Stabilization of a binary protein complex by intein-mediated cyclization

CY M. JEFFRIES,¹ STEPHEN C. GRAHAM,² PHILIPPA H. STOKES,¹
CHARLES A. COLLYER,¹ J. MITCHELL GUSS,¹ AND JACQUELINE M. MATTHEWS¹

¹School of Molecular and Microbial Biosciences, University of Sydney, New South Wales 2006, Australia

²Division of Structural Biology, Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, United Kingdom

(RECEIVED May 30, 2006; FINAL REVISION July 26, 2006; ACCEPTED August 2, 2006)

Abstract

The study of protein–protein interactions can be hampered by the instability of one or more of the protein complex components. In this study, we showed that intein-mediated cyclization can be used to engineer an artificial intramolecular cyclic protein complex between two interacting proteins: the largely unstable LIM-only protein 4 (LMO4) and an unstructured domain of LIM domain binding protein 1 (ldb1). The X-ray structure of the cyclic complex is identical to noncyclized versions of the complex. Chemical and thermal denaturation assays using intrinsic tryptophan fluorescence and dynamic light scattering were used to compare the relative stabilities of the cyclized complex, the intermolecular (or free) complex, and two linear versions of the intramolecular complex (in which the interacting domains of LMO4 and ldb1 were fused, via a flexible linker, in either orientation). In terms of resistance to denaturation, the cyclic complex is the most stable variant and the intermolecular complex is the least stable; however, the two linear intramolecular variants show significant differences in stability. These differences appear to be related to the relative contact order (the average distance in sequence between residues that make contacts within a structure) of key binding residues at the interface of the two proteins. Thus, the restriction of the more stable component of a complex may enhance stability to a greater extent than restraining less stable components.

Keywords: intein; circular protein; protein cyclization; LIM domain; LMO4; fusion protein; protein stability

In this postgenomic era, the characterization of protein–protein interactions is becoming increasingly important. Yet many biologically and medically important proteins (or their domains) are unfolded or unstable in the absence of a binding partner. This problem can often be overcome by the coexpression and/or copurification of pairs of proteins. Alternatively, some complexes can be stabilized by engineering a single polypeptide chain chimera in which a protein–protein complex can form in an intramolecular

rather than an intermolecular fashion. This approach has been remarkably successful for preparing stable, soluble binary complexes that contain LIM domains from LIM-only (LMO) and LIM-homeodomain proteins (Deane et al. 2001, 2002, 2003a,b, 2004; Lee et al. 2005). LIM domains are a class of zinc finger domains that ligate two zinc ions, mediate protein–protein interactions, and are responsible for coordinating specific protein–protein interactions that regulate cell fate (Bach 2000; Visvader et al. 2001; Kadmas and Beckerle 2004). Whereas recombinant forms of LIM domains from some proteins are sufficiently soluble and stable to permit structural analyses, those from LMO and LIM-homeodomain proteins have a strong tendency to aggregate. This tendency to aggregate is not fully alleviated by the presence of binding partners such as the LIM-interaction domain (LID) of

Reprint requests to: Jacqueline Matthews, School of Molecular and Microbial Biosciences, University of Sydney, New South Wales, Australia 2006; e-mail: j.matthews@mmmb.usyd.edu.au; fax: 61-2-93564726.

Article published online ahead of print. Article and publication date are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.062377006>.

ldb1, despite having affinity constants of up to $\sim 10^8/\text{M}$ (Deane et al. 2003a, 2004).

We have previously solved the crystal structure of FLINC4, a fusion of ldb1-LID and the N- and C-terminal LIM domains of LMO4 (Deane et al. 2003b, 2004). The structure revealed that the two halves of the LMO4:ldb1-LID complex are arranged in a head-to-tail fashion with the N terminus of each protein lying in close proximity (~ 10 Å) to the C terminus of its protein partner. A flexible peptide linker holds the C terminus of LMO4 and the N terminus of ldb1-LID together enforcing a chelate-like effect that stabilizes the LMO4:ldb1-LID interaction (Deane et al. 2001, 2004) without introducing strain or steric hindrance. Because the N and C termini of the chimera lie close together in space, the LMO4:ldb1-LID complex is in an ideal configuration to be stabilized by a linker at the opposite end of the complex (i.e., between the C terminus of ldb1-LID and the N terminus of LMO4 rather than the C terminus of LMO4 and the N terminus of Ldb1-LID). Moreover, it appeared to be an excellent candidate for creating a cyclic protein complex through the incorporation of two linkers.

