Interaction of the disordered Gab1 with tyrosine phosphatase Shp2

Dissertation

zur Erlangung des

Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät I

Institut für Biochemie und Biotechnologie

der Martin-Luther-Universität

Halle-Wittenberg,

MARTIN-LUTHER-UNIVERSITÄT
HALLE-WITTENBERG

vorgelegt von

Lisa Machner

There's a flame that leads all souls astray
No one's safe from its tender touch of pain
Everyday it's looking for new slaves
To celebrate the beauty of the grave
HIM

Disorder, Disorder, Disorder
System of a Down

Table of Contents

Introduction
Cell signalling

A necessity in early evolution was the communication between cells. This mechanism dates back to the origin of prokaryotes and has since evolved. In bacteria, extracellular quorum sensing [Latin *quōrum*, the genitive plural form of *quī* ("who, which"); the minimum number of members needed to conduct business and cast votes] is a cell density-dependent system that is utilised, e.g. to form biofilms that protect the bacteria and allow sharing of nutrients (Mhatre et al. 2014). The extracellular signalling is coupled to intracellular signalling, where cyclic dinucleotides spread the signal as second messengers (Fontaine et al. 2018).
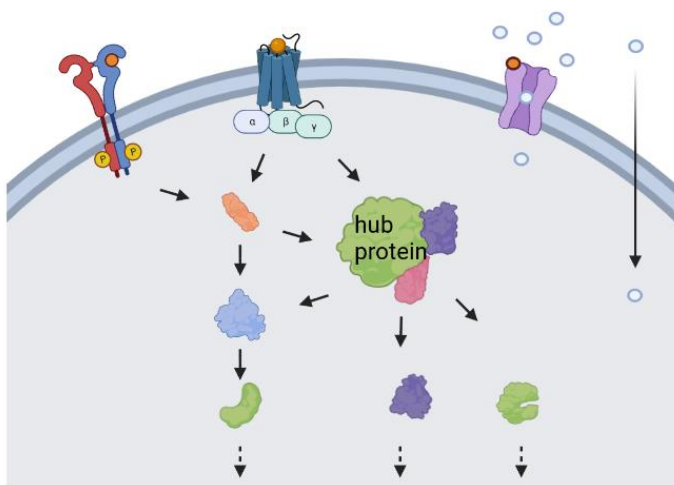
Figure 1 *Eukaryotic cell signalling through receptors. Extracellular signalling molecules (orange balls) bind to the receptors (from left to right: RTK, GPCR, LIC), which transduce the signal throughout the cell through protein-protein-interaction networks with or without integrated hub proteins as nods. Small or lipophilic molecules can diffuse through the membrane. Created with biorender.com.*

In eukaryotes, these systems evolved further to form complex interconnected pathways. Many signalling molecules bind to cell surface receptors in the first step of cell communication (reception). By contrast, small or lipophilic ligands, like steroid hormones, can diffuse through the membrane to bind to intracellular receptors in the cytosol, nucleus or organelle membranes. Three important classes of cell surface receptors are shown in **Figure 1**: enzyme-coupled receptors (e.g. receptor tyrosine kinases; RTKs) (Alberts et al. 2002), G-protein-coupled receptors (GPCRs) (Tuteja 2009) and ligand-gated ion channels (LICs) (Hille 2001). The signal from one ligand molecule is thereby transduced to several molecules, e.g. second messengers. These, in turn, act on a yet greater number of molecules further downstream, thus amplifying, relaying, and distributing the signal throughout the cell to elicit appropriate cellular responses. Multi-site docking proteins (hubs) act to integrate and cross-regulate signals from different pathways. The following section describes the Ras/MAPK and PI3K/AKT pathways with

a focus on the function of the regulating hub protein Gab1 and its implication in pathway crosstalk.

Ras/MAPK and PI3K/AKT pathway

The mitogen EGF (epidermal growth factor) binds to the extracellular domain of its cognate EGF receptor (EGFR), an RTK. This leads to the dimerisation of the receptor and autocatalytic phosphorylation of its intrinsically disordered intracellular tails. These phosphotyrosines serve as docking sites for other proteins. The further downstream signalling is depicted in **Figure 2**. One of these binding partners is Grb2 (Growth factor receptor-bound protein 2), which also binds to the Ras-guanine nucleotide exchange factor, Sos (Son of Sevenless), recruiting it to the membrane. Sos activates the membrane-localised Ras·GDP by replacing GDP with GTP, resulting in Ras·GTP. Ras·GTP, in turn, initiates the mitogen-activated protein kinase (MAPK) phosphorylation cascade by activating the serine-/threonine kinase Raf-1 (MAPKKK, MAPK kinase kinase), which transduces the signal further to MEK 1/ 2 (MAPKK, MAPK kinase) and ERK 1/ 2 (MAPK). ERK translocates into the nucleus and activates transcription factors responsible for cell proliferation, survival, and migration (Molina and Adjei 2006; McCubrey et al. 2007).
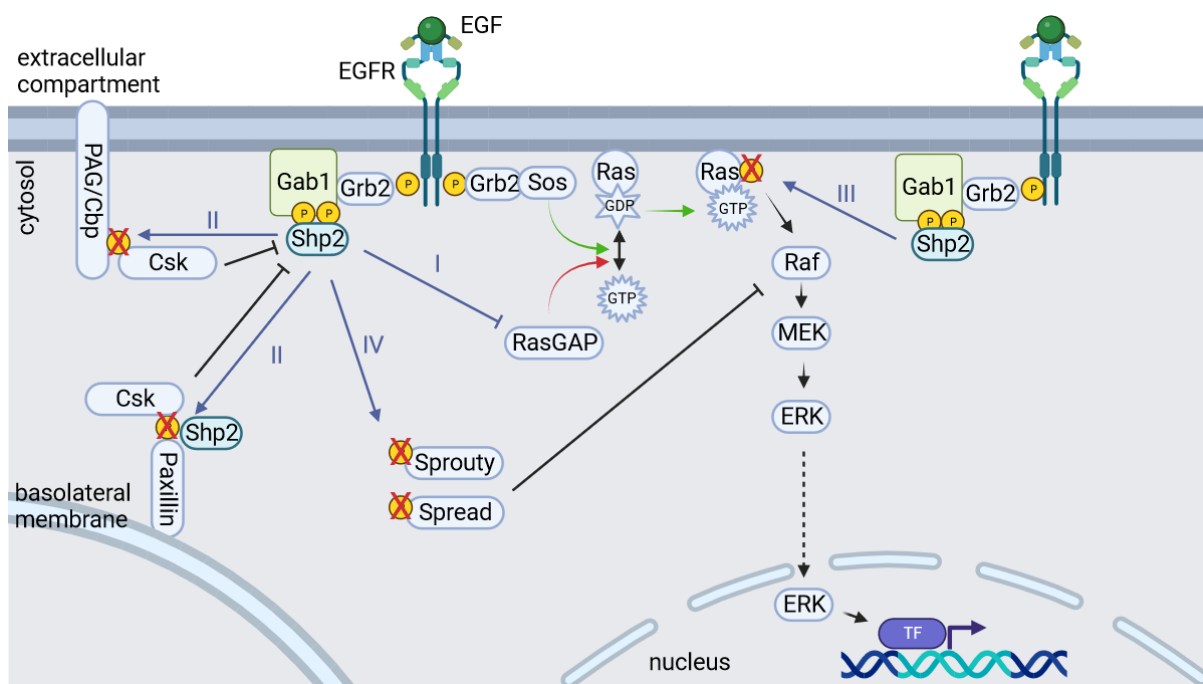
Figure 2 *Schematic of the Ras/MAPK pathway in context with Shp2 regulation. Upon binding of EGF to EGFR, the receptor dimerises and gets phosphorylated. Grb2 binds to those phosphosites and recruits Sos, a guanine nucleotide exchange factor. Sos catalyses the exchange from Ras·GDP to RAS·GTP, which leads to further activation of Raf, MEK and ERK. ERK transfers to the nucleus and regulates transcription. Membrane-associated Gab1 is bound to the receptor indirectly through Grb2. Src kinases phosphorylate Gab1, which recruits the phosphatase Shp2. Upon binding to Gab1, Shp2 gets activated and can dephosphorylate proteins in proximity (e.g. PAG/Cbp) or proteins at the basolateral membrane (e.g. Paxillin). The impact of those and other dephosphorylations are numbered in a roman system (blue arrows) and further discussed in the text. Created with biorender.com*

In addition, accessory proteins, such as Gab1 (Grb2-associated-binding protein1) and Shp2 (Src homology region 2 domain-containing phosphatase 2, also known as PTPN11) can regulate and enable crosstalk between pathways. Gab1 can be recruited by Grb2 (Bardelli et al. 1997; Nguyen et al. 1997; Holgado-Madruga et al. 1996) and further stabilised by membrane association with its PH domain (Pleckstrin Homology). Upon association, Gab1 becomes phosphorylated by Src family kinases (SFK) (Vemulapalli et al. 2021) and receptor kinases (Lehr2000, Lehr 1999, Gual 2000). Phosphorylated Gab1 can then bind to the phosphatase Shp2, leading to its activation (Cunnick et al. 2001).

Activated Shp2 has been reported to have several impacts on the Ras/MAPK pathway, illustrated in **Figure 2** (Niogret et al. 2019). (I) Shp2 dephosphorylates binding sites of p120RasGAP on Gab1 (Montagner et al. 2005) and the receptor (Agazie and Hayman 2003). This keeps p120RasGAP inactive, a Ras GTPase activating protein, and thereby prolongs the above-mentioned signalling cascade. (II) Activated Shp2 dephosphorylates the binding sites of Src-inactivating kinase, Csk, on paxillin (Ren et al. 2004) and PAG/Cbp (Zhang et al. 2004), which maintains the Gab1:Shp2 complex. It is further noteworthy that paxillin is a focal adhesion-associated protein on the basolateral membrane. This indicates the relocalisation of Gab1:Shp2 as a cytosolic complex (Furcht et al. 2015). (III) Shp2 positively regulates the pathway by dephosphorylating GTP-activated Ras on position 32. This phosphosite normally promotes engagement to the GTPase-activating protein. Dephosphorylation by Shp2 leads to prolonged signalling pathway activation (Bunda et al. 2015). (IV) Shp2 dephosphorylates SPROUTY and SPRED, which are negative regulators of the Ras/MAPK pathway (Hanafusa et al. 2004).

EGF also triggers other pathways, including phosphatidylinositol 3-kinase (PI3K)/Akt, regulating cell growth, migration, and apoptosis. PI3K (composed of the p85 (regulatory) and p110 (catalytic) subunits) binds to the activated receptor either directly or mediated through an adapter-complex, e.g. Grb2-Gab1 (Jiang and Ji 2019). The kinase converts the membrane lipid phosphatidylinositol - (3,4) - bisphosphate (PIP_2) to phosphatidylinositol (3,4,5) - triphosphate (PIP_3). The PH domains of the kinases Akt and PDK1 then bind to these newly formed lipids. Akt becomes activated in this membrane- bound complex and phosphorylates downstream targets, ultimately leading to upregulation of mTOR, a kinase known for its impact on cell survival, growth and proliferation. The pathway is illustrated in **Figure 3**, together with its effects on the MAPK pathway.
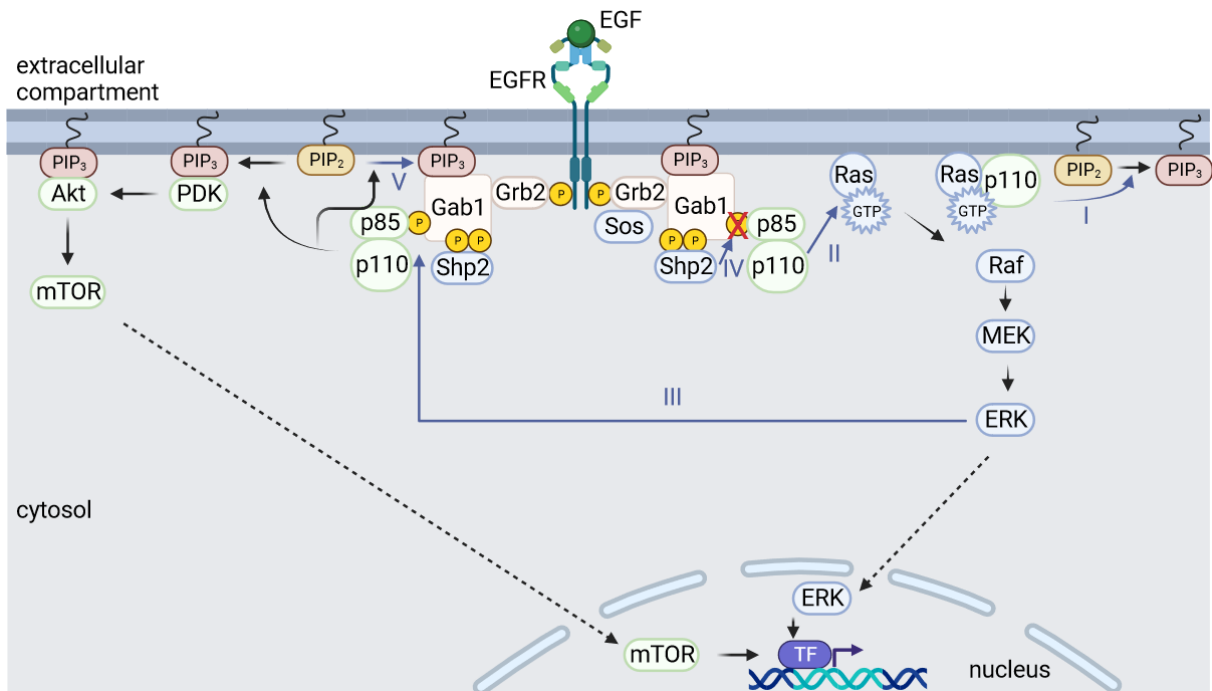
Figure 3 *Akt-PI3K pathway (green illustrated proteins) and the interplay with members of the Ras/MAPK pathway (blue illustrated proteins). Upon reception of EGF to the receptor, phosphorylation leads to the recruitment of Grb2. Gab1, which binds to phosphoinositol-3-phosphate (PIP3) with its PH domain and to Grb2 via its SH3 domains, can recruit p85 and Shp2, amongst others. p85 recruits the catalytic subunit p110 of PIP3K, which leads to the formation of PIP3 from PIP2. Other PH domain-containing proteins of that pathway (PDK and Akt) get activated and will, in turn, activate mTOR, which can transduce to the nucleus and regulate transcription. Implications of proteins involved in the Ras/MAPK pathway are referenced with roman numerals (blue arrows), corresponding to their description in the text. Created with biorender.com*

The MAPK and PI3K cell signalling pathways are intimately intertwined, as reported in a large body of literature. According to these reports, Ras·GTP can bring the catalytic subunit p110 to the membrane and activate it (I) (Rodriguez-Viciana et al. 1994; Chantley 2002), while PI3K can activate Ras and trigger the MAPK pathway (II) (Yart et al. 2002). The activated ERK can regulate the association of Gab1 with PI3K by phosphorylating Gab1, depending on the signalling growth factor, and thus up- or downregulate the PI3K pathway (III) (Yu et al. 2002; Zeke et al. 2022). Shp2, in turn, can dephosphorylate PI3K binding sites on Gab1 (IV) (Zhang et al. 2002). Elevated PI3K activity, in turn, promotes membrane association of Gab1 through its PH domain and thus leads to positive feedback in both pathways (V) (Rodrigues et al. 2000). Taken together, Gab1 has a crucial impact on the proliferation driving Ras/MAPK and the PI3K pathways, necessary for cell survival. On the other hand, this means that upregulation of Gab1 is also a critical contributor to tumorigenesis.

Gab1, Shp2, their homologues, IDPs and SH2 domains in cell signalling

Gab family proteins

The Grb2-associated binding protein (Gab) family was first discovered as a binding partner of Grb2 (Holgado-Madruga et al. 1996). Homologues Gab2 (Gu et al. 1998) and Gab3 (Wolf et al. 2002) were discovered later, and a putative fourth family member, Gab4, has been described as a transcript product. In *Drosophila* as well as *Caenorhabditis* orthologues Dos (Daughter of Sevenless) (Raabe et al. 1996) and Soc-1 (Suppressor of Clear), respectively, have been described and their schematic overview is depicted in **Figure 4**. The fundamental importance of Gab proteins is documented by the observation that Gab1-like proteins emerged around 1 billion years ago (Srivastava et al. 2010). Proteins of the Gab family take on crucial functions in signalling downstream of RTKs (Herbst et al. 1996; Schutzman et al. 2001). They are scaffolding adaptor proteins and possess, therefore, no enzymatic capacity. Instead, they bring associated enzymes into the proximity of their substrates. Common to all Gab family proteins is the folded N-terminal PH domain, a domain that binds to PIP_3 in the plasma membrane. C-terminally, a long intrinsically disordered (ID) tail with binding sites for a variety of signalling proteins provides this family with a typical tadpole-like topology. Phosphotyrosines on the ID-tail aid as binding sites for Src homology (SH2) domain (phosphotyrosine binding domain) containing proteins. Crk, PLCɣ, Rho GTPases and RasGAP binding sites, as well as for p85 PI3K and Shp2 are clustered (Liu and Rohrschneider 2002). Grb2 binds to Gab family members through its C-terminal SH3 domain (poly-proline binding domain). A tandem poly-proline motif in the Gab1 ID tail binds to the two SH3 domains present in Grb2 (Breithaupt et al. 2023). A Met receptor binding site (MBR) in Gab1 provides it with the ability to bind the Met-receptor directly. In contrast, recruitment to other receptors must be facilitated through Grb2.
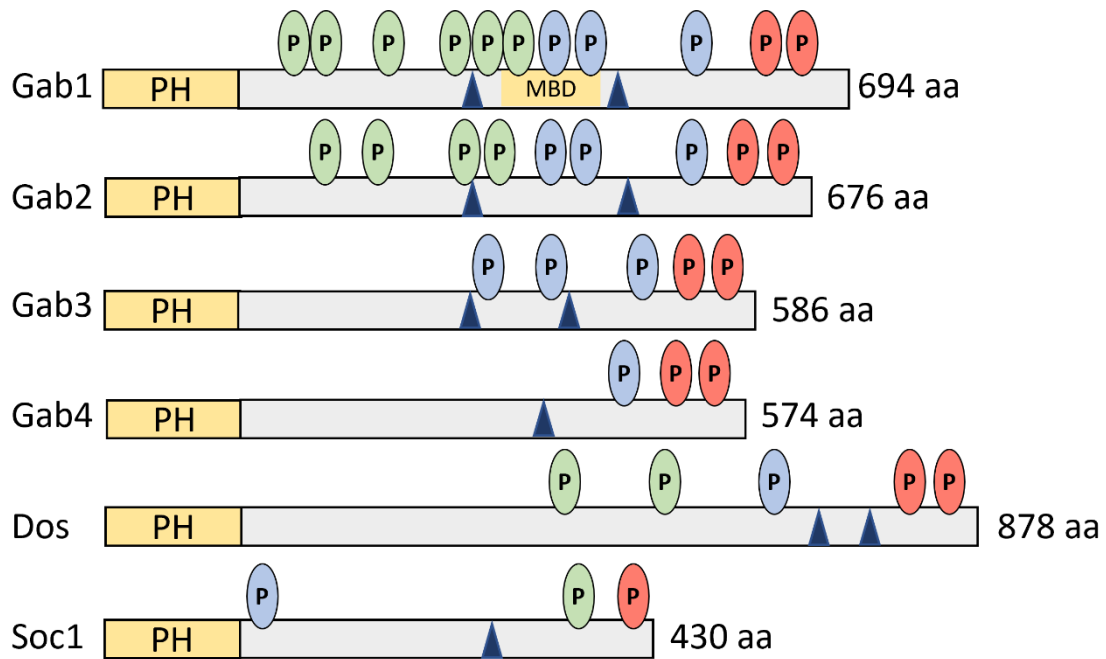
Figure 4 Schematic overview of Gab family proteins. Human homologues Gab1 – Gab4 (Gab4 is a putative protein – its PTMs are therefore assumed) and the orthologous of D. melanogaster (Daughter of Sevenless (Dos)) and C. elegans (Suppressor of Clear-1 (Soc-1)) are shown as bars. Common to all is the N-terminal conserved and folded PH domain (yellow). The intrinsically disordered tail (grey) shows several phosphorylation sites (oval with the letter P), as well as Grb2 SH3 domain binding sites (blue triangle). Gab1 has a Met binding domain (yellow box with MBD), which establishes direct contact with the Met receptor. The phosphorylation sites correspond to phosphotyrosines, which are SH2 domain binding partners. These sites are clustered on the tail according to the proteins they recruit. Proteins of the Jnk pathway (Crk, PLCγ, RasGAP and Rho GTPases) bind to the phosphotyrosines indicated in green. p85 of the (PI3K pathway) binds to the conserved phosphotyrosine sites indicated in blue. The red phosphotyrosines are conserved sites for the SH2 tandem domains of Shp2 (Ras/MAPK pathway). The length of each protein is indicated as well. The figure was modified from (Nakaoka and Komuro 2013)

As intrinsically disordered regions (IDRs), the tails of Gab family proteins do not have any distinct tertiary or quaternary structure, providing them with flexibility. The only secondary structural elements in the tail known so far are the poly-proline type II (PPII) and 3_{10} helices at the Grb2 binding motifs (Harkiolaki et al. 2009). Posttranslational modifications (PTMs) like phosphorylation on tyrosines, serines and threonines are typical for IDRs since enzymes, such as kinases (but also proteases), have sterically unhindered access. The N-terminal folding hypothesis (Simister et al. 2011) describes how tadpole-like signalling proteins potentially escape unregulated protein degradation. Upon synthesis on the ribosome, the N-terminal residues fold into the domain while incorporating regions of the tail into its centre. The clustered phosphotyrosines for specific SH2 domain-containing proteins thus appear in functional, specialised loops, reducing inherent flexibility, escaping proteolysis, and providing order (**Figure 5**). Another regulatory feature which supports this hypothesis is the translocation of Gab1. It has been shown that phosphorylation at S552 leads to translocation of Gab1 to the membrane (Eulenfeld and Schaper 2009). Phosphorylation of S552 leads to the release of the PH domain from the ID tail and opens up a free PIP3 binding site (Wolf et al. 2015).
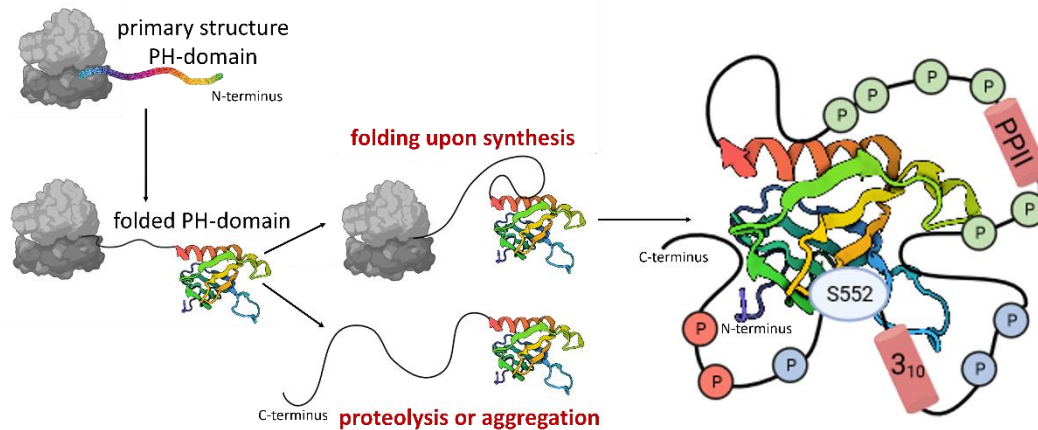
Figure 5 *N-terminal folding hypothesis. The primary structure of the designated PH domain (rainbow) folds during release from the ribosome (grey). The further synthesis could result in a tadpole-like free intrinsically disordered tail. This would be prone to degradation or aggregation. On the other hand, the N-terminal folding hypothesis postulates that the disordered tail folds upon release from the ribosome around the PH domain as a core. For Gab1 this leads to functional loops of this tail due to the clustering of the phosphorylation sites (letter P in the colours corresponding to* *Figure 4*). *Furthermore, S552 and the SH3domain binding sites (which are the only secondary structures known for the disordered tail of Gab1; 3_{10} and PPII) are depicted. The figure is adapted from (Simister et al. 2011; Lewitzky et al. 2012) and created with biorender.com.*

Intrinsically disordered proteins

Intrinsically disordered proteins (IDPs), or the shorter intrinsically disordered regions (IDRs) within a protein, lack secondary or tertiary structure. Their structural properties vary from extended (random coil-like) to collapsed (molten-globule-like), depending on their net charge, complexity and hydropathy. IDPs are found in all domains of life where an average of 2% of the archaeal, 4% of the bacterial and 33 % of the eucaryotic proteome are predicted to contain regions (of over 30 residues) of disorder (Ward et al. 2004). Differences within an evolutionary lineage are also described. For example, 44% of proteins in the human proteome contain regions of disorder longer than 30 residues (van der Lee et al. 2014). Extreme habitats seem to favour increased disorder, as discussed in (Xue et al. 2010), and even DNA and RNA-viruses show more extensive amounts of disorder compared to prokaryotes (Tokuriki et al. 2009). This might be due to the extreme compaction of the genome, the buffer effect of mutations introduced by low-fidelity polymerases and higher structural flexibility to interact with host proteins, making IDPs a better toolkit compared to folded proteins (Tokuriki et al. 2009). IDPs are often involved in biological processes primarily associated with differentiation, division, and protein synthesis (Dunker et al. 2015). In their role in cell signalling, they are found in one-to-many and many-to-one interactions with other proteins. For example, p53 interacts with more than 100 partners while adopting different conformations upon binding (one-to-many), whilst the structured protein, 14-3-3, undergoes many-to-one interactions with more than 200 different disordered regions (Alterovitz et al. 2020). These large interactomes of one protein provide a central node within a signalling

network of protein-protein interactions (PPI). Such nodes are called hub proteins and are further categorised as "party hubs", where a multitude of proteins bind simultaneously, or "data hubs", where they bind sequentially to the hub protein (Han et al. 2004).

IDPs can bind their partners in a multitude of manners. Disordered regions can undergo a transition towards an ordered state or may stay flexible in the complex. The structural transition can occur prior, during or after binding and can be further modulated by posttranslational modifications (PTMs). Whether an IDR is prone to undergo disorder-to-order transition or fuzziness can again be predicted from its sequence. IDPs in their native, crowded environment can also be seen in liquid-liquid phase separations, either alone or together with other proteins. These membrane-less compartments lead to enrichment and high local concentrations that can either be part of a physiological function, e.g. storage or signalling or might be an accelerator of disease (Wang et al. 2021).

SH2 domains

Besides HYB (Hakai pY-binding) and pyruvate kinase M2 and C2 domains, SH2 (Src-homology 2) and PTB domains (phosphotyrosine-binding) are among the most prominent phosphotyrosine binding domains. The PTB domains are around 100-170 amino acids (aa) long and have a central β-sandwich capped with an α-helix, similar to the fold of PH domains. They are found in approximately 60 human proteins (Uhlik et al. 2005). PTB domains recognize phosphotyrosines and select for N-terminal residues by the motif NPXpY (X = any amino acid; pY = phosphotyrosine). Worthy of note, some PTB domains bind targets independent of phosphorylation (Uhlik et al. 2005). The approximately 100 aa long SH2 domain is found in around 110 proteins and has a conserved fold of a central β-sheet with flanking α-helices to both sides **Figure 6**. In contrast to the PTB domain, it recognises pY plus additional C-terminal residues. SH2 domains and protein tyrosine kinases (PTKs) evolved together at the divergence of single-celled and multicellular organisms, suggesting that pY-signalling was a crucial driver of metazoan development (Shiu and Li 2004; Lim and Pawson 2010; Liu et al. 2011). In their role in pY-signalling, SH2 domains can facilitate localisation and complex formation as connectors, e.g. in the adaptor protein Grb2, and as functionally involved domains, e.g. as a conformation stabiliser in Shp2. However, SH2 domains are much more versatile. For example, the ancestral SH2 domains of Spt6 bind to phosphoserines/-threonines (pS/pT). Other SH2 domains, like the tensin family, have been reported to bind independently of phosphorylation on the tyrosine (Dai et al. 2011). Yet others bind their phosphopeptide in reverse orientation (Ng et al. 2008; Wang et al. 2018), and it has been reported that the N-SH2 domain of Shp2 can bind to two peptides

(Zhang et al. 2011) as well as recognise only N-terminal residues (Qin et al. 2005), like the PTB domain.
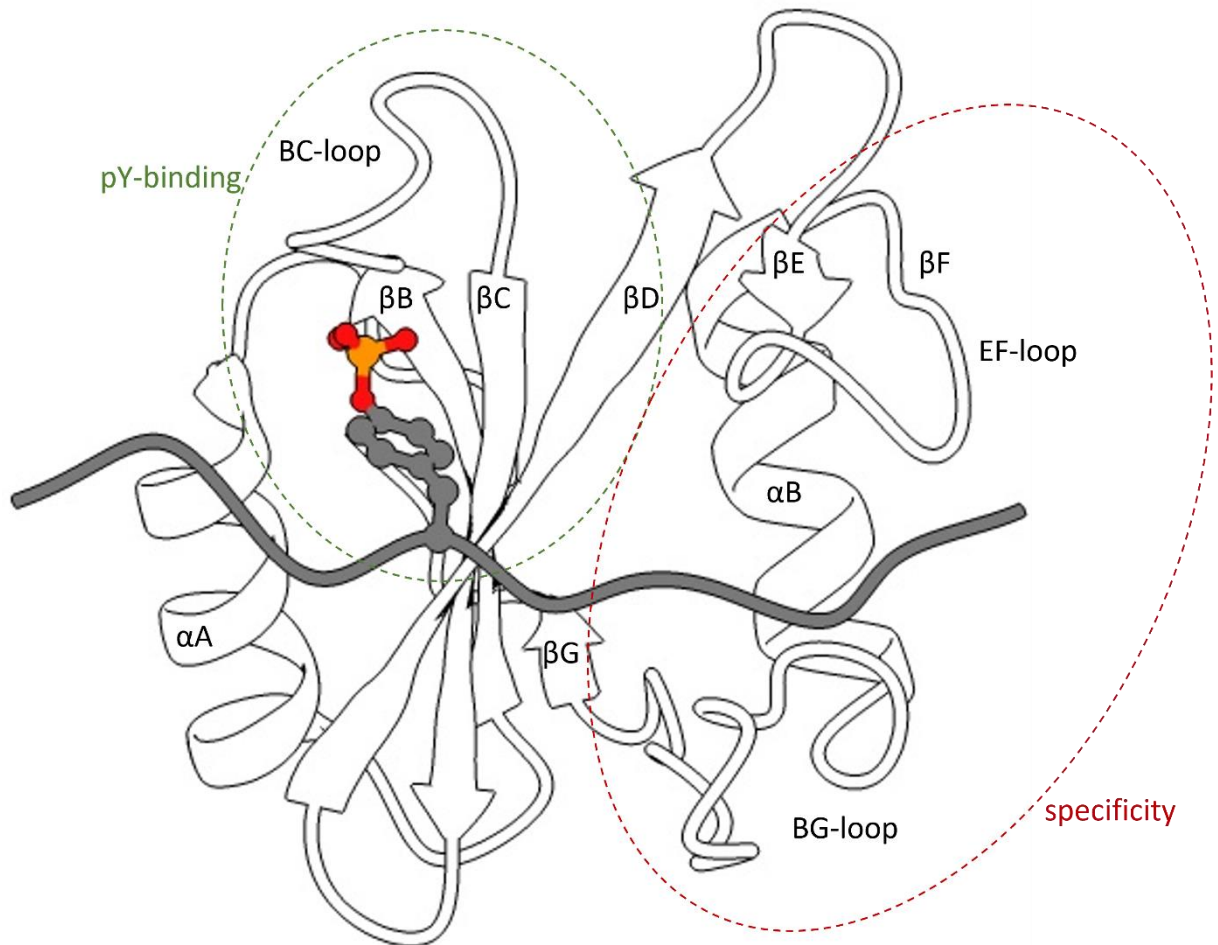
Figure 6 *Topology of the SH2 domain. N-SH2 of Shp2 (white, cartoon) bound to Gab1 fragment (grey, with phosphotyrosine as stick). Secondary structure elements are labelled (PDB: 4qsy (Gogl and Remenyi, unpublished)). The phosphotyrosine (coloured by element, phosphate orange, oxygen red, carbon grey) is coordinated by residues of the βB, βC, BC-loop and βD structural elements (green cycle). The specificity is provided through contact of the peptide to the EF- and BG-loop (red cycle). The illustration was done using chimeraX.*

SH2 domains have a highly conserved fold with the antiparallel β-sheet (βB-βD), continuing into two small antiparallel β-strands (βE, βF) connected by the DE- and EF-loops. The topology of the domain is shown in **Figure 6**. α-helices (αA and αB) are located N-terminally to the βB and C-terminally to the βF strand. Additional small β strands are also annotated (βG, βH). The secondary structure elements are connected with mobile loops that differ in length and amino acid composition and provide specificity and selectivity. The general binding is described to happen in a two-pronged way. The phosphotyrosine is bound in a groove (phosphotyrosine-binding pocket, green cycle in **Figure 6**) lined by βB, βC, βD, αA and the BC loop. Here, the phosphate is coordinated by ArgβB5 (βB) through two ionic bonds. Other positively charged amino acids in that groove are not as well conserved. In the SH2 domains of Shp1 and Shp2 pY gets further

stabilised by hydrogen bonds to serines of the BC-loop. A second binding site is constituted by the specificity pocket formed by CD-, DE-, EF-, BG-loops, βD and αB (red cycle in **Figure 6**). The bound peptide commonly adopts an extended conformation and orientates perpendicular to the central β-sheet with the C-terminal residues, with respect to pY, being coordinated in the specificity loop. The hydrophobic specificity pocket varies greatly and can even accommodate residues of position +5 /+6, relative to pY, in some cases (PLCg C-SH2 and Shp2), whereas, in most SH2 domains, shorter regions are covered (up to +3/ +4). The affinities measured for SH2 domains to their targets lie within 100 nM to 10 μM (Ladbury et al. 1995), thus showing modest affinity. Furthermore, the specificity for pY-binding sites is not very strong. This can be seen, for example, in the N- SH2 domain of Shp2, which binds to peptides from four different classes classified by binding sequence (Sweeney et al. 2005).

In ten proteins, two SH2 domains line up to build tandem SH2 domains. These show high affinity for their bisphosphorylated targets in the pM to nM range (Ottinger et al. 1998). They also have much greater specificity than single domains (Ottinger et al. 1998), presumably due to the local proximity of two binding sites. Additional selectivity may be provided through spacer length as well as differences in flexibility. The linker between tandem SH2 domains can vary greatly between proteins, from ten amino acids (aa) for Shp2, 200 aa for p85 and 65 aa for ZAP70. Most likely, these features are key regulating factors due to special restrictions.

It was recently shown that phosphatases can decrease the half-life of these complexes. Since the domains switch between binding with one of the SH2 domains and the tandem to the double-pY target, phosphatases act on the single domain bound state and dephosphorylate both pYs sequentially over time. This mechanism counteracts the tight binding to fine-tune kinase signalling (Goyette et al. 2022).

Shp2

The phosphotyrosine phosphatase Shp2 (PTPN11) (SH2 domain containing phosphatase 2) is the product of the *ptpn11 (*protein tyrosine phosphatase non-receptor type 11) gene. It possesses a tandem SH2 domain motif (comprised of the N-SH2 and C-SH2) interspaced by ten amino acids (aa). Followed by a phosphotyrosine phosphatase (PTP) domain and an intrinsically disordered tail. A homologue with the same domain structure and 68% sequence similarity is Shp1 (PTPN6) (Shen et al. 1991). Both proteins can take on two distinct conformations: an inactive, closed and an active, open. The current model of this conformational switch is depicted in **Figure 7**.
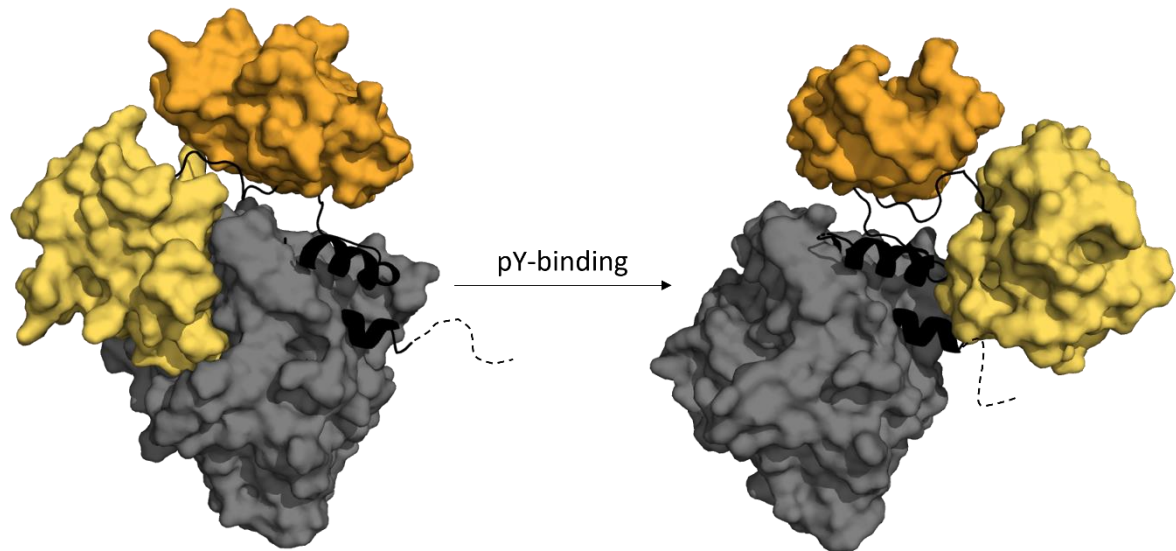
Figure 7 *Activation of Shp leads to conformational change. In the closed conformation, the DE loop of the N-SH2 domain is buried inside the PTP domain, blocking the active site of the phosphatase domain. Upon activation, e.g. through pY-peptide binding to the SH2 domains, the N-SH2 domain flips behind the C-SH2 domain. The C-SH2 domain also undergoes slight reorientation. Illustrated here are the inactive (closed) and active (open) conformation of Shp2 (PDB: 4dgp (Yu et al. 2013), 6crf (LaRochelle et al. 2018)) as surface models with the N-SH2 in yellow and the C-SH2 domain coloured in orange. The PTP domain is seen in grey. Linker regions between the domains are illustrated as cartoon in black. The C-terminal intrinsically disordered tail following the C-terminal helix of the PTP domain is implied as a dashed line. Illustration was done with PyMol.*

In the closed conformation, the N-SH2 domain blocks the active site of the PTP domain through its DE loop (Hof et al. 1998). Only in the activated state, either through mutation or peptide binding to the domain, the N-SH2 domain is released from the PTP domain and situated behind the C-SH2 domain. In this state, the active site of the PTP domain can accommodate substrates to catalyse the dephosphorylation. Despite a substantial number of structural publications, the activation mechanism of Shp2 is still not fully understood. Shp2 apparently can be activated by a high-affinity binding partner interacting solely with the N-SH2 domain (Hayashi et al. 2017; Sugimoto et al. 1994). For high-affinity N-SH2 activators, it has been postulated recently that a hydrogen bond network of the N-SH2 domain is perturbed upon binding of the phosphotyrosine. The perturbation propagates to the distant EF- and DE-loops, where conformational fluctuations lead to an opening of Shp2 and prevent the reclosure of the N-SH2:PTP interface (Marasco et al. 2021). After this initial binding of pY and the release of the N-SH2 domains from the phosphatase domain, the remaining C-terminal residues of the activating peptide can fully bind to the specificity pocket, leading to full activation. A central element of activation appears to be the opening of the central β-sheet, which relays the peptide-binding signal to the specificity loops (Anselmi and Hub 2021).

In addition, Shp2 can be activated by binding to bis-phosphorylated targets (Cunnick et al. 2001). One discriminating factor for either binding mode appears to be the spacer length between the two phosphotyrosines, as shown in **Figure 9**. Shp2 has a higher

affinity to binding partners with long pY-linkers of up to 49 aa as in IRS-1. Shorter linkers as the 14 aa in PD-1 (Marasco et al. 2020a; Patsoukis et al. 2020) or 11 aa in PDGFR (Ottinger et al. 1998) lead to spatial restrictions as the distance of the binding sites of the SH2 domains is fixed. In these cases, the N-SH2 and C-SH2 domains can bind to two distinct activator proteins in close proximity. This can be facilitated since these phosphorylated proteins are mostly dimerised receptors. This concept is shown in **Figure 8** for PD-1 but is also discussed for PDGFR (Ottinger et al. 1998; Patsoukis et al. 2020).
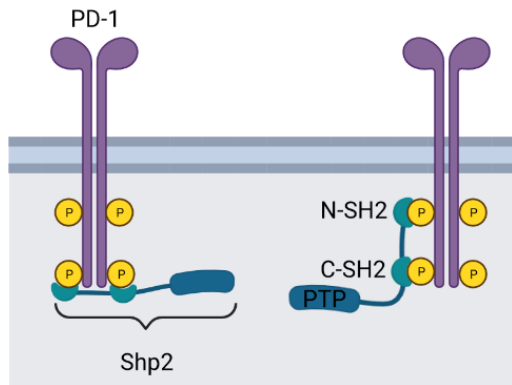
Figure 8 pY-linker length-dependent Shp binding modes. The 2:1 (receptor:Shp2) and the 1:1 binding modes are illustrated. Created with biorender.com.

```
IRS-1  NGLNYIDLDLVKDFKQCPQECTPEPQPPPPPPPHQPLGSGESSSTRRSSEDLSAYASISFQK
CagA   TEPIYAKVNKKKTGQVASPEEPIYTQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGLSASPEPIYATIDDLG
Gab1   KQVEYLDLDLDSGKSTPPRKQKSSGSGSSVADERVDYVVVDQQK
Gab2   GSVDYLALDFQPSSPSPHRKPSTSSVTSDEKVDYVQVDKEK
PD-1   FSVDYGELDFQWREKTPEPPVPCVPEQTEYATIVFPS
```

Figure 9 The linker regions between phosphotyrosine residues (pY) vary considerably. For CagA the typical EPIYA motif is underlined.

This 2:1 binding can still activate Shp2 but highlights the importance of a flexible linker in monomeric proteins. Even though Shp2 is a phosphotyrosine phosphatase when it is bound to its activating targets, like Gab1, the SH2 domains protect the phosphotyrosines from its enzymatic potential.

The phosphotyrosine phosphatase domain

Phosphotyrosine phosphatase domains (PTP) are well studied since the human genome encodes for 107 of those domains. The PTP domain of Shp2 is 270 aa long (247 – 517 aa). The catalytic active site (depicted in **Figure 10**) consists of the P-loop (458-467 aa), pY-recognition-loop (273-279 aa), WPD loop (Trp-Pro-Asp) (423-427 aa), Q-loop (504-510 aa) and E-loop (362–366 aa) and is occluded in the closed conformation by the DE-loop and adjacent βD and βE strands of the N-SH2 domain (58-63 aa). The P-loop (phosphate-binding loop) consists of the PTP signature motif $(C(X)_5R)$ with the catalytic cysteine and coordinating arginine (C459, R465), the pY-recognition loop discriminates

between phosphoserine (pS), phosphothreonine (pT) and pY (phosphotyrosine), the flexible WPD-loop provides the catalytic aspartate (D425), while its mobility is supported by the E-loop, lastly, the Q-loop's glutamine residue (not shown in **Figure 10**) coordinates water molecules necessary for the hydrolysis (Tautz et al. 2013; Hof et al. 1998; Yu et al. 2014).
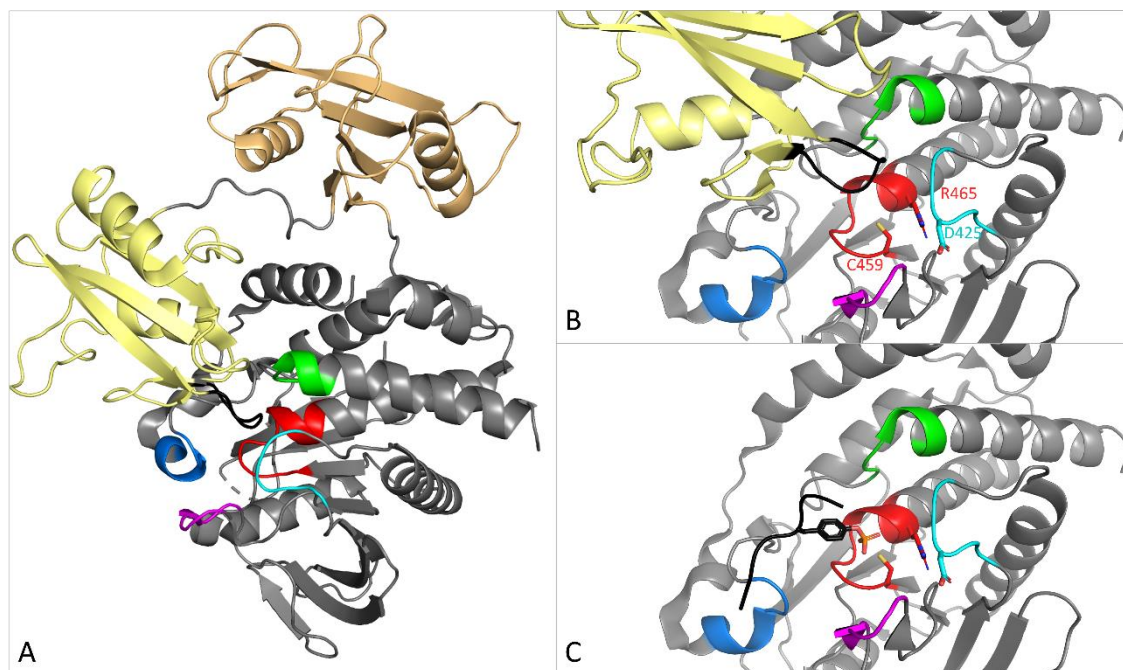
Figure 10 Catalytic site of Shp2. Shown is the catalytic site of Shp2 (cartoon) in closed (A, B) and substrate-bound state (C). The N-SH2 domain is shown in yellow with the DE-loop in black (A, B), the C-SH2 is depicted in orange, the PTP domain in grey (PDB: 4dgp (Yu et al. 2013)). The DE-loop is occluding the active site. The active site consists of the WPD-loop (red), the pY-recognition loop (blue), P-loop (red), Q-loop (green) and E-loop (magenta). In the substrate-bound state (C) the N- SH2 is not occupying the active site, instead, at the same position the peptide (black) is coordinated (IRS-1 peptide from the PDB: 7ppm (Zeke et al. 2022)). Created with PyMol.

The catalytical mechanism is depicted in **Figure 11**, showing the nucleophilic attack of C459 on the substrate's phosphate in the first step of the reaction. Here, the phosphate is coordinated by R465. D425 provides a proton to the releasing OH^- group, leading to free water. In the second step, a water molecule is deprotonated by D425, and the resulting OH^- group leads to the hydrolysis of the thiophosphate bond, which results in the release of free phosphate (Tautz and Sergienko 2013).
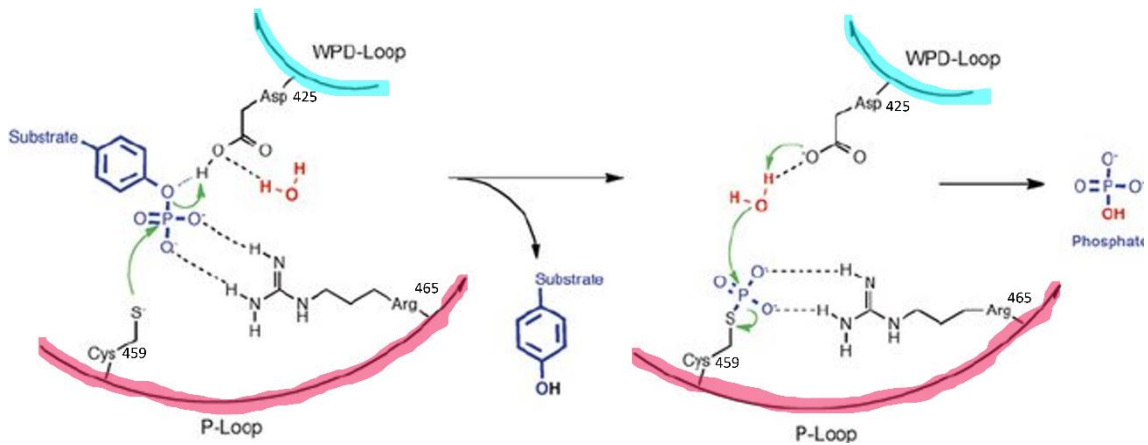
Figure 11 Catalytic mechanism of dephosphorylation. The P-loop (red) is needed to bind the phosphate through its R465, while C459 attacks the substrate's phosphate (blue) nucleophilically. The D425 acts as an acid in the first step, where it provides a proton to release a water molecule. In the second step, the thio-phosphate is hydrolysed by an OH- after D425 abstracted a proton from a water molecule. This attack leads to the release of free phosphate. Picture is taken and adapted from (Tautz and Sergienko 2013).

In general, Shp2 is a discriminating enzyme regarding its phosphatase substrates. A preference for negatively charged amino acids at -4 to -1 and serine/threonine phosphorylation at the +4 and +5 (relative to pY) followed by a proline was just recently shown (Zeke et al. 2022; Hartman et al. 2020), postulating how serine-/threonine kinases like the MAPKs might regulate the pathway crosstalk. One example is the phosphotyrosine binding site pY472 on Gab1 for the regulatory subunit of PI3K. This site has a threonine at the +4 position that, upon phosphorylation by MAPKs, makes pY472 a preferred target for the dephosphorylation of Shp2. This, in turn, would lead to a loss of binding capacity between Gab1 and PI3K (Zeke et al. 2022).

These very recent findings show that even though intensive research has already been carried out on Shp2, much functionally relevant information is still missing.

Differences between Shp1 and Shp2

Shp1 and Shp2 are the only cytosolic phosphotyrosine phosphatases (PTPs) containing a tandem SH2 domain, although others with one SH2 domain, like tensin and SHIP, are known. Both homologues share 53 % sequence identity throughout their entire length (EMBOSS – global pairwise alignment). **Figure 12** illustrates the identity between the domains.
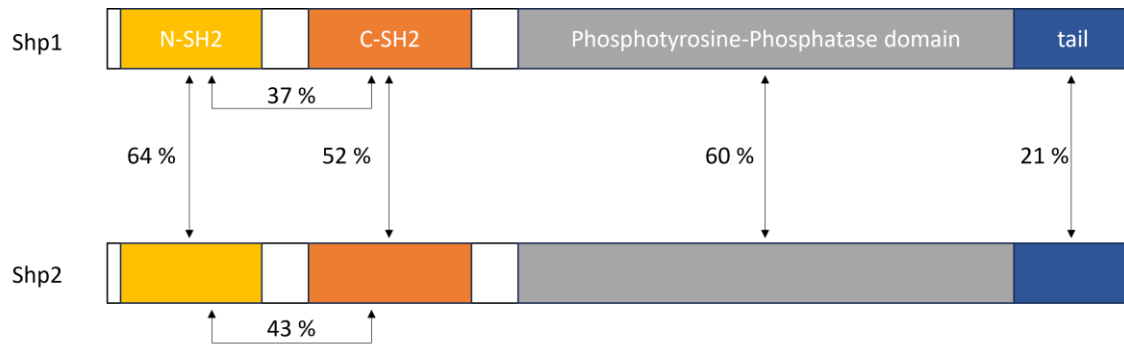
Figure 12 *Comparison of the sequence identity between Shp1 and Shp2. The two homologues are depicted schematically, with the N-SH2 in yellow, the C-SH2 in orange, the PTP domain in grey, and the intrinsically disordered tail in blue. The identity between the sequences of both homologues is indicated. The sequence identity was calculated using the EBLOSUM62 matrix with a gap penalty of 10 and a penalty for extension of 0.5.*

The two phosphatases behave quite differently in cell signalling. While Shp2 (inactive in the basal cellular state) is a positive regulator, as described above, Shp1 leads to the termination of cell signalling through its basal active state. The mechanism behind these differences is still not discovered. The following sections try to collect some reasons for this.

Some mechanisms appear to lie in the disordered tails of both homologues. As seen in **Figure 12**, the tail shows greater variation in its sequence. This may be explained by its disordered nature, which does not rely on a conserved fold. However, it provides the two proteins with specific features.

Shp1 is mainly expressed in hematopoietic cells, while Shp2 is more widely expressed (Neel et al. 2003). Subcellular localisation also seems to be a factor since Shp1 in hematopoietic cells is cytosolic, while in non-hematopoietic cells, it is located in the nucleus, where it positively regulates cell signalling (Poole and Jones 2005). To achieve different localisations of Shp1 depending on the cell type, it is speculated that the nucleus localisation sequence (NLS) of Shp1 is either a splicing variant, like Shp1-L, which lacks this motif (Poole and Jones 2005) or is affected by PTMs. S591 near the NLS might get phosphorylated and prevent nucleus entry (Craggs and Kellie 2001). The NLS is absent in Shp2, keeping it cytosolic. The same serine residue (S591) in its phosphorylated state leads to the decrease of Shp1 activity, probably by recruiting a phosphoserine-binding protein, which is yet unknown (Poole and Jones 2005). Although Shp2 also possesses this residue, it is presumably not associated with regulating its activity (Poole and Jones 2005). Shp1 is active under basal cellular conditions, which must be down-regulated upon activation of the signalling cascade(Jones et al. 2004). Therefore, the switch-off model through pS591 is widely accepted. How the constitutive active state is accomplished is only speculation, but binding to a phosphorylated peptide via its SH2 domain is likely (Jones et al. 2004; Sathish et al. 2001). Another feature that was

only described for Shp2 so far is its SUMOylation at K590 at the very C-terminal region of the ID tail. It is reportedly essential for Shp2:Gab1 complex formation and, thus, full activation of ERK (Deng et al. 2015). Interestingly, the SUMO binding sites of Gab1 are located at and near the PH domain. This provides another hint for the N-terminal folding hypothesis of Gab family proteins, which brings the Shp2 binding sites of Gab1 (pY627 and pY659) close to the SUMO-binding motifs (SIM57 and SIM164).

Another aspect of the different functions between both homologs is their selectivity for activating peptides. Both N-SH2 domains bind phosphopeptides that recognise the -2 to +6 position (Imhof et al. 2006; Sweeney et al. 2005). The N-SH2 domain of Shp1 prefers positively charged residue in the +4 and +5 positions, whereas Shp2 N-SH2 prefers hydrophobic residues in these positions (Imhof et al. 2006). Similarly, the Shp1 C-SH2 domain prefers positively charged side chains at +4 in contrast to the hydrophobic residues preferred by Shp2 C-SH2 (Imhof et al. 2006). The positions from -2 to +3 are surprisingly similar between the C-SH2 domains (Sweeney et al. 2005). The N-SH2 domain has four classes of binding partners, where the class I is similar to the C-SH2 domains (Sweeney et al. 2005), whereas the class IV peptides (the high-affinity binders) are similar to the motif found to bind Shp1 N-SH2 (Beebe et al. 2000). It was shown that most phosphopeptides bind with a clear preference to one of those four domains. Nevertheless, peptides with promiscuous binding to all four domains could also be identified (K_D=2-5 µM) (Sweeney et al. 2005). Although Gab1 is thought to be a specific target for Shp2, it could be shown in vitro and in vivo, that Shp1 is associated with the Shp2:Gab1 complex. How that interaction is facilitated is not known, nevertheless, it is most likely facilitated through a Shp1:Shp2 interaction (Wang et al. 2006).

Table 1 *Preferred binding motifs of the SH2 domains of Shp1 and Shp2. Amino acids written with capital letters are preferred over the ones written in lowercase. X can be any amino acid except glycine and proline. * (Sweeney et al. 2005); ** (Beebe et al. 2000)*

		-2	-1	pY	+1	+2	+3
N-SH2 Shp2*	Class I	Ile/ Leu/ Val/ met	X		Thr/ Val/ Ala	X	Ile/ Val/ Leu/ phe
	Class II	Trp	Met/ Thr/ val		tyr/ arg	Ile/ Leu	X
	Class III	Ile/ Val	X		Leu/ Met/ Thr	Tyr	Ala/ Pro/ Thr/ Ser/ gly
	Class IV	Ile/ Val/ Leu	X		Phe/ Met	X	Pro
N-SH2 Shp1**		Leu	Tyr/ His		Met/ Phe	X	Phe/ Met
C-SH2 Shp2*		Thr/ Val/ Ile/ tyr	X		Ala/ ser/ thr/ val	X	Ile/ val/ leu
C-SH2 Shp1*		Thr/ val/ ile	X		Ala/ thr	X	Leu/ met/ val

Structurally, both proteins show similar relative domain orientations of the N-SH2 and PTP domains in the unliganded closed states (**Figure 13**), indicating similar inactivation mechanisms, but the intervening C-SH2 domains adopt alternative positions. In contrast, the open (active) states show similar relative domain orientations of the C-SH2 domains but different positions of the N-SH2 domains. Whereas the open Shp2 state was achieved using the constitutively active E76K mutant protein (LaRochelle et al. 2018), it is not clear why Shp1 crystallised in the open form (Wang et al. 2011). It should be noted that full-length protein Shp1 was used, and sulphate ions from the crystallisation buffer were found in the ligand binding sites of both SH2 domains.
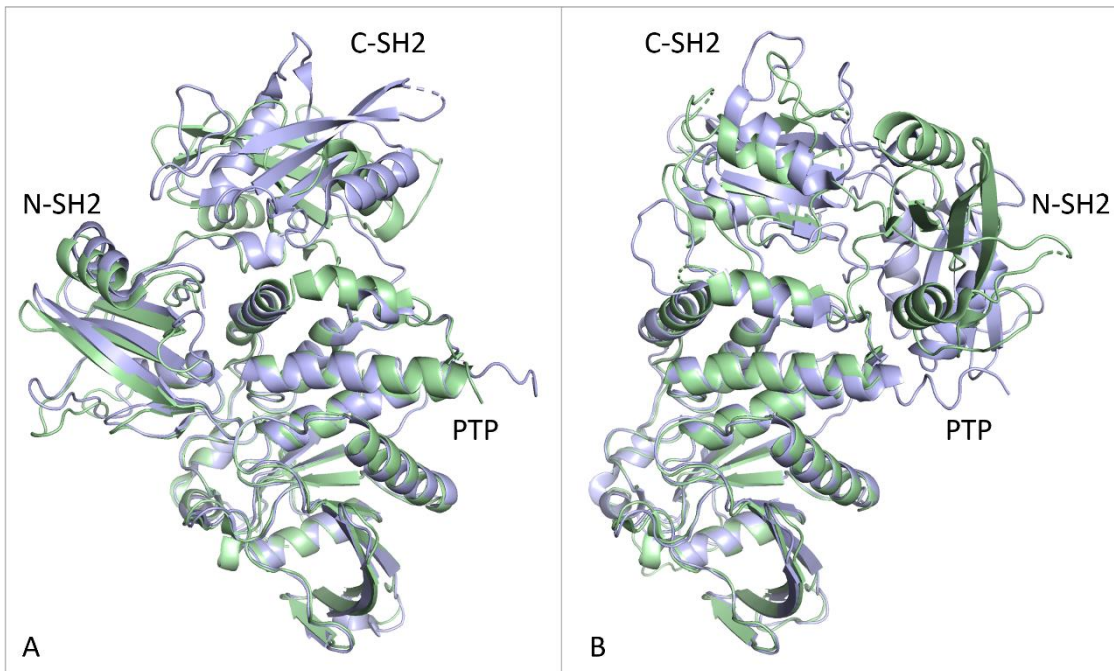
Figure 13 Superposition of the closed (A) and open (B) conformation of Shp1 and Shp2 (cartoon). A) In the closed conformation, the Shp1 (blue, PDB: 2b3o (Yang et al. 2003)) and Shp2 (green, PDB: 4dgp (Yu et al. 2013)) structures align well with alternative positions of the C-SH2 domain. B) In the open conformation, the C-SH2 domains adopt similar relative positions, but the N-SH2 domains show greater variance (Shp1, PDB: 3ps5 (Wang et al. 2011); Shp2, PDB: 6crf (LaRochelle et al. 2018)). Created with PyMol.

