UNIVERSITÄT ZU LÜBECK

From the Institute of Chemistry and Metabolomics
of the University of Lübeck
Director: Prof. Dr. rer. nat. Thomas Peters

On the Interaction of Norovirus Capsid Proteins with Glycans and Bile Acids

Zur Interaktion der Noroviruskapsidproteine mit Glykanen und Gallensäuren

Dissertation
for Fulfillment of
Requirements
for the Doctoral Degree
of the University of Lübeck

from the Department of Natural Sciences

Submitted by
Robert Creutznacher
from Berlin

Lübeck, 2020

First referee: Prof. Dr. Thomas Peters
Second referee: Prof. Dr. Christian Hübner
Date of oral examination: August 4, 2020
Approved for printing. Lübeck, August 5, 2020

Contents

Acknowledgments & Contributions **ix**

1 Abstract **1**

2 Zusammenfassung **3**

3 Introduction **5**

3.1 Noroviruses 5

3.1.1 Disease & Epidemiology 5

3.1.2 The Norovirus Particle 6

3.1.3 Murine Noroviruses 8

3.1.4 The Role of Glycans in Norovirus Infections 8

3.1.5 The Role of Bile Acids in Norovirus Infections 11

3.2 Asparagine Deamidation in Proteins 13

3.3 Biomolecular Nuclear Magnetic Resonance (NMR) Spectroscopy 14

3.3.1 The Chemical Shift 14

3.3.2 Relaxation & Molecular Motion 16

3.3.3 Isotopic Labeling 17

3.3.4 Protein Signal Assignment - 3D NMR Spectroscopy 18

3.3.5 Saturation Transfer Difference (STD) NMR Experiments 20

3.3.6 Paramagnetic Lanthanides in NMR Spectroscopy 21

3.4 Objectives 22

4 Results **23**

4.1 NMR Backbone Assignment of a GII.4 NoV Protruding Domain 23

4.2 A Post-Translational Modification Impairs Glycan Recognition 28

4.3 A Novel, Low-Affinity Binding Pocket for Bile Acids on GII.4 Norovirus Capsids . . . 37

4.4 Bile Acids Functionalize The Murine Norovirus Protruding Domain in Solution 41

4.5 Protein-Metal Interactions of Norovirus Capsids 50

4.6 Glycan Recognition of Human and Murine NoV Capsid Proteins 55

5 Discussion **63**

5.1 Backbone Assignment and Deamidation of Asn373 in Human GII.4 Noroviruses . . . 63

5.2 Interactions of Norovirus Capsids with Bile Acids . . . 65

5.3 Norovirus Protruding Domains are Metalloproteins . . . 68

5.4 Redefining the Glycan Code of Murine and Human Noroviruses . . . 70

5.5 Conclusions & Outlook . . . 72

6 Materials & Methods **75**

6.1 Expression Vectors . . . 75

6.1.1 Genes . . . 75

6.1.2 Polymerase Chain Reaction (PCR) . . . 75

6.1.3 Bacterial Growth Media . . . 76

6.1.4 Bacterial Transformation . . . 76

6.1.5 Bacterial Cryo Stocks . . . 77

6.2 Protein Biosynthesis . . . 77

6.2.1 Unlabeled Protruding Domain . . . 77

6.2.2 Isotope-Labeled Protruding Domain . . . 78

6.3 Protein Purification & Biochemical Analysis . . . 78

6.3.1 Cell Lysis . . . 78

6.3.2 Affinity Chromatography & Proteolytic Cleavage . . . 78

6.3.3 Size Exclusion Chromatography (SEC) . . . 79

6.3.4 Denaturing Gel Electrophoresis . . . 80

6.3.5 Chemical Denaturation & Refolding of the GII.4 Protruding Domain . . . 80

6.3.6 Ion Exchange Chromatography (IEX) . . . 80

6.3.7 Hydrophobic Interaction Chromatography (HIC) . . . 81

6.4 Nuclear Magnetic Resonance (NMR) Spectroscopy . . . 82

6.4.1 NMR samples . . . 82

6.4.2 Chemical Shift Referencing . . . 82

6.4.3 Backbone Amide Signal Assignment - 3D NMR Spectroscopy . . . 83

6.4.4 Potential Ligand Molecules Used in STD and CSP Experiments . . . 84

6.4.5 Chemical Shift Perturbation (CSP) Experiments . . . 86

6.4.6 Saturation Transfer Difference (STD) NMR Spectroscopy . . . 86

6.4.7 ^{15}N Backbone Relaxation . . . 87

6.4.8 Dimerization & GCDCA-Binding of the Murine NoV P-Domain . . . 88

6.5 Software . . . 89

6.6 Data Availability . . . 89

Bibliography **91**

Appendix **111**

6.7 Protein Biosynthesis and Purification . . . 111
6.8 Hydrophobic Interaction Chromatography . . . 113
6.9 Simultaneous Deamidation and Dimer Reassembly . . . 114
6.10 Source Code . . . 115
6.10.1 Chemical Shift Referencing . . . 115
6.10.2 Fitting of Simultaneous Deamidation and Dimer Reassembly Reaction Rates . 116
6.10.3 Analysis of TRACT experiments . . . 118
6.10.4 MNV-GCDCA Interaction . . . 119
6.10.5 Calcium Contamination . . . 121
6.10.6 Pulse program ^{15}N T_1 times . . . 122
6.10.7 Pulse program ^{15}N $T_{1\rho}$ times . . . 126
6.11 Curriculum Vitae . . . 130

Acknowledgments & Contributions

First and foremost, I would like to thank **Prof. Dr. Thomas Peters** for his guidance through the different stages of academic life since I was an undergraduate student. I do not take the given support for granted and I am grateful for the enjoyable work on the project.

All members of the Institute of Chemistry and Metabolomics are greatly thanked for creating such a welcoming atmosphere. In particular, I greatly appreciate the unwavering scientific and personal support from **Barbara Richer** and **Dr. Thorsten Biet**. I could not have wished for better colleagues. Further, I would like to thank **Dr. Karsten Seeger** and **Dr. Alvaro Mallagaray** for their practical advice and our helpful discussions, and **Dr. Hanne Peters** for managing the laboratory and for an open door whenever problems arose.

I am grateful for the fruitful collaborations within the DFG research unit *ViroCarb*, especially with **Prof. Dr. Stefan Taube** in murine norovirus research, with **Dr. Charlotte Uetrecht** in the field of native mass spectrometry, and with **Jasmin Dülfer**, **Marit Lingemann**, and **Jan Knickmann** in their groups. Likewise, I appreciate the collaboration with **Dr. Matthias Stein** and **Eric Schulze** who provided molecular dynamics simulations of the norovirus capsid protein.

Finally, I wish to express my gratitude towards **Patrick Ogrissek**, **Georg Wallmann**, **Thorben Maaß**, and **Clara Feldmann**. Supervising their theses was an invaluable lesson for myself and the source of many fond memories. Their contributions to this thesis are summarized below in Tab. 1.

Table 1: Contributions from students to this thesis. Data shown in this thesis that was obtained as part of supervised Bachelor's and Master's theses or during the students' subsequent work as the author's research assistants (German: *"wissenschaftliche Hilfskraft"*) is listed here. A detailed description of the supervised tasks is given here and, in a shortened version, again in the respective figure and table legends.

student	contribution to	supervised task
Patrick Ogrissek	Fig. 4.4, Fig. 4.18, Fig. 4.31, Tab. 4.2	denaturation and refolding of isotope labeled P-domain from strain GII.4 VA387
	Fig. 4.5, Tab. 4.1	purification, denaturation, and refolding of isotope labeled N373D P-domain from strain GII.4 Saga; peak picking of spectra for measurement of ^{15}N relaxation data
	Fig. 4.8	biosynthesis, purification, denaturation, refolding, and chromatographic separation of protein species of isotope labeled P-domain from strain GII.4 Saga
Georg Wallmann	Fig. 4.4, Fig. 4.18, Tab. 4.2	biosynthesis, purification, denaturation, and refolding of isotope labeled P-domain from strain GII.4 MI001
Clara Feldmann	Fig. 4.20, Tab. 4.3	biosynthesis and purification of isotope labeled P-domain from strain GV MNV07; hydrophobic interaction chromatography experiments
	Fig. 4.22, Fig. 6.4	hydrophobic interaction chromatography experiments
Thorben Maaß	Tab. 4.4	STD NMR experiments with murine NoV strain CW1 with GalNAc, A-trisaccharide, and Forssman-trisaccharide

1. Abstract

Norovirus infections are a major cause for acute gastroenteritis and responsible for over 700 million illnesses each year worldwide. Like other non-enveloped viruses, noroviruses must interact with their environment through a proteinaceous shell that consists of 180 copies of the major capsid protein. One domain of this protein, the protruding domain in its functional, dimeric form, is responsible for cell surface attachment. Noroviruses from different genogroups are restricted in their host tropism. For example, some noroviruses only infect rodents but not humans. These murine noroviruses resemble their human counterparts in terms of capsid structure and are thought to recognize related attachment factors or ligands. Two classes of ligands have been described for both human and murine norovirus protruding domain proteins: bile acids and glycans. The biological function of the interactions with these ligands is poorly understood. Bile acids have been linked to increased viral infectivity and glycans from the histo-blood group antigen family serve as cell attachment factors for human noroviruses. Recently, it has been described that both murine and human noroviruses interact with sialic acid-containing glycans, e.g. gangliosides.

In my thesis, I have characterized murine and human norovirus capsid proteins using nuclear magnetic resonance spectroscopy to establish a comprehensive picture of their interactions with ligands in solution. Chemical shift perturbation experiments revealed that the affinities between human norovirus protruding domains and histo-blood group antigens are significantly lower than reported, with dissociation constants in the mM-range. Spontaneous deamidation of an asparagine residue has been discovered in a surface loop of GII.4 noroviruses, abrogating glycan binding. This post-translational transformation of the capsid of the predominant human noroviruses into a binding-incompetent form has implications for future studies of viral glycan recognition and vaccine development. Surprisingly and contrary to previous studies, neither human nor murine norovirus capsid proteins were found to bind sialic acid-bearing glycans such as gangliosides. In fact, murine norovirus protruding domains do not interact with any of the carbohydrates tested. These findings redefine the glycan binding code of noroviruses.

Another focus of my thesis is on the role of bile acids. A novel, low-affinity binding site for bile acids has been identified near the C-terminus of the protruding domains of GII.4 noroviruses. Bile acids were also found to fulfill an important function for murine noroviruses as their presence is required for protruding domain dimerization and stability. Binding is also accompanied by major changes in protein dynamics, possibly linked to viral immune evasion. These unique mechanisms highlight the importance of bile acids as essential cofactors for norovirus infections.

2. Zusammenfassung

Jährlich können weltweit über 700 Millionen Erkrankungsfälle auf Infektionen mit Noroviren zurückgeführt werden. Noroviren interagieren, wie andere nicht-behüllte Viren auch, mit ihrer Umgebung mittels einer Proteinhülle. Diese wird aus 180 Kopien des viralen Kapsidproteins gebildet. Eine Domäne dieses Proteins – die *protruding domain* oder P-Domäne – ist verantwortlich für die Anheftung des Virus an Zelloberflächen. Noroviren unterschiedlicher Genogruppen zeigen einen eingeschränkten Wirtstropismus. Beispielhaft zu nennen sind murine Noroviren, die nur Nagetiere aber keine Menschen infizieren, ihren humanen Gegenstücken in ihrer Kapsidstruktur jedoch stark ähneln. Es wird davon ausgegangen, dass die P-Domänen humaner wie auch muriner Noroviren ähnliche Liganden erkennen. Zwei Arten von Liganden, Gallensäuren und Glykane, wurden für beide Proteine beschrieben. Die biologische Bedeutung dieser Interaktionen ist weitestgehend unverstanden; Gallensäuren beeinflussen die virale Infektiosität während Glykane aus der Gruppe der Histo-Blutgruppenantigene als Anheftungsfaktoren für humane Noroviren dienen. Darüber hinaus wurden sialinsäurehaltige Glykane, wie z.B. Ganglioside, als Liganden von murinen und humanen Noroviren beschrieben.

Mithilfe der NMR-Spektroskopie habe ich in dieser Arbeit murine und humane Noroviruskapsidproteine charakterisiert, um ein umfassendes Bild ihrer Ligandeninteraktionen in Lösung zu erlangen. Protein-NMR-Experimente zeigten, dass die Affinitäten zwischen Histo-Blutgruppenantigenen und den P-Domänen humaner Noroviren erheblich niedriger sind als bisher angenommen. Die entdeckte spontane Deamidierung eines Asparagins auf der Oberfläche des Proteins führt darüber hinaus zu einem vollständigen Verlust der Fähigkeit zur Glykanerkennung. Diese posttranslationale Transformation des Kapsids humaner Noroviren in eine bindungsunfähige Form ist potentiell bedeutend für zukünftige Untersuchungen der viralen Glykanerkennung und die Impfstoffentwicklung. Im Gegensatz zu Erkenntnissen vorheriger Studien binden sialinsäurehaltige Glykane nicht an die P-Domäne muriner oder humaner Noroviren. Tatsächlich findet überhaupt keine Interaktionen zwischen murinen Noroviruskapsidproteinen und einer Reihe hier getesteter Glykane statt. Daraus folgt, dass bisherige Modelle zur Glykanspezifität der Noroviren grundsätzlich revidiert werden müssen.

Weiterhin wurde in dieser Arbeit eine neue Bindetasche für Gallensäuren in humanen GII.4 Noroviruskapsiden identifiziert. Es zeigte sich darüber hinaus, dass Gallensäuren essentiell für die Dimerisierung und Stabilisierung der P-Domäne muriner Noroviren sind. Ihre Bindung geht einher mit umfangreichen Änderungen der Proteindynamik, die in möglichem Zusammenhang zur Umgehung des Wirtsimmunsystems stehen. Diese neuartigen Mechanismen geben einen weiteren Hinweis auf die Bedeutung von Gallensäuren als Kofaktoren der Norovirusinfektion.

3. Introduction

3.1 Noroviruses

3.1.1 Disease & Epidemiology

Noroviruses (NoV) are non-enveloped, single-stranded RNA viruses and belong to the family of *Caliciviridae* (Vinje et al., 2019). An approximately 8 kB large RNA genome serves as mRNA template for six non-structural proteins, including the RNA-dependent RNA polymerase, in one open-reading frame (ORF) and the major structural capsid protein VP1 in another. A third ORF encodes the minor structural protein VP2 with a hypothesized role in the release of the genome upon infection of target cells (Conley et al., 2019). The major capsid protein VP1 is used for differentiation of members of the genus *Norovirus*: phylogenetic analysis based on the VP1 amino acid sequence leads to the clustering of NoV into ten genogroups (termed GI-GX) which are then subdivided into genotypes (e.g. GII.4) (Chhabra et al., 2019). Further differentiation is achieved by designation of NoV *strains*, named after the location of their first occurrence (e.g. GII.4 Sydney). NoV can infect a variety of species, with GV NoV infecting mice and GI, GII and GIV NoV infecting humans. Among human NoV, genotype GII.4 has been predominant for almost two decades (van Beek et al., 2018).

Infection with NoV is a leading cause for acute gastroenteritis. Commonly, symptoms in otherwise healthy adults are mostly diarrhea and vomiting (Robilotti et al., 2015). NoV gastroenteritis is generally of a short duration, and most patients recover after 96 h. However, complications – including fatal outcomes – may arise in elderly patients, immunocompromised hosts, or very young children. This amounts to an estimate of 700 million illnesses worldwide each year, including 70,000 deaths among children below an age of 5 years (Bartsch et al., 2016). In Germany, NoV gastroenteritis ranked second among the most common infectious diseases in 2017 (Hofmann et al., 2020). The large case number is thought to be responsible for a global annual economic burden of about $60 billion through health system and societal costs (Bartsch et al., 2016). NoV outbreaks are mostly food-borne or arise from contaminated water sources (Robilotti et al., 2015). Especially in living situations with closed-space environments – cruise-ships, schools, military sites – NoV are easily transmitted through the fecal-oral route. NoV particles are environmentally stable and exposure to a single particle is estimated to carry a 50 % risk of infection (Teunis et al., 2008).

Treatment of NoV infections is focused on reducing dehydration. No antiviral agents or vaccines are commercially available (Cortes-Penfield et al., 2017, Mattison et al., 2018). Vaccine development is impaired by viral antigenic diversity and short-term strain-specific natural immunity. Research in this

area is mostly focused on non-replicating virus-like particles. Traditional vaccines based on attenuated viruses could not be developed due to the lack of cell culture systems for human NoV until recently.

Human NoV have been successfully cultivated in B cells (Jones et al., 2014) and human intestinal enteroids (Ettayebi et al., 2016). The latter are derived from intestinal stem cells and, after differentiation into a multicellular epithelium, support both infection and replication of a variety of human NoV strains, including the predominant ones (Estes et al., 2019). The enteroid system is, however, experimentally challenging, expensive, and sustained passaging of NoV is impaired. The lack of suitable animal models is equally challenging for research in NoV biology (Todd and Tripp, 2019), although experiments with primates, gnotobiotic pigs, immunocompromised mice (Taube et al., 2013), and zebrafish (Van Dycke et al., 2019) have been reported.

Human NoV appear to primarily infect cells in the small intestine, giving rise to the gastrointestinal symptoms of infection. However, the actual cell type targeted by NoV remains elusive. Reports on immune cells as the target cells (Jones et al., 2014, 2015) are controversial and intestinal enterocytes have been proposed as the most likely alternative (Atmar et al., 2018). No proteinaceous cell surface receptor for human NoV is currently known and the viral cell attachment process remains poorly understood (cf. chapter 3.1.4).

3.1.2 The Norovirus Particle

All *Caliciviridae*, including NoV, are non-enveloped viruses, i.e. their proteinaceous viral capsid is not surrounded by an additional membrane layer. Thus, the viral capsid itself must fulfill a variety of functions in the viral infection cycle (Flint et al., 2003). First and foremost, capsids serve to protect the viral genome from its environment. Some capsids not only encapsulate but also organize the viral genome at the interior face. After infection, the attachment to cell surface receptors, and thereby in a broader sense the viral tropism itself, is mediated by the capsid. Finally, the capsid must allow for the controlled release of the viral genome at the appropriate point in time.

The infectious NoV virion is thought to contain only three different elements: the RNA genome itself, linked to the VPg protein, and a small number of copies of the minor structural protein VP2, hypothesized to be important in the release of the viral genome (Sosnovtsev and Green, 2000, Conley et al., 2019). The capsid itself is formed by assembly of 180 copies of the 60 kDa major structural protein VP1 as shown in Fig. 3.1a (Vinje et al., 2019). The VP1 protein is comprised of two domains, separated by a flexible hinge (Prasad et al., 1999). The interior *shell-domain* (S-domain) is responsible for the overall icosahedral symmetry of the viral capsid by making intersubunit contacts with other S-domains. Three connected S-domains form one face of the icosahedron and, thereby, the asymmetrical unit that is repeated 60-times. Accordingly, the NoV capsid has a triangulation number of $T = 3$. Of note, a polymorphism of NoV capsids has recently been observed indicating that NoV capsids might also adopt $T = 1$ or $T = 4$ capsid symmetries (Jung et al., 2019).

The second domain of the VP1 protein is the *protruding domain* (P-domain). Two P-domains interact with each other and form 90 dimeric protrusions on top of the capsid's shell (Fig. 3.1b). Due to the icosahedral nature of the capsid, two different kinds of P-domain interaction can occur in the con-

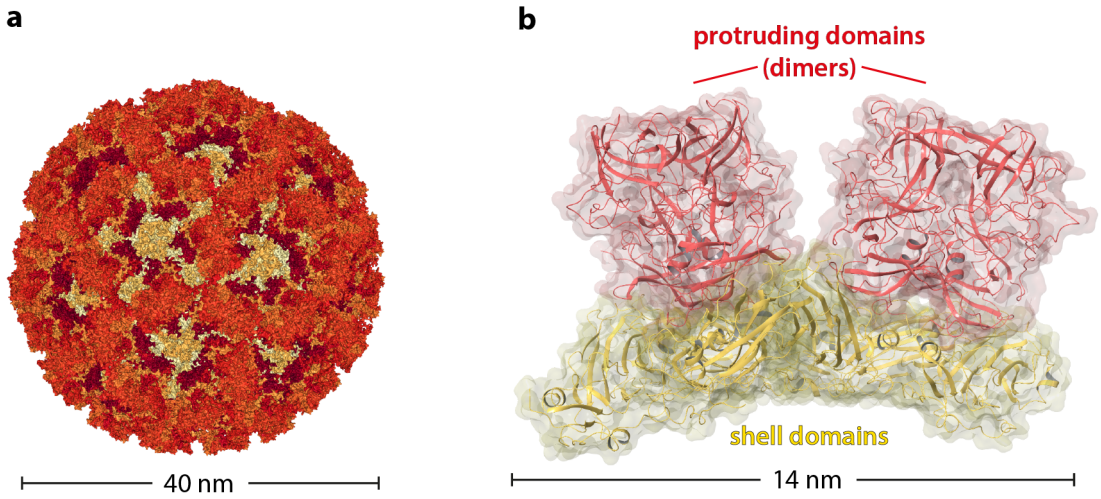


Figure 3.1: The norovirus capsid. Norovirus particles are assembled from 180 copies of the major structural protein VP1 and adopt a T=3 icosahedral symmetry (**a**, pdb 1IHM). The viral capsid presented in **a** is color-coded according to the sequence of the VP1 protein: the shell domain, forming a hollow sphere, appears in yellow and the N-terminal protruding domain in orange and red. An enlarged part of the capsid is shown in **b**. The VP1 shell domains (yellow) are arranged as a continuous surface that encapsulates the viral genome. The protruding domains are organized as homodimers, two of which are depicted here (red). Shell and protruding domains are separated by a flexible linker peptide. Here, both domains adopt a contracted conformation but other structures show protruding domains elevated from the shell domain.

text of the assembled particle (termed AB or CC dimers). In most NoV strains, the P-domains appear to rest on the shell domains. However, in some NoV the P-domain has been found to be ”floating” above the shell in an extended conformation (Smith and Smith, 2019). The biological function of this apparent flexibility of the NoV capsid remains unknown. Interactions of the capsid with the environment occur at the outward facing part of the P-domain dimers. Interactions with cell surface proteins and carbohydrates, responsible for cell attachment, are subject of chapters 3.1.3 - 3.1.4. The P-domain also harbors binding sites for antibodies of the host immune system and is, therefore, subject to high selection pressure that drives the VP1 sequence divergence and the emergence of new NoV strains (Mallory et al., 2019).

Several studies have addressed the inherent flexibility of the NoV P-domain. Access to some antibody epitopes was found to be limited by temperature-dependent protein dynamics (Lindesmith et al., 2018). Antibody binding might also be subject to allosteric regulation within the P-domain of human GII.4 NoV (Lindesmith et al., 2014) and murine NoV (Kolawole et al., 2017). Several surface loops on the apex of the murine NoV P-domain can adopt both ‘open’ and ‘closed’ conformations that were linked to the occlusion of another antibody binding site (Taube et al., 2010, Sherman et al., 2019).

Empty virus-like particles can be produced cell-free (Sheng et al., 2017), in insect cells (Jiang et al., 1992), or in plant cells (Diamos and Mason, 2018) by expression of the VP1 protein and subsequent self-assembly. These particles appear to maintain the native structure, antigenicity and carbohydrate binding activity of natural, infectious NoV (Pogan et al., 2018). The P-domain of human NoV can be synthesized recombinantly in *Escherichia coli* and appears to readily adopt its native, dimeric structure (68 kDa) without being attached to the S-domain (Tan et al., 2004). Again, both antibody and carbohydrate binding appear to be unaffected by bacterial expression of the soluble P-domain.

3.1.3 Murine Noroviruses

Murine NoV have been proposed as a model system to study NoV biology in general (Wobus et al., 2006). In the time period between their discovery (Karst et al., 2003, Wobus et al., 2004) and the development of the human intestinal enteroid system for human NoV(Estes et al., 2019), murine NoV were the only members of the genus *Norovirus* that readily replicated in cell culture. Additionally, the ability to use laboratory mice as an animal model is a unique advantage of murine NoV. The murine NoV genome is similar to that of other NoV, including the major capsid protein VP1, but has an additional ORF encoding a virulence factor (Vinje et al., 2019). Murine NoV can infect both immunocompromised and wild-type mice but remain fully asymptomatic in the latter (Baldridge et al., 2016). The cell tropism of murine NoV appears to differ from human NoV as they target intestinal macrophages, B cells, and dendritic cells and not enterocytes (Wobus, 2018).

Murine NoV are less genetically diverse than human NoV and comprise a single genogroup. The murine NoV does not appear to be able to overcome the species-barrier and has not been found in humans (Smith et al., 2012). Notably, a protein cell-surface receptor has been found that is essential for attachment and cell entry of murine NoV (Haga et al., 2016, Orchard et al., 2016). The *CD300lf* protein belongs to a large family of transmembrane receptors associated with regulation of the immune system and binding of lipids (Borrego, 2013). The binding site of the CD300lf protein was found on the top side of the murine NoV protruding domain (Kilic et al., 2018, Nelson et al., 2018). The murine NoV particle engages the same binding pocket as CD300lf's natural phospholipid ligand, described as an act of *viral mimicry*. Multivalency increases the strength of the capsid-receptor interaction as multiple CD300lf molecules are able to bind to the viral capsid simultaneously (Sherman et al., 2019). However, CD300lf appears not to be relevant for infections with human NoV. Two distinct binding sites for bivalent metal ions (Mg^{2+}, Ca^{2+}) have been identified on the murine NoV P-domain and their occupation was associated with increased viral infectivity (Nelson et al., 2018).

3.1.4 The Role of Glycans in Norovirus Infections

Glycans are currently described as *attachment factors*, but not receptors, of NoV (Tan and Jiang, 2014, Graziano et al., 2019). Attachment factors are considered non-essential, but beneficial for the infection event and concentrate viruses on cell surfaces without triggering the actual entry process. As required for a non-enveloped virus, NoV attachment is mediated by the viral capsid through protein-carbohydrate interactions. Among these glycan interactions, the binding of certain carbohydrates termed *histo-blood group antigens* (HBGAs) is best characterized. HBGAs are synthesized by stepwise, enzymatic linkage of monosaccharide subunits (Nordgren and Svensson, 2019). The HBGA family is comprised of both the ABO and Lewis carbohydrates, with an α(1,2)-fucosyltransferase (FUT2) or an α(1,3/4)-fucosyltransferase (FUT3) responsible for the attachment of the respective characteristic fucose molecules. Expression of the active FUT2 glycosyltransferase results in the formation of the H antigen core structure (blood group O) that can be further elongated to yield blood group A or B antigens. HBGAs are abundant on epithelial cell surfaces in the form of glycolipids or glycoproteins, especially in the gastrointestinal mucosa, but can also be secreted as soluble carbohydrates into body

fluids (Heggelund et al., 2017).

Biological evidence for the importance of HBGAs in human NoV infections comes from a variety of findings. In challenge studies with human volunteers, the susceptibility towards infection with a GI.1 NoV strain was linked to the ABO blood type: individuals with blood type B were found to have a decreased risk in infection (Hutson et al., 2002). In addition, humans that are homozygous recessive for the functional FUT2, termed *non-secretors*, are fully resistant against GI.1 and GII.4 NoV infections (Lindesmith et al., 2003, Frenck et al., 2012). However, the protection of non-secretors appears to be strongly dependent on the NoV genotype: susceptibility for GI.3 and GII.2 NoV strains did not depend on the secretor status (Lindesmith et al., 2005, Nordgren et al., 2010). The presence of antibodies blocking the interaction of the NoV capsid with HBGAs is correlated with protection from NoV infection (Cortes-Penfield et al., 2017). In the human intestinal enteroid cell culture, replication of GII.4 NoV only occurred in cells with a functional FUT2 (Estes et al., 2019).

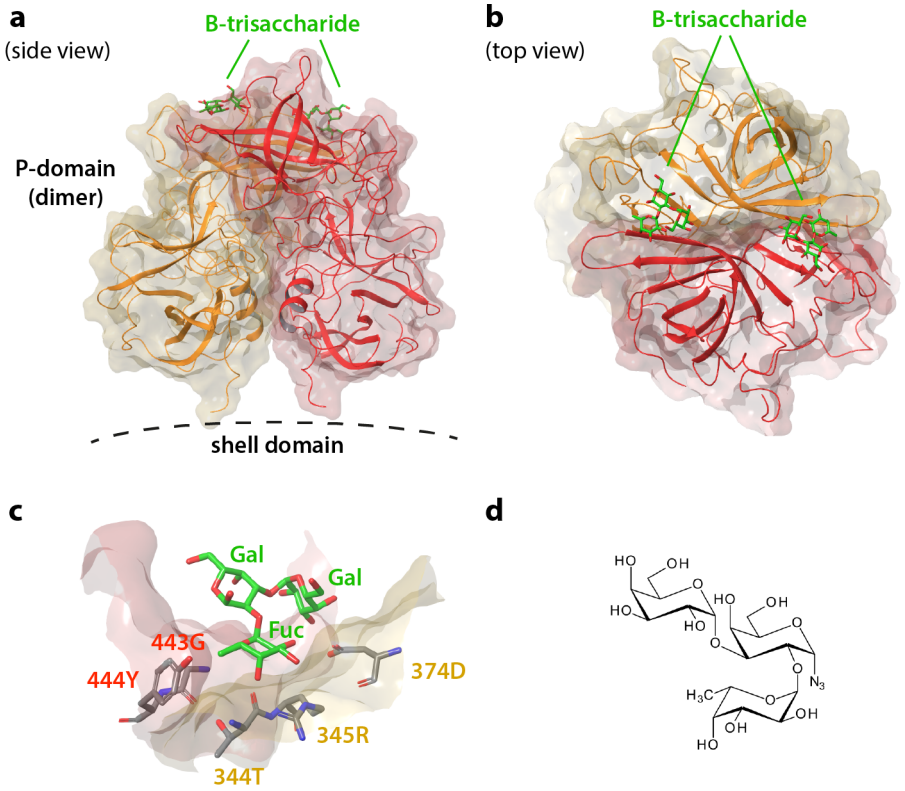


Figure 3.2: The glycan binding pocket of the GII.4 NoV protruding domain. A large number of human NoV strains interacts with carbohydrates from the histo blood group antigen (HBGA) family. Here, the HBGA binding pocket of GII.4 NoV is shown with the representative GII.4 Saga strain in complex with the blood group B-trisaccharide, shown in green (pdb 4X06). The two glycan binding sites are located at the outward facing part of the protruding domain homodimer (**a**, monomers depicted in orange and red, respectively). The N-terminal part of the proteins is shown at the bottom and would connect the protruding domain with the inner shell domain in the full virus capsid. HBGAs bind at the interface of both monomers (**b**) and have essential interactions with amino acids from both protein chains (**c**). The amino acids that comprise the canonical HBGA binding pocket are indicated with the color of their respective monomer chain. Most interactions are made with the α(1,2)-linked fucose moiety that is present in all HBGAs. This interaction is essential for HBGA recognition and makes fucose the minimal binding motif that is recognized by GII.4 NoV. The structure of the B-trisaccharide used in this thesis for interaction studies is shown in **d**, with an unnatural α-linked azide group in the second galactose preventing mutarotation.

Molecular details of the interactions between the viral capsid and HBGAs have been obtained by co-crystallization of HBGAs and recombinant P-domains (Tan and Jiang, 2011). Notably, analogous HBGA binding sites exist in GI and GII NoV as a result of convergent evolution but have distinct locations and amino acid compositions (Tan et al., 2009). Additional HBGA binding sites have been identified in some rare NoV strains (Liu et al., 2015, Qian et al., 2019). In GII.4 NoV, the HBGA binding site is located at the outward-facing part of the P-domain (Fig. 3.2). A P-domain dimer can engage two HBGA molecules with contributions from both monomers to each binding pocket. A variety of different HBGAs can access these binding pockets, as most interactions occur with the L-fucose moiety that is present in all HBGAs. This has lead to the identification of α(1,2)-linked L-fucose as the essential, minimal binding motif (Fiege et al., 2012).

It remains unclear whether soluble or surface-anchored HBGAs are responsible for the viral attachment process in vivo. In glycosphingolipids, HBGAs are linked to a ceramide that serves as a membrane anchor. In vitro, HBGAs are restricted in their conformational flexibility in the context of such glycosphingolipids (Nasir et al., 2017). Accordingly, the viral capsid's engagement of membrane embedded and soluble HBGAs might differ. As one NoV particle has a total of 180 binding pockets for HBGAs, multivalent interactions with several HBGA molecules can compensate for low intrinsic carbohydrate binding affinities and lead to firm surface attachment (Parveen et al., 2019). Another layer of complexity is added by the observation that NoV particles can interact with enteric bacteria that present HBGA-like carbohydrates on their outer surfaces (Jones et al., 2014, Li et al., 2015). Recently, it was proposed that the gastrointestinal microbiome of the host is critically involved in NoV infections (Walker and Baldridge, 2019, Karst and Wobus, 2015). Human milk contains a variety of soluble carbohydrates, some of which structurally mimic HBGAs including their essential α(1,2)-linked fucose moiety (Taube et al., 2018). Milk oligosaccharides have been shown to bind to the HBGA binding pocket in the NoV P-domain and might have antiviral properties (Koromyslova et al., 2017).

Capsid-glycan affinities have been investigated with a variety of biophysical techniques, including native mass spectrometry (Han et al., 2013, 2015), nuclear magnetic resonance spectroscopy (Hansman et al., 2012, Mallagaray et al., 2015, 2017), surface plasmon resonance (de Rougemont et al., 2011, Wegener et al., 2017), and a quartz crystal microbalance (Parveen et al., 2019). The monovalent protein-glycan interaction is generally considered to be weak. However, actual binding energies obtained in these studies vary considerably. As an example, the dissociation constant of the protein complex with the blood group B trisaccharide has been reported to be both 40 µM and 1.25 mM in GII.4 NoV (Wegener et al., 2017, Han et al., 2015).

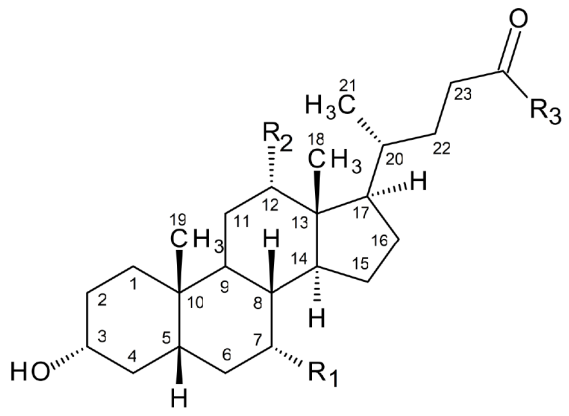
The glycan specificity of NoV can vary substantially between genogroups but also between related genotypes (de Rougemont et al., 2011). For instance, GIII bovine NoV do not bind HBGAs but recognize the αGal (Galili)-epitope that is not present in humans (Zakhour et al., 2009). Some rare human NoV were found to interact with terminal galactose moieties or heparan sulfates (Cong et al., 2019, Tamura et al., 2004). Sialic acid-containing glycans of the ganglioside family have been described as ligands for both human and murine NoV P-domains. Sialic acids are commonly found on cell surfaces and are being recognized as the most common attachment factor for pathogens, especially viruses

(Lehmann et al., 2006). Three other members of the *Caliciviridae* family are known to use sialic acid molecules for cell attachment (Kim et al., 2014, Stuart and Brown, 2007, Tan et al., 2015). Reports on interactions of NoV with sialic acids are contradictory and no structural model of the binding pocket is available yet. Two studies showed that 3'-sialyllactose, the oligosaccharide portion of the GM3 ganglioside, did *not* bind to human GII.4 NoV P-domains using nuclear magnetic resonance spectroscopy and native mass spectrometry (Fiege et al., 2012, Han et al., 2013). However, later studies did report binding using very similar experimental setups (Wegener et al., 2017, Han et al., 2014). As of now, the discrepancy between these experimental results has not been resolved. Of note, the recognition of sialic acid-containing glycans from the *Lewis* family by human NoV capsids is well established. However, this interaction does not involve the sialic acid molecule itself as it sticks out of the binding pocket (Singh et al., 2015, Fiege et al., 2012).

Similarly, there is cell culture-based evidence that murine NoV P-domains engage gangliosides as well (Taube et al., 2009, 2012). The infectivity of some murine NoV strains was found to be sensitive towards enzymatic cleavage of sialic acids from cell surfaces. Cell treatment with sialic-acid binding antibodies or lectins reduced attachment of murine NoV. As other, highly similar, murine NoV strains did not show this behavior, it was concluded that a binding pocket for sialic acids must exist on the top side of the P-domain based on an amino acid sequence comparison. This putative binding site would be located in a similar position as the HBGA binding pocket in structurally homologous human NoV P-domains. HBGAs were found to be not relevant in murine NoV infections (Nelson et al., 2018).

3.1.5 The Role of Bile Acids in Norovirus Infections

Bile acids are cholesterol-derived, amphipathic molecules that fulfill an essential role in the digestion of dietary lipids in the gastrointestinal tract (Payne et al., 2008). Synthesis of *primary bile acids* occurs in the liver where the hydrophobic steroid nucleus is hydroxylated and then conjugated with hydrophilic glycine or taurine head groups (Fig. 3.3).



	CA	DCA	CDCA	GCDCA
R_1	OH	H	OH	OH
R_2	OH	OH	H	H
R_3	OH	OH	OH	NH-$^{25}CH_2$-COOH

Figure 3.3: Bile acid structure. All human bile acids share the same steroid ring system but differ in three positions $R_{1\text{-}3}$. Bile acids are hydroxylated to different degrees, with cholic acid (CA) carrying the maximum of three hydroxyl groups at positions 3, 7 and 12. Hydroxylation occurs only on one side of the rigid steroid rings, producing a more hydrophilic face in all bile acid molecules. Other bile acids such as deoxycholic acid (DCA) and chenodeoxycholic acid (CDCA) are only hydroxylated twice and, therefore, less water-soluble. Glycochenodeoxycholic acid (GCDCA) is conjugated with a glycine molecule as a hydrophilic head group.

Hydroxylation occurs only on one side of the rigid steroid ring system, leading to the formation of a hydrophilic and a hydrophobic face. All bile acids are weak acids and are prone to micelle-formation or precipitation in aqueous solutions depending on environmental conditions such as pH or ionic strength (Hofmann and Mysels, 1992). Major primary bile acids in humans are cholic acid (CA) and chenodeoxycholic acid (CDCA). Conjugated bile acids are secreted by the gall bladder into the small intestine. There, their primary function is the solubilization of dietary lipids in mixed micelles. After passage through the small intestine, bile acids are deconjugated, converted into *secondary bile acids* as e.g. deoxycholic acid (DCA), reabsorbed, and recycled in the liver. Bile acid concentrations in the small intestine are, depending on the dietary status, in the mM-range (McLeod and Wiggins, 1968). In addition to their role in digestion, bile acids have more recently been described as ligands for membrane and nuclear receptors with complex roles in inflammation and metabolic processes (Chiang, 2013).

With the emergence of an *in vitro* cultivation system for human NoV, it was reported that bile is a necessary cofactor for successful replication of some NoV in human intestinal enteroid cell culture (Ettayebi et al., 2016). Bile was found to be essential for replication of GI.1, GII.3 and GII.17 NoV. Bile was, however, not required for replication of the epidemiologically predominant GII.4 strains, but did increase their replication efficiency. A similar role of bile was described previously for another member of the *Caliciviridae* family, the porcine enteric calicivirus (Chang et al., 2004). It remains unclear whether the requirement of bile for NoV replication is due to an effect on the infected cells or the NoV particle itself. Bile acids were found to mediate GII.3 NoV replication by multiple effects on a cellular level, including stimulation of a membrane receptor and changes in cell membrane composition (Murakami et al., 2020). A link between the cellular immune response and the processing of bile acids by the gut microbiome has been proposed to be relevant for murine NoV infections (Grau et al., 2020).

Complementary to these findings on a cellular level, bile acids were also shown to be ligands of the NoV capsid itself (Fig. 3.4).

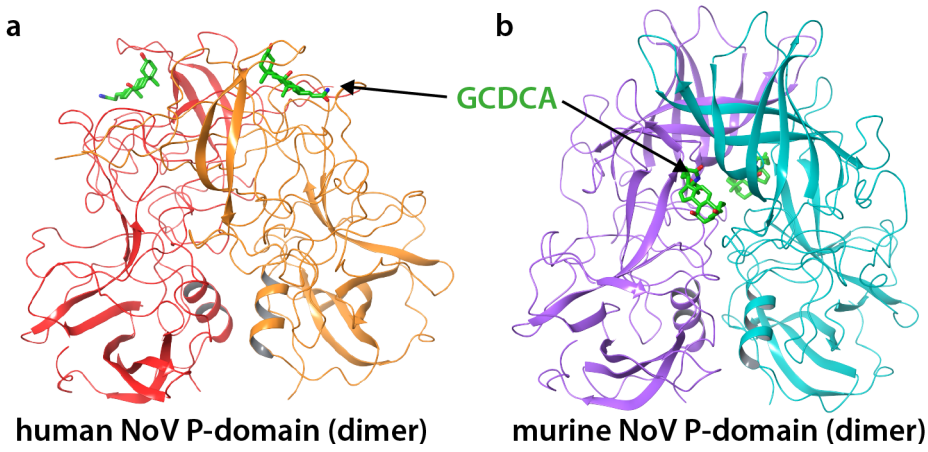


Figure 3.4: Bile acid binding pockets in rare human NoV and murine NoV P-domains. Human NoV P-domains of the rare GII.10 Vietnam strain have a binding pocket for glycochenodeoxycholic acid (GCDCA, green) on top of the P-domain (**a**, pdb 6GW1). Two symmetrical binding sites have been identified in close proximity to the glycan binding pockets (cf. Fig. 3.2). The respective P-domain monomers are depicted in orange and red. In contrast, murine NoV P-domains possess a structurally distinct binding pocket for GCDCA (**b**, pdb 6E47). Here, the binding sites are located between the two P-domain monomers (shown in purple and blue). Each GCDCA molecule makes essential interactions with amino acids from both monomers.

A high-affinity binding site for bile acids was found in rare human NoV strains near the glycan binding pocket but was absent in the prevalent GI.1, GII.4 and GII.17 strains (Kilic et al., 2019). In murine NoV, another binding pocket for glycochenodeoxycholic acid (GCDCA) was identified by both crystallography and cryo electron microscopy (Nelson et al., 2018, Sherman et al., 2019). This binding pocket is located at the interface of two monomers in the P-domain dimer (Fig. 3.4b). GCDCA was reported to increase the infectivity of murine NoV through enhancement of the affinity for the proteinaceous receptor CD300lf. Additionally, bile acids were shown to cause dramatic morphological changes in assembled murine NoV particles (Sherman et al., 2019). In the absence of bile acids, P-domains were elevated from the S-domain but collapsed onto the S-domain upon bile acid addition. Minor structural rearrangements within surface-exposed loops of the murine NoV P-domain are thought to be important for evasion of the host immune system.

3.2 Asparagine Deamidation in Proteins

Asparagine (Asn) residues in proteins and peptides can undergo a spontaneous post-translational cyclization reaction and subsequently form either aspartate (Asp) or iso-aspartate (isoAsp) (Müller, 2018). Cyclization occurs via a nucleophilic attack of a backbone amide nitrogen on a favorably oriented carbonyl group of a preceding Asn residue (Fig. 3.5). The succinimide intermediate undergoes rapid hydrolysis which results in the formation of isoAsp or Asp, typically with a ratio of 3:1 in model peptides (Nowak et al., 2018). This process increases a protein's negative charge and, in the case of isoAsp, introduces an isopeptide bond in the protein backbone. The deamidation of Asn is irreversible, whereas Asp and isoAsp can interconvert in an equilibrium reaction. In addition, racemization can occur, leading to the formation of unnatural, D-configurated reaction products (Radkiewicz et al., 1996).

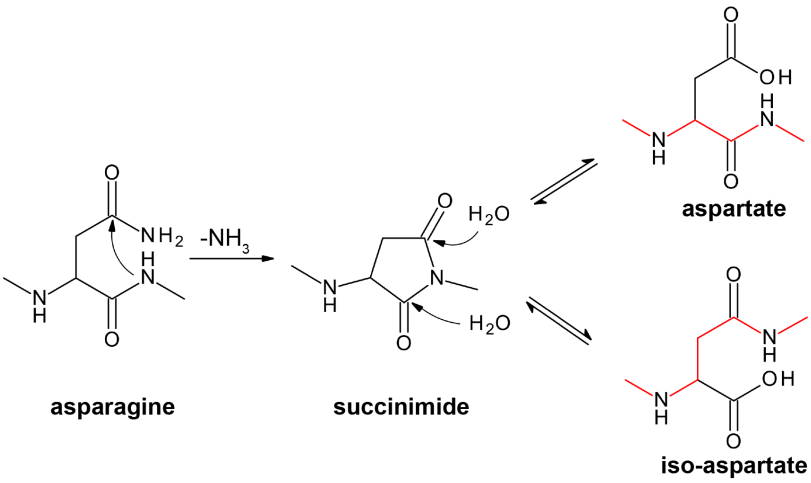


Figure 3.5: Deamidation of asparagine residues in proteins. A backbone amide nitrogen in a polypeptide chain can, under favorable circumstances, perform a nucleophilic attack on the C_γ atom of a preceding asparagine side chain, accompanied by the irreversible loss of ammonia. The cyclic succinimide intermediate is unstable and undergoes hydrolysis. Depending on the position of the water attack, one of two possible products is formed: either aspartate or iso-aspartate. In case of iso-aspartate formation, the backbone of the polypeptide chain (indicated by red bonds) now incorporates an additional carbon atom – the C_β of the former Asn side chain. The ring-opening reaction is reversible and, consequently, aspartate and iso-aspartate are in equilibrium with iso-aspartate being the predominant product.

Deamidation reactions have mostly been described in the context of aging or degraded proteins. The most eminent examples of such a loss-of-function are found in therapeutic monoclonal antibodies that can potentially lose antigen binding capabilities (Harris et al., 2001, Phillips et al., 2017, Nowak et al., 2018). On the contrary, a gain-of-function from Asn deamidation has rarely been observed with one example being the activation of a fibronectin-integrin binding site (Corti and Curnis, 2011). The rate of the deamidation reaction can vary dramatically depending on the Asn microenvironment. In order for the nucleophilic attack to occur, the Asn side chain must adopt a reactive conformation that minimizes the distance between nucleophile and electrophile (Plotnikov et al., 2017). Accordingly, the reaction rate depends on the conformational flexibility in the context of the polypeptide chain. This led first to the development of sequence-based prediction rules for deamidation that were mostly derived from model peptides (Robinson and Robinson, 2001a, Lorenzo et al., 2015). Attempts to account for the influence of a protein's tertiary structure resulted in structure-based algorithms (Jia and Sun, 2017, Sydow et al., 2014), supported by molecular and quantum mechanics approaches (Plotnikov et al., 2017). Additionally, deamidation rates are also influenced by environmental conditions such as ionic strength, pH, or temperature (Robinson and Robinson, 2001b, Stratton et al., 2001). Considering the wide range of the reaction's accessible time scales, it has been hypothesized that Asn deamidation can serve as a "molecular clock" in a variety of cellular processes (Robinson and Robinson, 2004).

Detection of a specific deamidation event in a given protein remains analytically challenging. Mass spectrometry allows for detection of the mass difference between Asn and isoAsp/Asp but cannot be used routinely for differentiation between both reaction products (Hao et al., 2017). Liquid chromatography techniques can be used for separation of the involved molecules based on their charge difference (Harris et al., 2001, Sandra et al., 2014). Nuclear magnetic resonance (NMR) spectroscopy has been used to characterize the properties of the respective molecular species, including the short-lived succinimide (Grassi et al., 2017, Tugarinov et al., 2002, Hagan et al., 2010).

3.3 Biomolecular Nuclear Magnetic Resonance (NMR) Spectroscopy

3.3.1 The Chemical Shift

The chemical shift is the fundamental physical quantity of nuclear magnetic resonance (NMR) spectroscopy. It originates in the property of certain nuclei to have a non-zero spin angular momentum and, thus, a magnetic moment. In an external magnetic field, the different spin states adopt different energies and corresponding energy levels are populated differently, giving rise to an observable net magnetization vector (Keeler, 2010). The energy difference between these spin states can be easily expressed in terms of a frequency, the *Larmor frequency*. In classical terms, this corresponds to the precession frequency the net magnetization vector adopts when it is moved away from the static magnetic field axis. Because of the spins' angular momentum, the net magnetization vector does not simply realign with the external field after this perturbation. Instead, the magnetic field imparts a torque on the tilted magnetization vector, causing a precession movement (Neuhaus and Williamson, 2000). In NMR spectroscopy, manipulation of the macroscopic magnetization vector is achieved by application

of short radio-frequency pulses perpendicular to the static magnetic field.

Circulations of surrounding electrons induce a secondary magnetic field that can exert a shielding effect on a given nucleus, modifying its precession frequency depending on its electronic environment (Facelli, 2011). This modified frequency, in relation to a reference frequency and expressed in a field-independent unit, is the *chemical shift*. Magnetic shielding is, among others, influenced by diamagnetic contributions of spherical electron clouds and paramagnetic effects from non-spherical distributions, e.g. of p-electrons. In a biological molecule such as a protein, ^{1}H, ^{15}N, and ^{13}C chemical shifts are the most relevant. An accurate *ab-initio* calculation of shielding tensors and the accurate prediction of chemical shifts remains challenging even for small molecules and is an unsolved problem for more complex entities such as proteins (Krivdin, 2017). However, with growing databases of protein structure models linked to experimental NMR chemical shift data, empirical prediction methods have improved in quality but are far from allowing a structure-based prediction of an NMR spectrum (Shen and Bax, 2010). Still, protein chemical shift data can report reliably on structural features of a protein as typical electronic arrangements are imposed upon a given nucleus by the geometric requirements of a folded protein (Saitô et al., 2010). This is why backbone and some side chain torsion angles – and thus the secondary structure – can be accurately predicted from chemical shift data (Shen and Bax, 2013). This empirical knowledge has led to attempts for *de novo* predictions of entire protein structures based on chemical shifts (Shen et al., 2008, Berjanskii et al., 2015, Shen and Bax, 2015).

Not only its geometry but also environmental effects can have pronounced, although poorly understood, effects on a protein's chemical shifts. For example, temperature (Tomlinson and Williamson, 2012), solvent conditions (e.g. salt content) (An et al., 2014), and single ionization events (e.g. through pH changes) (Platzer et al., 2014) all can cause unanticipated chemical shift changes that impair chemical shift predictions.

Protein chemical shift data is most useful for the characterization of protein-ligand interactions. Binding ligand molecules influence the electronic environment of the protein's NMR-active nuclei that are close to the respective binding site. If the identity of an NMR signal is known, i.e. the signal has been assigned to the nucleus that is causing it (cf. chapter 3.3.4), a ligand titration can reveal the binding epitope on the protein surface (Williamson, 2013). Experimental restraints of this type have been used for NMR-driven docking procedures to obtain structural models of protein-ligand complexes (van Zundert et al., 2016). Importantly, quantitative data about binding affinity (i.e. K_D values) can be derived from ligand titration experiments (Markin and Spyracopoulos, 2012). These *chemical shift perturbation* (CSP) experiments are currently the only biophysical technique that is capable of simultaneously characterizing binding sites with atomic resolution and obtaining quantitative binding data. Depending on the time scale of exchange between bound and unbound states and the chemical shift difference associated with these states, ligand binding can cause different spectral effects. In the *slow exchange* regime, the exchange rate constant k_{ex} is much smaller than the chemical shift difference, resulting in the observation of two separate NMR signals – one for each state. The intensity of each signal reports on the population of the respective state in the protein ensemble. Under *fast exchange* conditions, only one signal with an intermediate chemical shift can be observed because k_{ex} is greater

than the chemical shift difference. Its position is determined by the population-weighted average of bound and unbound states. Consequently, the observed shift reports on the fractional occupation of the binding pocket. Assuming a two-state binding model, the signal is expected to shift along a straight line connecting the signals belonging to the apo and saturated state depending on the ligand concentration. This relationship can be used to obtain binding isotherms by observation of the change in signal position (the CSP) in a series of ligand titration steps. If the exchange rate is of similar magnitude as the chemical shift difference (*intermediate exchange*), signals become broadened, sometimes even beyond detection.

