

Domain-Specific Incorporation of Noninvasive Optical Probes into Recombinant Proteins

Vasant Muralidharan,[†] Jaehyun Cho,[‡] Michelle Trester-Zedlitz,^{†,§,⊥} Lukasz Kowalik,[†] Brian T. Chait,[#] Daniel P. Raleigh,^{‡,§} and Tom W. Muir^{*,†}

Contribution from the Laboratory of Synthetic Protein Chemistry, The Rockefeller University, New York, New York 10021, Graduate Program in Biochemistry and Structural Biology, State University of New York at Stony Brook, Stony Brook, New York 11794-3400, Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, New York 10021, and Department of Chemistry, State University of New York at Stony Brook, Stony Brook, New York 11794-3400

Received June 8, 2004; E-mail: muirt@rockvax.rockefeller.edu

Abstract: An integrated approach is described that allows the domain-specific incorporation of optical probes into large recombinant proteins. The strategy is the combination of two existing techniques, expressed protein ligation (EPL) and in vivo amino acid replacement of tryptophans with tryptophan (Trp) analogues. The Src homology 3 (SH3) domain from the c-Crk-I adaptor protein has been labeled with a Trp analogue, 7-azatryptophan (7AW), using *Escherichia coli* Trp auxotrophs. Structural, biochemical, and thermodynamic studies show that incorporation of 7AW does not significantly perturb the structure or function of the isolated domain. Ligation of the 7AW-labeled SH3 domain to the c-Crk-I Src homology 2 (SH2) domain, via EPL, generated the multidomain protein, c-Crk-I, with a domain-specific label. Studies of this labeled protein show that the biochemical and thermodynamic properties of the SH3 domain do not change within the context of a larger multidomain protein. The technology described here is likely to be a useful tool in enhancing our understanding of the behavior of modular domains in their natural context, within multidomain proteins.

Introduction

Fluorescence spectroscopy is a powerful technique for studying various aspects of protein biophysics, such as protein folding, protein-protein interactions, and protein dynamics.^{1,2} 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.^{1,2} In proteins containing several Trp residues, 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.³ However, Trp residues are often part of the hydrophobic core of the protein; so, mutating them runs the risk of destabilizing the protein and/or affecting its function. Extrinsic fluorescent probes, such as dansyl or related

amino acid derivatives, can also be incorporated into proteins via modification of reactive residues, such as lysine and cysteine, or via chemical synthesis.^{4,5} This, again, carries the risk of destabilizing the protein and/or affecting its function. Moreover, external probes are usually not well suited for studying processes such as protein folding, due 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.⁶ This approach has the inherent advantage that Trp analogues are structurally similar to tryptophan and, in many instances, will have minimal effect on protein structure and function.⁷⁻⁹ Out of the several Trp analogues available, 7-azatryptophan (7AW) is ideally suited as an optical spectroscopic probe. 7AW has an absorption spectrum that is red-shifted by

[†] Laboratory of Synthetic Protein Chemistry, The Rockefeller University.

[‡] Graduate Program in Biochemistry and Structural Biology, State University of New York at Stony Brook.

[#] Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University.

[§] Department of Chemistry, State University of New York at Stony Brook.

[⊥] Current address: University of California, San Francisco, San Francisco, CA 94143.

(1) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 2nd ed.; Kluwer Academic/Plenum Publishers: New York, 1999.

(2) Beechem, J. M.; Brand, L. *Annu. Rev. Biochem.* **1985**, *54*, 43-71.

(3) Waldman, A. D.; Clarke, A. R.; Wigley, D. B.; Hart, K. W.; Chia, W. N.; Barstow, D.; Atkinson, T.; Munro, I.; Holbrook, J. J. *Biochim. Biophys. Acta* **1987**, *913*, 66-71.

(4) Mannuzzo, L. M.; Moronne, M. M.; Isacoff, E. Y. *Science* **1996**, *271*, 213-216.

(5) Cohen, B. E.; McAnaney, M. T.; Park, E. S.; Jan, Y. N.; Boxer, S. G.; Jan, L. Y. *Science* **2002**, *296*, 1700-1703.

(6) Schlesinger, S. *J. Biol. Chem.* **1968**, *243*, 3877-3883.

(7) Ross, J. B. A.; Szabo, A. G.; Hogue, C. W. *Methods Enzymol.* **1997**, *278*, 151-190.

(8) Ross, J. B. A.; Rusinova, E.; Luck, L. A.; Rousslang, K. W. *Top. Fluoresc. Spectrosc.* **2000**, *6*, 17-42.

(9) Twine, S. M.; Szabo, A. G. *Methods Enzymol.* **2003**, *360*, 104-127.

