Interrogating Protein-Ligand Interactions through Structural Mass Spectrometry, Computational Simulations and Biophysical Techniques

Dissertation

zur Erlangung des
Doktorgrades der Naturwissenschaften (*Dr. rer. nat.*)

der
Naturwissenschaftlichen Fakultät I - Biowissenschaften
der Martin-Luther-Universität
Halle-Wittenberg,

vorgelegt von

Herr Julian Bender
geboren am [redacted]
in [redacted]

Gutachter:
Prof. Dr. Carla Schmidt
Apl. Prof. Dr. Ralph Golbik
Prof. Dr. Florian Stengel

Verteidigt am: 29.06.2023

"The most exciting phrase to hear in science,
the one that heralds the most discoveries,
is not 'Eureka!' (I found it!)
but 'That's funny...'."

- Isaac Asimov

Contents

ZUSAMMENFASSUNG

Proteininteraktionen sind für die Organisation und Regulierung biochemischer Prozesse in lebenden Organismen unerlässlich. In der vorliegenden Arbeit werden Protein-Lipid- und Protein-Ribonukleinsäure (RNS)-Wechselwirkungen im Detail untersucht: Im Fokus stehen die Bindung von Lipiden durch das regulatorische Protein Syt-1 und die Interaktion des *turnip crinkle virus* (TCV)-Hüllproteins p38 mit RNS.

Synaptotagmin-1 (Syt-1) ist ein zentraler Regulator der Neurotransmitter-Exozytose eines Hauptvorgangs der Neurotransmission. Es löst die SNARE-vermittelte (engl. Abkürzung für *soluble N-ethylmaleimide-sensitive-factor-attachment receptor*) Fusion von synaptischen Vesikeln (SVs) mit der präsynaptischen Membran als Reaktion auf den Einstrom von Ca^{2+} in das Neuron aus. Bei diesem Prozess interagieren die Ca^{2+}-bindenden C2-Domänen von Syt-1 mit SNARE-Proteinen und Lipiden der SV- und Plasmamembranen. Viele molekulare Details dieser Interaktionen sind jedoch nicht geklärt, sodass der molekulare Mechanismus der Funktion von Syt-1 noch diskutiert wird.

Im ersten Teil dieser Arbeit wurden die Lipidpräferenzen von Syt-1 experimentell charakterisiert, wobei native Massenspektrometrie (MS), eine Methode, die sich besonders für die Analyse nichtkovalenter Komplexe eignet, und zwei komplementäre biophysikalische Techniken eingesetzt wurden. Zunächst wurde Syt-1 in voller Länge sowie verkürzte lösliche Varianten, die entweder die C2A-Domäne, die C2B-Domäne oder beide Domänen (C2AB) enthalten, rekombinant in *Escherichia coli* exprimiert, gereinigt und mit nativer MS charakterisiert. Syt-1 lag überwiegend als Monomer vor und für C2AB wurde ein geringer Grad an Dimerisierung beobachtet. Um die Interaktion von Syt-1 mit Membranlipiden zu untersuchen, wurde, unter Verwendung verschiedener Detergenzien und Standardproteine, eine auf nativer MS basierende Methode zur Bestimmung der Bindung von Lipiden an lösliche Proteine etabliert. Das hierbei verwendete Detergens C8E4 (Octyl-Tetraethylenglycolether) reduzierte die durchschnittliche Ladung der Protein-Ionen und ermöglichte die Analyse von Protein-Lipidkomplexen mittels nativer MS. Diese Methode wurde auf die löslichen Syt-1-Varianten angewandt und zeigte, dass C2AB vorzugsweise mit mehrfach phosphorylierten Phosphatidylinositolen interagierte, während die Varianten C2A oder C2B die Bindung an Phosphatidylserin und Phosphatidylinositol bevorzugten. Um ergänzende Daten zur Bindung von Syt-1 an Membranlipide zu erhalten, wurde außerdem (1) die Bindung von Syt-1 an immobilisierte Lipide und (2) die Interaktion von C2A und C2B mit intakten Membranen durch Liposomendichtegradientenzentrifugation untersucht.

Außerdem wurden die experimentellen Methoden durch eine 900 ns Molekulardynamiksimulation ergänzt, die eine räumliche und zeitliche Auflösung ermöglicht, wie sie mit den hier verwendeten experimentellen Methoden nicht möglich ist. Konkret wurde die Interaktion von C2A und C2B mit SV- und Plasmamembran-Modellmembranen, die in ihrer Zusammensetzung den Lipidmischungen der Liposomendichtegradientenzentrifugation entsprachen, in An- und Abwesenheit von Ca^{2+} simuliert. Die Simulationen lieferten Einblicke in die Lipidpräferenz von Syt-1, die strukturelle *loop*-Dynamik, die Membraninteraktionswinkel und die Membraninsertion in An- und Abwesenheit von Ca^{2+} und zeigten, dass C2A und C2B unterschiedlich mit Lipidmembranen interagieren. C2A bindet bevorzugt an die SV-Membran, während C2B bevorzugt mit der Plasmamembran interagiert. Die Anwesenheit von Ca^{2+} erhöht diese Membranaffinitäten und führt zu einer Änderung des Membraninteraktionswinkels für C2A. Insgesamt deuten die Experimente und Computersimulationen auf eine zentrale Rolle der negativ geladenen Phospholipide für die Membraninteraktion von Syt-1 hin.

Im zweiten Teil dieser Arbeit wurden die strukturellen Auswirkungen der Bindung von RNS an p38 analysiert. p38 spielt eine Doppelrolle, indem es das virale Kapsid bildet, aber auch als siRNA-Fänger (eng. Abkürzung *small interfering RNA* etwa kleine eingreifende RNS) fungiert, um die Wirtsabwehr zu behindern. Die RNS-Bindung findet in der löslichen Form von p38 statt, die meisten Strukturstudien zu p38 konzentrierten sich jedoch bislang auf das Protein in seiner Kapsidform.

Daher wurde die Oligomerisierung von p38 in Abhängigkeit von RNS unterschiedlicher Länge mit nativer MS untersucht. In Lösung bildet p38 Dimere, die leicht Untereinheiten austauschen. In Gegenwart von RNS erfolgt kein Austausch von Untereinheiten, möglicherweise aufgrund der Bindung von RNS an p38. Chemische Quervernetzung von p38 mit Molekülen unterschiedlicher Länge in An- und Abwesenheit von RNS und bei unterschiedlichen Temperaturen wurde verwendet, um Informationen über die Struktur von löslichem p38 zu gewinnen. Die identifizierten Querverbindungen zusammen mit computerbasiertem *Docking* von p38-Domänen und Molekulardynamiksimulationen der resultierenden Modelle weisen auf eine kompakte Struktur von p38 in Lösung hin. Allerdings kann die Existenz eines hochflexiblen Ensembles von p38-Konformeren nicht ausgeschlossen werden.

Insgesamt trägt diese Arbeit zum allgemeinen Verständnis von Proteininteraktionen mit Lipiden und RNA bei, indem sie die Fähigkeit von struktureller MS und Molekulardynamiksimulationen zur Untersuchung verschiedener Aspekte von Proteininteraktionen veranschaulicht.

Abstract

Protein interactions are essential for the organisation and regulation of biochemical processes in living organisms. In the present work, protein-lipid and protein-ribonucleic acid (RNA) interactions are studied in detail, namely the binding of lipids by the regulatory protein Synaptotagmin-1 (Syt-1) and the interaction of the turnip crinkle virus (TCV) coat protein p38 with RNA.

Syt-1 is a central regulator of neurotransmitter exocytosis, a main process during neurotransmission. It triggers the SNARE-mediated fusion of synaptic vesicles (SVs) with the presynaptic membrane in response to the influx of Ca^{2+}. In this process, the Ca^{2+}-binding C2 domains of Syt-1 interact with soluble N-ethylmaleimide-sensitive-factor attachment receptor (SNARE) proteins and lipids of the SV and plasma membranes. However, molecular details of these interactions are unclear and many aspects of the molecular mechanism underlying Syt-1 function are still under debate.

In the first part of this work, the lipid preferences of Syt-1 were experimentally characterised using native mass spectrometry (MS), a method particularly suited to the analysis of non-covalent complexes, as well as two complementary biophysical techniques. Full-length Syt-1 and truncated soluble variants containing the C2A domain, the C2B domain or both (C2AB) were recombinantly expressed in *Escherichia coli* (*E. coli*), purified and characterised using native MS. Syt-1 existed predominantly as a monomer and a low degree of dimerisation was observed only for C2AB. To investigate the interaction of Syt-1 with membrane lipids, a native MS assay to determine binding of lipids to soluble proteins was established using various detergents and standard proteins. The detergent octyl tetraethylene glycol ether (C8E4) used in this assay reduced the average charge of the protein ions and allowed the analysis of protein-lipid complexes by native MS. The assay was applied to the soluble Syt-1 variants, revealing that C2AB preferentially interacts with multiply phosphorylated phosphatidylinositol phosphates (PIPs), whereas the C2A or C2B variants preferred binding to phosphatidylserine (PS) or phosphatidylinositol (PI). To obtain complementary information on the interaction of Syt-1 with membrane lipids, we also examined (1) the binding of Syt-1 to immobilised lipids and (2) the interaction of C2A and C2B with intact membranes by liposome density gradient centrifugation.

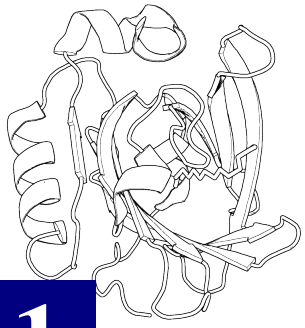
The experimental methods were complemented by a 900 ns molecular dynamics (MD) simulation providing a spatial and temporal resolution inaccessible to the experimental methods used before. In more detail, the interactions of C2A and C2B with SV and plasma membrane models identical

to those employed for liposome density gradient centrifugation were simulated in the presence and absence of Ca^{2+}. The simulations provided insights into the lipid preference of Syt-1, its loop dynamics, membrane interaction angles and membrane insertion and showed that C2A and C2B interact differently with lipid membranes. C2A preferentially binds to the SV membrane whereas C2B preferentially interacts with the plasma membrane. The presence of Ca^{2+} increases these membrane affinities and also leads to a shift in the membrane interaction angle for C2A. Overall, the experiments and simulations suggest a central role for negatively charged phospholipids in the membrane interaction of Syt-1.

In the second part of this work, the structural effects of RNA binding to p38 were analysed. p38 plays a dual role, forming the viral capsid but also acting as small interfering RNA (siRNA) scavenger to impede host defence. Scavenging occurs in the soluble form of p38, however, most structural studies of p38 to date have focused on the protein in its capsid form.

Therefore, the oligomerisation of p38 in the presence and absence of siRNA of different lengths was investigated using native MS. In solution, p38 forms dimers that readily exchange subunits, however, subunit exchange is impaired in the presence of RNA, possibly due to the binding of RNA to p38. Chemical cross-linking of p38 with molecules of different lengths in the presence and absence of RNA and at different temperatures was used to gain information about the structure of soluble p38. The identified cross-links together with computational docking of p38 domains and molecular dynamics simulations of the resulting models indicate a compact structure of p38 in solution. However, the existence of a highly flexible ensemble of p38 conformers cannot be excluded.

In summary, this thesis contributes to the general understanding of protein interactions with lipids and RNA and highlights the capacity of structural MS in combination with MD simulations for the study of protein interactions.



1 Introduction

The large variety of living organisms is based on interactions of five key molecule classes. (1) Deoxyribonucleic acid (DNA) serves as the central storage molecule for genetic information. (2) RNA acts as carrier for genetic information in form of messenger RNA and fulfils multiple functions as non-coding RNA. (3) Lipid membranes confine reactions within a cell and maintain defined reaction conditions differing from the surrounding medium. Furthermore, lipids act as highly efficient energy storage molecules. (4) Carbohydrates play a role in energy metabolism, act as structural components in cell walls and enable specific cell-cell recognition as glycosylations of lipids and proteins. (5) Proteins orchestrate the interactions of all molecule classes as they fulfil regulatory, signalling, structural and transport functions within the cell. Due to their central role in cellular metabolism, proteins are required to selectively recognise a large variety of interaction partners.

1.1 Proteins Engage in a Variety of Interactions

Proteins are involved in the majority of processes in living cells. Despite the limited building block repertoire of the twenty proteinogenic amino acids, this molecule class has adapted to diverse tasks ranging from catalysis of chemical reactions to the formation of membrane channels. This diversity requires selective interactions within a large span of molecule sizes. Large molecules such as structurally diverse messenger RNAs (mRNAs) in the nucleus are exported by a 28 subunit multiprotein complex that requires assembly in a defined and reproducible manner [1]. On the contrary, small molecules such as individual potassium ions are recognised and discriminated from slightly different sodium ions by specific channel proteins [2].

Part of the high diversity of protein interactions can be attributed to the existence of protein domains, parts of a protein chain that independently fold into defined three-dimensional structures [3, 4]. Proteins consist of a single domain (e.g. haemoglobin) or a combination of multiple domains inter-spaced with linker regions (e.g. pyruvate kinase). Moreover, discontinuous domains that are formed from distinct parts of the protein primary sequence evolved by merging of previously autonomous domains. Protein domains function as modular structures and building blocks for interaction networks through genetic recombination or duplication of the genetic material for intact

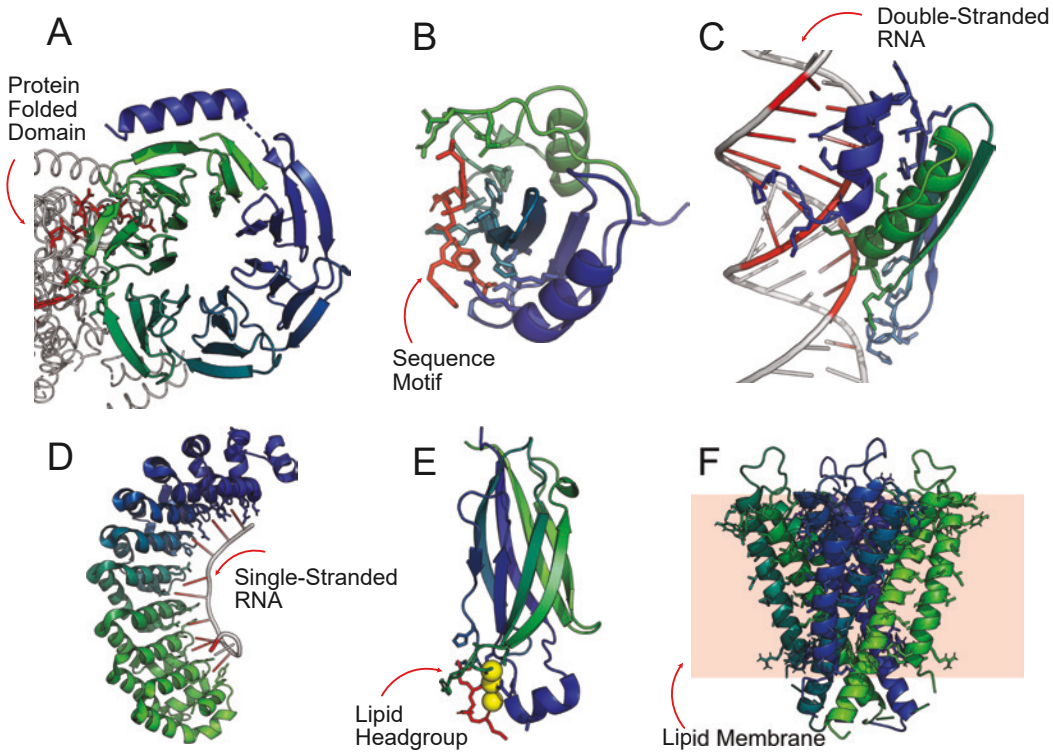


Figure 1: Protein Interaction Domains. Protein sequences are coloured from N- (blue) to C-terminus (green). Protein residues within 3.5 Å of interacting atoms (red) are shown as stick representation. Calcium ions (yellow) are shown as spheres. **(A)** WD40 domain of THO complex subunit 3 (PDB 7APK). **(B)** SH2 domain of Grb2 (PDB 1JYQ). **(C)** dsRBD of ADAR2 (PDB 2L2K). **(D)** PUF domain for cytosine recognition (PDB 2YJY). **(E)** C2 domain of phospholipase A2 bound to phosphatidylcholine (PDB 6IEJ). **(F)** Potassium channel KcsA (PDB 4LBE). The interaction surface of KcsA was obtained from the MemProtMD database [5].

domains. For example, one copy of a duplicated gene for a domain retains its original function (and that of the protein), whereas the second copy of the domain accumulates genetic mutations that eventually lead to new interaction specificity.

Additional diversity in protein interaction is added by post-translational modifications [6]. Proteins are, for example, modified by glycosylation of asparagine, serine, threonine, or their Carboxy-terminus, thereby changing stability and solubility. Phosphate groups are attached to serine, threonine, and tyrosine residues by kinases, introducing a negative charge that influences protein structure, activity and interactions with other proteins. Other common modifications include acylation that mediates membrane interactions of proteins and ubiquitination that renders proteins susceptible to proteasomal degradation. As post-translational modifications are dynamically changed during the life of a protein, they allow temporal fine-tuning and regulation of interactions.

1.1.1 Protein-Protein Interactions and Protein Oligomerisation

Protein domains enable proteins to interaction with a diverse set of different molecules. Furthermore, protein diversity is increased by complex formation between multiple copies of either the same or different proteins. This allows spatial segregation of proteins involved in a common reaction and is the basis of large multi-protein complexes such as molecular pumps (e.g. vesicular-type adenosinetriphosphatases [7]) or enzyme super-complexes (e.g. the pyruvate dehydrogenase complex [8]). Protein complexes also increase the effectiveness of each protein interacting with its target e.g. by favouring a certain orientation of the target. Thereby, protein complexes modulate interactions and provide local reaction environments that enable otherwise unfavourable reactions.

Protein-protein interactions include (1) domain-domain interactions, in which two folded domains share an interaction interface, (2) domain-motif interactions that involve a folded domain interacting with a linear amino acid sequence and (3) unstructured complexes. Domain-domain interactions usually mediate stable complexes. For example, WD40 domains form interaction hubs in many eukaryotic complexes such as in the transcription-export complex (**Figure 1A**) [1], immunoglobulin domains are involved in the formation of antibodies [9] and furin-like domains are involved in dimerisation and auto-phosphorylation of epidermal growth factor receptor [10]. In contrast, complexes mediated by domain-motif interactions are generally transient and, therefore, often found in signalling cascades. For example, the *Src homology 2* domain of the protein Grb-2 binds to phosphotyrosine residues of the epidermal growth factor receptor (**Figure 1B**) [11]. The SH3 domain of Grb-2 subsequently binds to polyproline rich motifs in Sos thereby transmitting the receptor signal [12]. A third class of protein interactions involves intrinsically disordered regions (IDRs) that lack a stable tertiary structure but nevertheless engage in complex formation. Protein-Protein interactions of IDRs include domain-motif interactions with folded domains, structure formation during the interaction (structure upon binding) and completely unstructured „fuzzy complexes“ that are often based on multivalent binding governed by electrostatic interactions [13].

1.1.2 Protein-RNA Interactions

Protein-RNA interactions involve binding to coding mRNA and non-coding RNA [14]. Protein binding to mRNA is important during localisation, translation, degradation, (de-)stabilisation and editing of mRNA sequences [15, 16]. Protein-RNA complexes involving non-coding small RNA play a major role in RNA-mediated silencing, a process in which protein-RNA silencing complexes perform gene-specific degradation of mRNA by base-pairing of the small RNA with the target mRNA [17]. Interestingly, a functional interaction between proteins and RNA also has an effect on the protein as it reduces for example protein aggregation [18].

RNA-binding domains distinguish between double-stranded RNA (dsRNA) and single-stranded RNA (ssRNA) with various levels of sequence specificity, partly through the formation of complex three-dimensional RNA-folds that are specifically recognised. However, most RNA-binding proteins preferentially bind to ssRNA in a sequence-independent manner [15]. Examples for RNA-binding protein domains include the dsRNA-binding domain without sequence specificity (**Figure 1C**) [19], RNA recognition motif-domains that bind to ssRNA with low sequence specificity [20], and *Pumilio/fem-3*-domains that bind to a specific consensus sequence in ssRNA (**Figure 1D**) [21].

In addition to folded interaction domains, IDRs play a role in RNA-binding e.g. by disorder-to-order transitions [22]. Furthermore, dynamic protein-RNA interactions involving IDRs play important roles in liquid-liquid phase-separation of ribonucleoprotein granules or in cellular signalling [23].

1.1.3 Protein-Lipid Interactions

The formation [24], maintenance [25] and degradation [26] of biological membranes is mediated through proteins. Integral membrane proteins are inserted into the lipid membrane and, therefore, restricted to the two-dimensional space of their membrane. Their functions involve the regulation of proton permeability [7], salt ions [2] or small molecules [27] across the membrane. The lipid-binding interfaces of integral membrane proteins are usually complex and cannot be classified into defined domain topologies [28]. Furthermore, lipid interactions of an integral membrane protein include specific binding of certain lipids but also involves contacts with the bulk lipid content of the membrane (**Figure 1E**). Whereas binding of structurally important lipids to integral membrane proteins can be resolved experimentally [29], insights into interactions with less tightly bound lipids usually requires computational simulations [5].

In contrast, peripheral membrane proteins are soluble proteins that temporarily bind to membranes [30]. They can influence a membrane, for example, by vesicle formation on the endoplasmic reticulum [31], lipid transport between membranes [32], and the selective degradation of lipids as part of signalling events [33]. Peripheral membrane proteins often contain lipid interaction domains that interact with lipid membranes through either hydrophobic interactions or by binding to anionic lipid headgroups through electrostatic interactions [30]. Some lipid interaction domains recognise a specific lipid type, whereas others bind to membranes with a certain membrane curvature. Several lipid-binding domains bind only to their target lipids when certain additional criteria, such as the binding of a signalling molecule, are met. This coincidence detection allows tight spatio-temporal control of the protein activity [34]. Examples of lipid-binding domains are C2 domains that bind membranes in a Ca^{2+}-dependent manner (**Figure 1F**) [35], *pleckstrin homology* domains that interact with phosphoinositides [36] or *bin, amphiphysin and Rvs* domains that bind nonspecifically to acidic membrane surfaces and sense or induce membrane curvature [37].

1.2 The Role of Synaptotagmin-1 in the Neuronal Synapse

The nervous system relies on neurotransmission, the signal transmission between neurons. During neurotransmission action potentials that are propagated within a neuron are translated into the exocytosis of neurotransmitters at the synapse, the interface between two connecting neurons (**Figure 2**). For this, neurotransmitters are stored inside SVs that undergo various priming and docking steps before they are ready for fusion [38]. After arrival of an action potential, influx of Ca^{2+} leads to the activation of a specialised protein machinery and to fusion of the SVs with the presynaptic membrane. Neurotransmitters are released into the synaptic cleft and recognised on the post-synaptic neuron by specialised receptors. Neurotransmitters can promote or reduce the formation of an action potential in the postsynaptic neuron, depending on the neurotransmitter. Eventually, fused vesicles are recycled through endocytosis and again loaded with neurotransmitters.

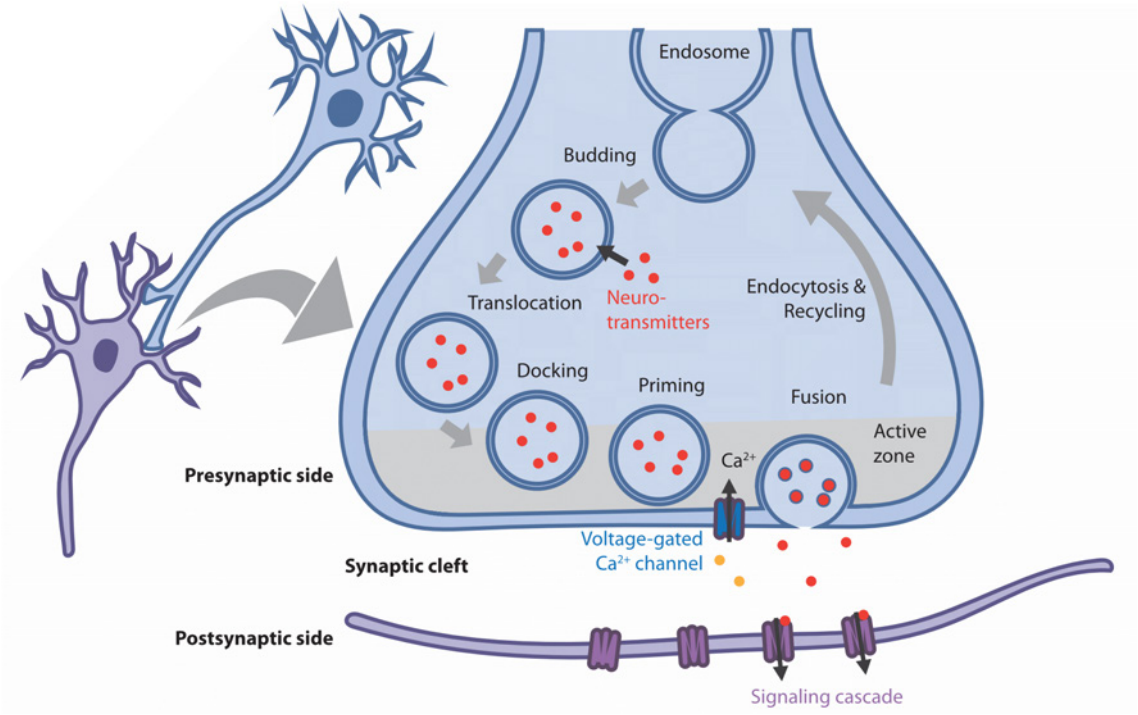


Figure 2: Synaptic Vesicle Exocytosis and Recycling. Synaptic vesicles are formed through recycling of endocytosed vesicles or by budding from the endosome, loaded with neurotransmitters and undergo a series of docking and priming steps on the presynaptic membrane. After Ca^{2+} influx, vesicles fuse with the presynaptic membrane and release neurotransmitters that activate the postsynaptic neuron. Vesicles are recycled through endocytosis.

1.2.1 Structural Changes in the SNARE Complex Drive Membrane Fusion

Fusion of two separated membranes involves at least three distinct steps [39]. An isolated membrane consists of two lipid leaflets that are in contact through their hydrophobic fatty acyl chains, whereas their polar head groups are exposed to the aqueous environment. During fusion, membranes first come into close contact and form local contact sites such as membrane protrusions at which both membranes are separated by only a thin hydration layer (**Figure 3A**). After overcoming this energy barrier, the membranes form an intermediate hemifusion structure in which the outer leaflets of both membranes have merged but the inner leaflets remain separated. The fusion pore eventually forms either directly from the hemifusion stalk, or after expansion of the stalk to a hemifusion diaphragm which again collapses into the fusion pore [39]. The mechanical force required for SV fusion at the presynaptic membrane is provided by structural changes of the SNARE complex that translate into bending of the two membranes. The SNARE complex is a universal membrane fusion machinery that is expressed in various organisms from yeast to humans. It mediates fusions of large vesicles fusing cell organelles as well as the exocytosis of small SVs. The SNARE complex consists of Syntaxin-1A (Stx-1A) and Synaptosomal-Associated Protein of 25 kDa (SNAP-25) residing on the presynaptic membrane, whereas Synaptobrevin-2 (Syb-2) is anchored to the SV (**Figure 3B**). Together they form a short four helix bundle with two helices contributed by SNAP-25 and one helix contributed by Syb-2 and Stx-1A [40, 41]. During membrane fusion, this bundle progressively extends towards the C-terminus of the SNARE proteins (zippering), thereby pulling the associated membranes proximity and leading to a hemifusion state [42, 43, 41]. Additional force for opening the initial fusion pore is gained from large SNARE-associated protein complexes that sterically drive the hemifusion structure towards pore opening and efficient release of neurotransmitters [44]. Energy for pore expansion is likely provided by relaxation of the high membrane curvature of the SVs [44].

Various proteins are involved in the regulation of the SNARE complex. Among them, Syt-1 and Complexin-1 (Cpx-1) control the zippering of the SNARE complex in response to Ca^{2+}. The function of Cpx-1 remains controversial as fusion-enhancing and fusion-reducing properties have been proposed for this protein, depending on the species under investigation [45]. According to one model of SNARE regulation, Syt-1 and Cpx-1 bind simultaneously to the SNARE complex (**Figure 3B**). In the resting state the accessory helix of Cpx-1 prevents full zippering of the four helix bundle [45]. Only upon Ca^{2+} influx Syt-1 dissociates from the complex and the accessory helix shifts away from the fusion pore enabling full zippering of the complex. Thereby, Cpx-1 acts as cofactor for the Syt-1-induced zippering of the SNARE complex.

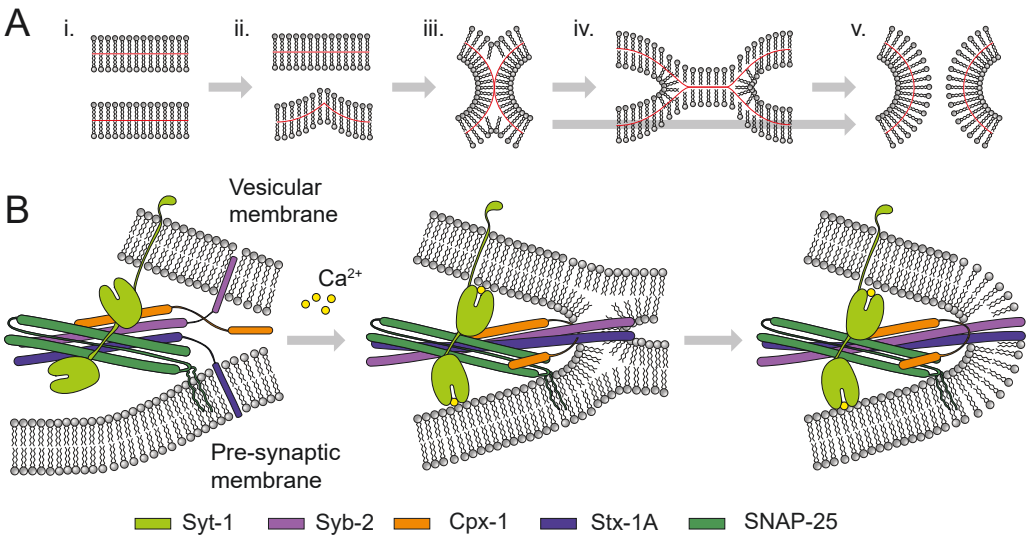


Figure 3: SNARE-Induced Membrane Fusion. **(A)** During membrane fusion, membranes come into close contact (i), form initial contacts (ii) and the hemifusion stalk (iii). The hemifusion stalk expands into the hemifusion diaphragm (iv) or directly forms the fusion pore (v). **(B)** The prefusion SNARE complex is bound to the C2B domain of Syt-1 while Cpx-1 locks the SNARE complex in a partially zipped state (left). Ca^{2+} binds to the Syt-1 C2 domains, compensates the negative charge of the Ca^{2+} binding loops (CBLs) and allows interaction of Syt-1 with lipid membranes (middle). Full zippering of the SNARE complex leads to hemifusion and complete fusion of SV membrane with the presynaptic membrane (right).

1.2.2 The Calcium Sensor Protein Synaptotagmin-1

Multiple synaptotagmin isoforms are expressed in the human brain in different regions correlating with different neurotransmitter release kinetics [46]. This thesis focuses on Syt-1, an isoform that shows intermediate release kinetics and is expressed in the anterior region of the mammalian brain. Syt-1 consists of two cytoplasmic C2 domains (termed C2A and C2B) connected by a flexible linker. C2A is anchored to the SV membrane via an IDR linker connected to a transmembrane helix. This IDR has been suggested to act as a sensor for the dielectric environment of the protein [47]. Electrostatic effects are not only involved in regulation of the IDR, but likely also play a role for membrane interactions of the C2 domains. The electrostatic surface potentials of both domains show large patches of opposed potentials [48, 49]. Calcium-binding to C2A and C2B occurs through negatively charged aspartate residues in the CBLs that coordinate two and three Ca^{2+} ions, respectively [48–50]. Furthermore, a lysine-rich, polybasic region close to the CBL of C2B interacts with anionic phospholipids or the SNARE complex [51–54]. A third important

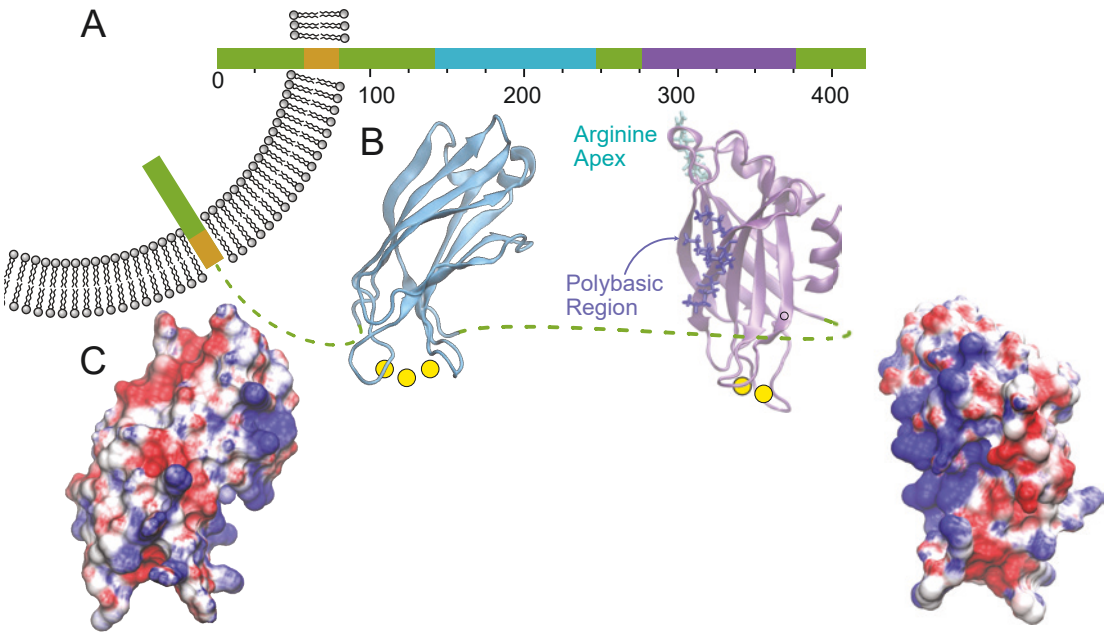


Figure 4: Overview of Syt-1 Structure. **(A)** Schematic linear representation of *Rattus norvegicus* Syt-1 sequence including of a transmembrane domain (orange), the C2A (cyan) and C2B (purple) domains. **(B)** Cartoon representations of C2A (PDB 1BYN) and C2B (PDB 1K5W). Bound number of Ca^{2+} ions (yellow) and C2B arginine apex and polybasic region are highlighted. **(C)** Surface representations of C2A (left) and C2B (right) coloured by surface potentials separated at ±5 kT/e.

interaction site involves two positively charged arginine residues located opposite the CBL of C2B, the arginine apex (AA), that has been proposed to interact with either the SNARE complex or with membranes during fusion pore expansion [55, 56].

Based on these potential interaction sites, various models for Syt-1 function during SV fusion have been proposed (reviewed in [57]). These models can be divided into models with emphasis on the Syt-1-SNARE interaction (SNARE-centric models) and models with focus on the interaction between Syt-1 and the lipid membranes (membrane-centric models).

SNARE-centric Models *Cpx-1 arrests the SNARE complex in a semi-zipped state and binding of Ca^{2+} to Syt-1 leads to displacement of (part of) Cpx-1 and zippering of the SNARE complex.* This model is based on experiment showing simultaneous binding of Syt-1 and Cpx-1 to the SNARE complex. For example, X-ray crystallography studies indicated binding of multiple Syt-1 molecules to the SNARE complex through two separate Syt-1/SNARE interfaces in the presence of Cpx-1 [58]. Binding of Ca^{2+} to Syt-1 then causes structural rearrangements and eventually SNARE by displacement of Cpx-1 from the SNARE complex. A combined nuclear magnetic resonance and MD study reported that Syt-1 C2B binds to Stx-1A and SNAP-25 via the polybasic region and to

the lipid membrane through the AA [59]. According to this study, C2B binds to the SV membrane through its CBLs after Ca^{2+} influx. Thereby, C2B simultaneously interacts with the SV and plasma membranes causes displacement or loss of structure of the accessory helix of Cpx-1, that prevents SNARE fusion. However, the importance of these interactions between Syt-1 and the SNARE *in vivo* is under debate. Several studies reported that the interactions disappear at physiological ion strengths [60, 54], whereas another study reported in favour of Syt-1-SNARE interactions that the AA of C2B interacts with the SNARE complex under the same conditions [61]. In contrast, Cpx-1 has a high affinity for the SNARE complex [60].

Syt-1 directly arrests SNARE complex zippering in the prefusion state. Ca^{2+}-binding then relocates Syt-1 to a membrane and thereby enables fusion. In support of this model, spontaneous neurotransmitter release increased in the absence of Syt-1 [62]. However, this model also depends on the existence of the Syt-1/SNARE interaction at physiological ion strength [60, 54, 61].

Membrane-centric Models *Syt-1 changes membrane properties by interacting with membrane phospholipids and thereby lowers the energy barrier for SNARE-mediated fusion.* The interaction of Syt-1 with negatively charged anionic phospholipids in the presence of Ca^{2+} is the basis for this hypothesis [54, 63–67]. Ca^{2+} coordination by the CBL drastically changes the electrostatic potential of Syt-1, reduces the electrostatic repulsion between the plasma membrane and the C2 domain and enables binding to membrane lipids such as PS through completion of the Ca^{2+} coordination sphere by the lipid head group [50, 68]. The presence of polyanionic lipids such as phosphatidylinositol-4,5-bisphosphate ($PI(4,5)P_2$) further increases the binding of Syt-1 to lipid membranes [69–72] most likely through binding to the polybasic region of Syt-1 [51, 53, 73, 52, 54, 71]. This suggests a multivalent binding mode for C2B in which the CBL inserts into the plasma membrane and the polybasic stretch simultaneously interacts with membrane lipids [54]. Destabilisation of the lipid membrane by Syt-1 is supported by studies showing the insertion of the CBLs of C2B into the inner leaflet of the plasma membrane [66, 74, 54]. Another indication membrane destabilisation by Syt-1 is the observation that C2B locally increases the concentration of PS and thereby induces positive curvature [75]. Furthermore, negative stain electron microscopy Syt-1 C2B together with liposomes showed the formation of tubules [74]. However, this result was not reproduced in later study using cryo electron microscopy [76].

Syt-1 tethers SVs to the presynaptic membrane by simultaneously interacting with both the SV and the plasma membrane through both of its C2 domains. In the absence of Ca^{2+}, negative charges on the C2 domains prevent close contact of the two membranes. This model is supported by both C2 domains face in opposing directions in the presence of membranes as required for binding to different membranes [77]. Furthermore, Syt-1 requires an asymmetric PS distribution to

effectively tether liposomes [78]. However, the presence of physiological concentrations of ATP and Mg^{2+} abolished binding of Syt-1 C2AB to liposomes mimicking the SV membrane supposedly by competing for the coordination of the Ca^{2+} ion [69, 79].

Syt-1 is the initial SV tethering protein and SNARE complex assembly occurs only after Ca^{2+}-binding. By binding to PI(4,5)P_2 via its polybasic region, Syt-1 locates SVs in proximity to their fusion site. Repulsion of negatively charged SVs and the plasma membrane prevents annealing and formation of the SNARE complex. Ca^{2+}-binding to Syt-1 is required to compensate the negative charge and allow the SNARE complex to assemble. This model is supported by experiments showing that Syt-1-tethered liposomes are 8 nm apart in the absence of Ca^{2+} but significantly anneal in its presence [80].

As outlined above, various models for Syt-1 function coexist to date. They can be classified as SNARE-centric or focusing on the membrane interaction of Syt-1. Whereas the SNARE-centric models depend on whether Syt-1 SNARE interactions exist *in vivo*, the importance of Syt-1-membrane contacts is generally accepted. Nevertheless, detailed functions of these interactions remain still unclear and therefore new and complementary methods for the analysis of protein-lipid interactions are needed.

1.3 p38: A Viral Swiss-Army Knife

A second class of protein interactions investigated in this thesis are protein-RNA interactions. A biological process in which RNA recognition plays a key role is the viral invasion of plant host cells and its detection by the host. Viral RNA thereby functions as a signature molecule for the host defence and various mechanisms to impede RNA recognition or the triggered defence have emerged in viruses. One example is the coat protein p38 of TCV that functions as a viral counter-defence.

1.3.1 A Virus-Host Arms Race

RNA silencing is central to genetic regulation in most eukaryotes. Its evolutionary origin likely lies in the defence against foreign nucleic acids such as viruses and transposable elements [81]. This function is still present in recent plants. Replication of RNA viruses, the largest class of plant viruses, involves (transient) synthesis of dsRNA molecules. Therefore, dsRNAs are a main target for the identification of an infection. dsRNAs are recognised by cellular Dicer-like proteins (DCLs) and processed into short siRNAs (**Figure 5**). In *Arabidopsis thaliana*, four DCLs are present of which DCL2, DCL3 and DCL4 are involved in the production of siRNAs of 22 nt, 24 nt and 21 nt size, respectively [82]. Subsequently, siRNA duplexes are loaded onto Argonaute proteins (AGOs) to form a RNA-induced silencing complex (RISC) [83]. During maturation of RISC, only a single

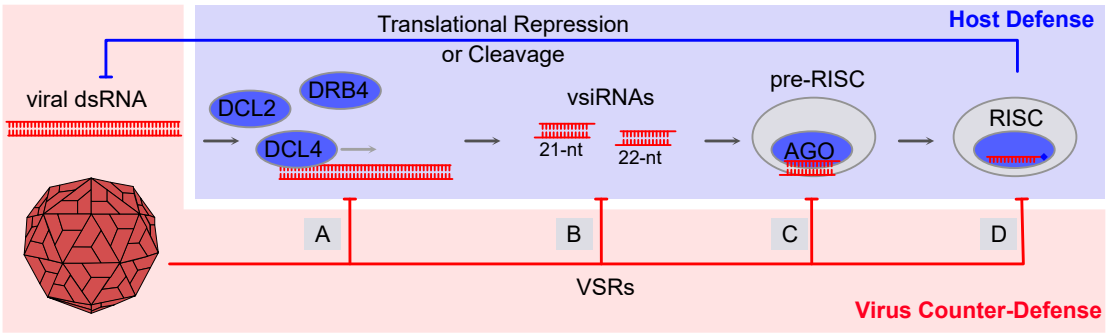


Figure 5: RNA Silencing and RNA Silencing Suppression. Viral dsRNA is recognised by DCLs assisted by double-stranded RNA-binding protein 4 and processed into short long siRNAs. These are loaded onto AGOs that are part of pre-RISCs. After maturation by retention and methylation of the siRNA guide strand, mature RISC target viral dsRNA for translation inhibition or degradation. VSRs target all parts of RNA interference: **(A)** Prevention dsRNA cleavage by DCLs; **(B)** Binding and sequestering of siRNAs; **(C)** Degradation of AGO; **(D)** Preventing of siRNA methylation and RISC maturation.

siRNA guide-strand is retained and protected against degradation by methylation [84]. Base pair complementarity of the guide-strand eventually enables the mature RISC to recognise target RNAs and mediate their selective repression by endonucleolytic cleavage or translational repression [85]. Thereby, RNA silencing in plants acts as a highly specific inhibitor of viral replication without triggering harmful defence system such as local necrosis.

In adaption to RNA silencing-mediated host defence, various viral suppressors of RNA silencing (VSRs) that attenuate or inhibit host defence have evolved independently in diverse virus genera [86]. Viral effectors for most steps in the RNA silencing pathway have been described. For example, the processing of dsRNA into siRNAs is reduced by the synthesis of viral decoy RNA that overloads the capacity of the cellular enzyme [87] (**Figure 5A**). In other viruses, the loading of RISC is prevented by sequestering processed siRNAs [88] (**Figure 5B**) or by degradation of AGO [89] (**Figure 5C**). The stability of the mature RISC is reduced by some VSRs e.g. by preventing methylation of the guide-strand RNA, leading to degradation of the guide strand [90] (**Figure 5D**). Selection pressure for resistance against VSRs has led to additional lines of defence for plant hosts such as permanent targeting of specific host genes by RNA interference [17]. If VSRs reduce RNA interference, gene products of the targeted host genes accumulate and mediate stronger (and more costly) defence mechanisms.

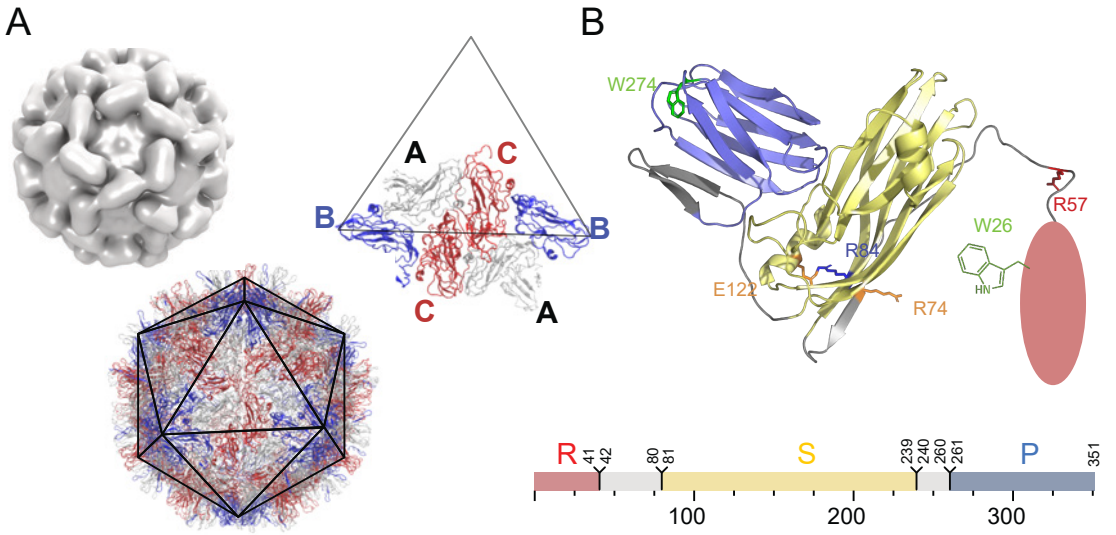


Figure 6: Structural Models of p38. **(A)** Cryo electron density map (top left) and structural model of the TCV capsid structure (bottom). p38 exists in three conformers, A (blue), B (white) and C (red) and forms A/B and C/C dimers in the capsid structure (right). **(B)** Crystal structure of a single p38 molecule (PDB 3ZXA). The structured S- (yellow) and P-domain (blue) are coloured in the model and the sequence representation (bottom). The flexible R-domain (red, indicated by circle) was not resolved by crystallography. Amino acids important for binding to AGO (W26 and W274, green) [89], dsRNA processing (R84, blue) [91], homo-dimerisation (R74 and E122, orange) [92, 89] and RISC loading (R57, red) [91] are shown.

1.3.2 Structure and Function of Turnip Crinkle Virus p38

Viral genomes are exposed to two opposing evolutionary pressures. Adaptation to host defence strategies by diversifying and enlarging their genome and maintaining a small genome size for fast reproduction. Therefore, viruses which combined multiple functionalities in a single protein had an evolutionary advantage. The TCV coat protein p38 is such a multi-functional protein. In addition to forming the icosahedral viral capsid [93, 94], it is a potent VSR. p38 interacts with cellular AGOs with two GW-motifs at its N- and C-terminal ends [89]. Thereby, p38 mimics cellular GW-containing proteins that engage in formation of the RISC and prevents RISC assembly. p38 suppresses the function of DCL4 which produces 21 nts siRNAs [92]. DCL4 function is to some degree rescued by DCL2 which produces 22 nts siRNAs. However, 22 nts siRNA levels were also reduced in the presence of p38. Likely, p38 interacts with long dsRNA as well as with double-stranded siRNA and, thereby, the protein prevents either dsRNA cleavage by DCL and/or sequesters cleaved siRNAs before incorporation into RISC [88]. Moreover, p38 likely acquired the ability to bind long dsRNA and inhibit DCL independently of siRNA-binding and RISC inhibition [91].

The functional diversity of this VSR is, however, not reflected by its structural characterisation. An initial structural model from X-ray crystallography [93] revealed the folded shell (S)- and projecting (P)-domains of p38 (**Figure 6**) but the flexible N-terminus of the protein was not resolved. The 180 copies of p38 that form the capsid were in three different conformers, namely A, B and C. These conformers are organised in 60 A/B and 30 C/C dimers by non-covalent interactions between the P-domains of each protein pair. The C/C dimers were suggested to form the β-annulus, a three-fold symmetric β-sheet formed by the N-terminus of p38. A more recent cryo electron microscopy study shed further light on the inner structure of the virion [94]. Three shells of electron densities corresponding to (1) the outer coat formed by the S- and P-domains, (2) a middle shell forming a cage-like structure and likely corresponding to the structured N-terminal regions of C/C dimers and (3) a central region with low electron density indicating a low occupancy or high structural dynamics of its components were observed. Based on the cryo electron microscopy analysis, a potential mechanism of capsid opening after host invasion was proposed. By entering the cell, a virus particle loses Ca^{2+} ions that were previously coordinated by negatively charged side chains between the protein subunits. This leads to electrostatic repulsion between the side chains and the capsid expanding by translation of the coat proteins. Thereby, pores are formed through which ssRNA leaves the capsid and is translated by host ribosomes. The authors suggest proteolytic degradation of p38 to play a role in disassembly of the capsid after cell entry, however, their analysis did not provide any further details on the fate of p38 after cell invasion. Therefore, despite a number of biochemical analyses suggesting p38 to be an active VSR, its structure in solution remains still unclear.

1.4 Aim of This Study

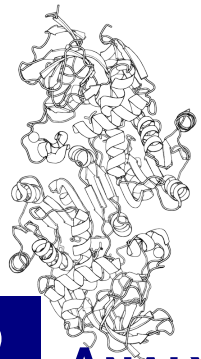
Protein interactions are central to the organisation and regulation of biochemical processes in living cells. In this thesis, protein-lipid interactions of the neuronal regulatory protein Syt-1 and protein-RNA interactions of the viral capsid protein p38 will be interrogated.

Lipid preferences of Syt-1 will be characterised experimentally using native MS, a technique that so far has not been applied to Syt-1 lipid interactions, of Syt-1 in the absence and presence of various lipids. For this, a lipid interaction assay for soluble proteins using native MS will be established and tested. Effects influencing lipid binding to soluble proteins during electrospray ionisation will be investigated and discussed. Complementary biophysical techniques such as binding of Syt-1 to immobilised lipids and density gradient centrifugation of Syt-1 in the presence of liposomes mimicking the SV and plasma membranes will be used to gain orthogonal information on Syt-1 lipid preferences. Molecular effects of Syt-1 membrane interactions will be investigated using computational simulations of Syt-1 interacting with the SV and plasma membranes in the presence and absence of Ca^{2+}. Detailed analysis of the simulated trajectories will suggest possible

lipid-interactions of Syt-1 during its initial interaction with a membrane and their effects on, for example, protein orientation. Through comparison of experimental results and computational simulation, molecular mechanisms for Syt-1-membrane interactions will be discussed.

In the second part of this thesis, native MS and chemical cross-linking will be applied to the soluble form of TCV p38 binding to RNA. In contrast to the structurally well characterised capsid form, the soluble form is likely very flexible and therefore inaccessible to e.g. x-ray crystallography. Native MS of isotopically labelled variants of p38 will be used to determine the oligomeric state of p38 in the presence and absence of siRNAs of different lengths and to interrogate the dynamic assembly and disassembly of p38 oligomers. Using chemical cross-linking at different temperatures and in the presence and absence of siRNA, conformational changes of p38 in response to these conditions will be characterised. The resulting cross-links will be used to generate structural models of p38 that will be compared with the structural organisation of p38 in the capsid form. The stability of the models in aqueous solution will be assessed using MD simulation and implications for the infection mechanism of TCV will be discussed.

Both parts of this thesis will contribute to advancing methodologies for the analysis of protein-ligand interactions and deepen the understanding of how ligand interactions mediate protein structure and dynamics.



2 Analysis of Protein Interactions

The previous chapter highlighted the biological importance of protein interactions. Protein-protein recognition in protein complexes such as the SNARE complex or binding of proteins to lipids or RNA play key roles in the biological function of Syt-1 and p38, respectively. To elucidate the those functions, protein interactions have to be determined and characterised experimentally. A common approach to determine protein interactions is protein purification though affinity tags and subsequent identification of co-purified interaction partners. For this, for example a protein of interest is immobilised through recognition by an antibody (co-immunoprecipitation) [95] or specific binding of an affinity tag that was genetically fused to the protein (affinity pull-down) [96]. After removal of low-affinity binders, only strong and therefore likely important complexes are retained and identified by western blot or mass spectrometry (see **Section 2.2**). Similarly, protein interactions involving RNA [97] and lipids [98] are identified.

2.1 Characterisation of Protein Interactions

To understand the role of protein interactions in a cellular context, knowledge of the interacting partners alone is not sufficient. The characterisation of structural, kinetic and thermodynamic aspects of protein interactions relies on various biochemical and biophysical methods as well as on computational tools for hypothesis generation. These include methods with a low spatial resolution such as circular dichroism spectroscopy that yields characteristic spectra indicating secondary structure of a protein [99] and small-angle X-ray scattering that informs on monodispersity of the sample, stoichiometry of a complex and its shape and radius of gyration [100]. In addition, high-resolution structural methods provide detailed information on the molecular organisation of a protein or protein complex. X-ray crystallography provides high-resolution structural information if the protein is amenable to crystallisation and cryo-electron microscopy recently has achieved local resolution comparable to X-ray crystallography without the requirement for crystals [101]. However, high resolution structural methods are often unsuccessful in the structural characterisation of highly flexible and dynamic protein complexes although some degree of structural dynamics can be resolved by cryo-electron microscopy [102]. A notable exception is nuclear magnetic resonance spectroscopy that provides high-resolution models with high temporal resolution, however, is limited

by the size of a protein complex [103]. Static high-resolution structures generated by these methods can serve as direct input for computational simulations that produce dynamic high-resolution models (**Section 2.3**). Moreover, results from techniques with different structural resolution are often complementary and low resolution methods may validate high resolution structures or guide model building [104].

Two structural methods particularly important in this thesis are MS that has matured from a versatile identification tool for small molecules and peptides to a structural biology technique on its own (**Section 2.2**) and MD simulations that generate high resolution dynamic models for comparison with experimental results (**Section 2.3**).

2.2 Mass Spectrometry

MS is a very versatile technique for low to medium resolution structural characterisation of proteins and protein complexes. In a nutshell, MS essentially determines the masses of individual molecules. However, mass cannot be measured directly by MS, instead, analyte molecules are ionised, transferred into the gas-phase and their mass-to-charge (*m/z*) ratio is detected. In contrast to small molecules that tolerate harsh ionisation conditions, the analysis of large biomolecules with MS was greatly facilitated by the invention of soft ionisation techniques in the 1980s [105–107]. Since then, various developments including the orbitrap, a small high-resolution mass analyser [108], and instrument modifications for the transfer of intact protein complexes into the gas-phase [109] have expanded the use of MS in structural biology.

2.2.1 Principles of Mass Spectrometry

The fundamental structure of the mass spectrometer is similar for all types of MS instrumentation. Analyte molecules are ionised in the ion source, enter one or several mass analysers by which they are separated according to their *m/z* ratio and a detector that records the *m/z*. Signals from the detector are computationally processed and the final mass spectrum is calculated.

Ion Sources Soft ionisation is a prerequisite for the analysis of biomolecules and mainly achieved by two techniques, namely matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI). In MALDI, analytes are co-crystallised within a matrix of light-absorbing molecules [106, 107]. Desorption of the matrix molecules by laser irradiation induces ionisation and gas-phase transfer of the analyte molecules that are subsequently carried into the mass spectrometer along an electric potential gradient. MALDI mostly generates singly charged peptide ions thereby simplifying the annotation of the resulting spectra. During ESI, analytes in solution are transferred into a

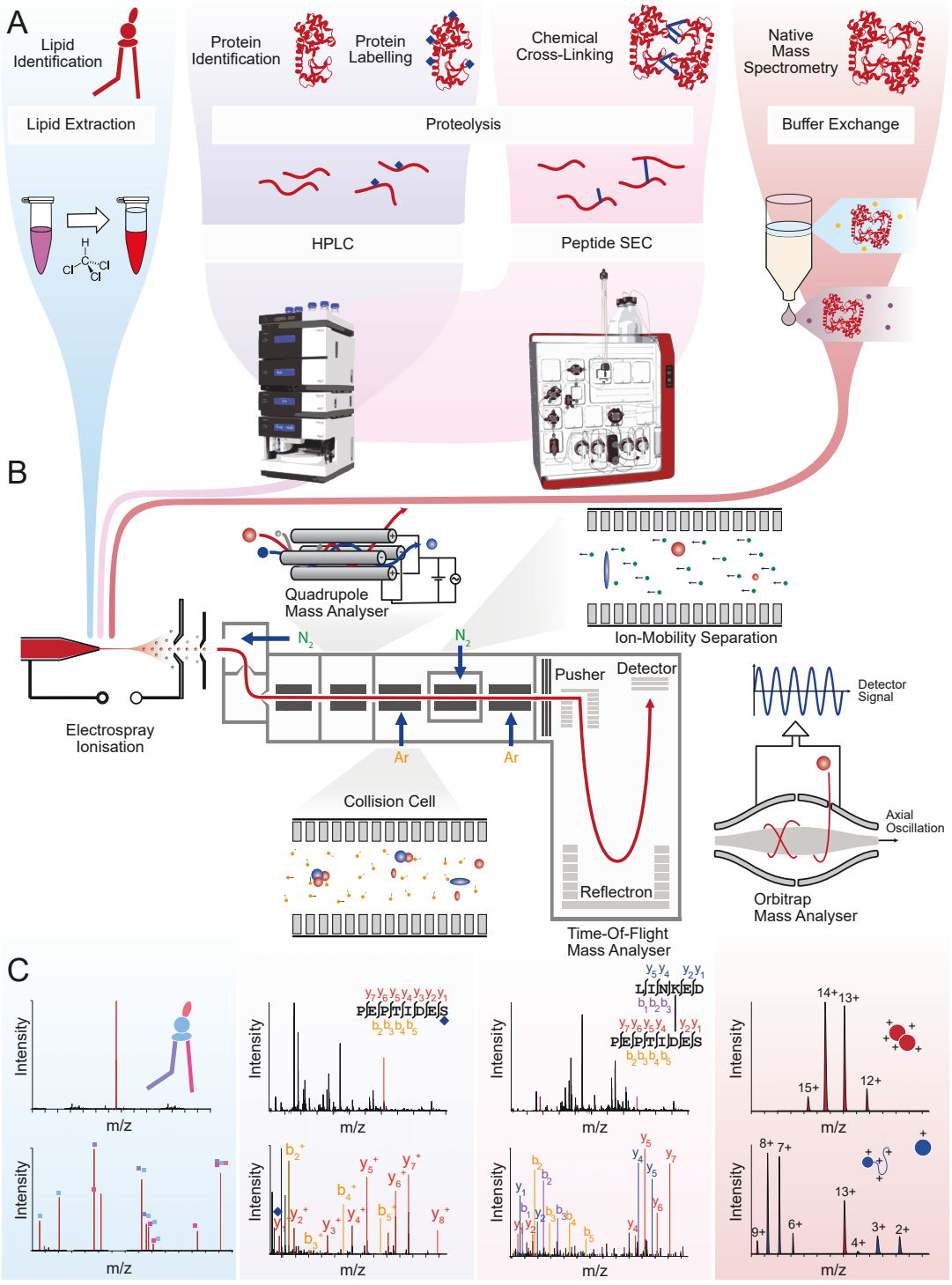


Figure 7: Workflow of Tandem Mass Spectrometry Experiments. Schematic workflows for lipid identification (blue background), identification of proteins and their post-translational modifications (violet), chemical cross-linking (magenta) and native MS (red) are shown. **(A)** Sample preparation: Lipids are extracted using organic solvents.

Figure 7 ***(cont.)*****:** Proteins are enzymatically hydrolysed, disulfide bridges in the resulting peptides are reduced and alkylated and peptide mixtures are separated by high-performance LC. Cross-linked proteins are hydrolysed similarly and cross-linked peptide pairs are enriched for example through peptide SEC. Intact proteins for native MS are transferred into ammonium acetate solution. **(B)** MS analysis. Components of the MS instrument are described along the path of the ions produced by ESI (left). In the quadrupole (centre, top), ions of defined *m/z* are selectively filtered (ion selection mode) or the paths of all ions are stabilised (RF-only mode). In the collision cell (centre, bottom), collisions of analyte ions with inert gas molecules produce intact analyte ions with reduced internal energy at low activation energy or fragment ions at higher energies. Ion mobility separation (right, top) separates ions according to their shape through low-energy collisions with an inert gas. Eventually, ions enter a TOF or orbitrap mass analyser coupled to a detector for recording of the mass spectrum. For details refer to the text. **(C)** Data analysis. Schematic representations of MS1 spectra (top) and fragment spectra (bottom) after ion selection in the quadrupole and collisional activation. Lipids (blue background) are identified based on their intact masses and the masses of characteristic fragment ions arising e.g. from dissociation of fatty acyl chains (purple and magenta) or the lipid head group (red). Proteins (purple background) are identified based on the intact peptide masses and fragment masses after collisional dissociation by comparison with theoretical spectra from a database. Post-translational modifications (blue diamond) are identified by a characteristic mass shift determined during database search. Cross-linked peptide pairs (violet background) require specialised software tools to manage the large size of that database and the higher complexity of the fragment spectra. In native MS (red background), the MS1 spectrum directly indicates the stoichiometry of a protein complex of known components. Collisional activation of a specific ion aids identification by confirming the presence of complex subunits.

conductive capillary at high electric potential with respect to the inlet of the mass spectrometer (**Figure 7B**) [110]. Thereby, droplets emerging at the tip of the capillary carry a net charge depending on instrument polarity and disperse into a fine jet due to charge repulsion on their way towards the instrument. Droplets continuously shrink by solvent evaporation and the surface area of the droplets decreases leading to charge repulsion between ions on the droplet. When charge repulsion is larger than the droplet surface tension, the droplets undergo charge-induced explosion or fission. Eventually, the charged analyte ions are released into the gas-phase and enter the mass spectrometer. As the number of charges residing in a final droplet is subject to stochastic variance, ESI generates a series of charged ions differing by one charge between consecutive peaks of the series. Application of ESI to biological molecules increased with the invention of nanoESI, a miniaturised version of ESI that generates smaller initial droplets [111]. NanoESI allows reduced sample consumption and improved tolerance of salt ions in the solvent and thereby increased sensitivity [112].

Mass Analysers After ionisation, analyte ions enter the mass spectrometer and are directed towards the first mass analyser by electrical ion guides. Mass analysers separate ions according to their *m/z* ratio and are often coupled to detectors that record the presence of ions after separation. Three types of mass analysers are commonly used, namely quadrupole, time-of-flight (TOF) and orbitrap mass analysers. Quadrupole mass analysers consist of four symmetrically oriented rods and every pair of opposing rods is electrically connected. A combination of direct current and alternating current is applied to the two pairs of rods generating an electrical quadrupole field that stabilises the trajectory of ions of a defined *m/z* ratios and thereby acts as a mass filter (**Figure 7B**). By variation of the magnitude of both potentials, ions of different *m/z* values are stabilised. When quadrupoles are used as stand-alone mass analyser, they are usually coupled to a photomultiplier detector that records the impact of the ions at defined *m/z* values. In contrast, quadrupoles operated without the direct current potential (RF-only mode) allow the passage of a broad range of *m/z* values. Quadrupoles are frequently coupled to TOF or orbitrap mass analysers in which they enable tandem MS experiments.

TOF mass analysers are conceptually the simplest type of mass analyser (**Figure 7B**). In these analysers, ions are accelerated by a defined potential and their flight time along a path of defined length is recorded. As the flight time is proportional to the square root of the *m/z* ratio, the *m/z* ratio can be directly calculated from such measurements. Modern TOF mass analysers often contain a reflectron that generates an inverse potential opposing the acceleration region and forces the ions on a parabolic trajectory. This increases resolution in two ways: (1) Velocity differences of ions of the same *m/z* (generated e.g. by spatial dispersion of in the acceleration region) are compensated and (2) the length of the flight path is increased. Similar to quadrupoles, TOF mass analysers are coupled to detectors recording the impact of ions such as photomultipliers.

In contrast to the other mass analysers, the orbitrap acts both as a mass analyser and a detector. In more detail, ions are accelerated laterally into the orbitrap by a specialised ion trap and drawn towards a central electrode by an electric potential (see **Figure 7B**). As a result of the lateral injection, ions do not collide with the central electrode but enter defined orbits around it. Moreover, the electric potential along the axial direction of the central electrode is lowest in the centre of the orbitrap, therefore, the orbiting ions oscillate with a frequency that is inversely proportional to the square root of their *m/z* value. The axially oscillating ions induce a current in the orbitrap that is recorded and converted into a single mass spectrum by Fourier transformation. Resolution of the orbitrap, therefore, is dependent on the time the ions orbit the central electrode, whereas resolution of the TOF detector varies with length of the drift-tube and precision of time measurement.

Tandem MS With very high mass resolution the sum formula, i.e. the composition, of an ion can be deduced from its mass alone. However, the sum formula is insufficient for many applications as for example peptides composed of the same amino acids linked in different sequences or lipids

with different acyl chains can share the same exact mass. Fragmentation of the ions for example by collisional activation in a mass spectrometer enables to discriminate such isomeric ions. For this, ions are transferred into a collision cell filled with inert gas molecules such as nitrogen, argon or helium, the ions are accelerated and collide with the gas molecules. Depending on the pressure in the collision cell and the kinetic energy of the ions, high-energy collisions breaking covalent bonds of the analyte or softer collisions disrupting non-covalent interactions occur. In both situations characteristic fragment ions are generated that ideally enable the identification of the analyte. Tandem MS instruments use collisional dissociation by combining a first mass analyser (typically a quadrupole) with a collision cell followed by a second mass analyser, usually an orbitrap or a TOF mass analyser, the latter followed by a detector. During MS1 acquisition, the second mass analyser/detector pair records *m/z* values of all ions while the quadrupole operates in RF-only mode. In tandem MS or MS2 experiments, the quadrupole is switched to ion selection mode. Thereby, only ions of a narrow *m/z* range corresponding to a parent ion pass the quadrupole, are fragmented in the collision cell and the spectrum recorded after the second mass analyser is characteristic for that parent ion. Tandem MS forms the basis for the identification of proteins and lipids and for the analysis of intact protein complexes by native MS.

Ion Mobility Spectrometry During ion-mobility spectrometry (IMS), analyte ions introduced into an ion mobility cell that contains relatively high pressures of inert gas molecules such as nitrogen so that the mean free path between collisions is short and the acquired energy during each collision is low. In drift tube IMS, ions are dragged through this gas-filled region by a constant potential (**Figure 7B**) [113]. Ions with a large surface area undergo more collisions than smaller and more compact ions and therefore take longer to traverse the cell. In more detail, ions are separated by their rotationally averaged collisional cross-section (i.e. their shape). Using IMS separation, the drift-time for every ion can be recorded in addition to *m/z* and intensity. This additional dimension is used to discriminate conformational and chemical isomers (molecules with identical composition but different structure) or to monitor unfolding of an ion during collisional activation [114, 115]. Travelling wave IMS cells are a second important type of IMS devices [116]. During travelling wave IMS, ions are propelled through the gas-filled tube by electric potential waves and collide with the inert gas molecules. Large ions undergo more collisions and are more likely to be slowed down to such an extent that the maximum potential of the travelling wave moves faster than the ion. After the maximum of the wave has passed, ions are slowed down until the potential gradient of the following wave accelerates the ions again. This process can occur multiple times until the ions eventually leave the IMS cell. Accordingly, smaller ions of similar *m/z* leave the IMS cell before their larger counterparts injected during the same IMS cycle. Compared to drift tube IMS, travelling wave IMS offers increased sensitivity as the ions take less time to traverse the mobility cell.

2.2.2 Mass Spectrometry for Identification

Proteomics The proteome is the entire set of proteins of one or multiple cells at a specific time point, under defined conditions [117]. Accordingly, MS-based proteomics is the study of composition and variability of the proteome. For this, proteins are first prepared for MS analysis e.g. by precipitation with organic solvents [118] (**Figure 7A**). Intact disulfide bridges in a protein are chemically reduced and the sulfhydryl groups are protected. Next, proteins are hydrolysed by treatment with sequence-specific proteases such as trypsin. The use of trypsin as protease has two advantages: (1) The *m/z* range amenable to quadrupole selection is limited and tryptic hydrolysis usually produces peptides corresponding to that limit. (2) Trypsin cleaves exclusively C-terminally of arginine and lysine residues [119] and the presence of basic residues in the resulting peptides improves ionisation in positive ion mode. Furthermore, the high sequence specificity of trypsin increases the reliability of data analysis. Tryptic hydrolysis of for example membrane proteins that contain only few polar residues leads to few peptides that can be identified with MS. In such cases, additional proteases such as chymotrypsin (C-terminal to aromatic amino acids) and GluC (C-terminal to glutamate) are used together with or instead of trypsin. The resulting peptides are separated using reversed-phase liquid chromatography (LC) to reduce complexity of the peptide mixture and thereby improve spectral quality during analysis. To date most proteomics applications use ESI as it allows direct coupling of the mass spectrometer to the chromatography column. Mass spectra are recorded for all eluting peptide ions. Peptide ions with the highest signal intensities are selected in the quadrupole and fragmented in a collision cell. Eventually, the resulting MS1 and MS2 spectra are computationally analysed and proteins are identified (see **Figure 7C**).

It is possible to determine peptide sequences *de novo* from the fragment ions as dissociation mainly occurs at the peptide bond [120]. Two ions generated by fragmentation at consecutive peptide bonds differ in only one amino acid whose identity is reflected in the mass difference between those ions. However, when the putative protein sequences are known (e.g. from genetic sequencing), identification is simplified by searching for the most likely candidate matching the experimental spectra. For this, a protein library is transformed into peptides *in silico* and theoretical fragment ions are calculated for each peptide. By scoring the similarity between experimental and theoretical spectra of peptide and fragment ions and statistical correction for false-positive assignments, even complex protein mixtures are reliably characterised [121]. In the same manner, post-translational modifications such as phosphorylation or acetylation are identified by incorporating the characteristic mass shifts of potential modifications into the database search [6].

Lipidomics The analysis of lipids by MS is complicated by the large diversity of this molecule class. To date, the LIPID MAPS database [122] lists over 40,000 different lipids. In comparison to proteins, chemical properties of lipids vary to a much larger extent. Therefore, a complete

characterisation of all lipids within a sample by MS usually requires the combination of spectra obtained under different conditions (e.g. in the absence and presence of adduct forming salt ions or at different pH values) or instrument parameters (e.g. instrument polarity) [123]. Identification of lipids by MS is often carried out by extraction of lipids using organic solvents (see **Figure 7A**) followed by ESI or MALDI coupled to tandem MS instruments with the capability of ion fragmentation [124]. In case of glycerophospholipids, dissociation of the head group after activation in the mass spectrometer gives rise to fragment ions characteristic for the lipid class (e.g. phosphatidylcholine (PC) vs. PS, see **Figure 7C**). Fragment ions of the same precursor ion matching to the mass of specific fatty acyl chains then determine the full lipid species. Similarly, members of other lipid classes are identified by fragmentation. The complexity of lipid mixtures can be reduced prior to MS analysis using liquid chromatography as done in many proteomics experiments. Moreover, the direct injection of lipids dissolved in organic solvent into the mass spectrometer without chromatographic preseparation (termed "shotgun lipidomics"), is widely applied. It benefits from theoretically unlimited analysis time for any given analyte in contrast to chromatography-based methods that are limited in analysis time by the elution time from the column. However, lipidomics is complicated by several lipid-specific challenges such as ion suppression of low-concentration lipids by more abundant ones and difficulties in resolving isobaric lipids [125]. Nevertheless, lipidomics plays a central role in the experimental identification of protein-lipid complexes [126].

2.2.3 Structural Mass Spectrometry

The field of structural mass spectrometry includes techniques that infer the three-dimensional structure of a protein or protein complex from the analysis of intact proteins (top down) and methods that encode structural information in chemical modifications prior to MS analysis (bottom up). Two main groups of reagents are employed in bottom up approaches: (1) Chemical labels that react with a single atom (e.g. a single protein residue) and (2) cross-linking reagents that covalently connect two atoms (e.g. two residues of two interacting proteins).

Labelling Techniques Labelling of proteins for structural analysis is similar to the detection of post-translational modifications in proteomics. Solvent accessible residues are modified, the modified protein is enzymatically hydrolysed and analysed by LC-MS/MS. Identification of the labelled residues during database search enables inferences on the overall structure of the molecule. During labelling by hydrogen-deuterium exchange, proteins are solubilised in deuterated solvent such as D_2O and solvent accessible protons exchange against deuterium atoms [127]. Deuteration in quenched by low pH values and the incorporation of deuterium into the peptide backbone amides is quantified by LC-MS/MS. To maintain quenching conditions and prevent back-exchange of the deuterium atoms for protons [128, 129], enzymatic hydrolysis is performed at low pH and low

temperatures requiring specialised equipment. Protein structures during interactions with proteins, lipids and nucleic acids have been investigated using this labelling technique [128]. Labelling by rapid footprinting involves labelling reagents with fast reaction rates for example through the formation of radical intermediates. In contrast to hydrogen-deuterium exchange, those labels are not prone to back-exchange increasing the reliability of the labelling. Hydroxyl radicals generated by various methods [130] are typically used. They react with solvent accessible amino acid side chains of a protein or protein complex, proteins are subsequently hydrolysed and the chemical modification is mapped and quantified by MS. Over the last decade, the use of laser flash photolysis to generate hydroxyl radicals [131] enabled the investigation of fast conformational transitions of proteins [132]. Covalent labelling involves chemical modifications directed towards a specific subset of amino acids and reacting within minutes to hours [133]. For this, proteins or protein complexes are incubated with the labelling reagent for a defined time period and modifications on the proteins are detected by conventional enzymatic hydrolysis and LC-MS/MS analysis. Key advantages of covalent labelling include its ease of use and the availability of reagents with various well-characterised reaction specificities [134]. Covalent labelling has been successfully employed for example to define binding sites for small molecules interacting with amyloid proteins [135].

Chemical Cross-Linking Chemical cross-linking takes the concept of covalent labelling a step further. Instead of tagging surface accessible residues, cross-linkers covalently connect two solvent accessible residues in proximity. A typical cross-linker molecule consists of two reactive groups connected by a spacer that defines the maximum linking distance of the cross-linker. Cross-linking reagents can consist of identical (homobifunctional) or different (heterobifunctional) reactive groups. In addition to reactive groups linked by a spacer, cross-linker molecules that do not introduce a spacer arm (zero-length cross-linkers) are commonly used. The repertoire of chemical cross-linkers includes molecules targeting primary amines [136], carboxylic acids [137] and cysteine residues. A commonly applied homobifunctional cross-linker is bis(sulfosuccinimidyl)glutarate (BS2G) (**Figure 8A**) that reacts with amine groups (and to a lower extend with serine, threonine and tyrosine residues [136]) through formation of an isopeptide bond. By hydrolysis of a reaction intermediate linked to a single residue, monolinks are formed. Another commonly used cross-linker is the zero-length cross-linker 1-ethyl-3-[3-(dimethylamino)propyl]carbodiimide (EDC) (**Figure 8B**). EDC activates carboxylic acids for example in glutamate and aspartate residues and the activated isourea intermediate reacts with primary amines by forming an isopeptide bond. Moreover, photo-reactive cross-linkers that react with every functional group in proximity are available [138]. For site-specific introduction of UV-reactive groups in a protein of interest, artificial amino acids were developed that utilise a bio-orthogonal codon available in specialised expression strains [139, 140].

The experimental workflow for MS identification of cross-links is similar to the identification of labelled residues and starts with enzymatic hydrolysis of the cross-linked proteins (see **Figure 7A**). In the resulting peptide mixture, cross-linked peptide pairs are usually low abundant and, therefore, enriched prior to MS analysis. For this, size-exclusion [141] and ion-exchange chromatography [142] have been applied, that separate the cross-linked peptide pairs from linear peptides by their larger hydrodynamic radius and charge, respectively. Moreover, enrichable cross-linkers allow specific enrichment for example using metal-affinity chromatography [143]. Next, cross-linked peptide pairs in the fractions from cross-link enrichment are separated by LC and analysed by tandem MS (see **Figure 7C**) before the resulting mass spectra are computationally analysed. In comparison to labelling, however, data analysis for cross-linking MS is more complicated. As all pairwise combinations of peptides have to be considered during generation of theoretical spectra, search space increases quadratically with the number of proteins. Therefore, specialised software tools for data analysis of cross-linking experiments have been developed [144–150]. More recently, advancements in MS instrumentation and data analysis rendered more complex samples amenable to cross-linking MS so that nowadays analysis of intact cell organelles is feasible [151, 152]. In addition to improved data analysis the cross-linking reagents that form characteristic mass patterns and allow filtering of relevant cross-link spectra were developed to simplify cross-link identifications. They include isotope-labelled cross-linker molecules [146] and gas-phase cleavable cross-linkers that characteristically fragment upon dissociation [153, 154].

