2006

Exploring the Biophysical Properties of Domains Within Multi-Domain Proteins

Vasant Muralidharan

EXPLORING THE BIOPHYSICAL PROPERTIES OF DOMAINS WITHIN MULTI-DOMAIN PROTEINS

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Vasant Muralidharan

June 2006

EXPLORING THE BIOPHYSICAL PROPERTIES OF DOMAINS WITHIN MULTI-DOMAIN PROTEINS

Vasant Muralidharan, Ph.D.

The Rockefeller University 2006

In the past few decades, considerable progress has been made in understanding the biophysical properties of proteins using small modular domains such as SH3 domains. However, there is a surprising lack of knowledge regarding how these properties are affected when the domain is placed back within its full-length multi-domain protein. Using a combination of expressed protein ligation (EPL) and *in vivo* amino acid replacement of tryptophans with tryptophan (Trp) analogues, we have developed an integrated approach that allows the domain-specific incorporation of optical probes into large recombinant proteins. The Src homology 3 (SH3) domain from the c-Crk-I adaptor protein has been labeled with a Trp analogue, 7-azatryptophan (7AW), using *E.coli* Trp auxotrophs. Biophysical analysis shows that incorporation of 7AW does not significantly perturb the structure or function of the isolated domain. Ligation of 7AW labeled SH3 domain to the c-Crk-I Src homology 2 (SH2) domain, via EPL, generated the multi-domain protein, c-Crk-I, with a domain specific label. Studies on this labeled protein show that the biochemical and thermodynamic properties of the SH3 domain do not change within the context of a larger multi-domain protein. We have also utilized this technique and segmental isotopic labeling to study the C-

terminal SH3 (cSH3) domain of the signaling adaptor protein, Crk-II. Several studies suggest that the cSH3 domain plays an important regulatory role in the protein. However, no structural information is available on this domain and relatively little is known about its binding partners. We have solved the solution NMR structure of the C-terminal SH3 domain and it adopts the standard SH3 fold comprising a five-stranded β-barrel. Thermodynamic and kinetic studies show that the domain folds in a reversible two-state manner and that the stability of the fold is similar to that observed for other SH3 domains. Studies on the cSH3 domain specifically labeled within Crk-II have provided the strongest evidence yet that the cSH3 domain interacts in an intramolecular fashion to regulate Crk-II. The techniques developed here should be applicable to many other multi-domain proteins. This sets the stage for a better understanding of the biophysical properties of domains within these complex systems.

Acknowledgements

I would like to thank Professor Tom Muir and Adjunct Professor Dan Raleigh for their extraordinary scientific mentorship, enthusiasm, optimism and support during the course of this work.

I am greatly indebted to all members of the Muir and Raleigh laboratories for their technical assistance, advice and for the exceptional work/play environment, particularly, Mr. Jaehyun Cho, Dr. Miquel Vila-Perello, Mr. Lukasz Kowalik, Dr. Michelle Trester-Zedlitz, Dr. Izabela Giriat, Dr. Mande Holford, Dr. Morgan Huse, Dr. Alessandra Romanelli, Dr. Roseanne Hofmann, Dr. Jen Ottesen, Mr. Mike Hahn, Mr. Ed Schwartz, Ms. Amy Tyszkiewicz, Mr. Rob Flavell, Dr. Baldiserra Giovani, Dr. Matt Pratt, Dr. Champak Chatterjee and Dr. Jean-Phillippe Pellois.

I gratefully acknowledge Dr. Kaushik Dutta for his intellectual and practical contributions to the latter half of this work and for teaching me protein NMR spectroscopy.

I would like to express my appreciation to Dr. Mike Goger and Dr. David Cowburn (President & CEO) for providing research facilities at the New York Structural Biology Center and for all their contributions to the latter half of this work.

Support from the Rockefeller University graduate program, under the direction of Dr. Sid Strickland and from the Albert Cass Travel Fellowship is gratefully acknowledged.

I would like to thank the members of my thesis committee, Professor Brian Chait and Professor Sandy Simon, for their constant support and guidance and Professor Stephen Kent for serving as my external examiner.

I would especially like to thank my fiancèe Ms. Nupur Kittur for her ongoing support and encouragement during difficult times and her exceptional culinary skills, which have never failed to lift my spirits.

I wish to thank my family, especially my brother, Dr. Shishir Karthik Muralidharan, my father, Mr. Muralidharan Balasubramanian and my mother, Mrs. Prema Muralidharan for their constant support and encouragement.

Table of Contents

List of Figures

List of Tables

Chapter 1– Introduction

Cell survival within an environment is governed by the ability to respond to various extracellular signals. Essentially, all cellular functions are modulated via networks of signaling proteins that transmit signals from the plasma membrane to the cytoplasm and the nucleus (Dueber et al., 2004; Pawson and Nash, 2003). The primary mechanism underlying signal transduction is the formation of transient protein-protein interactions. These protein-protein interactions can result in certain posttranslational modifications, targeting of the protein to specific subcellular location, release of chemical second messengers and modulation of protein activity. Over the last few decades, a large number of cellular proteins have been identified in eukaryotic cells that are involved in the process of signal transduction. However, if each of these proteins had a unique biological role, then there are simple not enough proteins to account for the sophisticated behavior of diverse cell types. This suggests that the signaling proteins must act in a combinatorial fashion to achieve such an exquisite variety of responses. Nature achieves this combinatorial expansion by utilizing a small number of modular protein domains that are joined together in various configurations to form larger signaling proteins (Schultz et al., 1998). Each domain has a characteristic functional property and consequently, the larger multi-domain protein has the potential to receive various signaling inputs and transmit several outputs (Lim, 2002). Signal transduction networks can then be built

up from various interacting pathways (Dueber et al., 2004; Pawson and Nash, 2003).

Modular Signaling Domains

There have been several (~80) modular signaling domains identified in eukaryotes (Schultz et al., 1998) that play an important role in signal transduction via protein-protein or protein-small molecule interactions (Pawson and Nash, 2003) (Table 1.1). Signaling proteins often contain many such modular domains covalently linked in various combinations to form larger multi-domain proteins. The linking of modular domains in this fashion can serve several purposes. A larger interaction surface and consequently, greater binding affinity, can be achieved if two domains recognize different sites in the same target (Pawson and Nash, 2000). On the other hand, the domains can interact with different targets thereby modulating the formation of large signaling complexes with several proteins (Dueber et al., 2004). This has been observed for signaling adaptor proteins such as Grb2, Nck and Crk, which recruit effector molecules to tyrosine kinases (Pawson and Scott, 1997). Modular domains can also interact in an intramolecular fashion to regulate the structure and function of a multi-domain protein, such as the Src tyrosine kinase (Lim, 2002).

Among the first modular signaling protein domains to be identified were the Src Homology 2 (SH2) and Src Homology 3 (SH3) domains (Kuriyan

and Cowburn, 1997). The SH2 domain is comprised of ~100 residues and binds phosphotyrosine residues in the context of specific sequences (Table1.1). The SH2 domain-peptide interaction has been well characterized and this interaction has moderate affinity ($K_d \approx$ 0.1-1.0 μM) (Kuriyan and Cowburn, 1997). SH3 domains, on the other hand, are small (~60 residues) protein-protein interaction domains that recognize proline-rich peptide motifs. They have been found within ~1000 proteins in various

Table 1.1 – Examples of modular signaling domains and their binding partners.

Modular Domain	Recognition Motif/Small Molecule[a]
SH3	P-X-X-P
WW	P-P-X-Y
EVH1	F-P-X-ϕ-P
SH2	pY-X-X-ϕ
PDZ	E-S/T-D-V-COOH
PTB	ϕ-X-N-P-X-pY
14-3-3	R-S-X-pS-X-P
FHA	pT-X-X-X
PH	Polyphosphoinositides
FYVE	Phosphatidylinositol-3-Phosphate
PX	Phosphatidylinositol-3-Phosphate
C1	Diacylglycerol
C2	Ca^{+2} ion and Acidic Phospholipids

[a]X= any amino acid, p= phosphorylated amino acid, ϕ= hydrophobic amino acid.

organisms ranging from yeast to humans (Bateman et al., 2004). The sequence identity between different SH3 domains is low, ranging from 15% to 45%, with the average pairwise identity being around 27% (Larson and Davidson, 2000) (Figure 1.1). Despite this, all SH3 domains fold in a similar manner to adopt a five-stranded β-barrel structure (currently 380 structures in PDB) (Figure 1.2). This incongruity (i.e. low sequence homology *vs.* high structural homology) has made SH3 domains an important model system for protein stability, dynamics and folding studies.

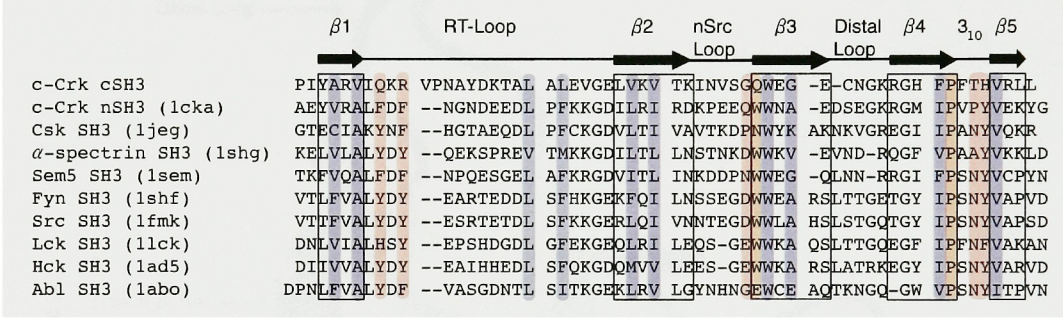

Figure 1.1. Primary sequence and secondary structure alignment of SH3 domains. Shaded in red are conserved residues involved in ligand binding and in blue are conserved residues of the hydrophobic core.

Indeed, the folding properties of several SH3 domains has been well characterized (Filimonov et al., 1999; Grantcharova and Baker, 1997; Guijarro et al., 1998; Lim, 1994; Plaxco et al., 1998; Riddle et al., 1999; Riddle et al., 1997; Viguera et al., 1994; Zhang et al., 1994). SH3 domains fold in a rapid two-state manner with no detectable intermediates (Grantcharova and Baker, 1997; Plaxco et al., 1998; Viguera et al., 1994).

The folding transition state of SH3 domains consists of a relatively small number of conformers in which the distal loop and the type-II β-turn are ordered and interact with each other (Grantcharova et al., 2000) (Figure 1.2). The transition state conformers also have a relatively well-formed hydrophobic core (Northey et al., 2002). Folding studies on SH3 domains have greatly extended our understanding of the forces governing the folding of a polypeptide chain into a three dimensional structure.

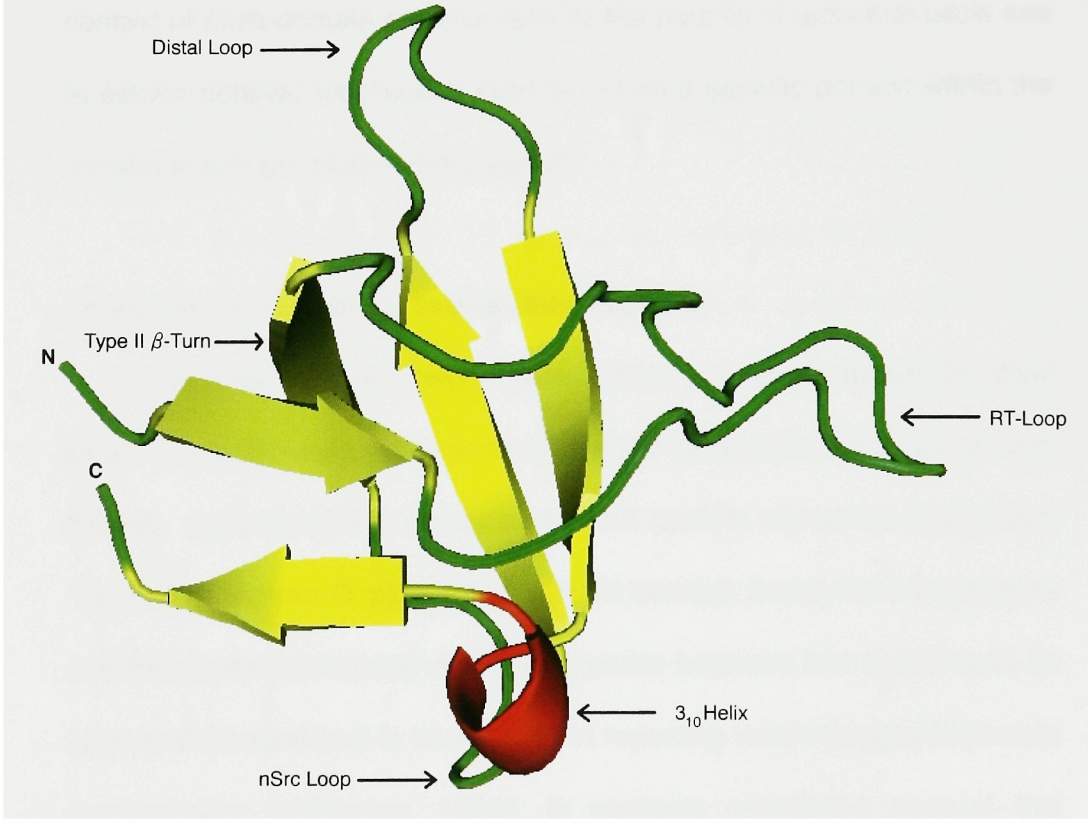

Figure 1.2. Structure of an SH3 domain. The structure of the middle SH3 domain from a signaling adaptor protein, Crk-II is shown (Wu et al., 1995) (1CKA). All SH3 domains adopt a similar five-stranded β-barrel fold.

Although considerable progress has been made in understanding the structural, functional and folding properties of proteins using small modular domains such as SH3 domains, there is a surprising lack of knowledge regarding how these properties are affected when the domain is placed back within its full-length protein. Since modular domains are always found in the context of multi-domain proteins, understanding structural and functional interplay between individual domains in this context is important. The lack of detailed biophysical analyses of domains within the context of multi-domain proteins reflects the paucity of tools that allow one to extract detailed biophysical information on a specific domain within the context of a large multi-domain protein.

Incorporation of Optical Probes Into Proteins

Traditionally, fluorescence spectroscopy has been utilized to great effect to study various aspects of protein biophysics such as protein folding, protein-protein interactions and protein dynamics (Lakowicz, 1999). Tryptophan (Trp) is the dominant intrinsic fluorophore in proteins and has been particularly useful as a probe because it is sensitive to its local environment and is thus useful in reporting local changes in protein conformation (Lakowicz, 1999). In proteins containing several Trp residues, as is often the case in multi-domain proteins, it is difficult to isolate the spectral contributions of individual fluorophores. Site-directed

mutagenesis is one possible solution to this problem, removing all but one of the Trp residues (Waldman, 1987). However, Trp residues are often part of the hydrophobic core of the protein and so mutating them runs the risk of destabilizing the protein and/or affecting its function. Extrinsic fluorescent probes like dansyl or related amino acid derivatives can also be incorporated into proteins, via modification of reactive residues like lysine and cysteine or via chemical synthesis (Cohen et al., 2002; Mannuzzu, 1996). This again carries the risk of destabilizing the protein and/or affecting its function. Moreover, external probes are usually not well suited to study processes such as protein folding, owing to their bulky nature.

Ideally one would want to replace Trp residues in proteins with molecules that cause minimal perturbation in protein structure and have a unique spectroscopic signature, as is the case with Trp analogues. The use of *Escherichia coli* Trp auxotrophs to incorporate Trp analogues into proteins in place of Trp residues is thus an attractive alternative to site-directed mutagenesis (Schlesinger, 1968). This approach has the inherent advantage that Trp analogues are structurally similar to tryptophan and in many instances have minimal effect on protein structure and function (Ross et al., 1997; Ross et al., 2000; Twine and Szabo, 2003). Indeed, different Trp analogues, such as 4-fluorotryptophan, 5-fluorotryptophan, 6-fluorotryptophan, 7-azatryptophan (7AW) and 5-hydroxytryptophan (5-

OHW) have been incorporated into several proteins (Ross et al., 2000) (Figure 1.3). The fluorotryptophan analogues are primarily utilized to incorporate ^{19}F into proteins for NMR studies (Danielson and Falke, 1996). The spectral properties of 5-OHW and 7AW are different from Trp, however, 5-OHW is insensitive to the environment (Wong and Eftink, 1998a).

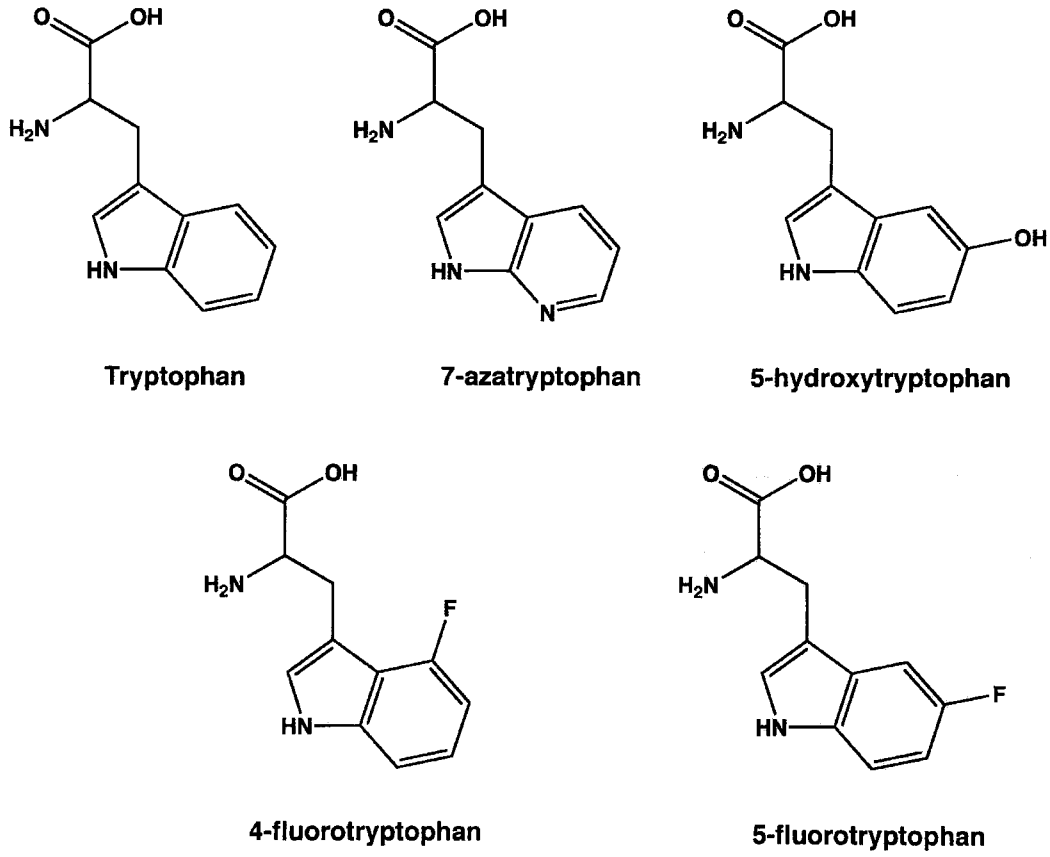

Figure 1.3. Tryptophan and its analogues. The most commonly used tryptophan analogues that have been incorporated into proteins using *E.coli* trp auxotrophs.

Out of the several Trp analogues, 7AW is ideally suited as a spectroscopic probe. The absorption spectrum of 7AW is red-shifted by 10-15 nm relative to the absorption spectra of Trp and the emission spectrum is red-shifted by up to 50 nm. Thus the 7AW labeled protein can be selectively excited and observed in the presence of Trp residues (Ross et al., 1997). 7AW is highly sensitive to its environment (Ross et al., 1997; Ross et al., 2000), and can be used as a probe to study protein folding (Wong and Eftink, 1998a; Wong and Eftink, 1998b). The use of *E.coli* Trp auxotrophs to label proteins results in the global replacement of all Trp residues in the protein with Trp analogues. This would be a disadvantage in proteins where one wants to study a specific Trp residue among many, like in the case of multi-domain proteins. One solution to the problem is to utilize nonsense suppression mutagenesis approaches (Steward, 1997). In principle, these limitations can also be overcome using expressed protein ligation (EPL) (Muir, 2003; Muir et al., 1998).

Expressed Protein Ligation

EPL is a protein semi-synthesis approach that allows synthetic and/or recombinant polypeptides to be chemoselectively and regioselectively joined together via a peptide bond (Muir, 2003). The approach takes advantage of two processes; native chemical ligation (Dawson et al., 1994) and protein splicing (Noren et al., 2000).

Over the last ~15 years, chemical ligation has emerged as a powerful technique in chemical biology, allowing unprotected molecules to be selectively linked together in an aqueous environment. Chemical ligation was initially introduced to solve the decades old problem of how to chemically synthesize proteins from smaller, synthetically accessible, peptide building blocks (Gaertner et al., 1992; Schnolzer and Kent, 1992). However, the basic strategy has found widespread use in other areas, including carbohydrate synthesis (Hang and Bertozzi, 2001) and nucleic acid synthesis (Burbulis et al., 2005; Dose and Seitz, 2005; Ficht et al., 2004; Lovrinovic et al., 2003). All chemical ligation strategies rely on the incorporation of "molecular velcro" at the appropriate locations within the reactive partners; that is, mutually reactive groups that are unreactive towards everything else in the molecules under the conditions used. Several chemoselective reactions are in common use in chemical biology, including oxime formation (Gaertner et al., 1992; Rose, 1994), the Staudinger ligation (Nilsson et al., 2000; Saxon et al., 2000; Saxon and Bertozzi, 2000) and Huisgen cycloaddition chemistry (Kolb and Sharpless, 2003).

By far the most commonly used peptide ligation approach is native chemical ligation (NCL) (Dawson et al., 1994) (Figure 1.4). In this reaction, two fully unprotected synthetic peptide fragments are reacted under neutral aqueous conditions with the formation of a normal (native) peptide

bond at the ligation site. The first step in NCL involves the chemoselective reaction between a peptide fragment containing an N-terminal cysteine residue (α-Cys) and a second peptide fragment containing a α-thioester group. The initial transthioesterification reaction is followed by a spontaneous intramolecular S→N acyl shift to generate an amide bond at the ligation junction. NCL is compatible with all naturally occurring functional groups found in proteins, including the sulfhydryl group of cysteine. Indeed, it is worth emphasizing that internal cysteine residues are permitted in both peptide segments due to the reversible nature of the initial transthioesterification step. NCL is a remarkably robust reaction that can be performed in the presence of chemical denaturants (Dawson et al., 1994; Hackeng et al., 1999; Hackeng et al., 1997), detergents (Bianchi et al., 1999), lipids (Clayton et al., 2004; Hunter and Kochendoerfer, 2004), reducing agents (Dawson et al., 1994; Hackeng et al., 1999; Hackeng et al., 1997), organic solvents (Kochendoerfer et al., 1999) and even crude cell extracts (Yeo et al., 2003). The approach has undergone a number of important refinements since it was introduced. These include; the development of auxiliary groups that in favorable cases overcome the requirement for an α-Cys (Canne et al., 1996; Offer et al., 2002), the identification of catalytic thiol cofactors (Dawson et al., 1997), and the development of solution-phase and solid-phase sequential ligation strategies that allow several peptides to be linked together in series (Bang

and Kent, 2004; Camarero et al., 1998; Cotton et al., 1999). In addition, new SPPS approaches for the preparation of peptide α-thioesters (Alsina et al., 1999; Botti et al., 2004; Camarero et al., 2004; Ingenito et al., 1999; Ohta et al., 2006; Shin et al., 1999; Swinnen and Hilvert, 2000), promise to make this class of peptide more accessible to the non-specialist.

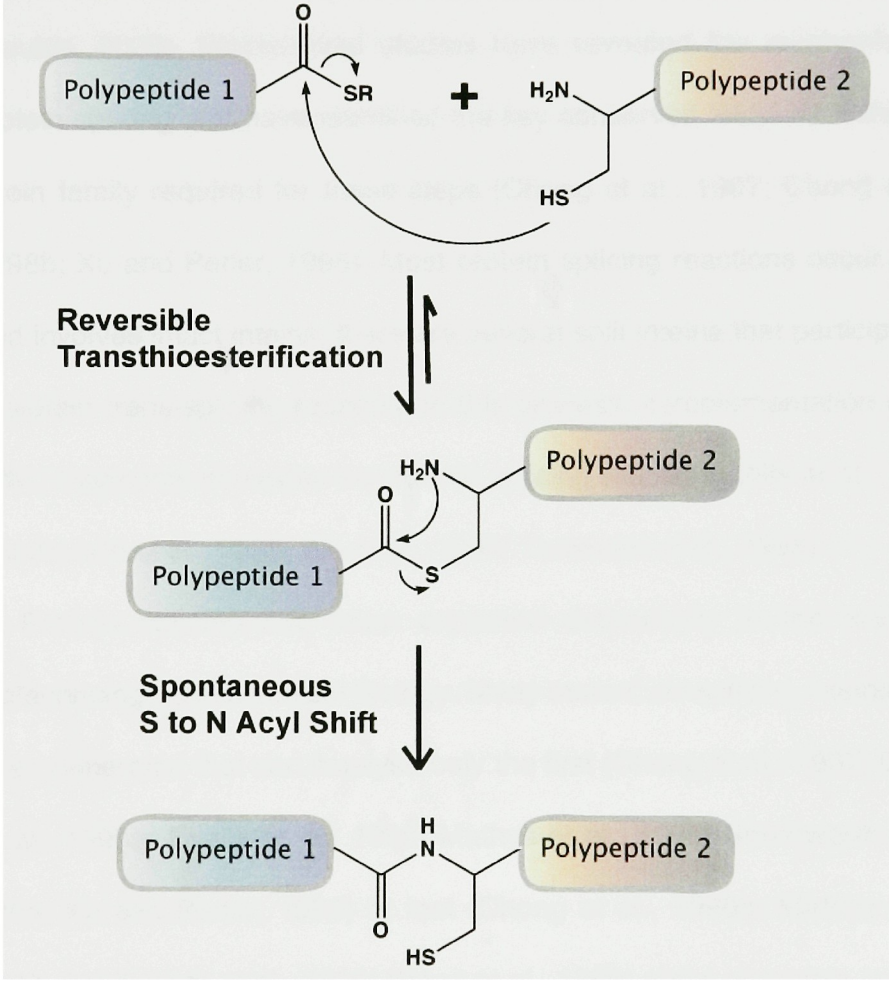

Figure 1.4. Principle of Native Chemical Ligation (NCL). The two reacting peptides, one with α-Cys and one with α-thioester, are unprotected. The reaction occurs under aqueous conditions around neutral pH.

Protein splicing is a post-translational process in which a precursor protein undergoes a series of self-catalyzed intramolecular rearrangements that result in the removal of an internal protein domain, termed an intein, and the ligation of the two flanking polypeptides, referred to as the N- and C-exteins (Figure 1.5). Inteins are autocatalytic and remarkably promiscuous for the sequences of the two flanking sequences (Paulus, 2000). Biochemical studies have revealed the mechanism of protein splicing and have identified the key conserved residues within the intein family required for these steps (Chong et al., 1997; Chong et al., 1998b; Xu and Perler, 1996). Most protein splicing reactions occur in *cis* and involves intact inteins; there are several split inteins that participate in a protein *trans*-splicing reaction. In this process, complementation of the intein fragments precedes the normal splicing reaction (Mills et al., 1998; Southworth et al., 1998; Wu et al., 1998; Yamazaki et al., 1998).

Protein splicing has been exploited extensively in the areas of biotechnology and chemical biology. Many mutant *cis*-splicing inteins have been generated that can catalyse only the first (Chong et al., 1997; Chong et al., 1998a; Evans et al., 1999; Mathys et al., 1999; Southworth et al., 1999; Xu and Perler, 1996) or last (Chong et al., 1998b; Mathys et al., 1999; Southworth et al., 1999; Wood et al., 1999) steps of protein splicing. In each case, a protein of interest is expressed as an in-frame N- or C-terminal fusion to the mutant intein, which is usually linked to an affinity

tag (Chong et al., 1997). Following purification, the protein of interest can be cleaved off the intein by addition of thiols (C-terminal cleavage, i.e. the protein corresponds to the N-extein) or by changing the pH and temperature of the solution (N-terminal cleavage, i.e. the protein corresponds to the C-extein). These intein-based expression systems can be used for the "traceless" purification of recombinant proteins (Chong et al., 1997). Importantly, intein-fusions provide a convenient route to recombinant proteins with an N-terminal cysteine or a C-terminal thioester.

In EPL, an N-terminal cysteine (Cys) containing protein or peptide is specifically ligated to the C-terminus of a recombinant α-thioester protein, generated using an intein fusion (Figure 1.6). Importantly, EPL has no inherent size limitation, in terms of the target protein, and the final yields are comparable to recombinant protein expression (Muir, 2003). Indeed, EPL has been used to selectively label specific domains within large multi-domain proteins with stable isotope probes for NMR (Camarero et al., 2002; Ottesen et al., 2003; Xu et al., 1999). In this strategy, a uniformly labeled recombinant protein fragment is ligated to an unlabeled protein fragment. The end result is a 'block' labeled protein sample that gives rise to heteronuclear NMR spectra containing a subset of the resonances that one would have observed from the corresponding uniformly labeled sample. In other words, there is significant reduction in spectral complexity.

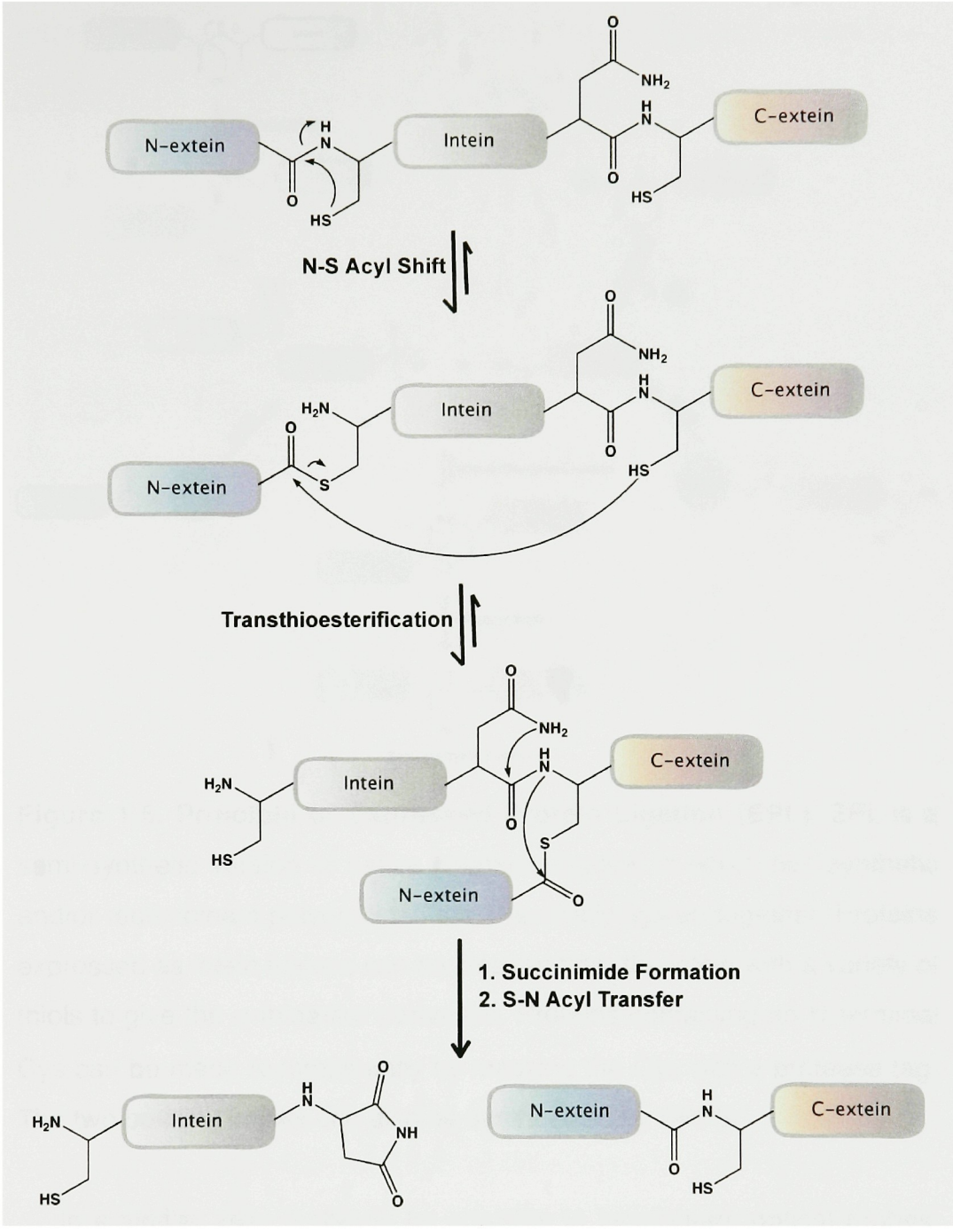

Figure 1.5. Mechanism of Protein Splicing. The central intein protein domain catalyses a series of intramolecular acyl transfers. These intramolecular rearrangements result in the joining of the two flanking proteins (termed, N-extein and C-extein) via a peptide bond.

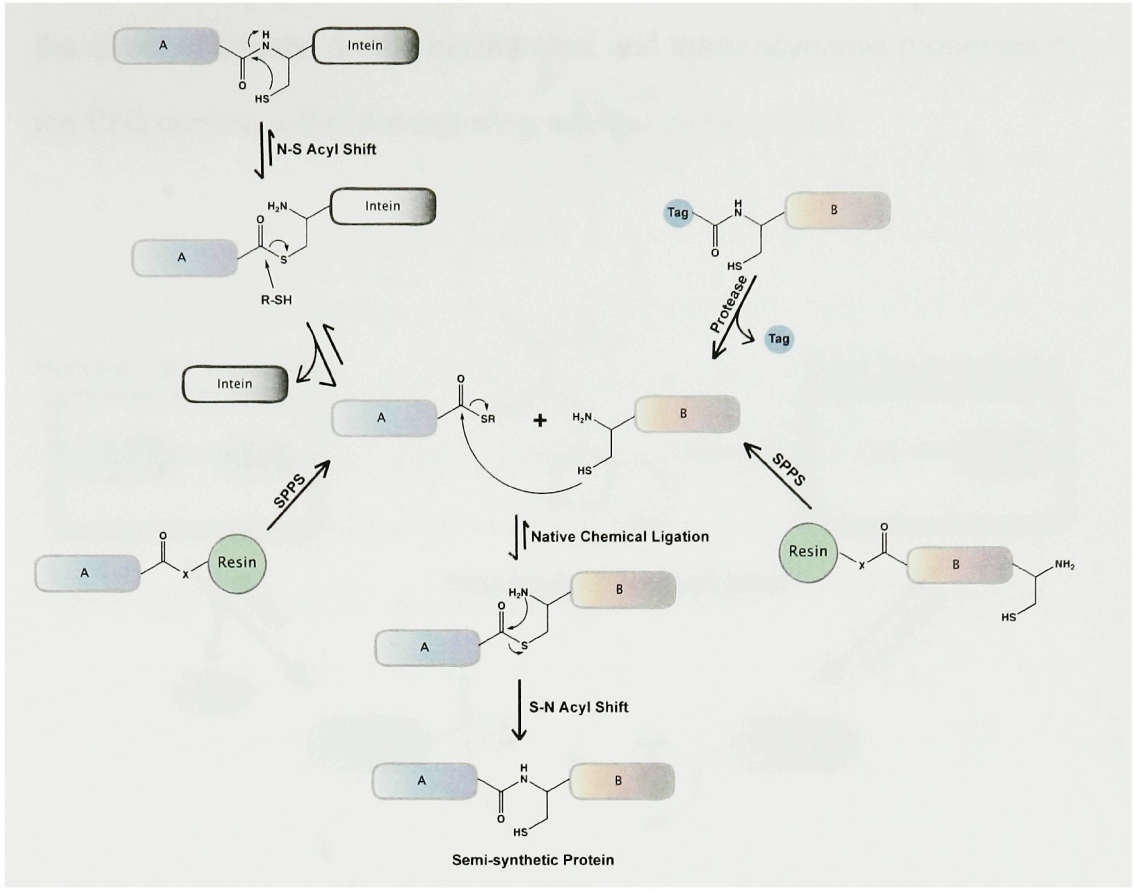

Figure 1.6. Principle of Expressed Protein Ligation (EPL). EPL is a semi-synthetic version of native chemical ligation in which two synthetic and/or recombinant polypeptides are chemically ligated together. Proteins expressed as intein fusions can be cleaved from the intein with a variety of thiols to give the α-thioester derivative. Proteins containing an N-terminal Cys can be made recombinantly by masking the Cys with a protease tag. The two building blocks can also be synthesized by SPPS.