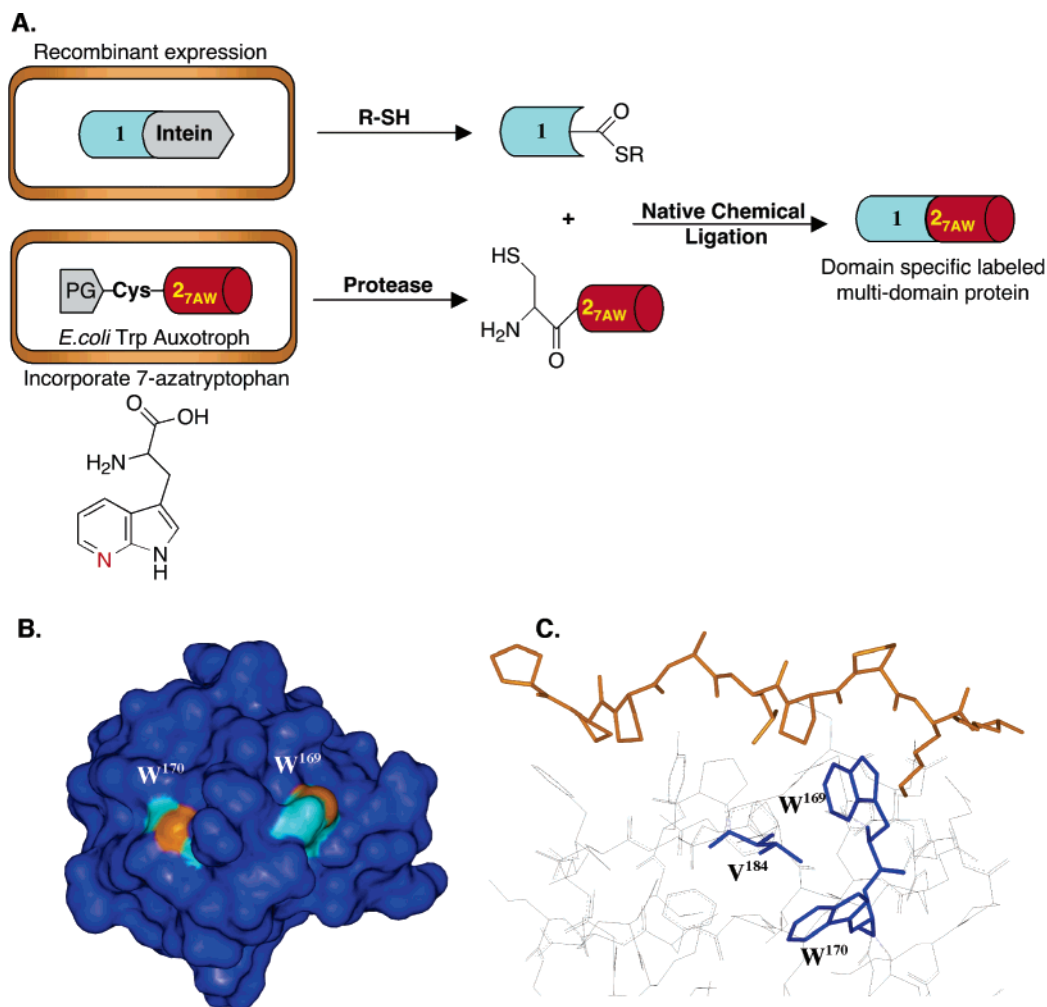


Figure 1. Domain-specific labeling of Crk-I. (A) Schematic representation of the domain-specific labeling procedure. PG = protecting group, Cys = cysteine, 7AW = 7-azatryptophan, 1 = domain 1, 2 = domain 2, 2_{7AW} = 7AW-labeled domain 2. (B) Structure of the SH3 domain of the c-Crk-I adapter protein (pdb code: 1CKA) shown with surface rendering (done using the Insight-II program), with the two Trp residues highlighted in cyan and the 7-positions (which are modified to an aza group in 7AW) of the Trp residues highlighted in orange. (C) Expanded view of the structure of the SH3 domain in complex with a poly-Pro peptide ligand (in orange) from the C3G protein (pdb code, 1CKA). Trp169, Trp170, and Val184 are highlighted in blue.

10–15 nm, relative to the absorption spectra of Trp, and an emission spectrum that is red-shifted by up to 50 nm. Thus, the labeled protein can be selectively excited and observed in the presence of Trp residues.⁷ 7AW is also highly sensitive to its environment,^{7,8} for example, allowing it to be used as a probe to study protein folding.^{10,11}

The use of *E. coli* Trp auxotrophs for labeling proteins results in the global replacement of all Trp residues in the protein with Trp analogues. This may be considered to be a disadvantage where one wants to study only a specific Trp residue among many, such as in the case of multidomain proteins. One solution to the problem might be nonsense suppression mutagenesis approaches.¹² In principle, these limitations can also be overcome using expressed protein ligation (EPL)^{13,14} in combination with the *in vivo* incorporation of non-natural Trp analogues using *E. coli* Trp auxotrophs. EPL is a protein semisynthesis approach that allows synthetic and/or recombinant polypeptides

to be chemoselectively and regioselectively joined together via a peptide bond.¹⁴ The approach takes advantage of two processes, native chemical ligation¹⁵ and protein splicing.¹⁶ 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. Importantly, EPL has no inherent size limitation, in terms of the target protein, and the final yields are comparable to recombinant protein expression.¹⁴ Indeed, EPL has been used to selectively label specific domains within large multidomain proteins with stable isotope probes for NMR.^{17–20} By extension, it should be possible to incorporate optical probes, such as 7AW, into recombinant proteins in a domain-specific manner (Figure 1A).

(10) Wong, C. Y.; Eftink, M. R. *Biochemistry* **1998**, *37*, 8938–8946.
 (11) Wong, C. Y.; Eftink, M. R. *Biochemistry* **1998**, *37*, 8947–8953.
 (12) Steward, L. E.; Collins, S. C.; Gilmore, M. A.; Carlson, J. E.; Ross, J. B. A.; Chamberlin, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 6–11.
 (13) Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6705–6710.
 (14) Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249–289.

(15) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776–779.
 (16) Noren, C. J.; Wang, J.; Perler, F. B. *Angew. Chem., Int. Ed.* **2000**, *39*, 450–466.
 (17) Xu, R.; Ayers, B.; Cowburn, D.; Muir, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 388–393.
 (18) Blaschke, U. K.; Cotton, G. J.; Muir, T. W. *Tetrahedron* **2000**, *56*, 9461–9470.
 (19) Ottesen, J. J.; Blaschke, U. K.; Cowburn, D.; Muir, T. W. *Biol. Magn. Reson.* **2002**, *20*, 35–51.
 (20) Camarero, J. A.; Shekhtman, A.; Campbell, E.; Chlenov, M.; Gruber, T. M.; Bryant, D.; Darst, S. A.; Cowburn, D.; Muir, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 8536–8541.

In the present study, we decided to apply this strategy to the multidomain adapter protein c-Crk-I.²¹ This protein represents a good model system for our studies because it contains two domains, an Src Homology 2 (SH2) domain followed by an Src Homology 3 (SH3) domain, each of which contains two Trp residues. We chose to concentrate our efforts on understanding the effect of context on the biochemical and thermodynamic properties of the SH3 domain of c-Crk-I. SH3 domains are small (~60 residues) modular proteins that have long been of interest for studying protein folding.^{22–26} This is because they all fold in a reversible two-state manner and have similar three-dimensional structures, yet, their sequence homology is low.

The folding and thermodynamic properties of the isolated Crk-I SH3 domain have been well studied.²⁷ However, nothing is known about the folding or thermodynamic properties of this SH3 domain in the context of the full-length protein, Crk-I. Indeed, this is generally the case for other SH3 domains, which are always part of larger proteins, such as kinases or adaptor proteins. Such studies have been hampered by the relative lack of techniques that can be used to elucidate such issues. In this study, we have developed a technique that allows us to gather domain-specific biophysical information from a multidomain context. Specifically, we have introduced the optical probe, 7AW, into the SH3 domain of Crk-I and used this probe to study the biochemical and thermodynamic properties of this domain in its native context. This methodology offers a general approach to the study of domains within multidomain proteins.

