universität wien

MASTERARBEIT

Titel der Masterarbeit

„Insights into the modulation of the interaction between the Unique and SH3 domains in Src Family kinases“

verfasst von

Simone Höfler, BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2015

Studienkennzahl lt. Studienblatt:	A 066 863
Studienrichtung lt. Studienblatt:	Masterstudium Biologische Chemie
Betreut von:	Prof. Dr. Robert Konrat

Contents

1 Abstract

The Fyn-related Kinase (FRK) or Rak is one of the leading members of a distinct family of non-receptor tyrosine kinases consisting of FRK, PTK6 or Brk (Breast Tumer Kinase) and SRMS (Src-related tyrosine kinase lacking C-terminal regulatory tyrosine and N-terminal myristoylation sites). In contrast to the structurally closely related Src Family Kinases (SFKs), members of the FRK/PTK6 family lack the N-terminal lipidation (myristoylation and/or palmitoylation) and also the Src-homology 4 (SH4) domain. Both families share the intrinsically disordered Unique domain, the Src-homology 3 and 2 (SH3 and SH2) domains and the C-terminal kinase domain. Previous research on the intrinsically disordered Unique domain of c-Src revealed that the Unique domain interacts with the SH3 domain and the cytoplasmatic lipid layer of the cellular membrane under the formation of a peptide loop within the Unique domain. This loop formation has also been observed in Fyn, another SFK. FRK naturally provides two cysteine residues in regions of the Unique domain that are involved in the loop formation in c-Src and which could therefere be involved in a loop formation by metal ion chelation or other mechanisms. In this work the first expression and purification of the N-terminal part of human FRK, containing Unique and SH3 domain, referred to as FRK(USH3), is described. A complete protocol involving expression in *E. coli*, successful refolding of an insoluble expressed construct and the purification of an aggregation prone protein for analysis by NMR spectroscopy was developed. Furthermore the first ^{1}H-^{15}N two-dimensional NMR spectrum of FRK(USH3) is reported.

2 Introduction

2.1. Non-receptor tyrosine kinases

Within the big enzymatic class of transferases, protein kinases constitute one of the most prominent subgroups. Kinases themselves can primarily be categorized based on the amino acid they transfer the phosphate group to. Therefore they can be separate into Serine/Threonine protein kinases and Tyrosine protein kinases. The latter, as well as the former, can further be distinguished based on whether the kinase acts as a transmembrane receptor or is located in the cytoplasma and functions as an intracellular receptor and/or as signal transducer. Relating to Tyrosine protein kinases (TPKs) this classification results in in the distinctions between Receptor Tyrosine Kinases (RTKs) and non-receptor Tyrosine Kinases (nRTKs).

2.1.1. Src Family Kinases (SFKs)

One of the best studied nRTKs is the human proto-oncogene c-Src, which is also the leading member of a family of closely related kinases, the Src Family Kinases (SFKs). The SFKs comprise eight known members which share the same modular structural composition and significant similarity in their primary sequences. Figure 2.1 shows a phylogenetic tree calculated based on the primary sequences (generated with *Clustal Omega* and *CustalW2 2013 Phylogeny*) and Figure 2.2 shows a graph illustrating features that distinguish the different kinases (Goujon *et al.* 2010; W. Li *et al.* 2015; McWilliam *et al.* 2013; Sievers *et al.* 2011). Both Figures all consider the later discussed FRK/PTK6 kinase family. The described members of the SFKs include c-Src, Yes, Fyn, Lyn, Lck, Hck, Blk and Fgr. C-Src, Yes and Lyn are ubiquitiously expressed though expression levels are found to be increased in certain tissues for each of the kinases. The remaining members are primarily expressed in hematopoietic cells.

All SFKs share a common modular structural composition comprising a N-terminal Src-homology (SH) 4 domain, followed by an intrinsically disordered Unique domain, an SH3 domain ,an SH2 domain, a kinase or SH1 domain and a C-terminal regulatory tail (Figure 2.3)(Thomas & Brugge 1997).

The unfolded SH4 domain contains a Glycine in position 2 which is myristoylated in the cellular context and directs the kinase to the cytoplasmatic membrane and all SFKs, excluding Src and Blk, are further subject to palmitylation on conserved Cysteine residues (Thomas & Brugge 1997). Nevertheless the myristoyl group alone doesn't seem the be sufficient to to ensure the localization of human c-Src to the membrane, but is necessary for functional activity within the cell (Roskoski 2015). Further is has very recently been shown, that the SH4 domain of the human c-Src is also involved in a direct interaction with the SH3 domain (Maffei *et al.* 2015).

The Unique domain is the 50-70 residues long flexibel linker between the SH4 domain (membrane

anchor) and the SH3 domain. It owes its name to the unique primary sequence in each of the SFKs. Nevertheless this intrinsically disordered domain contains several known phosphorylation sites in Lck, Hck, Lyn, Yes, Fgr, Fyn and Src, for some of which also a function is annotated (Amata *et al.* 2014). The Unique domain of the human c-Src was found to contain a partially structured region (residues 60-75), which was only recently found to have lipid binding properties and habors a region called the Unique Lipid Binding Region (UBLR) (Pérez, Gairí, *et al.* 2009). The ULBR was also reported to be be involved in an allosteric regulation involving the neighbouring SH3 domain and therefore seems to directly interact with it respectively. (Maffei *et al.* 2015; Pérez, Maffei, *et al.* 2013). Also the Unique domain of Lck (lymphocyte cell-specific protein tyrosine kinase) was found to form a ternary complex with either CD4 or CD8 T cell coreceptor. The ternary complex is based on the complexation of a Zn^{2+} ion involving two conserved N-terminal Cystein residues from each Lck and CD4 or CD8. This interaction between Lck and the T cell coreceptors is needed for normal T cell development and antigen-depended activation of mature T cells (Kim *et al.* 2003)

The Unique domain is followed by an SH3 domain which is composed of about 50 amino acids and can bind sequences that can adopt a left-handed helical conformation. The SH3 domain of the human Src consists of five anti-parallel β- sheets and two prominent loops, the RT and the n-Src loops. These loops are positioned at either end of the surface and contain mostly hydrophobic and aromatic residues, which form the recognition site for proline-rich sequences bearing the 'PxxP' motif. This sequence is abel to adapt a polyproline type II helical conformation, that forms a complex with the aromatic residues in the n-Src (Boggon & Eck 2004; Roskoski 2015).

Positioned C-terminal of the SH3 domain is the SH2 domain. This non-catalytic domain consists of approximately 100 amino acids and folds into a tertiary structure comopsed of a central anti-parallel β- sheet flanked by an α-helix on either side of the β- sheet. This domain has the ability to specifically bind phosphorylated tyrosine residues (pY), whereby the specificity depends on the sequence directly C-terminal (3-5 residues) of the pY. The preferential recognition/binding motif for SFKs SH2 is pTyr - Glu - Glu - Ile (pYEEI), but many other amino acids can inhabit the positions Y+1 and +2. The binding of the pY is mediated by two binding pockets , the pY binding pocket and a hydrophobic binding pocket.The pY binding pocket contains a conserved Arginine residue, which is essential for pY binding, as was shown by mutational studies (Sawyer 1998). The hydrophobic binding pocket contacts the Y+3 residue (Ile) (Boggon & Eck 2004; Thomas & Brugge 1997).

The SH2 domain is connected to the kinase or SH1 domain via a Linker domain which forms, even though it only contains one single proline residue a polyproline type II helix (Boggon & Eck 2004). The kinase domain exhibits a bilobal protein kinase fold. The N-terminal lobe consists of five β- strands and one single α- helix (which is referred to as C helix in Src) and the C-terminal lobe is mostly α-helical and contains the activation loop. The nucleotide binding and the phosphotransfer happens in the cleft between the two lobes (Boggon & Eck 2004) .

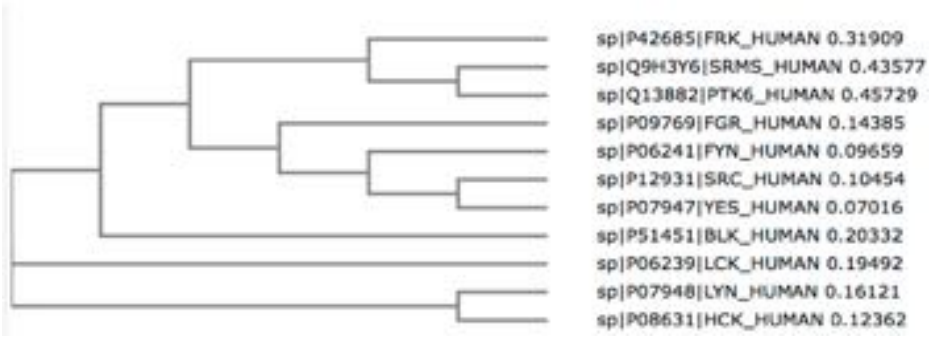


Figure 2.1: Phylogentic tree of the Src Family Kinases (SFKs) and the FRK/PTK6 Family generated based on the primary sequence of all proteins considered (Sequence Alignment Tools: Custal Omega and CustalW2 2013 Phylogeny)

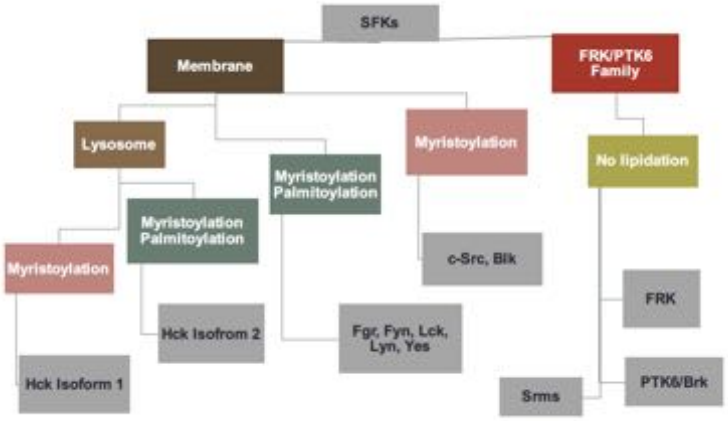


Figure 2.2: Tree diagram displaying SFKs and FRK/PTK6 kinases according to differentiating properties.

Regulatory mechanisms

Almost all SFKs are characterized as proto-oncogenes, therefore a tight regulation of their catalytic activity is of essential importance. Representative for the SFKs crystal structures of the active and the inactive conformations of c-Src (Williams *et al.* 1997; Xu, W *et al.* 1997) and Hck (Sicheri, F *et al.* 1997) have shed light on essential features of SFK regulation. The mentioned X-ray structures reveal that both the SH2 and the SH3 domain play a central role in assembling and stabilizing the inactive conformation of the kinase. In this conformation both domains are packed against the kinase domain on the opposite side of the catalytic center. On the one hand the SH3 domain binds to the left-handed

polyproline type II helix formed by the linker domain between the SH2 and the kinase domain and two regions flanking the hydrophobic binding areas of SH3 seem the contact the catalytic domain at the N-terminal lobe. The SH2 domain on the other hand binds the pY (in Src Y530) and the directly adjacent residues in the regulatory C-terminal tail. The interaction of the SH2 domain with the pY positions the SH2 domain close the the C-terminal lobe of the kinase domain on the opposite side of the catalytic cleft. These respective tyrosine residues in Src and the other SFKs are the primary sites of phosphorylation in vivo by the cytosolic tyrosine kinase Csk. The importance of this phosphorylation and the binding of the SH2 domain for SFKs regulation is supported by evidence showing that either a mutation of the tyrosine or its deletion are sufficient to activate the kinase (Cartwirght, CA *et al.* 1987; Reynolds *et al.* 1987). Further studies have also discovered that a disruption of the *csk* gene leads to an activation of at least three SFKs (Src, Fyn and Lyn) *in vivo* (Imamoto & Soriano 1993; Nada, S *et al.* 1993). Another phosphorylation that proved to be important for regulation of the kinase activity is concerned with a tyrosine residue within the catalytic domain. In c-Src this corresponds to Y419. In the oncogenic variants of Src this residue is constitutively phosphorylated, whereas it is not phosphorylated in inactive wild type Src (Thomas & Brugge 1997).

Summarizing the the interaction between the SH3 domain and the linker domain and the SH2 domain and the pY at the exposed C-terminus lock the catalytic domain in an inactive conformation, as they result in several structural adaptations within the kinase domain. *(1)* The C helix in the N-terminal lobe is displaced, which results in the removal of a catalytically important glutamic acid from the active side, *(2)* the activation loop adapts an α- helical conformation that excludes peptide binding and sequesters the tyrosine residue which is subject to phosphorylation and *(3)* the relative orientation of the N and C lobes seems to be constrained in a way that might not optimal for catalytic activity (Boggon & Eck 2004).

Importantly these intramolecular interactions are low affinity interactions and can be easily disrupted in the present of a high affinity ligand, dephosphorylation or phosphorylation of relevant residues. The lower affinity can be accounted for in the none optimal composition of the intramolecular ligands. The polyproline type II helix in the linker domain, which can be bound by the SH3 domain, is based on only one single proline residue unlike the preferential 'PxxP motif and in the C-terminal recognition motif the Y+3 position contains a glycine (in Src, Fyn and Yes) rather then the preferred isoleucine (Boggon & Eck 2004; Thomas & Brugge 1997).

The above mentioned interactions implicate that there are several possible mechanisms to activate SFKs. These include disturbance of intramolecular interactions of SH3 and SH2 domain by higher affinity ligands or dephosphorylation as well as phosphorylation of relevant tyrosine residues, which lead to the adaption of an open or active conformation of the kinase. Any or a combinations of this activation mechanisms can accompany the activation of a SFK by a potential binding partner.

Biological functions

Each of the Src Family Kinases (SFKs) are involved in a plethora of signaling pathways, from which only a fraction is known or described. Investigations into their interactions partners and targets showed that SFKs couple with many different receptor pathways, including immune recognition receptors (IRRs) and components of the major histocompartibility complex (MHC), integrins, receptor

tyrosine kinases (RTKs), G-protein coupled receptors, cytokine receptors and membrane channels (Thomas & Brugge 1997).

The prototypical IRRs are B-cell (BCR) and T-cell receptor (TCR/CD3) which, regardless of their different extracellular antigen-binding regions, both contain a immunoreceptor tyrosine-based activation motif (ITAM). This motif is defined by a consensus sequence Y-x-x-I/L-$_{(7\text{-}8)}$-Y-x-x-I/L . The tyrosine residues within the ITAM get phosphorylated upon receptor engagement and provide an important anchor for effector proteins, especially kinases, which contain SH2 domains (Isakov 1997). The antigene recognition in T-cells involves TCR/CD3 in complex with one of two co-receptors CD4 or CD8, which bind to MHC class II or class I proteins respectively (Davis & Bjorkman 1988). In the TCR-based signal transduction the SFKs Fyn and Lck are significantly involved. Fyn has been shown to coprecipitate and therefore to somehow interact with TCR/CD3 (Da Silva, AJ *et al.* 1992), whereby the first 10 amino acids of FynT (Fyn kinase variant expressed in T-lympocytes) are sufficient to interact with the ζ- subunit of TCR/CD3 (Gauen, Timson L. K. *et al.* 1992). More recent studies further revealed that Fyn is important for an optimal humoral response and that the none-catalytic functionalities of FynT and Pyk2 are involved in the late stage adhesion of T-lymphocytes after activation (Chaimowitz *et al.* 2013; Chapman *et al.* 2012). Lck on the other hand has been shown to directly interact with either one or the other co-receptor CD4 and CD8. This interactions is based on the chelation of a Zn^{2+} ion by four cysteine residues, where Lck provides two cysteines in the Unique domain and the other two residues are provided by the respective co-receptor (Kim *et al.* 2003).

Integrins are heterodimeric transmembrane receptors that mediate cell-matrix and cell-cell interactions and are involved in the regulation of cell adhesion and spreading, migration, proliferation, differentiation, etc.. c-Src is transiently activated upon engagement of integrins caused by cell attachment to a fibronectin matrix and recruited to focal adhesions (Kaplan, Kenneth B. *et al.* 1995). Though not fully understand the activation of Src is preceded by the dephosphorylation of Y530 by tyrosine phosphatases that might either be activated and/or relocalizied through the activation of intergin receptors. The localization of Src kinase to focal adhesions required only the presence of the N-terminal myristoylation and the SH3 domain (Kaplan, Kenneth B *et al.* 1994), but the association of Src with components of the intergrin complexes which harbor a tyrosine phosphorylation might involve the SH2 domain. One of these components is the focal adhesion kinase (FAK).

A previously mentioned class of protein tyrosine kinases are the receptor tyrosine kinases (RTKs). They comprise a diverse group of receptors, like epidermal growth factor receptors (EGFRs), fibroblast growth factor receptors (FGFRs), platelet-derived growth factor receptors (PDGFRs), ephrin (Eph) receptors and insulin receptor, etc.. In this class of receptors the binding of the ligand (soluble or membrane bound) causes the to receptor to dimerize. Following the dimerization the receptors phosphorylate each other and some intercellular effector proteins on certain tyrosine residues. Phosphorylated tyrosine residues outside the receptors cytosolic kinase domain create high-affinity docking sites for specific intercellular signaling proteins. This proteins, which bind to pY residues, usually contain a highly-conserved pY binding domain, which is either a SH2 domain or a PTB domain (phosphotyrosine-binding) (Alberts *et al.* 2008). SFKs are a group of effectors proteins that work downstream of RTKs. SFKs were shown to by phosphorylated by RTKs, phosphorylate RTKs, bind to RTKs and get activated by RTKs (Thomas & Brugge 1997). Src, Fyn and Yes can associate directly with PDGFRβ (Kypta, Robert M *et al.* 1990). The SH2 domain of c-Src binds to two pY residues in the

juxtamembrane region of the PDGFRα/β (Mori *et al.* 1993). Binding of the SH2 domain to pY disturbs the intramolecular interactions that stabilize the closed/inactive conformation of c-Src and could be the initial step of activation. Further Src is also phosphorylated on two serine residues and one tyrosine residue in the N-terminal non-catalytic part and one tyrosine residue in the catalytic domain by the PDGFR kinase domain (Broome & Hunter 1997).

