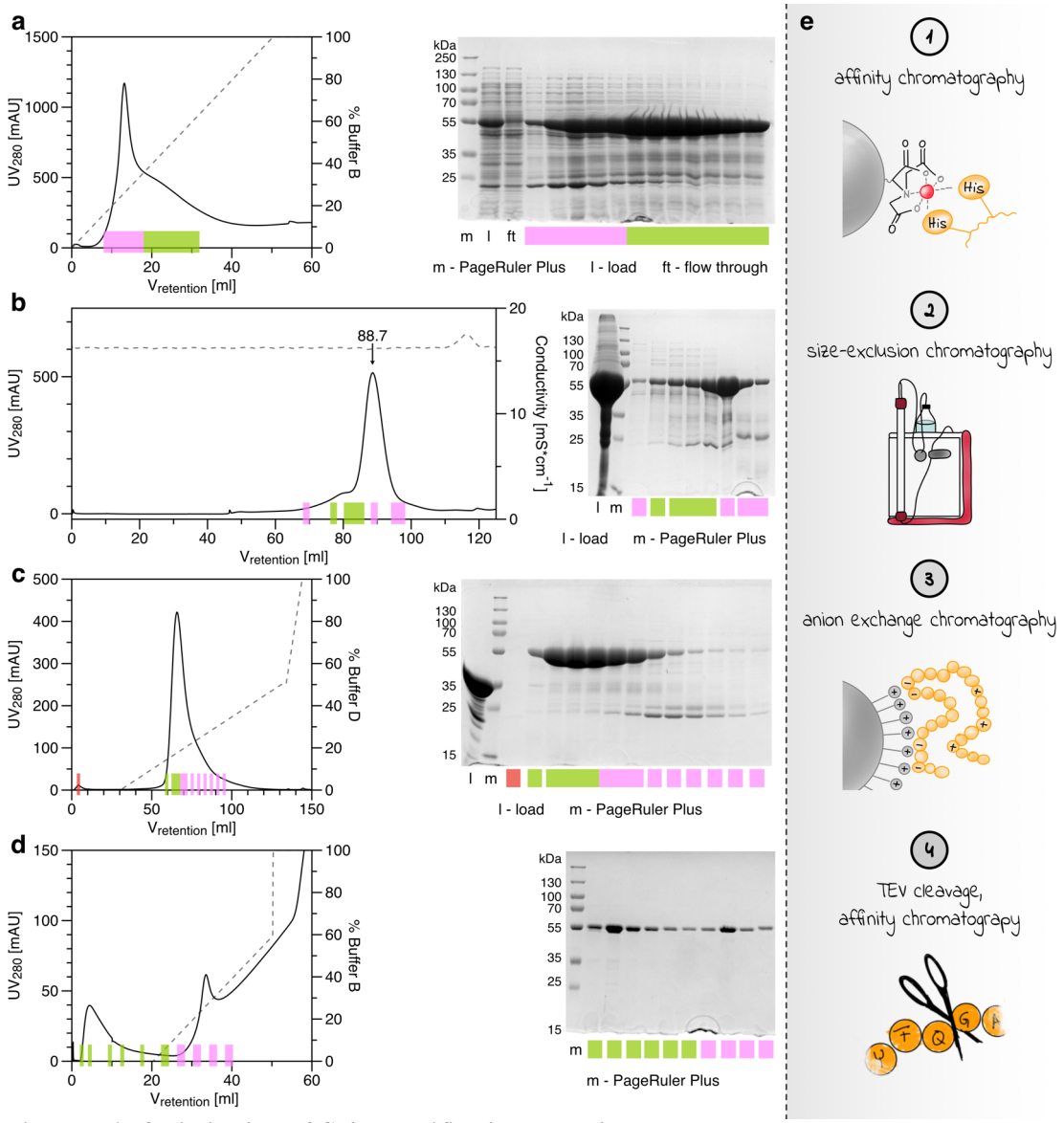


Figure 5-1: Optimization of CalpL purification procedure.
a) *UV_{280} chromatogram from Ni^{+} affinity chromatography (left) and the according SDS-PAGE analysis (right). Fractions that have been pooled and concentrated for further purifcation steps are indicated by a green bar and those that were discarded by a magenta bar.* ***b)-d)*** *Same depictions as in a) but showing results from preparative size exclusion chromatography (b), Anion-exchange chromatography (AIEX) (c) and Ni^{2+} affinity chromatography after TEV cleavage (d). All graphs show the retention volume on the x-axis and the UV_{280} absorbance (in mAU) on the y-axis. For methods using gradient elution, an additional y-axis on the right-hand side depicts the percentage of elution buffer. The value above the gel filtration peak annotates the exact retention volume at the maximum height.* ***e)*** *Schematic representation of the optimized purification process including affinity chromatography (1), SEC (2), AIEX (3) and tag removal by TEV protease cleavage (4).*

proportion of protein within the sample could not be bound by the primary antibody. **Figure 5-2 a** shows exactly the same samples but on a Coomassie stained polyacrylamide gel demonstrating the high sensitivity of Western Blot analyses. All protein bands run at the same height in the blot and the gel and show no observable impurities. A comparison of both images indicates that the extreme signal difference in the blot was exclusively caused by the presence of a low amount of tagged protein in the non-tagged (-) sample but not a difference in the amount of loaded protein.

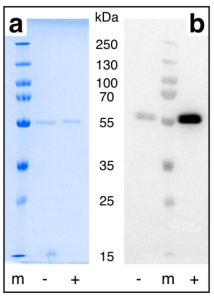


Figure 5-2: Western Blot analysis of CalpL. ***a)*** *Coomassie stained polyacrylamide gel showing CalpL with (+) and without (-) His$_{10}$ tag.* ***b)*** *Western Blot with primary anti-His antibody of the same samples shown in a).*

With the certainty of having the right protein in satisfactory purity, crystallization trials were started. First attempts using commercially available crystal screens already yielded crystals, and diffraction experiments utilizing these crystals indicated that they consist of protein, but did not reach sufficient quality standards. After several rounds of optimization, a self-designed screen based on JCSG+ condition D7 (0.2 M $LiSO_4$, 0.1 M Tris, pH 8.5, 40 % PEG400) led to the reliable formation of large and well-diffracting crystals.

5.1.2 Structural investigation of CalpL, a soluble monomeric protein

5.1.2.1 Experimental phasing by single wavelength anomalous dispersion (SAD)

To solve the crystallographic phase problem, most often molecular replacement is used. While nowadays it is relatively unlikely that an adequate model for molecular replacement can neither be found in the protein data bank (PDB) nor generated by AlphaFold (Jumper et al., 2021), we were faced with exactly this problem in the pre-AlphaFold days. To overcome this, we decided to try experimental phasing which requires an anomalous scatterer within the protein crystal of interest. Therefore, a selenomethionine (SeMet) derivative of CalpL was expressed and purified according to the optimized purification protocol, illustrated in **Figure 5-1**. SeMet CalpL showed a similar behavior as the native protein during affinity chromatography and gel filtration (**Figure 5-3 a**). After SEC, the purity could be increased by AIEX chromatography (not shown) and subsequent tag removal (**Figure 5-3 b**). Albeit TEV protease cleavage was only partially successful, the cleavage yielded enough protein for crystallization trials at 23.1 mg/ml (≙400 µM). In the previously mentioned optimization screen, crystal growth was obtained after 3 days of incubation at 20 °C in 0.1 M Tris, pH 8.0, 33.68 % PEG400, 0.11 M Li_2SO_4 (**Figure 5-3 c** i). Crystals were harvested with additional 15 % PEG400 for cryo protection and flash frozen to liquid nitrogen (**Figure 5-3 c** ii). X-ray diffraction datasets were recorded at the P13 beamline of the Petra III storage ring (DESY) in Hamburg with resolutions up to 2.1 Å for the SeMet derivative (**Figure 5-3 c** iii-iv) and 1.9 Å for native CalpL. For the experiment with SeMet crystals, the X-ray wavelength was set to λ=0.9795 Å. The resulting energy (12.6578 keV) corresponds to the absorption edge of selenium (Rose & Wang, 2016) which could be nicely visualized by an energy scan (**Figure 5-3 c** v). This experiment simultaneously proved the presence of selenomethionine within the crystals. The structure was solved by single wavelength

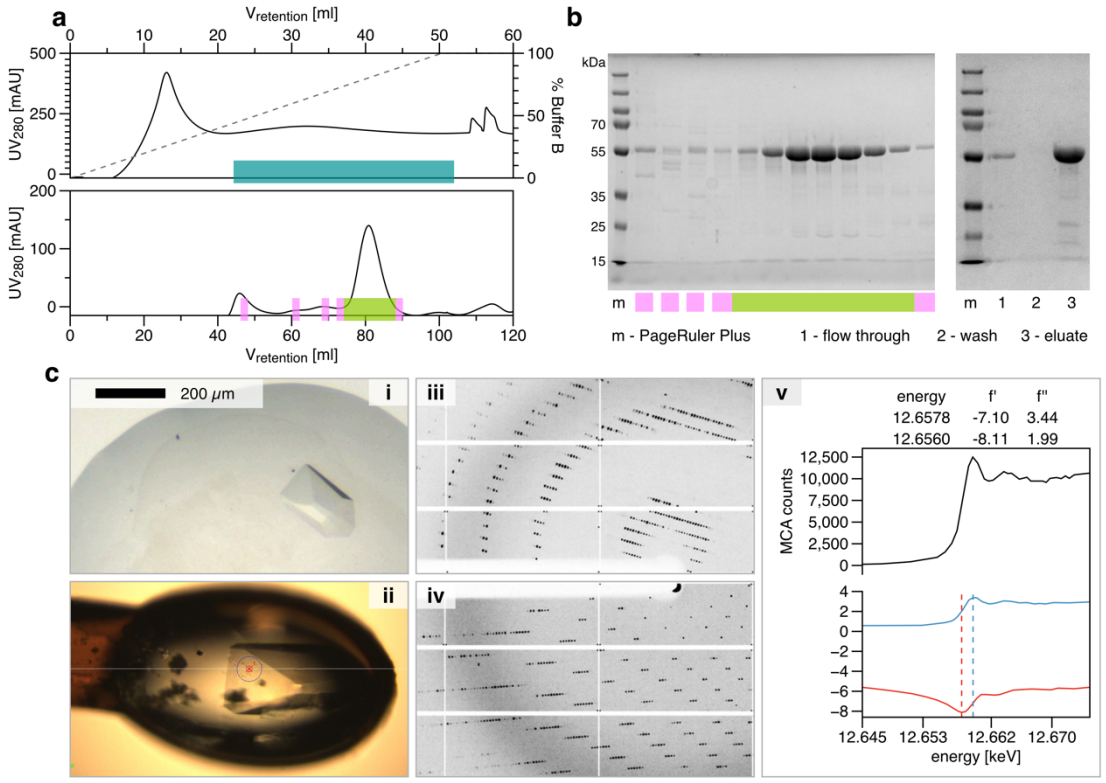


Figure 5-3: Crystallization of a selenomethionine CalpL derivative.
a) *Chromatograms from Ni^{2+} affinity chromatography (top) and size exclusion chromatography (bottom), respectively. The retention volumes are shown on the x-axis and the UV absorbance on the left y-axes, independently for each graph. The percentage of elution buffer is shown on the right-handed y-axis of the affinity chromatogram.* ***b)*** *SDS-PAGE analysis with samples from SEC (left) and affinity chromatography after TEV cleavage (right). The samples from SEC are indicated by colored bars according to the SEC chromatogram in a). Samples from IMAC are annotated below the gel.* ***c)*** *Images of the CalpL SeMet derivative crystal that led to structure determination before* ***(i)*** *and after harvesting* ***(ii)****. Diffraction images from the shown crystal obtained at different rotation angles* ***(iii, iv)****.* ***v)*** *Energy scan to visualize the absorption edge of selenium and identify the optimal X-ray energy for the data collection.*

anomalous dispersion phasing using the AutoSol wizard of the PHENIX software package (Zwart et al., 2008). Thereby, the location of anomalous scatterer atoms, phasing and initial model building were done automatically. Using Coot and PHENIX.refine, the structural model was subsequently optimized and refined to R_{work} and R_{free} values of 19.3 and 22.5, respectively (Afonine et al., 2012; V. B. Chen et al., 2010; Emsley & Cowtan, 2004). CalpL consists of three domains, an N-terminal domain (purple), a Lon protease domain with a catalytic dyad formed of S152 and K193 (green, active site positions are highlighted in magenta) and a C-terminal SAVED domain which forms a large cOA binding cavity (**Figure 5-4 a**). Overall, CalpL is an elongated, banana-shaped protein of roughly 100 Å in length and 35 Å in diameter. According to calculations by the APBS electrostatics plugin for PyMol (Jurrus et al., 2018), the main area of the surface exhibits a rather high partial charge, suggesting a highly soluble protein. While most sections show a negative partial charge, two positively charged areas of the SAVED domain stand out. On the upside, a ~22*22 $Å^2$ large and roughly 6 Å deep positively charged cavity and a positively charged patch of similar size but without deepening on the downside, directly opposite. The concave area on the upside fitted a cA_n molecule with $3 \leq n \leq 5$ in molecular dynamics simulations quite well (data not shown). Another positively charged area can be found on the tip of the N-terminal part of the protease (**Figure 5-4 b**). Notably, at all

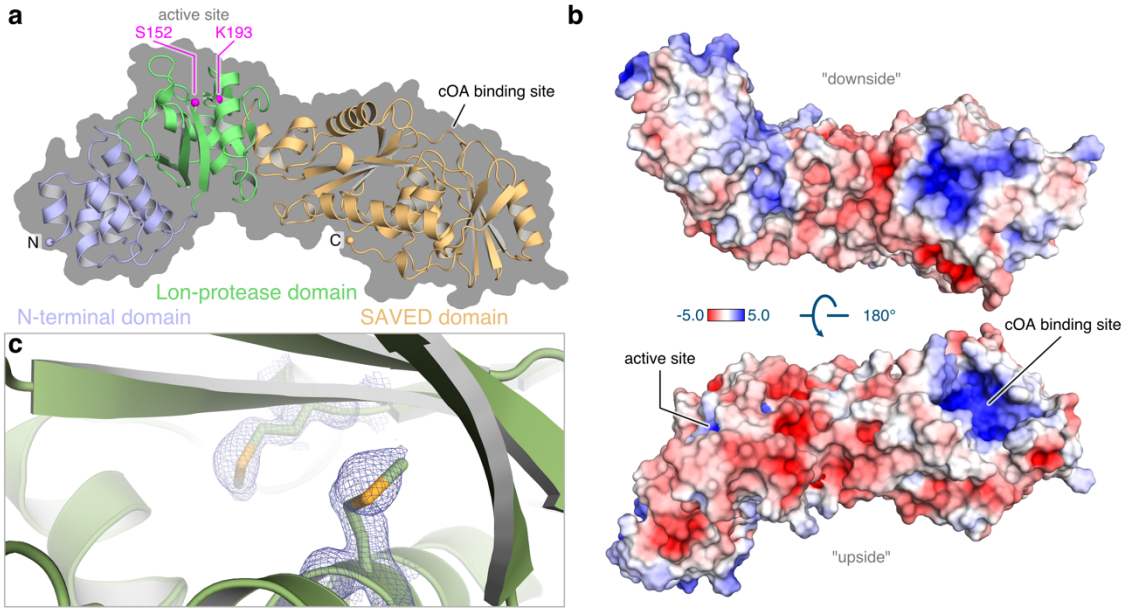


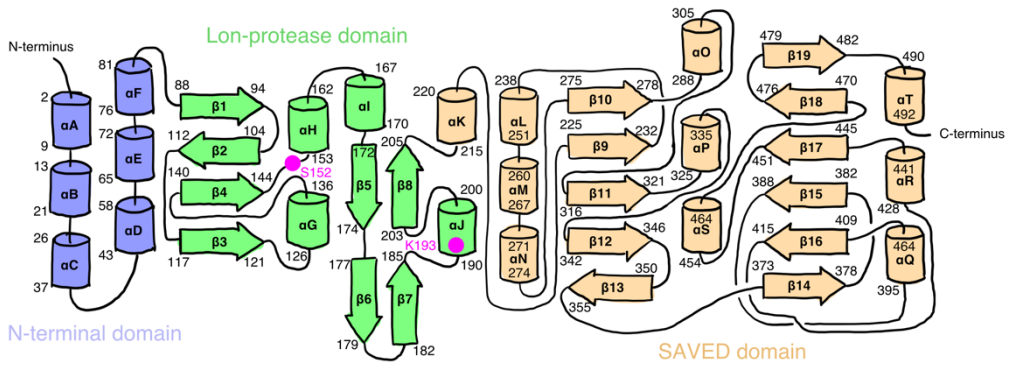
Figure 5-4: Structural overview of CalpL, solved by experimental phasing.
a) *Cartoon model of the CalpL structure showing the three domains in different colors. The positions of N-terminus and C-terminus are indicated as well as the active site residues (magenta spheres) and the cOA binding site.* ***b)*** *Electrostatic surface visualization of the upside (bottom) and the downside (top) of the protein. The partial charges are colored according to the included gradient legend. Active site position and cOA binding site are labeled.* ***c)*** *Magnified view onto two SeMet residues represented as sticks surrounded by the electron density map (blue mesh).*

methionine positions, the electron density map showed large spheres around the selenium atoms reflecting the higher atomic number of selenium compared to sulfur (**Figure 5-4 c**).

5.1.2.2 Structural homologies confirm the theory of a cOA activated Lon protease

To compare the structure-based findings to the predicted functions and architecture, we further characterized the individual domains. Moreover, we searched for proteins with structurally similar domains. The N-terminal domain of CalpL is made up of four α-helices (αA-αD) and shows only poor structural homologies, with the most interesting being a katanin domain which is essential for specific binding of polypeptides (K. Jiang et al., 2018). The Lon protease domain is formed by helices αG-αJ that, together with the helix bundle of the katanin-like domain, enwrap a four stranded β-sheet (β1- β4). A much higher structural conservation helped to find multiple structural homologs confirming the proteolytic function of this domain. By structural alignments, we identified S152 and K192 to form the characteristic catalytic Ser-Lys dyad of Lon proteases. The C-terminal domain represents the largest part of the protein and folds into a SAVED4-domain which is responsible for cOA binding (Makarova et al., 2020). Also here, structural conservation could be observed (**Figure 5-5**). The domain is built up of two CARF-like domains with a pseudo two-fold rotational axis running in between the two helices αP and αS. Interestingly, these two helices and the neighboring β-strands were predicted to form transmembrane helices (S. A. Shmakov et al., 2018). It is worth to mention, that an updated version of the THMHH 2.0 tool that uses deep learning methods (DeepTHMHH (Hallgren et al., 2022)) does not predict these helices to be membrane associated. The structural homologs of the protease domain, were later used to investigate the protease cleavage site at the target polypeptide (see **Section 5.3.1**).

a



b

Chain	Z	rmsd	%id	PDB Description
5ow5-A	6.4	2.8	11	KATANIN /.../ SUBUNIT B1;
5dlq-B	6.4	3.7	15	EXPORTIN-4;
7wtw-w	6.2	3.1	13	18S RRNA;
8huj-B	6.0	2.7	12	EUKAR. INITIATION FACTOR
6sl1-A	5.9	3.8	11	Ser/Thr-PROTEIN KINASE TEL1;
7z8v-D	5.8	3.1	12	CULLIN-1;
3icq-T	5.8	3.4	8	EXPORTIN-T;
7unk-A	5.7	3.3	6	IMPORTIN-4;
4m57-A	5.6	3.1	11	CHLOROPLAST PENTATRICO /.../
6fai-i	5.6	3.0	12	40S RIBOSOMAL PROTEIN /.../

Chain	Z	rmsd	%id	PDB Description
5h45-B	13.1	2.1	15	DNA REPAIR PROTEIN RADA;
8dvh-A	12.7	2.1	21	LON PROTEASE 2;
1xhk-B	12.6	2.1	20	PUTATIVE PROTEASE /.../
3k1j-A	12.3	2.2	15	ATP-DEPENDENT PROTEASE LON
6v11-D	12.2	2.0	17	LON PROTEASE;
7nfy-A	12.1	2.2	16	LON PROTEASE HOMOLOG,/.../
7fd4-A	11.8	2.1	16	LON PROTEASE;
5lkm-A	11.1	2.0	16	DNA REPAIR PROTEIN RADA;
4fw9-A	10.9	2.4	17	TTC1975 PEPTIDASE;
8emh-B	10.5	2.6	12	PROTEASE LON-RELATED BREX

Chain	Z	rmsd	%id	PDB Description
7rwk-A	19.1	3.6	13	SAVED DOMAIN-/.../ PROTEIN;
7rwm-C	18.2	3.2	12	SAVED DOMAIN-/.../ PROTEIN;
7qqk-A	17.8	3.6	8	TIR_SAVED FUSION PROTEIN;
6vm6-B	17.4	3.6	13	SAVED DOMAIN-/.../ PROTEIN;
7yib-A	15.6	3.7	10	CD-NTASE-ASSOCIATED PROTEIN
6sce-A	8.1	4.6	10	UNCHARACTERIZED PROTEIN;
8bao-A	5.8	2.5	14	DUF1887 FAMILY PROTEIN;
7ygh-B	5.6	2.7	8	CRISPR SYSTEM RING NUCLEASE
6isq-A	5.5	3.4	6	LIPASE B;
2i3d-B	5.5	3.5	14	HYPOTHETICAL PROTEIN /.../

Figure 5-5: Domain architecture of CalpL and structural similarity search of domains.
a) *Topology diagram of CalpL including information on secondary structure elements. The different domains are marked by different colors as indicated. Catalytic residues S152 and K193 are highlighted by magenta spheres.* ***b)*** *Top ten results of structural similarity searches for the individual domains using the DALI server (Holm et al., 2023). The most interesting entries are accentuated in yellow.*

5.1.3 Biochemical demonstration of a functional cOA activated protease

The structural analysis confirmed the theory of a cyclic nucleotide activated proteolytic enzyme. While molecular dynamic simulations validated that the SAVED domain would be suitable to bind cA_3, cA_4 or cA_5, only the assumption that the target must contain a flexible loop could be made by locating the catalytic dyad. Phage display experiments to search for peptide sequences that show affinities towards CalpL remained unsuccessful. Therefore, we tried a different approach. Using the WebFLAGs server the gene neighborhood of *calpL* was analyzed to detect a highly conserved open reading frame upstream but within the same operon. The hypothetically encoded protein product (271 amino acids, UniProt ID B2V8L8) without annotated function shared sequence similarities to the MazF toxin within its N-terminal half. We decided to express and purify the protein we called CalpT (for target) to test it for proteolytic degradation.

5.1.3.1 Expression and purification of CalpT

A synthetic, codon-optimized variant of the CalpT encoding gene (*CalpT*) was cloned into a pET-11a vector. The expression construct was designed similarly to that of CalpL expression construct (N-terminal His_{10} tag and a TEV protease cleavage sequence). The protein was expressed in *E. coli* BL21 (DE3) cells. After optimizing the expression yield, different buffer compositions were tested for lysis and affinity chromatography. Unfortunately, CalpT always tended to aggregate either directly after lysis or after elution with imidazole. Keeping the protein at 20 °C after lysis improved the aggregation behavior, but did not solve the instability problem completely. Testing buffer further buffer compositions finally solved this problem. According to surname of the colleague who provided the buffer, it was named ‘Buffer W’. Especially for crystallization purposes, the purification was further optimized to increase the purity, e.g. heparin

chromatography helped to drastically increase the homogeneity. However, for first cleavage assays, Ni^{2+} affinity chromatography followed by gel filtration provided a sufficient degree of purity (**Figure 5-6 a**, lane '2').

5.1.3.2 Design of a protease cleavage assay to identify target and activator at once

To investigate whether CalpT can be cleaved by CalpL, the two proteins were incubated for 1 h at 60 °C with different cOAs at 1.2:1:1.5 molar ratios. Subsequently, the different samples were analyzed by SDS-PAGE. The individual proteins migrated as expected in the polyacrylamide gel (CalpL ~55 kDa, CalpT~30 kDa). While for CalpT a weak band slightly below 25 kDa was observed, CalpL did not contain any noticeable impurities (Figure 5-6 a, 1-2). Incubation of both proteins without addition of cOA resulted in a combination of the individual protein bands. Addition of cA_4 to the incubated protein mixture led to a different pattern of bands (Figure 5-6 a, 5). The CalpT band in between 25 kDa and 35 kDa almost disappeared, two new bands arose, one clearly visible band slightly below 25 kDa and one very faint band at ~10 kDa. For the incubation with other cOAs, minimal activity of CalpL could be observed. To ensure, that CalpT degradation is caused by activated CalpL, the experiment was also done with a variant of CalpL (S152A) which lacks the nucleophilic serine of the catalytic dyad and showed no peptidase activity (Figure 5-6 a, 8-10). A functional cleavage could also be visualized by SEC-MALS experiments (Figure 5-6 b). The experiments did not only reveal that CalpL and CalpT for a stable heterodimeric complex in solution (purple), but furthermore, that upon activation with cA_4 the ~10 kDa cleavage fragment remained bound to CalpL while the ~23 kDa eluted in a separated peak (black). The full CalpT–L complex eluted significantly earlier than that of $CalpT_{10}$–L which in turn eluted earlier than CalpL alone (dashed green). In addition, the second cleavage fragment $CalpT_{23}$ eluted significantly

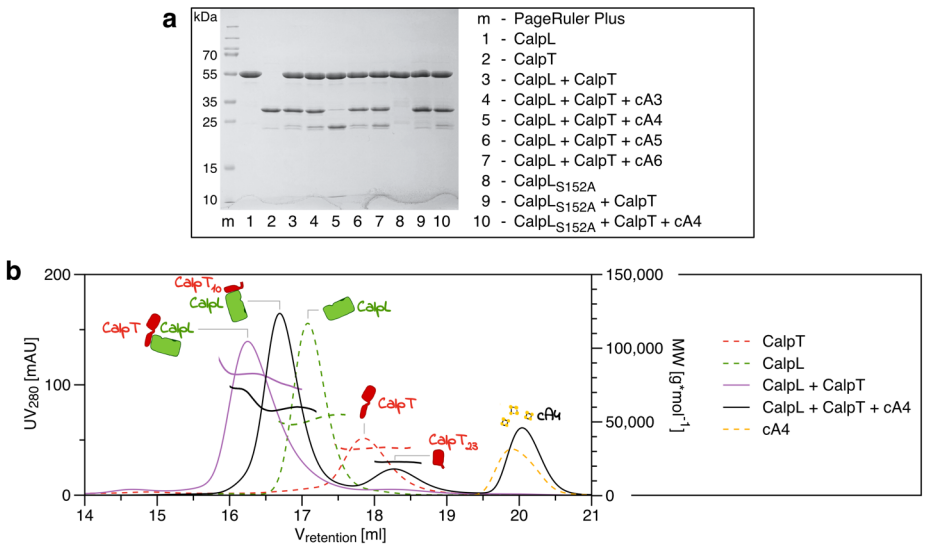


Figure 5-6: Protease activity assay to determine activator and target at once.
a) *SDS-PAGE analysis of samples from protease activity assay. A band at ~55 kDa shows CalpL and one in between 25 kDa and 35 kDa corresponds to CalpT. On the left-hand side the molecular weights corresponding to the molecular weight marker (M) are annotated. All lanes are labeled according to the caption on the right-hand side indicating the content of the sample.* ***b)*** *SEC-MALS analysis of the CalpT–L complex and the cleavage fragments resulting from activation with cA_4. The UV_{280} absorbance and the molecular weight are shown on the left-handed and right-handed y-axis, respectively. The x-axis shows the retention volume. UV chromatograms are shown for all x-values and the MALS data is only shown for x-values where an absorbance signal was observed. The color coding is explained to the right to the chromatograms. For orientation, small pictograms of the peak contents are illustrated.*

later than full length CalpT (dashed red). Analysis of the protein bands by peptide mass fingerprinting revealed that the 23 kDa cleavage fragment is the N-terminal part of CalpT and the smaller 10 kDa fragment is the C-terminal part of the protein (**Figure 9-2**).

With the knowledge that CalpL is a cA_4 activated protease that specifically cleaves CalpT into two distinct fragments, further questions came up: on the one hand we wanted to find out more on the cleavage specificity and the role of $CalpT_{10}$ and on the other hand how the cA_4 triggered activation mechanism of CalpL works. However, these questions needed individual investigations.

5.2 CalpL is activated by cA_4-induced oligomerization

The next chapter focuses on the biochemical and structural investigation of the cA_4-triggered activation mechanism. The detailed analysis includes binding studies as well as mutational analyses based on structural data.

5.2.1 CalpL specifically binds cA_4.

To address the observation from protease cleavage assay that CalpL is specifically activated by cA_4 from a quantitative point of view, binding of different cyclic oligoadenylates (cA_3, cA_4, cA_5, cA_6) to CalpL was compared by surface plasmon resonance (**Figure 5-7 a**). While nanomolar affinity binding ($K_d \approx 1$ nM) was observed for cA_4, all other cOAs showed significantly weaker binding affinities that could not be

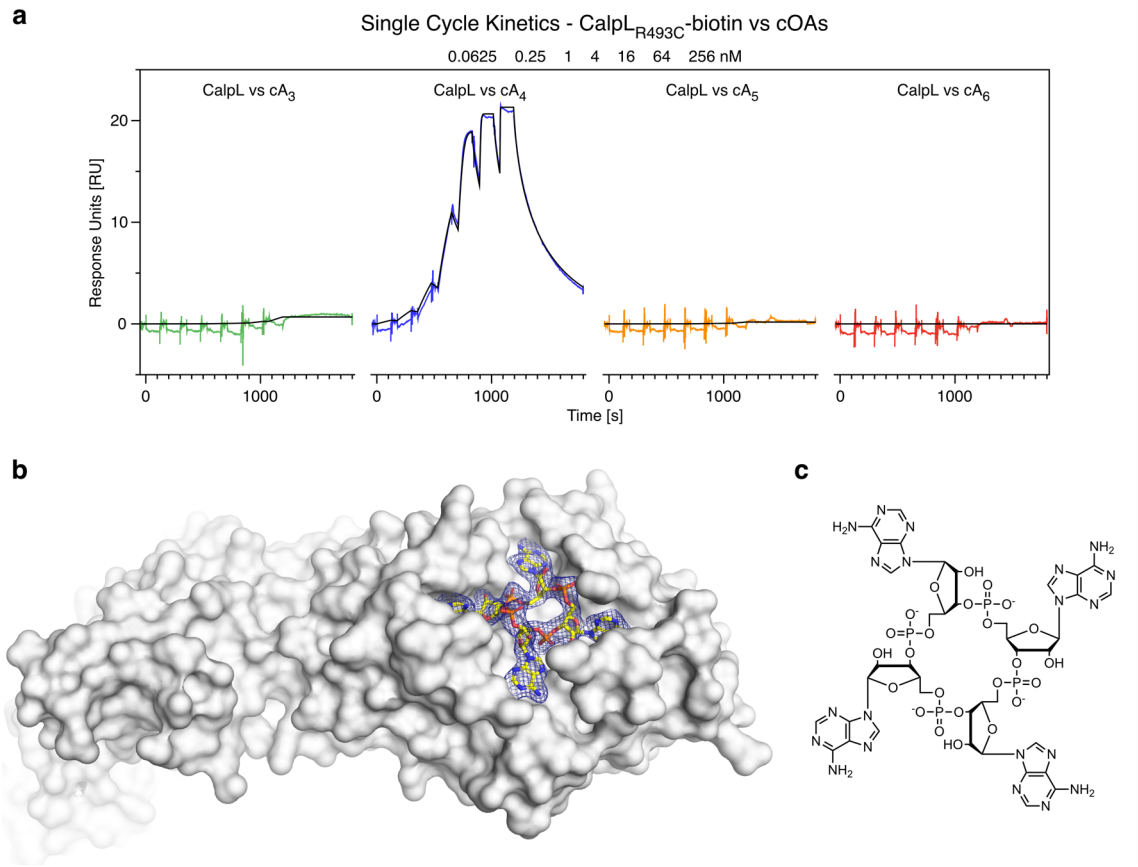


Figure 5-7: The SAVED domain of CalpL contains a deep cavity to specifically bind cA_4.
a) *Single cylce kinetics using CalpL as ligand and different cOAs as analytes. The response units are displayed against the time. The analyte concentration was increased from injection to injection from 625 pM to 256 nM.* ***b)*** *Crystal structure of CalpL binding cA_4. The cyclic nucleotide (stick representation) is shown inside its electron density (blue mesh). CalpL is visualized as surface model.* ***c)*** *Chemical structure of 3'-5' linked cyclic tetraadenylate (cA_4).*

accurately determined within the experimental concentration range. Notably, the single cycle kinetics experiment revealed fast association- and dissociation rates which appear in a steep rise directly after the injection start or a rapid drop in response units after the end of the injection. To study the cA_4-induced activation mechanism we aimed for a nucleotide bound CalpL structure. Several attempts of co-crystallization were unsuccessful, as either no crystals, only cA_4 crystals, or CalpL crystals without nucleotide grew. However, by changing the approach to soaking native CalpL crystals with cA_4 solution (5 mM), it was possible to elucidate the structure at 2.2 Å resolution. The nucleotide-free CalpL structure was used for molecular replacement and the final model was refined to R_{work} and R_{free} values of 23.6 and 26.7, respectively. A highly defined difference electron density was observed for the cyclic oligonucleotide located in the deep surface cavity on the "upside" of the SAVED domain of the protease (**Figure 5-7 b,** compare **Figure 5-4 b**). The structure demonstrates that cA_4 with its two-fold symmetry (**Figure 5-7 c**) is perfectly embedded in the positively charged surface cavity formed by the pseudo two-fold symmetric SAVED domain. Each two opposite adenine moieties are either horizontal or vertical to the CalpL surface with the latter ones being involved in a π-stacking with Y346 and Y475, respectively (**Figure 5-8 a**). Further electrostatic and hydrophobic interactions could be identified that target either the nucleobases, the ribose subunits or the phosphate groups. Some of these interactions are indicated in Figure 5-8 a. However, the structure did not show major conformational rearrangements compared to that of cA_4 free CalpL (RMSD=0.365 [over 471 Cα-atoms]). Only minor changes in the loops surrounding the binding pocket were identified (**Figure 5-8 b**).

5.2.1.1 Losing the ability to bind cA_4 leads to a loss of protease activity.

To prove that binding of cA_4 to the distinctive cavity is directly related to protease activity, we mutated some of the residues that form the substrate binding site and analyzed the effect on proteolytic cleavage in combination with binding studies. These residues are shown as magenta sticks in Figure 5-8 a. In order to follow the reaction, samples were taken for SDS-PAGE analysis at 1 min and 5 min, respectively, after the addition of cA_4. Wild type CalpL was able to degrade CalpT almost completely after just 1 min of incubation, whereas the protease-dead variant (S152A) did not show any cleavage. Some single mutants only slightly affected the cleavage activity, at least within the given reaction times. The most pronounced decrease in cleavage activity was observed for the Y346E mutant. A390F slowed down the degradation process, but after 5 min almost all CalpT was cleaved. The combination of two less active single mutants to a double mutant resulted in a further loss of proteolytic activity (**Figure 5-8 d**). In a second approach, a sterically challenging residue was introduced in the center of the surface cavity, which also reduced protease activity. However, when biotin was attached to the cysteine residues of the mutants, for S235C the activity was further decreased as expected, whereas biotinylating V454C increased the cleavage efficacy compared to the non-biotinylated mutant (**Figure 5-8 e**). The cA_4 binding behavior of the mutants that showed the strongest effect in the protease activity assays was then characterized by SPR measurements. This confirmed a relationship between cA_4 binding and proteolytic activity. Single cycle kinetics were performed using CalpL V454C (lower left), CalpL Y346E (upper right) and CalpL Y346E/A390F (lower right) as ligand and cA_4 as analyte. The measurement of wild type CalpL (upper left) serves as reference. While the mutation of valine 454 to a cysteine drastically weakened cA_4 binding, the single site mutation of tyrosine 346 to glutamic acid resulted in a reduced but still observable binding affinity. The introduction

of a second mutation (Y346E/A390F) almost abolished the ability to bind cA4. These results are fully consistent with the observations from the activity assays.

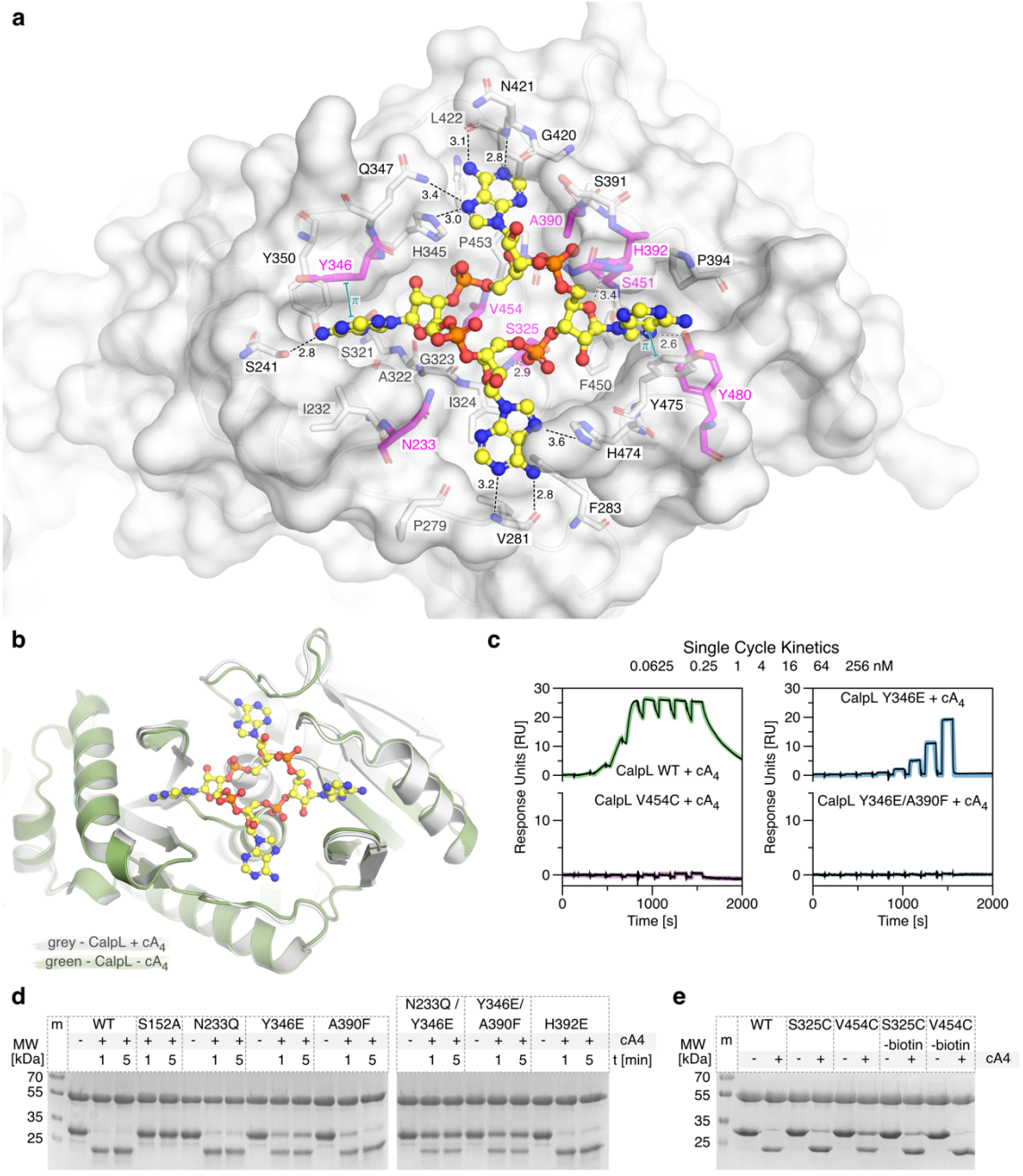


Figure 5-8: Structure guided and detailed investigation of the cOA binding site.
a) *Visualization of interacting residues within the cOA binding site. The surface of the CalpL SAVED domain is shown and cA4 is included as ball and stick model. Amino acid sidechains within 4 Å distance to the nucleotide are represented as sticks. Residues which were mutated for subsequent analyses are highlighted in magenta. All shown residues are labeled and some electrostzatic interactions are indicated by dashed lines. Bond lengths are given in Ångström.* ***b)*** *Cartoon representation of apo CalpL (green, PDB ID: 7qda) superimposed to CalpL binding cA4 (gray, PDB ID: 8b0r) to illustrate the minor changes between the two states.* ***c)*** *SPR binding curves of different cOA-binding site mutants of CalpL showing a drastically decreased affinity towards cA4 compared to the wild type protein (green). For each individual single cycle experiment, response units are displayed against the time. The concentration of eacht single injection is shown above all four graphs.* ***d)*** *SDS-PAGE analysis of samples from protease activity assays to investigate the effect of mutating a single (or two single) amino acid side chains onto protease activity. Wild type CalpL (WT) and a protease dead variant (S152A) were included as reference. All mutants are indicated. Samples have been taken after 1 min and 5 min after activation, respectively.* ***e)*** *Same as in d) but with mutants that introduce a sterically demanding residue in the middle of the cA4 binding cavity. The reaction time was set to 3 min in this assay.*

5.2.2 Binding of cA_4 to CalpL induces oligomerization

So far, we have analyzed in detail that the binding of the oligonucleotide is essential, as well as the presence of the two catalytic residues, serine 152 and lysine 193. However, since cA_4 binding did not induce any conformational changes of CalpL, at least in our structure, we still wondered what mechanism causes the activation of the protease. Scientific discussions with collaborators at the University of St. Andrews, Scotland, led to the idea that the binding of cyclic oligoadenylates might trigger oligomerization in an organized manner. A similar behavior was later published for a SAVED domain-containing immune effector where cA_3 activates the protein by acting as a molecular glue causing a helical assembly (Hogrel et al., 2022). Recall that we indicated a positively charged surface patch at the downside of the CalpL SAVED domain (see **Figure 5-4 b**). This surface area could be made for such an oligomerization as it could sit on top of another CalpL molecule bound to cA_4 by interacting with the negatively charged phosphate groups.

To test this hypothesis, dynamic light scattering (DLS) and small-angle X-ray scattering (SAXS) experiments were performed. The latter were carried out and analyzed by collaborators in Hamburg. The SAXS data revealed an increased molecular weight of CalpL in a cA_4-dependent and protein concentration-dependent manner. In addition, the data was used to produce a dimeric *ab initio* model. Thereby, it was important not to impose symmetry elements as we expected non-identical binding interfaces. The fitting algorithm generated stable solutions consisting of two stacked elongated shapes which were large enough to fit two CalpL monomers (**Figure 5-9 a**). To further validate the importance of the positively charged surface patch for the stacking, charge-flip mutants were designed that are located within this region (**Figure 5-9 c**). After successful mutagenesis, expression and purification of these protein variants they were tested for their ability to cleave CalpT (**Figure 5-9 d**). Arginine to glutamic acid mutations showed reduced (R361E) or even a complete loss (R338E) of proteolytic activity. Interestingly, for the R493C mutant, which was designed for site specific immobilization in SPR experiments and is located at the edge of the positively charged surface region, also showed a decreased activity. Both lysine to glutamic acid mutants tested did not exhibit significant changes in their cleavage efficiencies. DLS experiments confirmed a concentration dependent oligomerization behavior. An increasing particle size was observed upon addition of cA_4 to wild type CalpL whereas without addition of the cyclic nucleotide, the particle radii remained at the size calculated for a monomeric protein (**Figure 5-9 e**). The experiment was done in the same way for different mutants. Strikingly, the double mutant Y346E/A390F, for which no more cA_4 binding was observed, in turn lost its ability to oligomerize. Besides, one “downside” mutant (R361E) also resulted in a drastically decreased cA_4-dependent oligomerization. Taken together, the different techniques helped to validate that cA_4 induces oligomerization of CalpL, which then triggers the proteolytic cleavage.