Results

Design and Synthesis of c-Crk-I[SH3_{7AW}]. Tryptophan analogues, such as 5-hydroxytryptophan, fluorotryptophans, and 7AW, allow the local environment of Trp residues in proteins to be studied using either NMR or fluorescence spectroscopy. In this study, we decided to replace the two Trp residues in the SH3 domain of c-Crk-I, 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 1B,C), Trp169 is involved in ligand binding whereas Trp170 is an important component of the hydrophobic core of the protein.²⁸ In both of these Trp residues, the 7-position is solvent exposed in the unliganded form of the protein (Figure 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 cysteine-containing 7AW-labeled SH3 domain (here-

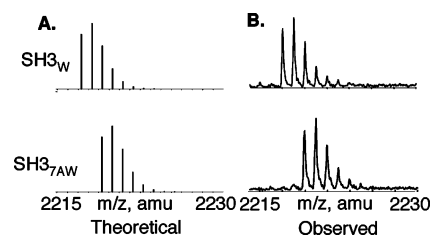


Figure 2. Calculated (assuming complete incorporation of 7AW) and observed isotope distributions of trypsin digest fragments of SH3_W and SH3_{7AW}. SH3_W peptide fragment: ¹⁶⁰IRDKPEEQWVNAEDSEGK¹⁷⁷ (expected mass = 2217.3 Da). SH3_{7AW} peptide fragment: ¹⁶⁰IRDKPEEQ-[7AW][7AW]VNAEDSEGK¹⁷⁷ (expected mass = 2219.3 Da). (A) Theoretical isotope distributions of the SH3_W peptide (top panel) and of the SH3_{7AW} peptide (bottom panel). (B) MALDI-TOF MS (reflectron) of the SH3_W peptide (top panel) and of the SH3_{7AW} peptide (bottom panel), with 4-hydroxycinnamic acid (4HCCA) as the matrix.

after referred to as SH3_{7AW}). Residues Gly124–Ser125, which lie within the short linker region (residues 118–132) 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 laboratory have shown that this mutation has no effect on the structure or function of the protein.^{18,19}

The SH2 domain of c-Crk-I (residues A1–G124) was generated under nonlabeling 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_{7AW} (residues C125–G206) was expressed as a His-tag fusion in an *E. coli* Trp auxotroph (*E. coli* CY15077²⁹) grown in minimal media supplemented with 7AW.⁶ Following cell lysis, the desired fusion protein was purified from the soluble fraction by Ni²⁺-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 ~1.5 mg/L. An unlabeled SH3 domain (hereafter referred to as SH3_W) was expressed under nonlabeling conditions using rich media and purified as above.

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 I160–K177). 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.³⁰ Figure 2 shows the theoretical and observed isotope distributions in the mass spectrum of the relevant peptide fragments from SH3_{7AW} and SH3_W. This comparison demonstrates that 7AW labeling in SH3_{7AW} was highly efficient ($\geq 93\%$ at both sites).

Initially, the SH3_{7AW} domain was also expressed as a gyrase A intein fusion. This intein contains a single Trp residue, which

- (21) Matsuda, M.; Tanaka, S.; Nagata, S.; Kojima, A.; Kurata, T.; Shibuya, M. *Mol. Cell. Biol.* **1992**, *12*, 3482–3489.
 (22) Viguera, A. R.; Martinez, J. C.; Filimonov, V. V.; Mateo, P. L.; Serrano, L. *Biochemistry* **1994**, *33*, 2142–2150.
 (23) Grantcharova, V. P.; Baker, D. *Biochemistry* **1997**, *36*, 15685–15692.
 (24) Prieto, J.; Wilms, M.; Jimenez, M. A.; Rico, M.; Serrano, L. *J. Mol. Biol.* **1997**, *268*, 760–778.
 (25) Martinez, J. C.; Serrano, L. *Nat. Struct. Biol.* **1999**, *6*, 1010–1016.
 (26) Riddle, D. S.; Grantcharova, V. P.; Santiago, J. V.; Alm, E.; Ruczinski, I.; Baker, D. *Nat. Struct. Biol.* **1999**, *6*, 1016–1024.
 (27) Camarero, J. A.; Fushman, D.; Sato, S.; Giriati, I.; Cowburn, D.; Raleigh, D. P.; Muir, T. W. *J. Mol. Biol.* **2001**, *308*, 1045–1062.
 (28) Wu, X.; Knudsen, B.; Feller, S. M.; Zheng, J.; Sali, A.; Cowburn, D.; Hanafusa, H.; Kuriyan, J. *Structure (London)* **1995**, *3*, 215–226.

- (29) Yanofsky, C.; Horn, V. J. *Bacteriol.* **1995**, *177*, 4451–4456.
 (30) Senechal, D. F.; Mendelson, R. A.; Stone, D. B.; Luck, L. A.; Rusinova, E.; Ross, J. B. *Anal. Biochem.* **2002**, *300*, 77–86.

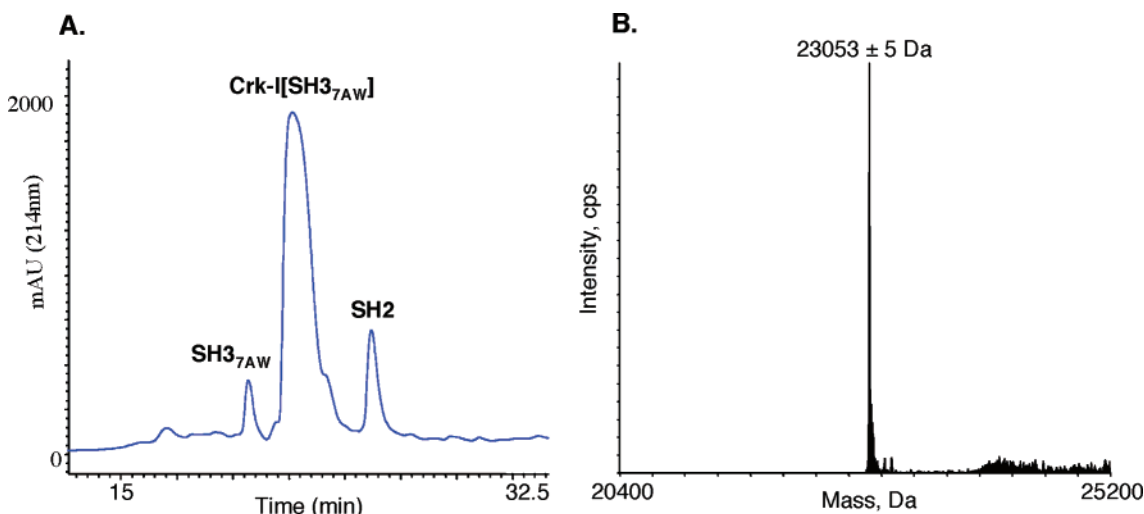


Figure 3. Semisynthesis of Crk-I[SH_{7AW}]. (A) RP-HPLC chromatogram of the crude mixture after ligation between the SH2 and the SH_{7AW} domains to give Crk-I[SH_{7AW}] after 5 days. The two starting materials and the product are indicated. (B) The reconstructed mass obtained from the charged species seen in the ESMS spectra of the ligation product, Crk-I[SH_{7AW}]. Expected mass = 23 048.5 Da.