Whether Shp1 and/or Shp2 can actually get activated by inorganic phosphate is not known. The SH2 domains of Shp2 show binding to inorganic phosphate and free phosphotyrosine, even with structural changes of the N-SH2 domain identical to phospho-peptide binding (Marasco et al. 2021). However, its relevance for activation was not shown per se. The crystallisation of the sulphate ion complexed Shp1 open state indicates that inorganic phosphate should actually be able to activate the protein. It would further be interesting to know whether Shp1, as a negative regulator of cell signalling, might be easier to activate than Shp2. Unfortunately, no data could be found that addresses that question.

Gab and Shp2 in physiology and pathology
Gab
The human Gab family proteins are widely expressed. Especially high expression of Gab1 and Gab2 is seen in tissues of the brain, gastrointestinal, gallbladder, male and female reproductive systems, skin and lymphatic system; Gab2 is also seen in hematopoietic cells (Wang et al. 2017). Gab3 is highly expressed in fat cell tissue (Human protein atlas) and macrophages, dendritic cells and T-cells (Wang et al. 2017). While Gab1 deficient mice die at mid-to-late gestation with developmental defects in the heart, liver, muscle, placenta, and skin, similar to impaired RTK signalling (Itoh et al. 2000; Sachs et al. 2000), Gab2$^{-/-}$ (Nishida et al. 2002) and Gab3$^{-/-}$ (Seiffert et al. 2003)

mice are vital and have an average life span. Gab2-deficient mice have reduced numbers of mast cells leading to impaired allergic reactions (Nishida et al. 2002), and double-knockout Gab2/3$^{-/-}$ mice develop colitis, whereas single knockouts do not (Wang et al. 2019b). These findings imply the importance of Gab family proteins in embryology for physiologic development. Due to their involvement in signalling, they are also critical factors in tumorigenesis. Gab1 is closely connected to the oncogene c-Met (Mood et al. 2006; Seiden-Long et al. 2008) and was found to be upregulated in head and neck (Hoeben et al. 2013), breast (Wang et al. 2019a) and colorectal cancer (Seiden-Long et al. 2008) as well as glioma (Liu et al. 2014; Shao et al. 2018) and prostate cancer (Human protein atlas). Recently, a Gab1 (1-528 aa)-ABL1 (28-1130 aa) fusion gene was found and predicted to be involved in paediatric malignances (Panagopoulos et al. 2020). Mutations of Gab1 found in cancer patients are attributed to binding sites for pathway- regulatory proteins (Ortiz-Padilla et al. 2013). Gab2 overexpression was reportedly found in breast cancer (Daly et al. 2002), and gene locus amplification has been attributed to a variety of cancers (Huang et al. 2002). Gab3 is implicated in ovarian (Berkel and Cacan 2021) and colorectal cancer (Xiang et al. 2017).

Shp2

While Gab proteins are overexpressed in cancer, Shp2 is upregulated in diabetes (Saint-Laurent et al. 2022). Nevertheless, Shp2 does have a significant impact on cancer development and other diseases through mutations. 50% of Noonan-Syndrome (NS) patients, characterised by low hight, chest deformity and heart defects, show germline mutations of Shp2 (Tartaglia et al. 2001). These mutations mostly disrupt the N-SH2:PTP interface or lead to a higher binding affinity for activating peptides (Keilhack et al. 2005; Tartaglia et al. 2006). These gain-of-function (GOF) mutations are in contrast to the loss-of-function (LOF) mutations seen in Noonan syndrome with multiple lentigines (NSML) (formally LEOPARD syndrome). Here, Shp2 makes up 90 % of germline mutations, which are causing the RASopathy similar to NS but with additional freckle-like markings. These LOF mutations are found in the PTP domain, reducing the phosphatase activity (Kontaridis et al. 2006). Why LOF and GOF mutations result in similar phenotypical output was investigated by (Yu et al. 2013; Yu et al. 2014). The mutations of the PTP domain weaken the interaction with the N-SH2, leading to hyperactivation that compensates for the impaired PTP activity. The aggressive leukaemia of infants and children, juvenile myelomonocytic leukaemia (JMML), can either be sporadic or associated with the above-mentioned diseases. Shp2 is mutated in 35% of sporadic JMML at the N-SH2:PTP interface, destabilising the closed conformation (Wang et al. 2009). Shp2 is also involved in stomach cancer in succession of infection with

Heliobacter pylori. The bacterial CagA protein is released into the cell, where it acts as an activator of Shp2 and upregulates processes thereafter (Hayashi et al. 2017).

Targeting Gab1 and Shp2

As discussed, Gab1 and Shp2 have been implicated in several cancers and other diseases and are, therefore, sought-after targets in therapy. Targeting an IDP, such as Gab1, is of utmost complexity since no binding groove for small molecules can be defined. On the other hand, targeting one player (a hub protein, e.g. Gab1) and eliminating the signalling of several downstream events can have advantages if all implications are known or the drug is selective to the affected tissue. One approach could be prohibiting the membrane localisation of Gab and thus its phosphorylation by RTKs by targeting the PH domain (Chen et al. 2015; Seda et al. 2021). Such drugs need to be highly specific for the Gab1 PH domain to prevent disruption of other PH domain- containing proteins, like IRS1. Stapled peptides are another idea to regulate the binding of Gab partners (Verdine and Hilinski 2012). Stapled peptides are forced into helicity by an internal crosslinker. These helical parts are essential for access to the cell, better thermodynamics, and prolonged half-life. Another part of the peptide is responsible for binding specifically to its target. Phosphorylated binding sites of Gab proteins could thus be covered and prevent downstream signalling more specifically (Verma et al. 2012).

Much more research has been done on inhibiting Shp2. In the early years from 2006 to 2015, researchers focused on targeting the catalytic site of the PTP domain. A great hurdle has been its homology to other PTPs (side effects) and its high polarity, leading to poor cell access of potential ligands. Novartis identified an allosteric site at the interface of all three domains that keeps Shp2 in its closed conformation (Chen et al. 2016). They developed the inhibitors Shp099 and TNO155 and entered into clinical trials (LaMarche et al. 2020). Alternative allosteric sites have been found at the N-SH2:PTP and C-SH2:PTP interfaces (Fodor et al. 2018; Wu et al. 2019). Applying two drugs that bind to different sites (like Shp099 and Shp504) improves the inhibitory effect (Fodor et al. 2018). Another approach is to target the N-SH2 and C-SH2 domains in order to block the binding sites for activating phosphopeptides, which has been achieved using nanobodies (Sha et al. 2013) and artificial peptides derived from native binding peptide IRS-1 (Bobone et al. 2021). Another recently evolving therapeutic strategy is Proteolysis Targeting Chimera (PROTAC) (Sakamoto et al. 2001), where a protein of interest (POI) is labelled with ubiquitin to clear it from the cell by proteolysis. For this, a molecule with a specific binding site for the POI connected by a linker to a recognition site for an endogenous E3-ligase, is designed. This approach has also been applied to Shp2 by

using inhibitors like Shp099 as specific Shp2-targeting entities (Wang et al. 2020). All these strategies focus on inhibiting Shp2, which is in line with most of the research done on Shp2's implication in cancer and other disease.

Nevertheless, evidence that Shp2 can also act as a tumour suppressor is emerging. This could be seen in liver and colorectal cancer (Chen et al. 2020; Bard-Chapeau et al. 2011). Elevated activity of Shp2 could also be of potential interest in tumours with hyperactivated STAT5, as Shp2 can dephosphorylate it and prevent further signal transduction (Chen et al. 2003; Rani and Murphy 2016; Verma et al. 2012). Therefore, strategies for activation of Shp2 should be considered as well.

Aim and concept of the study

The purpose of this thesis was to shine a light on the interaction of the disordered tail of Gab1 with its cognate binding partner Shp2. A detailed understanding of this particular interaction is necessary to further understand:

- how Shp2 discriminates its activators

- how this interaction can be targeted

- how Gab1 differs from other Shp2 activator proteins

- how a doubly phosphorylated peptide gets bound by an SH2 domain tandem

- how tandem SH2 domains with a short spacer bind to a partner with a long phosphotyrosine linker

As the interaction between both proteins is not only physiological but also found in pathology, structural und biochemical data are of importance. The overall challenge of this project was to generate structural data, which were intended to support findings and to complete the picture of the interaction. Three main questions were addressed during the project:

I)   How does the phosphatase Shp2 bind to the doubly phosphorylated Gab1 tail?

         The results addressing this question can be found in chapters 2, 3, 4 and 5.

II)   Does the intrinsically disordered tail of Gab1 adopt any structure prior to or upon binding to the SH2 domains of Shp2?

         The results addressing this question can be found in chapters 1, 2, 3 and 4.

III)   Does the phosphatase Shp2 undergo any structural changes upon binding to Gab1?

The results addressing this question can be found in chapters 2, 3, 4 and 5.

To address these questions, the following methods and concepts were applied:

I)   Constructs of the tandem SH2 domain, the single SH2 domains, binding to double phosphorylated Gab1 peptides and mono-phosphorylated Gab1 peptides were analysed. For this ITC, NMR, X-ray crystallography and electron diffraction on micro crystals (microED) were employed.

II)   Phosphorylated and unphosphorylated Gab1 peptides were used either as mono-Y/pY or double-Y/pY. Additionally, tandem and single SH2 domain proteins were used as complex partners. The structural changes of the intrinsically disordered Gab1 peptide were monitored by CD, NMR, and MicroED. In addition, computational prediction, e.g. by AlphaFold, was done.

III)   The dynamic changes of the single SH2 domains were analysed in complex with a doubly phosphorylated Gab1 peptide with NMR. microED and further NMR experiments on the doubly phosphorylated Gab1 peptide with tandem SH2 domains were also used. ITC was conducted on the doubly phosphorylated Gab1 peptide interacting with the tandem SH2 domains, which has either a binding-competent deprived mutant in the N- SH2 or C-SH2.

In the following **Figure 14** all constructs used in the thesis are schematically shown.
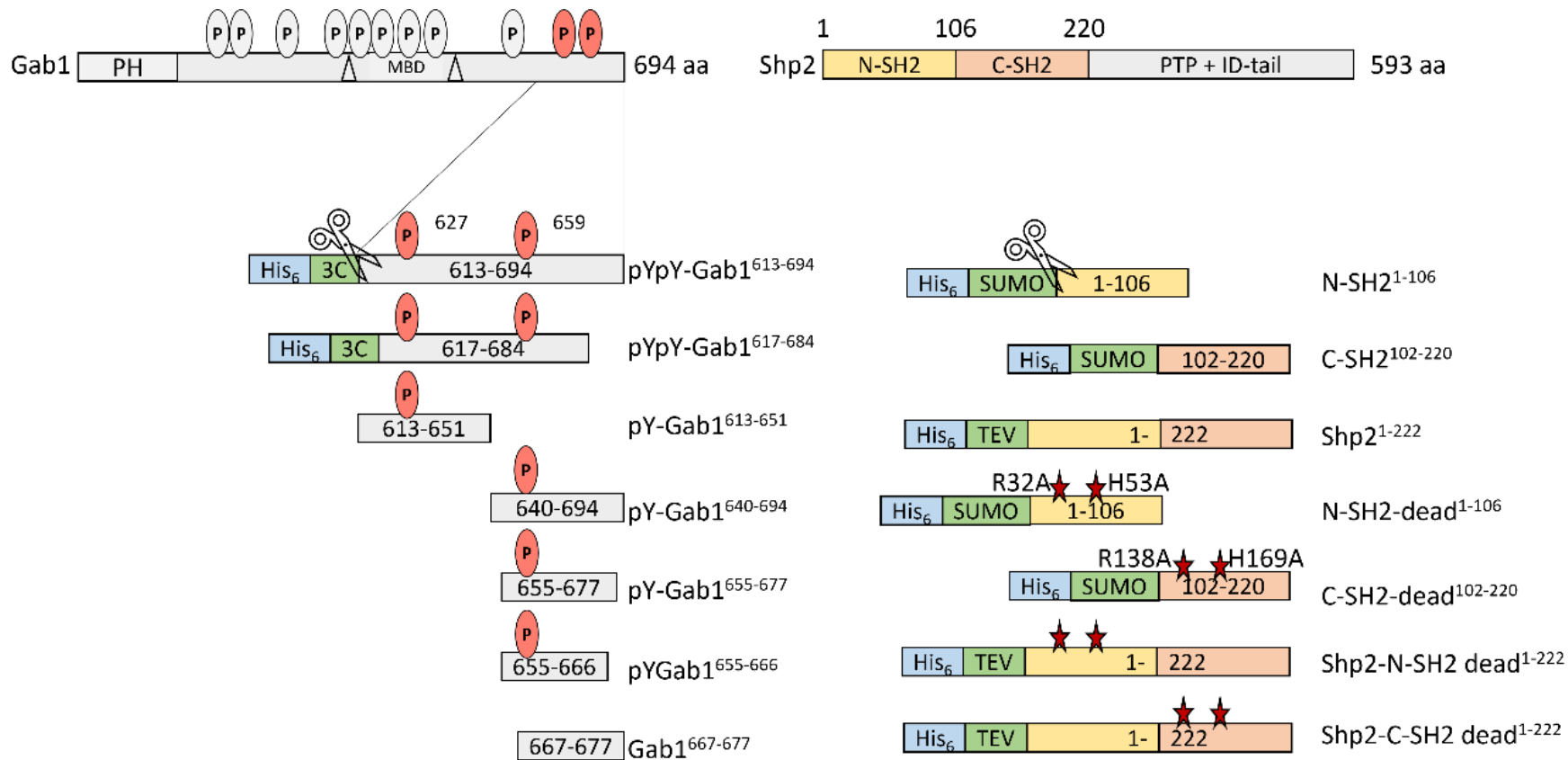
Figure 14 *Schematic overview of all used constructs. Gab1 and SHp2 constructs have been expressed and purified in case they are depicted with affinity tags. His6 (blue box) and protease recognition tags (green box) have been removed upon protease cleavage (3C, SUMO, TEV). Those Gab1 constructs have been used in the phosphorylated and unphosphorylated states. Both Gab1 and Shp2 constructs eligible to purify have been expressed with and without NMR-sensitive isotopes. Other constructs (Gab1 monophosphorylated peptides) have been ordered from commercial suppliers.*

Material and Methods

Equipment, chemicals and crystallisation kits

Equipment	Manufacturer
1,5 ml Axygen reaction tubes	Corning
500 MHz NMR spectrometer (Avance DRX)	Bruker
500 ml filter unit (22 µm)	Corning
600 MHz-NMR spekcrometer (Avance III)	Bruker
800 MHz-NMR spectrometer (Avance III)	Bruker
Äkta pure	GE Healthcare
autoclave VX150	Systec
balance Pioneer	OHAUS
balance Scout	OHAUS
BioPhotometer D30	Eppendorf
Blockthermostat 2099	Liebisch
Canoscan LiDE 110	Canon
centrifuge 5427 R	Eppendorf
centrifuge 5804 R	Eppendorf
centrifuge Avanti JXN-26	Beckman Coulter
centrifuge J6-MI	Beckman Coulter
centrifuge Megafuge 1.0R	Hereaus
CMOS-based CetaD camera	Thermo Fisher Scientific
Cryostream 800	Oxford Cryssystems
cuvette single-use	LABSOLUTE
dest. water system MilliQ	Merck
Digital IXUS 80 IS	Canon
EasyXtal 15-well Tool	QIAGEN
falcon tubes (15 ml, 50 ml)	Corning
filter unit for syringe (0.2 µm)	Sartorius
fraction collector F9-R	GE Healthcare
Glacios Cryo-TEM	Thermo Fisher Scientific
HiLoad 16/600 Superdex S30	GE Healthcare
HiLoad 16/600 Superdex S75	GE Healthcare
HiTrap Column (IMAC FF, Q HP)	GE Healthcare
holy grid R2/1	quantifoil
HyPix-Arc 150°	Rigaku
incubator 6030	Hereaus
incubator Innova 4200	New Brunswick Scientific
incubator Innova 43	New Brunswick Scientific
KL 1500 LCD	Leica
Macrosep Advanced Centrifugal Devices (1k MWCO)	PALL
Magnetrührer MR Hei-Standard	Heidolph
MicroCaliTC200	GE Healthcare
Microsys 4000	Zinsser Analytic
Mini-PROTEAN III Cell	Bio-Rad
Minstrel DTUV	Rigaku

MRC 3-well plates	SWISSCI
NanoDrop One Microvolume Spectrometer	Thermo Fisher Scientific
NanoVue Plus Spectrometer	GE Healthcare
PELCO easiGlow	TedPella
pH meter 765 Calimatic	Knick
pH meter SevenDirectSD20	Mettler Toledo
pipette tips	Corning
Plastipak syringe (5ml, 50 ml)	BD
PROTEANII xi Cell	Bio-Rad
QIAprep spin miniprep kit	Quiagen
QIAquick gel extraction kit	Quiagen
QIAquick PCR purification kit	Quiagen
Quick-Change site-directed mutagenesis kit	Agilent
RA-Micro7HF	Rigaku
rotor JA.25.50	Beckman Coulter
rotor JLA 16.250	Beckman Coulter
rotor JS-4.2	Beckman Coulter
Sonicator Sonoplus HD 2200	Bandelin
Spectra/Por dialyse tube (1, 3.5 kDa MWCO)	VWR Scientific
standard ashless blotting paper Ø55/20mm, Grade 595	TedPella
stereomicroscope M165C	Leica
TCl-CryoProbe	Bruker
tumbling shaker PMR-150	Grant-Bio
tumbling shaker S0500	VWR Scientific
Ultra-Micro-cuvette	Hellma Analytics
Ultrospec 7000 photometer	biochrom
vitrobot mark iv system	Thermo Fisher Scientific
Vivaspin 20 (3K, 5K, 10K MWCO)	Sartorius
water bath Sonorex RK100H	Bandelin

Chemicals	**Supplier**
1-methylpiperazine	Sigma-Aldrich
1,4-dimethylpiperazine	Sigma-Aldrich
30% acrylamide 29:1	Roth
30% acrylamide 37.5:1	Roth
ammonium acetate	Roth
^{15}N-ammonium chloride	Cortecnet
^{14}N-ammonium chloride	Sigma-Aldrich
ammonium sulphate hydrate	Roth
antipain - dihydrochloride	Roth
aprotinin	Roth
ammonium persulphate (APS)	Roth
adenosine-5'-triphosphat-di-sodium salt (ATP)	Roth
bacto-agar	DIFCO
benzonase nuclease (250 U / ul)	Sigma-Aldrich

beta-mercaptoethanole	Roth
BisTris	Roth
Blue Eye Prestained Protein Marker	Jena Bioscience
bromphenol blue	Bio-Rad
bovine serum albumin (BSA)	Serva
calcium chloride dihydrate	Sigma-Aldrich
coomassie brillant blue R-250	Serva
deuterium oxide (D2O)	Cortecnet
ethylene diamine	Merck
ion (III) chloride hexahydrate	Sigma-Aldrich
acidic acid	Roth
^{13}C-Glucose	Cortecnet
^{12}C-Glucose	Roth
glycerol	Roth
hydrochloric acid	Roth
Hydroxylamine (50% in water)	Sigma-Aldrich
imidazole	Roth
InstantBlue	Expedeon
isopropyl-beta-D-thiogalactopyranoside (IPTG)	Roth
kanamycin sulphate	Roth
potassium dihydrogen phosphate	Roth
cobalt(II) sulphate heptahydrate	Roth
Luria/Miller medium (LB)	Roth
leupeptin	Serva
Low-Range Rainbow marker 3.5 - 38 kDa)	Cytiva
lysozyme	Roth
Magnesium chloride hexahydrate	Roth
magnesium sulphate	Sigma-Aldrich
Mangan (II) chloride tetrahydrate	VWR Scientific
methanole	Roth
methylamine (50% in water)	Merck
sodium acetate	Roth
sodium carbonate	VWR Scientific
sodium chloride	Roth
sodium citrate dihydrate	Roth
sodium hydrogen carbonate	Roth
di sodium hydrogen phosphate dihydrate	Roth
sodium hydroxide	Roth
sodium molybdate dihydrate	Sigma-Aldrich
sodium trimethylsilylpropansulfonate (DSS)	Cortecnet
sodium vanadate	Sigma-Aldrich
nickel (II) sulphate hexahydrate	Sigma-Aldrich
polyethylene glycol 3350	Fluka Analytical
pepstain	Roche
Phos-tag	WAKO Chemicals

Phenylmethylsulfonyl fluoride (PMSF)	Roth
propanole	Thermo Fisher Scientific
sodium dodecylsulphate	Roth
Terrific Broth(TB) medium	Roth
Tetramethylethylenediamine (TEMED)	Roth
thiamin hydrochloride	Sigma-Aldrich
Tris	Roth
Tris (2-carboxyethyl) phosphine hydrochloride (TCEP)	Roth
Tricin	Roth
urea	Roth
citric acid	Monohydrat

Kit	Manufacturer
Classic 1-4	Jena Bioscience
Classic 5-8	Jena Bioscience
Classic 9-10	Jena Bioscience
Crystal	Hampton Research
Index	Hampton Research
JCSG++	Jena Bioscience
MIDASplus	Molecular Dimensions
Morpheus	Molecular Dimensions
Wizard Cryo	Molecular Dimensions

Molecular biology
CPEC

Circular polymerase extension cloning (CPEC) (Quan and Tian 2009) is a fast and inexpensive technique, as it is conducted without the need for restriction and ligase enzymes while relying on a PCR based method to generate sticky ends. To linearise the vector, primers for both ends of the cloning site must be designed to anneal at similar temperatures at opposite strands and synthesized. The mixture of the PCR reaction and the PCR protocol are the following:

Mixture for PCR reaction:

substance	concentration
Pfusion Buffer GC	1x
dNTPs 10 mM	0.2 mM
forward Primer	0.5 µM
backward Primer	0.5 µM
template	50 pg
polymerase Pfusion	1x
water	top up to 100 µl

PCR protocol:

Temperature	Time	reasoning
98°C	30 sec	initial melting
30x		
98°C	10 sec	melting
x°C	20 sec	annealing
72°C	140 sec note: extension time is calculated 15-30 sec /kb	amplification
72°C	5 min	final amplification
4°C	Hold	storage

The second step is the amplification of the insert, for which two primers covering the flanking sites of the insertable part, as well as the overlapping regions with the backbone, must be designed and synthesized. The primer pairs are between 30-40 bp long and should have similar annealing temperatures for the insert-part and the overlapping-part since the PCR reactions are sequential. The primers were checked for homo- and heterodimerisation, which would hinder the reaction, and designed accordingly. The PCR for insert amplification is divided into two phases: the short insert extraction and amplification from the vector and the later amplification of the insert with the overlapping ends. In the following tables, the reaction mixture and the PCR protocol are presented:

Mixture for PCR reaction:

substance	concentration
Pfusion Buffer GC	1x
dNTPs 10 mM	0.2 mM
forward primer	0.5 µM
backward primer	0.5 µM
template	50 pg
polymerase Pfusion	1x
water	top up to 55 µl

temperature	time	reasoning
98°C	30 sec	initial melting
5x		cycle to amplify the initial insert
98°C	10 sec	melting
x°C	20 sec	annealing to the insert
72°C	10 sec note: extension time is calculated 15-30 sec /kb	amplification
25x		cycle to amplify the final insert
98°C	10 sec	melting
x°C	20 sec	annealing of the full primer
72°C	10 sec note: extension time is calculated 15-30 sec /kb	amplification of final insert with overlapping ends
72°C	5 min	final amplification
4°C	hold	storage

The amplified and linearised vector and the insert were loaded and separated on agarose gels (0.8% for backbone; 1.6% for insert) at 120V for approximately 50 min. The amplified PCR products were extracted from the gel following the instructions of the gel extraction kit (Quiagen) after cutting out the PCR bands with a scalpel.

The extracted PCR products were eluted in water, and their concentration was measured using UV-spectroscopy. The final step is the ligation of the insert to the vector, which is the actual CPEC reaction. As a control for the success of the reaction, a negative control was conducted in parallel with only the backbone but no insert in the mixture.

Mixture for the CPEC reaction:

substance	concentration
Pfusion Buffer HF	1x
dNTPs 10 mM	0.2 mM
backbone	50-100 ng
insert	2x molar excess of backbone
polymerase Pfusion	1x
water	top up to 25 µl

temperature	time	reasoning
98°C	30 sec	initial melting
20x		
98°C	10 sec	melting
70°C	1 sec	cool to 70°C
70°C to Tm+3°C	0.1°/ sec	slow ramp annealing
Tm+3°C	90 sec	Annealing
72°C	140 sec note: extension time is calculated 15-30 sec /kb	amplification
72°C	5 min	final amplification
4°C	hold	storage

1 µl of the reaction and the negative control were used for the transformation into XL10 *E.coli* cells (see Transformation). Cells were incubated overnight at 37°C, and colony growth was observed for the reaction but not the control experiment. Three colonies were picked, incubated and used for plasmid extraction, which was used for sequencing of the final plasmid.

Transformation
Chemical competent *E.coli* cells (e.g. BL21(DE3), XL10, TOP10 etc.) were taken from the -80°C storage unit and thawed on ice for 10-20 minutes. 1 µl of plasmid was added to the cells, and the tube was gently mixed by careful snapping. The mixture was stored on ice for 30 minutes to allow the plasmids to accumulate at the prokaryotic membrane. The tube was carefully held in a prewarmed water bath at 42°C for 30 seconds and

directly placed into ice afterwards. After a short cooling period, 200 µl of TB medium was added to the mixture. The tube was incubated at 37°C while shaking at 225 rpm (innova 4200) for 1 hour. The entire mixture was applied on LB agar plates with the appropriate antibiotic and spread over the whole plate with a heat-sterilised spatula under a hood. The plate was incubated in a top-up position at 37°C overnight.

Miniprep and Plasmid extraction
Colonies grown on the LB plates were picked with a sterilised plastic tip under the hood and placed into a tube with 5 ml LB medium and the appropriate antibiotic. The tube was incubated at 37°C while shaking at 225 rpm (innova 4200) overnight.

The tube was centrifuged for 10 minutes at 4000 g, and the supernatant was discarded into an autoclavable glass bottle. The pelleted cells were resuspended and further processed as described in the protocol of the miniprep kit (Quiagen).

The plasmid was eluted in water, and the concentration was measured using UV-spectroscopy at the Nanoview.

Mutagenesis
Mutagenesis was conducted using the QuickChange Site-Directed Mutagenesis Kit (Agilent). For this, primers were designed using the following website: https://www.agilent.com/store/primerDesignProgram.jsp

The reaction mixture and PCR protocol were used as described:

substance	concentration
QuickChange LM Buffer	1x
QuickSolution	0.2 µl
dNTPs	0.4 mM
primer	0.5 µM
template	100 ng
QuickChange enzyme	1x
water	top up to 10 µl

temperature	time	reasoning
95°C	120 sec	initial melting
30x		
95°C	20 sec	melting
55°C	30 sec	annealing
65°C	210 sec note: extension time is calculated 30 sec /kb	amplification
65°C	5 min	final amplification
4°C	hold	storage

0.5 µl of DpnI enzyme was added to the mixture to digest the methylated mother plasmid during a 10-minute incubation at 37°C. 1 µl of the final reaction was used for the transformation of *E.coli* XL10 cells. After incubation at 37°C overnight, three colonies were picked for miniprep, plasmid extraction and sequencing.

Agarose-gel electrophoresis

To separate linearized DNA from PCR experiments, the DNA was loaded on an agarose matrix gel and an electric current was applied in a TAE buffer system. The DNA was supplemented with a loading dye to fill the well. The DNA migrates towards the anode due to its negatively charged phosphate sugar backbone. A DNA intercalating substance (GelRed™) was added to the agarose gel immediately before casting to detect the DNA on the gel. An appropriate gel percentage must be chosen for the DNA molecule to separate: a small percentage is used for larger constructs (e.g. backbone with around 5 kbp, 0.6 % agarose), and a larger percentage gel (e.g. 1.8 %) for smaller constructs like inserts of 200 bp.

The gel was made as follows: the appropriate mass of agarose was solved in 50 ml TAE buffer [40 mM Tris, 20 mM acidic acid, 1 mM EDTA] by heating in a microwave, with occasional careful shaking. After refilling to 50 ml with ddH$_2$O, 5 µl GelRed™ was added after cooling the matrix to 37°C. The gel was cast with the appropriate comb.

Expression and purification
Enzymes used for purification of Gab1 and Shp2 constructs
SUMO protease
The SUMO protease was expressed using the pET28b(+) vector system with an N- terminal His$_6$-tag and a resistance gene against kanamycin. The vector was ordered from addgene (plasmid #64697). The plasmid was transformed into BL21(DE3), and a colony was used for cultivation in 5 ml TB medium in the presence of 50 µg/ml kanamycin and 1% glucose at 225 rpm (innova 4200) and 37°C overnight. Another overnight incubation was set up in 100 ml TB medium under the same conditions. The density of the culture was measured at a wavelength of 600 nm, and 3.5 litres of freshly autoclaved TB medium, divided into seven flasks, was inoculated to an OD$_{600}$ of 0.05. The culture was cultivated after the addition of 1 % glucose and 50 µg/ml kanamycin at 37°C, 130 rpm (innova 43) until an OD$_{600}$ of 1.5 was reached. The expression was induced by the addition of 1 mM IPTG, and the cultivation continued for another 4 h. The cells were harvested by centrifugation in 1-litre buckets (BeckmanCoulter) at 2560 g in a precooled centrifuge (J6-MI; JS-4.2) set to 4°C for 30 min. The cells were resuspended in 35 ml lysis buffer [50 mM Tris/HCl pH 7.5^{RT}, 20% saccharose, 350 mM NaCl and 45 mM imidazole] per litre and stored until further use at -80°C.

After thawing the suspension on ice 1 mM TCEP and 0.2% Igepal were added, and cells were lysed by sonication for three cycles each 1 min (100% power, 50% interval (Bandelin Sonopuls HD 2200, booster horn SH 213G, TT13)). The protein was cooled during and in between cycles. To separate the supernatant from the cell debris, centrifugation at 30.000 g in a precooled centrifuge (Avanti JXN-26; JLA-16.250) was conducted two times one hour. The supernatant was applied to a Ni^{2+} immobilised IMAC FF (5ml) column through the sample pump of the Äkta pure system located in a 4°C cold room. The column was treated with a washing buffer [50 mM Tris/HCl pH 7.5^{RT}, 150 mM NaCl, 30 mM imidazole, 1 mM TCEP] until no further UV signal was detected. And the protein was eluted with a gradient of 30 mM – 300 mM of imidazole. The fractions containing the protein were pooled and dialysed against 4 litres of dialysis buffer [50 mM pH 7.5^{RT}, 150 mM NaCl and 1 mM mercaptoethanol] in a 3.5 kDa cut-off dialysis tube at 4°C, overnight. The concentration was measured using UV-spectroscopy and Glycerine to a final concentration of 50% (v/v) was added, and 1 mM DTT was freshly added before storing the protease at -80°C.

3C protease
For the expression and purification of the 3C protease, the vector containing an N-terminal GST-tag and an ampicillin resistance gene was transferred into BL21(DE3)

cells, from which a colony was used to inoculate 400 ml LB medium + 100 µg/ml ampicillin. The culture was shaken at 225 rpm (innova 4200) and 37°C overnight. Eight flasks containing each 560 ml TB medium and 100 µg/ml ampicillin were inoculated with 40 ml of the overnight culture to a final volume of 600 ml. The flasks were set to a 37°C incubator and shaken at 130 rpm (innova 43) until an optical density of 1.2 at 600 nm was measured. The flasks were transferred to a 4°C cold room, and the incubator was set to 20°C. After reaching the desired temperature, the flasks were implemented again into the incubator, and cultivation was continued for 3 hours after inducing the expression with 0.1 mM IPTG. The cells were harvested at 2560 g in a precooled centrifuge (J6-MI; JS-4.2) set to 4°C for 30 min.

To resuspend the cells, 20 ml of TPE buffer and 0.0007 mg/ml pepstatin were added per 1L bucket [1x PBS, 1% triton X-100, 100 mM EDTA, 250 mM NaCl, 5% glycerol] on ice. The resuspension was sonicated for three cycles of each 30 seconds at 100% power and 50% interval (Bandelin Sonopuls HD 2200, booster horn SH 213G, TT13)) on ice. To separate the cell debris, the lysed cells were centrifuged for 1 hour at 20.000 g (Avanti JXN-26; JLA-16.250) at 4°C. The supernatant was transferred to fresh GSH beads (30 ul per 1 ml supernatant) and incubated on a nutator in the cold room overnight. The incubated beads were centrifuged at 700 g for 5 minutes, and the supernatant was collected and later autoclaved. The beads have been extensively washed with TPE buffer and later with washing buffer [50 mM Tris/HCl pH 7.5^{RT}, 150 mM NaCl]. For each washing step, the buffer was added and nutated for 5 minutes, followed by centrifugation. This was continued until no further obvious protein contamination was seen in the discarded supernatant. The elution of the GST-tagged 3C protease was done in the presence of 100 mM GSH adjusted with 1 M Tris/HCl of 8.8 to a final pH of 7.0. The supernatant was dialysed against 3 x 500 ml dialysis buffer [50 mM Tris/HCl pH 7.5^{RT}, 150 mM NaCl, 1 mM EDTA, 1 mM DTT] in a 3.5 kDa cut-off dialysis tube while stirring at 4°C. The concentration was measured, and the protein was aliquoted and stored at -80°C after shock freezing in liquid nitrogen.

TEV protease
For expression of the His_6-tagged Tobacco Etch Virus (TEV) protease, the vector pRK793 (TEV S219V) was transformed into BL21(DE3)-RIL cells and incubated overnight on an LB plate with the necessary antibiotics. The vector has a resistance gene to ampicillin, while the bacterial stock has a resistance to chloramphenicol. BL21(DE3)-RIL cells have additional copies of tRNA encoding genes, elevating the available tRNA pool to overcome codon bias. 5 ml LB medium supplemented with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol were inoculated with a colony from the transformation and

shaken at 225 rpm (innova 4200) at 37 °C during the day. 600 μl of this culture was used to inoculate 600 ml antibiotic-supplemented LB medium for incubation overnight at the same parameters. To ten expression flasks with each 450 ml LB medium, 2 g/L glucose and antibiotics, 50 ml of the overnight culture was added and shaken at 130 rpm (innova 43) at 37°C until an OD_{600} of 0.9 was reached. The flasks were shortly stored in a 4°C cold room until the incubator's temperature had reached 30 °C. After 30 minutes of further incubation, 1 mM of IPTG was added to induce the expression. The cultures were harvested after 4 hours by centrifugation at 2560 g (J6-MI; JS-4.2) for 30 minutes at 4°C. The supernatant was discarded, and 40 ml of lysis buffer [25 mM Tris/HCl pH 7.5^{RT}, 200 mM NaCl, 10 mM imidazole, 1 mM TCEP] was added to each pellet derived from 1 L expression. The resuspension was transferred to 50 ml tubes and stored at - 80°C until further use.

After thawing the resuspended cells on ice, sonication of three cycles, each one minute at 100 % power and an interval of 50% (Bandelin Sonopuls HD 2200, booster horn SH 213G, TT13), was carried out while continuously cooling with ice. The lysed cells were centrifuged for 90 minutes at 20.000 g (Avanti JXN-26; JLA-16.250). The supernatant was applied to a 5 ml HiTrap IMAC FF column with immobilised cobalt ions. After extensive washing with lysis buffer, the elution of the His$_6$-tagged protein was done with a step gradient of 75 mM, 250 mM and 300 mM imidazole. The fractions of 250 mM imidazole were diluted (1:1) with dialysis buffer [25 mM Tris/HCl pH 7.5^{RT}, 200 mM NaCl and 10 mM imidazole, 1 mM TCEP] and dialysed in a 3.5 kDa dialysis tube overnight against 4 L dialysis buffer while stirring at 4°C. The protein concentration was measured using UV-Vis spectroscopy and aliquoted into 1.5 ml tubes, frozen in liquid nitrogen and stored at -80 °C until further use. The expression and purification of the TEV-protease (S219V) were done according to (Tropea et al. 2009).

Abl kinase
The Abl kinase was expressed together with the phosphatase YopH, to erase potential phosphorylation to the host proteome, done by the Abl kinase. Both vectors (pCDFDuet1-YopH (164-468) with a streptomycin resistance gene and pET21-muABL1 (138-534, F420V)-His with an ampicillin resistance gene) were co-transferred into BL21(DE3) and incubated on an LB agar plate with 100 μg/ml ampicillin and 50 μg/ml streptomycin at 37°C overnight. A colony was used to inoculate 50 ml TB medium supplemented with both antibiotics and cultivated at 37°C while shaking at 300 rpm (innova 4200) for 3 hours. This culture was used to inoculate 4x 490 ml TB medium to a final volume of 500 ml to set up a large expression. The antibiotics were added to the medium, and the cultures were placed in a preheated 37°C incubator to be shaken at 130 rpm (innova 43)

until an OD$_{600}$ of 1.2 was measured. The flasks were stored in a 4°C cold room while the incubator was cooled down to 18°C. After the continuation of the cultivation for one more hour, 0.2 mM IPTG was added, the shaking speed was reduced to 90 rpm (innova 43), and the cultivation was continued overnight. The cells were harvested into 1-litre buckets (BeckmanCoulter) at 2560 g (J6-MI, JS-4.2) at 4°C for 30 minutes. The supernatant was autoclaved, and the cell pellets were stored at -80°C.

The resuspension per bucket was done in 35 ml IMAC A buffer [50 mM Tris/HCl pH 8.0^{RT}, 500 mM NaCl, 5% glycerol, 25 mM imidazole] supplemented with the following inhibitor mix [0.01 mg/ml aprotinin, 0.005 mg/ml antipain, 0.0005 mg/ml leupeptin, 0.0007 mg/ml pepstatin, 0.2 mM PMSF, 0.1 mM Na$_3$VO$_4$, 0.1 mM Na$_2$MoO$_4$, 1 mg/ml lysozyme, 1 mM DTT], on ice. The cells were lysed by sonication for three cycles, each 1 minute-long exposure to 100% power, 50% interval (Bandelin Sonopuls HD 2200, booster horn SH 213G, TT13). Following centrifugation at 30.000 g (Avanti JXN-26; JLA-16.250) at 4°C for 60 minutes, the supernatant was transferred into a fresh tube on ice. A freshly prepared Ni^{2+}-immobilised IMAC FF (5ml) column was washed, and the supernatant was applied to the column via a sample pump of the Äkta pure system. The purification was done in a 4°C cold room. After extensively washing the column with IMAC A buffer, elution was done as a gradient from 25 mM - 275 mM imidazole. The flow rate was set to 5 ml/min and the fractions were automatically collected to 5 ml. Fractions containing the protein have been pooled and dialysed against 4 litres of dialysis buffer [20 mM Tris/HCl pH 8.0^{RT}, 25 mM NaCl, 5% glycerol, 1 mM DTT] at 4°C while stirring overnight. The sample was applied to a washed anion exchange chromatography column (HiTrap Q HP, 5ml). After the column was extensively flushed with the binding buffer [20 mM Tris/HCl pH 8.0^{RT}, 5% glycerol, 1 mM DTT] the elution was done as a gradient from 0 mM to 350 mM NaCl. The fractions containing the kinase were pooled and concentrated to a final volume of 5 ml. The concentrated protein was applied to a preequilibrated size exclusion chromatography column (HiLoad 16/600 Superdex 75 pg) with the SEC buffer [50 mM Tris/HCl pH 8.0^{RT}, 100 mM NaCl, 5 % glycerol, 1 mM DTT]. The flow rate was set to 1 ml/min, and the fractions were automatically collected to a volume of 5 ml. The protein eluted at 58.5 ml, which corresponds well to the 46 kDa protein. 15 mg per 1 L expression could be extracted. The kinase was aliquoted and stored at -80°C after shock freezing in liquid nitrogen.

Gab1 and Shp2 proteins characterised in this project
The following constructs are depicted in **Figure 14**. Monophosphorylated Gab1 peptides were synthesized by the commercial supplier ProteoGenix, France.

N-SH2$^{1\text{-}106}$ and C-SH2$^{102\text{-}220}$

The individual SH2 domains N-SH2$^{1\text{-}106}$ and C-SH2$^{102\text{-}220}$ were expressed as N- His$_6$- SUMO-fusion proteins using the pETSUMO plasmid (Invitrogen K300-01). pETSUMO-N-SH2$^{1\text{-}106}$ and pETSUMO-C-SH2$^{102\text{-}220}$ have been transformed into BL21(DE3) *E.coli* cells. A colony was picked of each and cultivated in 10 ml TB medium + 50 µg/ml kanamycin at 37°C, 225 rpm (innova 4200) overnight. 100 µl of this culture were used to inoculate 100 ml TB medium in the presence of 50 µg/ml kanamycin. Again, the culture was incubated as described above. The density of the bacterial culture was measured the next morning using the optical density at 600 nm (OD$_{600}$). 3L of sterilised TB medium + 50 µg/ml kanamycin, divided into six flasks, were inoculated with the overnight culture to an OD$_{600}$ of 0.05 and the flasks set into the shaking incubator at 37°C and 130 rpm (innova43). The bacterial growth was monitored, and at a final OD$_{600}$ of 1, taken to the cold room at 4°C. The incubator was set to 16°C, and the cultures were incubated for another 30 minutes before induction of expression using a final concentration of 0.5 mM IPTG. The incubation was continued overnight at 16°C while shaking at 130 rpm (innova 43). The cultivated medium was transferred into 1 L centrifugation buckets (BeckmanCoulter) and centrifuged at 2560 g (J6-MI, JS-4.2), 4°C for 30 minutes. The supernatant was discarded and autoclaved while the bacterial pellets were stored at -80°C until further use.

30-35 ml ice-cold lysis buffer [500 mM NaCl, 20 mM Tris-HCl pH 7.5^{RT}, 1 mM TCEP] including an inhibitor-mix [0.01 mg/ml aprotinin, 0.005 mg/ml antipain, 0.0005 mg/ml leupeptin, 0.0007 mg/ml pepstatin, 0.2 mM PMSF] was added to the bacterial pellets, resting on ice. Recurrent vortexing and pipetting resulted in the complete resuspension of the cells in the buffer. The solution was transferred into 50 ml tubes and exposed to ultrasonication waves for 3 x 1 minute ((100 % power, 50 % interval) (Bandelin Sonopuls HD 2200, booster horn SH 213G, TT13)), while cooling the suspension in between repetitions and during treatment. The lysed cell suspension was transferred into 250 ml centrifugation buckets (BeckmanCoulter) and centrifuged at 30 000 g, 4°C for 1 h (Avanti JXN-26; JLA16.250). The supernatant was transferred into a clean centrifugation bucket, and centrifugation was continued for another 1 h.

An immobilised metal affinity chromatography (IMAC) FastFlow (FF) column was cleaned with 250 mM imidazole after NiSO$_4$ incubation, followed by extensive washing with lysis buffer (without inhibitors). The supernatant was applied via a sample pump to a Ni^{2+}-IMAC column, attached to an Äkta pure system at a 5 ml/min flow rate in the cold room at 4°C. The chromatography was monitored using UV-spectroscopy at 220 nm, 260 nm and 280 nm. The flow-through was collected, and the column was washed with

lysis buffer until no UV absorption at 280 nm was detectable. The His_6-tag, which is located at the N-terminal region of the protein, chelates the column-bound Ni^{2+} ions, attaching the fusion protein to the column. To elute the protein, elution buffer [250 mM Imidazole, 500 mM NaCl, 20 mM Tris-HCl pH 7.5^{RT}, 1 mM TCEP] was applied to the column. Protein-containing fractions were identified by their elevated UV absorption at 280 nm, on passing the UV-detection chamber. The column was rinsed with water and stored overnight.

The concentration of the elution fraction was measured using UV-spectroscopy at 280 nm, and SUMO protease was added in a 1:20 (SUMO protease: POI) ratio. The mixture was dialysed in a 3.5 kDa cut-off dialysis tube immersed in a stirred 4L dialysis buffer [150 mM NaCl, 50 mM Tris-HCl pH7.5^{RT}, 1 mM TCEP] at 4°C, overnight. The SUMO protease recognises the folded SUMO domain, which is located C-terminal to the His_6- tag and cleaves to yield the protein of interest with a native N-terminus, i.e. with no amino acids from the tag-system.

The content of the dialysis tube was carefully poured into clean glassware and re-applied to a Ni^{2+}-IMAC column, equilibrated with dialysis buffer beforehand. The column was washed with dialysis buffer, and the flow-through was collected. After UV absorption at 280 nm was no further detectable, elution of the retarded protein (uncleaved fusion protein, tag-system and SUMO-protease) on the column was done by letting elution buffer pass the column.

The pooled flow-through fractions were concentrated to 5 ml using the vivaspin 5kDa MWCO concentrator and applied to a 5 ml sample loop. Upon attachment to the FPLC-system the sample loop released the protein on to a size exclusion chromatography (SEC) column, HiLoad 16/600 Superdex 75 pg. The flow rate was set to 1 ml/min, and 5 ml fractions were collected with the fraction collector into 15 ml tubes. The column and system were rinsed with water and SEC buffer, dependent on the further use of the protein – (**Table 2**) overnight beforehand. The protein of interest eluted at 79 ml (N- $SH2^{1-106}$) and 77 ml (C-$SH2^{102-220}$), which corresponds to a globular protein of 13 kDa and 14 kDa. The elution volume correlates with the actual sizes of the 11.9 kDa and 13.3 kDa proteins. The SEC elution fraction was concentrated to around 2 mM, aliquoted and freezed in liquid nitrogen for storage at -80°C.

Shp2 tandem SH2 domain construct $Shp2^{1-222}$

The Shp2 tandem SH2 domain construct was expressed with an N-terminal His_6-tag followed by a TEV-protease cleavage site using the pET28 plasmid. The pET28- $Shp2^{1-222}$ vector was freshly transformed into BL21(DE3) cells and incubated on an LB plate with 50 µg/ml kanamycin overnight at 37°C. A colony was picked to inoculate

10 ml TB medium with 50 µg/ml kanamycin. The culture was incubated overnight at 37°C while shaking at 225 rpm (innova 4200). 100 µl of this culture was used to inoculate 100 ml TB medium with 50 µg/ml kanamycin, and the culture was again incubated at 37°C while shaking overnight. The density of the culture was measured at a wavelength of 600 nm. 3 L of sterilised TB medium supplemented with 50 µg/ml kanamycin was divided into 6 Erlenmeyer flasks, inoculated with the overnight culture at an OD_{600} of 0.05 and incubated at 37°C while shaking at 130 rpm (innova 43) until an OD_{600} of approximately one was reached. The flasks were taken to the cold room to slow down further growth while the incubator was set to 18°C. The flasks were again placed in the incubator, and cultivation was continued for 30 minutes before IPTG was added to a final concentration of 0.5 mM. After overnight incubation, the culture was poured into 1L centrifugation buckets (BeckmanCoulter) and centrifuged for 30 minutes at 2560 g (J6-MI; JS-4.2). The supernatant medium was autoclaved while the bacterial pellets were stored at -80°C until further use.

To resuspend the bacterial pellets on ice, 30-35 ml of IMAC binding buffer [50 mM Tris- HCl pH 7.5^{RT}, 200 mM NaCl, 20 mM imidazole, 1 mM TCEP] supplemented with an inhibitor-mix [0.01 mg/ml aprotinin, 0.005 mg/ml antipain, 0.0005 mg/ml leupeptin, 0.0007 mg/ml pepstatin, 0.2 mM PMSF] was added per 1L bucket and pipetted and vortexed occasionally. The resuspension was transferred into 50 ml falcon tubes for sonication, which was done on ice for 3 cycles, each 1 minute long at 100% power and 50% interval (Bandelin Sonopuls HD 2200, booster horn SH 213G, TT13), storing the tube on ice after each cycle. The lysed cell suspension was transferred into 250 ml centrifugation buckets (BeckmanCoulter), and cell debris was separated at 30.000 g for a total of 3 h at 4°C (Avanti JXN-26; JLA-16.250). The supernatant was applied to a freshly Ni^{2+}- immobilised and prewashed IMAC FF column using the sample pump of the Äkta pure system. The purification was done at 4°C on the Äkta pure system while manually or automatically collecting the fractions. The real-time monitoring of protein-containing fractions was done using UV-spectroscopy at 220 nm, 260 nm and 280 nm on the device. The column was washed with IMAC-binding buffer until the UV-signal intensity was continuously low. To elute the fusion protein from the column, the imidazole concentration was increased to 500 mM while keeping the other buffer components constant. The elution was accompanied by an increase in the UV-signal intensity, and fractions were collected and stored on ice. The column was washed with binding buffer and stored at 4°C until the next day. The approximate protein concentration was measured using UV- spectroscopy and the elution buffer as a reference. TEV protease was added to the protein to a final ratio of 1:30 (TEV protease: fusion protein) and transferred into a 3.5 kDa cut-off dialysis tube. It was dialysed against 4L buffer [50 mM

Tris-HCl, pH7.5^{RT}, 150 mM NaCl, 1 mM TCEP] at 4°C overnight. The dialysed protein mixture was centrifuged, and the supernatant was applied to the prepared Ni^{2+}-IMAC attached to the Äkta pure system. The flowthrough was collected, and the column was further washed with dialysis buffer until the UV-signal intensity reached the baseline. The elution buffer was used to mobilise all Ni^{2+} chelating molecules, e.g. uncleaved protein, tag-system, TEV protease. The flowthrough and washing fractions were pooled and concentrated up to 5 ml by centrifugation using the vivaspin concentrator (10 kDa MWCO) and applied to a 5 ml sample loop. The protein was injected automatically to a prewashed HiLoad 16/600 Superdex 75 pg size exclusion chromatography column at a flowrate of 1ml/min and fraction size of 5 ml. The buffer used for preparing the column and eluting the protein depended on the further use of the protein (**Table 2**). The protein eluted at around 67 ml, corresponding to a 25 kDa globular protein, which is in line with the actual size of the tandem SH2 domain protein. The protein was concentrated to around 2 mM, aliquoted and stored at -80°C after freezing in liquid nitrogen. A total of 15 mg of pure protein could be extracted per Litre.

Gab1 peptides Gab1$^{613-694}$ and Gab1$^{617-684}$

Gab1 peptides Gab1$^{613-694}$ and Gab1$^{617-684}$ were produced with a His$_6$-3C-tag. The GST-tag, inherent to the vector, was not used for purification – purification was done according to Gruber et al. 2022. The pET42-hGab1$^{613-694}$ and pET42-hGab1$^{617-684}$ plasmids were transformed into BL21(DE3) cells and incubated on an LB plate with 50 µg/ml kanamycin at 37°C overnight. One colony was picked to inoculate 10 ml TB medium supplemented with 50 µg/ml kanamycin. The culture was incubated while shaking at 225 rpm (innova 4200) at 37°C overnight. 100 ml TB medium with 50 µg/ml kanamycin was inoculated with 100 µl of the overnight culture. The flask was placed into an incubator at 37°C while shaking at 225 rpm (innova 4200) overnight. The density of the culture was measured at a wavelength of 600 nm. 4L sterilised TB medium supplemented with 50 µg/ml kanamycin was divided into eight Erlenmeyer flasks. The overnight culture was added to a final OD$_{600}$ of 0.05, and the cultivation started at 37°C while shaking in the (innova 43) at 130 rpm until the bacterial growth reached an OD$_{600}$ of 1.3. To induce expression, IPTG was added to a final concentration of 0.5 mM IPTG, and incubation continued for 4 hours. The bacterial culture was transferred into 1L centrifugation buckets (BeckmanCoulter), and bacterial cells settled at 2560 g in a precooled centrifuge (J6-MI; JS-4.2) set to 4°C for 30 min. The supernatant medium was autoclaved while the bacterial pellet was stored until further use at -80°C.

35 ml lysis buffer [20 mM Tris/HCl pH 7.5^{RT}, 500 mM NaCl], supplemented with an inhibitor-mix [0.01 mg/ml aprotinin, 0.005 mg/ml antipain, 0.0005 mg/ml leupeptin,

0.0007 mg/ml pepstatin, 0.2 mM PMSF, 1 mM TCEP], was added per 1L-bucket and the cells were resuspended on ice, alternating with occasional vortexing and additional pipetting until the cells were fully resuspended. The suspension was transferred into 50 ml tubes and sonicated in three cycles for 1 minute each at 100% power and an interval of 50% (Bandelin Sonopuls HD 2200, booster horn SH 213G, TT13) while cooled on the ice during and in between the cycles. The lysed cell suspension was centrifuged two times for 30 and 60 min at 30.000 g (Avanti JXN-26; JLA-16.250) at 4°C in clean 250 ml centrifugation buckets (BeckmanCoulter). The cell debris was discarded, and the protein- rich supernatant was transferred into fresh 50 ml tubes and stored on ice.

7 ml of a freshly prepared Co2+ IMAC beads suspension was washed with lysis buffer and divided into falcon tubes filled with the lysed supernatant. The solution was incubated while gently shaking on a tumbling shaker at 4°C for 1 hour. After centrifugation at 500 g (5804 R, Eppendorf) for 5 minutes at 4°C, the beads settled, and the supernatant could be aspirated. After pooling the beads, they were washed several times with lysis buffer, including inverting for 5 minutes and short centrifugation at 500 g, and the supernatant was again discarded. After no obvious protein contamination could be detected, the tube was topped up to 20 ml with lysis buffer, and 5 mg/ml 3C protease was added. The tube was nutated in a 4°C cold room overnight. The 3C protease cleaves the His_6-tag at the following 3C protease recognition site to release the desired protein with glycine and proline as additional N-terminal amino acids into the supernatant while the tag and uncleaved fusion protein remained on the beads.

The supernatant was transferred into a clean tube stored on ice, and the remaining beads were extensively washed with lysis buffer, several times. The supernatant fractions were pooled. In the case where a phosphorylated protein was desired, the concentration of the supernatant was measured using UV-spectroscopy and 10 mM $MgCl_2$, 1 mM TCEP, 4 mM ATP were added together with a 1:200 (w/w) ratio of the murine ABL1 kinase to the tube. After gentle inversion, the reaction was placed into a 30°C warm water bath for at least 5 hours. To exchange the buffer for the following chromatography, the sample (for phosphorylated and unphosphorylated Gab1 peptides alike) was placed in a 3.5 kDa cut-off dialysis tube and stirred in 4L dialysis buffer [20 mM carbonate buffer pH 9.5, 0.1 mM Na_3VO_4, 0.1 mM Na_2MoO_4] at 4°C, overnight. For phosphorylated protein, an anion exchanger column (HiTrap Q HP 3x 5ml) was extensively washed with water followed by cleansing with pH 4 and pH 11 chromatography buffer [7.7 mM methylamine; 9.12 mM ethylenediamine; 6,4 mM 1-methylpiperazine; 1.3 mM 1,4-dimethylpiperazine; 5.8 mM BisTris; 9.3 mM hydroxylamine, pH adjusted with HCl]. The sample was applied with a sample pump onto the column, attached to the Äkta pure

system, located in a 4°C cold room. The flowthrough was collected, and a washing step was done with the pH 11 chromatography buffer. To elute the proteins attached to the positive column material by their net surface charge, the pH 11 buffer was continuously mixed with the pH4 buffer to provide a linear elution gradient from pH 11 to pH 4. The proteins eluting from the column were collected automatically in 5 ml fractions. The flow rate was set to 5ml/min. The fractions were loaded on a gel and separated via SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Fractions corresponding to 72 % (pYpY-Gab1$^{613-694}$) and 74 % (pYpY-Gab1$^{617-684}$) of buffer B (chromatofocussing buffer pH 4) contained the desired double phosphorylated Gab1 species. These fractions (or in cases without phosphorylation, dialysed protein) were applied to a size exclusion chromatography column (HiLoad 16/600 Superdex 75 pg) after concentration to a final volume of 5 ml. The flow rate was set to 1 ml/min, and the fractions were automatically collected to 5 ml. The buffer used depended on the later use of the protein. The fractions of the final elution peak were pooled and concentrated to 2-4 mM. The protein was aliquoted and stored at -80°C after freezing in liquid nitrogen. This procedure resulted in 6.5 mg pure double phosphorylated protein per litre of expression culture.

Table 2 *SEC-Buffers for individual experiments.*

experiment	buffer
ITC	20 mM citrate buffer pH 6.0
NMR	20 mM BisTris/HCl, 50 mM NaCl pH 6.0
crystallisation	50 mM Tris/HCl pH 7.5, 50 mM NaCl, 1 mM DTT

Protein characterisation
Polyacrylamide electrophoresis (PAGE)
To follow up on the progress of protein purification and analyse protein mixtures, electrophoresis of these molecules on polyacrylamide gels was performed. Proteins are denatured by temperature, reducing agents and SDS. As SDS binds to the protein in a constant ratio and shields the net charge of the protein, the SDS-covered protein chain migrates to the anode based on its size alone. Small molecules run faster, while larger proteins are retarded by the gel matrix. A gel with the appropriate acrylamide percentage (responsible for the gel matrix) is cast: the gel percentage is chosen inversely to the size of the molecule. Proteins do not separate and run uniformly (as a front) if they are too small for the chosen gel percentage. To analyse molecules of 9 kDa size like Gab1$^{613-694}$ and Gab1$^{617-684}$, the gel percentage was 18 %. For larger molecules from 11-25 kDa, like the SH2 domains a 15 % gel percentage was cast.