5.2.2.1 Oligomerization induced cleavage of CalpT occurs ‘in trans’

With the certainty that CalpL forms staggered oligomers with cA_4 acting as molecular glue, sandwiched in between the monomers, we wondered why this behavior is necessary for activation. A plausible explanation for such a mode of action would be an ‘in trans’ cleavage, where CalpL cleaves the CalpT molecule bound to a neighboring protease. To investigate this, mixtures of preformed CalpT–L complexes (e.g. CalpL WT + CalpT WT,

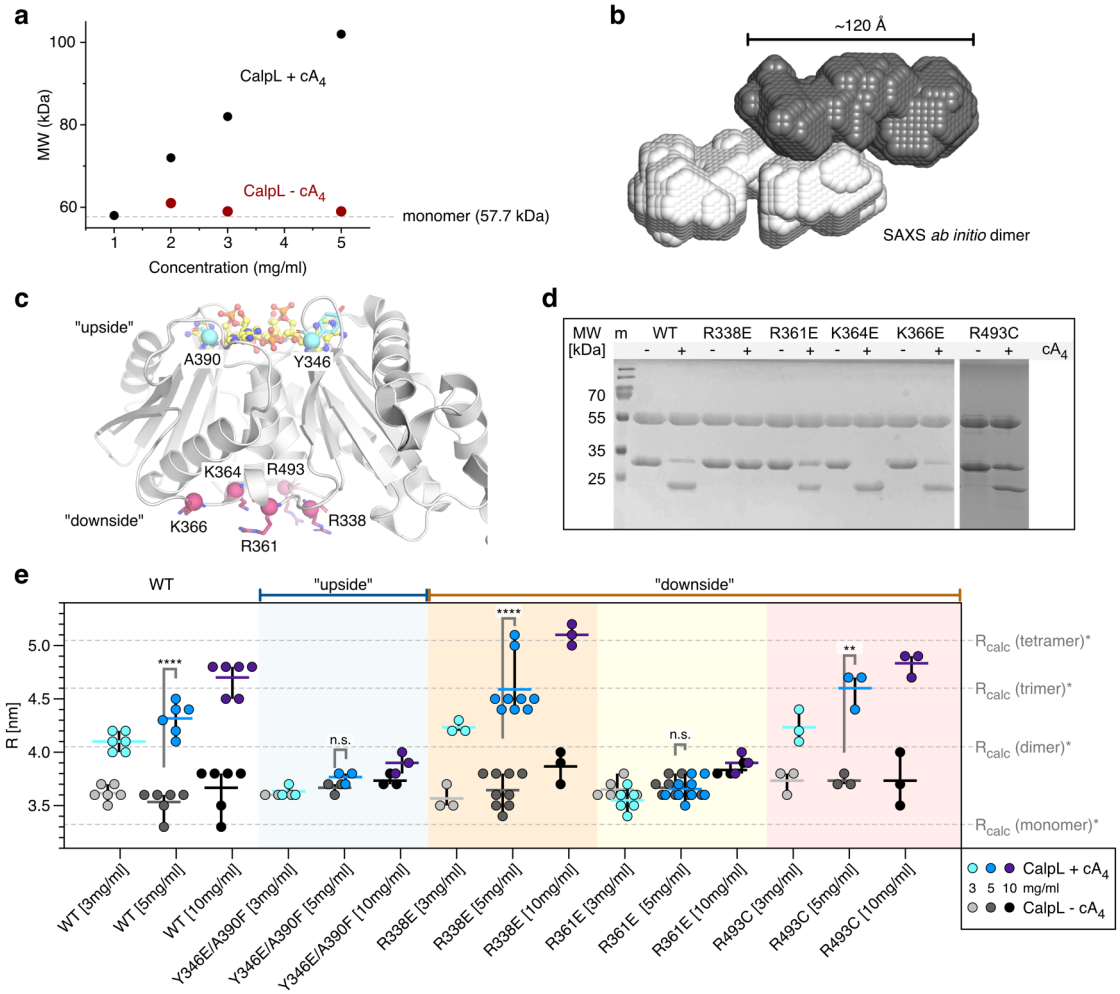


Figure 5-9: Binding of cyclic tetraadenylate triggers oligomerization of CalpL.
a) *Molecular weights derived from forward scattering, I_0, calculated from SAXS data plotted against the concentration. At each concentration one data set was recorded with (red) and without (black) an equimolar amount of cA_4. The molecular weight of monomeric CalpL is indicated by a dashed line.* ***b)*** Ab initio*/rigid-body model of a CalpL dimer generated based on SAXS data.* ***c)*** *Side view of the SAVED domain of CalpL, the "upside" is annotated as well as the "downside". Residues that were tested for their contribution in protease activity and oligomerization are highlighted and labeled.* ***d)*** *SDS-PAGE analysis of protease activity assay to test "downside" mutants for protease activity.* ***e)*** *Hydrodynamic radii of different CalpL mutants determined by dynamic light scattering at different protein concentrations and in the absence (gray to black) and presence (cyan to violet) of cA_4. Horizontal dashed lines indicate the calculated sizes of a CalpL monomer and modelled CalpL oligomers as indicated. The theoretical radii were computed using HullRad (Fleming & Fleming, 2018). The significance of the differences in hydrodynamic radii depending on the presence of cA4 was tested by two-tailed t-tests and is annotated (n.s. - $P > 0.5$; * - $P \leq 0.05$;** - $P \leq 0.01$; *** - $P \leq 0.001$;**** - $P \leq 0.0001$)*

CalpL S152A + CalpT WT, CalpL WT + CalpT A195E) were tested for cA_4-triggered proteolytic cleavage (**Figure 5-10 a**). CalpT A195E is a non-cleavable variant of the protease target, further details on this mutant can be found in **Section 5.3.1**. No cleavage was observed for single mixtures of CalpL S152A–CalpT WT or CalpL WT–CalpT A195E. However, mixing the two complexes, which are not capable of 'in cis' cleavage, in a 1:1 molar ratio just before the addition of cA_4, resulted in 50 % degraded CalpT (**Figure 5-10 a,** "50/50"). The remaining 50 % of uncleaved CalpT was caused by the A195E mutation. This could be confirmed by changing the ratio of the two complexes (**Figure 5-10 a,** "75/25"), where only 25 % of full-length CalpT remained. Addition of uncomplexed CalpL to a CalpL S152A–CalpT WT complex just prior to addition of the

nucleotide addition, led to a complete degradation of CalpT. These results demonstrated, that the target of CalpL is cleaved ‘in trans’ by a protease adjacent to the CalpT–L complex that actually contains the degraded CalpT molecule.

5.2.2.2 The directionality of proteolytic ‘in trans’ cleavage is defined

Knowing the above, another question immediately arose: is the cleavage direction defined within a staggered oligomer, or is it unidirectional? To examine this, the protease activity assay was further modified. One component (for instance CalpL or the complex of CalpL + CalpT) was immobilized on Streptavidin beads to sterically block the bottom side and the other component was provided freely in solution. Just before activation, the individually preincubated components were mixed (**Figure 5-10 b**). When all CalpT–L complexes were bound to the beads, no cleavage was observed. This made it clear, that there is no ‘in cis’ cleavage in addition to the ‘in trans’ mechanism. Furthermore, we could identify differences for the mixtures of immobilized CalpT–L complex and soluble protease complex (lane 4/5) compared to immobilized protease and soluble CalpT-L complex (lane 2/3). When the protease can only attack from the top, no cleavage of CalpT occurred (lane 4/5). In contrast, when the protease could only attack from the bottom, CalpT was completely degraded in the assay.

In conclusion, the addition of CalpT to CalpL results in a stable heterodimeric complex. This complex oligomerizes in an ordered manner upon addition of cyclic tetraadenylate. The oligomerization in turn, activates the protease–target complex and leads to a directional ‘in trans’ cleavage. The directionality is defined so that the protease cleaves exclusively ‘upwards’ (**Figure 5-10 c**).

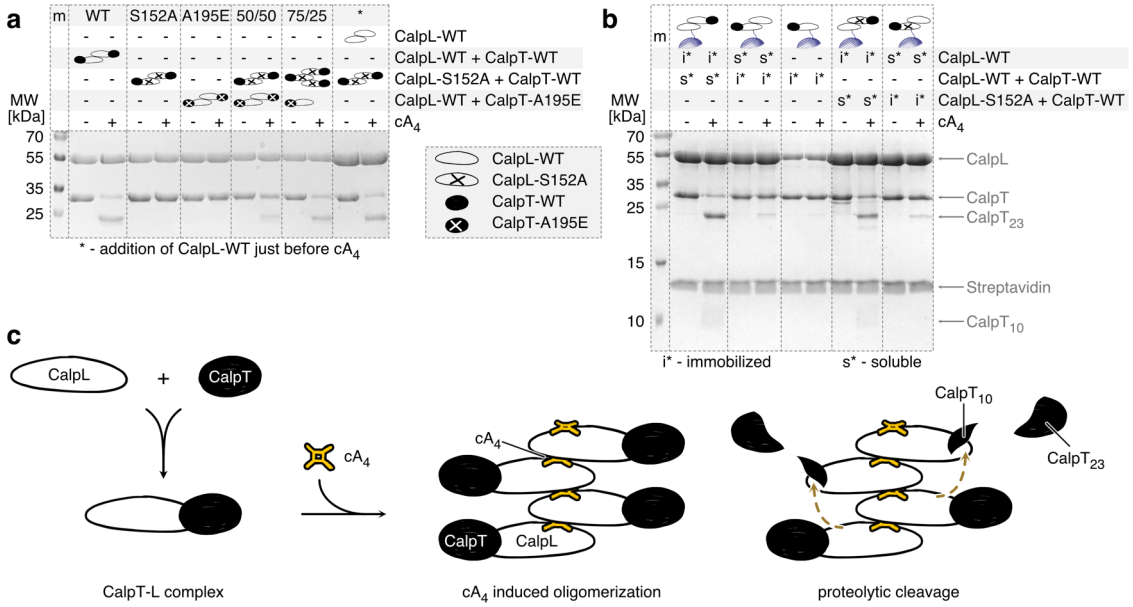


Figure 5-10: Determination of the cleavage directionality by protease activity assays.
a) *Protease activity assays (15 % SDS–PAGE, Coomassie stained) with different preformed CalpL–CalpT complexes, as indicated. Pictograms are included for visualization.* ***b)*** *Same as in a) but using biotinylated CalpL bound to Streptavidin beads to block the bottom side. The indications “i*” and “s*” describe which component was immobilized to the beads and which was free in solution, respectively.* ***c)*** *Schematic representationof the cA_4-induced cleavage mechanism. CalpL and CalpT form a stable complex (left. Addition of cA_4 leads to an ordered oligomerization (middle), which results in an activation and subsequent ‘in trans’ cleavage of CalpT bound to a protease on top of the enzymatically active one (right).*

5.3 Site specific cleavage requires CalpT–L complex formation

The detailed analysis of the cleavage site specificity as well as the role of CalpT–L complex formation for effective proteolytic cleavage is described in the following section. This study includes structural and biochemical analyses.

As the size of the CalpT cleavage fragments were in line with the analysis by peptide mass fingerprinting. CalpT was cleaved into a distinct larger N-terminal and a smaller C-terminal cleavage fragment upon incubation with CalpL and subsequent activation of the protease by cA_4 addition. To determine the specific cleavage sequence, we had a closer look into the structure and used the structure of the ATP-independent Lon protease of the yellowfin ascites virus for structural superposition (PDB ID: 4izj, (Chung & Paetzel, 2013)). This helped to align a peptide chain into an elongated surface cavity that ends at the protease catalytic dyad (see **Figure 5-11**).

5.3.1 Structure guided mutational analysis to investigate the cleavage site of CalpT

The mentioned superposition hinted the position of the P1 site with its amino acid side chain pointing towards a small hydrophobic cavity of the CalpL surface suggesting to fit only tiny hydrophobic residues such as alanine or glycine. From the P1 site, the peptide chain proceeds through the channel in N-terminal direction with every second side chain pointing towards the CalpL surface (**Figure 5-11**). After this structural modelling we used the AlphaFold2 software to structurally predict CalpT. The program built a two-domain protein with a ~23 kDa large N-terminal part and a C-terminal part of ~10 kDa. Both domains were connected by a flexible loop which was searched for alanine or glycine residues, respectively (**Figure 5-12 a**). Indeed, we found four alanine residues within the respective region (**Figure 5-12 a-b**) and created glutamic acid variants for each of them individually (A172E, A182E, A195E, A201E). Protease cleavage assays using all four variants were performed and observed no cleavage for CalpT A195E (**Figure 5-12 c**). Assuming this alanine to be the P1 residue, the amino acid sequence at the cleavage site would read V_{190}-L-R-H-V-A|S-T with the peptide bond in between A195 and S196 being

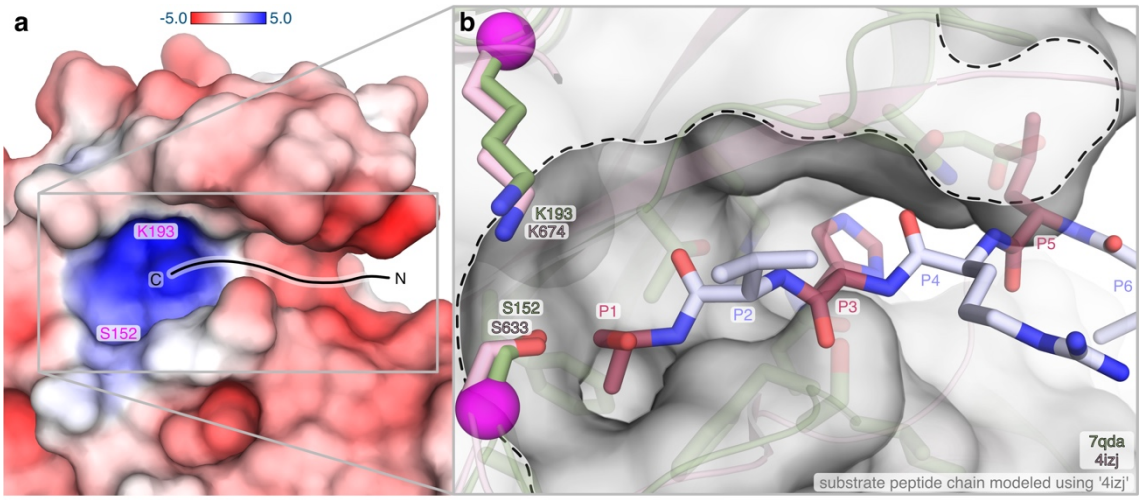


Figure 5-11: CalpL substrate channel superposition.
a) *Surface electrostatics of the Lon protease active site region. The positions of the catalytic dyad residues are indicated. A line marks the likely substrate channel with indicated N- and C- termini of the peptide.* ***b)*** *Magnification of the substrate channel including a superposition with the enzyme intermediate of yellowfin asciitis virus protease (PDB ID: 4izj). The protease catalytic residues are indicated and highlighted by magenta spheres. The substrate peptide chain sequence was modified to match the CalpT cleavage site sequence. Even and odd amino acid residues of the peptide chain are colored differently and annotated by P1-P6.*

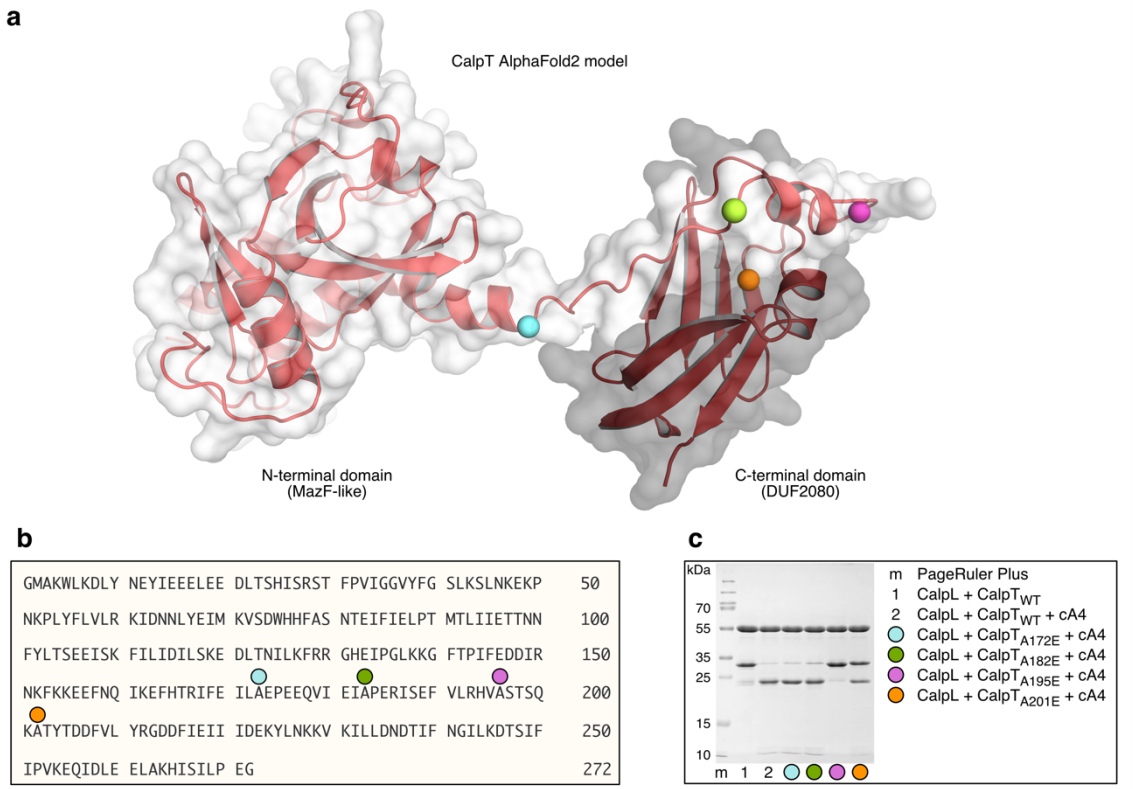


Figure 5-12: Structural prediction of CalpT helped to identify the protease cleavage site. ***a)*** *AlphaFold2 generated structural model of CalpT with its N-terminal MazF-like domain on the left-hand side (red, white) and its smaller C-terminal domain on the right-hand side (red, black). Alaninine residues within the unstructured loop in between both domains are highlighted by colored spheres.* ***b)*** *Amino acid sequence of CalpT starting with "G0" that remains from the TEV cleavage site. Alanine residues of the linking loop are marked by colored shperes above the respective position. The coor coding was done in the same way as in a).* ***c)*** *SDS-PAGE analysis of samples from a protease activity assay using wild type CalpL and CalpT A to E mutants. Wild type CalpT was used as reference with and without activation of the protease. For all CalpT mutants, the protease was activated. The color coding of the mutants was done in analogy to a) and b).*

the hydrolyzed one. Cleavage at that position would lead to an N-terminal fragment of 23.17 kDa size and a C-terminal fragment of 8.75 kDa size which is in agreement with the observations made in the previous experiments. Interestingly, alanine 195 as well as the directly neighboring amino acids are conserved among CalpT homologs.

Furthermore, SEC-MALS analysis was used to investigate if the mutants were able to form a stable complex with CalpL as it was observed for the wild type protein. All four glutamic acid variants were incubated with CalpL in a 1:1.2 (CalpL:CalpT) molar ratio and subsequently were run on a Superose6 gel filtration column. All mutants, except A201E showed complex peaks at the same retention volume as observed for CalpT and CalpL wild type proteins. However, the mixture of CalpL and CalpT A201E eluted in two separated peaks consistent with those of the individual protein samples (**Figure 5-13**). The experimental molecular weights obtained from multi-angle light scattering supported that the A201E variant is unable to form a stable complex with CalpL, at least at the used experimental conditions. Notably, this mutant did only show partial cleavage in the protease activity assay (compare **Figure 5-12 c**).

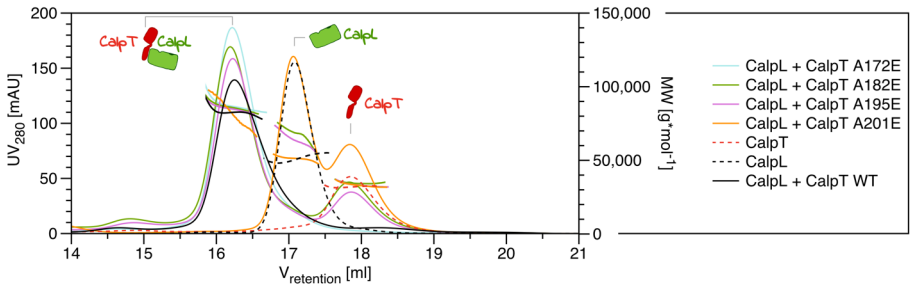


Figure 5-13: Analysis of the effect of CalpT cleavage site mutations on complex formation.
SEC MALS analysis of wild type CalpL with different A to E mutants of CalpT to investigate the effect of the respective mutation onto complex formation. The retention volume is annotated on the x-axis, the UV absorbance on the left y-axis and the molecular weight on the right y-axis. The color coding of the graphs is explained in the caption on the right-hand side.

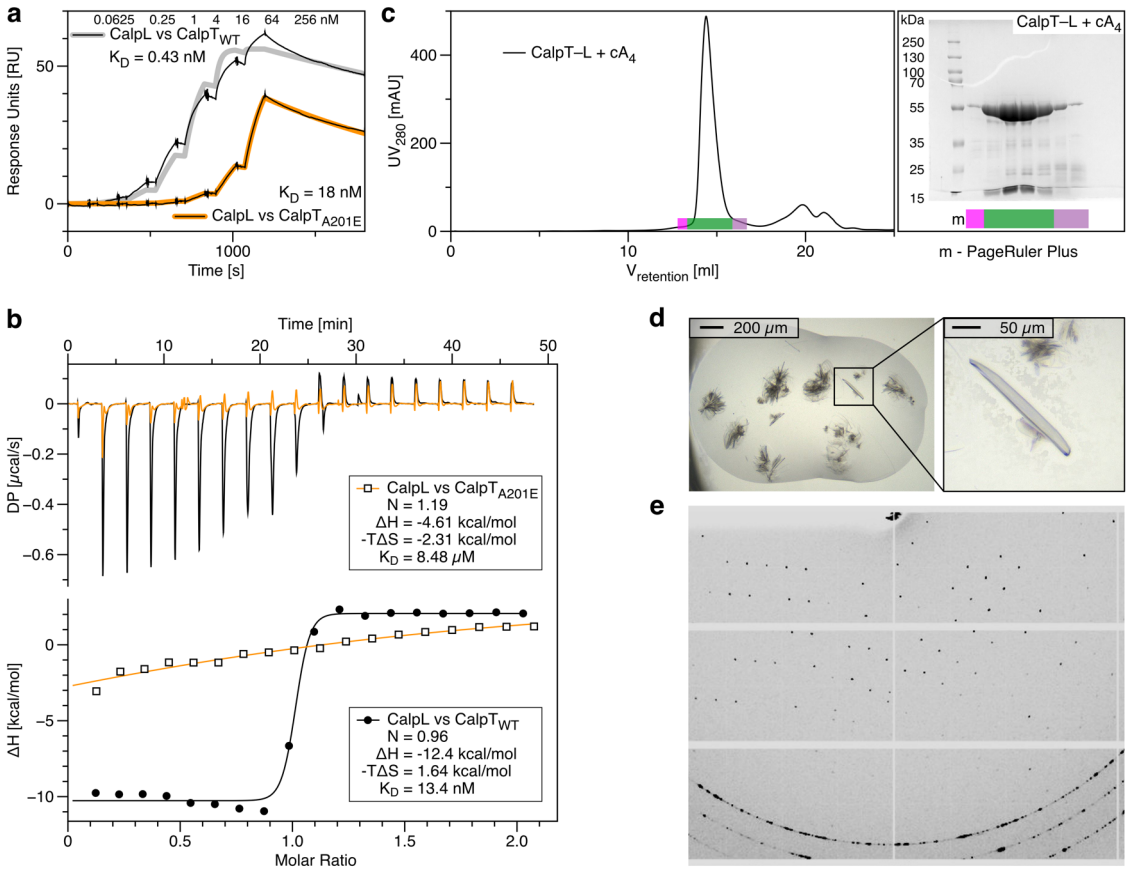


Figure 5-14: CalpT-L complex analysis and crystallization.
a) *Single cycle kinetics to quantitatively investigate the complex formation of CalpL and CalpT. The individual experiments for wild-type CalpT (gray) and the CalpT A201E mutant (orange) are overlaid in the graph of response units agains reaction time. The concentrations of the individual injections are shown at the top.* ***b)*** *ITC experiment for the CalpT-L complex formation. The differential power (DP) of the thermograms are applied against the reaction time (top) and the resulting binding curves as a function of the enthalpy against the molar ratio (below). The measurement of wild-type CalpT (black) and that of the A201E mutant (orange).* ***c)*** *Gel filtration chromatogram (UV_{280} against rention volume) of the $CalpT_{10}$-L complex (left). And the corresponding SDS-PAGE analysis of selected fractions (right). The fractions indicated by a green bar were pooled, concentrated and used for crystallization.* ***d)*** *Images of the protein crystals grown after 5 days of incubation at 20 °C. On the left-hand side, the whole drop is shown and on the right-hand side a magnification of the crystal which was used for X-ray scattering.* ***e)*** *Section of the diffraction image obtained from the scattering experiment using the crystal shown in d).*

5.3.2 Structural investigation of CalpT–L complex formation

Analytical gel filtration already showed complex formation of CalpT and CalpL qualitatively. Quantitative characterization of the interaction was achieved by SPR and ITC experiments. Both techniques suggested sub-nanomolar affinities of the two wild-type proteins, whilst the interaction of CalpL with the A201E mutant of CalpT was still observable but of drastically lower affinity (**Figure 5-14 a, b**). To learn more on the complex formation, we co-crystallized CalpL with CalpT. The two proteins were preincubated and the complex peak after size exclusion chromatography was pooled and concentrated. This procedure was done with and without addition of cA_4 to the protein mixture prior to the gel filtration step. However, only those crystallization experiments with the cleaved complex were successful.

5.3.2.1 Crystal structure of $CalpT_{10}$ in complex with CalpL

After size exclusion chromatography, the fractions which contained both proteins in a 1:1 ratio were pooled and concentrated to c≈30 mg/ml (**Figure 5-14 c**). After 5 days of incubation, crystals were obtained in condition E2 of the commercially available JCSG+ crystallization screen (Molecular Dimensions, Sheffield, UK) (**Figure 5-14 d**). Crystals were harvested with additional 35 % glycerol as cryoprotectant, flash frozen to liquid nitrogen and shipped to the Petra III electron storage ring in Hamburg. Diffraction data was collected (**Figure 5-14 e**) and the structure could be solved at 3.3 Å resolution by molecular replacement using the crystal structure of CalpL and the AlphaFold model of $CalpT_{10}$ as search models. Despite the 'low' resolution, all CalpT10 residues fitted well into the density except for residue S196. However, this residue was still included to the model (**Figure 9-3**). The structural data showed that the $CalpT_{10}$ fragment binds to the N-terminal 'katanin' domain of CalpL and the interaction interface area could be determined to A=691.6 $Å^2$ by the PISA online tool (Krissinel & Henrick, 2007). However, this tool commented that the interface might be a crystal packing artefact. To exclude this, we employed SEC-SAXS measurements of the complex which were in agreement with the crystal structure. The interface is highly charged, on the CalpL side mainly positively and on the CalpT side mainly negatively (**Figure 5-15 a**). As already suggested by the mutational SEC-MALS analysis (see **Figure 5-13)**, A201 of CalpT is part of the interface. Several more residues could be identified to contribute to the complex formation electrostatically and hydrophobically. The interactions are shown in detail in **Figure 5-15 b and c**. Interestingly, residue S196 of CalpT is located 37 Å far away from the protease active site and even stretching the loop would not make it fit there (**Figure 9-3**). This geometric arrangement excludes a self-cleavage within one heterodimeric complex without major conformational changes. This observation perfectly fits the results from the activity assays presented in **Section 5.2.2** which determined the cleavage to occur 'in trans'.

5.3.2.2 $CalpT_{10}$ is required for specific cleavage of the adjacent peptide sequence

As the A201E mutant of CalpT was cleaved with a drastically lower efficiency, we investigated whether this can be explained by the decreased affinity for complex formation. To test, if the heterodimeric complex formation of the protease and its target is a required condition, two different artificial constructs were designed. For one construct (NiS037), the two VcSiaP specific VHH antibodies (see **Section 2.1.2**) were linked by the amino acid sequence of the flexible loop region of CalpT that includes the protease

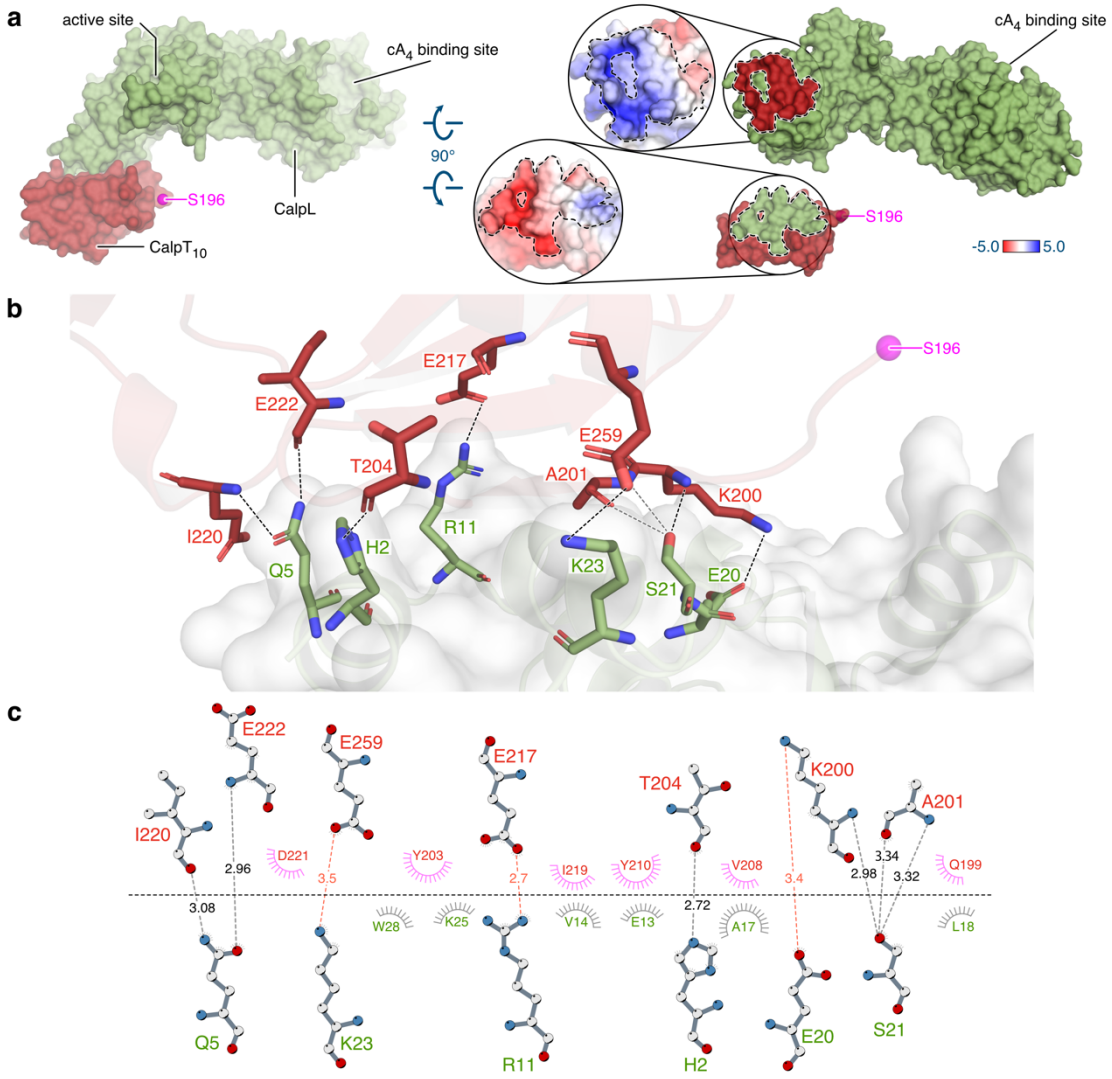


Figure 5-15: Structural analysis of CalpT$_{10}$–CalpL complex formation.
a) *Overview on the CalpT$_{10}$-CalpL complex structure. The two proteins are shown as surface model and significant regions are labeled. On the right-hand side, the two proteins are rotated to visualize the interface regions. These regions are also shown magnified and colored accoriding to the electrostatic surface potential as indicated.* ***b)*** *Magnification of the interaction interface of the structure. Amino acid side chains that contribute to the binding by electrostatic interactions are labeled and shown as sticks.* ***c)*** *Representation of interactions between CalpL and CalpT$_{10}$ visualized using LigPlot+ (Laskowski & Swindells, 2011). Electrostatic interactions (black) and salt bridges (red) are indicated by dashed lines and the respective bond lengths are given in Ångström (Å). Residues that undergo hydrophobic interactions are labeled in one letter code and their orientation is depicted. The color scheme is consistent from a) to c) (CalpT-red, CalpL-green).*

cleavage site. The other construct (NiS038) was designed similarly, but with the CalpT$_{10}$ sequence instead of the C-terminal VHH antibody (**Figure 5-16 a**). For the latter one, complex formation with CalpL as well as with VcSiaP could be observed in gel filtration experiments. Additionally, it could be demonstrated, that NiS038 could be cleaved by CalpL even when interacting to VcSiaP via the N-terminal nanobody (**Figure 5-16 b**). This showed, that the N-terminal part is not relevant for an efficient cleavage and even increasing that part by a factor of 2 does not lead to a decreased cleavage. In SPR measurements, a binding behavior of CalpL to NiS038 was observed that was similar to that of CalpT (**Figure 5-16 c**). The results could be also verified by isothermal titration calorimetry (data not shown). Both artificial constructs were also tested in protease activity assays to find that only NiS038 could be proteolytically degraded, while NiS037

remained as a whole. This suggested, that the C-terminal domain of CalpT, namely $CalpT_{10}$, is absolutely necessary for CalpT to be cleaved by CalpL. By this, specificity is introduced which is not necessarily based on a recognition sequence adjacent to the cleavage site, as known for e.g. TEV protease, but on a complementary domain architecture of $CalpT_{10}$ and the N-terminal domain of CalpL.

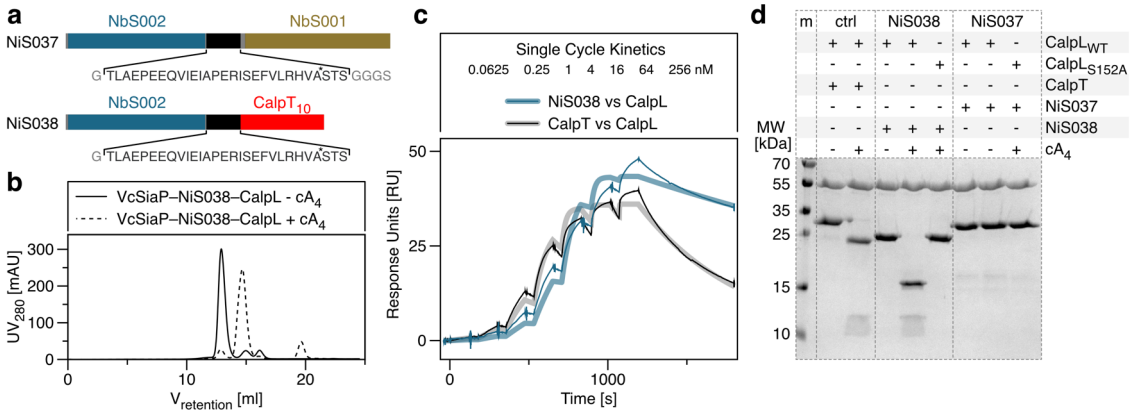


Figure 5-16: Investigating the role CalpT$_{10}$ with respect to an efficient proteolytic cleavage.
a) *Sketch of the two artificial protein constructs, NiS037 and NiS038. NbS001 and NbS002 are VcSiaP specific VHH antibodies (described in Part I) The linker sequence is denoted and the CalpL cleavage site is indicated by a '*'.* ***b)*** *Size exclusion chromatogram of the VcSiaP/NiS038/CalpL complex in absence (solid line) and presence (dashed line) of cA$_4$, respectively. The UV absorption is plotted against the retention volume.* ***c)*** *Single cycle kinetics using CalpL as ligand and CalpT (black, gray) or NiS038 (blue) as analyte. The response units are shown on the y-axis and the time on the x-axis.* ***d)*** *SDS-PAGE analysis of samples from protease activity assays. The components of each reaction mixture are indicated above the 15 % polyacryl amide gel. CalpL and CalpT wild type proteins were used as control ('ctrl').*

5.4 Elucidating the role of CalpT in a phage defense signaling cascade

After studying the cA$_4$ activated cleavage mechanism of CalpL and determining the required properties of the protease substrate, we next focused on the role of CalpT in an effective phage defense mechanism. Obviously, the site-specific proteolytic cleavage of the polypeptide needs to trigger a downstream effect in the bacterial immune response.

5.4.1 CalpT shares a structural similarity to MazE/F-like toxin-antitoxin systems

$CalpT_{10}$, characterized as a 'domain of unknown function', was found to play a key role in the essential complex formation with CalpL. However, the N-terminal domain, $CalpT_{23}$, did not affect the cleavage mechanism, but its structural homology to the MazF toxin (**Figure 5-17 a, b**), an endoribonuclease, provided a starting point for further analyses. As it is known for MazF to form dimeric assemblies in solution (Simanshu et al., 2013; Zorzini et al., 2016), a similar behavior for CalpT was studied *in silico* and *in vitro*. Structural prediction of a $CalpT_{23}$ dimer resulted in reasonable models with good confidence scores (**Figure 5-17 c**). To analyze this by PELDOR spectroscopy, a single cysteine was introduced at position 119 to attach an MTSSL spin label to the protein. If dimerization happens, a defined distance distribution should result from EPR experiments. However, neither without nor with cA$_4$-induced cleavage of CalpT such an observation was made (**Figure 5-17 d**). Additionally, the ability of CalpT to cleave DNA or RNA was tested in collaboration with scientist from Bonn and St. Andrews, UK. Incubation of CalpT-L in presence or absence of cA$_4$ with six different fluorescently labeled single stranded RNA sequences (ssRNAs) did not result in observable degradation of any of

these (**Figure 5-17 f-g**). To test more sequences at once, an RNA cleavage assay was designed that uses a library of ssRNAs to be tested for nucleolytic degradation. To build the ssRNA library, DNA oligos were designed with specific 3'- and 5' ends and 10 random bases in the middle. These oligos were transcribed by T7 RNA polymerase *in vitro* and the transcripts were incubated with CalpT-L in presence or absence of cyclic tetraadenylate, a control reaction was set up with MazF. Incubation of the RNA library with MazF indeed led to a depletion of sequences that contained the MazF recognition site (ACA) (**Figure 5-17 e** (left)). This proved that the assay worked in general. However, for CalpT-L, no signs for nuclease activity could be identified (**Figure 5-17 e** (right)).

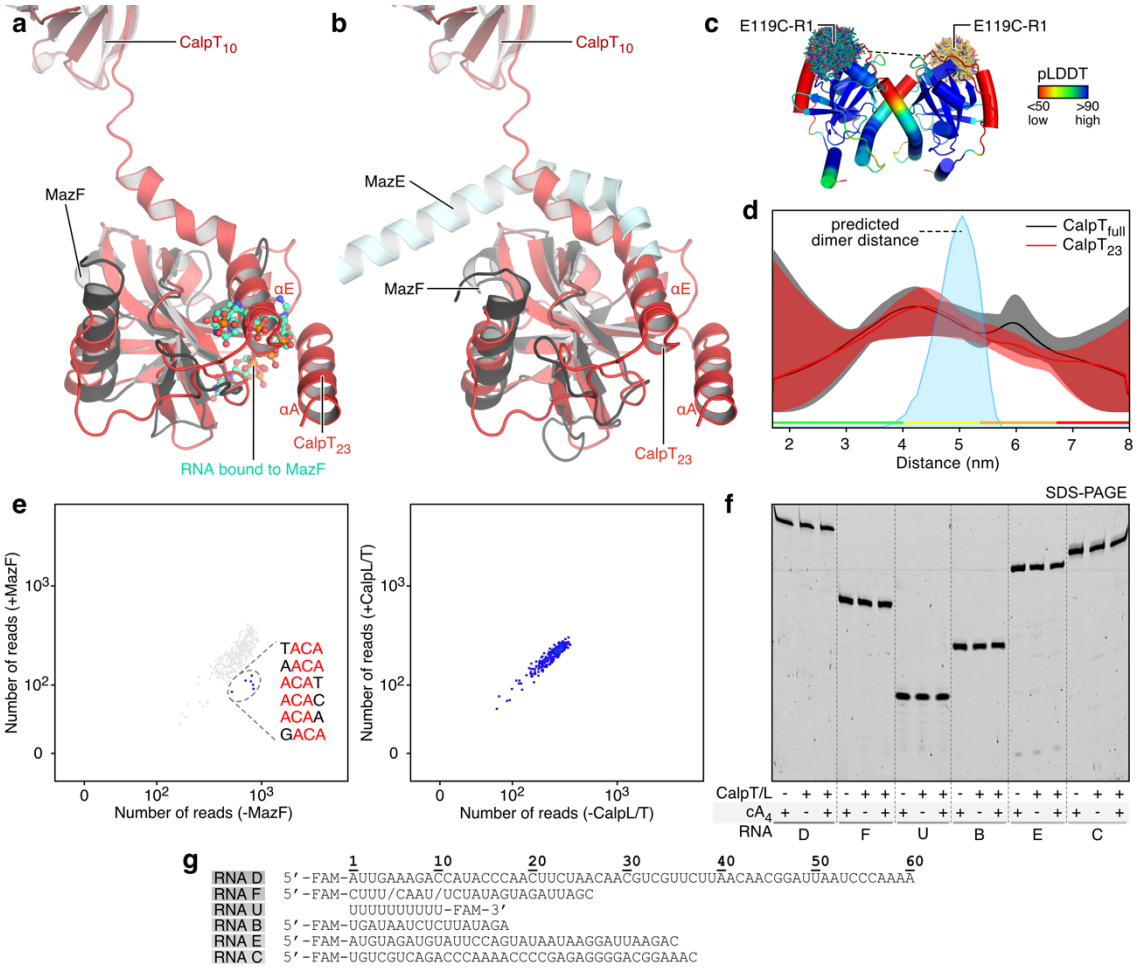


Figure 5-17: Investigating the structural homology of CalpT to a MazF toxin.
a) *Superposition of the CalpT alphafold model (red) with one monomer of MazF in complex with ssDNA (black/cyan, PDB ID: 5cr2). At the position where the nucleotide binds to MazF, the long αE helix of CalpT is located.* ***b)*** *Same as in a) but for CalpT and the MazE/F complex (black, palecyan, PDB ID: 4me7). The MazE helix is located at the same position as the αE helix of CalpT.* ***c)*** *Predicted structure of a possible CalpT dimer including an MTSSL spin label. The two protein chains are colored according to the prediction confidence as indicated (pLDDT, predicted local distance difference test (Tunyasuvunakool et al., 2021)).* ***d)*** *Consensus distributions determined by EPR spectroscopy of CalpL/T-E119C-R1 in presence (red) and absence (gray) of cA_4. The predicted distance distribution of a $CalpT_{23}$ dimer is shown in blue, it was calculated with mtsslWizard (Hagelueken et al., 2012).* ***e)*** *Illumina sequencing results to search a RNA library for cleaved sequences after incubation with MazF (left) and $CalpT_{23}$ (right). The proteins were incubated with a ssRNA library containing 10 random bases. After reverse transcription of the ssRNA these were sequenced. Compared to a control reaction without protein, sequences containing the target site were depleted. By this, the known recognition motif of MazF (ACA) could be found (left) while for the experiment with CalpL/T, no off diagonal sequences were identified.* ***f)*** *Fluorescence image of an SDS-PAGE to determine ribonuclease activity of CalpT after proteolytic cleavage. Six fluorescently labelled RNA sequences were tested (listed in g)). After 30 min of incubation at 60 °C, no cleavage was observed.* ***g)*** *Sequences of the RNA substrates tested in f).*

5.4.2 CalpT forms a stable complex with the putative ECF σ factor CalpS

Since the initial assumption that CalpT has nuclease activity like its structural homolog MazF could not be confirmed, we pursued another possible function of CalpT. Analysis of the gene neighborhood conservation as previously done to find CalpT as target for CalpL (see **Section 5.1.3**), drew our attention to a third conserved protein encoded by the operon (224 amino acids, 26.5 kDa, UniProt ID: B2V8L7). The encoded protein product, CalpS, showed structural similarities to ECF σ factors, which mediate the promoter identification by RNA polymerase (RNAP) and play a key role in transcriptional regulation during cellular stress (Paget, 2015; Singh et al., 2011). The AlphaFold model of CalpS is presented in **Figure 5-18 a**, and shows a protein comprised of two domains which are linked by a flexible loop. The individual domains have structural similarities to the σ_2- and σ_4- subunits which are responsible for specific binding of the DNA

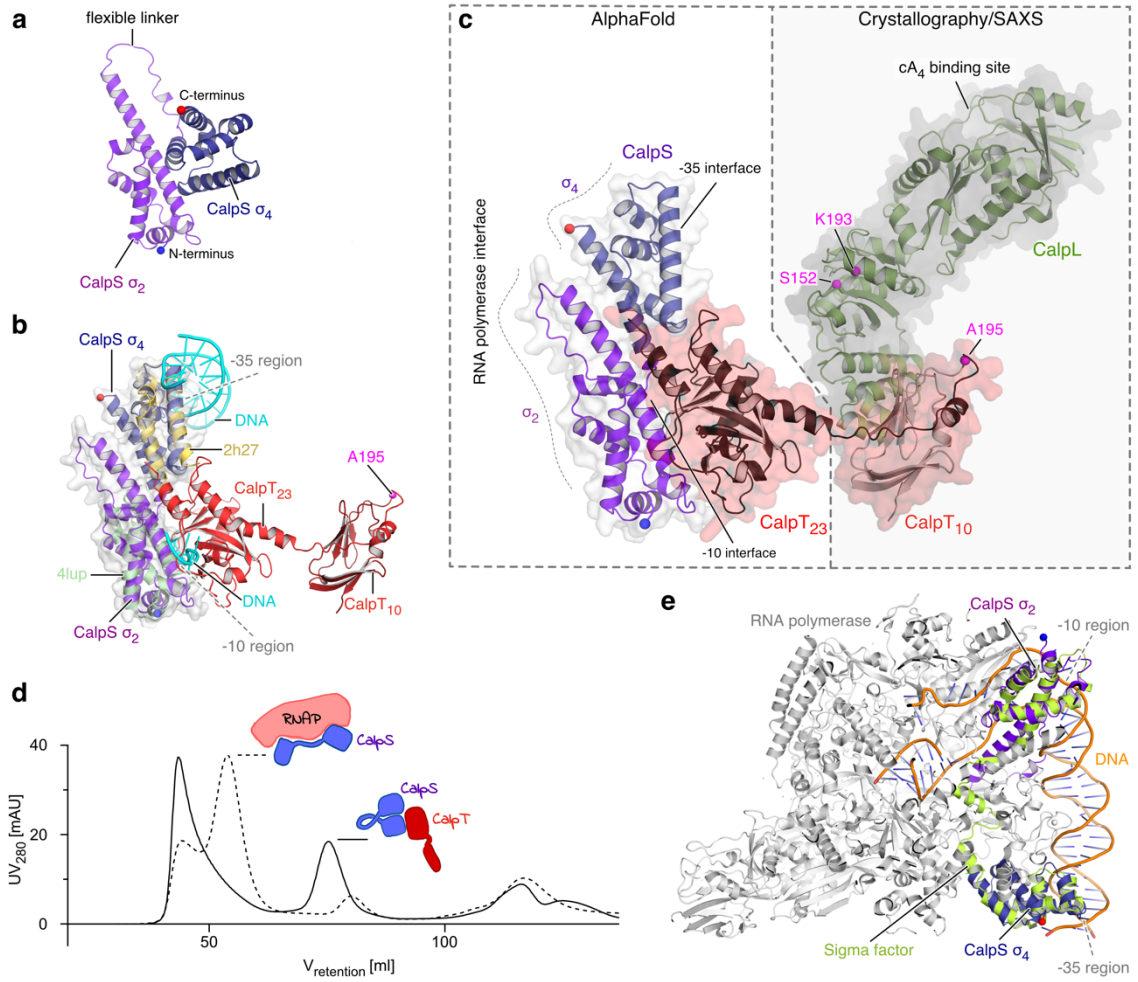


Figure 5-18: CalpT forms a complex with the ECF sigma factor CalpS.
a) *AlphaFold model of CalpS. The N-terminal and C-terminal ends are indicated as well as the σ_2- and σ_4 subunits.* ***b)*** *Predicted CalpS-T complex structure. All protein domains are labeled and the CalpT cleavage site A195 is highlighted. To identify the DNA binding sites of CalpS, the individual subdomains were superimposed with the experimental structures of sigma factor σ_2 and σ_4 domains bound to DNA fragments (used PDB IDs: 2h27 and 4lup).* ***c)*** *Model of the tripartite protein complex CalpS-T-L. The $CalpT_{10}$-L part consists of experimentally determined structures while the CalpS-T_{23} part was generated by AlphaFold. Proteins and characteristic sites of them are indicated.* ***d)*** *SEC chromatograms from the purification of CalpS (dashed line) and the result from CalpS CalpT coexpression. The retention volume is depicted on the x-axis and the UV absortption at λ=280 nm on the y-axis.* ***e)*** *Superposition of the σ_2- and σ_4 subunit of CalpS (blue-purple) on a complex structure of a sigma factor (green) in complex with RNA polymerase (gray) and DNA (orange) (PDB ID: 5zx2). The individual components are labeled. CalpS was cut in the flexible linker region to enable individual alignments of the two domains.*

at the -10 and -35 region, respectively. These regions are indicated in **Figure 5-18 c**, which depict a predicted complex of CalpL, CalpT, and CalpS. This prediction showed, that $CalpT_{23}$ could bind to CalpS to regulate its function in an anti-σ factor manner, namely by blocking the DNA binding site of the σ factor and trapping it in a conformation that does not allow binding to RNAP (**Figure 5-18 b, e**).

