„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
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 Tumor 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 thereby 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 cytoplasm and functions as an intracellular receptor and/or as signal transducer. Relating to Tyrosine protein kinases (TPKs) this classification results 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 CusatalW2 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 ubiquitously 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 it 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 flexible 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 UBLR was also reported to be involved in an allosteric regulation involving the neighboring 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 $\beta$-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 able to adopt 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 $\beta$-sheet flanked by an $\alpha$-helix on either side of the $\beta$-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 $\beta$-strands and one single $\alpha$-helix (which is referred to as C helix in Src) and the C-terminal lobe is mostly $\alpha$-helical and contains the activation loop. The nucleotide binding and the phosphotransfer happens in the cleft between the two lobes (Boggon & Eck 2004).
2.1. NON-RECEPTOR TYROSINE KINASES

CHAPTER 2. INTRODUCTION

Figure 2.1: Phylogenetic 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 to 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 (Cartwright, 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 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 histocompatibility 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 an immunoreceptor tyrosine-based activation motif (ITAM). This motif is defined by a consensus sequence Y-x-x-I/L-(7,8)-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 antigen 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-lymphocytes) 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 relocalized through the activation of integrin 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 integrin 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 receptor (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
2.1. NON-RECEPTOR TYROSINE KINASES

CHAPTER 2. INTRODUCTION

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 βγ- 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 βγ- 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 (signal transducer and activator of transcription), 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, Ca2+-channels and glutamate and N-acetylcholine receptors (Thomas & Brugge 1997).
2.1. NON-RECEPTOR TYROSINE KINASES

CHAPTER 2. INTRODUCTION

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 PKC\(\alpha\), 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 mediated 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 SFKs 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).

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 cytoplasm. 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, SL1M 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: KRLDEGGFFLTRRR), 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 be 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 sent for proteosomal degradation (Yim et al. 2009). Therefore phosphorylation leads to more stable PTEN which accumulates in the cytoplasm 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 happening in the G1 and S phase of the cell cycle, while FRK levels are highest during G1 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 G1 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 cellular membrane (Jin & Craven 2014).

**Figure 2.4:** Primary sequence of the Unique and SH3 domain of human FRK.
2.1. NON-RECEPTOR TYROSINE KINASES

CHAPTER 2. INTRODUCTION

The study described here is focused 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 induces conformational changes which delimitate a peptide loop within the Unique domain (Maffei et al. 2015; Pérez, Maffei, et al. 2013). To study the loop formation in c-Src, non-natural cysteines were introduced into the sequence that can be linked by the formation of a 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).

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 glutathione-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 the solubility of the whole GST-FRK(USH3) construct (Lebendiker & Danieli 2014). Plasmids 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 resolubilized with lysis buffer containing a high concentration 8M urea. For each condition both fraction were analyzed by SDS polyacrylamide gel electrophoresis (PAGE). Figure 3.1 shows the electrophoretic analysis for all samples collected at $25^\circ$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 lead 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.
### Table 3.1: Expression tests: tested conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Temperature (K)</th>
<th>Time (h)</th>
<th>IPTG [mM]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>288.15</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>288.15</td>
<td>5</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>288.15</td>
<td>o/n</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>288.15</td>
<td>o/n</td>
<td>0.1</td>
</tr>
<tr>
<td>5</td>
<td>288.15</td>
<td>o/n</td>
<td>0.5</td>
</tr>
<tr>
<td>6</td>
<td>288.15</td>
<td>o/n</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>288.15</td>
<td>o/n</td>
<td>0.1</td>
</tr>
<tr>
<td>8</td>
<td>288.15</td>
<td>o/n</td>
<td>0.5</td>
</tr>
<tr>
<td>9</td>
<td>288.15</td>
<td>o/n</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>293.15</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>11</td>
<td>293.15</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>12</td>
<td>293.15</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>293.15</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>14</td>
<td>293.15</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>15</td>
<td>293.15</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>293.15</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>17</td>
<td>293.15</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>18</td>
<td>293.15</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>293.15</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>20</td>
<td>293.15</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>21</td>
<td>293.15</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>293.15</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>23</td>
<td>293.15</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>24</td>
<td>293.15</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>293.15</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>26</td>
<td>293.15</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>27</td>
<td>293.15</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>303.15</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>29</td>
<td>303.15</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>30</td>
<td>303.15</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>31</td>
<td>303.15</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>32</td>
<td>303.15</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>33</td>
<td>303.15</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>303.15</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>35</td>
<td>303.15</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>36</td>
<td>303.15</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>37</td>
<td>310.15</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>38</td>
<td>310.15</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>39</td>
<td>310.15</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>40</td>
<td>310.15</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>41</td>
<td>310.15</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>42</td>
<td>310.15</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>43</td>
<td>310.15</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>44</td>
<td>310.15</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>45</td>
<td>310.15</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
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 resolubilized. 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-
3.2. GST-FRK(USH3) CAN BE SUCCESSFULLY REFOLDED