It has been noted that ligand binding not only causes CSPs in spatial proximity to the binding pocket but potentially also in distant regions of the protein (Boulton et al., 2015). These long-range effects are poorly understood but are thought to encode, sometimes subtle, allosteric effects (Aoto et al., 2016).

3.3.2 Relaxation & Molecular Motion

Relaxation describes the process by which a spin system returns to its equilibrium state, i.e. the net transverse magnetization (perpendicular to the external field) is lost and longitudinal magnetization assumes its thermal equilibrium value again after an external perturbation (Charlier et al., 2016). The time constants T_2 and T_1 give a phenomenological description of the time-dependency of the two respective processes.

On a microscopic level, a spin ensemble can exchange energy with the solution environment through local magnetic fields (Keeler, 2010). Local fields originate from a variety of mechanisms and are generally short-ranged and weaker than the external magnetic field but can still interact with a given spin's magnetic moment. As a consequence, not all spins of an ensemble are experiencing the same local fields at the same time and, therefore, are subject to random reorientations that finally lead to a loss in net magnetization. In semi-classical relaxation theory, the entirety of a spin ensemble's local fields are described with a continuous function, the *spectral density function* (Luginbühl and Wüthrich, 2002). A prerequisite for relaxation due to interactions with a local field is that the local field must oscillate with a similar frequency as that of a given spin's precession movement. The most important source of these field fluctuations in solution is thermal, and thus random, rotational movement characterized by the rotational correlation time τ_c. The availability of random field fluctuations of a certain frequency in the ensemble that are able to relax a perturbed nuclear spin is described by the spectral density function. In flexible biomolecules such as proteins, another source of field fluctuations becomes important: intramolecular protein dynamics, i.e. the interconversion between structurally distinct states (Jarymowycz and Stone, 2006). This constitutes a fundamental link between a protein's internal motions and its NMR-observable relaxation properties.

Two physical phenomena are the most important origins of local magnetic fields. As detailed in chapter 3.3.1, asymmetrically distributed electrons with their own magnetic moments induce a magnetic field at the position of the nucleus (Saitô et al., 2010). The asymmetry causes the local magnetic field to be orientation-dependent with regard to the external field. As a consequence, the chemical

shift itself depends on its spin's orientation but is ensemble-averaged in solution due to thermal motion. In biomolecules, this *chemical shift anisotropy* influences mostly ^{15}N and ^{13}C nuclei. *Dipolar interactions* occurring between NMR-active nuclei are another source of local fields. This type of interaction has a moderate distance dependency of r^{-3} but several spins in close spatial proximity can act simultaneously on one receiving spin, leading to efficient relaxation in dense spin networks.

For large, slowly tumbling molecules, low frequency fluctuations dominate the spectral density function (Neuhaus and Williamson, 2000). This leads to fast transverse relaxation as motions of this frequency are efficient in causing dephasing of transverse magnetization (Luginbühl and Wüthrich, 2002). The respective time constant T_2 decreases linearly with an increasing rotational correlation time and, thus, approximately with increasing molecular weight. As T_2 is inherently linked with the line width of an NMR signal, large proteins have broad signals, challenging to detect. As T_1 relaxation depends mostly on local field contributions at other frequencies, the ratio of T_1/T_2 for a given signal reports semi-quantitatively on the presence of intramolecular motions on the ps to ns time scale (Jarymowycz and Stone, 2006).

Under certain circumstances, dipolar interactions and chemical shift anisotropy are not independent from each other as their time-dependent field fluctuations are fundamentally correlated via the rotational movement of the molecule (Keeler, 2010). This is the case for ^{1}H-^{15}N spin pairs as found in a protein backbone's amide groups in which the ^{15}N nucleus is mostly relaxed through its own chemical shift anisotropy and the dipolar interaction with the proton. This *cross-correlation* between both types of field fluctuations can lead to both destructive and constructive interference effects. The ^{15}N spin undergoes *scalar coupling* with the attached proton and, therefore, its NMR signal is split into two lines – one associated with each spin state of the proton. As the spin state of the proton determines the direction of the local magnetic field it causes, one half of the ^{15}N doublet experiences constructive interference and the other one destructive interference. Consequently, one NMR signal has a decreased relaxation rate as both relaxation mechanisms partially cancel out each other which results in a sharpened NMR signal (Fernández and Wider, 2003). Similar mechanisms are operative on the doublet signal of the attached proton. That is why in the two-dimensional spectrum of a ^{1}H-^{15}N spin pair, one of the four signal components has a very small linewidth whereas the others are severely broadened. Selection of only the sharp signal component allows for the acquisition of *transverse-relaxation optimized* heteronuclear experiments (TROSY HSQC) in which all amide resonances give only one very sharp signal (Pervushin et al., 1997). Together with new isotopic labeling strategies (cf. chapter 3.3.3) and advances in spectrometer technology, TROSY-type NMR experiments have enabled the study of high-molecular weight proteins (Wider and Wüthrich, 1999).

3.3.3 Isotopic Labeling

Protein NMR spectroscopy is mostly based on two-dimensional, heteronuclear experiments. Consequently, the presence of at least two different NMR-active spin 1/2 nuclei is required. Whereas protons (^{1}H) are found ubiquitously in a protein, NMR-active ^{13}C or ^{15}N nuclei have to be introduced artificially into a protein during its biosynthesis as their natural abundance is typically too low for detection

(Haynes, 2011) (1.1 % and 0.4 %, respectively).

The majority of protein NMR studies relies on recombinant bacterial expression of the target protein. In these systems, isotopic labeling can be achieved easily by using bacterial growth media that contain only one type of nutrient molecule as a carbon and nitrogen source, most often commercially available ^{13}C-glucose or ^{15}N-ammonium chloride (Tugarinov et al., 2006). ^{13}C and ^{15}N are then incorporated into every cellular molecule, including the over-expressed target proteins. In cases, where bacterial expression is not possible, e.g. for reasons of toxicity or the requirement of post-translational modifications, more elaborate labeling techniques in other organisms have to be used at the cost of much lower yields (Goto and Kay, 2000).

As outlined in chapter 3.3.2, strong dipolar couplings to surrounding protons leads to fast relaxation of perturbed nuclei, severely impairing NMR experiments with proteins $>$ 25 kDa (Venters et al., 1996). To circumvent this problem, the proton density and, thus, the number of relaxation sources in a protein can be minimized by exchanging them with ^{2}H nuclei. This can be achieved by bacterial growth in deuterated water and additional ^{2}H-labeling of the glucose used as a carbon source. Protein deuteration results in great improvements in signal intensity and resolution, but limits the yield of bacterial expressions as isotope effects decrease the catalytic activity of bacterial metabolic enzymes (Gardner and Kay, 1998).

It should be noted that isotopic labeling is no strict requirement for protein NMR studies anymore as technical advances have made measurements of medium-sized proteins at natural ^{13}C and ^{15}N abundance feasible (Arbogast et al., 2016).

3.3.4 Protein Signal Assignment - 3D NMR Spectroscopy

A major bottle neck in structural studies on proteins with NMR spectroscopy is the signal assignment, i.e. finding the link between experimentally observable chemical shifts and the actual nucleus in the protein that is causing it. In large proteins, a *backbone* signal assignment aims at identifying the amide group that belongs to a certain ^{1}H,^{15}N cross peak in a TROSY HSQC spectrum (cf. chapter 3.3.2). This process typically requires [*U*-^{13}C,^{15}N]-labeled samples and the acquisition of three-dimensional (3D) NMR experiments (Cavanagh et al., 1995). In a first step, these experiments are used to link neighboring amino acids via sequential cross peaks. The actual identity of the involved amino acids remains unknown during this process. Sequences of connected signals are generated with the help of the ^{13}C chemical shift of their corresponding amino acids' C_α, C_β, or carbonyl atom. This process is illustrated in detail in Fig. 3.6. Some amino acids (Thr, Ser, Ala, Gly) have unique ^{13}C chemical shifts (Saitô et al., 2010) that allow for assignment of an amino acid *type* to the respective signal. There is a high probability that in long, connected series of signals, a combination of identifiable amino acids will occur that is unique in the protein sequence. At this point, *all* signals in the chain can be assigned to their corresponding amino acid, including the signals with ambiguous ^{13}C chemical shifts in between.

Sequential magnetization transfer between an amide's ^{15}N nucleus and the carbon atom of choice is realized through suitable polarization transfers steps with delays in the pulse sequence matching the scalar coupling constants corresponding to the targeted connection. For instance, the respective

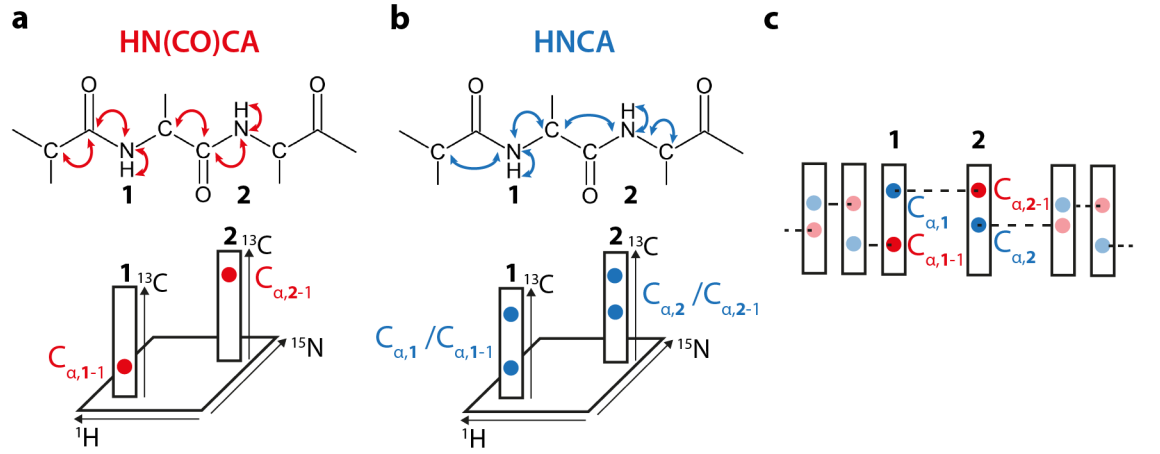


Figure 3.6: Principles of protein backbone signal assignment using 3D NMR spectroscopy. Here, one sequential assignment strategy based on two types of 3D experiments, the HN(CO)CA and HNCA experiments (**a** and **b**, respectively), is shown. For simplicity only two neighboring amino acids of the polypeptide chain are being considered (labeled **1** and **2**). Magnetization can be transferred step-wise from the amide proton to the directly linked ^{15}N and then onto neighboring ^{13}C atoms (indicated by arrows). 3D experiments differ in the kind of ^{13}C atom to which magnetization is being transferred. In the HN(CO)CA experiment, magnetization passes through the carbonyl atom (CO) onto the C_α (CA) of the preceding amino acid. Thus, in a HN(CO)CA experiment each amide HN resonance is being labeled with exactly one ^{13}C chemical shift - that of the C_α of the amino acid on its N-terminal side (i-1). A simplified view of the respective 3D spectrum is shown at the bottom in which the ^{13}C dimension is depicted only as strips perpendicular to the ^{1}H,^{15}N plane. Of course, it is not known a priori that amino acid **1** causes amide signal **1**. In the HNCA experiment, magnetization is not routed through the carbonyl atom but directly onto neighboring C_α atoms. In this case, each amide resonance is labeled with two ^{13}C chemical shifts, the one of its preceding (i-1) and of its own (i) C_α. An overlay (**c**) of the respective ^{13}C strips for each amide signal can be used to distinguish its $C_{\alpha,\,i}$ and $C_{\alpha,\,i-1}$ signals. Matching the $C_{\alpha,i-1}$ shift of one strip with the $C_{\alpha,i}$ shift of another strip identifies amino acids that are neighbors in the polypeptide chain. This process can be continued to yield extended series of connected amino acids through sequential linkage of their ^{13}C atoms (indicated by light blue and red colors).

^{1}J(^{15}N,^{13}CO) and ^{1}J(^{15}N,$^{13}C_\alpha$) coupling constants differ by a factor of 1.5 (Schmidt et al., 2011), enabling selective transfer to either carbon by adjustment of the respective delay (Salzmann et al., 1998). For reasons of sensitivity, the magnetization transfer starts at the amide proton and, in the absence of any other proton in a deuterated protein, is later transferred back to the same proton in an 'out-and-back' manner. This is indicated by the double-headed arrows in Fig. 3.6.

In practice, signal overlap of ^{1}H,^{15}N resonances limits the size of proteins amenable to a backbone assignment because the number of expected signals scales with the length of the peptide chain. As the spectral dispersion is limited in both dimensions, a practical size limit of approximately 80 kDa currently exists (Tugarinov et al., 2002). Recent advances in computational power, the field strength of NMR spectrometers, and the availability of higher-dimensional pulse sequences are challenging this limit. Spectra can be potentially detected with much higher resolution using processing methods that do not rely on classical Fourier transformation approaches (Hyberts et al., 2014). Direct detection methods can lead to resolution improvements in the ^{15}N dimension which might lead to reduced spectral crowding in some cases (Takeuchi et al., 2016). Ubiquitous protein deuteration is a prerequisite for the study of high-molecular weight proteins but poses another challenge when a backbone assignment is attempted (Tugarinov et al., 2006). After expression, most deuterons in amide groups readily exchange with protons as soon as they are exposed to aqueous buffers, causing an observable ^{1}H,^{15}N

signal. However, especially in large proteins some deuterons may be protected from solvent-exchange by secondary and tertiary structure elements – the respective amide groups stays invisible. In this case, the sequential linking of signals with 3D NMR spectra is prohibited in certain parts of the structure, severely impairing any assignment attempts.

3.3.5 Saturation Transfer Difference (STD) NMR Experiments

Saturation transfer difference (STD) NMR is a spectroscopic method that is used to characterize the interaction between small molecules and a protein (Mayer and Meyer, 1999). As a ligand-based NMR technique, typically only 1D proton NMR spectra have to be acquired and no expensive isotope-labeling of the protein is required.

Some protons of folded proteins, especially methyl group protons, have chemical shifts that rarely occur in most small molecules. This spectral window can be used to selectively irradiate protein resonances without affecting any signal of the putative ligand. This irradiation leads to *saturation* of certain protein signals, i.e. equalization of the populations of different spin energy levels and, thus, the disappearance of observable steady-state z-magnetization (Neuhaus and Williamson, 2000). Consequently, no resulting NMR signal is detected. Saturation is achieved by application of a long cascade of low-power radio-frequency pulses at the desired frequency (the *on-resonance experiment*). Within a spin network, saturation spreads quickly via a process called *spin diffusion* onto other, not directly irradiated spins as a consequence of efficient cross-relaxation in the slow tumbling limit. As saturation transfer occurs through space and not through bonds, bound ligand molecules can undergo intermolecular saturation transfer as well. Under favorable circumstances, saturated molecules will dissociate rapidly and, as the ligand is in large excess, other non-saturated molecules will enter the binding pocket (Meyer and Peters, 2003). In this way, a high number of saturated molecules can accumulate in solution and the effect is only lost slowly via relaxation processes. The reduction in signal intensity can be detected by subtraction of the on-resonance spectrum from a reference spectrum in which the irradiation frequency is placed far away from any protein or ligand resonance (the *off-resonance experiment*). Thus, an STD effect appears as a positive signal in a difference spectrum.

Originally, STD NMR has been developed for the quick identification of binders from mixtures of small molecules and a protein, as only the binding molecules would show an intensity decrease (Mayer and Meyer, 1999). Within one ligand molecule, protons can receive different amounts of saturation transfer depending on their orientation towards the protein. This allows to quantitatively describe binding epitopes on the ligand (Mayer and Meyer, 2001) and was later extended to validate and predict models of the bound ligand within its binding pocket (Jayalakshmi and Krishna, 2002). STD NMR is applied routinely in pharmaceutical research as a screening method (Gossert and Jahnke, 2016).

As too strong binding, characterized by low dissociation rates, prevents the accumulation of saturated molecules in solution, STD NMR is limited in the affinity range in which interactions can be studied. Typically, dissociation constants should be at least in the low µM range of dissociation constants. Direct irradiation of ligand signals can lead to false-positive identification of binding molecules. Therefore, negative control experiments are essential using ligand samples lacking the protein.

3.3.6 Paramagnetic Lanthanides in NMR Spectroscopy

Lanthanides are rare-earth metals and most of them have unpaired electrons rendering them paramagnetic. Due to the large magnetic moment of electrons, paramagnetic lanthanide ions as used as tags in protein NMR spectroscopy, especially when investigating metal-binding proteins (Otting, 2010). Two paramagnetic effects are of special importance: paramagnetic relaxation enhancements (PRE) and pseudocontact shifts (PCS).

Depending on the lanthanide's electronic configuration, PRE effects in metal-protein complexes are due to different mechanisms. All cause increased nuclear transverse relaxation through dipolar interactions between the nucleus and the unpaired electrons. This leads to NMR signals that are significantly broadened, sometimes beyond detection (Liu et al., 2014). When bound to a much larger protein, two different fluctuating fields originate from a paramagnetic ion that lead to nuclear relaxation (cf. chapter 3.3.2): one due to molecular motion and another due to electronic spin relaxation (Gueron, 1975). The PRE effect has a strong distance dependency of r^{-6}.

In an external magnetic field, the magnitude of the magnetic moment of some paramagnetic ions is anisotropic, i.e. it depends on the orientation in the external field (John and Otting, 2007). In a tumbling molecule, resulting dipolar interactions with a nuclear spin are consequently not averaged to zero and cause a changed effective magnetic field at the position of the nucleus. This results in an altered Larmor frequency of the nucleus and, hence, an observable chemical shift difference compared to a state in which the paramagnetic ion is absent – the PCS. In a fixed frame, e.g. a paramagnetic ion bound to a protein, different nuclei will experience different effects depending on their distance and orientation towards the metal (Liu et al., 2014). Paramagnetic ions have very different anisotropic components of their magnetic susceptibility tensors $\Delta\chi$. This can be exploited to experimentally derive structural restraints in a protein as PCS effects have a moderate distance dependency of r^{-3}, and the type of metal determines the range in which effects can be observed: short-range effects for cerium or europium (< 15 Å) or long-range effects up to 40 Å with dysprosium (Allegrozzi et al., 2000).

Both PCS and PRE effects have found numerous applications (Otting, 2008): assignment of NMR spectra in case of existing structural models (or the inverse: structure-determination if assigned spectra are available), characterization of ligand binding, protein dynamics or protein-protein interactions.

3.4 Objectives

This thesis aims at a comprehensive characterization of the interactions of murine and human NoV capsid proteins with various soluble ligands that were shown to influence the viral infection process. Specifically, questions from three areas of NoV biology had been identified:

Which glycans are being recognized by human and murine NoV? Where are the respective binding pockets for sialic-acid containing glycans such as gangliosides?

Is the bile-dependent cell culture behavior of the prevailing GII.4 NoV caused by binding of bile acids to a so far unknown binding pocket in the viral capsid? What is the link between bile acid binding of murine NoV capsids and bile's postulated role in viral immune escape?

Is the binding of metal ions, described as cofactors for murine and some human NoV, a general feature of NoV capsid proteins? If so, does it modify their interactions with glycans and bile acids?

The protruding domain proteins from several strains of both the predominant human GII.4 NoV and of murine NoV have been selected to address the above questions. NMR spectroscopy is well-suited for studies into interactions of NoV capsids with glycans, bile acids, and metal ions. Using this approach, a wide range of affinities are experimentally accessible, including low affinity carbohydrate binding. Ligand-based STD NMR and protein-based NMR approaches can be combined to screen for binding molecules but also perform in-depth characterization of selected interactions. Most importantly, chemical shift perturbation experiments can be used to elucidate so far unknown binding sites with atomic resolution under quasi-physiological conditions.

Chemical shifts of backbone amide groups are most sensitive towards environmental changes and give access to well-distributed reporter groups across a protein. Accordingly, a ^{1}H-^{15}N isotopic labeling with a TROSY HSQC read-out can be employed for NMR studies of proteins smaller than 80 kDa. For one of the selected strains, GII.4 Saga, recombinant expression of the protruding domain and its isotopic labeling had already been achieved (Mallagaray et al., 2015). However, the identification of novel binding pockets requires the backbone signal assignment of the protruding domain. Considering the high molecular weight of the dimeric protruding domain (73 kDa), a full backbone assignment is experimentally challenging. Protruding domains from two highly related GII.4 NoV strains, MI001 and VA387, as well as from the murine NoV strain CW1 had been previously described as ganglioside binders (Wegener et al., 2017, Han et al., 2014) but attempts to determine the corresponding binding pocket had not been successful. Establishing an isotopic labeling strategy is required as a first step to allow for protein-based NMR binding studies of these strains.

It is conspicuous that murine and human NoV are reported to share a number of features – bile acid and glycan binding to the capsid among them – but both viruses differ in their host and cell tropism and are linked to very different disease patterns. Thus, in a broader sense this thesis tries to contribute to the question to which extent murine NoV are a suitable model system for human NoV infections.

4. Results

4.1 NMR Backbone Assignment of a GII.4 NoV Protruding Domain

Isotope labeling of a protein's backbone amide groups allows for the acquisition of $^{1}H,^{15}N$ TROSY HSQC NMR spectra, showing one signal for each ^{1}H-^{15}N spin pair (Fig. 4.1a). However, it is not known *a priori* which amide signal corresponds to which amino acid in the protein sequence. In order to achieve a backbone *signal assignment*, 3D NMR spectra are commonly employed to sequentially connect NH signals of neighboring amino acids through the protein's carbon atoms (cf. chapter 3.6). Here, TROSY versions of 3D NMR spectra correlating NH cross peaks with carbonyl, C_α or C_β nuclei (or combinations thereof) were acquired with a sample of [*U*-$^{2}H,^{13}C,^{15}N$]-labeled GII.4 Saga P-domain.

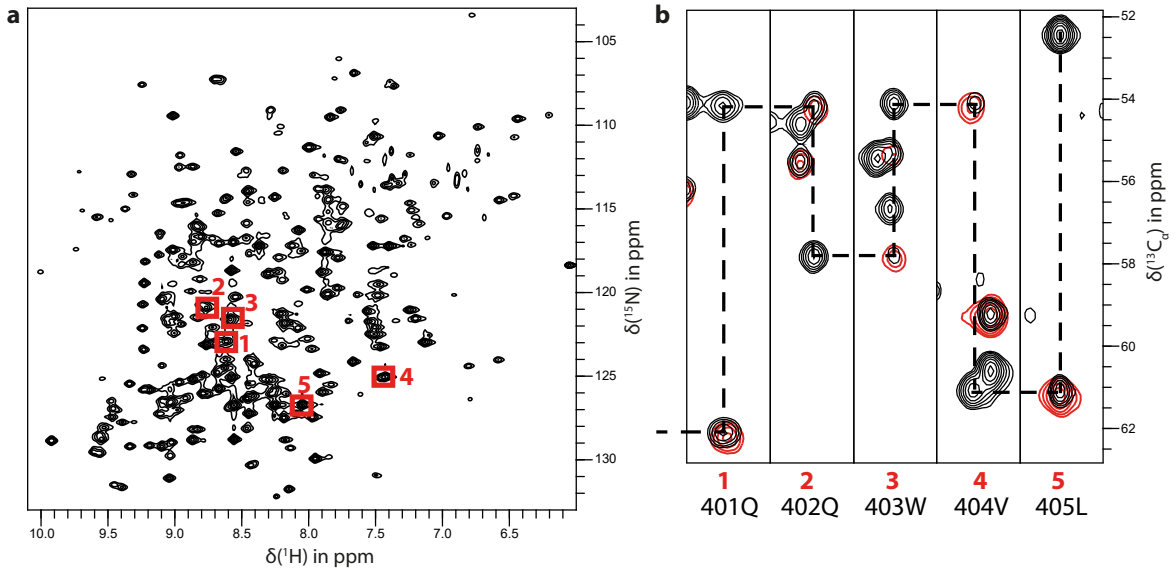


Figure 4.1: Assignment of backbone amide resonances using 3D NMR experiments. 3D protein NMR spectra correlate $^{1}H^{15}N$ cross peaks (**a**) with a third carbon dimension. The nature of the visible carbon resonance depends on the type of 3D spectrum chosen. Here, exemplary strips of HNCA spectra (**b**, black) show the C_α signal of the *i* and *i-1* residue for each $^{1}H^{15}N$ signal of [*U*-$^{2}H,^{13}C,^{15}N$]-labeled GII.4 Saga P-domain dimers (73 kDa) indicated. HN(CO)CA spectra (red) only display C_α signals of the preceding, *i-1*, amino acid. Matching *i-1* cross peaks of one strip with the *i* cross peaks of another yields a sequence of connected amino acids (dashed lines). The connectivities identified from different types of 3D NMR spectra can be used to assign NH signals to their corresponding amino acids. Using this method, an almost complete backbone assignment of the GII.4 Saga P-domain has been obtained (86 % of assignable amino acids). The TROSY HSQC and the HN(CO)CA spectrum have been acquired at 500 MHz, the HNCA spectrum at 900 MHz with a protein concentration of 360 µM in 20 mM sodium phosphate buffer (pH* 7.3). Protein concentrations given here and in the following chapters refer to the concentrations of the monomeric protruding domain species.

Fig. 4.1b shows exemplary data that connects five amino acids via their C_α nuclei in HNCA and HN(CO)CA experiments. Connections to neighboring amino acids were made only when each connection was supported by at least two different carbon atoms with matching chemical shifts. Unique combinations of amino acids with distinct chemical shifts (i.e. Gly, Ala, Ser, Thr) within chains of connected signals served as starting points and were used to assign amino acids in between. Some series of connected amino acids were obtained not containing unique combinations of identifiable amino acids either due to the given primary sequence or incomplete spectral information. In these cases, assignments where made only if at least five connected signals had C_α and C_β chemical shifts matching their amino acids' expected values (available from a chemical shift data base, see below). Using this approach, 243 (86 %) of the protruding domain's backbone amides have been assigned to a corresponding signal. Unassigned amino acids are mostly found within some surface-exposed loops.

In order to validate the signal assignment, obtained chemical shift information of ^{1}H, ^{15}N, and ^{13}C nuclei was used to predict the local secondary structure of the polypeptide chain using the TALOS-N web server (Fig. 4.2). Predictions based on the NMR assignment are in excellent agreement with the crystal structure, correctly mapping α-helix and β-sheet regions. Carbon chemical shifts were also compared to reference data for the different amino acid types deposited in the *Biological Magnetic Resonance Bank* (BMRB)[1]. Three outliers were identified (336Q, 349A, 400P) which can be explained by shielding and deshielding effects of nearby aromatic amino acid side chains.

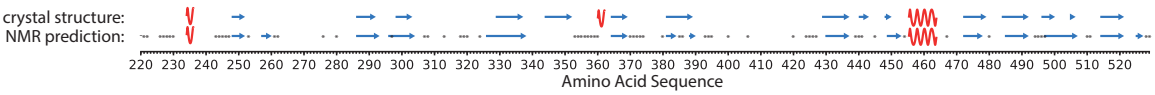


Figure 4.2: Chemical shift-based secondary structure prediction. Chemical shift information of assigned H, N, and C nuclei has been used to predict the local secondary structure for each residue using the TALOS-N web server. α-helices are depicted in red, β-sheets in blue. Asterisks indicate amino acids with missing assignment. Predictions are in good agreement with a crystal structure model (pdb 4X06), thereby validating the obtained chemical shift assignment. Chemical shift data was corrected for the deuterium isotope effect. Figure adapted from Fig. S10 in Mallagaray et al. (2019).

For a high-molecular weight protein such as the NoV protruding domain (73 kDa), backbone signal assignment is far from being trivial. It is imperative to deuterate the protein during its biosynthesis in order to reduce the number of relaxation sources. In turn, ^{2}H nuclei (deuterons) in the backbone amides must later be back-exchanged to observable ^{1}H nuclei (protons). Most amides' deuterium atoms readily exchange with protons from aqueous buffers and a high number of signals can be observed immediately after a buffer exchange (Fig. 4.3a, black spectrum). However, following a protocol for controlled chemical denaturation and refolding of the P-domain leads to the appearance of about 40 new NH signals (Fig. 4.3a, red spectrum). A complete backbone assignment was only possible employing a protein sample treated in this manner as the 3D experimental approach critically relies on sequential magnetization transfer through *observable* nuclei. Using this assignment, the resonances emerging after renaturation can be mapped on a crystal structure model (Fig. 4.3b). All affected amino acids can be found in regions of the protein protected from solvent exchange, mostly through intra-

[1] www.bmrb.wisc.edu, queried in March 2018

molecular hydrogen bonds in buried secondary structure elements such as α-helices and β-sheets. It is noteworthy that amino acids at the dimer interface do not exhibit significant protection from solvent exchange, indicating at least slow monomer-dimer exchange kinetics.

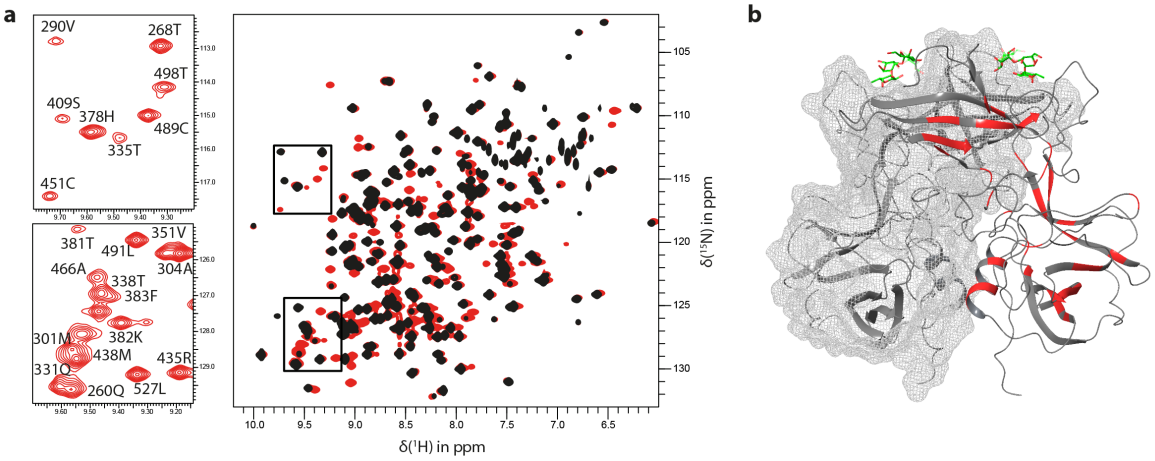


Figure 4.3: Backbone assignment reveals amino acids protected from DH exchange. Protein biosynthesis in deuterated media leads to ubiquitous incorporation of deuterons into the protein. $^{1}H^{15}N$ signals only become visible in TROSY HSQC spectra (**a**, black) after buffer exchange to aqueous buffers due to rapid exchange of amide deuterons with protons from the solvent. Denaturation and refolding of [U-^{2}H,^{15}N]-labeled GII.4 Saga P-domains in aqueous buffers leads to the appearance of approx. 40 additional $^{1}H^{15}N$ cross peaks (red). Based on the backbone signal assignment, these emerging resonances can be localized in regions of the protein with secondary structure elements (**b**, pdb 4X06) where hydrogen bonds protect deuterons from solvent exchange in the native structure. Samples contained 20 mM sodium phosphate buffer (pH* 7.3) with 360-600 µM protein concentration. Due to the different protein concentrations used, emerging signals have been identified in a non-quantitative approach by scaling of the spectra, using the signal of a surface-exposed amino acid (8 / 130 ppm). The denaturation-refolding protocol was developed in-house by Dr. Alvaro Mallagaray (Mallagaray et al., 2019).

The highly homologous P-domains of the strains GII.4 MI001 and GII.4 VA387 were synthesized, isotope labeled and refolded with similar yields (approximately 15 mg [U-^{2}H,^{15}N]-labeled protein after refolding from one liter of cell culture). Data on the synthesis and purification of human NoV P-domains is given in chapter 6.7 of the appendix. Both strains have more than 90 % sequence identity compared to the reference strain GII.4 Saga. For GII.4 VA387, an identical overall protein fold is found as judged from an existing crystal structure with an RMSD value of 0.8 Å (pdb entries 2OBR, 4OOX). A backbone signal assignment was obtained for the P-domain of the NoV strain GII.4 Saga only. Despite their high overall similarity, TROSY HSQC spectra of [U-^{2}H,^{15}N]-labeled samples of the respective P-domains do not allow for the transfer of assignments (Fig. 4.4). The signal assignment can only be transferred onto the homologous strains for some isolated signals. Interestingly, chemical shift differences can be identified even in regions without any sequence variation (Fig. 4.4, bottom). In the case of the conserved hexapeptide around Ile293, a variation in the $i + 3$ position apparently suffices to cause measurable chemical shift differences for Ile293. This demonstrates the sensitivity of ^{1}H,^{15}N chemical shifts towards subtle changes in their electronic environment.

^{15}N relaxation rates can be measured to probe an amino acid's dynamics on a ps- to ns-time scale (cf. chapter 3.3.2). In these experiments, a variable delay τ is introduced into the TROSY HSQC pulse sequence allowing for relaxation to occur. The resulting decrease in signal intensity as a function of

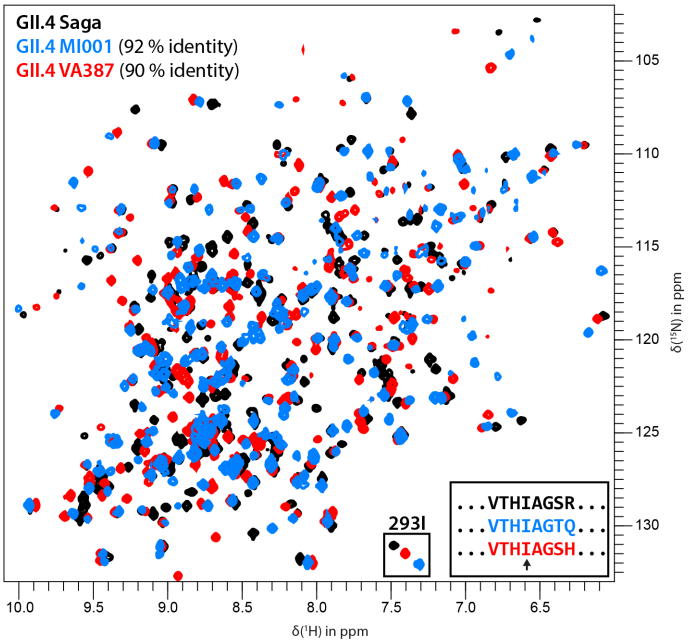


Figure 4.4: TROSY HSQC spectra of [*U*-^{2}H,^{15}N]-labeled P-domains of different GII.4 NoV strains. Isotope-labeled P-domains of the strains Saga (red), MI001 (blue) and VA387 (orange) can be synthesized recombinantly and display highly diverse HSQC fingerprint spectra (**a**) despite a sequence identity larger than 90 % (**b**). Each signal corresponds to one amide group of the protein backbone. Proteins were measured in 75 mM sodium phosphate buffer, 100 mM NaCl (pH* 7.3) with the following protein concentrations: 300 µM GII.4 Saga, 170 µM GII.4 VA387, and 240 µM GII.4 MI001. GII.4 MI001 P-domains were produced as part of the supervised Bachelor's thesis of G. Wallmann (Wallmann, 2019). GII.4 VA387 P-domains were denatured and refolded by P. Ogrissek as part of supervised work as a research assistant.

τ can be fitted to an exponential decay model for each signal, yielding characteristic relaxation times for almost all residues (Fig. 4.5a). The ratio of the relaxation times T_1 and T_2 is sensitive towards motions on the ps- to ns-time scale. Plotting this ratio against the amino acid sequence (Fig. 4.5b) reveals overall homogeneous relaxation times as expected for an isotropically tumbling rigid rotor in solution. Notably, some amino acid stretches display extraordinarily low T_1/T_2 values, indicative of high-frequency motions. The distribution of T_1/T_2 ratios in a protein is non-Gaussian. Accordingly, the $Q_{0.16}$ quantile has been selected as a threshold for T_1/T_2 minima as it corresponds to approximately one standard deviation in normally distributed data. Experimental relaxation data has been compared with data from a 1 µs molecular dynamics simulation (MD) (Fig. 4.5b bottom)[2]. The root mean square fluctuation (RMSF) of atomic positions, i.e. the deviation from their average positions over time, in the MD trajectory was used to identify regions of high flexibility. Indeed, RMSF maxima coincide with minima in T_1/T_2 values. All of the respective residues can be found in solvent-exposed, mobile loops or the termini of the polypeptide chain (Fig. 4.5c). Average ^{15}N T_1 and T_2 relaxation times are given in Tab. 4.1. A TRACT NMR experiment was used to obtain a lower limit for the rotational correlation time τ_c of the 73 kDa GII.4 Saga P-domain dimers. τ_c was found to be 34 ns and can be used to estimate the molecular weight of the protein as larger than 60 kDa as explained in detail in chapter 6.4.7, corroborating the dimeric nature of the protein.

[2]provided by Dr. M. Stein and E. Schulze, Max Planck Institute of Complex Technical Systems Magdeburg as published in Creutznacher et al. (2019)

Table 4.1: Average ^{15}N relaxation times of GII.4 Saga. Residue-specific T_1 and T_2 relaxation times were obtained for 207 amino acids (cf. Fig. 4.5). The average values are reported here. A TRACT experiment of 360 µM [*U*-^{2}H,^{15}N]-labeled GII.4 Saga in 75 mM sodium phosphate buffer, 100 mM NaCl (pH 7.3) was used to derive rotational correlation times τ_c.

τ_c	^{15}N T_1	^{15}N T_2
34 ns	2.5 s	23 ms

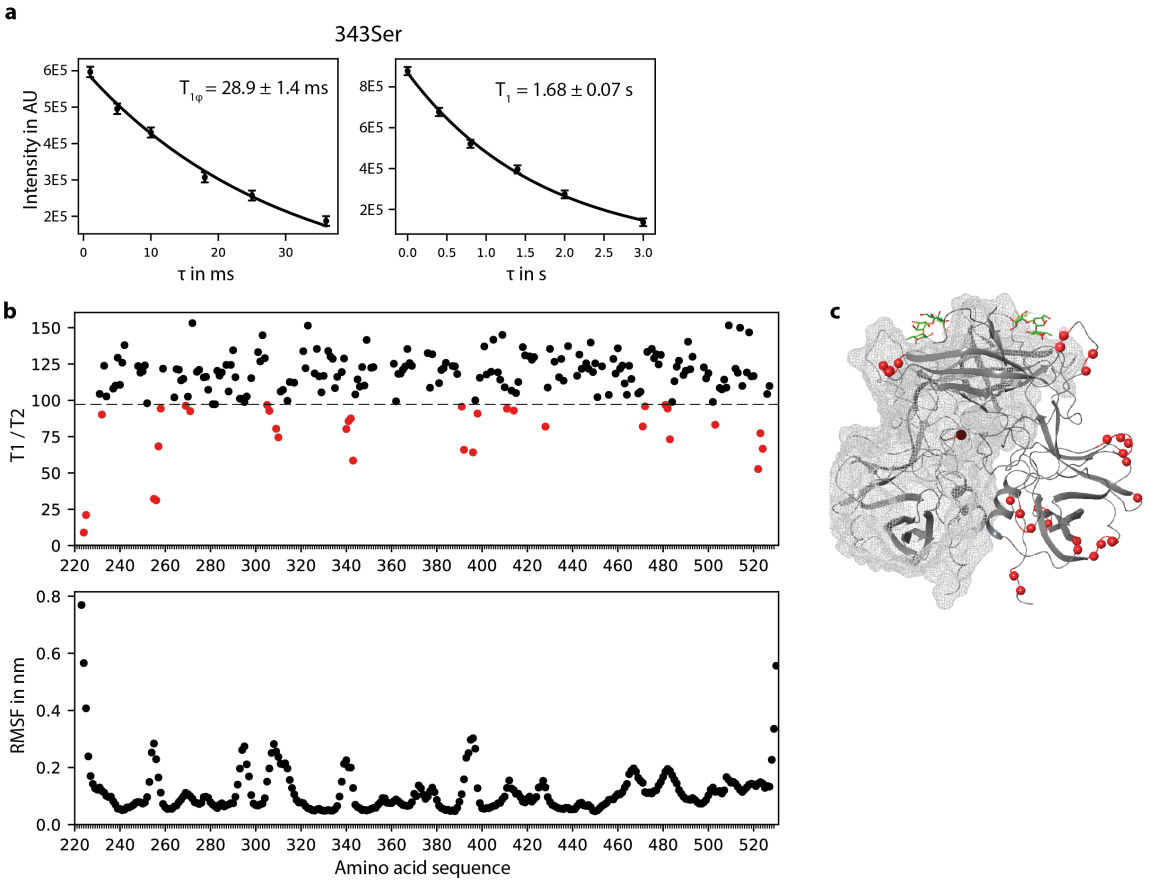


Figure 4.5: ^{15}N relaxation times and local protein dynamics. Characteristic T_1 and $T_{1\rho}$ relaxation times of individual ^{15}N backbone atoms can be extracted by fitting of signal intensities in a series of 2D TROSY HSQC spectra with increasing time delays τ to an exponential decay model. In $T_{1\rho}$ experiments, signal decay is probed in the presence of a static spin lock. Exemplary data for one amino acid is shown in **a**. Site-specific T_2 values can be calculated from the respective T_1 and $T_{1\rho}$ values according to Eq. 6.11. The ratio T_1/T_2 is sensitive towards local protein dynamics (**b**). Amino acids with T_1/T_2 ratios smaller than the $Q_{0.16}$ quantile are depicted in red. Such residues are highly dynamic on a ps- to ns-time scale and can be found in solvent-exposed loops (**c**, pdb 4X06). The relaxation behavior in regions with secondary structure elements matches that of a rigid rotor. NMR relaxation data is in excellent agreement with data from a 1 µs molecular dynamics (MD) simulation. Root mean squared fluctuations (RMSF) of the backbone atoms during the simulation give another measure for local flexibility (**b**, bottom). Monomer-averaged RMSF maxima coincide with T_1/T_2 minima. MD data has been provided by Dr. M. Stein and E. Schulze, Max Planck Institute of Complex Technical Systems Magdeburg and was published in Creutznacher et al. (2019). ^{15}N relaxation data has been acquired at 500 MHz with a sample containing 340 µM [*U*-^{2}H,^{15}N]-labeled Saga 373D P-domain in 20 mM sodium phosphate buffer (pH* 7.3) with contributions from P. Ogrissek as part of supervised work as a research assistant.

4.2 A Post-Translational Modification Impairs Glycan Recognition

TROSY HSQC spectra of GII.4 Saga P-domains acquired immediately after purification and some time thereafter show major differences (Fig. 4.6a). A large number of signals decrease in intensity with new signals arising at a different position. The spectral changes are irreversible, indicating covalent post-translational changes to the protein. In order to localize the source of the effect, Euclidean chemical shift differences between corresponding signals of the two species were calculated (Fig. 4.6b) and were mapped on a crystal structure model using the signal assignment (Fig. 4.6c). Chemical shift perturbations are centered in the outward-facing part of the P-domain near the glycan binding site.

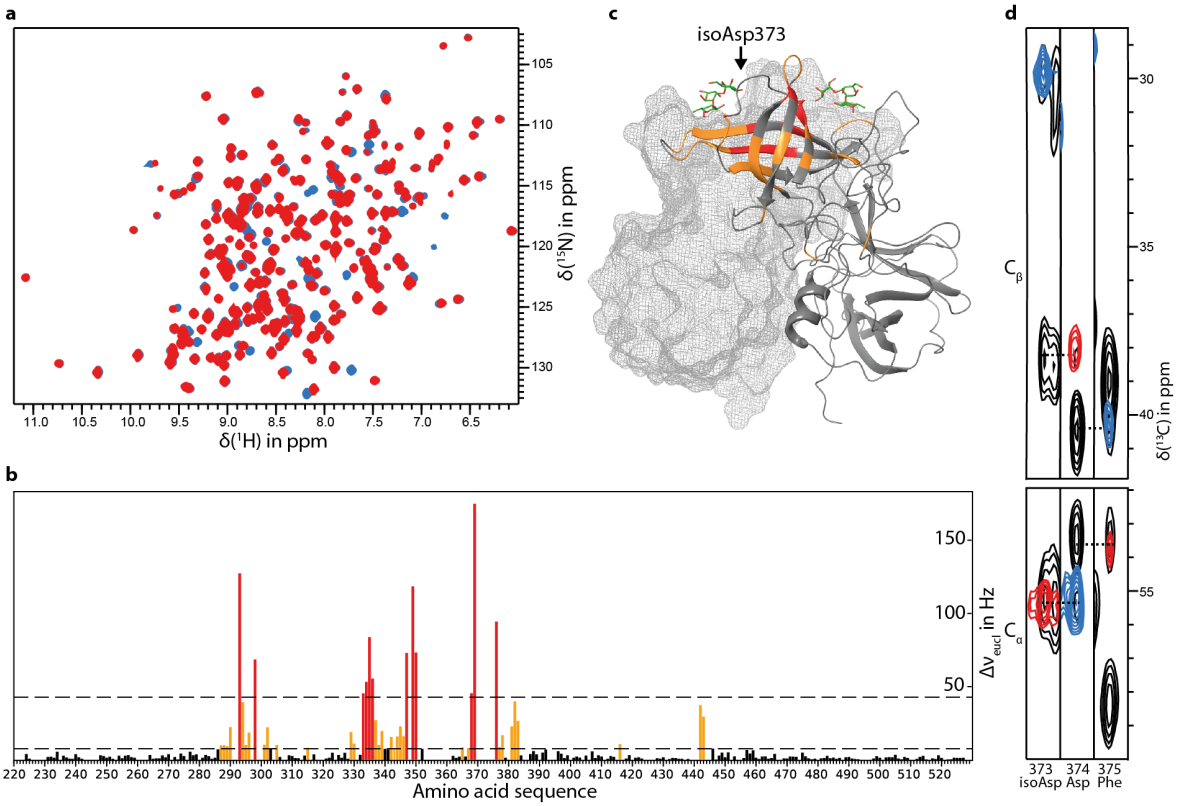


Figure 4.6: NMR experiments reveal spontaneous deamidation of an Asn residue within the HBGA binding pocket. Comparison of TROSY HSQC spectra (**a**) of freshly synthesized [*U*-^{2}H,^{15}N]-labeled GII.4 Saga P-domains (red) with those of aged samples (blue) shows the emergence of a second characteristic set of signals. This process is irreversible and is caused by the post-translational conversion of the protein into a different species. Using the backbone signal assignment, Euclidean chemical shift differences between corresponding signals can be calculated for almost all amino acids (**b**). Perturbations larger than 8 Hz are considered significant (orange), CSPs larger than two standard deviations are indicated in red. These affected amino acids are all located in the outward facing part of the protruding domain (**c**, pdb 4X06) in close proximity to the canonical glycan binding pocket. Inspection of strips (**d**) from the 3D HNCACB (black) and HN(CO)CACB (blue and red) spectra reveals an unusual phase inversion for one set of C_α and C_β signals, demonstrating the formation of an isopeptide bond at position 373 as a result of Asn deamidation. This is in excellent agreement with the observed chemical shift perturbations within the protein. Spectra were acquired with 210-300 µM P-domain in 75 mM sodium phosphate buffer, 100 mM NaCl (pH* 7.3). Figure adapted from Fig. 2 in Mallagaray et al. (2019).

Inspection of the 3D HN(CO)CACB spectrum in the affected region reveals the underlying molecular cause of this effect (Fig. 4.6d). Usually in this type of experiment, C_α and C_β signals can be distinguished by their phase, with one set of carbons giving positive signals and the other negative signals

(indicated by red and blue color in Fig. 4.6d). However, at position 374 the carbon signals of the preceding amino acid Asn373 display a peculiar phase inversion. This is unambiguous proof for the formation of an isopeptide bond at position 373 (Tugarinov et al., 2002). The asparagine residue 373 deamidates and forms an iso-aspartate (isoAsp) residue via a cyclic succinimide intermediate, introducing the former side chain C_β atom into the protein backbone. As this interchanges the order of C_α and C_β as seen from the following residue 374, step-wise magnetization transfer in the HN(CO)CACB experiment causes the observed phase inversion (cf. Fig. 3.6). Formation of an isoAsp residue at position 373 is in excellent agreement with the experimental CSPs. Interestingly, significant perturbations can still be detected 20 Å apart from the new isopeptide bond. Signals from the loop carrying the deamidation site (370-375) only become visible *after* deamidation. This indicates that the post-translational modification also significantly alters local protein dynamics.

In order to properly characterize wild type and deamidated protein species, a separation protocol based on cation exchange chromatography (IEX) has been developed. Deamidation of Asn373 (N) and subsequent formation of an iso-aspartate (iD) residue introduces a new negative charge into the protein. Due to this charge difference, both species can be eluted separately from the IEX resin with an increasing salt gradient (Fig. 4.7a). As the GII.4 P-domain forms stable homodimers (cf. Fig. 4.12) up to three different species can be detected: wild type protein, carrying Asn in each monomer (N/N); a mixed dimer with one isoAsp and one Asn (iD/N); and a fully deamidated dimer (iD/iD). Prolonged incubation of the protein at 37 °C shows that, as expected, the three species form consecutively and each new negative charge shifts the elution volume further to the left. The integral of the UV absorbance was used to quantify the amounts of each species during the incubation experiment (Fig. 4.7b). Fitting the decrease in N/N dimers to an exponential decay model yields a half-life of 1.6 d.

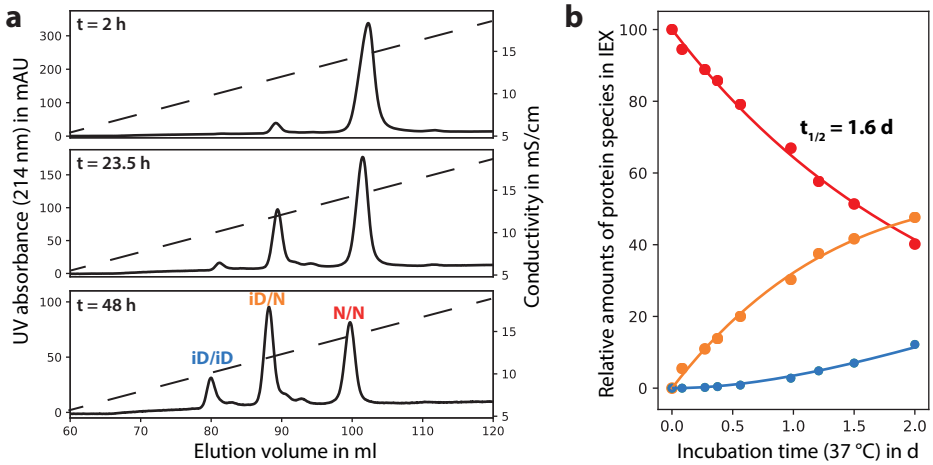


Figure 4.7: Deamidation kinetics of the GII.4 Saga P-domain determined by ion exchange chromatography. Deamidation of 373N and the subsequent formation of an iso-aspartate (iD) residue increases the net negative charge of the protein. Up to two deamidation reactions can occur within the homodimeric protein complex. Protein species with different charge states can be separated via cation exchange chromatography (IEX) and elute in order of decreasing negative charge (**a**) in a salt gradient (dashed line). The P-domain was incubated in 75 mM sodium phosphate buffer, 100 mM NaCl, pH 7.3 at 37 °C. Three major species can be identified upon prolonged incubation - wild type (N/N), once (iD/N) and fully deamidated (iD/iD) dimers. Relative amounts of these species can been quantified by integration of their respective UV absorbance at 214 nm. Data has been fitted to an exponential decay model (**b**), yielding a half-life of 1.6 d of the wild-type protein species at 37 °C. Figure adapted from Fig. S21 in Mallagaray et al. (2019).