The high popularity of cross-linking as structural technique stems in part from its applicability to virtually any protein in solution and its high complementarity to other structural techniques. For example, the combination of cryo-electron microscopy with chemical cross-linking has proven especially fruitful. For example, cross-links have been used to select the correct model from multiple solutions fitting to the experimental electron density [155, 156]. Furthermore, comparative quantitative cross-linking of proteins under different conditions can pinpoint molecular movements, e.g. induced by post-translational modifications [157]. In addition to protein-protein cross-linking, the interaction of proteins with RNA has also been targeted by cross-linking. By exploiting the intrinsic reactivity of nucleobases irradiated with UV light, interaction sites on the RNA [158] as well as on the protein are mapped [159]. Whereas the identification of interacting RNAs is achieved by enzymatic cleavage of the protein and sequencing the remaining RNA, the identification of the protein binding site by MS requires specialised software for data analysis [160]. Chemical cross-linking of proteins with lipids was shown using the amino-reactive cross-linker bis(sulfosuccinimidyl)suberate (BS3) reacting with phosphatidylethanolamine lipids [161] but is likely restricted to lipid classes and cross-linkers with complementary reactive groups. To extend the range of linkable lipids, photoactivatable groups have been introduced on proteins [162] and on several lipids [163, 164]. However, if only the identity of a lipid interacting with a protein and not position of the interaction is of interest, native MS is a good alternative to cross-linking.

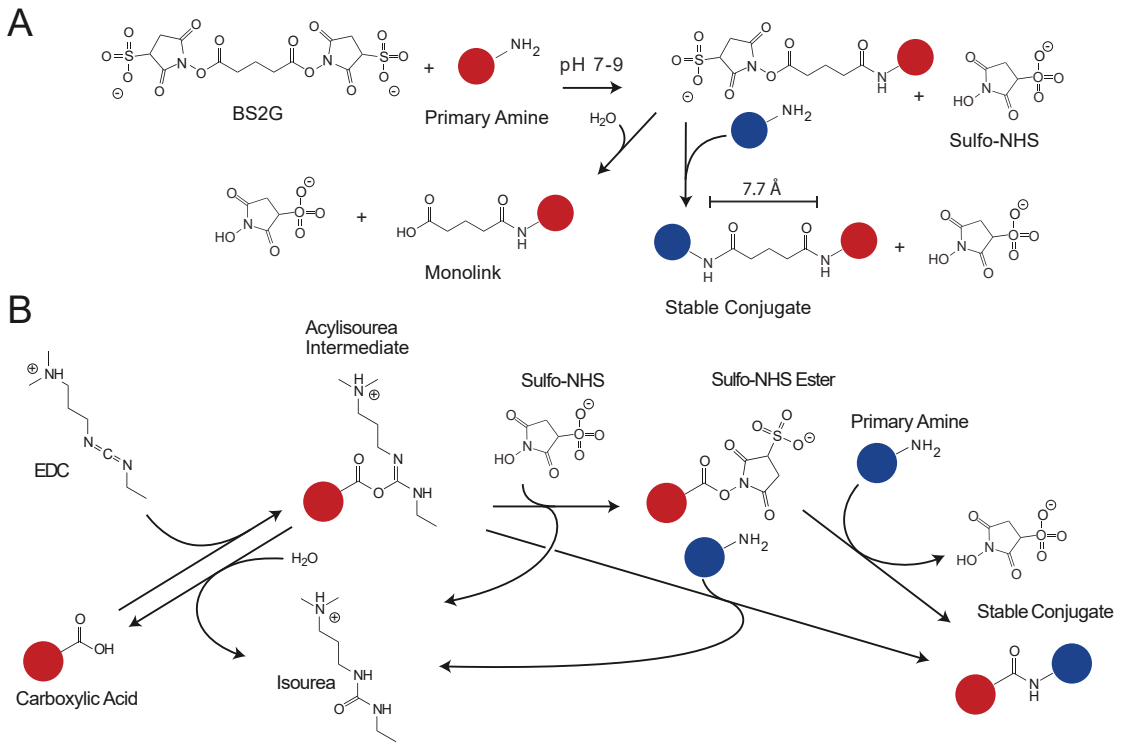


Figure 8: Cross-Linking Reaction Mechanisms of BS2G and EDC. (A) Homobifunctional succinimidyl cross-linkers such as BS2G and BS3 react with primary amines and form an isopeptide bond. **(B)** The zero-length cross-linker EDC activates carboxylic acids that react with primary amines through formation of an peptide bond. Addition of Sulfo-NHS stabilises a reactive intermediate and increases the reaction yield.

Native Mass Spectrometry Native MS is a MS technique that preserves non-covalent interactions, for example interactions between subunits of a protein complex, during transfer of the analyte into the gas-phase. Native MS generally relies on ESI ion sources with solvent conditions (i.e. pH value and ion strength) are adapted to preserve non-covalent interactions. A commonly employed solvent is aqueous ammonium acetate solution, a volatile electrolyte that can be used at physiological ion strengths, prevents low pH values during ionisation (acetate buffers around its pKa value of 4.76) [165] and does not interfere with MS analysis. Before MS analysis, the analyte, e.g. a protein complex, is transferred into ammonium acetate solution using size-exclusion chromatography, filtration devices or dialysis [166] (see **Figure 7A**). In contrast to denaturing MS, analytes are often injected directly into the mass spectrometer without preseparation but coupling to size-exclusion chromatography is possible [167]. Ions take up a characteristic number of charges during ESI that correlates with their surface area [168]. Thereby, the ion charge as such is already informative on structural properties such as folding or unfolding [169]. The mass spectrometers applied for native MS are mostly quadrupole/TOF and quadrupole/orbitrap hybrid tandem instruments.

For large ions to be effectively manipulated in these instruments higher pressures in the first ion stages and reduced quadrupole frequencies are required [170] for native MS instruments compared to mass spectrometers used for peptide analysis. During a native MS experiment, first an MS1 spectrum of all ions corresponding to a protein complex is recorded. As the charge states for the ions belonging to a given single species form a Gaussian distribution, related ions and their corresponding charges can be annotated manually, semi-manually [171] or fully automatic [172] (see **Figure 7C**). The annotated spectrum provides information about the mass of a non-covalent complex and therefore about its stoichiometry. Due to its unbiased nature, native MS also informs about novel interactions if, for example, an unexpected binding partner was retained after purification. To test a potential composition assigned in such an MS1 experiment, the intact complex ion is selected in the quadrupole and dissociated by increasing the collision voltage in the collision cell. The masses inferred from the resulting fragment ions correspond to smaller subcomplexes or complex subunits and therefore the identity of the protein complex is validated. In contrast to peptides, dissociation energy in native MS is too low to dissociate covalent bonds so that peripheral subunits are ejected first. In the process, the ejecting molecule unfolds and gathers charges from the complex due to its increased surface area [173] (see **Figure 7C**). In a mass spectrum, this is reflected by a high charged monomer and a low charged oligomer. This dissociation pattern can further be used to distinguish subcomplexes arising by dissociation in solution prior to MS analysis from subcomplexes formed after collisional activation. Thereby, native MS enables topological inferences on a protein complex [174].

Native MS is not restricted to soluble proteins but can be also applied to membrane proteins in the presence of suitable detergents. For this, the protein is either purified in or transferred into MS-compatible non-ionic detergents [175]. Subsequently, the detergent-solubilised protein is analysed as described for soluble proteins. However, to obtain well-resolved spectra, detergents interacting with the protein need to be removed by collisional activation. Even though the collision energy for detergent removal is usually higher compared than the collision energy employed for soluble protein analysis, protein ions remain typically intact as the detergent micelle captures most of the activation energy [176]. Not all detergents are applicable to all proteins and a suitable protein detergent combination needs to be determined experimentally. Interestingly, membrane proteins often retain tightly bound lipids during purification and MS analysis so that interacting lipids can be identified from native MS [174]. Protein-lipid interactions can also be determined by native MS analysis if the lipids are not retained after purification or not present in the expression host. By addition of lipid mixtures to a membrane protein of interest in the presence of detergents, protein-lipid complexes are formed and high-affinity binders can be identified by native MS [177, 178]

Ion Mobility Separation IMS separation is frequently coupled to native MS instruments to facilitate identification of complex mass spectra, monitoring unfolding during collisional activation or producing collisionally averaged collisional cross-sections that can be compared with structural models [179]. With increasing heterogeneity of the analyte mixture, charge states distributions from ESI tend to overlap, complicating correct mass assignment. If the different analytes have distinct shapes, the corresponding ions can be separated by IMS and peaks belonging to a common species can be assigned. Interactions of proteins with, for example, small molecules or lipids stabilise the protein structure in the gas-phase [115]. This is monitored using IMS separation and increasing the energy of the collision cell of a tandem mass spectrometer. In such a collision-induced unfolding (CIU) experiment, collision energy and drift time are recorded in the presence and absence of a putative ligand. For this, a protein is first analysed in the absence of ligand and drift times at increasing collisional activation corresponding to successive unfolding of the protein are recorded. Productive protein-ligand interactions are determined by a shift of such a drift-time transition to higher collisional activation when the ligand is present. IMS is also used to produce rotationally-averaged collision-averaged cross sections (CCSs), i.e. orientation-averaged projection areas of the molecule structure. Several tools to calculate CCSs from molecular structures have been published [180, 181]. In drift-tube IMS can be directly inferred from drift-time, whereas travelling wave instruments require calibration with a matching set of standard proteins for each analyte. The CCS of an ion is dependent on its charge state with lower charge states are generally considered more native than high charge states [179]. Experimentally derived CCSs are systematically smaller than CCSs obtained from structural models generated based on X-ray crystallography [182]. Nevertheless, CCSs have been successfully used for the generation and validation of structural models [183, 184].

2.3 Molecular Dynamics Simulations of Protein Interactions

Computational simulation of biomolecules allows to reproduce structural dynamics (i.e. conformational changes over time) *in silico* with spatial and temporal precision exceeding experimental techniques. Therefore, computational simulations are a valubale tool to generate hypotheses and interpret experimental results. For small systems and short time scales, quantum mechanics approaches relying on first principles are applicable. However, simulation of macromolecules such as proteins is prohibited by the high computational cost of such methods. In molecular mechanics, performance is increased by ignoring the explicit representation of electrons and approximating the interactions between nuclei in terms of classical mechanics, i.e. as particles connected by springs. Despite the large simplification, MD simulations accurately capture experimentally determined local and global dynamics of flexible proteins over long time scales [185] and to successfully

guide experimental design for complex systems [186]. However, every performance-increasing approximation comes with the cost of reduced accuracy and should be chosen based on the biological system under investigation.

To calculate MD simulations, the potential energy landscape of a system is calculated using molecular mechanics (**Section 2.3.1**). As forces can be derived from the potential energy, Newtons equations of motions are used to update positions and velocities of the atoms based on the forces corresponding to the potential (section 2.3.2). By iteratively updating potential, particle positions and velocities, trajectories corresponding to molecular movement are calculated.

2.3.1 Calculating Energy Potentials with Molecular Mechanics

In traditional MD simulations, atoms are represented by a charged particle with a given mass. Its movements are governed by the sum of all physical interactions with its environment. These can be broadly grouped into bonded interactions with neighbouring atoms connected by a covalent bond and non-bonded interactions with all atoms in proximity. Potentials in MD are generally calculated based on the positions of the particles. While different force fields (see below) differ in their detailed implementation of the potential energy calculation, the general principle is similar.

$$U = \underbrace{U_{bond} + U_{angle} + U_{dihedral} + U_{other}}_{bonded} + \underbrace{U_{vdW} + U_{coulomb}}_{non-bonded}$$

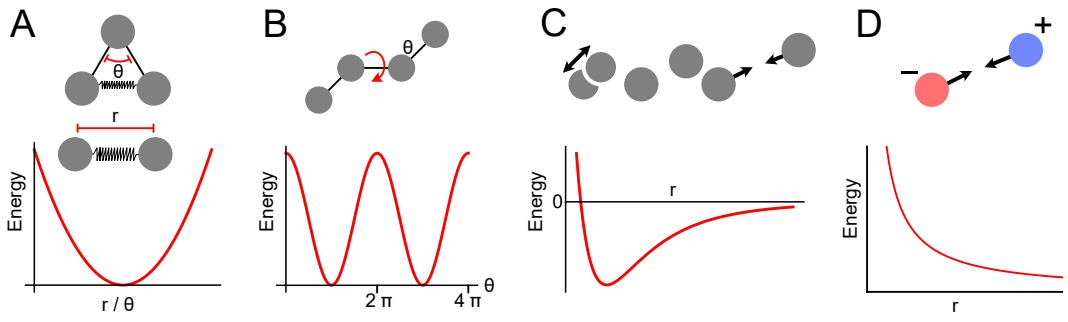


Figure 9: Molecular Mechanics Potential Terms. **(A)** Harmonic potential used for bond and angle potentials. **(B)** Periodic harmonic potential for dihedral angles. **(C)** 12-6 Lennard-Jones potential for van der Waals interactions. **(D)** Coulomb potential for electrostatic interactions.

Bonded Potentials The bond potential U_{bond} between two adjacent atoms and the angle potential U_{angle} between two adjacent bonds are modelled as a harmonic potential with a minimum at the equilibrium bond length and bond angle, respectively. The dihedral angle potential $U_{dihedral}$ describes the angle between two adjacent atoms with respect to their common bond. Other bonded potentials U_{other} include for example the improper dihedral potential to keep planar groups (e.g. aromatic rings) planar. The requirement of predefined bonded potentials (i.e. a fixed molecule topology) limits the applicability of MD simulations to systems in which no bonds are broken or formed i.e. no chemical reactions occur.

Non-bonded Potentials Non-bonded potentials describe interactions between atoms not connected by a bond or angle (i.e. atoms which are separated by more than three covalent bonds [1]). The van der Waals potential U_{vdW} consists of a repulsive term corresponding to atom shell repulsion and an attractive term describing the London force generated by transient dipole moments. Due to its computational efficiency, the 12-6 Lennard-Jones potential with the potential well depth ϵ, the size of a particle σ and the interatomic distance r is often used (see **Figure 9**).

$$U_{vdW} = 4\epsilon \left(\left[\frac{\sigma}{r} \right]^{12} - \left[\frac{\sigma}{r} \right]^{6} \right)$$

In a Lennard-Jones potential the repulsive term scales with $1/r^6$, whereas the attractive term scales with $1/r^{12}$. In practice, U_{vdW} is typically computed up to a maximum cut-off distance at which the potential is reasonably close to zero to be excluded from the calculations without introducing large errors.

Electrostatic interactions are governed by the Coulomb potential $U_{coulomb}$ that scales linearly with the charges (q_1 and q_2) of the involved atoms, by $1/r$ with interatomic distance and is inversely proportional to ϵ_0, the permittivity of space (see **Figure 9**).

$$U_{coulomb} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r}$$

Thereby, $U_{coulomb}$ significantly contributes to the potential energy of atoms separated by larger distances. Instead of using a plain cut-off as described above for U_{vdW}, long-range electrostatics potentials are typically treated by fast-electrostatics algorithms such as Particle Mesh Ewald [188]. For this purpose, $U_{coulomb}$ is decomposed into a short-range term computed explicitly up until a certain distance and a long-range term that is more efficiently computed in Fourier space.

[1] 1-4 interactions between atoms separated by three covalent bonds are treated differently by different force-fields. The CHARMM force field does fully include 1-4 interactions but modifies parameters for some atom types [187].

Force Fields and System Equilibration The precise equations and constants for the potential terms are bundled in force fields. Force fields are derived from quantum calculations or experimental results and optimised towards certain molecules or conditions. Therefore, the force field for a simulation should be chosen with respect to the target and conditions of the simulation and no entirely generalisable best force field is available. As force fields describe equilibrium conditions, a system for simulation needs to be in an equilibrium state prior to the actual MD simulation. For this, the system is first transformed to a local energy minimum by energy minimisation. This prevents very large local energies, e.g. from atom clashes, that influence subsequent simulation by overcompensation and explosion of the total energy of the system. However, the local energy minimum is usually not representative of an equilibrated system. Therefore, one or several equilibration MD simulations with some atom positions restrained by a defined force are usually carried out. Thereby, the system is transformed to an equilibrium state while keeping the desired initial conformation. After equilibration, a production MD simulation is performed without restraints.

2.3.2 Molecular Dynamics Simulations

Integrating Newtons Equations of Motion During an MD simulation, atom positions $X(t)$ are iteratively updated based on the molecular mechanics potential $U(X(t))$. For this, forces $F(X(t))$ acting on the atoms are computed for every time step t.

$$F(X(t)) = -\nabla U(X(t))$$

Using Newtons equations of motion, the forces are translated into positions and velocities for each atom after a time step. The Verlet scheme, an iterative version of the equations of motion, is the simplest approach for this.

$$X(t + \Delta t) = 2X(t) - X(t - \Delta t) + \frac{F(X(t))}{m}$$

$$V(t) = \frac{X(t + \Delta t) + X(t)}{2\Delta t}$$

$V(t)$ denotes the velocity of a particle at time t, m is the mass of the particle and Δt is the integration time step. Δt is chosen so that the fastest motion in the system, the fluctuation between a hydrogen and a heavy atom, is sampled. These oscillations occur at a frequency of 10×10^{14} Hz and restrict the integration time step to 0.5 fs. By constraining hydrogen-heavy atom and heavy atom-heavy atom oscillations using algorithms such as LINCS [189], the integration time step can be increased to 2 fs. Calculation of molecular mechanics potentials and updating of positions is repeated until the desired length of the simulation trajectory is reached.

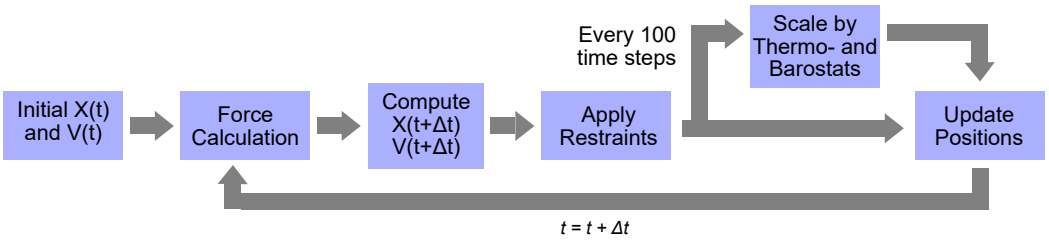


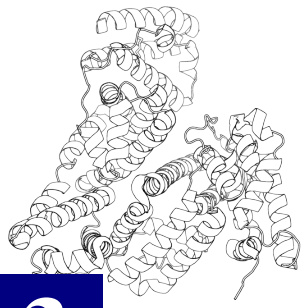
Figure 10: An MD Time Step. Calculations for the next time step $t + \Delta t$ are shown. Atom positions are termed $X(t)$ and atom velocities $V(t)$.

Sampling under Realistic Conditions Biological macromolecules exist in an aqueous environment such as the cytoplasm. In an MD simulation, the macromolecules are typically immersed in a box of defined dimensions containing water molecules and salt ions. Periodic boundary conditions are applied to imitate a continuous system, i.e. all atoms leaving the box on one side re-enter on the opposite site and interactions are computed across box boundaries. This prevents boundary artefacts, however, interactions between two atoms are counted twice, once across the inner of the box and once across periodic boundaries. The minimum image convention prevents this by stating that only interactions between the closest pair of atoms should be considered. In addition, periodic boundary conditions require that the simulation box has larger dimensions than the distance at which long-range interactions between two atoms are close to zero.

Biological systems exchange energy with their surroundings. Thermal energy released by a reaction is readily dissipated across a cell and volume changes are met by flexible membranes. For correct representation of these conditions, MD simulations of biological macromolecules are usually carried out in the canonical NVT (constant number of particles, volume and temperature) or in the isothermal-isobaric NPT ensemble (constant number of particles, pressure and temperature). To simulate a particular thermodynamic ensemble, temperature and pressure are controlled with thermostats and barostats, respectively. Thermostats like the Berendsen [190] and Nosé-Hoover [191, 192] thermostats work by modifying the equations of motion such that the velocities of the atoms are scaled. Barostats, in contrast, scale the positions of the atoms and, consequently, change the size of the simulation box. Commonly used barostats are, for instance, an adaptation of the Berendsen thermostat and the Parrinello-Rahman barostat [193].

The goal of an ideal MD simulation is to sample a system for a sufficient time so that the time average of any simulated quantity reasonably approximates the ensemble average (ergodic hypothesis). However, the time taken to sample a representative amount of the phase space (of all possible combinations of atom positions and velocities) depends on the potential landscape. If part of the phase space is separated by a high energy barrier, it will take more simulation steps to sample than on a planar potential landscape. Therefore, it is impossible to determine if sufficient simulation time was accumulated based on the simulation alone. Even though enhanced sampling methods to

overcome such energy barriers have been developed [194], the sampling problem as such remains. Therefore, the combination of MD simulations with experiments is particularly constructive. For example, MD simulations were employed to interpret results from native MS CIU experiments by calculating the stabilising effect of the presence of interacting lipids on the CCS of the protein [195]. Another example is the computational simulation of the IDR of Syt-1 and comparison of the obtained structural ensembles with distances determined by fluorescence resonance energy transfer spectroscopy revealing that the IDR may act as a sensor the low dielectric environment near the membrane surface [47].



3 Material and Methods

3.1 Material

Table 1: Chemicals.

Product Name	Manufacturer
Acetonitrile, LC-MS grade	Thermo Fisher Scientific, Waltham, USA
Agar	Carl Roth, Karlsruhe, Germany
Ni^{2+}-nitrilotriacetic acid agarose beads	Qiagen, Hilden, Germany
Ammonium acetate solution (7.5 M)	Sigma Aldrich, St. Louis, USA
Ammonium bicarbonate	Sigma Aldrich, St. Louis, USA
Bovine serum albumin (BSA)	Carl Roth, Karlsruhe, Germany
$CaCl_2$ x $6H_2O$	Sigma Aldrich, St. Louis, USA
CsI	Sigma Aldrich, St. Louis, USA
Dithiothreitol	Sigma Aldrich, St. Louis, USA
Ethylene diamine tetraacetic acid (EDTA) disodium salt, dihydrate	Carl Roth, Karlsruhe, Germany
Ethanol 99.8 % (v/v), LC-MS grade	Carl Roth, Karlsruhe, Germany
Formic acid	Thermo Fisher Scientific, Waltham, USA
Glycerol	Sigma Aldrich, St. Louis, USA
4-(2-Hydroxyethyl)-1-piperazine ethanesulfonic acid (HEPES)	Sigma Aldrich, St. Louis, USA
Iodoacetamide	Sigma Aldrich, St. Louis, USA
Imidazole	Sigma Aldrich, St. Louis, USA
Isopropyl-β-D-thiogalactosid	Carl Roth, Karlsruhe, Germany
KH_2PO_4	Carl Roth, Karlsruhe, Germany

Continued on next page

Table 1 – *Continued from previous page*

Product Name	**Manufacturer**
K_2HPO_4	Carl Roth, Karlsruhe, Germany
Kanamycin A	Carl Roth, Karlsruhe, Germany
$MgSO_4$ x $7H_2O$	Carl Roth, Karlsruhe, Germany
NaH_2PO_4 (monobasic, anhydrous)	Sigma Aldrich, St. Louis, USA
NaCl	Carl Roth, Karlsruhe, Germany
Sodium acetate	Sigma Aldrich, St. Louis, USA
Tris(2-carboxyethyl)phosphine (TCEP)	Sigma Aldrich, St. Louis, USA
Trifluoroacetic acid	Thermo Fisher Scientific, Waltham, USA
Tris(hydroxymethyl)aminomethane (TrisBase)	Sigma Aldrich, St. Louis, USA
Ultrapure water (Optima LC-MS grade)	Thermo Fisher Scientific, Waltham, USA

Table 2: Proteins.

Product Name	**Manufacturer**
Alcohol dehydrogenase (ADH)	Sigma Aldrich, St. Louis, USA
β-lactoglobulin (β-LG)	Sigma Aldrich, St. Louis, USA
Bovine serum albumin (BSA)	Sigma Aldrich, St. Louis, USA
Myoglobin	Sigma Aldrich, St. Louis, USA
Turnip crinkle virus (TCV) p38 wild-type	apl. Prof. Dr. Ralph Golbik, Halle (Saale)
Turnip crinkle virus (TCV) p38 W274A	apl. Prof. Dr. Ralph Golbik, Halle (Saale)

Table 3: Lipids and Soluble Lipid Head Groups.

Product Name	**Manufacturer**
Cholesterol (ovine)	Avanti Polar Lipids, Birmingham, USA
Di-oleyl-phosphatidylcholine (DOPC)	Avanti Polar Lipids, Birmingham, USA
Di-oleyl-phosphatidylglycerol (DOPG)	Avanti Polar Lipids, Birmingham, USA
Di-oleyl-phosphatidylinositol-3,4,5-trisphosphate (DOPI(3,4,5)P_3)	Avanti Polar Lipids, Birmingham, USA

Continued on next page

Table 3 – *Continued from previous page*

Product Name	**Manufacturer**
Dioleyl-phosphatidylserine (DOPS)	Avanti Polar Lipids, Birmingham, USA
Inositol-1,3,5-triphosphate (I(1,3,5)P_3)	Avanti Polar Lipids, Birmingham, USA
Inositol-1,4,5-triphosphate (I(1,4,5)P_3)	Avanti Polar Lipids, Birmingham, USA
Phosphatidylinositol-4-phosphate (PI(4)P) (brain)	Avanti Polar Lipids, Birmingham, USA
Phosphatidylinositol-4,5-bisphosphate (PI(4,5)P_2) (brain)	Avanti Polar Lipids, Birmingham, USA
Palmitoyl-oleyl-phosphatidylcholine (POPC)	Avanti Polar Lipids, Birmingham, USA
Palmitoyl-oleyl-phosphatidylethanolamine (POPE)	Avanti Polar Lipids, Birmingham, USA
Palmitoyl-oleyl-glycerol (Diacyl glycerol)	Avanti Polar Lipids, Birmingham, USA

Table 4: Detergents.

Product Name	**Manufacturer**	**CMC* [%]**
Octyl tetraethylene glycol ether (C8E4)	Glycon Biochemicals, Luckenwalde, Germany	0.25
3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS) $\geq$ 98%	Carl Roth, Karlsruhe, Germany	0.49
Lauryldimethylamine N-oxide (LDAO)	Glycon Biochemicals, Luckenwalde, Germany	0.023
Octyl glycoside (OG)	Glycon Biochemicals, Luckenwalde, Germany	0.53
Sodium cholate hydrate	Sigma Aldrich, St. Louis, USA	0.60
Tween20	Thermo Fisher Scientific, Waltham, USA	0.0074
Triton X-100	Sigma Aldrich, St. Louis, USA	0.02

* Critical micelle concentration

Table 5: Special Reagents.

Product Name	**Manufacturer**
Bis(sulfosuccinimidyl)glutarate (BS2G)	Thermo Fisher Scientific, Waltham, USA
1-ethyl-3-[3-(dimethylamino)propyl]carbodiimide (EDC)	Thermo Fisher Scientific, Waltham, USA
RapiGest	Waters Corporation, Milford, USA
N-hydroxysulfosuccinimide (Sulfo-NHS)	Thermo Fisher Scientific, Waltham, USA

Table 6: Cells.

Organism	**Strain**	**Supplier**
Escherichia coli (*E. coli*)	NEB5-alpha (DH5α derivative)	New England Biolabs
Escherichia coli (*E. coli*)	BL21(DE3)	New England Biolabs

Table 7: Plasmids.

Plasmid	**Backbone**	**Insert**	**Source**
pET28a Syt-1 FL	pET28a	*Rattus norvegicus* Syt-1 (1-421)	Prof. Dr. Reinhard Jahn, Göttingen
pET28a Syt-1 C2AB-His	pET28a	*Rattus norvegicus* Syt-1 (143-421)	This thesis
pET28a Syt-1 Cyt	pET28a	*Rattus norvegicus* Syt-1 (97-421)	This thesis
pET28a Syt-1 C2AB	pET28a	*Rattus norvegicus* Syt-1 (84-421)	Prof. Dr. Reinhard Jahn, Göttingen
pET28a Syt-1 C2A	pET28a	*Rattus norvegicus* Syt-1 (97-263)	Prof. Dr. Reinhard Jahn, Göttingen
pET28a Syt-1 C2B	pET28a	*Rattus norvegicus* Syt-1 (271-421)	Prof. Dr. Reinhard Jahn, Göttingen

Table 8: Antibodies.

Name	Epitope	Host	Supplier
Anti-His IgG	Hexahistidine tag	Mouse (monoclonal, clone HIS.H8)	Abcam, Cambridge, UK
Anti-mouse IgG peroxidase-coupled	Mouse IgG	Rabbit (polyclonal)	Sigma Aldrich, St. Louis, USA
Anti-Syt-1 cytoplasmic tail IgG	*Rattus norvegicus* Syt-1 residues 80 - 421	Mouse (monoclonal, clone 41.1)	Synaptic Systems, Göttingen, Germany

Table 9: Enzymes.

Product Name	Manufacturer
DNase I (bovine)	Sigma Aldrich, St. Louis, USA
Thrombin	Thermo Fisher Scientific, Waltham, USA
Trypsin	Promega, Mannheim, Germany
Trypsin	Roche, Mannheim, Germany

Table 10: Buffers, media and kits.

Product Name	Manufacturer
Bis-Tris gels 4-12% (Novex NuPAGE)	Thermo Fisher Scientific, Waltham, USA
Colloidal coomassie (InstantBlue)	Sigma Aldrich, St. Louis, USA
cOmplete EDTA-free protease inhibitor tablets	Roche, Basel, Switzerland
ECL detection kit	Thermo Fisher Scientific, Waltham, USA
LDS reducing agent (Novex NuPAGE)	Thermo Fisher Scientific, Waltham, USA
LDS sample buffer (Novex NuPAGE)	Thermo Fisher Scientific, Waltham, USA
MES SDS buffer (Novex NuPAGE)	Thermo Fisher Scientific, Waltham, USA
Phosphate buffered saline tablet	Sigma Aldrich, St. Louis, USA
SeeBlue Plus 2 prestained standard protein marker	Thermo Fisher Scientific, Waltham, USA

Continued on next page

Table 10 – *Continued from previous page*

Product Name	**Manufacturer**
Tryptone (peptone from casein)	Sigma Aldrich, St. Louis, USA
Yeast extract (total)	Sigma Aldrich, St. Louis, USA

Table 11: Instruments and Devices.

Name	**Product Name**	**Manufacturer**
Cell disruptor (Syt-1 FL)	CF1	Constant Systems, Daventry, UK
Cell disruptor (soluble variants)	Micron LAB 40	APV Systems, Unna, Germany
Chemiluminescence detector	LAS-4000	Fujifilm, Minato, Japan
Circular dichroism spectropolarimeter	J-810	JASCO Instruments, Tokyo, Japan
Fast protein liquid chromatography system	ÄKTA pure	General Electric, Boston, USA
Micropipette puller	Q 150R S	Quorum, Laughton, UK
Nano high-performance liquid chromatography	DionexUltiMate 3000 RSLCnano	Thermo Fisher Scientific, Waltham, USA
Polyacrylamide gel system	Novex Mini Cell	Invitrogen
Quadrupol-orbitrap mass spectrometer	Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer	Thermo Fisher Scientific, Waltham, USA
Quadrupol-TOF mass spectrometer	Q-TOF Ultima (modified for transmission of high masses)	Micromass, Wilmslow, UK (modified by MSVision, Almere, Netherlands)
Quadrupol-TOF mass spectrometer	Synapt G1 HDMS (modified for transmission of high masses)	Waters Corp, Milford, USA (modified by MSVision, Almere, Netherlands)
Sputter coater	Model P-1000	Sutter Instruments, Novato, USA
Ultracentrifuge	Optima MAX-XP (MLS 50 rotor)	Beckman Coulter, Brea, USA

Table 12: Chromatography Columns.

Product Name	Manufacturer
C18 capillary column (50 cm, HPLC column Acclaim PepMap 100, C18, 75 µm I.D., particle size 3 µm)	Thermo Fisher Scientific, Waltham, USA
HisTrap HP 5 mL	Cytiva Life Sciences, Marlborough, USA
HiTrap Q HP 5 mL	Cytiva Life Sciences, Marlborough, USA
HiTrap SP HP 5 mL	Cytiva Life Sciences, Marlborough, USA
Micro Bio-Spin Columns with Bio-Gel P-6	Biorad, Hercules, USA
Superdex peptide 3.2/300 column	Cytiva Life Sciences, Marlborough, USA
µ-Precolumn C18 (Acclaim PepMap 100, C18, 300 µm I.D., particle size 5 µm)	Thermo Fisher Scientific, Waltham, USA

Table 13: Consumables.

Product Name	Manufacturer
1.0 OD x 0.58 ID x 100 mm Borosilicate Glass Capilaries	Harvard Apparatus, Cambridge, USA
AmiconUltra-0.5	Merck, Darmstadt, Germany
Polyvinylidene fluoride blotting membrane	Carl Roth, Karlsruhe, Germany
Polyvinylidene fluoride lipid strips (Membrane Lipid Strips P-6002)	Echelon Biosciences, Salt Lake City, USA
Vivaspin 500	Vivaproducts, Littleton, USA
Whatman paper	Carl Roth, Karlsruhe, Germany

Table 14: Software and Servers.

Software	Version	Maintainer / Source	Link
Anaconda Python Distribution	4.9.2	Anaconda Inc.	anaconda.org/
CHARMM-GUI		Jo et al. [196]	charmm-gui.org
CroCo	1.3.1	This work	github.com/cschmidtlab/CroCo

Continued on next page

Table 14 – *Continued from previous page*

Software	Version	Maintainer / Source	Link
Gromacs	2020.4	Abraham et al. [197]	www.gromacs.org/
HADDOCK	2.4	van Zundert et al. [198]	wenmr.science.uu.nl/
MassLynx	4.1	Waters Corp.	
MaxQuant	1.6.3	Cox and Mann [121]	www.maxquant.org
Membrane Curvature Tool	0.0.2	Barreto-Ojeda [199]	github.com/MDAnalysis/membrane-curvature
pLink2	2.3.9	Chen et al. [144]	pfind.ict.ac.cn/software/pLink
PULSAR	1.0.0	Allison et al. [200]	pulsar.chem.ox.ac.uk/
PyMol	2.3.0	Schrödinger LLC	github.com/schrodinger/pymol-open-source
Python	3.8.3	Python Software Foundation	www.python.org/
ROSETTA	2021.16.61629	Mandell et al. [201]	www.rosettacommons.org/
scipy	1.7.0	Virtanen et al. [202]	scipy.org
SESCA	0.95	Nagy et al. [203]	www.mpibpc.mpg.de/sesca
UCSF Chimera	1.15	Pettersen et al. [204]	www.cgl.ucsf.edu/chimera/
UniDec	180321 to 210804	Marty et al. [172]	github.com/michaelmarty/UniDec
vmd	1.9.3	Humphrey et al. [205]	www.ks.uiuc.edu/Research/vmd/
Xlink Analyzer	1.1.4	Kosinski et al. [148]	www.embl-hamburg.de/XlinkAnalyzer/
XWalk	0.6	Kahraman et al. [206]	github.com/abxka/Xwalk

3.2 Methods

3.2.1 Expression and Purification of Synaptotagmin-1 Variants

Transformation of *Escherichia coli* Cells 200 µL chemically competent *E. coli* NEB5-alpha (for plasmid production and storage) or BL21 (for protein expression) cells were inoculated with 3 µL of purified plasmid (see **Table 6**) and incubated on ice for 30 min. Cells were heated to 42 °C for 10 s (BL21) or 30 s (NEB5-alpha) and placed on ice for 5 min. By heating and cooling, plasmid DNA enters the cells through transient pores in the cell membranes. Transformed cells were mixed

Table 15: Buffers and media used for protein purification.

Name	Composition	Application
Lysogeny Broth (LB) medium	10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract	Cell transformation
LB agar	20 g/L agar, 10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract	Cell transformation
Terrific Broth (TB) medium	12 g/L tryptone, 24 g/L yeast extract, 0.4 % (v/v) glycerol, 2.31 g/L KH_2PO_4, 12.54 g/L K_2HPO_4	Protein expression
Resuspension buffer	20 mM HEPES, 500 mM NaCl, 8 mM imidazole pH 7.4	Protein purification
Extraction buffer	20 mM HEPES, 500 mM NaCl, 8 mM imidazole, 10 % (w/v) cholate pH 7.4	Protein purification
Washing buffer	20 mM HEPES, 300 mM NaCl, 80 mM imidazole, 0.1 mM TCEP, 1 % (w/v) CHAPS pH 7.4	Protein purification
Elution buffer	20 mM HEPES, 300 mM NaCl, 400 mM imidazole, 0.1 mM TCEP, 1 % (w/v) CHAPS pH 7.4	Protein purification
Dialysis buffer*	20 mM HEPES, 300 mM NaCl, 1 mM EDTA, 0.1 mM TCEP, 1 % (w/v) CHAPS pH 7.4	Protein purification
Ion exchange chromatography buffer A*	20 mM HEPES, 0.1 mM TCEP, 1 % (w/v) CHAPS pH 7.4	Protein purification
Ion exchange chromatography buffer B*	20 mM HEPES, 1 M NaCl, 0.1 mM TCEP, 1 % (w/v) CHAPS pH 7.4	Protein purification
IMAC buffer A	20 mM HEPES, 300 mM NaCl, 0.1 mM TCEP, pH 7.4	Protein purification
IMAC buffer B	20 mM HEPES, 300 mM NaCl, 400 mM imidazole, 0.1 mM TCEP, pH 7.4	Protein purification
PBS-T	10 mM total phosphate, 2.7 mM KCl, 137 mM NaCl, 1 % (v/v) Triton X-100	Western Blot
TBS-T	50 mM Tris-HCl, 150 mM NaCl, 0.1 % (v/v) Tween20,	Lipid overlay assay

* Buffers were prepared without CHAPS for purification of soluble Syt-1 variants.

with 950 µL LB medium (**Table 15**) and incubated at 37 °C, 250 rpm for 60 min in a thermomixer. Cells were pelleted by centrifugation at 5000 ×g for 2 min, the pellet was resuspended in 100 µL LB medium and dispersed on prewarmed culture plates containing LB agar and 30 µg/(mL) of the selection antibiotic kanamycin. Cells were grown for 8 h - 16 h. Selected transformed clones were cultivated in 5 mL LB medium containing 30 µg/(mL) kanamycin. From these cultures, plasmid DNA for validation of plasmid identity was extracted or cells were mixed with 25 % (w/v) sterile glycerol, flash frozen in liquid nitrogen and stored at −80 °C.

Expression of Full-Length and Soluble Synaptotagmin-1 Variants of *Rattus norvegicus* Syt-1 fused to a hexahistidine tag were expressed as described [77, 54]. 200 mL LB medium (**Table 15**) were inoculated with transformed *E. coli* BL21(DE3) cells and incubated at 37 °C, 180 rpm for 8 h - 16 h. 500 mL TB medium (**Table 15**) were inoculated with 10 mL of the preculture and incubated at 37 °C, 180 rpm until an optical density of 0.9 - 1 at 600 nm was reached. Expression of Syt-1 was induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside and cells were pelleted by centrifugation at 4000 ×g and 4 °C for 20 min after 16 h incubation at 25 °C. Cells were resuspended in 10 mL resuspension buffer (**Table 15**) per litre of cell suspension and stored at −20 °C.

Purification of Full-Length and Soluble Variants of Synaptotagmin-1 Cell pellets were thawed on ice and supplied with 800 µg DNase I and 1 mM $MgSO_4$ per cell pellet corresponding to 1 L of culture medium. Subsequent purification of full-length Syt-1 was carried out in the presence of 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS) detergent to solubilise the hydrophobic transmembrane helix, whereas detergents were omitted for the purification of soluble Syt-1 variants. An equal volume of extraction buffer (**Table 15**) with or without detergent was added and cells were mixed with the buffer in a rotator for 30 min. Cells were disrupted by high pressure and two ethylene diamine tetraacetic acid (EDTA)-free protease inhibitor tablets dissolved in extraction buffer per 150 mL of suspension were added directly after cell lysis. The suspension was incubated on ice for 15 min and insoluble cell debris was pelleted by centrifugation at 27,660 ×g and 4 °C for 30 min. Full-length Syt-1 and the soluble variants, were purified by affinity chromatography using Ni^{2+}-loaded nitrilotriacetic acid (NTA)-coupled agarose beads for the full-length protein and commercially available affinity columns for the soluble variants. As the proteins were expressed fused to a hexahistidine tag, high-affinity binding of the tag to Ni^{2+}-NTA enabled selective purification of the proteins.

For purification of full-length Syt-1, 5 mL Ni^{2+}-loaded NTA-coupled agarose beads washed in washing buffer were added per 150 mL of the centrifugation supernatant. Syt-1 did bind to the agarose beads during 3 h incubation at 4 °C on a rotator. Beads were pelleted by centrifugation at 150 ×g and 4 °C for 10 min, the supernatant was removed and the beads were washed three times with 20 mL, 20 mL and 10 mL of washing buffer by repeated centrifugation and removal of the supernatant. After the last washing step, agarose beads were transferred into an empty polypropylene gravity-flow column. Syt-1 was eluted from the beads by addition of 30 mL of elution buffer. 2 mL fractions were collected and each fraction was supplemented with 1 mM tris(2-carboxyethyl)phosphine (TCEP) (final concentration) directly after elution and the presence of Syt-1 in the fractions was determined by polyacrylamide gel electrophoresis (see section 4.1.2). Fractions containing the highest amounts of Syt-1 were pooled and the hexahistidine tag was removed

by cleavage with the sequence-specific protease thrombin. For this, 50 µL 1 U/µL per 10 mL of protein solution were added and the solution was dialysed in 1 L of dialysis buffer (**Table 15**) for 16 h. The efficiency of thrombin cleavage was assessed by polyacrylamide gel electrophoresis.

For purification of soluble Syt-1 variants, the centrifugation supernatant was filtered through a 0.45 µm filter and loaded onto a HisTrap HP 5 mL column equilibrated in IMAC buffer A (**Table 15**) and bound Syt-1 was eluted by increasing concentrations of IMAC buffer B (**Table 15**). The presence of Syt-1 in the elution fractions was monitored by light absorption at 280 nm and polyacrylamide gel electrophoresis. Fractions with the highest Syt-1 content were combined and treated with thrombin as described. The concentration of imidazole in the protein buffer was reduced by dialysis in 2 L dialysis buffer without EDTA for 16 h (**Table 15**). Uncleaved Syt-1 molecules still attached to a hexahistidine tag were removed by repeating the affinity purification step and collecting the flow-through. The presence of Syt-1 was confirmed by gel electrophoresis and fractions with the highest concentration of Syt-1 were pooled and dialysed in 2 L of dialysis buffer containing EDTA to remove Ni^{2+} ions bound to the protein.

In a last step, proteins and nucleic acids co-purified with full-length and soluble Syt-1 variants were removed by cation-exchange chromatography on a HiTrap SP HP 5 mL column at a flow rate of 1 mL/min. For the soluble variants C2A and C2AB, an anion exchange HiTrap Q HP column was employed instead. The protein solution was filtered through a 0.45 µm filter and loaded onto the column equilibrated in ion exchange chromatography buffer A (**Table 15**). Elution at increasing concentrations of ion exchange chromatography buffer B (**Table 15**) was monitored by light absorption at 280 nm wavelength and the presence of Syt-1 in the fractions was verified by polyacrylamide gel electrophoresis. Fractions with the highest amount of Syt-1 were combined, the protein was concentrated to a final concentration of 1.5 mg/mL to 2 mg/mL using 10 kDa cellulose filtration devices and stored at −80 °C.

3.2.2 Biochemical and Biophysical Methods

Circular Dichroism Spectroscopy The protein concentration of Syt-1 soluble variants was adjusted to <1 mg/mL using 20 mM HEPES-NaOH pH 7.4, 300 mM NaCl, 0.1 mM TCEP. 50 µL of buffer were loaded into a 0.1 mm quartz cuvette maintained at 8 °C cell temperature and ellipticity was recorded from 250 nm - 198 nm using a JASCO J-810 spectropolarimeter. The instrument was operated in continuous scanning mode with a sensitivity of 100 mdeg and a data pitch of 1 nm. 64 scans were accumulated for each final spectrum were recorded at a scanning speed of 50 nm/min. A reference spectrum of the buffer without protein was subtracted from the recorded spectrum and a single binomial smoothing was applied using the JASCO SpectraManager software. Spectra were

truncated to wavelengths at which a photomultiplier voltage of less than 600 V was used for signal amplification. Raw ellipticity was converted to mean residue weighted ellipticity as described [207].

For calculation of theoretical circular dichroism (CD) spectra, a structural model of Syt-1 was obtained (PDB ID 5KJ7 chain E) [208] and the structures of Syt-1 residues 271 - 418, serving as reference for Syt-1 C2B, or residues 141 - 418, as reference for all other experimental CD spectra, were saved as pdb files. These structures were used as input for SESCA [203] operated with standard settings to calculate reference spectra.

Polyacrylamide Gel Electrophoresis Proteins were separated by gel electrophoresis on a 4 % - 12 % Bis-Tris polyacrylamide gel. For this, disulfide bridges were reduced by mixing with 10 × sample reducing agent containing 500 mM dithiothreitol and denatured with 4 × sample loading buffer containing lithium dodecyl sulfate for denaturation, glycerol for loading the sample on the gel and Coomassie G250 as tracking dye. The samples were heated to 70 °C for 10 min and loaded onto a 4 % - 12 % Bis-Tris polyacrylamide gel immersed in dodecylsulfate running buffer containing 2-(N-morpholino)ethanesulfonic acid as buffer and a constant voltage of 200 V was applied to the gel for 35 min. Negatively-charged dodecylsulfate-bound proteins migrated towards the anode and were separated according to their mobility in the gel matrix. Protein bands were visualised by staining with colloidal Coomassie solution for 8 h - 16 h and destaining with distilled water for 8 h - 16 h. Gels were documented using a scanner.

Western-Blot Analysis Unstained polyacrylamide gels were transferred into PBS-T buffer (see **Table 15**). Proteins were transferred from the gel onto polyvinylidene fluoride membranes using a semidry electroblotting system operated at 50 mA constant current per gel for 120 min. Free binding positions on the membrane were blocked using 1 % bovine serum albumin (BSA) in PBS-T buffer. Membranes were incubated with anti-Syt-1 antibody (see **Table 8**) at 8 °C for 8 h - 16 h. Free antibody was removed by washing the membrane three times with PBS-T buffer at room temperature for 30 min per washing step. The primary antibody was detected by incubation with a peroxidase-coupled anti-mouse IgG secondary antibody (see **Table 8**) at room temperature for 1 h, removal of free antibody by washing as described and visulisation by chemoluminescence detection using peroxidase-catalysed reaction of luminol and H_2O_2.

3.2.3 Characterisation of Protein-Lipid Interactions

Protein Lipid Overlay Assay Free binding positions on a prefabricated polyvinylidene fluoride lipid strip with 100 pmol immobilised triglycerides, diacylglycerol, phosphatidic acid, PS, phosphatidylethanolamine, phosphatidylcholine, phosphatidylglycerol (PG), cardiolipin, PI, phosphati-

dylinositol-4-phosphate (PI(4)P), PI(4,5)P_2, phosphatidylinositol-3,4,5-trisphosphate (PI(3,4,5)P_3), cholesterol, sphingomyelin and 3-sulfogalactosylceramide (sulfatide) per spot was blocked by incubation in 2 mg/mL BSA in TBS-T buffer (see **Table 15**) for 1 h. The blocking buffer was removed by 6 - 8 washing steps with TBS-T buffer for a total time 1 h. Either 2.8 µg/mL (75 nM) Syt-1 C2AB or 2.0 µg/mL (100 nM) Syt-1 C2B fused to a hexahistidine tag in the presence or absence of 500 µM $CaCl_2$ were added. The presence of the hexahistidine tag was required for subsequent detection of C2B as the antibody directed against the cytoplasmic part of Syt-1 did not bind to the C2B variant. The strips were incubated either with anti-Syt-1 antibody for experiments including Syt-1 C2AB or with anti-His IgG to recognise the hexahistidine tag on Syt-1 C2B for 8 h - 12 h (see **Table 8**). Lipid strips were again washed in TBS-T buffer and peroxidase-coupled anti-mouse IgG secondary antibody was added. Binding of secondary antibody took place during 1 h incubation and free antibody was removed by washing with TBS-T buffer binding of Syt-1 to a specific lipid was detected by chemiluminescence detection. Experiments were performed by Til Kundlacz.

Raw images (encoding higher luminescence as higher gray values) were loaded into ImageJ [209] and images were cropped to the lipid strip area. Background intensity was substracted using the background removal tool of ImageJ with a 50 px radius. Gray values were modified so that the intensity of the maximum gray value in the spots was equivalent to the maximum intensity of the picture format. For visualisation, the inverted gray lookup table and the viridis lookup table were used.

Preparation of Lipid Mixtures Lipid mixtures for flotation assays were prepared as described [210]. Briefly, for membranes mimicking the SV membrane, palmitoyl-oleyl-phosphatidylcholine (POPC), dioleyl-phosphatidylserine (DOPS), palmitoyl-oleyl-phosphatidylethanolamine (POPE) and cholesterol solubilised in chloroform were mixed at 38:18:20:20 molar ratio. Aliquots of this mixture containing 2.75 mg lipids were prepared and the solvent was evaporated under a dry nitrogen gas stream. Similarly, lipid mixtures mimicking the plasma membrane were composed of a 38:18:20:20:2:2 molar ratio of POPC:DOPS:POPE:cholesterol:PI(4,5)P_2:diacyl glycerol. Note that PI(4,5)P_2 was purchased solubilised in chloroform/methanol. Dried lipid mixtures were stored at −20 °C.

Malachite Green Phosphate Assay Correct mixing of lipids was validated using the malachite green phosphate assay. For this, lipids in a defined volume of aqueous solution were extracted using organic solvents and the solvent was evaporated. Lipids were reduced to ashes in the presence of sulphuric acid and perchloric acid and, by addition of molybdate solution, phosphomolybdate was formed. The basic dye malachite green was added, formed a complex with phosphomolybdate and thereby shifted the absorption maximum to 623 nm. By comparison with a titration curve

of inorganic phosphate, absolute inorganic phosphate concentrations and, using the known lipid-to-phosphate ratio for each lipid species, lipid concentrations in the samples were determined. Analyses were performed by Claudia Müller (working group of Prof. Kirsten Bacia).

Preparation of Proteoliposomes SV or plasma membrane lipid mixtures were dissolved in 2:1 methanol:chloroform to a final concentration of 4 mM lipids, transferred into spherical flasks and the solvent was removed by rotary evaporation. The resulting dry lipid film was rehydrated in 20 mM HEPES pH 7.4, 150 mM KCl, 0.1 mM TCEP, 1 mM EGTA or 20 mM HEPES pH 7.4, 150 mM KCl, 0.1 mM TCEP, 100 µM $CaCl_2$ to a final concentration of 2 mM lipids at room temperature for 1 h. Liposomes of defined size were prepared by extrusion through 100 nm polycarbonate membranes for 21 strokes. Syt-1 C2A or C2B were mixed with liposomes in 1:250 and 1:1000 protein:lipid molar ratio and incubated for 1 h at room temperature. Proteoliposomes were prepared by Melissa Frick.

Flotation Assay Proteoliposomes incubated with Syt-1 C2A or C2B were mixed with sucrose to a final concentration of 1 M. The solution was overlaid with $\frac{3}{4}$ v of 0.75 M sucrose dissolved in the corresponding liposome buffer with or without Ca^{2+}. 0.175 v of the respective buffer without sucrose was added. The assembled sucrose density gradient was centrifuged at 268,000 ×g for >2 h to separate the liposome-bound proteins migrating to lower densities from the unbound proteins sedimenting on the bottom of the ultracentrifugation tube. Fractions from the top and bottom of the tube were collected and the analysed by polyacrylamide gel electrophoresis. Flotation assays were performed by Melissa Frick.

3.2.4 Mass Spectrometry of Unmodified and Cross-Linked Proteins

Cross-Linking of TCV p38 The molar concentration of TCV p38 was determined prior to every cross-linking experiment by light absorption at 280 nm assuming a molar extinction coefficient of 57,410 $^{1}/_{M\,cm}$. A stock solution of 2.5 mM - 5 mM BS2G in 50 mM 4-(2-Hydroxyethyl)-1-piperazine ethanesulfonic acid (HEPES) pH 7.5 100 mM NaCl, 1 mM TCEP was prepared. p38 was mixed with buffer and BS2G stock solution to a final p38 concentration of 10 µM and a cross-linker concentration of 10 µM - 500 µM. The reaction of BS2G with lysine (and to a lower degree serine, threonine and tyrosine) residues took place during incubation at 25 °C, 1050 rpm in a thermomixer for 1 h. Free BS2G was quenched by addition of aqueous ammonium bicarbonate solution to a final concentration of 20 mM and incubating for another 15 min under the same conditions.

For EDC cross-linking of carboxylic acids to lysine residues, a 20 mM EDC stock solution and a 50 mM N-hydroxysulfosuccinimide (Sulfo-NHS) stock solution in 50 mM HEPES pH 7.5, 100 mM NaCl, 1 mM TCEP were prepared immediately before to the experiment. 10 µM p38 were cross-

linked with 0.5 mM - 1.5 mM EDC and 1.25 mM - 3.75 mM Sulfo-NHS by incubation at 25 °C, 1050 rpm for 40 min. Cross-links were identified after in-gel or in-solution hydrolysis of proteins followed by LC-MS analysis. Cross-linking experiments using EDC or BS2G were repeated three times for in-gel and in-solution hydrolysis, respectively.

For comparison of the RNA-free and the RNA-bound form of p38, cross-linking in the absence of RNA was repeated at 5 °C. For this, 10 µM p38 in 50 mM NaH_2PO_4/Na_2HPO_4 pH 7.5, 100 mM NaCl, 1 mM TCEP were cross-linked with 100 µM BS2G and 5 µM p38 in 50 mM NaH_2PO_4/Na_2HPO_4 pH 7.5, 100 mM NaCl were cross-linked with 2 mM EDC. For cross-linking of p38 with BS2G and EDC in the presence of RNA, 2.5 µM RNA and 5 µM p38 were mixed so that cross-linker concentrations of 50 µM BS2G or 2 mM EDC were reached. Reaction mixtures were incubated at 5 °C for 1 h. The BS2G reaction was quenched as described above and EDC cross-linking was stopped by addition of TCEP to a final concentration of 5 mM and tris(hydroxymethyl)aminomethan to a final concentration of 50 mM. The experiment was performed twice except for cross-linking with EDC in the presence of RNA that was performed three times. Cross-links were identified by LC-MS after in-solution hydrolysis and size-exclusion chromatography (SEC) enrichment of cross-linked peptide pairs.

In-Gel Protein Hydrolysis Protein bands were excised from a gel and sliced into 1 mm^2 pieces, washed with ultra pure water and incubated for 5 min at 1050 rpm, 26 °C in a thermomixer. The supernatant was removed and bands were dehydrated by covering with acetonitrile, incubated under the same conditions for 15 min and the supernatant was removed. Gel bands were dried in a vacuum concentrator for 5 min - 10 min. Disulfide bridges in the proteins were reduced by covering the gel bands with 10 mM dithiothreitol in 100 mM ammonium bicarbonate (pH 8.0) and heating for 50 min to 56 °C. Free dithiothreitol in the gel pieces was removed by removal of the supernatant, the gel pieces were covered with acetonitrile, incubated for 15 min and the supernatant was removed. Free cysteines were carbamidomethylated by addition of 55 mM iodoacetamide followed by incubation for 20 min in the dark to prevent photo-degradation of iodoacetamide. The supernatant was removed and gel bands were covered with 100 mM ammonium bicarbonate, incubated for 15 min and the an equal volume of acetonitrile was added. After 15 min incubation, the supernatant was removed, acetonitrile was added and gel pieces were dehydrated for 15 min under the same conditions. The supernatant was removed and gel pieces were dried using a vacuum concentrator. The following steps were performed on ice. Gel pieces were rehydrated with 15 µL hydrolysis buffer 1 (see **Table 16**) for 30 min - 45 min, more buffer was added if gel pieces were not completely rehydrated. To prevent dehydration, 10 µL - 50 µL hydrolysis buffer 2 without trypsin were added. Proteins were hydrolysed at 37 °C for 12 h - 16 h [211]. After hydrolysis, gel pieces were covered with water and incubated at 37 °C for 15 min. An equal volume of acetonitrile was added and the solution was incubated for 15 min under the same conditions. The supernatant was transferred to a fresh tube

Table 16: In-gel hydrolysis buffers

Reagent	Buffer 1	Buffer 2
Trypsin	12.5 µg/mL	-
NH_4HCO_3 (pH 8.0)	41.67 mM	47.62 mM
$CaCl_2$	4.17 mM	4.76 mM

and remaining peptides were extracted from the gel pieces by covering with 5 % formic acid and incubation for 15 min under the same conditions. An equal volume of acetonitrile was added, the incubation step was repeated and the supernatant was collected and pooled with that from the first peptide extraction step. Peptide solutions were concentrated to dryness in a vacuum concentrator and stored at −20 °C.

In-Solution Protein Hydrolysis Solubilised proteins were precipitated with ethanol. For this, proteins were mixed with 3 v ice-cold ethanol and $\frac{1}{10}$ v 3 M sodium acetate pH 5.3. During incubation at −20 °C for >2 h, proteins precipitated and were sedimented by centrifugation at 16,000 ×g and 4 °C for 30 min. Protein pellets were washed by addition of 1 mL 80 % (v/v) ice-cold ethanol and repeating the centrifugation step. Eventually, the supernatant was discarded and the protein pellet was dried in a vacuum concentrator.

25 mM ammonium bicarbonate pH 8.0 was freshly prepared. Protein pellets were dissolved in 10 µL 1 % (w/v) RapiGest surfactant [212] in 25 mM ammonium bicarbonate. 10 µL 50 mM dithiothreitol, 25 mM ammonium bicarbonate were added and the solution was incubated at 37 °C for 1 h to reduce disulfide bridges. Free cysteines were carbamidomethylated by addition of 10 µL 100 mM iodoacetamide and repeating the incubation step in the dark. Next, 70 µL trypsin solution corresponding to a protein:enzyme ratio of 1:20 to 1:40 was added. Thus, the total volume was increased to 100 µL and the concentration of RapiGest surfactant was reduced to <0.1 % (w/v). Proteolysis took place at 37 °C for 12 h - 16 h. Detergent was decomposed by addition of $\frac{1}{5}$ v 5 % Trifluoroacetic acid (TFA), incubation at 37 °C for 2 h and precipitation of the dissociated detergent by centrifugation at 16,000 ×g for 30 min. The peptide-containing supernatant was transferred to a new tube, dried in a vacuum concentrator and stored at −20 °C.

Size Exclusion Chromatography for Enrichment of Cross-Linked Peptide Pairs Peptides from enzymatic hydrolysis of cross-linked proteins were dissolved in 60 µL 30 % acetonitrile, 0.1 % TFA by 15 s sonification and subsequent removal of solids by centrifugation at 13,000 ×g. Cross-linked peptide pairs were separated from unmodified peptides and peptides decorated with partially hydrolysed cross-linker molecules by size exclusion chromatography on a Superdex peptide 3.2/300 GL column mounted to an ÄKTA pure 25M chromatography system. For this, the dissolved

peptide mixture was loaded onto the column equilibrated in 30 % acetonitrile, 0.1 % TFA and separated isocratically at a flow-rate of 0.05 mL/min. 50 µL fractions were collected using an automated fraction collector. Fractions of high content of cross-linked peptide pairs as judged by light absorbance at 214 nm, 254 nm and 280 nm were pooled and the solvent was evaporated using a vacuum concentrator. Dried fractions were directly subjected to LC-MS/MS analysis.

Liquid Chromatography Tandem Mass Spectrometry Peptides were dissolved in 2 % (v/v) acetonitrile, 0.1 % (v/v) formic acid by sonication. Remaining solids were removed by centrifugation at 21,100 ×g for 3 min and the supernatant was transferred into 0.3 mL polypropylene autosampler vials. Peptides were separated on an Ultimate3000 RSLC high-performance LC system (ThermoFisher Scientific) directly coupled to a QExactive Plus hybrid quadrupol-orbitrap mass spectrometer (ThermoFisher Scientific). For this, peptides were loaded onto a precolumn (Acclaim PepMap C18, 0.3 mm x 5 mm, 5 µm particle size, 100 Å pore size) and washed at 10 µL/min flow-rate with 2 % (v/v) acetonitrile, 0.05 % (v/v) TFA for 3 min. Subsequently, peptides were eluted onto an analytical column (Acclaim PepMap C18, 75 µm x 50 cm, 3 µm particle size, 100 Å pore size) equilibrated in 0.1 % (v/v) formic acid. Separation of peptides on the analytical column was achieved using a linear gradient of 4 % - 90 % solvent B (80 % (v/v) acetonitrile, 0.1 % (v/v) formic acid) at a flow-rate of 0.3 µL/min over 90 min for peptides generated by in-gel hydrolysis or 120 min for peptides from in-solution hydrolysis. Analytical column temperature was maintained at 35 °C. Eluted peptides were directly transferred into an online nanoESI ion source with the transfer capillary held at 2.8 kV and 275 °C. Precursor ion scans were acquired from 350 m/z - 1600 m/z at a resolution of 70,000. The automated gain control target, i.e., the number of ions collected before injection into the orbitrap, was set to 1×10^6. The 20 ions with highest signal intensity in a precursor scan were selected for tandem MS and fragmented with 30 % normalised collision energy. The automated gain control target for fragment ion spectra was set to 1×10^5 and resolution for fragment ion scans was 17,500. Ions with less than two or more than seven charges, unrecognised charge states and ions with *m/z* ratios that were selected for fragmentation during the last 30 s (dynamic exclusion time) were excluded. For cross-linking experiments, ions with two charges were also excluded.

Database Search for Peptide Identification For protein identification by LC-MS/MS, rawfiles were processed with MaxQuant [121]. Carbamidomethylation of cysteine residues (a result of peptide reduction and alkylation) was specified as fixed modification and oxidation of methionine as well as acetylation of the N-terminus of the protein as variable modifications. Theoretical peptides were generated from sequence files in fasta format using trypsin as proteolytic enzyme and considering up to two missed cleavages. For initial identification of full-length Syt-1 after purification, the complete *E. coli* proteome and the sequence of Syt-1 were used as a basis for peptide

generation. For identification of soluble Syt-1 variants and p38, only the protein sequences of the target proteins were used. In both cases, MaxQuant automatically included common contaminations in the search. A 20 ppm mass tolerance for precursor ion mass and 0.5 Da tolerance for fragment ions was allowed for database search.

Identification of Cross-Linked Peptide Pairs Spectra from LC-MS/MS analysis of cross-linked proteins were loaded directly from Thermo raw files using the pLink2 software. Standard settings for BS2G or EDC cross-linkers were used and up to three missed cleavages were allowed during theoretical peptide library generation. For cross-linking experiments performed at 5 °C using BS2G, cross-links between serine, threonine and tyrosine residues were considered in a separate programme run. Spectra of all replicates belonging to a single experiment were processed together to increase reliability of the false-discovery rate estimation during the search. Peptide dimensions were set to minimum 6 residues (or 600 Da) and a maximum of 60 residues (or 6000 Da). Oxidation of methionine was specified as a variable modification and carbamidomethylation of cysteine as a fixed modification. MS1 and MS2 tolerance were set to 20 ppm with a final filter tolerance of 10 ppm. A false-discovery rate of 5 % was applied. pLink results were collected and filtered using the in-house written CroCo software [213] and only cross-links identified in all replicates of a cross-linker concentration were retained. For data analysis of in-gel hydrolysed peptides, only cross-links present in all replicates of a certain cross-linker concentration and molecular weight (i.e. monomer or dimer) band were retained after initial filtering. Briefly, CroCo parses results files of common cross-link search engines, converts their content into an internal database and saves the result either as table in csv format or as input for various cross-link visualisation and analysis software programmes [213]. Moreover, CroCo is implemented as a Python module enabling scripting of the workflows to ensure consistent analysis of multiple cross-link search runs. To decrease the number of mass spectra for manual validation, only the three highest scoring peptide-spectrum matches (PSMs) according to pLink2 score were retained for each cross-link. Resulting PSMs were manually inspected by converting the data tables with CroCo and peptide spectrum matches were visualised with the pLabel software. Only PSMs that were manually validated in at least two replicates of a cross-linker concentration and, for in-gel hydrolysis, molecular weight band were retained. Cross-links originating from consecutive peptides isobaric to a peptide with a missed cleavage and a hydrolysed cross-linker modification [214] were removed. When the same mass spectra were processed in two separate searches, i.e. the first considering only lysine cross-links and the second search considering only cross-links between serine, threonine and tyrosine, and both searches yielded an annotation for the same spectrum, only the PSM involving lysine-lysine cross-links was retained. All identified cross-link PSMs were again inspected using pLabel to ascertain consistent spectral quality. For visualisation, input files for xiNet [147] were generated.

3.2.5 Mass Spectrometry of Intact Proteins and Protein Complexes

Buffer Exchange Proteins were transferred into 200 mM ammonium acetate using size exclusion chromatography or filtration devices. For buffer exchange using size-exclusion columns, Micro Bio-Spin Columns with Bio-Gel P-6 were equilibrated by four washing steps with 200 mM - 600 mM ammonium acetate solution. For this, 500 µL of the solution were added to the column and removed by centrifugation at 1000 ×g and 4 °C for 1 min. Up to 50 µL of protein solution were applied to the column and protein in ammonium acetate solution was eluted by 4 min centrifugation at 1000 ×g and 4 °C. For buffer exchange using filtration devices, polyethersulfone membranes (Vivaspin 500) or regenerated cellulose membranes (Amicon Ultra-0.5) were used. The filtration device was prefilled with ammonium acetate solution and protein solution to reach 500 µL total volume was added. The diluted protein solution was concentrated at 4 °C by centrifugation at 12,000 ×g (Vivaspin) or 14,000 ×g (Amicon). Dilution and concentration were repeated 4 - 6 times. The final protein concentration was determined by absorbance at 280 nm and adjusted to <20 µM with ammonium acetate solution.

For native MS of full-length Syt-1, the protein purified in CHAPS detergent was first transferred into 200 mM ammonium acetate, 2 × critical micelle concentration (cmc) (0.5 % (w/v)) C8E4 using a Superdex 200 10/300 size-exclusion column. Elution of the protein was monitored by absorbance at 280 nm, protein concentrations were determined for every fraction and fractions with the highest Syt-1 concentration were selected for native MS analysis. For this, protein concentration in these fractions was increased to >15 µM using a 30 kDa cut-off polyethersulfone filter and centrifugation at 12,000 ×g and 4 °C.

Preparation of Native MS Electrospray Capillaries Thin-walled borosilicate glass capillaries were heated and extruded using a Micropipette Puller (see [166] for details). Capillaries were gold-coated using a plasma sputter coater operated at a current of 80 mA for 150 s and stored at room temperature before use.

Native Mass Spectrometry of Syt-1 Samples were loaded into in-house prepared, gold-coated borosilicate capillaries and analysed on a Synapt G1 HDMS quadrupole TOF mass spectrometer modified for transmission of high mass ions [170] operated in TOF mode or on a Q-TOF Ultima modified for the transmission of high mass ions. Typical acquisition parameters for native MS on the Synapt G1 were: Capillary Voltage 1.3 kV - 1.7 kV; Capillary Temperature 80 °C; Backing Pressure 4 mbar - 8 mbar; Sampling Cone Voltage 30 V - 80 V; Extraction Cone Voltage 5 V - 80 V; Collision Cell Pressure 1.25×10^{-2} mbar; Trap Collision Cell Voltage 10 V; Transfer Collision Cell Voltage 10 V - 30 V. Typical settings for analyses on the Q-TOF Ultima were: Capillary Voltage 1.3 kV - 1.7 kV; Capillary Temperature 80 °C; Inlet Pirani Pressure 2 mbar; Cone Voltage 35 V;

Collision Cell Pressure 8×10^{-3} mbar; Collision Cell Voltage 10 V - 30 V. Mass spectra were calibrated using masses of salt clusters formed by 100 mg/mL aqueous CsI solution and processed with MassLynx (V4.1) and UniDec [172].

Collision-Induced Unfolding Samples were prepared as described for native MS and analysed on the Synapt G1 HDMS instrument operated in mobility TOF mode. Trap collision voltage was stepwise increased from 8 V - 150 V to sequentially unfold the ions. Typical instrument settings differing from native MS analysis were: Sampling Cone Voltage 30 V; Extraction Cone Voltage 0.3 V; Trap Collision Cell Voltage 8 V - 150 V; Trap DC Bias 30 V; Trap Wave Height 0.2 V; IMS Pressure 0.5 mbar; IMS DC Entrance 10 V; IMS Wave Height 10 V; IMS Wave Velocity 300 m/s; IMS DC Exit 2 V; Transfer DC Entrance 2 V; Transfer Wave height8 V; Transfer Wave Velocity 200 m/s; Transfer DC Exit 9 V; Transfer Collision Cell Voltage 8 V.

Native MS Analysis of Soluble Proteins in the Presence of Detergents Stock solutions of 10 × cmc C8E4, lauryldimethylamine N-oxide (LDAO) and 5 × cmc octyl glycoside (OG) were prepared in 200 mM ammonium acetate solution (see **Table 4**). Alcohol dehydrogenase (ADH), β-lactoglobulin (β-LG) and BSA were transferred into 200 mM ammonium acetate using size-exclusion columns. Protein and detergent solutions were mixed to reach final detergent concentrations of 2 × cmc for C8E4 and LDAO or 1 × cmc for OG and a final protein concentration of 10 µM. Protein detergent mixtures were analysed on a Synapt G1 HDMS mass spectrometer and data were analysed as described.

Native Mass Spectrometry for the Study of Protein-Lipid Interactions For analysis of binding of soluble lipid head groups to Syt-1 C2AB, 1 mM stock solutions of inositol-1,3,5-triphosphate ($I(1,3,5)P_3$) and inositol-1,4,5-triphosphate ($I(1,4,5)P_3$) in water were prepared and stored at −20 °C. Syt-1 C2AB-His was transferred into 600 mM ammonium acetate using 10 kDa polyethersulfone membrane filtration devices. Protein concentration was determined by absorbance at 280 nm and the protein was mixed with ligand at a defined ratio. Mixed samples were directly analysed by native MS on a Synapt G1 HDMS mass spectrometer as described.

To test the interaction of β-LG with POPE, 1 mM POPE stock solutions in 200 mM ammonium acetate containing 2 × cmc C8E4, LDAO or OG were prepared by sonication for 30 min. β-LG was transferred into 200 mM ammonium acetate and mixed with the lipid stock to a final concentration of 10 µM protein and 100 µM - 500 µM POPE at 2 × cmc C8E4 or LDAO and 1 × cmc OG, respectively. Protein-lipid mixtures were analysed on a Synapt G1 HDMS operated at sampling and extraction cone voltages of 30 V and 0.3 V, respectively. Experiments were performed by Til Kundlacz.

For experiments of Syt-1 binding to intact lipids, a stock solution of 10 × cmc C8E4 in 200 mM ammonium acetate was prepared. Syt-1 C2AB, C2A or C2B were transferred into 200 mM ammonium acetate solution using regenerated cellulose membrane filtration devices as described. For analysis of Syt-1 C2AB-lipid interactions, defined amounts of di-oleyl-phosphatidylcholine (DOPC), di-oleyl-phosphatidylglycerol (DOPG), di-oleyl-phosphatidylinositol (DOPI), DOPS, PI(4)P, PI(4,5)P_2 or di-oleyl-phosphatidylinositol-3,4,5-trisphosphate (DOPI(3,4,5)P_3) were sonicated in 200 mM ammonium acetate solution for 30 min to form lipid vesicles. Syt-1 C2AB was mixed with detergent and lipid vesicles were added. Protein lipid mixtures were analysed on a Synapt G1 HDMS at sampling and extraction cone voltages of 80 V and 80 V. For analysis of Syt-1 C2A and C2B, DOPC, DOPG, DOPI, DOPS, PI(4)P or PI(4,5)P_2 were directly solubilised in 2 × cmc C8E4 in 200 mM ammonium acetate before they were mixed with the proteins at a final concentration of 10 µM. Protein-lipid mixtures were analysed on a Synapt G1 HDMS operated at sampling and extraction cone voltages of 30 V and 0.3 V, respectively. Comparative analyses of multiple lipids was performed by Til Kundlacz.

Spectrum Annotation for Native MS Experiments For spectrum annotation of native MS spectra using MassLynx (V 4.1), raw-files were loaded into the programme and spectra were combined by summation. Sums of spectra were smoothed 2 × with a 10 - 50 points Savitzky-Golay filter. Peak centres were determined as the top value of a curve with a minimum peak width at half height of 80 points. Multiple peak centres belonging to the same Gaussian peak envelope were chosen manually and two adjacent peaks were selected for mass determination. Typical parameters for mass determination were: Window: 5 Da; Threshold: 0.5 % base peak intensity; Reject: 3 sd's.

For comparative analyses, spectra were deconvolved using UniDec [172]. Typical parameters for deconvolution were: Charge range: 1+ - 20+; Mass Range: 5000 Da - 60,000 Da; Sample Mass Every: 10 Da; Smooth nearby Points: some; Suppress Artifacts: Some; Peak Detection Threshold: 0.0125 - 0.03.

Data Analysis of Native MS Protein-Lipid Interaction Experiments Spectra were first deconvolved using UniDec using the same settings for all spectra. The peak area of each mass peak was extracted from multiple deconvolution results using the Data Collector Utility of UniDec with the standard settings. Experimental outliers were removed by fitting a saturation function

$$y = \frac{xa}{x+b} \quad \text{or} \quad y = 1 - \frac{xa}{x+b}$$

to the extracted peak areas of lipid-bound or lipid-free peaks, respectively. Variables are: lipid concentration x, extracted peak area y, and fitting constants a and b determining the slope and asymptote of the function. Points with a distance larger than two standard deviations from the this fit were considered as outliers. If any data point of a measurement was contained an outlier, all data points of that measurement (i.e. those of lipid-free and lipid-bound protein peaks) were excluded from subsequent analysis. From the filtered dataset, apparent dissociation constant ($K_{D\,app}$) values were calculated using the UniFit routine of UniDec. In the UniFit algorithm, the ligand-binding reaction is represented as a graph with nodes defining the experimentally determined intensities for each species. Between these nodes, edges are defined that describe the reaction mediating the transition between the nodes (e.g. P + L $\rightleftharpoons$ PL) and, therefore, the respective K_D value (e.g. $K_D=\frac{[P][L]}{[PL]}$). As native MS analysis only informs on the protein in its ligand-free and ligand-bound state, the concentration of free ligand is unknown. Therefore, the UniFit routine optimises the free ligand concentration and the K_D value s iteratively. For this, free ligand concentration is first estimated based on guessed K_D value s. Next, it is minimised with respect to the experimental intensities by comparison with the known (total) protein and ligand concentrations. Subsequently, K_D value s are optimised using the free ligand concentration from the previous step. Minimisations of free ligand concentration and K_D value s are repeated multiple times. Uncertainty of the fit was calculated using 100 iterations of bootstrapping. Typical parameters were: numtotprotein: 1; numtotlig: 2 - 3; nummaxsites: 2 - 3; bootnum: 100; prot: 'one'; lig: 'free'. $K_{D\,app}$ values were calculated assuming that all ligand binding reactions were independent (lig: free), i.e. a separate $K_{D\,app}$ value was calculated for every protein-ligand stoichiometry.

Data Analysis of Collision-Induced Unfolding Experiments Analysis of CIU experiments and extraction of peak drift times and width was performed with Pulsar [200]. Spectra were loaded using the Import Wizard and clustered by collision energy. Masses were assigned to the peaks using the Mass Fitter utility. Unfolding plots were automatically generated by Pulsar and exported as image file. Drift times were exported as csv file for further processing using the Export CCS function.

To normalise drift times, low-intensity (<10 % base peak intensity) peaks were first removed. For each charge state and collisional voltage, the observed drift time was divided by the drift time observed at low activation energy (30 V and 10 V collision voltage for ADH and myoglobin, respectively) resulting in relative drift times. Eventually, relative drift times were averaged by calculating the weighted arithmetic mean and standard deviation (using the drift intensity as weight) for each protein-detergent combination.

3.2.6 Computational Analysis

Mapping of Cross-Links onto a Structural Model The list of cross-links was converted to a suitable input format for xlink analyzer [148] using custom python scripts and the CroCo [213] software. Structures of p38 dimers from crystallography (provided by Prof. Dr. Milton T. Stubbs) or from docking (this study) were loaded into UCSF Chimera [204] and cross-links were projected on the structure using the Xlink Analyzer software [148].