G protein-coupled receptors (GPCRs) form the largest family of cell-surface receptors and recognize a plethora of different signals from the environment as well as signals from other cells. Nevertheless they share a fairly well conserved structure composed of a single polypeptide chain that crosses the cellular membrane seven times. The are coupled to G-proteins to transduce the signal into the interior of the cell. A heterotrimeric GTP-binding protein (G-protein) is activated upon ligand-binding induced structural rearrangements of the GCPR . The G-proteins consist of a α-, β- and γ- protein subunit. The activated GCPR acts as a guanine nucleotide exchange factor (GEF), which causes the α- subunit to release a bound GDP and allows for GTP binding. GTP binding induces a large conformational change and leads to the dissociations of the $\beta\gamma$- complex from the α- subunit. The α- subunit is a GTPase and its activity determines the time the G-protein stays active, as hydrolyzation of the bound GTP to GDP leads to inactivation of the complex. Both the α- subunit and the $\beta\gamma$- complex can regulate the activity of enzymatic effectors (Alberts *et al.* 2008). Src, Fyn and Yes have been shown to be activated by certain GCPRs and though the activation mechanism are not thoroughly understood in detail yet several mechanism permit GCPRs the activation of SFKs. This mechanisms comprise direct association of SFKs with GPCRs or receptor-associated proteins and transactivation of RTKs and focal adhesion complexes by GPCRs-mediated signals (D. K. Luttrell & L. M. Luttrell 2004; S. J. Parsons & J. T. Parsons 2004).

The class of cytokine receptors transduces signals from many kinds of local mediators, including cytokines, hormones and prolactin. This type of receptors does not contain any intrinsic tyrosine kinase activity, but is stably associated with cytoplasmic tyrosine kinases called Janus kinases (JAKs). Binding of a ligand crosslinks two separate receptor polypeptide chains to from a dimer, which brings the associated JAKs in close proximity and they first cross-phosphorylate each other and then the receptor on certain tyrosine residues. The pY on the receptor provide docking platforms for the SH2 domain of STATs (***s**ignal **t**ransducer and **a**ctivator of **t**ranscription*), which can get phosphorylated and activated by JAKs. Activated STATs form homo- and heterodimers and translocate into the nucleus, where the associated with other regulatory protein to regulate gene transcription (Alberts *et al.* 2008). In has been described in some cases that STATs can also be activated by SFKs, this happens mostly in a growth factor receptor - dependent manner (Silva 2004). Nevertheless a yet undefined role of Lck as been characterized in the interleukin-2 (IL-2) receptor signaling and cells transformed with v-Src show constitutively tyrosyl-phosphorylated STAT3 by v-Src in a JAK-independent manner (Hayakawa & Naoe 2006).

The last major receptor type which is coupling with SFKs are membrane channels. Channels function as important regulators of efflux and influx of small molecules and ions through the cell membrane. SFKs have been shown to couple with several ligand-gated and voltage-gated membrane channels, including K^+-channels, Ca^{2+}-channels and glutamate and *N*-acteylcholine receptors (Thomas & Brugge 1997).

Once SFKs are activated they transduce the received signal by phosphorylating other effector proteins on specific tyrosine residues or by kinase-independent mechanisms. Due to the plethora of pathways SFKs are involved in the number of phosphorylated target proteins ist vast. Some of the targets that have been identified for Src include FAK, tubulin, the protein serine/threonine kinase PKCσ, PLC-$\gamma$and protein phosphatases SHP-1 and SHP-2 (Thomas & Brugge 1997).

The nature of the SFK targets cause these kinases to be involved in cellular processes like adhesion and spreading of cell onto and along a fibronectin coated surface via an integrin meditated pathway, the formation and disassembly of focal adhesion (Src and Fyn), cell migration, cell cycle progression/cell proliferation and gene transcription just to name a few (D. K. Luttrell & L. M. Luttrell 2004; S. J. Parsons & J. T. Parsons 2004; Playford & Schaller 2004; Silva 2004; Thomas & Brugge 1997).

2.1.2. FRK/PTK6 family

The non-receptor tyrosine kinases of the FRK/PTK6 family are distantly related to the SFKs (Brauer & Tyner 2009). This tyrosine kinase family consists of three known members in vertebrates FRK (Fyn-related kinase) or Rak, PTK6 or Brk (Breast Tumor Kinase) and SRMS (Src-related tyrosine kinase lacking C-terminal regulatory tyrosine and N-terminal myristoylation sites) (Cance *et al.* 1994; Kohmura *et al.* 1994; P. Mitchell *et al.* 1994). They share the same modular structural composition consisting of a N-terminal Unique domain, followed by a SH3 and SH2 domain and a C-terminal Kinase domain (Figure 2.3)(Brauer & Tyner 2009; Serfas & Tyner 2003). The most remarkable feature that distinguishes the FRK/PTK6 from the SKFs is the of lack the N-terminal lipidation and the SH4 domain, which makes them not target to localization at the membrane but gives them some sort of flexibility in the their subcellular localization (Figure 2.2) (Cance *et al.* 1994; Goel, Miah, *et al.* 2013; Serfas & Tyner 2003). Besides the lack of a lipidation this kinases also share a common exon-intron structure which is distinct from the SFKs (H. Lee *et al.* 1998; Serfas & Tyner 2003). In a study published in 2008 it was suggested that a ancestral FRK/PTK6 kinase duplicated and diverged to give rise to the FRK/PTK6 family and SFKs (D'Aniello *et al.* 2008).

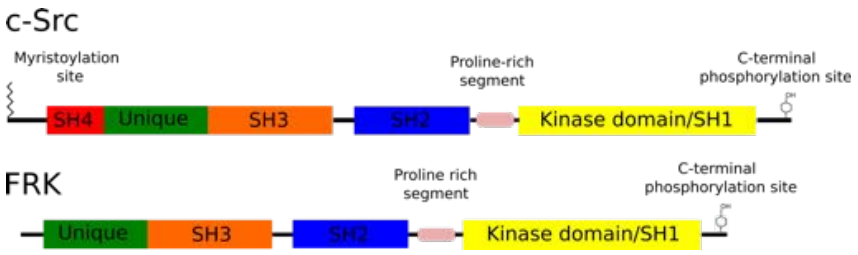


Figure 2.3: Domain structure of c-Src and FRK.

Compared to the SFKs, the FRK/PTK6 family has been subject to much less investigations and therefore not much is known about the biological function of any of the three members.

PTK6 is a moderately well studied member of the FRK/PTK6 family and was first characterized in human melanocytes, breast cancer cells and normal intestine (S. Lee *et al.* 1993; P. Mitchell *et al.* 1994; Siyanova *et al.* 1994). PTK6 has found to be expressed in 60% of breast carcinomas, whereas

it is not found in normal mammary gland, that is why it is believed to be a tissue specific oncogene (Goel & Lukong 2015; P. Mitchell *et al.* 1994). It has been shown that PTK6 sensitizes mammary epithelial cells to epidermal growth factor and therefore enhances the mitogenic response, whereas in normal tissue PTK6 promotes cellular differentiation and apoptosis (Kamalati, Jolin, Fry, *et al.* 2000; Kamalati, Jolin, P. J. Mitchell, *et al.* 1996; Serfas & Tyner 2003). Several interactions partners have been identified so far containing nuclear and cytoplasmatic proteins, which also indicates that PTK6 can localize into the nucleus and the cytoplasma. Interaction partners include signaling molecules like proto-oncogene Akt kinase and BSK, which are shown to be phosphorylated by PTK6, transcription factors like STAT3 and STAT5A/B and RNA-binding proteins like SAM68, SLM1 and SLM2 (Abeyrathna & Su 2015; Ikeda, Miyasaka, *et al.* 2009; Ikeda, Mizushima, *et al.* 2011; Lukong *et al.* 2005; Serfas & Tyner 2003).

SRMS is the least studied member of the FRK/PTK6 family, which was cloned for the first time in 1994 but was subsequently not further studied (Kohmura *et al.* 1994). Recent research revealed that SRMS is expressed in breast cancer tissue and can not be found in tissue of normal mammary gland, just like its family member PTK6 (Goel, Miah, *et al.* 2013). One special feature of SRMS is the lack of the regulatory C-terminal tail. The activity regulation by the tyrosine phosphorylation in the C-terminal tail is compensated for by the N-terminal region of SRMS (Goel, Miah, *et al.* 2013). So far only one direct substrate for SRMS has been identified, namely Dok1 (Goel, Miah, *et al.* 2013).

FRK is the leading members of this kinase family and the subject of the study presented. It was first isolated and described in primary human breast cancer cells by Cance *et al.* 1994, and in human melanocytes by S. Lee *et al.* 1993. The canonical sequence for the human FRK variant is isoform 1, which consists of 505 amino acids with a molecular weight of 58 254 Da and is encoded on chromosome 6q21-23(Cance *et al.* 1994; Yim *et al.* 2009). It is predominately expressed in epithelial-derived cell-lines and tissues, including kidney and liver, and breast and colon cell lines (Cance *et al.* 1994). The SH2 domain of human FRK contains a bipartite nuclear localization signal (NLS) (residues 168-181: **KR**LDEGG**FF**LT**RRR**), which could direct FRK to the nucleus, but so far no nuclear function of FRK has been described (Brauer & Tyner 2009; Cance *et al.* 1994). For the Unique domain two phosphorylation sites have been identified, namely on S37 and S40, and one on the interface between the Unique and SH3 domain on Y46 (Bian *et al.* 2014; Daub *et al.* 2008; Stokes *et al.* 2012). Y178, within the SH2 domain, has also been found to be subject to phosphorylation and Y387 and Y497 are thought to be the homologs to Y419 and Y530 respectively in c-Src, but these are just assumption based on sequence similarities (Bian *et al.* 2014). No structure of FRK or parts of it, neither X-ray nor NMR, has been published to this point .

Some biological functions and interaction partners have already been described for FRK and it has been shown to have a tumor suppressor function. PTEN (phosphatase and tensin homolog deleted from chromosom 10) has been shown to be phosphorylated by FRK (Yim *et al.* 2009). PTEN itself is a tumor suppressor that is often found to be mutated or deleted in different types of human cancer (J. Li *et al.* 1997).The main function of PTEN described so far is to antagonize phosphatidylinositol 3 kinase (PI3K) by dephosphorylation of PIP3 to PI(4,5)P2. The PIP3 produced by PI3K results in a recruitment of proteins containing pleckstrin-homology (PH) domains to the plasma membrane, which are responsible for the transmission of PI3K-dependent signaling(Carracedo & Pandolfi 2008). One

of these proteins is proto-oncogene Akt kinase, which gets activated at the membrane and mediates mainly responses concerned with cell growth and survival (Brauer & Tyner 2009; Carracedo & Pandolfi 2008). In case of PTEN loss, as can been seen in several human cancers, the antagonizing effect is lost which leads to a hyperactive PI3K and an upregulation of Akt activity (Carracedo & Pandolfi 2008). FRK has been shown to physically interact with PTEN and phosphorylate it on Y336 (Yim *et al.* 2009). This phosphorylation leads to an increase of PTEN stability as it prevents it binding to the ubiquitin E3 ligase NEDD4-1 , which can either lead to mono- or polyubiquitinated PTEN, with monoubiquitinated protein being translocated to to nucleus and polyubiquitinated being send for proteosomal degradation (Yim *et al.* 2009). Therefore phosphorylation leads to more stable PTEN which accumulates in the cytoplasma and can antagonize PI3K and therefore inhibit Akt activity (Brauer & Tyner 2009; Carracedo & Pandolfi 2008; Yim *et al.* 2009). Consequently FRK is supposed to inhibit cell growth when expressed and promote cell growth when deleted or mutated.

Furthermore FRK has been shown to associate with retinoblastoma (pRb) tumor suppressor via the A/B binding pocket of pRb *in vivo* and *in vitro* (Craven *et al.* 1995). This association is happing in the G_1 and S phase of the cell cycle, while FRK levels are highest during G_1 phase. Binding of FRK to pRb does not activate the kinase (Craven *et al.* 1995). FRK has also an growth inhibitory effect when overexpressed in NIH 3T3 cells and two different breast cancer cells lines and causes G_1 arrest in the cell cycle, which is independent of its association with pRb (Craven *et al.* 1995; Meyer *et al.* 2003).

Even though a tumor suppressor role is very strongly predicted for FRK no evidence was found in mice lacking FRK (Chandrasekharan *et al.* 2002). For *FRK -/-* mice only a mild phenotype in the circulating T3 hormones could be found (Chandrasekharan *et al.* 2002).

Also a role in the production and regulation of embryonal pancreatic beta cells has been investigated and shown for FRK (Åkerblom *et al.* 2007; Welsh *et al.* 2004). An inhibitory effect on the invasiveness and migration of human glioma cells via the c-Jun N-terminal protein kinase (JNK)/c-Jun pathway has also been reported for FRK (Zhou *et al.* 2012).

Most recent research reports the first interaction of FRK with a membrane receptor. It shows that FRK binds to epidermal growth factor receptor (EGFR), preferentially when bound to epidermal growth factor (EGF), via the SH3 and SH2 domain and phosphorylates EGFR at Y1173. This phosphorylation has been associated with increased EGFR uptake into the cell and overexpression of FRK indeed correlates with a lower level of EGFR at the celluar membrane (Jin & Craven 2014).

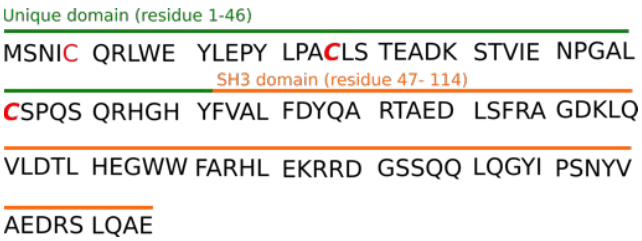


Figure 2.4: Primary sequence of the Unique and SH3 domain of human FRK.

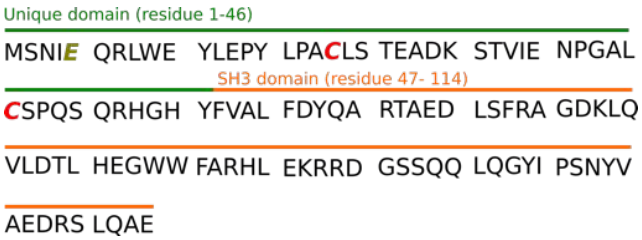


Figure 2.5: Primary sequence of the Unique and SH3 domain of human FRK with a cysteine to glutamic acid mutation at position two (C5E).

The study described here is focussed around the structural study of the Unique and the SH3 domain of the non-receptor tyrosine kinase FRK and the interaction between these two domains. An interaction between the Unique and SH3 domain has previously been described for c-Src which involves also interactions with the cellular membrane. The combination of these interactions induce conformational changes which delimitates a peptide loop within the Unique domain (Maffei *et al.* 2015; Pérez, Maffei, *et al.* 2013). To study the the loop formation in c-Src non-natural cysteines were introduced into the sequence that can be linked by the formation of disulfide bond.

In the case of FRK, the Unique domain has no obvious membrane interactions but it presents naturally a pair of cysteine residues (Figure 2.4), namely cysteine 18 and 36, that correspond to the positions of conserved aromatic residues in members of the closely related SFKs that participate in the long-range interactions involved in the delimitation and formation of the peptide loop in c-Src (Figure 2.6, *this figure is used with kind permission of Prof. Miquel Pons and Miguel Arbesú, Universitat de Barcelona. The alignment was generated using MUSCLE.*). Therefore the cysteine residues (18 and 36) might be involved in a similar structural rearrangement, namely the loop formation in the Unique domain, by metal ion chelation or the formation of a disulfide bridge. As only C18 and C36 should be involved the hypothesized interaction the third cysteine in position five was mutated into glutamic acid (Figure 2.5).

Figure 2.6: Primary sequence alignment of the Unique domain of c-Src homologues and SFKs inculding c-Src, Yes, Fyn and Fgr, highlighting conserved aromatic residues involved in a long-distance interaction (orange).

The aim is to establish an expression and purifying protocol for a construct consisting of the Unique and SH3 domain of human FRK, further referred to as FRK(USH3), and explore the properties and of this construct, including the interaction between the two domains in the open and cyclized forms. Cyclization will be attempted through Zn^{2+} complexation. This report shows the first described expression and purification of FRK(USH3) expressed in *E. coli*.

Nuclear magnetic resonance (NMR) spectroscopy has already been successfully used in the study of

the intrinsically disordered Unique domain of c-Src (Maffei *et al.* 2015; Pérez, Gairí, *et al.* 2009; Pérez, Maffei, *et al.* 2013). Likewise, NMR spectroscopy is being used as the main technique to study FRK(USH3) and will be complemented by Dynamic Light Scattering (DLS) and native polyacrylamide gelelectrophoresis (PAGE) studies.

3 Results

3.1. GST-FRK(USH3) is expressed into inclusion bodies

The coding sequence for the first 114 amino acids of the C5E mutant of FRK (Unique and SH3 domain) (later on referred to as FRK(USH3)), optimized for expression in *Escherichia coli*, was cloned in a pETM-30 vector (EMBL) to result in a gluthatione-S-transferase(GST)-FRK fusion protein. The gene sequence was already ordered with the first of the three cysteine residues mutated to glutamic acid (C5E). The GST fusion partner was chosen to enhance to solubility of the whole GST-FRK(USH3) construct (Lebendiker & Danieli 2014). Plasimds containing the correct sequence were transformed into bacteria of the *E. coli* BL21(DE3) strain. Successfully transformed bacteria were selected using the kanamycin resistance encoded in the plasmid and used to prepare an overnight (o/n) cultures in LB medium.

The glycerol cultures prepared from the o/n culture were used to perform expression test, which are a commonly used tool to determine the optimal conditions to express a new protein construct in a given expression system. In the case of GST-FRK(USH3) 45 different conditions were tested. Temperature during expression, expression time and isopropyl β-D-1-thiogalactopyranoside (IPTG) concentration used to induce expression were used as variables. Table 3.1 shows all 45 tested conditions.

For each condition the cells were harvested by centrifugation and resuspended in lysis buffer. Cell breakage was performed by repetitively freezing and thawing the cells. To determine whether the GST-FRK(USH3) construct is expressed in a soluble or in an insoluble from the insoluble cell debris was separated from the soluble fraction by centrifugation. The insoluble fraction was resolublilized with lysis buffer containing a high concentration 8M urea. For each condition both fraction were analyzed by SDS polyacrylamide gelelectrophoresis (PAGE). Figure 3.1 shows the electrophoretic analysis for all samples collected at 25°C (298.15K).

The analysis showed that 42 206 Da GST-FRK(USH3) construct is expressed in high quantities, which are similar in all tested conditions, but in insoluble form under all conditions. These observations leads to the conclusion, that GST-FRK(USH3) is not soluble when expressed in *E. coli* and in high quantities and therefore goes into inclusion bodies during cell lysis. Further this results show that the GST fusion tag does not enhance solubility of the FKR(USH3) construct sufficiently.