Surprisingly, cross peaks in a TROSY HSQC spectrum of the purified asymmetrical iD/N dimer do not completely match those of the symmetrical N/N or iD/iD counterparts (Fig. 4.8a). Some signals of the iD/N species have unique chemical shifts, not matching the chemicals shifts of the same amino acid in the N/N or iD/iD species. This observation implies that the electronic environment of some amino acids in the non-deamidated monomer is influenced by deamidation of the respective other monomer. As the involved distances are approximately 20 Å, this represents, direct or indirect, long-range effects. Amino acids with even greater distances do not show this behavior and their signals in the corresponding spectra are symmetrical (Fig. 4.8b).

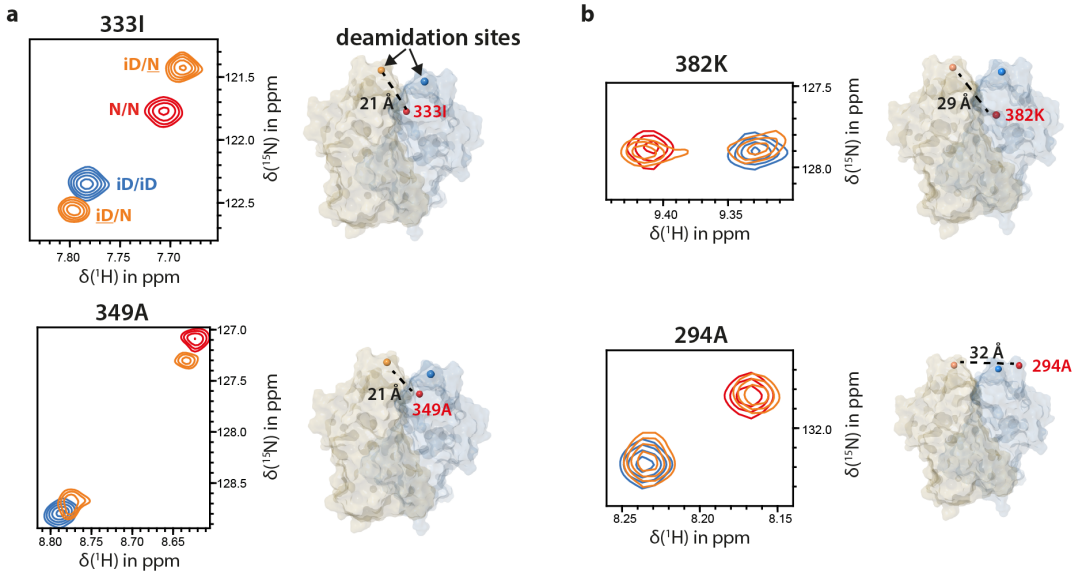


Figure 4.8: Asymmetrically deamidated GII.4 Saga P-domains give unique signals in TROSY HSQC spectra. Some signals (**a**) of the iD/N dimers (orange) do not match their counterparts in spectra of N/N (red) and iD/iD dimers (blue). This indicates an effect of the deamidation of one monomer on the electronic environment of the non-deamidated monomer despite distances larger than 20 Å (indicated in the corresponding crystal structure, pdb 4X06). However, signals of other amino acids with larger distances are symmetrical (**b**). Data was obtained at 293 K in 75 mM sodium phosphate buffer, 100 mM NaCl (pH* 7.3) as part of the supervised Bachelor's thesis of P. Ogrissek (Ogrissek, 2018). Concentrations were 360 µM for N/N and iD/iD and 270 µM for iD/N P-domains.

Asp374, the amino acid neighboring the deamidation site, is critically involved in glycan recognition. It forms a bidentate hydrogen bond with the fucose moiety that is an essential part of all histo-blood group antigens (HBGAs, cf. chapter 3.1.4). One model HBGA, the B-trisaccharide (cf. Fig. 3.2), has been titrated to [*U*-^{2}H,^{15}N]-labeled GII.4 Saga P-domain to characterize the impact of deamidation on ligand recognition. Titrating the non-deamidated wild type protein causes fast-exchange chemical shift perturbations (CSPs) (Fig. 4.9a) centered around the canonical glycan binding pocket (Fig. 4.9c and d). The largest CSPs have been fitted to the law of mass action to yield a K_D value of 5.6 mM (Fig. 4.9b). Repeating the experiment with an aged sample that was mostly deamidated reveals a dramatic decrease in affinity: the K_D value increases by an order of magnitude. Titration of the freshly purified protein was completed within less than four days in order to minimize the impact of the on-going deamidation reaction on the binding behavior. Ligand saturation of the binding pockets of the deamidated protein could not be achieved as it would require impractical amounts of ligand. Therefore, the derived K_D value of 60 mM has to be understood as an estimate.

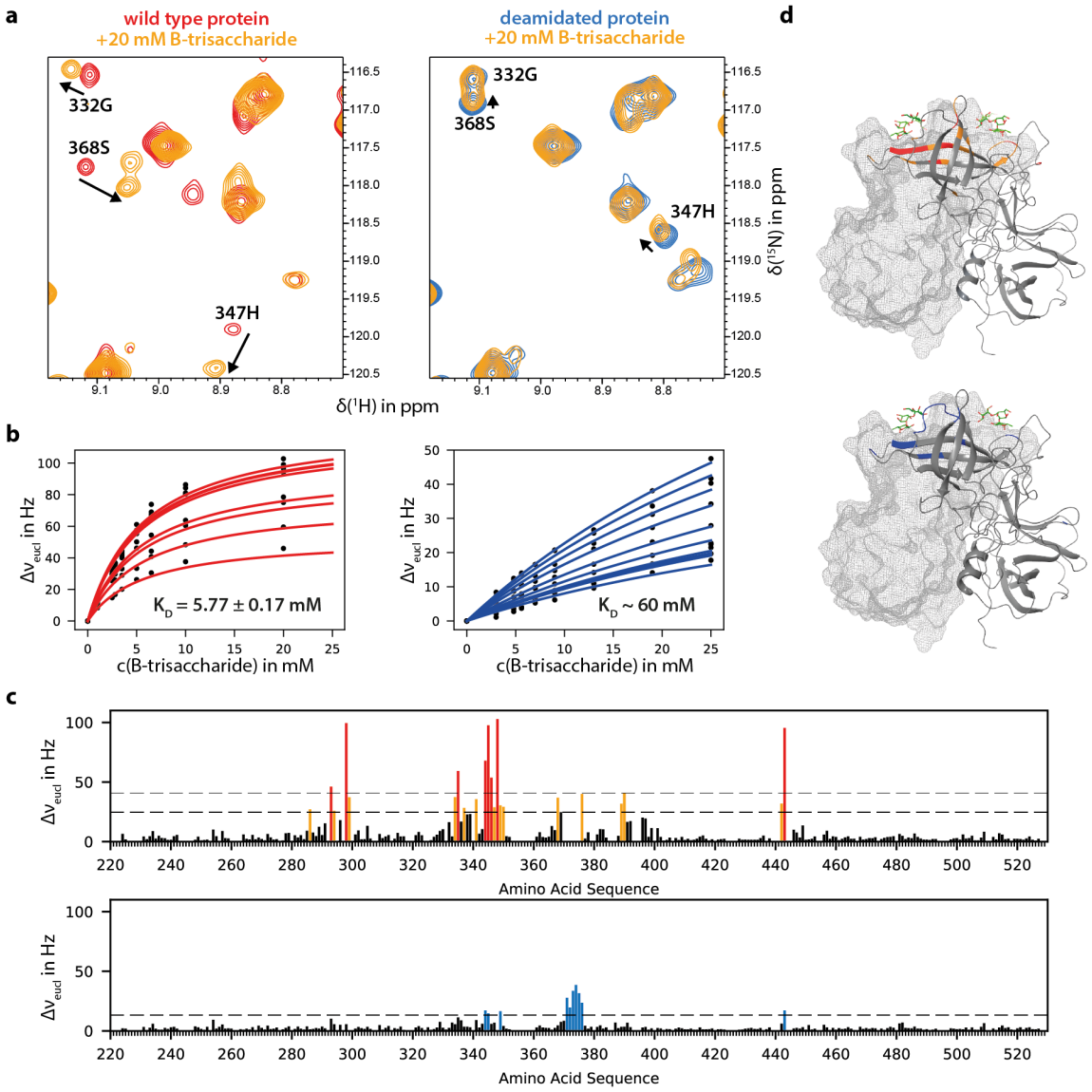


Figure 4.9: Glycan binding to a GII.4 NoV P-domain. Titration of [*U*-^{2}H,^{15}N]-labeled GII.4 Saga P-domains with increasing amounts of blood group B-trisaccharide – a model histo-blood group antigen (HBGA) – leads to chemical shift perturbations (CSPs) in TROSY HSQC spectra (**a**). Data for the non-deamidated protein is depicted in red, data for the fully deamidated protein in blue. Significant Euclidean CSPs can be fitted to the law of mass action to yield dissociation constants K_D for both protein species (**b**). The affinity of the wild type protein species was found to be in the low mM-range. The affinity for HBGAs decreases by an order of magnitude upon protein deamidation. The K_D for the deamidated protein is only an estimate as saturation of the binding pocket could not be reached. CSPs are centered around the known HBGA binding pocket (**c** and **d**, pdb 4X06). CSPs larger than one standard deviation are considered significant (orange), CSPs larger than two standard deviations were used for curve fitting (red and blue, respectively). NMR signals of the loop carrying the deamidation site (370-375) cannot be observed in the wild type protein possibly due to unfavorable relaxation properties or conformational exchange. NMR samples contained 210-300 µM P-domain in 75 mM sodium phosphate buffer, 100 mM NaCl (pH* 7.3). Figure adapted from Fig. 5 in Mallagaray et al. (2019).

Asparagine deamidation occurs via a cyclic succinimide intermediate (cf. Fig. 3.5). In principle, the ring-opening reaction can lead to the formation of either aspartate or iso-aspartate. Both species are expected to form as this reaction is reversible. However, only the iso-aspartate product could be detected in NMR spectra and IEX chromatograms (cf. Figs. 4.6 and 4.7). Site-directed mutagenesis was used to generate a GII.4 Saga point mutant that contains an aspartate at position 373. An NMR spectrum of the N373D mutant confirms that aspartate is, indeed, not formed as a product of this deamidation reaction in detectable quantities. A number of characteristic amide signals of 373D can be identified that do not overlap with signals of the 373N or 373iD proteins (Fig. 4.10a, red boxes). These signals can be used as a read-out for aspartate formation and could not be observed at any point in time during experiments with either 373N or 373iD. Additionally, the 373D mutant itself is stable and does not produce any NMR signals that correspond to 373iD over time. Incubation of the 373D protein at 25 °C for 2 weeks and subsequent IEX analysis reveals only a minor secondary species, populated less than 5 %. Consequently, the deamidation of Asn373 appears to be under strong *kinetic control* in favor of iso-aspartate formation. The 373D P-domain has a similar affinity for glycans as the Asn wild type (cf. Fig. 4.9). Fitting CSP data from a titration of the model HBGA B-trisaccharide results in a K_D value of 2.44 ± 0.05 mM (Fig. 4.10b). Affected signals correspond to amino acids surrounding the canonical glycan binding pocket identified in crystal structures.

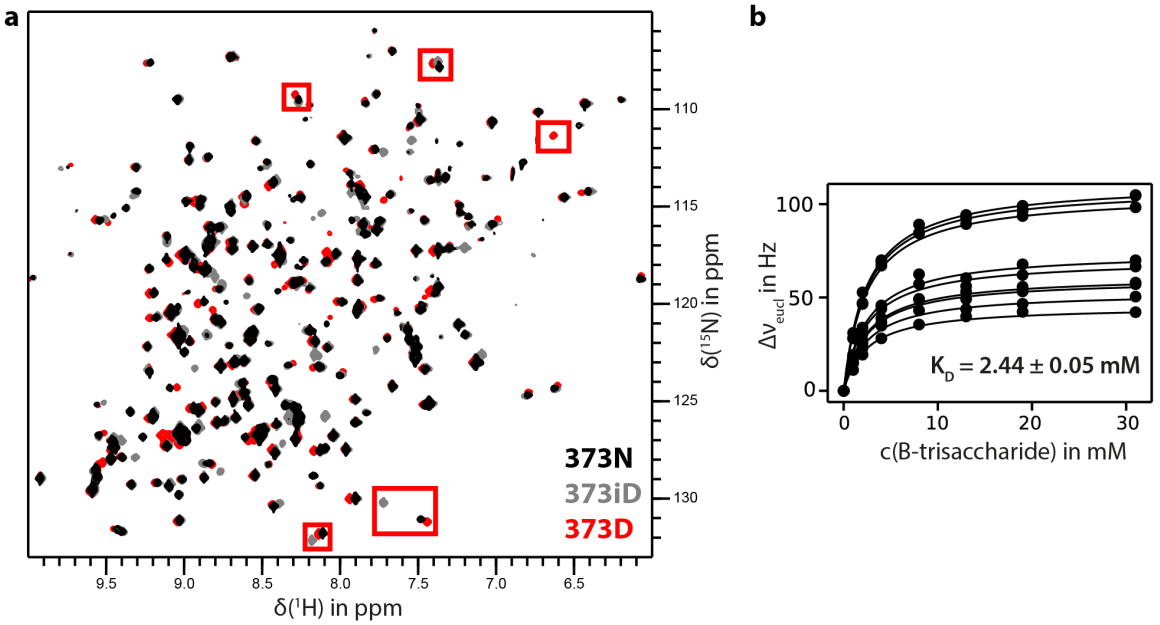


Figure 4.10: Deamidation reaction favors iso-aspartate, not aspartate formation. TROSY HSQC spectra of [*U*-^{2}H,^{15}N]-labeled GII.4 Saga P-domain reveal that the product of the deamidation reaction is solely isoAsp (**a**). An N373D point mutant of the protein (red) produces characteristic signals that do not overlap with any of the 373N (black) or 373iD (grey) signals (red boxes). These signals are never found in any of the spectra of the deamidating protein, thus indicating that the reaction equilibrium is strongly shifted towards isoAsp with Asp being below the limit of detection ($<$ 10%). Likewise, spectra of the 373D mutant do not display signals attributed to 373iD. A titration of α-N_3-B-trisaccharide to the 373D P-domain shows that the binding affinity for HBGAs is similar to that of the wild type 373N protein (**b**). The 373D P-domain can be used as a stable model system for ligand interaction studies. Spectra were acquired with 210-300 µM P-domain in 75 mM sodium phosphate buffer, 100 mM NaCl (pH* 7.3). The sample used for titration contained 240 µM GII.4 N373D Saga P-domain.

In order to determine whether the deamidation reaction at position 373 is specific for the P-domain of GII.4 Saga, other NoV P-domains were subjected to incubation experiments with subsequent IEX analysis as well (Fig. 4.11a and b). GII.4 MI001 is closely related to GII.4 Saga and has a very similar amino acid sequence in positions 370-376 (Fig. 4.11c), including Asn373. GII.10 Vietnam belongs to a different genotype and contains a longer glutamine residue instead. The local loop structure as well as the overall protein fold appear to be conserved between all strains as judged from the crystallographic models (Fig. 4.11c, right). Only GII.4 MI001 shows the characteristic pattern of three peaks in the IEX chromatogram, indicating deamidation (cf. Fig. 4.7). Apparently, the glutamine residue in GII.10 Vietnam cannot undergo deamidation and cyclization on the same time scale as a shorter asparagine residue.

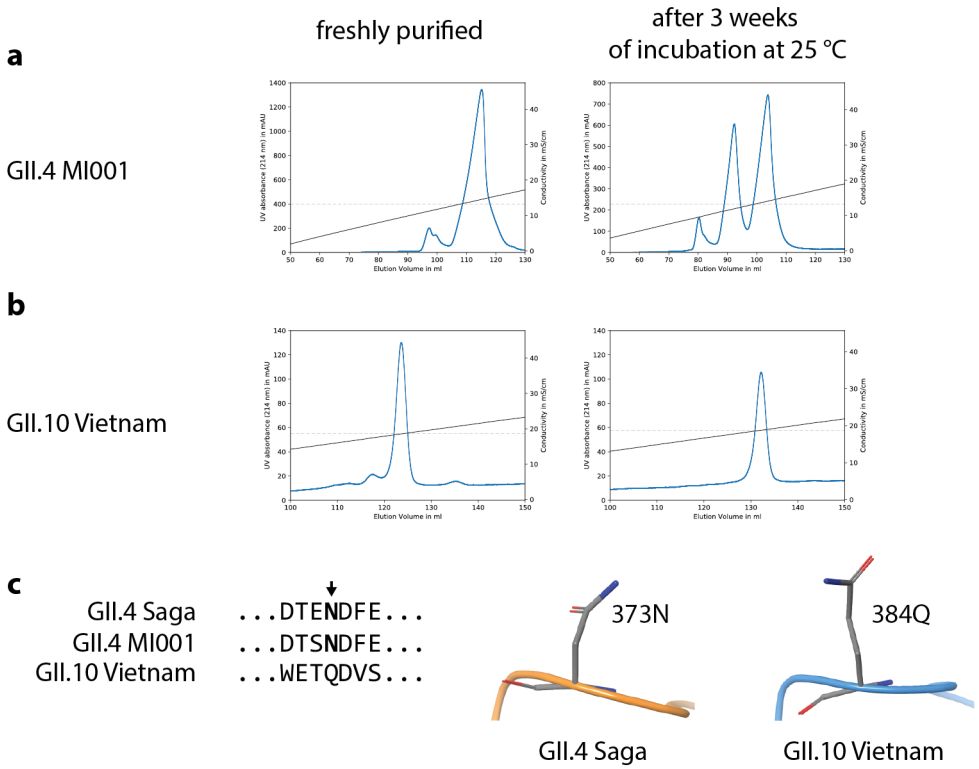


Figure 4.11: Deamidation of P-domains from different human NoV strains. GII.4 MI001 (**a**) and GII.10 Vietnam P-domains (**b**) were subjected to IEX chromatography directly after expression (left). The major protein species was purified, and re-analyzed after three weeks of incubation at 25 °C (right). Only the GII.4 MI001 P-domain shows the three distinct charge states indicative of deamidation after incubation. Solid black lines indicate solvent conductivity, dashed black lines mark the elution conductivity of the wild type protein species. Elution volumes cannot be compared due to slight changes in salt content in different buffer batches. The GII.4 Saga and GII.4 MI001 P-domain have high sequence similarity in the loop carrying the deamidation site 373N (**c**, pdb 4X06). The homologous loop in the GII.10 Vietnam P-domain displays the same overall conformation (right, pdb 3Q38) but carries a longer glutamine residue instead of an asparagine. Vietnam and MI001 P-domains were provided by Dr. Alvaro Mallagaray. Figure adapted from Fig. S22 in Mallagaray et al. (2019).

The time-dependent evolution of N/N into iD/N, and then iD/iD dimers as identified by IEX chromatography (cf. Fig. 4.7) can be used to determine reaction rates k of the macroscopic reaction Asn $\xrightarrow{k}$ isoAsp. The involved short-lived succinimide intermediate must be omitted as it cannot be detected in IEX chromatograms. Deamidation of the dimeric P-domain can then be thought of as an irreversible, consecutive reaction NN $\xrightarrow{k_1}$ iDN $\xrightarrow{k_2}$ iDiD. The system of differential equations

describing the evolution of such a system has an analytical solution (Laidler, 1987). However, IEX data can only be fitted if it is assumed that $k_1 \neq k_2$, i.e. the second deamidation reaction progresses with a different rate than the first. As it appears counter-intuitive that the deamidation reaction rates of the first monomer should differ dramatically from that of the second monomer, further experiments were conducted to test the simple model of consecutive deamidation reactions.

The model presented above assumes that the P-domain remains an intact dimer until the deamidation reaction has been completed. Dissociation and re-assembly of dimer molecules could obscure the deamidation reaction kinetics as, for example, N/N dimers could re-form from two dissociated iD/N molecules. To study the possibility of monomer-dimer exchange kinetics independently, two stable, i.e. not deamidating, variants of the GII.4 Saga P-domain were produced. A 373Q point mutant can be assumed to not undergo deamidation (cf. Fig. 4.11) and fully deamidated iD/iD dimers do not revert back into any other species. Both protein species were mixed 1:1, incubated at 25 °C, and subjected to IEX analysis in regular time intervals. 373Q/Q and 373iD/iD have a charge difference of 2 and, therefore, can be separated via IEX. If monomer-dimer exchange occurs, a mixed iD/Q species will form over time with an intermediate charge and elution behavior. Indeed, over the course of 40 days, mixed iD/Q dimers formed and the system reached the expected 1:2:1 equilibrium. Data can be fitted to exponential decay models and the dimer dissociation rate can be obtained as $k_{\text{off}} = 1.51 \pm 0.02 \times 10^{-6} s^{-1}$. This corresponds to a dimer half-life of 5.3 d. It has been assumed that neither the 373Q point mutation nor the isopeptide bond influence the dissociation and association reaction.

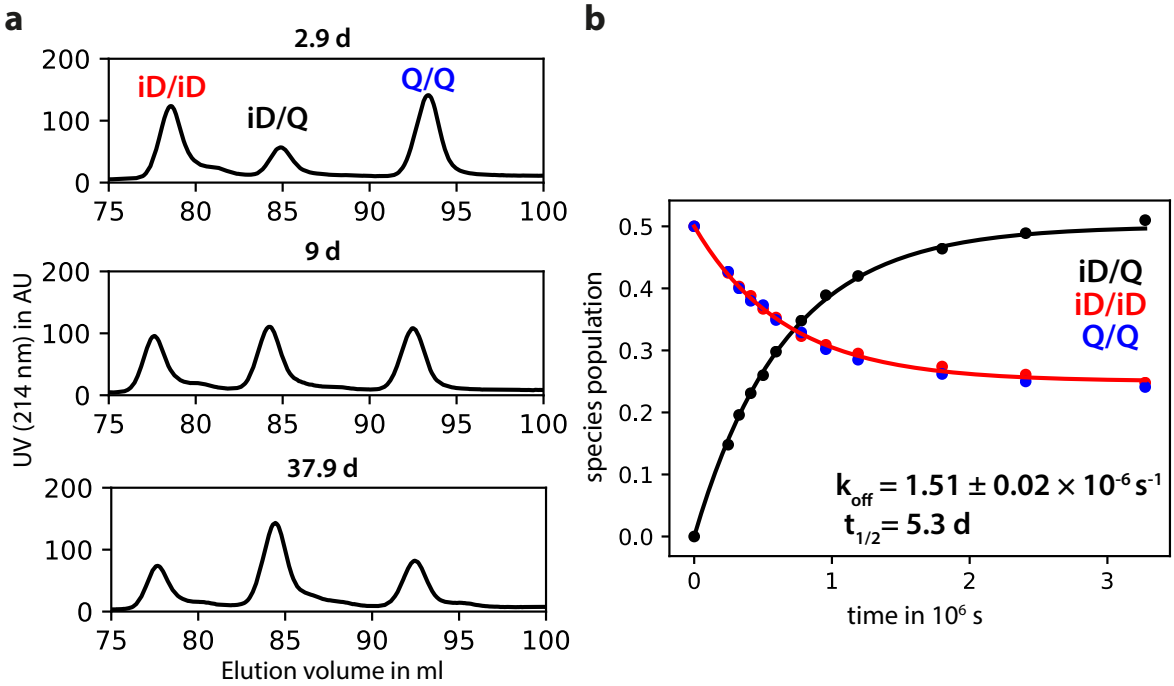


Figure 4.12: Dimer exchange kinetics of human NoV GII.4 Saga. Protruding domain samples which were deamidated at position 373 into a negatively charged iso-aspartate (iD) residue and point mutants with a neutral glutamine (Q) at this position were mixed 1:1 and incubated at 25 °C in 75 mM sodium phosphate buffer, 100 mM NaCl (pH 7.3). Exchange of monomer subunits between these otherwise stable dimers could be observed. The fraction of mixed 373 iD/Q dimers was quantified by ion exchange chromatography at selected time intervals (**a**). Homo- and mixed dimer fractions reached the expected 1:2:1 ratio after approx. 40 d (**b**). Curve fitting yields a dimer dissociation constant k_{off} of $1.5 \times 10^{-6} s^{-1}$.

The following set of reaction equations describes the system of irreversibly deamidating 373N proteins with simultaneous monomer-dimer exchange:

$$\mathrm{NN} \xrightarrow{2\,k_1} \mathrm{iDN} \xrightarrow{k_1} \mathrm{iDiD}$$

$$\mathrm{N} + \mathrm{N} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \mathrm{NN}, \mathrm{iD} + \mathrm{N} \underset{k_{off}}{\overset{2\,k_{on}}{\rightleftharpoons}} \mathrm{iDN}, \mathrm{iD} + \mathrm{iD} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \mathrm{iDiD}$$

$$\mathrm{N} \xrightarrow{k_1} \mathrm{iD}$$

Here, deamidation of the first and second monomer has been assumed to have the same reaction rate constant k_1. An additional correction factor of 2 is introduced in terms describing the deamidation of N/N dimers accounting for a probability argument. There are two possibilities how one iD/N molecule can form out of one N/N dimer, but only one possibility to form iD/iD out of one iD/N molecule. Similarly, the association constant for the formation of asymmetric iD/N dimers has to be modified. For large numbers of molecules N, it is twice as likely to form heterodimers than homodimers (N^2 and $\frac{N(N-1)}{2}$ possible combinations, respectively, cf. Goulet (2016) and chapter 6.3.6). The ordinary differential equations 4.1 - 4.5 describe the kinetics of this system of coupled reactions.

$$\frac{\mathrm{d}[\mathrm{NN}]}{\mathrm{d}t} = -(2k_1 + k_{\mathrm{off}})[\mathrm{NN}] + k_{\mathrm{on}}[\mathrm{N}]^2 \tag{4.1}$$

$$\frac{\mathrm{d}[\mathrm{iDN}]}{\mathrm{d}t} = 2k_1[\mathrm{NN}] - (k_1 + k_{\mathrm{off}})[\mathrm{iDN}] + 2k_{\mathrm{on}}[\mathrm{iD}][\mathrm{N}] \tag{4.2}$$

$$\frac{\mathrm{d}[\mathrm{iDiD}]}{\mathrm{d}t} = k_1[\mathrm{iDN}] - k_{\mathrm{off}}[\mathrm{iDiD}] + k_{\mathrm{on}}[\mathrm{iD}]^2 \tag{4.3}$$

$$\frac{\mathrm{d}[\mathrm{N}]}{\mathrm{d}t} = -k_1[\mathrm{N}] - 2k_{\mathrm{on}}[\mathrm{N}]^2 + 2k_{\mathrm{off}}[\mathrm{NN}] - 2k_{\mathrm{on}}[\mathrm{iD}][\mathrm{N}] + k_{\mathrm{off}}[\mathrm{iDN}] \tag{4.4}$$

$$\frac{\mathrm{d}[\mathrm{iD}]}{\mathrm{d}t} = k_1[\mathrm{N}] - 2k_{\mathrm{on}}[\mathrm{iD}]^2 + 2k_{\mathrm{off}}[\mathrm{iDiD}] - 2k_{\mathrm{on}}[\mathrm{iD}][\mathrm{N}] + k_{\mathrm{off}}[\mathrm{iDN}] \tag{4.5}$$

Two incubation experiments were performed to test this model: one starting with isolated N/N and one with asymmetrical iD/N dimers (Fig. 4.13). IEX experiments were performed after selected timer intervals. Re-formation of N/N dimers can be observed in the iD/N incubation experiment, confirming that monomer-dimer exchange is, indeed, occurring on a similar time scale as the deamidation reaction. The system of ordinary differential equations 4.1 - 4.5 was solved numerically using k_{off} derived in Fig. 4.12. Initial concentration values at $t = 0$ were set to 1 for the isolated starting species N/N or iD/N and zero for all the other species. Monomer concentrations [N] and [iD] remain negligibly small for $k_{\mathrm{on}} = 10^3 - 10^6 \mathrm{M}^{-1}\mathrm{s}^{-1}$, i.e. the equilibrium lies almost completely on the side of the dimers in an affinity range of 1.5 nM to 1.5 pM. Accordingly, varying k_{on} in the given range has no discernible effect on the solution and residuals of the curve fitting do not form a narrow minimum (shown in chapter 6.9). Thus, k_{on} was set arbitrarily to $10^3 \mathrm{M}^{-1}\mathrm{s}^{-1}$ as this is the order of magnitude determined in this thesis for the homologous murine NoV protruding domain (cf. chapter 4.4). The deamidation reaction rate k_1 was determined by least squares minimization with the data of both independent experiments. Data is in good agreement with the fit and k_1 was found to be $4.2 - 4.5 \times 10^{-7} s^{-1}$.

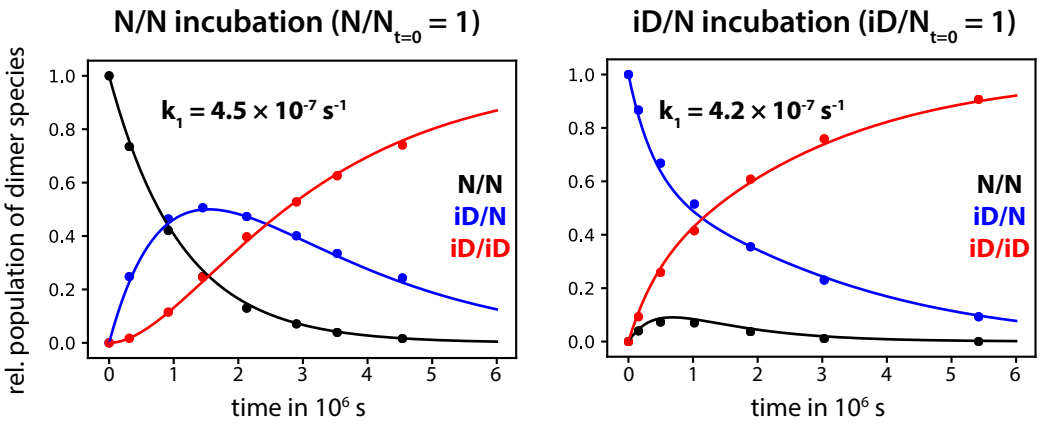


Figure 4.13: Deamidation reaction rate of the GII.4 Saga P-domain. Relative populations of protein species during incubation experiments at 25 °C in 75 mM sodium phosphate buffer, 100 mM NaCl (pH 7.3) were quantified by IEX. Two incubation experiments were performed: starting with wild type N/N dimers (left) or asymmetrically deamidated iD/N dimers (right). Emergence of the N/N species in the latter experiment is the result of on-going monomer-dimer exchange as the deamidation reaction itself is irreversible. Global fitting of experimental data to the numerical solution of the system of differential equations 4.1-4.5 (using the dimer dissociation rate obtained in Fig. 4.12) yields a deamidation reaction rate $k_1 = 4.5 \times 10^{-7} s^{-1}$. Residuals of the curve fitting are shown in chapter 6.9.

The exponential decay rate of N/N dimers strongly depends on the buffer composition (Fig. 4.14). At neutral pH, higher sodium chloride concentrations favor deamidation whereas the effect of the phosphate buffer concentration was found to be negligible. Acidic pH values increase the half-life of the N/N species dramatically. Accordingly, GII.4 P-domains were stored in 20 mM sodium acetate buffer, 100 mM NaCl (pH 4.9) at 4 °C after purification to prevent deamidation. The observed decay rate of N/N dimers is determined both by the deamidation and the dimer dissociation reaction with the macroscopic rate constant being $2k_1 + k_{\text{off}}$ (cf. Eq. 4.1). Arguably, the dissociation rate constant k_{off} cannot be assumed to be independent of the buffer composition and, accordingly, the previously derived value (cf. Fig. 4.12) cannot be used to dissect k_1 from the "observable" macroscopic constant $2k_1 + k_{\text{off}}$. Thus, only the half-life of the Asn373 wild-type species in different buffers has been reported.

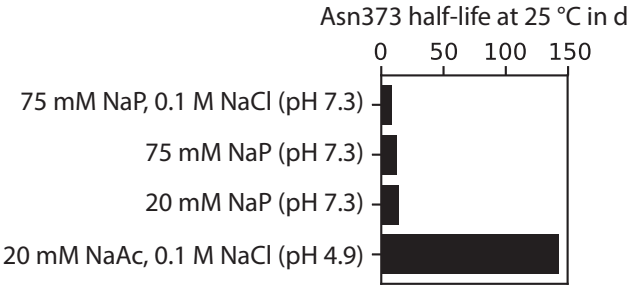


Figure 4.14: The half-life of non-deamidated GII.4 Saga P-domains depends on the buffer composition. Incubation experiments of the GII.4 Saga P-domain at 25 °C in different buffers and subsequent IEX analysis reveal strong pH and ionic strength dependencies of the half-life of the wild type N/N species. Neutral pH and high salt content favor the decay of the wild type species. The influence of the sodium phosphate buffer (NaP) itself is negligible. Half-lives are longest under acidic conditions in sodium acetate (NaAc) buffer. Deamidation reaction rates k_1 cannot be calculated because the dimer dissociation rate k_{off} cannot be assumed to be independent of the buffer composition. Therefore, only half-lives $ln(2)/(2k_1 + k_{\text{off}})$ can be given as obtained from applying an exponential decay model to the decrease in N/N dimers.

4.3 A Novel, Low-Affinity Binding Pocket for Bile Acids on GII.4 Norovirus Capsids

GII.4 Saga P-domains bind cholic acid, a representative bile acid (cf. Fig. 3.3), as evidenced by significant chemical shift perturbations (CSPs) upon titration to [*U*-^{2}H,^{15}N]-labeled protein (Fig. 4.15a). With an almost complete backbone signal assignment at hand, Euclidean CSPs can be mapped onto a crystallographic model of the P-domain dimer (Fig. 4.15c). CSPs are centered near the C-terminus of the protein with significant effects in three different peptide stretches of the protein (Fig. 4.15b). The largest CSPs can be found in signals from the surface-exposed amino acids 503-509, revealing a new binding pocket for bile acids on the human NoV capsid surface 30 Å apart from the HBGA binding site. The symmetry of the dimeric P-domain is not being broken by cholic acid-binding, as evident from the TROSY HSQC spectra. Therefore, it can be assumed that there are two symmetrical binding pockets per dimer.

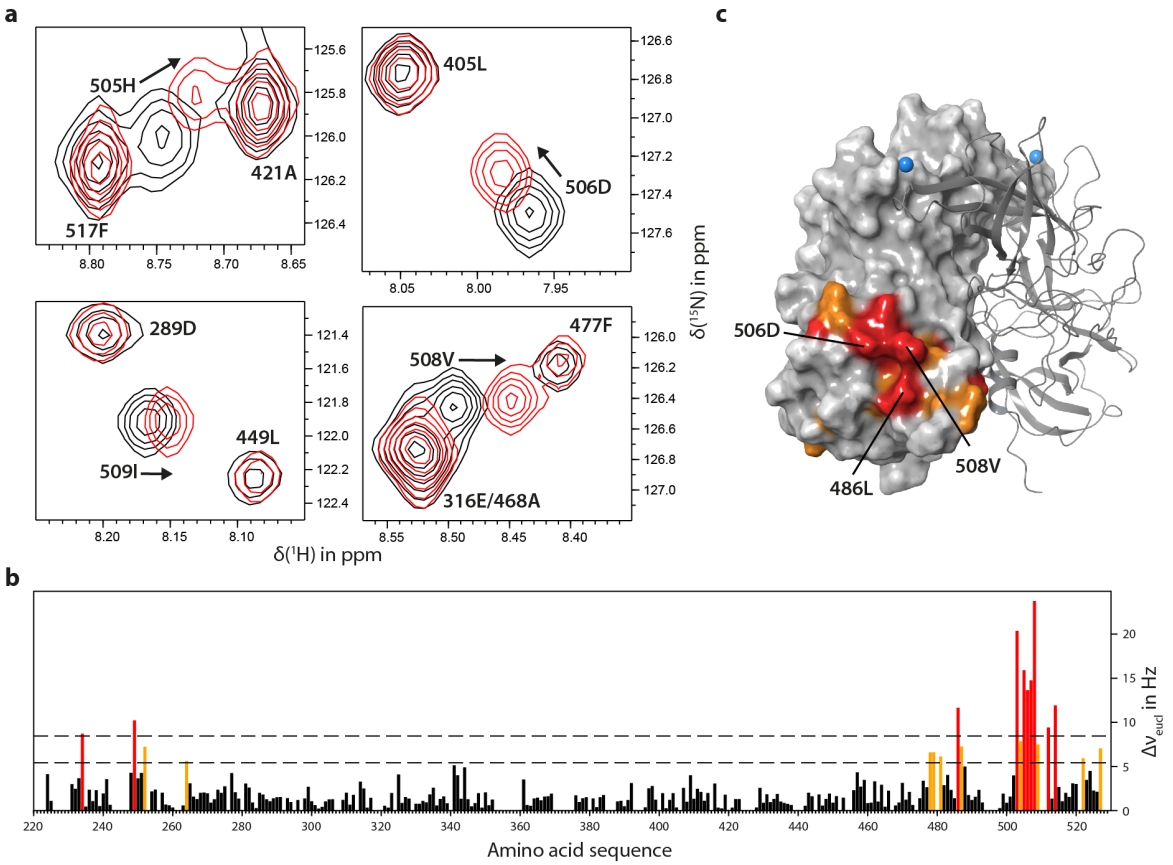


Figure 4.15: GII.4 NoV P-domains bind cholic acid. Addition of 8 mM cholic acid (CA) to the [*U*-^{2}H,^{15}N]-labeled GII.4 Saga P-domain causes significant chemical shift perturbations (**a**, spectrum after ligand addition in red). Exemplary regions are shown with arrows denoting the direction of CSPs. Euclidean CSPs can be calculated (**b**) and were mapped on a crystallographic model using the signal assignment (**c**, pdb 4X06). CSPs larger than one standard deviation are indicated in orange, those larger than two standard deviations in red. CSPs are centered around a region near the C-terminus of the protein, indicating a well-defined binding pocket for CA. For orientation, the canonical glycan binding pocket is depicted as a blue sphere, marking C6 of the bound fucose. The CSP experiment was performed in 75 mM sodium phosphate buffer, 100 mM NaCl (pH* 7.3) with a protein concentration of 200 µM. Figure adapted from Fig. 1 in Creutznacher et al. (2019).

As a complementary technique, saturation transfer difference experiments (STD) have been employed to verify cholic acid binding (Fig. 4.16). Observation of intensity changes of ligand signals upon saturation of protein resonances yields the binding epitope of the ligand. A STD spectrum of cholic acid in the presence of GII.4 Saga P-domains shows saturation transfer to almost all cholic acid signals (Fig. 4.16b). As seen in a 1D reference spectrum (Fig. 4.16a), most cholic acid signals show strong overlap and, thus, no attempt was made to map quantitative STD effects onto the ligand. However, it is evident that the complete cholic acid molecule makes contact with the protein surface – the affected, isolated signals indicated in Fig. 4.16a are well-distributed across the the ligand structure (cf. Fig 3.3). A negative control experiment without protein shows almost no STD effects (Fig. 4.16c), corroborating a specific interaction between cholic acid and the human NoV P-domain.

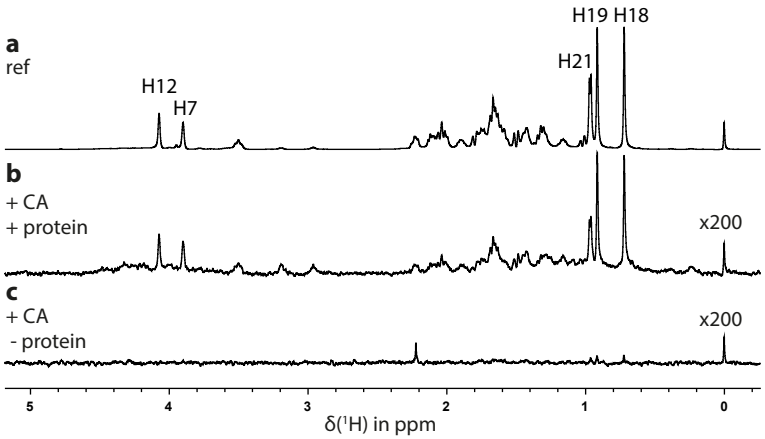


Figure 4.16: A saturation transfer difference (STD) NMR experiment confirms cholic acid binding. Comparison of a reference spectrum (**a**) with the STD spectrum (**b**) shows that almost all cholic acid (CA) protons receive saturation. A negative control without protein shows almost no direct irradiation effects (**c**). Spectra were acquired at 298 K with 90 µM GII.4 Saga 373D P-domain in 20 mM deuterated sodium phosphate buffer and 1 mM CA at 600 MHz for 17 h (4240 scans). Figure adapted from Fig. S10 in Creutznacher et al. (2019).

To test, whether the identified binding pocket is specific for cholic acid, other bile acids have been titrated to the [*U*-^{2}H,^{15}N]-labeled GII.4 Saga P-domain as well (Fig. 4.17). A selection of primary and secondary bile acids (cf. chapter 3.1.5) causes almost identical CSP profiles as cholic acid (cf. Fig. 4.15). Deoxycholic acid (Fig. 4.17a), glycochenodeoxycholic acid (GCDCA, Fig. 4.17b), and chenodeoxycholic acid (Fig. 4.17c) are all recognized by the protein and share the same binding pocket.

Titration experiments with different GII.4 NoV strains show that bile acid recognition is a conserved feature among NoV of this genogroup. Experiments have been performed with P-domains from the homologous human NoV strains GII.4 VA387 and GII.4 MI001 (Fig. 4.18). Protein backbone assignments cannot be readily transferred onto other proteins even in situations of high sequence identity (cf. Fig. 4.4). Comparison with the cholic acid titration to GII.4 Saga reveals an overall similar magnitude of CSPs, hinting at similar ligand affinities in all strains. In an isolated region of ^{1}H,^{15}N TROSY HSQC spectra of the respective NoV strains, a very similar CSP pattern can be discerned. The signals, most likely corresponding to 486L in all three strains, shift in a similar direction after bile acid addition as in the assigned reference strain GII.4 Saga. All amino acids with significant CSPs in the GII.4 Saga reference experiment are conserved in the other GII.4 strains investigated (Fig. 4.18b).

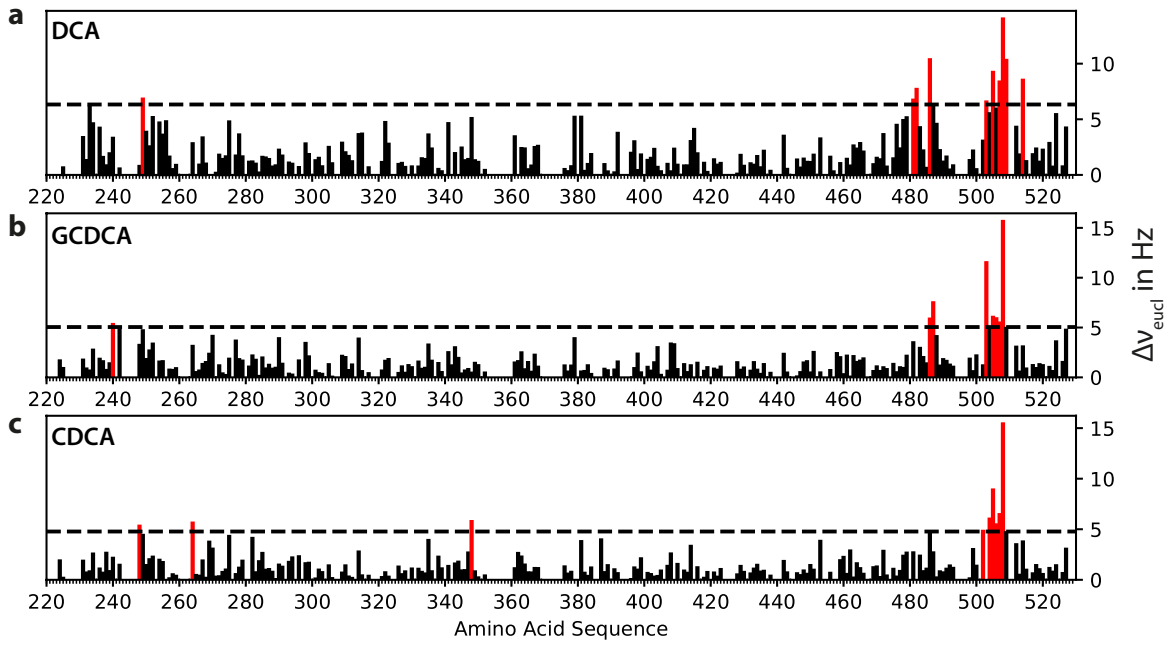


Figure 4.17: The NoV P-domain recognizes a variety of different bile acids. 2 mM deoxycholic acid (DCA, **a**), glycochenodeoxycholic acid (GCDCA, **b**) and chenodeoxycholic acid (CDCA, **c**) have been titrated to the [*U*-^{2}H,^{15}N]-labeled GII.4 Saga P-domain and Euclidean CSPs have been calculated. CSPs larger than two standard deviations are depicted in red. All bile acids cause a very similar CSP profile and, therefore, likely share the same binding pocket near the protein's C-terminus. Corresponding NMR spectra have been acquired in 75 mM sodium phosphate buffer, 100 mM NaCl (pH* 7.3) and a protein concentration of 100 µM. Figure adapted from Fig. S1 in Creutznacher et al. (2019).

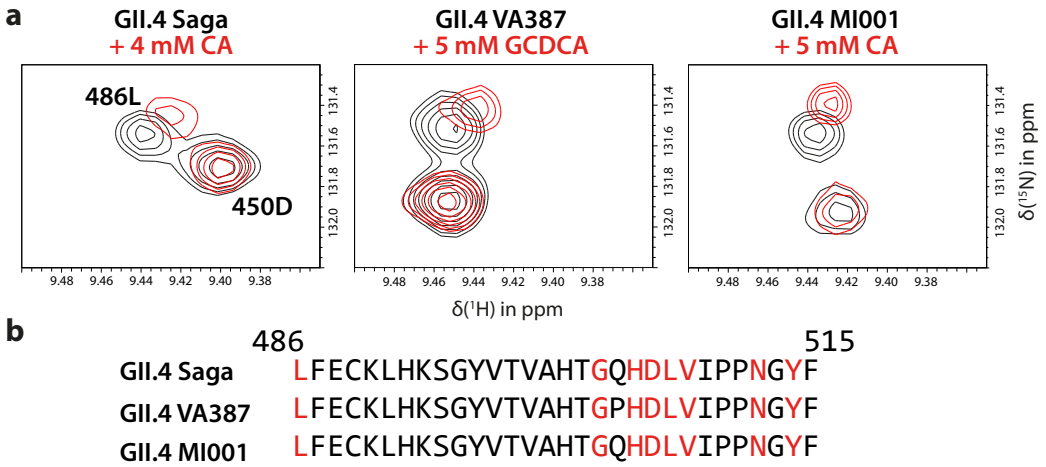


Figure 4.18: Low-affinity bile acid recognition is a conserved feature among GII.4 P-domains. P-domains from the different NoV strains GII.4 Saga, GII.4 VA387 and GII.4 MI001 all show similar CSPs when titrated with cholic acid or glycochenodeoxycholic acid (**a**). A signal assignment is only available for GII.4 Saga but can be transferred onto the other strains for some isolated signals (cf. Fig. 4.4). The amide signal of 486L appears at the same position in all different protein spectra and is part of the bile acid binding pocket in GII.4 Saga. Titration of bile acids causes CSPs similar in magnitude and direction in the other GII.4 strains as well, indicating a conserved binding pocket. All amino acids with significant CSPs in a cholic acid titration (cf. Fig. 4.15) are conserved among the three strains and are highlighted in red in a sequence alignment (**b**). Data was acquired in 75 mM sodium phosphate buffer, 100 mM NaCl (pH* 7.3) with the following protein concentrations: 200 µM GII.4 Saga, 170 µM GII.4 VA387, and 240 µM GII.4 MI001. GII.4 MI001 P-domains were produced as part of the supervised Bachelor's thesis of G. Wallmann (Wallmann, 2019). GII.4 VA387 P-domains were denatured and refolded by P. Ogrissek as part of supervised work as a research assistant.

The affinity of the GII.4 Saga P-domain towards cholic acid and GCDCA has been investigated with titration experiments (Fig. 4.19). Both molecules are low-affinity ligands with dissociation constants K_D in the mM-range. A similar K_D value of 9 mM for cholic acid can be derived by both CSP and STD titration experiments (Fig. 4.19a and b). GCDCA appears to have a slightly higher affinity with a K_D of 1.6 mM (Fig. 4.19d). However, both ligands have a tendency for self-aggregation at higher concentrations as evidenced by concentration-dependent chemical shift changes in the ligands' 1D proton signals (Fig. 4.19c and e). The bile acid concentrations at which chemical shift changes first appear, approximately 1-4 mM, can be interpreted as *critical micelle concentrations* (CMC) above which bile acid molecules exist as a mixture of monomers and higher-order aggregates. As it cannot be ruled out that binding of bile acid aggregates occurs in parallel, all K_D values reported here should be considered as *apparent* K_D values. It is noteworthy that chemical shift perturbations and saturation transfer already occur at bile acid concentrations well below the CMC.

