Crystal structure of the SH3 domain in human Fyn; comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin

Martin E.M.Noble¹, Andrea Musacchio,
Matti Saraste, Sara A.Courtneidge and
Rik K.Wierenga²

EMBL, Meyerhofstrasse 1, D-6900 Heidelberg, Germany

¹Present address: Laboratory of Molecular Biophysics, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

²Corresponding author

Communicated by S.A.Courtneidge

The Src-homology 3 (SH3) region is a protein domain consisting of ~60 residues. It occurs in a large number of eukaryotic proteins involved in signal transduction, cell polarization and membrane–cytoskeleton interactions. The function is unknown, but it is probably involved in specific protein–protein interactions. Here we report the crystal structure of the SH3 domain of Fyn (a Src family tyrosine kinase) at 1.9 Å resolution. The crystals have two SH3 molecules per asymmetric unit. These two Fyn SH3 domains are not related by a local twofold axis. The crystal structures of spectrin and Fyn SH3 domains as well as the solution structure of the Src SH3 domain show that these all have the same basic fold. A protein domain which has the same topology as SH3 is present in the prokaryotic regulatory enzyme BirA. The comparison between the crystal structures of Fyn and spectrin SH3 domains shows that a conserved surface patch, consisting mainly of aromatic residues, is flanked by two hairpin-like loops (residues 94–104 and 114–118 in Fyn). These loops are different in tyrosine kinase and spectrin SH3 domains. They could modulate the binding properties of the aromatic surface.

Key words: crystal structure/Fyn/protein–protein interactions/spectrin/Src

Introduction

The product of the *fyn* proto-oncogene is a member of the *src* family of cytoplasmic tyrosine kinases. This group of enzymes is involved in signal transduction pathways which begin at activated transmembrane receptors in the plasma membrane (see Koch *et al.*, 1991, and references therein). A remarkable feature of Fyn- and Src-like kinases is that beside the catalytic domain, they contain two small, mutually unrelated, non-catalytic domains. These are called SH2 and SH3 (Src-homology regions 2 and 3), and they have about 100 and 60 amino acids, respectively (see Koch *et al.*, 1991; Musacchio *et al.*, 1992a). Homologous domains are also found in a number of other proteins involved in signal transduction. These include the p21^{ras} GTPase activating protein (GAP) (Trahey *et al.*, 1988), phospholipase C γ (Stahl *et al.*, 1988), one subunit of phosphatidylinositol 3'

kinase (Otsu *et al.*, 1991) and Sem5 (Clark *et al.*, 1992). The latter belongs to a family of signalling proteins composed entirely of SH2 and SH3 domains. It has been suggested that the SH2 and SH3 domains function as molecular adaptors that mediate molecular recognition events between proteins involved in signal transduction pathways (Koch *et al.*, 1991).

The structure and phosphotyrosine-binding function of the SH2 domain have recently been established (Mayer *et al.*, 1991; Booker *et al.*, 1992; Overduin *et al.*, 1992; Waksman *et al.*, 1992, 1993; Eck *et al.*, 1993). In contrast, the functional role of SH3 remains largely unsolved. SH3 is not exclusively found in signalling proteins. Both in single-cell and in higher eukaryotes it is present in apparently unrelated proteins, including erythroid and non-erythroid α -spectrins (Sahr *et al.*, 1990; Wasenius *et al.*, 1989), a putative transcriptional regulator, Hs1 (Kitamura *et al.*, 1989), the p47 and p67 activators of the human neutrophil respiratory burst oxidase (Volpp *et al.*, 1989; Leto *et al.*, 1990) and proteins involved in yeast morphogenesis and fusion (McCaffrey *et al.*, 1987; Trueheart *et al.*, 1987; Drubin *et al.*, 1988, 1990; Chant *et al.*, 1991; Chenevert *et al.*, 1992) and regulation of *ras* activity (Broek *et al.*, 1987; Hughes *et al.*, 1990). The above mentioned proteins do not contain SH2 domains—SH2 and SH3 are not obligate partners. The presence of the SH3 domain in diverse proteins suggests that it carries out an important function in different cellular processes.

Mutations in, and deletions of, the SH3 domains in protein tyrosine kinases such as Src and Abl result in the activation of the transforming potential of these proteins. This implies that SH3 has an important role in the down-regulation of kinase activity (Jackson and Baltimore, 1989; Kato *et al.*, 1986; Seidel-Dugan *et al.*, 1992) either via interactions with other regulatory proteins or via an intramolecular mechanism.

Screening of an expression library for proteins that bind to the SH3 domain of the Abl tyrosine kinase have led to the isolation of two clones encoding proteins called 3BP-1 and 3BP-2 (Cicchetti *et al.*, 1992; Ren *et al.*, 1993). The sequence of 3BP-1 is related to a family of proteins that are known to have GAP activity. The SH3-binding site in 3BP-1 and 3BP-2 have been mapped to short proline-rich segments, but the sequences of these two proteins are otherwise unrelated (Cicchetti *et al.*, 1992; Ren *et al.*, 1993).

