Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk

Xiaodong Wu[1], Beatrice Knudsen[1], Stephan M Feller[1], Jie Zheng[1], Andrej Sali[2†], David Cowburn[1], Hidesaburo Hanafusa[1] and John Kuriyan[1,3*]

[1]The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA, [2]Department of Chemistry, Harvard University, Cambridge, MA 02138, USA and [3]Howard Hughes Medical Institute, 1230 York Avenue, New York, NY 10021, USA

Background: Proline-rich segments in the guanine nucleotide exchange factor C3G bind much more strongly to the N-terminal Src homology 3 domain (SH3-N) of the proto-oncogene product c-Crk than to other SH3 domains. The presence of a lysine instead of an arginine in the peptides derived from C3G appears to be crucial for this specificity towards c-Crk.

Results: In order to understand the chemical basis of this specificity we have determined the crystal structure of Crk SH3-N in complex with a high affinity peptide from C3G (PPPALPPKKR, K_d ~2 μM) at 1.5 Å resolution. The peptide adopts a polyproline type II helix that binds, as dictated by electrostatic complementarity, in reversed orientation relative to the orientation seen in the earliest structures of SH3–peptide complexes. A lysine in the C3G peptide is tightly coordinated by three acidic residues in the SH3 domain. In contrast, the co-crystal structure of c-Crk SH3-N and a peptide containing an arginine at the equivalent position (determined at 1.9 Å resolution) reveals non-optimal geometry for the arginine and increased disorder.

Conclusions: The c-Crk SH3 domain engages in an unusual lysine-specific interaction that is rarely seen in protein structures, and which appears to be a key determinant of its unique ability to bind the C3G peptides with high affinity.

Structure 15 February 1995, **3**:215–226

Key words: c-Crk, peptide binding, polyproline helix, SH3 domain

Introduction

The viral oncogene product Crk was the first member of the family of 'adapter proteins' to be discovered [1,2]. These proteins are important in cellular signal transduction. The surprising aspect of the discovery of Crk was the realization that despite its ability to transform cells and to modulate the activity of tyrosine kinases, it contains only modular units known as SH2 and SH3 domains (Src homology domains) that have no intrinsic catalytic or transcriptional activity [3]. It is now well established that the functions of adapter proteins are indeed a consequence of the SH2 and SH3 domains, which mediate protein–protein interactions via recognition of phosphorylated tyrosines and proline-rich sequences, respectively [2,4–6]. Other members of this family include Nck, and the various homologs of Grb2/Sem5 [6]. These proteins form multi-dentate complexes with diverse signalling proteins in response to external stimuli, and serve to localize these proteins at the sites of receptor phosphorylation. This has been demonstrated most clearly for Grb2/Sem5, which binds to tyrosine-phosphorylated proteins at the cell membrane mediated by its SH2 domain and, simultaneously, to the guanine nucleotide exchange factor, Sos, through its SH3 domain. Membrane localization of Sos results in the activation of Ras (for a review, see [2]).

The cellular form of the Crk protein (c-Crk) consists of an SH2 domain followed by two SH3 domains [7]. Screening of expression libraries for proteins that bind to the SH3 domains of c-Crk led to the identification of C3G, a 145–155 kDa protein that contains a region of homology to Ras guanine nucleotide exchange factors as well as a segment containing four homologous proline-rich stretches incorporating the PXXP motifs [8] that are characteristic of SH3 targets [9,10]. The interaction with C3G is mediated entirely by the N-terminal SH3 domain of c-Crk (SH3-N), which is quite divergent in sequence from the C-terminal SH3 domain. c-Crk is also known to bind to other proteins, such as Sos and the tyrosine kinases Abl and Arg, via its SH3 domains [11,12]. Alignment of proline-rich sequences from these Crk-binding proteins resulted in a preliminary consensus for c-Crk SH3-N binding: PXLPXK/R. Peptides from C3G, Abl and Arg contain lysine at the last position, whereas those from Sos contain arginine [10]. Despite the apparent similarity in their proline-rich SH3-binding sequences, C3G binds selectively to the c-Crk SH3-N domain whereas Sos binds to the SH3 domains of c-Crk, Grb2, Src and Nck.

In a filter-binding assay using the proline-rich segment of C3G fused to glutathione S-transferase (GST), the SH3 domains of Nck, Src, Abl, phospholipase Cγ and spectrin were unable to bind to the C3G segment [10]. The SH3 domains of Grb2 and the p85 subunit of phosphatidyl-inositol 3-kinase bound weakly to C3G in the

*Corresponding author. †Present address: The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.

filter-binding assays [9,10], but would not precipitate from solution in enzyme-linked immunosorbent assays [10]. Quantitative measurement of the interaction between various proline-rich peptides and SH3 domains by fluorescence spectroscopy has revealed that the high affinity C3G peptide, PPPALPPKKR (referred to as C3G in this paper), binds to the c-Crk SH3-N domain with a dissociation constant (K_d) of 1.9 μM, which is one of the highest affinities reported for an SH3 domain (Table 1) (B Knudsen *et al.*, unpublished data). The C3G peptide binds much more weakly to the Grb2 N-terminal SH3 domain, with a K_d of 142 μM. In contrast, a peptide from Sos which binds tightly to the c-Crk SH3-N domain (PPPVPPRRRR, referred to as Sos; K_d=5.2 μM) also binds tightly to Grb2 (K_d=3.5 μM) (B Knudsen *et al.*, unpublished data).

Table 1. Binding affinities of proline-rich peptides for the SH3 domains of of c-Crk (N-terminal domain), v-Crk and Grb2 (N-terminal domain).

Peptide												K_d (μM) c-Crk SH3-N	v-Crk SH3	Grb2 SH3-N
C3G	P	P	P	A	L	P	P	K	K	R		1.9	25.5	142.0
Sos		P	P	**P**	**V**	P	P	**R**	**R**	R	**R**	5.2	23.7	3.5
C3G (mutant)	P	P	P	A	L	P	P	**R**	K	R		17.2	23.5	27.9

Data are taken from an extensive study of peptide binding to the Crk SH3 domain, carried out by Knudsen *et al.* (unpublished data). The data shown are for glutathione S-transferase–SH3 fusion proteins, and are obtained by measurement of the changes in intrinsic tryptophan fluorescence upon peptide binding. Very similar results are obtained with the isolated Crk SH3 domain used in the crystallographic study (data not shown). All measurements were carried out in phosphate buffered saline at 18°C. Sites that differ from the C3G peptide are shown in bold.

The promiscuity displayed by the Sos peptide is typical of SH3-binding peptides, which most often only discriminate between broad classes of SH3 domains and rarely manifest the more specific discrimination shown by the C3G peptides. Mutagenesis of the C3G peptide showed that the presence of a lysine rather than an arginine as the first basic residue following the PXXP motifs is crucial for this specificity (B Knudsen *et al.*, unpublished data). Replacement of the first lysine by arginine in the C3G peptide, keeping everything else the same, reduces the affinity for c-Crk SH3-N by a factor of nine (K_d=17.2 μM) and significantly increases the affinity for Grb2 SH3-N (K_d=27.9 μM), leading to loss of specificity (Table 1). The critical role played by a lysine residue in the determination of specificity is noteworthy, because arginine rather than lysine is usually the residue of choice for the formation of electrostatic interactions with high specificity. Interestingly, the SH3 domain of the viral form of Crk, v-Crk, which differs from c-Crk SH3-N only at two amino acid positions, is not sensitive to the presence of arginine *versus* lysine in the peptides. This indicates that a particular localized interaction underlies the ability of c-Crk SH3-N to preferentially bind to the lysine-containing peptides. We therefore initiated a project to obtain high-resolution crystal structures of peptide complexes of the c-Crk SH3 domain in order to understand the mode of peptide binding in general and the role of the lysine in particular.

The general features of SH3 structure and the interaction with proline-rich peptides are now understood [5,13]. The structures of a number of uncomplexed SH3 domains have been determined and have established a highly conserved fold consisting of two antiparallel β-sheets packed against each other (reviewed in [5]). The discovery that SH3 domains bind to sequences that contain multiple prolines with characteristic PXXP motifs [8,14] was followed by the proposal that the SH3 surface could recognize a polyproline type II helix [15] and by the determination of the three-dimensional structures of peptide complexes of the p85 SH3 domain by NMR [16] and the Abl and Fyn SH3 domains by X-ray crystallography [17]. These first structures of SH3 complexes revealed that a conserved set of residues on the SH3 surface provide hydrophobic interactions with side chains presented by the polyproline type II helix.

In each of these structures, the peptide bound in the same 'plus' orientation. While this manuscript was in preparation, four papers have been published that describe the interaction of proline-rich peptides with the Src SH3 domain [18] and with the SH3 domains of mammalian Grb2 [19,20] and its *Caenorhabditis elegans* homolog Sem-5 [21]. The structures of Src SH3 domain bound to two peptides obtained by selection from a combinatorial library revealed that whereas one peptide bound in the 'plus' orientation, the other bound in the opposite, 'minus', orientation. The three structures of Grb2/Sem-5 SH3 domains show that the Sos peptides bind in the 'minus' orientation. The ability of different proline-rich peptides to bind in opposite orientations to the same binding sites on the SH3 surface can be understood in terms of the symmetry of the polyproline helix and the specific electrostatic and hydrophobic packing interactions observed in these complexes. Analyses of these structures have led to the enunciation of general rules for describing the mode of binding of proline-rich peptides to SH3 domains [18,21].