In a similar vein, it should be possible to incorporate optical probes, like 7AW, into large proteins in a domain specific manner to reduce spectroscopic complexity in a sample with multiple fluorophores (Figure 1.7). In this study we have developed and applied this strategy to study

the effect of context on the biochemical and thermodynamic properties of the SH3 domain within the signaling adaptor protein, Crk-I.

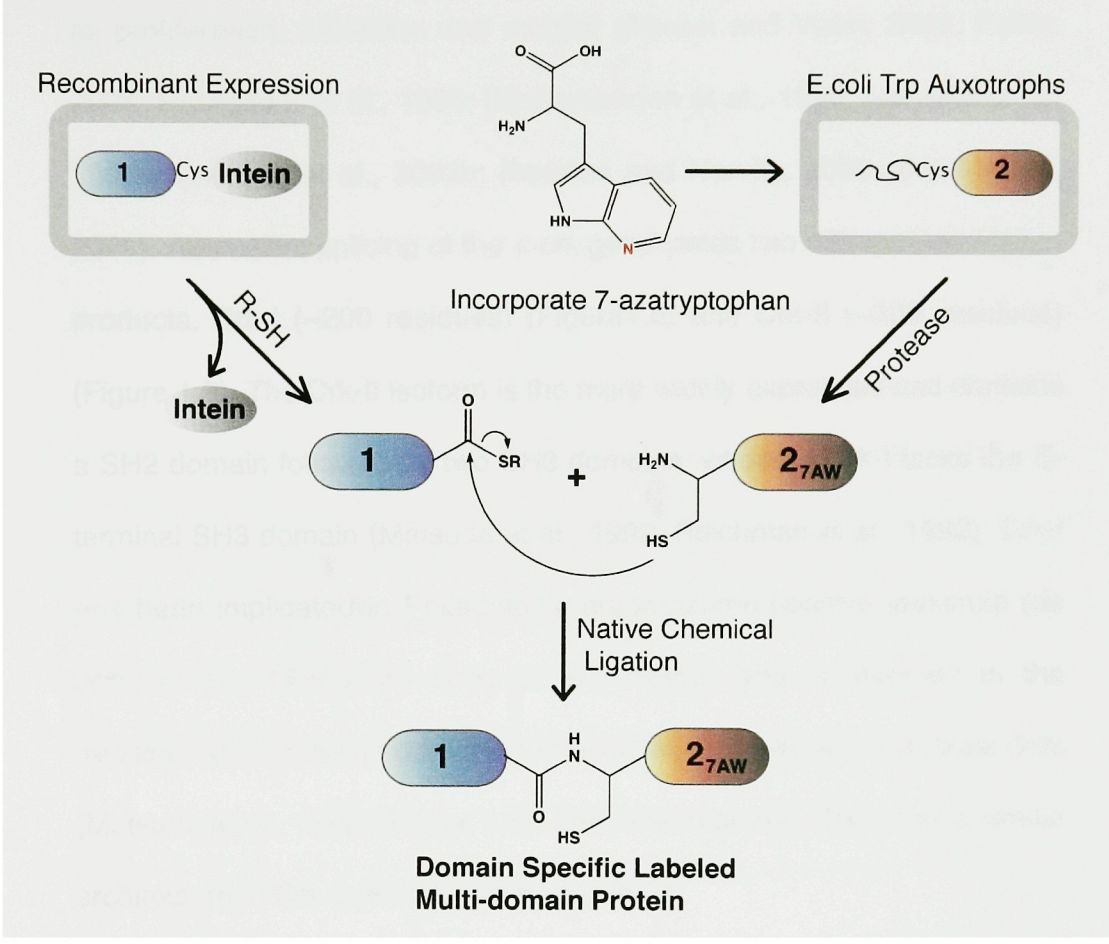

Figure 1.7. Domain specific labeling of multi-domain proteins. Schematic representation of the domain specific incorporation of 7-azatryptophan. Cys = Cysteine, 7AW = 7-azatryptophan, 1 = Domain 1, 2 =Domain 2, 2_{7AW} = 7AW labeled Domain 2.

Crk Family of Signaling Adaptor Proteins

The Crk family of signaling adaptor proteins are SH2 and SH3 domain containing proteins that have been implicated in cellular processes such as proliferation, adhesion and motility (Abassi and Vuori, 2002; Feller, 2001; Hasegawa et al., 1996; KizakaKondoh et al., 1996; Lamorte et al., 2002a; Lamorte et al., 2002b; Reddien and Horvitz, 2000; Zvara et al., 2001). Alternative splicing of the *c-crk* gene yields two different translation products, Crk-I (~200 residues) (Figure1.8) and Crk-II (~300 residues) (Figure 1.9). The Crk-II isoform is the more widely expressed and contains a SH2 domain followed by two SH3 domains, whereas Crk-I lacks the C-terminal SH3 domain (Matsuda et al., 1992; Reichman et al., 1992). Crk-I has been implicated in Philadelphia chromosome-positive leukemia (de Jong et al., 1995a; de Jong et al., 1995b) and is involved in the transformation of hematopoietic (Senechal et al., 1998) and fibroblast cells (Matsuda et al., 1992) (Figure 1.8). The related protein Crk-L has a similar architecture to Crk-II (ten Hoeve et al., 1993).

The primary role of Crk-II as a scaffolding protein is to bring together signaling proteins that bind to the SH2 and SH3 modular interaction domains. Accordingly, the SH2 domain recognizes pY-X-X-P motifs and mediates interactions with several proteins such as $p130^{cas}$, Cbl kinase and paxillin (de Jong et al., 1995a; Feller, 2001; Nojima et al., 1995; Schaller and Parsons, 1995) (Figure 1.9). In addition, the linker region

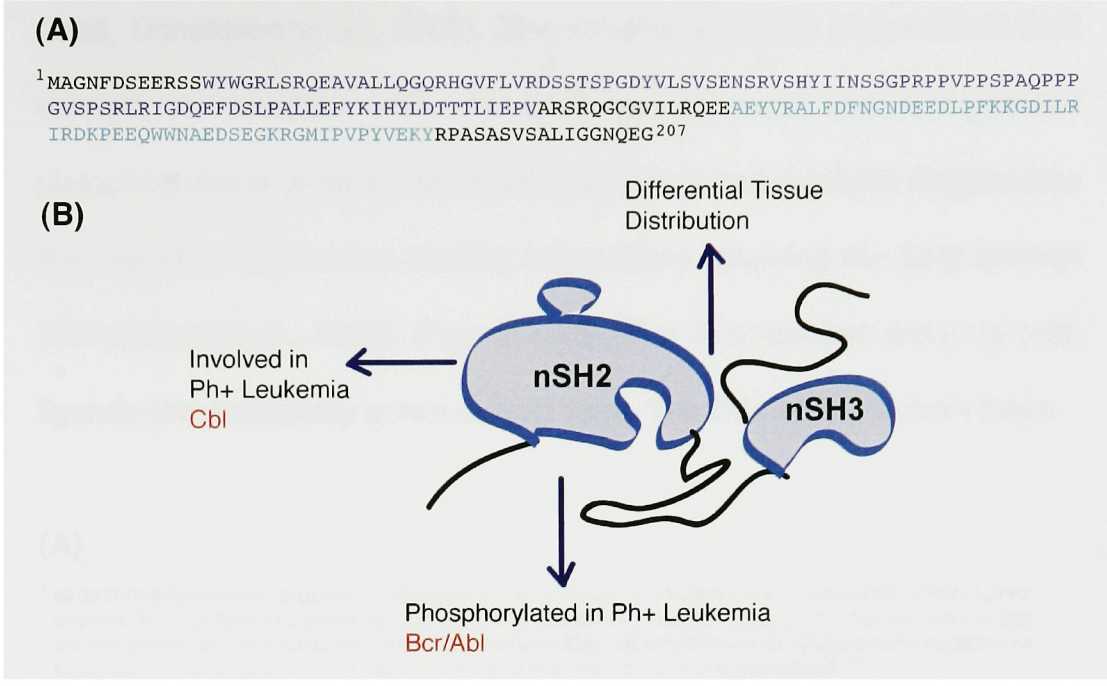

Figure 1.8. Domain architecture of the signaling adaptor protein, Crk-I. **(A)** The primary sequence of the Crk-I protein. The SH2 domain is highlighted in blue and the SH3 domain in cyan. **(B)** The proteins that are known to interact with the SH2 domain are shown in red. Crk-I is known to have a differential distribution in various tissue types in mice, though the functional significance of this distribution is unknown.

between the two SH3 domains of Crk-II contains the motif Y^{221}-A-Q-P, which can be phosphorylated by the protein tyrosine kinase, c-Abl. This phosphorylation creates an intramolecular docking site for the Crk-II SH2 domain (Feller et al., 1994). This triggers a structural change in the protein that is thought to regulate its function (Rosen et al., 1995). The Crk-II SH2 domain is unique in that it contains a proline-rich insertion within a surface loop that serves as a ligand for the SH3 domain of c-Abl (Anafi et al.,

1996; Donaldson et al., 2002). The solution structure of the Crk-II SH2 domain in complex with Abl-SH3 domain and the Crk-derived phosphotyrosine peptide has been determined and provides insights into the regulation of the two binding interactions involving the SH2 domain (Donaldson et al., 2002) (Figure 1.10). The SH2 domain binds to both ligands simultaneously at two distinct sites. The Crk-II SH2 domain has a

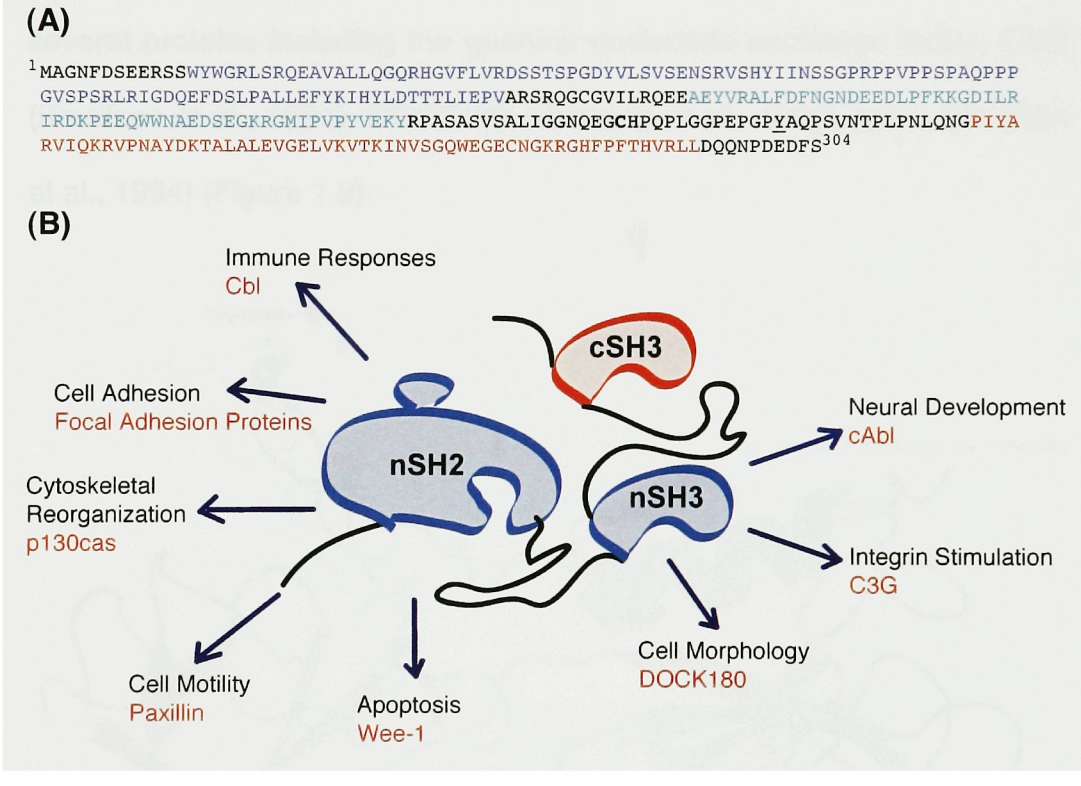

Figure 1.9. Domain architecture of the signaling adaptor protein, Crk-II. (A) The primary sequence of the Crk-II protein. The SH2 domain is highlighted in blue, the nSH3 domain in cyan and the cSH3 domain in red. ^{221}Y is underlined. **(B)** The proteins interacting with the nSH2 and nSH3 domains is shown along with the signaling events these interactions facilitate. Note that there are no functional protein-protein interactions attributed to the cSH3 domain.

unique poly-proilne loop, not present in other SH2 domains that is recognized by the Abl SH3. The structure shows how the SH2 domain of Crk can bring together distinct proteins via its two unique binding sites (Figure 1.10). The central SH3 domain of Crk-II (hereafter referred to as nSH3) has also been well characterized at the structural (Wu et al., 1995) and biochemical (Camarero et al., 2001) levels. The nSH3 domain recognizes P-X-X-P-X-K sequences and has been shown to interact with several proteins including the guanine nucleotide exchange factor, C3G (Knudsen et al., 1994), DOCK180 (Matsuda et al., 1996) and c-Abl (Ren et al., 1994) (Figure 1.9).

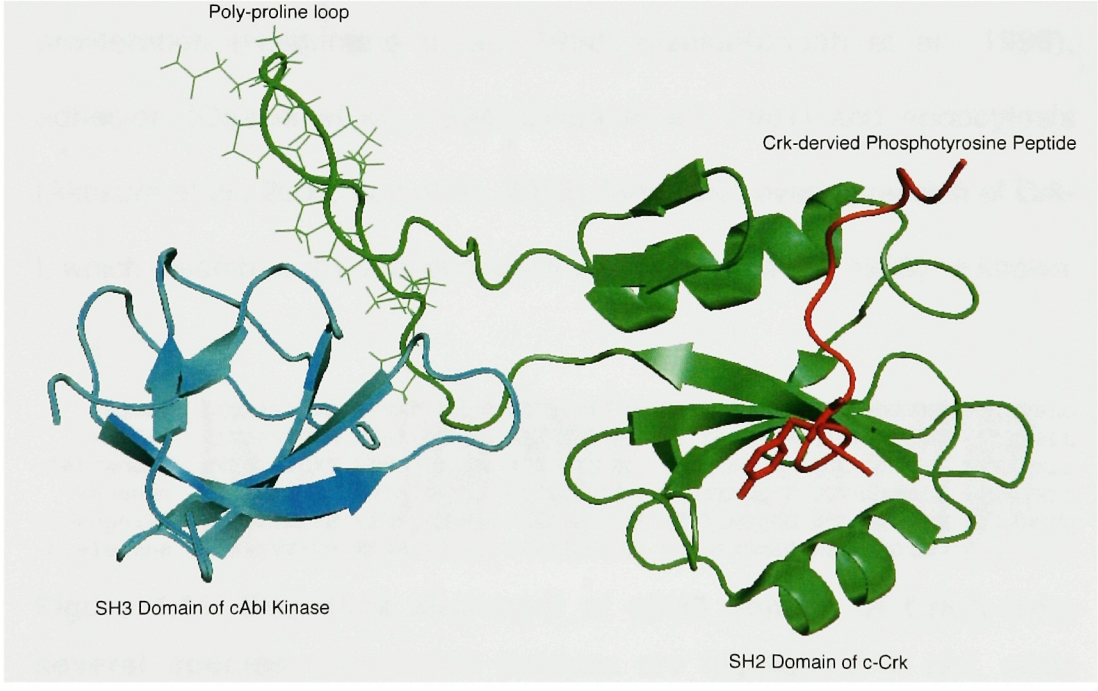

Figure 1.10. Structure of the SH2 domain of Crk (in green) in complex with SH3 domain of cAbl (in cyan) and Crk-derived phosphotyrosine peptide (in red).

However, the biophysical properties of the nSH3 domain are poorly understood within the context of the full-length Crk protein. Therefore, one goal of this study is to gain a detailed understanding of how the folding and functional properties of the nSH3 domain are affected by the context of the multi-domain protein, Crk-I.

Compared to the SH2 and nSH3 domains, the structure and function of the putative cSH3 domain of Crk-II is poorly understood. The sequence of this region of the protein is highly conserved in organisms ranging from fruitfly to man, suggesting that the domain plays an important functional role (Figure 1.11). Indeed, point mutations or deletions in the cSH3 domain are known to perturb a number of cellular processes, including proliferation (Hashimoto et al., 1998; KizakaKondoh et al., 1996), adhesion (Ogawa et al., 1994; Zvara et al., 2001) and endocytosis (Akakura et al., 2005; Sun et al., 2005). Moreover, overexpression of Crk-I, which resembles the viral oncogene v-Crk (Mayer et al., 1988), is known

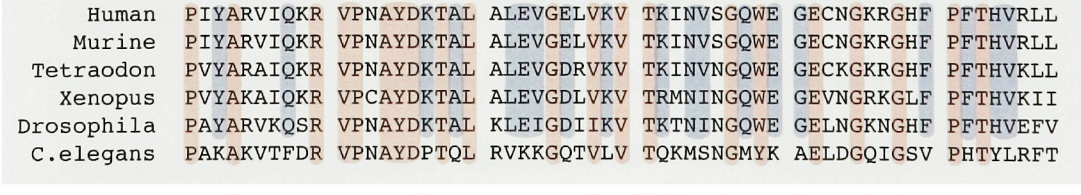

Figure 1.11. Sequence alignment of cSH3 domain of Crk-II from several species. Conserved residues are highlighted in red, while residues conserved in all organisms, except *C.elegans* are in blue.

to transform fibroblasts in culture (Matsuda et al., 1992). These studies have led to the hypothesis that the cSH3 domain negatively regulates the function of Crk-II (Feller, 2001; Ogawa et al., 1994). Further advancement of this model has been to a large extent frustrated by the inability to find classical P-X-X-P containing protein ligands for the cSH3 domain (Feller, 2001).

Although sequence alignment analyses strongly predict that the cSH3 adopts the canonical SH3 fold (Larson and Davidson, 2000), many of the conserved aromatic residues that constitute the ligand-binding surface in the SH3 domain family are mutated to polar, typically non-aromatic residues in the cSH3 domain (Figure 1.1). This raises the possibility that the cSH3 domain does not bind typical P-X-X-P ligands or that it binds them with reduced affinity. The only documented interaction partner for the cSH3 domain is the nuclear exportin, Crm-1, which is believed to bind to a putative nuclear export sequence (NES, residues L256-V266) within the cSH3 domain (Smith et al., 2002). Interestingly, based on sequence alignment analysis (Larson and Davidson, 2000), the predicted NES forms part of the central β-barrel in the protein and, furthermore, two of the residues (V264 & V266) are expected to be in the hydrophobic core of the protein (Figure 1.1). Thus, the mechanism by which Crm-1 recognizes this NES in the context of the native cSH3 structure is unclear. Therefore, one aspect of this work has been to solve the solution structure and study the

folding properties of the cSH3 domain of Crk-II. Progress towards utilizing domain specific labeling methods to label the cSH3 domain within the context of full-length Crk-II will also be described. These studies should clarify the role of the cSH3 domain in regulating the function of the Crk-II protein.

Objective of Thesis

Outlined below are the specific aims of the thesis proposal, originally submitted in May, 2002. Significant progress towards each of these aims is reported herein.

Specific aim 1 – Domain specific incorporation of 7AW into the SH3 domain of Crk-I. This will be one of the first examples where a modular domain whose biophysical properties have been well characterized in isolation will be studied within its native context of the full-length protein.

Specific aim 2 – Solve the solution structure and study the folding properties of the cSH3 domain of Crk-II. Begin to understand the role of the cSH3 domain in modulating the function of Crk-II and how this domain compares to other well-characterized SH3 domains from other proteins.

Specific aim 3 – Begin to explore the role of cSH3 domain within full-length Crk-II. Domain specific labeling methods will be utilized to label the cSH3 domain within the context of Crk-II. This should provide the important clues regarding the function of the cSH3 domain of Crk-II.

Chapter 2– Results

Section 2.1 – Domain specific incorporation of 7AW into the SH3 domain of Crk-I.

The past few decades have seen incredible advances in our understanding of the biophysical properties of proteins. This has been possible largely due to studies on small isolated protein domains such as the SH3 domain. Arguably, the next step in extending our understanding of protein biophysics is to study how these domains behave within the native context of large multi-domain proteins. The major obstacle here is the lack of techniques that allow us to incorporate labels that inform us on the biophysical properties of a specific domain within a larger protein. Thus, a study was undertaken to address this issue. Significant progress towards this goal is described herein.

Design and synthesis of Crk-I with a 7AW labeled SH3 domain

Tryptophan analogues, such as 5-hydroxytryptophan, fluorotryptophans and 7-AW, allow the local environment of Trp residues in proteins to be studied using either NMR or fluorescence spectroscopies. In this study, we decided to replace the two Trp residues in the SH3 domain of c-Crk-I, residues Trp169 and Trp170, with 7AW. As seen in the crystal structure of this domain in complex with a poly-Pro ligand from the protein, C3G (Figure 2.1A), Trp169 is involved in ligand binding whereas Trp170 is an important component of the hydrophobic core of the protein (Wu et al., 1995). In both these Trp residues, the 7-position is solvent exposed in the unliganded form of the protein (Figure 2.1B). This suggested to us that replacing these Trp residues with 7AW would not cause major perturbations in the structure of this SH3 domain. The effect of 7AW labeling on ligand binding was more of an open question.

To make Crk-I with a 7AW labeled SH3 domain (hereafter referred to as Crk-I[$SH3_{7AW}$]) by EPL requires the generation of two recombinant protein building blocks; an α-thioester derivative encompassing the unlabeled SH2 domain and an N-terminal Cys containing 7AW labeled SH3 domain (hereafter referred to as $SH3_{7AW}$). Residues Gly124-Ser125, which lie within the short linker region (residues 119-133) between the two domains, were chosen as our ligation junction. The chemistry requires mutation of the serine to a cysteine, however, previous studies in our lab

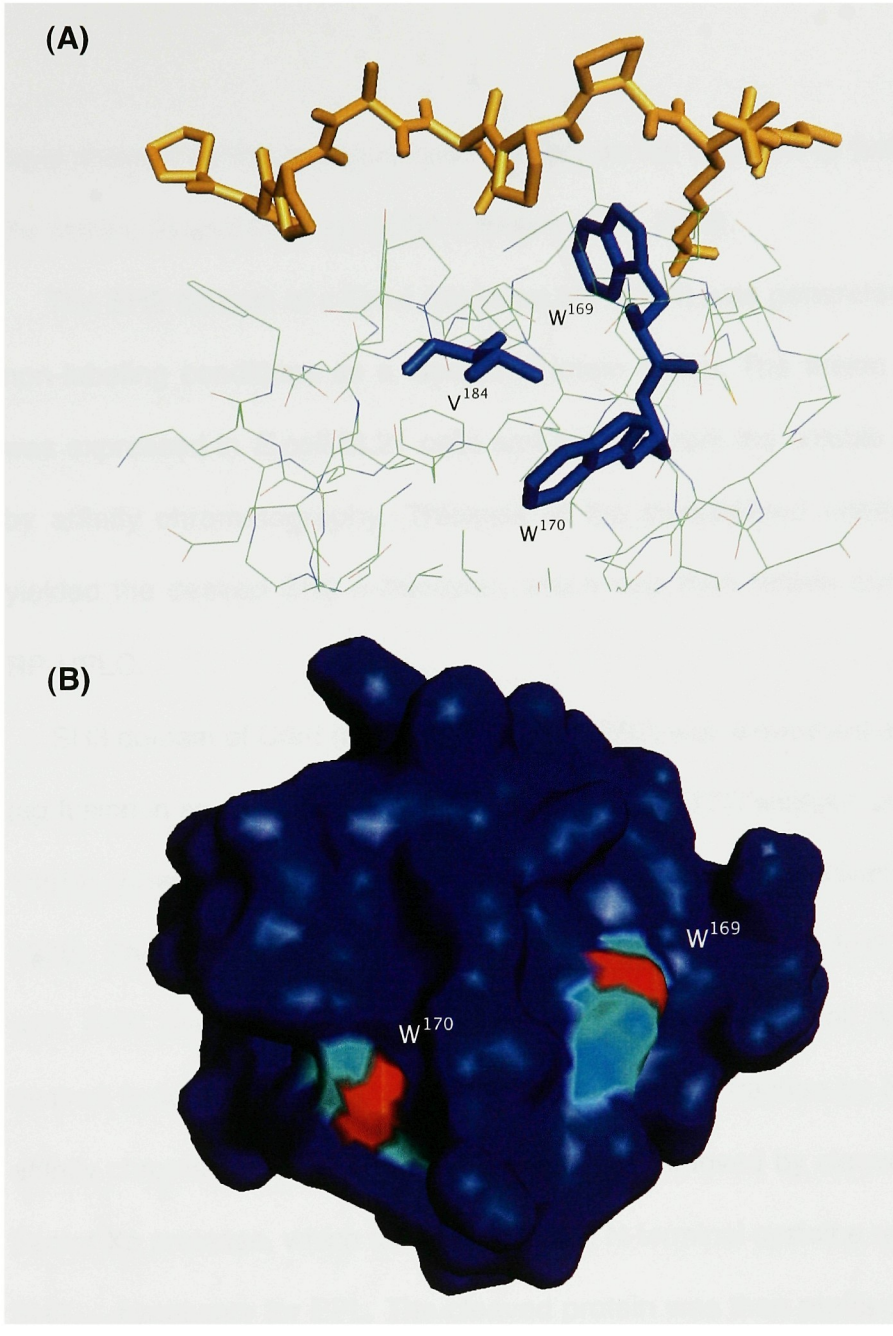

Figure 2.1 Structure of the SH3 domain of Crk-I (pdb: 1CKA). **(A)** Expanded view of the structure of the SH3 domain of in complex with a poly-Pro peptide ligand (in orange) from the C3G protein. Trp 169, Trp 170 and Val 184 are highlighted in blue. **(B)** Structure of the SH3 domain with surface rendering. The 7-positions of the two Trp residues are highlighted in red.

have shown that this mutation has no effect on the structure or function of the protein (Blaschke et al., 2000; Ottesen et al., 2003).

The SH2 domain of c-Crk-I (residues M1-G124) was generated under non-labeling conditions as a Gyrase A intein fusion. The fusion protein was expressed in *E.coli* BL21 cells and purified from the soluble fraction by affinity chromatography. Thiolysis of the immobilized intein fusion yielded the desired SH2 α-thioester, which was then further purified by RP-HPLC.

SH3 domain of Crk-I (residues C125 to G207) was expressed as a His-tag fusion in an *E.coli* Trp auxotroph (*E.coli* CY15077(Yanofsky and Horn, 1995)) grown in minimal media supplemented with 7AW (Schlesinger, 1968). This procedure replaced the two Trp residues in the SH3 domain with 7AW (hereafter referred to as $SH3_{7AW}$). Following cell lysis, the desired fusion protein was purified from the soluble fraction by Ni^{2+}-NTA affinity chromatography. The His-tag was then removed by cleavage with Factor Xa protease, which also exposed the N-terminal cysteine residue in $SH3_{7AW}$ necessary for EPL. The cleaved protein was then purified by RP-HPLC in an overall yield of ~1.5mg/L. Unlabeled SH3 domain (hereafter referred to as $SH3_W$) was expressed under non-labeling conditions using rich media and purified as above.

The recombinant building blocks in hand, Crk-I[$SH3_{7AW}$] was then

assembled via EPL. The ligation reaction was initiated by mixing the purified SH2 α-thioester and $SH3_{7AW}$ domains in the presence of 6M Guanidinium chloride. Use of denaturating conditions was found to increase the yields in this system, however, in general the ligation approach can be performed under native conditions (Muir, 2003). The reaction was monitored by RP-HPLC, which indicated near complete conversion after 5 days (Figure 2.2A). The ligation product, Crk-I[$SH3_{7AW}$], was then purified by RP-HPLC and characterized by ESMS (Figure 2.2). Purified Crk-I[$SH3_{7AW}$]was refolded as described previously (Blaschke et al., 2000).

To determine the extent of Trp replacement with 7AW, the purified $SH3_{7AW}$ and $SH3_W$ domains were each cleaved with trypsin and the fragments analyzed by MALDI-TOF mass spectrometry. Of particular interest was the mass of an 18 amino acid fragment that contained both Trp residues in the protein (residues I161-K178). Trp and 7AW differ by 1 Da, thus the $SH3_{7AW}$ derived fragment should be 2 Da heavier than the corresponding $SH3_W$ fragment, assuming both Trp residues are replaced. Interpretation of the mass spectral data is complicated by the presence of the minor isotopes of C, H, N and O, which must be considered when calculating the incorporation efficiency (Senear et al., 2002). Figure 2.3 shows the theoretical and observed isotope distributions in the mass spectrum of the relevant peptide fragment from $SH3_{7AW}$ and $SH3_W$. This

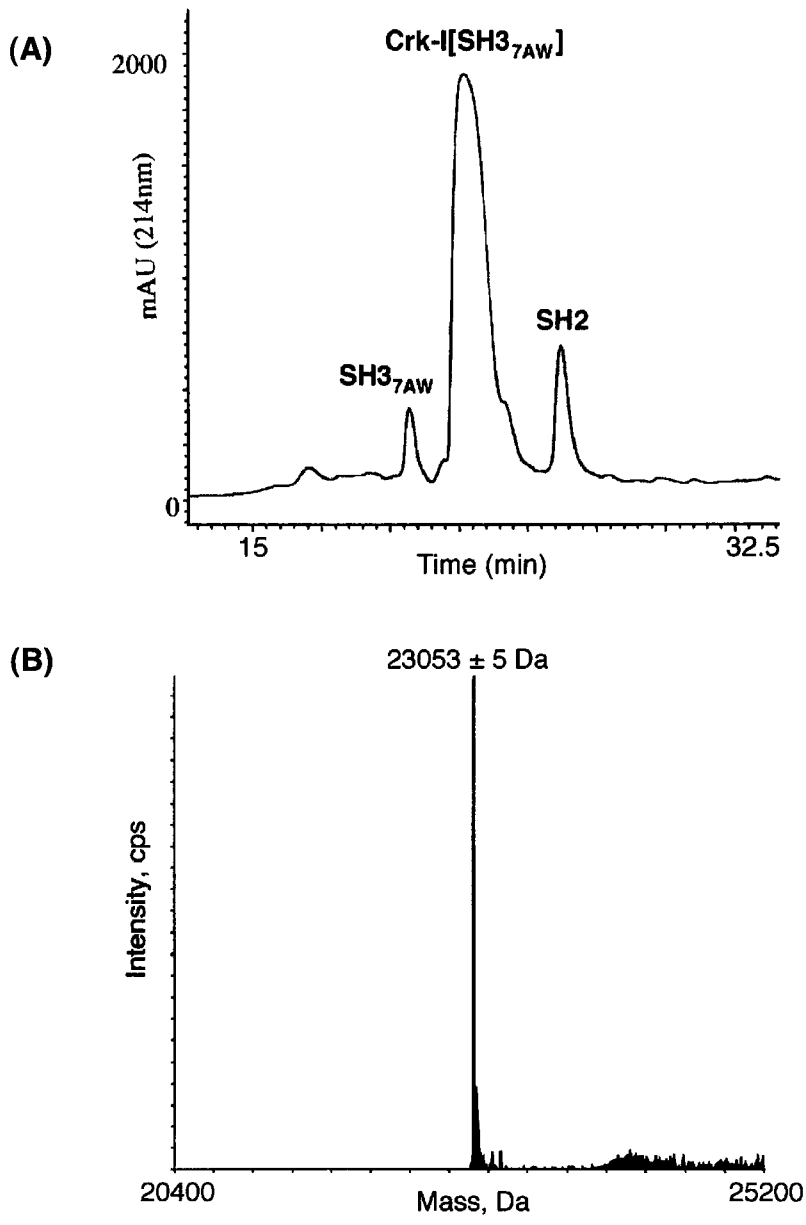

Figure 2.2. Semi-synthesis of Crk-I[SH3$_{7AW}$]. (A) RP-HPLC chromatogram of the crude mixture after ligation between the SH2 and the SH3$_{7AW}$ domain to give Crk-I[SH3$_{7AW}$] after 5 days. The two starting materials and the product are indicated. The gradient was 30%-45% Buffer B (90% acetonitrile/0.1% TFA in water) **(B)** The observed molecular mass calculated from the charged species seen in ESMS spectra of the ligation product, Crk-I[SH3$_{7AW}$]. Expected mass: 23048.5 Da.

comparison demonstrates that 7AW labeling in $SH3_{7AW}$ was highly efficient, ≥93% at both sites.

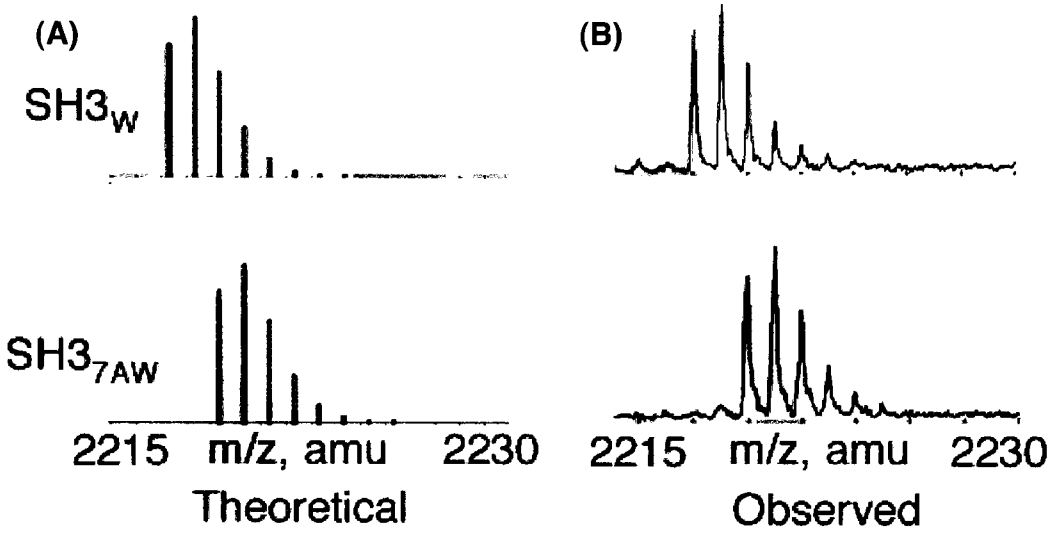

Figure 2.3. Efficiency of 7AW incorporation. Calculated (assuming complete incorporation of 7AW) and observed isotope distributions of trypsin digest fragments of $SH3_W$ and $SH3_{7AW}$. $SH3_W$ peptide fragment: ^{161}IRDKPEEQWWNAEDSEGK^{178} (expected mass: 2217. 3Da) and $SH3_{7AW}$ peptide fragment: ^{161}IRDKPEEQ[7AW][7AW]NAEDSEGK^{178} (expected mass: 2219.3Da). **(A)** Theoretical isotope distributions of $SH3_W$ peptide and of $SH3_{7AW}$ peptide (lower panel). **(B)** MALDI-TOF MS (reflectron) of $SH3_W$ peptide and of $SH3_{7AW}$ peptide (lower panel) with 4-hydroxycinnaminnic acid (4HCCA) as the matrix.