Cross-Linking-Guided Docking of p38 S- and P-domains The connecting loops between the S- and P-domains (residues 239 - 249) were removed using PyMol and new PDB-files corresponding to a single S-domain or the P-P dimer were generated. Cross-link tables containing all cross-links identified after treatment with BS2G in the presence and absence of RNA were converted into a suitable tbl-file using a custom python script and the CroCo [213] software to serve as input for HADDOCK [198]. Version 2.4 of the HADDOCK server was used as it supports docking of more than two molecules in a single run. The cross-links in the tbl-file were specified as unambiguous restraints and the „tight center of mass restraints“ option was selected. All other settings were at their default values. To calculate solvent-accessible surface distances (i.e. paths that are not blocked by the presence of protein atoms), the software Xwalk [206] was used. The residue numbers and chain identifiers of the amino acids connected by the loops were specified as input and the maximum calculated distance was specified as 100 Å.

Molecular Dynamics Simulations of p38 Structural Models The capsid dimer structure (provided by Prof. Dr. Milton T. Stubbs) and the two representative models from docking were transferred into a cubic box of 160 Å edge length filled with TIP3P water. Na^+ and Cl^- ions were added to a final concentration of 150 mM neutralising the charge of the protein. The Gromacs software [197] and the CHARMM36m force field [215] obtained from `http://mackerell.umaryland.edu/charmm_ff.shtml` were used. Steepest gradient energy minimisation was performed until the maximum force was <1000 $\frac{kJ}{mol\,nm}$. Two equilibration steps were performed, (1) in the NVT ensemble using the V-rescale modified Berendsen thermostat with separate coupling groups for protein and water for 500 ps and (2) in the NPT ensemble keeping the thermostat and adding isotropic pressure coupling by the Berendsen barostat for 5 ns. Next, simulation runs were performed using the Nosé-Hoover barostat with 1 ps time constant for water and ions and 2.5 ps for the protein. The pressure was isotropically controlled by the Parinello-Rahman barostat with a time constant of 5 ps. The particle mesh Ewald approximation with a 12 Å cut-off was used for calculating long-range electrostatics interactions and H-bonds were constrained using the LINCS algorithm. Simulations were carried out for a total of 300 ns simulation time.

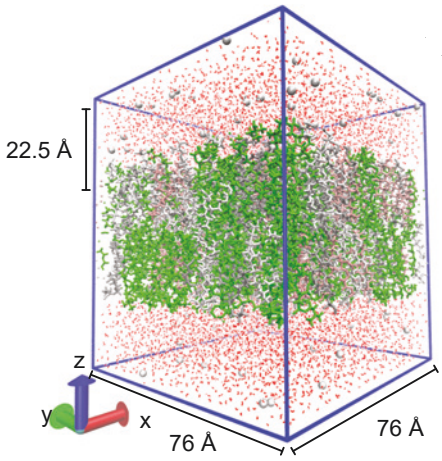


Figure 11: The SV Membrane Model Generated by CHARMM-GUI. POPC (green), DOPS, POPE (grey) and cholesterol (salmon) lipids, water molecules (red lines), K^+ and Cl^- ions (white) are shown.

Simulations of Soluble C2A and C2B Domains Initial coordinates of Syt-1 in the presence of Ca^{2+} were obtained from PDB ID 5KJ7 chain K. Coordinates for the atoms of the C2 domains (residues 141-263 for C2A and residues 272-418 for C2B) were extracted from the pdb file. To obtain starting coordinates of the Ca^{2+}-free structures, calcium ions were removed from the models. Syt-1 structures in the presence or absence of Ca^{2+} were placed in a box of TIP3p water and charge was neutralised with K^+ and Cl^- at a final concentration of 150 mM. Molecular dynamics simulations in the absence of lipids were performed using the Gromacs engine and the CHARMM36m force field. The particle mesh Ewald method with a 10 Å cut-off was used for calculating long-range electrostatics. H-Bonds were constrained using the LINCS algorithm enabling a time step of 2 fs. The system was relaxed until the maximum force was <1000 $\frac{kJ}{mol\,nm}$ and equilibrated using a two-step protocol. (1) A 500 ps equilibration in the NVT ensemble with all heavy atoms restrained by 1000 $\frac{kJ}{mol\,nm^2}$ was performed using the V-rescale modified Berendsen thermostat and separate coupling groups for protein and solvent. (2) The restrained simulation was continued in the NPT ensemble with pressure coupling calculated by the Berendsen barostat for 500 ps. Eventually, all heavy atom restraints were removed and the proteins were simulated for 10 ns in the NPT ensemble using the same settings. The last frame of this simulation was used for assembling the protein-lipid system (see below).

Simulations of Synaptic Vesicle and Plasma Membranes Lipid bilayers were simulated in aqueous solvent in the absence of protein. Lipid bilayer input files for Gromacs were created using CHARMM-GUI [196, 216] specifying a box size of 76 Å edge length and 22.5 Å thickness of the water layer (see **Figure 11**). For the SV membrane, 38 POPC molecules, 18 DOPS molecules, 20 POPE molecules and 20 cholesterol molecules per leaflet were used. The plasma membrane was

modelled using 38 POPC molecules, 18 DOPS molecules, 20 POPE molecules, 20 cholesterol molecules, 2 PI(4,5)P_2 (POPI25) molecules and 2 diacylglycerol (POGL) molecules per leaflet. 150 mM KCl and a simulation temperature of 300 K were specified. The force field parameters supplied by CHARMM-GUI were used for MD simulation of the lipids. Relaxation and equilibration of the lipid bilayer was performed using a protocol supplied by CHARMM-GUI. This included increasing the particle mesh Ewald cut-off for long-range electrostatics to 12 Å for all subsequent simulations. Phospholipid phosphate atoms and the hydroxyl oxygen of cholesterol positions were restrained by a force constant of 1000 $\mathrm{kJ/mol\,nm}$. The dihedral angle of the central C2 atom of the glycerol backbone of phospholipids was restrained to -120deg ± 2.5deg by a force constant of 1000 $\mathrm{kJ/mol\,rad^2}$. Similarly, the unsaturated C-C double bond of the fatty acyl chains was restrained to ensure planarity. The dihedral angle of the DOPS serine C_α was restrained to 120deg ± 2.5deg and the inositol ring carbons of PI(4,5)P_2 to a dihedral angle of 60deg ± 2.5deg. The restraints were gradually reduced during a set of six equilibration simulations. (1) A short 125 ps simulation in the NVT ensemble at 300 K using the V-rescale modified Berendsen thermostat and separate coupling groups for protein and solvent. (2) A 125 ps simulation with force constants reduced to 400 $\mathrm{kJ/mol\,nm}$ and 400 $\mathrm{kJ/mol\,rad^2}$, respectively. (3) A 125 ps simulation in the NPT ensemble using the Berendsen barostat to maintain 1 bar pressure and reducing the dihedral restraints to 200 $\mathrm{kJ/mol\,rad^2}$. (4) A 500 ps simulation with position restraints reduced to 200 $\mathrm{kJ/mol\,nm}$. (5) A 500 ps simulation with 40 $\mathrm{kJ/mol\,nm}$ position restraints and 100 $\mathrm{kJ/mol\,rad^2}$ dihedral restraints, respectively. (6) A 500 ps simulation with all heavy atom restraints released. Eventually, the system was simulated for 10 ns in the NPT ensemble without restraints using the Nosé-Hoover thermostat and the Parrinello-Rahman barostat. The last frame of this simulation was used for assembling the protein-lipid system.

Simulations of Synaptotagmin-Membrane Interactions Equilibrated lipid membrane structures were arranged in the *xy*-plane and the equilibrated protein was placed at 4 Å distance above the plane using vmd [205] controlled from a python script (`https://github.com/Eigenstate/vmd-python`). For each protein-membrane combination, three input structures with the C2 domains oriented at 0deg, 30deg and 60deg with respect to the z-axis were generated. In each of these structures, the CBLs of the protein faced towards the membrane and the tilting angle was calculated based on a vector from the centre of mass of the protein to the centre of mass of residue 172 of C2A or residue 305 of C2B (both located in the CBLs of the C2 domain), respectively. Potential clashes from assembly of the structures were removed by energy minimisation converging when the maximum force was below 1000 $\mathrm{kJ/mol\,nm}$. Subsequent equilibration followed the six-step protocol supplied by CHARMM-GUI and described above for the lipid bilayer. The production run was performed in the NPT ensemble using the Nosé-Hoover thermostat to maintain 300 K temperature with a coupling constant of 1 ps for water and ions and 2.5 ps for protein and lipid groups, respectively. The pressure was controlled using the Parinello-Rahman barostat employing

semi-isotropic coupling with a time constant of 5 ps. Initial equilibration of the individual proteins and lipids and the first four equilibration steps of the protein-lipid systemwere calculated on an AMD Ryzen 7 3700X processor running Ubuntu 20.04 and Gromacs version 2020.4. The last two steps of the equilibration of the protein-lipid system and production simulations were computed on an Intel Xeon E5-2680v3 processor running CentOS 7 and Gromacs version 2020.4 until >900 ns simulation time were accumulated. If not stated otherwise, the last 200 ns of each simulation were used for analysis.

Analysis of Structural Dynamics Analysis of the simulation was performed using MDAnalysis [217], a python module for the analysis of MD simulations. Trajectories of 1 ns time step per frame were generated with Gromacs and loaded into MDAnalysis. For root-mean-square deviation (RMSD) analysis, a rotational and translational alignment was performed by comparing the positions of the C_α atoms of the protein with the respective positions in the first frame of the trajectory. For each comparison the RMSD was calculated using the following equation with the position x_i of an atom i and the position x_i^{ref} of the same atom in the reference frame.

$$RMSD(x, x^{\text{ref}}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} |x_i - x_i^{\text{ref}}|^2)}$$

The lowest RMSD obtained during alignment was returned. RMSDs for Syt-1 MD simulations were averaged over all three protein starting orientations.

For root-mean-square fluctuation (RMSF) calculation, an average protein structure was calculated by RMSD-guided alignment of all frames of the last 200 ns of each simulation to a frame in the middle of the time range. The protein structure of every frame of the trajectory was then aligned to the average protein structure. The RMSFs of the protein C_α positions were calculated according to the following equation with the coordinates X_i of each C_α atom index i and the coordinates of the same atom in the averaged structure $\langle X_i \rangle$.

$$RMSF(X_i) = \sqrt{\langle (X_i - \langle X_i \rangle)^2 \rangle}$$

To calculate RMSDs of a flexible region defined by RMSF analysis, residues flanking the flexible region that showed low fluctuation in RMSF analysis were RMSD aligned with respect to the first frame of the simulation. RMSDs of the flexible region were calculated as described.´

Minimum and Centre of Mass Distances Distances between all C_α atoms were calculated using the distance_array algorithm implemented within MDAnalysis using the minimum image convention for calculating distances. The minimum of the calculated distances was then calculated for each simulation step. Centre of mass distances were calculated similarly, using the centres of mass of C_α atoms as input for distance_array.

Angle Calculation The membrane interaction angle of the C2 domains during the simulation was calculated between the z-axis and a vector connecting residues Tyr151 and Pro179 in C2A or Tyr282 and Pro 310 in C2B. The tyrosine and proline residues are located at homologous positions at the periphery of the β-sandwich fold in both C2 domains with the proline residues located closer to the CBLs. The angle α was calculated by the following equation using a normalised vector $\hat{z}$ parallel to the z-axis and the normalised Pro-Tyr vector $\hat{v}$.

$$\alpha = \frac{180}{\pi} * cos^{-1}(\hat{z} \cdot \hat{v})$$

To average angles from simulations of the three replicates of each protein membrane combination, the median angle was calculated for every initial orientation, assuming that protein-lipid interactions were stabilised at this time point. The mean and standard deviations of the three medians angles were calculated.

Protein-Lipid Contact Analysis To determine protein-lipid contacts, the closest lipid molecules for every protein residue were determined for every frame of each simulation. For this, a k-dimensional tree analysis implemented in scipy [202] initialised with the positions of the lipid atoms was used. The tree was queried with the positions of the protein atoms specifying that only nearest neighbours with a maximum distance of 3 Å are returned. The resulting list of atom-atom contacts was filtered for contacts with a distance of less than 2.6 Å, the distance of a hydrogen bond. Duplicates of protein-lipid contacts (i.e. the same combination of protein residue number and lipid molecule) were removed. To exclude transient contacts, protein-lipid contacts that were not observed continuously for a time span of at least 10 ns were not considered. To compare the number of lipid contacts between simulations, counts for each lipid were divided by the sum of counts of all lipids in a simulation. These normalised counts were then, for each lipid species, divided by the relative number of lipids, i.e. the number of lipids of this type relative to the total number of lipids in the membrane. The values were summed up for every simulation, resulting in a total number of lipid contacts per simulation per lipid. The total numbers of protein-lipid contacts were averaged across the three replicates and standard deviations were calculated. Structural representations of lipid contacts were generated by summing the raw lipid contacts per residue of the three replicate simulations. The summed lipid contacts were saved as structure factor together with the coordinates of C2A and C2B before the simulation. Visualisation was performed with PyMol.

Membrane Penetration Analysis To analyse membrane penetration, lipid molecules belonging to the upper or lower leaflet of the membrane bilayer were determined using the LeafletFinder algorithm [217]. For every 1 ns step, all atoms were translated so that (1) the C2 domain was centred laterally in the box and (2) the centre of mass of the phosphor atoms of one leaflet was at a defined z-coordinate z_{mem}. Penetration depth was then calculated as the distance between z_{mem} and the C2 domain C_α atom with the lowest z-coordinate. For simulations in which the protein interacted with the lower leaflet, the same procedure was repeated for the second leaflet. This is necessary as periodic boundary conditions allow the protein to move away from the upper membrane leaflet, leave the simulation box and enter again below the lower leaflet. For comparison, distances were normalised to that values >0 that indicate the protein is located within the lipid bilayer. Distances calculated from different starting orientations were plotted in separate histograms (see **Figures S14 and S15**). To summarise behaviour of multiple simulations starting with different orientations of the protein, histograms of all corresponding simulations were summed (see **Figure 38**).

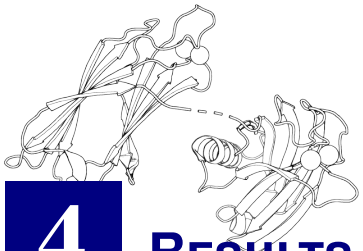
Analysis of Membrane Bending and Curvature To determine membrane curvature, all atoms of simulations with Syt-1 C2 domains were translated for each frame of each trajectory so that (1) the C2 domains were centred laterally and (2) the membrane leaflets were located completely within the simulation box. For control simulations without protein, only the leaflets were translated completely inside the box. Lipid atoms were wrapped inside the box for both sets of simulations. Surface positions, Gaussian curvature and mean curvature of phospholipid phosphor atoms were calculated using the Membrane Curvature tool of MDAnalysis specifying binning into 8 x 8 bins. Briefly, Gaussian curvature K of a surface at certain point is the product of the two principal curvatures (a measure how the surface bends in a direction) at the point. In contrast, the mean curvature H of a point is the mean of the principal curvatures.

$$K = k_1 * k_2 \qquad H = \frac{1}{2}(k_1 + k_2)$$

For comparison between simulations, surface positions were reported relative to the lowest phosphor atom coordinate. To determine the variation in orientation of the C2 domain during the analysis time range, the angle between a projection on the xy plane of the vector connecting residues Tyr151 and Pro179 in C2A or Tyr282 and Pro310 in C2B and the y-axis was calculated for every frame and minimum and maximum angles were reported.

Simulation of Protein and Detergent Concentrations in Electrospray Droplets Following an approach from Benesch et al. [173], the numbers of protein molecules and detergent micelles in electrospray droplets of 18 nm diameter were simulated. For this, a number of proteins and detergent micelles corresponding to protein concentrations of 0 µM - 250 µM and micelle concentrations of 0 µM - 250 µM were randomly placed in a volume of 1 μm^3. 100,000 points in the volume were

randomly chosen and the numbers of protein molecules and detergent micelles within a radius of 9 nm of these points were calculated using a k-dimensional tree analysis implemented in scipy [202]. Expected ratios between lipid-bound and lipid-free proteins α was calculated by dividing the number of droplets containing one protein and at least one detergent micelle by the number of droplets containing only one protein.



4 Results

4.1 Protein-Lipid Interactions of Synaptotagmin-1

To investigate lipid preferences of Syt-1, binding of the individual soluble C2 domains to lipids was characterised. In contrast to the analysis of soluble proteins such as peripheral membrane proteins or the Syt-1 C2 domains, the analysis of lipid interactions of integral membrane proteins by native MS is well established. A membrane protein solubilised with an MS compatible detergent, is transferred into the gas phase by electrospray ionisation and the detergent is removed by high-energy collisional activation after ionisation [175]. Ideally, co-purified lipids of structural importance remain tightly bound after removal of detergents and are observed in the mass spectrum as additional peaks shifted to higher masses [218, 126]. However, detergents not only function as inert barriers shielding transmembrane regions from the aqueous environment, but likely also play a role in mediating the exchange between protein-bound and free lipids [177]. Important questions on the way to interrogating Syt-1-lipid interactions were if the presence of detergents has any, and if so, what effects on soluble proteins and which detergent is best suited for mediating lipid transfer to a soluble protein[1].

4.1.1 Native Mass Spectrometry of Soluble Proteins in the Presence of Detergents

Three detergents commonly employed in native MS of membrane proteins, namely C8E4, LDAO and OG (see **Figure 12A** for structural formulae) were selected and their effects on three soluble proteins were investigated. These were the tetrameric ADH complex, β-LG, a known lipid-binding protein from milk [220], and BSA that contains solvent-exposed hydrophobic patches. The standard proteins were mixed with the detergents to reach a protein concentration of 10 µM and a final detergent concentration of 2 × cmc for C8E4 and LDAO or 1 × cmc for OG. The cmc is the concentration

[1]The work on protein-detergent interactions was published in Til Kundlacz*, **Julian Bender***and Carla Schmidt. Effects of non-ionic and zwitterionic detergents on soluble proteins during native mass spectrometry experiments. International Journal of Mass Spectrometry (2021), doi:10.1016/j.ijms.2021.116652
*equally contributing

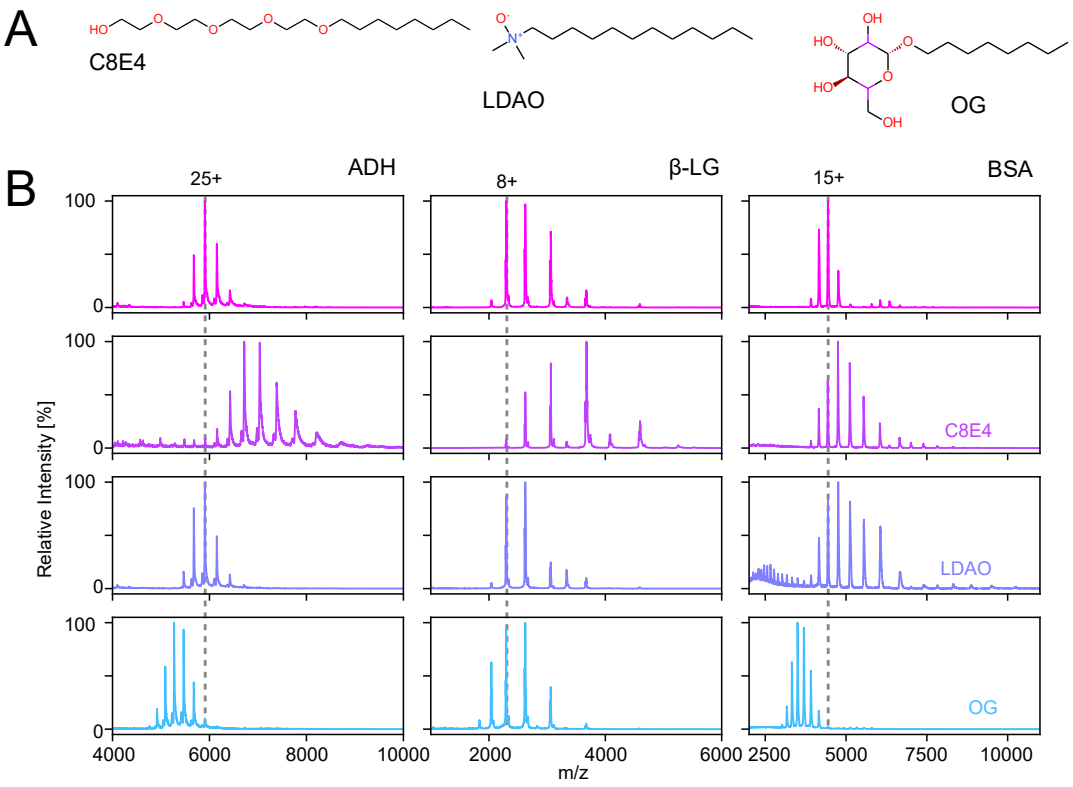


Figure 12: Effects of Detergents on the Charge Distribution of Standard Proteins. (**A**) Structural formulae of the employed detergents. (**B**) Native mass spectra of 10 µM ADH, β-LG and BSA acquired in the absence (top row, pink) or presence of 2 × cmc of the detergents C8E4 (second row, magenta), LDAO (third row, blue) or 1 × cmc of OG (bottom row, cyan). The charge state of the peak with the highest intensity for each detergent-free form is indicated.

above which detergents form micelles that are required for solubilisation of hydrophobic molecules. Analysis of membrane proteins by native MS is generally performed at detergent concentrations above the cmc [175]. The following experiments were performed together with Til Kundlacz.

Detergents Modulate Protein Charge **Figure 12B** shows native mass spectra of the different protein-detergent combinations. In the absence of detergent, ADH and BSA show symmetric peak distributions with an intensity maximum at the 25+ and 15+ charge state, respectively. β-LG showed a slightly asymmetric peak distribution with a maximum at the 8+ charge state. For β-LG and BSA, a low intensity peak distribution corresponding to dimers of the proteins was observed. In the presence of 2 × cmc of C8E4, the highest intensity charge states shifted to lower charges for all three proteins. The main charge state of ADH was 22+ (25+ without detergent), of β-LG monomer 5+ (8+ without detergent) and of BSA monomer 14+ (15+ without detergent). Moreover, the observed peak distributions of ADH and BSA broadened in the presence of C8E4, i.e. the number of peaks

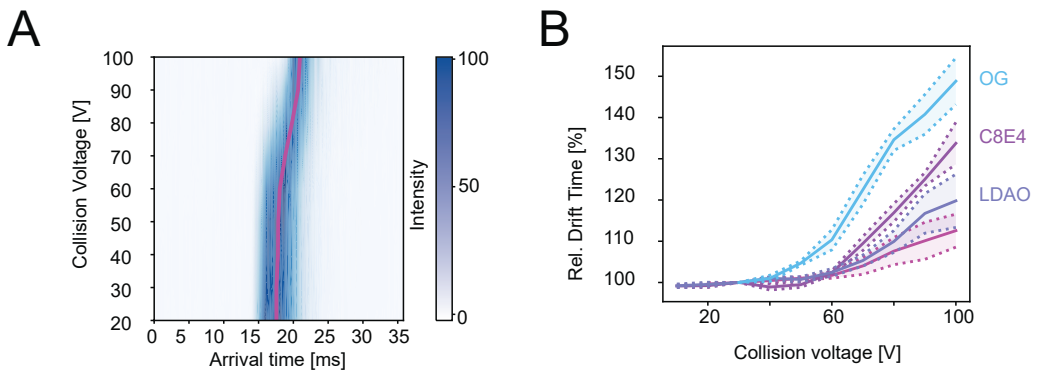


Figure 13: Collision-Induced Unfolding of ADH in the Presence of Detergents. (**A**) Collision-induced unfolding of the 8+ charge state of ADH in the absence of detergent. Drift time distributions of the ion were recorded at various activation energies. A shift to higher drift times indicates unfolding of the protein at higher collisional activation. (**B**) Relative drift times of ADH normalised to the respective drift times at 30 V collision voltage in the presence of the three detergents. Mean values (solid line) and single standard deviation (dashed line and shaded area) are indicated.

increased from four to seven for ADH and from five to eight for BSA monomer. Broadening of the distribution was less pronounced for β-LG, possibly related to the low overall charge of the protein. Native MS analysis in the presence of LDAO showed that the most intense charge state of ADH remained 25+, the main charge state of β-LG decreased slightly from 8+ to 7+ and the main charge state of BSA shifted from 15+ to 14+. This is in line with previous observations showing that LDAO acts as a charge-reducing detergent [176], although the effect on soluble proteins is lower than on membrane proteins. In contrast to its low effect on the charge state of the most intense peak, LDAO showed substantial influence on the width of the peak distribution of BSA that included more charge states in the presence of the detergent than in its absence, similar to the effect of C8E4. However, a series of highly charged BSA ions was observed at low *m/z* indicating the unfolding of the protein in the presence of LDAO. Lastly, the effect of 1 × cmc OG on the standard proteins was investigated. In the presence of this detergent, the main charge states of the main peaks increased for ADH by three (to 28+) and for BSA by four (to 19+). Again, only a minimal effect was observed on the charge state distribution of β-LG for which the main charge state was shifted from 8+ to 7+. Therefore, OG overall acts as a charge-increasing additive during native MS analysis.

Protein Stability is Influenced by Detergents To further investigate how detergents influence on protein stability, IMS of the standard proteins in the presence and absence of detergent was employed. **Figure 13A** shows a CIU experiment of the 8+ state of ADH in the absence of detergent. For this, the collision voltage in the trap cell of the mass spectrometer was increased stepwise from 10 V - 100 V in separate experiments. After passing the collision cell, the ions were separated by

IMS and their drift time at the detector was recorded. Collision-induced unfolding was detected by a shift in drift time due to the increase in CCS of the unfolded protein. For the 8+ charge state of ADH in the absence of detergent, unfolding was detected at approximately 75 V. At this voltage, the main drift time shifted from approximately 17.5 ms to 21 ms indicating the unfolding of the ion. Ideally, comparison of drift time distributions of the same charge state in the presence of the different detergents enables direct assessment of their stabilising and destabilising properties. However, the charge-modulating effects of the detergents complicated comparison as similar charge states were different in intensity. To compare the stabilising effect of detergents nevertheless, the increase in drift time was calculated for every charge state relative to the drift time at low collisional activation (see **Section 3.2** for details). The increases of relative drift time of all charge states were then averaged, allowing a comparison of the average collision voltage at which unfolding occurs. Moreover, drift times are typically converted to CCSs using standard proteins of known CCS for calibration to compare the results with theoretical CCSs of structural models. However, the presence of detergent required sampling and extraction cone voltages of 80 V which interfered with reliable drift time calibration. The published correlation of mass-dependent flight time (t'_d) and theoretical CCSs corrected for charge (Ω') [200] could not be reproduced using these settings. Nonetheless, absolute CCSs are not necessary for assessment of relative stabilising or destabilising effects of the detergents.

A plot of the normalised drift time of ADH in the presence and absence of the three detergents shows an increase in drift time above approximately 70 V in the absence of detergents (**Figure 13B**).´ Relative drift times for LDAO and C8E4 increased at a similar collision voltage. The relative drift time of ADH in the presence of OG, however, increased at lower collision voltages >40 V. This indicates that in the presence of OG, ADH is more like unfold under collisional activation, possibly related to the supercharging effect of the detergent as extensive supercharging leads to electrostatically-driven unfolding after transfer into the gas phase [221, 222]. It was suggested that Coulombic repulsion is prevented in the electrospray droplet due to solvation effects [221]. However, charge repulsion after desolvation may lead to metastable ions. As the degree of destabilisation is proportional to the charge density, it is likely that ADH in the presence of OG is destabilised to a degree that does not lead to unfolding at standard collision voltages but only at higher collisional activation energies employed during collision-induced unfolding. Presumably, the high charge density on the surface of ADH increases the likelihood of protein unfolding by charge migration. Thereby, locally unfolded regions are stabilised as they increase the distance between charges and minimise Coulombic repulsion [173].

The analysis of the standard proteins in the presence of detergents indicated that C8E4 has a stabilising effect on proteins in the gas phase. Accordingly, the question if this stabilising effect also includes the preservation of non-covalent interactions was addressed next. Binding to co-factors is often mediated by multiple contacts at a defined spatial orientation. Therefore, conformational

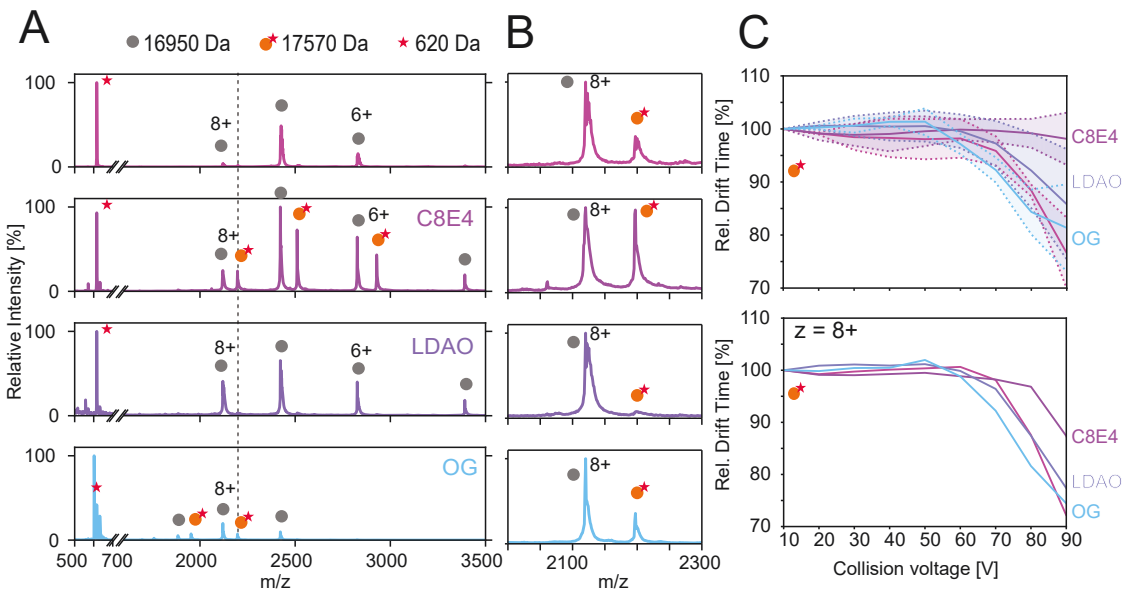


Figure 14: Stabilisation of Holo-Myoglobin by Detergents. (**A**) Native MS of myoglobin in the absence (pink) and presence of 2 × cmc of the detergents C8E4 (magenta) and LDAO (purple) or 1 × cmc of OG (cyan). Charge state distributions corresponding to the mass of the heme-free apo-myoglobin (grey spheres), the isolated heme group (red stars) and the heme-containing holo-myoglobin (yellow spheres with red stars) are indicated. **(B)** Zoom on the 8+ state shown in (A). (**C**) Average drift times of holo-myoglobin normalised to the respective drift times at 10 V in the absence and presence of the three detergents (top). Drift-time distributions of the 8+ charge state (bottom). Note that intensities for this charge state vary between the different detergent conditions (dashed line in (A)).

stability during MS analysis was investigated by monitoring co-factor binding. For this, myoglobin, a small globular protein involved in oxygen transport in muscular tissue stably binding a heme as prosthetic group [223], was chosen. **Figure 14A** shows the native mass spectrum of myoglobin in the absence and presence of detergents. In the absence of detergent, a charge state distribution corresponding to the mass of heme-free myoglobin of 16.95 kDa and a single peak corresponding a singly charged free heme group (617 Da) was observed. Moreover, a second peak distribution corresponding to the mass of heme-bound protein (17.57 kDa) was detected, albeit at low intensity (**Figure 14B**). In the presence of C8E4, the mass spectrum showed two charge distributions corresponding to heme-free and heme-bound myoglobin at approximately the same intensity. Interestingly, presence of the detergent did not shift the charge state distribution significantly towards lower charges. Similar to the low effect on β-LG, preservation of the main charge state in the presence of C8E4 is likely a consequence of the overall low charge of myoglobin. In the presence of OG, heme-free and heme-bound myoglobin were detected. However, the charge state distribution was shifted to higher charges (i.e. lower *m/z* values). Moreover, the intensity of holo-myoglobin was lower than in the presence of C8E4. In contrast to the other detergents, LDAO had no effect on the

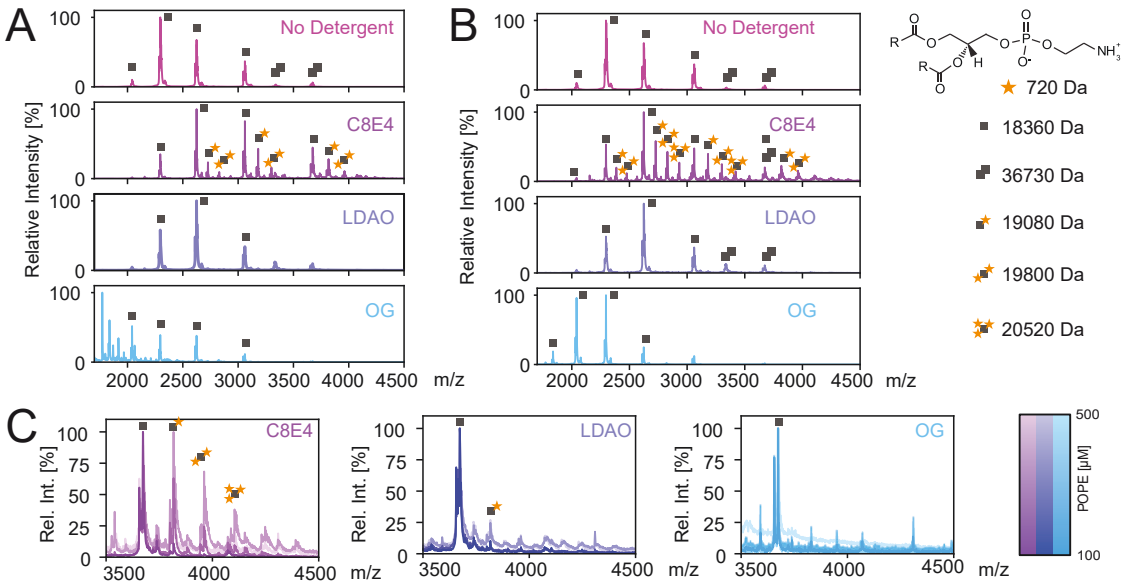


Figure 15: Stabilisation of β-LG-Lipid Interactions by the Detergent C8E4. **(A)** Native MS of 10 µM β-LG and 100 µM POPE in the absence (top) and presence of 2 × cmc of the detergents C8E4 and LDAO or 1 × cmc of OG. Peaks corresponding to the lipid-free protein (black squares) and to β-LG binding to one or two POPE molecules (black squares with one or two stars) are shown. **(B)** Native MS of β-LG and 200 µM POPE in the absence of presence of detergents. See (A) for details. **(C)** Monitoring the 5+ charge state of β-LG in the presence of detergents and increasing POPE concentrations of 100 µM - 500 µM shows peaks corresponding to up to three lipid bound to β-LG in the presence of 2 × cmc C8E4 (left), up to one lipid bound in the presence of 2 × cmc LDAO (middle) and no protein-lipid peaks the presence of 1 × cmc OG (right).

stabilisation of the holo-protein. A peak distribution corresponding to the heme-free protein was almost exclusively present in the spectrum. Similar to the analysis of the effects of the detergents on ADH, CIU experiments were performed with myoglobin and relative drift times were calculated (**Figure 14C**). In contrast to the unfolding events seen for ADH, compaction (i.e. a decrease of drift time) of the protein upon collisional activation was observed for myoglobin, likely a result of the protein collapsing once the heme-group was displaced from the binding pocket. This collapse of myoglobin occured at 60 V - 70 V collision voltage in the absence of detergent and in the presence of LDAO and OG. In the presence of C8E4, however, the relative drift time remained stable up to 90 V collision voltage. As the charge-modulating effects of the detergents on myoglobin were lower than those observed for ADH, drift times of the 8+ charge state were compared directly (**Figure 14C**). in the presence of C8E4, holo-myoglobin was observed at collison voltages <80 V, whereas it collapsed at voltages exceeding 50 V - 60 V in the presence of LDAO and OG or in the absence of detergent, similar to what was observed using relative drift times. In summary, based on the native MS and IMS analysis, C8E4 and OG stabilise holo-myoglobin, whereas LDAO showed no stabilisation of tertiary structure of the protein.

C8E4 Mediates Lipid-Transfer to Soluble Proteins Native MS analysis of myoglobin and ADH in the presence of C8E4 showed stabilising effects of the detergent on the proteins. Therefore, stabilisation of non-covalent protein-lipid interactions by the detergent was investigated next. A facilitating effect of detergents on the transfer of lipids to soluble proteins has been described previously, however, this effect was considered to be unspecific [177]. In this thesis, β-LG that specifically binds to hydrophobic molecules such as POPE [224] was added. POPE was first solubilised with C8E4 and than the protein was added to the mixed lipid detergent micelles (see **Section 3.2** for details). **Figure 15A** shows the mass spectra of β-LG and POPE in the presence of the detergents. Mass spectra for LDAO and OG were similar to those acquired in the absence of lipids and did not show protein-lipid complexes. However, in the presence of 2 × cmc C8E4, peaks corresponding to the masses of up to two bound POPE molecules were observed. When the POPE concentration was increased from 100 µM to 200 µM, the number of lipid molecules binding to β-LG increased in the presence of C8E4 to up to three molecules (**Figure 15B**). The measurement was repeated with increasing lipid concentrations up to 500 µM (**Figure 15C**), however, no protein-lipid complexes were observed in the presence of OG and only one bound lipid was detected in the presence of 500 µM LDAO. This indicates that gas phase lipid binding is indeed facilitated in the presence of C8E4 via a so far unknown mechanism. The experiment also demonstrates that the lipid transfer efficiency to soluble proteins depends on the class of detergent present.

In summary, the analysis of the effects of detergents on soluble proteins indicates that C8E4 stabilises protein tertiary and quatenary structure and also mediates non-covalent protein-lipid contacts in the gas phase. Therefore, C8E4 was selected as additive for the analysis of Syt-1-lipid interactions in the following experiments.

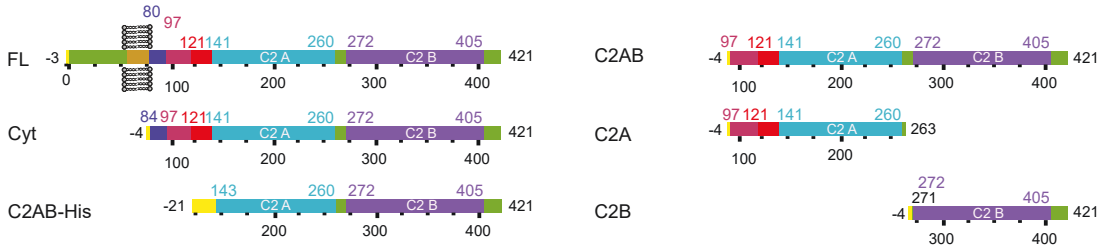


Figure 16: Schematic Representation of Syt-1 Variants Used in This Thesis. N-terminal extensions remaining from protein expression and purification (yellow), residues belonging to the vesicular region (green), the transmembrane helix (orange), the intrinsically disordered linker (containing a basic region (blue), multiple KD-repeats (dark red) and an acidic region (red)), the C2A (cyan) domain, the connecting linker (green), the C2B (purple) domain and the C-terminal flexible region (green) are shown.

4.1.2 Expression and Purification of Synaptotagmin-1

Syt-1 consists of a short N-terminal lumenal region, a transmembrane helix that anchors the protein to the SV membrane and a large cytoplasmic part. The cytoplasmic part of Syt-1 is highly flexible and dynamic. It is composed of two folded C2 domains (termed C2A and C2B) that are connected with each other and the transmembrane helix through flexible linkers (see **Figure 16**). Adjacent to the transmembrane helix on the cytoplasmic side, five palmitoylated cysteine residues of structural importance for Syt-1 oligomerisation were reported [225, 226]. The intrinsically disordered linker connecting C2A and the transmembrane helix (residues 80 - 141) contains many charged residues [47]. More precisely, this region is composed of a basic N-terminal region (residues 80 - 96), a central region with lysine-aspartic acid repeats (residues 97 - 120) and a C-terminal acidic region (residues 121 - 141).

Purification of Full-Length Synaptotagmin-1 and Truncated Variants *E. coli* cells were transformed with a plasmid encoding full-length Syt-1 from *Rattus norvegicus* (residues 1 - 421) fused to a hexahistidine tag for purification (see **Figure S2** for a representative plasmid map, plasmid kindly provided by Prof. Reinhard Jahn, Göttingen). To analyse lipid-interactions of the C2 domains in the absence of the membrane anchor, the construct was shortened generating two variants, namely Syt-1 Cyt, which contains the intrinsically disordered linker, both C2 domains with their interconnecting linker and the C-terminal flexible region (residues 84 - 421), and Syt-1 C2AB-His which contains only the connected C2 domains and the C-terminal flexible region (residues 143 - 421). Moreover, an established construct termed C2AB and equivalent to Syt-1 Cyt with a shortened N-terminal linker (residues 94 - 421) and two shorter variants, C2A encoding for the C2A domain and part of the linker region (residues 97 - 263) and C2B encoding for the C2B domain and the C-terminal flexible region (residues 271 - 421) were obtained (all plasmids kindly provided by Prof. Reinhard Jahn, Göttingen).

To investigate lipid-interactions of Syt-1, plasmid-encoded protein variants were recombinantly expressed and purified (see **Figure 17**). Purification was achieved using metal affinity chromatography with either Ni^{2+}-NTA coated agarose beads or with commercially available Ni^{2+}-NTA columns for the full-length protein and the soluble variants, respectively. First, *E. coli* cells expressing the full-length protein were lysed in the presence of cholate detergent, whereas cells expressing the soluble variants were lysed without detergent. Second, proteins fused to a hexahistidine tag were captured by affinity purification. Full-length Syt-1 was immobilised using agarose beads decorated with Ni^{2+}-NTA as described [72]. The hexahistidine tag used for purification was then removed by thrombin cleavage. In contrast, capture and enrichment of soluble variants was achieved using commercially available immobilised Ni^{2+}-NTA affinity columns. The hexahistidine tag was removed and the protein was loaded again on the same Ni^{2+}-NTA affinity column. Thereby, proteins binding

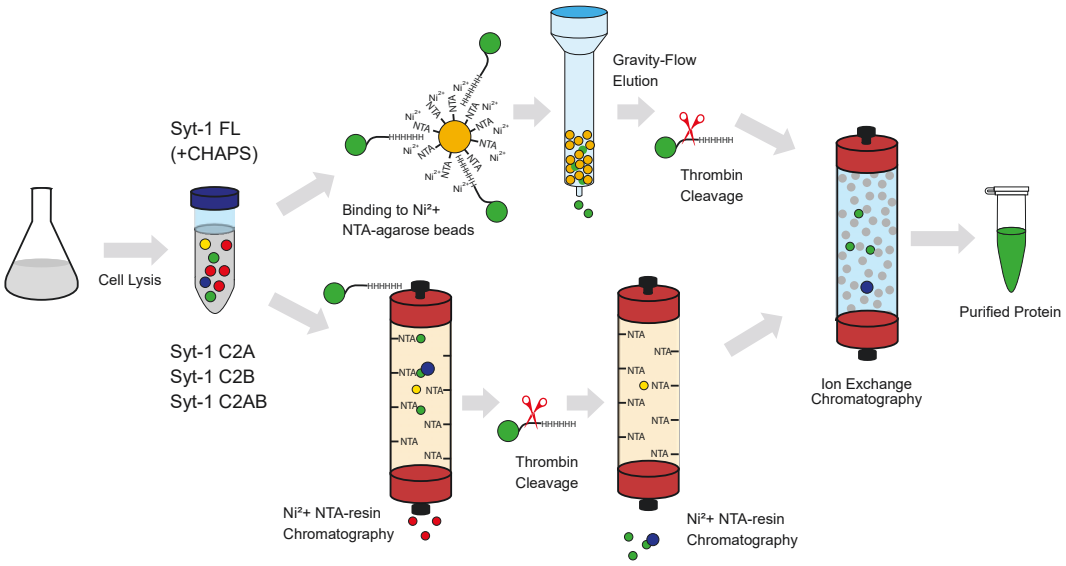


Figure 17: Purification Workflow for Syt-1 Variants. Full-length Syt-1 was purified using immobilisation on agarose beads via a hexahistidine tag and subsequent manual elution from the beads (top). The tag was removed by thrombin cleavage after elution. Soluble variants were purified via a two-step protocol using fast protein liquid chromatography and metal-affinity purification columns (bottom). For this, the tag was removed between two rounds of chromatography. Eventually, all proteins were subjected to ion-exchange chromatography (right) to remove bound nucleotides.

non-specifically to the column and uncleaved Syt-1 were retained and removed from the purified protein. Third, full-length and soluble variants were subjected to ion-exchange chromatography before protein solutions were concentrated and stored.

The purification of Syt-1 variants is exemplified in **Figure 18**. After immobilisation of full-length Syt-1 on agarose beads and removal of low-affinity binders, the protein was eluted from the beads by addition of imidazole competing with the hexahistidine tag for complexation of Ni^{2+} ions. **Figure 18A** shows a stained polyacrylamide gel of fractions of the affinity purification. Gel bands corresponding to the mass of Syt-1 at 49 kDa molecular weight are visible in the elution fractions 4 - 16. After thrombin cleavage, a shift of this band to lower mass was observed indicating successful removal of the hexahistidine tag (**Figure 18A**, right). Ion strength was reduced by dialysis and contaminants were removed by strong cation-exchange chromatography (**Figure 18B**). To verify the identity of the purified protein, western blots of the proteins after affinity chromatography were performed using antibodies directed against the cytoplasmic part of Syt-1 and against the hexahistidine tag (**Figure 18C**). Both antibodies confirmed the presence of Syt-1. Note that the antibody directed against the hexahistidine tag did not show a signal after thrombin cleavage, confirming the successful removal of the hexahistidine tag. This indicates that the 49 kDa protein is indeed Syt-1 and that thrombin cleavage was successful.

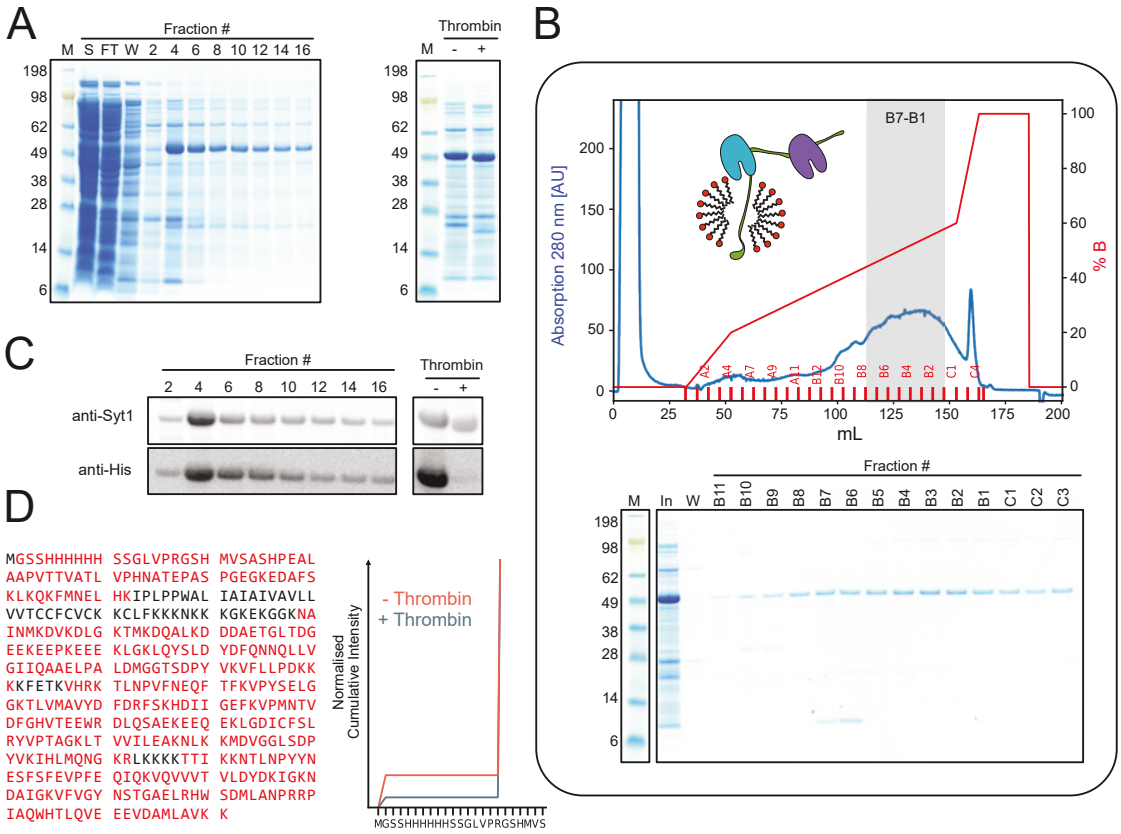


Figure 18: Purification of Recombinantly Expressed Full-Length Syt-1. **(A)** Ni^{2+} affinity purification of the cell lysate containing hexahistidine tagged Syt-1 in the presence of CHAPS. The tagged protein was immobilised on Ni^{2+}-NTA agarose beads. Washing steps and protein elution were performed manually by gravity-flow chromatography (left panel). The hexahistidine tag was removed by thrombin cleavage (right panel). **(B)** Contaminations were removed from the protein by strong cation exchange chromatography. The chromatogram (top) and the gel of fractions B11-C3 (bottom) are shown. In: Input; M: Marker; W: Wash. **(C)** Western blot analysis of fractions from Ni^{2+} affinity purification and before and after thrombin cleavage. Binding of antibodies binding to the cytoplasmic part of Syt-1 (top) and to the hexahistidine tag (bottom) are shown. **(D)** Protein sequence coverage (left panel) determined after tryptic hydrolysis, LC-MS/MS analysis and database search. Residues identified by at least one peptide are coloured in red. Quantification of N-terminal peptides (right panel) shows lower intensities of residues corresponding to the hexahistidine tag after (blue) than before (red) thrombin cleavage.

´ Protein identity and thrombin cleavage were also confirmed by LC-MS/MS (**Figure 18D**). 87 % sequence coverage indicate that the protein was completely expressed and corresponds to the expected amino acid sequence. Surprisingly, residues of the removed hexahistidine tag were identified after thrombin cleavage. Quantification of residue intensities by summation of the

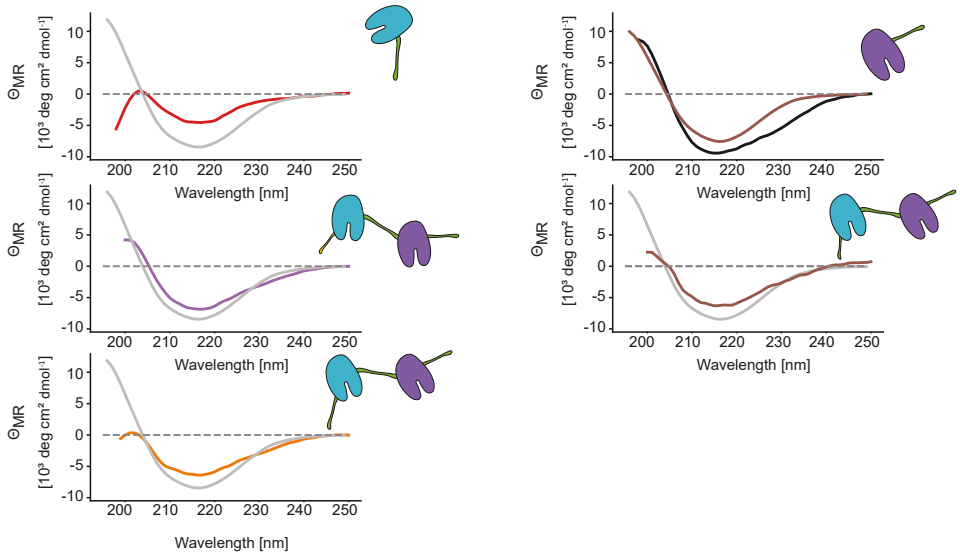


Figure 19: CD Spectroscopy of Soluble Syt-1 Variants. Mean residue ellipticities (θ_{MR}) for Syt-1 C2A (top left), C2B (top right), C2AB-His with uncleaved hexahistidine tag (middle left), C2AB (middle right) and Cyt (bottom left) are shown. For comparison, theoretical spectra [203] of Syt-1 residues 141 - 418 (grey) and residues 271 - 418 (black) based on a published structure of Syt-1 (PDB: 5KJ7) are shown.

corresponding peptide intensities for every residue revealed a decrease in intensity for residues of the hexahistidine tag after treatment with the protease. The identification of the hexahistidine tag after purification can, therefore, be explained by a higher sensitivity of MS analysis compared to the western blot. Uncleaved Syt-1 is likely present after purification albeit at very low concentrations that do not affect subsequent experiments. Similar to the described procedure, other variants were purified (**Figures S3, S4, S5, S6, S7**). Note that no thrombin cleavage and, moreover, a bimodal elution peak after ion-exchange chromatography was observed for Syt-1 C2AB-His.

Characterisation of Purified Variants Having confirmed expression and purification of the Syt-1 variants, correct folding of the variants was validated by far-UV CD spectroscopy (**Figure 19**). This technique uses optical activity of chiral centres in the protein backbone to provide information on protein secondary structure. During CD spectroscopy, changes in the intensities of left and right polarised components of a circularly polarised light wave interacting with the protein are reported as ellipticity [99]. Raw ellipticity is converted to mean residue ellipticity taking into account the protein concentration and the length of a protein and therefore allowing the comparison between proteins of different molecular weight. Spectra for all Syt-1 variants showed a local minimum at 218 nm, a characteristic of anti-parallel β-sheet secondary structure (**Figure 19**). For a more detailed interpretation, theoretical spectra [203] based on a published structure of Syt-1 (residues 141 - 418, PDB: 5KJ7 [208]) were calculated. The experimental spectra agree well with the theoretical spectra

in terms of the minimum at 218 nm and the magnitude of mean residue ellipticity at wave lengths >205 nm. Differences between theoretical and experimental spectra below 205 nm can be related to the N-terminal intrinsically disordered linker that is present in all soluble variants except C2B. As disordered peptides generally show a negative ellipticity at around 195 nm and low absolute ellipticity around 210 nm, the observed differences can be the result of the additional contribution of the disordered linker that lowers the observed ellipticity low wavelengths <205 nm. For C2B that does not include the disordered linker, a different theoretical spectrum based on residues 271 - 418 of the model was calculated and agreed with the experimental results at wavelengths between 195 nm - 250 nm. In summary, the purified Syt-1 variants were correctly expressed, folded and suitable for further analysis.

Prior to the analysis of Syt-1-lipid interactions, purity and intact mass of the purified proteins were verified using native MS, i.e. a technique that preserves non-covalent interactions during ionisation and enables the detection of protein complexes. The native mass spectrum of Syt-1 C2A (**Figure 20A**) shows a single peak distribution corresponding a mass of 19,520 Da and charge states ranging from 7+ to 10+. This mass agrees well with the theoretical mass of 19,522.19 Da for Syt-1 C2A, calculated based on its amino acid sequence. Similarly, a single distribution corresponding to monomeric Syt-1 C2B (**Figure 20B**) with an experimentally determined mass of 17,620 Da (17,609.52 Da from sequence) was observed. In contrast, mass spectra of Syt-1 C2AB and Cyt (**Figure 20C and D**) each show two charge state distributions corresponding to monomeric and dimeric Syt-1. The corresponding masses of 75,020 Da and 78,100 Da agree well with the theoretical masses calculated from protein sequence of 74,977.94 Da for Syt-1 C2AB and 78,037.72 Da for Syt-1 Cyt. Similar to the masses of the observed dimers, monomer masses agree well with those expected from sequence after successful thrombin cleavage (37,488.97 Da and 39,018.86 Da vs. 37,510 Da and 39,050 Da experimentally, respectively). Native MS analysis of Syt-1 C2AB-His revealed two sub-distributions of the monomer peak distribution corresponding to masses of 34,169 Da and 34,342 Da compared with a theoretical mass of 34,122.19 Da for Syt-1 C2AB-His with the hexahistidine tag attached. Note that a shift after thrombin cleavage was not observed for this variant (see **Figure S5**) indicating that thrombin cleavage was not successful. Moreover, the additional mass shift of 173 Da indicated a post-translational modification of the protein, presumably phosphogluconylation of the hexahistidine tag that corresponds to a mass shift of 178 Da [227]. Interestingly, the mass of the observed dimer of 68,511 Da is between the masses of the modified and the unmodified dimer, indicating that the post-translational modification does not impair dimerisation. In summary, monomer masses determined using native MS agree well with those expected from sequence and a low degree of dimerisation observed for Syt-1 variants containing both C2 domains.

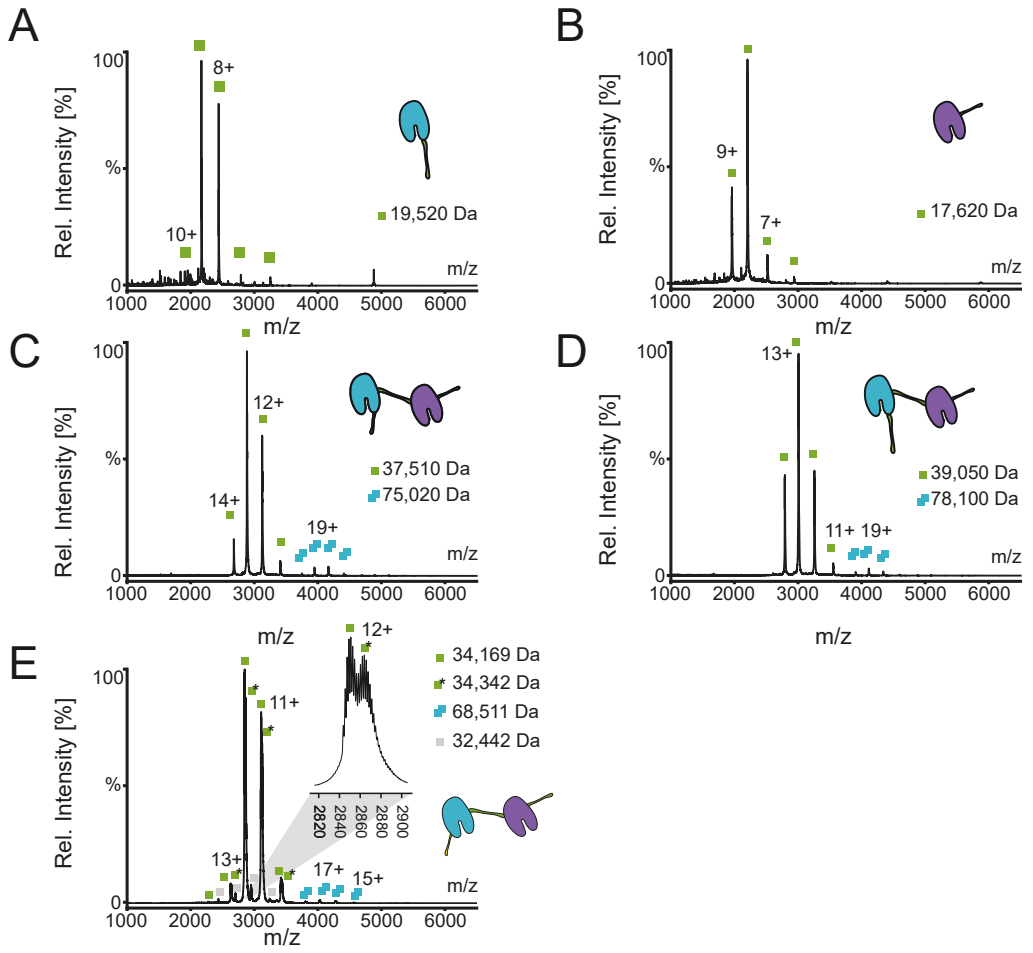


Figure 20: Native MS of Syt-1 Soluble Variants. Native mass spectra of the soluble variants are shown. **(A)** Syt-1 C2A, **(B)** Syt-1 C2B, **(C)** Syt-1 C2AB, **(D)** Syt-1 Cyt and **(E)** Syt-1 C2AB-His. The two peak distributions with a mass difference of approximately 180 Da (phosphogluconoylation of the hexahistidine tag) are shown as inset. Peaks assigned to a monomer (green squares) or dimer (cyan double squares) peak distribution are highlighted with the charge indicated. Masses derived from the peak distributions are shown.

Next, the oligomeric state of full-length Syt-1 was analysed by native MS. As detergent was required to solubilise the membrane helix, Syt-1 was purified in CHAPS detergent. However, CHAPS is incompatible with native MS analysis. Therefore, Syt-1 was transferred into a buffer containing C8E4, a detergent which is compatible with native MS analyis and can be removed by collisional activation [175]. The obtained mass spectrum (**Figure 21A**) shows three peak distributions assigned to monomeric and dimeric Syt-1. Interestingly, splitting of the monomer peak into four populations corresponding to masses of 47,970 Da, 48,272 Da, 48,672 Da and 49,075 Da was observed, all larger than the mass expected from Syt-1 sequence (47,680.22 Da). These masses likely correspond to Syt-1 retaining interactions with other molecules such as detergents or lipids. The first peak

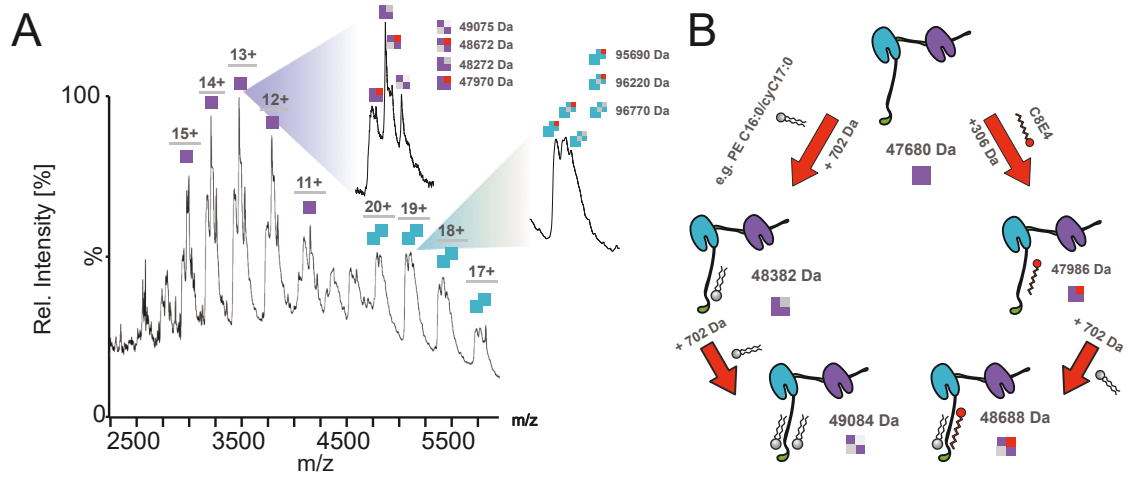


Figure 21: Native MS of Full-Length Syt-1. **(A)** Mass spectrum of full-length Syt-1 transferred into 2 × cmc C8E4. Peaks and charge states corresponding to masses of monomeric (purple squares) and dimeric (turquoise double squares) Syt-1 are assigned. Monomer and dimer peaks split into four populations (insets) with masses (see legend) >300 Da larger than expected from the sequence. **(B)** Tentative assignment of observed masses to non-covalent binding of up to two molecules of C8E4 and an unidentified molecule of approximately 700 Da, likely a membrane lipid.

observed in the spectrum corresponds in mass to Syt-1 retaining a single C8E4 molecule (mass shift of 306 Da, see **Figure 21B**). Additional shifts of approximately 700 Da correspond in mass to the protein binding to *E. coli* lipids. The same mass shifts were observed for the putative Syt-1 dimer, supporting the idea that the protein retains detergents and lipids after purification. Note that these mass shifts cannot originate from Syt-1 retaining the uncleaved hexahistine tag (theoretical mass of 49,431.08 Da).

In summary, full-length Syt-1 and soluble variants of the protein were purified and their correct expression and folding verified. Soluble variants predominantly formed monomers even though a low degree of dimerisation was observed by native MS. The mass of full-length Syt-1 was verified by native MS revealing that the protein likely retains detergent and lipid molecules from the purification.

4.1.3 Lipid Selectivity of Synaptotagmin-1

Interactions of Syt-1 with lipids in the SV or the plasma membranes are important for its biological function. For example, the C2B domain of Syt-1 interacts specifically with the plasma membrane [54], mediated by interaction with membrane-specific PI(4,5)P_2 lipids. Syt-1-lipid interactions were suggested to occur via two independent mechanisms: The first mechanism involves coordination of multiple Ca^{2+} ions by the C2 domains [50, 228, 48] and interactions of the Ca^{2+} with negatively

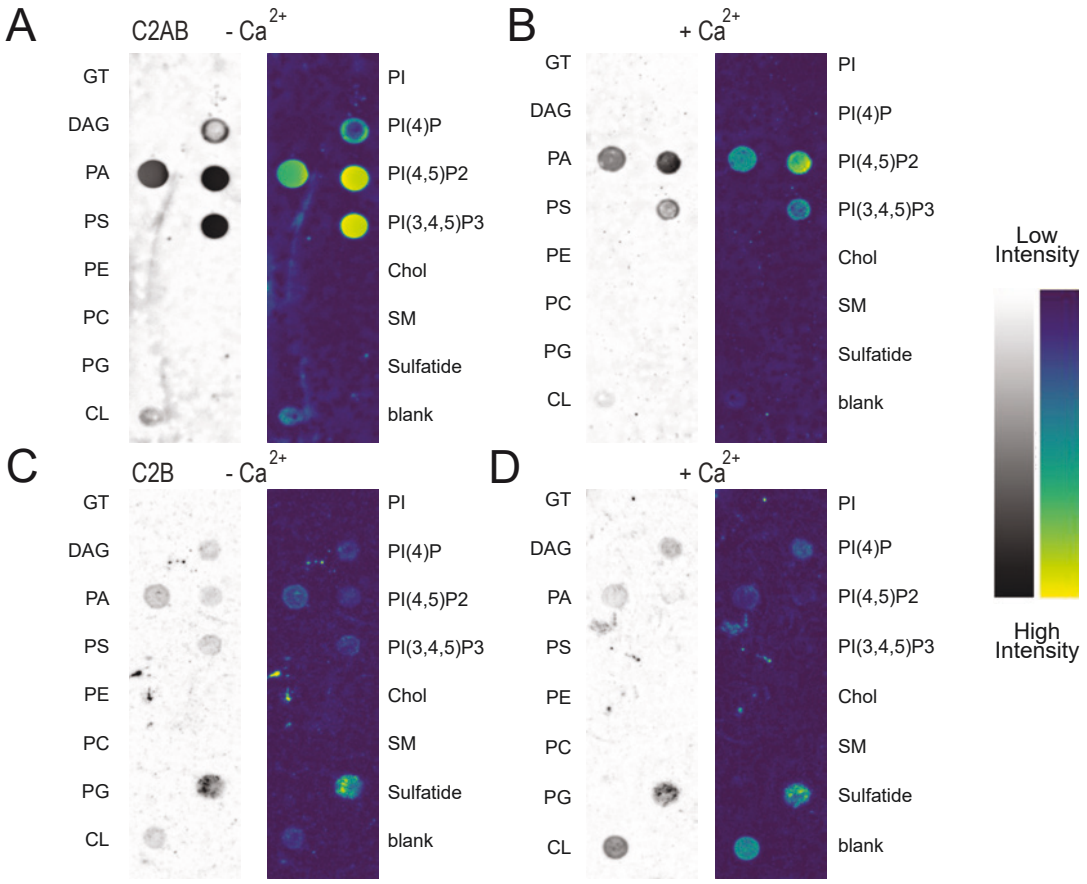


Figure 22: Protein Lipid Overlay Assay to Determine Lipid Binding of Syt-1 C2AB and C2B. Syt-1 C2AB or C2B with a hexahistidine tag were added to lipid strips containing immobilised lipids of different classes (see labels). Binding was detected with antibodies recognising Syt-1 (C2AB) or the hexahistidine tag (C2B). Pixel intensities were normalised to the highest pixel value on a lipid spot and visualised using black-white (left) and multi-colour (right) lookup tables. (**A**) Syt-1 C2AB-His binding to immobilised lipids in the absence of Ca^{2+}. (**B**) Syt-1 C2AB-His binding to immobilised lipids in the presence of 500 µM $CaCl_2$. (**C**) Syt-1 C2B binding to immobilised lipids in the absence of Ca^{2+}. (**D**) Syt-1 C2B binding to immobilised lipids in the presence of 500 µM $CaCl_2$.

charged phospholipids in the membrane thereby completing the Ca^{2+} coordination sphere [228]. The second mechanism involves the polybasic patch of C2B that specifically interacts with $PI(4,5)P_2$ in a Ca^{2+}-independent manner [71]. Next, preferences of Syt-1 for different lipid classes were investigated using immobilised lipids, native MS and binding of Syt-1 to intact lipid membranes.

Protein Lipid Overlay Assay A commonly applied technique for determining the lipid head group-specificity of proteins is the protein lipid overlay assay [229] (see **Section 3.2** for details). Briefly, this approach is based on immobilising lipids of different classes on a polyvinylidene fluoride

membrane. A protein is added and binds to the various lipids with different affinities. Similar to a western blot, the presence of the protein is detected using antibodies. In this thesis, interaction of Syt-1 with triglycerides, diacylglycerol, phosphatidic acid, PS, phosphatidylethanolamine, phosphatidylcholine, PG, cardiolipin, PI, PI(4)P, PI(4,5)P_2, PI(3,4,5)P_3, cholesterol, sphingomyelin and 3-sulfogalactosylceramide lipids was analysed using the protein lipid overlay assay.

To discriminate Ca^{2+}-dependent and Ca^{2+}-independent lipid interactions of Syt-1, the protein lipid overlay assay was performed twice, in the absence and presence of Ca^{2+} (**Figure 22**). The experiments were performed together with Til Kundlacz. In the absence of Ca^{2+}, Syt-1 C2AB showed a preference for anionic lipids by binding to the three PIPs and to phosphatidic acid (**Figure 22A**). All these lipids contain at least two negative charges at neutral pH, indicating that charge plays an important role for Syt-1-lipid interactions. A low degree of affinity was also observed for cardiolipin, most likely a non-natural interaction as cardiolipin in eukaryotes is mostly located in the inner mitochondrial membrane. In contrast, interaction between Syt-1 and the singly charged lipid PS was not identified under these conditions. In contrast, binding to multiply charged PI(4,5)P_2 and PI(3,4,5)P_3 as well as to phosphatidic acid was observed (**Figure 22B**). Interactions with PI(4)P were not detected, indicating that the presence of Ca^{2+} reduced the overall binding efficiency of Syt-1 C2AB to immobilised lipids. The protein lipid overlay assay was also performed with Syt-1 C2B, a variant including the C2B domain and the C-terminal extension. In the absence of Ca^{2+}, interactions with PI(4)P, PI(4,5)P_2, PI(3,4,5)P_3 and phosphatidic acid as well as binding to cardiolipin and sulfatides were observed (**Figure 22C**). Similar to C2AB, addition of Ca^{2+} reduced the observed affinities and did not induce a shift of the interacting pattern to different lipids (**Figure 22D**). Interestingly, the weak interaction of C2B with PI(4)P and phosphatidic acid was observed in the presence and absence of Ca^{2+}, whereas binding to PI(4,5)P_2 and PI(3,4,5)P_3 was only detected in the absence of Ca^{2+}. The protein lipid overlay assay, thereby, showed that recombinantly expressed Syt-1 variants have an affinity for PIPs. However, no Ca^{2+}-induced shift in affinity towards singly negatively charged lipids such as PS was observed using this approach.

Interaction of Synaptotagmin-1 with Soluble PIP_2 Head Groups Next, interactions of head groups of the anionic PIPs with Syt-1 were analysed using native MS. In contrast to intact phospholipids, lipid head groups are soluble in aqueous solution, have a lower mass than the intact lipids, are easier to handle and, therefore, well suited for initial studies. To investigate whether Syt-1 discriminates between different PIPs, binding of the two positional isomers I(1,3,5)P_3 and I(1,4,5)P_3 to Syt-1 C2AB-His was analysed. **Figure 23A** shows mass spectra of 10 µM soluble Syt-1 C2AB-His in the absence and presence of 50 µM of I(1,4,5)P_3. The mass spectrum in the absence of the head group shows two peak distributions corresponding in mass to Syt-1 fused to the N-terminal hexahistidine tag with and without phosphogluconoylation of the tag (see **Section 4.1.2** for details). In the presence of I(1,4,5)P_3, binding of up to four I(1,4,5)P_3 molecules to

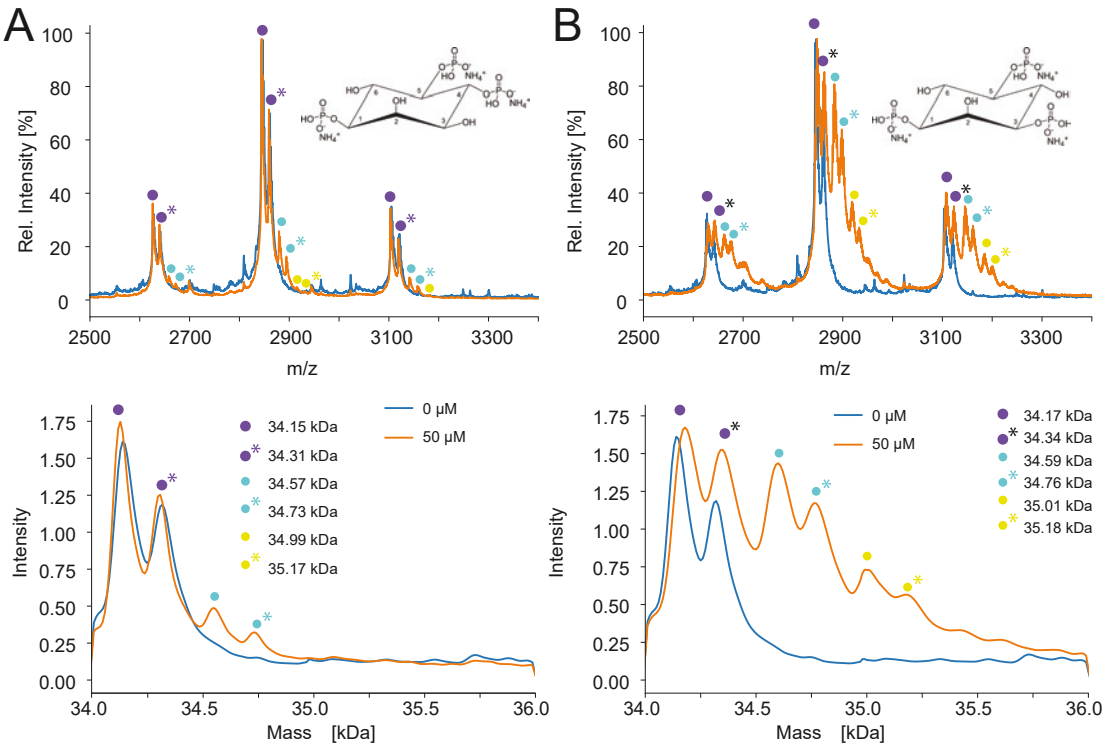


Figure 23: Binding of Inositoltriphosphates to Syt-1 C2AB-His. **(A)** Native MS of Syt-1 C2AB-His (top) in the absence (blue) and presence (orange) of I(1,4,5)P_3 (structure shown in inset). Peaks corresponding to unmodified Syt-1 C2AB-His (circles) and to Syt-1 C2AB-His with a mass modification from post-translational phosphogluconoylation (circles with asterisks) were observed. Mass spectra and deconvolved mass spectra (bottom) show distinct peaks for the binding of one (turquoise) or two (yellow) inositoltriphosphate molecules. **(B)** Binding of Syt-1 to I(1,3,5)P_3. See (A) for details.

Syt-1 was observed, however, due to the low signal intensity of the peaks corresponding to the larger complexes, analysis was focused on the binding of up to two I(1,4,5)P_3 molecules. For quantification, deconvolved charge-free spectra were calculated using the UniDec software [172]. In contrast to direct peak annotation, deconvolution sums intensities from multiple corresponding peaks and generates a continuous intensity distribution of masses encoded in the spectrum. This simplifies quantitative comparison of peaks corresponding to the same mass across different spectra. **Figure 23A** (bottom panel) shows the deconvolved mass spectra of Syt-1 C2AB-His in the absence and presence of I(1,4,5)P_3. In total four peaks corresponding to the unmodified and the modified proteoform of Syt-1 C2AB-His with and without a single bound I(1,4,5)P_3 molecule were calculated. Peaks corresponding to two I(1,4,5)P_3 molecules binding to the protein were only observed at higher head group concentration.