Moreover all tested condition show similar protein yield, therefore the condition for expression was chosen with 298.15K, 0.5mM IPTG and o/n expression, due to best integration into the workflow.

	Temperature (K)	Time (h)	IPTG [mM]
condition 1	288.15	2	0.1
condition 2			0.5
condition 3			1
condition 4		5	0.1
condition 5			0.5
condition 6			1
condition 7		o/n	0.1
condition 8			0.5
condition 9			1
condition 10	293.15	2	0.1
condition 11			0.5
condition 12			1
condition 13		5	0.1
condition 14			0.5
condition 15			1
condition 16		o/n	0.1
condition 17			0.5
condition 18			1
condition 19	298.15	2	0.1
condition 20			0.5
condition 21			1
condition 22		5	0.1
condition 23			0.5
condition 24			1
condition 25		o/n	0.1
condition 26			0.5
condition 27			1
condition 28	303.15	2	0.1
condition 29			0.5
condition 30			1
condition 31		5	0.1
condition 32			0.5
condition 33			1
condition 34		o/n	0.1
condition 35			0.5
condition 36			1
condition 37	310.15	2	0.1
condition 38			0.5
condition 39			1
condition 40		5	0.1
condition 41			0.5
condition 42			1
condition 43		o/n	0.1
condition 44			0.5
condition 45			1

Table 3.1: Expression tests: tested conditions

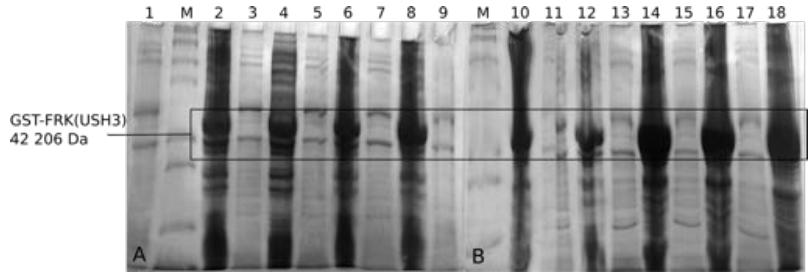


Figure 3.1: SDS-PAGE gels of all samples collected at 25°C (298.15K). The dominant bands within the black box represent the 42 206 Da GST-FRK(USH3) construct. Lanes marked with "M" contain a protein standard (Mark 12^{TM} Unstained Standard by ThermoFisher Scientific). (1) supernatant, 0.1mM IPTG, 2h; (2) pellet, 0.1mM IPTG, 2h; (3) supernatant, 0.5mM IPTG, 2h; (4) pellet, 0.5mM IPTG, 2h; (5) supernatant, 1mM IPTG, 2h; (6) pellet, 1mM IPTG, 2h; (7) supernatant, 0.1mM IPTG, 5h; (8) pellet, 0.1mM IPTG, 5h; (9) supernatant, 0.5mM IPTG, 5h; (10) pellet, 0.5mM IPTG, 5h; (11) supernatant, 1mM IPTG, 5h; (12) pellet, 1mM IPTG, 5h; (13) supernatant, 0.1mM IPTG, o/n; (14) pellet, 0.1mM IPTG, o/n; (15) supernatant, 0.5mM IPTG, o/n; (16) pellet, 0.5mM IPTG, o/n; (17) supernatant, 1mM IPTG, o/n; (18) pellet, 1mM IPTG, o/n

3.2. GST-FRK(USH3) can be successfully refolded

As shown in the previous section the GST-FRK(USH3) construct is expressed into inclusion bodies in insoluble form under all tested condition. Therefore the pellet collected after cell lysis and centrifugation needed to be resolublilized. In order to do so a Tris-HCl buffer with 8M urea in combination with mild sonication was used. This treatment resulted in resolubilization as well as denaturation of the protein of interest. To remove the GST fusion tag and to study FRK(USH3) in its natural conformation, native conditions, therefore removal of the detergent and refolding of the protein into its native conformation, needed to be restored. As the Unique domain of FRK is supposed to be intrinsically disordered, only the GST fusion tag and the SH3 domain needed to be refolded. For both domains several refolding strategies are described in the literature and three of these strategies were tested for GST-FRK(USH3): (1) direct dilution, (2) indirect dilution and (3) on-column refolding (Chen *et al.* 2009; Grantcharova & Baker 1997; Hutchinson & Chase 2006; D.-W. Li *et al.* 2004; Matsumura *et al.* 2013; Plaxco *et al.* 1998; Singh & Panda 2005; Vallejo & Rinas 2004; Wingfield *et al.* 2001).

Direct dilution is the quickest and most straightforward methodology. Here the solution containing the denatured protein of interest is diluted, at least 10-fold, into a refolding buffer containing no denaturant, which leads to a rapid reduction of the denaturant concentration and a dilution of the protein itself. This strategy usually works for small proteins or protein domains with simple fold-

ing dynamics, like the SH3 domain (Grantcharova & Baker 1997) but leads to aggregation and/or incorrect folding and precipitation of bigger and more complex proteins.

A very similar strategy is refolding by indirect dilution. The difference to a direct dilution is, that here the refolding buffer containing no denaturant is added to the protein solution containing a high concentration of denaturant. In this case the detergent is slowly diluted together with the protein. The main difference of indirect dilution compared to a direct dilution is that the detergent is slowly removed.

A very convenient way to refold a protein is to combine the refolding with a purification step. This can be done by gelfiltration or affinity purification. Here an affinity purification approach was used, where the protein was first bound to the column resin in its denatured form, then washed with buffers with step-wise or gradually reduced denaturant concentrations and finally eluted from the column purified and in its native conformation (Hutchinson & Chase 2006).

In all of the above described refolding techniques it is important to dilute the protein sufficiently to kinetically disfavor aggregation of intermediate folding states, which can lead to irreversible precipitation of large aggregates of misfolded proteins. To prevent aggregation several chemical additive have been described in the literature, which are generally clustered into three group: denaturants (protein destabilizers), protein stabilizers and protein aggregation inhibitors (H. Yamaguchi & Miyazaki 2014). Denaturants like urea and guanidinium chloride denature proteins by a chaotropic effect at high concentrations, but have been shown to have an inhibitory effect on protein aggregation during refolding at low concentrations(S. Yamaguchi *et al.* 2013). Additives which are used to stabilize the protein are glycerol, polyethylene glycol (PEG), sucrose and terhalose. Glycerol, for example, increases the protein stability by favoring hydrophobic interactions and increasing the solvent ordering around the protein (H. Yamaguchi & Miyazaki 2014). The most commonly used protein aggregation inhibitor is the amino acid L-arginine, which has been shown to increase protein solubility and inhibit protein aggregation during refolding, despite the exact mechanism is not yet understood (Tischer *et al.* 2010).Other aggregation inhibitors are glycineamid and proline.

Each of the described techniques was tested in the refolding of GST-FRK(USH3) as described in Chapter 5 (Materials and Methods). To prevent protein aggregation during the refolding process L-arginine was added to the refolding buffers.

During refolding regardless of the technique a smaller or greater fraction of the protein of interest was lost due to aggregation, misfolding and precipitation. The best yield of refolded and soluble GST-FRK(USH3) could be achieved by indirect dilution with refolding buffer 2 and 3. Table 3.2 shows the exact yield and loses for each technique, when GST-FRK(USH3) is expressed in LB.

	inital protein (mg)	soluble fraction (mg)	soluble fraction (%)	precipitation (mg)	precipitation (%)
direct dilution	34.53	6.98	20.2	27.55	79.8
indirect dilution (refolding buffer 1)	34.28	27.31	79.7	6.97	20.3
indirect dilution (refolding buffer 2 and 3)	34.14	34.07	99.8	0.07	0.2
on-column refolding	34.51	0.67	1.9	33.84	98.1

Table 3.2: Protein yields after refolding. Indirect dilution using refolding buffer 2 or 3 yields the highest proportion of successfully refolded GST-FRK(USH3) (grey).

3.3. FRK(USH3) tends to form aggregates

After successful refolding ^{15}N-labeled GST-FRK(USH3) is obtained highly diluted in volumes of at least 100mL and needed to be further purified. In a first purification attempt the refolded protein was concentrated using centrifugal concentrators (Amicon), but ^{1}H-^{15}N-HSQC and ^{1}H-^{15}N-SOFAST-HMQC spectra of the purified FRK(USH3) (the identity of the protein was verified my mass spectrometry, see Figure 3.2) showed that the protein was present in an aggregated form. To prevent the protein from aggregating during purification, steps involving concentration of the construct by centrifugal concentration were excluded and substituted to avoid the creation of high local concentration which would favor the formation of aggregates. By avoiding centrifugal concentration and keeping the protein at low concentrations during the entire purification, NMR spectra of reasonable quality, showing no aggregation could be obtained, which verified that the previously observed aggregation is indeed concentration-dependent. This tendency to aggregate and even precipitate also shows when the GST fusion tag is removed by cleavage with tobacco etch virus (TEV) protease at protein concentrations higher that 0.8mg/mL as visible insoluble aggregates are formed. This suggests that the fusion tag increases the solubility of FRK(USH3) at least to small extent.

In an attempt to increase the solubility and decrease the tendency to aggregate both cysteines residues in position 18 and 36 were derivatised with iodoacetic acid which introduced two additional negative charges and thereby increased the netto charge of the protein from -2 to -4 and inactivates the reactive thiol group in the cysteine side chain. The successful derivatization of both cysteins was verified by mass spectrometry (Figure 3.10. Indeed the solubility and the quality of the recorded NMR spectra could be sightly increased as expected. The disadvantage of the so far established purification protocol (described in Chapter 5 Materials and Methods) is that the highest obtainable protein concentration, which doesn't show any aggregation is around 13μM for both the non-derivatised and the derivatised protein construct. In order to be able to obtain a resonance assignment for FRK(USH3) concentrations of at least 100μM and higher are needed, which prevented the establishment of a resonance assignment so far. Furthermore 99.7% of the protein is lost during the purifications due to precipitation and aggregation (Table 5.1).

To monitor and analyze aggregation of FRK(USH3) in the none-derivatised and derivatised form three different techniques were applied, (1) NMR spectroscopy, (2) Dynamic Light Scattering (DLS) and (3) Native polyacrylamide gelelectrophoresis (PAGE).

	inital protein (mg)	final protein (mg)	protein loss (%)
non-derivatised FRK(USH3)	20	0.06	99.7
derivatised FRK(USH3)	20	0.06	99.7

Table 3.3: Proteins yields from 100mL of minimal medium expression from the expression to the final protein sample.

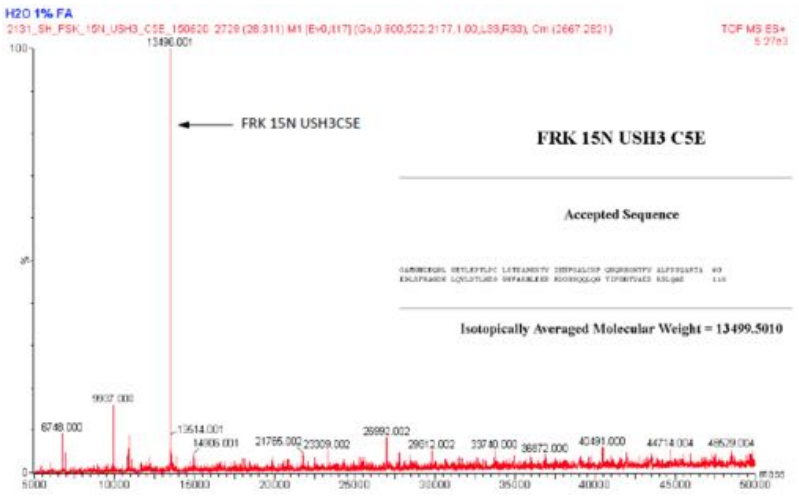


Figure 3.2: Mass spectrometric analysis of FRK(USH3): mass spectrum of FRK(USH3)

3.3.1. Nuclear Magnetic Resonance (NMR) spectroscopy

Theoretical Background

NMR spectroscopy was the main technique used to study the novel FRK construct, consisting if the entire N-terminal Unique domain and the subsequent SH3 domain. In this spectroscopy technique properties of nuclear spins and their behavior in an external magnetic field are exploited to extract structural and dynamic information about the molecule in question.

A *wavefunction* is a mathematical function which contains all the information about a given system, therefore if the *wavefunction* of the system under consideration is known, anything about the system can be deduced. In order to access this information quantum mechanical *operators* are needed, as in quantum mechanics they represent ”observables”, therefore things that can be measured. *Operators* have *eigenfunctions* with corresponding *eigenvalues* associated with them (*eigenvalue equation*, Equation 3.1).

$$\text{(operator) acting on (eigenfunction)} = \text{(eigenvalue) x (eigenfunction)} \tag{3.1}$$

The *Hamiltonian operator*, H, represents the observable quantity energy, which is the relevant observable in NMR spectroscopy. An important postulate in quantum mechanics states that if the value of an observable is measured, the results will always be one the eigenvalues of the operator representing the observable (Keeler 2010). Therefore the energy values measured will correspond to one of the eigenvalues of the Hamiltonian or in other words the eigenvalues of the Hamiltonian are the energy levels available to the system in question. The energy of interaction between a nuclear spin in a magnetic field, along the z-axis, of the strength B_0 with this magnetic field is represented by the Hamilitonian

in Equation 3.2.

$$\hat{H}_{onespin} = -\gamma B_0 \hat{I}_z \tag{3.2}$$

With γ being the gyromagnetic ratio (usually given in $rads^{-1}T^{-1}$), which is a specific, fundamental property of each nucleus and $\hat{I}_z$ being the operator representing the z-component of the nuclear spin angular momentum with interacts with the magnetic field along the z-axis. The nuclear angular momentum is a vector quantity with a given direction and magnitude. As can be deduced from Equation 3.1 all eigenfunctions of $\hat{I}_z$ are also eigenfunctions of $\hat{H}$, as the only difference between them is the multiplication by two constants. Therefore also all eigenvalues of $\hat{I}_z$ are the same for $\hat{H}$ but multiplied by $-\gamma B_0$. The number of eigenvalues for $\hat{I}_z$ is dependent on the nuclear spin angular momentum quantum number I. $\hat{I}_z$ has (2I + 1) eigenfunctions, which are each characterized by the quantum number m. The quantum number m can have values between -I and +I in integer steps. For a ^{1}H nucleus with $I = \frac{1}{2}$, m can be $-\frac{1}{2}$ or $+\frac{1}{2}$. With Ψ_m being the general eigenfunction the eigenvalue equation results as shown in Equation 3.3.

$$\hat{I}_z \Psi_m = m\hbar \Psi_m \tag{3.3}$$

With $\hbar$ being the reduced Planck constant.
Therefore Ψ_m has the eigenvalue $m\hbar$. Taking together Equation 3.2 and 3.3 we obtain Equation 3.4, from which we can extract the eigenvalues E_m of the Hamiltonian (Equation 3.5).

$$\hat{H}_{onespin} \Psi_m = -m\hbar\gamma B_0 \Psi_m \tag{3.4}$$

$$E_m = m\hbar\gamma B_0 \tag{3.5}$$

As m depens on the nuclear spin angular momentum quantum number I, also the number of available energy levels depends on I of a given system or nucleus respectively. In case of 1^H two different energy levels are available, were $m = +\frac{1}{2}$ is labeled the α state and $m = -\frac{1}{2}$ the β state. Therefore the transition between this two states corresponds to following energy difference (Equation 3.6).

$$\Delta E_{\alpha\to\beta} = \hbar\gamma B_0 \tag{3.6}$$

The frequency, in Hz and $rads^1$, corresponding to ΔE is given in Equation 3.7 and 3.8.

$$\nu_{\alpha\to\beta} = \gamma B_0 / 2\pi \tag{3.7}$$

$$\omega_{\alpha\to\beta} = \gamma B_0 \tag{3.8}$$

The Lamor frequency, in Hz, of a given spin is defined in Equation 3.9.

$$\nu_{\alpha\to\beta} = -\nu_0 \tag{3.9}$$

Therefore for systems consisting only of one single spin with $I = \frac{1}{2}$, like ^{1}H, there is only one allowed transition resulting in one single line in the spectrum at minus the Lamor frequency. Based on the theory presented above systems consisting of more than one spin and might therefore be coupled can be analyzed and discussed.

To account for chemical shift perturbations caused by differences in the chemical environment of nuclei of the same isotope in a given the Lamor frequency is redefined to include the chemical shift (Equation 3.10).

$$\nu_0 = -\gamma(1 + 10^{-6} * \delta)B_0/2\pi \tag{3.10}$$

With δ being the chemical shift in ppm.

The energy difference $\Delta E_{\alpha\rightarrow\beta}$, as can be seen in Equation 3.6, is dependent on the external magnetic field strength B_0. Therefore the corresponding Lamor frequency is dependent on the effective magnetic field strength (B_{eff}) that a given nucleus in a system experiences. This B_{eff} is dependent on the chemical environment of this nucleus and is therefore not the same for all nuclei of the same isotope in a system. The resulting differences in ν_0 contain information on the chemical environment and the spatial position of the nuclei. Equation 3.11 accounts for B_{eff}.

$$\Delta E_{\alpha\rightarrow\beta} = \hbar\gamma B_{eff} \tag{3.11}$$

$$B_{eff} = (1 - \sigma)B_0 \tag{3.12}$$

With σ being the shielding constant.

Besides the *nuclear spin angular moment* there is also a *nuclear spin magnetic momentum*, which means that there is a small magnetic field generated by the nucleus. This nuclear magnetic moment interacts with the applied external field it is placed in when an NMR experiment is to be conducted. The energetically most favorable orientation for the magnetic moments of a system is parallel to the external field, which is by convention along the z-axis. But the alignment of the magnetic moments is opposed to the random thermal motion of the molecules in the system, which strives to establish a random orientation. Due to a slight energetic advantage for the magnetic moments to align with the field a slight net alignment is created and the whole sample gets a magnetic moment. This is referred to als bulk magnetization along the z-axis, which is a vector quantity. When a sample is placed into a magnetic field this bulk magnetization as to be establish first and reach an equilibrium. The process be with the spins of a system come to equilibrium in a magnetic field is referred to as *relaxation*.

