Src homology domain-mediated protein interactions

Lindfors, H.E.

Chapter 1

General introduction

Signal transduction and protein-protein interactions

Cells in multicellular organisms constantly receive and react to external stimuli. Signals from outside the cell initiate signal transduction cascades within the cell, and in this way diverse external signals are detected, amplified and integrated to generate cellular responses such as changes in enzyme activity, gene expression or ion-channel activity. An appropriate response to external signals is necessary for the proper functioning of individual cells and for multicellular life to be possible. A central feature of signal transduction and many biological processes is the ability of proteins to bind each other in a highly specific manner. Studying protein complex formation in detail and understanding the forces that drive the interaction is therefore of great interest.

Protein complexes vary greatly in their properties, with equilibrium dissociation constants (K_d) spanning many orders of magnitude. Some proteins form stable complexes, interacting for a long time, whereas others interact only briefly. The properties of protein complexes are related to their biological functions. Antibody-antigen complexes or enzyme-inhibitor complexes require tight binding and high specificity, to ensure a proper immune response or strict control of enzyme activity. In contrast, proteins involved in signal transduction cascades or in electron transfer often need to interact with multiple partners and maintain a high turnover. Consequently, these protein complexes tend to be more transient and to display a lower binding affinity.

In our current understanding of protein complex formation at least two steps are involved, with the first step being the formation of an encounter complex. This involves the proteins coming together, mainly with the help of long-range electrostatic forces, to form a loosely-bound intermediate state. From the encounter complex the proteins can either dissociate or proceed to form a final complex involving short-range interactions such as hydrogen bonding, van der Waals forces

and the hydrophobic effect. The role of the encounter complex is to accelerate the rate of specific complex formation, by reducing the dimensionality in the diffusional search process and increasing the lifetime of macromolecular collisions [1].

The equilibrium between the encounter complex and the productive complex varies between protein complexes. Some protein complexes exist mainly as a specific, well-defined complex, whereas in other protein-protein interactions the encounter complex is populated for a significant part of the time. Some electron transfer complexes can even exist purely as an encounter complex, never proceeding to form a specific complex [2].

Phosphotyrosine signalling and modular proteins

In order for multicellular life to be possible cell proliferation, differentiation, adhesion and motility need to be strictly controlled. Many of these processes are regulated by tyrosine phosphorylation, which is believed to have been necessary for the transition from single-cell to multicellular organisms [3-5]. Tyrosine phosphorylation, the covalent addition of a phosphate group to the hydroxide group in the side chain of a tyrosine residue in a protein, is regulated by two groups of enzymes: protein tyrosine kinases and protein tyrosine phosphatases. Protein tyrosine kinases catalyze the transfer of a phosphate group from ATP to a tyrosine residue, and this action is opposed by protein tyrosine phosphatases that catalyze the reaction of phosphate removal. Addition of a phosphate group to a tyrosine residue creates a high-affinity binding site for Src homology 2 (SH2) domains. This leads to the formation of new protein complexes, and thereby, to the transmission of the signal. Tyrosine phosphorylation signalling can therefore be considered to consist of three components: A ‘writer’ (the kinase), a ‘reader’ (the SH2 domain) and an ‘eraser’ (the phosphatase), which can be combined to generate

remarkably diverse signalling responses [6], including hormone-, growth factor-, immune-, and adhesion-based signalling [7-10]. Because of their involvement in so many signalling pathways and regulatory events protein tyrosine kinases are important drug targets. Many human diseases are recognized to be associated with abnormal phosphorylation of cellular proteins resulting from disregulation of kinase activity [11].

Despite its importance, tyrosine phosphorylation is still a relatively rare event in cells compared to the more common serine/threonine phosphorylation, and it was discovered only in 1979 [12]. Like many important scientific discoveries the finding of protein tyrosine phosphorylation was a serendipitous event. In the processes of determining whether a protein was phosphorylated on serine or threonine residues, Tony Hunter used an old buffer in which the pH had changed to a point that allowed phosphotyrosine to be separated from phosphothreonine [13].