In these recent structures of the Src [18] and Grb2/Sem5 [19–21] peptide complexes, the orientation of the peptide is determined primarily by electrostatic interactions between arginine residues in the peptide and acidic residues localized at one end of the peptide-binding surface of the SH3 domains. Many SH3 domains have acidic residues in this region. For example, most SH3 domains have a glutamate or aspartate residue, corresponding to Glu172 in Sem-5, that is observed to hydrogen bond to an arginine residue in the Sos peptide [21]. Likewise, the central interactions mediated by PXXP motifs involve highly conserved

hydrophobic residues in the SH3 domains. Consequently, the basis for sequence discrimination between proline-rich peptides and particular SH3 domains requires further elaboration.

Here we present high-resolution crystal structures of c-Crk SH3-N bound to two high affinity peptides, one from C3G and one from Sos. Both peptides bind in the minus orientation, and the general features of the peptide interaction are very similar to those seen for Sos peptides bound to Grb2/Sem5 [19–21]. However, comparison of the interactions between the c-Crk SH3 domain and a lysine (in the C3G peptide) and an arginine (in the Sos peptide) sheds light on the structural basis for the preferential binding of c-Crk SH3-N to C3G and illustrates how subtle variations in the structure of a conserved interface can lead to binding specificity.

Results and discussion

Structures of c-Crk SH3-N in complex with C3G and Sos peptides

Crystals of the two peptides bound to c-Crk diffract to very high resolution, and the datasets used have been limited to Bragg spacings of 1.5 Å and 1.9 Å for the C3G and Sos complexes, respectively, by experimental constraints such as the time available for data collection. Fig. 1a shows electron density at 1.5 Å resolution for the C3G peptide bound to the c-Crk SH3 domain. Strong density is seen for the first eight residues of the peptide (PPPALPPK) and somewhat weaker density for the lysine at position 9. The C-terminal arginine is disordered and not visible in the electron density. The peptide adopts a left-handed polyproline type II helix over most of its length (Fig. 1b, Table 2), and is bound with the C-terminal residues near the RT loop (so called

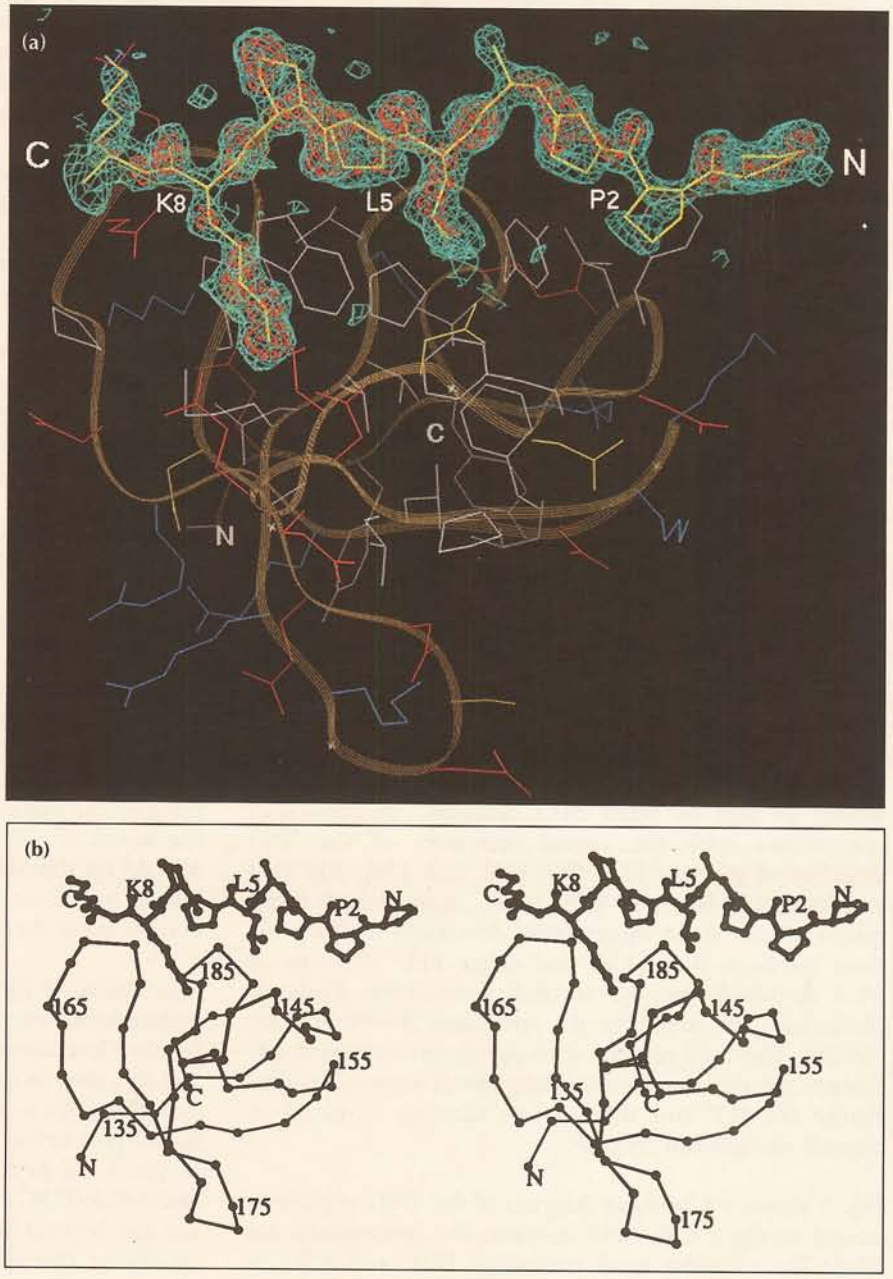

Fig. 1. **(a)** Difference electron-density map for the C3G peptide bound to c-Crk SH3-N. The C3G peptide was removed from the final model of the Crk–C3G complex and an electron-density map was calculated using (F_o–F_c) coefficients and model phases, using data between 10.0 Å and 1.5 Å spacings. Positional refinement without peptide (100 steps) preceded the calculation of the map. Electron-density contours at the 2.0σ (in blue) and 4.0σ (in red) levels are shown. The C3G peptide is shown in yellow within the density, with Pro2, Leu5 and Lys8 residues labeled. The backbone of c-Crk SH3-N is represented as a ribbon. Acidic side chains are red, basic side chains are blue, polar side chains are yellow and non-polar side chains are white. **(b)** Stereo diagram of the c-Crk SH3-N domain in complex with the C3G peptide (PPPALPPKKR). The Cα backbone of the SH3 domain is shown, as well as a ball-and-stick representation of the C3G peptide. Every tenth residue in the SH3 domain is numbered, and Pro2, Leu5 and Lys8 of C3G are labeled. (Figure generated using MOLSCRIPT [47].)

Table 2. Backbone torsion angles and structural deviations in the C3G and Sos peptides.

Position	Peptide	Residue	φ	ψ	Deviation from ideal PPII helix (Å)	Deviation between C3G and Sos (Å)
P_4	C3G	Proline	ND	161.1	4.74	
P_3	C3G	Proline	−62.9	168.9	2.49	0.59
	Sos	Proline	ND	166.8	1.56	
P_2	C3G	Proline	−78.8	175.6	0.35	0.35
	Sos	Proline	−71.2	164.6	0.38	
P_1	C3G	Alanine	−60.4	138.8	0.24	0.61
	Sos	Proline	−64.4	137.7	0.37	
P_0	C3G	Leucine	−81.9	124.2	0.27	0.48
	Sos	Valine	−81.3	121.0	0.35	
P_{-1}	C3G	Proline	−68.4	152.7	0.26	0.42
	Sos	Proline	−61.9	150.0	0.29	
P_{-2}	C3G	Proline	−70.8	151.9	0.07	0.35
	Sos	Proline	−68.9	160.6	0.09	
P_{-3}	C3G	Lysine	−91.5	143.0	0.67	0.28
	Sos	Arginine	−97.5	160.3	0.43	
P_{-4}	C3G	Lysine	−118.4	ND	0.40	0.41
	Sos	Arginine		ND	0.40	

The positions in the polyproline helix are shown in Fig. 3. The φ and ψ angles for an ideal polyproline type II (PPII) helix are −78° and +149°, respectively [46]. An ideal PPII helix was generated and aligned with the central residues (P_1 to P_{-2}) of the C3G and Sos helices, respectively. The deviations in Cα positions between the ideal and actual helices are shown. The last column shows the deviations in Cα positions between the C3G and Sos peptides after the protein structures are superimposed by least-squares. ND, not determined.

because this loop in the tyrosine kinase Src contains critical arginine and threonine residues) and the n-Src loop (the site of an insertion in the sequence of neuronal Src) (Fig. 1). A striking aspect of the electron density is that the strongest density features associated with the peptide are for the lysine side chain at position 8. The electron density is stronger for the terminal atoms of the side chain than for the backbone, indicating unusually tight interactions between the lysine and the SH3 domain.

The overall fold of the c-Crk SH3-N domain is very similar to that of other SH3 domains. In particular, comparison with the crystal structures of the SH3 domains of spectrin [22], Fyn [23], Lck [24], Abl [17] and Nck (X Wu and J Kuriyan, unpublished data), results in root mean square (rms) deviations in Cα positions between the c-Crk and other SH3 domains of ~0.5 Å–0.6 Å for the secondary-structural elements (deviations are larger for the structures determined by NMR). The only regions with significant structural difference are the loops connecting the β-strands, in particular the RT and n-Src loops that are involved in peptide recognition (Fig. 2).