CHAPTER 3. RESULTS

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 losses for each technique, when GST-FRK(USH3) is expressed in LB.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Initial protein (mg)</th>
<th>Soluble fraction (mg)</th>
<th>Soluble fraction (%)</th>
<th>Precipitation (mg)</th>
<th>Precipitation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>direct dilution</td>
<td>34.53</td>
<td>6.98</td>
<td>20.2</td>
<td>27.55</td>
<td>79.8</td>
</tr>
<tr>
<td>indirect dilution (buffer 1)</td>
<td>34.28</td>
<td>27.31</td>
<td>79.7</td>
<td>6.97</td>
<td>20.3</td>
</tr>
<tr>
<td>indirect dilution (buffer 2 and 3)</td>
<td>34.14</td>
<td>34.07</td>
<td>99.8</td>
<td>0.07</td>
<td>0.2</td>
</tr>
<tr>
<td>on-column refolding</td>
<td>34.51</td>
<td>0.67</td>
<td>1.9</td>
<td>33.84</td>
<td>98.1</td>
</tr>
</tbody>
</table>

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 by 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 than 0.8mg/mL as visible insoluble aggregates are formed. This suggests that the fusion tag increases the solubility of FRK(USH3) at least to some 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 slightly 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).

<table>
<thead>
<tr>
<th></th>
<th>initial protein (mg)</th>
<th>final protein (mg)</th>
<th>protein loss (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-derivatised FRK(USH3)</td>
<td>20</td>
<td>0.06</td>
<td>99.7</td>
</tr>
<tr>
<td>derivatised FRK(USH3)</td>
<td>20</td>
<td>0.06</td>
<td>99.7</td>
</tr>
</tbody>
</table>

Table 3.3: Proteins yields from 100mL of minimal medium expression from the expression to the final protein sample.
3.3.1. Nuclear Magnetic Resonance (NMR) spectroscopy

Theoretical Background

NMR spectroscopy was the main technique used to study the novel FRK construct, consisting of 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 \textit{wavefunction} is a mathematical function which contains all the information about a given system, therefore if the \textit{wavefunction} of the system under consideration is known, anything about the system can be deduced. In order to access this information quantum mechanical \textit{operators} are needed, as in quantum mechanics they represent "observables", therefore things that can be measured. \textit{Operators} have \textit{eigenfunctions} with corresponding \textit{eigenvalues} associated with them \textit{(eigenvalue equation, Equation 3.1)}.

\begin{equation}
\text{(operator) acting on (eigenfunction) } = \text{(eigenvalue) x (eigenfunction)}
\end{equation}

The \textit{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 Hamiltonian

\begin{equation}
H = \sum_\sigma \left( \frac{\sigma \cdot B_0}{\hbar} \right) I_\sigma
\end{equation}
in Equation 3.2.

\[ \hat{H}_{\text{onespin}} = -\gamma B_0 \hat{I}_z \]  

(3.2)

With \( \gamma \) being the gyromagnetic ratio (usually given in \( \text{rads}^{-1}\text{T}^{-1} \)), 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 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 \(^1\text{H}\) nucleus with \( I = \frac{1}{2} \), \( m \) can be \(-\frac{1}{2}\) or \(+\frac{1}{2}\). With \( \Psi_m \) being the general eigenfunction the eigenvalue equation results as shown in Equation 3.3.

\[ \hat{I}_z \Psi_m = mh \Psi_m \]  

(3.3)

With \( h \) being the reduced Planck constant. Therefore \( \Psi_m \) has the eigenvalue \( mh \). 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).