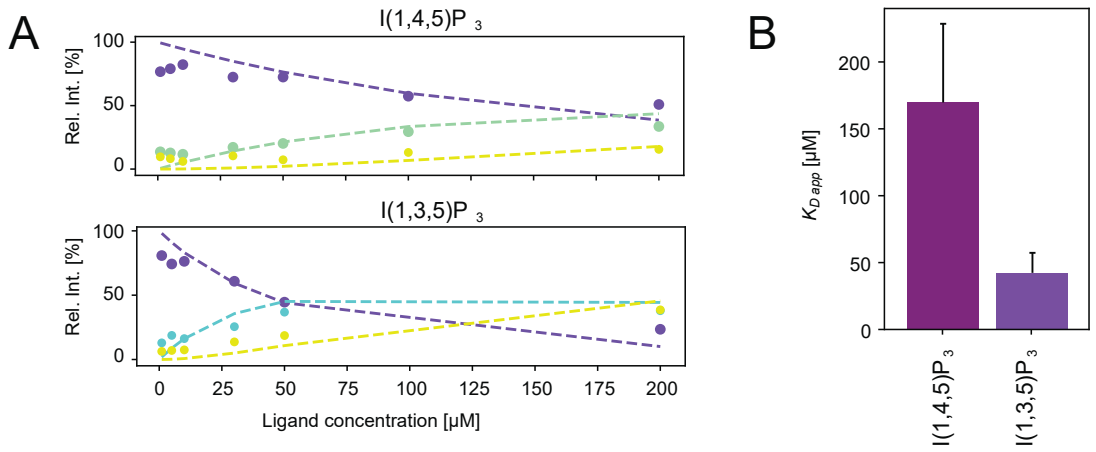


Figure 24: Quantification of Interactions of Inositoltriphosphates with Syt-1 C2AB-His. (A) Extracted peak areas for free protein (purple) and protein bound to one (turquoise) or two (yellow) molecules of I(1,4,5)P_3 (top) or I(1,3,5)P_3 (bottom), respectively. Global fits from UniFit for each protein ligand combination are shown as dashed line. **(B)** Bar plot of $K_{D\,app}$ values for both ligands. Fit confidence from bootstrapping is shown as error bar.

Analysis of Syt-1 in the presence of 50 µM I(1,3,5)P_3 also revealed peaks corresponding to Syt-1 ligand complexes. Interestingly, binding to I(1,3,5)P_3 was observed at lower ligand concentrations than binding to I(1,4,5)P_3 (**Figure 23B**). For example, comparing the spectrum of 50 µM I(1,4,5)P_3 (**Figure 23A**, orange) with that of the same concentration of I(1,3,5)P_3 (**Figure 23B**, orange), a clear difference in the intensity of a 1:1 complex of Syt-1 and the respective inositolphosphate (turquoise circles) is observed. For I(1,4,5)P_3 the intensity is approximately half the intensity of the ligand-free form, for I(1,3,5)P_3, however, the complex and the free protein show roughly the same intensity. For quantitative comparison of the affinity of Syt-1 to the two inositolphosphates, peak areas for every mass corresponding to a Syt-1-ligand complex were extracted and plotted against the ligand concentration (**Figure 24A**). $K_{D\,app}$ values were determined for binding of the first, second and third ligand molecule using the UniFit routine implemented within UniDec [172]. UniFit calculates the $K_{D\,app}$ values involved in ligand binding by global fitting, i.e. by determining values that minimise the difference between the experimentally observed intensities and those derived from the fitting model for all experimental spectra simultaneously. Eventually, fit confidences are calculated in UniFit by bootstrapping, i.e. by randomly selecting a subset of the input data and calculating the corresponding $K_{D\,app}$ values. This is repeated multiple times to calculate the final confidence. The $K_{D\,app}$ values obtained using this approach must be discriminated from in-solution dissociation constants as for example gas phase dissociation of the complexes after ionisation [230] results in systematically higher $K_{D\,app}$ values from native MS experiments. Nevertheless, comparison between multiple $K_{D\,app}$ values obtained from native MS is reliable as long as activation in the gas phase is similar for the corresponding measurements. By native MS analysis of the

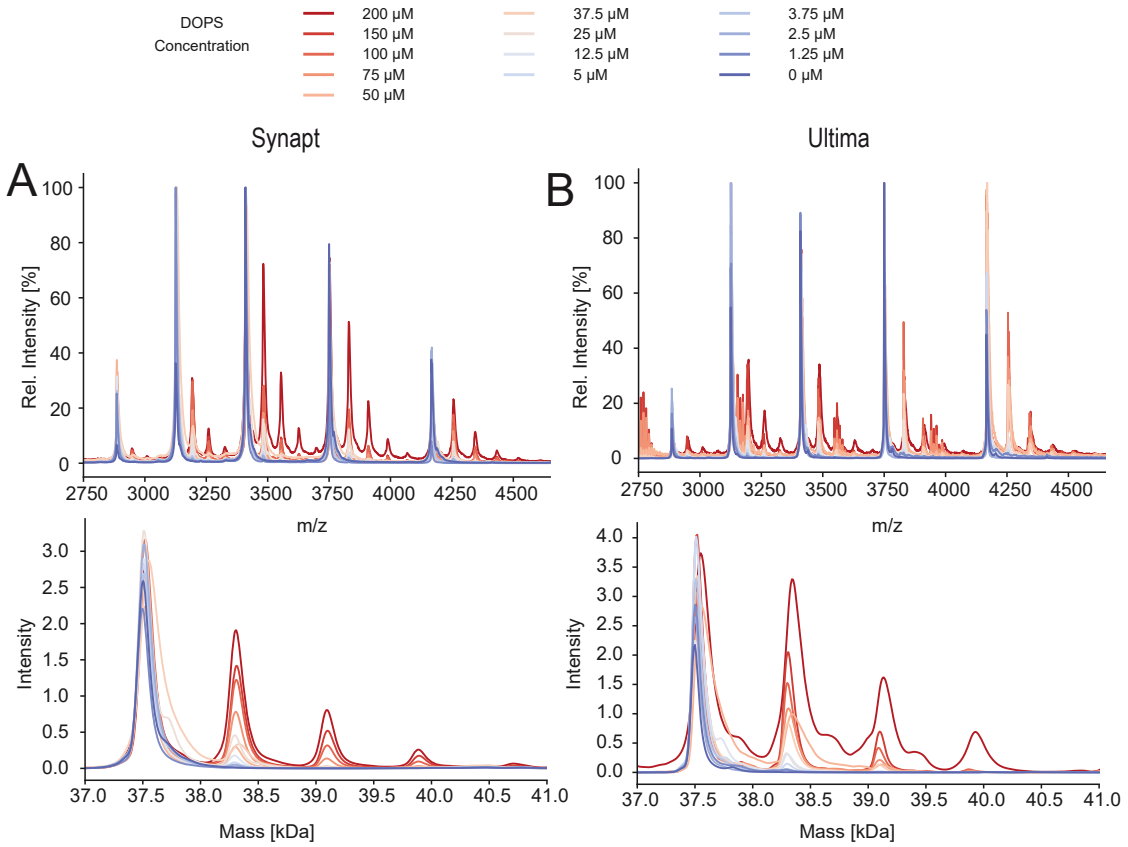


Figure 25: Binding of Syt-1 C2AB to DOPS. (A) 5 µM Syt-1 C2AB in the presence of 2 × cmc C8E4 and increasing concentrations of DOPS (blue to red) analysed by native MS on a Synapt G1 instrument modified for high-mass transmission. Up to four lipid molecules in complex with the protein were observed. Mass spectra (top) and deconvolved spectra (bottom) are shown. **(B)** The same protein lipid mixtures analysed on a Q-TOF Ultima modified for high-mass transmission.

inositolphasphate lipid head groups interacting with Syt-1, $K_{D\ app}$ values of approximately 170 µM and 40 µM were obtained for I(1,4,5)P_3 and I(1,3,5)P_3, respectively (**Figure 24B and Table 17**). Altogether, this experiment indicates that native MS analysis of soluble protein-ligand interactions allows the determination of different ligand affinities.

Native Mass Spectrometry of Synaptotagmin-1 Binding to Intact Lipids Analysis of full-length Syt-1 revealed additional peaks in the mass spectrum, indicating binding of lipids or detergents after solubilisation and purification (**Figure 21**). However, lipid interactions with the transmembrane helix cannot be discriminated from interactions with the soluble C2 domains by native MS analysis of the full-length protein. Having shown that presence of C8E4 enables lipid transfer to soluble standard proteins (**Section 4.1.1**), detergent-mediated lipid-transfer of different

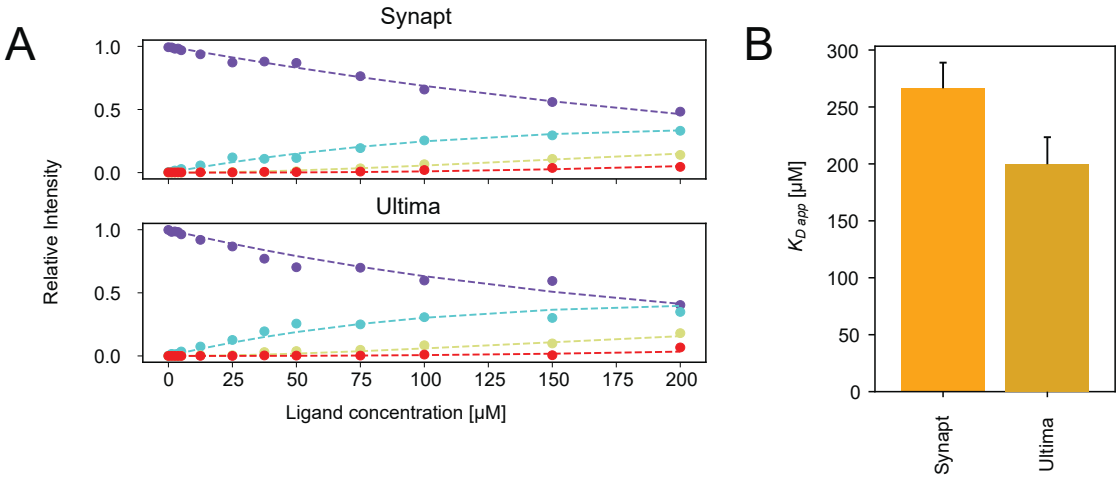


Figure 26: Quantification of Syt-1 C2AB Interacting with DOPS. (A) Extracted peak areas of lipid-free (purple) peaks and peaks corresponding to Syt-1 binding to one (turquoise), two (yellow) or three (red) lipid molecules. The result of the global fit is shown as dashed line. **(B)** Bar plot representation of the affinity constants of Syt-1 binding to one DOPS molecule derived from the analysis. Fit confidence derived from bootstrapping is shown as error bars.

intact lipids was applied to truncated soluble variants of Syt-1, namely C2AB, C2A and C2B. According to literature, Syt-1 preferably interacts with negatively charged phospholipids such as PS or PI(4,5)P_2. Therefore, DOPS, a highly pure, readily available and synthetically produced lipid was selected as potential binding partner of Syt-1. Binding of DOPS to Syt-1 was analysed on two different native MS instruments, namely a Synapt G1 HDMS hybrid quadrupole TOF instrument and Q-TOF Ultima hybrid quadrupole TOF instrument. Ionisation on the Synapt G1 is considered to involve a certain degree of collisional activation and unfolding. This property has been exploited for the analysis of membrane proteins that require more collisional activation for removal of the detergent micelle [231]. As Syt-1-lipid binding experiments included detergents but also soluble proteins that are more susceptible to unfolding, the Q-TOF Ultima, an instrument considered producing less activating ionisation conditions, was chosen for comparison. Note that, in general, higher collision voltages were necessary on both instruments to obtain resolved peaks compared to analysis of the soluble variants in the absence of detergent, possibly related to the stabilising properties of detergent micelles (see **Section 5.1.1**).

5 µM Syt-1 C2AB were mixed with 2 × cmc C8E4 and increasing concentrations of DOPS vesicles produced by sonification of the lipid in ammonium acetate solution and analysed on the Synapt G1 instrument (**Figure 25A**). As observed previously for soluble PI(4,5)P_2 head groups and for POPE binding to β-LG, presence of protein-lipid complexes was indicated by additional peaks shifted to higher *m/z* ratios. For C2AB in the presence of DOPS, up to four bound DOPS molecules were observed at a lipid concentration of 200 µM. Moreover, the number of bound lipids correlated with

the charge state of the protein. Fewer lipids were bound to highly charged Syt-1 ions than to Syt-1 ions with low charge states. As lower charge states correlate with a more native structure of the protein, this observation suggests that lipid binding is not solely dependent on the overall charge but rather requires native-like protein-lipid interactions. Using the Q-TOF Ultima, the same number of DOPS molecules binding to Syt-1 were observed (**Figure 25B**). In contrast to the spectra obtained on the Synapt G1, a similar number of bound lipids was determined for all charge states, possibly due to the gentler ionisation conditions of the instrument. After deconvolution and peak area extraction, $K_{D\,app}$ values were determined using UniDec (**Figure 26A**). Note that $K_{D\,app}$ value analysis was performed with mass peaks corresponding to binding of one and two lipids as peak intensities for binding of a third lipid were too low for reliable fitting for the spectra acquired on both instrument (see **Figures 25A and B**). For Syt-1 C2AB analysed using the Synapt G1 mass spectrometer, $K_{D\,app}$ values of approximately 260 µM was obtained for the first lipid and of approximately 420 µM for a second DOPS binding to a Syt-1 DOPS complex (**Figure 26B**). $K_{D\,app}$ values for the same interaction analysed on Q-TOF Ultima mass spectrometer were approximately 200 µM and 480 µM for binding to the first and second DOPS molecule. Note that slightly higher affinities were determined using the Q-TOF Ultima mass spectrometer, in line with this instrument showing less gas phase dissociation so that more lipids are retained at lower lipid concentrations. Nevertheless, the $K_{D\,app}$ values obtained from analyses on both instruments were overall in agreement suggesting that instrument parameters only play a minor role in determining lipid affinities. However, the experiment also suggests that $K_{D\,app}$ values are best compared between mass spectra obtained under similar conditions. Subsequent experiments were, therefore, performed with the Synapt G1 instrument.

After successfully determining DOPS affinities of Syt-1 C2AB, the analysis was extended to members of other important lipid classes, namely DOPC, DOPG, DOPI, PI(4)P, PI(4,5)P_2 and DOPI(3,4,5)P_3. It should be noted that PI(4)P and PI(4,5)P_2 lipids were purchased purified from brain tissue and therefore contained a natural distribution of fatty acyl chains, whereas synthetic DOPI(3,4,5)P_3 containing two oleyl chains was used as PI(3,4,5)P_3 lipid. Syt-1 was analysed in the presence of C8E4 detergent and increasing concentrations of the lipids, an experiment performed together with Til Kundlacz (see **Figure 27**). Note that the actual lipid concentrations after solubilisation were verified in this experiment using the phosphate assay (see **Section 3.2**). The mass spectra were deconvolved, peak areas for lipid-free and the protein-lipid complexes were extracted and $K_{D\,app}$ values were determined (**Figures 27A and B**). Affinity of C2AB to DOPC was low, reflected by a high $K_{D\,app}$ value for the interaction with the first lipid molecule (see **Table 17**). DOPG, DOPI and DOPS showed similar $K_{D\,app}$ values of around 350 µM, whereas analysis in the presence PIs yielded higher affinities. A single PI(4)P lipid was bound with an affinity of approximately 170 µM, whereas binding to PI(4,5)P_2 and DOPI(3,4,5)P_3 corresponded to lower $K_{D\,app}$ values of approximately 50 µM. Data quality of the PI(4,5)P_2 interaction experiments was sufficient to determine the $K_{D\,app}$ value of the second lipid binding that was double the $K_{D\,app}$ value of C2AB binding to the first

Table 17: $K_{D\,app}$ values of Syt-1-Lipid Interactions. $K_{D\,app}$ values for binding of one ligand ($K_{Dapp,1}$) and for binding of a second ligand to a ligand-bound Syt-1 ($K_{Dapp,2}$) are shown. Values were derived from global fitting of native mass spectra obtained at multiple Syt-1 ligand ratios. Errors were calculated from 100 random bootstrapping samples. Only values with a relative error below 100 % are shown.

Syt-1 variant	Detergent	Ligand	$K_{Dapp,1}$	$K_{Dapp,2}$
C2AB-His	No Detergent	I(1,3,5)P_3	42.3 ± 14.4 µM	
		I(1,4,5)P_3	170.0 ± 58.8 µM	
C2AB	C8E4	DOPS	364 ± 64 µM	427 ± 166 µM
		DOPC	6560 ± 677 µM	
		DOPG	372 ± 65 µM	
		DOPI	324 ± 30 µM	
		PI(4)P	178 ± 21 µM	
		PI(4,5)P_2	54 ± 9 µM	102 ± 75 µM
		DOPI(3,4,5)P_3	58 ± 12 µM	
C2A	C8E4	DOPC	338 ± 31 µM	372 ± 103 µM
		DOPG	180 ± 24 µM	
		DOPI	87 ± 18 µM	
		DOPS	103 ± 10 µM	
		PI(4)P	191 ± 12 µM	360 ± 82 µM
		PI(4,5)P_2	204 ± 8 µM	269 ± 39 µM
C2B	C8E4	DOPC	258 ± 18 µM	448 ± 150 µM
		DOPG	226 ± 21 µM	248 ± 115 µM
		DOPI	155 ± 6 µM	247 ± 70 µM
		DOPS	82 ± 10 µM	
		PI(4)P	284 ± 32 µM	
		PI(4,5)P_2	175 ± 13 µM	250 ± 207 µM

PI(4,5)P_2 lipid. This indicates affinity of Syt-1 to PI(4,5)P_2 is lowered after binding to the first lipid, i.e. the Syt-1-PI(4,5)P_2 interaction shows negative cooperativity. Overall, the determined lipid preference was DOPC$\ll$DOPG<DOPS<DOPI<PI(4)P<DOPI(3,4,5)P_3<PI(4,5)P_2. This is in line with the expectation of Syt-1 preferentially binding to negatively charged lipids (see **Section 5.1**).

Syt-1 was suggested to exert its function by bridging the SV and the plasma membrane in the neuron after Ca^{2+} influx [77]. This would involve either both domains interacting with both membranes or C2A and C2B having distinct membrane preferences. To interrogate lipid-binding of Syt-1 in detail, truncated variants of Syt-1 including the individual C2A or C2B domains were investigated using

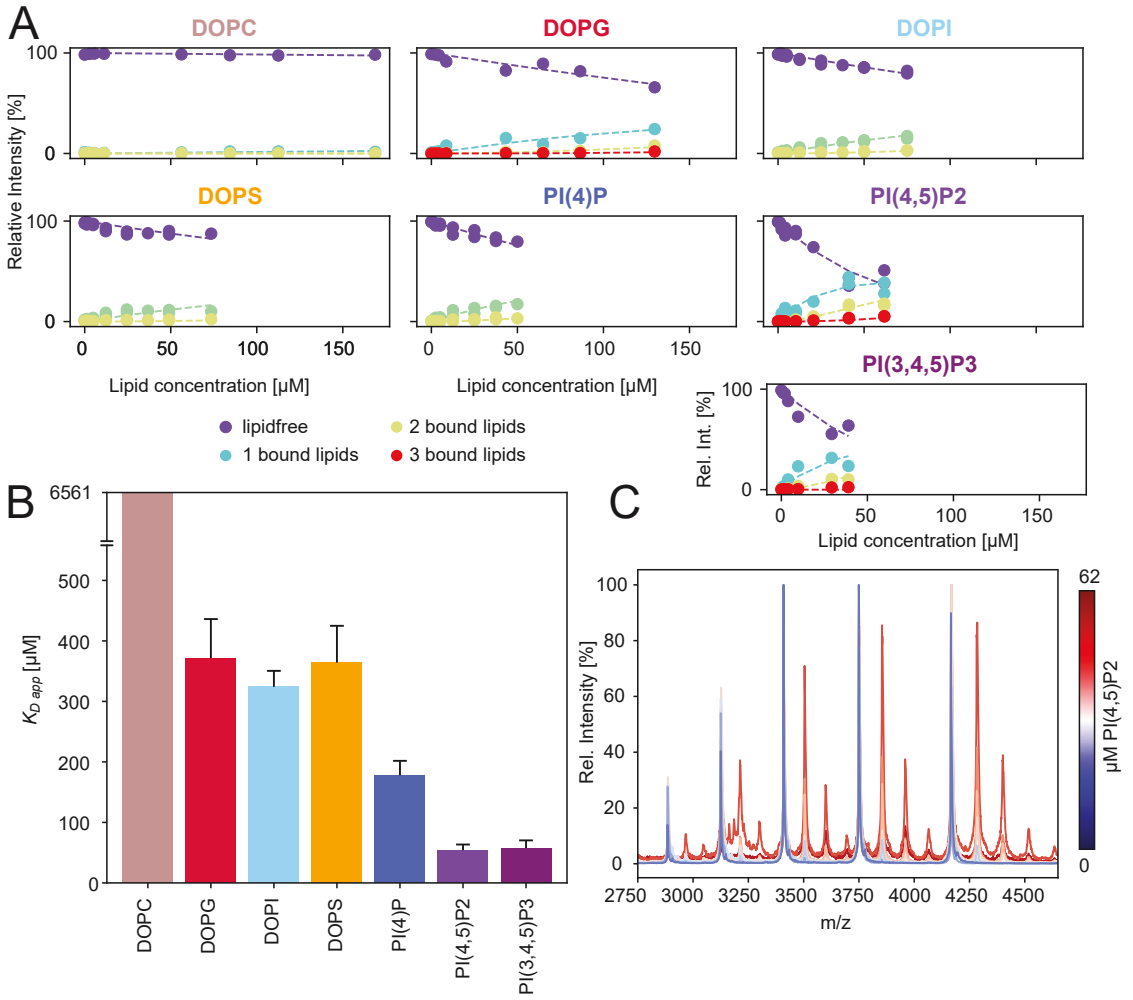


Figure 27: Quantification of Syt-1 C2AB-Lipid Complexes Involving Various Lipids. (A) Extracted peak areas from native MS analysis of 5 µM Syt-1 C2AB in the presence of 2 × cmc C8E4 and increasing concentrations of the lipids DOPC, DOPG, DOPI, DOPS, PI(4)P, PI(4,5)P_2 and DOPI(3,4,5)P_3. Native mass spectra were first deconvolved and peak areas corresponding to the ligand-free (purple) protein and protein-lipid complexes involving one (turquoise), two (yellow) or three (red) lipid molecules were extracted separately. The results of global fits to these intensities are shown (dashed lines). **(B)** Bar plot of the $K_{D\ app}$ values of Syt-1 C2AB to the different lipids. Only the affinities for binding to the first lipid are shown (refer to **Table 17** for detailed values for affinities for all bound lipids). Fit confidence derived from bootstrapping is shown as error bars. **(C)** Example spectra of binding of Syt-1 to PI(4,5)P_2. Increasing lipid concentrations from 0 µM (blue) to 62 µM (red) are indicated.

native MS. For this experiment, the workflow was slightly modified compared to the one previously described. Instead of forming lipid vesicles by sonification and adding them to the protein in the presence of detergent, lipids were solubilised with detergent before they were added to the protein (see **Section 3.2**). Moreover, sampling and extraction cone voltage were lowered compared to the

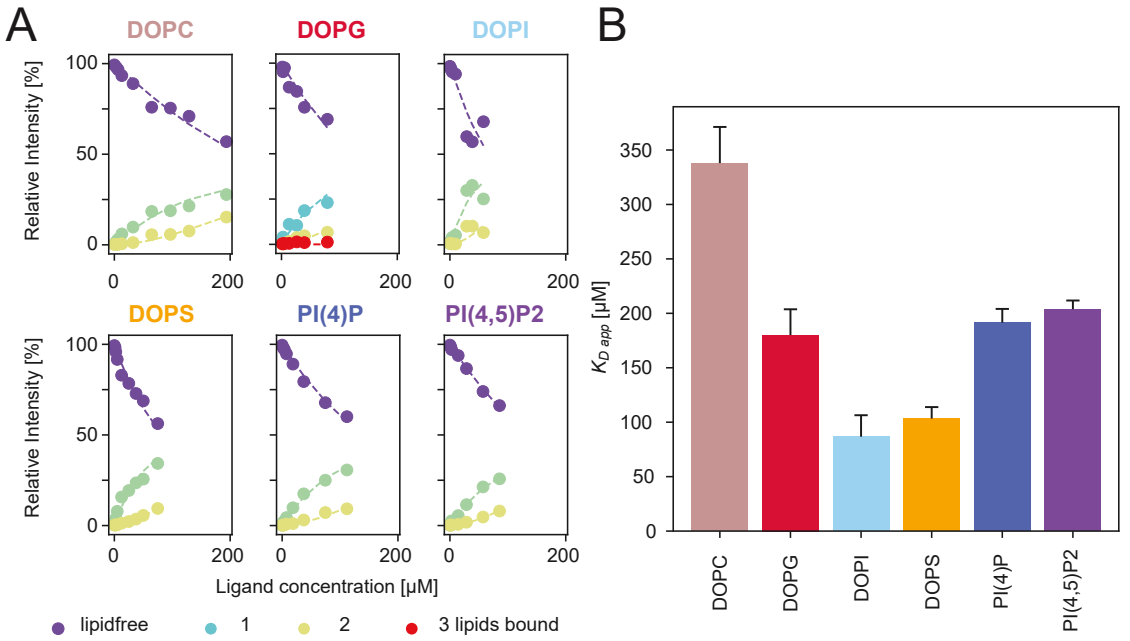


Figure 28: Quantification of Syt-1 C2A Lipid-Binding. **(A)** Extracted peak areas from native MS analysis of 5 µM Syt-1 C2A in the presence of 2 × cmc C8E4 and increasing concentrations of the lipids DOPC, DOPG, DOPI, DOPS, PI(4)P and PI(4,5)P_2. **(B)** Bar plot of the apparent affinities of Syt-1 C2A to the different lipids. Only the affinity for binding the first lipid is shown (refer to **Table 17** for detailed values for affinities for all bound lipids). Fit confidence derived from bootstrapping is shown as error bars.

previous experiment to reduce collisional activation of the protein during ionisation as much as possible (see **Section 3.2** for details). Proteins and lipids were mixed, analysed by native MS and, after extraction of peak areas, $K_{D\ app}$ values were calculated for lipid interactions of Syt-1 C2A and C2B, respectively (**Figures 28A and 29A**). This experiment was performed together with Til Kundlacz.

Similar to the results obtained with C2AB, the lowest observed affinity for C2A corresponded to the interaction with DOPC (**Figure 28B**). However, the absolute value of approximately 300 µM was lower than that calculated for C2AB. Note that, in contrast to the analysis of C2AB, binding of a second DOPC molecule was observed for C2A and C2B. C2A showed a reduced affinity (higher $K_{D\ app}$ value) for PI(4,5)P_2 compared to C2AB (approximately 200 µM vs. 50 µM). Interestingly, affinity of C2A for PI(4)P was of the same magnitude as for PI(4,5)P_2 in contrast to C2AB that showed a higher affinity for PI(4,5)P_2. Affinity of C2A for DOPG was approximately 200 µM, similar to Syt-1 C2AB. The highest lipid affinities of C2A ,corresponding to $K_{D\ app}$ values of approximately 100 µM, were observed for the singly charged lipids DOPI and DOPS, with a slightly higher affinity for DOPI. Overall, lipid preference of C2A was DOPC<PI(4,5)P_2<PI(4)P<DOPG<DOPS<DOPI.

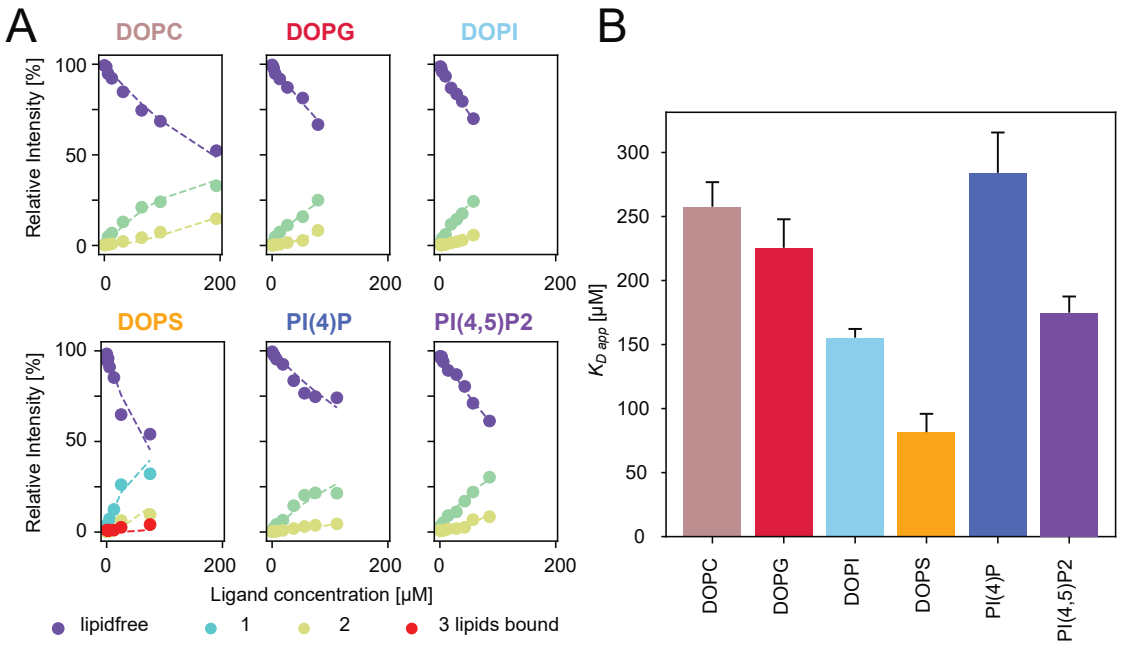


Figure 29: Quantification of Syt-1 C2B Lipid-Binding. **(A)** Extracted peak areas from native MS analysis of 5 µM Syt-1 C2B in the presence of 2 × cmc C8E4 and increasing concentrations of the lipids DOPC, DOPG, DOPI, DOPS, PI(4)P and PI(4,5)P_2. **(B)** Bar plot of the apparent affinities of Syt-1 C2B to the different lipids. Only the affinity for binding the first lipid is shown (refer to **Table 17** for detailed values for affinities for all bound lipids). Fit confidence derived from bootstrapping is shown as error bars.

Native MS analysis of C2B showed lipid affinities of similar magnitude, however with slightly different lipid preferences (**Figure 29B**). The highest lipid affinity of approximately 75 µM was calculated for C2B binding to DOPS, comparable to that calculated for C2A binding to DOPS. However, affinity of C2B to DOPI was approximately 150 µM, lower than C2A binding to DOPI ($K_{D\ app}$ value of approximately 100 µM). Affinity of C2B for PI(4,5)P_2 of approximately 175 µM was comparable to that observed for PI(4,5)P_2 and PI(4)P for C2A. In contrast, the $K_{D\ app}$ value obtained for binding of C2B to PI(4)P was approximately 300 µM and, therefore, similar to that observed for DOPG and DOPC. Overall, lipid preference of Syt-1 C2B was PI(4)P<DOPC<DOPG<PI(4,5)P_2<DOPI<DOPS.

In summary, the experiment showed that lipid preferences for the isolated domains differ from those observed for the C2AB variant. Whereas the tandem-C2 variant clearly preferred multiply charged PI(4,5)P_2 and DOPI(3,4,5)P_3 lipids, C2A favoured DOPI and C2B DOPS (see **Section 5.1** for discussion).

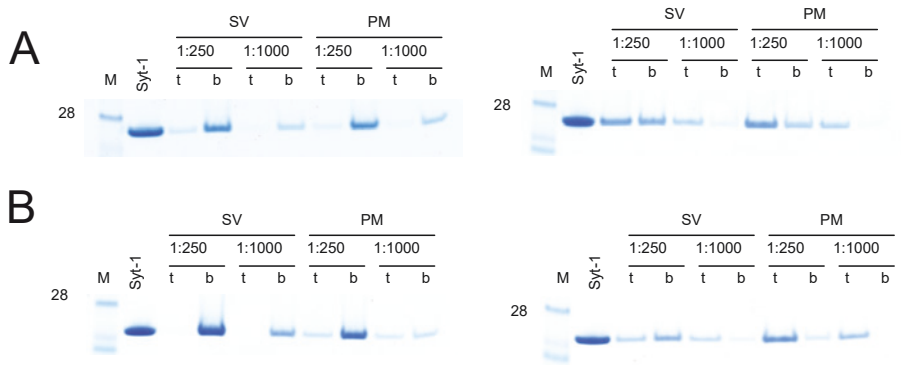


Figure 30: Liposome Flotation Assay of the Soluble Syt-1 Variants C2A and C2B. (A) Interaction of Syt-1 C2A in the absence of Ca^{2+} (left) and presence (right) of 100 µM $CaCl_2$ with liposomes mimicking the SV and plasma membranes at a protein-lipid ratio of 1:250 or 1:1000. Proteins binding to the liposomes float on top (t) of a sucrose density gradient, whereas free protein accumulates at the bottom (b) of the ultracentrifugation tube. **(B)** Interaction of Syt-1 C2B in the absence (left) and presence (right) of Ca^{2+}.

Flotation Assays of the Soluble C2 Domains During the protein lipid overlay assay, lipid acyl chains are not accessible for the protein. In contrast, detergent-mediated lipid transfer and native MS analysis enable binding of head groups and lipid acyl chains to the protein. However, interactions with an intact lipid bilayer may still differ from binding of isolated lipids to Syt-1, for example due to simultaneous interactions of a single protein with multiple lipids. To investigate Syt-1 interactions with intact double layered membranes in the absence of detergent, a flotation assay with soluble Syt-1 variants and liposomes of defined composition was performed together with Melissa Frick (**Figures 30 and S8**). For this, liposomes of defined size and composition were generated by extrusion through a polycarbonate membrane and incubated with the soluble variants Syt-1 C2A or C2B at protein-lipid ratios of 1:250 or 1:1000. Two lipid compositions were employed, mimicking the SV membrane and the plasma membrane composition, respectively (see **Section 3.2**). The SV model membrane was composed of POPC, DOPS, POPE and cholesterol, whereas the model for the plasma membrane contained $PI(4,5)P_2$ and diacylglycerol in addition to lipids of the SV model membrane. Syt-1 was added to the prepared liposomes, protein lipid mixtures were subjected to density-gradient centrifugation, fractions from the top and bottom of the ultracentrifugation tubes were collected and subjected to polyacrylamide gel electrophoresis. In this assay, proteins interacting with the liposomes co-migrate to the top of the tube, whereas unbound protein accumulates at the bottom. Therefore, the amount of protein in the top fractions correlates with the affinity of the protein for the respective lipid composition of the liposomes. In the absence of Ca^{2+}, no binding to liposomes mimicking SV or plasma membranes was observed for Syt-1 C2A (**Figure 30A**). Note that, to ensure absence of Ca^{2+}, 1 mM EGTA was added to the buffer for this experiment. In contrast, in the presence of Ca^{2+}, a clear interaction of C2A with liposomes mimicking either the SV or plasma membrane was found, indicating that the assay

captures a Ca^{2+}-induced shift in lipid affinity. Similarly, Syt-1 C2B showed no binding to liposomes mimicking the SV membrane in the absence of Ca^{2+}, however, a low degree of binding to liposomes mimicking the plasma membrane was found (**Figure 30B**). This is in line with the supposed Ca^{2+}-independent interaction of the polybasic region of C2B with the polyanionic $PI(4,5)P_2$ in the plasma membrane. In the presence of Ca^{2+}, C2B interacted with liposomes mimicking either the SV or the plasma membrane. Notably, C2B was almost exclusively present in the top fractions for liposomes mimicking the plasma membrane, indicating that it specifically interacts with plasma membrane lipids.

In summary, three biophysical techniques were employed to characterise the lipid preferences of the soluble C2 domains of Syt-1. Using the lipid overlay assay, interactions of the C2 domains with several biologically relevant lipid classes were tested and a preference of Syt-1 for negatively charged PIPs was found (**Figure 22**). This was investigated closer using native MS of Syt-1 C2AB-His in the presence of two soluble PIP head groups (**Figures 23 and 24**). Next, using detergent-mediated lipid transfer to soluble proteins, interactions of Syt-1 with various intact lipids were studied and $K_{D\ app}$ values were determined (**Table 17**). Experiments recapitulated the published affinity of the protein for multiply charged PIPs, in line with the lipid overlay assay (**Figure 27**). With native MS analysis of the isolated C2 domains, diverging lipid preferences for C2A and C2B were observed, namely DOPI and DOPS for C2A and C2B, respectively (**Figures 28 and 29**). To analyse Syt-1 binding to intact lipid bilayers, liposomes were used as membrane mimetics and revealed a preference of C2A for the SV membrane and of C2B for the plasma membrane in the presence of Ca^{2+} (**Figure 30**).

4.1.4 Molecular Dynamics Simulations of Interactions with Lipid Bilayers

Distinct membrane preferences for the C2A and C2B domain in the presence of Ca^{2+} were observed using the flotation assay. To investigate Syt-1-membrane interactions in more detail, MD simulations that provide dynamic models of molecular interactions were performed. For this, trajectories of the two Syt-1 C2 domains in the presence of SV and plasma membrane model membranes and Ca^{2+} ions were computed. For details on the simulations see **Section 3.2.6**. Briefly, structures of Syt-1 C2 domains in complex with Ca^{2+} were obtained from a crystal structure of the Syt-1 SNARE complex [208] and lipid bilayers of defined composition were generated using CHARMM-GUI [196]. For analysis of Ca^{2+} free Syt-1, Ca^{2+} ions were removed from the structures. Protein structures and membrane models were relaxed and equilibrated separately (see **Section 3.2** for details) in a simulation box containing water molecules and salt ions. The C2 domains were then placed on the membranes at 0deg, 30deg and 60deg angles with respect to the membrane normal to sample different initial orientations (subsequently referred to as replicates). The combined protein-membrane system was equilibrated and production runs corresponding to 900 ns total simulated time were performed.

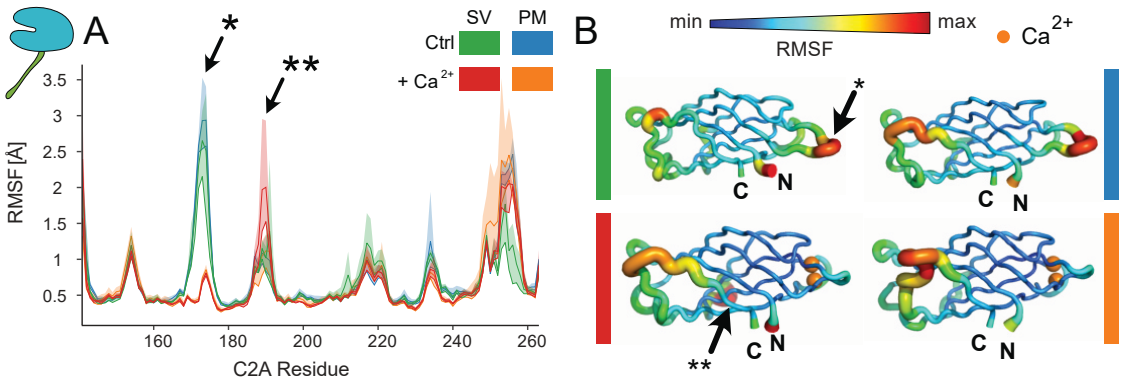


Figure 31: RMSF of Syt-1 C2A. Regions of high flexibility or with differences between the simulations are highlighted (arrows). Residues 169 - 176 (calcium binding loop) are labelled with an asterisk (∗) and residues 187 - 192 (polybasic stretch) with two asterisks (∗∗). **(A)** C_α RMSF of Syt-1 141-263. See inset for colour scheme. The mean RMSF (solid line) and the range of RMSFs per residue (transparent area) of three replicates is shown. **(B)** Sausage representation of Syt-1 141-263 RMSF. See (A) for colour scheme. Higher RMSFs are indicated by larger diameter and warmer colours of the protein backbone representation. Calcium ions (orange spheres) are shown.

Protein Backbone Dynamics C_α RMSDs of the proteins, a measure of structural similarity to a reference, were calculated with respect to the first frame of each trajectory. RMSDs for all simulations were below 2 Å during the whole trajectory (**Figure S9**) suggesting that the C2 domains did not undergo large conformational changes in the presence of membranes. Interestingly, simulations of C2A in the absence of Ca^{2+} showed a slightly higher RMSD than simulations in the presence of Ca^{2+} (**Figure S9A**). This was not observed for C2B (**Figure S9B**) for which similar RMSDs were calculated in the presence and absence of Ca^{2+}. Having confirmed that the proteins remain structurally stable during the simulation, local backbone flexibility was analysed. For this, C_α RMSFs, i.e. the time average of the RMSD per residue, were calculated for the structures sampled within the last 200 ns of each simulation. The RMSF describes the local flexibility of a residue with respect to the average conformation of the C2 domain during the inspected time range (see **Section 3.2.6** for details). The mean RMSF of C2A averaged over the three replicates (i.e. initial orientations) indicated regions of high flexibility in the Syt-1 C2A sequence as well as rigid regions with RMSFs below 1 Å (**Figure 31A**). To visualise the position of the flexible regions on the C2 domains, the calculated RMSFs were mapped onto the C2 domain structures (**Figure 31B**, see **Figure S10** for the visualisation of all simulations). Two regions in C2A showed shifts in RMSF depending on the presence of Ca^{2+}, namely residues 169 - 176 (one of the Ca^{2+} binding loops) and residues 187 - 192 (C2A polybasic stretch, see below). For residues 169 - 176, RMSF was reduced from 3 Å in the absence of Ca^{2+} to <1 Å in its presence. Mapping the residues to the structure of C2A showed that they are located within one of the CBLs of Syt-1. Accordingly, reduction of flexibility in the simulation with Ca^{2+} present is in agreement with this loop participating in Ca^{2+}

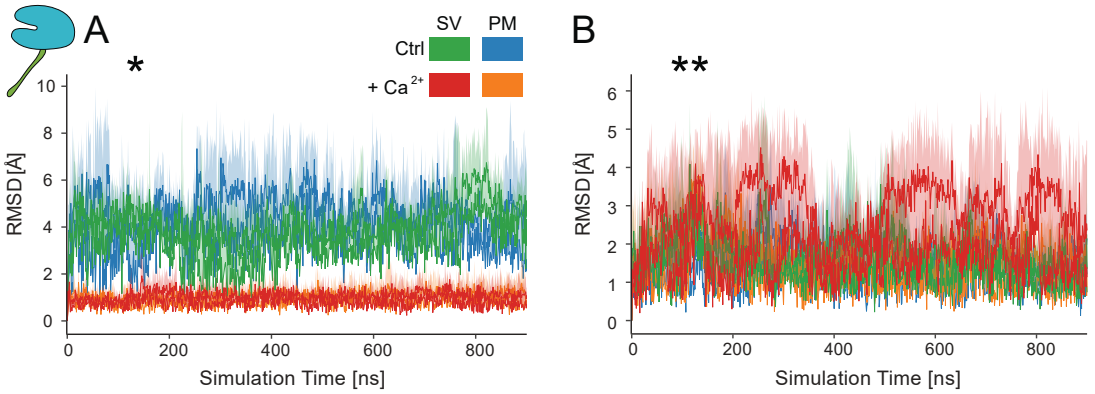


Figure 32: Temporal Evolution of RMSD of Syt-1 C2A Flexible Regions. Structures were aligned based on the positions with low RMSF residues flanking the regions. Mean (solid line) and range (transparent area) of RMSDs of three replicates are shown. **(A)** Evolution of RMSD of Syt-1 residues 172-174. **(B)** Evolution of RMSD of Syt-1 residues 189-190.

coordination. On the contrary, residues 187 - 192 showed a higher RMSF in the simulations of Ca^{2+}-bound Syt-1 C2A interacting with the SV membrane compared with all other simulations. These residues correspond to the basic sequence $_{187}$PDKKKK$_{192}$ and are located at a position homologous to the polybasic region described for C2B ([57], see below). In C2B, the polybasic region is part of the central β-sandwich fold, however, the corresponding residues in C2A form a flexible loop. Flexibility of a third region of high RMSF including residues 248 - 259 was similar in all sets of simulations involving C2A. This region is located in a loop opposite to the CBLs.

The RMSF is a time average and, therefore, does not inform whether there is a single conformational change of a certain residue or whether that shows multiple fast conformational transitions. To investigate the temporal evolution of the flexibility of the loops, RMSDs of the most flexible residues in each loop were calculated. For this, residues of low RMSF flanking a flexible region were aligned with respect to the position of these residues in the first frame of the trajectory (see **Section 3.2.6** for details) and RMSDs of C_α atoms in the flexible region were calculated without further alignment. **Figure 32A** shows the analysis for residues 172 - 174 located within one of the CBLs. The simulations of Ca^{2+}-bound C2A showed a low RMSD below 2 Å, whereas the Ca^{2+} free simulations showed larger RMSDs of more than 4 Å over the whole duration of the simulation. This indicates that Ca^{2+} remained stably bound during the simulation and stabilised the CBLs. Plotting the temporal evolution of the RMSD of residues 189 - 190 (located in the basic loop) for the simulation of Ca^{2+}-bound C2A interacting with the SV membrane, a recurring shift between <3 Å and 4 Å of the RMSD was observed (**Figure 32B**). This indicates that the increase in RMSF of residues 189 - 190 was caused by multiple sequential events. In contrast, RMSDs of these residues in all other simulations generally remained <3 Å.

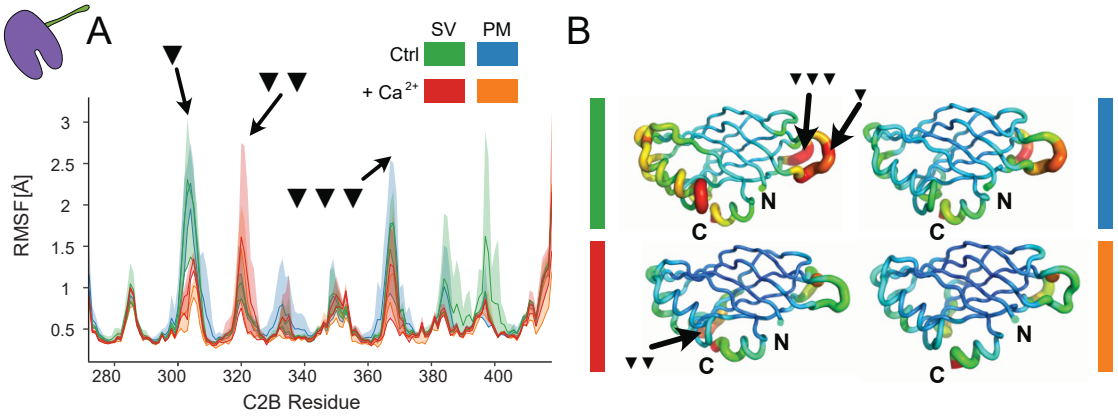


Figure 33: RMSF of Syt-1 C2B. Regions of high flexibility or with differences between the simulations are highlighted (arrows). Residues 300 - 308 are labelled with ▼, residues 317 - 324 with ▼▼ and residues 365 - 370 with ▼▼▼. **(A)** C_α RMSF of Syt-1 272-418. **(B)** Sausage representation of Syt-1 272-418 RMSF. See (A) for colour scheme. Higher RMSFs are indicated by larger diameter and warmer colours of the protein backbone representation. Calcium ions (orange spheres) are shown.

Similar to the analysis of C2A, multiple flexible regions separated in the primary sequence by regions of high stability were observed for C2B (**Figure 33**). Again, a region with a notable difference between the RMSFs of Ca^{2+}-free and the Ca^{2+}-bound simulations corresponding to one of the CBLs (residues 300 - 308) was identified (see **Figure 33**). Another region of high RMSF was located within residues 317 - 324, close to the polybasic region of C2B (residues 324 - 327) [57]. High flexibility (i.e. high RMSFs) was observed for this region in all simulations of Ca^{2+} bound C2B interacting with both membrane models. A third region of high RMSF (residues 365 - 370) corresponds to the second CBL of C2B. In contrast to residues 300 - 308, the CBL discussed above, high flexibility of residues 365 - 370 was observed in all simulations irrespective of the presence of Ca^{2+}. This indicates that residues 365 - 370 were not spatially restrained in the presence of Ca^{2+} and the corresponding loop, therefore, plays a minor role in the coordination of the ion. Moreover, two flexible regions were located within residues 382 - 388 and 395 - 399. The increase in RMSF in these regions was attributed to a single simulation (C2B in the absence of Ca^{2+} interacting with the SV membrane). Nevertheless, the two flexible regions correspond to the termini of one of the three C2B alpha-helices, possibly indicating dynamics of the helix in proximity to the membrane.

Temporal evolution of RMSDs of residues within the flexible loops were analysed for C2B as well. **Figure 34B** shows RMSDs of residues 302 - 305 located within one of the CBLs. In contrast to Ca^{2+}-free C2A for which a high RMSD starting shortly after the beginning of the simulation was calculated, the RMSD of the CBL of C2B increased over the last 500 ns from <3 Å to approximately 4 Å. As the molecular structures of the C2 domains in the absence of Ca^{2+} were generated from Ca^{2+}-bound crystal structures, the observed transition may be the result of slow relaxation after

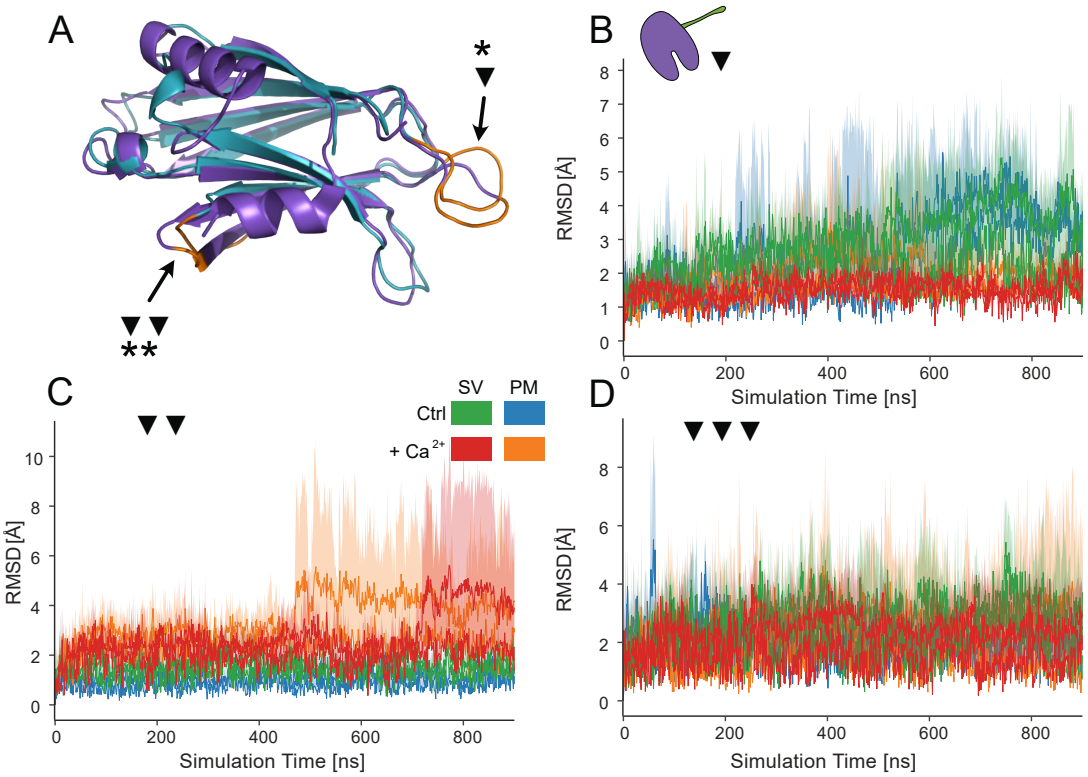


Figure 34: Temporal Evolution of RMSD of Syt-1 C2B Flexible Regions. Mean (solid line) and range (transparent area) of RMSDs of three initial orientations are shown. **(A)** Loops with changes in flexibility upon Ca^{2+}-binding are shown for C2A (cyan) and C2B (purple). **(B)** Evolution of RMSD of Syt-1 C2B residues 302-305. Structures were aligned based on the positions with low RMSF residues flanking the regions. **(C)** Evolution of RMSD of Syt-1 C2B residues 320-321. **(D)** Evolution of RMSD of Syt-1 C2B residues 367-368.

removal of the ion. Residues 320 - 321 located close to the polybasic region showed a low RMSD for the first 400 ns of the simulation (**Figure 34C**). After 500 ns the RMSD of Ca^{2+}-bound C2B interacting with the plasma membrane drastically increased to up to 8 Å. This increase was observed once per simulation, in contrast to Ca^{2+}-bound C2A interacting with the SV membrane for which multiple sequential changes in RMSD were determined (compare **Figures 32B and 34C**). Moreover, the increase in RMSD was observed in a single simulation (starting with the C2 domain oriented parallel to the membrane normal), whereas the two other simulations of the same protein-membrane system showed RMSDs of <4 Å. Similarly, during interactions with the SV membrane, the RMSD of the same residues 320 - 321 of Ca^{2+}-bound C2B increased after 750 ns simulation time (**Figure 34C**). Again, the increase in RMSD corresponded to a single simulation (with the protein initially oriented parallel to the membrane normal) increasing to 8 Å RMSD. Residues 367 - 368, located within the second CBL of C2B, showed no change in RMSD neither between the different conditions nor

over the course of the trajectory (**Figure 34D**). As the dynamics of this loop are not modulated by the interaction conditions sampled here, the loop is possibly not involved in Syt-1 membrane interactions.

In summary, analysis of protein dynamics of Syt-1 C2A and C2B revealed two protein loops modulated by the presence of Ca^{2+}. One loop was part of the CBLs and showed reduced dynamics in the presence of Ca^{2+} and a second loop on the opposite site of both C2 domains with increased dynamics in the presence of Ca^{2+}.

Membrane Contact Angle In addition to local effects such as flexibility of loops, protein-membrane interactions also influence the and orientation of a protein. As shown by the RMSD analyses (see **Figure S9**), the overall conformations of the Syt-1 C2 domains remained stable during the MD simulation. However, the relative orientation of the C2 domains with respect to the membrane might change, depending on the protein-membrane interactions.

To determine whether the orientation of the C2 domains is influenced by the lipid contents of the membrane or the presence of Ca^{2+} ions, average contact angles between the membrane normal and a vector parallel to the β-sandwich fold of the C2 domains were calculated. Residues Tyr151 and Pro179 in C2A or Tyr282 and Pro310 in C2B were chosen for calculation of the C2 domain vector. The median angle within the final 200 ns of each simulation was determined (see **Section 3.2.6**). While the angles of the replicate simulations were very consistent for C2A interacting with the plasma membrane in the absence of Ca^{2+}, differences of up to 30deg between the replicates were observed for Ca^{2+}-bound C2A interacting with the plasma membrane. All other simulations showed differences between the replicates falling within these two extremes (see **Figures S12 and S13** for details). This is reflected by the standard deviations of the mean contact angles shown in **Figure 35**. Simulations of C2A without Ca^{2+} ions showed a contact angle of approximately 70deg (**Figure 35A**). In the presence of Ca^{2+}, this angle was reduced to approximately 50deg in the presence of the SV and plasma membranes. Interestingly, a similar effect was observed for C2B (**Figure 35B**) albeit to a lower extent. Simulations without Ca^{2+} showed an average contact angle of 90deg, whereas those in the presence of Ca^{2+} aligned at an average angle of 80deg. The vector defined by Pro39 and Tyr11 is not exactly parallel to the β-sandwich structure of the C2 domains but provides a robust measure to compare between the simulations. To demonstrate how the intact C2 domains were oriented under the four different angles observed in this thesis, **Figures 35C and D** show representative snapshots of simulations of C2A and C2B in the absence and presence of Ca^{2+}, respectively. In line with the average contact angle, C2A adopted an intermediate orientation with the CBLs facing the membrane even in the absence of the ion. In the simulations including Ca^{2+}, C2A showed a more upright orientation with only the CBLs in close contact to the membrane. C2B, however, displayed a domain orientation nearly parallel to the membrane with a large surface exposed to the lipids during all simulations, irrespective of the presence of Ca^{2+}.

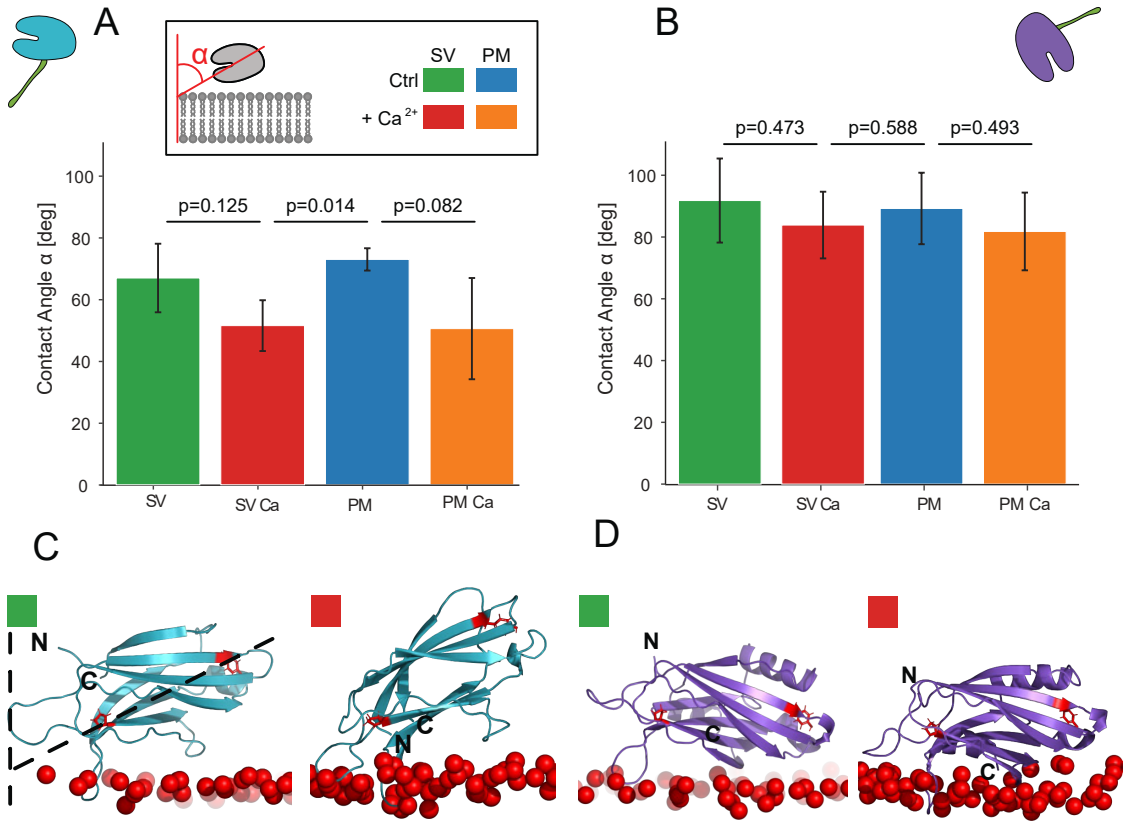


Figure 35: Average Membrane Contact Angles of Syt-1. Schematic representation of calculated contact angles and colour code matrix are shown (inset). **(A)** Mean contact angle of C2A interacting with SV membrane in the absence (green) and presence (red) of Ca^{2+} and with the plasma membrane in the absence (blue) and presence of Ca^{2+} (orange). Mean contact angle, standard deviation of medians of three simulations (as error bars) and p-values from a pairwise two-sided t-test are shown. **(B)** Mean contact angles of C2B interacting with the membranes. For details see (A). **(C)** Examples of C2A (cyan) interacting with the SV membrane in the absence (left) and presence (right) of Ca^{2+}. The vectors between which angles were calculated are shown schematically (left). Tyr151 and Pro179 are shown in red stick representation and phosphor atoms of membrane bilayer phosphate head groups are shown as red spheres. Frames with angles corresponding to the average angle of a protein membrane combination are shown. **(D)** Examples of C2B (purple) interacting with the SV membrane. Tyr282 and Pro310 are shown in red stick representation. For details see (C).

In summary, MD simulations of Syt-1 C2A and C2B in the presence of model membranes indicate a Ca^{2+}-dependent change of the membrane contact angle for C2A, whereas C2B aligned parallel to the membrane with only a small effect of Ca^{2+} on the membrane interaction angle. Possibly, this is involved in the mechanism of Syt-1 function (see discussion at **Section 5.1.4**).

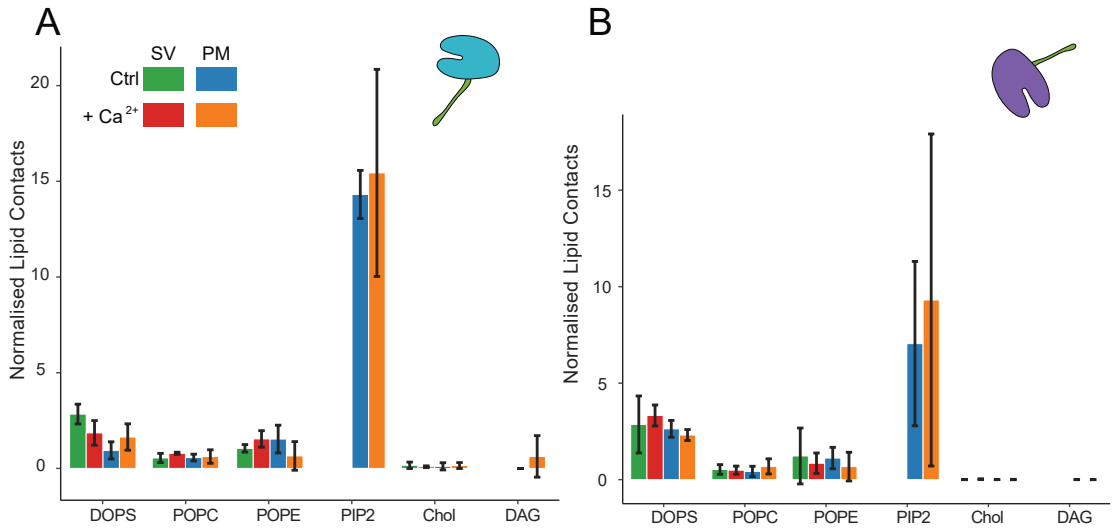


Figure 36: Syt-1 Lipid Preference During MD Simulations. Simulations of Syt-1 interacting with the SV and plasma membrane in the absence (green and blue) and presence (red and orange) of Ca^{2+} are shown. Error bars denote a single standard deviation after averaging across three simulations with different initial orientations. **(A)** Lipid preference of Syt-1 C2A. **(B)** Lipid preference of C2B.

Protein-Lipid Contacts of the C2 Domains Specific protein-lipid interactions are crucial for directing proteins to their target membranes. The C2 domains of Syt-1 were suggested to bind negatively charged lipids, namely PS and PI(4,5)P_2 [54]. Therefore, lipid contacts of the C2 domains in the simulations presented in this thesis were investigated. In the analysis, lipid contacts were defined as one protein atom located within a distance of 2.6 Å to a lipid atom. To focus on stable interactions, only contacts identified in all simulation frames in a 10 ns period were considered in the analysis (see **Section 3.2** for details). For a protein-lipid interaction to occur, a C2 domain first has to come into close contact with the membrane. The proportion of simulation time, and, therefore, the absolute number of frames, in which the C2 domains were in contact to the membrane differed between the simulations for example due to different initial placement of the C2 domains in the replicate simulations. Therefore, the obtained contacts were normalised for each simulation. First, the number of contacts, i.e. the absolute numbers of frames in which the protein interacted with a certain lipid type, were calculated for each lipid. The number of contacts for each lipid was then divided by the sum of contact numbers of all lipids to calculate relative contact numbers. Second, relative contact numbers were divided by the relative number of lipids in a membrane (i.e. the ratio between the number of molecules of this lipid type and the total number of lipids in the membrane). Thereby, relative contact numbers were corrected for stochastic contacts with highly abundant lipids. Moreover, the normalised lipid contacts permitted comparison between simulations of different membrane compositions.

Figure 36 shows the normalised lipid contacts averaged for the three replicates of the same protein membrane combination. For Syt-1 C2A (**Figure 36A**) the analysis shows that DOPS was the most contacted lipid in the absence of PI(4,5)P_2. However, in its presence (i.e. in simulations of C2A interacting with the plasma membrane), PI(4,5)P_2 contact numbers were significantly higher than those for any other lipid. Contact numbers with the zwitterionic bulk lipid POPC were consistently low for all simulations. Binding to zwitterionic POPE was more frequent, however, remained lower than DOPS under most conditions. Similarly, contact analysis for Syt-1 C2B (**Figure 36B**) showed a preference for PI(4,5)P_2 over all other lipids. Interestingly, the number of lipid contacts with DOPS was higher for C2B than for C2A. Notably, the addition of Ca^{2+} ions to the system did not change the number of lipid contacts significantly for both C2A and C2B interacting with the SV and the plasma membrane in comparison to simulations without Ca^{2+}. In summary, our analysis reproduces the published preference of Syt-1 for anionic lipids [54], however, it remains elusive how Ca^{2+} ions play a role in lipid selectivity.

A modulating effect of Ca^{2+} on Syt-1 orientation during membrane interaction was observed (**Figure 35**), however, no change in lipid preference upon Ca^{2+}-binding was observed (**Figure 36**). Nevertheless, binding of Ca^{2+} might influence the exact residues of Syt-1 interacting with the membrane. To investigate this assumption, the number of lipid contacts for each lipid type were counted on a per-residue basis during lipid contact analysis and summed for replicates of the same condition. **Figure 37A** shows the lipid-contacting residues of Syt-1 C2A for interactions with DOPS and PI(4,5)P_2. Distribution of residues interacting with DOPS remained similar in the presence and absence of Ca^{2+} for both the SV and the plasma membrane. They were located within one of the CBLs (residues 231 - 237), within a proximate loop containing an arginine residue (residue 199), involved a lysine residue located in one of the β-sheets (residue 182) or a basic stretch on the lateral side of C2A (residues 189 - 194). Note that the CBL involved in C2A lipid binding treated in this paragraph (residues 231 - 237) is different from the CBL showing high flexibility in the absence of Ca^{2+} (residues 169 - 176, see above). For C2A interacting with the plasma membrane in the presence of Ca^{2+}, lipid interaction sites remained similar to those in the absence of Ca^{2+}. In contrast, for C2A interacting with the SV membrane in the presence of Ca^{2+}, the main part of the interaction shifted from the basic stretch to the CBLs. Similarly, interactions of C2A with PI(4,5)P_2 occured mostly through the basic stretch but also involved the CBLs (see **Figure 37A**). Moreover, contacts involving Lys213, a residue that was not involved in DOPS interactions, were identified. Again, in the presence of Ca^{2+}, interactions shifted towards the CBLs.

Lipid contacting residues in C2B did not change considerably between the presence and absence of Ca^{2+} (**Figure 37B**). Interactions with DOPS mostly involved the polybasic region of C2B (residues 322 - 327) and neighbouring residues such as Lys312 and, interestingly, Glu341. Moreover, the CBLs (residues 364 - 398 and 307 - 309) and neighbouring residues (331 - 332) were in close contact to DOPS lipids. The polybasic region of C2B was the main interaction site for DOPS in the presence

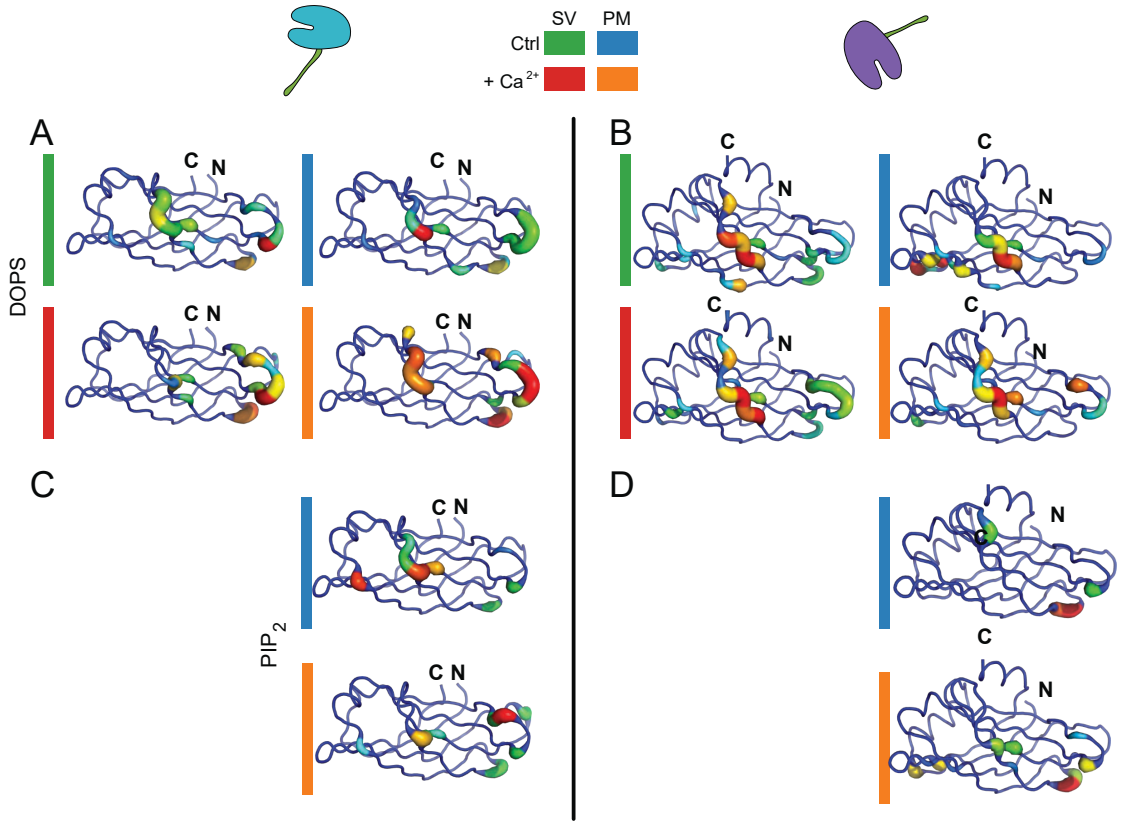


Figure 37: Spatial Distribution of Lipid-Contacting Residues of Syt-1. The sum of lipid contacts of three simulations starting with different initial orientations are shown as sausage representation. Warmer colours indicate a higher number of lipid contacts. Lipid contacts are shown separately for Syt-1 interacting with the SV and plasma membrane in the absence (green and blue) and presence (red and orange) of Ca^{2+}. **(A)** Distribution of Syt-1 C2A residues interacting with DOPS. **(B)** Distribution of Syt-1 C2B residues interacting with DOPS. **(C)** Distribution of Syt-1 C2A residues interacting with PI(4,5)P_2. Note that the SV membrane contains no PI(4,5)P_2. **(D)** Distribution of Syt-1 C2B residues interacting with PI(4,5)P_2.

and absence of Ca^2. However, in simulations involving the plasma membrane in the absence of Ca^{2+}, interactions of C2B with DOPS involved several basic residues on the face of C2B opposing the CBLs. These residues included Arg281, Lys288, Ser344, Arg398 and Arg399. For C2B interacting with PI(4,5)P_2, interaction sites in the absence of Ca^{2+} were Lys331 and Lys332, Lys366 as well as Arg322 close to the polybasic region. Moreover, Lys366 is located in a CBL that showed high flexibility irrespective of the presence of Ca^{2+} in the RMSF analysis (see **Figure 34C**) indicating that PI(4,5)P_2 contacts at this position might be the result of the flexibility of the loop rather than of a specific interaction. PI(4,5)P_2-binding residues in C2B changed slightly in the presence of Ca^{2+}. The loop involving Lys332 and Asn333 remained the main interaction site, however, interactions

with Lys366 were also identified. Additional contacts involved Arg281 and Arg398 located on the face opposing the CBLs (both taking part in the interaction with DOPS in the plasma membrane) as well as Lys325 and Lys313, located within and close to the polybasic region.

In summary, small differences in the interacting residues of C2A and C2B in the absence and presence of Ca^{2+} were identified. Most interactions with negatively charged phospholipids occurred in the CBLs and the polybasic regions of the C2 domains. Notably, additional contacts of C2B with DOPS in the absence of Ca^{2+} in a region of C2B opposite to the CBLs were detected.

Membrane Penetration of C2A and C2B A postulated mechanism of Syt-1 function involves neutralisation of the negative charge of the CBLs by Ca^{2+}-binding, enabling the C2 domains to penetrate into a leaflet of the SV or plasma membrane [54]. To investigate whether this is recapitulated by the MD simulations, membrane penetration of the C2 domains were calculated. The phosphor atoms of the membrane phospholipids were chosen as proxy for the membrane position and penetration depths (i.e. the maximum of distances between protein C_α atoms located within the membrane and their closest membrane phosphate atoms, see (**Section 3.2** for details) were calculated for structures sampled during the last 200 ns of each simulation.

For C2A, this analysis yielded a consistent picture in the absence of Ca^{2+} (**Figures S14 and 38A**). Penetration depths up to 5 Å were observed for all three replicates of each simulation. Simulations diverged more for C2A in the presence of Ca^{2+}. Analysis of Ca^{2+}-bound C2A interacting with the SV membrane yielded histograms of penetration depth with three distinct maxima at approximately 1 Å, 4 Å and 6 Å for the three replicates. Analysis of Ca^{2+}-bound C2A interacting with the plasma membrane produced a maximum penetration depth of approximately 5 Å, however, one of the three replicates showed no penetration at all. Nevertheless, comparing the simulations of different conditions (**Figure 38A**) a shift of penetration depth for C2A from approximately 1 Å in the absence of Ca^{2+} to 5 Å in its presence is visible. This increased membrane insertion was observed to a similar degree for C2A binding to the SV and the plasma membranes. For Ca^{2+}-bound C2A interacting with the plasma membrane, a penetration depth distribution showing free states and states with membrane penetration was observed in a single simulation. This can originate from unstable insertion of the C2 domain to the membrane, i.e. short time periods with the protein inserted into the membrane alternating with short time periods with no insertion. Alternatively, initial membrane insertion can have occurred within the 200 ns time frame of the analysis, so that only part of the analysed simulation frames showed membrane penetration. To differentiate between these two cases, the maximum penetration depths for C2A was plotted against simulation time (**Figure 38C**). Overall, maximum penetration of C2A in the absence of Ca^{2+} was low with the deepest penetrating protein atom located close to the lipid phosphate layer. Nevertheless, most of the analysed frames showed protein C_α below the phosphate layer. In the presence of Ca^{2+}, C2A domains were generally inserted deeper into the membrane. One simulation of C2A interacting

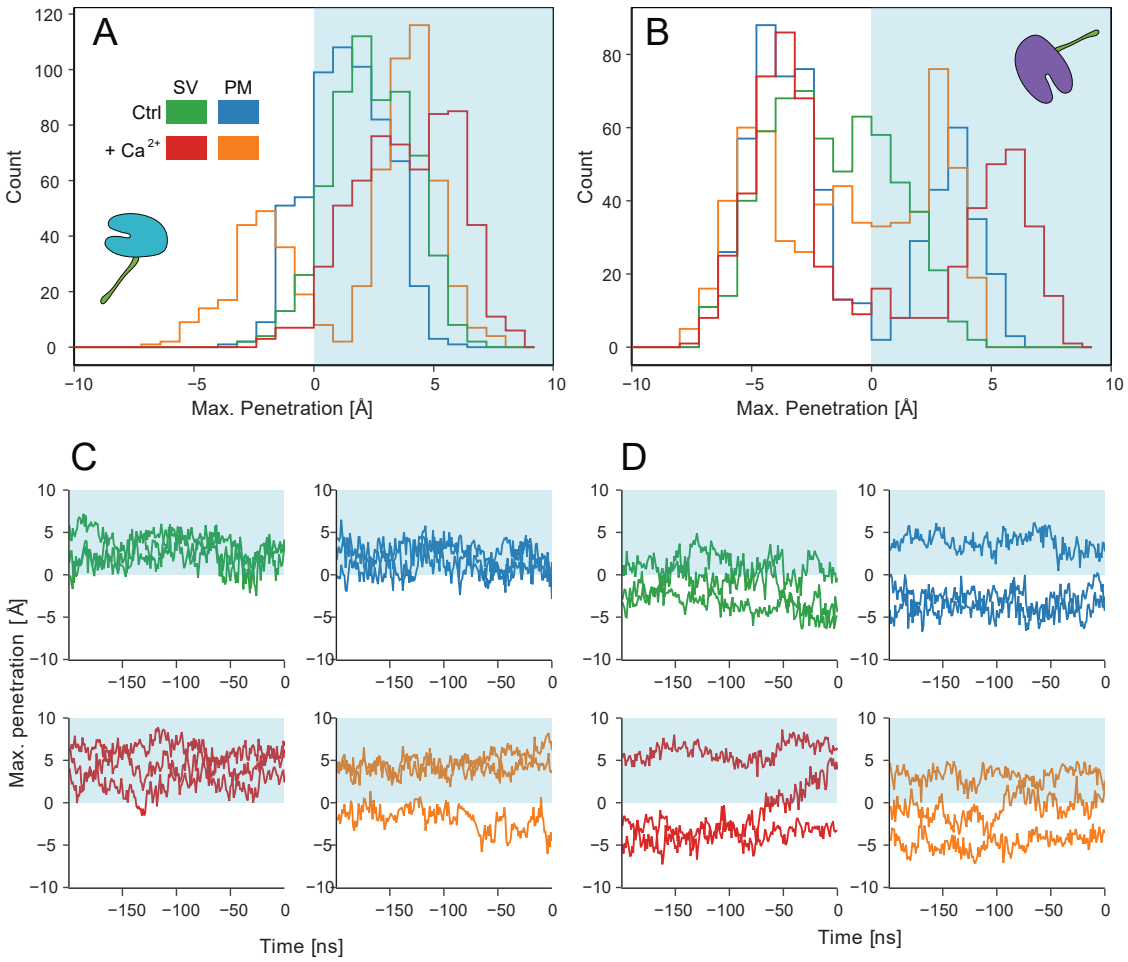


Figure 38: Membrane Penetration of Syt-1. Sums of histograms of three simulations with different Syt-1 starting orientations are shown for the last 200 ns of each simulation. Histograms for Syt-1 interacting with the SV (green) or plasma membrane (blue) in the absence or presence (red and orange) of Ca^{2+} are depicted. **(A)** Histogram of membrane insertion depths of Syt-1 C2A. **(B)** Histogram of membrane insertion depths of Syt-1 C2B. **(C)** Temporal evolution of membrane penetration of Syt-1 C2A interacting with the SV (green) or plasma membrane (blue) in the absence or presence (red and orange, respectively) of Ca^{2+}. **(D)** Temporal evolution of membrane penetration of Syt-1 C2B. Distances corresponding to Syt-1 the membrane bilayer are shown as light blue area.

with the plasma membrane showed no insertion but remained close to the phosphate layer for 100 ns. This simulation shows that close contacts alone are not sufficient for integration that likely also depends on additional factors such as correct orientation and local membrane composition.