Initially, the $SH3_{7AW}$ domain was also expressed as a Gyrase A intein fusion. This intein contains a single Trp residue, which was also replaced by 7AW in the labeling procedure. Our studies indicate that replacement of the Trp residue in Gyrase A intein with 7AW has no effect on the functioning of the intein; thiolysis of the 7AW-containing fusion protein proceeded to a similar degree as for the wild-type intein (Figure 2.4). Thus, we can also generate 7AW labeled proteins with C-terminal thioesters. Importantly, this will allow us to label any part of the protein and apply EPL to its fullest extent.

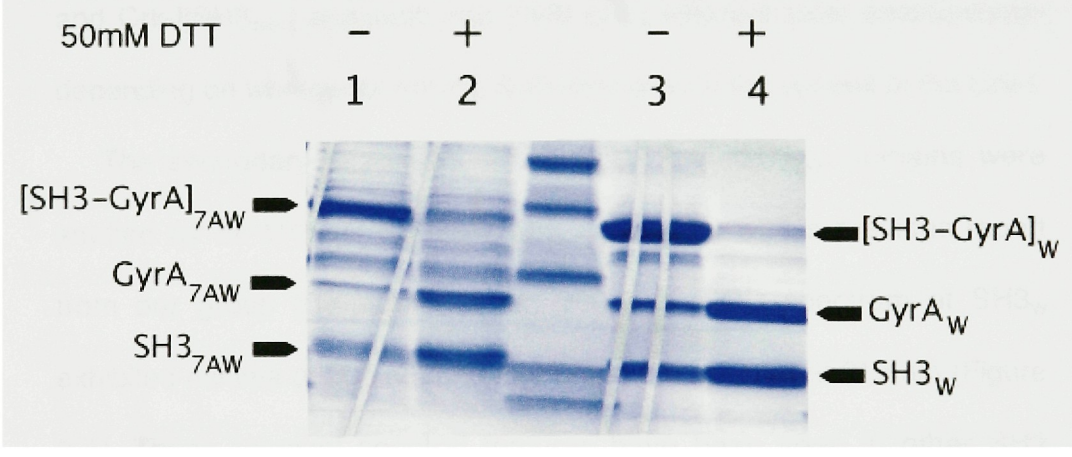

Figure 2.4. Cleavage of 7AW labeled Gyrase A intein (GyrA) **with DTT.** The SH3-GyrA intein fusion protein labeled with 7AW ($[SH3\text{-}GyrA]_{7AW}$, lanes 1 and 2) and the unlabeled intein fusion protein ($[SH3\text{-}GyrA]_{W}$, lanes 3 and 4) were incubated for one hour either with DTT (lanes 2 and 4) or without DTT (lanes 1 and 3) and analyzed by SDS-PAGE. Cleavage of the intein fusion proteins with DTT yielded $SH3_{7AW}$, 7AW labeled GyrA ($GyrA_{7AW}$), $SH3_{W}$ and unlabeled GyrA ($GyrA_{W}$). Coomassie stained SDS-PAGE is shown above.

Structural characterization of $SH3_W$, $SH3_{7AW}$ and Crk-I[$SH3_{7AW}$]

A combination of fluorescence, CD and NMR spectroscopies were used to gauge the affect of 7AW labeling on the structural and folding properties of the Crk-I SH3 domain. The fluorescence emission spectra of $SH3_W$, $SH3_{7AW}$ and Crk-I[$SH3_{7AW}$] upon excitation at 310 nm are shown in Figure 2.5. $SH3_{7AW}$ and Crk-I[$SH3_{7AW}$] exhibited broad emission spectra with maxima at 388 nm and 377 nm, respectively, whereas $SH3_W$ was essentially non-fluorescent under these excitation conditions. Thus, as expected, the 7AW labeled SH3 domain can be selectively excited in the presence of Trp. The slight difference in the emission spectra of $SH3_{7AW}$ and Crk-I[$SH3_{7AW}$] suggests that 7AW sees different local environments depending on whether or not the SH3 domain is in the context of the Crk-I.

The secondary structures of the $SH3_W$ and $SH3_{7AW}$ domains were studied by far-UV CD spectroscopy. Consistent with a previous report from our group (Camarero et al., 2001), the CD spectrum of $SH3_W$ exhibited minima at 203 nm and 229 nm and a maxima at 219 nm (Figure 2.6). These unusual spectral features have been seen in other SH3 domains (Viguera et al., 1995) and are likely due to aromatic contributions (Camarero et al., 2001; Manning and Woody, 1989). The CD spectrum of $SH3_{7AW}$ differs slightly from that of $SH3_W$ (Figure 2.6), however, it does so in the regions attributed to aromatic contributions (the minimum at 229 nm is shifted to 232 nm and the maximum is shifted from 219 nm to 224 nm),

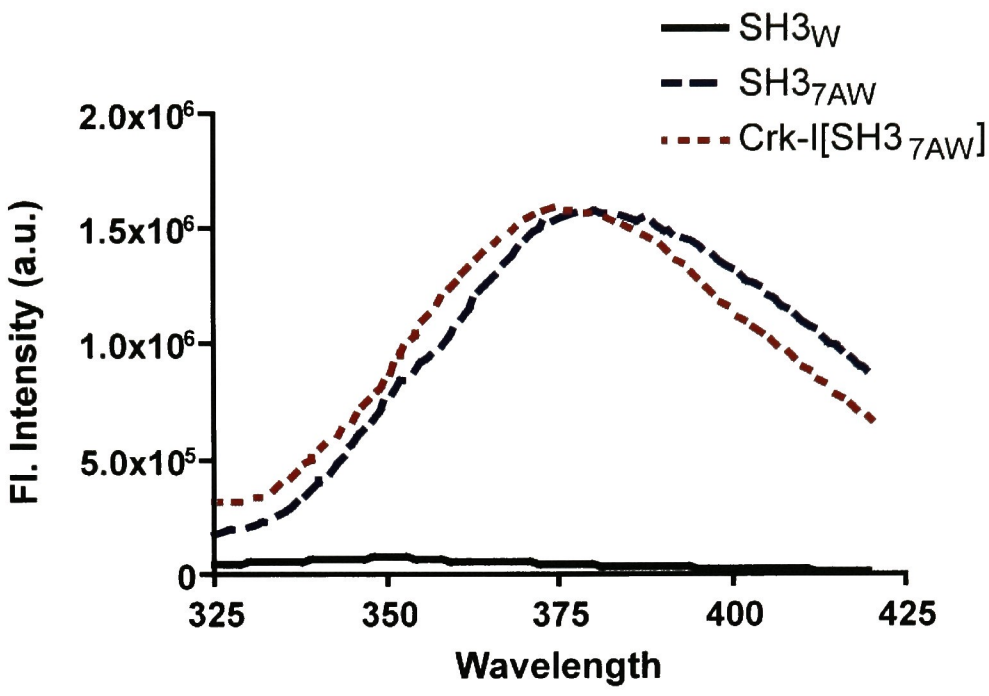

Figure 2.5. Fluorescence emission spectra of $SH3_W$, $SH3_{7AW}$and Crk-1[$SH3_{7AW}$]. The excitation wavelength was 310 nm. $SH3_W$ has virtually no fluorescence emission when excited at this wavelength.

possibly reflecting the differences in the aromaticity of 7AW *vs.* Trp. Importantly, the CD spectra of $SH3_W$ and $SH3_{7AW}$ are otherwise similar, indicating that 7AW incorporation does not cause major structural changes in the domain. Furthermore, the mid-point (T_m) of temperature denaturation of $SH3_W$ and $SH3_{7AW}$ was determined by following the change in the CD signal at 220 nm (Figure 2.7). Thermal denaturation of $SH3_W$ and $SH3_{7AW}$ showed 2-state behavior and was reversible. The T_m of $SH3_W$ and $SH3_{7AW}$ were similar (332.2 ± 0.4 and 330.0 ± 0.6 K, respectively), reinforcing the notion that 7AW is structurally tolerated in the c-Crk-I SH3 domain.

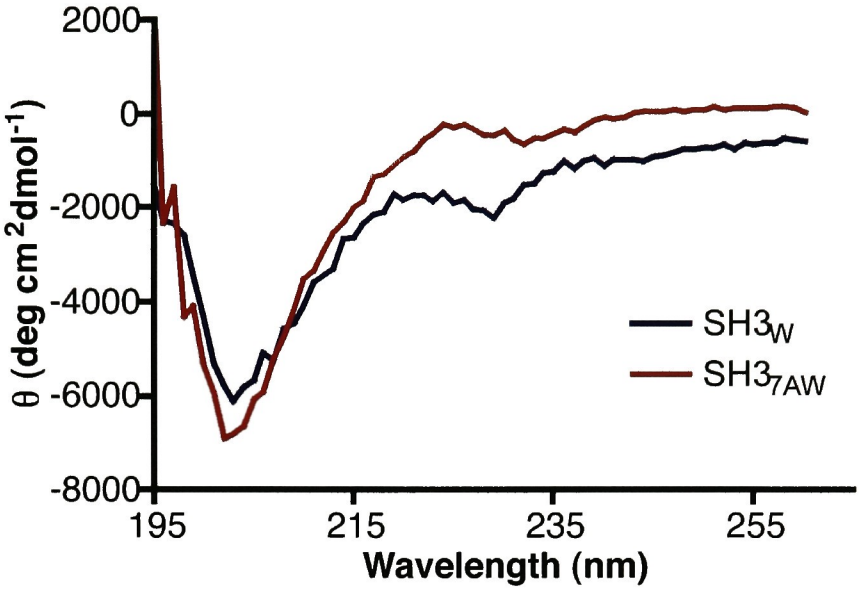

Figure 2.6. Far UV CD spectra of $SH3_W$ and $SH3_{7AW}$. The CD spectra of the two SH3 domains differs in the region (around 230 nm) attributed to contributions from the two Trp residues.

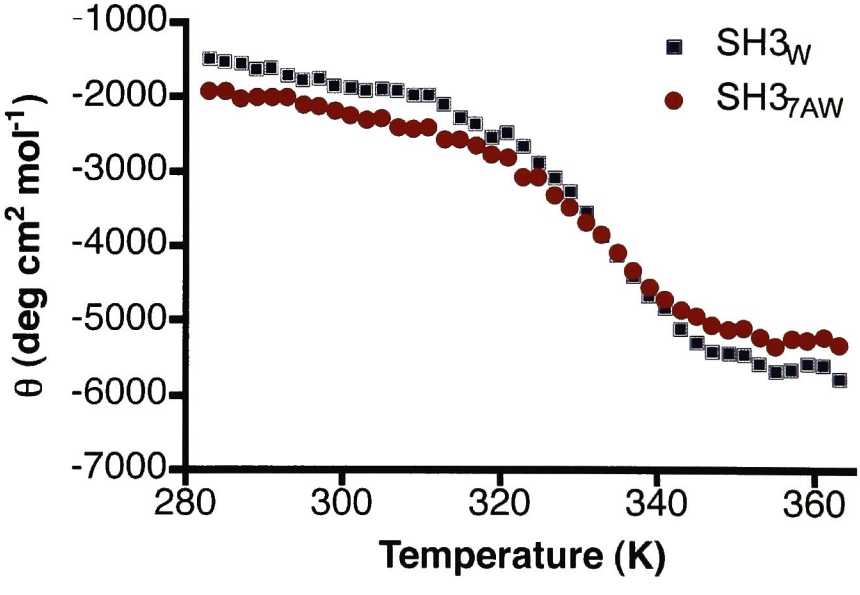

Figure 2.7. Thermal denaturation curves of $SH3_W$ and $SH3_{7AW}$. CD signal was recorded at 220nm. Thermal unfolding of both domains was found to be completely reversible.

Homonuclear ^{1}H NMR spectroscopy was used to further investigate the effects of 7AW incorporation (Figure 2.8 and 2.9). The NOESY spectra of $SH3_W$ and $SH3_{7AW}$ are very similar (Figure 2.9). The α-fingerprint regions of both spectra are well dispersed and indicative of a well-defined protein fold with significant contribution from β-sheets. Additional diagnostic features of a native fold in the $SH3_W$ and $SH3_{7AW}$ domains are the presence of NOESY cross-peaks between Trp170 and Val184 and upfield chemical shifts at around –0.85 ppm for γ-methyl protons of Val184 (Figure 2.8). These result from the side-chain of Val184 being tightly packed against the indole ring of Trp170 within the hydrophobic core of the protein (Figure 2.1A) (Wu et al., 1995). Thus, the CD and homonuclear NMR data suggests that the $SH3_W$ and $SH3_{7AW}$ domains share a common fold and that the incorporation of 7AW is well tolerated.

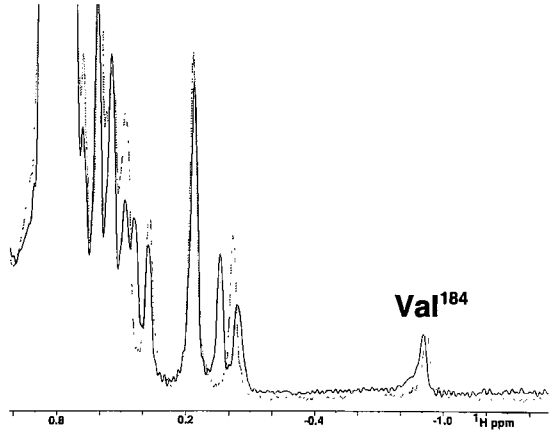

Figure 2.8. Overlaid of ^{1}H NMR spectra of $SH3_W$ (in gray) **and $SH3_{7AW}$** (in black). The spectra were acquired on a 400 MHz spectrometer. The signal corresponding to the γ-methyl protons of Val184 are indicated.

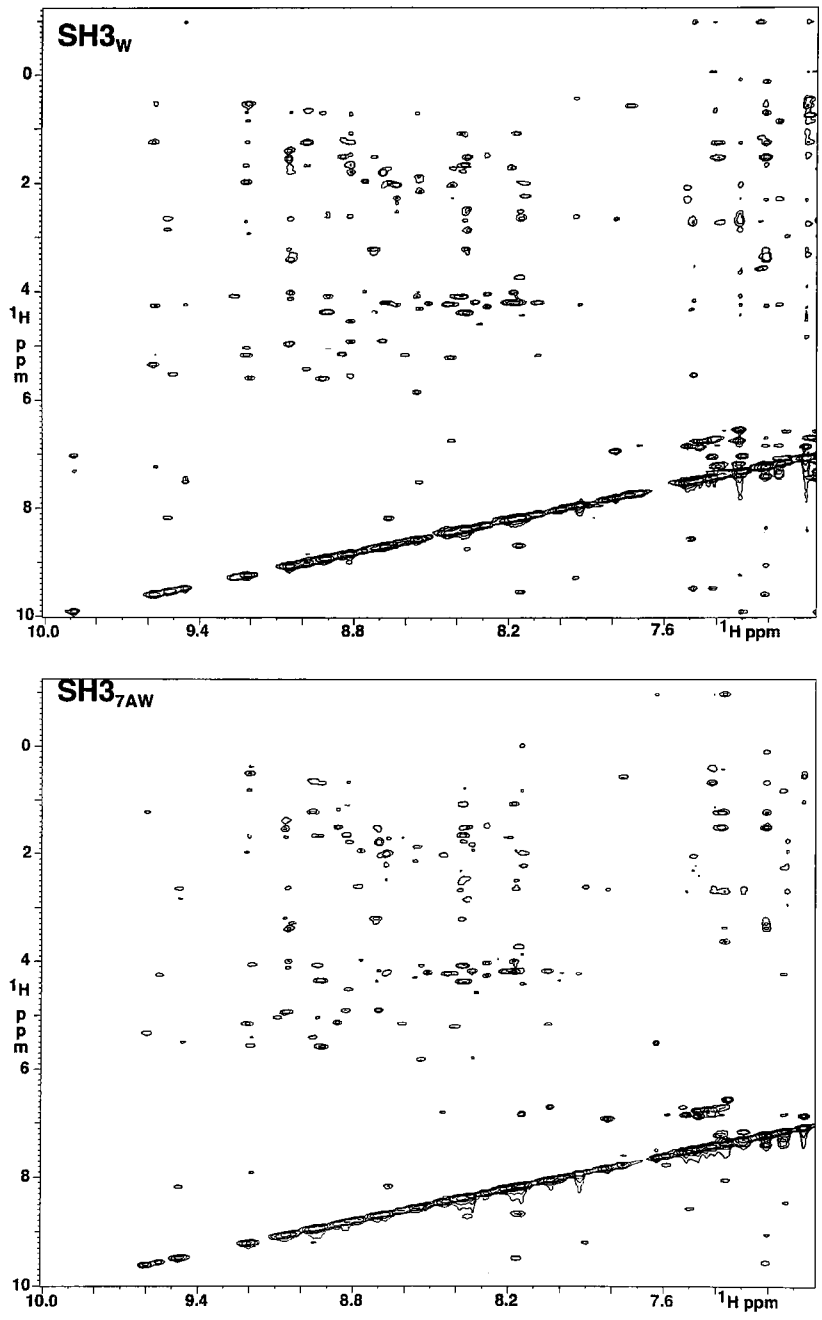

Figure 2.9. Amide and aromatic region of the ^{1}H NOESY spectra of $SH3_W$ and $SH3_{7AW}$. The spectra were acquired on a 500 MHz spectrometer.

Ligand binding properties of $SH3_W$ and $SH3_{7AW}$ and Crk-I[$SH3_{7AW}$]

The ligand binding properties of $SH3_W$, $SH3_{7AW}$ and Crk-I[$SH3_{7AW}$] were studied using a fluorescence based binding assay. The ligand used in these experiments was a fluorescein-containing analogue of a proline-rich peptide derived from the protein, C3G. The binding isotherm was followed either by the change in Trp fluorescence, in the case of $SH3_W$, or by the change in 7AW fluorescence, in the case of $SH3_{7AW}$ and Crk-I[$SH3_{7AW}$], as a function of added ligand to the protein (Figure 2.10). As shown in Table 2.1, these studies indicate that there are no significant differences in the ligand binding properties of $SH3_W$ and $SH3_{7AW}$. Thus, replacement of the two Trp residues in the SH3 domain by 7AW appears to be a silent mutation in terms of the functional properties of the molecule. This finding answers a key question posed at the outset of our studies, namely whether incorporation of the aza group at the 7-position of the indole ring of Trp169, which is ideally positioned to contact the ligand, would perturb the ligand binding properties of the protein. The data summarized in Table 2.1 also reveals that the ligand binding activity of the SH3 domain does not change dramatically in the context of Crk-I. However, it should be stressed that a small peptide ligand was used in these studies and not a native protein ligand.

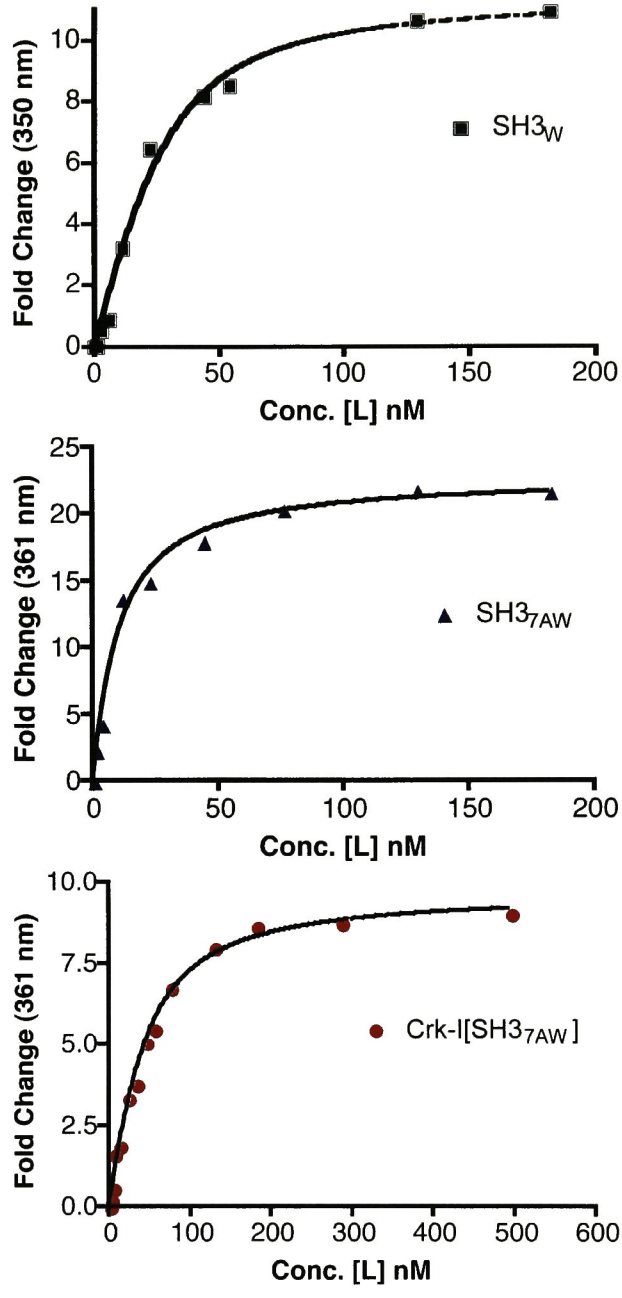

Figure 2.10. Binding isotherms of poly-Pro ligand from C3G to $SH3_W$, $SH3_{7AW}$ and Crk-I[$SH3_{7AW}$], as determined by change in either tryptophan or 7AW fluorescence. The solid curve represents the best fit for equation 1 (page 113). Shown here is a one representative curve out of three experiments.

Table 2.1. Dissociation Constants of the Various SH3 Constructs for the Poly-Pro ligand from C3G (H-PPPPLPPKRRK[ε-Flouresceine]G-NH_2)[a]

SH3 construct	K_d (nM)
$SH3_W$ domain	10.2 ± 1.6
$SH3_{7AW}$ domain	13.8 ± 2.7
Crk-I[$SH3_{7AW}$]	22.5 ± 1.3

[a]All experiments were performed at 25 °C in 20 mM sodium phosphate, 100 mM sodium chloride, pH 7.0. The numbers following ± symbol represent the standard errors to the fit.

Equilibrium stabilities of SH2, $SH3_W$ and $SH3_{7AW}$ and Crk-I[$SH3_{7AW}$]

The equilibrium stabilities of the SH2, $SH3_W$ and $SH3_{7AW}$ domains and Crk-I[$SH3_{7AW}$] were determined using the change in fluorescence of either Trp or 7AW residues upon the addition of chemical denaturants. All the constructs exhibited well-defined unfolding curves (Figure 2.11) and from this we elucidated the equilibrium stabilities (ΔG^0_{H2O}) and m values for all the proteins (Table 2.2). The equilibrium stabilities of $SH3_W$ matched closely with the published values for this domain (Camarero et al., 2001).

The ΔG^0_{H2O} value for the SH2 domain reveals that this protein is unusually stable (Table 2.2) (Scalley-Kim et al., 2003; Tzeng et al., 2000).

The estimated values of ΔG^0_{H2O} for $SH3_W$ and $SH3_{7AW}$, are similar to that of other SH3 domains (Grantcharova and Baker, 1997; Lim et al., 1994; Viguera et al., 1994) and to each other. The fact that the $SH3_W$ and $SH3_{7AW}$ domains have similar stabilities validates the labeling procedure and provides further evidence that the labeled and the unlabeled domains are comparable. Analysis of the ΔG^0_{H2O} values reveals that the SH3 domain when in the context of c-Crk-I has a similar stability to the isolated domain. Thus, equilibrium stability measurements do not reveal any major change in the thermodynamic properties of the SH3 domain in isolation versus in c-Crk-I. This represents the first direct comparison of the stability of an isolated SH3 domain to its stability when incorporated in an intact multi-domain protein.

Table 2.2. The thermodynamics of unfolding for the various constructs.[a]

Protein	$\Delta G°_{H2O}$ (kcal mol^{-1})	m (kcal mol^{-1} M^{-1})
SH2	8.1 ± 0.4	3.2 ± 0.1
$SH3_W$	3.9 ± 0.4	2.6 ± 0.2
$SH3_{7AW}$	3.2 ± 0.3	2.3 ± 0.1
Crk1[$SH3_{7AW}$]	3.7 ± 0.4	2.5 ± 0.2

[a]All experiments were performed at 25 °C in 20 mM sodium phosphate, 100 mM sodium chloride, pH 7.0. The numbers following ± symbol represent the standard errors to the fit.

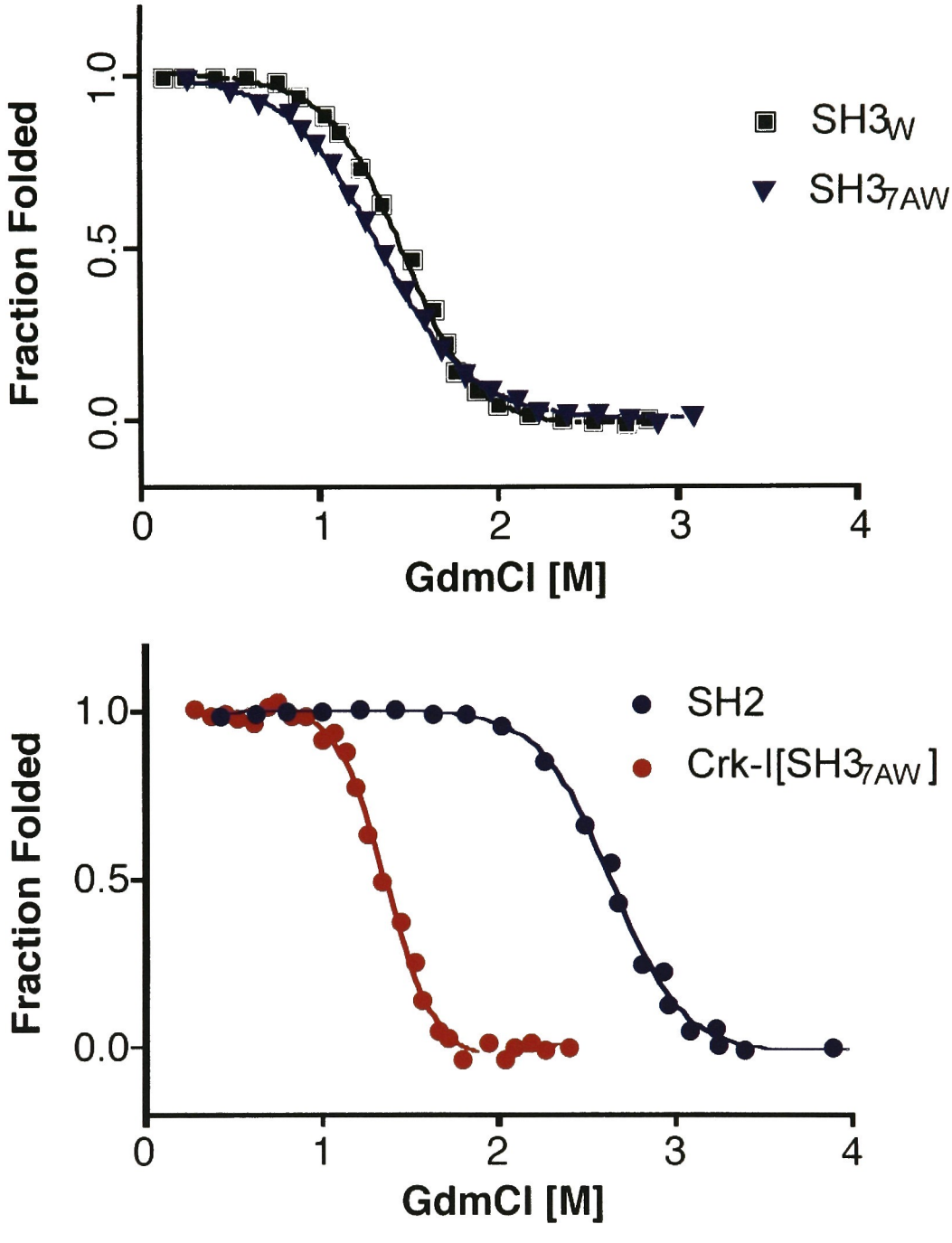

Figure 2.11. Chemical denaturation of $SH3_W$, $SH3_{7AW}$, SH2 domain and Crk-I[$SH3_{7AW}$], followed by fluorescence spectroscopy. The solid curve represents the best fit for eqs. 3-4.

Section 2.2 -- The solution structure and folding properties of the cSH3 domain of Crk-II.

The C-terminal SH3 domain of Crk-II is unique among SH3 domains in that all attempts to find a poly-proline ligand for this domain have failed. This has deepened the mystery surrounding the role that this domain plays in regulating the function of Crk-II. As a first step towards understanding the role of the cSH3 in regulating Crk-II function, a series of structural and biophysical studies on the isolated domain were carried out. Significant progress towards this goal is described herein. These studies also offer some insight into the likely structural and thermodynamic consequences of point mutations in the cSH3 domain that are known to deregulate Crk-II function.

Solution Structure of the Crk-II cSH3 domain

Sequence homology suggests that the cSH3 domain of Crk-II extends from P237 to L294 (Figure 1.1). However, sequence alignment analysis lacks the ability to predict domain boundaries with absolute certainty. Indeed, it has been observed in the nSH3 domain of Crk-II that minor truncations at the N-terminus can have serious consequences on the folding of the domain (Camarero et al., 2001). Therefore, a slightly longer cSH3 domain construct was used in our studies, extending from L230 to S304. The cSH3 domain was over expressed in *E.coli* as a His-tag fusion linked via a Tev protease cleavage site. Following affinity purification, the tag was cleaved and the cSH3 domain purified by ion exchange chromatography (Note: Use of the Tev protease leaves a glycine residue at the N-terminus of the protein). The purified cSH3 domain was exchanged into the NMR buffer (see methods) and concentrated to typically 0.5 mM for NMR analysis. An initial estimate of the rotational diffusion tensor was made using the ^{15}N relaxation rates R_1, R_2 and ^{15}N-^{1}H NOE values after removing the overlapping residues and those that had potential chemical exchange contribution. The best model that fits the NMR data was an isotropic model with the overall correlation time (τ_c) of 5.27 ns. This strongly suggests that the cSH3 domain behaves as a monomer in solution.

Uniformly ^{15}N and ^{13}C labeled cSH3 domain was used to obtain the

main-chain and side-chain assignments of the ^{1}H, ^{15}N and ^{13}C resonances using a standard set of triple resonance 3D experiments (Cavanagh, 1996) (see Appendix 1 for assignments). All backbone assignments were made with the exception of the amide nitrogen and amide proton resonances of L230, V247, and V271 (Figure 2.12). There are five proline residues in the cSH3 domain and they are in trans configuration based on H_{α}-H_{δ}/ H_{δ}' NOE data.

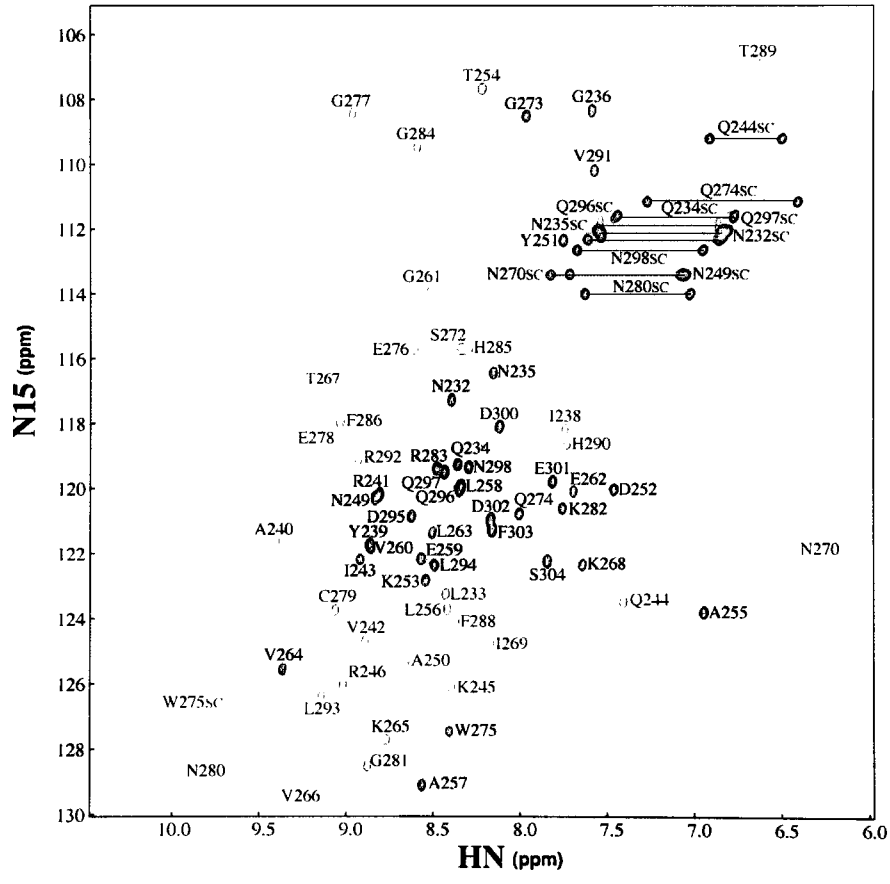

Figure 2.12. Assigned HSQC spectrum of cSH3 domain. The spectrum was assigned using a set of standard 3D NMR experiments. The side chain assignments are denoted as 'SC'.

An initial assessment of the cSH3 secondary structure was made using the deviations of assigned chemical shifts from random coil (Wishart and Sykes, 1994) (Figure 2.13). This indicates that the domain consists of five β -strands as seen for other SH3 domains. The steady state ^{15}N-[^{1}H] heteronuclear overhauser effect (NOE) indicates that, except for residues at the N-terminus (L230-G236) and the C-terminus (L294-S304), the domain is well ordered with relatively little flexibility in the intervening sequence compared to the termini of the domain (Figure 2.14). Therefore residues L230 to N235 at the N-terminus and residues P299 to S304 at the C-terminus were excluded from the final structure calculation.

All NOE cross peaks (both assigned and unassigned) in the ^{15}N NOESY-HSQC and ^{13}C NOESY-HSQC spectra were used by ARIA/CNS. The final set of structure calculations were done using a total of 1074 unambiguous restraints (538 intraresidue, 238 sequential, 91 short range, 17 medium range and 190 long range). Furthermore, 291 ambiguous restraints were also incorporated in the structure calculations. In addition, 96 dihedral (49 ϕ and 47 ψ) and 46 hydrogen bond restraints were used in the structure calculations. There was no increase of the target function upon adding the hydrogen bond restraints, thereby indicating that these restraints were consistent with the distance restraints. Structural statistics and analysis are given in Table 2.3.

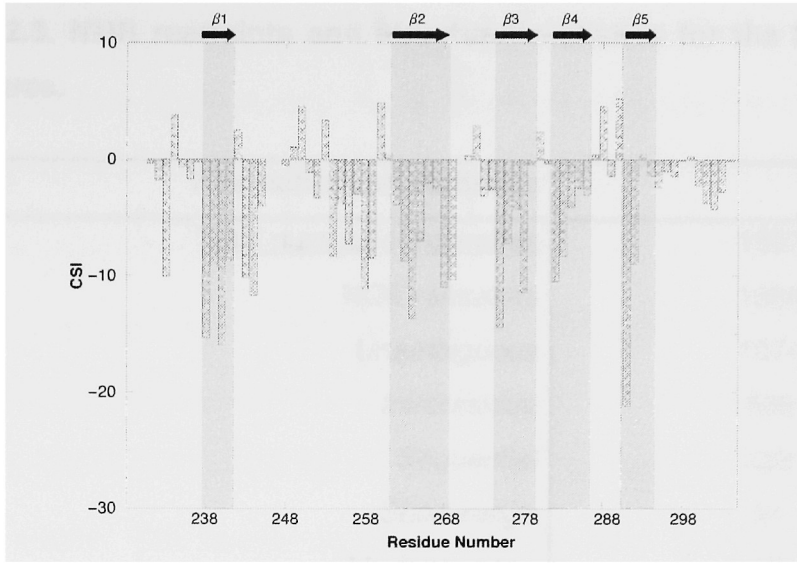

Figure 2.13. Chemical shift indexing (CSI) of the $^{13}C'$, $^{13}C_\alpha$ and ^{15}N assignments of the cSH3 domain. The assignments were compared to a database to predict the secondary structure of the domain.