We have previously presented the crystal structure of the SH3 domain of α -spectrin at 1.8 Å resolution (Musacchio *et al.*, 1992b). Here we report the crystal structure of the SH3 domain of the Fyn tyrosine kinase, solved at a resolution of 1.9 Å. Another group has recently solved the solution structure of the Src SH3 domain using NMR techniques (Yu *et al.*, 1992). We compare these three structures, discuss their mutual similarity and in particular the structural details of the proposed ligand-binding surface.

Results and discussion

Structure of Fyn SH3

The Fyn SH3 domain, amino acid residues 81–144 in the sequence reported by Kawakami *et al.* (1986) and Semba *et al.* (1986), was expressed with a pET vector in the *Escherichia coli* cytoplasm. The construction introduced only a single extra residue (N-terminal methionine) into the 64 residue protein. The protein was purified and crystallized as described in Materials and methods.

The structure of the SH3 domain of human Fyn has been solved independently using the SIRAS (single isomorphous replacement, including anomalous scattering) method. The two heavy atom sites are clustered together at a crystal contact region. One site is close to OH of Tyr132 (2.5 Å) and OE2 of Glu116 (2.6 Å), and the other site is near OH of Tyr132 (3.5 Å) and OE2 of Glu121 (3.7 Å). Subsequently the structure has been refined at 1.9 Å. The final R-factor is 18.0% for data between 5 and 1.9 Å. The structure agrees well with standard stereochemical parameters: the r.m.s. difference from standard bond lengths is 0.02 Å, and from standard bond angles it is 2.9° (Table I). The PROCHECK program suite (Morris *et al.*, 1992) confirms that the structure also agrees well with various non-refined stereochemical parameters, scoring better than or within 0.2 σ on all of the analysed parameters when compared with other proteins of the same nominal resolution.

The crystallographic asymmetric unit consists of two molecules of Fyn SH3. These two molecules are not related by a local two-fold axis. We will refer to these two molecules as Fyn SH3-1 and SH3-2. For both molecules electron density exists for 59 residues, which correspond to residues 84–142 in the sequence of human Fyn.

The crystal structure of spectrin SH3 (Musacchio *et al.*, 1992b), and the more recently published NMR structure of Src SH3 (Yu *et al.*, 1992) have shown that the basic fold is formed from two approximately orthogonal β -sheets, each consisting of three strands. Figure 1 shows the topology of the SH3 domain and the nomenclature used in this paper. As in spectrin and Src SH3, a surface patch of conserved residues, consisting in its centre of four aromatic residues, is also present in Fyn SH3. These aromatic residues with exposed side chains are Tyr91+Tyr137 (bottom pair) and Trp119+Tyr132 (top pair), as shown in Figure 2.

The outer surface of Fyn SH3 is 53% hydrophobic compared with 54% for spectrin SH3, as calculated from the contribution of hydrophobic atoms (only C) and polar atoms (N and O) to the molecular surface. This is in the normal range for globular proteins (Janin *et al.*, 1988; Wierenga *et al.*, 1992). The conserved patch of aromatic residues seems to be more hydrophobic, although polar atoms also contribute to this surface (see Figure 8 below).

The two Fyn SH3 molecules have a similar fold over most of the structure, except for residues 112–118 (Figure 3). The r.m.s. coordinate difference between SH3-1 and SH3-2 is 0.3 Å for the 48 C α -atoms of residues 86–140, excluding residues 112–118. The latter residues do not superimpose very well. In Fyn SH3-1 this region forms most of the strand β b2 and the loop between strands β b2 and β c. It corresponds to the region where neuronal Src has a six residue insertion (after Glu116 in human Fyn SH3) (Martinez *et al.*, 1987). For this reason, we refer to this loop as the n-Src loop (Figure 1). The largest coordinate difference is observed for

Table I.

(a) Statistics of data collected

	Native	Uranyl
Number of measured reflections	37 654	16 950
Number of unique reflections	10 190	3330
Overall merging R-factor (%)	5.8	4.9
Maximum resolution (Å)	1.9	2.8
Completeness	86% (∞ –1.9 Å)	89% (∞ –2.8 Å)
	75% (2.0–1.9 Å)	67% (3.0–2.8 Å)

(b) Refinement statistics

Resolution used	5.0–1.9 Å
Reflections used	9808
Number of protein atoms	956
Number of water molecules	53
R-factor	18.0%
r.m.s. deviation from ideal bond lengths	0.02 Å
r.m.s. deviation from ideal bond angles	2.9°
χ^2 imperfection ^a	23.8°
Residues with non-allowed ϕ/ψ	0
r.m.s. B-difference between bonded atoms	7.3 Å ²

^ar.m.s. deviation from Ponder and Richards (1987) side chain torsion angle database.

