Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Post-translational modification of the N-terminal His tag interferes with the crystallization of the wild-type and mutant SH3 domains from chicken src tyrosine kinase

Kristine M. Kim,[a]† Eugene C. Yi,[b] David Baker[c] and Kam Y. J. Zhang[a]*

[a]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA, [b]The Institute for Systems Biology, Seattle, 4225 Roosevelt Way NE, Seattle, Washington 98105, USA, and [c]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA

† Current address: Immunex Corporation, 51 University Street, Seattle, Washington 98101, USA.

Correspondence e-mail: kzhang@fhcrc.org

Received 9 November 2000
Accepted 12 February 2001

Structural studies of the wild type and mutants of the src SH3 domain were initiated to elucidate the correlation of the native-state topology with protein thermostability and folding kinetics. An extra mass of 178 Da arising from the post-translational modification at the N-terminal His tag was observed. The spontaneous α-*N*-6 gluconoylation at the amino group of the His-tagged SH3 domain contributed to the observed extra mass. The partial modification of the N-terminal His-tag produced heterogeneity, both in size and in charge, in the *Escherichia coli* expressed SH3 domain. The removal of the His tag from the SH3 domain was essential for the crystallization of both wild-type and mutant src SH3. Both the wild type and the W43I mutant were crystallized by hanging-drop vapor diffusion and are in the hexagonal space group $P6_522$ with one molecule in the asymmetric unit. Data sets were collected to 1.8 and 1.95 Å resolution for the the wild type and the W43I mutant, respectively.

1. Introduction

Protein structures are determined by the non-bonded interatomic interactions. The relative strength of these interactions can be estimated by introducing mutations into protein sequences and characterizing their effects on protein thermodynamic stability (Fersht, 1988; Leatherbarrow & Fersht, 1986). Carefully selected mutations that eliminate targeted interactions have revealed a wealth of information regarding the effect of amino-acid substitution on stability (Jackson *et al.*, 1993; Serrano *et al.*, 1990). The design and interpretation of mutations can be based on the wild-type structure only if the structural change triggered by the mutation is minimum, local and predictable. However, some single amino-acid substitutions can cause substantive local arrangement of side chains and even trigger global conformation changes that are difficult to predict based on the wild-type structure. Under these circumstances, the three-dimensional structural information of the mutants is required in order to obtain the energetics of the non-bonded interatomic interactions based on thermodynamic stability data.

Small globular proteins are ideal systems for the investigation of the effect of a single amino-acid substitution on protein thermostability and folding kinetics. The SH3 domain from chicken src tyrosine kinase (src SH3) is an ideal simple system for such studies. The folding kinetics and thermostability of the src SH3 domain have been well characterized (Grantcharova *et al.*, 1998; Riddle *et al.*, 1997, 1999; Tsai *et al.*, 1999). Recently, the effects of systematic point mutations of src SH3 (mutation of 52 of 57 residues) on the protein folding and unfolding rates and protein stability have been reported (Riddle *et al.*, 1999). We have initiated the structural studies of wild type and a series of mutant src SH3 domains to obtain a measure of non-bonded interatomic interactions by combining the structural information with the thermodynamic and kinetic data.

We report here the purification, crystallization and preliminary X-ray diffraction studies of wild-type and mutant src SH3. In addition, the characterization of the post-translational modification with an extra 178 Da mass at the N-terminal His tag of SH3 and its effect on crystallization will also be described.