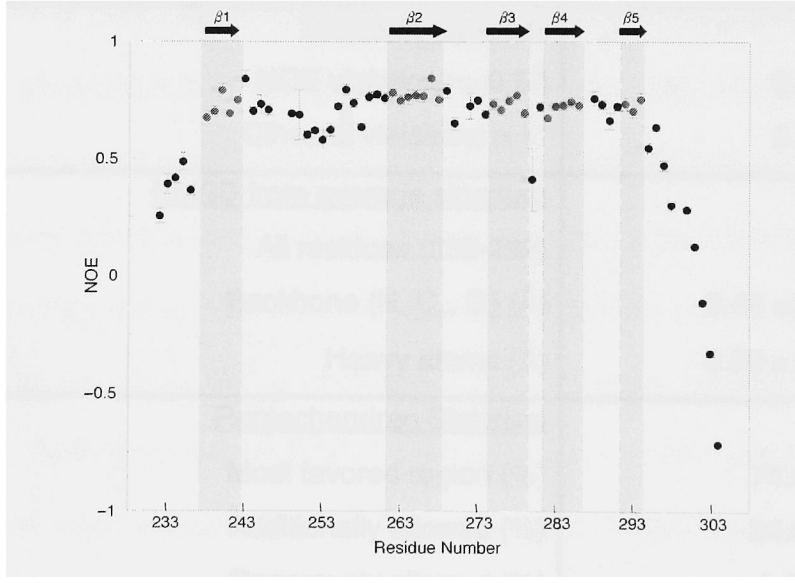

Figure 2.14. Hetero-nuclear NOE measurements for the cSH3 domain. Residues with NOE values 0.7 or more are expected to be well ordered and residues with NOE values less than 0.7 are expected to be in flexible loops or linker regions (Kay et al., 1989).

Table 2.3. NMR restraints and structural statistics for the 20 best structures.

Restraints and statistics	
Total number of restraints	1507
NOE restraints	1365
Unambiguous	1074
Intraresidue	538
Sequential	238
Short-range	91
Medium-range	17
Long-range	190
Ambiguous	291
Dihedral angle restraints	96
Hydrogen bond restraints	46
Structure Statistics	
NOE violations > 0.5Å	0
Dihedral violations > 5°	0
RMSD from average structure	
All residues (236-298)	
Backbone (N, C_α, C') (Å)	0.48 ± 0.12
Heavy atoms (Å)	0.99 ± 0.12
Ramachandran Statistics	
Most favored region (%)	74.8
Additionally allowed (%)	24.6
Generously allowed (%)	0.6
Disallowed (%)	0.0

The 20 calculated structures with the lowest target function values are superimposed in Figure 2.15A. These structures exhibit no distance and dihedral restraint violations greater than 0.5 Å and 5° respectively. The structure ensemble has average pairwise atomic rmsd of 0.68 ± 0.17 Å for backbone residues and 1.39 ± 0.21 Å for all heavy atom residues. The root mean square deviation of the ensemble with respect to the average structure was 0.48 ± 0.12 Å for the main-chain and 0.99 ± 0.12 Å for the heavy atoms. The quality of the structures were analyzed using PROCHECK and WHATIF. 74.8% of all non-glycine and non-proline residues are in the most favored regions of the ϕ-ψ space, with none of the residues in the disallowed regions (Table 2.3). The three-dimensional structure of the Crk-II cSH3 domain has five β-strands (residues β1=I238-V242, β2=E262-K268, β3=W275-C279, β4=K282-F286 and β5=R292-L294) organized in an anti-parallel β-barrel structure (Figure 2.15B). The secondary structure of the cSH3 domain is similar to that determined by homology modeling (Reichman et al., 2005). Analogous to all other SH3 domains, the stands are connected by a series of loops, namely; residues I243 to G261 (by convention the RT-loop), residues I269 to Q274 (the n-Src loop) and residues N280 to G281 (the distal loop). The last two β-strands are connected by a short 3_{10}-helix from F288 to H290.

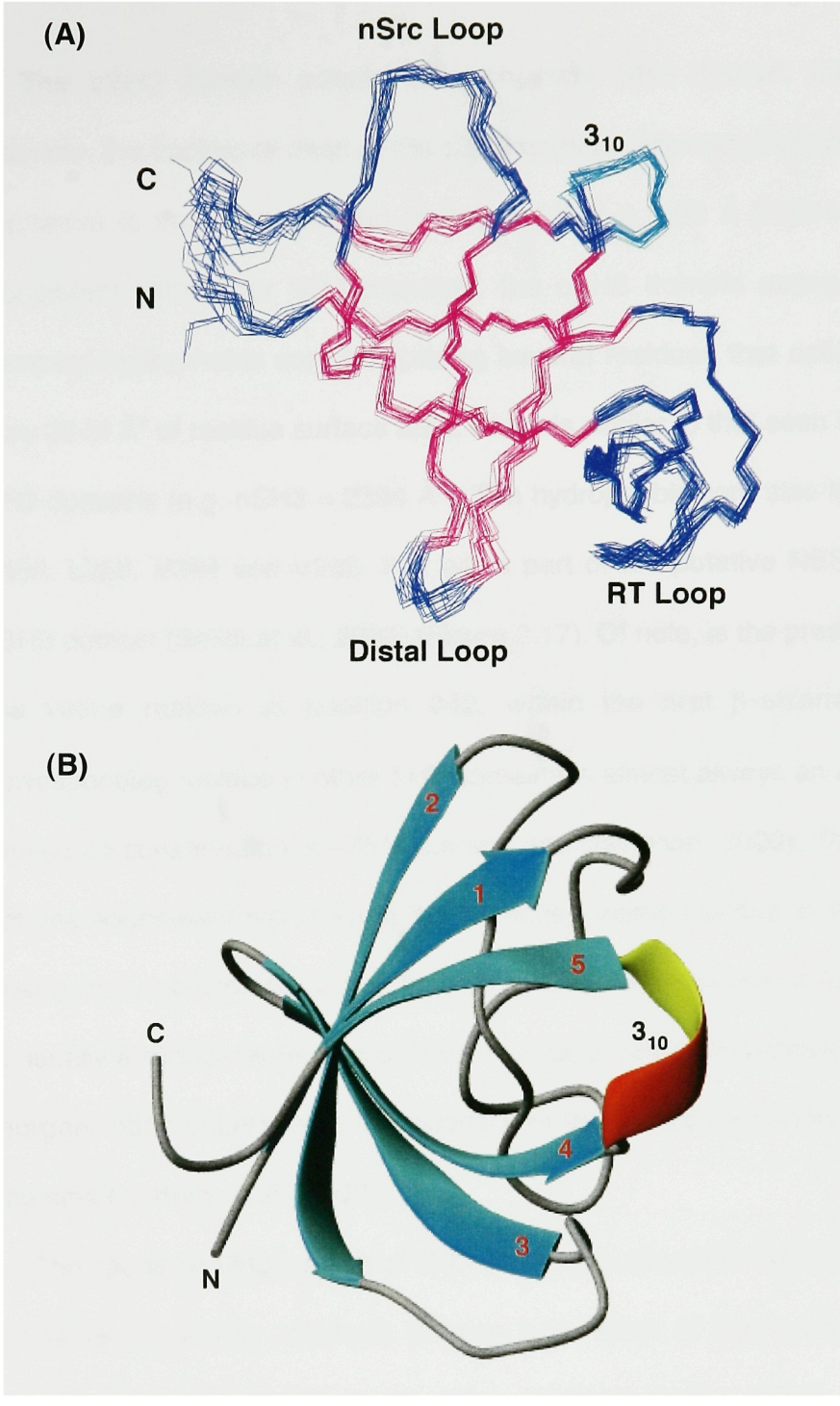

Figure 2.15. Solution structure of the cSH3 domain. The structures were drawn using MOLMOL (Koradi et al., 1996). **(A)** The 20 lowest energy structures are superimposed. β-strands are shown in red and the 3_{10}-helix is shown in cyan. **(B)** Ribbon diagram of the cSH3 domain.

The cSH3 domain adopts the canonical SH3 domain fold. For example, the backbone rmsd of the cSH3 domain obtained after structural alignment to the nSH3 domain (1cka) of Crk-II is 1.29 Å (Figure 2.16). Consistent with other SH3 domains, the cSH3 domain possesses a compact hydrophobic core comprising several residues that collectively bury 2249 $Å^2$ of residue surface area, which is similar to that seen in other SH3 domains (e.g. nSH3 = 2355 $Å^2$). The hydrophobic core also includes L256, L258, V264 and V266, that are a part of the putative NES in the cSH3 domain (Smith et al., 2002) (Figure 2.17). Of note, is the presence of the valine residue at position 242, within the first β-strand. The corresponding residue in other SH3 domains is almost always an alanine, indeed its conservation is ~75% (Larson and Davidson, 2000). The extra volume associated with having the branched valine residue in the core seems to be accommodated, at least in part, by the presence of A240 that is usually a slightly larger hydrophobic residue in other SH3 domains. This reorganization underscores the plasticity of the hydrophobic core of SH3 domains (Northey et al., 2002).

The ligand-binding surface of SH3 domains is characterized by several conserved aromatic residues (Figure 1.1). Most of these conserved residues are mutated to more polar residues in the cSH3 domain, specifically, Q244, R246, Q274, and H290 (Figure 2.18A). Thus, a question at the outset of this study was whether the

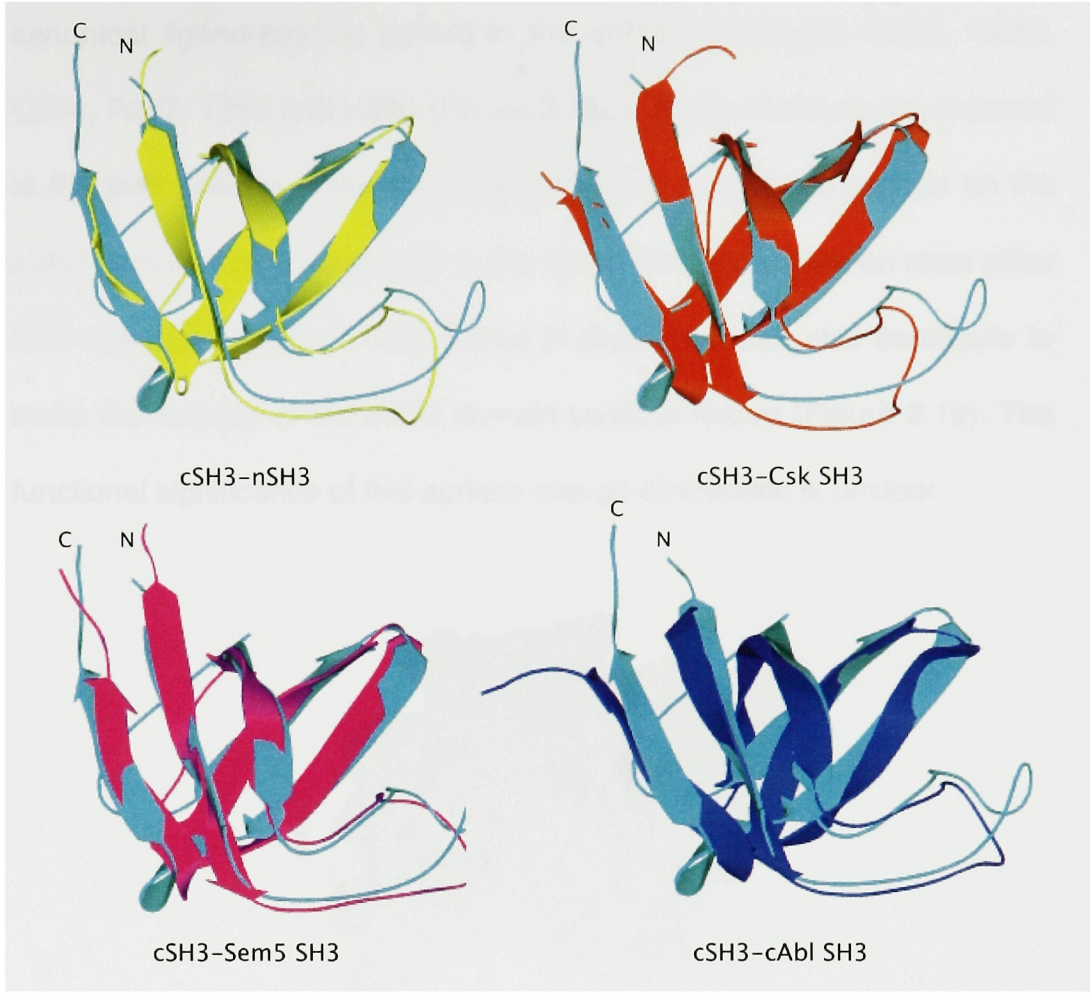

Figure 2.16. Structural alignment of the cSH3 domain with other SH3 domains. The cSH3 domain is shown in cyan. The nSH3 domain of Crk-II is shown in yellow, pdb code:1cka (rmsd= 1.29 Å). Csk SH3 domain is shown in red, pdb code:1jeg (rmsd= 1.24 Å). cAbl SH3 domain is shown in blue, pdb code:1abo (rmsd= 1.36 Å). Sem5 SH3 domain is shown in magenta, pdb code:1sem (rmsd= 1.28 Å).

conserved aromatic residues on the ligand-binding surface have been replaced with non-conserved aromatic residues or is the surface indeed more polar. Based on structural alignment, the residues that form a

canonical ligand-binding pocket in the cSH3 domain are Q244, G273, Q274, P287, T289 and H290 (Figure 2.18B). These residues are exposed to the bulk solvent in the structure forming a contiguous surface on the cSH3 domain that corresponds to the ligand binding surface on most other SH3 domains (Figure 2.18B). Some of these residues also contribute to make the surface of the cSH3 domain basic in nature (Figure 2.19). The functional significance of this surface charge distribution is unclear.

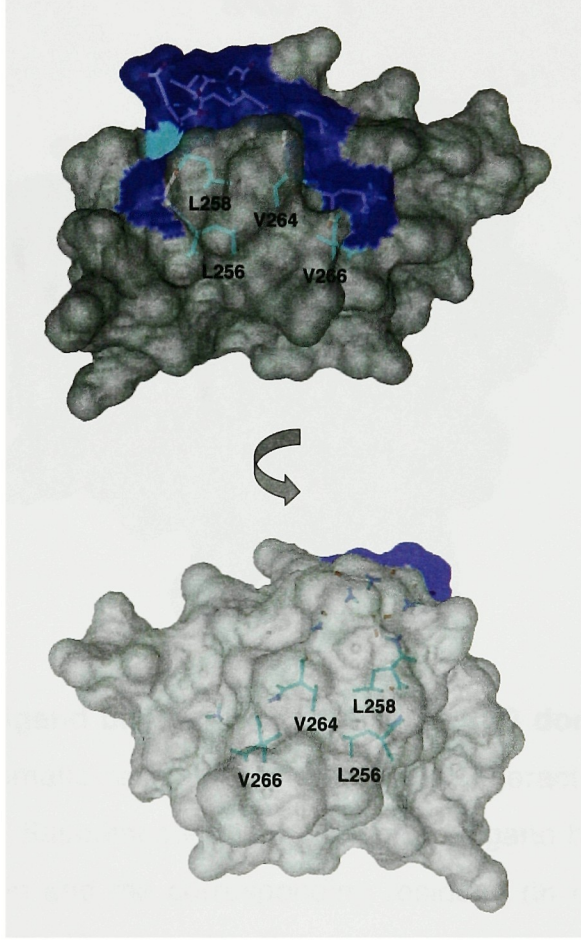

Figure 2.17. Surface representation of the putative NES (highlighted in blue and cyan) **in the cSH3 domain.** The residues in cyan are part of the hydrophobic core of the domain and are indicated.

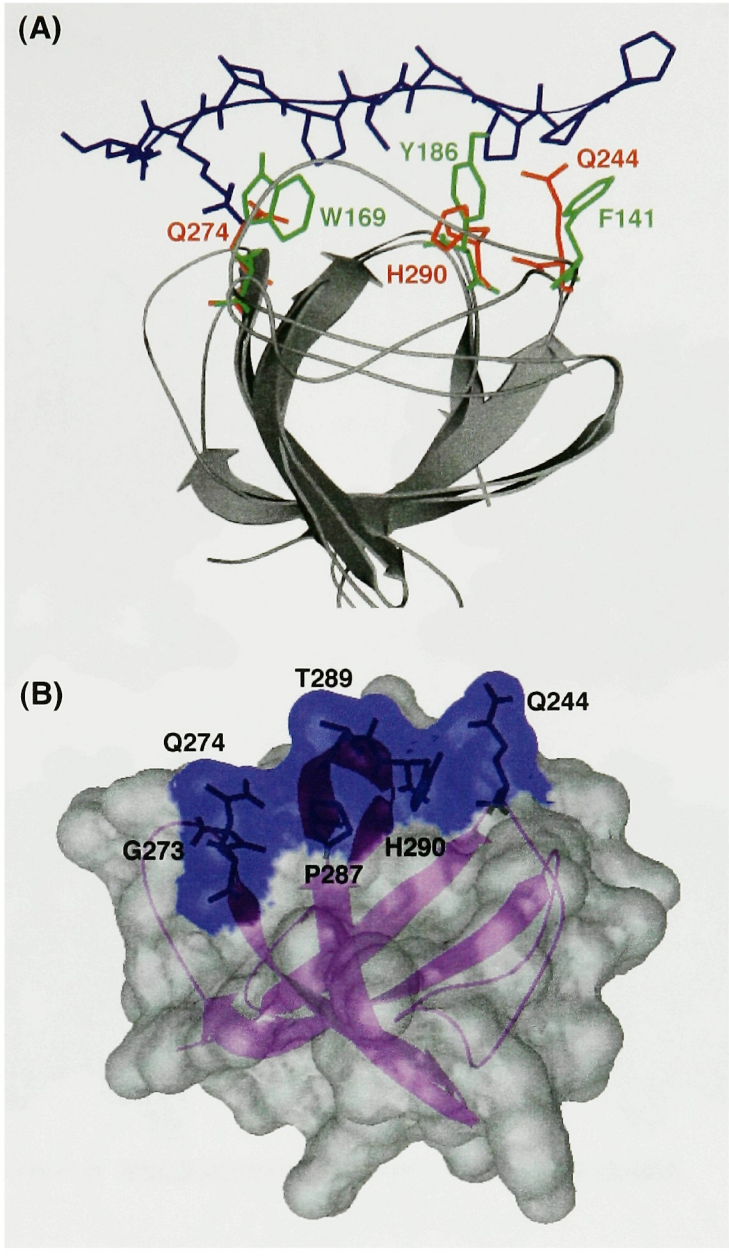

Figure 2.18. The ligand binding pocket of the cSH3 domain. (A) The ligand and the aromatic residues (in green) that interact in the nSH3 domain are shown. Superimposed is the canonical ligand binding pocket of the cSH3 domain and the corresponding residues (in red) lining the pocket in the cSH3 domain. **(B)** The residues that form the entire canonical ligand binding pocket in the cSH3 domain are shown in surface representation and are indicated.

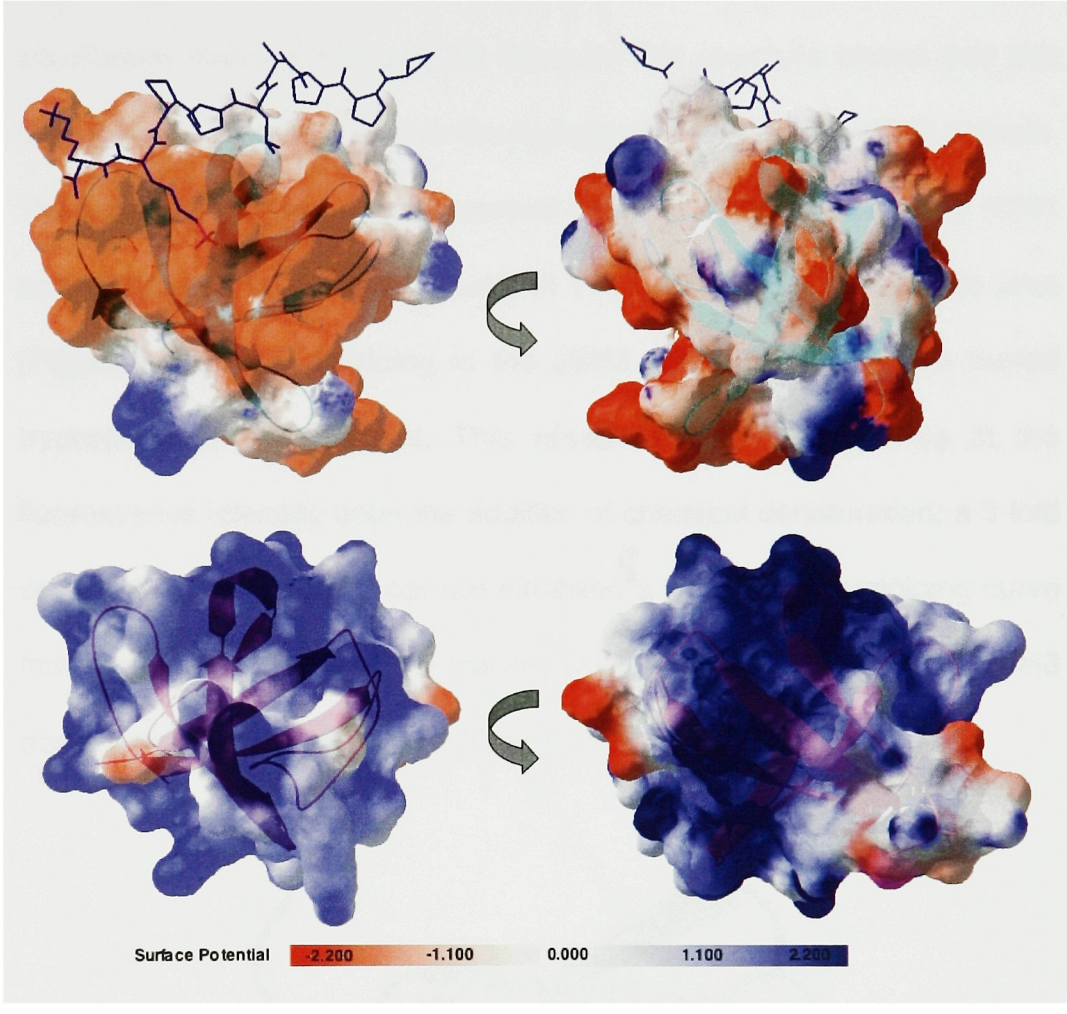

Figure 2.19. Surface charge distribution of the nSH3 domain (top panel, 1cka) **and the cSH3 domain** (bottom panel). The electrostatic surface representation of the two structures was drawn using Swisspdb viewer (Guex and Peitsch, 1997). The poly-proline ligand is shown for the nSH3 domain.

Folding Properties of the cSH3 domain of Crk-II

The change in the fluorescence signal of the tryptophan residue upon addition of chemical denaturants was used to determine the equilibrium stability of the cSH3 domain. The structure shows that this tryptophan residue is buried in the hydrophobic core of the cSH3 domain. The tryptophan packs against several residues (I238, V266, F288, V291 and L293) and results in the burial of 1117 $Å^2$ hydrophobic surface area (Figure 2.20). The unfolding of the cSH3 domain exposes this buried tryptophan to bulk solvent. This results in a robust change in the fluorescence intensity upon the addition of chemical denaturation; a 3-fold decrease at 340 nm. The domain exhibited a well-defined unfolding curve from which we elucidated the stability (ΔG^0_{H2O}) and m value for the cSH3 domain using equation 3 and 4 (Figure 2.21, Table 2.4).

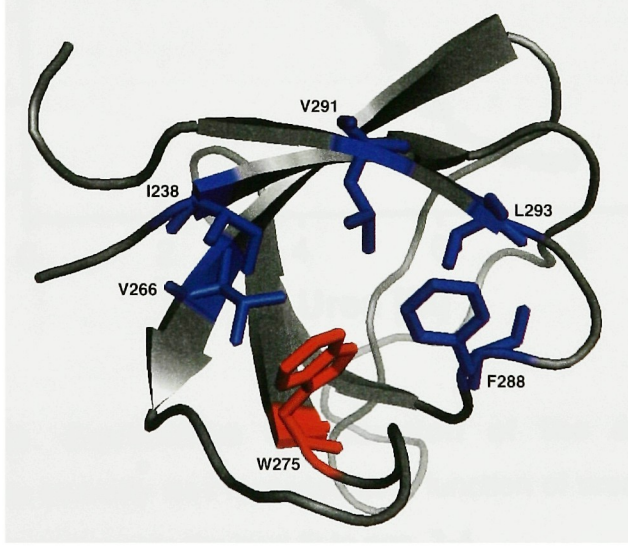

Figure 2.20. The packing of W275 into the hydrophobic core of the cSH3 domain. The residues that pack along with W275 are indicated.

Equilibrium denaturation measurements show that the cSH3 domain folds in a reversible two-state manner. The equilibrium stability of the cSH3 domain was found to be 3.3 kcal mol^{-1}, which is similar to that observed for other SH3 domains (Camarero et al., 2001; Grantcharova and Baker, 1997; Guijarro et al., 1998; Plaxco et al., 1998; Viguera et al., 1994) (Table 2.4). The m-value, which is proportional to the change in solvent exposed hydrophobic surface area upon unfolding, is 0.66 kcal mol^{-1} M^{-1}, which is slightly lower than that seen for other SH3 domains (Camarero et al., 2001; Viguera et al., 1994) (Table 2.4).

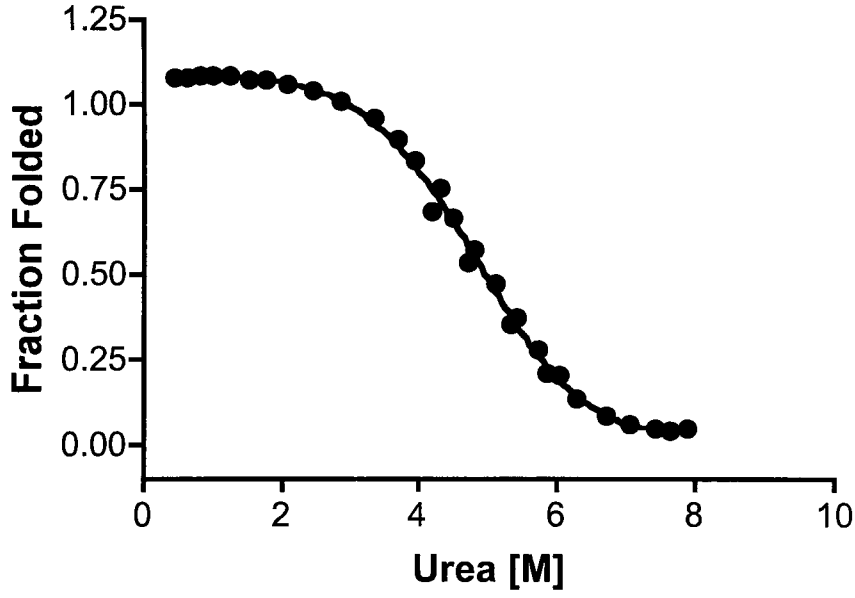

Figure 2.21. Equilibrium denaturation of the cSH3 domain. Fluorescence intensity was recorded as a function of urea concentration. The solid line represents the best fit to eqs. 3-4.

The folding kinetics of the cSH3 domain was measured using fluorescence-detected stopped-flow experiments (Figure 2.22). Interestingly, the data can be fit to a single exponential decay, i.e., the refolding of the cSH3 domain is monophasic. This was surprising since the cSH3 domain contains five proline residues, which was expected to result in a slow-phase in the folding amplitude due to proline *cis-trans* isomerization as has been observed in most other SH3 domains (Camarero et al., 2001; Grantcharova and Baker, 1997; Guijarro et al., 1998; Viguera et al., 1994). The absence of the slow phase in the folding

Table 2.4. Folding thermodynamics and kinetics for the cSH3 domain[a]

Equilibrium Thermodynamics	
ΔG^0_{H2O} (kcal mol^{-1})	3.3 ± 0.3
m-value (kcal mol^{-1} M^{-1})	0.7 ± 0.1
Folding Kinetics	
k_f (sec^{-1})	82.6 ± 2.9
k_u (sec^{-1})	0.2 ± 0.03
ΔG^0_{kin} (kcal mol^{-1})	3.7
m_f-value (kcal mol^{-1} M^{-1})	0.6 ± 0.01
m_u-value (kcal mol^{-1} M^{-1})	0.2 ± 0.01
m_{total}-value (kcal mol^{-1} M^{-1})	0.8

[a]All experiments were performed at 25 °C in 20 mM sodium phosphate and 100 mM sodium chloride at pH 7.2. The numbers following the ± represent standard errors to the fit.

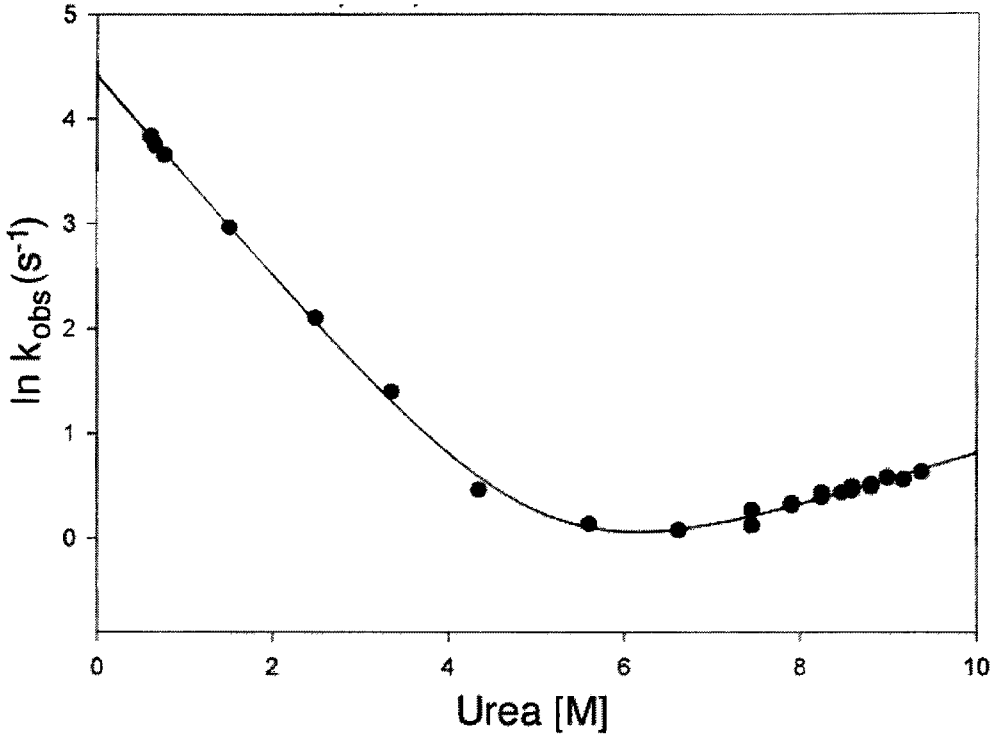

Figure 2.22. Folding kinetics of the cSH3 domain. Chevron plot of the natural log of the observed rate constant as a function of urea concentration. The solid line represents the best fit to eq. 5.

amplitude indicates either that the proline residues are all in the trans configuration in the unfolded state or that the observed tryptophan fluorescence is unaffected by the nature of proline isomerization. This situation is reminiscent of the Fyn SH3 domain, which also exhibits monophasic refolding behavior despite containing two proline residues that are trans in the native state (Plaxco et al., 1998). Figure 2.22 shows the plot of the natural logarithm of the observed rate constant, k_{obs} versus urea concentration. The chevron plots show a classic V-shape curve, as

expected for a protein that folds in a two-state manner. At low denaturation concentration we did not observe any rollover, i.e., deviation from linearity. The ΔG^0 value and the m-value calculated from the kinetic measurements, 3.66 kcal mol^{-1} and 0.71 kcal mol^{-1} M^{-1}, are in good agreement with the equilibrium measurements (Table 2.4). The ratio of m_f / m_{total}, denoted β_T, is a measure of the relative compactness of the transition state. The calculated value for cSH3 is 0.78 indicating a compact transition state ensemble. This value is similar to that observed with other SH3 domains. For example, the value for nSH3 domain of Crk-II is 0.71 (Camarero et al., 2001). The kinetic data is also consistent with an apparent two-state folding model.

Section 2.3 – The functional role of the cSH3 domain within full-length Crk-II.

Several lines of evidence point to an important role of the C-terminal SH3 domain (cSH3) in regulating the function of Crk-II. The structure offers us important clues to the function of the cSH3 domain and explains the results of some of the mutational studies performed on Crk-II. However, understanding the detailed nature of the functional interplay between the cSH3 domain and the rest of Crk-II requires the use of domain specific labeling. In this section, progress towards that goal will be described.

Design and synthesis of Crk-II with a cSH3 domain specific label

To explore the role of the cSH3 domain within Crk-II, two domain specific labeling strategies were employed. The first strategy was to incorporate 7AW specifically into the cSH3 domain and utilize fluorescence spectroscopy to study the domain. There is a single Trp residue in the cSH3 domain (W275), which is buried in the hydrophobic core. As can be seen from the structure, the 7-position of Trp 275 is also in the hydrophobic core of the protein (Figure 2.23). This suggests that replacing this residue with 7AW might perturb the structure of the cSH3 domain. The extent of this perturbation was an open question. With this in mind a second strategy was undertaken. In this approach the cSH3 domain was segmentally labeled with stable isotopes and utilize NMR spectroscopy.

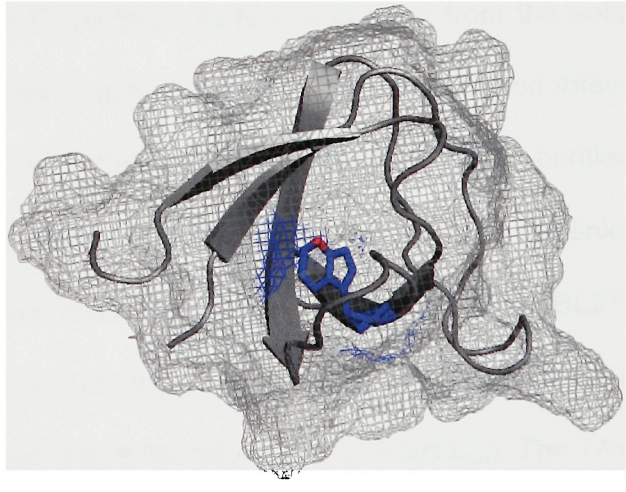

Figure 2.23. Structure of the cSH3 domain of Crk-II with surface rendering showing Trp 275 (in blue). The 7-position is highlighted in red.

To make Crk-II with a 7AW or isotope labeled cSH3 domain (hereafter referred to as Crk-II[lcSH3$_{7AW}$] and Crk-II[lcSH3$_{15N}$]) by EPL requires the generation of two recombinant protein building blocks; an α-thioester derivative encompassing the unlabeled SH2 domain and N-terminal SH3 domain (nSH3) and an N-terminal Cys containing labeled cSH3 domain. Based on previous studies in our lab (Blaschke et al., 2000; Ottesen et al., 2003), Gly207-Ser208, which lies in the linker region between the two SH3 domains, was chosen as the ligation site. Ser208 was mutated to a cysteine, which is required for the ligation chemistry. Based on previous work, this does not have any effect on the structure or function of the protein, at least *in vitro* (Blaschke et al., 2000; Ottesen et al., 2003).

The construct encompassing the SH2 domain and the nSH3 domain, which is the same as Crk-I (residues M1- G207), was made under non-labeling conditions as a Gyrase A intein fusion. The fusion protein was expressed in *E.coli* BL21 cells and purified from the soluble fraction by affinity chromatography. Thiolysis of the immobilized intein fusion yielded the desired Crk-I α-thioester, which was then further purified by RP-HPLC.