Fig. 3 shows a schematic diagram of the C3G peptide as bound to the c-Crk SH3 domain. For consistency, we adopt the notation used previously [21], and refer to sites on the polyproline helix as P_{-1}, P_0, P_{+1} etc., where Leu5 in the C3G peptide occupies the P_0 position (Fig. 3). Using this notation, there are six binding sites for the peptide on the c-Crk SH3 surface, corresponding to P_{+3} (Pro2 in C3G), P_{+2} (Pro3), P_0 (Leu5), P_{-1} (Pro6), P_{-3} (Lys8) and P_{-4} (Lys9). The requirement for proline residues at certain positions in the peptide has been explained by considering the disposition of non-proline side chains with respect to the SH3 surface at various positions along the polyproline type II helix [18,21]. The polyproline helix has a triangular cross-section, and side chains along two of the edges interact with the surface of the SH3 domain. Non-proline side chains along one edge tend to adopt conformations that extend away from the SH3 surface, leading to poorer interactions (referred to as 'external packing' [21]). Proline side chains at this edge pack well against the SH3 surface. Non-proline side chains on the other edge of the helix point into the SH3 surface, and can replace proline at these positions (referred to as 'internal packing' [21]). The third edge of the helix points away from the SH3 surface and can be substituted with non-proline residues, although proline may be favored here to increase the proline content of the peptides with consequent stabilization of the helix. The characteristic PXXP motif that is a feature of SH3 target sites is thus seen to result from the need to preserve proline residues at sites of external packing, and the positions of the prolines will differ depending on the orientation of the peptide [18,21]. For C3G bound to c-Crk SH3, sites P_{+2} and P_{-1} correspond to external packing positions and have proline residues. The sites at P_{+3}, P_0 and P_{-3} correspond to internal packing positions, and are occupied by a proline, a leucine and a lysine, respectively (Fig. 3).

Interactions between the C3G peptide and the SH3 domain are illustrated in Fig. 4a. The peptide residues at positions P_{-1}, P_0, P_{+2} and P_{+3} are closely packed against conserved hydrophobic residues in the SH3 domain (Phe141, Phe143, Pro183, Tyr186, Trp169 and Pro185). There are three prominent ridges formed on the surface of the SH3 domain by the side chains of these hydrophobic groups (Fig. 4b), and their spacing is such that a close fit arises between the side chains presented by the polyproline helix and the SH3 surface [15], resulting in the burial of 670 $Å^2$ of surface area on the peptide and 450 $Å^2$ on the SH3 domain. These interactions are very similar to those observed in the Src and Grb2/Sos complexes in the minus orientation [19–21].

The ability of the polyproline helix to make equivalent hydrophobic interactions in either orientation is reflected by the close overlap between the residues at the P_0, P_{+2} and P_{+3} sites in the structure of the Abl SH3 domain in complex with a peptide from the 3BP1 protein (bound in the plus orientation) [17], and corresponding residues in the C3G peptide (Fig. 2). However, the Abl/3BP1 and c-Crk/C3G structures diverge significantly at the P_{-3} site and beyond because of fundamentally different interactions in this region. The cleft between the RT and

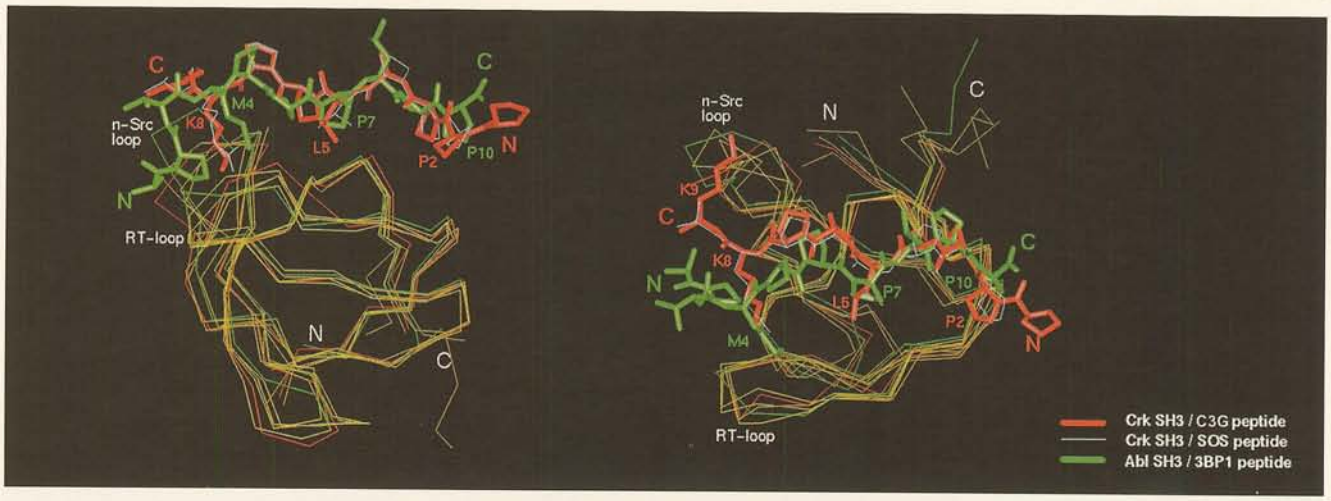

Fig. 2. Comparison of various uncomplexed and peptide-complexed SH3 structures determined by X-ray crystallography. The Cα backbone of a number of SH3 domains, aligned by least-squares superposition, are shown in this figure. The uncomplexed SH3 domains are from Fyn [23], Lck [24], spectrin [22] and Nck (X Wu and J Kuriyan, unpublished data) (all in yellow) and the three complexed SH3 structures are c-Crk SH3/C3G (in red), c-Crk SH3/Sos (in white) and Abl SH3/3BP1 (in green), [17]. To emphasize the reversed orientation of the C3G and Sos peptides with respect to the 3BP1 peptide, their N and C termini are labeled.

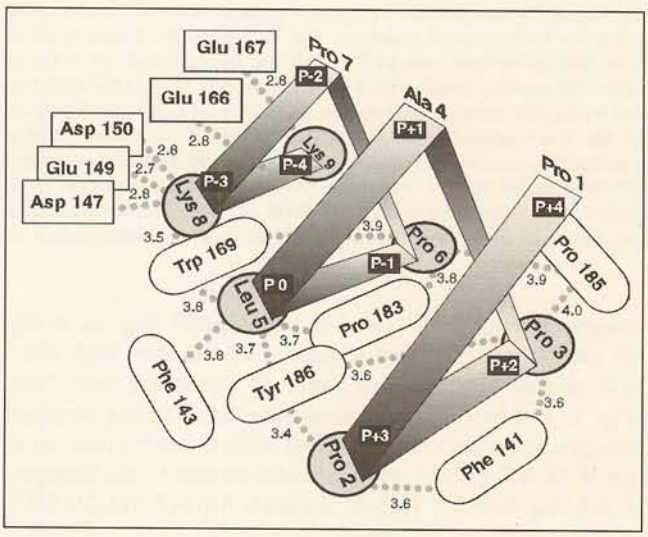

Fig. 3. Schematic diagram showing interactions between the C3G peptide and the c-Crk SH3-N domain. The C3G peptide is represented by the left-handed shaded ribbon, and peptide residues that interact with the SH3 domain are indicated by shaded circles. Residues in c-Crk SH3-N that interact with the peptide are indicated by oval boxes (for hydrophobic residues) and rectangular boxes (for acidic residues). Distances (in Å) between interacting residues are shown for the nearest pair of carbon atoms (for hydrophobic interactions) and between donor atom and acceptor atom for hydrogen bonds.

n-Src loops in Abl is hydrophobic, allowing the N-terminal hydrophobic region of the 3BP1 peptide to wrap around the SH3 domain in a non-helical conformation [17]. The presence of five acidic residues in the RT and n-Src loops of the c-Crk SH3 domain would prevent the hydrophobic N terminus of the peptide from interacting in this region. Instead, the two lysines in the peptide are bound here, and make a number of hydrogen-bonding and packing interactions (Fig. 4).

The Sos peptide is bound to c-Crk SH3-N in a very similar conformation as seen for the C3G peptide. There are six differences in amino acid sequence between the two peptides. The proline at position P_{+4} in C3G is missing in Sos, proline is substituted for alanine at P_{+1}, valine for leucine at P_0, arginine for lysine at P_{-3} and P_{-2} and an extra arginine is present after P_{-4}. The deviations in Cα positions between the two peptides are in the range from ~0.3–0.6 Å (Table 2), after the protein structures have been superimposed. Deformations in the Sos peptide with respect to the C3G peptide around the P_0 site result in a movement of the valine (in Sos) towards the SH3 surface, so that interactions are similar to those seen for the leucine side chain in C3G. Both structures overlap reasonably well with an ideal polyproline type II helix (Table 2). If the residues at positions P_1 to P_{-2} are superimposed upon the ideal helix, the deviations in Cα positions between the ideal and actual helices are in the range 0.07–0.37 Å for these residues. The ends of the peptide are distorted away from the ideal helix by as much as 2.5 Å (Table 2), with the distortions resulting in an increase in the interactions with the SH3 surface.