To test whether the predicted complexes also exist *in vitro*, a synthetic, codon-optimized variant of *calpS* was cloned into a pBADHisTEV expression vector containing an N-terminal His_6 tag and a TEV cleavage site. However, different constructs were designed in different expression vectors, including an MBP-fusion protein to increase the expression yield and help to keep the protein soluble and a protein without an affinity tag for coexpression studies. The protein was expressed in *E. coli* BL21 (DE3) and *E. coli* C43 cells. Although the expression yielded only small amounts of protein, peptide mass fingerprinting confirmed its accuracy. Strikingly, mass spectrometric analysis showed that CalpS coelutes with RNAP from the *E.coli* expression system (Sequence identities: RpoA=39.6 %, RpoB=50.4 %, RpoC=52.7 %, RpoZ=29.9 %). In contrast, coexpression of CalpS with CalpT prevented copurification of CalpS with RNAP (**Figure 5-18 d**). However, when CalpS was copurified with RNAP, CalpS was still able to bind CalpT and, mediated by the anti-σ factor, also CalpL. It was also observed, that CalpT could still be proteolytically cleaved by CalpL when in complex with CalpS and RNAP (**Figure 9-4**). These experimental data confirmed the hypothesis that CalpL, CalpT and CalpS form a complex that prevents the σ factor, CalpS, from binding RNAP and thus, from transcriptional regulation. To further strengthen the structural model of the CalpS-T complex, some charge-flip mutants of residues that are part of the predicted complex interface were generated. And indeed, for CalpS R80E, no complex formation with CalpT was observed in analytical gel filtration experiments (**Figure 5-19**).

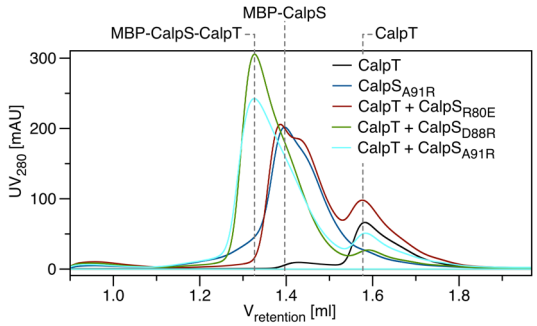


Figure 5-19: Mutational CalpS-T complex analysis. *Analytical SEC of CalpS-CalpT interface mutants. The UV absorption is plotted against the retention volume. Dashed lines indicate the retention volume for the individual proteins and their complex as annotated. The color coding is described in the legend.*

5.4.3 Experimental validation of the full-length CalpT AlphaFold prediction

All previous attempts to crystallize CalpT in its full length did not work out, neither for the protein alone nor in complex with CalpL or CalpS. Only the C-terminal part, $CalpT_{10}$, was observed by co-crystallization of CalpT with CalpL. Co-crystallization trials without addition of cA_4, yielded crystals after more than 50 days. Interestingly, all crystals that diffracted well enough for further analysis only contained CalpL and $CalpT_{10}$. However, we did not surrender trying to solve the structure, and finally, succeeded with co-crystallization of CalpT with a VHH antibody.

The primary idea to generate VHH antibodies against the CalpL–CalpT complex arose from the thought they might stabilize the CalpT-L complex, compete for cA_4 binding, or bind the N-terminal domain of CalpT and provide hints for its function. For the immunization, the protease-dead mutant S152A of CalpL and wild-type CalpT were

preincubated and run on gel filtration, just as for co-crystallization trials. Generation, expression and purification of VHH antibodies was done in a similar way as described in **Section 2.1.1.2**. By biochemical analysis, 2 CalpL targeting VHHs (NbS020-NbS021) and 4 CalpT targeting VHHs (NbS022-NbS025) were identified. Although the analysis is not discussed further below, it can be summarized that the two CalpL-specific antibodies bind to different epitopes, while the four CalpT-binding VHHs target overlapping regions. All nanobodies bound their antigen with nanomolar affinities, or even stronger and some of the them showed interesting effects, which need further analysis.

Individual co-crystallization of all four CalpT specific VHHs with their target protein was only successful in the case of CalpT–NbS023, which is astonishing as they showed a similar binding behavior towards the same target region on the C-terminal CalpT domain. Crystals were obtained after 3 days of incubation at 20 °C in condition G4 of the Morpheus crystallization screen (Molecular Dimensions) (**Figure 5-20 a**). The crystals were harvested without cryo protection and a diffraction dataset was recorded at DESY, Hamburg (**Figure 5-20 b, c**). The structure was solved at 2.1 Å resolution by molecular

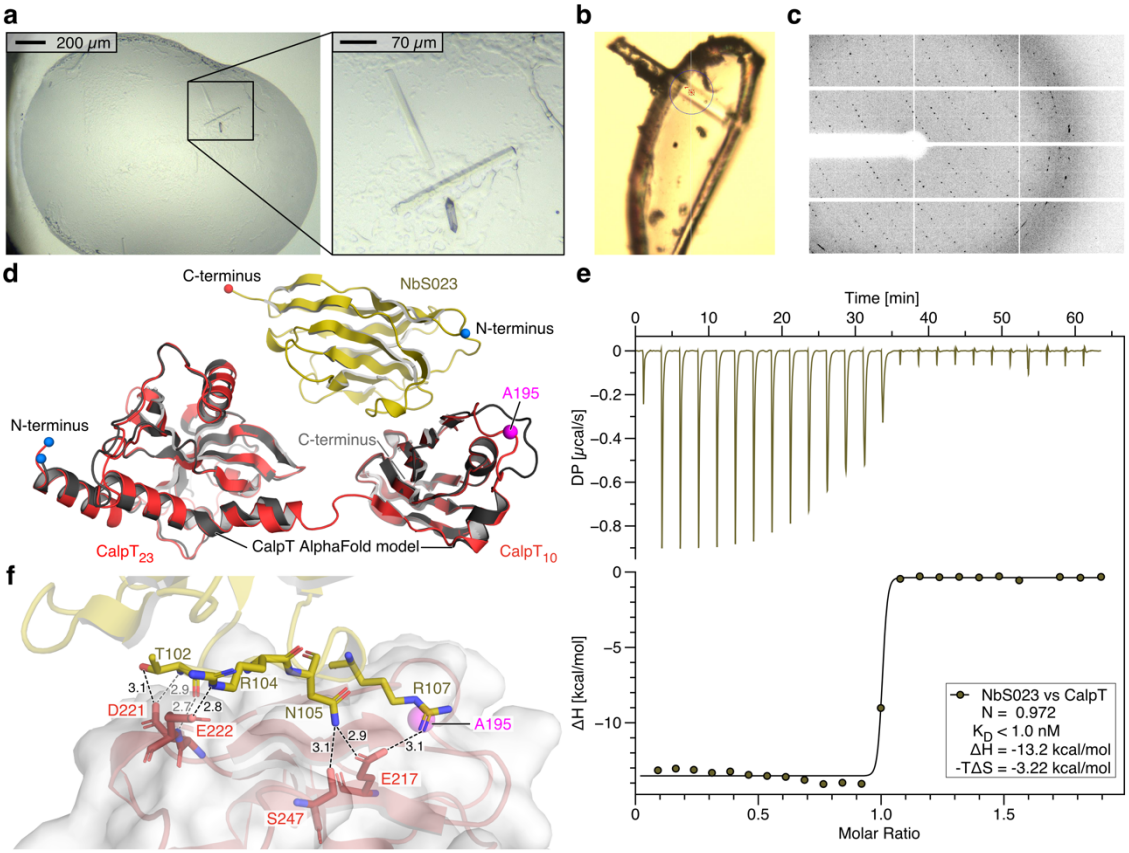


Figure 5-20: VHH antibody stabilizes CalpT for crystallization.
a) *Pictures of CalpT–NbS023 complex crystals (left) including a magnification (right). Scale bars are included in the top left corner.* ***b)*** *Image of the crystal inside the cryo loop. The position where the Xray beam hit the crystal is indicated by a blue circle.* ***c)*** *Diffraction image obtianed from a test diffraction.* ***d)*** *Structural model of the CalpT–NbS023 complex (red and yellow, respectively). The two domains of the AlphaFold model (black) were superimposed onto the experimental structural individually. All protein domains, termini and the CalpT cleavage site A195 (magenta) are indicated.* ***e)*** *ITC thermogram from the titration of NbS023 to CalpT (top) and the resulting binding curve (bottom). For the thermogram, the differential power (DP) was plotted against the reaction time, and for the binding curve, the enthalpy (ΔH) against the molar ratio. Resulting binding data are shown in the box in the lower right corner.* ***f)*** *Magnification of the interaction interface of CalpT (red) and NbS023 (yellow). Residues that form electrostatic interactions are labeled and shown as sticks, interactions are indicated by dashed lines. Bond lengths are given in Ånström.*

replacement using the CalpT10 structure, the CalpT23 AlphaFold model and the NbS001 structure as search models. Like biochemical analysis already suggested, the VHH antibody was observed to bind to the C-terminal domain of CalpT. Both CalpT domains showed a high similarity to the AlphaFold model (r.m.s.d. for $CalpT_{23}$=0.486 over 133 Cα-atoms, r.m.s.d. for $CalpT_{10}$=0.329 over 68 Cα-atoms) (**Figure 5-20 d**). The two domains of the structural model were superimposed independently after cutting the flexible connection loop. The high affinity binding (see **Figure 5-20 e**), was observed to be a result of a 630.1 $Å^2$ large interaction interface. The most electrostatic interactions were identified between the CDR3 region of the nanobody and $CalpT_{10}$ residues E217, D221, E222, and S247 (**Figure 5-20 f**). However, an additional, albeit distant, interaction was identified between the VHH (E89) and the CalpT23 domain (K118, E119) by analyzing the structure with the LigPlot+ software (Laskowski & Swindells, 2011), as well as with the PDBePISA tool (Krissinel & Henrick, 2007) (**Figure 5-21 a**). The latter one was also used to analyze all interfaces within the crystal. The largest area (1329.6 $Å^2$) occurred between two CalpT monomers and stabilizes the flexible linker region within the crystal from one side. The other side of the same loop was stabilized by a NbS023 symmetry mate by another large area (682.2 $Å^2$) of crystal contacts (**Figure 5-21 b**). These contacts which are based on the crystal packing, might be the reason for a successful crystallization approach exclusively with NbS023.

Since The VHH antibody binds to CalpT10 and the interaction surface overlaps with that of the CalpT–L complex, we investigated the effect of all VHHs on the protease activity. Therefore, CalpT and CalpL, respectively, were preincubated with one of the six VHHs, and subsequently, the other components were added to the reaction mixture. The protease activity assay was performed at 37 °C and did not show significant effects for NbS020 and NbS022. For NbS021, which targets CalpL, an increased proteolytic activity

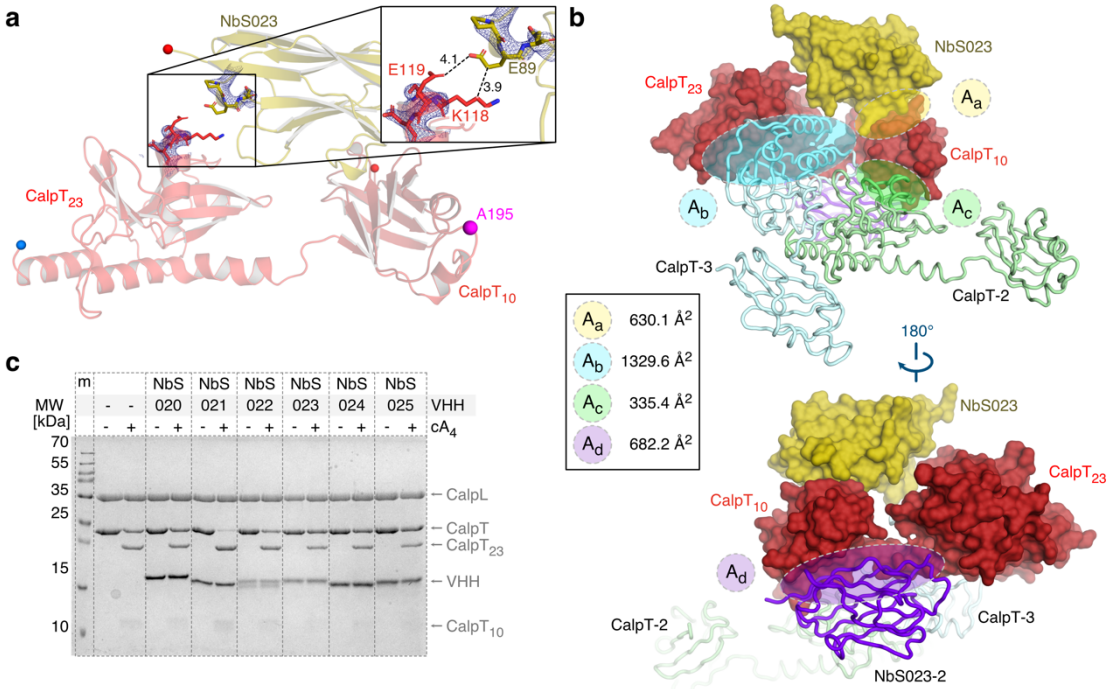


Figure 5-21: Crystal packing analysis of CalpT-NbS023 and activity assay.
a) *Overview on the CalpT–NbS023 complex structure showing an additional interaction between $CalpT_{23}$ and the VHH. Residues are labeled within the magnification in the top right corner. The distances between the residues are given in Å.* ***b)*** *The complex (surface) and symmetry mates of the individual chains (ribbon) to illustrate the different interfaces A_a, A_b, A_c, and A_d (highlighted by colored speheres). The areas of the interfaces are annotated, all components are labeled.* ***c)*** *SDS-PAGE analysis of samples from a protease activity assay after preincubation with the different VHHs.*

was observed. However, this effect could not be further investigated, as first attempts to crystallize the complex failed. Preincubation of CalpT with NbS023, NbS024 or NbS025 led to a slightly decreased proteolytic cleavage. Which might be caused by the overlapping interaction interfaces of the VHHs and the protease. However, these findings require further and more detailed investigation.

6. Discussion Part II

Structural and functional characterization of the CRISPR-associated Lon protease (CalpL) from *Sulfurihydrogenibium* sp. YO3AOP1 revealed that CalpL contains a typical Ser-Lys catalytic dyad (Amerik et al., 1991; Kirthika et al., 2023) and is activated by cyclic tetraadenylate (cA_4)-induced oligomerization.

The binding of cA_4 to SAVED domain-containing effector proteins was previously known only for CBASS systems (N. Liu et al., 2022), but within the last two years more and more CRISPR systems have been reported to be activated in a similar manner (Hogrel et al., 2022; Steens et al., 2024). To unravel the role of CalpL in bacterial immunity, we identified conserved genes within the same operon as CalpL and tested their interactions using *in silico* approaches, biochemical and biophysical methods, as well as X-ray crystallography.

6.1 Analysis of the gene neighborhood conservation of CalpL

Functionally related bacterial proteins are often encoded by genes that are localized in close proximity in the genomic DNA (Rogozin et al., 2002). To find possible candidates for protease targets, we used the WebFlags server for gene neighborhood conservation (Saha et al., 2020). This approach led to identify CalpT and CalpS, an anti-sigma factor – sigma factor pair that controls bacterial transcription upon activation by proteolytic cleavage. Apart of that, we observed CalpL homologs that showed significant differences in their sequence length, or the presence of *calpT and calpS* genes (**Figure 6-1**).

6.1.1 Subtype classification according to the CRISPR locus

6.1.1.1 Absence of Cas1 and Cas2 and presence of a signature gene

By analyzing the gene neighborhood of organisms that encode CalpL homologs, it becomes evident that some do contain *cas1* (magenta) and *cas2* (blue) genes, which are important for spacer acquisition (Barrangou, 2013; Nuñez et al., 2014) and some do not, such as *Sulfurihydrogenibium* sp. (see **Figure 4-9**; **Figure 6-1**). CRISPR loci that lack those genes are most often assigned to the subtypes III-b or IV (Makarova & Koonin, 2015). Such CRISPR systems, depend on the astonishing versatility of alternative adaption pathways (Majumdar et al., 2015; Mohanraju et al., 2016; Nuñez et al., 2015; Peters et al., 2017). The presence of the type III signature gene, *cas10* (orange), in the CRISPR locus and the absence of the adaption module suggest the system to be of subtype

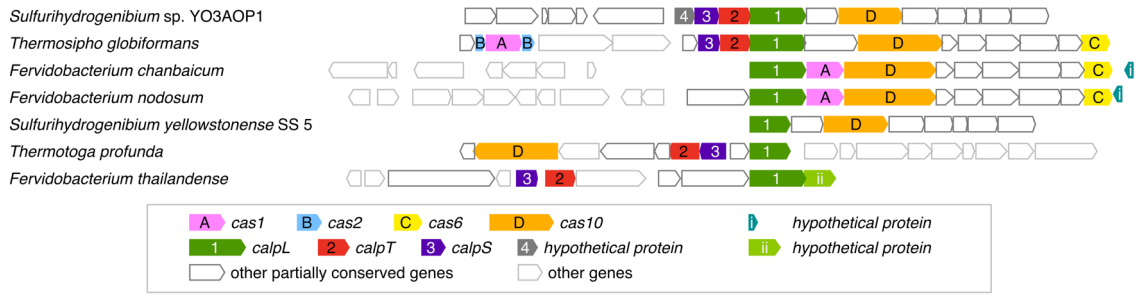


Figure 6-1: Gene neighborhood analysis revealed interesting CalpL homologs.
The conservation of the genomic context of calpL among different organisms was analyzed using the webFlaGs online tool (Saha et al., 2020). Like in Figure 4-9, Sulfurihydrogenibium *sp. YO3AOP1 is shown on top and the CalpL encoding gene and its homologs are shown in green (1). The genes encoding for CalpT and CalpS are indicated in red (2) and purple (3), respectively. Further conserved ORFs are highlighted and indicated in the legend.*

III b (Makarova et al., 2015; Makarova & Koonin, 2015). Furthermore, the locus encodes Cmr proteins which are important in this subtype of CRISPR immunity (Hale et al., 2009, 2012; Makarova et al., 2015).

6.1.2 CalpL homologs in other organisms show similarities and differences

During the gene neighborhood analysis, we recognized that some organisms lack genes for the anti-sigma factor–sigma factor pair, *calpT* and *calpS*, and furthermore, some CalpL homologs seemed to differ in their sequence length. Some of these organism and their *calpL*-containing operons are depicted in **Figure 6-1**.

6.1.2.1 Organisms that contain *calpL* genes but lack CalpS and CalpT encoding genes

Upon taking a closer look at systems that lack the *calpT* and *calpS* genes, we recognized a gene ‚i' (teal), downstream of *cas6* (yellow), that was exclusively present in these organisms. We then predicted the complex structures of these hypothetical proteins with the corresponding CalpL homolog from *Fervidobacterium chanbaicum*, *Fc*CalpL, and *Fervidobacterium nodosum*, *Fn*CalpL, respectively (**Figure 6-2 a, b**). In each case, the AlphaFold2 prediction of the complex showed binding of the small hypothetical protein to the N-terminal domain of CalpL with the same interface as $CalpT_{10}$. Both protease homologs contain a Ser-Lys catalytic dyad and showed evidence of cA_4-induced oligomerization in predictions with the AlphaFold3 server (**Figure 6-2 c,**

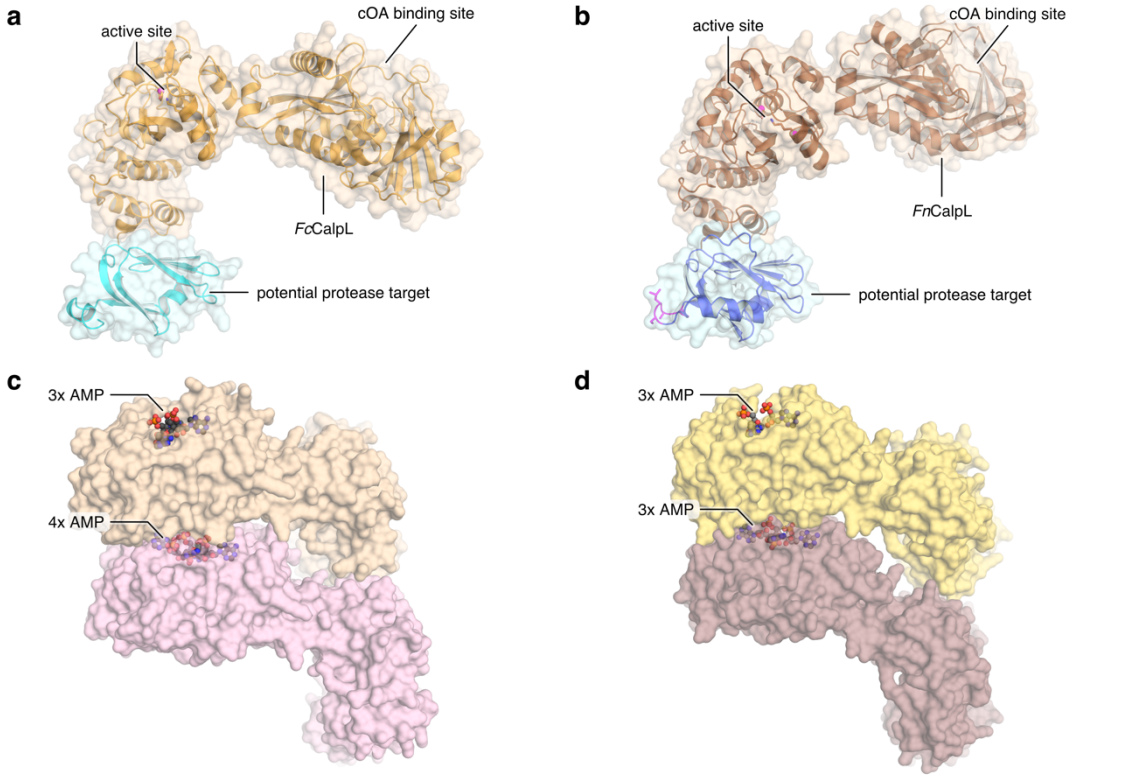


Figure 6-2: CalpL homologs are predicted to bind a different potential protease target.
a) *Predicted heterodimeric complex of* Fc*CalpL (tan) and the hypothetical protein encoded (turquoise) by gene 'i' (**Figure 6-1**, teal) from* Fervidobacterium chanbaicum. *The cOA binding site, and the Ser-Lys catalytic dyad are indicated.* ***b)*** *Same as a) but for the protein from* Fervidobacterium nodosum. ***c)*** *AlphaFold3 (Abramson et al., 2024) model of a cOA-induced dimer of* Fc*CalpL. Eight AMP molecules were used for the prediction to mimic cOA.* ***d)*** *Same as a) but for the protein from* Fervidobacterium nodosum.

d) (Abramson et al., 2024). At the position where CalpT is cleaved (H193-VAST-S198), the sequence of the potential protease target reads ILPS (*F. chanbaicum*) and VALS (*F. nodosum*), respectively. In contrast to the CalpS-T-L system from *Sulfurihydrogenibium* sp. YO3AOP1, proteolytic cleavage would not lead to the release of the N-terminal T23 domain, but only to the release of a small peptide sequence at the C-terminus of the protein. Peptides are known to be important in eukaryotic immune processes (Quax et al., 2013; Sui & Guo, 2021) and are also utilized by phages and some bacteria for communication purposes (Stokar-Avihail et al., 2019; Sturme et al., 2002). Thus, the protease-initiated release of a small peptide sequence could be part of the CRISPR-mediated immune strategy of these organisms.

Interestingly, some genes further downstream, CalpS and CalpT encoding genes, are found and both the protease–anti-sigma factor complex and the anti-sigma factor–sigma factor complex were predicted with high confidence scores. So either, the protease is multi-selective and capable of cleaving both targets to orchestrate a broader immune response, or the system evolved from one another and the two organisms still carry the genes for both targets. In any case, these analyses provided interesting aspects which might be worth studying in detail.

6.1.2.2 CalpL homologs with significantly shorter amino acid sequences

In the CRISPR locus of *Sulfurihydrogenibium yellowstonenense* a significantly shorter *calpL* variant can be found. A structural model showed that the encoded protein is very similar to CalpL from *Sulfurihydrogenibium* sp. YO3AOP1 (*Ss*CalpL), but lacks the N-terminal domain (**Figure 6-3 a**), which we found to be essential for the binding and subsequent proteolytic cleavage of CalpT. According to the WebFlags analysis, this locus does not encode an anti-sigma factor–sigma factor pair (**Figure 6-1**). Another interesting observation was made when examining the type III locus of *Thermotoga profunda*. Again, a CalpL variant that misses the N-terminal domain was predicted (**Figure 6-3 b**, dark gray). Since the locus contains a *calpT* and *calpS* variant, we wondered how an efficient cleavage of the anti-sigma factor could be achieved without the N-terminal target-binding domain. Upon having a closer look at the short ORF directly upstream of the predicted shorter *calpL* variant, I recognized a similarity to the missing N-terminal domain. An AlphaFold3 prediction including the translation of this ORF yielded a full-length CalpL structure (**Figure 6-3 b**, slate-blue). A further structure prediction suggested that this full-length homolog is able to bind the CalpT homolog, which furthermore includes a possible cleavage sequence (VAAT) within a compatible loop region (**Figure 6-3 b**, cyan). In the case of *T. profundum*, the hypothetical shorter *calpL* gene could be a prediction error, or could also be a product of a gene fission. Such events occur to a significant extent in thermophillic organisms, like *T. profundum*, and often result in genes that are split in between specific domains (Snel et al., 2000). If a CalpL protein that consists of two individual peptide chains is still functional or if the encoding genes are only leftovers from exchanged genetic material requires further investigation. In *S. yellowstonense*, the absence of genes encoding a sigma-factor–anti-sigma factor pair argues for a non-functional evolutionary remnant (Gil & Latorre, 2012; Molina & Nimwegen, 2008). In each case, the existence of significantly different but yet similar gene variants highlights the exchange and adaption of genetic material between different prokaryotic organisms (Diard & Hardt, 2017; Molina & Nimwegen, 2008).

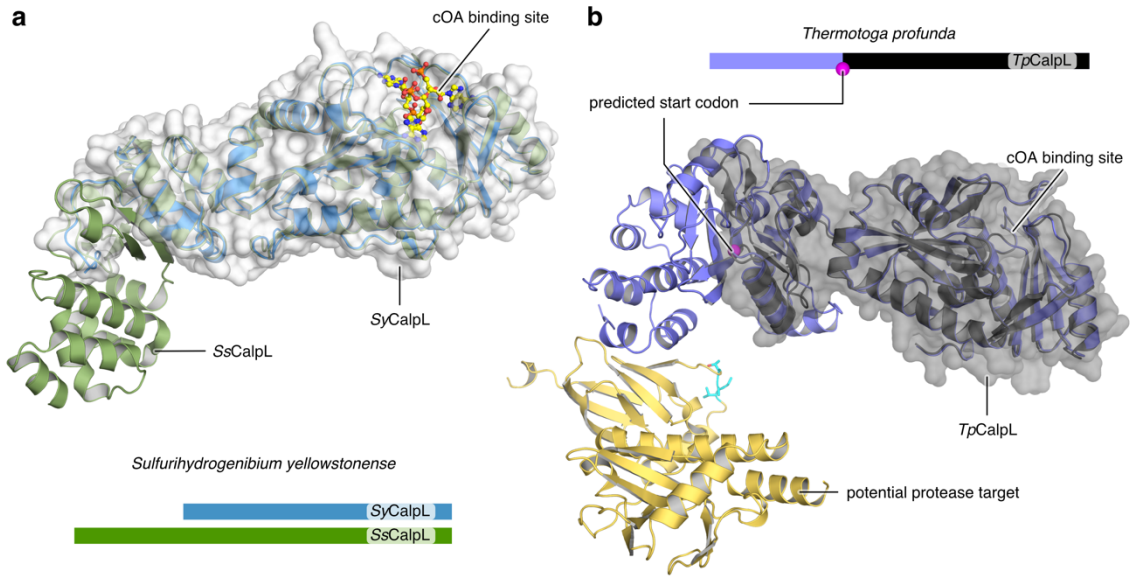


Figure 6-3: CalpL variants that are predicted to contain only the protease- and SAVED domain. ***a)*** *A structural model of a CalpL homolog from* Sulfurihydrogenibium yellowstonense *(blue cartoon, gray surface), aligned to* Ss*CalpL (green cartoon), lacks the N-terminal domain, that is essential for binding of the protease target.* ***b)*** *The CRISPR locus of* Thermotoga profunda *also encodes a homolog protease that misses the N-terminal domain according to* in silico *predictions (dark gray cartoon and surface, a magenta sphere marks the sequence start). When the translation of the sequence directly upstream of that* calpL *gene is included for the structure prediction (slate-blue cartoon), the product becomes very similar to full-length* Ss*CalpL. Notably, the full length construct is predicted to bind a CalpT homolog (yellow) that contains a possible protease cleavage site (cyan sticks). Colored bars represent the sequence lengths of the constructs.*

6.1.2.3 A CalpL homolog that shows evidence for a different mode of action

When examining the *calpL* variant from *Fervidobacterium thailandense* it becomes clear that two adjacent gene products share sequence similarity (29.6 % amino acid identity) but differ significantly in their length (495 vs. 259 amino acids) (**Figure 6-1** '1' and 'ii'). Structural models of the two hypothetical proteins revealed that the smaller protein (UniProt ID: A0A1E3G190) contains only a SAVED domain (hereafter: *Ft*soloSAVED) whereas the larger protein (UniProt ID: A0A1E3G168) (hereafter: *Ft*CalpL) is a structural equivalent of *Ss*CalpL. Notably, an AlphaFold3 prediction including eight AMP molecules (to mimic cOA) and both protein sequences showed a heterodimer, held together by four AMP molecules (**Figure 6-4 a**), just like the predicted *Ss*CalpL oligomer. The heterodimeric complex contains one soloSAVED protein on top of the cOA binding site (binding four AMP molecules) of *Ft*CalpL. Investigation of the interacting surface areas by focusing on the electrostatic surface potential showed, similar to the analysis of SsCalpL (see **Figure 5-4 b**), compatible interfaces in terms of charge and size (**Figure 6-4 b**). Interestingly, the *in silico* analysis of possible homodimeric complexes yielded a reliable model for a *Ft*soloSAVED dimer, but failed to predict a dimeric *Ft*CalpL. One possible reason for this could be a non-compatible positively-charged patch on the *Ft*CalpL downside, which does not match the size of the opposing cOA binding site, whereas this requirement seems to be fulfilled for *Ft*soloSAVED (**Figure 9-5**).

At first glance, it remains unclear how such a system might be able to proteolytically cleave its target, as we proved the oligomerization to be necessary for an effective proteolytic cleavage (see **Section 5.2.2**). But 'in trans' cleavage could also occur without

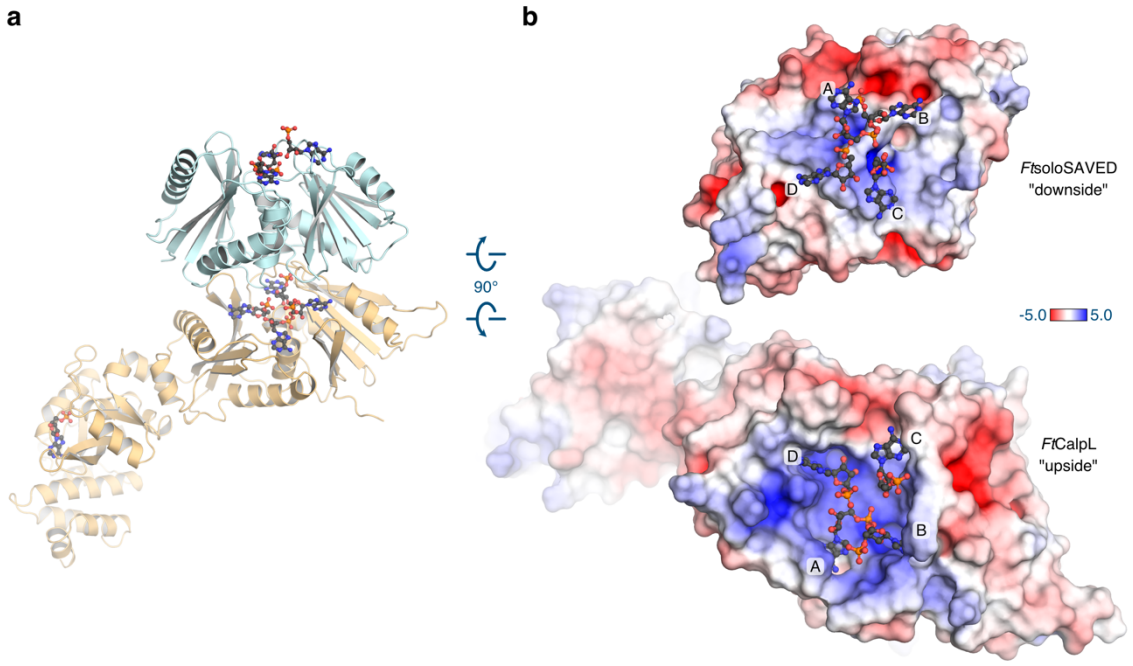


Figure 6-4: Dimer of CalpL and a SAVED-only variant of it, encoded by two neighboring genes. ***a)*** *AlphaFold3 model of the heterodimeric complex from CalpL (wheat) and soloSAVED (cyan) from* Fervidobacterium thailandense *held together by four AMP molecules that mimic cyclic oligoadenlyte.* ***b)*** *'Open book' visualization of the two interacting proteins to highlight the interfaces. The bases of the four AMP molecules are annotated to highlight the complementary regions. Note that the four bridging AMP molecules were copied to be shown on both sides The surface is colored according to its electrostatic potential as indicated.*

oligomerization, it might be only essential that the target peptide accesses the protease active site. The lower the affinity of the peptide chain to the active site and the lower the local concentration of them two, the less probable such an event becomes. However, there is evidence that proteolytic activity even occurs in the absence of cA_4

in case of *Ss*CalpL. On the one hand, we observed cleavage of CalpT after several months in crystallization trials of the CalpL-T complex in absence of the second messenger. On the other hand, at higher concentrations in long-term protease activity assays ($t_{incubation}$>1 h), sometimes a faint band at the height of $CalpT_{23}$ was observed (see **Figure 5-10 d**, **Figure 5-12 c**, **Figure 5-16 d**). Thus, it appears that *Ss*CalpL and its target have chosen a way to increase the local concentration by oligomerization in order to optimize the proteolytic efficiency. However, it remains unclear whether cA_4 has another activating impact on the protease besides inducing oligomerization, for example by triggering conformational changes. In contrast to such a mode of activation, other known Lon-proteases are independent of activators, but rely only on the correct geometry of the catalytically active residues (Chung & Paetzel, 2013; Paetzel et al., 2002; Wlodawer et al., 2022). The stacking of soloSAVED on top of CalpL may be required for nucleophilic cleavage of the second messenger molecule, as it was recently shown that the cooperation of two stacked SAVED domains is essential for cOA degradation (Binder*, Schneberger* et al., 2024; Smalakyte et al., 2024). Degradation of these oligonucleotide-based second messengers is in turn an important 'off-switch' to return from the bacterial 'infected state' to the 'normal state' (Athukoralage et al., 2018, 2019; Athukoralage & White, 2021).

6.1.2.4 Another CalpL homologs shows evidence for being functional as a monomer

A similar strategy seems to be used by *Aquifex* sp. that encodes a CalpL variant with an additional C-terminal domain, which we discovered by a FoldSeek search (Kempen et al., 2024) using our CalpL structure as input. The N-terminal domain, as well as the protease active site and the SAVED domain share the same structural fold and 28 % amino acid identity. Interestingly, the SAVED domain contains a positively charged potential cOA binding site, but no positively charged surface patch on the opposite (**Figure 6-5 a**). Instead, the additional C-terminal domain, which is connected to the SAVED domain by a flexible loop, contains such a positively charged region. Based on this features, one could imagine that this domain could cover the cOA-bound SAVED domain like a lid. Another FoldSeek search revealed a structural similarity of this additional domain to one monomer of an anti-CRISPR viral ring nuclease (PDB ID: 6scf). This ring nuclease allows viruses to evade type III CRISPR immunity by specifically binding and degrading cA_4 (Athukoralage et al., 2020). Thus, this ring nuclease domain may allow the enzyme to cleave the second messenger molecule without oligomerization. However, as discussed in **Section 6.1.2.3**, it remains puzzling how the monomeric protease is activated by cOA.

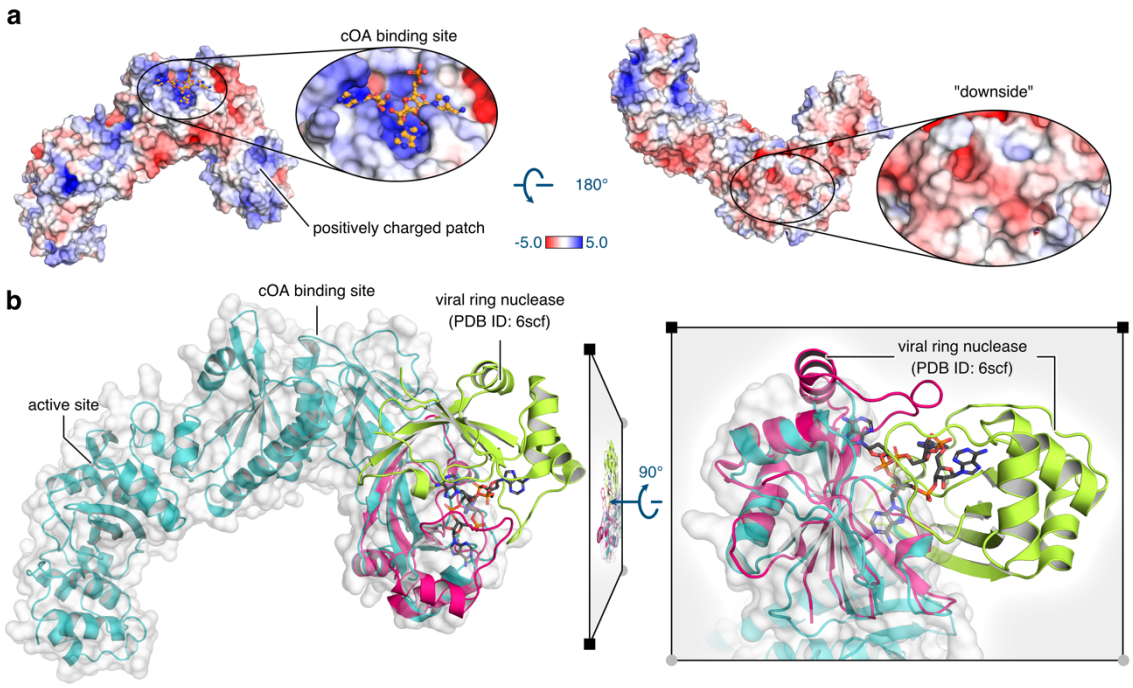


Figure 6-5: Structural investigation of a CalpL variant that harbors an additional domain. ***a)*** *Surface representation of an AlphaFold3 model of a CalpL variant from* Aquifex *sp. (*As*CalpL) showing the 'upside' (left) and the 'downside' (right). A magnification of the potential cOA binding site is included as well a magnification of the opposing surface region. The surface was colored according to its electrostatic potential.* ***b)*** *Cartoon representation of the structure shown in a) (teal) with indicated cOA binding site and protease active site. The crystal structure of a dimeric viral ring nuclease (PDB ID: 6scf; magenta, lime) was superimposed onto the C-terminal domain of* As*CalpL. The right-hand side shows a more detailed view onto the superposition. The viewing direction and the rotation of the projection is indicated.*

6.2 Mechanistic details of the CalpS-T-L mediated immune response

6.2.1 Proteolytic cleavage by the Lon protease, CalpL

Our results highlighted important requirements for a functional protease. Mutation of either residue of the Ser-Lys catalytic dyad to an alanine resulted in a loss of protease activity, consistent with previous analyses of Lon proteases (Amerik et al., 1991; Chung & Paetzel, 2013; Wlodawer et al., 2022). By mutational analysis, we showed that the 3-dimensional substrate channel on the surface of CalpL, terminating at the protease active site, determines substrate specificity to a certain extend. Thus, the unstructured loop of the protease target must be complementary in structure and charge. This mechanism is similar, although not as extensive, to that reported for a Lon protease that exclusively degrades misfolded proteins by recognizing unstructured polypeptides by burying the protease active site in a hydrophobic core of an oligomer (Cha et al., 2010; Gur & Sauer, 2008).

In addition, our data revealed that the N-terminal CalpL domain defines the absolute substrate specificity by providing a binding site for $CalpT_{10}$. The interaction of these two domains is essential for target recognition and cleavage. These findings are in line with previously published analyses of Lon proteases (Tsilibaris et al., 2006; Tzeng et al., 2021).

Another important aspect of CalpL is the cA_4-induced oligomerization behavior. We have shown, that mutants of CalpL, that are either unable to bind the second messenger or to oligomerize upon cOA binding, have a significantly reduced (up to complete loss) protease activity. However, without a high-resolution structure of the CalpL oligomer, the exact reason for this need for oligomerization remains speculative.

6.2.1.1 Oligomerization of CalpL

The interplay between cA_4-dependent filament formation and the proteolytic activity of CalpL remains the most puzzling aspect. Several Lon proteases assemble into hexamers (Botos et al., 2005, 2019; Cha et al., 2010), but there is no evidence that this quaternary structure is essential for a functional active site. Instead, the crystal structures of a Lon protease from *Yellowtail ascites* virus revealed a monomeric enzyme without indications for the formation of multimers (Chung & Paetzel, 2013). On the other hand, Rudyak et al. reported on a relationship between Mg^{2+} concentration, oligomerization state and activity of the Lon protease from *Mycobacterium smegmatis* (Rudyak et al., 2001). However, the general activation of these enzymes seems to be allosterically induced by the substrate itself (Chung & Paetzel, 2013; A. L. Goldberg et al., 1994). For this allosteric activation, it may only be necessary to "place" the substrate peptide into the protease active site. Chung et al. showed that for the *Y. ascites* Lon protease, the target peptide interacts in an anti-parallel fashion with one side of the substrate binding groove and forms a parallel β-sheet with the other side of the groove. This stabilizes the active site and results in an optimized geometry of the catalytic residues (Chung & Paetzel, 2013). A similar mode of stabilization is also conceivable for CalpL, since the peptide-binding groove and surface architecture of the two different proteases are very similar. If this is the only mode of activation required, oligomerization of CalpL may simply be a strategy to increase the local effective concentration and thus the likelihood of the target peptide entering the protease active site. However, structural models of CalpL filaments,

as well as a recently published experimental structure (Smalakyte et al., 2024), show that it would be very difficult to place the substrate into the peptide-binding groove of the protease underneath without any further conformational rearrangements. If the proteolytic activity of the CalpL homologs from other organisms, which show evidence for being monomeric (*As*CalpL, see **Figure 6-5**) or only oligomerize to achieve cOA degradation (*Ft*CalpL, see **Figure 6-4**), would be proven, this would suggest a different mode of activation, at least for those specific enzymes. To address whether cA_4 binding induces conformational changes, FRET experiments might be appropriate. Moreover, by introducing artificial disulfide bridges to remove interdomain flexibility, i.e. by fixing the absolute orientation of the N-terminal domain towards the protease domain of CalpL (see **Figure 5-4 a**), and testing these mutants for proteolytic activity, valuable information could be obtained.

6.2.1.2 Detailed mechanism of the proteolytic process according to Ser Lys proteases

Although the active site of Lon proteases has been extensively characterized in the past (Chung & Paetzel, 2013; Wlodawer et al., 2022), a detailed mechanism of the underlying chemical processes is lacking in the literature. Based on current knowledge of the catalytic dyad of Lon proteases and the functional mechanism of other serine proteases (Dodson & Wlodawer, 1998; Radisky et al., 2006), I developed a working hypothesis for the individual and reversible chemical reactions, that in sum lead to irreversible hydrolysis of the peptide bond.