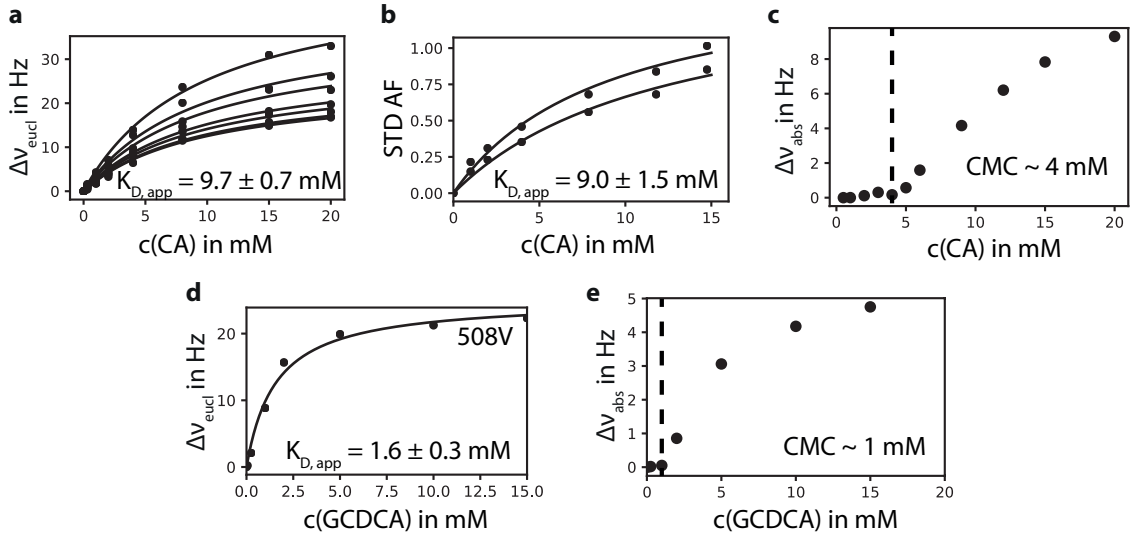


Figure 4.19: Bile acids are low-affinity ligands of GII.4 NoV. Cholic acid (CA) and glycochenodeoxycholic acid (GCDCA) have been titrated to the GII.4 Saga P-domain to characterize their binding affinity. CA binds with a weak affinity of $K_D = 9$ mM as judged from independent global fitting of data from CSP and STD experiments (**a** and **b**, respectively). GCDCA shows a slightly higher affinity $K_D = 1.6$ mM (**d**). Both CA and GCDCA show concentration-dependent chemical shift perturbations in their proton NMR signals. This indicates micelle formation at higher ligand concentrations. Critical micelle concentrations (CMC) can be derived from their H19 ligand-signal CSPs and were found to be 4 mM and 1 mM for CA and GCDCA, respectively (**c**, **e**). Affinity data can only give *apparent* K_D values as both binding and ligand-aggregation occur in similar concentration ranges. CSP experiments were acquired with 100 µM (**d**) or 200 µM (**a**) protein samples in 75 mM sodium phosphate buffer, 100 mM NaCl (pH 7.3). STD experiments were performed at 278 K on a 600 MHz NMR spectrometer with 90 µM GII.4 Saga 373D P-domain in 20 mM fully deuterated sodium phosphate buffer (pH* 7.47). STD amplification factors have been calculated for protons H7 and H12 of CA. CSP experiments for CA and GCDCA were performed at 298 K on a 500 MHz NMR spectrometer. CSP data for CA larger than two standard deviations was used for global fitting. CSPs for GCDCA were found to be of lower overall magnitude and only the strongest CSP has been used for data fitting. Figure adapted from Figs. 3, S4, and S9 in Creutznacher et al. (2019).

4.4 Bile Acids Functionalize The Murine Norovirus Protruding Domain in Solution

Protruding domains of the murine NoV capsid can be synthesized and isotope-labeled similarly to the homologous P-domains of human NoV. However, acquisition of a $^{1}H,^{15}N$ TROSY HSQC spectrum of the murine NoV P-domain reveals poor protein stability: whereas the human NoV P-domain is stable for prolonged periods of time in buffers at neutral pH, the murine NoV P-domain rapidly degrades under these conditions (Fig. 4.20). The protein yield after purification decreases accordingly (3 mg [U-$^{2}H,^{15}N$]-labeled protein per liter cell culture). The respective TROSY HSQC spectrum shows almost no dispersed backbone amide signals and substantial line broadening (Fig. 4.20a). Proton chemical shifts are mostly centered around the random coil-value of 8.3 ppm (Rehm et al., 2002), indicating the lack of a stable tertiary structure. In addition, repeated measurements of a single NMR sample reveal ongoing protein unfolding and aggregation as evident from further irreversible losses in chemical shift dispersion. Surprisingly, the stability of the murine NoV P-domain strongly depends on the buffer pH. Lowering the sample pH to 6.0 enables the acquisition of well-dispersed TROSY HSQC spectra with little signs of protein aggregation even after several measurements (Fig. 4.20b).

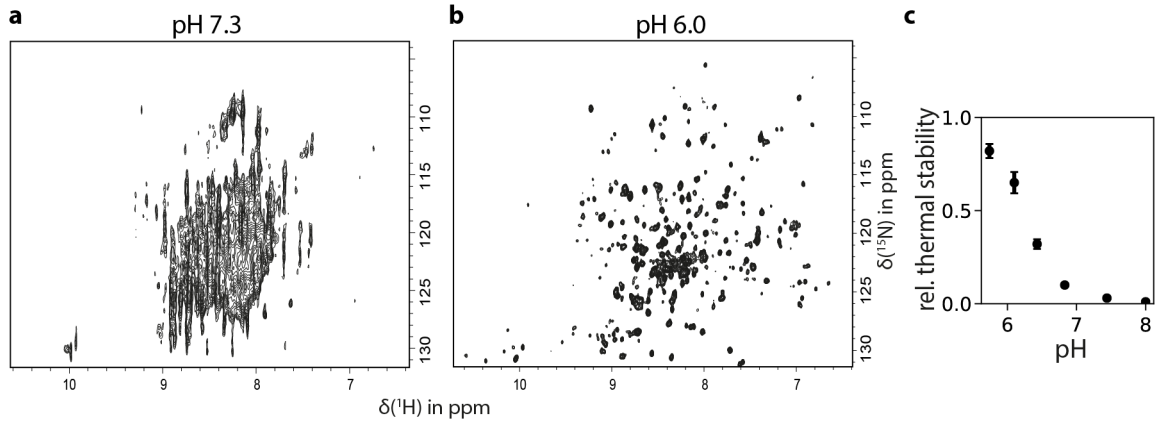


Figure 4.20: pH-dependent stability of murine NoV P-domains. Purification of murine NoV P-domains in buffers at pH 7.3 leads to irreversible protein unfolding and aggregation (**a**). A TROSY HSQC spectrum of [U-$^{2}H,^{15}N$]-labeled MNV07 P-domain in 20 mM sodium phosphate buffer (pH* 7.3) shows little chemical shift dispersion and strong line broadening. In contrast, samples at pH* 6 give spectra with dispersed, sharp NH signals indicating a well-ordered and stable protein (**b**). The thermal stability of murine NoV P-domains decreases with increasing pH (**c**). P-domains from the strain CR10 were subjected to isothermal denaturation in 75 mM sodium phosphate buffer, 100 mM NaCl at different pH values at 45 °C and subsequent hydrophobic interaction chromatography (HIC). The UV absorption in HIC experiments can be used to quantify the amount of non-denatured protein. UV integrals were normalized against a non-heat-treated control. Data shown here were acquired as part of the supervised Bachelor's thesis and work as a research assistant of C. Feldmann (Feldmann, 2019). Protein concentrations were 15 µM (**a**) and 30 µM (**b**).

To investigate the apparent pH-dependent stability systematically, the murine NoV P-domain was subjected to isothermal denaturation experiments (Fig. 4.20c). Incubation of the P-domain at 45 °C for 10 min in phosphate buffers with pH-values ranging from 5.7 to 8.0 leads to protein degradation. The residual amount of natively folded protein can be quantified by subjecting the protein sample to *hydrophobic interaction chromatography* (HIC) experiments after the forced denaturation (cf. chapter

6.8 for exemplary HIC data). Indeed, thermal stability sharply decreases when the buffer pH is increased from slightly acidic conditions to neutral values and above. Further NMR experiments were thus performed under acidic conditions in acetate buffers at pH 5.3.

TROSY HSQC spectra of the murine NoV P-domain change with protein concentration (Fig. 4.21). Higher protein concentrations lead to the appearance of new $^{1}H,^{15}N$ signals not visible at lower concentrations. Likewise, other signals decrease in relative intensity at higher concentrations. The spectra remain well-dispersed with none of the spectral characteristics of protein aggregation (cf. Fig. 4.20). This concentration-dependent behavior indicates homodimerization of the murine NoV P-domain at high protein concentrations. This is in stark contrast to human NoV P-domains which readily dimerize almost independent of the protein concentration used (cf. Fig. 4.12). Inspection of the TROSY HSQC spectra reveals several signal pairs in close proximity in which one signal decreases and the other signal increases in relative intensity (Fig. 4.21a, right). These signals can be attributed to the respective monomeric and dimeric protein species. The appearance of separate resonances for monomers and dimers in slow exchange can be used to estimate an upper limit for the monomer-dimer exchange rate of $k_{ex} << 10$ Hz based on the respective chemical shift differences (cf. chapter 3.3.1).

Size exclusion chromatography (SEC) experiments corroborate the existence of a concentration-dependent dimerization reaction (Fig. 4.21b). Samples of the murine NoV P-domain were subjected to SEC experiments at different concentrations. Higher protein concentrations lead to a shift of the protein's UV absorption maximum to smaller elution volumes, i.e. larger hydrodynamic radii. Apparently, monomers and dimers interconvert too fast to be separated by SEC experiments and elute with an intermediate elution volume instead. Normalization of UV chromatograms shows that, in addition to the elution volume, the peak shape also changes at higher protein concentrations (Fig. 4.21 right). A noticeable peak tailing occurs at concentrations higher than 6 µM. This concentration can be used to derive an estimate for the dimer association rate k_{on} of 860 M^{-1} s^{-1} as explained in chapter 6.3.3.

Using the SEC elution behavior of standard proteins, a linear relationship can be derived between elution volumes and apparent molecular weights. The apparent molecular weight of the murine NoV P-domain species shifts from 39.6 kDa at the lowest protein concentration to 58.4 kDa at the highest concentration used (Fig. 4.21c). This is in good agreement with the molecular weights of the monomeric and dimeric P-domain (33.2 kDa and 66.4 kDa, respectively). The relationship between the concentration of the protein sample loaded onto the column and the apparent molecular weight can be used to derive the dissociation constant K_D of the dimerization reaction based on the law of mass action, resulting in an apparent K_D value of 40 µM. Similarly, the absolute intensity of dimer signals in TROSY HSQC spectra at different total protein concentrations can be used to obtain an independent estimate of the dissociation constant (Fig. 4.21c, right). Intensity data of nine dimer signals were used to derive a K_D value of 31 µM (cf. chapter 6.4.8 for details). Using the value for the association rate constant k_{on} obtained above by chromatography, a monomer-dimer exchange rate k_{ex} can be estimated (Eq. 6.5). For the NMR experiment series, the total protein concentration was varied between 25 and 200 µM. In this range, k_{ex} is 0.4-0.7 Hz and, thus, in good agreement with the upper limit for k_{ex} based on the chemical shift difference.

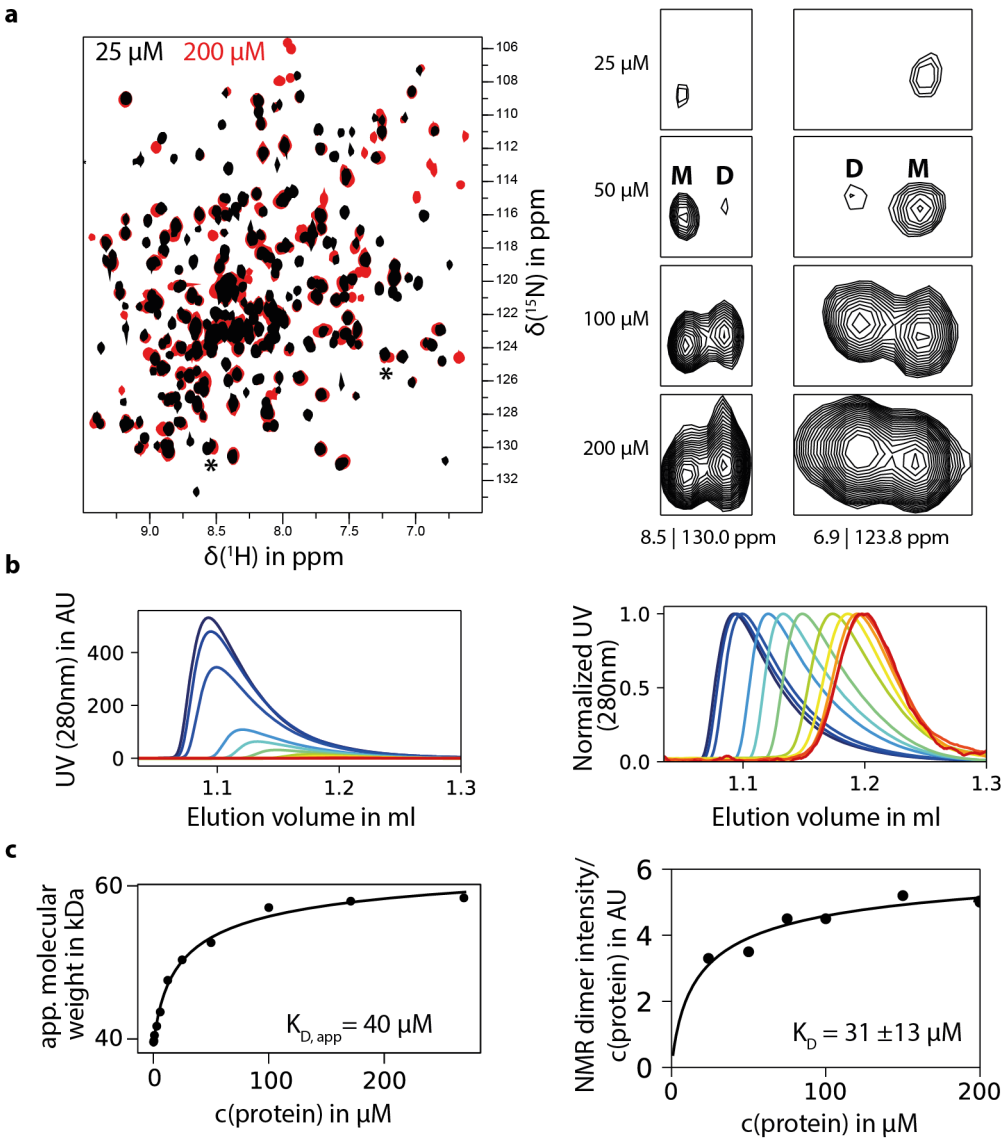


Figure 4.21: Concentration-dependent dimerization of the murine NoV P-domain in solution. TROSY HSQC spectra of the [*U*-^{2}H,^{15}N]-labeled CW1 P-domain show protein concentration-dependent changes (**a**, black: 25 µM, red: 200 µM). Several new backbone NH resonances appear with increasing protein concentrations while others decrease in relative intensity. These signals can be assigned to the homodimer (D) or the monomer (M), respectively. Two such signal pairs are marked by asterisks and are shown in detail (right). The murine NoV P-domain has been subjected to size exclusion chromatography at increasing concentrations (**b**), ranging from 250 nM (red) to 270 µM (blue). The elution volume strongly depends on the protein concentration and can be associated with an increase of apparent molecular weights from 39.6 kDa (monomer) to 58.4 kDa (dimer). Normalized chromatograms (right) reveal a noticeable peak tailing at higher concentrations, indicating ongoing monomer-dimer exchange on the time scale of the chromatographic separation. Fitting the concentration-dependent apparent molecular weight against the law of mass action yields an apparent K_D value of the dimerization reaction of 40 µM (**c**), matching the fit of the averaged NMR dimer signal intensity ($n = 9$) that results in $K_D = 31 \pm 13$ µM. To obtain the K_D from NMR data, normalized intensities were divided by the photometrically determined protein concentration and fitted to the law of mass action (Eq. 6.12). Spectra were acquired in 20 mM sodium acetate, 100 mM NaCl (pH* 5.3).

As the murine NoV P-domain does not readily dimerize in solution, it was necessary to investigate whether the isolated P-domain is actually able to adopt its functional, binding-competent form in solution. The bile acid glycochenodeoxycholic acid (GCDCA) is a known ligand of the murine NoV P-domain with two binding pockets per P-domain dimer (cf. chapter 3.1.5). The presence of GCDCA dramatically changes various biophysical properties of the P-domain, thereby verifying that the protein is indeed able to recognize its ligand in solution (Fig. 4.22). Addition of GCDCA to the running buffer in size exclusion experiments prevents the protein concentration-dependent elution behavior described above (Fig. 4.22a). The apparent molecular weight of the P-domain remains constant at 64 kDa independent of the protein concentration used. This experimental value matches the molecular weight of the dimer, 66 kDa, very well. The binding of GCDCA appears to shift the homodimerization equilibrium almost completely to the side of the dimeric protein species. Consequently, peak tailing – indicating fast monomer-dimer exchange – cannot be observed anymore.

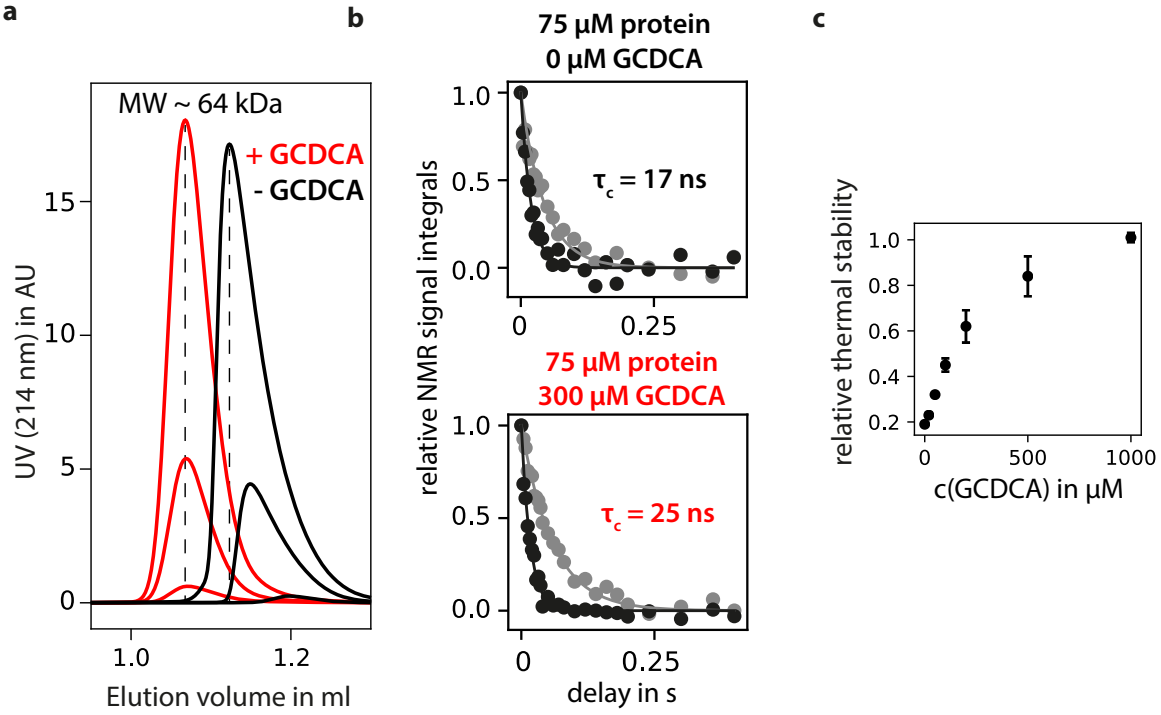


Figure 4.22: Murine NoV P-domains readily dimerize in presence of glycochenodeoxycholic acid (GCDCA). Size exclusion chromatograms of the CW1 P-domain in the presence and absence of GCDCA at varying protein concentrations (**a**, red and black curves, respectively) reveal a major change in elution behavior. The presence of saturating amounts of GCDCA shifts the apparent molecular weight of the P-domain to that of the dimer (64 kDa) regardless of the protein concentration used (250 nM, 12.5 µM, and 50 µM). The disappearance of the protein concentration-dependent elution behavior indicates a dramatic stabilization of the dimeric protein in presence of GCDCA. TRACT NMR experiments measure average ^{15}N relaxation rates of backbone amide signals (**b**, black and gray curves, respectively) and, thereby, rotational correlation times τ_c. In the absence of GCDCA, 75 µM [*U*-^{2}H,^{15}N]-labeled murine NoV P-domains form a mixture of monomers and dimers at a ratio of 1:1 (using a K_D value of 31 µM, cf. Fig. 4.21). Accordingly, the measured τ_c corresponds to the averaged value of monomers and dimers. Adding saturating amounts of GCDCA leads to an increase in τ_c corroborating the complete dimerization of the P-domain sample through GCDCA addition. GCDCA increases the thermal stability of the P-domain in a concentration-dependent manner (**c**). Performing isothermal denaturation for 30 min at 45 °C and analysis of non-denatured P-domain via HIC shows complete protection with 1 mM GCDCA when compared to a non-heat-treated control. HIC data were acquired by C. Feldmann as part of supervised work as a research assistant. NMR samples for TRACT experiments contained 20 mM sodium acetate, 100 mM NaCl (pH* 5.3).

The average rotational correlation time τ_c of [*U*-^{2}H,^{15}N]-labeled proteins can be measured via TRACT NMR experiments and is sensitive towards changes in hydrodynamic radius and thereby to changes in molecular weight. Indeed, addition of GCDCA shifts the average τ_c of a 75 µM protein solution from 17 ns to 25 ns (Fig. 4.22b). As the protein forms a mixture of monomers and dimers in the absence of GCDCA of approximately 1:1 (assuming a dimerization K_D of 31 µM, cf. Fig. 4.21), the respective correlation time of 17 ns is only a description of the ensemble average. In contrast, the measured value in the presence of GCDCA should almost exclusively reflect the properties of the dimeric, ligand-bound protein.

The thermal stability of the murine NoV P-domain under different solution conditions can be assessed by isothermal denaturation and subsequent quantification of residual non-denatured protein using hydrophobic interaction chromatography (cf. Fig. 4.20 and chapter 6.8). Incubating the protein under acidic conditions with increasing amounts of GCDCA shows a strong protection of ligand-binding against thermal denaturation (Fig. 4.22c). The protective effect is dose-dependent; at a GCDCA concentration of 1 mM the P-domain is able to withstand thermal denaturation for 30 min at 45 °C without losses. Qualitatively, heightened protein stability in presence of GCDCA can also be observed at lower temperatures: prolonged NMR experiments with the unbound protein result in visible precipitation, limiting the lifetime of these protein samples even under stabilizing acidic conditions. No such denaturation could be observed after NMR experiments in the presence of GCDCA.

To further elucidate the molecular basis of these phenomena, GCDCA was titrated stepwise to [*U*-^{2}H,^{15}N]-labeled murine NoV P-domains and a series of TROSY HSQC spectra was acquired (Fig. 4.23). At the given protein concentration of 100 µM, the P-domain forms a mixture of monomers and dimers with a ratio of 1:2 in the absence of GCDCA (again, assuming a K_{D1} of 31 µM). When comparing the beginning and end point of the GCDCA titration, it is evident that, without exception, all backbone amide signals are subject to major changes in either signal position, signal intensity or both (Fig. 4.23a). Particularly, all dimer signals increase in signal intensity with increasing GCDCA concentrations until a stationary value is reached. A small subset of dimer signals additionally displays chemical shift perturbations (CSPs) (Fig. 4.23b). In contrast, all signals attributed to the monomer decrease in intensity until they are no longer visible (Fig. 4.23b, c). Monomer signals were not found to show CSPs.

Taken together, this suggests a coupled equilibrium of dimerization and two consecutive, independent ligand-binding reactions as the simplest possible model (Fig. 4.23d). The exchange between monomers and dimers is slow on the proton chemical shift time scale. In contrast, ligand-binding induces CSPs and is, therefore, a fast-exchange process. A number of signals only become visible after GCDCA is added (Fig. 4.23c). These signals reflect only the bound dimer species as the corresponding unbound dimer signals are invisible. The monomeric P-domain does not appear to be binding-competent as only dimer signals display CSPs. This is in agreement with the observed disappearance of monomer signals when an equimolar protein to ligand ratio is reached. Removal of unbound dimers from the reaction equilibrium by binding causes further monomers to dimerize until they become completely depleted.

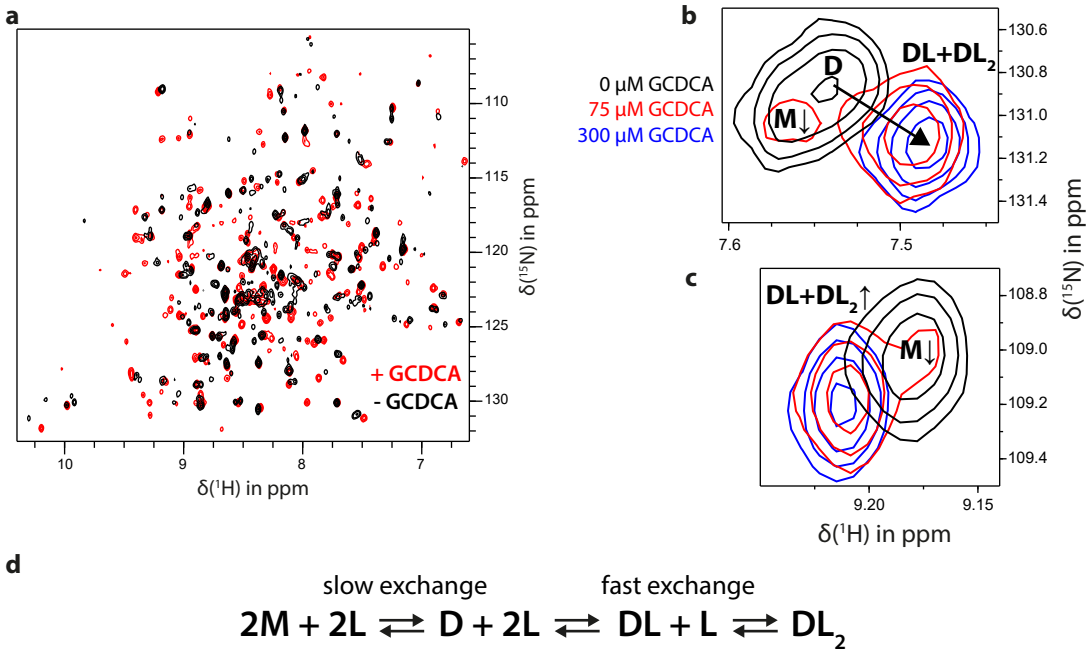


Figure 4.23: Coupled equilibria of dimerization and ligand binding of murine NoV P-domains. TROSY HSQC spectra of 100 µM [*U*-^{2}H,^{15}N]-labeled CW1 P-domains with and without GCDCA show major overall changes (**a**). Exemplary sections of the spectra are shown in **b** and **c**. Signals attributable to the monomer (M) disappear with increasing GCDCA concentrations, whereas dimer signals (D) show an increase in signal intensity. This is indicated by upward and downward facing arrows. Additionally, some dimer signals display chemical shift perturbations (CSPs) with increasing GCDCA concentrations (**b**). CSP and signal intensity data can be explained with a coupled equilibrium of dimerization and consecutive, independent ligand binding reactions (**d**). As visible in the apo spectrum (black), monomers and unbound dimers are in slow exchange and cause separate signals. The occurrence of CSPs upon ligand addition is indicative of a fast exchange ligand binding process. Spectra were acquired with samples in 20 mM sodium acetate, 100 mM NaCl (pH* 5.3)

To obtain the GCDCA dissociation constant K_{D2}, it is necessary to derive an equation that describes the momentary concentration of one involved protein species as a function of the total ligand concentration using the dimerization and ligand-binding dissociation constants.

$$2\,M + 2\,L \overset{K_{D1}}{\rightleftharpoons} D + 2\,L \overset{K_{D2}}{\rightleftharpoons} DL + L \overset{K_{D2}}{\rightleftharpoons} DL_2$$

Here, M represents monomeric P-domains, L is the ligand GCDCA, and D is the P-domain dimer. The dissociation constants K_{D1} and K_{D2} can then be defined according to the respective equilibria:

$$K_{D1} = \frac{[M]^2}{[D]}, K_{D2} = \frac{[D][L]}{[DL]} = \frac{[DL][L]}{[DL_2]} \tag{4.6}$$

Eq. 4.7-4.8 describe the mass balance of ligand and protein molecules.

$$L_t = [L] + [DL] + 2[DL_2] \tag{4.7}$$

$$P_t = [M] + 2[D] + 2[DL] + 2[DL_2] \tag{4.8}$$

Rearranging Eq. 4.6 and substituting the $[DL]$ and $[DL_2]$ terms in Eq. 4.7 yields:

$$0 = \frac{2[D][L]^2}{K_{D2}^2} + \left(1 + \frac{[D]}{K_{D2}}\right)[L] - L_t \tag{4.9}$$

Substitution of $[D]$ in Eq. 4.9 with the dimerization term of Eq. 4.6 results in an equation describing the free ligand concentration $[L]$ as a function of K_D values, the total ligand concentration L_t and the monomer concentration $[M]$.

$$[L] = -\frac{K_{D2}^2 + \frac{K_{D2}[M]^2}{K_{D1}}}{\frac{4[M]^2}{K_{D1}}} + \sqrt{\left(\frac{K_{D2}^2 + \frac{K_{D2}[M]^2}{K_{D1}}}{\frac{4[M]^2}{K_{D1}}}\right)^2 + \frac{L_t K_{D2}^2}{\frac{2[M]^2}{K_{D1}}}} \tag{4.10}$$

Similarly, substitution of $[D]$, $[DL]$, and $[DL_2]$ terms in Eq. 4.8 leads to Eq. 4.11.

$$0 = [M] + \frac{2[M]^2}{K_{D1}} + \frac{2[M]^2}{K_{D1}K_{D2}}[L] + \frac{2[M]^2}{K_{D1}K_{D2}^2}[L]^2 - P_t \tag{4.11}$$

In Eq. 4.11, $[L]$ can be substituted with Eq. 4.10 to obtain a function that relates the monomer concentration with the K_D values and the known concentrations L_t and P_t. This equation can be solved numerically for $[M]$ as detailed in chapter 6.4.8. This approach can be used to simulate the concentrations of the different protein species as a function of the total ligand concentration if K_{D1} and K_{D2} are known (Fig. 4.24). As expected from the NMR data, both monomer and unbound dimer concentrations approach zero (black and orange curves) when protein and ligand molecules are present in equimolar amounts.

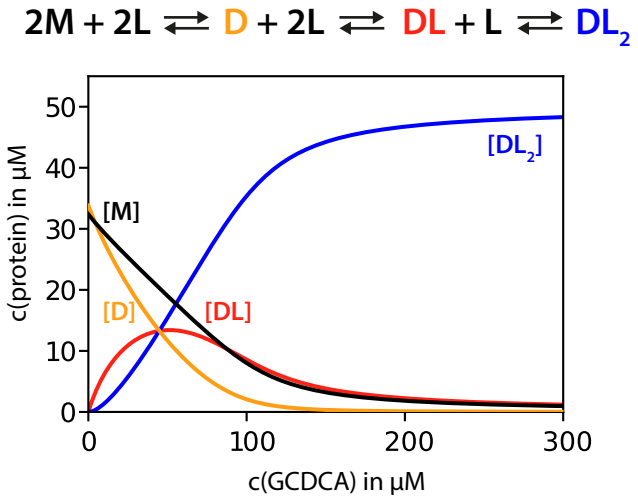


Figure 4.24: Simulation of a coupled dimerization and ligand binding equilibrium. Using the numerical solution of Eq. 4.11, the concentrations of the different protein species were determined in dependence of an increasing total ligand concentration. Simulation parameters were as follows: a total protein concentration of 100 µM, a K_{D1} of 31 µM, and a K_{D2} of 5 µM.

In order to obtain a K_{D2} value from the NMR data presented in Fig. 4.23, it is necessary to convert the two experimental NMR observables, signal intensity and signal position, into actual concentrations of the different protein species. These experimental protein concentrations can then be used for a least-squares minimization against the numerical solution of the equations derived above. However, linking NMR data to real protein concentrations is not straightforward.

As K_{D1} is known from previous experiments, so is the concentration of monomers at the beginning of the titration. Accordingly, experimental monomer concentrations $[M]$ can be obtained by

normalization of the respective monomer signal intensity at the beginning of the titration to 32 µM. All monomer signals have disappeared at the highest GCDCA concentration and all dimer signals have reached their maximum intensity as the system is saturated with the ligand. Therefore, dimer signal intensities at the titration end point reflect a dimer concentration of 50 µM. However, no class of NMR signals solely reflects the concentration of a single dimer species (D, DL, or DL_2). Instead, some dimer signal intensities contain information on the sum of concentrations of dimers bound once and twice , i.e. $[DL]+[DL_2]$. This information comes from the dimer signals that only become visible *after* GCDCA addition (Fig. 4.25a). Consequently, the unbound dimer species D does not contribute to the overall intensity of these signals. In fact, the concentration of the asymmetrically bound DL species needs to be modified with a factor of 0.5 to account for its invisible, unbound monomer.

With these experimental protein concentrations at hand, the data can now be used for curve fitting to derive the ligand dissociation constant K_{D2}. Monomer and bound dimer intensity data were used for global fitting to the numerical solution of Eq. 4.10-4.11 independently. As K_{D2} was varied, squared residuals were calculated for data from the individual signals and then summed for all signals in a data class (Fig. 4.25b). Summed residuals form a narrow minimum around a K_{D2} of 5 µM for both monomer and dimer intensity data sets. Averaged intensity data sets are in good agreement with the numerical solution assuming a K_{D1} of 31 µM and a K_{D2} of 5 µM (Fig. 4.25c).

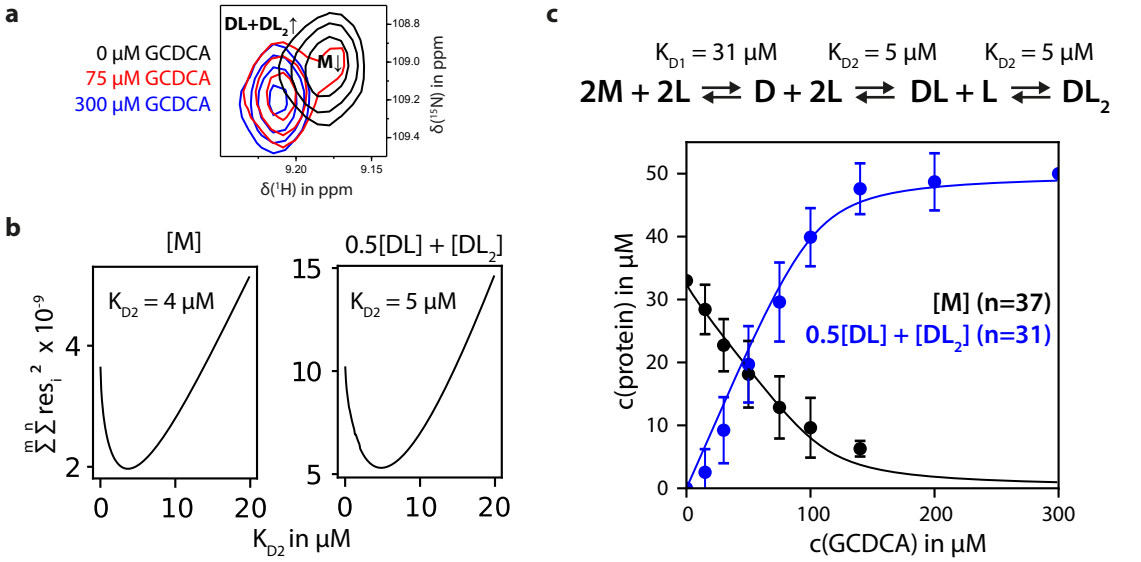


Figure 4.25: Global fitting of experimental NMR intensity data yields a ligand dissociation constant. Fitting of intensity data of the different species (M or DL/DL_2) (**a**) yields a K_D of 5 µM for GCDCA binding with steep minima in the respective sum of squared residuals (**b**). Assuming K_{D2}= 5 µM, fits are in good agreement with averaged experimental intensity data of the different signal species (**d**). Error bars reflect the standard deviations of the averaged signals from each data class. The NMR data shown in **a** are identical to data shown in Fig. 4.23.

Without a backbone signal assignment, the interpretation of CSP data is more complicated (Fig. 4.26a). There are different population-weighted averages a given CSP could potentially reflect. Chemical shift perturbations are influenced by the position of an amino acid relative to the ligand binding pocket. The geometry of the murine NoV P-domain and its two symmetrical GCDCA binding pockets (cf. Fig. 3.4) allows for two possible scenarios. Some signals will be influenced by the occupation of

both pockets corresponding to amino acids located in between them. Other amino acids will be located such that only the occupation of the closest binding pocket causes a CSP. Without a backbone signal assignment available, it is not known *a priori* to which of these two classes a particular signal belongs. Thus, data can only be interpreted in relation to these two boundary scenarios as described in detail in chapter 6.4.8 (Fig. 4.23d, red). Again, data were normalized to 50 µM at the end point. Residuals for CSP data were calculated according to the two boundary cases outlined above (Fig. 4.26b). Depending on the model, the minimum of the summed residuals appears at 1-13 µM. As expected, averaged CSP data points lie in between the two boundary curves (Fig. 4.26c).

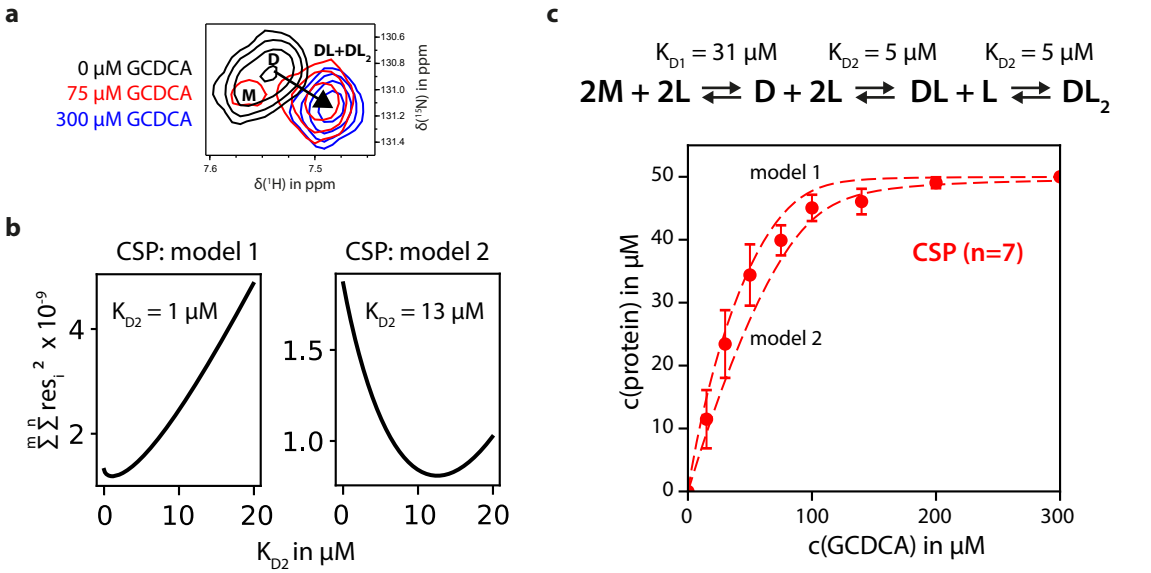


Figure 4.26: Global fitting of experimental NMR CSP data confirms the ligand dissociation constant. As there is no signal assignment available, only boundary cases can be described for CSP data: depending on their position in relation to the two GCDCA binding pockets some nuclei will experience perturbations by occupation of only one binding pocket (model 1) while others can be influenced by occupation of both pockets (model 2, cf. chapter 6.4.8). Depending on the chosen model, global fitting of CSP data gives K_D values of either 1 µM or 13 µM (**b**). As expected, averaged CSP data (**c**) lies between the theoretical curves of the two boundary cases (dashed lines). Notably, K_D values derived from intensity data (cf. Fig. 4.25) lie between these two CSP boundary values. The NMR data shown in **a** are identical to data shown in Fig. 4.23.

4.5 Protein-Metal Interactions of Norovirus Capsids

Both human and murine [*U*-^{2}H,^{15}N]-labeled NoV P-domains were subjected to titration experiments with various metal salts to investigate the interactions of the NoV capsid with metal ions. First, the human NoV P-domain from GII.4 Saga was titrated with an equimolar amount of $ZnCl_2$ leading to major overall changes in a TROSY HSQC spectrum (Fig. 4.27a).

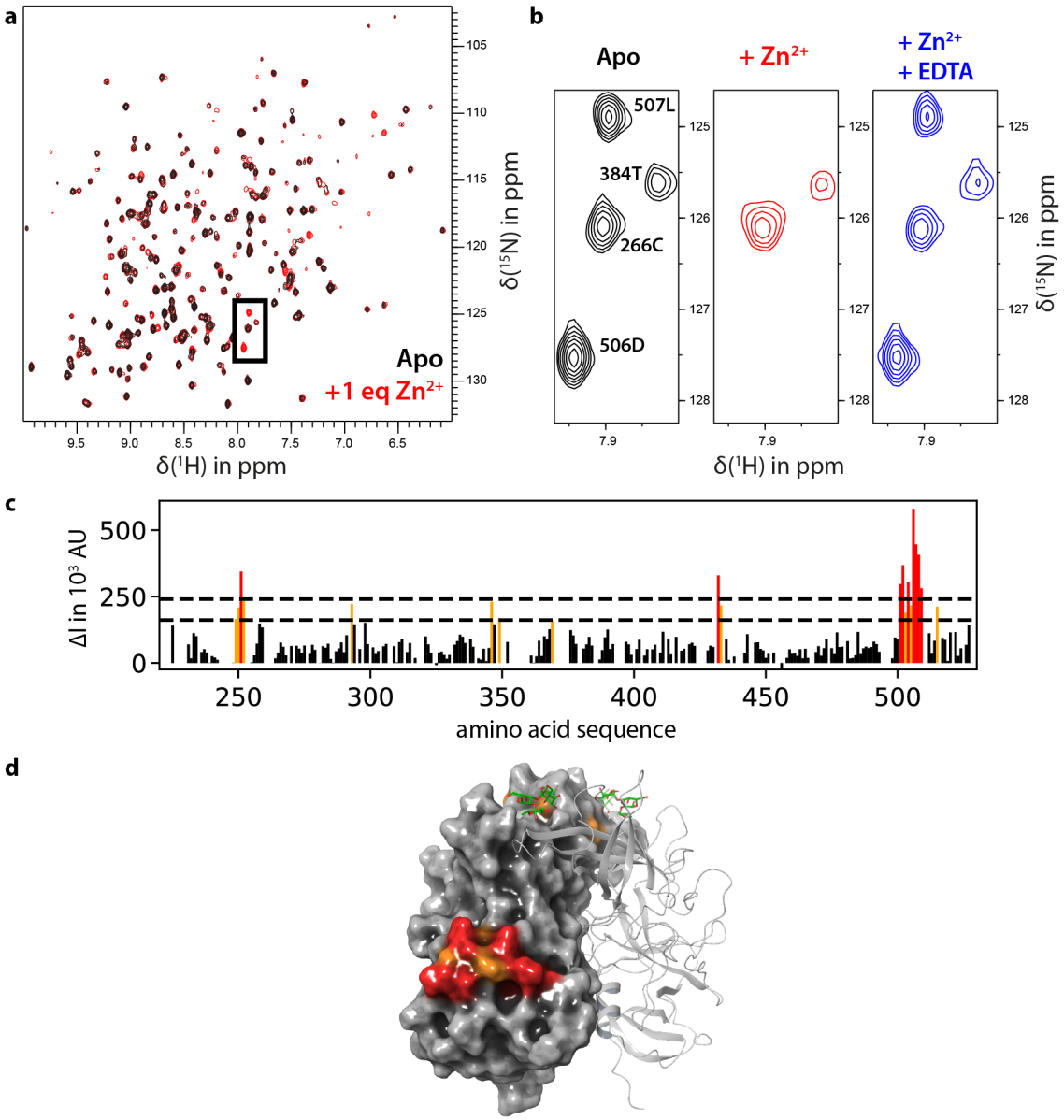


Figure 4.27: The human NoV P-domain binds Zn^{2+}. Titration of 120 µM $ZnCl_2$ to 120 µM [*U*-^{2}H,^{15}N]-labeled GII.4 Saga 373D P-domain (one equivalent assuming one binding pocket per monomer) leads to the disappearance of several NH signals (**a**). One region of the respective spectra is shown in detail in **b**. Signals of the amino acids 501-509 are no longer visible after Zn^{2+}-titration but can be recovered by complexation of Zn^{2+} with 2 mM EDTA. Calculating intensity differences between the apo and Zn^{2+} spectrum reveals at least two affected regions (**c** and **d**, pdb 4X06). Differences larger than one standard deviation are indicated in orange and those larger than two standard deviations in red. Most prominent effects can be seen near the C-terminus of the protein revealing a well-defined Zn^{2+} binding pocket. A secondary, lower affinity binding pocket can be found at the top of the GII.4 P-domain near the canonical HBGA binding pocket. For orientation, a bound B-trisaccharide molecule is depicted in green. Spectra were acquired in 75 mM Tris, 100 mM NaCl (pH* 7.3).

An appreciable number of signals is strongly reduced in intensity (Fig. 4.27b). No new resonances appear in the spectrum. This effect was quantified by calculating intensity differences ΔI between the respective spectra (Fig. 4.27c). The effects are localized in a C-terminal region of the P-domain (Fig. 4.27d) in close proximity to, but not identical with, the novel bile acid binding pocket described in chapter 4.3. Apparently, there is also a secondary, lower-affinity Zn^{2+}-binding site near the HBGA binding pocket at the outward facing part of the protruding domain showing minor effects. Complexation of Zn^{2+} with the chelating agent EDTA reverses the effects and restores signal intensities.

Similarly, the murine NoV P-domain has been titrated with both $MgCl_2$ and $CaCl_2$. The experiment was performed in presence of saturating amounts of the bile acid GCDCA to ensure complete dimerization of the protein (cf. chapter 4.4). Several amide resonances display strong CSPs upon addition of either Mg^{2+} or Ca^{2+} (Fig. 4.28, red and blue spectra, respectively). Both metal ions affect an identical set of signals, indicating that both metals likely share a binding pocket. $CaCl_2$ was titrated to a [U-^{2}H,^{15}N]-labeled sample to derive the dissociation constant K_D of the MNV-Ca^{2+} interaction. Signals shift in a linear fashion as expected for a protein-ligand system in fast exchange starting from the GCDCA-saturated apo state.

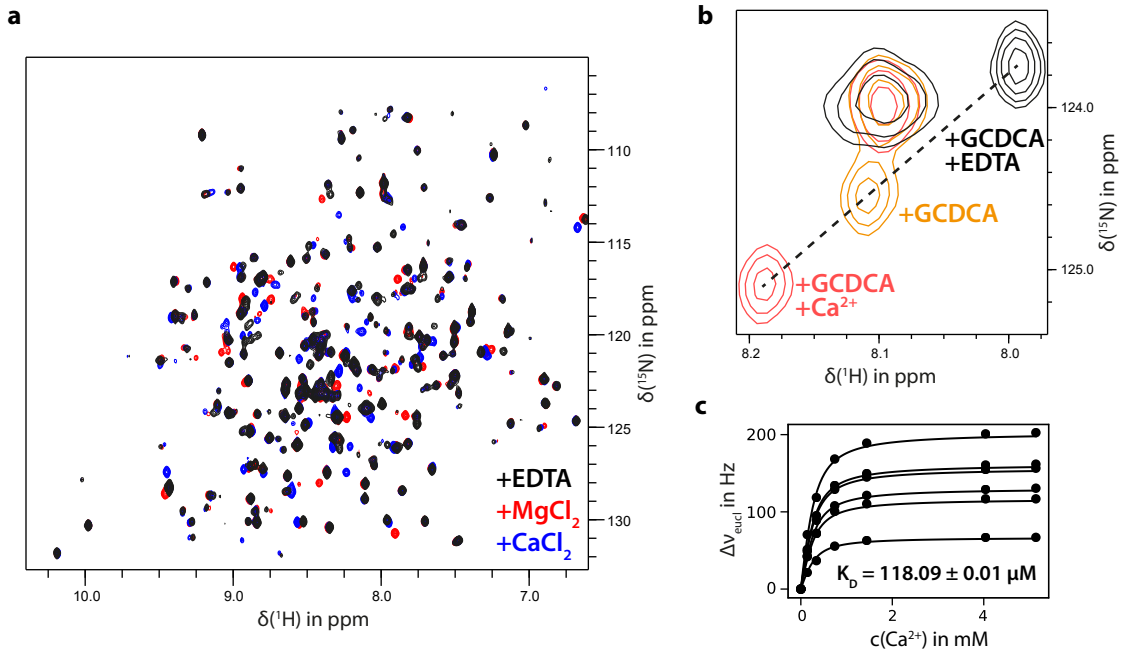


Figure 4.28: The murine NoV P-domain binds bivalent cations. Titration of 5 mM $MgCl_2$ or $CaCl_2$ to 150 µM, GCDCA-saturated [U-^{2}H,^{15}N]-labeled CW1 P-domain shows chemical shift perturbations in TROSY HSQC spectra (**a**). Identical signals are affected by both Mg^{2+} (red) and Ca^{2+} (blue) indicating a shared binding pocket. Chelation of all bivalent metals with 2 mM EDTA shifts signals in the opposite direction as the Ca^{2+} titration with the signal from the GCDCA-bound starting points in between (**b**). Apparently, GCDCA contains a high amount of contaminating calcium salts. Therefore, all GCDCA-treated samples inadvertently contain an appreciable amount of Ca^{2+} as well. Euclidean CSPs can be calculated from a Ca^{2+} titration series with the EDTA-state considered as the starting point. Due to the Ca^{2+}-contamination, the total Ca^{2+} concentration during the titration is not known precisely. Using the amount of added Ca^{2+} ions, a global least-squares minimization of CSP data larger than two standard deviations can be used to derive both the amount of initial contaminants and the dissociation constant K_D of the interaction. This approach yields a K_D of 118 µM. With 300 µM GCDCA, the initial Ca^{2+} contamination was found to be 140 µM. Spectra were acquired in 20 mM sodium acetate, 100 mM NaCl (pH* 5.3), Ca^{2+} was titrated to a sample containing 250 µM P-domain.

However, complexation of Ca^{2+} with EDTA does not restore the original 'apo' spectrum. Instead, all affected signals are shifted further into the opposite direction (Fig. 4.28b). This is clear evidence for a fundamental contamination with Ca^{2+} ions originating from the GCDCA stock solution. Indeed, revisiting the GCDCA titration experiment (cf. Fig. 4.23) reveals a secondary set of fast-exchange CSPs that is distinct from those from GCDCA binding. The respective signals do not reach saturation and correspond to those signals further affected by the dedicated Ca^{2+}-titration. This corroborates that GCDCA stock solutions contain a large amount of Ca^{2+} ions. To quantify the amount of contaminating Ca^{2+} as well as the K_D of the interaction itself, CSP data were fitted to the law of mass action while simultaneously fitting the amount of unknown, additional Ca^{2+} from the GCDCA stock solution (Fig. 4.28c). Peak positions from the EDTA-spectrum have been used as the starting point for the CSP calculation. Fitting yields a K_D of 118 µM and an initial Ca^{2+} concentration of 140 µM.

To test whether the binding pocket for bivalent metals in the murine NoV P-domain also accepts paramagnetic, trivalent metals, different lanthanide salts have been titrated (Fig. 4.29). Again, spectra were acquired while saturating amounts of GCDCA were present. La^{3+} has been used as a diamagnetic reference ion with similar chemical properties. Titration of $LaCl_3$ leads to chemical shift perturbations indicative of fast-exchange between metal-bound and unbound protein species.