2. Methods

2.1. Protein expression and purification

Cloning of the chicken src SH3 gene in plasmid pET-15b (Novagen) and the expression of proteins in BL21(DE3)pLysS have been described previously (Grantcharova *et al.*, 1998). The overexpressed src SH3 contained a cleavable histidine (His) tag at the N-terminus of the protein. All protein purification was carried out at 277 K. Typically, cell pellets from 6 l culture were suspended in 1:5(*w*/*v*) PEB (20 m*M* Tris pH 8.0, 0.2 *M* NaCl, 5 m*M* imidazole, 0.5 m*M* PMSF, 2.5 m*M* benzamidine) containing DNaseI (10 µg ml^{-1}), lysozyme (0.2 mg ml^{-1}) and streptomycin sulfate [0.1%(*w*/*v*)]. Cells were lysed by pulse sonication and the cell debris was removed by

centrifugation (30 min at 31 000*g*). The supernatant was loaded onto an Ni-NTA Sepharose (Qiagen) column pre-equilibrated with PEB and washed with PEB containing 40 m*M* imidazole. Proteins were eluted with a linear gradient of 40–150 m*M* imidazole over six column volumes. Pooled proteins were dialyzed against DB (20 m*M* Tris pH 8, 200 m*M* NaCl, 0.2 m*M* PMSF) and concentrated to about 15 mg ml^{-1}. The N-terminal His tag was cleaved using thrombin and was removed using Ni-NTA Sepharose. The protein was then concentrated and fractionated on a Superdex 75 gel-filtration column (Pharmacia Biotech Inc.) equilibrated with DB to remove residual thrombin. Proteins were concentrated using disc membrane YM3 under pressurized N_2 (Amicon).

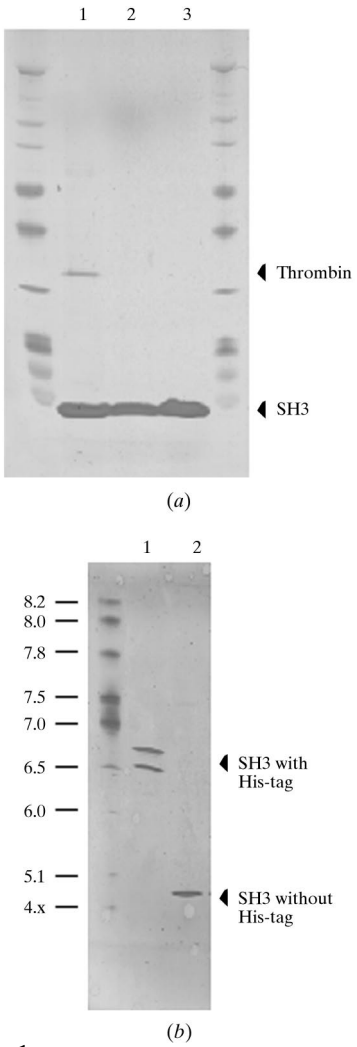

Figure 1
Analysis of recombinant src SH3 by electrophoresis. (*a*) Silver-stained SDS–PAGE of wild-type SH3 treated with thrombin (lane 1), SH3 after removal of thrombin and His-tagged fragments (lane 2) and SH3 from affinity chromatography and before thrombin treatment (lane 3). (*b*) Silver-stained IEF electrophoresis analyses of recombinant SH3 fused with and without His tag are shown in lanes 1 and 2, respectively.

2.2. Characterization of protein

Proteins were analyzed by either tricine SDS–PAGE (Schagger & von Jagow, 1987) or PhastGel gradient 8–25 on a PHAST system (Amersham Pharmacia Biotech). The cleavage of the His tag as well as the isoelectric points of the proteins were analyzed by isoelectric focusing (IEF) using Pharmacia Ampholite PAGplate gel (pH 3–9) with prestained isoelectric markers (BioRad) as the standards. All gels were silver stained. Mass measurements were performed by both electrospray ionization (ESI-MS) mass spectrometry (Finnigan LCQ quadrupole ion-trap Finnigan-MAT, San Jose, CA, USA) and matrix-assisted laser desorption/ionization (MALDI–TOF) mass spectrometry (Bioperseptive, Farmington, MA, USA). The matrix used for MALDI–TOF was recrystallized α-cyano-4-hydroxycinnamic acid (Sigma). The sample was added to a saturated solution of matrix-prepared 2:1 aqueous 0.1% trifluoroacetic acid:acetonitrile. 1 ml of the solution was placed on the probe of the mass spectrometer and allowed to dry before analysis. The polydispersity of the proteins was analyzed by dynamic light scattering (DLS) using DynaPro MS (Protein Solutions Ltd) equipped with temperature controller and fixed angle. All samples were filtered through 0.1 m*M* membrane before analysis. Lysozyme (20 mg ml^{-1}) in 40 m*M* sodium acetate pH 5.5, 0.5%(*v*/*v*) NaCl and BSA in PBS (2 mg ml^{-1}; Pierce) were used as controls.