The distance between the catalytically active serine and lysine needs to be in the range of a hydrogen bond to enable the coordination of the Ser Hδ by the free electron pair of Lys Nζ. Then, the free electron pair of Oγ can nucleophilically attack the carboxylic

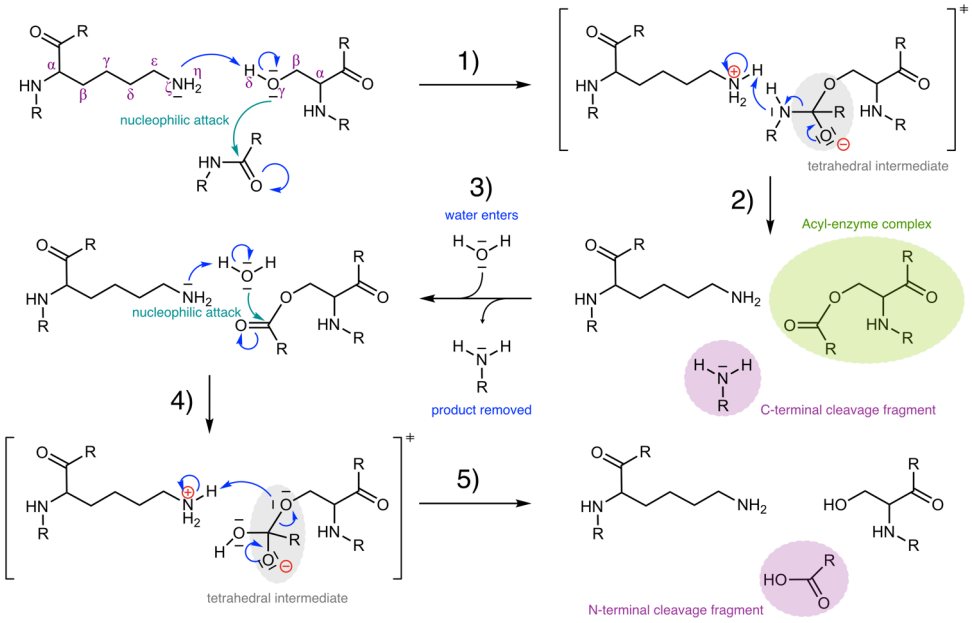


Figure 6-6: Possible mechanism of proteolytic cleavage for Lon proteases.
Mechanism of the enzymatic hydrolysis of a peptide bond by Lon proteases. The atoms of the catalytic residues are labeled by greek letters on the top left. Rearrangement of electrons is indicated by blue arrows, nucleophilic attacks at the substrate are shown by teal arrows, and tetrahedral intermediates are highlighted by grey spheres. The acyl-enzyme complex (green) as well as the C-terminal and N-terminal cleavage fragments (puprple) are higlighted by spheres. The chemical mechanism was drawn using ChemDraw and is based on known mechanisms for other Serine proteases (Dodson & Wlodawer, 1998; Radisky et al., 2006).

carbon of the peptide chain (**Figure 6-6**.1). Rearrangement of electrons via a tetrahedral intermediate leads to the removal of the C-terminal cleavage fragment and formation of an acyl-enzyme complex (**Figure 6-6**.2). Next, the deacylating water enters to nucleophilically attack the carboxyl carbon of the acyl-enzyme complex (**Figure 6-6**.3). Again, this leads to the formation of a tetrahedral intermediate (**Figure 6-6**.4) followed by electron rearrangement releasing the N-terminal cleavage fragment and recovering the catalytic residues (**Figure 6-6**.5). It should be noted that the proposed mechanism shows a deprotonated lysine side chain which is unusual at a physiological pH. However, different scenarios could lead to a deprotonated lysine side chain. The microenvironment of the lysine side-chain can affect the protonation state, for example, in a hydrophobic environment a drastically decreased pK_a value for the amine (Nζ) group of the lysine could be observed (Takayama et al., 2008). A transiently deprotonated lysine could also be stabilized by nearby amino acid residues with Brønsted base characteristics (Wallerstein et al., 2015) or even metal ions (Dudev & Lim, 2002).

While lysine (K193 in CalpL) and serine (S152 in CalpL) form the catalytic dyad, more recent publications have also highlighted the role of a highly conserved threonine- or serine residue (T175 in CalpL) (**Figure 6-7**). The Oγ of this amino acid is in hydrogen bond distance of the lysine Nζ and is proposed to be essential for the coordination of the deacylating water. For the nucleophilic attack on the carbonyl of the ester intermediate, it is essential that the water is placed in hydrogen bond distance and at a suitable Bürgi-Dunitz angle of 107° (Burgi et al., 1973). Interestingly, the threonine could also serve as nucleophile for the initial attack on the peptide chain. In contrast to the nucleophilic attack by the catalytic serine, the threonine mediated attack would occur from the backside of the peptide chain and the deacylating water would have to enter from the other side (Chung & Paetzel, 2013; Paetzel et al., 2002). Although such activity would definitely be less favored in terms of geometry of the intermediate states, it may occur partially. This would explain the minimal proteolytic activity of the CalpL S152A mutant. However, it should be noted that a possible activity of this mutant was only observed in crystallization experiments after an incubation time of t>200 days, and the observation was not validated by mass spectrometry. Furthermore, it remains to be tested whether a T175A mutant is indeed catalytically less active and whether a double mutant of S152A and T175A is completely inactive. To mechanistically analyze the proteolytic process of CalpL in detail, structural analyses with intermediate states would be interesting as presented for Lon protease from *Yellowfin ascites* virus (Chung et al 2013). Time-resolved crystallography might be an interesting experiment (Moffat, 1996), but it would clearly be difficult to obtain CalpL crystals in an active, i.e. oligomeric, state and moreover, to initiate the proteolytic process from the outside.

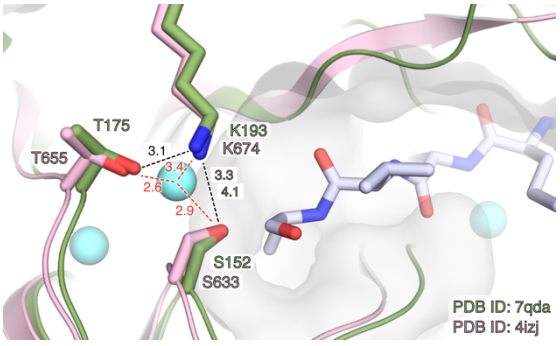


Figure 6-7: Detailed view on the catalytic dyad. *The catalytic residues S152 and K193 of CalpL are superimposed onto the coresponding residues of the Lon protease from* Yellowfin ascites *(4izj). The conserved T175 on the backside is indicated and the distances of all residues to the proposed deacylating water are given in Ångström.*

6.2.2 Validating the sigma factor–anti-sigma factor hypothesis

It took some time to finally identify CalpT as an anti-sigma factor regulating CalpS. *In silico* analysis of the operon using the WebFlags server as well as predicting the CalpS-T complex with AlphaFold and searching for structural relatives with FoldSeek and HHpred, provided valuable information to formulate a hypothesis of an anti-sigma–sigma factor pair and to design experiments to prove this hypothesis. When investigating the predicted interaction of the two proteins with mutants, we observed that CalpS R80E was unable to form the complex (see **Figure 5-19**). However, some of the interface mutants showed no effect on the complex formation. This observation may be explained by a closer look onto the predicted structure. While arginine 80 is in the center of the interface and coordinates E15 and D73 of CalpT, D88 and A91 are located at the edge of the interface, where mutations could be tolerated to a certain extent, since an orientation of the amino acid side chains away from the interface would not interfere with the predicted complex (**Figure 6-8**).

As is typical for the inhibition of sigma factors by their anti-sigma factor, CalpT binds to the σ2 and σ4 domains (Paget, 2015). This complex formation prevents CalpS from binding to RNA polymerase and furthermore, the DNA binding interface is occupied by CalpT.

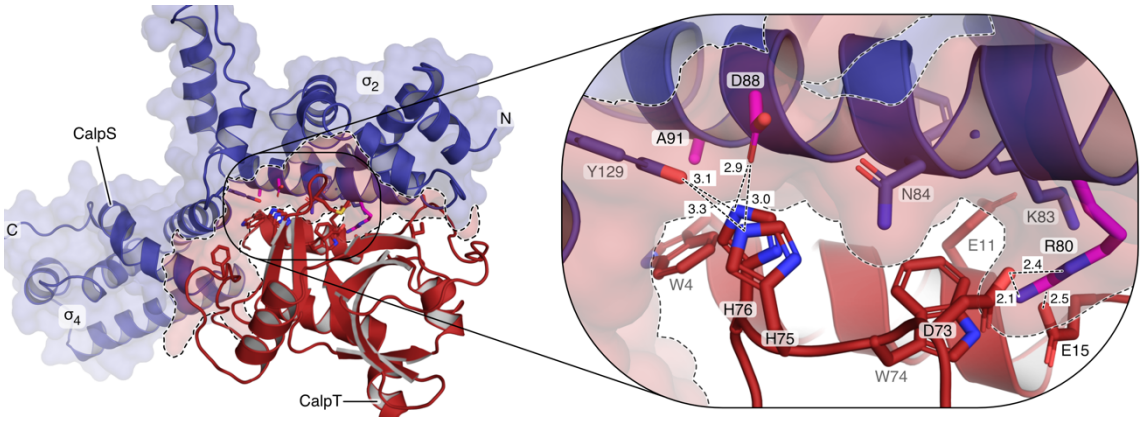


Figure 6-8: The CalpS-T interaction interface as predicted by AlphaFold2.
The predicted heterodimeric complex of CalpS (blue) and CalpT (red) shows an interface that covers both the σ2 and σ4 domains of CalpS. While the CalpS R80E mutant lost the ability to form a stable complex, mutation of residues D88 or A91, which are located at the edge of the interface, did not affect the interaction, as judged by gel filtration experiments. Residues that may be important for the interface are labeled and shown as sticks, possible electrostatic interactions are indicated by dashed lines and the lengths are given in Ångström. The part of the CalpS surface that is within 4 Å of CalpT is colored red and framed by a dashed line.

6.2.3 A working hypothesis of the CalpS-T-L mediated antiviral signaling cascade

After validation of the computational hypotheses, and recapitulation of all experimental results, we aimed to propose a working mechanism of the CalpS-CalpT-CalpL mediated signaling cascade that utilizes second messenger signaling and controlled proteolysis to combat phage attack. Signaling cascades are frequently initiated by proteases and often play important roles in cellular stress response. For example, bacterial toxin–antitoxin systems are activated by proteolytic degradation

(Schuster & Bertram, 2013), and Lon proteases have even been reported to participate in nutritionally controlled DNA replication (Leslie et al., 2015). In eukaryotes, proteases also control cellular pathways, such as the Gasdermin D mediated pore formation, which is initiated by proteolytic cleavage of the two-domain protein by inflammatory caspases (Kayagaki et al., 2015; J. Shi et al., 2015).

The presented antiviral signaling cascade starts with the recognition of viral transcription products by the CRISPR type III complex, which leads to the activation of the Cas10 subunit to synthesize cyclic tetraadenylate (cA_4) from ATP (Kazlauskiene et al., 2016; Niewoehner et al., 2017) (**Figure 6-9**.1/2). The second messenger molecule binds to the SAVED domain of CalpL, which is part of the preformed tripartite CalpS-T-L complex (**Figure 6-9**.3). The binding of cA_4 changes the electrostatics of the SAVED domain, allowing head-to-tail assembly of SAVED domains. This cOA-induced oligomerization leads to proteolytic cleavage of CalpT. The strictly cOA-dependent oligomerization was demonstrated by SAXS and DLS experiments (see **Figure 5-9**) and emerged as a common activation mechanism in SAVED domain containing effector proteins in CBASS defense systems (Hogrel et al., 2022; Lowey et al., 2020), and - more recently - also in CRISPR systems (Smalakyte et al., 2024; Steens et al., 2024).

Furthermore, we now know that the oligomerization also activates the nucleolytic cleavage of cA_4 by CalpL (Binder*, Schneberger* et al., 2024). Such active degradation of the second messenger has been observed for CARF domain containing enzymes and is an essential step to ensure complete escape from the 'infected' state (Athukoralage et al., 2018; Brown et al., 2020; Garcia-Doval et al., 2020). While $CalpT_{10}$ remains bound to the N-terminal CalpL domain, the $CalpT_{23}$-CalpS complex is released (**Figure 6-9**.4). Further proteolysis of the anti-σ factor, as reported for other members of the ECF σ factor family (Hughes & Mathee, 1998; Paget, 2015), releases the σ factor, CalpS, which can adapt a conformation to bind RNA polymerase (**Figure 6-9**.5). Subsequently, the RNA

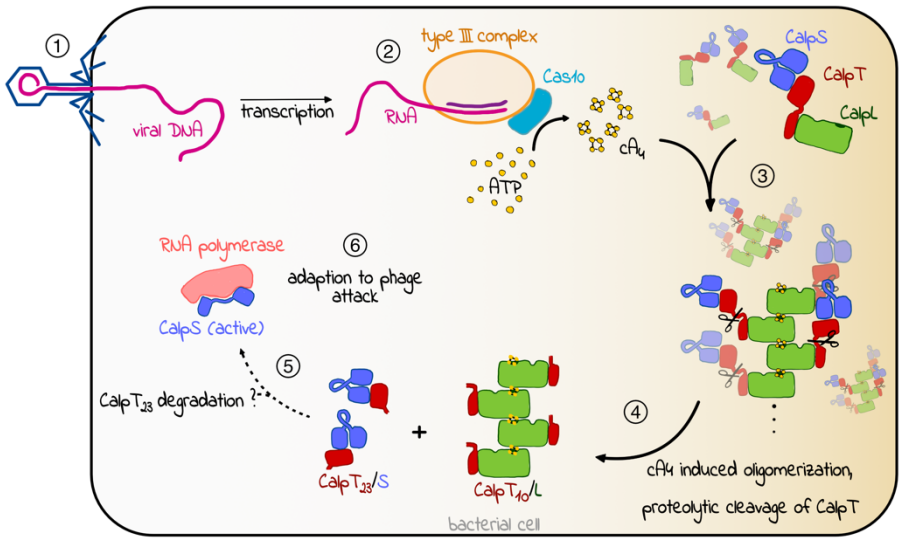


Figure 6-9: Working hypothesis for the CalpS-T-L mediated antiviral signaling cascade.
*Phage attack and injection of viral DNA (**1**). Upon transcription of viral DNA to RNA, the latter is detected by the type III CRISPR complex. Binding of viral RNA activates the complex, and the Cas10 subunit of the ribonucleoprotein (RNP) converts ATP to cA_4 (**2**). The second messenger binds to preformed CalpL–CalpT–CalpS complexes and triggers oligomerization (**3**). This, in turn, leads to proteolytic cleavage of CalpT, and release of the $CalpT_{23}$–CalpS fragment (**4**). $CalpT_{23}$ is probably degraded by proteases, which enables CalpS to adapt its active conformation and to bind to the RNAP (**5**). The σ factor, CalpS, directs the RNAP to a specific promoter at the genomic DNA to initiate transcription at that postion. This process enables the bacterium to counteract the phage infection (**6**).*

polymerase holoenzyme complex (including CalpS) is able to initiate the expression of specific, most likely survival-promoting, genes (**Figure 6-9**.6) (Bergkessel, 2021; Murakami et al., 2002). The entire CRISPR type III signaling cascade thus combines detection of viral infection with transcriptional regulation by second messenger signaling and proteolytic activity. Other CRISPR-based immune responses also affect the transcription machinery in order to adapt to phage attack, but the effector proteins involved are not related to the Calp proteins (Hu et al., 2022; Strecker et al., 2022). Interestingly, Strecker et al. found that the σ factor, CASP-σ, has a high affinity a sequence similar to the promoter region of the *cas1* and *cas2* genes and may boost spacer acquisition. A similar scenario is unlikely for Calp systems, since many related operons lack these genes (see Figure 4-9, **Figure 6-1**). Future experiments will address the identification of the DNA target sequence of CalpS to learn more about the detailed mode of action of how transcriptional changes can help to survive a phage attack.

Interestingly, the Cas10 proteins found in the Calp operons do not contain an HD nuclease domain, which normally degrades viral transcripts concomitantly with the cOA synthesis of the Palm domain (Jia & Patel, 2021; Kazlauskiene et al., 2017), suggesting that the cOA-induced signaling cascade and the resulting transcriptional changes are sufficient for antiviral immunity. It remains puzzling why second messenger signaling has evolved for this system, since the type III operon of *Sulfurihydrogenibium* sp. shows no evidence for any other CARF- or SAVED domain containing effector protein. Thus, cOA production activates only one specific enzyme and the potential of simultaneous activation of multiple effectors remains unused. More questions are left unanswered, ranging from the mechanistic details of how proteolytic cleavage is activated by cA_4 binding, to the proteolysis of $CalpT_{23}$ to release CalpS, up to how the escape from the 'infected' state is achieved in detail.

However, the results presented here uncovered a completely new CRISPR-based defense strategy that shows an evolutionary link between bacterial CBASS systems and CRISPR systems, but also highlights how complex bacterial immune mechanisms can become, reminiscent of complex eukaryotic mechanisms.

6.3 Outlook

Since the cA_4-triggered oligomerization of CalpL is a mechanistic key element for the protease activity as well as for the nuclease activity, it is of great interest to study the details of this homomeric assembly. Unfortunately, the oligomers turned out to be unstable under the conditions we tried, even using non-hydrolysable F-cA_4. Stabilization of the filaments could be achieved by different approaches. Lysine-specific covalent cross-linking using BS3 cross-linking reagents would be one strategy that has been used in our group to stabilize $NLRP3^{PYD}$ oligomers (Hochheiser et al., 2022). In order to observe stabilized CalpL filaments, the protein must be treated with either cA_4 or F-cA_4, followed immediately by the addition of BS3 cross-linking reagent. The length of the oligomers could be adjusted by varying the concentration of CalpL or cross-linker. It should also be noted, that cA_4 is degraded and therefore the incubation time between cA_4 treatment and BS3 addition will affect the assembly. However, if F-cA4, or nuclease deficient CalpL mutants are used, this dependence should not be present, while the binding affinities are most likely not affected (Binder*, Schneberger* et al., 2024). Another strategy would be to artificially introduce cysteine residues at positions that allow the formation of intermolecular disulfides when the oligomerization is induced (**Figure 6-10**). Such an approach was already used to trap specific conformational states of TRAP transporters (Peter et al., 2024). The size of oligomers can be adjusted by varying the same parameters as described for the cross-linker approach. The selection of suitable positions for cysteines is not trivial, as even tiny changes in the geometry could lead to an inactive oligomer. Oligomer structures of CalpL homologs (Smalakyte et al., 2024) and AlphaFold 3 predictions could be used to design different cysteine mutants. Cryogenic electron microscopy (cryo-EM) could then be used to obtain structures of stabilized filaments. Interestingly, both the recently published filament structure as well as the structural model predicted by AlphaFold 3 suggests a non-staggered assembly (**Figure 6-10 b**), which is in contrast to the staggered oligomer (**Figure 6-10 a**) we modelled based on SAXS experiments. In order to differentiate between the two possibilities, FRET experiments could be performed. When attaching fluorescence labels to the N-terminal domain, the distance between the labels differs significantly for the different orientations. Thus, a staggered oligomer would lead to a much larger distance in between the labels (and hence, a reduced energy transfer) than its non-staggered equivalent. Of course, two different labels which are suited for FRET must be used. Two individual labeling reactions with either the one or the other label must be done and a mix of the two samples can then be used for the FRET experiments.

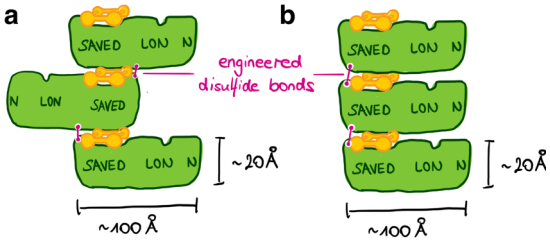


Figure 6-10: Sketch of CalpL filaments.
*CalpL (green) oligomers could be stabilized by engineered disulfide bonds (magenta). Our SAXS data suggested a staggered oligomer (**a**), but recent literature rather suggests a non-staggered cA_4-induced stacking (**b**).*

Fluorescent-based assays could also provide information on the dynamics of the oligomerization and furthermore, also on the dynamics of the proteolytic cleavage. Total internal reflection fluorescence (TIRF) microscopy can be used to track the movement of fluorescent-labeled proteins. When biotinylated CalpL is immobilized on a streptavidin-coated surface, fluorescent-labelled CalpT binds to the protease and creates a static florescence signal (**Figure 6-11 a**). This experiment would show whether the

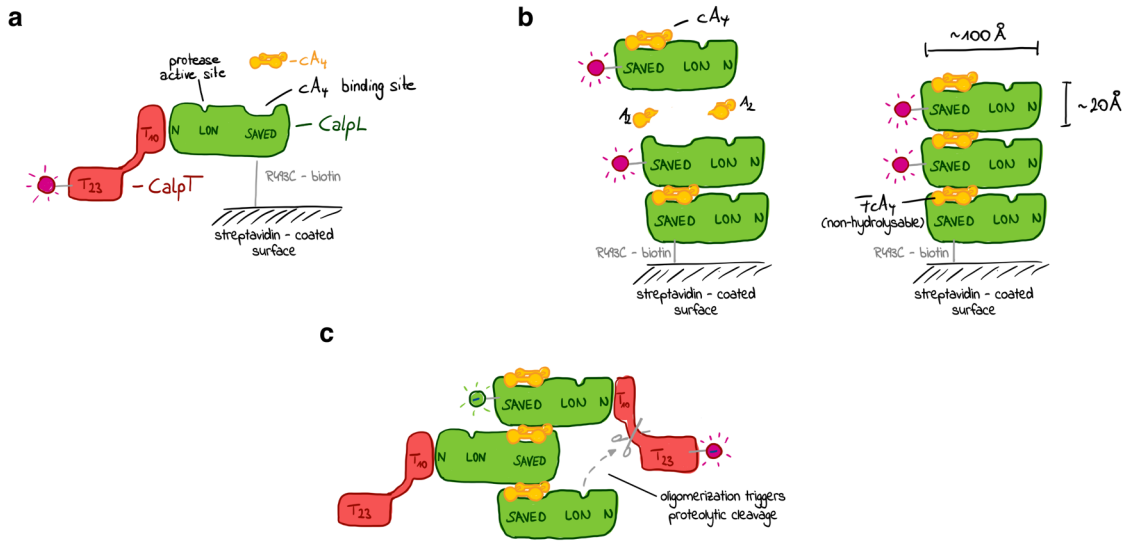


Figure 6-11: Illustrations of planned TIRF experiments to study CalpL oligomerization.
A biotinylated CalpL variant is immobilized on a streptavidin-coated surface. Using fluorescent-labelled CalpL (green) or CalpT (red), the dynamics of the interactions and proteolytic cleavage, depending on presence or absence of cA_4 (yellow), can be investigated.

immobilization strategy is successful. Thereafter, fluorescent-labelled CalpL can be used instead of CalpT and different interaction times of the labelled protein with the immobilized CalpL would be expected, depending on the presence or absence of cA_4 and F-cA_4, respectively (**Figure 6-11 b**). When CalpL and CalpT are both labelled with different dyes, the proteolytic cleavage of CalpT would lead to dissociation of one fluorescence signal (magenta), while the other signal (green) stays bound in the CalpL filament (**Figure 6-11 c**). Note that non-hydrolysable F-cA_4 should be used to prevent dissociation of the CalpL stacks. If it turns out that fluorescent labeling of CalpL or CalpT alters proteolytic activity, one could use the CalpL- or CalpT-specific nanobodies and label them.

Since we assume that in cells CalpL is not present as a monomer but in a heterotrimeric 1:1:1 complex with the CalpT and CalpS proteins, oligomers of this complex would provide further interesting information. In particular, the question of how the target peptide sequence of CalpT can reach the protease active site of CalpL could be investigated. By looking at structural models of a CalpT-L oligomer, it is difficult to imagine how the target sequence can reach the active site without conformational changes of CalpL. One possible region that could act as a hinge for conformational rearrangements is the transition between the N-terminal- and Lon protease domains (**Figure 6-12**). To see if conformational flexibility in between these two domains is important for the function, one could fix the relative conformation of the two domains by engineering interdomain disulfide bridges (**Figure 6-12**, magenta, i/ii). The ability of these mutants to proteolytically cleave CalpT can be investigated in protease cleavage assays.

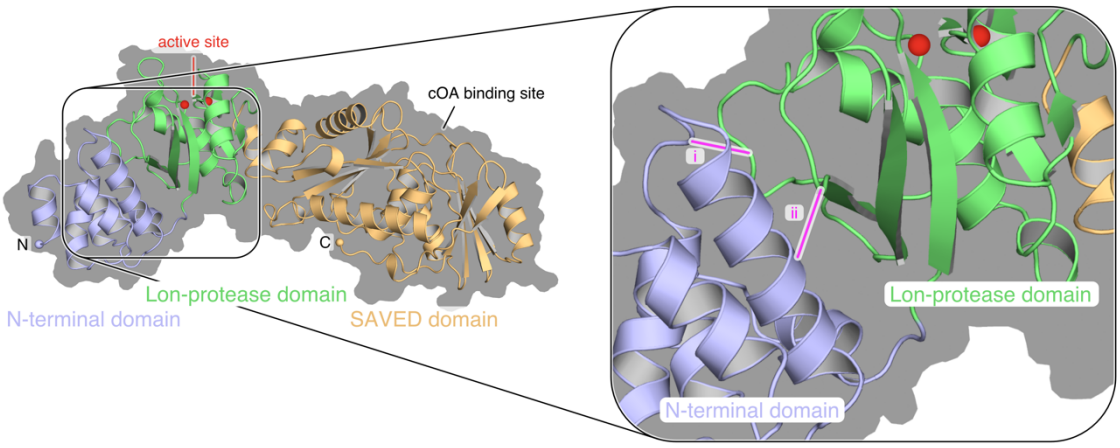


Figure 6-12: Preventing possible conformational changes by engineered disulfide bridges.
The domain architecture of CalpL reveals a possible hinge region in between the N-terminal domain (purple) and the Lon-protease domain (green). The orientation of these two domains relative to each other could be fixed by engineering disulfide bridges (i or ii, magenta).

7. Methods

The upcoming chapter contains a comprehensive list of the methods employed, accompanied by a detailed explanation of the underlying workflow. The initial section primarily focuses on preliminary tasks, including fundamental molecular genomics and cloning techniques used to prepare expression constructs, along with general procedures for protein biochemistry. Followed by a section that provides elaborate protocols for the recombinant expression and purification of all the proteins discussed, with some of these protocols being developed and fine-tuned during this work. The third part outlines the application of general analytical methods utilized to characterize the proteins and their complexes described herein.

7.1 Molecular genetics and cloning

7.1.1 Polymerase Chain Reaction (PCR)

Polymerase chain reaction (PCR) was used for amplification of DNA-fragments containing the protein-coding sequences or parts of it. Primer pairs were designed to anneal to the ends of the desired sequence and to contain overhang regions containing recognition sites for selected restriction enzymes. This allowed targeted ligation of the amplified sequence into the multiple cloning site (MCS) of an expression vector. The reaction mix of a standard PCR and the standard protocol are listed in **Table 7-1** and **Table 7-2**, respectively.

Table 7-1: PCR reaction mix

Component	Stock concentration	Final concentration
Template DNA	variable	50-100 ng
Primer forward	100 µM	1 µM
Primer reverse	100 µM	1 µM
dNTPs	2.5 mM	200 µM
DNA polymerase buffer	5x or 10x	1x
DNA polymerase	5 units/µl	0.05 units/µl
ddH_2O		Fill ut to 50 µl

Table 7-2: PCR thermocycler program

Number of cycles	Step	Temperature [°C]	Time [s]
1	Initial denaturation	95	300
30	Denaturation	95	60
	Annealing	T_m -5	60
	Elongation	72	30/kb
1	Final elongation	72	900

7.1.2 Restriction enzyme digestion of DNA

Restriction enzymes are endonucleases which are part of the bacterial or archaeal immune system. In molecular biology they are frequently used to specifically digest double-stranded DNA sequences. The enzymatic cleavage produced 5’ or 3’ protruding ends (sticky ends) that were used for subsequent ligation (see Chapter 7.1.4). For

digestion, 45 µl of the DNA was supplemented with 5 µl of CutSmart® buffer and after addition of the restriction enzymes, the reaction mix was incubated at 37 °C for 1 h. Subsequently, the endonucleases were heat inactivated according to the instructions of the manufacturer (New England Biolabs). The results were analyzed by agarose gel electrophoresis.

7.1.3 Agarose gel electrophoreses

To separate DNA samples by their size, agarose gel electrophoresis was performed. Samples were filled into an agarose gel (1 % agarose in TAE buffer) supplemented with DNA intercalating dye (peqGREEN). The gel was run in an electrophoresis chamber filled with TAE buffer for 40 min at a constant voltage of 100 V. DNA bands were detected by UV excitation of the used DNA-binding dye and documented using a ChemiDoc™ XRS+ imaging system. If the DNA was needed for cloning purposes, the according bands were cut out with a scalpel and the contained DNA was extracted and purified using the ExtractMe DNA Clean-up & Gel-Out Kit according to the provided protocol. Afterwards, the concentration was determined by a NanoDrop 2000 spectrophotometer at λ=260 nm and the solution was stored at -20 °C.

7.1.4 Ligation of DNA

The digested DNA amplicon was ligated into the linearized destination vector using a DNA ligase from bacteriophage T4. To catalyze phosphodiester bond formation of the two compatible sticky ends, 50 ng of linearized vector was mixed with a 10-fold molar excess of insert. After addition of buffer, water and T4 DNA ligase, the reaction mix (20 µl in total) was incubated at 25 °C for 20 min and then at 16 °C overnight. After heat inactivation of the ligase at 65 °C, 10 µl of the solution was transformed into chemically competent *E. coli* cells.

7.1.5 Site-directed mutagenesis

To study the importance of specific amino acid residues, the triplet code of the protein encoding gene was modified by site-directed mutagenesis. For example, this was done to determine the exact cleavage site of CalpT, to observe the effects on complex formation between two proteins, or to introduce cysteines to site-specifically attach labels. For this purpose, a pair of primers was designed that contain non-overlapping and overlapping regions containing the triplet modification within the latter region. These regions were chosen in a way, that the melting temperature of non-overlapping sequences ($T_{m,no}$) is 5-10 °C higher than that of the overlapping sequences ($T_{m,pp}$). The design of primers as well as the thermocycler program were based on the protocol established by Liu and Naismith (H. Liu & Naismith, 2008). The PCR reaction mix was pipetted as already mentioned in **Section 7.1.1** and the thermocycler program is described in **Table 7-3**.

Table 7-3: PCR thermocycler program for site-directed mutagenesis

Number of cycles	Step	Temperature [°C]	Time [s]
1	Initial denaturation	95	300
3	Denaturation Annealing Elongation	95 $T_{m, pp}$ - 5 72	60 30 30/kb
25	Denaturation Annealing Elongation	95 $T_{m, no}$ -5 72	60 30 30/kb
2	Denaturation Annealing Elongation	95 42 72	120 60 30/kb
1	Final elongation	72	900

7.1.6 Transformation of plasmid DNA into bacteria

After ligation of a DNA fragment into a plasmid or subsequent to site-directed mutagenesis, the newly generated plasmid DNA was transformed into chemically competent *E. coli* cells. The cells used in this work and their application for either *in vivo* amplification and selection or protein expression are listed in **Table 8-6**. All competent cells were prepared in our laboratories according to a standard protocol using $CaCl_2$ treatment and stored at -80 °C as 50 µl aliquots.

The cells were thawed on ice and 50-100 ng of plasmid DNA was added, for transformation directly from a ligation reaction mixture, 10 µl of the reaction mixture were added without determining the exact concentration. After incubation on ice for 10 min, was followed by 45 sec incubation at 42 °C in a water bath. Subsequently, the cells were cooled on ice, 1 ml of LB media (without antibiotics) was added and the cell suspension was incubated at 37 C and 800 rpm shaking for 1 h. After this time, which is needed to allow the cells the expression of the antibiotic resistance genes encoded on the transformed plasmid, the cell suspension was centrifuged at 5,000 r.c.f. for 5 min. Excess media was discarded, the cell pellet was resuspended in the remaining (~100 µl) media and the suspension was streaked out on an LB agar plates supplemented with the appropriate antibiotics. The plates were incubated at 37 °C overnight. After successful transformation, a single colony was transferred to a liquid culture containing the same antibiotics. These liquid cultures were either used for subsequent plasmid extraction and sequencing, or served as preculture for protein expression.

To optimize the work-flow efficiency, glycerol stocks (1.5 ml containing 20 % glycerol, 80 % cell suspension) were prepared from precultures of expression strains so that new precultures could be set up directly from this stocks which were stored at -80 °C.

7.1.7 Plasmid DNA isolation and sequencing

To check plasmid DNA, 50-100 ng were transformed into *E. coli* DH5α, or *E. coli* β10 cells as described above. After successful transformation, a single colony was transferred from the agar plate to 4 ml of LB media supplemented with the appropriate antibiotics. The cell suspension was incubated at 37 °C for at least 10 h or overnight and subsequently, plasmid DNA was extracted and purified using the ExtractMe Plasmid mini kit (Blirt) or the Plasmid DNA purification kit (Machery-Nagel). DNA isolation and clean up was performed step-wise according to the protocol of the manufacturer. Finally, the

DNA concentration was determined by measuring the absorption at λ=260 nm with a NanoDrop 2000 spectrophotometer. A sample containing 12 µl plasmid DNA solution with a concentration in between 40-100 ng/µl was sent for external sanger sequencing to Microsynth SEQLAB or Eurofins Genomics. The sequencing results were analyzed using SnapGene.

7.2 General methods for protein biochemistry

7.2.1 SDS-PAGE

To separate proteins by their size and visualize their purity, sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) was used. This standard technique was used to check the protein purification procedure, to visualize proteolytic cleavage, to analyze complex formation and to prepare samples for peptide mass fingerprinting. By addition of SDS sample buffer and subsequent incubation of the samples at 92 °C for 5 min, the protein samples were denatured. Afterwards, the samples were loaded to self-prepared discontinuous polyacrylamide gels. The gels were run in a Mini-Protean Tetra Cell (Bio-Rad) electrophoresis system filled with SDS running buffer at a constant voltage of 250 V for 37 min. Thereafter, the gels were boiled for 1 min in Coomassie staining solution and incubated for 10 min at slow mixing, followed by boiling for 5 min in destaining solution and incubation for 1 h at room temperature and slow shaking. Coomassie staining was not done for subsequent Western Blot analysis. Coomassie stained gels were imaged with a ChemiDoc™ XRS+ (Bio-Rad) imaging system. Either 12 % or 15 % polyacrylamide gels were used, the ingredients for casting the gels are listed in **Table 7-4**.

Table 7-4: Composition of polyacrylamide gels.

Gel	Ingredient	Volume
5 % stacking gel	Acrylamide (30 %)	225 µl
	Stacking gel buffer	175 µl
	H2O	900 µl
	TEMED	1.3 µl
	APS (10 %)	13.25 µl
12 % separation gel	Acrylamide (30 %)	2.1 ml
	Separation gel buffer	1.46 ml
	H2O	1.69 ml
	TEMED	1.76 µl
	APS (10 %)	58.88 µl
15 % separation gel	Acrylamide (30 %)	2.63 ml
	Separation gel buffer	1.46 ml
	H2O	1.16 ml
	TEMED	1.76 µl
	APS (10 %)	58.88 µl

7.2.2 Western Blot

For the specific detection of proteins following SDS-PAGE, Western Blot was used. During this work, all analyses were done with a primary His_6 affinity tag antibody and a horseradish peroxidase (HRP) conjugated goat anti-mouse secondary antibody. Prior to the blotting, approximately 100 ng of protein were loaded to and run on a polyacrylamide

gel as described above. As already mentioned, the gel was not stained with Coomassie staining solution, but directly used for the blotting procedure. The blotting membrane (nitrocellulose membrane and the blotting paper were equilibrated in the transfer buffer. The components were assembled in a semi-dry blotting chamber so that the gel was on top of the membrane and the two were sandwiched by 2 blotting papers on each side. Potential air bubbles between the different layers were removed. The blotting chamber was assembled in a way that the membrane was facing the cathode and the polyacrylamide gel was on the side of the anode. Blotting was carried out at a constant current of 160 mA for 45 min. Subsequently, the blotting membrane was incubated in PBS-T with 5 % (w/V) milk powder for 1 h at room temperature and gentle agitation. Then, the membrane was washed three times with 10 ml of PBS-T for 10 minutes before it was incubated at 4 °C and gentle agitation overnight in 5 ml of a PBS-T solution containing a 1:10,000 dilution of the primary antibody. After another three washing steps wit 10 ml PBS-T each, the membrane was incubated for 1 h at room temperature and gentle agitation in 5 ml of a PBS-T solution containing a 1:5,000 dilution of the secondary antibody. Subsequently, the membrane was washed two times with 10 ml of PBS-T and then with 10 ml of PBS before it was developed with Western Blotting detection reagent (Invitrogen) and documented with a ChemiDoc^{TM} XRS+ (Bio-Rad) imaging system.

7.2.3 Labeling of proteins

7.2.3.1 Spin labeling

For the analysis of a potential dimerization of CalpT, a single cysteine mutant (CalpT E119C) was designed to attach an MTSSL spin label via a maleimide reaction at the cysteine position. First, the mutant was expressed and purified as described for CalpT but without removal of the His_{10} affinity tag. A volume of 250 µl of a 315 µM protein solution was bound to Ni^{2+}-NTA beads and the mixture was loaded to a gravity column. The flow thorough was discarded, the protein loaded beads were washed with 10 ml of reduction buffer (25 mM Tris, 250 mM NaCl, 1 mM TCEP, 10 % glycerol, pH 8.0) and subsequently with 20 ml of wash buffer (25 mM Tris, 250 mM NaCl, 10 % glycerol, pH 8.0). The protein was eluted from the beads in spin label-containing elution buffer (25 mM Tris, 250 mM NaCl, 1 M imidazole, 0.6 mM MTSSL, 10 % glycerol, pH 8.0). Free spin label and imidazole was separated using a PD10 desalting column. To reduce background signal in the PELDOR measurements, the sample buffer was exchanged to contain D_2O instead of H_2O as solvent. The buffer exchange was done using a PD10 desalting column equilibrated with wash buffer (see above) prepared with D_2O. In addition, the buffer exchange procedure was done for wild-type CalpL and the protease deficient CalpL S152A mutant (Buffer: 20 mM Tris, 50 mM NaCl, pH 8.0). In a final step, the spin labelled E119C mutant of CalpT was proteolytically cleaved by CalpL to yield spin labelled $CalpT_{23}$. Samples of full length CalpT, full length CalpT in complex with CalpL S152A and CalpT23 were sent to collaboration partners at the University of St. Andrews to be analyzed by pulsed EPR spectroscopy.

7.2.4 Mass spectrometry

For peptide mass fingerprinting, protein bands from SDS-PAGE analysis were excised and sent to BSRC Mass Spectrometry & Proteomics Facility at the University of St. Andrews, Scotland. There, the gel bands were cut into 1 mm^3 cubes and were destained with ethanol, acetonitrile and 25 mM ammonium bicarbonate. Reduction of the samples with 10 mM dithioerythritol, was followed by alkylation using iodoacetamide (20 mM). Shrinking of the gel pieces (acetonitrile) was followed by soaking in 25 mM ammonium bicarbonate supplemented with 2 ng/µl trypsin and peptide digestion was carried out at 37 °C over night. Then, peptide extraction was done with 1% formic acid before the sample was loaded onto a Eksigent 2D ultra nano HPLC with Sceix 5600+ mass spectrometer. The Acclaim Pepmap 100 trap (Thermo Scientific, 20 mm × 75 µm) and column (150 mm × 75 µm) were in trap elute configuration with flow rates of 5 µl/min and 0.3 µl/min, respectively. The peptides were loaded onto the trap and washed for 5 min with loading buffer (0.05% trifluoroacetic acid) before the trap was switched in line with the column and the peptides eluted with a linear gradient over 20 min of 98 % A to 98 % B, in which A is water with 0.1 % formic acid and B is 80 % acetonitrile, 20 % water and 0.1 % formic acid. The eluent was sprayed directly into the nanosource of the mass spectrometer. Mass spectrometry (MS) data were collected from 400 to 1,250 m/z in positive ionization for 150 ms. Data-dependant acquisition mode was utilized to collect MS/MS data from 100 to 2,000 m/z on the 20 strongest peptides with 2–5+ charge states. The peak list was extracted from the .wiff file using MSconvert, and the .mgf file was searched against an in-house database of 7,000 protein sequences to which the sequences of the proteins of interest were added. The following settings were used in the Mascot search: trypsin and semi-trypsin as digest enzymes; fixed modification of carbamidomethyl (c); and variable modification of oxidation (M). The MS tolerance was set at 20 ppm and MS/MS at 0.1 Da.

7.3 Recombinant protein expression and purification

7.3.1 Proteins for TRAP project

Table 7-5: UniProt IDs TRAP project

Protein	**Organism (origin)**	**#AA / MW (kDa)**	**UniProt ID**
VcSiaP	*Vibrio cholerae*	321[1] / 35.982[1] 299[2] / 33.755[2] 303[3] / 34.072[3]	Q9KR64
HiSiaP	*Haemophilus influenzae*	329[1] / 36.513[1] 306[2] / 34.165[2] 310[3] / 34.495[3]	P44542

[1] UniProt sequence (includes signal sequence);

[2] UniProt sequence without signal sequence;

[3] recombinant expression construct after TEV cleavage.

7.3.1.1 TRAP transporter P-domains

The detailed procedure for protein expression and purification of VcSiaP and HiSiaP was described before (Glaenzer et al., 2017). The genes for both proteins were cloned into a pBADHisTEV vector containing a TEV cleavable N-terminal His_6 tag

(Huanting Liu, University of St Andrews). All proteins were expressed in M9-minimal media to prevent sialic acid binding during the expression. For protein purification, Ni^{2+}-affinity chromatography was followed by size-exclusion chromatography. The His tag was removed by incubation of a 1:50 mass ratio of TEV-protease to protein at 4 °C overnight. The purified protein was collected in the flow through of another affinity chromatography before it was concentrated, flash frozen in $N_{2(l)}$ and stored at -80 °C.

7.3.2 Proteins for CRISPR project

Table 7-6: UniProt IDs CRISPR project

Protein	**Organism (origin)**	**#AA / MW (kDa)**	**UniProt ID**
CalpL	*Sulfurihydrogenibium* sp. (strain YO3AOP1)	496 / 57.811	B2V8L9
CalpT	*Sulfurihydrogenibium* sp. (strain YO3AOP1)	271 / 31.844	B2V8L8
CalpS	*Sulfurihydrogenibium* sp. (strain YO3AOP1)	224 / 26.561	B2V8L7

7.3.2.1 CalpL

The codon-optimized gene for CalpL (UniProt ID: B2V8L9) was cloned into pET-11a vector containing an N-terminal His_{10}-TEV tag. All CalpL constructs were expressed in lysogeny broth (LB) medium. Therefore, *E. coli* BL21(DE3) cells were grown at 37 °C until an optical density at λ=600 nm (OD_{600}) of 0.6–0.8 was reached. At that cell density, expression was started by induction with 0.4 mM Isopropyl β- d-1-thiogalactopyranoside (IPTG), and the cell suspension was incubated at 30 °C for 4.5 h with shaking. Cells were collected by centrifugation at 4,000 r.c.f. for 25 min at 20 °C and resuspended in lysis buffer (20 mM Tris, 50 mM NaCl, pH 8.0). The cells were lysed by sonification (Sonopuls HD3100, Bandelin) and cell debris was removed by centrifugation at 75,000 r.c.f. for 45 min at 10 °C. For protein purification, Ni^{2+}-affinity chromatography (20 mM Tris, 50 mM NaCl, pH 8.0; 500 mM imidazole was included for elution) using Ni^{2+}-NTA resin beads and a gravity flow column was followed by size-exclusion chromatography (SEC) (20 mM Tris, 50 mM NaCl, pH 8.0) using a Superdex 200 16/600 column on an ÄKTA chromatography system (both: GE Healthcare). In case the purity was not sufficient, anion exchange chromatography (AIEX) was run on a HiPrep Q XL column (GE Healthcare) using gradient elution (20 mM Tris, 0-500 mM NaCl, pH 8.0). For Tag cleavage, overnight incubation at 4 °C with a 50:1 ratio (m/m) of protein to TEV protease (20 mM Tris, 50 mM NaCl, pH 8.0) was performed. A second Ni^{2+}-affinity chromatography was done to remove the TEV protease and noncleaved CalpL. The protein purity was checked by SDS–PAGE analysis after each purification step. After successful purification, the protein was concentrated, flash-frozen in liquid nitrogen and stored at −80 °C in 20 mM Tris, 50 mM NaCl, pH 8.0.

For structure determination, a selenomethionine derivative of CalpL was prepared using *E. coli* B834 cells and the SelenoMethionine Medium Complete kit (Molecular Dimensions) according to the instructions. Protein expression and purification were done in the same way as for the native protein.

7.3.2.2 CalpT

Prior to protein expression, the codon-optimized synthetic gene (BioCat) for CalpT (UniProt ID: B2V8L8) including an N-terminal His_{10}-TEV tag was cloned into pET-11a vector and transformed into chemically competent *E. coli* BL21(DE3).

For the expression culture, 1 l of LB media was supplemented with 1 ml of Ampicillin (100 mg/ml) and 5 ml of overnight preculture. The cells were grown at 37 °C and continuous shaking at 130 rpm until an OD_{600} of 0.6-0.8 was reached. Subsequently, the protein expression was started upon induction with 0.4 mM IPTG and incubation at 30 °C with shaking at 130 rpm. After 4.5 h, the cells were harvested by centrifugation at 4,000 r.c.f. for 25 min at 20 °C. To extract the protein, the cells were resuspended in lysis buffer (25 mM Tris, 500 mM NaCl, 10% glycerol and 1 mM DTT, pH 8.0) and sonicated (Sonopuls HD3100, Bandelin). By centrifugation at 75,000 r.c.f. for 45 min at 20 °C, cell debris was removed and the soluble fraction was filtered through a 0.8 µm syringe filter.

To purify CalpT, Ni^{2+}-affinity chromatography (25 mM Tris, 500 mM NaCl, 1 mM DTT, 10% glycerol, pH 8.0; 40 mM and 1 M imidazole was included for wash and elution, respectively) using Ni^{2+}-NTA beads and a gravity flow column was followed by heparin chromatography (25 mM Tris, 0-1 M NaCl (gradient elution, 0 %-40 % over 3.6 column volumes), 1 mM DTT and 10% glycerol, pH 8.0) using a HiPrep Heparin FF 16/10 column on an ÄKTA chromatography system (both: GE Healthcare). Before loading the sample to the column, the protein solution was diluted 50-fold with no salt buffer (25 mM Tris, 1 mM DTT and 10% glycerol, pH 8.0). As a final purification step, size-exclusion chromatography was performed on a Superdex 200 16/600 column on an ÄKTA chromatography system (both: GE Healthcare) (25 mM Tris, 500 mM NaCl, 1 mM DTT and 10% glycerol, pH 8.0). Removal of the N-terminal His tag was done by overnight incubation at 20 °C with a 20:1 ratio (m/m) of protein to TEV protease (25 mM Tris, 500 mM NaCl, 1 mM DTT and 10% glycerol, pH 8.0). Tag cleavage was only done for crystallization purposes, but not for any functional assays. A second Ni^{2+}-affinity chromatography step was used to separate the TEV protease and noncleaved protein. The purity of the protein was checked by SDS–PAGE after each purification step. After successful purification, the protein was concentrated, flash-frozen in liquid nitrogen and stored at −80 °C in 25 mM Tris, 500 mM NaCl, 1 mM DTT and 10% glycerol, pH 8.0. All purification steps for CalpT were done at room temperature as the protein tended to aggregate at 4 °C.