Fig. 1. Topology diagram of Fyn SH3. The secondary structure assignment was calculated with the DSSP program (Kabsch and Sander, 1983), using the coordinates of Fyn SH3-1. The 3_{10} helix begins with Pro134. The RT-Src loop refers to the observation that point mutations of residues Arg95 and Thr96 in Src have been shown to have an effect on the transforming potential (Kato *et al.*, 1986). The n-Src loop is the site for insertions in the neuronal Src isoenzyme (Pyper and Bolen, 1990; Martinez *et al.*, 1987). The distal loop is away from the conserved surface patch.

the C α -atoms of residues Leu112 (3.6 Å) and Asn113 (3.5 Å). In Fyn SH3-2 these residues have moved outwards, away from β c, such that the strand β b2 has effectively been reduced to only one residue, Gln110. In Fyn SH3-2 two well defined water molecules have filled up the resulting space; these two water molecules mediate the interactions between residues Leu112 and Asn113 and strand β c. Interestingly, this site of structural diversity has a high intrinsic flexibility in both Fyn SH3 molecules in a temperature factor plot (Figure 4), although both loops have well defined electron density.

Fig. 2. α -trace of Fyn SH3-1. All atoms of the 14 conserved residues (see the marked residues of the spectrin SH3 sequence in Figure 4) are also shown. In this view Trp119 and Tyr132 are at the top of the picture and Tyr91 and Tyr137 are at the bottom of the picture.

Fig. 4. Plots of average main chain B-factors. Thick line: Fyn SH3-1; Thin continuous line: Fyn SH3-2; thin dashed line: spectrin SH3. The high B-factor regions are in the RT-Src loop (position 96), the n-Src loop (position 114) and the distal loop (position 126).

Comparison of SH3 domains

Fyn SH3 is two residues longer than the spectrin domain due to the insertion of Thr126 and Thr127. The sequences of Fyn SH3 and spectrin SH3 are 33% identical (19 identical residues out of 57 equivalent residues). The sequence identity between the SH3 domains in Fyn and Src is 77% (Figure 5). As is shown in Figure 3, the spectrin α -trace closely follows the α -trace of Fyn SH3-1. The r.m.s. coordinate difference in the superposition of spectrin SH3 and Fyn SH3-1 is 1.1 Å for the 51 α -atoms of the (Fyn) residues between 86 and 140, excluding residues 124–127. The largest deviations are in the loop near residues Thr126 and Thr127. This loop is rather far away from the conserved patch, and it is referred to as the distal loop in Figure 1.

The residues whose side chains are seen to point into the hydrophobic core of the Fyn SH3 molecule are Ala89, Phe103, Phe109, Ala122, Ile133 and Val138 (Figure 6). This set of amino acids is also obtained when one selects

for residues whose side chains do not contribute to the molecular surface. In spectrin SH3 the equivalent residues are Ala11, Met25, Leu31, Val44, Val53 and Val58. Apparently, the volume of the hydrophobic core is not conserved; in Fyn SH3 (23 side chain atoms) it is significantly larger than in spectrin SH3 (18 side chain atoms). By contrast, the surface patch of the proposed ligand-binding site (Musacchio *et al.*, 1992b; Yu *et al.*, 1992) shows considerable conservation: the 14 conserved residues marked in Figure 5 (at the spectrin SH3 sequence) are identical except for residues Glu22, Asp29 and Phe52, which are changed into Asp100, Glu107 and Tyr132, respectively, in Fyn and Src SH3 domains.

A domain in BirA has the SH3 fold

A supersecondary structure similar to the SH3 fold is also present in the BirA protein (Wilson *et al.*, 1992). This bacterial protein is bifunctional. It regulates the transcription

^{&} spectrin	[°] Fyn	1	2	secondary structure nomenclature	[°] Src
Lys6	Val84				
Glu7	Thr85				
Leu8	Leu86	β	β	βa	Thr85
Val9	Phe87	β	β	βa	Phe86
Leu10	Val88	β	β	βa	Val87
Ala11*	Ala89	β	β	βa	Ala88
Leu12*	Leu90				Leu89
Tyr13*	Tyr91				Tyr90
Asp14*	Asp92				Asp91
Tyr15*	Tyr93				Tyr92*
Glu16	Glu94				Glu93
Glu17	Ala95				Ser94
Lys18	Arg96			RT-Src loop	Arg95*
Ser19	Thr97				Thr96*
Pro20	Glu98				Glu97
Arg21	Asp99				Thr98*
Glu22*	Asp100				Asp99
Val23	Leu101				Leu100
Thr24	Ser102				Ser101
Met25	Phe103				Phe102
Lys26	His104				Lys103
Lys27	Lys105				Lys104
Gly28*	Gly106				Gly105
Asp29*	Glu107				Glu106
Ile30	Lys108	β	β	βb1	Arg107
Leu31	Phe109	β	β	βb1	Leu108
Thr32	Gln110	β	β	βb2	Gln109
Leu33	Ile111	β	β	βb2	Ile110
Leu34	Leu112	β	β	βb2	Val111*
Asn35	Asn113	β	β	βb2	Asn112
Ser36	Ser114				Asn113*
Thr37	Ser115			n-src loop	Thr114*
Asn38	Glu116				Glu115*
Lys39	Gly117				Gly116
Asp40	Asp118				Asp117
Trp41*	Trp119	β	β	βc	Trp118*
Trp42*	Trp120	β	β	βc	Trp119*
Lys43	Glu121	β	β	βc	Leu120*
Val44	Ala122	β	β	βc	Ala121
Glu45	Arg123	β	β	βc	His122
Val46	Ser124	β	β	βc	Ser123
Asn47	Leu125				Leu124
	Thr126			distal loop	Thr125
Asp48	Gly128				Thr126
Arg49	Glu129				Gly127
Gln50	Thr130	β	β	βd	Gln128
Gly51*	Gly131	β	β	βd	Thr129
Phe52*	Tyr132	β	β	βd	Gly130
Val53	Ile133	β	β	βd	Tyr131*
Pro54*	Pro134				Ile132
Ala55	Ser135				Pro133
Ala56	Asn136			3 ₁₀ -helix	Ser134
Tyr57*	Tyr137				Asn135*
Val58	Val138	β	β	βe	Tyr136*
Lys59	Ala139	β	β	βe	Val137
Lys60	Pro140	β	β	βe	Ala138
Leu61	Val141				Pro139
Asp62	Asp142				Ser140