2.3. Crystallization, data collection and reduction

All crystallizations were conducted by vapor diffusion in a hanging drop using Crystal Systems Q plates (Hampton Research) at 298 K with protein concentrations ranging from 35 to 60 mg ml^{-1}. Drops consisted of equal volumes of protein and reservoir solutions (well volume = 0.75 ml). Both the wild type and the W43I mutant of SH3 crystallized in 0.1 *M* MES at pH 6, 5–10% PEG 400 and 50–55% saturated $(NH_4)_2SO_4$. Crystals were transferred to cryoprotectant buffer in four sequential steps at the final concentration of 30%(*w*/*v*) sucrose. For X-ray analysis, crystals were mounted in loops on magnetic pins, flash-frozen in liquid nitrogen and mounted on a goniostat under a cryostream (MSC). In-house diffraction data were collected with a Rigaku R-AXIS IV image-plate detector and a Rigaku RU-200B rotating-anode generator with a copper anode (Cu $K\alpha$) operating at 50 kV and 100 mA. Synchrotron data were collected on beamline 5.0.2 at the Advanced Light Source (Lawrence Berkeley National Laboratory, Berkeley, CA, USA). The acquired data were reduced and scaled using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

3. Results and discussion

3.1. Purification and characterization

Proteins isolated from the soluble fraction by His-tag affinity followed by size-exclusion chromatography resulted in homogeneous protein when analyzed by SDS–PAGE (Fig. 1*a*). However, analysis by isoelectric focusing electrophoresis revealed two distinct protein bands at pI 6.75 and 6.5; the band with pI 6.5 is in excellent agreement with the calculated pI of 6.45 (Fig. 1*b*). Treatment of protein with thrombin, however, produced a single protein band on IEF near the expected pI of 4.96 for the His-tag-cleaved SH3. These observations suggest that the band with pI of 6.5 is the unmodified His-tagged SH3, while the band with pI of 6.75 represents SH3 with post-translational modification at His-tagged fragment.

Analysis by MALDI–TOF MS showed a singly charged ion corresponding to a mass of 8678.5 Da, the mass expected for src SH3 with the f-Met residue removed by endogenous methionine aminopeptidase (data not shown). In addition, a singly charged ion corresponding to an additional 178 Da in mass, 8857 Da, was observed (data not shown). An ESI-MS spectrum also produced a distinct pattern of ions corresponding to a mass of 8856 $\pm$ 1 Da, 178 Da larger than the expected mass of SH3, indicating that the observed molecular ion with +178 Da by MALDI–TOF is not an experimental artifact (Fig. 2*a*). Analysis by infusion ESI-MS of wild-type SH3 treated with thrombin resulted in ions corresponding to the expected mass for the His-tag fragment (1769 Da) and the His-tag-cleaved SH3 domain (6928 Da; Fig. 2*b*). In addition, ions corresponding to 1947 $\pm$ 1 Da, an extra 178 Da mass for the His-tag fragment, were also observed, indicating a post-translational modification at the N-terminus His-tag fragment consistent with observations made by IEF. Recently, Geoghegan *et al.* (1999) also observed 178 Da extra mass in His-tagged fusion proteins and hypothesized that spontaneous α-*N*-6 gluconoylation at the amino group of His-tagged protein is responsible for the observed extra mass.

Table 1
Summary of X-ray diffraction statistics and crystallographic parameters for wild-type and W43I SH3.

Values in parentheses represent the highest resolution shell.