\[ \hat{H}_{\text{onespin}} \Psi_m = -m h \gamma B_0 \Psi_m \]  

(3.4)

\[ E_m = m h \gamma B_0 \]  

(3.5)

As \( m \) depends 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 \(^1\text{H}\) two different energy levels are available, were \( m = +\frac{1}{2} \) is labeled the \( \alpha \) state and \( m = -\frac{1}{2} \) the \( \beta \) state. Therefore the transition between this two states corresponds to following energy difference (Equation 3.6).

\[ \Delta E_{\alpha \leftrightarrow \beta} = h \gamma B_0 \]  

(3.6)

The frequency, in Hz and \( \text{rads}^1 \), corresponding to \( \Delta E \) is given in Equation 3.7 and 3.8.

\[ \nu_{\alpha \leftrightarrow \beta} = \gamma B_0 / 2\pi \]  

(3.7)

\[ \omega_{\alpha \leftrightarrow \beta} = \gamma B_0 \]  

(3.8)

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

\[ \nu_{\alpha \leftrightarrow \beta} = -\nu_0 \]  

(3.9)

Therefore for systems consisting only of one single spin with \( I = \frac{1}{2} \), like \(^1\text{H}\), 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).

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

(3.10)

With $\delta$ 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 $\nu_0$ contain information on the chemical environment and the spatial position of the nuclei. Equation 3.11 accounts for $B_{eff}$.

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

(3.11)

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

(3.12)

With $\sigma$ 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 Single-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, Kupé, 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 187$\mu$M (Figure 3.3) and 113$\mu$M (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 110$\mu$M. 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.
3.3. FRK(USH3) TENDS TO FORM AGGREGATES

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
3.3. **FRK(USH3) TENDS TO FORM AGGREGATES**

As a consequence of the poor quality of the spectra at high concentrations a new protocol was established, 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 13 mM, 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 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

---

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

$^1$H-$^{15}$N HSQC spectrum of GST, 110μM, 298.15K
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
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 10.50\textmu M, 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 tryptophane residues in the SH3 domain around 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
CHAPTER 3. RESULTS

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.

Figure 3.10: Mass spectrometric analysis of derivatised FRK(USH3)
3.3. *FRK(USH3) TENDS TO FORM AGGREGATES*  

**CHAPTER 3. RESULTS**

**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) (10.50µM)
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. 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).

$$D_t = \frac{k_B T}{6 \pi \eta R_H}$$  \hspace{1cm} (3.13)

With $D_t$ being the diffusion coefficient, $k_B$ the Boltzman’s constant ($1.381 \times 10^{-23}$ J/K), $\eta$ 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 caused 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).

$$g_2(q; \tau) = \frac{\langle I(t)I(t+\tau) \rangle}{\langle I(t) \rangle^2}$$  \hspace{1cm} (3.14)

With $g_2(q; \tau)$ being the second order autocorrelation function at a specific wave vector, $q$, and delay time, $\tau$, 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 and 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.

$$g_1(q; \tau) = \exp(-\Gamma \tau)$$  \hspace{1cm} (3.15)

With $\Gamma$ being the decay rate. The relationship between the decay rate $\Gamma$ and $D_t$ through the wave
vector, \( q \), is given by Equation 3.4 and 3.5.

\[
\Gamma = q^2 D_t
\]  
\[
q = \frac{4\pi n_0}{\lambda} \sin\left(\frac{\Theta}{2}\right)
\]

With \( n_0 \) being the refractive index of the sample, \( \lambda \) the wavelength of the incident laser and \( \Theta \) the angle at which the scattered light is detected. The vector \( q \) is a function of angle \( \Theta \) and the incident wavelength \( \lambda \) 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 multimodal 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).

<table>
<thead>
<tr>
<th></th>
<th>none-derivatised FRK(USH3)</th>
<th>derivatised FRK(USH3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z-average (nm)</td>
<td>363.0</td>
<td>277.4</td>
</tr>
<tr>
<td>Polydispersity Index (PDI)</td>
<td>0.491</td>
<td>0.862</td>
</tr>
</tbody>
</table>

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.
3.3. FRK(USH3) TENDS TO FORM AGGREGATES

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. Native PAGE

The separation of protein in native PAGE is based on the same separation principal as denaturing SDS-PAGE, except 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.