Lysine-specific interactions with the C3G peptide

There are four acidic residues in the RT loop of the murine c-Crk SH3-N domain: Asp147, Glu148, Glu149, and Asp150. Of these, Glu148 points away from the peptide and does not participate directly in binding interactions. The other three residues approach each other closely, and present three oxygen atoms that form a nearly equilateral triangle (oxygen–oxygen distances in the range 4.3–4.6 Å), with the lysine nitrogen at the geometric center and at a distance of 1.0 Å from the plane of the oxygens (nitrogen–oxygen distances in the range 2.7–2.8 Å). The binding site is thus remarkably well suited for interaction with lysine, since each of the three protons of the sp^3 hybridized amino group can be donated to carboxylate

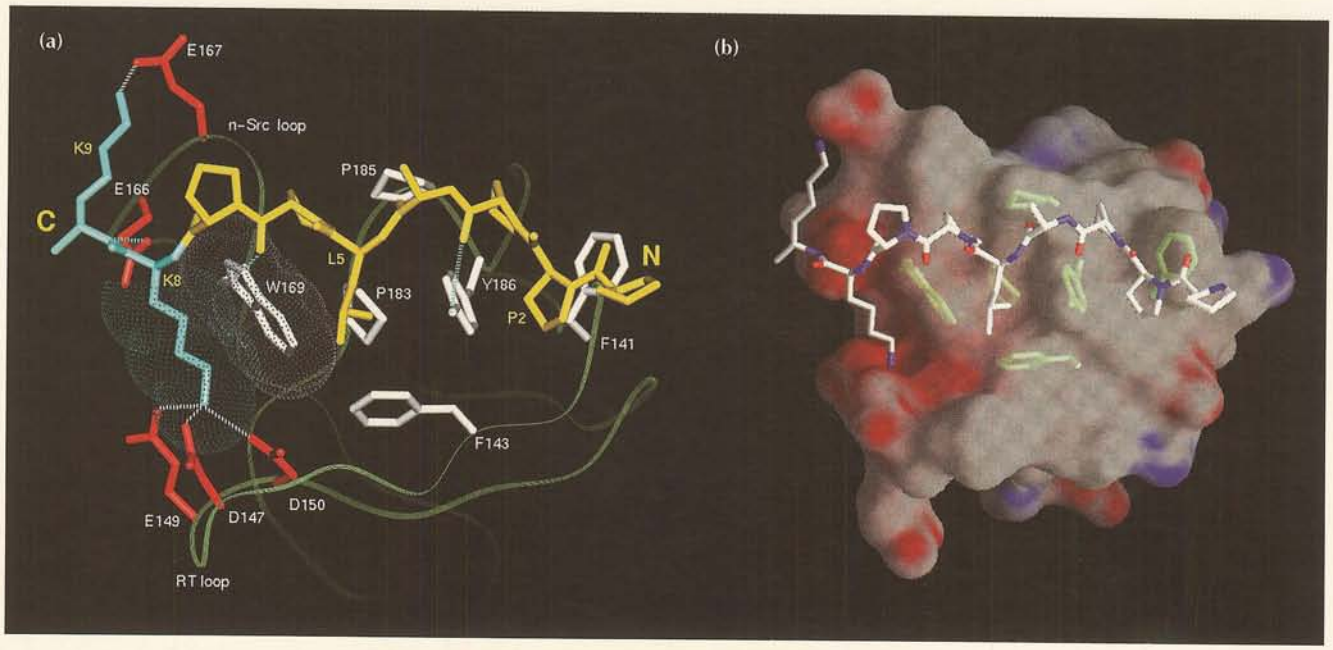

Fig. 4. The interface between c-Crk SH3-N domain and the C3G peptide. **(a)** Important hydrophobic and polar interactions between c-Crk SH3-N and the C3G peptide. The polypeptide backbone of the SH3 domain is represented by a green ribbon, and SH3 side chains that interact with the peptide are colored red for acidic residues and white for hydrophobic residues. The RT and n-Src loops flanking the C-terminal region of the C3G peptide are labeled. The C3G peptide is colored yellow, except for Lys8 and Lys9, which are colored blue. The van der Waals surfaces of Lys8 (in C3G) and Trp169 (in Crk) are indicated by small dots. Hydrogen bonds associated with the side chains of the C3G peptide are indicated by dashed white lines, and hydrogen bonds made to the backbone of C3G are drawn in blue. [Figure generated using QUANTA (Molecular Simulations, Inc.).] **(b)** The molecular surface of the c-Crk/C3G complex, calculated with the peptide removed and displayed using GRASP [48]. The surface is colored according to the local electrostatic potential calculated in the absence of peptide, assuming a 0.10 M NaCl concentration in the solvent. Note the negatively charged pocket (red) surrounding the Lys8 side chain of the C3G peptide and the neutral region (gray) interacting with Pro2–Pro6 of C3G. The hydrophobic side chains underneath the SH3 surface are displayed in green. The orientation in (b) is the same as in (a), for ease of identification of the residues.

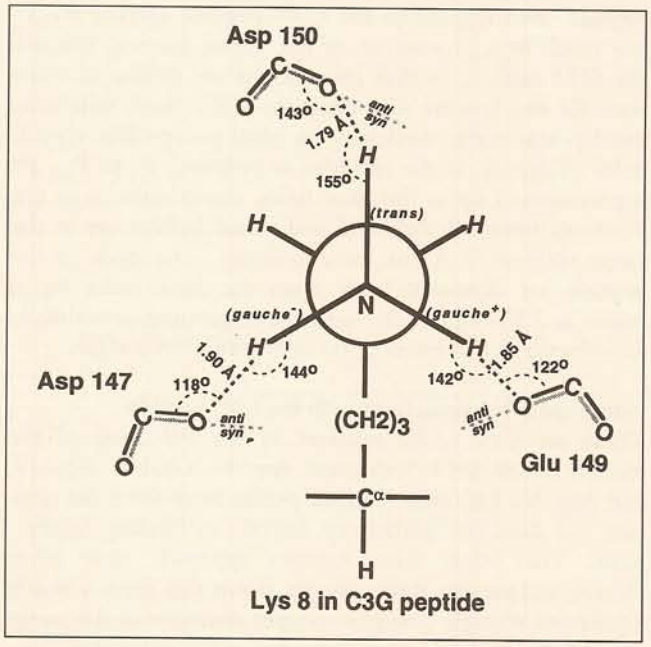

Fig. 5. Schematic diagram of the stereochemistry of the lysine–carboxylate interaction at position P_{-3} (after Ippolito *et al.* [26]). The amino group of the peptide lysine is shown in a Newman projection, and the relative disposition of the carboxylate groups and the hydrogen-bonding hydrogens in the C3G/Crk complex are shown.

oxygens from Asp147, Glu149 and Asp150 (Fig. 5). A stable interaction results, as suggested by the very high electron density level for the atoms of the lysine side chain (Fig. 1) and by the low temperature factor of the terminal nitrogen atom (6.3 Å^2 compared with the side chain average of 18 Å^2) and the oxygen atoms nearest to the nitrogen (6.3 Å^2 for Asp150, 13.5 Å^2 for both Asp147 and Glu149).

The lysine is in a fully extended conformation, and it approaches the triad of oxygen atoms directly from above (Fig. 4), leading to near-optimal geometry for simultaneous hydrogen bonding in the *trans*, *gauche*$^+$, and *gauche*$^-$ orientations (Fig. 5). In particular, each of the nitrogen-hydrogen bonds is almost coplanar with the corresponding carboxyl group, and the hydrogen–oxygen distances are within the range expected for tight hydrogen-bonding interactions (~1.8 Å [25]). One feature that is somewhat less than optimal is that nitrogen–hydrogen bonds approach residues Asp147 and Glu149 in the slightly less favored *anti* rather than *syn* orientation (Fig. 5) [26]. However, the carbon–oxygen–hydrogen angles are consistent with the appropriate interactions between the hydrogen and the lone pair electrons of the oxygen in either the *syn* or *anti* orientation [26].

This precise alignment of the lysine appears to be maintained by other interactions between the peptide and the

SH3 domain. The polypeptide backbone of the peptide is held in place by the formation of a hydrogen bond between the side chain of Glu166 and the backbone amide nitrogen at position 9 of the peptide, and by a conserved hydrogen bond [21] between the side chain nitrogen of Trp169 and the peptide carbonyl group at position 6 (Fig. 4). In addition, the methylene groups of the lysine side chain are tightly packed against the hydrophobic surface of the side chain of Trp169 (Fig. 4), with carbon–carbon distances of ~3.5 Å.

The lysine–carboxylate interaction is rarely seen in protein structures

The formation of simultaneous hydrogen bonds between a lysine side chain and three carboxylate groups appeared to us to be an uncommon interaction, and so a systematic search of the entire protein databank was carried out for other instances of this interaction. For the purposes of this search, the 'Lys/3O' motif was defined as a cluster formed by the terminal nitrogen atom from a lysine side chain (the Nζ atom), and three oxygen atoms from the carboxylate groups of different aspartate or glutamate side chains. The only geometric constraint placed on the search was that the three distances between the Nζ and oxygen atoms be <3.7 Å. All instances of this motif in all 2943 release and pre-release coordinate sets deposited in the Brookhaven Protein Databank [27] on November 1, 1994, were found by brute-force search. No attempt was made to generate multi-subunit complexes from the symmetry operations given for some coordinate sets, although inter-subunit interactions within an asymmetric unit are included. The proteins with the Lys/3O motif were organized into families such that any two proteins from the same and different families had greater than or less than 30% sequence identity, respectively. Only one representative protein structure from each family was retained for further analysis.