7.3.2.3 CalpS

One codon optimized gene for CalpS was designed in a pBADHisTEV vector including an N-terminal His_6 tag and TEV cleavage site. The Synthesis was done by BioCat. This gene was subsequently also cloned into a pET-28a vector to yield a MPB-fusion protein with a TEV protease cleavage site in between MPB and CalpS. A second construct was cloned by Haotian Chi (University of St. Andrews) into a pEHisTev vector with a cleavable octa-histidine tag at the N-terminus of the protein. Furthermore, the same construct was cloned into a pCDFDuet vector yielding a tag-less protein. The different expression vectors were tested one after another. The so far most promising expression and purification procedure will be described in the following. Protein expression was done using *E. coli C43*(DE3) cells as expression strain. The main culture was inoculated with overnight preculture (10 ml/l) of high cell density. Cells were grown in lysogeny broth (LB) media at 37 °C at 130 rpm shaking. At an optical density of $OD_{600}\approx0.7$, the

protein expression was induced by addition of IPTG (c_{end}=0.4 mM). Expression was carried out at 16 °C over night. Thereafter, the cells were harvested by centrifugation at 10 °C, 4,000 r.c.f. for 25 min, resuspended in 5 ml lysis buffer per gram of cell pellet (50 mM Tris, 500 mM NaCl, 20 mM imidazole, 10 % glycerol, pH 7.5, supplemented with EDTA-free protease inhibitor (Roche)) and lysozyme (1 mg/ml). The cells were lysed by sonification and the lysate was centrifuged at 10 °C, 75,000 r.c.f. for 45 min. The lysate was filtered using a 0.45 µm syringe filter and loaded to Ni^{2+}-NTA beads (or Amylose beads for MBP affinity chromatography). As for CalpL and CalpT, affinity chromatography (50 mM Tris, 500 mM NaCl, 10 % glycerol, pH 7.5, 500 mM imidazole for elution) was followed by size-exclusion chromatography (50 mM Tris, 250 mM NaCl, 10 % glycerol, pH 7.5). If desired, the affinity tag was cleaved off using TEV protease and another affinity chromatography was performed to yield untagged CalpS in the flow through. All purification steps were checked by SDS-PAGE analysis. Purified protein was concentrated, aliquoted at 50-100 µl volume and flash frozen in liquid Nitrogen, before the aliquots were stored at -80 °C.

7.3.3 VHH antibodies

7.3.3.1 Generation of VHH antibodies

The 'Landesuntersuchungsamt Rheinland-Pfalz (23 177- 07/A 17-20-005 HP) authorized all immunizations described in this work. Within 12 weeks, one alpaca (Vicugna pacos) was immunized by six subcutaneous injections á 200 µg antigen (1:1 (v/v) mixture of protein solution and GERBU-FAMA adjuvant). Subsequently, peripheral blood mononuclear cells (PBMCs) were isolated from 100 ml of blood, their mRNA was extracted and reverse transcribed to cDNA. To generate a library for phage display, VHH sequences were amplified by PCR and cloned into a phagemid vectors. The phage display was done using *E. coli TG1* cells in combination with VCSM13 helper phages to enrich specific VHHs. The bait protein was therefore immobilized to magnetic streptavidin beads. After two rounds of panning, *E. coli ER2738* cells were infected with the cleared and enriched phages and individual clones were selected to be grown in a 96-well plate. The supernatants from these small-scale expressions were tested for specific binding VHHs by ELISA. Hits were identified, sequenced and grouped according to their sequence similarity.

7.3.3.2 Expression and purification of VHH antibodies

All VHHs were generated by the Core Facility Nanobodies of the University of Bonn. The nanobody encoding genes were provided in a pHEN6 vector for TRAP transporter SBP specific VHHs and in a pSBinit2 for CalpT–L specific VHHs, respectively. All constructs contained an N-terminal pelB signaling sequence to achieve periplasmic expression at non-reducing conditions and a C-terminal HA-His_6 tag for intracellular recognition and purification purposes. The plasmids were transformed into chemically competent *E. coli* WK6 cells and *E. coli* MC1061 cells, respectively. Cells were grown in 2 l Terrific Broth (TB) media (including 100 µg/ml Ampicillin and 25 µg/ml Chloramphenicol, respectively) inoculated with 25 ml of an overnight preculture and incubated at 37 °C and shaking until reaching an OD_{600} of 0.7. Protein expression was induced by adding 0.4 mM IPTG (25 ml of a 1 % L(+)-Arabinose stock solution for CalpT–L specific VHHs) followed by overnight incubation at 30 °C (22 °C for CalpT–L

specific VHHs) and 130 rpm shaking. The cells were harvested by centrifugation at 5,000 r.c.f. and 10 °C for 25 min. The cell pellet was resuspended in 25 ml TES buffer (200 mM Tris, 0.65 mM EDTA, 500 mM sucrose, pH 8.0) and incubated for at least 1.5 h and slow mixing at 4 °C. For osmotic lysis, the cell suspension was diluted with 75 ml of 0.25-fold concentrated TES buffer and incubated at 4 °C overnight with slow mixing. Subsequently, the suspension was centrifuged at 10,000 r.c.f. and 10 °C for 1 h, the supernatant was decanted to a new centrifuge beaker and once again centrifuged at the same conditions. The periplasmic extract was mixed with equilibrated Ni^{2+}-NTA beads and incubated for 2 h at 4 °C and slow mixing. Afterwards, Ni^{2+}-affinity chromatography was carried out on a gravity flow column. The flow through was discarded and the beads were washed with 100 ml wash buffer (50 mM Tris, 300 mM NaCl, 10 mM imidazole, pH 8.0). The protein was eluted in 15 ml elution buffer (50 mM Tris, 50 mM NaCl, 500 mM imidazole, pH 8.0) and concentrated to a final volume of 4 ml using an Amicon 3 kDa MWCO. The protein solution was loaded onto a HiLoad Superdex 75 16/600 gel filtration column for size exclusion chromatography (50 mM Tris, 50 mM NaCl, pH 8.0) on an ÄKTA chromatography system (both: GE Healthcare). Protein containing fractions were pooled, concentrated, flash frozen in $N_{2(l)}$ and stored at -80 °C. After each purification step, the purity was checked by SDS-PAGE analysis.

7.3.3.3 Optimization of VHH antibody constructs for crystallization

To improve the crystallization behavior, VcSiaP specific VHHs were cloned into a pET-28a vector containing a pelB signaling sequence and an N-terminal His_6-TEV tag. The VHH encoding sequences were amplified by PCR and assembled into pET-28a-pelB-His_6-TEV by traditional cloning using the restriction enzymes *BamH*I and *Eco*RI or *BamH*I and *Xho*I for NbS001 and NbS002, respectively. The PCR products and the target vector were digested with the respective restriction enzymes and ligated at a molar ratio of 3:1. Positive clones were identified by double enzyme digestion and the correct sequences confirmed by sequencing at Microsynth AG (CH). The expression constructs were transformed into chemically competent *E. coli BL21*(DE3) cells. Expression and purification were performed in the same way as described above (Chapter 7.3.3.2).

7.4 Analytical methods

7.4.1 Analytical size-exclusion chromatography and SEC-MALS measurements

7.4.1.1 Analytical size-exclusion chromatography

Size-exclusion chromatography cannot only be used for protein purification, but also to characterize complex formation with affinities stronger than about 1 mM (Stevens & Schiffer, 1981). Analytical SEC was either carried out using a Superdex 200 (or Superose6) increase 10/300 gel filtration column on an ÄKTA chromatography system, or on a Superdex 200 (or Superose6) increase 3.2/300 gel filtration column installed on a 1260 Infinity II HPLC system (Agilent). Column types, buffer compositions and concentrations that were used for the experiments are listed in **Table 7-7**.

Table 7-7: Specific information for analytical size exclusion chromatography.

Experiment name / *buffer composition*	**Column /** additional information	**Protein**	**Conc.**
VcSiaP-VHH complex (including VcSiaP S14C T192C) 50 mM Tris, 50 mM NaCl, pH 8.0	**SD200 increase 3.2/300** Flow rate: 0.07 ml/min Injection volume: 20 µl	VcSiaP NbS001 NbS002	29.4 µM 32.3 µM 32.3 µM
CalpS-T-L complex 20 mM Tris, 250 mM NaCl, 1 mM DTT, 10% glycerol, pH 8.0	**Superose6 increase 10/300** *all samples were incubated at 50 °C for 60 min, SEC was done at RT Flow rate: 0.4 ml/min Injection volume: approx. 300 µl	CalpL CalpT CalpS cA_4	182 µM 182 µM 182 µM 200 µM
CalpS-T complex 20 mM Tris, 250 mM NaCl, 1 mM DTT, 10% glycerol, pH 8.0	**SD200 increase 3.2/300** Flow rate: 0.07 ml/min Injection volume: 20 µl	CalpT CalpS	20.0 µM 20.0 µM

7.4.1.2 Multi-angle light scattering (MALS)

Multi-angle light scattering (MALS) is a powerful biophysical technique for determining the size and molecular weight of macromolecules in solution. The scattered light from a particle is measured at multiple angles, providing valuable information on the mentioned properties. To characterize the heteromultimeric protein complexes, SEC-MALS runs were performed at room temperature on an Agilent 1260 Infinity II Prime Bio LC system coupled to a Wyatt miniDAWN MALS detector, an Optilab rEX refractive index detector and a Superose6 increase 10/300 or Superdex 200 increase 10/300 chromatography column (both: GE Healthcare) equilibrated with the corresponding sample buffer. All data acquisition and evaluation were carried out using ASTRA 8 software (Wyatt Technologies). Detailed information on the sample concentration and the experimental setup are listed in **Table 7-8**. All concentrations were adjusted by dilution with the appropriate running buffer and all samples were centrifuged at 15,000 r.c.f. for 10 min before injection into the instrument.

Table 7-8: Specific information for multi-angle light scattering experiments.

Experiment name / buffer composition	**Column /** additional information	**Protein**	**Conc.**
VcSiaP (HiSiaP) & NbS001-003 50 mM Tris, 50 mM NaCl, pH 8.0	**Superose6 increase 10/300** Flow rate: 0.5 ml/min Injection volume: 50 µl *HiSiaP was only used in combination with NbS003	VcSiaP HiSiaP NbS001 NbS002 NbS003	120 µM 120 µM 120 µM 120 µM 120 µM
HiSiaP & NbS004-011 50 mM Tris, 50 mM NaCl, pH 8.0	**Superdex200 increase 10/300** Flow rate: 0.5 ml/min Injection volume: 50 µl	HiSiaP NbS004-NbS011	30 µM 30 µM
CalpT-L complex analysis 25 mM Tris, 500 mM NaCl, 1 mM DTT, 10% glycerol, pH 8.0	**Superose6 increase 10/300** *all samples were incubated at 50 °C for 60 min, SEC was done at RT Flow rate: 0.4 ml/min Injection volume: 50 µl	CalpL CalpT cA_4	51 µM 59 µM 60 µM

7.4.2 Thermal nano-DSF

The stability of proteins can be monitored by their intrinsic fluorescence which changes upon unfolding of the tertiary structure caused by the application of thermal or chemical denaturant gradients. Multiple external factors, such as small molecules, macromolecular binding partners, or buffer compositions can affect the conformational stability. Nano differential scanning fluorometry (nano-DSF) can be used to analyze how changes in the buffer composition, binding of small molecules, or the binding of macromolecular molecules stabilize or destabilize a protein.

Thermal denaturation curves were determined using a Prometheus NT.48 thermal nanoDSF device in combination with the PR.ThermControl software (both: NanoTemper Technologies). The sample concentration was set to 1 mg/ml of VcSiaP and potential binding partners were added at a 1.1x molar excess. All samples were centrifuged at 14,000 r.c.f. and 4 °C for 10 min and then loaded to nanoDSF grade standard glass capillaries (NanoTemper Technologies). In all experiments, samples were loaded as technical duplicates to exclude errors that can potentially occur from tiny air bubbles inside the capillary. As starting temperature T_{start}=20 °C and end temperature T_{end}=90 °C were set with a heating rate of 1 °C or 1.5 °C per minute. The different heating rates did not show effects on the observed denaturation curves.

7.4.3 Dynamic Light Scattering (DLS)

Dynamic light scattering is a technique to determine the hydrodynamic radius of a particle in solution. During the experimental procedure, all samples were centrifuged at 15,000 r.c.f. for 10 min to remove aggregates and measured in a DynaPro NanoStar (Wyatt Technology) DLS device using the appropriate single use cuvettes. For each condition, three measurements were done at a sample temperature of T=25 °C and three measurement cycles of each 20 single data acquisitions with acquisition times of t=3 s. The device was controlled and evaluation of the data was done with the DYNAMICS® software (Wyatt Technology).

7.4.3.1 Visualizing heteromeric protein–protein interactions

To investigate changes of the hydrodynamic radius upon complex formation, protein solutions were initially prepared at a final concentration of c=40 µM of each component. For an improved signal, the concentrations were set to or c=114 µM in a second measurement. No significant changes were observed upon increasing the concentration. For each condition, a minimum of three measurements was done at a sample temperature of T=25 °C and three measurement cycles of each 20 single data acquisitions with acquisition times of t=3 s.

7.4.3.2 Investigating cA_4-induced oligomerization of CalpL

To investigate the cA_4-induced oligomerization of CalpL, samples at different concentrations were prepared (17.3 µM, 52.0 µM, 86.6 µM and 173 µM, either with (1:1 molar ratio) or without cA_4). The samples were measured using a DynaPro NanoStar (Wyatt Technology) DLS device using the appropriate single use cuvettes as described above. For initial measurements, the samples were measured directly after addition of cA_4, and after 5, 10, 15, 30, 60 min. For incubation times less than 10 min, the sample was not centrifuged after, but only before, cA_4 addition. To investigate the significance

of the radii difference between wild-type protein and mutants, a two-tailed unpaired t test was performed. The resulting significances are depicted in the figures by "n.s." ($p \geq 0.05$), "*" ($p < 0.05$), "**" ($p < 0.01$), and "***" ($p < 0.001$), and "****" ($p < 0.0001$).

7.4.4 Isothermal Titration Calorimetry (ITC) measurements

Isothermal titration calorimetry is a powerful technique to quantitatively determine thermodynamic parameters involved in molecular interactions. During a titration experiment, ITC directly measures changes in the total heat energy resulting from a binding event in solution providing valuable insights into the binding affinity (K_D), stoichiometry (n), enthalpy (ΔH) and entropy (ΔS) of a reaction. This method is suitable for the investigation of a broad range of molecular interactions, such as protein–ligand, protein–protein, and enzyme–substrate interactions in the range of 10^{-2} M$\geq K_D \geq 10^{-9}$ M (Lottspeich & Engels, 2012). ITC measurements were performed on a MicroCal PEAQ-ITC (Malvern Panalytical, UK) using the corresponding software for instrument control and data analysis. The parameters of the individual experiments are listed in **Table 7-9**, **Table 7-10**, **Table 7-11**, **Table 7-12**, and **Table 7-13**. Prior to each measurement, the sample cell was equilibrated three times with water and subsequently three times with protein buffer. Protein solutions were transferred into the sample cell with a 500 µl Hamilton syringe the concentration of remaining sample solution was determined by UV absorption at $\lambda = 280$ nm with a NanoDrop 2000 (Thermo Scientific, US). The titration syringe was loaded automatically. All titration experiments were done at 25 °C.

Table 7-9: Detailed information VcSiaP-VHH ITC experiments.

Experiment name / buffer composition	**Analyte** concentration	**Titrant** concentration	**# of injections /** injection volume
NbS001 vs. VcSiaP	**VcSiaP**	**NbS001**	
50 mM Tris, 50 mM NaCl,	176 µM	2.35 mM	**13** / 3 µl
pH 8.0	117 µM	906 µM	**13** / 3 µl
VcSiaP vs. NbS001	**NbS001**	**VcSiaP**	
50 mM Tris, 50 mM NaCl,	120 µM	1.17 mM	**13** / 3 µl
pH 8.0			
NbS002 vs. VcSiaP	**VcSiaP**	**NbS002**	
50 mM Tris, 50 mM NaCl,	109 µM	869 µM	**19** / 2 µl
pH 8.0	117 µM	1.00 mM	**19** / 2 µl
	114 µM	1.09 mM	**13** / 3 µl
	35.5 µM	364 µM	**13** / 3 µl
NbS002 vs. VcSiaP W73A	**VcSiaP W73A**	**NbS002**	
50 mM Tris, 50 mM NaCl,	49.4 µM	526 µM	**13** / 3 µl
pH 8.0	49.5 µM	520 µM	**13** / 3 µl
NbS002 vs. VcSiaP[NbS001]	**VcSiaP[NbS001]**	**NbS002**	
50 mM Tris, 50 mM NaCl,	40 µM	400 µM	**13** / 3 µl
pH 8.0			

Table 7-10: Detailed information SiaP-Neu5Ac ITC experiments.

Experiment name / buffer composition	**Analyte** concentration	**Titrant** concentration	**# of injections /** injection volume
Neu5Ac vs. VcSiaP	**VcSiaP**	**Neu5Ac**	
50 mM Tris, 50 mM NaCl,	107 µM	1.2 mM	**37** / 1 µl
pH 8.0	108 µM	1.2 mM	**37** / 1 µl
	543 µM	5.0 mM	**37** / 1 µl
	100 µM	1.2 mM	**19** / 2 µl
Neu5Ac vs. VcSiaP W73A	**VcSiaP W73A**	**Neu5Ac**	
50 mM Tris, 50 mM NaCl,	53.6 µM	500 µM	**19** / 2 µl
pH 8.0	47.6 µM	500 µM	**19** / 2 µl
Neu5Ac vs. HiSiaP	**HiSiaP**	**Neu5Ac**	
50 mM Tris, 50 mM NaCl,	49.4 µM	500 µM	**13** / 3 µl
pH 8.0	95.0 µM	1.0 mM	**19** / 2 µl
	50.0 µM	500 µM	**19** / 2 µl

Table 7-11: Detailed information competitive VcSiaP-VHH/Neu5Ac ITC experiments.

Experiment name / buffer composition	**Analyte** concentration	**Titrant** concentration	**# of injections /** injection volume
Neu5Ac vs. VcSiaP[NbS001]	**VcSiaP[NbS001]**	**Neu5Ac**	
50 mM Tris, 50 mM NaCl,	120 µM [120 µM]	1.2 mM	**13** / 3 µl
pH 8.0			
Neu5Ac vs. VcSiaP[NbS002]	**VcSiaP[NbS002]**	**Neu5Ac**	
50 mM Tris, 50 mM NaCl,	120 µM [120 µM]	1.2 mM	**13** / 3 µl
pH 8.0	120 µM [120 µM]	1.2 mM	**13** / 3 µl
	120 µM [60 µM]	1.2 mM	**13** / 3 µl
NbS002 vs. VcSiaP[Neu5Ac]	**VcSiaP[Neu5Ac]**	**NbS002**	
50 mM Tris, 50 mM NaCl,	60.0 µM	600 µM	**13** / 3 µl
pH 8.0	48.1 µM	519 µM	**13** / 3 µl
NbS002 vs. VcSiaP$_{CC*}$ [1)]	**VcSiaP$_{CC*}$** [1)]	**NbS002**	
50 mM Tris, 50 mM NaCl,	43.6 µM	500 µM	**13** / 3 µl
pH 8.0			

[1)] "VcSiaP$_{CC*}$" denotes double cysteine mutant VcSiaP S14C T192C preincubated with sialic acid. Disulfide bond formation was forced by addition of H_2O_2 to the substrate bound protein (see **Section 7.6**)

Table 7-12: Detailed information on CalpL-CalpT titration experiments

Experiment name / buffer composition	**Analyte** concentration	**Titrant** concentration	**# of injections /** injection volume
CalpT vs. CalpL	**CalpT**	**CalpL**	
20.75 mM Tris, 117.5 mM NaCl, 1.5 %	38.4 µM	415 µM	**19** / 2 µl
glycerol, 0.15 mM DTT, pH 8.0	50.2 µM	627 µM	**25** / 1.5 µl
CalpT A201E vs. CalpL	**CalpT A201E**	**CalpL**	
20.75 mM Tris, 117.5 mM NaCl, 1.5 %	37.5 µM	415 µM	**19** / 2 µl
glycerol, 0.15 mM DTT, pH 8.0			

Table 7-13: Detailed information on HiSiaP-VHH ITC experiments.

Experiment name / buffer composition	**Analyte** concentration	**Titrant** concentration	**# of injections /** injection volume
NbS003 vs. HiSiaP	**HiSiaP**	**NbS003**	
50 mM Tris, 50 mM NaCl, pH 8.0	50 µM	513 µM	**13** / 3 µl
	54 µM	514 µM	**13** / 3 µl
HiSiaP vs. NbS003	**NbS003**	**HiSiaP**	
50 mM Tris, 50 mM NaCl, pH 8.0	124 µM	1.24 mM	**19** / 2 µl
	124 µM	1.24 mM	**19** / 2 µl
	124 µM	1.24 mM	**19** / 2 µl
HiSiaP vs. NbS004	**NbS004**	**HiSiaP**	
50 mM Tris, 50 mM NaCl, pH 8.0	115 µM	1.20 mM	**19** / 2 µl
HiSiaP vs. NbS005	**NbS005**	**HiSiaP**	
50 mM Tris, 50 mM NaCl, pH 8.0	102 µM	1.04 mM	**19** / 2 µl
HiSiaP vs. NbS006	**NbS006**	**HiSiaP**	
50 mM Tris, 50 mM NaCl, pH 8.0	102 µM	1.02 mM	**19** / 2 µl
HiSiaP vs. NbS009	**NbS009**	**HiSiaP**	
50 mM Tris, 50 mM NaCl, pH 8.0	107 µM	935 mM	**19** / 2 µl
HiSiaP vs. NbS010	**NbS010**	**HiSiaP**	
50 mM Tris, 50 mM NaCl, pH 8.0	91.6 µM	1.02 mM	**19** / 2 µl
HiSiaP vs. NbS011	**NbS010**	**HiSiaP**	
50 mM Tris, 50 mM NaCl, pH 8.0	99.9 µM	1.02 mM	**19** / 2 µl
Neu5Ac vs. HiSiaP[NbS004]	**HiSiaP[NbS004]**	**Neu5Ac**	
50 mM Tris, 50 mM NaCl, pH 8.0	50 µM [55 µM]	500 µM	**19** / 2 µl
Neu5Ac vs. HiSiaP[NbS005]	**HiSiaP[NbS005]**	**Neu5Ac**	
50 mM Tris, 50 mM NaCl, pH 8.0	50 µM [55 µM]	500 µM	**19** / 2 µl
Neu5Ac vs. HiSiaP[NbS006]	**HiSiaP[NbS006]**	**Neu5Ac**	
50 mM Tris, 50 mM NaCl, pH 8.0	50 µM [55 µM]	500 µM	**19** / 2 µl
Neu5Ac vs. HiSiaP[NbS009]	**HiSiaP[NbS009]**	**Neu5Ac**	
50 mM Tris, 50 mM NaCl, pH 8.0	50 µM [55 µM]	500 µM	**19** / 2 µl
Neu5Ac vs. HiSiaP[NbS010]	**HiSiaP[NbS010]**	**Neu5Ac**	
50 mM Tris, 50 mM NaCl, pH 8.0	50 µM [55 µM]	500 µM	**19** / 2 µl
Neu5Ac vs. HiSiaP[NbS011]	**HiSiaP[NbS011]**	**Neu5Ac**	
50 mM Tris, 50 mM NaCl, pH 8.0	50 µM [55 µM]	500 µM	**19** / 2 µl
	50 µM [25 µM]	500 µM	**19** / 2 µl
HiSiaP[Neu5Ac] vs. NbS004	**NbS004**	**HiSiaP[Neu5Ac]**	
50 mM Tris, 50 mM NaCl, pH 8.0	46.8 µM	500 µM	**19** / 2 µl
HiSiaP[Neu5Ac] vs. NbS005	**NbS005**	**HiSiaP[Neu5Ac]**	
50 mM Tris, 50 mM NaCl, pH 8.0	51.6 µM	500 µM	**19** / 2 µl
HiSiaP[Neu5Ac] vs. NbS006	**NbS006**	**HiSiaP[Neu5Ac]**	
50 mM Tris, 50 mM NaCl, pH 8.0	48.6 µM	500 µM	**19** / 2 µl
HiSiaP[Neu5Ac] vs. NbS009	**NbS009**	**HiSiaP[Neu5Ac]**	
50 mM Tris, 50 mM NaCl, pH 8.0	47.8 µM	500 µM	**19** / 2 µl
HiSiaP[Neu5Ac] vs. NbS010	**NbS010**	**HiSiaP[Neu5Ac]**	
50 mM Tris, 50 mM NaCl, pH 8.0	54.4 µM	500 µM	**19** / 2 µl
HiSiaP[Neu5Ac] vs. NbS011	**NbS011**	**HiSiaP[Neu5Ac]**	
50 mM Tris, 50 mM NaCl, pH 8.0	48.8 µM	500 µM	**19** / 2 µl

7.4.4.1 Sequential ITC experiments

To analyze the competitive effect when titrating NbS002 to an analyte solution of VcSiaP and Neu5Ac, a multi-step sequential ITC experiment was designed. In a first step, sialic acid was titrated to VcSiaP. The volume that was added to the initially loaded 280 µl was removed from the cell (36.4 µl). The "new" analyte concentration was estimated by using the dilution of the initial concentration (V_0=280 µl, V_{end}=316.4 µl). To estimate the concentration of the titrant in the analyte solution, the "new" analyte concentration was multiplied by the final molar ratio resulting from the titration. The syringe of the ITC device was washed and then, another titrant was loaded. In a second titration step, NbS002 was used as analyte. After finishing this step, the excess cell volume was again removed and "new" concentrations of the analyte contents were calculated. In a final titration step, HiSiaP was used as titrant to detect free sialic within the analyte solution. Further details of the experimental parameters are listed in **Table 7-14**.

Table 7-14: Detailed information for sequential VcSiaP-VHH/Neu5Ac ITC experiments.

Experiment name / buffer composition		**Analyte** concentration	**Titrant** concentration	**# of injections /** injection volume
VcSiaP wild type		**VcSiaP**	**Neu5Ac**	
50 mM Tris,	**1-1**	49.4 µM	500 µM	**19** / 2 µl
50 mM NaCl, pH 8.0	**2-1**	46.3 µM	500 µM	**19** / 2 µl
		VcSiaP[Neu5Ac]	**NbS002**	
	1-2	43.7 µM [83 µM] [1)]	495 µM	**19** / 2 µl
	2-2	41 µM [82 µM] [1)]	491 µM	**19** / 2 µl
		VcSiaP[Neu5Ac/NbS002]	**HiSiaP**	
	1-3	38.7 µM [73 µM/81 µM][1)]	520 µM	**19** / 2 µl
	2-3	36.3 µM [72 µM/80 µM][1)]	544 µM	**19** / 2 µl
VcSiaP W73A		**VcSiaP** [2)]	**Neu5Ac**	
50 mM Tris,	**1-1**	49.4 µM	500 µM	**19** / 2 µl
50 mM NaCl, pH 8.0	**2-1**	53.6 µM	500 µM	**19** / 2 µl
		VcSiaP[2)][Neu5Ac] [2)]	**NbS002**	
	1-2	43.7 µM [85 µM] [1)]	491 µM	**19** / 2 µl
	2-2	47.4 µM [85 µM] [1)]	493 µM	**19** / 2 µl
		VcSiaP[Neu5Ac/NbS002] [2)]	**HiSiaP**	
	1-3	38.7 µM [75 µM/85 µM][1)]	544 µM	**19** / 2 µl
	2-3	41.9 µM [75 µM/83 µM][1)]	527 µM	**19** / 2 µl

[1)] concentrations are estimated using the maximum cell volume (280 µl), the amount of added titrant (observed final molar ratio from experiment) and the volume that has been removed from the cell prior to the next titration (36.4 µl).

[2)] highlights experiments where VcSiaP W73A mutant was used

7.4.5 Surface plasmon resonance (SPR) measurements

All SPR measurements were done on a Biacore 8K instrument (GE Healthcare) using a streptavidin-functionalized sensor chip (Serie S Sensor Chip SA, GE Healthcare). Data were recorded at a rate of 10 Hz and a flow cell temperature of 25 °C. The experiments were conducted in 25 mM Tris, 250 mM NaCl, 5% glycerol and 0.05% Tween-20, pH 8.0. For the immobilization three initial injections of 1 M NaCl in 50 mM NaOH (10 µl min^{-1}, 60 s) were followed by injection of the biotinylated CalpL construct,

R493C-biotin (86 nM, 5 µl min^{-1}, 180 s). Binding of cOA and CalpT was measured as single-cycle kinetic experiments. For the cOAs, a series of seven different concentrations (0.086, 0.26, 0.78, 2.33, 7, 21 and 63 nM) were injected at a flow rate of 30 µl min^{-1} (contact time of 120 s, dissociation time of 600 s). For CalpT and NiS038, a series of seven different concentrations (0.0625, 0.25, 1, 4, 16, 64 and 256 nM) was injected applying the same parameters as above. The recorded data were double referenced by reference flow cell and blank cycle subtraction, and data were analyzed and fitted using Biacore Insight Evaluation software.

7.4.6 Protease activity assay

For the protease activity assays, CalpL (4.64 µM) and CalpT (5.11 µM) were prepared in 25 mM Tris, 500 mM NaCl, 5 % glycerol, 1 mM DTT, pH 8.0 and incubated for 1 h at 60 °C. Subsequently, cOAs (5.11 µM) were added and the mixture was incubated for another 1 h at 60 °C. For SDS–PAGE analysis, 3 µl of 4× SDS loading buffer was added to 9 µl of the sample. The mixture was heated for 5 min at 94 °C and 10 µl was loaded onto a 15 % polyacrylamide gel, which was run at 250 V for 40 min. To investigate the effect of mutants, the incubation times and temperatures where changed, these changes are indicated in the according results.

7.5 Crystallization and Structure Determination

7.5.1 Crystallization of proteins and protein complexes

All crystallization experiments were done using the vapor diffusion method. Sitting-drop crystallization plates (MRC 2 Well Crystallization Plate, Jena Bioscience) were set up by hand or using a gryphon crystallization robot (Art Robbins, US). For all experiments, a 1:1 ratio of mother liquor:protein solution was used. The plates were incubated at 20 °C in a Rock Imager 1000 crystallization hotel (Formulatrix, US) and imaged automatically in predefined time intervals. Crystals were harvested and flash frozen in $N_{2(l)}$. Detailed information on the crystallization condition, protein concentrations, and cryo protectant solution are listed in **Table 7-15**. Crystals were stored and sent at liquid nitrogen temperature.

Table 7-15: Crystallization conditions that yielded the crystals used for structure determination.

Protein (complexes) / buffer composition	**PDB ID**	**Crystallization condition**	**Cryo protectant**	**Additional Information**
VcSiaP[NbS001] 50 mM Tris, 50 mM NaCl, pH 8.0	**9fvc**	**Morpheus®** B-9		590 µM VcSiaP 750 µM NbS001 0.3 µl + 0.3 µl drops
$VcSiaP_{holo}$[NbS001] 50 mM Tris, 50 mM NaCl, pH 8.0	**-**	**ProPlex** G12	35 % (v/v) glycerol	667 µM VcSiaP 1.4 mM Neu5Ac 695 µM NbS001 0.3 µl + 0.3 µl drops
VcSiaP[NbS002] 50 mM Tris, 50 mM NaCl, pH 8.0	**9fvb**	**Morpheus®** D1		440 µM VcSiaP 470 µM NbS002 0.3 µl + 0.3 µl drops

VcSiaP[NbS001, NbS002] 50 mM Tris, 50 mM NaCl, pH 8.0	**-**	**Morpheus®** G12		532 µM VcSiaP 539 µM NbS001 539 µM NbS002 0.3 µl + 0.3 µl drops
VcSiaP$_{W73A, holo}$[NbS002] 50 mM Tris, 50 mM NaCl, pH 8.0	**9fve**	**ProPlex** G3	35 % (v/v) glycerol	405 µM VcSiaP 411 µM NbS002 2 mM Neu5Ac 0.2 µl + 0.2 µl drops
CalpL 20 mM Tris, 50 mM NaCl, pH 8.0	**-**	0.1 M Tris pH 8.0 38.8% PEG400 0.29 M Li_2SO_4	+ 5 % (v/v) PEG400	346 µM CalpL
CalpL (SeMet derivative) 20 mM Tris, 50 mM NaCl, pH 8.0	**7qda**	0.1 M Tris pH 8.0 38.8% PEG400 0.29 M Li_2SO_4	+ 5 % (v/v) PEG400	350 µM CalpL
CalpL-T_{10} 50 mM Tris, 50 mM NaCl, pH 8.0	**8b0u**	**JCSG+** E2	35 % (v/v) glycerol	ca. 440 µM CalpL ca. 440 µM CalpT after incubation with cA_4 followed by SEC
CalpL[cA_4]	**8b0r**	0.1 M Tris pH 8.0 38.8% PEG400 0.29 M Li_2SO_4	+ 5 % (v/v) PEG400 + 5mM cA_4	346 µM CalpL 3 min soaking
CalpT[NbS023]	**-**	**Morpheus®** G4		233 µM CalpT 256 µM NbS023

7.5.2 X-ray diffraction and data analysis

7.5.2.1 Acquisition of X-ray diffraction data

Diffraction data was collected either at beamline PX10 (Zuerich, Switzerland) with Pilatus 2M detector ([1]) or at beamline P13 operated by EMBL Hamburg at the PETRA III storage ring (DESY) with an Eiger 16M detector ([2]). Detailed information on the recorded data sets can be found in **Table 7-16**.

7.5.2.2 Processing of X-ray diffraction data

Processing of the X-ray diffraction data was carried out with the following software: AutoProc (Vonrhein & Bricogne, 2008) ,XDS (Kabsch et al., 2010), POINTLESS (P. Evans, 2006), AIMLESS(P. R. Evans & Murshudov, 2013).

7.5.3 Molecular replacement and experimental phasing

7.5.3.1 Molecular replacement to solve the phase problem

All structures were solved by molecular replacement with phenix.phaser (McCoy et al., 2007) using the structural models listed in **Table 7-16** as search model, except for CalpL which was solved by single-wavelength anomalous diffraction as described in the next section.

Table 7-16: Quality of the datasets and final refinements used for structure determination.

Protein (complexes)		Highest resolution	PDB ID for MR	R_{work}/R_{free}
VcSiaP[NbS001]	([1])	**2.6 Å**	4mag / 5ovw	0.195/0.257
$VcSiaP_{holo}$[NbS001]	([2])	**1.9 Å**	7a5q / 9fvc	0.192/0.232
VcSiaP[NbS002]	([1])	**2.1 Å**	4mag / 5ovw	0.195/0.257
VcSiaP[NbS001, NbS002]	([2])	**2.3 Å**	4mag / 9fvb / 9fvc	0.207/0.259
$VcSiaP_{W73A, holo}$[NbS002]	([2])	**2.8 Å**	7a5q / 9fvb	0.229/0.265
CalpL	([2])	**1.6 Å**	7qda	0.176/0.201
CalpL (SeMet derivative)	([2])	**2.1 Å**	SAD phasing[3]	0.193/0.228
$CalpL\text{-}T_{10}$	([2])	**3.3 Å**	7qda / AF2[4]	0.195/0.230
$CalpL[cA_4]$	([2])	**2.2 Å**	7qda	0.236/0.267
CalpT[NbS023]	([2])	**2.2 Å**	AF2[4] / 5ovw	0.240/0.282

[1] - Diffraction data collected at SLS (Zurich, Switzerland)
[2] - Diffraction data collected at DESY (Hamburg, Germany)
[3] - SeMet derivative of CalpL used for SAD phasing, as explained in detail
[4] - AlphaFold2 model was used for molecular replacement job

7.5.3.2 Single-wavelength anomalous diffraction (SAD) phasing

Since no suitable structural model for molecular replacement (sequence identity ≥ 30 %) was available from the PDB, the phase problem for the CalpL apo structure was solved by single-wavelength anomalous diffraction (SAD). A selenomethionine derivative of the protein had to be produced beforehand, the procedure is described in **Section 7.3.2.1**. Crystallization was achieved using the optimized crystallization screen based on condition D7 of the commercially available JCSG+ crystallization screen (Molecular Dimensions, UK). The crystals were harvested with additional 5 % (v/v) PEG 400 for cryoprotection. Prior to recording a diffraction data set, an energy scan was performed to identify the absorption edge of selenium to confirm the presence of the heavy atom within the crystal. A diffraction dataset was collected at beamline P13 of the PETRA III storage ring (Hamburg) using an X-ray wavelength of λ=0.9795 Å (≙12.659 keV), just above the absorption edge of selenium. The diffraction data were automatically processed using XDS and the structure was solved using phenix.autosol, and refined using phenix.refine (Liebschner et al., 2019).

7.5.4 Model building and refinement

All structural models were optimized with phenix.refine (Adams et al., 2010), COOT (Emsley & Cowtan, 2004) and ISOLDE (Croll, 2018). After each refinement step, the model quality was checked with MolProbity (V. B. Chen et al., 2010). All figures were prepared in PyMOL.

7.6 Design of a VcSiaP mutant that can be locked in its closed conformation

VcSiaP S14C T192C was incubated with sialic acid. After reducing the cysteine residues with 1 mM TCEP, separation of the reducing agent was followed by addition of H_2O_2 to force disulfide bond formation. This technique was already used in our lab for HiSiaP (Peter et al., 2024).

8. Materials

All used chemicals, enzymes and other consumables, as well as used hardware are listed below. Moreover, this section contains a summary of all buffers and crystallization screens that have been used during this study.

8.1 Chemicals

Table 8-1: List of used chemicals and the corresponding manufacturer.

Chemicals	Supplier / Manufacturer
Acrylamide (Rotiphorese Gel 30)	Carl Roth
Agar	Sigma-Aldrich (Merck)
Agarose	Carl Roth
Ammonium peroxydisulfate (APS)	Carl Roth
Ampicillin, sodium salt	Carl Roth
$CaCl_2$	Carl Roth
Coomassie Brilliant Blue R-250	Carl Roth
D(+)-Glucose	Carl Roth
D(+)-Maltose monohydrate	Carl Roth
D_2O	Deutero GmbH
EDTA (Ethlyenediaminetetraacetic acid)	Carl Roth
Glycerol	Carl Roth
Imidazole	Carl Roth
IPTG (Isopropyl-β-D-1-thiogalactopyra- noside)	Carl Roth
Kanamycin sulfate	Carl Roth
KH_2PO_4	Carl Roth
L(+)-arabinose	Carl Roth
$MgSO_4$	Carl Roth
Milk powder	Carl Roth
***N*-acetylneuraminic acid (Neu5Ac, Sialic Acid)**	Carbosynth Limited
Na_2HPO_4	Carl Roth
NaCl	Carl Roth
$NiSO_4$	Carl Roth
Nitrogen (liquid)	Linde GmbH
Sodium dodecylsulfate (SDS), 20%	AppliChem GmbH
Tetramethylethylenediamine (TEMED)	Carl Roth
Tris(hydroxymethyl)aminomethan (Tris)	Carl Roth
Tween 20	Carl Roth
β -Mercaptoethanol	AppliChem GmbH
cOmplete ULTRA, EDTA-free protease inhibitor	Roche
Cyclic oligoadenylates (cA_3, cA_4, 2'F-cA_4, cA_5, cA_6)	BioLog
Chloramphenicol	Carl Roth
Streptomycin sulfate	Carl Roth

Manufacturers and their place of business:

AppliChem GmbH (Darmstadt, Germany), Carbosynth Ltd. (UK), Biolog Life Science Institute GmbH & Co. KG (Bremen, Germany), Biosynth Ltd (, Carl Roth GmbH & Co. KG (Karlsruhe, Germany), Deutero GmbH (Kastellaun, Germany), Linde GmbH (Pullach, Germany), New England Biolabs (USA), Roche (Switzerland), Sigma-Aldrich (USA),

8.2 Consumables

Table 8-2: List of consumables and their corresponding manufacturers.

Consumables	Supplier / Manufacturer
96-deep well block (2 ml)	VWR
96-well microplates, round bottom	Greiner
Adhesive foil for crystallization plates	Molecular Dimensions
Blotting paper	Whatman plc
Falcon tubes (15 ml, 50 ml)	Greiner
Filter paper (0.22 µm)	Merck Millipore
HisPur™ Ni-NTA Resin	Thermo Fisher Scientific
Korasilon™ silicon paste	Carl Roth
Microcuvette kit for the NanoStar, disposable	Wyatt Technologies
MRC 2 Well Crystallization Plate (96-well)	Jena Bioscience
PCR tubes	Sarstedt
pH indicator paper	Carl Roth
Pipette tips (20 µl, 200 µl, 1000 µl)	VWR
Prometheus standard capillaries	NanoTemper Technologies
Reaction tubes (0.5 ml, 1 ml, 2 ml)	Eppendorf
Series S sensor chip SA	Cytiva
Serological pipettes (5 ml, 10 ml, 25 ml, 50 ml)	Sarstedt
SnakeSkin™ dialysis tubing (3.5kDa MWCO)	Thermo Fisher Scientific
SuperSignal™ West Pico PLUS Chemiluminescent Substrate	Invitrogen
Syringe filters (0.22 µm and 0.45 µm)	Carl Roth
Syringes (5 ml, 10 ml, 20 ml)	Carl Roth

Manufacturers and their place of business:
Cytiva (USA), Eppendorf (Hamburg, Germany), Greiner (Kremsmünster, Austria),
Jena Bioscience (Jena, Germany), Merck Millipore (USA),
Molecular Dimensions (UK), NanoTemper Technologies (München, Germany),
Sarstedt (Nürnberg, Germany), Thermo Fisher Scientific (USA),
VWR (Darmstadt, Germany), Whatman plc (UK), Wyatt Technologies (USA),

8.3 Enzymes, buffers and more

Table 8-3: List of used enzymes and buffers including the corresponding manufacturer.

Enzymes and buffers	Enzyme type	Supplier / Manufacturer
Restriction enzymes	Restriction Enzymes	New England Biolabs
OneTaq® DNA Polymerase	DNA Polymerase	New England Biolabs
Phusion DNA Polymerase	DNA Polymerase	New England Biolabs
Q5 DNA Polymerase	DNA Polymerase	New England Biolabs
T4 DNA Ligase	DNA Ligase	New England Biolabs
Polymerase buffers	Buffer	New England Biolabs
Restriction enzyme buffers	Buffer	New England Biolabs
T4 DNA Ligase buffer	Buffer	New England Biolabs

8.4 Buffers and media

8.4.1 Protein purification buffers

Table 8-4: List of protein purification buffers and their composition.

VcSiaP / HiSiaP buffers	
Buffer A (lysis/SEC)	50 mM Tris, 50 mM NaCl, pH 8
Ni^{2+}-elution buffer	50 mM Tris, 50 mM NaCl, 500 mM imidazole, pH 8
No salt buffer	50 mM Tris, pH 8
High salt buffer	50 mM Tris, 1 M NaCl, pH 8
P-domain nanobody buffers	
TES buffer	200 mM Tris, 0.65 mM EDTA, 0.5 M sucrose, pH 8
0.25x TES buffer	50 mM Tris, 0.16 mM EDTA, 125 mM sucrose, pH 8
Buffer A (SEC)	50 mM Tris, 50 mM NaCl, pH 8
Ni^{2+}-wash buffer	50 mM Tris, 300 mM NaCl, 20 mM imidazole, pH 8
Ni^{2+}-elution buffer	50 mM Tris, 50 mM NaCl, 500 mM imidazole, pH 8

CalpL buffers	
Buffer A (lysis/SEC)	20 mM Tris, 50 mM NaCl, pH 8
Ni^{2+}-wash buffer	20 mM Tris, 50 mM NaCl, 20 mM imidazole, pH 8
Ni^{2+}-elution buffer	20 mM Tris, 50 mM NaCl, 500 mM imidazole, pH 8
No salt buffer	20 mM Tris, pH 8
High salt buffer	20 mM Tris, 1 M NaCl, pH 8
CalpT buffers	
Buffer W (lysis/SEC)	25 mM Tris, 500 mM NaCl, 1 mM DTT, 10 % glycerol, pH 8
Ni^{2+}-wash buffer	25 mM Tris, 500 mM NaCl, 40 mM imidazole, 1 mM DTT, 10 % glycerol, pH 8
Ni^{2+}-elution buffer	25 mM Tris, 250 mM NaCl, 750 mM imidazole, 1 mM DTT, 10 % glycerol, pH 8
No salt buffer	25 mM Tris, 1 mM DTT, 10 % glycerol, pH 8
High salt buffer	25 mM Tris, 1 M NaCl, 1 mM DTT, 10 % glycerol, pH 8
CalpS buffers	
Buffer W (lysis/SEC)	25 mM Tris, 500 mM NaCl, 1 mM DTT, 10 % glycerol, pH 8
Ni^{2+}-wash buffer	25 mM Tris, 500 mM NaCl, 40 mM imidazole, 1 mM DTT, 10 % glycerol, pH 8
Ni^{2+}-elution buffer	25 mM Tris, 250 mM NaCl, 750 mM imidazole, 1 mM DTT, 10 % glycerol, pH 8
No salt buffer	25 mM Tris, 1 mM DTT, 10 % glycerol, pH 8
High salt buffer	25 mM Tris, 1 M NaCl, 1 mM DTT, 10 % glycerol, pH 8

CalpT-L nanobody buffers	
TES buffer	200 mM Tris, 0.65 mM EDTA, 0.5 M sucrose, pH 8
0.25x TES buffer	50 mM Tris, 0.16 mM EDTA, 125 mM sucrose, pH 8
Buffer A (SEC)	50 mM Tris, 50 mM NaCl, pH 8
Ni^{2+}-wash buffer	50 mM Tris, 300 mM NaCl, 20 mM imidazole, pH 8
Ni^{2+}-elution buffer	50 mM Tris, 50 mM NaCl, 500 mM imidazole, pH 8
15 % Buffer W	20.75 mM Tris, 117.5 mM NaCl, 0.15 mM DTT, 1.5 % glycerol, pH 8

8.4.1 Media for protein expression

Table 8-5: List of media for cultivation of* E. coli *cells.

Luria-Broth medium	
10 g/l	Trypton
5 g/l	Yeast extract
5 g/l	NaCl
Terrific-Broth medium	
12 g/l	Trypton
24 g/l	Yeast extract
4 ml/l	Glycerol
12.54 g/l	K_2HPO_4
2.31 g/l	KH_2PO_4
M9-minimal medium	
1 g/l	$(NH)_4Cl$
0.5 g/l	NaCl
50 ml/l	Glycerol
8.49 g/l	Na_2HPO_4
3 g/l	KH_2PO_4
pH	7.4

8.5 Bacterial strains

Table 8-6:* E. coli *strains used in this work.

Strain	**Usage**
β10	Mini cultures for sequencing
DH5α	Mini cultures for sequencing
C43 (DE3)	Expression of VcSiaP, HiSiaP, CalpS
BL21 (DE3)	Expression of CalpL, CalpT, CalpS
WK6	Expression of SiaP specific VHH antibodies
MC1061	Expression of CalpT–L specific VHH antibodies

8.6 Plasmids and Oligonucleotides

Table 8-7: Plasmids used in this work and their characteristics.