The artificial cyclization of recombinant polypeptides can be facilitated by the use of inteins (Camarero and Muir 1999; Camarero et al. 2001a; Xu and Evans 2001). These are self-associating discrete protein domains found in

a number of pro-proteins that catalyze a *cis*-splicing event internal to a polypeptide chain (Paulus 2000; Perler 2002, 2005). Splicing produces an excised intein domain with the simultaneous ligation of two flanking regions (the exteins) to form a mature protein. Inteins themselves can be split into two functional peptides, Int_N and Int_C , which comprise the N- and C-terminal sequences of the domain (Mills et al. 1998; Williams et al. 2002). Cloning a target gene between a reverse-order split intein sequence with the subsequent production of a fusion pro-protein in the order Int_C -Target- Int_N can result in the generation of a cyclized protein in vivo (Fig. 1; Xu and Evans 2001; Williams et al. 2002).

We used the synthetic *Synechocystis* sp. PCC6803 DnaB split mini-intein developed by Dixon and colleagues (Williams et al. 2002) to generate cyclized FLINC4. To assess differences in the structure and stability of different forms of the complex, we also produced a number of non-cyclized LMO4:ldb1-LID intramolecular complexes, FLINC4 and flip-FLINC4 and Xa-FLINC4 (Fig. 2B,C,E). Xa-FLINC4 contains a Factor Xa protease site in the linker and was treated with Factor Xa protease to generate an intermolecular complex, cut-Xa-FLINC4 (Fig. 2F), thus completing the LMO4:ldb1-LID series of protein complexes. Whereas the X-ray structure of the cyclic complex is identical to its non-cyclized intra- and intermolecular counterparts, the cyclic complex is more stable than the other variants.

Figure 1. Proposed intein-mediated in vivo protein cyclization mechanism. A target gene is cloned between the C- and N-terminal sequences of *Synechocystis* sp. PCC6803 DnaB split mini-intein to produce an Int_C -Target- Int_N protein fusion in vivo, in this instance a LMO4 (blue) ldb1-LID (yellow) chimera (A). The Int_C and Int_N domains spontaneously associate (B) and catalyze an N-to-S acyl transfer at the end of Int_N to generate a thioester intermediate (C). *Trans*-esterification of the thioester (via nucleophilic attack of an Int_C serine γO) cleaves Int_N and produces an Int_C -Target protein lariat intermediate (D). Asparagine side chain cyclization liberates the circular protein as a lactone (E), which then undergoes a spontaneous O-to-N acyl shift to form a stable lactam (F).

Figure 2. LMO4/ldb1-LID protein constructs. (A) Schematic diagram indicating the domains from LIM-only protein 4 and LIM-domain binding protein 1 used to generate intramolecular protein chimeras between the two proteins. (B) FLINC4, a fusion of LMO4 residues 16–152 and the LIM-interaction domain (LID, residues 300–339) of ldb1 (ldb1-LID) connected via a ldb1-LID N-terminal/LMO4 C-terminal Gly/Ser (N/C) linker. (C) Flip-FLINC4, a fusion of ldb1-LID and LMO4 connected via a ldb1-LID C-terminal/LMO4 N-terminal Gly/Ser rich (C/N) linker. (D) cz-FLINC4, a cyclized version of FLINC4 that contains the intein-derived C/N linker. (E) Uncut Xa-FLINC4, FLINC4 with a Factor Xa protease site in the linker between LMO4 and ldb1-LID. (F) Cut-Xa-FLINC4, an intermolecular or free complex between LMO4 and ldb1-LID produced by Factor Xa treatment of Xa-FLINC4.

Results and Discussion

The unusual purification of cz-FLINC4

The first hint of an increase in the stability of cz-FLINC4 over its noncyclized counterparts came during protein expression and purification. To obtain significant quantities of soluble protein, the noncyclized FLINC4 variants could only be expressed at lower temperatures (20°C–25°C) as affinity-tagged glutathione-*S*-transferase (GST)-fusion proteins, whereas cz-FLINC4 was optimally produced in an untagged state at 37°C. It was not necessary to carry out any steps to specifically cyclize cz-FLINC4; cyclization occurred spontaneously *in vivo* and no other intermediates were observed by SDS-PAGE (Fig. 3A). Initial attempts at purifying cz-FLINC4 from whole-cell lysates by ion-exchange chromatography were unsuccessful, presumably because of unidentified interactions between the cyclized protein and components of the lysate. Instead, a series of selective precipitation and resuspension steps, similar to those used to purify proteins from thermophilic organisms, were used to partially purify cz-FLINC4. This protocol included heating whole-cell lysates at 55°C in 30% (vol/vol) ethanol/buffer (wherein many contaminants were precipitated, but cz-FLINC4 remained in solution) followed by cold-induced precipitation

at –20°C. The cold-induced protein aggregates were easily dissolved in low ionic strength buffer. Thereafter, ion-exchange chromatography was an effective separation procedure, yielding >90% pure protein as judged by SDS-PAGE, with a final yield of ~1 mg/mL bacterial cell culture.