	Wild type		W43I	
	Home source†	ALS‡	Home source†	ALS‡
Resolution (Å)	2.2	1.8	2.1	1.9
Unique reflections	5070	8284	5388	7197
Completeness (%)	96.7 (95.6)	98.8 (98.4)	97.4 (96.9)	99.2 (97.8)
$I/\sigma(I)$	22 (9.7)	23.8 (6.4)	32.4 (3.9)	19.0 (1.8)
R_{merge} (%)	7.7 (21.3)	4.3 (25.4)	5.2 (37.1)	7.9 (49.7)
Space group	$P6_522$		$P6_522$	
Unit-cell parameters (Å,°)	$a = b = 46.7$, $c = 127.7$, $\alpha = \beta = 90$, $\gamma = 120$		$a = b = 46.9$, $c = 128.5$, $\alpha = \beta = 90$, $\gamma = 120$	
Asymmetric unit	Monomer		Monomer	

† Diffraction data collected with a Rigaku R-AXIS image-plate detector. ‡ Diffraction data collected at the Advanced Light Source, Berkeley, USA.

They found that gluconoylation occurs on the N-terminal His-tag but not on the C-terminal His-tag and that the leading sequence GSS before the N-terminal His tag makes it more susceptible to gluconoylation. Our pET15b plasmid for SH3 expression contains a leading sequence of MGSSHHHHHH at the N-terminus (Fig. 2*b*). The removal of the f-Met residue by endogenous methionine aminopeptidase produced the more susceptible form of N-terminal His tag, GSSHHHHHH, in the *E. coli* expressed SH3 domain. The presence of an additional protein band very closely spaced near the expected pI and the additional 178 Da mass was also observed for all other mutants and confirmed that the minor heterogeneity is from the modification at the His-tagged fragments by mass spectrometry.