The main sites of tyrosine phosphorylation in the cell are focal adhesions, the sites of attachment of the cell to the extracellular matrix (ECM). At focal adhesions integrin receptors link ECM proteins to the actin cytoskeleton involving a multitude of signalling and adaptor proteins (Fig. 1.1). Focal adhesions perform at least two important functions in the cell, they transmit force or tension at adhesion sites in order to maintain strong attachments to the ECM, and they are of central importance in many signalling pathways that regulate cell growth, survival and gene expression [14].

Many eukaryotic signalling proteins are modular proteins, containing several individually folded domains connected by linker regions. These domains can be protein-interaction domains or domains with a catalytic function. Common for these signalling proteins is that their activity is tightly regulated. The activity is normally low under basal conditions, but the proteins can be activated by specific ligands binding to the protein-interaction domains. This way the activity is

intrinsically coupled to protein localization. Recognition of short peptide sequences by modular interaction domains plays a central role in regulating cellular behaviour, since it is via these protein-protein interactions that the assembly of signalling protein complexes and larger protein networks can occur [15;16].

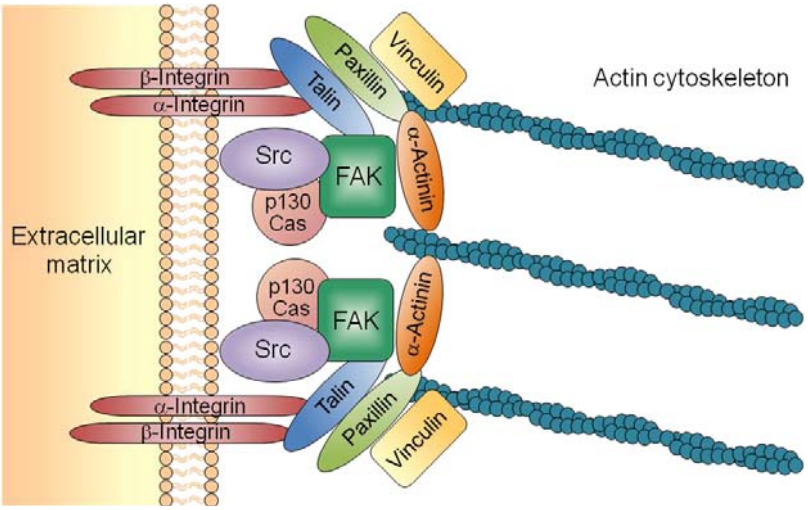

Figure 1.1. Focal adhesions. At focal adhesions integrin receptors bind to extracellular matrix proteins. A number of proteins, including talin, paxillin, vinculin and α-actinin, bind to the cytoplasmic tails of integrins, linking the integrins to the actin cytoskeleton. The large protein complexes also contain signalling proteins such as FAK and Src that promote focal adhesion turnover and cell motility.

Two proteins with a central role at focal adhesions are the non-receptor protein tyrosine kinases focal adhesion kinase (FAK) and Src kinase. FAK and Src are involved in a number of processes such as cell proliferation, survival and migration [17;18]. Increased activity and expression of Src and FAK has been demonstrated in many human cancers and is implicated in increased metastatic potential and invasiveness of tumour cells [19-28].

Proteins

Src

In 1911, Peyton Rous discovered that a cell-free filtrate of a chicken tumour was able to induce the formation of tumours in other chickens. Rous concluded that the tumour was caused by a 'filterable agent', as viruses were known at the time [29]. The idea that the virus, later called Rous sarcoma virus, could cause cancer was at first controversial, but it was later proven that the cancer-causing ability of the virus could be attributed to a viral gene, v-src [30]. A cellular counterpart of v-Src, Src, was subsequently discovered and found to be conserved in the vertebrate genome, indicating that the v-src gene had been incorporated into the viral genome through recombination. V-Src differs from Src in deletions at the C-terminal and in point mutations throughout the gene [31]. Unlike v-Src, Src is not constitutively active and is poorly transforming under normal conditions, but can act as an oncogene when activated [32;33]. This makes Src a proto-oncogene, the first of many to be discovered [34].