Trypsin proteolysis

Partial proteolysis by trypsin of purified cz-FLINC4 with subsequent SDS-PAGE analysis showed that the protein chimera was indeed circular. During SDS-PAGE circular proteins display an increased electrophoretic mobility than expected for their molecular mass because of the more compact nature of their denatured states (Iwai and Pluckthun 1999; Scott et al. 1999; Iwai et al. 2001). This was observed for cz-FLINC4 (Fig. 3B, lanes 2–3). After exposure to low levels of trypsin (that preferentially cuts cz-FLINC4 in the intein produced linker [Fig. 2D]), however, cz-FLINC4 shows a decrease in electrophoretic mobility consistent with conversion to a noncyclized form (Fig. 3B, lanes 3–8) (Iwai and Pluckthun 1999; Iwai et al. 2001; Williams et al. 2002).

Protein complex stability

It is generally expected that cyclized proteins (both naturally occurring and engineered) should be more stable than their noncyclized counterparts (Iwai and Pluckthun 1999; Scott et al. 1999; Camarero et al. 2001b; Iwai et al. 2001; Mulvenna et al. 2006), provided that no significant structural perturbations or strain are introduced as a consequence of the cyclization process (Williams et al. 2005). To gauge differences in the comparative stability of each complex, cz-FLINC4, FLINC4, flip-FLINC4, Xa-FLINC4, and cut-Xa-FLINC4 were each subjected to chemical denaturation by guanidine

Figure 3. Expression and trypsin proteolysis of cz-FLINC4. (A) Expression of cz-FLINC4 in BL21 (DE3) pMICO. (Lane 2) Cell lysate showing the over expression of cz-FLINC2 (arrow). The prominent bands above and below cz-FLINC4 (*) are proteins derived from the pMICO plasmid. (B) Trypsin proteolysis. (Lane 2) FLINC4; (lane 3) cz-FLINC4; (lanes 4–8) cz-FLINC4 after treatment with trypsin for indicated times. (Lane 1) Molecular weight marker (MW); protein sizes (kDa) are shown to the left of each image.

hydrochloride (GdnHCl) monitored by tryptophan fluorescence. Tryptophan fluorescence experiments report directly on the environments of the two Trp residues in LMO4, Trp19 and Trp43, both of which lie in the structured regions of the N-terminal LIM domain of LMO4 (LIM1). All of the complexes have essentially identical tryptophan emission spectra and show a red shift in emission wavelength maxima (λ_{max}) from ~ 340 nm to 355 nm upon complete unfolding (data not shown). Denaturation curves over the range 0–6 M GdnHCl (Fig. 4A) were not reversible and could not be fitted by equations that describe two-state equilibrium denaturation. However, they indicate a striking difference in the stabilities of the different forms of the LMO4:ldb1-LID complex. The free complex (cut-Xa-FLINC4) appears least stable, the circular form (cz-FLINC4) most stable, and the linear FLINC4 chimera of intermediate stability. Curiously, the second noncyclized chimera, flip-FLINC4, appears just as resistant to GdnHCl denaturation as cz-FLINC4. In both cz-FLINC4 and flip-FLINC4 there is a linker between the C terminus of ldb1-LID and the N terminus of LMO4-LIM1. Thus, it was possible that this denaturation assay reported specifically on the stabilization of LMO4-LIM1 by fusion through this linker.

To provide an alternative measure of relative stability FLINC4, flip-FLINC4, and cz-FLINC4 were also subjected to denaturation by heat and monitored for aggregation by dynamic light scattering (Fig. 4B). In these experiments we can readily detect the formation of large soluble aggregates that arise through the accumulation of unfolding intermediates. All proteins tested were monomeric at low temperatures, although experimental molecular weights estimated by this method were slightly high

Table 1. X-ray data collection and refinement statistics of cz-FLINC4 (molecular replacement)

Space group	p3
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	61.33, 61.33, 93.02
α , β , γ (°)	90, 90, 120
Resolution (Å)	10 (1.65) ^a
<i>R</i> _{sym} or <i>R</i> _{merge}	0.034 (0.33)
<i>I</i> / σ <i>I</i>	36.3 (2.21)
Completeness (%)	99.5 (99.0)
Redundancy	3.9 (2.2)
Resolution (Å)	1.65
No. reflections	44,295
<i>R</i> _{work} / <i>R</i> _{free}	0.167/0.194
No. atoms	2711
Protein	2448
Ligand/ion	8
Water	244
RMSD bond lengths (Å)	0.014
RMSD bond angles (°)	1.398

Data set derived from one crystal.