Table 3 *Composition of SDS-gels*

ingredient	separating gel	stacking gel
1 M Tris HCl pH 8.8^{RT}	5.65 ml	-
1 M Tris HCl pH 6.8^{RT}	-	0.8 ml
10 % SDS	0.15 ml	75 µl
30% acrylamide 37.5:1	9 ml (18%)/ 7.5 ml (15%)	1 ml
50 % glycerol	-	675 µl
water	top up to 15 ml	top up to 7.5 ml
TEMED	15 µl	7.5 µl
10 % APS	150 µl	75 µl

For phosphorylated proteins, a 15 % phostag gel was cast, which interacts with the phosphate group, allowing the separation of different phosphorylated species.

Table 4 *Composition of phostag gel*

ingredient	separating gel	stacking gel
30% acrylamide (29:1)	4 ml	750 µl
1 M Tris HCl pH 8.8^{RT}	2 ml	-
1 M Tris HCl pH 6.8^{RT}	-	1.25 ml
10 mM MnCl$_2$	80 µl	-
5 mM phostag solution	80 µl	-
10% SDS	80 µl	50 µl
water	top up to 8 ml	top up to 5 ml
10 % APS	40 µl	25 µl
TEMED	8 µl	5 µl

As a running buffer, a laemmli system was used.

Table 5 *Recipe for laemmli running buffer*

ingredient	1 x laemmli Buffer
Tris	0.025 M
glycerol	0.19 M
SDS	0.1 %

Isothermal titration calorimetry

To characterise the binding of macromolecules, isothermal titration calorimetry (ITC) is a favoured method. It can provide direct determination of the enthalpy (ΔH), the stoichiometry (N) and the affinity ($K_A=1/K_D$) (**Figure 15**). By further calculation of the fitted parameters, the entropy (ΔS) and the Gibbs free energy can also be extracted using ΔG = -RTlnK_A and ΔS= (ΔH-ΔG)/T. This method is the only one that allows the exact determination of several thermodynamic binding parameters in one experiment. Nevertheless, measurements are time- and sample-consuming, the affinity range is in the nM-μM range, and kinetics (k_{on} and k_{off} rates) cannot be determined. The method relies on the heat release or absorption of a reaction (exothermal or endothermal reaction). In an experiment, the ligand is stored in a syringe, while the binding partner is placed in the measurement cell. Throughout the experiment, the ligand (normally 10 times higher in concentration than the binding partner) is periodically injected (normally with a constant volume) into the binding partner under constant stirring. When reactants interact, they release or absorb heat into the system. A reference cell, which is filled with distilled water or buffer, is used to detect the energy needed to adjust the measurement cell to the temperature of the reference cell (μcal/sec). This energy is seen as a spike in the titration profile (**Figure 15** upper panel). Upon addition of the excess ligand to the measurement cell, the binding partner gets further saturated, leading to decreasing thermal changes until full saturation. Thermal changes detected after full saturation display the dilution energy of the ligand and should be subtracted with control measurements. After the integration of the spikes of the titration profile (kcal/mol), a fit of the resulting sigmoidal curve (**Figure 15** lower panel) provides the determination of the above-mentioned parameters.
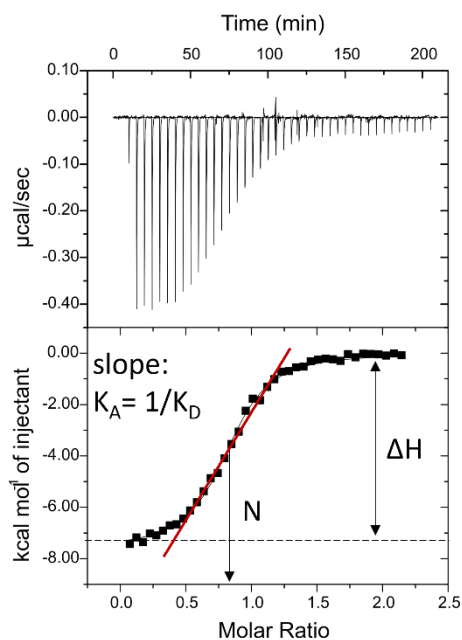
Figure 15 *ITC thermograph and analysis from those data. ΔH, N and K_A can be directly derived. Curve from a measurement of C-SH2$^{102-220}$:pY-Gab1$^{655-677}$.*

ITC measurements were done using a MicroCal™ iTC200 device. For all measurements, a 20 mM citrate buffer pH 6.0 was used, as the enthalpy of ionisation is low, resulting in low artificial heat development. For the final step in protein purification, the ligand and binding partner were both eluted or dialysed against this buffer. 350 µl of the binding partner (the different SH2 domain constructs), with a concentration of 20 µM - 100 µM, was carefully injected into the measurement cell. 40 µl of ligand (the different Gab1 constructs), with a concentration of 200 µM – 1 mM, were transferred into the syringe. For calculation of the concentration needed, the following equation was used, where [M] is the concentration of the binding partner and K_A is the association constant.

(1) $10 < K_A[M] < 1000$

The concentration of the syringe is normally 10 times higher than the binding partner. The titration volume was normally set to 2 µl, resulting in 19 injections per experiment. The first injection of 0.5 µl is discarded from the data analysis but essential to the experiment, as the small volume is needed to minimise the impact of potential equilibration artefacts. The temperature was kept constant at 25°C. The data analysis was performed in MicroCal ITC-ORIGIN analysis software; the data were fitted to a software-integrated formula for a one-set of sites binding mode.

UV/Vis-spectroscopy

To determine the concentration of DNA and protein preparations and detect potential contamination, UV/Vis-spectroscopy was used. The method relies on the absorption of

light by a molecule of a specific wavelength in the ultraviolet and visible spectra. Molecules absorb light of a wavelength energetic enough to excite an electron of a bond at a higher energy level. The n-> pi* and pi->pi* transitions of the peptide bond absorb light of 190 nm - 220 nm, the aromatic ring systems of tryptophan and tyrosine side chain absorb light of a wavelength of 280 nm, while nucleic acids absorb at 200 nm and 260 nm. To measure a concentration of a solution, a cuvette with a path length of 1 cm was used. The cuvette was placed in a double-beam spectrophotometer together with a buffer reference to eliminate any real-time fluctuations in the measurement. The absorbance detected is the difference between the incoming light intensity (I_0) and the transmitted intensity (I) and is expressed in relation to the concentration of the sample in the Beer-Lambert law.

$$(2)\ E = -\lg\left(\frac{I_0}{I}\right) = \varepsilon * c * d$$

where E is the extinction, ε is the extinction coefficient [L*cm^{-1}*mol^{-1}], c is the concentration [mol*L^{-1}], and d is the length of the pathway [cm]. The extinction coefficient can either be calculated from the sequence or experimentally determined. For this thesis, it was calculated from the sequence with the Expasy tool ProtParam (Gasteiger et al.).

CD-spectroscopy

Circular-dichroism spectroscopy (CD-spectroscopy) is an absorbance-based method to e.g. determine the secondary structural content of a macromolecule, including proteins. The technique relies on the same principle as UV-Vis spectroscopy, with the difference that circularly polarised light of both directions is used to illuminate the sample. Chiral sample molecules absorb left- and right-handed light differently. This absorption is measured over a spectral width from 190 nm to 260 nm (far-UV). The resulting spectrum shows the resulting ellipticity for each wavelength, which is influenced by secondary structural features. The intensity and energy of the n->pi* (210-220 nm) and the pi->pi* (190 nm) transition of the peptide bond depends on the psi and phi angles of the backbone and, thus, on the secondary structure. The mean residual ellipticity (MRW) [degree * cm^2*dmol^{-1}] was calculated from the measured ellipticity [Θ].

$$(3)\ MRW = \frac{[\theta]}{c*d*N_{aa}*10}$$

Curve fitting and comparison to reference spectra is carried out to gain insights into the structural composition of the protein. For this thesis, the resulting spectra from the Gab peptides used were typical for a random coil, and no further analysis was done. The experiments were measured at a JASCO J-815 device (group of Prof. Dr. rer. nat. Jochen Balbach) at 20°C, in a range of 260-190 nm, a scanning speed of 50 nm/min and a band

width of 1 nm with 10 accumulations for each measurement. Hellma-cuvettes with a layer thickness of 1 mm or 10 mm were used.

Structure determination

Standard techniques used to determine the structure of a biomolecule are X-ray crystallography, NMR and (Cryo)-electron microscopy (EM). All three methods have their limitations. Large kDa molecules can be analysed with Cryo-EM, while low kDa molecules can be addressed with NMR. Crystallisation is not limited to the size of a molecule but inherently relies on its ability to form a well-ordered crystal, which prohibits the structural analysis of intrinsically disordered proteins unless they adopt a stabilised conformation, e.g., through a binding partner or artificial crystal contacts. Cryo-EM techniques are constantly improving, which by now allows assumptions on IDRs in combination with computational analysis. The only technique out of the three suitable for the structural determination of free IDPs with atomic resolution is NMR, besides mass spectrometry in combination with cross-linking and MD simulations.

NMR

Nuclear magnetic resonance (NMR) addresses the individual Larmor frequency of the nuclear spin of atoms within a molecule and can thus provide information on the measured sample at atomic resolution. The information derived from the measurement depends on the experiment chosen. They can provide knowledge on the structural position of atoms, their distances to other NMR active nuclei, and dynamic behaviour on different time scales, which help address questions from folding (CPMG) up to side chain rotation (relaxation parameters R_1 and R_2, hetNOE) to name a few possible experiments. Shorter, less complex measurements can already support the understanding of macromolecular complex formation (chemical shift perturbation; CSP) and secondary structure formation (chemical shift index; CSI). NMR provides a great toolbox to study intrinsically disordered proteins, especially with recent developments of ^{13}C-NMR (Felli and Pierattelli 2022). Still, IDPs remain a challenging target for NMR as low complexity regions, a high number of proline residues and the unfolded state of the protein provide low dispersion spectra with overlapping peaks. For this thesis, ^{15}N-NMR experiments were used on labelled and double phosphorylated pYpY-Gab1$^{613-694}$ either alone or with SH2 domains as binding partners.

Devices and general procedure

All measurements were planned together with Dr. rer. nat. Tobias Gruber (groups of Prof. Dr. rer. nat. Jochen Balbach and Prof. Stephan M. Feller, PhD), The sample was prepared by the author. The prepared sample was measured, processed, and experiments other than assignment, CSP (chemical shift perturbation) and CSI (chemical

shift index) were analysed mainly through Dr. rer. nat Tobias Gruber. The 2D and 3D experiments were measured on a Bruker Avance III 800 MHz spectrometer equipped with a CP-TCI cryoprobe at 25°C. Additional magnetic fields were applied when measuring the relaxation rates used to calculate the order parameter S^2. Avance III 600 MHz and a Bruker DRK spectrometer with a proton frequency of 500 MHz were also used for those experiments. The proteins were measured in 20 mM BisTris/HCl, 50 mM NaCl pH 6.0 10% D2O (v/v) with sodium trimethylsilylpropanesulfonate (DSS) as standard in each sample and were referenced on it. All spectra have been processed using NMRPipe and analysed in NMRView.

Two-dimensional experiments
Two-dimensional experiments were used to address the quality of the sample as well as for assignment and CSP experiments. ^{15}N-^1H TROSY and ^{15}N-^1H HSQC spectra were used, which provide information on the backbone amide.

Three-dimensional experiments
For the assignment of the two-dimensional spectra, three-dimensional experiments were conducted, namely HNCA, HNCACB, HN(CO)CACB, HNCO and HN(CA)CO. For these, double labelled ^{15}N, ^{13}C- protein samples were prepared at around 500-750 µM. These experiments allow magnetisation transfer on several nuclei within the backbone and to the Cβ of the side chain as well as from preceding residues, allowing a sequential arrangement, which enables the assignment of the cross-peaks in the corresponding two-dimensional spectra.

hNOE
The ^1H-^{15}N steady-state heteronuclear NOE experiment was measured in a proton-saturated and unsaturated state. The intensity of the hNOE derives from the ratio of the saturated and unsaturated intensity for each ^{15}N-^1H cross peak. Decreased hNOE, compared to the average hNOE, is indicative of fast motion on the ps-ns time scale.

$$(4)\ hNOE = \frac{Intensity_{\ saturated}}{Intensity_{\ unsaturated}}$$

R_1 and R_2
The longitudinal relaxation time (T_1), whose inverse provides the relaxation rate 1 (R_1 =1/T_1), describes the process of the net magnetisation returning to its equilibrium as a function of time. The magnetisation from the z to -z-axis was inverted by a 180° pulse followed by varying delay times of 0 ms, 50 ms, 100 ms, 150 ms, 200 ms, 300 ms, 500 ms, 750 ms, 1000 ms and 1500 ms. For detection, the magnetisation is flipped to the xy-axis by a 90° pulse. The transversal relaxation (T_2), providing the relaxation rate 2

($R_2=1/T_2$), as an inverse of T_2, describes the decay of the net transversal magnetisation over time. The magnetisation is excited by a 90° pulse, and a series of measurements with varying spin-lock periods of 0 ms, 8 ms, 16 ms, 32 ms, 48 ms, 64 ms, 80 ms, 96 ms, 112 ms, 160 ms is recorded. The intensity of each peak in a series of experiments was determined by NMRview (Johnson 2004) or PINT (Niklasson et al. 2017) and fitted to the following equation.

$$(5)\ I(t) = I_0 * e^{(-Rt)}$$

Calculation of order parameter
The order parameter S^2 was calculated using the input of R_1, R_2 and hNOE measured at 800 MHz, 600 MHz and 500 MHz. using the Lipari-Szabo model-free analysis (Lipari and Szabo 1982) processed in the relax suite (nmr-relax.com).

CSP calculation
To determine the change in the chemical environment of each backbone amide upon the addition of a potential binding partner, chemical shift perturbation (CSP) was done. The isotope-labelled molecule is measured in a two-dimensional experiment in its free state as well as in solution together with its potential binding partner. After assigning the cross- peaks, the chemical shift position in the ^1H and ^{15}N dimensions of peaks corresponding to the same amino acid were extracted and used to calculate the CSP according to the following formula (Gruber 2014 // 2015; Williamson 2013):

$$(6)\ \sqrt{\frac{\Delta\delta(^1H)^2 + \frac{1}{25}\Delta\delta(^{15}N)^2}{2}}$$

Only CSP values within the same system can be compared: small shifts indicate minor influence upon binding, while large perturbations demonstrate an involvement of the amino acid in the binding event, either at the binding site and/or in induced conformational changes.

CSI calculation
To get an idea of the potential of a peptide to form certain secondary structure elements, the chemical shift index (CSI) is a valid analysis method. A measured spectrum of a protein is compared to a calculated chemical shift dataset of the same protein in the denatured state. The dataset used for the calculation considered phosphorylation at the tyrosine residues (Hendus-Altenburger et al. 2019). The CSI is calculated as seen in equation 7.

$$(7)\ \mathrm{CSI} = (C\alpha - C\beta)_{exp} - (C\alpha - C\beta)_{calc}$$

Positive values indicate helical tendency, while negative values show regions of probable extended conformation (Luca et al. 2001).

Crystallisation

The crystallisation of molecules is a naturally occurring process in nature and an established procedure in laboratories. Common to all crystals is a periodic arrangement of atoms leading to a solid state. Knowledge of the crystal form, the diffraction pattern and other physical properties of the crystal can be used to identify the crystallised substance. For macromolecules, crystallisation is commonly used to identify the atomic structure.

As this thesis focused on the structural determination of an IDP fragment (Gab1) with a binding partner (Shp2), crystallisation was one method of choice. The resulting crystals were either used for x-ray diffraction or electron diffraction, which will be further described below.

General biophysical process of crystallisation

Each molecule has a unique solubility curve, describing the solubility (solvated mass of molecule per volume solvent) against, e.g. the temperature or pressure. As the solubility of many molecules increases with increasing temperatures, cooling the solution carefully results in supersaturation. This can either lead to uncontrolled precipitation or nucleation (in the so-called labile state) and crystal growth (in the metastable state) and must thus be carefully controlled. This solubility curve can either be experimentally determined to yield desired crystals or, as in most cases, the solubility properties of the undersaturated solution are varied blindly with high throughput testing (initial screen) of precipitants, ionic strength, pH and other additives (**Figure 16**).
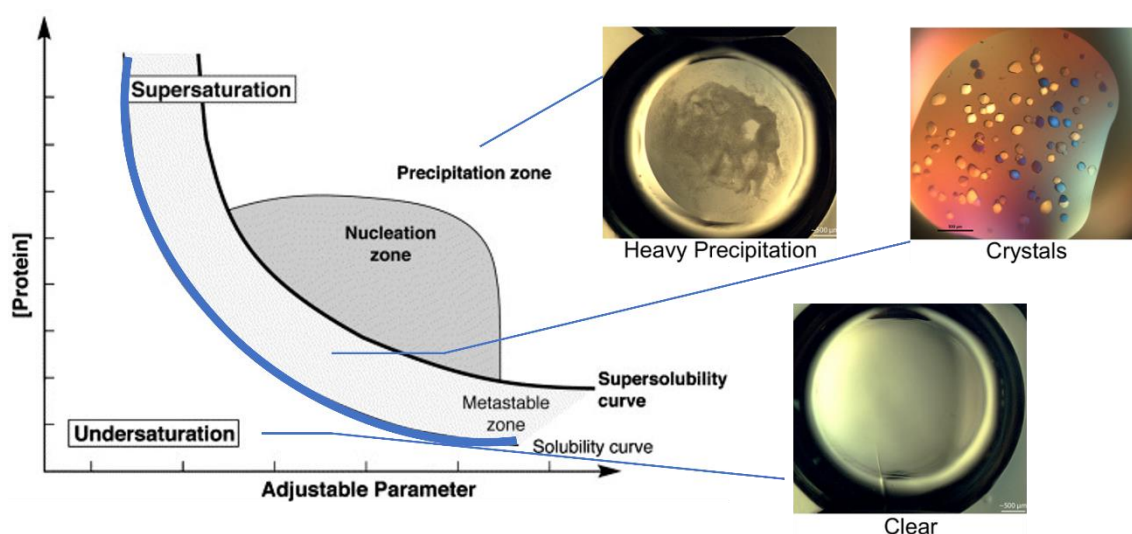
Figure 16 Solubility curve of an idealised protein. The protein concentration and other parameters (e.g. temperature, pressure, pH, ionic strength etc.) are responsible for the solubility of the protein. This combination can either lead to

To proceed from the undersaturated phase to the nucleation and subsequently to the metastable phase, the method of choice in this thesis was vapour diffusion. A well is filled with reservoir solution, and a drop of protein mixed with mother liquor (mostly in a 1:1 volume) is placed on a coverslip or a drop shelf. The system is tightly sealed and stored at a constant temperature. As the precipitant concentration of the reservoir solution is higher than in the protein drop, diffusion of water from the small sample drop to the reservoir solution occurs, increasing the protein concentration in the drop. If the protein and precipitant concentrations are sufficient, this can lead to supersaturation, resulting in nucleation and/or precipitate. Otherwise, the drop stays clear. Depending on the degree of supersaturation, it can either result in many tiny crystals or in a few but large ones. Once a condition is found that can yield crystals, seeding is an option to have tight control over the growth process. To validate and optimise the crystallisation conditions selected from the initial screen, the experiment is set up in a larger volume while varying the buffer composition (e.g., pH, ionic strength, precipitant concentration). This fine screen is used to find the environment that yields the most suitable crystals for the following diffraction experiment.

In general, crystals form upon repeated identical interactions of molecules (crystal contacts). These mostly weak interactions (ΔH^0_{cryst}) will only contribute slightly to the negative free Gibbs energy of crystallisation (ΔG^0_{cryst}) needed. Crystallisation is driven entropically as the release of the solvent from the solvation shell of the individual molecules is positive ($\Delta S^0_{solvent}$), although the ordered packing also results in negative $\Delta S^0_{protein}$ (Derewenda and Godzik 2017). The thermodynamic of crystallisation is described with the following equation (8):

$$(8) \quad \Delta G^0_{cryst} = \Delta H^0_{cryst} - T(\Delta S^0_{protein} + \Delta S^0_{solvent})_{cryst}$$

Upon formation of a well-ordered protein crystal, diffraction experiments can be facilitated. Therefore, a radiation source, either X-ray, electrons or neutrons, will interact with the planes of the crystal and lead to diffraction. During the experiment, the phase information of the resulting wave is lost, and only the amplitude can be derived from the intensity of the diffraction peak. Nevertheless, the information of both is needed to recreate the original 'image' (structure of the molecule). This so-called phase-problem can be solved either *ab initio* with a high diffracting crystal of at least 1.2 Å or experimentally by collecting additional data on a heavy-atom modified crystal (isomorphous replacement and anomalous scattering). Another method is through molecular replacement, where the phases will be provided by an existing structure.

Crystallisation of N-SH2$^{1-106}$:pY-Gab1$^{613-651}$

The crystallisation of N-SH2$^{1-106}$:pY-Gab1$^{613-651}$ was done as described in the master's thesis of (Judith Kniest 2021).

For the initial screen, nine assay kits (see material and method part) were screened in 96- well plates, resulting in 864 tested conditions. The wells of the microtiter plates were filled manually with the buffer solutions from the kits with a volume of 50 ul. The pipetting robot Zinser Analytic Microsys 4000 was used to load the protein-buffer drop, with a volume of 400 nl, on the respective drop shelf of the crystallisation plates. The protein concentration of N-SH2$^{1-106}$ was 20 mg/ml, while the phosphorylated synthetically derived pY-Gab1$^{613-651}$ construct was added in a 1.5 molar excess. The plates were sealed with UV-transparent foil and stored in a plate hotel at 12°C. The crystal growth was monitored automatically within the plate hotel and checked after one week manually.

The initial screen showed 22 conditions where crystals could be observed. From those, five were picked for fine screening. The 15-well EasyXtal plates were filled with 500 µl of reservoir solution, and the protein:buffer drops with a volume of 1µl + 1µl were placed on the screwing cap. The drops contained either the complex, the N-SH2$^{1-106}$ domain or only the buffer as control). The complex was further screened at different protein concentrations of 20 mg/ml, 18 mg/ml and 15 mg/ml. Crystals grew within 12 hours to 3 days.

Crystallisation of the tandem SH2 domain (Shp2$^{1-222}$) with pYpY-Gab1$^{617-684}$

To understand the structural basis of the C-SH2:Gab1 interaction, crystallisation was facilitated. Unfortunately, the crystallisation experiments of the free C-SH2$^{98-220}$ and C- SH2$^{102-220}$ domains have not been successful. Both constructs in complex with pY-Gab1$^{640-694}$ and pY-Gab1$^{651-685}$ phosphopeptides were not fruitful either. An additional Shp2$^{1-222}$ tandem construct with a binding-inefficient N-SH2 domain, the N- SH2- dead$^{1-222}$, was used in a complex with pY-Gab1$^{651-685}$, which did also not yield crystals in an initial screen. The only system that was successfully crystallised was Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$.

For the crystallisation of the Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$ complex, the initial screen was conducted as described above for N-SH2$^{1-106}$:pY-Gab1$^{613-651}$. After one week, two conditions Wizard Cryo (D8/D9) (0.1M Tris pH 8.5, 25% PEG3350 and 0.1M HEPES pH 7.5, 25% PEG 3350) resulted in thin needle-like crystals. The crystal growth could be reproduced in a fine screen, varying the pH and PEG3350 concentration over 30 conditions per initial screen hit (**Figure 53**). Interestingly, the needle-like crystals started to dissolve and grew into a more compact, disc-like crystal form (**Figure 54**). These were fished and tested in the in-house X-ray source. The tested crystal diffracted to 3.5 Å, but

the diffraction pattern demonstrated severe twinning. These compact crystal forms could not be reproduced, and the crystals harvested from fine screens always appeared in a needle-like shape. Since these crystals were hard to fish and not optimal for X-ray diffraction experiments due to their size, shape, and diffraction quality, optimisation was done intensively with no successful outcome. Seeding, additive screens and matrix seeding screens were performed using the Oryx pipetting robot, and different temperatures were tried. In the end, electron diffraction on microcrystals (microED) was applied for data collection to overcome the crystal shape limitation.

To analyse the content of the crystals, these were fished, washed in fresh crystallisation buffer, collected into a reaction tube and dissolved in SDS-loading buffer. The sample was loaded onto a gel, revealing the presence of a 25 kDa protein (Shp2$^{1-222}$) and an 8 kDa protein (pYpY-Gab1$^{617-684}$). The stained protein band of the smaller protein was cut out and handed for further Edman degradation and protein sequencing to the Mass spectrometry/Proteomics in-house core facility (Matt Fuszard, PhD). The resulting gel and sequencing profile can be seen in supplemental **Figure 55**.

Data collection and processing
X-ray
A single crystal containing N-SH2$^{1-106}$:pY-Gab1$^{613-651}$ in a tetragonal shape could be fished from the fine screen (25% PEG 3350, 0.2 M NaCl, 0.1 M BisTris pH 5.5). A dataset was acquired using the in-house x-ray source (rotating copper anode, x-ray wavelength of λ= 1.54 Å) coupled to a Rigaku HyPix-Arc 150° detector, together with Dr. rer. nat. Christoph Parthier (group of Prof. Milton T. Stubbs II, PhD). 10% ethylene glycole was added as cryo-protectant The crystal was fished and mounted under cryogenic conditions. The data set was acquired during continuous rotation to measure at all 360°.

Further data analysis, processing, molecular replacement, model building and refinement were done together with M.Sc. Judith Kniest, under close supervision of Dr. rer. nat. Constanze Than-Breithaupt (group of Prof. Milton T. Stubbs II, PhD). The dataset was processed in XDS (Kabsch 2010), molecular replacement was done in PHASER (McCoy et al. 2007), refinement in PHENIX (Liebschner et al. 2019) and model building in COOT (Emsley et al. 2010). The deposited PDB structure of N-SH2$^{1-106}$:Gab1$^{621-633}$ (PDB: 4qsy (Gogl and Remenyi, unpublished)), without the Gab1-chain, was used as the template for the molecular replacement.

microED
To collect a dataset from the Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$ complex, which grew in thin, 400 μm long needles, an alternative to x-ray diffraction was needed. X-rays do not interact very strongly with matter, resulting in the need for a high number of asymmetric units in

each dimension. In microED, electrons are used as the radiation source (in our case 200 keV Thermo Scientific Glacios Cryo-Transmission Electron Microscope), while the crystal is immobilised under cryogenic conditions on a glow-discharged quantifoil holy grid R2/1. The method was established at the ZIK HALOmem, Halle, together with Farzad Hamdi, PhD (group of Jun.-Prof. Dr. Panagiotis Kastritis) and Dr. rer. nat. Alaa M.A. Shaikhqasem (group of Prof. Milton T. Stubbs II, PhD), following the methodology established in the lab of Prof.Tamir Gonen, PhD (UCLA) (Clabbers et al. 2022).

The crystals of Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$, grown at 12 °C within one week, were pipetted from the hanging-drop set-up (condition: 0.1 M Tris, pH 8.8 12°C, 29% PEG3350). The skin covering the protein drop was gently removed, and the drop was carefully transferred and pooled with others in an autoclaved 1.5 ml tube. To blot the crystal on the grid, a glow-discharged (60 seconds, 20 mA, PELCO easiGlow) grid was placed into a Vitrobot Mark IV (ThermoFisher). 3.5 ul of the crystal sample was carefully applied to the front of the grid, while 3.5 ul of the crystallisation buffer (diluted 1:1 with distilled water) was pipetted to the back. The back was blotted with standard ashless blotting paper (Ø55/20mm, Grade 595), while the front was blotted against parafilm. Blotting was done for 25 seconds at 4°C and 95% humidity. The grid was plunged into ethane and moved to the loading cassette. The blotting was done by Dr. rer. nat. Fotios Kyrillis (group of Jun.-Prof. Dr. Panagiotis Kastritis) with scientific advice by PD Dr. rer. nat. Annette Meister.

The blotted grid was autoloaded into the Glacios Cryo-TEM (ThermoScientific), and the grid was screened at low magnification STEM mode (detector: Falcon IIIEC) to find crystals that were thinly covered in ice. Electron diffraction was detected using a scintillator-based Ceta-D camera by Farzad Hamdi, PhD (group of Jun.-Prof. Dr. Panagiotis Kastritis). Data were collected on two crystals orientated perpendicular to each other. For each crystal, a total tilt range of 140° (-70° to 70°) was acquired while measuring in smaller wedges of 20° to 40° range each, while the centre was measured from -30° to 30°. For each wedge, a new spot of at least 1 µm distance was chosen on the crystal to avoid radiation damage. The electron dose was fixed to approximately 0.046 e$^-$/Å^2/s, while the total cumulative dose was set to 3-4 e$^-$/A^2 per acquisition.

Data processing was carried out using XDS (Kabsch 2010) together with Dr. rer. nat. Alaa M.A. Shaikhqasem (group of Prof. Milton T. Stubbs II, PhD). Molecular replacement, model building and refinement was performed using the CCP4i2 suite (Potterton et al. 2018) with PHASER (McCoy et al. 2007), REFMAC5 (Murshudov et al. 2011) and COOT (Emsley et al. 2010), together with by Dr. rer. nat. Constanze Than-Breithaupt (group of Prof. Milton T. Stubbs II, PhD) and Prof. Milton T. Stubbs II, PhD. For molecular

replacement, the Shp2$^{1-222}$ crystal structure in complex with a phosphorylated TXNIP-peptide was used (PDB: 5df6, (Liu et al. 2016)). The model was prepared as follows: The phospho-peptides were deleted, and the domains were separated into the N-SH2 (residues 1-104) and the C-SH2 (residues 108-220) domains. The crystal belonged to space group P2$_1$2$_1$2$_1$, with one complex in the asymmetric unit: TFZ scores were obtained of 23.1 for ensemble 1 (N-SH2) and 10.4 for ensemble 2 (C- SH2). For refinement, a restrained jelly body was performed using form factors corresponding to electron diffraction. The initial model building was guided by the preexisting structures of the N-SH2$^{1-106}$:Gab1$^{621-633}$ (PDB: 4qsy (Gogl and Remenyi, unpublished)) and the C- SH2$^{108-220}$:TXNIP (PDB: 5df6, (Liu et al. 2016)) allowing positioning of the phosphorylated Gab1 peptides. After refinement, additional density was visible, especially at the C-SH2 domain surface. A Gab1 fragment from an AlphaFold2 (Jumper et al. 2021) generated model (without PDB-templates, MSA only; ColabFold (Mirdita et al. 2022), Notebook: AlphaFold2_complexes (Evans et al. 2021) of Shp2$^{1-593}$:Gab1$^{611-694}$ fitted the electron density. The coordinates were further modelled into the density by hand using COOT.

After further cycles of manual building and refinement, the structure of the complex was solved to 3.2 Å from a dataset with 89 % completeness, with modelled structures of the Shp2-fragment (aa 4-221) and two Gab1 fragments from aa 622-631 and aa 653-672.

Results and Discussion

Chapter 1: Secondary structure of the C-terminal region of Gab1 in the unbound state

The Gab1 protein has a folded PH domain and a predicted long, disordered tail (**Figure 17**). Along the tail, the only secondary structure elements that have been identified are the poly-proline type II (PPII) and 3_{10} helices at the Grb2 binding sites (Harkiolaki et al. 2009) (**Figure 5**). The Shp2 binding site of Gab1 is located at the very C-terminal region of the disordered tail (Gab1$^{613-694}$). Thus, this protein fragment is expected to be mainly disordered. Computational analysis of the structure of Gab1$^{613-694}$ was done by several disorder and secondary structure predictors. For all these tools, the unphosphorylated sequence was used for calculation.
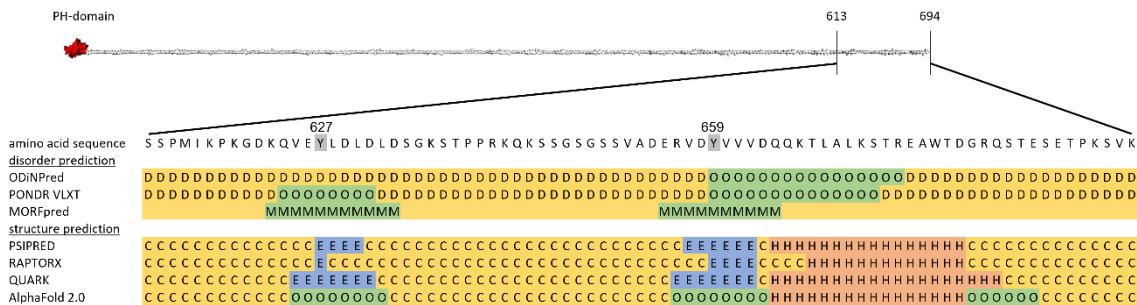
Figure 17 *Summary of the output from disorder and structure prediction tools. Coiled (C) and disordered (D) regions are coloured in yellow, while extended (E) and helical regions (H) are depicted in blue and red, respectively. Ordered (O) regions and those predicted to be MoRF (M) are coloured in green. Illustration is taken and modified from (Gruber et al. 2022).*

Some programmes predicted ordered regions for the amino acids comprising the two tyrosines (Y) (**Figure 17**). An extended conformation is predicted for residues -2 to +3 /+4, relative to the tyrosines. Interestingly, a larger region of order is calculated for the amino acids C-terminal to Y659, where a helix is predicted for aa 664-680.

To investigate its secondary structure experimentally, circular dichroism (CD) in the far- UV range and predictions were used. The Gab1$^{613-694}$ fragment in its unphosphorylated and doubly phosphorylated state shows one sharp minimum at 200 nm and another weak minimum at around 210-240 nm as depicted in **Figure 18, A, B**. This result was measured by Dr. rer. nat. Tobias Gruber and published in Gruber et al. 2022.
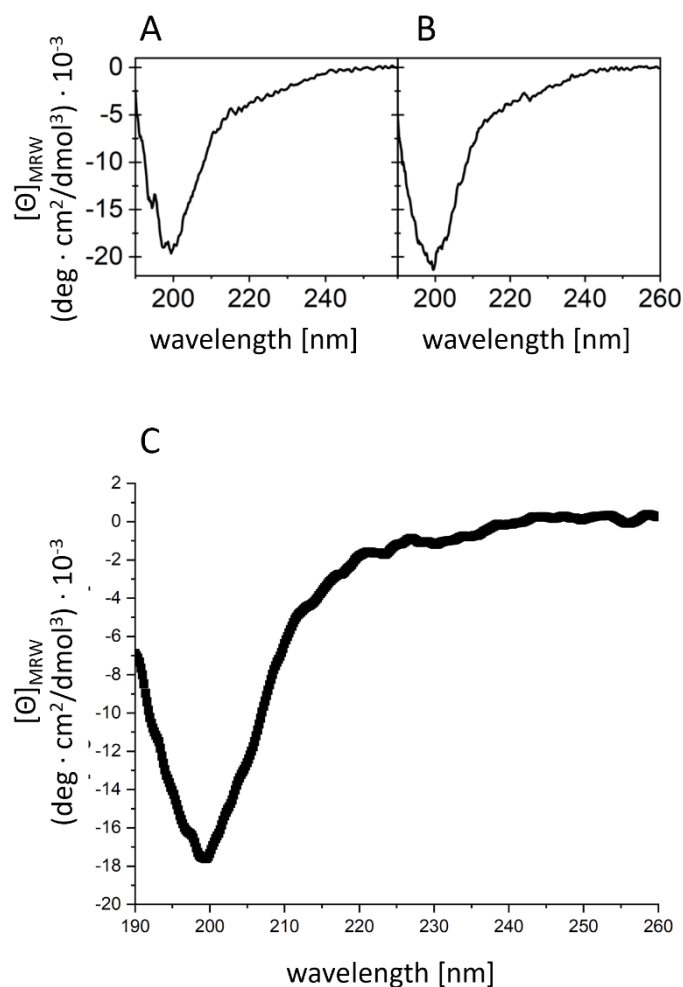
Figure 18 *CD-spectra of Gab1-fragments indicate a disordered protein. A) CD-spectra of unphosphorylated and B) phosphorylated Gab1$^{613-694}$ (Gruber et al. 2022). C) CD-Spectrum of phosphorylated pY-Gab1$^{655-677}$. Figures created in Origin.*

To investigate the structure of the region around Y659 (predicted to be helical – see **Figure 17**), a phosphopeptide from aa 655 to 677 (pY-Gab1$^{655-677}$) was measured by CD. This peptide shows the same typical IDP spectrum as the unphosphorylated and phosphorylated Gab1$^{613-694}$, with a negative ellipticity at 200 nm and a weak minimum from 210-240 nm (**Figure 18**, C). Indicating that the C-terminal fragment of Gab1 in solution is disordered, in contrast to its prediction.

^{15}N, ^{1}H TROSY NMR-measurements of the unbound Gab1$^{613-694}$ in its unphosphorylated and phosphorylated state was facilitated by Dr. rer. nat. Tobias Gruber (groups of Prof. Dr. rer. nat. Jochen Balbach and Prof. Stephan M. Feller, PhD) (Gruber et al. 2022).

Every assigned peak corresponds to one amino acid. For the unphosphorylated and the phosphorylated free states of Gab1$^{613-694}$, a low dispersion of the proton chemical shifts of these ^{15}N-^{1}H cross peaks is seen in **Figure 19**.
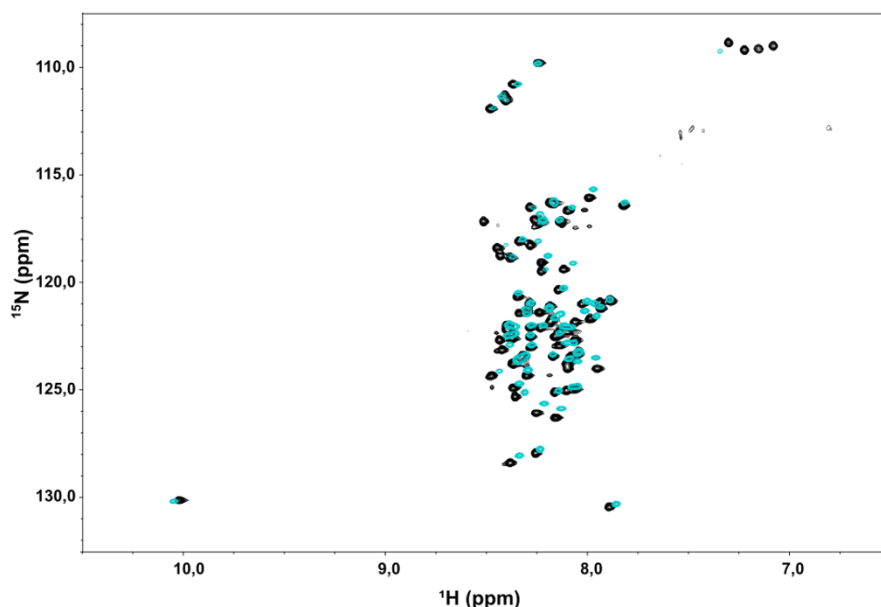
Figure 19 *Superposition of the ^{15}N, ^{1}H-TROSY spectra of unbound unphosphorylated (black) and phosphorylated (turquoise) Gab1[613-694]. Slight chemical shift changes can be seen upon phosphorylation, but the low dispersion, typical for an IDP, is still evident. Data also published in (Gruber et al. 2022). Created with NMRView.*

This is indicative of an IDP since no secondary structure or tertiary fold is present. Instead, every amino acid experiences the same chemical environment and shows similar values for the amide protons at around 8.1-8.3 ppm; these values represent the typical chemical shifts of a randomly coiled protein. Minor changes for the regions around the two tyrosines are present upon phosphorylation, but besides that, no structural changes are seen (supplemental **Figure 49**). This coincides with the findings from the CD experiments.

1.2 Discussion

1.2.1 Prediction

Computational prediction of secondary or tertiary structures is based either on a homology-based or an *ab initio* approach. For homology modelling, the structure of a homologue protein is used as a template, and the query protein is adjusted, assuming that two homologous proteins fold identically (PSIPRED uses this approach (Jones 1999)). If no close homologues are found, threading can be used for sequence identities of under 25%. This technique relies on the principle that the structure has a higher conservation than the sequence and that only a limited number of folds exist in nature. With the help of a scoring function, the optimal relationship between sequence and structure is calculated (RaptorX relies on a combination of threading and *ab initio* (Källberg et al. 2012)). For *ab initio* calculations, only the sequence is used to predict the structure based on physicochemical properties extracted from the amino acid composition. This template-free method can identify novel folds (QUARK works *ab initio*

(Mortuza et al. 2021)). All three structure prediction methods come up with similar results for Gab1.

Disordered proteins, by definition, sample the conformational space and have thus no defined coordinates. Therefore, structural templates as needed for homology modelling do not exist unless they are found in a structure-induced conformation, e.g., due to protein binding, homology modelling for IDPs is therefore complicated. Nevertheless, through the information on the evolutionary conservation of sequences, IDRs can be extracted, as these regions evolve faster than in structured homologues. Moreover, structure predictors are trained on structured proteins, and IDRs perform, therefore, worse (with low accuracy) in those tools (Hsu et al. 2020).

Disorder predictors have evolved to fill the gap. As disordered proteins are hard to analyse, their prediction provides first insights into the system. IDPs have some common features in their sequence, which suit a sequence-based approach, e.g., they are low in complexity, high in overall charge, proline-rich and have a high hydrophilic-to-hydrophobic ratio. This allows the general identification of disordered regions. Disorder and order predictors initially identified IDRs also from the absence of modelled amino acids in X-ray crystallographic data. Later, disorder predictors where trained on other experimental data, e.g. from CD-spectroscopy and (small-angle x-ray scattering) SAXS, which indicate disorder globally but not per residue. A sensitive experimental method to get per residue information of a disordered region are chemical shifts from NMR experiments. PONDR-VLXT was trained on the sequence and experimental data from those methods (Romero et al. 2001). ODiNPRED, is sequence-based but relies on NMR data alone (Dass et al. 2020).

Besides those general disorder predictors, further regulatory regions of IDPs can be predicted like MoRFs (molecular recognition features – that fold upon binding) and SLiMs (short linear sequence motifs – which are disordered interaction sites). For Gab1 the regions around both tyrosines are predicted to be MoRFs thus, they fold upon binding. SLiMs could not be detected.

With the release of AlphaFold2 (Jumper et al. 2021), artificial neural network learning was introduced to the structure prediction community. It relies on multiple sequence alignments (MSA) and, optionally, on structural templates. The MSA is used to comprehend mutational events of amino acids that have taken place throughout evolution and correlate them with the spatial proximity of those amino acids in a folded protein. AlphaFold2 scored best in template-free and template-based methods in the CASP14 (Critical Assessment of Techniques for Protein structure Prediction 2020)

competition (Bertoline et al. 2023). The Critical Assessment of Protein Intrinsic Disorder (CAID) experiment is a new community-based test, similar to CASP, to validate the accuracy of predicted disorder and regions of conditional folding (Necci et al. 2021). SPOT-Disoder2 and ANCHOR2 have each scored best in the above-mentioned categories. AlphaFoldDB and AlphaFold2 were separately assessed and can compete with the best predictors (Piovesan et al. 2022; Wilson et al. 2022). The pLDDT-score for each amino acid is a good indicator of disorder, and a combination with the relative solvent accessibility provides a good measure of the sites for binding regions.

Overall, for the prediction of unphosphorylated Gab1[613-694], all used calculators resulted in similar outputs. The extended prediction at the tyrosines is likely driven by the surrounding hydrophobic amino acids, while the helical propensity is due to the enrichment of amino acids with a low energetic penalty for helix formation. Whereas the regions predicted to be disordered have some prolines and hydrophilic amino acids, while the charges are balanced and found in the structured parts.

1.2.2 CD
To investigate the structural composition experimentally, an easy, fast and low-on-sample concentration needed method is circular dichroism (CD).

ForpYpY-Gab1[613-694] the CD spectrum is identical in the unphosphorylated and phosphorylated states. Both show a typical CD-spectrum of a disordered protein (Gruber et al. 2022). Moreover, the predicted helical phosphopeptide region (pY-Gab1[655-677]) exhibits an IDP-like spectrum. This could be due to missing amino acids (the prediction showed helicity for aa 664-680) to the instability of individual helices in water. The penalty of breaking H- bonds with the water and forming entropically unfavoured H-bonds between the peptide bonds keeps the peptide in its disordered state (Yang and Honig 1995). In a hydrophobic environment, as found in the context of a tertiary structure, a membrane, or a hydrophobic solvent like TFE, these peptides can adopt a helical structure. (Gruber et al. 2022) showed that pYpY-Gab1[613-694] adopts the typical CD-spectrum of a mostly α-helical protein in the presence of 30% TFE. NMR measurements of ^{15}N, ^{13}C-labelled pYpY-Gab1[613-694] at 25% TFE showed helical propensity at the region of aa 655-680 (Gruber et al. 2022).

1.2.3 NMR
The 2D-NMR spectra clearly show typical IDP-like low dispersion of the peaks. This is in unity with the findings of the CD-measurements. Nevertheless, upon phosphorylation, slight chemical shift changes can be seen around the phosphorylation sites (supplemental **Figure 49**).

Chapter 2: Gab1 binds to the SH2 domains of Shp2

It is known from the literature that Gab1 is an activator of Shp2, after phosphorylation at Y627 and Y659 (Cunnick et al. 2001). To further investigate the interaction between these two proteins, isothermal titration calorimetry (ITC) was conducted on different complexes, from which the affinity and thermodynamic parameters can be derived. To investigate the binding on a more detailed structural level, nuclear magnetic resonance spectroscopy (NMR) was used, which allows insights per residue into structural and environmental changes of isotope-labelled proteins upon binding.

2.1 ITC measurements of Gab1 and Shp2 demonstrate moderate affinity of the single domains but tight binding when tandem SH2 domains are present

2.1.1 tandem SH2 domain

The interaction of $Shp2^{1-222}$ to bis-phosphorylated $pYpY\text{-}Gab1^{613-694}$ was measured by ITC, revealing a high-affinity binding of $K_D = 37$ nM, assuming a simple one-binding mode. The thermodynamic parameters ΔH, ΔG and $-\Delta T \cdot S$ derived from these measurements are shown in **Figure 20**. The reaction is driven by enthalpy, as is expected from interactions of SH2 domain with phosphopeptides.

	$Shp2^{1-222}$: $Gab1^{613-694}$	$N\text{-}SH2^{1-106}$: $Gab1^{613-94}$	$C\text{-}SH2^{102-220}$: $Gab1^{613-94}$
N	0.98 ± 0.07	1.25 ± 0.08	1.14 ± 0.10
K_D [nM]	37 ± 13	3089 ± 1881	2712 ± 156
$-\Delta S \cdot T$	9550.73 ± 477.86	1507.89 ± 658.72	4084.65 ± 280.74
ΔH	-19743.33 ± 273.41	-5924.25 ± 369.58	-11676.66 ± 192.58
ΔG	-10144.00 ± 180.40	-7432.14 ± 315.58	-7592.01 ± 101.79
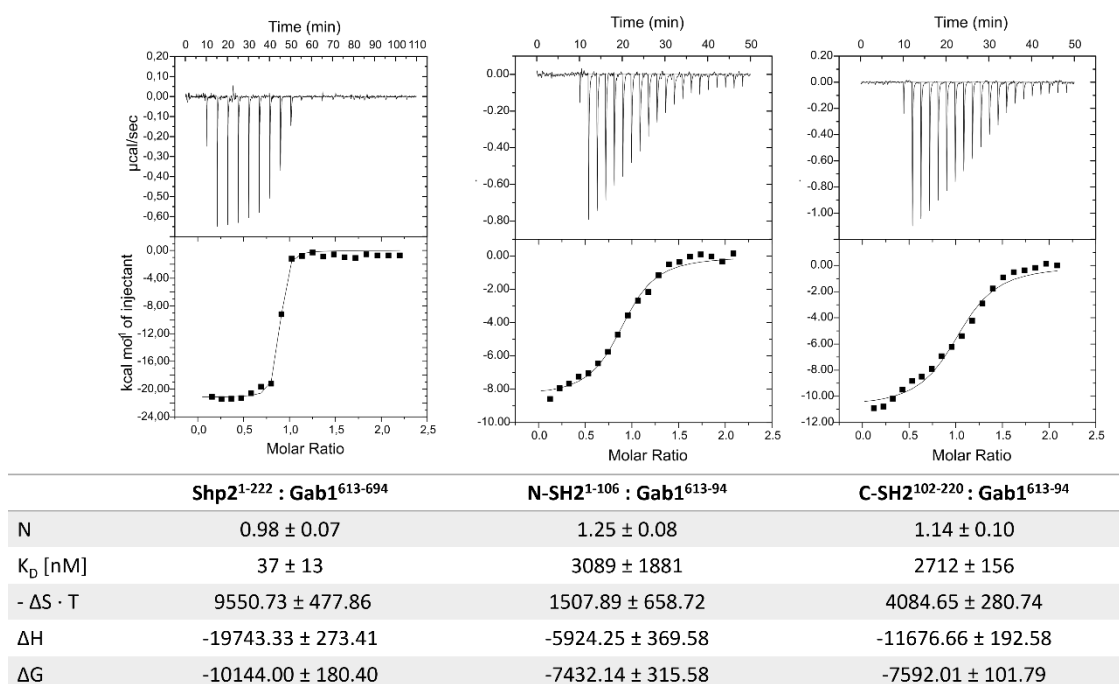
Figure 20 ITC thermographs and thermodynamic parameters derived from interactions of double phosphorylated pYpY-Gab1$^{613-694}$ to Shp2$^{1-222}$, N-SH2$^{1-106}$, and C-SH2$^{102-220}$. n=3-4. Replicates of those measurements can be found in supplemental *Figure 50*. Analysed with Origin.

2.1.2 single SH2 domains

The binding of double phosphorylated $pYpY\text{-}Gab1^{613-694}$ to the single SH2 domains, $N\text{-}SH2^{1-106}$ and $C\text{-}SH2^{102-220}$, showed decreased affinity of $K_D = 3.0 \pm 1.8$ µM and

2.7 ± 0.37 µM, respectively, compared to the interaction to Shp2[1-222], although the curves appear to show small deviations from a 1:1 binding mode. The thermodynamic contributions for both single reactions are again enthalpy driven but reduced compared to the reaction with the tandem SH2 construct. The reaction of N-SH2[1-106] showed decreased enthalpy and entropy compared to the interaction of pYpY-Gab1[613-694] with the C-SH2[102-220]. This indicates that additional interactions between C-SH2[102-220] and pYpY- Gab1[613-694], together with a more significant reduction in the flexibility of the system is present. The sum of the enthalpic and entropic terms of both single domain interactions do not add up to the once measured for the interaction of the tandem.

When measuring the interaction of the single SH2 domains with mono-phosphorylated Gab1 peptides, pY-Gab1[613-651] and pY-Gab1[640-694], a clear preference for the N-SH2 domain towards the pY627 and the C-SH2 domain towards pY659 is seen. Although, the N-SH2 domain might, very weakly, interact with the pY659 site – indicative from the thermograph (**Figure 21**). These data have been measured, analysed and first presented in the thesis of (Judith Kniest 2021)
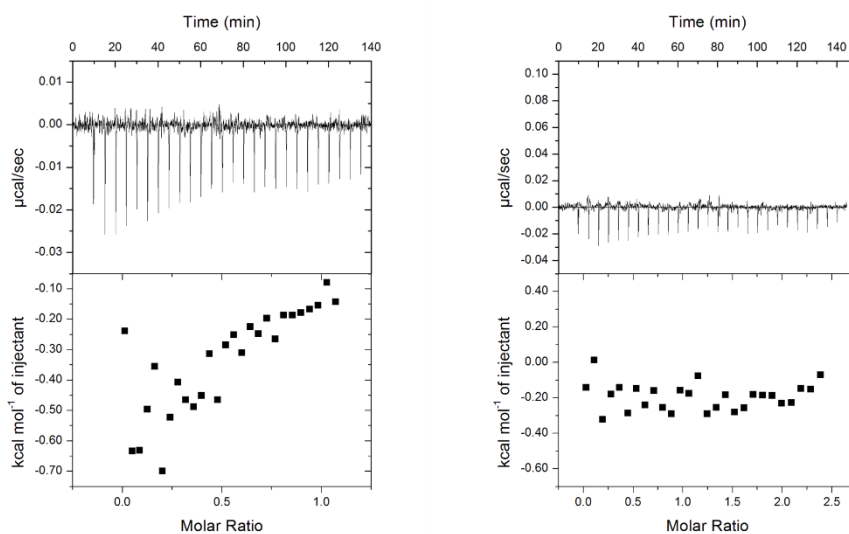
N-SH2[1-106] : Gab1[640-694]   C-SH2[102-220] : Gab1[613-651]

Figure 21 ITC reaction of the single SH2 domains with monophosphorylated Gab1 peptides. Replicates of those measurements can be found in the supplemental Figure 50. Analysed with Origin.

2.2 NMR measurements demonstrate a preferred binding orientation of the domains and reveal structural changes upon binding

2.2.1 *15N, 13C pYpY-Gab1[613-694] binding to single domains*
[15N],[13C]-labelled phosphorylated pYpY-Gab1[613-694] was phosphorylated and used for 2D and 3D NMR-measurements as descried in chapter 1.

To gain information on the bound Gab1, $^{15}N,^{13}C$-labelled and phosphorylated pYpY- Gab1$^{613-694}$ was measured in complex with the single SH2 domains. For both complexes, HNCA, HNCOCA and HNCOCACB spectra and the 2D ^{15}N, ^{1}H-TROSY have been measured and assigned. When adding either the N-SH2$^{1-106}$ or the C-SH2$^{102-220}$ to the ^{15}N, ^{13}C- labelled, bis-phosphorylated pYpY-Gab1$^{613-694}$, higher dispersion is seen, indicating that regions of the IDP are now within a new chemical environment. The superposition of all three spectra is seen in **Figure 22**.
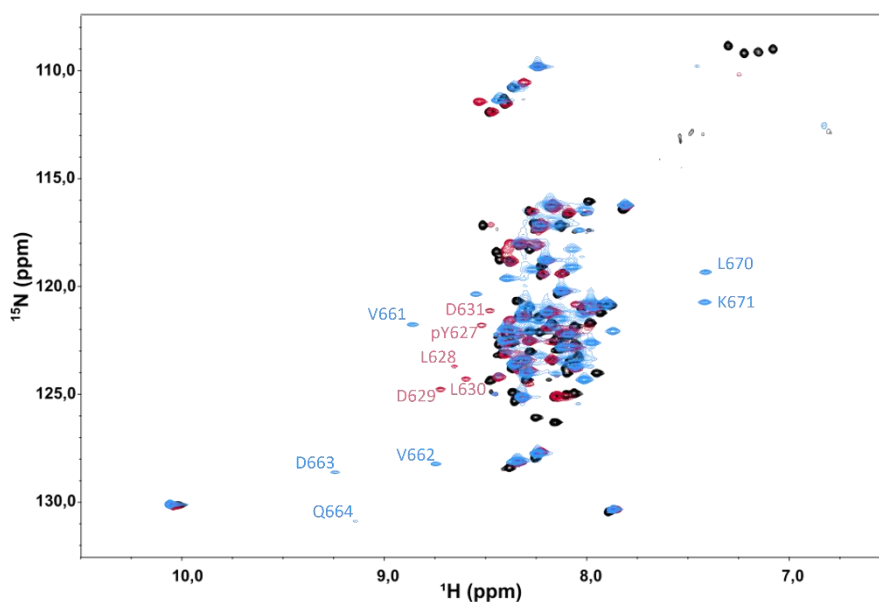
Figure 22 Superposition of ^{15}N, ^{1}H TROSY-spectra of the free phosphorylated pYpY-Gab1$^{613-694}$ (black) and in complex either with the N-SH2$^{1-106}$ (red) or the C-SH2$^{106-220}$ (blue) domain. The spectra of both complexes show higher dispersion of some amino acids, which are labelled where assigned. Most of the peaks show low dispersion, like the unbound molecule, indicating an intrinsic disorder. Created with NMRView.

That can be due to the binding of a domain and, thus a different chemical environment or to structural reorientation. The difference in the position of a specific peak upon some experimental changes (addition of binding partner, pH titration etc.) is called chemical shift perturbation (CSP). In total, 85 % (63 aa/ 74 aa (without prolines)) of the bis-phosphorylated pYpY-Gab1$^{613- 694}$ peptide could be assigned in complex with the N- SH2$^{1- 106}$ domain and 69 % (51 aa/ 74 aa (without prolines)) for the C-SH2$^{106-220}$ domain. A larger unassigned region, aa 637-653, in the bound state to the C-SH2$^{106-220}$ domain is noteworthy. This pY-linker region could be in some unsuitable state that cannot be detected in this experiment, resulting in loss of peak information.