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 degree similar to that for the wild-type intein (data not shown). 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.

With the recombinant building blocks in hand, Crk-I[SH_{7AW}] was then assembled via EPL. The ligation reaction was initiated by mixing the purified SH2 α -thioester and SH_{7AW} domains in the presence of 6 M guanidinium chloride. The use of denaturing conditions was found to increase the yields in this system; however, in general, the ligation approach can be performed under native conditions.¹⁴ The reaction was monitored by RP-HPLC, which indicated near-complete conversion after 5 days (Figure 3A). The ligation product, Crk-I[SH_{7AW}], was then purified by RP-HPLC and characterized by ESMS (Figure 3). Purified Crk-I[SH_{7AW}] was refolded as described previously.¹⁸

Structural Characterization of SH_{3W}, SH_{37AW}, and Crk-I[SH_{7AW}]. A combination of fluorescence, CD, and NMR spectroscopies were used to gauge the affect of 7AW labeling on the folding properties of the Crk-I SH3 domain. The fluorescence emission spectra of SH_{3W}, SH_{37AW}, and Crk-I[SH_{7AW}] upon excitation at 310 nm are shown in Figure 4A. SH_{37AW} and Crk-I[SH_{7AW}] exhibited broad emission spectra with maxima at 388 and 377 nm, respectively, whereas SH_{3W} was essentially nonfluorescent 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 SH_{37AW} and Crk-I[SH_{7AW}] suggests that 7AW sees different local environments depending on whether the SH3 domain is in the context of the Crk-I.

The secondary structures of the SH_{3W} and SH_{37AW} domains were studied by far-UV CD spectroscopy. Consistent with a previous report from our group,²⁷ the CD spectrum of SH_{3W} exhibited minima at 203 and 229 nm and a maxima at 219 nm (Figure 4B). These unusual spectral features have been seen in

other SH3 domains³¹ and are likely due to aromatic contributions.^{27,32} The CD spectrum of SH_{37AW} differs slightly from that of SH_{3W} (Figure 4B); 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 to 224 nm), possibly reflecting the differences in the aromaticity of 7AW versus Trp. Importantly, the CD spectra of SH_{3W} and SH_{37AW} are otherwise similar, indicating that the incorporation of 7AW does not cause major structural changes in the domain. Furthermore, we determined the midpoint (T_m) of the temperature denaturation of SH_{3W} and SH_{37AW} by following the change in the CD signal at 220 nm (Figure 4C). Thermal denaturation of both SH_{3W} and SH_{37AW} showed a two-state behavior and was reversible. The T_m values of the SH_{3W} and SH_{37AW} 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.

Homonuclear ¹H NMR spectroscopy was used to further investigate the effects of 7AW incorporation. The NOESY spectra of SH_{3W} and SH_{37AW} are very similar (Figure 5A,B). 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 SH_{3W} and SH_{37AW} domains are the presence of NOESY cross-peaks between Trp170 and Val184 and upfield chemical shifts at around -0.85 ppm for the γ -methyl protons of Val184 (Figure 5C). 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 1C).²⁸ Thus, the CD and homonuclear NMR data suggest that the SH_{3W} and SH_{37AW} domains share a common fold, and that the incorporation of 7AW is well tolerated.

Ligand-Binding Properties of SH_{3W}, SH_{37AW}, and Crk-I[SH_{7AW}]. The ligand-binding properties of SH_{3W}, SH_{37AW}, and Crk-I[SH_{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

(31) Viguera, A. R.; Blanco, F. J.; Serrano, L. *J. Mol. Biol.* **1995**, *247*, 670–681.

(32) Manning, M. C.; Woody, R. W. *Biochemistry* **1989**, *28*, 8609–8613.

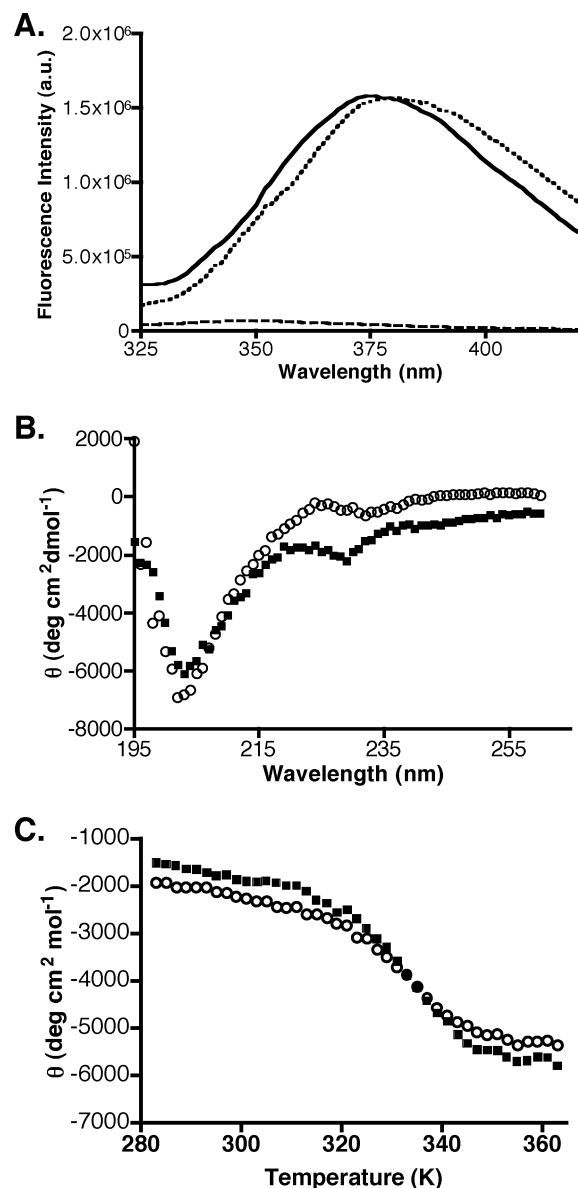


Figure 4. Spectroscopic characterization of SH3_W, SH3_{7AW}, and Crk-I[SH3_{7AW}]. (A) Fluorescence emission spectra of SH3_W (---), SH3_{7AW} (---), and Crk-I[SH3_{7AW}] (—). The excitation wavelength was 310 nm. (B) Far-UV CD spectra of SH3_W (■) and SH3_{7AW} (○). (C) Thermal denaturation curves of SH3_W (■) and SH3_{7AW} (○) domains. The CD signal was recorded at 220 nm.