As has been shown above, the information we want to get from an NMR experiment is the Lamor frequency of the individual nuclei. In order to do so the bulk magnetization is tipped away a certain angle from the z-axis, there the magnetization vector starts the precess around the z-axis, therefore the applied field, with a precession frequency which equals the Lamor frequency (Equation 3.7, 3.8 and 3.9). To move the magnetization away from the z-axis into the xy-plane (transverse plane) a short-lived radiofrequency field is generated by specific pulses. The precession frequency is then recorded after the pulse, when the spins relax back into the equilibrium state, as *free induction decay* (FID). This time-dependent signal is then converted into frequency-dependent signal by Fourier transformation (Keeler 2010).

For the analysis of proteins by NMR spectroscopy the most common type of experiments are heteronuclear correlation spectra using ^{15}N (introduced by isotope enrichment) and ^{1}H as nuclei. This type of experiments make use of the fact that chemical shifts of different types of nuclei can be correlated. One of the used nuclei is very often ^{1}H and as only one of the nuclei can be observed in the

experiment it will mostly likely be the proton. This is due to the fact that the highest sensitivity can be achieved when the nucleus with the highest Lamor frequency, which is in most cases the proton, is observed. Because of historical reasons experiments which observe the proton are called inverse experiments. In the investigation of FRK(USH3) two types of ^{1}H and ^{15}N correlation spectra are used.

The HSQC (Heteronuclear Signle-Quantum Correlation) experiment is commonly used to record one-bond correlation spectra between ^{1}H and ^{15}N, for biological purposes (for technical details consult cited literature) (Bodenhausen & Ruben 1980; Keeler 2010). The other type of recorded experiment is the SOFAST-HMQC (Band-Selective Optimized-Flip-Angle Short Transient Heteronuclear Multiple Quantum Correlation) experiment, which is a variation of the basic HMQC experiment (Keeler 2010; Schanda & Brutscher 2005; Schanda, Kupe, *et al.* 2005). Both experiments result in basically the same spectrum, but differ in the way relaxation effects the two experiments (Keeler 2010). In both experiment types one cross peak represent one amino acid or rather the one-bond coupling between the amino-proton and the nitrogen. Therefore prolines can not be detected as well as the N-terminal amino acid.

Aggregation

The spectra shown in Figure 3.3 and 3.4 display the first ever 2D-dimensional NMR spectra recorded for the C5E FRK(USH3) construct at concentrations of 187μM (Figure 3.3) and 113μM (Figure 3.4). The spectra are poorly resolved even though the protein concentration is in a reasonable range. Therefore the poor resolution can not be accounted for by a low concentration. To see whether the poor quality of the spectra could be due to some sort of aggregation or oligomerization of FRK(USH3), spectra of a protein known to oligomerize were recorded und compared to the FRK(USH3) spectra. Figure 3.4 shows a ^{1}H-^{15}N HSQC spectrum of the glutathione-S-transferase (GST) encoded in the pETM-30 vector (EMBL). GST is known to at least dimerize, as has been shown before, and was therefore used as reference to see if the spectra of GST show the same characteristics as the spectra recorded for FRK(USH3) (Fabrini *et al.* 2009). As can be seen when Figure 3.3 and 3.4 are compared to Figure 3.5, the spectra of the two very distinct proteins show essentially the same characteristics, *(1)* poorly resolved peaks and *(2)* peaks with very different intensities. Based on this observation it was deduced that FRK(USH3) forms indeed some kind of oligomer or aggregate at concentrations higher that 110μM. Thus FRK(USH3) needed to be kept at lower concentrations and high concentrations at any point of the protocol needed to be avoided, as explained above.

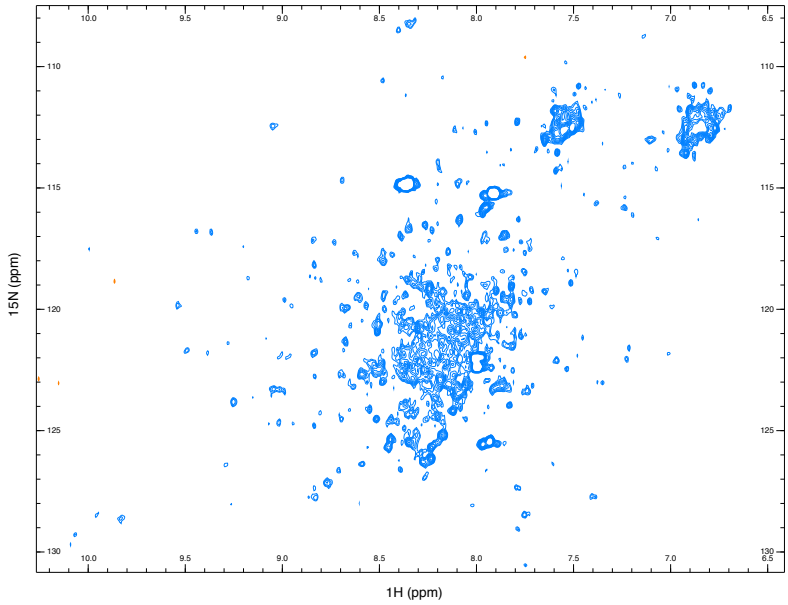


Figure 3.3: Spectrum A.1: ^{1}H-^{15}N HSQC spectrum of FRK(USH3), 187µM, 298.15K

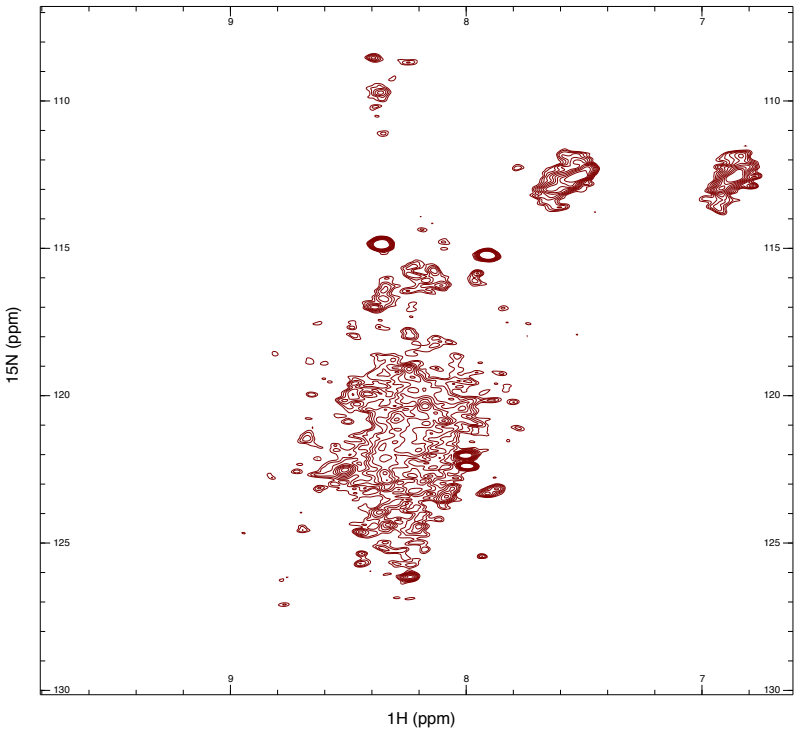


Figure 3.4: Spectrum B.1: ^{1}H-^{15}N SOFAST-HMQC spectrum of FRK(USH3), 113µM, 298.15K

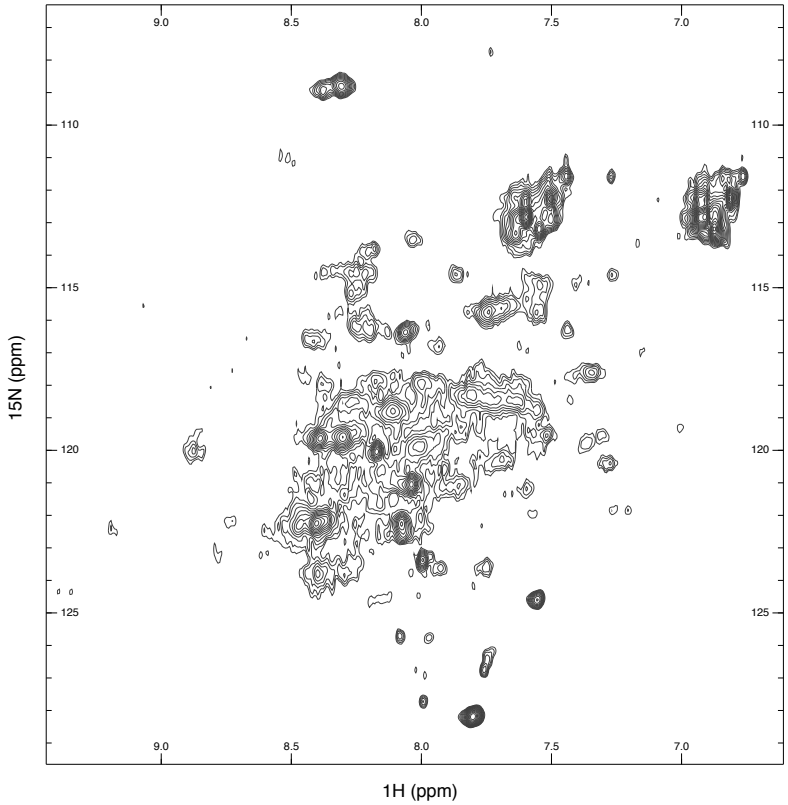


Figure 3.5: Spectrum C: ^{1}H-^{15}N HSQC spectrum of GST, 110µM, 298.15K

Zn^{2+} binding

As a consequence of the the poor quality of the spectra at high concentrations a new protocol was establish, which essentially avoided to generate higher concentrations of FRK(USH3) during any point of the purification (see Chapter 6 Materials and Methods). FRK(USH3) prepared according to the new protocol resulted in a sample with a concentration of about 13µM, Figure 3.6 shows the ^{1}H-^{15}N SOFAST-HMQC spectrum recorded at 25°C (298.15K). An obvious improvement in the quality of the spectra is visible and the the spectrum does not suggest any aggregation or oligomerization. Nevertheless the protein concentration is very low and only a fraction of the the expected 110 resonances (the construct consists of 116 amino acids, including five prolines) , representing one amino acid each, is observable. Only 50 peaks could be unambiguously characterized. The distribution of these is within the characteristic region for intrinsically disordered proteins (IDPs) or regions and as the Unique domain is an intrinsically disordered region with the protein, it can be deduced that hardly only resonances corresponding to amino acids in the Unique domain can be observed. Furthermore the Unique domain in this construct contains 48 residues, containing four prolines, therefore 43 peaks are expected for the Unique domain only.

Even with the improved quality of the spectrum the very low protein concentration, which does not show aggregation, is not suitable to attempt to establish a resonance assignment. Nevertheless $ZnCl_2$ was added to the FRK(USH3) solution and the effect of the Zn^{2+} ions, which could interacts with the thiol groups in the cysteine side chains, on the position of the resonances was observed by recording a ^{1}H-^{15}N SOFAST-HMQC spectrum (Figure 3.7). As can be seen in Figure 3.8 and 3.9 most

of the peaks disappear or experience a shift (chemical shift perturbation), which is due to changes in the chemical environment of the effected residues. Because of the lack of an assignment not many deductions can be made from this observation, except for the fact that the Zn2+ is indeed binding to FRK(USH3), but which residues are involved, besides cysteine 18 and 36, and with structural changes are induced by the binding can not be said.

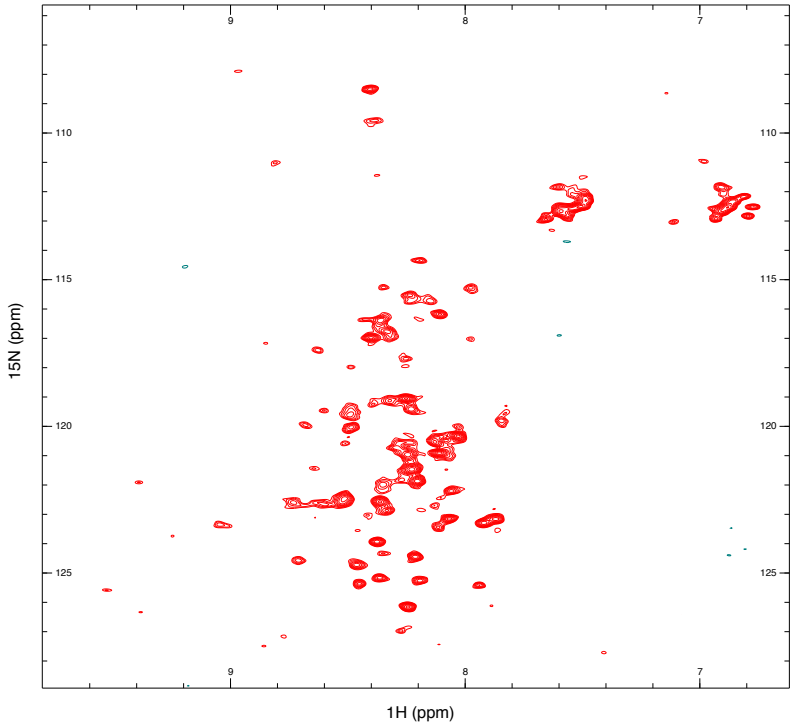


Figure 3.6: Spectrum D.2: ^{1}H-^{15}N SOFAST-HMQC spectrum of FRK(USH3), 13.25µM, 298.15K

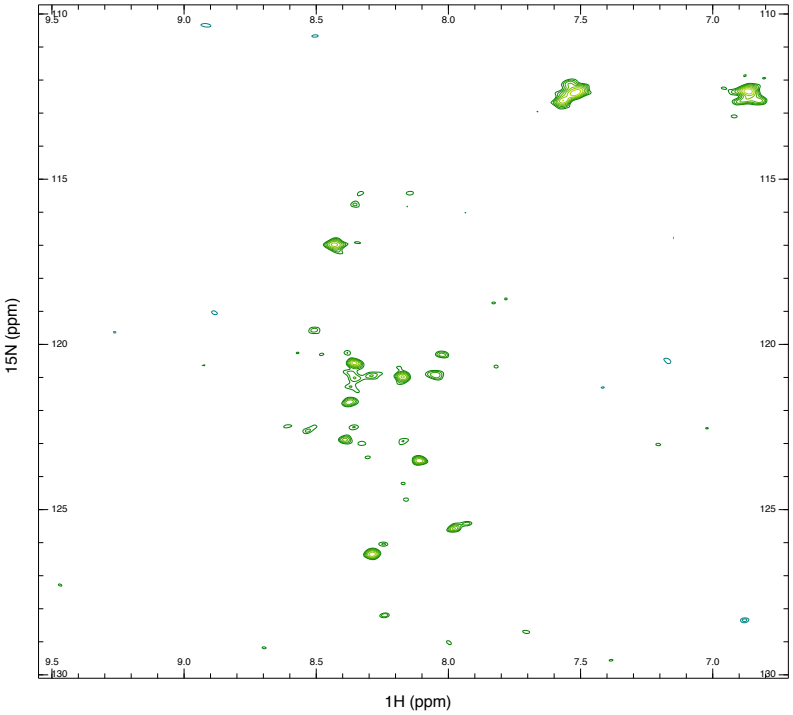


Figure 3.7: Spectrum D.3: ^{1}H-^{15}N SOFAST-HMQC spectrum of FRK(USH3) (13µM) with 200µM $ZnCl_2$, 298.15K

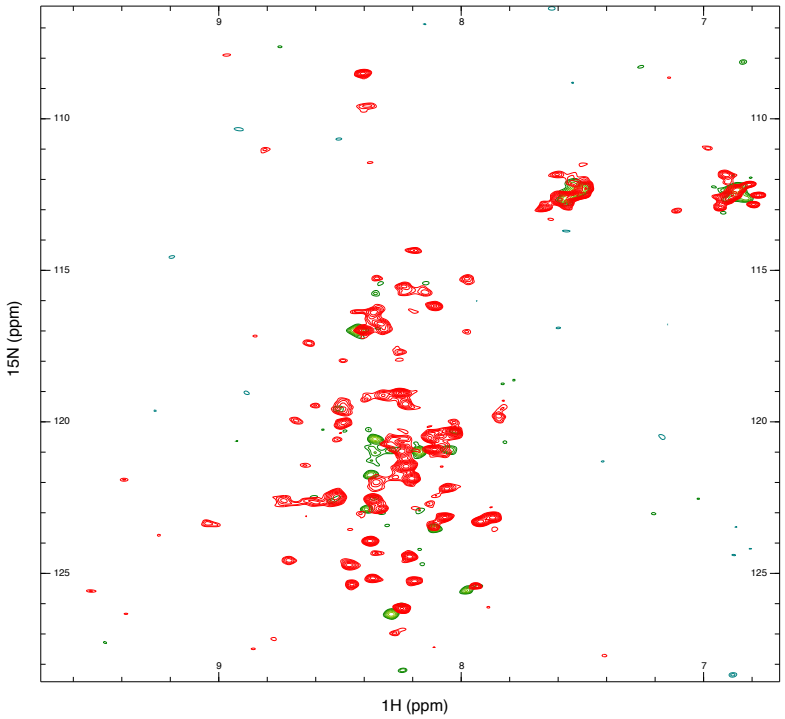


Figure 3.8: Overlay of spectrum D.2 and D.3: Overlay of the ^{1}H-^{15}N SOFAST-HMQC spectra of FRK(USH3) (13.25µM) (red) and FRK(USH3)(13.25µM) with 200µM $ZnCl_2$ (green), 298.15K