^aHighest resolution shell is shown in parentheses. FLINC4 (PDB ID, 1RUT) was used as an initial molecular replacement model.

(cz-FLINC4, estimated 24 kDa, theoretical 21.2 kDa; FLINC4, 23 vs. 20.2 kDa; flip-FLINC4, 27.0 vs. 19.9 kDa). This overestimation of molecular mass likely arises from the deviation from an ideal sphere that is intrinsic to these molecules; the structured regions of FLINC4 are elongated with molecular dimensions of $80 \times 20 \times 20$ Å. Again, FLINC4 was less stable than cz-FLINC4 showing signs of aggregation at 30°C–32.5°C, $\sim 10^\circ\text{C}$ – 12°C lower than the cyclic complex (aggregation initiated at 40°C–45°C). In these experiments, flip-FLINC4 begins to aggregate at temperatures only marginally lower than cz-FLINC4 but shows a much higher tendency to aggregate than the circular protein at temperatures above the initiation of aggregation (Fig. 4B,C).

X-ray structure of cz-FLINC4

To determine whether the structure of the LMO4:ldb1-LID complex was affected by cyclization, we solved the X-ray structure of cz-FLINC4 to a resolution of 1.65 Å (Table 1). The X-ray structures of both cz-FLINC4 and FLINC4 are essentially identical (*C*_α RMSD 0.2 Å; Fig. 5). For both proteins, the anisotropic temperature factor refinement and analysis shows a noticeable elevation in temperature factors for those residues of LIM2 (LMO4_{80–146}) and the corresponding LIM2-binding region of ldb1 (residues 300–310; Fig. 6), which suggests increased conformational flexibility in this region.

Previous investigations using NMR spectroscopy have shown that Xa-FLINC4 is indistinguishable from the “free” intermolecular cut-Xa-FLINC4 complex (Deane et al. 2004). Similarly, an NMR spectrum of flip-FLINC4

Figure 4. cz-FLINC4 is more stable than FLINC4. (A) Chemical denaturation of cz-FLINC4 (filled squares), FLINC4 (circles), flip-FLINC4 (diamonds), and cut-Xa-FLINC4 (open squares) monitored by tryptophan fluorescence. Xa-FLINC4 (triangles) is also included to show that the presence of the Factor Xa protease site in the fusion linker has no significant effect on stability. (B) Thermal denaturation of cz-FLINC4, FLINC4, and flip-FLINC4 monitored by dynamic light scattering from 20°C–60°C (symbol key as in A). (Inset) Close-up view of the thermally induced molecular aggregation temperatures of cz-FLINC4 (40°C–45°C), FLINC4 (32.5°C–35°C), and flip-FLINC4 (40°C–42.5°C). Error bars show ± 1 standard deviation from the mean.

Figure 5. cz-FLINC4 and FLINC4 are essentially identical. Structural alignment of cz-FLINC4 chain X (blue) and FLINC4 (green): C_{α} RMSD, 0.2 Å. Minor variations include: the absence of side chain alternate conformations in cz-FLINC4 (LMO4: Asp40, Gln56, Gln92; ldb1: Met302, Glu317, Glu326); the presence of alternate amino acid side chain conformations in cz-FLINC4 (LMO4: Arg74); and alternate side chain rotamers in cz-FLINC4 (LMO4-Ile131, ldb1-Thr308). Zinc atoms (two per LIM domain) are shown as blue spheres.

is almost identical to that of FLINC4 (data not shown). Combined, these observations indicate that the structured regions of the intermolecular, intramolecular, and cyclic complexes are all homologous.