Simulations of C2B membrane insertion were more diverse with only a single simulation per condition showing membrane insertion (**Figures S15 and 38B**). In the absence of Ca^{2+}, C2B showed shallow penetration for one simulation involving the SV and for one simulation involving

the plasma membrane. However, penetration into the plasma membrane was deeper than into the SV membrane. C2 domains in the other two simulations of each condition were located close to the membrane without clearly penetrating into the membrane (also see **Figure 38D**). In the presence of Ca^{2+}, C2B was stably inserted into the SV membrane in one simulation, whereas in a second, C2B constantly remained at approximately 5 Å maximum distance to the membrane. In the third simulation, interestingly, C2B did not bind to the upper leaflet of the membrane close to which it was initially placed. In contrast, the C2 domain moved away from the membrane and contacted the lower leaflet of a mirror image of the membrane which it penetrated during analysis time (see **Figure 38D**). Thereby, this simulation is an example of initial membrane insertion occurring within analysis time. The simulations of C2B interacting with the plasma membrane in the presence of Ca^{2+} similarly showed heterogeneous membrane penetration. In one simulation, C2B inserted into the plasma membrane up to approximately 5 Å, in the second it remained stably at 5 Å distance to the membrane, whereas the third simulation showed close contact of C2B with the membrane but no penetration.

Overall, the simulations presented in this work show heterogeneity in membrane insertion of both C2 domains, however, simulations involving C2A were more consistent. Analysis of the insertion depth over time indicates transitions of the C2 domains from unbound to membrane-inserted states but no dissociation of the C2 domains from a membrane-bound to a free state. The membrane-inserted C2 domains are likely energetically favourable, in line with the observation that the C2 domains did not dissociate in the MD simulations once they bound to a membrane. However, this energetically favourable state might not have been sampled in all simulations possibly because the simulation time is short compared to the time periods in which membrane binding and dissociation of Syt-1 occurs.

Membrane Curvature in the Presence of C2 Domains Penetration depth analysis performed in this work cannot discriminate between apparent membrane penetration by integration into the membrane (i.e. the membrane remaining planar and the protein inserting between the lipid molecules) and local bending of the membrane whereby the protein remains outside the lipid bilayer. To investigate these two possibilities, the effect of the presence of the C2 domains of Syt-1 on the respective membranes was analysed.

Figure S16 shows the average phosphor atom z-positions for 64 evenly sized bins of the proximal (i.e. protein facing) and distal leaflets. For comparison between simulations, average surface z-positions were calculated by shifting all z-coordinates so that the lowest coordinate is zero. For most simulations, the upper and the lower leaflet remained planar. The simulations with the most pronounced change in surface z-positions compared to membranes in the absence of protein were selected for every membrane-protein combination (**Figure 39**). While larger surface z-position changes were observed for the proximal leaflets (i.e. the leaflets closest to the C2 domains), the

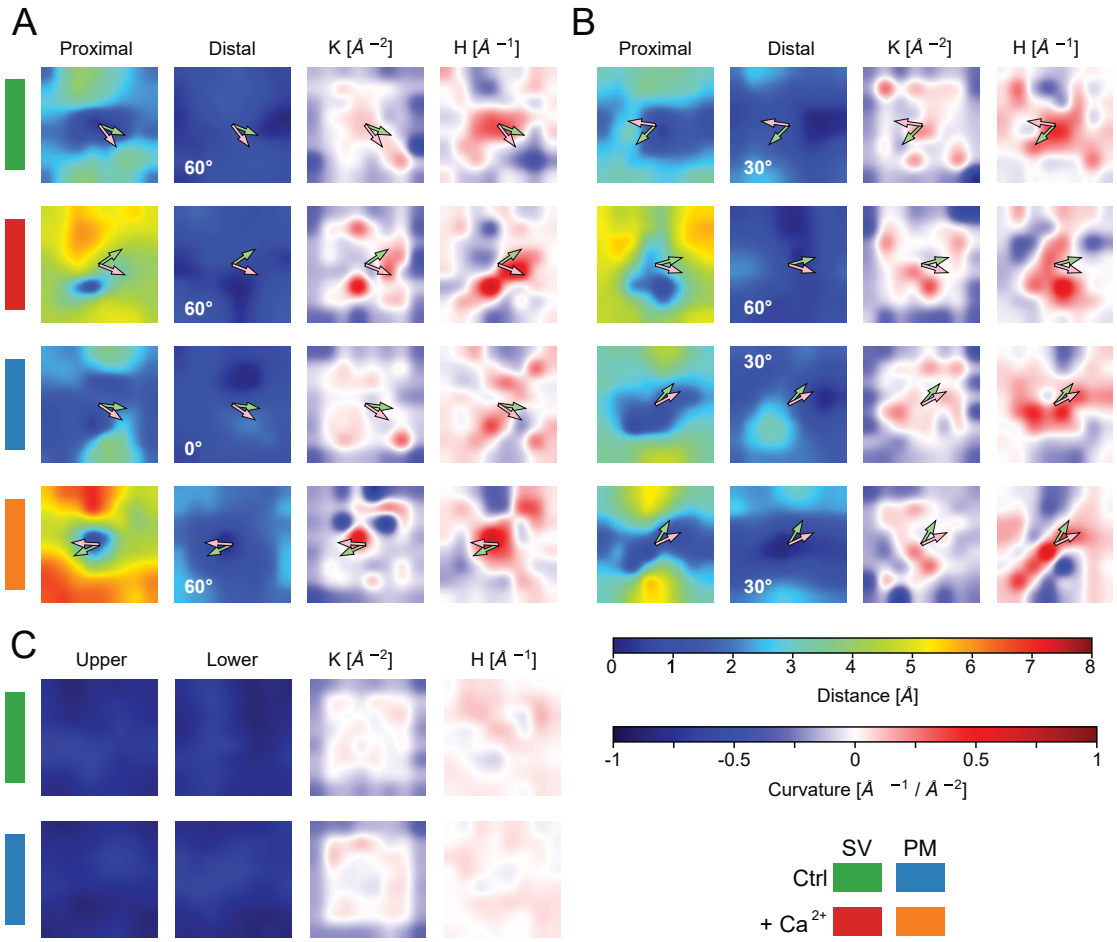


Figure 39: Effects of Syt-1 on the Lipid Membrane. 8 out of 24 simulations of Syt-1 interacting with the SV and plasma membranes in the absence (green and blue) and presence (red and orange) of Ca^{2+} are shown. See **Figure S16** for all simulations. Minimum (green) and maximum (pink) lateral angles of the C2 domain (calculated based on Tyr151 and Pro179 in C2A or Tyr282 and Pro310 in C2B) and colour bars indicating surface elevation and curvature are shown. **(A)** Membrane surface z-positions defined by lipid headgroup phosphor atoms proximal and distal to the C2A interaction site. Gaussian curvature (K) and mean curvature (H) were derived from the proximal surface. **(B)** Same for membranes in the presence of C2B. **(C)** Same for SV (green) and plasma membrane (blue) in the absence of protein.

distal leaflets showed only minor changes of <4 Å compared to membranes in the absence of protein (**Figure 39C**). This indicates that, within the time range investigated here, Syt-1 mainly influenced lipids in the proximal leaflet. Moreover, a change in surface z-positions of up to 7 Å was observed for simulations of C2A in the presence of Ca^{2+} compared to up to 5 Å for simulations in the absence of Ca^{2+} (**Figure 39A**). A similar difference in maximum membrane z-position was determined for C2B (**Figure 39B**), although the effect was less pronounced than for C2A. To evaluate the qualitative observation that Syt-1 C2 domains influence lipid positions, Gaussian and

mean curvature were calculated for each simulation. Both measures describe the two principal curvatures of a point on a surface and indicate the surface geometry. For example, if Gaussian curvature is $K > 0$, the underlying surface has an elliptic point and is of sphere-like geometry forming either an invagination or a protrusion. Mean curvature indicates the average curvature of a point and its magnitude is, therefore, informative on the absolute curvature of the surface. For example, the surface around a point with Gaussian curvature $K > 0$ and mean curvature of $H = 0.5$ can be approximated by a sphere with radius $r = 1/0.5\text{Å}^{-1} = 2\text{Å}$. Indeed, calculated curvatures for the SV and plasma membrane centres in the presence of Syt-1 C2A and Ca^{2+} indicate an elliptic point, i.e. the formation of a cavity on the membrane bilayer. Local mean curvature in the centre of the cavity of $H \approx 1$ is equivalent to a sphere with 1 Å radius.

Overall, analysis of membrane curvature in the presence of C2 domains showed the formation of local cavities in the presence of Ca^{2+} for C2A and C2B interacting with the SV and plasma membranes. Interestingly, membrane penetration was observed previously for all simulations showing modulation of lipid *z*-positions (**Figure 38**), however, not *vice versa*.

In summary, the MD simulations and analyses presented in this section show a decrease in flexibility of the first CBLs in the presence of Ca^{2+}, whereas another loop on the opposite side of the C2 domain (involving residues 187 - 192 in C2A and 321 - 322 in C2B) increases in flexibility in the presence of Ca^{2+} (see **Figure 34D**). In contrast, flexibility of the second CBL remained constant irrespective of the presence of Ca^{2+}. Analysis of the interaction angle of the C2 domains (**Figure 35**) showed that C2A existed in a semi-upright 45deg orientation that shifted towards reduced membrane contact (i.e. a more upright position) in the presence of Ca^{2+}. Conversely, C2B was oriented parallel to the membrane even in the presence of Ca^{2+}. Using lipid contacts as proxy for lipid preference (**Figure 36**), preferred interaction of both C2A and C2B with DOPS and PI(4,5)P_2, irrespective of the presence of Ca^{2+}, was observed. Lipid contacting residues were similar in the absence or presence of Ca^{2+} and interactions with negatively charged phospholipids were primarily loacted in the CBLs and the polybasic region of the C2 domains. Using the lipid phosphates to identify the position of the membrane leaflets, a Ca^{2+}-dependent shift in penetration depth was detected for C2A (**Figure 38**). Analysis of C2B was less conclusive, as the C2 domain did not penetrate the membrane in most of the simulations, possibly due to insufficient sampling. Eventually, the effect of the presence of C2 domains on the lipid membrane was assessed (**Figure 39**) showing a correlation between membrane penetration and modulation of phosphate *z*-positions.

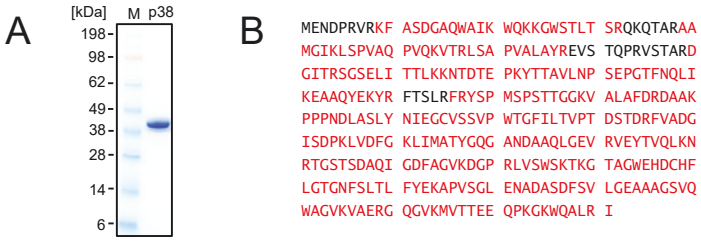


Figure 40: Sequence Coverage of p38. **(A)** p38 separated by gel electrophoresis was analysed by **(B)** LC-MS/MS and identified tryptic peptides were mapped onto the primary sequence. Amino acids identified during MS analysis are coloured red.

4.2 Structural Characterisation of the RNA-Binding Viral Coat Protein p38

The viral coat protein p38 plays a dual role during the life cycle of TCV. It forms the viral capsid coat and functions as VSR to circumvent host defence during invasion. The capsid form of p38 has structurally been characterised by X-ray crystallography [93] and cryo electron microscopy [94]. However, less is known about the conformations and the oligomeric states the protein adopts in solution and, most importantly, during VSR activity. One possible explanation for this lack of structural information is an increase in structural flexibility of p38 during transition to the active form that complicates structural analyses. Therefore, native MS and chemical cross-linking MS were chosen to investigate the dynamic behaviour of p38 and MD simulations were employed to test the implications of the structural models generated from cross-linking restraints *in silico*.

First, the sequence of recombinantly purified p38 (provided by apl. Prof. Dr. Ralph Golbik) was confirmed by LC-MS/MS analysis (see **Section 3.2** for details). For this, proteins within the provided sample were separated by polyacrylamide gel electrophoresis (**Figure 40 A**). The gel band corresponding to the mass of p38 was excised, proteins were hydrolysed by trypsin and analysed by LC-MS/MS. The resulting mass spectra were compared with an *in silico* library of peptides derived from the known sequence of p38. A sequence coverage of 91.2 % confirmed the presence and correct expression of p38 (**Figure 40B**). However, no peptides corresponding to the N-terminal R-domain of p38 (residues 1-41) and the following linker region (residues 42-81) were identified (see **Figure 6** for a schematic representation of the structural organisation of p38). Some unidentified residues are located within clusters of lysine and arginine residues. In these regions, tryptic hydrolysis yields very short peptides that cannot be detected by LC-MS/MS.

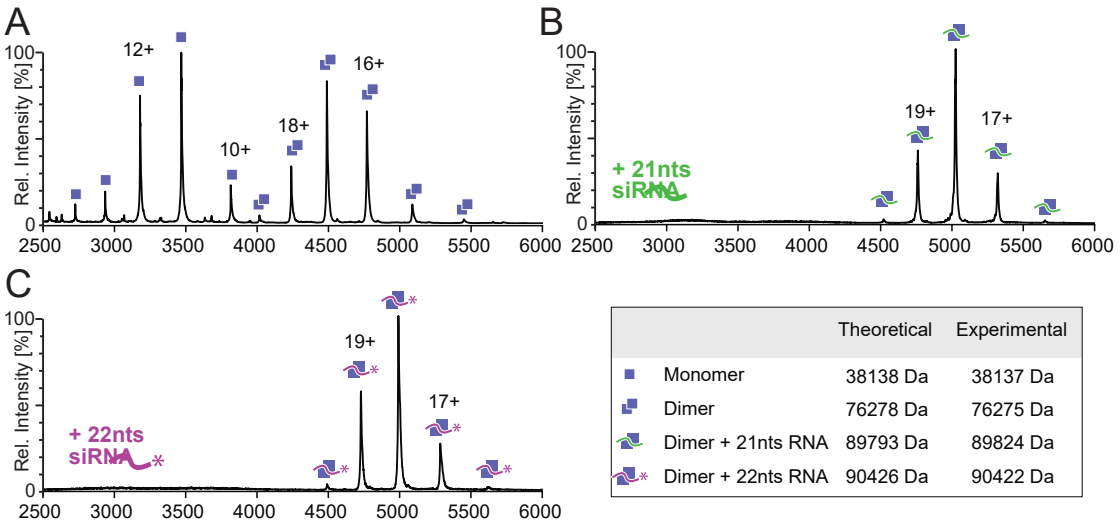


Figure 41: Native MS of p38 in the Presence and Absence of siRNA. See **Table 18** for experimental conditions. **(A)** Native MS of p38 in the absence of RNA. Peaks corresponding in mass to monomeric (single square) and dimeric (two squares) p38 were observed. Charge states for selected peaks are indicated. **(B)** Addition of 21 nts siRNA shifted the spectrum to a single peak distribution (blue squares with green line) corresponding in mass to the RNA-bound p38 dimer. No peaks corresponding to RNA-free monomeric or dimeric p38 were detected. **(C)** Addition of 22 nts siRNA shifted the spectrum to a distribution corresponding in mass to the 22 nts RNA-bound p38 dimer (blue squares with purple line). Experimentally determined and theoretically expected masses are shown (lower right).

4.2.1 Native MS of p38 Binding to siRNAs

The oligomeric state of p38 before and after interaction with siRNA was investigated using native MS. With native MS, multiple co-existing oligomeric states of a protein and binding of RNA molecules can be deduced from the determined masses in an unbiased manner. Therefore, native MS is an ideal tool to determine changes in the oligomeric state of p38 upon interaction with different forms of siRNA.

p38 (provided by apl. Prof. Dr. Ralph Golbik), was transferred into aqueous ammonium acetate solution (see **Section 3.2** for details), protein concentration was determined by UV absorbance and adjusted to <5 µM with ammonium acetate solution. p38 was loaded onto gold-coated borosilicate capillaries and analysed on Synapt G1 HDMS mass spectrometer modified for the transmission of high masses. Two peak distributions with similar peak intensities corresponding to the monomeric and dimeric forms of p38 were observed (**Figures 41A and S17A**). This result confirmed the correct expression of intact p38 including the N-terminus that was not identified by LC-MS/MS. As the spectrum was acquired at a low protein concentration below 5 µM, dimer formation is likely a

Table 18: Conditions for p38 Native MS Experiments.

Protein	RNA concentration	p38 concentration	Figure
p38 WT		4.7 µM ^{14}N	41A, 42A
p38 WT		5.9 µM ^{15}N	42A
p38 WT	<1 µM 21 nts ⋆	2.1 µM ^{14}N	41B, 42B
p38 WT	<1 µM 21 nts ⋆	2.2 µM ^{15}N	42B
p38 WT	<1 µM 22 nts ⋆	4.5 µM ^{14}N	41C, 42C
p38 WT	<1 µM 22 nts ⋆	5.7 µM ^{15}N	42C
p38 WT		2.6 µM ^{14}N, 2.7 µM ^{15}N	42A
p38 WT	<1 µM 21 nts⋆	1.2 µM ^{14}N, 1.2 µM ^{15}N	42B
p38 WT	<1 µM 22 nts⋆	2.6 µM ^{14}N, 2.7 µM ^{15}N	42C
p38 W274A		22.8 µM	43A
p38 W274A	1.45 µM 21 nts	20.9 µM	43B
WT and W274A		17 µM p38 W274A, 1.2 µM ^{15}N p38 WT	43C
WT and W274A	1.6 µM 21 nts	17 µM p38 W274A, 1.2 µM ^{15}N p38 WT	43D

⋆ Free RNA was removed by filtration before analysis.

specific event and not caused by the presence of multiple proteins in a single electrospray droplet, an artefact generated by high protein concentrations during electrospray [173]. Subsequently, siRNAs of 21 nts and 22 nts length were added to the protein. siRNAs are target sequences for the host RISC binding to these siRNAs would be a possible mechanism of the VSR activity of p38. The protein-RNA mixture was incubated on ice for 10 min and free RNA was removed by filtration before native MS analysis was performed (**Figures 41B and C** and **Table 18**). During native MS with either siRNAs, only a single peak distribution was observed corresponding in mass to a p38 dimer bound to 21 nts and 22 nts siRNA, respectively. This indicates that binding of p38 to RNA stabilises a dimeric form of p38 and is in line with findings from Prof. Dr. Hauke Lilie who showed that p38 forms a dimer in solution using analytical ultracentrifugation (oral communication, data not shown).

To gain detailed insights into the molecular process of p38 dimer formation, a combination of unlabelled p38 and p38 produced with ^{15}N as nitrogen source was used (both provided by apl. Prof. Dr. Ralph Golbik). Due to the mass shift introduced by the stable isotope labelling, both populations

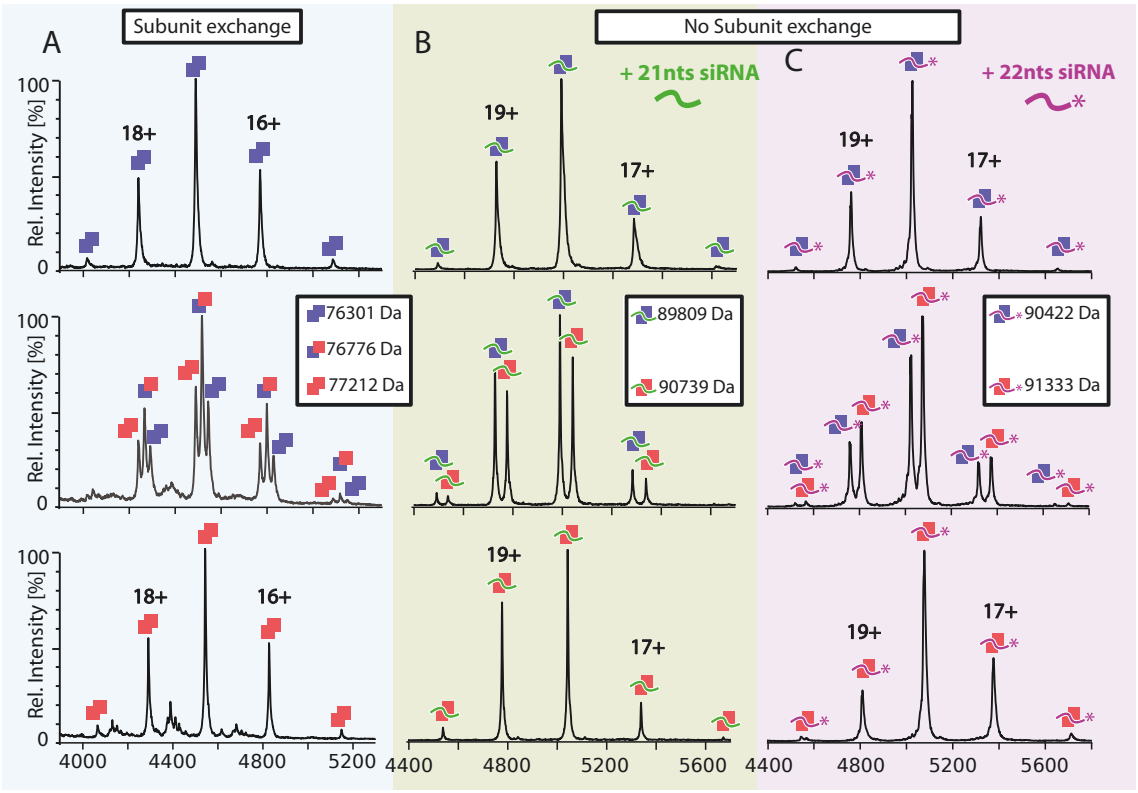


Figure 42: Subunit Exchange Between p38 Dimers is Prevented in the Presence of siRNA. See **Table 18** for experimental conditions. **(A)** p38 dimers exchange subunits. ^{15}N labelled p38 (bottom, red) showed a similar dimer peak distribution as the ^{14}N p38 (top, blue). Mixing at equimolar ratio (middle) yielded a peak distribution with corresponding masses matching the unlabelled dimer, the labelled dimer and a dimer of one labelled and one unlabelled subunit (double squares, red and blue) at 1:2:1 intensity ratio. **(B)** Incubation of ^{14}N (top) and ^{15}N p38 (bottom) with 21 nts RNA resulted in two peak series corresponding to masses of the RNA-bound dimers. Mixing at equimolar ratio (middle) resulted in a peak series corresponding to masses of ^{14}N p38 dimer with RNA and ^{15}N p38 dimer with RNA but no signal for a mixed dimer. **(C)** Addition of equimolar amount of RNA-free ^{15}N p38 to 22 nts RNA-bound ^{14}N p38 (top) resulted in two separate peaks series corresponding to the RNA-bound form of both dimers (middle). ^{14}N (top) and ^{15}N (bottom) p38 bound to 22 nts RNA are shown for comparison.

can be distinguished by native MS. First, ^{15}N p38 was analysed by native MS in the absence of RNA, confirming that ^{14}N and ^{15}N p38 show similar charge state distributions (**Figure 42A and S18A** and **Table 18**). Second, both forms were mixed at equimolar ratio to test the formation of mixed dimers between the two populations. Indeed, peak distributions with a 1:2:1 intensity ratio were observed corresponding in mass to a ^{14}N/^{14}N dimer, a mixed ^{14}N/^{15}N and a ^{15}N/^{15}N dimer. As both forms were mixed immediately prior to MS analysis, the experiment indicates fast (<5 min) exchange of monomers between the differentially labelled dimers. Next, the experiment was repeated in the presence of 21 nts (**Figures 42B, 42C** and **Table 18**) and 22 nts siRNAs (**S18B and S18C** and

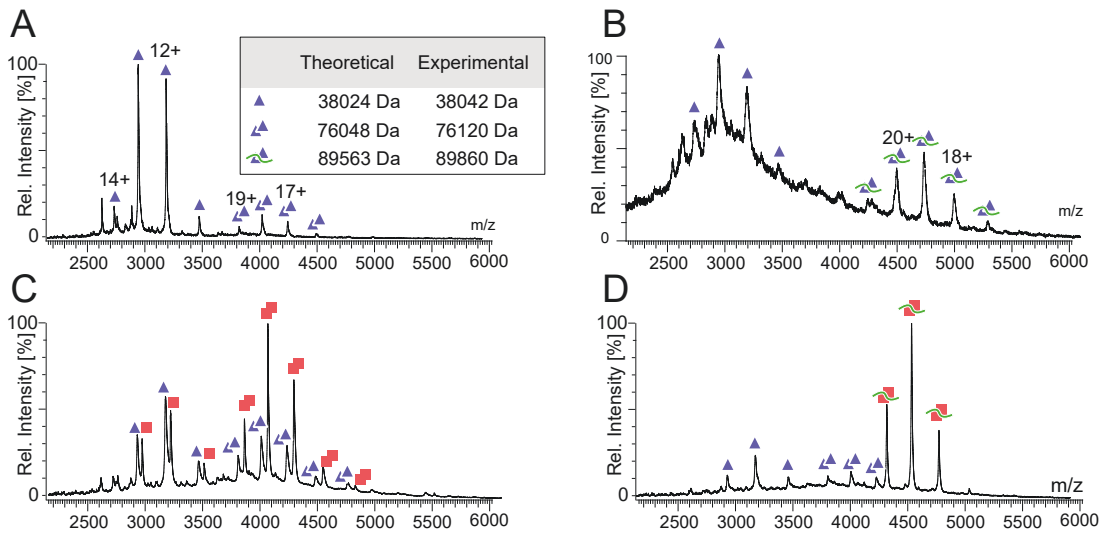


Figure 43: Native MS of p38 W274A in the Presence and Absence of RNA. See **Table 18** for experimental conditions. **(A)** Analysis of p38 W274A in the absence of RNA yielded two peak distributions corresponding to p38 monomers (triangles) and dimers (double triangles). **(B)** Addition of 21 nts RNA lead to the formation of RNA-bound p38 dimers (double triangles with green line). **(C)** Addition of ^{15}N p38 (red) to ^{14}N p38 W274A. Peak distributions corresponding to monomers and dimers of proteins were recorded. **(D)** Addition of 21 nts RNA to (C) shifted ^{15}N p38 completely to the RNA-bound form, whereas no RNA-bound dimer was detected for p38 W274A.

Table 18). As observed for ^{14}N p38, only RNA-bound dimers were identified and no peaks for p38 without the nucleic acid were observed (top and bottom spectra). In contrast to mixing of RNA-free p38, native MS analysis of an equimolar mixture of the differentially labelled RNA-bound dimers (middle spectrum) showed a spectrum corresponding to the sum of its components. The absence of a ^{14}N/^{15}N mixed-isotope RNA-bound form indicates that the p38 dimer, after capturing the nucleic acid, remains stably associated and monomer exchange between dimers does not take place or is greatly reduced compared to the RNA-free form.

Two GW motifs in p38 have been suggested to play a role in p38 binding to the AGO proteins and for its VSR function [91]. One motif is located in the N-terminal R-domain, whereas the second motif is located in the folded P-domain. The effect of mutating the tryptophan residue in this second GW motif to alanine was analysed. Similar to the analysis of the wild-type variant, p38 W274A (provided by apl. Prof. Dr. Ralph Golbik) was transferred into ammonium acetate solution and the oligomeric state of the protein was analysed by native MS. The spectrum of the protein in the absence of RNA showed two peak distributions corresponding to the masses of monomeric and dimeric p38 W274A (**Figure 43A**). In contrast to the wild-type protein (**Figure 41A**), the peak series corresponding to the monomeric protein was of higher intensity than that of the dimer. This implies a lower propensity of dimer formation when the second GW motif is inactivated. Next, 21 nts siRNA

was added to the protein to assess if RNA-induced dimers are also formed by the mutated protein (**Figure 43B**). Indeed, a peak distribution corresponding to the mass expected for p38 W274A bound to 21 nts siRNA and a second peak distribution that could not be unambiguously annotated but likely corresponds to the RNA-free monomer was observed. Subsequently, the formation of heterodimers between the wild-type protein and p38 W274A was tested. Mixing of equimolar amounts of ^{14}N p38 W274A and ^{15}N p38 wild-type followed by native MS analysis resulted in a spectrum corresponding to that of the ^{15}N wild-type form (not shown). However, increasing the W274A to wild-type ratio to 10:1 resulted in four distinct peak distributions corresponding to monomers and dimers of p38 W274A and wild-type (**Figure 43C**). No mixed dimer was observed in this spectrum. Interestingly, addition of 21 nts siRNA at a wild-type:W274A:RNA ratio of approximately 1:10:1 affected only the peaks of p38 wild-type (**Figure 43D**). In comparison to the spectrum without RNA, no peaks corresponding to RNA-free ^{15}N p38 were observed. Instead, a new peak series matching the RNA-bound dimer was recorded and no mixed dimers between the mutant and the wild-type protein were observed. This indicates that, although p38 W274A is capable of binding to 21 nts siRNA, the fast subunit exchange between dimers is impaired by the mutation. Moreover, binding to 21 nts siRNA is likely lower in the mutant compared to the wild-type protein.

4.2.2 Structural Characterisation of p38 by Chemical Cross-Linking

To characterise the structure p38 adopts in solution, chemical cross-linking was employed. Thermal unfolding CD studies performed previously by apl. Prof. Dr. Ralph Golbik indicated that the N-terminal R-domain of p38 undergoes temperature-dependent structural changes. To monitor changes in p38 structure during this transition, chemical cross-linking was performed at 25 °C and at 5 °C.

Identification of Cross-Linked Peptide Pairs The amount of structural information a cross-link holds depends on the distance it bridges with cross-links from short cross-linker molecules generally encoding more information. However, cross-linking efficiency also decreases with the distance a cross-linker can bridge as the likelihood of a second reactive residue in sufficient distance is lower for shorter cross-linkers. Therefore, cross-linker length is always a trade-off between overall efficiency and information content per cross-link. Therefore, two different cross-linker molecules were employed for the study of p38, namely BS2G and EDC. BS2G has a spacer arm distance of 7.7 Å and reacts with primary amines such as lysine residues. Note that the cross-linker also shows reactivity towards serine, threonine and tyrosine residues. EDC, on the other hand, links lysine residues in direct contact with carboxylic acids such as glutamic acid, aspartic acid and the C-terminus of the protein. For identification of cross-linked residue pairs, cross-linked proteins were

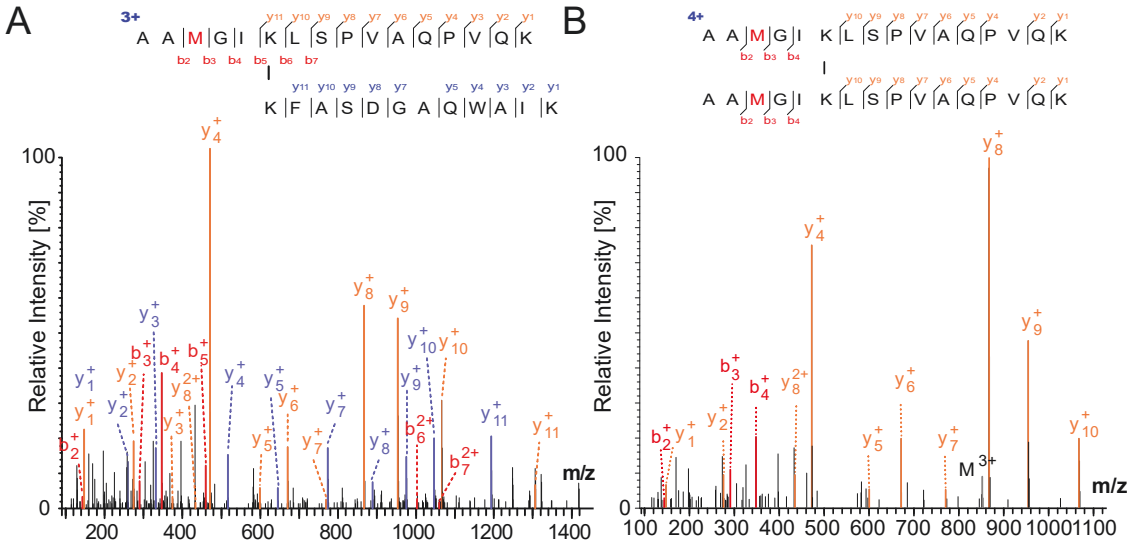


Figure 44: Annotated Cross-Link Spectra After In-Gel Hydrolysis. (A) Annotated spectrum of the cross-linked peptide pair $_{39}$AAMGIKLSPVAQPVQK$_{54}$-$_{9}$KFASDGAQWAIK$_{20}$. Peptides were analysed by LC-MS/MS after cross-linking of 10 µM p38 using 100 µM BS2G at 25 °C, excision of the monomer gel band and in-gel hydrolysis (see **Figure 45A**). **(B)** Annotated spectrum for the homomultimeric peptide pair $_{39}$AAMGIKLSPVAQPVQK$_{54}$-$_{39}$AAMGIKLSPVAQPVQK$_{54}$. Peptides were obtained after excision of the dimer band of 10 µM p38 treated with 30 µM BS2G (see **Figure 45B**). Observed fragment ions are mapped to the peptide sequences (below).

enzymatically hydrolysed and the resulting cross-linked peptide pairs were analysed by LC-MS/MS. Similar to protein identification by LC-MS/MS, (cross-linked) peptide ions were selected in the quadrupole and fragmented by collisional activation. The resulting fragment ions were compared to a database generated by *in silico* fragmentation of all possible peptide pair combinations modified with the cross-linker mass.

As an example, a representative cross-link spectrum obtained after collisional activation of the peptide pair $_{39}$AAMGIKLSPVAQPVQK$_{54}$-$_{9}$KFASDGAQWAIK$_{20}$ is shown (**Figure 44A**). Peaks matching the expected masses from the database are highlighted and the positions of the fragmentation events are indicated in the peptide pair sequence. In this spectrum, both peptides are covered by series of consecutive fragment ions. This is a clear indicator of a correct spectrum annotation as random (false-positive) matches occur at every position with equal probability. The cross-link positions in both peptides (residue numbers 44 and 9, respectively) are not overlapping and separated by 34 residues in the primary sequence of p38. Therefore, the cross-link can either derive from self-interaction of a single copy of p38 (intra-protein cross-link) or from interaction of multiple copies of p38 (inter-protein cross-link). In contrast, the cross-link PSM of the peptide pair $_{39}$AAMGIKLSPVAQPVQK$_{54}$-$_{39}$AAMGIKLSPVAQPVQK$_{54}$ contains two identical peptide

sequences (**Figure 44B**). This cross-link can only derive from interaction of multiple copies of p38 and, therefore, indicates oligomerisation of p38. In general, cross-links with non-overlapping sequences originating from multiple copies of the same protein (inter-protein cross-link) are indistinguishable from interactions of the same residues within a single polypeptide chain (intra-protein cross-link). Note that, in this thesis, cross-links with non-overlapping sequence are shown as intra-protein cross-links mapped onto a single primary sequence, although it should be kept in mind that they can also be the result of inter-protein contacts.

Cross-Linking at Room Temperature Initially, p38 was treated with increasing amounts of BS2G and EDC at room temperature before separating cross-linked proteins by polyacrylamide gel electrophoresis (**Figure 45A** and **Figure 45E**). p38 treated with BS2G showed a gel band corresponding to the molecular weight of monomeric p38 at approximately 38 kDa (marked with $*$) under all conditions. In addition, a second band corresponding to the molecular weight of the protein dimer (marked with $**$) was visible at higher cross-linker concentration. Similarly, 0.5 mM - 1.5 mM EDC were applied to 10 µM p38 and the cross-linked proteins were separated by gel electrophoresis. In addition to the gel bands for monomeric p38, faint bands corresponding to dimeric p38 increasing with cross-linker concentration were detected with EDC.

Gel bands corresponding to monomeric and dimeric p38 were excised, proteins were hydrolysed and the resulting peptides were analysed by LC-MS/MS. Cross-linked peptide pairs were identified for the three lowest cross-linker concentrations. Experiments were carried out in three replicates and the resulting spectra were analysed using pLink software [144]. Only cross-links manually inspected and validated in at least two replicates were retained (see **Section 3.2.4** for details). The cross-links obtained from the monomer bands at cross-linker concentrations ranging from 10 µM - 100 µM, corresponding to a molar BS2G excess of 1 - 10 x, show various contacts within the N-terminal R-domain, indicating a high flexibility of this region (**Figure 45B**). Cross-links between the R- and the N-terminal half of the S-domain as well as between the R- and P-domains indicate proximity of the R-domain to both regions, again confirming its structural flexibility. With increasing cross-linker concentration the number of cross-links from the R- to the S- and P-domain increased. Cross-link analysis of peptides extracted from the p38 dimer band (**Figure 45C**) yielded a similar picture of intra-protein connectivity. Moreover, several cross-linked peptide pairs with overlapping sequences (see above) were identified at equimolar cross-linker concentration confirming that p38 forms a dimer in solution. Gel bands corresponding to monomeric and dimeric p38 after EDC cross-linking were excised, proteins were hydrolysed and peptides analysed by LC-MS/MS. In contrast to BS2G, however, EDC cross-links were only identified in the monomer gel band. This can be attributed to the low protein concentration in the dimer gel bands that was likely insufficient for identification of cross-linked peptide pairs. From the monomer band, four cross-links were identified in total (**Figure 45F**), both located within the folded S- and P-domains.

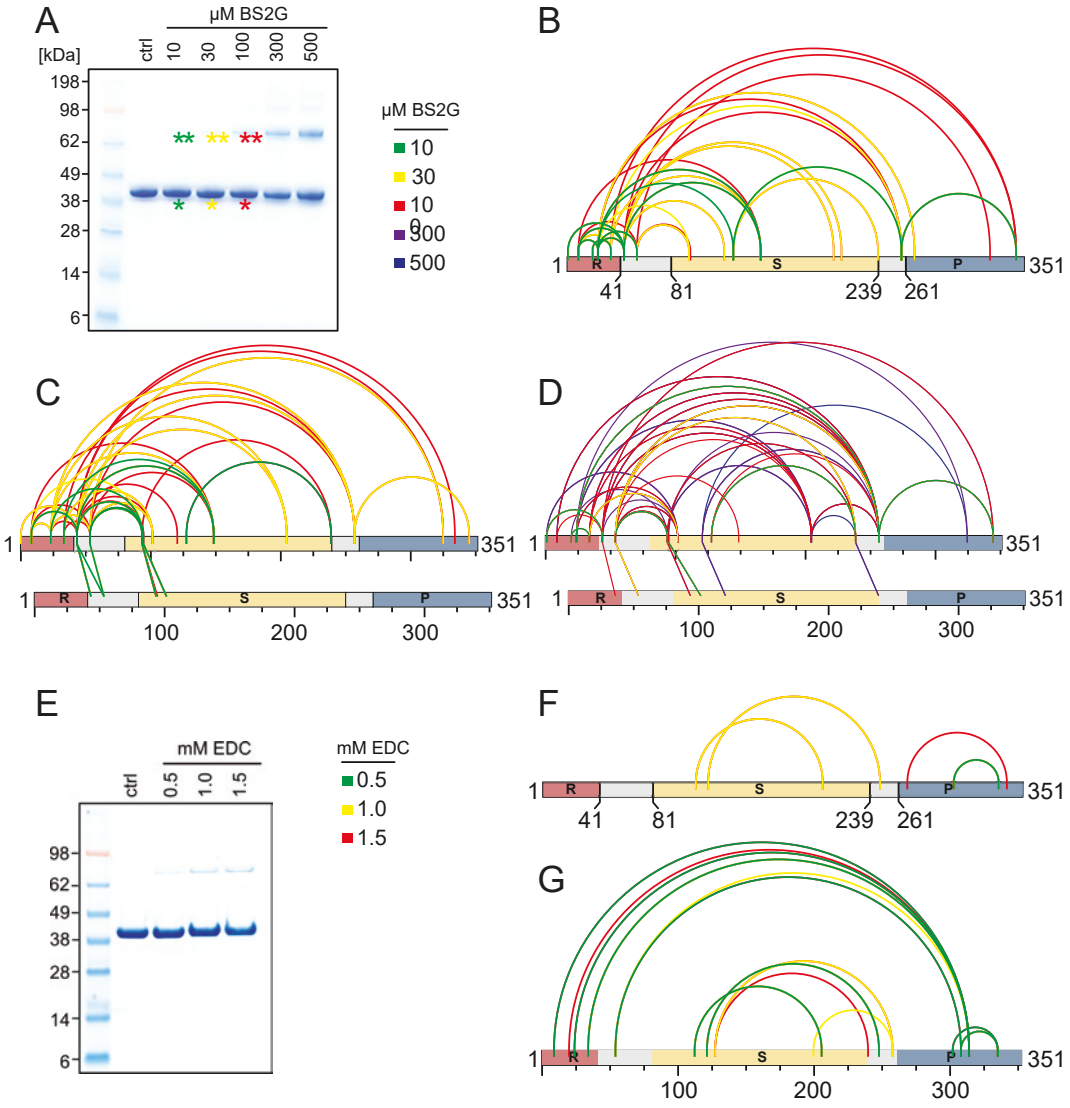


Figure 45: Cross-Linking of p38 at 25 °C. Cross-link PSMs were filtered to retain only PSMs identified in all three replicates and manually validated in at least two of three replicates. **(A)** Polyacrylamide gel electrophoresis of 10 µM p38 without (ctrl) and with increasing concentrations of BS2G. Cross-linked p38 shifted towards higher molecular masses. **(B)** Cross-links identified after LC-MS/MS analysis of peptides obtained by in-gel hydrolysis of p38 treated with 10 µM (green), 30 µM (yellow) and 100 µM (red) BS2G. The corresponding bands of the polyacrylamide gel electrophoresis are indicated by an asterisk (∗) in panel (A). **(C)** Cross-links identified after LC-MS/MS analysis of peptides obtained by in-gel hydrolysis of the gel bands corresponding to p38 dimers (bands are indicated by two asterisks (∗∗) in panel (A)). Cross-links from peptides with overlapping sequence are shown as inter-protein cross-links between two bar representations of the p38 primary sequence. **(D)** Cross-links identified after in-solution hydrolysis of 10 µM p38 cross-linked with 10 µM (green), 30 µM (yellow), 100 µM (red), 300 µM (purple) and 500 µM (blue) BS2G.

Figure 45 ***(cont.)*****: (E)** Polyacrylamide gel electrophoresis of 10 µM p38 with increasing concentrations of EDC. **(F)** Cross-links identified after LC-MS analysis of peptides obtained by in-gel hydrolysis of p38 treated with 0.5 mM (green), 1 mM (yellow) and 1.5 mM (red) EDC and excised from the monomer bands of the polyacrylamide gel electrophoresis. **(G)** Cross-links identified after in-solution hydrolysis and SEC separation of 10 µM p38 cross-linked with 0.5 mM (green), 1 mM (yellow) and 1.5 mM (red) EDC.

To complement in-gel cross-link analysis, direct hydrolysis of cross-linked p38 followed by LC-MS/MS analysis was employed. The cross-links obtained after treatment of 10 µM p38 with 10 µM - 500 µM BS2G and in-solution hydrolysis confirmed the interactions observed with in-gel analysis (**Figure 45D**). Interestingly, the BS2G concentrations required to identify the majority of cross-links in samples treated with in-solution hydrolysis was higher than in samples hydrolysed in-gel. For example, whereas most cross-links in the R-domain were identified at 10 µM BS2G concentration in the in-gel analysis, many of those cross-links were only identified at >30 µM cross-linker concentration when employing in-solution hydrolysis without enrichment of cross-linked peptide pairs. A possible explanation for this increase in cross-linker concentration is the lack of preseparation after in-solution hydrolysis resulting in a higher concentration of non-protein ions that complicated spectral assignment and scoring for these samples. As the final list of reported cross-links is determined by a fixed false-discovery rate that is applied after sorting the matches by score, overall lower scores of the correct PSMs result in less cross-links in the final list. In-solution analysis of 10 µM p38 cross-linked with 0.5 mM - 1.5 mM EDC revealed a total of 13 cross-links within the S- and P-domains and between the R-domain and the P-domain (**Figure 45G**). Thereby, cross-linking with EDC confirmed the high flexibility of the linker between the R- and S-domains and shows that the R-domain forms close contacts with the P-domain. In summary, cross-linking with EDC and BS2G results in a consistent interaction pattern after in-gel and in-solution hydrolysis.

Cross-Linking at 5 °C Cross-linking can produce complex mixtures of cross-linked peptide pairs, linear peptides and small molecules such as buffer salts or non-reactive cross-linker molecules that can interfere with cross-link identification. Complexity can be reduced prior to LC-MS/MS analysis by separating the hydrolysed peptides using SEC (**Figure 46**). Larger cross-linked peptide pairs can access less of the column volume and elute first (region I). This can, for example be monitored by light absorption of the peptide bond at 214 nm. Linear peptides elute next (region II) followed by other components such as salt ions (region III). The putative cross-link peptides are collected (grey shaded area) and analysed by LC-MS/MS. A spectrum of the peptide $_{242}$GSTSDAQIGDFAGVKDGPR$_{261}$-$_{24}$KGWSTLTSR$_{32}$ identified after peptide SEC is shown in **Figure 46B**.

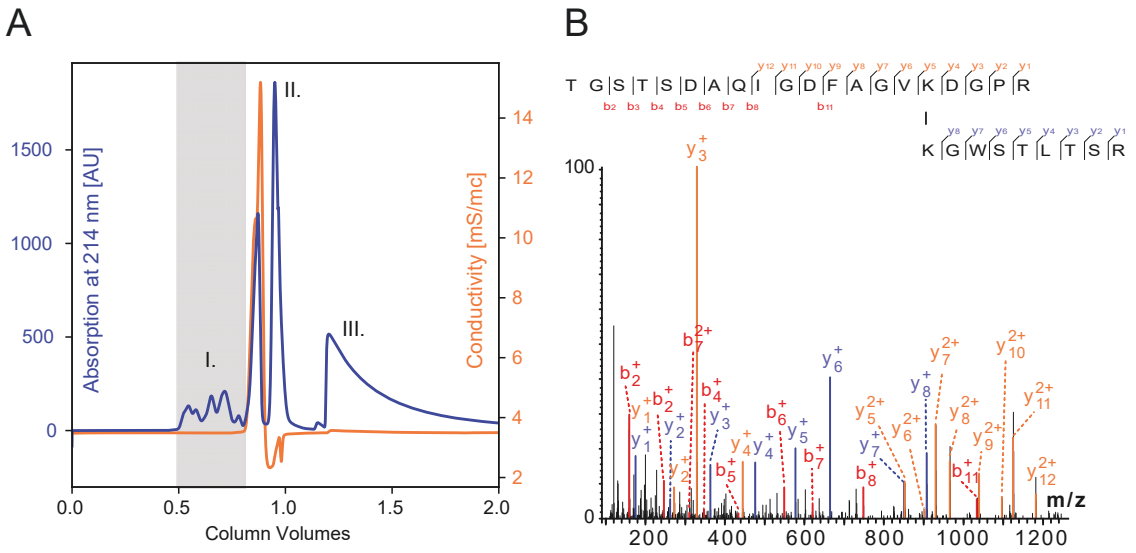


Figure 46: Chromatogram of Peptide SEC and Resulting Cross-Link PSM. (A) Chromatogram of SEC to separate long cross-linked peptide pairs (I) from shorter peptides (II) and buffer salts (III). p38 was cross-linked with BS2G at 5 °C in the presence of siRNA (see **Figure 47**). The fractions analysed by LC-MS are highlighted in grey. **(B)** Annotated spectrum of the cross-linked peptide pair $_{242}$GSTSDAQIGDFAGVKDGPR$_{261}$-$_{24}$KGWSTLTSR$_{32}$ identified from elution fractions shown in (A).

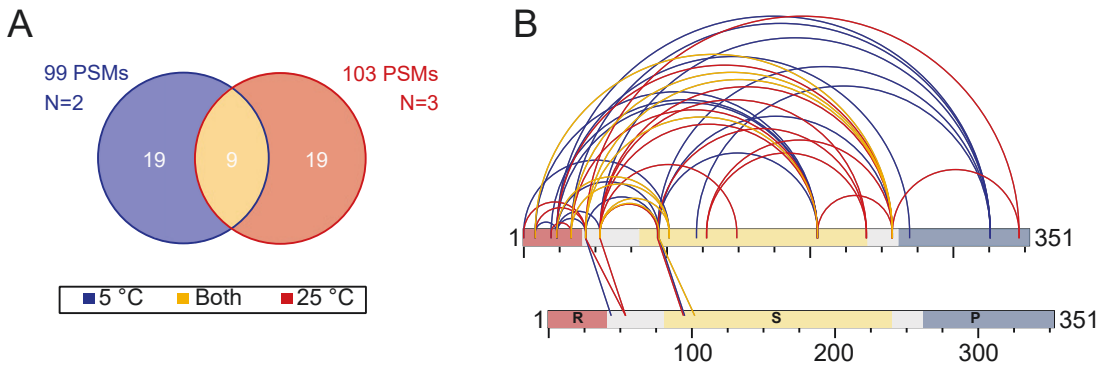


Figure 47: Cross-Linking of p38 with BS2G at 5 °C. (A) Cross-links identified after in-solution hydrolysis of 10 µM p38 treated with 100 µM BS2G at 5 °C. Cross-links identified only at 5 °C (blue) are compared to cross-links obtained from cross-linking at 25 °C (red, see **Figure 45D**). Cross-links found under both conditions are shown (yellow). **(B)** Area-proportional Venn diagram of cross-links identified at 5 °C (blue), 25 °C (red) and under both conditions (yellow). The numbers of validated PSMs and performed replicates for each condition are shown. Cross-links were manually validated in at least two replicates.

To characterise temperature-dependent structural changes in p38, the protein was analysed by cross-linking at 5 °C. For this, tryptic in-solution hydrolysis was employed followed by enrichment of cross-linked peptide pairs by SEC. In total, 28 cross-links were identified from in-solution hydrolysis of p38 cross-linked with BS2G at 5 °C. To compare the identified cross-links with those at 25 °C, a Venn diagram grouping residue-residue pairs as identified at 5 °C, 25 °C or under both conditions was generated (**Figure 47A**). Nine interactions identified at a 5 °C were also present at 25 °C, whereas 19 cross-links identified at 25 °C were not reproduced at 5 °C. Cross-links observed at both temperatures connected the R-domain either to the N-terminal half of the S-domain or to the linker between S- and P-domains (**Figure 47B**). Interactions identified only at 25 °C are evenly distributed across the primary sequence of p38 indicating no specific structure. In contrast, the cross-links exclusively observed at 5 °C are contacts between single lysine residue in the P-domain and multiple residues in the R- and S-domains as well as between the S- and R-domains. In summary, similar connections from the R- to the S- and P-domains were identified at both temperatures, although different residues were detected in the different experiments. Moreover, a change in the interaction pattern was not observed for the R-domain that was supposed to exhibit increased secondary structure at 5 °C.

Cross-Linking in the Presence of siRNA A key function of p38 is counteracting the antiviral response of the host, presumably by scavenging siRNAs. After demonstrating the ability of recombinant p38 to bind siRNA of different lengths (see **Section 4.2.1**) the effect of siRNA on the p38 structure was analysed by cross-linking MS. For this, p38 was incubated with 21 nts siRNA at 2:1 stoichiometry corresponding to an RNA-bound dimer as identified by native MS and ultracentrifugation (performed by Prof. Dr. Hauke Lilie). For comparison with the cross-links obtained in the absence of siRNA, the reaction was carried out at 5 °C. A Venn diagramm summarising the residue-pairs identified in the presence and absence of siRNA is shown in **Figure 48A**. 14 unique contacts located within the R-domain and between the R- and S-domains were observed under both conditions, likely static interactions that remain unchanged after siRNA-binding (**Figure 48B**). A total of 15 cross-links was only identified in the absence of siRNA. These cross-links involve residues in the N-terminal part of p38 but also include connections from the R-domain to the P-domain. In contrast, 30 unique cross-links identified only in the presence of siRNA were distributed along the sequence of p38 and involved long-range R- to P-domain contacts as well as several medium-range interactions within the S-domain. Interestingly, two new homo-oligomeric cross-links were detected in the C-terminus of the S-domain after addition of siRNA. Analysis of p38 with EDC yielded 5 cross-links identified in the presence and absence of 21 nts RNA under both conditions, 12 cross-links were only identified in the presence and 7 only in the absence of 21 nts siRNA (**Figure 48C**). Overall, the cross-links confirmed the interaction pattern observed after cross-linking with BS2G (**Figure 48D**). Notably, a EDC cross-link indicating homo-oligomerisation via the linker region between the S- and P-domains was consistently identified

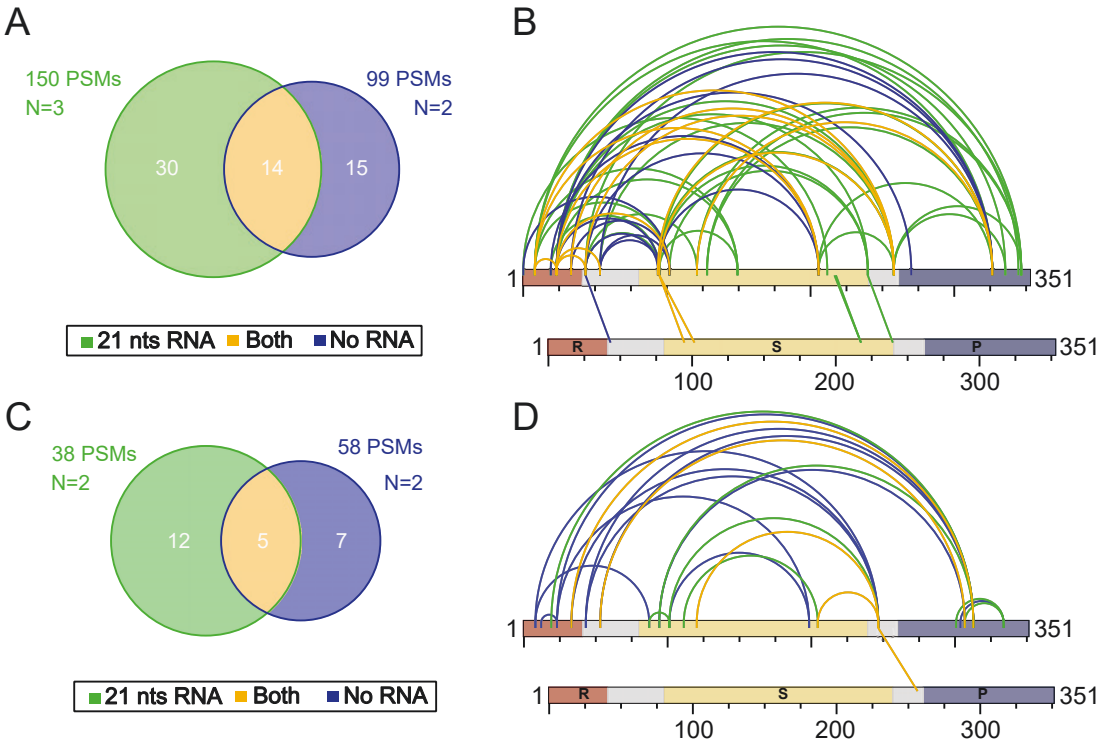


Figure 48: Cross-Linking of p38 in the Presence and Absence of 21 nts siRNA at 5 °C. (A) Area-proportional Venn diagram of cross-links identified with BS2G. The number of validated PSMs and performed replicates for each condition are shown. **(B)** Cross-links identified with BS2G after in-solution hydrolysis of 10 µM p38 treated with 100 µM BS2G at 5 °C in the presence of 5 µM 21 nts siRNA. **(C)** Area-proportional Venn diagram of cross-links identified with EDC. **(D)** Cross-links identified with EDC after in-solution hydrolysis of 5 µM p38 treated with 2 mM EDC and 5 mM Sulfo-NHS at 5 °C in the presence of 2.5 µM 21 nts siRNA. Cross-links identified only in the presence (green) or absence (blue) of RNA or under both conditions (yellow) are shown. Cross-links were manually validated in at least two replicates.

under all conditions. Therefore, homo-oligomerisation via the C-terminus of the S-domain does not uniquely occur in the presence of siRNA as suggested by the BS2G cross-links. Overall, cross-linking in the presence of siRNA did not reveal distinct cross-link contacts indicating that RNA-binding has no major effect on p38 structure. However, the number of identified cross-links was higher in the presence of siRNA than in the absence.

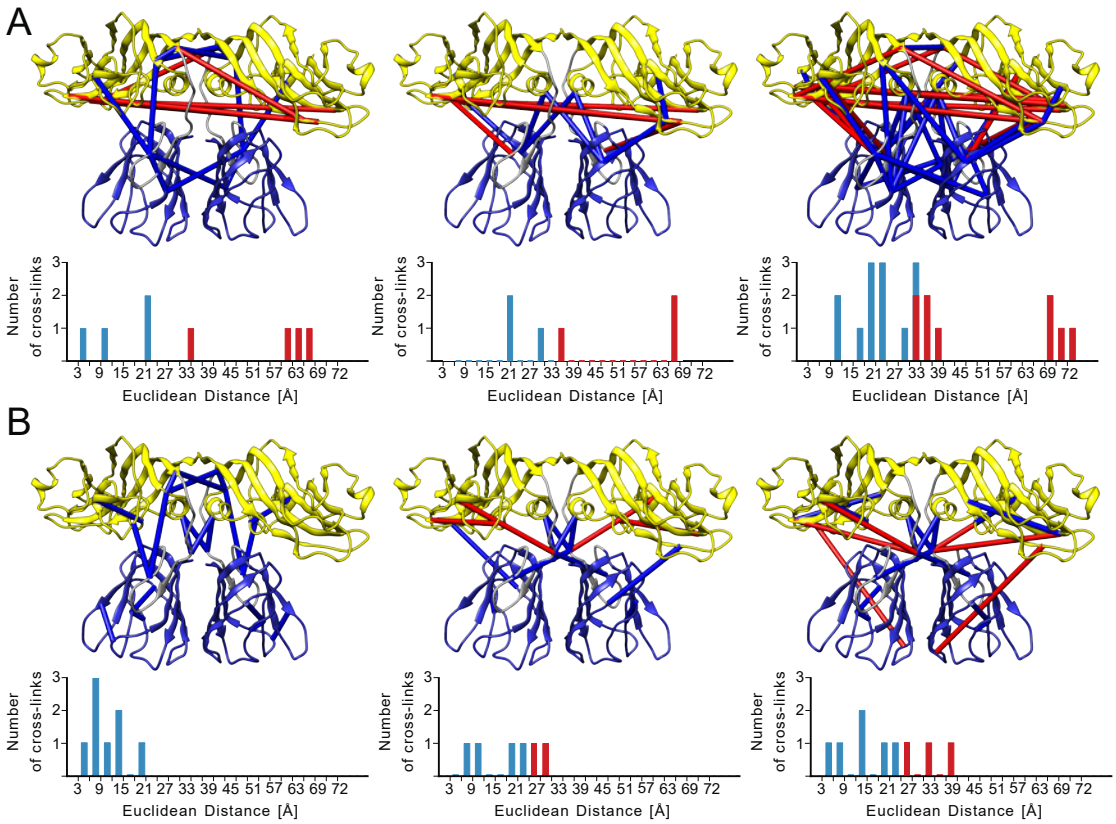


Figure 49: Mapping the Identified Cross-Linked Residues onto a Structural Model of the p38 dimer Derived from Crystallography. (A) Cross-links obtained with BS2G at 25 °C (left), at 5 °C (middle) and at 5 °C in the presence of 21 nts siRNA (right) were mapped on the dimer structure (blue and red lines). Contacts exceeding a distance of 31 Å are coloured red. Histograms of $C_\alpha - C_\alpha$ distances for each cross-link structure combination are shown below the structures. **(B)** Mapping of cross-links obtained with EDC at 25 °C (left), at 5 °C (middle) and at 5 °C in the presence of 21 nts siRNA (right). Contacts exceeding a distance of 25 Å are coloured red. The S- (yellow) and P-domains (blue) are colour-coded.

4.2.3 Cross-Linking-Guided Model Building

To interpret the obtained cross-links in a structural context, distances of the cross-linked residues calculated from a structural model were compared with the maximum distances the cross-linker molecules were expected to bridge. For this, a model of the p38 dimer derived from a structure of the viral capsid generated from X-ray crystallography (provided by Prof. Dr. Milton T. Stubbs) was used. Euclidean distances between the C_α atoms of cross-linked residues in the structural model were calculated and compared with a theoretical upper distance limit of each cross-linker. By comparison with the maximum distance of 35 Å previously suggested for BS3 (11.4 Å spacer arm) [232], a maximum distance of 31 Å for BS2G (7.7 Å spacer arm) was assumed in this thesis. For

EDC, a distance limit of 25 Å was used as described [233]. The difference between this limit and the length of the cross-linker spacer arm accounts for the side chain-C_α distance and for flexibility of the protein backbone.

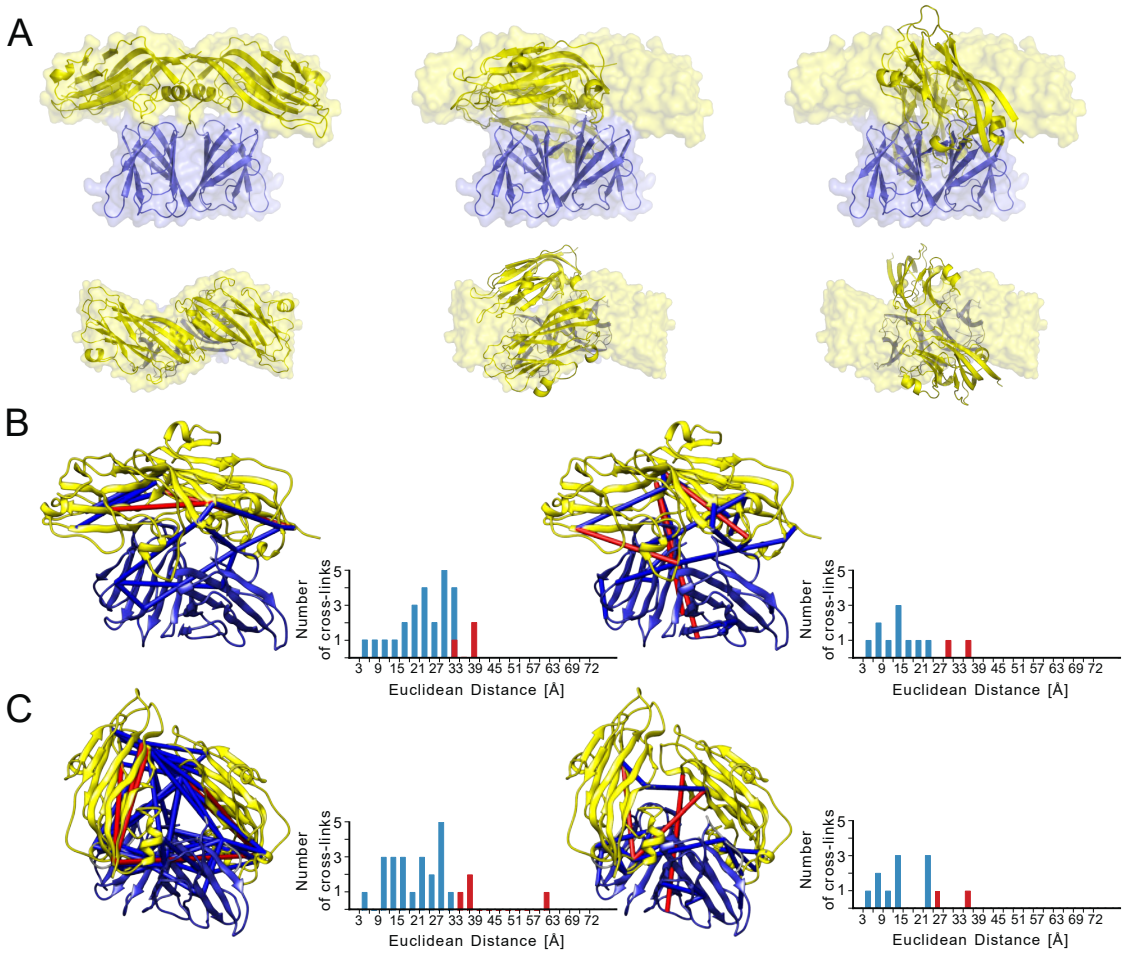


Figure 50: Docking of Separated S- and P-domains using BS2G Cross-Links as Restraints. **(A)** Front (top) and top views (bottom) on cartoon representations of the dimer model derived from crystallography (left) and docking models 1 (middle) and 2 (right) (cartoon representation). The dimer model from crystallography (surface representation) is shown for comparison. Models were aligned with respect to the P-P dimer. **(B)** Mapping of all cross-links identified with BS2G (left) and EDC (right) on model 1. **(C)** Mapping of cross-links on model 2. Note that EDC cross-links were not used to guide docking. Furthermore, note that the cross-links projected onto the model structure are visualised between all copies of the interacting residues and, therefore, the number of cross-links in the model can be larger than in the histogram. Contacts exceeding a distance of 25 Å or 31 Å for BS2G or EDC, respectively, are coloured in red. The S- (yellow) and P-domains (blue) are shown.

Comparison of Cross-Links with the Capsid Structure Mapping of the cross-links obtained with BS2G onto the dimer model showed that the cross-links were distributed across the surface of the model (**Figure 49A**). Similar regions of the model were cross-linked at 25 °C in the absence of RNA, at 5 °C in the absence of RNA and at 5 °C in the presence of 21 nts siRNA. Notably, most cross-links in the P-P dimer were in agreement with the model, indicating that the P-P-dimer likely forms the dimerisation interface in the capsid as well as in solution. Cross-links disagreeing with the structure were mostly located between the two S-domains. For all three conditions, at least one cross-link corresponding to a distance of >50 Å in the model indicated a more compact conformation of the S-domains. For example, the cross-link Lys94-Lys94 implied a short distance between S-domains of two copies p38 as it was derived from overlapping peptide sequences. Similarly, several additional cross-links between the S-domains violating the maximum distance cut-off originated from overlapping peptide sequences. Moreover, in the experiments performed at 5 °C, multiple cross-links between the S- and P-domains were incompatible with the model, suggesting an overall more compact conformation of p38.

Interestingly, mapping the cross-links obtained with EDC onto the structure yielded fewer cross-links exceeding the 25 Å cut-off distance (**Figure 49B**). Whereas the cross-links obtained at 25 °C were completely in agreement with the structural model, two cross-links exceeding the cut-off were identified for the experiments at 5 °C in the absence of RNA and three for the experiments at 5 °C in the presence of RNA. In the absence of RNA, one cross-link connecting the S-domain to Asp247, a residue located within the flexible loop region connecting the S- and P-domains, was identified, namely Lys205-Asp247. Similarly, in the presence of RNA, Asp247 was cross-linked to Lys102 and Lys205 suggesting a closer interaction between the S-domains and the central flexible loop possibly through slight rotation of the S-domains. In the absence of RNA, the second cross-link exceeding the distance cut-off connected Lys102 and Asp199. Whereas Lys102 is located within the β-sandwich fold of an S-domain, Asp199 is part of one of the flexible loops connecting the β-sheets in the S-domain. Therefore, violation of the distance cut-off by approximately 2 Å is most likely the result of structural dynamics of the flexible loop that are not mirrored by the static p38 dimer model. In the presence of RNA, the third cross-link disagreeing with the model was Lys95-Glu313 that connected the S- and P-domains over a distance of approximately 40 Å. Similar to the BS2G cross-links, this cross-link indicates that p38 exists either in a more compact conformation or that cross-links originate from multiple states (see below and discussion in **Section 5.2**).

Cross-Link-Guided Docking of Domains The observed discrepancy between the identified cross-links and the p38 dimer model can be rationalised in two ways: (1) p38 exhibits high conformational flexibility so that only a subset of cross-linking restraints is fulfilled by a single structure but all cross-links are explained by the ensemble of structures. (2) A single overall more compact p38 structure that is different from the capsid model and that can explain most of

Table 19: Solvent-Accessible Surface Distances and Euclidean Distances Between the S- and P-domains in the Docking Models. Distances from the C_α atom of Lys239 to the C_α atom of Q249 were calculated. Pairs with distances >100 Å were not computed.

Model	From Chain	To Chain	Euclidean Distance [Å]	SASD* [Å]
Model 1	A	A	39.1	79.5
	A	B	23.5	90.8
	B	A	28.8	55.2
	B	B	-	>100
Model 2	A	A	28.9	47.2
	A	B	32.7	58.1
	B	A	41.8	82.4
	B	B	31.4	78.5

* Solvent-accessible surface distance

the cross-links exists in solution. To gain an idea whether a single conformation is compatible with the identified cross-links, cross-links were used as guiding restraints for docking. For this, the S-domains were separated from the two P-domains of the p38 dimer model by removing the connecting loop regions. Next, the three separate structures (i.e. the two S-domains and the P-P dimer) were docked simultaneously using HADDOCK [198, 234] guided by all BS2G cross-links (**Figure 50**, see **Section 3.2**). The generated models were clustered by RMSD resulting in four models in two clusters and scored according to energetic terms. The top scoring model from each cluster was used for subsequent analysis (termed model 1 and model 2).

Comparison of models 1 and 2 with the dimer structure generated in reference to the viral capsid (**Figure 50A**) showed that, in both docking models, the S-domains were rotated by 90deg along the S-P axis. However, the positions of the S-domain in the two models differed by their tilt and relative rotation. Whereas the S-domains were positioned flat on the P-P dimer with a large contact area between the two S-domains in model 1 (**Figure 50A**, middle), they were tilted by approximately 45deg in model 2 and shared a smaller contact area (**Figure 50A**, right). Moreover, the S-domains in cluster 1 were oriented parallel, i.e. their termini faced in the same direction. In contrast, the S-domains in cluster 2 were both facing towards the P-domains. To evaluate the compatibility of the generated models with the lengths of the removed loop regions, solvent-accessible surface distances between Lys239 of the S-domain and Gln249 of the P-domain were calculated using the Xwalk software [206] (see **Table 19**). All possible combinations of the two S- and two P-domains show solvent-accessible distances >40 Å. Assuming an amino acid contour length of 4.02 Å [235],

the removed loop region $_{240}$NRTGSTSDA$_{248}$ can bridge a maximum distance of 36.18 Å at full chain extension. Therefore, the length of the connecting linker region is incompatible with the positioning the S-domains during docking. However, flexibility of amino acids adjacent to the loop possibly increases the maximum allowed distance between the S- and P-domains (see **Section 5.2** for discussion).

Next, agreement of the docked models with the observed cross-links was analysed by mapping the connected residues on the models and calculating Euclidean distances. Analysis of model 1 (**Figure 50B**) showed that most cross-links observed after cross-linking with BS2G were in agreement with the structure. However, three experimentally observed contacts between residues showed distances >30 Å. Two cross-links located within a single S-domain, namely Lys239-Lys94 and Lys239-Lys95, corresponding to distances of 36 Å and 39 Å, respectively, were identified. Lys239, Lys94 and Lys95 are located within the folded β-sandwich of the S-domain suggesting that the large distances in the model are not the result of protein flexibility but rather of interactions between two different S-domains. A third cross-link, namely Lys239-Lys239, connected the two S-domains at a distance of 32 Å, thereby exceeding the cut-off distance by only 1 Å and possibly within the range of distances explained by protein dynamics. Similarly, most of the cross-links obtained with EDC were in agreement with the model. The calculated distance of the cross-link Lys102-Asp199 within the S-domains exceeded the cut-off distance by 3.6 Å (28.6 Å vs. 25 Å cut-off distance) implying a different interaction interface between the S-domains. A second distance violation was identified for the cross-link Lys95-Glu313 connecting the S- and P-domains and corresponding to a distance of 35.2 Å in the model.

Mapping the same cross-links on model 2 similarly showed reasonable agreement between the cross-links and the structure (**Figure 50C**). In total four cross-link violated the maximum distance. Three of these, namely Lys239-Lys94, Lys239-Lys95 and Lys239-Lys102, were located in the S-domains and exceeded the maximum distance by <4 Å (calculated distances were 36.0 Å, 38.7 Å and 36.2 Å, respectively). Note that the cross-link Lys239-Lys102 corresponded to a distance of 29.7 Å in model 1. The fourth incompatible cross-link was Lys239-Lys239. The cross-link that closely missed the maximum distance in model 1, corresponded to a distance of 62.7 Å in model 2. EDC cross-links were also in agreement with model 2, only two cross-link distances exceeded their maximum distance. One cross-link, Lys102-Asp199, missed the cut-off by less than 1 Å (25.4 Å vs. 25 Å maximum distance). The second cross-link, Lys95-Glu313, that was also incompatible with model 1, corresponded to a distance of 35.5 Å in model 2. This cross-link alone suggests an inverse orientation of the P-P dimer, however, a similar cross-link Lys95-Asp307 connected the same lysine residue to the opposite face of the P-P-dimer. Therefore, it is unlikely that both cross-linking restraints can be fulfilled in a single structural model. As a control, the cross-links were mapped onto a model of the intact viral capsid (kindly provided by Prof. Dr. Milton T. Stubbs). As expected from the tight packing of the p38 monomers in the capsid, most BS2G cross-links were

in agreement with the model (**Figure S19A**), only three cross-link distances exceeded the distance cut-off slightly by less than 1 Å. Likewise, most EDC cross-links were in agreement with the model (**Figure S19B**), only the cross-link Lys95-Glu313 that proved incompatible with a single structure for the docking models, corresponded to a distance of 40 Å.

MD Simulation of the Structural Models To test the stability of the docking models in comparison with the dimer model derived from the p38 capsid structure, a short 300 ns MD simulation of all three models was performed. Analysis of the capsid dimer model showed that RMSDs of both monomers (calculated with respect to the initial model structure) remained below 5 Å even though the two domains are connected by a flexible linker (**Figure 51A**). In addition to RMSDs, minimum distances between C_α atoms of the S- and P-domains were extracted (**Figure 51B**). The two P-domains of the capsid model remained at a distance of approximately 5 Å, whereas the S-domains opened slightly by shifting from 6 Å to 8 Å distance during the simulation. This indicates that the P-domain interface present in the viral capsid structure possibly also exists in solution, at least for the first 300 ns investigated in the simulations. Models 1 and 2 exhibited more dynamics in the simulations. RMSDs of one monomer of model 1 remained at 4 Å, whereas the RMSD of the second monomer increased from 4 Å to 7 Å after 300 ns (**Figure 51C**). This increase in RMSD was accompanied by a slight tilt movement of the S-domains with respect to the P-P-dimer (compare insets of **Figures 51C and D**). Nevertheless, the distance between the S- and P-domains remained stable at approximately 5 Å over the course of the simulation (**Figure 51D**), indicating that p38 retained the compact initial conformation. In contrast, model 2 showed an immediate increase of RMSD of both p38 monomers to 6 Å - 10 Å accompanied by an outward movement of the S-domains (compare insets of **Figures 51E and F**). This separation of S-domains was quantified by the S-domain distance (**Figure 51F**) increasing to approximately 23 Å. In addition to minimum C_α atom distances, distances between centres of mass of the S- and P-domains were calculated (**Figure S20**). Similarly, these distances indicated a compact conformation for model 1, whereas the capsid model and model 2 exhibited short distances between the P-domains and a large separation between the S-domains.