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 cases of SH3_{7AW} and Crk-I[SH3_{7AW}], as a function of added ligand to the protein (Figure 6).

As shown in Table 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 1 also reveal that the ligand-binding activity of the SH3 domain does not change

Table 1. Dissociation Constants (K_d) of the Various SH3 Constructs for the Poly-Pro Ligand from C3G (H-PPPLPPKRRK[ϵ -fluorescein]G-NH₂)^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 of the experiments were performed at 25 °C in 20 mM sodium phosphate and 100 mM sodium chloride at pH 7.0. The numbers following the ± represent the standard errors of the fit.

Table 2. Thermodynamics of Unfolding for the Various Constructs^a

protein	$\Delta G^0_{H_2O}$ (kcal mol ⁻¹)	m (kcal mol ⁻¹ M ⁻¹)
SH2	8.03 ± 0.4	3.16 ± 0.1
SH3 _W	3.90 ± 0.4	2.62 ± 0.2
SH3 _{7AW}	3.24 ± 0.3	2.32 ± 0.1
Crk-I[SH3 _{7AW}]	3.66 ± 0.4	2.54 ± 0.2

^a All of the experiments were performed at 25 °C in 20 mM sodium phosphate and 100 mM sodium chloride at pH 7.0. The numbers following the ± represent the standard errors of the fit.

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.

Equilibrium Stabilities of SH2, SH3_W, 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 tryptophan or 7AW residues upon the addition of chemical denaturants. All of the constructs exhibited well-defined unfolding curves (Figure 7), and from this, we elucidated the equilibrium stabilities ($\Delta G^0_{H_2O}$) and m values for all of the proteins (Table 2). The equilibrium stabilities of SH3_W matched closely with the published values for this domain.²⁷

The $\Delta G^0_{H_2O}$ value for the SH2 domain reveals that this protein is unusually stable (Table 2).^{33,34} The estimated values of $\Delta G^0_{H_2O}$ for SH3_W and SH3_{7AW} are similar to those of other SH3 domains^{23,35,36,50} and to each other. The fact that the SH3_W and SH3_{7AW} domains have similar stabilities validates the labeling

- (33) 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.; Cheng, J. W. *Protein Sci.* **2000**, *9*, 2377–2385.
- (34) Scalley-Kim, M.; Minard, P.; Baker, D. *Protein Sci.* **2003**, *12*, 197–206.
- (35) Lim, W. A.; Fox, R. O.; Richards, F. M. *Protein Sci.* **1994**, *3*, 1261–1266.
- (36) Rath, A.; Davidson, A. R. *Protein Sci.* **2000**, *9*, 2457–2469.
- (37) Soumillion, P.; Jaspers, L.; Vervoort, J.; Fastrez, J. *Protein Eng.* **1995**, *8*, 451–456.
- (38) Shen, F.; Triezenberg, S. J.; Hensley, P.; Porter, D.; Knutson, J. R. *J. Biol. Chem.* **1996**, *271*, 4827–4837.
- (39) Mohammadi, F.; Prentice, G. A.; Merrill, A. R. *Biochemistry* **2001**, *40*, 10273–10283.
- (40) There are a few examples of multidomain proteins in which Trp is present in one domain (refs 41 and 42).
- (41) Sato, S.; Kuhlman, B.; Wu, W. J.; Raleigh, D. P. *Biochemistry* **1999**, *38*, 5643–5650.
- (42) Sato, S.; Luisi, D. L.; Raleigh, D. P. *Biochemistry* **2000**, *39*, 4955–4962.
- (43) Knudsen, B. S.; Feller, S. M.; Hanafusa, H. *J. Biol. Chem.* **1994**, *269*, 32781–32787.
- (44) Cotton, G. J.; Ayers, B.; Xu, R.; Muir, T. W. *J. Am. Chem. Soc.* **1999**, *121*, 1100–1101.
- (45) Sato, S.; Xiang, S.; Raleigh, D. P. *J. Mol. Biol.* **2001**, *312*, 569–577.
- (46) Krutchinsky, A. N.; Zhang, W.; Chait, B. T. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 493–504.
- (47) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1996**, *68*, 850–858.
- (48) Tzeng, S. R.; Karas, M.; Strupat, K.; Pramanik, B. N.; Nagabhushan, T. L.; Hillenkamp, F. *Anal. Chem.* **1994**, *66*, 2062–2070.
- (49) Anafi, M.; Rosen, M. K.; Gish, G. D.; Kay, L. E.; Pawson, T. *J. Biol. Chem.* **1996**, *271*, 21365–21374.
- (50) Viguera, A. R.; Arrondo, J. L.; Musacchio, A.; Saraste, M.; Serrano, L. *Biochemistry* **1994**, *33*, 10925–10933.