![Image of native PAGE gel showing protein migration](image_url)
3.3. FRK(USH3) TENDS TO FORM AGGREGATES

CHAPTER 3. RESULTS

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).
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 happening 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 $^1H$,$^{15}N$ two-dimensional correlation spectra
could only be recorded at protein concentrations around 10$\mu$M. 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 gelelectrophore-
sis (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 inves-
tigated 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 CSE, 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$^\text{TM}$ PCR Kit (New England Biolabs, Inc.) following the manufactures introductions. All primers were evaluated using the OligoEvaluator$^\text{TM}$ online tool (Sigma-Aldrich$^\text{R}$) and pursued from Sigma-Aldrich$^\text{R}$ (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.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Length (nt)</th>
<th>Tm (C)</th>
<th>GC %</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward(NcoI) S2A</td>
<td>27</td>
<td>76.5</td>
<td>51.8</td>
<td>TATACCATGGCCAACATTGAGCGCCG</td>
</tr>
<tr>
<td>Reverse(XhoI)</td>
<td>30</td>
<td>73.1</td>
<td>50</td>
<td>TATACCTCGAGTTATTCGGCCTGCAGAGAGC</td>
</tr>
<tr>
<td>Forward A2S</td>
<td>24</td>
<td>76.8</td>
<td>58.3</td>
<td>CAGGGCGCCATGTCCAACATTGAG</td>
</tr>
<tr>
<td>Reverse A2S</td>
<td>24</td>
<td>76.8</td>
<td>58.3</td>
<td>CTCAATGTGGACATGGCGCCCTG</td>
</tr>
</tbody>
</table>

Table 5.1: Primer specifications and sequence

5.2 ProteinExpression 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 50mL lysis buffer (100mM Tris-HCl, 150mM NaCl, 0.5mM EDTA, 1% NaN₃, 20mM 2-mercaptoethanol, PMSF (phenyl-methylsulfonyl 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 ⇠ -180°C) and thawing (water bath ⇠ 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 ¹⁵N-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 Na₂HPO₄, 22mM KH₂PO₄, 8,5mM NaCl, 0.2% glucose, 1mM MgSO₄, 0.3mM CaCl₂, 1µg thiamine, 1µg biotin, trace elements) supplemented with ¹⁵N (0.5g/ml of ¹⁵NH₄Cl) 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 ¹⁵N-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 slowly 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, 0.5mM 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 ¹⁵N-labeled FRK(USH3)

After establishing a robust refolding protocol the highly diluted protein was dialyzed (3.5kDa cut-off, Spectrum Laboratories, 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 His₆-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 Laboratories, Inc.) against TEV cleavage buffer (50mM Tris-HCl, 50mM NaCl, 1mM DTT, 0.5mM EDTA, pH 8.0) o/n at 4°C.

For removal of the His₆-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 shaking. 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 target protein gelfiltration using a hand-packed Superdex 75 26&60 (GE Healthcare Bio-Sciences) was applied. The Column was operated using an AKTA 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 (NaH₂PO₄, Na₂HPO₄), 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 lyophilized 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$_2$PO$_4$, Na$_2$HPO$_4$), 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.

<table>
<thead>
<tr>
<th>Sample concentration</th>
<th>none-derivatised FRK(USH3)</th>
<th>derivatised FRK(USH3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample concentration</td>
<td>5µM (diluted with 1% formic acid)</td>
<td>1pmol/µL (diluted in 1% formic acid)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chromatograph</th>
<th>Acquity UPLC Binary Sol MGR (Waters Corporation)</th>
<th>nanoAcquity UPLC (Waters Corporation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Column</td>
<td>BioSuite pPhenyl 1000RPC 2.0x75mm, 10µm</td>
<td>Trap loading, separation on column</td>
</tr>
<tr>
<td>Eluents</td>
<td>A: Water 1% formic acid</td>
<td>A: Water 0.1% formic acid</td>
</tr>
<tr>
<td>Gradient</td>
<td>5% to 80% of B in 60min. Total chromatographic time: 86min</td>
<td>10% to 80% B in 30min + 80% to 90% in 5min</td>
</tr>
<tr>
<td>Flow rate</td>
<td>100µL/min</td>
<td>300µL/min</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Mass spectrometer</th>
<th>none-derivatised FRK(USH3)</th>
<th>derivatised FRK(USH3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polarity</td>
<td>ESI positive</td>
<td>ESI positive</td>
</tr>
<tr>
<td>Capillary voltage</td>
<td>3000V</td>
<td>34V</td>
</tr>
<tr>
<td>Capillary temperature</td>
<td>120°C</td>
<td>200°C</td>
</tr>
<tr>
<td>Tube Lens</td>
<td>100V</td>
<td>100V</td>
</tr>
<tr>
<td>m/z range</td>
<td>400-5000</td>
<td>400-2000</td>
</tr>
<tr>
<td>Data Processing</td>
<td>MassLynx software V4.1. SCN639 (Waters Inc.)</td>
<td>Xcalibur software vs 2.0SR2 (Thermo Scientific)</td>
</tr>
</tbody>
</table>