The MD simulations suggest that the capsid model is stable for short periods of time in aqueous solution in the absence of the intact capsid. However, the large separation of the S-domain in this model does not agree with the identified cross-links. Models 1 and 2 were generated on the basis of these cross-links and, therefore, are overall more compact than the capsid model. Interestingly, the compact conformation of model 1 was retained in the 300 ns simulation, indicating that the generated model is stable at least for short time scales. In contrast, this model 2 proved unstable as it changed from an initial compact form to a more extended structure through separation of the S-domains.

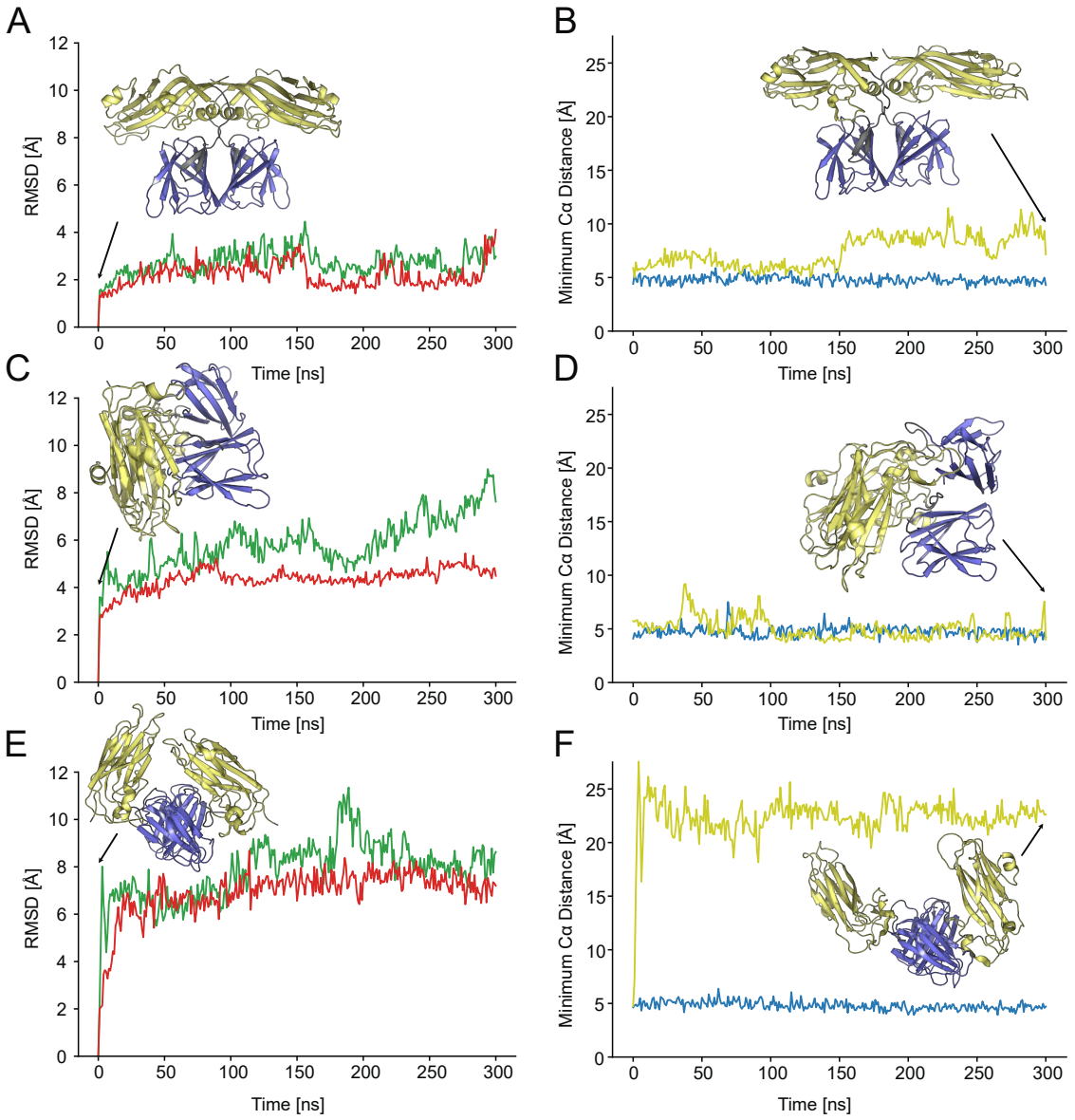
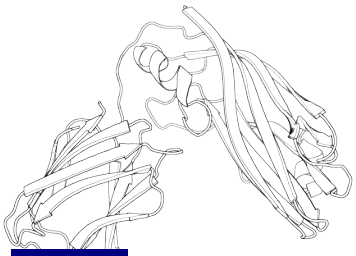


Figure 51: MD Simulations of Structural Models of p38. RMSDs of the two p38 monomers (red and green) and the initial structure (at t=0 ns) are shown (left). Minimum distances between C_α atoms of the S- (yellow, residues 81 - 239) and P-domains (blue, residues 261 - 351) along the trajectory and the structural model after 300 ns simulation time (right) are shown. **(A) and (B)** p38 model generated from the capsid X-ray structure. **(C) and (D)** Cluster 1 obtained from cross-linking guided docking. **(E) and (F)** Cluster 2 obtained from cross-linking guided docking.

In summary, native MS analyis of p38 showed that the protein exists as a dimer in solution that is stabilised after binding to 21 nts and 22 nts siRNA (**Figure 41**). Moreover, p38 dimers readily exchange p38 monomers in solution in the absence of RNA and subunit exchange is reduced after

binding to siRNA (**Figure 42**). p38 was cross-linked with BS2G and EDC at different temperatures and in the absence and presence of RNA. Most of the cross-links identified in this thesis are in agreement with (1) the structure of an intact capsid (**Figure S19**), however, native MS analysis and polyacrylamide gel electrophoresis of cross-linked p38 did not indicate that p38 formed capsids in solution. Cross-linking MS and cross-link guided model building indicate that most of the observed cross-links can be explained by (2) a compact dimer model similar to cluster 1. MD simulations presented in this thesis did not rule out the existence of such a compact dimer. Moreover, an ensemble of multiple, possibly highly dynamic p38 dimer structures (3) that each satisfy a subset of the identified cross-links is a likely explanation for the identified cross-links.



5 Discussion

5.1 Interaction of Synaptotagmin-1 with Lipids

The precise mechanism by which Syt-1 regulates SNARE-mediated membrane fusion is a matter of ongoing debate as highlighted by the variety of coexisting hypotheses on this matter (see **Section 1.2**). However, most of the models agree that Syt-1-lipid interactions are important in this process. For example, Syt-1-lipid contacts are involved in local remodelling the membrane [236] and during membrane-specific recruitment of the C2 domains [54]. In this thesis, lipid preferences of Syt-1 were analysed using various biophysical approaches including native MS and MD simulations.

5.1.1 Detergent-Mediated Lipid Transfer to Soluble Proteins

Native MS analysis of protein-lipid interactions has been extensively performed with membrane proteins purified in or transferred into MS-compatible detergents that can be removed by collisional activation in the gas phase. Lipids retained during protein purification and detergent exchange are observed as additional peaks during native MS analysis. The presence of detergents is not required for native MS analysis of soluble proteins or lipid vesicles such as liposomes *per se* [237]. However, lipid transfer from liposomes to soluble proteins was not observed in this publication. Therefore, the use of detergents to destabilise the lipid vesicle and facilitate lipid transfer to a soluble protein such as the soluble variants of Syt-1 was investigated. Effects of the detergents C8E4, LDAO and OG on the well-characterised soluble model proteins ADH, β-LG, BSA and myoglobin was analysed by native MS. The observed effects included the decrease or increase of the average charge states observed by MS (**Figure 12**), destabilisation of non-covalent interactions (**Figure 14**) and detergent-mediated lipid transfer (**Figure 15**).

LDAO Does not Affect Charge Distributions of Soluble Proteins The non-ionic detergents reduced (C8E4) or increased (OG) the number of charges acquired during native MS analysis independent of protein size and oligomeric state (**Figure 12**). However, the magnitude of charge-modulation was higher for larger proteins that acquired a higher number of charges during electrospray ionisation. In contrast, the zwitterionic detergent LDAO did not affect the charge

distributions of the standard proteins (see **Section 4.1.1**). Moreover, an additional broad charge state distribution likely corresponding to partial unfolding was observed during the analysis of BSA the presence of LDAO (**Figure 12B**).

The charge-reducing properties of LDAO were described previously for native MS experiments of membrane proteins [176, 238]. Therefore, charge reduction is possibly related to interactions with hydrophobic regions of these proteins. In line with this assumption, the additive trimethylamine-N-oxide (TMAO) that is structurally similar to LDAO (**Figure 52A**) reduced ion charges during native MS analysis of the soluble protein myoglobin [239]. LDAO differs from TMAO by a dodecyl chain instead of the methyl group present in TMAO supporting the hypothesis that hydrophobic interactions play for its charge-reducing effect on membrane proteins. Similar interactions of LDAO with hydrophobic surface regions of BSA leading to destabilisation in solution possibly explain the unfolding of the protein in the presence of LDAO. Furthermore, the intensity of peaks corresponding to heme-bound myoglobin was reduced in the presence of LDAO (**Figure 14B**), similarly indicating destabilisation of the holo-protein. In contrast, LDAO did not reduce stability of the tetrameric model protein ADH (**Figure 13B**), suggesting that destabilisation is likely limited to proteins with accessible hydrophobic surface areas.

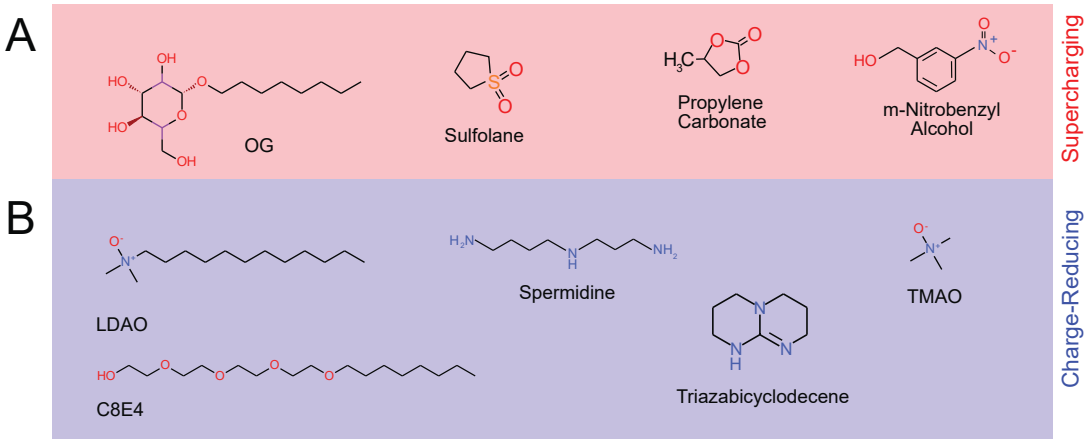


Figure 52: Structural Formulae of Charge-Modulating Reagents. Structure formulae of **(A)** supercharging reagents and **(B)** charge-reducing reagents.

Protein Supercharging The term supercharging describes the increase of protein ion charge states in the presence of supercharging reagents. Similarly, soluble proteins analysed in this thesis in the presence of the detergent OG showed an increase in the number of charges (**Figure 12B**). Analysis of the molecular basis of supercharging is challenging as the effect occurs in the transient nano-metre sized droplets of the final stages of electrospray ionisation. Nevertheless, three models have been proposed to explain its molecular basis.

The surface tension model argues that supercharging agents such as sulfolane, propylene carbonate or m-nitrobenzyl alcohol (**Figure 52A**) increase droplet surface tension during the final stages of ionisation, thereby increasing the final droplet size [240]. According to the charged residue model of electrospray ionisation [241], the charges of the final electrospray droplet are completely transferred onto the protein. Therefore, increasing the size of the final droplet directly increases the number of charges on the protein ion. However, OG was shown experimentally to reduce surface tension in aqueous solution [242]. Moreover, several other known supercharging reagents have been shown to lower surface tension suggesting that supercharging is largely independent of surface tension [243]. The unfolding model is based on the assumption that supercharging reagents lead to unfolding of the protein thereby increasing the surface area that can accomodate charges [244]. However, solution-phase unfolding is characterised by broad charge-state distributions in native MS in contrast to the narrow native-like charge distribution observed in this thesis for proteins in the presence of OG. This suggests that unfolding is neither the cause for the observed charge increase of standard proteins. The charge trapping model [245, 221] proposes that supercharging reagents accumulate in the outer shell of the electrospray droplet and prevent loss of charge carriers such as protons, Na^+ or NH_4^+ by restricting their access to the surface of the droplet. Due to the low volatility of the supercharging reagents, this charge-accumulating effect increases during shrinkage of the droplet until the solvent-free protein ion is eventually released into the gas phase. In this thesis, the presence of OG reduced the activation voltage required for unfolding of ADH (**Figure 13**) in line with a high charge-density destabilising the protein complex. However, the drift times recorded for the 26+ charge state of ADH in the presence and absence of OG are similar (**Figure S1**), indicating that ADH adopts a native conformation despite supercharging. In summary, supercharging mediated by OG as observed in this thesis is best explained by the charge trapping model.

Protein Charge Reduction In contrast to the effect of OG, native MS analysis in the presence of C8E4 reduced the number of charges of all protein ions investigated (**Figure 12B**). Charge-reducing properties during native MS have been described previously for various additives and two possible mechanisms have been proposed.

The concept of activation-dependent charge reduction was suggested based on the charge-reducing properties of TMAO (**Figure 52B**) [239]. According to this mechanism, charge is reduced by gas phase collisions of TMAO with the charged protein ion after complete desolvation. However, collision voltage-dependent charge reduction was not observed for ADH in the presence of C8E4 indicating that activation-dependent charge reduction is not the underlying mechanism [219]. Nevertheless, voltage-dependent charge reduction prior to the collision cell cannot be ruled out. The second mechanism, charge reduction by proton capture, was described for charge-reducing agents with a high gas phase basicity such as triazabicyclodecene and polyamines such as spermidine (**Figure 52B**) [246, 247]. These additives dissociate from the protein after capturing a proton either

during the final stages of electrospray ionisation or after collisional activation of the desolvated protein-additive complex. Similarly, a model for C8E4 acting as a charge-carrier and removing protons from the electrospray droplet by evaporation of its protonated form was proposed [176]. The same study reported that the charge-reducing effect of the tetraethylene glycol head group of C8E4 is lower than that of the intact detergent [176]. This implies that the hydrophobic octyl chain of C8E4 is likely involved in charge reduction, possibly by mediating evaporation through contact with the surrounding vacuum at the electrospray droplet surface or by direct interaction with the protein. The latter is supported by the observation in this thesis that lipid transfer to soluble β-LG was most pronounced in the presence of C8E4 (**Figure 15**). In summary, charge reduction of soluble proteins in the presence of C8E4 most likely occurs through proton capture by the detergent. However, proton capture does not explain the experimentally observed lipid transfer to soluble proteins in the presence of C8E4 (see below).

Structural Stabilisation in the Presence of Detergents Reduced stability of ADH in the presence of OG (**Figure 13**) was observed in this thesis. However, C8E4 and OG also showed a stabilising effect by preserving holo-myoglobin (**Figure 14**). In light of the previously discussed mechanisms for charge-modulation, destabilisation of ADH by OG is likely the consequence of increased charge-density and the resulting Coulombic repulsion. Preservation of the myoglobin-heme complex seems contradictory to the destabilising effect of OG. However, preservation of holo-myoglobin by OG might be explained by the charge trapping model. Accordingly, OG prevents ejection of charge-carriers such as protons from the electrospray droplet by accumulating in the outer shell of the droplet. Possibly, not only charge carriers but also the heme cofactor was maintained in the droplet and eventually released into the gas phase together with myoglobin.

C8E4 showed a stabilising effect on holo-myoglobin even though only minimal charge reduction was observed (the main charge state of myoglobin was 7+ in the presence and absence of C8E4, see **Figure 14**). Therefore, the underlying principle of myoglobin stabilisation through C8E4 is not based on the charge-reducing properties of the detergent. Detergents were previously suggested lowering the impact of collisional activation on membrane proteins by capturing activation energy [248]. Possibly, this energy-reducing effect also plays a role for soluble proteins in the presence of detergents by reducing activation of holo-myoglobin and consequently preventing ejection of the heme-group. However, it remains to be elucidated why preservation of holo-myoglobin was observed in the presence of C8E4 but not in the presence of LDAO even though both detergents should capture activation energy. Likely, the destabilising effects described above for LDAO counteract its energy-reducing effect and thus induce ejection of the heme group from holo-myoglobin.

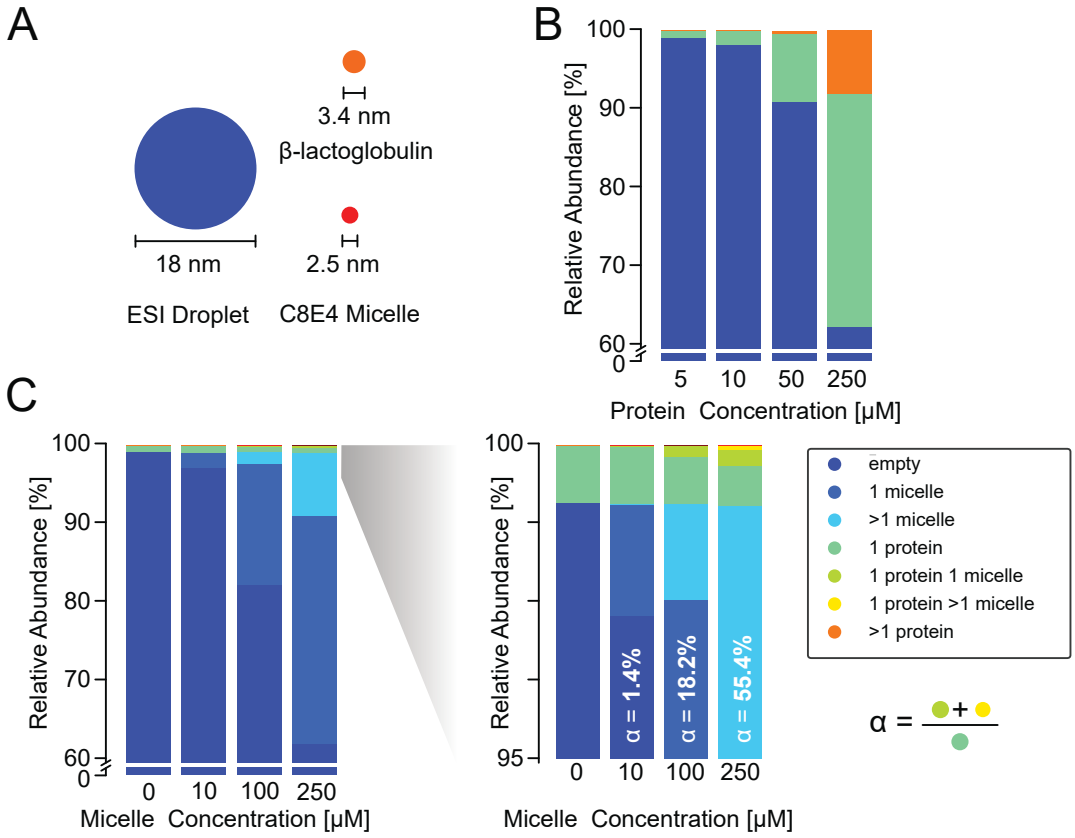


Figure 53: Nonspecific Aggregation during Electrospray Ionisation. **(A)** Schematic representation of the sizes of the final electrospray droplet [249], the C8E4 micelle [250] and β-LG (PDB 3NPO). **(B)** Stochastic model of protein aggregation during electrospray ionisation similar to Benesch et al. [173]. Defined concentrations of protein were randomly placed in a box and numbers of proteins were counted from 100,000 randomly sampled volumes equalling the electrospray droplet volume. **(C)** Same as (B) for a model including 5 µM protein and varying amounts of detergent micelles. The ratio α between final droplets containing a protein and at least one detergent micelle to droplets containing a single protein is shown.

Lipid Transfer to β-Lactoglobulin Using β-LG as model protein for soluble lipid-binding proteins, lipid transfer from detergent-lipid micelles to a soluble protein was interrogated (**Figure 15**). Comparing C8E4, LDAO and OG, protein-lipid complexes were observed exclusively in the presence of C8E4. This was unexpected as solubilisation of lipids in the different detergents was considered to be similar and, therefore, might provide insights into the underlying mechanism of detergent-mediated lipid transfer. No protein-lipid complexes were formed in the presence of OG. Supercharging presumably prevented formation of protein-lipid complexes under these conditions. However, protein-lipid complexes were only observed for β-LG analysed in the presence of C8E4 despite β-LG analysed in the presence of C8E4 and OG showed several common charge states. This

indicates that lipid transfer is likely a specific property of C8E4 possibly independent of charge manipulation by the detergent. A mechanism for lipid transfer to soluble and membrane proteins was suggested previously [177]. Accordingly, mixed lipid detergent micelles are formed and interact in the electrospray droplets with the protein. The hypothesis further proposes that protein-lipid complexes of soluble proteins are primarily caused by lipid detergent micelles present in the final electrospray droplets leading to the formation of non-specific aggregates simply due to stochastic combination of multiple small proteins and detergent micelles in the much larger ESI droplet (**Figure 53A**). This is based on the fact that analyte concentrations >50 µM result in a significant amount of final electrospray droplets containing more than one analyte ion [173] (see **Figure 53B**). In this thesis, detergent micelle concentrations of 15 µM (LDAO) and 100 µM (C8E4) were employed and protein-lipid peaks were only observed in the presence of C8E4 in agreement with the hypothesis. Note that OG was used at 1 × cmc as higher concentrations of the detergent interfered with native MS analysis. Therefore, no free OG micelles were expected. Moreover, assuming uniform distribution of lipids in the detergent micelles, on average 1 - 5 lipids are expected in every C8E4 micelle and 7 - 38 lipid molecules in every LDAO micelle at the lipid concentrations employed in this thesis (see **Figure 15C**). To explore this model quantitatively, a random sampling approach similar to the approach suggesting formation of artificial oligomers at high protein concentrations [173] was used (see **Section 3.2.6** for details). However, two different analyte types, namely protein and micelles, were modelled (**Figure 53C**). Using a protein concentration of 5 µM and detergent concentrations of 0 µM - 250 µM the ratio α between the final electrospray droplets containing one protein and at least one detergent micelle and droplets containing only one protein and no detergent was calculated. At 100 µM detergent micelle concentration, as employed for the analysis of lipid transfer by C8E4, a ratio of $\alpha = 0.18$ was calculated, i.e. for every five droplets containing a protein without lipid, one droplet containing a protein and at least one detergent micelle is expected. If the reason for the observed protein-lipid transfer is non-specific association of proteins with lipids present in the final electrospray droplets, the peak-intensity ratio of the lipid-free protein peak and the peaks corresponding to binding of at least one lipid should be <20 %. However, peak intensity ratios observed in this thesis were higher for all observed charge states (**Figure 15**). Moreover, comparison of different lipid classes binding to soluble variants of Syt-1 revealed distinct binding curves for the different lipid classes (**Figures 27, 28 and 29**). This indicates that the formation of protein-lipid complexes in the presence of C8E4 is not only a result of stochastic association of protein and detergent micelles. On the contrary, interactions between the protein and the lipid, for example through electrostatic effects, are likely involved in the observed protein-lipid contacts. Detergents used for native MS analysis are characterised by weak interactions with the protein, as observed e.g. during detergent removal by collisional activation in the gas phase. The presence of protein-lipid complexes thereby suggests that protein-lipid interactions are stronger than protein-detergent interactions as protein-lipid but no protein-detergent complexes were observed in this thesis. A likely mechanism for those protein-lipid interactions are lipid head group interactions.

In summary, the experiments on charging effects of the detergents C8E4, LDAO and OG presented in this thesis show that the presence of LDAO does not affect protein charge, whereas OG leads to native supercharging, i.e. the observation of higher average charge states with a native-like narrow charge state distribution. This is most likely caused by OG preventing the evaporation of charge carriers in the electrospray droplet. C8E4 showed charge-reducing and lipid transferring properties during native MS experiments. Charge reduction is most likely mediated by protonation of C8E4 and ejection of the charged detergent from the electrospray droplet. A mechanism for C8E4-mediated lipid transfer to β-LG remains elusive, however, it likely involves specific protein-lipid interactions possibly between the lipid head-groups and the protein.

5.1.2 Characterisation of Synaptotagmin-1 in the Absence of Lipids

To experimentally interrogate the lipid-preferences of Syt-1, the full-length protein was recombinantly expressed in *E. coli* and purified in CHAPS detergent (**Figure 18**). Indeed, native MS analysis of Syt-1 monomers and dimers (**Figure 21**) indicated binding of endogenous lipids and detergent molecules to the protein. This is in agreement with previous research showing that native MS captures protein-lipid interactions of membrane proteins purified in the presence of detergents [126]. However, analysis of full-length Syt-1 complicates the interrogation of protein-lipid interactions by two factors: (1) It is not possible to distinguish between lipid binding to the transmembrane helix and to the soluble C2 domains and (2) CHAPS detergent is incompatible with native MS analysis so that additional steps to remove the detergent are required prior to native MS analysis. Therefore, analysis of protein-lipid interactions was performed with soluble variants encoding the C2A domain, the C2B domain or both.

Proteins were expressed, purified and correct folding of the soluble variants was assessed using CD spectroscopy (**Figure 19**). Oligomerisation of the variants was analysed by native MS revealing low amounts of Syt-1 dimers for variants including both C2 domains (**Figure 20**). The formation of oligomers of Syt-1 was previously reported in the presence of polyphosphates and Mg^{2+} [79, 251]. It was suggested that multiple copies of Syt-1 interact via C2B-C2B contacts thereby forming a circular structure that is involved in the membrane fusion process. In contrast, native MS analysis in this thesis showed that Syt-1 dimers were only observed for the tandem-C2 variants containing both C2 domains (**Figure 20**). Native MS of the individual C2A or C2B domains showed that the isolated C2 domains exist exclusively as monomers. This suggests that the low degree of dimerisation identified in this thesis occurs through interactions between C2A and C2B. In more detail, these interactions might involve the two C2 domains themselves but possibly also include interactions with the flexible linker regions (i.e. the intrinsically disorderered linker in the C2A variant or the

C-terminal region in the C2B variant). Even though the presence of polyphosphates and Mg^{2+} might shift the interaction to C2B-C2B contacts as previously reported, it seems likely that Syt-1 oligomerisation in the absence of polyphosphates and Mg^{2+} occurs via C2A-C2B contacts.

Native MS of soluble Syt-1 variants was performed at a protein concentration of <20 µM to minimise formation of artificial oligomer peaks through the presence of multiple copies of Syt-1 in the same final electrospray droplet [173]. In contrast, quantitative immunoblotting of purified synaptosomes revealed a physiological Syt-1 concentration of 70 µM [252] and quantification of the protein content of SVs showed that approximately 15 copies of Syt-1 are present in each SV [253]. Therefore, oligomerisation of Syt-1 at concentrations >20 µM *in vivo* cannot be excluded. Even though Syt-1 is one of the most abundant proteins in SVs, it accounts for only 7 % of total SV proteins [254]. Therefore, the effective concentration of Syt-1 oligomers is presumably low *in vivo* due to molecular crowding on the SV. In line with this, a previous study investigating SV proteins by chemical cross-linking MS [152] reported several cross-links involving Syt-1. However, no cross-links with overlapping peptide sequences were identified that unambiguously show the formation of Syt-1 homo-oligomers. This highlights that Syt-1 oligomers are likely of low abundance *in vivo* similar to what was observed by native MS in this thesis.

In summary, recombinantly expressed Syt-1 purified in the presence of detergent retained endogenous lipids or detergent molecules. To focus on the lipid interactions of the soluble C2 domains, soluble variants of Syt-1 were expressed and purified. A low degree of dimerisation of Syt-1 observed by native MS of the full-length protein and the soluble tandem-C2 variants likely occurs through C2A/C2B-interactions as the individual C2 domains were exclusively monomeric. Overall, comparison of native MS results presented in this thesis and previous research indicates that Syt-1 most likely exists as a monomer.

5.1.3 Analysis of Synaptotagmin-1 Lipid-Interactions *in vitro*

Having confirmed the correct expression of soluble Syt-1 variants, their binding affinities to lipids and lipid mixtures of increasing complexity were characterised using various biophysical techniques. The lipid overlay assay was used to screen multiple lipid head groups simultaneously. Native MS of lipid head groups and intact lipids was used to quantify lipid affinities followed by the liposome flotation assay to probe interactions of Syt-1 with intact lipid bilayers.

Interactions with Immobilised Lipids The protein lipid overlay assay detects interactions of a soluble protein with immobilised lipids using antibody staining. Application to Syt-1 C2AB and C2B (**Figure 22**) revealed preferential binding to the negatively charged PIPs PI(4)P, PI(4,5)P_2 and PI(3,4,5)P_3. PI(4)P was reported as part of the SV membrane in which Syt-1 is anchored,

$PI(4,5)P_2$ is located in the plasma membrane and $PI(3,4,5)P_3$ plays a role in clathrin-mediated endocytosis of SVs [255]. Accordingly, interactions with PI(4)P and $PI(4,5)P_2$ are in agreement with the physiological location of Syt-1, whereas interactions with $PI(3,4,5)P_3$ are so far not associated with the function of Syt-1. The lipid overlay assay also indicated interactions with other lipids such as phosphatidic acid, 3-sulfogalactosylceramide and cardiolipin. Similar to PIPs these lipids carry negative charges at physiological pH. Phosphatidic acid carries two negative charges and is present in the plasma membrane as a product of phospholipase C-mediated hydrolysis of $PI(4,5)P_2$. However, it is readily transferred to the endoplasmic reticulum membrane [256]. 3-sulfogalactosylceramide is primarily located in the extracellular leaflet of the neuronal plasma membrane [257] and is, therefore, inaccessible for intracellular Syt-1. However, it structurally resembles PIPs and carries a single negative charge at physiological pH. Cardiolipin carries two negative charges above pH 7.5 and is located exclusively in the inner mitochondrial membrane and, similarly, is likely not a biological target of Syt-1. These interactions indicate that the main factor determining Syt-1 binding to immobilised lipids is the lipid head group charge. Interestingly, Syt-1 binding to phosphatidyl glycerol or PS was not observed even though these lipids are negatively charged. Whereas negative charges on the inositol head group of the PIPs, the phosphate group of phosphatidic acid, the sulfate group of 3-sulfogalactosylceramide and the two phosphate groups of cardiolipin are all accessible, negative charges in PG and PS are protected by the esterified glycerol and serine moieties, respectively. Therefore, steric hindrance by the glycerol and serine groups alone or in combination with the restricted mobility of the lipid due to immobilisation possibly prevented binding of C2 domains to PS and PG in the lipid overlay assay. However, Ca^{2+} was suggested to form a complex with the CBLs of Syt-1 that eventually interacts with the carboxyl group of PS [228]. Accordingly, increased binding of Syt-1 to PS was expected in the presence of Ca^{2+}, however, an interaction with PS was not observed. This might be related to reduced lipid mobility after immobilisation. Stable interactions of Syt-1 with PS possibly require partial insertion of the CBLs of Syt-1 into the membrane as indicated by the presence of hydrophobic residues on the tips of the CBLs [48]. Such an interaction is likely disfavoured if lipid mobility is reduced through immobilisation. In line with this hypothesis, a previously published study of C2A by the lipid overlay assay does not indicate binding to PS in the presence of Ca^{2+} either [52]. Interestingly, the study showed a preference of C2AB for singly phosphorylated PIPs [52], whereas the lipid overlay assays presented in this thesis suggest a preference of Syt-1 C2AB for multiply phosphorylated PIPs. The latter is in agreement with Syt-1 preferentially binding to $PI(4,5)P_2$ [54].

Moreover, selectivity between the different negatively charged lipids might not be required for Syt-1 function as for example the PIPs are strictly separated to different membrane compartments [255]. Instead, Syt-1 might be sensitive to the local negative charge density. In line with this idea, local $PI(4,5)P_2$ concentrations of up to 82 % were reported in nanometre-sized micro domains at the inner leaflet of the plasma membrane [258]. Therefore, selectivity for negatively charged lipids as described in this thesis might be sufficient to target Syt-1 to the plasma membrane.

Previous research suggested that Syt-1 interacts with PI(4,5)P_2 primarily through the polybasic stretch of C2B [54]. Comparison between the lipid overlay assays of C2AB and C2B (**Figure 22A and C**), however, indicates that presence of both C2-domains in C2AB increases the interaction with the immobilised lipids compared to C2B alone, indicating that not only the C2B domain mediates binding to PI(4,5)P_2. The higher lipid affinity of C2AB possibly originates from the C2A domain increasing the residency time of C2B in proximity to the lipids through additional charge-mediated interactions. Analysis of C2AB and C2B in the presence of 500 µM Ca^{2+} (**Figure 22B and D**) showed binding to similar lipids, however, with lower affinity compared to the experiment in the absence of Ca^{2+}. In summary, the lipid overlay assay suggests preferential binding of Syt-1 C2AB to negatively charged lipids with accessible charged residues in the presence of Ca^{2+}.

Interactions with Lipid Head Groups In the lipid overlay assay immobilised lipids were employed to interrogate protein interactions with the lipid head groups. Native MS is a technique that is increasingly employed to characterise protein complexes including interactions with smaller molecules such as lipids [218, 126]. Therefore, binding of Syt-1 to lipid head groups was assessed using native MS. As non-covalent complexes are preserved during electrospray ionisation using this technique, protein-ligand complexes can be deduced from adduct peaks corresponding to the mass of the protein-lipid complex. Using the Syt-1 C2AB variant that includes both C2 domains, affinity towards the PIP head groups I(1,4,5)P_3 and I(1,3,5)P_3 was analysed (**Figure 24**). The intensity of Syt-1 in complex with lipid head groups increased with the head group concentration. To quantify this observation, mass spectra were deconvolved, peak areas were extracted and a ligand binding model was fitted (**Figure 23**). Manual inspection of native mass spectra and calculation of $K_{D\,app}$ values revealed a higher affinity of Syt-1 to I(1,3,5)P_3 than to I(1,4,5)P_3. This is surprising as I(1,4,5)P_3 is the head group of PI(4,5)P_2, a known interaction partner of Syt-1 [54], whereas I(1,3,5)P_3 is the head group of phosphatidylinositol 3,5-bisphosphate, a lipid involved in vesicle budding from late endosomes [259] that so far has not been associated with Syt-1. However, a similar preference for immobilised PI(3,5)P_2 over PI(4,5)P_2 was previously reported for intact and truncated Syt-1 [52]. The lipid overlay assay showed that lipid head group affinity of Syt-1 is correlated with the number of negative charges and their accessibility for Syt-1. Phosphate groups in I(1,3,5)P_3 are evenly distributed along the inositol ring, thereby maximising the negatively charged area Syt-1 can interact with. I(1,4,5)P_3 in contrast has two phosphate groups esterified to adjacent carbon atoms in the ring, resulting in an asymmetric charge distribution. Possibly, the larger negatively charged interaction area of I(1,3,5)P_3 is the basis for the increased affinity of Syt-1 for this lipid head group observed in this thesis and in the literature.

The $K_{D\,app}$ values calculated in this thesis for lipid head groups (**Figure 23**) and for intact lipids (see below) indicate lower affinity towards PIPs than K_D values previously determined for the interaction of Syt-1 with PI(4,5)P_2 in solution [54]. This discrepancy likely arises from at least two

differences between the previous analysis and this thesis. First, native MS experiments systematically report $K_{D\,app}$ values higher than K_D values from experiments performed in solution as gas phase transfer of the protein-ligand complex and subsequent collisional activation of the protein-ligand complex by atmospheric gas molecules leads to the loss of bound ligands [230]. This is of special concern for hydrophobic interactions after transfer into the hydrophobic gas phase [260, 261]. Nevertheless, affinities of hydrophobic protein-ligand interactions such as binding of ubiquitin to ubiquitin-binding domains determined using native MS were in agreement with affinities derived from other biophysical techniques [262]. Moreover, the use of charge-reducing agents such as imidazole or C8E4 in this thesis was suggested to reduce collisional activation of protein-ligand complexes, thereby increasing the reliability of the resulting $K_{D\,app}$ values [263]. Second, native MS reports on the interaction of few (up to two in this experiment) lipid head groups with Syt-1, whereas reported K_D values were derived from interactions of Syt-1 with intact liposomes containing up to 5 % $PI(4,5)P_2$ [54]. Therefore, the low affinity of Syt-1 towards a single lipid head group identified in this study might result in a high avidity (i.e. the combinatorial effect of multiple binding interactions) of the protein interacting with multiple lipids simultaneously. In summary, Syt-1 showed distinct affinities towards different PIP head groups, possibly related to the negative charge density on these head groups. The determined $K_{D\,app}$ values were lower than in experiments involving intact liposomes, likely a result from loss of ligands during native MS ionisation, increased avidity by Syt-1 interacting with multiple $PI(4,5)P_2$s in liposomes or a combination of both.

Interactions with Intact Lipids As discussed above, systematic errors in the determination of protein-ligand affinities as a result of collision-induced loss of bound ligands are of concern during native MS experiments. To investigate this effect in more detail, two native MS instruments with different degrees of in-source activation, namely the Waters Synapt G1 HDMS and the Micromass Q-TOF Ultima, were compared. For this, detergent-mediated lipid transfer was used to investigate binding of intact DOPS to Syt-1 (**Figure 25**) in the presence of the detergent C8E4 that showed the most effective lipid transfer to the standard protein β-LG (**Figure 15**). During native MS analysis with both instruments, binding of up to four DOPS molecules to Syt-1 was observed at 200 µM DOPS concentration. Moreover, deconvolution, extraction of peak areas and fitting showed slight differences in the obtained $K_{D\,app}$ values (**Figure 26**). The value obtained for mass spectra recorded on the less-activating Q-TOF Ultima (200 ± 25 µM) was lower than the $K_{D\,app}$ value calculated for spectra recorded on the Waters Synapt G1 (266 ± 25 µM). This shows that collisional activation in the gas phase indeed reduces the apparent lipid affinity. However, the observed difference in $K_{D\,app}$ values between the two instruments is small compared to the absolute $K_{D\,app}$ values indicating that loss of bound lipids due to gas phase effects contributes only moderately to the observed $K_{D\,app}$ value. Moreover, in comparative studies involving multiple lipids binding to the same protein and analysed with the same instrument, collisional activation is similar for all protein-lipid complexes and, therefore, reliable determination of lipid preferences can be achieved.

Comparative Analysis of Synaptotagmin-1 Lipid Binding Relative affinites of Syt-1 for lipids of seven lipid classes were analysed by native MS (**Figure 27**). Syt-1 did not interact with the zwitterionic lipid DOPC, showed intermediate affinity for the lipids DOPG, DOPI and DOPS, and high affinities for multiply phosphorylated PIPs. This affinity series corresponds to the charge of the lipids at neutral pH. DOPC is net neutral, DOPG, DOPI and DOPS carry a single negative charge, and PI(4)P, PI(4,5)P_2 and DOPI(3,4,5)P_3 carry two, three, and four negative charges, suggesting that lipid selectivity of Syt-1 in the absence of Ca^{2+} is mainly driven by electrostatic interactions. This is in line with the lipid overlay assays of Syt-1 binding to lipids described above. However, comparison of the absolute $K_{D\,app}$ value of intact PI(4,5)P_2 (54 µM) with that of its head group I(1,4,5)P_3 (170 µM, **Figure 24**) shows higher affinity towards the intact lipid, possibly indicating additional contributions from the fatty acyl chain. Contributions of C8E4 during native MS analysis to the preference for intact PI(4,5)P_2, however, cannot be ruled out. In summary, the lipid preference of Syt-1 C2AB is proportional to the amount of negative charge on the lipid head groups.

The lipid affinity of Syt-1 C2AB is either based on a high lipid affinity of a single C2 domain or on additive lipid affinities of both C2 domains. To investigate the lipid preferences of C2A and C2B separately, short soluble Syt-1 variants consisting of either C2A or C2B were analysed using detergent-mediated lipid transfer and native MS (**Figure 28 and 29**). Both C2 domains showed a preference for DOPS, a lipid for which the tandem-C2 variant showed medium affinity. Moreover, a correlation between net charge and $K_{D\,app}$ value was not observed for the C2 domains even though DOPC, that is net uncharged, was among the weakest binders for both domains. Remarkably, C2A and C2B showed moderate affinity towards PI(4)P and PI(4,5)P_2, even though PI(4,5)P_2 is the known target of Syt-1 and showed strong binding to C2AB. The observed differences between the analysis of Syt-1 C2AB and the separate C2 domains may be explained by two effects. First, the increased apparent lipid affinity of Syt-1 C2AB is based on additive affinities of both C2 domains. PI(4)P and PI(4,5)P_2 binding to the isolated C2 domains showed a $K_{D\,app}$ value similar to DOPG. However, the $K_{D\,app}$s value of PI(4)P and PI(4,5)P_2 binding to Syt-1 C2AB were notably lower than the $K_{D\,app}$ value of DOPG binding to C2AB. Moreover, PI(4)P and PI(4,5)P_2 are the only lipids without a defined fatty acyl distribution employed in this experiment. Whereas all other lipids contain two oleyl (18:1) chains, PI(4)P and PI(4,5)P_2 were purchased purified from brain tissue containing a naturally occurring distribution of acyl chains. The most prevalent fatty acyl chain types in these lipids are stearyl (18:0) and highly unsaturated arachidonyl (20:4) chains[1]. Interactions of both C2 domains of C2AB with PI(4)P or PI(4,5)P_2, possibly involving their fatty acyl chains, would explain the higher affinity of C2AB for these lipids. These interactions either occur simultaneously or in quick succession as the distance between the C2 domains is restricted through their connecting flexible linker. A second possible explanation for the differences in $K_{D\,app}$ values between C2AB and the isolated C2 domains is that analysis conditions for the isolated C2-domains were less

[1] Fatty acyl chain composition obtained from distributor.

activating than for C2AB. Lipid binding to the isolated C2 domains was analysed with lowered instrument sample and extraction cone voltages compared to the analysis of Syt-1 C2AB to reduce in-source activation of the protein-lipid complexes. However, as demonstrated by the comparison of two native MS instruments with differing in-source activation (**Figure 25**), reduced activation of the complex leads to higher $K_{D\ app}$ values. A combination of the additive affinities of the C2 domains during Syt-1 C2AB binding to PIPs and increased $K_{D\ app}$ values after analysis at less activating ionisation conditions are likely the reason for the different lipid preferences of the tandem-C2 variant and the isolated C2-domains.

In summary, comparative analysis of multiple lipids binding to Syt-1 revealed that net charge is a major determinant of the lipid affinity of Syt-1 C2AB. Analysis of the separate C2 domains showed increased overall affinities, most likely due to different instrument conditions, and reduced affinities towards PI(4)P and PI(4,5)P_2, probably due to joint interactions of the C2 domains in the tandem-C2 variant with these lipids or their fatty acyl chains.

Interaction of Synaptotagmin-1 C2 Domains with Intact Lipid Bilayers Interactions of the individual C2-domains with liposomes of defined composition mimicking the SV or plasma membranes were analysed (**Figure 30**). In contrast to binding to isolated lipids investigated before using native MS, liposomes enable simultaneous interactions of Syt-1 with multiple lipids. Density-gradient centrifugation was used to discriminate between liposome-bound and free Syt-1. In the absence of Ca^{2+}, binding to liposomes was not identified for C2A and a low degree of binding to liposomes mimicking the plasma membrane was observed for C2B. In contrast, in the presence of Ca^{2+}, C2A predominantly interacted with the SV membrane, whereas C2B bound to the plasma membrane in line with previous research suggesting that Ca^{2+} is an affinity switch for Syt-1 [54]. Moreover, the experiment supports previous research proposing that Syt-1 C2 domains have distinct affinities for the SV and plasma membranes [72]. Furthermore, the flotation assay verified the correct, functional expression of the soluble C2 domains analysed in this thesis.

Binding of C2B to liposomes mimicking the plasma membrane was observed in the absence of Ca^{2+} (**Figure 30**). However, C2B did not bind to liposomes mimicking the SV membrane. The main difference between the two liposome compositions was the presence of PI(4,5)P_2 and diacyl glycerol in liposomes mimicking the plasma membrane. This suggests that diacyl glycerol or, most likely, PI(4,5)P_2 is responsible for binding of C2B. Interestingly, native MS analysis of C2B interacting with isolated lipids (**Figure 29**) implied that C2B has a higher affinity for DOPS than for PI(4,5)P_2. However, DOPS was present in both model membranes mimicking the SV and plasma membranes and, according to native MS lipid preference, C2B would be expected to bind to liposomes of both compositions. This contradiction possibly originates from C2B interacting with multiple PI(4,5)P_2 head groups on the liposome membranes simultaneously in contrast to native MS analysis that reports interactions with individual lipids. Moreover, $K_{D\ app}$ values calculated for C2A and

C2B binding to PI(4,5)P_2 were similar. Nevertheless, C2A did not bind to liposomes mimicking the plasma membrane in the flotation assay, whereas C2B did (**Figures 30 and S8**). Therefore, the flotation assay and the native MS experiments likely report on slightly different interactions. Possibly, the lipid affinities observed in native MS include additional contributions involving, for example, the lipid acyl chains.

In summary, the flotation assay showed distinct membrane preferences for C2A, preferring the SV membrane, and C2B, preferring the plasma membrane, in the presence of Ca^{2+}. In the absence of Ca^{2+} only minor interactions of both C2 domains with the plasma membrane were found. This result differs from the native MS analysis of the isolated C2 domains. This is, most likely, a result of (1) different interactions playing a role during native MS analysis and in the flotation assay or (2) the C2 domains interacting with multiple lipids of the liposomes simultaneously during the flotation assay.

5.1.4 Analysis of Synaptotagmin-1 Lipid-Interactions *in silico*

To gain structural insights into membrane interactions of Syt-1 C2A and C2B, MD simulations of the C2 domains in the presence of planar lipid bilayers were performed. For this, the same lipid compositions used previously for the flotation assay of C2A and C2B interacting with liposomes (see **Section 4.1.3**) were used. Atomistic simulations of the C2 domains interacting with the lipid membranes in the presence and absence of Ca^{2+} were computed, corresponding to 900 ns simulation time per protein-membrane combination. The time range is similar to previously reported all-atom simulations of membrane interactions of soluble proteins [264, 265] and therefore well suited to investigate the initial interaction of Syt-1 with lipid membranes.

It should be noted that SVs and liposomes are spherical membranes with a radius of 25 nm [266] and 50 nm, respectively. Therefore, modelling the SV membrane as a planar membrane introduces an error. However, mapped on quadratic plane of 76 Å side length as used during the simulations in this thesis, the radii of SVs and liposomes correspond to differences of the axial lipid coordinates of approximately 2.9 Å and 1.4 Å, respectively, between the edges and the centre of the plane. This difference is small compared to the diameter of a lipid leaflet (approximately 20 Å) and smaller than the length of, for example, the PS head group (approximately 5 Å). Therefore, the error introduced by ignoring the curved character of liposomes and SVs is likely small. In contrast, the plasma membrane exhibits a low curvature *in vivo* and, therefore, is correctly represented by a planar model membrane.

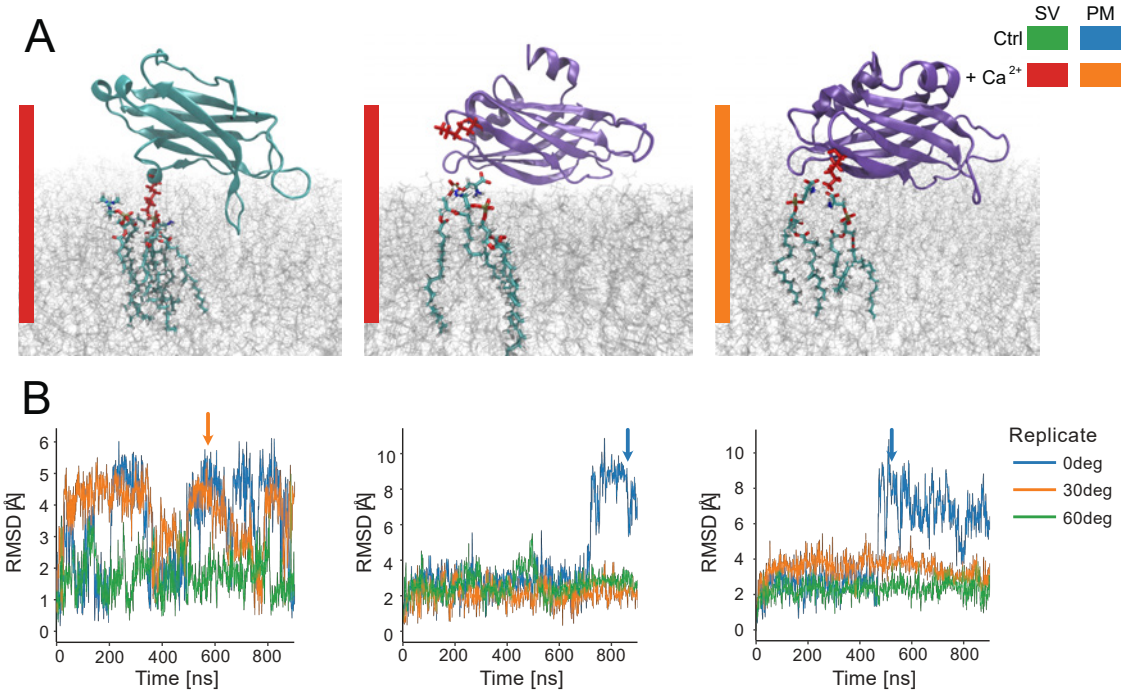


Figure 54: Contacts Between the Polybasic Regions of Syt-1 C2A and C2B and the Membrane **(A)** Snapshots of Ca^{2+}-bound Syt-1 C2A (left, 566 ns simulation time) interacting with the SV membrane, C2B interacting with the SV membrane (middle, 861 ns simulation time) and C2B interaction with the plasma membrane (right, 515 ns simulation time). Lipid molecules in proximity to Syt-1 are shown in stick representation. C2A Lys190 and C2B Lys321 are coloured (red). **(B)** Corresponding RMSD plots for all three replicates of a protein-membrane combination. The time point corresponding to the snapshot in (A) is indicated (arrow).

Calcium-Binding Modulates Loop Dynamics in C2A and C2B Structural stability of all simulated C2 domains was maintained over the course of the simulation as indicated by RMSDs of <2 Å (**Figure S9**). Whereas global RMSDs remained stable over the trajectory, several loops of C2A and C2B showed considerable dynamics in RMSF analysis (**Figures 31 and 33**). In particular, two homologous loops in C2A and C2B showed opposite dynamics in the presence and absence of Ca^{2+}. To analyse the temporal changes in loop position in more detail, protein structures were aligned with respect to residues flanking the flexible loops and RMSDs were calculated. A CBL involving residues number 172 - 174 in C2A and 302 - 305 in C2B showed increased dynamics with loop RMSDs >4 Å in the absence of Ca^{2+} (**Figures 32A and 34B**). In the presence of Ca^{2+}, the loop RMSD was reduced to <2 Å in line with aspartic acid residues in the loops participating in Ca^{2+} complexation. In contrast, increased dynamics in the presence of Ca^{2+} were observed for a second loop involving positions 189 - 190 in C2A and homologous residues number 320 - 321 in C2B located on the face of the C2 domains opposite to the CBLs. In C2B, this loop is located close to the polybasic region that is involved in interactions with PIPs [57]. Interestingly, increased dynamics in

the homologous loop in C2A were observed in two simulations involving the SV membrane which does not contain PIPs (**Figures 54A and 32B**), whereas an increase in dynamics was not observed in the presence of the plasma membrane. In contrast, increased dynamics in C2B were observed in two simulations involving the SV membrane or the plasma membrane (**Figures 54B and C** and **Figure 34C**). Close inspection of the simulation trajectories revealed that the time point of increase in RMSD corresponded to a parallel orientation of C2B with respect to the membrane and with the polybasic region near the membrane surface, suggesting that the increased dynamics correlated with lipid contacts. However, manual inspection of close lipids for the three simulations did not indicate a favourite lipid interaction for all three simulations (**Figure 54**). In summary, analysis of local dynamics of Syt-1 C2A and C2B revealed reduced dynamics of one CBL in both C2 domains and increased dynamics for a loop located in the polybasic region of C2B and in a homologous region in C2A, respectively, in the presence of Ca^{2+}.

Calcium-Binding Modulates the Membrane Contact Angle Binding of Ca^{2+} to Syt-1 is considered the trigger for SNARE complex zippering and membrane fusion. Therefore, it is likely that Ca^{2+} modulates the membrane interactions of Syt-1 for example by favouring interactions with negatively charged lipids [50, 68]. To investigate the initial stages of the Syt-1 membrane interaction, membrane contact angles between isolated C2A or C2B domains and the SV and plasma model membranes were analysed (**Figure 35**). For this, the angles between the membrane normal and a vector along the C2 domains was calculated (see **Section 3.2** for details). MD simulations indicated a Ca^{2+}-dependent change of the contact angle of C2A, whereas C2B remained nearly parallel to the membrane and showed only slight differences in the contact angle between simulations in the presence and absence of Ca^{2+}.

Previous studies on C2 domains interacting with lipid membranes showed that changes in membrane contact angle correlate with protein-membrane interactions. For example, the tilt angle of the C2 domain of protein kinase C_α depends on the presence of $PI(4,5)P_2$ in the target membrane [267]. In more detail, the Ca^{2+}-bound C2 domain was aligned parallel to the membrane in the absence of $PI(4,5)P_2$, whereas the presence of $PI(4,5)P_2$ correlated with a more upright orientation of the C2 domain (corresponding to lower angles as calculated in this thesis). This is in contrast to the simulations of Syt-1 in this thesis which showed that the membrane contact angle for Ca^{2+}-bound C2A or C2B was similar during contact with the $PI(4,5)P_2$-free SV membrane or the $PI(4,5)P_2$-containing plasma membrane. A similar analysis of the C2A domain of Synaptotagmin-7 concluded that the tilt angle correlated with the membrane insertion depth of a phenylalanine residue in one of the CBLs [268]. Larger contact angles (i.e. the C2 domain located parallel to the membrane) correlated with less insertion, whereas a more upright orientation of the protein enabled the residue to contact the hydrophobic lipid tails. Interestingly, a similar correlation was found in this thesis. C2B, which was aligned parallel to the membrane under most conditions showed

less membrane penetration than C2A (see below). Moreover, the presence of Ca^{2+} correlated with deeper membrane insertion and lower tilt angles for C2A and C2B, although the effect was significantly stronger for C2A. This suggests that different insertion of the CBLs of C2A and C2B into the membrane might be the reason for the different contact angles observed in this thesis.

In contrast to the computational analysis presented in this thesis, site-directed spin-labelling and electron paramagnetic resonance (EPR) spectroscopy experiments suggest that Syt-1 C2A and C2B adopt an upright orientation on membranes composed of PC and PS [269, 270, 54]. The difference between *in vitro* experiments and *in silico* simulations of C2B may have several reasons. First, the simulation possibly did not sample the part of the conformations interrogated by EPR spectroscopy due to insufficient computation time suggesting that the reorientation of C2B occurs on time scales longer than 900 ns. In line with this, 4 µs MD simulations of C2A and C2B on a lipid bilayer containing PS and $PI(4,5)P_2$ [265] showed a similar orientation of the C2 domains as reported in this thesis. However, if sampling would be the only reason for the discrepancies between MD simulations and EPR analyses, a similar effect would be expected for C2A that showed an upright orientation in the simulations and according to EPR analyses. Second, an upright orientation of C2B as observed by EPR might require additional steps that occur at longer time scales than treated in the MD simulations. This could involve for example lipid diffusion and reorientation in the membrane or conformational changes in C2B. Indeed, a previous publication suggested that a reorientation of the C-terminal helix of C2B upon Ca^{2+}-binding is involved in membrane curvature modulation [271]. Possibly, such a conformational change is not only a prerequisite for membrane bending but also for the orientation of C2B observed by EPR spectroscopy.

Lipid Interactions of Synaptotagmin-1 Lipid contacts during MD simulations of Syt-1 C2A and C2B interacting with the SV and plasma model membranes were analysed to infer lipid preferences of the C2 domains (**Figure 36**). To correct for lipid concentrations in the membranes and to compare lipids preferences between simulations, a normalised contact number was calculated (see **Section 3.2.6** for details). The highest number of normalised lipid contacts observed for C2A and C2B involved contacts to $PI(4,5)P_2$ followed by DOPS. This is in line with experimental observations that $PI(4,5)P_2$ and PS are important for Syt-1 membrane interactions in the presence of Ca^{2+} [54]. The *in silico* lipid contact analysis further agrees with the lipid overlay assay (**Figure 22**) and the native MS experiments involving Syt-1 C2AB that indicated a preference for PIPs. Interestingly, it is not in agreement with the native MS analysis of the isolated C2A and C2B domains (see above for a discussion of differences between the analysis of C2AB and the isolated C2 domains).

A difference in normalised lipid contacts between simulations of the C2 domains in the presence and absence of Ca^{2+} was not observed. Ca^{2+}-binding to the CBLs is considered to neutralise the negative charge of the loops, thereby favouring the interaction of Syt-1 with negatively charged membrane

lipids [68]. Previously reported stopped-flow experiments of the C2AB variant interacting with lipid vesicles of various compositions indicated that the presence of Ca^{2+} reduces the dissociation rate of Syt-1 from the membrane instead of changing the association rate [54]. The MD simulations performed in this thesis cover only a short time period with few membrane binding or dissociation events (see **Figures 38C and D**). Therefore, the simulation time is likely insufficient to distinguish binding and dissociation rates and an effect of Ca^{2+} on the dissociation rate cannot be detected in the simulations. The simulations are, therefore, most likely representative of an initial Syt-1 membrane association unaffected by the presence of Ca^{2+}.

Most contacts of C2A and C2B with DOPS and PI(4,5)P_2 occurred via the CBLs and the polybasic region (**Figure 37**). Previously, binding of Syt-1 to the membrane via the polybasic region of C2B was studied using lysine to alanine mutations in the polybasic region of C2B. These mutations reduced binding of Syt-1 to vesicles with [52] or without PI(4,5)P_2 [272] and reduced affinities of Syt-1 for PI(4,5)P_2 and PS in the presence and absence of Ca^{2+} [53, 54] implying that the polybasic region is important for PI(4,5)P_2 and PS affinity. It should be noted that another study reported that the mutation does not affect binding of Syt-1 to vesicles containing PS or PI(4,5)P_2 [72]. Lipid contact analysis performed in this thesis suggests that interactions with PS and PI(4,5)P_2 in the presence and absence of Ca^{2+} are mediated by the polybasic region of C2B (residues 324 - 327) and a similar basic region on the surface of C2A (residues 189 - 192). Interestingly, most contacts of DOPS with C2B were made via the polybasic region, whereas contacts with PI(4,5)P_2 occurred predominantly via the CBLs. In contrast, contacts between DOPS and C2A happened mostly through interactions involving the CBLs and contacts with PI(4,5)P_2 occurred to a similar extent in the CBLs and the polybasic region. Contact numbers between PS and the CBLs were generally higher in simulations involving Ca^{2+} than in simulations without Ca^{2+} in line with Ca^{2+}-binding increasing the CBL affinity for PS [68]. Interactions with PI(4,5)P_2 via the CBLs were previously suggested for C2A [273, 70] and C2B [53] as well as for the C2AB fragment [52] in the presence of Ca^{2+}. Indeed, simulations of C2A showed increased PI(4,5)P_2 contact frequencies in the presence of Ca^{2+}. However, a similar trend was not observed for C2B.

Overall, MD simulations of C2A and C2B indicate similar lipid preference and contact regions for PS and PI(4,5)P_2 as previously reported. Lipid contact numbers calculated in this thesis suggest a major role of the CBLs for binding to PS and PI(4,5)P_2. Similarly, contacts with the polybasic region of C2B and a homologous region in C2A were observed for PS and PI(4,5)P_2. This supports previous reports that the polybasic region is important for PI(4,5)P_2-binding [53, 54], however, raises the question if the polybasic region is only involved in mediating Syt-1 contacts with PI(4,5)P_2 or if it also plays a role in binding to PS.

Penetration of Syt-1 Into Lipid Membranes MD simulations of Syt-1-membrane interactions were prepared using a lipid membrane compositions similar to the flotation assay (**Figure 30**) to directly compare experiment and simulation. Penetration depth analysis of the MD simulations indicated a deeper insertion of C2A in the SV and plasma membranes in the presence than in the absence of Ca^{2+} (**Figure 38**A) . Similarly, the flotation assay of C2A in the presence of liposomes revealed low binding of C2A to liposomes mimicking the SV and plasma membranes in the absence of Ca^{2+}, whereas binding was observed for both liposome compositions in the presence of Ca^{2+}. Taken together, experiment and simulations indicate that binding of C2A to liposomes in the presence of Ca^{2+} is mediated by insertion of the C2 domain into the membrane. In contrast, experimental analysis and simulation of C2B interacting with membranes of both compositions diverge. Whereas the flotation assay suggests a weak interaction of C2B with liposomes composed of plasma membrane lipids in the absence of Ca^{2+} that increased in the presence of Ca^{2+} (**Figure 30B**), MD simulations of C2B suggest a similar insertion into the plasma membrane in the absence and presence of Ca^{2} (**Figure 38B**). Furthermore, the largest membrane insertion was determined for C2B interacting with the SV membrane in the presence of Ca^{2+}. However, binding of C2B to liposomes composed of SV lipids was weaker than to liposomes composed of plasma membrane lipids. Possibly, the difference between experimental results and theoretical simulations of Syt-1 C2B membrane is the result of the different time scales of the two methods. Whereas the liposome flotation assay reports on interactions that are stable for several minutes to hours, the MD simulations cover interactions in the nanosecond to microsecond timescale. Therefore, effects such as protein conformational changes [271] or lipid diffusion are not covered by the MD simulations (also see discussion on membrane contact angles). Overall, MD simulations of Syt-1 C2A are in good agreement with the experimental flotation assay, whereas simulations of C2B do not correlate with the experimental results. Possibly, membrane interactions of C2B occur on longer timescales than sampled in the MD simulations.

Previous EPR spin-label experiments showed penetration of C2A [269] and C2B [270] into lipid bilayers composed of PC and PS in the presence of Ca^{2+}. The maximum insertion depth of the spin-labels below the lipid phosphate layer in these studies was 5.3 ± 3.0 Å and 3.8 ± 3.0 Å for C2A and C2B, respectively. This is in reasonable agreement with the maximum insertion depth of 6 Å that was calculated based on MD simulations of C2A and C2B interacting with the SV membrane in the presence of Ca^{2+} (**Figure 38**). Interestingly, the EPR studies reported a lower membrane penetration depth for C2B compared to C2A, whereas MD simulations suggest that both C2 domains penetrate into the membrane to a similar distance. However, C2A inserted into the membrane in all three replicates, whereas C2B inserted into the membrane only in a single replicate. EPR analysis of the C2AB variant interacting with lipid membranes containing $PI(4,5)P_2$ showed insertion of the spin-label up to 9 Å below the lipid phosphate bilayer [54]. In contrast, penetration into the plasma membrane was below 5 Å and 6 Å for C2A and C2B (**Figure 38**). Even though the penetration depths calculated for the EPR measurement are likely larger than C_α distances calculated from the

MD simulations due to the additional length of the spin label, MD simulations presented in this thesis and previous EPR analyses suggest that C2AB likely inserts deeper into the lipid membrane than the C2A or C2B domains individually.

A recent publication on MD simulations of Syt-1 C2A and C2B interacting with model membranes reported insertions of C2A and C2B into the plasma membrane of 15 Å in the absence of Ca^{2+} and of 20 Å in the presence of Ca^{2+} [265]. A similar tendency of deeper insertion of Syt-1 into SV and plasma membrane model bilayers in the presence than in the absence of Ca^{2+} was observed in the MD simulations of C2A (**Figure 38**). Interestingly, no clear effect of Ca^{2+} was observed for C2B in this thesis, probably related to the overall shorter length of the trajectories of 900 ns compared to 4 µs in the publication. In contrast to the publication in which distances between any protein atom and the lipid phosphate layer of the membrane were reported [265], insertion depths in this thesis were calculated as distances between C_α atoms and the phosphate layer. However, assuming a maximum distance of 8 Å between the C_α atom of a residue and any side chain atom, the previously published simulations still suggest insertion of Syt-1 up to 12 Å into the membrane bilayer and therefore a deeper membrane penetration than observed in this thesis.

Effects of Synaptotagmin-1 on the Membrane Surface Analysis of the effect of Syt-1 on the membrane surface revealed a change in membrane phosphate z-positions in the presence of Ca^{2+}-bound Syt-1 C2A and C2B (**Figures 39A and B**). A similar effect with lower magnitude was observed in the absence of Ca^{2+}. This was unexpected as the analysis of membrane bending curvature MD simulations generally requires a finite membrane in at least one direction to account for the curvature induced shrinking of the membrane area. Moreover, measures to enhance lipid diffusion [271] are generally required to observe membrane-modulating effects. However, mean curvatures calculated in **Figure 39** correspond to spheres with radii of approximately 1 Å - 2 Å. Most likely, the membrane curvature observed in this thesis resulted from local displacement of lipid molecules for example by insertion of the C2 domains into the lipid bilayers. In line with this assumption, a larger membrane curvature and a deeper insertion was identified for C2A compared to C2B (compare **Figures 38 and 39**). Moreover, simulations in which lipid displacement was observed also showed membrane penetration but not *vice versa*.

5.2 Structural Characterisation of the Viral Coat Protein p38

The TCV coat protein p38 interacts with different forms of RNA by (1) forming the viral capsid that encloses the viral genomic RNA and (2) mediating the evasion of the host-defence by interaction with small siRNAs. p38 consists of the R- S- and P-domains which are all interconnected by flexible linkers. Although structural characterisation of the capsid structure enabled important insights into

possible infection mechanisms [94], the structure of p38 in solution remains elusive. To structurally investigate the solution conformation of p38, the protein was recombinantly purified by apl. Prof. Dr. Ralph Golbik and characterised by structural MS.

The identity of purified p38 was first validated by gel electrophoresis and MS-based identification. In contrast to previous studies investigating p38 purified from chinese cabbage [274], no proteolytic degradation of the protein was observed (**Figure 40**). Moreover, proteolysis and LC-MS/MS analysis as well as native MS analysis of the intact protein confirmed the correct expression of p38.

Native MS Analyses Oligomerisation of wild-type p38 was analysed by native MS and showed that p38 formed monomers and dimers in solution in the absence of RNA (**Figure 41**). This confirms previous findings from analytical gel filtration, chemical cross-linking and gel electrophoresis [274] as well as structural considerations [89] that suggested that p38 forms a dimer in solution. Moreover, the p38 W274A mutant with reduced VSR activity [89] was analysed using native MS. Trp274 is located in the periphery of the P-domain, away from the P-P dimer interface and, therefore, a mutation of this residue is unlikely to impair p38 dimerisation directly. However, native MS analysis of p38 W274A suggested a lower degree of dimer formation in the mutant protein than in the wild-type protein (**Figure 43A**). This does not imply that the VSR-deficient phenotype of p38 W274A is caused by impairing p38 dimerisation. However, it might indicate a structural impact of this mutation on the P-P dimer interface. In line with this, the coexistence of separate p38 wild-type and p38 W274A dimers was observed (**Figure 43C**), possibly indicating structural differences in the P-P dimerisation interface between the wild-type and the mutant.

Next, binding of siRNAs to p38 was analysed. Native MS analysis of p38 with siRNAs of 21 nts and 22 nts length showed that p38 formed a 2:1 protein-RNA complex (**Figures 41B and C**). Higher-order oligomers were not observed implying that p38 mainly forms dimers in its RNA-bound state. In contrast to native MS of wild-type p38, analysis of the W274A mutant in the presence of 21 nts siRNA showed dimeric and monomeric p38 W274A (**Figure 43B**). The W274A mutant was not considered to be RNA-binding deficient, as the mutation was suggested to impair interactions of p38 with AGO [89]. However, the observation of a monomeric form of the protein in the presence of RNA indicates that RNA-binding was reduced, possibly a result of the lower degree of dimerisation observed by native MS in the absence of RNA.

Using unlabelled and ^{15}N labelled p38, the exchange of subunits in the p38 homo-dimer was analysed (**Figure 42**). In the absence of RNA, three complexes of masses corresponding to a ^{14}N/^{14}N, ^{14}N/^{15}N and ^{15}N/^{15}N dimer, respectively, were observed at 1:2:1 intensity ratio. As mixing of the differentially labelled forms was performed immediately prior to MS analysis, this suggests fast exchange of monomer subunits between different p38 dimers. Closer inspection of the P-P

interface of p38 in the capsid structure showed few polar contacts, mainly hydrophobic amino acids and two tyrosine residues in close contact. Possibly, a p38 monomer can successively displace a second monomer while keeping the buried surface area constant, leading to the observed fast exchange of monomers in the p38 dimer. After addition of 21 nts and 22 nts siRNA, fast subunit exchange was completely abolished (**Figures 42B and C**). This indicates that binding to siRNA stabilises the dimerisation interface so that an RNA-bound subunit is less likely displaced by a second copy of p38. Such an additive effect of siRNA-mediated stabilisation of the dimer was also observed for the W274A mutant for which the degree of dimerisation increased after addition of siRNA (**Figure 43A and B**). RNA possibly binds to the R-domain of both p38 monomers and thereby prevents the displacement of one monomer and, by that, subunit exchange.

Chemical Cross-Linking MS To gain information on the structural ensemble of p38 in solution, the protein was analysed by chemical cross-linking with BS2G and EDC. The experiment was carried out under three conditions, namely at 25 °C in the absence of RNA (**Figure 45**), at 5 °C in the absence of RNA (**Figure 47**) and at 5 °C in the presence of 21 nts siRNA (**Figure 48**). Detection of cross-linked peptide pairs in the samples cross-linked with BS2G confirmed the presence of p38 dimers in solution under all investigated conditions, in line with the results from native MS. Moreover, most of the five lysine residues located in the basic R-domain were found to be cross-linked with BS2G under all conditions. Cross-links to these residues were identified after cross-linking with low BS2G concentrations (**Figure 45A**). As solvent accessibility is a prerequisite for cross-linking, this indicates that lysine residues in the R-domain are easily accessible from the aqueous environment. Similarly, cross-links involving residues in the R-domain showed contacts to the C-terminal half of the P-domain and to most of the S-domain. Taken together, this indicates a high flexibility of p38 in general and of the R-domain in particular. Interestingly, a change in the cross-linked residues in R-domain was not observed between experiments carried out at 25 °C and 5 °C (**Figure 47**), a temperature at which an increasing secondary structure was determined by CD spectroscopy. Although the 11.4 Å cross-linker disulfosuccinimidyl suberate was previously considered for discriminating α-helices and β-sheets [275], discriminating unfolded and folded parts of the R-domain was not expected. However, a decrease in the number of contacts between the R- and P-domains was not observed at 5 °C either (**Figure 47**), implying that increased secondary structure does not impede the flexibility of the linker between the R- and S-domains. Similarly, a change in the number or positions of cross-links in the presence of RNA that is considered to interact with the basic R-domain through its phosphate backbone was not seen (**Figure 48**). Overall, the observed cross-linking pattern was similar for all three conditions and using both cross-linkers. Even though the precise cross-link positions between the different conditions varied (see **Figures 47A, 48A and 48C**), cross-links generally connected residues at similar positions of the primary sequence of p38.

Comparison of Cross-Links with Structural Models To interpret the cross-links obtained between the S- and P-domains, cross-linked residues were mapped onto a model of dimeric p38 (kindly provided by Prof. Dr. Milton T. Stubbs). Calculation of residue-residue distances in the model revealed that approximately half of the BS2G cross-links corresponded to distances above the maximum theoretical distance (**Figure 49A**). Most strikingly, cross-linked residue-pairs connecting two S-domains such as Lys94-Lys94 that were consistently identified under all experimental conditions were separated by >55 Å in the capsid dimer model. In contrast, most cross-links identified with EDC were in agreement with the model (**Figure 49B**). This is consistent with short-range cross-links often occurring within folded domains that are similar in all models presented here. However, cross-links with EDC were also identified between the S- and P-domains suggesting that some cross-links are informative on the relative positioning of the folded domains in solution. Nevertheless, a few EDC cross-links in disagreement with the model were determined and indicated a closer arrangement of the S-domains and between the S- and P-domains. However, native MS (**Figure 41A**), analytical ultracentrifugation (performed by Prof. Dr. Hauke Lilie) and previous research [274, 89] suggest that p38 forms dimers in solution. Furthermore, higher-order oligomers were not observed after gel electrophoresis of cross-linked p38 (**Figure 45A and E**), rendering the presence of p38 capsid structures in solution highly unlikely.

Cross-links disagreeing with the maximum theoretical distance when mapped on the p38 dimer model can originate from multiple sequential contacts (i.e. from a high flexibility) of p38 or from a less flexible but more compact conformation different from the capsid model. To investigate if a single compact dimer model compatible with the experimental cross-links can be generated, cross-links obtained using BS2G as cross-linking reagent were used to guide docking of the separated S-domains to the P-P dimer (**Figure 50**). The number of cross-links in agreement with the model was higher for the resulting docking models than for the capsid model. Nevertheless, contradicting cross-links such as Lys95-Glu313 and Lys95-Asp307 that cannot be fulfilled by a single model and several residue-residue distances incompatible with the cross-links between the S- and P-domains were also present in these models. Moreover, as the models were docked without the connecting loop sequences, the solvent-accessible surface distances corresponding to these loops in the models were calculated (**Table 19**). All calculated S- to P-domain distances exceed the maximum contour length of the loop peptide $_{240}$NRTGSTSDA$_{248}$. However, it should be noted that Gln249, the residue C-terminal to the removed loop sequence, is located within a short β-sheet structure attached to the β-barrel of the P-domain. Detachment of the short β-sheet would increase the length of the flexible loop connecting S- and P-domain by 14 residues, thereby allowing distances of up to 92.5 Å between the S- and P-domains that are compatible with most solvent-accessible surface distances calculated for the docking models.

The generated models were subjected to short MD simulations to test their structural stability. The model of the p38 dimer derived from the capsid structure showed only minimal structural changes over the simulation time. This conformational stability is in line with the structural stability of the viral capsid and, moreover, indicated that the transition from capsid to the RNA-binding soluble dimer occurs on longer timescales than those sampled in the simulation [276]. MD simulations of the docking models revealed that the model 1 was structurally stable and remained in its compact conformation for the 300 ns simulation time. In contrast, the model 2 showed an increase of inter S-domain distance immediately after the start of the simulation, suggesting that this conformation likely does not exist in solution (**Figure 51**). Therefore, a compact conformation of the p38 dimer in agreement with most of the experimental cross-links is stable over short time periods. However, the simulations cannot rule out other explanations for the observed cross-links. Most cross-links are in agreement with a structural model of the intact p38 capsid (**Figure S18**). The presence of a capsid during cross-linking is very unlikely as native MS implied that p38 mainly exists in a monomer-dimer equilibrium in solution (**Figure 41**) and no higher-order oligomers were observed after polyacryalmide gel electrophoresis of cross-linked p38 (**Figure 45**). A more likely explanation for the observed cross-links is high structural flexibility of p38 i.e. the experimental cross-links are representative of multiple different conformations of p38. High flexibility of the p38 dimer would, furthermore, explain the identification of contradictory cross-links incompatible with a single structure such as Lys95-Glu313 and Lys95-Asp307. To discriminate between the existence of a single compact and multiple extended conformations, additional experimental information such as radii of gyration is required (see **Section 5.3**).

5.3 Conclusion and Outlook

Proteins play central roles in cell integrity, metabolism and signalling that involve binding to various biomolecules such as RNA and lipids (see **Section 1.1.1**). Protein-lipid interactions are of particular importance for the neuronal protein Syt-1, whereas protein-RNA interactions are crucial for the function of the viral coat protein p38. In this thesis, binding of Syt-1 to lipids was analysed using native MS and computational modelling, moreover, the structure of p38 was characterised in the presence and absence of RNA.

Syt-1 Lipid Preference *in vitro* The neuronal regulatory protein Syt-1 controls the SNARE complex-mediated fusion of SVs with the presynaptic membrane in response to Ca^{2+} influx during neurotransmission. The function of Syt-1 involves interactions with lipids of the SV or plasma membranes, however, many open questions regarding the lipid specificity and membrane interaction of Syt-1 remain. In order to biophysically characterise Syt-1, full-length Syt-1 and truncated soluble variants were recombinantly expressed (**Figure 18**), their correct expression was validated

(**Figure 19**) and their oligomeric state was analysed using native MS (**Figure 20**). Full-length Syt-1 and truncated variants existed mainly in a monomeric state, however, low amounts of dimers, presumably mediated through C2A/C2B interactions, were detected. Syt-1 consists of a transmembrane helix and two cytoplasmic C2 domains connected by flexible linkers. Regulation of membrane fusion by Syt-1 involves Ca^{2+}-dependent membrane-interactions of the C2 domains. To analyse lipid preferences of the isolated C2 domains, a native MS assay to characterise lipid interactions of soluble proteins was established.