This search of the entire contents of the protein databank resulted in the identification of only 10 unique protein structures in which a lysine amino nitrogen has hydrogen-bonding interactions with three carboxyl oxygens, confirming that this motif is indeed extremely rare. The 10 structures identified by this search are: glycogen phosphorylase (for example, protein databank entry 1abb [28]), HIV reverse transcriptase (entry 3hvt [29]), aspartate aminotransferase (entry 1asn [30]), mengo virus coat protein (entry 2mev [31]), chloromuconate cycloisomerase or muconate lactonizing enzyme (entry 1chr [32]), the nitrogenase Mo–Fe protein (entry 1min [33]), chorismate mutase (entry 2cht [34]), xylose isomerase mutant (entry 1xyl [35]), enolase (entry 4enl [36]) and actinidin (entry 1aec [37]). A detailed examination of these 10 structures revealed that in many cases the lysine is part of a network of basic and acidic residues that does not quite resemble the Lys-Glu-Asp interaction seen in the SH3 domain. In two others, there are clusters that resemble the SH3 Lys/3O motif (3hvt and 2cht), but they are in regions of the protein with high temperature factors. The cluster in actinidin does resemble the SH3 case closely, but here the lysine (Lys17) has four potential ligands, including three glutamates and one serine, and an additional lysine coordinates one of the glutamates [37].

It is not surprising to find that protein structures avoid bringing together three carboxyl groups without more than one compensating basic residue nearby. Indeed, the closest analogy to the SH3–lysine interaction may be the structure of a mutant xylose isomerase, with the lysine replacing the magnesium ion at a metal-binding site of the enzyme [35] (Fig. 6). In an interesting protein engineering experiment, a glutamate residue that is required for metal coordination in xylose isomerase was mutated to lysine, in order to abolish metal binding. Little structural change is induced in the protein as a consequence of this mutation, and the amino group of the lysine replaces the magnesium ion and coordinates two aspartate side chains and one glutamate side chain [35]. Although there is striking similarity between this interaction and that observed in the SH3–C3G complex, it is interesting to note that the orientation of the lysine residue in the mutant xylose

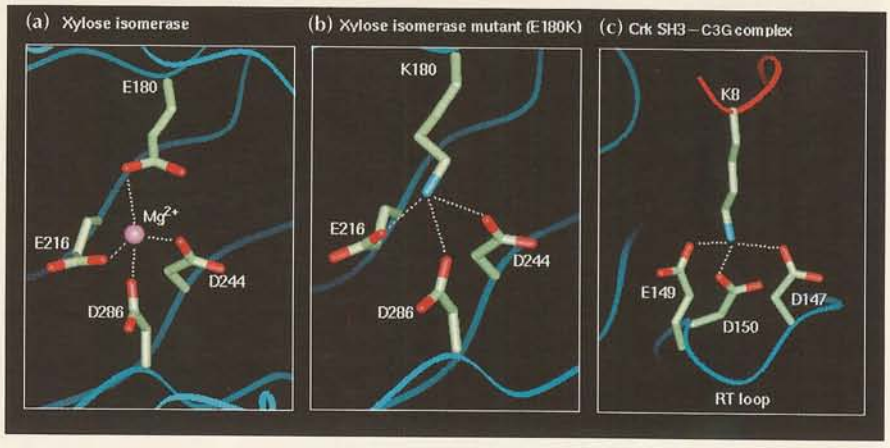

Fig. 6. Comparison of the lysine-binding site in Crk to a metal-binding site in xylose isomerase. In each of the panels, residues of interest are shown as stick figures, the protein backbone is shown as a blue ribbon, and hydrogen bonds and metal coordination are indicated by broken lines. **(a)** A Mg^{2+}-binding site in wild-type xylose isomerase (protein databank entry, 1xya, [35]). **(b)** The lysine-specific interactions in a mutant xylose isomerase are shown, in which the Glu180 was changed to a lysine (entry 1xyl, [35]). Both structures of xylose isomerase were determined at 1.8 Å resolution. **(c)** The lysine-specific interactions in the Crk/C3G complex. The red ribbon represents the peptide backbone and the blue ribbon represents the SH3 backbone. Note the extended conformation of the lysine and the fact that it approaches the triad of carboxylate groups from directly above, thus allowing the formation of a more optimal hydrogen-bonding network in Crk/C3G than seen in the xylose isomerase mutant.

isomerase is not as optimal as in the SH3 domain, because it approaches the triad of carboxylate groups at an angle and is not in a fully extended conformation (Fig. 6).

Interaction with arginine at position P_{-3} in the Sos peptide

The binding mode of the Sos peptide is very similar to that seen with the C3G peptide (Fig. 2) and the arginine side chain at P_{-3} extends across the face of the Trp169 side chain and interacts with the acidic residues in the RT loop. However, a notable difference is that in contrast to lysine, the arginine side chain is not well ordered and does not adopt an optimal conformation for hydrogen-bonding interactions with the acidic groups or for hydrophobic interactions with the tryptophan side chain. One terminal nitrogen of the arginine side chain makes a strong hydrogen bond with Asp150 (nitrogen–oxygen distance of 2.4 Å). The second nitrogen interacts less optimally with Asp147 (nitrogen–oxygen distance of 3.1 Å), and the imino nitrogen is not involved in hydrogen-bonding interactions with the protein. Unlike the lysine in the C3G peptide, the arginine in Sos does not adopt a low energy extended conformation (the $\chi 3$ torsion angle is in a *gauche* conformation) and does not pack tightly against the tryptophan (the closest van der Waals contact is 4.2 Å, in contrast to 3.5 Å in C3G). Finally, electron density for the arginine side chain is relatively weak, indicating conformational flexibility.

The extent of conformational disorder in the two peptide complexes was estimated by carrying out a number of independent crystallographic refinements of the structures after the introduction of random perturbations in the atomic positions (Fig. 7) [38]. In order to make the comparisons meaningful, these refinements were carried out using data to 1.9 Å spacings for both complexes. Atomic positions in the RT loop and the P_{-3} position of the peptide were randomly displaced such that the rms deviation from the starting model was ~1.7 Å. Many perturbed structures were generated independently in this way, and subjected to crystallographic least-squares refinement. In order to increase the radius of convergence of the procedure, the refinements were initiated at 3.5 Å, with the resolution limit being gradually increased to 1.9 Å during the course of the refinement. Excessive shifts in water molecule positions were avoided by restraining the water oxygens to their original positions by a harmonic constant. A total of 1000 steps of conjugate-gradient optimization were carried out. Refinements that failed to converge to R-values within 0.3 percentile points of the original value were rejected. A set of 24 structures was generated in this way for each of the two peptide complexes. A more exhaustive search procedure would involve simulated annealing and multiple conformers for the protein [38], but this was not attempted.

Despite the relatively large random shifts introduced into the structures prior to refinement, the amino group of the lysine as well as the carboxylate groups of the three acidic side chains in the c-Crk/C3G complex