The cSH3 domain (C208-S304, which includes the linker region C208-P229) was expressed as a His-tag fusion in *E.coli* BL21 cells grown in minimal media supplemented with $^{15}NH_4Cl$ to generate a uniformly ^{15}N labeled domain (hereafter referred to as lcSH3$_{15N}$). The 7AW labeled cSH3 domain (hereafter referred to as lcSH3$_{7AW}$) was also expressed as a His-

tag fusion protein in *E.coli* Trp auxotrophic cells (Muralidharan et al., 2004) grown in minimal media supplemented with 7AW. This replaced the single Trp residue in the cSH3 domain (W275) with 7AW. To gauge the effect of 7AW incorporation into the cSH3 domain, a shorter construct (L230-S304), previously used to solve its structure, was also labeled with 7AW (hereafter referred to as $cSH3_{7AW}$), which was purified as in section 2.2. Following cell lysis, the desired fusion proteins were purified from the soluble fraction by Ni^{2+}-NTA affinity chromatography. The His-tags were then removed by cleavage with Factor Xa protease, which also exposed the N-terminal cysteine residue in the labeled cSH3 domains necessary for EPL. The cleaved proteins were then purified by RP-HPLC with an overall yield of ~1mg/L for $lcSH3_{7AW}$ and ~4 mg/L for $lcSH3_{15N}$ (Figure 2.24).

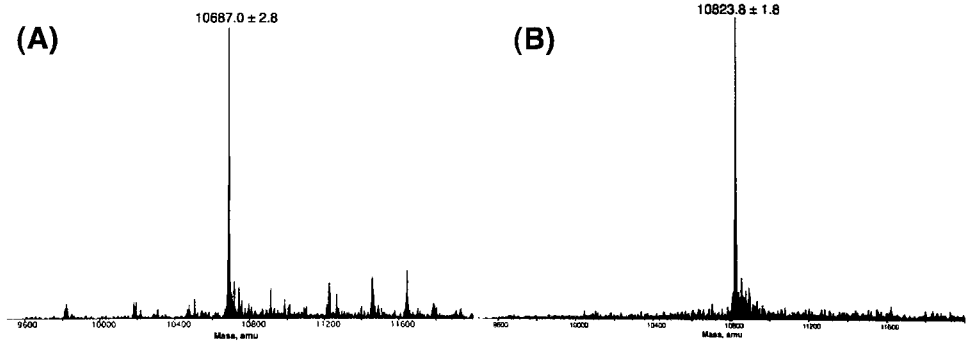

Figure 2.24. ESMS characterization of $lcSH3_{7AW}$ and $lcSH3_{15N}$. (A) The reconstructed mass spectrum of $lcSH3_{7AW}$. The indicated mass was calculated from the charged species seen in the parent ESMS spectrum. Expected mass: 10688.0 Da. **(B)** The reconstructed mass spectrum of $lcSH3_{15N}$. The indicated mass was calculated from the charged species seen in the parent ESMS spectrum. Expected mass: 10822.0 Da.

Crk-II with the cSH3 domain specifically labeled with either ^{15}N or 7AW was assembled via EPL by mixing purified Crk-I *a*-thioester with the labeled cSH3 domain (Figure 2.25). Both reactions were carried out in the presence of 6M guanidinium chloride. The reactions were monitored by RP-HPLC and SDS-PAGE and were both near completion after 7 days (Figure 2.26 and 2.27). The ligation products, Crk-II[IcSH3$_{7AW}$] (Figure 2.26) and Crk-II[IcSH3$_{15N}$] (Figure 2.27), were purified by RP-HPLC and characterized by ESMS. The lyophilized ligated proteins were dissolved in buffer containing 6 M guanidinium chloride and refolded as described previously (Blaschke et al., 2000).

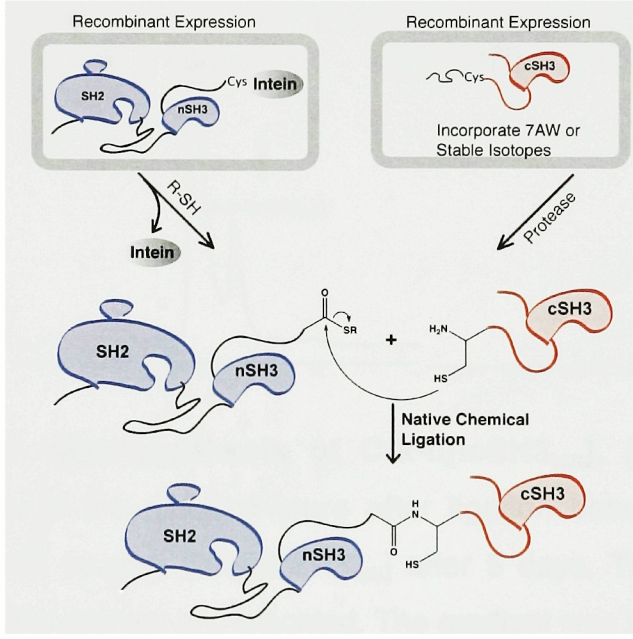

Figure 2.25. Domain specific incorporation of labels into the cSH3 domain of Crk-II. Schematic representation of the strategy to label the cSH3 domain within full-length Crk-II.

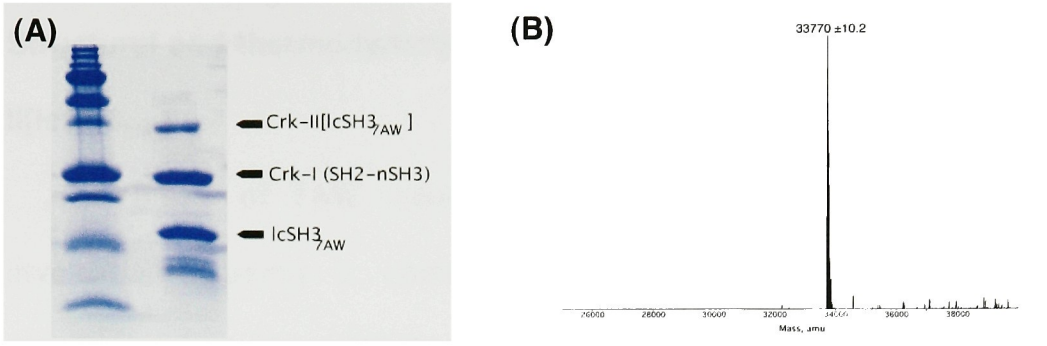

Figure 2.26. Semi-synthesis of Crk-II[IcSH3$_{7AW}$]. (A) SDS-PAGE analysis of the crude mixture after ligation between Crk-I and IcSH3$_{7AW}$ to give Crk-II[IcSH3$_{7AW}$] after 3 days. The reaction was allowed to proceed for 7 days after which it was near completion based on HPLC. The two starting materials and the ligation product are indicated. **(B)** The reconstructed mass spectrum of the ligation product. The indicated mass was calculated from the charged species seen in the parent ESMS spectrum. Expected mass: 33773.6 Da.

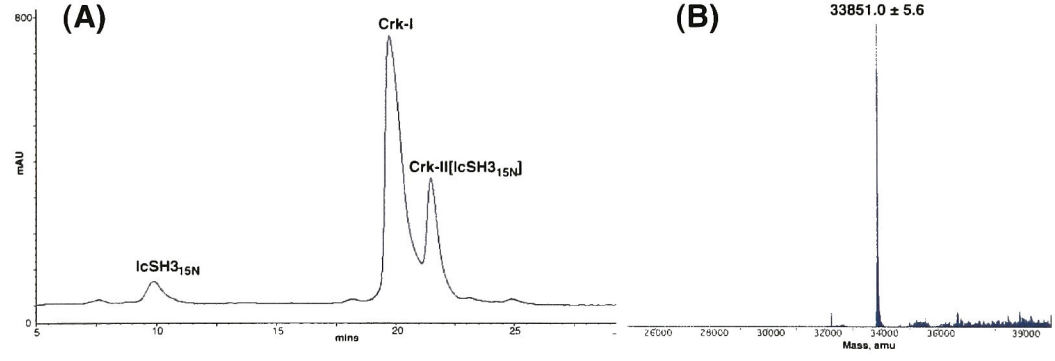

Figure 2.27. Semi-synthesis of Crk-II[IcSH3$_{15N}$]. (A) RP-HPLC chromatogram of the crude mixture after ligation between Crk-I and IcSH3$_{15N}$ domain to give Crk-II[IcSH3$_{15N}$] after 5 days. The two starting materials and the product are indicated. The gradient was 38%-42% buffer B (90% acetonitrile/0.1% TFA in water) **(B)** The reconstructed mass spectrum of the ligation product. The indicated mass was calculated from the charged species seen in parent ESMS spectrum. Expected mass: 33907.6 Da (assuming complete incorporation of ^{15}N).

Structural and thermodynamic characterization of $cSH3_{7AW}$ and Crk-II[$cSH3_{7AW}$]

The effect of 7AW incorporation into the cSH3 domain was investigated using a combination of NMR and fluorescence spectroscopies. The NOESY spectra of unlabeled cSH3 domain and $cSH3_{7AW}$ are similar (Figure 2.28). The α-fingerprint regions of both spectra are well dispersed and indicative of a well-defined protein fold with significant contribution from β-sheets. This suggests that the incorporation of 7AW does not change the overall fold of the cSH3 domain.

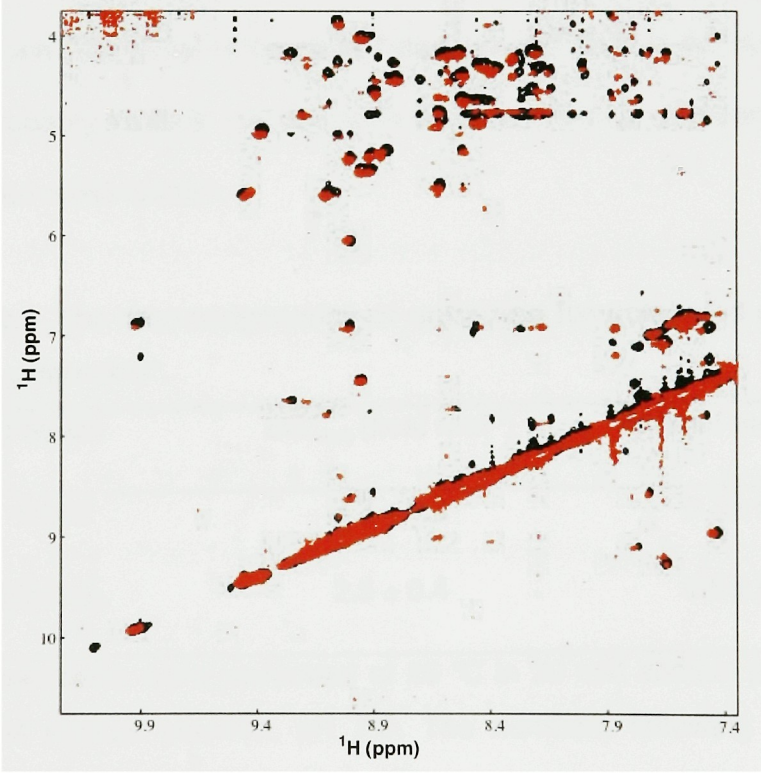

Figure 2.28. Amide and aromatic region of ^{1}H NOESY spectra of cSH3 (in black) **and $cSH3_{7AW}$** (in red) **domains.** The spectra were acquired on a 700 MHz spectrometer.

The equilibrium stability of $cSH3_{7AW}$ was determined using the change in 7AW fluorescence as a function of increasing denaturant concentration. The $cSH3_{7AW}$ domain exhibited a well-defined unfolding curve from which the equilibrium stability (ΔG^0_{H2O}) and the m-value of this domain were elucidated (Table 2.5 and Figure 2.29). These measurements show that the incorporation of 7AW into the cSH3 domain destabilizes the domain by ~ 1 kcal mol^{-1}. Thus, even though the overall structure of the cSH3 domain is not affected by the incorporation of 7AW, there is an energetic cost associated with the incorporation. This is presumably due to the fact that the Trp residue in the cSH3 domain is buried in the hydrophobic core of the domain. The burial of the polar 7-aza group in 7AW in the hydrophobic core probably causes the destabilization of $cSH3_{7AW}$ compared to the unlabeled cSH3 domain.

Table 2.5. The thermodynamics of unfolding for unlabeled cSH3 and $cSH3_{7AW}$ domains[a].

Protein	$\Delta G°_{H2O}$ (kcal mol^{-1})	m (kcal mol^{-1} M^{-1})
cSH3	3.3 ± 0.3	0.7 ± 0.1
$cSH3_{7AW}$	2.5 ± 0.4	0.5 ± 0.1

[a]All experiments were performed at 25 °C in 20 mM sodium phosphate, 100 mM sodium chloride, pH 7.0. The numbers following ± symbol represent the standard errors to the fit.

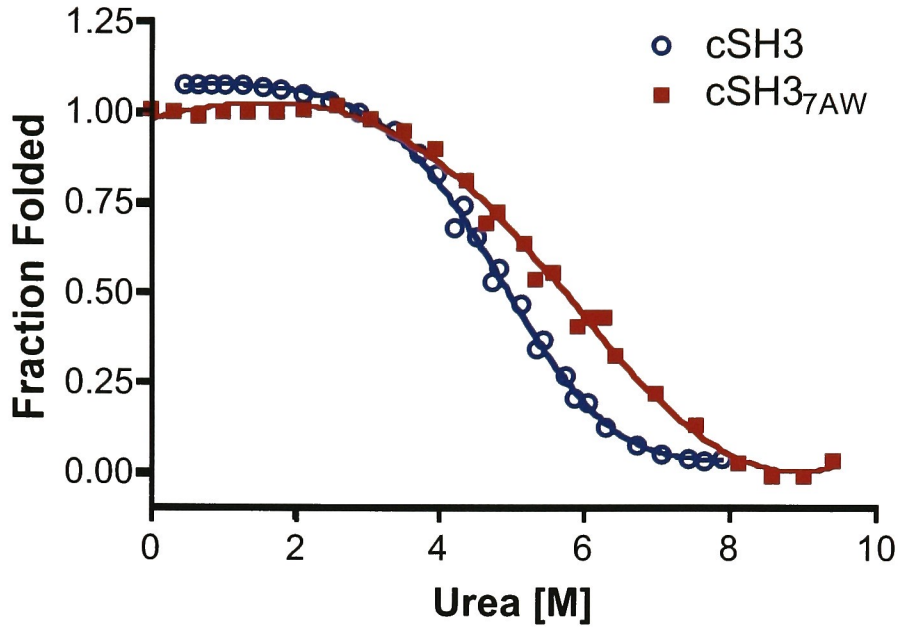

Figure 2.29. Chemical denaturation of unlabeled cSH3 domain and cSH3$_{7AW}$ followed by fluorescence spectroscopy. The solid curve represents the best fit for eqs. 3-4.

The measurement of the equilibrium stability of Crk-II[lcSH3$_{7AW}$] was also attempted by measuring the change in 7AW fluorescence as a function of changing denaturant concentration. However, the 7AW labeled cSH3 domain did not exhibit a well-defined unfolding curve and that precluded any estimation of the equilibrium stability of the cSH3 domain within the context of Crk-II (Figure 2.30).

The fluorescence emission spectra of Crk-II[lcSH3$_{7AW}$] is broader than that of cSH3$_{7AW}$ (Figure 2.31). The emission maxima of cSH3$_{7AW}$ is at 371 nm while that of Crk-II[lcSH3$_{7AW}$] is at 377 nm. At these maxima, the

fluorescence intensity of Crk-II[lcSH3$_{7AW}$] is ~22% less than that of cSH3$_{7AW}$. This strongly suggests that 7AW is present in a different local environment when the cSH3 domain is within the context of Crk-II.

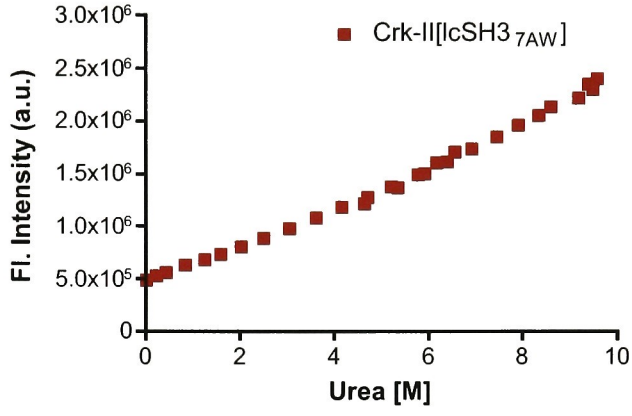

Figure 2.30. The change in 7AW fluorescence of Crk-II[lcSH3$_{7AW}$] as a function of increasing amount of denaturant. This data could not be fit to eqs. 3-4.

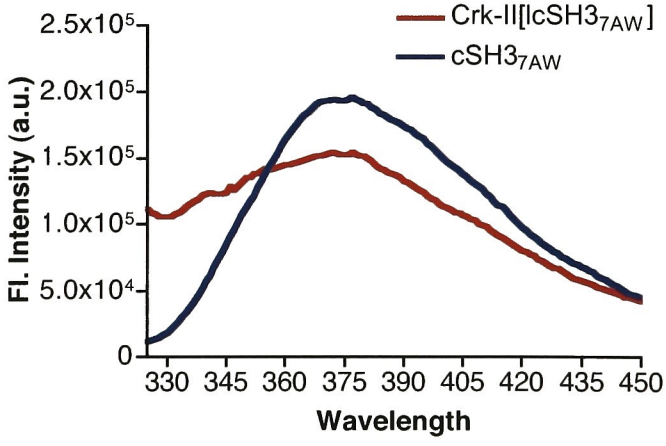

Figure 2.31. Fluorescence emission spectra of cSH3$_{7AW}$ and Crk-II[lcSH3$_{7AW}$]. The excitation wavelength was 310 nm. The concentrations of both proteins were 1 μM (see methods for concentration determination).

Structural characterization of $lcSH3_{15N}$ and Crk-II[$lcSH3_{15N}$]

Previous attempts to find a PXXP ligand for the cSH3 domain have failed (Feller, 2001). This has led to the hypothesis that the cSH3 domain of Crk-II regulates the function of the protein via an intramolecular interaction (Feller, 2001; Ogawa et al., 1994). The Crk-II protein has two proline rich regions, which could interact with the cSH3 domain in an intrasteric fashion (Figure 2.32). To investigate the possibility of an intramolecular interaction, segmental isotopic labeling was utilized, where the cSH3 domain was labeled within the context of full-length Crk-II. NMR spectroscopy was utilized to study the changes in the structure of the cSH3 domain within Crk-II compared to the isolated cSH3 domain.

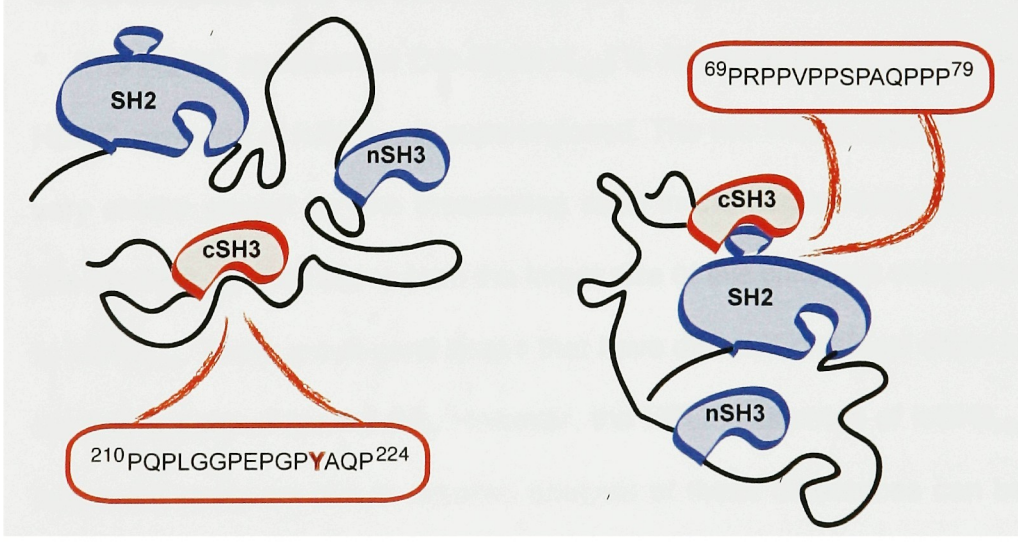

Figure 2.32. Possible intramolecular binding sites of the cSH3 domain within Crk-II. Shown are just two possible binding interactions, assuming a canonical PXXP-type interaction. However, it is possible that the cSH3 domain participates in non-canonical interactions. cAbl phosphorylates Crk-II on Y221 (highlighted in red).

The HSQC spectrum of $lcSH3_{15N}$ is shown in Figure 2.33. Overlaid is the HSQC spectrum of the cSH3 domain without the linker region that was used to solve the solution solution structure of the domain. The difference between these two cSH3 domain constructs is that lcSH3 has an additional 22 residues (C208-P229) from the linker region between the two SH3 domains of Crk-II. The comparison shows that there is some minor perturbation in the core SH3 domain upon addition of the linker. In particular, I269, N270 and T289 show the largest chemical shift perturbation compared to the cSH3 domain. There are also some peaks that had improved line shapes in the presence of the linker, namely A250, K253, S272 and N280. Other than these residues, there is little change in the HSQC spectrum of the cSH3 domain upon addition of the linker.

The HSQC spectrum of Crk-II[$lcSH3_{15N}$] is shown in Figure 2.34. The HSQC spectrum of $lcSH3_{15N}$ is superimposed. The two HSQC spectra look very similar except for line broadening seen in the segmentally labeled Crk-II[$lcSH3_{15N}$]. This arises from the larger size of this construct compared to $lcSH3_{15N}$. There are several peaks that have different chemical shifts in the two spectra (Figure 2.34). However, the HSQC spectrum of $lcSH3_{15N}$ has to be assigned before detailed analysis of these differences can be made. Efforts at assigning the linker residues are underway. Comparing the HSQC spectra of Crk-II[$lcSH3_{15N}$] and the cSH3 domain shows that there are some residues with different chemical shifts between the two

spectra (Figure 2.35). Of note, is the peak assigned to the side chain of the indole ring of W275. This residue is not perturbed by the addition of the linker, but only when the domain is in the context of Crk-II.

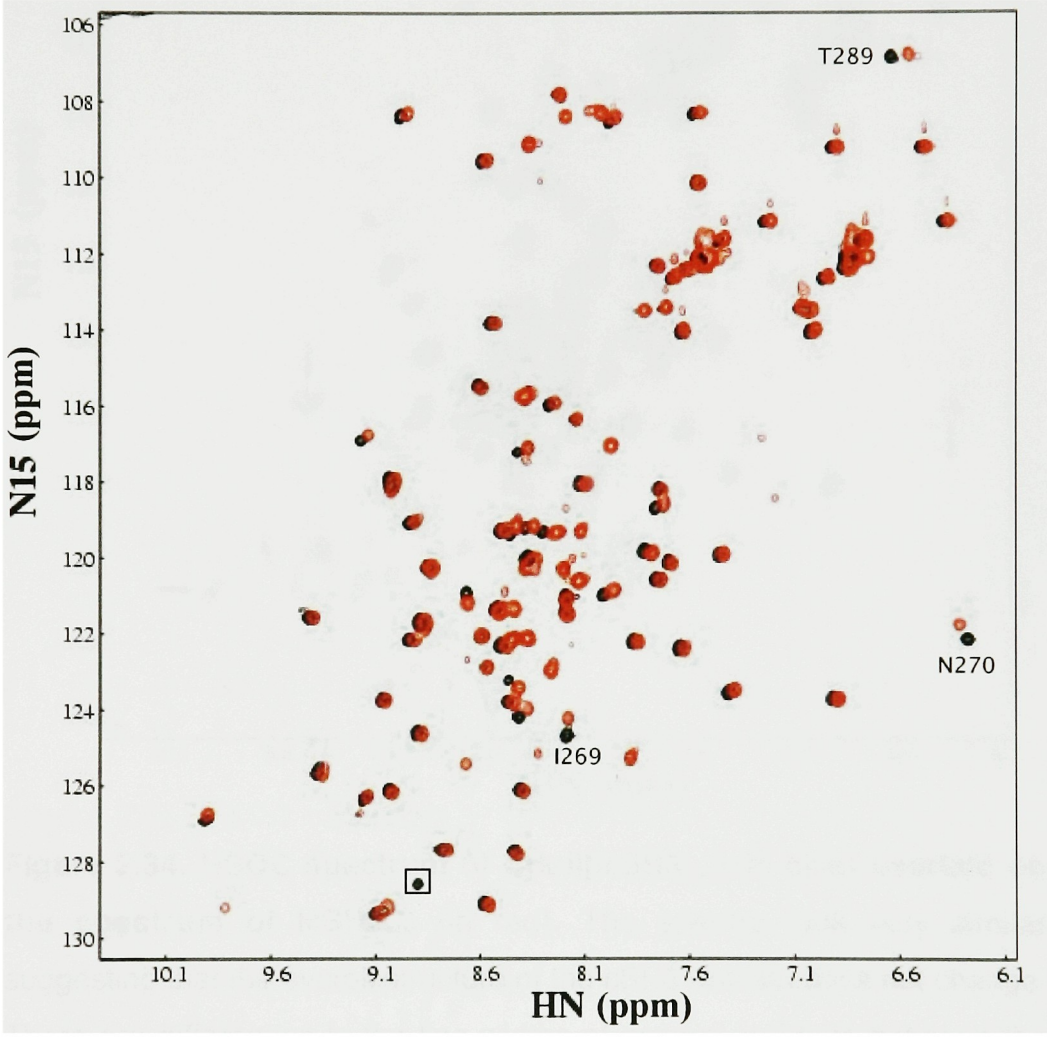

Figure 2.33. HSQC spectrum of $IcSH3_{15N}$ (in red) **overlaid on the spectrum of the cSH3 domain** (in black)**.** The two spectra look very similar and the linker does not perturb the overall structure of the cSH3 domain. The residues that have different chemical shifts in the two spectra are indicated (assigned in spectrum of the cSH3 domain). The peak for G281 (box) is also indicated and is folded out of the spectrum of $IcSH3_{15N}$.

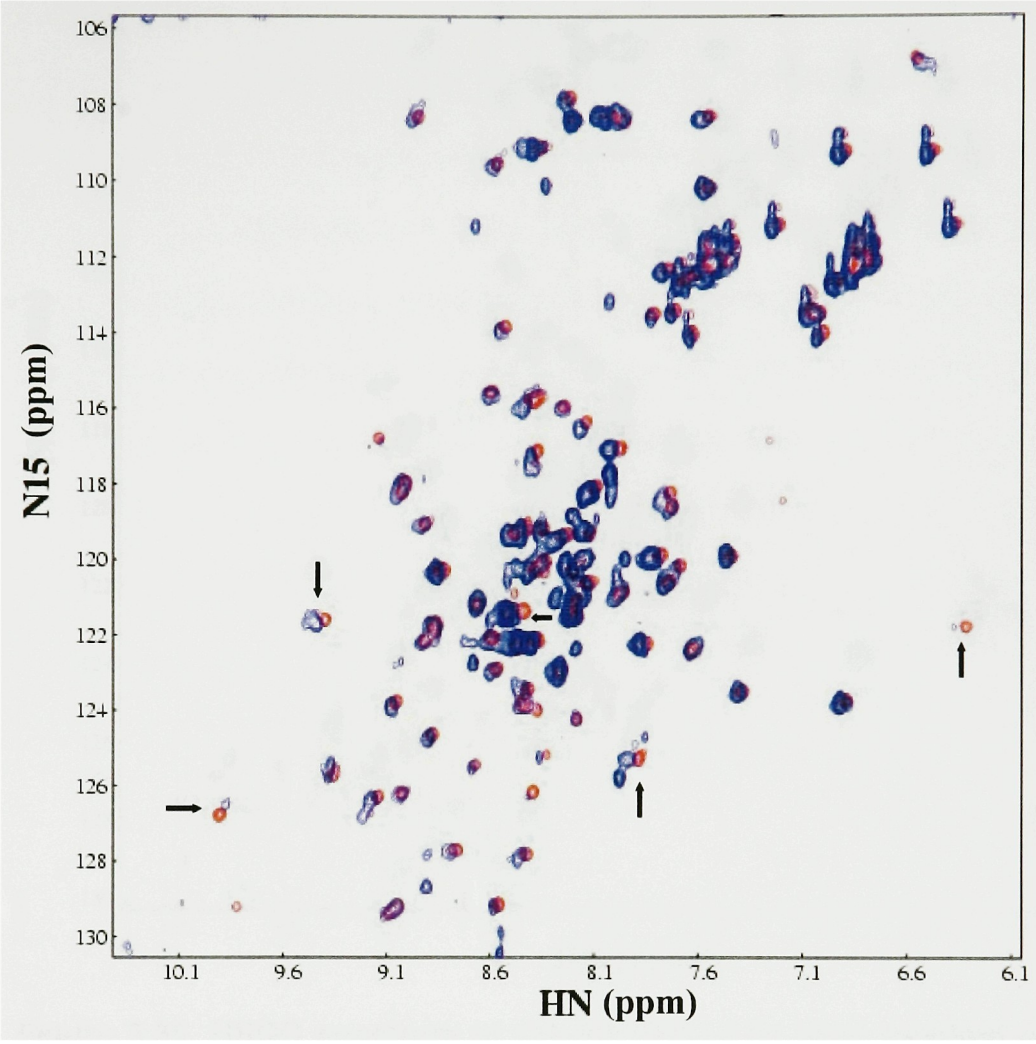

Figure 2.34. HSQC spectrum of Crk-II[lcSH3$_{15N}$] (in blue) **overlaid on the spectrum of lcSH3$_{15N}$** (in red). The spectra look very similar suggesting that the overall structure of the cSH3 domain does not change. There are differences in the line widths of several residues between the two spectra. This arises due to the larger size of Crk-II[lcSH3$_{15N}$]. The peaks that have obviously different chemical shifts in the two spectra are indicated. Since the HSQC spectrum of lcSH3 has not been assigned the nature of these changes cannot be interpreted.

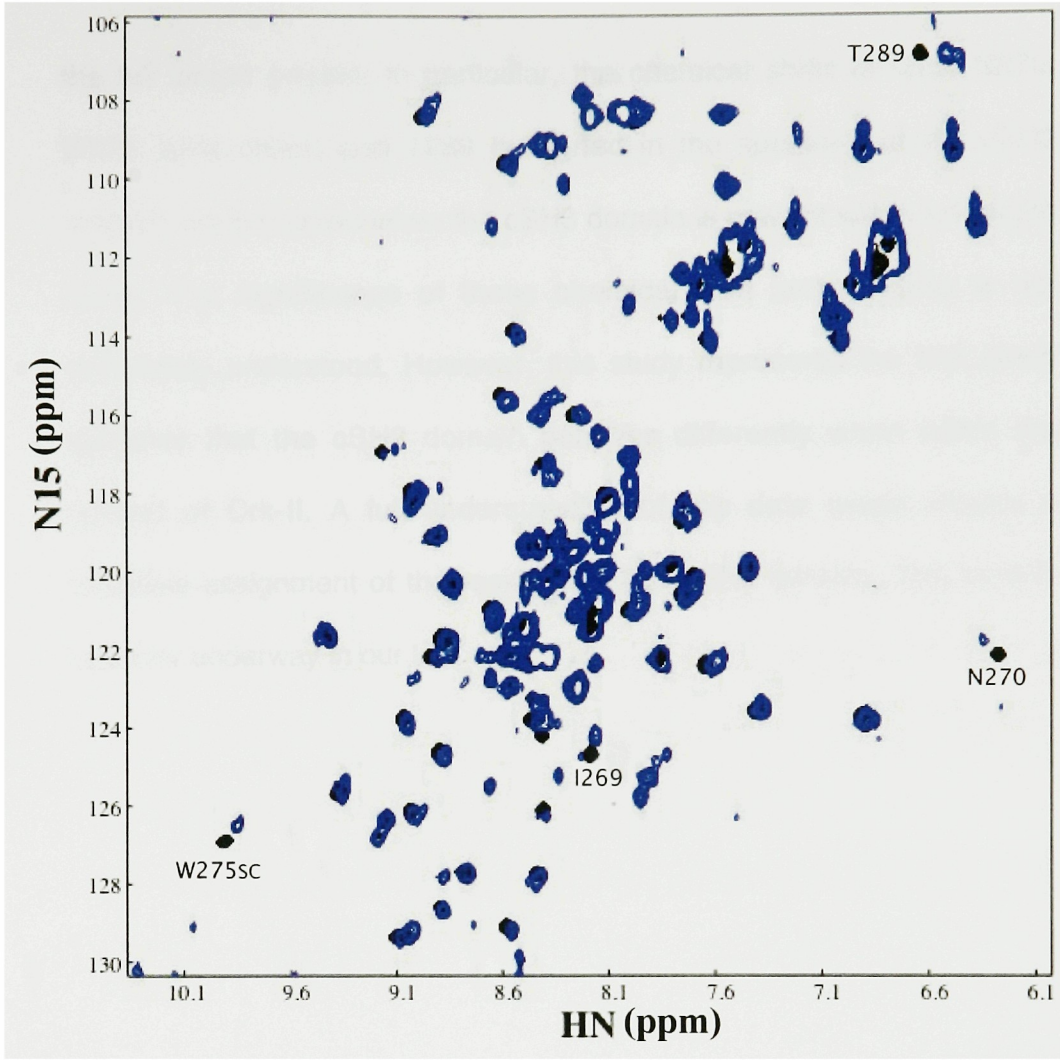

Figure 2.35. HSQC spectrum of Crk-II[IcSH3$_{15N}$] (in blue) **overlaid on the spectrum of the cSH3** (in black). The two spectra look similar except for the differences in line widths. The residues that have different chemical shifts in the two spectra have been indicated. The residues have been assigned in the HSQC spectrum of the cSH3 domain.

The changes between the HSQC spectra of the cSH3 domain, IcSH3$_{15N}$ and Crk-II[IcSH3$_{15N}$] suggests that there are some differences between the domain in isolation and when the domain is in the context of

the full length protein. In particular, the chemical shifts of I269, N270, W275 (side chain) and T289 (assigned in the spectrum of the cSH3 domain) are perturbed when the cSH3 domain is present within full-length Crk-II. The significance of these chemical shift perturbations is not completely understood. However, this study represents the first direct evidence that the cSH3 domain behaves differently when within the context of Crk-II. A full understanding of this data would require a complete assignment of the residues in the lcSH3 domain. This work is currently underway in our lab.

Chapter 3– Discussion

Small modular protein domains have long been used as model systems to understand the biophysical properties of proteins. In particular, studies on SH3 domains have greatly extended our knowledge of the principles underlying protein folding and structure. However, very little is known about how their biophysical properties are affected when the domain is part of a larger multi-domain protein. This is important because modular domains like the SH3 domain are always a part of larger multi-domain proteins. Thus a detailed understanding of how the functional interplay between domains affects their biophysical properties is essential. The lack of a generally applicable spectroscopic technique that allows the extraction of biophysical information from a specific domain within a multi-domain protein has hindered this sort of detailed analysis.

Domain Specific Incorporation of Optical Probes

Modifying proteins in a site-specific manner with unnatural amino acids, particularly ones that can be used as spectroscopic probes, is of considerable interest to the protein chemistry community. The introduction of Trp analogues into recombinant proteins expressed in *E. coli* Trp auxotrophs has proven a particularly useful strategy for uniformly labeling proteins with spectroscopic probes (Mohammadi et al., 2001; Ross et al., 2000; Shen et al., 1996; Soumillion P, 1995; Twine and Szabo, 2003;

Wong and Eftink, 1998a; Wong and Eftink, 1998b). Current approaches do not, however, allow one to incorporate Trp analogues into proteins containing multiple tryptophans in a site-specific or domain-specific manner. This limitation precludes the extraction of domain specific biophysical information for large multi-domain proteins (Sato et al., 1999; Sato et al., 2000). In this study, this problem was addressed by integrating *in vivo* Trp labeling and protein ligation techniques. This has allowed the domain specific incorporation of an optical probe, 7AW, into the SH3 domain of the adaptor protein, c-Crk-I.

The SH3 domain of c-Crk-I is small, like most other SH3 domains, and has been studied extensively with regards to its biochemical (Knudsen et al., 1994), structural (Wu et al., 1995) and biophysical (Camarero et al., 2001) properties. Its folding properties are similar to other SH3 domains such as the src SH3 and α-spectrin SH3 domains (Camarero et al., 2001). These attributes make the c-Crk-I SH3 domain an attractive model system to explore the effect of context on the functional and thermodynamic properties of domains.