Plasmid	Characteristics	Source
pBadHisTEV	N-terminal His_6 tag TEV protease cleavage site Ampicillin resistance pBAD promoter	Huanting Liu (University of St. Andrews)
pET11a	Ampicillin resistance Lac promoter	BioCat
pET28a	Kanamycin resistance Lac promoter	Novagen
pHEN6	N-terminal pelB signal peptide C-terminal HA-His_6 sequence Ampicillin resistance Lac promoter	Core Facility Nanobodies (University of Bonn)
pSBinit2	N-terminal pelB signal peptide C-terminal HA-His_6 sequence Chloramphenicol resistance pBAD promoter	Core Facility Nanobodies (University of Bonn)

Table 8-8: Oligonucleotides for site-directed mutagenesis used in this work.

Primer	Sequence
VcSiaP_W73A	GGGGCTTGCGATACCGCGAGCAGAAGCGGT
	GTATCGCAAGCCCCATACGACCAAACTCAGCATAAGT
NbS2_F101A	GATCACGCACACGTAACACATCGAAAGTATGACTACTGG
	TTACGTGTGCGTGATCTGCTGCACAGTAATAAATGGCC
CalpT_A189E	CCTGGAGGAACCGGAAGAACAGGTTATTGAAATTGCACC
	CCGGTTCCTCCAGGATTTCAAAAATGCGGGTATGAAATTCTTT
CalpT_A199E	GAAATTGAGCCGGAACGTATTAGTGAATTTGTGCTGCGC
	GTTCCGGCTCAATTTCAATAACCTGTTCTTCCGGTTCG
CalpT_A212E	CATGTTGAGAGTACCAGTCAGAAAGCAACCTATACCGATG
	CTGGTACTCTCAACATGGCGCAGCACAAATTCACTAATAC
CalpT_A218E	CAGAAAGAGACCTATACCGATGATTTTGTTCTGTATCGTGGC
	GGTATAGGTCTCTTTCTGACTGGTACTGGCAACATGGC
CalpT_E105C	CAGCtgcGAAATTAGTAAATTCATTCTGATCG
	AATTTCgcaGCTGGTCAGGTAAAAATTATTGG
CalpT_E119C	GCAAAtgcGATCTGACAAATATTCTGAAATTC
	GATCgcaTTTGCTCAGGATGTCGATC
CalpT_150N	ATTCGCaacAAATTCAAAAAGGAAGAATTCAAC
	GAATTTgttGCGAATATCATCTTCAAAAATCG
CalpT_K151C	GCAACtgcTTCAAAAAGGAAGAATTCAACCAG
	TTTTGAAgcaGTTGCGAATATCATCTTCAAAAATCGG
CalpL_N233Q	catatcCAGaagccggaattagccctg
	ggcttCTGgatatgaataaagatgactggcag
CalpL_H345R	catttatCGAtatcagaaggaatatcacaaagtg
	ctgataTCGataaatgatgctggcttgc
CalpL_Y346E	tttatcatGAGcagaaggaatatcacaaagtgattg
	cttctgCTCatgataaatgatgctggcttg
CalpL_A390F	atcttTTCagccataatccgatagaaaagggtc
	ggctGAAaagataaatgataatcatcagagggt
CalpL_R338E	caacGAAcaagccagcatcatttatcattatcag
	cttgTTCgttgccgagaatgactccg
CalpL_R361E	cccgGAAaaaattaaagagaaaaagagcgagtttgag
	ctttaattttTTCcgggttatccgtaaggtcaatcac
CalpL_K364E	gaaaaattGAAgagaaaaagagcgagtttg
	ctttttctcTTCaatttttctcgggttatccgtaag
CalpL_K366E	attaaagagGAAaagagcgagtttgagaaaatttctgttaataaaaac
	gctcttTTCctctttaatttttctcgggttatccg
CalpS_R80E	gctatattcagGAAatgaccaaaaattttctggcagatgtgtatgcaag
	ggtcatTTCctgaatatagctaaccagaccgctctgctg
CalpS_D88R	ctggcaCGTgtgtatgcaagcgttaaactgatgagc
	catacacACGtgccagaaaatttttggtcatacgctgaatatag
CalpS_A91R	gtgtatCGAagcgttaaactgatgagcgaaaatg
	cgctTCGatacacatctgccagaaaatttttgg
CalpT_W4E	CTTTCAGctcTTTGGCCATACCCTGGAAATACAG
	CCAAAgagCTGAAAGATCTGTATAATGAATACATTGAGGAG
CalpT_E15R	CAGTTCcctCTCAATGTATTCATTATACAGATCTTTCAGCCATTTG
	CATTGAGaggGAACTGGAAGAAGATCTGACCAGTCATATTAG
CalpT_D72R	GCCAacgACTCACCTTCATGATTTCATACAGATTATTATC
	GTGAGTcgtTGGCATCATTTTGCCAGCAATAC
CalpT_M90E	GGGTctcGGTCGGCAGTTCAATAAAAATTTCGGTATTG
	CGACCgagACCCTGATTATTGAAACCACCAATAATTTTTAC
CalpT_I111R	CGATCAGacgGAATTTACTAATTTCTTCGCTGGTCAG
	GTAAATTCcgtCTGATCGACATCCTGAGCAAAGAAG
CalpL_S152A	ggcaaaGCTtatcagctggcagtagcagcg
	gctgataAGCtttgccagaaaagatattatcaaagatgacgaaaaag
VcSiaP_A43C	caagcTGTcagcttggtgatgatcgtgc
	gctgACAgcttgggtacaaagcgagtttg
VcSiaP_S169C	gtcattcTGTgaagtttatttagcgctgcagaccaatg
	ctaaataaacttcACAgaatgacatcgggggttggcg
VcSiaP_T192C	caacaattaaaTGCatgaagttctatgaagtgcaaaagaacttagcc
	gaacttcatGCAtttaattgttggtagcgggttttcttgcc

Table 8-9: Oligonucleotides used for PCR amplification of genes or gene fragments.

Primer	Sequence
CalpL_EcoRI_rev	GATCGAATTCtcagaacggggagc
CalpL_SacI_fwd	GATCGAGCTCATGcatattaaacagctgctcaa
CalpL_SalI_rev	GATCGTCGACtcagaacggggagc
CalpL_XhoI_fwd	GATCCTCGAGATGcatattaaacagctgctcaa
CalpS_BamHI_fwd	GATCGGATCcatgagcggcaatgatc
CalpS_BamHI_rev	CTAGggatccttagccgttac
CalpS_NcoI_fwd	GATCccatgggcatgagc
CalpS_SacI_rev	GATCGAGCTcttagccgttactgttcac
CalpT_195_SacI_rev	gatcgagctcTTAGGCAACATGGCGC
CalpT_BamHI_fwd	gatcggatccATGGCCAAATGGCTGAA
CalpT_BamHI_fwd	gatcggatccATGGCCAAATGGCTGAA
CalpT_BamHI_rev	GATCGGATCCTTAACCTTCCG
CalpT_NcoI_fwd	ctagCCATGGCCAAATGGCTGAA
CalpT_NcoI_fwd	ctagCCATGGCCAAATGGCTGAA
CalpT_Nterm-His_NdeI_fwd	CTAGCATATGGCCAAATGGCTGAAA
CalpT_Nterm+His_NdeI_fwd	CTAGCATATGCATCATCATCATCACC
CalpT-Cterm_fwd_SpeI	gatcactAGTCAGAAAGCAACCTATACC
CalpT-Cterm_rev_XhoI	gatcctcgagTTAACCTTCCGGCAGAATG
CalpT-full_SacI_rev	gatcgagctcTTAACCTTCCGGCAG

8.7 Instruments and Columns for protein purification and analysis

Table 8-10: Instruments used for protein purification and analysis.

Instrument	Name	Manufacturer
Cell sonicator	SONOPULS HD3100	Bandelin electronic
Centrifuge	Avanti JXN 26	Beckman Coulter
Crystal loops	Micro loops	MiTeGen
Crystallization imager	Rock Imager 1000	Fromulatrix
Crystallization robot	Crystal Gryphon LCP	Art Robbins Instruments
DLS	DynaPro NanoStar	Wyatt
FPLC	Äkta avant	GE Healtcare
FPLC	Äkta pure	GE Healtcare
Gel imager	ChemiDoc XRS+	Bio-Rad Laboratories
HPLC	1260 Infinity II	Agilent
Incubators	Multitron Minitron	Infors HT
ITC	MicroCal PEAQ-ITC	Malvern Panalytical
MALS	miniDAWN	Wyatt
nanoDSF	Prometheus NT.48	NanoTemper
PCR thermocycler	Mastercycler Nexus SX1	Eppendorf
Pipetting robot	epMotion 5073	Eppendorf
SEC columns	SD75 16/600; 10/300; 3.2/300 SD200 16/600; 10/300; 3.2/300 Superose6 16/600; 10/300	Cytiva
Spectrophotometer	NanoDrop DN-2000C	Thermo Scientific

SPR	Biacore 8K	GE Healthcare
Table top centrifuge (large)	Centrifuge 5804 R	Eppendorf
Table top centrifuge (small)	Centrifuge 5427 R	Eppendorf
Ultrasonic Bath	USC100T	VWR
Water purification system	Milli-Q Direct	Merck Millipore

8.8 Software

Table 8-11: List of used software.

Software name	Developer / Company
Affinity Designer	Serif Ltd
AlphaFold2	Jumper et al., 2021
AlphaFold3	Abramson et al., 2024
Biacore Insight Evaluation	GE Healthcare
CCP4i	(Collaborative Computational Project, Number 4, 1994)
ChemDraw 20.0	PerkinElmer
ChimeraX-1.5	(RBVI), UCSF
COOT 0.9.8.8	Emsley et al., 2010
DataGraph	Visual Data Tools
DYNAMICS	Wyatt Technologies
FoldSeek	Kempen et al., 2024
GraphPad Prism 10	Dotmatics
HullRad	Fleming and Fleming, 2018
ISOLDE	Tristan Croll at Altos Labs
LigPlot	Laskowski and Swindells, 2011
MtsslWizard	Hagelueken et al., 2015
Phaser	McCoy et al., 2007
Phenix	Liebschner et al., 2019
ProtParam (Expasy)	Swiss Institute of Bioinformatics
PyMOL	Schrödinger LLC
SnapGene	Dotmatics
WebFlags	Saha et al., 2020
XDS	Kabsch et al., 2010

9. Additional Material

Below, a collection of additional figures and tables can be found. These materials have been referenced within the main text to gain further information. However, these figures might not be essential for following the story.

9.1 Additional Figures (Part I)

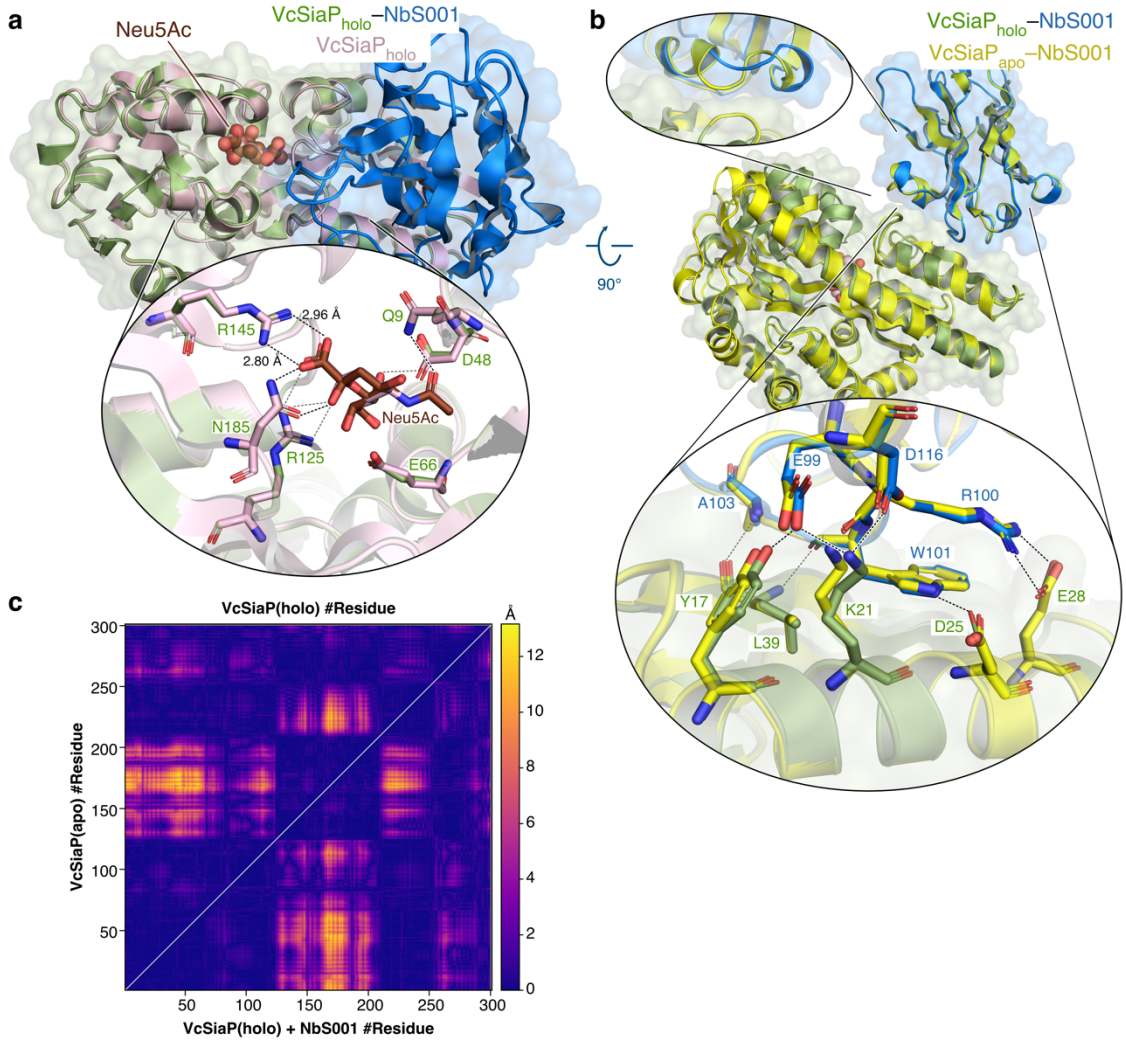


Figure 9-1: Demonstration of the functional closing mechanism of the VcSiaP/NbS001 complex. ***a)*** *Comparison of nanobody free and -bound VcSiaP$_{holo}$ structures. A magnification shows the interactions (dashed lines) between the SBP and sialic acid, the bond lengths of the strong coordination of R145 to the carboxyl group was measured.* ***b)*** *Comparison of substrate free and -bound VcSiaP/NbS001 complex structures.* ***c)*** *Difference distance matrices generated with MtsslWizard (Hagelueken et al., 2012) to demonstrate a fully functional closing mechanism showing that for VcSiaP$_{apo}$ and VcSiaP$_{holo}$ in the upper left triagle and that of VcSiaP$_{apo}$ and VcSiaP$_{holo}$/NbS001 in the lower right part.*

9.2 Additional Tables (Part I)

Extended Data Table 1: Crystallographic data of VcSiaP$_{apo}$–nanobody 1:1 complexes.

	VcSiaP_apo-NbS001	**VcSiaP_apo-NbS002**
Wavelength	0.999 Å	0.999 Å
Resolution range	47.18 - 2.64 (2.71 - 2.64)	46.40 - 2.05 (2.13 - 2.05)
Space group	P 21 21 2	C 1 2 1
Unit cell	169.6 72.4 85.6 90 90 90	151.3 50.3 134.0 90 114.9 90
Total reflections	63055 (6023)	114870 (10986)
Unique reflections	31540 (3021)	57561 (5513)
Multiplicity	2.0 (2.0)	2.0 (2.0)
Completeness (%)	99.61 (95.33)	99.54 (95.99)
Mean I/sigma(I)	6.90 (1.34)	7.48 (1.04)
Wilson B-factor	35.22	26.91
R-merge	0.09713 (0.5301)	0.09801 (0.7936)
R-meas	0.1374 (0.7496)	0.1386 (1.122)
R-pim	0.09713 (0.5196)	0.09801 (0.7936)
CC1/2	0.988 (0.638)	0.992 (0.436)
CC*	0.997 (0.883)	0.998 (0.779)
Reflections used in refinement	31516 (3009)	57530 (5499)
Reflections used for R-free	2000 (191)	1997 (189)
R-work	0.1947 (0.2644)	0.1952 (0.3056)
R-free	0.2569 (0.3665)	0.2572 (0.3534)
CC(work)	0.959 (0.826)	0.962 (0.728)
CC(free)	0.919 (0.586)	0.918 (0.487)
Number of non-hydrogen atoms	6701	7186
macromolecules	6644	6647
ligands	0	106
solvent	171	493
Protein residues	852	846
RMS(bonds)	0.013	0.013
RMS(angles)	1.36	1.20
Ramachandran favored (%)	97.04	97.85
Ramachandran allowed (%)	2.61	2.03
Ramachandran outliers (%)	0.36	0.12
Rotamer outliers (%)	1.54	0.56
Clashscore	7.06	6.04
Average B-factor	44.72	34.38
macromolecules	44.77	34.26
ligands	-	48.22
solvent	38.79	34.66
PDB accession code	9fvc	9fvb

Statistics for the highest-resolution shell are shown in parentheses.

Extended Data Table 2: Crystallographic data of VcSiaP W73A$_{holo}$–NbS002 complex.

	VcSiaP W73A_holo-NbS002
Wavelength	0.97626 Å
Resolution range	33.8 - 2.81 (2.91 - 2.81)
Space group	C 1 2 1
Unit cell	223.6 153.1 210.5 90 90 90
Total reflections	340130 (33893)
Unique reflections	170810 (17066)
Multiplicity	2.0 (2.0)
Completeness (%)	99.05 (99.39)
Mean I/sigma(I)	8.09 (2.41)
Wilson B-factor	44.90
R-merge	0.06463 (0.3428)
R-meas	0.0914 (0.4848)
R-pim	0.06463 (0.3428)
CC1/2	0.994 (0.731)
CC*	0.999 (0.919)
Reflections used in refinement	170742 (17064)
Reflections used for R-free	2002 (200)
R-work	0.2296 (0.2955)
R-free	0.2649 (0.3374)
CC(work)	0.934 (0.771)
CC(free)	0.882 (0.607)
Number of non-hydrogen atoms	39979
macromolecules	39703
ligands	528
solvent	0
Protein residues	5063
RMS(bonds)	0.018
RMS(angles)	1.97
Ramachandran favored (%)	97.37
Ramachandran allowed (%)	1.89
Ramachandran outliers (%)	0.74
Rotamer outliers (%)	3.02
Clashscore	15.29
Average B-factor	45.04
macromolecules	45.07
ligands	40.05
solvent	
PDB accession code	9fve

Statistics for the highest-resolution shell are shown in parentheses.

Extended Data Table 3: Table1 for VcSiaP[NbS001/NbS002] and VcSiaP$_{holo}$[NbS001].

	VcSiaP_apo-NbS001-NbS002	**VcSiaP_holo-NbS001**
Wavelength	0.97626 Å	0.826555 Å
Resolution range	52.74 - 2.28 (2.36 - 2.28)	37.75 - 1.96 (2.03 - 1.96)
Space group	P 1	P 31 2 1
Unit cell	55.2 66.6 98.0 101.7 104.2 96.5	71.5 71.5 190.4 90 90 120
Total reflections	167391 (17385)	82893 (8192)
Unique reflections	57307 (5711)	41447 (4096)
Multiplicity	2.9 (3.0)	2.0 (2.0)
Completeness (%)	96.25 (96.74)	99.86 (99.80)
Mean I/sigma(I)	9.34 (2.37)	14.06 (1.03)
Wilson B-factor	52.11	43.17
R-merge	0.06688 (0.4817)	0.02306 (0.6818)
R-meas	0.08171 (0.591)	0.03261 (0.9642)
R-pim	0.04618 (0.3374)	0.02306 (0.6818)
CC1/2	0.995 (0.817)	1 (0.483)
CC*	0.999 (0.948)	1 (0.807)
Reflections used in refinement	57263 (5705)	41443 (4092)
Reflections used for R-free	2805 (269)	2004 (197)
R-work	0.2073 (0.3501)	0.1924 (0.3351)
R-free	0.2590 (0.4301)	0.2324 (0.3789)
CC(work)	0.966 (0.810)	0.968 (0.670)
CC(free)	0.931 (0.622)	0.970 (0.601)
Number of non-hydrogen atoms	8625	3527
macromolecules	8520	3332
ligands	0	147
solvent	315	384
Protein residues	1094	427
RMS(bonds)	0.013	0.013
RMS(angles)	1.28	1.23
Ramachandran favored (%)	96.03	97.87
Ramachandran allowed (%)	3.79	2.13
Ramachandran outliers (%)	0.18	0.00
Rotamer outliers (%)	1.32	0.00
Clashscore	9.22	4.30
Average B-factor	64.32	46.70
macromolecules	64.36	46.51
ligands		52.22
solvent	61.44	48.56

Statistics for the highest-resolution shell are shown in parentheses.

9.3 Additional Figures (Part II)

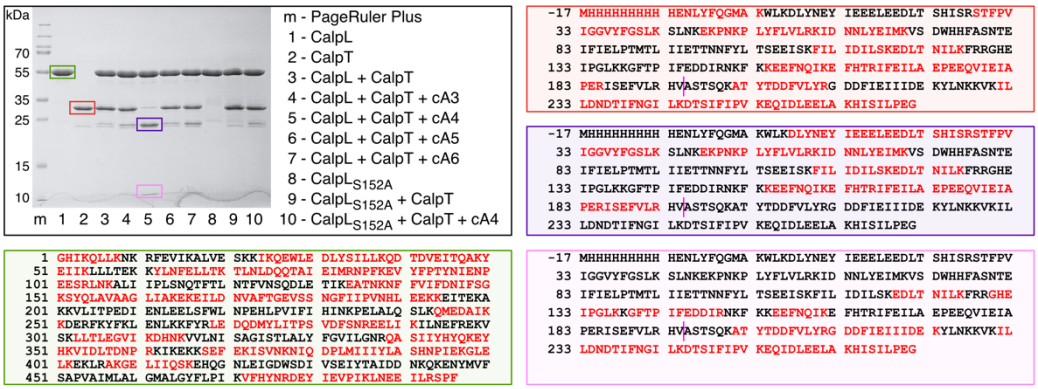


Figure 9-2: Mass spectrometric analysis of CalpL, CalpT and the proteolytiy cleavage fragments.
15 % Polyacryl amide gel with samples from a protease activity assay. The indicated gel bands were cut out and sent for mass spectrometric analysis. The results from peptide mass fingerprinting are shown in the colored boxes. The box color corresponds to the color used to highlight the cutted gel band. The complete protein sequences are shown and covered peptide sequences are highlighted red.

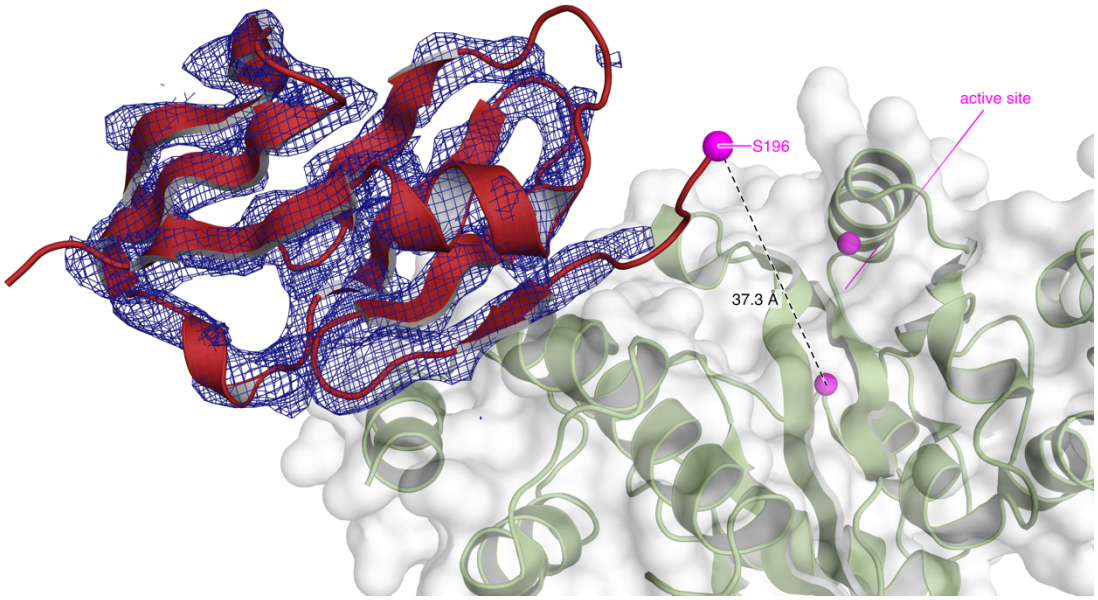


Figure 9-3: Additional illustration of the CalpT$_{10}$–CalpL complex.
CalpT$_{10}$ is shown as cartoon representation within its electron density and CalpL is presented as cartoon and transparent surface model. CalpT residue S196 as well as the catalytic residues S152 and K193 of CalpL are highlighted as magenta spheres. The shortest distance between the cleavage site of CalpT (S196) and the protease active site was measured to 37.3 Å as indicated by a dashed line.

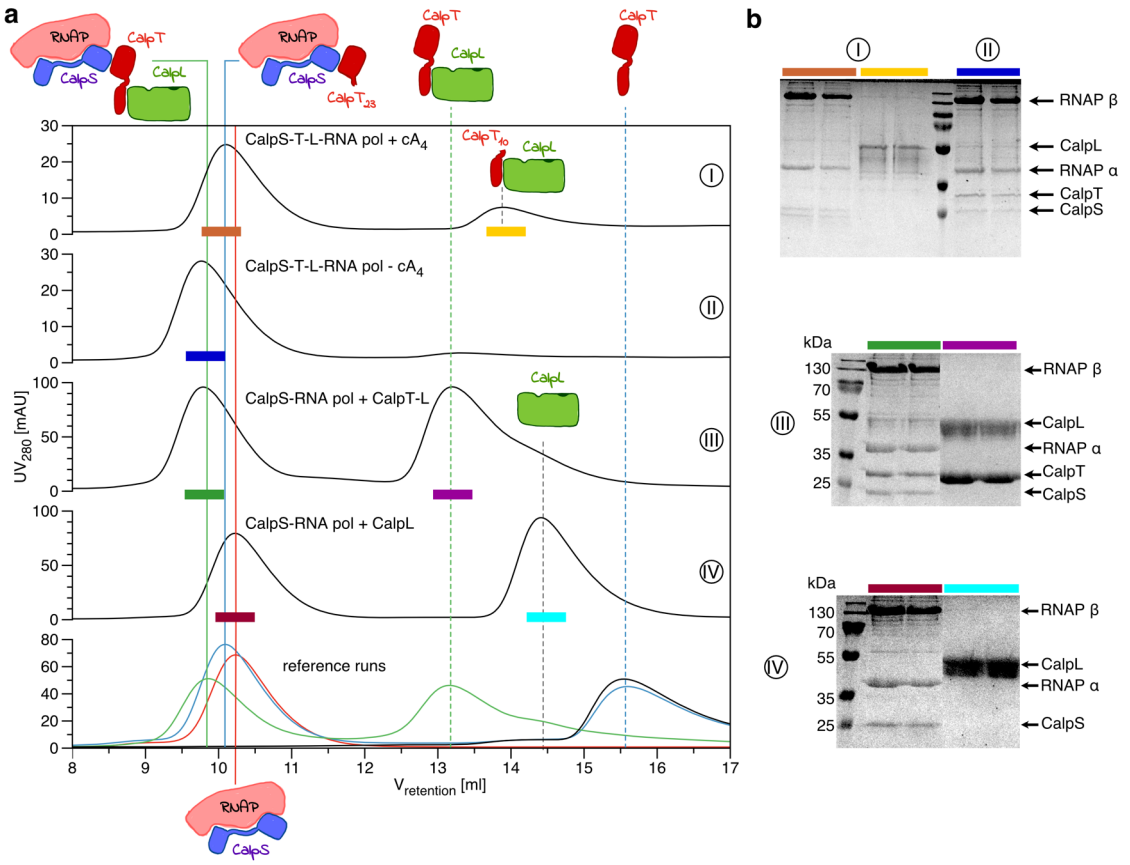


Figure 9-4: Further analysis of CalpS-T-L and RNA polymerase complexes.
a) *Results from analytical gel filtration experiments of CalpS with and without CalpT and CalpL, respectively. Each chromatogram peak is indicated by a pictogram to illustrate the components. Colored bars below the chromatograms refer to the respective gel samples depicted in b). For all graphs, the UV absorpiton at λ=280 nm was plotted against the retention volume. The contents which were preincubated prior to the SEC runs are annotated.* ***b)*** *SDS-PAGE analyses of samples taken from the analytical gel filtration runs shown in a).*

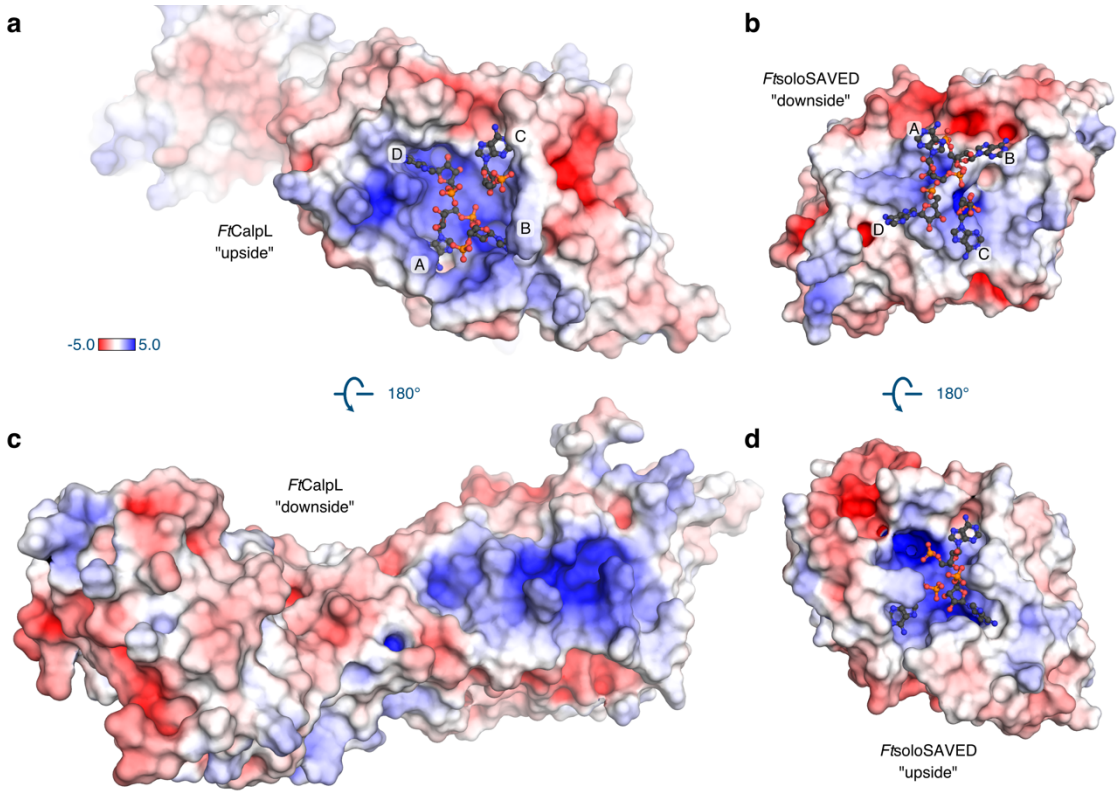


Figure 9-5: Electrostatic surface analysis of the* Ft*CalpL and* Ft*soloSAVED.
a)-b) *Open book' visualization of the two interacting proteins CalpL and soloSAVED from an AlphaFold3 model of the heterodimeric complex from* Fervidobacterium thailandense *held together by four AMP molecules that mimic cyclic oligoadenlyte. The surfaces are colored by their electrostatuc potential. The bases of the four AMP molecules are annotated to highlight the complementary regions. Note that the four bridging AMP molecules were copied to be shown on both sides.* ***c)-d)*** *The same as shown in a)-b), but both proteins are rotated by 180° as indicated.*

9.4 Additional Tables (Part II)

Extended Data Table 4: Crystallographic data of CalpL and CalpL in complex with cA_4.

	CalpL (SeMet derivative)	**CalpL + cA_4**
Resolution range	54.06 - 2.1 (2.18 - 2.1)	49.2 - 2.2 (2.28 - 2.2)
Space group	P 32 2 1	P 32 2 1
Unit cell	68.9 68.9 255.4 90.0 90.0 120.0	69.5 69.5 256.3 90.0 90.0 120.0
Unique reflections	41842 (3732)	37614 (3685)
Multiplicity	2.0 (2.0)	2.0 (2.0)
Completeness (%)	99.02 (90.28)	99.98 (100)
Mean I/sigma(I)	23.30 (2.36)	8.37 (2.33)
R-merge	0.0540 (0.344)	0.0347 (0.246)
R-work	0.1932	0.2450
R-free	0.2255	0.2747
Number of non-hydrogen atoms	4487	4246
macromolecules	4175	4096
ligands	69	113
solvent	243	37
RMS(bonds)	0.003	0.002
RMS(angles)	0.567	0.535
B-factors		
macromolecules	43.7	41.7
ligands	-	48.1
solvent	46.6	41.8
PDB accession code	7qda	8b0r

Statistics for the highest-resolution shell are shown in parentheses.

Extended Data Table 5: Table1 for CalpL[CalpT] and CalpT[NbS023] complexes.

	CalpL-T_{10}	**CalpT[NbS023]**
Resolution range	49.3 - 3.29 (3.41 - 3.29)	48.85 - 2.183 (2.261 - 2.183)
Space group	P 21	P 21 21 21
Unit cell	89.8 96.6 131.7 90.0 98.2 90.0	54.0 85.3 114.3 90 90 90
unique reflections	33758 (3319)	26053 (922)
Multiplicity	6.8 (6.9)	2.0 (2.0)
Completeness (%)	99.12 (99.34)	98.4 (83.9)
Mean I/sigma(I)	9.92 (1.29)	7.77 (0.76)
R-merge	0.1467 (1.559)	0.03488 (0.6088)
R-work	0.1968	0.2329
R-free	0.2313	0.2876
Number of non-hydrogen atoms	9441	3249
macromolecules	9392	3191
ligands	47	69
solvent	2	39
RMS(bonds)	0.002	0.004
RMS(angles)	0.533	0.670
B-factors		
macromolecules	116.6	83.21
ligands	-	124.74
solvent	69.3	82.15
PDB accession code	8b0u	-

Statistics for the highest-resolution shell are shown in parentheses.

References

Aboudounya, M. M., & Heads, R. J. (2021). COVID-19 and Toll-Like Receptor 4 (TLR4): SARS-CoV-2 May Bind and Activate TLR4 to Increase ACE2 Expression, Facilitating Entry and Causing Hyperinflammation. *Mediators of Inflammation*, *2021*, 8874339. https://doi.org/10.1155/2021/8874339

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., … Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, *630*(8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, *66*, 213–221. https://doi.org/10.1107/s0907444909052925

Ades, S. E., Connolly, L. E., Alba, B. M., & Gross, C. A. (1999). The Escherichia coli ςE-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti-ς factor. *Genes & Development*, *13*(18), 2449–2461. https://doi.org/10.1101/gad.13.18.2449

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., & Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D: Biological Crystallography*, *68*, 352–367. https://doi.org/10.1107/s0907444912001308

Akiba, H., Tamura, H., Kiyoshi, M., Yanaka, S., Sugase, K., Caaveiro, J. M. M., & Tsumoto, K. (2019). Structural and thermodynamic basis for the recognition of the substrate-binding cleft on hen egg lysozyme by a single-domain antibody. *Scientific Reports*, *9*. https://doi.org/10.1038/s41598-019-50722-y

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular Biology of the Cell* (4th edition). Garland Science. https://www.ncbi.nlm.nih.gov/books/NBK26860/

Almagro-Moreno, S., & Boyd, E. F. (2009). Sialic acid catabolism confers a competitive advantage to pathogenic Vibrio cholerae in the mouse intestine. *Infection and Immunity*, *77*, 3807–3816. https://doi.org/10.1128/iai.00279-09

Amerik, A. Yu., Antonov, V. K., Gorbalenya, A. E., Kotova, S. A., Rotanova, T. V., & Shimbarevich, E. V. (1991). Site-directed mutagenesis of La protease. *FEBS Letters*, *287*(1–2), 211–214. https://doi.org/10.1016/0014-5793(91)80053-6

Angata, T., & Varki, A. (2002). Chemical diversity in the sialic acids and related α-keto acids: An evolutionary perspective. *Chemical Reviews*, *102*, 439–469. https://doi.org/10.1021/cr000407m

Arbabi-Ghahroudi, M. (2017). Camelid single-domain antibodies: Historical perspective and future outlook. *Frontiers in Immunology*, *8*. https://doi.org/10.3389/fimmu.2017.01589

Arbabi-Ghahroudi, M., Desmyter, A., Wyns, L., Hamers, R., & Muyldermans, S. (1997). Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. *FEBS Letters*, *414*(3), 521–526. https://doi.org/10.1016/s0014-5793(97)01062-4

Arbabi-Ghahroudi, M., Tanha, J., & MacKenzie, R. (2005). Prokaryotic expression of antibodies. *Cancer and Metastasis Reviews*, *24*(4), 501–519. https://doi.org/10.1007/s10555-005-6193-1

Arike, L., & Hansson, G. C. (2016). The Densely O-Glycosylated MUC2 Mucin Protects the Intestine and Provides Food for the Commensal Bacteria. *Journal of Molecular Biology*, *428*(16), 3221–3229. https://doi.org/10.1016/j.jmb.2016.02.010

Asaadi, Y., Jouneghani, F. F., Janani, S., & Rahbarizadeh, F. (2021). A comprehensive comparison between camelid nanobodies and single chain variable fragments. *Biomarker Research*, *9*(1), 87. https://doi.org/10.1186/s40364-021-00332-6

Athukoralage, J. S., Graham, S., Grüschow, S., Rouillon, C., & White, M. F. (2019). A Type III CRISPR Ancillary Ribonuclease Degrades Its Cyclic Oligoadenylate Activator. *Journal of Molecular Biology*, *431*(15), 2894–2899. https://doi.org/10.1016/j.jmb.2019.04.041

Athukoralage, J. S., McMahon, S. A., Zhang, C., Grüschow, S., Graham, S., Krupovic, M., Whitaker, R. J., Gloster, T. M., & White, M. F. (2020). An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. *Nature*, *577*(7791), 572–575. https://doi.org/10.1038/s41586-019-1909-5

Athukoralage, J. S., Rouillon, C., Graham, S., Grüschow, S., & White, M. F. (2018). Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate. *Nature*, *562*(7726), 277–280. https://doi.org/10.1038/s41586-018-0557-5

Athukoralage, J. S., & White, M. F. (2021). Cyclic oligoadenylate signaling and regulation by ring nucleases during type III CRISPR defense. *RNA*, *27*, 855–867. https://doi.org/10.1261/rna.078739.121

Bakaletz, L. O., & Novotny, L. A. (2018). Nontypeable Haemophilus influenzae (NTHi). *Trends in Microbiology*, *26*(Infect. Immun. 73 2005), 727–728. https://doi.org/10.1016/j.tim.2018.05.001

Barrangou, R. (2013). CRISPR-Cas systems and RNA-guided interference. *Wiley Interdisciplinary Reviews: RNA*, *4*(3), 267–278. https://doi.org/10.1002/wrna.1159

Barthel, T., Wollenhaupt, J., Lima, G. M. A., Wahl, M. C., & Weiss, M. S. (2022). Large-Scale Crystallographic Fragment Screening Expedites Compound Optimization and Identifies Putative Protein–Protein Interaction Sites. *Journal of Medicinal Chemistry*, *65*(21), 14630–14641. https://doi.org/10.1021/acs.jmedchem.2c01165

Bell, A., Severi, E., Owen, C. D., Latousakis, D., & Juge, N. (2023). Biochemical and structural basis of sialic acid utilization by gut microbes. *Journal of Biological Chemistry*, *299*, 102989. https://doi.org/10.1016/j.jbc.2023.102989

Benjamin, D. C., Berzofsky, J. A., East, I. J., Gurd, F. R. N., Hannum, C., Leach, S. J., Margoliash, E., Michael, J. G., Miller, A., Prager, E. M., Reichlin, M., Sercarz, E. E., Smith-Gill, S. J., Todd, P. E., & Wilson, A. C. (1984). The Antigenic Structure of Proteins: A Reappraisal. *Annual Review of Immunology*, *2*(1), 67–101. https://doi.org/10.1146/annurev.iy.02.040184.000435

Benjamin, D. C., & Perdue, S. S. (1996). Site-Directed Mutagenesis in Epitope Mapping. *Methods*, *9*(3), 508–515. https://doi.org/10.1006/meth.1996.0058

Berger, E. A., & Heppel, L. A. (1974). Different Mechanisms of Energy Coupling for the Shock-sensitive and Shock-resistant Amino Acid Permeases of Escherichia coli. *Journal of Biological Chemistry*, *249*(24), 7747–7755. https://doi.org/10.1016/s0021-9258(19)42031-0

Bergkessel, M. (2021). Bacterial transcription during growth arrest. *Transcription*, *12*, 232–249. https://doi.org/10.1080/21541264.2021.1968761

Berntsson, R. P. A., Smits, S. H. J., Schmitt, L., Slotboom, D. J., & Poolman, B. (2010). A structural classification of substrate-binding proteins. *FEBS Letters*, *584*, 2606–2617. https://doi.org/10.1016/j.febslet.2010.04.043

Bever, C. S., Dong, J. X., Vasylieva, N., Barnych, B., Cui, Y., Xu, Z. L., Hammock, B. D., & Gee, S. J. (2016). VHH antibodies: emerging reagents for the analysis of environmental chemicals. *Analytical and Bioanalytical Chemistry*, *408*, 5985–6002. https://doi.org/10.1007/s00216-016-9585-x

Binder*, S. C., Schneberger*, N., Schmitz, M., Engeser, M., Geyer, M., Rouillon, C., & Hagelueken, G. (2024). The SAVED domain of the type III CRISPR protease CalpL is a ring nuclease. *Nucleic Acids Research*, gkae676. https://doi.org/10.1093/nar/gkae676

Blanc, M. R., Anouassi, A., Abed, M. A., Tsikis, G., Canepa, S., Labas, V., Belghazi, M., & Bruneau, G. (2009). A one-step exclusion-binding procedure for the purification of functional heavy-chain and mammalian-type γ-globulins from camelid sera. *Biotechnology and Applied Biochemistry*, *54*(4), 207–212. https://doi.org/10.1042/ba20090208

Bohacek, R. S., McMartin, C., & Guida, W. C. (1996). The art and practice of structure-based drug design: A molecular modeling perspective. *Medicinal Research Reviews*, *16*(1), 3–50. https://doi.org/10.1002/(sici)1098-1128(199601)16:1<3::aid-med1>3.0.co;2-6

Bosdriesz, E., Magnúsdóttir, S., Bruggeman, F. J., Teusink, B., & Molenaar, D. (2015). Binding proteins enhance specific uptake rate by increasing the substrate–transporter encounter rate. *The FEBS Journal*, *282*(12), 2394–2407. https://doi.org/10.1111/febs.13289

Botos, I., Lountos, G. T., Wu, W., Cherry, S., Ghirlando, R., Kudzhaev, A. M., Rotanova, T. V., Val, N. de, Tropea, J. E., Gustchina, A., & Wlodawer, A. (2019). Cryo-EM structure of substrate-free E. coli Lon protease provides insights into the dynamics of Lon machinery. *Current Research in Structural Biology*, *1*, 13–20. https://doi.org/10.1016/j.crstbi.2019.10.001

Botos, I., Melnikov, E. E., Cherry, S., Kozlov, S., Makhovskaya, O. V., Tropea, J. E., Gustchina, A., Rotanova, T. V., & Wlodawer, A. (2005). Atomic-resolution Crystal Structure of the Proteolytic Domain of Archaeoglobus fulgidus Lon Reveals the Conformational Variability in the Active Sites of Lon Proteases. *Journal of Molecular Biology*, *351*(1), 144–157. https://doi.org/10.1016/j.jmb.2005.06.008

Bouchet, V., Hood, D. W., Li, J., Brisson, J. R., Randle, G. A., Martin, A., Li, Z., Goldstein, R., Schweda, E. K. H., Pelton, S. I., Richards, J. C., & Moxon, E. R. (2003). Host-derived sialic acid is incorporated into Haemophilus influenzae lipopolysaccharide and is a major virulence factor in experimental otitis media. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 8898–8903. https://doi.org/10.1073/pnas.1432026100

Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V., & Oost, J. van der. (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, *321*(5891), 960–964. https://doi.org/10.1126/science.1159689

Brown, S., Gauvin, C. C., Charbonneau, A. A., Burman, N., & Lawrence, C. M. (2020). Csx3 is a cyclic oligonucleotide phosphodiesterase associated with type III CRISPR–Cas that degrades the second messenger cA4. *Journal of Biological Chemistry*, *295*(44), 14963–14972. https://doi.org/10.1074/jbc.ra120.014099

Brüssow, H., & Hendrix, R. W. (2002). Phage Genomics Small Is Beautiful. *Cell*, *108*(1), 13–16. https://doi.org/10.1016/s0092-8674(01)00637-7

Burdette, D. L., Monroe, K. M., Sotelo-Troha, K., Iwig, J. S., Eckert, B., Hyodo, M., Hayakawa, Y., & Vance, R. E. (2011). STING is a direct innate immune sensor of cyclic di-GMP. *Nature*, *478*(7370), 515–518. https://doi.org/10.1038/nature10429

Burgess, R. R., Travers, A. A., Dunn, J. J., & Bautz, E. K. F. (1969). Factor Stimulating Transcription by RNA Polymerase. *Nature*, *221*(5175), 43–46. https://doi.org/10.1038/221043a0

Burgi, H. B., Dunitz, J. D., & Shefter, E. (1973). Geometrical reaction coordinates. II. Nucleophilic addition to a carbonyl group. *Journal of the American Chemical Society*, *95*(15), 5065–5067. https://doi.org/10.1021/ja00796a058

Campagne, S., Marsh, M. E., Capitani, G., Vorholt, J. A., & Allain, F. H.-T. (2014). Structural basis for −10 promoter element melting by environmentally induced sigma factors. *Nature Structural & Molecular Biology*, *21*(3), 269–276. https://doi.org/10.1038/nsmb.2777

Campbell, E. A., Greenwell, R., Anthony, J. R., Wang, S., Lim, L., Das, K., Sofia, H. J., Donohue, T. J., & Darst, S. A. (2007). A Conserved Structural Module Regulates Transcriptional

Responses to Diverse Stress Signals in Bacteria. *Molecular Cell*, *27*(5), 793–805. https://doi.org/10.1016/j.molcel.2007.07.009

Campbell, E. A., Tupy, J. L., Gruber, T. M., Wang, S., Sharp, M. M., Gross, C. A., & Darst, S. A. (2003). Crystal Structure of Escherichia coli σE with the Cytoplasmic Domain of Its Anti-σ RseA. *Molecular Cell*, *11*(4), 1067–1078. https://doi.org/10.1016/s1097-2765(03)00148-5

Carte, J., Wang, R., Li, H., Terns, R. M., & Terns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & Development*, *22*(24), 3489–3496. https://doi.org/10.1101/gad.1742908

Casadevall, A., & Pirofski, L. (1999). Host-Pathogen Interactions: Redefining the Basic Concepts of Virulence and Pathogenicity. *Infection and Immunity*, *67*(8), 3703–3713. https://doi.org/10.1128/iai.67.8.3703-3713.1999

Cellular gatekeepers. (2016). *Nature Structural & Molecular Biology*, *23*(6), 463–463. https://doi.org/10.1038/nsmb.3246

Cha, S., An, Y. J., Lee, C. R., Lee, H. S., Kim, Y., Kim, S. J., Kwon, K. K., Donatis, G. M. D., Lee, J., Maurizi, M. R., & Kang, S. G. (2010). Crystal structure of Lon protease: molecular architecture of gated entry to a sequestered degradation chamber. *The EMBO Journal*, *29*(20), 3520–3530. https://doi.org/10.1038/emboj.2010.226

Chames, P., & Baty, D. (2009). Bispecific antibodies for cancer therapy. *Current Opinion in Drug Discovery & Development*, *12*(2), 276–283.