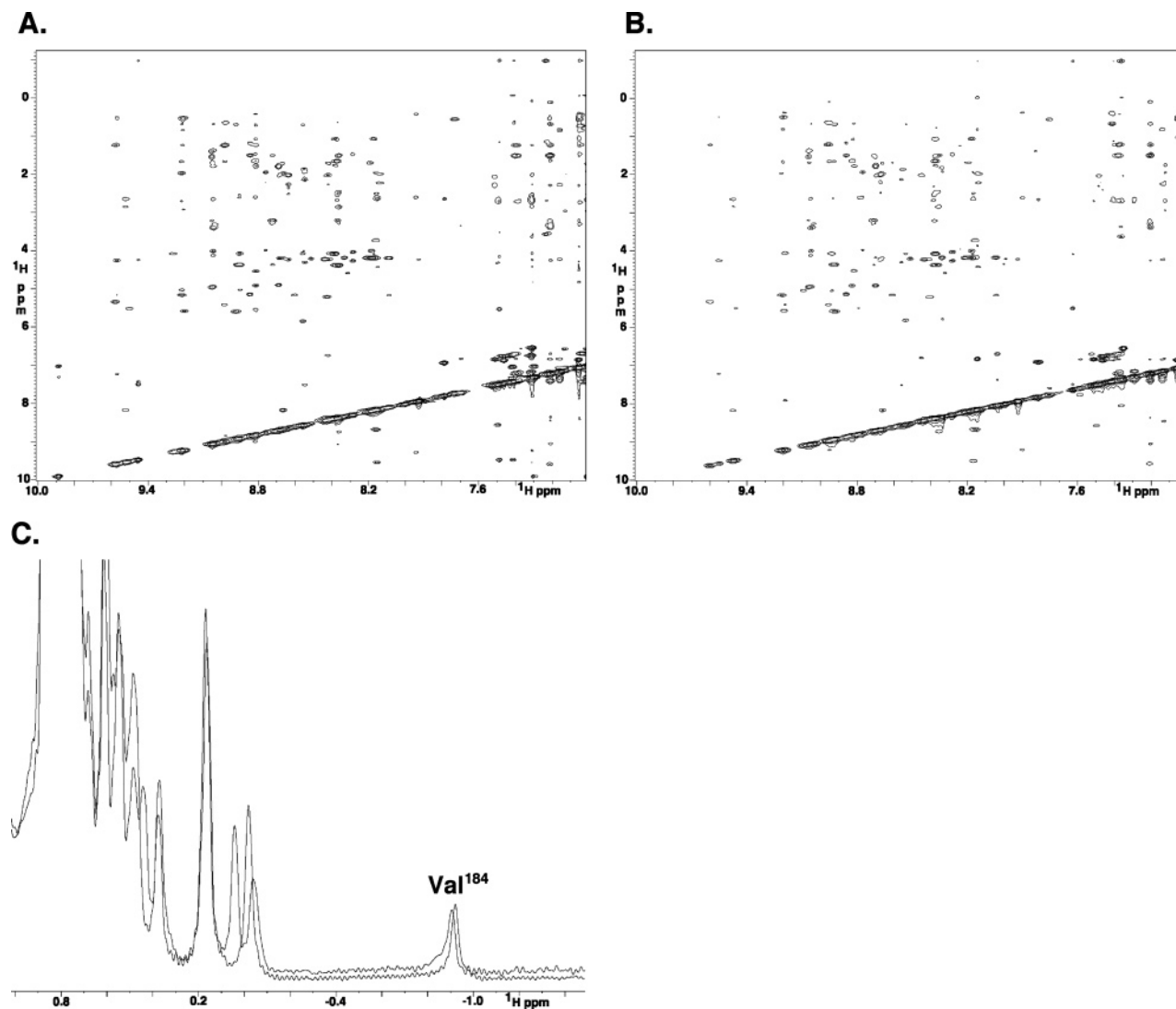


Figure 5. Amide and aromatic regions of the ^1H NOESY spectra at 500 MHz of (A) SH3_W, (B) SH3_{7AW}, and (C) the overlay of the ^1H NMR spectra at 400 MHz of SH3_W (in gray) and SH3_{7AW} (in black). The signal corresponding to the γ -methyl protons of Val184 are indicated.

procedure and provides further evidence that the labeled and the unlabeled domains are comparable. Analysis of the $\Delta G^0_{\text{H}_2\text{O}}$ values reveals that the SH3 domain, when in the context of c-Crk-I, has a stability similar to that of 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 multidomain protein.

Discussion

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 to be a particularly useful strategy for uniformly labeling proteins with spectroscopic probes.^{8–11,37–39} 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 multidomain

proteins.^{40–42} In this work, we have addressed this problem 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, similar to most other SH3 domains, and has been studied extensively with regards to its biochemical,⁴³ structural,²⁸ and biophysical properties.²⁷ Its folding properties are similar to those of other SH3 domains, such as the src SH3 and α -spectrin SH3 domains.²⁷ These attributes make the c-Crk-I SH3 domain an attractive model system for exploring the effect of context on the functional and thermodynamic properties of domains.

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 ($\geq 93\%$) at both sites, as indicated by mass spectrometric analysis of tryptic peptides (Figure 2). Such high efficiency has been seen before in proteins with two or more Trp residues, though the exact reason for this remains unknown.³⁰ Analysis of the structure of the c-Crk-I SH3 domain reveals that the 7-position of both Trp residues points

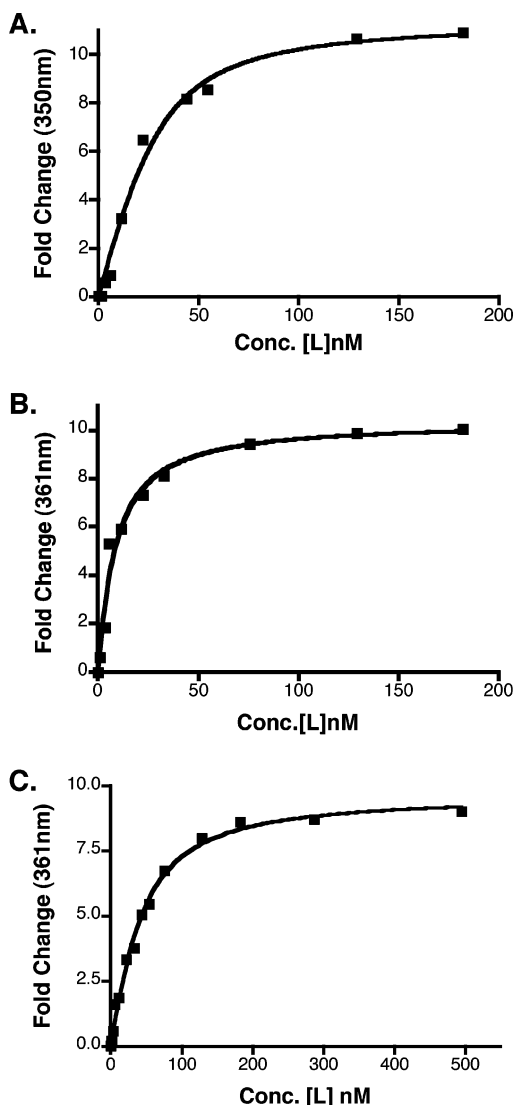


Figure 6. Binding isotherms of the poly-Pro ligand of C3G to (A) SH3_w, (B) SH3_{7AW}, and (C) Crk-I[SH3_{7AW}], as determined by a change in either tryptophan or 7AW fluorescence.