FLINC4 and flip-FLINC4

Why is flip-FLINC4 apparently more stable than FLINC4? In trying to answer this question, several properties of these proteins must be considered. First, the sequences of the unstructured linkers differ. However, FLINC4 and Xa-FLINC4 have identical denaturation profiles (Fig. 4A), suggesting that the precise composition of an unstructured linker appears to have little effect on stability. Second, there are the lengths of the linker sequences to consider. Including the designed linkers and regions from the domains that lack electron density (and are presumably unstructured), the unstructured linkers of flip-FLINC4 and FLINC4 are 24 and 16 residues in length, respectively (whereas the LMO4 N-terminal and C-terminal linkers of cz-FLINC4 are 22 and 15 residues, respectively). Although polymer theory suggests that increasing loop length should result in a decrease in stability, this is inconsistent with our data. It has been shown previously that the insertion of unstructured loops into proteins can have very little effect on protein stability (Ladurner and Fersht 1997). Finally, a localized thermodynamic effect encompassing the LMO4-LIM1:ldb1-LID portion of the interaction may dominate overall complex stability. Although both LIM1 and LIM2 domains are required for high-affinity binding of LMO4 to ldb1 ($K_A \sim 10^8/M$), LMO4-LIM1 is primarily responsible for binding ($K_A \sim 10^6/M$) (Deane et al. 2003a, 2004). Thus, restraining the more stable end of the complex (LIM1:ldb1-LID) may enhance complex stability to a greater extent than restraining the less stable end (LIM2:ldb1-LID).

It has been shown previously that circular permutants of a protein domain can have a range of stabilities (Lindberg et al. 2006). These circular permutants are

domain variants that have the same order of amino acids in the polypeptide sequence but with the N and C termini in different positions; the permutants were designed such that the N and C termini should lie in close proximity and that the structures be essentially identical. An extensive analysis of the folding properties of these permutants showed that their folding nuclei and rates of folding and unfolding were significantly different. This study also showed that increased folding rates and stabilities both correlated with decreased contact order. This is a topological parameter that is the average distance in sequence between residues that make contacts within a structure (Plaxco et al. 1998). If we simply consider the relative contact order of residues at the interface of the LMO4:ldb1-LID (i.e., only those residues in ldb1-LID that contact LMO4 and vice versa; intradomain contacts should be same in all cases), there is little difference between FLINC4 and flip-FLINC4 (0.53 and 0.52, respectively). However, the relative contact order for residues at the LMO4-LIM1:ldb1-LID interface is very different (0.57 vs. 0.33, respectively). The contact orders of rate-determining transition state folding ensembles of model

Figure 6. Individual nonhydrogen atom temperature factor comparisons between cz-FLINC4 (black line) and FLINC4 (gray line). There is an increase in temperature factors when contrasting LIM domains 1 and 2 of cz-FLINC4 or FLINC4. Average temperature factors are shown in parentheses. The temperature factors of ldb1-LID decrease markedly as the LID-protein traverses the length of LMO4.

protein domains are $\sim 30\%$ larger than, but correlate with, the native states for the same proteins (Paci et al. 2005). Thus, if the “hotspot” contacts at the LMO4:ldb1-LID interface (Deane et al. 2004) make contributions to transition state ensembles for complex formation that are analogous to folding nuclei, then the decreased contact order of those residues in flip-FLINC4 could be associated with increased on-rates of association and a consequent increased stability over its permutant, FLINC4.

In conclusion, we have shown that generating intramolecular protein complexes can have a marked improvement on the stability of those proteins compared with intermolecular complexes, although levels of stability can vary according to the design of the engineered complex. Furthermore, we have demonstrated that it is possible to engineer a cyclic polypeptide comprising two different proteins through intein-mediated cyclization complex, which in this system can form a cyclic complex that is more stable than noncyclic variants. This work has important implications for the generation of stable protein complexes in which one or both components of the complex is unstable, and where the N and C termini of each protein are close to the C and N termini of their respective partner protein.