Incorporation of 7AW into the SH3 domain of Crk-I

There are two Trp residues (Trp169 and Trp170) in this domain that we chose to replace with 7AW by bacterial expression in a suitable *E.coli* Trp auxotrophic strain. 7AW incorporation was found to be efficient (≥93%) at

both sites as indicated by mass spectrometric analysis of tryptic peptides (Figure 2.3). Such high efficiency has been seen before in proteins with two or more Trp residues, though the exact reason for this remains unknown (Senear et al., 2002). Analysis of the structure of the c-Crk-I SH3 domain reveals that the 7-position of both Trp residues points out into solution. Hence, the incorporation of 7AW was not expected to cause structural perturbation in the system (Figure 2.1). Indeed, CD and homonuclear ^{1}H NMR spectroscopies indicate that the $SH3_{W}$ and $SH3_{7AW}$ domains have the same global fold (Figure 2.6, 2.8 and 2.9). This conclusion is also supported by thermal (Figure 2.7) and chemical (Figure 2.11) denaturation experiments, which indicate that the labeled and unlabeled proteins have similar thermodynamic stabilities (~3 kcal mol^{-1}) (Table 2.2). Lastly, the dissociation constants of the $SH3_{W}$ and $SH3_{7AW}$ domains, for the peptide ligand used in this study, are essentially the same (Figure 2.10 and Table 2.1). Taken together, these results strongly suggest that the structural changes caused by the incorporation of 7AW into the SH3 domain of c-Crk-I are very minor, if any.

In contrast, incorporation of 7AW into the cSH3 domain of Crk-II has an energetic cost of ~1 kcal mol^{-1} (Table 2.5 and Figure 2.29). Interestingly, the m-value, which is proportional to the change in solvent exposed surface area upon unfolding, decreases when both SH3 domains are labeled with 7AW. However, the m-value of the cSH3 domain

decreases by ~27% while that of the Crk-I SH3 domain decreases by only ~12% when 7AW is incorporated in these domains. One might expect the decrease to be greater in the Crk-I SH3 domain, where two Trp residues are replaced by 7AW. One possible reason for this unexpected result is that when 7AW in incorporated into the cSH3 domain, it increases the solvent accessible surface area in the native state relative to the unfolded state. This might not occur in the Crk-I SH3 domain, where both 7-aza groups point out into solution. It is also possible that we are populating a hidden folding intermediate, which would also explain the larger decrease in the m-value. However, the second scenario seems unlikely since we have no spectroscopic evidence of any folding intermediates. In the cSH3 domain the single Trp residue is buried and the 7-position points into the hydrophobic core (Figure 2.25). However, the overall fold of the cSH3 domain is maintained when labeled with 7AW (Figure 2.28). In the case of staphylococcal nuclease, this kind of destabilization has also been observed when the two Trp residues in that protein are replaced with 7AW (Wong and Eftink, 1998a; Wong and Eftink, 1998b). The 7-positions of both Trp residues in the nuclease are also buried in the hydrophobic core of the protein. This underscores the need for careful thermodynamic and structural analysis before labeling a protein with 7AW. In particular, it depends on where the 7-position is in the structure of the protein. The recent availability of other azatryptophan analogues such as 6AW (Twine

and Szabo, 2003) should allow us to choose a label that may cause the least perturbation based on the structure of the protein.

Expressed protein ligation was successfully used to prepare Crk-I[$SH3_{7AW}$] from its component domains. The ligation reaction was extremely efficient (Figure 2.2) allowing multi-milligram amounts of the labeled protein to be isolated. Although not relevant for synthesis of Crk-I[$SH3_{7AW}$], our studies also indicate that the GyrA intein is functional when its lone Trp residue is replaced by 7AW (Figure 2.4). Importantly, this allows the preparation of α-thioester proteins containing the Trp analogue. Thus, our domain specific labeling strategy will be applicable to N-terminal, C-terminal, and even internal (Cotton et al., 1999) domains of large proteins.

Biophysical Properties of the SH3 domain within Crk-I

Access to large amounts of Crk-I[$SH3_{7AW}$] has allowed us to investigate the effect of context on the thermodynamic and functional properties of the SH3 domain by exploiting the spectral characteristics of the 7AW probe. These properties do not change dramatically when the domain is present in the Crk-I protein (Tables 2.1 and 2.2). Interestingly, the emission maxima for $SH3_{7AW}$ and Crk-I[$SH3_{7AW}$] differ by around 13 nm (Figure 2.5). The blue-shift observed in the Crk-I[$SH3_{7AW}$] spectrum suggests that 7AW is present in a slightly different environment (perhaps more hydrophobic)

in the context of Crk-I[$SH3_{7AW}$] compared to the isolated domain. Conceivably, there might be some interaction between the SH2 and the SH3 domains in Crk-I, though this does not have a significant effect on either the ligand binding or equilibrium stability of the SH3 domain. Thus, these interactions, if present are very weak at best.

The SH2 domain of Crk-I is extremely stable (Table 2.2). Indeed, the free energy of unfolding is significantly higher than has been reported for other SH2 domains (ΔG^0_{H2O} ~3 kcal mol^{-1}) (Scalley-Kim et al., 2003; Tzeng et al., 2000). As can be seen in Figure 2.11, the SH2 domain remains fully folded at intermediate denaturant concentrations at which the SH3 domain is completely unfolded, e.g. 2 M GdmCl. This implies that in the context of Crk-I[$SH3_{7AW}$], the SH3 domain is unfolding while tethered to the presumably folded SH2 domain. Interestingly, this has no effect on the thermodynamics of the SH3 domain. Of course, it is possible that the stability of the SH2 domain changes when in the context of Crk-I. However, this seems unlikely since there are only two domains in the protein; any effect caused by one domain on the other should be reflected in the equilibrium stability of both domains. Thus we can conclude that the thermodynamic properties of the SH3 domain remain unchanged regardless of the presence of a fully folded SH2 domain physically linked to the SH3 domain. These equilibrium denaturation studies suggest that in c-Crk-I, where the domains are linked via flexible non-structured linker

regions, the domains fold independently of each other.

The Signaling Adaptor Protein Crk-II

Crk-II is an alternative spliced form of the Crk-I protein and contains an SH2 domain followed by two SH3 domains (nSH3 and cSH3). Crk-II has been implicated in several cellular processes like proliferation, motility, phagocytosis and adhesion (Abassi and Vuori, 2002; Akakura et al., 2005; Hasegawa et al., 1996; KizakaKondoh et al., 1996; Lamorte et al., 2002a; Lamorte et al., 2002b; Reddien and Horvitz, 2000; Sun et al., 2005; Zvara et al., 2001) (Figure 1.9). The major difference between Crk-I and Crk-II is that, the latter has a C-terminal SH3 domain (cSH3) that is not present in the former. The cSH3 domain is hypothesized to negatively regulate the function of Crk-II via an intramolecular binding event (Feller, 2001; Ozawa and Umezawa, 2001). However, understanding the role of the cSH3 domain in regulating the function first requires structural characterization of this domain.

Solution Structure of the cSH3 domain

Sequence homology analysis lacks the ability to predict domain boundaries with accuracy. Therefore, we used a slightly longer cSH3 domain construct (L230 to S304) than that predicted by homology to solve the NMR structure of the domain. This domain was found to be

monomeric at concentrations required for NMR spectroscopy (0.3-0.4 mM). Steady state ^{15}N-{^{1}H} heteronuclear Overhauser effect (NOE) measurements (Figure 2.14) and chemical shift indexing data (Figure 2.13) showed that there are five β-strands in the cSH3 domain. This data also showed that the domain boundaries predicted by sequence homology (P237-Q296) were actually quite accurate. This allowed us to exclude residues L230 to N235 at the N-terminus and residues P299 to S304 at the C-terminus from final structure calculations. In the present work, we have found that the cSH3 domain of Crk-II has the canonical five-stranded β-barrel SH3 structure (Figure 2.15). We have also characterized the thermodynamic stability and kinetic folding properties of the cSH3 domain (Figures 2.22 and 2.23). This allowed us to better understand some of the mutational studies performed on Crk-II. The folding properties of the cSH3 domain are also comparable to other SH3 domains.

Ligand Binding Surface of the cSH3 Domain

In spite of the low sequence conservation, the majority of the SH3 domains bind to proline-rich sequences (PXXP) via key conserved aromatic residues that form a contiguous hydrophobic patch on the surface of the protein. Burial of these hydrophobic residues accounts for the moderately high affinity (usually low μM) associated with the PXXP binding interaction; ligand binding typically results in the burial of 800-

1200Å^2 of surface area, most of which is hydrophobic in nature (Ghose et al., 2001; Lee et al., 1996). The specificity of ligand binding arises out of ancillary interactions usually involving polar or charged residues flanking the core PXXP sequence (Ghose et al., 2001; Wu et al., 1995). By contrast, these conserved hydrophobic residues are replaced by non-aromatic and mostly polar residues in the cSH3 domain (Figure 1.1 and Figure 2.18). Indeed, our structure shows that these residues form a continuous surface that corresponds to the ligand-binding surface on other SH3 domains (Figure 2.18). Consequently, the energy derived from the burial of hydrophobic residues upon ligand binding (Clackson and Wells, 1995) is no longer available in the case of the cSH3 domain. This suggests that the cSH3 domain probably does not bind to PXXP sequences with high affinity, assuming the interaction is canonical. It is also possible that the domain binds to other sequences, as has been observed in several SH3 domains including the Pix SH3 domain (Manser et al., 1998), the Eps8 SH3 domain (Mongiovi et al., 1999) and the Hbp SH3 domain (Kato et al., 2000). It is therefore not surprising that efforts to find a PXXP ligand for this domain have not yet been fruitful (Feller, 2001).

The Putative Nuclear Export Sequence in the cSH3 Domain

The only known interaction involving the cSH3 domain is with the nuclear export factor, Crm1 (Smith et al., 2002). Affinity pull-down assays

indicate that the Crk-II-Crm1 interaction is dependent upon a putative NES within the cSH3 domain (L256-V266). Several residues in this sequence are part of the hydrophobic core of the domain, namely, L256, L258, V264 and V266 (Figure 2.17). Recognition of this NES by Crm1 would therefore require unfolding of the cSH3 domain to expose the sequence. We determined the free energy of unfolding (ΔG^0_{H2O}) of the cSH3 domain to be 3.66 kcal mol^{-1} (Figure 2.21). Based on this equilibrium stability, Crm1 would have to bind to the NES with low nanomolar affinity to compensate for the free energy loss associated with the unfolding of the domain.

Crm1 is thought to bind NES-containing cargoes and RanGTP in a cooperative manner with affinities in the nanomolar range (Petosa et al., 2004). Thus, there could be enough binding energy associated with a Crm1-RanGTP-Crk ternary complex to maintain the cSH3 domain in an unfolded state (Petosa et al., 2004). This is an intriguing idea and there are certainly examples of binding events that maintain proteins in the unfolded or partially folded state (Haslbeck et al., 2005; Spiess et al., 2004). At the same time, it is worth considering the possibility that Crm1 actually binds to the surface of the folded cSH3 domain. Indeed, the principle line of evidence for the involvement of an NES in the CRM1-cSH3 interaction comes from the inability of Crm1 to bind a cSH3 mutant in which residues 263, 264 and 266 are replaced with alanine. As noted above, V264 and V266 are part of the hydrophobic core of the protein and

thus the native fold of this triple mutant would be expected to be dramatically destabilized. Consequently, any binding interaction requiring the native SH3 fold would also be weakened. Although Crm1 does not contain any PXXP motifs, there are several examples of SH3 domains that bind to non-canonical ligands (Kato et al., 2000; Manser et al., 1998; Mongiovi et al., 1999) and given the unusual nature of the binding pocket in cSH3, such a possibility cannot be discounted in this case. Clearly, further biochemical and structural analysis of the Crm1-cSH3 interaction is merited. We have also made mutants of the cSH3 domain where L263, V264 and V266 have been replaced with alanine. The investigation of the thermodynamics and the potential Crm1 interaction of these mutants are ongoing.

Mutational Studies on the cSH3 Domain

A recent study also alludes to the possibility that the cSH3 domain might be involved in the activation of the Abl kinase by Crk-II (Reichman et al., 2005). Abl kinase is known to interact with both the SH2 domain and the nSH3 domain of Crk-II (Anafi et al., 1996; Donaldson et al., 2002; Ren et al., 1994). The authors of this study have identified a conserved sequence in the RT-loop of the cSH3 domain ($P^{248}NAY^{251}$) that might play a role in the activation of Abl kinase (Reichman et al., 2005). However, deletion of this sequence alone does not have an effect on Abl kinase

activation (Reichman et al., 2005). Only when combined with other mutations in Crk-II, does the deletion of PNAY sequence show an inhibitory effect on Abl kinase (Reichman et al., 2005). In the structure of the cSH3 domain, the PNAY sequence points into solution and so is available for potential interactions. Several recent studies have shown that mutating the single tryptophan residue in the cSH3 domain can lead to the disruption of the normal function of Crk-II (Akakura et al., 2005; Sun et al., 2005). This tryptophan is an integral part of the hydrophobic core of the domain (Figure 2.20). The fluorescence of this tryptophan residue shows a robust decrease upon the addition of denaturants, which allowed us to follow the folding of this domain. We also replaced the tryptophan residue with 7AW, which destabilized the cSH3 domain by ~1 kcal mol^{-1}, about a 25% decrease in the overall free energy of folding (Figure 2.29 and Table 2.5). This strongly suggests that the mutation of the tryptophan residue unfolds the domain, which leads to the observed deregulation of the function of Crk-II (Akakura et al., 2005; Sun et al., 2005). We have also mutated this tryptophan to alanine and the thermodynamic characterization of this mutant is ongoing. We are also using this mutant to investigate the putative Crm1-cSH3 domain interaction.

Regulation of Crk-II by the cSH3 Domain

Several lines of evidence indicate that the cSH3 domain acts to negatively regulate the function of Crk-II. Specifically, mutations that disrupt the structure of the cSH3 domain interfere with the regulation of Crk-II leading to an abrogation of phagocytosis (Sun et al., 2005) and endocytosis (Akakura et al., 2005). The mechanism by which the cSH3 domain regulates Crk function is unknown. There are no known PXXP-type binding partners for this domain, despite considerable efforts to identify such ligands. As a consequence, several groups have speculated on the possibility of intrasteric regulation (Feller, 2001; Ogawa et al., 1994). In this scheme, the cSH3 domain would participate in an intramolecular interaction with another region of the protein, perhaps involving the proline-rich sequences within either the SH2 domain or the extended linker between the two SH3 domains.

Several SH3 domains are known to bind ligands in an intramolecular fashion (Andreotti et al., 1997; Barila and Superti-Furga, 1998; Groemping et al., 2003) and one of the factors governing self-association is the affinity of the SH3 domain for the ligand - lower affinities have been shown to favor intramolecular binding events (Laederach et al., 2003). Thus, it is intriguing to note that the canonical ligand-binding surface of the cSH3 domain is atypical, being much less hydrophobic and hence likely to lead to reduced affinity for PXXP sequences (Figure 2.18). An unexpected

feature of the cSH3 domain was the basic nature of its surface (Figure 2.19). Interestingly, the nSH3 domain of Crk-II has an acidic surface (Figure 2.19). Such a distribution of charges has not been seen in proteins with multiple SH3 domains for which structures are available. This poses obvious questions about an interaction between the two SH3 domains of Crk-II.

Domain Specific Labeling to Explore The Mechanism of Regulation of Crk-II by the cSH3 Domain

Domain specific labeling methods were utilized to explore the possibility of an intramolecular interaction mediated by the cSH3 domain. The cSH3 domain was labeled with 7AW and ^{15}N within the context of full length Crk-II (Figures 2.26 and 2.27). Labeling of the cSH3 domain with 7AW replaces the single tryptophan (W275) residue in the domain (Figure 2.23). Although the difference between 7AW and tryptophan is minor, this replacement does destabilize the cSH3 domain by ~1 kcal Mol^{-1}, a 25% decrease in the free energy of folding (Table 2.5 and Figure 2.29). The destabilization is probably the result of burial of the polar 7-aza group into the hydrophobic core of the domain. The overall fold of the $cSH3_{7AW}$ domain is maintained despite the energetic cost (Figure 2.28). This underscores the plasticity of the hydrophobic core of SH3 domains.

Thermodynamic analysis of Crk-II[$SH3_{7AW}$] was attempted to analyze

the energetics involved in the putative interaction between the cSH3 domain and the rest of Crk-II (Figure 2.30). However, the change in 7AW fluorescence as a function of increasing amounts of denaturant did not result in a curve that could be analyzed. It is possible that under different pH, salt or temperature conditions we will get a curve that will allow thermodynamic analysis of Crk-II[$SH3_{7AW}$]. Ongoing work in the Muir laboratory should address these issues. The fluorescence emission of 7AW is very sensitive to its local environment. The steady-state fluorescence spectrum of Crk-II[$SH3_{7AW}$] shows a ~22% decrease in intensity compared to that of $cSH3_{7AW}$ (Figure 2.31). Thus, we can conclude that 7AW is present in a different environment in the $cSH3_{7AW}$ domain compared to Crk-II[$cSH3_{7AW}$].

The HSQC spectra of cSH3, $lcSH3_{15N}$ domains and Crk-II[$lcSH3_{15N}$] are similar suggesting that the fold of the cSH3 domain is not changed by context (Figures 2.33–2.35). However, there are some key differences between the spectra. There are several residues with different chemical shifts in the $lcSH3_{15N}$ domain compared to Crk-II[$lcSH3_{15N}$]. Analysis of this data requires assignment of the HSQC spectrum of the linker region in $lcSH3_{15N}$. There are a few residues that show an obvious chemical shift differences in the cSH3 domain compared to $lcSH3_{15N}$ and Crk-II[$lcSH3_{15N}$] (Figures 2.33, 2.35 and 3.1). Only one of these residues (T289) is in the canonical ligand binding pocket of the cSH3 domain (Figure 3.1).

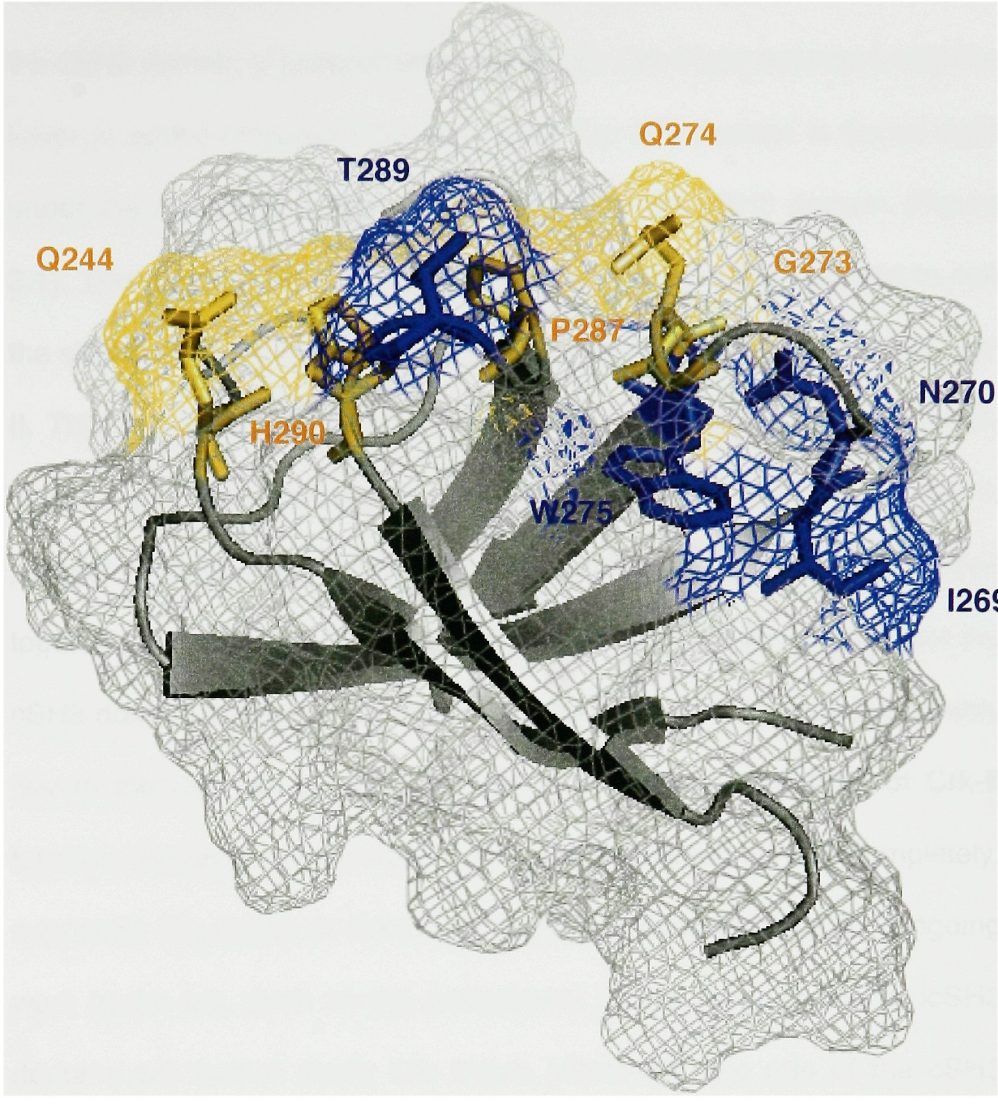

Figure 3.1. Chemical shift perturbation data mapped on the structure of the cSH3 domain. The residues that show obvious chemical shift perturbation in the cSH3 domain compared to the domain within Crk-II are indicated in blue. The residues that form the canonical PXXP-type ligand binding pocket are shown in orange. T289 is the only residue that is part of the canonical ligand binding pocket and also shows chemical shift perturbation when the cSH3 domain is within the context of full-length Crk-II.

Interestingly, the chemical shift of the Trp residue is perturbed only when the cSH3 domain is present within full length Crk-II and not when only the linker is added (Figures 2.33 and 2.35). Trp 275 residue is buried right under the canonical ligand binding pocket of the cSH3 domain (Figure 3.1). This suggests that the context of full-length Crk-II is required to see the effects of an interaction between the cSH3 domain and the rest of Crk-II. This Trp residue in the cSH33 domain was replaced by 7AW in full-length Crk-II. The fluorescence emission spectra of 7AW differ when it is present in the cSH3 domain and when the domain is within Crk-II. Taken together, fluorescence and NMR spectroscopies strongly suggest that the cSH3 domain interacts with the rest of Crk-II. This interaction is probably key to the role of the cSH3 domain as a negative regulator of Crk-II function. However, we still need to map the interaction surface completely, especially the region in Crk-II that the cSH3 domain binds to. Ongoing work in the lab, such as the assignment of the linker region of lcSH3 domain, will further clarify this issue. Ultimately, the role of the cSH3 domain in Crk-II regulation will be best addressed by high-resolution structural and thermodynamic analysis on the full-length protein.

Conclusions

We have developed an approach that allows domain-specific incorporation of tryptophan analogues within modular, multi-domain proteins. In the present example, the methodology was used to probe the effect of context on the biophysical properties of the SH3 domain of c-Crk-1. Comparison of the ligand binding and thermodynamic properties of this domain in isolation versus in the context of c-Crk-1 indicates that there are few, if any, inter-domain interactions in the protein.

We have also solved the solution structure and characterized the folding properties of the cSH3 domain of Crk-II. This has led to a better understanding of the results of several mutational studies on this domain and how it might regulate Crk-II. Domain specific labeling techniques were also applied to generate a 7AW labeled and isotope labeled cSH3 domain within full length Crk-II. Studies on these segmentally labeled molecules have offered us several clues as to how the cSH3 domain might regulate Crk-II. These studies have provided the strongest evidence yet that the cSH3 domain interacts with the rest of Crk-II. Ongoing work in the lab will further clarify the mechanism of regulation of Crk-II by the cSH3 domain.

The techniques developed here should be applicable to many other multi-domain proteins, thereby providing the means to dissect the biophysical properties of these complex systems.

Chapter 4 – Materials and Methods

General Methods

All amino acid derivatives and resins were purchased from Novabiochem (San Diego, CA) and Peninsula Laboratories (Belmont, CA). All other chemicals were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO). Restriction endonucleases were purchased from New England Biolabs (Beverly, MA). Analytical RP-HPLC was carried out using a Hewlett-Packard 1100 series instrument with 214 nm and 280 nm detection. The column used for analytical RP-HPLC runs was a Vydac C18 column with a constant flow rate of 1mL/min. Preparative RP-HPLC was performed on a Waters DeltaPrep 4000 system using a Vydac C18 semi-preparative column at a flow rate of 5 mL/min. All runs used a linear gradient of buffer A (0.1% (v/v) aqueous TFA) versus buffer B (0.1% TFA (v/v), 90% (v/v) acetonitrile). Electrospray mass spectrometry (ESMS) was performed on a Sciex API-100 single quadrupole electrospray mass spectrometer. MALDI-ToF MS was performed on either a Voyager-DE STR instrument (PE Biosystems, Foster City, CA) or a modified Sciex prototype QqTOF (Centaur) instrument (Krutchinsky AN, 2000). These instruments are equipped with nitrogen lasers delivering pulses of ultraviolet light (wavelength 337 nm) at 3 and 20 Hz respectively to the sample spot.

Cloning and Protein Expression

Recombinant SH2 (residues 1-124) ethyl α-thioester was prepared as previously described (Blaschke et al., 2000). The sequence encoding the Crk-I SH3 domain (residues C125 to G207, note that the numbering refers to the full-length murine c-Crk sequence) was isolated from Crk-II cDNA by PCR using a 5'-primer (5'- GCCGCGCTCCAGATAGAAGGGCGGTG CGGAGTGATTCTCAGGCAGGAGGAGGCAGAG-3') encoding a *Xho*I restriction site and a 3'-primer (5'- GCGCCGAACTAGTGCATCTCCCG TGATGCAACCCTCCTGGTTACCTCCAATCAGAGCCG-3') encoding a *Spe*I restriction site. This PCR product was cloned into the plasmid pTXB1 (New England Biolabs) to make the plasmid pTXB1-Xa-SH3-GyrA-CBD, which encodes a Factor Xa site N-terminal to the SH3 domain and Gyrase A intein-Chitin Binding Domain (CBD) tag C-terminal to the SH3 domain. The plasmid, pTXB1-Xa-SH3-GyrA-CBD is incompatible with 7AW labeling primarily because it uses the T7 RNA polymerase for protein expression, which is inactive when labeled with 7AW (Ross et al., 1997). Therefore, the sequence encoding the SH3-GyrA intein fusion was isolated from this plasmid by PCR using the same 5'-primer as when constructing pTXB1-Xa-SH3-GyrA-CBD and a 3'-primer (5'- GCGCCGAAGCTTCTAAGCGTGGCTCACGAACCCGTTCGTGATAAAC GC-3') encoding a *Hind*III restriction site. The PCR product was cloned

into the plasmid pTrcHis (Invitrogen), which is compatible with 7AW labeling, to make pTrcHis-Xa-SH3-GyrA, which encodes a Factor Xa site between a poly-His tag and the SH3-GyrA intein fusion. We also constructed another plasmid, pTrcHis-Xa-SH3, that encodes a stop codon between the SH3 domain and GyrA intein, using site-directed mutagenesis (Qiagen). All the plasmids were shown to be free of mutations in the protein encoding region by DNA sequencing.

For expression of the unlabeled SH3 domain, $SH3_W$, the *Escherichia Coli* Trp auxtrophic strain CY15077 (kindly provided by the Yale University Genetic Stock Center (Yanofsky and Horn, 1995)) transformed with pTrcHis-Xa-SH3, was grown to mid-log phase in Luria-Bertani medium containing 100 μg/mL ampicillin. Protein expression was induced with 1 mM IPTG at 37 °C for 6 hours, after which cells were harvested and lysed by passage through a French press. The His-tagged fusion protein was purified by affinity chromatography using Ni^{2+} NTA resin (Novagen) and dialyzed into Factor Xa cleavage buffer (100 mM sodium phosphate at pH 7.2, 100 mM NaCl, and 1 mM $CaCl_2$). The fusion protein was then cleaved with Factor Xa (Amersham Biosciences) using 100 units of protease/mg of protein, over 8-10 hours. The reaction was quenched by the addition of 20 mM DTT and the cleaved protein further purified by semi-preparative RP-HPLC using a linear gradient of 25%-45% solvent B over 60 minutes. The purified protein was characterized by ESMS as $SH3_W$ (Expected mass:

9429.5 Da, Observed mass: 9430.4 ± 3.2 Da). The overall yield of $SH3_W$ domain was 10 mg/L. Protein concentration was determined by UV absorption (λ =280 nm); $SH3_W$ domain, ε = 15400 M^{-1} cm^{-1}.

For expression of the labeled SH3 domain, $SH3_{7AW}$, an overnight culture of the *E.coli* CY15077 cells transformed with the plasmid pTrcHis-Xa-SH3 was diluted 1:20 into 6 litres of Luria-Bertani (LB) broth containing 100 μg/mL ampicillin. The cells were grown at 37 °C to an A_{600} of 0.9-1.0, harvested and then washed with M9 minimal media supplemented with 0.1 mM $CaCl_2$, 1 mM $MgSO_4$, 0.5% glucose, 0.1% thiamine, 100 μg/mL ampicillin and 0.6% glycerol. The culture was then resuspended in 6 L of M9 minimal media supplemented as above and allowed to grow for another 40 minutes to deplete the cellular stores of Trp. Protein expression was then induced with the addition of 1 mM IPTG and 80 μg/mL of D, L-7AW (Sigma) and the cells were harvested after 4 h of growth. Subsequent purification and cleavage protocols were the same as for $SH3_W$. Purified $SH3_{7AW}$ was characterized by ESMS (Expected mass: 9431.5 Da and Observed mass: 9433.3 ± 1.9 Da). The overall yield of $SH3_{7AW}$ domain was 1.5-2 mg/L of culture. Protein concentration was determined by UV absorption (λ =280 nm); $SH3_{7AW}$, ε = 22840 M^{-1} cm^{-1}.

The cSH3 domain (residues L230-S304) was isolated from Crk-II cDNA by PCR using a 5'-primer (5'-CCGCTCGAGGAGAACCTGTA CTTCCAGGGCCTCCCTAACCTCCAGAATGGGCCCATTTATGCCAGG-

3') encoding a *Xho* I restriction site and a 3'-primer (5'-CCCAAGCTTTCAGCTGAAGTCCTCATCGGG-3') encoding a *Hind* III restriction site. The 5'-primer also encoded a poly-His tag and a Tev protease site. The PCR product was introduced into pTrcHis plasmid (Invitrogen) using *Xho*I and *Hind*III restriction sites. The resulting plasmid, pTrcHis-Tev-cSH3, was free of mutations in the protein-encoding region as determined by DNA sequencing. The plasmid pTrcHis-Tev-cSH3 encodes a Tev protease cleavage site between the poly-His tag and the cSH3 domain.

For expression of unlabeled cSH3 domain, the plasmid was transformed into *E.coli* BL21 cells (Invitrogen) and was grown to mid-log phase in Luria-Bertani medium containing 100 μg/mL ampicillin. Protein expression was induced with the addition of 1 mM IPTG at 37 °C for 6 h. The cells were then lysed by passage through a French press. Uniformly ^{13}C and ^{15}N labeled cSH3 domain was produced using the same procedure as above, except the cells were grown in M9 minimal media supplemented with 2 g/L ^{13}C-glucose and 1 g/L $^{15}NH_4Cl$ (Isotec) as the sole carbon and nitrogen source, respectively. For labeling the cSH3 domain with 7AW, the same procedure as mentioned earlier for making the Crk-I $SH3_{7AW}$ was followed. The His-tagged protein was then purified by affinity chromatography using Ni^{2+}- NTA resin (Novagen) and dialyzed into Tev cleavage buffer (20mM Tris, pH 8, 0.5mM EDTA and 1mM DTT).

Cleavage of the poly-His tag by Tev protease (Invitrogen) results in the addition of a glycine residue at the N-terminus of the cSH3 domain. Tev cleavage was initiated by the addition of 1 U of protease per 50 mg of fusion protein. After 3 days the cleaved product, cSH3, was purified by ion exchange chromatography using a HiTrap Q sepharose column (Amersham Biosciences) and a linear gradient of 10 mM Tris, pH 7.5 versus 10 mM Tris, pH 7.5, 1M NaCl at a flow rate of 1mL/min. The purity of the cSH3 domain was assessed to be >95% by HPLC and ESMS (Expected mass: 9005.6 Da, Observed mass: 8997.5 ± 3.2 Da). The purified cSH3 domain was dialyzed into the NMR buffer (10mM sodium phosphate, pH 7.2, 50mM NaCl, 5mM DTT-d_{10}, 0.1% NaN_3, 1mM EDTA and protease inhibitor cocktail (Roche) to a concentration of 0.3-0.4 mM. For fluorescence spectroscopy of $cSH3_{7AW}$, the protein was dialyzed into fluorescence buffer (20mM sodium phosphate, pH 7.2, 50mM NaCl). The concentrations of the cSH3 domains were determined by UV absorption (λ= 280). cSH3 domain, ε= 8370 M^{-1} cm^{-1} and $cSH3_{7AW}$, ε= 8470 M^{-1} cm^{-1}. The final yield of the purified cSH3 domain was ~ 5 mg/L and that of $cSH3_{7AW}$ was ~1 mg/L.

The lcSH3 domain (residues C208-S304) was isolated from Crk-II cDNA by PCR using a 5'-primer (5'- CCGCTCGAGATAGAAGGG CGGTGCCACCCACAGCCACTGGGTGGGCCGG-3') encoding a *Xho* I restriction site and a 3'-primer (5'- CCCAAGCTTTCAGC

TGAAGTCCTCATCGGG-3') encoding a *Hind* III restriction site. The 5'-primer also encoded a poly-His tag and a Factor Xa protease site. The resulting plasmid, pTrcHis-Xa-lcSH3 was shown to be free of mutations in the protein encoding region by DNA sequencing. For expression of unlabeled lcSH3 domain, the plasmid was transformed into *E.coli* BL21 cells (Invitrogen) and was grown to mid-log phase in Luria-Bertani medium containing 100 μg/mL ampicillin. Protein expression was induced with the addition of 1 mM IPTG at 37 °C for 6 h. The cells were then lysed by passage through a French press. Uniformly ^{15}N labeled lcSH3 domain was produced using the same procedure as above, except the cells were grown in M9 minimal media supplemented with 1 g/L $^{15}NH_4Cl$ (Isotec) as the sole nitrogen source. For labeling the lcSH3 domain with 7AW, the same procedure as mentioned earlier for making the Crk-I $SH3_{7AW}$ was followed. The His-tagged protein was then purified by affinity chromatography using Ni^{2+}- NTA resin (Novagen) and dialyzed into Factor Xa cleavage buffer (100 mM sodium phosphate at pH 7.2, 100 mM NaCl, and 1 mM $CaCl_2$). The fusion protein was then cleaved with Factor Xa (Amersham Biosciences) using 100 units of protease/mg of protein over 8-10 hours. This exposed the N-terminal cysteine residue required for ligation. The reaction was quenched by the addition of 20 mM DTT and the cleaved protein further purified by semi-preparative RP-HPLC using a linear gradient of 32%-38% solvent B over 60 minutes. The purified

protein was characterized by ESMS (for $IcSH3_{15N}$, Expected mass: 10822.0 Da, Observed mass: 10823.8 ± 1.8 Da and for $IcSH3_{7AW}$, Expected mass: 10688.0 Da, Observed mass: 10687 ± 2.8 Da). The final yield of purified $IcSH3_{15N}$ was ~7 mg/L and that of $IcSH3_{7AW}$ was ~2 mg/L. Protein concentrations were determined by UV absorption (λ =280 nm); $IcSH3_{15N}$, $\varepsilon = 10822\ M^{-1}\ cm^{-1}$; $IcSH3_{7AW}$, $\varepsilon = 9630\ M^{-1}\ cm^{-1}$.