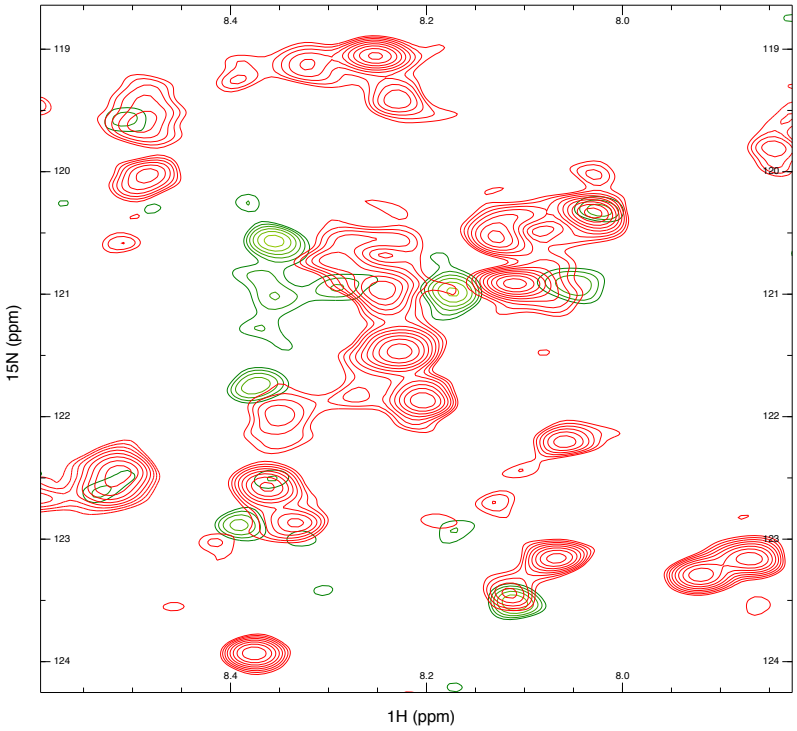


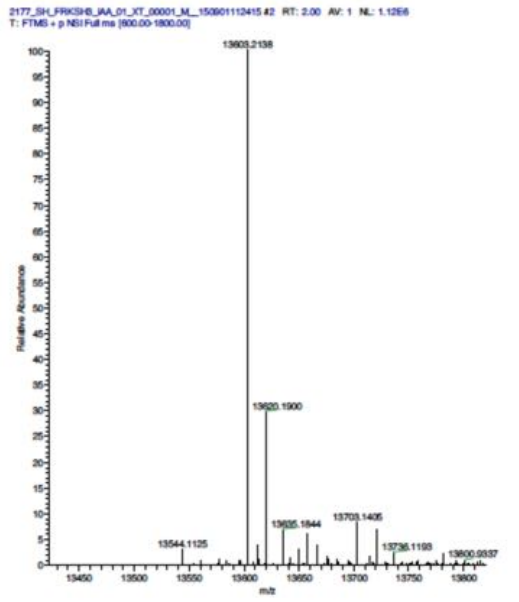
Figure 3.9: Overlay of spectrum D.2 and D.3 (Detail): Overlay of the ^{1}H-^{15}N SOFAST-HMQC spectra of FRK(USH3) (13.25µM) (red) and FRK(USH3)(13.25µM) with 200µM $ZnCl_2$ (green), 298.15K - Detail

None-derivatised and derivatised FRK(USH3)

In order to achieve higher concentrations without the protein aggregating, further improve the quality of the spectra and attempt to reach conditions which could be used to establish a resonance assignment, the cysteine residues were derivatised with iodoacetic acid. This treatment had two effects, *(1)* the reactive thiol side chains are masked and *(2)* two additional negative charges are introduced, which results in an increase of the protein netto charge from -2 to -4. Mass spectrometric analysis showed that the derivatisation was complete for both cysteine residues with 99.7% (Figure 3.10). Unfortunately the final concentration that could be obtained for the derivatised FRK(USH3) was 10.50µM, therefore in same range as for the none-derivatised protein. But even though the spectra were recorded for the same concentration range a higher quality spectrum could be obtained, with more peaks unambiguously characterizable (82 peaks compared to 56 in the none-derivatised protein) (Figure 3.13 and 3.14).

Further the ^{1}H-^{15}N HSQC and ^{1}H-^{15}N SOFAST-HMQC spectra were compared for the same sample, which are on the one hand the none-derivatised FRK(USH3) (Figure 3.11) and on the other hand the derivatised FRK(USH3) (Figure 3.12). The ^{1}H-^{15}N SOFAST-HMQC spectra for both proteins show the same peak distribution as in Figure 3.6 and therefore show mainly peaks corresponding to the intrinsically disorder Unique domain. In the ^{1}H-^{15}N HSQC spectrum resonances that can be assigned to the so far invisible SH3 domain can be observed, for example peaks corresponding to three three tryptophane residues in the SH3 domain arround 10ppm in the ^{1}H dimension and 129ppm in the ^{15}N dimension are clearly observable. Taken in account technical details about both types of experiments and the fact that the SH3 domain is not visible in the SOFAST-HMQC experiments but in HSQC experiments

suggests that the domain behaves as a much larger protein than it is suppose to be (Bodenhausen & Ruben 1980; Schanda & Brutscher 2005; Schanda, Kupe, *et al.* 2005). One possible explanation for this behavior could be some sort of association between two or more SH3 domains, which does not effect the Unique domain.



Sample	Elemental composition	Calculated mass (Da) %15N (protein)			Experimental mass (Da)
		0	99.7	100	
Protein	C590H892O183N16653	13326.4758	13488.8565	13492.8659	13602.2138
Protein + 1 IAA	C592H895O184N16753	13383.4972	13545.8779	13549.8873	
Protein + 2 IAA	C594H898O185N16853	13440.5187	13602.8994	13606.9088	

Figure 3.10: Mass spectrometric analysis of derivatised FRK(USH3)

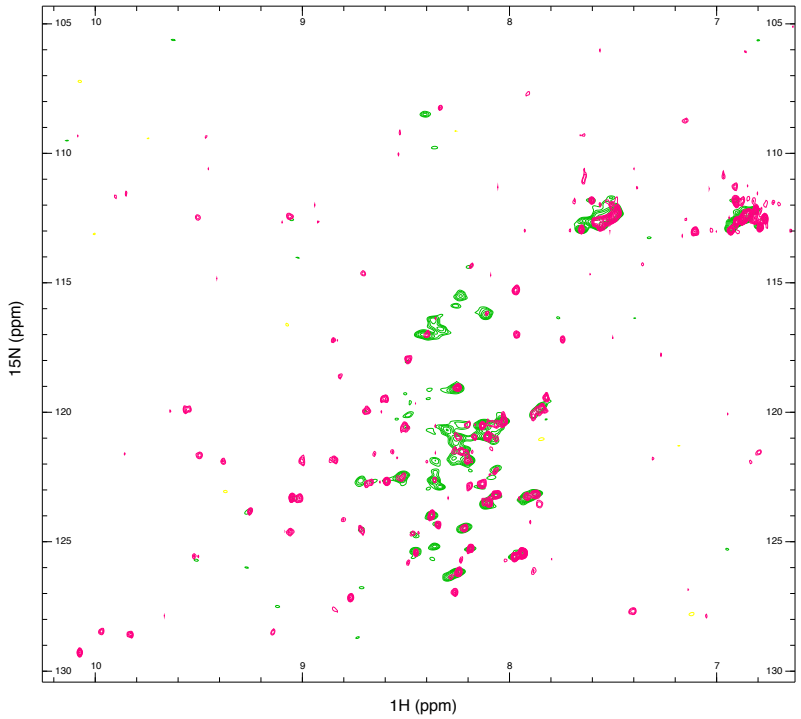


Figure 3.11: Overlay of spectrum E.1 and E.2: Overlay of the ^{1}H-^{15}N HSQC spectrum (pink) of FRK(USH3) (13.35µM) with the ^{1}H-^{15}N SOFAST-HMQC spectrum (green) of FRK(USH3) (13.35µM)

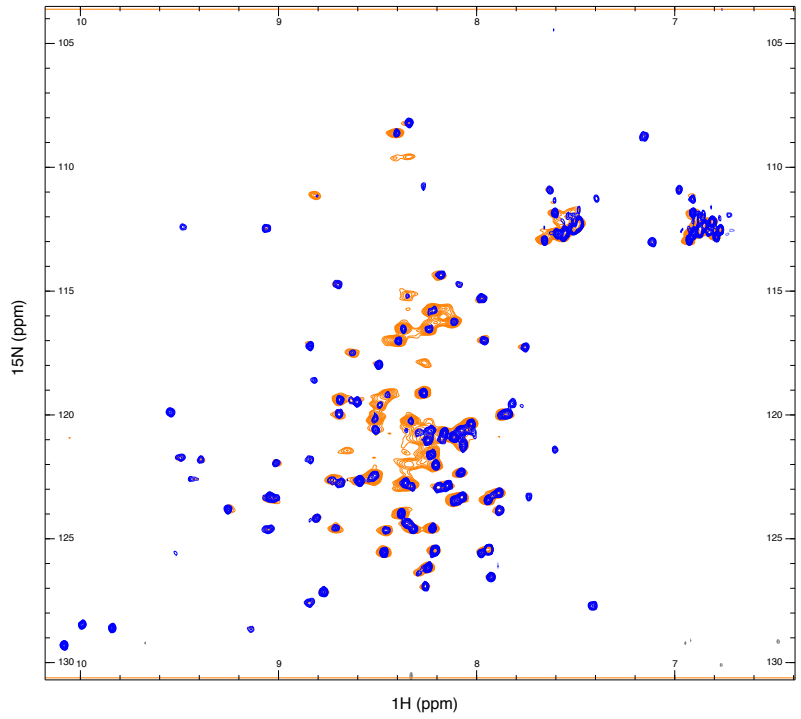


Figure 3.12: Overlay of spectrum F.1 and F.2: Overlay of the ^{1}H-^{15}N HSQC spectrum (blue) of derivatised FRK(USH3) (10.50µM) and the ^{1}H-^{15}N SOFAST-HMQC spectrum (orange) of derivatised FRK(USH3) (10.50M)

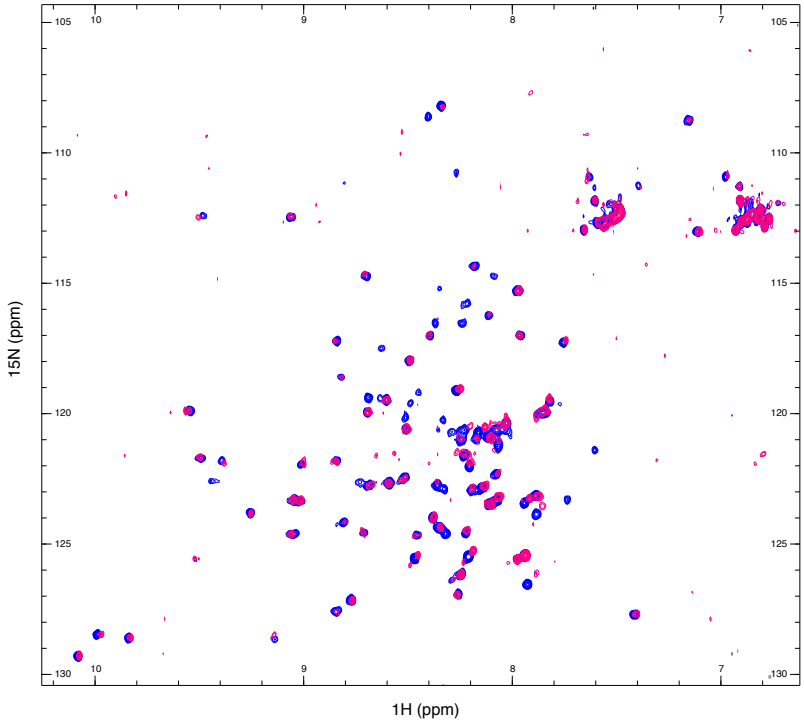


Figure 3.13: Overlay of spectrum E.1 and F.1: Overlay of the ^{1}H-^{15}N HSQC spectra of FRK(USH3) (13.25µM) (pink) and derivatised FRK(USH3) (10.50µM) (blue)

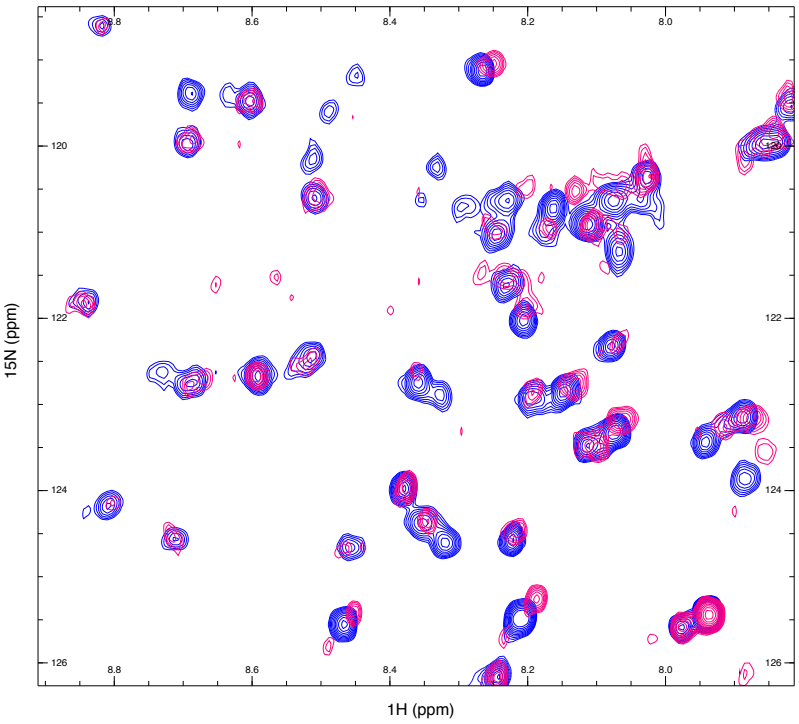


Figure 3.14: Overlay of spectrum E.1 and F.1 (Detail): Overlay of the ^{1}H-^{15}N HSQC spectra of FRK(USH3) (13.25µM) (pink) and derivatised FRK(USH3) (10.50µM) (blue) - Detail

3.3.2. Dynamic Light Scattering

Dynamic Light Scattering (DLS) measurements were performed to determine possible differences in the hydrodynamic radii (R_H) of the derivatised and the none-derivatised FRK(USH3). The motivation to determine the hydrodynamic radii of both proteins was to see whether one or the other is present in a more compacted state in solution that the other and also whether there is some extent of aggregation or oligomerization happening even though it can not be observed in the NMR spectra. As DLS is very sensitive in the detection of aggregates and a suitable method to determine the hydrodynamic radius it was the method of choice.

DLS is generally used to determine the mutual translational diffusion coefficient D_t, which is correlated with the hydrodynamic radius R_H. For hard spheres this relationship is given by the Stokes-Einstein relation (Equation 3.1).

$$D_t = k_B T / 6\pi\eta R_H \tag{3.13}$$

With D_t being the diffusion coefficient, k_B the Boltzman's constant ($1.381 * 10^{-23}$ J/K), η the viscosity and R_H the hydrodynamic radius.

In general when visible light is scattered by small particles it is scattered in all direction (Rayleigh scattering) as long as the particle is small compared to the wavelength. If the scattered light source is monochromatic (e.g. laser) the intensity of the scattered light fluctuates over time. This fluctuation is cause by brownian motion, which results in a displacement of the particles from their initial position over time. Equation 3.1 shows that D_t is indirect proportional to R_H, therefore it can be deduced that smaller molecules or particles move faster in solution than bigger molecules or particles, which is a characteristic of brownian motion. Hence the recorded intensity fluctuations contain information about the time-scale of the movement of the particles scattering the light source. The intensity trace can be fitted with an autocorrelation function (Equation 3.2).

$$g^2(q;\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle^2} \tag{3.14}$$

With $g^2(q;\tau)$ being the second order autocorrelation function at a specific wave vector,q, and delay time, τ, and I is the intensity of the scattered light. At short time delays between the recorded scattered intensity the correction is very high, as the particles do not have time to move to a greater extent, this correlation decays exponentially over time as the particles deviate more and more from their initial position until there is no correlation between the scattered intensity between the initial an the final states (Lorber *et al.* 2012). The decay of the correlation has a direct relationship with the motion of the particles and therefore with D_t and by Equation 3.1 with R_H.

In order to extract information about the scattering particles from the establish autocorrelation function different mathematical approaches are applied. In case of monodisperse samples the first order autocorrelation function can be written as a single exponential as in Equation 3.3.

$$g^1(q;\tau = exp(-\Gamma\tau) \tag{3.15}$$

With Γ being the decay rate. The relationship between the decay rate Γ and D_t through the wave

vector, q, is given by Equation 3.4 and 3.5.

$$\Gamma = q^2 D_t \tag{3.16}$$

$$q = \frac{4\pi n_0}{\lambda} sin\left(\frac{\Theta}{2}\right) \tag{3.17}$$

With n_0 being the refractive index of the sample, λ the wavelength of the incident laser and Θ the angle at which the scattered light is detected. The vector q is a function of angle Θ and the incident wavelength λ and varies with the experimental conditions (e.g. temperature, pressure, concentration). If D_t of a particle is extracted its dimensions can be determined using Equation 3.1 (Berne & Pecora 2000; Lorber *et al.* 2012).

In order to extract as much useful information as possible several numerical methods have been developed. For monodisperse samples the cumulant method provides a good fit, whereas the CONTIN method developed by Steven Provencher is ideal for heterodisperse, polydisperse and multimodol samples (Koppel 1972; Provencher 1982).

Measurements were conducted and analyzed as described in Chapter 5 (Materials and Methods). Table 3.4 displays the Z-average as hydrodynamic diameter and the polydispersity index (PDI) measured for the none-derivatised und derivatised FRK(USH3).

	none-derivatised FRK(USH3)	**derivatised FRK(USH3)**
Z-average (nm)	363.0	277.4
Polydispersity Index (PDI)	0.491	0.862

Table 3.4: Measured Z-average (d.nm) and PDI for none-derivatised and derivatised FRK(USH3)

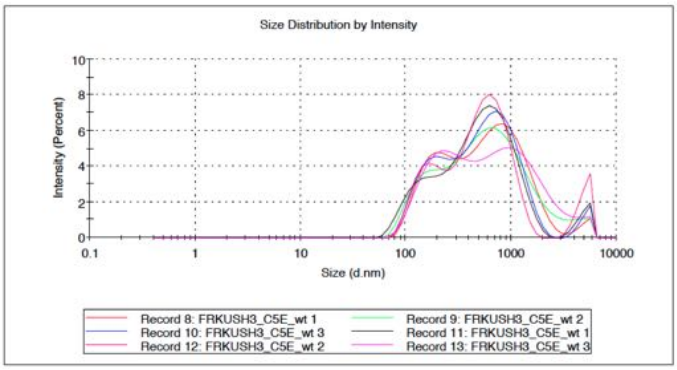


Figure 3.15: Size distribution by intensity for all measurements conducted of none-derivatised FRK(USH3) in solution.

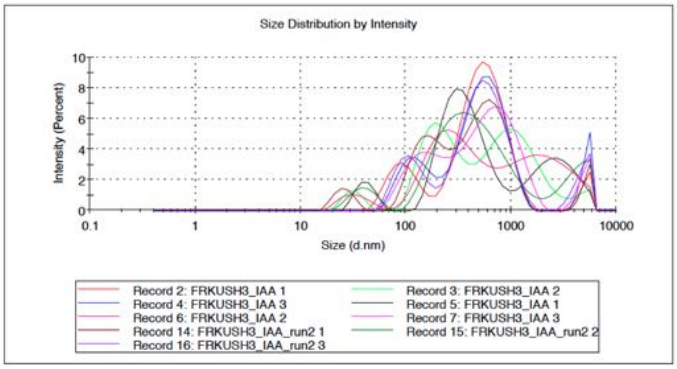


Figure 3.16: Size distribution by intensity for all measurements conducted of derivatised FRK(USH3) in solution.