Src has a molecular weight of 60 kDa. It is expressed ubiquitously, but with the highest levels in the brain, osteoclasts and platelets [35]. It is a member of the Src family of protein tyrosine kinases that also includes Fyn, Yes, Lck, Hck, Blk, Fgr, Lyn, Yrk, Brk and Srm [36]. The members of the Src family share a conserved domain structure consisting of an N-terminal myristoylated SH4-domain followed by a region unique to each family member, an SH3 domain, an SH2 domain, a kinase domain and a C-terminal regulatory region [37]. The myristoylation facilitates attachment of Src to membranes. The SH3 domain and the SH2 domain are involved in protein-protein recognition, and facilitate the interaction of Src with its substrates.

The Src SH3 domain is about 60 amino acid residues in size. It has a β-barrel structure consisting of five antiparallel β-strands and two loops, known as the RT

and n-Src loops [38] (Fig. 1.2). SH3 domains bind to sequences that can adopt a left-handed helical conformation. These sequences often contain a characteristic proline-rich motif, PxxP. Src SH3 target sequences can be divided into two classes, which bind in opposite orientations to the SH3 domain. The binding orientation is largely determined by the position of an arginine residue close to the proline-rich core motif [39].

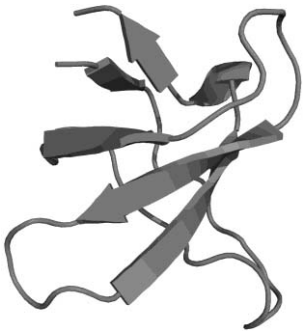

Figure 1.2. Solution structure of the Src SH3 domain (PDB entry 1SRL [40]).

The SH2 domain of Src contains about 100 aminoacids. It recognises and binds to sequences containing a phosphorylated tyrosine residue. The structure consists of a central β-sheet flanked by two α-helices, with connecting loops in-between [41] (Fig. 1.3). The preferred sequence for Src SH2 domain-binding is pYEEI [42], and the binding has been described by the 'two-pronged plug two-holed socket' model, where the phosphotyrosine is inserted into a pocket containing a conserved arginine residue, and the isoleucine at position pY+3 binds to a hydrophobic pocket [43].

The Src kinase domain consists of a small amino-terminal lobe, with a predominantly antiparallel β-sheet structure, and a larger carboxyl-terminal lobe that is mostly α-helical. The catalytic site is situated in a cleft between the two lobes [44].

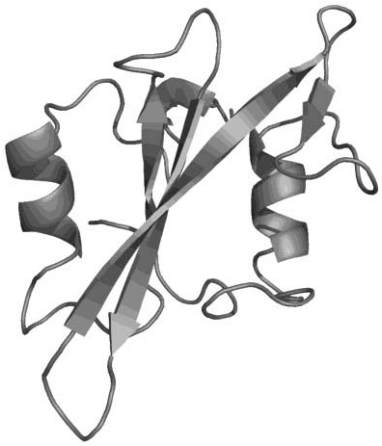

Figure 1.3. Solution structure of the Src SH2 domain (PDB entry 1HCS [45]).

Activation and regulation of Src

Src is an important component in many signalling pathways, and can be activated in different ways, including activation by receptor tyrosine kinases [46]. Several mechanisms of activation and regulation of Src have been proposed [47].

The C-terminal part of Src contains a regulatory tyrosine residue, Y529 (if not stated otherwise mouse Src numbering is used throughout this work), which can be phosphorylated by the tyrosine kinase c-Src terminal kinase (Csk). When Y529 is phosphorylated the SH2 domain binds to this region, while at the same time the SH3 domain binds to the linker region between the SH2 domain and the kinase domain. Together these intramolecular interactions cause the protein to assume a closed, inactive conformation [48;49]. When the C-terminal phosphate is removed, Src assumes an open, active form (Fig. 1.4). In contrast to Src, v-Src lacks the negative-regulatory element, and is constitutively active. Full activation of Src also requires the phosphorylation of a tyrosine residue in the kinase domain, Y418, through autophosphorylation [50].