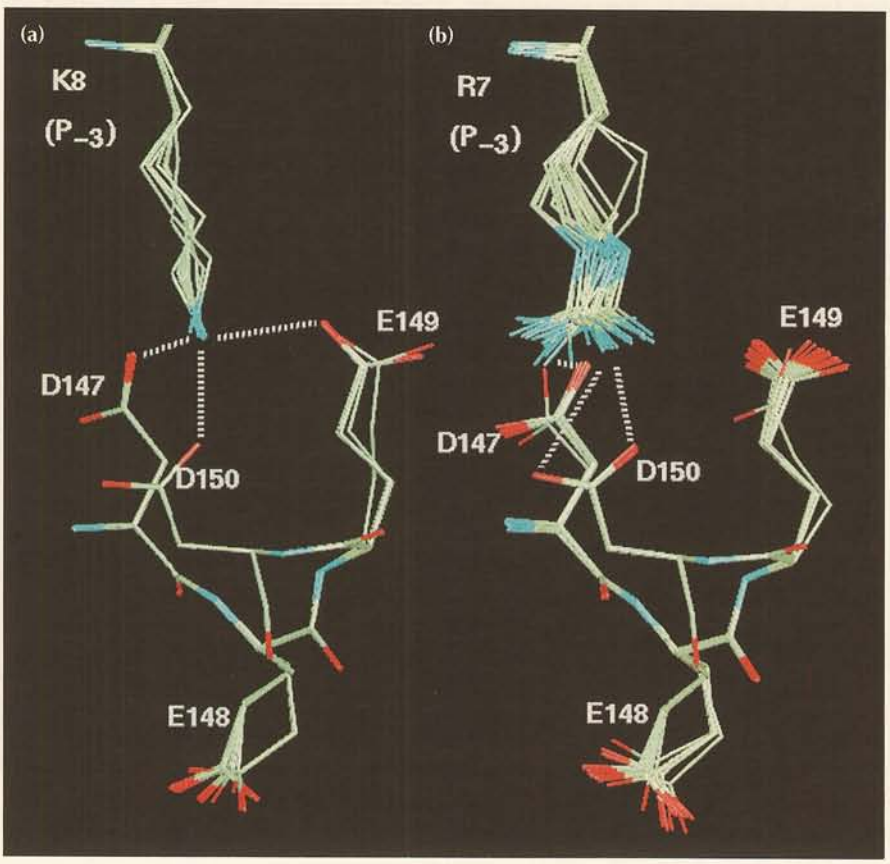

Fig. 7. Results of multiple refinements of the structures of **(a)** the c-Crk/C3G complex and **(b)** the c-Crk/Sos complex. To estimate the structural uncertainties associated with the lysine and arginine interactions, the coordinates of residues 147–150 (the RT loop) of Crk SH3 and Lys8 at P_{-3} in C3G peptide (Arg7 at P_{-3} in Sos peptide) of the final structure were repeatedly perturbed by random displacements, resulting in an rms deviation of 1.7 Å from the original model. The resulting structures were independently refined, using data to 1.9 Å for c-Crk/C3G and c-Crk/Sos. The results of 24 independent refinements for each structure are shown. All the structures shown have R-values that are comparable with the original structure (see text). Hydrogen bonds are indicated by dashed lines. Oxygen atoms are colored red, nitrogens, blue and carbons, white.

```
                          | ||
                bbbbbbb  RT loop        bbbbb     n-src loop      bbbbb distal bbbbb  3 10   bbbb
Crk-N        AEY VRALFDF NGNDEEDLPFKKGDI LRIRD KP                EEQ WWNAE DSE G  KRGMI PVPYV EKYR
v-Crk        VEY VRALFDF KGNDDGDLPFKKGDI LKIRD KP                EEQ WWNAE DMD G  KRGMI PVPYV EKC
Crkl         LEY VRTLYDF PGNDAEDLPFKKGEI LVIIE KP                EEQ WWSAR NKD G  RVGMI PVPYV EKL
Nck-N     AEEVV VVAKFDY VAQQEQELDIKKNER LWLLD D                  SKS WWRVR NSM N  KTGFV PSNYV ERKN SA
Lck          DNL VIALHSY EPSHDGDLGFEKGEQ LRILE Q                 SGE WWKAQ SLTTG  QEGFE PFNFV AKA
Fyn          VTL FVALYDY EARTGDDLSFHKGEK FQILN SS                EGD WWEAR SLTTG  ETGYI PSNYV APVD
Spec      DETGKEL VLALYDY QEKSPREVTMKKGDI LTLLN ST               NKD WWKVE VN  D  RQGFV PAAYV KKLD
Abl          L FVALYDF VASGDNTLSITKGEK LRVLG YNH                 NGE WCEAQ TKN    GQGWV PSNYI TPVN SL
Grb2-N       ME AIAKYDF KATADDELSFKRGDI LKVLN EEC                 DQN WYKAE LN  G  KDGFI PKNYI EMKP HP
PI3Kb     MSAEGYQ YRALYDY KKEREEDIDLHLGDI LTVNK GSLVALGFSDGQEARPEEIG WLNGY NETTG  ERGDF PGTYV EYIG RK
```

Fig. 8. Sequence alignment of several SH3 domains. The sequences of different SH3 domains are aligned based on the secondary structure. The top line provides the notations for secondary structure: 'b' for β-sheet and 3_{10} for 3_{10}-helix. The most highly conserved residues that interact with the ligand are in bold-faced type. Residues that are conserved in many SH3 domains and that interact with the ligand are underlined. The residues in italics represent those that are specific for c-Crk SH3 and closely related domains. The SH3 sequences shown are as follows. crk-N: mouse c-Crk, N-terminal SH3 domain (residues 134–191; Genbank accession s72408) [49]. v-Crk: viral Gag-Crk oncogene product from the CT10 virus (residues 370–426; Genbank accession Y00302) [1]. Crkl: human Crk-like protein, N-terminal SH3 domain (residues 125–181; Genbank accession X59656) [50]. Nck-N: human Nck protein, N-terminal SH3 domain (residues 2–62; Genbank accession X17576) [51]. Lck: human Lck tyrosine kinase (residues 51–119; Genbank accession M36881) [24]. Fyn: human Fyn tyrosine kinase (residues 84–142) [23]. Spec: chicken spectrin (residues 2–62) [22]. Abl: Abl tyrosine kinase (residues 84–141; Genbank accession U07563) [17]. Grb2-N: human Grb2 protein, N-terminal SH3 domain (residues 1–59; Genbank accession M96995) [52]. PI3Kb: p85 subunit of human phosphatidylinositol 3-OH kinase (residues 1–80) [16].

consistently return to their original positions and are thus quite precisely localized (Fig. 7a). In contrast to the tight clustering of all 24 structures in the C3G complex, the arginine side chain and the carboxylate groups are less well localized in the c-Crk/Sos complex (Fig. 7b). The disorder in the arginine side chain is probably underestimated by this method, as the electron density shows evidence for an alternative arginine conformation (not revealed by the limited radius of convergence of the refinement method) that makes fewer hydrogen bonds than those depicted in Fig. 7b.

These results demonstrate that the arginine side chain in the crystal structure does not hydrogen bond with the side chain of Glu149 (Fig. 7b). Most interestingly, in the SH3 domain of the viral oncogene product v-Crk, the residue at the corresponding position is a glycine (Fig. 8). This SH3 domain has a 13-fold lower binding affinity for the C3G peptide (K_d=25.5 μM) than does the c-Crk SH3-N domain (K_d=1.9 μM) and also does not discriminate between lysine and arginine at the P_{-3} position (Table 1). The comparison of the binding modes of lysine and arginine (Fig. 7) explains this effect very nicely, as the loss of Glu149 in v-Crk removes a side chain that interacts with lysine, but not with arginine. Likewise, the loss of Glu148 in Crkl (Fig. 8), which does discriminate between lysine and arginine [10], does not affect peptide binding because this side chain points away from the binding site (Fig. 7).

Why does the lysine-containing C3G peptide bind so much more strongly to the c-Crk SH3 domain than to other SH3 domains such as Grb2? Ionic interactions do not often make a major contribution to protein stability, and are thought to be destabilizing in most situations, in contrast to hydrophobic packing interactions [39]. However, the involvement of charged side chains in protein complexes can assure specificity in the interaction, because the energetic penalty for desolvating the ions has to be compensated for, at least partially, through ion-pairing and hydrogen bonding. In comparing the affinity of the C3G peptide for c-Crk and other SH3 domains, the extent to which the SH3 domain can compensate for the desolvation of lysine *versus* arginine has to be considered. The difference in binding affinity between the C3G peptide and a variant that contains a single Lys→Arg mutation corresponds to ~1.3 kcal mol^{-1} (Table 1), which is roughly equivalent to the loss of a hydrogen bond [40]. For c-Crk, we observe three strong hydrogen bonds for lysine in the C3G peptide, and two or three less optimal hydrogen bonds for arginine in the Sos peptide (Fig. 7). Analysis of an extensive list of SH3 sequences shows that the presence of three acidic residues in the RT loop that can form the lysine-specific interaction is unique to c-Crk and its close relatives (see Fig. 8 for an alignment of a limited set of sequences). In the case of arginine binding to SH3 domains with one or two acidic side chains in the RT loop, the guanidinium group of the arginine is able to make two or three hydrogen-bonding interactions with these side chains [21]. Lysine, on the other hand, cannot make multiple hydrogen-bonding interactions unless the carboxylate groups approach each other closely and in a relatively precise orientation, as seen in the c-Crk SH3 domain structure.

Biological implications

Src homology 3 (SH3) and SH2 domains are small modular units that are important for the generation of protein–protein interactions in cellular signal transduction pathways. SH3 domains bind to proline-rich peptides. Binding is mediated by a relatively small interaction surface, and without the benefit of the multiple hydrogen bonds that anchor phosphotyrosine side chains to SH2 domains. SH3–peptide interactions are therefore

of low affinity (~1 μM in the best cases). The conserved framework of side chains on the SH3 surface allows for efficient packing with polyproline type II helices, but the reliance on hydrophobic interactions, which are generally less specific than those involving hydrogen bond formation, results in SH3–peptide recognition being promiscuous. Ion-pairing interactions between the peptide and the SH3 domain have been shown previously to select one of two possible orientations of a peptide, with corresponding restrictions on the placement of non-proline residues [18–21]. However, in the general case where arginine residues in the peptide are used to form the ion pairs, the presence of acidic patches on the surfaces of a large set of SH3 domains results in the binding of particular peptides, such as those from the guanine nucleotide exchange factor Sos, to a large number of SH3 domains with comparable affinity.

For many SH3 domains it may be the case that biological specificity is only achieved by exploiting multiple binding sites and by tertiary interactions with the parent protein and its target. However, proline-rich peptides from another guanine nucleotide exchange factor, C3G, are able to distinguish the c-Crk SH3 domain from a number of others by using a lysine rather than an arginine at the key P_{-3} position in the polyproline helix. The lysine-specific binding site seen in the c-Crk SH3 domain described here is extremely rare in protein structures, and it may have evolved in the SH3 domain because the limited sites of variability on the proline-rich peptide make it advantageous to discriminate between lysine and arginine while exploiting electrostatic complementarity. As specific interactions between other SH3 domains and target peptides are discovered, subtle yet precise differences in their interactions may again turn out to be crucial for discriminating between SH3-binding sites.

Materials and methods

Expression and purification

The DNA sequence coding for the N-terminal SH3 domain of c-Crk was amplified using the polymerase chain reaction (PCR) with a mouse cDNA clone as a template. The amplified fragment encodes 58 residues (134–191) of the mouse c-Crk protein plus an initial methionine. This fragment was cloned into the *Escherichia coli* expression vector pET3a (Novagen). Cultures of *E. coli* BL21(DE3) strains, transformed by the resulting plasmid, were grown at 37°C in LB medium supplemented with 100 μg ml^{-1} ampicillin [41]. After 3 h of induction with isopropyl-thio-β-D-galactoside (IPTG) at 37°C, the cell pellet from 1 l of culture was resuspended in 100 ml of standard buffer (20 mM Tris, pH 8.0, 2 mM dithiothreitol, 0.5 mM EDTA) supplemented with 50 μg ml^{-1} leupeptin (Boehringer-Mannheim, Germany) and 1% (v/v) aprotinin (Sigma). Cells were lysed by sonication on ice and the suspension was centrifuged for 20 min at 15000 rpm at 4°C in a Beckmann JA17 rotor, and the clear supernatant was recovered. Protein purification proceeded along standard lines using a Q-Sepharose Fast Flow column (Pharmacia), a mono-Q HR 10/10 column (Pharmacia) and a Superdex75 HiLoad 16/60 column (Pharmacia). The protein was concentrated to 85 mg ml^{-1} using ultrafiltration (Amicon). SDS-PAGE electrophoresis using Coomassie blue staining reveals a single band at the expected molecular weight (7 kDa).