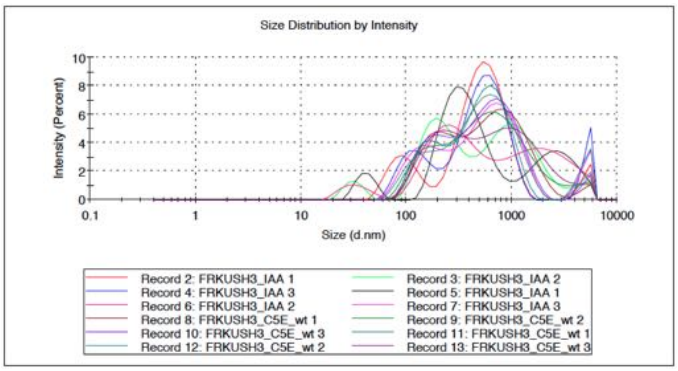


Figure 3.17: Size distribution by intensity for all measurements conducted of none-derivatised and derivatised FRK(USH3) in solution.

As can be seen in Figure 3.15, 3.16 and 3.17 the measurements of the derivatised FRK(USH3) are poorly reproducible, which could be partly due to the high polydispersity, as indicated by the high PDI, of the sample. In contrast measurements of the none-derivatised protein are nicely reproducible and also give a lower PDI. Due to the poor quality of the results obtained for the derivatised protein any conclusions drawn might be unreliable, but what can be seen in both cases is, that there is an unnegligible fraction of protein in aggregated or oligomerized form present in solution. The calculated Z-average for the hydrodynamic diameter shows an average smaller diameter for derivatised FRK(USH3) compared to none-derivatised FRK(USH3), which could indicate an aggregation inhibiting effect of the additional charged and the removal of the cysteine side chains.

3.3.3. Native PAGE

The separation of protein in native PAGE is based on the same separation principal as denaturing SDS-PAGE, expect that protein are prepared in a none-denaturing and sometimes none-reducing sample buffer and are not denatured by heat as well. This results in the maintenance of the secondary structure and the native charge density of the protein. Therefore the migration of a given protein in not only dependent on its molecular weight, like in denaturing SDS-Page, but rather on the isoelectric point (pI) and the hydrodynamic radius (R_H).

Here native PAGE was used to analyze possible relative differences in the hydrodynamic radius between the none-derivatised and the derivatised FRK(USH3). As the pI of the protein of interest is the second parameter influencing the migration in the native gel, the pI of both proteins had to be taken into account. None-derivatised has a theoretical pI of 5.09 and the derivatised version of FRK(USH3) has a theoretical pI of 4.87 (pI values were calculated based on the primary sequence, where the derivatised cysteine residues were replaced by glutamic acid, by the Protpharam online tool (Gasteiger *et al.* 2005)). As the electrophoretic separation was conducted at basic pH both protein are positively charged and migrate towards the cathode.

Figures 3.18 and 3.19 shows that the derivatised variant of FRK(USH3) indeed migrates a bit further in both the 12% and the 15% polyacrylamide separation gel. This difference in migration could be due to either a slightly small hydrodynamic radius or a slightly lower pI. As the derivatised FRK(USH3) shows a slightly lower pI its migrational behavior would be explainable by its pI, hence no conclusions can be drawn from this experiment concerning possible differences in hydrodynamic radii.

In order to rule out differences in migration caused by the difference in pI a two-dimensional analysis (native PAGE coupled with isoelectric focusing) might provide more information.

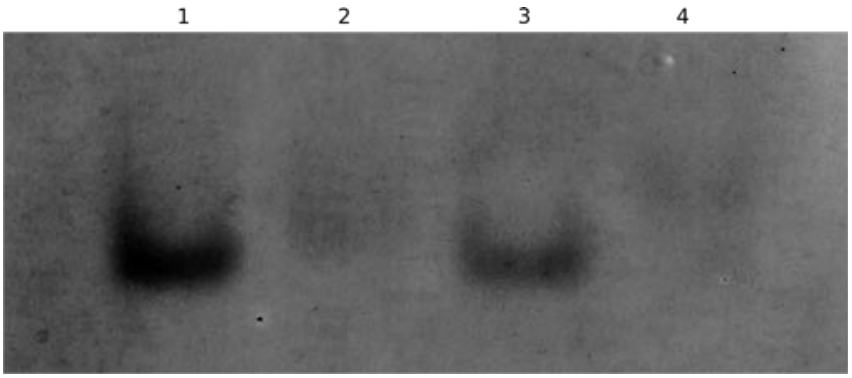


Figure 3.18: 15% Native PAGE gel of none-derivatised and derivatised FRK(USH3).(1) derivatised FRK(UAH3); (2) none-derivatised FRK(USH3); (3) derivatised FRK(USH3) under reducing conditions; (4) none-derivatised FRK(USH3) under reducing conditions.

Figure 3.19: 12% Native PAGE gel of none-derivatised and derivatised FRK(USH3).(1+2) derivatised FRK(UAH3); (3+4) none-derivatised FRK(UAH3).

4 Discussion

Here the first expression and purification of the N-terminal fragment (residues 1-114) of human FRK, referred to as FRK(USH3), from expression in *E. coli* is reported. This fragment contains the intrinsically disordered Unique domain and the structurally well defined Src-homology 3 (SH3) domain. Further the first ever recorded two-dimensional heteronuclear NMR spectrum of FRK(USH3) is presented.

In c-Src, the leading member of the Src family kinases (SFKs), Unique and SH3 domain have been shown to interact. This interaction together with another interaction involving the inner lipid layer of the cellular membrane cooperate to delimitate a peptide loop within the Unique domain, that interacts with the SH3 domain (Maffei *et al.* 2015; Pérez, Gairí, *et al.* 2009; Pérez, Maffei, *et al.* 2013). This interaction has also been found in another member of the SFKs, namely Fyn, and is under investigation for Yes (unpublished data). Therefore it is believed to be a general feature of the SFKs.

The non-receptor tyrosine kinase FRK/PTK6 family is structurally closely related the the SFKs as the kinases share the same modular structure consisting of Unique, SH3, SH2 and kinase domain, but lacking the N-terminal SH4 domain as well as a lipidation (Serfas & Tyner 2003). As no lipidation is present, these kinases are not naturally targeted to the cellular membrane and therefore have more flexibility in their cellular localization.

The loop formation observed in SFKs is believed to be also happing in FRK, the leading member of the FRK/PTK6 family. This hypothesis is formed around the fact that the Unique domain of FRK naturally presents two cysteine residues in regions that are involved in the loop formation in c-Src. Therefore the loop on the Unique domain of FRK could be delimited by metal ion chelation between the two cysteine residues.

In order to test this hypothesis *in vitro* the sequence coding for the C5E mutant of the N-terminal fragment of the human FRK, containing Unique and SH3 domain, was cloned into an expression vector and expressed as glutathione-S-transferase (GST) fusion protein in *E. coli*. Even though the construct was expressed as GST fusion protein, the fusion tag could not enhance solubility of the construct enough and GST-FRK(USH3) was obtained in inclusion bodies under 45 tested expression conditions. As a result the insoluble construct had to resolubilized by using high concentrations of urea and was consequently obtained as a denatured protein.

The refolding of the natively folded domains of the construct, namely GST and SH3 domain, required highly diluted protein solutions and a very slow reduction of urea together with protein aggregation inhibiting additives (L-arginine). Indirect dilution was finally established as the most successful method by which high yields of refolded GST-FRK(USH3) could be obtained.

After the fusion tag was removed by proteolytic cleavage the released FRK(USH3) showed decreased solubility at moderately high concentration and consequently needed to be kept at low con-

centrations for further purification. The low solubility was revealed to be correlated with a high tendency to aggregate and precipitate at higher concentrations.

This posed limitations for the application of NMR spectroscopy as a method of investigation. Due to FRK(USH3) tendency to aggregate properly resolved ^{1}H-^{15}N two-dimensional correlation spectra could only be recorded at protein concentrations around 10μM. The quality of the spectra but not the protein concentration could be increased by derivatising both cysteine residues in the Unique domain with iodoacetic acid. This modification introduced two additional negative charged and lead to an increase of the protein's netto charge from -2 to -4 and masked the thiol groups in the cysteine side chains. The establishment of a resonance assignment was not possible due to the low concentration at which the protein did not show aggregation.

Other techniques like Dynamic Light Scattering (DLS) and native polyacrylamide gelelectrophoresis (PAGE) gave indications that the derivatised protein has a lower tendency to aggregate and might therefore point into a promising direction to achieve higher concentrations of non-aggregated FRK(USH3) in the future and also reduce the substantial loss of protein during the purification. Higher concentrations would possibly allow the establish a resonance assignment.

Due to the lack of an assignment the initially made hypothesis proposing a loop formation within the Unique domain of FRK involving metal ion chelation by two cysteine residues could not be investigated in detail. A NMR spectrum of FRK(USH3) in the presence of Zn^{2+} ions reveals only that the metal ion indeed induces structural changes within the Unique domain.

Even though FRK(USH3) poses difficulties and limitations to be overcome in future work, this study shows that FRK(USH3) is indeed a workable protein construct and as the knowledge about its structural and functional properties is limited up to this moment, opens up possibilities to explore this so far mostly unregarded protein from a structural perspective and also gain insights into the functional space occupied by intrinsically disorders regions within the FRK/PTK6 family.

5 Materials and Methods

5.1. Cloning and Mutagenesis

Cloning and Mutagenesis

The sequence optimized cDNA (342bp) encoding the human C5E mutant of the Unique and SH3 domain of FRK (FRK USH3 C5E, residues 1-114) (ProteoGenix SAS, France) was cloned into a pETM-30 expression vector (EMBL) downstream a TEV cleavage site to be expressed as a His_6-GST fusion protein.The T4 DNA Ligase used was pursued from New England Biolabs, Inc.. For successful cloning, using *NcoI* and *XhoI* restriction enzymes (New England Biolabs, Inc.), a mutation had to be introduced at position two from the original serine to an alanine (S2A). This mutation was afterwards reversed by site-directed mutagenesis using the Q5 High-Fidelity® PCR Kit (New England Biolabs, Inc.) following the manufactures introductions. All primers were evaluated using the OligoEvaluator™online tool (Sigma-Aldrich®) and pursued from Sigma-Aldrich® (Table 5.1). All nucleic acid concentrations were measured with a NanoDrop Spectrometer ND-1000 (NanoDrop Technologies, Inc.).
The sequencing results (obtained from Macrogen, Inc.) confirmed the accuracy of the construct sequence.

Primer Name	Length (nt)	Tm (C)	GC %	Sequence (5'-3')
Foward(NcoI) S2A	27	76.5	51.8	TATACCATGGCCAACATTGAGCAGCGC
Reverse(XhoI)	30	73.1	50	TATACTCGAGTTATTCGGCCTGCAGAGAGC
Foward A2S	24	76.8	58.3	CAGGGCGCCATGTCCAACATTGAG
Reverse A2S	24	76.8	58.3	CTCAATGTTGGACATGGCGCCCTG

Table 5.1: Primer specifications and sequence

5.2. Protein Expression and Purification

5.2.1. GST-FRK(USH3) Expression

Transformation

The ligation product was transformed into XL-Gold omnimax and BL21(DE3) *E. Coli* strains (Novagen) using the heat shock method. Successfully transformed XL-Gold omnimax cells were used to amplify the plasmid encoding His_6-GST-FRK(USH3), whereas successfully transformed BL21(DE3) cells were used to express His_6-GST-FRK(USH3).
As the plasmid harbors a kanamycin resistance, plasmid-containing cells were selected by growing

them on kanamycin-supplemented agar plates. Successfully transformed BL21(DE3) cells were grown in an LB (Laboratorios CONDA) overnight (o/n) culture and glycerol stocks were prepared.

Expression tests

Expression tests were performed for a total of 45 conditions (Table 3.1), in which isopropyl β-D-1-thiogalactopyranoside (IPTG) concentration (0.1mM, 0.5mM, 1mM), temperature (15°C, 20°C, 25°C, 30°C, 37°C) and expression time (2h, 5h, o/n) were systematically changed. All tests were preformed with an equal number of bacteria for each of the conditions. The cells were grown in 5mL of sterile LB medium supplemented with kanamycin (1mg/mL) at 37°C, until an optical density at 600nm (OD600) between 0.6-0.8 and then expression was induced with IPTG.
1mL-samples were taken after certain time points, the cells were immediately centrifuged for 15min, 10k rpm at 4°C. The medium was removed and the cell pellet was resuspended in 50μL lysis buffer (100mM Tris-HCl, 150mM NaCl, 0,5mM EDTA, 1% NaN_3, 20mM 2-mercaptoethanol, PMSF (phenylmethylsulfonyl fluoride), PIC (protease inhibitor cocktail containing AEBSF, EDTA,Bestatin, Pepstatin and E64), 1mM dithiothreitol (DTT), pH 8.0) resuspended pellets were frozen at -20°C.

The cells were broken by cycles of freezing (liquid nitrogen $\sim$ -180°C) and thawing (water bath $\sim$ 40°C). After complete cell breakage all samples where centrifuged for 10min, 13k rpm at 4°C and the supernatant was separated from the insoluble components in the pellet.The pellets were resuspended in lysis buffer with 8M urea and all fraction were thereafter analyzed by 20% SDS-polyacrylamide gelelectrophoresis (PAGE).

Expression of ^{15}N-labeled GST-FRK(USH3)

An o/n LB pre-culture of 20mL was grown from 200μL of the glycerol stock at 37°C. For 1L of expression 10mL of pre-culture were used. The bacteria were grown in M9 minimal medium (42mM Na_2HPO_4, 22mM KH_2PO_4, 8,5mM NaCl, 0.2% glucose, 1mM $MgSO_4$, 0.3mM $CaCl_2$, 1μg thiamine, 1μg biotin, trace elements) supplemented with ^{15}N (0.5g/ml of $^{15}NH_4Cl$) and kanamycin (1g/mL) until an OD600 between 0.6-0.8 at 37°C and 135rpm. The expression of the target protein was induced at 25°C by the addition of IPTG to a final concentration of 0.5mM. Expression was left to process o/n.
The cell were harvested (centrifugation 30min, 4k g at 4°C) and the cell pellet was resuspended in 20mL ice cold lysis buffer. The cells were broken by sonication and treated with DNase for 30min at 4°C on ice. The cell lysate was centrifuged for 30min, 25k rpm at 4°C and the supernatant was discarded. The pellet was resuspended in 10mL lysis buffer with 8M urea using mild sonication. The solubilized pellet containing the denatured protein of interest was centrifuged again for 30min, 25k rpm at 4°C to remove cellular membrane components. The supernatant collected after the second centrifugation was frozen at -80°C in aliquots of 1mL.

5.2.2. FRK(USH3) Purification

Refolding of ^{15}N-labeled FRK(USH3)

To reduce the urea concentration and restore the native conformation of the construct several strategies were tried *(1)*direct dilution, *(2)* indirect dilution and *(3)* on-column refolding.

For refolding by direct dilution one aliquot (1mL) was diluted 1:2 in lysis buffer containing 8M urea and the protein solution was dropped slowing into 20mL of stirring refolding buffer 1 (50mM Tris-HCl, 150mM NaCl, 0.5M urea, 0.1M L-Arginine, 1mM DTT, pH 7.0) at room temperature and stirred for another two hours after complete addition of the protein solution. For indirect dilution one aliquot (1mL) was diluted 1:10 in lysis buffer containing 8M urea and the resulting solution was slowly diluted further by adding a total of 90mL of refolding buffer 1, refolding buffer 2 (50mM Tris-HCl, 150mM NaCl, 0.5M urea, 0.1M L-Arginine, 1mM DTT, pH 8.0) and refolding buffer 3 (50mM Tris-HCl, 50mM NaCl, 0.5M urea, 0.1M L-Arginine, 1mM DTT, 0,5M EDTA, pH 8.0) over several hours at 4°C.

The protein was attempted to be refolded using an IMAC (immobilized metal affinity column) (Ni-NTA Superflow Cartridges 5mL by QIAGEN). The column was equilibrated with 100mL of lysis buffer containing 8M urea (flow 10mL/min) and the protein was loaded and bound onto the column with of flow of 2mL/min. To refold, the column with the bound protein was washed with lysis buffer with stepwise reduced urea concentration (6M, 5M, 4M, 3M, 2M, 1M and 0M) in 12mL steps and eluted with lysis buffer containing 400mM imidazol and lysis buffer containing 8M urea and 400mM imidazol.
Indirect dilution was chosen to be used to refold GST-FRK(USH3). All protein concentrations were measured using a Nanodrop Spectrometer ND-1000 device by NanoDrop Technologies, Inc.

Purification of ^{15}N-labeled FRK(USH3)

After establishing a robust refolding protocol the highly diluted protein was dialyzed (3.5kDa cut-off, Spectrum Laboratiories, Inc.) against an IMAC compatible buffer (50mM Tris-HCl, 150mM NaCl, 10mM 2-mercaptoethanol, pH 8.0) o/n at 4°C. After changing the buffer the protein solution was loaded onto a pre-equilibrated IMAC (Ni-NTA Superflow Cartridges 5mL by QIAGEN) with a constant flow rate of 2mL/min at 4°C and the native His_6-GST-USH3 FRK was eluted in 20mL of IMAC elution buffer (50mM Tris-HCl, 150mM NaCl, 400mM imidazol, 10mM 2-mercaptoethanol, pH 8.0). For further procedures the buffer was exchanged by dialysis (3.5kDa cut-off, Spectrum Laboratiories, Inc.) against TEV cleavage buffer (50mM Tris-HCl, 50mM NaCl, 1mM DTT, 0.5 EDTA, pH 8.0) o/n at 4°C.

For removal of the His_6-GST-tag the protein solution was diluted to a final protein concentration between 0.5-0.7 mg/mL and split into aliquots of 5mL. TEV protease (pursued from the IRB Barcelona) was added to each aliquot to reach a final target to protease ratio of 1:50. The proteolytic digest was left to process o/n at 4°C under slight shacking. The cleavage resulted in a two-amino-acid N-terminal extented FRK(USH3) construct (**GA** MSNIE QRLWE YLEPY LPCLS TEADK STVIE NPGAL CSPQS QRHGH YFVAL FDYQA RTAED LSFRA GDKLQ VLDTL HEGWW FARHL EKRRD GSSQQ LQGYI PSNYV AEDRS LQAE,116 amino acids).