**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-ISCT, supported by grant PRB2 (IPT13/0001 - ISCIII-SGPF / FEDER)"
5.4. NMR Spectroscopy

Sample preparation

Samples of 250\( \mu \)L or 300\( \mu \)L in NMR buffer (20mM Sodium phosphate (NaH\(_2\)PO\(_4\), Na\(_2\)HPO\(_4\)), 1mM DTT or10mM 2-mercaptoethanol pH 7.0) and 10% D\(_2\)O 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 \( \varepsilon = 25 440 \) (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 6°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 200\( \mu \)M 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).

<table>
<thead>
<tr>
<th>Spectrum</th>
<th>Sample</th>
<th>Sample Volume (( \mu )L)</th>
<th>Sample concentration (( \mu )M)</th>
<th>Experiment Type</th>
<th>Num. Scans</th>
<th>Temperature (K)</th>
<th>Experimental Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1</td>
<td>none-derivatised FRK(USH3)</td>
<td>250</td>
<td>187</td>
<td>HSQC</td>
<td>24</td>
<td>298.15</td>
<td>1h 58min</td>
</tr>
<tr>
<td>A.2</td>
<td>none-derivatised FRK(USH3)</td>
<td>250</td>
<td>187</td>
<td>SOFAST-HMQC</td>
<td>16</td>
<td>298.15</td>
<td>25min</td>
</tr>
<tr>
<td>A.3</td>
<td>none-derivatised FRK(USH3)</td>
<td>250</td>
<td>187</td>
<td>HSQC</td>
<td>24</td>
<td>278.15</td>
<td>1h 58min</td>
</tr>
<tr>
<td>A.4</td>
<td>none-derivatised FRK(USH3)</td>
<td>250</td>
<td>187</td>
<td>SOFAST-HMQC</td>
<td>16</td>
<td>278.15</td>
<td>25min</td>
</tr>
<tr>
<td>B.1</td>
<td>none-derivatised FRK(USH3)</td>
<td>250</td>
<td>113</td>
<td>SOFAST-HMQC</td>
<td>128</td>
<td>298.15</td>
<td>2h 30min</td>
</tr>
<tr>
<td>B.2</td>
<td>none-derivatised FRK(USH3)</td>
<td>250</td>
<td>113</td>
<td>SOFAST-HMQC</td>
<td>128</td>
<td>278.15</td>
<td>2h 30min</td>
</tr>
<tr>
<td>C</td>
<td>GST</td>
<td>250</td>
<td>110</td>
<td>HSQC</td>
<td>64</td>
<td>298.15</td>
<td>1h 10min</td>
</tr>
<tr>
<td>D.1</td>
<td>none-derivatised FRK(USH3)</td>
<td>300</td>
<td>12.25</td>
<td>SOFAST-HMQC</td>
<td>64</td>
<td>278.15</td>
<td>1h 14min</td>
</tr>
<tr>
<td>D.2</td>
<td>none-derivatised FRK(USH3)</td>
<td>300</td>
<td>12.25</td>
<td>SOFAST-HMQC</td>
<td>64</td>
<td>278.15</td>
<td>1h 14min</td>
</tr>
<tr>
<td>D.3</td>
<td>none-derivatised FRK(USH3)</td>
<td>300</td>
<td>12.25</td>
<td>SOFAST-HMQC</td>
<td>64</td>
<td>278.15</td>
<td>1h 14min</td>
</tr>
<tr>
<td>D.4</td>
<td>none-derivatised FRK(USH3) + Zn(^{2+}) and EDTA</td>
<td>300</td>
<td>12.25</td>
<td>SOFAST-HMQC</td>
<td>64</td>
<td>278.15</td>
<td>1h 14min</td>
</tr>
<tr>
<td>E.1</td>
<td>none-derivatised FRK(USH3)</td>
<td>300</td>
<td>12.25</td>
<td>SOFAST-HMQC</td>
<td>64</td>
<td>298.15</td>
<td>51min</td>
</tr>
<tr>
<td>E.2</td>
<td>none-derivatised FRK(USH3)</td>
<td>300</td>
<td>12.25</td>
<td>HSQC</td>
<td>64</td>
<td>298.15</td>
<td>5h 5min</td>
</tr>
<tr>
<td>F.1</td>
<td>derivatised FRK(USH3)</td>
<td>300</td>
<td>10.50</td>
<td>SOFAST-HMQC</td>
<td>64</td>
<td>298.15</td>
<td>51min</td>
</tr>
<tr>
<td>F.2</td>
<td>derivatised FRK(USH3)</td>
<td>300</td>
<td>10.50</td>
<td>HSQC</td>
<td>64</td>
<td>298.15</td>
<td>5h 15min</td>
</tr>
</tbody>
</table>