Materials and methods

Protein constructs and purification

Flip-FLINC4 (ldb1 residues 300–339, a GGSGGHMGSGG linker, and LMO4 residues 18–150) was generated using overlap extension PCR as described previously for FLINC4 (Deane et al. 2003b). FLINC4, Xa-FLINC4, and flip-FLINC4 proteins were expressed and purified as described previously (Deane et al. 2001, 2003b). An aliquot of Xa-FLINC4 (1 mg in 20 mM Tris, ~ 200 mM NaCl, 4 mM β -mercaptoethanol at pH 8.0) was digested at 10°C overnight with 20 μg of Factor Xa (New England Biolabs) to generate cut-Xa-FLINC4. cz-FLINC4 was generated by ligating the FLINC4 gene construct (LMO4 residues 18–152, a GGSGGSGGSGG linker, and ldb1 residues 300–339) into a split mini-intein cassette (Williams et al. 2002), which was subcloned into the constitutive protein expression vector pCY76 (Yang et al. 2003). Protein was expressed by *Escherichia coli* BL21 (DE3) *recA* in Luria broth (LB) supplemented with ampicillin (60 $\mu\text{g}/\text{mL}$) and thymine (25 $\mu\text{g}/\text{mL}$) for 24 h at 37°C . Initial overexpression trials were carried out in BL21 (DE3) *recA* containing an extra chromosomal element, pMICO (argU, ileX, and glyT; Cinquin et al. 2001), in which case chloramphenicol (30 $\mu\text{g}/\text{mL}$) was added. The harvested cell pellet was dissolved in lysis buffer (20 mM Tris-HCl, 150 mM NaCl, 1 mM ZnSO_4 , 0.1% vol/vol β -mercaptoethanol at pH 8.0) to which ethanol was added to a final volume of 30%. The suspension was heated to 55°C for 5 min and centrifuged for 1 h at 4°C to remove major contaminants. The supernatant was incubated at -20°C for 12 h to precipitate cz-FLINC4, and the protein was further purified by anion-exchange chromatography (Deane et al. 2001, 2003b).

Trypsin proteolysis

cz-FLINC4 (2 mg/mL) was treated with trypsin (127 U; Sigma) on ice. Protein aliquots (80 μL) were removed at 15 sec, 30 sec,

1 min, 2 min, and 5 min and the proteolytic reaction stopped by the addition of 20 μL of $5\times$ SDS-PAGE loading buffer (containing 2% SDS) followed by heating to 90°C for 5 min. Samples were analyzed by 15% SDS-PAGE.

Denaturation experiments

Proteins (~ 10 μM) in 20 mM Tris, 150 mM NaCl, 2 mM EDTA (pH 8.0) and GdnHCl (0–6 M) were subjected to tryptophan fluorescence (excitation 295 nm) at 20°C on a Cary Eclipse Fluorescence Spectrophotometer (Varian Inc.). Emission spectra (325–370 nm) were collected at 0.5-nm intervals using excitation and emission slit widths of 10 nm. Reported data are the average of three scans recorded at 30 nm/min and processed using Scan Software v.1.1. Temperature dependent denaturation/aggregation states of proteins were measured using a Dynapro Dynamic Light Scattering temperature controlled microsamplers (Protein Solutions) with an 830-nm laser diode. Light scattering intensities were measured from 20°C – 60°C in steps of 5°C or 2.5°C . A total of 200 scattering intensity acquisitions were recorded for each temperature tested (10 acquisitions of 1 sec per measurement, 20 measurements per temperature). Protein samples (~ 6 mg/mL in 20 mM Tris, 200 mM NaCl, 4 mM BME at pH 8.0) were equilibrated for 3 min in a sealed 15- μL quartz cuvette at the respective temperature prior to recording. Data were processed using Dynamics Dynapro Control Software v.6.3.40.

Crystallization, X-ray diffraction, data processing, and refinement

Hanging drop vapor diffusion was used to grow crystals of cz-FLINC4. Protein solutions (2 μL , 8 mg/mL in 20 mM Tris, 200 mM NaCl at pH 8.0) were mixed with 2 μL of 1 M $(\text{NH}_4)_2\text{SO}_4$, 100 mM Tris (pH 8.5), 15% vol/vol glycerol and equilibrated at 23°C against 500 μL of reservoir solution. Diffraction quality crystals grew over a period of 2–3 d. Crystals were cryoprotected by brief soaking in reservoir solution supplemented with an additional 10% vol/vol glycerol before being flash cryo-cooled in a stream of cold (100 K) nitrogen gas (Cryosystems). X-rays ($\text{Cu K}\alpha$, 1.5418 Å) were generated using a Rigaku RU-200H rotating anode generator (Rigaku MSC) and were collimated and focused with Osmic mirror optics (Rigaku MSC). Diffraction data were recorded in three passes ($d = 110$ mm, 1.6 Å; $d = 165$ mm, 1.7 Å; $d = 315$ mm, 3.1 Å) at 100 K on a mar345image-plate detector (Marresearch) through a rotation of 0° – 50° (ϕ increment 0.25°). All data to 1.65 Å were integrated and scaled with DENZO and SCALEPACK (HKL Suite; Otwinowski and Minor 1997). Molecular replacement was performed against data to 3.0 Å in Phaser (Storoni et al. 2004) using the structure of noncyclized FLINC4 (PDB ID, 1RUT) stripped of metal ions, solvent molecules, and alternate conformers as a starting model. Because of poor electron density resolved for half of the initial molecular replacement model, residues encompassing the second LIM domain of LMO4 (residues 80–146) and ldb1 residues 300–310 were removed from the model and rebuilt de novo into unbiased $F_o - F_c$ maps. Molecular visualization and manual model building were performed in Coot (Emsley and Cowtan 2004). Refinement cycles were carried out with REFMAC5 (Murshudov et al. 1997) with hydrogen atoms modeled in “riding” positions. Final rounds of refinement included the modeling of anisotropic temperature factors of all nonhydrogen atoms. Automated water building was performed using ARP/wARP (Perrakis et al.