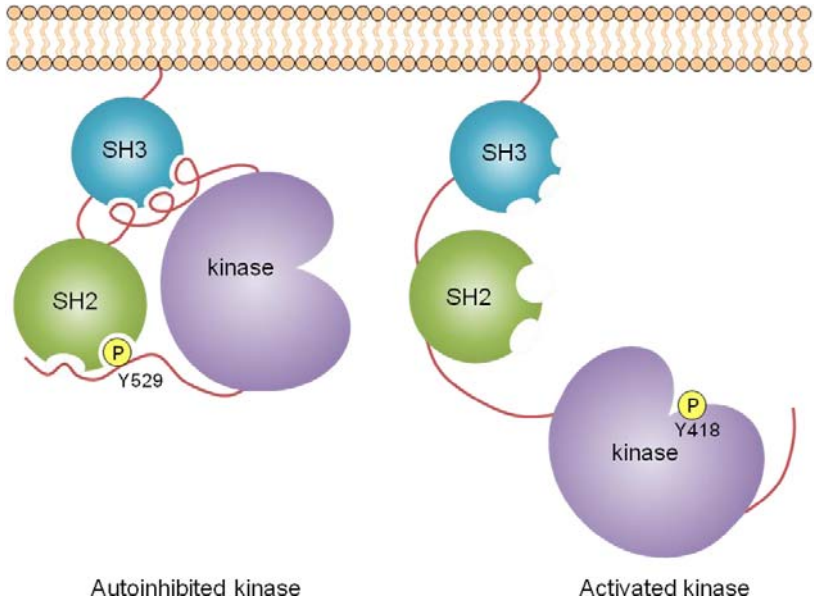

Figure 1.4. Activation of Src. Phosphorylation of Y529 in the C-terminal tail of Src by Csk causes the tail to bind to the SH2 domain. Together with interactions between the SH3 domain and the SH2-kinase linker, this locks the protein into an inactive conformation. Removal of the phosphate group by cellular phosphatases and interactions of the SH3 and SH2 domains with external ligands opens the protein up into an active conformation. Phosphorylation of Y418 in the kinase domain stabilizes the active conformation and is required for full activation of Src. Figure adapted from [51].

A likely mechanism for activation of Src is the removal of the C-terminal phosphate group by protein tyrosine phosphatases. Elevated levels of the protein tyrosine phosphatase PTP1B, which is able to dephosphorylate Src, has been found in breast cancer cell lines [52].

Competition between the low-affinity intramolecular binding sites for the SH2 and SH3 domains and high affinity binding sites in other proteins is another possible mechanism of activation. Upon binding of a ligand by the SH2 or SH3 domain, the closed, inactive conformation of Src would be disrupted and the protein would assume an open, active form instead. The use of domains for autoinhibition enhances the specificity – since the SH2 and SH3 domains already have

intramolecular binding sites, only specific binding sites of higher affinity can bind to the domains and activate the protein [53].

Focal adhesion kinase

FAK, initially identified in 1992 [54;55], is the only member in the FAK family of nonreceptor tyrosine kinases apart from PYK2. FAK can be found in the majority of tissues and cell types, and is evolutionary conserved in mammals and lower eukaryotic organisms [56].

FAK contains a FERM (band 4.1, ezrin, radixin, moesin homology) domain, a tyrosine-kinase domain and a focal adhesion targeting (FAT) domain. The crystal structure of the kinase domain has been determined and displays the typical protein kinase bilobal architecture; with the smaller N-terminal lobe containing a five-stranded antiparallel β-sheet and a single α-helix, and the larger C-terminal lobe being mostly α-helical [57] (Fig. 1.5).

The FAT domain is a four-helix bundle required for localization of FAK to focal adhesions via binding to paxillin [58;59]. The FERM domain is a three-lobed domain thought to mediate protein-protein interactions by binding to cytoplasmic domains of transmembrane receptors, such as the cytoplasmic region of β-integrin subunits [60-62].