In **Figure 23 A**, lower panel, the CSPs upon addition of the N-SH2$^{1-106}$ domain are plotted against the residue numbers of pYpY-Gab1$^{613-694}$. Some regions could not be assigned and are indicated by a red bar. Minor perturbations are seen for regions of aa 613-624, aa 645-655 and aa 670-690. These amino acids have a similar chemical environment compared to the free state. On the other hand, regions aa 627-636 and aa 656-663,

63

which flank the two phosphorylation sites pY627 and pY659 show more significant chemical shift changes.
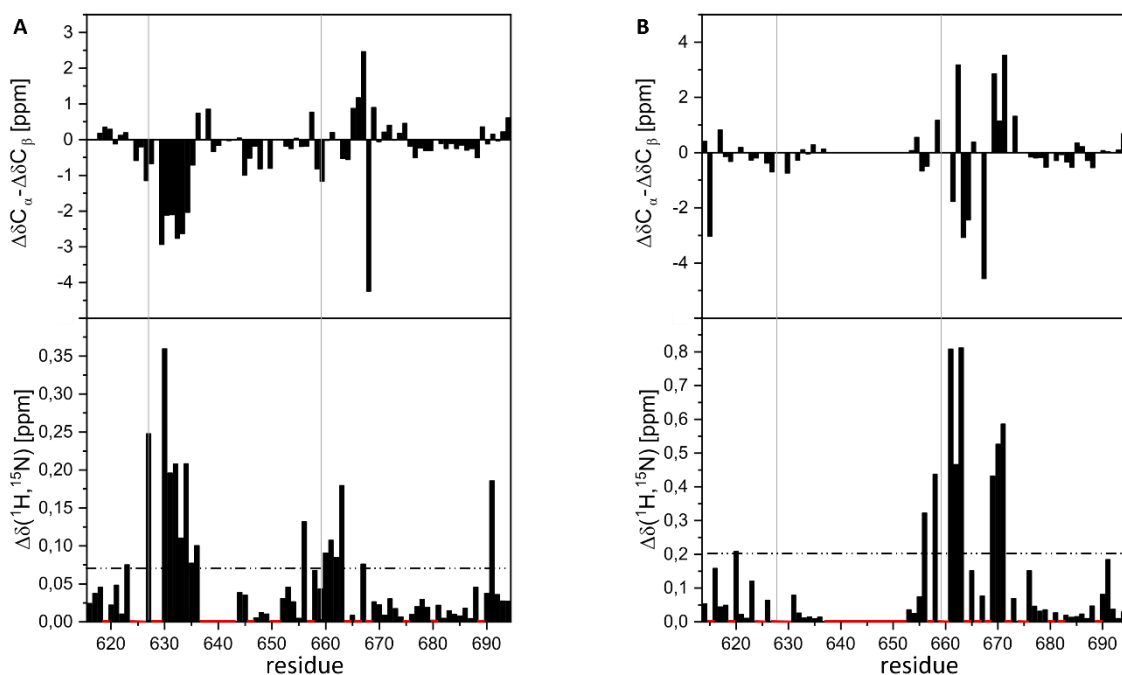
Figure 23 *Chemical Shift Indices (CSI) (top) and Chemical Shift Perturbations (CSP) (bottom) of the phosphorylated* *[15]N, ^{13}C-pYpY-Gab1[613-694] in complex with the N-SH2 (A) and the C-SH2 (B) domain. Red bars indicate regions that have not been assigned or could not be analysed. As CSP analysis requires two assigned spectra (free and bound), regions which do not yield CSP values are higher compared to CSI analysis, which compares the data of the complex to a computational data set. Grey vertical lines indicate the pY627 and pY659 positions. Black horizontal dashed lines indicate the threshold of 1σ. Created with Origin.*

As CSPs can be caused by binding and the Chemical Shift Index (CSI), which is sensitive to the torsion angle of the backbone and side chain and thus describes the secondary structure, were calculated for each residue. By comparing the measured chemical shift values of Cα and Cβ atoms of each assigned amino acid to those values of a perfect random coiled phosphorylated protein (Hendus-Altenburger et al. 2019; Luca et al. 2001), one can derive information on the secondary structure constitution of the observed protein: Negative CSI values smaller than -0.7 ppm are attributed to extended conformations; positive values greater than 0.7 ppm tend to be helical, while values between -0.7 ppm and 0.7 ppm correspond to disorder. Thus, an extended conformation of bis-phosphorylated ^{15}N, [13]C pYpY-Gab1[613-694] can be inferred for the region of aa 630-635 upon interaction with N- SH2[1- 106], which was not present in the unbound pYpY-Gab1 fragment (Gruber et al. 2022). The positive values for aa 665-668 (already present in the unbound state (Gruber et al. 2022)) could hint towards a helical conformation, although more than four amino acids would be needed for a complete α-helical turn. This region lies within a stretch of elevated CSP values (**Figure 23 A**; lower panel), suggesting that binding to the domain elicits structural reorganisation. T667, as a single amino acid, has

a very prominent negative CSI value, which can not be attributed to an extended conformation but might be otherwise important for the interaction. Interestingly, this prominent CSI was not seen when comparing the unbound phosphorylated pYpY- Gab1[613-694] to the same reference dataset (data not shown). This implies that T667 might be involved in some interaction with the domain. Apart from these findings, pYpY-Gab1[613-694] stays disordered, indicated by its small CSI values.

For the interaction of the C-SH2[102-220] domain to phosphorylated ^{15}N, [13]C- pYpY- Gab1[613- 694], the CSPs are elevated at the region of aa 656-676, although the threshold of significant elevation states the region aa 657-671. The region can be divided into two parts: aa 657-665 and aa 666-671. Some elevated values can be seen for the regions around aa 620 and aa 690. The CSI shows disorder for the whole protein except for the stretch of aa 650-672. Here, fluctuating high positive and negative CSI values were calculated. A clear secondary structural tendency is not seen from this data, although it can be speculated that, again, two regions can be derived. One with more negative values from aa 661-665, indicating extended conformation and another part with more positive values from aa 669-673, hinting towards helicity. In the unbound state, positive values can be seen for residues 665-668, as described above (Gruber et al. 2022).

2.2.2 Dynamics of Gab upon binding
The labelled and doubly phosphorylated pYpY-Gab1[613-694] was also used for measurements of the relaxation rates, R_1 and R_2, and hNOE, which report on the dynamical behaviour of the backbone on a ps-ns time scale.
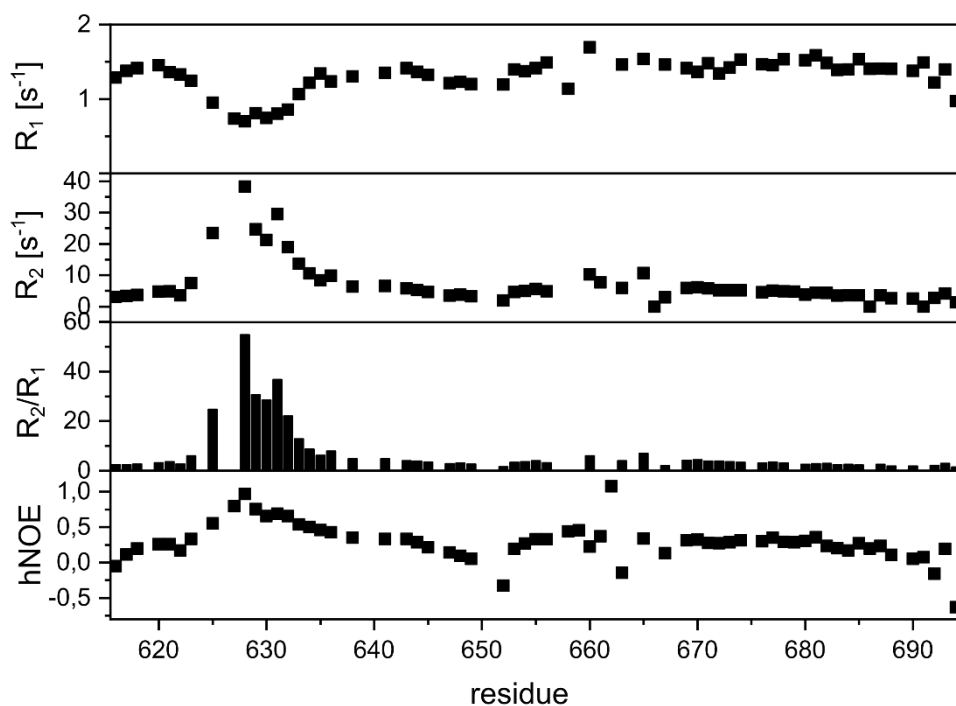
Figure 24 *Relaxation rates (R_1 and R_2) and hNOE from ^{15}N, ^1H pYpY-Gab1$^{613-694}$ in complex with the N-SH2$^{1-106}$ domain. Residual numbers of pYpY-Gab1$^{613-694}$ are shown on the x-axis. Created with Origin.*

Free pYpY-Gab1$^{613-694}$ protein (Gruber et al. 2022) shows low R_2/R_1 rates (around 3) throughout the whole sequence, indicating no persistent secondary structure or preferred conformation. hNOE values are reduced at the termini and for aa 640-660 (Gruber et al. 2022), indicating that those regions exhibit higher flexibility. Upon addition of the N- SH2$^{1-106}$ domain, the R_2/R_1 rates increase for the region of aa 623-636, suggesting a higher conformational restriction compared to the overall tumbling of the molecule (**Figure 24**). This is supported by the elevated hNOE values for the same region. The linker region of aa 640-660 shows the identical decay as for the free pYpY-Gab1$^{613-694}$, which again implicates disorder.
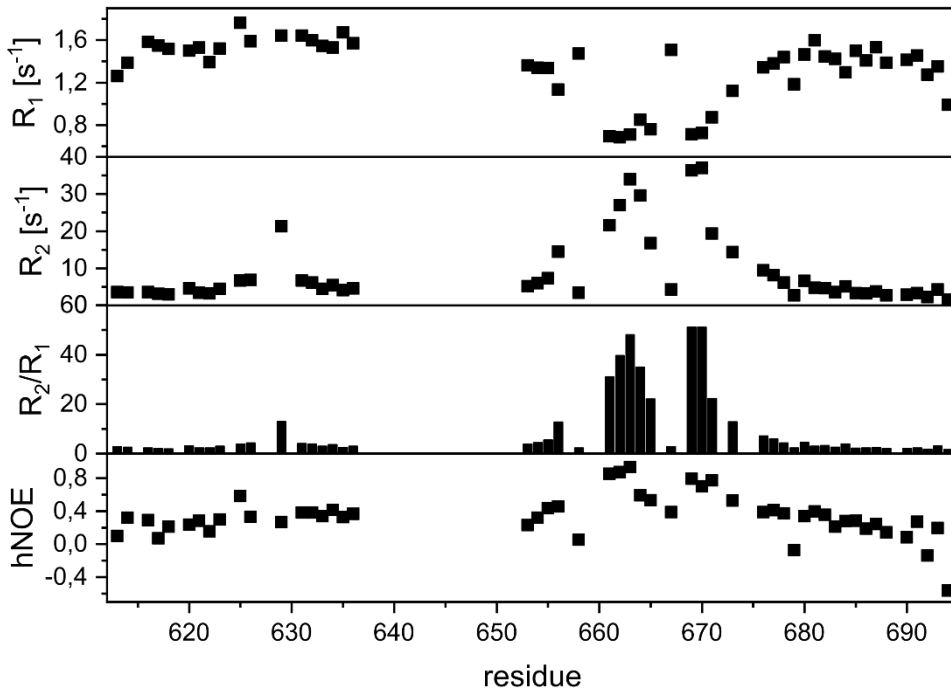
Figure 25 *Relaxation rates (R_1 and R_2) and hNOE from ^{15}N, ^1H pYpY-Gab1$^{613-694}$ in complex with the C-SH2$^{106-220}$ domain. Residues 637-653 could not be assigned, and therefore, no relaxation data could be quantified. Created with Origin.*

In complex with the C-SH2$^{102-220}$ domain, the R_2/R_1 values are increased from aa 656-680, indicative of a higher conformational restriction (**Figure 25**). Interestingly, a dip around aa 667 suggests that these ordered regions are separated locally. The remaining regions of pYpY-Gab1$^{613-694}$ are close to the ratios of the free state (Gruber et al. 2022). The hNOE values for this region increase with an identical dip as the R_2/R_1 rates. Additionally, the terminal residues experience a decrease in their hNOE values, while the region of aa 637-653 could not be assigned and thus has no R_2/R_1 and hNOE data.

2.3 Discussion

Far western blot analysis have shown that the Gab1-Shp2 interaction involves pY627 binding to N-SH2 and pY659 to C-SH2 (Cunnick et al. 2001). The here presented ITC and NMR results for the interaction between the N- and C-SH2 domain of Shp2 with pYpY-Gab1$^{613-694}$ confirm this hypothesis that the N-SH2 binds to residues around pY627 and the C-SH2 to residues around pY659. There are hardly any cooperativity effects between both binding events, leading to a mainly extended Gab1 conformation for residues around the pY binding sites with reduced local dynamics and an unexpected long binding site for the C-SH2 in particular. Using combinational peptide libraries (Sweeney et al. 2005), the SH2 domains of Shp2 exhibit overlapping target peptide preferences. While the N-SH2 domain favours (I/L/V/m)X**pY**(T/V/A)X(I/V/L/f), the C-SH2 prefers (T/V/I/y)X**pY**(A/s/t/v)X(I/v/l), where X can be any amino acid except glycine and proline, and lower case letters are less frequently observed, i.e both belong to class I specific motifs. Thus, both Gab1 phosphotyrosines could be capable of binding to either

site (pY627: V-E-pY-L-D-L and pY659: V-D-pY-V-V-V). In this project, ITC and NMR were used to investigate this further. ITC measurements of single domains to bis-phosphorylated pYpY-Gab1$^{613-694}$ could be fitted with a 1:1 stoichiometry. If each domain were to bind to both pY-motifs, a stoichiometry of 1:2 (N=0.5) would be expected. The chemical shift perturbation measurements (CSP) by NMR further supported specificity of each domain to their cognate pY residue. The N-SH2$^{1-106}$ domain binds the pY627, while the C-SH2$^{106-220}$ domain interacts with the pY659. Although NMR as well as ITC also hint toward a slight tendency for the N-SH2$^{1-106}$ domain to bind pY659. ITC data, which show slight titration in the thermograph, which cannot be fitted, might explain why, in NMR experiments with high protein concentrations, elevated CSPs for the pY659 binding region could also be seen, although this interaction could only be transient. At least from the binding preference (Sweeney et al. 2005), this would not be surprising, although these motifs only defy the range from -2 to +3 (relative to the phosphotyrosine), which does not resemble the native situation. On the other hand, the elevated CSPs might also originate from long-range secondary structure induction due to the binding of the N-SH2$^{1-106}$ domain at pY627. A hypothetical 2:1 (domain:peptide) binding of the N- SH2$^{1-106}$ to both phosphotyrosines could be possible in light of the previously described simultaneous interaction of N-SH2 to a class IV peptide (Sweeney et al. 2005). Here, two peptides are bound and form an antiparallel short β-sheet elongation of the domain (Zhang et al. 2011). The authors further showed that a double phosphorylated peptide of this class IV peptide interspaced by a glycine-linker showed elevated affinity. It could potentially be possible for Gab1 to bind in this way to the N-SH2, which would also give rise to a 1:1 stoichiometry in ITC. The experiment of the single SH2 domains with monophosphorylated Gab1 peptides also showed that while the C-SH2 domain does not react to pY627, the N-SH2 might very weakly interact with pY659. On the other hand, this 2:1 binding was only proven for class IV peptides.

The binding sites implicated by the high CSPs, are -4 to +6, for the N-SH2$^{1-106}$ and -3 to +17, for the C-SH2$^{106-220}$ domain. This is longer than that reported in literature: the SH2 domains of Shp2 were shown to accommodate IRS1 and PD-1 phosphopeptides with residues up to pY+5 (Marasco et al. 2021). CSP analysis of phosphopeptides is not commonly facilitated. Mostly, the SH2 domain is analysed upon binding by NMR. Unfortunately, that makes it hard to validate these results with others since chemical shifts are very sensitive to the environmental changes of a specific amino acid. A result measured by NMR might thus slightly vary from the output of other experimental techniques, like ITC. One study measured the NMR peak intensity of a Shp2 binding partner (KIR3DL1) upon interaction. Here, the binding can be seen up to pY+6 (Cheng et al. 2019). Thus, the interaction site for the N-SH2$^{1-106}$ domain on Gab1 is "typical",

whereas the C-SH2[106-220] binding site is much longer. In both cases, the region adjacent to the phosphotyrosine appears to be in an extended conformation, as observed for other SH2:phophopeptide interactions (Waksman et al. 1992). The extensive binding site C- terminally to pY659 is particularly intriguing, not having been seen for other SH2 domains. Based on the CSP and CSI calculations, this region can be divided into pY+6 (extended conformation) and the region pY+7 to +17 (helical conformation). This topic will be further discussed in chapter 4

The affinities measured for the binding of bis-phosphorylated pYpY-Gab1[613-694] to either the single SH2 domains or the tandem SH2 domains are within the range stated in the literature (nM - µM affinity). The IRS-1[1169-1178] peptide, with a K_D of 14 nM (Sugimoto et al. 1994) is one of the highest affinity binders to the N-SH2 domain of Shp2; affinities of around 100 nM have been determined for CagA-EPIYA-D (Hayashi et al. 2017) and PD- 1[244-254] (Marasco et al. 2021) and 30 nM for Gab2[608-620] (Visconti et al. 2020; Bonetti et al. 2018) (all measured at 150 mM NaCl). The interaction to the C-SH2 domain of Shp2 is less well studied; PD-1[219-227] binds with K_D=13 nM (Marasco et al. 2021) and IRS- 1[1219-1228] with K_D=110 nM (Sugimoto et al. 1994). Both studies further showed that both domains can bind to both investigated binding sites, although with a preference and a lesser affinity of at least one order of magnitude, resulting mostly in only µM tight interactions with the second domain. All these measurements were made with short peptides, in contrast to the present 83 aa bis-phosphorylated Gab1 ligand. Thus, a direct comparison to the data mentioned above is impossible and might explain the difference in binding strength, especially for the interaction of the N-SH2 to Gab2 (Visconti et al. 2020) and Gab1 (this thesis). Of course, a much greater contributing factor in the differential binding affinity between Gab1 and Gab2 is probably its sequence: although both sequences are identical at position: -2, +1, +3, +4, they vary in all other positions (Gab1/Gab2:   G/S-D/T-K/G-Q/S-V-E/D-pY-L-D/A-L-D-L/F-D/Q).   Unfortunately,   the thermodynamic parameters are not listed for the above-listed complexes. It might be that the entropy for binding a short or a longer ligand results in different binding strengths. The interaction of Gab1 with Shp2 is enthalpically driven. Thus, changes to the system by pH and salt concentration must be considered. For Gab2 (Bonetti et al. 2018), it was shown that an increase in salt leads to a decrease in the k_{on} rate while having constant k_{off} rates, demonstrating that the interaction is driven by the electrostatic attraction. This makes comparisons of affinities even harder. A study investigating the interaction of the tandem SH2 domains of Shp2 with short Gab1 mono-phosphorylated peptides revealed that the binding of the peptide containing pY627 (pY-Gab1[621-632]) had an affinity of 1.5 nM, while the peptide containing pY659 (pY-Gab1[653-664]) was measured to bind with a K_D of 27 nM (Koncz et al. 2001). These measurements, using peptides spanning the

interaction site from –6 to +5, diverge even more from the measurements presented in this thesis, using single domains with double phosphorylated pYpY-Gab1[613-694]. Additionally, the binding strength for both peptides varies, which is not the case for the here presented measurements. It can be assumed that additional interactions beyond the pY659+5 residue are found in the native complex. This is in line with the NMR results of this chapter; further information on that will be presented in chapter 4.

Nevertheless, the affinity reported here of around 3 µM for each domain is lower than for other phosphopeptides to those domains, in line with activation data for Shp2. Shp2 can be activated by binding to one high-affinity phosphopeptide-region alone, as is the case for IRS-1 (Sugimoto et al. 1994) and PD-1 (Patsoukis et al. 2020). In contrast, the activation by Gab1 requires both pY627 and pY659 binding to both domains (Cunnick et al. 2001). Inorganic phosphate and phosphotyrosine alone have affinities for the N-SH2 domain in the µM range (Marasco et al. 2021; Bonetti et al. 2018). Binding of pY is thought to be entropically driven due to the release of solvent molecules upon interaction (Ladbury et al. 1995) and contributes to around 50% of the total free energy (Bradshaw et al. 1999). As mentioned above, the interaction of Gab1 and Shp2 is enthalpically driven, with a moderately unfavourable entropic term. Some SH2:phosphopeptide interactions show slightly favourable entropic contributions (Bradshaw et al. 1999; McNemar et al. 1997), with minor effects of residues C-terminal to pY on the total free energy: in these cases, backbone and side chain contributions are negligible, and the reaction is driven by the pY-binding (Bradshaw et al. 1999). On the other hand, studies on the SH2 domains of p85 in complex with CD28 phosphopeptides showed highly unfavourable entropic terms that are compensated by strong negative enthalpies (Inaba et al. 2017). An elevated affinity for tandem SH2 domains is typically due to their proximal binding capacity and slower k_{off} rate (Ottinger et al. 1998; Goyette et al. 2022).

In the complex of Gab1 and Shp2, both energetic components are moderate. The unfavourable entropic term might stem from restriction in rotational freedom. For the interaction of the bis-phosphorylated pYpY-Gab1[613-694] to the Shp2[1-222] tandem SH2 domains, an increase in affinity of two orders of magnitude with respect to the single domains with a large negative enthalpy has been measured. The enthalpic and entropic contributions, which are not additive, suggest that the reaction to the tandem SH2 domains is highly complex and is not solely explained by the simultaneous binding and restriction of the flexible Gab1; see more on that in chapter 5. This situation is similar to IRS-1 binding to Shp2 (Ottinger et al. 1998), where exchange of the PEG-spacer (linker between the two pY) results in similar affinities for Shp2, indicating no specific interactions of the linker. The length of the linker is, however, important for binding affinity:

reducing the length of photoswitchable spacers in ITAM peptidomimetics results in a significant decrease in binding to tandem SH2 domains from Syk kinase (Kuil et al. 2009).

Chapter 3: Structural investigations of the interaction between N-SH2 and Gab1

The N-SH2 domain of Shp2 is the subject of extensive research interest as it is associated not only with the binding to phosphorylated peptides but especially with the activation of the phosphatase and, thus, its implications in cell regulation and pathology. The N-SH2 domain blocks the active site of the PTP domain with its DE-loop as illustrated in **Figure 26**, so that it must be displaced for activation.
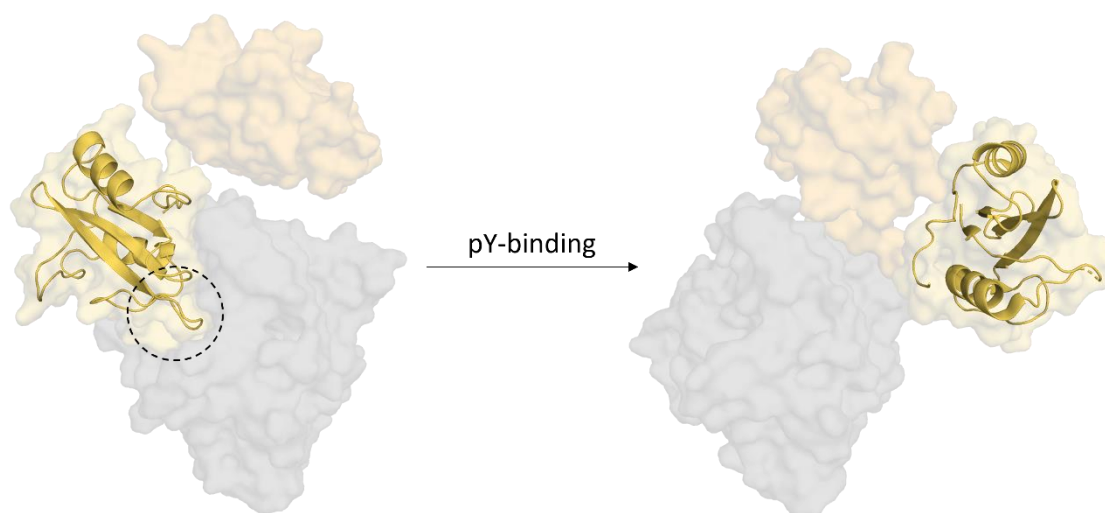
Figure 26 *Activation of the phosphatase Shp2. It is associated with the release of the N-SH2 domain (yellow, cartoon and surface) from the PTP domain (grey, surface). The DE-loop (marked with a black circle) of the N-SH2 domain blocks the active site of the phosphatase domain in a closed conformation (left; PDB:4dgp; (Yu et al. 2013). In the constitutively active Shp2 mutant E76K (LaRochelle et al. 2018), the N-SH2 domain relocates to adopt the open conformation. The C-SH2 domain rotates slightly in the process of activation. Illustrated with PyMol.*

3.1 Crystallisation and structural analysis of N- SH2$^{1- 106}$:pY- Gab1$^{613- 651}$

To analyse the N-SH2:Gab1 complex structure, protein crystallisation was facilitated (together with M.Sc. Judith Kniest, supervised master's student). The N-SH2$^{1-106}$ domain was used in initial screens in complex with a synthesized mono-phosphorylated Gab1 peptide (pY-Gab1$^{613-651}$). An unpublished crystal structure of Gab1 complexed by N-SH2 deposited in the PDB (PDB:4qsy, (Gogl and Remenyi, unpublished)) used the same N- SH2 construct, N-SH2$^{1-106}$, in complex with a much shorter phosphorylated pY- Gab1$^{621- 633}$. By using a longer Gab1-construct, a more native setting can be produced, and potential additional interactions might be investigateable. Data collection of a single crystal, which grew in a tetragonal form, revealed a space group of P4$_3$2$_1$2 with one molecule in the asymmetric unit and was solved to 2.1 Å by molecular replacement with the N-SH2 domain of the 4qsy-PDB structure (Gogl and Remenyi, unpublished) (in cooperation with Dr. rer. nat. Constanze Breithaupt-Than and Dr. rer. nat. Christoph Parthier, Research Group of Prof. Stubbs).

Table 6 *Crystallographic data from the N-SH2$^{1-106}$: pY-Gab1$^{613-651}$ complex*

	N- SH2$^{1-106}$: pY-Gab1$^{613-651}$	N- SH2$^{1-106}$: pY-Gab1$^{621-633}$ (PDB:4qsy)
Space group	P4$_3$2$_1$2	P4$_3$2$_1$2
Unit cell	a = b= 60.78 Å c = 69.85 Å α = β = ɣ = 90°	a = b = 61.78 Å c = 74.19Å α = β = ɣ = 90°
Resolution (Å)	27.87 – 2.08	43.69 – 2.10
Number of reflections	118443	8894
Number of unique refelctions	8364	
CC1/2	99.8	
Rmeas (%)	13.7	
I/Iσ	14.0	2.77
Completeness (%)	99.7	99.7
Wilson B-factor (Å^2)	31.7	33.9
Rfree (%)	27.2	22.9
R (%)	22.5	20.0
Ramachandran (%) (favoured, allowed, outliers)	97.0, 2.9, 0.0	97.0; 3.0; 0.0

The N-SH2 domain consists of two α-helices flanking a central β-sheet, composed of three anti-parallel β-strands connected by loops. N-SH2 residues N-terminally to 6 and C-terminally to 104 are not visible in electron density. The Gab1-peptide is bound orthogonal to the β-sheet in an extended conformation. Similarly, Gab1 residues N- terminal to 625 and C- terminal to 632 are not seen in the electron density and are presumably flexible. Thus, only eight amino acids from -2 to +5 of the Gab1 peptide could be modelled; the remaining amino acids of the 38 aa peptide could potentially occupy the free space between the symmetry-related molecules (supplemental **Figure 52**).

The Gab1 peptide binds to the 'front' of the domain (for clarification of the nomenclature used to describe the faces of the domain, see supplemental **Figure 64**), i.e. its structure is not influenced by crystal contacts (supplemental **Figure 51 A**). It adopts an extended conformation with stabilising backbone H-bonds between the peptide and domain. The amide group of Gab1-L628 (+1) H-bonds the carbonyl oxide of H53; the carbonyl group

of D629 (+2) forms a stabilising H-bond with the amide of K91, and the amide of D631 (+4) forms one with the carbonyl group of K89.

The phosphotyrosine pY627 is coordinated by H-bonds from the hydroxyl-groups of S34, S36 and T42 as well as through the backbone of K35 and S36. R32 provides a partner for an ionic bond. The side chain of K55 stabilises the phenyl ring through cation-pi interaction at a 3.8 Å distance, as seen in **Figure 27**.
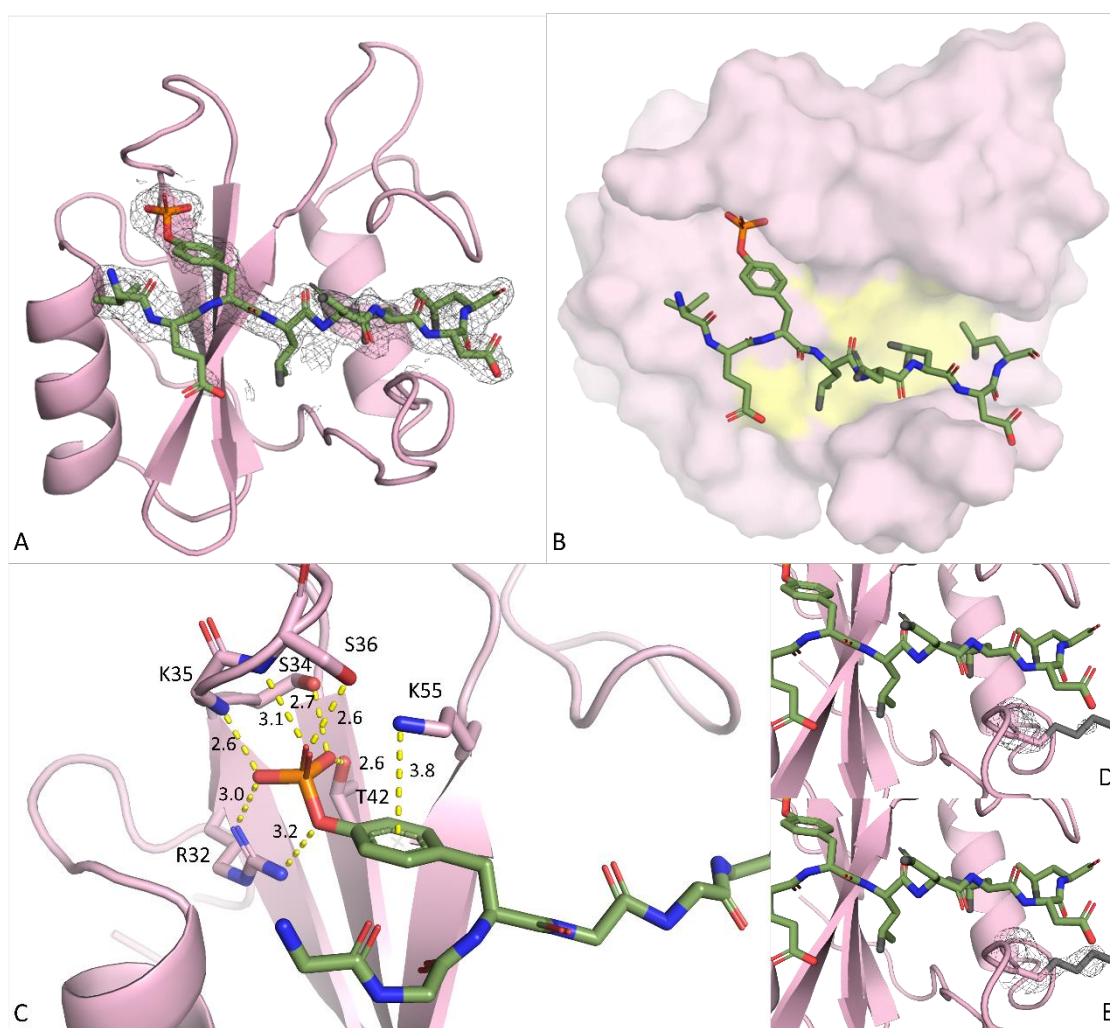
Figure 27 Interaction of the N-SH2$^{1-106}$ domain (pink) with the Gab1 fragment (green, stick) from aa 625-632 (position -2 to +5). A) Cartoon representation of the N-SH2 domain in complex with the Gab1 fragment. The density at $\sigma_{2Fo-Fc}=1$ is shown for the peptide as a grey mesh. B) The N-SH2 domain is shown as a surface model with the hydrophobic patch coloured in yellow. Side chains of the +1, +3 and +5 positions of Gab1 face towards this patch. C) Interactions of pY627 to the domain. Amino acids involved (sticks) are labelled as are their distances to atoms of the pY residue (Å). D) The side chain D631 is in proximity to the side chain K89 of the domain (stick, coloured in grey, as these atoms have an occupancy of zero). Shown is the density at $\sigma_{2Fo-Fc}=1$ for this residue as mesh. E) same position as in D) but with $\sigma_{2Fo-Fc}=0.5$ shown as mesh. Illustrated with PyMol.

Although amino acids C-terminal to pY627 are expected to contribute to specificity, side chain interactions can only be assumed, as the leucines and important interaction partners could be not fully modelled (**Figure 27**). The hydrophobic leucines at positions +1, +3 and +5 occupy a hydrophobic patch formed by T52, I54, L65, Y81, L88 and I96

(coloured yellow in the figure) and are thus protected from interaction with the solvent Residue +1 (L628) faces T52, I54 and I96 of the hydrophobic patch, while the +3 position (L630) is in neighbourhood of I54, L65, L88 and I96 and the +5 position (L632) is surrounded by L65, Y81 and L88. The amino acid at +4 (D631) faces towards the side chain of K89, although density is only seen at a contour level of $\sigma_{2Fo-Fc} = 0.5$, where it would be within the distance to form an H-bond.

3.2 Crystallisation and structural investigation of the tandem SH2:pYpY-Gab1$^{617-684}$ complex by electron diffraction

A second crystal structure was solved of the tandem SH2 domain Shp2$^{1-222}$ in complex with bis-phosphorylated pYpY-Gab1$^{617-684}$ in cooperation with Fazard Hamdi, PhD (Research Group of Jun.-Prof. Kastritis), Dr. rer. nat. Alaa Shaikhqasem and Dr. rer. nat. Constanze Breithaupt (Research Group of Prof. Stubbs). Crystals of the pYpY- Gab1:tandem SH2 domain complex grew within one week in two conditions of an initial screen. The needle-shaped crystals could be reproduced in in a fine screen using a larger volume. These crystals failed to yield X-ray diffraction patterns suitable for structure solution, and optimisation through seeding and additive screens was not successful, either.

Data could, however, be collected using electron diffraction (microED). The crystals were applied to a grid used for standard electron microscopy, back-blotted and vitrified. The grid was transferred to a cryo-electron microscope, and data were collected from two crystals perpendicular to one another. A data set of 89% completeness could be collected and further processed by molecular replacement. The structure was solved to 3.2Å (for further information on crystallisation, data collection and processing, see Material and Methods).

The final model consists of SH2 residues 4-221 (with density missing for residues N- terminal to R4, and the C-terminal residue N222 and between T153 and K164 (CD-loop in C-SH2)) and bis-phosphorylated-Gab1 (pY627 and pY659) residues D622-D631 and A653-S672, amino acids 617-621, 632-652 (the pY-linker) and 673-684 could not be resolved. Mass spectrometry analyses (Edman degradation) showed that at least aa 621- 684 were present in the crystal (supplemental **Figure 55**). In this chapter, the focus will be on the N-SH2:Gab1 interaction; further information on the microED structure will be presented in the following chapters 4 and 5.

Table 7 Crystallographic table for the Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$ complex

Data collection	
Excitation voltage	200 kV
Wavelength (Å)	0.025
Total dose per crystal (e$^-$/Å2)	4.8
No of crystals used	2
Total angular range collected (°)	138
Merging statistics	
Space group	P2$_1$2$_1$2$_1$
Unit cell dimensions	
a,b,c (Å)	30.51, 82.15, 118.16
α=β=ɣ (°)	90
Resolution (Å)	33.7 – 3.2 (3.5 – 3.2)
Total reflections	56119 (13409)
Total unique reflections	4747 (1091)
Completeness	89.3 (90)
I/σ(I)	3.29 (1.47)
CC$_{1/2}$	85 (56)
R-meas (%)	80.2 (207)
Model Refinement	
R$_{work}$ / R$_{free}$	33.6 / 39.0
RMSD bonds (Å)	0.0019
RMSD angles (°)	1.27
Ramachandran (%) **(favoured, allowed, outlier)**	95.07, 4.48, 0.45

An eight amino acid-long Gab1-peptide is bound to the N-SH2 domain, ranging from -5 to +4. It is in extended conformation, as seen in supplemental **Figure 51,B**. This is a rather long peptide fragment bound to the N-SH2 domain, as normally only regions from -2 to +5 are modelled to be domain-bound (Anselmi et al. 2020).

The phosphotyrosine of the Gab1-peptide gets stabilised by the hydroxyl-groups of S34 an T42, the carboxyl group from the backbone of K35. The side chain of S36 could not be modelled, in contrast to the structure of N-SH2$^{1-106}$:pY-Gab1$^{613-651}$. R32 coordinates the negative phosphate through an ionic bond. The cation-pi interaction cannot be seen since the side chain of K55 is not fully built (**Figure 28**).
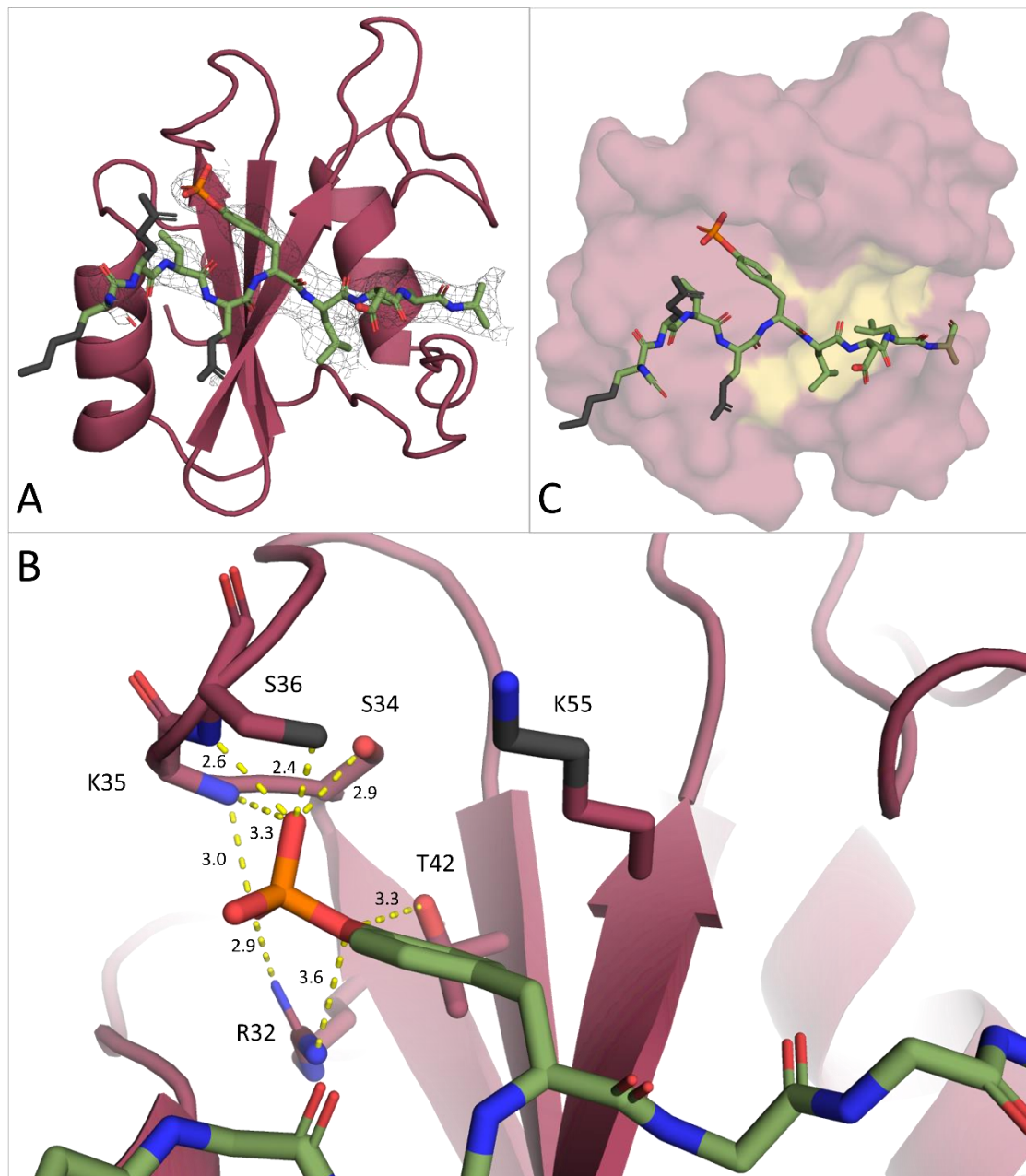
Figure 28 *Interaction of the N-SH2 domain (raspberry) to the Gab1 fragment (green, stick) from the Shp2[1-222]:pYpY-Gab1[617-684] structure solved by microED. A) The N-SH2 domain is shown as cartoon with the bound stick model of the Gab1 fragment covered in density at σ_{2Fo-Fc}=1, shown as mesh. Atoms with an occupancy of zero are coloured in grey. B) The interaction of pY627 to the residues of the domain (stick) is shown. Amino acids involved are labelled, and interactions are indicated with yellow dashed lines stating the distance in angstrom. C) The domain is shown as a surface model bound to the peptide. The hydrophobic patch is indicated in yellow. Residues +1 and +3 are pointed towards that surface. Illustrated with PyMol.*

As with the crystal structure solved by X-ray of the N-SH2[1-106]:pY-Gab1[613-651], the side chains of the residues C-terminal to the phosphotyrosine are not seen in direct interaction to the domain, except for two backbone-backbone interactions between L628 (+1) and H53; D629 (+2) and K99 as well as A631 (+4) and K89. These interactions seem not to provide specificity. The residues modelled for positions +1 and +3 face the hydrophobic surface of the domain (**Figure 28**).

The N-SH2:Gab1 structures observed in the single and tandem domain crystals exhibit the same overall conformation of the flexible EF- and BG-loops, described to be essential for specificity. They align structurally also with other N-SH2:phosphopeptides complexes from literature (PDB: 1aya, 1ayb, 1ayc, 6roz, 6roy, 5x94 (Hayashi et al. 2017), 5x7b (Hayashi et al. 2017)), including the single N- SH2 in its free state (PDB: 1ayd) (Lee et al. 1994) (**Figure 29, A**).
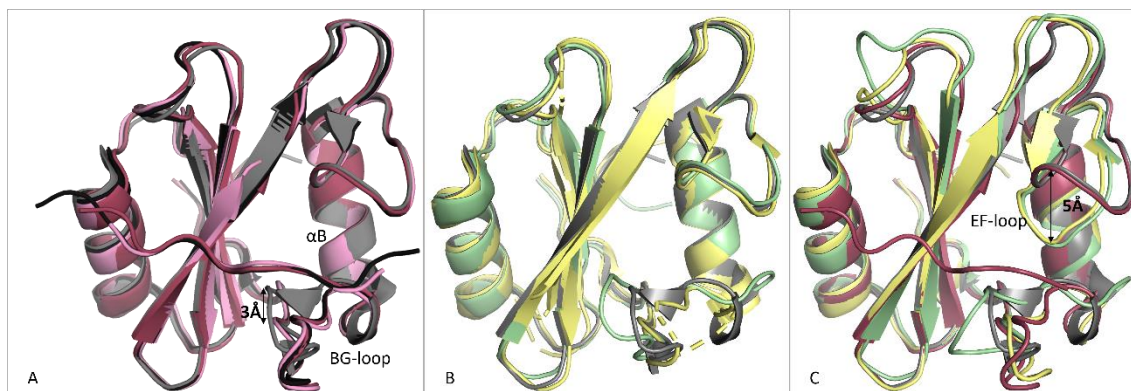
Figure 29 Superposition of N-SH2 domains of different complexes. A) The apo N-SH2 domain (PDB: 1ayd, grey) and the solved models of N-SH2:Gab1 from this thesis (N-SH2$^{1-106}$:pY-Gab1$^{613-651}$ complex in pink and Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$ complex in raspberry) indicate that apo and bound states are nearly identical. B) The bound N-SH2:Gab1 model (raspberry) derived from the Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$ crystal to the structures of the N-SH2 domain of the open Shp1 (green, PDB: 3ps5 (Wang et al. 2011)) and Shp2 (yellow, PDB: 6crf (LaRochelle et al. 2018); both models of the two molecules in the asymmetric unit are shown, as their BG- loop varies) structures and the apo- structure (grey) are presented. C) The Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$ complex (raspberry), with the N-SH2 domains taken from the closed Shp1 (green) and Shp2 (yellow) structures (PDB: 2b3o (Yang et al. 2003), 4dgp (Yu et al. 2013)) and the apo domain (grey) are depicted. Note the different conformations of the EF-loop in the closed structures (PTP-bound, inactive state). All models are shown in cartoon presentation. Illustrated with PyMol.

The overall conformation of apo N-SH2 and the N-SH2 bound to Gab1 are highly similar, as seen in **Figure 29 A**, except for the orientation of the BG-loop immediately following the αB helix, which differs between the apo structure and the peptide bound state. The C- terminal part of the αB helix elongates into the BG-loop. The BG-loop moves around 3 Å to accommodate residues L628 (+1) and L629 (+3) and is not involved in any crystal contacts.

In the structures of the open Shp1 and Shp2 conformation (PDB: 3ps5 (Wang et al. 2011), 6crf (LaRochelle et al. 2018)), the N-SH2 exhibits the same conformation as does the single apo domain, except the BG-loop is not completely modelled. Both are thus also similar to the bound N-SH2 domain (**Figure 29, B**)

Interestingly, in the closed conformation of Shp2 (PDB: 4dgp (Yu et al. 2013)), where the DE-loop (N58-Y63) is buried within the PTP domain, the above-described variant BG-loop is similar to the N-SH2:Gab1 structures, whereas the EF-loop is displaced by around 5 Å (**Figure 29, C**). The same conformation of the EF-loop is found in the closed Shp1 structure (PDB: 2b3o (Yang et al. 2003)), while the BG-loop is similar to that in the apo

domain (PDB: 1ayd), leading to a wider open binding site – which nevertheless is not as open as in the Gab1-bound structures. The N-SH2 of the closed Shp1 (PDB: 2b3o (Yang et al. 2003)) is the only structure where the BC-loop (responsible for pY-binding) is positioned upwards while in all other structures (apo, bound, closed, open) it is folded downwards. Those differences were shown to not be associated with physiological aspects but rather with crystallographic artefacts (Anselmi and Hub 2021).

Nevertheless, both structures (PDB: 4dgp (Yu et al. 2013), 2b3o (Yang et al. 2003)) in the closed conformation (where the DE-loop is buried in the PTP domain) show a tightly packed central β-sheet (measured between the carbonyl C of G39 and the amide N of N58). This conformational rearrangement was shown to be associated with the activation of Shp2 – the peptide binds, leading to a spread of the β-sheet and thus the release from the PTP-domain (Anselmi and Hub 2021; Marasco et al. 2021). This opening of the β-sheet can also be seen in the Gab1-bound structures of the N-SH2.

3.3 Analysing the complex of N-SH2:Gab1 via NMR

In contrast to a model derived from crystallography, which can be affected artificially by crystal contacts, NMR can provide insights into the system in solution. ^{15}N, ^{13}C labelled N-SH2$^{1-106}$ was measured in the free and bis-phosphorylated pYpY-Gab1$^{613-694}$ bound state and assigned using tripple-NMR experiments. The free state was assigned to 89 % (91 aa/ 102 aa (without prolines)) and the bound state to 85 % (87 aa/102 aa (without prolines)).

3.3.1 CSP

Chemical shift perturbations (CSP) of the N-SH2 domain were calculated upon peptide binding. The tip of the αA-helix (especially G13), the BC-loop (K35 – F41), βC (especially F41 and L43) and βD (T52 - I56), and the DE (N58 - D61) and EF (Y66 – G68) loops show high CSP values, indicating that these amino acids are affected by the binding of doubly phosphorylated pYpY-Gab1$^{613-694}$. The CSPs are plotted on the structure in **Figure 30**; a per-residue presentation of these CSP values and the 2D-TROSY can be found in the supplemental part (**Figure 57** and **Figure 58**).
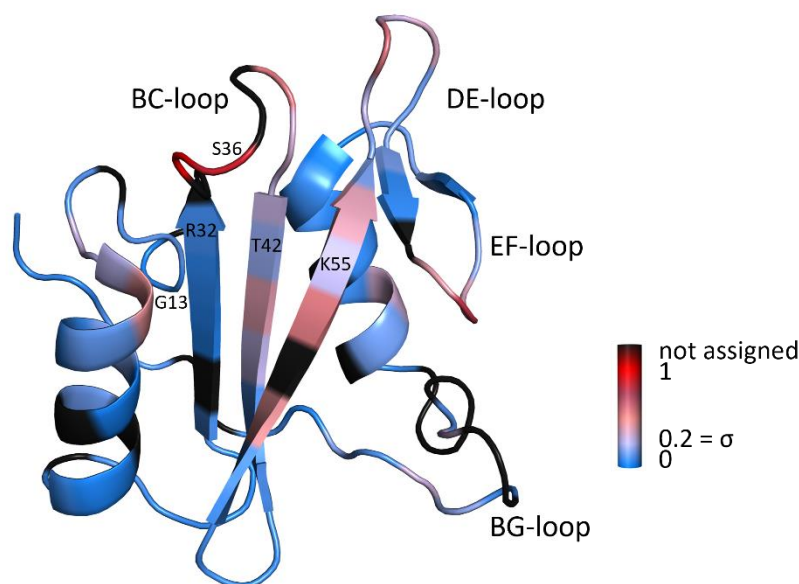
Figure 30 *CSP values derived from the differences measured from the free and pYpY-Gab1^{613-694} bound N-SH2 domain plotted on the N-SH2 domain structure. Small CSPs [in ppm] in blue indicate that those regions are mostly unaffected by the interaction. Red indicates CSPs of 1 ppm, which were the largest CSPs measured for this system. These regions, especially the BC-loop, which is involved in pY-coordination, are mainly affected by the interaction. The colour bar to the right indicates the CSP intensity relative to the colour. Black residues could not be assigned or were not able to be analysed. The threshold of σ is indicated. Illustrated with PyMol.*

3.3.2 Dynamic

NMR can also be used to obtain dynamical information on a system. Measurements of the relaxation rates R_1 and R_2, as well as the hNOEs of amide protons at different magnetic field strengths, are sensitive to the motion of the molecule on the ps-ns time-scale, from which the order parameter S^2 can be derived using the Lipari and Szabo model-free analysis. High values of the order parameter, which ranges between 0 and 1, indicate stable and rigid conformations, while small order parameters indicate higher mobility.

An unexpected but reproducible overall dependency of the R_2 rates on the magnetic field was noticed the N-SH2 domain in complex to pY-Gab1^{613-651} (supplemental **Figure 59**), indicating large exchange terms between at least two conformational states. Whereas lowering the protein concentration had no effect, an increase in ionic strength to 250 mM NaCl uncoupled the dependency. Thus, the S^2 values derived for the Gab1-bound state, calculated from the original low salt measurement to make them comparable to the free state, shown in **Figure 31** must be taken with caution.
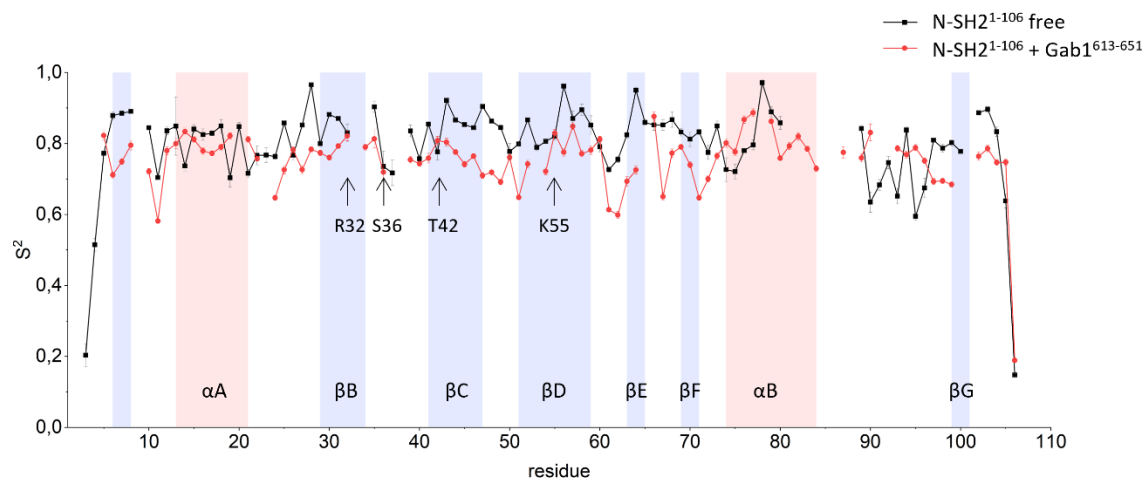
Figure 31 *Order parameter S^2 plotted against the N-SH2$^{1-106}$ domain residue number for the free complex (black), and the pY-Gab1$^{613-651}$ bound state. The secondary structure elements of the domain are indicated in blue (strand) and red (helix). Illustrated with Origin.*

Overall, the N-SH2 domain can be described as a folded entity. The backbone motion in secondary structure elements shows slightly higher rigidity than the connecting loops. The free SH2 domain has an average S^2 of 0.79, which decreases to 0.75 in the Gab1-bound state. Thus, the bound state exhibits slightly more global flexibility, particularly for the central β-sheet and the EF-loop, whilst the helices are not much affected by binding. On the other hand, the BG-loop becomes more rigid upon binding. This loop and the amino acids responsible for binding the phosphotyrosine (e.g. T42 (S^2: 0.77), whose hydroxyl group acts as an H-donor to the phosphate) show higher mobility than their neighbours (F41 S^2: 0.85 L43 S^2: 0.92) in the free N-SH2 domain. Similarly, R32, S36 and K55 show identical low-order parameters in the bound as in the free state, compared to their neighbouring amino acids, indicating that residues designated to bind to the phosphotyrosine show higher motion to better adopt to the binding partner. How the domain behaves in response to binding the C-terminal residues of Gab1 cannot be derived from these data since essential amino acids, e.g., H53 and K91 (interaction to the peptide's backbone), L65, Y81 and L88 (hydrophobic patch), could not be assigned and have thus no S^2 value. Interestingly, residues D61 and Y62 of the DE-loop show high backbone dynamics in both the bound and unbound states. Although these are not involved in Gab1 binding, they are found in the N-SH2:PTP interface in Shp2, where they stabilise the inactivate conformation.

3.4 Discussion

Immense research has been conducted on the N-SH2 domain of Shp2. This is due to its two crucial regulatory mechanisms: the binding of the activating peptide and the stabilisation of the closed conformation. Shp2's implication in diseases was thus a great accelerator to study this domain.

Binding of residues N-terminal to pY627

The SH2 domains of Shp2 and its homolog Shp1 share 64% sequence identity (**Figure 12**), whereas the specificity providing BG-loop between both homologues shares only 29% sequence identity. This is probably the reason for their selectivity in binding partners. Other than that, both N- and C-SH2 of both homologues have a glycine residue at the tip of the αA helix (αA2 position), whereas it is mostly an arginine in other SH2 domains. This allows the domain to bind peptides with a hydrophobic amino acid at the -2 position, which can be buried inside the domain compared to the normally present arginine side chain. Thus, the glycine provides a selective feature while losing the ion pair-providing arginine for stabilising the phosphate. The -1 position of Gab1, a glutamic acid, faces the solvent in all crystallographic structures (of this thesis, as well as the 4qsy-PDB structure (Gogl and Remenyi, unpublished)). Nevertheless, in MD simulations, this side chain was able to interact electrostatically with K35 from the pY-binding loop (BC-loop). This side chain is normally seen in a solvent-exposed position (as in this thesis), while the backbone is involved in the phosphate coordination. In those simulations, the K35 side chain was turned towards the phosphotyrosine and was also involved in its coordination (Anselmi et al. 2020). The microED structure of the tandem SH2 domain in complex with the bis-phosphorylated Gab1 showes the N-terminal part of the peptide (positions -5 to -3) facing away from the domain.

Phosphate binding

The binding of the phosphate group is a fundamental step in the activation of Shp2 binding of phosphotyrosine alone is capable of inducing chemical shift perturbations (CSP) in distal parts of the N-SH2 domain (Marasco et al. 2021). The authors described an H-bond network from the pY-binding region to the EF- and BG-loop, which provide access to the binding sites for C-terminal amino acids and, thus, full binding and activation by distributing the signal of "binding" (opening of the binding site) to those loops. The phosphate in the N-SH2 domain of Shp2 is coordinated by H-bonds (S34, S36, T42, backbone of K35) and a salt bridge by R32 (position βB5), as well as a cation- pi interaction between the phenyl ring and K55 (or an ionic interaction to the phosphate).

The arginine in the βB5 strand (R32), part of the FLVR motif, is highly conserved and present in nearly every SH2 domain (Liu et al. 2012), its mutation leads to a dramatic decrease in affinity (Bradshaw et al. 1999), and it was determined for selective binding of phosphotyrosine over phospho-serine or threonine (Kuriyan and Cowburn 1997; Mayer and Gupta 1998).

A second positively charged amino acid is K55 (βD6 position) in the N-SH2. In the previously deposited Gab1:N-SH2 structure (PDB: 4qsy (Gogl and Remenyi, unpublished)) as well as in other structures, the lysine side chain is in proximity to building a salt bridge to the phosphate (Lee et al. 1994). The N-SH2$^{1-106}$:pY-Gab1$^{613-651}$ structure, nevertheless, shows it to be involved in a cation-pi interaction (3.8 Å). This interaction has also been shown, although not discussed, in a complex of the N-SH2 with PD-1 (PDB: 6roy) (Marasco et al. 2020a) (4.7 Å distance) and in the interaction to PDGFR (PDB: 1aya, (Lee et al. 1994)) (4.1 Å distance) both at similar distances as in this thesis. For both structures, one molecule of the asymmetric unit showed cation-pi interaction, while in the other molecule, K55 was poorly modelled or in a distal position to the negative charge of the phosphate, indicating that this amino acid is not necessarily involved in distinctly stabilising the phosphate. In the tandem SH2 domain structure in complex with pYpY-Gab1$^{617-684}$, K55 was also not modelled, indicating that together with the rotamer in the 4qsy-PDB structure (Gogl and Remenyi, unpublished), K55 is not necessarily involved in the complexation of the phosphotyrosine, neither through a salt bridge nor through cation-pi interaction.

Interestingly, binding of a pY-peptide to Shp2 mutant T42A leads to a hyperactivation of the enzyme. A loss of the hydroxyl-group would intuitively be expected to lead to a decrease in affinity, but the opposite is the case, with an increase of two orders of magnitude from 100 nM to 1 nM (Toto et al. 2020), through a reduction in the dissociation rate. Interestingly, T42S does not restore the WT binding and results in an affinity of 5 nM. These findings suggest that the mutant complex has a longer half-life, leading to prolonged activation by minimal activator concentration. The authors postulate that T42 might shield the negative charge of the phosphate from K55 (βD6 position), which could potentially interact with the phosphate (as discussed above). Alternatively, the T42A and T42S mutations might destabilise the β-sheet (Anselmi and Hub 2020). Molecular dynamics (MD) simulations suggested that opening and spreading of the β-sheet upon phosphate binding is responsible for Shp2 activation and that destabilisation of the β- sheet might result in higher activation.

The long aliphatic Gab1 L628 side chain at position +1 occupies a hydrophobic patch formed by T52, I54, I96 and adjacent residue Gab1 L630 (position +3), which would be only partially filled by shorter side chains like alanine in PD-1 ((Marasco et al. 2020a) PDB: 6roz). The hydrophobic patch extends to residues L65, Y81 and L88 (Marasco et al. 2021), which shields hydrophobic amino acids of the peptide, particularly in positions +1, +3 and +5. In other SH2 domains, three (L65, Y81, L88) out of these six amino acids are conserved, which are mostly responsible for providing a hydrophobic environment

for the +3 residue, making this larger hydrophobic area a specific feature of Shp2 and Shp1, where besides +3 also the +1 and +5 positions are required to be hydrophobic (Diop et al. 2022; Marasco et al. 2021) and **Table 1**. The leucine L632 at the +5 position is facing the hydrophobic surface of the domain. Activators of Shp1 and Shp2 nearly always have a hydrophobic (mostly aromatic) side chain at this position (Marasco et al. 2021). This seems to be a critical selector, as CagA binding motifs vary in affinity to the N-SH2 domain by two orders of magnitude, depending on having a phenylalanine or an aspartic acid residue at this position (Hayashi et al. 2017).