Fig. 5. The sequences of chicken spectrin, human Fyn and chicken Src SH3 domains. Notes: [&]The listed sequence covers the structured part of spectrin SH3. Fourteen conserved residues identified by Musacchio *et al.* (1992a,b) are marked with (*). The renumbered spectrin sequence (Musacchio *et al.*, 1992b) corresponds to residues 969–1025 of the full length chicken α -spectrin (Wasenius *et al.*, 1989). [°]The tabulated residues cover the region of Fyn SH3, which is well defined by the electron density map. The bold residues are the conserved, exposed aromatic residues: Tyr91+Tyr137 are the 'bottom' pair and Trp119+Tyr132 the 'top' pair, as shown in Figure 2. Numbers 1 and 2 refer to Fyn SH3-1 and SH3-2. [°]The tabulated residues refer to the Src SH3 peptide, for which the NMR structure is described by Yu *et al.* (1992). The marked residues (*) have been identified in the NMR experiment as being close to the binding site for proline-rich peptides. The β -strands of spectrin and Fyn SH3 domains have been identified by the DSSP program (Kabsch and Sander, 1983). The secondary structure of the Src SH3 is taken from Yu *et al.* (1992).

of the *E. coli* biotin biosynthesis operon and it carries out a catalytic conversion of biotin. It has three domains, an N-terminal DNA-binding domain, which is followed by a biotin-binding domain, and a small C-terminal domain (44 residues) with an unknown function. The latter has a structure which is very similar to the SH3 fold. 30 residues of this domain can be superimposed on Fyn SH3 with an r.m.s. deviation of 1.8 Å for equivalent C α -atoms. Only two of these 30 residues have identical side chains. The RT-Src loop (Figure 1) is missing in the BirA domain (Figure 5). The absence of sequence identity suggests that the BirA domain and SH3 are not functionally related, despite the surprising structural similarity.

SH3 fold

All SH3 domains have the same topology, which is shown in Figure 1. For the crystal structures, the only variation between secondary structure seems to affect strand β b2 in the two forms of Fyn SH3. This strand is also shorter in the Src domain (Figure 4) despite the fact that in general the secondary structure assignment of the Src SH3 made by Yu *et al.* (1992) is more extensive than the secondary structure calculated according to Kabsch and Sander (1983) for the Fyn SH3.

The SH3 fold is a newly discovered one. So far it is clearly shared by only one other protein domain, the C-terminal domain in BirA (Wilson *et al.*, 1992); in addition, topological similarity between the R67 dihydrofolate reductase, a small enzyme encoded by an R-plasmid in Gram-negative bacteria (Matthews *et al.*, 1986), and SH3 has been pointed out by Murzin (1992). It is no wonder that the attempts to predict this novel structure *ab initio* have only been partially successful (see Rost and Sander, 1992; Benner *et al.*, 1992, 1993).

In the structures of Fyn and spectrin SH3 a few residues at the N- and C-terminus are not very well defined. These residues are apparently important for linking these domains to the rest of the protein molecule. The structured part of the Fyn and spectrin SH3 covers exactly the same fragment (Figure 5). It is not known if these SH3 domains are flexibly linked to the rest of the protein or if there is a tight association as in the BirA protein (Wilson *et al.*, 1992).

Ligand-binding surface

The proposed location of a ligand-binding site on SH3 (Figure 8) is based on two arguments. First, the residues that are conserved in the SH3 sequences cluster on one side of the molecule (Musacchio *et al.*, 1992a,b). Second, two peptides which have been mapped to the SH3-binding segment of 3BP-1 (Cicchetti *et al.*, 1992), perturb the NMR signals of a number of Src SH3 residues (Yu *et al.*, 1992). The perturbed residues are either among those conserved in the sequence or amino acids in the vicinity of conserved residues. However, the NMR experiments were carried out with millimolar concentrations of peptides, and it is not known what the actual binding constants of these peptides to SH3 might be.

The aromatic bottom and top pairs of residues in Fyn SH3-1 and SH3-2 are all involved in crystal contacts. This further suggests that these aromatic residues can be involved in protein–protein interactions. The aromatic patch consists predominantly of tyrosines and tryptophans. It is of interest

Fig. 6. A stereodiameter showing all protein atoms of Fyn SH3-1. The residues of the hydrophobic core (see text) as well as the N- and C-termini are labelled.