Crystallization with peptides

The two peptides used in this work, C3G (PPPALPPKKR) and Sos (PPPVPPRRRR) were synthesized as part of a project to determine the peptide binding specificity of the c-Crk SH3 domain [10] (B Knudsen *et al.*, unpublished data), and made available to us for crystallization. Peptides were synthesized by the Peptide Technology Group at Amgen Boulder Inc., Boulder, CO, using conventional technology. The peptides were purified by high-performance liquid chromatography and lyophilized prior to resuspension in buffer for crystallization, and their identities were confirmed by mass spectroscopy. Crystallization of the c-Crk SH3-N/peptide complexes was performed using the hanging drop technique at 21°C. The best crystals were obtained by vapor diffusion against a reservoir solution containing 0.1 M sodium acetate pH 4.6, 0.2 M ammonium acetate, 28–32% polyethylene glycol (PEG) 4K. The c-Crk SH3-N protein solution (1 μl at 85 mg ml^{-1}, 12 mM) was first mixed with 1.5 μl of C3G or Sos peptide solution (10 mg ml^{-1}, 9.0 mM, in water) and then mixed with 1.0 μl of reservoir solution. Crystals were reproducibly obtained over a period of several days, with final dimensions of 2.0×0.2×0.2 mm^3 in the best cases. Interestingly, all attempts to obtain crystals in the absence of high-affinity peptides were unsuccessful.

Crystals of c-Crk SH3 in complex with either peptide were obtained in the same crystal system (tetragonal $P4_1$; a=47.2 Å, c=29.4Å for the C3G complex and a=47.4 Å, c=29.5 Å for the Sos complex). X-ray diffraction data were measured using a Rigaku R-AXIS IIC imaging plate area detector, mounted on a Rigaku RU200 rotating anode X-ray generator. Each data set was collected from a single crystal flash-frozen at −160°C. The crystals were cryo-protected by transferring them to a solution containing 0.1 M sodium acetate, pH 4.6, 0.2 M ammonium acetate, 35% PEG 4K supplemented with 10% glycerol. Data for c-CrkSH3/C3G were collected using a crystal-to-detector distance of 60 mm and exposure times of 10 min for 2° oscillations (for data to 1.50 Å spacings) and a total of 80 frames was collected. Data for the Sos complex were collected with a crystal-to-detector distance of 100 mm (for data to 1.90 Å spacings), exposures of 60 min and 3° oscillations. Data processing and reduction were carried out with the HKL, DENZO and SCALEPACK programs (Z Otwinowski and W Minor, unpublished programs). Statistics for the final data sets are given in Table 3.

Structure determination and refinement

The structure of the c-Crk/C3G complex was determined by molecular replacement [42] using the X-PLOR program [43]. The crystal structures of two SH3 domains, the Lck SH3 domain [24] and the Nck N-terminal SH3 domain (X Wu and J Kuriyan, unpublished data), were independently used as search models. No model for bound peptide was included in the molecular replacement calculations. Rotation function searches followed by Patterson correlation refinement [42] gave similar results with either the Lck or the Nck SH3 domain as the model. The highest peak in the rotation search corresponded to an orientation of the model that gave a strong

Table 3. Statistics for the data collection and refinement.

	c-CrkSH3/C3G	c-CrkSH3/Sos
Data collection statistics		
Space group	$P4_1$	$P4_1$
Unit cell	a = 47.2 Å, c = 29.4 Å	a = 47.4 Å, c = 29.5 Å
Resolution (Å)	20–1.50	20–1.90
No. of observed reflections	66868	50442
No. of unique reflections	10512	5289
Completeness of data to 1.90 Å	89.1%	95.5%
Completeness of data between 1.90–1.50 Å	73.8%	–
$I/\sigma(I)$ in all shells	29.6	21.9
$I/\sigma(I)$ in outer shell	16.6	5.4
R_{sym}	4.0%	9.4%
Refinement statistics		
R_{cryst} [$\|F\|>2\sigma(F)$]	17.4% (6.0–1.50 Å)	18.3% (6.0–1.90 Å)
No. of unique reflections [$\|F\|>2\sigma(F)$]	7667	4616
No. of molecules in asymmetric unit	1	1
No. of non-hydrogen atoms in final model	664	608
No. of water molecules in final model	119	68
Rms deviation of bond lengths (Å)	0.012	0.011
Rms deviation of bond angles (°)	1.9	1.9

signal in translation functions calculated using $P4_1$ (but not $P4_3$) as the space group. Consistent results were obtained using various resolution ranges. The peak in the translation function was 10σ above the mean value when using data from 8.0–2.5 Å, and the next highest peak was below 5σ. This solution was used for further structural refinement.

The structure was refined following standard procedures using X-PLOR [44] and O [45]. Charges for the side-chain atoms of lysine, arginine, glutamate and aspartate were set to zero, and thus the refinement procedure is not expected to introduce any bias towards ion-pairing interactions. Rigid-body refinement in X-PLOR of the translation-function solution reduced the R-value from 50.2% to 49.7% (all data with $|F|>2\sigma$, 8.0–2.5 Å). This model was then subjected to 200 cycles of conjugate-gradient optimization, which reduced the R-value to 29.7% (8.0–2.5 Å). Electron-density maps calculated at this stage were interpretable in terms of the sequence of c-Crk SH3-N, and strong density corresponding to the bound peptide was visible. In particular, very strong density was observed for the side chain of Lys8 in the peptide, as well as for Leu5 and Pro2. The unambiguous placement of the lysine and leucine side chains made it obvious that the peptide orientation was reversed relative to that seen previously in the p85, Abl and Fyn SH3 domains [16,17], and a model for peptide residues 1–8 was built into the density. The atomic model for Lck SH3 [24] was modified to correspond to the sequence of c-Crk SH3-N, and rebuilt based on the electron-density map. Several cycles of model building and crystallographic refinement, including simulated annealing, resulted in a final model that has an R-value of 17.4% (for data between 6.0 Å and 1.5 Å resolution) with very good stereochemistry (see Table 3). The model contains 57 residues of c-Crk SH3-N (residues 134–190), nine residues of the peptide (no electron density is seen for the terminal arginine) and 119 water molecules.

The refined structure of the c-Crk SH3-N domain (with peptide removed) was used to initiate refinement against data for the c-Crk SH3-N/Sos complex, as the two crystal forms are isomorphous. During the initial rigid-body refinement, the R-value dropped significantly from 51.8% to 33.6%, using data from 6.0 Å to 1.9 Å spacings. Positional refinement reduced the R-value further, to 26.8%. An ($|F_o|-|F_c|$) electron-density map, with model phases, revealed the presence of bound peptide. In contrast to the C3G complex, where the density for the side chain of Lys8 was the strongest feature in the electron-density map, the arginine side chain in the Sos peptide does not have particularly strong density. Nevertheless, the placement of the peptide residues was unambiguous. Straightforward model building and refinement yielded the final model which contains 57 residues of c-Crk SH3-N (residues 134–190), eight residues of Sos (residues 1–8; no electron density is present for the last two arginine residues) and 68 water molecules. The Arg8 side chain of Sos is not visible in electron-density maps, and so an alanine residue was built at that position. The final R-value is 18.3 % (using data from 6.0–1.9 Å, see Table 3).

The atomic coordinates and structure factors have been deposited in the Brookhaven protein databank. Coordinates are also available directly from the authors by e-mail (wux@rockefeller.edu).

Acknowledgements: We thank John Mayer, Sarah K Burrell and Thomas J Zamborelli of Amgen, Inc. for providing the peptides used in this study, and Stephen K Burley, Joseph L Kim and David Wilson for useful suggestions. A study describing the binding affinities of proline-rich peptides, referred to in the text as B Knudsen *et al.*, has been submitted for publication: Knudsen, B., Zheng, J., Feller, S.M., Mayer, J.P., Burrell, S.K., Cowburn, D. & Hanafusa, H. (1995). Affinity and specificity requirements for the first Src homology 3 domain of the Crk protein. Technical assistance with data collection provided by Ramakoti Suresh is gratefully acknowledged. This work was partially supported by grants from the National Institutes of Health (GM-47021 to DC and CA-44356 to HH).