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 (NaH₂PO₄, Na₂HPO₄), 1mM DTT, pH 7.0). Measurements were performed at 25°C (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 polyacrylamide 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.

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

Table 5.5: Native PAGE gel composition
## Abbreviations and acronyms

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


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.


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 Proteinkonstruks 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 verwen-
det um Unterschiede zwischen dem derivatisierten und nicht-derivatisierten Konstrukt zu analysieren.
7.2. Curriculum Vitae

SIMONE HÖFLER

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

Education

<table>
<thead>
<tr>
<th>Year</th>
<th>Degree</th>
<th>Institution</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013 - 2015</td>
<td>M.Sc. in Biological Chemistry</td>
<td>University of Vienna</td>
<td>Austria</td>
</tr>
<tr>
<td>2010 - 2013</td>
<td>B.Sc. in Biology/Molecular Biology</td>
<td>University of Vienna</td>
<td>Austria</td>
</tr>
</tbody>
</table>

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 stoichiometry 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 Adapotions 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](http://doi.org/10-1002/anie.201308389)
### Research Skills

<table>
<thead>
<tr>
<th>Area</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DNA manipulation</strong></td>
<td>In-depth theoretical and practical knowledge in the areas of DNA cloning and PCR-based site-specific mutagenesis;</td>
</tr>
<tr>
<td><strong>Protein expression</strong></td>
<td>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;</td>
</tr>
<tr>
<td><strong>Protein purification</strong></td>
<td>Experience with purification from inclusion bodies and soluble proteins by means of different chromatographic methods (seize exclusion, ion exchange and IMAC);</td>
</tr>
<tr>
<td><strong>Protein refolding</strong></td>
<td>Experience in refolding from inclusion body proteins by means of direct and indirect dilution, dialysis and solid-state refolding;</td>
</tr>
<tr>
<td><strong>NMR spectroscopy</strong></td>
<td>Theoretical background knowledge and a basic knowledge in processing spectra (NMRPipe, Sparky, CcpNmr, rNMR) and interpretation;</td>
</tr>
<tr>
<td><strong>Isothermal Titration</strong></td>
<td>Work experience with the calorimeter MicroCal IT200 MicroCalorimeter (GE Healthcare) and subsequence data processing with Origin 7 SR4 Software (OriginLab);</td>
</tr>
<tr>
<td><strong>Transmission Electron Microscopy (TEM)</strong></td>
<td>Basic skills in sample preparation;</td>
</tr>
<tr>
<td><strong>Cell culture</strong></td>
<td>Basic skills in mammalian cell culture;</td>
</tr>
<tr>
<td><strong>EtPEx</strong></td>
<td>Basic skills in creating documents;</td>
</tr>
</tbody>
</table>