Fig. 7. α -trace of the BirA domain (thin line, residues 270–317) and Fyn SH3-1 (thick line). Same view as in Figure 2.

that the antigen-binding sites of antibodies also have a marked preference for these residues (Padlan, 1990; Chothia, 1991; Mian *et al.*, 1991).

The patch of conserved aromatic residues is flanked by two loops: the n-Src loop and the RT-Src loop (Figures 1 and 2). The residues in the RT-Src loop and the n-Src loop are not conserved between Fyn, Src and spectrin SH3s. For example, the RT-Src loop side chain Asp99 in Fyn is replaced with Thr98 in Src and with Arg21 in spectrin SH3, and the residues Ser115-Glu116-Gly117 in Fyn SH3 (Thr114-Glu115-Gly116 in Src) are replaced with residues Thr37-Asn38-Lys39 in spectrin SH3. The side chains of residues Arg21 and Asn38 of spectrin SH3 cover the top aromatic pair Trp41, Phe52, whereas the shorter residue at position 21 and the changed conformation of the equivalent residues at positions 37–39 result in the surface of the Trp119, Tyr132 pair in Fyn SH3 being much more exposed than the Trp41, Phe52 pair in the spectrin SH3.

The ligand-binding properties of the aromatic surface could be modulated by the residues of the RT-Src and n-Src loops. Such differences in the proposed ligand-binding site between the SH3 domains in tyrosine kinases and spectrin may indicate that they recognize slightly different proteins.

Ligand

Based on the homology between the partial sequence of 3BP-1 and proteins that have GAP activity, Cicchetti *et al.* (1992) have suggested that the SH3 domains may associate

with GAPs. The diversity of GTP-binding proteins and their regulatory factors matches with the occurrence of SH3 in cytoskeletal and signalling proteins, but the hypothesis that GTP-binding proteins or GAPs could interact with SH3 domains (see Mayer and Baltimore, 1993; Musacchio *et al.*, 1992a) needs to be scrutinized by future studies. With the possible exception of 3BP-1, peptides related to the putative proline-rich SH3-binding sites APTMPPLPP and PPAYPPPPVP, which have been identified by mutagenesis and a filter-binding assay (Cicchetti *et al.*, 1992; Ren *et al.*, 1993), do not occur in proteins belonging to GAP families. However, related peptides are common in diverse viral, microbial, plant and animal protein sequences in the current databases. There is no clear pattern in the occurrence of this motif in particular protein families, and neither has a proper quantitative determination of binding affinities been carried out. Therefore it remains to be seen whether such hydrophobic, proline-rich peptides are true SH3 ligands. However, it seems likely that the aromatic patch on the SH3 surface is indeed involved in protein–protein interactions.

Materials and methods

Expression and purification

The DNA sequence coding for the SH3 domain of *fyn* was amplified as a 200 bp fragment using PCR from a human cDNA clone (Semba *et al.*, 1986), making use of the two oligonucleotides, 5'-CGCGGCCCATGGG-AACAGGAGTGACACTCTTGTC-3' and 5'-GCGCGCAAGCTTGGAT-CCCTACTAAGAGTCAACTGGAGCC-3'.

Fig. 3. Superposition of Fyn SH3-1 (red) and SH3-2 (magenta) as well as spectrin SH3 (yellow). The view is approximately the same as that in Figure 2. Residues Leu112 and Asn113 (strand β 2) of Fyn SH3-2 adopt a unique conformation, not observed in Fyn SH3-1 or spectrin SH3.

Fig. 8. The molecular surface near the conserved aromatic residues. (A) Stereopair. (B) Monopicture with labels: Trp119 and Tyr132 are the top pair and Tyr91 and Tyr137 the bottom pair, as also shown in Figure 2. The hydrophobic atoms (only C) are in red, polar atoms (N and O) are in blue. The view is similar to that in Figure 2.

The amplified fragment codes for 64 residues (81–144) of the entire molecule. An *Nco*I site constructed at the 5' end of the coding sequence provides the AUG codon necessary for initiation of translation. A *Hind*III site was also introduced at the 3' end of the amplified fragment. The engineered sites allowed cloning into a modified pET vector (see Van der

Oost *et al.*, 1992). The plasmid pETfyn1 was used to express a 65 residue protein, comprising 64 amino acids of the Fyn SH3 domain, and an N-terminal methionine.

Expression in *E.coli* BL21(DE3) was carried out as described by Van der Oost *et al.* (1992). The bacterial pellet from a 6 l culture was resuspended

Fig. 9. An SDS-polyacrylamide gel showing purification of the Fyn SH3 domain. Lanes 1 and 2: whole-cell and soluble extracts of *E. coli* expressing the protein; lanes 3–5: pooled fractions containing Fyn SH3 after chromatography on Q-Sepharose, Superdex and Mono Q columns, respectively. Molecular weight markers are shown on the left, and the position of the 7.3 kDa SH3 band is indicated. See Materials and methods for details.

in 50 ml of standard Tris buffer (20 mM Tris-HCl buffer, pH 8.2, containing 1 mM DTT, 0.1 mM phenylmethyl sulfonyl fluoride and 0.1% benzamide) and the cells were broken by three cycles through a French press at 4°C. The suspension was centrifuged for 1 h at 40 000 r.p.m. in a Beckmann Ti45 rotor, and the clear supernatant was recovered.