The aspartic acid D629 at position +2 faces the solvent, based on MD simulations, this residue was suggested to interact with the domains K91 side chain, which was undefined in all N-SH2:Gab1 structures (within this thesis and PDB:4qsy (Gogl and Remenyi, unpublished)) (Anselmi et al. 2020).

Similarly, the aspartic acid D631 at position +4 is oriented away from the surface of the domain. Again, MD simulations of the N-SH2:Gab1 and N-SH2:IRS1 suggested interaction with K89 (Anselmi et al. 2020), but this side chain is also undefined. In the 4qsy-PDB structure (Gogl and Remenyi, unpublished), the +6 residue is modelled, although the orientation of the side chain and backbone is not clear. Despite using longer constructs of the Gab1-peptide (aa 613-651 and 617-684), residues N-terminal to -3/ -5 and C-terminal to +4/+5 are not defined, suggesting flexibility. This is in agreement with MD simulations, where the root mean square fluctuations (RMSF) of the bound peptide have indicated that only residues up to +4 or +5 (Gab1) are stably embedded in the domain (Anselmi et al. 2020). In general, the Gab1-peptide is in contact with amino acids in the αA, βB, BC-loop, βD, BG- loop and αB helix.

Compared to the Gab2 sequence (Gab1/Gab2: G/S-D/T-K/G-Q/S-V-E/D-pY-L-D/A-L-D-L/F-D/Q) on a structural basis, a similar interaction between both homologues to the N-SH2 domain could be expected (at least for structurally solved residues -4 to +6 (PDB: 4qsy (Gogl and Remenyi, unpublished)). Except for the Gab2-glycine (-4), which presumably might be better situated facing V14 (in contrast to Gab1-K623), the other differences between both sequences should not result in a large variance between the affinities (nM for Gab2 and µM for Gab1 – as discussed in Chapter 2).

3.4.2 Dynamics

MD simulations suggest that the complex is more flexible than seen in crystal structures (Anselmi and Hub 2021), with important implications for the activation mechanism of Shp2. The simulations suggest that the+5 binding pocket closed in the crystal structure of autoinhibited Shp2 (Hof et al. 1998), remains accessible in the inactive state. Since

crystal structures show solid models, more native insights can be derived from NMR experiments; therefore the N-SH2:Gab1 complex was analysed using CSPs and dynamics.

The chemical shift perturbations (CSP) of the N-SH2 domain in complex with pY-Gab1[613-651] are elevated in those regions (αA, βB, BC-loop, βD, BG-loop and αB), which were involved in binding as seen from the crystal structures. Nevertheless, additional and absent interactions might be concluded from NMR. R32, essential for binding the phosphate, its amide backbone is seemingly unaffected by the solution binding for the Gab1:N-SH2 (this thesis) and the N-SH2:PD-1 complex (Marasco et al. 2020b; Marasco et al. 2021) as well as for other SH2 domain complexes, (Feng et al. 2018). The order parameter of R32 upon binding is also not changing, independent of binding Gab1 (this thesis) or PD-1 (Marasco et al. 2021). Nevertheless, it is interesting to note that in the dynamic measurements, the residues destined to bind the phosphotyrosine (R32, S36, T42, K55) exhibit greater flexibility already in the free state compared to their neighbours.

According to CSP data, both DE- and EF-loops are also involved in binding. Consistently, these loops gain flexibility upon binding, as seen in the order parameters. In the autoinhibited state, these loops mediate the interaction between the N-SH2 and PTP domains so that increased flexibility in these regions might contribute to the release of the N-SH2 domain to activate Shp2. Although the overall domain becomes more mobile upon binding of Gab1, the helices as well as the BG-loop, are stabilised upon binding. The BG-loop might rigidify because of the potential side chain interactions of the +2 and +4 residue seen by MD (Anselmi et al. 2020), as discussed above, and the verified backbone interactions between those residues. In contrast to the PD-1 bound state, which does not change in mobility upon peptide binding, the domain in complex with Gab1 gains some additional flexibility. The surprising (yet reproducible) dependency of the R_2 rates on the magnetic field upon Gab1 binding suggests a large exchange term between at least two states, exchanging on a millisecond time scale, which could affect the order parameters derived from these data. Marasco et al. 2021 did not describe a similar effect resulting from the PD-1 binding, where they obtained an average S^2 value of 0.86 for the apo-N-SH2 compared to S^2 of 0.79 reported in this thesis. This discrepancy might derive from the lower pH used in this thesis (pH 6.0 compared to pH 6.8), destabilising the domain. Whether the deviation might result from the N-SH2:Gab1 system or from the experimental set up must be further investigated. Nevertheless, as will be presented in the next chapter, it is unique to the N-SH2 domain under the same experimental conditions.

Chapter 4: Structural investigations of C-SH2 in complex with Gab1 reveal the ordering of the bound Gab1

As seen from the NMR experiments presented and discussed in chapter 2, the Gab1- pY659 binding sequence for the C-SH2 domain was longer than expected, spanning 20 amino acids from aa 657-676, with a clear division into two parts: aa 657-665 (i) and aa 666-676 (ii). CSI calculations suggested that the first stretch (i) is probably extended while the second region (ii) could be helical when bound. Such long binding regions are not known for SH2 domains so far so different approaches were taken to study this interaction.

4.1 Binding of C-SH2 to pY659-Gab1 peptides

The interaction between the C-SH2$^{102-220}$ domain and three (phospho)-Gab1-peptides of different lengths was studied using ITC (**Figure 32**). The pY-Gab1$^{655-677}$ peptide spans the whole binding site (i+ii), pY-Gab1$^{655-666}$ covers the first region (i), and Gab1$^{667-677}$ covers the second region (ii). This peptide is not phosphorylated as it does not have Y659 in its sequence.
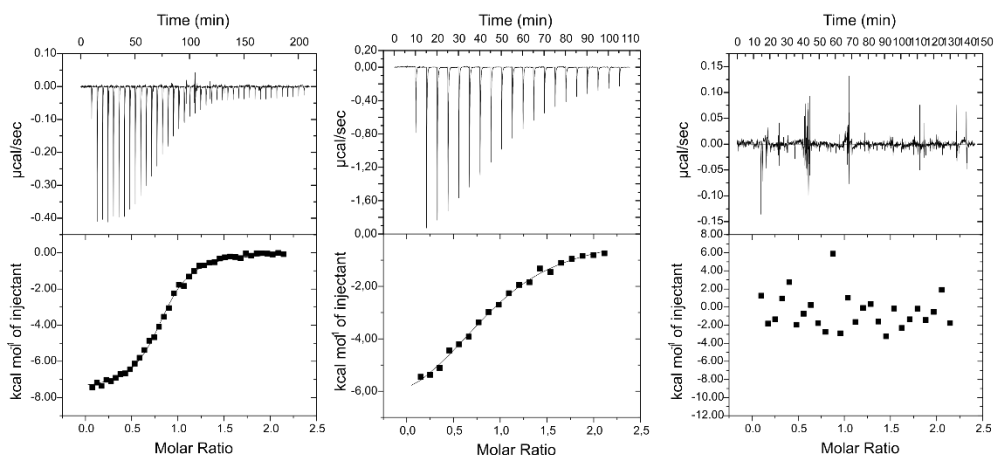
	C-SH2$^{102-220}$:Gab1$^{655-677}$	C-SH2$^{102-220}$:Gab1$^{655-666}$	C-SH2$^{102-220}$:Gab1$^{667-677}$
N	0.95 ± 0.11	1.0 ± 0.1	-
K_D [nM]	4233 ± 295	53605 ± 8071	-
- ΔS · T	515.60 ± 301.54	860.65 ± 129.03	-
ΔH	-7273.33 ± 630.67	-6927.75 ± 346.08	-
ΔG	-7332.16 ± 40.73	-5833.53 ± 85.21	-

Figure 32 *ITC measurements of the C-SH2$^{102-220}$ domain with Gab1 fragments. Replicates of those measurements can be found in the supplemental **Figure 61**. Analysed with Origin.*

The interaction with the longest peptide, pY-Gab1$^{655-677}$, showed a similar affinity to the reaction to double-phosphorylated pYpY-Gab1$^{613-694}$ (K_D 4.2 ± 0.3 µM, compared to K_D of 2.7 ± 0.37 µM). The stoichiometry showed a 1:1 binding in an enthalpically driven reaction, with nearly no entropic contribution. Differences in the entropy and enthalpy compared to the pYpY-Gab1$^{613-694}$ (**Figure 20**) might result from the much longer

construct. Interestingly, the reaction to Gab1[655-666] was less affine with a K_D of 53.0 ± 8.0 µM, while having similar thermodynamic, although slightly less unfavourable entropic, contributions. The unphosphorylated Gab1[667-677], corresponding to the second binding site (ii), showed no measurable change in enthalpy in the presence of C-SH2[102-220] domain.

4.2 CSP

HSQC experiments using the ^{15}N, ^{13}C-labelled C-SH2[102-220] were measured in the free state and bound to all three peptides as well as the doubly phosphorylated pYpY- Gab1[613- 694] (supplemental **Figure 63**). The HSQC of the C-SH2[102-220] : Gab1[667-677] was identical to the free C-SH2[102-220] spectrum (supplemental **Figure 62**), in accordance with the ITC experiments. The calculated chemical shift perturbations for the two phosphopeptides are shown in **Figure 33**.
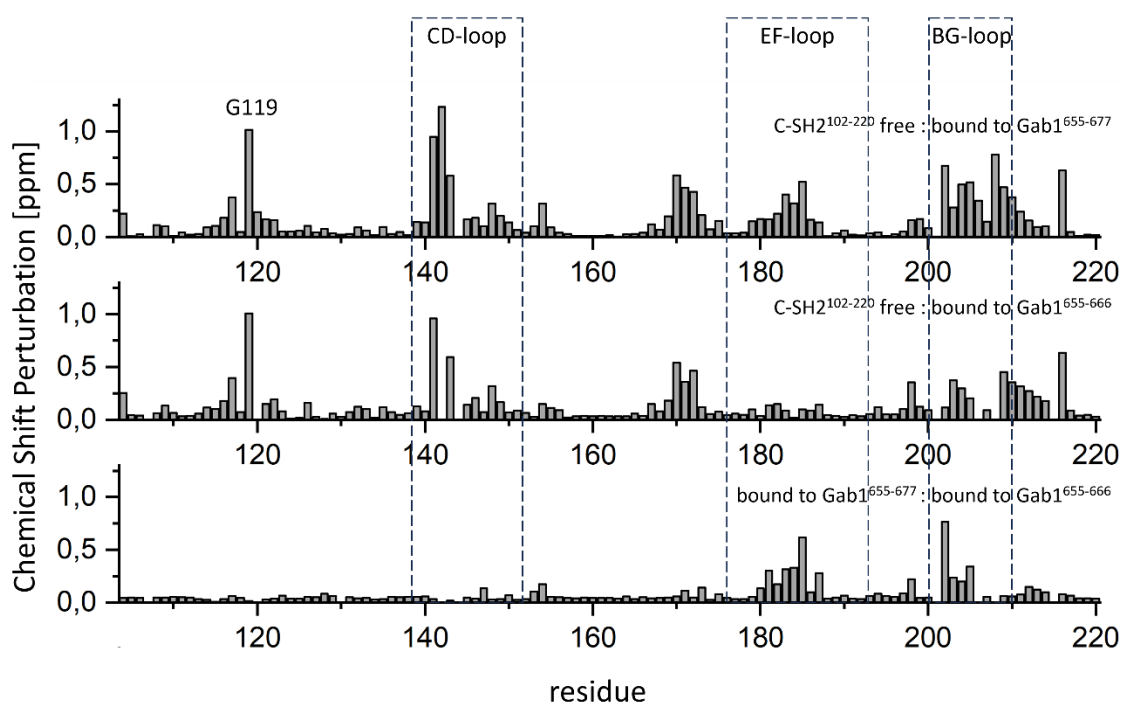
Figure 33 Chemical Shift Perturbations (CSP) in ppm for C-SH2[102-220] in complex with the pY-Gab1[655-677] (top) and the pY-Gab1[655-666] (middle). The lower panel shows the CSPs when comparing both bound states to one another. y-axis: CSP in ppm, x-axis: the C-SH2[102-220] domain residue number of. Illustrated in Origin.

The binding of pY-Gab1[655-677], resulted in a similar HSQC-spectrum to that measured for the complex of C-SH2[102-220] with doubly phosphorylated pYpY-Gab1[613-694] (supplemental **Figure 62, F**). Residues of the helix αA (especially G119), the pY-binding region (BC-loop), the βC and βD strand, part of the CD-loop, as well as the EF- and BG-loop show elevated CSP values, indicating a change in the chemical environment upon binding. These identical CSPs (between binding pY-Gab1[655-677] and pYpY-Gab1[613-694]) align with the ITC measurements of both complexes and show that the complete binding to the C-SH2 domain is restored when the pY-Gab1[655-677] fragment is present. The shorter

phosphopeptide, pY-Gab1$^{655-666}$, also resulted in chemical shift changes (CSP), although different from the longer peptides (e.g. with pYpY-Gab1$^{613-694}$ or pY-Gab1$^{655-677}$). As seen in **Figure 33**, residues at the pY-binding site (around aa 120, aa 140, and aa 170) and the specificity region (regions around aa 180-190 and aa 200-210) are affected by short peptide binding. Comparing data for both phosphopeptides (**Figure 33 lower panel**), it is evident that the pY-binding is equivalent, as no differences in the chemical shifts between both states are present. On the other hand, residues of the EF (aa 179-187)- and BG (aa 200-210)-loop experience different changes in chemical environments upon binding pY- Gab1$^{655-666}$ compared to pY-Gab1$^{655-677}$, leading to a greater variance between both spectra. The EF-loop, for example, is merely slightly involved in the binding, as the residues in this region experience only minor changes. These differences point towards the importance of the Gab1-residues 667-677 and correspond to the drop in affinity of the short peptide to the C-SH2 domain. To investigate whether binding is possible through allosteric effects, an HSQC measurement was made upon adding Gab1$^{667-677}$ to the complex of C-SH2$^{102-220}$:pY-Gab1$^{655-666}$. The resulting spectra were identical to the one of the C-SH2$^{102-220}$:pY-Gab1$^{655-666}$ complex (supplemental **Figure 62, D**), indicating that the Gab1$^{667-677}$ C-terminal peptide corresponding to the second binding region (ii) can only complement the binding when it is covalently attached to the pY-Gab1$^{655-666}$ phosphopeptide.

4.3 Dynamic

As with the interaction of N-SH2$^{1-106}$ with Gab1 (chapter 3), dynamics on the ps-ns time scale were investigated via relaxation rate and hNOE experiments to yield order parameters (S^2=0: entirely flexible, S^2=1: entirely rigid; **Figure 34**). In the free state, the C-SH2 domain is largely rigid with an average order parameter of S^2=0.78; excluding the long and rather flexible C-SH2-specific CD-loop (aa 154-165, only partially assigned; 6 out of 12 aa), the average S^2 increases to 0.80.
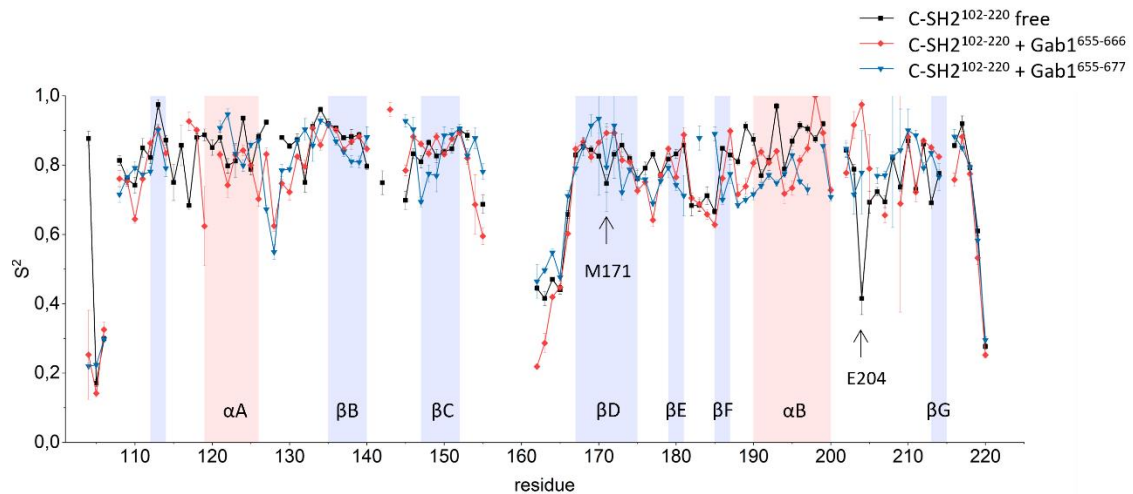
Figure 34 *Order parameter S^2 plotted against the residue of the C-SH2$^{102-220}$ domain for the free domain (black), the pY-Gab1$^{655-677}$ (blue) and pY-Gab1$^{655-666}$ (red) bound complex. The secondary structure elements of the domain are indicated in blue (strand) and red (helix) for better orientation. Residues 107, 141-144, 156-161, 201 and 215 were not assigned, and therefore no S^2 values could be calculated. Illustrated in Origin.*

Regions of highest flexibility are the BC, CD (where assigned), EF, and BG-loop. Interestingly, M171 in the βD strand and E204 in the BG-loop have lower S^2 values than their adjacent residues in the free state. In the pY-Gab1$^{655-666}$ and pY-Gab1$^{655-677}$ bound state, the overall domain flexibility increases slightly with an S^2 of 0.78, excluding the CD-loop. For both peptides, a gain in mobility is seen for the AB-loop (connecting αA with βB), βD and the BG-loop. Nevertheless, differences between both complexes are present. In the pY-Gab1$^{655-666}$ bound state, the EF-loop is as flexible as in the free state, while the CD-loop increases in mobility. In the pY-Gab1$^{655-677}$ bound state, however, the EF-loop becomes more rigid, the CD-loop maintains similar flexibility as in the free domain, while the central β- sheet is slightly more flexible than in the other two states.

4.4 microED

Crystals were obtained from the Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$ complex in two initial screen conditions that were not suitable for X-ray diffraction but yielded electron scattering patterns that allowed structure solution (chapter 3).

4.4.1 Overall structure of the C-SH2 domain and its interaction with pY659 Gab1

Density for the doubly phosphorylated pYpY-Gab1$^{617-684}$ was clearly visible, with the N-terminal pY627-Gab1 fragment bound to the N-SH2 domain reaching from aa 622-631 and a longer C-terminal pY659-Gab1 fragment bound to the C-SH2, covering residues aa 653-672 (**Figure 35**); the intervening residues were not visible (mass spectrometry data of the crystal confirmed the presence of the linker region (supplemental **Figure 55**)) and are presumably disordered in the crystal. Whereas Gab1 residues 653-663 bind to the C-SH2 in a typical extended conformation, residues 664 to 672 adopt a helical conformation. The two binding sites of the tandem domain are oriented in the same

direction, with a significant buried interface between the N- and C-SH2 domains (presented and discussed in chapter 5), where the tandem takes on an orientation in which both binding sites face in the same direction.

4.4.2 SH2 domain

The C-SH2 domain has two α-helices (αA and αB), that flank the central β-sheet (βB, βC, βD-strand).
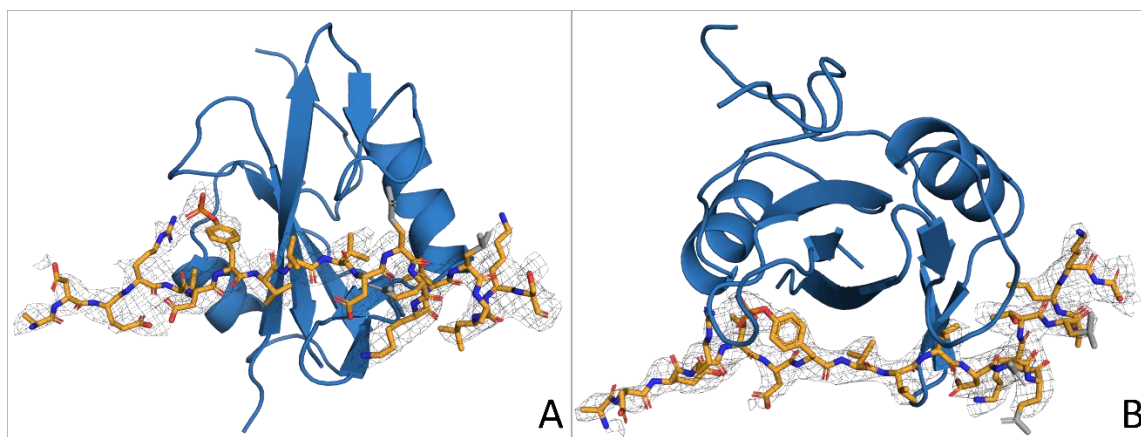
Figure 35 *C-SH2 domain (blue, cartoon) in complex with the pY659-Gab1 fragment (orange, stick). The peptide is covered in its coulomb potential, contoured at $\sigma_{2Fo-Fc}=1$, shown as mesh for both illustrations. Atoms with an occupancy of zero are shown in grey. A) overall depiction of the interaction in the typical orientation. B) Top-down orientation emphasizing the C-terminal helix. Amino acids facing the solvent are not entirely modelled (grey). Illustrated with PyMol.*

The long CD-loop, which connects the βC- and βD-strands could not be modelled in the crystal structure, probably due to its disorder (**Figure 35**), whereas the specificity determining EF- and BG-loops, that interact with the Gab1 peptide are clearly defined.
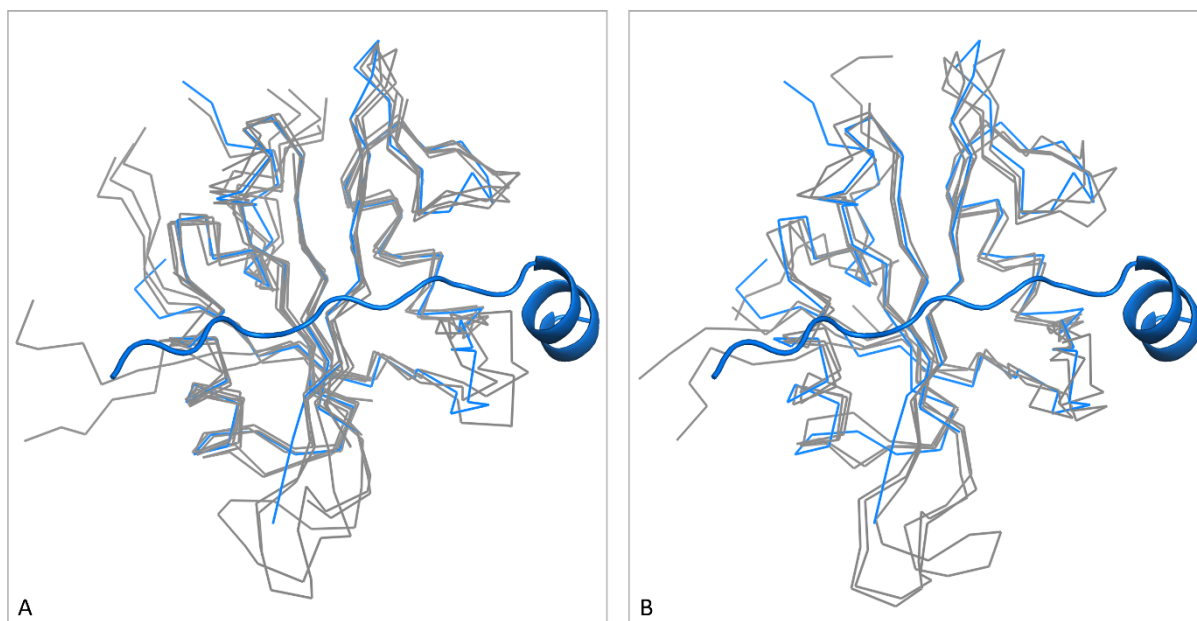
Figure 36 The C-SH2 domain in complex with Gab1 (blue) compared with other Shp C-SH2 structures. The C-SH2 domain solved by microED (ribbon, marine blue) in complex with the Gab1 fragment (cartoon, marine blue) is seen in an overlay with A) the C-SH2 domain of the closed (PDB:4dgp Shp2 (Yu et al. 2013); 2b3o, Shp1 (Yang et al. 2003)) and open (PDB: 6crf, Shp2 (LaRochelle et al. 2018); 3ps5, Shp1 (Wang et al. 2011)) conformation of Shp1 and Shp2 (ribbon, grey) and B) the isolated C-SH2 domain of Shp1 and Shp2 in the free (PDB:1x6c, Shp1 (Sato et al. unpublished) and bound (PDB: 6r5g (Marasco et al. 2020a) (Shp2:PD-1); 2yu7 (Kasai et al. unpublished) (Shp1:NKG2A) states (only the domain is depicted for clarity of the illustration) solved by NMR (ribbon, grey). Illustrated with PyMol.

Overall, the backbone orientation is similar to the C-SH2 in the closed conformation (PDB:4dgp, Shp2 (Yu et al. 2013); 2b3o, Shp1 (Yang et al. 2003)) and in the open conformation (PDB: 6crf, Shp2 (LaRochelle et al. 2018); 3ps5, Shp1 (Wang et al. 2011)), both in Shp2 and in Shp1 (**Figure 36**). The long CD-loop varies between two structures (PDB: 4dgp, Shp2 (Yu et al. 2013); 3ps5, Shp1 (Wang et al. 2011)), where it is solvent exposed and not involved in crystal contacts. No crystal structure of the isolated Shp2 C-SH2 domain has been published so far: an NMR structure in complex with a PD-1 phosphopeptide fragment is available (PDB: 6r5g (Marasco et al. 2020a)), as well as the Shp1 C-SH2 domain in the absence (PDB:1x6c, (Sato et al. unpublished)) and presence (PDB: 2yu7 (Kasai et al. unpublished)) of a phosphopeptide ligand. The same overall backbone orientation is found in all three NMR structures, with the CD loop in defined (although different) conformations. Thus, the missing density for the CD-loop in the present structure could indicate a variety of superimposed loop structures or an increase in flexibility upon pYpY-Gab1$^{617-684}$ binding.

The pY659-Gab1 fragment reaches from aa 653-672, corresponding to position -6 to +13 relative to pY659. Gab1 peptide amino acids E655-V662 (-4 to +3) adopt an extended conformation (supplemental **Figure 60**), with the side chains of residues -6 to -5 and the residues at -1 facing towards the solvent. E655, at position -4, interacts with crystal water through its backbone, while the side chain points towards K120 without direct contact.

The side chains of -2 and -3 face the domain, where V657 (-2) is buried deep while facing the hydrophobic G119 in the αA helix. G119 is similar to G13 in the N- SH2, a Shp-specific feature that allows for hydrophobic amino acids of the binding partner in the -2 position. The side chain of R656 at the -3 position of the Gab1 peptide is located ambiguously, with probably more than one conformation being present. It is either facing towards the domain and is thus in interaction to the phosphate, or it might face outside of the domain. Amino acids 664-672 have torsion angles that populate the α-helical region of the plot (supplemental **Figure 60**).

The phosphate group of pY659 is complexed by the C-SH2 domain by hydrogen bonds with the hydroxyl group of S140 and S142, as well as the backbone of Q141 (**Figure 37**). An ionic interaction is formed with the side chain of R138 (position βB6). These interactions are equivalent to the N-SH2 domain.
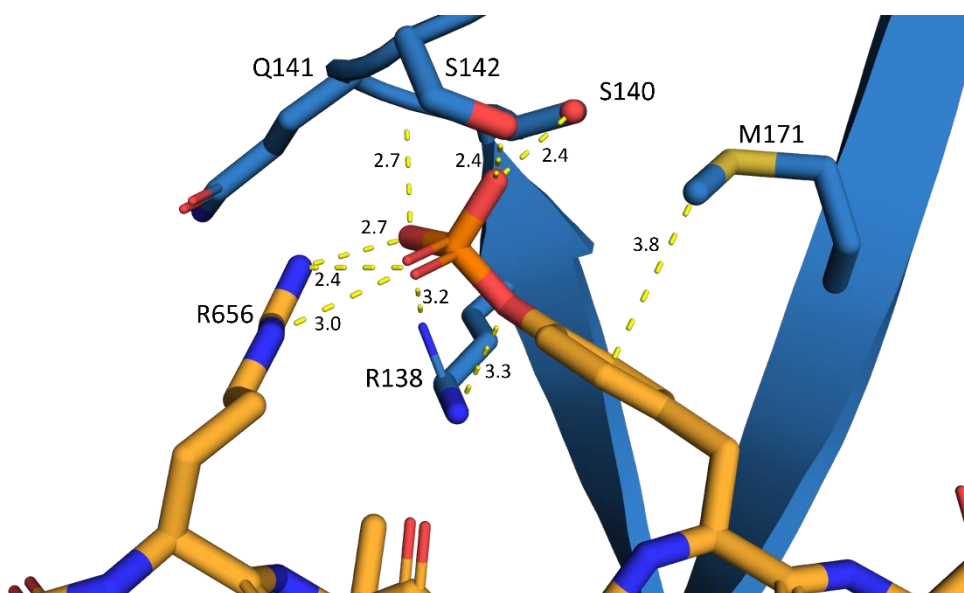
Figure 37 *Interaction of the pY659 with the residues of the C-SH2 domain and R656. The domain is shown as cartoon with involved residues depicted as sticks in blue. These residues are labelled, and interactions to the pY659 residue of Gab1 are indicated with yellow dashed lines and stated distances in angstrom. The Gab1 peptide is shown in orange and as stick model. Illustrated with PyMol.*

Specific to the C-SH2 domain in complex to Gab1 is the -3 position (R656), in the inward position, which is further stabilised by the side chain of Q141. The tyrosyl ring is coordinated by a methionine-pi interaction with M171, which is located at position βD6.

The interactions of the peptide's further residues to the domain are illustrated in **Figure 38**. As for the N-SH2 domain, the C-SH2 domain has a hydrophobic patch formed by V170, V181, Y197, M202, L210
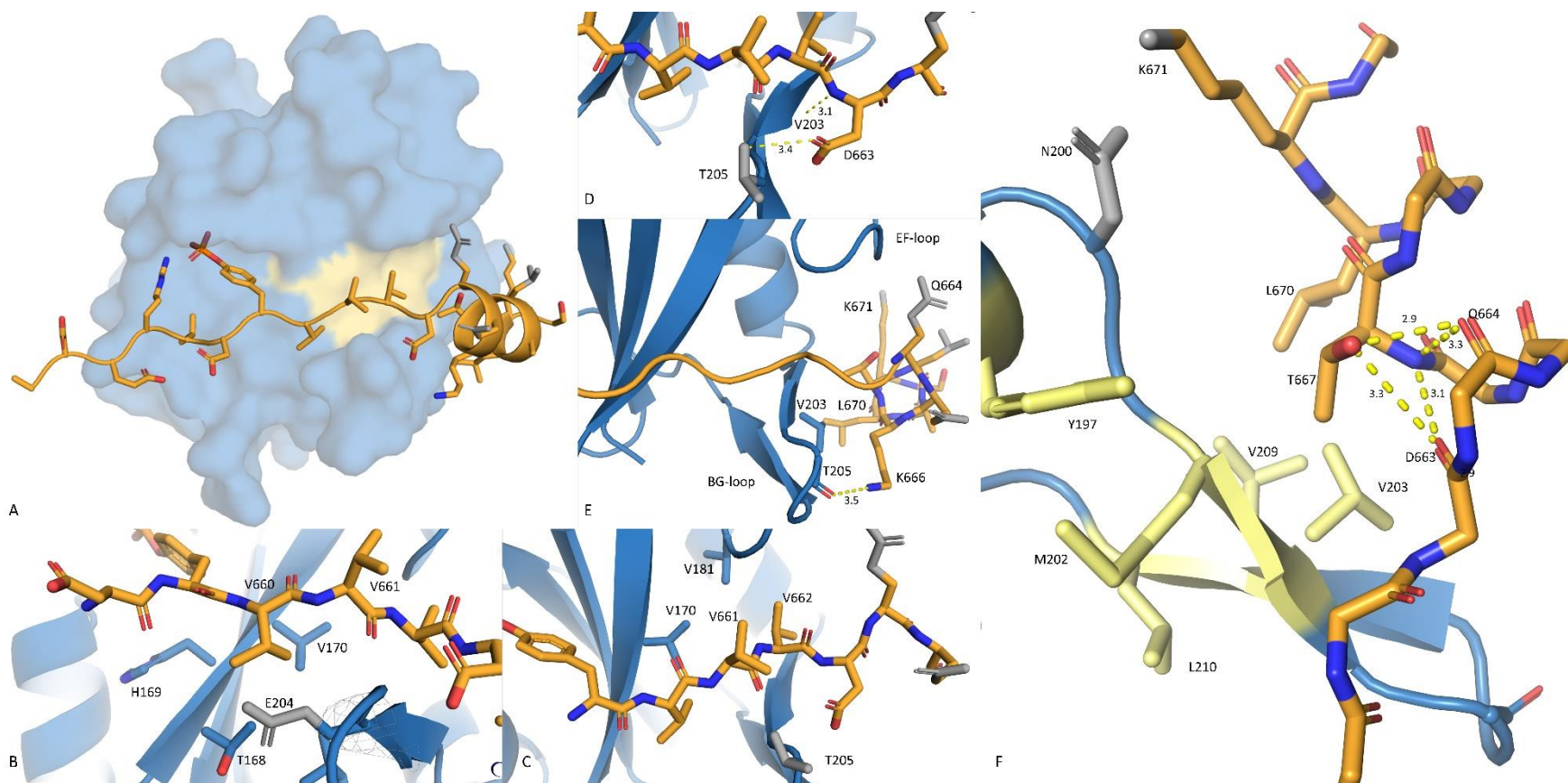
Figure 38 *Interaction of the C-SH2 domain (blue) with the pY659-Gab1 fragment (orange cartoon and stick). Residues that have an occupancy of zero are coloured in grey. A) The C-SH2 domain is illustrated as a surface model with the yellow region indicating the hydrophobic patch (V170, V181, Y197, M202, L210). The +3 residue (V662) of Gab1 faces this region, while the +5 position (Q664) points towards the EF loop, avoiding the hydrophobic patch. B) The +1 residue (V660) NH-group is hydrogen bonded to the backbone carbonyl-oxygen of H169 and is hydrophobically shielded by the side chains of T168, V170 and the aliphatic portion of E204. C) The +2 position (V661) is solvent exposed. Its backbone carbonyl oxygen-hydrogen bonds to the backbone amide of T205. The +3 position (V662) is in van-der-Waals' contact to residues V170 and V181. D) The side chain of the +4 position (D663) makes a hydrogen bond to the T205 hydroxyl group, while its backbone amide stabilises the helical turn by H-bonding to the carbonyl oxygen of T667 (+8) and the amide of V203. E) The +5 position (Q664) is the first helical amino acid and faces away from the domain's hydrophobic patch to point towards the EF loop. F) The +8 side chain (T667) is hydrogen bonding through its hydroxyl group with the backbone groups of Q664 and D663. The +11 position (L670) is in the hydrophobic environment of V203 and V209 (BG-loop). The +12 position (K671) could interact with N200. Residues involved in the hydrophobic patch and its extension are coloured in yellow. Illustrated with PyMol.*

93

The amide group of the +1 position (backbone of V660) is in polar interaction with the carbonyl oxygen of the H169 backbone. The side chain faces the hydrophobic residues V170 and L210. It is probably additionally shielded from the solvent by the CH2-groups of E204 and hydrophobic interaction with the methyl group of T168. The side chain of V661 (position +2) is solvent-exposed. The backbone might facilitate a polar interaction with the backbone amide of T205. V662 (position +3) faces the domain, where V170 and V181 form a hydrophobic surface. The carboxyl group of D663 (position +4) could hydrogen bond to the hydroxyl group of the T205 side chain as well as to the backbone of T205 and V203, situated in the BG loop. The main chain carbonyl group is complexed by the amide of the T667 main chain and its hydroxyl group of the side chain. The +5 (Q664) position is the tip of the newly formed α-helix of Gab1. The side chain of Q664 is pointing upwards towards the EF loop to G183. Unfortunately, not the entire side chain is covered in density, making it impossible to make assumptions about interactions to the specificity loop. The side chain of Q665 (+6) is solvent-exposed and not completely modelled, nevertheless, its backbone stabilises the helical turn by interacting with the backbone of A669. The side chain of K666 (+7) points towards the EF-loop, where it interacts with the backbone of T205. The main chain carbonyl group is H-bonding the amide backbone of L670 (+11) to stabilise the helical conformation. T667 (+8) faces the domain while being shielded from electrostatic interactions. The hydroxyl group interacts with the carbonyl groups of the Gab1-backbone of residue D663 and Q664 as well as with the amide of the L668. The backbone carbonyl group is again stabilising the helical turn by interacting to the amide groups of K671 (+12) and additionally to L670 (+11). The side chains of L668 (+9) and A669 (+10) are solvent-exposed, and the leucine side chain is not completely modelled. Both stabilise the helical conformation through backbone interactions (L668 to Q664 and S672, while A669 interacts with Q665). L670 (position +11) is in hydrophobic interaction with V203 and V209 and further shielded by the proline side chain (P201), all situated in the BG-loop. K671 (+12), although the NH3 group is not modelled, the side chain faces the C-SH2 domains' αB-EF-loop transition, where it could interact with N200 electrostatically. S673 (+13) is solvent-exposed and the last modelled amino acid of the pY659-Gab1 peptide.

Overall, the binding of pY659-Gab1 to the C-SH2 domain is more extensive than that between pY627-Gab1 and the N-SH2 domain as well as to other structures of the C-SH2 domain complexed with a peptide. The NKG2A peptide (PDB: 2yu7 (Kasai et al. unpublished)) provides also the N-terminal amino acids up to -7. These show similar orientation facing the domain or the solvent as in Gab1, except for the -3 position (glutamic acid) facing outwards, which corresponds to the phosphate interacting R656 in Gab1. Another exception is the leucine at the +5 position, which faces the hydrophobic

patch, while Q664 is pointing towards the EF-loop. Residues C-terminal of the +5 position do not interact with the domain and bend away from it. Another peptide with residues modelled up to -6 is the EPIYA-motif D of CagA (PDB: 5x94 (Hayashi et al. 2017)). Interestingly, residues N-terminal of -2 are in a lower position as for Gab1, leading to an additional interaction to the αA helix. The orientation of the side chains up to +5 is similar to Gab1; again, with the hydrophobic phenylalanine (+5) facing the domain, the C-terminal +6 residue is in further extension of the elongated peptide. The same peptide but with an aspartic acid at the +5 position (CagA EPIYA-motif C (PDB: 5x7b (Hayashi et al. 2017) is only solved from -1 to +5. The side chain of the aspartic acid at +5 is not modelled. For the PD-1 peptide (PDB: 6r5g (Marasco et al. 2020a)), the first two amino acids (-4 and -3) as well as the C-terminal proline at +6 are bent away. Again the hydrophobic phenylalanine in the +5 position is facing the hydrophobic patch. The TXNIP peptide bound to the C-SH2 domain (PDB: 5df6 (Liu et al. 2016)), modelled from -7 to +4, shows a similar backbone and side chain orientation in the C-terminal to -3. The N-terminal region is again positioned lower as Gab1 but without additional interactions to the domain and in the extension of the elongation (not bent). From these data, it is evident that the N-terminal parts of the peptides are often solved up to the -6 position, in contrast to the peptides binding the N-SH2 domain, where the density is not strong enough beyond the -2 position (Anselmi et al. 2020) (exception: the -5 position of pY627-Gab1 bound to the tandem SH2 domains).

4.5 Discussion

Many structural studies have focussed on the N-SH2 domains of Shp phosphatases, as phosphopeptide binding must be associated directly with uncoupling from the catalytic domain. In contrast, structural information on the C-SH2 domains is scarce, although the importance of bidentate phosphopeptide ligands for Shp phosphatase activation strongly suggests that binding to the C-SH2 is of crucial importance. This makes additional findings presented in this thesis on the C-SH2 domain in complex with its native binding partner Gab1 especially important.

In the closed state of Shp2 (PDB: 4dgp (Yu et al. 2013)), the peptide-binding site of the C-SH2 domain is solvent-exposed and thus accessible to phosphorylated binding partners, while the opposite face of the domain faces the PTP domain. Few contacts are made between the N-SH2 and C-SH2 domains, largely mediated by the covalent linker between them. A similar arrangement is found in the closed conformation of Shp1 (PDB: 2b3o (Yang et al. 2003)). Interestingly, the position of the N-SH2 domain - and the closure of the EF-loop, restricting access to the +3 site - is nearly identical in both closed structures (**Figure 39, A-C**). Upon activation, the N-SH2 domain is released from the

PTP domain active site. In the constitutively active Shp2 mutant E76K (PDB: 6crf (LaRochelle et al. 2018)), the N-SH2 swings behind the C-SH2 domain, which also reorientates itself, although not as dramatically as the N-SH2 domain (**Figure 39**, F & G). An analogous rearrangement is seen for the active conformation of Shp1 (PDB: 3ps5 (Wang et al. 2011)), although the N-SH2 domain is positioned in a considerably different way to E76K-Shp2 (**Figure 39 D & E**). In a ΔN-SH2 Shp2 structure (PDB: 6cmq (Pádua et al. 2018)), classified as an open conformation, four molecules in the asymmetric unit show variance in the orientation of the C-SH2 domain, suggesting that dissociation of the N-SH2 domain from the PTP domain in turn loosens the interface between the C-SH2 and PTP domain.

The C-SH2 domain shows similar backbone orientations in all free, bound and Shp1 and Shp2 structures available, which is probably due to the relative accessibility of the domain in the open and closed conformations. The ligand binding sites are open, in contrast to the N-SH2 domain in the closed conformation and can directly accommodate a peptide without conformational reorganisation. As shown for the PD-1:Shp2 interaction, the C-SH2 domain recruits the bidentate peptide and reorientation of the SH2 domains is necessary to accommodate the 24 aa-long pY-linker of PD-1 (Marasco et al. 2020a). Since Gab1 has a longer linker (31 aa), such a restriction does not apply. As can be seen from **Figure 39**, in the closed conformation, the Gab1 linker (aa 632-653) needs to span a distance of 68 Å, which is within reach for Gab1 but definitely not for PD-1.
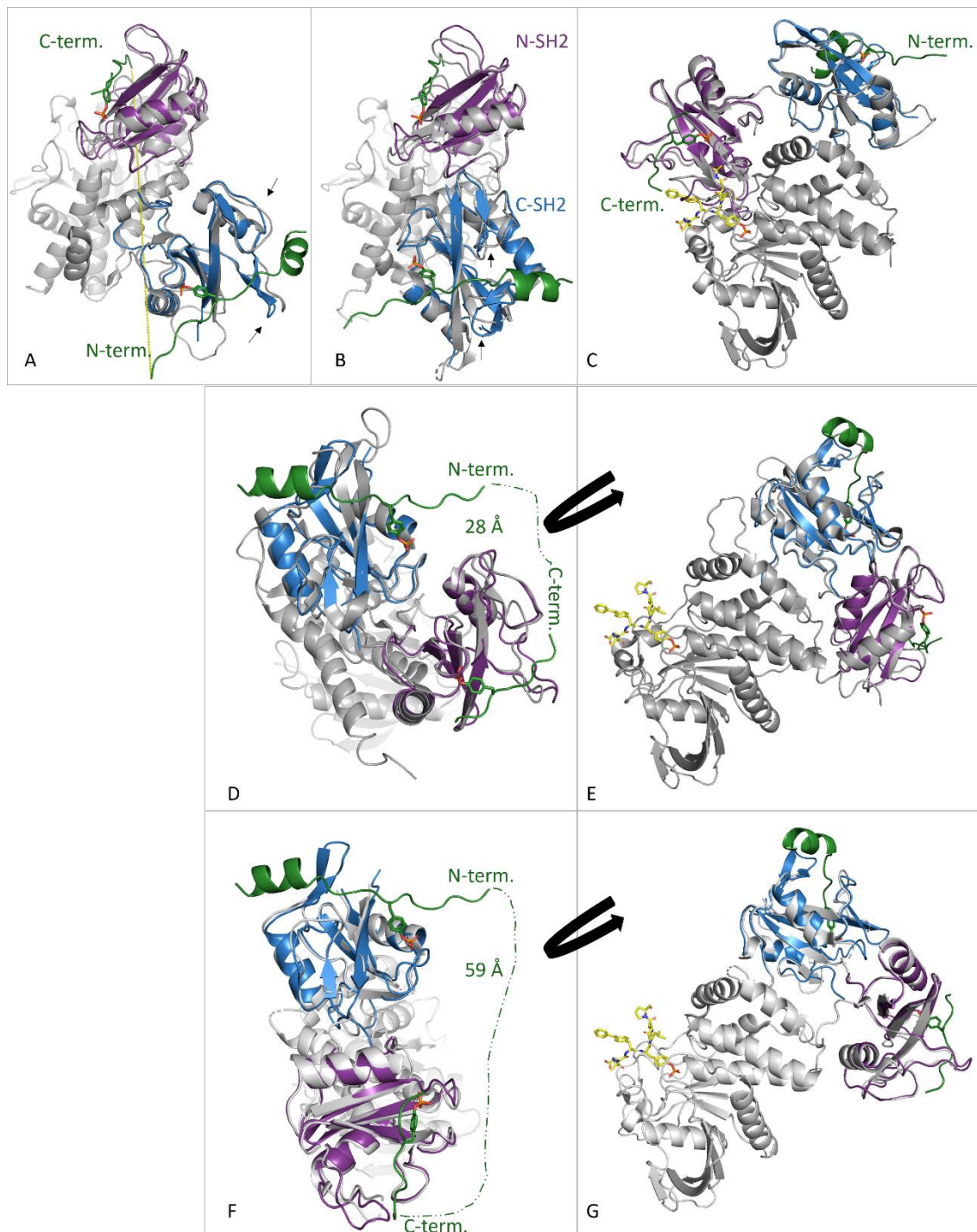
Figure 39 *Model of activation. Shp2 and Shp1 models (grey) in their closed and open conformations superimposed with the N-SH2:Gab1 (purple: green) and C-SH2:Gab1 (blue: green) bound state (from this thesis). A) Closed Shp2 (PDB: 4dgp (Yu et al. 2013)); the dashed line in yellow indicates the 68Å between residue 632 (pY627-Gab1) and 653 (pY659-Gab1). Black arrows indicate the +3 binding site. B) Closed conformation of Shp1 (PDB: 2b3o (Yang et al. 2003)) in the same orientation as A. Black arrows indicate the +3 binding site. C) Structure of the closed Shp2 (PDB: 4dgp (Yu et al. 2013)); The N-SH2 domain occupies the active site of the PTP domain. For illustration purposes, a substrate (yellow, stick – taken from PDB: 1een (Sarmiento et al. 2000)) is bound to the active site in C, E and G – all depicted in the same orientation. D) & E) Active Shp1 (PDB: 3ps5 (Wang et al. 2011)). The distance between 632 and 653 shrinks to 28 Å (straight line). F) & G) Active Shp2 (PDB: 6crf (LaRochelle et al. 2018)). The distance between 632 and 653 is 59 Å (straight line). E) & G) The active site is accessible for the substrate.*

As seen from chapter 2, C-SH2 binding to Gab1 affects two regions: aa 657-665 (i) and aa 666-676 (ii), and the two Gab1 phosphopeptides, pY-Gab1$^{655\text{-}666}$ and pY-Gab1$^{655\text{-}677}$

show distinctive binding behaviour towards the C-SH2 domain. ITC experiments showed that the short peptide has a lower affinity while the longer Gab1 peptide could restore the K_D of the interaction between C-SH2$^{102-220}$:pYpY-Gab1$^{613-694}$. Additionally, NMR experiments showed that the binding site for the phosphotyrosine was identical (as seen in the chemical shifts and dynamic of the domain), but the interactions with the EF- and BG-loops were different. The enhanced affinity for the longer peptide is explained by the extended interface afforded by the pY659-Gab1 C-terminal helix (residues Q664-S672). In canonical SH2 binding modes, residue Q664 (+5 position) is expected to face the domain. The SH2 domains of Shp2 and Shp1 each have a large hydrophobic surface (illustrated as a yellow surface on the illustrations of the N-SH2 (**Figure 27, B**) and C-SH2 domain (**Figure 38, A**)) facing the activator peptide's +1, +3 and +5 side chain, so that hydrophobic amino acids are favoured over charged ones (Marasco et al. 2020a). Indeed, an investigation of CagA peptides with an EPIYA-motif binding to the Shp2 C-SH2 domain showed that a hydrophobic phenylalanine side chain at position +5 results in an affinity of 1.4 µM, while replacements with the negatively charged aspartate residue decreases the K_D to 60.5 µM (Hayashi et al. 2017). Thus, the presence of the polar Q664 at the +5 position in the two tested Gab1 peptides would be expected to be unfavourable, as seen in the pY-Gab1$^{655-666}$ peptide. For pY-Gab1$^{655-677}$ bound, however, Q664 is removed from the interface through the formation of the helix so that instead, T667 and L670 juxtaposition the hydrophobic patch (V170, V181, Y197, M202, L210) and its extension within the BG-loop (V203 and V209) on the surface of C-SH2. In the N-SH2, the EF-loop's K89 and K91 side chains have been postulated to interact with the charged aspartate side chains in the +2 and +4 positions, based on MD simulations (Anselmi et al. 2020). In the C-SH2 domain, the corresponding residues are V203 and T205. These smaller and more apolar residues fit better to the long side chains of K666, T667 and L670 located in the Gab1-helix, which face towards the EF-loop. Further, the hydrophobic V661 (pY659-Gab1) at the +2 position is not confronted with charged lysine side chains (K89, K91 of the N-SH2), while the aspartate at the +4 (D663) position can still build polar contacts with the hydroxyl group of T205. These structural features probably result in the preference of the C-SH2 domain towards pY659-Gab1. NMR experiments further showed that both binding regions (i – extended; ii – helical) need to be covalently linked in order to interact correctly with the C-SH2 domain. From the CD-spectra of the pY-Gab1$^{655-677}$ peptide (chapter 1 **Figure 18**), it is evident that in the unbound state the peptide is disordered. Only upon binding to the domain will it adopt its helical conformation.

To the knowledge of the author, other helical interactions to the SH2 domains of Shp2 or Shp1 are so far not described. The elevated affinity propagated through the helical

conformation in pY-Gab1$^{655-677}$ compared to pY-Gab1$^{655-666}$ and thus the avoidance of the charged +5 position (Q664) forced towards the hydrophobic patch, could potentially also have an impact on the activity of Shp2. Unfortunately, this could not be tested within this thesis. Since the +5 position is almost exclusively a hydrophobic, mostly aromatic residue in Shp2-activating peptides (Marasco et al. 2021), CagA EPIYA-motif C (aspartic acid residue compared to phenylalanine in EPIYA-motif D (Hayashi et al. 2017)) could potentially also adopt a helical conformation upon binding. Interestingly, this region is also predicted to be helical by AlphaFold (AlphaFold Protein Structure Database (Jumper et al. 2021; Varadi et al. 2022); entry: Q7WS60 – residue pY208). Nevertheless, predictions as a complex either with the tandem SH2 domain or the C-SH2 domain do not show a valid binding (no binding in the tandem; pY binding in the specificity loops of C-SH2 – data not shown). Nevertheless, since CagA is a pathogenic protein interacting with the Shp2 host phosphatase, structural analysis of this interaction might be beneficial.

The dynamic of the domain backbone upon binding the pY-Gab1$^{655-666}$ peptide shows a flexible CD-loop, which is more rigid in the free and pY-Gab1$^{655-677}$ bound state. It could be that the elevated flexibility in the unfavoured reaction to pY-Gab1$^{655-666}$ is needed to facilitate the binding of such a non-ideal target. Unfortunately, this could not be validated by ITC as both reactions have similar entropic terms. The purpose of the long-C-SH2-specific CD- loop is still not understood, and further research needs to be done.

Chapter 5: novel SH2 domain interface on binding Gab1 suggests a similar open conformation of Shp2 and Shp1

The small molecule orally available Shp2 inhibitor SHP099 acts allosterically, stabilising the interface between the N-SH2, C-SH2 and PTP domains of Shp2 in the closed conformation to suppress activation (PDB: 5ehr (Chen et al. 2016)) . These interfaces are dissolved upon activation, and new ones are formed; studying the buried surface areas in each conformation allows analysis of the energetic terms contributing to activation and can further help in the development of specific targeting molecules.