Chanishvili, N. (2012). Chapter 1 Phage Therapy—History from Twort and d'Herelle Through Soviet Experience to Current Approaches. *Advances in Virus Research*, *83*, 3–40. https://doi.org/10.1016/b978-0-12-394438-2.00001-3

Charbonneau, A. A., Eckert, D. M., Gauvin, C. C., Lintner, N. G., & Lawrence, C. M. (2021). Cyclic Tetra-Adenylate (cA4) Recognition by Csa3; Implications for an Integrated Class 1 CRISPR-Cas Immune Response in Saccharolobus solfataricus. *Biomolecules*, *11*(12), 1852. https://doi.org/10.3390/biom11121852

Chaudhuri, A., Halder, K., & Datta, A. (2022). Classification of CRISPR/Cas system and its application in tomato breeding. *Theoretical and Applied Genetics*, *135*(2), 367–387. https://doi.org/10.1007/s00122-021-03984-y

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, *66*, 12–21. https://doi.org/10.1107/s0907444909042073

Chen, X., & Varki, A. (2010). Advances in the biology and chemistry of sialic acids. *ACS Chemical Biology*. https://doi.org/10.1021/cb900266r

Chien, N. C., Roberts, V. A., Giusti, A. M., Scharff, M. D., & Getzoff, E. D. (1989). Significant structural and functional change of an antigen-binding site by a distant amino acid substitution: proposal of a structural mechanism. *Proceedings of the National Academy of Sciences*, *86*(14), 5532–5536. https://doi.org/10.1073/pnas.86.14.5532

Chowdhury, N., Norris, J., McAlister, E., Lau, S. Y. K., Thomas, G. H., & Boyd, E. F. (2012). The VC1777–VC1779 proteins are members of a sialic acid-specific subfamily of TRAP transporters (SiaPQM) and constitute the sole route of sialic acid uptake in the human pathogen Vibrio cholerae. *Microbiology*, *158*(Pt_8), 2158–2167. https://doi.org/10.1099/mic.0.059659-0

Chung, I. Y. W., & Paetzel, M. (2013). Crystal Structures of Yellowtail Ascites Virus VP4 Protease TRAPPING AN INTERNAL CLEAVAGE SITE TRANS ACYL-ENZYME COMPLEX IN A NATIVE SER/LYS DYAD ACTIVE SITE*. *Journal of Biological Chemistry*, *288*(18), 13068–13081. https://doi.org/10.1074/jbc.m112.386953

Cohen, D., Melamed, S., Millman, A., Shulman, G., Oppenheimer-Shaanan, Y., Kacen, A., Doron, S., Amitai, G., & Sorek, R. (2019). Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature*, *574*, 691–695. https://doi.org/10.1038/s41586-019-1605-5

Coker, J. K., Moyne, O., Rodionov, D. A., & Zengler, K. (2021). Carbohydrates great and small, from dietary fiber to sialic acids: How glycans influence the gut microbiome and affect human health. *Gut Microbes*, *13*(1), 1869502. https://doi.org/10.1080/19490976.2020.1869502

Conrath, K. E., Lauwereys, M., Wyns, L., & Muyldermans, S. (2001). Camel Single-domain Antibodies as Modular Building Units in Bispecific and Bivalent Antibody Constructs*. *Journal of Biological Chemistry*, *276*(10), 7346–7350. https://doi.org/10.1074/jbc.m007734200

Corfield, T. (1992). Bacterial sialidases—roles in pathogenicity and nutrition. *Glycobiology*, *2*(6), 509–521. https://doi.org/10.1093/glycob/2.6.509

Cortez-Retamozo, V., Backmann, N., Senter, P. D., Wernery, U., Baetselier, P. D., Muyldermans, S., & Revets, H. (2004). Efficient Cancer Therapy with a Nanobody-Based Conjugate. *Cancer Research*, *64*(8), 2853–2857. https://doi.org/10.1158/0008-5472.can-03-3935

Croll, T. I. (2018). ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D*, *74*(6), 519–530. https://doi.org/10.1107/s2059798318002425

Crooks, G. E., Hon, G., Chandonia, J.-M., & Brenner, S. E. (2004). WebLogo: A Sequence Logo Generator. *Genome Research*, *14*(6), 1188–1190. https://doi.org/10.1101/gr.849004

Darby, J. F., Hopkins, A. P., Shimizu, S., Roberts, S. M., Brannigan, J. A., Turkenburg, J. P., Thomas, G. H., Hubbard, R. E., & Fischer, M. (2019). Water Networks Can Determine the Affinity of Ligand Binding to Proteins. *Journal of the American Chemical Society*, *141*, 15818–15826. https://doi.org/10.1021/jacs.9b06275

Davidson, A. L., Dassa, E., Orelle, C., & Chen, J. (2008). Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems. *Microbiology and Molecular Biology Reviews*, *72*(2), 317–364. https://doi.org/10.1128/mmbr.00031-07

Davies, J., & Riechmann, L. (1994). 'Camelising' human antibody fragments: NMR studies on VH domains. *FEBS Letters*, *339*(3), 285–290. https://doi.org/10.1016/0014-5793(94)80432-x

Davies, J. S., Currie, M. J., North, R. A., Scalise, M., Wright, J. D., Copping, J. M., Remus, D. M., Gulati, A., Morado, D. R., Jamieson, S. A., Newton-Vesty, M. C., Abeysekera, G. S., Ramaswamy, S., Friemann, R., Wakatsuki, S., Allison, J. R., Indiveri, C., Drew, D., Mace, P. D., & Dobson, R. C. J. (2023). Structure and mechanism of a tripartite ATP-independent periplasmic TRAP transporter. *Nature Communications*, *14*(1), 1120. https://doi.org/10.1038/s41467-023-36590-1

Deng, L., Chen, X., & Varki, A. (2013). Exploration of sialic acid diversity and biology using sialoglycan microarrays. *Biopolymers*, *99*, 650–665. https://doi.org/10.1002/bip.22314

Desmyter, A., Spinelli, S., Payan, F., Lauwereys, M., Wyns, L., Muyldermans, S., & Cambillau, C. (2002). Three Camelid VHH Domains in Complex with Porcine Pancreatic α-Amylase INHIBITION AND VERSATILITY OF BINDING TOPOLOGY*. *Journal of Biological Chemistry*, *277*(26), 23645–23650. https://doi.org/10.1074/jbc.m202327200

Desmyter, A., Transue, T. R., Ghahroudi, M. A., Thi, M.-H. D., Poortmans, F., Hamers, R., Muyldermans, S., & Wyns, L. (1996). Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nature Structural Biology*, *3*(9), 803–811. https://doi.org/10.1038/nsb0996-803

Diard, M., & Hardt, W. D. (2017). Evolution of bacterial virulence. *FEMS Microbiology Reviews*, *41*, 679–697. https://doi.org/10.1093/femsre/fux023

Dingus, J. G., Tang, J. C., Amamoto, R., Wallick, G. K., & Cepko, C. L. (2022). A general approach for stabilizing nanobodies for intracellular expression. *ELife*, *11*, e68253. https://doi.org/10.7554/elife.68253

Dodson, G., & Wlodawer, A. (1998). Catalytic triads and their relatives. *Trends in Biochemical Sciences*, *23*(9), 347–352. https://doi.org/10.1016/s0968-0004(98)01254-7

Dolk, E., Vaart, M. van der, Hulsik, D. L., Vriend, G., Haard, H. de, Spinelli, S., Cambillau, C., Frenken, L., & Verrips, T. (2005). Isolation of Llama Antibody Fragments for Prevention of

Dandruff by Phage Display in Shampoo. *Applied and Environmental Microbiology*, *71*(1), 442–450. https://doi.org/10.1128/aem.71.1.442-450.2005

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., & Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. *Science*, *359*(6379). https://doi.org/10.1126/science.aar4120

Duckworth, D. H. (1976). "Who discovered bacteriophage?" *Bacteriological Reviews*, *40*(4), 793–802. https://doi.org/10.1128/br.40.4.793-802.1976

Dudev, T., & Lim, C. (2002). Factors Governing the Protonation State of Cysteines in Proteins: An Ab Initio/CDM Study. *Journal of the American Chemical Society*, *124*(23), 6759–6766. https://doi.org/10.1021/ja012620l

Dumoulin, M., Conrath, K., Meirhaeghe, A. V., Meersman, F., Heremans, K., Frenken, L. G. J., Muyldermans, S., Wyns, L., & Matagne, A. (2002). Single-domain antibody fragments with high conformational stability. *Protein Science*, *11*(3), 500–515. https://doi.org/10.1110/ps.34602

Dy, R. L., Richter, C., Salmond, G. P. C., & Fineran, P. C. (2014). Remarkable Mechanisms in Microbes to Resist Phage Infections. *Annual Review of Virology*, *1*(1), 1–25. https://doi.org/10.1146/annurev-virology-031413-085500

Emmerson, C. D., Vlist, E. J. van der, Braam, M. R., Vanlandschoot, P., Merchiers, P., Haard, H. J. W. de, Verrips, C. T., Henegouwen, P. M. P. van B. en, & Dolk, E. (2011). Enhancement of Polymeric Immunoglobulin Receptor Transcytosis by Biparatopic VHH. *PLoS ONE*, *6*(10), e26299. https://doi.org/10.1371/journal.pone.0026299

Emsley, P., & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, *60*, 2126–2132. https://doi.org/10.1107/s0907444904019158

Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W., & Jhoti, H. (2016). Twenty years on: the impact of fragments on drug discovery. *Nature Reviews Drug Discovery*, *15*(9), 605–619. https://doi.org/10.1038/nrd.2016.109

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography*, *62*, 72–82. https://doi.org/10.1107/s0907444905036693

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography*, *69*, 1204–1214. https://doi.org/10.1107/s0907444913000061

Fischer, M., Hopkins, A. P., Severi, E., Hawkhead, J., Bawdon, D., Watts, A. G., Hubbard, R. E., & Thomas, G. H. (2015). Tripartite ATP-independent Periplasmic (TRAP) Transporters Use an Arginine-mediated Selectivity Filter for High Affinity Substrate Binding*. *Journal of Biological Chemistry*, *290*(45), 27113–27123. https://doi.org/10.1074/jbc.m115.656603

Fischer, M., Zhang, Q. Y., Hubbard, R. E., & Thomas, G. H. (2010). Caught in a TRAP: substrate-binding proteins in secondary transport. *Trends in Microbiology*, *18*(10), 471–478. https://doi.org/10.1016/j.tim.2010.06.009

Fleming, P. J., & Fleming, K. G. (2018). HullRad: Fast Calculations of Folded and Disordered Protein and Nucleic Acid Hydrodynamic Properties. *Biophysical Journal*, *114*, 856–869. https://doi.org/10.1016/j.bpj.2018.01.002

Flynn, J. M., Levchenko, I., Sauer, R. T., & Baker, T. A. (2004). Modulating substrate choice: the SspB adaptor delivers a regulator of the extracytoplasmic-stress response to the AAA+ protease ClpXP for degradation. *Genes & Development*, *18*(18), 2292–2301. https://doi.org/10.1101/gad.1240104

Fortier, L.-C., & Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*, *4*(5), 354–365. https://doi.org/10.4161/viru.24498

Forward, J. A., Behrendt, M. C., Wyborn, N. R., Cross, R., & Kelly, D. J. (1997). TRAP transporters: a new family of periplasmic solute transport systems encoded by the dctPQM genes

of Rhodobacter capsulatus and by homologs in diverse gram-negative bacteria. *Journal of Bacteriology*, *179*(17), 5482–5493. https://doi.org/10.1128/jb.179.17.5482-5493.1997

Fox, K. L., Cox, A. D., Gilbert, M., Wakarchuk, W. W., Li, J., Makepeace, K., Richards, J. C., Moxon, E. R., & Hood, D. W. (2006). Identification of a Bifunctional Lipopolysaccharide Sialyltransferase in Haemophilus influenzae INCORPORATION OF DISIALIC ACID*. *Journal of Biological Chemistry*, *281*(52), 40024–40032. https://doi.org/10.1074/jbc.m602314200

Frank, S. A. (2002). *Immunology and Evolution of Infectious Disease*. https://doi.org/10.1354/books/unregistered/9780691220161

Garcia-Doval, C., Schwede, F., Berk, C., Rostøl, J. T., Niewoehner, O., Tejero, O., Hall, J., Marraffini, L. A., & Jinek, M. (2020). Activation and self-inactivation mechanisms of the cyclic oligoadenylate-dependent CRISPR ribonuclease Csm6. *Nature Communications*, *11*. https://doi.org/10.1038/s41467-020-15334-5

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). *The Proteomics Protocols Handbook*. 571–607. https://doi.org/10.1385/1-59259-890-0:571

Gelkop, S., Sobarzo, A., Brangel, P., Vincke, C., Romão, E., Fedida-Metula, S., Strom, N., Ataliba, I., Mwiine, F. N., Ochwo, S., Velazquez-Salinas, L., McKendry, R. A., Muyldermans, S., Lutwama, J. J., Rieder, E., Yavelsky, V., & Lobel, L. (2018). The Development and Validation of a Novel Nanobody-Based Competitive ELISA for the Detection of Foot and Mouth Disease 3ABC Antibodies in Cattle. *Frontiers in Veterinary Science*, *5*, 250. https://doi.org/10.3389/fvets.2018.00250

Georjon, H., & Bernheim, A. (2023). The highly diverse antiphage defence systems of bacteria. *Nature Reviews Microbiology*, *21*(10), 686–700. https://doi.org/10.1038/s41579-023-00934-x

Gil, R., & Latorre, A. (2012). Factors Behind Junk DNA in Bacteria. *Genes*, *3*(4), 634–650. https://doi.org/10.3390/genes3040634

Glaenzer, J., Peter, M. F., Thomas, G. H., & Hagelueken, G. (2017). PELDOR Spectroscopy Reveals Two Defined States of a Sialic Acid TRAP Transporter SBP in Solution. *Biophysical Journal*, *112*, 109–120. https://doi.org/10.1016/j.bpj.2016.12.010

Gleditzsch, D., Pausch, P., Müller-Esparza, H., Özcan, A., Guo, X., Bange, G., & Randau, L. (2019). PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biology*, *16*(4), 504–517. https://doi.org/10.1080/15476286.2018.1504546

Glockshuber, R., Malia, M., Pfitzinger, I., & Plueckthun, A. (1990). A comparison of strategies to stabilize immunoglobulin Fv-fragments. *Biochemistry*, *29*(6), 1362–1367. https://doi.org/10.1021/bi00458a002

Goldberg, A. L., Moerschell, R. P., Hachung, C., & Maurizi, M. R. (1994). ATP-dependent protease La (Lon) from Escherichia coli. *Methods in Enzymology*, *244*, 350–375. https://doi.org/10.1016/0076-6879(94)44027-1

Goldberg, G. W., Jiang, W., Bikard, D., & Marraffini, L. A. (2014). Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature*, *514*(7524), 633–637. https://doi.org/10.1038/nature13637

Gonzalez-Sapienza, G., Rossotti, M. A., & Rosa, S. T. (2017). Single-Domain Antibodies As Versatile Affinity Reagents for Analytical and Diagnostic Applications. *Frontiers in Immunology*, *8*, 977. https://doi.org/10.3389/fimmu.2017.00977

Greenberg, A. S., Avila, D., Hughes, M., Hughes, A., McKinney, E. C., & Flajnik, M. F. (1995). A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature*, *374*(6518), 168–173. https://doi.org/10.1038/374168a0

Gross, R., Vavre, F., Heddi, A., Hurst, G. D. D., Zchori-Fein, E., & Bourtzis, K. (2009). Immunity and symbiosis. *Molecular Microbiology*, *73*(5), 751–759. https://doi.org/10.1111/j.1365-2958.2009.06820.x

Gruber, T. M., & Gross, C. A. (2003). MULTIPLE SIGMA SUBUNITS AND THE PARTITIONING OF BACTERIAL TRANSCRIPTION SPACE. *Annual Review of Microbiology*, *57*(1), 441–466. https://doi.org/10.1146/annurev.micro.57.030502.090913

Gur, E., & Sauer, R. T. (2008). Recognition of misfolded proteins by Lon, a AAA+ protease. *Genes & Development*, *22*(16), 2267–2277. https://doi.org/10.1101/gad.1670908

Hagar, J. A., Powell, D. A., Aachoui, Y., Ernst, R. K., & Miao, E. A. (2013). Cytoplasmic LPS Activates Caspase-11: Implications in TLR4-Independent Endotoxic Shock. *Science*, *341*(6151), 1250–1253. https://doi.org/10.1126/science.1240988

Hagelueken, G., Ward, R., Naismith, J. H., & Schiemann, O. (2012). MtsslWizard: In Silico Spin-Labeling and Generation of Distance Distributions in PyMOL. *Applied Magnetic Resonance*, *42*, 377–391. https://doi.org/10.1007/s00723-012-0314-0

Hale, C. R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A. M., Glover, C. V. C., Graveley, B. R., Terns, R. M., & Terns, M. P. (2012). Essential Features and Rational Design of CRISPR RNAs that Function with the Cas RAMP Module Complex to Cleave RNAs. *Molecular Cell*, *45*(3), 292–302. https://doi.org/10.1016/j.molcel.2011.10.023

Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., & Terns, M. P. (2009). RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell*, *139*(5), 945–956. https://doi.org/10.1016/j.cell.2009.07.040

Hallgren, J., Tsirigos, K. D., Pedersen, M. D., Armenteros, J. J. A., Marcatili, P., Nielsen, H., Krogh, A., & Winther, O. (2022). *DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks*. https://doi.org/10.1101/2022.04.08.487609

Hallström, T., & Riesbeck, K. (2010). Haemophilus influenzae and the complement system. *Trends in Microbiology*, *18*(6), 258–265. https://doi.org/10.1016/j.tim.2010.03.007

Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hammers, C., Songa, E. B., Bendahman, N., & Hammers, R. (1993). Naturally occurring antibodies devoid of light chains. *Nature*, *363*, 446–448. https://doi.org/10.1038/363446a0

Harmsen, M. M., & Haard, H. J. D. (2007). Properties, production, and applications of camelid single-domain antibody fragments. *Applied Microbiology and Biotechnology*, *77*, 13–22. https://doi.org/10.1007/s00253-007-1142-2

Haugen, S. P., Berkmen, M. B., Ross, W., Gaal, T., Ward, C., & Gourse, R. L. (2006). rRNA Promoter Regulation by Nonoptimal Binding of σ Region 1.2: An Additional Recognition Element for RNA Polymerase. *Cell*, *125*(6), 1069–1082. https://doi.org/10.1016/j.cell.2006.04.034

He, L., Luo, D., Yang, F., Li, C., Zhang, X., Deng, H., & Zhang, J.-R. (2018). Multiple domains of bacterial and human Lon proteases define substrate selectivity. *Emerging Microbes & Infections*, *7*(1), 149. https://doi.org/10.1038/s41426-018-0148-4

Hecker, M., Pané-Farré, J., & Völker, U. (2007). SigB-Dependent General Stress Response in Bacillus subtilis and Related Gram-Positive Bacteria. *Annual Review of Microbiology*, *61*(1), 215–236. https://doi.org/10.1146/annurev.micro.61.080706.093445

Hediger, M. A., Romero, M. F., Peng, J.-B., Rolfs, A., Takanaga, H., & Bruford, E. A. (2004). The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins. *Pflugers Arch-Eur J Physiol*, *447*, 465–468. https://doi.org/10.1007/s00424-003-1192-y

Higgins, C. F., Hyde, S. C., Mimmack, M. M., Gileadi, U., Gill, D. R., & Gallagher, M. P. (1990). Binding protein-dependent transport systems. *Journal of Bioenergetics and Biomembranes*, *22*(4), 571–592. https://doi.org/10.1007/bf00762962

Hochheiser, I. V., Behrmann, H., Hagelueken, G., Rodríguez-Alcázar, J. F., Kopp, A., Latz, E., Behrmann, E., & Geyer, M. (2022). Directionality of PYD filament growth determined by the transition of NLRP3 nucleation seeds to ASC elongation. *Science Advances*, *8*(19), eabn7583. https://doi.org/10.1126/sciadv.abn7583

Hodges, R. S., Heaton, R. J., Parker, J. M., Molday, L., & Molday, R. S. (1988). Antigen-antibody interaction. Synthetic peptides define linear antigenic determinants recognized by monoclonal antibodies directed to the cytoplasmic carboxyl terminus of rhodopsin. *The Journal of Biological Chemistry*, *263*(24), 11768–11775.

Hogrel, G., Guild, A., Graham, S., Rickman, H., Grüschow, S., Bertrand, Q., Spagnolo, L., & White, M. F. (2022). Cyclic nucleotide-induced helical structure activates a TIR immune effector. *Nature*, *608*(7924), 808–812. https://doi.org/10.1038/s41586-022-05070-9

Holliger, P., & Hudson, P. J. (2005). Engineered antibody fragments and the rise of single domains. *Nature Biotechnology*, *23*, 1126–1136. https://doi.org/10.1038/nbt1142

Holm, L., Laiho, A., Törönen, P., & Salgado, M. (2023). DALI shines a light on remote homologs: One hundred discoveries. *Protein Science*, *32*(1), e4519. https://doi.org/10.1002/pro.4519

Hood, D. W., Cox, A. D., Gilbert, M., Makepeace, K., Walsh, S., Deadman, M. E., Cody, A., Martin, A., Månsson, M., Schweda, E. K. H., Brisson, J., Richards, J. C., Moxon, E. R., & Wakarchuk, W. W. (2001). Identification of a lipopolysaccharide α-2,3-sialyltransferase from Haemophilus influenzae. *Molecular Microbiology*, *39*(2), 341–351. https://doi.org/10.1046/j.1365-2958.2001.02204.x

Hood, D. W., Makepeace, K., Deadman, M. E., Rest, R. F., Thibault, P., Martin, A., Richards, J. C., & Moxon, E. R. (1999). Sialic acid in the lipopolysaccharide of Haemophilus influenzae: Strain distribution, influence on serum resistance and structural characterization. *Molecular Microbiology*, *33*, 679–692. https://doi.org/10.1046/j.1365-2958.1999.01509.x

Hood, D. W., Randle, G., Cox, A. D., Makepeace, K., Li, J., Schweda, E. K. H., Richards, J. C., & Moxon, E. R. (2004). Biosynthesis of Cryptic Lipopolysaccharide Glycoforms in Haemophilus influenzae Involves a Mechanism Similar to That Required for O-Antigen Synthesis. *Journal of Bacteriology*, *186*(21), 7429–7439. https://doi.org/10.1128/jb.186.21.7429-7439.2004

Hopfner, K. P., & Hornung, V. (2020). Molecular mechanisms and cellular functions of cGAS–STING signalling. *Nature Reviews Molecular Cell Biology*, *21*, 501–521. https://doi.org/10.1038/s41580-020-0244-x

Hu, C., Beljouw, S. P. B. van, Nam, K. H., Schuler, G., Ding, F., Cui, Y., Rodríguez-Molina, A., Haagsma, A. C., Valk, M., Pabst, M., Brouns, S. J. J., & Ke, A. (2022). Craspase is a CRISPR RNA-guided, RNA-activated protease. *Science*, *377*(6612), 1278–1285. https://doi.org/10.1126/science.add5064

Huang, L., Reekmans, G., Saerens, D., Friedt, J.-M., Frederix, F., Francis, L., Muyldermans, S., Campitelli, A., & Hoof, C. V. (2005). Prostate-specific antigen immunosensing based on mixed self-assembled monolayers, camel antibodies and colloidal gold enhanced sandwich assays. *Biosensors and Bioelectronics*, *21*(3), 483–490. https://doi.org/10.1016/j.bios.2004.11.016

Hughes, K. T., & Mathee, K. (1998). THE ANTI-SIGMA FACTORS. *Microbiology*, *52*(1), 231–286. https://doi.org/10.1146/annurev.micro.52.1.231

Hussack, G., Hirama, T., Ding, W., MacKenzie, R., & Tanha, J. (2011). Engineered Single-Domain Antibodies with High Protease Resistance and Thermal Stability. *PLoS ONE*, *6*(11), e28218. https://doi.org/10.1371/journal.pone.0028218

Hussein, H. A., Borrel, A., Geneix, C., Petitjean, M., Regad, L., & Camproux, A.-C. (2015). PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. *Nucleic Acids Research*, *43*(W1), W436–W442. https://doi.org/10.1093/nar/gkv462

Ihssen, J., & Egli, T. (2004). Specific growth rate and not cell density controls the general stress response in Escherichia coli. *Microbiology*, *150*(6), 1637–1648. https://doi.org/10.1099/mic.0.26849-0

Ishikawa, H., & Barber, G. N. (2008). STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature*, *455*(7213), 674–678. https://doi.org/10.1038/nature07317

Janeway Jr., C. A., & Medzhitov, R. (2002). INNATE IMMUNE RECOGNITION. *Immunology*, *20*(1), 197–216. https://doi.org/10.1146/annurev.immunol.20.083001.084359

Janin, J., & Rodier, F. (1995). Protein–protein interaction at crystal contacts. *Proteins: Structure, Function, and Bioinformatics*, *23*(4), 580–587. https://doi.org/10.1002/prot.340230413

Jansen, Ruud., Embden, Jan. D. A. van, Gaastra, Wim., & Schouls, Leo. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, *43*(6), 1565–1575. https://doi.org/10.1046/j.1365-2958.2002.02839.x

Jermyn, W. S., & Boyd, E. F. (2002). Characterization of a novel Vibrio pathogenicity island (VPI-2) encoding neuraminidase (nanH) among toxigenic Vibrio cholerae isolates. *Microbiology*, *148*(11), 3681–3693. https://doi.org/10.1099/00221287-148-11-3681

Jia, N., Jones, R., Yang, G., Ouerfelli, O., & Patel, D. J. (2019). CRISPR-Cas III-A Csm6 CARF Domain Is a Ring Nuclease Triggering Stepwise cA4 Cleavage with ApA>p Formation Terminating RNase Activity. *Molecular Cell*, *75*(5), 944-956.e6. https://doi.org/10.1016/j.molcel.2019.06.014

Jia, N., & Patel, D. J. (2021). Structure-based functional mechanisms and biotechnology applications of anti-CRISPR proteins. *Nature Reviews Molecular Cell Biology*, *22*(8), 563–579. https://doi.org/10.1038/s41580-021-00371-9

Jiang, F., & Doudna, J. A. (2015). CRISPR–Cas9 Structures and Mechanisms. *Annual Review of Biophysics*, *46*(1), 1–25. https://doi.org/10.1146/annurev-biophys-062215-010822

Jiang, K., Faltova, L., Hua, S., Capitani, G., Prota, A. E., Landgraf, C., Volkmer, R., Kammerer, R. A., Steinmetz, M. O., & Akhmanova, A. (2018). Structural Basis of Formation of the Microtubule Minus-End-Regulating CAMSAP-Katanin Complex. *Structure*, *26*(3), 375-382.e4. https://doi.org/10.1016/j.str.2017.12.017

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Johnston, J. W., Coussens, N. P., Allen, S., Houtman, J. C. D., Turner, K. H., Zaleski, A., Ramaswamy, S., Gibson, B. W., & Apicella, M. A. (2008). Characterization of the N-Acetyl-5-neuraminic Acid-binding Site of the Extracytoplasmic Solute Receptor (SiaP) of Nontypeable Haemophilus influenzae Strain 2019*. *Journal of Biological Chemistry*, *283*(2), 855–865. https://doi.org/10.1074/jbc.m706603200

Jore, M. M., Lundgren, M., Duijn, E. van, Bultema, J. B., Westra, E. R., Waghmare, S. P., Wiedenheft, B., Pul, Ü., Wurm, R., Wagner, R., Beijer, M. R., Barendregt, A., Zhou, K., Snijders, A. P. L., Dickman, M. J., Doudna, J. A., Boekema, E. J., Heck, A. J. R., Oost, J. van der, & Brouns, S. J. J. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Structural & Molecular Biology*, *18*(5), 529–536. https://doi.org/10.1038/nsmb.2019

Jovčevska, I., & Muyldermans, S. (2020). The Therapeutic Potential of Nanobodies. *BioDrugs*, *34*(1), 11–26. https://doi.org/10.1007/s40259-019-00392-z

Juge, N. (2012). Microbial adhesins to gastrointestinal mucus. *Trends in Microbiology*, *20*(1), 30–39. https://doi.org/10.1016/j.tim.2011.10.001

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., … Baker, N. A. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Science*, *27*(1), 112–128. https://doi.org/10.1002/pro.3280

Kabsch, W. (2010). *XDS*. *Acta Crystallographica Section D Biological Crystallography*, *66*, 125–132. https://doi.org/10.1107/s0907444909047337

Kang-Pettinger, T., Walker, K., Brown, R., Cowan, R., Wright, H., Baravalle, R., Waters, L. C., Muskett, F. W., Bowler, M. W., Sawmynaden, K., Coombs, P. J., Carr, M. D., & Hall, G. (2023). Identification, binding, and structural characterization of single domain anti-PD-L1 antibodies inhibitory of immune regulatory proteins PD-1 and CD80. *Journal of Biological Chemistry*, *299*, 102769. https://doi.org/10.1016/j.jbc.2022.102769

Kapsalis, C., Wang, B., Mkami, H. E., Pitt, S. J., Schnell, J. R., Smith, T. K., Lippiat, J. D., Bode, B. E., & Pliotas, C. (2019). Allosteric activation of an ion channel triggered by modification of mechanosensitive nano-pockets. *Nature Communications*, *10*(1), 4619. https://doi.org/10.1038/s41467-019-12591-x

Karpowich, N. K., Huang, H. H., Smith, P. C., & Hunt, J. F. (2003). Crystal Structures of the BtuF Periplasmic-binding Protein for Vitamin B12 Suggest a Functionally Important Reduction in Protein Mobility upon Ligand Binding*. *Journal of Biological Chemistry*, *278*(10), 8429–8434. https://doi.org/10.1074/jbc.m212239200

Katritch, V., & Abagyan, R. (2011). GPCR agonist binding revealed by modeling and crystallography. *Trends in Pharmacological Sciences*, *32*(11), 637–643. https://doi.org/10.1016/j.tips.2011.08.001

Kayagaki, N., Stowe, I. B., Lee, B. L., O'Rourke, K., Anderson, K., Warming, S., Cuellar, T., Haley, B., Roose-Girma, M., Phung, Q. T., Liu, P. S., Lill, J. R., Li, H., Wu, J., Kummerfeld, S., Zhang, J., Lee, W. P., Snipas, S. J., Salvesen, G. S., … Dixit, V. M. (2015). Caspase-11 cleaves gasdermin D for non-canonical inflammasome signalling. *Nature*, *526*(7575), 666–671. https://doi.org/10.1038/nature15541

Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G., & Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science*, *357*, 605–609. https://doi.org/10.1126/science.aao0100

Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, Č., & Siksnys, V. (2016). Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Molecular Cell*, *62*(2), 295–306. https://doi.org/10.1016/j.molcel.2016.03.024

Keitel, T., Kramer, A., Wessner, H., Scholz, C., Schneider-Mergener, J., & Hohne, W. (1997). Crystallographic Analysis of Anti-p24 (HIV-1) Monoclonal Antibody Cross-Reactivity and Polyspecificity. *Cell*, *91*(6), 811–820. https://doi.org/10.1016/s0092-8674(00)80469-9

Kelly, D. (2001). The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. *FEMS Microbiology Reviews*, *25*(4), 405–424. https://doi.org/10.1016/s0168-6445(01)00061-4

Kempen, M. van, Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L. M., Söding, J., & Steinegger, M. (2024). Fast and accurate protein structure search with Foldseek. *Nature Biotechnology*, *42*(2), 243–246. https://doi.org/10.1038/s41587-023-01773-0

Kirthika, P., Lloren, K. K. S., Jawalagatti, V., & Lee, J. H. (2023). Structure, Substrate Specificity and Role of Lon Protease in Bacterial Pathogenesis and Survival. *International Journal of Molecular Sciences*, *24*(4), 3422. https://doi.org/10.3390/ijms24043422

Koenig, P. A., Das, H., Liu, H., Kümmerer, B. M., Gohr, F. N., Jenster, L. M., Schiffelers, L. D. J., Tesfamariam, Y. M., Uchima, M., Wuerth, J. D., Gatterdam, K., Ruetalo, N., Christensen, M. H., Fandrey, C. I., Normann, S., Tödtmann, J. M. P., Pritzl, S., Hanke, L., Boos, J., … Schmidt, F. I. (2021). Structure-guided multivalent nanobodies block SARS-CoV-2 infection and suppress mutational escape. *Science*, *371*. https://doi.org/10.1126/science.abe6230

Kontermann, R. E. (2009). Strategies to Extend Plasma Half-Lives of Recombinant Antibodies. *BioDrugs*, *23*(2), 93–109. https://doi.org/10.2165/00063030-200923020-00003

Kranzusch, P. J., Wilson, S. C., Lee, A. S. Y., Berger, J. M., Doudna, J. A., & Vance, R. E. (2015). Ancient Origin of cGAS-STING Reveals Mechanism of Universal 2′,3′ cGAMP Signaling. *Molecular Cell*, *59*(6), 891–903. https://doi.org/10.1016/j.molcel.2015.07.022

Krissinel, E., & Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*, *372*(3), 774–797. https://doi.org/10.1016/j.jmb.2007.05.022

Labrie, S. J., Samson, J. E., & Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nature Reviews Microbiology*, *8*(5), 317–327. https://doi.org/10.1038/nrmicro2315

Lamont, J. T. (1992). Mucus: The Front Line of Intestinal Mucosal Defense. *Annals of the New York Academy of Sciences*, *664*(1), 190–201. https://doi.org/10.1111/j.1749-6632.1992.tb39760.x

Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. *Journal of Chemical Information and Modeling*, *51*, 2778–2786. https://doi.org/10.1021/ci200227u

Lau, R. K., Ye, Q., Birkholz, E. A., Berg, K. R., Patel, L., Mathews, I. T., Watrous, J. D., Ego, K., Whiteley, A. T., Lowey, B., Mekalanos, J. J., Kranzusch, P. J., Jain, M., Pogliano, J., & Corbett, K. D. (2020). Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. *Molecular Cell*, *77*(4), 723-733.e6. https://doi.org/10.1016/j.molcel.2019.12.010

Lavoie, T. B., Mohan, S., Lipschultz, C. A., Grivel, J.-C., Li, Y., Mainhart, C. R., Kam-Morgan, L. N. W., Drohan, W. N., & Smith-Gill, S. J. (1999). Structural differences among monoclonal antibodies with distinct fine specificities and kinetic properties. *Molecular Immunology*, *36*(17), 1189–1205. https://doi.org/10.1016/s0161-5890(99)00130-3

Leenay, R. T., Maksimchuk, K. R., Slotkowski, R. A., Agrawal, R. N., Gomaa, A. A., Briner, A. E., Barrangou, R., & Beisel, C. L. (2016). Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Molecular Cell*, *62*(1), 137–147. https://doi.org/10.1016/j.molcel.2016.02.031

Lehninger, A. L. (1979). *Biochemie* (2nd ed.). Verlag Chemie GmbH.

Leslie, D. J., Heinen, C., Schramm, F. D., Thüring, M., Aakre, C. D., Murray, S. M., Laub, M. T., & Jonas, K. (2015). Nutritional Control of DNA Replication Initiation through the Proteolysis and Regulated Translation of DnaA. *PLoS Genetics*, *11*(7), e1005342. https://doi.org/10.1371/journal.pgen.1005342

Levin, M. C., Lee, S. M., Kalume, F., Morcos, Y., Dohan, F. C., Hasty, K. A., Callaway, J. C., Zunt, J., Desiderio, D. M., & Stuart, J. M. (2002). Autoimmunity due to molecular mimicry as a cause of neurological disease. *Nature Medicine*, *8*(5), 509–513. https://doi.org/10.1038/nm0502-509

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., … Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology*, *75*, 861–877. https://doi.org/10.1107/s2059798319011471

Liu, H., & May, K. (2012). Disulfide bond structures of IgG molecules. *MAbs*, *4*(1), 17–23. https://doi.org/10.4161/mabs.4.1.18347

Liu, H., & Naismith, J. H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology*, *8*, 91. https://doi.org/10.1186/1472-6750-8-91

Liu, N., Pang, X., Zhang, H., & Ji, P. (2022). The cGAS-STING Pathway in Bacterial Infection and Bacterial Immunity. *Frontiers in Immunology*, *12*. https://doi.org/10.3389/fimmu.2021.814709

Lonetto, M. A., Brown, K. L., Rudd, K. E., & Buttner, M. J. (1994). Analysis of the Streptomyces coelicolor sigE gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. *Proceedings of the National Academy of Sciences*, *91*(16), 7573–7577. https://doi.org/10.1073/pnas.91.16.7573

Losick, R., & Pero, J. (1981). Cascades of sigma factors. *Cell*, *25*(3), 582–584. https://doi.org/10.1016/0092-8674(81)90164-1

Lottspeich, F., & Engels, J. W. (2012). *Bioanalytik* (3rd ed.). Spektrum Akademischer Verlag.

Lowey, B., Whiteley, A. T., Keszei, A. F. A., Morehouse, B. R., Mathews, I. T., Antine, S. P., Cabrera, V. J., Kashin, D., Niemann, P., Jain, M., Schwede, F., Mekalanos, J. J., Shao, S., Lee,

A. S. Y., & Kranzusch, P. J. (2020). CBASS Immunity Uses CARF-Related Effectors to Sense 3′–5′- and 2′–5′-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection. *Cell*, *182*(1), 38-49.e17. https://doi.org/10.1016/j.cell.2020.05.019

Luepke, K. H., Suda, K. J., Boucher, H., Russo, R. L., Bonney, M. W., Hunt, T. D., & Mohr, J. F. (2017). Past, Present, and Future of Antibacterial Economics: Increasing Bacterial Resistance, Limited Antibiotic Pipeline, and Societal Implications. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, *37*(1), 71–84. https://doi.org/10.1002/phar.1868

Maillard, A. P., Girard, E., Ziani, W., Petit-Härtlein, I., Kahn, R., & Covès, J. (2014). The Crystal Structure of the Anti-σ Factor CnrY in Complex with the σ Factor CnrH Shows a New Structural Class of Anti-σ Factors Targeting Extracytoplasmic Function σ Factors. *Journal of Molecular Biology*, *426*(12), 2313–2327. https://doi.org/10.1016/j.jmb.2014.04.003

Majumdar, S., Zhao, P., Pfister, N. T., Compton, M., Olson, S., Glover, C. V. C., Wells, L., Graveley, B. R., Terns, R. M., & Terns, M. P. (2015). Three CRISPR-Cas immune effector complexes coexist in Pyrococcus furiosus. *RNA*, *21*(6), 1147–1158. https://doi.org/10.1261/rna.049130.114

Makarova, K. S., Anantharaman, V., Grishin, N. V., Koonin, E. V., & Aravind, L. (2014). CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Frontiers in Genetics*, *5*, 102. https://doi.org/10.3389/fgene.2014.00102

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J. M., Wolf, Y. I., Yakunin, A. F., Oost, J. van der, & Koonin, E. V. (2011). Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology*, *9*(6), 467–477. https://doi.org/10.1038/nrmicro2577

Makarova, K. S., & Koonin, E. V. (2015). CRISPR, Methods and Protocols. *Methods in Molecular Biology*, *1311*, 47–75. https://doi.org/10.1007/978-1-4939-2687-9_4

Makarova, K. S., Timinskas, A., Wolf, Y. I., Gussow, A. B., Siksnys, V., Venclovas, Č., & Koonin, E. V. (2020). Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. *Nucleic Acids Research*, *48*, 8828–8847. https://doi.org/10.1093/nar/gkaa635

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J. J., Charpentier, E., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Terns, R. M., Terns, M. P., White, M. F., Yakunin, A. F., Garrett, R. A., Oost, J. van der, … Koonin, E. V. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, *13*(11), 722–736. https://doi.org/10.1038/nrmicro3569

Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2018). Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? *The CRISPR Journal*, *1*, 325–336. https://doi.org/10.1089/crispr.2018.0033

Makarova, K. S., Wolf, Y. I., Snir, S., & Koonin, E. V. (2011). Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. *Journal of Bacteriology*, *193*(21), 6039–6056. https://doi.org/10.1128/jb.05535-11

Manglik, A., Kobilka, B. K., & Steyaert, J. (2017). Nanobodies to Study G Protein-Coupled Receptor Structure and Function. *Annual Review of Pharmacology and Toxicology*, *57*, 19–37. https://doi.org/10.1146/annurev-pharmtox-010716-104710

Mao, B., Pear, M. R., McCammon, J. A., & Quiocho, F. A. (1982). Hinge-bending in L-arabinose-binding protein. The "Venus's-flytrap" model. *Journal of Biological Chemistry*, *257*(3), 1131–1133. https://doi.org/10.1016/s0021-9258(19)68161-5

Maqbool, A., Horler, R. S. P., Muller, A., Wilkinson, A. J., Wilson, K. S., & Thomas, G. H. (2015). The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. *Biochemical Society Transactions*, *43*(5), 1011–1017. https://doi.org/10.1042/bst20150135

Marinelli, F., Kuhlmann, S. I., Grell, E., Kunte, H.-J., Ziegler, C., & Faraldo-Gómez, J. D. (2011). Evidence for an allosteric mechanism of substrate release from membrane-transporter accessory

binding proteins. *Proceedings of the National Academy of Sciences*, *108*(49), E1285–E1292. https://doi.org/10.1073/pnas.1112534108

Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science*, *322*(5909), 1843–1845. https://doi.org/10.1126/science.1165771

Marraffini, L. A., & Sontheimer, E. J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, *463*(7280), 568–571. https://doi.org/10.1038/nature08703

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, *40*, 658–674. https://doi.org/10.1107/s0021889807021206

Mireku, S. A., Sauer, M. M., Glockshuber, R., & Locher, K. P. (2017). Structural basis of nanobody-mediated blocking of BtuF, the cognate substrate-binding protein of the Escherichia coli vitamin B12 transporter BtuCD. *Scientific Reports*, *7*, 1–12. https://doi.org/10.1038/s41598-017-14512-8

Mitchell, J. E., Zheng, D., Busby, S. J. W., & Minchin, S. D. (2003). Identification and analysis of 'extended –10' promoters in Escherichia coli. *Nucleic Acids Research*, *31*(16), 4689–4695. https://doi.org/10.1093/nar/gkg694

Moffat, K. (1996). Time-resolved macromolecular crystallography. *Synchrotron Radiation News*, *9*(6), 15–18. https://doi.org/10.1080/08940889608602913

Mohanraju, P., Makarova, K. S., Zetsche, B., Zhang, F., Koonin, E. V., & Oost, J. van der. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science*, *353*(6299), aad5147. https://doi.org/10.1126/science.aad5147

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, *155*(3), 733–740. https://doi.org/10.1099/mic.0.023960-0

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular Evolution*, *60*(2), 174–182. https://doi.org/10.1007/s00239-004-0046-3

Mojica, F. J. M., Díez-Villaseñor, C., Soria, E., & Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology*, *36*(1), 244–246. https://doi.org/10.1046/j.1365-2958.2000.01838.x

Mojica, F. J. M., Ferrer, C., Juez, G., & Rodríguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. *Molecular Microbiology*, *17*(1), 85–93. https://doi.org/10.1111/j.1365-2958.1995.mmi_17010085.x

Molina, N., & Nimwegen, E. van. (2008). The evolution of domain-content in bacterial genomes. *Biology Direct*, *3*(1), 51. https://doi.org/10.1186/1745-6150-3-51

Morehouse, B. R., Govande, A. A., Millman, A., Keszei, A. F. A., Lowey, B., Ofir, G., Shao, S., Sorek, R., & Kranzusch, P. J. (2020). STING cyclic dinucleotide sensing originated in bacteria. *Nature*, *586*(7829), 429–433. https://doi.org/10.1038/s41586-020-2719-5

Motshwene, P. G., Moncrieffe, M. C., Grossmann, J. G., Kao, C., Ayaluru, M., Sandercock, A. M., Robinson, C. V., Latz, E., & Gay, N. J. (2009). An Oligomeric Signaling Platform Formed by the Toll-like Receptor Signal Transducers MyD88 and IRAK-4*. *Journal of Biological Chemistry*, *284*(37), 25404–25411. https://doi.org/10.1074/jbc.m109.022392

Müller, A., Severi, E., Mulligan, C., Watts, A. G., Kelly, D. J., Wilson, K. S., Wilkinson, A. J., & Thomas, G. H. (2006). Conservation of Structure and Mechanism in Primary and Secondary Transporters Exemplified by SiaP, a Sialic Acid Binding Virulence Factor from Haemophilus influenzae *. *Journal of Biological Chemistry*, *281*(31), 22212–22222. https://doi.org/10.1074/jbc.m603463200

Mulligan, C., Fischer, M., & Thomas, G. H. (2011). Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS Microbiology Reviews*, *35*, 68–86. https://doi.org/10.1111/j.1574-6976.2010.00236.x