This material was applied to a Q-Sepharose high performance anion-exchange column (XK-26, 60 ml bed volume), previously equilibrated with standard Tris buffer. The flow rate was 4 ml/min. A linear gradient of NaCl from 0 to 0.5 M in 480 ml was used to elute the proteins. The peak fraction containing the SH3 domain eluted at 400 mM NaCl. It was precipitated with 75% saturated ammonium sulphate, resuspended in a small volume of standard Tris buffer and loaded on to a Superdex 75 HiLoad gel filtration column (120 ml, Pharmacia). This column was equilibrated with the standard Tris buffer. The flow rate was 1 ml/min. SH3 eluted in two peaks, centred at 85 and 105 ml, indicating a possible dimer-monomer equilibrium. These peaks were pooled and applied to a Mono Q HR 10/10 column (Pharmacia) equilibrated with standard Tris buffer without protease inhibitors. Elution occurred at ~300 mM NaCl. The fraction containing highly purified Fyn SH3 domain (Figure 9) was concentrated with a Microsep Centrifugal Microconcentrator (Filtron Technology Corporation) to ~4 mg/ml. The yield of purified protein was ~3 mg from 6 l of bacterial cells. A 10 kDa impurity that was present in the preparation (Figure 9), does not apparently influence crystallization.

Crystallization, data collection and heavy atom derivatives

Crystallization was performed using the hanging drop technique at 4°C. The best crystals were obtained with vapour diffusion against a reservoir solution containing 4 M sodium formate (pH 8.3) and 0.6% β -mercaptoethanol. The protein (4 mg/ml in 20 mM Tris, 1 mM DTT, pH 8.2) was mixed 1:1 with this solution. Crystals with variable size formed in ~1 week. The presence of β -mercaptoethanol was not essential for crystal growth, but it improved the size and shape of the crystals. The best crystals reached the dimensions of $0.4 \times 0.4 \times 0.3$ mm³. The crystals are well ordered and diffract at least to 1.8 Å.

Data were collected using a FAST television area detector mounted on an Elliott GX21 area detector running at 40 mA, 70 kV. Rotation frames of 0.15° around an arbitrary axis were collected with an exposure time of 45 s per frame. Data collection was controlled with the program MADNES (Messerschmidt and Pflugrath, 1987), with data processing performed by the XDS package (Kabsch, 1988). Two batches of reflections, covering a total of 190° of data were scaled together by the program XSCALE. Statistics of the final data set are given in Table I. The space group is C2, with cell dimensions $a = 73.5$ Å, $b = 48.8$ Å, $c = 42.9$ Å, $\beta = 98.0^\circ$. V_m is ~2.7 Å³/dalton for two molecules per asymmetric unit.

Crystals that are transferred to the mother liquor containing heavy metal salts, tend to dissolve. This complicated the soaking experiments. The single heavy metal derivative was obtained by a macro-seeding technique. Crystals to be seeded were soaked overnight in 4 M sodium formate (pH 8.3), in the absence of β -mercaptoethanol. Hanging drops were prepared against a reservoir buffer containing 4 M sodium formate, also in the absence of β -mercaptoethanol. Drops were formed by mixing 1 μ l of protein solution, 2 μ l of reservoir buffer and 1 μ l of a solution containing 10 mM uranyl acetate dissolved in 4 M sodium formate. These drops were immediately seeded and kept in the dark at 4°C for about a week. After this a crystal was mounted for crystallographic analysis. The cell dimensions are very similar to native crystals ($a = 73.7$ Å, $b = 48.6$ Å, $c = 43.0$ Å, $\beta = 98.1^\circ$). The difference R-factor between structure factors of the native and uranyl derivative dataset is 16.2% at 3 Å resolution.

Structure determination and refinement

For the subsequent calculations the programs of the CCP4 package (CCP4, 1979) were used, unless otherwise stated. Attempts to solve the Fyn SH3 structure by molecular replacement using the refined coordinates of spectrin SH3 as a search model, proved unsuccessful. The difference Patterson calculated using the uranyl data could be interpreted in terms of a two site derivative. Phases calculated from one of the sites were used to locate the second site by double difference Fourier. The second site transpired to have a y coordinate nearly equal to that of the first ($\Delta Y = 0.5$ Å). The result of this is that the heavy atom structure had an approximate centre of symmetry, so that noncentric reflections of the protein will be assigned incorrect phases. An anomalous Patterson for the uranyl derivative indicated the existence of a strong anomalous signal out to 3 Å resolution, and it was hoped that this would provide enough information to allow correct phasing also of noncentric terms.