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](https://www.researchgate.net/profile/Simone_Hoefler)
- **LinkedIn**  
  [https://es.linkedin.com/pub/simone-hoefler/a2/754/976](https://es.linkedin.com/pub/simone-hoefler/a2/754/976)
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Expression tests: tested conditions</td>
<td>15</td>
</tr>
<tr>
<td>3.2</td>
<td>Protein yields after refolding. Indirect dilution using refolding buffer 2 or 3 yields the highest proportion of successfully refolded GST-FRK(USH3) (grey).</td>
<td>17</td>
</tr>
<tr>
<td>3.3</td>
<td>Proteins yields from 100mL of minimal medium expression from the expression to the final protein sample.</td>
<td>18</td>
</tr>
<tr>
<td>3.4</td>
<td>Measured Z-average (d.nm) and PDI for none-derivatised and derivatised FRK(USH3)</td>
<td>32</td>
</tr>
<tr>
<td>5.1</td>
<td>Primer specifications and sequence</td>
<td>38</td>
</tr>
<tr>
<td>5.2</td>
<td>LC and nanoLC conditions for none-derivatised and derivatised FRK(USH3)</td>
<td>41</td>
</tr>
<tr>
<td>5.3</td>
<td>MS conditions for none-derivatised and derivatised FRK(USH3)</td>
<td>41</td>
</tr>
<tr>
<td>5.4</td>
<td>List of conducted NMR experiments</td>
<td>42</td>
</tr>
<tr>
<td>5.5</td>
<td>Native PAGE gel composition</td>
<td>43</td>
</tr>
</tbody>
</table>
List of Figures

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) ......................................................... 5

2.2 Tree diagram displaying SFKs and FRK/PTK6 kinases according to differentiating properties. ................................................................. 5

2.3 Domain structure of c-Src and FRK. ................................................................. 9

2.4 Primary sequence of the Unique and SH3 domain of human FRK. ................................. 11

2.5 Primary sequence of the Unique and SH3 domain of human FRK with a cysteine to glutamic acid mutation at position two (CSE). ................................................................. 12

2.6 Primary sequence alignment of the Unique domain of c-Src homologues and SFKs includ- ing c-Src, Yes, Fyn and Fgr, highlighting conserved aromatic residues involved in a long-distance interaction (orange). ................................................................. 12

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 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 ........................................ 16

3.2 Mass spectrometric analysis of FRK(USH3): mass spectrum of FRK(USH3) ................... 19

3.3 Spectrum A.1: 1H-15N HSQC spectrum of FRK(USH3), 187μM, 298.15K .......................... 23

3.4 Spectrum B.1: 1H-15N SOFAST-HMQC spectrum of FRK(USH3), 113μM, 298.15K .......... 23

3.5 Spectrum C: 1H-15N HSQC spectrum of GST, 110μM, 298.15K ............................... 24

3.6 Spectrum D.2: 1H-15N SOFAST-HMQC spectrum of FRK(USH3), 13.25μM, 298.15K .... 25

3.7 Spectrum D.3: 1H-15N SOFAST-HMQC spectrum of FRK(USH3) (13μM) with 200μM ZnCl2, 298.15K ................................................................. 26

3.8 Overlay of spectrum D.2 and D.3: Overlay of the 1H-15N SOFAST-HMQC spectra of FRK(USH3) (13.25μM) (red) and FRK(USH3)(13.25μM) with 200μM ZnCl2 (green), 298.15K ................................................................. 26
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 ZnCl₂ (green), 298.15K - Detail ................................................................. 27
3.10 Mass spectrometric analysis of derivatised FRK(USH3) ......................................................... 28
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) ................................................................. 29
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) (10.50μM) ................................................................. 29
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) ................................. 30
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 ................................. 30
3.15 Size distribution by intensity for all measurements conducted of none-derivatised FRK(USH3) in solution. .................................................................................................................. 32
3.16 Size distribution by intensity for all measurements conducted of derivatised FRK(USH3) in solution. .................................................................................................................. 33
3.17 Size distribution by intensity for all measurements conducted of none-derivatised and derivatised FRK(USH3) in solution. .................................................................................. 33
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. .......... 34
3.19 12% Native PAGE gel of none-derivatised and derivatised FRK(USH3), (1+2) derivatised FRK(UAH3); (3+4) none-derivatised FRK(UAH3). ...................................................... 35