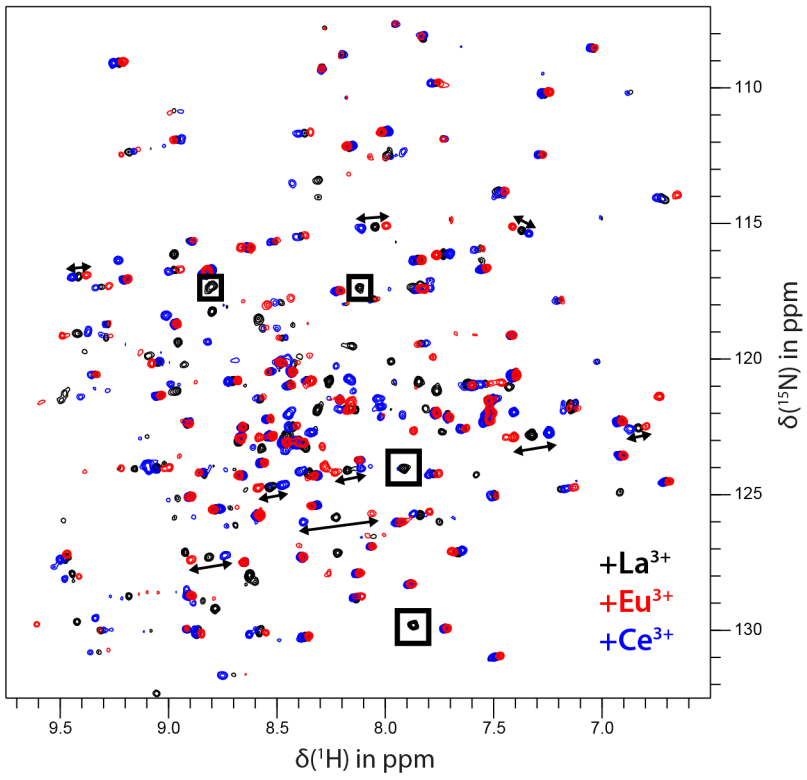


Figure 4.29: The calcium binding pocket of the murine NoV P-domain accepts paramagnetic metals. Titration of the paramagnetic lanthanide ions Eu^{3+} and Ce^{3+} to [*U*-^{2}H,^{15}N]-labeled CW1 P-domain causes pseudocontact shifts (PCS, black arrows) and paramagnetic relaxation enhancement (PRE, black boxes) in a TROSY HSQC spectrum. A spectrum with the diamagnetic La^{3+} ions was used as a reference. PRE effects lead to the disappearance of signals previously identified to constitute the binding pocket for Ca^{2+} ions, demonstrating that both lanthanides and bivalent metals share the same binding pocket. Spectra were acquired on a 600 MHz NMR spectrometer. Samples contained 50 µM protein, 300 µM GCDCA, 840 µM $CaCl_2$, and 300 µM $LaCl_3$, $EuCl_3$, or $CeCl_3$, respectively, in 20 mM sodium acetate, 100 mM NaCl (pH* 5.3).

The two paramagnetic ions Ce^{3+} and Eu^{3+} bind specifically as well and produce characteristic paramagnetic effects. Some signals are broadened beyond detection due to *paramagnetic relaxation enhancement* (PRE, black boxes), coinciding with those signals affected by Mg^{2+}- and Ca^{2+}-addition (cf. Fig. 4.28). As PRE effects have a strong distance dependence, signals of amino acids close to the paramagnetic ion are affected most. Thus, it can be concluded that lanthanide ions and bivalent metal ions share the same binding pocket. All samples were prepared with close to saturating amounts of $CaCl_2$. Considering the pronounced effects of lanthanide addition, it appears likely that lanthanide binding occurs with higher affinity than Ca^{2+} binding and that lanthanides are able to displace Ca^{2+} from its binding pocket.

A large number of signals is shifted in opposing directions by Ce^{3+} and Eu^{3+}, demonstrating *pseudocontact shifts* (PCS, black arrows). The magnitude of a PCS in an amide resonance is both distance- and orientation-dependent in relation to the bound metal. Both lanthanides are expected to cause PCS on the same line but in different directions based on their respective magnetic susceptibility anisotropy tensors $\Delta\chi$. The magnitude of the PCS also depends on the degree of saturation of the binding pocket as binding occurs on the fast exchange time scale. Almost all PCS effects occur along parallel lines as the orientations and distances relative to the bound lanthanide ion of attached ^{1}H and ^{15}N nuclei, separated by only one bond, are highly similar. As the dimeric P-domain contains two metal binding pockets, PCS may also occur between the two monomeric units.

PCS effects can also be identified in 1D proton NMR spectra of the ligand molecule GCDCA (Fig. 4.30). In presence of the murine NoV P-domain, titration of paramagnetic lanthanide ions leads to shifts of up to 5 Hz in ^{1}H signals of GCDCA. As the ligand GCDCA is in fast exchange, the position of its ^{1}H signals corresponds to the position of the population-weighted average of bound and unbound ligand molecules. Apparently, the bound GCDCA comes close to the coordinated lanthanide ions and, thus, experiences PCS effects from both paramagnetic Ce^{3+} and Eu^{3+} ions. Despite the 6-fold excess of the ligand, the magnitude of the PCS effect on the bound species is large enough to be still visible in the averaged chemical shift. Again, due to the different signs of the respective $\Delta\chi$-tensors, signals shift in opposite directions with both ions. Not all GCDCA ^{1}H signals are subject to PCS effects: whereas the signal of H25 displays clear PCS, the neighboring signal of H7 is not affected (Fig. 4.30a). This can be explained qualitatively by a crystal structure model of the murine NoV P-domain and GCDCA obtained in the presence of Mg^{2+} (Fig. 4.30b, pdb: 6E47). As lanthanide ions and Mg^{2+} share the same binding pocket, atomic distances between the lanthanide ions and GCDCA can be estimated. H25 has a distance of 13 Å to the metal binding site, while H7 is further apart (16 Å). These distances are in very good agreement with the NMR data, as the maximum distance at which PCS effects with Ce^{3+} and Eu^{3+} are observable is thought to be 15 Å (cf. chapter 3.3.6). This experiment also corroborates that the neighboring binding pockets for metals and GCDCA identified by crystallography are also those that are being occupied in solution and have been characterized here.

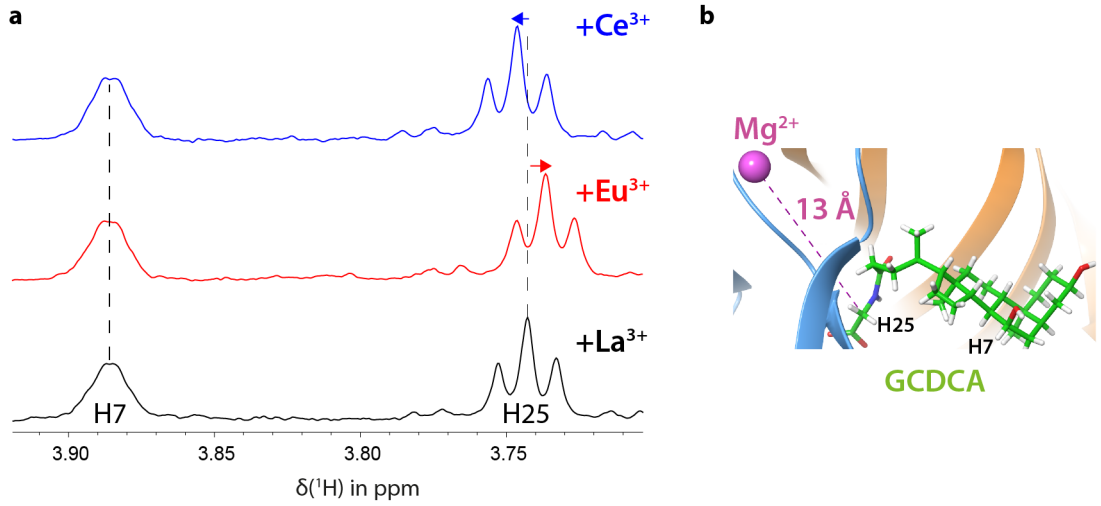


Figure 4.30: Paramagnetic effects reveal the spatial proximity between metal- and bile acid-binding sites in the murine NoV P-domain in solution. Inspection of 1D proton NMR spectra of GCDCA in the presence of the murine NoV P-domain and paramagnetic ions shows pseudo-contact shifts (PCS) in some GCDCA ^{1}H signals (**a**). The chemical shift of GCDCA signals reflects the average chemical shifts of bound and unbound species. Apparently, PCS effects in the bound state are still visible in the population-weighted average. Thus, GCDCA and the metal ions must come into close proximity (< 15 Å) when both molecules are bound by the P-domain. Comparison with an existing crystal structure model of the murine P-domain with GCDCA and $MgCl_2$ shows a distance of 13 Å between H25 and Mg^{2+} (**b**, pdb: 6E47, Nelson et al. (2018)). Lanthanide ions have been shown to occupy the same binding pocket as Mg^{2+} in Fig. 4.29. NMR samples were identical to those described in Fig. 4.29 and were measured at 600 MHz.

4.6 Glycan Recognition of Human and Murine NoV Capsid Proteins

The glycan binding specificity of human and murine NoV capsid proteins was investigated with protein-based CSP as well as ligand-based STD experiments. A detailed list of carbohydrates investigated here is given in Tab. 6.6 and Fig. 6.1 including their structural formulas and a description of the abbreviated glycan notation used throughout this chapter.

Human GII.4 NoV P-domains recognize fucose-containing HBGAs as already shown in Fig. 4.9 for the blood group B-trisaccharide. The affinity is significantly attenuated by spontaneous deamidation of Asn373 (cf. chapter 4.2). Binding of the blood group A-trisaccharide to [*U*-^{2}H,^{15}N]-labeled, N/N P-domains causes CSPs similar to the ones observed upon addition of B-trisaccharide (Fig. 4.31a). The dissociation constant K_D for the A-trisaccharide has been obtained from global fitting of CSP data and was found to be 4.3 mM (Fig. 4.31d), almost identical to that of the B-trisaccharide (cf. Fig. 4.9).

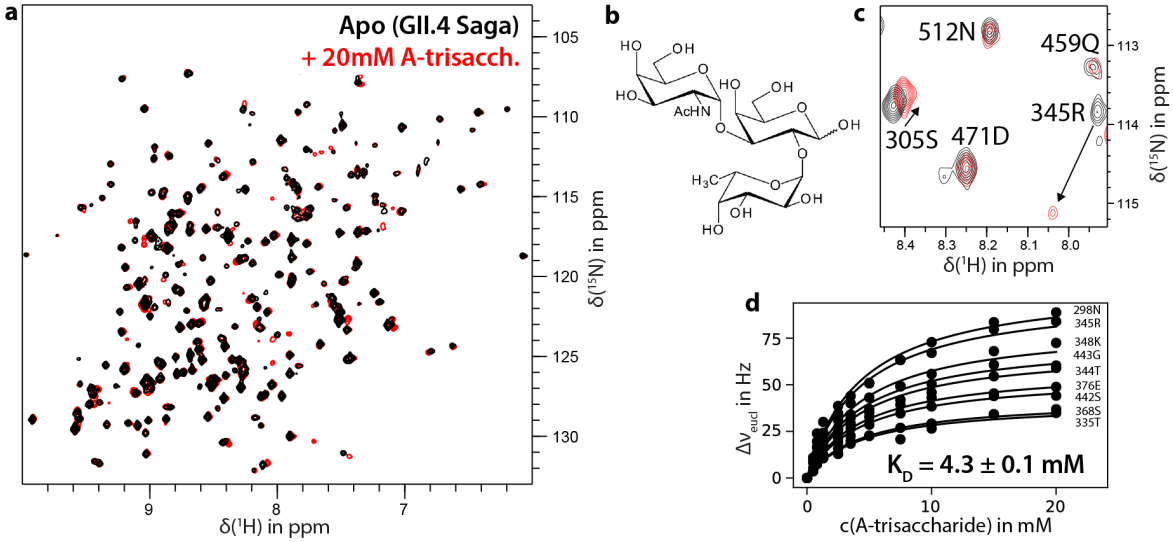


Figure 4.31: HBGA binding of human GII.4 P-domains. Titration of the blood group A-trisaccharide (**b**) to [*U*-^{2}H,^{15}N]-labeled P-domains from GII.4 Saga results in significant CSPs of signals of amino acids located in the the HBGA binding pocket (**a**). A representative section of the spectra is shown in **c**, highlighting significant CSPs of some signals. Global fitting of significant CSPs to the law of mass action yields a K_D value of 4.3 mM (**d**). The sample has been prepared in 75 mM sodium phosphate buffer, 100 mM NaCl (pH* 7.3) with a protein concentration of 230 µM.

In contrast, all GII.4 NoV P-domains fail to recognize any sialic acid-containing glycan. Titration of 3'-sialyllactose (hereafter abbreviated as 3'SL) to GII.4 VA387 does not cause any chemical shift perturbations, unambiguously demonstrating that there is no protein-ligand interaction (Fig. 4.32a). CSP experiments have been performed under a large number of experimental conditions, to rule out that binding might be dependent on the GII.4 NoV strain investigated, or the sialylated glycan used for titration, or the sample buffer (Tab. 4.2). None of the P-domains from the three GII.4 strains Saga, MI001 and VA387 binds 3'SL. Also, spectra from the fully deamidated P-domain of GII.4 Saga do not show any CSPs. Sialic acid (α-D-Neu5Ac-OMe) is not recognized either. Finally, it was shown that there is also no binding under acidic conditions using sodium acetate buffer, or under neutral conditions using ammonium acetate buffer as used for native mass spectrometry.

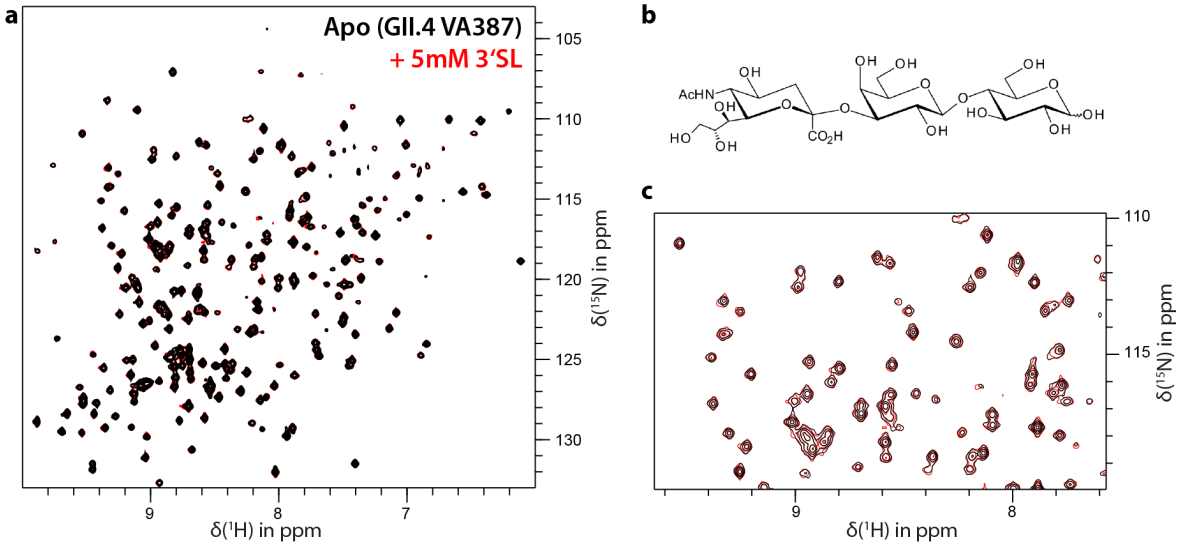


Figure 4.32: There is no interaction of human GII.4 P-domains with sialylated glycans. 3'-sialyllactose (the glycan moiety of the GM3 ganglioside, **b**) does not cause any CSPs of signals of [*U*-^{2}H,^{15}N]-labeled GII.4 VA387 P-domains (**a**). An enlarged section of the spectra is shown in **c**. Samples conditions are described in Tab. 4.2.

Table 4.2: GII.4 NoV P-domains do not bind sialic acid-containing glycans. Titrations of sialic acid (α-D-Neu5Ac-OMe) or 3'-sialyllactose (3'SL) to [*U*-^{2}H,^{15}N]-labeled GII.4 NoV P-domains do not lead to any changes in respective TROSY HSQC spectra regardless of buffer conditions, deamidation status (estimated on basis of the spectra), or strain investigated. Data described previously in the author's Master's thesis (Creutznacher, 2016) or in the supervised Bachelor's thesis of G. Wallmann (Wallmann, 2019) are indicated by [a] and [b], respectively. The VA387 P-domain ([c]) was denatured and refolded by P. Ogrissek during supervised work as a research assistant.

NoV strain (deamidation status)	carbohydrate (concentration)	buffer	carbohydrate binding
GII.4 Saga (>90 % 373N, 600 µM)[a]	α-D-Neu5Ac-OMe (5 mM)	20 mM sodium phosphate, pH 7.4	no binding
GII.4 Saga (>90 % 373N, 860 µM)	3'SL (10 mM)	75 mM sodium phosphate, 100 mM NaCl, pH 7.3	no binding
GII.4 Saga (>80 % 373N, 250 µM)	3'SL (16 mM)	100 mM sodium acetate, pH 4.6	no binding
GII.4 Saga (>90 % 373N, 110 µM)	3'SL (10 mM)	200 mM ammonium acetate , pH 7.3	no binding
GII.4 Saga (>90 % 373iD, 100 µM)	3'SL (10 mM)	75 mM sodium phosphate, 100 mM NaCl, pH 7.3	no binding
GII.4 MI001 (>90 % 373N, 80 µM)[b]	3'SL (5 mM)	50 mM sodium phosphate, 20 mM NaCl, pH 7.4	no binding
GII.4 MI001 (>90 % 373N, 180 µM)[b]	α-D-Neu5Ac-OMe (8 mM)	75 mM sodium phosphate, 100 mM NaCl, pH 7.3	no binding
GII.4 VA387 (>90 % 373N, 170 µM)[c]	3'SL (5 mM)	75 mM sodium phosphate, 100 mM NaCl, pH 7.3	no binding
GII.4 VA387 (>90 % 373N, 170 µM)[c]	3'SL (5 mM)	5 mM GCDCA, 75 mM sodium phosphate, 100 mM NaCl, pH 7.3	no binding
GII.4 VA387 (>90 % 373N, 170 µM)[c]	3'SL (5 mM)	20 mM sodium acetate, 100 mM NaCl, pH 5.3	no binding

Murine NoV P-domains do not form stable dimers in the absence of GCDCA (cf. chapter 4.4). Therefore, binding of fucosylated or sialylated glycans to mixtures of monomers and homodimers was first tested with no GCDCA being present (Tab. 4.3). Ligand-based STD experiments did not show binding of the respective fucose or sialic acid monosaccharides (α-L-Fuc-OMe, α-D-Neu5Ac-OMe) to P-domains from the MNV07 strain. Protein-based CSP experiments were then used to test for interactions with more complex carbohydrates. No ligand-specific CSPs could be detected in experiments with the GM3 trisaccharide, the GD1a hexasaccharide, or the blood group B-trisaccharide. As the stability of the murine NoV P-domain is strongly dependent on the buffer pH (cf. Fig. 4.20), binding was only investigated under acidic pH values in sodium acetate or sodium phosphate buffers. The B-trisaccharide solution used for titration was found to contain Mn^{2+} impurities arising from the chemo-enzymatic synthesis. This results in changes of TROSY HSQC spectra as bivalent metal ions have been shown to bind to the murine NoV P-domain (cf. Fig. 4.28). Specifically, Mn^{2+} causes paramagnetic relaxation enhancements in signals belonging to the metal ion binding pocket (cf. Fig. 4.29). No additional effects indicative of actual glycan binding were detected.

Table 4.3: The murine NoV P-domain does not recognize fucosylated or sialylated glycans in the absence of GCDCA. Murine NoV P-domains form mixtures of monomers and homodimers in the absence of GCDCA depending on the protein concentration (cf. chapter 4.4). Neither monomers nor homodimers bind sialic acid- or fucose-containing glycans. Carbohydrate binding has been investigated using ligand-based STD and protein-based CSP experiments for different murine NoV strains under different buffer conditions. Data obtained as part of the supervised Bachelor's thesis of C. Feldmann are indicated by [a] (Feldmann, 2019). CSP data showing binding of Mn^{2+}-impurities, but not glycans, are labeled with [b].

NoV strain (concentration)	carbohydrate (concentration)	buffer	experiment (temperature)	carbohydrate binding
MNV07 [a] (24 µM)	α-D-Neu5Ac-OMe (2 mM)	20 mM sodium phosphate, 100 mM NaCl, pH 6	STD (277 K)	no binding
CW1 (200 µM)	3'SL (10 mM)	20 mM sodium acetate, 100 mM NaCl, pH 5.3	CSP (298 K)	no binding
MNV07 [a] (28 µM)	GD1a (glycan) (2.5 mM)	20 mM sodium phosphate, 100 mM NaCl, pH 6	CSP (293 K)	no binding
CW1 (40 µM)	α-L-Fuc-OMe (4 mM)	20 mM sodium acetate, 100 mM NaCl, pH 5.3	STD (298 K)	no binding
MNV07 (41 µM)	B-trisaccharide (5 mM)	20 mM sodium acetate, 100 mM NaCl, pH 5.3	CSP (298 K)	no binding [b]
CW1 (200 µM)	B-trisaccharide (10 mM)	20 mM sodium acetate, 100 mM NaCl, pH 5.3	CSP (298 K)	no binding [b]

Bile acid-binding shifts the dimerization equilibrium of the murine NoV P-domain almost quantitatively towards the dimeric state. Additionally, the binding of bile acids is accompanied by changes in structural equilibria within the P-domain itself (cf. chapter 4.4). To test whether the murine NoV P-domain can interact with glycans in its bile acid-bound state, CSP and STD experiments have been performed in the presence of saturating amounts of GCDCA (Tab. 4.4). No binding could be detected for any sialylated glycan (Neu5Ac, 3'SL, the GM1a glycan) or HBGA (fucose, H-disaccharide, A-trisaccharide, B-trisaccharide). In addition, there is no interaction with galactose-based carbohydrates from other glycan epitopes (galactose, GalNAc, Forssman-trisaccharide, Galili-disaccharide). Again, contaminations of glycans with bivalent metal ions causes CSPs not due to glycan binding. Particularly, 3'SL was found to contain appreciable amounts of Ca^{2+} ions. All significant CSPs can be explained by metal-binding as evident from dedicated metal ion-titration experiments (cf. Fig 4.28). STD experiments do not suffer from limitations by inadvertent addition of metal ions. None of the STD experiments show any saturation transfer from the P-domain to the tested glycans (Fig. 4.33). However, signals from the simultaneously present GCDCA do undergo saturation transfer from the protein, corroborating the existence of a GCDCA-P-domain interaction.

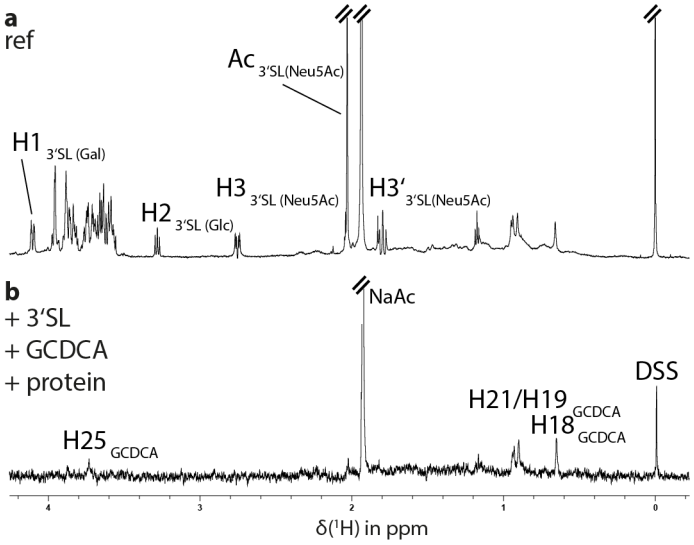


Figure 4.33: Murine NoV P-domains do not interact with the carbohydrate moiety of GM3 (3'SL) in the presence of GCDCA. The 3'SL carbohydrate does not receive saturation from the murine MNV07 P-domain. Comparison of a reference proton spectrum (**a**) with a STD spectrum (**b**) shows only saturation transfer to protons of GCDCA. None of the 3'SL signals show any response (prominent signals from all three monosaccharide units are indicated in **a**). The sample was prepared with a P-domain concentration of 80 µM in 20 mM sodium phosphate buffer, 100 mM NaCl (pH 5.3) with 2 mM 3'-sialyllactose (3'SL) and 250 µM GCDCA. STD spectra were acquired with an on-resonance frequency of -3 ppm and a measurement time of 14 h on a 600 MHz NMR spectrometer at 298 K. Bile acid and 3'SL signal assignments have been obtained from Ijare et al. (2005) and Wegener et al. (2017), respectively. High intensity signals from acetate, the 3'SL N-acetyl group and DSS-d_6 have been cropped.

Table 4.4: Murine NoV P-domains do not bind carbohydrates in the presence of GCDCA. GCDCA binding shifts the homo-dimerization equilibrium of the murine NoV P-domain towards the dimeric state (cf. chapter 4.4). Using STD and CSP experiments, a variety of potential glycan ligands has been tested: fucosylated HBGAs, sialylated gangliosides, the Galili- and Forssman-epitopes. No binding could be detected for any of these carbohydrates for different murine NoV strains. CSP data interpretation requires caution as various glycans were found to contain appreciable amounts of bivalent metals that are bound by the murine NoV P-domain (cf. chapter 4.5). Respective metal contaminants are indicated as follows: [a]: Mg^{2+}, [b]: Ca^{2+}, [c]: Mn^{2+}. CSP data was compared with dedicated metal ion titrations (cf. Fig. 4.28) to rule out simultaneous binding of both contaminants and carbohydrates. Data indicated by [d] were obtained as part of the supervised Master's thesis of T. Maaß (Maaß, 2019).

NoV strain	carbohydrate (concentration)	buffer	experiment (temperature)	carbohydrate binding
CW1 (72 µM)	α-D-Neu5Ac-OMe (40 mM)	300 µM GCDCA, 20 mM sodium acetate, 100 mM NaCl, pH 5.3	CSP (298 K)	no binding [a]
MNV07 (80 µM)	3'SL (2 mM)	250 µM GCDCA, 20 mM sodium acetate, 100 mM NaCl, pH 5.3	STD (277 K & 298 K)	no binding
CW1 (72 µM)	3'SL (50 mM)	300 µM GCDCA, 20 mM sodium acetate, 100 mM NaCl, pH 5.3	CSP (298 K)	no binding [b]
CW1 (100 µM)	GM1a (glycan) (4 mM)	300 µM GCDCA, 20 mM sodium acetate, 100 mM NaCl, pH 5.3	CSP (298 K)	no binding [b]
CW1 [d] (50 µM)	GalNAc (4 mM)	300 µM GCDCA, 20 mM sodium acetate, 100 mM NaCl, pH 5.3 (D_2O)	STD (298 K)	no binding
CW1 (100 µM)	α-L-Fuc-OMe (12 mM) Galactose (20 mM)	300 µM GCDCA, 20 mM sodium acetate, 100 mM NaCl, pH 5.3	CSP (298 K)	no binding
CW1 (100 µM)	Galili-Disaccharide (50 mM)	300 µM GCDCA, 20 mM sodium acetate, 100 mM NaCl, pH 5.3	CSP (298 K)	no binding [a]
CW1 (100 µM)	H-Disaccharide (50 mM)	300 µM GCDCA, 20 mM sodium acetate, 100 mM NaCl, pH 5.3	CSP (298 K)	no binding [a]
MNV07 (73 µM)	B-trisaccharide (1.8 mM)	140 µM GCDCA, 20 mM sodium acetate, 100 mM NaCl, pH 5.3	STD (277 K)	no binding
CW1 (200 µM)	B-trisaccharide (0.5 mM)	300 µM GCDCA, 20 mM sodium acetate,	CSP (298 K)	no binding [c]

Table 4.4: (continued)

NoV strain	carbohydrate (concentration)	buffer	experiment (temperature)	carbohydrate binding
		100 mM NaCl, pH 5.3		
CW1 [d] (50 µM)	A-trisaccharide (2 mM)	300 µM GCDCA, 20 mM sodium acetate, 100 mM NaCl, pH 5.3 (D_2O)	STD (298 K)	no binding
CW1 [d] (50 µM)	Forssman-trisaccharide (2 mM)	300 µM GCDCA, 20 mM sodium acetate, 100 mM NaCl, pH 5.3 (D_2O)	STD (298 K)	no binding

To rule out any interdependence between metal and glycan binding, CSP experiments with HBGAs and gangliosides where performed with deliberately saturated or depleted metal ion-binding pockets (Tab. 4.5). Titration of sialic acid or 3'SL does not result in any spectral changes in the presence of saturating amounts of $CaCl_2$. Chelation of bivalent metals with EDTA does not impact glycan recognition either. Specifically, EDTA suppresses the disappearance of signals upon B-trisaccharide titration linked to paramagnetic relaxation enhancement due to a contamination with Mn^{2+} ions (cf. Tab. 4.3).

Table 4.5: Saturation with or depletion of metal ions, or buffer pH cannot induce glycan recognition of murine NoV P-domains. Neither saturating amounts of Ca^{2+} nor metal chelation using EDTA do not enable HBGA or sialoside recognition. No binding could be detected using protein-based CSP experiments.

NoV strain (concentration)	carbohydrate (concentration)	buffer	experiment (temperature)	carbohydrate binding
CW1 (70 µM)	α-D-Neu5Ac-OMe (10 mM)	300 µM GCDCA, 75 mM sodium phosphate, 100 mM NaCl, 1 mM $CaCl_2$, pH 6	CSP (298 K)	no binding
CW1 (70 µM)	3'SL (10 mM)	300 µM GCDCA, 20 mM sodium acetate, 100 mM NaCl, 1 mM $CaCl_2$, pH 5.3	CSP (298 K)	no binding
CW1 (190 µM)	α-D-Neu5Ac-OMe (19 mM)	600 µM GCDCA, 20 mM sodium acetate, 100 mM NaCl, 3 mM EDTA, pH 5.3	CSP (298 K)	no binding
CW1 (70 µM)	3'SL (10 mM)	300 µM GCDCA, 20 mM sodium acetate, 100 mM NaCl, 1 mM $CaCl_2$, 2 mM EDTA, pH 5.3	CSP (298 K)	no binding
CW1 (190 µM)	B-trisaccharide (10 mM)	600 µM GCDCA, 20 mM sodium acetate, 100 mM NaCl, 3.4 mM EDTA, pH 5.3	CSP (298 K)	no binding

5. Discussion

5.1 Backbone Assignment and Deamidation of Asn373 in Human GII.4 Noroviruses

The 73 kDa dimer of the human GII.4 NoV P-domain is among the largest protein complexes for which an NMR backbone signal assignment has been obtained. The current record in this regard is set by the 81 kDa malate synthase G (Tugarinov et al., 2002). Assignment of the GII.4 P-domain dimer clearly benefited from the homodimeric nature of the complex, reducing the number of expected signals and, thus, potential signal overlap compared to a single-chain protein of this size. For very large molecules, selective isotope labeling of methyl groups is a powerful alternative to backbone labeling (Schütz and Sprangers, 2020). This extends the molecular weight limit of solution NMR spectroscopy well beyond the 100 kDa mark albeit at the cost of a reduced number of atomic probes .

Interestingly, the above-mentioned, backbone assigned malate synthase G was also found to contain an Asn deamidation site. This raises the intricate question, how often Asn deamidation occurs in proteins and has gone unnoticed due to the limited number of analytical techniques capable of detecting this reaction. Asn deamidation in malate synthase G had only very subtle effects on ^{1}H,^{15}N TROSY HSQC spectra. Only four signals were found to be influenced by this reaction. In contrast, a much larger number of signals was affected in the GII.4 P-domain (cf. Fig 4.6a) indicating that deamidation causes long-range, indirect structural changes in this case. In the wild-type protein, no ^{1}H,^{15}N signals could be identified that belong to the loop carrying the deamidation site. In contrast, all of these amino acids gave high intensity signals *after* deamidation hinting at a major change in loop dynamics. This is in good agreement with findings from hydrogen-deuterium exchange mass spectrometry that revealed an increase in protein flexibility in the entire upper P-domain after deamidation (Mallagaray et al., 2019).

How can the accompanying decrease in HBGA binding affinity by one order of magnitude be explained? A recently obtained crystal structure model of the deamidated P-domain in complex with fucose shows that the immediate neighbor of Asn373 changes its orientation towards the bound carbohydrate (Fig. 5.1). Asp374 is critically involved in HBGA recognition of the wild-type protein (cf. Fig. 3.2) as it forms a bidentate hydrogen bond with the fucose OH-2 and OH-3 hydroxyl groups. Formation of an isopeptide bond leads to rotation of the Asp sidechain and the subsequent loss of one hydrogen bond. Taken together with the observed change in protein dynamics, this might – at least partially – explain, why Asn deamidation impairs glycan recognition. Consequently, previously re-

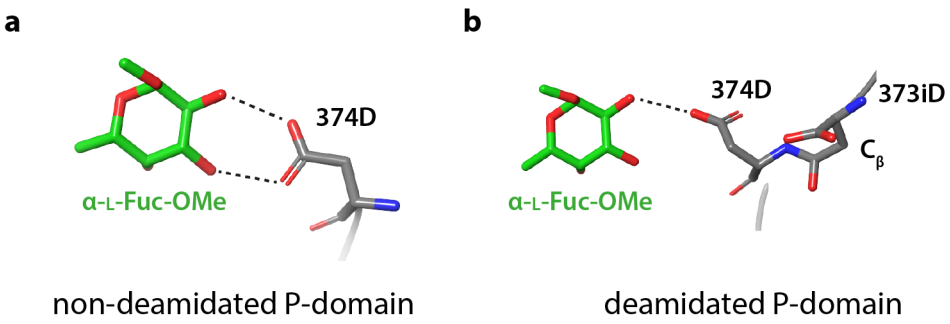


Figure 5.1: Deamidation of Asn373 causes structural changes in the HBGA binding pocket. A crystal structure model of the deamidated P-domain of GII.4 Saga (pdb 6H9V) reveals the reorientation of Asp374, the amino acid directly neighboring the deamidation site. Comparing the orientation of the Asp sidechain in the wild-type protein (**a**) with the deamidated protein (**b**) shows that deamidation results in the loss of a hydrogen bond with the fucose moiety (dashed lines). To some extent, this might explain the observed loss in binding affinity for HBGAs upon deamidation. The C_β atom of the Asn373 sidechain has become a part of the protein backbone, corroborating that iso-aspartate is the major, likely exclusive, deamidation reaction product.

ported affinity values are potentially incorrect as deamidated proteins could accumulate over time and experiments were then performed on undefined mixtures of different protein species with varying affinities. In my thesis, I have shown that the N373D point mutant of the P-domain is also capable of binding HBGAs. This mutant is stable for prolonged periods of time, making it the preferred protein for future affinity studies.

Although deamidation occurs in a surface-exposed loop, the observed reaction rate is 20-times faster in the context of an entire protein than for a comparable model peptide under similar buffer conditions (Robinson et al., 2004). At 37 °C, the Asn in the pentapeptide `GENDG` has a half-life time of 32 d, whereas 1.6 d have been observed for the respective Asn in the GII.4 NoV P-domain. As a protein's tertiary structure has been considered to slow down, not enhance, the deamidation reaction rate (Robinson and Robinson, 2008), this finding strengthens the emerging view that peptide-derived empirical rules cannot predict actual deamidation rates in proteins (Plotnikov et al., 2017). In addition, in peptides both Asp and isoAsp are expected to form in a ratio of 1:3 and subsequently interconvert in an equilibrium reaction (Robinson and Robinson, 2004). In the P-domain, however, Asp is no detectable reaction product at pH 7.3 and an Asp point mutant at position 373 does not show signs of interconversion into isoAsp.

The biological importance of the capsid's deamidation for the viral life cycle remains unknown. Approximately two thirds of all GII.4 NoV strains contain an Asn in position 373 of their capsid and are, therefore, likely to deamidate (Mallagaray et al., 2019). It was shown here that at least two of these GII.4 strains, Saga and MI001, are undergoing this reaction. Considering the fast deamidation reaction rate and long environmental stability of NoV particles (Robilotti et al., 2015), it appears plausible that deamidated capsids can occur during the infection process. As GII.4 NoV undergo fast epochal evolution, it is worthwhile speculating if the continued presence of an Asn residue in position 373 might be evidence of some beneficial function. Asn373 has been described as part of the immunodominant antibody epitope A in GII.4 NoV (Mallory et al., 2019). Antibody recognition is known to be sensitive towards minor changes in this epitope and position 373 in particular has been identified as critical

for binding of some antibodies (Lindesmith et al., 2017). A major structural change by means of a newly formed isopeptide bond at this position might, therefore, lead to evasion from the host immune system. Asn deamidation might have further consequences for vaccine development. As most vaccine candidates are based on recombinant, non-infectious GII.4 virus-like particles (Cortes-Penfield et al., 2017), it appears critical to identify whether non-deamidated or deamidated particles are actually required to elicit an effective immune response.

5.2 Interactions of Norovirus Capsids with Bile Acids

Using chemical shift perturbation (CSP) experiments, a previously unknown, low-affinity binding pocket for bile acids has been identified in GII.4 NoV capsids. The existence of this binding site was validated by both protein-based and ligand-based NMR experiments and it has been located near the C-terminus of the protein. This binding pocket was shown to be permissive for primary and secondary bile acids and exists in at least three different GII.4 NoV strains (Saga, MI001 and VA387). Binding to a previously reported, high-affinity binding site for bile acids in rare human NoV strains (cf. Fig. 3.4) can be excluded as no CSPs could be observed at the outward-facing part of the P-domain.

The interaction was found to be of low affinity in the mM-range, similar to that of HBGAs. This raises the question of the relevance of such a binding pocket in a biological context. Whereas low-affinity glycan binding can be compensated for by multivalent interactions leading to firm surface attachment (Parveen et al., 2019), such a mechanism appears unlikely for soluble bile acids in binding pockets that are much less exposed to the exterior. However, bile acid concentrations in the gastrointestinal tract are very high and certainly in the same order of magnitude as the K_D of the interaction (cf. chapter 3.1.5). This makes – at least partial – occupation of this binding pocket upon viral passage of the small intestine likely. How and whether this process is linked to the observed increase in GII.4 NoV replication efficiency upon bile acid treatment in cell culture experiments (Ettayebi et al., 2016) remains unclear.

The biophysical properties of bile acids complicate the determination of accurate K_D values in solution. Being prone to aggregation in the relevant concentration range, binding of both, monomeric and aggregated, bile acids to the identified binding pocket is possible, potentially obscuring binding isotherms. Although critical *micelle* concentrations (CMC) are commonly reported and have been estimated here by simple proton NMR spectroscopy, bile acids do not adopt classical, spherical micelles. Instead, bile acid 'micelles' are highly polydisperse aggregates and, at concentrations much higher than the CMC, bile acid monomers and dimers are still the dominant species (Verde and Frenkel, 2010).

Lastly, it should be noted that binding was observed to the soluble P-domain and not to a complete virus particle. The binding pocket is located near the C-terminus of the protein and, thereby, in close proximity to where the shell domain would be in the intact capsid. In context of the emerging evidence that the P-domains might adopt different states with regard to the viral shell (termed 'floating' P-domains, cf. chapter 3.1.2), the observed bile acid binding mode might not reflect 'natural' bile acid binding. Similarly, soluble, recombinant P-domains are truncated and are missing the last C-terminal amino acids of the natural capsid to prevent aggregation (Tan et al., 2006). These C-terminal 'tails'

could well influence bile acid recognition. However, bile acid binding to GII.4 virus-like particles has been confirmed and similar affinities were reported (Creutznacher et al., 2019). These particles contain both the natural C-terminus as well as the shell domain, suggesting low-affinity bile acid binding is a common feature of intact GII.4 NoV particles.

Bile acid binding to another, structurally distinct binding pocket has already been described for murine NoV P-domains (Nelson et al., 2018). In this thesis, several consequences of this binding process have been identified. Most importantly, the murine NoV P-domain can only form stable dimers in the presence of its bile acid ligand. This reveals a striking difference to human NoV P-domains: whereas these P-domains form extremely stable dimers without the involvement of any ligands (k_{off}: $10^{-6}s^{-1}$), the dissociation rate of unbound murine NoV P-domains is approximately four orders of magnitude larger. The high-affinity binding of the bile acid GCDCA at the dimer interface shifts the dimerization equilibrium almost completely to the dimeric state by depletion of unbound dimers from the monomer-dimer equilibrium. As the monomeric P-domain was shown to be binding-incompetent, saturation with GCDCA consequently leads to complete dimerization. The back-reaction towards monomers is severely impaired: this would require both GCDCA molecules to dissociate consecutively from the complex which is unlikely considering the two K_D values in the low µM range.

Predictions derived from the GCDCA binding model of coupled dimerization and two consecutive, independent ligand binding events are in agreement with the experimental NMR data. However, given the complex nature of the involved spectral changes other models might also be applicable. Especially the assumption that both GCDCA binding events are independent and not cooperative needs to be addressed by future studies. In an assembled viral capsid, the P-domain dimers will, most likely, not fully dissociate into long-lived monomers in the absence of GCDCA. All P-domains are anchored to their respective shell domains that form additional interactions with each other, thereby keeping the P-domains in close spatial proximity. However, the P-domain is separated from the shell by a flexible linker peptide and it appears likely that ligand-induced modulation of the interactions between P-domains is another mode of murine NoV capsid dynamics.

Another aspect of the complex interaction with GCDCA is the apparent change in protein dynamics upon binding as evident from changes in TROSY HSQC spectra (cf. Fig. 4.23). Some, but not all, dimer signals become only visible after GCDCA titration, with no corresponding signal of the unbound dimer state being detectable. This behavior can have several explanations, all of which indicate major changes in protein dynamics (cf. chapter 3.3.2):

(1) Certain parts of the protein could undergo conformational exchange on the same time scale as the chemical shift difference between bound and unbound states. Accordingly, the corresponding signals are broadened beyond detection. These intermediate exchange kinetics on the µs to ms time scale are altered by ligand binding and the system moves into slow exchange with the minor, unbound dimer species being undetectable.

(2) Parts of the unbound dimer might have unfavorably short relaxation times and become more dynamic on the ps to ns time scale as the ligand is binding. Higher frequency motions on this time scale can increase a signal's intensity considerably.

(3) As the protein is synthesized in deuterated media, there might be regions in the protein where amide groups are protected from back-exchange with protons from the aqueous solvent. This phenomenon would be analogous to the human NoV P-domain in which a high number of signals is invisible without a controlled denaturation-refolding reaction (cf. Fig. 4.3). Changes in solvent accessibility through conformational changes upon GCDCA binding could expose certain amide resonance to the solvent and, thereby, trigger proton-deuterium exchange.

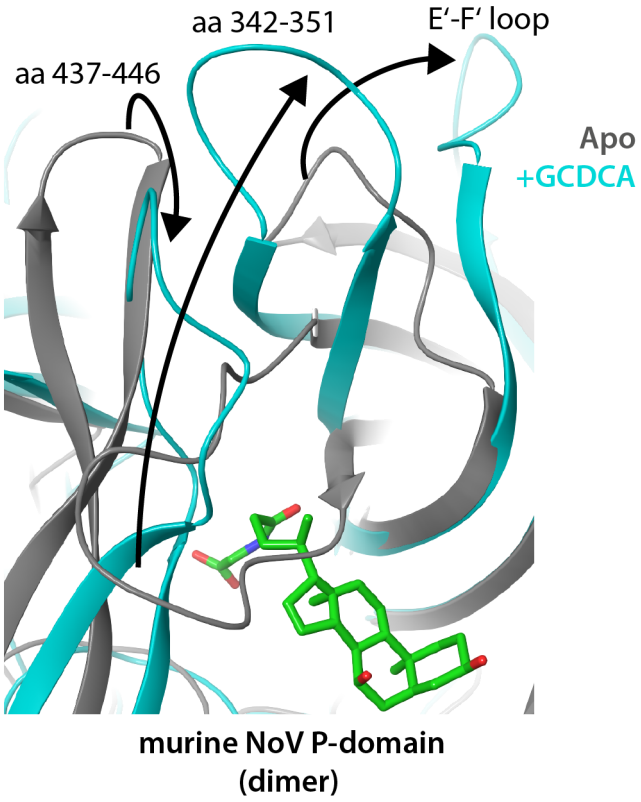


Figure 5.2: Structural changes in crystal structure models of the murine NoV P-domain upon GCDCA binding. The aligned crystal structure models of the murine NoV P-domain with (blue, pdb 6E47, Nelson et al. (2018)) and without (black, pdb 3LQ6, Taube et al. (2010)) a bound GCDCA molecule are shown. Several major structural rearrangements around the binding pocket are highlighted with arrows. In the apo state, a loop blocks access to the GCDCA binding site and is found pointing upwards after binding (middle, amino acids 342-351). The loop comprising the bivalent metal ion binding site in the GCDCA-bound state (left, aa 437-446) can undergo reorientation as well. In the back, the E'-F' loop is seen adapting the 'open' conformation in absence of GCDCA and the 'closed' state after ligand binding. As recognition of the murine NoV P-domain by two different antibodies occurs preferentially in the open state (Kolawole et al., 2017), GCDCA might shift the structural equilibrium towards a state capable of immune evasion.

A backbone signal assignment and dedicated relaxation time measurements on different time scales are required to narrow down these possible explanations. An overlay of two crystal structure models of the murine NoV P-domain with and without bound GCDCA, however, supports the notion that protein dynamics are impacted by GCDCA-binding (Fig. 5.2). Several surface-exposed loops on the apex of the P-domain change their orientation upon ligand recognition. This includes the loop with the metal ion binding site and another loop blocking the GCDCA binding pocket in the apo structure that is moved upwards. Most importantly, these movements have an indirect effect on the distal E'-F' loop, that is critical for antibody recognition. It has been described in previous studies that the dy-

namic E'-F' loop can adopt two different conformations in the P-domain, termed 'open' and 'closed', in the absence of any ligands (Taube et al., 2010). Two antibodies directed against the murine NoV P-domain were shown to bind preferentially to the 'open' conformation using a combination of cryo-electron microscopy and a molecular dynamics simulation (Kolawole et al., 2017). Point mutations, impacting the population of these states, have been suggested to allosterically control antibody recognition and, thereby, viral immune escape. In the crystal structure model with the bound GCDCA, the E'-F' loop is shifted into the 'closed' conformation. This difference in 'open'-'closed' exchange kinetics could explain the observed appearance of NMR signals upon GCDCA binding as outlined above as hypothesis (1). Taken together, it appears plausible that GCDCA-binding triggers an allosteric mechanism making the viral capsid unrecognizable by the host immune system.

Many NoV capsids have pH-dependent stability profiles (Pogan et al., 2018). Here, it was shown that the murine NoV P-domain is more resistant against thermal denaturation under acidic conditions and becomes unstable at neutral pH. For enteric viruses, acid resistance is a necessary feature and murine NoV were, indeed, shown to be infectious over a wide range of pH values (Cannon et al., 2006). However, no decrease in infectivity at neutral pH values has been found, highlighting a potential stabilizing influence of the viral shell domain that is absent in the recombinant P-domain. Similarly, the presence of GCDCA has a strong protective effect on capsid stability (cf. Fig. 4.22). It is a known property of many proteins that ligand binding increases the thermal stability of the native protein when it is selective for the folded protein only (Bai et al., 2019). It remains an open question whether pH- and ligand-dependent protein stability is a property of the dimeric P-domain or whether both effects are due to depletion of monomeric P-domains that might expose a hydrophobic, aggregation-prone dimer interface.

5.3 Norovirus Protruding Domains are Metalloproteins

In a rare GII.2 human NoV strain, occupation of a Zn^{2+} binding pocket on the apex of the P-domain has been linked to an increase in capsid stability (Jung et al., 2019). The investigated strain is believed to be binding-incompetent for HBGAs in its apo state and it was postulated that Zn^{2+}-binding might act as an allosteric cofactor that enables HBGA recognition. NoV of the genotype GII.4 can bind HBGA without cofactors, but partial conservation of the proposed Zn^{2+} binding site between both genotypes lead to the question whether Zn^{2+}-binding is relevant for HBGA recognition in the predominant GII.4 NoV strains.

In fact, the GII.4 NoV P-domain was found to bind Zn^{2+} ions but not in the binding pocket described above. Instead, spectral changes upon Zn^{2+}-titration revealed a binding pocket near the C-terminus of the protein. The nature of the spectral changes is somewhat unusual. Unexpectedly, affected backbone amide signals did not show CSPs but were broadened beyond detection. No new signals, indicative of slow exchange between apo and bound state (cf. chapter 3.3.1), appeared in the spectrum. In analogy to the discussion in chapter 5.2, several possible reasons for this behavior can be identified: intermediate exchange kinetics between free and metal-bound species (1), unfavorable relaxation times due to a decreased flexibility of the binding pocket (2), and changes of the amide pro-

ton exchange rate with the solvent (3) (Rezaei-Ghaleh et al., 2011). In addition, a contamination of the $ZnCl_2$ titration stock with paragmagnetic, high-affinity metals could be possible as discussed below for the murine NoV P-domain. A Zn^{2+} binding pocket in a similar region of the P-domain has been identified in bovine NoV (Singh et al., 2016). Increasing the amount of Zn^{2+} leads to the disappearance of other signals located in various regions of the P-domain, including one centered around the HBGA binding pocket. This points to the existence of further, lower affinity Zn^{2+} binding sites. Complexation of bivalent metal ions with EDTA lead to a reversal of all spectral changes, indicating that they were caused by specific binding of Zn^{2+}. Apparently, the GII.4 NoV P-domain binds zinc at two sites, one close to the known HBGA binding pocket and the other neighboring the low-affinity bile acid binding pocket near the protein's C-terminus. Due to the unfavorable broadening of signals in the HBGA and bile acid binding sites, the role of metal binding in potential allosteric regulation of ligand recognition could not be investigated using CSP experiments. Instead, ligand-based NMR techniques are preferable for future studies as they do not suffer from these limitations.

In murine NoV, the presence of Ca^{2+} or Mg^{2+} was linked to an increase in viral infectivity (Nelson et al., 2018). Two binding pockets have been described in the P-domain: one between the P-domain and the protein receptor CD300lf (hereafter: binding site **I**) and one close to the bile acid binding pocket at the dimer interface (**II**) (Fig. 5.3). As binding pocket **I** involves interactions with the receptor itself, its affinity will likely be reduced in the absence of the receptor. Although occupation of this pocket is most likely causing the metal-dependent increase in infectivity, this has not yet been demonstrated directly. It was shown here that only binding site **II** can be occupied in solution in absence of the receptor. All significant CSPs (cf. Fig. 4.28) follow the same binding isotherm, showing that at least Ca^{2+} binds to one distinct position in each P-domain with moderate affinity (K_D: 100 µM). In total, six different metal ions were able to bind (Mg^{2+}, Ca^{2+}, Mn^{2+}, La^{3+}, Ce^{3+}, and Eu^{3+}). Using the spectral effects of bound paramagnetic metal ions, it could be demonstrated that GCDCA and the bound metal ion are in close proximity (< 15 Å), ruling out binding to site **I** (Allegrozzi et al., 2000).