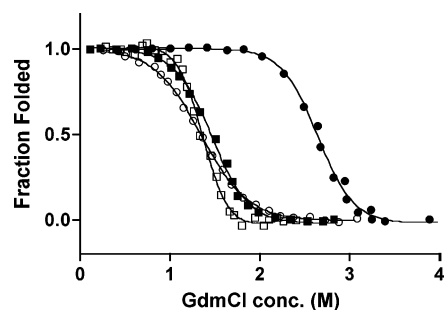


Figure 7. Chemical denaturation of SH3_w (■), the SH2 domain (●), SH3_{7AW} (○), and c-Crk-I[SH3_{7AW}] (□) followed by fluorescence spectroscopy. The solid curve represents the best fit for eqs 3 and 4.

out into solution. Hence, the incorporation of 7AW was not expected to cause structural perturbation in the system (Figure 1B). Indeed, CD and homonuclear ¹H NMR spectroscopies indicate that the SH3_w and SH3_{7AW} domains have the same global fold (Figures 4B and 5). This conclusion is also supported by thermal (Figure 4C) and chemical (Figure 7) denaturation experiments, which indicate that the labeled and unlabeled proteins have similar thermodynamic stabilities. Finally, the

dissociation constants of the SH3_w and SH3_{7AW} domains, for the peptide ligand used in this study, are essentially the same (Figure 6 and Table 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 there are any. Since SH3 domains from different proteins have very similar structures, we believe that this strategy will prove useful for studying SH3 domains present in other multidomain proteins.

Expressed protein ligation was successfully used for preparing Crk-I[SH3_{7AW}] from its component domains. The ligation reaction was extremely efficient (Figure 3A), allowing multi-milligram amounts of the labeled protein to be isolated. Although it is not relevant for the synthesis of Crk-I[SH3_{7AW}], our studies also indicate that the gyrase A intein is functional when its lone Trp residue is replaced by 7AW. 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 internal⁴⁴ domains of large proteins.

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 1 and 2). Interestingly, the emission maxima for SH3_{7AW} and Crk-I[SH3_{7AW}] differ by around 13 nm (Figure 4A). 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). Indeed, the free energy of unfolding is significantly higher than that reported for other SH2 domains ($\Delta G^0_{\text{H}_2\text{O}} \sim 3$ kcal mol⁻¹).^{33,34} As can be seen in Figure 7, 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 nonstructured linker regions, the domains fold independently of each other. This type of independent folding behavior has been observed in one other two-domain protein, namely, the ribosomal protein, L9; however, in this case, the two domains are linked by a fairly rigid α -helix.^{41,42,45}

Conclusions

We have developed an approach that allows domain-specific incorporation of noninvasive tryptophan analogues within modular, multidomain 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-I. Comparison of the ligand-binding and thermodynamic properties of this domain in isolation versus in the context of c-Crk-I indicates that there are little, if any, interdomain interactions in the protein. Our domain-specific labeling procedure should be applicable to many other multidomain systems, thereby providing the means to dissect the folding properties of these complex systems.

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 Company (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 and 280 nm detection. The column used for analytical RP-HPLC was a Vydac C18 column with a constant flow rate of 1 mL/min. Preparative RP-HPLC was performed on a Waters DeltaPrep 4000 system using a Vydac C18 semipreparative 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) and 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 Biosystem, Foster City, CA) or a modified Sciex prototype QqTOF (Centaur) instrument.⁴⁶ 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.¹⁸ For construction of the plasmid, pTrcHis-Xa-SH3, which encodes a factor Xa cleavage sequence between a poly-His affinity tag and the Crk-I SH3 domain (residues C125–G206; note that the numbering refers to the full-length murine c-Crk sequence without the initial methionine residue), the appropriate PCR product was cloned into a pTrcHis plasmid (Invitrogen) using *Xho*I and *Hind*III restriction sites. The resulting plasmid was shown to be free of mutations in the protein encoding region by DNA sequencing. For expression of the unlabeled SH3 domain, SH3_w, the *E. coli* Trp auxotrophic strain CY15077 (provided by the Yale University Genetic Stock Center²⁹), transformed with pTrcHis-Xa-SH3, was grown to a mid-log phase in Luria–Bertani medium containing 100 μ g/mL ampicillin. Protein expression was induced with 1 mM IPTG at 37 °C for 6 h, after which cells were harvested and lysed by passage through a French press. The His-tagged fusion protein was purified by affinity chromatography using Ni²⁺-NTA resin (Novagen) and dialyzed into Factor Xa cleavage buffer (100 mM sodium phosphate (pH 7.2), 100 mM NaCl, and 1 mM CaCl₂). The fusion protein was then cleaved with Factor Xa (Amersham Biosciences) using 100 units of protease/milligram of protein for 8–10 h. The reaction mixture was quenched by the addition of 20 mM DTT, and the cleaved protein was further purified by semipreparative RP-HPLC using a linear gradient of 25–45% solvent B for 60 min. The purified protein was characterized by ESMS as SH3_w (expected mass = 9429.5 Da; observed mass = 9430.4 \pm 3.2 Da).

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 L of Luria–Bertani (LB) broth containing 100 μ g/mL ampicillin. The cells were grown at 37 °C to an A₆₀₀ of 0.9–1.0, harvested, and then washed with M9 minimal

media supplemented with 0.1 mM CaCl₂, 1 mM MgSO₄, 0.5% glucose, 0.1% thiamine, 100 μ g/mL ampicillin, and 0.6% glycerol. The culture was then resuspended in 6 L of M9 minimal media, supplemented as described above, and allowed to grow for another 40 min to deplete the cellular stores of Trp. Protein expression was then induced with the addition of 1 mM IPTG and 80 μ g/mL of D,L-7AW (Sigma), and the cells were then harvested after 4 h of growth. Subsequent purification and cleavage protocols were the same as described above for SH3_w. Purified SH3_{7AW} was characterized by ESMS (expected mass = 9431.5 Da; observed mass = 9433.3 \pm 1.9 Da). The overall yield of the SH3_{7AW} domain was 1.5–2 mg/L of culture.