The subsequent SIRAS map (3 Å resolution) map was of very low quality but could be used to build a partial polyaniline model of one molecule using the BONES options of program O (Jones *et al.*, 1991). In an iterative process, this polyaniline model was improved and brought to include increasingly more of the entire molecule, using phase combination of phases from the partial model and the SIRAS phases. By this process, the electron density improved to the stage where the amino acid sequence became visible, and a model corresponding to ~80% of one molecule could be built. Electron density was apparent corresponding to the presumed second molecule in the asymmetric unit, but was of insufficient quality to allow interpretation. The eventual partial model of the first molecule was found to be inadequate to solve the rotation function for finding the orientation of the second molecule. However, the region of electron density corresponding to the first molecule could be used successfully in molecular replacement calculations to find the orientation and position of the second molecule (M. Noble, R. Abagyan, R. Wierenga, manuscript in preparation).

In order to test the validity of this solution, the 'a' suite of O programs was used to calculate the electron density correlation between molecule 1 and the presumed molecule 2. After improvement of the operator using `a_r_t_improve`, this correlation was 0.40 as calculated from the original SIRAS map, using the available partial model for defining the mask. In order to remove unwanted bias from the model built so far, we used the averaged map to build *ab initio* the Fyn SH3 molecule, using the BONES option of O. Subsequently cyclic averaging and solvent flattening, with phase combination, was used to improve the model. In the final stages of model building, the two SH3 molecules were built in ($2F_o - F_c$), α_c maps. The structure was then refined using XPLOR (Brünger *et al.*, 1987) and TNT (Tronrud *et al.*, 1987), first at 2.6 Å and eventually at 1.9 Å. In the final model molecules 1 and 2 have 59 residues. In both molecules there is no density for three residues at the N-terminus and two residues at the C-terminus. There are 53 waters in the final model; these water molecules are hydrogen-bonded to other polar atoms. The refinement statistics are summarized in Table I.

The structures were analysed with the WHAT IF program (Vriend, 1990) and molecular surfaces were calculated with the MSP package (Connolly, 1985), using parameters as described previously (Wierenga *et al.*, 1992).

Acknowledgements

We thank Dr B.W.Matthews for providing us with the coordinates of the BirA protein. The coordinates of the spectrin and Fyn SH3 domains will be deposited in the Brookhaven Protein Data Bank.