For this, a detergent enabling native MS analysis of peripheral membrane proteins interacting with isolated lipids was identified by assessing the effects of three detergents on well characterised standard proteins. The detergent C8E4 reduced the average charge of protein ions, presumably through proton capture by the detergent (**Figure 12**). Moreover, soluble protein-lipid complexes were observed by native MS analysis of β-LG in the presence of C8E4 (**Figure 15**). Formation of protein-lipid complexes during native MS analysis in the presence of C8E4 likely involves specific interactions between the protein and the lipid as the experimentally determined ratios between lipid-bound and lipid-free protein peaks were higher than expected by stochastic association (**Figure 53**). In contrast, the two other detergents, namely OG and LDAO, did not lead to the formation of protein-lipid complexes of soluble proteins.

Native MS analysis of the soluble Syt-1 C2AB variant interacting with intact lipids in the presence of C8E4 showed that net charge is a major factor determining lipid binding and that C2AB preferentially binds $PI(4,5)P_2$ and $PI(3,4,5)P_3$ lipids (**Figure 27**). In contrast, the shorter C2A and C2B variants preferentially bound to PS and showed overall higher lipid binding affinities. Using the liposome flotation assay (**Figure 30**) affinities of the C2A and C2B variants for intact lipid membranes were interrogated, showing that both variants have low membrane affinity in the absence of Ca^{2+}. In the presence of Ca^{2+}, C2A preferentially interacted with the SV membrane, whereas C2B bound to the plasma membrane.

Syt-1 Lipid Interactions *in silico* To gain insights into the interactions of C2A and C2B with intact membranes, 900 ns MD simulations of C2A and C2B interacting with SV and plasma membrane model membranes were computed. Simulations showed reduced dynamics of one CBL in the presence of Ca^{2+}, in line with its Ca^{2+}-chelating function, and increased dynamics of a second loop on the opposite side of C2A and C2B under the same conditions (**Figure 34**). Calculation of the average membrane interaction contact angle of the C2 domains interacting with SV or plasma membrane models showed that binding to Ca^{2+} shifts the contact angle of C2A to more upright (i.e. less membrane parallel) orientations (**Figure 35**). C2B was aligned parallel to the lipid bilayer in the simulations with only a minor effect of Ca^{2+} on the contact angle in contrast to experimental evidence from EPR spin labelling [269, 270, 54] but in line with a previous *in silico* analysis [265]. Most likely, EPR analysis reported tilt angles in an equilibrated C2B-membrane system,

whereas MD simulations captured the initial membrane contacts of C2B and, thereby, provide a distinct perspective on the protein-membrane interaction of Syt-1. By calculating the number of contacts between the C2 domains and membrane lipids, a preference of C2A and C2B for $PI(4,5)P_2$ and binding to DOPS was observed (**Figure 36**). However, a difference in lipid affinity between simulations in the presence and absence of Ca^{2+} was not determined. Protein membrane contacts were mapped on the surface of the C2 domains (**Figure 37**) and negatively charged phospholipids predominantly interacted with the CBLs and the polybasic regions of C2A and C2B as previously suggested. Membrane insertion depths of the C2 were Ca^{2+}-dependent for interactions of C2A with the SV and plasma membranes (**Figure 38**) in line with *in vitro* analysis using the flotation assay (**Figure 30**) und with analyses of membrane lipid positions (**Figure 39**). Only a minority of simulations involving C2B showed membrane insertion and no clear trend for a Ca^{2+}-dependence of insertion depth emerged.

Implications for Models of Syt-1 Function The biophysical characterisation of Syt-1 presented in this thesis supports some models of Syt-1 function presented in **Section 1.2**. Results in this thesis suggests that Syt-1 predominantly interacts with negatively charged lipids, mainly PS and $PI(4,5)P_2$. Therefore, interactions are most likely mediated through binding to the lipid head groups. Syt-1 was suggested to modulate membrane properties and destabilise the membrane for SNARE-mediated fusion [66, 74, 54] or to induce membrane curvature [75, 74]. During MD simulations, Syt-1 inserted into the membranes (**Figures 38 and 39**) a process that *in vivo* likely modifies the local membrane properties. However, conclusions on Syt-1-induced membrane curvature require longer simulation time or increased lipid diffusion and, therefore, were not possible with this set of simulations.

Furthermore, Syt-1 was proposed to interact simultaneously with the SV and plasma membranes in the presence of Ca^{2+}, thereby tethering SVs to the presynaptic membrane [77, 78]. In line with this assumption, the liposome flotation assay presented in this thesis showed that C2A and C2B have opposite membrane preferences in the presence of Ca^{2+} (**Figure 30**). In the absence of Ca^{2+}, results from the liposome flotation assay suggest that C2A does not bind to membranes, whereas C2B showed a low degree of binding to the plasma membrane. Moreover, native MS analyses revealed a preference of both C2 domains for PS in the absence of Ca^{2+}, thereby indicating no membrane selectivity of the variants in the absence Ca^{2+} (**Figures 28 and 29**). Interestingly, MD simulations did also not show selectivity of C2A and C2B for the SV and plasma membranes, indicating that membrane selectivity likely occurs on longer time scales than interrogated by MD simulations. Overall, the results presented in this thesis indicate that Syt-1 does not interact specifically with the SV and plasma membranes in the absence of Ca^{2+}. However, Syt-1 binds to the SV membrane via C2A and to the plasma membrane via C2B in the presence of Ca^{2+}. Membrane selectivity of the C2 domains is presumably derived from reduced dissociation of membrane-bound protein rather than

from increased affinity for a certain membrane composition as suggested previously [54], thereby explaining high similarity of lipid binding profiles of C2A and C2B observed by MD simulation (**Figure 36**).

Future Directions for the Analysis of Synaptic Membrane Fusion Comparison of $K_{D\ app}$ values of C2AB binding to $PI(4,5)P_2$ (**Figure 27**) and to its head group $I(1,4,5)P_3$ (**Figure 24**) suggests that the fatty acyl chain contributes to the affinities observed during native MS analysis. Therefore, the contribution of fatty acyl chain length and saturation on lipid affinities determined using native MS should be investigated in future experiments. Furthermore, the SNARE protein Synaptobrevin-2 and the SNARE-associated protein Complexin-1 were suggested previously to directly bind to lipid membranes [277, 278]. Assessment of the lipid affinities of Synaptobrevin-2 and Complexin-1 and comparison with the lipid preferences of Syt-1 presented in this thesis could reveal a common lipid preference of these proteins that could be involved in mutual regulation of the SNARE complex and Syt-1.

Analysis of the MD simulations showed a good agreement between experiment and simulation (see **Section 5.1.4**). However, some discrepancies between the flotation assay of the C2B domain and the corresponding MD simulations were observed. These differences likely originate from the different timescales investigated in the simulations and the experiments. Most importantly, conformational changes of the protein and diffusion of lipids should be included in the simulation of Syt-1 membrane interactions. Investigation of protein movements can be achieved by increasing the simulation time until the conformational change is sampled or by biasing the potential landscape towards a certain known conformation [279]. Increasing the simulation time most likely involves a certain amount of coarse graining of the simulations, i.e. representing multiple atoms as a single bead [280] to reduce computational load. Alternatively, a putative structural model after the conformational change could be generated and used as input for simulations [271]. Second, lateral diffusion of lipids in a bilayer occurs in the micrometer per second timescale [281] and is inaccessible to atomistic MD simulations. Various tools to increase lipid diffusion such as representation of the membrane core as an organic solvent [282] or reducing the overall spatial resolution [283] were put forward to overcome this challenge. Therefore, these tools should be included in future simulations of Syt-1-lipid interactions to judge the influence of Syt-1 on local lipid composition.

p38 Protein-RNA Interactions Using native MS, the oligomeric state of p38 in solution was determined. In the absence of RNA, the protein exists in monomeric and dimeric states, whereas binding to RNA shifts the equilibrium to the dimeric RNA-bound form (**Figure 41**). RNA-binding of the p38 dimer was observed for siRNA of 21 nts and 22 nts length and stabilised the p38 dimer as drastically reduced subunit exchange of p38 monomers was observed (**Figure 42**). In contrast to wild-type p38, the W274A mutant showed reduced dimerisation in native MS indicating an effect of

the mutation on the p38 dimer interface (**Figure 43**). Chemical cross-linking of p38 in the presence and absence of RNA was used to characterise its structure in solution (**Figure 48**). Only a minor difference in the cross-linking pattern was observed between the two conditions indicating that p38 structure in the presence and absence of RNA is similar. The identified cross-links were used for structural modelling of the relative positions of the S- and P-domains of p38. The resulting compact models are in agreement with most of the cross-links and MD simulations of the models showed that one model was structurally stable for short periods of time. This suggests that p38 may adopt a compact conformation, possibly stabilised by binding of RNA, additional contacts between the R-domain and the S- and P-domains or both. Alternatively, p38 may exist in a conformational ensemble of multiple interchanging conformations, each of which satisfies a subset of the identified cross-linking restraints.

Challenges in Structural Analysis of p38 by Chemical Cross-Linking Using the cross-linking restraints presented in this thesis, a compact model for p38 in solution fulfilling most of the cross-link restraints was proposed. However, the existence of an ensemble of p38 conformations in which no single structure is in agreement with all observed cross-links is equally possible. This ambiguity arises from three methodological challenges: First, experimentally determined cross-links are derived from all peptide pairs in accessible distance of a structural ensemble. It is, therefore, not possible to distinguish cross-links originating from different conformations of a residue pair even if the conformations are largely different. To face this challenge, cross-links can be used for validation of a structural ensemble produced for example by MD simulation rather than as restraints for structure determination [284]. Future studies of p38, therefore, should include extensive MD simulations of p38 in solution to define the conformational landscape the protein can explore and compare it with the experimental cross-links. Second, cross-links from non-overlapping peptide sequences can derive from the interaction of two copies of p38 in the dimer or from contacts within a single copy of p38. Therefore, a correctly identified cross-link with non-overlapping peptide sequences may support two largely different molecular structures. Note that a similar cross-link from peptides with overlapping sequences can only derive from interactions of two copies of p38. It has been proposed to use isotope-labelled versions of the oligomerising protein to categorise cross-links into intra- or inter-protein groups [285]. Peptides resulting from inter-protein cross-links show a characteristic intensity pattern of the differentially labelled peptide combinations, whereas intra-protein cross-links appear as two distinct peaks corresponding to the labelled or unlabelled peptides, respectively. However, software to automatically analyse spectra from such an experiment is still sparse and its application is, therefore, limited to very low numbers of cross-links that can be manually annotated. Nevertheless, cross-linking of isotopically labelled p38 and manual spectrum annotation might be feasible for selected cross-links. Third, cross-linking restraints are unilateral, i.e. they only define a maximum distance leaving all possible shorter distances in a model valid. Therefore, cross-linking restraints are generally more likely fulfilled in compact structures such

as the compact model of the p38 dimer presented in this thesis. To experimentally elucidate the dimensions of the p38 dimer in solution, additional restraints should be generated and included into the modelling procedure. This could be, for example, donor-acceptor distances determined by fluorescence resonance energy transfer experiments or radii of gyration obtained by small-angle X-ray scattering. In contrast to cross-linking restraints, these experimental techniques report distance distributions with an upper and a lower boundary and, therefore, allow discriminating a compact soluble p38 dimer from an ensemble of extended conformations.

Future Directions for the Analysis of TCV Infection Although not used for structural modelling, the distance restraints obtained for the R-domain in this thesis are a good starting point for defining structure and contact points of the R-domain. The reliability of computational modelling of disordered regions has increased with recent force-fields [215], making a computational simulation of the flexible linker between the R and S domains feasible. After calculation of such a trajectory, the observed residue-residue distances should be compared with the experimental cross-links reported in this thesis. In addition to structural characterisation of the R-domain, a detailed molecular model of the p38 dimer interacting with RNA in solution would provide insights into p38 function. A first step in this direction was made in this thesis by generating a possible p38 dimer model based on cross-linking restraints. Moreover, a reduced subunit exchange in the p38 dimer after siRNA-binding was reported in this thesis. This may provide a starting point for the structural analysis of the p38 dimer using X-ray crystallography as reduced subunit exchange between the p38 is possibly accompanied by reduced dynamics and, therefore, better crystal formation. Another future experiment on the function of p38 during infection should investigate the preference of p38 for different RNAs. During infection, the host enzymes DCL4 and DCL2 process the viral dsRNA into siRNAs of 21 nts and 22 nts length. In this thesis, binding of both lengths of RNAs to p38 was observed (**Figure 41**). As DCL2 was reported to act as a rescuing mechanism after the inhibition of DCL4 during viral infection [92], the affinities of p38 to siRNA of both lengths should be investigated. For example, competition experiments using both forms of siRNA or other RNAs involved in viral infection such as dsRNA could provide insights into the RNA preference and, therefore, into the primary target of p38.

Overall, this thesis contributes to the understanding of protein-protein, protein-lipid and protein-RNA interactions by investigating structural changes related to these protein interactions. For example, binding of Ca^{2+} to Syt-1 C2A interacting with lipid membranes affected the orientation of the protein and binding of RNA to the p38 dimer reduced the subunit exchange of p38 monomers that was observed in p38 dimers in the absence of RNA. Due to the important role that protein-lipid and protein-RNA interactions play in the organisation and regulation of living cells, the biological and methodological results presented in this thesis hopefully add to the growing understanding of biology on a molecular level.

Bibliography

[1] T. Pühringer, U. Hohmann, L. Fin, B. Pacheco-Fiallos, U. Schellhaas, J. Brennecke, C. Plaschka, *eLife* **2020**, *9*, e61503, DOI 10.7554/elife.61503.

[2] M. Hellgren, L. Sandberg, O. Edholm, *Biophysical Chemistry* **2006**, *120*, 1–9, DOI 10.1016/j.bpc.2005.10.002.

[3] A. Andreeva, E. Kulesha, J. Gough, A. G. Murzin, *Nucleic Acids Research* **2019**, *48*, D376–D382, DOI 10.1093/nar/gkz1064.

[4] I. Sillitoe, N. Bordin, N. Dawson, V. P. Waman, P. Ashford, H. M. Scholes, C. S. M. Pang, L. Woodridge, C. Rauer, N. Sen, M. Abbasian, S. Le Cornu, S. D. Lam, K. Berka, I. H. Varekova, R. Svobodova, J. Lees, C. A. Orengo, *Nucleic Acids Research* **2020**, *49*, D266–D273, DOI 10.1093/nar/gkaa1079.

[5] T. D. Newport, M. S. P. Sansom, P. J. Stansfeld, *Nucleic Acids Research* **2018**, *47*, D390–D397, DOI 10.1093/nar/gky1047.

[6] O. N. Jensen, *Nature Reviews Molecular Cell Biology* **2006**, *7*, 391–403, DOI 10.1038/nrm1939.

[7] Y. M. Abbas, D. Wu, S. A. Bueler, C. V. Robinson, J. L. Rubinstein, *Science* **2020**, *367*, 1240–1246, DOI 10.1126/science.aaz2924.

[8] F. L. Kyrilis, D. A. Semchonok, I. Skalidis, C. Tüting, F. Hamdi, F. J. O'Reilly, J. Rappsilber, P. L. Kastritis, *Cell Reports* **2021**, *34*, 108727, DOI 10.1016/j.celrep.2021.108727.

[9] J. Deisenhofer, *Biochemistry* **1981**, *20*, 2361–2370, DOI 10.1021/bi00512a001.

[10] H. Ogiso, R. Ishitani, O. Nureki, S. Fukai, M. Yamanaka, J.-H. Kim, K. Saito, A. Sakamoto, M. Inoue, M. Shirouzu, S. Yokoyama, *Cell* **2002**, *110*, 775–787, DOI 10.1016/s0092-8674(02)00963-7.

[11] P. Nioche, W.-Q. Liu, I. Broutin, F. Charbonnier, M.-T. Latreille, M. Vidal, B. Roques, C. Garbay, A. Ducruix, *Journal of Molecular Biology* **2002**, *315*, 1167–1177, DOI 10.1006/jmbi.2001.5299.

[12] M. Wittekind, C. Mapelli, B. T. Farmer, K.-L. Suen, V. Goldfarb, J. Tsao, T. Lavoie, M. Barbacid, C. A. Meyers, L. Mueller, *Biochemistry* **1994**, *33*, 13531–13539, DOI 10.1021/bi00250a004.

[13] M. Fuxreiter, *Current Opinion in Structural Biology* **2018**, *54*, 19–25, DOI 10.1016/j.sbi.2018.09.008.

[14] R. W. Carthew, E. J. Sontheimer, *Cell* **2009**, *136*, 642–655, DOI 10.1016/j.cell.2009.01.035.

[15] A. Re, T. Joshi, E. Kulberkyte, Q. Morris, C. T. Workman in *RNA Sequence, Structure, and Function: Computational and Bioinformatic Methods*, (Eds.: J. Gorodkin, W. L. Ruzzo), Humana Press, Totowa, NJ, **2014**, pp. 491–521, DOI 10.1007/978-1-62703-709-9_23.

[16] M. W. Hentze, A. Castello, T. Schwarzl, T. Preiss, *Nature Reviews Molecular Cell Biology* **2018**, *19*, 327–341, DOI 10.1038/nrm.2017.130.

[17] N. Pumplin, O. Voinnet, *Nature Reviews Microbiology* **2013**, *11*, 745–760, DOI 10.1038/nrmicro3120.

[18] P. A. Temussi, G. G. Tartaglia, A. Pastore, *BioEssays* **2021**, *43*, 2100031, DOI 10.1002/bies.202100031.

[19] R. Stefl, F. C. Oberstrass, J. L. Hood, M. Jourdan, M. Zimmermann, L. Skrisovska, C. Maris, L. Peng, C. Hofr, R. B. Emeson, F. H.-T. Allain, *Cell* **2010**, *143*, 225–237, DOI 10.1016/j.cell.2010.09.026.

[20] F. C. Oberstrass, *Science* **2005**, *309*, 2054–2057, DOI 10.1126/science.1114066.

[21] S. Dong, Y. Wang, C. Cassidy-Amstutz, G. Lu, R. Bigler, M. R. Jezyk, C. Li, T. M. T. Hall, Z. Wang, *Journal of Biological Chemistry* **2011**, *286*, 26732–26742, DOI 10.1074/jbc.m111.244889.

[22] A. T. Phan, V. Kuryavyi, J. C. Darnell, A. Serganov, A. Majumdar, S. Ilin, T. Raslin, A. Polonskaia, C. Chen, D. Clain, R. B. Darnell, D. J. Patel, *Nature Structural & Molecular Biology* **2011**, *18*, 796–804, DOI 10.1038/nsmb.2064.

[23] P. A. Chong, J. D. Forman-Kay, *Current Opinion in Structural Biology* **2016**, *41*, 180–186, DOI 10.1016/j.sbi.2016.08.001.

[24] A. E. Rusiñol, Z. Cui, M. H. Chen, J. E. Vance, *Journal of Biological Chemistry* **1994**, *269*, 27494–27502, DOI 10.1016/s0021-9258(18)47012-3.

[25] J. T. Best, P. Xu, T. R. Graham, *Current Opinion in Cell Biology* **2019**, *59*, 8–15, DOI 10.1016/j.ceb.2019.02.004.

[26] H. Schulze, T. Kolter, K. Sandhoff, *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2009**, *1793*, Lysosomes, 674–683, DOI https://doi.org/10.1016/j.bbamcr.2008.09.020.

[27] L. Kuhlen, P. Abrusci, S. Johnson, J. Gault, J. Deme, J. Caesar, T. Dietsche, M. T. Mebrhatu, T. Ganief, B. Macek, S. Wagner, C. V. Robinson, S. M. Lea, *Nature Structural & Molecular Biology* **2018**, *25*, 583–590, DOI 10.1038/s41594-018-0086-9.

[28] A. Elofsson, G. von Heijne, *Annual Review of Biochemistry* **2007**, *76*, 125–140, DOI 10.1146/annurev.biochem.76.052705.163539.

[29] M. T. Agasid, C. V. Robinson, *Current Opinion in Structural Biology* **2021**, *69*, 78–85, DOI 10.1016/j.sbi.2021.03.010.

[30] A. Whited, A. Johs, *Chemistry and Physics of Lipids* **2015**, *192*, 51–59, DOI 10.1016/j.chemphyslip.2015.07.015.

[31] C. Schwieger, A. Meister, S. Daum, A. Blume, K. Bacia, *Polymers* **2017**, *9*, 612, DOI 10.3390/polym9110612.

[32] D. B. Iaea, I. Dikiy, I. Kiburu, D. Eliezer, F. R. Maxfield, *Biochemistry* **2015**, *54*, 4623–4636, DOI 10.1021/acs.biochem.5b00618.

[33] L. Dreßler, F. Michel, I. Thondorf, J. Mansfeld, R. Golbik, R. Ulbrich-Hofmann, *Phytochemistry* **2017**, *138*, 57–64, DOI 10.1016/j.phytochem.2017.02.024.

[34] M. A. Lemmon, *Nature Reviews Molecular Cell Biology* **2008**, *9*, 99–111, DOI 10.1038/nrm2328.

[35] Y. Hirano, Y.-G. Gao, D. J. Stephenson, N. T. Vu, L. Malinina, D. K. Simanshu, C. E. Chalfant, D. J. Patel, R. E. Brown, *eLife* **2019**, *8*, e44760, DOI 10.7554/elife.44760.

[36] K. M. Ferguson, M. A. Lemmon, J. Schlessinger, P. B. Sigler, *Cell* **1995**, *83*, 1037–1046, DOI 10.1016/0092-8674(95)90219-8.

[37] C. David, P. S. McPherson, O. Mundigl, P. de Camilli, *Proceedings of the National Academy of Sciences* **1996**, *93*, 331–335, DOI 10.1073/pnas.93.1.331.

[38] T. C. Südhof, *Annual Review of Neuroscience* **2004**, *27*, 509–547, DOI 10.1146/annurev.neuro.26.041002.131412.

[39] L. V. Chernomordik, M. M. Kozlov, *Nature Structural & Molecular Biology* **2008**, *15*, 675–683, DOI 10.1038/nsmb.1455.

[40] F. Li, D. Kümmel, J. Coleman, K. M. Reinisch, J. E. Rothman, F. Pincet, *Journal of the American Chemical Society* **2014**, *136*, 3456–3464, DOI 10.1021/ja410690m.

[41] J. Shin, X. Lou, D.-H. Kweon, Y.-K. Shin, *Biochemical Journal* **2014**, *459*, 95–102, DOI 10.1042/bj20131668.

[42] A. Stein, G. Weber, M. C. Wahl, R. Jahn, *Nature* **2009**, *460*, 525–528, DOI 10.1038/nature08156.

[43] M. Kyoung, A. Srivastava, Y. Zhang, J. Diao, M. Vrljic, P. Grob, E. Nogales, S. Chu, A. T. Brunger, *Proceedings of the National Academy of Sciences* **2011**, *108*, E304–E313, DOI 10.1073/pnas.1107900108.

[44] H. J. Risselada, A. Mayer, *Biochemical Journal* **2020**, *477*, 243–258, DOI 10.1042/bcj20190050.

[45] T. Trimbuch, C. Rosenmund, *Nature Reviews Neuroscience* **2016**, *17*, 118–125, DOI 10.1038/nrn.2015.16.

[46] J. Xu, T. Mashimo, T. C. Südhof, *Neuron* **2007**, *54*, 567–581, DOI 10.1016/j.neuron.2007.05.004.

[47] M. E. Fealey, B. P. Binder, V. N. Uversky, A. Hinderliter, D. D. Thomas, *Biophysical Journal* **2018**, *114*, 550–561, DOI 10.1016/j.bpj.2017.12.013.

[48] I. Fernandez, D. Araç, J. Ubach, S. H. Gerber, O.-H. Shin, Y. Gao, R. G. Anderson, T. C. Südhof, J. Rizo, *Neuron* **2001**, *32*, 1057–1069, DOI 10.1016/s0896-6273(01)00548-7.

[49] R. Sutton, B. A. Davletov, A. M. Berghuis, T. C. Südhof, S. R. Sprang, *Cell* **1995**, *80*, 929–938, DOI 10.1016/0092-8674(95)90296-1.

[50] J. Ubach, X. Zhang, X. Shao, T. C. Südhof, J. Rizo, *The EMBO Journal* **1998**, *17*, 3921–3930, DOI 10.1093/emboj/17.14.3921.

[51] D. Araç, X. Chen, H. A. Khant, J. Ubach, S. J. Ludtke, M. Kikkawa, A. E. Johnson, W. Chiu, T. C. Südhof, J. Rizo, *Nature Structural & Molecular Biology* **2006**, *13*, 209–217, DOI 10.1038/nsmb1056.

[52] M. Vrljic, P. Strop, R. C. Hill, K. C. Hansen, S. Chu, A. T. Brunger, *Biochemistry* **2011**, *50*, 9998–10012, DOI 10.1021/bi200998y.

[53] G. van den Bogaart, K. Meyenberg, U. Diederichsen, R. Jahn, *Journal of Biological Chemistry* **2012**, *287*, 16447–16453, DOI 10.1074/jbc.m112.343418.

[54] A. Pérez-Lara, A. Thapa, S. B. Nyenhuis, D. A. Nyenhuis, P. Halder, M. Tietzel, K. Tittmann, D. S. Cafiso, R. Jahn, *eLife* **2016**, *5*, e15886, DOI 10.7554/eLife.15886.

[55] A. Zhou, K. D. Brewer, J. Rizo, *Biochemistry* **2013**, *52*, 3446–3456, DOI 10.1021/bi400230u.

[56] S. B. Nyenhuis, N. Karandikar, V. Kiessling, A. J. B. Kreutzberger, A. Thapa, B. Liang, L. K. Tamm, D. S. Cafiso, *Nature Communications* **2021**, *12*, DOI 10.1038/s41467-021-21090-x.

[57] Y. Park, J.-K. Ryu, *FEBS Letters* **2018**, *592*, 3480–3492, DOI 10.1002/1873-3468.13193.

[58] Q. Zhou, P. Zhou, A. L. Wang, D. Wu, M. Zhao, T. C. Südhof, A. T. Brunger, *Nature* **2017**, *548*, 420–425, DOI 10.1038/nature23484.

[59] K. D. Brewer, T. Bacaj, A. Cavalli, C. Camilloni, J. D. Swarbrick, J. Liu, A. Zhou, P. Zhou, N. Barlow, J. Xu, A. B. Seven, E. A. Prinslow, R. Voleti, D. Häussinger, A. M. J. J. Bonvin, D. R. Tomchick, M. Vendruscolo, B. Graham, T. C. Südhof, J. Rizo, *Nature Structural & Molecular Biology* **2015**, *22*, 555–564, DOI 10.1038/nsmb.3035.

[60] Y. Park, J. B. Seo, A. Fraind, A. Pérez-Lara, H. Yavuz, K. Han, S.-R. Jung, I. Kattan, P. J. Walla, M. Choi, D. S. Cafiso, D.-S. Koh, R. Jahn, *Nature Structural & Molecular Biology* **2015**, *22*, 815–823, DOI 10.1038/nsmb.3097.

[61] S. Wang, Y. Li, C. Ma, *eLife* **2016**, *5*, e14211, DOI 10.7554/elife.14211.

[62] M. C. Chicka, E. Hui, H. Liu, E. R. Chapman, *Nature Structural & Molecular Biology* **2008**, *15*, 827.

[63] C. Li, B. Ullrich, J. Z. Zhang, R. G. W. Anderson, N. Brose, T. C. Südhof, *Nature* **1995**, *375*, 594–599, DOI 10.1038/375594a0.

[64] B. A. Davletov, T. C. Südhof, *Journal of Biological Chemistry* **1993**, *268*, 26386–26390.

[65] E. R. Chapman, A. F. Davis, *Journal of Biological Chemistry* **1998**, *273*, 13995–14001, DOI 10.1074/jbc.273.22.13995.

[66] D. Z. Herrick, S. Sterbling, K. A. Rasch, A. Hinderliter, D. S. Cafiso, *Biochemistry* **2006**, *45*, 9668–9674, DOI 10.1021/bi060874j.

[67] G. Schiavo, Q.-M. Gu, G. D. Prestwich, T. H. Sollner, J. E. Rothman, *Proceedings of the National Academy of Sciences* **1996**, *93*, 13327–13332, DOI 10.1073/pnas.93.23.13327.

[68] D. Murray, B. Honig, *Molecular Cell* **2002**, *9*, 145–154, DOI 10.1016/s1097-2765(01)00426-9.

[69] Y. Park, J. M. Hernandez, G. van den Bogaart, S. Ahmed, M. Holt, D. Riedel, R. Jahn, *Nature Structural & Molecular Biology* **2012**, *19*, 991–997, DOI 10.1038/nsmb.2375.

[70] X. Zhang, J. Rizo, T. C. Südhof, *Biochemistry* **1998**, *37*, 12395–12403, DOI 10.1021/bi9807512.

[71] J. Bai, W. C. Tucker, E. R. Chapman, *Nature Structural & Molecular Biology* **2003**, *11*, 36–44, DOI 10.1038/nsmb709.

[72] A. Radhakrishnan, A. Stein, R. Jahn, D. Fasshauer, *Journal of Biological Chemistry* **2009**, *284*, 25749–25760, DOI 10.1074/jbc.M109.042499.

[73] Z. Wang, H. Liu, Y. Gu, E. R. Chapman, *The Journal of Cell Biology* **2011**, *195*, 1159–1170, DOI 10.1083/jcb.201104079.

[74] E. Hui, C. P. Johnson, J. Yao, F. M. Dunning, E. R. Chapman, *Cell* **2009**, *138*, 709–721, DOI 10.1016/j.cell.2009.05.049.

[75] A. L. Lai, H. Huang, D. Z. Herrick, N. Epp, D. S. Cafiso, *Journal of Molecular Biology* **2011**, *405*, 696–706, DOI 10.1016/j.jmb.2010.11.015.

[76] A. B. Seven, K. D. Brewer, L. Shi, Q.-X. Jiang, J. Rizo, *Proceedings of the National Academy of Sciences* **2013**, *110*, E3243–E3252.

[77] D. Z. Herrick, W. Kuo, H. Huang, C. D. Schwieters, J. F. Ellena, D. S. Cafiso, *Journal of Molecular Biology* **2009**, *390*, 913–923, DOI 10.1016/j.jmb.2009.06.007.

[78] W. Vennekate, S. Schroder, C.-C. Lin, G. van den Bogaart, M. Grunwald, R. Jahn, P. J. Walla, *Proceedings of the National Academy of Sciences* **2012**, *109*, 11037–11042, DOI 10.1073/pnas.1116326109.

[79] M. N. Zanetti, O. D. Bello, J. Wang, J. Coleman, Y. Cai, C. V. Sindelar, J. E. Rothman, S. S. Krishnakumar, *eLife* **2016**, *5*, e17262, DOI 10.7554/elife.17262.

[80] C.-C. Lin, J. Seikowski, A. Pérez-Lara, R. Jahn, C. Höbartner, P. J. Walla, *Nature Communications* **2014**, *5*, DOI 10.1038/ncomms6859.

[81] N. G. Bologna, O. Voinnet, *Annual Review of Plant Biology* **2014**, *65*, 473–503, DOI 10.1146/annurev-arplant-050213-035728.

[82] I. R. Henderson, X. Zhang, C. Lu, L. Johnson, B. C. Meyers, P. J. Green, S. E. Jacobsen, *Nature Genetics* **2006**, *38*, 721–725, DOI 10.1038/ng1804.

[83] H. Vaucheret, *Trends in Plant Science* **2008**, *13*, 350–358, DOI 10.1016/j.tplants.2008.04.007.

[84] B. Yu, *Science* **2005**, *307*, 932–935, DOI 10.1126/science.1107130.

[85] T. Iki, M. Yoshikawa, M. Nishikiori, M. C. Jaudal, E. Matsumoto-Yokoyama, I. Mitsuhara, T. Meshi, M. Ishikawa, *Molecular Cell* **2010**, *39*, 282–291, DOI 10.1016/j.molcel.2010.05.014.

[86] T. Csorba, L. Kontra, J. Burgyán, *Virology* **2015**, *479-480*, 85–103, DOI 10.1016/j.virol.2015.02.028.

[87] T. Blevins, R. Rajeswaran, M. Aregger, B. K. Borah, M. Schepetilnikov, L. Baerlocher, L. Farinelli, F. Meins, T. Hohn, M. M. Pooggin, *Nucleic Acids Research* **2011**, *39*, 5003–5014, DOI 10.1093/nar/gkr119.

[88] Z. Mérai, Z. Kerényi, S. Kertész, M. Magna, L. Lakatos, D. Silhavy, *Journal of Virology* **2006**, *80*, 5747–5756, DOI 10.1128/jvi.01963-05.

[89] J. Azevedo, D. Garcia, D. Pontier, S. Ohnesorge, A. Yu, S. Garcia, L. Braun, M. Bergdoll, M. A. Hakimi, T. Lagrange, O. Voinnet, *Genes & Development* **2010**, *24*, 904–915, DOI 10.1101/gad.1908710.

[90] L.-Y. Wang, S.-S. Lin, T.-H. Hung, T.-K. Li, N.-C. Lin, T.-L. Shen, *Molecular Plant-Microbe Interactions* **2012**, *25*, 648–657, DOI 10.1094/mpmi-06-11-0155.

[91] T. Iki, M.-A. Tschopp, O. Voinnet, *RNA* **2017**, *23*, 639–654, DOI 10.1261/rna.060434.116.

[92] A. Deleris, J. Gallego-Bartolome, J. Bao, K. D. Kasschau, J. C. Carrington, O. Voinnet, *Science* **2006**, *313*, 68–71, DOI 10.1126/science.1128214.

[93] J. Hogle, A. Maeda, S. Harrison, *Journal of Molecular Biology* **1986**, *191*, 625–638, DOI 10.1016/0022-2836(86)90450-x.

[94] S. E. Bakker, R. J. Ford, A. M. Barker, J. Robottom, K. Saunders, A. R. Pearson, N. A. Ranson, P. G. Stockley, *Journal of Molecular Biology* **2012**, *417*, 65–78, DOI 10.1016/j.jmb.2012.01.017.

[95] M. Selbach, M. Mann, *Nature Methods* **2006**, *3*, 981–983, DOI 10.1038/nmeth972.

[96] A.-C. Gingras, M. Gstaiger, B. Raught, R. Aebersold, *Nature Reviews Molecular Cell Biology* **2007**, *8*, 645–654, DOI 10.1038/nrm2208.

[97] N. Hosoda, F. Lejeune, L. E. Maquat, *Molecular and Cellular Biology* **2006**, *26*, 3085–3097, DOI 10.1128/mcb.26.8.3085-3097.2006.

[98] J. Borch, P. Roepstorff, J. Møller-Jensen, *Molecular & Cellular Proteomics* **2011**, *10*, O110.006775, DOI 10.1074/mcp.o110.006775.

[99] N. J. Greenfield, *Nature Protocols* **2006**, *1*, 2876–2890, DOI 10.1038/nprot.2006.202.

[100] H. D. Mertens, D. I. Svergun, *Journal of Structural Biology* **2010**, *172*, 128–141, DOI 10.1016/j.jsb.2010.06.012.

[101] K. M. Yip, N. Fischer, E. Paknia, A. Chari, H. Stark, *Nature* **2020**, *587*, 157–161, DOI 10.1038/s41586-020-2833-4.

[102] K. Murata, M. Wolf, *Biochimica et Biophysica Acta (BBA) - General Subjects* **2018**, *1862*, 324–334, DOI 10.1016/j.bbagen.2017.07.020.

[103] W. Becker, K. C. Bhattiprolu, N. Gubensäk, K. Zangger, *ChemPhysChem* **2018**, *19*, 895–906, DOI 10.1002/cphc.201701253.

[104] A. M. Lau, A. Politis in *Multiprotein Complexes: Methods and Protocols*, (Ed.: A. Poterszman), Springer US, New York, NY, **2020**, pp. 221–241, DOI 10.1007/978-1-0716-1126-5_12.

[105] C. M. Whitehouse, R. N. Dreyer, M. Yamashita, J. B. Fenn, *Analytical Chemistry* **1985**, *57*, 675–679, DOI 10.1021/ac00280a023.

[106] M. Karas, D. Bachmann, U. Bahr, F. Hillenkamp, *International Journal of Mass Spectrometry and Ion Processes* **1987**, *78*, 53–68, DOI 10.1016/0168-1176(87)87041-6.

[107] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida, T. Matsuo, *Rapid Communications in Mass Spectrometry* **1988**, *2*, 151–153.

[108] A. Makarov, *Analytical Chemistry* **2000**, *72*, 1156–1162, DOI 10.1021/ac991131p.

[109] J. A. Loo, *Mass Spectrometry Reviews* **1997**, *16*, 1–23, DOI 10.1002/(SICI)1098-2787(1997)16:1<1::AID-MAS1>3.0.CO;2-L.

[110] M. Yamashita, J. B. Fenn, *The Journal of Physical Chemistry* **1984**, *88*, 4451–4459.

[111] R. Körner, M. Wilm, K. Morand, M. Schubert, M. Mann, *Journal of the American Society for Mass Spectrometry* **1996**, *7*, 150–156, DOI 10.1016/1044-0305(95)00626-5.

[112] M. Wilm, M. Mann, *Analytical Chemistry* **1996**, *68*, 1–8, DOI 10.1021/ac9509519.

[113] H. H. Hill, W. F. Siems, R. H. S. Louis, D. G. McMinn, *Analytical Chemistry* **1990**, *62*, 1201A–1209A, DOI 10.1021/ac00222a716.

[114] C. Hinz, S. Liggi, J. L. Griffin, *Current Opinion in Chemical Biology* **2018**, *42*, 42–50, DOI https://doi.org/10.1016/j.cbpa.2017.10.018.

[115] S. M. Dixit, D. A. Polasky, B. T. Ruotolo, *Current Opinion in Chemical Biology* **2018**, *42*, 93–100, DOI 10.1016/j.cbpa.2017.11.010.

[116] A. A. Shvartsburg, R. D. Smith, *Analytical Chemistry* **2008**, *80*, 9689–9699, DOI 10.1021/ac8016295.

[117] M. R. Wilkins, A. A. Gooley in *Proteome Research: New Frontiers in Functional Genomics*, (Eds.: M. R. Wilkins, K. L. Williams, R. D. Appel, D. F. Hochstrasser), Springer, Berlin, Heidelberg, **1997**, pp. 35–64, DOI 10.1007/978-3-662-03493-4_3.

[118] H. Steen, M. Mann, *Nature Reviews Molecular Cell Biology* **2004**, *5*, 699–711, DOI 10.1038/nrm1468.

[119] J. V. Olsen, S.-E. Ong, M. Mann, *Molecular & Cellular Proteomics* **2004**, *3*, 608–614, DOI 10.1074/mcp.t400003-mcp200.

[120] K. F. Medzihradszky, R. J. Chalkley, *Mass Spectrometry Reviews* **2013**, *34*, 43–63, DOI 10.1002/mas.21406.

[121] J. Cox, M. Mann, *Nature Biotechnology* **2008**, *26*, 1367–1372, DOI 10.1038/nbt.1511.

[122] E. Fahy, S. Subramaniam, R. C. Murphy, M. Nishijima, C. R. Raetz, T. Shimizu, F. Spener, G. van Meer, M. J. Wakelam, E. A. Dennis, *Journal of Lipid Research* **2009**, *50*, S9–S14, DOI 10.1194/jlr.r800095-jlr200.

[123] X. Han, R. W. Gross, *Mass Spectrometry Reviews* **2005**, *24*, 367–412, DOI 10.1002/mas.20023.

[124] B. Brügger, *Annual Review of Biochemistry* **2014**, *83*, 79–98, DOI 10.1146/annurev-biochem-060713-035324.

[125] C. Hu, Q. Duan, X. Han, *Proteomics* **2020**, *20*, 1900070, DOI https://doi.org/10.1002/pmic.201900070.

[126] E. Pyle, C. Guo, T. Hofmann, C. Schmidt, O. Ribiero, A. Politis, B. Byrne, *Analytical Chemistry* **2019**, *91*, 13071–13079, DOI 10.1021/acs.analchem.9b03271.

[127] Z. Zhang, D. L. Smith, *Protein Science* **1993**, *2*, 522–531, DOI 10.1002/pro.5560020404.

[128] G. R. Masson, J. E. Burke, N. G. Ahn, G. S. Anand, C. Borchers, S. Brier, G. M. Bou-Assaf, J. R. Engen, S. W. Englander, J. Faber, R. Garlish, P. R. Griffin, M. L. Gross, M. Guttman, Y. Hamuro, A. J. R. Heck, D. Houde, R. E. Iacob, T. J. D. Jørgensen, I. A. Kaltashov, J. P. Klinman, L. Konermann, P. Man, L. Mayne, B. D. Pascal, D. Reichmann, M. Skehel, J. Snijder, T. S. Strutzenberg, E. S. Underbakke, C. Wagner, T. E. Wales, B. T. Walters, D. D. Weis, D. J. Wilson, P. L. Wintrode, Z. Zhang, J. Zheng, D. C. Schriemer, K. D. Rand, *Nature Methods* **2019**, *16*, 595–602, DOI 10.1038/s41592-019-0459-y.

[129] H.-M. Zhang, G. M. Bou-Assaf, M. R. Emmett, A. G. Marshall, *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 520–524, DOI 10.1016/j.jasms.2008.11.010.

[130] G. Xu, M. R. Chance, *Chemical Reviews* **2007**, *107*, 3514–3543, DOI 10.1021/cr0682047.

[131] D. M. Hambly, M. L. Gross, *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 2057–2063, DOI 10.1016/j.jasms.2005.09.008.

[132] J. Chen, D. L. Rempel, B. C. Gau, M. L. Gross, *Journal of the American Chemical Society* **2012**, *134*, 18724–18731, DOI 10.1021/ja307606f.

[133] V. L. Mendoza, R. W. Vachet, *Mass Spectrometry Reviews* **2008**, *28*, 785–815, DOI 10.1002/mas.20203.

[134] M. Barth, J. Bender, T. Kundlacz, C. Schmidt, *Journal of Proteomics* **2020**, *222*, 103793, DOI 10.1016/j.jprot.2020.103793.

[135] T. Liu, T. M. Marcinko, P. A. Kiefer, R. W. Vachet, *Analytical Chemistry* **2017**, *89*, 11583–11591, DOI 10.1021/acs.analchem.7b02915.

[136] S. Mädler, C. Bich, D. Touboul, R. Zenobi, *Journal of Mass Spectrometry* **2009**, *44*, 694–706, DOI 10.1002/jms.1544.

[137] A. Leitner, L. A. Joachimiak, P. Unverdorben, T. Walzthoeni, J. Frydman, F. Förster, R. Aebersold, *Proceedings of the National Academy of Sciences* **2014**, *111*, 9455–9460, DOI 10.1073/pnas.1320298111.

[138] A. Belsom, G. Mudd, S. Giese, M. Auer, J. Rappsilber, *Analytical Chemistry* **2017**, *89*, 5319–5324, DOI 10.1021/acs.analchem.6b04938.

[139] J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proceedings of the National Academy of Sciences* **2002**, *99*, 11020–11024, DOI 10.1073/pnas.172226299.

[140] M. Suchanek, A. Radzikowska, C. Thiele, *Nature Methods* **2005**, *2*, 261–268, DOI 10.1038/NMETH752.

[141] A. Leitner, R. Reischl, T. Walzthoeni, F. Herzog, S. Bohn, F. Förster, R. Aebersold, *Molecular & Cellular Proteomics* **2012**, *11*, M111.014126, DOI 10.1074/mcp.m111.014126.

[142] O. Rinner, J. Seebacher, T. Walzthoeni, L. N. Mueller, M. Beck, A. Schmidt, M. Mueller, R. Aebersold, *Nature Methods* **2008**, *5*, 315–318, DOI 10.1038/nmeth.1192.

[143] B. Steigenberger, R. J. Pieters, A. J. R. Heck, R. A. Scheltema, *ACS Central Science* **2019**, *5*, 1514–1522, DOI 10.1021/acscentsci.9b00416.

[144] Z.-L. Chen, J.-M. Meng, Y. Cao, J.-L. Yin, R.-Q. Fang, S.-B. Fan, C. Liu, W.-F. Zeng, Y.-H. Ding, D. Tan, L. Wu, W.-J. Zhou, H. Chi, R.-X. Sun, M.-Q. Dong, S.-M. He, *Nature Communications* **2019**, *10*, DOI 10.1038/s41467-019-11337-z.

[145] M. Götze, J. Pettelkau, S. Schaks, K. Bosse, C. H. Ihling, F. Krauth, R. Fritzsche, U. Kühn, A. Sinz, *Journal of the American Society for Mass Spectrometry* **2011**, *23*, 76–87, DOI 10.1007/s13361-011-0261-2.

[146] A. Leitner, T. Walzthoeni, R. Aebersold, *Nature Protocols* **2013**, *9*, 120–137, DOI 10.1038/nprot.2013.168.

[147] C. W. Combe, L. Fischer, J. Rappsilber, *Molecular & Cellular Proteomics* **2015**, *14*, 1137–1147, DOI 10.1074/mcp.o114.042259.

[148] J. Kosinski, A. von Appen, A. Ori, K. Karius, C. W. Müller, M. Beck, *Journal of Structural Biology* **2015**, *189*, 177–183, DOI 10.1016/j.jsb.2015.01.014.

[149] F. Liu, D. T. S. Rijkers, H. Post, A. J. R. Heck, *Nature Methods* **2015**, *12*, 1179–1184, DOI 10.1038/nmeth.3603.

[150] Z. A. Chen, J. Rappsilber, *Nature Protocols* **2018**, *14*, 171–201, DOI 10.1038/s41596-018-0089-3.

[151] F. Liu, P. Lössl, B. M. Rabbitts, R. S. Balaban, A. J. Heck, *Molecular & Cellular Proteomics* **2018**, *17*, 216–232, DOI 10.1074/mcp.ra117.000470.

[152] S. Wittig, M. Ganzella, M. Barth, S. Kostmann, D. Riedel, Á. Pérez-Lara, R. Jahn, C. Schmidt, *Nature Communications* **2021**, *12*, DOI 10.1038/s41467-021-21102-w.

[153] A. Kao, C.-L. Chiu, D. Vellucci, Y. Yang, V. R. Patel, S. Guan, A. Randall, P. Baldi, S. D. Rychnovsky, L. Huang, *Molecular & Cellular Proteomics* **2011**, *10*, M110.002170, DOI 10.1074/mcp.m110.002212.

[154] C. Iacobucci, M. Götze, C. H. Ihling, C. Piotrowski, C. Arlt, M. Schäfer, C. Hage, R. Schmidt, A. Sinz, *Nature Protocols* **2018**, *13*, 2864–2889, DOI 10.1038/s41596-018-0068-8.

[155] C. Schmidt, H. Urlaub, *Current Opinion in Structural Biology* **2017**, *46*, 157–168, DOI 10.1016/j.sbi.2017.10.005.

[156] S. V. Faull, A. M. C. Lau, C. Martens, Z. Ahdash, K. Hansen, H. Yebenes, C. Schmidt, F. Beuron, N. B. Cronin, E. P. Morris, A. Politis, *Nature communications* **2019**, *10*, 3814, DOI 10.1038/s41467-019-11772-y.

[157] C. Schmidt, M. Zhou, H. Marriott, N. Morgner, A. Politis, C. V. Robinson, *Nature Communications* **2013**, *4*, DOI 10.1038/ncomms2985.

[158] M. Hafner, M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano, A.-C. Jungkamp, M. Munschauer, A. Ulrich, G. S. Wardle, S. Dewell, M. Zavolan, T. Tuschl, *Cell* **2010**, *141*, 129–141, DOI 10.1016/j.cell.2010.03.009.

[159] A. Stützer, L. M. Welp, M. Raabe, T. Sachsenberg, C. Kappert, A. Wulf, A. M. Lau, S.-S. David, A. Chernev, K. Kramer, A. Politis, O. Kohlbacher, W. Fischle, H. Urlaub, *Nature Communications* **2020**, *11*, DOI 10.1038/s41467-020-19047-7.

[160] K. Kramer, T. Sachsenberg, B. M. Beckmann, S. Qamar, K.-L. Boon, M. W. Hentze, O. Kohlbacher, H. Urlaub, *Nature Methods* **2014**, *11*, 1064–1070, DOI 10.1038/nmeth.3092.

[161] A. Scacioc, C. Schmidt, T. Hofmann, H. Urlaub, K. Kühnel, Á. Pérez-Lara, *Scientific Reports* **2017**, *7*, 14008, DOI 10.1038/s41598-017-14337-5.

[162] B. Oberfeld, J. Brunner, P. Dimroth, *Biochemistry* **2006**, *45*, 1841–1851, DOI 10.1021/bi052304+.

[163] S. Lindner, K. Gruhle, R. Schmidt, V. M. Garamus, D. Ramsbeck, G. Hause, A. Meister, A. Sinz, S. Drescher, *Langmuir* **2017**, *33*, 4960–4973, DOI 10.1021/acs.langmuir.7b00228.

[164] J. Gubbens, E. Ruijter, L. E. de Fays, J. M. A. Damen, B. de Kruijff, M. Slijper, D. T. Rijkers, R. M. Liskamp, A. I. de Kroon, *Chemistry & Biology* **2009**, *16*, 3–14, DOI 10.1016/j.chembiol.2008.11.009.

[165] L. Konermann, *Journal of the American Society for Mass Spectrometry* **2017**, *28*, 1827–1835, DOI 10.1007/s13361-017-1739-3.

[166] H. Hernández, C. V. Robinson, *Nature Protocols* **2007**, *2*, 715–726, DOI 10.1038/nprot.2007.73.

[167] K. Muneeruddin, J. J. Thomas, P. A. Salinas, I. A. Kaltashov, *Analytical Chemistry* **2014**, *86*, 10692–10699, DOI 10.1021/ac502590h.

[168] J. F. de la Mora, *Analytica Chimica Acta* **2000**, *406*, 93–104, DOI 10.1016/s0003-2670(99)00601-7.

[169] C. Santambrogio, A. Natalello, S. Brocca, E. Ponzini, R. Grandori, *Proteomics* **2018**, *19*, 1800060, DOI 10.1002/pmic.201800060.

[170] F. Sobott, H. Hernández, M. G. McCammon, M. A. Tito, C. V. Robinson, *Analytical Chemistry* **2002**, *74*, 1402–1407, DOI 10.1021/ac0110552.

[171] N. Morgner, C. V. Robinson, *Analytical Chemistry* **2012**, *84*, 2939–2948, DOI 10.1021/ac300056a.

[172] M. T. Marty, A. J. Baldwin, E. G. Marklund, G. K. A. Hochberg, J. L. P. Benesch, C. V. Robinson, *Analytical Chemistry* **2015**, *87*, 4370–4376, DOI 10.1021/acs.analchem.5b00140.

[173] J. L. P. Benesch, B. T. Ruotolo, D. A. Simmons, C. V. Robinson, *Chemical Reviews* **2007**, *107*, 3544–3567, DOI 10.1021/cr068289b.

[174] M. Zhou, N. Morgner, N. P. Barrera, A. Politis, S. C. Isaacson, D. Matak-Vinković, T. Murata, R. A. Bernal, D. Stock, C. V. Robinson, *Science* **2011**, *334*, 380–385, DOI 10.1126/science.1210148.

[175] A. Laganowsky, E. Reading, J. T. S. Hopper, C. V. Robinson, *Nature Protocols* **2013**, *8*, 639–51, DOI 10.1038/nprot.2013.024.

[176] E. Reading, I. Liko, T. M. Allison, J. L. P. Benesch, A. Laganowsky, C. V. Robinson, *Angewandte Chemie International Edition* **2015**, *54*, 4577–4581, DOI 10.1002/anie.201411622.

[177] M. Landreh, J. Costeira-Paulo, J. Gault, E. G. Marklund, C. V. Robinson, *Analytical Chemistry* **2017**, *89*, 7425–7430, DOI 10.1021/acs.analchem.7b00922.

[178] K. Gupta, J. Li, I. Liko, J. Gault, C. Bechara, D. Wu, J. T. S. Hopper, K. Giles, J. L. P. Benesch, C. V. Robinson, *Nature Protocols* **2018**, *13*, 1106–1120, DOI 10.1038/nprot.2018.014.

[179] A. Konijnenberg, A. Butterer, F. Sobott, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2013**, *1834*, 1239–1256, DOI 10.1016/j.bbapap.2012.11.013.

[180] A. A. Shvartsburg, M. F. Jarrold, *Chemical Physics Letters* **1996**, *261*, 86–91, DOI 10.1016/0009-2614(96)00941-4.

[181] E. G. Marklund, M. T. Degiacomi, C. V. Robinson, A. J. Baldwin, J. L. Benesch, *Structure* **2015**, *23*, 791–799, DOI 10.1016/j.str.2015.02.010.

[182] E. Jurneczko, P. E. Barran, *The Analyst* **2011**, *136*, 20–28, DOI 10.1039/c0an00373e.

[183] Z. Hall, A. Politis, C. V. Robinson, *Structure* **2012**, *20*, 1596–1609, DOI 10.1016/j.str.2012.07.001.

[184] J. D. Eschweiler, M. A. Farrugia, S. M. Dixit, R. P. Hausinger, B. T. Ruotolo, *Structure* **2018**, *26*, 599–606.e3, DOI 10.1016/j.str.2018.03.001.

[185] K. Lindorff-Larsen, N. Trbovic, P. Maragakis, S. Piana, D. E. Shaw, *Journal of the American Chemical Society* **2012**, *134*, 3787–3791, DOI 10.1021/ja209931w.

[186] C. Bae, C. Anselmi, J. Kalia, A. Jara-Oseguera, C. D. Schwieters, D. Krepkiy, C. W. Lee, E.-H. Kim, J. I. Kim, J. D. Faraldo-Gómez, K. J. Swartz, *eLife* **2016**, *5*, e11273, DOI 10.7554/elife.11273.

[187] A. D. Mackerell Jr., *Journal of Computational Chemistry* **2004**, *25*, 1584–1604, DOI https://doi.org/10.1002/jcc.20082.

[188] T. Darden, D. York, L. Pedersen, *The Journal of Chemical Physics* **1993**, *98*, 10089–10092, DOI 10.1063/1.464397.

[189] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *Journal of Computational Chemistry* **1997**, *18*, 1463–1472, DOI 10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.0.co;2-h.

[190] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *The Journal of Chemical Physics* **1984**, *81*, 3684–3690, DOI 10.1063/1.448118.

[191] S. Nosé, *The Journal of Chemical Physics* **1984**, *81*, 511–519, DOI 10.1063/1.447334.

[192] W. G. Hoover, *Physical Review A* **1985**, *31*, 1695–1697, DOI 10.1103/physreva.31.1695.

[193] M. Parrinello, A. Rahman, *Journal of Applied Physics* **1981**, *52*, 7182–7190, DOI 10.1063/1.328693.

[194] R. C. Bernardi, M. C. Melo, K. Schulten, *Biochimica et Biophysica Acta (BBA) - General Subjects* **2015**, *1850*, 872–877, DOI 10.1016/j.bbagen.2014.10.019.

[195] M. Landreh, E. G. Marklund, P. Uzdavinys, M. T. Degiacomi, M. Coincon, J. Gault, K. Gupta, I. Liko, J. L. P. Benesch, D. Drew, C. V. Robinson, *Nature Communications* **2017**, *8*, 13993, DOI 10.1038/ncomms13993.

[196] S. Jo, T. Kim, V. G. Iyer, W. Im, *Journal of Computational Chemistry* **2008**, *29*, 1859–1865, DOI 10.1002/jcc.20945.

[197] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX* **2015**, *1-2*, 19–25, DOI 10.1016/j.softx.2015.06.001.

[198] G. van Zundert, J. Rodrigues, M. Trellet, C. Schmitz, P. Kastritis, E. Karaca, A. Melquiond, M. van Dijk, S. de Vries, A. Bonvin, *Journal of Molecular Biology* **2016**, *428*, 720–725, DOI 10.1016/j.jmb.2015.09.014.

[199] E. Barreto-Ojeda, MDAnalysis Membrane Curvature Tool, **2021**, DOI 10.5281/ZENODO.5553452.

[200] T. M. Allison, E. Reading, I. Liko, A. J. Baldwin, A. Laganowsky, C. V. Robinson, *Nature Communications* **2015**, *6*, DOI 10.1038/ncomms9551.

[201] D. J. Mandell, E. A. Coutsias, T. Kortemme, *Nature Methods* **2009**, *6*, 551–552, DOI 10.1038/nmeth0809-551.

[202] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, *Nature Methods* **2020**, *17*, 261–272, DOI 10.1038/s41592-019-0686-2.

[203] G. Nagy, M. Igaev, N. C. Jones, S. V. Hoffmann, H. Grubmüller, *Journal of Chemical Theory and Computation* **2019**, *15*, 5087–5102, DOI 10.1021/acs.jctc.9b00203.

[204] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *Journal of Computational Chemistry* **2004**, *25*, 1605–1612, DOI 10.1002/jcc.20084.

[205] W. Humphrey, A. Dalke, K. Schulten, *Journal of Molecular Graphics* **1996**, *14*, 33–38, DOI 10.1016/0263-7855(96)00018-5.

[206] A. Kahraman, L. Malmström, R. Aebersold, *Bioinformatics* **2011**, *27*, 2163–2164, DOI 10.1093/bioinformatics/btr348.

[207] S. M. Kelly, T. J. Jess, N. C. Price, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2005**, *1751*, 119–139, DOI 10.1016/j.bbapap.2005.06.005.

[208] A. Y. Lyubimov, M. Uervirojnangkoorn, O. B. Zeldin, Q. Zhou, M. Zhao, A. S. Brewster, T. Michels-Clark, J. M. Holton, N. K. Sauter, W. I. Weis, A. T. Brunger, *eLife* **2016**, *5*, e18740, DOI 10.7554/elife.18740.

[209] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, *Nature Methods* **2012**, *9*, 676–682, DOI 10.1038/nmeth.2019.

[210] X. Liu, A. B. Seven, J. Xu, V. Esser, L. Su, C. Ma, J. Rizo, *Nature Protocols* **2017**, *12*, 2014–2028, DOI 10.1038/nprot.2017.068.

[211] A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen, M. Mann, *Nature Protocols* **2006**, *1*, 2856–2860, DOI 10.1038/nprot.2006.468.

[212] Y.-Q. Yu, M. Gilar, P. J. Lee, E. S. P. Bouvier, J. C. Gebler, *Analytical Chemistry* **2003**, *75*, 6023–6028, DOI 10.1021/ac0346196.

[213] J. Bender, C. Schmidt, *Bioinformatics* **2019**, *36*, 1296–1297, DOI 10.1093/bioinformatics/btz732.

[214] C. Iacobucci, A. Sinz, *Analytical Chemistry* **2017**, *89*, 7832–7835, DOI 10.1021/acs.analchem.7b02316.

[215] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, A. D. MacKerell, *Nature Methods* **2016**, *14*, 71–73, DOI 10.1038/nmeth.4067.

[216] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks, A. D. MacKerell, J. B. Klauda, W. Im, *Journal of Chemical Theory and Computation* **2015**, *12*, 405–413, DOI 10.1021/acs.jctc.5b00935.

[217] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, O. Beckstein, *Journal of Computational Chemistry* **2011**, *32*, 2319–2327, DOI 10.1002/jcc.21787.

[218] J. Bender, C. Schmidt, *Biological Chemistry* **2019**, *400*, 813–829, DOI doi:10.1515/hsz-2018-0443.

[219] T. Kundlacz, J. Bender, C. Schmidt, *International Journal of Mass Spectrometry* **2021**, *468*, 116652, DOI 10.1016/j.ijms.2021.116652.

[220] I. Hasni, P. Bourassa, H.-A. Tajmir-Riahi, *The Journal of Physical Chemistry B* **2011**, *115*, 6683–6690, DOI 10.1021/jp200045h.

[221] H. Metwally, R. G. McAllister, V. Popa, L. Konermann, *Analytical Chemistry* **2016**, *88*, 5345–5354, DOI 10.1021/acs.analchem.6b00650.

[222] H. J. Sterling, A. F. Kintzer, G. K. Feld, C. A. Cassou, B. A. Krantz, E. R. Williams, *Journal of the American Society for Mass Spectrometry* **2011**, *23*, 191–200, DOI 10.1007/s13361-011-0301-y.

[223] M. F. Perutz, M. G. Rosssmann, A. F. Cullis, H. Murihead, W. G, A. C. North, *Nature* **1960**, *185*, 416–422, DOI 10.1038/185416a0.

[224] P. Martins, F. Gomes, W. Vaz, M. Moreno, *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2008**, *1778*, 1308–1315, DOI https://doi.org/10.1016/j.bbamem.2008.02.011.

[225] U. Heindel, M. F. Schmidt, M. Veit, *FEBS Letters* **2003**, *544*, 57–62, DOI https://doi.org/10.1016/S0014-5793(03)00449-6.

[226] M. Fukuda, E. Kanno, Y. Ogata, K. Mikoshiba, *Journal of Biological Chemistry* **2001**, *276*, 40319–40325, DOI 10.1074/jbc.m105356200.

[227] K. F. Geoghegan, H. B. Dixon, P. J. Rosner, L. R. Hoth, A. J. Lanzetti, K. A. Borzilleri, E. S. Marr, L. H. Pezzullo, L. B. Martin, P. K. LeMotte, A. S. McColl, A. V. Kamath, J. G. Stroh, *Analytical Biochemistry* **1999**, *267*, 169–184, DOI https://doi.org/10.1006/abio.1998.2990.

[228] A. Honigmann, G. van den Bogaart, E. Iraheta, H. J. Risselada, D. Milovanovic, V. Mueller, S. Müllar, U. Diederichsen, D. Fasshauer, H. Grubmüller, S. W. Hell, C. Eggeling, K. Kühnel, R. Jahn, *Nature Structural & Molecular Biology* **2013**, *20*, 679–686, DOI 10.1038/nsmb.2570.

[229] S. Dowler, G. Kular, D. R. Alessi, *Science's STKE* **2002**, *2002*, pl6, DOI 10.1126/stke.2002.129.pl6.

[230] C. L. Hunter, A. G. Mauk, D. J. Douglas, *Biochemistry* **1997**, *36*, 1018–1025, DOI 10.1021/bi961993+.

[231] A. Laganowsky, E. Reading, T. M. Allison, M. B. Ulmschneider, M. T. Degiacomi, A. J. Baldwin, C. V. Robinson, *Nature* **2014**, *510*, 172–175, DOI 10.1038/nature13419.

[232] C. Schmidt, J. A. Macpherson, A. M. Lau, K. W. Tan, F. Fraternali, A. Politis, *Analytical Chemistry* **2017**, *89*, 1459–1468, DOI 10.1021/acs.analchem.6b02875.

[233] L. Xia, Z. Ma, J. Tong, Y. Tang, S. Li, S. Qin, R. Lou, S. Zhao, X. Lei, W. Shui, *Analytica Chimica Acta* **2020**, *1102*, 53–62, DOI 10.1016/j.aca.2019.12.036.

[234] E. Karaca, A. S. Melquiond, S. J. de Vries, P. L. Kastritis, A. M. Bonvin, *Molecular & Cellular Proteomics* **2010**, *9*, 1784–1794, DOI 10.1074/mcp.m000051-mcp201.

[235] S. R. K. Ainavarapu, J. Brujić, H. H. Huang, A. P. Wiita, H. Lu, L. Li, K. A. Walther, M. Carrion-Vazquez, H. Li, J. M. Fernandez, *Biophysical Journal* **2007**, *92*, 225–233, DOI 10.1529/biophysj.106.091561.

[236] R. Sorkin, M. Marchetti, E. Logtenberg, M. C. Piontek, E. Kerklingh, G. Brand, R. Voleti, J. Rizo, W. H. Roos, A. J. Groffen, G. J. Wuite, *Biophysical Journal* **2019**, *118*, 543–656, DOI 10.1016/j.bpj.2019.12.021.

[237] M. Frick, C. Schwieger, C. Schmidt, *Angewandte Chemie International Edition* **2021**, *60*, 11523–11530, DOI 10.1002/anie.202101242.

[238] E. Reading, T. A. Walton, I. Liko, M. T. Marty, A. Laganowsky, D. C. Rees, C. V. Robinson, *Chemistry & Biology* **2015**, *22*, 593–603, DOI 10.1016/j.chembiol.2015.04.016.

[239] M. Kaldmäe, N. Österlund, D. Lianoudaki, C. Sahin, P. Bergman, T. Nyman, N. Kronqvist, L. L. Ilag, T. M. Allison, E. G. Marklund, M. Landreh, *Journal of the American Society for Mass Spectrometry* **2019**, *30*, 1385–1388, DOI 10.1007/s13361-019-02177-8.

[240] A. T. Iavarone, E. R. Williams, *Journal of the American Chemical Society* **2003**, *125*, 2319–2327, DOI 10.1021/ja021202t.

[241] J. V. Iribarne, *The Journal of Chemical Physics* **1976**, *64*, 2287, DOI 10.1063/1.432536.

[242] K. Shinoda, T. Yamaguchi, R. Hori, *Bulletin of the Chemical Society of Japan* **1961**, *34*, 237–241, DOI 10.1246/bcsj.34.237.

[243] S. H. Lomeli, I. X. Peng, S. Yin, R. R. O. Loo, J. A. Loo, *Journal of the American Society for Mass Spectrometry* **2010**, *21*, 127–131, DOI 10.1016/j.jasms.2009.09.014.

[244] H. J. Sterling, M. P. Daly, G. K. Feld, K. L. Thoren, A. F. Kintzer, B. A. Krantz, E. R. Williams, *Journal of the American Society for Mass Spectrometry* **2010**, *21*, 1762–1774, DOI 10.1016/j.jasms.2010.06.012.

[245] L. Konermann, H. Metwally, Q. Duez, I. Peters, *The Analyst* **2019**, *144*, 6157–6171, DOI 10.1039/c9an01201j.

[246] M. I. Catalina, R. H. H. van den Heuvel, E. van Duijn, A. J. R. Heck, *Chemistry - A European Journal* **2005**, *11*, 960–968, DOI 10.1002/chem.200400395.

[247] J. Lyu, Y. Liu, J. W. McCabe, S. Schrecke, L. Fang, D. H. Russell, A. Laganowsky, *Analytical Chemistry* **2020**, *92*, 11242–11249, DOI 10.1021/acs.analchem.0c01826.

[248] A. J. Borysik, D. J. Hewitt, C. V. Robinson, *Journal of the American Chemical Society* **2013**, *135*, 6078–6083, DOI 10.1021/ja401736v.

[249] P. Kebarle, L. Tang, *Analytical Chemistry* **1993**, *65*, 972A–986A, DOI 10.1021/ac00070a001.

[250] M. Zulauf, J. Rosenbusch, *The Journal of Physical Chemistry* **1983**, *87*, 856–862, DOI 10.1021/j100228a032.

[251] J. Wang, F. Li, O. D. Bello, C. V. Sindelar, F. Pincet, S. S. Krishnakumar, J. E. Rothman, *eLife* **2017**, *6*, e27441, DOI 10.7554/elife.27441.

[252] B. G. Wilhelm, S. Mandad, S. Truckenbrodt, K. Kröhnert, C. Schäfer, B. Rammner, S. J. Koo, G. A. Claßen, M. Krauss, V. Haucke, H. Urlaub, S. O. Rizzoli, *Science* **2014**, *344*, 1023–1028, DOI 10.1126/science.1252884.

[253] S. Takamori, M. Holt, K. Stenius, E. A. Lemke, M. Grønborg, D. Riedel, H. Urlaub, S. Schenck, B. Brügger, P. Ringler, S. A. Müller, B. Rammner, F. Gräter, J. S. Hub, B. L. D. Groot, G. Mieskes, Y. Moriyama, J. Klingauf, H. Grubmüller, J. Heuser, F. Wieland, R. Jahn, *Cell* **2006**, *127*, 831–846, DOI 10.1016/j.cell.2006.10.030.

[254] E. Chapman, R. Jahn, *Journal of Biological Chemistry* **1994**, *269*, 5735–5741, DOI 10.1016/s0021-9258(17)37523-3.

[255] M. R. Wenk, P. D. Camilli, *Proceedings of the National Academy of Sciences* **2004**, *101*, 8262–8269, DOI 10.1073/pnas.0401874101.

[256] Y. J. Kim, M.-L. Guzman-Hernandez, E. Wisniewski, T. Balla, *Developmental Cell* **2015**, *33*, 549–561, DOI 10.1016/j.devcel.2015.04.028.

[257] S. Xiao, C. V. Finkielstein, D. G. S. Capelluto in *Lipid-Mediated Protein Signaling*, (Ed.: D. G. Capelluto), Springer Netherlands, Dordrecht, **2013**, pp. 27–40, DOI 10.1007/978-94-007-6331-9_3.

[258] G. van den Bogaart, K. Meyenberg, H. J. Risselada, H. Amin, K. I. Willig, B. E. Hubrich, M. Dier, S. W. Hell, H. Grubmüller, U. Diederichsen, R. Jahn, *Nature* **2011**, *479*, 552–555, DOI 10.1038/nature10545.

[259] G. D. Paolo, P. D. Camilli, *Nature* **2006**, *443*, 651–657, DOI 10.1038/nature05185.

[260] J. L. P. Benesch, C. V. Robinson, *Nature* **2009**, *462*, 576–577, DOI 10.1038/462576a.

[261] L. Liu, D. Bagal, E. N. Kitova, P. D. Schnier, J. S. Klassen, *Journal of the American Chemical Society* **2009**, *131*, 15980–15981, DOI 10.1021/ja9060454.

[262] K. Sokratous, L. V. Roach, D. Channing, J. Strachan, J. Long, M. S. Searle, R. Layfield, N. J. Oldham, *Journal of the American Chemical Society* **2012**, *134*, 6416–6424, DOI 10.1021/ja300749d.

[263] J. T. S. Hopper, C. V. Robinson, *Angewandte Chemie International Edition* **2014**, *53*, 14002–14015, DOI 10.1002/anie.201403741.

[264] L. Michaeli, I. Gottfried, M. Bykhovskaia, U. Ashery, *Traffic* **2017**, *18*, 825–839, DOI 10.1111/tra.12528.

[265] M. Bykhovskaia, *Biophysical journal* **2021**, *120*, 642–661, DOI 10.1016/j.bpj.2020.12.025.

[266] R. Jahn, T. C. Sudhof, *Annual Review of Neuroscience* **1994**, *17*, 219–246, DOI 10.1146/annurev.ne.17.030194.001251.

[267] C.-L. Lai, K. E. Landgraf, G. A. Voth, J. J. Falke, *Journal of Molecular Biology* **2010**, *402*, 301–310, DOI 10.1016/j.jmb.2010.07.037.

[268] N. L. Chon, J. R. Osterberg, J. Henderson, H. M. Khan, N. Reuter, J. D. Knight, H. Lin, *Biochemistry* **2015**, *54*, 5696–5711, DOI 10.1021/acs.biochem.5b00422.

[269] A. A. Frazier, C. R. Roller, J. J. Havelka, A. Hinderliter, D. S. Cafiso, *Biochemistry* **2002**, *42*, 96–105, DOI 10.1021/bi0268145.

[270] E. Rufener, A. A. Frazier, C. M. Wieser, A. Hinderliter, D. S. Cafiso, *Biochemistry* **2004**, *44*, 18–28, DOI 10.1021/bi048370d.

[271] Z. Wu, K. Schulten, *Biophysical Journal* **2014**, *107*, 1156–1166, DOI 10.1016/j.bpj.2014.07.041.

[272] L. Li, O.-H. Shin, J.-S. Rhee, D. Araç, J.-C. Rah, J. Rizo, T. Südhof, C. Rosenmund, *Journal of Biological Chemistry* **2006**, *281*, 15845–15852, DOI 10.1074/jbc.m600888200.

[273] J. Guillen, C. Ferrer-Orta, M. Buxaderas, D. Perez-Sanchez, M. Guerrero-Valero, G. Luengo-Gil, J. Pous, P. Guerra, J. C. Gomez-Fernandez, N. Verdaguer, S. Corbalan-Garcia, *Proceedings of the National Academy of Sciences* **2013**, *110*, 20503–20508, DOI 10.1073/pnas.1316179110.

[274] J. S. Golden, S. C. Harrison, *Biochemistry* **1982**, *21*, 3862–3866.

[275] M. Fioramonte, A. M. dos Santos, S. McIIwain, W. S. Noble, K. G. Franchini, F. C. Gozzo, *Proteomics* **2012**, *12*, 2746–2752, DOI 10.1002/pmic.201200040.

[276] T. Schlick, E. Barth, M. Mandziuk, *Annual Review of Biophysics and Biomolecular Structure* **1997**, *26*, 181–222, DOI 10.1146/annurev.biophys.26.1.181.

[277] N.-A. Lakomek, H. Yavuz, R. Jahn, Á. Pérez-Lara, *Proceedings of the National Academy of Sciences* **2019**, *116*, 8699–8708, DOI 10.1073/pnas.1813194116.

[278] F. Seiler, J. Malsam, J. M. Krause, T. H. Söllner, *FEBS Letters* **2009**, *583*, 2343–2348, DOI 10.1016/j.febslet.2009.06.025.

[279] A. Laio, M. Parrinello, *Proceedings of the National Academy of Sciences* **2002**, *99*, 12562–12566, DOI 10.1073/pnas.202427399.

[280] P. J. Stansfeld, M. S. Sansom, *Journal of Chemical Theory and Computation* **2011**, *7*, 1157–1166, DOI 10.1021/ct100569y.

[281] G. Lindblom, G. Orädd, *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2009**, *1788*, 234–244, DOI 10.1016/j.bbamem.2008.08.016.

[282] J. L. Baylon, J. V. Vermaas, M. P. Muller, M. J. Arcario, T. V. Pogorelov, E. Tajkhorshid, *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2016**, *1858*, 1573–1583, DOI 10.1016/j.bbamem.2016.02.027.

[283] W. Pezeshkian, S. J. Marrink, *Current Opinion in Cell Biology* **2021**, *71*, 103–111, DOI 10.1016/j.ceb.2021.02.009.

[284] M. T. Degiacomi, C. Schmidt, A. J. Baldwin, J. L. Benesch, *Structure* **2017**, *25*, 1751–1757.e5, DOI 10.1016/j.str.2017.08.015.

[285] T. Taverner, N. E. Hall, A. Richard, R. J. Simpson, *Journal of Biological Chemistry* **2002**, *277*, 46487–46492.

[286] T. C. Erren, J. V. Groß, U. Wild, P. Lewis, D. M. Shaw, *EMBO reports* **2016**, *18*, 18–20, DOI 10.15252/embr.201643618.

Supplementary Material

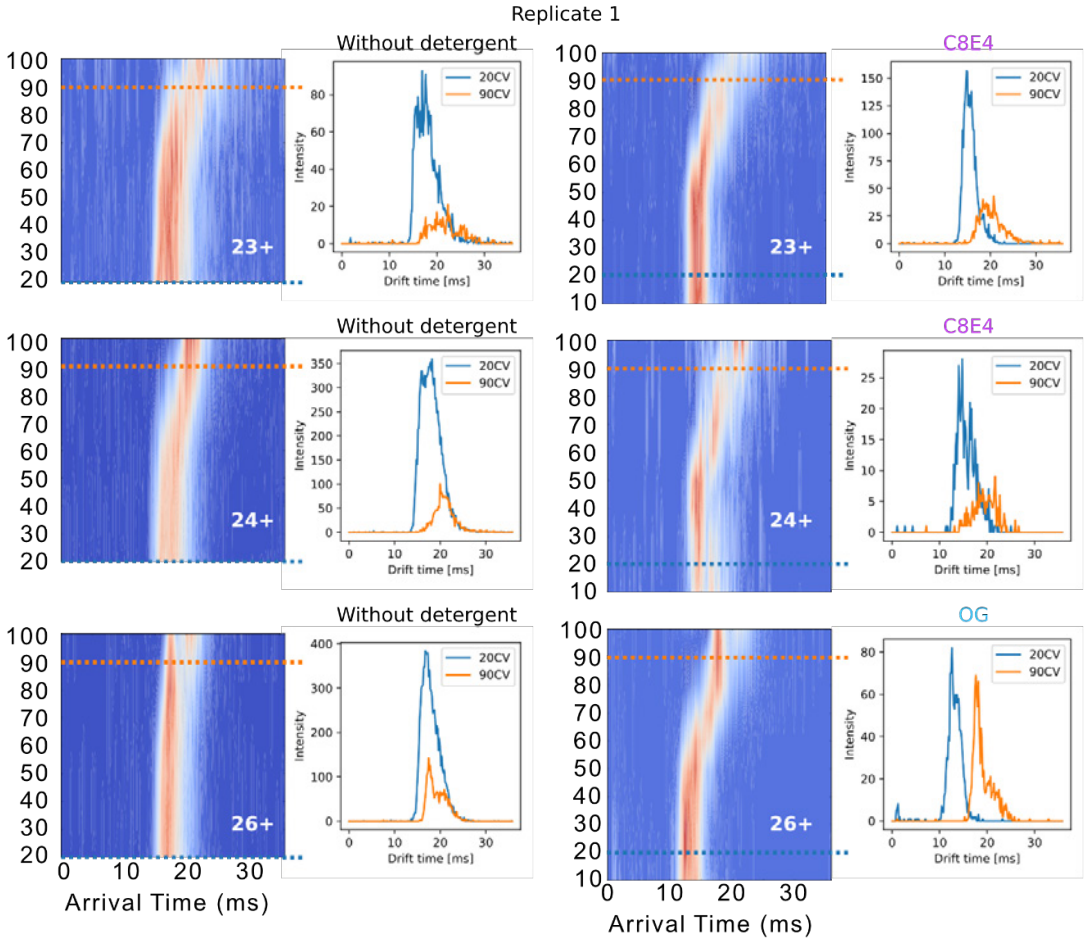


Figure S1: Comparison of CIU of the Same Charge States of ADH Observed in the Absence and in the Presence of C8E4 or OG Detergents. CIU plots (left) and extracted drift times at collisional voltages of 20 V and 90 V (right) are shown. ADH was analysed as described. See **Figure 13** for details. Figure adapted from Kundlacz et al.