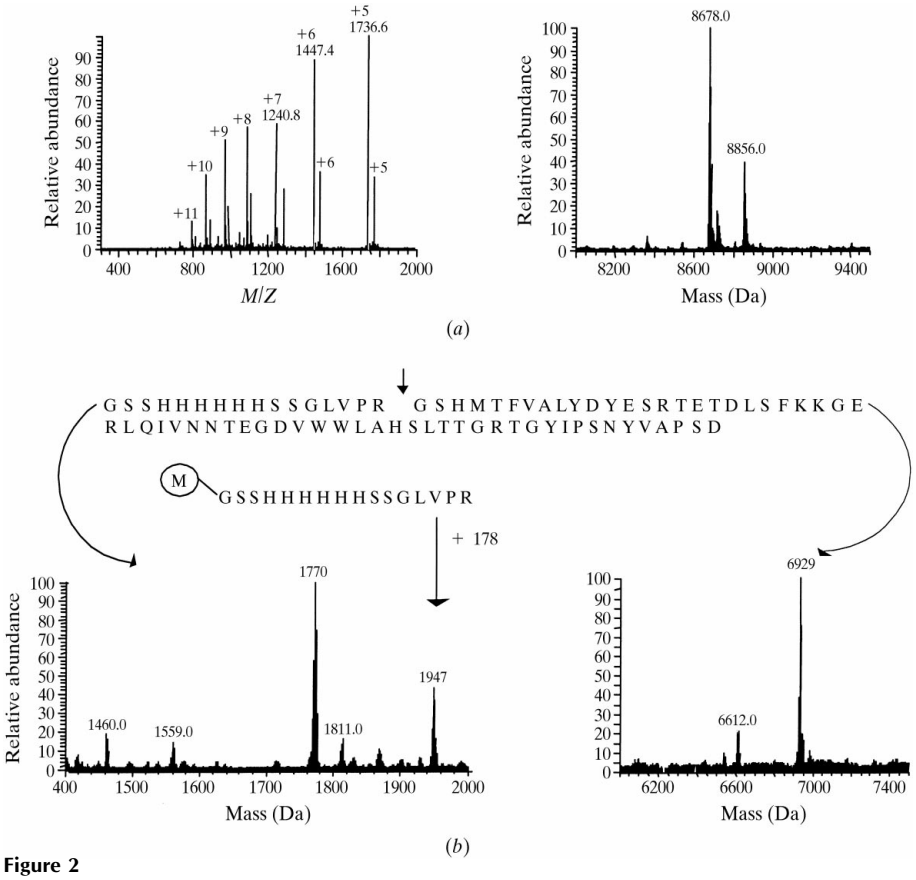

Figure 2
The extra 178 Da mass in His-tagged SH3 is from the modification of the N-terminal His tag. (*a*) ESI-MS spectrum showing molecular ions corresponding to the expected mass of His-tagged wild-type SH3 (8678.5 Da) and 8856 Da, 178 Da larger than the expected mass. (*b*) Infusion ESI-MS of wild-type SH3 after cleavage of the His tag with thrombin. The cleavage site by thrombin is shown by an arrow above the amino-acid sequence of recombinant wild-type src SH3.

3.2. Crystallization

Analysis by DLS indicated that proteins both with and without His tag are monodisperse (data not shown). However, the His-tagged proteins of both wild-type and mutant SH3 domains failed to crystallize after repeated attempts. We have also tried to crystallize the His-tagged SH3 domains in the presence of the divalent metal ion Zn^{2+}, which has been shown to be able to coordinate with His tags and immobilize them in crystals (O'Neill *et al.*, 2001). At the end, the removal of N-terminal His tag was essential for crystallization. The failure of crystallization of His-tagged SH3 domains may be attributed to the heterogeneities introduced by the post-translational modification at the N-terminal His tag rather than the flexibility of the N-terminal His tag. Similarly, interference of the His tag in the crystallization of His-tagged fusion proteins was also observed for ornithine decarboxylase (Grishin *et al.*, 1996) and HIV-1 integrase (Hickman & Davies, 1997). In contrast, some proteins retaining a His tag at either of the termini have been successfully crystallized (Mol *et al.*, 1996; Montoya *et al.*, 1997; Hakansson *et al.*, 2000; Hofmann *et al.*, 2000). Thus, design of recombinant proteins fused to affinity tags for rapid and selective purification on affinity columns may introduce heterogeneities arising from the post-translational modification of the His tag and may subsequently interfere with protein crystallization. Therefore, the heterogeneity introduced by post-translation modification of the N-terminal His tag should be examined as a possible cause when His-tagged recombinant proteins fail to crystallize. The post-translational modification can be identified by IEF electrophoresis and mass spectrometry. The deletion of the His tag could effectively eliminate the heterogeneity created by post-translational modification.

For the crystallization of both wild-type and mutant SH3 domain, the His tag introduced at the N-terminus was removed by thrombin cleavage. The optimal crystal growth occurred at 60.5 mg ml^{-1} protein in 0.1 *M* MES pH 6, 5% PEG 400, 50–55% saturated $(NH_4)_2SO_4$ for the wild type and at 48 mg ml^{-1} protein in 0.1 *M* MES pH 6, 10% PEG 400, 40–45% saturated $(NH_4)_2SO_4$ for the W43I SH3 domain. The crystals for both wild-type and W43I SH3 were hexagonal (Fig. 3). Interestingly, the crystal growth rate of W43I, which is less

stable and has higher folding and unfolding rates than the wild type, was much faster than the wild type (3–4 d compared with four weeks).

3.3. Data collection and analysis

Data sets were collected to 1.8 and 1.95 Å resolution for the wild type and W43I mutant, respectively; both proteins have similar crystallographic parameters. The X-ray diffraction statistics and crystallographic parameters are summarized in Table 1. The wild-type unit-cell volume of 24 1940 Å, which can accommodate 12 molecules of molecular weight 6928 Da with a reasonable value for V_M (Matthews, 1968) of 2.91 Å^3 Da^{-1}, indicates that there are 1–2 molecules of SH3 in the crystallographic asymmetric unit. Similarly, the V_M for W43I SH3 obtained from the unit-cell volume of 245 900 Å with a molecular weight of 6855 Da was 2.99 Å^3 Da^{-1}, indicating 1–2 molecules of W43I SH3 in the asymmetric unit. Preliminary structure determination by molecular replacement, however, indicates that both the wild type and the W43I mutant contain one molecule in the asymmetric unit.