The linker region connecting the FERM and the catalytic domain contains a proline-rich site which forms a binding-motif for Src family SH3 domains [63;64]. In the same linker the major autophosphorylation site in FAK, Y397, is situated. When phosphorylated, it forms a high-affinity binding site for the SH2 domains of Src family kinases, the p85 subunit of phosphatidylinositol 3-kinase (PI3K) and growth factor receptor-bound protein 7(Grb7) [65-69].

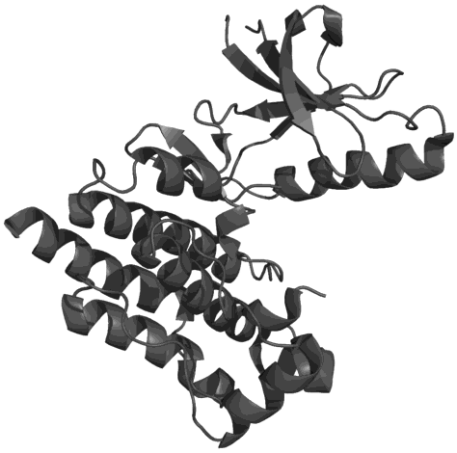

Figure 1.5. Crystal structure of the FAK kinase domain (PDB entry 1MP8 [70]).

Regulation of FAK and interaction with Src

Evidence is mounting that the FERM domain of FAK can interact with the catalytic domain, acting as an autoinhibitor of FAK activity [71-74]. A crystal structure of FAK including the FERM domain, linker region and catalytic domain shows FAK in an autoinhibited state [75]. In this structure the FERM domain binds the kinase domain, blocking access to the active site and to the kinase activation loop, as well as sequestering the Y397 phosphorylation site. This gives rise to a model of FAK activation where the FERM domain is displaced by competitive binding of an activating protein, such as the cytoplasmic regions of β-integrins or growth factor receptors. After FERM domain displacement Y397 is rapidly autophosphorylated and the PxxP sequence in the same linker region is exposed, enabling binding of the SH2 and SH3 domains of Src (Fig. 1.6). The interaction of Src and FAK leads to phosphorylation of other tyrosine residues in FAK and full activation of both proteins. The FAK-Src complex further phosphorylates various adaptor proteins, affecting a number of downstream signalling cascades [76].

Phosphatidylinositol 3-kinase (PI3K)

Phosphatidylinositol 3-kinases, also known as phosphoinositide 3-kinases, are a family of enzymes that phosphorylate inositol lipids at the 3' position of the inositol ring to generate the 3-phosphoinositides PI(3)P, PI(3,4) P2 and PI(3,4,5) P3 [77]. The resulting phosphoinositides act as second messengers in signal transduction cascades controlling cellular activities such as proliferation, differentiation, chemotaxis, survival, trafficking, and glucose homeostasis. PI3Ks therefore play a central role in many processes in the cell, and deregulated PI3K signalling is implicated in diseases such as cancer and diabetes [78].