Chemical shift perturbation experiments with metalloproteins can easily lead to false-positive binding data: if the putative ligand contains non-declared metal ion impurities, the spectral effects of metal binding can be interpreted easily as binding of the investigated ligand molecule. This was found to be highly relevant for the murine NoV P-domain as a surprising number of tested glycans contained appreciable amounts of Mg^{2+}, Ca^{2+}, or Mn^{2+}. Only by comparison with dedicated metal ion titrations or deliberate saturation or depletion of the respective metal ion binding sites can reliable binding data be obtained. This kind of misinterpretation of binding data has been noted in other cases, e.g. in isothermal titration calorimetry (ITC) experiments (Morreale et al., 2017). As commercially available GCDCA was found to contain great amounts of Ca^{2+} (approximately one half of the total amount of substance) and its binding to the murine NoV P-domain was previously characterized with ITC (Nelson et al., 2018), existing binding data should be regarded with caution. In the NMR analysis of GCDCA binding presented here, binding of Ca^{2+} and GCDCA could be differentiated based on their different CSP profiles. Of note, the previous ITC analysis also neglected the accompanying process of P-domain dimerization that must have obstructed resulting thermograms. What is the origin

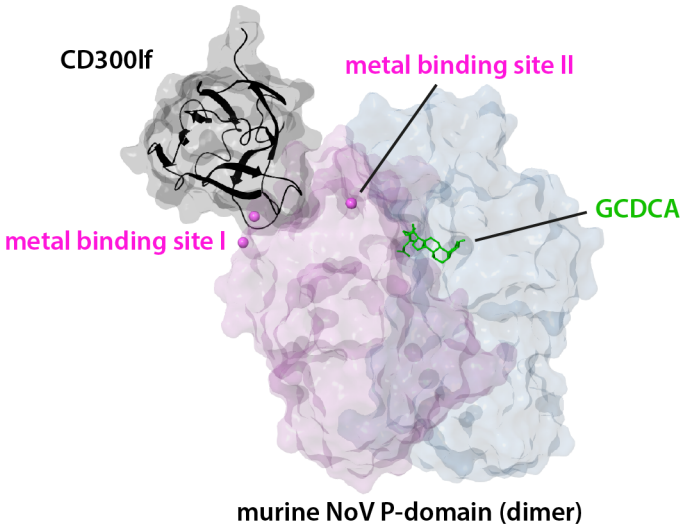


Figure 5.3: The metal binding pockets in murine NoV P-domains in the presence of the receptor CD300lf. The murine NoV P-domain dimer (purple and blue) harbors two binding sites for bivalent metals (purple spheres) as identified by crystallography (pdb 6E47). One binding site (**I**) is located at the interface to the protein receptor CD300lf (black) and accepts two metal ions in close proximity. The other binding site (**II**) is located near the GCDCA binding pocket (green). It has been shown in this thesis that metal binding occurs exclusively at site **II** in the absence of CD300lf. This binding site accepts a large number of bivalent and trivalent metal ions with moderate affinity. For clarity, only half of the binding sites for metals, GCDCA, and CD300lf in the symmetrical P-domain dimer have been displayed.

of these contaminating metal ions? At least for two tested molecules, an explanation can be given. Bile naturally contains a high amount of calcium ions (with insoluble calcium-bile salts being the reason for gallstones) (Hofmann and Mysels, 1992) and commercial bile acids, including GCDCA, are isolated natural products. The blood group B trisaccharide was synthesized chemo-enzymatically from an H-disaccharide precursor, using a recombinant blood group B glycosyltransferase as catalyst of the reaction. Blood group glycosyltransferases are well known to require Mn^{2+} ions as cofactors (Sindhuwinata et al., 2013). Both commercial and in-house synthesized chemicals have been purified by HPLC but, evidently, enough bivalent metal ions remained to significantly occupy the binding pocket on the murine NoV P-domains.

5.4 Redefining the Glycan Code of Murine and Human Noroviruses

Commonly, carbohydrates undergo only weak interactions with proteins, making the determination of accurate dissociation constants difficult (Bewley and Shahzad-ul Hussan, 2013). Isothermal titration calorimetry is the most widespread technique for analysis of high-affinity interactions in solution but it is of limited use for the characterization of low affinity interactions (Turnbull and Daranas, 2003). Using chemical shift perturbation NMR experiments, it was shown here that the interaction between human GII.4 NoV P-domains and glycans from the HBGA family are in the low mM-range and, thus, 10- to 100-times weaker than previously thought (cf. chapter 3.1.4). It is noteworthy that in a recent study using membrane-embedded blood group B antigens and a quartz crystal microbalance monitoring dissipation, the monovalent K_D of the interaction has been estimated as 2 mM (Parveen et al., 2019)

which is in excellent agreement with the NMR data presented here. Essentially, only two different techniques have been used to quantify HBGA binding affinity in earlier studies: native mass spectrometry (MS) (Han et al., 2013, 2015) and saturation transfer difference (STD) NMR spectroscopy (Hansman et al., 2012, Mallagaray et al., 2015, 2017, Wegener et al., 2017).

In native MS, folded proteins are ionized together with varying amounts of putative ligand molecules and the appearance of heavier protein-ligand complexes is used to derive binding site occupancies and, thereby, K_D values (Sun et al., 2006). Notably, the method requires subtraction of 'unspecific' interaction contributions by use of an unrelated, non-binding 'reference protein' – often cytochrome C (12 kDa) – that is ionized in parallel. 'Unspecific' interactions are thought to arise from various gas phase processes after electrospray ionization that cannot occur in solution (Robinson et al., 1996, Wang et al., 2003). The identified amount of molecules that are 'bound unspecifically' to the reference protein is then subtracted from the bound molecules to the actual target protein. It appears likely that for certain combinations of target proteins and ligand molecules, the type of reference protein used is of critical importance, a sometimes ignored aspect. As a consequence, derived binding site occupancies might deviate substantially from their real values in solution and, in extreme cases, this can lead to false-positively identified ligands (see below).

The reason for previous over-estimations of binding affinities by STD NMR remain mostly elusive. In these experiments, several step-like discontinuities have been observed in HBGA binding isotherms that were initially linked to cooperative occupation of several binding pockets (Mallagaray et al., 2015). Curve fitting to an extended Hill equation lead to the low μM affinity values reported in these studies. No such steps in binding isotherms have been found in the CSP titration experiments performed here. Consequently, binding follows the predictions from a two-state model with two independent HBGA binding pockets with no assumptions on binding cooperativity required. Qualitatively, fitting of the reported STD titration curves to this simpler model would possibly yield affinities of the correct order of magnitude, at the cost of ignoring the observed systematic step-like deviations. Unfortunately, at the moment no explanation can be given for the origin of these steps and efforts need to be made to replicate these results with stringent control of sample conditions as pH and deamidation status of the protein during the titration.

Both native MS and STD NMR experiments have also been used to identify charged, sialic acid-containing gangliosides as ligands for the GII.4 NoV P-domain (Han et al., 2014, Wegener et al., 2017). In this thesis, it is demonstrated that these reports were false-positive for the reasons outlined above. CSP NMR experiments unambiguously prove that there is no interaction between the P-domain and sialic acid or the GM3 carbohydrate. Three different GII.4 NoV P-domains were tested in a variety of buffers, including proteins from the strains MI001 and VA387 that were subject of the studies mentioned above. As outlined in chapter 5.3, there are many ways in which CSP experiments can falsely identify binders. A false-negative result of CSP experiments, however, appears highly unlikely (if not impossible), assuming correct protein folding and the presence of all necessary cofactors. Thus, earlier studies showing no ganglioside-type glycan-capsid interactions were correct (Fiege et al., 2012, Han et al., 2013).

Similarly, the murine NoV P-domain has also been described as a ganglioside-type glycan binding protein before (Taube et al., 2009, 2012) but was shown here not to interact with ganglioside glycans in vitro. The initial report on ganglioside recognition was based on a variety of cell culture-based experiments, including the use of sialidases, inhibitors of ganglioside synthesis, and interaction-blocking antibodies or lectins and their effects on viral infectivity (cf. chapter 3.1.4). There are several possible reasons that could explain the discrepancy between in vivo and in vitro experiments.

(1) An unidentified cofactor could be missing that is present in cell culture but was not included in the NMR experiments. GCDCA and bound metal ions have been excluded as such cofactors, but others could potentially exist.

(2) Recognition of sialosides could be pH-dependent. Studies with the soluble murine NoV P-domain are restricted to slightly acidic pH values (5-6) for reasons of protein stability. Binding could theoretically occur at higher or lower pH values. The relevance of binding under these conditions is questionable as the mean intestinal pH in mice and rats corresponds very well to the pH range studied here (McConnell et al., 2008).

(3) The putative binding pocket could be in a region that is only present in the fully assembled capsid (e.g. at the interface to the shell domain or within the shell domain itself) and not, as initially claimed, at the topside of the P-domain. The binding pocket on top of the P-domain was postulated based on four amino acid exchanges that would eliminate the sialic acid-dependent cell culture behavior. The CW1 murine NoV strain used here is identical with the 'binding' strain described in Taube et al. (2012), whereas the MNV07 strain is a hybrid with two out of four amino acids corresponding to the 'binding' and the others to the 'non-binding' strain. However, no strain-dependent differences were observed here. Both strains were found not to bind sialic acid-containing or any other glycans. Of note, three out the four identified amino acids belong to the loop coordinating the bivalent metal ions within the P-domain (cf. Fig. 5.3, binding site **II**). The relevance of this observation remains to be seen.

(4) Sialylated glycans may not undergo direct interactions with the capsid but are involved at another point in the infection cycle. Expression of the protein receptor CD300lf is sufficient to render cells permissive for murine NoV infections (Haga et al., 2016). No such connection has been made with any glycan, raising the question if they are truly needed as attachment factors for murine NoV. Notably, members of the CD300 family carry potential glycosylation sites (Comas-Casellas et al., 2012).

5.5 Conclusions & Outlook

Biophysical and biochemical characterization of the protruding domains of human and murine NoV capsids has revealed both a set of shared properties but also a surprising number of fundamental differences between both proteins. Considering their extraordinary structural similarity, as visible for instance in Fig. 3.4, the comparative study presented here sheds light on the question whether properties of viral capsid proteins are transferable onto their homologs.

On a molecular biologicy level, both proteins behave similarly and can be synthesized recombinantly and purified in an almost identical manner. Yet, their difference in protein stability is remark-

able: whereas the human NoV P-domain is stable at room temperature for weeks, the life time of the murine NoV P-domain is limited to a few days. This property also manifests itself in the range of pH values in which the two proteins can be studied. The human NoV P-domain tolerates a broad range of pH values well – in striking contrast to the murine NoV P-domain with its small stability window at slightly acidic conditions. These differences could well be linked to the fundamental finding that the proteins' tendencies for self-association and dimerization are vastly different. The dimer dissociation rate of murine NoV P-domain homodimers is approximately four orders of magnitude larger than that of human NoV P-dimers. On the contrary, a post-translational degradation reaction such as the asparagine deamidation identified in GII.4 NoV appears to have no equivalent in murine NoV so far.

All NoV capsid proteins have evolved to interact with small molecules that are abundantly present in the gastrointestinal tracts of mice and Man. The P-domains of both the predominant human GII.4 NoV and murine NoV can bind bile acids but molecular details of these interactions differ greatly. Firstly, the herein identified binding pocket in GII.4 NoV capsids is located in a separate region of the protein and, naturally, involves other amino acids. This is accompanied by a significant difference in ligand affinity. Murine NoV bind bile acids with a dissociation constant in the low μM-range – a factor of thousand better than human GII.4 NoV. The biological importance of bile acid binding of GII.4 NoV remains unclear, although bile acid concentrations in the gastrointestinal tract are high enough for significant occupation of this pocket in vivo. For the murine NoV, however, certain conclusions can be made about the importance of bile acid binding in viral infection. The observed drastic change in protein dynamics in connection with existing structural models implies a role in viral immune escape. On another level, bile acid-mediated strengthening of the interaction between two P-domain molecules might be relevant in capsid assembly or disassembly. There is also growing evidence for metal ions as cofactors in the infection processes of murine and, at least some, human NoV. Here, initial characterizations of the different metal ion binding sites have been made, highlighting the role of NoV P-domains as metalloproteins. Finally, the data assembled here redefines the role of glycan binding for both viruses. Apparently, previous reports on the interactions between murine NoV P-domains and carbohydrates were false-positive and glycans are not at all involved as attachment factors for murine NoV. Human NoV bind soluble histo-blood group antigens but it could be shown that their affinity has been overestimated in previous studies. Importantly, human NoV capsids do not bind to sialylated glycans.

As human NoV infection in general does require HBGA binding and murine NoV capsid proteins were shown not to bind glycans at all, this study questions the use of murine NoV capsid proteins as a model system for their human NoV equivalents. Despite this fact, sophisticated mechanisms for viral immune escape and capsid stabilization have apparently evolved in murine NoV that are worth investigating further. Considering that murine NoV are amenable to a large range of experimental approaches from modern biology – from structural studies in vitro to cell culture experiments and in vivo trials in laboratory mice – murine NoV will certainly remain a model virus of broad significance.

From a structural biology perspective, NMR investigations into the mechanism governing the murine NoV's immune evasion are highly desirable. These studies could make use of already described

antibody escape mutants that are far away from the actual antibody binding epitope. This, of course, would require the challenging backbone NMR signal assignment as already obtained here for a human NoV P-domain. Assignment of the murine NoV P-domain, however, might substantially benefit from its capability to bind lanthanide ions as their paramagnetic effects can be used to guide the assignment. With an assignment at hand, studies into the dynamics of the P-domain will become feasible and might elucidate the nature of the observed drastic changes upon ligand binding. The binding of paramagnetic metal ions can be further exploited to identify other possible small molecule ligands in simple proton NMR screening experiments as established here in a first pilot experiment.

For human NoV, several questions arise as a consequence of the identified deamidation reaction. What is the molecular explanation for the exceptionally fast reaction rate, in particular with respect to comparable model peptides? Are there differences between the strains investigated here and others with varying amino acid sequences in the respective loop? Can asparagine deamidation be linked to a change in antibody recognition, thereby establishing a novel mechanism for viral immune escape?

Here, a set of experimental approaches has been established to characterize both human and murine P-domain dimerization reactions. This enables further studies into environmental factors influencing capsid assembly and disassembly, at least on the level of the isolated protruding domain. Little is known so far about this process, and especially the role of gastrointestinal as well as intracellular pH variations is worthwhile investigating.

6. Materials & Methods

6.1 Expression Vectors

6.1.1 Genes

Table 6.1: Protruding domain sequences and molecular characteristics. Amino acid sequences are given for all NoV P-domain proteins used in this thesis. GenBank (Benson et al., 2013) identifiers were obtained from www.ncbi.nlm.nih.gov. Two to four amino acids were added N-terminally to yield enzymatic cleavage sites necessary for protein purification. Molecular weights and UV molar extinction coefficients ϵ_{280} refer to the monomeric P-domain and were calculated using the ProtParam tool (www.web.expasy.org/protparam/) (Gasteiger et al., 2005).

NoV strain	GenBank ID	VP1 amino acids (N-terminal additions)	molecular weight in kDa	ϵ_{280} in $\text{M}^{-1}\text{cm}^{-1}$
GII.4 Saga 2006	AB447457	225-530 (GPGS)	34.1	35410
GII.4 MI001	KC631814	225-530 (GPGS)	34.1	33920
GII.4 VA387	AY038600	225-529 (GPGS)	34.0	36900
GII.10 Vietnam	AF504671	224-538 (GPGS)	34.7	40910
GV CW1	DQ285629	228-530 (GP)	33.2	46870
GV MNV07	AET79296	228-530(GPGS)	33.4	46870
GV CR10	ABU55613	228-530(GPGS)	33.4	46870

pMAL-c2X expression plasmids (Walker et al., 2010) contained the genes of human and murine NoV protruding domains (P-domains) (Tab. 6.1). The vectors carried genes for ampicillin resistance and isopropyl thiogalactoside-inducable (IPTG) *tac* promoters. The expressed polypeptide is a fusion protein of maltose-binding protein (MBP) and the P-domain, separated by two His_8-tags and a cleavage site of the HRV 3C protease (`LEVLFQGP`).

6.1.2 Polymerase Chain Reaction (PCR)

Point mutants of the P-domain from GII.4 Saga were generated by site-directed mutagenesis. Methylated template DNA was obtained from *E. coli* DH5α cells carrying the plasmid with the wild type DNA sequence as explained in chapter 6.1.5. 0.47 µL of 100 ng µL^{-1} template were mixed with 5 µL 10x *Pfu* polymerase buffer (Promega), 1.25 µL of the respective forward and reverse primers carrying the mutation sites for N373D or N373Q given in Tab. 6.2 (Eurofins), 1 µL 10 mM dNTP mix (New England Biolabs), 41.28 µL sterile H_2O and 1 µL *Pfu* polymerase (Promega) in sterile 0.5 mL reaction tubes on ice. The PCR was performed using the following incubation times and temperatures: 3 min at 95 °C, 19 times (50 s at 95 °C, 30 s at 65 °C, 18 min at 72 °C), 9 min at 72 °C for a total of 6 h 23

min. Template DNA was digested by addition of 1 µL *DPN1* (Promega) to the reaction mixture and incubation at 37 °C for 1 h.

Table 6.2: Primers for site-directed mutagenesis of the GII.4 Saga protruding domain.

mutant	sequence (5' to 3')
N373Q for	CAGTTTAGCACCGATACCGAACAGGATTTTGAAACCCATCAG
N373Q rev	CTGATGGGTTTCAAAATCCTGTTCGGTATCGGTGCTAAACTG
N373D for	CAGTTTAGCACCGATACCGAAGACGATTTTGAAACCCATCAG
N373D rev	CTGATGGGTTTCAAAATCCTGTTCGGTATCGGTGCTAAACTG

6.1.3 Bacterial Growth Media

Table 6.3: Media used for bacterial growth and protein biosynthesis.

LB medium	$10\,g\,L^{-1}$ peptone, $5\,g\,L^{-1}$ yeast extract, $10\,g\,L^{-1}$ NaCl, pH 7.3, $100\,\mu g\,mL^{-1}$ ampicillin
TB medium	$12\,g\,L^{-1}$ tryptone, $23.8\,g\,L^{-1}$ yeast extract, 10 mL glycerol, $0.55\,g\,L^{-1}$ NaCl, $3.3\,g\,L^{-1}$ KH_2PO_4, $16.63\,g\,L^{-1}$ $Na_2HPO_4\,x\,12\,H_2O$, $1\,g\,L^{-1}$ NH_4Cl, 1 mM $MgCl_2$, 0.1 mM $CaCl_2$, 0.2 % w/v glucose, $10\,\mu g\,mL^{-1}$ vitamin B1, 0.4 % w/v casamino acids, $100\,\mu g\,mL^{-1}$ ampicillin
minimal medium ([*U*-^{2}H,^{15}N]-labeling) in 99.8 % D_2O, pH* 7.4	$1.88\,g\,L^{-1}$ KH_2PO_4, $5.42\,g\,L^{-1}$ Na_2HPO_4, $1\,g\,L^{-1}$ NaCl, 4 mM $MgSO_4$ 0.1 mM $CaCl_2$ $10\,\mu g\,mL^{-1}$ vitamin B1, 1 mM $MgCl_2$ $100\,\mu g\,mL^{-1}$ ampicillin 24 % v/v OD2 DN full medium (Silantes) $0.4\,mg\,L^{-1}$ vitamin B2 $4\,mg\,L^{-1}$ vitamins B3, B5-B7, B9, B12 $4\,mg\,L^{-1}$ choline chloride

6.1.4 Bacterial Transformation

To yield chemically competent *E. coli* BL21 (DE3) (Jeong et al., 2009), cells were grown overnight in LB medium (Tab. 6.3) without ampicillin at 37 °C. 100 mL LB medium without ampicillin were

Table 6.4: Buffers used to produce chemically competent *E. coli* cells.

CM1	10 mM sodium acetate (pH 5.6), 50 mM $MnCl_2$, 5 mM NaCl
CM2	10 mM sodium acetate (pH 5.6), 5 mM $MnCl_2$, 70 mM $CaCl_2$, 5 % v/v glycerol

inoculated with 1 mL of this culture and grown to an optical density at 600 nm (OD_{600}) of 0.4. The culture was incubated on ice for 10 min and a cell pellet was obtained by centrifugation. The pellet was resuspended in 20 mL chilled CM1 buffer (Tab. 6.4) and incubated on ice for 20 min. Again, a cell pellet was produced by centrifugation, resuspended in 2 mL chilled CM2 buffer and immediately stored at -80 °C.

LB agar plates were produced by adding 15 g L^{-1} agar to LB medium, sterilization, and addition of 100 µg mL^{-1} ampicillin. For bacterial transformations, 100 ng plasmid DNA (cf. chapter 6.1.1) or 1 µL PCR reaction product were added to thawed, competent *E. coli* BL21 (DE3) cells. Cells were incubated on ice for 15 min and subjected to a heat shock at 42 °C for 1 min. 1 mL LB medium without ampicillin were added and cells were grown at 37 °C for 1 h. 100 µL cell culture were spread on LB agar plates and incubated overnight at 37 °C.

6.1.5 Bacterial Cryo Stocks

3 mL LB medium (Tab. 6.3) were inoculated with bacterial clones from LB agar plates. Cells were grown at 37 °C until an OD_{600} of 1.5 was reached. Sterile glycerol was added with a final concentration of 15 % v/v and cryo stocks were stored at -80 °C.

To verify the success of bacterial transformations, 10 mL LB medium (Tab. 6.3) were inoculated from cryo stocks. Cells were grown overnight at 37 °C. Plasmid DNA was extracted using the NucleoSpin Plasmid (NoLid) kit (Macherey-Nagel) following the manufacturer's instructions. DNA was sequenced by *LGC Genomics* with commercial *malE* and *pTrcHis rev* sequencing primers (Eurofins).

6.2 Protein Biosynthesis

6.2.1 Unlabeled Protruding Domain

50 mL LB medium were inoculated with *E. coli* BL21 (DE3) transformed with plasmids encoding the NoV P-domain from bacterial cryo stocks (cf. chapters 6.1.1-6.1.5). The growth temperature was 37 °C unless stated otherwise. After 16 h, cells were diluted in TB medium (Tab. 6.3) to an OD_{600} of 0.05. Cells were grown until an OD_{600} of 1.5 was reached. Then, the growth temperature was reduced to 16 °C and expression was induced with 1 mM IPTG. Growth was continued for 24 h until the stationary phase was reached and cells were harvested by centrifugation. Cell pellets were stored at -80 °C.

6.2.2 Isotope-Labeled Protruding Domain

For synthesis of [*U*-^{2}H,^{15}N]-labeled P-domains, minimal media were used with fully deuterated glucose and $^{15}NH_4Cl$ (Deutero) as sole carbon and nitrogen sources, respectively (Tab. 6.3). NaCl, KH_2PO_4, Na_2HPO_4, deuterated glucose, and $^{15}NH_4Cl$ were dissolved in D_2O (Deutero) and lyophilized prior to use. 50 mL LB medium were inoculated and cells were grown overnight at 37 °C. A fraction of the medium corresponding to an OD_{600} of 0.05 in 40 mL was centrifuged and the bacterial pellet was resuspended in minimal medium . Cells were grown until an OD_{600} of 0.4 was reached and then diluted to the final culture volume (0.5-1 L). At an OD_{600} of 0.8, expression was induced with 1 mM IPTG and growth was continued at 16 °C until the stationary phase was reached. Murine NoV P-domains were synthesized in media without addition of isotope-labeled DN OD2 media (Silantes).

6.3 Protein Purification & Biochemical Analysis

6.3.1 Cell Lysis

Frozen cell pellets of human and murine NoV P-domains were resuspended in 75 mM phosphate buffer, 100 mM NaCl, pH 7.3 or 20 mM sodium acetate, 100 mM NaCl, pH 5.3, respectively. Additionally, resuspension buffers contained 160 mU mL^{-1} benzonase (Novagen), 0.27 µg mL^{-1} leupeptin (Roth) and aprotinin (Roth), and 16 µg mL^{-1} lysozyme (Merck). Cells were lysed using a high-pressure homogenizer (Thermo) operating at 1400 psi. The lysate was clarified by centrifugation at 9000 g and 4 °C for 1 h; the remaining cell pellet was discarded.

6.3.2 Affinity Chromatography & Proteolytic Cleavage

All chromatographic purification steps were performed on an ÄKTA pure system (GE Life Sciences) at 4 °C. UV absorption was monitored at 280 and 214 nm. Two 5 ml HisTrap HP affinity columns (GE Life Sciences) were connected in series and equilibrated in 20 mM sodium phosphate buffer, 100 mM NaCl, 20 mM imidazole (pH 7.3) for human NoV P-domains or 20 mM sodium acetate, 100 mM NaCl (pH 5.3) for murine NoV P-domains. Cell lysate was loaded onto the columns using the system pump with a flow-rate of 5 mL min^{-1}. Columns were washed with equilibration buffer until a steady UV baseline was reached. Proteins were eluted step-wise, using first 10 % and then 100 % of 20 mM sodium phosphate buffer, 100 mM NaCl, 500 mM imidazole (pH 7.3). The fusion protein of MBP and P-domain (cf. chapter 6.1.1) eluted in the last step; protein identity and purity was characterized by gel electrophoresis (cf. chapter 6.3.4).

The fusion protein was collected, 5 mM β-mercaptoethanol and 200 µL of a 0.4 mg mL^{-1} HRV 3C protease solution (synthesized in-house) were added. The mixture was dialyzed overnight against affinity chromatography equilibration buffer using dialysis tubing with 10 kDa cutoff (ThermoFisher). Completeness of the proteolytic cleavage reaction was verified by gel electrophoresis. The solution was again subjected to affinity chromatography; the cleaved P-domain does not bind to the column and elutes immediately whereas protease, fusion protein, and MBP are being retained on the column.

6.3.3 Size Exclusion Chromatography (SEC)

Preparative SEC

P-domain solutions were concentrated to a final volume of 3.5 mL using 50 mL centrifugal filters with 10 kDa cutoff (Merck). The protein solution was applied onto a preparative Superdex 16/600 200 pg size exclusion column (GE Life Sciences) operating with a flow-rate of 1 mL min^{-1} using a 7 mL sample loop. 20 mM sodium phosphate buffer (pH 7.3) or 20 mM sodium acetate, 100 mM NaCl (pH 5.3) were selected as running buffers for human and murine NoV P-domains, respectively. The purified protein was quantified using a UV spectrophotometer (Peqlab) with 280 nm wavelength and the respective molar extinction coefficients given in Tab. 6.1. Proteins were stored in their respective running buffers at 4 °C.

Analytical SEC

Analytical size exclusion chromatography has been performed with a Superdex 75 Increase 3.2/300 column (GE Life Sciences) and 20 mM sodium acetate, 100 mM NaCl (pH 5.3) as running buffer with a flow-rate of 0.075 mL min^{-1} at 278 K. To study the effect of ligand binding to the murine NoV P-domain, 100 µM glycochenodeoxycholic acid (GCDCA, Sigma Aldrich) were added to the running buffer. Protein samples with concentrations from 250 nM to 270 µM were applied with a 20 µL sample loop using a complete loop filling technique. UV absorption was monitored at 280 and 214 nm simultaneously. A molecular weight calibration was performed with a mixture of 1.5 mg mL^{-1} conalbumin (75 kDa, Sigma), 4 mg mL^{-1} ovalbumin (44 kDa, Sigma), 1.5 mg mL^{-1} carbonic anhydrase (29 kDa, Sigma), 4 mg mL^{-1} cytochrome c (12.3 kDa, Sigma), and 1 mg mL^{-1} aprotinin (6.5 kDa, Roth). Apparent molecular weights of murine NoV P-domains as a function of feed protein concentration were fitted to Eq. 6.1 to yield an apparent dissociation constant K_D of the protein-protein interaction as well as apparent molecular weights of the pure monomer and dimer (modified from Eq. 4 of Thiele and Huttner (1998) or Eq. 4 of Åkerud et al. (2002)).

$$MW_{obs} = MW_D + \frac{\sqrt{K_D(K_D + 8P_t)} - K_D}{4P_t}(MW_M - MW_D) \tag{6.1}$$

MW_{obs} is the observed molecular weight according to the calibration of the column using standard proteins, MW_D and MW_M are the apparent molecular weights of the monomers and dimers, respectively, and P_t is the total protein concentration applied onto the column .

The dimer association rate k_{on} can be estimated from the appearance of peak tailing at protein concentrations higher than 6 µM using the relationship derived in Yu et al. (2006):

$$N_{k\,\mathrm{on}} = \frac{L * k_{on} * c}{u_0} = 10 \tag{6.2}$$

with L being the length of the column (30 cm), c being the lowest feed concentration at which tailing appears (6 µM), and u_0 being the linear flow rate (0.9325 cm min^{-1}).

In a monomer-dimer equilibrium 2 M $\rightleftharpoons$ D, the dissociation constant K_D and the total protein

concentration P_t are defined as:

$$K_D = \frac{[M]^2}{[D]}, P_t = [M] + 2[D] \tag{6.3}$$

The exchange rate k_{ex} is defined as (Åkerud et al., 2002):

$$k_{ex} = k_{off} + k_{on}[M] \tag{6.4}$$

It follows, that:

$$k_{ex} = k_{off} + k_{on}\left(-\frac{K_D}{4} + \sqrt{\frac{K_D^2}{16} + \frac{K_D P_t}{2}}\right) \tag{6.5}$$

6.3.4 Denaturing Gel Electrophoresis

Samples were diluted 1:1 with reducing sample buffer (12 % v/v glycerol, 3 % w/v sodium dodecyl sulfate, 1 mg mL^{-1} bromphenol blue, 300 mg mL^{-1} urea, 180 mM Tris-OH, pH 6.8) and incubated at 80 °C for 10 min. Samples were loaded on 15-well, 12 % acrylamide gels (Mini-PROTEAN TGX gels, Bio-Rad). Gel electrophoresis was performed at 180 V with 6.05 g L^{-1} Tris, 28.8 g L^{-1} glycine, and 1 g L^{-1} sodium dodecyl sulfate as running buffer. Gels were stained using Coomassie Brilliant Blue staining solution (Bio-Rad).

6.3.5 Chemical Denaturation & Refolding of the GII.4 Protruding Domain

Denaturation and refolding of GII.4 P-domains was performed at 4 °C. Protein samples were prepared in 20 mM sodium phosphate buffer (pH 7.3) at a protein concentration of 7 mg mL^{-1}. 1 mL of the protein solution was transferred into 35 mL of unfolding buffer (4 M guanidine chloride, 0.5 M Tris, 0.3 M NaCl, 10 mM β-mercaptoethanol (BME), pH 7.3) and incubated for 4 h. Then, refolding was started by addition of 35 mL of stabilization buffer (3 M guanidine chloride, 0.5 M Tris, 1 M L-proline, 0.4 M D-sucrose, 10 mM BME, pH 7.3) to the reaction mixture and subsequent dialysis against 2 L of refolding buffer (0.5 M Tris, 0.5 M L-proline, 0.2 M D-sucrose, 10 mM BME, pH 7.3) overnight. The protein solution was then dialyzed step-wise against 2 L of diluted refolding buffer (0.25 M Tris, 0.25 M L-proline, 0.1 M D-sucrose, 5 mM BME, pH 7.3) for 8 h, 2 L of 50 mM Tris, 0.45 M NaCl, 5 mM BME (pH 7.3) once for 8 h and twice for 24 h.

6.3.6 Ion Exchange Chromatography (IEX)

Separation of fully-, partially, and non-deamidated P-domain species was achieved by cation exchange chromatography using a 6 mL Resource S column (GE Life Sciences) at 4 °C. Protein samples were prepared in 20 mM sodium acetate buffer (pH 4.9) and eluted using a 78.4 mL linear salt gradient up to 195 mM NaCl with a flowrate of 1 mL min^{-1}. The deamidation reaction rate at 310 K was determined by incubation of 280 µg GII.4 Saga P-domain in 125 µL aliquots of 75 mM sodium phosphate buffer, 100 mM NaCl (pH 7.3) for up to 48 h and subsequent analytical ion exchange chromatography after selected time intervals. The buffer dependency of the deamidation reaction rate was determined by

incubation of 0.4-0.75 mg protein in 200-300 µL aliquots in 75 mM sodium phosphate buffer, 100 mM NaCl (pH 7.3); 75 mM sodium phosphate buffer (pH 7.3); 20 mM sodium phosphate buffer (pH 7.3); or 20 mM sodium acetate (pH 4.9) at 298 K. GII.4 MI001 and GII.10 Vietnam were tested for deamidation by incubation in 25 mM Tris, 300 mM NaCl (pH 7.3) for three weeks at 298 K. N373D GII.4 Saga P-domain was tested for interconversion into isoAsp by incubation of purified protein aliquots with a concentration of 1.3 mg mL^{-1} in 75 mM sodium phosphate buffer, 100 mM NaCl (pH 7.3) for 382 h at 25 °C and subsequent IEX analysis.

Baseline correction and quantification of peak integrals of the native P-domain species was performed using *Unicorn 7* software (GE Life Sciences). Experimental half-life times were obtained by fitting the normalized UV integrals of the non-deamidated species at different time points to an exponential decay model.

The dimer dissociation rate k_{off} of GII.4 Saga P-domains was determined by mixing two dimer species with different net charges and subsequent IEX analysis. Fully deamidated GII.4 Saga dimers (iD/iD) were obtained by preparative IEX. A 373Q point mutant was generated by site-directed mutagenesis (cf. chapter 6.1.2). Both types of dimers were mixed 1:1 with a total protein concentration of 1.24 mg mL^{-1} in 75 mM sodium phosphate buffer, 100 mM NaCl, 0.02 % v/v NaN_3 (pH 7.3) and incubated at 298 K. The increase in the fraction of iD/Q heterodimers $f_{iD/Q}$ was fitted to Eq. 6.6 (Chevreux et al., 2011).

$$f_{iD/Q} = \frac{1}{2}\left(1 - e^{-k_{\text{off}}t}\right) \tag{6.6}$$

Real macroscopic reaction rates of the deamidation reaction were obtained from two data sets of GII.4 Saga in 75 mM sodium phosphate buffer, 100 mM NaCl (pH 7.3) at 298 K taking into account the simultaneous monomer-dimer exchange. The experiment series was started with either non-deamidated N/N dimers (cf. description of buffer dependency above) or iD/N heterodimers. Experiments were performed with protein concentrations of 55 µM and 35 µM, respectively. The system of ordinary differential equations (ODE) 4.1 - 4.5 was derived on the basis of an ODE system describing monomer-dimer exchange without a deamidation reaction described in Goulet (2016). Section 3.2.2 of the aforementioned study justifies the introduction of the correction factor to account for the increased likelihood of heterodimer formation. The system of ODEs was solved numerically with the *Python* source code presented in chapter 6.10.2. Briefly, *Scipy*'s (Oliphant, 2007) `odeint` solver was used for solving the ODE. Initial values at $t = 0$ for the different species were set according to the starting protein species used: either $[\text{NN}]_{t=0} = 1$ or $[\text{iDN}]_{t=0} = 1$. The deamidation reaction rate k_1 was varied from $0.01 - 10 \times 10^{-6} s^{-1}$ and squared residuals were calculated between experimental and simulated data. k_1 was determined by a least-squares approach.

6.3.7 Hydrophobic Interaction Chromatography (HIC)

A 1 mL HiTrap Butyl-HP column (GE Life Sciences) was used for hydrophobic interaction chromatography of the murine NoV P-domain. CR10 and CW1 protein samples (240 and 270 µg mL^{-1} for pH and GCDCA data sets, respectively) were prepared in different buffers and then subjected to isother-

mal denaturation at 45 °C. The effect of GCDCA binding was studied using 20-1000 µM GCDCA in 20 mM sodium acetate, 100 mM NaCl (pH 5.3) with 30 min incubation time, while experiments on pH-dependence were performed in 75 mM sodium phosphate buffer, 100 mM NaCl with pH values ranging from 5.7 to 8 with 10 min incubation time. Prior to application onto the column, 750 mM ammonium sulfate were added from a concentrated stock solution. The protein bound to the column in 20 mM sodium acetate, 750 mM ammonium sulfate (pH 5.3) and was eluted using a linear gradient over 5 column volumes up to 100 % of 20 mM sodium acetate (pH 5.3). A flow rate of 3 mL min^{-1} was used. The UV integral at 214 nm of a non-heat-treated control sample was normalized to 1. Experiments were performed as duplicates, the respective percentage of deviation is given as error bars.

6.4 Nuclear Magnetic Resonance (NMR) Spectroscopy

6.4.1 NMR samples

NMR samples were prepared in 3 mm NMR tubes with a volume of 160 µL. NMR experiments with [*U*-^{2}H,^{15}N]-labeled proteins were acquired with samples containing 10 % D_2O. Samples with human NoV P-domains contained 200 µM deuterated 2,2-dimethyl-2-silapentane-5-sulfonate-d_6 (DSS-d_6, Sigma-Aldrich) and 300 µM imidazole (Roth) to monitor the sample pH (Baryshnikova et al., 2008). Murine NoV P-domain samples contained 500 µM DSS-d_6. Protein concentrations and sample buffers are given in the figure legends of the respective data sets. Sample pH values were measured using an Orion 8220BNWP pH-electrode (ThermoFisher) with an Orion Star A221 pH-meter (ThermoFisher) after calibration with pH 4, 7 and 10 standard solutions (Roth). pH values of samples with varying amounts of D_2O are given as uncorrected pH-meter readings pH* (Krezel and Bal, 2004). Saturation transfer difference (STD) NMR experiments were measured with unlabeled proteins. Concentrations of potential ligands and proteins in chapter 4.6 are given in the respective tables. Samples for STD negative control experiments were prepared with identical small molecule concentrations without the addition of proteins.

6.4.2 Chemical Shift Referencing

Proton chemical shifts were set to 0 ppm at the position of the DSS-d_6 signal. Chemical shift referencing in indirect ^{13}C and ^{15}N dimensions was performed according to Eq. 6.7 using the source code in chapter 6.10.1.

$$SR_X = BF_H + SR_H * \Xi - BF_X \tag{6.7}$$

SR_X is the reference shift applied to the indirect dimensions, BF_H and BF_X are the base frequencies for the respective dimensions in Hz, SR_H is the reference shift of the direct dimension, Ξ is the relative gyromagnetic ratio of the respective nucleus and protons.

6.4.3 Backbone Amide Signal Assignment - 3D NMR Spectroscopy

3D NMR spectra of GII.4 Saga were obtained with a 180 µM [*U*-^{2}H,^{13}C,^{15}N]-labeled sample of GII.4 Saga P-dimers produced in-house and provided by Dr. Alvaro Mallagaray. Spectra were acquired on 900 and 500 MHz spectrometers (Bruker) equipped with TCI cryogenic probes using TROSY versions of standard 3D pulse programs (Salzmann et al., 1998, Eletsky et al., 2001) at 298 K. Experimental details are given in Tab. 6.5. Assignment data was deposited with the BMRB (entry ID: 27445)[1]. The TALOS-N web server[2] was used for chemical shift-based secondary structure prediction with corrections for the deuterium isotope effect.

Table 6.5: Parameters for 3D NMR pulse programs. TD denotes the number of increments in the respective dimensions, O1 is the center of the spectral window, SW is the sweep width, and NS is the number of scans.

experiment	**field strength**	**F3 (TD; O1; SW)**	**F2 (TD; O1; SW)**	**F1 (TD; O1; SW)**	**NS**
HNCO	900 MHz	1024; 4.71 ppm; 14 ppm	64; 117 ppm; 31 ppm	128; 176 ppm; 18 ppm	16
HN(CA)CO	900 MHz	2048; 4.71 ppm; 14 ppm	64; 117 ppm; 31 ppm	96; 176 ppm; 18 ppm	32
HN(CA)CO	500 MHz	2048; 4.65 ppm; 13 ppm	64; 117.5 ppm; 32 ppm	128; 174 ppm; 16 ppm	32
HNCA	900 MHz	2048; 4.71 ppm; 13 ppm	64; 117 ppm; 31 ppm	128; 54 ppm; 30 ppm	16
HN(CO)CA	500 MHz	2048; 4.65 ppm; 13 ppm	64; 117.5 ppm; 32 ppm	128; 54 ppm; 30 ppm	16
HNCACB	900 MHz	2048; 4.71 ppm; 14 ppm	64; 117 ppm; 31 ppm	96; 40 ppm; 80 ppm	32
HN(CO)CACB	900 MHz	2048; 4.71 ppm; 14 ppm	64; 117 ppm; 31 ppm	128; 40 ppm; 80 ppm	16
HN(CO)CACB	500 MHz	2048; 4.65 ppm; 13 ppm	64; 117.5 ppm; 32 ppm	128; 40 ppm; 80 ppm	16

[1] www.bmrb.wisc.edu/data_library/summary/?bmrbId=27445
[2] https://spin.niddk.nih.gov/bax/software/TALOS-N/

6.4.4 Potential Ligand Molecules Used in STD and CSP Experiments

Table 6.6: Potential ligand molecules used in STD and CSP titration experiments. Carbohydrate nomenclature follows the recommendations of the Consortium for Functional Glycomics (2004). Ganglioside carbohydrates were obtained as sodium salts with a minimal purity of 90 % as specified by the manufacturer by proton NMR spectroscopy. Carbohydrates were anomeric mixtures unless stated otherwise. Bile acids were obtained as sodium salts with purities of 97 % as specified by the manufacturer by TLC or HPLC. Purities and contaminations of metal salts with other bivalent metal ions are as follows: $MgCl_2$: $\geq$ 99%, $\leq$ 0.01% Ca, $\leq$ 0.0005% Mn, $\leq$ 0.001% Zn; $CaCl_2$: $\geq$ 99%, $\leq$ 0.005% Mg; $ZnCl_2$: $\geq$ 98%, $\leq$ 0.001% Mg, $\leq$ 0.001% Ca; $LaCl_3$: $\geq$ 99.9%; $EuCl_3$: $\geq$ 99.99%; $CeCl_3$: $\geq$ 99.9%.

compound	synonyms	source
Galα1-3(Fucα1-2)Galα1-N_3	B-trisaccharide	in-house chemo-enzymatic synthesis as detailed in Mallagaray et al. (2019)
Neu5Acα2-3Galβ1-4Glc	3'SL, 3'-sialyllactose	Elicityl
2-O-Methyl-α-D-N-acetylneuraminic acid	α-D-Neu5Ac-OMe	Carboysynth
Neu5Acα2-3Galβ1-3GalNAc β1-4(Neu5Acα2-3)Galβ1-4Glc	GD1a (glycan)	Elicityl
Methyl α-L-fucopyranoside	α-L-Fuc-OMe	Carbosynth
Galβ1-3GalNAc β1-4(Neu5Acα2-3)Galβ1-4Glc	GM1a (glycan)	Elicityl
N-acetylgalactosamine	GalNAc	Fluka
galactose	Gal	Fluka
Galα1-3Galβ1-OMe	Galili disaccharide	Dextra
Fucα1-2Galβ1-OMe	H-disaccharide	in-house synthesis by W. Hellebrandt
GalNAcα1-3(Fucα1-2)Gal	A-trisaccharide	Carbosynth
GalNAcα1-3GalNAcβ1-3Gal	Forssman trisaccharide	Elicityl
cholic acid	CA	Sigma-Aldrich
deoxycholic acid	DCA	Sigma-Aldrich
glycochenodeoxycholic acid	GCDCA	Sigma-Aldrich
chenodeoxycholic acid	CDCA	Sigma-Aldrich
$ZnCl_2$	-	Merck
$CaCl_2$	-	Sigma
$MgCl_2$	-	Roth
$LaCl_3$	-	Sigma-Aldrich
$EuCl_3$	-	Sigma-Aldrich
$CeCl_3$	-	Sigma-Aldrich

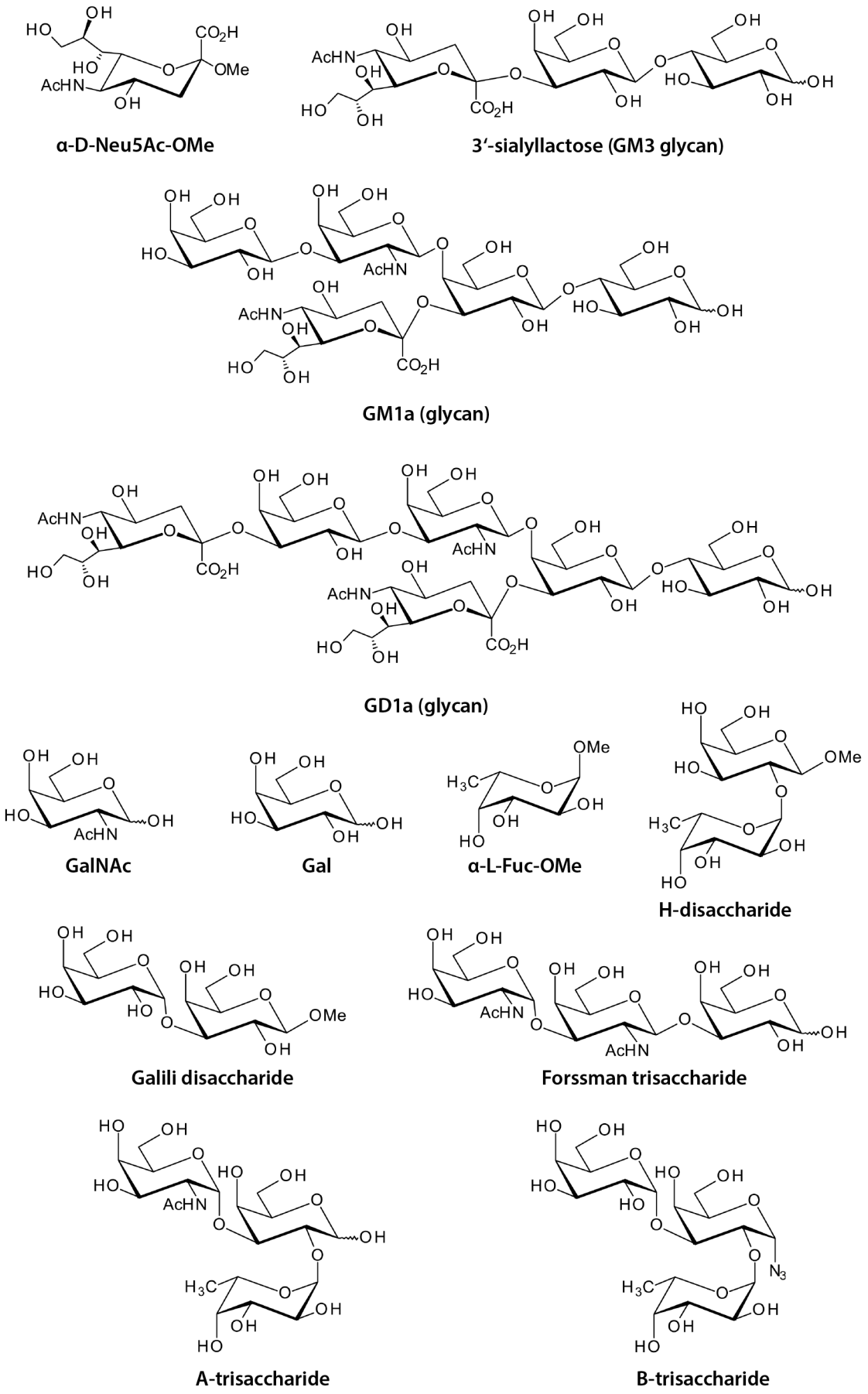


Figure 6.1: Synthetic oligosaccharides used in this thesis.

6.4.5 Chemical Shift Perturbation (CSP) Experiments

^{1}H,^{15}N TROSY HSQC spectra (Pervushin et al., 1997, Schulte-Herbruggen and Sorensen, 2000) were acquired with 128 ms acquisition time and a spectral window of 16 ppm in the direct dimension (Bruker pulse program: `trosyetf3gpsi`). In the indirect dimension, the spectral window was set to 35 ppm with 430 increments for human NoV P-domains and 256 increments for murine NoV P-domains. The relaxation delay was 1.5 s. The number of scans was chosen according to the protein concentration of the sample, ranging from 8 to 48 corresponding to measurement times of 1.6 to 6 h. Unless stated otherwise, spectra were acquired at 298 K on a Bruker Avance III 500 MHz spectrometer with TCI cryo probe. Potential ligand molecules (cf. Tab. 6.6) were titrated from highly concentrated, pH-adjusted stock solutions in the same buffer as the NMR sample to minimize dilution effects and pH-artifacts.

For each titration point, Euclidean chemical shift perturbations $\Delta\nu_{eucl}$ (CSPs) can be calculated between the position of a signal in fast exchange and the corresponding signal in the unbound state according to Eq. 6.8.

$$\Delta\nu_{eucl} = \sqrt{(\Delta\delta_H\omega_H)^2 + (\Delta\delta_N\omega_N)^2} \tag{6.8}$$

$\Delta\delta$ is the chemical shift difference for the given dimension and ω is the respective spectrometer frequency. Observed chemical shift differences of a signal in fast exchange at a given ligand concentration $\Delta\nu_{obs}$ are related to the fractional occupancy of a binding pocket (and thereby to the dissociation constant K_D) via Eq. 6.9 (Williamson, 2013).

$$\Delta\nu_{obs} = \Delta\nu_{max}\frac{(P_t + L_t + K_D) - \sqrt{(P_t + L_t + K_D)^2 - 4P_tL_t}}{2P_t} \tag{6.9}$$

$\Delta\nu_{max}$ is the maximum CSP at saturation of the binding pocket for each signal, P_t is the total protein concentration, and L_t is the total ligand concentration at each titration point. Titration curves for global non-linear least squares fitting were selected according to the magnitude of the CSP at the highest ligand concentration: CSPs larger than the sum of the mean of all CSPs and two standard deviations were used to derive a global K_D value.

For analysis of the titration of $CaCl_2$ to murine NoV P-domains already containing an unknown amount of Ca^{2+}-ions (due to contamination of the GCDCA stock solution), CSP data were fitted using the source code in chapter 6.10.5. Significant CSP data were fitted globally with varying amounts of Ca^{2+}-ions added to those ions titrated on purpose. The concentration at the minimum sum of squared residuals for fitted CSP data was assumed to correspond to the amount of contaminating ions.

6.4.6 Saturation Transfer Difference (STD) NMR Spectroscopy

STD NMR spectra were acquired at 600 MHz with relaxation delays of 5 s (Bruker pulse program: `stddiffesgp.3`). The number of scans and the resulting experiment time were as follows: 1720 - 4240 (7 h - 17 h) for the cholic acid titration to GII.4 Saga 373D P-domains; 1712 (7 h) for the experiment with the MNV07 strain and GM3 at 277 K; 3584 - 4000 (14h - 16 h) for the remaining STD experiments

of Tab. 4.3 and 4.5. On-resonance frequencies and saturation pulse strengths were optimized for each individual sample in negative control experiments without protein and were between -2 and -4 ppm. Off-resonance frequencies were set to 200 ppm. A 2 s pulse cascade of 50 ms Gaussian pulses was used for saturation of protein resonances. A 20 ms spin-lock filter was used for suppression of protein signals.

Saturation transfer from GII.4 Saga 373D P-domains to cholic acid was measured with an on-resonance frequency of -4 ppm, a protein concentration of 90 µM and a maximum cholic acid concentration of 15 mM. Saturation transfer was quantified for the isolated H7 and H12 cholic acid signals by calculation of the STD amplification factor STD_{AF} according to Eq. 6.10 (Mayer and Meyer, 2001).

$$STD_{AF} = \frac{I_{\text{off}} - I_{\text{on}}}{I_{\text{off}}} \frac{c_L}{c_P} \tag{6.10}$$

I_{off} and I_{on} are signal intensities in the respective off- and on-resonance experiment; c_L/c_P is the ligand excess. Data were fitted against Eq. 6.9 with STD_{AF} replacing the observable $\Delta\nu_{obs}$ to obtain K_D values.

6.4.7 ^{15}N Backbone Relaxation

^{15}N backbone relaxation data were obtained using TROSY-based pulse schemes for measurement of T_1 and $T_{1\rho}$ relaxation times (Lakomek et al., 2012) and are given in chapters 6.10.6 - 6.10.7. Spectra were acquired with 256 increments in the indirect dimension and a relaxation delay of 3 s at 500 MHz for 45 h. Delays in the pulse sequence were 0 s, 0.4 s, 0.8 s, 1.4 s, 2 s, and 3 s for determination of T_1 and 1 ms, 5 ms, 10 ms, 18 ms, 25 ms, and 36 ms for determination of $T_{1\rho}$. Both experiments contained a spin-lock temperature compensation element of up to 36 ms. The spin-lock field strength ω in the $T_{1\rho}$ experiment was 1.4 kHz, the carrier frequency in the ^{15}N dimension was 117.3 ppm. T_2 relaxation times were calculated according to Eq. 6.11.

$$\frac{1}{T_2} = \frac{1}{T_{1\rho} \sin^2\theta} - \frac{1}{T_1 \tan^2\theta} \tag{6.11}$$

The angle θ is given by $\theta = \omega/\Omega$ with Ω being each signal's offset in the ^{15}N dimension. Signal-specific intensity data was fitted to exponential decay models with the time constants being T_1 and $T_{1\rho}$, respectively. Intensity errors were estimated from the standard deviation of background noise in regions of the spectrum containing no signals.