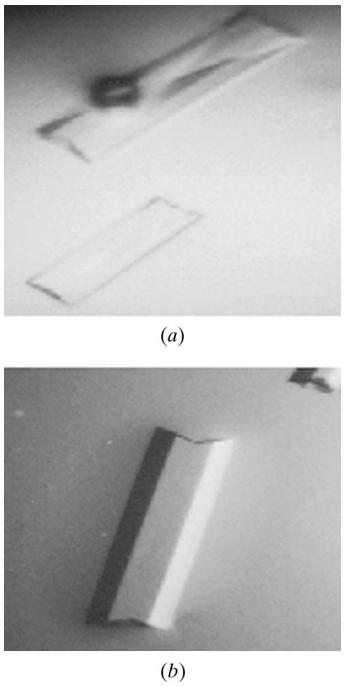

Figure 3
Crystals of (*a*) wild-type and (*b*) W43I src SH3 domain.

We gratefully acknowledge Drs Barry Stoddard, Ping-wei Li and Roland Strong for aid in data collection and fruitful discussions. This work was supported by Fred Hutchinson Cancer Research Foundation.

References

Fersht, A. R. (1988). *Biochemistry*, **27**, 1577–1580.
Geoghegan, K. F., Dixon, H. B., Rosner, P. J., Hoth, L. R., Lanzetti, A. J., Borzilleri, K. A., Marr, E. S., Pezzullo, L. H., Martin, L. B., LeMotte, P. K., McColl, A. S., Kamath, A. V. & Stroh, J. G. (1999). *Anal. Biochem.* **267**, 169–184.
Grantcharova, V. P., Riddle, D. S., Santiago, J. V. & Baker, D. (1998). *Nature Struct. Biol.* **5**, 714–720.
Grishin, N. W., Osterman, A. L., Goldsmith, E. J. & Phillips, M. A. (1996). *Proteins*, **24**, 272–273.
Hakansson, K., Broder, D., Wang, A. H. & Miller, C. G. (2000). *Acta Cryst.* D**56**, 924–926.
Hickman, A. B. & Davies, D. R. (1997). *Current Protocols in Protein Science*, Vol. 2, pp. 17.3.1–17.3.15, edited by J. Coligan, B. Dunn, H. Ploegh, D. Speicher & P. Wingfield. New York: John Wiley & Sons.
Hofmann, A., Proust, J., Dorowski, A., Schantz, R. & Huber, R. (2000). *J. Biol. Chem.* **275**, 8072–8082.
Jackson, S. E., elMasry, N. & Fersht, A. R. (1993). *Biochemistry*, **32**(42), 11270–11278.
Leatherbarrow, R. J. & Fersht, A. R. (1986). *Protein Eng.* **1**, 7–16.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mol, C. D., Harris, J. M., McIntosh, E. M. & Tainer, J. A. (1996). *Structure*, **4**, 1077–1092.
Montoya, G., Svensson, C., Luirink, J. & Sinning, I. (1997). *Proteins*, **28**, 285–288.
O'Neill, J. W., Kim, D. E., Baker, D. & Zhang, K. Y. J. (2001). *Acta Cryst.* D**57**, 480–487.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I. & Baker, D. (1999). *Nature Struct. Biol.* **6**, 1016–1024.
Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., Doshi, N., Grantcharova, V. P., Yi, Q. & Baker, D. (1997). *Nature Struct. Biol.* **4**, 805–809.
Schagger, H. & von Jagow, G. (1987). *Anal. Biochem.* **166**, 368–379.
Serrano, L., Horovitz, A., Avron, B., Bycroft, M. & Fersht, A. R. (1990). *Biochemistry*, **29**(40), 9343–9352.
Tsai, J., Levitt, M. & Baker, D. (1999). *J. Mol. Biol.* **291**(1), 215–225.