Mulligan, C., Geertsma, E. R., Severi, E., Kelly, D. J., Poolman, B., & Thomas, G. H. (2009). The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 1778–1783. https://doi.org/10.1073/pnas.0809979106

Mulligan, C., Kelly, D. J., & Thomas, G. H. (2007). Tripartite ATP-Independent Periplasmic Transporters: Application of a Relational Database for Genome-Wide Analysis of Transporter Gene Frequency and Organization. *Journal of Molecular Microbiology and Biotechnology*, *12*(3–4), 218–226. https://doi.org/10.1159/000099643

Mulligan, C., Leech, A. P., Kelly, D. J., & Thomas, G. H. (2012). The Membrane Proteins SiaQ and SiaM Form an Essential Stoichiometric Complex in the Sialic Acid Tripartite ATP-independent Periplasmic (TRAP) Transporter SiaPQM (VC1777–1779) from Vibrio cholerae *. *Journal of Biological Chemistry*, *287*(5), 3598–3608. https://doi.org/10.1074/jbc.m111.281030

Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., & Darst, S. A. (2002). Structural Basis of Transcription Initiation: An RNA Polymerase Holoenzyme-DNA Complex. *Science*, *296*(5571), 1285–1290. https://doi.org/10.1126/science.1069595

Muyldermans, S. (2013). Nanobodies: Natural Single-Domain Antibodies. *Annual Review of Biochemistry*, *82*(1), 775–797. https://doi.org/10.1146/annurev-biochem-063011-092449

Muyldermans, S., Atarhouch, T., Saldanha, J., Barbosa, J. A. R. G., & Hamers, R. (1994). Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Engineering, Design and Selection*, *7*(9), 1129–1135. https://doi.org/10.1093/protein/7.9.1129

Muyldermans, S., Cambillau, C., & Wyns, L. (2001). Recognition of antigens by single-domain antibody fragments: the superfluous luxury of paired domains. *Trends in Biochemical Sciences*, *26*(4), 230–235. https://doi.org/10.1016/s0968-0004(01)01790-x

Nguyen, P. T., Lai, J. Y., Lee, A. T., Kaiser, J. T., & Rees, D. C. (2018). Noncanonical role for the binding protein in substrate uptake by the MetNI methionine ATP Binding Cassette (ABC) transporter. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E10596–E10604. https://doi.org/10.1073/pnas.1811003115

Nie, R., Stark, S., Symersky, J., Kaplan, R. S., & Lu, M. (2017). Structure and function of the divalent anion/Na+ symporter from Vibrio cholerae and a humanized variant. *Nature Communications*, *8*(1), 15009. https://doi.org/10.1038/ncomms15009

Niewoehner, O., Garcia-Doval, C., Rostøl, J. T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L. A., & Jinek, M. (2017). Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature*, *548*, 543–548. https://doi.org/10.1038/nature23467

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nature Structural & Molecular Biology*, *21*(6), 528–534. https://doi.org/10.1038/nsmb.2820

Nuñez, J. K., Lee, A. S. Y., Engelman, A., & Doudna, J. A. (2015). Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature*, *519*(7542), 193–198. https://doi.org/10.1038/nature14237

Ohtsubo, K., & Marth, J. D. (2006). Glycosylation in Cellular Mechanisms of Health and Disease. *Cell*, *126*, 855–867. https://doi.org/10.1016/j.cell.2006.08.019

Overington, J. P., Al-Lazikani, B., & Hopkins, A. L. (2006). How many drug targets are there? *Nature Reviews Drug Discovery*, *5*(12), 993–996. https://doi.org/10.1038/nrd2199

Padlan, E. A. (1994). Anatomy of the antibody molecule. *Molecular Immunology*, *31*(3), 169–217. https://doi.org/10.1016/0161-5890(94)90001-9

Paetzel, M., Dalbey, R. E., & Strynadka, N. C. J. (2002). Crystal Structure of a Bacterial Signal Peptidase Apoenzyme IMPLICATIONS FOR SIGNAL PEPTIDE BINDING AND THE SER-LYS DYAD MECHANISM*. *Journal of Biological Chemistry*, *277*(11), 9512–9519. https://doi.org/10.1074/jbc.m110983200

Paget, M. S. (2015). Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules*, *5*(3), 1245–1265. https://doi.org/10.3390/biom5031245

Paget, M. S., & Helmann, J. D. (2003). The σ70family of sigma factors. *Genome Biology*, *4*(1), 203. https://doi.org/10.1186/gb-2003-4-1-203

Pardon, E., Laeremans, T., Triest, S., Rasmussen, S. G. F., Wohlkönig, A., Ruf, A., Muyldermans, S., Hol, W. G. J., Kobilka, B. K., & Steyaert, J. (2014). A general protocol for the generation of Nanobodies for structural biology. *Nature Protocols*, *9*, 674–693. https://doi.org/10.1038/nprot.2014.039

Park, B. S., Song, D. H., Kim, H. M., Choi, B.-S., Lee, H., & Lee, J.-O. (2009). The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature*, *458*(7242), 1191–1195. https://doi.org/10.1038/nature07830

Pawluk, A., Davidson, A. R., & Maxwell, K. L. (2018). Anti-CRISPR: discovery, mechanism and function. *Nature Reviews Microbiology*, *16*(1), 12–17. https://doi.org/10.1038/nrmicro.2017.120

Peltola, H. (2000). Worldwide Haemophilus influenzae Type b Disease at the Beginning of the 21st Century: Global Analysis of the Disease Burden 25 Years after the Use of the Polysaccharide Vaccine and a Decade after the Advent of Conjugates. *Clinical Microbiology Reviews*, *13*(2), 302–317. https://doi.org/10.1128/cmr.13.2.302

Peter, M. F., Gebhardt, C., Glaenzer, J., Schneberger, N., Boer, M. de, Thomas, G. H., Cordes, T., & Hagelueken, G. (2021). Triggering Closure of a Sialic Acid TRAP Transporter Substrate Binding Protein through Binding of Natural or Artificial Substrates. *Journal of Molecular Biology*, *433*. https://doi.org/10.1016/j.jmb.2020.166756

Peter, M. F., Gebhardt, C., Mächtel, R., Muñoz, G. G. M., Glaenzer, J., Narducci, A., Thomas, G. H., Cordes, T., & Hagelueken, G. (2022). Cross-validation of distance measurements in proteins by PELDOR/DEER and single-molecule FRET. *Nature Communications*, *13*(1), 4396. https://doi.org/10.1038/s41467-022-31945-6

Peter, M. F., Ruland, J. A., Depping, P., Schneberger, N., Severi, E., Moecking, J., Gatterdam, K., Tindall, S., Durand, A., Heinz, V., Siebrasse, J. P., Koenig, P. A., Geyer, M., Ziegler, C., Kubitscheck, U., Thomas, G. H., & Hagelueken, G. (2022). Structural and mechanistic analysis of a tripartite ATP-independent periplasmic TRAP transporter. *Nature Communications*, *13*. https://doi.org/10.1038/s41467-022-31907-y

Peter, M. F., Ruland, J. A., Kim, Y., Hendricks, P., Schneberger, N., Siebrasse, J. P., Thomas, G. H., Kubitscheck, U., & Hagelueken, G. (2024). Conformational coupling of the sialic acid TRAP transporter HiSiaQM with its substrate binding protein HiSiaP. *Nature Communications*, *15*(1), 217. https://doi.org/10.1038/s41467-023-44327-3

Peters, J. E., Makarova, K. S., Shmakov, S., & Koonin, E. V. (2017). Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proceedings of the National Academy of Sciences*, *114*(35), E7358–E7366. https://doi.org/10.1073/pnas.1709035114

Pleschberger, M., Saerens, D., Weigert, S., Sleytr, U. B., Muyldermans, S., Sára, M., & Egelseer, E. M. (2004). An S-Layer Heavy Chain Camel Antibody Fusion Protein for Generation of a Nanopatterned Sensing Layer To Detect the Prostate-Specific Antigen by Surface Plasmon Resonance Technology. *Bioconjugate Chemistry*, *15*(3), 664–671. https://doi.org/10.1021/bc049964w

Poljak, R. J., Amzel, L. M., AVEY, H. P., Becka, L. N., & Nisonoff, A. (1972). Structure of Fab New at 6 Å Resolution. *Nature New Biology*, *235*(57), 137–140. https://doi.org/10.1038/newbio235137a0

Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, *151*(3), 653–663. https://doi.org/10.1099/mic.0.27437-0

Püschel, G., Kühn, H., Kietzmann, T., Höhne, W., Christ, B., Doenecke, D., & Koolman, J. (2011). *Taschenlehrbuch Biochemie* (1st ed.). Georg Thieme Verlag KG.

Quax, T. E. F., Voet, M., Sismeiro, O., Dillies, M.-A., Jagla, B., Coppée, J.-Y., Sezonov, G., Forterre, P., Oost, J. van der, Lavigne, R., & Prangishvili, D. (2013). Massive Activation of Archaeal Defense Genes during Viral Infection. *Journal of Virology*, *87*(15), 8419–8428. https://doi.org/10.1128/jvi.01020-13

Quick, M., & Javitch, J. A. (2007). Monitoring the function of membrane transport proteins in detergent-solubilized form. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 3603–3608. https://doi.org/10.1073/pnas.0609573104

Rabus, R., Jack, D. L., Kelly, D. J., & Saier, M. H. (1999). TRAP transporters: An ancient family of extracytoplasmic solute-receptor-dependent secondary active transporters. *Microbiology*, *145*, 3431–3445. https://doi.org/10.1099/00221287-145-12-3431

Radisky, E. S., Lee, J. M., Lu, C.-J. K., & Koshland, D. E. (2006). Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. *Proceedings of the National Academy of Sciences*, *103*(18), 6835–6840. https://doi.org/10.1073/pnas.0601910103

Ridley, R. G. (1988). *Antibodies: A Laboratory Manual.* (0 87969 314 2 & D. Lane, Eds.; Vol. 54). Cold Spring Harbor Laboratory.

Riesenfeld, C. S., Schloss, P. D., & Handelsman, J. (2004). METAGENOMICS: Genomic Analysis of Microbial Communities. *Annual Review of Genetics*, *38*(1), 525–552. https://doi.org/10.1146/annurev.genet.38.072902.091216

Robbe, C., Capon, C., Coddeville, B., & Michalski, J.-C. (2004). Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochemical Journal*, *384*(2), 307–316. https://doi.org/10.1042/bj20040605

Robbe-Masselot, C., Maes, E., Rousset, M., Michalski, J.-C., & Capon, C. (2009). Glycosylation of human fetal mucins: a similar repertoire of O-glycans along the intestinal tract. *Glycoconjugate Journal*, *26*(4), 397–413. https://doi.org/10.1007/s10719-008-9186-9

Rogozin, I. B., Makarova, K. S., Murvai, J., Czabarka, E., Wolf, Y. I., Tatusov, R. L., Szekely, L. A., & Koonin, E. V. (2002). Connected gene neighborhoods in prokaryotic genomes. *Nucleic Acids Research*, *30*(10), 2212–2223. https://doi.org/10.1093/nar/30.10.2212

Rosa, L. T., Bianconi, M. E., Thomas, G. H., & Kelly, D. J. (2018). Tripartite ATP-independent periplasmic (TRAP) transporters and Tripartite Tricarboxylate Transporters (TTT): From uptake to pathogenicity. *Frontiers in Cellular and Infection Microbiology*, *8*. https://doi.org/10.3389/fcimb.2018.00033

Rose, J. P., & Wang, B.-C. (2016). SAD phasing: History, current impact and future opportunities. *Archives of Biochemistry and Biophysics*, *602*, 80–94. https://doi.org/10.1016/j.abb.2016.03.018

Rostøl, J. T., Xie, W., Kuryavyi, V., Maguin, P., Kao, K., Froom, R., Patel, D. J., & Marraffini, L. A. (2021). The Card1 nuclease provides defence during type III CRISPR immunity. *Nature*, *590*(7847), 624–629. https://doi.org/10.1038/s41586-021-03206-x

Rouillon, C., Athukoralage, J. S., Graham, S., Grüschow, S., & White, M. F. (2018). Control of cyclic oligoadenylate synthesis in a type III CRISPR system. *ELife*, *7*, e36734. https://doi.org/10.7554/elife.36734

Rouillon*, C., Schneberger*, N., Chi, H., Blumenstock, K., Vela, S. D., Ackermann, K., Moecking, J., Peter, M. F., Boenigk, W., Seifert, R., Bode, B. E., Schmid-Burgk, J. L., Svergun, D., Geyer, M., White, M. F., & Hagelueken, G. (2023). Antiviral signalling by a cyclic nucleotide activated CRISPR protease. *Nature*, *614*(1), 168–174. https://doi.org/10.1038/s41586-022-05571-7

Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., Graham, S., Robinson, C. V., Spagnolo, L., & White, M. F. (2013). Structure of the CRISPR Interference Complex CSM Reveals Key Similarities with Cascade. *Molecular Cell*, *52*(1), 124–134. https://doi.org/10.1016/j.molcel.2013.08.020

Roux, K. H., Greenberg, A. S., Greene, L., Strelets, L., Avila, D., McKinney, E. C., & Flajnik, M. F. (1998). Structural analysis of the nurse shark (new) antigen receptor (NAR): Molecular convergence of NAR and unusual mammalian immunoglobulins. *Proceedings of the National Academy of Sciences*, *95*(20), 11804–11809. https://doi.org/10.1073/pnas.95.20.11804

Ruddigkeit, L., Deursen, R. van, Blum, L. C., & Reymond, J.-L. (2012). Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *Journal of Chemical Information and Modeling*, *52*(11), 2864–2875. https://doi.org/10.1021/ci300415d

Rudyak, S. G., Brenowitz, M., & Shrader, T. E. (2001). Mg2+-Linked Oligomerization Modulates the Catalytic Activiy of the Lon (La) Protease from Mycobacterium smegmatis †. *Biochemistry*, *40*(31), 9317–9323. https://doi.org/10.1021/bi0102508

Saerens, D., Ghassabeh, G. H., & Muyldermans, S. (2008). Single-domain antibodies as building blocks for novel therapeutics. *Current Opinion in Pharmacology*, *8*(5), 600–608. https://doi.org/10.1016/j.coph.2008.07.006

Saha, C. K., Pires, R. S., Brolin, H., Delannoy, M., & Atkinson, G. C. (2020). FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation. *Bioinformatics*, *37*(9), 1312–1314. https://doi.org/10.1093/bioinformatics/btaa788

Samai, P., Pyenson, N., Jiang, W., Goldberg, G. W., Hatoum-Aslan, A., & Marraffini, L. A. (2015). Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell*, *161*(5), 1164–1174. https://doi.org/10.1016/j.cell.2015.04.027

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Research*, *39*(21), 9275–9282. https://doi.org/10.1093/nar/gkr606

Sasnauskas, G., & Siksnys, V. (2020). CRISPR adaptation from a structural perspective. *Current Opinion in Structural Biology*, *65*, 17–25. https://doi.org/10.1016/j.sbi.2020.05.015

Schauer, R. (1985). Sialic acids and their role as biological masks. *Trends in Biochemical Sciences*, *10*(9), 357–360. https://doi.org/10.1016/0968-0004(85)90112-4

Scheepers, G. H., Nijeholt, J. A. L. a, & Poolman, B. (2016). An updated structural classification of substrate-binding proteins. *FEBS Letters*, *590*(23), 4393–4401. https://doi.org/10.1002/1873-3468.12445

Schuster, C. F., & Bertram, R. (2013). Toxin–antitoxin systems are ubiquitous and versatile modulators of prokaryotic cell fate. *FEMS Microbiology Letters*, *340*(2), 73–85. https://doi.org/10.1111/1574-6968.12074

Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biology*, *14*(8), e1002533. https://doi.org/10.1371/journal.pbio.1002533

Setty, T. G., Cho, C., Govindappa, S., Apicella, M. A., & Ramaswamy, S. (2014). Bacterial periplasmic sialic acid-binding proteins exhibit a conserved binding site. *Acta Crystallographica Section D: Biological Crystallography*, *70*, 1801–1811. https://doi.org/10.1107/s139900471400830x

Severi, E., Hood, D. W., & Thomas, G. H. (2007). Sialic acid utilization by bacterial pathogens. *Microbiology*, *153*(9), 2817–2822. https://doi.org/10.1099/mic.0.2007/009480-0

Severi, E., Randle, G., Kivlin, P., Whitfield, K., Young, R., Moxon, R., Kelly, D., Hood, D., & Thomas, G. H. (2005). Sialic acid transport in Haemophilus influenzae is essential for lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite ATP-independent periplasmic transporter. *Molecular Microbiology*, *58*, 1173–1185. https://doi.org/10.1111/j.1365-2958.2005.04901.x

Shaw, J. G., Hamblin, M. J., & Kelly, D. J. (1991). Purification, characterization and nucleotide sequence of the periplasmic C4-dicarboxylate-binding protein (DctP) from Rhodobacter capsulatus. *Molecular Microbiology*, *5*(12), 3055–3062. https://doi.org/10.1111/j.1365-2958.1991.tb01865.x

Shi, J., Zhao, Y., Wang, K., Shi, X., Wang, Y., Huang, H., Zhuang, Y., Cai, T., Wang, F., & Shao, F. (2015). Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. *Nature*, *526*(7575), 660–665. https://doi.org/10.1038/nature15514

Shi, Y. (2013). Common Folds and Transport Mechanisms of Secondary Active Transporters. *Annual Review of Biophysics*, *42*(1), 51–72. https://doi.org/10.1146/annurev-biophys-083012-130429

Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. (2018). Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E5307–E5316. https://doi.org/10.1073/pnas.1803440115

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O. O., Gootenberg, J. S., Makarova, K. S., Wolf, Y. I., Severinov, K., Zhang, F., & Koonin, E. V. (2017). Diversity and evolution of class 2 CRISPR–Cas systems. *Nature Reviews Microbiology*, *15*(3), 169–182. https://doi.org/10.1038/nrmicro.2016.184

Sifniotis, V., Cruz, E., Eroglu, B., & Kayser, V. (2019). Current Advancements in Addressing Key Challenges of Therapeutic Antibody Design, Manufacture, and Formulation. *Antibodies*, *8*(2), 36. https://doi.org/10.3390/antib8020036

Silas, S., Mohr, G., Sidote, D. J., Markham, L. M., Sanchez-Amat, A., Bhaya, D., Lambowitz, A. M., & Fire, A. Z. (2016). Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase–Cas1 fusion protein. *Science*, *351*(6276), aad4234. https://doi.org/10.1126/science.aad4234

Simanshu, D. K., Yamaguchi, Y., Park, J.-H., Inouye, M., & Patel, D. J. (2013). Structural Basis of mRNA Recognition and Cleavage by Toxin MazF and Its Regulation by Antitoxin MazE in Bacillus subtilis. *Molecular Cell*, *52*(3), 447–458. https://doi.org/10.1016/j.molcel.2013.09.006

Singh, S. S., Typas, A., Hengge, R., & Grainger, D. C. (2011). Escherichia coli σ 70 senses sequence and conformation of the promoter spacer region. *Nucleic Acids Research*, *39*(12), 5109–5118. https://doi.org/10.1093/nar/gkr080

Smalakyte, D., Ruksenaite, A., Sasnauskas, G., Tamulaitiene, G., & Tamulaitis, G. (2024). Filament formation activates protease and ring nuclease activities of CRISPR SAVED-Lon. *BioRxiv*, 2024.05.08.593097. https://doi.org/10.1101/2024.05.08.593097

Snel, B., Bork, P., & Huynen, M. (2000). Genome evolution gene fusion versus gene fission. *Trends in Genetics*, *16*(1), 9–11. https://doi.org/10.1016/s0168-9525(99)01924-1

Srinivasan, N. (2010). Telling apart friend from foe: discriminating between commensals and pathogens at mucosal sites. *Innate Immunity*, *16*(6), 391–404. https://doi.org/10.1177/1753425909357577

Staroń, A., Sofia, H. J., Dietrich, S., Ulrich, L. E., Liesegang, H., & Mascher, T. (2009). The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family. *Molecular Microbiology*, *74*(3), 557–581. https://doi.org/10.1111/j.1365-2958.2009.06870.x

Steens, J. A., Bravo, J. P. K., Salazar, C. R. P., Yildiz, C., Amieiro, A. M., Köstlbacher, S., Prinsen, S. H. P., Andres, A. S., Patinios, C., Bardis, A., Barendregt, A., Scheltema, R. A., Ettema, T. J. G., Oost, J. van der, Taylor, D. W., & Staals, R. H. J. (2024). Type III-B CRISPR-Cas cascade of proteolytic cleavages. *Science*, *383*(6682), 512–519. https://doi.org/10.1126/science.adk0378

Steens, J. A., Bravo, J. P. K., Salazar, C. R. P., Yildiz, C., Amieiro, A. M., Köstlbacher, S., Prinsen, S. H. P., Patinios, C., Bardis, A., Barendregt, A., Scheltema, R. A., Ettema, T. J. G., Oost, J. van der, Taylor, D. W., & Staals, R. H. J. (2023). Type III-B CRISPR-Cas signaling-based cascade of proteolytic cleavages. *BioRxiv*, 2023.06.23.546230. https://doi.org/10.1101/2023.06.23.546230

Stern, A., & Sorek, R. (2011). The phage-host arms race: Shaping the evolution of microbes. *BioEssays*, *33*(1), 43–51. https://doi.org/10.1002/bies.201000071

Stevens, F. J., & Schiffer, M. (1981). Computer simulation of protein self-association during small-zone gel filtration. Estimation of equilibrium constants. *The Biochemical Journal*, *195*, 213–219. https://doi.org/10.1042/bj1950213

Stokar-Avihail, A., Tal, N., Erez, Z., Lopatina, A., & Sorek, R. (2019). Widespread Utilization of Peptide Communication in Phages Infecting Soil and Pathogenic Bacteria. *Cell Host & Microbe*, *25*(5), 746-755.e5. https://doi.org/10.1016/j.chom.2019.03.017

Strecker, J., Demircioglu, F. E., Li, D., Faure, G., Wilkinson, M. E., Gootenberg, J. S., Abudayyeh, O. O., Nishimasu, H., Macrae, R. K., & Zhang, F. (2022). RNA-activated protein cleavage with a CRISPR-associated endopeptidase. *Science*, *378*, 874–881. https://doi.org/10.1126/science.add7450

Streltsov, V. A., Varghese, J. N., Carmichael, J. A., Irving, R. A., Hudson, P. J., & Nuttall, S. D. (2004). Structural evidence for evolution of shark Ig new antigen receptor variable domain antibodies from a cell-surface receptor. *Proceedings of the National Academy of Sciences*, *101*(34), 12444–12449. https://doi.org/10.1073/pnas.0403509101

Sturme, M. H. J., Kleerebezem, M., Nakayama, J., Akkermans, A. D. L., Vaughan, E. E., & Vos, W. M. de. (2002). Cell to cell communication by autoinducing peptides in gram-positive bacteria. *Antonie van Leeuwenhoek*, *81*(1–4), 233–243. https://doi.org/10.1023/a:1020522919555

Sui, L., & Guo, H.-C. (2021). ERAP1 binds peptide C-termini of different sequences and/or lengths by a common recognition mechanism. *Immunobiology*, *226*(4), 152112. https://doi.org/10.1016/j.imbio.2021.152112

Sukupolvi-Petty, S., Grass, S., & StGeme, J. W. (2006). The Haemophilus influenzae Type b hcsA and hcsB Gene Products Facilitate Transport of Capsular Polysaccharide across the Outer Membrane and Are Essential for Virulence. *Journal of Bacteriology*, *188*(11), 3870–3877. https://doi.org/10.1128/jb.01968-05

Sun, L., Wu, J., Du, F., Chen, X., & Chen, Z. J. (2013). Cyclic GMP-AMP Synthase Is a Cytosolic DNA Sensor That Activates the Type I Interferon Pathway. *Science*, *339*(6121), 786–791. https://doi.org/10.1126/science.1232458

Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D. L., Pulcini, C., Kahlmeter, G., Kluytmans, J., Carmeli, Y., Ouellette, M., Outterson, K., Patel, J., Cavaleri, M., Cox, E. M., Houchens, C. R., Grayson, M. L., Hansen, P., Singh, N., … Zorzet, A. (2018). Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The Lancet Infectious Diseases*, *18*(3), 318–327. https://doi.org/10.1016/s1473-3099(17)30753-3

Takayama, Y., Castañeda, C. A., Chimenti, M., García-Moreno, B., & Iwahara, J. (2008). Direct Evidence for Deprotonation of a Lysine Side Chain Buried in the Hydrophobic Core of a Protein. *Journal of the American Chemical Society*, *130*(21), 6714–6715. https://doi.org/10.1021/ja801731g

Takeuchi, N., Wolf, Y. I., Makarova, K. S., & Koonin, E. V. (2012). Nature and Intensity of Selection Pressure on CRISPR-Associated Genes. *Journal of Bacteriology*, *194*(5), 1216–1225. https://doi.org/10.1128/jb.06521-11

Tamulaitis, G., Venclovas, Č., & Siksnys, V. (2017). Type III CRISPR-Cas Immunity: Major Differences Brushed Aside. *Trends in Microbiology*, *25*(1), 49–61. https://doi.org/10.1016/j.tim.2016.09.012

Trakhanov, S., Vyas, N. K., Luecke, H., Kristensen, D. M., Ma, J., & Quiocho, F. A. (2005). Ligand-Free and -Bound Structures of the Binding Protein (LivJ) of the Escherichia coli ABC Leucine/Isoleucine/Valine Transport System: Trajectory and Dynamics of the Interdomain Rotation and Ligand Specificity †. *Biochemistry*, *44*(17), 6597–6608. https://doi.org/10.1021/bi047302o

Treviño-Quintanilla, L. G., Freyre-González, J. A., & Martínez-Flores, I. (2013). Anti-Sigma Factors in E. coli: Common Regulatory Mechanisms Controlling Sigma Factors Availability. *Current Genomics*, *14*(6), 378–387. https://doi.org/10.2174/1389202911314060007

Tsilibaris, V., Maenhaut-Michel, G., & Melderen, L. V. (2006). Biological roles of the Lon ATP-dependent protease. *Research in Microbiology*, *157*(8), 701–713. https://doi.org/10.1016/j.resmic.2006.05.004

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., Pritzel, A., Figurnov, M., Ronneberger, O., Bates, R., Kohl, S. A. A., … Hassabis, D. (2021). Highly accurate protein structure prediction for the human proteome. *Nature*, *596*(7873), 590–596. https://doi.org/10.1038/s41586-021-03828-1

Tzeng, S.-R., Tseng, Y.-C., Lin, C.-C., Hsu, C.-Y., Huang, S.-J., Kuo, Y.-T., & Chang, C.-I. (2021). Molecular insights into substrate recognition and discrimination by the N-terminal domain of Lon AAA+ protease. *ELife*, *10*, e64056. https://doi.org/10.7554/elife.64056

Valenciano-Bellido, S., Caaveiro, J. M. M., Nakakido, M., Kuroda, D., Aikawa, C., Nakagawa, I., & Tsumoto, K. (2023). Targeting hemoglobin receptors IsdH and IsdB of Staphylococcus aureus with a single VHH antibody inhibits bacterial growth. *Journal of Biological Chemistry*, *299*(9), 104927. https://doi.org/10.1016/j.jbc.2023.104927

Vanden Broeck, D., Horvath, C., & Wolf, M. J. S. D. (2007). Vibrio cholerae: Cholera toxin. *The International Journal of Biochemistry & Cell Biology*, *39*, 1771–1775. https://doi.org/10.1016/j.biocel.2007.07.005

van der Linden, R. H. J., Frenken, L. G. J., Geus, B. de, Harmsen, M. M., Ruuls, R. C., Stok, W., Ron, L. de, Wilson, S., Davis, P., & Verrips, C. T. (1999). Comparison of physical chemical properties of llama VHH antibody fragments and mouse monoclonal antibodies. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, *1431*(1), 37–46. https://doi.org/10.1016/s0167-4838(99)00030-8

Van Regenmortel, M. H. (1998). From absolute to exquisite specificity. Reflections on the fuzzy nature of species, specificity and antigenic sites. *Journal of Immunological Methods*, *216*(1–2), 37–48. https://doi.org/10.1016/s0022-1759(98)00069-6

Ventola, C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. *P & T : A Peer-Reviewed Journal for Formulary Management*, *40*(4), 277–283.

Verdon, G., Oh, S., Serio, R. N., & Boudker, O. (2014). Coupled ion binding and structural transitions along the transport cycle of glutamate transporters. *ELife*, *3*, e02283. https://doi.org/10.7554/elife.02283

Vighi, G., Marcucci, F., Sensi, L., Cara, G. D., & Frati, F. (2008). Allergy and the gastrointestinal system. *Clinical & Experimental Immunology*, *153*(Supplement_1), 3–6. https://doi.org/10.1111/j.1365-2249.2008.03713.x

Vimr, E., Lichtensteiger, C., & Steenbergen, S. (2000). Sialic acid metabolism's dual function in Haemophilus influenzae. *Molecular Microbiology*, *36*, 1113–1123. https://doi.org/10.1046/j.1365-2958.2000.01925.x

Vimr, E. R., Kalivoda, K. A., Deszo, E. L., & Steenbergen, S. M. (2004). Diversity of Microbial Sialic Acid Metabolism. *Microbiology and Molecular Biology Reviews*, *68*(1), 132–153. https://doi.org/10.1128/mmbr.68.1.132-153.2004

Vonrhein, C., & Bricogne, G. (2008). AutoPROC – a framework for automated data processing. *Acta Crystallographica Section A Foundations of Crystallography*, *64*(a1), C78–C78. https://doi.org/10.1107/s010876730809750x

Vu, K. B., Ghahroudi, M. A., Wyns, L., & Muyldermans, S. (1997). Comparison of llama VH sequences from conventional and heavy chain antibodies. *Molecular Immunology*, *34*(16–17), 1121–1131. https://doi.org/10.1016/s0161-5890(97)00146-6

Wallerstein, J., Weininger, U., Khan, M. A. I., Linse, S., & Akke, M. (2015). Site-Specific Protonation Kinetics of Acidic Side Chains in Proteins Determined by pH-Dependent Carboxyl 13C NMR Relaxation. *Journal of the American Chemical Society*, *137*(8), 3093–3101. https://doi.org/10.1021/ja513205s

Wang, B., & Brand-Miller, J. (2003). The role and potential of sialic acid in human nutrition. *European Journal of Clinical Nutrition*, *57*, 1351–1369. https://doi.org/10.1038/sj.ejcn.1601704

Wang, S., Xie, J., Pei, J., & Lai, L. (2023). CavityPlus 2022 Update: An Integrated Platform for Comprehensive Protein Cavity Detection and Property Analyses with User-friendly Tools and Cavity Databases. *Journal of Molecular Biology*, *435*(14), 168141. https://doi.org/10.1016/j.jmb.2023.168141

Wang, Y., Guo, Y., Li, G., Liu, C., Wang, L., Zhang, A., Yan, Z., & Song, C. (2021). The push-to-open mechanism of the tethered mechanosensitive ion channel NompC. *ELife*, *10*, e58388. https://doi.org/10.7554/elife.58388

Ward, E. S., Güssow, D., Griffiths, A. D., Jones, P. T., & Winter, G. (1989). Binding activities of a repertoire of single immunoglobulin variable domains secreted from Escherichia coli. *Nature*, *341*(6242), 544–546. https://doi.org/10.1038/341544a0

Whiteley, A. T., Eaglesham, J. B., Mann, C. C. de O., Morehouse, B. R., Lowey, B., Nieminen, E. A., Danilchanka, O., King, D. S., Lee, A. S. Y., Mekalanos, J. J., & Kranzusch, P. J. (2019). Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature*, *567*(7747), 194–199. https://doi.org/10.1038/s41586-019-0953-5

Wilson, I. A., & Stanfield, R. L. (1994). Antibody-antigen interactions: new structures and new conformational changes. *Current Opinion in Structural Biology*, *4*(6), 857–867. https://doi.org/10.1016/0959-440x(94)90267-4

Wlodawer, A., Sekula, B., Gustchina, A., & Rotanova, T. V. (2022). Structure and the Mode of Activity of Lon Proteases from Diverse Organisms. *Journal of Molecular Biology*, *434*(7), 167504. https://doi.org/10.1016/j.jmb.2022.167504

Wollenhaupt, J., Metz, A., Barthel, T., Lima, G. M. A., Heine, A., Mueller, U., Klebe, G., & Weiss, M. S. (2020). F2X-Universal and F2X-Entry: Structurally Diverse Compound Libraries for Crystallographic Fragment Screening. *Structure*, *28*(6), 694-706.e5. https://doi.org/10.1016/j.str.2020.04.019

Wu, J., Sun, L., Chen, X., Du, F., Shi, H., Chen, C., & Chen, Z. J. (2013). Cyclic GMP-AMP Is an Endogenous Second Messenger in Innate Immune Signaling by Cytosolic DNA. *Science*, *339*(6121), 826–830. https://doi.org/10.1126/science.1229963

Xia, P., Dutta, A., Gupta, K., Batish, M., & Parashar, V. (2022). Structural basis of cyclic oligoadenylate binding to the transcription factor Csa3 outlines crosstalk between Type-III & Type-I CRISPR systems. *Journal of Biological Chemistry*, *298*(Biopolymers 95 2011), 101591. https://doi.org/10.1016/j.jbc.2022.101591

Xiao, Y., Ng, S., Nam, K. H., & Ke, A. (2017). How type II CRISPR–Cas establish immunity through Cas1–Cas2-mediated spacer integration. *Nature*, *550*(7674), 137–141. https://doi.org/10.1038/nature24020

Xu, Y., Wang, S., Hu, Q., Gao, S., Ma, X., Zhang, W., Shen, Y., Chen, F., Lai, L., & Pei, J. (2018). CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. *Nucleic Acids Research*, *46*(Web Server issue), gky380-. https://doi.org/10.1093/nar/gky380

Zamyatina, A., & Heine, H. (2020). Lipopolysaccharide Recognition in the Crossroads of TLR4 and Caspase-4/11 Mediated Inflammatory Pathways. *Frontiers in Immunology*, *11*, 585146. https://doi.org/10.3389/fimmu.2020.585146

Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.-V., Naismith, J. H., Spagnolo, L., & White, M. F. (2012). Structure and Mechanism of the CMR Complex for CRISPR-Mediated Antiviral Immunity. *Molecular Cell*, *45*(3), 303–313. https://doi.org/10.1016/j.molcel.2011.12.013

Zhong, B., Yang, Y., Li, S., Wang, Y.-Y., Li, Y., Diao, F., Lei, C., He, X., Zhang, L., Tien, P., & Shu, H.-B. (2008). The Adaptor Protein MITA Links Virus-Sensing Receptors to IRF3 Transcription Factor Activation. *Immunity*, *29*(4), 538–550. https://doi.org/10.1016/j.immuni.2008.09.003

Zorzini, V., Mernik, A., Lah, J., Sterckx, Y. G. J., Jonge, N. D., Garcia-Pino, A., Greve, H. D., Versées, W., & Loris, R. (2016). Substrate Recognition and Activity Regulation of the Escherichia coli mRNA Endonuclease MazF* ♦. *Journal of Biological Chemistry*, *291*(21), 10950–10960. https://doi.org/10.1074/jbc.m116.715912

Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., Terwilliger, T. C., & Adams, P. D. (2008). Structural Proteomics, High-Throughput Methods. *Methods in Molecular Biology*, *426*, 419–435. https://doi.org/10.1007/978-1-60327-058-8_28

List of Abbreviations

ABC	ATP binding cassette
ADP	adenosine diphosphate
AMP	adenosine monophosphate
ASD	Anti-sigma domain
ATP	adenosine triphosphate
A_x	absorbance at x nm
cA_n	cyclic n x (adenylate)
CARF	CRISPR associated Rossmann Fold
Cas	CRISPR associated
CBASS	cyclic oligonucleotide-based antiphage signaling system
CDR	complementary determining region
cGAMP	cyclic guanosine monophosphate–adenosine monophosphate
cGAS	cyclic GMP-AMP synthase
cOA	cyclic oligoadenylate
CRISPR	clustered and regularly interspaced short palindromic repeats
DESY	Deutsches Elektronen Synchrotron
DLS	dynamic light scattering
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
ELISA	enzyme-linked immunosorbent assay
HiSiaP	P-domain of the TRAP-transporter from *Haemophilus influenzae*
HiSiaQM	QM-domains of the TRAP-transporter from *Haemophilus influenzae*
IANBD	N,N'-Dimethyl-N-(iodoacetyl) -N'-(7-nitrobenz-2-oxa-1,3-diazol- 4-yl)ethyle ylenediamine
ITC	isothermal titration calorimetry
MALS	multi-angle light scattering
nanoDSF	nano differential scanning fluorometry
Neu5Ac	N-acetyl neuraminic acid
PDB	protein data bank
PDB ID	PDB identifier / accession code
RNA	ribonucleic acid
RNAP	RNA-Polymerase
RNase	ribonuclease
SAD	single-wavelength anomalous diffraction
SAVED	SMODS-associated and fused to various effector domains
SBP	substrate binding protein
SEC	size-exclusion chromatography
SLS	Swiss Light Source
SMODS	second messenger oligonucleotide or dinucleotide synthetase
STING	Stimulator of interferon genes
TRAP	tripartite ATP-independent periplasmic
VcSiaP	P-domain of the TRAP-transporter from *Vibrio cholerae*
VcSiaQM	QM-domains of the TRAP-transporter from *Vibrio cholerae*

List of Tables

Table 2-1: Binding parameters of the 1:1 complex formation of SiaPs and their specific VHHs. 31
Table 7-1: PCR reaction mix 119
Table 7-2: PCR thermocycler program 119
Table 7-3: PCR thermocycler program for site-directed mutagenesis 121
Table 7-4: Composition of polyacrylamide gels. 122
Table 7-5: UniProt IDs TRAP project 124
Table 7-6: UniProt IDs CRISPR project 125
Table 7-7: Specific information for analytical size exclusion chromatography. 129
Table 7-8: Specific information for multi-angle light scattering experiments. 129
Table 7-9: Detailed information VcSiaP-VHH ITC experiments. 131
Table 7-10: Detailed information SiaP-Neu5Ac ITC experiments. 132
Table 7-11: Detailed information competitive VcSiaP-VHH/Neu5Ac ITC experiments. 132
Table 7-12: Detailed information on CalpL-CalpT titration experiments 132
Table 7-13: Detailed information on HiSiaP-VHH ITC experiments. 133
Table 7-14: Detailed information for sequential VcSiaP-VHH/Neu5Ac ITC experiments. 134
Table 7-15: Crystallization conditions that yielded the crystals used for structure determination. ..135
Table 7-16: Quality of the datasets and final refinements used for structure determination. 137
Table 8-1: List of used chemicals and the corresponding manufacturer. 139
Table 8-2: List of consumables and their corresponding manufacturers. 140
Table 8-3: List of used enzymes and buffers including the corresponding manufacturer. 140
Table 8-4: List of protein purification buffers and their composition. 141
Table 8-5: List of media for cultivation of E. coli cells. 142
Table 8-6: E. coli strains used in this work. 142
Table 8-7: Plasmids used in this work and their characteristics. 143
Table 8-8: Oligonucleotides for site-directed mutagenesis used in this work. 144
Table 8-9: Oligonucleotides used for PCR amplification of genes or gene fragments. 145
Table 8-10: Instruments used for protein purification and analysis. 145
Table 8-11: List of used software. 146
Extended Data Table 1: Crystallographic data of VcSiaP$_{apo}$–nanobody 1:1 complexes. 148
Extended Data Table 2: Crystallographic data of VcSiaP W73A$_{holo}$–NbS002 complex. 149
Extended Data Table 3: Table1 for VcSiaP[NbS001/NbS002] and VcSiaP$_{holo}$[NbS001]. 150
Extended Data Table 4: Crystallographic data of CalpL and CalpL in complex with cA$_4$. 154
Extended Data Table 5: Table1 for CalpL[CalpT] and CalpT[NbS023] complexes. 154

List of Figures

Figure 1-1: Chemical structure of sialic acid. 8
Figure 1-2: Overview of important key elements of the sialic acid utilization in bacteria. 9
Figure 1-3: LPS induced TLR4 pathway....10
Figure 1-4: LPS composition of different H. influenzae strains dependent on Neu5Ac availability..11
Figure 1-5: Comparing overview of three different transporter families. 13
Figure 1-6: Different types of transport mechanisms across the cytoplasmic membrane.....14
Figure 1-7: Working model for the mechanism of transport of TRAP transporters. 15
Figure 1-8: Conformational changes of ABC transporter substrate binding proteins.....16
Figure 1-9: Domain architecture and closing transition of sialic acid TRAP transporter SBPs. 17
Figure 1-10: Natural or artificial substrates trigger the same structural rearrangements in VcSiaP...18
Figure 1-11: Overview on IgG antibodies and stable antigen binding fragments thereof. 20
Figure 1-12: Overview on heavy chain only antibodies and VHH antibodies.....21
Figure 2-1: Purification of VcSiaP. 25
Figure 2-2: From immunization to hit identification and sequence analysis of VHH antibodies.....27
Figure 2-3: Overview on expression and purification of VHH antibodies. 28
Figure 2-4: Analysis of 1:1 complex formation of VcSiaP and NbS001 and NbS002, respectively..29
Figure 2-5: Biochemical characterization of HiSiaP binding VHH antibody NbS003.....30
Figure 2-6: Biochemical analysis of HiSiaP–VHH complexes and determination of epitopes. 32
Figure 2-7: Epitope binning experiments for VcSiaP nanobodies. 33
Figure 2-8: Structural analysis of VcSiaP/VHH 1:1 complexes. 34
Figure 2-9: Structural analysis of the heterotrimeric VcSiaP/NbS001/NbS002 complex. 36
Figure 2-10: Visualization of ITC experiments to analyze VHH binding in presence of sialic acid..37
Figure 2-11: Substrate binding of VcSiaP is inhibited by preincubation with NbS002.....38
Figure 2-12: Allosteric inhibition by preventing conformational changes. 39
Figure 2-13: VcSiaP mutant W73A is able to bind Neu5Ac in presence of NbS002. 40
Figure 2-14: Schematic representation of the events during the sequential ITC analysis.....41
Figure 2-15: Results from sequential ITC analysis of wild type VcSiaP and W73A. 42
Figure 2-16: Structural analysis of $VcSiaP_{W73A}$ bound to sialic acid and NbS002. 44
Figure 3-1: Different types of stabilization for the CDR3 region of VcSiaP specific Nanobodies. ...48
Figure 3-2: Overview on the dimeric assembly of the VcSiaP[NbS001, NbS002] complex.....49
Figure 3-3: Structural modelling reveals a hydrophobic loop next to the NbS002 binding site. 50
Figure 3-4: Elevator mechanism including an allosterical triggered SBP opening mechanism. 52
Figure 3-5: Overview of predicted druggable surface cavities of VcSiaP. 53
Figure 4-1: Classification of CRISPR systems. 62
Figure 4-2: Representation of an exemplary CRISPR locus. 63
Figure 4-3: Similarities and differences in the mode of action of different CRISPR types. 64
Figure 4-4: Immune signaling by cyclic nucleotides – an overview. 66
Figure 4-5: Chemical structures of different 3',5'-linked cyclic oligoadenylates.....67
Figure 4-6: Schematic representation of the architecture of CARF- and SAVED domains. 69
Figure 4-7: CRISPR loci of type III systems containing predicted CRISPR-associated genes. 70
Figure 4-8: The active site of a Lon protease. 71
Figure 4-9: Analysis of the gene neighborhood of calpL reveals functionally linked genes. 72
Figure 4-10: Domain architecture and structural arrangement of bacterial sigma factors. 73
Figure 5-1: Optimization of CalpL purification procedure. 78
Figure 5-2: Western Blot analysis of CalpL. 79
Figure 5-3: Crystallization of a selenomethionine CalpL derivative. 80
Figure 5-4: Structural overview of CalpL, solved by experimental phasing. 81
Figure 5-5: Domain architecture of CalpL and structural similarity search of domains. 82
Figure 5-6: Protease activity assay to determine activator and target at once.....83
Figure 5-7: The SAVED domain of CalpL contains a deep cavity to specifically bind cA_4. 84
Figure 5-8: Structure guided and detailed investigation of the cOA binding site. 86
Figure 5-9: Binding of cyclic tetraadenylate triggers oligomerization of CalpL. 88
Figure 5-10: Determination of the cleavage directionality by protease activity assays. 89
Figure 5-11: CalpL substrate channel superposition. 90
Figure 5-12: Structural prediction of CalpT helped to identify the protease cleavage site. 91

Figure 5-13: Analysis of the effect of CalpT cleavage site mutations on complex formation. 92
Figure 5-14: CalpT-L complex analysis and crystallization. 92
Figure 5-15: Structural analysis of $CalpT_{10}$–CalpL complex formation. 94
Figure 5-16: Investigating the role $CalpT_{10}$ with respect to an efficient proteolytic cleavage. 95
Figure 5-17: Investigating the structural homology of CalpT to a MazF toxin. 96
Figure 5-18: CalpT forms a complex with the ECF sigma factor CalpS. 97
Figure 5-19: Mutational CalpS-T complex analysis. 98
Figure 5-20: VHH antibody stabilizes CalpT for crystallization. 99
Figure 5-21: Crystal packing analysis of CalpT-NbS023 and activity assay. 100
Figure 6-1: Gene neighborhood analysis revealed interesting CalpL homologs. 103
Figure 6-2: CalpL homologs are predicted to bind a different potential protease target 104
Figure 6-3: CalpL variants that are predicted to contain only the protease- and SAVED domain. 106
Figure 6-4: Dimer of CalpL and a SAVED-only variant of it, encoded by two neighboring genes. 107
Figure 6-5: Structural investigation of a CalpL variant that harbors an additional domain. 108
Figure 6-6: Possible mechanism of proteolytic cleavage for Lon proteases. 110
Figure 6-7: Detailed view on the catalytic dyad. 111
Figure 6-8: The CalpS-T interaction interface as predicted by AlphaFold2. 112
Figure 6-9: Working hypothesis for the CalpS-T-L mediated antiviral signaling cascade. 113
Figure 6-10: Sketch of CalpL filaments. 115
Figure 6-11: Illustrations of planned TIRF experiments to study CalpL oligomerization. 116
Figure 6-12: Preventing possible conformational changes by engineered disulfide bridges. 117
Figure 9-1: Demonstration of the functional closing mechanism of the VcSiaP/NbS001 complex. 147
Figure 9-2: Mass spectrometric analysis of CalpL, CalpT and the proteolytiy cleavage fragments. 151
Figure 9-3: Additional illustration of the $CalpT_{10}$–CalpL complex. 151
Figure 9-4: Further analysis of CalpS-T-L and RNA polymerase complexes. 152
Figure 9-5: Electrostatic surface analysis of the FtCalpL and FtsoloSAVED. 153