Rotational correlation times τ_c of [*U*-^{2}H,^{15}N]-labeled human and murine NoV P-domains have been estimated using TRACT experiments (Lee et al., 2006). The relaxation delay was set to 2 s. Experiments were measured for 20 min and 80 min for human and murine NoV P-domains, respectively with 25 increasing delays in the pulse sequence of up to 0.4 s at 298 K. Data were integrated from 8 to 9.5 ppm for human NoV P-domains and from 6-9.5 ppm for murine NoV P-domains. Data was normalized to the highest intensity, and fitted to an exponential decay model for determination of average ^{15}N R_α and R_β relaxation rates. The difference in relaxation rates was used to calculate τ_c according to Eq. (3)-(5) of Lee et al. (2006). The respective source code can be found in chapter 6.10.3. Molecular

weight estimates based on the rotational correlation time have been obtained using the reference data published in Su et al. (2007).

6.4.8 Dimerization & GCDCA-Binding of the Murine NoV P-Domain

Dimerization and binding of glycochenodeoxycholic acid (GCDCA) were studied using TROSY HSQC experiments with the murine NoV P-domain of the CW1 strain (cf. chapter 6.4.5).

A series of spectra of the CW1 P-domain with concentrations from 25-200 µM was acquired. Nine signals have been identified whose relative intensities increase with increasing protein concentration, whereas these of immediate neighbors decreases (6.9|123.8 ppm, 8.5|130 ppm, 10.0|130.2 ppm, 8.8|125.6 ppm, 8.6|126 ppm, 7.5|131 ppm, 8.8|128.9 ppm, 9|112 ppm, 8.1|128.6 ppm). Intensities of these dimer signals in slow exchange have been normalized to each signal's highest value and were then averaged. Averaged data were divided by the total protein concentration as determined by UV absorption and then fitted to Eq. 6.12 to obtain the dissociation constant K_D of the dimerization reaction.

$$\frac{[D]}{P_t} = O_{max}\frac{P_t + \frac{K_D}{4} - \sqrt{\frac{K_D^2}{16} + \frac{P_t K_D}{2}}}{2P_t} \tag{6.12}$$

$[D]$ is the absolute dimer concentration, O_{max} is a scaling factor accounting for the relationship between measured signal intensity and the underlying protein concentration and P_t is the total protein concentration. The equation can be derived by solving the quadratic equation given in Eq. 6.3 for $[D]$.

GCDCA has been titrated step-wise to a sample with 100 µM protein concentration in 20 mM sodium acetate, 100 mM NaCl (pH 5.3). Monomer and dimer signals in the apo state have been identified by comparison with the concentration series described above. Backbone amide signals were categorized according to their behavior with increasing GCDCA concentrations. Data analysis was performed as described in chapter 4.4. The source code for data fitting is given in chapter 6.10.4. The numerical solution of Eq. 4.10-4.11 for the monomer concentration was obtained using *SciPy*'s `fsolve` function. All other protein concentrations were calculated according to the definition of the respective equilibrium constants in Eq. 4.6. Eq. 6.13-6.14 were used for fitting of CSP data, with Eq. 6.13 describing signals influenced by occupation of both binding pockets and Eq. 6.14 describing signals perturbed by occupation of only the closest binding pocket (Huma et al., 2014).

$$\Delta\nu_{obs} = \Delta\nu_{max}\frac{[DL] + [DL_2]}{[D] + [DL] + [DL_2]} \tag{6.13}$$

$$\Delta\nu_{obs} = \Delta\nu_{max}\frac{[DL] + 2[DL_2]}{2[D] + 2[DL] + 2[DL_2]} \tag{6.14}$$

$\Delta\nu_{obs}$ is the observed Euclidean CSP in Hz, $\Delta\nu_{max}$ is the maximum perturbation at saturation of the binding pocket, $[D]$, $[DL]$, $[DL_2]$ refer to the concentrations of the respective unbound or bound dimer species.

6.5 Software

NMR spectra were acquired and processed using *TopSpin* v3.6 (Bruker). Signal positions and intensities were extracted with *CcpNmr Analysis* v2.4.2 (Vranken et al., 2005). Amino acid sequence alignments were performed with *Unipro UGENE* v1.32 and the *ClustalW* allignment algorithm (Thompson et al., 1994, Larkin et al., 2007). *Maestro* v11.6 (Schrodinger) was used for analysis of crystallographic models.

Data were fitted with *Python* scripts (v2.7.15), using *SciPy*'s (v1.2.1) (Oliphant, 2007) `curve_fit` function for local fitting and `least_squares` for global fitting whenever analytical solutions of the respective equations were available. Fit parameter errors are calculated from the functions' estimated covariance matrices as detailed in the respective documentation[3].

6.6 Data Availability

Experimental raw data, including all NMR spectra and chromatograms, have been deposited together with the respective figure in which they have been used on the server of the Institute of Chemistry & Metabolomics (`\StaffOnly\creutznacher\Raw_data_PhD_thesis`).

[3] https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html, queried in December 2019

Bibliography

M. Allegrozzi, I. Bertini, M. B. L. Janik, Y.-M. Lee, G. Liu, and C. Luchinat. Lanthanide-induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 Å from the metal ion. *Journal of the American Chemical Society*, 122(17):4154–4161, 2000. ISSN 0002-7863. doi: 10.1021/ja993691b.

L. An, Y. Wang, N. Zhang, S. Yan, A. Bax, and L. Yao. Protein apparent dielectric constant and its temperature dependence from remote chemical shift effects. *Journal of the American Chemical Society*, 136(37): 12816–9, 2014. ISSN 1520-5126 (Electronic) 0002-7863 (Linking). doi: 10.1021/ja505852b.

P. C. Aoto, B. T. Martin, and P. E. Wright. NMR characterization of information flow and allosteric communities in the MAP kinase p38γ. *Scientific Reports*, 6:28655, 2016. ISSN 2045-2322. doi: 10.1038/srep28655.

L. W. Arbogast, R. G. Brinson, and J. P. Marino. *Chapter One - Application of Natural Isotopic Abundance 1H–13C- and 1H–15N-Correlated Two-Dimensional NMR for Evaluation of the Structure of Protein Therapeutics*, 566:3–34. Academic Press, 2016. ISBN 0076-6879. doi: 10.1016/bs.mie.2015.09.037.

R. L. Atmar, S. Ramani, and M. K. Estes. Human noroviruses: recent advances in a 50-year history. *Curr Opin Infect Dis*, 31(5):422–432, 2018. ISSN 0951-7375. doi: 10.1097/qco.0000000000000476.

N. Bai, H. Roder, A. Dickson, and J. Karanicolas. Isothermal analysis of thermofluor data can readily provide quantitative binding affinities. *Sci Rep*, 9(1):2650, 2019. ISSN 2045-2322. doi: 10.1038/s41598-018-37072-x.

M. T. Baldridge, H. Turula, and C. E. Wobus. Norovirus regulation by host and microbe. *Trends in Molecular Medicine*, 22(12): 1047–1059, 2016. ISSN 1471-499X (Electronic) 1471-4914 (Linking). doi: 10.1016/j.molmed.2016.10.003.

S. M. Bartsch, B. A. Lopman, S. Ozawa, A. J. Hall, and B. Y. Lee. Global economic burden of norovirus gastroenteritis. *PLoS One*, 11(4):e0151219, 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0151219.

O. K. Baryshnikova, T. C. Williams, and B. D. Sykes. Internal pH indicators for biomolecular NMR. *Journal of Biomolecular NMR*, 41:5–7, 2008. ISSN 0925-2738. doi: 10.1007/s10858-008-9234-6.

D. A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers. Genbank. *Nucleic Acids Res*, 41:36–42, 2013. ISSN 0305-1048. doi: 10.1093/nar/gks1195.

M. Berjanskii, D. Arndt, Y. Liang, and D. S. Wishart. A robust algorithm for optimizing protein structures with NMR chemical shifts. *J Biomol NMR*, 63(3):255–264, 2015. ISSN 1573-5001 (Electronic) 0925-2738 (Linking). doi: 10.1007/s10858-015-9982-z.

C. A. Bewley and S. Shahzad-ul Hussan. Characterizing carbohydrate-protein interactions by nuclear magnetic resonance spectroscopy. *Biopolymers*, 99(10):796–806, 2013. ISSN 0006-3525. doi: 10.1002/bip.22329.

F. Borrego. The CD300 molecules: an emerging family of regulators of the immune system. *Blood*, 121 (11):1951–1960, 2013. ISSN 0006-4971. doi: 10.1182/blood-2012-09-435057.

S. Boulton, M. Akimoto, R. Selvaratnam, A. Bashiri, and G. Melacini. A tool set to map allosteric networks through the NMR chemical shift covariance analysis. *Scientific Reports*, 4:7306, 2015. ISSN 2045-2322. doi: 10.1038/srep07306.

J. L. Cannon, E. Papafragkou, G. W. Park, J. Osborne, L. A. Jaykus, and J. Vinje. Surrogates for the study of norovirus stability and inactivation in the environment: A comparison of murine norovirus and feline calicivirus. *J Food Prot*, 69(11):2761–5, 2006. ISSN 0362-028X. doi: 10.4315/0362-028x-69.11.2761.

J. Cavanagh, W. Fairbrother, A. Palmer, and N. Skelton. *Protein NMR Spectroscopy: Principles and Practice*, pages 532–553. Elsevier Science, 1995. ISBN 9780080515298.

K.-O. Chang, S. V. Sosnovtsev, G. Belliot, Y. Kim, L. J. Saif, and K. Y. Green. Bile acids are essential for porcine enteric calicivirus replication in association with down-regulation of signal transducer and activator of transcription 1. *Proceedings of the National Academy of Sciences*, 101(23): 8733–8738, 2004. ISSN 0027-8424. doi: 10.1073/pnas.0401126101.

C. Charlier, S. F. Cousin, and F. Ferrage. Protein dynamics from nuclear magnetic relaxation. *Chemical Society Reviews*, 45(9):2410–2422, 2016. ISSN 0306-0012. doi: 10.1039/C5CS00832H.

G. Chevreux, C. Atmanene, P. Lopez, J. Ouazzani, A. Van Dorsselaer, B. Badet, M.-A. Badet-Denisot, and S. Sanglier-Cianférani. Monitoring the dynamics of monomer exchange using electrospray mass spectrometry: The case of the dimeric glucosamine-6-phosphate synthase. *Journal of The American Society for Mass Spectrometry*, 22(3):431–439, 2011. ISSN 1879-1123. doi: 10.1007/s13361-010-0054-z.

P. Chhabra, M. de Graaf, G. I. Parra, M. C. Chan, K. Green, V. Martella, Q. Wang, P. A. White, K. Katayama, H. Vennema, M. P. G. Koopmans, and J. Vinje. Updated classification of norovirus genogroups and genotypes. *J Gen Virol*, 100(10):1393–1406, 2019. ISSN 0022-1317. doi: 10.1099/jgv.0.001318.

J. Y. Chiang. Bile acid metabolism and signaling. *Compr Physiol*, 3(3):1191–212, 2013. ISSN 2040-4603. doi: 10.1002/cphy.c120023.

E. Comas-Casellas, A. Martinez-Barriocanal, F. Miro, A. Ejarque-Ortiz, J. Schwartz, S., M. Martin, and J. Sayos. Cloning and characterization of CD300d, a novel member of the human CD300 family of immune receptors. *J Biol Chem*, 287(13):9682–93, 2012. ISSN 0021-9258. doi: 10.1074/jbc.M111.279224.

X. Cong, X.-m. Sun, J.-x. Qi, H.-b. Li, W.-g. Chai, Q. Zhang, H. Wang, X.-y. Kong, J. Song, L.-l. Pang, M. Jin, D.-d. Li, M. Tan, and Z.-j. Duan. GII.13/21 noroviruses recognize glycans with a terminal β-galactose via an unconventional glycan binding site. *Journal of Virology*, 93(15):e00723–19, 2019. 1098-5514 (Electronic) 0022-538X (Linking). doi: 10.1128/jvi.00723-19.

M. J. Conley, M. McElwee, L. Azmi, M. Gabrielsen, O. Byron, I. G. Goodfellow, and D. Bhella. Calicivirus VP2 forms a portal-like assembly following receptor engagement. *Nature*, 565(7739):377–381, 2019. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). doi: 10.1038/s41586-018-0852-1.

N. W. Cortes-Penfield, S. Ramani, M. K. Estes, and R. L. Atmar. Prospects and challenges in the development of a norovirus vaccine. *Clinical Therapeutics*, 39(8):1537–1549, 2017. ISSN 0149-2918. doi: 10.1016/j.clinthera.2017.07.002.

A. Corti and F. Curnis. Isoaspartate-dependent molecular switches for integrin-ligand recognition. *J Cell Sci*, 124(Pt 4):515–22, 2011. ISSN 1477-9137 (Electronic) 0021-9533 (Linking). doi: 10.1242/jcs.077172.

R. Creutznacher. *NMR Studies into the Binding of Human Norovirus P-Dimers to L-Fucose based Precision Oligomers*. master's thesis, University of Luebeck, 2016.

R. Creutznacher, E. Schulze, G. Wallmann, T. Peters, M. Stein, and A. Mallagaray. Chemical-shift perturbations reflect bile acid binding to norovirus coat protein: Recognition comes in different flavors. *Chembiochem*, 2019. ISSN 1439-7633 (Electronic) 1439-4227 (Linking). doi: 10.1002/cbic.201900572.

A. de Rougemont, N. Ruvoen-Clouet, B. Simon, M. Estienney, C. Elie-Caille, S. Aho, P. Pothier, J. Le Pendu, W. Boireau, and G. Belliot. Qualitative and quantitative analysis of the binding of GII.4 norovirus variants onto human blood group antigens. *Journal of Virology*, 85(9): 4057–4070, 2011. ISSN 1098-5514 (Electronic) 0022-538X (Linking). doi: 10.1128/JVI.02077-10.

A. G. Diamos and H. S. Mason. High-level expression and enrichment of norovirus virus-like particles in plants using modified geminiviral vectors. *Protein Expr Purif*, 151:86–92, 2018. ISSN 1046-5928. doi: 10.1016/j.pep.2018.06.011.

A. Eletsky, A. Kienhofer, and K. Pervushin. TROSY NMR with partially deuterated proteins. *J Biomol NMR*, 20(2):177–80, 2001. ISSN 0925-2738. doi: 10.1023/a:1011265430149.

M. K. Estes, K. Ettayebi, V. R. Tenge, K. Murakami, U. Karandikar, S. C. Lin, B. V. Ayyar, N. W. Cortes-Penfield, K. Haga, F. H. Neill, A. R. Opekun, J. R. Broughman, X. L. Zeng, S. E. Blutt, S. E. Crawford, S. Ramani, D. Y. Graham, and R. L. Atmar. Human norovirus cultivation in nontransformed stem cell-derived human intestinal enteroid cultures: Success and challenges. *Viruses*, 11(7):638, 2019. ISSN 1999-4915. doi: 10.3390/v11070638.

K. Ettayebi, S. E. Crawford, K. Murakami, J. R. Broughman, U. Karandikar, V. R. Tenge, F. H. Neill, S. E. Blutt, X. L. Zeng, L. Qu, B. Kou, A. R. Opekun, D. Burrin, D. Y. Graham, S. Ramani, R. L. Atmar, and M. K. Estes. Replication of human noroviruses in stem cell-derived human enteroids. *Science*, 353 (6306):1387–1393, 2016. ISSN 1095-9203 (Electronic) 0036-8075 (Linking). doi: 10.1126/science.aaf5211.

J. C. Facelli. Chemical shift tensors: Theory and application to molecular structural problems. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 58(3-4): 176–201, 2011. ISSN 1873-3301 (Electronic) 0079-6565 (Linking). doi: 10.1016/j.pnmrs.2010.10.003.

C. Feldmann. *NMR-spektroskopische Charakterisierung einer möglichen Glykan-Bindungstasche des murinen Noroviruskapsidproteins*. bachelor's thesis, University of Luebeck, 2019.

C. Fernández and G. Wider. TROSY in NMR studies of the structure and function of large biological macromolecules. *Current opinion in structural biology*, 13(5): 570–80, 2003. ISSN 0959-440X. doi: 10.1016/j.sbi.2003.09.009.

B. Fiege, C. Rademacher, J. Cartmell, P. I. Kitov, F. Parra, and T. Peters. Molecular details of the recognition of blood group antigens by a human norovirus as determined by STD NMR spectroscopy. *Angew Chem Int Ed Engl*, 51(4):928–32, 2012. ISSN 1521-3773 (Electronic) 1433-7851 (Linking). doi: 10.1002/anie.201105719.

S. J. Flint, V. R. Racaniello, L. W. Enquist, and A. M. Skalka. *Principles of virology: molecular biology, pathogenesis, and control of animal viruses*. ASM Press, Washington, 2nd edition, pages 91–150, 2003. ISBN 1555812597.

R. Frenck, D. I. Bernstein, M. Xia, P. Huang, W. Zhong, S. Parker, M. Dickey, M. McNeal, and X. Jiang. Predicting susceptibility to norovirus GII.4 by use of a challenge model involving humans. *J Infect Dis*, 206(9):1386–93, 2012. ISSN 0022-1899. doi: 10.1093/infdis/jis514.

K. H. Gardner and L. E. Kay. The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annual Review of Biophysics and Biomolecular Structure*, 27:357–406, 1998. ISSN 1056-8700. doi: 10.1146/annurev.biophys.27.1.357.

E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch. *Protein Identification and Analysis Tools on the ExPASy Server*, pages 571–607. Humana Press, Totowa, NJ, 2005. ISBN 978-1-59259-890-8. doi: 10.1385/1-59259-890-0:571.

A. D. Gossert and W. Jahnke. NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 97:82–125, 2016. ISSN 1873-3301 (Electronic) 0079-6565 (Linking). doi: 10.1016/j.pnmrs.2016.09.001.

N. K. Goto and L. E. Kay. New developments in isotope labelling strategies for protein solution NMR spectroscopy. *Curr. Opin. Str. Biol.*, 10(5): 585–592, 2000. ISSN 0959-440X doi: 10.1016/s0959-440x(00)00135-4.

D. Goulet. Modeling, simulating, and parameter fitting of biochemical kinetic experiments. *SIAM Review*, 58(2):331–353, 2016. ISSN 0036-1445. doi: 10.1137/151004707.

L. Grassi, C. Regl, S. Wildner, G. Gadermaier, C. G. Huber, C. Cabrele, and M. Schubert. Complete NMR assignment of succinimide and its detection and quantification in peptides and intact proteins. *Analytical Chemistry*, 89(22): 11962–11970, 2017. 1520-6882 (Electronic) 0003-2700 (Linking). doi: 10.1021/acs.analchem.7b01645.

K. R. Grau, S. Zhu, S. T. Peterson, E. W. Helm, D. Philip, M. Phillips, A. Hernandez, H. Turula, P. Frasse, V. R. Graziano, C. B. Wilen, C. E. Wobus, M. T. Baldridge, and S. M. Karst. The intestinal regionalization of acute norovirus infection is regulated by the microbiota via bile acid-mediated priming of type III interferon. *Nat Microbiol*, 5(1):84–92, 2020. ISSN 2058-5276. doi: 10.1038/s41564-019-0602-7.

V. R. Graziano, J. Wei, and C. B. Wilen. Norovirus attachment and entry. *Viruses*, 11(6):495, 2019. ISSN 1999-4915. doi: 10.3390/v11060495.

M. Gueron. Nuclear relaxation in macromolecules by paramagnetic ions: a novel mechanism. *Journal of Magnetic Resonance*, 19(1):58–66, 1975. ISSN 0022-2364. doi: 10.1016/0022-2364(75)90029-3.

K. Haga, A. Fujimoto, R. Takai-Todaka, M. Miki, Y. H. Doan, K. Murakami, M. Yokoyama, K. Murata, A. Nakanishi, and K. Katayama. Functional receptor molecules CD300lf and CD300ld within the CD300 family enable murine noroviruses to infect cells. *Proceedings of the National Academy of Sciences*, 113(41):E6248-E6255, 2016. ISSN 1091-6490 (Electronic) 0027-8424 (Linking). doi: 10.1073/pnas.1605575113.

R. M. Hagan, R. Bjornsson, S. A. McMahon, B. Schomburg, V. Braithwaite, M. Buhl, J. H. Naismith, and U. Schwarz-Linek. NMR spectroscopic and theoretical analysis of a spontaneously formed Lys-Asp isopeptide bond. *Angew Chem Int Ed Engl*, 49(45):8421–5, 2010. ISSN 1521-3773 (Electronic) 1433-7851 (Linking). doi: 10.1002/anie.201004340.

L. Han, P. I. Kitov, E. N. Kitova, M. Tan, L. Wang, M. Xia, X. Jiang, and J. S. Klassen. Affinities of recombinant norovirus P dimers for human blood group antigens. *Glycobiology*, 23(3): 276–285, 2013. ISSN 1460-2423 (Electronic) 0959-6658 (Linking). doi: 10.1093/glycob/cws141.

L. Han, M. Tan, M. Xia, E. N. Kitova, X. Jiang, and J. S. Klassen. Gangliosides are ligands for human noroviruses. *Journal of the American Chemical Society*, 136(36): 12631–12637, 2014. ISSN 1520-5126 (Electronic) 0002-7863 (Linking). doi: 10.1021/ja505272n.

L. Han, E. N. Kitova, M. Tan, X. Jiang, B. Pluvinage, A. B. Boraston, and J. S. Klassen. Affinities of human histo-blood group antigens for norovirus capsid protein complexes. *Glycobiology*, 25:(2) 170–180, 2015. ISSN 1460-2423 (Electronic) 0959-6658 (Linking). doi: 10.1093/glycob/cwu100.

G. S. Hansman, S. Shahzad-ul Hussan, J. S. McLellan, G.-Y. Chuang, I. Georgiev, T. Shimoike, K. Katayama, C. A. Bewley, and P. D. Kwong. Structural basis for norovirus inhibition and fucose mimicry by citrate. *Journal of Virology*, 86(1):284, 2012. ISSN 1098-5514 (Electronic) 0022-538X (Linking) doi: 10.1128/JVI.05909-11.

P. Hao, S. S. Adav, X. Gallart-Palau, and S. K. Sze. Recent advances in mass spectrometric analysis of protein deamidation. *Mass spectrometry reviews*, 36:(6) 677–692, 2017. ISSN 1098-2787 (Electronic) 0277-7037 (Linking). doi: 10.1002/mas.21491.

R. J. Harris, B. Kabakoff, F. D. Macchi, F. J. Shen, M. Kwong, J. D. Andya, S. J. Shire, N. Bjork, K. Totpal, and A. B. Chen. Identification of multiple sources of charge heterogeneity in a recombinant antibody. *Journal of Chromatography B: Biomedical Sciences and Applications*, 752(2): 233–245, 2001. ISSN 1387-2273. doi: 10.1016/S0378-4347(00)00548-X.

W. M. Haynes. *CRC handbook of chemistry and physics*. CRC Press, 92nd edition, pages 11-4 – 11-5, 2011.

J. E. Heggelund, A. Varrot, A. Imberty, and U. Krengel. Histo-blood group antigens as mediators of infections. *Current Opinion in Structural Biology*, 44:190–200, 2017. ISSN 1879-033X (Electronic) doi: 10.1016/j.sbi.2017.04.001.

A. F. Hofmann and K. J. Mysels. Bile acid solubility and precipitation in vitro and in vivo: the role of conjugation, pH, and Ca2+ ions. *J Lipid Res*, 33(5):617–26, 1992. ISSN 0022-2275.

F. M. Hofmann, E. Olawumi, M. Michaelis, F. Hofmann, and U. Stossel. Challenges in infection epidemiology: On the underreporting of norovirus gastroenteritis cases in Germany. *Int J Environ Res Public Health*, 17(1):314, 2020. ISSN 1660-4601. doi: 10.3390/ijerph17010314.

Z. E. Huma, J. P. Ludeman, B. L. Wilkinson, R. J. Payne, and M. J. Stone. NMR characterization of cooperativity: fast ligand binding coupled to slow protein dimerization. *Chem. Sci.*, 5(7):2783–2788, 2014. ISSN 2041-6520. doi: 10.1039/c4sc00131a.

A. M. Hutson, R. L. Atmar, D. Y. Graham, and M. K. Estes. Norwalk virus infection and disease is associated with ABO histo-blood group type. *J Infect Dis*, 185(9):1335–7, 2002. ISSN 0022-1899. doi: 10.1086/339883.

S. G. Hyberts, H. Arthanari, S. A. Robson, and G. Wagner. Perspectives in magnetic resonance: NMR in the post-FFT era. *Journal of Magnetic Resonance*, 241:60–73, 2014. ISSN 1090-7807. doi: 10.1016/j.jmr.2013.11.014.

O. B. Ijare, B. S. Somashekar, Y. Jadegoud, and G. A. Nagana Gowda. 1H and 13C NMR characterization and stereochemical assignments of bile acids in aqueous media. *Lipids*, 40(10):1031–1041, 2005. ISSN 0024-4201. doi: 10.1007/s11745-005-1466-1.

V. A. Jarymowycz and M. J. Stone. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem Rev*, 106(5):1624–71, 2006. ISSN 0009-2665. doi: 10.1021/cr040421p.

V. Jayalakshmi and N. Krishna. Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand–receptor complexes. *Journal of Magnetic Resonance*, 155(1): 106–118, 2002. ISSN 1090-7807. doi: 10.1006/jmre.2001.2499.

H. Jeong, V. Barbe, C. H. Lee, D. Vallenet, D. S. Yu, S. H. Choi, A. Couloux, S. W. Lee, S. H. Yoon, L. Cattolico, C. G. Hur, H. S. Park, B. Segurens, S. C. Kim, T. K. Oh, R. E. Lenski, F. W. Studier, P. Daegelen, and J. F. Kim. Genome sequences of Escherichia coli B strains REL606 and BL21(DE3). *J Mol Biol*, 394(4):644–52, 2009. ISSN 0022-2836. doi: 10.1016/j.jmb.2009.09.052.

L. Jia and Y. Sun. Protein asparagine deamidation prediction based on structures with machine learning methods. *PLOS ONE*, 12:e0181347, 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0181347.

X. Jiang, M. Wang, D. Y. Graham, and M. K. Estes. Expression, self-assembly, and antigenicity of the norwalk virus capsid protein. *Journal of virology*, 66(11):6527–6532, 1992. ISSN 0022-538X.

M. John and G. Otting. Strategies for measurements of pseudocontact shifts in protein NMR spectroscopy. *ChemPhysChem*, 8(16):2309–2313, 2007. ISSN 1439-4235. doi: 10.1002/cphc.200700510.

M. K. Jones, M. Watanabe, S. Zhu, C. L. Graves, L. R. Keyes, K. R. Grau, M. B. Gonzalez-Hernandez, N. M. Iovine, C. E. Wobus, J. Vinje, S. A. Tibbetts, S. M. Wallet, and S. M. Karst. Enteric bacteria promote human and mouse norovirus infection of B cells. *Science*, 346(6210): 755–759, 2014. ISSN 1095-9203 (Electronic) 0036-8075 (Linking). doi: 10.1126/science.1257147.

M. K. Jones, K. R. Grau, V. Costantini, A. O. Kolawole, M. de Graaf, P. Freiden, C. L. Graves, M. Koopmans, S. M. Wallet, S. A. Tibbetts, S. Schultz-Cherry, C. E. Wobus, J. Vinjé, and S. M. Karst. Human norovirus culture in B cells. *Nature Protocols*, 10(12): 1939–1947, 2015. ISSN 1750-2799. doi: 10.1038/nprot.2015.121.

J. Jung, T. Grant, D. R. Thomas, C. W. Diehnelt, N. Grigorieff, and L. Joshua-Tor. High-resolution cryo-EM structures of outbreak strain human norovirus shells reveal size variations. *Proceedings of the National Academy of Sciences*, 116(26):12828–12832, 2019. ISSN 1091-6490 (Electronic) 0027-8424 (Linking). doi: 10.1073/pnas.1903562116.

S. M. Karst and C. E. Wobus. A working model of how noroviruses infect the intestine. *PLOS Pathogens*, 11(2): e1004626, 2015. ISSN 1553-7374 (Electronic) 1553-7366 (Linking). doi: 10.1371/journal.ppat.1004626.

S. M. Karst, C. E. Wobus, M. Lay, J. Davidson, and H. W. Virgin. STAT1-dependent innate immunity to a Norwalk-like virus. *Science*, 299(5612):1575–8, 2003. ISSN 1095-9203 (Electronic) 0036-8075 (Linking). doi: 10.1126/science.1077905.

J. Keeler. *Understanding NMR Spectroscopy*. WILEY-VCH Verlag, 2nd edition, pages 23–45, 241–313, 2010.

T. Åkerud, E. Thulin, R. L. Van Etten, and M. Akke. Intramolecular dynamics of low molecular weight protein tyrosine phosphatase in monomer–dimer equilibrium studied by NMR: A model for changes in dynamics upon target binding. *Journal of Molecular Biology*, 322(1):137–152, 2002. ISSN 0022-2836. doi: 10.1016/S0022-2836(02)00714-3.

T. Kilic, A. Koromyslova, V. Malak, and G. S. Hansman. Atomic structure of the murine norovirus protruding domain and soluble CD300lf receptor complex. *J Virol*, 92(11):e00413-18, 2018. ISSN 1098-5514 (Electronic) 0022-538X (Linking). doi: 10.1128/JVI.00413-18.

T. Kilic, A. Koromyslova, and G. S. Hansman. Structural basis for human norovirus capsid binding to bile acids. *J Virol*, 93(2):e01581-18, 2019. ISSN 1098-5514 (Electronic) 0022-538X (Linking). doi: 10.1128/JVI.01581-18.

D.-S. Kim, M. Hosmillo, M. M. Alfajaro, J.-Y. Kim, J.-G. Park, K.-Y. Son, E.-H. Ryu, F. Sorgeloos, H.-J. Kwon, S.-J. Park, W. S. Lee, D. Cho, J. Kwon, J.-S. Choi, M.-I. Kang, I. Goodfellow, and K.-O. Cho. Both α2,3- and α2,6-linked sialic acids on O-linked glycoproteins act as functional receptors for porcine sapovirus. *PLOS Pathogens*, 10(6):e1004172, 2014. ISSN 1553-7374 (Electronic) 1553-7366 (Linking). doi: 10.1371/journal.ppat.1004172.

A. O. Kolawole, H. Q. Smith, S. A. Svoboda, M. S. Lewis, M. B. Sherman, G. C. Lynch, B. M. Pettitt, T. J. Smith, and C. E. Wobus. Norovirus escape from broadly neutralizing antibodies is limited to allostery-like mechanisms. *mSphere*, 2(5):e00334–17, 2017. ISSN 2379-5042. doi: 10.1128/mSphere.00334-17.

A. Koromyslova, S. Tripathi, V. Morozov, H. Schroten, and G. S. Hansman. Human norovirus inhibition by a human milk oligosaccharide. *Virology*, 508:81–89, 2017. ISSN 1096-0341 (Electronic) 0042-6822 (Linking). doi: 10.1016/j.virol.2017.04.032.

A. Krezel and W. Bal. A formula for correlating pKa values determined in D2O and H2O. *J Inorg Biochem*, 98(1):161–6, 2004. ISSN 0162-0134. doi: 10.1016/j.jinorgbio.2003.10.001.

L. B. Krivdin. Calculation of 15N NMR chemical shifts: Recent advances and perspectives. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 102-103:98–119, 2017. ISSN 1873-3301 (Electronic) 0079-6565 (Linking). doi: 10.1016/j.pnmrs.2017.08.001.

K. Laidler. *Chemical Kinetics*. Harper and Row, 3rd edition, pages 280-281, 1987. ISBN 0-06-043862-2.

N. A. Lakomek, J. Ying, and A. Bax. Measurement of 15N relaxation rates in perdeuterated proteins by TROSY-based methods. *J Biomol NMR*, 53(3):209–21, 2012. ISSN 1573-5001 (Electronic) 0925-2738 (Linking). doi: 10.1007/s10858-012-9626-5.

M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. Clustal W and clustal X version 2.0. *Bioinformatics*, 23(21):2947–2948, 2007. ISSN 1367-4803. doi: 10.1093/bioinformatics/btm404.

D. Lee, C. Hilty, G. Wider, and K. Wüthrich. Effective rotational correlation times of proteins from NMR relaxation interference. *Journal of Magnetic Resonance*, 178(1): 72–76, 2006. ISSN 1090-7807. doi: 10.1016/j.jmr.2005.08.014.

F. Lehmann, E. Tiralongo, and J. Tiralongo. Sialic acid-specific lectins: occurrence, specificity and function. *Cell Mol Life Sci*, 63(12):1331–1354, 2006. ISSN 1420-9071. doi: 10.1007/s00018-005-5589-y.

D. Li, A. Breiman, J. le Pendu, and M. Uyttendaele. Binding to histo-blood group antigen-expressing bacteria protects human norovirus from acute heat stress. *Frontiers in Microbiology*, 6:659, 2015. ISSN 1664-302X. doi: 10.3389/fmicb.2015.00659.

L. Lindesmith, C. Moe, S. Marionneau, N. Ruvoen, X. Jiang, L. Lindblad, P. Stewart, J. LePendu, and R. Baric. Human susceptibility and resistance to Norwalk virus infection. *Nat Med*, 9(5): 548–553, 2003. ISSN 1078-8956. doi: 10.1038/nm860.

L. Lindesmith, C. Moe, J. LePendu, J. A. Frelinger, J. Treanor, and R. S. Baric. Cellular and humoral immunity following snow mountain virus challenge. *Journal of Virology*, 79(5):2900–9, 2005. ISSN 0022-538X. doi: 10.1128/JVI.79.5.2900-2909.2005.

L. C. Lindesmith, E. F. Donaldson, M. Beltramello, S. Pintus, D. Corti, J. Swanstrom, K. Debbink, T. A. Jones, A. Lanzavecchia, and R. S. Baric. Particle conformation regulates antibody access to a conserved GII.4 norovirus blockade epitope. *Journal of Virology*, 88(16):8826–42, 2014. ISSN 1098-5514 (Electronic) 0022-538X (Linking). doi: 10.1128/JVI.01192-14.

L. C. Lindesmith, P. D. Brewer-Jensen, M. L. Mallory, K. Debbink, E. W. Swann, J. Vinjé, and R. S. Baric. Antigenic characterization of a novel recombinant GII.p16-GII.4 Sydney norovirus strain with minor sequence variation leading to antibody escape. *The Journal of Infectious Diseases*, 217 (7):1145–1152, 2017. ISSN 0022-1899. doi: 10.1093/infdis/jix651.

L. C. Lindesmith, M. L. Mallory, K. Debbink, E. F. Donaldson, P. D. Brewer-Jensen, E. W. Swann, T. P. Sheahan, R. L. Graham, M. Beltramello, D. Corti, A. Lanzavecchia, and R. S. Baric. Conformational occlusion of blockade antibody epitopes, a novel mechanism of GII.4 human norovirus immune evasion. *mSphere*, 3(1):e00518-17, 2018. ISSN 2379-5042. doi: 10.1128/mSphere.00518-17.

W. Liu, Y. Chen, X. Jiang, M. Xia, Y. Yang, M. Tan, X. Li, and Z. Rao. A unique human norovirus lineage with a distinct HBGA binding interface. *PLoS Pathog*, 11(7):e1005025, 2015. ISSN 1553-7374 (Electronic) 1553-7366 (Linking). doi: 10.1371/journal.ppat.1005025.

W.-M. Liu, M. Overhand, and M. Ubbink. The application of paramagnetic lanthanoid ions in NMR spectroscopy on proteins. *Coordination Chemistry Reviews*, 273-274:2–12, 2014. ISSN 0010-8545. doi: 10.1016/j.ccr.2013.10.018.

J. R. Lorenzo, L. G. Alonso, and I. E. Sánchez. Prediction of spontaneous protein deamidation from sequence-derived secondary structure and intrinsic disorder. *PLOS ONE*, 10(12): e0145186, 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0145186.

P. Luginbühl and K. Wüthrich. Semi-classical nuclear spin relaxation theory revisited for use with biological macromolecules. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 40(3):199–247, 2002. ISSN 0079-6565. doi: 10.1016/S0079-6565(01)00043-7.

T. Maaß. *Stable isotope labeling schemes for the assignment of NMR spectra of murine Norovirus P-domain proteins*. master's thesis, University of Luebeck, 2019.

A. Mallagaray, J. Lockhauserbäumer, G. Hansman, C. Uetrecht, and T. Peters. Attachment of norovirus to histo blood group antigens: A cooperative multistep process. *Angewandte Chemie International Edition*, 54(41): 12014–12019, 2015. ISSN 1521-3773 (Electronic) 1433-7851 (Linking). doi: 10.1002/anie.201505672.

A. Mallagaray, C. Rademacher, F. Parra, G. Hansman, and T. Peters. Saturation transfer difference nuclear magnetic resonance titrations reveal complex multistep-binding of L-fucose to norovirus particles. *Glycobiology*, 27(1):80–86, 2017. ISSN 1460-2423 (Electronic) 0959-6658 (Linking). doi: 10.1093/glycob/cww070.

A. Mallagaray, R. Creutznacher, J. Dulfer, P. H. O. Mayer, L. L. Grimm, J. M. Orduna, E. Trabjerg, T. Stehle, K. D. Rand, B. S. Blaum, C. Uetrecht, and T. Peters. A post-translational modification of human norovirus capsid protein attenuates glycan binding. *Nat Commun*, 10(1):1320, 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-09251-5.

M. L. Mallory, L. C. Lindesmith, R. L. Graham, and R. S. Baric. GII.4 human norovirus: Surveying the antigenic landscape. *Viruses*, 11(2):177, 2019. ISSN 1999-4915. doi: 10.3390/v11020177.

C. J. Markin and L. Spyracopoulos. Increased precision for analysis of protein–ligand dissociation constants determined from chemical shift titrations. *Journal of Biomolecular NMR*, 53(2):125–138, 2012. ISSN 1573-5001. doi: 10.1007/s10858-012-9630-9.

C. P. Mattison, C. V. Cardemil, and A. J. Hall. Progress on norovirus vaccine research: public health considerations and future directions. *Expert Rev Vaccines*, 17(9):773–784, 2018. ISSN 1476-0584. doi: 10.1080/14760584.2018.1510327.

M. Mayer and B. Meyer. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie International Edition*, 38:1784–1788, 1999. ISSN 1433-7851. doi: 10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q.

M. Mayer and B. Meyer. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *Journal of the American Chemical Society*, 123(25): 6108–6117, 2001. ISSN 0002-7863 (Print). doi: 10.1021/ja0100120.

E. L. McConnell, A. W. Basit, and S. Murdan. Measurements of rat and mouse gastrointestinal pH, fluid and lymphoid tissue, and implications for in-vivo experiments. *Journal of Pharmacy and Pharmacology*, 60(1):63–70, 2008. ISSN 0022-3573. doi: 10.1211/jpp.60.1.0008.

G. M. McLeod and H. S. Wiggins. Bile-salts in small intestinal contents after ileal resection and in other malabsorption syndromes. *The Lancet*, 291(7548):873–876, 1968. ISSN 0140-6736. doi: 10.1016/S0140-6736(68)90235-3.

B. Meyer and T. Peters. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angewandte Chemie International Edition*, 42(8): 864–890, 2003. ISSN 1433-7851. doi: 10.1002/anie.200390233.

M. M. Müller. Post-translational modifications of protein backbones: Unique functions, mechanisms, and challenges. *Biochemistry*, 57(2): 177–185, 2018. ISSN 1520-4995 (Electronic) 0006-2960 (Linking). doi: 10.1021/acs.biochem.7b00861.

F. E. Morreale, A. Testa, V. K. Chaugule, A. Bortoluzzi, A. Ciulli, and H. Walden. Mind the metal: A fragment library-derived zinc impurity binds the E2 ubiquitin-conjugating enzyme Ube2T and induces structural rearrangements. *Journal of Medicinal Chemistry*, 60(19):8183–8191, 2017. ISSN 0022-2623. doi: 10.1021/acs.jmedchem.7b01071.

K. Murakami, V. R. Tenge, U. C. Karandikar, S. C. Lin, S. Ramani, K. Ettayebi, S. E. Crawford, X. L. Zeng, F. H. Neill, B. V. Ayyar, K. Katayama, D. Y. Graham, E. Bieberich, R. L. Atmar, and M. K. Estes. Bile acids and ceramide overcome the entry restriction for GII.3 human norovirus replication in human intestinal enteroids. *Proc Natl Acad Sci U S A*, 117(3):1700–1710, 2020. ISSN 1091-6490 (Electronic) 0027-8424 (Linking). doi: 10.1073/pnas.1910138117.

W. Nasir, M. Frank, A. Kunze, M. Bally, F. Parra, P. G. Nyholm, F. Hook, and G. Larson. Histo-blood group antigen presentation is critical for binding of norovirus VLP to glycosphingolipids in model membranes. *ACS Chem Biol*, 12(5):1288–1296, 2017. ISSN 1554-8937 (Electronic) 1554-8929 (Linking). doi: 10.1021/acschembio.7b00152.

C. A. Nelson, C. B. Wilen, Y. N. Dai, R. C. Orchard, A. S. Kim, R. A. Stegeman, L. L. Hsieh, T. J. Smith, H. W. Virgin, and D. H. Fremont. Structural basis for murine norovirus engagement of bile acids and the CD300lf receptor. *Proc Natl Acad Sci U S A*, 115(39):E9201–E9210, 2018. ISSN 1091-6490 (Electronic) 0027-8424 (Linking). doi: 10.1073/pnas.1805797115.

D. Neuhaus and M. P. Williamson. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*. Wiley VCH, 2nd edition, pages 3–33, 2000. ISBN 978-0-471-24675-6.

J. Nordgren and L. Svensson. Genetic susceptibility to human norovirus infection: An update. *Viruses*, 11(3):226, 2019. ISSN 1999-4915. doi: 10.3390/v11030226.

J. Nordgren, E. Kindberg, P. E. Lindgren, A. Matussek, and L. Svensson. Norovirus gastroenteritis outbreak with a secretor-independent susceptibility pattern, Sweden. *Emerg Infect Dis*, 16(1):81–7, 2010. ISSN 1080-6040. doi: 10.3201/eid1601.090633.

C. Nowak, A. Tiwari, and H. Liu. Asparagine deamidation in a complementarity determining region of a recombinant monoclonal antibody in complex with antigen. *Anal Chem*, 90(11):6998–7003, 2018. ISSN 1520-6882 (Electronic) 0003-2700 (Linking). doi: 10.1021/acs.analchem.8b01322.

P. Ogrissek. *Einfluss von Ligandenbindung und posttranslationaler Modifikation auf die Dynamik des Norovirus Kapsidproteins*. bachelor's thesis, University of Luebeck, 2018.

T. E. Oliphant. Python for scientific computing. *Computing in Science and Engineering*, 9(3):10–20, 2007. ISSN 1558-366X. doi: 10.1109/MCSE.2007.58.

R. C. Orchard, C. B. Wilen, J. G. Doench, M. T. Baldridge, B. T. McCune, Y.-C. J. Lee, S. Lee, S. M. Pruett-Miller, C. A. Nelson, D. H. Fremont, and H. W. Virgin. Discovery of a proteinaceous cellular receptor for a norovirus. *Science*, 353(6302): 933–936, 2016. ISSN 1095-9203 (Electronic) 0036-8075 (Linking). doi: 10.1126/science.aaf1220.

G. Otting. Prospects for lanthanides in structural biology by NMR. *Journal of Biomolecular NMR*, 42 (1):1–9, 2008. ISSN 1573-5001. doi: 10.1007/s10858-008-9256-0.

G. Otting. Protein NMR using paramagnetic ions. *Annual Review of Biophysics*, 39(1):387–405, 2010. ISSN 1936-122X. doi: 10.1146/annurev.biophys.093008.131321.

N. Parveen, G. E. Rydell, G. Larson, V. P. Hytonen, V. P. Zhdanov, F. Hook, and S. Block. Competition for membrane receptors: Norovirus detachment via lectin attachment. *J Am Chem Soc*, 141(41): 16303–16311, 2019. ISSN 1520-5126 (Electronic) 0002-7863 (Linking). doi: 10.1021/jacs.9b06036.

C. M. Payne, C. Bernstein, K. Dvorak, and H. Bernstein. Hydrophobic bile acids, genomic instability, darwinian selection, and colon carcinogenesis. *Clin Exp Gastroenterol*, 1:19–47, 2008. ISSN 1178-7023 (Print) 1178-7023.

K. Pervushin, R. Riek, G. Wider, and K. Wuthrich. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences*, 94:12366–12371, 1997. ISSN 0027-8424. doi: 10.1073/pnas.94.23.12366.

J. J. Phillips, A. Buchanan, J. Andrews, M. Chodorge, S. Sridharan, L. Mitchell, N. Burmeister, A. D. Kippen, T. J. Vaughan, D. R. Higazi, and D. Lowe. Rate of asparagine deamidation in a monoclonal antibody correlating with hydrogen exchange rate at adjacent downstream residues. *Analytical Chemistry*, 89(4): 2361–2368, 2017. ISSN 1520-6882 (Electronic) 0003-2700 (Linking). doi: 10.1021/acs.analchem.6b04158.

G. Platzer, M. Okon, and L. P. McIntosh. pH-dependent random coil 1H, 13C, and 15N chemical shifts of the ionizable amino acids: a guide for protein pKa measurements. *Journal of Biomolecular NMR*, 60(2-3): 109–129, 2014. ISSN 1573-5001 (Electronic) 0925-2738 (Linking). doi: 10.1007/s10858-014-9862-y.

N. V. Plotnikov, S. K. Singh, J. C. Rouse, and S. Kumar. Quantifying the risks of asparagine deamidation and aspartate isomerization in biopharmaceuticals by computing reaction free-energy surfaces. *The Journal of Physical Chemistry B*, 121(4): 719–730, 2017. ISSN 1520-6106. doi: 10.1021/acs.jpcb.6b11614.

R. Pogan, J. Dulfer, and C. Uetrecht. Norovirus assembly and stability. *Curr Opin Virol*, 31:59–65, 2018. ISSN 1879-6265 (Electronic) 1879-6257 (Linking). doi: 10.1016/j.coviro.2018.05.003.

B. V. V. Prasad, M. E. Hardy, T. Dokland, J. Bella, M. G. Rossmann, and M. K. Estes. X-ray crystallographic structure of the norwalk virus capsid. *Science*, 286(5438): 287–290, 1999. ISSN 0036-8075. doi: 10.1126/science.286.5438.287.

Y. Qian, M. Song, X. Jiang, M. Xia, J. Meller, M. Tan, Y. Chen, X. Li, and Z. Rao. Structural adaptations of norovirus GII.17/13/21 lineage through two distinct evolutionary paths. *Journal of Virology*, 93(1): e01655–18, 2019. ISSN 1098-5514 (Electronic) 0022-538X (Linking). doi: 10.1128/JVI.01655-18.

J. L. Radkiewicz, H. Zipse, S. Clarke, and K. N. Houk. Accelerated racemization of aspartic acid and asparagine residues via succinimide intermediates: An ab initio theoretical exploration of mechanism. *Journal of the American Chemical Society*, 118(38): 9148–9155, 1996. ISSN 0002-7863. doi: 10.1021/ja953505b.

T. Rehm, R. Huber, and T. A. Holak. Application of NMR in structural proteomics. *Structure*, 10(12): 1613–1618, 2002. ISSN 09692126. doi: 10.1016/s0969-2126(02)00894-8.

N. Rezaei-Ghaleh, K. Giller, S. Becker, and M. Zweckstetter. Effect of zinc binding on β-amyloid structure and dynamics: implications for Aβ aggregation. *Biophysical journal*, 101(5):1202–1211, 2011. ISSN 1542-0086 (Electronic) 0006-3495 (Linking). doi: 10.1016/j.bpj.2011.06.062.

E. Robilotti, S. Deresinski, and B. A. Pinsky. Norovirus. *Clin Microbiol Rev*, 28(1):134–64, 2015. ISSN 0893-8512. doi: 10.1128/cmr.00075-14.

C. V. Robinson, E. W. Chung, B. B. Kragelund, J. Knudsen, R. T. Aplin, F. M. Poulsen, and C. M. Dobson. Probing the nature of noncovalent interactions by mass spectrometry. A study of protein–CoA

ligand binding and assembly. *Journal of the American Chemical Society*, 118(36):8646–8653, 1996. ISSN 0002-7863. doi: 10.1021/ja960211x.

N. E. Robinson and A. B. Robinson. Prediction of protein deamidation rates from primary and three-dimensional structure. *Proceedings of the National Academy of Sciences*, 98(8): 4367–4372, 2001a. ISSN 0027-8424. doi: 10.1073/pnas.071066498.

N. E. Robinson and A. B. Robinson. Deamidation of human proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 98(22): 12409–13, 2001b. ISSN 0027-8424. doi: 10.1073/pnas.221463198.

N. E. Robinson and A. B. Robinson. *Molecular clocks: Deamidation of asparaginyl and glutaminyl residues in pepides and proteins*. Peptides. Althouse Press, 2004. ISBN 1-59087-250-0.

N. E. Robinson and A. B. Robinson. Use of merrifield solid phase peptide synthesis in investigations of biological deamidation of peptides and proteins. *Biopolymers*, 90(3):297–306, 2008. ISSN 0006-3525. doi: 10.1002/bip.20852.

N. E. Robinson, Z. W. Robinson, B. R. Robinson, A. L. Robinson, J. A. Robinson, M. L. Robinson, and A. B. Robinson. Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. *The Journal of Peptide Research*, 63(5):426–436, 2004. ISSN 1397-002X. doi: 10.1111/j.1399-3011.2004.00151.x.

H. Saitô, I. Ando, and A. Ramamoorthy. Chemical shift tensor – the heart of NMR: Insights into biological aspects of proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 57(2):181–228, 2010. ISSN 1873-3301 (Electronic) 0079-6565 (Linking). doi: 10.1016/j.pnmrs.2010.04.005.

M. Salzmann, K. Pervushin, G. Wider, H. Senn, and K. Wüthrich. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci U S A*, 95(23):13585–90, 1998. ISSN 0027-8424. doi: 10.1073/pnas.95.23.13585.

K. Sandra, I. Vandenheede, and P. Sandra. Modern chromatographic and mass spectrometric techniques for protein biopharmaceutical characterization. *J Chromatogr A*, 1335:81–103, 2014. ISSN 1873-3778 (Electronic) 0021-9673 (Linking). doi: 10.1016/j.chroma.2013.11.057.

J. M. Schmidt, S. Zhou, M. L. Rowe, M. J. Howard, R. A. Williamson, and F. Löhr. One-bond and two-bond J couplings help annotate protein secondary-structure motifs: J-coupling indexing applied to human endoplasmic reticulum protein ERp18. *Proteins: Structure, Function, and Bioinformatics*, 79 (2):428–443, 2011. ISSN 0887-3585. doi: 10.1002/prot.22893.

S. Schütz and R. Sprangers. Methyl TROSY spectroscopy: A versatile NMR approach to study challenging biological systems. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 116:56–84, 2020. ISSN 0079-6565. doi: 10.1016/j.pnmrs.2019.09.004.

T. Schulte-Herbruggen and O. W. Sorensen. Clean TROSY: compensation for relaxation-induced artifacts. *J Magn Reson*, 144(1):123–8, 2000. ISSN 1090-7807. doi: 10.1006/jmre.2000.2020.