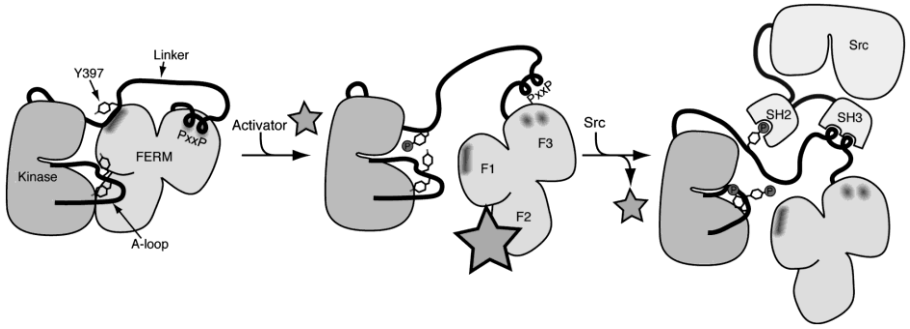

Figure 1.6. Model of activation of FAK and interaction with Src. In the inactive state the FERM domain blocks the access to the kinase domain active site, while sequestering the PxxP and Y397 regions in the FERM-kinase linker. In this model binding of a partner protein to the FERM domain is proposed to be the first step in FAK activation, freeing the kinase domain to autophosphorylate Y397 in cis or in trans. Src recruitment occurs via binding of the SH2 domain to the phosphorylated Y397 and binding of the SH3 domain to the proline-rich region. Phosphorylation of tyrosines Y576 and Y577 in the FAK activation loop by Src leads to full activation of FAK and prevents inhibition by the FERM domain. From Ref. [79] copyright (2007), with permission from Elsevier.

PI3Ks can be divided into different classes depending on their structure and substrate specificity. Class I_A PI3Ks are heterodimers consisting of a catalytic domain with a molecular weight of around 110 kDa (the p110 subunit) and an

adaptor/regulatory subunit known as the p85 subunit, which contains two SH2 domains and one SH3 domain [80]. The SH2 domains bind to phosphotyrosine residues generated by tyrosine kinases, allowing for translocation of PI3K to the membranes where its lipid substrates can be found.

Compared to the SH3 domain of Src, the PI3K p85 subunit SH3 domain contains a 15 aminoacid insertion, and the sequence identity of the two domains is only 21%. Despite these differences, the protein structures are remarkably similar [81;82]. Screening of a combinatorial peptide library for binding to the PI3K p85 SH3 domain lead to the identification of the consensus sequence RXLPPRP [83]. Like the Src SH3 domain the PI3K SH3 domain binds peptides in a left-handed type II polyproline helical conformation [84].

Methods used to study protein complexes

NMR chemical shift perturbation mapping

NMR is a powerful technique for mapping the binding site of a protein upon complex formation with another protein or a ligand [85;86]. In chemical shift perturbation mapping a two-dimensional NMR spectrum such as a [^{15}N,^{1}H]-HSQC spectrum is normally recorded of the free protein, which needs to be ^{15}N-labelled. In the spectrum each peak corresponds to an amide group in the protein, such as the protein backbone amides for all amino acid residues except prolines. N-H groups in sidechains of asparagine, glutamine, histidine and tryptophan residues may also give rise to crosspeaks. After addition of an unlabelled binding partner to the ^{15}N-labelled protein another NMR spectrum is recorded. Nuclei situated at the binding interface may experience a change in their chemical environment upon binding. The chemical shifts in both the nitrogen and proton dimensions are sensitive to this change and the position of the resonance in the spectrum will change. The average

chemical shift perturbation in the nitrogen and proton dimensions can be calculated for each residue using Eq. 1, where $\Delta\delta^{N}{}_{binding}$ and $\Delta\delta^{H}{}_{binding}$ are the chemical shift perturbations of the amide nitrogen and amide proton, respectively.

$$\Delta\delta_{avg} = \sqrt{\frac{(\Delta\delta^{N}_{binding}/5)^2 + (\Delta\delta^{H}_{binding})^2}{2}} \quad (1)$$

By mapping the chemical shift changes onto the protein structure information can be obtained about the binding interface. If the binding induces structural or conformational changes in the protein, chemical shift perturbations can also be seen for residues situated away from the binding site.

In order to determine the binding constant the unlabelled binding partner is titrated into the ^{15}N-labelled protein and a 2D NMR spectrum is recorded at each titration point. The chemical shift perturbations caused by binding can be followed if the chemical exchange rate is large compared to the chemical shift difference between the free and bound forms, measured in radians per second. From the chemical shift perturbations during the titration the binding constant can be determined. The use of deuterated ^{15}N-labelled protein together with TROSY experiments extends the limit of the method to protein complexes of a molecular weight above 100 kDa [87].