To separate the protease and the tag from the traget protein gelfiltration using a hand-packed Superdex 75 26&60 (GE Healthcare Bio-Sciences) was applied. The Column was operated using an ÄKTA FPLC system and a UNICORN 5.20 (Build-500) - Workstation (GE Healthcare Life Sciences). The column (column volume of 180mL) was clean with 220mL of milliQ water and 0.5M NaOH before equilibration with 250mL FPLC buffer (20mM Sodium phosphate (NaH_2PO_4, Na_2HPO_4), 50mM NaCl, 1mM DTT, pH 7.5). The sample was injected in injections of 5-7mL. Gelfiltration purification was performed at 4°C. Fractions containing purified FRK(USH3) were combined and split into aliquots

of equal volumes and frozen o/n in carbonic ice at -80°C for lyophilization. Frozen aliquots were lyophilizied for at least 24 and afterwards slowly resuspended in either 1,5mL or 6mL total volume of milliQ water and dialyzed into NMR buffer (20mM Sodium phosphate (NaH_2PO_4, Na_2HPO_4), 1mM DTT or 10mM 2-mercaptoethanol pH 7.0) o/n at 4°C.

A fraction of the protein was derivatized at the cysteine residues with iodoacetic acid (Sigma-Aldrich®). After purification by the affinity column (Ni-NTA) one third of the obtained yield was taken and treated with an access of DTT for 4h at room temperature to assure that thiol groups are reduced. A 1M iodoacetic acid solution was prepared in 1M NaOH. After complete reduction 125µL of the 1M iodoacetic acid solution were added to the protein solution (1mg/25µL of protein) and reaction left to process for 30min in the dark at room temperature. The reaction was quenched with an access of DTT. The iodoacetic acid and the DTT access was removed by dialyzing the derivatized protein solution against TEV cleavage buffer o/n at 4°C. Further purifications steps followed the above described protocol.

5.3. Mass Spectrometry

Samples of none-derivatized and derivatized FRK(USH3) were analyzed by LC-MS to assure their identity. All analysis were performed by the IRB Barcelona Mass Spectrometry Core Facility*.
Analysis specifications can be found in Table 5.2 for the chromatographic separation and in Table 5.3 for the mass spectrometric analysis.

	none-derivatised FRK(USH3)	**derivatised FRK(USH3)**
Sample concentration	5µM (diluted with 1% formic acid)	1pmol/µL (diluted in 1% formic acid)
Chromatgraph	Acquity UPLC Binary Sol MGR (Waters Corporation)	nanoAcquity UPLC (Waters Corporation)
Autosampler	Acquity UPLC Autosampler MO (Waters Corporation)	N/A
Column	BioSuite pPhenyl 1000RPC 2.0x75mm; 10µm	Trap loading, separation on column Trap column: 180µm x 2cm C4 Symmetry 300 (Waters Corporation) Analytical column: BEH130™ C4 75µm x 10cm, 1.7µm (Waters Corporation)
Eluents	A: Water 1% formic acid B: Acetonnitrile 1% formic acid	A: Water 0.1% formic acid B: Acteonnitrile 0.1% formic acid
Gradient	5% to 80% of B in 60min. Total chromatographic time: 80min	10% 1to 80% B in 30min + 80% to 90% in 5min
Flow rate	100µL/min	300nL/min

Table 5.2: LC and nanoLC conditions for none-derivatised and derivatised FRK(USH3)

	none-derivatised FRK(USH3)	**derivatised FRK(USH3)**
Mass spectrometer	LCT-Premier (Waters Corporation)	LTQ-FT Ultra (Thermo Scientific)
Polarity	ESI positive	ESI positive
Capillary voltage	3000V	34V
Capillary temperature	120C	200C
Tube Lens	100V	100V
m/z range	400-5000	400-2000
Data Processing	MassLynx software V4.1. SCN639 (Waters Inc.)	Xcalibur software vs 2.0SR2 (Thermo Scientific)

Table 5.3: MS conditions for none-derivatised and derivatised FRK(USH3)

"Mass spectrometry/Proteomics was performed at the IRB Barcelona Mass Spectrometry Core Facility, which actively participates in the BMBS European COST Action BM 1403 and is a member of Proteored, PRB2-ISCIII, supported by grant PRB2 (IPT13/0001 - ISCIII-SGEFI / FEDER)"

5.4. NMR Spectroscopy

Sample preparation

Samples of 250µL or 300µL in NMR buffer (20mM Sodium phosphate (NaH_2PO_4, Na_2HPO_4), 1mM DTT or10mM 2-mercaptoethanol pH 7.0) and 10% D_2O were prepared. The final protein concentration of all samples was measured with a Nanodrop Spectrometer ND-1000 (NanoDrop Technologies, Inc.) using the theoretical mass of 13 334 Da and the theoretical extinction coefficient ε= 25 440 (both were calculated based on the primary sequence using the ProtPharam online tool (Gasteiger *et al.* 2005)). The complete sample volume was transferred into 5mm Shigemi NMR tubes.

NMR measurements

All experiments were performed on a Bruker 600 MHz Advance III spectrometer equipped with a TCI CryoProbe (*Unitat de RMN, Universitat de Barcelona, Bruker Corp.*). For each sample either a ^{1}H-^{15}N-HSQC (heteronuclear single quantum correlation) or a ^{1}H-^{15}N-SOFAST-HMQC (band-selective optimized-flip-angle short transient heteronuclear multiple quantum correlation) or both spectra were recorded at either 25°C (298.15 K) or 5°C (278.15 K) or both temperatures (Bodenhausen & Ruben 1980; Schanda & Brutscher 2005; Schanda, Kupe, *et al.* 2005). Table 5.4 shows a details list of all spectra recorded. To observe possible changes (chemical shift perturbations) upon complexation of Zn^{2+} ions a sample of previously measured FRK(USH3) was saturated with 200µM $ZnCl_2$ and a ^{1}H-^{15}N-SOFAST-HMQC spectrum at 25°C (298.15 K) was recorded. To test the reversability of the Zn^{2+} complexation the same sample was saturated with EDTA and again a ^{1}H-^{15}N-SOFAST-HMQC spectrum at 25°C (298.15 K) was recorded. DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) was used as a reference compound.

All recorded spectra were processed and analyzed with Bruker TopSpin 3.2 (Bruker Corp.), rNMR (ver 1.1.9) and CcpNmr Analysis (Version 2.4.2) (Lewis *et al.* 2009; Vranken *et al.* 2005).

Spectrum	Sample	Sample Volume (µL)	Sample concentration (µM)	Experiment Type	Num. Scans	Temperature (K)	Experimental Time
A.1	none-derivatised FRK(USH3)	250	187	HSQC	24	298.15	1h 58min
A.2	none-derivatised FRK(USH3)	250	187	SOFAST-HMQC	16	298.15	25min
A.3	none-derivatised FRK(USH3)	250	187	HSQC	24	278.15	1h 58min
A.4	none-derivatised FRK(USH3)	250	187	SOFAST-HMQC	16	278.15	25min
B.1	none-derivatised FRK(USH3)	250	113	SOFAST-HMQC	128	298.15	2h 30min
B.2	none-derivatised FRK(USH3)	250	113	SOFAST-HMQC	128	278.18	2h 30min
C	GST	250	110	HSQC	64	298.15	1h 18min
D.1	none-derivatised FRK(USH3)	300	13.25	SOFAST-HSQC	64	278.15	1h 14min
D.2	none-derivatised FRK(USH3)	300	13.25	SOFAST-HMQC	64	298.15	1h 14min
D.3	none-dervatised FRK(USH3)+Zn^{2+}	300	13.25	SOFAST-HMQC	64	298.15	1h 14min
D.4	none-derivatised FRK(USH3) +Zn^{2+} and EDTA	300	13.25	SOFAST-HMQC	64	298.15	1h 14min
E.1	none-derivatised FRK(USH3)	300	13.25	SOFAST-HMQC	64	298.15	51min
E.2	none-derivatised FRK(USH3)	300	13.25	HSQC	64	298.15	5h 5min
F.1	derivatised FRK(USH3)	300	10.50	SOFAST-HMQC	64	298.15	51min
F.2	derivatised FRK(USH3)	300	10.50	HSQC	64	298.15	5h 15min

Table 5.4: List of conducted NMR experiments

5.5. Dynamic Light Scattering (DLS)

Dynamic Light Scattering (DLS) for none-derivatised and derivatised FRK(USH3) was measured for

samples of 60µL in NMR buffer (20mM Sodium phosphate (NaH_2PO_4, Na_2HPO_4), 1mM DTT, pH 7.0). Measurements were performed at 25C (298.15K) with a duration of 70 seconds on a Zetasizer Nano S (Malvern Instruments Ltd) and analyzed with the associated Zetasizer software (Ver. 7.03). Buffer alone was measured as a blank and each sample was measured six times.

5.6. Native PAGE

Native polyacrylamid gelelectrophresis (PAGE) gels were prepared according to Table 5.5. One pure sample containing of either none-derivatised or derivatised FRK(USH3) were used for the native PAGE analysis. Each sample was diluted 1:2 with 2x sample buffer (62.5mM TrisHCl, 25% glycerol, 1% bromophenol blue, pH 6.8) and separated using a Mini PROTEAN Tetra Cell electrophoresis system (Biorad) with 80V of applied voltage at 4°C.
After separation all gels were stained for 30-40 min with Coomassie Brilliant Blue and de-stained o/n at room temperature.

	stacking gel	**separation gel (12%)**	**separation gel (15%)**
0.375M TrisHCl, pH 8.8	4.275mL	5.89mL	4.89mL
30% Acrylamid/Bis Solution (Biorad)	0.67mL	4mL	5mL
Ammonium persulfate (APS)	50µL	100µL	100µL
Tetramethylethylenediamine (TEMED)	5µL	10µL	10µL

Table 5.5: Native PAGE gel composition

6 Abbreviations and acronyms

AEBSF ...	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
BCR ...	B-cell Receptor
DLS ...	Dynamic Light Scattering
DTT ...	Dithiothreitol
EDTA ...	Ethylenediaminetetraacetic acid
FRK(USH3) ...	FRK construct containing the Unique and SH3 domain (residues 1-114)
GST ...	Glutathione S-transferases
GST-FRK(USH3) ...	Construct containing a GST fusion tag N-terminal of the FRK construct
HSQC ...	Heteronuclear Single Quantum Correlation
IDP ...	Intrinsically Disordered Proteins
IPTG ...	Isopropyl β-D-1-thiogalactopyranoside
IRR ...	Immune recognition receptor
LC ...	Liquid chromatography
MS ...	Mass spectrometry
NMR ...	Nuclear magnetic resonance
nRTKs ...	non-receptor receptor tyrosine kinases
PAGE ...	Polyacrylamide gelelectrophoresis
PMSF ...	Phenylmethylsulfonyl fluoride
PTKs ...	Protein tyrosine kinases
RTKs ...	Receptor tyrosine kinases
SDS ...	Sodium dodecyl sulfate
SFKs ...	Src family kinases
SH1 ...	Src-homology 1
SH2 ...	Src-homology 2
SH3 ...	Src-homology 3
SOFAST-HMQC ...	Band-Selective Optimized-Flip-Angle Short Transient Heteronuclear Multiple Quantum Correlation
TCR ...	T-cell Receptor
TEV ...	Tobacco Etch Virus
TPKs ...	Tyrosine protein kinases

Bibliography

1. Abeyrathna, P. & Su, Y. The critical role of Akt in cardiovascular function. *Vascular Pharmacology*. doi:http://dx.doi.org/10.1016/j.vph.2015.05.008 (2015).
2. Åkerblom, B., Annerén, C. & Welsh, M. A role of FRK in regulation of embryonal pancreatic beta cell formation. *Molecular and Cellular Endocrinology* **270,** 73–78. ISSN: 03037207 (2007).
3. Alberts, B. *et al. Molecular Biology of the Cell* 5th ed. Chap. 15 (Garland Science, Taylor and Francis Group, LLC, 2008).
4. Amata, I., Maffei, M. & Pons, M. Phosphorylation of unique domains of Src family kinases. *Frontiers in Genetics* **5,** 1–6 (2014).
5. Berne, B. J. & Pecora, R. *Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics* (General Publishing Company, Ltd.,30 Lesmill Road, Don Mills, Toronto, Ontario, 2000).
6. Bian, Y. *et al.* An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome. *Journal of Proteomics* **96,** 253–262. ISSN: 1874-3919 (2014).
7. Bodenhausen, G. & Ruben, D. J. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* **69,** 185–189 (1980).
8. Boggon, T. J. & Eck, M. J. Structure and regulation of Src family kinases. *Oncogene* **23,** 7918–7927 (2004).
9. Brauer, M. P. & Tyner, A. L. RAKing in AKT: A Tumor Suppressor Function for the Intracellular Tyrosine Kinase FRK. *Cell Cycle* **8,** 2728–2732 (2009).
10. Broome, M. & Hunter, T. The PDGF receptor phosphorylates Tyr 138 in the c-Src SH3 domain in vivo reducing peptide ligand binding. *Oncogene* **14,** 17–34 (1997).
11. Cance, G., Craven, R. J., Xu, L., Alitalo, K. & Liu, E. T. Rak, a novel nuclear tyrosine kinase expressed in epithelial cells. **5,** 1347–1355 (1994).
12. Carracedo, a. & Pandolfi, P. P. The PTEN-PI3K pathway: of feedbacks and cross-talks. *Oncogene* **27,** 5527–5541 (2008).
13. Cartwirght, CA, Eckhart, W., Simon, S. & Kaplan, P. Cell transformation by pp60c-src mutated in the carboxy-terminal regulatory domain. *Cell* **49,** 83–91 (1987).

14. Chaimowitz, N. S., Falanga, Y. T., Ryan, J. J. & Conrad, D. H. Fyn Kinase Is Required for Optimal Humoral Responses. *PLoS ONE* **8,** 4–11 (2013).

15. Chandrasekharan, S. *et al.* Characterization of mice deficient in the Src family nonreceptor tyrosine kinase Frk/rak. *Molecular and cellular biology* **22,** 5235–5247 (2002).

16. Chapman, N. M., Yoder, A. N. & Houtman, J. C. D. Non-Catalytic Functions of Pyk2 and Fyn Regulate Late Stage Adhesion in Human T Cells. *PLoS ONE* **7** (2012).

17. Chen, J. *et al.* Cooperative effects of urea and L-arginine on protein refolding. *Protein expression and purification* **66,** 82–90 (2009).

18. Craven, R. J., Cance, W. G. & Liu, E. T. The nuclear tyrosine kinase Rak associates with the retinoblastoma protein pRb. *Cancer Research* **55,** 3969–3972 (1995).

19. Da Silva, AJ, Yamamoto, M., Zalvan, C. & Rudd, C. Engagement of the TcR/CD3 complex stimulates p59fyn(T) activity: detection of associated proteins at 72 and 120-130 kD. *Molecular Immunology* **29,** 1417–25 (1992).

20. D'Aniello, S. *et al.* Gene expansion and retention leads to a diverse tyrosine kinase superfamily in amphioxus. *Molecular Biology and Evolution* **25,** 1841–1854. ISSN: 07374038 (2008).

21. Daub, H. *et al.* Kinase-Selective Enrichment Enables Quantitative Phosphoproteomics of the Kinome across the Cell Cycle. *Molecular Cell* **231,** 438–448 (2008).

22. Davis, M. M. & Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature* **334,** 395–402 (1988).

23. Fabrini, R. *et al.* Monomer00E200880092Dimer Equilibrium in Glutathione Transferases: A Critical Re-Examination. *Biochemistry* **48,** 10473–10482 (2009).

24. Gasteiger, E. *et al. The Proteomics Protocols Handbook* (ed Walker, J. M.) 571–607 (Humana Press Inc., Totowa, NJ, 2005).

25. Gauen, Timson L. K., Kong, a. N., Samelson, L. E. & Shaw, a. S. p59fyn tyrosine kinase associates with multiple T-cell receptor subunits through its unique amino-terminal domain. *Molecular and cellular biology* **12,** 5438–5446 (1992).

26. Goel, R. K., Miah, S., *et al.* The unique N-terminal region of SRMS regulates enzymatic activity and phosphorylation of its novel substrate docking protein 1. *FEBS Journal* **280,** 4539–4559 (2013).

27. Goel, R. K. & Lukong, K. E. Tracing the footprints of the breast cancer oncogene BRK ? Past till present. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1856,** 39–54. ISSN: 0304419X (2015).

28. Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL2013EBI. *Nucleic Acids Research* **38,** W695–W699 (2010).

29. Grantcharova, V. P. & Baker, D. Folding Dynamics of the src SH3 Domain. *Biochemistry* **36,** 15685–15692 (1997).

30. Hayakawa, F. & Naoe, T. SFK-STAT pathway: An alternative and important way to malignancies. *Annals of the New York Academy of Sciences* **1086,** 213–222 (2006).

31. Hutchinson, M. H. & Chase, H. a. Adsorptive refolding of histidine-tagged glutathione S-transferase using metal affinity chromatography. *Journal of Chromatography A* **1128,** 125–132 (2006).

32. Ikeda, O., Miyasaka, Y., *et al.* STAP-2 is phosphorylated at tyrosine-250 by Brk and modulates Brk-mediated STAT3 activation. *Biochemical and Biophysical Research Communications* **384,** 71–75. ISSN: 0006-291X (2009).

33. Ikeda, O., Mizushima, A., *et al.* Involvement of STAP-2 in Brk-mediated phosphorylation and activation of STAT5 in breast cancer cells. *Cancer Science* **102,** 756–761. ISSN: 1349-7006 (2011).

34. Imamoto, A. & Soriano, P. Disruption of the csk gene, encoding a negative regulator of Src family tyrosine kinases, leads to neural tube defects and embryonic lethality in mice. *Cell* **73,** 1117–24 (1993).

35. Isakov, N. Immunoreceptor tyrosine-based activation motif (ITAM), a unique module linking antigen and Fc receptors to their signaling cascades. *Journal of Leukocyte Biology* **61,** 6–16 (1997).

36. Jin, L. & Craven, R. J. The Rak/Frk tyrosine kinase associates with and internalizes the epidermal growth factor receptor. *Oncogene* **33,** 326–35 (2014).