Y. Shen and A. Bax. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *Journal of biomolecular NMR*, 48(1): 13–22, 2010. ISSN 1573-5001 (Electronic) 0925-2738 (Linking). doi: 10.1007/s10858-010-9433-9.

Y. Shen and A. Bax. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of biomolecular NMR*, 56(3): 227–41, 2013. ISSN 0925-2738. doi: 10.1007/s10858-013-9741-y.

Y. Shen and A. Bax. Homology modeling of larger proteins guided by chemical shifts. *Nature Methods*, 12(8): 747–750, 2015. ISSN 1548-7105 (Electronic) 1548-7091 (Linking). doi: 10.1038/nmeth.3437.

Y. Shen, O. Lange, F. Delaglio, P. Rossi, J. M. Aramini, G. Liu, A. Eletsky, Y. Wu, K. K. Singarapu, A. Lemak, A. Ignatchenko, C. H. Arrowsmith, T. Szyperski, G. T. Montelione, D. Baker, and A. Bax. Consistent blind protein structure generation from NMR chemical shift data. *Proceedings of the National Academy of Sciences*, 105(12):4685–90, 2008. ISSN 1091-6490 (Electronic) 0027-8424 (Linking). doi: 10.1073/pnas.0800256105.

J. Sheng, S. Lei, L. Yuan, and X. Feng. Cell-free protein synthesis of norovirus virus-like particles. *RSC Advances*, 7(46):28837–28840, 2017. ISSN 2046-2069. doi: 10.1039/c7ra03742b.

M. B. Sherman, A. N. Williams, H. Q. Smith, C. Nelson, C. B. Wilen, D. H. Fremont, H. W. Virgin, and T. J. Smith. Bile salts alter the mouse norovirus capsid conformation: Possible implications for cell attachment and immune evasion. *J Virol*, 93(19):e00970-19, 2019. ISSN 1098-5514 (Electronic) 0022-538X (Linking). doi: 10.1128/JVI.00970-19.

N. Sindhuwinata, L. L. Grimm, S. Weißbach, S. Zinn, E. Munoz, M. M. Palcic, and T. Peters. Thermodynamic signature of substrates and substrate analogs binding to human blood group B galactosyltransferase from isothermal titration calorimetry experiments. *Biopolymers*, 99(10): 784–795, 2013. ISSN 0006-3525. doi: 10.1002/bip.22297.

B. K. Singh, M. M. Leuthold, and G. S. Hansman. Human noroviruses' fondness for histo-blood group antigens. *Journal of virology*, 89(4): 2024–40, 2015. ISSN 1098-5514 (Electronic) 0022-538X (Linking). doi: 10.1128/JVI.02968-14.

B. K. Singh, A. Koromyslova, and G. S. Hansman. Structural analysis of bovine norovirus protruding domain. *Virology*, 487:296–301, 2016. ISSN 0042-6822. doi: 10.1016/j.virol.2015.10.022.

D. B. Smith, N. McFadden, R. J. Blundell, A. Meredith, and P. Simmonds. Diversity of murine norovirus in wild-rodent populations: species-specific associations suggest an ancient divergence. *J Gen Virol*, 93(Pt 2):259–66, 2012. ISSN 1465-2099 (Electronic) 0022-1317 (Linking). doi: 10.1099/vir.0.036392-0.

H. Q. Smith and T. J. Smith. The dynamic capsid structures of the noroviruses. *Viruses*, 11(3): 235, 2019. ISSN 1999-4915. doi: 10.3390/v11030235.

S. V. Sosnovtsev and K. Y. Green. Identification and genomic mapping of the ORF3 and VPg proteins in feline calicivirus virions. *Virology*, 277(1):193–203, 2000. ISSN 0042-6822. doi: 10.1006/viro.2000.0579.

L. P. Stratton, R. M. Kelly, J. Rowe, J. E. Shively, D. D. Smith, J. F. Carpenter, and M. C. Manning. Controlling deamidation rates in a model peptide: Effects of temperature, peptide concentration, and additives. *Journal of Pharmaceutical Sciences*, 90(12):2141–2148, 2001. ISSN 00223549. doi: 10.1002/jps.1165.

A. D. Stuart and T. D. Brown. Alpha2,6-linked sialic acid acts as a receptor for feline calicivirus. *J Gen Virol*, 88(Pt 1):177–86, 2007. ISSN 0022-1317 (Print) 0022-1317. doi: 10.1099/vir.0.82158-0.

X. C. Su, S. Jergic, K. Ozawa, N. D. Burns, N. E. Dixon, and G. Otting. Measurement of dissociation constants of high-molecular weight protein-protein complexes by transferred 15N-relaxation. *J Biomol NMR*, 38(1):65–72, 2007. ISSN 0925-2738 (Print) 0925-2738. doi: 10.1007/s10858-007-9147-9.

J. Sun, E. N. Kitova, W. Wang, and J. S. Klassen. Method for distinguishing specific from nonspecific protein–ligand complexes in nanoelectrospray ionization mass spectrometry. *Analytical Chemistry*, 78(9): 3010–3018, 2006. ISSN 0003-2700. doi: 10.1021/ac0522005.

J. F. Sydow, F. Lipsmeier, V. Larraillet, M. Hilger, B. Mautz, M. Mølhøj, J. Kuentzer, S. Klostermann, J. Schoch, H. R. Voelger, J. T. Regula, P. Cramer, A. Papadimitriou, and H. Kettenberger. Structure-based prediction of asparagine and aspartate degradation sites in antibody variable regions. *PLoS ONE*, 9(6): e100736, 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0100736.

K. Takeuchi, H. Arthanari, M. Imai, G. Wagner, and I. Shimada. Nitrogen-detected TROSY yields comparable sensitivity to proton-detected TROSY for non-deuterated, large proteins under physiological salt conditions. *Journal of Biomolecular NMR*, 64(2): 143–151, 2016. ISSN 1573-5001 (Electronic) 0925-2738 (Linking). doi: 10.1007/s10858-016-0015-3.

M. Tamura, K. Natori, M. Kobayashi, T. Miyamura, and N. Takeda. Genogroup II noroviruses efficiently bind to heparan sulfate proteoglycan associated with the cellular membrane. *J Virol*, 78(8): 3817–26, 2004. ISSN 0022-538X. doi: 10.1128/jvi.78.8.3817-3826.2004.

M. Tan and X. Jiang. Norovirus–host interaction: Multi-selections by human histo-blood group antigens. *Trends in Microbiology*, 19(8):382–388, 2011. ISSN 1878-4380 (Electronic) 0966-842X (Linking). doi: 10.1016/j.tim.2011.05.007.

M. Tan and X. Jiang. Histo-blood group antigens: a common niche for norovirus and rotavirus. *Expert Reviews in Molecular Medicine*, 16:e5, 2014. ISSN 1462-3994. doi: 10.1017/erm.2014.2.

M. Tan, W. Zhong, D. Song, S. Thornton, and X. Jiang. E. coli-expressed recombinant norovirus capsid proteins maintain authentic antigenicity and receptor binding capability. *Journal of Medical Virology*, 74(4):641–649, 2004. ISSN 0146-6615. doi: 10.1002/jmv.20228.

M. Tan, J. Meller, and X. Jiang. C-terminal arginine cluster is essential for receptor binding of norovirus capsid protein. *Journal of Virology*, 80(15):7322–7331, 2006. ISSN 0022-538X. doi: 10.1128/JVI.00233-06.

M. Tan, M. Xia, Y. Chen, W. Bu, R. S. Hegde, J. Meller, X. Li, and X. Jiang. Conservation of carbohydrate binding interfaces – evidence of human HBGA selection in norovirus evolution. *PLoS ONE*, 4(4):e5058, 2009. ISSN 1932-6203. doi: 10.1371/journal.pone.0005058.

M. Tan, C. Wei, P. Huang, Q. Fan, C. Quigley, M. Xia, H. Fang, X. Zhang, W. Zhong, J. S. Klassen, and X. Jiang. Tulane virus recognizes sialic acids as cellular receptors. *Sci Rep*, 5:11784, 2015. ISSN 2045-2322. doi: 10.1038/srep11784.

S. Taube, J. W. Perry, K. Yetming, S. P. Patel, H. Auble, L. Shu, H. F. Nawar, C. H. Lee, T. D. Connell, J. A. Shayman, and C. E. Wobus. Ganglioside-linked terminal sialic acid moieties on murine macrophages function as attachment receptors for murine noroviruses. *J Virol*, 83(9):4092–101, 2009. ISSN 1098-5514 (Electronic) 0022-538X (Linking). doi: 10.1128/JVI.02245-08.

S. Taube, J. R. Rubin, U. Katpally, T. J. Smith, A. Kendall, J. A. Stuckey, and C. E. Wobus. High-resolution X-ray structure and functional analysis of the murine norovirus 1 capsid protein protruding domain. *J Virol*, 84(11):5695–705, 2010. ISSN 1098-5514 (Electronic) 0022-538X (Linking). doi: 10.1128/JVI.00316-10.

S. Taube, J. W. Perry, E. McGreevy, K. Yetming, C. Perkins, K. Henderson, and C. E. Wobus. Murine noroviruses bind glycolipid and glycoprotein attachment receptors in a strain-dependent manner. *Journal of Virology*, 86(10):5584–5593, 2012. ISSN 1098-5514 (Electronic) 0022-538X (Linking). doi: 10.1128/JVI.06854-11.

S. Taube, A. O. Kolawole, M. Hohne, J. E. Wilkinson, S. A. Handley, J. W. Perry, L. B. Thackray, R. Akkina, and C. E. Wobus. A mouse model for human norovirus. *mBio*, 4(4):e00450-13, 2013. ISSN 2150-7511. doi: 10.1128/mBio.00450-13.

S. Taube, A. Mallagaray, and T. Peters. Norovirus, glycans and attachment. *Curr Opin Virol*, 31:33–42, 2018. ISSN 1879-6257. doi: 10.1016/j.coviro.2018.04.007.

P. F. M. Teunis, C. L. Moe, P. Liu, S. E. Miller, L. Lindesmith, R. S. Baric, J. Le Pendu, and R. L. Calderon. Norwalk virus: How infectious is it? *Journal of Medical Virology*, 80(8):1468–1476, 2008. ISSN 0146-6615. doi: 10.1002/jmv.21237.

C. Thiele and W. B. Huttner. The disulfide-bonded loop of chromogranins, which is essential for sorting to secretory granules, mediates homodimerization. *J Biol Chem*, 273(2):1223–31, 1998. ISSN 0021-9258. doi: 10.1074/jbc.273.2.1223.

J. D. Thompson, D. G. Higgins, and T. J. Gibson. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22(22):4673–80, 1994. ISSN 0305-1048. doi: 10.1093/nar/22.22.4673.

K. V. Todd and R. A. Tripp. Human norovirus: Experimental models of infection. *Viruses*, 11(2):151, 2019. ISSN 1999-4915. doi: 10.3390/v11020151.

J. H. Tomlinson and M. P. Williamson. Amide temperature coefficients in the protein G B1 domain. *J Biomol NMR*, 52(1):57–64, 2012. ISSN 1573-5001 (Electronic) 0925-2738 (Linking). doi: 10.1007/s10858-011-9583-4.

V. Tugarinov, R. Muhandiram, A. Ayed, and L. E. Kay. Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase G. *J Am Chem Soc*, 124(34):10025–35, 2002. ISSN 0002-7863. doi: 10.1021/ja0205636.

V. Tugarinov, V. Kanelis, and L. E. Kay. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nature Protocols*, 1(2):749–754, 2006. ISSN 1750-2799. doi: 10.1038/nprot.2006.101.

W. B. Turnbull and A. H. Daranas. On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *Journal of the American Chemical Society*, 125(48):14859–14866, 2003. ISSN 0002-7863. doi: 10.1021/ja036166s.

J. van Beek, M. de Graaf, H. Al-Hello, D. J. Allen, K. Ambert-Balay, N. Botteldoorn, M. Brytting, J. Buesa, M. Cabrerizo, M. Chan, F. Cloak, I. Di Bartolo, S. Guix, J. Hewitt, N. Iritani, M. Jin, R. Johne, I. Lederer, J. Mans, V. Martella, L. Maunula, G. McAllister, S. Niendorf, H. G. Niesters, A. T. Podkolzin, M. Poljsak-Prijatelj, L. D. Rasmussen, G. Reuter, G. Tuite, A. Kroneman, H. Vennema, and M. P. G. Koopmans. Molecular surveillance of norovirus, 2005–16: an epidemiological analysis of data collected from the NoroNet network. *The Lancet Infectious Diseases*, 18(5):545–553, 2018. ISSN 1474-4457 (Electronic) 1473-3099 (Linking). doi: 10.1016/s1473-3099(18)30059-8.

J. Van Dycke, A. Ny, N. Conceição-Neto, J. Maes, M. Hosmillo, A. Cuvry, I. Goodfellow, T. C. Nogueira, E. Verbeken, J. Matthijnssens, P. de Witte, J. Neyts, and J. Rocha-Pereira. A robust human norovirus replication model in zebrafish larvae. *PLOS Pathogens*, 15(9):e1008009, 2019. ISSN 1553-7374 (Electronic) 1553-7366 (Linking). doi: 10.1371/journal.ppat.1008009.

G. C. P. van Zundert, J. P. G. L. M. Rodrigues, M. Trellet, C. Schmitz, P. L. Kastritis, E. Karaca, A. S. J. Melquiond, M. van Dijk, S. J. de Vries, and A. M. J. J. Bonvin. The HADDOCK2.2 web server: User-friendly integrative modeling of biomolecular complexes. *Journal of Molecular Biology*, 428 (4):720–725, 2016. ISSN 0022-2836. doi: 10.1016/j.jmb.2015.09.014.

R. A. Venters, B. T. Farmer, C. a. Fierke, and L. D. Spicer. Characterizing the use of perdeuteration in NMR studies of large proteins: 13C, 15N and 1H assignments of human carbonic anhydrase II. *Journal of molecular biology*, 264(5):1101–16, 1996. ISSN 0022-2836. doi: 10.1006/jmbi.1996.0699.

A. V. Verde and D. Frenkel. Simulation study of micelle formation by bile salts. *Soft Matter*, 6(16): 3815–3825, 2010. ISSN 1744-683X. doi: 10.1039/C0SM00011F.

J. Vinje, M. K. Estes, P. Esteves, K. Y. Green, K. Katayama, N. J. Knowles, Y. L'Homme, V. Martella, H. Vennema, and P. A. White. ICTV virus taxonomy profile: Caliciviridae. *J Gen Virol*, 100(11): 1469–1470, 2019. ISSN 0022-1317. doi: 10.1099/jgv.0.001332.

W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function and Genetics*, 59(4): 687–696, 2005. ISSN 0887-3585. doi: 10.1002/prot.20449.

F. C. Walker and M. T. Baldridge. Interactions between noroviruses, the host, and the microbiota. *Curr Opin Virol*, 37:1–9, 2019. ISSN 1879-6257. doi: 10.1016/j.coviro.2019.04.001.

I. H. Walker, P. C. Hsieh, and P. D. Riggs. Mutations in maltose-binding protein that alter affinity and solubility properties. *Appl Microbiol Biotechnol*, 88(1):187–97, 2010. ISSN 0175-7598. doi: 10.1007/s00253-010-2696-y.

G. Wallmann. *Charakterisierung der Ligandenbindung eines humanen Noroviruskapsidproteins mithilfe von Kernspinresonanzspektroskopie*. bachelor's thesis, University of Luebeck, 2019.

W. Wang, E. N. Kitova, and J. S. Klassen. Influence of solution and gas phase processes on protein–carbohydrate binding affinities determined by nanoelectrospray fourier transform ion cyclotron resonance mass spectrometry. *Analytical Chemistry*, 75(19):4945–4955, 2003. ISSN 0003-2700. doi: 10.1021/ac034300l.

H. Wegener, A. Mallagaray, T. Schöne, T. Peters, J. Lockhauserbaumer, H. Yan, C. Uetrecht, G. S. Hansman, and S. Taube. Human norovirus GII.4(MI001) P dimer binds fucosylated and sialylated carbohydrates. *Glycobiology*, 27(11):1027–1037, 2017. ISSN 1460-2423 (Electronic) 0959-6658 (Linking). doi: 10.1093/glycob/cwx078.

G. Wider and K. Wüthrich. NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Current Opinion in Structural Biology*, 9(5):594–601, 1999. ISSN 0959-440X. doi: 10.1016/S0959-440X(99)00011-1.

M. P. Williamson. Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc*, 73:1–16, 2013. ISSN 1873-3301 (Electronic) 0079-6565 (Linking). doi: 10.1016/j.pnmrs.2013.02.001.

C. E. Wobus. The dual tropism of noroviruses. *J Virol*, 92(16):e01010-17, 2018. ISSN 1098-5514 (Electronic) 0022-538X (Linking). doi: 10.1128/JVI.01010-17.

C. E. Wobus, S. M. Karst, L. B. Thackray, K.-O. Chang, S. V. Sosnovtsev, G. Belliot, A. Krug, J. M. Mackenzie, K. Y. Green, and H. W. I. V. Virgin. Replication of norovirus in cell culture reveals a tropism for dendritic cells and macrophages. *PLOS Biology*, 2(12):e432, 2004. ISSN 1545-7885 (Electronic) 1544-9173 (Linking). doi: 10.1371/journal.pbio.0020432.

C. E. Wobus, L. B. Thackray, and H. W. Virgin. Murine norovirus: a model system to study norovirus biology and pathogenesis. *Journal of Virology*, 80(11):5104–5112, 2006. ISSN 0022-538X. doi: 10.1128/JVI.02346-05.

C.-M. Yu, S. Mun, and N.-H. L. Wang. Theoretical analysis of the effects of reversible dimerization in size exclusion chromatography. *Journal of Chromatography A*, 1132(1):99–108, 2006. ISSN 0021-9673. doi: 10.1016/j.chroma.2006.07.017.

M. Zakhour, N. Ruvoen-Clouet, A. Charpilienne, B. Langpap, D. Poncet, T. Peters, N. Bovin, and J. Le Pendu. The alphaGal epitope of the histo-blood group antigen family is a ligand for bovine norovirus Newbury2 expected to prevent cross-species transmission. *PLoS Pathog*, 5(7):e1000504, 2009. ISSN 1553-7366. doi: 10.1371/journal.ppat.1000504.

Appendix

6.7 Protein Biosynthesis and Purification

The following chapter shows the step-wise purification of recombinant NoV P-domains on the example of the [*U*-^{2}H,^{15}N]-labeled GII.4 VA387 P-domain. Experimental details can be found in chapters 6.3.2, 6.3.3, and 6.3.4. Both human and murine NoV P-domain behave almost identically and can be purified with the same strategy.

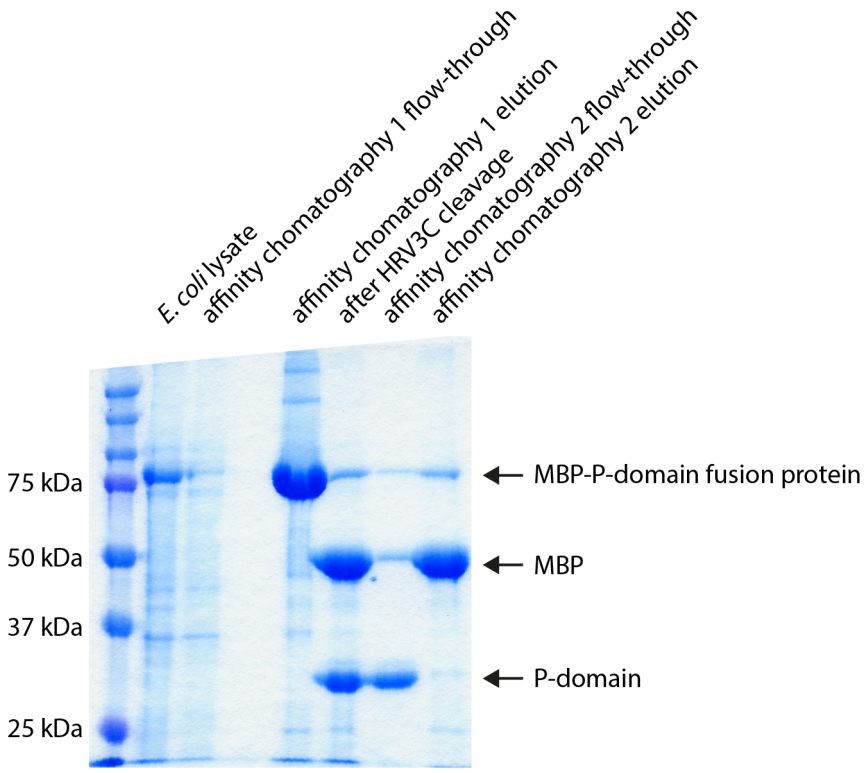


Figure 6.2: Denaturing SDS page gel of samples taken during the purification of the recombinant P-domain from GII.4 VA387. The gel was obtained as detailed in chapter 6.3.4. Briefly, the MBP-P-domain fusion protein binds to a NiNTA affinity column and can be selectively eluted using high imidazole concentrations. The elution fraction can be proteolytically digested yielding MBP and the P-domain as cleavage products. A second affinity chromatography can separate both species. The flow-through of the second affinity chromatography is then collected, concentrated, and further purified by preparative size exclusion chromatography (cf. Fig. 6.3).

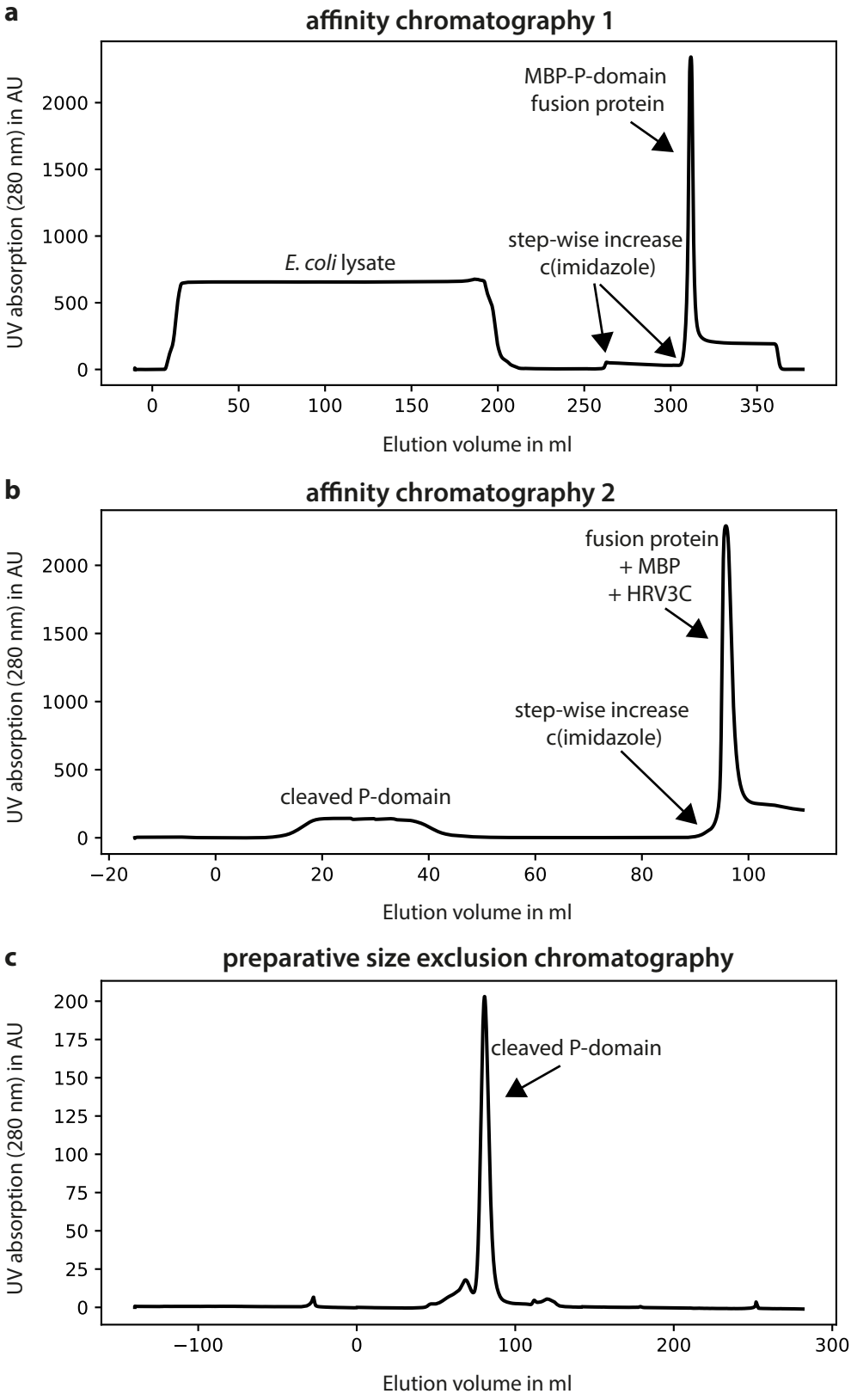


Figure 6.3: UV chromatograms of the three chromatographic purification steps of the GII.4 VA387 P-domains. The UV absorption was monitored at 280 nm. The respective protein species described in the legend of Fig. 6.2 have been labeled. Experimental details are given in chapters 6.3.2-6.3.3.

6.8 Hydrophobic Interaction Chromatography

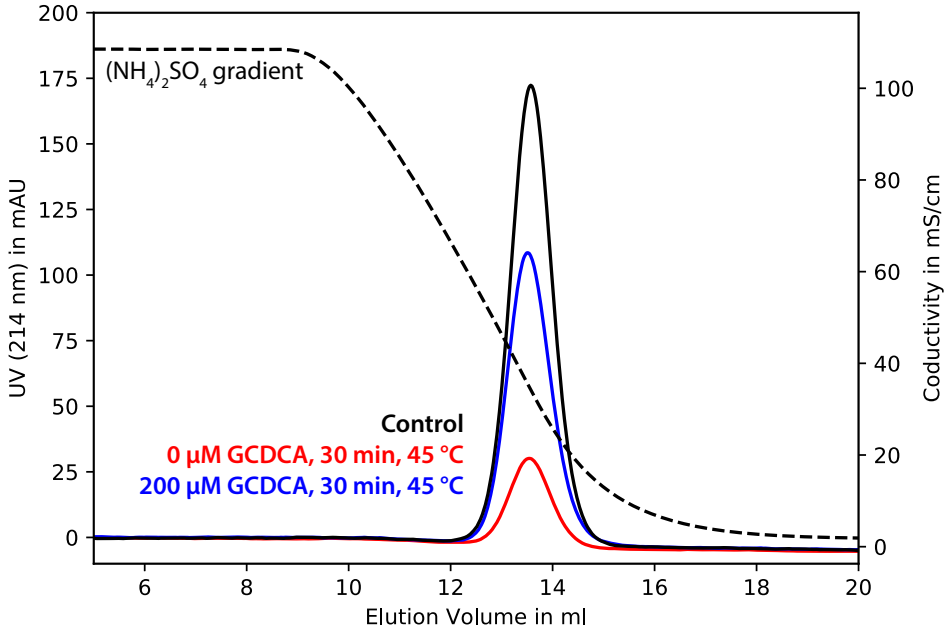


Figure 6.4: Exemplary UV chromatograms from hydrophobic interaction chromatography (HIC) experiments after isothermal denaturation of the CW1 murine NoV P-domain with varying amounts of GCDCA. Experimental detail are given in chapter 6.3.7. UV absorption has been monitored at 214 nm. The ammonium sulfate gradient is depicted as a dashed line. A non-denatured sample served as control (black). Samples were prepared with varying amounts of GCDCA in 20 mM sodium acetate, 100 mM NaCl (pH 5.3), denatured for 30 min at 45 °C, and then subjected to HIC analysis. HIC data were acquired by C. Feldmann as part of supervised work as a research assistant.

6.9 Simultaneous Deamidation and Dimer Reassembly

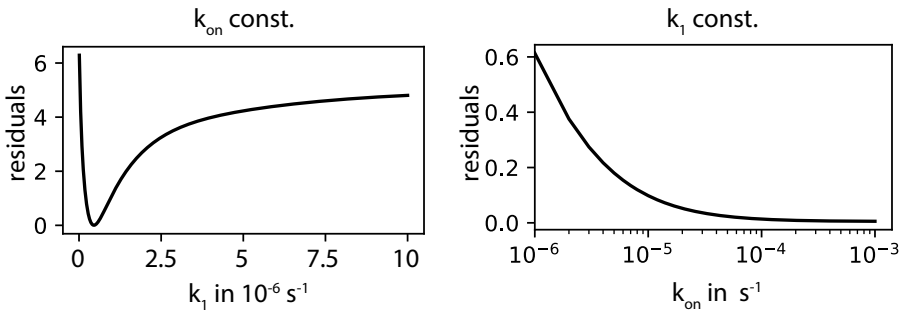


Figure 6.5: Residuals from fitting of the deamidation reaction rate in Fig. 4.13. Fitting of experimental IEX data to the numerical solution of the system of differential equations describing simultaneous deamidation and dimer reassembly allows for determination of the deamidation rate k_1 but not of the association rate k_{on}. Choosing a fixed and arbitrary $k_{on} = 10^3 \mathrm{s}^{-1}$, variation of k_1 leads to a narrow minimum in squared residual of the fit at $4.5 \times 10^{-7} s^{-1}$. In contrast, fixing k_1 at this value and varying k_{on} over a meaningful range does not produce any well-defined, narrow minimum. Accordingly, k_{on} cannot be determined from curve fitting of the IEX data and its variation has no discernible effect on the solution.

6.10 Source Code

6.10.1 Chemical Shift Referencing

```
"""
(C) Robert Creutznacher
v1.1 (09.01.2017)
"""

if int(GETPROCDIM() ) != 2:
        ERRMSG("2D Data Only", "Error")
        EXIT()

result = INPUT_DIALOG("Choose X Nucleus",
"Input '15N' or '13C', 2D Experiments Only",
["Nucleus = "],
["15N"], [""], ["1"],
["Accept","Close"], ['a','c'], 10)
if result <> None:
        if result[0] == "15N":
                X = result[0]
                BF_X = GETPAR('BF3')
                Xi = 0.101329118
        elif result[0] == "13C":
                X = result[0]
                BF_X = GETPAR('BF2')
                Xi = 0.251449530

        else:
                ERRMSG("Input wasn't '15N' or '13C'", "Error")
                EXIT()
        print "Calibrated with " + result[0]

BF_1 = 1E6 * float(GETPAR('BF1')) #1H base frequency in Hz
SR_H = GETPAR('SR', 2)
BF_X = 1E6 * float(BF_X)
SR_X = (float(BF_1) + float(SR_H))*Xi - BF_X

print "BF_1 in Hz: " +  str(BF_1)
print "SR_H in Hz: " +  str(SR_H)
print "BF_X in Hz: " + str(BF_X)
print "SR_X in Hz: " + str(SR_X)
XCMD("1 SR " + str(SR_X))
```

6.10.2 Fitting of Simultaneous Deamidation and Dimer Reassembly Reaction Rates

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.integrate import odeint

def f(y, t, params): #Eq. 4.1-4.5 of the main text
    NN, iDN, iDiD, N, iD = y
    k_off, k_on, k1 = params                # unpack parameters
    f1 = -(2*k1+k_off)*NN + k_on*N**2
    f2 = 2*k1*NN - (k1+k_off)*iDN + 2*k_on*iD*N
    f3 = k1*iDN - k_off*iDiD + k_on*iD**2
    f4 = -k1*N -2*k_on*N**2 + 2*k_off*NN -2*k_on*N*iD + k_off*iDN
    f5 = k1*N -2*k_on*iD**2 + 2*k_off*iDiD -2*k_on*N*iD + k_off*iDN
    derivs = [f1,f2,f3,f4,f5]          # list of dy/dt=f functions
    return derivs

#Experimental data
"""
#asymmetrical
t_exp =np.asarray([0.,1.55E+05,4.97E+05,1.02E+06,1.89E+06,\
                   3.03E+06,5.42E+06])
NN_exp=np.asarray([0.000,0.040,0.073,0.070,0.037,0.011,0.000])
iDN_exp=np.asarray([1.000,0.867,0.668,0.515,0.355,0.230,0.093])
iDiD_exp=np.asarray([0.,0.093,0.259,0.415,0.608,0.759,0.907])
"""
#NN start
t_exp =np.asarray([0.,3.13E+05,9.16E+05,1.45E+06,2.13E+06,\
                   2.90E+06,3.53E+06,4.54E+06])
NN_exp=np.asarray([1.,0.735,0.421,0.248,0.130,0.071,0.039,0.016])
iDN_exp=np.asarray([0.,0.248,0.464,0.506,0.473,0.401,0.335,0.243])
iDiD_exp=np.asarray([0.,0.017,0.115,0.246,0.397,0.528,0.626,0.741])

# Initial values
NN_0 = 1.0
idN_0 = 0.0
iDiD_0 = 0.0
N_0 = 0.
iD_0 = 0.

# Bundle initial conditions for ODE solver
y0 = [NN_0, idN_0, iDiD_0, N_0, iD_0]

# Make time array for solution
tStop = 0.6E7
tInc = 1000.
t = np.arange(0., tStop, tInc)

K1_arr = np.arange(0.01E-6,10E-6,0.01E-6)
res = []

for i in range(len(K1_arr)):
    # Parameters
    k_on     = 1E3
    k_off    = 1.51E-6
    k1       = K1_arr[i]

    # Bundle parameters for ODE solver
```

```
    params = [k_off, k_on, k1]

    # Call the ODE solver
    psoln = odeint(f, y0, t, args=(params,))
    NN_sim      =  psoln[:,0][np.where(np.in1d(t, t_exp))]
    iDN_sim     =  psoln[:,1][np.where(np.in1d(t, t_exp))]
    iDiD_sim    =  psoln[:,2][np.where(np.in1d(t, t_exp))]

    res.append( np.sum( (np.subtract(NN_exp,NN_sim)**2 + \
            np.subtract(iDN_exp,iDN_sim)**2 + \
           np.subtract(iDiD_exp,iDiD_sim)**2) ) )

res = np.asarray(res)
ind = np.unravel_index(np.argmin(res, axis=None), res.shape)

print res[ind]

k1 = K1_arr[ind[0]]
print k1

params = [k_off, k_on, k1]
psoln = odeint(f, y0, t, args=(params,))

# Plot results
fig, ax1 = plt.subplots (1,1)

ax1.plot(t, psoln[:,0], c = 'black', label = 'NN')
ax1.plot(t, psoln[:,1], c = 'blue', label = 'iDN')
ax1.plot(t, psoln[:,2], c = 'red', label = 'iDiD' )
ax1.set_xlabel('time')
ax1.set_ylabel('dimers')
ax1.scatter(t_exp,NN_exp,c = 'black', s =16 )
ax1.scatter(t_exp,iDN_exp,c = 'blue', s =16 )
ax1.scatter(t_exp,iDiD_exp,c = 'red', s =16 )

plt.tight_layout()
plt.show()
```

6.10.3 Analysis of TRACT experiments

```
from matplotlib import pyplot as plt
import numpy as np
from math import sqrt
from math import cos

y_N = -2.7116E07
y_H =  2.67513E08

h    = 1.0545E-34
r    = 1.02E-10                #according to TRACT paper
w_H = 600.23E06
B0   = w_H/1E6/42.567
w_N = abs(y_N/y_H)*w_H*2*np.pi
dS   = 160E-06                 #value from TRACT paper
angle = (3*cos(17*np.pi/180)**2-1)

u0   = 1.2566E-06

def R1(tau):
    p = u0 * y_H*y_N*h/(8*np.pi *sqrt(2)* r**3)
    d =  y_N*B0*dS/(sqrt(2)*3)
    J0 = 0.4*tau
    J_w= 0.4*tau/(1+(w_N*tau)**2)
    return 2*p*d * (4*J0+3*J_w) *angle

tau=np.arange(1E-09,100E-09,1E-09)
R_diff = 64-13 #insert Ra and Rb here
y = R1(tau)

plt.axhline(y=(R_diff), xmin=0, xmax=1, ls = '--', c='black')

plt.plot(tau, y)
plt.show()
```

6.10.4 MNV-GCDCA Interaction

```
from matplotlib import pyplot as plt
import numpy as np
from scipy.optimize import fsolve
import pandas as pd

#imput data normalized to unity
path = r'C:\Users\rober\Downloads\CW1_GCDCA\M.csv'
df = pd.read_table(path, sep = ';')
df = df.set_index('keys')

L_loop = np.arange(0, 300E-6, 1E-6)
P_t = 100E-6
sol_M = []
sol_D = []
sol_DL = []
sol_DLL = []

def f(M): # Eq. 4.11
    L = - (KD2**2 + KD2*M**2/KD1)/(4*M**2/KD1) +\
        np.sqrt( ((KD2**2 + KD2*M**2/KD1)/(4*M**2/KD1))**2 +\
        (KD2**2*L_t)/(2*M**2/KD1))
    return M + 2* M**2/KD1 + 2* M**2/(KD1*KD2) * L + \
        2* M**2/(KD1*KD2**2) * L**2 - P_t

def L_free(M, KD1, KD2, L_t): # Eq. 4.10
    L = - (KD2**2 + KD2*M**2/KD1)/(4*M**2/KD1) +\
        np.sqrt( ((KD2**2 + KD2*M**2/KD1)/(4*M**2/KD1))**2 +\
        (KD2**2*L_t)/(2*M**2/KD1))
    return L

c_GCDCA = [0,15E-6,30E-6,50E-6,75E-6,100E-6,140E-6]

# Initialize arrays of residuals
res_M = 0
res_DL = 0
res_arr = []
all_res = []
KD1 = 31E-6 # Dimerisation KD known from other experiments

#arrays of GCDCA binding KDs to be tested
KD2_arr = np.arange(0.01E-6,20E-6,0.1E-6)
KD_out =[]
df_out = pd.DataFrame()
res_total_all_signals = 0
res_total_all_signals_out = []

for k2 in range(len(KD2_arr)):
    KD2 = KD2_arr[k2]
    KD_out.append(KD2)
    all_res = []
    KD_out =[]
    sol_M = []
    sol_D = []
    sol_DL = []
    sol_DLL = []
    res_total_all_signals = 0
```

```
    #iterate through M intensity curves
    for j in df.iterrows():
        norm =32E-6/j[1][0]
        a = j[1].multiply(norm) #starting [M] - KD1 is known

        #iterate through individual curve, each data point
        for i in range(len(a)):
            L_t = c_GCDCA[i]

            if i == 0:
                start = P_t
            else:
                start = sol_M[i-1]

            #numerical solution for M, derive other
            #concentrations based on KD definitions

            M = abs(fsolve(f,start)[0])
            D = abs((M**2/KD1))
            DL = M**2/KD1 * L_free(M, KD1, KD2, L_t )/KD2
            DLL = (DL * L_free(M, KD1, KD2, L_t )/KD2)
            res = (a[i]-M)**2 # residuals for [M]
            #res = (a[i]-(50E-6*(DL + DLL)/(D + DL + DLL)))**2
            #res = (a[i]-(50E-6*(DL + 2*DLL)/(2*D + \
                # 2*DL + 2*DLL)))**2
            #res = (a[i] - (0.5*DL + DLL))**2
            sol_M.append(M)

            res_arr.append(res)
            res = 0

        res_total_all_signals += np.nansum(res_arr)
        res_arr=[]
    res_total_all_signals_out.append(res_total_all_signals)
res_total_all_signals_out = \
    np.asarray(res_total_all_signals_out)

fig, ax = plt.subplots (1,1)

ax.plot(1E6*KD2_arr,res_total_all_signals_out, c= 'black')
plt.tight_layout()
plt.show()
```

6.10.5 Calcium Contamination

```
from matplotlib import pyplot as plt
import numpy as np
from scipy.optimize import least_squares
import pandas as pd

def func(L, KD,CSP_max): # Williamson et al., 2013
    P_t =250E-6
    CSP = CSP_max * ( (P_t+L+KD) - \
        np.sqrt( (P_t+L+KD)**2 - 4 * P_t * L) )/ (2*P_t)
    return CSP

def err_global(arg, xdata, ydata):
    kd= arg[0]
    oarr = arg[1:]
    error = []

    for k,y in enumerate(ydata):
        o = oarr[int(k)]
        e = 0
        for i,l in enumerate(xdata):
            e += ((func(l,kd,o)-y[i]))**2
        error.append(e)
    return error

path = r'C:\Users\rober\CW1_GCDCA_Ca\Data_forfit1.csv'

#possible contaminations to test
contam = np.arange(0,1E-3,1E-6)
res_tot = []

df_data = pd.read_table(path, sep = ';')
df_data = df_data.drop('keys', axis=1)
ydata = df_data.values
x0 = np.append([100E-6], np.array(df_data["5"].values))
xdata = np.array([0,0,0.2E-3,0.6E-3,1.3E-3,3.9E-3,5E-3])

for i in range(len(contam)):
    x = contam[i]

    #EDTA spectrum as starting point,
    #other spectra contamination with x
    x_add = np.sum([xdata,[0,x,x,x,x,x,x]],axis = 0)
    res = least_squares(err_global,x0,args = ([x_add, ydata]))
    res_tot.append(np.sum(res.fun))

plt.scatter(contam, res_tot)
plt.show()
```

6.10.6 Pulse program ^{15}N T_1 times

```
#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

/*#define LABEL_CN*/
#define TEMP_COMPENSATION

"in0=inf1*0.5"

# ifdef LABEL_CN
"d0=97u-p4*2+p7*0.66-p1*0.5"
"d25=20m-p15*0.5-p4*4-8u"
#else
"d0=100u+p7*0.66-p1*0.5"
"d25=20m-p15*0.5"
#endif /*LABEL_CN*/

"d11=30m"
"DELTA=2.65m"
"DELTA1=2.65m"
"DELTA2=2.65m-p22-p11-300u"
"DELTA3=2.65m-p23-p10-300u"
"DELTA4=260u-p24-p1*0.66"

"d27=p24+35u"

"l1=1"
"l2=1"

"cnst21=176"
"cnst22=56"
"cnst18=-800"

"spoff4=bf2*((cnst22-cnst21)/1000000)"

1       ze
        1m
2       d11 do:f2
        1m LOCKH_OFF
        3m
3       1m
        1m
4       3m
5       2m BLKGRAD
        10u pl1:f1
        10u pl4:f2
        10u pl7:f3
        (p7 ph0):f3
        10u

# ifdef TEMP_COMPENSATION

"d17=d1-p18"

        10u fq=cnst18(bf ppm):f3
        10u pl8:f3
```

```
        (p18 ph0):f3
        10u
        10u fq=0:f3
        d17
# else
        d1
# endif
        1m UNBLKGRAD
        10u pl7:f3

        (p7 ph0):f3
        5u
        p20:gp6
        200u

        (p1 ph0):f1
        5u
        DELTA gron0
        5u groff
        (center(p1*2 ph0):f1 (p7*2 ph0):f3)
        5u
        DELTA gron0
        5u groff

        (p1 ph5):f1 (p7 ph0):f3
        5u
        DELTA1 gron1
        5u groff
        (center (p1*2 ph0):f1 (p7*2 ph0):f3)
        5u
        DELTA1 gron1
        5u groff
        (p7 ph6):f3
        5u

        (p1 ph2):f1
        5u
        p21:gp7
        100u
        100u pl0:f1

if "l2==1" goto 77

70      d25*0.5
#  ifdef LABEL_CN
        3u pl4:f2
        (p4*2 ph0 3u 3u pl2 p4*2:sp4 ph0):f2
#  endif
        d25*0.5
        (p15:sp5 ph0):f1
        d25*0.5
#  ifdef LABEL_CN
        3u pl4:f2
        (p4*2 ph0 3u 3u pl2 p4*2:sp4 ph0):f2
# endif
        d25*0.5
        lo to 70 times c
```

```
 77      3u
         3u pl4:f2
         3u pl1:f1
         if "l1==1"
         {
         (p7 ph7):f3
         10u
         p25:gp5
         200u
         (p7*2 ph7):f3
         10u
         p25:gp5*-1
         }
         else
         {
         (p7 ph17):f3
         10u
         p25:gp5*-1
         200u
         (p7*2 ph17):f3
         10u
         p25:gp5
         }

         d0
# ifdef LABEL_CN
         (p4*2 ph0 3u 3u pl2 p4*2:sp4 ph0):f2
# endif
         d0
         if "l1==1"
         {
         (p1 ph1):f1
         3u
         3u pl0:f1
         (p11:sp11 ph11:r):f1
         6u
         }
         else
         {
         (p1 ph3):f1
         3u
         3u pl0:f1
         (p11:sp11 ph13:r):f1
         6u
         }
         5u pl1:f1
         DELTA2
         p22:gp2
         300u
         (center (p1*2 ph0):f1 (p7*2 ph0):f3)
         7u
         p22:gp2
         DELTA2
         300u pl0:f1

         (p11:sp12 ph12:r):f1
         5u
         3u pl1:f1
```

```
        if "l1==1"
        {
        (p1 ph0):f1 (p7 ph1):f3
        }
        else
        {
        (p1 ph0):f1 (p7 ph3):f3
        }
        DELTA3
        p23:gp3
        200u
        100u pl10:f1
        (center(p10 ph10:r 5u pl1 p1*2 ph0 5u
                        pl10 p10 ph10:r):f1 (p7*2 ph0 d27):f3)
        5u
        p23:gp3
        DELTA3
        DELTA4
        (p7 ph0):f3
        5u
        p24:gp4
999     5u
        5u pl31:f2
        20u BLKGRAMP
        go=2 ph31 cpds2:f2
        1m do:f2
        1m LOCKH_OFF
        d11 wr #0 if #0 zd
        1m ivc
        1m iu2
        lo to 3 times l6
        1m iu1
        1m igrad EA
        1m ru2
        lo to 4 times 2
        1m id0
        1m ru1
        lo to 5 times l3
1m
1m BLKGRAD
exit

ph0=0
ph1=1
ph2=2
ph3=3
ph5=1
ph6=1 1 1 1  3 3 3 3
ph10=2
ph11=3
ph12=0
ph13=1
ph7=1 0 3 2
ph17=1 2 3 0
ph31=1 2 3 0  3 0 1 2
```

6.10.7 Pulse program ^{15}N $T_{1\rho}$ times

```
#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

/*#define LABEL_CN*/
#define TEMP_COMPENSATION

"in0=inf1*0.5"

# ifdef LABEL_CN
"d0=97u-p4*2+p7*0.66-p1*0.5"
#else
"d0=100u+p7*0.66-p1*0.5"
#endif /*LABEL_CN*/

"d11=30m"
"DELTA=2.65m"
"DELTA1=2.65m"
"DELTA2=2.65m-p22-p11-300u"
"DELTA3=2.65m-p23-p10-300u"
"DELTA4=260u-p24-p1*0.66"

"d27=p24+35u"

"l1=1"
"l2=1"

"cnst21=176"
"cnst22=56"
"cnst18=-800"

"spoff4=bf2*((cnst22-cnst21)/1000000)"

1       ze
        1m
2       d11 do:f2
        1m LOCKH_OFF
        3m
3       1m
        1m
4       3m
5       2m BLKGRAD
        10u pl1:f1
        10u pl4:f2
        10u pl7:f3
        (p7 ph0):f3
        10u

# ifdef TEMP_COMPENSATION
"p17=p18+1m-vp"
"d17=d1-p17"
# endif

"d8=p8+vp*0.25-p1*2-2u"
"d9=vp*0.5-p1*4-24u"
"d29=p8"
```

```
#ifdef TEMP_COMPENSATION
        10u fq=cnst18(bf ppm):f3
        10u pl8:f3
        (p17 ph0):f3
        10u
        10u fq=0:f3
        d17
#else
        d1
#endif
        1m UNBLKGRAD
        10u pl7:f3

        (p7 ph0):f3
        5u
        p20:gp6
        200u

        (p1 ph0):f1
        5u
        DELTA gron0
        5u  groff
        (center(p1*2 ph0):f1 (p7*2 ph0):f3)
        5u
        DELTA gron0
        5u groff

        (p1 ph5):f1  (p7 ph0):f3
        5u
        DELTA1 gron1
        5u groff
        (center (p1*2 ph0):f1 (p7*2 ph0):f3)
        5u
        DELTA1 gron1
        5u groff
        (p7 ph6):f3
        5u
        (p1 ph2):f1
        5u
        p21:gp7
        100u
        100u pl8:f3

# ifdef LABEL_CN
        (d8 p1 ph0 3u p1*2.3 ph1 3u p1 ph0 3u d9*0.5 gron9 3u
                        groff 6u d9*0.5 gron9*-1 3u groff 6u p1 ph0 3u
                        p1*2.3 ph1 3u p1 ph0):f1 (p8:sp8 ph0 3u pl8 vp
                        ph0 p8:sp9 ph0):f3 (d29 d9 p4*2 ph0 3u 3u pl2
                        p4*2:sp4 ph0):f2
#else
        (d8 p1 ph0 3u p1*2.3 ph1 3u p1 ph0 3u d9*0.5 gron9 3u
                        groff 6u d9*0.5 gron9*-1 3u groff 6u p1 ph0 3u
                        p1*2.3 ph1 3u p1 ph0):f1 (p8:sp8 ph0 3u pl8 vp
                        ph0 p8:sp9 ph0):f3
#endif
        5u
        p21:gp8
```

```
        200u

        3u pl7:f3
        3u pl4:f2
        3u pl1:f1
        if "l1==1"
        {
        (p7 ph7):f3
        10u
        p25:gp5
        200u
        (p7*2 ph7):f3
        10u
        p25:gp5*-1
        }
        else
        {
        (p7 ph17):f3
        10u
        p25:gp5*-1
        200u
        (p7*2 ph17):f3
        10u
        p25:gp5
        }
        d0
# ifdef LABEL_CN
        (p4*2 ph0 3u 3u pl2 p4*2:sp4 ph0):f2
# endif
        d0
        if "l1==1"
        {
        (p1 ph1):f1
        3u
        3u pl0:f1
        (p11:sp11 ph11:r):f1
        6u
        }
        else
        {
        (p1 ph3):f1
        3u
        3u pl0:f1
        (p11:sp11 ph13:r):f1
        6u
        }
        5u pl1:f1
        DELTA2
        p22:gp2
        300u
        (center (p1*2 ph0):f1 (p7*2 ph0):f3)
        7u
        p22:gp2
        DELTA2
        300u pl0:f1
        (p11:sp12 ph12:r):f1
        5u
        3u pl1:f1
```

```
        if "l1==1"
        {
        (p1 ph0):f1 (p7 ph1):f3
        }
        else
        {
        (p1 ph0):f1 (p7 ph3):f3
        }
        DELTA3
        p23:gp3
        200u
        100u pl10:f1
        (center(p10 ph10:r 5u pl1 p1*2 ph0 5u pl10 p10
                    ph10:r):f1 (p7*2 ph0 d27):f3)
        5u
        p23:gp3
        DELTA3
        DELTA4
        (p7 ph0):f3
        5u
        p24:gp4
999     5u
        5u pl31:f2
        20u BLKGRAMP
        go=2 ph31 cpds2:f2
        1m do:f2
        d11 wr #0 if #0 zd
        1m LOCKH_OFF
        2m ivp
        lo to 3 times l6
        3m iu1
        lo to 4 times 2
        1m id0
        1m ru1
        lo to 5 times l3
1m
1m BLKGRAD
exit

ph0=0
ph1=1
ph2=2
ph3=3
ph5= 1
ph6= 1 1 1 1  3 3 3 3
ph9=2
ph10=2
ph11=3
ph12=0
ph13=1
ph7=1 0 3 2
ph17=1 2 3 0
ph31=1 2 3 0  3 0 1 2
```