The N-SH2:C-SH2 interfaces of all conformations known to date are presented in **Figure 40**. In the closed conformation (PDB:4dgp (Yu et al. 2013)), the domains form an interface of 223 Å (PISA; (Krissinel and Henrick 2007)), in which the interdomain linker plays a predominant role. The interaction is further stabilised by charged and polar interactions of R5 (back side of the N-SH2 domain, relative to the binding site (front side) of the activator peptide – for clarification of the faces of the domain see supplemental **Figure 64**) to amino acids of the αB helix of the C-SH2 domain (back side). In the open conformation (PDB: 6crf (LaRochelle et al. 2018)), the relative orientation of the C-SH2 is tilted so that the N-terminal flexible region of the N-SH2 domain (back side) faces amino acids of the βC and βF strands of the C-SH2 (top). In contrast to the closed conformation, there are no charged interactions; the surface area comprises 236 Å (PISA; (Krissinel and Henrick 2007)) for one molecule in the asymmetric unit and 99 Å for the second molecule. The larger area of the first molecule is probably a result of crystal packing, as the N-SH2 domain is stabilised by crystal contacts to two other symmetry mates. The SH2 domains are not involved in crystal packing for the second molecule in the asymmetric unit and are thus not squeezed together. Interestingly, the open conformation of Shp1 (PDB: 3ps5 (Wang et al. 2011)) differs, with a three times greater surface area of 641 Å (PISA; (Krissinel and Henrick 2007)). The domains are oriented so that their top and back sides can interact; interestingly, a hydrophobic core of W4, F5, F39, Y61, L72 of the N-SH2 faces P142 of the C-SH2. This hydrophobic interaction is supported by salt bridges and H-bonds of R3 and R7 to several residues located in the C-SH2 domain. Additionally, the hydrophobic amino acids P105, Y111, F145, and Y176, though not involved in interaction with the N-SH2 domain, are buried and shielded from the solvent. These interactions do thus contribute to the activation with a loss in solvation-free energy of ΔG^{solv} = -2.3 kcal/mol (Krissinel and Henrick 2007). To date, three structures of the Shp2 tandem bound to phosphopeptides have been deposited (PDB:5df6 (Liu et al. 2016), 5x94 (Hayashi et al. 2017), 5x7b (Hayashi et al. 2017)), which could also potentially correspond to an activated conformation. In all three structures, residues of the αA helix of the N-SH2 are facing amino acids located in the

linker and the DE-loop. Depending on the tilt angle of the C-SH2 to the N-SH2, additional buried amino acids are located in the BC-loop of the C-SH2 and the N-terminal region of the N-SH2. In the structure in which this last effect is not possible due to its overall orientation (PDB:5x94 (Hayashi et al. 2017)), a hydrophobic interaction of I11, L19 and L177 is instead formed and contributes the most to the Gibbs-free-solvation energy. In none of those structures are the peptide ligands bound to the N-SH2 and C-SH2 domains covalently linked, providing the system with additional freedom resulting in three different conformational orientations of the domains, none of which resembles the open conformation of Shp2 or Shp1.
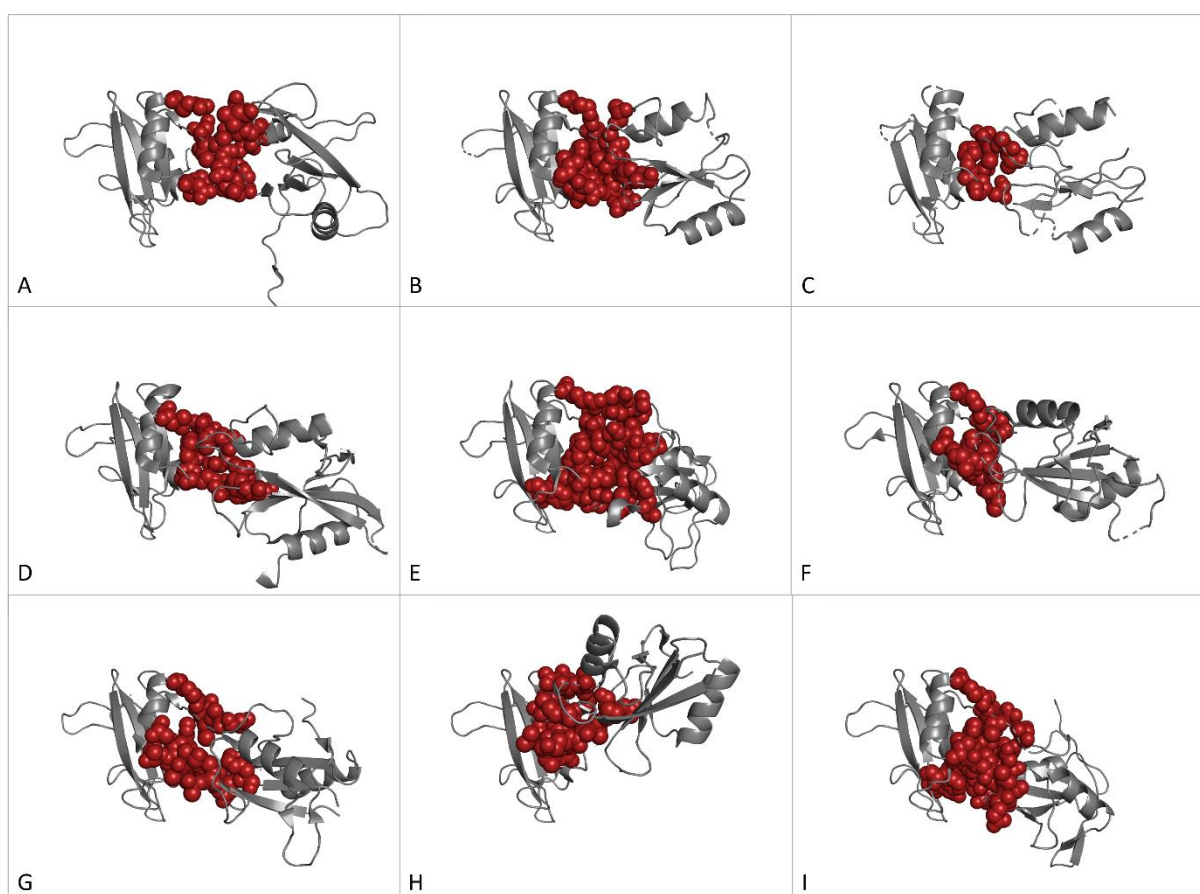
Figure 40 *No common interface between the N-SH2 and C-SH2 has been published to date. The interacting residues from the N-SH2 and C-SH2 domain (cartoon, grey) sown for each structure were calculated using PISA (Krissinel and Henrick 2007) (sphere, red). All structures have been aligned according to the N-SH2 domain. A) closed Shp2 (PDB: 4dgp (Yu et al. 2013)), B) open Shp2 (PDB: 6crf_molecule1 (LaRochelle et al. 2018)) C) open Shp2 (PDB: 6crf_molecule2 (LaRochelle et al. 2018)), D) closed Shp1 (PDB: 2b3o (Yang et al. 2003)), E) open Shp1 (PDP: 3ps5 (Wang et al. 2011)), F) Shp2-tandem in complex with TXNIP peptide (PDB: 5df6 (Liu et al. 2016)), G) Shp2-tandem in complex with CagA peptide (PDB: 5x7b (Hayashi et al. 2017)), H) Shp2-tandem in complex with CagA-peptide (PDB: 5x94 (Hayashi et al. 2017)), I) microED structure of $Shp2^{1-222}:pYpY-Gab1^{617-684}$. The PTP domain and binding partners are not depicted. Illustrated with PyMol.*

5.1 microED

The crystal structure of the tandem SH2 domains ($Shp2^{1-222}$) bound to $pYpY-Gab1^{617-684}$ revealed the interface between the domains to consist of a hydrophobic core formed by F6, F41, Y63, L74 and P144, supported by salt bridges and H-bonds between the N-

terminal region of the N-SH2 domain to amino acids in the linker and the βB-strand of the C-SH2. The overall orientation and the stated interactions are remarkably similar to the above-described open conformation of Shp1 (PDB: 3ps5 (Wang et al. 2011)) (**Figure 41**), leading to an interface area of 414 Å. Since the Shp1 structure shows some additionally involved amino acids, it has a greater interface area. Notably, the amino acids involved in the hydrophobic interface are conserved between both homologues.
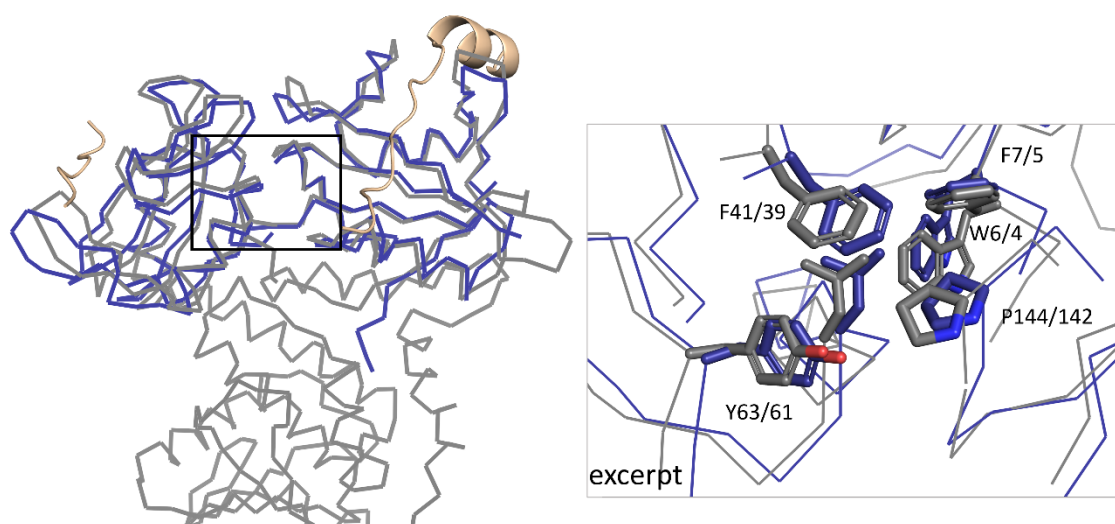
Figure 41 *Structural alignment of the open conformation of Shp1 (grey, ribbon; PDB: 3ps5 (Wang et al. 2011)) with the model solved from the Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$ complex (blue, ribbon; Gab1 fragments are coloured in wheat and shown as cartoon). The excerpt shows the amino acids involved in the hydrophobic interface from Shp1 and Shp2. The residues are labelled as (Shp2/Shp1). Illustrated with PyMol.*

The orientation of the SH2-tandem domains changes drastically between the Gab1 bound structure and the closed Shp2 conformation (PDB: 4dgp (Yu et al. 2013)). While the Gab1-bound conformation shows the above-shown hydrophobic interaction, the SH2 domains in the closed conformation are only minorly attached to one another, mainly by charged interactions between the N-SH2's R5 and the C-SH2's S189 and D192 side chains. The domains are tilted, as seen in **Figure 42**.
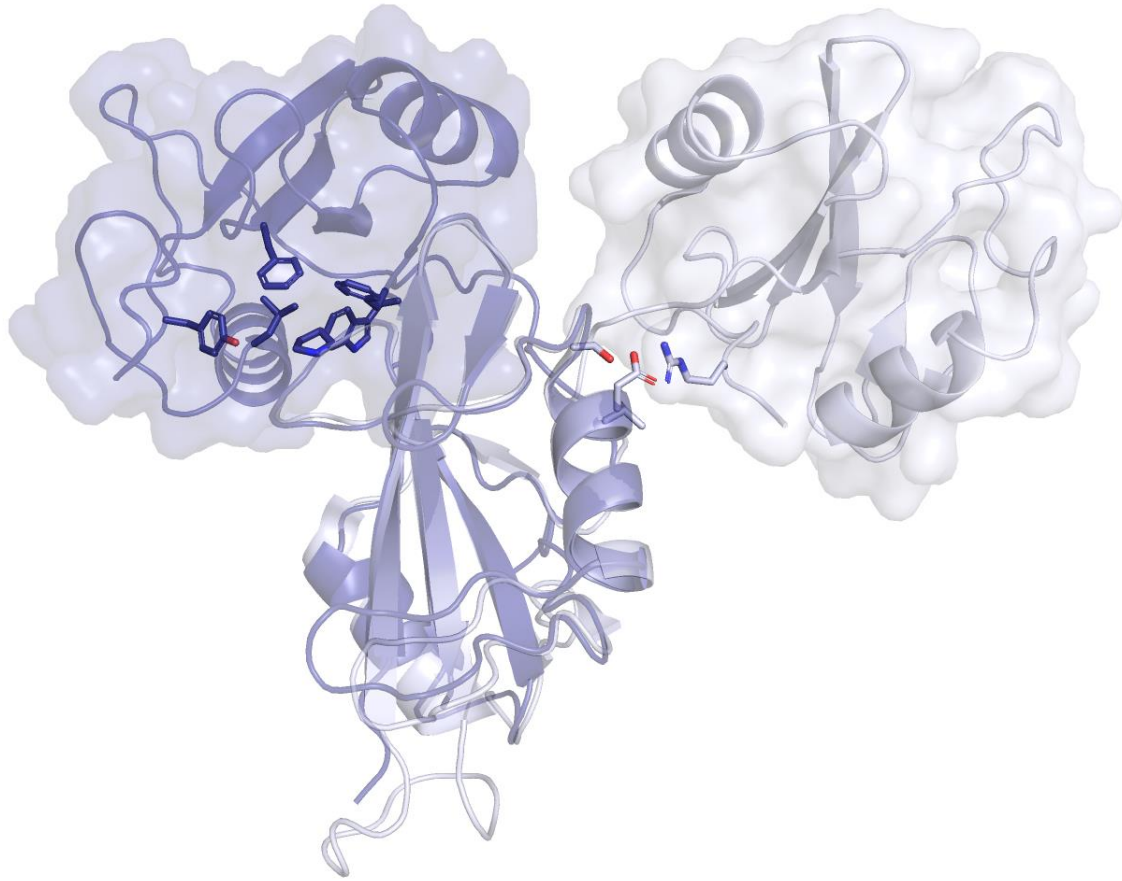
Figure 42 *Structural alignment of the SH2 domain tandem derived from the closed Shp2 conformation (grey; PDB: 4dgp (Yu et al. 2013)) and the Gab1 bound structure (blue, Gab1 not depicted). Both tandem SH2 domains are aligned on the N-SH2 domain, demonstrating that the C-SH2 domains have distinct orientations in both structures. The residues shown as sticks are the ones involved in inter-domain interaction. Illustrated with PyMol.*

5.2 AlphaFold

After completion of the structural work described here, the structure of the complex was predicted using AlphaFold2 (Jumper et al. 2021). The computational prediction of $Shp2^{1-593}$:$YYGab1^{611-694}$ showed a striking similarity to the experimentally derived $Shp2^{1-222}$:$pYpYGab1^{617-684}$ structure. It should be noted that no PDB-templates were used for this calculation. All the more astonishing, both structures align well.
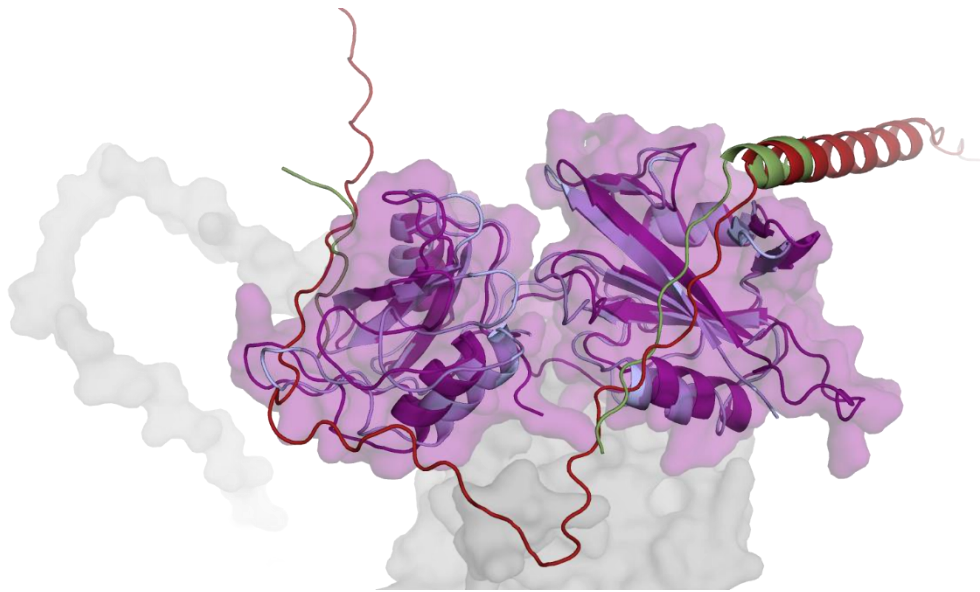
Figure 43 *Superposition of the AlphaFold2 model of Shp2$^{1-593}$:YY-Gab1$^{611-694}$ (Shp2 PTP domain: grey surface; Shp2 SH2- tandem: deep purple, surface and cartoon; Gab1: red, cartoon) to the experimentally solved structure of the Gab1 bound (green, cartoon) Shp2 SH2 tandem (light purple, cartoon). The overall alignment is not as smooth as for the Shp1 structure in the open conformation. Nevertheless, the similarity between the predicted and the experimentally solved structure is evident. Illustrated with PyMol.*

Nevertheless, the AlphaFold model does not align as good as the model of the open Shp1 structure (PDB: 3ps5 (Wang et al. 2011)). However, the striking similarity between both models provides additional confidence that the here-solved structure might actually represent an open and active conformation (**Figure 43**).
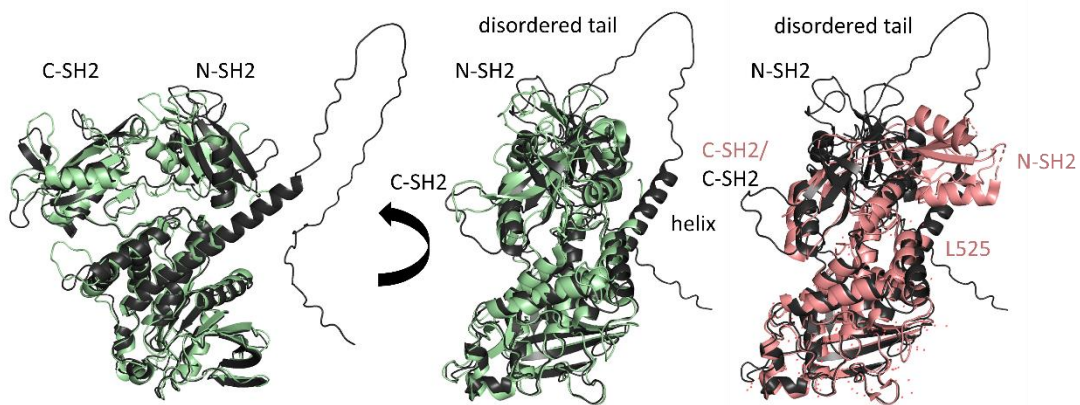
Figure 44 *Superposition of the AlphaFold2 model of Shp2$^{1-593}$:YY-Gab1$^{611-694}$ (black) to the open conformation of the full-length Shp1$^{1-595}$ (PDB: 3ps5 (Wang et al. 2011), green) and the open Shp2 (PDB: 6crf (LaRochelle et al. 2018); salmon) conformations. All structures are depicted as cartoons. The first two graphics show the well-aligned AlphaFold model of the potentially activated Shp2 (black) to the activated Shp1 structure (green). The second orientation is achieved through rotation, indicated with the arrow. The black C-terminal helix of the AlphaFold model is seen in a distinct position to the N-SH2 domain (that reorientates due to the activation of Shp). The third graphic is a superposition of the AlphaFold and activated Shp2 models (PDB: 6crf (LaRochelle et al. 2018)) aligned to the PTP domain. Both the C-SH2 domain and the PTP domain align well. The N-SH2 domain of the experimentally solved structure, nevertheless, move beyond the predicted C-terminal helix of Shp2 of the AlphaFold model. Illustrated with PyMol.*

When comparing the Gab1-bound conformation to the published open conformation of Shp2, a few things must be addressed. Firstly, the published Shp2 structure was solved

using a truncated construct E76K-Shp2[1-525] (PDB: 6crf (LaRochelle et al. 2018)). The AlphaFold prediction suggests that the remaining C-terminal 68 residues of Shp2 extend (at least in part) as a helix, as observed in full-length Shp1 (PDB: 3ps5 (Wang et al. 2011)). In the predicted model of the activated Shp2[1-593], the N-SH2 of the truncated structure would clash with this C-terminal α-helix (**Figure 44**). Potentially, the helix could act as a barrier to the N-SH2 domain, which is in movement due to the activation of the protein. Secondly, the experimentally solved open conformation (PDB: 6crf (LaRochelle et al. 2018)) was achieved using the gain of function mutation E76K. The side chain is in a rather acidic environment formed by E227, E553 and the C-terminal carboxyl group of L525 introduced by the above-mentioned truncation, which would be unfavourable with E76 in wild-type Shp2 (**Figure 45**).
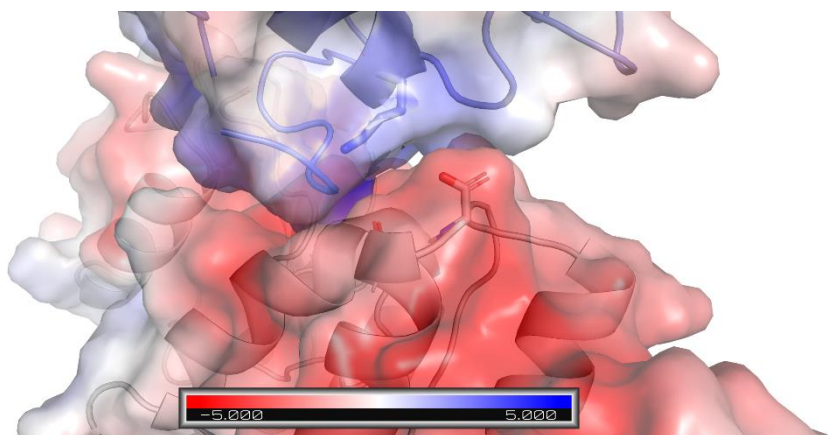
Figure 45 Electrostatic interaction of the K67 side chain to the truncated C-terminus of the PTP domain. The experimentally solved open conformation of Shp2 (PDB: 6crf (LaRochelle et al. 2018)) was achieved by introducing the gain of function mutation E76K. The charge flip leads to an interaction of K67 with the C-terminal carboxyl group of the truncated PTP domain. Illustrated with PyMol.

5.3 NMR

The tandem Shp2[1-222] was labelled with ^{15}N isotope and measured by NMR in complex with unlabelled, phosphorylated pYpY-Gab1[613-694]. The resulting 2D TROSY spectrum was compared to spectra of the free tandem and the labelled single domains bound to phosphorylated pYpY-Gab1[613-694]. The spectra of the free and bound Shp2[1-222] differ dramatically (supplemental **Figure 65**). Superposition of the single domain spectra in their bound state to the bound Shp2[1-222] (**Figure 46**; full spectra in supplemental **Figure 66**) reveals that nearly all peaks align (e.g. A50, K124, G158 and G163). Since a full assignment of the bound Shp2[1-222] complex was not made, a detailed analysis of chemical shift perturbations (CSP) cannot be performed. However, comparing the peak positions in all three spectra allowed their estimation, bearing in mind the critical assessment that the exact amino acid cannot be assigned with certainty.
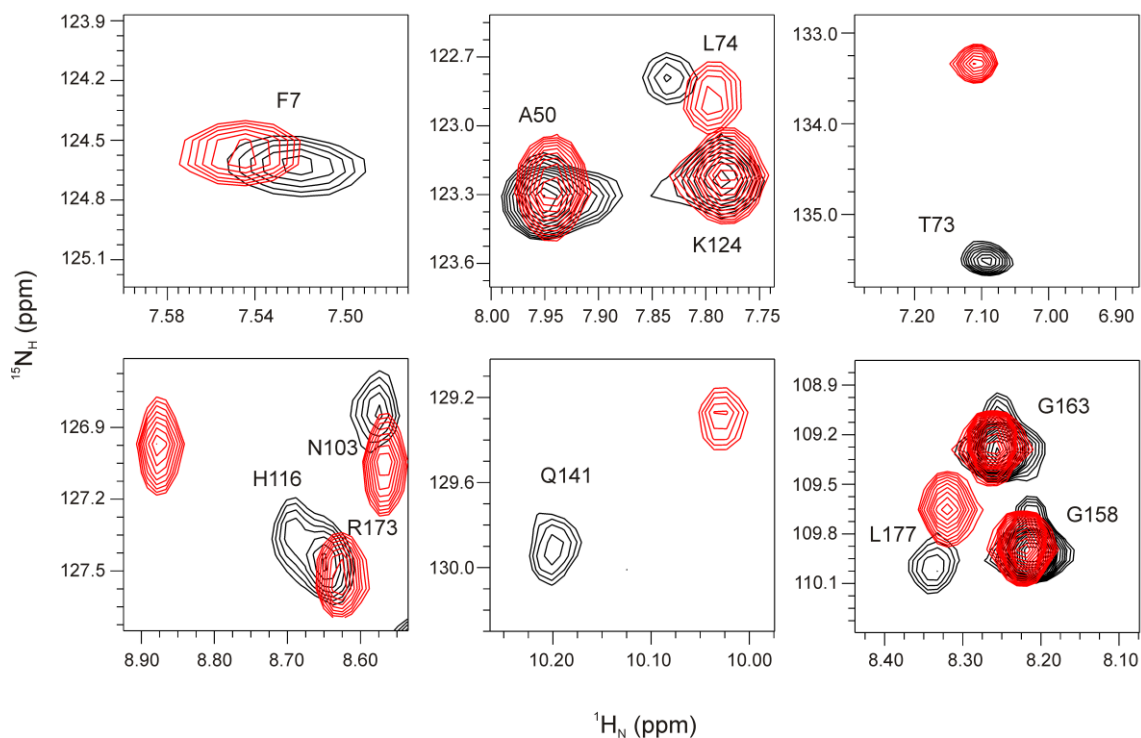
Figure 46 *Excerpt of the ^{15}N, ^1H TROSY spectra of bound Shp2$^{1-222}$ (black) and the bound single SH2 domains (for clarity, both spectra are coloured in red). Residues corresponding to the cross-peaks are labelled. No, minor and heavy chemical shift changes can be seen. Analysed with NMRView.*

No variations in peak positions should result from the binding of the phosphorylated peptide to the binding site, as the binding should be identical in the single and tandem domain systems. Therefore, variations of the spectra should theoretically only result from additional effects of the interaction that are not seen in the single domain systems, e.g., interface formation, remote conformational changes or peripheral binding of Gab1 outside of the binding sites.

The largest chemical shift changes can be assigned to the amino acids of the interdomain-linker (aa 102-106), e.g. N103. This is to be expected as the linker is the C- or N-terminal border for the single N-SH2$^{1-106}$ and C-SH2$^{102-220}$ constructs, respectively, so that chemical shifts in the tandem construct must be different.
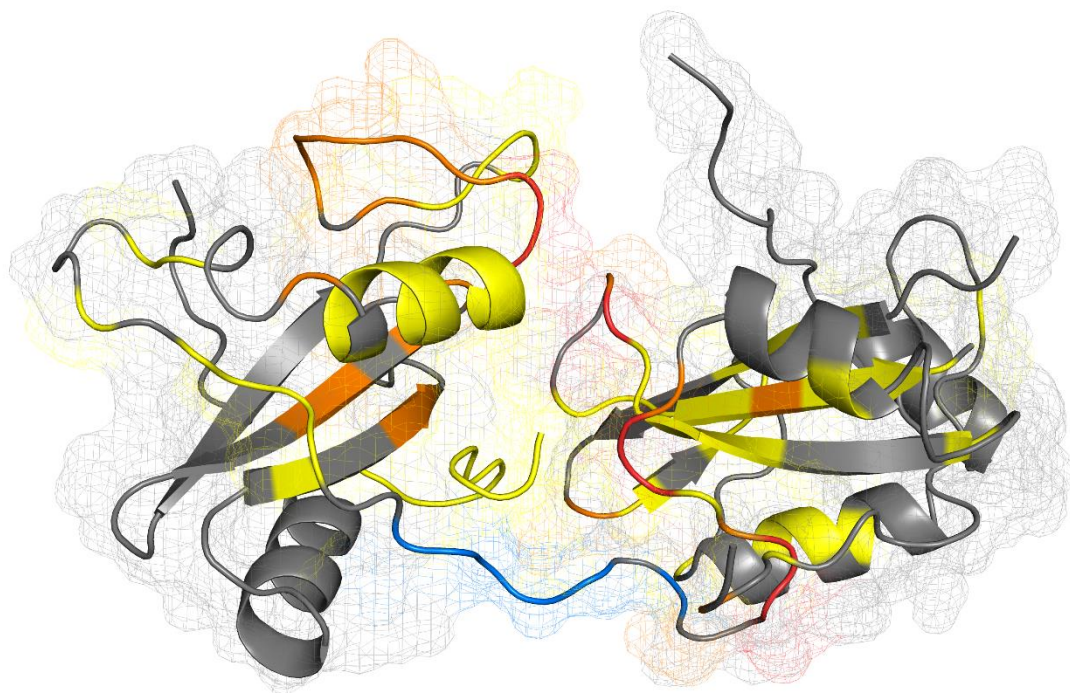
Figure 47 *Chemical shift changes plotted onto the Gab1 bound Shp2 SH2 tandem (cartoon and and mesh). Large chemical shift changes that correspond to amino acids of the linker are coloured in blue. Slight (yellow), medium (orange), and stark chemical shift perturbations (red) can be seen mainly in the interface and the β-sheets. Illustrated in PyMol.*

Slight and medium chemical shift changes can be witnessed in the β-sheet core of both domains, the EF and αB of the N-SH2 domain (**Figure 47**). Overall, the N-SH2 domain seems to experience a greater change in its chemical environment. The peaks corresponding to residues A72, T73, R111, W112, G115, H116, and Q141 (of which T73, H116 and Q141 are present in the interface of both domains, calculated by PISA; R111, W112 and G115 are located in a loop region close to the N-SH2) show stark chemical shift changes. Several slight shifts, like for F7, L74 and L177, could also be seen and lie within the newly formed interface. As seen from **Figure 47**, these amino acids are solvent exposed in all so far deposited Shp2 structures (PDB: 4dgp (Yu et al. 2013), 6crf (LaRochelle et al. 2018), 5x7b (Hayashi et al. 2017), 5x94 (Hayashi et al. 2017), 5df6 (Liu et al. 2016)). If this were the case, they should not experience such large chemical shift changes compared to the free state, where they should be solvent-exposed. Mapping the chemical shift changes to the pYpY-Gab1[617-684] bound tandem structure solved in this thesis shows these amino acids to be involved in the interface calculated by PISA or adjacent to this area. The only exceptions are R111 and W112, which are not part of the interface but are the elongation of the interdomain linker and thus involved in domain reorientation.

5.4 ITC

Formation of the interface described here should be driven by the hydrophobic effect and s should be visible in ITC measurements. Therefore, the affinity of Shp2[1-222] tandem constructs with one active and one inactive domain (N-SH2 dead: R32A, H53A; C-SH2 dead: R138A, H169A - (Hayashi et al. 2017)) to double phosphorylated pYpY-Gab1[613- 694] was measured. The tandem constructs have one native SH2 domain as well as a double-mutated second SH2 domain that should be devoid of phosphate-binding activity. The corresponding constructs are: N-SH2-dead[1-222] (with an active C-SH2 domain) and C-SH2-dead[1-222] (with an active N-SH2 domain). In a control experiment, to test whether these domains are binding-inefficient, the mutations were analysed in the single SH2 domain constructs, N-SH2-dead[1-106] and C-SH2-dead[102-220]. Neither individual dead domain showed any interaction with the pYpY-Gab1[613-694], indicating that the mutations were sufficient for blocking the binding (supplement, **Figure 67**). The chromatographic behaviour of those dead-single domains and the dead-tandem constructs was identical to the native constructs on a size exclusion column (supplemental **Figure 68**; **Figure 69**; **Figure 70**), indicating that the folding of the tandem domains was unlikely to have been affected by the mutations.

	Shp2[1-222] : Gab1[613-694]	Shp2[1-222]-N-SH2 dead : Gab1[613-94]	Shp2[1-222]-C-SH2 dead : Gab1[613-94]
N	0.98 ± 0.07	1.01 ± 0.07	1.04 ± 0.04
K_D [nM]	37 ± 13	293 ± 56	128 ± 14
- ΔS · T	9550.73 ± 477.86	11081.24 ± 506.17	11111.05 ± 393.53
ΔH	-19743.33 ± 273.41	-19970 ± 539.62	-20516.66 ± 344.51
ΔG	-10144.00 ± 180.40	-8922 ± 119.34	-9339.33 ± 51.54
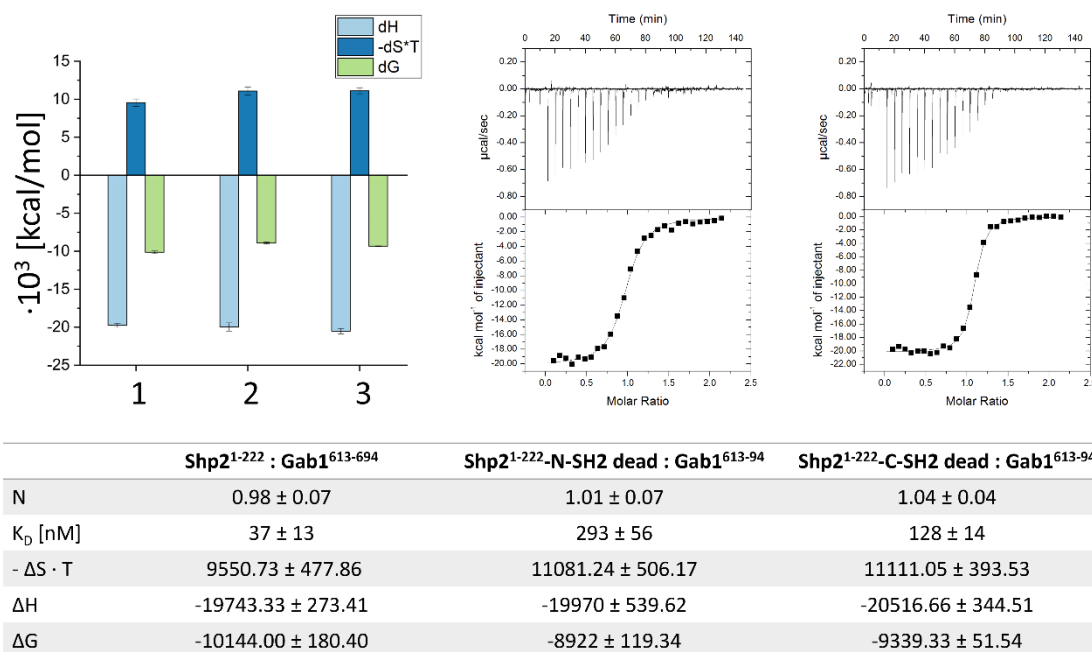
Figure 48 *ITC measurements and the derived parameters from the interaction of the Shp2[1-222]- N-SH2 dead and -C-SH2 dead constructs in reaction to the pYpY-Gab1[613-694] doubly phosphorylated peptide. Replicates of those measurements can be found in supplemetal **Figure 67**. Analysed with Origin.*

The doubly phosphorylated pYpY-Gab1[613-694] interacted with both N-SH2/C-SH2-dead-tandem constructs, although with a decreased affinity of one order of magnitude compared to the native Shp2[1-222]:pYpY-Gab1[613-694] reaction (**Figure 48**). If the interaction of the phosphorylated Gab1 sites to the single remaining active SH2 domain binding

region of each construct would be the only factor determining the interaction, the affinities should be similar to the ones of the single SH2 domains to pYpY-Gab1[613-694] (compare chapter 2 **Figure 20**). This is not the case. Instead, the thermodynamic contributions of enthalpy and entropy are similar to the Shp2[1-222]:pYpY-Gab1[613-694], indicating that although one domain of each tandem construct was unable to bind the phosphorylated site of pYpY-Gab1[613-694], the interaction showed a similar thermodynamic profile with a reduction in affinity.

5.5 Discussion

Knowledge of the conformational landscape of Shp2 – particularly in its active and inactive forms – is essential for studying the system. The structures published so far vary in the position of their SH2 domain orientation upon activation and must thus be considered with caution. Due to its large size, Shp2[1-593] is not readily accessible to NMR studies, so that structures of the complete molecule have relied upon crystallographic methods. The structure of Shp2 possessing a constitutively active oncogenic mutation E76K revealed an open conformation, with the mutant found in a novel N-SH2:PTP interface (PDB:6crf (LaRochelle et al. 2018)). However, the use of a C-terminally truncated Shp2 (1-525) construct, which exhibits a shortened C-terminal helix and ensuing intrinsically disordered region, allows the N-SH2 domain to adopt positions that are not necessarily native. Interestingly, an open conformation has been reported for the related enzyme Shp1 (PDB: 3ps5 (Wang et al. 2011)) in full-length Shp1[1-595] that shows the corresponding N- and C-SH2 domains in very different positions. It is not clear, however, why they obtained an open conformation – the presence of sulphate ions in the two SH2 pY binding sites suggests that the high ammonium sulphate concentrations and/or detergents used for crystallisation might have been responsible.

Since the crystallisation of full-length Shp2 in the presence of activating peptides has not been successful, a divide-and-conquer approach can be used to gain insights into the activated structure. As in this thesis, several studies of the tandem SH2 domains in complex with activating peptides have been published. For the three cases in which this approach has been followed (PDB: 5df6 (Liu et al. 2016), 5x7b (Hayashi et al. 2017), 5x94 (Hayashi et al. 2017)7b), three different relative orientations of the SH2-domains were observed, presumably due to the use of short, unlinked activator peptides, providing the system with additional flexibility.

In the crystal structure of the complex of Shp2[1-222]:pYpYGab1[617-684] determined via microED, the overall orientation of the SH2 domains differs from all other Shp2 structures deposited so far. On the other hand, it aligns well with the activated Shp1 structure, with the same hydrophobic interface between the N-SH2 and C-SH2 domains. The

hydrophobic interface is conserved in many SH2 domains and might be a general interaction platform. In the closed conformation of Shp2, N-SH2 residues from this surface are involved in the N-SH2:PTP interface, contributing also to the inactive form. In the closed conformation of ZAP70 (PDB:2ozo (Deindl et al. 2007)), these residues are in the proximity of the kinase domain, although not completely buried in the interface. These conserved hydrophobic amino acids on the surface of the SH2 domains are located at the back side of the domain and not at the peptide interacting site, except for the pY-binding pocket (P144 C-SH2 equivalent), which provides the domains with their individual selectivity for specific peptides through the EF- and BG-loop at the front of the domain.

The chemical shift changes of the pYpY-Gab1[613-694] bound tandem Shp2[1-222] compared to the single bound SH2 domains presented here further support the crystal structure, as residues showing large changes in their chemical environment become located in the interface between the N-SH2 and C-SH2 domain, while for all other structures published of a potentially activated Shp2, these residues face the solvent and would thus not experience those great chemical changes upon binding resulting in the CSPs measured here. A similar experiment conducted in the presence of the double phosphorylated activating protein PD-1 (Marasco et al. 2020a) records chemical shift changes that show an interface of the N-SH2 and C-SH2 when both phosphotyrosine motifs are bound. Although the CSPs are not further specified, the authors state that they located to the SH2-domain interface seen in the Shp1 open (PDB: 3ps5 (Wang et al. 2011)) and the Shp2 closed (PDB: 2Shp (Hof et al. 1998)) structure. A paramagnetic NMR study of lanthanide labelled unbound tandem domains of Shp2 revealed two conformations in solution (Julia Amber Tartaglia 2019): A larger population of around 60% resembling the closed conformation and a smaller second population with an unknown orientation of the two domains. The specific residues are again not stated, making a comparison and analysis with the here-found microED structure only vaguely possible. Nevertheless, the resulting novel conformation (of 40% of the population) shows similarity in the orientation to the Gab1 bound Shp2 structure. These results show that without the PTP domain, the tandem SH2 domains are in an equilibrium between the closed and an open conformation.

According to Pádua et al. 2018 the open Shp1 and Shp2 structures differ in their conformation. Evaluation of chemical shift perturbations of the N-SH2 domain in the tandem construct or truncated (aa 1-529) constitutively active mutant E76K indicates that only amino acids located at the top of the αB helix show differences between the "free" N-SH2 in the tandem and the PTP-interacting construct. These amino acids are involved

in the N-SH2:PTP interface in the E76K-Shp2 open state (PDB:6crf (LaRochelle et al. 2018)). The truncated E76K mutated active conformation of Shp2 (aa 1-529) seems identical in solution as in the crystallographic environment. In the same analysis, a crystal structure of ΔN-SH2 Shp2 (aa 105-529) (PDB:6cmq (Pádua et al. 2018)) showed slight variations in the C-SH2 domain in the four molecules in the asymmetric unit, indicating an ensemble of conformations. The open Shp2 (PDB: 6crf (LaRochelle et al. 2018)) and open Shp1 (PDB:3ps5 (Wang et al. 2011)) structures align with the ΔN-SH2 structure. This indicates that both might be eligible active forms of Shp, although one has to keep in mind that the main difference is a true full-length WT construct (Shp1; PDB: 3ps5 (Wang et al. 2011)), while the other is a truncated and mutated variant (Shp2; PDB: 6crf (LaRochelle et al. 2018))

The use of a C-terminally truncated Shp2[1-525] structure (PDB:6crf (LaRochelle et al. 2018)) is problematic. With AlphaFold, the truncated residues are predicted to form a long C-terminal helix followed by a disordered segment, similar to the full-length Shp1[1-595] structure. An elongated helix would clash with the N-SH2 domain found in open E76K structure(s). The lysine side chain of E76K is in a negatively charged region involving amino acids E225, E227 and E523, which would clearly be unfavourable in the presence of wild-type E76; this favourable environment for the lysyl side chain is additionally stabilised by the L525 carboxylate introduced via the truncation. In addition, the mutation E76K, is proposed to lead to the release of the N-SH2 from the PTP domain by altering the hydrogen bond network to the PTP domain (Hou et al. 2023). Thus, there are reasons to believe that the combination of truncation and mutation might lead to a different open structure to that experienced by the wild-type enzyme.

ITC

The ITC measurements show a significant unfavourable entropic and favourable enthalpic contribution upon binding of pYpY-Gab1[613-694] to WT-Shp2[1-222] as well as to the SH2-dead tandem constructs, compared to the single domain binding to the bis-phosphorylated Gab1 (**Figure 20**). To derive an accurate theory from those data is, of course, not possible as energetics are complex. The binding to two domains instead of one results in a higher number of newly formed bonds, which also leads to a release of solvent molecules, favouring both enthalpy and entropy. Entropically unfavoured processes, as seen from the here presented measurements of bis-phosphorylated Gab1 with tandem SH2 domains, but not – to that extent- for the single domains is puzzling. Of course, this could be due to the helix formation of Gab1 upon binding. In that case, the binding of the single C-SH2 domain should result in a similar energetic profile, as NMR data (chapter 2) demonstrate helicity upon binding, while the interaction with the

C-SH2-dead tandem should not cause such reduction in entropy. Therefore, a depletion in the systems flexibility upon binding could be the origin and could arise from the reduction of Gab1 mobility upon binding to two sites or from the newly formed patch between both SH2 domains upon binding. However, as Gab1 is still highly flexible outside the binding sites, and this effect is also seen in the SH2-dead tandem measurements, it is most likely an effect derived from the restriction in motion of the SH2 domains upon binding to either or both sites of Gab1.

Taken together from the data presented here and from the structure of the open Shp1 conformation (PDB: 3ps5 (Wang et al. 2011), an activation model for full-length Shp2 upon binding of its activator proteins can be assumed. The binding results in movement of the N-SH2 domain away from the PTP domain and the open conformation is stabilised by interactions of the N-SH2 domain and the C-SH2 domain, which was unreported so far and could be a potential binding site target for drug development, specific to the activated Shp phosphatase. In the closed conformation, both domains interact mostly through the linker, which is seen in both homologues. Therefore, a similar structure between both in the activated form would not be surprising, especially since Shp1 and Shp2 show an almost identical interface, which is conserved (Wang et al. 2011).

For further validation of this interface, activation assays of interface-mutated Shp1 and Shp2 could be facilitated as well as further structural analysis of the full-length Shp2 in the activated form.

Summary

The presented and discussed results help to answer the following questions with partly unexpected outcomes. This project is the first to present a structure of the Shp2 tandem SH2 domains bound to a bis-phosphorylated, intrinsically disordered activator. Furthermore, interaction studies of this native 84 aa long activator were conducted instead of using short mono-phosphorylated peptides (commonly utilised), which are connected with an artificial PEG-linker.

I) How does the phosphatase Shp2 bind to the doubly phosphorylated Gab1 tail?

The results addressing this question can be found in chapters 2, 3, 4 and 5.

In principle, a tandem SH2 domain construct could bind a double phosphorylated target in a sequential or parallel mode. Additionally, there could be a preference from one SH2 domain to one specific phosphotyrosine or an unpreferred mixture of interactions. ITC experiments showed that both single domains bind with similar affinity to the doubly phosphorylated peptide. NMR experiments of the labelled bis-phosphorylated Gab1 with either of the domains described the exact binding site for both domains. These data show the preference for the N-SH2 domain to pY627 and C-SH2 to pY659. The orientation of the interaction was further validated from the crystal structure of $Shp2^{1-222}$:pYpY- $Gab1^{617-684}$, solved by electron diffraction experiments. The structural details of the interaction were elucidated from the crystal structure of the N- $SH2^{1-106}$:pY- $Gab1^{613-651}$ and the above-mentioned crystal. A parallel binding mode might still be possible since both interactions are similar in affinity. It is often discussed that the C-SH2 domain initiates the binding, while the binding to the N-SH2 is most important for the activation of Shp2. This would be one unexpected finding, which might be specific to the interaction of Shp2 and Gab1. Further research to validate this finding should be done.

II) Does the intrinsically disordered tail of Gab1 adopt any structure prior to or upon binding to the SH2 domains of Shp2?

The results addressing this question can be found in chapters 1, 2, 3 and 4.

It was demonstrated that the Gab1 fragment studied in this project is intrinsically disordered in solution independent of its phosphorylation status, shown by NMR and CD. Interestingly, the predicted α-helix C-terminally to the pY659 binding site could be experimentally proven in a crystal structure on the $Shp2^{1-222}$:pYpY$Gab1^{617-684}$ complex, solved from electron diffraction data. A peptide corresponding to this binding site with a C-terminal elongation encompassing the length of the helix did not show an α- helical CD spectrum in solution. This indicates that the disorder-to-order transition is happening

during or after the binding. Further NMR and ITC experiments showed that the pY659 binding site and the helix-forming region must be covalently linked; an allosteric interaction is thereby not witnessed. Besides the newly induced helix upon binding, the pY-binding sites adopt an extended conformation in the SH2 domains' binding grove. This was seen in both crystal structures. The remainder of the Gab1 peptide is expected to stay disordered when bound to the tandem SH2 domains. This is due to the missing coulomb potential for those regions in the microED structure. Still, further experiments can be done to structurally analyse the complex in solution and with a full-length Shp2 protein as a binding partner.

III) Does the phosphatase Shp2 undergo any structural changes upon binding to Gab1?

The results addressing this question can be found in chapters 2, 3, 4 and 5.

It is expected and known from the literature that Shp2 undergoes a conformational change upon binding to a double phosphorylated activator, thereby moving the N-SH2 domain from its PTP-active site blocking position, and thus activating the phosphatase. As this thesis focussed on the SH2 tandem interaction with a doubly phosphorylated activator, it was not expected to add to these findings. Nevertheless, a novel interface of the N-SH2 and C-SH2 domain in the Gab1-bound state was found by mciroED, NMR and ITC. This state most likely represents an activated, open conformation of Shp2. The same orientation could be found in a previously published open Shp1 structure as well as in an AlphaFold predicted Gab1:Shp2 complex. Nevertheless, this is in contrast with the currently assumed open Shp2 conformation. Whether the here presented novel conformation represents one of many open conformations or is actually challenging the current structure must be the subject of further research.

Additionally, this thesis investigated the dynamics of the domains in the free and Gab1-bound states, showing that their flexibility slightly increases upon binding.

Outlook

To address still open questions of the interaction between Shp2 and Gab1, further research must focus on the generation of NMR solution structure data of the tandem SH2 domain in complex with doubly phosphorylated Gab1. This could help to validate that an artificial crystal environment does not induce the structure solved by microED. Furthermore, mutations and their effect on the structure could be easier to follow, allowing helpful insights into regulatory and, thus, druggable sites of the complex. Since the conserved residues of the interface located in the N-SH2 domain are all involved in the PTP:N-SH2 interface of the closed conformation, only P144 of the C-SH2 would be a useful mutation site. However, P144 is located in the phosphotyrosine-binding loop (CD-loop) of the C-SH2 domain. Nevertheless, initial testing demonstrated that mutations did not influence the binding of the domain to phosphorylated Gab1. Further binding and activity studies with those mutations in the tandem and full-length Shp2 will provide functional proof for that interface. Further, testing helix-truncated Gab1 constructs in those assays might shine a light on the regulatory importance of the disorder-to-order formation of the IDP, beyond the higher affinity. Structural data of the full-length Shp2 and Gab1 by Cryo-EM might help address how the disordered tail interacts with the phosphatase and further provide insights into the structure of tadpole-like IDPs (N-terminal folding hypothesis, SUMOylation sites of Shp2 and Sumo binding sites on Gab1; see Chapter "Differences between Shp1 and Shp2"). To validate whether the findings of this thesis are specific to the Gab1:Shp2 interaction or are also applicable to the other Gab family members or even to other native activators of Shp2, further structural and interaction data are needed. These interaction data could also be extended to kinetic studies to address the binding event in a time-resolved manner.

References

Agazie, Yehenew M.; Hayman, Michael J. (2003): Molecular mechanism for a role of SHP2 in epidermal growth factor receptor signaling. In *Molecular and cellular biology* 23 (21), pp. 7875–7886. DOI: 10.1128/MCB.23.21.7875-7886.2003.

Alberts, Bruce; Johnson, Alexander; Lewis, Julian; Raff, Martin; Roberts, Keith; Walter, Peter (2002): Molecular Biology of the Cell. 4th edition. Signaling through Enzyme-Linked Cell-Surface Receptors. New York: Garland Science.

Alterovitz, Wei-Lun; Faraggi, Eshel; Oldfield, Christopher J.; Meng, Jingwei; Xue, Bin; Huang, Fei et al. (2020): Many-to-one binding by intrinsically disordered protein regions. In *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* 25, pp. 159–170.

Anselmi, Massimiliano; Calligari, Paolo; Hub, Jochen S.; Tartaglia, Marco; Bocchinfuso, Gianfranco; Stella, Lorenzo (2020): Structural Determinants of Phosphopeptide Binding to the N-Terminal Src Homology 2 Domain of the SHP2 Phosphatase. In *Journal of chemical information and modeling* 60 (6), pp. 3157–3171. DOI: 10.1021/acs.jcim.0c00307.

Anselmi, Massimiliano; Hub, Jochen S. (2020): An allosteric interaction controls the activation mechanism of SHP2 tyrosine phosphatase. In *Scientific reports* 10 (1), p. 18530. DOI: 10.1038/s41598-020-75409-7.

Anselmi, Massimiliano; Hub, Jochen S. (2021): The loops of the N-SH2 binding cleft do not serve as allosteric switch in SHP2 activation. In *Proceedings of the National Academy of Sciences of the United States of America* 118 (17). DOI: 10.1073/pnas.2025107118.

Bard-Chapeau, Emilie A.; Li, Shuangwei; Ding, Jin; Zhang, Sharon S.; Zhu, Helen H.; Princen, Frederic et al. (2011): Ptpn11/Shp2 acts as a tumor suppressor in hepatocellular carcinogenesis. In *Cancer cell* 19 (5), pp. 629–639. DOI: 10.1016/j.ccr.2011.03.023.

Bardelli, A.; Longati, P.; Gramaglia, D.; Stella, M. C.; Comoglio, P. M. (1997): Gab1 coupling to the HGF/Met receptor multifunctional docking site requires binding of Grb2 and correlates with the transforming potential. In *Oncogene* 15 (25), pp. 3103–3111. DOI: 10.1038/sj.onc.1201561.

Beebe, K. D.; Wang, P.; Arabaci, G.; Pei, D. (2000): Determination of the binding specificity of the SH2 domains of protein tyrosine phosphatase SHP-1 through the screening of a combinatorial phosphotyrosyl peptide library. In *Biochemistry* 39 (43), pp. 13251–13260. DOI: 10.1021/bi0014397.

Berkel, Caglar; Cacan, Ercan (2021): GAB2 and GAB3 are expressed in a tumor stage-, grade- and histotype-dependent manner and are associated with shorter progression-free survival in ovarian cancer. In *Journal of cell communication and signaling* 15 (1), pp. 57–70. DOI: 10.1007/s12079-020-00582-3.

Bertoline, Letícia M. F.; Lima, Angélica N.; Krieger, Jose E.; Teixeira, Samantha K. (2023): Before and after AlphaFold2: An overview of protein structure prediction. In *Frontiers in bioinformatics* 3, p. 1120370. DOI: 10.3389/fbinf.2023.1120370.

Bobone, Sara; Pannone, Luca; Biondi, Barbara; Solman, Maja; Flex, Elisabetta; Canale, Viviana Claudia et al. (2021): Targeting Oncogenic Src Homology 2 Domain-Containing Phosphatase 2 (SHP2) by Inhibiting Its Protein-Protein Interactions. In *Journal of medicinal chemistry* 64 (21), pp. 15973–15990. DOI: 10.1021/acs.jmedchem.1c01371.

Bonetti, Daniela; Troilo, Francesca; Toto, Angelo; Travaglini-Allocatelli, Carlo; Brunori, Maurizio; Gianni, Stefano (2018): Mechanism of Folding and Binding of the N-Terminal SH2 Domain from SHP2. In *J. Phys. Chem. B* 122 (49), pp. 11108–11114. DOI: 10.1021/acs.jpcb.8b05651.

Bradshaw, J. M.; Mitaxov, V.; Waksman, G. (1999): Investigation of phosphotyrosine recognition by the SH2 domain of the Src kinase. In *Journal of molecular biology* 293 (4), pp. 971–985. DOI: 10.1006/jmbi.1999.3190.

Breithaupt, Constanze; Gruber, Tobias; Mandel, Katharina; Lewitzky, Marc; Meister, Annette; Jochen, Balbach et al. (2023): Self-assembly of Grb2 meshworks revealed by Grb2-Gab1 497-528 complex structure. DOI: 10.1101/2023.06.17.545433.

Bunda, Severa; Burrell, Kelly; Heir, Pardeep; Zeng, Lifan; Alamsahebpour, Amir; Kano, Yoshihito et al. (2015): Inhibition of SHP2-mediated dephosphorylation of Ras suppresses oncogenesis. In *Nature communications* 6, p. 8859. DOI: 10.1038/ncomms9859.

Chantley, Lewis C. (2002): The Phosphoinositide 3-Kinase Pathway. In *Science (New York, N.Y.)* 296 (5573), pp. 1655–1657. DOI: 10.1126/science.296.5573.165.

Chayen, Naomi E. (2005): Methods for separating nucleation and growth in protein crystallisation. In *Progress in biophysics and molecular biology* 88 (3), pp. 329–337. DOI: 10.1016/j.pbiomolbio.2004.07.007.

Chen, Lu; Du-Cuny, Lei; Moses, Sylvestor; Dumas, Sabrina; Song, Zuohe; Rezaeian, Abdol Hossein et al. (2015): Novel inhibitors induce large conformational changes of GAB1 pleckstrin homology domain and kill breast cancer cells. In *PLoS computational biology* 11 (1), e1004021. DOI: 10.1371/journal.pcbi.1004021.

Chen, Xin; Fu, Xueqi; Zhao, Wanke; Ho, Wan-Ting Tina; Xing, Shu; Zhao, Zhizhuang Joe (2020): Loss of tyrosine phosphatase SHP2 activity promotes growth of colorectal carcinoma HCT-116 cells. In *Signal transduction and targeted therapy* 5 (1), p. 83. DOI: 10.1038/s41392-020-0192-0.

Chen, Ying-Nan P.; LaMarche, Matthew J.; Chan, Ho Man; Fekkes, Peter; Garcia-Fortanet, Jorge; Acker, Michael G. et al. (2016): Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases. In *Nature* 535 (7610), pp. 148–152. DOI: 10.1038/nature18621.

Chen, Yuhong; Wen, Renren; Yang, Shoua; Schuman, James; Zhang, Eric E.; Yi, Taolin et al. (2003): Identification of Shp-2 as a Stat5A phosphatase. In *The Journal of biological chemistry* 278 (19), pp. 16520–16527. DOI: 10.1074/jbc.M210572200.

Cheng, Hong; Schwell, Vered; Curtis, Brett R.; Fazlieva, Ruzaliya; Roder, Heinrich; Campbell, Kerry S. (2019): Conformational Changes in the Cytoplasmic Region of KIR3DL1 upon Interaction with SHP-2. In *Structure (London, England : 1993)* 27 (4), 639-650.e2. DOI: 10.1016/j.str.2019.01.009.

Clabbers, Max T. B.; Shiriaeva, Anna; Gonen, Tamir (2022): MicroED: conception, practice and future opportunities. In *IUCrJ* 9 (Pt 2), pp. 169–179. DOI: 10.1107/S2052252521013063.

Craggs, G.; Kellie, S. (2001): A functional nuclear localization sequence in the C-terminal domain of SHP-1. In *The Journal of biological chemistry* 276 (26), pp. 23719–23725. DOI: 10.1074/jbc.M102846200.

Cunnick, J. M.; Mei, L.; Doupnik, C. A.; Wu, J. (2001): Phosphotyrosines 627 and 659 of Gab1 constitute a bisphosphoryl tyrosine-based activation motif (BTAM) conferring binding and activation of SHP2. In *The Journal of biological chemistry* 276 (26), pp. 24380–24387. DOI: 10.1074/jbc.M010275200.

Dai, Kun; Liao, Shanhui; Zhang, Jiahai; Zhang, Xuecheng; Tu, Xiaoming (2011): Solution structure of tensin2 SH2 domain and its phosphotyrosine-independent interaction with DLC-1. In *PloS one* 6 (7), e21965. DOI: 10.1371/journal.pone.0021965.

Daly, Roger J.; Gu, Haihua; Parmar, Jayamala; Malaney, Suzann; Lyons, Ruth J.; Kairouz, Rania et al. (2002): The docking protein Gab2 is overexpressed and estrogen regulated in human breast cancer. In *Oncogene* 21 (33), pp. 5175–5181. DOI: 10.1038/sj.onc.1205522.

Dass, Rupashree; Mulder, Frans A. A.; Nielsen, Jakob Toudahl (2020): ODiNPred: comprehensive prediction of protein order and disorder. In *Scientific reports* 10 (1), p. 14780. DOI: 10.1038/s41598-020-71716-1.

Deindl, Sebastian; Kadlecek, Theresa A.; Brdicka, Tomas; Cao, Xiaoxian; Weiss, Arthur; Kuriyan, John (2007): Structural basis for the inhibition of tyrosine kinase activity of ZAP-70. In *Cell* 129 (4), pp. 735–746. DOI: 10.1016/j.cell.2007.03.039.

Deng, Rong; Zhao, Xian; Qu, YingYing; Chen, Cheng; Zhu, Changhong; Zhang, Hailong et al. (2015): Shp2 SUMOylation promotes ERK activation and hepatocellular carcinoma development. In *Oncotarget* 6 (11), pp. 9355–9369. DOI: 10.18632/oncotarget.3323.

Derewenda, Zygmunt S.; Godzik, Adam (2017): The "Sticky Patch" Model of Crystallization and Modification of Proteins for Enhanced Crystallizability. In *Methods in molecular biology (Clifton, N.J.)* 1607, pp. 77–115. DOI: 10.1007/978-1-4939-7000-1_4.

Diop, Awa; Santorelli, Daniele; Malagrinò, Francesca; Nardella, Caterina; Pennacchietti, Valeria; Pagano, Livia et al. (2022): SH2 Domains: Folding, Binding and Therapeutical Approaches. In *International journal of molecular sciences* 23 (24). DOI: 10.3390/ijms232415944.

Dunker, A. Keith; Bondos, Sarah E.; Huang, Fei; Oldfield, Christopher J. (2015): Intrinsically disordered proteins and multicellular organisms. In *Seminars in cell & developmental biology* 37, pp. 44–55. DOI: 10.1016/j.semcdb.2014.09.025.

Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. (2010): Features and development of Coot. In *Acta crystallographica. Section D, Biological crystallography* 66 (Pt 4), pp. 486–501. DOI: 10.1107/S0907444910007493.

Eulenfeld, René; Schaper, Fred (2009): A new mechanism for the regulation of Gab1 recruitment to the plasma membrane. In *Journal of cell science* 122 (Pt 1), pp. 55–64. DOI: 10.1242/jcs.037226.

Evans, Richard; O'Neill, Michael; Pritzel, Alexander; Antropova, Natasha; Senior, Andrew; Green, Tim et al. (2021): Protein complex prediction with AlphaFold-Multimer. DOI: 10.1101/2021.10.04.463034.

Felli, Isabella C.; Pierattelli, Roberta (2022): 13C Direct Detected NMR for Challenging Systems. In *Chemical reviews* 122 (10), pp. 9468–9496. DOI: 10.1021/acs.chemrev.1c00871.

Feng, Chao; Roy, Amitava; Post, Carol Beth (2018): Entropic allostery dominates the phosphorylation-dependent regulation of Syk tyrosine kinase release from immunoreceptor tyrosine-based activation motifs. In *Protein science : a publication of the Protein Society* 27 (10), pp. 1780–1796. DOI: 10.1002/pro.3489.

Fodor, Michelle; Price, Edmund; Wang, Ping; Lu, Hengyu; Argintaru, Andreea; Chen, Zhouliang et al. (2018): Dual Allosteric Inhibition of SHP2 Phosphatase. In *ACS chemical biology* 13 (3), pp. 647–656. DOI: 10.1021/acschembio.7b00980.

Fontaine, Benjamin M.; Duggal, Yashasvika; Weinert, Emily E. (2018): Exploring the Links between Nucleotide Signaling and Quorum Sensing Pathways in Regulating Bacterial Virulence. In *ACS infectious diseases* 4 (12), pp. 1645–1655. DOI: 10.1021/acsinfecdis.8b00255.

Furcht, Christopher M.; Buonato, Janine M.; Lazzara, Matthew J. (2015): EGFR-activated Src family kinases maintain GAB1-SHP2 complexes distal from EGFR. In *Science signaling* 8 (376), ra46. DOI: 10.1126/scisignal.2005697.

Gasteiger, Elisabeth; Hoogland, Christine; Gattiker, Alexandre; Duvaud, S'everine; Wilkins, Marc R.; Appel, Ron D.; Bairoch, Amos: Protein Identification and Analysis Tools on the ExPASy Server, pp. 571–607. DOI: 10.1385/1-59259-890-0:571.

Goyette, Jesse; Depoil, David; Yang, Zhengmin; Isaacson, Samuel A.; Allard, Jun; van der Merwe, P. Anton et al. (2022): Dephosphorylation accelerates the dissociation of ZAP70 from the T cell receptor. In *Proceedings of the National Academy of Sciences of the United States of America* 119 (9). DOI: 10.1073/pnas.2116815119.

Gruber, Tobias (2014 // 2015): Biophysikalische Charakterisierung der Proteinfaltung und Lipidinteraction der humanen AmphiphysinII/Bin1 N-BAR-Domäne // Biophysikalische Charakterisierung der Proteinfaltung und Lipidinteraktion der humanen AmphiphysinII/Bin1 N-BAR-Domäne. Dissertation. Martin-Luther Universität Halle-Wittenberg, Halle/Saale, Germany.

Gruber, Tobias; Lewitzky, Marc; Machner, Lisa; Weininger, Ulrich; Feller, Stephan M.; Balbach, Jochen (2022): Macromolecular Crowding Induces a Binding Competent Transient Structure in Intrinsically Disordered Gab1. In *Journal of molecular biology* 434 (5), p. 167407. DOI: 10.1016/j.jmb.2021.167407.

Gu, H.; Pratt, J. C.; Burakoff, S. J.; Neel, B. G. (1998): Cloning of p97/Gab2, the major SHP2-binding protein in hematopoietic cells, reveals a novel pathway for cytokine-induced gene activation. In *Molecular cell* 2 (6), pp. 729–740. DOI: 10.1016/S1097-2765(00)80288-9.