37. Kamalati, T., Jolin, H. E., Fry, M. J. & Crompton, M. R. Expression of the BRK tyrosine kinase in mammary epithelial cells enhances the coupling of EGF signalling to PI 3-kinase and Akt, via erbB3 phosphorylation. *Oncogene* **19,** 5471–5476. ISSN: 0950-9232 (2000).

38. Kamalati, T., Jolin, H. E., Mitchell, P. J., *et al.* Sensitizes Mammary Epithelial Cells to Epidermal Growth Factor *. **271,** 30956–30963 (1996).

39. Kaplan, Kenneth B., Swedlow, J. R., Morgan, D. O. & Varmus, H. E. c-Src enhances the spreading of *src-/-* fibrobasts on fibronectin by kinase-independent mechanism. *Genes and Development* **9,** 1505–1517 (1995).

40. Kaplan, Kenneth B *et al.* Association of the amino-terminal half of c-Src with focal adhesions alters their properties and is regulated by phosphorylation of tyrosine 527. **13,** 4745–4756 (1994).

41. Keeler, J. *Understanding NMR Spectroscopy* 2nd ed. Chap. 2,3,4,8,9 (John Wiley Sons, Ltd., 2010).

42. Kim, P. W., Sun, Z.-Y. J., Blacklow, S. C., Wagner, G. & Eck, M. J. A zinc clasp structure tethers Lck to T cell coreceptors CD4 and CD8. *Science (New York, N.Y.)* **301,** 1725–1728 (2003).

43. Kohmura, N., Yagi, T., Tomooka, Y. & Oyanagi, M. A Novel Nonreceptor Tyrosine Kinase , Srm : Cloning and Targeted Disruption. **14,** 6915–6925 (1994).

44. Koppel, D. E. Analysis of Macromolecular Polydispersity in Intensity Correlation Spectroscopy: The Method of Cumulants. **57,** 4814–4820 (1972).

45. Kypta, Robert M, Goldberg, Y., Ulug, E. T. & Courtneidge, S. A. Association between the PDGf receptor and members of the *src* family of tyrosine kinases. *Cell* **62,** 481–492 (1990).

46. Lebendiker, M. & Danieli, T. Production of prone-to-aggregate proteins. *FEBS Letters* **588,** 236–246 (2014).

47. Lee, H., Kim, M., Lee, K. H., Kang, K. N. & Lee, S. T. Exon-intron structure of the human PTK6 gene demonstrates that PTK6 constitutes a distinct family of non-receptor tyrosine kinase. *Molecules and cells* **8,** 401–407. ISSN: 10168478 (1998).

48. Lee, S., Strunk, K. & Spritz, R. A survey of protein tyrosine kinase mRNAs expressed in normal human melanocytes. *Oncogene* **8,** 3403–3410. ISSN: 0950-9232 (Dec. 1993).

49. Lewis, I. a., Schommer, S. C. & Markley, J. L. rNMR: Open source software for identifying and quantifying metabolites in NMR spectra. *Magnetic Resonance in Chemistry* **47** (2009).

50. Li, J. *et al.* PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science (New York, N.Y.)* **275,** 1943–1947. ISSN: 0036-8075 (1997).

51. Li, W. *et al.* The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research*. doi:10.1093/nar/gkv279 (2015).

52. Li, D.-W. *et al.* Refolding and characterization of recombinant human GST-PD-1 fusion protein expressed in Escherichia coli. *Acta biochimica et biophysica Sinica* **36,** 141–146 (2004).

53. Lorber, B., Fischer, F., Bailly, M., Roy, H. & Kern, D. Protein analysis by dynamic light scattering: Methods and techniques for students. *Biochemistry and Molecular Biology Education* **40,** 372–382 (2012).

54. Lukong, K. E., Larocque, D., Tyner, A. L. & Richard, S. Tyrosine Phosphorylation of Sam68 by Breast Tumor Kinase Regulates Intranuclear Localization and Cell Cycle Progression. *Journal of Biological Chemistry* **280,** 38639–38647 (2005).

55. Luttrell, D. K. & Luttrell, L. M. Not so strange bedfellows: G-protein-coupled receptors and Src family kinases. *Oncogene* **23,** 7969–7978 (2004).

56. Maffei, M. *et al.* The SH3 Domain Acts as a Scaffold for the N-Terminal Intrinsically Disordered Regions of c-Src. *Structure,* 893–902 (2015).

57. Matsumura, Y. *et al.* Transient Helical Structure during PI3K and Fyn SH3 Domain Folding. *The Journal of Physical Chemistry B* **117,** 4836–4843 (2013).

58. McWilliam, H. *et al. Analysis Tool Web Services from the EMBL-EBI* 2013. doi:10.1093/nar/gkt376.

59. Meyer, T. *et al.* Breast cancer cell line proliferation blocked by the Src-related Rak tyrosine kinase. *International Journal of Cancer* **104,** 139–146. ISSN: 00207136 (2003).

60. Mitchell, P. *et al.* Cloning and characterisation of cDNAs encoding a novel non-receptor tyrosine kinase, brk, expressed in human breast tumours. *Oncogene* **9,** 2383–90 (1994).

61. Mori, S. *et al.* Identification of two juxtamembrane autophosphorylation sites in the PDGF beta-receptor; involvement in the interaction with Src family tyrosine kinases. *The EMBO journal* **12,** 2257–2264 (1993).

62. Nada, S *et al.* Constitutive activation of Src family kinases in mouse embryos that lack Csk. *Cell* **73,** 1125–35 (1993).

63. Parsons, S. J. & Parsons, J. T. Src family kinases, key regulators of signal transduction. *Oncogene* **23,** 7906–7909 (2004).

64. Pérez, Y., Gairí, M., Pons, M. & Bernadó, P. Structural characterization of the natively unfolded N-terminal domain of human c-Src kinase: insights into the role of phosphorylation of the unique domain. *Journal of molecular biology* **391,** 136–48 (2009).

65. Pérez, Y., Maffei, M., *et al.* Lipid binding by the Unique and SH3 domains of c-Src suggests a new regulatory mechanism. *Scientific reports* **3,** 1295 (2013).

66. Plaxco, K. W. *et al.* The Folding Kinetics and Thermodynamics of the Fyn-SH3 Domain. *Biochemistry* **37,** 2529–2537 (1998).

67. Playford, M. P. & Schaller, M. D. The interplay between Src and integrins in normal and tumor biology. *Oncogene* **23,** 7928–7946 (2004).

68. Provencher, S. W. CONTIN: A general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Computer Physics Communications* **27,** 229–242 (1982).

69. Reynolds, a. B. *et al.* Activation of the oncogenic potential of the avian cellular src protein by specific structural alteration of the carboxy terminus. *The EMBO journal* **6,** 2359–2364 (1987).

70. Roskoski, R. Src protein-tyrosine kinase structure, mechanism, and small molecule inhibitors. *Pharmacological Research* **94,** 9–25 (2015).

71. Sawyer, T. K. Src Homology-2 Domains : Structure , Mechanisms , and Drug Discovery. *Biopolymers (Peptide Science)* **47,** 243–261 (1998).

72. Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proetins on the time scale of seconds. *Jacs* **127,** 8014–8015 (2005).

73. Schanda, P., Kupe, E. & Brutscher, B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *Journal of Biomolecular NMR* **33,** 199–211 (2005).

74. Serfas, M. S. & Tyner, A. L. Brk, Srm, Frk, and Src42A Form a Distinct Family of Intracellular Src-Like Tyrosine Kinases. *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics* **13,** 409–419 (2003).

75. Sicheri, F, Moarefi, I. & Kuriyan, J. Crystal structure of the the Src family tyrosine kinase Hck. *Nature* **385,** 602–9 (1997).

76. Sievers, F. *et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega* 2011. doi:10.1038/msb.2011.75.

77. Silva, C. M. Role of STATs as downstream signal transducers in Src family kinase-mediated tumorigenesis. *Oncogene* **23,** 8017–8023 (2004).

78. Singh, S. M. & Panda, A. K. Solubilization and refolding of bacterial inclusion body proteins. *Journal of bioscience and bioengineering* **99,** 303–10 (2005).

79. Siyanova, E., Serfas, M., Mazo, I. & Tyner, A. Tyrosine kinase gene expression in the mouse small intestine. *Oncogene* **9,** 2053–2057. ISSN: 0950-9232 (July 1994).

80. Stokes, M. P. *et al.* PTMScan Direct: Identification and Quantification of Peptides from Critical Signaling Proteins by Immunoaffinity Enrichment Coupled with LC-MS/MS. *Molecular Cellular Proteomics* **11,** 187–201 (2012).

81. Thomas, S. M. & Brugge, J. S. Cellular functions regulated by Src family kinases. *Annual review of cell and developmental biology* **13,** 513–609 (1997).

82. Tischer, A., Lilie, H., Rudolph, R. & Lange, C. L-arginine hydrochloride increases the solubility of folded and unfolded recombinant plasminogen activator rPA. *Protein Science* **19,** 1783–1795 (2010).

83. Vallejo, L. F. & Rinas, U. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microbial cell factories* **3,** 11 (2004).

84. Vranken, W. F. *et al.* The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins* **59,** 687–696 (2005).

85. Welsh, M. *et al.* The tyrosine kinase FRK/RAK participates in cytokine-induced islet cell cytotoxicity. *The Biochemical journal* **382,** 261–268 (2004).

86. Williams, J. C. *et al.* The 2.35 A crystal structure of the inactivated form of chicken Src: a dynamic molecule with multiple regulatory interactions. *Journal of molecular biology* **274,** 757–775 (1997).

87. Wingfield, P. T., Palmer, I. & Liang, S. M. Folding and purification of insoluble (inclusion body) proteins from Escherichia coli. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* **Chapter 6,** Unit 6.5 (2001).

88. Xu, W, Harrison, S. C. & Eck, M. J. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* **385,** 595–601 (1997).

89. Yamaguchi, H. & Miyazaki, M. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules* **4,** 235–51 (2014).

90. Yamaguchi, S., Yamamoto, E., Mannen, T., Nagamune, T. & Nagamune, T. Protein refolding using chemical refolding additives. *Biotechnology Journal* **8,** 17–31. ISSN: 1860-7314 (2013).

91. Yim, E. K. *et al.* Rak Functions as a Tumor Suppressor by Regulating PTEN Protein Stability and Function. *Cancer Cell* **15,** 304–314. ISSN: 15356108 (2009).

92. Zhou, X. *et al.* FRK controls migration and invasion of human glioma cells by regulating JNK/c-Jun signaling. *Journal of Neuro-Oncology* **110,** 9–19. ISSN: 0167594X (2012).

7 Appendix

7.1. German Abstract

Die Fyn-related Kinase (FRK) oder Rak ist eines der besser erforschten Mitglieder einer Familie von Nicht-Rezeptor Tyrosinkinasen, die sich aus FRK, PTK6/Brk (Breast Tumor Kinase) und SRMS (Src-related tyrosine kinase lacking C-terminal regulatory and N-terminal myristoylation sites) zusammensetzen. Im Gegensatz zu den strukturell nahe verwandten Src Family Kinasen (SFKs) haben die Mitglieder der FRK/PTK6 Familie keine N-terminale Lipidierung (Myristoylierung und/oder Palmitoylierung) und ebenso fehlt ihnen die Src-homolgy 4 (SH4) Domäne. Beiden Kinasefamilien sind die intrinsically disordered Unique Domäne, die Src-homolgy 3 und 2 (SH3 and SH2) Domänen sowie die C-terminale Kinasedomäne gemein.

Vorangegangene Untersuchungen der intrinsically disordered Unique Domäne von c-Src zeigten, dass diese Domäne mit der anschließenden SH3 Domäne und der cytoplasmatischen Seite der Zellmembran unter der Formation einer Peptidschlaufe innerhalb der Domäne interagiert. Diese Schlaufenbildung wurde u.a. auch in Fyn, einer weitern Src Family Kinase, beobachtet.

FRK besitzt zwei Cysteine in der Primärstruktur seiner Unique Domäne die in Regionen sind in denen sich konservierte aromatische Aminosäuren in den SFKs befinden, welche in die Formation der Peptidschlaufe involviert sind. Daher liegt die Hypothese nahe, dass diese beiden Cysteine an einer ähnlichen Schlaufenformation in FRK beteiligt sind, die zum Beispiel durch die Chelatierung von Metalionen zwischen den beiden Cysteinen bewirkt werden könnte.

Im Rahmen dieser Arbeit werden die erste Expression sowie Aufreinigung des N-terminalen Teils des humanen FRK Proteins, bestehend aus Unique und SH3 Domäne, beschrieben. Dieses Konstrukt wird in dieser Arbeit als FRK(USH3) bezeichnet.Ein vollständiges Protokoll, bestehend aus der Expression in *E.coli*, erfolgreichem Falten eines unlöslich-exprimierten Proteinkonstrukts und der Aufreinigung eines zur Aggregation neigendem Proteins für die Analyse über NMR Spektroskopie,wurde entwickelt. Des Weiteren wird das erste ^{1}H-^{15}N zwei-dimensionale NMR Spektrum von FRK(USH3) präsentiert.

Es konnte gezeigt werden, dass der N-terminal Teil von FRK bei Proteinkonzentrationen höher als 100µM eine starke Tendenz zur Aggregation zeigt und daher nur in einem sehr niedrigen Konzentrationsbereich für die Analyse mittels NMR Spektroskopie verwendet werden kann. Derivatisierungsversuche, die in einer Maskierung der Thiolgruppen in den Cysteinseitenketten und der Erhöhung der Nettoladung des Konstrukts resultierten, konnten die Löslichkeit geringfügig erhöhen und die Aggregationstendenz senken. Dadurch konnte die Qualität der aufgenommenen NMR Spektren verbessert

werden. Dynamic Light Scattering (DLS) und Native Polyacrylamidgelelektrophorese wurden verwendet um Unterschiede zwischen dem derivatisierten und nicht-derivatisierten Konstrukt zu analysieren.

7.2. Curriculum Vitae

SIMONE HÖFLER

Lindenstraβe 19C
Böheimkirchen, 3071, AUSTRIA
Phone: +43 664 1041176
Email: hoefler.si@gmail.com
Born: May 24, 1991, St. Pölten, Austria
Nationality: Austria

Education

2013 - 2015	**M.Sc. in Biological Chemistry**, *University of Vienna, Austria*
2010 - 2013	**B.Sc. in Biology/Molecular Biology**, *University of Vienna, Austria*

Research Experience

February to August 2015	**M.Sc. Student**, *BioNMR Group, University of Barcelona, Spain* *Supervisor*: Prof. Dr. Miquel Pons *Title*: Insights into the modulation of the interaction between the Unique and SH3 domains in Src Family kinases *Short Description*: The project is focussed on the study of the Unique and SH3 domain of Fyn-related Kinase (FRK). It involved the design and construction of an appropriate expression construct as well as site-specific mutagenesis. Expression tests were performed to develop and optimize an expression protocol. The purification of the construct involved solubilization and refolding from inclusion bodies. The interaction between the Unique and SH3 domain was studied by NMR spectroscopy.

July to August 2014	**Research Intern**, *Institute of Molecular Biotechnology (IMBA), Vienna, Austria* *Supervisor*: Thomas Marlovits, PhD *Short Description*: Investigation of protein complex formation by the needle forming protein in the type III secretion system of *Salmonella* using Transmission Electron Microscopy (TEM).
May to June 2013	**B.Sc. Student**, *Max F. Perutz Laboratories (MFPL), University of Vienna, Austria* *Supervisor*: Prof. Dr. Robert Konrat *Title*: NMR Analysis of Extracellular Matrix Protein Osteopontin reveals Binding of MAX Homodimer and Thermodynamic Analysis shows Binding Thermodynamics of Heparin and Osteopontin *Short Description*: The project involved the expression and purification of the intrinsically disordered protein (IDP) osteopontin (OPN). The potential binding of OPN to MAX (part of the Myc/MAX transcription factor) homodimer, which are no physiological binding partners, based on electrostatic interaction between the dipole of the homodimer and the charged residues in the IDP was investigated by NMR spectroscopy. Further the binding thermodynamics and stoichometry of OPN and it's physiological binding partner were analyzed by isothermal titration calorimetry at different temperatures.

Awards, Grants and Training

2011, 2012, 2013 and 2014	**Performance Scholarship** awarded by the *University of Vienna*

Publications

2014	Kurzbach, Dennis; Schwarz, Thomas C.; Platzer, Gerald; Höfler, Simone; Hinterberger, Dariush; Konrat, Robert (2014), **Compensatory Adaptions of Structural Dynamics in an Intrinsically Disordered Protein Complex**. *Angewandte Chemie (International Ed. in English), 53(15), 3840-3.* http://doi.org/10-1002/anie.201308389

Research Skills

DNA manipulation	In-depth theoretical and practical knowledge in the areas of DNA cloning and PCR- based site-specific mutagenesis;
Protein expression	Independent work competence in bacterial expression of structured and intrinsically disordered proteins for structural purposes using BL21(DE3) as well as Rosetta(DE3)pLysS *E. coli* strains and different expressions vector (pETM-30, pET16b and pET11d); Expression in full medium and minimal medium, as well as of unlabelled and labelled (^{15}N) protein;
Protein purification	Experience with purification from inclusion bodies and soluble proteins by means of different chromatographic methods (seize exclusion, ion exchange and IMAC);
Protein refolding	Experience in refolding from inclusion body proteins by means of direct and indirect dilution, dialysis and solid-state refolding;
NMR spectroscopy	Theoretical background knowledge and a basic knowledge in processing spectra (NMRPipe, Sparky, CcpNmr, rNMR) and interpretation;
Isothermal Titration Calorimetry (ITC)	Work experience with the calorimeter MicroCal IT200 MicroCalorimeter (GE Healthcare) and subsequence data processing with Origin 7 SR4 Software (OriginLab);
Transmission Electron Microscopy (TEM)	Basic skills in sample preparation;
Cell culture	Basic skills in mammalian cell culture;
LaTeX	Basic skills in creating documents;

Additional Qualifications and Skills

2015	**COUSERA** online course
	Course title: An Introduction to American Law
	Provider: University of Pennsylvania, USA
2015	**COUSERA** online course
	Course title: Programming for Everyone (Python)
	Provider: University of Michigan, USA

Languages

German	Native language
English	C1 Level (European Reference Frame)
Spanish	B1 Level (European Reference Frame)

Websites

Researchgate	`https://www.researchgate.net/profile/Simone_Hoefler`
LinkedIn	`https://es.linkedin.com/pub/simone-hoefler/a2/754/976`

List of Tables

List of Figures