References

1. Mayer, B.J., Hamaguchi, M. & Hanafusa, H. (1988). A novel viral oncogene with structural similarity to phospholipase C. *Nature* **332**, 272–275.
2. Pawson, T. & Schlessinger, J. (1993). SH2 and SH3 domains. *Curr. Biol.* **3**, 434–442.
3. Mayer, B.J. & Hanafusa, H. (1990). Mutagenic analysis of the v-crk oncogene: requirement for SH2 and SH3 domains and correlation between increased cellular phosphotyrosine and transformation. *J. Virol.* **64**, 3581–3589.
4. Koch, C.A., Anderson, D., Moran, M.F., Ellis, C. & Pawson, T. (1991). SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science* **252**, 668–674.
5. Kuriyan, J. & Cowburn, D. (1993). Structures of SH2 and SH3 domains. *Curr. Opin. Struct. Biol.* **3**, 828–837.
6. Feller, S.M., Ren, R., Hanafusa, H. & Baltimore, D. (1994). SH2 and SH3 domains as molecular adhesives: the interactions of Crk and Abl. *Trends Biochem. Sci.* **19**, 453–459.
7. Reichmann, C.T., Mayer, B.J., Keshav, S. & Hanafusa, H. (1992). The product of the cellular Crk gene consists primarily of SH2 and SH3 regions. *Cell Growth Differentiation* **3**, 451–460.
8. Ren, R., Mayer, B.J., Cicchetti, P. & Baltimore, D. (1993). Identification of a ten-amino acid proline-rich SH3 binding site. *Science* **259**, 1157–1161.
9. Tanaka, S., *et al.*, & Matsuda, M. (1994). C3G, a guanine nucleotide-releasing protein expressed ubiquitously, binds to the Src homology 3 domains of CRK and GRB2/ASH proteins. *Proc. Natl. Acad. Sci. USA* **91**, 3443–3447.
10. Knudsen, B.S., Feller, S.M. & Hanafusa, H. (1994). Four proline-rich sequences of the guanine-nucleotide exchange factor, C3G, bind with unique specificity to the first Src homology 3 domain of Crk. *J. Biol. Chem.* **269**, 32781–32787.

11. Feller, S.M., Knudsen, B. & Hanafusa, H. (1994). c-Abl kinase regulates the protein binding activity of c-Crk. *EMBO J.* **13**, 2341–2351.
12. Ren, R., Ye, Z.-S., & Baltimore, D. (1994). Abl protein-tyrosine kinase selects the Crk adapter as a substrate using SH3-binding sites. *Genes Dev.* **8**, 783–795.
13. Saraste, M. & Musacchio, A. (1994). Backwards and forwards binding. *Nat. Struct. Biol.* **1**, 835–837.
14. Cicchetti, P., Mayer, B.J., Theil, G. & Baltimore, D. (1992). Identification of a protein that binds to the SH3 region of Abl and is similar to Bcr and GAP-rho. *Science* **257**, 803–806.
15. Lim, W.A. & Richards, F.M. (1994). Critical residues in an SH3 domain from Sem-5 suggest a mechanism for proline-rich peptide recognition. *Nat. Struct. Biol.* **1**, 221–225.
16. Yu, H., Chen, J.K., Feng, S., Dalgarno, D.C., Brauer, A.W. & Schreiber, S.L. (1994). Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76**, 933–945.
17. Musacchio, A., Saraste, M. & Wilmanns, M. (1994). High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. *Nat. Struct. Biol.* **1**, 546–551.
18. Feng, S., Chen, J.K., Yu, H., Simon, J.A. & Schreiber, S.A. (1994). Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3–ligand interactions. *Science* **266**, 1241–1247.
19. Goudreau, N., *et al.*, & Roques, B.P. (1994). NMR structure of the N-terminal SH3 domain of GRB2 and its complex with a proline rich peptide from Sos. *Nat. Struct. Biol.* **1**, 898–907.
20. Terasawa, H., *et al.*, & Inagaki, F. (1994). Structure of the N-terminal SH3 domain of GRB2 complexed with a peptide from the guanine nucleotide releasing factor Sos. *Nat. Struct. Biol.* **1**, 891–897.
21. Lim, W.A., Richards, F.M. & Fox, R.O. (1994). Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature* **372**, 375–379.
22. Musacchio, A., Noble, M., Pauptit, R., Wierenga, R. & Saraste, M. (1992). Crystal structure of a Src-homology 3 (SH3) domain. *Nature* **359**, 851–855.
23. Noble, M.E.M., Musacchio, A., Saraste, M., Courtneidge, S.A. & Wierenga, R.K. (1993). Crystal structure of the SH3 domain in human Fyn; comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin. *EMBO J.* **12**, 2617–2624.
24. Eck, M., Atwell, S.K., Shoelson, S.E. & Harrison, S.C. (1994). Crystal structure of the regulatory domains of the Src-family tyrosine kinase lck. *Nature* **368**, 764–769.
25. Taylor, R. & Kennard, O. (1984). Hydrogen-bond geometry in organic crystals. *Accounts Chem. Res.* **17**, 320–326.
26. Ippolito, J.A., Alexander, R.S. & Christianson, D.W. (1990). Hydrogen bond stereochemistry in protein structure and function. *J. Mol. Biol.* **215**, 457–471.
27. Bernstein, F., *et al.*, & Tasumi, M. (1978). The protein data bank: a computer-based archival file for macromolecular structures. *Arch. Biochem. Biophys.* **185**, 584–591.
28. Leonidas, D.D., Oikonomakos, N.G., Papageorgiou, A.C., Acharya, K.R., Barford, D. & Johnson, L.N. (1992). Control of phosphorylase B conformation by a modified cofactor: crystallographic studies on R-state glycogen phosphorylase reconstituted with pyridoxal 5′-diphosphate. *Protein Sci.* **1**, 1112–1122.
29. Smerdon, S.J., *et al.*, & Steitz, T.A. (1994). Structure of the binding site for nonnucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type I. *Proc. Natl. Acad. Sci. USA* **92**, 3911–3915.
30. Jager, J., Moser, M., Sauder, U. & Jansonius, J.N. (1994). Crystal structures of *Escherichia coli* aspartate aminotransferase in two conformations. Comparison of an unliganded open and two liganded closed forms. *J. Mol. Biol.* **239**, 285–305.
31. Krishnaswamy, S. & Rossmann, M.G. (1990). Structural refinement and analysis of Mengo virus. *J. Mol. Biol.* **211**, 803–844.
32. Goldman, A., Ollis, D.L. & Steitz, T.A. (1987). Crystal structure of muconate lactonizing enzyme at 3 Å resolution. *J. Mol. Biol.* **194**, 143–153.
33. Chan, M.K., Kim, J. & Rees, D.C. (1993). The nitrogenase FeMo-cofactor and P-cluster pair: 2.2 Å resolution structures. *Science* **260**, 792–794.
34. Chook, Y.M., Ke, H. & Lipscomb, W.N. (1993). Crystal structures of the monofunctional chorismate mutase from *Bacillus subtilis* and its complex with a transition state analog. *Proc. Natl. Acad. Sci. USA* **90**, 8600–8603.
35. Allen, K.N., *et al.*, & Ringe, D. (1994). The role of the divalent metal ion in sugar binding, ring opening and isomerization by D-xylose isomerase: replacement of a catalytic metal by an amino-acid. *Biochemistry* **33**, 1488–1494.
36. Lebioda, L. & Stec, B. (1991). Mechanism of enolase: the crystal structure of enolase-Mg^{2+}-phosphoglycerate/phosphoenolpyruvate complex at 2.2 Å resolution. *Biochemistry* **30**, 2817–2822.
37. Baker, E.N. (1980). Structure of actinidin, after refinement at 1.7 Å resolution. *J. Mol. Biol.* **141**, 441–484.
38. Kuriyan, J., Osapay, K., Burley, S.K., Brünger, A.T., Hendrickson, W.A. & Karplus, M. (1991). Probing disorder in high resolution protein structures by simulated annealing. *Proteins* **10**, 340–358.
39. Tidor, B. & Karplus, M. (1991). Simulation analysis of the stability mutant R96H of T4 lysozyme. *Biochemistry* **30**, 3217–3228.
40. Fersht, A.R., Leatherbarrow, R.J. & Wells, T.N.C. (1986). Structure and activity of the tyrosyl-tRNA synthetase: the hydrogen bond in catalysis and specificity. *Philos. Trans. R. Soc. Lond. A* **317**, 305–320.
41. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular Cloning.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
42. Brünger, A.T. (1990). Extension of molecular replacement: a new search strategy based on Patterson correlation refinement. *Acta Crystallogr. A* **46**, 46–57.
43. Brünger, A.T. (1988). *X-PLOR Manual.* Howard Hughes Medical Institute and Dept. of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT.
44. Brünger, A.T., Krukowski, A. & Erickson, J.W. (1990). Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr. A* **46**, 585–593.
45. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119.
46. Ramachandran, G.N. & Sasisekharan, V. (1968). Conformations of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283–437.
47. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.
48. Nicholls, A., Sharp, K.A., & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296.
49. Ogawa, S., *et al.*, & Hirai, H. (1994). The C-terminal SH3 domain of the mouse c-Crk protein negatively regulates tyrosine-phosphorylation of Crk-associated p130 in rat 3Y1 cells. *Oncogene* **9**, 1669–1678.
50. ten Hoeve, J., Morris. C., Heisterkamp, N. & Groffen, J. (1993). Isolation and chromosomal localization of CRKL, a human crk-like gene. *Oncogene* **8**, 2469–2474.
51. Lehmann, J.M., Reithmuller, G. & Johnson, J.P. (1990). Nck, a melanoma cDNA encoding a cytoplasmic protein consisting of the src homology units SH2 and SH3. *Nucleic Acids Res.* **18** 1048.
52. Lowenstein, E.J., *et al.*, & Schlessinger, J. (1992). The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. *Cell* **70**, 431–442.

Received: **13 Dec 1994**; revisions requested: **9 Jan 1995**; revisions received: **16 Jan 1995**. Accepted: **17 Jan 1995**.