Han, Jing-Dong J.; Bertin, Nicolas; Hao, Tong; Goldberg, Debra S.; Berriz, Gabriel F.; Zhang, Lan V. et al. (2004): Evidence for dynamically organized modularity in the yeast protein-protein interaction network. In *Nature* 430 (6995), pp. 88–93. DOI: 10.1038/nature02555.

Hanafusa, Hiroshi; Torii, Satoru; Yasunaga, Takayuki; Matsumoto, Kunihiro; Nishida, Eisuke (2004): Shp2, an SH2-containing protein-tyrosine phosphatase, positively regulates receptor tyrosine kinase signaling by dephosphorylating and inactivating the inhibitor Sprouty. In *The Journal of biological chemistry* 279 (22), pp. 22992–22995. DOI: 10.1074/jbc.M312498200.

Harkiolaki, Maria; Tsirka, Theodora; Lewitzky, Marc; Simister, Philip C.; Joshi, Dhira; Bird, Louise E. et al. (2009): Distinct binding modes of two epitopes in Gab2 that interact with the SH3C domain of Grb2. In *Structure (London, England : 1993)* 17 (6), pp. 809–822. DOI: 10.1016/j.str.2009.03.017.

Hartman, Zachary; Geldenhuys, Werner J.; Agazie, Yehenew M. (2020): A specific amino acid context in EGFR and HER2 phosphorylation sites enables selective binding to the active site of Src homology phosphatase 2 (SHP2). In *The Journal of biological chemistry* 295 (11), pp. 3563–3575. DOI: 10.1074/jbc.RA119.011422.

Hayashi, Takeru; Senda, Miki; Suzuki, Nobuhiro; Nishikawa, Hiroko; Ben, Chi; Tang, Chao et al. (2017): Differential Mechanisms for SHP2 Binding and Activation Are Exploited by Geographically Distinct Helicobacter pylori CagA Oncoproteins. In *Cell reports* 20 (12), pp. 2876–2890. DOI: 10.1016/j.celrep.2017.08.080.

Hendus-Altenburger, Ruth; Fernandes, Catarina B.; Bugge, Katrine; Kunze, Micha B. A.; Boomsma, Wouter; Kragelund, Birthe B. (2019): Random coil chemical shifts for serine, threonine and tyrosine phosphorylation over a broad pH range. In *Journal of biomolecular NMR* 73 (12), pp. 713–725. DOI: 10.1007/s10858-019-00283-z.

Herbst, R.; Carroll, P. M.; Allard, J. D.; Schilling, J.; Raabe, T.; Simon, M. A. (1996): Daughter of sevenless is a substrate of the phosphotyrosine phosphatase Corkscrew and functions during sevenless signaling. In *Cell* 85 (6), pp. 899–909. DOI: 10.1016/S0092-8674(00)81273-8.

Hille, Bertil (2001): Ion channels of exciteable membranes. 3rd edition. Sunderland: Sinauer Associates, Inc.

Hoeben, A.; Martin, D.; Clement, P. M.; Cools, J.; Gutkind, J. S. (2013): Role of GRB2-associated binder 1 in epidermal growth factor receptor-induced signaling in head and neck squamous cell carcinoma. In *International journal of cancer* 132 (5), pp. 1042–1050. DOI: 10.1002/ijc.27763.

Hof, P.; Pluskey, S.; Dhe-Paganon, S.; Eck, M. J.; Shoelson, S. E. (1998): Crystal structure of the tyrosine phosphatase SHP-2. In *Cell* 92 (4), pp. 441–450. DOI: 10.1016/s0092-8674(00)80938-1.

Holgado-Madruga, M.; Emlet, D. R.; Moscatello, D. K.; Godwin, A. K.; Wong, A. J. (1996): A Grb2-associated docking protein in EGF- and insulin-receptor signalling. In *Nature* 379 (6565), pp. 560–564. DOI: 10.1038/379560a0.

Hou, Yingnan; Lu, Xiaoli; Xu, Ziyao; Qu, Jiarun; Huang, Jing (2023): How a single mutation alters the protein structure: a simulation investigation on protein tyrosine phosphatase SHP2. In *RSC advances* 13 (7), pp. 4263–4274. DOI: 10.1039/d2ra07472a.

Hsu, Claire C.; Buehler, Markus J.; Tarakanova, Anna (2020): The Order-Disorder Continuum: Linking Predictions of Protein Structure and Disorder through Molecular Simulation. In *Scientific reports* 10 (1), p. 2068. DOI: 10.1038/s41598-020-58868-w.

Huang, Xin; Gollin, Susanne M.; Raja, Siva; Godfrey, Tony E. (2002): High-resolution mapping of the 11q13 amplicon and identification of a gene, TAOS1, that is amplified and overexpressed in oral cancer cells. In *Proceedings of the National Academy of Sciences of the United States of America* 99 (17), pp. 11369–11374. DOI: 10.1073/pnas.172285799.

Imhof, Diana; Wavreille, Anne-Sophie; May, Andreas; Zacharias, Martin; Tridandapani, Susheela; Pei, Dehua (2006): Sequence Specificity of SHP-1 and SHP-2 Src Homology 2 Domains // Sequence specificity of SHP-1 and SHP-2 Src homology 2 domains. Critical roles of residues beyond the pY+3 position. In *The Journal of biological chemistry* 281 (29), pp. 20271–20282. DOI: 10.1074/jbc.M601047200.

Inaba, Satomi; Numoto, Nobutaka; Ogawa, Shuhei; Morii, Hisayuki; Ikura, Teikichi; Abe, Ryo et al. (2017): Crystal Structures and Thermodynamic Analysis Reveal Distinct Mechanisms of CD28 Phosphopeptide Binding to the Src Homology 2 (SH2) Domains of Three Adaptor Proteins. In *The Journal of biological chemistry* 292 (3), pp. 1052–1060. DOI: 10.1074/jbc.M116.755173.

Itoh, M.; Yoshida, Y.; Nishida, K.; Narimatsu, M.; Hibi, M.; Hirano, T. (2000): Role of Gab1 in heart, placenta, and skin development and growth factor- and cytokine-induced extracellular signal-regulated kinase mitogen-activated protein kinase activation. In *Molecular and cellular biology* 20 (10), pp. 3695–3704. DOI: 10.1128/mcb.20.10.3695-3704.2000.

Jiang, Wei; Ji, Meiju (2019): Receptor tyrosine kinases in PI3K signaling: The therapeutic targets in cancer. In *Seminars in cancer biology* 59, pp. 3–22. DOI: 10.1016/j.semcancer.2019.03.006.

Johnson, Bruce A. (2004): Using NMRView to visualize and analyze the NMR spectra of macromolecules. In *Methods in molecular biology (Clifton, N.J.)* 278, pp. 313–352. DOI: 10.1385/1-59259-809-9:313.

Jones, D. T. (1999): Protein secondary structure prediction based on position-specific scoring matrices. In *Journal of molecular biology* 292 (2), pp. 195–202. DOI: 10.1006/jmbi.1999.3091.

Jones, Matthew L.; Craik, Johnathan D.; Gibbins, Jonathan M.; Poole, Alastair W. (2004): Regulation of SHP-1 tyrosine phosphatase in human platelets by serine phosphorylation at its C terminus. In *The Journal of biological chemistry* 279 (39), pp. 40475–40483. DOI: 10.1074/jbc.M402970200.

Judith Kniest (2021): Untersuchung der Komplexbildung phosphorylierter, C-terminaler Gab1-Peptidfragmente mit SH2-Domänen der Phosphatase Shp2. Abschlussarbeit zur Erlangung des akademischen Grades. Martin-Luther Universität Halle-Wittenberg, Halle/Saale, Germany. Institut für Biochemie und Biotechnologie.

Julia Amber Tartaglia (2019): Structural Dynamics of SHP2 Activation utilizing IRS-1 and Paramagnetic Enhanced NMR Spectroscopy. Master's thesis. Brandeis University, Waltham, Massachusetts. The Faculty of the Graduate School of Arts and Sciences; Department of Biochemistry. Available online at https://hdl.handle.net/10192/36778.

Jumper, John; Evans, Richard; Pritzel, Alexander; Green, Tim; Figurnov, Michael; Ronneberger, Olaf et al. (2021): Highly accurate protein structure prediction with AlphaFold. In *Nature* 596 (7873), pp. 583–589. DOI: 10.1038/s41586-021-03819-2.

Kabsch, Wolfgang (2010): XDS. In *Acta crystallographica. Section D, Biological crystallography* 66 (Pt 2), pp. 125–132. DOI: 10.1107/S0907444909047337.

Källberg, Morten; Wang, Haipeng; Wang, Sheng; Peng, Jian; Wang, Zhiyong; Lu, Hui; Xu, Jinbo (2012): Template-based protein structure modeling using the RaptorX web server. In *Nature protocols* 7 (8), pp. 1511–1522. DOI: 10.1038/nprot.2012.085.

Keilhack, Heike; David, Frank S.; McGregor, Malcolm; Cantley, Lewis C.; Neel, Benjamin G. (2005): Diverse biochemical properties of Shp2 mutants. Implications for disease phenotypes. In *The Journal of biological chemistry* 280 (35), pp. 30984–30993. DOI: 10.1074/jbc.M504699200.

Koncz, G.; Tóth, G. K.; Bökönyi, G.; Kéri, G.; Pecht, I.; Medgyesi, D. et al. (2001): Co-clustering of Fcgamma and B cell receptors induces dephosphorylation of the Grb2-associated binder 1 docking protein. In *European journal of biochemistry* 268 (14), pp. 3898–3906. DOI: 10.1046/j.1432-1327.2001.02295.x.

Kontaridis, Maria I.; Swanson, Kenneth D.; David, Frank S.; Barford, David; Neel, Benjamin G. (2006): PTPN11 (Shp2) mutations in LEOPARD syndrome have dominant negative, not activating, effects. In *The Journal of biological chemistry* 281 (10), pp. 6785–6792. DOI: 10.1074/jbc.M513068200.

Krissinel, Evgeny; Henrick, Kim (2007): Inference of macromolecular assemblies from crystalline state. In *Journal of molecular biology* 372 (3), pp. 774–797. DOI: 10.1016/j.jmb.2007.05.022.

Kuil, Joeri; van Wandelen, Loek T. M.; Mol, Nico J. de; Liskamp, Rob M. J. (2009): Switching between low and high affinity for the Syk tandem SH2 domain by irradiation of azobenzene containing ITAM peptidomimetics. In *Journal of peptide science : an official publication of the European Peptide Society* 15 (10), pp. 685–691. DOI: 10.1002/psc.1173.

Kuriyan, J.; Cowburn, D. (1997): Modular peptide recognition domains in eukaryotic signaling. In *Annual review of biophysics and biomolecular structure* 26, pp. 259–288. DOI: 10.1146/annurev.biophys.26.1.259.

Ladbury, J. E.; Lemmon, M. A.; Zhou, M.; Green, J.; Botfield, M. C.; Schlessinger, J. (1995): Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal. In *Proceedings of the National Academy of Sciences of the United States of America* 92 (8), pp. 3199–3203. DOI: 10.1073/pnas.92.8.3199.

LaMarche, Matthew J.; Acker, Michael; Argintaru, Andreea; Bauer, Daniel; Boisclair, Julie; Chan, Homan et al. (2020): Identification of TNO155, an Allosteric SHP2 Inhibitor for the Treatment of Cancer. In *Journal of medicinal chemistry* 63 (22), pp. 13578–13594. DOI: 10.1021/acs.jmedchem.0c01170.

LaRochelle, Jonathan R.; Fodor, Michelle; Vemulapalli, Vidyasiri; Mohseni, Morvarid; Wang, Ping; Stams, Travis et al. (2018): Structural reorganization of SHP2 by oncogenic mutations and implications for oncoprotein resistance to allosteric inhibition. In *Nature communications* 9 (1), p. 4508. DOI: 10.1038/s41467-018-06823-9.

Lee, C. H.; Kominos, D.; Jacques, S.; Margolis, B.; Schlessinger, J.; Shoelson, S. E.; Kuriyan, J. (1994): Crystal structures of peptide complexes of the amino-terminal SH2 domain of the Syp tyrosine phosphatase. In *Structure (London, England : 1993)* 2 (5), pp. 423–438. DOI: 10.1016/S0969-2126(00)00044-7.

Lewitzky, Marc; Simister, Philip C.; Feller, Stephan M. (2012): Beyond 'furballs' and 'dumpling soups' - towards a molecular architecture of signaling complexes and networks. In *FEBS letters* 586 (17), pp. 2740–2750. DOI: 10.1016/j.febslet.2012.04.029.

Liebschner, Dorothee; Afonine, Pavel V.; Baker, Matthew L.; Bunkóczi, Gábor; Chen, Vincent B.; Croll, Tristan I. et al. (2019): Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. In *Acta crystallographica. Section D, Structural biology* 75 (Pt 10), pp. 861–877. DOI: 10.1107/S2059798319011471.

Lim, Wendell A.; Pawson, Tony (2010): Phosphotyrosine signaling: evolving a new cellular communication system. In *Cell* 142 (5), pp. 661–667. DOI: 10.1016/j.cell.2010.08.023.

Lipari, Giovanni; Szabo, Attila (1982): Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. In *J. Am. Chem. Soc.* 104 (17), pp. 4559–4570. DOI: 10.1021/ja00381a010.

Liu, Bernard A.; Engelmann, Brett W.; Nash, Piers D. (2012): The language of SH2 domain interactions defines phosphotyrosine-mediated signal transduction. In *FEBS letters* 586 (17), pp. 2597–2605. DOI: 10.1016/j.febslet.2012.04.054.

Liu, Bernard A.; Shah, Eshana; Jablonowski, Karl; Stergachis, Andrew; Engelmann, Brett; Nash, Piers D. (2011): The SH2 domain-containing proteins in 21 species establish the provenance and scope of phosphotyrosine signaling in eukaryotes. In *Science signaling* 4 (202), ra83. DOI: 10.1126/scisignal.2002105.

Liu, Hui; Li, Gang; Zeng, Weitao; Zhang, Pengxing; Fan, Feiyan; Tu, Yanyang; Zhang, Yongsheng (2014): Combined detection of Gab1 and Gab2 expression predicts clinical outcome of patients with glioma. In *Medical oncology (Northwood, London, England)* 31 (8), p. 77. DOI: 10.1007/s12032-014-0077-6.

Liu, Yan; Rohrschneider, Larry R. (2002): The gift of Gab. In *FEBS letters* 515 (1-3), pp. 1–7. DOI: 10.1016/s0014-5793(02)02425-0.

Liu, Yanli; Lau, Johnathan; Li, Weiguo; Tempel, Wolfram; Li, Li; Dong, Aiping et al. (2016): Structural basis for the regulatory role of the PPxY motifs in the thioredoxin-interacting protein TXNIP. In *The Biochemical journal* 473 (2), pp. 179–187. DOI: 10.1042/BJ20150830.

Luca, S.; Filippov, D. V.; van Boom, J. H.; Oschkinat, H.; Groot, H. J. de; Baldus, M. (2001): Secondary chemical shifts in immobilized peptides and proteins: a qualitative basis for structure refinement under magic angle spinning. In *Journal of biomolecular NMR* 20 (4), pp. 325–331. DOI: 10.1023/A:1011278317489.

Marasco, M.; Berteotti, A.; Weyershaeuser, J.; Thorausch, N.; Sikorska, J.; Krausze, J. et al. (2020a): Molecular mechanism of SHP2 activation by PD-1 stimulation. In *Science advances* 6 (5), eaay4458. DOI: 10.1126/sciadv.aay4458.

Marasco, Michelangelo; Kirkpatrick, John; Nanna, Vittoria; Sikorska, Justyna; Carlomagno, Teresa (2021): Phosphotyrosine couples peptide binding and SHP2 activation via a dynamic allosteric network. In *Computational and structural biotechnology journal* 19, pp. 2398–2415. DOI: 10.1016/j.csbj.2021.04.040.

Marasco, Michelangelo; Kirkpatrick, John P.; Carlomagno, Teresa (2020b): 1H, 13C, 15N chemical shift assignments of SHP2 SH2 domains in complex with PD-1 immune-tyrosine motifs. In *Biomolecular NMR assignments* 14 (2), pp. 179–188. DOI: 10.1007/s12104-020-09941-y.

Mayer, B. J.; Gupta, R. (1998): Functions of SH2 and SH3 domains. In *Current topics in microbiology and immunology* 228, pp. 1–22. DOI: 10.1007/978-3-642-80481-6_1.

McCoy, Airlie J.; Grosse-Kunstleve, Ralf W.; Adams, Paul D.; Winn, Martyn D.; Storoni, Laurent C.; Read, Randy J. (2007): Phaser crystallographic software. In *Journal of applied crystallography* 40 (Pt 4), pp. 658–674. DOI: 10.1107/S0021889807021206.

McCubrey, James A.; Steelman, Linda S.; Chappell, William H.; Abrams, Stephen L.; Wong, Ellis W. T.; Chang, Fumin et al. (2007): Roles of the Raf/MEK/ERK pathway in cell growth, malignant transformation and drug resistance. In *Biochimica et biophysica acta* 1773 (8), pp. 1263–1284. DOI: 10.1016/j.bbamcr.2006.10.001.

McNemar, C.; Snow, M. E.; Windsor, W. T.; Prongay, A.; Mui, P.; Zhang, R. et al. (1997): Thermodynamic and structural analysis of phosphotyrosine polypeptide binding to Grb2-SH2. In *Biochemistry* 36 (33), pp. 10006–10014. DOI: 10.1021/bi9704360.

Mhatre, Eisha; Monterrosa, Ramses Gallegos; Kovács, Akos T. (2014): From environmental signals to regulators: modulation of biofilm development in Gram-positive bacteria. In *Journal of basic microbiology* 54 (7), pp. 616–632. DOI: 10.1002/jobm.201400175.

Mirdita, Milot; Schütze, Konstantin; Moriwaki, Yoshitaka; Heo, Lim; Ovchinnikov, Sergey; Steinegger, Martin (2022): ColabFold: making protein folding accessible to all. In *Nature methods* 19 (6), pp. 679–682. DOI: 10.1038/s41592-022-01488-1.

Molina, Julian R.; Adjei, Alex A. (2006): The Ras/Raf/MAPK Pathway. In *Journal of Thoracic Oncology* 1 (1), pp. 7–9. DOI: 10.1016/S1556-0864(15)31506-9.

Montagner, Alexandra; Yart, Armelle; Dance, Marie; Perret, Bertrand; Salles, Jean-Pierre; Raynal, Patrick (2005): A novel role for Gab1 and SHP2 in epidermal growth factor-induced Ras activation. In *The Journal of biological chemistry* 280 (7), pp. 5350–5360. DOI: 10.1074/jbc.M410012200.

Mood, Kathleen; Saucier, Caroline; Bong, Yong-Sik; Lee, Hyun-Shik; Park, Morag; Daar, Ira O. (2006): Gab1 is required for cell cycle transition, cell proliferation, and transformation induced by an oncogenic met receptor. In *Molecular biology of the cell* 17 (9), pp. 3717–3728. DOI: 10.1091/mbc.e06-03-0244.

Mortuza, S. M.; Zheng, Wei; Zhang, Chengxin; Li, Yang; Pearce, Robin; Zhang, Yang (2021): Improving fragment-based ab initio protein structure assembly using low-accuracy contact-map predictions. In *Nature communications* 12 (1), p. 5011. DOI: 10.1038/s41467-021-25316-w.

Murshudov, Garib N.; Skubák, Pavol; Lebedev, Andrey A.; Pannu, Navraj S.; Steiner, Roberto A.; Nicholls, Robert A. et al. (2011): REFMAC5 for the refinement of macromolecular crystal structures. In *Acta crystallographica. Section D, Biological crystallography* 67 (Pt 4), pp. 355–367. DOI: 10.1107/S0907444911001314.

Nakaoka, Yoshikazu; Komuro, Issei (2013): Gab docking proteins in cardiovascular disease, cancer, and inflammation. In *International journal of inflammation* 2013, p. 141068. DOI: 10.1155/2013/141068.

Necci, Marco; Piovesan, Damiano; Tosatto, Silvio C. E. (2021): Critical assessment of protein intrinsic disorder prediction. In *Nature methods* 18 (5), pp. 472–481. DOI: 10.1038/s41592-021-01117-3.

Neel, Benjamin G.; Gu, Haihua; Pao, Lily (2003): The 'Shp'ing news: SH2 domain-containing tyrosine phosphatases in cell signaling. In *Trends in biochemical sciences* 28 (6), pp. 284–293. DOI: 10.1016/S0968-0004(03)00091-4.

Ng, Cherlyn; Jackson, Rebecca A.; Buschdorf, Jan P.; Sun, Qingxiang; Guy, Graeme R.; Sivaraman, J. (2008): Structural basis for a novel intrapeptidyl H-bond and reverse binding of c-Cbl-TKB domain substrates. In *The EMBO journal* 27 (5), pp. 804–816. DOI: 10.1038/emboj.2008.18.

Nguyen, L.; Holgado-Madruga, M.; Maroun, C.; Fixman, E. D.; Kamikura, D.; Fournier, T. et al. (1997): Association of the multisubstrate docking protein Gab1 with the hepatocyte growth factor receptor requires a functional Grb2 binding

120

site involving tyrosine 1356. In *The Journal of biological chemistry* 272 (33), pp. 20811–20819. DOI: 10.1074/jbc.272.33.20811.

Niklasson, Markus; Otten, Renee; Ahlner, Alexandra; Andresen, Cecilia; Schlagnitweit, Judith; Petzold, Katja; Lundström, Patrik (2017): Comprehensive analysis of NMR data using advanced line shape fitting. In *Journal of biomolecular NMR* 69 (2), pp. 93–99. DOI: 10.1007/s10858-017-0141-6.

Niogret, Charlène; Birchmeier, Walter; Guarda, Greta (2019): SHP-2 in Lymphocytes' Cytokine and Inhibitory Receptor Signaling. In *Frontiers in immunology* 10, p. 2468. DOI: 10.3389/fimmu.2019.02468.

Nishida, Keigo; Wang, Lin; Morii, Eiichi; Park, Sung Joo; Narimatsu, Masahiro; Itoh, Shousaku et al. (2002): Requirement of Gab2 for mast cell development and KitL/c-Kit signaling. In *Blood* 99 (5), pp. 1866–1869. DOI: 10.1182/blood.v99.5.1866.

Ortiz-Padilla, C.; Gallego-Ortega, D.; Browne, B. C.; Hochgräfe, F.; Caldon, C. E.; Lyons, R. J. et al. (2013): Functional characterization of cancer-associated Gab1 mutations. In *Oncogene* 32 (21), pp. 2696–2702. DOI: 10.1038/onc.2012.271.

Ottinger, E. A.; Botfield, M. C.; Shoelson, S. E. (1998): Tandem SH2 domains confer high specificity in tyrosine kinase signaling. In *The Journal of biological chemistry* 273 (2), pp. 729–735. DOI: 10.1074/jbc.273.2.729.

Pádua, Ricardo A. P.; Sun, Yizhi; Marko, Ingrid; Pitsawong, Warintra; Stiller, John B.; Otten, Renee; Kern, Dorothee (2018): Mechanism of activating mutations and allosteric drug inhibition of the phosphatase SHP2. In *Nature communications* 9 (1), p. 4507. DOI: 10.1038/s41467-018-06814-w.

Panagopoulos, Ioannis; Gorunova, Ludmila; Andersen, Kristin; Tafjord, Svetlana; Lund-Iversen, Marius; Lobmaier, Ingvild et al. (2020): Recurrent Fusion of the GRB2 Associated Binding Protein 1 (GAB1) Gene With ABL Proto-oncogene 1 (ABL1) in Benign Pediatric Soft Tissue Tumors. In *Cancer genomics & proteomics* 17 (5), pp. 499–508. DOI: 10.21873/cgp.20206.

Patsoukis, Nikolaos; Duke-Cohan, Jonathan S.; Chaudhri, Apoorvi; Aksoylar, Halil-Ibrahim; Wang, Qi; Council, Asia et al. (2020): Interaction of SHP-2 SH2 domains with PD-1 ITSM induces PD-1 dimerization and SHP-2 activation. In *Communications biology* 3 (1), p. 128. DOI: 10.1038/s42003-020-0845-0.

Piovesan, Damiano; Monzon, Alexander Miguel; Tosatto, Silvio C. E. (2022): Intrinsic protein disorder and conditional folding in AlphaFoldDB. In *Protein science : a publication of the Protein Society* 31 (11), e4466. DOI: 10.1002/pro.4466.

Poole, Alastair W.; Jones, Matthew L. (2005): A SHPing tale: perspectives on the regulation of SHP-1 and SHP-2 tyrosine phosphatases by the C-terminal tail // A SHPing tale: Perspectives on the regulation of SHP-1 and SHP-2 tyrosine phosphatases by the C-terminal tail. In *Cellular signalling* 17 (11), pp. 1323–1332. DOI: 10.1016/j.cellsig.2005.05.016.

Potterton, Liz; Agirre, Jon; Ballard, Charles; Cowtan, Kevin; Dodson, Eleanor; Evans, Phil R. et al. (2018): CCP4i2: the new graphical user interface to the CCP4 program suite. In *Acta crystallographica. Section D, Structural biology* 74 (Pt 2), pp. 68–84. DOI: 10.1107/S2059798317016035.

Qin, Chuanguang; Wavreille, Anne-Sophie; Pei, Dehua (2005): Alternative mode of binding to phosphotyrosyl peptides by Src homology-2 domains. In *Biochemistry* 44 (36), pp. 12196–12202. DOI: 10.1021/bi050066o.

Quan, Jiayuan; Tian, Jingdong (2009): Circular polymerase extension cloning of complex gene libraries and pathways. In *PloS one* 4 (7), e6441. DOI: 10.1371/journal.pone.0006441.

Raabe, T.; Riesgo-Escovar, J.; Liu, X.; Bausenwein, B. S.; Deak, P.; Maröy, P.; Hafen, E. (1996): DOS, a novel pleckstrin homology domain-containing protein required for signal transduction between sevenless and Ras1 in Drosophila. In *Cell* 85 (6), pp. 911–920. DOI: 10.1016/S0092-8674(00)81274-X.

Rani, Aradhana; Murphy, John J. (2016): STAT5 in Cancer and Immunity. In *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research* 36 (4), pp. 226–237. DOI: 10.1089/jir.2015.0054.

Ren, Yuan; Meng, Songshu; Mei, Lin; Zhao, Z. Joe; Jove, Richard; Wu, Jie (2004): Roles of Gab1 and SHP2 in paxillin tyrosine dephosphorylation and Src activation in response to epidermal growth factor. In *The Journal of biological chemistry* 279 (9), pp. 8497–8505. DOI: 10.1074/jbc.M312575200.

Rodrigues, G. A.; Falasca, M.; Zhang, Z.; Ong, S. H.; Schlessinger, J. (2000): A novel positive feedback loop mediated by the docking protein Gab1 and phosphatidylinositol 3-kinase in epidermal growth factor receptor signaling. In *Molecular and cellular biology* 20 (4), pp. 1448–1459. DOI: 10.1128/MCB.20.4.1448-1459.2000.

Rodriguez-Viciana, P.; Warne, P. H.; Dhand, R.; Vanhaesebroeck, B.; Gout, I.; Fry, M. J. et al. (1994): Phosphatidylinositol-3-OH kinase as a direct target of Ras. In *Nature* 370 (6490), pp. 527–532. DOI: 10.1038/370527a0.

Romero, Pedro; Obradovic, Zoran; Li, Xiaohong; Garner, Ethan C.; Brown, Celeste J.; Dunker, A. Keith (2001): Sequence complexity of disordered protein. In *Proteins* 42 (1), pp. 38–48. DOI: 10.1002/1097-0134(20010101)42:1<38::aid-prot50>3.0.co;2-3.

Sachs, M.; Brohmann, H.; Zechner, D.; Müller, T.; Hülsken, J.; Walther, I. et al. (2000): Essential role of Gab1 for signaling by the c-Met receptor in vivo. In *The Journal of cell biology* 150 (6), pp. 1375–1384. DOI: 10.1083/jcb.150.6.1375.

Saint-Laurent, Céline; Mazeyrie, Laurène; Tajan, Mylène; Paccoud, Romain; Castan-Laurell, Isabelle; Valet, Philippe et al. (2022): The Tyrosine Phosphatase SHP2: A New Target for Insulin Resistance? In *Biomedicines* 10 (9). DOI: 10.3390/biomedicines10092139.

Sakamoto, K. M.; Kim, K. B.; Kumagai, A.; Mercurio, F.; Crews, C. M.; Deshaies, R. J. (2001): Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. In *Proceedings of the National Academy of Sciences of the United States of America* 98 (15), pp. 8554–8559. DOI: 10.1073/pnas.141230798.

Sarmiento, M.; Puius, Y. A.; Vetter, S. W.; Keng, Y. F.; Wu, L.; Zhao, Y. et al. (2000): Structural basis of plasticity in protein tyrosine phosphatase 1B substrate recognition. In *Biochemistry* 39 (28), pp. 8171–8179. DOI: 10.1021/bi000319w.

Sathish, J. G.; Johnson, K. G.; Fuller, K. J.; LeRoy, F. G.; Meyaard, L.; Sims, M. J.; Matthews, R. J. (2001): Constitutive association of SHP-1 with leukocyte-associated Ig-like receptor-1 in human T cells. In *Journal of immunology (Baltimore, Md. : 1950)* 166 (3), pp. 1763–1770. DOI: 10.4049/jimmunol.166.3.1763.

Schutzman, J. L.; Borland, C. Z.; Newman, J. C.; Robinson, M. K.; Kokel, M.; Stern, M. J. (2001): The Caenorhabditis elegans EGL-15 signaling pathway implicates a DOS-like multisubstrate adaptor protein in fibroblast growth factor signal transduction. In *Molecular and cellular biology* 21 (23), pp. 8104–8116. DOI: 10.1128/MCB.21.23.8104-8116.2001.

Seda, Vaclav; Vojackova, Eva; Ondrisova, Laura; Kostalova, Lenka; Sharma, Sonali; Loja, Tomas et al. (2021): FoxO1-GAB1 axis regulates homing capacity and tonic AKT activity in chronic lymphocytic leukemia. In *Blood* 138 (9), pp. 758–772. DOI: 10.1182/blood.2020008101.

Seiden-Long, Isolde; Navab, Roya; Shih, Warren; Li, Ming; Chow, Jane; Zhu, Chang Qi et al. (2008): Gab1 but not Grb2 mediates tumor progression in Met overexpressing colorectal cancer cells. In *Carcinogenesis* 29 (3), pp. 647–655. DOI: 10.1093/carcin/bgn009.

Seiffert, Martina; Custodio, Joseph M.; Wolf, Ingrid; Harkey, Michael; Liu, Yan; Blattman, Joseph N. et al. (2003): Gab3-deficient mice exhibit normal development and hematopoiesis and are immunocompetent. In *Molecular and cellular biology* 23 (7), pp. 2415–2424. DOI: 10.1128/MCB.23.7.2415-2424.2003.

Sha, Fern; Gencer, Emel Basak; Georgeon, Sandrine; Koide, Akiko; Yasui, Norihisa; Koide, Shohei; Hantschel, Oliver (2013): Dissection of the BCR-ABL signaling network using highly specific monobody inhibitors to the SHP2 SH2 domains. In *Proceedings of the National Academy of Sciences of the United States of America* 110 (37), pp. 14924–14929. DOI: 10.1073/pnas.1303640110.

Shao, Nai-Yuan; Wang, Dong-Xing; Wang, Yin; Li, Ya; Zhang, Zhi-Qing; Jiang, Qin et al. (2018): MicroRNA-29a-3p Downregulation Causes Gab1 Upregulation to Promote Glioma Cell Proliferation. In *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* 48 (2), pp. 450–460. DOI: 10.1159/000491776.

Shen, S. H.; Bastien, L.; Posner, B. I.; Chrétien, P. (1991): A protein-tyrosine phosphatase with sequence similarity to the SH2 domain of the protein-tyrosine kinases. In *Nature* 352 (6337), pp. 736–739. DOI: 10.1038/352736a0.

Shiu, Shin-Han; Li, Wen-Hsiung (2004): Origins, lineage-specific expansions, and multiple losses of tyrosine kinases in eukaryotes. In *Molecular biology and evolution* 21 (5), pp. 828–840. DOI: 10.1093/molbev/msh077.

Simister, Philip C.; Schaper, Fred; O'Reilly, Nicola; McGowan, Simon; Feller, Stephan M. (2011): Self-organization and regulation of intrinsically disordered proteins with folded N-termini. In *PLoS biology* 9 (2), e1000591. DOI: 10.1371/journal.pbio.1000591.

Srivastava, Mansi; Simakov, Oleg; Chapman, Jarrod; Fahey, Bryony; Gauthier, Marie E. A.; Mitros, Therese et al. (2010): The Amphimedon queenslandica genome and the evolution of animal complexity. In *Nature* 466 (7307), pp. 720–726. DOI: 10.1038/nature09201.

Sugimoto, S.; Wandless, T. J.; Shoelson, S. E.; Neel, B. G.; Walsh, C. T. (1994): Activation of the SH2-containing protein tyrosine phosphatase, SH-PTP2, by phosphotyrosine-containing peptides derived from insulin receptor substrate-1. In *The Journal of biological chemistry* 269 (18), pp. 13614–13622.

Sweeney, Michael C.; Wavreille, Anne-Sophie; Park, Junguk; Butchar, Jonathan P.; Tridandapani, Susheela; Pei, Dehua (2005): Decoding protein-protein interactions through combinatorial chemistry: sequence specificity of SHP-1, SHP-2, and SHIP SH2 domains. In *Biochemistry* 44 (45), pp. 14932–14947. DOI: 10.1021/bi051408h.

Tartaglia, M.; Mehler, E. L.; Goldberg, R.; Zampino, G.; Brunner, H. G.; Kremer, H. et al. (2001): Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. In *Nature genetics* 29 (4), pp. 465–468. DOI: 10.1038/ng772.

Tartaglia, Marco; Martinelli, Simone; Stella, Lorenzo; Bocchinfuso, Gianfranco; Flex, Elisabetta; Cordeddu, Viviana et al. (2006): Diversity and functional consequences of germline and somatic PTPN11 mutations in human disease. In *American journal of human genetics* 78 (2), pp. 279–290. DOI: 10.1086/499925.

Tautz, Lutz; Critton, David A.; Grotegut, Stefan (2013): Protein tyrosine phosphatases: structure, function, and implication in human disease. In *Methods in molecular biology (Clifton, N.J.)* 1053, pp. 179–221. DOI: 10.1007/978-1-62703-562-0_13.

Tautz, Lutz; Sergienko, Eduard A. (2013): High-throughput screening for protein tyrosine phosphatase activity modulators. In *Methods in molecular biology (Clifton, N.J.)* 1053, pp. 223–240. DOI: 10.1007/978-1-62703-562-0_14.

Tokuriki, Nobuhiko; Oldfield, Christopher J.; Uversky, Vladimir N.; Berezovsky, Igor N.; Tawfik, Dan S. (2009): Do viral proteins possess unique biophysical features? In *Trends in biochemical sciences* 34 (2), pp. 53–59. DOI: 10.1016/j.tibs.2008.10.009.

Toto, Angelo; Malagrinò, Francesca; Visconti, Lorenzo; Troilo, Francesca; Gianni, Stefano (2020): Unveiling the Molecular Basis of the Noonan Syndrome-Causing Mutation T42A of SHP2. In *International journal of molecular sciences* 21 (2). DOI: 10.3390/ijms21020461.

Tropea, Joseph E.; Cherry, Scott; Waugh, David S. (2009): Expression and purification of soluble His(6)-tagged TEV protease. In *Methods in molecular biology (Clifton, N.J.)* 498, pp. 297–307. DOI: 10.1007/978-1-59745-196-3_19.

Tuteja, Narendra (2009): Signaling through G protein coupled receptors. In *Plant signaling & behavior* 4 (10), pp. 942–947. DOI: 10.4161/psb.4.10.9530.

Uhlik, Mark T.; Temple, Brenda; Bencharit, Sompop; Kimple, Adam J.; Siderovski, David P.; Johnson, Gary L. (2005): Structural and evolutionary division of phosphotyrosine binding (PTB) domains. In *Journal of molecular biology* 345 (1), pp. 1–20. DOI: 10.1016/j.jmb.2004.10.038.

van der Lee, Robin; Buljan, Marija; Lang, Benjamin; Weatheritt, Robert J.; Daughdrill, Gary W.; Dunker, A. Keith et al. (2014): Classification of intrinsically disordered regions and proteins. In *Chemical reviews* 114 (13), pp. 6589–6631. DOI: 10.1021/cr400525m.

Varadi, Mihaly; Anyango, Stephen; Deshpande, Mandar; Nair, Sreenath; Natassia, Cindy; Yordanova, Galabina et al. (2022): AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. In *Nucleic acids research* 50 (D1), D439-D444. DOI: 10.1093/nar/gkab1061.

Vemulapalli, Vidyasiri; Chylek, Lily A.; Erickson, Alison; Pfeiffer, Anamarija; Gabriel, Khal-Hentz; LaRochelle, Jonathan et al. (2021): Time-resolved phosphoproteomics reveals scaffolding and catalysis-responsive patterns of SHP2-dependent signaling. In *eLife* 10. DOI: 10.7554/eLife.64251.

Verdine, Gregory L.; Hilinski, Gerard J. (2012): Stapled peptides for intracellular drug targets. In *Methods in enzymology* 503, pp. 3–33. DOI: 10.1016/B978-0-12-396962-0.00001-X.

Verma, Sheetal; Vaughan, Tamisha; Bunting, Kevin D. (2012): Gab adapter proteins as therapeutic targets for hematologic disease. In *Advances in hematology* 2012, p. 380635. DOI: 10.1155/2012/380635.

Visconti, Lorenzo; Malagrinò, Francesca; Pagano, Livia; Toto, Angelo (2020): Understanding the Mechanism of Recognition of Gab2 by the N-SH2 Domain of SHP2. In *Life (Basel, Switzerland)* 10 (6). DOI: 10.3390/life10060085.

Waksman, G.; Kominos, D.; Robertson, S. C.; Pant, N.; Baltimore, D.; Birge, R. B. et al. (1992): Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. In *Nature* 358 (6388), pp. 646–653. DOI: 10.1038/358646a0.

Wang, Bin; Zhang, Lei; Dai, Tong; Qin, Ziran; Lu, Huasong; Zhang, Long; Zhou, Fangfang (2021): Liquid-liquid phase separation in human health and diseases. In *Signal transduction and targeted therapy* 6 (1), p. 290. DOI: 10.1038/s41392-021-00678-1.

Wang, Mingliang; Lu, Jianfeng; Wang, Mi; Yang, Chao-Yie; Wang, Shaomeng (2020): Discovery of SHP2-D26 as a First, Potent, and Effective PROTAC Degrader of SHP2 Protein. In *Journal of medicinal chemistry* 63 (14), pp. 7510–7528. DOI: 10.1021/acs.jmedchem.0c00471.

Wang, Ning; Li, Zhe; Ding, Ronghua; Frank, Gerald D.; Senbonmatsu, Takaaki; Landon, Erwin J. et al. (2006): Antagonism or synergism. Role of tyrosine phosphatases SHP-1 and SHP-2 in growth factor signaling. In *The Journal of biological chemistry* 281 (31), pp. 21878–21883. DOI: 10.1074/jbc.M605018200.

Wang, Rui; Leung, Pete Y. M.; Huang, Feng; Tang, Qingzhuang; Kaneko, Tomonori; Huang, Mei et al. (2018): Reverse Binding Mode of Phosphotyrosine Peptides with SH2 Protein. In *Biochemistry* 57 (35), pp. 5257–5269. DOI: 10.1021/acs.biochem.8b00677.

Wang, Siying; Yu, Wen-Mei; Zhang, Wanming; McCrae, Keith R.; Neel, Benjamin G.; Qu, Cheng-Kui (2009): Noonan syndrome/leukemia-associated gain-of-function mutations in SHP-2 phosphatase (PTPN11) enhance cell migration and angiogenesis. In *The Journal of biological chemistry* 284 (2), pp. 913–920. DOI: 10.1074/jbc.M804129200.

Wang, Wei; Liu, Lijun; Song, Xi; Mo, Yi; Komma, Chandrasekhar; Bellamy, Henry D. et al. (2011): Crystal structure of human protein tyrosine phosphatase SHP-1 in the open conformation. In *Journal of cellular biochemistry* 112 (8), pp. 2062–2071. DOI: 10.1002/jcb.23125.

Wang, Xiao; Peng, Jing; Yang, Ziqiang; Zhou, Pei-Jie; An, Na; Wei, Lianzi et al. (2019a): Elevated expression of Gab1 promotes breast cancer metastasis by dissociating the PAR complex. In *Journal of experimental & clinical cancer research : CR* 38 (1), p. 27. DOI: 10.1186/s13046-019-1025-2.

Wang, Zhengqi; Feng, Gen-Sheng; Bunting, Kevin D. (2017): Unique Roles for Grb2-Associated Binding (Gab) Adaptors in Hematopoietic Stem Cell Homeostasis and Repopulation. In *Blood* 130, p. 2431. DOI: 10.1182/blood.V130.Suppl_1.2431.2431.

Wang, Zhengqi; Vaughan, Tamisha Y.; Zhu, Wandi; Chen, Yuhong; Fu, Guoping; Medrzycki, Magdalena et al. (2019b): Gab2 and Gab3 Redundantly Suppress Colitis by Modulating Macrophage and CD8+ T-Cell Activation. In *Frontiers in immunology* 10, p. 486. DOI: 10.3389/fimmu.2019.00486.

Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. (2004): Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. In *Journal of molecular biology* 337 (3), pp. 635–645. DOI: 10.1016/j.jmb.2004.02.002.

Williamson, Mike P. (2013): Using chemical shift perturbation to characterise ligand binding. In *Progress in nuclear magnetic resonance spectroscopy* 73, pp. 1–16. DOI: 10.1016/j.pnmrs.2013.02.001.

Wilson, Carter J.; Choy, Wing-Yiu; Karttunen, Mikko (2022): AlphaFold2: A Role for Disordered Protein/Region Prediction? In *International journal of molecular sciences* 23 (9). DOI: 10.3390/ijms23094591.

Wolf, Alexandra; Eulenfeld, René; Bongartz, Hannes; Hessenkemper, Wiebke; Simister, Philip C.; Lievens, Sam et al. (2015): MAPK-induced Gab1 translocation to the plasma membrane depends on a regulated intramolecular switch. In *Cellular signalling* 27 (2), pp. 340–352. DOI: 10.1016/j.cellsig.2014.11.017.

Wolf, Ingrid; Jenkins, Brendan J.; Liu, Yan; Seiffert, Martina; Custodio, Joseph M.; Young, Paul; Rohrschneider, Larry R. (2002): Gab3, a new DOS/Gab family member, facilitates macrophage differentiation. In *Molecular and cellular biology* 22 (1), pp. 231–244. DOI: 10.1128/MCB.22.1.231-244.2002.

Wu, Xiaoqin; Xu, Gang; Li, Xiaobo; Xu, Weiren; Li, Qianjin; Liu, Wei et al. (2019): Small Molecule Inhibitor that Stabilizes the Autoinhibited Conformation of the Oncogenic Tyrosine Phosphatase SHP2. In *Journal of medicinal chemistry* 62 (3), pp. 1125–1137. DOI: 10.1021/acs.jmedchem.8b00513.

Xiang, Shihao; Wang, Na; Hui, Pingping; Ma, Jiali (2017): Gab3 is required for human colorectal cancer cell proliferation. In *Biochemical and biophysical research communications* 484 (4), pp. 719–725. DOI: 10.1016/j.bbrc.2017.01.095.

Xue, Bin; Williams, Robert W.; Oldfield, Christopher J.; Dunker, A. Keith; Uversky, Vladimir N. (2010): Archaic chaos: intrinsically disordered proteins in Archaea. In *BMC systems biology* 4 Suppl 1 (Suppl 1), S1. DOI: 10.1186/1752-0509-4-S1-S1.

Yang, A. S.; Honig, B. (1995): Free energy determinants of secondary structure formation: I. alpha-Helices. In *Journal of molecular biology* 252 (3), pp. 351–365. DOI: 10.1006/jmbi.1995.0502.

Yang, Jian; Liu, Lijun; He, Dandan; Song, Xi; Liang, Xiaoshan; Zhao, Zhizhuang Joe; Zhou, G. Wayne (2003): Crystal structure of human protein-tyrosine phosphatase SHP-1. In *The Journal of biological chemistry* 278 (8), pp. 6516–6520. DOI: 10.1074/jbc.M210430200.

Yart, Armelle; Chap, Hugues; Raynal, Patrick (2002): Phosphoinositide 3-kinases in lysophosphatidic acid signaling: regulation and cross-talk with the Ras/mitogen-activated protein kinase pathway. In *Biochimica et biophysica acta* 1582 (1-3), pp. 107–111. DOI: 10.1016/S1388-1981(02)00144-0.

Yu, Cheng Fang; Liu, Zhen-Xiang; Cantley, Lloyd G. (2002): ERK negatively regulates the epidermal growth factor-mediated interaction of Gab1 and the phosphatidylinositol 3-kinase. In *The Journal of biological chemistry* 277 (22), pp. 19382–19388. DOI: 10.1074/jbc.M200732200.

Yu, Zhi-Hong; Xu, Jie; Walls, Chad D.; Chen, Lan; Zhang, Sheng; Zhang, Ruoyu et al. (2013): Structural and mechanistic insights into LEOPARD syndrome-associated SHP2 mutations. In *The Journal of biological chemistry* 288 (15), pp. 10472–10482. DOI: 10.1074/jbc.M113.450023.

Yu, Zhi-Hong; Zhang, Ruo-Yu; Walls, Chad D.; Chen, Lan; Zhang, Sheng; Wu, Li et al. (2014): Molecular basis of gain-of-function LEOPARD syndrome-associated SHP2 mutations. In *Biochemistry* 53 (25), pp. 4136–4151. DOI: 10.1021/bi5002695.

Zalewski, Jenna K.; Heber, Simone; Mo, Joshua H.; O'Conor, Keith; Hildebrand, Jeffrey D.; VanDemark, Andrew P. (2017): Combining Wet and Dry Lab Techniques to Guide the Crystallization of Large Coiled-coil Containing Proteins. In *Journal of visualized experiments : JoVE* (119). DOI: 10.3791/54886.

Zeke, András; Takács, Tamás; Sok, Péter; Németh, Krisztina; Kirsch, Klára; Egri, Péter et al. (2022): Structural insights into the pSer/pThr dependent regulation of the SHP2 tyrosine phosphatase in insulin and CD28 signaling. In *Nature communications* 13 (1), p. 5439. DOI: 10.1038/s41467-022-32918-5.

Zhang, Si Qing; Tsiaras, William G.; Araki, Toshiyuki; Wen, Gengyun; Minichiello, Liliana; Klein, Ruediger; Neel, Benjamin G. (2002): Receptor-specific regulation of phosphatidylinositol 3'-kinase activation by the protein tyrosine phosphatase Shp2. In *Molecular and cellular biology* 22 (12), pp. 4062–4072. DOI: 10.1128/MCB.22.12.4062-4072.2002.

Zhang, Si Qing; Yang, Wentian; Kontaridis, Maria I.; Bivona, Trever G.; Wen, Gengyun; Araki, Toshiyuki et al. (2004): Shp2 regulates SRC family kinase activity and Ras/Erk activation by controlling Csk recruitment. In *Molecular cell* 13 (3), pp. 341–355. DOI: 10.1016/s1097-2765(04)00050-4.

Zhang, Yanyan; Zhang, Jinjin; Yuan, Chunhua; Hard, Ryan L.; Park, In-Hee; Li, Chenglong et al. (2011): Simultaneous binding of two peptidyl ligands by a SRC homology 2 domain. In *Biochemistry* 50 (35), pp. 7637–7646. DOI: 10.1021/bi200439v.

Supplement
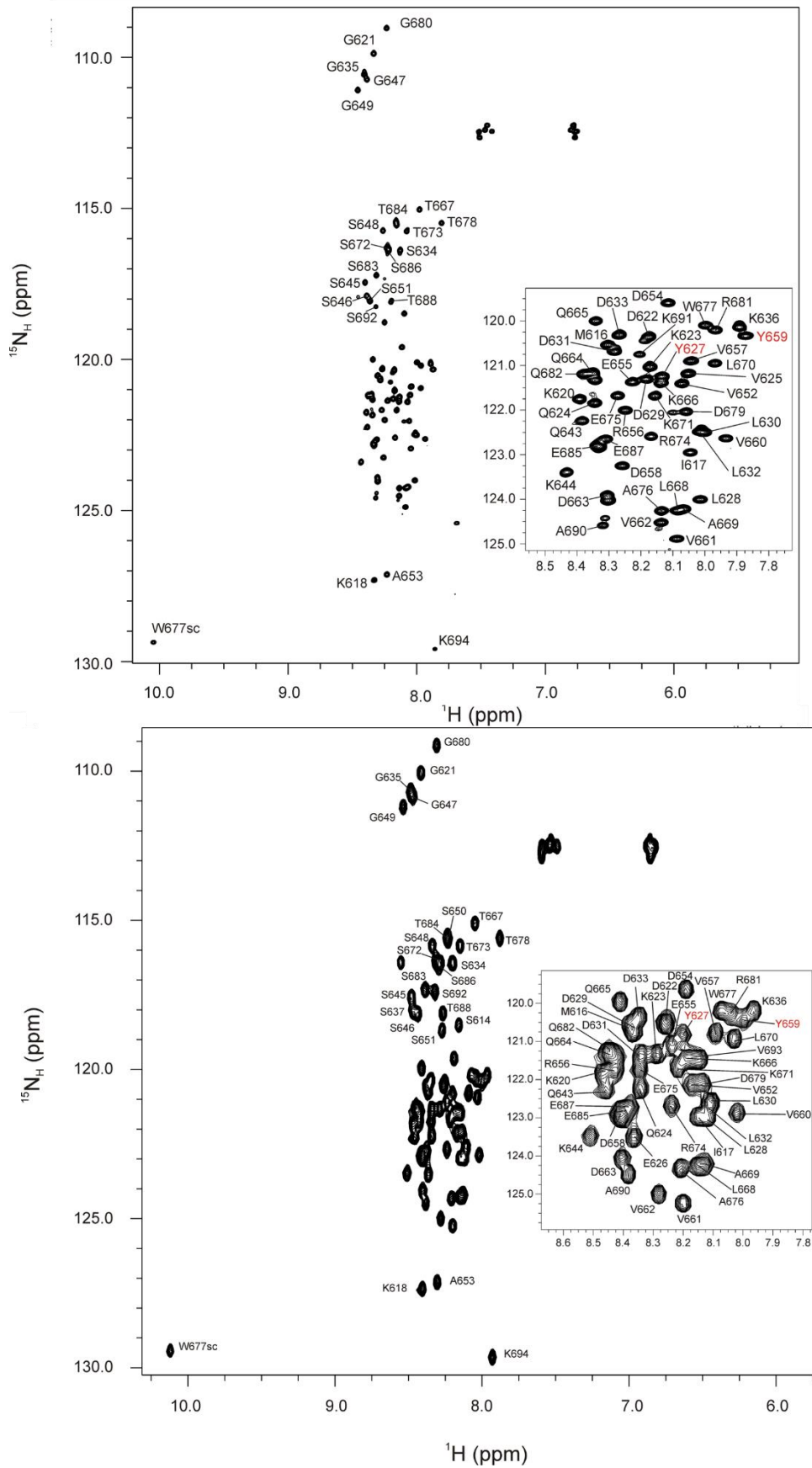
Figure 49 ^{15}N HSQC of unphosphorylated (top) and phosphorylated (bottom) Gab1$^{613-694}$. Assignment done by Dr. rer. nat Tobias Gruber

N-SH2$^{1-106}$:Gab1$^{613-694}$	
N	0.93; 1.14; 1.31; 1.22; 1.34
K$_A$ [*10^5 M]	3.23; 3.35; 6.12; 1.54; 1.94
- ΔS · T	1416; 2608; 960; 1043
ΔH	-6124; -5285; -6119; -6169
ΔG	-7543; -7839; -7079; -7212

C-SH2$^{102-220}$:Gab1$^{613-694}$	
N	1.28; 1.05; 1.1
K$_A$ [*10^5 M]	3.35; 4.58; 3.13
- ΔS · T	-4084; -3697; -4203
ΔH	-11900; -11430; -11700
ΔG	-7592; -7732; -7496

Shp2$^{1-122}$:Gab1$^{613-694}$	
N	0.88; 1.03; 1.03
K$_A$ [*10^7 M]	1.78; 3.72; 3.72
- ΔS · T	-10226; -9112; -9112
ΔH	-20130; -19550; -19550
ΔG	-9903; -10337; -10192
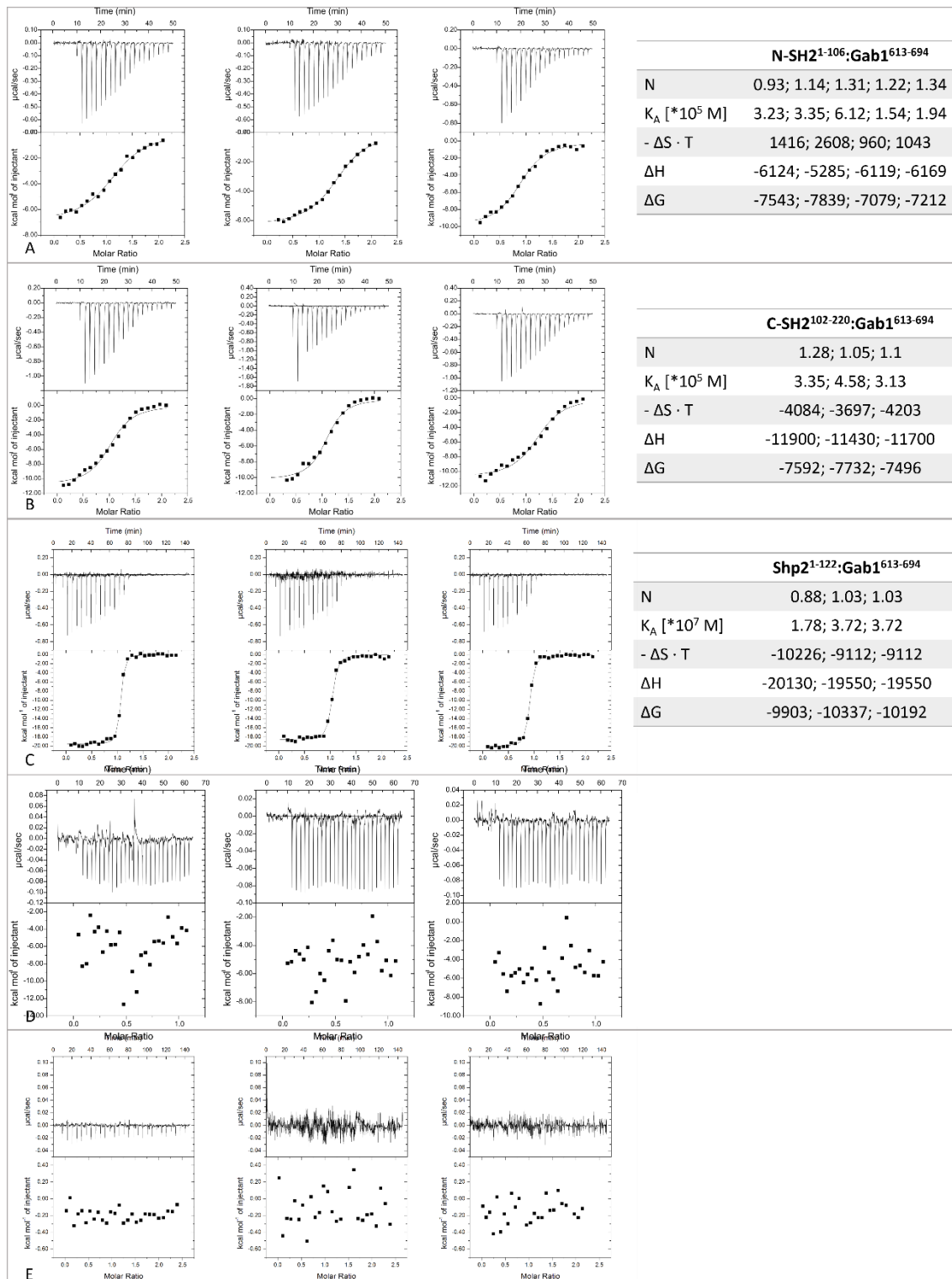
Figure 50 *Replicates of ITC measurements. A) N-SH2$^{1-106}$; B) C-SH2$^{102-220}$ and C) Shp2$^{1-222}$ with pYpY-Gab1$^{613-694}$. D) N-SH2$^{1-106}$ and pY-Gab1$^{651-694}$. E) C-SH2$^{102-220}$ and pY-Gab1$^{613-651}$.*

Ramachandran blot of the Gab1-peptides from the complex of N- SH2$^{1-106}$:pY- Gab1$^{613-651}$ (Figure 51, A) and Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$ (Figure 51, B):

The Ramachandran blot of the peptide shows, with the exception of D631 (+4), that it is in the extended conformation as the amino acids take on psi and phi angles that correspond to the upper left-hand corner of the blot. Stabilising H-bonds between the

peptides and domains backbone can be seen at the +1, +2 and +4 positions towards H53, K91 and K89, respectively.
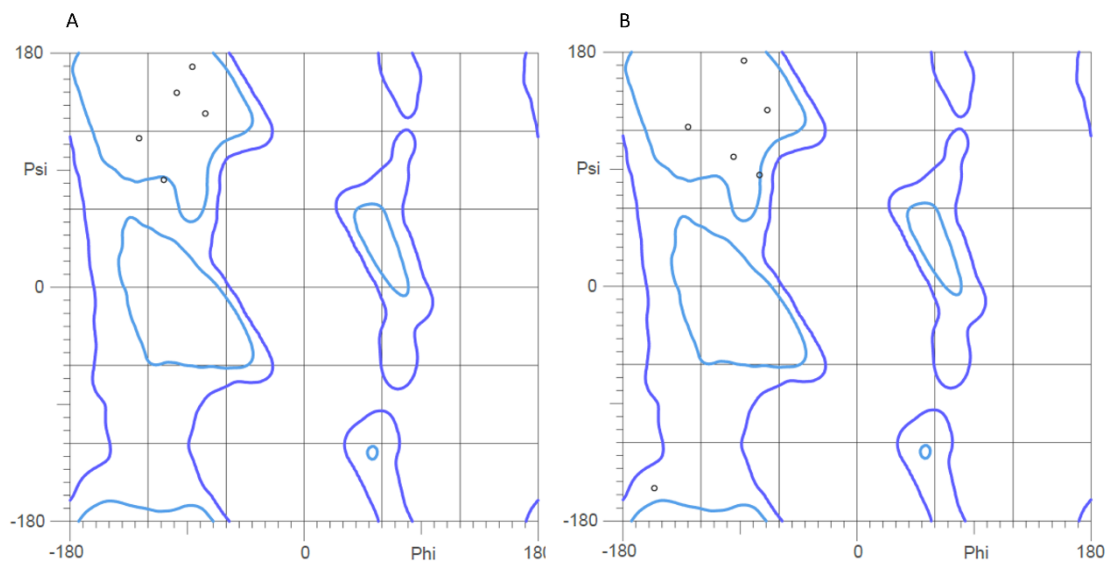
Figure 51 Ramachandran plot of the Gab1 fragment bound to the N-SH2 domain in the N-SH2$^{1-106}$: pY-Gab1$^{613-651}$ structure (A) and the Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$ complex (B).