Crk-I (residues M1-G207; SH2 and nSH3 domains) was isolated from Crk-II cDNA by PCR using a 5'-primer (5'- GGGAATTCCATATGG CGGGCAACTTCGACTCGG-3') encoding a *Nde* I restriction site and a 3'-primer (5'- GGACTAGTGCATCTCCCGTGATGCAACCCTCCTGGTT ACCTCCAATCAGAGCC -3') encoding a *Spe* I restriction site. The resulting plasmid, pTXB1-Crk1-GyrA-CBD was shown to be free of mutations in the protein encoding region by DNA sequencing. The plasmid pTXB1-Crk1-GyrA-CBD encodes for Crk-I followed by the Gyrase A intein and a chitin binding domain (CBD). For expression of the Crk-I-GyrA-CBD fusion protein, the plasmid was transformed into *E.coli* BL21 cells (Invitrogen) and was grown to mid-log phase in Luria-Bertani medium containing 100 μg/mL ampicillin. Protein expression was induced with the addition of 0.5 mM IPTG at 37 °C for 6 h. The cells were then lysed by passage through a French press. The Crk-I-GyrA-CBD fusion protein was then bound to chitin beads (Novagen). The Crk-I protein was cleaved from GryA-CBD by incubating the fusion protein bound beads in a buffer

containing 100 mM sodium phosphate at pH 7.2, 100 mM NaCl and 2% ethanthiol, overnight at room temperature. After on-bead cleavage of the intein to make Crk-I α-thioester, the proteins were eluted with buffer containing 6 M guanidinium chloride. Crk-I α-thioester was further purified by semi-preparative RP-HPLC using a linear gradient of 35%-55% solvent B over 60 minutes. The purified protein was characterized by ESMS (Expected mass: 23090.0 Da, Observed mass: 23097.1 ± 4.5 Da). The final yield of purified Crk-I-thioester was ~9 mg/L. Crk-I concentration was determined by UV absorption (λ =280 nm), ε = 33920 M^{-1} cm^{-1}.

Semi-synthesis of Crk-I[$SH3_{7AW}$], Crk-II[$IcSH3_{7AW}$] and Crk-II[$IcSH3_{15N}$]

Semi-synthesis of Crk-I[$SH3_{7AW}$]. Crk-I[$SH3_{7AW}$] was obtained by chemical ligation of SH2-thioester and $SH3_{7AW}$. Ligation reactions were initiated by dissolving the purified, lyophilized proteins to a final concentration of 0.2 mM each, in a buffer containing 100 mM sodium phosphate at pH 7.2, 100 mM NaCl, 6M guanidinium chloride, 2% MESNA, 2% ethanthiol and 1 mM $CaCl_2$. Ligation reactions were allowed to proceed for 4-5 days at 25 °C with constant stirring. The product, Crk-I[$SH3_{7AW}$] (residues M1 to G207) was purified by semi-preparative RP-HPLC using a linear gradient of 35% to 45% solvent B over 60 minutes and characterized by ESMS (Expected mass: 23048.5 Da, Observed mass: 23053 ± 5 Da) . In a typical ligation, 6 mg of purified Crk-I[$SH3_{7AW}$] was obtained from the reaction of 7 mg of

SH2-thioester and 3 mg of $SH3_{7AW}$. Crk-I[$SH3_{7AW}$] concentration was determined by UV absorption (λ =280 nm), ε = 33000 M^{-1} cm^{-1}.

Semi-synthesis of Crk-II[$lcSH3_{15N}$] and Crk-II[$lcSH3_{7AW}$]. Crk-I-thioester was chemically ligated to $lcSH3_{15N}$ and $lcSH3_{7AW}$ to obtain Crk-II[$lcSH3_{15N}$] and Crk-II[$lcSH3_{7AW}$] respectively. The ligation reaction to make Crk-II[$lcSH3_{15N}$] was initiated by dissolving the purified, lyophilized Crk-I-thioester to a final concentration of 1 mM and $lcSH3_{15N}$ to a final concentration of 0.4 mM. The ligation reaction to make Crk-II[$lcSH3_{7AW}$] was initiated by dissolving the purified, lyophilized Crk-I-thioester to a final concentration of 0.5 mM and $lcSH3_{7AW}$ to a final concentration of 0.4 mM . The ligations were carried out in a buffer containing 100 mM sodium phosphate at pH 7.2, 100 mM NaCl, 6M guanidinium chloride, 2% MESNA and 2% ethanthiol. Ligation reactions were allowed to proceed for 5-7 days at 25 °C with constant stirring. The products, Crk-II[$lcSH3_{15N}$] and Crk-II[$lcSH3_{7AW}$] (residues M1 to S304), were purified by semi-preparative RP-HPLC using a linear gradient of 39% to 43% solvent B over 60 minutes and characterized by ESMS (for Crk-II[$lcSH3_{15N}$], Expected mass: 33907.6 Da, Observed mass: 33851.0 ± 5.6 Da and for Crk-II[$lcSH3_{7AW}$], Expected mass: 33773.6 Da, Observed mass: 33770.0 ± 10.2 Da). In a typical ligation, 5 mg of purified Crk-II[$lcSH3_{15N}$] was obtained from the reaction of 20 mg of Crk-I-thioester and 8 mg of $lcSH3_{15N}$. The ligation reaction to make Crk-II[$lcSH3_{7AW}$] was performed once, yielding 1 mg of

Crk-II[lcSH3$_{7AW}$] from the reaction of 12 mg of Crk-I-thioester and 4 mg of lcSH3$_{7AW}$. Protein concentrations were determined by UV absorption (λ =280 nm); Crk-II[lcSH3$_{15N}$], ε = 43560 M^{-1} cm^{-1}; Crk-II[lcSH3$_{7AW}$], ε = 39310 M^{-1} cm^{-1}. The low yields were primarily due to the difficulty in separating Crk-I from Crk-II on RP-HPLC. This was especially acute when making Crk-II[lcSH3$_{7AW}$]. A better way of separating out Crk-I from ligated Crk-II (possibly by ion-exchange chromatography) is needed and will be addressed by ongoing work.

The extinction coefficients (ε) for all 7AW containing proteins was calculated by using the equation:

ε =(# of Tyr x 1210) + (# of Trp x 5500) + (# of 7AW x 6000)

All protein concentrations were determined in a buffer containing 6 M guanidinium chloride.

Tryptic digestion and mass spectrometry of SH3$_W$ and SH3$_{7AW}$

Purified SH3$_W$ and SH3$_{7AW}$ were run on SDS-PAGE using standard procedures and the respective bands excised for in-gel trypsin digestion (Shevchenko et al., 1996). Extractions of the digest were made as described in Shevchenko *et al.* (Shevchenko et al., 1996).The peptides thus derived were then prepared for MALDI-MS using 4-hydroxycinnaminnic acid (4HCCA) as the matrix. The ultrathin-layer method (Tsarbopoulos et al., 1994) was utilized for samples prepared in

4HCCA for MALDI-TOF analyses. The MALDI-TOF mass spectrometer was operated in the reflectron, delayed extraction mode. The resulting data was smoothed and calibrated using Data Explorer (PE Biosystems).

Fluorescence and Circular Dichroism Spectroscopy

Steady-state fluorescence spectra of all proteins were recorded using a Spex Fluorolog-3 spectroflourimeter. The spectra were recorded at 25 °C using an excitation wavelength of 295 nm (for Trp fluorescence), 310 nm (for 7AW fluorescence) and 5 nm slit widths. In all cases, protein concentration was approximately 1 μM and the protein samples were prepared in 20 mM sodium phosphate at pH 7.2, 100 mM NaCl buffer. CD spectra were recorded on an Aviv 62DS spectropolarimeter using a 0.1 cm cuvette. Protein concentrations were 50 μM in a buffer containing 5 mM sodium phosphate, 20 mM NaCl buffer at pH 7.2. Spectra were recorded at 25 °C in 0.5 nm steps from 260 to 195 nm and averaged over 4 seconds at each wavelength. The raw CD data (σ) was converted to mean residue ellipticity ($[\theta]$) using the following equation:

$$[i] = \frac{(100)(v)}{[P]nl}$$

where, σ is the CD signal, [P] is the protein concentration in mM, n is the number of amino acids and l is the cuvette pathlength in cm.

NMR spectroscopy

NMR samples were prepared by dissolving the purified lyophilized protein in a buffer containing 20 mM sodium phosphate (pH 7.2), 20 mM DTT-d_{10}, 100 mM NaCl, 10% (v/v) ^2H_2O and 0.1% (v/v) NaN_3 to a final concentration of 0.5-1.0 mM. 1D and 2D ^1H NMR experiments on the nSH3 domain (labeled and unlabeled) were performed on Bruker DPX-400 and DMX-500 spectrometers at 25 °C. A mixing time of 90 ms was used for the TOCSY measurements and 150 ms for NOESY measurements. Spectra were assigned based on those reported by Anafi *et al.* (Anafi et al., 1996) for the Crk nSH3 domain in the context of murine c-Crk and using assignments kindly provided by David Fushman (personal communication). The spectra were analyzed using XWINNMR (Bruker Instruments).

All NMR experiments on the cSH3, lcSH3 domains and Crk-II[lcSH3$_{15N}$] were recorded at 25 °C using Bruker Avance 500, 600, 700 and 800 MHz spectrometers (equipped with CryoProbes) and Varian Inova 600 spectrometers, equipped with triple resonance probes. A mixing time of 150 ms was used for NOESY measurements on unlabeled cSH3 and cSH3$_{7AW}$ domains. All NMR data was processed using NMRpipe and visualized using NMRDraw (Delaglio et al., 1995). NMRView was used to analyze all NMR data (Johnson, 2004). ^1H chemical shifts were referenced to water at 4.75 ppm (at 25 °C) whereas the $\gamma^{13}C/\gamma^1H$ and $\gamma^{15}N/\gamma^1H$ ratios

were used for indirect referencing of the ^{13}C and ^{15}N chemical shifts. ^{1}H-^{15}N-HSQC, ^{1}H-^{13}C-HSQC, 3D-^{15}N-NOESY-HSQC, HNCO, HNCACB, CBCA(CO)NH experiments were used to obtain the sequential assignment of the backbone (Sattler et al., 1999). Side-chain carbon and proton assignments were obtained using C(CO)NH, HC(CO)NH, HBHA(CO)NH and HCCH-TOCSY experiments. For assigning the aromatic resonances, ^{1}H-^{13}C HSQC, (Hβ)Cβ(CγCδ)Hδ and (Hβ)Cβ(CγCδCε)Hε (Yamazaki et al., 1993) and 3D ^{13}C NOESY-HSQC data was utilized.

Dihedral Angle, Distance, Orientation Restraints and Structure Calculations

The backbone dihedral angles (ϕ and ψ) were calculated by analyzing the $^{13}C_{\alpha}$, $^{13}C_{\beta}$, $^{13}C'$ and ^{15}N chemical shifts with the TALOS program that predicts the backbone torsion angles from the amino acid sequence and chemical shift information (Cornilescu et al., 1999a). The dihedral angles ϕ were also obtained from the HNHA experiment (Kuboniwa et al., 1994). Hydrogen bond restraints were obtained from 2D ^{3}J-HNCO (Cornilescu et al., 1999b). Inter-β strand NOEs were also observed for these hydrogen bonds. During later stages of structure calculations, these hydrogen bonds were added as explicit restraints. These restraints were defined as 1.8-2.3 Å for the H-O distance and 2.8-3.3 Å for the N-O distance.

Distance restraints were derived from 3D ^{13}C-edited NOESY (150 ms mixing time) and 3D ^{15}N-edited NOESY (150 ms mixing time) experiments. Using NMRView, the NOESY cross peak volumes/intensities were obtained and converted into distance restraints using the symmetry ambiguous distance restraints (ADR) protocol within the ARIA program (Habeck et al., 2004; Linge et al., 2003a). This protocol accounts for the ambiguity in the NOEs arising from signal overlap and symmetry degeneracy. A list of ADR is used to calculate molecular conformations and then used to filter the possible assignments. This iterative protocol generates a structure ensemble with consistent NOE assignments. The ARIA/CNS structure calculations were performed using both manually assigned peaks (unambiguous and ambiguous) and unassigned peaks. During the various iterations of peak assignments and structure calculations, ARIA/CNS either reject the violated peaks or re-assign these peaks to a different chemical shift resonance that correlated well with the structure. Thus at the end of 8^{th} iteration one would end up with set of unambiguous and ambiguous assignments which do not violate. Structure calculations were performed using the Cartesian dynamics simulated anneling protocol within ARIA/CNS. The following refinement protocol was used: a high temperature dynamics 2,000K (10,000 steps), followed by Cartesian dynamics cooling stage from 2,000K to 1,000K (6,000 steps) which was followed by second cooling stage from 1,000K to 50K (4,000

steps). 40 lowest energy structures were further refined in water using the protocol described by Linge et. al. (Linge et al., 2003b). The NMR structures were displayed and analyzed using MOLMOL (Koradi et al., 1996), PROCHECK (Laskowski et al., 1996), WHAT IF (Vriend, 1990), SwissPDB Viewer (Guex and Peitsch, 1997) and PyMOL (DeLano Scientific). The electrostatic potential were calculated SwissPDB Viewer and GRASP (Nicholls et al., 1991). The hydrophobic core residue surface areas were calculated using MOLMOL (Koradi et al., 1996).

Ligand binding assays

A fluorescence based titration assay was used to measure the affinity constants of the $SH3_W$ domain, $SH3_{7AW}$ domain and Crk-I[$SH3_{7AW}$] protein for a peptide ligand. All experiments were carried out at 25 °C in a stirred 1 cm pathlength cuvette using a Spex Fluorolog-3 spectroflourimeter. Excitation for the $SH3_W$ domain was at 295 nm, whereas the 7AW labeled samples were excited at 310 nm. In all cases, protein concentration was ~0.1 μM in 20 mM sodium phosphate (pH 7.2), 100 mM NaCl. The ligand used in these studies, H-PPPPLPPKRRK[ε-flourescein]G-NH_2, was based on a poly-proline sequence from the C3G protein (Knudsen et al., 1994) and was prepared by solid-phase peptide synthesis using Boc chemistry. Analysis of the changes in the fluorescence of the protein solution upon addition of defined

concentrations of ligand allowed us to determine the dissociation constants. Dissociation constants were calculated by assuming the formation of a 1:1 complex (Viguera AR, 1994) and by fitting the corrected fluorescence intensities (using non-linear least-squares analysis, GraphPad Prism v4.0a) to the equation (Lew J, 1997):

$$F = \frac{\{\Delta F\}\{P + K_d + L - [\sqrt{(P + K_d + L)^2 - 4PL}]\}}{2P} \qquad \text{(Eq.1)}$$

where F and ΔF are the change and maximum change in protein fluorescence respectively, P is the total protein concentration, L is the total ligand concentration and K_d is the equilibrium dissociation constant. All experiments were repeated three times. Protein and peptide concentrations were determined by UV absorption; $SH3_W$, $SH3_{7AW}$ and Crk-I[$SH3_{7AW}$] as earlier; poly-Pro ligand (λ = 495 nm), ε = 83000 M^{-1} cm^{-1}.

Equilibrium unfolding measurements

Equilibrium thermal denaturation experiments were performed on an Aviv 62DS spectropolarimeter using a 0.1 cm cuvette. The CD signal was recorded at each temperature at 220 nm and the signal was averaged for 4 seconds. The temperature was varied in 2 °C steps, in both directions, between 10 °C to 90 °C with 5 min. incubation at each temperature. Protein concentrations were kept at 50 μM and determined as described earlier. Protein samples were dissolved in 5 mM sodium

phosphate at pH 7.2, 20 mM NaCl. The CD signal at each temperature was fit to the following equation to determine the T_m values (Ramsay and Eftink, 1994):

$$Y_T = \left(\frac{1}{1 + e^{\frac{\left[-\Delta H\left(1 - \frac{T}{T_m}\right)\right]}{RT}}}\right) Y_N + \left[1 - \left(\frac{1}{1 + e^{\frac{\left[-\Delta H\left(1 - \frac{T}{T_m}\right)\right]}{RT}}}\right)\right] Y_U \qquad \text{(Eq. 2)}$$

where, Y_T is the CD signal at temperature T, T is the temperature, ΔH is the enthalpy of unfolding, T_m is the mid-point of the temperature transition, R is the gas constant, Y_N is the CD signal of the native protein and Y_U is the CD signal of the unfolded protein.

Equilibrium chemical denaturation experiments were performed on a Spex Fluorolog-3 spectrofluorimeter at 25 °C in a 1 cm pathlength cuvette with constant stirring. The fluorescence signals were recorded at 355 nm for the SH2 domain and 340 nm for the Crk-I $SH3_W$, cSH3 domains. The Trp residues in these domains were excited at 295 nm. The 7AW labeled proteins were excited at 310 nm and the fluorescence signals were recorded at 361 nm for $SH3_{7AW}$, at 388 nm for Crk-I[$SH3_{7AW}$], at 398 nm for $cSH3_{7AW}$ and at 400 nm for Crk-II[$lcSH3_{7AW}$]. Fluorescence signals were averaged for 60 sec at each guanidinium or urea concentration after a 5 min equilibration period with constant stirring. Protein concentrations were 5 μM (for cSH3 domain), 1 μM (for SH2, $SH3_W$, $SH3_{7AW}$, $cSH3_{7AW}$ domains and Crk-II[$cSH3_{7AW}$]) or 0.5 μM (for Crk-I[$SH3_{7AW}$]) in 20 mM sodium phosphate (pH 7.2), 100 mM NaCl.

Concentrations of guanidinium chloride (denoted [GdmCl]) and urea (denoted [Urea]) were determined by measuring the refractive index of the solution and using the following equations:

$$[\text{Urea}] = (117.66)(\gamma_x - \gamma_i) + (29.753)(\gamma_x - \gamma_i)^2 + (185.56)(\gamma_x - \gamma_i)^3$$

$$[\text{GdmCl}] = (54.147)(\gamma_x - \gamma_i) + (36.68)(\gamma_x - \gamma_i)^2 + (-91.6)(\gamma_x - \gamma_i)^3$$

where, γ_x is the measured refractive index of the solution and γ_i is the refractive index of the solution without any denaturant.

The measured fluorescence signals at each GdmCl or urea concentration were fit to the following equation (Pace, 1986):

$$F_D = \frac{\left\{a_N + b_N[D] + (a_U + b_U[D])e^{\frac{-\Delta G_U[D]}{RT}}\right\}}{1 + e^{\frac{-\Delta G_U[D]}{RT}}} \qquad \text{(Eq. 3)}$$

where,

$$\Delta G_U[D] = \Delta G^0_{H_2O} - m[D] \qquad \text{(Eq. 4)}$$

F_D is the measured fluorescence signal. α_N, β_N, α_U, and β_U are parameters that define the fluorescence signals of the native state (N) and the unfolded state (U) as a function of denaturant, [D]. ΔG^0_{H20} is the Gibbs free energy for unfolding in the absence of denaturant. Fraction folded at a given denaturant concentration was determined by using F_D, α_N, β_N, α_U, and β_U.

Stopped-flow measurements

Stopped-flow fluorescence measurements were performed using an

Applied Photophysics SX.18MV stopped-flow reaction analyzer equipped for asymmetric mixing at a ratio of 10:1. Fluorescence measurements were made with an excitation wavelength of 279 nm using a 305 nm cut-off filter. The refolding and unfolding were initiated by 11-fold dilution of a 50 μM solution of the cSH3 domain. All solutions contained 10 mM sodium phosphate buffer (pH 7.3) and 50 mM sodium chloride. The temperature of the syringes and the flow-cell was maintained at 25 °C with a circulating water bath. The resulting curves were fitted with a single exponential in order to determine rate constants for each reaction. The plots of ln k_{obs} versus denaturant concentration were fit to the following equation to determine the folding and unfolding rate constants in the absence of denaturant, k_f and k_u:

$$\ln k_{obs} = \ln \left\{ k_f \left(e^{\frac{m_f[D]}{RT}} \right) + k_u \left(e^{\frac{m_u[D]}{RT}} \right) \right\} \qquad \text{(Eq. 5)}$$

where m_f and m_u are constants that describe the change in ln k_f and ln k_u as a function of the concentration of denaturant, [D].

Appendix 1-

Chemical Shift Assignments of the cSH3 domain of Crk-II

Residue Number	Residue	Atom	Chemical Shift (ppm)
231	**Pro**	CA	62.133
		HA	4.4
		CA	31.254
		HB2	1.705
		HB1	2.232
		CG	26.747
		HG2	1.711
		HG1	2.018
		CD	49.676
		HD2	3.469
		HD1	3.882
		C	174.885
232	**Asn**	N	117.264
		HN	8.392
		CA	51.791
		HA	4.617
		CB	38.809
		HB2	2.717
		HB1	2.844
		CG	180.134
		ND2	112.238
		HD21	7.532
		HD22	6.84
		C	175.334
233	**Leu**	N	123.247
		HN	8.428
		CA	54.48
		HA	4.393
		CB	40.484
		HB2	1.515
		HB1	1.543
		CG	26.553
		HG	1.449
		CD1	24.923
		HD11	0.709
		CD2	22.989
		C	177.943

234	**Gln**	N	119.262
		HN	8.358
		CA	56.738
		HA	4.11
		CB	28.059
		HB1	2.015
		CG	33.361
		HG1	2.329
		CD	180.191
		NE2	111.587
		HE21	7.438
		HE22	6.769
		C	176.356
235	**Asn**	N	116.429
		HN	8.15
		CA	52.179
		HA	4.831
		CB	37.233
		HB2	2.69
		HB1	2.941
		CG	177.03
		ND2	112.312
		HD21	7.608
		C	174.509
236	**Gly**	N	108.343
		HN	7.584
		CA	43.79
		HA2	3.881
		HA1	4.319
237	**Pro**	CA	62.285
		HA	4.368
		CB	31.922
		HB2	1.497
		HB1	2.016
		CG	26.473
		HG1	2.001
		CD	48.917
		HD1	3.59
		C	175.644
238	**Ile**	N	118.161
		HN	7.74
		CA	58.785

		HA	4.454
		CB	39.778
		HB	1.796
		CG1	25.815
		HG12	1.062
		HG11	1.412
		CD1	12.965
		HD11	0.75
		CG2	17.338
		HG21	0.893
		C	173.588
239	**Tyr**	N	121.753
		HN	8.866
		CA	56.453
		HA	5.553
		CB	41.558
		HB2	2.806
		HB1	2.895
		CD1	133.213
		HD1	7.031
		CE1	117.823
		HE1	6.76
		C	174.645
240	**Ala**	N	121.621
		HN	9.402
		CA	49.501
		HA	5.155
		CB	22.928
		HB1	1.207
		C	174.047
241	**Arg**	N	120.198
		HN	8.814
		CA	52.262
		HA	5.221
		CB	31.921
		HB1	1.676
		CG	26.006
		HG1	1.537
		CD	41.883
		HD2	3.081
		HD1	3.129
		C	176.128

242	**Val**	N	124.657
		HN	8.889
		CA	63.509
		HA	4.015
		CB	31.311
		HB	2.275
		HG21	0.716
		CG1	23.262
		HG11	0.889
		C	177.716
243	**Ile**	N	122.186
		HN	8.92
		CA	60.149
		HA	4.556
		CB	38.952
		HB	2.05
		CG1	26.068
		HG12	1.007
		HG11	0.599
		CD1	13.236
		HD11	0.717
		CG2	18.232
		HG21	0.596
		C	174.602
244	**Gln**	N	123.474
		HN	7.407
		CA	54.483
		HA	4.305
		CB	30.77
		HB2	1.36
		HB1	1.585
		CG	32.422
		HG1	2.158
		CD	178.783
		NE2	109.205
		HE21	6.909
		HE22	6.495
		C	172.157
245	**Lys**	N	126.103
		HN	8.394
		CA	56.272
		HA	3.879

		CB	33.004
		HB2	1.683
		HB1	1.938
		CG	24.128
		HG2	1.254
		HG1	1.632
		CD	29.219
		HD1	1.697
		CE	41.315
		HE2	2.713
		HE1	2.973
		C	175.287
246	**Arg**	N	126.031
		HN	9.018
		CA	55.04
		HA	4.663
		CB	33.26
		HB2	1.423
		HB1	1.859
		HG1	1.591
247	**Val**	N	129.211
		HN	9.047
		HA	4.481
		HG11	0.908
248	**Pro**	CA	61.571
		HA	4.234
		CB	32.258
		HB2	1.879
		HB1	2.049
		CG	26.434
		HG2	1.561
		HG1	1.879
		CD	50.335
		HD2	3.761
		HD1	3.88
		C	175.579
249	**Asn**	N	120.197
		HN	8.812
		CA	52.357
		HA	4.526
		CB	38.606
		HB2	2.637

		HB1	2.709
		ND2	113.393
		HD21	7.711
		HD22	7.065
		C	176.225
250	**Ala**	N	125.323
		HN	8.657
		CA	53.568
		HA	3.905
		CB	17.7
		HB1	1.159
		C	177.663
251	**Tyr**	N	112.342
		HN	7.748
		CA	56.676
		HA	4.477
		CB	36.93
		HB2	2.99
		HB1	3.175
		CD1	133.369
		HD1	7.142
		CE1	118.463
		HE1	6.857
		C	175.23
252	**Asp**	N	120.01
		HN	7.459
		CA	51.913
		HA	4.873
		CB	39.705
		HB2	2.455
		HB1	2.948
		C	176.314
253	**Lys**	N	122.83
		HN	8.544
		CA	57.05
		HA	4.247
		CB	31.455
		HB2	1.88
		HB1	2.01
		CG	24.267
		HG1	1.557
		CD	28.094

		HD1	1.714
		CE	41.549
		HE1	3.047
		C	177.327
254	**Thr**	N	107.679
		HN	8.214
		CA	60.759
		HA	4.549
		CB	68.538
		HB	4.626
		CG2	21.613
		HG21	1.25
		C	173.028
255	**Ala**	N	123.797
		HN	6.943
		CA	50.549
		HA	3.373
		CB	18.797
		HB1	1.189
		C	176.421
256	**Leu**	N	123.699
		HN	8.419
		CA	53.657
		HA	4.188
		CB	43.301
		HB2	0.907
		HB1	1.439
		CG	25.436
		HG	1.323
		CD1	27.908
		HD11	0.707
		CD2	24.038
		HD21	0.576
		HD22	0.576
		C	175.922
257	**Ala**	N	129.104
		HN	8.564
		CA	51.041
		HA	4.345
		CB	18.146
		HB1	1.35
		C	176.035

258	**Leu**	N	119.937
		HN	8.337
		CA	53.047
		HA	4.77
		CB	44.796
		HB2	1.497
		HB1	1.624
		CG	22.33
		HG	1.502
		CD1	24.933
		HD11	0.537
		HD22	0.45
		C	175.507
259	**Glu**	N	122.158
		HN	8.569
		CA	52.761
		HA	4.574
		CB	30.392
		HB2	1.552
		HB1	1.897
		CG	34.608
		HG2	2.131
		HG1	2.268
		C	175.177
260	**Val**	N	121.748
		HN	8.862
		CA	65.012
		HA	2.902
		CB	30.387
		HB	1.833
		CG2	20.651
		CG1	22.499
		HG11	0.84
		C	177.133
261	**Gly**	N	113.836
		HN	8.547
		CA	43.968
		HA2	3.553
		HA1	4.443
		C	174.564
262	**Glu**	N	120.074
		HN	7.692

		CA	56.551
		HA	4.172
		CB	29.996
		HB2	1.851
		HB1	2.164
		CG	37.789
		HG2	2.065
		HG1	2.295
		C	174.466
263	**Leu**	N	121.378
		HN	8.505
		CA	53.226
		HA	4.962
		CB	43.428
		HB2	1.512
		HB1	1.732
		CG	26.629
		HG	1.536
		CD1	24.605
		HD11	0.81
		CD2	23.722
		HD22	0.748
		C	175.599
264	**Val**	N	125.568
		HN	9.367
		CA	59.573
		HA	4.43
		CB	34.259
		HB	1.592
		CG2	20.379
		CG1	21.519
		HG11	0.422
		C	174.531
265	**Lys**	N	127.716
		HN	8.769
		CA	54.521
		HA	4.292
		CB	32.991
		HB2	1.706
		HB1	1.82
		CG	24.39
		HG1	0.842

		CD	28.902
		HD1	1.566
		CE	41.412
		HE1	2.781
		C	175.028
266	**Val**	N	129.4
		HN	9.11
		CA	63.485
		HA	4.19
		CB	30.975
		HB	2.054
		HG22	0.891
		CG1	21.6
		HG11	1.139
		C	175.214
267	**Thr**	N	116.966
		HN	9.153
		CA	61.236
		HA	4.428
		CB	68.431
		HB	4.222
		CG2	21.239
		HG21	1.06
		C	175.506
268	**Lys**	N	122.348
		HN	7.641
		CA	56.858
		HA	4.122
		CB	35.622
		HB2	1.46
		HB1	1.718
		CG	24.419
		HG1	1.201
		CD	28.488
		HD1	1.598
		CE	41.446
		HE1	2.932
		C	173.266
269	**Ile**	N	124.757
		HN	8.171
		CA	59.719
		HA	2.883

		CB	37.969
		HB	1.068
		CG1	26.105
		HG12	0.656
		HG11	-0.624
		CD1	12.17
		HD11	0.088
		CG2	15.33
		HG21	0.088
		C	174.124
270	**Asn**	N	122.208
		HN	6.272
		CA	51.406
		HA	5.002
		CB	41.054
		HB1	2.944
		ND2	113.403
		HD21	7.821
		HD22	7.065
271	**Val**	CA	63.62
		HA	3.985
		CB	30.887
		HB	2.204
		HG22	0.993
		CG1	19.944
		HG11	1.008
		C	176.246
272	**Ser**	N	115.684
		HN	8.334
		CA	58.234
		HA	4.281
		CB	62.981
		HB1	3.956
		C	175.399
273	**Gly**	N	108.51
		HN	7.961
		CA	44.443
		HA2	3.707
		HA1	4.418
		C	172.432
274	**Gln**	N	120.776
		HN	8.004

		CA	53.874
		HA	4.652
		CB	27.429
		HB2	1.792
		HB1	1.927
		CG	33.198
		HG2	2.056
		HG1	2.191
		CD	179.373
		NE2	111.154
		HE21	7.267
		HE22	6.407
		C	174.755
275	**Trp**	N	127.463
		HN	8.407
		CA	53.611
		HA	4.854
		CB	31.083
		HB2	2.052
		HB1	3.097
		CD1	121.362
		HD1	6.911
		NE1	127
		HE1	9.898
		CZ2	115.646
		HZ2	7.156
		CH2	125.268
		HH2	7.152
		CZ3	120.722
		HZ3	6.944
		CE3	120.9
		HE3	7.503
		C	173.576
276	**Glu**	N	115.689
		HN	8.607
		CA	53.726
		HA	5.262
		CB	32.457
		HB2	1.731
		HB1	1.988
		CG	35.919
		HG1	2.015

		C	176.588
277	**Gly**	N	108.438
		HN	8.961
		CA	46.204
		HA2	4.265
		HA1	5.02
		C	170.617
278	**Glu**	N	117.972
		HN	9.033
		CA	53.162
		HA	5.58
		CB	33.188
		HB2	1.878
		HB1	1.988
		CG	35.634
		HG1	2.119
		C	175.407
279	**Cys**	N	123.743
		HN	9.063
		CA	57.848
		HA	4.696
		CB	28.774
		HB1	2.598
		C	174.747
280	**Asn**	N	129.084
		HN	9.804
		CA	53.64
		HA	4.493
		CB	36.935
		HB2	2.814
		HB1	3.124
		CG	177.75
		ND2	113.989
		HD21	7.625
		HD22	7.019
		C	174.774
281	**Gly**	N	103.541
		HN	8.879
		CA	44.774
		HA2	3.645
		HA1	4.161
		C	173.571

282	**Lys**	N	120.605
		HN	7.756
		CA	53.817
		HA	4.642
		CB	33.759
		HB1	1.822
		CG	24.194
		HG1	1.446
		CD	28.112
		HD2	1.721
		HD1	1.823
		CE	41.627
		HE1	3.047
		C	174.363
283	**Arg**	N	119.375
		HN	8.477
		CA	53.598
		HA	5.507
		CB	32.701
		HB2	1.751
		HB1	1.831
		CG	26.546
		HG1	1.564
		CD	42.848
		HD1	3.169
		C	175.974
284	**Gly**	N	109.473
		HN	8.588
		CA	45.41
		HA2	4.212
		HA1	4.393
		C	171.243
285	**His**	N	115.706
		HN	8.296
		CA	53.295
		HA	6.073
		CB	32.888
		HB1	3.002
		CD2	120.514
		HD2	6.94
		CE1	138.69
		HE1	7.696

		C	176.291
286	**Phe**	N	118.414
		HN	9.028
		CA	54.889
		HA	5.116
		CB	38.436
		HB2	2.68
		HB1	3.005
		CD1	133.619
		HD1	6.906
		CE1	130.046
		HE1	6.95
		CZ	128.812
		HZ	7.275
287	**Pro**	CA	60.23
		HA	4.893
		CB	30.869
		HB2	2.014
		HB1	2.22
		CG	26.25
		HG1	1.578
		CD	49.702
		HD2	3.243
		HD1	3.682
		C	179.368
288	**Phe**	N	124.101
		HN	8.368
		CA	58.551
		HA	3.488
		CB	35.161
		HB1	1.281
		CD1	132.159
		HD1	6.929
		CE1	130.173
		HE1	7.294
		C	174.661
289	**Thr**	N	106.933
		HN	6.625
		CA	61.768
		HA	3.563
		CB	68.727
		HB	4.145

		CG2	21.362
		HG21	0.796
		C	175.884
290	**His**	N	118.625
		HN	7.729
		CA	56.816
		HA	4.218
		CB	29.041
		HB2	2.823
		HB1	3.746
		CD2	120.49
		HD2	6.925
		CE1	138.699
		HE1	7.578
		C	174.863
291	**Val**	N	110.189
		HN	7.571
		CA	57.612
		HA	5.338
		CB	35.002
		HB	1.885
		CG2	17.814
		HG22	0.534
		CG1	20.496
		HG11	0.833
		C	172.706
292	**Arg**	N	119.117
		HN	8.929
		CA	53.294
		HA	4.797
		CB	32.556
		HB1	1.688
		CG	26.373
		HG1	1.584
		CD	42.693
		HD1	3.201
		C	175.751
293	**Leu**	N	126.376
		HN	9.143
		CA	55.504
		HA	4.388
		CB	41.276

		HB2	1.685
		HB1	1.893
		CG	26.962
		HG	2.054
		CD1	24.989
		HD11	1.174
		CD2	23.183
		HD22	1.092
		C	177.034
294	**Leu**	N	122.361
		HN	8.491
		CA	54.097
		HA	4.463
		CB	41.167
		HB2	1.53
		HB1	1.697
		CG	27.048
		HG	1.632
		CD1	25.035
		HD11	0.81
		CD2	22.399
		C	177.017
295	**Asp**	N	120.856
		HN	8.624
		CA	53.503
		HA	4.651
		CB	40.714
		HB2	2.657
		HB1	2.799
		C	176.371
296	**Gln**	N	119.972
		HN	8.343
		CA	54.973
		HA	4.309
		CB	28.832
		HB2	2.008
		HB1	2.155
		CG	33.48
		HG1	2.367
		CD	180.19
		NE2	112.061
		HE21	7.546

		HE22	6.843
		C	175.756
297	**Gln**	N	119.498
		HN	8.433
		CA	55.295
		HA	4.282
		CB	28.685
		HB2	1.996
		HB1	2.111
		CG	33.456
		HG1	2.335
		NE2	112.058
		HE21	7.546
		HE22	6.805
		C	175.187
298	**Asn**	N	119.341
		HN	8.294
		CA	52.912
		HA	4.924
		CB	38.41
		HB2	2.641
		HB1	2.839
		CG	176.92
		ND2	112.643
		HD21	7.669
		HD22	6.947
299	**Pro**	CA	63.064
		HA	4.367
		CB	31.36
		HB2	1.934
		HB1	2.215
		CG	26.425
		HG1	1.935
		CD	50.064
		HD1	3.728
		C	176.423
300	**Asp**	N	118.08
		HN	8.116
		CA	53.786
		HA	4.545
		CB	40.269
		HB2	2.547

		HB1	2.714
		C	175.949
301	**Glu**	N	119.786
		HN	7.813
		CA	55.573
		HA	4.17
		CB	29.919
		HB2	1.846
		HB1	1.919
		CG	35.663
		HG1	2.11
		C	175.178
302	**Asp**	N	120.952
		HN	8.168
		CA	53.292
		HA	4.544
		CB	40.64
		HB2	2.492
		HB1	2.608
		C	175.527
303	**Phe**	N	121.28
		HN	8.158
		CA	56.784
		HA	4.608
		CB	38.582
		HB2	2.892
		HB1	3.173
		CD1	132.027
		HD1	7.178
		CE1	131.444
		HE1	7.274
		C	174.69
304	**Ser**	N	122.226
		HN	7.843
		CA	59.766
		CB	64.272

References

Abassi, Y. A., and Vuori, K. (2002). Tyrosine 221 in Crk regulates adhesion-dependent membrane localization of Crk and Rac and activation of Rac signaling. EMBO J *21*, 4571-4582.