1999). Structure validation was checked using PROCHECK (Laskowski et al. 1993) and Ramachandran Plot statistics (Most favored, 99.7%; Allowed, 100%; Outliers, 0%) calculated with MOLPROBITY (Lovell et al. 2003). Tertiary structure alignment and C α RMSD calculations were calculated using Combinatorial Extension (Shindyalov and Bourne 1998). The coordinates of the crystal structure have been deposited in the Protein Data Bank (PDB ID, 2DFY).

Relative contact order

Relative contact order (CO) is the average sequence distance between all residue contacts between LMO4 and ldb1 in the FLINC4 structure⁶ (1RUT) normalized by the total sequence length as determined by:

$$CO = \frac{1}{L \times N} \sum \Delta S_{ij}$$

N is the total number of contacts, ΔS_{ij} is the sequence separation, in residues, between contacting residues i and j , and L is the total number of residues in the construct.

Acknowledgments

We thank Nicholas Dixon and Neil Williams for providing the split mini-intein expression cassette and Janet Deane for comments on the manuscript. C.M.J., S.C.G., and P.H.S. were supported by Australian Postgraduate Awards. J.M.M. is funded by the Viertel Foundation. This work was funded by grants from the Australian NHMRC and the University of Sydney Cancer Research Fund.

References

- Bach, I. 2000. The LIM domain: Regulation by association. *Mech. Dev.* **91**: 5–17.
- Camarero, J.A. and Muir, T.W. 1999. Biosynthesis of a head-to-tail cyclized protein with improved biological activity. *J. Am. Chem. Soc.* **121**: 5597–5598.
- Camarero, J.A., Fushman, D., Cowburn, D., and Muir, T.W. 2001a. Peptide chemical ligation inside living cells: *in vivo* generation of a circular protein domain. *Bioorg. Med. Chem.* **9**: 2479–2484.
- Camarero, J.A., Fushman, D., Sato, S., Giriat, L., Cowburn, D., Raleigh, D.P., and Muir, T.W. 2001b. Rescuing a destabilized protein fold through backbone cyclization. *J. Mol. Biol.* **308**: 1045–1062.
- Cinquin, O., Christopherson, R.I., and Menz, R.I. 2001. A hybrid plasmid for expression of toxic malarial proteins in *Escherichia coli*. *Mol. Biochem. Parasitol.* **117**: 245–247.
- Deane, J.E., Sum, E., Mackay, J.P., Lindeman, G.J., Visvader, J.E., and Matthews, J.M. 2001. Design, production and characterization of FLIN2 and FLIN4: The engineering of intramolecular ldb1:LMO complexes. *Protein Eng.* **14**: 493–499.
- Deane, J.E., Visvader, J.E., Mackay, J.P., and Matthew, J.M. 2002. ¹H, ¹⁵N and ¹³C assignments of FLIN4, an intramolecular LMO4:ldb1 complex. *J. Biomol. NMR* **23**: 165–166.
- Deane, J.E., Mackay, J.P., Kwan, A.H., Sum, E.Y., Visvader, J.E., and Matthews, J.M. 2003a. Structural basis for the recognition of ldb1 by the N-terminal LIM domains of LMO2 and LMO4. *EMBO J.* **22**: 2224–2233.
- Deane, J.E., Maher, M.J., Langley, D.B., Graham, S.C., Visvader, J.E., Guss, J.M., and Matthews, J.M. 2003b. Crystallization of FLINC4, an intramolecular LMO4-ldb1 complex. *Acta Crystallogr. D Biol. Crystallogr.* **59**: 1484–1486.
- Deane, J.E., Ryan, D.P., Sunde, M., Maher, M.J., Guss, J.M., Visvader, J.E., and Matthews, J.M. 2004. Tandem LIM domains provide synergistic binding in the LMO4:ldb1 complex. *EMBO J.* **23**: 3589–3598.
- Emsley, P. and Cowtan, K. 2004. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**: 2126–2132.
- Iwai, H. and Pluckthun, A. 1999. Circular β -lactamase: Stability enhancement by cyclizing the backbone. *FEBS Lett.* **459**: 166–172.
- Iwai, H., Lingel, A., and Pluckthun, A. 2001. Cyclic green fluorescent protein produced *in vivo* using an artificially split PI-Pf1 intein from *Pyrococcus furiosus*. *J. Biol. Chem.* **276**: 16548–16554.