Semisynthesis 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 (pH 7.2), 100 mM NaCl, 6 M guanidinium chloride, 2% MESNA, 2% ethanethiol, and 1 mM CaCl₂. Ligation reactions were allowed to proceed for 4–5 days at 25 °C, while the mixture was constantly stirred. The product, Crk-I[SH3_{7AW}] (residues A1–G206), was purified by semipreparative RP-HPLC using a linear gradient of 35–45% solvent B over 60 min and characterized by ESMS (Figure 3B). 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}.

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.⁴⁷ Extractions of the digest were made as described in Shevchenko et al.⁴⁷ The peptides thus derived were then prepared for MALDI-MS using 4-hydroxycinnamic acid (4HCCA) as the matrix. The ultrathin layer method⁴⁸ 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 were smoothed and calibrated using Data Explorer (PE Biosystem).

Fluorescence and Circular Dichroism Spectroscopy. Fluorescence spectra of all of the proteins were recorded using a Spex Fluorolog-3 spectrofluorimeter. The spectra were recorded at 25 °C using an excitation wavelength of 310 nm and 5 nm slit widths. In all cases, protein concentration was approximately 1 μ M, and the protein samples were prepared in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl buffer. CD spectra were recorded on an Aviv 62DS spectropolarimeter using a 0.1 cm cuvette. Protein concentrations were 50 μ M in a buffer containing 5 mM sodium phosphate and 20 mM NaCl buffer (pH 7.2). Spectra were recorded at 25 °C in 0.5 nm steps from 260 to 195 nm and averaged over 4 s at each wavelength. Protein concentrations were determined by UV absorption (λ = 280 nm: SH3_w domain, ϵ = 15 400 M⁻¹ cm⁻¹; SH3_{7AW} domain, ϵ = 22 840 M⁻¹ cm⁻¹).

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) ²H₂O, and 0.1% (v/v) NaN₃ to a final concentration of 0.5–1.0 mM. ¹H NMR experiments (1D and 2D) were performed on Bruker DPX-400 and DMX-500 spectrometers at 25 °C. Mixing times of 90 ms for the TOCSY measurements and 150 ms for the NOESY measurements were used. Spectra were assigned based on those values reported by Anafi et al.⁴⁹ for the Crk SH3 domain in the context of murine c-Crk and on assignments provided by David Fushman (personal communication). The spectra were analyzed using XWINNMR (Bruker Instruments).

Ligand-Binding Assays. A fluorescence-based titration assay was used to measure the affinity constants of the SH3_w domain, the SH3_{7AW} domain, and the Crk-I[SH3_{7AW}] protein for a peptide ligand. All of the experiments were carried out at 25 °C in a stirred 1 cm path length cuvette using a Spex Fluorolog-3 spectrofluorimeter. 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 \sim 0.1 μ M in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl. The ligand

used in these studies, H-PPPPLPPKRRK[ϵ -fluorescein]G-NH₂, was based on a poly-proline sequence from the C3G protein⁴³ 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 complex⁵⁰ and by fitting the corrected fluorescence intensities (using nonlinear least-squares analysis, Graph-Pad Prism v4.0a) to the equation:⁵¹

$$F = [\Delta F] [P + K_d + L - \{(P + K_d + L)^2 - 4PL\}^{0.5}] [2P]^{-1} \quad (1)$$

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 of the experiments were repeated three times. Protein and peptide concentrations were determined by UV absorption (SH3 and SH3_{7AW} domains as described above: Crk-I[SH3_{7AW}] ($\lambda = 280$ nm), $\epsilon = 33\,000$ M⁻¹ cm⁻¹; ligand ($\lambda = 495$ nm), $\epsilon = 83\,000$ 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 s. The temperature was varied in 2 °C steps, in both directions, between 10 and 90 °C with a 5 min incubation at each temperature. Protein concentrations were maintained at 50 μ M and determined as described above. Protein samples were dissolved in 5 mM sodium phosphate (pH 7.2) and 20 mM NaCl. The CD signal at each temperature was fit to the following equation to determine the T_m values:⁵²

$$Y_T = 1[1 + e^{\{-\Delta H(1-T/T_m)/(RT)\}}]^{-1} Y_N + [1 - 1[1 + e^{\{-\Delta H(1-T/T_m)/(RT)\}}]^{-1}] Y_U \quad (2)$$

where Y_T is the CD signal at temperature T . ΔH is the enthalpy of unfolding. T_m is the midpoint 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 path length cuvette, while the mixture was constantly stirred. The fluorescence signals were recorded at 355 nm for the SH2 domain and 340 nm for the SH3_w domain, and 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} and at 388 nm for Crk-I[SH3_{7AW}]. Fluorescence signals were averaged for 60 s at each guanidinium concentration after a 5 min equilibration period, while the mixture was constantly stirred. Protein concentration was either 1 μ M (for SH2, SH3_w, and SH3_{7AW}) or 0.5 μ M (for Crk-I[SH3_{7AW}]) in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl. Guanidinium chloride (GdmCl) concentrations were determined by measuring the refractive index of the solution. The measured fluorescence signals at each GdmCl concentration were fit to the following equation:⁵³

$$F_{\text{GdmCl}} = \{\alpha_N + \beta_N[\text{GdmCl}] + (\alpha_D + \beta_D[\text{GdmCl}]) e^{-\Delta G_U[\text{GdmCl}]/RT}\} \{1 + e^{-\Delta G_U[\text{GdmCl}]/RT}\}^{-1} \quad (3)$$

where

$$\Delta G_U[\text{GdmCl}] = \Delta G_{\text{H}_2\text{O}}^0 - m[\text{GdmCl}] \quad (4)$$

F_{GdmCl} is the measured fluorescence signal. α_N , β_N , α_D , and β_D are parameters that define the fluorescence signals of the native state (N) and the denatured state (D) as a function of [GdmCl]. $\Delta G_{\text{H}_2\text{O}}^0$ is Gibb's free energy for unfolding in the absence of GdmCl. Fraction folding at a given GdmCl concentration was determined by using F_{GdmCl} , α_N , β_N , α_D , and β_D .

Acknowledgment. We thank Jennifer Ottesen for helpful suggestions on the preparation of the proteins and Yufeng Wei for help with the NMR spectroscopy. This work was supported in part by the National Health Institutes, Grants GM55843-07 (T.W.M.) and RR00862 (B.T.C.).

JA0466199

(51) Lew, J.; Coruh, N.; Tsigelny, I.; Garrod, S.; Taylor, S. S. *J. Biol. Chem.* **1997**, *272*, 1507–1513.

(52) Ramsay, G. D.; Eftink, M. R. *Methods Enzymol.* **1994**, *240*, 615–645.

(53) Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266–280.