Akakura, S., Kar, B., Singh, S., Cho, L., Tibrewal, N., Sanokawa-Akakura, R., Reichman, C., Ravichandran, K. S., and Birge, R. B. (2005). C-terminal SH3 domain of CrkII regulates the assembly and function of the DOCK180/ELMO Rac-GEF. J Cell Physiol *204*, 344-351.

Alsina, J., Yokum, T. S., Albericio, F., and Barany, G. (1999). Backbone Amide Linker (BAL) Strategy for N(alpha)-9-Fluorenylmethoxycarbonyl (Fmoc) Solid-Phase Synthesis of Unprotected Peptide p-Nitroanilides and Thioesters(1). J Org Chem *64*, 8761-8769.

Anafi, M., Rosen, M. K., Gish, G. D., Kay, L. E., and Pawson, T. (1996). A potential SH3 domain-binding site in the Crk SH2 domain. J Biol Chem *271*, 21365-21374.

Andreotti, A. H., Bunnell, S. C., Feng, S., Berg, L. J., and Schreiber, S. L. (1997). Regulatory intramolecular association in a tyrosine kinase of the Tec family. Nature *385*, 93-97.

Bang, D., and Kent, S. B. H. (2004). A one-pot total synthesis of crambin. Angew Chem Intl Ed Engl *43*, 2534-2538.

Barila, D., and Superti-Furga, G. (1998). An intramolecular SH3-domain interaction regulates c-Abl activity. Nat Genet *18*, 280.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., *et al.* (2004). The Pfam protein families database. Nucleic Acids Res *32*, D138-141.

Bianchi, E., Ingenito, R., Simon, R. J., and Pessi, A. (1999). Engineering and Chemical Synthesis of a Transmembrane Protein: The HCV Protease

Cofactor Protein NS4A. J Am Chem Soc *121*, 7698-7699.

Blaschke, U. K., Cotton, G. J., and Muir, T. W. (2000). Synthesis of multi-domain proteins using expressed protein ligation: Strategies for segmental isotopic labeling of internal regions. Tetrahedron *56*, 9461-9470.

Botti, P., Villain, M., Manganiello, S., and Gaertner, H. (2004). Native chemical ligation through in situ O to S acyl shift. Org Lett *6*, 4861-4864.

Burbulis, I., Yamaguchi, K., Gordon, A., Carlson, R., and Brent, R. (2005). Using protein-DNA chimeras to detect and count small numbers of molecules. Nat Methods *2*, 31-37.

Camarero, J. A., Cotton, G. J., Adeva, A., and Muir, T. W. (1998). Chemical ligation of unprotected peptides directly from a solid support. Journal Of Peptide Research *51*, 303-316.

Camarero, J. A., Fushman, D., Sato, S., Giriat, I., Cowburn, D., Raleigh, D. P., and Muir, T. W. (2001). Rescuing a destabilized protein fold through backbone cyclization. J Mol Biol *308*, 1045-1062.

Camarero, J. A., Hackel, B. J., deYoreo, J. J., and Mitchell, A. R. (2004). Fmoc-Based Synthesis of Peptide alpha-Thioesters Using an Aryl Hydrazine Support. J Org Chem *69*, 4145-4151.

Camarero, J. A., Shekhtman, A., Campbell, E. A., Chlenov, M., Gruber, T. M., Bryant, D. A., Darst, S. A., Cowburn, D., and Muir, T. W. (2002). Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. Proc Natl Acad Sci USA *99*, 8536-8541.

Canne, L. E., Bark, S. J., and Kent, S. B. H. (1996). Extending the applicability of native chemical ligation. J Am Chem Soc *118*, 5891-5896.

Cavanagh, J., Fairbrother, J.W., Palmer III, A.G. and Skelton, N.J. (1996). Protein NMR Spectroscopy; Principles and Practice, First edn (San Diego:

Academic Press).

Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., Perler, F. B., Benner, J., Kucera, R. B., Hirvonen, C. A., *et al.* (1997). Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. Gene *192*, 271-281.

Chong, S., Montello, G. E., Zhang, A., Cantor, E. J., Liao, W., Xu, M. Q., and Benner, J. (1998a). Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. Nucl Acids Res *26*, 5109-5115.

Chong, S., Williams, K. S., Wotkowicz, C., and Xu, M. Q. (1998b). Modulation of protein splicing of the Saccharomyces cerevisiae vacuolar membrane ATPase intein. J Biol Chem *273*, 10567-10577.

Clackson, T., and Wells, J. A. (1995). A Hot Spot of Binding Energy in a Hormone-Receptor Interface. Science *267*, 383-386.

Clayton, D., Shapovalov, G., Maurer, J. A., Dougherty, D. A., Lester, H. A., and Kochendoerfer, G. G. (2004). Total chemical synthesis and electrophysiological characterization of mechanosensitive channels from Escherichia coli and Mycobacterium tuberculosis. Proc Natl Acad Sci USA *101*, 4764-4769.

Cohen, B. E., McAnaney, T. B., Park, E. S., Jan, Y. N., Boxer, S. G., and Jan, L. Y. (2002). Probing Protein Electrostatics with a Synthetic Fluorescent Amino Acid. Science *296*, 1700-1703.

Cornilescu, G., Delaglio, F., and Bax, A. (1999a). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR *13*, 289-302.

Cornilescu, G., Hu, J. S., and Bax, A. (1999b). Identification of the hydrogen bonding network in a protein by scalar couplings. J Am Chem

Soc *121*, 2949-2950.

Cotton, G. J., Ayers, B., Xu, R., and Muir, T. W. (1999). Insertion of a synthetic peptide into a recombinant protein framework: A protein biosensor. J Am Chem Soc *121*, 1100-1101.

Danielson, M. A., and Falke, J. J. (1996). Use of 19F NMR to probe protein structure and conformational changes. Annu Rev Biophys Biomol Struct *25*, 163-195.

Dawson, P. E., Churchill, M. J., Ghadiri, M. R., and Kent, S. B. H. (1997). Modulation of reactivity in native chemical ligation through the use of thiol additives. J Am Chem Soc *119*, 4325-4329.

Dawson, P. E., Muir, T. W., Clarklewis, I., and Kent, S. B. H. (1994). Synthesis Of Proteins By Native Chemical Ligation. Science *266*, 776-779.

de Jong, R., Haataja, L., Voncken, J. W., Heisterkamp, N., and Groffen, J. (1995a). Tyrosine phosphorylation of murine Crkl. Oncogene *11*, 1469-1474.

de Jong, R., ten Hoeve, J., Heisterkamp, N., and Groffen, J. (1995b). Crk1 is complexed with tyrosine-phosphorylated Cb1 in Ph-positive leukemia. J Biol Chem *270*, 21468-21471.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR *6*, 277-293.

Donaldson, L. W., Gish, G., Pawson, T., Kay, L. E., and Forman-Kay, J. D. (2002). Structure of a regulatory complex involving the Abl SH3 domain, the Crk SH2 domain, and a Crk-derived phosphopeptide. Proc Natl Acad Sci USA *99*, 14053-14058.

Dose, C., and Seitz, O. (2005). Convergent synthesis of peptide nucleic

acids by native chemical Ligation. Org Lett *7*, 4365-4368.

Dueber, J. E., Yeh, B. J., Bhattacharyya, R. P., and Lim, W. A. (2004). Rewiring cell signaling: the logic and plasticity of eukaryotic protein circuitry. Curr Opin Struct Biol *14*, 690.

Evans, T. C., Jr., Benner, J., and Xu, M.-Q. (1999). The in Vitro Ligation of Bacterially Expressed Proteins Using an Intein from Methanobacterium thermoautotrophicum. J Biol Chem *274*, 3923-3926.

Feller, S. M. (2001). Crk family adaptors - signalling complex formation and biological roles. Oncogene *20*, 6348-6371.

Feller, S. M., Knudsen, B., and Hanafusa, H. (1994). C-Abl Kinase Regulates The Protein-Binding Activity Of C-Crk. EMBO J *13*, 2341-2351.

Ficht, S., Mattes, A., and Seitz, O. (2004). Single-nucleotide-specific PNA-peptide ligation on synthetic and PCR DNA templates. J Am Chem Soc *126*, 9970-9981.

Filimonov, V. V., Azuaga, A. I., Viguera, A. R., Serrano, L., and Mateo, P. L. (1999). A thermodynamic analysis of a family of small globular proteins: SH3 domains. Biophys Chem *77*, 195-208.

Gaertner, H. F., Rose, K., Cotton, R., Timms, D., Camble, R., and Offord, R. E. (1992). Construction of Protein Analogs by Site-Specific Condensation of Unprotected Fragments. Bioconjug Chem *3*, 262-268.

Ghose, R., Shekhtman, A., Goger, M. J., Ji, H., and Cowburn, D. (2001). A novel, specific interaction involving the Csk SH3 domain and its natural ligand. Nat Struct Biol *8*, 998-1004.

Grantcharova, V. P., and Baker, D. (1997). Folding dynamics of the src SH3 domain. Biochemistry *36*, 15685-15692.

Grantcharova, V. P., Riddle, D. S., and Baker, D. (2000). Long-range order in the src SH3 folding transition state. Proc Natl Acad Sci U S A *97*, 7084-7089.

Groemping, Y., Lapouge, K., Smerdon, S. J., and Rittinger, K. (2003). Molecular basis of phosphorylation-induced activation of the NADPH oxidase. Cell *113*, 343-355.

Guex, N., and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis *18*, 2714-2723.

Guijarro, J. I., Morton, C. J., Plaxco, K. W., Campbell, I. D., and Dobson, C. M. (1998). Folding kinetics of the SH3 domain of PI3 kinase by real-time NMR combined with optical spectroscopy. J Mol Biol *276*, 657-667.

Habeck, M., Rieping, W., Linge, J. P., and Nilges, M. (2004). NOE assignment with ARIA 2.0: the nuts and bolts. Methods Mol Biol *278*, 379-402.

Hackeng, T. M., Griffin, J. H., and Dawson, P. E. (1999). Protein synthesis by native chemical ligation: Expanded scope by using straightforward methodology. Proc Natl Acad Sci USA *96*, 10068-10073.

Hackeng, T. M., Mounier, C. M., Bon, C., Dawson, P. E., Griffin, J. H., and Kent, S. B. H. (1997). Total chemical synthesis of enzymatically active human type II secretory phospholipase A2. Proc Natl Acad Sci USA *94*, 7845-7850.

Hang, H. C., and Bertozzi, C. R. (2001). Chemoselective approaches to glycoprotein assembly. Acc Chem Res *34*, 727-736.

Hasegawa, H., Kiyokawa, E., Tanaka, S., Nagashima, K., Gotoh, N., Shibuya, M., Kurata, T., and Matsuda, M. (1996). DOCK180, a major CRK-binding protein, alters cell morphology upon translocation to the cell membrane. Mol Cell Biol *16*, 1770-1776.

Hashimoto, Y., Katayama, H., Kiyokawa, E., Ota, S., Kurata, T., Gotoh, N., Otsuka, N., Shibata, M., and Matsuda, M. (1998). Phosphorylation of CrkII adaptor protein at tyrosine 221 by epidermal growth factor receptor. J Biol Chem *273*, 17186-17191.

Haslbeck, M., Franzmann, T., Weinfurtner, D., and Buchner, J. (2005). Some like it hot: the structure and function of small heat-shock proteins. Nat Struct Mol Biol *12*, 842.

Hunter, C. L., and Kochendoerfer, G. G. (2004). Native chemical ligation of hydropobic peptides in lipid bilayer systems. Bioconjug Chem *15*, 437-440.

Ingenito, R., Bianchi, E., Fattori, D., and Pessi, A. (1999). Solid phase synthesis of peptide C-terminal thioesters by Fmoc/t-Bu chemistry. J Am Chem Soc *121*, 11369-11374.

Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol Biol *278*, 313-352.

Kato, M., Miyazawa, K., and Kitamura, N. (2000). A deubiquitinating enzyme UBPY interacts with the Src homology 3 domain of Hrs-binding protein via a novel binding motif PX(V/I)(D/N)RXXKP. J Biol Chem *275*, 37481-37487.

Kay, L. E., Torchia, D. A., and Bax, A. (1989). Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry *28*, 8972-8979.

KizakaKondoh, S., Matsuda, M., and Okayama, H. (1996). CrkII signals from epidermal growth factor receptor to Ras. Proc Natl Acad Sci USA *93*, 12177-12182.

Knudsen, B. S., Feller, S. M., and Hanafusa, H. (1994). Four Proline-Rich Sequences Of The Guanine-Nucleotide Exchange Factor C3G Bind With

Unique Specificity To The First Src Homology-3 Domain Of Crk. J Biol Chem *269*, 32781-32787.

Kochendoerfer, G. G., Salom, D., Lear, J. D., Wilk-Orescan, R., Kent, S. B. H., and DeGrado, W. F. (1999). Total Chemical Synthesis of the Integral Membrane Protein Influenza A Virus M2: Role of Its C-Terminal Domain in Tetramer Assembly. Biochemistry *38*, 11905-11913.

Kolb, H. C., and Sharpless, K. B. (2003). The growing impact of click chemistry on drug discovery. Drug Discovery Today *8*, 1128.

Koradi, R., Billeter, M., and Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph *14*, 51-55, 29-32.

Krutchinsky AN, Z. W., Chait BT (2000). Rapidly switchable matrix-assisted laser desorption/ionization and electrospray quadrupole-time-of-flight mass spectrometry for protein identification. J Am Soc Mass Spectrom *11*, 493-504.

Kuboniwa, H., Grzesiek, S., Delaglio, F., and Bax, A. (1994). Measurement of H-N-H-Alpha J-Couplings in Calcium-Free Calmodulin Using New 2D and 3D Water-Flip-Back Methods. J Biomol NMR *4*, 871-878.

Kuriyan, J., and Cowburn, D. (1997). Modular peptide recognition domains in eukaryotic signaling. Annu Rev Biophys Biomol Struct *26*, 259-288.

Laederach, A., Cradic, K. W., Bruce Fulton, D., and Andreotti, A. H. (2003). Determinants of Intra versus Intermolecular Self-association Within the Regulatory Domains of Rlk and Itk. J Mol Biol *329*, 1011.

Lakowicz, J. R. (1999). Principles of Fluorescence Spectroscopy, Second edn (New York: Kluwer Academic/Plenum Publishers).

Lamorte, L., Rodrigues, S., Naujokas, M., and Park, M. (2002a). Crk synergizes with epidermal growth factor for epithelial invasion and morphogenesis and is required for the Met morphogenic program. J Biol Chem *277*, 37904-37911.

Lamorte, L., Royal, I., Naujokas, M., and Park, M. (2002b). Crk adapter proteins promote an epithelial-mesenchymal-like transition and are required for HGF-mediated cell spreading and breakdown of epithelial adherens junctions. Mol Biol Cell *13*, 1449-1461.

Larson, S. M., and Davidson, A. R. (2000). The identification of conserved interactions within the SH3 domain by alignment of sequences and structures. Protein Sci *9*, 2170-2180.

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR *8*, 477-486.

Lee, C. H., Saksela, K., Mirza, U. A., Chait, B. T., and Kuriyan, J. (1996). Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain. Cell *85*, 931-942.

Lew J, C. N., Tsigelny I, Garrod S, Taylor SS (1997). Synergistic binding of nucleotides and inhibitors to cAMP-dependent protein kinase examined by acrylodan fluorescence spectroscopy. J Biol Chem *272*, 1507-1513.

Lim, A. W., Fox, R. O., and Richards, F. M. (1994). Stability and peptide binding affinity of an SH3 domain from the C.elegan siganling protein Sem-5. Protein Sci *3*, 1261-1266.

Lim, W. A. (2002). The modular logic of signaling proteins: building allosteric switches from simple binding domains. Curr Opin Struct Biol *12*, 61.

Linge, J. P., Habeck, M., Rieping, W., and Nilges, M. (2003a). ARIA:

automated NOE assignment and NMR structure calculation. Bioinformatics *19*, 315-316.
Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M., and Nilges, M. (2003b). Refinement of protein structures in explicit solvent. Proteins *50*, 496-506.

Lovrinovic, M., Seidel, R., Wacker, R., Schroeder, H., Seitz, O., Engelhard, M., Goody, R. S., and Niemeyer, C. M. (2003). Synthesis of protein-nucleic acid conjugates by expressed protein ligation. Chemical Communications, 822-823.

Manning, M. C., and Woody, R. W. (1989). Theoretical study of the contribution of aromatic side chains to the circular dichroism of basic bovine pancreatic trypsin inhibitor. Biochemistry *28*, 8609-8613.

Mannuzzu, L. M., Moronne, M.M., Isacoff, E.Y. (1996). Direct Physical Measure of Conformational Rearrangement Underlying Potassium Channel Gating. Science *271*, 213-216.

Manser, E., Loo, T. H., Koh, C. G., Zhao, Z. S., Chen, X. Q., Tan, L., Tan, I., Leung, T., and Lim, L. (1998). PAK kinases are directly coupled to the PIX family of nucleotide exchange factors. Mol Cell Biol *1*, 183-192.

Mathys, S., Evans, T. C., Chute, I. C., Wu, H., Chong, S., Benner, J., Liu, X. Q., and Xu, M. Q. (1999). Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. Gene *231*, 1-13.

Matsuda, M., Ota, S., Tanimura, R., Nakamura, H., Matuoka, K., Takenawa, T., Nagashima, K., and Kurata, T. (1996). Interaction between the amino-terminal SH3 domain of CRK and its natural target proteins. J Biol Chem *271*, 14468-14472.

Matsuda, M., Tanaka, S., Nagata, S., Kojima, A., Kurata, T., and Shibuya, M. (1992). Two species of human CRK cDNA encode proteins with distinct biological activities. Mol Cell Biol *12*, 3482-3489.

Mayer, B. J., Hamaguchi, M., and Hanafusa, H. (1988). A novel viral oncogene with structural similarity to phospholipase C. Nature *332*, 272-275.

Mills, K. V., Lew, B. M., Jiang, S.-q., and Paulus, H. (1998). Protein splicing in trans by purified N- and C-terminal fragments of the Mycobacterium tuberculosis RecA intein. Proc Natl Acad Sci USA *95*, 3543-3548.

Mohammadi, F., Prentice, G. A., and Merrill, A. R. (2001). Protein-Protein Interaction Using Tryptophan Analogues: Novel Spectroscopic Probes for Toxin-Elongation Factor-2 Interactions. Biochemistry *40*, 10273-10283.

Mongiovi, A. M., Romano, P. R., Panni, S., Mendoza, M., Wong, W. T., Musacchio, A., Cesareni, G., and Di Fiore, P. P. (1999). A novel peptide-SH3 interaction. EMBO J *18*, 5300-5309.

Muir, T. W. (2003). Semisynthesis of proteins by expressed protein ligation. Annu Rev Biochem *72*, 249-289.

Muir, T. W., Sondhi, D., and Cole, P. A. (1998). Expressed protein ligation: a general method for protein engineering. Proc Natl Acad Sci U S A *95*, 6705-6710.

Muralidharan, V., Cho, J. H., Trester-Zedlitz, M., Kowalik, L., Chait, B. T., Raleigh, D. P., and Muir, T. W. (2004). Domain-specific incorporation of noninvasive optical probes into recombinant proteins. J Am Chem Soc *126*, 14004-14012.

Nicholls, A., Sharp, K. A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281-296.

Nilsson, B. L., Kiessling, L. L., and Raines, R. T. (2000). Staudinger ligation: a peptide from a thioester and azide. Org Lett *2*, 1939-1941.

Nojima, Y., Morino, N., Mimura, T., Hamasaki, K., Furuya, H., Sakai, R., Sato, T., Tachibana, K., Morimoto, C., Yazaki, Y., and et al. (1995). Integrin-mediated cell adhesion promotes tyrosine phosphorylation of p130Cas, a Src homology 3-containing molecule having multiple Src homology 2-binding motifs. J Biol Chem *270*, 15398-15402.

Noren, C. J., Wang, J. M., and Perler, F. B. (2000). Dissecting the chemistry of protein splicing and its applications. Angew Chem Intl Ed Engl *39*, 450-466.

Northey, J. G., Di Nardo, A. A., and Davidson, A. R. (2002). Hydrophobic core packing in the SH3 domain folding transition state. Nat Struct Biol *9*, 126-130.

Offer, J., Boddy, C. N. C., and Dawson, P. E. (2002). Extending synthetic access to proteins with a removable acyl transfer auxiliary. J Am Chem Soc *124*, 4642-4646.

Ogawa, S., Toyoshima, H., Kozutsumi, H., Hagiwara, K., Sakai, R., Tanaka, T., Hirano, N., Mano, H., Yazaki, Y., and Hirai, H. (1994). The C-terminal SH3 domain of the mouse c-Crk protein negatively regulates tyrosine-phosphorylation of Crk associated p130 in rat 3Y1 cells. Oncogene *9*, 1669-1678.

Ohta, Y., Itoh, S., Shigenaga, A., Shintaku, S., Fujii, N., and Otaka, A. (2006). Cysteine-Derived S-Protected Oxazolidinones: Potential Chemical Devices for the Preparation of Peptide Thioesters. Org Lett *8*, 467-470.

Ottesen, J. J., Blaschke, U. K., Cowburn, D., and Muir, T. W. (2003). Segmental Isotopic Labeling: Prospects for a new tool to study the structure-function relationships in multi-domain proteins. Biol Mag Res *20*, 35-51.

Ozawa, T., and Umezawa, Y. (2001). Detection of protein-protein interactions in vivo based on protein splicing. Current Opinion In Chemical Biology *5*, 578-583.

Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. Methods Enzymol *131*, 266-280.

Paulus, H. (2000). Protein splicing and related forms of protein autoprocessing. Annu Rev Biochem *69*, 447-496.

Pawson, T., and Nash, P. (2000). Protein-protein interactions define specificity in signal transduction. Genes Dev *14*, 1027-1047.

Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. Science *300*, 445-452.

Pawson, T., and Scott, J. D. (1997). Signaling through scaffold, anchoring, and adaptor proteins. Science *278*, 2075-2080.

Petosa, C., Schoehn, G., Askjaer, P., Bauer, U., Moulin, M., Steuerwald, U., Soler-Lopez, M., Baudin, F., Mattaj, I. W., and Muller, C. W. (2004). Architecture of CRM1/Exportin1 suggests how cooperativity is achieved during formation of a nuclear export complex. Mol Cell *16*, 761-775.

Plaxco, K. W., Guijarro, J. I., Morton, C. J., Pitkeathly, M., Campbell, I. D., and Dobson, C. M. (1998). The folding kinetics and thermodynamics of the Fyn-SH3 domain. Biochemistry *37*, 2529-2537.

Ramsay, G. D., and Eftink, M. R. (1994). Analysis of multidimensional spectroscopic data to monitor unfolding of proteins. Methods Enzymol *240*, 615-645.

Reddien, P. W., and Horvitz, H. R. (2000). CED-2/CrkII and CED-10/Rac control phagocytosis and cell migration in Caenorhabditis elegans. Nat Cell Biol *2*, 131-136.

Reichman, C., Singh, K., Liu, Y., Singh, S., Li, H., Fajardo, J. E., Fiser, A., and Birge, R. B. (2005). Transactivation of Abl by the Crk II adapter protein requires a PNAY sequence in the Crk C-terminal SH3 domain.

Oncogene.

Reichman, C. T., Mayer, B. J., Keshav, S., and Hanafusa, H. (1992). The Product Of The Cellular Crk Gene Consists Primarily Of Sh2 And Sh3 Regions. Cell Growth & Differentiation *3*, 451-460.

Ren, R., Ye, Z. S., and Baltimore, D. (1994). Abl protein-tyrosine kinase selects the Crk adapter as a substrate using SH3-binding sites. Genes Dev *8*, 783-795.

Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I. I., and Baker, D. (1999). Experiment and theory highlight role of native state topology in SH3 folding. Nat Struct Biol *6*, 1016-1024.

Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., Doshi, N., Grantcharova, V. P., Yi, Q., and Baker, D. (1997). Functional rapidly folding proteins from simplified amino acid sequences. Nat Struct Biol *4*, 805-809.

Rose, K. (1994). Facile Synthesis of Homogeneous Artificial Proteins. J Am Chem Soc *116*, 30-33.

Rosen, M. K., Yamazaki, T., Gish, G. D., Kay, C. M., Pawson, T., and Kay, L. E. (1995). Direct demonstration of an intramolecular SH2-phosphotyrosine interaction in the Crk protein. Nature *374*, 477-479.

Ross, J. B., Szabo, A. G., and Hogue, C. W. (1997). Enhancement of protein spectra with tryptophan analogs: fluorescence spectroscopy of protein-protein and protein-nucleic acid interactions. Methods Enzymol *278*, 151-190.

Ross, J. B. A., Rusinova, E., Luck, L. A., and Rousslang, K. W. (2000). Spectral enhancement of proteins by in vivo incorporation of tryptophan analogues. Topics in Fluorescence Spectroscopy *6*, 17-42.

Sato, S., Kuhlman, B., Wu, W. J., and Raleigh, D. P. (1999). Folding of the

multidomain ribosomal protein L9: the two domains fold independently with remarkably different rates. Biochemistry *38*, 5643-5650.

Sato, S., Luisi, D. L., and Raleigh, D. P. (2000). pH jump studies of the folding of the multidomain ribosomal protein L9: the structural organization of the N-terminal domain does not affect the anomalously slow folding of the C-terminal domain. Biochemistry *39*, 4955-4962.

Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulse field gradients. Prog Nuc Mag Res Spec *34*, 93-158.

Saxon, E., Armstrong, J. I., and Bertozzi, C. R. (2000). A "traceless" Staudinger ligation for the chemoselective synthesis of amide bonds. Org Lett *2*, 2141-2143.

Saxon, E., and Bertozzi, C. R. (2000). Cell surface engineering by a modified Staudinger reaction. Science *287*, 2007-2010.

Scalley-Kim, M., Minard, P., and Baker, D. (2003). Low free energy cost of very long loop insertions in proteins. Protein Sci *12*, 197-206.

Schaller, M. D., and Parsons, J. T. (1995). pp125FAK-dependent tyrosine phosphorylation of paxillin creates a high-affinity binding site for Crk. Mol Cell Biol *15*, 2635-2645.

Schlesinger, S. (1968). The Effect of Amino Acid Analogues on Alkaline Phosphatase Formation in Escherichia coli K-12. II. REPLACEMENT OF TRYPTOPHAN BY AZATRYPTOPHAN AND BY TRYPTAZAN. J Biol Chem *243*, 3877-3883.

Schnolzer, M., and Kent, S. B. (1992). Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease. Science *256*, 221-225.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A *95*, 5857-5864.

Senear, D. F., Mendelson, R. A., Stone, D. B., Luck, L. A., Rusinova, E., and Ross, J. B. (2002). Quantitative analysis of tryptophan analogue incorporation in recombinant proteins. Anal Biochem *300*, 77-86.

Senechal, K., Heaney, C., Druker, B., and Sawyers, C. L. (1998). Structural requirements for function of the Crk1 adapter protein in fibroblasts and hematopoietic cells. Mol Cell Biol *18*, 5082-5090.

Shen, F., Triezenberg, S. J., Hensley, P., Porter, D., and Knutson, J. R. (1996). Transcriptional activation domain of the herpesvirus protein VP16 becomes conformationally constrained upon interaction with basal transcription factors. J Biol Chem *271*, 4827-4837.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem *68*, 850-858.

Shin, Y., Winans, K. A., Backes, B. J., Kent, S. B. H., Ellman, J. A., and Bertozzi, C. R. (1999). Fmoc-based synthesis of peptide-(alpha)thioesters: Application to the total chemical synthesis of a glycoprotein by native chemical ligation. J Am Chem Soc *121*, 11684-11689.

Smith, J. J., Richardson, D. A., Kopf, J., Yoshida, M., Hollingsworth, R. E., and Kornbluth, S. (2002). Apoptotic regulation by the Crk adapter protein mediated by interactions with Wee1 and Crm1/exportin. Mol Cell Biol *22*, 1412-1423.

Soumillion P, J. L., Vervoort J, Fastrez J (1995). Biosynthetic incorporation of 7-azatryptophan into the phage lambda lysozyme: estimation of tryptophan accessibility, effect on enzymatic activity and protein stability. Protein Eng *8*, 451-456.

Southworth, M. W., Adam, E., Panne, D., Byer, R., Kautz, R., and Perler, F. B. (1998). Control of protein splicing by intein fragment reassembly. EMBO J *17*, 918-926.

Southworth, M. W., Amaya, K., Evans, T. C., Xu, M. Q., and Perler, F. B. (1999). Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein. Biotechniques *27*, 110-114, 116, 118-120.

Spiess, C., Meyer, A. S., Reissmann, S., and Frydman, J. (2004). Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. Trends in Cell Biology *14*, 598.

Steward, L. E., Collins, S.C., Gilmore, M.A., Carlson, J.E., Ross, J.B.A., Chamberlin, R.A. (1997). In Vitro Site-Specific Incorporation of Fluorescent Probes into b-Galactosidase. J Am Chem Soc *119*, 6-11.

Sun, H., Shen, Y., Dokainish, H., Holgado-Madruga, M., Wong, A., and Ireton, K. (2005). Host adaptor proteins Gab1 and CrkII promote InlB-dependent entry of Listeria monocytogenes. Cell Microbiol *7*, 443-457.

Swinnen, D., and Hilvert, D. (2000). Facile, Fmoc-compatible solid-phase synthesis of peptide C-terminal thioesters. Org Lett *2*, 2439-2442.

ten Hoeve, J., Morris, C., Heisterkamp, N., and Groffen, J. (1993). Isolation and chromosomal localization of CRKL, a human crk-like gene. Oncogene *8*, 2469-2474.

Tsarbopoulos, A., Karas, M., Strupat, K., Pramanik, B. N., Nagabhushan, T. L., and Hillenkamp, F. (1994). Comparative mapping of recombinant proteins and glycoproteins by plasma desorption and matrix-assisted laser desorption/ionization mass spectrometry. Anal Chem *66*, 2062-2070.

Twine, S. M., and Szabo, A. G. (2003). Fluorescent amino acid analogs. Methods Enzymol *360*, 104-127.

Tzeng, S. R., Pai, M. T., Lung, F. D., Wu, C. W., Roller, P. P., Lei, B., Wei, C. J., Tu, S. C., Chen, S. H., Soong, W. J., and Cheng, J. W. (2000). Stability and peptide binding specificity of Btk SH2 domain: molecular basis for X-linked agammaglobulinemia. Protein Sci *9*, 2377-2385.

Viguera AR, A. J., Musacchio A, Saraste M, Serrano L (1994). Characterization of the interaction of natural proline-rich peptides with five different SH3 domains. Biochemistry *33*, 10925-10933.

Viguera, A. R., Blanco, F. J., and Serrano, L. (1995). The order of secondary structure elements does not determine the structure of a protein but does affect its folding kinetics. J Mol Biol *247*, 670-681.

Viguera, A. R., Martinez, J. C., Filimonov, V. V., Mateo, P. L., and Serrano, L. (1994). Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition. Biochemistry *33*, 2142-2150.

Vriend, G. (1990). What If - a Molecular Modeling and Drug Design Program. Journal of Molecular Graphics *8*, 52-6.

Waldman, A. D., Clarke, A.R., Wigley, D.B., Hart, K.W., Chia, W.N., Barstow, D., Atkinson, T., Munro, I., Holbrook, J.J. (1987). The use of site-directed mutagenesis and time-resolved fluorescence spectroscopy to assign the fluorescence contributions of individual tryptophan residues in Bacillus stearothermophilus lactate dehydrogenase. Biochim Biophys Acta *913*, 66-71.

Wishart, D. S., and Sykes, B. D. (1994). The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR *4*, 171-180.

Wong, C.-Y., and Eftink, M. R. (1998a). Incorporation of Tryptophan Analogs into Staphylococcal Nuclease, Its V66W Mutant, and.DELTA.137-149 Fragment: Spectroscopic Studies. Biochemistry *37*, 8938-8946.

Wong, C.-Y., and Eftink, M. R. (1998b). Incorporation of Tryptophan Analogs into Staphylococcal Nuclease: Stability toward Thermal and Guanidine-HCl Induced Unfolding. Biochemistry *37*, 8947-8953.

Wood, D. W., Wu, W., Belfort, G., Derbyshire, V., and Belfort, M. (1999). A genetic system yields self-cleaving inteins for bioseparations. Nat Biotech *17*, 889.

Wu, H., Hu, Z., and Liu, X.-Q. (1998). Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. Proc Natl Acad Sci USA *95*, 9226-9231.

Wu, X., Knudsen, B., Feller, S. M., Zheng, J., Sali, A., Cowburn, D., Hanafusa, H., and Kuriyan, J. (1995). Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. Structure (London) *3*, 215-226.

Xu, M. Q., and Perler, F. B. (1996). The mechanism of protein splicing and its modulation by mutation. EMBO J *15*, 5146-5153.

Xu, R., Ayers, B., Cowburn, D., and Muir, T. W. (1999). Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. Proc Natl Acad Sci U S A *96*, 388-393.

Yamazaki, T., Formankay, J. D., and Kay, L. E. (1993). 2-Dimensional NMR Experiments for Correlating C-13-Beta and H-1-Delta/Epsilon Chemical-Shifts of Aromatic Residues in C-13-Labeled Proteins Via Scalar Couplings. J Am Chem Soc *115*, 11054-11055.

Yamazaki, T., Otomo, T., Oda, N., Kyogoku, Y., Uegaki, K., Ito, N., Ishino, Y., and Nakamura, H. (1998). Segmental Isotope Labeling for Protein NMR Using Peptide Splicing. J Am Chem Soc *120*, 5591-5592.

Yanofsky, C., and Horn, V. (1995). Bicyclomycin sensitivity and resistance affect Rho factor-mediated transcription termination in the tna operon of Escherichia coli. J Bacteriol *177*, 4451-4456.

Yeo, D. S., Srinivasan, R., Uttamchandani, M., Chen, G. Y., Zhu, Q., and Yao, S. Q. (2003). Cell-permeable small molecule probes for site-specific labeling of proteins. Chem Commun (Camb), 2870-2871.

Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1994). Backbone 1H and 15N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. J Biomol NMR *4*, 845-858.

Zvara, A., Fajardo, J. E., Escalante, M., Cotton, G., Muir, T., Kirsch, K. H., and Birge, R. B. (2001). Activation of the focal adhesion kinase signaling pathway by structural alterations in the carboxyl-terminal region of c-Crk II. Oncogene *20*, 951-961.

Publications related to Ph.D. thesis

Muralidharan, V., Cho, J. H., Trester-Zedlitz, M., Kowalik, L., Chait, B. T., Raleigh, D. P., and Muir, T. W. (2004). Domain-specific Incorporation of Noninvasive Optical Probes into Recombinant Proteins. J Am Chem Soc *126*, 14004-14012.

Muralidharan, V. and Muir, T. W. (2006). Protein Ligation: An Enabling Technology for the Biophysical Analysis of Proteins. Nat Methods *in review.*

Muralidharan, V., Dutta, K., Cho, J. H., Vila-Perello, M., Raleigh, D. P., Cowburn, D. and Muir, T. W. (2006). Solution Structure and Folding Characteristics of the C-terminal SH3 Domain of c-Crk-II. Biochemistry *in review.*