Paramagnetic relaxation enhancement NMR

Paramagnetic relaxation enhancement (PRE) NMR is a technique that can be used to determine the structure and dynamics of protein complexes. It is based on the fact that magnetic dipolar interactions between the spins of a nucleus and the unpaired electrons of a paramagnetic centre lead to an increase in the relaxation rate of the nuclear magnetization [88]. The PRE effect is proportional to r^{-6}, where

r is the distance between the unpaired electron and the nucleus. Some metalloproteins contain intrinsic paramagnetic centres, to other proteins a paramagnetic probe can be attached through site-specific labelling. The relaxation rates of the nuclei in the other protein are then measured, and an increase in the relaxation rate indicates that the nucleus has been in the vicinity of the probe. The larger the relaxation effect, the closer that part of the unlabelled protein came to the probe on the other protein. The paramagnetic contribution to the transverse relaxation rate, $R_{2,para}$, can be determined for the amide proton of each aminoacid residue in the protein. This can subsequently be converted into a distance between the paramagnetic centre and the amide proton using Eq. 2:

$$r = \sqrt[6]{\frac{\gamma^2 g^2 \beta^2 \tau_c}{20 R_{2,para}} \left(4 + \frac{3}{1 + \omega_h^2 \tau_c^2} \right)} \qquad (2)$$

where r is the distance between the paramagnetic centre and a given amide proton, τ_c is the correlation time of the dipolar interaction of the electron and the nucleus, ω_h is the proton Larmor frequency, γ is the proton gyromagnetic ratio, g is the electronic g-factor and β is the Bohr magneton. Because of the large magnetic moment of the unpaired electron the PRE effects are large and can provide long-range distance restraints of up to 35 Å [89]. The distance restraints can be used in docking calculations to determine the relative orientation of the macromolecules in the complex. The strong distance-dependence of the PRE enables the detection of protein complex orientations that are populated only a small fraction of the time. This has been exploited to study encounter complexes involved in protein-nucleic acid and protein-protein interactions [90-94].

Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a method that is suitable for studying both low-affinity and high-affinity macromolecular interactions. In an ITC experiment a macromolecule is placed in a sample cell, a ligand is injected in a programmed sequence of steps, and the tiny amounts of heat associated with the non-covalent interactions involved in binding are measured. From this, the affinity, the change in enthalpy and the stoichiometry of binding can be estimated and the change in entropy can be calculated, providing a complete thermodynamic characterization of the interaction [95;96].

Scope and outline of thesis

In order to learn more about how protein tyrosine kinases function, it is important not to focus on the kinase domain alone, but also on the interaction with other domains. The main topic of this thesis is the interaction of Src and FAK, mediated via the SH2 and SH3 domains of Src, and the goal is to outline the details of this interaction. To this end, a number of model systems of the FAK-Src interaction are studied, ranging from peptide-protein interaction studies to binding studies involving isolated protein domains. In addition, the interaction of a PI3K SH3 domain with a photocleavable peptide is investigated.

In chapter 2, the interaction of peptides derived from the SH3 domain binding site in FAK with the Src SH3 domain is studied, using paramagnetic relaxation enhancement nuclear magnetic resonance spectroscopy (PRE NMR) together with chemical shift perturbation analysis. In chapter 3, the binding of peptides from the SH2 domain binding site of FAK to the Src SH2 domain are studied using (PRE) NMR and isothermal titration calorimetry (ITC). Chapter 4 contains a study of peptides containing both SH2 domain- and SH3 domain-binding sites interacting

with a Src SH3-SH2 domain fragment, and the effect of decreasing the distance between the binding sites in the peptide is investigated using NMR and ITC. In chapter 5, the expression of the catalytic domain of FAK in insect cells is described, together with the purification and characterization of the protein. In chapter 6, an NMR binding study of the FAK catalytic domain with the Src SH2 domain is presented. In chapter 7, the construction of GFP-labelled SH3 or SH2 domain-containing phosphotyrosine reporter constructs is described, and the behaviour of the constructs in mammalian cells is characterized. In chapter 8, the interaction of the SH3 domain of the p85 subunit of PI3K with a photocleavable peptide is studied, investigating what effect modifying the peptide has on the interaction. Finally, chapter 9 contains a general discussion of the results presented in the previous chapters.