For the complex of N-SH2$^{1-106}$:pY-Gab1$^{613-651}$ (Figure 52): The domain has direct contact with other symmetry mates via αA and αB as well as βB βC βD βE, all positioned at the 'back' of the domain. Additional H-bonds to the symmetry mates via crystallographic water molecules, and the EF and BG-loop are present.
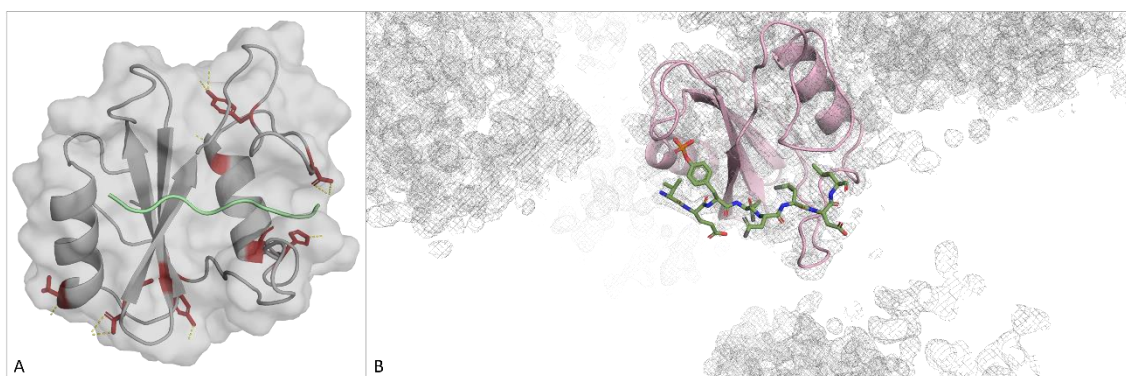
Figure 52 Crystal contacts and molecule in electron density cloud. A) N-SH2$^{1-106}$ (grey, surface and cartoon)in complex with the solved Gab1 fragment (green, cartoon). Residues involved in crystal contacts are coloured in red and depicted as stick model. B) The solved structure of the N-SH2$^{1-106}$ (light pink, cartoon) in complex with the Gab1 fragment (green, stick) is depicted in the surrounding density cloud of the symmetry mates. The 'back side' of the domain is involved in crystal contacts; the peptide is in an unoccupied area. The remaining amino acids of the flexible peptide potentially occupy this space.

Shp2¹⁻²²² after 1 day in
0.1M Tris pH 8.0 ; 25% PEG 3350

Shp2¹⁻²²²:Gab1⁶¹⁷⁻⁶⁸⁴ after 1 day in
0.1M Tris pH 8.0 ; 25% PEG 3350
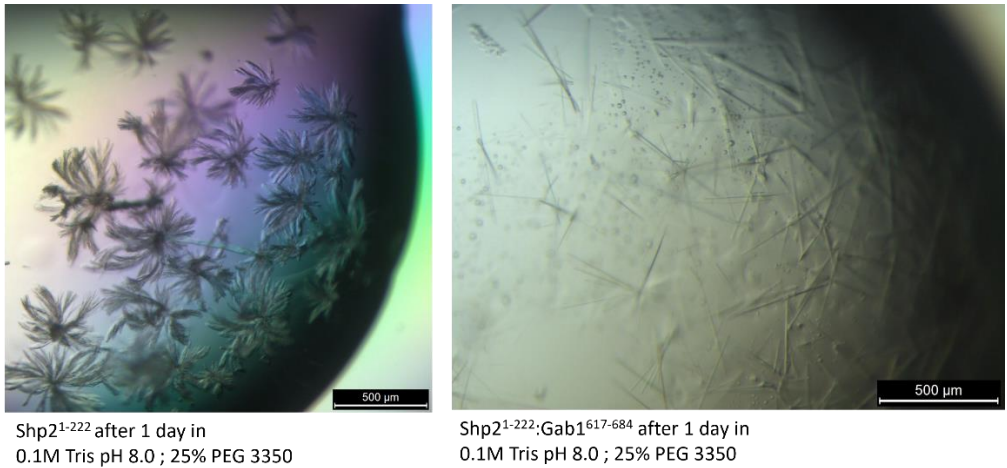
Figure 53 *Crystals of Shp2$^{1-222}$ and in complex to pYpY-Gab1$^{617-684}$ in the same condition*

0.1M Tris pH 8.0 ; 25% PEG 3350

0.1M Hepes pH 7.6; 23% PEG 3350

0.1M Hepes pH 7.4; 21% PEG 3350

0.1M Hepes pH 7.5; 25% PEG 3350
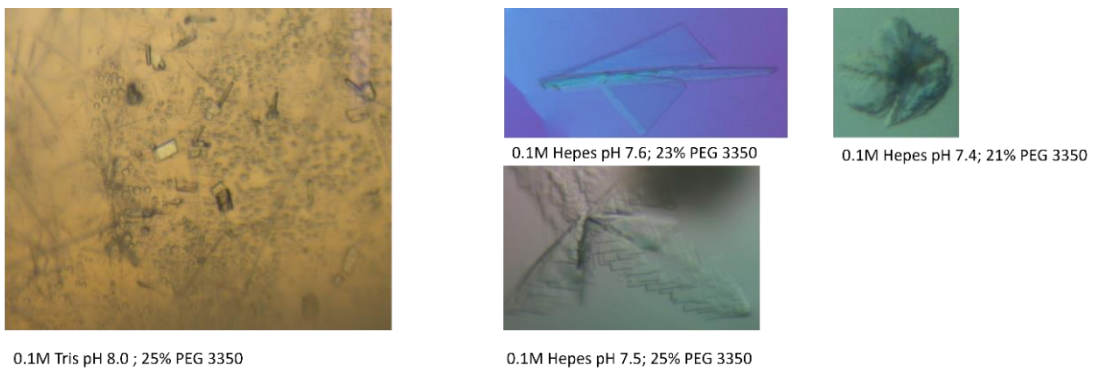
Figure 54 *Crystals of Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$ that dissolved the needle-like form and reappeared as more compact crystals.*
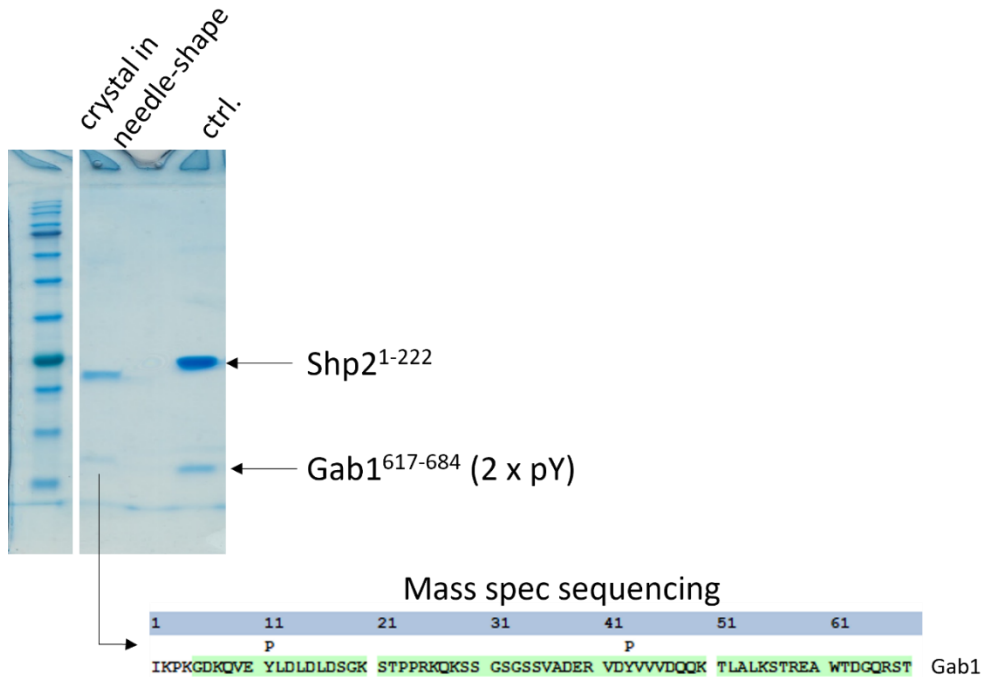
Figure 55 *Mass spectrometry data and SDS-gel of Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$ crystal. Crystals were washed in crystallisation buffer, dissolved in sample buffer and loaded on 15% gel. The protein band of the Gab1 complex partner was provided to the Proteomic/ Mass-Spectrometry Core facility (Dr. Matt Fuszard). The sequenced regions (edman degradation) are illustrated in green of the doubly phosphorylated pYpY-Gab1$^{617-684}$.*

For the complex of Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$: The complex makes crystal contacts via its N-SH2 domain's αA, BC, EF and BG-loops as well as interactions with symmetry mates in the αB and EF-loop of the C-SH2 and the linker connecting the two domains as well as K666, located in the Gab1-helix (**Figure 56**). Further, the back side of the C-SH2 domain is in juxtaposition with the extended part of the pY659-Gab fragment.
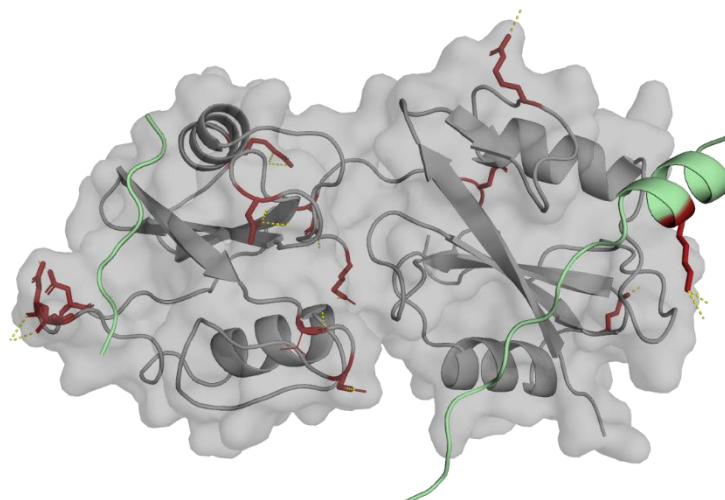
Figure 56 *Crystal contacts of the Shp2$^{1-222}$:pYpY-Gab1$^{617-684}$ structure solved by microED. The tandem SH2 domains are coloured in grey and is shown as surface. The structure in general is depicted as cartoon. The Gab1 fragments are tainted in green. Residues that are involved in crystal packing are shown as sticks coloured in red. The yellow dashed lines indicate the direction of interaction to the crystal mate.*
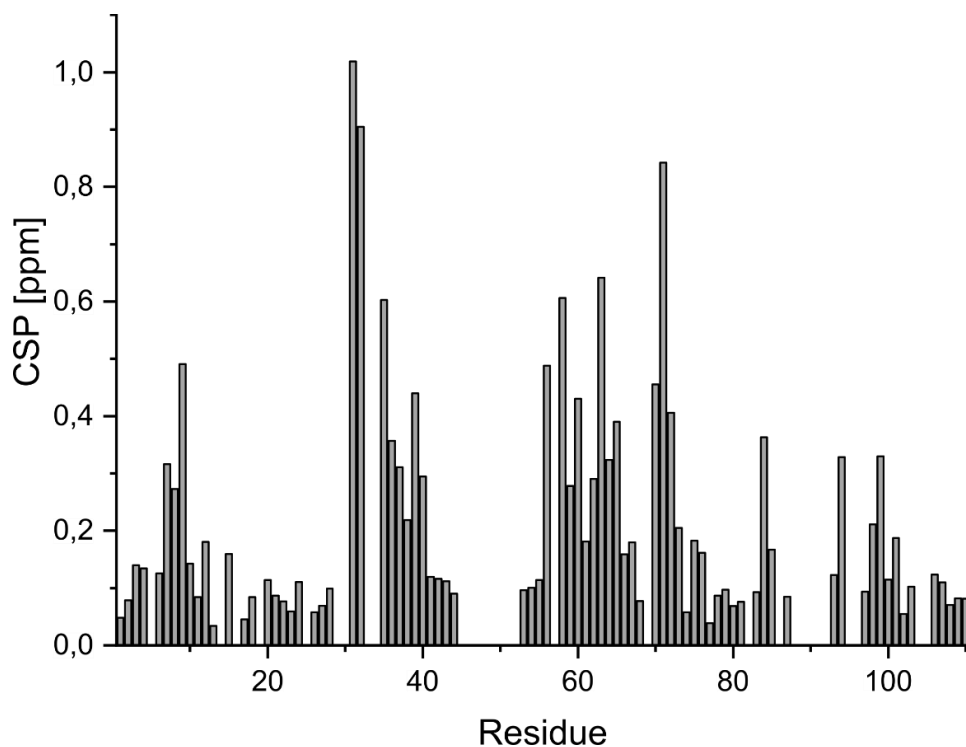
Figure 57 *Chemical shift perturbation (CSP) upon binding to pYpY-Gab1$^{613-694}$ are blotted on the residual number of N-SH2$^{1-106}$.*
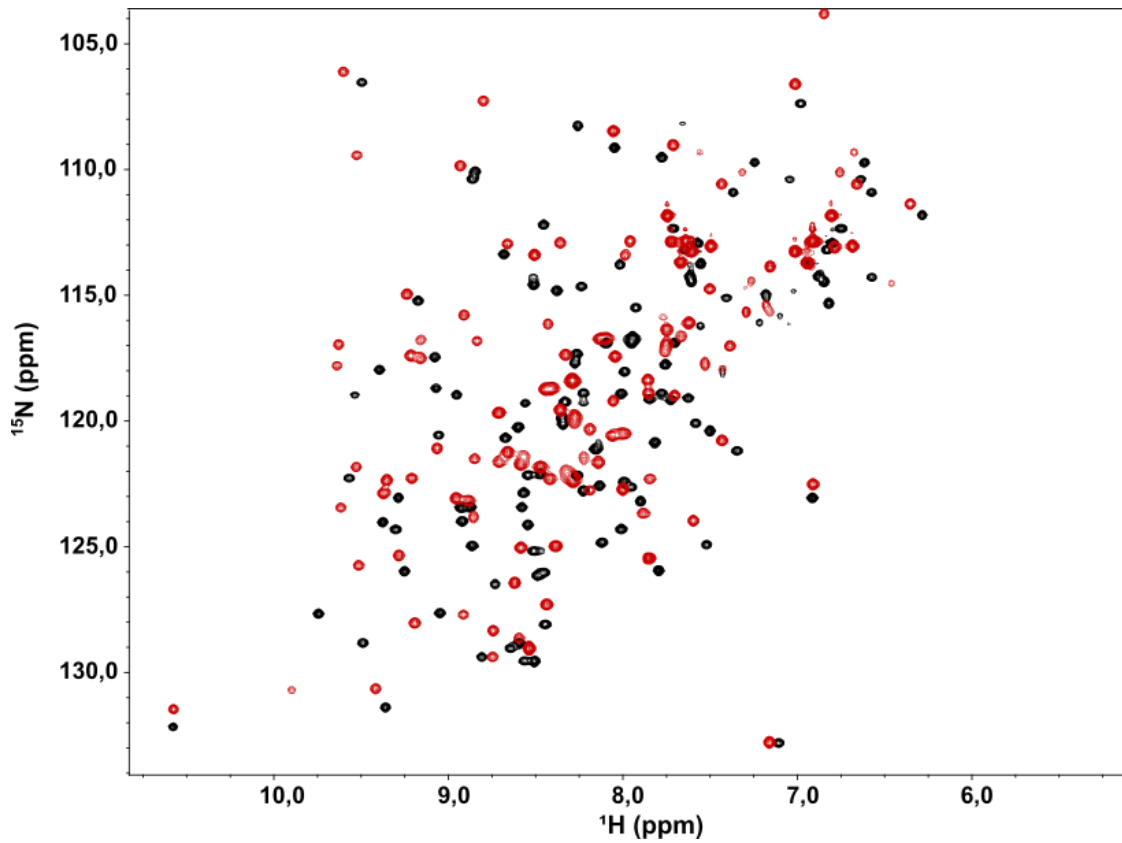
Figure 58 *Superposition of the TROSY spectra of ^{15}N-N-SH2$^{1-106}$ in the free (black) and Gab1$^{613-694}$ bound state (red).*
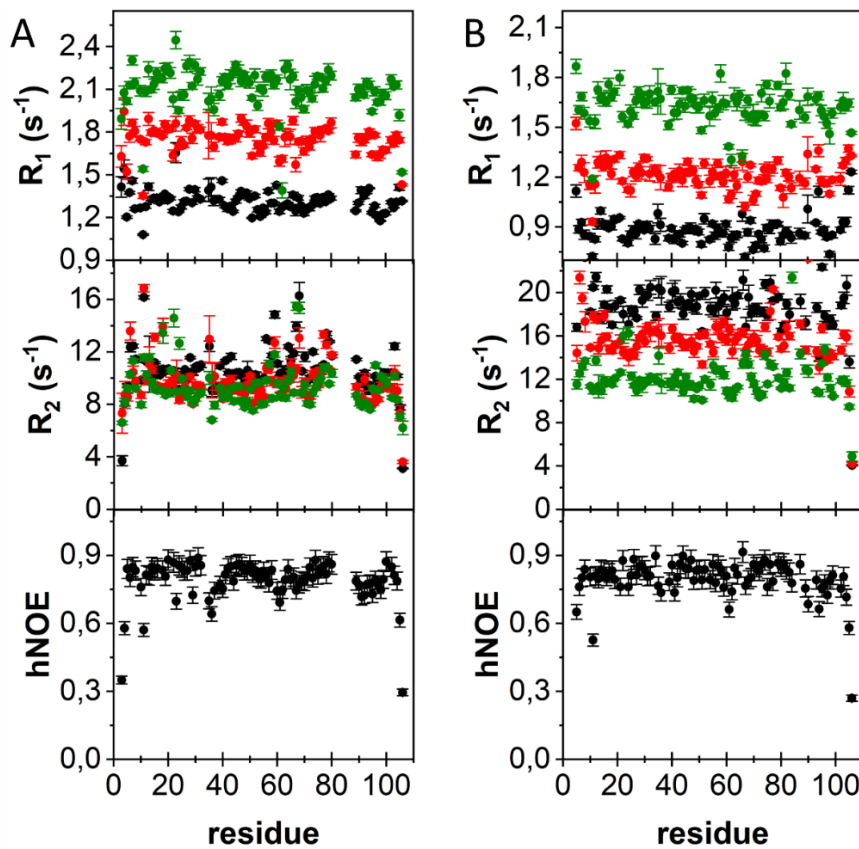
Figure 59 *R_1 and R_2 rates measured at 800 MHz, 600 MHz, 500 MHz and hNOE values from A) free N-SH2$^{1-106}$ and B) pY-Gab1$^{613-651}$ bound N-SH2$^{1-106}$ domain. Note the dependency of the R_2 rates upon addition of Gab1 on the magnetic field strength.*
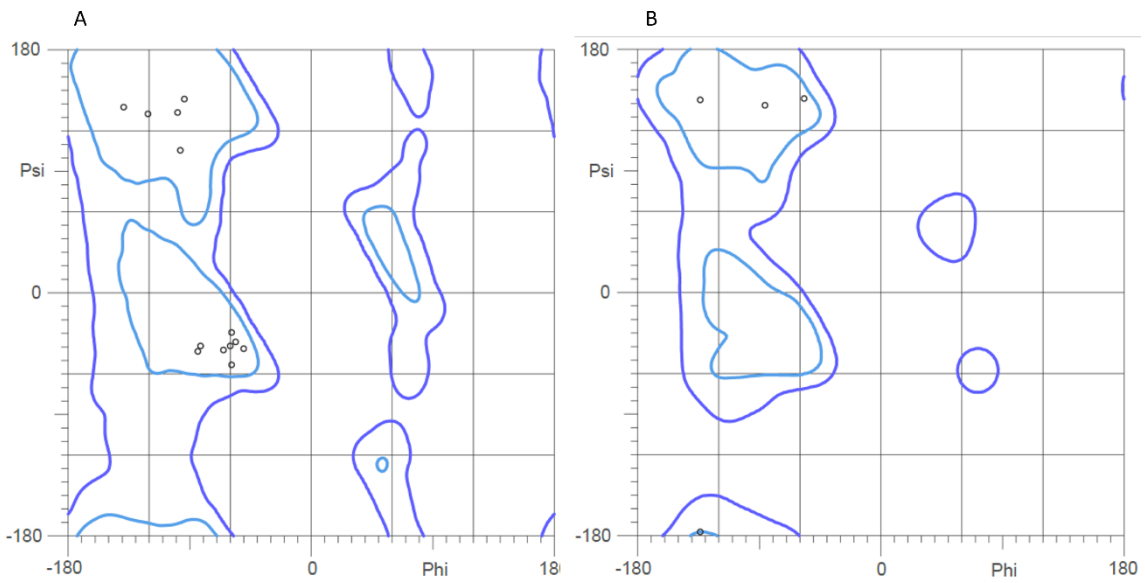
Figure 60 *Ramachandran plot of the Gab1 fragment bound to the C-SH2 domain of the Shp2^1-222^:pYpY-Gab1^617-684 structure.*

	C-SH2^102-220^:Gab1^655-666		
N	1.13; 0.93; 1.05; 0.92		
K_A [*10^4 M]	1.91; 1.50; 2.24; 1.97		
- ΔS · T	-733; -810; -1037		
ΔH	-6575; -7497; -6746; -6896		
ΔG	-5841; -5699; -5935; -5858		

	C-SH2^102-220^:Gab1^655-677		
N	1.15; 0.87; 0.97; 0.84		
K_A [*10^5 M]	2.41; 2.55; 2.16; 2.16		
- ΔS · T	-605; -109; -831		
ΔH	-6450; -7982; -7388; -8108		
ΔG	-7341; -7376; -7278; -7276		
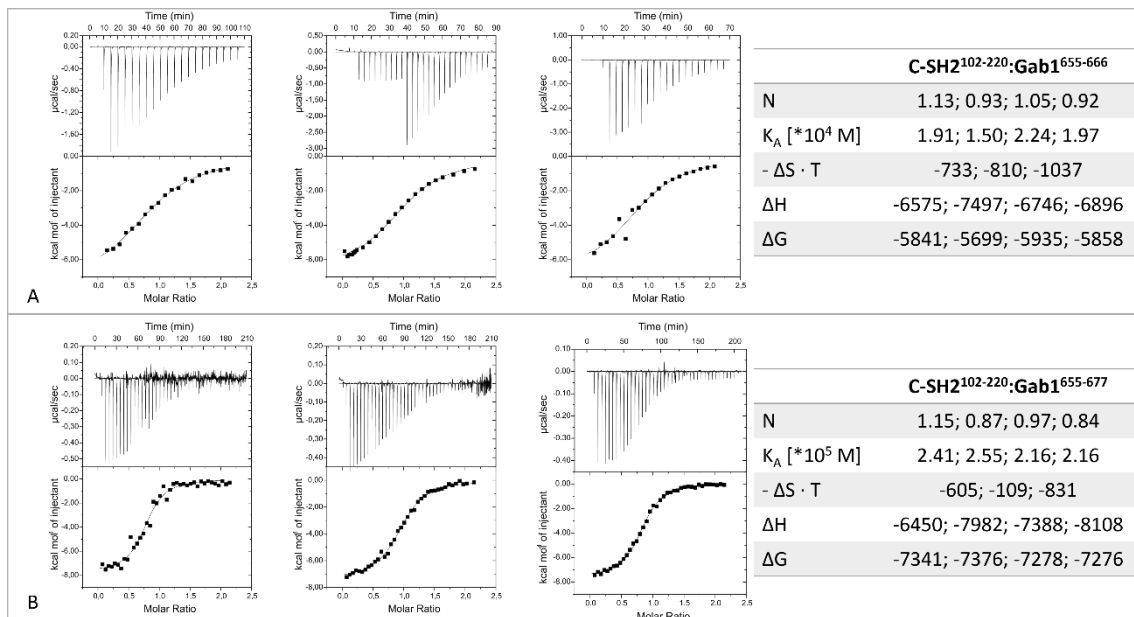
Figure 61 *Replicates of the ITC measurements of the C-SH2^102-220 with pY-Gab1^655-666 (A) and pY-Gab1^655-677 (B).*

131
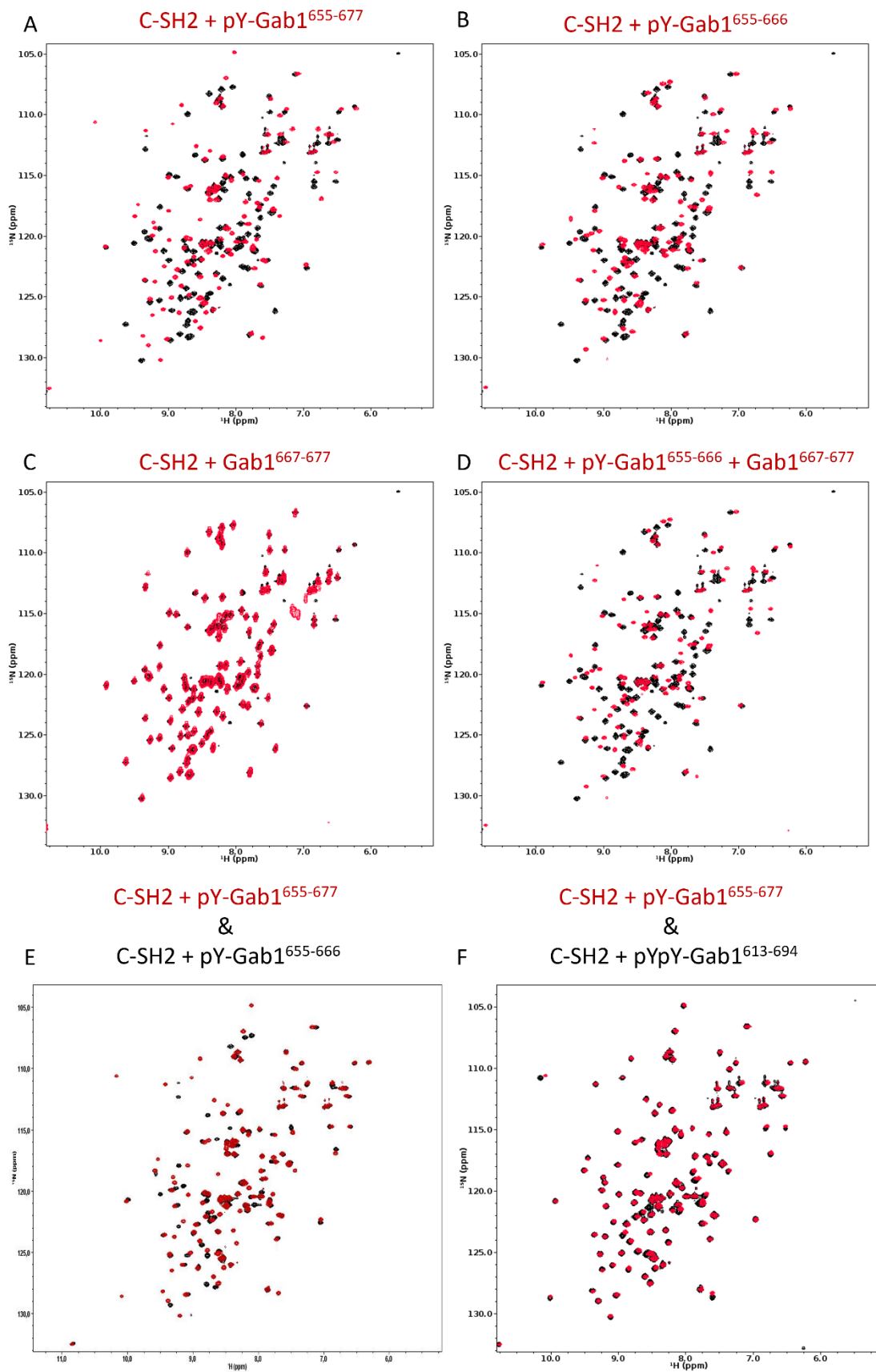
Figure 62 *HSQC-Spectra of free C-SH2 (black) and Gab1-fragment bound C-SH2 (red). The free C-SH2$^{102-220}$ is coloured in black, while the C-SH2$^{102-220}$ bound to A) pY-Gab1$^{655-677}$ B) pY-Gab1$^{655-666}$ C) Gab1$^{667-677}$ D) pY-Gab1$^{655-666}$ and Gab1$^{667-677}$ in red. E) Overlay of the pY-Gab1$^{655-677}$ (red) and pY-Gab1$^{655-666}$ (black) bound states of the C-SH2$^{102-220}$ to show deviation between both spectra. F) Overlay of the C-SH2$^{102-220}$ domain bound to pY-Gab1$^{655-677}$ (red) and to doubly phosphorylated pYpY-Gab1$^{613-694}$ (black).*

132
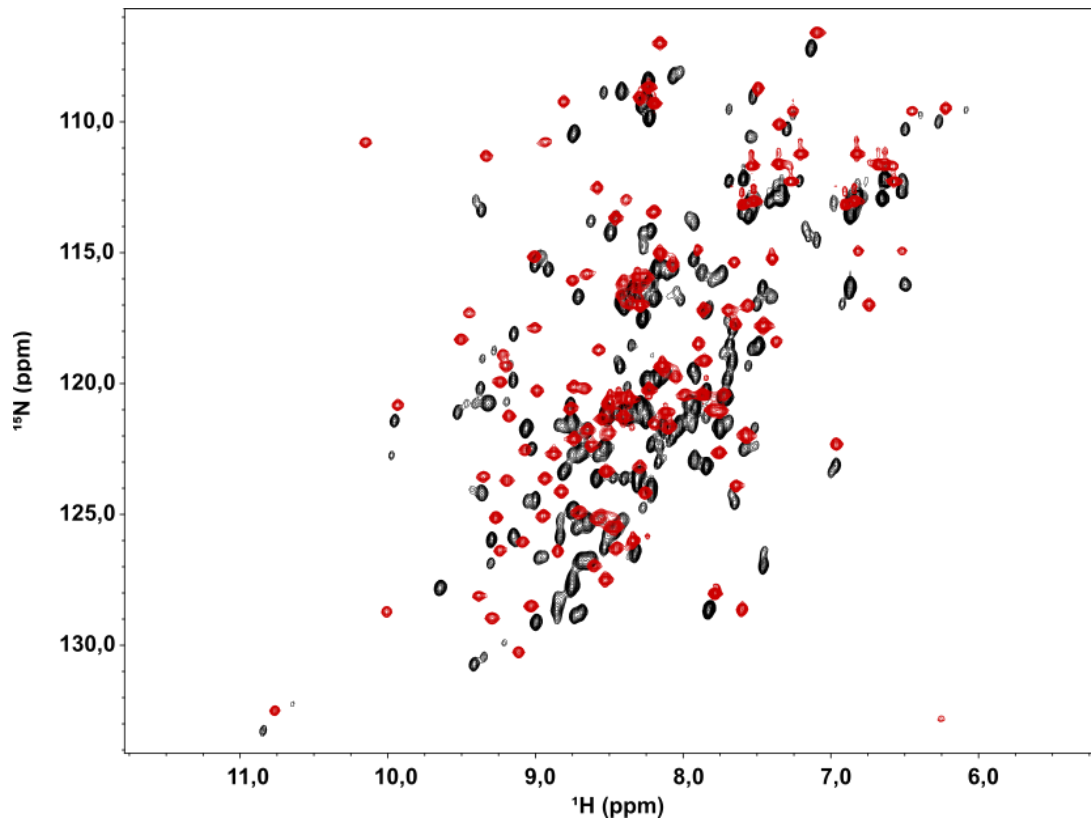
Figure 63 TROSY of ^{15}N C-SH2$^{102-220}$ in the free (black) and pYpY-Gab1$^{613-694}$ bound state (red).
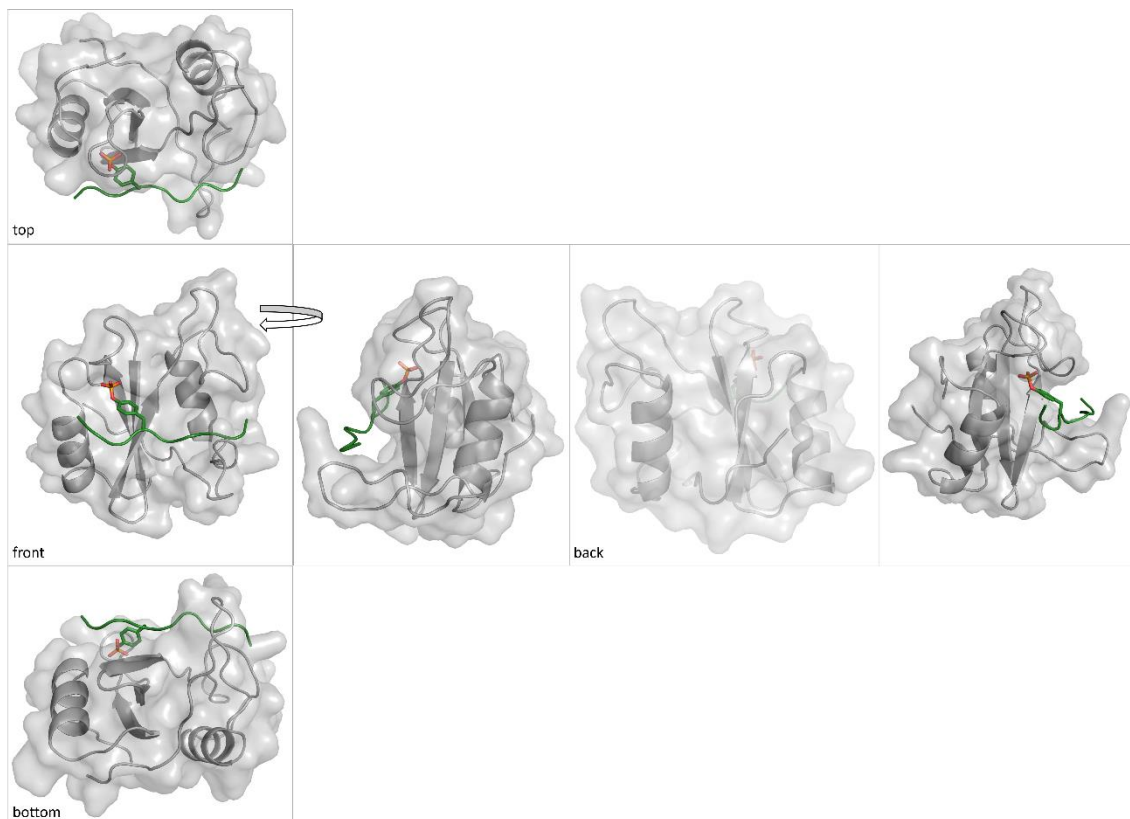
Figure 64 Faces of the SH2-domain. N-SH2 (grey, cartoon and surface) bound to Gab1 (green). The binding site is the 'front' of the domain; accordingly, the other faces in counterclockwise rotation are depicted.
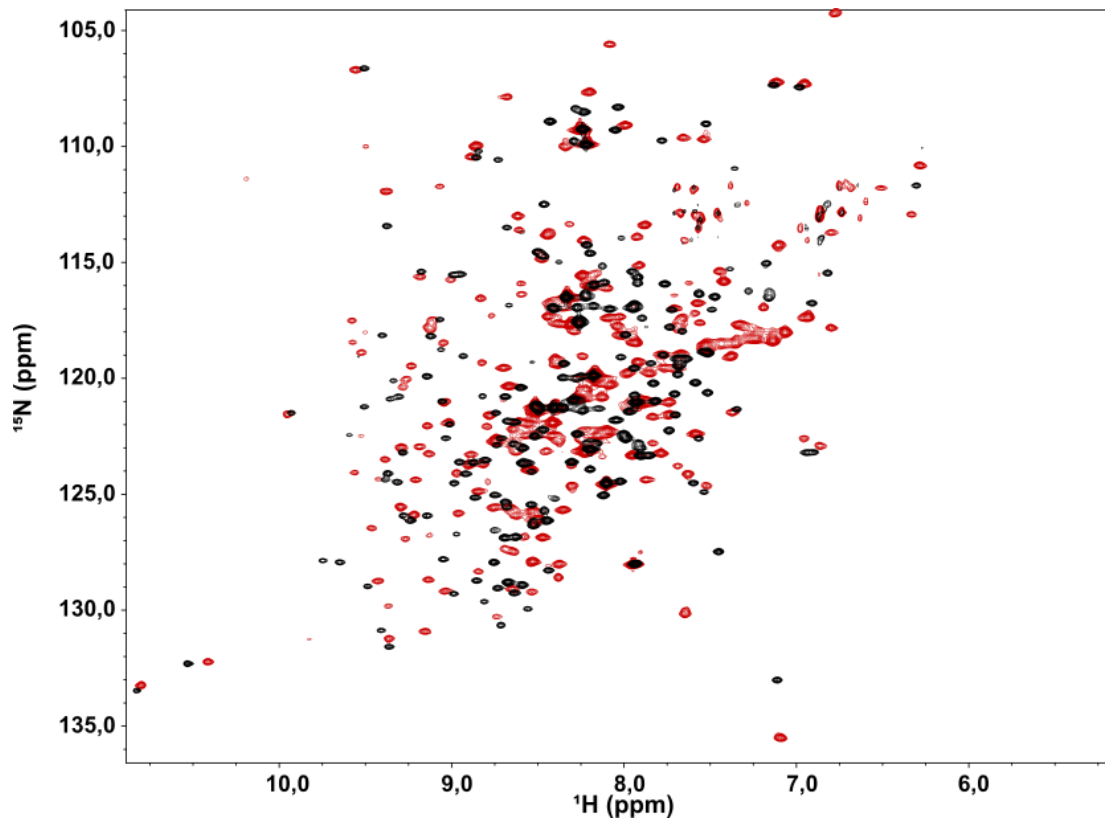
Figure 65 *Superposition of the ^{15}N-TROSY of Shp2$^{1-222}$ in the free (black) and Gab1$^{613-694}$ bound state (red).*
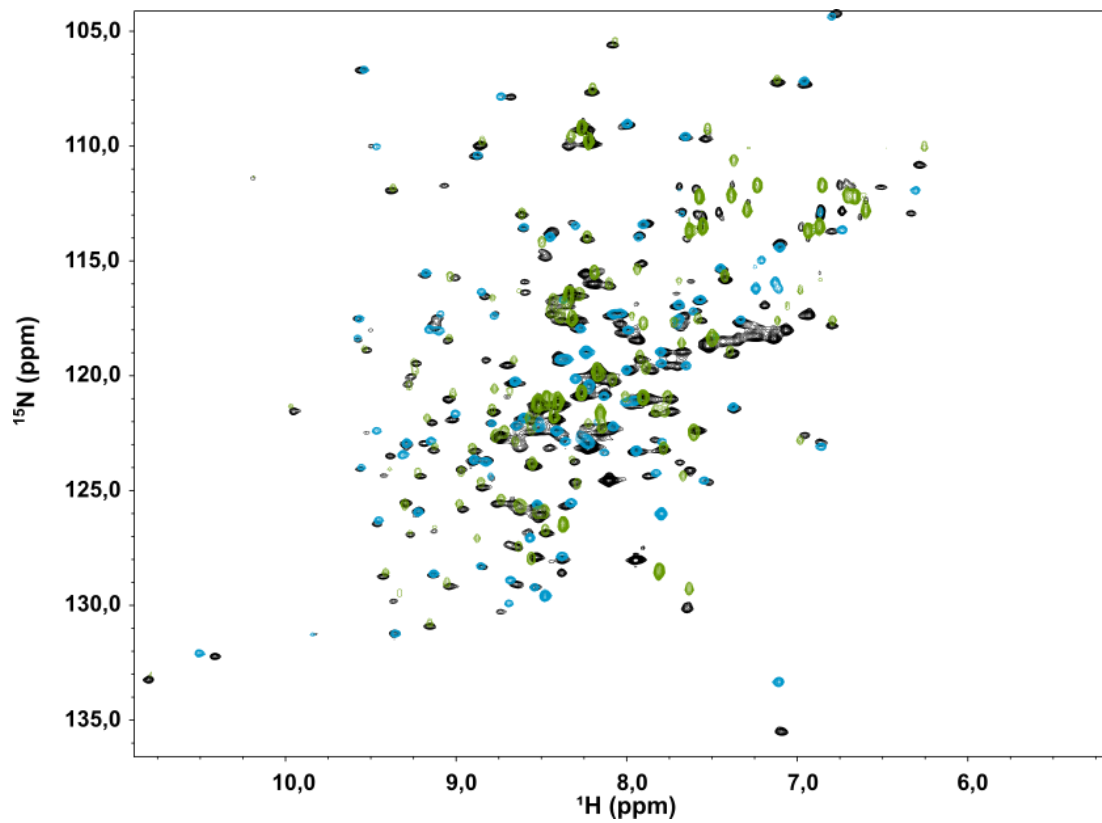
Figure 66 *Superposition of the TROSY spectra of the bound Shp2$^{1-222}$ (black) with the Gab1$^{613-694}$ bound N-SH2$^{1-106}$ (blue) and C-SH2$^{102-220}$ domain (green).*
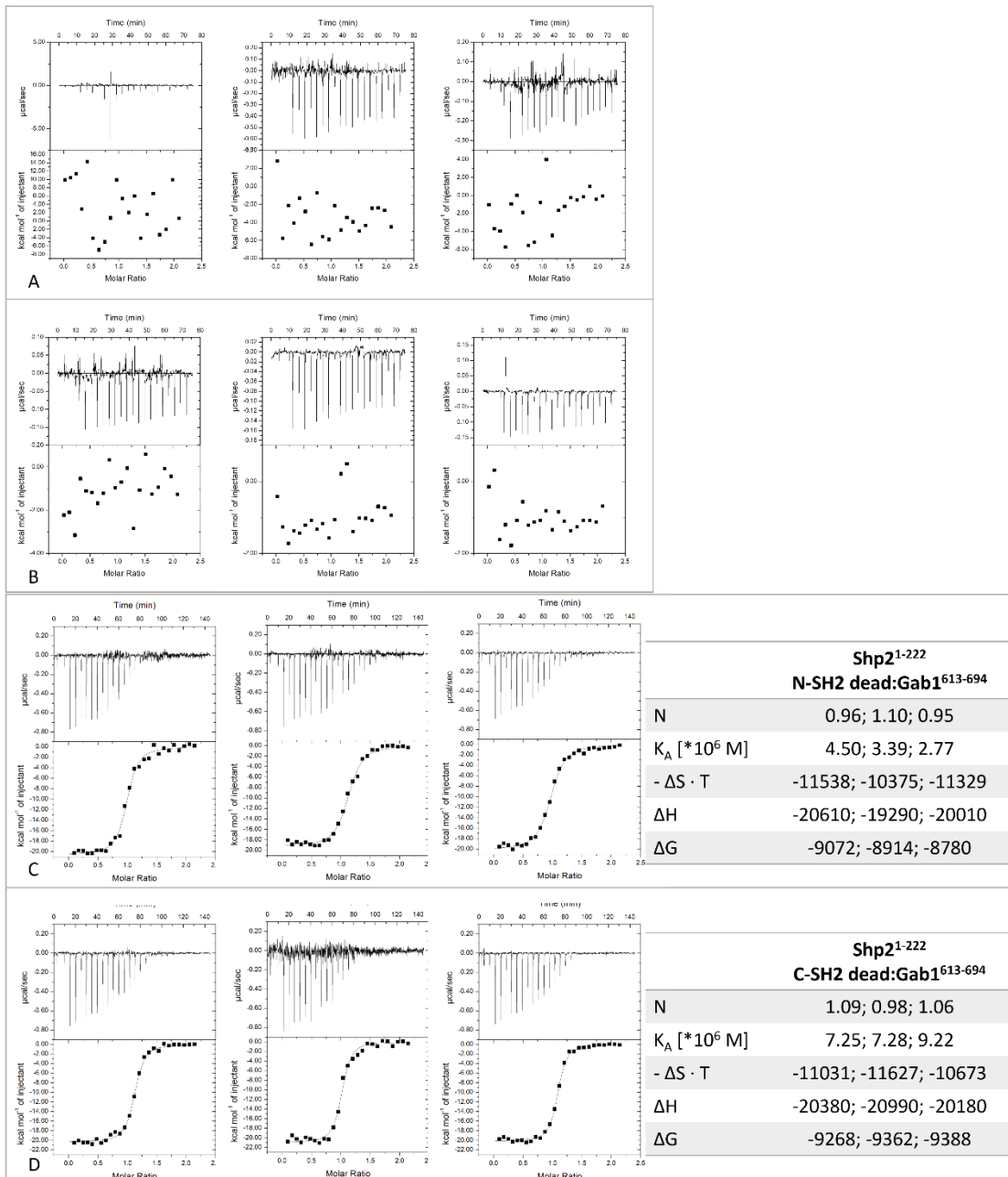
Figure 67 Replicates of ITC measurements of A) N-SH2[1-106 R32A H53A] B) C-SH2[102-220 R138A H169A] and pYpY-Gab1[613-694], as well as C) Shp2[1-222] N-SH2 dead and D) Shp2[1-222] C-SH2 dead with pYpY-Gab1[613-694].
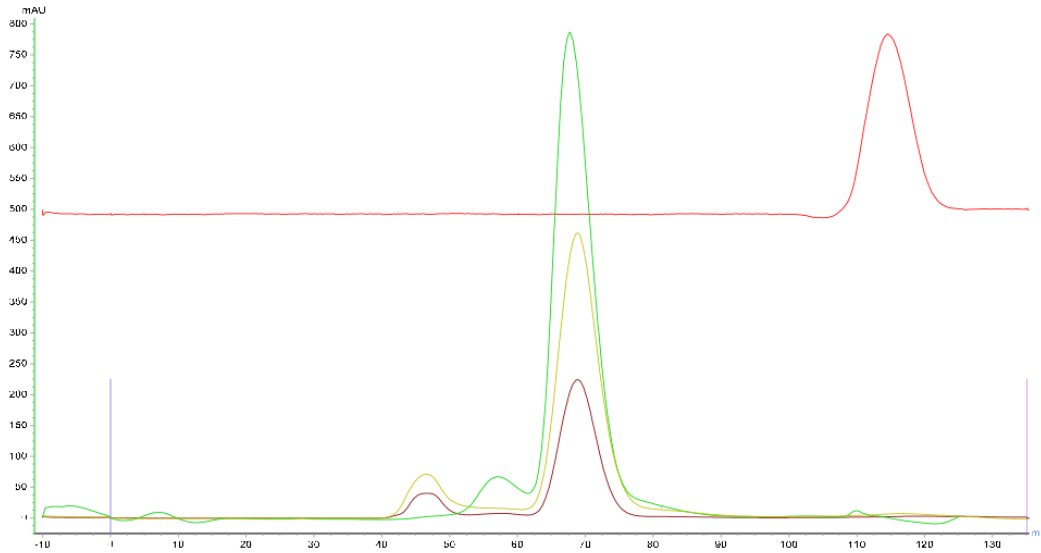
Figure 68 *Overlay SEC-elution profile (peak 65-75 ml) of the Shp2$^{1-222}$ constructs WT (green), N-SH2 dead (and) and C-SH2 dead (brown); conductivity in red.*
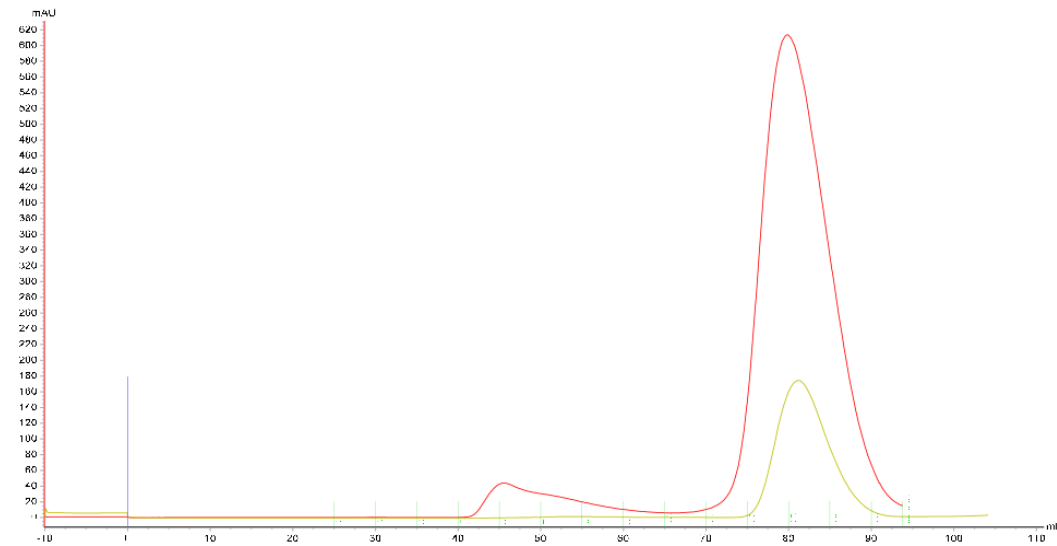
Figure 69 *Overlay SEC-elution profile of the single N-SH2$^{1-106}$ WT (red) and the dead construct (sand).*
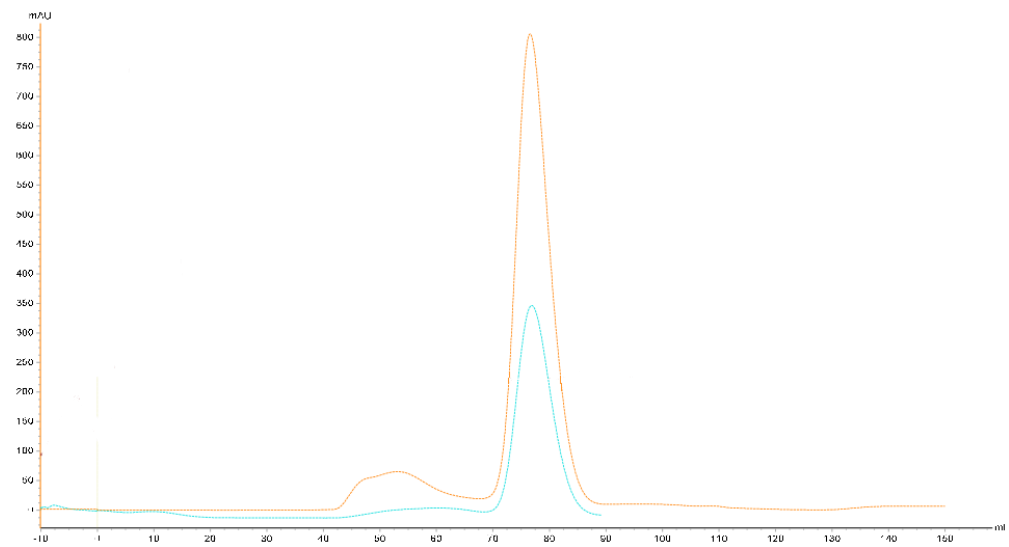
Figure 70 *Overlay SEC elution profile of the single C-SH2$^{102-220}$ WT (orange) and the dead construct (blue).*

Table of Figures and Tables

List of abbreviations

3C protease	Precision protease
Å	Angstrom
aa	Amino Acid
APS	Ammonium persulphate
CD	Circular dichroism
CSI	Chemical Shift Index
CSP	Chemical Shift Perturbation
Da	Dalton
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
E. coli	Escherichia coli
EDTA	Ethylene diamine tetraacetic acid
EGFR	Epidermal growth factor receptor
EM	Electron microscopy
Gab	Grb2-associated binder (Gab) protein
Grb2	Growth factor receptor-bound protein 2
GSH	Gluthathione sepharose
GST	Glutathione S-transferase tag
HGF	Hepatocyte growth factor
His$_6$-tag	Hexahistidine tag
hNOE	heteronuclear nuclear Overhauser effect
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl-β -D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
KD	Dissociation constant
LB	Luria-Bertani medium
microED	Electron diffraction on micro crystals
MoRF	Molecular recognition features
MW	Molecular weight
MWCO	Molecular weight cut off
N	Stoichiometry
NMR	Nuclear magnetic resonance
N-SH2/C-SH2	N-terminal and C-terminal SH2 domain of Shp
OD$_{600}$	Optical density at a wavelength of 600 nm
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein Data Bank
PEG	Polyethylene glycol
PH domain	Pleckstrin homology domain
PMSF	Phenylmethansulfonyl fluoride
PTP	Phosphotyrosine Phosphatase domain
pY	Phosphotyrosine
rpm	Rotations per minute
RT	Room temperature
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
SH2 domain	Src homology 2 domain
SH3 domain	Src homology 3 domain
Shp2	Src homology region 2 domain-containing phosphatase 2
SLiM	Short linear sequence motifs
TB	Terrific broth medium
TEMED	Tetramethyl ethylene diamine
TEV	Tobacco etch virus
Tris	Tris-(hydroxymethyl)-aminomethan
UV	Ultraviolet
WT	Wildtype
Y	Tyrosine

Eidesstattliche Erklärung

Hiermit erkläre ich, diese Dissertation selbstständig und ohne fremde Hilfe verfasst zu haben. Es wurden nur die angegebenen Hilfsmittel und Quellen verwendet, inhaltlich oder wörtlich entnommene Stellen sind als solche markiert. Weiter erkläre ich, dass ich die vorliegende Dissertation an keener anderen wissenschaftlichen Einrichtung zur Bewerbung eines Doktorgrades eingereicht habe. Ich selbst habe mich noch nicht um einen Doktorgrad beworben.

Halle/ Saale, 21.09.2023   Lisa Machner

List of Publications

Involved during funding of the RTG2467

Machner, Lisa; Shaikhqasem, Alaa; Hamdi, Farzad; Gruber, Tobias; Breithaupt, Constanze; Kniest, Judith; Lewitzky, Marc; Parthier, Christoph; Kyrilis, Fotios; Hesselbarth, Julia; Balbach, Jochen; Kastritis, Panagiotis; Stubbs, Milton; Feller, Stephan M..: Novel open conformation of Shp2 upon binding to Gab1. Manuscript

Gruber, Tobias; Lewitzky, Marc; **Machner, Lisa**; Weininger, Ulrich; Feller, Stephan M.; Balbach, Jochen (2022): Macromolecular Crowding Induces a Binding Competent Transient Structure in Intrinsically Disordered Gab1. In *Journal of molecular biology* 434 (5), p. 167407. DOI: 10.1016/j.jmb.2021.167407.

Piersimoni, Lolita; Abd El Malek, Marina; Bhatia, Twinkle; Bender, Julian; Brankatschk, Christin; Calvo Sánchez, Jaime; Dayhoff, Guy W.; Di Ianni, Alessio; Figueroa Parra, Jhonny Oscar; Garcia-Martinez, Dailen; Hesselbarth, Julia; Köppen, Janett; Lauth, Luca M.; Lippik, Laurin; **Machner, Lisa**; Sachan, Shubhra; Schmidt, Lisa; Selle, Robin; Skalidis, Ioannis; Sorokin, Oleksandr; Ubbiali, Daniele; Voigt, Bruno; Wedler, Alice; Wei, Alan An Jung; Zorn, Peter; Dunker, Alan Keith; Köhn, Marcel; Sinz, Andrea; Uversky, Vladimir N.. (2022): Lighting up Nobel Prize-winning studies with protein intrinsic disorder. In *Cellular and molecular life sciences : CMLS* 79 (8), p. 449. DOI: 10.1007/s00018-022-04468-y.

Lauth, Luca M.; Voigt, Bruno; Bhatia, Twinkle; **Machner, Lisa**; Balbach, Jochen; Ott, Maria (2022): Heparin promotes rapid fibrillation of the basic parathyroid hormone at physiological pH. In *FEBS letters* 596 (22), pp. 2928–2939. DOI: 10.1002/1873-3468.14455.

Earlier publications

Hartlage-Rübsamen, Maike; Bluhm, Alexandra; Moceri, Sandra; **Machner, Lisa**; Köppen, Janett; Schenk, Mathias; Hilbrich, Isabel; Holzer, Max; Weidenfeller, Martin; Richter, Franziska; Coras, Roland; Serrano, Geidy E.; Beach, Thomas G.; Schilling, Stephan; Hörsten, Stephan von; Xiang, Wei; Schulze, Anja; Roßner, Steffen (2021): A glutaminyl cyclase-catalyzed α-synuclein modification identified in human synucleinopathies. In *Acta neuropathologica* 142 (3), pp. 399–421. DOI: 10.1007/s00401-021-02349-5.

Hartlage-Rübsamen, Maike; Ratz, Veronika; Zeitschel, Ulrike; Finzel, Lukas; **Machner, Lisa**; Köppen, Janett; Schulze, Anja; Demuth, Hans-Ulrich; Hörsten, Stephan von; Höfling, Corinna; Roßner, Steffen (2019): Endogenous mouse huntingtin is highly abundant in cranial nerve nuclei, co-aggregates to Abeta plaques and is induced in reactive astrocytes in a transgenic mouse model of Alzheimer's disease. In *Acta neuropathologica communications* 7 (1), p. 79. DOI: 10.1186/s40478-019-0726-2.

Köppen, Janett; Schulze, Anja; **Machner, Lisa**; Wermann, Michael; Eichentopf, Rico; Guthardt, Max; Hähnel, Angelika; Klehm, Jessica; Kriegeskorte, Marie-Christin; Hartlage-Rübsamen, Maike; Morawski, Markus; Hörsten, Stephan von; Demuth, Hans-Ulrich; Roßner, Steffen; Schilling, Stephan (2020): Amyloid-Beta Peptides Trigger Aggregation of Alpha-Synuclein In Vitro. In *Molecules (Basel, Switzerland)* 25 (3). DOI: 10.3390/molecules25030580.

Curriculum Vitae

Persönliche Daten

Name   Machner, Lisa

Ausbildung

03/2023-08/2023   Weiterbildung "LifeScience
                                    Management"; ATV GmbH"
02/2019-02/2023   Promotionsstudentin im RTG2467
10/2015-03/2018   Masterstudium "Biochemie" Martin-
                                    Luther Universität Halle-Wittenberg
10/2012-09/2015   Bachelorstudium "Biochemie" Martin-
                                    Luther Universität Halle-Wittenberg
08/2006-06/2012   Max-Delbrück Gymnasium, Berlin