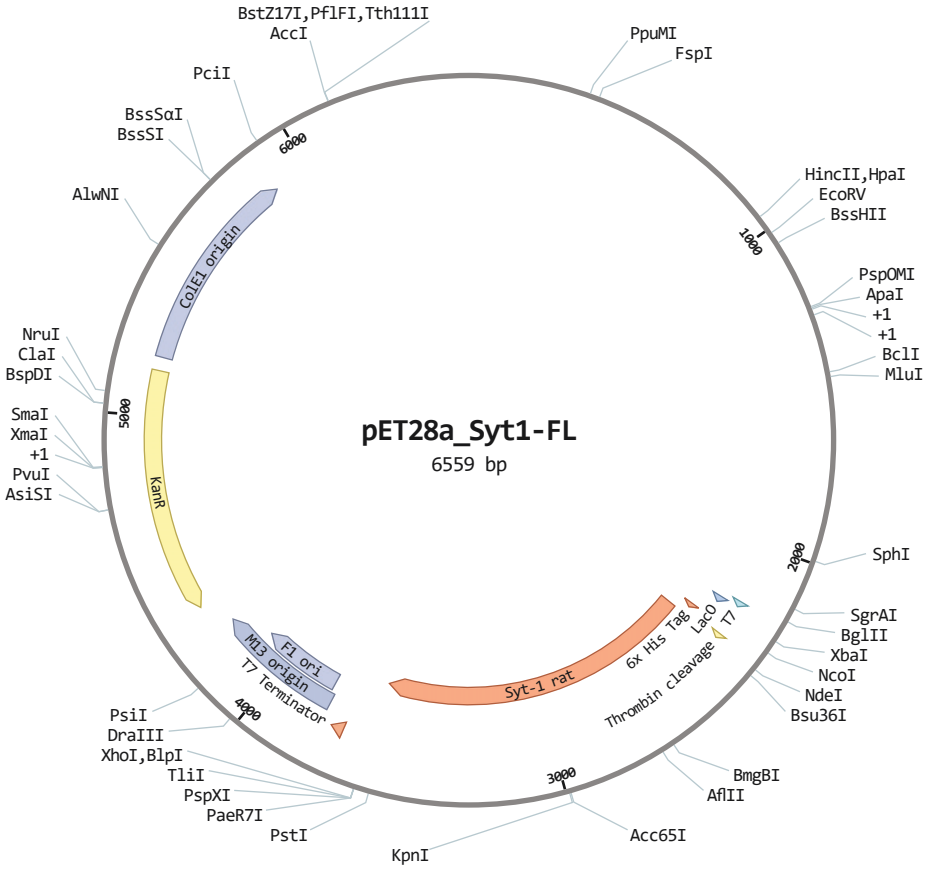


Figure S2: Plasmid Map of pET28a Encoding Lull-Length Syt-1. Annotated coding sequences are indicates as arrows. Restriction enzyme cleavage sites are annotated by the respective enzyme name. All soluble variants of Syt-1 were encoded using the same plasmid architecture.

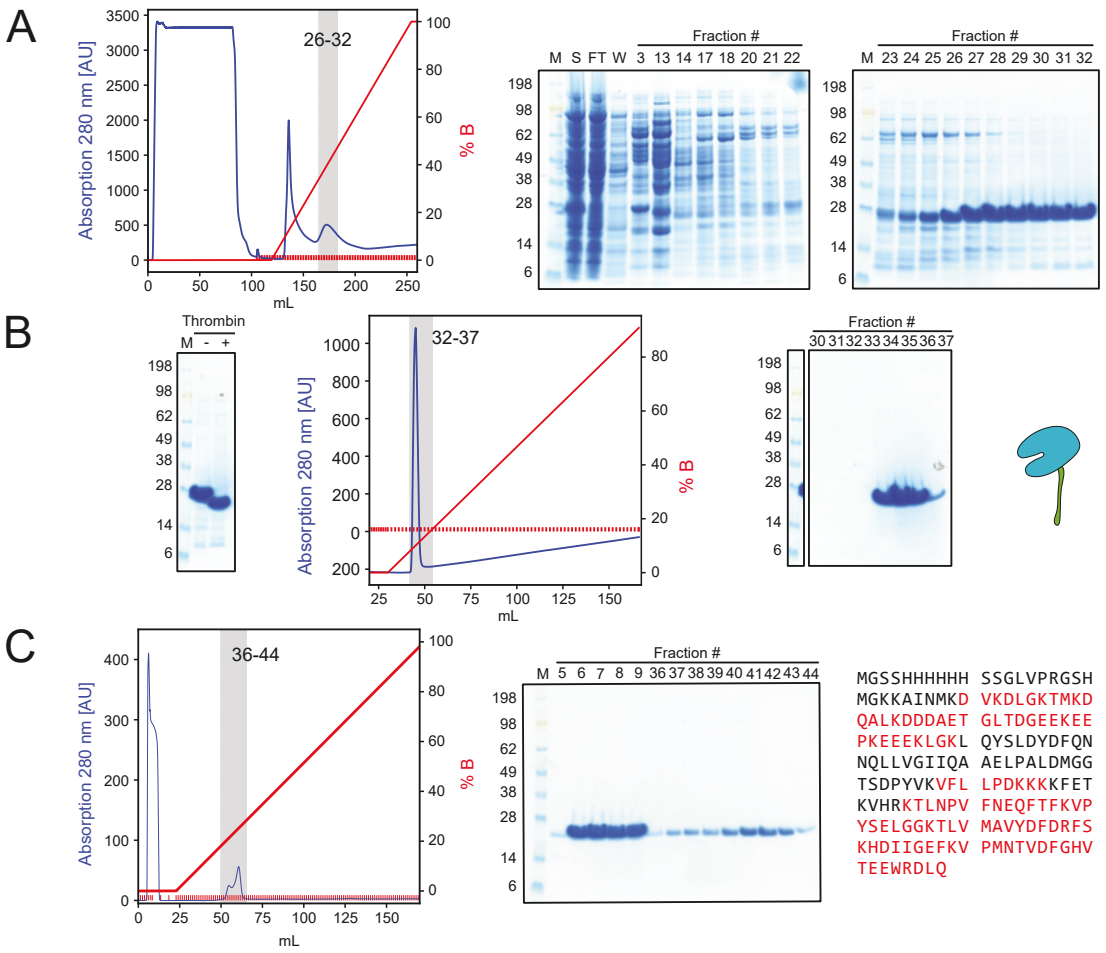


Figure S3: Purification of Syt-1 C2A. (A) Immobilised Ni^{2+} affinity chromatography of lysate of cells expressing Syt-1 C2A fused to a hexahistidine tag. The corresponding chromatogram (left) and stained gels from polyacrylamide gel electrophoresis of the collected fractions (right) are shown. Fractions combined for the next purification step are indicated by grey background. **(B)** Removal of the hexahistidine tag by proteolysis using thrombin (left) and subsequent Ni^{2+} affinity chromatography (chromatogram: middle, gel electrohoresis of fractions: right). After removal of the tag, Syt-1 eluted from the column at low imidazole concentration. **(C)** Anion exchange chromatography to remove bound nucleotides from the protein (left and middle). Fractions eluting from the column were combined, concentrated and stored. Sequence coverage determined from LC-MS/MS (right). Residues present in identified peptides are coloured red. M: Marker; S: Supernatant from centrifugation of cell disruption; W: Wash; FT: Flow-Through.

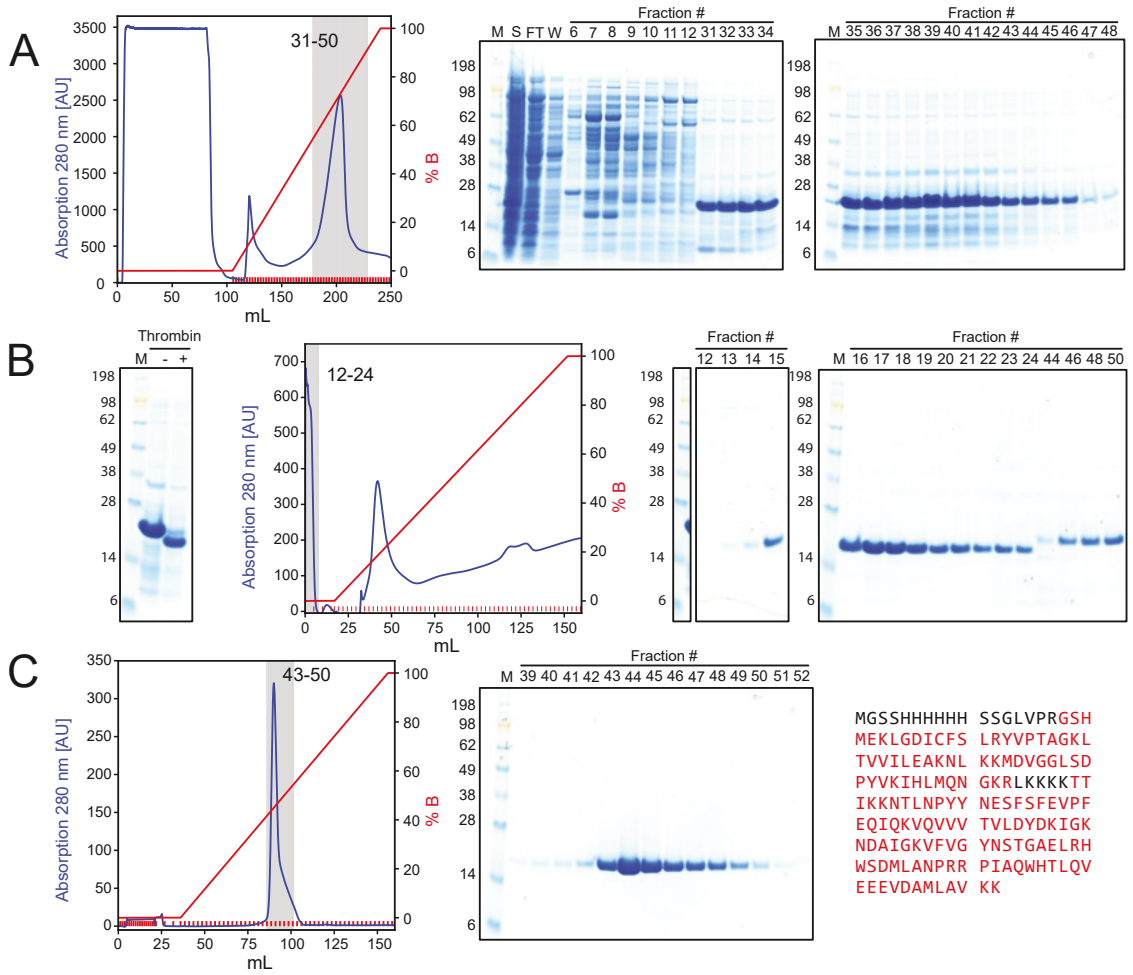


Figure S4: Purification of Syt-1 C2B. (A) Immobilised Ni^{2+} affinity chromatography of lysate of cells expressing Syt-1 C2B fused to a hexahistidine tag. **(B)** Removal of the hexahistidine tag by proteolysis using thrombin (left) and subsequent Ni^{2+} affinity chromatography (chromatogram: middle, gel electrohoresis of fractions: right). **(C)** Cation exchange chromatography to remove bound nucleotides from the protein (left and middle). Sequence coverage determined from LC-MS/MS (right). Residues present in identified peptides are coloured red. See figure S3 for details and abbreviations.

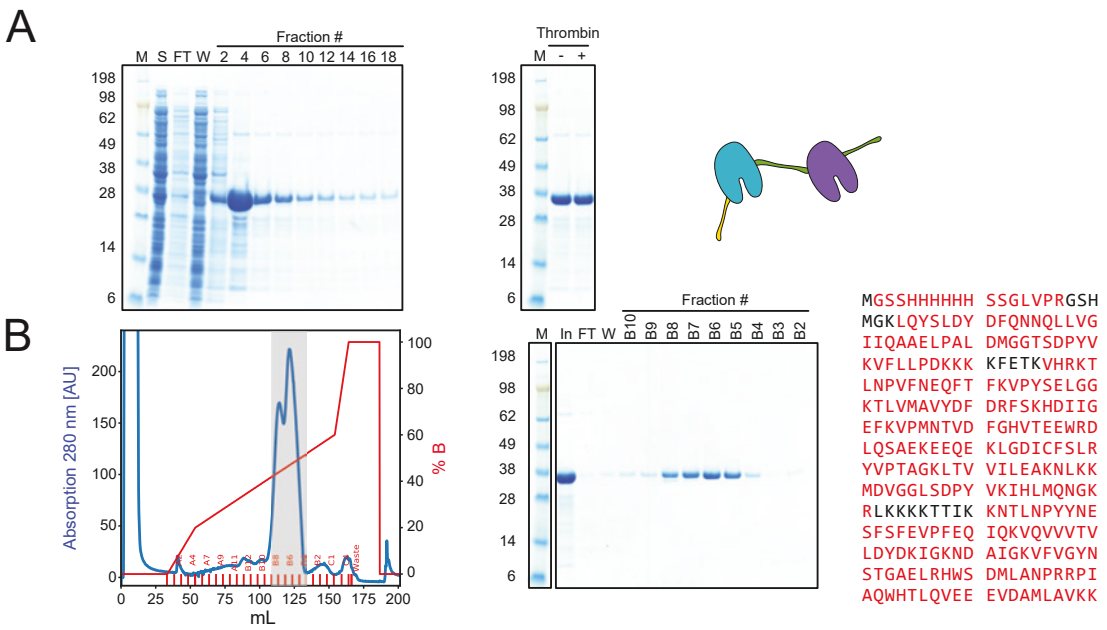


Figure S5: Purification of Syt-1 C2AB-His. **(A)** Ni^{2+} affinity purification of lysate of cells expressing Syt-1 C2AB-His fused to a hexahistidine tag. Tagged proteins were immobilised on Ni^{2+}-chelating nitrilotriacetic acid-agarose beads and washing and protein elution were performed manually by gravity-flow chromatograpohy (left). The hexahistidine tag could not be removed by proteolytic cleavage using thrombin protease (right). **(B)** Cation exchange chromatography to remove bound nucleotides from the protein (left and middle). Sequence coverage determined from LC-MS/MS (right). Residues present in identified peptides are coloured red. See figure S3 for details and abbreviations.

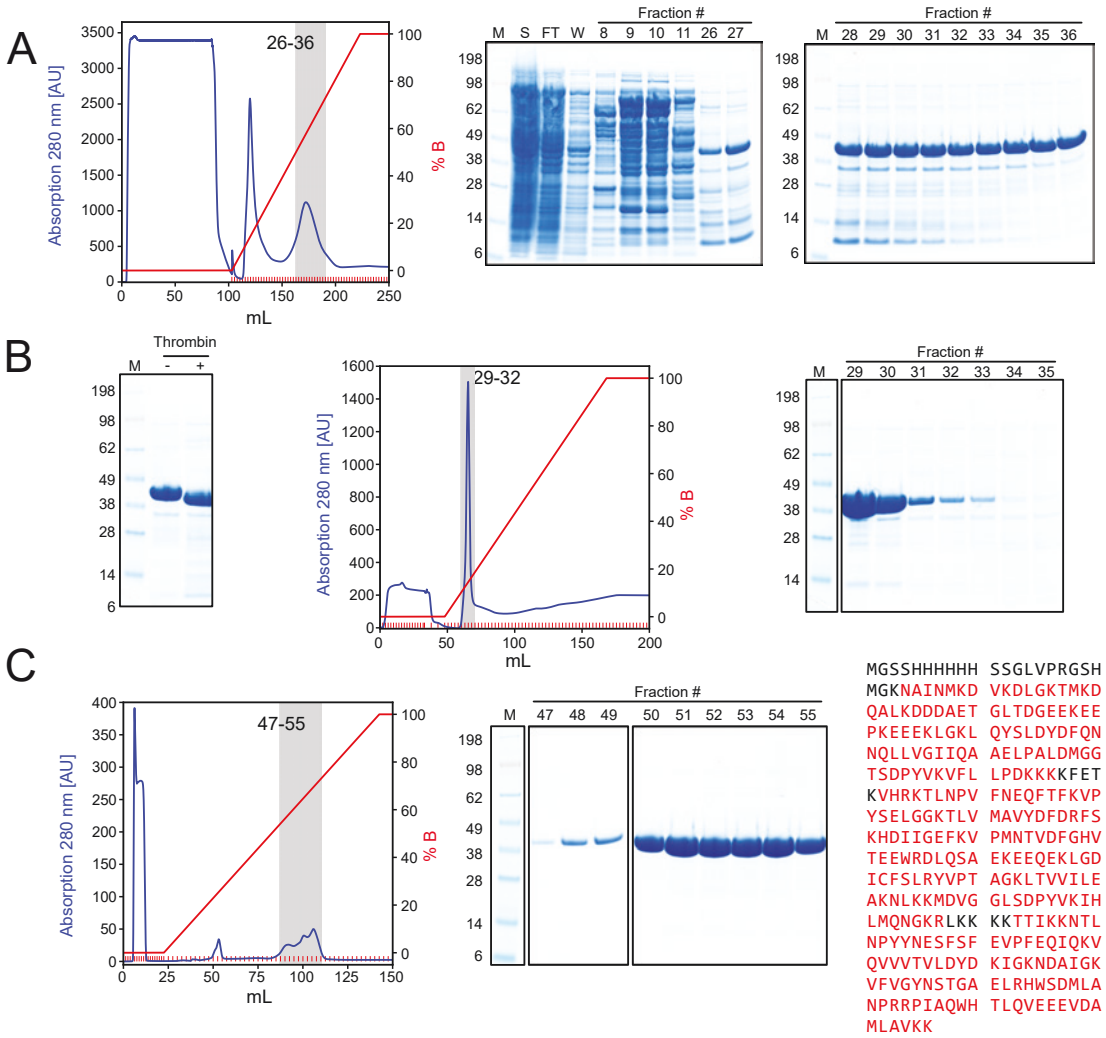


Figure S6: Purification of Syt-1 C2AB. (A) Immobilised Ni^{2+} affinity chromatography of lysate of cells expressing Syt-1 C2AB fused to a hexahistidine tag. **(B)** Removal of the hexahistidine tag by proteolysis using thrombin (left) and subsequent Ni^{2+} affinity chromatography (chromatogram: middle, gel electrohoresis of fractions: right). **(C)** Anion exchange chromatography to remove bound nucleotides from the protein (left and middle). Sequence coverage determined from LC-MS/MS (right). Residues present in identified peptides are coloured red. See figure S3 for details and abbreviations.

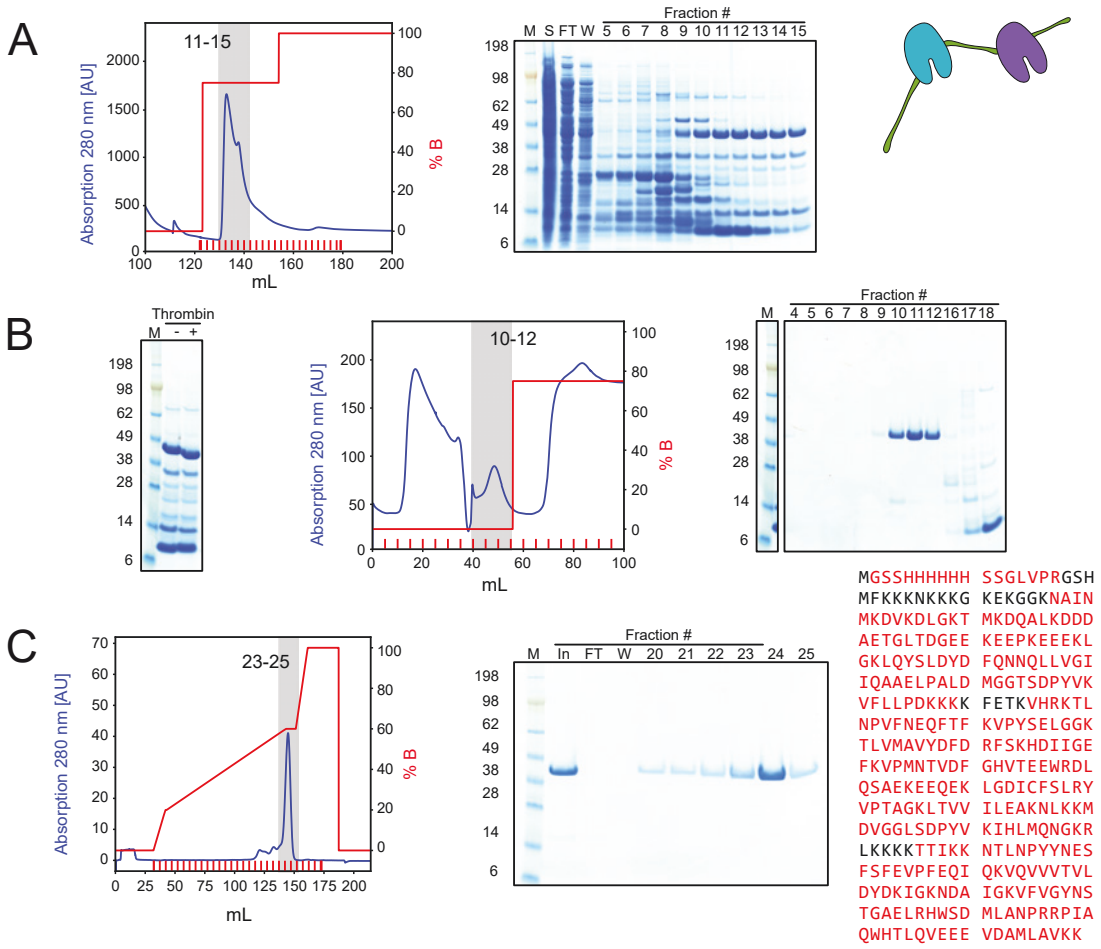


Figure S7: Purification of Syt-1 Cyt. **(A)** Immobilised Ni^{2+} affinity chromatography of lysate of cells expressing Syt-1 Cyt fused to a hexahistidine tag. **(B)** Removal of the hexahistidine tag by proteolysis using thrombin (left) and subsequent Ni^{2+} affinity chromatography (chromatogram: middle, gel electrohoresis of fractions: right). **(C)** Cation exchange chromatography to remove bound nucleotides from the protein (left and middle). Sequence coverage determined from LC-MS/MS (right). Residues present in identified peptides are coloured red. See figure S3 for details and abbreviations.

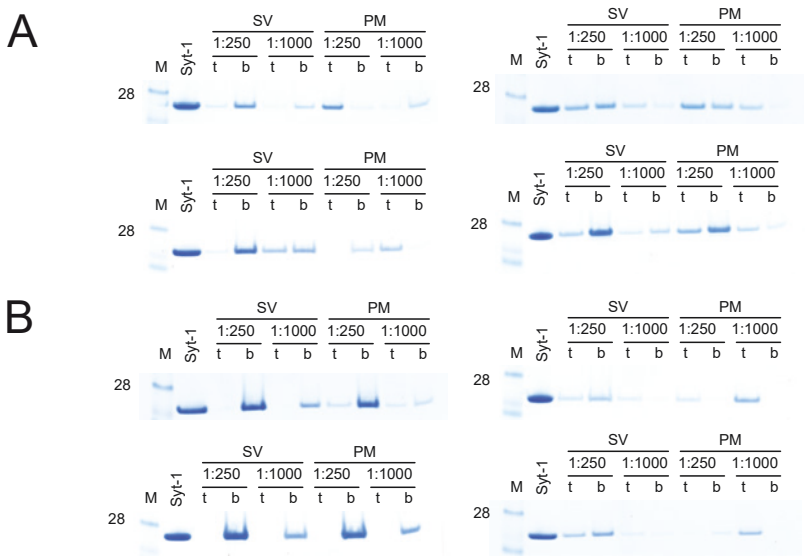


Figure S8: Replicates of the Flotation Assay of the Soluble Syt-1 Variants C2A and C2B. (A) Interaction of Syt-1 C2A in the absence of Ca^{2+} and presence of 1 mM of EGTA (left) and presence (right) of 100 µM $CaCl_2$ with liposomes of compositions mimicking the SV and plasma membranes. Proteins binding to the liposomes float on top (t) of a sucrose density gradient, whereas free protein accumulates at the bottom (b) of the centrifugation tube. **(B)** Interaction of Syt-1 C2B in the absence (left) and presence (right) of Ca^{2+}. Experiments were performed by Melissa Frick.

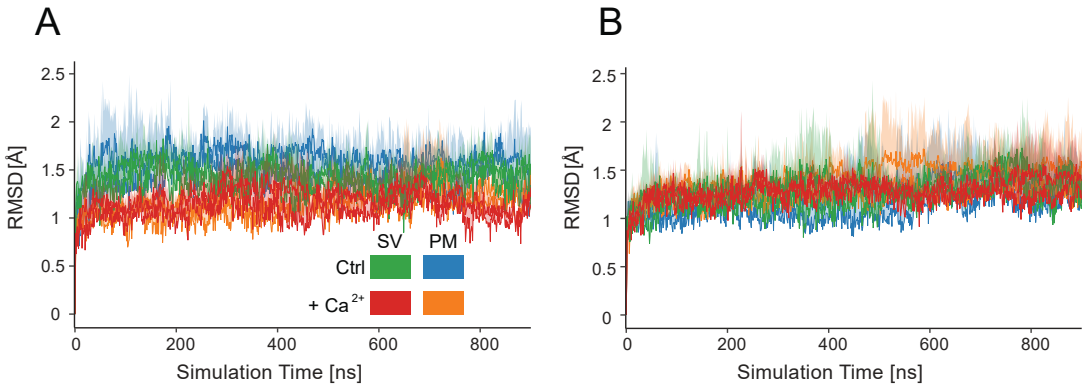


Figure S9: RMSD of Syt-1 C2A and C2B. (A) C_α RMSD of Syt-1 141-263 aligned to the first frame of the directory. **(B)** C_α RMSD of Syt-1 272-418. RMSDs are shown for each C2 domain interacting with the SV membrane in the presence (red) and absence (green) of bound Ca^{2+} ions as well as interacting with the plasma membrane in its presence (orange) and absence (blue) (See inset for colour scheme). The mean RMSD (solid line) and the range of RMSDs per time step (transparent area) of three replicates is shown.

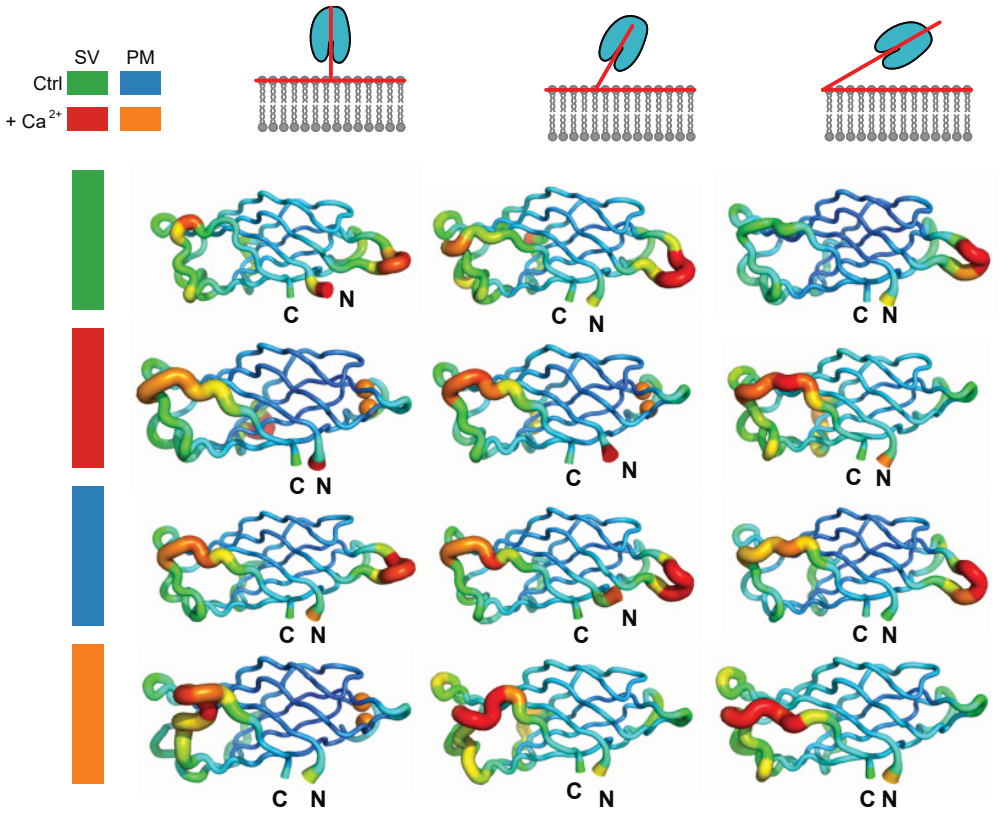


Figure S10: RMSF of Syt-1 C2A Mapped onto the Structure. The RMSF of Syt-1 141-263 as sausage representation of the final protein structure per simulation. Higher RMSFs are indicated by larger diameter and warmer colours of the protein backbone representation. Structures are shown for three simulations of the C2A domain aligned parallel to the z axis (left), at 30deg angle (middle) and 60deg angle (right) to the z axis. First row (green): Syt-1 C2A interacting with the SV membrane; Second row (red): C2A with two Ca^{2+} ions bound interacting with the SV membrane; Third row (blue): Syt-1 C2A interacting with the plasma membrane; Fourth row (orange): C2A with two Ca^{2+} ions bound interacting with the plasma membrane;

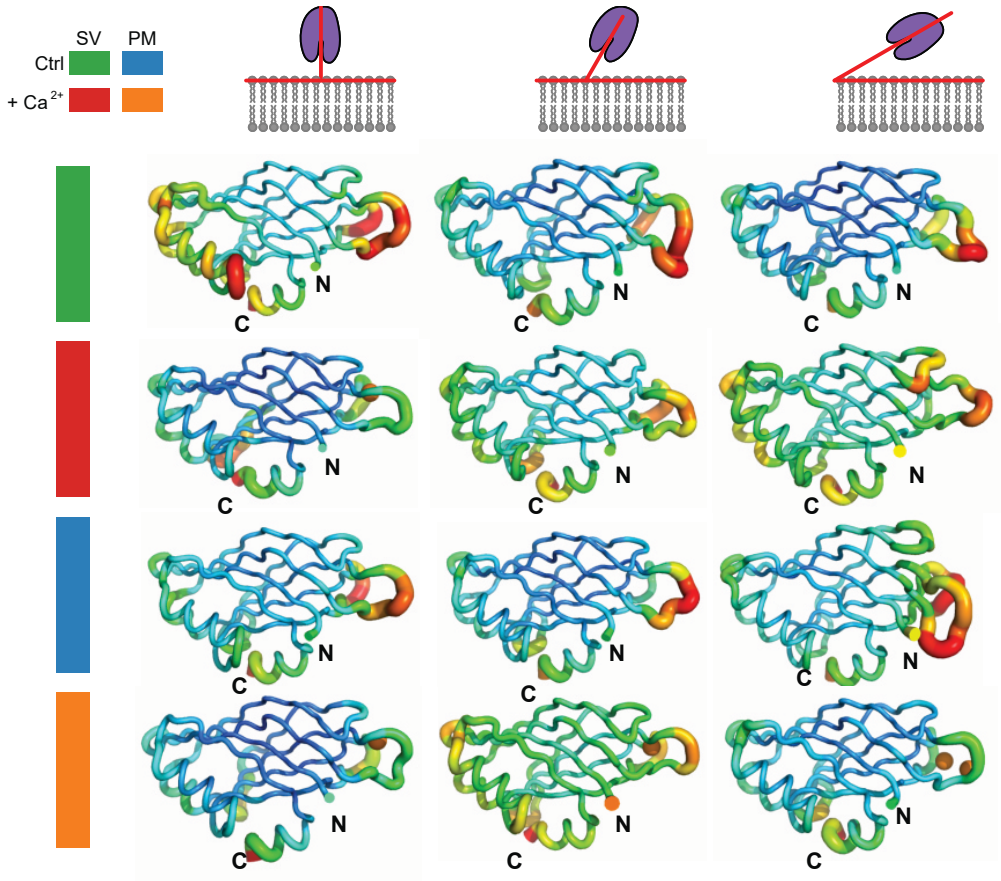


Figure S11: RMSF of Syt-1 C2B Mapped onto the Structure. The RMSF of Syt-1 272-418 as sausage representation of the final protein structure per simulation. Higher RMSFs are indicated by larger diameter and warmer colours of the protein backbone representation. Structures are shown for three simulations of the C2B domain aligned parallel to the z axis (left), at 30deg angle (middle) and 60deg angle (right) to the z axis. First row (green): Syt-1 C2B interacting with the SV membrane; Second row (red): C2B with two Ca^{2+} ions bound interacting with the SV membrane; Third row (blue): Syt-1 C2B interacting with the plasma membrane; Fourth row (orange): C2B with two Ca^{2+} ions bound interacting with the plasma membrane;

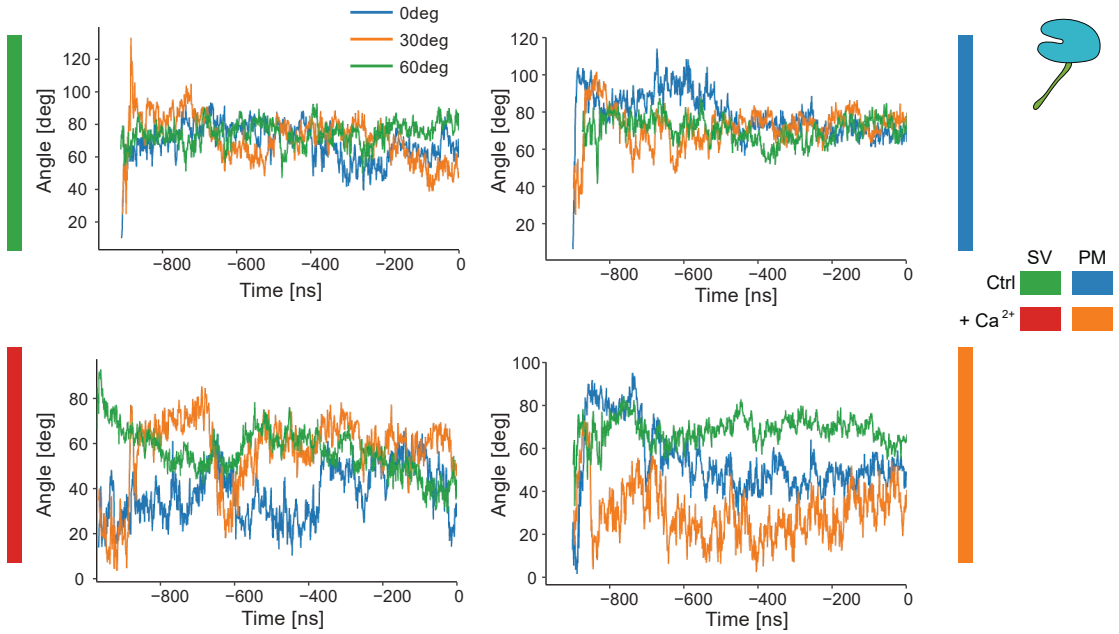


Figure S12: Membrane Contact Angles of C2A Separated by Initial Orientation. Angles are shown separately for interactions with the SV membrane in the absence (top left, green) and presence of Ca^{2+} (lower left, red) and for interaction with the plasma membrane in the absence (top right, blue) and presence of Ca^{2+} (lower right, orange). On the x-axis, the time before each trajectory ends is indicated.

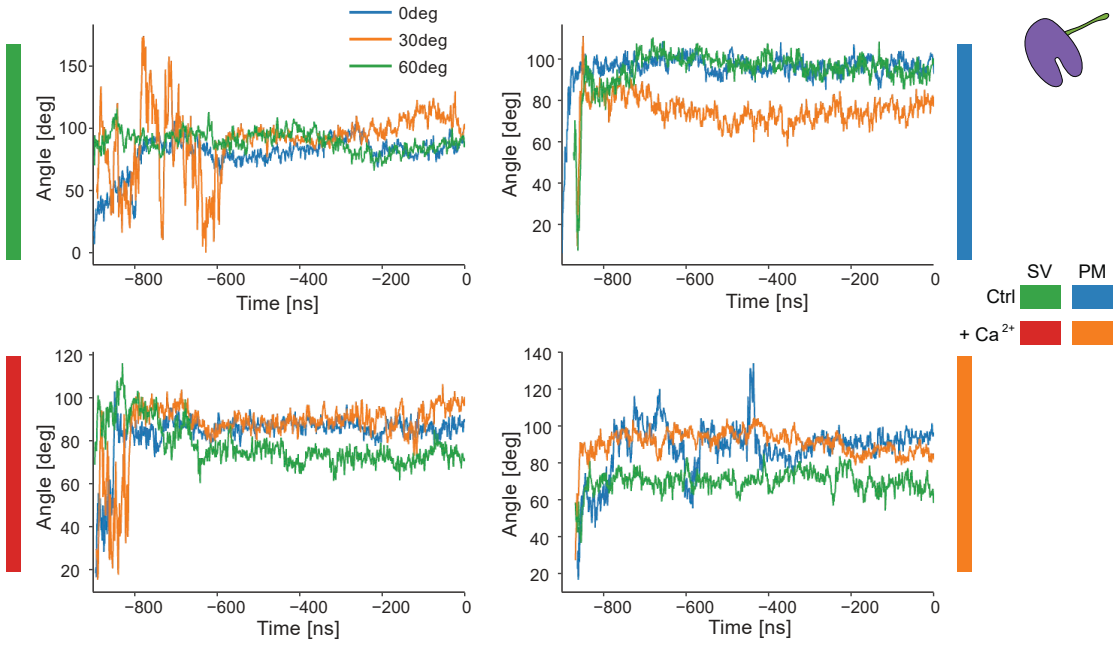


Figure S13: Membrane Contact Angles of C2B Separated by Initial Orientation. Angles are shown separately for interactions with the SV membrane in the absence (top left, green) and presence of Ca^{2+} (lower left, red) and for interaction with the plasma membrane in the absence (top right, blue) and presence of Ca^{2+} (lower right, orange). On the x-axis, the time before each trajectory ends is indicated.

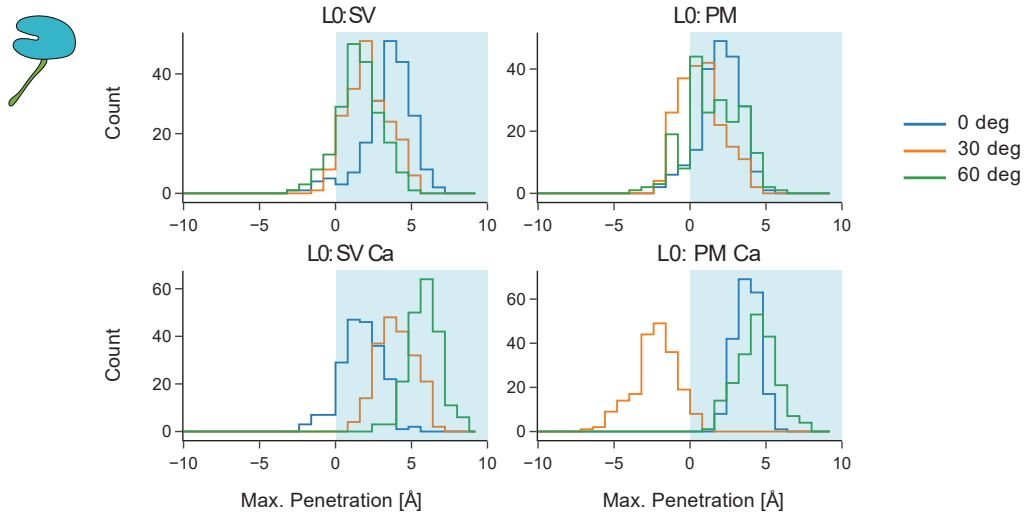


Figure S14: Membrane Penetration Histograms of Syt-1 C2A Simulations. Maximum membrane penetration for every simulation starting with different orientations of the C2 domain: 0deg (blue), 30deg (orange) and 60deg (green) with respect to the membrane normal. Distances are calculated with respect to the phosphate atoms of the upper leaflet (L0) of each membrane. Distances corresponding to Syt-1 entering the membrane are shown as light blue area.

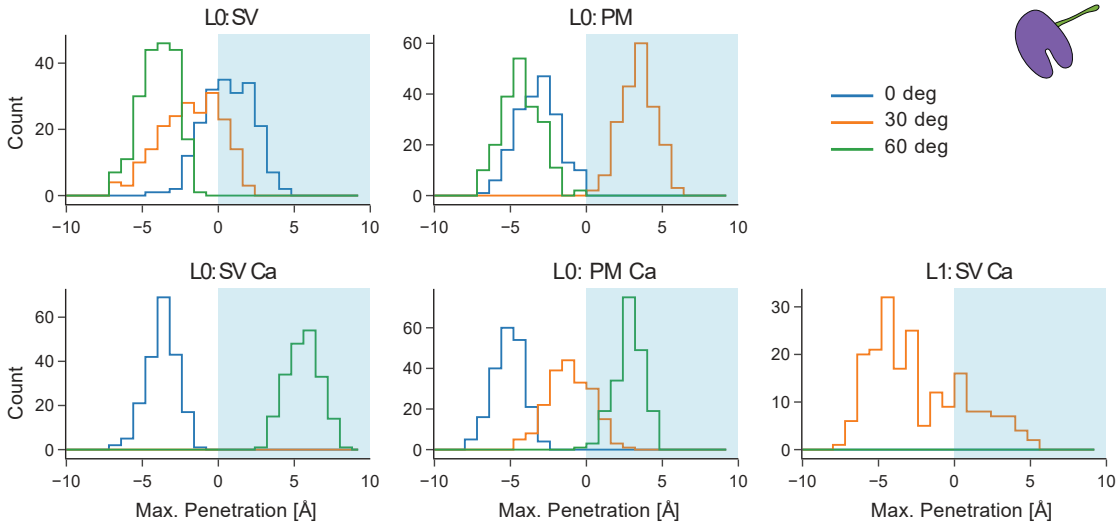


Figure S15: Membrane Penetration Histograms of Syt-1 C2B Simulations. Maximum membrane penetration for every simulation starting with different orientations of the C2 domain: 0deg (blue), 30deg (orange) and 60deg (green) with respect to the membrane normal. Distances are calculated with respect to the phosphate atoms of the upper (L0) or lower leaflet (L1) of each membrane. Distances corresponding to Syt-1 entering the membrane are shown as light blue area.

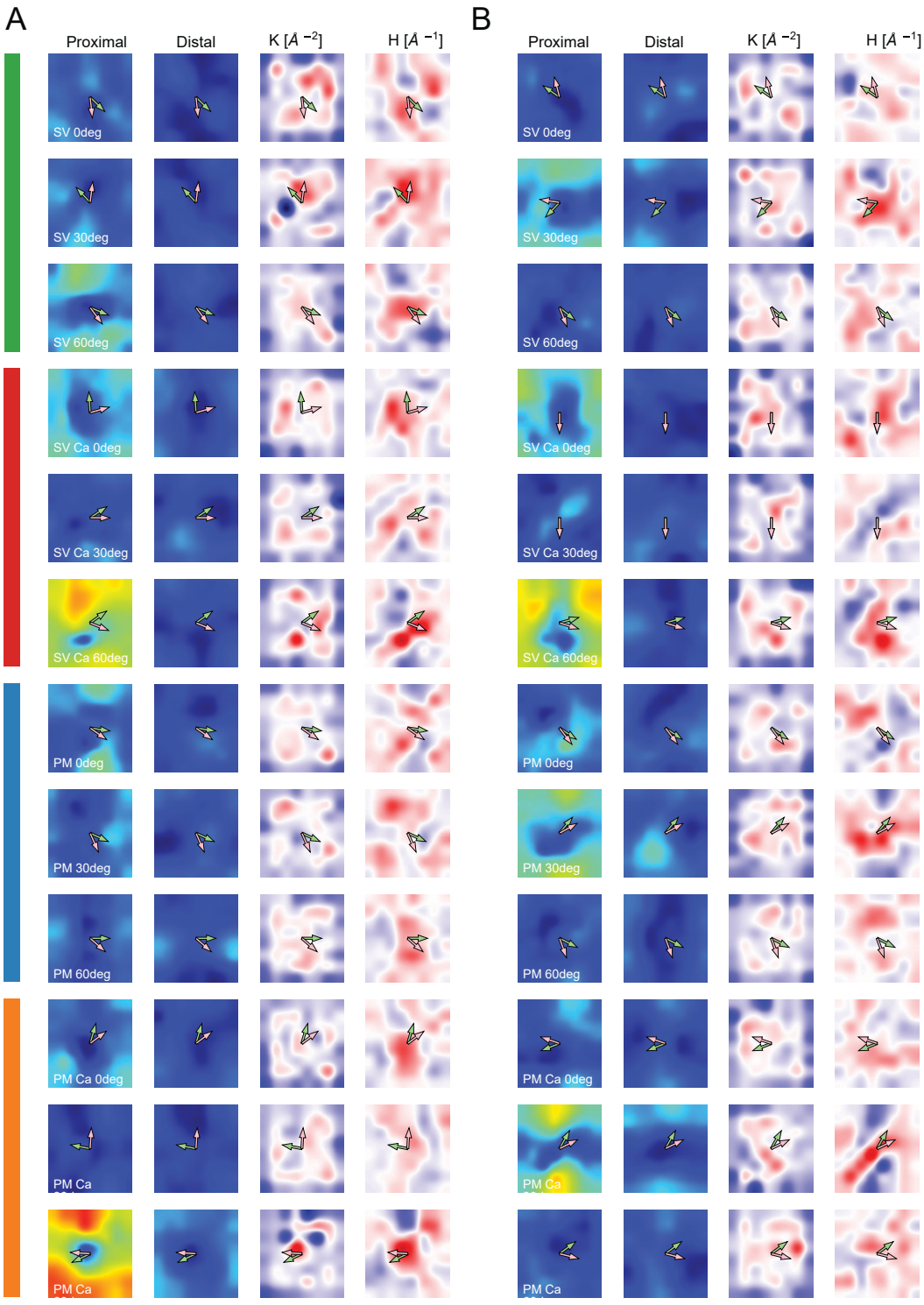
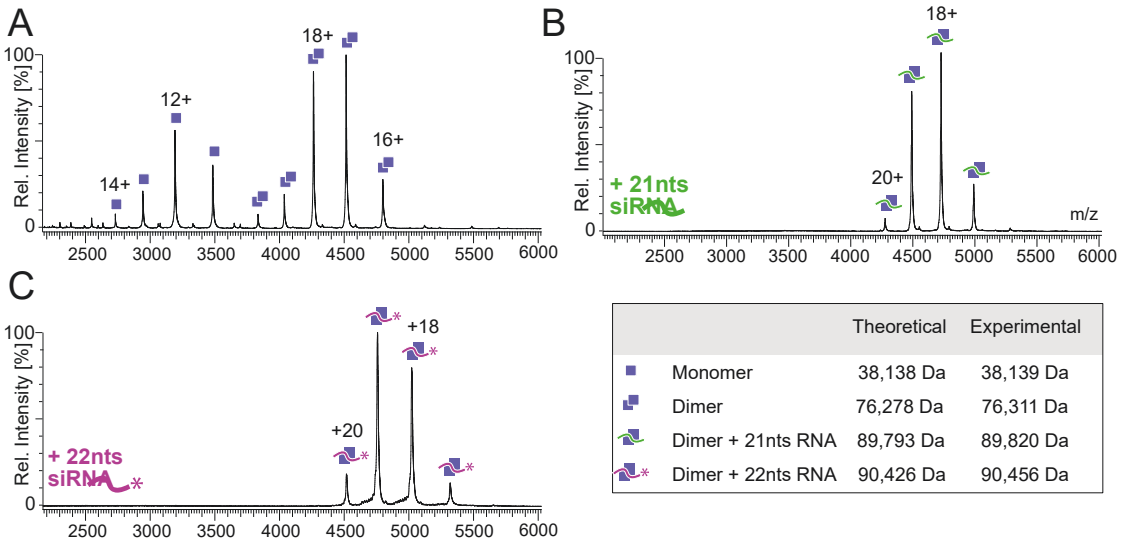


Figure S16: Syt-1-Induced Membrane Curvature in all Simulations. **(A)** Membrane surface elevations defined by lipid headgroup phosphor atoms proximal and distal to the C2A interaction site. **(B)** Same for membranes in the presence of C2B. For colour bars and simulations in the absence of protein refer to **Figure 39**.



	Theoretical	Experimental
Monomer	38,138 Da	38,139 Da
Dimer	76,278 Da	76,311 Da
Dimer + 21nts RNA	89,793 Da	89,820 Da
Dimer + 22nts RNA	90,426 Da	90,456 Da

Figure S17: Replicate of Native MS of p38 in the Presence and Absence of siRNA. (A) Native MS of p38 in the absence of RNA. Peaks corresponding in mass to monomeric (single square) and dimeric (two squares) p38 are observed. Charge states for selected peaks are indicated. **(B)** Mixing of 21 nts siRNA with p38. Peaks corresponding in mass to RNA-bound p38 dimer (blue squares with green line) are shown. **(C)** Mixing of 22 nts siRNA with p38. Peaks corresponding in mass to RNA-bound p38 dimer (blue squares with purple line) are shown. Experimentally determined and theoretically expected masses are shown (lower right).

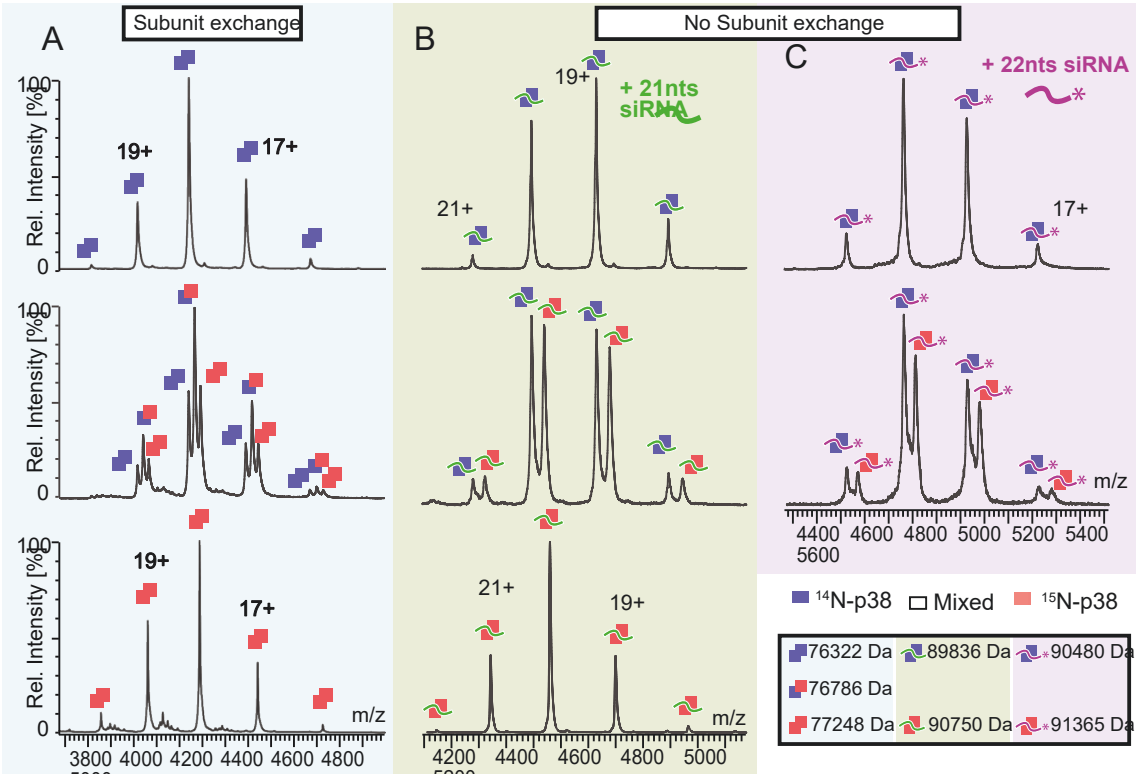


Figure S18: Replicate of Native MS Analysis of Subunit Exchange Between p38 Dimers. (A) Native MS of 5 µM ^{15}N p38 (bottom, red) and 7.5 µM ^{14}N (top, blue). Mixing at equimolar ratio (middle) yields a peak distribution corresponding to masses matching the unlabeled dimer(blue), the labelled dimer (red) and a dimer of one labelled and one unlabelled subunit (red and blue) at 1:2:1 intensity ratio. **(B)** 13 µM ^{14}N p38 preincubated with excess of 21 nts siRNA and free RNA removed by filtration (top) was mixed with an equimolar volume of 6 µM RNA-free ^{15}N resulting in a peak series corresponding to masses of ^{14}N p38 dimer with RNA and ^{15}N p38 dimer with RNA but no signal for a mixed dimer. A spectrum of ^{15}N p38 bound to 21 nts RNA (bottom) is shown for comparison. **(C)** 3.75 µM ^{14}N p38 dimer preincubated with 22 nts RNA-bound and with free RNA removed by filtration (top). Addition of equimolar volume of 6 µM RNA-free ^{15}N p38 (middle) results in two separate peaks series corresponding to the RNA-bound form of both dimers.

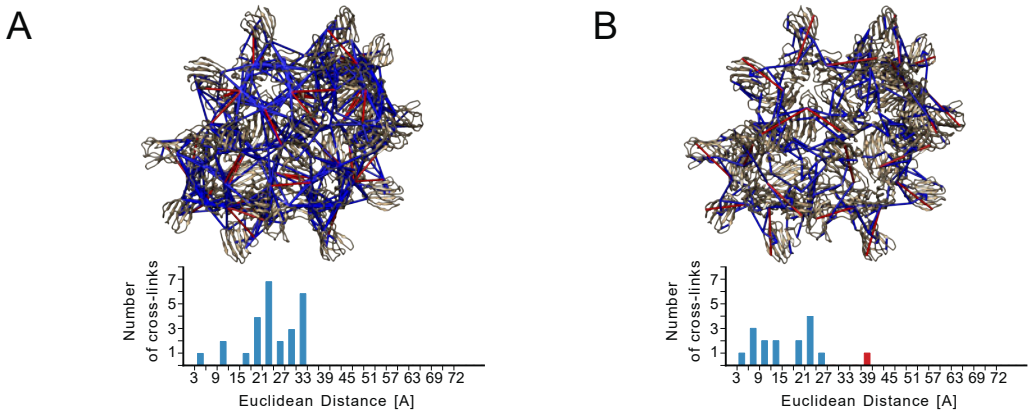


Figure S19: Mapping BS2G and EDC Cross-Links on the Capsid Model. **(A)** Mapping of cross-links identified with BS2G onto the model of the p38 viral capsid (top) and histogram of distances (bottom). Note that three cross-links violated the distance cut-off of 31 Å. **(B)** Mapping of cross-links identified with EDC onto the model of the p38 viral capsid (top) and histogram of distances (bottom). Note that one cross-link violated the distance cut-off of 25 Å. Model for the capsid structure was kindly provided by Prof. Dr. Milton T. Stubbs .

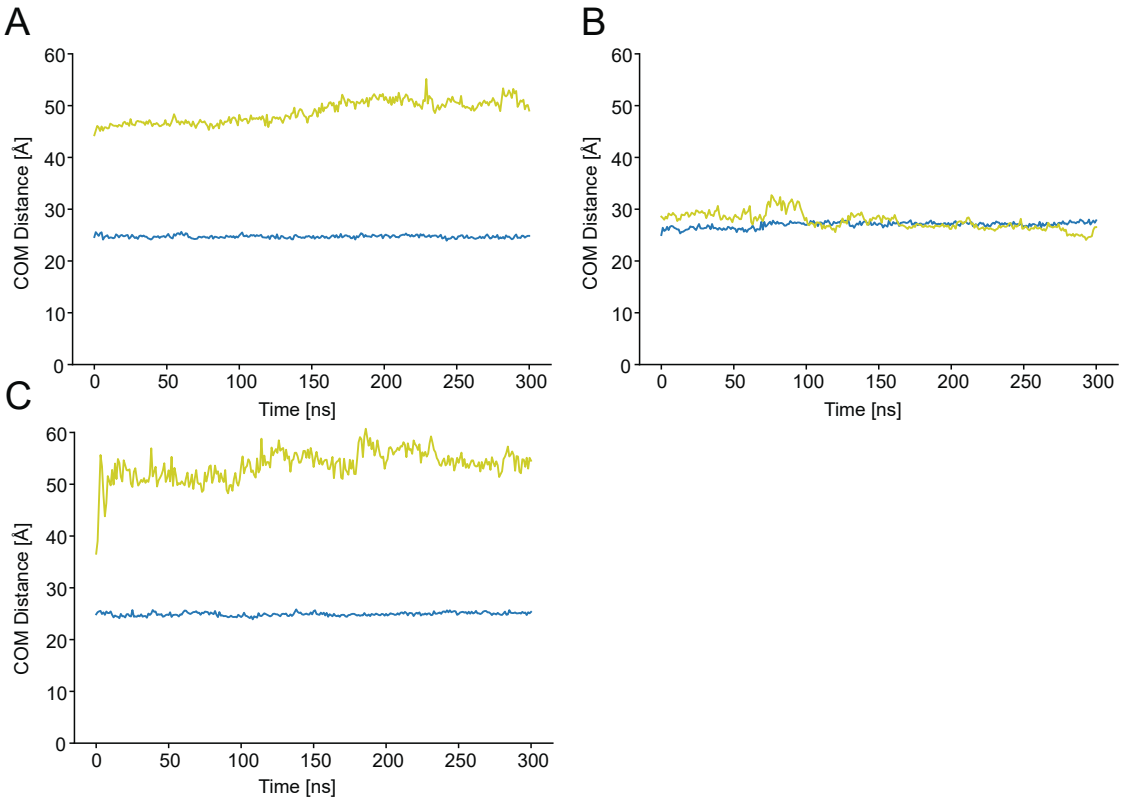


Figure S20: Centre of Mass Distances from MD Simulation of p38 Structural Models. Centre of mass distances between the S- (yellow, residues 81 - 239) and P-domains (blue, residues 261 - 351) along the trajectory. **(A)** p38 model generated from the capsid X-ray structure. **(B)** Cluster 1 obtained from cross-linking guided docking. **(C)** Cluster 2 obtained from cross-linking guided docking.

List of Figures

List of Tables

List of Abbreviations

θ_{MR} Mean Residue Ellipticity
β-**LG** β-lactoglobulin
Cpx-1 Complexin-1
SNAP-25 Synaptosomal-Associated Protein Of 25 kDa
Stx-1A Syntaxin-1A
Syb-2 Synaptobrevin-2
Syt-1 Synaptotagmin-1
E. coli *Escherichia coli*
m/z Mass-to-charge

AA Arginine Apex
ADH Alcohol Dehydrogenase
AGO Argonaute Protein

BS2G Bis(sulfosuccinimidyl)glutarate
BS3 Bis(sulfosuccinimidyl)suberate
BSA Bovine Serum Albumin

C8E4 Octyl Tetraethylene Glycol Ether
CBL Ca^{2+} Binding Loop
CCS Collision-averaged Cross Section
CD Circular Dichroism
CHAPS 3-((3-cholamidopropyl) Dimethylammonio)-1-propanesulfonate
CIU Collision-induced Unfolding
cmc Critical Micelle Concentration

DCL Dicer-like Protein
DNA Deoxyribonucleic Acid
DOPC Di-oleyl-phosphatidylcholine
DOPG Di-oleyl-phosphatidylglycerol
DOPI Di-oleyl-phosphatidylinositol
DOPI(3,4,5)P_3 Di-oleyl-phosphatidylinositol-3,4,5-trisphosphate
DOPS Dioleyl-phosphatidylserine
dsRNA Double-stranded RNA

EDC 1-ethyl-3-[3-(dimethylamino)propyl]carbodiimide
EDTA Ethylene Diamine Tetraacetic Acid
EGTA Ethylene Glycol Tetraacetic Acid
EPR Electron Paramagnetic Resonance
ESI Electrospray Ionisation

HEPES 4-(2-Hydroxyethyl)-1-piperazine Ethanesulfonic Acid

I(1,3,5)P_3 Inositol-1,3,5-triphosphate
I(1,4,5)P_3 Inositol-1,4,5-triphosphate
IDR Intrinsically Disordered Region
IMS Ion-mobility Spectrometry

$K_{D\ app}$ Apparent Dissociation Constant

LC Liquid Chromatography
LDAO Lauryldimethylamine N-oxide

MALDI Matrix-assisted Laser Desorption/ionisation
MD Molecular Dynamics
mRNA Messenger RNA
MS Mass Spectrometry

NTA Nitrilotriacetic Acid

OG Octyl Glycoside

PC Phosphatidylcholine
PG Phosphatidylglycerol
PI Phosphatidylinositol
PI(3,4,5)P_3 Phosphatidylinositol-3,4,5-trisphosphate
PI(4)P Phosphatidylinositol-4-phosphate
PI(4,5)P_2 Phosphatidylinositol-4,5-bisphosphate
PIP Phosphatidylinositol Phosphate
POPC Palmitoyl-oleyl-phosphatidylcholine
POPE Palmitoyl-oleyl-phosphatidylethanolamine
PS Phosphatidylserine
PSM Peptide Spectrum Match

RISC RNA-induced Silencing Complex
RMSD Root-mean-square Deviation
RMSF Root-mean-square Fluctuation
RNA Ribonucleic Acid

SEC Size-exclusion Chromatography
siRNA Small Interfering RNA
SNARE Soluble N-ethylmaleimide-sensitive-factor Attachment Receptor
ssRNA Single-stranded RNA
Sulfo-NHS N-hydroxysulfosuccinimide
SV Synaptic Vesicle

TCEP Tris(2-carboxyethyl)phosphine
TCV Turnip Crinkle Virus
TFA Trifluoroacetic Acid
TMAO Trimethylamine-N-oxide
TOF Time-of-flight

VSR Viral Suppressor Of RNA Silencing

CURRICULUM VITAE

NAME: Julian Bender, M.Sc.

PLACE AND DATE OF BIRTH: [redacted]

ADDRESS: [redacted]

NATIONALITY: German

EMAIL: [redacted]

Education

Since AUGUST 2017	PhD studies **Martin Luther University**, Halle (Saale) Thesis: "Biophysical Characterisation of Protein-Ligand Interactions" Combination of structural mass spectrometry techniques with computational methods for the characterisation of protein interactions Advisor: Jun.-Prof. Dr. Carla SCHMIDT
MAI 2017	Master's Degree in BIOCHEMISTRY & BIOPHYSICS, **Albert Ludwig University**, Freiburg Final Grade: 1.0 Thesis: "Study of protein import machineries by structural proteomics approaches" Application of chemical cross-linking and native mass spectrometry for the structural characterisation of peroxisomal proteins Advisor: Prof. Dr. Bettina WARSCHEID
JULY 2014	Bachelor's Degree in BIOLOGY, **Albert Ludwig University, Freiburg** Final Grade: 1.3 Thesis: "Protein import into peroxisomes through Pex14p" Analysis of protein-protein-interactions by chemical cross-linking, affinity purification and high-resolution mass spectrometry Advisor: Prof. Dr. Bettina WARSCHEID

July 2010	**Christian-Wirth-Schule Usingen**, Abitur Final Grade: 1.3

Awards & grants

Apr 2019	Scholarship of the Studienstiftung des deutschen Volkes
Mar 2019	Student Bursary for „Celebration of Native Mass Spectrometry “symposium
May 2018	Reduced fees for EMBO Workshop „Integrative Modelling of Biomolecular Interactions“
Dec 2017	ASMS Sanibel Student Stipend award
Dec 2017	German Academic Exchange Service (DAAD) conference travel stipend
July 2017	Freiburg university Alumni prize for academic achievement during Master’s studies
May 2017	bioss stipend
July 2015 - 2017	Scholarship of the Studienstiftung des deutschen Volkes
Sept 2014 - June 2015	*Deutschlandstipendium* - Federal German Scholarship
July 2014	Freiburg university Alumni prize for academic achievement during Bachelor’s studies
July 2010	Karl-von-Frisch prize for excellent abitur in life sciences

Conferences and Symposia

Feb 2020	Annual Meeting of the Biophysical Society, San Diego, CA, USA (poster presentation)
Nov 2019	HALOmem International Meeting, Halle (poster presentation and talk)
Nov 2019	Symposium on Structural Proteomics, Göttingen
Apr 2019	Mosbacher Kolloquium of the Gesellschaft für Biochemie und Molekularbiologie *High resolution imaging of cells and molecules*, Mosbach (poster presentation)
Mär 2019	Symposium *Celebration of Native Mass Specrometry*, Oxford, UK (poster presentation)
Dez 2018	Methods course *Modeling Proteins guided by Electron Paramagnetic Resonance and Mass Spectrometry*, Leipzig
Juli 2018	Integrative Modeling of Biomolecular Interactions: *European Molecular Biology Organisation* summer course, Barcelona, Spain (poster presentation and short talk)
Juni 2018	Annual conference of the *American Society for Mass Spectrometry*, San Diego (CA), USA (poster presentation)
Mar 2018	European Mass Spectrometry Conference, Saarbruecken (poster presentation)
Jan 2018	Sanibel conference: Molecular Modeling and Quantum Mechanical Calculations in Mass Spectrometry, St. Petersburg (FL) (poster presentation)
Oct 2017	Symposium on Structural Proteomics (SSP), Vienna (poster presentation and short talk)

Sept 2017	International Fall Conference of the German Society for Biochemistry and Molecular Biology: *Molecular Basis of Life*, Bochum (poster presentation)
Mar 2017	International Symposium: Coupling and Modification of Proteins, Freiburg (poster presentation)
Sept 2016	5th Open European Peroxisome Meeting (OEPM), Vienna (poster presentation)

Work

Feb 2022	Research associate, Chair of Biochemistry II, Julius-Maximilians-University of Würzburg
Aug 2017 - Jan 2022	Research associate, IWE ZIK HALOmem, Martin Luther University Halle-Wittenberg
May 2017 - Aug 2017	Research assistant, Institute of Biology II, Albert-Ludwigs-University Freiburg
Dec 2015 - Mar 2016	Internship in the laboratory of Prof. Dame Carol Robinson, Department of Physical and Theoretical Chemistry, University of Oxford, United Kingdom
Sept 2014 - Mar 2015	Research assistant, Institute of Biology II, Albert-Ludwigs-University Freiburg
Jul 2013 - Jan 2014	Research assistant, Institute of Biology II, Albert-Ludwigs-University Freiburg

Societies and Activities

2018 - today	Gesellschaft Deutscher Chemiker
May 2015 - Oct 2015	iGEM competition on synthetic biology Protein purification and glass surface derivatisation for a diagnostic chip device based on cell-free expression and label-free detection. Worked in a self-organised and self-financed student team.
2014 - today	Gesellschaft für Biochemie und Molekularbiologie

Halle (Saale),
February 14, 2022,

Julian Bender

List of Publications

10. S. Wieland, A. Balmes*, **J. Bender***, J. Kitzinger*, F. Meyer*, A.F.R.M. Ramsperger*, F. Roeder*, C. Tengelmann*, B. H. Wimmer*, C. Laforsch, H. Kress. From properties to toxicity: comparing microplastics to other airborne microparticles. Journal of Hazardous Materials (2022), doi: 10.1016/j.jhazmat.2021.128151

9. T. Kundlacz*, **J. Bender***, C. Schmidt. Effects of non-ionic and zwitterionic detergents on soluble proteins during native mass spectrometry experiments. International Journal of Mass Spectrometry (2021), doi: 10.1016/j.ijms.2021.116652

8. **J. Bender**, C. Schmidt. Exploring Phosphoinositide Binding Using Native Mass Spectrometry. Methods in Molecular Biology (2021), doi: 10.1007/978-1-0716-1142-5_11

7. P. Lill, T. Hansen, D. Wendscheck, B. U. Klink, T. Jeziorek, D. Vismpas, J. Miehling, **J. Bender**, A. Schummer, F. Drepper, W. Girzalsky, B. Warscheid, R. Erdmann, C. Gatsogiannis. Towards the molecular architecture of the peroxisomal receptor docking complex. Proceedings of the National Academy of Sciences (2020), doi: 10.1073/pnas.2009502117.

6. L. Sänger, **J. Bender**, K. Rostowski, R. Golbik, H. Lilie, C. Schmidt, S.-E. Behrens, S. Friedrich. Alternatively spliced isoforms of AUF1 regulate a miRNA-mRNA interaction differentially through their YGG motif. RNA Biology (2020), doi: 10.1080/15476286.2020.1822637.

5. M. Barth, **J. Bender**, T. Kundlacz, C. Schmidt. Evaluation of different labelling strategies for structural analysis of protein complexes. Journal of Proteomics (2020), doi: 10.1016/j.jprot.2020.103793.

4. **J. Bender**, C. Schmidt. The CroCo cross-link converter: a user-centred tool to convert results from cross-linking mass spectrometry experiments. Bioinformatics (2019), doi: 10.1093/bioinformatics/btz732.

3. **J. Bender**, C. Schmidt. Mass spectrometry of membrane protein complexes. Biological Chemistry (2019), doi: 10.1515/hsz-2018-0443

*equally contributing

2. **J. Bender***, S. Bognar*, M. Camagna*, J. A. M. Donauer*, J. W. Eble*, R. Emig*, S. Fischer*, R. Jesser*, L. Keilholz*, D. M. U. Kokotek*, J. Neumann*, S. Nicklaus*, R. R. Q. P. T. Oude Weernink*, L. G. Stühn*, N. Wössner*, S. D. Krämer, P. Schwenk, N. Gensch, G. Roth, M. H. Ulbrich. Multiplexed antibody detection from blood sera by immobilization of in vitro expressed antigens and label-free readout via imaging reflectometric interferometry (iRIf). Biosensors and Bioelectronics (2018), doi: 10.1016/j.bios.2018.05.022

1. A. Chan, A. Schummer, S. Fischer, T. Schröter, L. D. Cruz-Zaragoza, **J. Bender**, F. Drepper, S. Oeljeklaus, W.-H. Kunau, W. Girzalsky, B. Warscheid, R. Erdmann. Pex17p-dependent assembly of Pex14p/Dyn2p-subcomplexes of the peroxisomal protein import machinery. European Journal of Cell Biology (2016), doi: 10.1016/j.ejcb.2016.10.004

Selbstständigkeitserklärung

Ich versichere hiermit, die vorliegende Dissertation selbstständig und ohne unerlaubte fremde Hilfe angefertigt zu haben. Ich habe keine anderen als die im Literaturverzeichnis angeführten Quellen genutzt und sämtliche Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht. Ebenfalls sind alle von anderen Personen bereitgestellten Materialien oder erbrachten Dienstleistungen als solche gekennzeichnet. Die vorgelegte Arbeit ist weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt worden. Ich habe keine früheren erfolglosen Promotionsversuche unternommen.

Halle (Saale),
14. Februar 2022

Julian Bender

Acknowledgements

First and foremost, I would like to thank Jun.-Prof. Dr. Carla Schmidt for her support and for giving me the opportunity to work on Synaptotagmin-1. I witnessed the challenges and joys of a young emerging research group and learned a lot, not only scientifically. I also want to thank the reviewers of this thesis as science without peer review would not be what it is in the first place.

Next, I want to thank 'the group' as science is always a communal effort and most of this work would not have been possible without them (mentions in order of appearance). Dr. Tommy Hofmann who is never short of ideas, only of caffeinated drinks; Sabine Wittig whose idea of a smartphone-compatible cross-linking analysis app called Xinder sadly never saw the light of day; Melissa Frick to whom I am not only grateful for giving us a lift back home every Christmas but also for scientific discussions and museum visits in pre-COVID-19 times; Marie Barth who keeps the (proteomics) lab running under all circumstances; Julia Hesselbarth who has the power to ask seemingly simple questions that reliably lead to complex scientific discussions; Dr. Christian Schwieger who knows the best 'shortcuts' in science and real life that open up new perspectives. Special thanks go to Til Kundlacz with whom I worked together since he started with the mysteries of native MS during his Master's thesis and who didn't stop exploring since. I likewise had the pleasure to supervise Julia Kamischke and Patrick Pieczyk during their research group internship and Master thesis, respectively, a time from which I learned at least as much as they hopefully did.

Several people were involved directly in the work presented in this thesis. Asst. Prof. Dr. Matteo Degiacomi not only sparked my interest in molecular dynamics simulations but was also of great help interpreting the Synaptotagmin simulations. He introduced me to Dr. Lucas Powell-Rudden who meticulously answered all my naive questions concerning simulations, introduced me to Gromacs on the command-line and helped me find the right combination of pre-fabricated software tools and home-made scripts. Prof. Dr. Reinhard Jahn shared the plasmids encoding Synaptotagmin-1 and Dr. Angél Pérez-Lara helped me with setting up the initial purification of intact Synaptotagmin. apl. Prof. Dr. Ralph Golbik initiated the structural characterisation of p38 and was of great help for the interpretation of circular dichroism spectra. And his sharing of his experiences of the earlier days of science was both captivating and insightful. Dr. Tamilarasan Selvaraj and Raphael Janke also contributed to the work on p38. Several collaborative projects not covered in this thesis nevertheless contributed greatly to my personal and scientific growth during my PhD. For this, I would like

to thank Dr. Mona Grimmer, Dr. Matthias Reimers, Dr. Susann Friedrich, Prof. Dr. Sven-Erik Behrens, Dr. Tim Rasmussen, Dr. Marc Lewitzky, Prof. Dr. Stephan Feller, Dr. Jennica Träger, Prof. Dr. Dariush Hinderberger, Dr. Archna Archna and Prof. Dr. Milton Stubbs.

I also want to thank the scientific environment that greatly contributed to sparking ideas and testing hypotheses. I had the luck to be involved in three great groups that each in their on manner shaped the way I look onto science now. First, there is the HALOmem research centre (sometimes referred to as the HALOmem family) with its founding fathers Prof. Dr. Milton T. Stubbs and Prof. Dr. Jochen Balbach but also their research groups as well as the groups of Prof. Dr. Kirsten Bacia and Jun.-Prof. Dr. Panagiotis Kastritis who I want to thank for memorable meetings in Gohrisch, Goslar and of course Lutherstadt Wittenberg. And not one of the inspiring HALOmem retreats and supergroup meetings would have been possible without the organisational efforts of Dr. Ulla Niesbach-Klösgen, Marie Dietze and Fabienne Sommerkamp. Second, I was part of the the DFG research training group 2467 focused on intrinsically disordered proteins. Seminars and retreats within the RTG gave me the opportunity to practise ideas and broaden my horizon on intrinsically disordered proteins. I want to thank all RTG students, in particular Christin Brankatschk, Ioannis Skalidis and Alan Wei, the students' representatives. Keep going. Third, the German Scholarship Foundation (Studienstiftung des Deutschen Volkes) supported these studies not only financially but also by orginising workshops and courses on a diverse set of topics inside and outside science. I would like to thank the Lebenswissenschaftliches Kolleg (life science college) of the Studienstiftung and in particular the working group of Prof. Dr. Holger Kress who once again convinced me of the magic of small things. Furthermore, I would like to mention the inspiring EMBO workshop on integrative computational modelling in Barcelona organised by Prof. Dr. Alexandre Bonvin that I attended during the summer of 2018 and that advanced my interest in computational modelling approaches. A thorough education in mass spectrometry and biochemistry basics makes PhD work a lot easier. Hence, I want to thank the group of Prof. Dr. Bettina Warscheid for bringing me into first contact with mass spectrometers and in particular Dr. Friedel Drepper for sharing his always curious attitude towards technical problems with me. Last but not least I want to mention Jerre van der Horst who contributed to this thesis by providing support for the native mass spectrometers in a swift and approachable manner and through great discussions on what was going wrong.

In addition to scientific support, getting through a PhD also requires personal support. I want to thank my wife Sarah for aiding me keep on track during the time and for going with me through frustrations and joy. In line with proper treatment of animal contributions to scientific research[⋆] I hereby credit our dog Enite who reminded me of my necessary break times. Eventually, I want to thank my parents and my brother who initially put me on the aforementioned track towards science that lead me to where I am today.

[⋆] T. C. Erren et al., *EMBO reports* **2016**, *18*, 18–20, DOI 10.15252/embr.201643618