References

- Benner, S.A., Cohen, M.A. and Gerloff, D. (1992) *Nature*, **359**, 781.
- Benner, S.A., Cohen, M.A. and Gerloff, D. (1993) *J. Mol. Biol.*, **229**, 295–305.
- Booker, G.W., Breeze, A.L., Downing, A.K., Panayotou, G., Gout, I., Waterfield, M.D. and Campbell, I.D. (1992) *Nature*, **358**, 684–687.
- Broek, D., Toda, T., Michaeli, T., Levin, L., Birchmeier, C., Zoller, M., Powers, S. and Wigler, M. (1987) *Cell*, **48**, 789–799.
- Brünger, A.T., Kuriyan, J. and Karplus, M. (1987) *Science*, **235**, 458–460.
- CCP4 (1979) Daresbury Laboratory, Warrington.
- Chant, J., Corrado, K., Pringle, J.R. and Herskowitz, I. (1991) *Cell*, **65**, 1213–1224.
- Chenevert, J., Corrado, K., Bender, A., Pringle, J. and Herskowitz, I. (1992) *Nature*, **356**, 77–79.
- Chothia, C. (1991) *Curr. Opin. Struct. Biol.*, **1**, 53–59.
- Cicchetti, P., Mayer, B.J., Thiel, G. and Baltimore, D. (1992) *Science*, **257**, 803–806.
- Clark, S.G., Stern, M.J., and Horvitz, H.R. (1992) *Nature*, **356**, 340–344.
- Connolly, M.L. (1985) *J. Am. Chem. Soc.*, **107**, 1118–1124.
- Drubin, D.G., Miller, K.G. and Botstein, D.J. (1988) *Mol. Cell Biol.*, **107**, 2551–2561.
- Drubin, D.G., Mulholland, J., Zhu, Z. and Botstein, D. (1990) *Nature*, **343**, 288–290.
- Eck, M.J., Shoelson, S.E. and Harrison, S.C. (1993) *Nature*, **362**, 87–91.
- Hughes, D.A., Fukui, Y. and Yamamoto, M. (1990) *Nature*, **344**, 355–357.
- Jackson, P. and Baltimore, D. (1989) *EMBO J.*, **8**, 449–456.
- Janin, J., Miller, S. and Chothia, C. (1988) *J. Mol. Biol.*, **204**, 155–164.
- Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) *Acta Crystallogr.*, **A47**, 110–119.
- Kabsch, W. (1988) *J. Appl. Crystallogr.*, **21**, 916–924.
- Kabsch, W. and Sander, C. (1983) *Biopolymers*, **22**, 2577–2637.
- Kato, J., Takeja, T., Grandori, C., Iba, H., Levy, J.B. and Hanafusa, H. (1986) *Mol. Cell Biol.*, **6**, 4155–4160.
- Kawakami, T., Pennington, C.Y. and Robbins, K.C. (1986) *Mol. Cell Biol.*, **6**, 4195–4201.
- Kitamura, D., Kaneko, H., Miyagoe, Y., Ariyasu, T. and Watanabe, T. (1989) *Nucleic Acids Res.*, **17**, 9367–9379.
- Koch, C.A., Anderson, D., Moran, M.F., Ellis, C. and Pawson, T. (1991) *Science*, **252**, 668–674.
- Leto, T.L., Lomax, K.J., Volpp, B.D., Nunoi, H., Sechler, J.M.G., Nauseef, W.M., Clark, R.A., Gallin, J.I. and Malech, H.L. (1990) *Science*, **248**, 727–730.
- Martinez, R., Mathey-Preudt, B., Bernards, A. and Baltimore, D. (1987) *Science*, **237**, 411–415.
- Matthews, D.A., Smith, S.L., Baccanari, D.P., Burchall, J.J., Oatley, S.J. and Kraut, J. (1986) *Biochemistry*, **25**, 4194–4204.
- Mayer, B.J. and Baltimore, D. (1993) *Trends Cell Biol.*, **3**, 8–13.
- Mayer, B.J., Jackson, P.K. and Baltimore, D. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 627–631.
- McCaffrey, G., Clay, F.J., Kelsay, K. and Sprague, G.F., Jr (1987) *Mol. Cell Biol.*, **7**, 2680–2690.
- Mian, S., Bradwell, A.R. and Olson, A.J. (1991) *J. Mol. Biol.*, **217**, 133–151.
- Messerschmidt, A. and Pflugrath, J.W. (1987) *J. Appl. Crystallogr.*, **20**, 306–315.
- Morris, A.L., MacArthur, M.W., Hutchinson, E.G. and Thornton, J.M. (1992) *Proteins*, **12**, 345–364.
- Murzin, A.G. (1992) *Nature*, **360**, 635.
- Musacchio, A., Gibson, T., Lehto, V.-P. and Saraste, M. (1992a) *FEBS Lett.*, **307**, 55–61.
- Musacchio, A., Noble, M.E.M., Paupit, R., Wierenga, R. and Saraste, M. (1992b) *Nature*, **359**, 851–855.
- Otsu, M. et al. (1991) *Cell*, **65**, 91–104.
- Overduin, M., Rios, C.B., Mayer, B.J., Baltimore, D. and Cowburn, D. (1992) *Cell*, **70**, 697–704.
- Padlan, E.A. (1990) *Proteins*, **7**, 112–124.
- Ponder, J.W. and Richards, F.M. (1987) *J. Mol. Biol.*, **193**, 775–791.
- Pyper, J.M. and Bolen, J.B. (1990) *Mol. Cell Biol.*, **10**, 2035–2040.
- Ren, R., Mayer, B.J., Cicchetti, P. and Baltimore, D. (1993) *Science*, **259**, 1157–1161.
- Rost, B. and Sander, C. (1992) *Nature*, **360**, 540.
- Sahr, K.E. et al. (1990) *J. Biol. Chem.*, **265**, 4434–4443.
- Seidel-Dugan, C., Meyer, B.E., Thomas, S.M. and Brugge, J.S. (1992) *Mol. Cell Biol.*, **12**, 1835–1845.
- Semba, K., Nishizawa, M., Miyajima, N., Yoshida, M.C., Sukegawa, J., Yamanashi, Y., Sasaki, M., Yamamoto, M.C. and Toyoshima, K. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 5459–5463.
- Stahl, M.L., Ferenz, C.R., Kelleher, K.L., Kriz, R.W. and Knopf, J.L. (1988) *Nature*, **332**, 269–272.
- Trahey, M. et al. (1988) *Science*, **242**, 1697–1700.
- Tronrud, D.E., Ten Eyck, L.F. and Matthews, B.W. (1987) *Acta Crystallogr.*, **43A**, 489–501.
- Trueheart, J., Boeke, J.D. and Fink, G.R. (1987) *Mol. Cell Biol.*, **7**, 2316–2328.
- Van der Oost, J. et al. (1992) *EMBO J.*, **11**, 3209–3217.
- Volpp, B.D., Nauseef, W.M., Donelson, J.E., Moser, D.R. and Clark, R.A. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 7195–7199.
- Vriend, G. (1990) *J. Mol. Graph.*, **8**, 52–56.
- Waksman, G. et al. (1992) *Nature*, **358**, 646–653.
- Waksman, G., Shoelson, S.E., Pant, N., Cowburn, D. and Kuriyan, J. (1993) *Cell*, **72**, 779–790.
- Wasenius, V., Saraste, M., Salven, P., Erämaa, M., Holm, L., Lehto, V.-P. (1989) *J. Cell Biol.*, **108**, 79–93.
- Wierenga, R.K., Noble, M.E.M. and Davenport, R.C. (1992) *J. Mol. Biol.*, **224**, 1115–1126.
- Wilson, K.P., Shewchuk, L.M., Brennan, R.G., Otsuka, A.J. and Matthews, B.W. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 9257–9261.
- Yu, H., Rosen, M.K., Shin, T.B., Seidel-Dugan, C., Brugge, J.S. and Schreiber, S.L. (1992) *Science*, **258**, 1665–1668.

Received on January 15, 1993; revised on March 29, 1993