- Kadmas, J.L. and Beckerle, M.C. 2004. The LIM domain: From the cytoskeleton to the nucleus. *Nat. Rev. Mol. Cell Biol.* **5**: 920–931.
- Ladurner, A.G. and Fersht, A.R. 1997. Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. *J. Mol. Biol.* **273**: 330–337.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**: 283–291.
- Lee, C., Nancarrow, A.L., Bach, I., Mackay, J.P., and Matthews, J.M. 2005. ¹H, ¹⁵N and ¹³C assignments of an intramolecular Lhx3:ldb1 complex. *J. Biomol. NMR* **33**: 198.
- Lindberg, M.O., Haglund, E., Hubner, I.A., Shakhnovich, E.I., and Oliveberg, M. 2006. Identification of the minimal protein-folding nucleus through loop-entropy perturbations. *Proc. Natl. Acad. Sci.* **103**: 4083–4088.
- Lovell, S.C., Davis, I.W., Arendall 3rd, W.B., de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. 2003. Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins* **50**: 437–450.
- 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.* **95**: 3543–3548.
- Mulvenna, J.P., Wang, C., and Craik, D.J. 2006. CyBase: A database of cyclic protein sequence and structure. *Nucleic Acids Res.* **34**: D192–D194.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**: 240–255.
- Otwinowski, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**: 307–326.
- Paci, E., Lindorff-Larsen, K., Dobson, C.M., Karplus, M., and Vendruscolo, M. 2005. Transition state contact orders correlate with protein folding rates. *J. Mol. Biol.* **352**: 495–500.
- Paulus, H. 2000. Protein splicing and related forms of protein autoprocesing. *Annu. Rev. Biochem.* **69**: 447–496.
- Perler, F.B. 2002. InBase: The intein database. *Nucleic Acids Res.* **30**: 383–384.
- Perler, F.B. 2005. Protein splicing mechanisms and applications. *IUBMB Life* **57**: 469–476.
- Perrakis, A., Morris, R., and Lamzin, V.S. 1999. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**: 458–463.
- Plaxco, K.W., Simons, K.T., and Baker, D. 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**: 985–994.
- Scott, C.P., Abel-Santos, E., Wall, M., Wahnon, D.C., and Benkovic, S.J. 1999. Production of cyclic peptides and proteins *in vivo*. *Proc. Natl. Acad. Sci.* **96**: 13638–13643.
- Shindyalov, I.N. and Bourne, P.E. 1998. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11**: 739–747.
- Storoni, L.C., McCoy, A.J., and Read, R.J. 2004. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D Biol. Crystallogr.* **60**: 432–438.
- Visvader, J.E., Venter, D., Hahm, K., Santamaria, M., Sum, E.Y., O'Reilly, L., White, D., Williams, R., Armes, J., and Lindeman, G.J. 2001. The LIM domain gene LMO4 inhibits differentiation of mammary epithelial cells *in vitro* and is overexpressed in breast cancer. *Proc. Natl. Acad. Sci.* **98**: 14452–14457.
- Williams, N.K., Prosser, P., Liepinsh, E., Line, I., Sharipo, A., Littler, D.R., Curmi, P.M., Otting, G., and Dixon, N.E. 2002. *In vivo* protein cyclization promoted by a circularly permuted *Synechocystis* sp. PCC6803 DnaB mini-intein. *J. Biol. Chem.* **277**: 7790–7798.
- Williams, N.K., Liepinsh, E., Watt, S.J., Prosser, P., Matthews, J.M., Attard, P., Beck, J.L., Dixon, N.E., and Otting, G. 2005. Stabilization of native protein fold by intein-mediated covalent cyclization. *J. Mol. Biol.* **346**: 1095–1108.
- Xu, M.Q. and Evans Jr., T.C. 2001. Intein-mediated ligation and cyclization of expressed proteins. *Methods* **24**: 257–277.
- Yang, H., Carr, P.D., McLoughlin, S.Y., Liu, J.W., Horne, I., Qiu, X., Jeffries, C.M., Russell, R.J., Oakeshott, J.G., and Ollis, D.L. 2003. Evolution of an organophosphate-degrading enzyme: A comparison of natural and directed evolution. *Protein Eng.* **16**: 135–145.